Convex Optimization

Euclidean Distance

Geometry
Dattorro

convex optimization

euclidean distance

gometry

Meboo
Convex Optimization
†
Euclidean Distance Geometry $2\varepsilon$

Mεβoo Publishing


This is version 2015.04.11: available in print at Lulu.com, as conceived, in color.

cybersearch:

I. semidefinite program
II. rank constraint
III. cardinality minimization
IV. audio signal optimization
V. convex geometry
VI. convex cones
VII. distance matrix

programs by Matlab

typesetting by WinEdt

with donations from MATHWORKS, SIAM, and AMS.

This searchable electronic color pdfBook is click-navigable within the text by page, section, subsection, chapter, theorem, example, definition, cross reference, citation, equation, figure, table, and hyperlink. A pdfBook has no electronic copy protection and can be read and printed by most computers. The publisher hereby grants the right to reproduce this work in any format but limited to personal use.

© 2005-2015 Meboo Publishing USA
for Jennie Columba

Antonio

& Sze Wan
EDM = S_h \cap \left( S_c^\perp - S_+ \right)
Prelude

The constant demands of my department and university and the ever increasing work needed to obtain funding have stolen much of my precious thinking time, and I sometimes yearn for the halcyon days of Bell Labs.

—Steven Chu, Nobel laureate [84]

Convex Analysis is an emerging calculus of inequalities while Convex Optimization is its application. Analysis is inherently the domain of a mathematician while Optimization belongs to the engineer. A convex optimization problem is conventionally regarded as minimization of a convex objective function subject to an artificial convex domain imposed upon it by the problem constraints. The constraints comprise equalities and inequalities of convex functions whose simultaneous solution set generally constitutes the imposed convex domain: called feasible set.

It is easy to minimize a convex function over any convex subset of its domain because any local minimum must be a global minimum. But it is difficult to find the maximum of a convex function over some convex domain because there can be many local maxima; although this has practical application (Eternity II §4.6.0.0.15, §C.5), it is not a convex problem. Tremendous benefit accrues when a mathematical problem can be transformed to an equivalent convex optimization, primarily because any locally optimal solution is then guaranteed globally optimal. An optimal solution is a best solution to the problem posed; a certificate can be obtained guaranteeing that no better solution exists.

---

[0.1] Solving a nonlinear system, for example, by instead solving an equivalent convex optimization problem is therefore highly preferable and what motivates geometric programming; a form of convex optimization invented in 1960s [62] [82] that has driven great advances in the electronic circuit design industry. [35, §4.7] [258] [415] [418] [106] [190] [199] [200] [201] [202] [203] [274] [275] [320]
Recognizing a problem as convex is an acquired skill; that being, to know when an objective function is convex and when constraints specify a convex feasible set. The challenge, which is indeed an art, is how to express difficult problems in a convex way: perhaps, problems previously believed nonconvex. Practitioners in the art of Convex Optimization engage themselves with discovery of which hard problems can be transformed into convex equivalents; because, once convex form of a problem is found, then a globally optimal solution is close at hand — the hard work is finished: Finding convex expression of a problem is itself, in a very real sense, its solution.

Yet, that skill acquired by understanding the geometry and application of Convex Optimization will remain more an art for some time to come; the reason being, there is generally no unique transformation of a given problem to its convex equivalent. This means, two researchers pondering the same problem are likely to formulate a convex equivalent differently; hence, one solution is likely different from the other although any convex combination of those two solutions remains optimal. Any presumption of only one right or correct solution becomes nebulous. Study of equivalence & sameness, uniqueness, and duality therefore pervade study of Optimization.

It can be difficult for the engineer to apply convex theory without an understanding of Analysis. These pages comprise my journal over a ten year period bridging gaps between engineer and mathematician; they constitute a translation, unification, and cohering of about four hundred papers, books, and reports from several different fields of mathematics and engineering. Although beacons of historical accomplishment are cited throughout, much of what is written here will not be found elsewhere. Care to detail, clarity, accuracy, consistency, and typography accompanies removal of ambiguity and verbosity out of respect for the reader. But the book is nonlinear in its presentation. Consequently there is much indexing, cross referencing, linkage to online sources, and background material provided in the text, footnotes, and appendices so as to be more self-contained and to provide understanding of fundamental concepts.

Looking toward the future, there remains much to be done in the area of machine computation if mathematical Optimization is to become fully embraced by the signal processing community. Wordlength of contemporary computers and numerical burdens upon them prohibit real time solution and accuracy sufficient to embed optimization problems within a recursive mathematical setting. When optimization problems constitute only intermediate solution to much larger problems, acquiring only a “few digits” accuracy can throw off subsequent dependent calculations. Barrier methods of solution are the principle obstacle to accuracy while simplex methods are the principle setback to speed. Novel, not hybrid, methods of solution are needed.

— Jon Dattorro
Stanford, California
2015
Convex Optimization
Euclidean Distance Geometry

In my career, I found that the best people are the ones that really understand the content, and they’re a pain in the butt to manage. But you put up with it because they’re so great at the content. And that’s what makes great products; it’s not process, it’s content.

—Steve Jobs, 1995

1 Overview

2 Convex geometry
  2.1 Convex set
  2.2 Vectorized-matrix inner product
  2.3 Hulls
  2.4 Halfspace, Hyperplane
  2.5 Subspace representations
  2.6 Extreme, Exposed
  2.7 Cones
  2.8 Cone boundary
  2.9 Positive semidefinite (PSD) cone
  2.10 Conic independence (c.i.)
  2.11 When extreme means exposed
  2.12 Convex polyhedra
  2.13 Dual cone & generalized inequality

3 Geometry of convex functions
  3.1 Convex function
  3.2 Practical norm functions, absolute value
  3.3 Inverted functions and roots
  3.4 Affine function
  3.5 Epigraph, Sublevel set
  3.6 Gradient
  3.7 Convex matrix-valued function
  3.8 Quasiconvex
  3.9 Salient properties
# List of Figures

## 1 Overview

1. Sigma delta quantizer ........................................... 22
2. *Orion nebula* .................................................. 23
3. Application of trilateration is localization .................. 24
4. Molecular conformation ......................................... 25
5. Face recognition ................................................. 26
6. *Swiss roll* ..................................................... 27
7. USA map reconstruction ......................................... 28
8. Honeycomb, Hexabenzocoronene molecule ..................... 29
9. Robotic vehicles ................................................ 30
10. Reconstruction of David ........................................ 32
11. David by distance geometry .................................... 32

## 2 Convex geometry

12. Slab .................................................................. 35
13. Open, closed, convex sets ....................................... 37
14. Intersection of line with boundary ............................ 39
15. Tangentials ........................................................ 41
16. Inverse image ..................................................... 44
17. Inverse image under linear map ................................ 45
18. *Tesseract* ........................................................ 47
19. Linear injective mapping of Euclidean body ................. 49
20. Linear noninjective mapping of Euclidean body ............ 50
21. Convex hull of a random list of points ....................... 54
22. Hulls .................................................................. 56
23. Two Fantopes ...................................................... 58
24. Nuclear Norm Ball ............................................... 60
25. Convex hull of rank-1 matrices ................................ 61
26. A simplicial cone ................................................ 64
27. Hyperplane illustrated $\partial H$ is a partially bounding line 65
28. Hyperplanes in $\mathbb{R}^2$ ........................................ 67
29. Affine independence ............................................. 70

13
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>( { z \in C \mid a^T z = \kappa_i } )</td>
<td>71</td>
</tr>
<tr>
<td>31</td>
<td>Hyperplane supporting closed set</td>
<td>72</td>
</tr>
<tr>
<td>32</td>
<td>Minimizing hyperplane over affine set in nonnegative orthant</td>
<td>79</td>
</tr>
<tr>
<td>33</td>
<td>Maximizing hyperplane over convex set</td>
<td>80</td>
</tr>
<tr>
<td>34</td>
<td>Closed convex set illustrating exposed and extreme points</td>
<td>83</td>
</tr>
<tr>
<td>35</td>
<td>Two-dimensional nonconvex cone</td>
<td>86</td>
</tr>
<tr>
<td>36</td>
<td>Nonconvex cone made from lines</td>
<td>86</td>
</tr>
<tr>
<td>37</td>
<td>Nonconvex cone is convex cone boundary</td>
<td>86</td>
</tr>
<tr>
<td>38</td>
<td>Union of convex cones is nonconvex cone</td>
<td>87</td>
</tr>
<tr>
<td>39</td>
<td>Truncated nonconvex cone ( \mathcal{X} )</td>
<td>87</td>
</tr>
<tr>
<td>40</td>
<td>Cone exterior is convex cone</td>
<td>87</td>
</tr>
<tr>
<td>41</td>
<td>Not a cone</td>
<td>88</td>
</tr>
<tr>
<td>42</td>
<td>Minimum element, Minimal element</td>
<td>91</td>
</tr>
<tr>
<td>43</td>
<td>( \mathcal{K} ) is a pointed polyhedral cone not full-dimensional</td>
<td>94</td>
</tr>
<tr>
<td>44</td>
<td>Exposed and extreme directions</td>
<td>98</td>
</tr>
<tr>
<td>45</td>
<td>Positive semidefinite cone</td>
<td>100</td>
</tr>
<tr>
<td>46</td>
<td>Convex Schur-form set</td>
<td>102</td>
</tr>
<tr>
<td>47</td>
<td>Projection of truncated PSD cone</td>
<td>104</td>
</tr>
<tr>
<td>48</td>
<td>Circular cone showing axis of revolution</td>
<td>113</td>
</tr>
<tr>
<td>49</td>
<td>Circular section</td>
<td>114</td>
</tr>
<tr>
<td>50</td>
<td>Polyhedral inscriptions</td>
<td>115</td>
</tr>
<tr>
<td>51</td>
<td>Conically (in)dependent vectors</td>
<td>122</td>
</tr>
<tr>
<td>52</td>
<td>Pointed six-faceted polyhedral cone and its dual</td>
<td>124</td>
</tr>
<tr>
<td>53</td>
<td>Minimal set of generators for halfspace about origin</td>
<td>125</td>
</tr>
<tr>
<td>54</td>
<td>Venn diagram for cones and polyhedra</td>
<td>129</td>
</tr>
<tr>
<td>55</td>
<td>Simplex</td>
<td>130</td>
</tr>
<tr>
<td>56</td>
<td>Two views of a simplicial cone and its dual</td>
<td>132</td>
</tr>
<tr>
<td>57</td>
<td>Two equivalent constructions of dual cone</td>
<td>136</td>
</tr>
<tr>
<td>58</td>
<td>Dual polyhedral cone construction by right angle</td>
<td>137</td>
</tr>
<tr>
<td>59</td>
<td>( \mathcal{K} ) is a halfspace about the origin</td>
<td>138</td>
</tr>
<tr>
<td>60</td>
<td>Iconic primal and dual objective functions</td>
<td>139</td>
</tr>
<tr>
<td>61</td>
<td>Orthogonal cones</td>
<td>142</td>
</tr>
<tr>
<td>62</td>
<td><em>Blades</em> ( \mathcal{K} ) and ( \mathcal{K}^* )</td>
<td>144</td>
</tr>
<tr>
<td>63</td>
<td>Membership w.r.t ( \mathcal{K} ) and orthant</td>
<td>152</td>
</tr>
<tr>
<td>64</td>
<td>Shrouded polyhedral cone</td>
<td>155</td>
</tr>
<tr>
<td>65</td>
<td>Simplicial cone ( \mathcal{K} ) in ( \mathbb{R}^2 ) and its dual</td>
<td>160</td>
</tr>
<tr>
<td>66</td>
<td>Monotone nonnegative cone ( \mathcal{K}_{M^+} ) and its dual</td>
<td>168</td>
</tr>
<tr>
<td>67</td>
<td>Monotone cone ( \mathcal{K}_M ) and its dual</td>
<td>170</td>
</tr>
<tr>
<td>68</td>
<td>Two views of monotone cone ( \mathcal{K}_M ) and its dual</td>
<td>171</td>
</tr>
<tr>
<td>69</td>
<td>First-order optimality condition</td>
<td>174</td>
</tr>
<tr>
<td>70</td>
<td>Geometry of convex functions</td>
<td>185</td>
</tr>
<tr>
<td>71</td>
<td>Convex functions having unique minimizer</td>
<td>187</td>
</tr>
<tr>
<td>72</td>
<td>Minimum/Minimal element, dual cone characterization</td>
<td>188</td>
</tr>
<tr>
<td>73</td>
<td>1-norm ball ( B_1 ) from compressed sensing/compressive sampling</td>
<td>192</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>Cardinality minimization, signed <em>versus</em> unsigned variable</td>
</tr>
<tr>
<td>74</td>
<td>1-norm variants</td>
</tr>
<tr>
<td>75</td>
<td>Affine function</td>
</tr>
<tr>
<td>76</td>
<td>Supremum of affine functions</td>
</tr>
<tr>
<td>77</td>
<td>Epigraph</td>
</tr>
<tr>
<td>78</td>
<td>Gradient in $\mathbb{R}^2$ evaluated on grid</td>
</tr>
<tr>
<td>79</td>
<td>Quadratic function convexity in terms of its gradient</td>
</tr>
<tr>
<td>80</td>
<td>Contour plot of convex real function at selected levels</td>
</tr>
<tr>
<td>81</td>
<td>Iconic quasiconvex function</td>
</tr>
<tr>
<td>82</td>
<td>Quasiconcave monotonic function $xu$</td>
</tr>
<tr>
<td>83</td>
<td>Arbitrary magnitude analog filter design</td>
</tr>
<tr>
<td>84</td>
<td>Venn diagram of convex program classes</td>
</tr>
<tr>
<td>85</td>
<td>Visualizing positive semidefinite cone in high dimension</td>
</tr>
<tr>
<td>86</td>
<td>Primal/Dual transformations</td>
</tr>
<tr>
<td>87</td>
<td>Projection <em>versus</em> convex iteration</td>
</tr>
<tr>
<td>88</td>
<td>Trace heuristic</td>
</tr>
<tr>
<td>89</td>
<td>Sensor-network localization</td>
</tr>
<tr>
<td>90</td>
<td>2-lattice of sensors and anchors for localization example</td>
</tr>
<tr>
<td>91</td>
<td>3-lattice of sensors and anchors for localization example</td>
</tr>
<tr>
<td>92</td>
<td>4-lattice of sensors and anchors for localization example</td>
</tr>
<tr>
<td>93</td>
<td>5-lattice of sensors and anchors for localization example</td>
</tr>
<tr>
<td>94</td>
<td>Uncertainty ellipsoids orientation and eccentricity</td>
</tr>
<tr>
<td>95</td>
<td>2-lattice localization solution</td>
</tr>
<tr>
<td>96</td>
<td>3-lattice localization solution</td>
</tr>
<tr>
<td>97</td>
<td>4-lattice localization solution</td>
</tr>
<tr>
<td>98</td>
<td>5-lattice localization solution</td>
</tr>
<tr>
<td>99</td>
<td>10-lattice localization solution</td>
</tr>
<tr>
<td>100</td>
<td>100 randomized noiseless sensor localization</td>
</tr>
<tr>
<td>101</td>
<td>100 randomized sensors localization</td>
</tr>
<tr>
<td>102</td>
<td>Nonnegative spectral factorization</td>
</tr>
<tr>
<td>103</td>
<td>Regularization curve for convex iteration</td>
</tr>
<tr>
<td>104</td>
<td>1-norm heuristic</td>
</tr>
<tr>
<td>105</td>
<td>Sparse sampling theorem</td>
</tr>
<tr>
<td>106</td>
<td>Signal dropout</td>
</tr>
<tr>
<td>107</td>
<td>Signal dropout reconstruction</td>
</tr>
<tr>
<td>108</td>
<td>Simplex with intersecting line problem in compressed sensing</td>
</tr>
<tr>
<td>109</td>
<td>Geometric interpretations of sparse-sampling constraints</td>
</tr>
<tr>
<td>110</td>
<td>Permutation matrix column-norm and column-sum constraint</td>
</tr>
<tr>
<td>111</td>
<td>MAX CUT problem</td>
</tr>
<tr>
<td>112</td>
<td>Shepp-Logan phantom</td>
</tr>
<tr>
<td>113</td>
<td>MRI radial sampling pattern in Fourier domain</td>
</tr>
<tr>
<td>114</td>
<td>Aliased phantom</td>
</tr>
<tr>
<td>115</td>
<td>Neighboring-pixel stencil on Cartesian grid</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>116</td>
<td>Differentiable almost everywhere</td>
<td>329</td>
</tr>
<tr>
<td>117</td>
<td><em>Eternity</em> II game-board grid</td>
<td>331</td>
</tr>
<tr>
<td>118</td>
<td><em>Eternity</em> II demo-game piece illustrating edge-color ordering</td>
<td>334</td>
</tr>
<tr>
<td>119</td>
<td><em>Eternity</em> II vectorized demo-game-board piece descriptions</td>
<td>335</td>
</tr>
<tr>
<td>120</td>
<td><em>Eternity</em> II difference and boundary parameter construction</td>
<td>337</td>
</tr>
<tr>
<td>121</td>
<td>Sparsity pattern for composite <em>Eternity</em> II variable matrix</td>
<td>338</td>
</tr>
<tr>
<td>122</td>
<td>MIT logo</td>
<td>346</td>
</tr>
<tr>
<td>123</td>
<td>One-pixel camera</td>
<td>347</td>
</tr>
<tr>
<td>124</td>
<td>One-pixel camera - compression estimates</td>
<td>347</td>
</tr>
<tr>
<td>125</td>
<td>Straight line through three direction vectors by midpoint fit</td>
<td>353</td>
</tr>
<tr>
<td>126</td>
<td>Convex hull of three points</td>
<td>356</td>
</tr>
<tr>
<td>127</td>
<td>Complete dimensionless EDM graph</td>
<td>358</td>
</tr>
<tr>
<td>128</td>
<td>Fifth Euclidean metric property</td>
<td>360</td>
</tr>
<tr>
<td>129</td>
<td>Fermat point</td>
<td>367</td>
</tr>
<tr>
<td>130</td>
<td>Arbitrary hexagon in $\mathbb{R}^3$</td>
<td>369</td>
</tr>
<tr>
<td>131</td>
<td>Kissing number</td>
<td>370</td>
</tr>
<tr>
<td>132</td>
<td>Trilateration</td>
<td>374</td>
</tr>
<tr>
<td>133</td>
<td>This EDM graph provides unique isometric reconstruction</td>
<td>378</td>
</tr>
<tr>
<td>134</td>
<td>Two sensors $\bullet$ and three anchors $\circ$</td>
<td>378</td>
</tr>
<tr>
<td>135</td>
<td>Two discrete linear trajectories of sensors</td>
<td>379</td>
</tr>
<tr>
<td>136</td>
<td>Coverage in cellular telephone network</td>
<td>381</td>
</tr>
<tr>
<td>137</td>
<td>Contours of equal signal power</td>
<td>381</td>
</tr>
<tr>
<td>138</td>
<td>Depiction of molecular conformation</td>
<td>383</td>
</tr>
<tr>
<td>139</td>
<td>Square diamond</td>
<td>390</td>
</tr>
<tr>
<td>140</td>
<td>Orthogonal complements in $S^N$ abstractly oriented</td>
<td>392</td>
</tr>
<tr>
<td>141</td>
<td>Elliptope $\mathcal{E}^3$</td>
<td>408</td>
</tr>
<tr>
<td>142</td>
<td>Elliptope $\mathcal{E}^2$ interior to $S^2_+$</td>
<td>409</td>
</tr>
<tr>
<td>143</td>
<td>Smallest eigenvalue of $-V^T_NDV_N$</td>
<td>413</td>
</tr>
<tr>
<td>144</td>
<td>Some entrywise EDM compositions</td>
<td>413</td>
</tr>
<tr>
<td>145</td>
<td>Map of United States of America</td>
<td>424</td>
</tr>
<tr>
<td>146</td>
<td>Largest ten eigenvalues of $-V^T_NOV_N$</td>
<td>426</td>
</tr>
<tr>
<td>147</td>
<td><em>Relative-angle inequality tetrahedron</em></td>
<td>431</td>
</tr>
<tr>
<td>148</td>
<td>Nonsimplicial pyramid in $\mathbb{R}^3$</td>
<td>435</td>
</tr>
</tbody>
</table>

## 5 Euclidean Distance Matrix

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>127</td>
<td>Convex hull of three points</td>
<td>356</td>
</tr>
<tr>
<td>128</td>
<td>Complete dimensionless EDM graph</td>
<td>358</td>
</tr>
<tr>
<td>129</td>
<td>Fifth Euclidean metric property</td>
<td>360</td>
</tr>
<tr>
<td>130</td>
<td>Fermat point</td>
<td>367</td>
</tr>
<tr>
<td>131</td>
<td>Arbitrary hexagon in $\mathbb{R}^3$</td>
<td>369</td>
</tr>
<tr>
<td>132</td>
<td>Kissing number</td>
<td>370</td>
</tr>
<tr>
<td>133</td>
<td>Trilateration</td>
<td>374</td>
</tr>
<tr>
<td>134</td>
<td>This EDM graph provides unique isometric reconstruction</td>
<td>378</td>
</tr>
<tr>
<td>135</td>
<td>Two sensors $\bullet$ and three anchors $\circ$</td>
<td>378</td>
</tr>
<tr>
<td>136</td>
<td>Two discrete linear trajectories of sensors</td>
<td>379</td>
</tr>
<tr>
<td>137</td>
<td>Coverage in cellular telephone network</td>
<td>381</td>
</tr>
<tr>
<td>138</td>
<td>Contours of equal signal power</td>
<td>381</td>
</tr>
<tr>
<td>139</td>
<td>Depiction of molecular conformation</td>
<td>383</td>
</tr>
<tr>
<td>140</td>
<td>Square diamond</td>
<td>390</td>
</tr>
<tr>
<td>141</td>
<td>Orthogonal complements in $S^N$ abstractly oriented</td>
<td>392</td>
</tr>
<tr>
<td>142</td>
<td>Elliptope $\mathcal{E}^3$</td>
<td>408</td>
</tr>
<tr>
<td>143</td>
<td>Elliptope $\mathcal{E}^2$ interior to $S^2_+$</td>
<td>409</td>
</tr>
<tr>
<td>144</td>
<td>Smallest eigenvalue of $-V^T_NDV_N$</td>
<td>413</td>
</tr>
<tr>
<td>145</td>
<td>Some entrywise EDM compositions</td>
<td>413</td>
</tr>
<tr>
<td>146</td>
<td>Map of United States of America</td>
<td>424</td>
</tr>
<tr>
<td>147</td>
<td>Largest ten eigenvalues of $-V^T_NOV_N$</td>
<td>426</td>
</tr>
<tr>
<td>148</td>
<td><em>Relative-angle inequality tetrahedron</em></td>
<td>431</td>
</tr>
<tr>
<td>149</td>
<td>Nonsimplicial pyramid in $\mathbb{R}^3$</td>
<td>435</td>
</tr>
</tbody>
</table>

## 6 Cone of distance matrices

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>Relative boundary of cone of Euclidean distance matrices</td>
<td>439</td>
</tr>
<tr>
<td>151</td>
<td>Example of $V_X$ selection to make an EDM</td>
<td>444</td>
</tr>
<tr>
<td>152</td>
<td>Vector $V_X$ spirals</td>
<td>446</td>
</tr>
<tr>
<td>153</td>
<td>Three views of translated negated elliptope</td>
<td>452</td>
</tr>
<tr>
<td>154</td>
<td>Halfline $T$ on PSD cone boundary</td>
<td>455</td>
</tr>
<tr>
<td>155</td>
<td>Vectorization and projection interpretation example</td>
<td>456</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

156 Intersection of EDM cone with hyperplane ........................................ 459
157 Neighborhood graph ................................................................. 460
158 *Trefoil knot* untied ................................................................. 462
159 *Trefoil ribbon* .......................................................... 464
160 Orthogonal complement of geometric center subspace ......................... 468
161 EDM cone construction by flipping PSD cone .................................. 469
162 Decomposing member of polar EDM cone ..................................... 472
163 Ordinary dual EDM cone projected on $S^3_h$ ................................ 477

7 **Proximity problems** 479
164 Pseudo-Venn diagram for EDM ................................................. 481
165 Elbow placed in path of projection ............................................. 482
166 Convex envelope ................................................................. 497

A **Linear algebra** 511
167 Geometrical interpretation of full SVD ...................................... 537

B **Simple matrices** 545
168 Four fundamental subspaces for any dyad ................................ 546
169 Four fundamental subspaces for doublet .................................... 550
170 Four fundamental subspaces for elementary matrix ....................... 551
171 Gimbal ................................................................. 558

D **Matrix calculus** 575
172 Convex quadratic bowl in $\mathbb{R}^2 \times \mathbb{R}$ ................................ 584

E **Projection** 601
173 Action of pseudoinverse .......................................................... 602
174 Nonorthogonal projection of $x \in \mathbb{R}^3$ on $\mathcal{R}(U) = \mathbb{R}^2$ .......... 607
175 Biorthogonal expansion of point $x \in \text{aff } K$ .............................. 617
176 Dual interpretation of projection on convex set ................................ 631
177 Projection on orthogonal complement ....................................... 634
178 Projection on dual cone ......................................................... 636
179 Projection product on convex set in subspace ................................ 640
180 von Neumann-style projection of point $b$ ................................... 643
181 Alternating projection on two halfspaces .................................... 644
182 Distance, feasibility, optimization ............................................ 645
183 Alternating projection on nonnegative orthant and hyperplane .......... 648
184 Geometric convergence of iterates in norm .................................. 648
185 Distance between PSD cone and iterate in $A$ ............................... 652
186 Dykstra's alternating projection algorithm ................................. 653
187 Polyhedral normal cones ........................................................ 654
LIST OF FIGURES

188  Normal cone to ellipope ......................... 656
189  Normal-cone progression ........................ 657
List of Tables

2 Convex geometry
   Table 2.9.2.3.1, rank versus dimension of $S_+^3$ faces 107
   Table 2.10.0.0.1, maximum number of c.i. directions 122
   Cone Table 1 166
   Cone Table S 166
   Cone Table A 167
   Cone Table 1* 170

4 Semidefinite programming
   Faces of $S_+^3$ corresponding to faces of $S^3$ 243

5 Euclidean Distance Matrix
   Précis 5.7.2: affine dimension in terms of rank 399

B Simple matrices
   Auxiliary V-matrix Table B.4.4 556

D Matrix calculus
   Table D.2.1, algebraic gradients and derivatives 593
   Table D.2.2, trace Kronecker gradients 594
   Table D.2.3, trace gradients and derivatives 595
   Table D.2.4, logarithmic determinant gradients, derivatives 597
   Table D.2.5, determinant gradients and derivatives 598
   Table D.2.6, logarithmic derivatives 598
   Table D.2.7, exponential gradients and derivatives 599
Chapter 1
Overview

Convex Optimization
Euclidean Distance Geometry

People are so afraid of convex analysis.

——Claude Lemaréchal, 2003

In layman’s terms, the mathematical science of Optimization is the study of how to make a good choice when confronted with conflicting requirements. The qualifier convex means: when an optimal solution is found, then it is guaranteed to be a best solution; there is no better choice.

Any convex optimization problem has geometric interpretation. If a given optimization problem can be transformed to a convex equivalent, then this interpretive benefit is acquired. That is a powerful attraction: the ability to visualize geometry of an optimization problem. Conversely, recent advances in geometry and in graph theory hold convex optimization within their proofs’ core. [426] [330]

This book is about convex optimization, convex geometry (with particular attention to distance geometry), and nonconvex, combinatorial, and geometrical problems that can be relaxed or transformed into convex problems. A virtual flood of new applications follows by epiphany that many problems, presumed nonconvex, can be so transformed: [11] [12] [35, §4.3, p.316-322] [61] [96] [156] [159] [288] [309] [317] [373] [374] [423] [426] e.g, sigma delta analog-to-digital (A/D) audio converter demodulation (Figure 1).

Euclidean distance geometry is, fundamentally, a determination of point conformation (configuration, relative position or location) by inference from interpoint distance information. By inference we mean: e.g, given only distance information, determine whether there corresponds a realizable conformation of points; a list of points in some dimension that attains the given interpoint distances. Each point may represent simply location or, abstractly, any entity expressible as a vector in finite-dimensional Euclidean space; e.g, distance geometry of music [113].
CHAPTER 1. OVERVIEW

Figure 1: Sigma delta quantization is predominant technology [2, p.6] for analog-to-digital audio conversion. Input signal $u(t)$ is continuous. Delay $z^{-1}$ here is analog, perhaps implemented by sample/hold circuit at MHz rate of $\hat{y}_i$ samples. Observing only $\hat{y}_i$, samples $u$ can be reconstructed by finding a feasible point in the set of linear inequalities representing this coarse quantizer recursion. $R$ is a lower triangular matrix of ones. [103]

It is a common misconception to presume that some desired point conformation cannot be recovered in absence of complete interpoint distance information. We might, for example, want to realize a constellation given only interstellar distance (or, equivalently, parsecs from our Sun and relative angular measurement; the Sun as vertex to two distant stars); called stellar cartography, an application evoked by Figure 2. At first it may seem $O(N^2)$ data is required, yet there are many circumstances where this can be reduced to $O(N)$.

If we agree that a set of points may have a shape (three points can form a triangle and its interior, for example, four points a tetrahedron), then we can ascribe shape of a set of points to their convex hull. It should be apparent: from distance, these shapes can be determined only to within a rigid transformation (rotation, reflection, translation).

Absolute position information is generally lost, given only distance information, but we can determine the smallest possible dimension in which an unknown list of points can exist; that attribute is their affine dimension (a triangle in any ambient space has affine dimension 2, for example). In circumstances where stationary reference points are also provided, it becomes possible to determine absolute position or location; e.g., Figure 3.

Geometric problems involving distance between points can sometimes be reduced to convex optimization problems. Mathematics of this combined study of geometry and optimization is rich and deep. Its application has already proven invaluable discerning organic molecular conformation by measuring interatomic distance along covalent bonds; e.g., Figure 4. [91] [363] [146] [47] Many disciplines have already benefitted and simplified consequent to this theory; e.g, distance-based pattern recognition (Figure 5), localization in wireless sensor networks by measurement of intersensor distance along channels of communication, [48] [421] [46] wireless location of a radio-signal source such as a cell phone by multiple measurements of signal strength, the global positioning system (GPS), and multidimensional scaling which is a numerical representation of qualitative data by finding a low-dimensional scale.

Euclidean distance geometry together with convex optimization have also found application in artificial intelligence:
Figure 2: Orion nebula. (astrophotography by Massimo Robberto)

- to machine learning by discerning naturally occurring manifolds in:
  - Euclidean bodies (Figure 6, §6.7.0.0.1),
  - Fourier spectra of kindred utterances [222],
  - and photographic image sequences [404],

- to robotics; e.g., automatic control of vehicles maneuvering in formation. (Figure 9)

by chapter

We study the many manifestations and representations of pervasive convex Euclidean bodies. In particular, we make convex polyhedra, cones, and dual cones visceral through illustration in chapter 2 Convex geometry where geometric relation of polyhedral cones to nonorthogonal bases (biorthogonal expansion) is examined. It is shown that coordinates are unique in any conic system whose basis cardinality equals or exceeds spatial dimension; for high cardinality, a new definition of conic coordinate is provided in Theorem 2.13.12.0.1. The conic analogue to linear independence, called conic independence, is introduced as a tool for study, analysis, and manipulation of cones; a natural extension and next logical step in progression: linear, affine, conic. We explain conversion between halfspace- and vertex-description of a convex cone, we motivate the dual cone and provide formulae for finding it, and we show how first-order optimality conditions or alternative systems of linear inequality or linear matrix inequality can be explained by dual generalized inequalities with respect to convex cones. Arcane theorems of alternative generalized inequality are, in fact, simply derived from cone membership relations; generalizations of algebraic Farkas’ lemma translated to geometry of convex cones.
Figure 3: Application of trilateration (§5.4.2.8) is localization (determining position) of a radio signal source in 2 dimensions; more commonly known by radio engineers as the process “triangulation”. In this scenario, anchors $\hat{x}_2, \hat{x}_3, \hat{x}_4$ are illustrated as fixed antennae. [219] The radio signal source (a sensor • $x_1$) anywhere in affine hull of three antenna bases can be uniquely localized by measuring distance to each (dashed white arrowed line segments). Ambiguity of lone distance measurement to sensor is represented by circle about each antenna. Trilateration is expressible as a semidefinite program; hence, a convex optimization problem. [331]

Any convex optimization problem can be visualized geometrically. Desire to visualize in high dimension [Sagan, Cosmos – The Edge of Forever; 22:55'] is deeply embedded in the mathematical psyche. [1] Chapter 2 provides tools to make visualization easier, and we teach how to visualize in high dimension. The concepts of face, extreme point, and extreme direction of a convex Euclidean body are explained here; crucial to understanding convex optimization. How to find the smallest face of any pointed closed convex cone containing convex set $C$, for example, is divulged; later shown to have practical application to presolving convex programs. The convex cone of positive semidefinite matrices, in particular, is studied in depth:

- We interpret, for example, inverse image of the positive semidefinite cone under affine transformation. (Example 2.9.1.0.2)
- Subsets of the positive semidefinite cone, discriminated by rank exceeding some lower bound, are convex. In other words, high-rank subsets of the positive semidefinite cone boundary united with its interior are convex. (Theorem 2.9.2.9.3) There is a closed form for projection on those convex subsets.
- The positive semidefinite cone is a circular cone in low dimension; Geršgorin discs specify inscription of a polyhedral cone into it. (Figure 50)
Figure 4: Distance data collected via nuclear magnetic resonance (NMR) helped render this 3-dimensional depiction of a protein molecule. At the beginning of the 1980s, Kurt Wüthrich [Nobel laureate] developed an idea about how NMR could be extended to cover biological molecules such as proteins. He invented a systematic method of pairing each NMR signal with the right hydrogen nucleus (proton) in the macromolecule. The method is called sequential assignment and is today a cornerstone of all NMR structural investigations. He also showed how it was subsequently possible to determine pairwise distances between a large number of hydrogen nuclei and use this information with a mathematical method based on distance-geometry to calculate a three-dimensional structure for the molecule. [410] [193]

Chapter 3 Geometry of convex functions observes Fenchel’s analogy between convex sets and functions: We explain, for example, how the real affine function relates to convex functions as the hyperplane relates to convex sets. Partly a toolbox of practical useful convex functions and a cookbook for optimization problems, methods are drawn from the appendices about matrix calculus for determining convexity and discerning geometry.

Chapter 4. Semidefinite programming has recently emerged to prominence because it admits a new class of problem previously unsolvable by convex optimization techniques, and because it theoretically subsumes other convex techniques: linear programming and quadratic programming and second-order cone programming. (p.240) Semidefinite programming is reviewed with particular attention to optimality conditions for prototypical primal and dual problems, their interplay, and a perturbation method for rank reduction of optimal solutions (extant but not well known). Positive definite Farkas’ lemma is derived, and we also show how to determine if a feasible set belongs exclusively to a positive semidefinite cone boundary. An arguably good three-dimensional polyhedral analogue to the positive semidefinite cone of $3 \times 3$ symmetric matrices is introduced: a new tool for visualizing coexistence of low- and high-rank optimal solutions in six isomorphic dimensions and a mnemonic aid for understanding semidefinite programs. We find a minimal cardinality Boolean solution to an instance of $Ax = b$: 
CHAPTER 1. OVERVIEW

Figure 5: This coarsely discretized triangulated algorithmically flattened human face (made by Kimmel & the Bronsteins [236]) represents a stage in machine recognition of human identity; called face recognition. Distance geometry is applied to determine discriminating-features.

\[
\begin{align*}
\text{minimize} & \quad \|x\|_0 \\
\text{subject to} & \quad Ax = b \\
& \quad x_i \in \{0, 1\}, \quad i = 1 \ldots n
\end{align*}
\]  

(711)

The sensor-network localization problem is solved in any dimension in this chapter. We introduce a method of convex iteration for constraining rank in the form rank\(G \leq \rho\) and cardinality in the form \(\text{card} \, x \leq k\). Cardinality minimization is applied to a discrete image-gradient of the Shepp-Logan phantom, from Magnetic Resonance Imaging (MRI) in the field of medical imaging, for which we find a new lower bound of 1.9% cardinality. We show how to handle polynomial constraints, and how to transform a rank-constrained problem to a rank-1 problem.

The EDM is studied in chapter 5 Euclidean Distance Matrix: its properties and relationship to both positive semidefinite and Gram matrices. We relate the EDM to the four classical properties of Euclidean metric; thereby, observing existence of an infinity of properties of the Euclidean metric beyond triangle inequality. We proceed by deriving the fifth Euclidean metric property and then explain why furthering this endeavor is inefficient because the ensuing criteria (while describing polyhedra in angle or area, volume, content, and so on \(ad \infinium\)) grow linearly in complexity and number with problem size.

Reconstruction methods are explained and applied to a map of the United States; \(e.g.,\) Figure 7. We also experimentally test a conjecture of Borg & Groenen by reconstructing a distorted but recognizable isotonic map of the USA using only ordinal (comparative) distance data: Figure 146e-f. We demonstrate an elegant method for including dihedral (or torsion) angle constraints into a molecular conformation problem. We explain why
trilateration (a.k.a. localization) is a convex optimization problem. We show how to recover relative position given incomplete interpoint distance information, and how to pose EDM problems or transform geometrical problems to convex optimizations; e.g., kissing number of packed spheres about a central sphere (solved in $\mathbb{R}^3$ by Isaac Newton).

The set of all Euclidean distance matrices forms a pointed closed convex cone called the EDM cone: $\text{EDM}^N$. We offer a new proof of Schoenberg’s seminal characterization of EDMs:

$$D \in \text{EDM}^N \iff \begin{cases} -V_N^T D V_N \succeq 0 \\ D \in S^N_h \end{cases}$$

(988)

Our proof relies on fundamental geometry; assuming, any EDM must correspond to a list of points contained in some polyhedron (possibly at its vertices) and vice versa. It is known, but not obvious, this Schoenberg criterion implies nonnegativity of the EDM entries; proved herein.
Figure 7: About five thousand points along borders constituting United States were used to create an exhaustive matrix of interpoint distance for each and every pair of points in an ordered set (a list); called Euclidean distance matrix. From that noiseless distance information, it is easy to reconstruct map exactly via Schoenberg criterion (988). §5.13.1.0.1, confer Figure 146) Map reconstruction is exact (to within a rigid transformation) given any number of interpoint distances; the greater the number of distances, the greater the detail (as it is for all conventional map preparation).

We characterize the eigenvalue spectrum of an EDM, and then devise a polyhedral spectral cone for determining membership of a given matrix (in Cayley-Menger form) to the convex cone of Euclidean distance matrices; id est, a matrix is an EDM if and only if its nonincreasingly ordered vector of eigenvalues belongs to a polyhedral spectral cone for EDM$^N$:

$$D \in \text{EDM}^N \iff \left\{ \lambda \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \in \left[ \begin{array}{c} \mathbb{R}^N_+ \\ \mathbb{R}^- \end{array} \right] \cap \partial \mathcal{H} \right\} \quad (1206)$$

We will see: spectral cones are not unique.

In chapter 6 Cone of distance matrices we explain a geometric relationship between the cone of Euclidean distance matrices, two positive semidefinite cones, and the elliptope. We illustrate geometric requirements, in particular, for projection of a given matrix on a positive semidefinite cone that establish its membership to the EDM cone. The faces of the EDM cone are described, but still open is the question whether all its faces are exposed as they are for the positive semidefinite cone.

The Schoenberg criterion,

$$D \in \text{EDM}^N \iff \left\{ -V_N^T D V_N \in \mathbb{S}^{N-1}_+ \right\} \quad (988)$$
Figure 8: (a) These bees construct a honeycomb by solving a convex optimization problem (§5.4.2.2.6). The most dense packing of identical spheres about a central sphere in 2 dimensions is 6. Sphere centers describe a regular lattice. (b) A hexabenzocoronene molecule (diameter: 1.4nm) imaged by noncontact atomic force microscopy using a microscope tip terminated with a single carbon monoxide molecule. The carbon-carbon bonds in the imaged molecule appear with different contrast and apparent lengths. Based on these disparities, the bond orders and lengths of the individual bonds can be distinguished. (Image by Leo Gross)

for identifying a Euclidean distance matrix, is revealed to be a discretized membership relation (dual generalized inequalities, a new Farkas'-like lemma) between the EDM cone and its ordinary dual: $\text{EDM}^N$. A matrix criterion for membership to the dual EDM cone is derived that is simpler than the Schoenberg criterion:

$$D^* \in \text{EDM}^N \Leftrightarrow \delta(D^* \mathbf{1}) - D^* \succeq 0 \quad (1356)$$

There is a concise equality, relating the convex cone of Euclidean distance matrices to the positive semidefinite cone, apparently overlooked in the literature; an equality between two large convex Euclidean bodies:

$$\text{EDM}^N = \mathbb{S}^N_h \cap \left( \mathbb{S}_c^N \perp - \mathbb{S}_c^N \right) \quad (1350)$$

Seemingly innocuous problems in terms of point position $x_i \in \mathbb{R}^n$ like

$$\text{minimize} \sum_{i,j \in I} (\|x_i - x_j\| - h_{ij})^2 \quad (1390)$$

$$\text{minimize} \sum_{i,j \in I} (\|x_i - x_j\|^2 - h_{ij})^2 \quad (1391)$$

are difficult to solve. So, in chapter 7 Proximity problems, we instead explore methods of solution to a few fundamental and prevalent Euclidean distance matrix proximity problems; the problem of finding that distance matrix closest, in some sense, to a given
matrix $H = [h_{ij}]$:

\[
\begin{align*}
\text{minimize}_{D} & \quad ||V(D-H)V||_F^2 \\
\text{subject to} & \quad \text{rank}(VDV) \leq \rho \\
& \quad D \in \text{EDM}^N
\end{align*}
\]

\[
\begin{align*}
\text{minimize}_{\sqrt{D}} & \quad ||\sqrt{D} - H||_F^2 \\
\text{subject to} & \quad \text{rank}(VDV) \leq \rho \\
& \quad \sqrt{D} \in \sqrt{\text{EDM}}^N
\end{align*}
\]

\[
\begin{align*}
\text{minimize}_{D} & \quad ||D - H||_F^2 \\
\text{subject to} & \quad \text{rank}(VDV) \leq \rho \\
& \quad D \in \text{EDM}^N
\end{align*}
\]

\[
\begin{align*}
\text{minimize}_{\sqrt{D}} & \quad ||V(\sqrt{D} - H)V||_F^2 \\
\text{subject to} & \quad \text{rank}(VDV) \leq \rho \\
& \quad \sqrt{D} \in \sqrt{\text{EDM}}^N
\end{align*}
\]

We apply a convex iteration method for constraining rank. Known heuristics for rank minimization are also explained. We offer new geometrical proof, of a famous discovery by Eckart & Young in 1936 [136], with particular regard to Euclidean projection of a point on that generally nonconvex subset of the positive semidefinite cone boundary comprising all semidefinite matrices having rank not exceeding a prescribed bound $\rho$. We explain how this problem is transformed to a convex optimization for any rank $\rho$.  

Figure 9: Nanocopter swarm. Robotic vehicles in concert can move larger objects or localize a plume of gas, liquid, or radio waves. [145]
appendices

We presume a reader already comfortable with elementary vector operations; [14, §3] formally known as analytic geometry. [412] Toolboxes are provided, in the form of appendices and code, so as to be more self-contained:

- linear algebra (appendix A is primarily concerned with proper statements of semidefiniteness for square matrices),
- simple matrices (dyad, doublet, elementary, Householder, Schoenberg, orthogonal, etcetera, in appendix B),
- collection of known analytical solutions to some important optimization problems (appendix C),
- matrix calculus remains somewhat unsystematized when compared to ordinary calculus (appendix D concerns matrix-valued functions, matrix differentiation and directional derivatives, Taylor series, and tables of first- and second-order gradients and matrix derivatives),
- an elaborate exposition offering insight into orthogonal and nonorthogonal projection on convex sets (the connection between projection and positive semidefiniteness, for example, or between projection and a linear objective function in appendix E),
- MATLAB code on Wilkinson [394] to discriminate EDMs, to determine conic independence, to reduce or constrain rank of an optimal solution to a semidefinite program, compressed sensing (compressive sampling) for digital image and audio signal processing, and two distinct methods of reconstructing a map of the United States: one given only distance data, the other given only comparative distance.
Figure 10: Three-dimensional reconstruction of David from distance data.

Figure 11: Digital Michelangelo Project, Stanford University. Measuring distance to David by laser rangefinder. Spatial resolution is 0.29mm.
Chapter 2

Convex geometry

Convexity has an immensely rich structure and numerous applications. On the other hand, almost every “convex” idea can be explained by a two-dimensional picture.

—Alexander Barvinok [27, p.vii]

We study convex geometry because it is the easiest of geometries. For that reason, much of a practitioner’s energy is expended seeking invertible transformation of problematic sets to convex ones.

As convex geometry and linear algebra are inextricably bonded by linear inequality (asymmetry), we provide much background material on linear algebra (especially in the appendices) although a reader is assumed comfortable with [341] [343] [212] or any other intermediate-level text. The essential references to convex analysis are [209] [318]. The reader is referred to [340] [27] [403] [43] [63] [315] [370] for a comprehensive treatment of convexity. There is relatively less published pertaining to convex matrix-valued functions. [225] [213, §6.6] [305]

2.1 Convex set

A set \( \mathcal{C} \) is convex iff for all \( Y, Z \in \mathcal{C} \) and \( 0 \leq \mu \leq 1 \)

\[
\mu Y + (1 - \mu)Z \in \mathcal{C}
\] (1)

Under that defining condition on \( \mu \), the linear sum in (1) is called a convex combination of \( Y \) and \( Z \). If \( Y \) and \( Z \) are points in real finite-dimensional Euclidean vector space [237] [412] \( \mathbb{R}^n \) or \( \mathbb{R}^{m \times n} \) (matrices), then (1) represents the closed line segment joining them. Line segments are thereby convex sets; \( \mathcal{C} \) is convex iff the line segment connecting any two points in \( \mathcal{C} \) is itself in \( \mathcal{C} \). Apparent from this definition: a convex set is a connected set. [268, §3.4, §3.5] [43, p.2] A convex set can, but does not necessarily, contain the origin \( \mathbf{0} \).

Dattorro, Convex Optimization † Euclidean Distance Geometry 2e, Meboo, v2015.04.11. 33
An ellipsoid centered at \( x = a \) (Figure 14 p.39), given matrix \( C \in \mathbb{R}^{m\times n} \)

\[
\{ x \in \mathbb{R}^n \mid \| C(x - a) \|^2 = (x - a)^T C^T C (x - a) \leq 1 \}
\] (2)

is a good icon for a convex set.\(^{2,1}\)

### 2.1 subspace

A nonempty subset \( \mathcal{R} \) of real Euclidean vector space \( \mathbb{R}^n \) is called a **subspace** (§2.5) if every vector\(^{2,2}\) of the form \( \alpha x + \beta y \), for \( \alpha, \beta \in \mathbb{R} \), is in \( \mathcal{R} \) whenever vectors \( x \) and \( y \) are. \(^{260, §2.3}\) A subspace is a convex set containing the origin, by definition. \(^{318, p.4}\) Any subspace is therefore open in the sense that it contains no boundary, but closed in the sense \(^{268, §2}\)

\[
\mathcal{R} + \mathcal{R} = \mathcal{R}
\] (3)

It is not difficult to show

\[
\mathcal{R} = -\mathcal{R}
\] (4)

as is true for any subspace \( \mathcal{R} \), because \( x \in \mathcal{R} \iff -x \in \mathcal{R} \). Given any \( x \in \mathcal{R} \)

\[
\mathcal{R} = x + \mathcal{R}
\] (5)

The intersection of an arbitrary collection of subspaces remains a subspace. Any subspace not constituting the entire ambient vector space \( \mathbb{R}^n \) is a **proper subspace**; e.g.,\(^{2,3}\) any line (of infinite extent) through the origin in two-dimensional Euclidean space \( \mathbb{R}^2 \). The vector space \( \mathbb{R}^n \) is itself a conventional subspace, inclusively, \(^{237, §2.1}\) although not proper.

### 2.1.2 linear independence

Arbitrary given vectors in Euclidean space \( \{ \Gamma_i \in \mathbb{R}^n, \ i = 1 \ldots N \} \) are **linearly independent** (l.i.) if and only if, for all \( \zeta \in \mathbb{R}^N \) (\( \zeta_i \in \mathbb{R} \))

\[
\Gamma_1 \zeta_1 + \cdots + \Gamma_{N-1} \zeta_{N-1} - \Gamma_N \zeta_N = 0
\] (6)

has only the **trivial solution** \( \zeta = 0 \); in other words, iff no vector from the given set can be expressed as a linear combination of those remaining.

Geometrically, two nontrivial vector subspaces are linearly independent iff they intersect only at the origin.

#### 2.1.2.1 preservation of linear independence

(\textit{confer} §2.4.2.4, §2.10.1) Linear transformation preserves linear dependence. \(^{237, p.86}\) Conversely, linear independence can be preserved under linear transformation. Given \( Y = [y_1 \ y_2 \ \cdots \ y_N] \in \mathbb{R}^{N \times N} \), consider the mapping

\[
T(\Gamma) : \mathbb{R}^{n \times N} \to \mathbb{R}^{n \times N} \triangleq \Gamma Y
\] (7)

\(^{2,1}\)This particular definition is slablike (Figure 12) in \( \mathbb{R}^n \) when \( C \) has nontrivial nullspace.

\(^{2,2}\)A vector is assumed, throughout, to be a column vector.

\(^{2,3}\)We substitute abbreviation \textit{e.g} in place of the Latin \textit{exempli gratia}. 

---

\( \overline{\mathcal{R}} \) denotes the closure of \( \mathcal{R} \).

\( a \) is centroid of \( \mathcal{R} \).

\( C \) is symmetric, positive definite.

\( \mathbb{R}^n \) is a Euclidean vector space.

\( \mathcal{R} \) is nonempty.

\( \mathbb{R}^m \) is a Euclidean vector space.

\( \mathcal{R} \) is full dimensional.

\( \mathcal{R} \) is full rank.

\( \mathcal{R} \) is open.

\( \mathcal{R} \) is closed.

\( \mathcal{R} \) is compact.

\( \mathcal{R} \) is connected.

\( \mathcal{R} \) is bounded.

\( \mathcal{R} \) is convex.

\( \mathcal{R} \) is a convex set.

\( \mathcal{R} \) is a convex subset.

\( \mathcal{R} \) is a convex hull.

\( \mathcal{R} \) is a convex polyhedron.

\( \mathcal{R} \) is a convex polytope.

\( \mathcal{R} \) is a convex polygon.

\( \mathcal{R} \) is a convex polyhedron.

\( \mathcal{R} \) is a convex polytope.

\( \mathcal{R} \) is a convex polygon.

\( \mathcal{R} \) is a convex polyhedron.

\( \mathcal{R} \) is a convex polytope.

\( \mathcal{R} \) is a convex polygon.
2.1. CONVEX SET

\{ y \in \mathbb{R}^2 \mid c \leq a^T y \leq b \} \quad (2056)

Figure 12: A slab is a convex Euclidean body infinite in extent but not affine. Illustrated in \( \mathbb{R}^2 \), it may be constructed by intersecting two opposing halfspaces whose bounding hyperplanes are parallel but not coincident. Because number of halfspaces used in its construction is finite, slab is a polyhedron (§2.12). (Cartesian axes + and vector inward-normal, to each halfspace-boundary, are drawn for reference.)

whose domain is the set of all matrices \( \Gamma \in \mathbb{R}^{n \times N} \) holding a linearly independent set columnar. Linear independence of \( \{ \Gamma y_i \in \mathbb{R}^n, \; i = 1 \ldots N \} \) demands, by definition, there exist no nontrivial solution \( \zeta \in \mathbb{R}^N \) to

\[
\Gamma y_1 \zeta_1 + \cdots + \Gamma y_{N-1} \zeta_{N-1} - \Gamma y_N \zeta_N = 0
\]

By factoring out \( \Gamma \), we see that triviality is ensured by linear independence of \( \{ y_i \in \mathbb{R}^N \} \).

2.1.3 Orthant:

name given to a closed convex set that is the higher-dimensional generalization of quadrant from the classical Cartesian partition of \( \mathbb{R}^2 \); a Cartesian cone. The most common is the nonnegative orthant \( \mathbb{R}^n_+ \) or \( \mathbb{R}^{n \times n}_+ \) (analogue to quadrant I) to which membership denotes nonnegative vector- or matrix-entries respectively; e.g.,

\[
\mathbb{R}^n_+ \triangleq \{ x \in \mathbb{R}^n \mid x_i \geq 0 \; \forall i \}
\] (9)

The nonpositive orthant \( \mathbb{R}^n_- \) or \( \mathbb{R}^{n \times n}_- \) (analogue to quadrant III) denotes negative and 0 entries. Orthant convexity\(^2 4\) is easily verified by definition (1).

2.1.4 affine set

A nonempty affine set (from the word affinity) is any subset of \( \mathbb{R}^n \) that is a translation of some subspace. Any affine set is convex and open so contains no boundary: e.g., empty set \( \emptyset \), point, line, plane, hyperplane (§2.4.2), subspace, etcetera. The intersection of an arbitrary collection of affine sets remains affine.

\(^2 4\) All orthants are selfdual simplicial cones. (§2.13.5.1, §2.12.3.1.1)
CHAPTER 2. CONVEX GEOMETRY

2.1.4.0.1 Definition. Affine subset.
We analogize affine subset to subspace, defining it to be any nonempty affine set of vectors; an affine subset of \( \mathbb{R}^n \).

For some parallel subspace \( \mathcal{R} \) and any point \( x \in \mathcal{A} \)

\[
\mathcal{A} \text{ is affine } \iff \mathcal{A} = x + \mathcal{R} = \{ y \mid y - x \in \mathcal{R} \}
\]

Affine hull of a set \( C \subseteq \mathbb{R}^n \) (§2.3.1) is the smallest affine set containing it.

2.1.5 dimension

Dimension of an arbitrary set \( S \) is Euclidean dimension of its affine hull: [403, p.14]

\[
\dim S \triangleq \dim \text{aff } S = \dim \text{aff}(S - s), \ s \in S
\]

the same as dimension of the subspace parallel to that affine set \( \text{aff } S \) when nonempty. Hence dimension (of a set) is synonymous with affine dimension. [209, A.2.1]

2.1.6 empty set versus empty interior

Emptiness \( \emptyset \) of a set is handled differently than interior in the classical literature. It is common for a nonempty convex set to have empty interior; e.g., paper in the real world:

- An ordinary flat sheet of paper is a nonempty convex set having empty interior in \( \mathbb{R}^3 \) but nonempty interior relative to its affine hull.

2.1.6.1 relative interior

Although it is always possible to pass to a smaller ambient Euclidean space where a nonempty set acquires an interior [27, ¶II.2.3], we prefer the qualifier relative which is the conventional fix to this ambiguous terminology. [2.7] So we distinguish interior from relative interior throughout: [340] [403] [370]

- Classical interior \( \text{int } C \) is defined as a union of points: \( x \) is an interior point of \( C \subseteq \mathbb{R}^n \) if there exists an open ball of dimension \( n \) and nonzero radius centered at \( x \) that is contained in \( C \).

- Relative interior \( \text{rel int } C \) of a convex set \( C \subseteq \mathbb{R}^n \) is interior relative to its affine hull. [2.8]

---

2.5 The popular term affine subspace is an oxymoron.
2.6 Two affine sets are parallel when one is a translation of the other. [318, p.4]
2.7 Superfluous mingling of terms as in relatively nonempty set would be an unfortunate consequence.
From the opposite perspective, some authors use the term full or full-dimensional to describe a set having nonempty interior.
2.8 Likewise for relative boundary (§2.1.7.2), although relative closure is superfluous. [209, A.2.1]
2.1. CONVEX SET

Figure 13: (a) Closed convex set. (b) Neither open, closed, or convex. Yet PSD cone can remain convex in absence of certain boundary components (§2.9.2.9.3). Nonnegative orthant with origin excluded (§2.6) and positive orthant with origin adjoined [318, p.49] are convex. (c) Open convex set.

Thus defined, it is common (though confusing) for \( \text{int} \mathcal{C} \) the interior of \( \mathcal{C} \) to be empty while its relative interior is not: this happens whenever dimension of its affine hull is less than dimension of the ambient space (\( \dim \text{aff} \mathcal{C} < n \); e.g., were \( \mathcal{C} \) paper) or in the exception when \( \mathcal{C} \) is a single point; [268, §2.2.1]

\[
\text{rel int}\{x\} \triangleq \text{aff}\{x\} = \{x\}, \quad \text{int}\{x\} = \emptyset, \quad x \in \mathbb{R}^n
\]  

(12)

In any case, closure of the relative interior of a convex set \( \mathcal{C} \) always yields closure of the set itself;

\[
\overline{\text{rel int} \mathcal{C}} = \overline{\mathcal{C}}
\]  

(13)

Closure is invariant to translation. If \( \mathcal{C} \) is convex then \( \text{rel int} \mathcal{C} \) and \( \overline{\mathcal{C}} \) are convex. [209, p.24] If \( \mathcal{C} \) has nonempty interior, then

\[
\text{rel int} \mathcal{C} = \text{int} \mathcal{C}
\]  

(14)

Given the intersection of convex set \( \mathcal{C} \) with affine set \( \mathcal{A} \)

\[
\text{rel int}(\mathcal{C} \cap \mathcal{A}) = \text{rel int}(\mathcal{C}) \cap \mathcal{A} \iff \text{rel int}(\mathcal{C}) \cap \mathcal{A} \neq \emptyset
\]  

(15)

Because an affine set \( \mathcal{A} \) is open

\[
\text{rel int} \mathcal{A} = \mathcal{A}
\]  

(16)
2.1.7 classical boundary

(confer §2.1.7.2) Boundary of a set \( C \) is the closure of \( C \) less its interior;
\[
\partial C = \overline{C} \setminus \text{int} C
\]  
(17)

[57, §1.1] which follows from the fact
\[
\text{int} C = C \iff \partial \text{int} C = \partial C
\]  
(18)

and presumption of nonempty interior.\(^\text{2.9}\) Implications are:
- \( \text{int} C = \overline{C} \setminus \partial C \)
- a bounded open set has *boundary* defined but not contained in the set
- interior of an open set is equivalent to the set itself;

from which an open set is defined: [268, p.109]
\[
C \text{ is open } \iff \text{int} C = C
\]  
(19)
\[
C \text{ is closed } \iff \text{int} \overline{C} = C
\]  
(20)

The set illustrated in Figure 13b is not open because it is not equivalent to its interior, for example, it is not closed because it does not contain its boundary, and it is not convex because it does not contain all convex combinations of its boundary points.

2.1.7.1 Line intersection with boundary

A line can intersect the boundary of a convex set in any dimension at a point demarcating the line’s entry to the set interior. On one side of that entry-point along the line is the exterior of the set, on the other side is the set interior. In other words,
- starting from any point of a convex set, a move toward the interior is an immediate entry into the interior. [27, §II.2]

When a line intersects the interior of a convex body in any dimension, the boundary appears to the line to be as thin as a point. This is intuitively plausible because, for example, a line intersects the boundary of the ellipsoids in Figure 14 at a point in \( \mathbb{R} \), \( \mathbb{R}^2 \), and \( \mathbb{R}^3 \). Such thinness is a remarkable fact when pondering visualization of convex polyhedra (§2.12, §5.14.3) in four Euclidean dimensions, for example, having boundaries constructed from other three-dimensional convex polyhedra called *faces*.

We formally define *face* in (§2.6). For now, we observe the boundary of a convex body to be entirely constituted by all its faces of dimension lower than the body itself. Any face of a convex set is convex. For example: The ellipsoids in Figure 14 have boundaries composed only of zero-dimensional faces. The two-dimensional slab in Figure 12 is an unbounded *polyhedron* having one-dimensional faces making its boundary. The three-dimensional bounded polyhedron in Figure 21 has zero-, one-, and two-dimensional polygonal faces constituting its boundary.

\(^{2.9}\) Otherwise, for \( x \in \mathbb{R}^n \) as in (12), [268, §2.1-§2.3]
\[
\text{int} \{x\} = \emptyset = \emptyset
\]
the empty set is both open and closed.
2.1. CONVEX SET

Figure 14: (a) Ellipsoid in $\mathbb{R}$ is a line segment whose boundary comprises two points. Intersection of line with ellipsoid in $\mathbb{R}$, (b) in $\mathbb{R}^2$, (c) in $\mathbb{R}^3$. Each ellipsoid illustrated has entire boundary constituted by zero-dimensional faces; in fact, by vertices (§2.6.1.0.1). Intersection of line with boundary is a point at entry to interior. These same facts hold in higher dimension.


2.1.7.1.1 Example. Intersection of line with boundary in $\mathbb{R}^6$.

The convex cone of positive semidefinite matrices $\mathbb{S}^3_+$ (§2.9), in the ambient subspace of symmetric matrices $\mathbb{S}^3$ (§2.2.2.0.1), is a six-dimensional Euclidean body in isometrically isomorphic $\mathbb{R}^6$ (§2.2.1). Boundary of the positive semidefinite cone, in this dimension, comprises faces having only the dimensions 0, 1, and 3; id est, $\{\rho(\rho+1)/2, \rho = 0, 1, 2\}$.

Unique minimum-distance projection $PX$ (§E.9) of any point $X \in \mathbb{S}^3$ on that cone $\mathbb{S}^3_+$ is known in closed form (§7.1.2). Given, for example, $\lambda \in \mathbb{R}^3_+$ and diagonalization (§A.5.1) of exterior point $X = Q\Lambda Q^T \in \mathbb{S}^3$, where $Q \in \mathbb{R}^{3\times 3}$ is an orthogonal matrix, then the projection on $\mathbb{S}^3_+$ in $\mathbb{R}^6$ is

$$PX = Q\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \in \mathbb{S}^3_+$$

This positive semidefinite matrix $PX$ nearest $X$ thus has rank 2, found by discarding all negative eigenvalues in $\Lambda$. The line connecting these two points is $\{X + (PX - X)t \mid t \in \mathbb{R}\}$ where $t = 0 \Leftrightarrow X$ and $t = 1 \Leftrightarrow PX$. Because this line intersects the boundary of the positive semidefinite cone $\mathbb{S}^3_+$ at point $PX$ and passes through its interior (by assumption), then the matrix corresponding to an infinitesimally positive perturbation of $t$ there should reside interior to the cone (rank 3). Indeed, for $\varepsilon$ an arbitrarily small positive constant,

$$X + (PX - X)t|_{t=1+\varepsilon} = Q(\Lambda + (P\Lambda - \Lambda)(1+\varepsilon))Q^T = Q\begin{bmatrix} \lambda_1 & 0 \\ 0 & \varepsilon\lambda_3 \end{bmatrix} \in \text{int}\, \mathbb{S}^3_+$$

\[\square\]

2.1.7.1.2 Example. Tangential line intersection with boundary.

A higher-dimensional boundary $\partial C$ of a convex Euclidean body $C$ is simply a dimensionally larger set through which a line can pass when it does not intersect the body’s interior. Still, for example, a line existing in five or more dimensions may pass tangentially (intersecting no point interior to $C$ [229, §15.3]) through a single point relatively interior to a three-dimensional face on $\partial C$. Let’s understand why by inductive reasoning.

Figure 15a shows a vertical line-segment whose boundary comprises its two endpoints. For a line to pass through the boundary tangentially (intersecting no point relatively interior to the line-segment), it must exist in an ambient space of at least two dimensions. Otherwise, the line is confined to the same one-dimensional space as the line-segment and must pass along the segment to reach the end points.

Figure 15b illustrates a two-dimensional ellipsoid whose boundary is constituted entirely by zero-dimensional faces. Again, a line must exist in at least two dimensions to tangentially pass through any single arbitrarily chosen point on the boundary (without intersecting the ellipsoid interior).
2.1. **CONVEX SET**

Figure 15: Line tangential: (a) (b) to relative interior of a zero-dimensional face in $\mathbb{R}^2$, (c) (d) to relative interior of a one-dimensional face in $\mathbb{R}^3$. 
Now let’s move to an ambient space of three dimensions. Figure 15c shows a polygon rotated into three dimensions. For a line to pass through its zero-dimensional boundary (one of its vertices) tangentially, it must exist in at least the two dimensions of the polygon. But for a line to pass tangentially through a single arbitrarily chosen point in the relative interior of a one-dimensional face on the boundary as illustrated, it must exist in at least three dimensions.

Figure 15d illustrates a solid circular cone (drawn truncated) whose one-dimensional faces are halflines emanating from its pointed end (vertex). This cone’s boundary is constituted solely by those one-dimensional halflines. A line may pass through the boundary tangentially, striking only one arbitrarily chosen point relatively interior to a one-dimensional face, if it exists in at least the three-dimensional ambient space of the cone.

From these few examples, we deduce a general rule (without proof):

- A line may pass tangentially through a single arbitrarily chosen point relatively interior to a $k$-dimensional face on the boundary of a convex Euclidean body if the line exists in dimension at least equal to $k + 2$.

Now the interesting part, with regard to Figure 21 showing a bounded polyhedron in $\mathbb{R}^3$; call it $\mathcal{P}$: A line existing in at least four dimensions is required in order to pass tangentially (without hitting $\text{int} \mathcal{P}$) through a single arbitrary point in the relative interior of any two-dimensional polygonal face on the boundary of polyhedron $\mathcal{P}$. Now imagine that polyhedron $\mathcal{P}$ is itself a three-dimensional face of some other polyhedron in $\mathbb{R}^4$. To pass a line tangentially through polyhedron $\mathcal{P}$ itself, striking only one point from its relative interior $\text{rel int} \mathcal{P}$ as claimed, requires a line existing in at least five dimensions.\footnote{This rule can help determine whether there exists unique solution to a convex optimization problem whose feasible set is an intersecting line; e.g., the trilateration problem (§5.4.2.2.8).}

It is not too difficult to deduce:

- A line may pass through a single arbitrarily chosen point interior to a $k$-dimensional convex Euclidean body (hitting no other interior point) if that line exists in dimension at least equal to $k + 1$.

In layman’s terms, this means: a being capable of navigating four spatial dimensions (one Euclidean dimension beyond our physical reality) could see inside three-dimensional objects.

### 2.1.7.2 Relative boundary

The classical definition of boundary of a set $\mathcal{C}$ presumes nonempty interior:

$$\partial \mathcal{C} = \overline{\mathcal{C}} \setminus \text{int} \mathcal{C}$$  \hspace{1cm} (17)

More suitable to study of convex sets is the relative boundary; defined \[209, \text{§A.2.1.2}\]

$$\text{rel} \partial \mathcal{C} \triangleq \overline{\mathcal{C}} \setminus \text{rel int} \mathcal{C}$$  \hspace{1cm} (24)

boundary relative to affine hull of $\mathcal{C}$.\footnote{This rule can help determine whether there exists unique solution to a convex optimization problem whose feasible set is an intersecting line; e.g., the trilateration problem (§5.4.2.2.8).}
### 2.1. CONVEX SET

In the exception when \( C \) is a single point \( \{x\} \),

\[
\text{rel } \partial \{x\} = \{x\} \setminus \{x\} = \emptyset, \quad x \in \mathbb{R}^n \tag{25}
\]

A bounded convex polyhedron (\S 2.3.2, \S 2.12.0.0.1) in subspace \( \mathbb{R} \), for example, has boundary constructed from two points, in \( \mathbb{R}^2 \) from at least three line segments, in \( \mathbb{R}^3 \) from convex polygons, while a convex \emph{polychoron} (a bounded polyhedron in \( \mathbb{R}^4 \)) has boundary constructed from three-dimensional convex polyhedra. A halfspace is partially bounded by a hyperplane; its interior therefore excludes that hyperplane. An affine set has no relative boundary.

#### 2.1.8 intersection, sum, difference, product

**2.1.8.0.1 Theorem.** \emph{Intersection.} [318, \S 2, thm.6.5]

Intersection of an arbitrary collection of convex sets \( \{C_i\} \) is convex. For a finite collection of \( N \) sets, a necessarily nonempty intersection of relative interior

\[
\bigcap_{i=1}^N \text{rel int } C_i = \text{rel int } \bigcap_{i=1}^N C_i \tag{26}
\]

equals relative interior of intersection. And for a possibly infinite collection,

\[
\bigcap C_i = \bigcap C_i.
\]

In converse this theorem is implicitly false insofar as a convex set can be formed by the intersection of sets that are not. Unions of convex sets are generally not convex. [209, p.22]

**Vector sum** of two convex sets \( C_1 \) and \( C_2 \) is convex [209, p.24]

\[
C_1 + C_2 = \{x + y \mid x \in C_1, y \in C_2\} \tag{26}
\]

but not necessarily closed unless at least one set is closed and bounded.

By additive inverse, we can similarly define **vector difference** of two convex sets

\[
C_1 - C_2 = \{x - y \mid x \in C_1, y \in C_2\} \tag{27}
\]

which is convex. Applying this definition to nonempty convex set \( C_1 \), its selfdifference \( C_1 - C_1 \) is generally nonempty, nontrivial, and convex; e.g., for any convex cone \( K \), (\S 2.7.2) the set \( K - K \) constitutes its affine hull. [318, p.15]

**Cartesian product** of convex sets

\[
C_1 \times C_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \mid x \in C_1, \ y \in C_2 \right\} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \tag{28}
\]

remains convex. The converse also holds; \emph{id est}, a Cartesian product is convex iff each set is. [209, p.23]

Convex results are also obtained for scaling \( \kappa C \) of a convex set \( C \), rotation/reflection \( QC \), or translation \( C + \alpha \); each similarly defined.

Given any operator \( T \) and convex set \( C \), we are prone to write \( T(C) \) meaning

\[
T(C) \triangleq \{T(x) \mid x \in C\} \tag{29}
\]

Given linear operator \( T \), it therefore follows from (26),

\[
T(C_1 + C_2) = \{T(x + y) \mid x \in C_1, y \in C_2\} = \{T(x) + T(y) \mid x \in C_1, y \in C_2\} = T(C_1) + T(C_2) \tag{30}
\]
2.1.9 inverse image

While *epigraph* (§3.5) of a convex function must be convex, it generally holds that inverse image (Figure 16) of a convex function is not. The most prominent examples to the contrary are affine functions (§3.4):

2.1.9.0.1 Theorem. *Inverse image.*

Let $f$ be a mapping from $\mathbb{R}^{p \times k}$ to $\mathbb{R}^{m \times n}$.

- The image of a convex set $\mathcal{C}$ under any affine function
  \[ f(\mathcal{C}) = \{ f(X) \mid X \in \mathcal{C} \} \subseteq \mathbb{R}^{m \times n} \quad (31) \]

  is convex.

- Inverse image of a convex set $\mathcal{F}$,
  \[ f^{-1}(\mathcal{F}) = \{ X \mid f(X) \in \mathcal{F} \} \subseteq \mathbb{R}^{p \times k} \quad (32) \]

  a single- or many-valued mapping, under any affine function $f$ is convex. \( \diamond \)

In particular, any affine transformation of an affine set remains affine. [318, p.8] Ellipsoids are invariant to any [sic] affine transformation.
2.2. VECTORIZED-MATRIX INNER PRODUCT

Figure 17: (confer Figure 173) Action of linear map represented by $A \in \mathbb{R}^{m \times n}$:
Component of vector $x$ in nullspace $\mathcal{N}(A)$ maps to origin while component in rowspace $\mathcal{R}(A^T)$ maps to range $\mathcal{R}(A)$. For any $A \in \mathbb{R}^{m \times n}$, $A^T A x = x_p$ and $A A^T A x = b$ ($\S E$) and inverse image of $b \in \mathcal{R}(A)$ is a nonempty affine set: $x_p + \mathcal{N}(A)$.

Although not precluded, this inverse image theorem does not require a uniquely invertible mapping $f$. Figure 17, for example, mechanizes inverse image under a general linear map. Example 2.9.1.0.2 and Example 3.5.0.0.2 offer further applications.

Each converse of this two-part theorem is generally false; id est, given $f$ affine, a convex image $f(C)$ does not imply that set $C$ is convex, and neither does a convex inverse image $f^{-1}(F)$ imply set $F$ is convex. A counterexample, invalidating a converse, is easy to visualize when the affine function is an orthogonal projector [341] [260]:

2.1.9.0.2 Corollary. Projection on subspace.\(^\text{2.11}\)

Orthogonal projection of a convex set on a subspace or nonempty affine set is another convex set.

Again, the converse is false. Shadows, for example, are umbral projections that can be convex when the body providing the shade is not.

2.2 Vectorized-matrix inner product

Euclidean space $\mathbb{R}^n$ comes equipped with a vector inner-product

$$\langle y, z \rangle \triangleq y^T z = \|y\| \|z\| \cos \psi \tag{33}$$

\(^\text{2.11}\) For hyperplane representations see $\S 2.4.2$. For projection of convex sets on hyperplanes see [403], §6.6. A nonempty affine set is called an affine subset ($\S 2.1.4.0.1$). Orthogonal projection of points on affine subsets is reviewed in $\S E.4$. 

CHAPTER 2. CONVEX GEOMETRY

where $\psi$ (997) represents angle (in radians) between vectors $y$ and $z$. We prefer those angle brackets to connote a geometric rather than algebraic perspective; e.g., vector $y$ might represent a hyperplane normal ($\S$2.4.2). Two vectors are orthogonal (perpendicular) to one another if and only if their inner product vanishes (iff $\psi$ is an odd multiple of $\pi$):

$$y \perp z \iff \langle y, z \rangle = 0$$  \hfill (34)

When orthogonal vectors each have unit norm, then they are orthonormal.

A vector inner-product defines Euclidean norm (vector 2-norm, $\S$A.7.1)

$$\|y\|_2 = \|y\| \triangleq \sqrt{y^T y}, \quad \|y\| = 0 \iff y = 0$$  \hfill (35)

For linear operator $A$, its adjoint $A^T$ is a linear operator defined by [237, $\S$3.10]

$$\langle y, A^T z \rangle \triangleq \langle Ay, z \rangle$$  \hfill (36)

For linear operation on a vector, represented by real matrix $A$, the adjoint operator $A^T$ is its transposition. This operator is selfadjoint when $A = A^T$.

Vector inner-product for matrices is calculated just as it is for vectors; by first transforming a matrix in $\mathbb{R}^{p \times k}$ to a vector in $\mathbb{R}^{pk}$ by concatenating its columns in the natural order. For lack of a better term, we shall call that linear bijective (one-to-one and onto $\S$237, App.A1.2) transformation vectorization. For example, the vectorization of $Y = [y_1 \ y_2 \ldots \ y_k] \in \mathbb{R}^{p \times k}$ [176] [337] is

$$\text{vec } Y \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} \in \mathbb{R}^k$$  \hfill (37)

Then the vectorized-matrix inner-product is trace of matrix inner-product; for $Z \in \mathbb{R}^{p \times k}$, [63, $\S$2.6.1] [209, $\S$0.3.1] [414, $\S$8] [377, $\S$2.2]

$$\langle Y, Z \rangle \triangleq \text{tr}(Y^T Z) = \text{vec}(Y)^T \text{vec } Z$$  \hfill (38)

where ($\S$A.1.1)

$$\text{tr}(Y^T Z) = \text{tr}(Z Y^T) = \text{tr}(Y Z^T) = \text{tr}(Z^T Y) = 1^T (Y \circ Z) 1$$  \hfill (39)

and where $\circ$ denotes the Hadamard product $^{2.12}$ of matrices [168, $\S$1.1.4]. The adjoint $A^T$ operation on a matrix can therefore be defined in like manner:

$$\langle Y, A^T Z \rangle \triangleq \langle AY, Z \rangle$$  \hfill (40)

Take any element $C_1$ from a matrix-valued set in $\mathbb{R}^{p \times k}$, for example, and consider any particular dimensionally compatible real vectors $v$ and $w$. Then vector inner-product of $C_1$ with $vw^T$ is

$$\langle vw^T, C_1 \rangle = \langle v, C_1 w \rangle = v^T C_1 w = \text{tr}(wv^T C_1) = 1^T ((vw^T) \circ C_1) 1$$  \hfill (41)

$^{2.12}$ Hadamard product is a simple entrywise product of corresponding entries from two matrices of like size; id est, not necessarily square. A commutative operation, the Hadamard product can be extracted from within a Kronecker product. [212, p.475]
2.2. VECTORIZED-MATRIX INNER PRODUCT

Figure 18: (a) Cube in $\mathbb{R}^3$ projected on paper-plane $\mathbb{R}^2$. Subspace projection operator is not an isomorphism because new adjacencies are introduced. (b) Tesseract is a projection of hypercube in $\mathbb{R}^4$ on $\mathbb{R}^3$.

Further, linear bijective vectorization is *distributive* with respect to Hadamard product; *id est*,

$$\text{vec}(Y \circ Z) = \text{vec}(Y) \circ \text{vec}(Z) \quad (42)$$

2.2.0.0.1 Example. Application of inverse image theorem.

Suppose set $\mathcal{C} \subseteq \mathbb{R}^{p \times k}$ were convex. Then for any particular vectors $v \in \mathbb{R}^p$ and $w \in \mathbb{R}^k$, the set of vector inner-products

$$\mathcal{Y} \triangleq v^\top \mathcal{C} w = \langle vw^\top, \mathcal{C} \rangle \subseteq \mathbb{R} \quad (43)$$

is convex. It is easy to show directly that convex combination of elements from $\mathcal{Y}$ remains an element of $\mathcal{Y}$. Instead given convex set $\mathcal{Y}$, $\mathcal{C}$ must be convex consequent to inverse image theorem 2.1.9.0.1.

More generally, $vw^\top$ in (43) may be replaced with any particular matrix $Z \in \mathbb{R}^{p \times k}$ while convexity of set $\langle Z, \mathcal{C} \rangle \subseteq \mathbb{R}$ persists. Further, by replacing $v$ and $w$ with any particular respective matrices $U$ and $W$ of dimension compatible with all elements of convex set $\mathcal{C}$, then set $U^\top \mathcal{C} W$ is convex by the inverse image theorem because it is a linear mapping of $\mathcal{C}$.

2.2.1 Frobenius’

2.2.1.0.1 Definition. *Isomorphic.*

An *isomorphism* of a vector space is a transformation equivalent to a linear bijective mapping. Image and inverse image under the transformation operator are then called *isomorphic vector spaces.*

To verify that, take any two elements $\mathcal{C}_1$ and $\mathcal{C}_2$ from the convex matrix-valued set $\mathcal{C}$, and then form the vector inner-products (45) that are two elements of $\mathcal{Y}$ by definition. Now make a convex combination of those inner products; *videlicet*, for $0 \leq \mu \leq 1$

$$\mu \langle vw^\top, \mathcal{C}_1 \rangle + (1 - \mu) \langle vw^\top, \mathcal{C}_2 \rangle = \langle vw^\top, \mu \mathcal{C}_1 + (1 - \mu) \mathcal{C}_2 \rangle$$

The two sides are equivalent by linearity of inner product. The right-hand side remains a vector inner-product of $vw^\top$ with an element $\mu \mathcal{C}_1 + (1 - \mu) \mathcal{C}_2$ from the convex set $\mathcal{C}$; hence, it belongs to $\mathcal{Y}$. Since that holds true for any two elements from $\mathcal{Y}$, then it must be a convex set.
Isomorphic vector spaces are characterized by preservation of adjacency; id est, if \( v \) and \( w \) are points connected by a line segment in one vector space, then their images will be connected by a line segment in the other. Two Euclidean bodies may be considered isomorphic if there exists an isomorphism, of their vector spaces, under which the bodies correspond. \([379, \S L.1]\) Projection (\( \mathbb{E} \)) is not an isomorphism, Figure 18 for example; hence, perfect reconstruction (inverse projection) is generally impossible without additional information.

When \( Z = Y \in \mathbb{R}^{p \times k} \) in (38), Frobenius’ norm is resultant from vector inner-product; (confer (1774))

\[
\|Y\|_F^2 = \|\text{vec} Y\|_2^2 = \langle Y, Y \rangle = \text{tr}(Y^T Y) = \sum_{i,j} Y_{ij}^2 = \sum_i \lambda(Y^T Y)_i = \sum_i \sigma(Y)_i^2
\]

where \( \lambda(Y^T Y)_i \) is the \( i \)th eigenvalue of \( Y^T Y \), and \( \sigma(Y)_i \) the \( i \)th singular value of \( Y \).

Were \( Y \) a normal matrix (\( \mathbb{A}.5.1 \)), then \( \sigma(Y)_i = |\lambda(Y)_i| \) \([425, \S 8.1]\] thus

\[
\|Y\|_F^2 = \sum_i \lambda(Y)_i^2 = \|\lambda(Y)\|_2^2 = \langle \lambda(Y), \lambda(Y) \rangle = \langle Y, Y \rangle
\]

The converse (45) \( \Rightarrow \) normal matrix \( Y \) also holds. \([212, \S 2.5.4]\)

Frobenius’ norm is the Euclidean norm of vectorized matrices. Because the metrics are equivalent, for \( X \in \mathbb{R}^{p \times k} \)

\[
\|\text{vec} X - \text{vec} Y\|_2 = \|X - Y\|_F
\]

and because vectorization (37) is a linear bijective map, then vector space \( \mathbb{R}^{p \times k} \) is isometrically isomorphic with vector space \( \mathbb{R}^k \) in the Euclidean sense and vec is an isometric isomorphism of \( \mathbb{R}^{p \times k} \). Because of this Euclidean structure, all known results from convex analysis in Euclidean space \( \mathbb{R}^n \) carry over directly to the space of real matrices \( \mathbb{R}^{p \times k} \); e.g, norm function convexity (\( \mathbb{S}3.2 \)).

2.2.1.1 Injective linear operators

Injective mapping (transformation) means one-to-one mapping; synonymous with uniquely invertible linear mapping on Euclidean space.

- Linear injective mappings are fully characterized by lack of nontrivial nullspace.

2.2.1.1.1 Definition. Isometric isomorphism.

An isometric isomorphism of a vector space having a metric defined on it is a linear bijective mapping \( T \) that preserves distance; id est, for all \( x, y \in \text{dom} T \)

\[
\|Tx - Ty\| = \|x - y\|
\]

Then isometric isomorphism \( T \) is called a bijective isometry.

Unitary linear operator \( Q : \mathbb{R}^k \rightarrow \mathbb{R}^k \), represented by orthogonal matrix \( Q \in \mathbb{R}^{k \times k} \) (\( \mathbb{B}.5.2 \)), is an isometric isomorphism; e.g, discrete Fourier transform via (882). Suppose \( T(X) = UXQ \) is a bijective isometry where \( U \) is a dimensionally compatible orthonormal
2.2. VECTORIZED-MATRIX INNER PRODUCT

\[ \dim \text{dom} T = \dim \mathcal{R}(T) \]

Figure 19: Linear injective mapping \( T \mathbf{x} = A \mathbf{x} : \mathbb{R}^2 \to \mathbb{R}^3 \) of Euclidean body remains two-dimensional under mapping represented by skinny full-rank matrix \( A \in \mathbb{R}^{3 \times 2} \); two bodies are isomorphic by Definition 2.2.1.0.1.

\[ \parallel U(\mathbf{X} - \mathbf{Y})Q \parallel_F = \parallel \mathbf{X} - \mathbf{Y} \parallel_F \quad (48) \]

Then we also say Frobenius’ norm is orthogonally invariant; meaning, for \( \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{p \times k} \)

Yet isometric operator \( T : \mathbb{R}^2 \to \mathbb{R}^3 \), represented by \( A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \) on \( \mathbb{R}^2 \), is injective but not a surjective map to \( \mathbb{R}^3 \). [237, §1.6, §2.6] This operator \( T \) can therefore be a bijective isometry only with respect to its range.

Any linear injective transformation on Euclidean space is uniquely invertible on its range. In fact, any linear injective transformation has a range whose dimension equals that of its domain. In other words, for any invertible linear transformation \( T \) [ibidem]

\[ \dim \text{dom}(T) = \dim \mathcal{R}(T) \quad (49) \]

e.g., \( T \) represented by skinny-or-square full-rank matrices. (Figure 19) An important consequence of this fact is:

- Affine dimension, of any \( n \)-dimensional Euclidean body in domain of operator \( T \), is invariant to linear injective transformation.

2.2.1.1.2 Example. Noninjective linear operators.
Mappings in Euclidean space created by noninjective linear operators can be characterized in terms of an orthogonal projector (§E). Consider noninjective linear operator

\[ \text{any matrix } U \text{ whose columns are orthonormal with respect to each other } (U^TU = I); \text{ these include the orthogonal matrices.} \]
Figure 20: Linear noninjective mapping $PTx = A^\dagger Ax : \mathbb{R}^3 \to \mathbb{R}^3$ of three-dimensional Euclidean body $B$ has affine dimension 2 under projection on rowspace of fat full-rank matrix $A \in \mathbb{R}^{2 \times 3}$. Set of coefficients of orthogonal projection $TB = \{Ax \mid x \in B\}$ is isomorphic with projection $P(TB)$ [sic].

$Tx = Ax : \mathbb{R}^n \to \mathbb{R}^m$ represented by fat matrix $A \in \mathbb{R}^{m \times n}$ ($m < n$). What can be said about the nature of this $m$-dimensional mapping?

Concurrently, consider injective linear operator $Py = A^\dagger y : \mathbb{R}^m \to \mathbb{R}^n$ where $\mathcal{R}(A^\dagger) = \mathcal{R}(A^T)$. $P(Ax) = PTx$ achieves projection of vector $x$ on the row space $\mathcal{R}(A^T)$. (§E.3.1) This means vector $Ax$ can be succinctly interpreted as coefficients of orthogonal projection.

Pseudoinverse matrix $A^\dagger$ is skinny and full-rank, so operator $P_y$ is a linear bijection with respect to its range $\mathcal{R}(A^\dagger)$. By Definition 2.2.1.0.1, image $P(TB)$ of projection $PT(B)$ on $\mathcal{R}(A^T)$ in $\mathbb{R}^n$ must therefore be isomorphic with the set of projection coefficients $TB = \{Ax \mid x \in B\}$ in $\mathbb{R}^m$ and have the same affine dimension by (49). To illustrate, we present a three-dimensional Euclidean body $B$ in Figure 20 where any point $x$ in the nullspace $\mathcal{N}(A)$ maps to the origin.

### 2.2.2 Symmetric matrices

#### 2.2.2.0.1 Definition. Symmetric matrix subspace.

Define a subspace of $\mathbb{R}^{M \times M}$: the convex set of all symmetric $M \times M$ matrices;

$$S^M \triangleq \left\{ A \in \mathbb{R}^{M \times M} \mid A = A^T \right\} \subseteq \mathbb{R}^{M \times M} \quad (50)$$

This subspace comprising symmetric matrices $S^M$ is isomorphic with the vector space $\mathbb{R}^{M(M+1)/2}$ whose dimension is the number of free variables in a symmetric $M \times M$ matrix. The orthogonal complement [341] [260] of $S^M$ is

$$S^{M,\perp} \triangleq \left\{ A \in \mathbb{R}^{M \times M} \mid A = -A^T \right\} \subset \mathbb{R}^{M \times M} \quad (51)$$
2.2. VECTORIZED-MATRIX INNER PRODUCT

the subspace of antisymmetric matrices in $\mathbb{R}^{M \times M}$; id est,

$$S^M \oplus S^M_\perp = \mathbb{R}^{M \times M}$$

where unique vector sum $\oplus$ is defined on page 662. \triangle

All antisymmetric matrices are hollow by definition (have 0 main diagonal). Any square matrix $A \in \mathbb{R}^{M \times M}$ can be written as a sum of its symmetric and antisymmetric parts: respectively,

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$$

The symmetric part is orthogonal in $\mathbb{R}^{M^2}$ to the antisymmetric part; videlicet,

$$\text{tr}((A + A^T)(A - A^T)) = 0$$

In the ambient space of real matrices, the antisymmetric matrix subspace can be described

$$S^{M_\perp} = \left\{ \frac{1}{2}(A - A^T) \mid A \in \mathbb{R}^{M \times M} \right\} \subset \mathbb{R}^{M \times M}$$

because any matrix in $S^M$ is orthogonal to any matrix in $S^{M_\perp}$. Further confined to the ambient subspace of symmetric matrices, $S^{M_\perp}$ would become trivial.

2.2.2.1 Isomorphism of symmetric matrix subspace

When a matrix is symmetric in $S^M$, we may still employ the vectorization transformation (37) to $\mathbb{R}^{M^2}$; vec, an isometric isomorphism. We might instead choose to realize in the lower-dimensional subspace $\mathbb{R}^{M(M+1)/2}$ by ignoring redundant entries (below the main diagonal) during transformation. Such a realization would remain isomorphic but not isometric. Lack of isometry is a spatial distortion due now to disparity in metric between $\mathbb{R}^{M^2}$ and $\mathbb{R}^{M(M+1)/2}$. To realize isometrically in $\mathbb{R}^{M(M+1)/2}$, we must make a correction: For $Y = [Y_{ij}] \in S^M$ we take symmetric vectorization [225, §2.2.1]

$$\text{svec } Y \triangleq \begin{bmatrix}
Y_{11} \\
\sqrt{2}Y_{12} \\
Y_{22} \\
\sqrt{2}Y_{13} \\
\sqrt{2}Y_{23} \\
\vdots \\
Y_{MM}
\end{bmatrix} \in \mathbb{R}^{M(M+1)/2}$$

where all entries off the main diagonal have been scaled. Now for $Z \in S^M$

$$\langle Y, Z \rangle \triangleq \text{tr}(Y^T Z) = \text{vec}(Y)^T \text{vec } Z = 1^T (Y \circ Z) 1 = \text{svec}(Y)^T \text{svec } Z$$

Then because the metrics become equivalent, for $X \in S^M$

$$\|\text{svec } X - \text{svec } Y\|_2 = \|X - Y\|_F$$
and because symmetric vectorization (56) is a linear bijective mapping, then svec is an isometric isomorphism of the symmetric matrix subspace. In other words, $S^M$ is isometrically isomorphic with $\mathbb{R}^{M(M+1)/2}$ in the Euclidean sense under transformation svec.

The set of all symmetric matrices $S^M$ forms a proper subspace in $\mathbb{R}^{M \times M}$, so for it there exists a standard orthonormal basis in isometrically isomorphic $\mathbb{R}^{M(M+1)/2}$

\[
\{E_{ij} \in S^M\} = \left\{ e_i e_i^T, \quad i = j = 1 \ldots M \right\}
\]

where $M(M+1)/2$ standard basis matrices $E_{ij}$ are formed from the standard basis vectors

\[
e_i = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}, \quad j = 1 \ldots M \quad \in \mathbb{R}^M
\]

Thus we have a basic orthogonal expansion for $Y \in S^M$

\[
Y = \sum_{j=1}^{M} \sum_{i=1}^{j} \langle E_{ij}, Y \rangle E_{ij}
\]

whose unique coefficients

\[
\langle E_{ij}, Y \rangle = \begin{cases} Y_{ii}, & i = 1 \ldots M \\ Y_{ij} \sqrt{2}, & 1 \leq i < j \leq M \end{cases}
\]

(62)

 correspond to entries of the symmetric vectorization (56).

2.2.3  Symmetric hollow subspace

2.2.3.0.1  Definition.  Hollow subspaces.  [364]

Define a subspace of $\mathbb{R}^{M \times M}$: the convex set of all (real) symmetric $M \times M$ matrices having 0 main diagonal:

\[
\mathbb{R}_h^{M \times M} \triangleq \left\{ A \in \mathbb{R}^{M \times M} \mid A = A^T, \quad \delta(A) = 0 \right\} \subset \mathbb{R}^{M \times M}
\]

where the main diagonal of $A \in \mathbb{R}^{M \times M}$ is denoted (§A.1)

\[
\delta(A) \in \mathbb{R}^M
\]

(1497)

Operating on a vector, linear operator $\delta$ naturally returns a diagonal matrix: $\delta(\delta(A))$ is a diagonal matrix. Operating recursively on a vector $\Lambda \in \mathbb{R}^N$ or diagonal matrix $\Lambda \in S^N$, operator $\delta(\delta(\Lambda))$ returns $\Lambda$ itself:

\[
\delta^2(\Lambda) \equiv \delta(\delta(\Lambda)) = \Lambda
\]

(1499)

The subspace $\mathbb{R}_h^{M \times M}$ (63) comprising (real) symmetric hollow matrices is isomorphic with subspace $\mathbb{R}^{M(M-1)/2}$; its orthogonal complement is

\[
\mathbb{R}_h^{M \times M \perp} \triangleq \left\{ A \in \mathbb{R}^{M \times M} \mid A = -A^T + 2\delta^2(A) \right\} \subset \mathbb{R}^{M \times M}
\]

(64)
2.2. VECTORIZED-MATRIX INNER PRODUCT

the subspace of antisymmetric antihollow matrices in \( \mathbb{R}^{M \times M} \); id est,

\[
\mathbb{R}_h^{M \times M} \oplus \mathbb{R}^{M \times M}_h = \mathbb{R}^{M \times M}
\]  

(65)

Yet defined instead as a proper subspace of ambient \( \mathbb{S}^M \)

\[
\mathbb{S}^M_h \overset{\triangle}{=} \left\{ A \in \mathbb{S}^M \mid \delta(A) = 0 \right\} = \mathbb{R}^{M \times M}_h \subset \mathbb{S}^M
\]  

(66)

the orthogonal complement \( \mathbb{S}^M_h \) of symmetric hollow subspace \( \mathbb{S}^M_h \),

\[
\mathbb{S}^M_h \overset{\triangle}{=} \left\{ A \in \mathbb{S}^M \mid A = \delta^2(A) \right\} \subseteq \mathbb{S}^M
\]  

(67)

called symmetric antihollow subspace, is simply the subspace of diagonal matrices; id est,

\[
\mathbb{S}^M_h \oplus \mathbb{S}^M_h = \mathbb{S}^M
\]  

(68)

Any matrix \( A \in \mathbb{R}^{M \times M} \) can be written as a sum of its symmetric hollow and antisymmetric antihollow parts: respectively,

\[
A = \left( \frac{1}{2} (A + A^T) - \delta^2(A) \right) + \left( \frac{1}{2} (A - A^T) + \delta^2(A) \right)
\]  

(69)

The symmetric hollow part is orthogonal to the antisymmetric antihollow part in \( \mathbb{R}^{M^2} \); videlicet,

\[
\text{tr} \left( \left( \frac{1}{2} (A + A^T) - \delta^2(A) \right) \left( \frac{1}{2} (A - A^T) + \delta^2(A) \right) \right) = 0
\]  

(70)

because any matrix in subspace \( \mathbb{R}^{M \times M}_h \) is orthogonal to any matrix in the antisymmetric antihollow subspace

\[
\mathbb{R}^{M \times M}_h \overset{\triangle}{=} \left\{ \frac{1}{2} (A - A^T) + \delta^2(A) \mid A \in \mathbb{R}^{M \times M} \right\} \subseteq \mathbb{R}^{M \times M}
\]  

(71)

of the ambient space of real matrices; which reduces to the diagonal matrices in the ambient space of symmetric matrices

\[
\mathbb{S}^M_h \overset{\triangle}{=} \left\{ \delta^2(A) \mid A \in \mathbb{S}^M \right\} = \left\{ \delta(u) \mid u \in \mathbb{R}^M \right\} \subseteq \mathbb{S}^M
\]  

(72)

In anticipation of their utility with Euclidean distance matrices (EDMs) in §5, for symmetric hollow matrices we introduce the linear bijective vectorization \( \text{dvec} \) that is the natural analogue to symmetric matrix vectorization \( \text{svec} \) (56): for \( Y = [Y_{ij}] \in \mathbb{S}^M_h \)

\[
\text{dvec} Y \overset{\triangle}{=} \sqrt{2} \begin{bmatrix} Y_{12} \\ Y_{13} \\ Y_{23} \\ Y_{14} \\ Y_{24} \\ \vdots \\ Y_{M-1,M} \end{bmatrix} \in \mathbb{R}^{M(M-1)/2}
\]  

(73)
Figure 21: Convex hull of a random list of points in $\mathbb{R}^3$. Some points from that generating list reside interior to this convex polyhedron (§2.12). [405, Convex Polyhedron] (Avis-Fukuda-Mizukoshi)

Like $\text{svec}$, $\text{dvec}$ is an isometric isomorphism on the symmetric hollow subspace. For $X \in S^M_h$

$$\| \text{dvec } X - \text{dvec } Y \|_2 = \| X - Y \|_F$$

(74)

The set of all symmetric hollow matrices $S^M_h$ forms a proper subspace in $\mathbb{R}^{M \times M}$, so for it there must be a standard orthonormal basis in isometrically isomorphic $\mathbb{R}^{M(M-1)/2}$

$$\{ E_{ij} \in S^M_h \} = \left\{ \frac{1}{\sqrt{2}} (e_i e_j^T + e_j e_i^T) , \ 1 \leq i < j \leq M \right\}$$

(75)

where $M(M-1)/2$ standard basis matrices $E_{ij}$ are formed from the standard basis vectors $e_i \in \mathbb{R}^M$.

The symmetric hollow majorization corollary A.1.2.0.2 characterizes eigenvalues of symmetric hollow matrices.

## 2.3 Hulls

We focus on the affine, convex, and conic hulls: convex sets that may be regarded as kinds of Euclidean container or vessel united with its interior.

### 2.3.1 Affine hull, affine dimension

Affine dimension of any set in $\mathbb{R}^n$ is the dimension of the smallest affine set (empty set, point, line, plane, hyperplane (§2.4.2), translated subspace, $\mathbb{R}^n$) that contains it. For nonempty sets, affine dimension is the same as dimension of the subspace parallel to that affine set. [318, §1] [209, §A.2.1]

Ascribe the points in a list $\{ x_\ell \in \mathbb{R}^n, \ \ell = 1 \ldots N \}$ to the columns of matrix $X$:

$$X = [x_1 \cdots x_N] \in \mathbb{R}^{n \times N}$$

(76)
In particular, we define affine dimension \( r \) of the \( N \)-point list \( X \) as dimension of the smallest affine set in Euclidean space \( \mathbb{R}^n \) that contains \( X \);

\[
r \triangleq \dim \text{aff} X
\]  

(77)

Affine dimension \( r \) is a lower bound sometimes called embedding dimension. \([364][195]\)

That affine set \( \mathcal{A} \) in which those points are embedded is unique and called the affine hull \([340], \S 2.1\];

\[
\mathcal{A} \triangleq \text{aff} \{ x_\ell \in \mathbb{R}^n, \ell = 1 \ldots N \} = \text{aff} X = x_1 + \mathcal{R}\{x_\ell - x_1, \ell = 2 \ldots N \} = \{ Xa | a^T 1 = 1 \} \subseteq \mathbb{R}^n
\]  

(78)

for which we call list \( X \) a set of generators. Hull \( \mathcal{A} \) is parallel to subspace

\[
\mathcal{R}\{x_\ell - x_1, \ell = 2 \ldots N \} = \mathcal{R}(X - x_1 1^T) \subseteq \mathbb{R}^n
\]  

(79)

where

\[
\mathcal{R}(A) = \{ Ax | \forall x \}\]

(142)

Given some arbitrary set \( \mathcal{C} \) and any \( x \in \mathcal{C} \)

\[
\text{aff} \mathcal{C} = x + \text{aff}(\mathcal{C} - x)
\]  

(80)

where \( \text{aff}(\mathcal{C} - x) \) is a subspace.

\[
\text{aff} \emptyset \triangleq \emptyset
\]  

(81)

The affine hull of a point \( x \) is that point itself;

\[
\text{aff}\{x\} = \{x\}
\]  

(82)

Affine hull of two distinct points is the unique line through them. (Figure 22) The affine hull of three noncollinear points in any dimension is that unique plane containing the points, and so on. The subspace of symmetric matrices \( \mathbb{S}^m \) is the affine hull of the cone of positive semidefinite matrices; \( \S 2.9\)

\[
\text{aff} \mathbb{S}_+^m = \mathbb{S}^m
\]  

(83)

### 2.3.1.0.1 Example. Affine hull of rank-1 correlation matrices.

[228]

The set of all \( m \times m \) rank-1 correlation matrices is defined by all the binary vectors \( y \) in \( \mathbb{R}^m \) (confer \S 5.9.1.0.1)

\[
\{ yy^T \in \mathbb{S}_+^m | \delta(yy^T) = 1 \}
\]  

(84)

Affine hull of the rank-1 correlation matrices is equal to the set of normalized symmetric matrices; \( \text{id est} \),

\[
\text{aff} \{ yy^T \in \mathbb{S}_+^m | \delta(yy^T) = 1 \} = \{ A \in \mathbb{S}^m | \delta(A) = 1 \}
\]  

(85)
Figure 22: Given two points in Euclidean vector space of any dimension, their various hulls are illustrated. Each hull is a subset of range; generally, $A, C, K \subseteq \mathcal{R} \ni 0$. (Cartesian axes drawn for reference.)
2.3. HULLS

2.3.1.0.2 Exercise. Affine hull of correlation matrices.
Prove (85) via definition of affine hull. Find the convex hull instead. ▽

2.3.1 Partial order induced by $\mathbb{R}_+^N$ and $\mathbb{S}_+^M$

Notation $a \succeq 0$ means vector $a$ belongs to nonnegative orthant $\mathbb{R}_+^N$ while $a \succ 0$ means vector $a$ belongs to the nonnegative orthant’s interior $\text{int} \mathbb{R}_+^N$. $a \succeq b$ denotes comparison of vector $a$ to vector $b$ on $\mathbb{R}^N$ with respect to the nonnegative orthant; id est, $a \succeq b$ means $a - b$ belongs to the nonnegative orthant but neither $a$ or $b$ is necessarily nonnegative. With particular respect to the nonnegative orthant, $a \succeq b \iff a_i \geq b_i \ \forall \ i$ (369).

More generally, $a \succeq_K b$ denotes comparison with respect to pointed closed convex cone $K$, whereas comparison with respect to the cone’s interior is denoted $a \succ_K b$. But equivalence with entrywise comparison does not generally hold, and neither $a$ or $b$ necessarily belongs to $K$. (2.7.2.2)

The symbol $\geq$ is reserved for scalar comparison on the real line $\mathbb{R}$ with respect to the nonnegative real line $\mathbb{R}_+$ as in $a^T y \geq b$. Comparison of matrices with respect to the positive semidefinite cone $\mathbb{S}_+^M$, like $I \succeq A \succeq 0$ in Example 2.3.2.0.1, is explained in §2.9.0.1.

2.3.2 Convex hull

The convex hull [209, §A.1.4] [318] of any bounded list or set of $N$ points $X \in \mathbb{R}^{n \times N}$ forms a unique bounded convex polyhedron (confer §2.12.0.0.1) whose vertices constitute some subset of that list;

$$\mathcal{P} \triangleq \text{conv} \{x_\ell, \ell = 1 \ldots N\} = \text{conv} X = \{Xa | a^T 1 = 1, \ a \succeq 0\} \subseteq \mathbb{R}^n \quad (86)$$

Union of relative interior and relative boundary (§2.1.7.2) of the polyhedron comprise its convex hull $\mathcal{P}$, the smallest closed convex set that contains the list $X$; e.g., Figure 21. Given $\mathcal{P}$, the generating list $\{x_\ell\}$ is not unique. But because every bounded polyhedron is the convex hull of its vertices, $\{340, \text{§}2.12.2\}$ the vertices of $\mathcal{P}$ comprise a minimal set of generators.

Given some arbitrary set $\mathcal{C} \subseteq \mathbb{R}^n$, its convex hull $\text{conv} \mathcal{C}$ is equivalent to the smallest convex set containing it. (confer §2.4.1.1.1) The convex hull is a subset of the affine hull;

$$\text{conv} \mathcal{C} \subseteq \text{aff} \mathcal{C} = \text{aff} \partial \mathcal{C} = \text{aff} \mathcal{C} = \text{aff} \text{conv} \mathcal{C} \quad (87)$$

Any closed bounded convex set $\mathcal{C}$ is equal to the convex hull of its boundary;

$$\mathcal{C} = \text{conv} \partial \mathcal{C} \quad (88)$$

$$\text{conv} \emptyset \triangleq \emptyset \quad (89)$$

\textbf{2.15} An arbitrary set $\mathcal{C}$ in $\mathbb{R}^n$ is bounded iff it can be contained in a Euclidean ball having finite radius. [119, §2.2] (confer §5.7.3.0.1) The smallest ball containing $\mathcal{C}$ has radius $\inf_{x \in \mathcal{C}} \sup_{y \in \mathcal{C}} \|x - y\|$ and center $x^*$ whose determination is a convex problem because $\sup_{y \in \mathcal{C}} \|x - y\|$ is a convex function of $x$; but the supremum may be difficult to ascertain.
Figure 23: Two Fantopes. Circle (radius $1/\sqrt{2}$), shown here on boundary of positive semidefinite cone $S^2_+$ in isometrically isomorphic $\mathbb{R}^3$ from Figure 45, comprises boundary of a Fantope (90) in this dimension ($k = 1$, $N = 2$). Lone point illustrated is Identity matrix $I$, interior to PSD cone, and is that Fantope corresponding to $k = 2$, $N = 2$. (View is from inside PSD cone looking toward origin.)
2.3. Hulls

2.3.2.0.1 Example. Hull of rank-$k$ projection matrices. [150] [298] [12, §4.1] [303, §3] [249, §2.4] [250] Convex hull of the set comprising outer product of orthonormal matrices has equivalent expression: for $1 \le k \le N$ (§2.9.0.1)

$$\text{conv}\left\{UU^T \mid U \in \mathbb{R}^{N \times k}, U^TU = I \right\} = \left\{ A \in S^N \mid I \succeq A \succeq 0, \langle I, A \rangle = k \right\} \subset S^N_+$$ (90)

This important convex body we call Fantope (after mathematician Ky Fan). In case $k = 1$, there is slight simplification: ((1703), Example 2.9.2.7.1)

$$\text{conv}\left\{UU^T \mid U \in \mathbb{R}^N, U^TU = 1 \right\} = \left\{ A \in S^N \mid A \succeq 0, \langle I, A \rangle = 1 \right\}$$ (91)

This particular Fantope is called spectahedron. [sic] [157, §5.1] In case $k = N$, the Fantope is Identity matrix $I$. More generally, the set

$$\left\{UU^T \mid U \in \mathbb{R}^{N \times k}, U^TU = I \right\}$$ (92)

comprises the extreme points (§2.6.0.0.1) of its convex hull. By (1545), each and every extreme point $UU^T$ has only $k$ nonzero eigenvalues $\lambda$ and they all equal 1; id est, $\lambda(UU^T)_{1:k} = \lambda(U^TU) = 1$. So Frobenius’ norm of each and every extreme point equals the same constant

$$\|UU^T\|_F^2 = k$$ (93)

Each extreme point simultaneously lies on the boundary of the positive semidefinite cone (when $k < N$, §2.9) and on the boundary of a hypersphere of dimension $k(N - \frac{k}{2} + \frac{1}{2})$ and radius $\sqrt{k(1 - \frac{k}{N})}$ centered at $\frac{k}{N}I$ along the ray (base 0) through the Identity matrix $I$ in isomorphic vector space $\mathbb{R}^{N(N+1)/2}$ (§2.2.2.1).

Figure 23 illustrates extreme points (92) comprising the boundary of a Fantope, the boundary of a disc corresponding to $k = 1$, $N = 2$; but that circumscription does not hold in higher dimension. (§2.9.2.8)

2.3.2.0.2 Example. Nuclear norm ball: convex hull of rank-1 matrices.

From (91), in Example 2.3.2.0.1, we learn that the convex hull of normalized symmetric rank-1 matrices is a slice of the positive semidefinite cone. In §2.9.2.7 we find the convex hull of all symmetric rank-1 matrices to be the entire positive semidefinite cone.

In the present example we abandon symmetry; instead posing, what is the convex hull of bounded nonsymmetric rank-1 matrices:

$$\text{conv}\{uv^T \mid \|uv^T\| \le 1, u \in \mathbb{R}^m, v \in \mathbb{R}^n\} = \{ X \in \mathbb{R}^{m \times n} \mid \sum_i \sigma(X)_i \le 1 \}$$ (94)

where $\sigma(X)$ is a vector of singular values. (Since $\|uv^T\| = \|u\|\|v\|$ (1694), norm of each vector constituting a dyad $uv^T$ (§B.1) in the hull is effectively bounded above by 1.)
\( \{ \text{svec } X \mid \| X \|_2^* \leq 1 \} \)

\[ X = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \]

Figure 24: *Nuclear norm* is a sum of singular values; \( \| X \|_2^* \triangleq \sum \sigma(X)_i \). Nuclear norm ball, in the subspace of \( 2 \times 2 \) symmetric matrices, is a truncated cylinder in isometrically isomorphic \( \mathbb{R}^3 \).
2.3.  HULLS

Figure 25: $w_p^T$ is a convex combination of normalized dyads $||w_p^T|| = 1$; similarly for $x_p^T$. Any point in line segment joining $x_p^T$ to $w_p^T$ is expressible as a convex combination of two to four points indicated on boundary.

Proof. (⇐) Suppose $\sum \sigma(X)_i \leq 1$. Decompose $X = U\Sigma V^T$ by SVD (§A.6) where $U = [u_1 \ldots u_{\min\{m,n\}}] \in \mathbb{R}^{m \times \min\{m,n\}}$, $V = [v_1 \ldots v_{\min\{m,n\}}] \in \mathbb{R}^{n \times \min\{m,n\}}$, and whose sum of singular values is $\sum \sigma(X)_i = \text{tr} \Sigma = \kappa \leq 1$. Then we may write $X = \sum \frac{1}{\sqrt{\kappa}} u_i \sqrt{\kappa} v_i^T$ which is a convex combination of dyads each of whose norm does not exceed 1. (Srebro)

(⇒) Now suppose we are given a convex combination of dyads $X = \sum \alpha_i u_i v_i^T$ such that $\sum \alpha_i = 1$, $\alpha_i \geq 0 \forall i$, and $||u_i v_i^T|| \leq 1 \forall i$. Then by triangle inequality for singular values [213, cor.3.4.3] $\sum \sigma(X)_i \leq \sum \sigma(u_i v_i^T) = \sum \alpha_i ||u_i v_i^T|| \leq \sum \alpha_i$.

Given any particular dyad $w_p^T$ in the convex hull, because its polar $-w_p^T$ and every convex combination of the two belong to that hull, then the unique line containing those two points $\pm w_p^T$ (their affine combination (78)) must intersect the hull’s boundary at the normalized dyads $\{\pm w_p^T | \|w_p^T\| = 1\}$. Any point formed by convex combination of dyads in the hull must therefore be expressible as a convex combination of dyads on the boundary: Figure 25.

\[
\text{conv}\{w^T | \|w^T\| \leq 1, \ u \in \mathbb{R}^m, \ v \in \mathbb{R}^n\} \equiv \text{conv}\{w^T | \|w^T\| = 1, \ u \in \mathbb{R}^m, \ v \in \mathbb{R}^n\}
\]  

(95)

id est, dyads may be normalized and the hull’s boundary contains them:

\[
\partial\{X \in \mathbb{R}^{m \times n} | \sum_i \sigma(X)_i \leq 1\} = \{X \in \mathbb{R}^{m \times n} | \sum_i \sigma(X)_i = 1\} \supset \{w^T | \|w^T\| = 1, \ u \in \mathbb{R}^m, \ v \in \mathbb{R}^n\}
\]  

(96)

Normalized dyads constitute the set of extreme points (§2.6.0.0.1) of this nuclear norm ball (confer Figure 24) which is, therefore, their convex hull. \[\square\]
2.3.2.0.3 Exercise. Convex hull of outer product.
Describe the interior of a Fantope.
Find the convex hull of nonorthogonal projection matrices (§E.1.1):
\[
\{ U V^T | U \in \mathbb{R}^{N \times k}, V \in \mathbb{R}^{N \times k}, V^T U = I \}
\] (97)

Find the convex hull of nonsymmetric matrices bounded under some norm:
\[
\{ U V^T | U \in \mathbb{R}^{m \times k}, V \in \mathbb{R}^{n \times k}, \| U V^T \| \leq 1 \}
\] (98)

2.3.2.0.4 Example. Permutation polyhedron.
A permutation matrix \( \Xi \) is formed by interchanging rows and columns of Identity matrix \( I \). Since \( \Xi \) is square and \( \Xi^T \Xi = I \), the set of all permutation matrices is a proper subset of the nonconvex manifold of orthogonal matrices (§B.5). In fact, the only orthogonal matrices having all nonnegative entries are permutations of the Identity:
\[
\Xi^{-1} = \Xi^T, \quad \Xi \geq 0
\] (99)

And the only positive semidefinite permutation matrix is the Identity. [343, §6.5 prob.20]

Regarding the permutation matrices as a set of points in Euclidean space, its convex hull is a bounded polyhedron (§2.12) described (Birkhoff, 1946)
\[
\text{conv}\{ \Xi = \Pi_i(I \in \mathbb{S}^n) \in \mathbb{R}^{n \times n}, i = 1 \ldots n! \} = \{ X \in \mathbb{R}^{n \times n} | X^T 1 = 1, X 1 = 1, X \geq 0 \}
\] (100)

where \( \Pi_i \) is a linear operator here representing the \( i \)th permutation. This polyhedral hull, whose \( n! \) vertices are the permutation matrices, is also known as the set of doubly stochastic matrices. The permutation matrices are the minimal cardinality (fewest nonzero entries) doubly stochastic matrices. The only orthogonal matrices belonging to this polyhedron are the permutation matrices.

It is remarkable that \( n! \) permutation matrices can be described as the extreme points (§2.6.0.0.1) of a bounded polyhedron, of affine dimension \((n-1)^2\), that is itself described by \( 2n \) equalities \((2n-1) \) linearly independent equality constraints in \( n^2 \) nonnegative variables). By Carathéodory’s theorem, conversely, any doubly stochastic matrix can be described as a convex combination of at most \((n-1)^2+1\) permutation matrices. [212, §8.7] [57, thm.1.2.5] This polyhedron, then, can be a device for relaxing an integer, combinatorial, or Boolean optimization problem.²¹⁶ [68] [294, §3.1]

2.3.2.0.5 Example. Convex hull of orthonormal matrices.
Consider rank-\( k \) matrices \( U \in \mathbb{R}^{n \times k} \) such that \( U^T U = I \). These are the orthonormal matrices; a closed bounded submanifold, of all orthogonal matrices, having dimension \( nk - \frac{1}{2} k(k+1) \) [54]. Their convex hull is expressed, for \( 1 \leq k \leq n \)
\[
\text{conv}\{ U \in \mathbb{R}^{n \times k} | U^T U = I \} = \{ X \in \mathbb{R}^{n \times k} | \| X \|_2 \leq 1 \} = \{ X \in \mathbb{R}^{n \times k} | \| X^T a \| \leq \| a \| \ \forall a \in \mathbb{R}^{n} \}
\] (101)

²¹⁶ Relaxation replaces an objective function with its convex envelope or expands a feasible set to one that is convex. Dantzig first showed in 1951 that, by this device, the so-called assignment problem can be formulated as a linear program. [327] [27, §II.5]
By Schur complement (§A.4), the spectral norm \( \|X\|_2 \) constraining largest singular value \( \sigma(X) \) can be expressed as a semidefinite constraint

\[
\|X\|_2 \leq 1 \iff \begin{bmatrix}
I & X \\
X^T & I
\end{bmatrix} \succeq 0
\]

(102)

because of equivalence \( X^T X \preceq I \iff \sigma(X) \leq 1 \) with singular values. (1648) (1532) (1533)

When \( k = n \), matrices \( U \) are orthogonal and their convex hull is called the spectral norm ball which is the set of all contractions. [213, p.158] [339, p.313] The orthogonal matrices then constitute the extreme points (§2.6.0.0.1) of this hull. Hull intersection with the nonnegative orthant \( \mathbb{R}^{n \times n}_+ \) contains the permutation polyhedron (100).

\[\square\]

### 2.3.3 Conic hull

In terms of a finite-length point list (or set) arranged columnar in \( X \in \mathbb{R}^{n \times N} \) (76), its conic hull is expressed

\[
\mathcal{K} \triangleq \text{cone} \left\{ x_\ell, \ell = 1 \ldots N \right\} = \text{cone} X = \{ Xa \mid a \succeq 0 \} \subseteq \mathbb{R}^n
\]

(103)

id est, every nonnegative combination of points from the list. Conic hull of any finite-length list forms a polyhedral cone [209, §A.4.3] (§2.12.1.0.1; e.g., Figure 52a); the smallest closed convex cone (§2.7.2) that contains the list.

By convention, the aberration [340, §2.1]

\[
\text{cone} \emptyset \triangleq \{ 0 \}
\]

(104)

Given some arbitrary set \( \mathcal{C} \), it is apparent

\[
\text{conv} \mathcal{C} \subseteq \text{cone} \mathcal{C}
\]

(105)

### 2.3.4 Vertex-description

The conditions in (78), (86), and (103) respectively define an affine combination, convex combination, and conic combination of elements from the set or list. Whenever a Euclidean body can be described as some hull or span of a set of points, then that representation is loosely called a vertex-description and those points are called generators.

### 2.4 Halfspace, Hyperplane

A two-dimensional affine subset is called a plane. An \((n - 1)\)-dimensional affine subset of \( \mathbb{R}^n \) is called a hyperplane. [318] [209] Every hyperplane partially bounds a halfspace (which is convex, but not affine, and the only nonempty convex set in \( \mathbb{R}^n \) whose complement is convex and nonempty).
Figure 26: A simplicial cone (§2.12.3.1.1) in $\mathbb{R}^3$ whose boundary is drawn truncated; constructed using $A \in \mathbb{R}^{3 \times 3}$ and $C = 0$ in (286). By the most fundamental definition of a cone (§2.7.1), entire boundary can be constructed from an aggregate of rays emanating exclusively from the origin. Each of three extreme directions corresponds to an edge (§2.6.0.0.3); they are conically, affinely, and linearly independent for this cone. Because this set is polyhedral, exposed directions are in one-to-one correspondence with extreme directions; there are only three. Its extreme directions give rise to what is called a vertex-description of this polyhedral cone; simply, the conic hull of extreme directions. Obviously this cone can also be constructed by intersection of three halfspaces; hence the equivalent halfspace-description.
\[ \mathcal{H}_+ = \{ y \mid a^T(y - y_p) \geq 0 \} \]

\[ \mathcal{H}_- = \{ y \mid a^T(y - y_p) \leq 0 \} \]

\[ \mathcal{N}(a^T) = \{ y \mid a^T y = 0 \} \]

\[ \partial \mathcal{H} = \{ y \mid a^T(y - y_p) = 0 \} = \mathcal{N}(a^T) + y_p \]

Figure 27: Hyperplane illustrated \( \partial \mathcal{H} \) is a line partially bounding halfspaces \( \mathcal{H}_- \) and \( \mathcal{H}_+ \) in \( \mathbb{R}^2 \). Shaded is a rectangular piece of semiinfinite \( \mathcal{H}_- \) with respect to which vector \( a \) is outward-normal to bounding hyperplane; vector \( a \) is inward-normal with respect to \( \mathcal{H}_+ \). Halfspace \( \mathcal{H}_- \) contains nullspace \( \mathcal{N}(a^T) \) (dashed line through origin) because \( a^T y_p > 0 \). Hyperplane, halfspace, and nullspace are each drawn truncated. Points \( c \) and \( d \) are equidistant from hyperplane, and vector \( c - d \) is normal to it. \( \Delta \) is distance from origin to hyperplane.
2.4.1 Halfspaces $\mathcal{H}_+$ and $\mathcal{H}_-$

Euclidean space $\mathbb{R}^n$ is partitioned in two by any hyperplane $\partial \mathcal{H}$; id est, $\mathcal{H}_- + \mathcal{H}_+ = \mathbb{R}^n$. The resulting (closed convex) halfspaces, both partially bounded by $\partial \mathcal{H}$, may be described

$$\mathcal{H}_- = \{ y \mid a^T y \leq b \} = \{ y \mid a^T(y - y_p) \leq 0 \} \subset \mathbb{R}^n$$  \hspace{1cm} (106)

$$\mathcal{H}_+ = \{ y \mid a^T y \geq b \} = \{ y \mid a^T(y - y_p) \geq 0 \} \subset \mathbb{R}^n$$  \hspace{1cm} (107)

where nonzero vector $a \in \mathbb{R}^n$ is an outward-normal to the hyperplane partially bounding $\mathcal{H}_-$ while an inward-normal with respect to $\mathcal{H}_+$. For any vector $y - y_p$ that makes an obtuse angle with normal $a$, vector $y$ will lie in the halfspace $\mathcal{H}_-$ on one side (shaded in Figure 27) of the hyperplane while acute angles denote $y$ in $\mathcal{H}_+$ on the other side.

An equivalent more intuitive representation of a halfspace comes about when we consider all the points in $\mathbb{R}^n$ closer to point $d$ than to point $c$ or equidistant, in the Euclidean sense; from Figure 27,

$$\mathcal{H}_- = \{ y \mid \|y - d\| \leq \|y - c\| \}$$  \hspace{1cm} (108)

This representation, in terms of proximity, is resolved with the more conventional representation of a halfspace (106) by squaring both sides of the inequality in (108):

$$\mathcal{H}_- = \left\{ y \mid (c - d)^T y \leq \frac{\|c\|^2 - \|d\|^2}{2} \right\} = \left\{ y \mid (c - d)^T \left( y - \frac{c + d}{2} \right) \leq 0 \right\}$$  \hspace{1cm} (109)

2.4.1.1 PRINCIPLE 1: Halfspace-description of convex sets

The most fundamental principle in convex geometry follows from the geometric Hahn-Banach theorem [260, §5.12] [19, §1] [139, §I.1.2] which guarantees any closed convex set to be an intersection of halfspaces.

2.4.1.1.1 Theorem. Halfspaces.  \hspace{1cm} [209, §A.4.2b] [43, §2.4]

A closed convex set in $\mathbb{R}^n$ is equivalent to the intersection of all halfspaces that contain it. $\Diamond$

Intersection of multiple halfspaces in $\mathbb{R}^n$ may be represented using a matrix constant $A$

$$\bigcap_i \mathcal{H}_- = \{ y \mid A^T y \leq b \} = \{ y \mid A^T(y - y_p) \leq 0 \}$$  \hspace{1cm} (110)

$$\bigcap_i \mathcal{H}_+ = \{ y \mid A^T y \geq b \} = \{ y \mid A^T(y - y_p) \geq 0 \}$$  \hspace{1cm} (111)

where $b$ is now a vector, and the $i^{th}$ column of $A$ is normal to a hyperplane $\partial \mathcal{H}_i$ partially bounding $\mathcal{H}_i$. By the halfspaces theorem, intersections like this can describe interesting convex Euclidean bodies such as polyhedra and cones, giving rise to the term halfspace-description.
Figure 28: (a)-(d) Hyperplanes in $\mathbb{R}^2$ (truncated) redundantly emphasize: hyperplane movement opposite to its normal direction minimizes vector inner-product. This concept is exploited to attain analytical solution of linear programs by visual inspection; e.g., §2.4.2.6.2, §2.5.1.2.2, §3.4.0.0.2, [63, exer.4.8-exer.4.20]. Each graph is also interpretable as contour plot of a real affine function of two variables as in Figure 75. (e) $|\beta|/\|\alpha\|$ from $\partial H = \{ x \mid \alpha^T x = \beta \}$ represents radius of hypersphere about 0 supported by any hyperplane with same ratio $|\text{inner product}|/\|\alpha\|$. 

(a) $a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(b) $b = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

(c) $c = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

(d) $d = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(e) $e = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Figure 28: (a)-(d) Hyperplanes in $\mathbb{R}^2$ (truncated) redundantly emphasize: hyperplane movement opposite to its normal direction minimizes vector inner-product. This concept is exploited to attain analytical solution of linear programs by visual inspection; e.g., §2.4.2.6.2, §2.5.1.2.2, §3.4.0.0.2, [63, exer.4.8-exer.4.20]. Each graph is also interpretable as contour plot of a real affine function of two variables as in Figure 75. (e) $|\beta|/\|\alpha\|$ from $\partial H = \{ x \mid \alpha^T x = \beta \}$ represents radius of hypersphere about 0 supported by any hyperplane with same ratio $|\text{inner product}|/\|\alpha\|$.
2.4.2 Hyperplane $\partial \mathcal{H}$ representations

Every hyperplane $\partial \mathcal{H}$ is an affine set parallel to an $(n-1)$-dimensional subspace of $\mathbb{R}^n$; it is itself a subspace if and only if it contains the origin.

$$\dim \partial \mathcal{H} = n - 1 \quad (112)$$

so a hyperplane is a point in $\mathbb{R}$, a line in $\mathbb{R}^2$, a plane in $\mathbb{R}^3$, and so on. Every hyperplane can be described as the intersection of complementary halfspaces; [318, §19]

$$\partial \mathcal{H} = \mathcal{H}_- \cap \mathcal{H}_+ = \{ y \mid a^T y \leq b , \ a^T y \geq b \} = \{ y \mid a^T y = b \} \quad (113)$$

a halfspace-description. Assuming normal $a \in \mathbb{R}^n$ to be nonzero, then any hyperplane in $\mathbb{R}^n$ can be described as the solution set to vector equation $a^T y = b$ (illustrated in Figure 27 and Figure 28 for $\mathbb{R}^2$);

$$\partial \mathcal{H} \triangleq \{ y \mid a^T y = b \} = \{ y \mid a^T (y - y_p) = 0 \} = \{ Z \xi + y_p \mid \xi \in \mathbb{R}^{n-1} \} \subset \mathbb{R}^n \quad (114)$$

All solutions $y$ constituting the hyperplane are offset from the nullspace of $a^T$ by the same constant vector $y_p \in \mathbb{R}^n$ that is any particular solution to $a^T y = b$; id est,

$$y = Z \xi + y_p \quad (115)$$

where the columns of $Z \in \mathbb{R}^{n \times n-1}$ constitute a basis for the nullspace $\mathcal{N}(a^T) = \{ x \in \mathbb{R}^n \mid a^T x = 0 \}$. 2.17

Conversely, given any point $y_p$ in $\mathbb{R}^n$, the unique hyperplane containing it having normal $a$ is the affine set $\partial \mathcal{H}$ (114) where $b$ equals $a^T y_p$ and where a basis for $\mathcal{N}(a^T)$ is arranged in $Z$ columnar. Hyperplane dimension is apparent from dimension of $Z$; that hyperplane is parallel to the span of its columns.

2.4.2.0.1 Exercise. Hyperplane scaling.

Given normal $y$, draw a hyperplane $\{ x \in \mathbb{R}^2 \mid x^T y = 1 \}$. Suppose $z = \frac{1}{2} y$. On the same plot, draw the hyperplane $\{ x \in \mathbb{R}^2 \mid x^T z = 1 \}$. Now suppose $z = 2 y$, then draw the last hyperplane again with this new $z$. What is the apparent effect of scaling normal $y$? ▼

2.4.2.0.2 Example. Distance from origin to hyperplane.

Given the (shortest) distance $\Delta \in \mathbb{R}^+$ from the origin to a hyperplane having normal vector $a$, we can find its representation $\partial \mathcal{H}$ by dropping a perpendicular. The point thus found is the orthogonal projection of the origin on $\partial \mathcal{H}$ (§5.0.0.5), equal to $a \Delta / ||a||$ if the origin is known a priori to belong to halfspace $\mathcal{H}_-$ (Figure 27), or equal to $-a \Delta / ||a||$ if the origin belongs to halfspace $\mathcal{H}_+$; id est, when $\mathcal{H}_- \ni 0$

$$\partial \mathcal{H} = \{ y \mid a^T (y - a \Delta / ||a||) = 0 \} = \{ y \mid a^T y = ||a|| \Delta \} \quad (116)$$

We will later find this expression for $y$ in terms of nullspace of $a^T$ (more generally, of matrix $A^T$ (144)) to be a useful trick (a practical device) for eliminating affine equality constraints, much as we did here.
or when $H_+ \ni 0$

$$\partial H = \{y \mid a^T(y + a\Delta/\|a\|) = 0\} = \{y \mid a^Ty = -\|a\|\Delta\} \quad (117)$$

Knowledge of only distance $\Delta$ and normal $a$ thus introduces ambiguity into
the hyperplane representation.

### 2.4.2.1 Matrix variable

Any halfspace in $\mathbb{R}^{mn}$ may be represented using a matrix variable. For variable $Y \in \mathbb{R}^{m \times n}$, given constants $A \in \mathbb{R}^{m \times n}$ and $b = \langle A, Y_p \rangle \in \mathbb{R}$

$$H_- = \{Y \in \mathbb{R}^{mn} \mid \langle A, Y \rangle \leq b\} = \{Y \in \mathbb{R}^{mn} \mid \langle A, Y - Y_p \rangle \leq 0\} \quad (118)$$

$$H_+ = \{Y \in \mathbb{R}^{mn} \mid \langle A, Y \rangle \geq b\} = \{Y \in \mathbb{R}^{mn} \mid \langle A, Y - Y_p \rangle \geq 0\} \quad (119)$$

Recall vector inner-product from §2.2:

$$\langle A, Y \rangle = \text{tr}(A^T Y) = \text{vec}(A)^T \text{vec}(Y).$$

Hyperplanes in $\mathbb{R}^{mn}$ may, of course, also be represented using matrix variables.

$$\partial H = \{Y \mid \langle A, Y \rangle = b\} = \{Y \mid \langle A, Y - Y_p \rangle = 0\} \subset \mathbb{R}^{mn} \quad (120)$$

Vector $a$ from Figure 27 is normal to the hyperplane illustrated. Likewise, nonzero vectorized matrix $A$ is normal to hyperplane $\partial H$;

$$A \perp \partial H \text{ in } \mathbb{R}^{mn} \quad (121)$$

### 2.4.2.2 Vertex-description of hyperplane

Any hyperplane in $\mathbb{R}^n$ may be described as affine hull of a minimal set of points

$$\partial H = \text{aff}\{x_\ell \in \mathbb{R}^n, \ell = 1\ldots n\}, \quad \text{dim aff}\{x_\ell \ \forall \ell\} = n-1$$

$$= x_1 + \mathcal{R}\{x_\ell - x_1, \ell = 2\ldots n\}, \quad \text{dim } \mathcal{R}\{x_\ell - x_1, \ell = 2\ldots n\} = n-1 \quad (122)$$

where

$$\mathcal{R}(A) = \{Ax \mid \forall x\} \quad (142)$$

### 2.4.2.3 Affine independence, minimal set

For any particular affine set, a minimal set of points constituting its vertex-description is an affinely independent generating set and vice versa.

Arbitrary given points $\{x_i \in \mathbb{R}^n, i = 1\ldots N\}$ are affinely independent (a.i.) if and only if, over all $\zeta \in \mathbb{R}^N \ni \zeta^T \mathbf{1} = 1$, $\zeta_k = 0 \in \mathbb{R}$ (confer §2.1.2)

$$x_i \zeta_i + \cdots + x_j \zeta_j - x_k = 0, \quad i \neq \cdots \neq j \neq k = 1\ldots N \quad (123)$$
Figure 29: Of three points illustrated, any one particular point does not belong to affine hull $A_i \ (i \in 1, 2, 3$, each drawn truncated) of points remaining. Three corresponding vectors in $\mathbb{R}^2$ are, therefore, affinely independent (but neither linearly or conically independent).

has no solution $\zeta$; in words, iff no point from the given set can be expressed as an affine combination of those remaining. We deduce

\[ \text{l.i.} \implies \text{a.i.} \quad (124) \]

Consequently, $\{x_i, i=1\ldots N\}$ is an affinely independent set if and only if $\{x_i - x_1, i=2\ldots N\}$ is a linearly independent (l.i.) set. [215, §3] (Figure 29) This is equivalent to the property that the columns of $\begin{bmatrix} X \\ 1^T \end{bmatrix}$ (for $X \in \mathbb{R}^{n \times N}$ as in (76)) form a linearly independent set. [209, §A.1.3]

Two nontrivial affine subsets are affinely independent iff their intersection is empty $\{\emptyset\}$ or, analogously to subspaces, they intersect only at a point.

### 2.4.2.4 Preservation of affine independence

Independence in the linear (§2.1.2.1), affine, and conic (§2.10.1) senses can be preserved under linear transformation. Suppose a matrix $X \in \mathbb{R}^{n \times N}$ (76) holds an affinely independent set in its columns. Consider a transformation on the domain of such matrices

\[ T(X) : \mathbb{R}^{n \times N} \to \mathbb{R}^{n \times N} \triangleq XY \quad (125) \]

where fixed matrix $Y \triangleq [y_1, y_2, \cdots, y_N] \in \mathbb{R}^{N \times N}$ represents linear operator $T$. Affine independence of $\{Xy_i \in \mathbb{R}^n, i=1\ldots N\}$ demands (by definition (123)) there exist no solution $\zeta \in \mathbb{R}^N \ni \zeta^T 1 = 1, \zeta_k = 0$, to

\[ Xy_i \zeta_i + \cdots + Xy_j \zeta_j - Xy_k = 0, \quad i \neq \cdots \neq j \neq k = 1\ldots N \quad (126) \]
2.4. HALFSPACE, HYPERPLANE

By factoring out $X$, we see that is ensured by affine independence of $\{y_i \in \mathbb{R}^N\}$ and by $\mathcal{R}(Y) \cap \mathcal{N}(X) = \mathbf{0}$ where

$$\mathcal{N}(A) = \{x \mid Ax = \mathbf{0}\} \quad (143)$$

2.4.2.5 Affine maps

Affine transformations preserve affine hulls. Given any affine mapping $T$ of vector spaces and some arbitrary set $\mathcal{C}$ [318, p.8]

$$\text{aff}(T \mathcal{C}) = T(\text{aff} \mathcal{C}) \quad (127)$$

2.4.2.6 PRINCIPLE 2: Supporting hyperplane

The second most fundamental principle of convex geometry also follows from the geometric Hahn-Banach theorem [260, §5.12] [19, §1] that guarantees existence of at least one
Figure 31: (a) Hyperplane \( \partial H_- \) (128) supporting closed set \( Y \subset \mathbb{R}^2 \). Vector \( a \) is inward-normal to hyperplane with respect to halfspace \( H_+ \), but outward-normal with respect to set \( Y \). A supporting hyperplane can be considered the limit of an increasing sequence in the normal-direction like that in Figure 30. (b) Hyperplane \( \partial H_+ \) nontraditionally supporting \( Y \). Vector \( \tilde{a} \) is inward-normal to hyperplane now with respect to both halfspace \( H_+ \) and set \( Y \). Tradition [209] [318] recognizes only positive normal polarity in support function \( \sigma_Y \) as in (129); \emph{id est}, normal \( a \), figure (a). But both interpretations of supporting hyperplane are useful.
2.4. HALFSPACE, HYPERPLANE

hyperplane in \( \mathbb{R}^n \) supporting a full-dimensional convex set at each point on its boundary.

The partial boundary \( \partial H \) of a halfspace that contains arbitrary set \( Y \) is called a supporting hyperplane \( \partial H \) to \( Y \) when the hyperplane contains at least one point of \( Y \). [318, §11]

2.4.2.6.1 Definition. Supporting hyperplane \( \partial H \).

Assuming set \( Y \) and some normal \( a \neq 0 \) reside in opposite halfspaces (Figure 31a), then a hyperplane supporting \( Y \) at point \( y_p \in \partial Y \) is described

\[
\partial H^- = \{ y \mid a^T(y - y_p) = 0, \ y_p \in \mathcal{Y}, \ a^T(z - y_p) \leq 0 \ \forall z \in \mathcal{Y} \} \tag{128}
\]

Given only normal \( a \), the hyperplane supporting \( Y \) is equivalently described

\[
\partial H^- = \{ y \mid a^T y = \sup \{ a^T z \mid z \in \mathcal{Y} \} \} \tag{129}
\]

where real function

\[
\sigma_Y(a) = \sup \{ a^T z \mid z \in \mathcal{Y} \} \tag{554}
\]

is called the support function for \( Y \).

Another equivalent but nontraditional representation for a supporting hyperplane is obtained by reversing polarity of normal \( a \); (1765)

\[
\partial H^+ = \{ y \mid \tilde{a}^T(y - y_p) = 0, \ y_p \in \mathcal{Y}, \ \tilde{a}^T(z - y_p) \geq 0 \ \forall z \in \mathcal{Y} \} \tag{130}
\]

where normal \( \tilde{a} \) and set \( \mathcal{Y} \) both now reside in \( \mathcal{H}_+ \) (Figure 31b).

When a supporting hyperplane contains only a single point of \( \mathcal{Y} \), that hyperplane is termed strictly supporting. [2.21]

A full-dimensional set that has a supporting hyperplane at every point on its boundary, conversely, is convex. A convex set \( C \subset \mathbb{R}^n \), for example, can be expressed as the intersection of all halfspaces partially bounded by hyperplanes supporting it; videlicet, [260, p.135]

\[
C = \bigcap_{a \in \mathbb{R}^n} \{ y \mid a^T y \leq \sigma_C(a) \} \tag{131}
\]

by the halfspaces theorem (§2.4.1.1.1).

There is no geometric difference between supporting hyperplane \( \partial H^+ \) or \( \partial H^- \) or \( \partial H \) and an ordinary hyperplane \( \partial H \) coincident with them.

2.18 It is customary to speak of a hyperplane supporting set \( C \) but not containing \( C \); called nontrivial support. [318, p.100] Hyperplanes in support of lower-dimensional bodies are admitted.

2.19 Normal \( a \) belongs to \( \mathcal{H}_+ \) by definition.

2.20 Useful for constructing the dual cone; e.g., Figure 57b. Tradition would instead have us construct the polar cone; which is, the negative dual cone.

2.21 Rockafellar terms a strictly supporting hyperplane tangent to \( Y \) if it is unique there; [318, §18, p.169] a definition we do not adopt because our only criterion for tangency is intersection exclusively with a relative boundary. Hiriart-Urruty & Lemaréchal [209, p.44] (confer [318, p.100]) do not demand any tangency of a supporting hyperplane.

2.22 If vector-normal polarity is unimportant, we may instead signify a supporting hyperplane by \( \partial \mathcal{H} \).
2.4.2.6.2 Example. Minimization over hypercube.
Consider minimization of a linear function over a hypercube, given vector $c$

$$\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad -1 \preceq x \preceq 1
\end{align*}$$

(132)

This convex optimization problem is called a linear program because the objective of minimization $c^T x$ is a linear function of variable $x$ and the constraints describe a polyhedron (intersection of a finite number of halfspaces and hyperplanes).

Any vector $x$ satisfying the constraints is called a feasible solution. Applying graphical concepts from Figure 28, Figure 30, and Figure 31, $x^* = -\text{sgn}(c)$ is an optimal solution to this minimization problem but is not necessarily unique. It generally holds for optimization problem solutions:

$$\text{optimal } \Rightarrow \text{feasible}$$

(133)

Because an optimal solution always exists at a hypercube vertex (§2.6.1.0.1) regardless of value of nonzero vector $c$ in (132) [97, p.158] [16, p.2], mathematicians see this geometry as a means to relax a discrete problem (whose desired solution is integer or combinatorial, confer Example 4.2.3.1.1). [249, §3.1] [250]

2.4.2.6.3 Exercise. Unbounded below.
Suppose instead we minimize over the unit hypersphere in Example 2.4.2.6.2; $\|x\| \leq 1$. What is an expression for optimal solution now? Is that program still linear?

Now suppose minimization of absolute value in (132). Are the following programs equivalent for some arbitrary real convex set $C$? (confer (516))

$$\begin{align*}
\text{minimize} & \quad |x| \\
\text{subject to} & \quad -1 \leq x \leq 1 \\
x & \in C
\end{align*} \quad = \quad 
\begin{align*}
\text{minimize} & \quad \alpha + \beta \\
\text{subject to} & \quad 1 \geq \beta \geq 0 \\
& \quad 1 \geq \alpha \geq 0 \\
& \quad \alpha - \beta \in C
\end{align*}$$

(134)

Many optimization problems of interest and some methods of solution require nonnegative variables. The method illustrated below splits a variable into parts; $x = \alpha - \beta$ (extendible to vectors). Under what conditions on vector $a$ and scalar $b$ is an optimal solution $x^*$ negative infinity?

$$\begin{align*}
\text{minimize} & \quad \alpha - \beta \\
\text{subject to} & \quad \beta \geq 0 \\
& \quad \alpha \geq 0 \\
& \quad a^T \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = b
\end{align*}$$

(135)

Minimization of the objective function entails maximization of $\beta$. ▼

---

$\text{2.23}$ The term program has its roots in economics. It was originally meant with regard to a plan or to efficient organization or systematization of some industrial process. [97, §2]

$\text{2.24}$ The objective is the function that is argument to minimization or maximization.
2.5. SUBSPACE REPRESENTATIONS

2.4.2.7 PRINCIPLE 3: Separating hyperplane

The third most fundamental principle of convex geometry again follows from the geometric Hahn-Banach theorem [260, §5.12] [19, §1] [139, §1.1.2] that guarantees existence of a hyperplane separating two nonempty convex sets in $\mathbb{R}^n$ whose relative interiors are nonintersecting. *Separation* intuitively means each set belongs to a halfspace on an opposing side of the hyperplane. There are two cases of interest:

1) If the two sets intersect only at their relative boundaries (§2.1.7.2), then there exists a separating hyperplane $\partial \mathcal{H}$ containing the intersection but containing no points relatively interior to either set. If at least one of the two sets is open, conversely, then the existence of a separating hyperplane implies the two sets are nonintersecting. [63, §2.5.1]

2) A *strictly separating hyperplane* $\partial \mathcal{H}$ intersects the closure of neither set; its existence is guaranteed when intersection of the closures is empty and at least one set is bounded. [209, §A.4.1]

2.4.3 Angle between hyperspaces

Given halfspace-descriptions, *dihedral* angle between hyperplanes or halfspaces is defined as the angle between their defining normals. Given normals $a$ and $b$ respectively describing $\partial \mathcal{H}_a$ and $\partial \mathcal{H}_b$, for example

$$\varangle(\partial \mathcal{H}_a, \partial \mathcal{H}_b) \triangleq \arccos\left(\frac{\langle a, b \rangle}{\|a\|\|b\|}\right) \text{ radians} \quad (136)$$

2.5 Subspace representations

There are two common forms of expression for Euclidean subspaces, both coming from elementary linear algebra: *range form* $\mathcal{R}$ and *nullspace form* $\mathcal{N}$; a.k.a, vertex-description and halfspace-description respectively.

The fundamental vector subspaces associated with a matrix $A \in \mathbb{R}^{m \times n}$ [341, §3.1] are ordinarily related by orthogonal complement

$$\mathcal{R}(A^T) \perp \mathcal{N}(A), \quad \mathcal{N}(A^T) \perp \mathcal{R}(A) \quad (137)$$

$$\mathcal{R}(A^T) \oplus \mathcal{N}(A) = \mathbb{R}^n, \quad \mathcal{N}(A^T) \oplus \mathcal{R}(A) = \mathbb{R}^m \quad (138)$$

and of dimension:

$$\dim \mathcal{R}(A^T) = \dim \mathcal{R}(A) = \text{rank } A \leq \min\{m, n\} \quad (139)$$

with complementarity (a.k.a *conservation of dimension*)

$$\dim \mathcal{N}(A) = n - \text{rank } A, \quad \dim \mathcal{N}(A^T) = m - \text{rank } A \quad (140)$$

These equations (137)-(140) comprise the *fundamental theorem of linear algebra*. [341, p.95, p.138]
From these four fundamental subspaces, the rowspace and range identify one form of subspace description (range form or vertex-description (§2.3.4))

\[
\mathcal{R}(A^T) \triangleq \text{span } A^T = \{ A^T y \mid y \in \mathbb{R}^m \} = \{ x \in \mathbb{R}^n \mid A^T y = x, \ y \in \mathcal{R}(A) \} \quad (141)
\]

\[
\mathcal{R}(A) \triangleq \text{span } A = \{ Ax \mid x \in \mathbb{R}^n \} = \{ y \in \mathbb{R}^m \mid Ax = y, \ x \in \mathcal{R}(A^T) \} \quad (142)
\]

while the nullspaces identify the second common form (nullspace form or halfspace-description (113))

\[
\mathcal{N}(A) \triangleq \{ x \in \mathbb{R}^n \mid Ax = 0 \} = \{ x \in \mathbb{R}^n \mid x \perp \mathcal{R}(A^T) \} \quad (143)
\]

\[
\mathcal{N}(A^T) \triangleq \{ y \in \mathbb{R}^m \mid A^Ty = 0 \} = \{ y \in \mathbb{R}^m \mid y \perp \mathcal{R}(A) \} \quad (144)
\]

Range forms (141) (142) are realized as the respective span of the column vectors in matrices \( A^T \) and \( A \), whereas nullspace form (143) or (144) is the solution set to a linear equation similar to hyperplane definition (114). Yet because matrix \( A \) generally has multiple rows, halfspace-description \( \mathcal{N}(A) \) is actually the intersection of as many hyperplanes through the origin; for (143), each row of \( A \) is normal to a hyperplane while each row of \( A^T \) is a normal for (144).

**2.5.0.0.1 Exercise. Subspace algebra.**

Given

\[
\mathcal{R}(A) + \mathcal{N}(A^T) = \mathcal{R}(B) + \mathcal{N}(B^T) = \mathbb{R}^m \quad (145)
\]

prove

\[
\mathcal{R}(A) \supseteq \mathcal{N}(B^T) \iff \mathcal{N}(A^T) \subseteq \mathcal{R}(B) \quad (146)
\]

\[
\mathcal{R}(A) \supseteq \mathcal{R}(B) \iff \mathcal{N}(A^T) \subseteq \mathcal{N}(B^T) \quad (147)
\]

* e.g, Theorem A.3.1.0.6.

**2.5.1 Subspace or affine subset...**

Any particular vector subspace \( \mathbb{R}_p \) can be described as nullspace \( \mathcal{N}(A) \) of some matrix \( A \) or as range \( \mathcal{R}(B) \) of some matrix \( B \).

More generally, we have the choice of expressing an \( n - m \)-dimensional affine subset of \( \mathbb{R}^n \) as the intersection of \( m \) hyperplanes, or as the offset span of \( n - m \) vectors:

**2.5.1.1 ...as hyperplane intersection**

Any affine subset \( \mathcal{A} \) of dimension \( n - m \) can be described as an intersection of \( m \) hyperplanes in \( \mathbb{R}^n \); given *fat* \( m \leq n \) full-rank \( \text{rank} = \min\{m, n\} \) matrix

\[
A \triangleq \begin{bmatrix}
    a_1^T \\
    \vdots \\
    a_m^T
\end{bmatrix} \in \mathbb{R}^{m \times n} \quad (148)
\]
2.5. **SUBSPACE REPRESENTATIONS**

and vector \( b \in \mathbb{R}^m \),

\[
\mathcal{A} \triangleq \{ x \in \mathbb{R}^n \mid Ax = b \} = \bigcap_{i=1}^m \{ x \mid a_i^T x = b_i \}
\]

(149)

a halfspace-description. (113)

For example: The intersection of any two independent\(^{2.25} \) hyperplanes in \( \mathbb{R}^3 \) is a line, whereas three independent hyperplanes intersect at a point. In \( \mathbb{R}^4 \), the intersection of two independent hyperplanes is a plane (Example 2.5.1.2.1), whereas three hyperplanes intersect at a line, four at a point, and so on. \( \mathcal{A} \) describes a subspace whenever \( b = 0 \) in (149).

For \( n > k \)

\[
\mathcal{A} \cap \mathbb{R}^k = \{ x \in \mathbb{R}^n \mid Ax = b \} \cap \mathbb{R}^k = \bigcap_{i=1}^m \{ x \in \mathbb{R}^k \mid a_i(1:k)^T x = b_i \}
\]

(150)

The result in §2.4.2.2 is extensible; \textit{id est}, any affine subset \( \mathcal{A} \) also has a vertex-description:

### 2.5.1.2 ... as span of nullspace basis

Alternatively, we may compute a basis for nullspace of matrix \( A \) (§E.3.1) and then equivalently express affine subset \( \mathcal{A} \) as its span plus an offset: Define

\[
Z \triangleq \text{basis} \mathcal{N}(A) \in \mathbb{R}^{n \times n - \text{rank} A}
\]

so \( AZ = 0 \). Then we have a vertex-description in \( Z \),

\[
\mathcal{A} = \{ x \in \mathbb{R}^n \mid Ax = b \} = \left\{ Z \xi + x_p \mid \xi \in \mathbb{R}^{n - \text{rank} A} \right\} \subseteq \mathbb{R}^n
\]

(152)

the offset span of \( n - \text{rank} A \) column vectors, where \( x_p \) is any particular solution to \( Ax = b \); \textit{e.g.}, \( \mathcal{A} \) describes a subspace whenever \( x_p = 0 \).

### 2.5.1.2.1 Example. Intersecting planes in 4-space.

Two planes can intersect at a point in four-dimensional Euclidean vector space. It is easy to visualize intersection of two planes in three dimensions; a line can be formed. In four dimensions it is harder to visualize. So let’s resort to the tools acquired.

Suppose an intersection of two hyperplanes in four dimensions is specified by a fat full-rank matrix \( A_1 \in \mathbb{R}^{2 \times 4} \) \((m = 2, n = 4)\) as in (149):

\[
\mathcal{A}_1 \triangleq \left\{ x \in \mathbb{R}^4 \mid \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} x = b_1 \right\}
\]

(153)

\(^{2.25}\)Any number of hyperplanes are called \textit{independent} when defining normals are linearly independent. This misuse departs from independence of two affine subsets that demands intersection only at a point or not at all. (§2.1.4.0.1)
The nullspace of \( A_1 \) is two dimensional (from \( Z \) in (152)), so \( A_1 \) represents a plane in four dimensions. Similarly define a second plane in terms of \( A_2 \in \mathbb{R}^{2\times 4} \):

\[
A_2 \triangleq \left\{ x \in \mathbb{R}^4 \mid \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} x = b_2 \right\}
\]

If the two planes are affinely independent and intersect, they intersect at a point because \( \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \) is invertible:

\[
A_1 \cap A_2 = \left\{ x \in \mathbb{R}^4 \mid \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right\}
\]

\[\square\]

### 2.5.1.2.2 Exercise. Linear program.

Minimize a hyperplane over affine set \( A \) in the nonnegative orthant

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad x \succeq 0
\end{align*}
\]

where \( A = \{ x \mid Ax = b \} \). Two cases of interest are drawn in Figure 32. Graphically illustrate and explain optimal solutions indicated in the caption. Why is \( \alpha^* \) negative in both cases? Is there solution on the vertical axis? What causes objective unboundedness in the latter case (b)? Describe all vectors \( c \) that would yield finite optimal objective in (b).

Graphical solution to linear program

\[
\begin{align*}
\text{maximize} & \quad c^T x \\
\text{subject to} & \quad x \in \mathcal{P}
\end{align*}
\]

is illustrated in Figure 33. Bounded set \( \mathcal{P} \) is an intersection of many halfspaces. Why is optimal solution \( x^* \) not aligned with vector \( c \) as in Cauchy-Schwarz inequality (2153)?

\[\nabla\]

### 2.5.2 Intersection of subspaces

The intersection of nullspaces associated with two matrices \( A \in \mathbb{R}^{m\times n} \) and \( B \in \mathbb{R}^{k\times n} \) can be expressed most simply as

\[
\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}\left( \begin{bmatrix} A \\ B \end{bmatrix} \right) \triangleq \{ x \in \mathbb{R}^n \mid \begin{bmatrix} A \\ B \end{bmatrix} x = 0 \}
\]

nullspace of their rowwise concatenation.

Suppose the columns of a matrix \( Z \) constitute a basis for \( \mathcal{N}(A) \) while the columns of a matrix \( W \) constitute a basis for \( \mathcal{N}(B) \). Then [168, §12.4.2]

\[
\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{R}(ZW)
\]
2.5. SUBSPACE REPRESENTATIONS

Figure 32: Minimizing hyperplane over affine set $\mathcal{A}$ in nonnegative orthant $\mathbb{R}^2_+$ whose extreme directions (§2.8.1) are the nonnegative Cartesian axes. Solutions are visually ascertainable: (a) Optimal solution is $\bullet$. (b) Optimal objective $\alpha^* = -\infty$. 
CHAPTER 2. CONVEX GEOMETRY

Figure 33: Maximizing hyperplane $\partial H$, whose normal is vector $c \in \mathcal{P}$, over polyhedral set $\mathcal{P}$ in $\mathbb{R}^2$ is a linear program \((157)\). Optimal solution $x^*$ at $\bullet$.

If each basis is orthonormal, then the columns of $ZW$ constitute an orthonormal basis for the intersection.

In the particular circumstance $A$ and $B$ are each positive semidefinite \([22, \S 6]\), or in the circumstance $A$ and $B$ are two linearly independent dyads (\S B.1.1), then

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}(A + B), \quad \begin{cases} A, B \in \mathcal{S}_+^M \\ \text{or} \\ A + B = u_1v_1^T + u_2v_2^T \quad \text{(l.i.)} \end{cases}$$

(160)

2.5.3 Visualization of matrix subspaces

Fundamental subspace relations, such as

$$\mathcal{R}(A^T) \perp \mathcal{N}(A), \quad \mathcal{N}(A^T) \perp \mathcal{R}(A)$$

(137)

are partially defining. But to aid visualization of involved geometry, it sometimes helps to vectorize matrices. For any square matrix $A$, $s \in \mathcal{N}(A)$, and $w \in \mathcal{N}(A^T)$

$$\langle A, ss^T \rangle = 0, \quad \langle A, ww^T \rangle = 0$$

(161)

because $s^TA s = w^TA w = 0$. This innocuous observation becomes a sharp instrument for visualization of diagonalizable matrices (\S A.5.1): for rank-$\rho$ matrix $A \in \mathbb{R}^{M \times M}$

$$A = SAS^{-1} = [s_1 \cdots s_M] \Lambda \begin{bmatrix} w_1^T \\ \vdots \\ w_M^T \end{bmatrix} = \sum_{i=1}^M \lambda_i s_i w_i^T$$

(1629)
where nullspace eigenvectors are real by Theorem A.5.0.0.1 and where
\[ \mathcal{R}\{s_i \in \mathbb{R}^M \mid \lambda_i = 0\} = \mathcal{R}\left(\sum_{i=\rho+1}^{M} s_i s_i^T\right) = \mathcal{N}(A) \]
\[ \mathcal{R}\{w_i \in \mathbb{R}^M \mid \lambda_i = 0\} = \mathcal{R}\left(\sum_{i=\rho+1}^{M} w_i w_i^T\right) = \mathcal{N}(A^T) \] (162)

Define an unconventional basis among column vectors of each summation:

\[
\text{basis } \mathcal{N}(A) \subseteq \sum_{i=\rho+1}^{M} s_i s_i^T \subseteq \mathcal{N}(A) \\
\text{basis } \mathcal{N}(A^T) \subseteq \sum_{i=\rho+1}^{M} w_i w_i^T \subseteq \mathcal{N}(A^T) \] (163)

We shall regard a \textit{vectorized subspace} as vectorization of any $M \times M$ matrix whose columns comprise an \textit{overcomplete} basis for that subspace; \textit{e.g.}, §E.3.1

\[
\text{vec basis } \mathcal{N}(A) = \text{vec } \sum_{i=\rho+1}^{M} s_i s_i^T \\
\text{vec basis } \mathcal{N}(A^T) = \text{vec } \sum_{i=\rho+1}^{M} w_i w_i^T \] (164)

By this reckoning, vec basis $\mathcal{R}(A) = \text{vec } A$ but is not unique. Now, because

\[
\langle A , \sum_{i=\rho+1}^{M} s_i s_i^T \rangle = 0, \quad \langle A , \sum_{i=\rho+1}^{M} w_i w_i^T \rangle = 0 \] (165)

then vectorized matrix $A$ is normal to a hyperplane (of dimension $M^2 - 1$) that contains both vectorized nullspaces (each of whose dimension is $M - \rho$);

\[
\text{vec } A \perp \text{vec basis } \mathcal{N}(A), \quad \text{vec basis } \mathcal{N}(A^T) \perp \text{vec } A \] (166)

These vectorized subspace orthogonality relations represent a departure (absent $^T$) from fundamental subspace relations (137) stated at the outset.

2.6. EXTREME, EXPOSED

2.6.0.0.1 Definition. \textit{Extreme point.}

An extreme point $x_\varepsilon$ of a convex set $\mathcal{C}$ is a point, belonging to its closure $\overline{\mathcal{C}}$ [43, §3.3], that is not expressible as a convex combination of points in $\overline{\mathcal{C}}$ distinct from $x_\varepsilon$; \textit{id est}, for $x_\varepsilon \in \overline{\mathcal{C}}$ and all $x_1, x_2 \in \overline{\mathcal{C}} \setminus x_\varepsilon$

\[
\mu x_1 + (1-\mu)x_2 \neq x_\varepsilon, \quad \mu \in [0, 1] \] (167)

$\triangle$
In other words, \( x \in C \) is an extreme point of \( C \) if and only if \( x \) is not a point relatively interior to any line segment in \( C \). \([370, \S 2.10]\)

Borwein & Lewis offer: \([57, \S 4.1.6]\) An extreme point of a convex set \( C \) is a point \( x \in C \) whose relative complement \( C \setminus x \) is convex.

The set consisting of a single point \( C = \{x\} \) is itself an extreme point.

\[2.6.0.0.2\] Theorem. Extreme existence. \([318, \S 18.5.3]\) \([27, \S II.3.5]\)

A nonempty closed convex set containing no lines has at least one extreme point. \(\Diamond\)

\[2.6.0.0.3\] Definition. Face, edge. \([209, \S A.2.3]\)

- A face \( F \) of convex set \( C \) is a convex subset \( F \subseteq \bar{C} \) such that every closed line segment \( x_1x_2 \) in \( C \), having a relatively interior point \( x \in \text{rel int } x_1x_2 \) in \( F \), has both endpoints in \( F \). The zero-dimensional faces of \( C \) constitute its extreme points. The empty set \( \emptyset \) and \( \bar{C} \) itself are conventional faces of \( C \). \([318, \S 18]\)

- All faces \( F \) are extreme sets by definition; \( id \ ext \), for \( F \subseteq \bar{C} \) and all \( x_1, x_2 \in \bar{C} \setminus F \)

\[ \mu x_1 + (1 - \mu)x_2 \notin F, \quad \mu \in [0, 1] \] \(168\)

- A one-dimensional face of a convex set is called an edge. \(\triangle\)

Dimension of a face is the penultimate number of affinely independent points (\(\S 2.4.2.3\)) belonging to it;

\[ \text{dim } F = \sup_{\rho} \text{dim} \{x_2 - x_1, x_3 - x_1, \ldots, x_{\rho} - x_1 \mid x_i \in F, \ i = 1 \ldots \rho \} \] \(169\)

The point of intersection in \( \bar{C} \) with a strictly supporting hyperplane identifies an extreme point, but not \(\text{vice versa}\. The \text{nonempty} intersection of any supporting hyperplane with \( \bar{C} \) identifies a face, in general, but not \(\text{vice versa}\. To acquire a converse, the concept \text{exposed face} requires introduction:

\[2.6.1\] Exposure

\[2.6.1.0.1\] Definition. Exposed face, exposed point, vertex, facet. \([209, \S A.2.3, \S A.2.4]\)

- \( F \) is an exposed face of an \( n \)-dimensional convex set \( C \) iff there is a supporting hyperplane \( \partial H \) to \( C \) such that

\[ F = \bar{C} \cap \partial H \] \(170\)

Only faces of dimension \(-1\) through \( n - 1 \) can be exposed by a hyperplane.

- An exposed point, the definition of vertex, is equivalent to a zero-dimensional exposed face; the point of intersection with a strictly supporting hyperplane.
Figure 34: Closed convex set in $\mathbb{R}^2$. Point A is exposed hence extreme; a classical vertex. Point B is extreme but not an exposed point. Point C is exposed and extreme; zero-dimensional exposure makes it a vertex. Point D is neither an exposed or extreme point although it belongs to a one-dimensional exposed face. \[209, \S A.2.4\] \[340, \S 3.6\] Closed face $AB$ is exposed; a facet. The arc is not a conventional face, yet it is composed entirely of extreme points. Union of all rotations of this entire set about its vertical edge produces another convex set in three dimensions having no edges; but that convex set produced by rotation about horizontal edge containing D has edges.
• A facet is an \((n-1)\)-dimensional exposed face of an \(n\)-dimensional convex set \(C\); facets exist in one-to-one correspondence with the \((n-1)\)-dimensional faces.\(^{2.26}\)

\[\{\text{exposed points}\} = \{\text{extreme points}\} \subseteq \{\text{faces}\}\]

\[\triangle\]

2.6.1.1 Density of exposed points

For any closed convex set \(C\), its exposed points constitute a \textit{dense} subset of its extreme points; \cite{318, 345, 340, §3.6, p.115} dense in the sense \cite{405} that closure of that subset yields the set of extreme points.

For the convex set illustrated in Figure 34, point B cannot be exposed because it relatively bounds both the facet \(AB\) and the closed quarter circle, each bounding the set. Since B is not relatively interior to any line segment in the set, then B is an extreme point by definition. Point B may be regarded as the limit of some sequence of exposed points beginning at vertex C.

2.6.1.2 Face transitivity and algebra

Faces of a convex set enjoy transitive relation. If \(\mathcal{F}_1\) is a face (an extreme set) of \(\mathcal{F}_2\) which in turn is a face of \(\mathcal{F}_3\), then it is always true that \(\mathcal{F}_1\) is a face of \(\mathcal{F}_3\). (The parallel statement for exposed faces is false. \cite{318, §18}) For example, any extreme point of \(\mathcal{F}_2\) is an extreme point of \(\mathcal{F}_3\); in this example, \(\mathcal{F}_2\) could be a face exposed by a hyperplane supporting polyhedron \(\mathcal{F}_3\). \cite[def.115/6 p.358]{233} Yet it is erroneous to presume that a face, of dimension 1 or more, consists entirely of extreme points. Nor is a face of dimension 2 or more entirely composed of edges, and so on.

For the polyhedron in \(\mathbb{R}^3\) from Figure 21, for example, the nonempty faces exposed by a hyperplane are the vertices, edges, and facets; there are no more. The zero-, one-, and two-dimensional faces are in one-to-one correspondence with the exposed faces in that example.

2.6.1.3 Smallest face

Define the smallest face \(\mathcal{F}\), that contains some element \(G\), of a convex set \(\mathcal{C}\):

\[\mathcal{F}(\mathcal{C} \ni G)\]

\textit{videlicet}, \(\mathcal{C} \ni \text{rel int } \mathcal{F}(\mathcal{C} \ni G) \ni G\). An affine set has no faces except itself and the empty set. The smallest face, that contains \(G\), of intersection of convex set \(\mathcal{C}\) with an affine set \(\mathcal{A}\) \cite[§2.4]{249, 250}.

\[\mathcal{F}((\mathcal{C} \cap \mathcal{A}) \ni G) = \mathcal{F}(\mathcal{C} \ni G) \cap \mathcal{A}\]

equals intersection of \(\mathcal{A}\) with the smallest face, that contains \(G\), of set \(\mathcal{C}\).

\(^{2.26}\)This coincidence occurs simply because the facet’s dimension is the same as the dimension of the supporting hyperplane exposing it.
2.6.1.4 Conventional boundary

(confer §2.1.7.2) Relative boundary

\[ \text{rel } \partial C = \overline{C} \setminus \text{rel int } C \quad (24) \]

is equivalent to:

2.6.1.4.1 Definition. Conventional boundary of convex set. [209, §C.3.1]

The relative boundary \( \partial C \) of a nonempty convex set \( C \) is the union of all exposed faces of \( \overline{C} \). \( \triangle \)

Equivalence to (24) comes about because it is conventionally presumed that any supporting hyperplane, central to the definition of exposure, does not contain \( C \). [318, p.100] Any face \( F \) of convex set \( C \) (that is not \( C \) itself) belongs to \( \text{rel } \partial C \). (§2.8.2.1)

2.7 Cones

In optimization, convex cones achieve prominence because they generalize subspaces. Most compelling is the projection analogy: Projection on a subspace can be ascertained from projection on its orthogonal complement (Figure 177), whereas projection on a closed convex cone can be determined from projection instead on its algebraic complement (§2.13, Figure 178, §E.9.2); called the polar cone.

2.7.0.0.1 Definition. Ray. The one-dimensional set

\[ \{ \zeta \Gamma + B \mid \zeta \geq 0, \ \Gamma \neq 0 \} \subset \mathbb{R}^n \quad (173) \]

defines a halfline called a ray in nonzero direction \( \Gamma \in \mathbb{R}^n \) having base \( B \in \mathbb{R}^n \). When \( B = 0 \), a ray is the conic hull of direction \( \Gamma \); hence a convex cone. \( \triangle \)

Relative boundary of a single ray, base 0 in any dimension, is the origin because that is the union of all exposed faces not containing the entire set. Its relative interior is the ray itself excluding the origin.

2.7.1 Cone defined

A set \( \mathcal{X} \) is called, simply, cone if and only if

\[ \Gamma \in \mathcal{X} \Rightarrow \zeta \Gamma \in \overline{\mathcal{X}} \quad \text{for all } \zeta \geq 0 \quad (174) \]

where \( \overline{\mathcal{X}} \) denotes closure of cone \( \mathcal{X} \). An example of such a cone is the union of two opposing quadrants; \( e.g., \mathcal{X} = \{ x \in \mathbb{R}^2 \mid x_1 x_2 \geq 0 \} \) which is not convex. [403, §2.5] Similar examples are shown in Figure 35 and Figure 39.

All cones can be defined by an aggregate of rays emanating exclusively from the origin (but not all cones are convex). Hence all closed cones contain the origin 0 and are
Figure 35: (a) Two-dimensional nonconvex cone drawn truncated. Boundary of this cone is itself a cone. Each half is itself a convex cone. (b) This convex cone (drawn truncated) is a line through the origin in any dimension. It has no relative boundary, while its relative interior comprises entire line.

Figure 36: This nonconvex cone in $\mathbb{R}^2$ is a pair of lines through the origin. [260, §2.4] Because the lines are linearly independent, they are algebraic complements whose vector sum is $\mathbb{R}^2$ a convex cone.

Figure 37: Boundary of a convex cone in $\mathbb{R}^2$ is a nonconvex cone; a pair of rays emanating from the origin.
2.7. CONES

Figure 38: Union of two pointed closed convex cones is nonconvex cone $\mathcal{X}$.

Figure 39: Truncated nonconvex cone $\mathcal{X} = \{ x \in \mathbb{R}^2 \mid x_1 \geq x_2, x_1 x_2 \geq 0 \}$. Boundary is also a cone. [260, §2.4] (Cartesian axes drawn for reference.) Each half (about the origin) is itself a convex cone.

Figure 40: Nonconvex cone $\mathcal{X}$ drawn truncated in $\mathbb{R}^2$. Boundary is also a cone. [260, §2.4] Cone exterior is convex cone.
unbounded, excepting the simplest cone \( \{0\} \). The empty set \( \emptyset \) is not a cone, but its conic hull is;

\[
\text{cone } \emptyset = \{0\}
\]  

(104)

### 2.7.2 Convex cone

We call set \( K \) a **convex cone** iff

\[
\Gamma_1, \Gamma_2 \in K \implies \zeta \Gamma_1 + \xi \Gamma_2 \in \overline{K} \text{ for all } \zeta, \xi \geq 0 \quad (175)
\]

*id est*, if and only if any conic combination of elements from \( K \) belongs to its closure. Apparent from this definition, \( \zeta \Gamma_1 \in \overline{K} \) and \( \xi \Gamma_2 \in \overline{K} \) \( \forall \zeta, \xi \geq 0 \); meaning, \( K \) is a cone. Set \( K \) is convex since, for any particular \( \zeta, \xi \geq 0 \)

\[
\mu \zeta \Gamma_1 + (1 - \mu) \xi \Gamma_2 \in \overline{K} \quad \forall \mu \in [0,1] \quad (176)
\]

because \( \mu \zeta, (1 - \mu) \xi \geq 0 \). Obviously,

\[
\{X\} \supset \{K\} \quad (177)
\]

the set of all convex cones is a proper subset of all cones. The set of convex cones is a narrower but more familiar class of cone, any member of which can be equivalently described as the intersection of a possibly (but not necessarily) infinite number of hyperplanes (through the origin) and halfspaces whose bounding hyperplanes pass through the origin; a halfspace-description (§2.4). Convex cones need not be full-dimensional.
2.7. CONES

More familiar convex cones are Lorentz cone (confer Figure 48)\(^{2.27}\)

\[
\mathcal{K}_\ell = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_\ell \leq t \right\}, \quad \ell = 2
\]

and polyhedral cone (§2.12.1.0.1); e.g., any orthant generated by Cartesian half-axes (§2.1.3). Esoteric examples of convex cones include the point at the origin, any line through the origin, any ray having the origin as base such as the nonnegative real line \( \mathbb{R}_+ \) in subspace \( \mathbb{R} \), any halfspace partially bounded by a hyperplane through the origin, the positive semidefinite cone \( \mathbb{S}^M_+ \) (191), the cone of Euclidean distance matrices \( \text{EDM}^N \) (971) (§6), completely positive semidefinite matrices \( \{CC^T \mid C \succeq 0\} \) [41, p.71], any subspace, and Euclidean vector space \( \mathbb{R}^n \).

2.7.2.1 cone invariance

More Euclidean bodies are cones, it seems, than are not.\(^{2.28}\) This class of convex body, the convex cone, is invariant to scaling, linear and single- or many-valued inverse linear transformation, vector summation, and Cartesian product, but is not invariant to translation. [318, p.22]

2.7.2.1.1 Theorem. Cone intersection (nonempty).

- Intersection of an arbitrary collection of convex cones is a convex cone. [318, §2, §19]
- Intersection of an arbitrary collection of closed convex cones is a closed convex cone. [268, §2.3]
- Intersection of a finite number of polyhedral cones (Figure 52 p.124, §2.12.1.0.1) remains a polyhedral cone.

The property pointedness is associated with a convex cone; but,

- pointed cone \( \Leftrightarrow \) convex cone (Figure 37, Figure 38)

2.7.2.1.2 Definition. Pointed convex cone. (confer §2.12.2.2)

A convex cone \( \mathcal{K} \) is pointed iff it contains no line. Equivalently, \( \mathcal{K} \) is not pointed iff there exists any nonzero direction \( \Gamma \in \mathcal{K} \) such that \( -\Gamma \in \mathcal{K} \). If the origin is an extreme point of \( \overline{\mathcal{K}} \) or, equivalently, if

\[
\mathcal{K} \cap -\mathcal{K} = \{0\}
\]

then \( \mathcal{K} \) is pointed, and vice versa. [340, §2.10] A convex cone is pointed iff the origin is the smallest nonempty face of its closure.

\[^{2.27}\text{a.k.a. second-order cone, quadratic cone, circular cone (§2.9.2.8.1), unbounded ice-cream cone united with its interior.}\]

\[^{2.28}\text{confer Figures: 26 35 36 37 40 39 41 43 45 52 56 59 61 62 64 65 66 67 68 150 163 188}\]
Then a pointed closed convex cone, by principle of separating hyperplane (§2.4.2.7), has a strictly supporting hyperplane at the origin. The simplest and only bounded [403, p.75] convex cone $K = \{0\} \subseteq \mathbb{R}^n$ is pointed, by convention, but not full-dimensional. Its relative boundary is the empty set $\emptyset$ (25) while its relative interior is the point $0$ itself (12). The pointed convex cone that is a halfline, emanating from the origin in $\mathbb{R}^n$, has relative boundary $0$ while its relative interior is the halfline itself excluding $0$. Pointed are any Lorentz cone, cone of Euclidean distance matrices $\text{EDM}_N$ in symmetric hollow subspace $\mathbb{S}^N_{h}$, and positive semidefinite cone $\mathbb{S}^M_+$ in ambient $\mathbb{S}^M$.

2.7.2.1.3 Theorem. Pointed cones. [57, §3.15, exer.20]
A closed convex cone $K \subset \mathbb{R}^n$ is pointed if and only if there exists a normal $\alpha$ such that the set

$$C \triangleq \{ x \in K \mid \langle x, \alpha \rangle = 1 \}$$

(180)
is closed, bounded, and $K = \text{cone} C$. Equivalently, $K$ is pointed if and only if there exists a vector $\beta$ normal to a hyperplane strictly supporting $K$; id est, for some positive scalar $\epsilon$

$$\langle x, \beta \rangle \geq \epsilon \| x \| \quad \forall x \in K$$

(181)

If closed convex cone $K$ is not pointed, then it has no extreme point. Yet a pointed closed convex cone has only one extreme point [43, §3.3]: the exposed point residing at the origin; its vertex. Pointedness is invariant to Cartesian product by (179). And from the cone intersection theorem it follows that an intersection of convex cones is pointed if at least one of the cones is; implying, each and every nonempty exposed face of a pointed closed convex cone is a pointed closed convex cone.

2.7.2.2 Pointed closed convex cone induces partial order

Relation $\preceq$ represents partial order on some set if that relation possesses\footnote{2.29 nor does it have extreme directions (§2.8.1).}()

- reflexivity ($x \preceq x$)
- antisymmetry ($x \preceq z, z \preceq x \Rightarrow x = z$)
- transitivity ($x \preceq y, y \preceq z \Rightarrow x \preceq z$),

$x \preceq y, y \preceq z \Rightarrow x \preceq z$

A pointed closed convex cone $K$ induces partial order on $\mathbb{R}^n$ or $\mathbb{R}^{m \times n}$, [22, §1] [334, p.7] essentially defined by vector or matrix inequality:

$$x \preceq_K z \iff z - x \in K$$

(182)

$$x \prec_K z \iff z - x \in \text{rel int} K$$

(183)

\footnote{2.30 A set is totally ordered if it further obeys a comparability property of the relation: for each and every $x$ and $y$ from the set, $x \preceq y$ or $y \preceq x$; e.g., one-dimensional real vector space $\mathbb{R}$ is the smallest unbounded totally ordered and connected set.}
Figure 42: (confer Figure 71) (a) Point $x$ is the minimum element of set $C_1$ with respect to cone $K$ because cone translated to $x \in C_1$ contains entire set. (Cones drawn truncated.) (b) Point $y$ is a minimal element of set $C_2$ with respect to cone $K$ because negative cone translated to $y \in C_2$ contains only $y$. These concepts, minimum/minimal, become equivalent under a total order.
Neither $x$ or $z$ is necessarily a member of $\mathcal{K}$ for these relations to hold. Only when $\mathcal{K}$ is a nonnegative orthant $\mathbb{R}_n^+$ do these inequalities reduce to ordinary entrywise comparison ($\S2.13.4.2.3$) while partial order lingers. Inclusive of that special case, we ascribe nomenclature generalized inequality to comparison with respect to a pointed closed convex cone.

We say two points $x$ and $y$ are comparable when $x \preceq y$ or $y \preceq x$ with respect to pointed closed convex cone $\mathcal{K}$. Visceral mechanics of actually comparing points, when cone $\mathcal{K}$ is not an orthant, are well illustrated in the example of Figure 65 which relies on the equivalent membership-interpretation in definition $(182)$ or $(183)$.

Comparable points and the minimum element of some vector- or matrix-valued partially ordered set are thus well defined, so nonincreasing sequences with respect to cone $\mathcal{K}$ can therefore converge in this sense: Point $x \in \mathcal{C}$ is the (unique) minimum element of set $\mathcal{C}$ with respect to cone $\mathcal{K}$ iff for each and every $z \in \mathcal{C}$ we have $x \preceq z$; equivalently, iff $\mathcal{C} \subseteq x + \mathcal{K}$.\footnote{Borwein & Lewis [57, §3.3 exer.21] ignore possibility of equality to $x + \mathcal{K}$ in this condition, and require a second condition: . . . and $\mathcal{C} \subset y + \mathcal{K}$ for some $y$ in $\mathbb{R}^n$ implies $x \in y + \mathcal{K}$.}

A closely related concept, minimal element, is useful for partially ordered sets having no minimum element: Point $x \in \mathcal{C}$ is a minimal element of set $\mathcal{C}$ with respect to pointed closed convex cone $\mathcal{K}$ if and only if $(x - \mathcal{K}) \cap \mathcal{C} = x$. (Figure 42) No uniqueness is implied here, although implicit is the assumption: $\dim \mathcal{K} \geq \dim \text{aff } \mathcal{C}$. In words, a point that is a minimal element is smaller (with respect to $\mathcal{K}$) than any other point in the set to which it is comparable.

Further properties of partial order with respect to pointed closed convex cone $\mathcal{K}$ are not defining:

\begin{align*}
\text{homogeneity } & (x \preceq y, \lambda \geq 0 \Rightarrow \lambda x \preceq \lambda z), & (x \prec y, \lambda > 0 \Rightarrow \lambda x < \lambda z) \\
\text{additivity } & (x \preceq z, u \preceq v \Rightarrow x + u \preceq z + v), & (x \prec z, u \preceq v \Rightarrow x + u \prec z + v)
\end{align*}

\section{Definition. Proper cone: a cone that is}

- pointed
- closed
- convex
- full-dimensional.

A proper cone remains proper under injective linear transformation. [93, §5.1] Examples of proper cones are the positive semidefinite cone $\mathbb{S}_+^M$ in the ambient space of symmetric matrices ($\S2.9$), the nonnegative real line $\mathbb{R}_+$ in vector space $\mathbb{R}$, or any orthant in $\mathbb{R}^n$, and the set of all coefficients of univariate degree-$n$ polynomials nonnegative on interval $[0, 1]$ [63, exmp.2.16] or univariate degree-$2n$ polynomials nonnegative over $\mathbb{R}$ [63, exer.2.37].
2.8 Cone boundary

Every hyperplane supporting a convex cone contains the origin. [209, §A.4.2] Because any supporting hyperplane to a convex cone must therefore itself be a cone, then from the cone intersection theorem (§2.7.2.1.1) it follows:

2.8.0.0.1 Lemma. **Cone faces.**

Each nonempty exposed face of a convex cone is a convex cone.

2.8.0.0.2 Theorem. **Proper-cone boundary.**

Suppose a nonzero point Γ lies on the boundary ∂K of proper cone K in \( \mathbb{R}^n \). Then it follows that the ray \( \{ ζ Γ | ζ ≥ 0 \} \) also belongs to ∂K.

Proof. By virtue of its propriety, a proper cone guarantees existence of a strictly supporting hyperplane at the origin. [318, cor.11.7.3][2.32] Hence the origin belongs to the boundary of K because it is the zero-dimensional exposed face. The origin belongs to the ray through Γ, and the ray belongs to K by definition (174). By the cone faces lemma, each and every nonempty exposed face must include the origin. Hence the closed line segment \( 0Γ \) must lie in an exposed face of K because both endpoints do by Definition 2.6.1.4.1. That means there exists a supporting hyperplane \( ∂H \) to K containing \( 0Γ \). So the ray through Γ belongs both to K and to \( ∂H \). \( ∂H \) must therefore expose a face of K that contains the ray; id est,

\[
\{ ζ Γ | ζ ≥ 0 \} \subseteq K \cap ∂H \subseteq ∂K
\]

(184)

Proper cone \{ 0 \} in \( \mathbb{R}^0 \) has no boundary (24) because (12)

\[ \text{rel int}\{ 0 \} = \{ 0 \} \]

(185)

The boundary of any proper cone in \( \mathbb{R} \) is the origin.

The boundary of any convex cone whose dimension exceeds 1 can be constructed entirely from an aggregate of rays emanating exclusively from the origin.

2.8.1 Extreme direction

The property **extreme direction** arises naturally in connection with the pointed closed convex cone \( K \subset \mathbb{R}^n \), being analogous to extreme point. [318, §18, p.162][2.33] An extreme direction \( Γ_ε \) of pointed K is a vector corresponding to an edge that is a ray \( \{ ζ Γ_ε, ζ ≥ 0 \} \) emanating from the origin. [2.34] Nonzero direction \( Γ_ε \) in pointed K is extreme if and only if

\[
ζ_1 Γ_1 + ζ_2 Γ_2 \neq Γ_ε \quad ∀ ζ_1, ζ_2 ≥ 0, \quad ∀ Γ_1, Γ_2 ∈ K \{ ζ Γ_ε ∈ K | ζ ≥ 0 \}
\]

(186)

---

2.32 Rockafellar’s corollary yields a supporting hyperplane at the origin to any convex cone in \( \mathbb{R}^n \) not equal to \( \mathbb{R}^n \).

2.33 We diverge from Rockafellar’s extreme direction: “extreme point at infinity”.

2.34 An edge (§2.6.0.0.3) of a convex cone is not necessarily a ray. A convex cone may contain an edge that is a line; e.g., a wedge-shaped polyhedral cone (\( K^* \) in Figure 43).
Figure 43. $\mathcal{K}$ is a pointed polyhedral cone not full-dimensional in $\mathbb{R}^3$ (drawn truncated in a plane parallel to the floor upon which you stand). Dual cone $\mathcal{K}^*$ is a *wedge* whose truncated boundary is illustrated (drawn perpendicular to the floor). In this particular instance, $\mathcal{K} \subset \text{int} \mathcal{K}^*$ (excepting the origin). (Cartesian coordinate axes drawn for reference.)
2.8. CONE BOUNDARY

In words, an extreme direction in a pointed closed convex cone is the direction of a ray, called an extreme ray, that cannot be expressed as a conic combination of directions of any rays in the cone distinct from it.

An extreme ray is a one-dimensional face of \( \mathcal{K} \). By (105), extreme direction \( \Gamma_\varepsilon \) is not a point relatively interior to any line segment in \( \mathcal{K} \setminus \{ \zeta \Gamma_\varepsilon \in \mathcal{K} \mid \zeta \geq 0 \} \). Thus, by analogy, the corresponding extreme ray \( \{ \zeta \Gamma_\varepsilon \in \mathcal{K} \mid \zeta \geq 0 \} \) is not a ray relatively interior to any plane segment\(^{2.35}\) in \( \mathcal{K} \).

2.8.1.1 extreme distinction, uniqueness

An extreme direction is unique, but its vector representation \( \Gamma_\varepsilon \) is not because any positive scaling of it produces another vector in the same (extreme) direction. Hence an extreme direction is unique to within a positive scaling. When we say extreme directions are distinct, we are referring to distinctness of rays containing them. Nonzero vectors of various length in the same extreme direction are therefore interpreted to be identical extreme directions.\(^{2.36}\)

The extreme directions of the polyhedral cone in Figure 26 (p.64), for example, correspond to its three edges. For any pointed polyhedral cone, there is a one-to-one correspondence of one-dimensional faces with extreme directions.

The extreme directions of the positive semidefinite cone (§2.9) comprise the infinite set of all symmetric rank-one matrices. \[ 22, \text{§6} \] [205, §III] It is sometimes prudent to instead consider the less infinite but complete normalized set, for \( M > 0 \) (confer (235))

\[ \{ z \in \mathbb{S}^M \mid \| z \| = 1 \} \quad (187) \]

The positive semidefinite cone in one dimension \( M = 1 \), \( \mathbb{S}_+ \) the nonnegative real line, has one extreme direction belonging to its relative interior; an idiosyncrasy of dimension 1.

Pointed closed convex cone \( \mathcal{K} = \{ 0 \} \) has no extreme direction because extreme directions are nonzero by definition.

- If closed convex cone \( \mathcal{K} \) is not pointed, then it has no extreme directions and no vertex. \[ 22, \text{§1} \]

Conversely, pointed closed convex cone \( \mathcal{K} \) is equivalent to the convex hull of its vertex and all its extreme directions. \[ 318, \text{§18, p.167} \] That is the practical utility of extreme direction; to facilitate construction of polyhedral sets, apparent from the extremes theorem:

2.8.1.1.1 Theorem. (Klee) Extremes. \[ 340, \text{§3.6} \] [318, §18, p.166] (confer §2.3.2, §2.12.2.0.1) Any closed convex set containing no lines can be expressed as the convex hull of its extreme points and extreme rays.

It follows that any element of a convex set containing no lines may be expressed as a linear combination of its extreme elements; e.g., §2.9.2.7.1.

\(^{2.35}\)A planar fragment; in this context, a planar cone.

\(^{2.36}\)Like vectors, an extreme direction can be identified with the Cartesian point at the vector’s head with respect to the origin.
2.8.1.2 Generators

In the narrowest sense, generators for a convex set comprise any collection of points and directions whose convex hull constructs the set.

When the extremes theorem applies, the extreme points and directions are called generators of a convex set. An arbitrary collection of generators for a convex set includes its extreme elements as a subset; the set of extreme elements of a convex set is a minimal set of generators for that convex set. Any polyhedral set has a minimal set of generators whose cardinality is finite.

When the convex set under scrutiny is a closed convex cone, conic combination of generators during construction is implicit as shown in Example 2.8.1.2.1 and Example 2.10.2.0.1. So, a vertex at the origin (if it exists) becomes benign.

We can, of course, generate affine sets by taking the affine hull of any collection of points and directions. We broaden, thereby, the meaning of generator to be inclusive of all kinds of hulls.

Any hull of generators is loosely called a vertex-description. (§2.3.4) Hulls encompass subspaces, so any basis constitutes generators for a vertex-description; span basis $\mathbb{R}(A)$.

2.8.1.2.1 Example. Application of extremes theorem.

Given an extreme point at the origin and $N$ extreme rays \( \{\zeta \Gamma_i, i=1\ldots N \mid \zeta \geq 0\} \) (§2.7.0.0.1), denoting the $i$th extreme direction by $\Gamma_i \in \mathbb{R}^n$, then their convex hull (86) is

\[
\mathcal{P} = \{ [0 \quad \Gamma_1 \quad \Gamma_2 \cdots \Gamma_N] a \zeta \mid a^T 1 = 1, \quad a \succeq 0, \quad \zeta \geq 0 \}
\]

\[
= \{ [\Gamma_1 \quad \Gamma_2 \cdots \Gamma_N] a \zeta \mid a^T 1 \leq 1, \quad a \succeq 0, \quad \zeta \geq 0 \}
\]

\[
= \{ [\Gamma_1 \quad \Gamma_2 \cdots \Gamma_N] b \mid b \succeq 0 \} \subset \mathbb{R}^n
\]

(188)

a closed convex set that is simply a conic hull like (103).

2.8.2 Exposed direction

2.8.2.0.1 Definition. Exposed point & direction of pointed convex cone. [318, §18] (confer §2.6.1.0.1)

- When a convex cone has a vertex, an exposed point, it resides at the origin; there can be only one.
- In the closure of a pointed convex cone, an exposed direction is the direction of a one-dimensional exposed face that is a ray emanating from the origin.
- \{exposed directions\} $\subseteq$ \{extreme directions\}

For a proper cone in vector space $\mathbb{R}^n$ with $n \geq 2$, we can say more:

\[
\{\text{exposed directions}\} = \{\text{extreme directions}\}
\]

(189)

It follows from Lemma 2.8.0.0.1 for any pointed closed convex cone, there is one-to-one correspondence of one-dimensional exposed faces with exposed directions; id est, there is no one-dimensional exposed face that is not a ray base $0$. 
The pointed closed convex cone $\mathbb{EDM}^2$, for example, is a ray in isomorphic subspace $\mathbb{R}$ whose relative boundary ($\S2.6.1.4.1$) is the origin. The conventionally exposed directions of $\mathbb{EDM}^2$ constitute the empty set $\emptyset \subset \{\text{extreme direction}\}$. This cone has one extreme direction belonging to its relative interior; an idiosyncrasy of dimension 1.

### 2.8.2.1 Connection between boundary and extremes

#### 2.8.2.1.1 Theorem. Exposed. [318, §18.7] (confer §2.8.1.1.1)

Any closed convex set $C$ containing no lines (and whose dimension is at least 2) can be expressed as closure of the convex hull of its exposed points and exposed rays.

From Theorem 2.8.1.1.1,

$$\begin{align*}
\text{rel } \partial C &= C \setminus \text{rel int } C \\
&= \text{conv}\{\text{exposed points and exposed rays}\} \setminus \text{rel int } C \\
&= \text{conv}\{\text{extreme points and extreme rays}\} \setminus \text{rel int } C
\end{align*}$$

Thus each and every extreme point of a convex set (that is not a point) resides on its relative boundary, while each and every extreme direction of a convex set (that is not a halfline and contains no line) resides on its relative boundary because extreme points and directions of such respective sets do not belong to relative interior by definition.

The relationship between extreme sets and the relative boundary actually goes deeper: Any face $F$ of convex set $C$ (that is not $C$ itself) belongs to $\text{rel } \partial C$, so $\dim F < \dim C$. [318, §18.1.3]

#### 2.8.2.2 Converse caveat

It is inconsequent to presume that each and every extreme point and direction is necessarily exposed, as might be erroneously inferred from the conventional boundary definition ($\S2.6.1.4.1$); although it can correctly be inferred: each and every extreme point and direction belongs to some exposed face.

Arbitrary points residing on the relative boundary of a convex set are not necessarily exposed or extreme points. Similarly, the direction of an arbitrary ray, base 0, on the boundary of a convex cone is not necessarily an exposed or extreme direction. For the polyhedral cone illustrated in Figure 26, for example, there are three two-dimensional exposed faces constituting the entire boundary, each composed of an infinity of rays. Yet there are only three exposed directions.

Neither is an extreme direction on the boundary of a pointed convex cone necessarily an exposed direction. Lift the two-dimensional set in Figure 34, for example, into three dimensions such that no two points in the set are collinear with the origin. Then its conic hull can have an extreme direction $B$ on the boundary that is not an exposed direction, illustrated in Figure 44.
Figure 44: Properties of extreme points carry over to extreme directions. [318, §18] Four rays (drawn truncated) on boundary of conic hull of two-dimensional closed convex set from Figure 34 lifted to $\mathbb{R}^3$. Ray through point A is exposed hence extreme. Extreme direction B on cone boundary is not an exposed direction, although it belongs to the exposed face cone $\{A, B\}$. Extreme ray through C is exposed. Point D is neither an exposed or extreme direction although it belongs to a two-dimensional exposed face of the conic hull.
2.9 Positive semidefinite (PSD) cone

The cone of positive semidefinite matrices studied in this section is arguably the most important of all non-polyhedral cones whose facial structure we completely understand.

—Alexander Barvinok [27, p.78]

2.9.0.0.1 Definition. Positive semidefinite cone.

The set of all symmetric positive semidefinite matrices of particular dimension $M$ is called the positive semidefinite cone:

$$S^M_+ \triangleq \left\{ A \in S^M \mid A \succeq 0 \right\}$$

$$= \bigcap_{\|y\|=1} \left\{ A \in S^M \mid \langle yy^T, A \rangle \geq 0 \right\} \equiv \{ A \in S^M_+ \mid \text{rank } A \leq M \} \quad (191)$$

formed by the intersection of an infinite number of halfspaces (§2.4.1.1) in vectorized variable $A$, each halfspace having partial boundary containing the origin in isomorphic $\mathbb{R}^{M(M+1)/2}$. It is a unique immutable proper cone in the ambient space of symmetric matrices $S^M$.

The positive definite (full-rank) matrices comprise the cone interior

$$\text{int } S^M_+ = \left\{ A \in S^M \mid A > 0 \right\}$$

$$= \bigcap_{\|y\|=1} \left\{ A \in S^M \mid \langle yy^T, A \rangle > 0 \right\} \equiv \{ A \in S^M_+ \mid \text{rank } A = M \} \quad (192)$$

while all singular positive semidefinite matrices (having at least one 0 eigenvalue) reside on the cone boundary (Figure 45): (§A.7.5)

$$\partial S^M_+ = \left\{ A \in S^M \mid A \succeq 0, \ A \neq 0 \right\}$$

$$= \left\{ A \in S^M \mid \text{min}\{\lambda(A)_i, \ i=1 \ldots M\} = 0 \right\}$$

$$= \left\{ A \in S^M_+ \mid \langle yy^T, A \rangle = 0 \text{ for some } \|y\| = 1 \right\}$$

$$= \{ A \in S^M_+ \mid \text{rank } A < M \} \quad (193)$$

where $\lambda(A) \in \mathbb{R}^M$ holds the eigenvalues of $A$.

The only symmetric positive semidefinite matrix in $S^M_+$ having $M$ 0-eigenvalues resides at the origin. (§A.7.3.0.1)

\[\text{finite in number when } M>1. \text{ Because } y^T A y = y^T A^T y, \text{ matrix } A \text{ is almost always assumed symmetric. (§A.2.1)}\]
Minimal set of generators are the extreme directions: \( \text{svec}\{yy^T \mid y \in \mathbb{R}^M\} \)

Figure 45: (d’Aspremont) Truncated boundary of PSD cone in \( S^2 \) plotted in isometrically isomorphic \( \mathbb{R}^3 \) via svec (56); 0-contour of smallest eigenvalue (193). Lightest shading is closest, darkest shading is farthest and inside shell. Entire boundary can be constructed from an aggregate of rays (§2.7.0.0.1) emanating exclusively from origin: 
\[
\{ \kappa \begin{bmatrix} z_1^2 & \sqrt{2} z_1 z_2 & z_2^2 \end{bmatrix}^T \mid \kappa \in \mathbb{R}, \ z \in \mathbb{R}^2 \}
\]
A circular cone in this dimension (§2.9.2.8), each and every ray on boundary corresponds to an extreme direction but such is not the case in any higher dimension (confer Figure 26). PSD cone geometry is not as simple in higher dimensions [27, §II.12] although PSD cone is selfdual (376) in ambient real space of symmetric matrices. [205, §II] PSD cone has no two-dimensional face in any dimension, its only extreme point residing at \( 0 \).


2.9. POSITIVE SEMIDEFINITE (PSD) CONE

2.9.0.1 Membership

Observe notation $A \succeq 0$ denoting a positive semidefinite matrix, meaning matrix $A$ belongs to the positive semidefinite cone in the subspace of symmetric matrices whereas $A \succ 0$ denotes membership to that cone’s interior. Notation $A \succ 0$, denoting a positive definite matrix, can be read: symmetric matrix $A$ exceeds the origin with respect to the positive semidefinite cone interior. These notations further imply that coordinates for orthogonal expansion of a positive (semi)definite matrix must be its (nonnegative) positive eigenvalues when expanded in its eigenmatrices. Generalizing comparison on the real line, the notation $A \succeq B$ denotes comparison with respect to the positive semidefinite cone; $A \succeq B \iff A - B \in S^+_M$ but neither matrix $A$ or $B$ necessarily belongs to the positive semidefinite cone. Yet, $A \succeq B, \ B \succeq 0 \Rightarrow A \succeq 0$; id est, $A \in S^+_M$. (confer Figure 65)

2.9.0.1.1 Example. Equality constraints in semidefinite program (680).

Employing properties of partial order (§2.7.2.2) for the pointed closed convex positive semidefinite cone, it is easy to show, given $A + S = C$

\[
S \succeq 0 \iff A \preceq C \\
S \succ 0 \iff A \prec C
\]

2.9.1 Positive semidefinite cone is convex

The set of all positive semidefinite matrices forms a convex cone in the ambient space of symmetric matrices because any pair satisfies definition (175); [212, §7.1] videlicet, for all $\zeta_1, \zeta_2 \geq 0$ and each and every $A_1, A_2 \in S^M$

\[
\zeta_1 A_1 + \zeta_2 A_2 \succeq 0 \iff A_1 \succeq 0, \ A_2 \succeq 0
\]

a fact easily verified by the definitive test for positive semidefiniteness of a symmetric matrix (§A):

\[
A \succeq 0 \iff x^T A x \geq 0 \text{ for each and every } \|x\| = 1
\]

id est, for $A_1, A_2 \succeq 0$ and each and every $\zeta_1, \zeta_2 \geq 0$

\[
\zeta_1 x^T A_1 x + \zeta_2 x^T A_2 x \geq 0 \text{ for each and every normalized } x \in \mathbb{R}^M
\]

The convex cone $S^+_M$ is more easily visualized in the isomorphic vector space $\mathbb{R}^{M(M+1)/2}$ whose dimension is the number of free variables in a symmetric $M \times M$ matrix. When $M = 2$ the PSD cone is semiinfinite in expanse in $\mathbb{R}^3$, having boundary illustrated in Figure 45. When $M = 3$ the PSD cone is six-dimensional, and so on.

---

\(^{2.38}\) the same as nonnegative definite matrix.
2.9.1.0.1 Example. Sets from maps of positive semidefinite cone.

The set

\[ C = \{ X \in S^n \times x \in \mathbb{R}^n \mid X \succeq xx^T \} \]  \hspace{1cm} (198)

is convex because it has Schur-form; (§A.4)

\[ X - xx^T \succeq 0 \iff f(X, x) \triangleq \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \]  \hspace{1cm} (199)

e.g., Figure 46. Set \( C \) is the inverse image (§2.1.9.0.1) of \( S_n^{n+1} \) under affine mapping \( f \).

The set \( \{ X \in S^n \times x \in \mathbb{R}^n \mid X \preceq xx^T \} \) is not convex, in contrast, having no Schur-form.

Yet for fixed \( x = x_p \), the set

\[ \{ X \in S^n \mid X \preceq x_p x_p^T \} \]  \hspace{1cm} (200)

is simply the negative semidefinite cone shifted to \( x_p x_p^T \).

\[ \square \]

2.9.1.0.2 Example. Inverse image of positive semidefinite cone.

Now consider finding the set of all matrices \( X \in S^N \) satisfying

\[ AX + B \succeq 0 \]  \hspace{1cm} (201)

given \( A, B \in S^N \). Define the set

\[ \mathcal{X} \triangleq \{ X \mid AX + B \succeq 0 \} \subseteq S^N \]  \hspace{1cm} (202)

which is the inverse image of the positive semidefinite cone under affine transformation \( g(X) \triangleq AX + B \). Set \( \mathcal{X} \) must therefore be convex by Theorem 2.1.9.0.1.
Yet we would like a less amorphous characterization of this set, so instead we consider its vectorization (37) which is easier to visualize:

\[
\text{vec} \ g(X) = \text{vec}(AX) + \text{vec} B = (I \otimes A) \text{vec} X + \text{vec} B \tag{203}
\]

where

\[
I \otimes A \triangleq Q\Lambda Q^T \in \mathbb{S}^{N^2} \tag{204}
\]

is block-diagonal formed by Kronecker product (§A.1.1 no.31, §D.1.2.1). Assign

\[
x \triangleq \text{vec} X \in \mathbb{R}^{N^2}
\]

\[
b \triangleq \text{vec} B \in \mathbb{R}^{N^2} \tag{205}
\]

then make the equivalent problem: Find

\[
\text{vec} \mathcal{X} = \{x \in \mathbb{R}^{N^2} \mid (I \otimes A)x + b \in \mathcal{K}\} \tag{206}
\]

where

\[
\mathcal{K} \triangleq \text{vec} \mathbb{S}^N_+ \tag{207}
\]

is a proper cone isometrically isomorphic with the positive semidefinite cone in the subspace of symmetric matrices; the vectorization of every element of \(\mathbb{S}^N_+\). Utilizing the diagonalization (204),

\[
\text{vec} \mathcal{X} = \{x \mid \Lambda Q^T x \in Q^T (\mathcal{K} - b)\}
\]

\[
= \{x \mid \Phi Q^T x \in \Lambda^t Q^T (\mathcal{K} - b)\} \subseteq \mathbb{R}^{N^2} \tag{208}
\]

where \(\dagger\) denotes matrix pseudoinverse (§E) and

\[
\Phi \triangleq \Lambda^t \Lambda \tag{209}
\]

is a diagonal projection matrix whose entries are either 1 or 0 (§E.3). We have the complementary sum

\[
\Phi Q^T x + (I - \Phi)Q^T x = Q^T x \tag{210}
\]

So, adding \((I - \Phi)Q^T x\) to both sides of the membership within (208) admits

\[
\text{vec} \mathcal{X} = \{x \mid \Lambda Q^T x \in \Lambda^t Q^T (\mathcal{K} - b) + (I - \Phi)Q^T x\}
\]

\[
= \{x \mid \{x \mid \Phi Q^T x \in \Lambda^t Q^T (\mathcal{K} - b) \} \oplus (I - \Phi)\mathbb{R}^{N^2}\}
\]

\[
= \{x \mid \Phi Q^T x \in \Lambda^t Q^T (\mathcal{K} - b) \oplus Q(I - \Phi)\mathbb{R}^{N^2}\}
\]

\[
= (I \otimes A)^\dagger (\mathcal{K} - b) \oplus \mathcal{N}(I \otimes A) \tag{211}
\]

where we used the facts: linear function \(Q^T x\) in \(x\) on \(\mathbb{R}^{N^2}\) is a bijection, and \(\Phi \Lambda^t = \Lambda^t\).

\[
\text{vec} \mathcal{X} = (I \otimes A)^\dagger \text{vec}(\mathbb{S}^N_+ - B) \oplus \mathcal{N}(I \otimes A) \tag{212}
\]

In words, set \(\text{vec} \mathcal{X}\) is the vector sum of the translated PSD cone (linearly mapped onto the rowspace of \(I \otimes A\) (§E)) and the nullspace of \(I \otimes A\) (synthesis of fact from §A.6.3 and §A.7.3.0.1). Should \(I \otimes A\) have no nullspace, then \(\text{vec} \mathcal{X} = (I \otimes A)^{-1} \text{vec}(\mathbb{S}^N_+ - B)\) which is the expected result. \(\square\)
Figure 47: (a) Projection of truncated PSD cone $S^2_+$, truncated above $\gamma = 1$, on $\alpha \beta$-plane in isometrically isomorphic $\mathbb{R}^3$. View is from above with respect to Figure 45. (b) Truncated above $\gamma = 2$. From these plots we might infer, for example, line $\{[0 \ 1/\sqrt{2} \ \gamma]^T \mid \gamma \in \mathbb{R}\}$ intercepts PSD cone at some large value of $\gamma$; in fact, $\gamma = \infty$. 

$$\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$$
2.9. Positive semidefinite cone boundary

For any symmetric positive semidefinite matrix $A$ of rank $\rho$, there must exist a rank $\rho$ matrix $Y$ such that $A$ be expressible as an outer product in $Y$; [341, §6.3]

$$A = YY^T \in S^M_+ \quad \text{rank } A = \text{rank } Y = \rho \quad Y \in \mathbb{R}^{M \times \rho}$$

Then the boundary of the positive semidefinite cone may be expressed

$$\partial S^M_+ = \left\{ A \in S^M_+ \mid \text{rank } A < M \right\} = \left\{ YY^T \mid Y \in \mathbb{R}^{M \times M-1} \right\}$$

Because the boundary of any convex body is obtained with closure of its relative interior ($\S2.1.7$, §2.1.7.2), from (192) we must also have

$$S^M_+ = \left\{ A \in S^M_+ \mid \text{rank } A = M \right\} = \left\{ YY^T \mid Y \in \mathbb{R}^{M \times M} \right\}$$

2.9.2.1 rank $\rho$ subset of the positive semidefinite cone

For the same reason (closure), this applies more generally; for $0 \leq \rho \leq M$

$$\left\{ A \in S^M_+ \mid \text{rank } A = \rho \right\} = \left\{ A \in S^M_+ \mid \text{rank } A \leq \rho \right\}$$

For easy reference, we give such generally nonconvex sets a name: rank $\rho$ subset of a positive semidefinite cone. For $\rho < M$ this subset, nonconvex for $M > 1$, resides on the positive semidefinite cone boundary.

2.9.2.1.1 Exercise. Closure and rank $\rho$ subset.

Prove equality in (216).

For example,

$$\partial S^M_+ = \left\{ A \in S^M_+ \mid \text{rank } A = M - 1 \right\} = \left\{ A \in S^M_+ \mid \text{rank } A \leq M - 1 \right\}$$

In $S^2$, each and every ray on the boundary of the positive semidefinite cone in isomorphic $\mathbb{R}^3$ corresponds to a symmetric rank-1 matrix (Figure 45), but that does not hold in any higher dimension.

2.9.2.2 Subspace tangent to open rank $\rho$ subset

When the positive semidefinite cone subset in (216) is left unclosed as in

$$\mathcal{M}(\rho) \triangleq \left\{ A \in S^N_+ \mid \text{rank } A = \rho \right\}$$

In $\mathbb{R}^3$, each and every ray on the boundary of the positive semidefinite cone in isomorphic $\mathbb{R}^3$ corresponds to a symmetric rank-1 matrix (Figure 45), but that does not hold in any higher dimension.
then we can specify a subspace tangent to the positive semidefinite cone at a particular member of manifold \( \mathcal{M}(\rho) \). Specifically, the subspace \( \mathcal{R}_\mathcal{M} \) tangent to manifold \( \mathcal{M}(\rho) \) at \( B \in \mathcal{M}(\rho) \) \cite[5, prop.1.1]{196}

\[
\mathcal{R}_\mathcal{M}(B) \triangleq \{ XB + BX^T \mid X \in \mathbb{R}^{N \times N} \} \subseteq \mathbb{S}^N
\]

has dimension

\[
\dim \text{svec} \mathcal{R}_\mathcal{M}(B) = \rho \left( N - \frac{\rho - 1}{2} \right) = \rho(N - \rho) + \frac{\rho(\rho + 1)}{2}
\]

Tangent subspace \( \mathcal{R}_\mathcal{M} \) contains no member of the positive semidefinite cone \( \mathbb{S}_+^N \) whose rank exceeds \( \rho \).

Subspace \( \mathcal{R}_\mathcal{M}(B) \) is a hyperplane supporting \( \mathbb{S}_+^N \) when \( B \in \mathcal{M}(N-1) \). Another good example of tangent subspace is given in \S\E.7.2.0.2 by \( \rho(N-1) \); \( \mathcal{R}_\mathcal{M}(1^T) = \mathbb{S}_c^N \), orthogonal complement to the geometric center subspace. (Figure 160 p.468)

### 2.9.2.3 Faces of PSD cone, their dimension versus rank

Each and every face of the positive semidefinite cone, having dimension less than that of the cone, is exposed. \cite[6]{256}, \cite[2.3.4]{225} Because each and every face of the positive semidefinite cone contains the origin (\S\2.8.0.0.1), each face belongs to a subspace of dimension the same as the face.

Define \( \mathcal{F}(\mathbb{S}_+^M \ni A) \) \cite{171} as the smallest face, that contains a given positive semidefinite matrix \( A \), of positive semidefinite cone \( \mathbb{S}_+^M \). Then matrix \( A \), having ordered diagonalization \( A = QA \Lambda Q^T \in \mathbb{S}_+^M \) \( \S\A.5.1 \), is relatively interior to \cite[27, \SII.12]{119}, \cite[31.5.3]{249}, \cite[2.4]{250} \cite[2.39]{27}, \cite[II.12]{249}

\[
\mathcal{F}(\mathbb{S}_+^M \ni A) = \{ X \in \mathbb{S}_+^M \mid \mathcal{N}(X) \supseteq \mathcal{N}(A) \}
\]

\[
= \{ X \in \mathbb{S}_+^M \mid \langle Q(I - \Lambda\Lambda^T)Q^T, X \rangle = 0 \}
\]

\[
= \{ QA\Lambda^\Psi\Lambda\Lambda^\Psi^T \mid \Psi \in \mathbb{S}_+^M \}
\]

\[
\simeq \mathbb{S}_+^{\text{rank} A}
\]

which is isomorphic with convex cone \( \mathbb{S}_+^{\text{rank} A} \); \( e.g. \), \( QA\mathbb{S}_+^M Q^T = \mathbb{S}_+^M \). The larger the nullspace of \( A \), the smaller the face. \( \text{(140)} \) Thus dimension of the smallest face that contains given matrix \( A \) is

\[
\dim \mathcal{F}(\mathbb{S}_+^M \ni A) = \text{rank}(A)(\text{rank}(A) + 1)/2
\]

in isomorphic \( \mathbb{R}^{M(M+1)/2} \), and each and every face of \( \mathbb{S}_+^M \) is isomorphic with a positive semidefinite cone having dimension the same as the face. Observe: not all dimensions are represented, and the only zero-dimensional face is the origin. The positive semidefinite cone has no facets, for example.

\textit{For} \( X \in \mathbb{S}_+^M \), \( A = QA \Lambda Q^T \in \mathbb{S}_+^M \), show \( \mathcal{N}(X) \supseteq \mathcal{N}(A) \Leftrightarrow \langle Q(I - \Lambda\Lambda^T)Q^T, X \rangle = 0. \)

\textit{Given} \( \langle Q(I - \Lambda\Lambda^T)Q^T, X \rangle = 0 \Leftrightarrow \mathcal{R}(X) \perp \mathcal{N}(A). \) \( \S\A.7.4 \)

\textit{(\Rightarrow)} Assume \( \mathcal{N}(X) \supseteq \mathcal{N}(A) \), then \( \mathcal{R}(X) \perp \mathcal{N}(X) \supseteq \mathcal{N}(A). \)

\textit{(\Leftarrow)} Assume \( \mathcal{R}(X) \perp \mathcal{N}(A) \), then \( XQ(I - \Lambda\Lambda^T)Q^T = 0 \Rightarrow \mathcal{N}(X) \supseteq \mathcal{N}(A). \) \( \blacklozenge \)
2.9. POSITIVE SEMIDEFINITE (PSD) CONE

2.9.2.3.1 Table: Rank $k$ versus dimension of $S^+_3$ faces

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\dim F(S^+_3 \ni \text{rank-}k \text{ matrix})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>boundary</td>
<td>$\leq 1$</td>
</tr>
<tr>
<td></td>
<td>$\leq 2$</td>
</tr>
<tr>
<td>interior</td>
<td>$\leq 3$</td>
</tr>
</tbody>
</table>

For positive semidefinite cone $S^+_3$ in isometrically isomorphic $\mathbb{R}^3$ depicted in Figure 45, rank-2 matrices belong to the interior of that face having dimension 3 (the entire closed cone), rank-1 matrices belong to relative interior of a face having dimension 2.40 1, and the only rank-0 matrix is the point at the origin (the zero-dimensional face).

2.9.2.3.2 Exercise. Bijective isometry.
Prove that the smallest face of positive semidefinite cone $S^+_M$, containing a particular full-rank matrix $A$ having ordered diagonalization $Q\Lambda Q^T$, is the entire cone: id est, prove $Q S^+_M Q^T = S^+_M$ from (221).

2.9.2.4 rank-$k$ face of PSD cone

Any rank-$k < M$ positive semidefinite matrix $A$ belongs to a face, of positive semidefinite cone $S^+_M$, described by intersection with a hyperplane: for ordered diagonalization of $A = Q\Lambda Q^T \in S^+_M \ni \text{rank}(A) = k < M$

$$F(S^+_M \ni A) = \{ X \in S^+_M \mid \langle Q(I - \Lambda\Lambda^T)Q^T, X \rangle = 0 \}$$

$$= \left\{ X \in S^+_M \mid \left\langle Q \left( I - \left[ \begin{array}{c} I \in S^k \\ 0^T \\ 0 \end{array} \right] \right) Q^T, X \right\rangle = 0 \right\}$$

(223)

$$\simeq S^+_M \cap \partial H^+$$

Faces are doubly indexed: continuously indexed by orthogonal matrix $Q$, and discretely indexed by rank $k$. Each and every orthogonal matrix $Q$ makes projectors $Q(:, k+1:M)Q(:, k+1:M)^T$ indexed by $k$, in other words, each projector describing a normal\textsuperscript{2.41} $\text{svec}(Q(:, k+1:M)Q(:, k+1:M)^T)$ to a supporting hyperplane $\partial H^+$ (containing the origin) exposing a face (§2.11) of the positive semidefinite cone containing rank-$k$ (and less) matrices.

2.9.2.4.1 Exercise. Simultaneously diagonalizable means commutative.
Given diagonalization of rank-$k \leq M$ positive semidefinite matrix $A = Q\Lambda Q^T$ and any particular $\Psi \succeq 0$, both in $S^M$ from (221), show how $I - \Lambda\Lambda^T$ and $\Lambda\Lambda^T \Psi \Lambda\Lambda^T$ share a complete set of eigenvectors.

\textsuperscript{2.40} The boundary constitutes all the one-dimensional faces, in $\mathbb{R}^3$, which are rays emanating from the origin.

\textsuperscript{2.41} Any vectorized nonzero matrix $\in S^+_M$ is normal to a hyperplane supporting $S^+_M$ (§2.13.1) because PSD cone is selfdual. Normal on boundary exposes nonzero face by (329) (330).
2.9.2.5 PSD cone face containing principal submatrix

A principal submatrix of a matrix $A \in \mathbb{R}^{M \times M}$ is formed by discarding any particular subset of its rows and columns having the same indices. There are $M!/((M-1)!)$ principal $1 \times 1$ submatrices, $M!/(2!(M-2)!)$ principal $2 \times 2$ submatrices, and so on, totaling $2^M - 1$ principal submatrices including $A$ itself. Principal submatrices of a symmetric matrix are symmetric. A given symmetric matrix has rank $\rho$ iff it has a nonsingular principal $\rho \times \rho$ submatrix but none larger. [308, §5-10] By loading vector $y$ in test $y^T A y$ (§A.2) with various binary patterns, it follows that any principal submatrix must be positive (semi)definite whenever $A$ is (Theorem A.3.1.0.4). If positive semidefinite matrix $A \in S^+_M$ has principal submatrix of dimension $\rho$ with rank $r$, then rank $A \leq M - \rho + r$ by (1618).

Because each and every principal submatrix of a positive semidefinite matrix in $S^+_M$ is positive semidefinite, then each principal submatrix belongs to a certain face of positive semidefinite cone $S^+_M$ by (222). Of special interest are full-rank positive semidefinite principal submatrices, for then description of smallest face becomes simpler. We can find the smallest face, that contains a particular complete full-rank principal submatrix of $A$, by embedding that submatrix in a $0$ matrix of the same dimension as $A$: Were $\Phi$ a binary diagonal matrix

$$
\Phi = \delta^2(\Phi) \in S^+_M, \quad \Phi_{ii} \in \{0, 1\}
$$

having diagonal entry 0 corresponding to a discarded row and column from $A \in S^+_M$, then any principal submatrix so embedded can be expressed $\Phi A \Phi$; id est, for an embedded principal submatrix $\Phi A \Phi \in S^+_M \ni \text{rank } \Phi A \Phi = \text{rank } \Phi \leq \text{rank } A$

$$
\mathcal{F}(S^+_M \ni \Phi A \Phi) = \{ X \in S^+_M \mid \mathcal{N}(X) \supseteq \mathcal{N}(\Phi A \Phi) \}$$

$$
= \{ X \in S^+_M \mid \langle I - \Phi, X \rangle = 0 \}$$

$$
= \{ \Phi \Psi \Phi \mid \Psi \in S^+_M \}
$$

$$
\simeq S^{\text{rank } \Phi}_ M
$$

Smallest face that contains an embedded principal submatrix, whose rank is not necessarily full, may be expressed like (221): For embedded principal submatrix $\Phi A \Phi \in S^+_M \ni \text{rank } \Phi A \Phi \leq \text{rank } \Phi$, apply ordered diagonalization instead to

$$
\hat{\Phi}^T A \hat{\Phi} = U \Upsilon U^T \in S^+_{\text{rank } \Phi}
$$

where $U^{-1} = U^T$ is an orthogonal matrix and $\Upsilon = \delta^2(\Upsilon)$ is diagonal. Then

$$
\mathcal{F}(S^+_M \ni \Phi A \Phi) = \{ X \in S^+_M \mid \mathcal{N}(X) \supseteq \mathcal{N}(\Phi A \Phi) \}$$

$$
= \{ X \in S^+_M \mid \langle \hat{\Phi} U (I - \Upsilon \Upsilon^T) U^T \hat{\Phi}^T + I - \Phi, X \rangle = 0 \}$$

$$
= \{ \hat{\Phi} U \Upsilon U^T \Upsilon^T U \Upsilon U^T \hat{\Phi}^T \mid \Psi \in S^+_{\text{rank } \Phi} \}
$$

$$
\simeq S^{\text{rank } \Phi A \Phi}
$$

---

2.42 To express a leading principal submatrix, for example, $\Phi = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$. 
2.9. POSITIVE SEMIDEFINITE (PSD) CONE

where binary diagonal matrix $\Phi$ is partitioned into nonzero and zero columns by permutation $\Xi \in \mathbb{R}^{M \times M}$:

$$\Phi \Xi^T \triangleq \begin{bmatrix} \hat{\Phi} & 0 \\ \end{bmatrix} \in \mathbb{R}^{M \times M}, \quad \text{rank} \ \hat{\Phi} = \text{rank} \ \Phi, \quad \Phi = \hat{\Phi} \hat{\Phi}^T \in \mathbb{S}^+_M, \quad \hat{\Phi}^T \hat{\Phi} = I \quad (228)$$

Any embedded principal submatrix may be expressed

$$\Phi A \Phi = \hat{\Phi} \hat{\Phi}^T A \hat{\Phi} \hat{\Phi}^T \in \mathbb{S}^+_M \quad (229)$$

where $\hat{\Phi}^T A \hat{\Phi}$ extracts the principal submatrix whereas $\hat{\Phi} \hat{\Phi}^T A \hat{\Phi} \hat{\Phi}^T$ embeds it.

2.9.2.5.1 Example. Smallest face containing disparate elements.

Smallest face formula (221) can be altered to accommodate a union of points $\{A_i \in \mathbb{S}^+_M\}$:

$$\mathcal{F}(\mathbb{S}^+_M \supset \bigcup_i A_i) = \left\{ X \in \mathbb{S}^+_M \mid \mathcal{N}(X) \supset \bigcap_i \mathcal{N}(A_i) \right\} \quad (230)$$

To see that, imagine two vectorized matrices $A_1$ and $A_2$ on diametrically opposed sides of the positive semidefinite cone $\mathbb{S}^+_2$ boundary pictured in Figure 45. Regard svec $A_1$ as normal to a hyperplane in $\mathbb{R}^3$ containing a vectorized basis for its nullspace: svec basis $\mathcal{N}(A_1)$ (25.3). Similarly, there is a second hyperplane containing svec basis $\mathcal{N}(A_2)$ having normal svec $A_2$. While each hyperplane is two-dimensional, each nullspace has only one affine dimension because $A_1$ and $A_2$ are rank-1. Because our interest is only that part of the nullspace in the positive semidefinite cone, then by

$$\langle X, A_i \rangle = 0 \iff XA_i = A_i X = 0, \quad X, A_i \in \mathbb{S}^+_M \quad (1674)$$

we may ignore the fact that vectorized nullspace svec basis $\mathcal{N}(A_i)$ is a proper subspace of the hyperplane. We may think instead in terms of whole hyperplanes because equivalence (1674) says that the positive semidefinite cone effectively filters that subset of the hyperplane, whose normal is $A_i$, constituting $\mathcal{N}(A_i)$.

And so hyperplane intersection makes a line intersecting the positive semidefinite cone $\mathbb{S}^+_2$ but only at the origin. In this hypothetical example, smallest face containing those two matrices therefore comprises the entire cone because every positive semidefinite matrix has nullspace containing 0. The smaller the intersection, the larger the smallest face. □

2.9.2.5.2 Exercise. Disparate elements.

Prove that (230) holds for an arbitrary set $\{A_i \in \mathbb{S}^+_M \forall i \in I\}$. One way is by showing $\bigcap \mathcal{N}(A_i) \cap \mathbb{S}^+_M = \text{conv}(\{A_i\}) \cap \mathbb{S}^+_M$; with perpendicularity $\perp$ as in (372).\textsuperscript{2.43} ▼

2.9.2.6 face of all PSD matrices having same principal submatrix

Now we ask what is the smallest face of the positive semidefinite cone containing all matrices having a complete principal submatrix in common; in other words, that face containing all PSD matrices (of any rank) with particular entries fixed — the smallest

\textsuperscript{2.43}Hint: (1674) (2005).
face containing all PSD matrices whose fixed entries correspond to some given embedded principal submatrix $\Phi A \Phi$. To maintain generality, \footnote{\ref{footnote:2.44}} we move an extracted principal submatrix $\hat{\Phi} \in \mathbb{S}^\text{rang } \Phi$ into leading position via permutation $\Xi$ from (\ref{eq:2.44}): for $A \in \mathbb{S}^\text{M}_+$

$$
\Xi A \Xi^T \triangleq \begin{bmatrix}
\hat{\Phi}^T A \hat{\Phi} & B \\
B^T & C
\end{bmatrix} \in \mathbb{S}^\text{M}_+
$$

(231)

By properties of partitioned PSD matrices in §A.4.0.1, basis $\mathcal{N}\left(\begin{bmatrix}
\hat{\Phi}^T A \hat{\Phi} & B \\
B^T & C
\end{bmatrix}\right) \supseteq \begin{bmatrix} 0 \\ I - C C^T \end{bmatrix}$

(232)

Hence $\mathcal{N}(\Xi X \Xi^T) \supseteq \mathcal{N}(\Xi A \Xi^T) \not\subseteq \text{span} \begin{bmatrix} 0 \\ I \end{bmatrix}$ in a smallest face $\mathcal{F}$ formula \footnote{\ref{footnote:2.45}} because all PSD matrices, given fixed principal submatrix, are admitted: Define a set of all PSD matrices $\mathcal{S} \triangleq \{ A = \Xi^T \begin{bmatrix} \hat{\Phi}^T A \hat{\Phi} & B \\
B^T & C
\end{bmatrix} \Xi \succeq 0 \mid B \in \mathbb{R}^{\text{rank } \Phi \times M - \text{rank } \Phi}, C \in \mathbb{S}^{M - \text{rank } \Phi}_+ \}$ having fixed embedded principal submatrix $\Phi A \Phi = \Xi^T \begin{bmatrix} \hat{\Phi}^T A \hat{\Phi} & 0 \\
0^T & 0
\end{bmatrix} \Xi$. So

$$
\mathcal{F}\left(\mathbb{S}^\text{M}_+ \supseteq \mathcal{S}\right) = \{ X \in \mathbb{S}^\text{M}_+ \mid \mathcal{N}(X) \supseteq \mathcal{N}(\mathcal{S}) \} = \{ X \in \mathbb{S}^\text{M}_+ \mid \langle \hat{\Phi} U (I - YY^T) U^T \hat{\Phi}^T, X \rangle = 0 \} = \{ \Xi^T \begin{bmatrix} U T & 0 \\
0^T & I
\end{bmatrix} \Psi \begin{bmatrix} T T^T & 0 \\
0^T & I
\end{bmatrix} \Xi \mid \Psi \in \mathbb{S}^\text{M}_+ \} \tag{234}
$$

(234)

$\Xi = I$ whenever $\Phi A \Phi$ denotes a leading principal submatrix. Smallest face of the positive semidefinite cone, containing all matrices having the same full-rank principal submatrix ($YY^T = I$, $Y \succeq 0$), is the entire cone (Exercise \ref{exercise:2.9.3.2}).

\subsection*{2.9.2.7 Extreme directions of positive semidefinite cone}

Because the positive semidefinite cone is pointed (§2.7.2.1.2), there is a one-to-one correspondence of one-dimensional faces with extreme directions in any dimension $M$; \textit{id est}, because of the cone faces lemma (§2.8.0.0.1) and direct correspondence of exposed faces to faces of $\mathbb{S}_+^\text{M}$, it follows: there is no one-dimensional face of the positive semidefinite cone that is not a ray emanating from the origin.

Symmetric dyads constitute the set of all extreme directions: For $M > 1$

$$
\{ yy^T \in \mathbb{S}^\text{M} \mid y \in \mathbb{R}^M \} \subset \partial \mathbb{S}^\text{M}_+
$$

(235)\footnote{\ref{footnote:2.44} to fix any principal submatrix; not only leading principal submatrices. \ref{footnote:2.45} meaning, more pertinently, $I - \Phi$ is dropped from (\ref{eq:2.45}).}
2.9. POSITIVE SEMIDEFINITE (PSD) CONE

this superset of extreme directions (infinite in number, confer (187)) for the positive
semidefinite cone is, generally, a subset of the boundary. By extremes theorem 2.8.1.1.1,
the convex hull of extreme rays and origin is the positive semidefinite cone: (2.8.1.2.1)

\[
\text{conv}\{yy^T \in S^M \mid y \in \mathbb{R}^M\} = \left\{ \sum_{i=1}^{\infty} b_i z_i z_i^T \mid z_i \in \mathbb{R}^M, \ b \succeq 0 \right\} = S^M_+ \tag{236}
\]

For two-dimensional matrices \((M = 2, \text{ Figure 45})\)

\[
\{yy^T \in S^2 \mid y \in \mathbb{R}^2\} = \partial S^2_+ \tag{237}
\]

while for one-dimensional matrices, in exception, \((M = 1, \text{ §2.7})\)

\[
\{yy \in S \mid y \neq 0\} = \text{int} S_+ \tag{238}
\]

Each and every extreme direction \(yy^T\) makes the same angle with the Identity matrix
in isomorphic \(\mathbb{R}^{M(M+1)/2}\), dependent only on dimension; videlicet2.46

\[
\varangle(yy^T, I) = \arccos \frac{\langle yy^T, I \rangle}{\|yy^T\|_F \|I\|_F} = \arccos \left( \frac{1}{\sqrt{M}} \right) \quad \forall y \in \mathbb{R}^M \tag{239}
\]

This means the positive semidefinite cone broadens in higher dimension.

2.9.2.7.1 Example. Positive semidefinite matrix from extreme directions.

Diagonalizability (§A.5) of symmetric matrices yields the following results:

Any positive semidefinite matrix (1532) in \(S^M\) can be written in the form

\[
A = \sum_{i=1}^{M} \lambda_i z_i z_i^T = \hat{A} \hat{A}^T = \sum_{i} \hat{a}_i \hat{a}_i^T \succeq 0, \quad \lambda \succeq 0 \tag{240}
\]

a conic combination of linearly independent extreme directions \((\hat{a}_i \hat{a}_i^T \text{ or } z_i z_i^T)\) where \(\|z_i\| = 1\), where \(\lambda\) is a vector of eigenvalues.

If we limit consideration to all symmetric positive semidefinite matrices bounded via unity trace

\[
\mathcal{C} \triangleq \{ A \succeq 0 \mid \text{tr} A = 1 \} \tag{91}
\]

then any matrix \(A\) from that set may be expressed as a convex combination of linearly
independent extreme directions;

\[
A = \sum_{i=1}^{M} \lambda_i z_i z_i^T \in \mathcal{C}, \quad 1^T \lambda = 1, \quad \lambda \succeq 0 \tag{241}
\]

Implications are:

1. set \(\mathcal{C}\) is convex (an intersection of PSD cone with hyperplane),
2. because the set of eigenvalues corresponding to a given square matrix \(A\) is unique
   (§A.5.0.1), no single eigenvalue can exceed 1; \(\text{id est}, I \succeq A\)
3. and the converse holds: set \(\mathcal{C}\) is an instance of Fantope (91).

2.46 Analogy with respect to the EDM cone is considered in [195, p.162] where it is found: angle is not
constant. Extreme directions of the EDM cone can be found in §6.4.3.2. The cone’s axis is 
\(-E = 11^T - I\) (1176).
2.9.2.7.2 Exercise. Extreme directions of positive semidefinite cone.
Prove, directly from definition (186), that symmetric dyads (235) constitute the set of all extreme directions of the positive semidefinite cone.

2.9.2.8 Positive semidefinite cone is generally not circular

Extreme angle equation (239) suggests that the positive semidefinite cone might be invariant to rotation about its axis of revolution; id est, a circular cone. We investigate this now:

2.9.2.8.1 Definition. Circular cone: 2.47
a pointed closed convex cone having hyperspherical sections orthogonal to its axis of revolution about which the cone is invariant to rotation.

A conic section is the intersection of a cone with any hyperplane. In three dimensions, an intersecting plane perpendicular to a circular cone’s axis of revolution produces a section bounded by a circle. (Figure 48) A prominent example of a circular cone in convex analysis is Lorentz cone (178). We also find that the positive semidefinite cone and cone of Euclidean distance matrices are circular cones, but only in low dimension.

The positive semidefinite cone has axis of revolution that is the ray (base 0) through the Identity matrix $I$. Consider a set of normalized extreme directions of the positive semidefinite cone: for some arbitrary positive constant $a \in \mathbb{R}_+$

$$\{yy^T \in \mathbb{S}^M \ | \ \|y\| = \sqrt{a}\} \subset \partial \mathbb{S}^M_+ \quad (242)$$

The distance from each extreme direction to the axis of revolution is radius

$$R \triangleq \inf_c \|yy^T - cI\|_F = a \sqrt{1 - \frac{1}{M}} \quad (243)$$

which is the distance from $yy^T$ to $\frac{a}{M}I$; the length of vector $yy^T - \frac{a}{M}I$.

Because distance $R$ (in a particular dimension) from the axis of revolution to each and every normalized extreme direction is identical, the extreme directions lie on the boundary of a hypersphere in isometrically isomorphic $\mathbb{R}^{M(M+1)/2}$. From Example 2.9.2.7.1, the convex hull (excluding vertex at the origin) of the normalized extreme directions is a conic section

$$C \triangleq \text{conv}\{yy^T \ | \ y \in \mathbb{R}^M, \ y^T y = a\} = \mathbb{S}_+^M \cap \{A \in \mathbb{S}^M \ | \ \langle I, A \rangle = a\} \quad (244)$$

orthogonal to Identity matrix $I$;

$$\langle C - \frac{a}{M}I, I \rangle = \text{tr}(C - \frac{a}{M}I) = 0 \quad (245)$$

2.47 A circular cone is assumed convex throughout, although not so by other authors. We also assume a right circular cone.
Figure 48: This solid circular cone in $\mathbb{R}^3$ continues upward infinitely. Axis of revolution is illustrated as vertical line through origin. $R$ represents radius: distance measured from an extreme direction to axis of revolution. Were this a Lorentz cone, any plane slice containing axis of revolution would make a right angle.
CHAPTER 2. CONVEX GEOMETRY

Figure 49: Illustrated is a section, perpendicular to axis of revolution, of circular cone from Figure 48. Radius $R$ is distance from any extreme direction to axis at $\frac{a}{M}I$. Vector $\frac{a}{M}11^T$ is an arbitrary reference by which to measure angle $\theta$.

**Proof.** Although the positive semidefinite cone possesses some characteristics of a circular cone, we can show it is not by demonstrating shortage of extreme directions; *id est*, some extreme directions corresponding to each and every angle of rotation about the axis of revolution are nonexistent: Referring to Figure 49, \[ (246) \]

$$\cos \theta = \frac{\langle \frac{a}{M}11^T - \frac{a}{M}I, yy^T - \frac{a}{M}I \rangle}{a^2(1 - \frac{1}{M})}$$

Solving for vector $y$ we get

$$a(1 + (M - 1)\cos \theta) = (1^Ty)^2 \quad (247)$$

which does not have real solution $\forall 0 \leq \theta \leq 2\pi$ in every matrix dimension $M$. ♦

From the foregoing proof we can conclude that the positive semidefinite cone might be circular but only in matrix dimensions 1 and 2. Because of a shortage of extreme directions, conic section (244) cannot be hyperspherical by the *extremes theorem* (§2.8.1.1.1, Figure 44).

2.9.2.8.2 Exercise. *Circular semidefinite cone.*

Prove the positive semidefinite cone to be circular in matrix dimensions 1 and 2 while it is a rotation of Lorentz cone (178) in matrix dimension 2.  \[ ^{2.48} \]

\[^{2.48}\text{Hint: Given cone } \left\{ \begin{bmatrix} \alpha \\ \beta \sqrt{2} \\ \gamma \end{bmatrix} \mid \sqrt{\alpha^2 + \beta^2} \leq \gamma \right\}, \text{ show } \frac{1}{\sqrt{2}} \begin{bmatrix} \gamma + \alpha \\ \beta \\ \gamma - \alpha \end{bmatrix} \text{ is a vector rotation that is positive semidefinite under the same inequality.} \]
2.9. POSITIVE SEMIDEFINITE (PSD) CONE

$p = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$

$p = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$p = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

\[ p_1 A_{11} \geq \pm p_2 A_{12} \]
\[ p_2 A_{22} \geq \pm p_1 A_{12} \]

Figure 50: Proper polyhedral cone $\mathcal{K}$, created by intersection of halfspaces, inscribes PSD cone in isometrically isomorphic $\mathbb{R}^3$ as predicted by Geršgorin discs theorem for $A = [A_{ij}] \in \mathbb{S}^2$. Hyperplanes supporting $\mathcal{K}$ intersect along boundary of PSD cone. Four extreme directions of $\mathcal{K}$ coincide with extreme directions of PSD cone.
2.9.2.8.3 Example. Positive semidefinite cone inscription in three dimensions.

Theorem. Geršgorin discs. For \( p \in \mathbb{R}_+^m \) given \( A = [A_{ij}] \in S^m \), then all eigenvalues of \( A \) belong to the union of \( m \) closed intervals on the real line:

\[
\lambda(A) \in \bigcup_{i=1}^{m} \left\{ \xi \in \mathbb{R} \mid |\xi - A_{ii}| \leq \frac{1}{p_i} \sum_{j=1}^{m} p_j |A_{ij}| \right\} = \bigcup_{i=1}^{m} [A_{ii} - \xi_i, A_{ii} + \xi_i] \quad (248)
\]

Furthermore, if a union of \( k \) of these \( m \) intervals forms a connected region that is disjoint from all the remaining \( n-k \) intervals, then there are precisely \( k \) eigenvalues of \( A \) in this region.

To apply the theorem to determine positive semidefiniteness of symmetric matrix \( A \), we observe that for each \( i \) we must have

\[
A_{ii} \geq \xi_i \quad (249)
\]

Suppose \( m = 2 \)

so \( A \in S^2 \). Vectorizing \( A \) as in (56), \( \text{svec} A \) belongs to isometrically isomorphic \( \mathbb{R}^3 \). Then we have \( m^2 = 4 \) inequalities, in the matrix entries \( A_{ij} \) with Geršgorin parameters \( p = [p_i] \in \mathbb{R}_+^2 \),

\[
p_1 A_{11} \geq \pm p_2 A_{12} \\
p_2 A_{22} \geq \pm p_1 A_{12} \quad (251)
\]

which describe an intersection of four halfspaces in \( \mathbb{R}^{m(m+1)/2} \). That intersection creates the proper polyhedral cone \( K \) whose construction is illustrated in Figure 50. Drawn truncated is the boundary of the positive semidefinite cone \( \text{svec} S_+^2 \) and the bounding hyperplanes supporting \( K \).

Created by means of Geršgorin discs, \( K \) always belongs to the positive semidefinite cone for any nonnegative value of \( p \in \mathbb{R}_+^m \). Hence any point in \( K \) corresponds to some positive semidefinite matrix \( A \). Only the extreme directions of \( K \) intersect the positive semidefinite cone boundary in this dimension; the four extreme directions of \( K \) are extreme directions of the positive semidefinite cone. As \( p_1/p_2 \) increases in value from 0, two extreme directions of \( K \) sweep the entire boundary of this positive semidefinite cone. Because the entire positive semidefinite cone can be swept by \( K \), the system of linear inequalities

\[
Y^T \text{svec} A \leq \begin{bmatrix} p_1 & \pm p_2/\sqrt{2} & 0 \\ 0 & \pm p_1/\sqrt{2} & p_2 \end{bmatrix} \text{svec} A \succeq 0 \quad (252)
\]

(which made dynamic) can replace a semidefinite constraint \( A \succeq 0 \); id est, for

\[
K = \{ \mathbf{z} \mid Y^T \mathbf{z} \geq 0 \} \subset \text{svec} S_+^m \quad (253)
\]

given \( p \) where \( Y \in \mathbb{R}^{m(m+1)/2 \times m^2} \)

\[
\text{svec} A \in K \Rightarrow A \in S^m_+ \quad (254)
\]
but

\[ \exists p \ni Y^T svec A \succeq 0 \iff A \succeq 0 \quad (255) \]

In other words, diagonal dominance \[212, \text{p}349, \S7.2.3\]

\[ A_{ii} \geq \sum_{j=1, j \neq i}^m |A_{ij}|, \quad \forall i = 1 \ldots m \quad (256) \]

is generally only a sufficient condition for membership to the PSD cone. But by dynamic weighting \( p \) in this dimension, diagonal dominance was made necessary and sufficient. 

\[ \square \]

In higher dimension \((m > 2)\), boundary of the positive semidefinite cone is no longer constituted completely by its extreme directions (symmetric rank-one matrices); its geometry becomes intricate. How all the extreme directions can be swept by an inscribed polyhedral cone, \[2.49\] similarly to the foregoing example, remains an open question.

\[ \textbf{2.9.2.8.4 Exercise. Dual inscription.} \]
Find dual proper polyhedral cone \( K^* \) from Figure \textit{50}. \n
\[ \downarrow \]

\[ \textbf{2.9.2.9 Boundary constituents of the positive semidefinite cone} \]

\[ \textbf{2.9.2.9.1 Lemma. Sum of positive semidefinite matrices.} \quad (\text{confer} (1548)) \]
For \( A, B \in S^+_M \)

\[ \text{rank}(A + B) = \text{rank}(\mu A + (1 - \mu)B) \quad (257) \]
over open interval \((0, 1)\) of \( \mu \). \n
\[ \diamond \]

\[ \textbf{Proof.} \] Any positive semidefinite matrix belonging to the PSD cone has an eigenvalue decomposition that is a positively scaled sum of linearly independent symmetric dyads. By the \textit{linearly independent dyads definition} in \S B.1.1.0.1, rank of the sum \( A + B \) is equivalent to the number of linearly independent dyads constituting it. Linear independence is insensitive to further positive scaling by \( \mu \). The assumption of positive semidefiniteness prevents annihilation of any dyad from the sum \( A + B \). 

\[ \downarrow \]

\[ \textbf{2.9.2.9.2 Example. Rank function quasiconcavity.} \quad (\text{confer} \S3.8) \]
For \( A, B \in \mathbb{R}^{m \times n} \) \[212, \S0.4\]

\[ \text{rank } A + \text{rank } B \geq \text{rank}(A + B) \quad (258) \]
that follows from the fact \[341, \S3.6\]

\[ \dim \mathcal{R}(A) + \dim \mathcal{R}(B) = \dim \mathcal{R}(A + B) + \dim (\mathcal{R}(A) \cap \mathcal{R}(B)) \quad (259) \]

For \( A, B \in S^+_M \)

\[ \text{rank } A + \text{rank } B \geq \text{rank}(A + B) \geq \min \{ \text{rank } A, \text{rank } B \} \quad (1548) \]

\[ \text{\[2.49\]It is not necessary to sweep the entire boundary in higher dimension.} \]
that follows from the fact
\[ \mathcal{N}(A + B) = \mathcal{N}(A) \cap \mathcal{N}(B), \quad A, B \in S^+_M \] (160)

Rank is a *quasiconcave* function on \( S^+_M \) because the right-hand inequality in (1548) has the concave form (645); *videlicet*, Lemma 2.9.2.9.1. \( \Box \)

From this example we see, unlike convex functions, *quasiconvex* functions are not necessarily continuous. (§3.8) We also glean:

2.9.2.9.3 Theorem. Convex subsets of positive semidefinite cone.
Subsets of the positive semidefinite cone \( S^+_M \), for \( 0 \leq \rho \leq M \)
\[ S^+_M(\rho) \triangleq \{ X \in S^+_M \mid \text{rank} X \geq \rho \} \] (260)
are pointed convex cones, but not closed unless \( \rho = 0 \): *id est*, \( S^+_M(0) = S^+_M \). \( \Box \)

**Proof.** Given \( \rho \), a subset \( S^+_M(\rho) \) is convex if and only if convex combination of any two members has rank at least \( \rho \). That is confirmed by applying identity (257) from Lemma 2.9.2.9.1 to (1548); *id est*, for \( A, B \in S^+_M(\rho) \) on closed interval \([0, 1]\) of \( \mu \)
\[ \text{rank}(\mu A + (1-\mu)B) \geq \min\{\text{rank} A, \text{rank} B\} \] (261)
It can similarly be shown, almost identically to proof of the lemma, any conic combination of \( A, B \) in subset \( S^+_M(\rho) \) remains a member; *id est*, \( \forall \zeta, \xi \geq 0 \)
\[ \text{rank}(\zeta A + \xi B) \geq \min\{\text{rank}(\zeta A), \text{rank}(\xi B)\} \] (262)
Therefore, \( S^+_M(\rho) \) is a convex cone. \( \Diamond \)

Another proof of convexity can be made by projection arguments:

2.9.2.10 Projection on \( S^+_M(\rho) \)
Because these cones \( S^+_M(\rho) \) indexed by \( \rho \) (260) are convex, projection on them is straightforward. Given a symmetric matrix \( H \) having diagonalization \( H = Q \Lambda Q^T \in S^M \) (§A.5.1) with eigenvalues \( \Lambda \) arranged in nonincreasing order, then its *Euclidean projection* (minimum-distance projection) on \( S^+_M(\rho) \)
\[ P_{S^+_M(\rho)}H = Q \Upsilon^* Q^T \] (263)
corresponds to a map of its eigenvalues:
\[ \Upsilon^*_{ii} = \begin{cases} \max\{\epsilon, \Lambda_{ii}\}, & i = 1 \ldots \rho \\ \max\{0, \Lambda_{ii}\}, & i = \rho + 1 \ldots M \end{cases} \] (264)
where \( \epsilon \) is positive but arbitrarily close to 0.
2.9. POSITIVE SEMIDEFINITE (PSD) CONE

2.9.2.10.1 Exercise. Projection on open convex cones.

Prove (264) using Theorem E.9.2.0.1. ▼

Because each $H \in S^M$ has unique projection on $S^M_+(\rho)$ (despite possibility of repeated eigenvalues in $\Lambda$), we may conclude it is a convex set by the Bunt-Motzkin theorem (§E.9.0.0.1).

Compare (264) to the well-known result regarding Euclidean projection on a rank $\rho$ subset of the positive semidefinite cone (§2.9.2.1)

\[
P_{S^M_+ \setminus S^M_+(\rho+1)} H = QT^*Q^T
\]

As proved in §7.1.4, this projection of $H$ corresponds to the eigenvalue map

\[
\Upsilon_{ii}^* = \begin{cases} 
\max \{0, \Lambda_{ii}\} & \text{for } i = 1 \ldots \rho \\
0 & \text{for } i = \rho + 1 \ldots M
\end{cases}
\]

Together these two results (264) and (1423) mean: A higher-rank solution to projection on the positive semidefinite cone lies arbitrarily close to any given lower-rank projection, but not \emph{vice versa}. Were the number of nonnegative eigenvalues in $\Lambda$ known \emph{a priori} not to exceed $\rho$, then these two different projections would produce identical results in the limit $\epsilon \to 0$.

2.9.2.11 Uniting constituents

Interior of the PSD cone $\text{int} S^M_+$ is convex by Theorem 2.9.2.9.3, for example, because all positive semidefinite matrices having rank $M$ constitute the cone interior.

All positive semidefinite matrices of rank less than $M$ constitute the cone boundary; an amalgam of positive semidefinite matrices of different rank. Thus each nonconvex subset of positive semidefinite matrices, for $0 < \rho < M$

\[
\{Y \in S^M_+ \mid \text{rank } Y = \rho\}
\]

having rank $\rho$ successively 1 lower than $M$, appends a nonconvex constituent to the cone boundary; but only in their union is the boundary complete: (confer §2.9.2)

\[
\partial S^M_+ = \bigcup_{\rho=0}^{M-1} \{Y \in S^M_+ \mid \text{rank } Y = \rho\}
\]

The composite sequence, the cone interior in union with each successive constituent, remains convex at each step; \emph{id est}, for $0 \leq k \leq M$

\[
\bigcup_{\rho=k}^{M} \{Y \in S^M_+ \mid \text{rank } Y = \rho\}
\]

is convex for each $k$ by Theorem 2.9.2.9.3.
2.9.2.12 Peeling constituents

Proceeding the other way: To peel constituents off the complete positive semidefinite cone boundary, one starts by removing the origin; the only rank-0 positive semidefinite matrix. What remains is convex. Next, the extreme directions are removed because they constitute all the rank-1 positive semidefinite matrices. What remains is again convex, and so on. Proceeding in this manner eventually removes the entire boundary leaving, at last, the convex interior of the PSD cone; all the positive definite matrices.

2.9.2.12.1 Exercise. Difference $A - B$.

What about a difference of matrices $A, B$ belonging to the PSD cone? Show:

- Difference of any two points on the boundary belongs to the boundary or exterior.
- Difference $A - B$, where $A$ belongs to the boundary while $B$ is interior, belongs to the exterior.

2.9.3 Barvinok’s proposition

Barvinok posits existence and quantifies an upper bound on rank of a positive semidefinite matrix belonging to the intersection of the PSD cone with an affine subset:

2.9.3.0.1 Proposition. Affine intersection with PSD cone. [27, §II.13] [25, §2.2]

Consider finding a matrix $X \in S^N$ satisfying

$$X \succeq 0, \quad \langle A_j, X \rangle = b_j, \quad j = 1 \ldots m$$

(269)

given nonzero linearly independent (vectorized) $A_j \in S^N$ and real $b_j$. Define the affine subset

$$\mathcal{A} \triangleq \{ X \mid \langle A_j, X \rangle = b_j, \quad j = 1 \ldots m \} \subseteq S^N$$

(270)

If the intersection $\mathcal{A} \cap S^N_+$ is nonempty given a number $m$ of equalities, then there exists a matrix $X \in \mathcal{A} \cap S^N_+$ such that

$$\text{rank } X (\text{rank } X + 1)/2 \leq m$$

(271)

whence the upper bound$^{2.50}$

$$\text{rank } X \leq \left\lfloor \frac{\sqrt{8m+1} - 1}{2} \right\rfloor$$

(272)

Given desired rank instead, equivalently,

$$m < (\text{rank } X + 1)(\text{rank } X + 2)/2$$

(273)

$^{2.50}$ §4.1.2.2 contains an intuitive explanation. This bound is itself limited above, of course, by $N$; a tight limit corresponding to an interior point of $S^N_+$. 

An extreme point of $\mathcal{A} \cap \mathbb{S}_+^N$ satisfies (272) and (273). (confer §4.1.2.2) A matrix $X \triangleq R^T R$ is an extreme point if and only if the smallest face, that contains $X$, of $\mathcal{A} \cap \mathbb{S}_+^N$ has dimension 0; \cite[§2.4]{249} \textit{id est}, iff (171)

\[
\dim \mathcal{F}(\mathcal{A} \cap \mathbb{S}_+^N \ni X) = \text{rank}(X)(\text{rank}(X) + 1)/2 - \text{rank}\left[ \text{svec \mathbf{R} A_1 R^T} \, \text{svec RA}_2 R^T \cdots \text{svec RA}_m R^T \right] \]  

equals 0 in isomorphic $\mathbb{R}^{N(N+1)/2}$.

Now the intersection $\mathcal{A} \cap \mathbb{S}_+^N$ is assumed bounded: Assume a given nonzero upper bound $\rho$ on rank, a number of equalities $m = (\rho + 1)(\rho + 2)/2$ (275) and matrix dimension $N \geq \rho + 2 \geq 3$. If the intersection is nonempty and bounded, then there exists a matrix $X \in \mathcal{A} \cap \mathbb{S}_+^N$ such that

$\text{rank} X \leq \rho$ (276)

This represents a tightening of the upper bound; a reduction by exactly 1 of the bound provided by (272) given the same specified number $m$ (275) of equalities; \textit{id est},

$\text{rank} X \leq \frac{\sqrt{8m + 1} - 1}{2} - 1$ (277)

2.10 Conic independence (c.i.)

In contrast to extreme direction, the property \textit{conically independent direction} is more generally applicable; inclusive of all closed convex cones (not only pointed closed convex cones). Arbitrary given directions $\{\Gamma_i \in \mathbb{R}^n, \, i = 1 \ldots N\}$ comprise a \textit{conically independent set} if and only if (confer §2.1.2, §2.4.2.3)

$\Gamma_i \zeta_i + \cdots + \Gamma_j \zeta_j - \Gamma_\ell = 0, \quad i \neq \cdots \neq j \neq \ell = 1 \ldots N$ (278)

has no solution $\zeta \in \mathbb{R}_+^N$ ($\zeta_i \in \mathbb{R}_+$); in words, \textit{iff} no direction from the given set can be expressed as a conic combination of those remaining; \textit{e.g.}, Figure 51 \cite[conic independence test (278) MATLAB]. Arranging any set of generators for a particular closed convex cone in a matrix columnar,

$X \triangleq [\Gamma_1 \Gamma_2 \cdots \Gamma_N] \in \mathbb{R}^{n \times N}$ (279)

then this test of conic independence (278) may be expressed as a set of linear \textit{feasibility problems}: for $\ell = 1 \ldots N$

\[
\begin{array}{l}
\text{find} \quad \zeta \in \mathbb{R}^N \\
\text{subject to} \quad X\zeta = \Gamma_\ell \\
\qquad \zeta \succeq 0 \\
\qquad \zeta_\ell = 0
\end{array} \]  

(280)
(a) affinely and conically independent,  
(b) affinely independent but not conically independent,  
(c) conically independent but not affinely independent.  
None of the examples exhibits linear independence. (In general, a.i. ⇔ c.i.)

If feasible for any particular $\ell$, then the set is not conically independent.

To find all conically independent directions from a set via (280), generator $\Gamma_\ell$ must be removed from the set once it is found (feasible) conically dependent on remaining generators in $X$. So, to continue testing remaining generators when $\Gamma_\ell$ is found to be dependent, $\Gamma_\ell$ must be discarded from matrix $X$ before proceeding. A generator $\Gamma_\ell$ that is instead found conically independent of remaining generators in $X$, on the other hand, is conically independent of any subset of remaining generators. A c.i. set thus found is not necessarily unique.

It is evident that linear independence (l.i.) of $N$ directions implies their conic independence:

- l.i. $\Rightarrow$ c.i.

which suggests, number of l.i. generators in the columns of $X$ cannot exceed number of c.i. generators. Denoting by $k$ the number of conically independent generators contained in $X$, we have the most fundamental rank inequality for convex cones

$$\dim \text{aff} K = \dim \text{aff}[0 \ X] = \text{rank } X \leq k \leq N$$

(281)

Whereas $N$ directions in $n$ dimensions can no longer be linearly independent once $N$ exceeds $n$, conic independence remains possible:

2.10.0.1 Table: Maximum number of c.i. directions

<table>
<thead>
<tr>
<th>dimension $n$</th>
<th>sup $k$ (pointed)</th>
<th>sup $k$ (not pointed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
</tbody>
</table>

Assuming veracity of this table, there is an apparent vastness between two and three dimensions. The finite numbers of conically independent directions indicate:
2.10. CONIC INDEPENDENCE (C.I.)

- Convex cones in dimensions 0, 1, and 2 must be polyhedral. (§2.12.1)

Conic independence is certainly one convex idea that cannot be completely explained by a two-dimensional picture as Barvinok suggests [27, p.vii].

From this table it is also evident that dimension of Euclidean space cannot exceed the number of conically independent directions possible;

- \( n \leq \text{sup } k \)

2.10.1 Preservation of conic independence

Independence in the linear (§2.1.2.1), affine (§2.4.2.4), and conic senses can be preserved under linear transformation. Suppose a matrix \( X \in \mathbb{R}^{n \times N} \) (279) holds a conically independent set columnar. Consider a transformation on the domain of such matrices

\[
T(X) : \mathbb{R}^{n \times N} \rightarrow \mathbb{R}^{n \times N} \triangleq XY
\]

(282)

where fixed matrix \( Y \triangleq [y_1 \ y_2 \cdots y_N] \in \mathbb{R}^{N \times N} \) represents linear operator \( T \). Conic independence of \( \{Xy_i \in \mathbb{R}^n, \ i=1\ldots N\} \) demands, by definition (278),

\[
Xy_i \zeta_i + \cdots + Xy_j \zeta_j - Xy_\ell = 0, \quad i \neq \cdots \neq j \neq \ell = 1\ldots N
\]

(283)

have no solution \( \zeta \in \mathbb{R}_+^N \). That is ensured by conic independence of \( \{y_i \in \mathbb{R}^N\} \) and by \( \mathcal{R}(Y) \cap \mathcal{N}(X) = \emptyset \); seen by factoring out \( X \).

2.10.1.1 linear maps of cones

[22, §7] If \( \mathcal{K} \) is a convex cone in Euclidean space \( \mathcal{R} \) and \( T \) is any linear mapping from \( \mathcal{R} \) to Euclidean space \( \mathcal{M} \), then \( T(\mathcal{K}) \) is a convex cone in \( \mathcal{M} \) and \( x \preceq y \) with respect to \( \mathcal{K} \) implies \( T(x) \preceq T(y) \) with respect to \( T(\mathcal{K}) \). If \( \mathcal{K} \) is full-dimensional in \( \mathcal{R} \), then so is \( T(\mathcal{K}) \) in \( \mathcal{M} \).

If \( T \) is a linear bijection, then \( x \preceq y \iff T(x) \preceq T(y) \). If \( \mathcal{K} \) is pointed, then so is \( T(\mathcal{K}) \). And if \( \mathcal{K} \) is closed, so is \( T(\mathcal{K}) \). If \( \mathcal{F} \) is a face of \( \mathcal{K} \), then \( T(\mathcal{F}) \) is a face of \( T(\mathcal{K}) \).

Linear bijection is only a sufficient condition for pointedness and closedness; convex polyhedra (§2.12) are invariant to any linear or inverse linear transformation [27, §1.9] [318, p.44, thm.19.3].

2.10.2 Pointed closed convex \( \mathcal{K} \& \text{ conic independence} \)

The following bullets can be derived from definitions (186) and (278) in conjunction with the extremes theorem (§2.8.1.1.1):

- The set of all extreme directions from a pointed closed convex cone \( \mathcal{K} \subset \mathbb{R}^n \) is not necessarily a linearly independent set, yet it must be a conically independent set; (compare Figure 26 on page 64 with Figure 52a)

- \( \{\text{extreme directions}\} \Rightarrow \{\text{c.i.}\} \)
CHAPTER 2. CONVEX GEOMETRY

Figure 52: (a) A pointed polyhedral cone (drawn truncated) in $\mathbb{R}^3$ having six facets. The extreme directions, corresponding to six edges emanating from the origin, are generators for this cone; not linearly independent but they must be conically independent. (b) The boundary of dual cone $\mathcal{K}^*$ (drawn truncated) is now added to the drawing of same $\mathcal{K}$. $\mathcal{K}^*$ is polyhedral, proper, and has the same number of extreme directions as $\mathcal{K}$ has facets.
2.10. CONIC INDEPENDENCE (C.I.)

When a conically independent set of directions from pointed closed convex cone $\mathcal{K}$ is known to comprise generators, conversely, then all directions from that set must be extreme directions of the cone;

- $\{\text{extreme directions}\} \iff \{\text{c.i. generators of pointed closed convex } \mathcal{K}\}$

Barker & Carlson [22, §1] call the extreme directions a minimal generating set for a pointed closed convex cone. A minimal set of generators is therefore a conically independent set of generators, and vice versa.

An arbitrary collection of $n$ or fewer distinct extreme directions from pointed closed convex cone $\mathcal{K} \subset \mathbb{R}^n$ is not necessarily a linearly independent set; e.g., dual extreme directions (482) from Example 2.13.11.0.3.

- $\{\leq n \text{ extreme directions in } \mathbb{R}^n\} \not\Rightarrow \{\text{l.i.}\}$

Linear dependence of few extreme directions is another convex idea that cannot be explained by a two-dimensional picture as Barvinok suggests [27, p.vii]; indeed, it only first comes to light in four dimensions! But there is a converse: [340, §2.10.9]

- $\{\text{extreme directions}\} \iff \{\text{l.i. generators of closed convex } \mathcal{K}\}$

---

This converse does not hold for nonpointed closed convex cones as Table 2.10.0.0.1 implies; e.g., ponder four conically independent generators for a plane ($n=2$, Figure 51).
2.10.2.0.1 Example. Vertex-description of halfspace \( H \) about origin.

From \( n + 1 \) points in \( \mathbb{R}^n \) we can make a vertex-description of a convex cone that is a halfspace \( H \), where \( \{ x_\ell \in \mathbb{R}^n, \ell = 1 \ldots n \} \) constitutes a minimal set of generators for a hyperplane \( \partial H \) through the origin. An example is illustrated in Figure 53. By demanding the augmented set \( \{ x_\ell \in \mathbb{R}^n, \ell = 1 \ldots n+1 \} \) be affinely independent (we want vector \( x_{n+1} \) not parallel to \( \partial H \)), then

\[
H = \bigcup_{\zeta \geq 0} (\zeta x_{n+1} + \partial H) \\
= \{ \zeta x_{n+1} + \text{cone}\{ x_\ell \in \mathbb{R}^n, \ell = 1 \ldots n \} | \zeta \geq 0 \} \\
= \text{cone}\{ x_\ell \in \mathbb{R}^n, \ell = 1 \ldots n+1 \}
\]

(284)

a union of parallel hyperplanes. Cardinality is one step beyond dimension of the ambient space, but \( \{ x_\ell \ \forall \ell \} \) is a minimal set of generators for this convex cone \( H \) which has no extreme elements.

2.10.2.0.2 Exercise. Enumerating conically independent directions.

Do Example 2.10.2.0.1 in \( \mathbb{R} \) and \( \mathbb{R}^3 \) by drawing two figures corresponding to Figure 53 and enumerating \( n + 1 \) conically independent generators for each. Describe a nonpointed polyhedral cone in three dimensions having more than 8 conically independent generators. 

(confer Table 2.10.0.0.1)

2.10.3 Utility of conic independence

Perhaps the most useful application of conic independence is determination of the intersection of closed convex cones from their halfspace-descriptions, or representation of the sum of closed convex cones from their vertex-descriptions.

- \( \bigcap \mathcal{K}_i \) A halfspace-description for the intersection of any number of closed convex cones \( \mathcal{K}_i \) can be acquired by pruning normals; specifically, only the conically independent normals from the aggregate of all the halfspace-descriptions need be retained.

- \( \sum \mathcal{K}_i \) Generators for the sum of any number of closed convex cones \( \mathcal{K}_i \) can be determined by retaining only the conically independent generators from the aggregate of all the vertex-descriptions.

Such conically independent sets are not necessarily unique or minimal.

2.11 When extreme means exposed

For any convex full-dimensional polyhedral set in \( \mathbb{R}^n \), distinction between the terms \textit{extreme} and \textit{exposed} vanishes [340, §2.4] [119, §2.2] for faces of all dimensions except \( n \); their meanings become equivalent as we saw in Figure 21 (discussed in §2.6.1.2). In other words, each and every face of any polyhedral set (except the set itself) can be exposed by a hyperplane, and \textit{vice versa}; \textit{e.g.}, Figure 26.

Lewis [256, §6] [225, §2.3.4] claims nonempty extreme proper subsets and the exposed subsets coincide for \( S^n_+ \); \textit{id est}, each and every face of the positive semidefinite cone, whose
2.12 CONVEX POLYHEDRA

Dimension is less than dimension of the cone, is exposed. A more general discussion of cones having this property can be found in [351]; e.g., Lorentz cone (178) [21, §II.A].

2.12 Convex polyhedra

Every polyhedron, such as the convex hull (86) of a bounded list $X$, can be expressed as the solution set of a finite system of linear equalities and inequalities, and vice versa. [119, §2.2]

2.12.0.0.1 Definition. Convex polyhedra, halfspace-description.

A convex polyhedron is the intersection of a finite number of halfspaces and hyperplanes;

$$\mathcal{P} = \{y \mid Ay \succeq b, Cy = d\} \subseteq \mathbb{R}^n$$ (285)

where coefficients $A$ and $C$ generally denote matrices. Each row of $C$ is a vector normal to a hyperplane, while each row of $A$ is a vector inward-normal to a hyperplane partially bounding a halfspace.

By the halfspaces theorem in §2.4.1.1.1, a polyhedron thus described is a closed convex set possibly not full-dimensional; e.g., Figure 21. Convex polyhedra are finite-dimensional comprising all affine sets (§2.3.1), polyhedral cones, line segments, rays, halfspaces, convex polygons, solids [233, def.104/6 p.343], polychora, polytopes, etcetera.

It follows from definition (285) by exposure that each face of a convex polyhedron is a convex polyhedron.

Projection of any polyhedron on a subspace remains a polyhedron. More generally, image and inverse image of a convex polyhedron under any linear transformation remains a convex polyhedron; [27, §I.9] [318, thm.19.3] the foremost consequence being, invariance of polyhedral set closedness.

When $b$ and $d$ in (285) are $0$, the resultant is a polyhedral cone. The set of all polyhedral cones is a subset of convex cones:

2.12.1 Polyhedral cone

From our study of cones, we see: the number of intersecting hyperplanes and halfspaces constituting a convex cone is possibly but not necessarily infinite. When the number is finite, the convex cone is termed polyhedral. That is the primary distinguishing feature between the set of all convex cones and polyhedra: all polyhedra, including polyhedral cones, are finitely generated [318, §19]. (Figure 54) We distinguish polyhedral cones in the set of all convex cones for this reason, although all convex cones of dimension 2 or less are polyhedral.

\[\text{We consider only convex polyhedra throughout, but acknowledge the existence of concave polyhedra. [405, Kepler-Poinsot Solid]}\]

\[\text{Some authors distinguish bounded polyhedra via the designation polytope. [119, §2.2]}\]
2.12.1.0.1 Definition. Polyhedral cone, halfspace-description.\(^{2.54}\) (confer (103))

A polyhedral cone is the intersection of a finite number of halfspaces and hyperplanes about the origin:

\[
\mathcal{K} = \{ y \mid Ay \succeq 0, Cy = 0 \} \subseteq \mathbb{R}^n \quad (a)
\]

\[
= \{ y \mid Ay \succeq 0, Cy \succeq 0, Cy \preceq 0 \} \quad (b)
\]

\[
= \left\{ y \mid \begin{bmatrix} A \\ C \\ -C \end{bmatrix} y \succeq 0 \right\} \quad (c)
\]

where coefficients \(A\) and \(C\) generally denote matrices of finite dimension. Each row of \(C\) is a vector normal to a hyperplane containing the origin, while each row of \(A\) is a vector inward-normal to a hyperplane containing the origin and partially bounding a halfspace.

A polyhedral cone thus defined is closed, convex (§2.4.1.1), has only a finite number of generators (§2.8.1.2), and can be not full-dimensional. (Minkowski) Conversely, any finitely generated convex cone is polyhedral. (Weyl) [340, §2.8] [318, thm.19.1]

2.12.1.0.2 Exercise. Unbounded convex polyhedra.

Illustrate an unbounded polyhedron that is not a cone or its translation.

From the definition it follows that any single hyperplane through the origin, or any halfspace partially bounded by a hyperplane through the origin is a polyhedral cone. The most familiar example of polyhedral cone is any quadrant (or orthant, §2.1.3) generated by Cartesian half-axes. Esoteric examples of polyhedral cone include the point at the origin, any line through the origin, any ray having the origin as base such as the nonnegative real line \(\mathbb{R}_+\) in subspace \(\mathbb{R}^n\), polyhedral flavor (proper) Lorentz cone (316), any subspace, and \(\mathbb{R}^n\). More polyhedral cones are illustrated in Figure 52 and Figure 26.

2.12.2 Vertices of convex polyhedra

By definition, a vertex (§2.6.1.0.1) always lies on the relative boundary of a convex polyhedron. [233, def.115/6 p.358] In Figure 21, each vertex of the polyhedron is located at an intersection of three or more facets, and every edge belongs to precisely two facets [27, §VI.1 p.252]. In Figure 26, the only vertex of that polyhedral cone lies at the origin.

The set of all polyhedral cones is clearly a subset of convex polyhedra and a subset of convex cones (Figure 54). Not all convex polyhedra are bounded; evidently, neither can they all be described by the convex hull of a bounded set of points as defined in (86).

Hence a universal vertex-description of polyhedra in terms of that same finite-length list \(X\) (76):

\(^{2.54}\)Rockafellar [318, §19] proposes affine sets be handled via complementary pairs of affine inequalities; e.g., antisymmetry \(Cy \preceq d\) and \(Cy \succeq d\) which can present severe difficulty to some interior-point methods of numerical solution.
2.12. CONVEX POLYHEDRA

Bounded polyhedra
convex polyhedra
convex cones
polyhedral cones

Figure 54: Polyhedral cones are finitely generated, unbounded, and convex.

2.12.2.0.1 Definition. Convex polyhedra, vertex-description. (confer §2.8.1.1.1)

Denote the truncated $a$-vector

$$a_{i:ℓ} = \begin{bmatrix} a_i \\ \vdots \\ a_ℓ \end{bmatrix}$$

(287)

By discriminating a suitable finite-length generating list (or set) arranged columnar in $X \in \mathbb{R}^{n \times N}$, then any particular polyhedron may be described

$$P = \{ Xa \mid a_{i:ℓ}^T 1 = 1, \ a_{m:N} \succeq 0, \ \{1\ldots k\} \cup \{m\ldots N\} = \{1\ldots N\} \}$$

(288)

where $0 \leq k \leq N$ and $1 \leq m \leq N+1$. Setting $k=0$ removes the affine equality condition. Setting $m=N+1$ removes the inequality.

Coefficient indices in (288) may or may not be overlapping, but all the coefficients are assumed constrained. From (78), (86), and (103), we summarize how the coefficient conditions may be applied;

affine sets $\rightarrow a_{i:ℓ}^T 1 = 1$

polyhedral cones $\rightarrow a_{m:N} \succeq 0$

$\leftarrow$ convex hull $(m \leq k)$

(289)

It is always possible to describe a convex hull in the region of overlapping indices because, for $1 \leq m \leq k \leq N$

$$\{a_{m:k} \mid a_{m:k}^T 1 = 1, \ a_{m:k} \succeq 0\} \subseteq \{a_{m:k} \mid a_{i:ℓ}^T 1 = 1, \ a_{m:N} \succeq 0\}$$

(290)

Members of a generating list are not necessarily vertices of the corresponding polyhedron; certainly true for (86) and (288), some subset of list members reside in the polyhedron’s relative interior. Conversely, when boundedness (86) applies, convex hull of the vertices is a polyhedron identical to convex hull of the generating list.

2.12.2.1 Vertex-description of polyhedral cone

Given closed convex cone $\mathcal{K}$ in a subspace of $\mathbb{R}^n$ having any set of generators for it arranged in a matrix $X \in \mathbb{R}^{n \times N}$ as in (279), then that cone is described setting $m=1$ and $k=0$ in
\[ S = \{ s \mid s \geq 0, \ 1^T s \leq 1 \} \]

Figure 55: Unit simplex \( S \) in \( \mathbb{R}^3 \) is a unique solid tetrahedron but not regular.

vertex-description (288):

\[ \mathcal{K} = \text{cone } X = \{ Xa \mid a \geq 0 \} \subseteq \mathbb{R}^n \quad (103) \]

a conic hull of \( N \) generators.

2.12.2.2 Pointedness

(§2.7.2.1.2) [340, §2.10] Assuming all generators constituting the columns of \( X \in \mathbb{R}^{n \times N} \) are nonzero, polyhedral cone \( \mathcal{K} \) is pointed if and only if there is no nonzero \( a \geq 0 \) that solves \( Xa = 0 \); id est, iff

\[
\begin{align*}
\text{find} & \quad a \\
\text{subject to} & \quad Xa = 0 \\
& \quad 1^T a = 1 \\
& \quad a \geq 0
\end{align*}
\]

is infeasible or iff \( \mathcal{N}(X) \cap \mathbb{R}_+^N = \emptyset \). Otherwise, the cone will contain at least one line and there can be no vertex; id est, the cone cannot otherwise be pointed. Any subspace, Euclidean vector space \( \mathbb{R}^n \), or any halfspace are examples of nonpointed polyhedral cone; hence, no vertex. This null-pointedness criterion \( Xa = 0 \) means that a pointed polyhedral cone is invariant to linear injective transformation.

Examples of pointed polyhedral cone \( \mathcal{K} \) include: the origin, any 0-based ray in a subspace, any two-dimensional V-shaped cone in a subspace, any orthant in \( \mathbb{R}^n \) or \( \mathbb{R}^{m \times n} \); e.g, nonnegative real line \( \mathbb{R}_+ \) in vector space \( \mathbb{R} \).

\[ \text{If } \text{rank } X = n, \text{ then the dual cone } \mathcal{K}^* (§2.13.1) \text{ is pointed. } (309) \]
2.12. CONVEX POLYHEDRA

2.12.3 Unit simplex

A peculiar subset of the nonnegative orthant with halfspace-description

\[ S \triangleq \{ s \mid s \succeq 0, \ 1^T s \leq 1 \} \subseteq \mathbb{R}^n_+ \]  

is a unique bounded convex full-dimensional polyhedron called unit simplex (Figure 55) having \( n + 1 \) facets, \( n + 1 \) vertices, and dimension

\[ \dim S = n \]  

The origin supplies one vertex while heads of the standard basis \( \{ e_i, i=1\ldots n \} \) in \( \mathbb{R}^n \) constitute those remaining;\(^{2,56} \) thus its vertex-description:

\[
S = \text{conv} \{ \mathbf{0}, \{ e_i, i=1\ldots n \} \} = \{ [0\ e_1\ e_2\ \cdots\ e_n]a \mid a^T 1 = 1,\ a \succeq 0 \} \]  

2.12.3.1 Simplex

The unit simplex comes from a class of general polyhedra called simplex, having vertex-description: \[ \text{conv} \{ x_\ell \in \mathbb{R}^n \mid \ell = 1\ldots k+1,\ \dim \text{aff} \{ x_\ell \} = k,\ n \geq k \]  

So defined, a simplex is a closed bounded convex set possibly not full-dimensional. Examples of simplices, by increasing affine dimension, are: a point, any line segment, any triangle and its relative interior, a general tetrahedron, any five-vertex polychoron, and so on.

2.12.3.1.1 Definition. Simplicial cone.

A proper polyhedral (§2.7.2.2.1) cone \( K \) in \( \mathbb{R}^n \) is called simplicial iff \( K \) has exactly \( n \) extreme directions;\(^{21} \) equivalently, iff proper \( K \) has exactly \( n \) linearly independent generators contained in any given set of generators. \( \triangle \)

- simplicial cone ⇒ proper polyhedral cone

There are an infinite variety of simplicial cones in \( \mathbb{R}^n \); e.g, Figure 26, Figure 56, Figure 66. Any orthant is simplicial, as is any rotation thereof.

2.12.4 Converting between descriptions

Conversion between halfspace- (285) (286) and vertex-description (86) (288) is nontrivial, in general,\(^{[16]} \)\(^{[119]} \)\(^{\S 2.2} \) but the conversion is easy for simplices.\(^{[63]} \)\(^{\S 2.2.4} \) Nonetheless, we tacitly assume the two descriptions to be equivalent.\(^{[318]} \)\(^{\S 19 \text{ thm.19.1}} \) We explore conversions in \( \S 2.13.4, \S 2.13.9, \) and \( \S 2.13.11 \):

\(^{2,56} \)In \( \mathbb{R}^0 \) the unit simplex is the point at the origin, in \( \mathbb{R} \) the unit simplex is the line segment \([0,1] \), in \( \mathbb{R}^2 \) it is a triangle and its relative interior, in \( \mathbb{R}^3 \) it is the convex hull of a tetrahedron (Figure 55), in \( \mathbb{R}^4 \) it is the convex hull of a pentatope \( [405], \) and so on.
Figure 56: Two views of a simplicial cone and its dual in $\mathbb{R}^3$ (second view on next page). Semiinfinite boundary of each cone is truncated for illustration. Each cone has three facets (confer §2.13.11.0.3). (Cartesian axes are drawn for reference.)
2.13 Dual cone & generalized inequality & biorthogonal expansion

These three concepts, dual cone, generalized inequality, and biorthogonal expansion, are inextricably melded; meaning, it is difficult to completely discuss one without mentioning the others. The dual cone is critical in tests for convergence by contemporary primal/dual methods for numerical solution of conic problems. For unique minimum-distance projection on a closed convex cone $K$, the negative dual cone $-K^*$ plays the role that orthogonal complement plays for subspace projection. Indeed, $-K^*$ is the algebraic complement in $\mathbb{R}^n$;

$$K \oplus -K^* = \mathbb{R}^n \quad (2121)$$

where $\oplus$ denotes unique orthogonal vector sum.

One way to think of a pointed closed convex cone is as a new kind of coordinate system whose basis is generally nonorthogonal; a conic system, very much like the familiar Cartesian system whose analogous cone is the first quadrant (the nonnegative orthant). Generalized inequality $\succeq_K$ is a formalized means to determine membership to any pointed closed convex cone $K$ (§2.7.2.2) whereas biorthogonal expansion is, fundamentally, an expression of coordinates in a pointed conic system whose axes are linearly independent but not necessarily orthogonal. When cone $K$ is the nonnegative orthant, then these three concepts come into alignment with the Cartesian prototype: biorthogonal expansion becomes orthogonal expansion, the dual cone becomes identical to the orthant, and generalized inequality obeys a total order entrywise.

2.13.1 Dual cone

For any set $K$ (convex or not), the dual cone $K^*$ [115, §4.2]

$$K^* \triangleq \{ y \in \mathbb{R}^n \mid \langle y, x \rangle \geq 0 \text{ for all } x \in K \} \quad (296)$$

is a unique cone that is always closed and convex because it is an intersection of halfspaces (§2.4.1.1.1). Each halfspace has inward-normal $x$, belonging to $K$, and boundary containing the origin; e.g., Figure 57a.

When cone $K$ is convex, there is a second and equivalent construction: Dual cone $K^*$ is the union of each and every vector $y$ inward-normal to a hyperplane supporting $K$ (§2.4.2.6.1); e.g., Figure 57b. When $K$ is represented by a halfspace-description such as (286), for example, where

$$A \triangleq \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad C \triangleq \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix} \in \mathbb{R}^{p \times n} \quad (297)$$

$2.57$ Namely, projection on a subspace is ascertainable from projection on its orthogonal complement (Figure 177).

$2.58$ The dual cone is the negative polar cone defined by many authors; $K^* = -K^\circ$. [209, §A.3.2] [318, §14] [42] [27] [340, §2.7]
then the dual cone can be represented as the conic hull
\[ K^* = \text{cone}\{a_1, \ldots, a_m, \pm c_1, \ldots, \pm c_p\} \] (298)
a vertex-description, because each and every conic combination of normals from the half-space-description of \( K \) yields another inward-normal to a hyperplane supporting \( K \).

\( K^* \) can also be constructed pointwise using projection theory from §E.9.2: for \( P_K x \) the Euclidean projection of point \( x \) on closed convex cone \( K \)

\[ -K^* = \{ x - P_K x \mid x \in \mathbb{R}^n \} = \{ x \in \mathbb{R}^n \mid P_K x = 0 \} \] (2122)

Perhaps the most instructive graphical method of dual cone construction is cut-and-try. Find the dual of each polyhedral cone from Figure 58 by using dual cone equation (296).

2.13.1.0.2 Exercise. Dual cone definitions.
What is \( \{ x \in \mathbb{R}^n \mid x^T z \geq 0 \ \forall z \in \mathbb{R}^n \} \)?
What is \( \{ x \in \mathbb{R}^n \mid x^T z \geq 1 \ \forall z \in \mathbb{R}^n \} \)?
What is \( \{ x \in \mathbb{R}^n \mid x^T z \geq 1 \ \forall z \in \mathbb{R}_+^n \} \)? ▼

As defined, dual cone \( K^* \) exists even when the affine hull of the original cone is a proper subspace; \( \text{id est} \), even when the original cone is not full-dimensional 2.59.

To further motivate our understanding of the dual cone, consider the ease with which convergence can be ascertained in the following optimization problem (301p):

2.13.1.0.3 Example. Dual problem. (confer §4.1)
Duality is a powerful and widely employed tool in applied mathematics for a number of reasons. First, the dual program is always convex even if the primal is not. Second, the number of variables in the dual is equal to the number of constraints in the primal which is often less than the number of variables in the primal program. Third, the maximum value achieved by the dual program is often equal to the minimum of the primal. [310, §2.1.3]
When not equal, the dual always provides a bound on the primal optimal objective. For convex problems, the dual variables provide necessary and sufficient optimality conditions:

Essentially, Lagrange duality theory concerns representation of a given optimization problem as half of a minimax problem. [318, §36] [63, §5.4] Given any real function \( f(x, z) \)

\[ \min_x \max_z f(x, z) \geq \max_z \min_x f(x, z) \] (299)
always holds. When

\[ \min_x \max_z f(x, z) = \max_z \min_x f(x, z) \] (300)

2.59 Rockafellar formulates dimension of \( K \) and \( K^* \). [318, §14.6.1] His monumental work Convex Analysis has not one figure or illustration. See [27, §III.16] for a good illustration of Rockafellar’s recession cone [43].
Figure 57: Two equivalent constructions of dual cone $\mathcal{K}^*$ in $\mathbb{R}^2$: (a) Showing construction by intersection of halfspaces about $\mathbf{0}$ (drawn truncated). Only those two halfspaces whose bounding hyperplanes have inward-normal corresponding to an extreme direction of this pointed closed convex cone $\mathcal{K} \subset \mathbb{R}^2$ need be drawn; by (368). (b) Suggesting construction by union of inward-normals $y$ to each and every hyperplane $\partial \mathcal{H}_+$ supporting $\mathcal{K}$. This interpretation is valid when $\mathcal{K}$ is convex because existence of a supporting hyperplane is then guaranteed (§2.4.2.6).
2.13. DUAL CONE & GENERALIZED INEQUALITY

Figure 58: Dual cone construction by right angle. Each extreme direction of a proper polyhedral cone is orthogonal to a facet of its dual cone, and vice versa, in any dimension. (§2.13.6.1)  
(a) This characteristic guides graphical construction of dual cone in two dimensions: It suggests finding dual-cone boundary \( \partial \mathcal{K}^* \) by making right angles with extreme directions of polyhedral cone. The construction is then pruned so that each dual boundary vector does not exceed \( \pi/2 \) radians in angle with each and every vector from polyhedral cone. Were dual cone in \( \mathbb{R}^2 \) to narrow, Figure 59 would be reached in limit.  
(b) Same polyhedral cone and its dual continued into three dimensions. (confer Figure 66)
Figure 59: Polyhedral cone $\mathcal{K}$ is a halfspace about origin in $\mathbb{R}^2$. Dual cone $\mathcal{K}^*$ is a ray base $0$, hence not full-dimensional in $\mathbb{R}^2$; so $\mathcal{K}$ cannot be pointed, hence has no extreme directions. (Both convex cones appear truncated.)

we have strong duality and then a saddle value [160] exists. (Figure 60) [315, p.3] Consider primal conic problem (p) (over cone $\mathcal{K}$) and its corresponding dual problem (d): [304, §3.3.1] [249, §2.1] [250] given vectors $\alpha, \beta$ and matrix constant $C$

$$\begin{align*}
\text{(p)} & \text{ minimize } & \alpha^T x & \text{ maximize } & \beta^T z \\
& \text{ subject to } & x \in \mathcal{K} & \text{ subject to } & y \in \mathcal{K}^* (d) \\
& & C x = \beta & C^T z + y = \alpha
\end{align*}$$

Observe: the dual problem is also conic, and its objective function value never exceeds that of the primal;

$$\begin{align*}
\alpha^T x & \geq \beta^T z \\
x^T (C^T z + y) & \geq (Cx)^T z \\
x^T y & \geq 0
\end{align*}$$

which holds by definition (296). Under the sufficient condition that (301p) is a convex problem\textsuperscript{2.60} satisfying Slater’s condition (p.247), then equality

$$x^* y^* = 0$$

is achieved; which is necessary and sufficient for optimality (§2.13.10.1.5); each problem (p) and (d) attains the same optimal value of its objective and each problem is called a strong dual to the other because the duality gap (optimal primal–dual objective difference)

\textsuperscript{2.60}In this context, problems (p) and (d) are convex if $\mathcal{K}$ is a convex cone.
2.13. DUAL CONE & GENERALIZED INEQUALITY

Figure 60: (Drawing by Lucas V. Barbosa.) This serves as mnemonic icon for primal and dual problems, although objective functions from conic problems \(301p\) \(301d\) are linear. When problems are strong duals, duality gap is 0; meaning, functions \(f(x)\), \(g(z)\) (dotted) kiss at saddle value as depicted at center. Otherwise, dual functions never meet \((f(x) > g(z))\) by (299).
becomes 0. Then (p) and (d) are together equivalent to the minimax problem
\[
\begin{align*}
\text{minimize} \quad & \alpha^T x - \beta^T z \\
\text{subject to} \quad & x \in K, \quad y \in K^* \\
& Cx = \beta, \quad CTz + y = \alpha
\end{align*}
\]
whose optimal objective always has the saddle value 0 (regardless of the particular convex cone \( K \) and other problem parameters). [371, §3.2] Thus determination of convergence for either primal or dual problem is facilitated.

Were convex cone \( K \) polyhedral (§2.12.1), then problems (p) and (d) would be linear programs. Selfdual nonnegative orthant \( K \) yields the primal prototypical linear program and its dual. Were \( K \) a positive semidefinite cone, then problem (p) has the form of prototypical semidefinite program (680) with (d) its dual. It is sometimes possible to solve a primal problem by way of its dual; advantageous when the dual problem is easier to solve than the primal problem, for example, because it can be solved analytically, or has some special structure that can be exploited. [63, §5.5.5] (§4.2.3.1)

2.13.1.1 Key properties of dual cone

- For any cone, \((-K)^* = -K^*\)
- For any cones \( K_1 \) and \( K_2 \), \( K_1 \subseteq K_2 \Rightarrow K_1^* \supseteq K_2^* \) [340, §2.7]
- (Cartesian product) For closed convex cones \( K_1 \) and \( K_2 \), their Cartesian product \( K = K_1 \times K_2 \) is a closed convex cone, and
\[
K^* = K_1^* \times K_2^*
\]
where each dual is determined with respect to a cone’s ambient space.

- (conjugation) [318, §14] [115, §4.5] [340, p.52] When \( K \) is any convex cone, dual of the dual cone equals closure of the original cone:
\[
K^{**} = \overline{K}
\]

If any cone \( K \) is full-dimensional, then \( K^* \) is pointed; \( K \) full-dimensional \( \Rightarrow \) \( K^* \) pointed (309)

\[
K = \{ x \in \mathbb{R}^n \mid \langle y, x \rangle \geq 0 \ \forall y \in K^* \}
\]
2.13. DUAL CONE & GENERALIZED INEQUALITY

If the closure of any convex cone $\mathcal{K}$ is pointed, conversely, then $\mathcal{K}^*$ is full-dimensional:

$$\overline{\mathcal{K}} \text{ pointed } \Rightarrow \mathcal{K}^* \text{ full-dimensional} \quad (310)$$

Given that a cone $\mathcal{K} \subseteq \mathbb{R}^n$ is closed and convex, $\mathcal{K}$ is pointed if and only if $\mathcal{K}^* - \mathcal{K}^* = \mathbb{R}^n$; id est, iff $\mathcal{K}^*$ is full-dimensional. [57, §3.3 exer.20]

- (vector sum) [318, thm.3.8] For convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$

$$\mathcal{K}_1 + \mathcal{K}_2 = \text{conv}(\mathcal{K}_1 \cup \mathcal{K}_2) \quad (311)$$

is a convex cone.

- (dual vector-sum) [318, §16.4.2] [115, §4.6] For convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$

$$\mathcal{K}_1^* \cap \mathcal{K}_2^* = (\mathcal{K}_1 + \mathcal{K}_2)^* = (\mathcal{K}_1 \cup \mathcal{K}_2)^* \quad (312)$$

- (closure of vector sum of duals)\(^\text{2.61}\) For closed convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$

$$(\mathcal{K}_1 \cap \mathcal{K}_2)^* = \mathcal{K}_1^* + \mathcal{K}_2^* = \text{conv}(\mathcal{K}_1^* \cup \mathcal{K}_2^*) \quad (313)$$

[340, p.96] where operation closure becomes superfluous under the sufficient condition $\mathcal{K}_1 \cap \text{int} \mathcal{K}_2 \neq \emptyset$ [57, §3.3 exer.16, §4.1 exer.7].

- (Krein-Rutman) Given closed convex cones $\mathcal{K}_1 \subseteq \mathbb{R}^m$ and $\mathcal{K}_2 \subseteq \mathbb{R}^n$ and any linear map $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$, then provided int $\mathcal{K}_1 \cap A \mathcal{K}_2 \neq \emptyset$ [57, §3.3.13, confer §4.1 exer.9]

$$(A^{-1} \mathcal{K}_1 \cap A \mathcal{K}_2)^* = A^T \mathcal{K}_1^* + \mathcal{K}_2^* \quad (314)$$

where dual of cone $\mathcal{K}_1$ is with respect to its ambient space $\mathbb{R}^m$ and dual of cone $\mathcal{K}_2$ is with respect to $\mathbb{R}^n$, where $A^{-1} \mathcal{K}_1$ denotes inverse image ($\S 2.1.9.0.1$) of $\mathcal{K}_1$ under mapping $A$, and where $A^T$ denotes adjoint operator. The particularly important case $\mathcal{K}_2 = \mathbb{R}^n$ is easy to show: for $A^T A = I$

$$(A^T \mathcal{K}_1)^* \triangleq \{ y \in \mathbb{R}^n \mid x^T y \geq 0 \ \forall x \in A^T \mathcal{K}_1 \} \quad (315)$$

- $\mathcal{K}$ is proper if and only if $\mathcal{K}^*$ is proper.

- $\mathcal{K}$ is polyhedral if and only if $\mathcal{K}^*$ is polyhedral. [340, §2.8]

- $\mathcal{K}$ is simplicial if and only if $\mathcal{K}^*$ is simplicial. (§2.13.9.2) A simplicial cone and its dual are proper polyhedral cones (Figure 66, Figure 56), but not the converse.

- $\mathcal{K} \oplus -\mathcal{K}^* = \mathbb{R}^n \Leftrightarrow \mathcal{K}$ is closed and convex. (2121)

- Any direction in a proper cone $\mathcal{K}$ is normal to a hyperplane separating $\mathcal{K}$ from $-\mathcal{K}^*$.

\(^\text{2.61}\)These parallel analogous results for subspaces $\mathcal{R}_1, \mathcal{R}_2 \subseteq \mathbb{R}^n$; [115, §4.6]

$$(\mathcal{R}_1 + \mathcal{R}_2)^\perp = \mathcal{R}_1^\perp \cap \mathcal{R}_2^\perp$$

$$(\mathcal{R}_1 \cap \mathcal{R}_2)^\perp = \mathcal{R}_1^\perp + \mathcal{R}_2^\perp$$

$\mathcal{R}^\perp = \mathcal{R}$ for any subspace $\mathcal{R}$. 

\[\]
Figure 61: When convex cone $\mathcal{K}$ is any one Cartesian axis, its dual cone is the convex hull of all axes remaining; its orthogonal complement. In $\mathbb{R}^3$, dual cone $\mathcal{K}^*$ (drawn tiled and truncated) is a hyperplane through origin; its normal belongs to line $\mathcal{K}$. In $\mathbb{R}^2$, dual cone $\mathcal{K}^*$ is a line through origin while convex cone $\mathcal{K}$ is that line orthogonal.
2.13. DUAL CONE & GENERALIZED INEQUALITY

2.13.1.2 Examples of dual cone

When \( K \) is \( \mathbb{R}^n \), \( K^\ast \) is the point at the origin, and vice versa.

When \( K \) is a subspace, \( K^\ast \) is its orthogonal complement, and vice versa. (\$E.9.2.1, Figure 61)

When cone \( K \) is a halfspace in \( \mathbb{R}^n \) with \( n > 0 \) (Figure 59 for example), the dual cone \( K^\ast \) is a ray (base \( 0 \)) belonging to that halfspace but orthogonal to its bounding hyperplane (that contains the origin), and vice versa.

When convex cone \( K \) is a closed halfplane in \( \mathbb{R}^3 \) (Figure 62), it is neither pointed or full-dimensional; hence, the dual cone \( K^\ast \) can be neither full-dimensional or pointed.

When \( K \) is any particular orthant in \( \mathbb{R}^n \), the dual cone is identical; \( \text{id est} \), \( K = K^\ast \).

When \( K \) is any quadrant in subspace \( \mathbb{R}^2 \), \( K^\ast \) is a wedge-shaped polyhedral cone in \( \mathbb{R}^3 \); \( \text{e.g.} \), for \( K \) equal to quadrant I, \( K^\ast = \left[ \mathbb{R}^2 + \mathbb{R} \right] \).

When \( K \) is a polyhedral flavor Lorentz cone (\( \text{confer} \) (178))

\[ K_\ell = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_\ell \leq t \right\}, \quad \ell \in \{1, \infty\} \] (316)

its dual is the proper cone [63, exmp.2.25]

\[ K_q = K_\ell^\ast = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_q \leq t \right\}, \quad \ell \in \{1, 2, \infty\} \] (317)

where \( \|x\|_\ell = \|x\|_q \) is that norm dual to \( \|x\|_\ell \) determined via solution to \( 1/\ell + 1/q = 1 \).

Figure 64 illustrates \( K = K_1 \) and \( K^\ast = K_1^\ast = K_\infty \) in \( \mathbb{R}^2 \times \mathbb{R} \).

2.13.2 Abstractions of Farkas’ lemma

2.13.2.0.1 Corollary. Generalized inequality and membership relation. [209, \$A.4.2]

Let \( K \) be any closed convex cone and \( K^\ast \) its dual, and let \( x \) and \( y \) belong to a vector space \( \mathbb{R}^n \). Then

\[ y \in K^\ast \iff \langle y, x \rangle \geq 0 \text{ for all } x \in K \quad (318) \]

which is, merely, a statement of fact by definition of dual cone (296). By closure we have conjugation: [318, thm.14.1]

\[ x \in K \iff \langle y, x \rangle \geq 0 \text{ for all } y \in K^\ast \quad (319) \]

which may be regarded as a simple translation of Farkas’ lemma [140] as in [318, \$22] to the language of convex cones, and a generalization of the well-known Cartesian cone fact

\[ x \geq 0 \iff \langle y, x \rangle \geq 0 \text{ for all } y \geq 0 \quad (320) \]

for which implicitly \( K = K^\ast = \mathbb{R}^n_+ \) the nonnegative orthant.

Membership relation (319) is often written instead as dual generalized inequalities, when \( K \) and \( K^\ast \) are pointed closed convex cones,

\[ x \geq 0 \iff \langle y, x \rangle \geq 0 \text{ for all } y \geq 0 \quad (321) \]
Figure 62: $\mathcal{K}$ and $\mathcal{K}^*$ are halfplanes in $\mathbb{R}^3$; *blades*. Both semiinfinite convex cones appear truncated. Each cone is like $\mathcal{K}$ from Figure 59 but embedded in a two-dimensional subspace of $\mathbb{R}^3$. (Cartesian coordinate axes drawn for reference.)
meaning, coordinates for biorthogonal expansion of $x$ (§2.13.7.1.2, §2.13.8) \[377\] must be nonnegative when $x$ belongs to $\mathcal{K}$. Conjugating, \[
y \succeq 0 \iff \langle y, x \rangle \geq 0 \text{ for all } x \preceq 0 \quad (322)\]

When pointed closed convex cone $\mathcal{K}$ is not polyhedral, coordinate axes for biorthogonal expansion asserted by the corollary are taken from extreme directions of $\mathcal{K}$; expansion is assured by Carathéodory’s theorem (§E.6.4.1.1).

We presume, throughout, the obvious:

\[
x \in \mathcal{K} \iff \langle y, x \rangle \geq 0 \text{ for all } y \in \mathcal{K}^* \quad (319)\]
\[
x \in \mathcal{K} \iff \langle y, x \rangle \geq 0 \text{ for all } y \in \mathcal{K}^*, \|y\| = 1 \quad (323)\]

\[2.13.2.0.2 \text{ Exercise.} \quad \text{Dual generalized inequalities.}\]

Test Corollary 2.13.2.0.1 and (323) graphically on the two-dimensional polyhedral cone and its dual in Figure 58.

\[\text{\color{red}{\textit{(confer} \S 2.7.2.2)} \textit{\}When pointed closed convex cone } \mathcal{K} \text{ is implicit from context:}\]

\[
x \geq 0 \iff x \in \mathcal{K} \]
\[
x > 0 \iff x \in \text{rel int } \mathcal{K} \quad (324)\]

Strict inequality $x > 0$ means coordinates for biorthogonal expansion of $x$ must be positive when $x$ belongs to $\text{rel int } \mathcal{K}$. Strict membership relations are useful; e.g., for any proper cone\[2.62\] $\mathcal{K}$ and its dual $\mathcal{K}^*$

\[
x \in \text{int } \mathcal{K} \iff \langle y, x \rangle > 0 \text{ for all } y \in \mathcal{K}^*, y \neq 0 \quad (325)\]
\[
x \in \mathcal{K}, x \neq 0 \iff \langle y, x \rangle > 0 \text{ for all } y \in \text{int } \mathcal{K}^* \quad (326)\]

Conjugating, we get the dual relations:

\[
y \in \text{int } \mathcal{K}^* \iff \langle y, x \rangle > 0 \text{ for all } x \in \mathcal{K}, x \neq 0 \quad (327)\]
\[
y \in \mathcal{K}^*, y \neq 0 \iff \langle y, x \rangle > 0 \text{ for all } x \in \text{int } \mathcal{K} \quad (328)\]

Boundary-membership relations for proper cones are also useful:

\[
x \in \partial \mathcal{K} \iff \exists y \neq 0 \ni \langle y, x \rangle = 0, y \in \mathcal{K}^*, x \in \mathcal{K} \quad (329)\]
\[
y \in \partial \mathcal{K} \iff \exists x \neq 0 \ni \langle y, x \rangle = 0, x \in \mathcal{K}, y \in \mathcal{K}^* \quad (330)\]

which are consistent; e.g., $x \in \partial \mathcal{K} \iff x \notin \text{int } \mathcal{K}$ and $x \in \mathcal{K}$.

\[2.62\text{ An open cone } \mathcal{K} \text{ is admitted to (325) and (328) by (19).}\]
CHAPTER 2. CONVEX GEOMETRY

2.13.2.0.3 Example. Linear inequality. [347, §4] (confer §2.13.5.1.1)
Consider a given matrix \( A \) and closed convex cone \( \mathcal{K} \). By membership relation we have
\[
Ay \in \mathcal{K}^* \iff x^T Ay \geq 0 \quad \forall x \in \mathcal{K}
\]
\[
\iff y^T z \geq 0 \quad \forall z \in \{ A^T x \mid x \in \mathcal{K} \}
\]
\[
\iff y \in \{ A^T x \mid x \in \mathcal{K} \}^*
\]
This implies
\[
\{ y \mid Ay \in \mathcal{K}^* \} = \{ A^T x \mid x \in \mathcal{K} \}^*
\]
When \( \mathcal{K} \) is the selfdual nonnegative orthant (§2.13.1), for example, then
\[
\{ y \mid Ay \succeq 0 \} = \{ A^T x \mid x \succeq 0 \}^*
\]
and the dual relation
\[
\{ y \mid Ay \succeq 0 \}^* = \{ A^T x \mid x \succeq 0 \}
\]
comes by a theorem of Weyl (p.128) that yields closedness for any vertex-description of a polyhedral cone.

2.13.2.1 Null certificate, Theorem of the alternative
If in particular \( x_p \notin \mathcal{K} \) a closed convex cone, then construction in Figure 57b suggests there exists a supporting hyperplane (having inward-normal belonging to dual cone \( \mathcal{K}^* \)) separating \( x_p \) from \( \mathcal{K} \); indeed, (319)
\[
x_p \notin \mathcal{K} \iff \exists y \in \mathcal{K}^* \ni \langle y, x_p \rangle < 0
\]
Existence of any one such \( y \) is a certificate of null membership. From a different perspective,
\[
x_p \in \mathcal{K}
\]
or in the alternative
\[
\exists y \in \mathcal{K}^* \ni \langle y, x_p \rangle < 0
\]
By alternative is meant: these two systems are incompatible; one system is feasible while the other is not.

2.13.2.1.1 Example. Theorem of the alternative for linear inequality.
Myriad alternative systems of linear inequality can be explained in terms of pointed closed convex cones and their duals.
Beginning from the simplest Cartesian dual generalized inequalities (320) (with respect to nonnegative orthant \( \mathbb{R}^n_+ \)),
\[
y \succeq 0 \iff x^T y \geq 0 \quad \text{for all } x \succeq 0
\]
Given \( A \in \mathbb{R}^{n \times m} \), we make vector substitution \( y \leftarrow A^T y \)
\[
A^T y \succeq 0 \iff x^T A^T y \geq 0 \quad \text{for all } x \succeq 0
\]
Introducing a new vector by calculating $b \triangleq Ax$ we get

$$A^T y \succeq 0 \iff b^T y \succeq 0, \quad b = Ax \text{ for all } x \succeq 0$$ (339)

By complementing sense of the scalar inequality:

$$A^T y \succeq 0$$

or in the alternative

$$b^T y < 0, \quad \exists b = Ax, \quad x \succeq 0$$ (340)

If one system has a solution, then the other does not; define a convex cone $K = \{ y \mid A^T y \succeq 0 \}$, then $y \in K$ or in the alternative $y \not\in K$.

Scalar inequality $b^T y < 0$ is movable to the other side of alternative (340), but that requires some explanation: From results in Example 2.13.2.0.3, the dual cone is $K^* = \{ Ax \mid x \succeq 0 \}$. From (319) we have

$$y \in K \iff b^T y \geq 0 \text{ for all } b \in K^*$$ (341)

Given some $b$ vector and $y \in K$, then $b^T y < 0$ can only mean $b \not\in K^*$. An alternative system is therefore simply $b \in K^*$: [209, p.59] (Farkas/Tucker)

$$A^T y \succeq 0, \quad b^T y < 0$$

or in the alternative

$$b = Ax, \quad x \succeq 0$$ (342)

Geometrically this means: either vector $b$ belongs to convex cone $K^*$ or it does not. When $b \not\in K^*$, then there is a vector $y \in K$ normal to a hyperplane separating point $b$ from cone $K^*$.

For another example, from membership relation (318) with affine transformation of dual variable we may write: Given $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$

$$b - Ay \in K^* \iff x^T (b - Ay) \geq 0, \quad \forall x \in K$$ (343)

$$A^T x = 0, \quad b - Ay \in K^* \implies x^T b \geq 0, \quad \forall x \in K$$ (344)

From membership relation (343), conversely, suppose we allow any $y \in \mathbb{R}^m$. Then because $-x^T Ay$ is unbounded below, $x^T (b - Ay) \geq 0$ implies $A^T x = 0$: for $y \in \mathbb{R}^m$

$$A^T x = 0, \quad b - Ay \in K^* \iff x^T (b - Ay) \geq 0, \quad \forall x \in K$$ (345)

In toto,

$$b - Ay \in K^* \iff x^T b \succeq 0, \quad A^T x = 0 \quad \forall x \in K$$ (346)

Vector $x$ belongs to cone $K$ but is also constrained to lie in a subspace of $\mathbb{R}^n$ specified by an intersection of hyperplanes through the origin $\{ x \in \mathbb{R}^n \mid A^T x = 0 \}$. From this, alternative systems of generalized inequality with respect to pointed closed convex cones $K$ and $K^*$.
CHAPTER 2. CONVEX GEOMETRY

\[ Ay \preceq b \]

or in the alternative
\[ x^Tb < 0, \ A^T x = 0, \ x \succeq 0 \]

derived from (346) simply by taking the complementary sense of the inequality in \( x^Tb \).
These two systems are alternatives; if one system has a solution, then the other does not. \[ 2.63 \] [318, p.201]

By invoking a strict membership relation between proper cones (325), we can construct a more exotic alternative strengthened by demand for an interior point;

\[ b - Ay \succ 0 \quad \Leftrightarrow \quad x^Tb > 0, \ A^T x = 0 \quad \forall x \succeq 0, \ x \neq 0 \]

From this, alternative systems of generalized inequality [63, pp.50,54,262]

\[ Ay \prec b \]

or in the alternative
\[ x^Tb \leq 0, \ A^T x = 0, \ x \succeq 0, \ x \neq 0 \]

derived from (348) by taking complementary sense of the inequality in \( x^Tb \).

And from this, alternative systems with respect to the nonnegative orthant attributed to Gordan in 1873: [169] [57, §2.2] substituting \( A \leftarrow A^T \) and setting \( b = 0 \)

\[ A^T y \prec 0 \]

or in the alternative
\[ Ax = 0, \ x \succeq 0, \ ||x||_1 = 1 \]

Ben-Israel collects related results from Farkas, Motzkin, Gordan, and Stiemke in Motzkin transposition theorem. [34] \[ \square \]

\[ 2.63 \] If solutions at \( \pm \infty \) are disallowed, then the alternative systems become instead mutually exclusive with respect to nonpolyhedral cones. Simultaneous infeasibility of the two systems is not precluded by mutual exclusivity; called a weak alternative. Ye provides an example illustrating simultaneous infeasibility with respect to the positive semidefinite cone: \( x \in \mathbb{S}^2, \ y \in \mathbb{R}, \ A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ and \ b = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \) where \( x^T b \) means \( \langle x, b \rangle \). A better strategy than disallowing solutions at \( \pm \infty \) is to demand an interior point as in (349) or Lemma 4.2.1.1.2. Then question of simultaneous infeasibility is moot.
2.13. **DUAL CONE & GENERALIZED INEQUALITY**

2.13.3 Optimality condition

(see §2.13.10.1) The general first-order necessary and sufficient condition for optimality of solution \(x^*\) to a minimization problem ((301p) for example) with real differentiable convex objective function \(f(x) : \mathbb{R}^n \rightarrow \mathbb{R}\) is [317, §3]

\[
\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in \mathcal{C}, \quad x^* \in \mathcal{C} \tag{351}
\]

where \(\mathcal{C}\) is a convex feasible set and where \(\nabla f(x^*)\) is the gradient (§3.6) of \(f\) with respect to \(x\) evaluated at \(x^*\). In words, negative gradient is normal to a hyperplane supporting the feasible set at a point of optimality. (Figure 69)

Direct solution to variation inequality (351), instead of the corresponding minimization, spawned from calculus of variations. [260, p.178] [139, p.37] One solution method solves an equivalent fixed point-of-projection problem

\[
x = P_{\mathcal{C}}(x - \nabla f(x)) \tag{352}
\]

that follows from a necessary and sufficient condition for projection on convex set \(\mathcal{C}\) (Theorem E.9.1.0.2)

\[
P(x^* - \nabla f(x^*)) \in \mathcal{C}, \quad \langle x^* - \nabla f(x^*) - x^*, x - x^* \rangle \leq 0 \quad \forall x \in \mathcal{C} \tag{2107}
\]

Proof of equivalence [381, Complementarity problem] is provided by Németh. Given minimum-distance projection problem

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \|x - y\|^2 \\
\text{subject to} & \quad x \in \mathcal{C}
\end{align*} \tag{353}
\]

on convex feasible set \(\mathcal{C}\) for example, the equivalent fixed point problem

\[
x = P_{\mathcal{C}}(x - \nabla f(x)) = P_{\mathcal{C}}(y) \tag{354}
\]

is solved in one step.

In the unconstrained case (\(\mathcal{C} = \mathbb{R}^n\)), optimality condition (351) reduces to the classical rule (p.212)

\[
\nabla f(x^*) = 0, \quad x^* \in \text{dom } f \tag{355}
\]

which can be inferred from the following application:

**2.13.3.0.1 Example.** *Optimality for equality constrained problem.*

Given a real differentiable convex function \(f(x) : \mathbb{R}^n \rightarrow \mathbb{R}\) defined on domain \(\mathbb{R}^n\), a full-rank matrix \(C \in \mathbb{R}^{p \times n}\), and vector \(d \in \mathbb{R}^p\), the convex optimization problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad Cx = d
\end{align*} \tag{356}
\]

preumably nonempty set of all variable values satisfying all given problem constraints; the set of feasible solutions.
is characterized by the well-known necessary and sufficient optimality condition [63, §4.2.3]

\[ \nabla f(x^\ast) + C^T \nu = 0 \]  

(357)

where \( \nu \in \mathbb{R}^p \) is the eminent Lagrange multiplier. [316] [260, p.188] [239] In other words, solution \( x^\ast \) is optimal if and only if \( \nabla f(x^\ast) \) belongs to \( \mathcal{R}(C^T) \).

Via membership relation, we now derive condition (357) from the general first-order condition for optimality (351): For problem (356)

\[ \mathcal{C} \triangleq \{ x \in \mathbb{R}^n \mid Cx = d \} = \{ Z \xi + x_p \mid \xi \in \mathbb{R}^{n-\text{rank}C} \} \]

(358)

is the feasible set where \( Z \in \mathbb{R}^{n \times n-\text{rank}C} \) holds basis \( \mathcal{N}(C) \) columnar, and \( x_p \) is any particular solution to \( C x = d \). Since \( x^\ast \in \mathcal{C} \), we arbitrarily choose \( x_p = x^\ast \) which yields an equivalent optimality condition

\[ \nabla f(x^\ast)^T Z \xi \geq 0 \quad \forall \xi \in \mathbb{R}^{n-\text{rank}C} \]

(359)

when substituted into (351). But this is simply half of a membership relation where the cone dual to \( \mathbb{R}^{n-\text{rank}C} \) is the origin in \( \mathbb{R}^{n-\text{rank}C} \). We must therefore have

\[ Z^T \nabla f(x^\ast) = 0 \iff \nabla f(x^\ast)^T Z \xi \geq 0 \quad \forall \xi \in \mathbb{R}^{n-\text{rank}C} \]

(360)

meaning, \( \nabla f(x^\ast) \) must be orthogonal to \( \mathcal{N}(C) \). These conditions

\[ Z^T \nabla f(x^\ast) = 0, \quad Cx^\ast = d \]

(361)

are necessary and sufficient for optimality.

\[ \Box \]

2.13.4 Discretization of membership relation

2.13.4.1 Dual halfspace-description

Halfspace-description of dual cone \( \mathcal{K}^\ast \) is equally simple as vertex-description

\[ \mathcal{K} = \text{cone}(X) = \{ Xa \mid a \geq 0 \} \subseteq \mathbb{R}^n \]  

(103)

for corresponding closed convex cone \( \mathcal{K} \): By definition (296), for \( X \in \mathbb{R}^{n \times n} \) as in (279), (confer (286))

\[ \mathcal{K}^\ast = \{ y \in \mathbb{R}^n \mid z^T y \geq 0 \text{ for all } z \in \mathcal{K} \} \]

\[ = \{ y \in \mathbb{R}^n \mid z^T y \geq 0 \text{ for all } z = Xa, \ a \geq 0 \} \]

\[ = \{ y \in \mathbb{R}^n \mid a^T X^T y \geq 0, \ a \geq 0 \} \]

\[ = \{ y \in \mathbb{R}^n \mid X^T y \geq 0 \} \]

(362)

that follows from the generalized inequality and membership corollary (320). The semi-infinity of tests specified by all \( z \in \mathcal{K} \) has been reduced to a set of generators for \( \mathcal{K} \) constituting the columns of \( X \); id est, the test has been discretized.

Whenever cone \( \mathcal{K} \) is known to be closed and convex, the conjugate statement must also hold; id est, given any set of generators for dual cone \( \mathcal{K}^\ast \) arranged columnar in \( Y \), then the consequent vertex-description of dual cone connotes a halfspace-description for \( \mathcal{K} \): [340, §2.8]

\[ \mathcal{K}^\ast = \{ Ya \mid a \geq 0 \} \iff \mathcal{K}^{**} = \mathcal{K} = \{ z \mid Y^T z \geq 0 \} \]

(363)
2.13. DUAL CONE & GENERALIZED INEQUALITY

2.13.4.2 First dual-cone formula

From these two results (362) and (363) we deduce a general principle:

- From any given vertex-description (103) of closed convex cone $K$, a halfspace-description of the dual cone $K^*$ is immediate by matrix transposition (362); conversely, from any given halfspace-description (286) of $K$, a dual vertex-description is immediate (363). [318, p.122]

Various other converses are just a little trickier. (§2.13.9, §2.13.11)

We deduce further: For any polyhedral cone $K$, the dual cone $K^*$ is also polyhedral and $K^{**} = K$. [340, p.56]

The generalized inequality and membership corollary is discretized in the following theorem inspired by (362) and (363):

2.13.4.2.1 Theorem. Discretized membership. (confer §2.13.2.0.1)\textsuperscript{2,65}

Given any set of generators, (§2.8.1.2) denoted by $G(K)$ for closed convex cone $K \subseteq \mathbb{R}^n$, and any set of generators denoted $G(K^*)$ for its dual such that $K = \text{cone} G(K)$, $K^* = \text{cone} G(K^*)$ (364)
then discretization of the generalized inequality and membership corollary (p.143) is necessary and sufficient for certifying cone membership: for $x$ and $y$ in vector space $\mathbb{R}^n$

$$x \in K \iff \langle \gamma^*, x \rangle \geq 0 \text{ for all } \gamma^* \in G(K^*)$$

$$y \in K^* \iff \langle \gamma, y \rangle \geq 0 \text{ for all } \gamma \in G(K)$$

\textbf{Proof.} $K^* = \{ G(K^*) a \mid a \geq 0 \}$. $y \in K^* \iff y = G(K^*) a$ for some $a \geq 0$.

$x \in K \iff \langle y, x \rangle \geq 0 \forall y \in K^* \iff \langle G(K^*) a, x \rangle \geq 0 \forall a \geq 0$ (319). $a \equiv \sum_i \alpha_i e_i$ where $e_i$ is the $i^{th}$ member of a standard basis of possibly infinite cardinality. $\langle G(K^*) a, x \rangle \geq 0 \forall a \geq 0$ $\iff \sum_i \alpha_i \langle G(K^*) e_i, x \rangle \geq 0 \forall \alpha_i \geq 0 \iff \langle G(K^*) e_i, x \rangle \geq 0 \forall i$. Conjugate relation (366) is similarly derived.

\textbf{2.13.4.2.2 Exercise. Discretized dual generalized inequalities.}

Test Theorem 2.13.4.2.1 on Figure 58a using extreme directions there as generators.

\textbf{\textdagger}

From the \textit{discretized membership theorem} we may further deduce a more surgical description of closed convex cone that prescribes only a finite number of halfspaces for its construction when polyhedral: (Figure 57a)

$$K = \{ x \in \mathbb{R}^n \mid \langle \gamma^*, x \rangle \geq 0 \text{ for all } \gamma^* \in G(K^*) \}$$

$$K^* = \{ y \in \mathbb{R}^n \mid \langle \gamma, y \rangle \geq 0 \text{ for all } \gamma \in G(K) \}$$

\textsuperscript{2,65}Stated in [22, §1] without proof for pointed closed convex case.
CHAPTER 2. CONVEX GEOMETRY

Figure 63: $x \succeq 0$ with respect to $\mathcal{K}$ but not with respect to nonnegative orthant $\mathbb{R}_+^2$ (pointed convex cone $\mathcal{K}$ drawn truncated).

2.13.4.2.3 Exercise. Partial order induced by orthant.

When comparison is with respect to the nonnegative orthant $\mathcal{K} = \mathbb{R}_+^n$, then from the discretized membership theorem it directly follows:

$$x \preceq z \iff x_i \leq z_i \quad \forall \ i$$  \hspace{1cm} (369)

Generate simple counterexamples demonstrating that this equivalence with entrywise inequality holds only when the underlying cone inducing partial order is the nonnegative orthant; e.g., explain Figure 63.

2.13.4.2.4 Example. Boundary membership to proper polyhedral cone.

For a polyhedral cone, test (329) of boundary membership can be formulated as a linear program. Say proper polyhedral cone $\mathcal{K}$ is specified completely by generators that are arranged columnar in

$$X = \begin{bmatrix} \Gamma_1 & \cdots & \Gamma_N \end{bmatrix} \in \mathbb{R}^{n \times N}$$  \hspace{1cm} (279)

id est, $\mathcal{K} = \{ xa \mid a \succeq 0 \}$. Then membership relation

$$c \in \partial \mathcal{K} \iff \exists y \neq 0 \quad \langle y, c \rangle = 0, \ y \in \mathcal{K}^*, \ c \in \mathcal{K}$$  \hspace{1cm} (329)

may be expressed$^{2.66}$

\begin{align*}
\text{find} & \quad y \neq 0 \\
\text{subject to} & \quad c^T y = 0 \\
& \quad X^T y \succeq 0 \\
& \quad X a = c \\
& \quad a \succeq 0
\end{align*}  \hspace{1cm} (370)

This linear feasibility problem has a solution iff $c \in \partial \mathcal{K}$.

$^{2.66}$Nonzero $y \in \mathbb{R}^n$ may be realized in many ways; e.g., via constraint $1^T y = 1$ if $c \neq 1$.  \hspace{1cm} \Box
2.13. DUAL CONE & GENERALIZED INEQUALITY

2.13.4.3 smallest face of pointed closed convex cone

Given nonempty convex subset \( C \) of a convex set \( K \), the smallest face of \( K \) containing \( C \) is equivalent to intersection of all faces of \( K \) that contain \( C \). By (308), membership relation (329) means that each and every point on boundary \( \partial K \) of proper cone \( K \) belongs to a hyperplane supporting \( K \) whose normal \( y \) belongs to dual cone \( K^* \). It follows that the smallest face \( F \), containing \( C \subseteq \partial K \subseteq \mathbb{R}^n \) on boundary of proper cone \( K \), is the intersection of all hyperplanes containing \( C \) whose normals are in \( K^* \);

\[
F(K \ni c) = \{ x \in K \mid x \perp K^* \cap c^\perp \}
\]  

(371)

where

\[
C^\perp \triangleq \{ y \in \mathbb{R}^n \mid \langle z, y \rangle = 0 \quad \forall z \in C \}
\]  

(372)

When \( C \cap \text{int} K \neq \emptyset \) then \( F(K \ni c) = K \).

2.13.4.3.1 Example. Finding smallest face of cone.

Suppose polyhedral cone \( K \) is completely specified by generators arranged columnar in

\[
X = [ \Gamma_1 \cdots \Gamma_N ] \in \mathbb{R}^{n \times N}
\]  

(279)

To find its smallest face \( F(K \ni c) \) containing a given point \( c \in K \), by the discretized membership theorem 2.13.4.2.1, it is necessary and sufficient to find generators for the smallest face. We may do so one generator at a time:

Consider generator \( \Gamma_i \). If there exists a vector \( z \in K^* \) orthogonal to \( c \) but not to \( \Gamma_i \), then \( \Gamma_i \) cannot belong to the smallest face of \( K \) containing \( c \). Such a vector \( z \) can be realized by a linear feasibility problem:

\[
\begin{align*}
\text{find} & \quad z \in \mathbb{R}^n \\
\text{subject to} & \quad c^T z = 0 \\
& \quad X^T z \succeq 0 \\
& \quad \Gamma_i^T z = 1
\end{align*}
\]  

(373)

If there exists a solution \( z \) for which \( \Gamma_i^T z = 1 \), then

\[
\Gamma_i \notin F(K \ni c) \quad \{ z \in \mathbb{R}^n \mid X^T z \succeq 0, \ c^T z = 0 \}
\]  

(374)

so \( \Gamma_i \notin F(K \ni c) \); solution \( z \) is a certificate of null membership. If this problem is infeasible for generator \( \Gamma_i \in K \), conversely, then \( \Gamma_i \in F(K \ni c) \) by (371) and (362) because \( \Gamma_i \perp K^* \cap c^\perp \); in that case, \( \Gamma_i \) is a generator of \( F(K \ni c) \).

Since the constant in constraint \( \Gamma_i^T z = 1 \) is arbitrary positively, then by theorem of the alternative there is correspondence between (373) and (347) admitting the alternative linear problem: for a given point \( c \in K \)

\[
\begin{align*}
\text{find} & \quad a, \mu \\
\text{subject to} & \quad \mu c - \Gamma_i = X a \\
& \quad a \succeq 0
\end{align*}
\]  

(375)

Now if this problem is feasible (bounded) for generator \( \Gamma_i \in K \), then (373) is infeasible and \( \Gamma_i \in F(K \ni c) \) is a generator of the smallest face that contains \( c \).
2.13.4.3.2 Exercise. Finding smallest face of pointed closed convex cone.
Show that formula (371) and algorithms (373) and (375) apply more broadly; id est, a full-dimensional cone \( K \) is an unnecessary condition.2.68

2.13.4.3.3 Exercise. Smallest face of positive semidefinite cone.
Derive (221) from (371).

2.13.5 Dual PSD cone and generalized inequality
The dual positive semidefinite cone \( K^* \) is confined to \( \mathbb{S}^M \) by convention:
\[
\mathbb{S}^M_+^* \triangleq \{ Y \in \mathbb{S}^M \ | \ \langle Y, X \rangle \geq 0 \text{ for all } X \in \mathbb{S}^M_+ \} = \mathbb{S}^M_+ \tag{376}
\]
The positive semidefinite cone is selfdual in the ambient space of symmetric matrices [63, exmp.2.24] [40] [205, §II]: \( K = K^* \).

Dual generalized inequalities with respect to the positive semidefinite cone in the ambient space of symmetric matrices can therefore be simply stated: (Fejér)
\[
X \succeq 0 \iff \text{tr}(Y^TX) \geq 0 \text{ for all } Y \succeq 0 \tag{377}
\]
Membership to this cone can be determined in the isometrically isomorphic Euclidean space \( \mathbb{R}^{M^2} \) via (38). (§2.2.1) By the two interpretations in §2.13.1, positive semidefinite matrix \( Y \) can be interpreted as inward-normal to a hyperplane supporting the positive semidefinite cone.

The fundamental statement of positive semidefiniteness, \( y^TXy \geq 0 \forall y \) (§A.3.0.0.1), evokes a particular instance of these dual generalized inequalities (377):
\[
X \succeq 0 \iff \langle yy^T, X \rangle \geq 0 \forall yy^T(\succeq 0) \tag{1524}
\]
Discretization (§2.13.4.2.1) allows replacement of positive semidefinite matrices \( Y \) with this minimal set of generators comprising the extreme directions of the positive semidefinite cone (§2.9.2.7).

2.13.5.1 selfdual cones
From (131) (a consequence of the halfspaces theorem, §2.4.1.1.1), where the only finite value of the support function for a convex cone is 0 [209, §C.2.3.1], or from discretized definition (368) of the dual cone we get a rather self evident characterization of selfdual cones:
\[
K = K^* \iff K = \bigcap_{\gamma \in \mathcal{G}(K)} \{ y \ | \ \gamma^T y \geq 0 \} \tag{378}
\]
In words: Cone \( K \) is selfdual iff its own extreme directions are inward-normals to a (minimal) set of hyperplanes whose intersection constructs it. This means each extreme direction of \( K \) is normal to a hyperplane exposing one of its own faces; a necessary but insufficient condition for selfdualness (Figure 64, for example).

2.68 Hint: A hyperplane, with normal in \( K^* \), containing cone \( K \) is admissible.
Figure 64: Two (truncated) views of a polyhedral cone $\mathcal{K}$ and its dual in $\mathbb{R}^3$. Each of four extreme directions from $\mathcal{K}$ belongs to a face of dual cone $\mathcal{K}^*$. Cone $\mathcal{K}$, shrouded by its dual, is symmetrical about its axis of revolution. Each pair of diametrically opposed extreme directions from $\mathcal{K}$ makes a right angle. An orthant (or any rotation thereof; a simplicial cone) is not the only selfdual polyhedral cone in three or more dimensions; [21, §2.A.21] e.g., consider an equilateral having five extreme directions. In fact, every selfdual polyhedral cone in $\mathbb{R}^3$ has an odd number of extreme directions. [23, thm.3]
Selfdual cones are necessarily full-dimensional. [31, §I] Their most prominent representatives are the orthants (Cartesian cones), the positive semidefinite cone $\mathbb{S}_+^n$ in the ambient space of symmetric matrices (376), and Lorentz cone (178) [21, §II.A] [63, exmp.2.25]. In three dimensions, a plane containing the axis of revolution of a selfdual cone (and the origin) will produce a slice whose boundary makes a right angle.

2.13.5.1.1 Example. Linear matrix inequality. (confer §2.13.2.0.3)

Consider a peculiar vertex-description for a convex cone $K$ defined over a positive semidefinite cone ($\text{instead of a nonnegative orthant as in definition (103)}$): for $X \in \mathbb{S}_+^n$ given $A_j \in \mathbb{S}_+^n$, $j = 1 \ldots m$

$$K = \left\{ \begin{bmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{bmatrix} \mid X \succeq 0 \right\} \subseteq \mathbb{R}^m$$

$$= \left\{ \begin{bmatrix} \text{svec}(A_1)^T \\ \vdots \\ \text{svec}(A_m)^T \end{bmatrix} \text{svec} X \mid X \succeq 0 \right\} \triangleq \{ A \text{svec} X \mid X \succeq 0 \}$$

where $A \in \mathbb{R}^{m \times n(n+1)/2}$, and where symmetric vectorization svec is defined in (56). Cone $K$ is indeed convex because, by (175)

$$A \text{svec} X_{p1}, A \text{svec} X_{p2} \in K \Rightarrow A(\zeta \text{svec} X_{p1} + \xi \text{svec} X_{p2}) \in K \text{ for all } \zeta, \xi \geq 0$$

since a nonnegatively weighted sum of positive semidefinite matrices must be positive semidefinite. (§A.3.1.0.2) Although matrix $A$ is finite-dimensional, $K$ is generally not a polyhedral cone (unless $m = 1$ or 2) simply because $X \in \mathbb{S}_+^n$.

Theorem. Inverse image closedness. [209, prop.A.2.1.12] [318, thm.6.7]

Given affine operator $g : \mathbb{R}^m \to \mathbb{R}^p$, convex set $D \subseteq \mathbb{R}^m$, and convex set $C \subseteq \mathbb{R}^p$ \hspace{1cm} $g^{-1}(\text{rel int } C) \neq \emptyset$, then

$$\text{rel int } g(D) = g(\text{rel int } D), \quad \text{rel int } g^{-1}C = g^{-1}(\text{rel int } C), \quad g^{-1}C = g^{-1}C$$

By this theorem, relative interior of $K$ may always be expressed

$$\text{rel int } K = \{ A \text{ svec } X \mid X \succ 0 \}$$

Because $\dim(\text{aff } K) = \dim(A \text{ svec } \mathbb{S}_+^n)$ (127) then, provided the vectorized $A_j$ matrices are linearly independent,

$$\text{rel int } K = \text{int } K$$

meaning, cone $K$ is full-dimensional $\Rightarrow$ dual cone $K^*$ is pointed by (309). Convex cone $K$ can be closed, by this corollary:
2.13. DUAL CONE & GENERALIZED INEQUALITY

Corollary. Cone closedness invariance. [58, §3] [53, §3]
Given linear operator $A : \mathbb{R}^p \to \mathbb{R}^m$ and closed convex cone $\mathcal{X} \subseteq \mathbb{R}^p$, convex cone $\mathcal{K} = A(\mathcal{X})$ is closed ($A(\mathcal{X}) = A(\mathcal{X})$) if and only if

$$\mathcal{N}(A) \cap \mathcal{X} = \{0\} \quad \text{or} \quad \mathcal{N}(A) \cap \mathcal{X} \nsubseteq \text{rel} \partial \mathcal{X}$$

(383)

Otherwise, $\mathcal{K} = A(\mathcal{X}) \supseteq A(\mathcal{X}) \supseteq A(\mathcal{X})$. [318, thm.6.6]

If matrix $A$ has no nontrivial nullspace, then $A \text{svec } X$ is an isomorphism in $X$ between cone $\mathbb{S}^n_+$ and range $\mathcal{R}(A)$ of matrix $A$; (§2.2.1.0.1, §2.10.1.1) sufficient for convex cone $\mathcal{K}$ to be closed and have relative boundary

$$\text{rel } \partial \mathcal{K} = \{ A \text{svec } X \mid X \succeq 0, \ X \not\succ 0 \}$$

(384)

Now consider the (closed convex) dual cone:

$$\mathcal{K}^* = \{ y \mid \langle z, y \rangle \geq 0 \text{ for all } z \in \mathcal{K} \} \subseteq \mathbb{R}^m$$
$$= \{ y \mid \langle z, y \rangle \geq 0 \text{ for all } z = A \text{svec } X \text{, } X \succeq 0 \}$$
$$= \{ y \mid \langle A \text{svec } X, y \rangle \geq 0 \text{ for all } X \succeq 0 \}$$
$$= \{ y \mid \langle \text{svec } X, A^T y \rangle \geq 0 \text{ for all } X \succeq 0 \}$$
$$= \{ y \mid \text{svec}^{-1}(A^T y) \succeq 0 \}$$

(385)

that follows from (377) and leads to an equally peculiar halfspace-description

$$\mathcal{K}^* = \{ y \in \mathbb{R}^m \mid \sum_{j=1}^m y_j A_j \succeq 0 \}$$

(386)

The summation inequality with respect to positive semidefinite cone $\mathbb{S}^n_+$ is known as linear matrix inequality. [61] [159] [272] [374] Although we already know that the dual cone is convex (§2.13.1), inverse image theorem 2.1.9.0.1 certifies convexity of $\mathcal{K}^*$ which is the inverse image of positive semidefinite cone $\mathbb{S}^n_+$ under linear transformation $g(y) \triangleq \sum y_j A_j$.

And although we already know that the dual cone is closed, it is certified by (381). By the inverse image closedness theorem, dual cone relative interior may always be expressed

$$\text{rel int } \mathcal{K}^* = \{ y \in \mathbb{R}^m \mid \sum_{j=1}^m y_j A_j \succ 0 \}$$

(387)

Function $g(y)$ on $\mathbb{R}^m$ is an isomorphism when the vectorized $A_j$ matrices are linearly independent; hence, uniquely invertible. Inverse image $\mathcal{K}^*$ must therefore have dimension equal to $\dim(\mathcal{R}(A^T) \cap \text{svec } \mathbb{S}^n_+)$ (49) and relative boundary

$$\text{rel } \partial \mathcal{K}^* = \{ y \in \mathbb{R}^m \mid \sum_{j=1}^m y_j A_j \succeq 0, \sum_{j=1}^m y_j A_j \not\succ 0 \}$$

(388)

When this dimension equals $m$, then dual cone $\mathcal{K}^*$ is full-dimensional

$$\text{rel int } \mathcal{K}^* = \text{int } \mathcal{K}^*$$

(14)

which implies: closure of convex cone $\mathcal{K}$ is pointed (309). □
CHAPTER 2. CONVEX GEOMETRY

2.13.6 Dual of pointed polyhedral cone

In a subspace of \( R^n \), now we consider a pointed polyhedral cone \( K \) given in terms of its extreme directions \( \Gamma_i \) arranged columnar in

\[
X = [\Gamma_1 \ \Gamma_2 \ \cdots \ \Gamma_N] \in \mathbb{R}^{n \times N}
\]

(279)

The extremes theorem (§2.8.1.1.1) provides the vertex-description of a pointed polyhedral cone in terms of its finite number of extreme directions and its lone vertex at the origin:

2.13.6.0.1 Definition. Pointed polyhedral cone, vertex-description.

Given pointed polyhedral cone \( K \) in a subspace of \( R^n \), denoting its \( i \)th extreme direction by \( \Gamma_i \in \mathbb{R}^n \) arranged in a matrix \( X \) as in (279), then that cone may be described: (86) (confer (188) (292))

\[
K = \left\{ \begin{bmatrix} 0 & X \end{bmatrix} a \zeta \mid a^T 1 = 1, \ a \geq 0, \ \zeta \geq 0 \right\}
\]

(389)

that is simply a conic hull (like (103)) of a finite number \( N \) of directions. Relative interior may always be expressed

\[
\text{rel int} \ K = \{ Xb \mid b > 0 \} \subset \mathbb{R}^n
\]

(390)

but identifying the cone’s relative boundary in this manner

\[
\text{rel } \partial K = \{ Xb \mid b \geq 0, \ b \neq 0 \}
\]

(391)

holds only when matrix \( X \) represents a bijection onto its range; in other words, some coefficients meeting lower bound zero \( (b \in \partial \mathbb{R}^N_+) \) do not necessarily provide membership to relative boundary of cone \( K \).

Whenever cone \( K \) is pointed, closed, and convex (not only polyhedral), then dual cone \( K^* \) has a halfspace-description in terms of the extreme directions \( \Gamma_i \) of \( K \):

\[
K^* = \left\{ y \mid \gamma^T y \geq 0 \ \text{for all } \gamma \in \{ \Gamma_i \ , \ i = 1 \ldots N \} \subseteq \text{rel } \partial K \right\}
\]

(392)

because when \( \{ \Gamma_i \} \) constitutes any set of generators for \( K \), the discretization result in §2.13.4.1 allows relaxation of the requirement \( \forall x \in K \) in (296) to \( \forall \gamma \in \{ \Gamma_i \} \) directly.\(^{2.69}\)

That dual cone so defined is unique, identical to (296), polyhedral whenever the number of generators \( N \) is finite

\[
K^* = \left\{ y \mid X^T y \geq 0 \right\} \subseteq \mathbb{R}^n
\]

(362)

and is full-dimensional because \( K \) is assumed pointed. But \( K^* \) is not necessarily pointed unless \( K \) is full-dimensional (§2.13.1.1).

\(^{2.69}\)The extreme directions of \( K \) constitute a minimal set of generators. Formulae and conversions to vertex-description of the dual cone are in §2.13.9 and §2.13.11.
2.13 DUAL CONE & GENERALIZED INEQUALITY

2.13.6.1 Facet normal & extreme direction

We see from (362) that the conically independent generators of cone $\mathcal{K}$ (namely, the extreme directions of pointed closed convex cone $\mathcal{K}$ constituting the $N$ columns of $X$) each define an inward-normal to a hyperplane supporting dual cone $\mathcal{K}^*$ (§2.4.2.6.1) and exposing a dual facet when $N$ is finite. Were $\mathcal{K}^*$ pointed and finitely generated, then by closure the conjugate statement would also hold; id est, the extreme directions of pointed $\mathcal{K}^*$ each define an inward-normal to a hyperplane supporting $\mathcal{K}$ and exposing a facet when $N$ is finite. Examine Figure 58 or Figure 64, for example.

We may conclude, the extreme directions of proper polyhedral $\mathcal{K}$ are respectively orthogonal to the facets of $\mathcal{K}^*$; likewise, the extreme directions of proper polyhedral $\mathcal{K}^*$ are respectively orthogonal to the facets of $\mathcal{K}$.

2.13.7 Biorthogonal expansion by example

2.13.7.0.1 Example. Relationship to dual polyhedral cone.

Simplicial cone $\mathcal{K}$ illustrated in Figure 65 induces a partial order on $\mathbb{R}^2$. All points greater than $x$ with respect to $\mathcal{K}$, for example, are contained in the translated cone $x + \mathcal{K}$. The extreme directions $\Gamma_1$ and $\Gamma_2$ of $\mathcal{K}$ do not make an orthogonal set; neither do extreme directions $\Gamma_3$ and $\Gamma_4$ of dual cone $\mathcal{K}^*$; rather, we have the biorthogonality condition

$$\Gamma_1^T \Gamma_1 = \Gamma_2^T \Gamma_2 = 0$$
$$\Gamma_3^T \Gamma_1 \neq 0, \Gamma_4^T \Gamma_2 \neq 0$$

Biorthogonal expansion of $x \in \mathcal{K}$ is then

$$x = \Gamma_1 \frac{\Gamma_1^T x}{\Gamma_3^T \Gamma_1} + \Gamma_2 \frac{\Gamma_2^T x}{\Gamma_4^T \Gamma_2}$$

where $\Gamma_1^T x/(\Gamma_3^T \Gamma_1)$ is the nonnegative coefficient of nonorthogonal projection (§E.6.1) of $x$ on $\Gamma_1$ in the direction orthogonal to $\Gamma_3$ ($y$ in Figure 175 p.617), and where $\Gamma_2^T x/(\Gamma_4^T \Gamma_2)$ is the nonnegative coefficient of nonorthogonal projection of $x$ on $\Gamma_2$ in the direction orthogonal to $\Gamma_4$ ($z$ in Figure 175); they are coordinates in this nonorthogonal system. Those coefficients must be nonnegative $x \succeq_\mathcal{K} 0$ because $x \in \mathcal{K}$ (324) and $\mathcal{K}$ is simplicial.

If we ascribe the extreme directions of $\mathcal{K}$ to the columns of a matrix

$$X \triangleq [\Gamma_1 \ \Gamma_2]$$

then we find that the pseudoinverse transpose matrix

$$X^\dagger = \begin{bmatrix} \Gamma_3 & \frac{1}{\Gamma_3^T \Gamma_1} \\ \Gamma_4 & \frac{1}{\Gamma_4^T \Gamma_2} \end{bmatrix}$$

holds the extreme directions of the dual cone. Therefore

$$x = XX^\dagger x$$
Figure 65: (confer Figure 175) Simplicial cone $\mathcal{K} \in \mathbb{R}^2$ and its dual $\mathcal{K}^*$ drawn truncated. Conically independent generators $\Gamma_1$ and $\Gamma_2$ constitute extreme directions of $\mathcal{K}$ while $\Gamma_3$ and $\Gamma_4$ constitute extreme directions of $\mathcal{K}^*$. Dotted ray-pairs bound translated cones $\mathcal{K}$. Point $x$ is comparable to point $z$ (and vice versa) but not to $y$; $z \preceq_{\mathcal{K}} x \iff z - x \in \mathcal{K} \iff z - x \succeq_{\mathcal{K}} 0$ iff there exist nonnegative coordinates for biorthogonal expansion of $z - x$. Point $y$ is not comparable to $z$ because $z$ does not belong to $y \pm \mathcal{K}$. Translating a negated cone is quite helpful for visualization: $u \preceq_{\mathcal{K}^*} w \iff u \in w - \mathcal{K} \iff u - w \preceq_{\mathcal{K}} 0$. Points need not belong to $\mathcal{K}$ to be comparable; e.g., all points less than $w$ (w.r.t $\mathcal{K}$) belong to $w - \mathcal{K}$. 
is biorthogonal expansion (394) (§E.0.1), and biorthogonality condition (393) can be expressed succinctly (§E.1.1)\textsuperscript{2.70}

\[ X^\dagger X = I \quad (403) \]

Expansion \( w = XX^\dagger w \), for any particular \( w \in \mathbb{R}^n \) more generally, is unique w.r.t \( X \) if and only if the extreme directions of \( K \) populating the columns of \( X \in \mathbb{R}^{n \times N} \) are linearly independent; \textit{id est}, iff \( X \) has no nullspace.

\[ 2.13.7.0.2 \text{ Exercise. Visual comparison of real sums.} \]

Given \( y \preceq x \) with respect to the nonnegative orthant, draw a figure showing a negated shifted orthant (like the cone in Figure 65) illustrating why \( 1^T y \leq 1^T x \) for \( y \) and \( x \) anywhere in \( \mathbb{R}^2 \). Incorporate two hyperplanes into your drawing, one containing \( y \) and another containing \( x \) with reference to Figure 28. Does this result hold in higher dimension? ▼

\[ 2.13.7.1 \text{ Pointed cones and biorthogonality} \]

Biorthogonality condition \( X^\dagger X = I \) from Example 2.13.7.0.1 means \( \Gamma_1 \) and \( \Gamma_2 \) are linearly independent generators of \( K \) (§B.1.1.1); generators because every \( x \in K \) is their conic combination. From §2.10.2 we know that means \( \Gamma_1 \) and \( \Gamma_2 \) must be extreme directions of \( K \).

A biorthogonal expansion is necessarily associated with a pointed closed convex cone; pointed, otherwise there can be no extreme directions (§2.8.1). We will address biorthogonal expansion with respect to a pointed polyhedral cone not full-dimensional in §2.13.8.

\[ 2.13.7.1.1 \text{ Example. Expansions implied by diagonalization. (confer §6.4.3.2.1)} \]

When matrix \( X \in \mathbb{R}^{M \times M} \) is diagonalizable (§A.5),

\[ X = S \Lambda S^{-1} = \begin{bmatrix} s_1 \cdots s_M \end{bmatrix} \Lambda \begin{bmatrix} w_1^T \\ \vdots \\ w_M^T \end{bmatrix} = \sum_{i=1}^{M} \lambda_i s_i w_i^T \quad (1629) \]

coordinates for biorthogonal expansion are its eigenvalues \( \lambda_i \) (contained in diagonal matrix \( \Lambda \)) when expanded in \( S \):

\[ X = SS^{-1}X = \begin{bmatrix} s_1 \cdots s_M \end{bmatrix} \begin{bmatrix} w_1^TX \\ \vdots \\ w_M^TX \end{bmatrix} = \sum_{i=1}^{M} \lambda_i s_i w_i^T \quad (397) \]

Coordinate values depend upon geometric relationship of \( X \) to its linearly independent eigenmatrices \( s_i w_i^T \). (§A.5.0.3, §B.1.1)

\[ \text{Possibly confusing is the fact that formula } XX^\dagger x \text{ is simultaneously: the orthogonal projection of } x \text{ on } \mathcal{R}(X) \text{ (2004), and a sum of nonorthogonal projections of } x \in \mathcal{R}(X) \text{ on the range of each and every column of full-rank } X \text{ skinny-or-square (§E.5.0.0.2).} \]
Eigenmatrices $s_i w_i^T$ are linearly independent dyads constituted by right and left eigenvectors of diagonalizable $X$ and are generators of some pointed polyhedral cone $K$ in a subspace of $\mathbb{R}^{M \times M}$.

When $S$ is real and $X$ belongs to that polyhedral cone $K$, for example, then coordinates of expansion (the eigenvalues $\lambda_i$) must be nonnegative.

When $X = QAQ^T$ is symmetric, coordinates for biorthogonal expansion are its eigenvalues when expanded in $Q$; id est, for $X \in \mathbb{S}^M$

$$X = QQ^T X = \sum_{i=1}^{M} q_i q_i^T X = \sum_{i=1}^{M} \lambda_i q_i q_i^T \in \mathbb{S}^M$$

becomes an orthogonal expansion with orthonormality condition $Q^T Q = I$ where $\lambda_i$ is the $i$th eigenvalue of $X$, $q_i$ is the corresponding $i$th eigenvector arranged columnar in orthogonal matrix

$$Q = [q_1 \ q_2 \ \cdots \ q_M] \in \mathbb{R}^{M \times M}$$

and where eigenmatrix $q_i q_i^T$ is an extreme direction of some pointed polyhedral cone $K \subset \mathbb{S}^M$ and an extreme direction of the positive semidefinite cone $\mathbb{S}^M_+$.

- Orthogonal expansion is a special case of biorthogonal expansion of $X \in \text{aff} \ K$ occurring when polyhedral cone $K$ is any rotation about the origin of an orthant belonging to a subspace.

Similarly, when $X = QAQ^T$ belongs to the positive semidefinite cone in the subspace of symmetric matrices, coordinates for orthogonal expansion must be its nonnegative eigenvalues (1532) when expanded in $Q$; id est, for $X \in \mathbb{S}^M_+$

$$X = QQ^T X = \sum_{i=1}^{M} q_i q_i^T X = \sum_{i=1}^{M} \lambda_i q_i q_i^T \in \mathbb{S}^M_+$$

where $\lambda_i \geq 0$ is the $i$th eigenvalue of $X$. This means matrix $X$ simultaneously belongs to the positive semidefinite cone and to the pointed polyhedral cone $K$ formed by the conic hull of its eigenmatrices.

**2.13.7.1.2 Example. Expansion respecting nonpositive orthant.**

Suppose $x \in K$ any orthant in $\mathbb{R}^n$. Then coordinates for biorthogonal expansion of $x$ must be nonnegative; in fact, absolute value of the Cartesian coordinates.

Suppose, in particular, $x$ belongs to the nonpositive orthant $K = \mathbb{R}^n_-$. Then biorhogonal expansion becomes orthogonal expansion

$$x = XX^T x = \sum_{i=1}^{n} -e_i (-e_i^T x) = \sum_{i=1}^{n} -e_i |e_i^T x| \in \mathbb{R}^n_-$$

and the coordinates of expansion are nonnegative. For this orthant $K$ we have orthonormality condition $X^T X = I$ where $X = -I$, $e_i \in \mathbb{R}^n$ is a standard basis vector, and $-e_i$ is an extreme direction (§2.8.1) of $K$.

$\text{An orthant is simplicial and selfdual.}$
2.13. DUAL CONE & GENERALIZED INEQUALITY

Of course, this expansion \( x = X X^T x \) applies more broadly to domain \( \mathbb{R}^n \), but then the coordinates each belong to all of \( \mathbb{R} \).

2.13.8 Biorthogonal expansion, derivation

Biorthogonal expansion is a means for determining coordinates in a pointed conic coordinate system characterized by a nonorthogonal basis. Study of nonorthogonal bases invokes pointed polyhedral cones and their duals; extreme directions of a cone \( K \) are assumed to constitute the basis while those of the dual cone \( K^* \) determine coordinates.

Unique biorthogonal expansion with respect to \( K \) relies upon existence of its linearly independent extreme directions: Polyhedral cone \( K \) must be pointed; then it possesses extreme directions. Those extreme directions must be linearly independent to uniquely represent any point in their span.

We consider nonempty pointed polyhedral cone \( K \) possibly not full-dimensional; \( \text{id est} \), we consider a basis spanning a subspace. Then we need only observe that section of dual cone \( K^* \) in the affine hull of \( K \) because, by expansion of \( x \), membership \( x \in \text{aff} K \) is implicit and because any breach of the ordinary dual cone into ambient space becomes irrelevant (§2.13.9.3). Biorthogonal expansion

\[
x = X X^\dagger x \in \text{aff} K = \text{aff} \text{cone}(X)
\]

is expressed in the extreme directions \( \{\Gamma_i\} \) of \( K \) arranged columnar in

\[
X = [\Gamma_1 \ \Gamma_2 \ \cdots \ \Gamma_N] \in \mathbb{R}^{n \times N}
\]

under assumption of biorthogonality

\[
X^\dagger X = I
\]

where \( \dagger \) denotes matrix pseudoinverse (§E). We therefore seek, in this section, a vertex-description for \( K^* \cap \text{aff} K \) in terms of linearly independent dual generators \( \{\Gamma_i^*\} \subset \text{aff} K \) in the same finite quantity\(^{2.72}\) as the extreme directions \( \{\Gamma_i\} \) of

\[
K = \text{cone}(X) = \{Xa \ | \ a \geq 0\} \subseteq \mathbb{R}^n
\]

We assume the quantity of extreme directions \( N \) does not exceed the dimension \( n \) of ambient vector space because, otherwise, expansion w.r.t \( K \) could not be unique; \( \text{id est} \), assume \( N \) linearly independent extreme directions hence \( N \leq n \) (X skinny\(^{2.73}\), or-square full-rank). In other words, fat full-rank matrix \( X \) is prohibited by uniqueness because of existence of an infinity of right inverses;

- polyhedral cones whose extreme directions number in excess of the ambient space dimension are precluded in biorthogonal expansion.

2.13.8.1 \( x \in K \)

Suppose \( x \) belongs to \( K \subseteq \mathbb{R}^n \). Then \( x = X a \) for some \( a \geq 0 \). Coordinate vector \( a \) is unique only when \( \{\Gamma_i\} \) is a linearly independent set.\(^{2.74}\) Vector \( a \in \mathbb{R}^N \) can take the

\(^{2.72}\)When \( K \) is contained in a proper subspace of \( \mathbb{R}^n \), the ordinary dual cone \( K^* \) will have more generators in any minimal set than \( K \) has extreme directions.

\(^{2.73}\)“Skinny” meaning thin; more rows than columns.

\(^{2.74}\)Conic independence alone (§2.10) is insufficient to guarantee uniqueness.
form \( a = Bx \) if \( \mathcal{R}(B) = \mathbb{R}^N \). Then we require \( Xa = XBx = x \) and \( Bx = Bxa = a \). The pseudoinverse \( B = X^\dagger \in \mathbb{R}^{N \times N} \) is suitable when \( X \) is skinny-or-square and full-rank. In that case rank \( X = N \), and for all \( c \geq 0 \) and \( i=1 \ldots N \)

\[
a \geq 0 \iff X^\dagger Xa \geq 0 \iff a^T X^T X^\dagger T c \geq 0 \iff \Gamma_i^T X^\dagger T c \geq 0 \quad (404)
\]

The penultimate inequality follows from the *generalized inequality and membership corollary*, while the last inequality is a consequence of that corollary’s discretization (§2.13.4.2.1). From (404) and (392) we deduce

\[
\mathcal{K}^* \cap \text{aff} \mathcal{K} = \text{cone}(X^\dagger T) = \{ X^\dagger T c \mid c \geq 0 \} \subseteq \mathbb{R}^n \quad (405)
\]

is the vertex-description for that section of \( \mathcal{K}^* \) in the affine hull of \( \mathcal{K} \) because \( \mathcal{R}(X^\dagger T) = \mathcal{R}(X) \) by definition of the pseudoinverse. From (309), we know \( \mathcal{K}^* \cap \text{aff} \mathcal{K} \) must be pointed if \( \text{rel int} \mathcal{K} \) is logically assumed nonempty with respect to \( \text{aff} \mathcal{K} \).

Conversely, suppose full-rank skinny-or-square matrix \( (N \leq n) \)

\[
X^\dagger T \triangleq [ \Gamma_1^* \quad \Gamma_2^* \quad \ldots \quad \Gamma_N^* ] \in \mathbb{R}^{N \times N} \quad (406)
\]

comprises the extreme directions \( \{ \Gamma_i^* \} \subseteq \text{aff} \mathcal{K} \) of the dual-cone intersection with the affine hull of \( \mathcal{K} \). From the *discretized membership theorem* and (313) we get a partial dual to (392); *id est*, assuming \( x \in \text{aff cone} X \)

\[
x \in \mathcal{K} \iff \gamma^T X x \geq 0 \quad \text{for all } \gamma^* \in \{ \Gamma_i^*, i=1 \ldots N \} \subseteq \partial \mathcal{K}^* \cap \text{aff} \mathcal{K} \quad (407)
\]

that leads to a partial halfspace-description,

\[
\mathcal{K} = \{ x \in \text{aff cone} X \mid X^\dagger T x \geq 0 \} \quad (409)
\]

For \( \gamma^* = X^\dagger T e_i \), any \( x = Xa \), and for all \( i \) we have \( e_i^T X^\dagger T X a = e_i^T a \geq 0 \) only when \( a \geq 0 \). Hence \( x \in \mathcal{K} \).

When \( X \) is full-rank, then unique biorthogonal expansion of \( x \in \mathcal{K} \) becomes (402)

\[
x = XX^\dagger x = \sum_{i=1}^N \Gamma_i \Gamma_i^T x \quad (410)
\]

\[\text{2.75} \]

Intuitively, any nonnegative vector \( a \) is a conic combination of the standard basis \( \{ e_i \in \mathbb{R}^N \} \); \( a \geq 0 \iff a_i \geq 0 \) for all \( i \). The last inequality in (404) is a consequence of the fact that \( x = Xa \) may be any extreme direction of \( \mathcal{K} \), in which case \( a \) is a standard basis vector; \( a = e_i \geq 0 \). Theoretically, because \( e_i \geq 0 \) defines a pointed polyhedral cone (in fact, the nonnegative orthant in \( \mathbb{R}^N \)), we can take (404) one step further by discretizing \( c \):

\[
a \geq 0 \iff \Gamma_i^T \Gamma_j^* \geq 0 \quad \text{for } i,j=1 \ldots N \iff X^\dagger T X \geq 0 \]

In words, \( X^\dagger T X \) must be a matrix whose whose entries are each nonnegative.

\[\text{2.76} \]

When closed convex cone \( \mathcal{K} \) is not full-dimensional, \( \mathcal{K}^* \) has no extreme directions.
whose coordinates $a = \Gamma_i^T x$ must be nonnegative because $K$ is assumed pointed, closed, and convex. Whenever $X$ is full-rank, so is its pseudoinverse $X^\dagger$. (§E) In the present case, the columns of $X^\dagger T$ are linearly independent and generators of the dual cone $K^\ast \cap \text{aff } K$; hence, the columns constitute its extreme directions. (§2.10.2) That section of the dual cone is itself a polyhedral cone (by (286) or the cone intersection theorem, §2.7.2.1.1) having the same number of extreme directions as $K$.

2.13.8.2 \quad x \in \text{aff } K

The extreme directions of $K$ and $K^\ast \cap \text{aff } K$ have a distinct relationship; because $X^\dagger X = I$, then for $i, j = 1 \ldots N$, $\Gamma_i^T \Gamma_i^* = 1$, while for $i \neq j$, $\Gamma_i^T \Gamma_j^* = 0$. Yet neither set of extreme directions, $\{\Gamma_i\}$ nor $\{\Gamma_i^*\}$, is necessarily orthogonal. This is a biorthogonality condition, precisely, [377, §2.2.4] [212] implying each set of extreme directions is linearly independent. (§B.1.1.1)

Biorthogonal expansion therefore applies more broadly; meaning, for any $x \in \text{aff } K$, vector $x$ can be uniquely expressed $x = Xb$ where $b \in \mathbb{R}^N$ because $\text{aff } K$ contains the origin. Thus, for any such $x \in \mathcal{R}(X)$ (confer §E.1.1), biorthogonal expansion (410) becomes $x = X X^\dagger X b = X b$.

2.13.9 \quad \text{Formulae finding dual cone}

2.13.9.1 \quad \text{Pointed } K, \text{ dual, } X \text{ skinny-or-square full-rank}

We wish to derive expressions for a convex cone and its ordinary dual under the general assumptions: pointed polyhedral $K$ denoted by its linearly independent extreme directions arranged columnar in matrix $X$ such that

$$\operatorname{rank}(X \in \mathbb{R}^{n \times N}) = N = \dim \text{aff } K \leq n \quad (411)$$

The vertex-description is given:

$$K = \{Xa \mid a \succeq 0\} \subseteq \mathbb{R}^n \quad (412)$$

from which a halfspace-description for the dual cone follows directly:

$$K^\ast = \{y \in \mathbb{R}^n \mid X^T y \succeq 0\} \quad (413)$$

By defining a matrix

$$X^\perp \triangleq \text{basis } \mathcal{N}(X^T) \quad (414)$$

(a columnar basis for the orthogonal complement of $\mathcal{R}(X)$), we can say

$$\text{aff cone } X = \text{aff } K = \{x \mid X^\dagger x = 0\} \quad (415)$$

meaning $K$ lies in a subspace, perhaps $\mathbb{R}^n$. Thus a halfspace-description

$$K = \{x \in \mathbb{R}^n \mid X^\dagger x \succeq 0, \ X^\perp x = 0\} \quad (416)$$
and a vertex-description\textsuperscript{2.77} from (313)
\[
\mathcal{K}^* = \{ [X^\top X^\perp - X^\perp]b \mid b \succeq 0 \} \subseteq \mathbb{R}^n
\] (417)

These results are summarized for a pointed polyhedral cone, having linearly independent generators, and its ordinary dual:

<table>
<thead>
<tr>
<th>Cone Table 1</th>
<th>(\mathcal{K})</th>
<th>(\mathcal{K}^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>vertex-description</td>
<td>(X)</td>
<td>(X^\top, \pm X^\perp)</td>
</tr>
<tr>
<td>halfspace-description</td>
<td>(X^\top, X^\perp)</td>
<td>(X^\top)</td>
</tr>
</tbody>
</table>

2.13.9.2 Simplicial case

When a convex cone is simplicial (§2.12.3), Cone Table 1 simplifies because then \(\text{aff } \mathcal{K} = \mathbb{R}^n\): For square \(X\) and assuming simplicial \(\mathcal{K}\) such that
\[
\text{rank}(X \in \mathbb{R}^{n \times N}) = N \triangleq \dim \text{aff } \mathcal{K} = n
\] (418)
we have

<table>
<thead>
<tr>
<th>Cone Table S</th>
<th>(\mathcal{K})</th>
<th>(\mathcal{K}^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>vertex-description</td>
<td>(X)</td>
<td>(X^\top)</td>
</tr>
<tr>
<td>halfspace-description</td>
<td>(X^\top)</td>
<td>(X^\top)</td>
</tr>
</tbody>
</table>

For example, vertex-description (417) simplifies to
\[
\mathcal{K}^* = \{X^\top b \mid b \succeq 0\} \subset \mathbb{R}^n
\] (419)

Now, because \(\dim \mathcal{R}(X) = \dim \mathcal{R}(X^\top)\), (§E) the dual cone \(\mathcal{K}^*\) is simplicial whenever \(\mathcal{K}\) is.

2.13.9.3 Cone membership relations in a subspace

It is obvious by definition (296) of ordinary dual cone \(\mathcal{K}^*\), in ambient vector space \(\mathcal{R}\), that its determination instead in subspace \(\mathcal{S} \subseteq \mathcal{R}\) is identical to its intersection with \(\mathcal{S}\); \textit{id est}, assuming closed convex cone \(\mathcal{K} \subseteq \mathcal{S}\) and \(\mathcal{K}^* \subseteq \mathcal{R}\)
\[
(\mathcal{K}^* \text{ were ambient } \mathcal{S}) \equiv (\mathcal{K}^* \text{ in ambient } \mathcal{R}) \cap \mathcal{S}
\] (420)
because
\[
\{y \in \mathcal{S} \mid \langle y, x \rangle \geq 0 \text{ for all } x \in \mathcal{K}\} = \{y \in \mathcal{R} \mid \langle y, x \rangle \geq 0 \text{ for all } x \in \mathcal{K}\} \cap \mathcal{S}
\] (421)
From this, a constrained membership relation for the ordinary dual cone \(\mathcal{K}^* \subseteq \mathcal{R}\), assuming \(x, y \in \mathcal{S}\) and closed convex cone \(\mathcal{K} \subseteq \mathcal{S}\)
\[
y \in \mathcal{K}^* \cap \mathcal{S} \iff \langle y, x \rangle \geq 0 \text{ for all } x \in \mathcal{K}
\] (422)
\textsuperscript{2.77}These descriptions are not unique. A vertex-description of the dual cone, for example, might use four conically independent generators for a plane (§2.10.0.0.1, Figure 51) when only three would suffice.
2.13. \textit{DUAL CONE & GENERALIZED INEQUALITY}

By closure in subspace $S$ we have conjugation (§2.13.1.1):

$$x \in \mathcal{K} \iff \langle y, x \rangle \geq 0 \text{ for all } y \in \mathcal{K}^* \cap S$$

(423)

This means membership determination in subspace $S$ requires knowledge of dual cone only in $S$. For sake of completeness, for proper cone $\mathcal{K}$ with respect to subspace $S$ (confer (325))

$$x \in \text{int } \mathcal{K} \iff \langle y, x \rangle > 0 \text{ for all } y \in \mathcal{K}^* \cap S, \ y \neq 0$$

(424)

$$x \in \mathcal{K}, \ x \neq 0 \iff \langle y, x \rangle > 0 \text{ for all } y \in \text{int } \mathcal{K}^* \cap S$$

(425)

(By closure, we also have the conjugate relations.) Yet when $S$ equals $\text{aff } \mathcal{K}$ for $\mathcal{K}$ a closed convex cone

$$x \in \text{rel int } \mathcal{K} \iff \langle y, x \rangle > 0 \text{ for all } y \in \mathcal{K}^* \cap \text{aff } \mathcal{K}, \ y \neq 0$$

(426)

$$x \in \mathcal{K}, \ x \neq 0 \iff \langle y, x \rangle > 0 \text{ for all } y \in \text{rel int}(\mathcal{K}^* \cap \text{aff } \mathcal{K})$$

(427)

2.13.9.4 \textbf{Subspace } $S = \text{aff } \mathcal{K}$

Assume now a subspace $S$ that is the affine hull of cone $\mathcal{K}$: Consider again a pointed polyhedral cone $\mathcal{K}$ denoted by its extreme directions arranged columnar in matrix $X$ such that

$$\text{rank}(X \in \mathbb{R}^{n \times N}) = N \equiv \dim \text{aff } \mathcal{K} \leq n$$

(411)

We want expressions for the convex cone and its dual in subspace $S = \text{aff } \mathcal{K}$:

<table>
<thead>
<tr>
<th>Cone Table A</th>
<th>$\mathcal{K}$</th>
<th>$\mathcal{K}^* \cap \text{aff } \mathcal{K}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>vertex-description</td>
<td>$X$</td>
<td>$X^\dagger$</td>
</tr>
<tr>
<td>halfspace-description</td>
<td>$X^\dagger, X^\perp$</td>
<td>$X^T, X^\perp$</td>
</tr>
</tbody>
</table>

When $\dim \text{aff } \mathcal{K} = n$, this table reduces to Cone Table S. These descriptions facilitate work in a proper subspace. The subspace of symmetric matrices $\mathbb{S}^N$, for example, often serves as ambient space.\footnote{The dual cone of positive semidefinite matrices $\mathbb{S}^N_+ = \mathbb{S}^N_+$ remains in $\mathbb{S}^N$ by convention, whereas the ordinary dual cone would venture into $\mathbb{R}^{N \times N}$.}

2.13.9.4.1 \textbf{Exercise.} \textit{Conically independent columns and rows.}

We suspect the number of conically independent columns (rows) of $X$ to be the same for $X^\dagger$, where $\dagger$ denotes matrix pseudoinverse (§E). Prove whether it holds that the columns (rows) of $X$ are c.i. $\iff$ the columns (rows) of $X^\dagger$ are c.i. \hfill ▼
Figure 66: Simplicial cones.  (a) Monotone nonnegative cone $\mathcal{K}_{M+}$ and its dual $\mathcal{K}_{M+}^*$ (drawn truncated) in $\mathbb{R}^2$.  (b) Monotone nonnegative cone and boundary of its dual (both drawn truncated) in $\mathbb{R}^3$. Extreme directions of $\mathcal{K}_{M+}^*$ are indicated.
2.13. DUAL CONE & GENERALIZED INEQUALITY

2.13.9.4.2 Example. Monotone nonnegative cone. [63, exer.2.33] [365, §2]

Simplicial cone (§2.12.3.1.1) $K_{M+}$ is the cone of all nonnegative vectors having their entries sorted in nonincreasing order:

$$K_{M+} = \{ x \mid x_1 \geq x_2 \geq \cdots \geq x_n \geq 0 \} \subseteq \mathbb{R}_+^n$$

$$= \{ x \mid (e_i - e_{i+1})^T x \geq 0, \; i = 1 \ldots n-1, \; e_n^T x \geq 0 \}$$

$$= \{ x \mid X^T x \geq 0 \}$$

a halfspace-description where $e_i$ is the $i^{\text{th}}$ standard basis vector, and where

$$X^T \triangleq [e_1 - e_2 \; e_2 - e_3 \; \cdots \; e_n] \in \mathbb{R}^{n \times n}$$

(429)

For any vectors $x$ and $y$, simple algebra demands

$$x^T y = \sum_{i=1}^n x_i y_i = (x_1 - x_2)y_1 + (x_2 - x_3)(y_1 + y_2) + (x_3 - x_4)(y_1 + y_2 + y_3) + \cdots$$

$$+ (x_{n-1} - x_n)(y_1 + \cdots + y_{n-1}) + x_n(y_1 + \cdots + y_n)$$

(430)

Because $x_i - x_{i+1} \geq 0 \forall i$ by assumption whenever $x \in K_{M+}$, we can employ dual generalized inequalities (322) with respect to the selfdual nonnegative orthant $\mathbb{R}_+^n$ to find the halfspace-description of dual monotone nonnegative cone $K_{M+}^*$. We can say $x^T y \geq 0$ for all $X^T x \geq 0$ [sic] if and only if

$$y_1 \geq 0, \; y_1 + y_2 \geq 0, \; \cdots, \; y_1 + y_2 + \cdots + y_n \geq 0$$

(431)

id est,

$$x^T y \geq 0 \; \forall X^T x \geq 0 \iff X^T y \geq 0$$

(432)

where

$$X = [e_1 \; e_1 + e_2 \; e_1 + e_2 + e_3 \; \cdots \; 1] \in \mathbb{R}^{n \times n}$$

(433)

Because $X^T x \geq 0$ connotes membership of $x$ to pointed $K_{M+}$, then by (296) the dual cone we seek comprises all $y$ for which (432) holds; thus its halfspace-description

$$K_{M+}^* = \{ y \geq 0 \} = \{ y \mid \sum_{i=1}^k y_i \geq 0, \; k = 1 \ldots n \} = \{ y \mid X^T y \geq 0 \} \subseteq \mathbb{R}^n$$

(434)

The monotone nonnegative cone and its dual are simplicial, illustrated for two Euclidean spaces in Figure 66.

From §2.13.6.1, the extreme directions of proper $K_{M+}$ are respectively orthogonal to the facets of $K_{M+}^*$. Because $K_{M+}^*$ is simplicial, the inward-normals to its facets constitute the linearly independent rows of $X^T$ by (434). Hence the vertex-description for $K_{M+}$ employs the columns of $X$ in agreement with Cone Table S because $X^T = X^{-1}$. Likewise, the extreme directions of proper $K_{M+}^*$ are respectively orthogonal to the facets of $K_{M+}$ whose inward-normals are contained in the rows of $X^T$ by (428). So the vertex-description for $K_{M+}^*$ employs the columns of $X^T$.  

\[2.79\] With $X^T$ in hand, we might concisely scribe the remaining vertex- and halfspace-descriptions from the tables for $K_{M+}$ and its dual. Instead we use dual generalized inequalities in their derivation.

\[428\]
2.13.9.4.3 Example. Monotone cone. (Figure 67, Figure 68)

Full-dimensional but not pointed, the monotone cone is polyhedral and defined by the halfspace-description

\[
\mathcal{K}_M \triangleq \{ x \in \mathbb{R}^n \mid x_1 \geq x_2 \geq \cdots \geq x_n \} = \{ x \in \mathbb{R}^n \mid X^* \mathbf{t} \geq 0 \} \tag{435}
\]

Its dual is therefore pointed but not full-dimensional;

\[
\mathcal{K}_M^* = \{ X^* b \triangleq [e_1 - e_2 \ e_2 - e_3 \ \cdots \ e_{n-1} - e_n] b \mid b \succeq 0 \} \subset \mathbb{R}^n \tag{436}
\]

the dual cone vertex-description where the columns of \( X^* \) comprise its extreme directions. Because dual monotone cone \( \mathcal{K}_M^* \) is pointed and satisfies

\[
\text{rank}(X^* \in \mathbb{R}^{n \times N}) = N \triangleq \text{dim aff} \mathcal{K}^* \leq n \tag{437}
\]

where \( N = n - 1 \), and because \( \mathcal{K}_M \) is closed and convex, we may adapt Cone Table 1 (p.166) as follows:

<table>
<thead>
<tr>
<th>Cone Table 1*</th>
<th>( \mathcal{K}^* )</th>
<th>( \mathcal{K}^{**} = \mathcal{K} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>vertex-description</td>
<td>( X^* )</td>
<td>( X^<em>, X^</em> \mathbf{t}, X^* \mathbf{t}^{\perp}, \pm X^* \mathbf{t}^{\perp} )</td>
</tr>
<tr>
<td>halfspace-description</td>
<td>( X^\dagger, X^* \mathbf{t}^{\perp} )</td>
<td>( X^* \mathbf{t}, X^* \mathbf{t}^{\perp} )</td>
</tr>
</tbody>
</table>

The vertex-description for \( \mathcal{K}_M \) is therefore

\[
\mathcal{K}_M = \{ [X^\dagger \ X^* \mathbf{t}^{\perp} \ -X^* \mathbf{t}^{\perp}] a \mid a \succeq 0 \} \subset \mathbb{R}^n \tag{438}
\]
2.13. DUAL CONE & GENERALIZED INEQUALITY

Figure 68: Two views of monotone cone $\mathcal{K}_M$ and its dual $\mathcal{K}_M^*$ (drawn truncated) in $\mathbb{R}^3$. Monotone cone is not pointed. Dual monotone cone is not full-dimensional. (Cartesian coordinate axes are drawn for reference.)
CHAPTER 2. CONVEX GEOMETRY

where \( X^* \perp = 1 \) and

\[
X^* \dagger = \frac{1}{n} \begin{bmatrix}
n - 1 & -1 & -1 & \cdots & -1 & -1 \\
n - 2 & n - 2 & -2 & \cdots & -2 & -2 \\
\vdots & n - 3 & n - 3 & \cdots & -(n - 4) & -3 \\
3 & \vdots & n - 4 & \cdots & -(n - 3) & -(n - 3) \\
2 & 2 & \cdots & \cdots & 2 & -(n - 2) & -(n - 2) \\
1 & 1 & 1 & \cdots & 1 & 1 & -(n - 1)
\end{bmatrix} \in \mathbb{R}^{n-1 \times n} \tag{439}
\]

while

\[
K_M = \{ y \in \mathbb{R}^n \mid X^* \dagger y \succeq 0, X^* \perp^T y = 0 \} \tag{440}
\]

is the dual monotone cone halfspace-description.

\[ \square \]

2.13.9.4.4 Exercise. Inside the monotone cones. Mathematically describe the respective interior of the monotone nonnegative cone and monotone cone. In three dimensions, also describe the relative interior of each face. ▼

2.13.9.5 More pointed cone descriptions with equality condition

Consider pointed polyhedral cone \( K \) having a linearly independent set of generators and whose subspace membership is explicit; \( \text{id est} \), we are given the ordinary halfspace-description

\[
K = \{ x \mid Ax \succeq 0, Cx = 0 \} \subseteq \mathbb{R}^n \tag{286a}
\]

where \( A \in \mathbb{R}^{m \times n} \) and \( C \in \mathbb{R}^{p \times n} \). This can be equivalently written in terms of nullspace of \( C \) and vector \( \xi \):

\[
K = \{ Z\xi \in \mathbb{R}^n \mid AZ\xi \succeq 0 \} \tag{441}
\]

where \( \mathcal{R}(Z \in \mathbb{R}^{n \times n - \text{rank } C}) \triangleq \mathcal{N}(C) \). Assuming (411) is satisfied

\[
\text{rank } X \triangleq \text{rank } (A^T)^\dagger \in \mathbb{R}^{n - \text{rank } C \times m} = m - \ell = \text{dim aff } K \leq n - \text{rank } C \tag{442}
\]

where \( \ell \) is the number of conically dependent rows in \( AZ \) which must be removed to make \( AZ \) before the Cone Tables become applicable. Then results collected there admit assignment \( \hat{X} \triangleq (AZ)^\dagger \in \mathbb{R}^{n - \text{rank } C \times m - \ell} \), where \( \hat{A} \in \mathbb{R}^{m - \ell \times n} \), followed with linear transformation by \( Z \). So we get the vertex-description, for full-rank \((\hat{A}Z)^\dagger\) skinny-or-square,

\[
K = \{ Z(\hat{A}Z)^\dagger b \mid b \succeq 0 \} \tag{443}
\]

From this and (362) we get a halfspace-description of the dual cone

\[
K^* = \{ y \in \mathbb{R}^n \mid (Z^T\hat{A}^T)^\dagger Z^T y \succeq 0 \} \tag{444}
\]

\[ \text{When the conically dependent rows (§2.10) are removed, the rows remaining must be linearly independent for the Cone Tables (p.19) to apply.} \]
From this and Cone Table 1 (p.166) we get a vertex-description, (1666)

\[ K^* = \{ [Z^T(\hat{A}Z)^T] C^T - C^T ]c \mid c \succeq 0 \} \] (445)

Yet because

\[ K = \{ x \mid Ax \succeq 0 \} \cap \{ x \mid Cx = 0 \} \] (446)

then, by (313), we get an equivalent vertex-description for the dual cone

\[ K^* = \{ x \mid Ax \succeq 0 \}^* + \{ x \mid Cx = 0 \}^* = \{ [A^T C^T - C^T]b \mid b \succeq 0 \} \] (447)

from which the conically dependent columns may, of course, be removed.

### 2.13.10 Dual cone-translate

(E.10.3.2.1) First-order optimality condition (351) inspires a dual-cone variant: For any set \( K \), the negative dual of its translation by any \( a \in \mathbb{R}^n \) is

\[-(K - a)^* = \{ y \in \mathbb{R}^n \mid \langle y, x - a \rangle \leq 0 \text{ for all } x \in K \} \triangleq \mathcal{K}^\perp(a)\]

\[ = \{ y \in \mathbb{R}^n \mid \langle y, x \rangle \leq 0 \text{ for all } x \in K - a \} \] (448)

a closed convex cone called normal cone to \( K \) at point \( a \). From this, a new membership relation like (319):

\[ y \in -(K - a)^* \iff \langle y, x - a \rangle \leq 0 \text{ for all } x \in K \] (449)

and by closure the conjugate, for closed convex cone \( K \),

\[ x \in K \iff \langle y, x - a \rangle \leq 0 \text{ for all } y \in -(K - a)^* \] (450)

### 2.13.10.1 first-order optimality condition - restatement

(confers §2.13.3) The general first-order necessary and sufficient condition for optimality of solution \( x^* \) to a minimization problem with real differentiable convex objective function \( f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) over convex feasible set \( C \) is [317, §3]

\[-\nabla f(x^*) \in -(C - x^*)^* , \quad x^* \in C \] (451)

id est, the negative gradient (§3.6) belongs to the normal cone to \( C \) at \( x^* \) as in Figure 69.

### 2.13.10.1.1 Example. Normal cone to orthant.

Consider proper cone \( K = \mathbb{R}^n_+ \), the selfdual nonnegative orthant in \( \mathbb{R}^n \). The normal cone to \( \mathbb{R}^n_+ \) at \( a \in K \) is (2189)

\[ K^\perp_{\mathbb{R}^n}(a \in \mathbb{R}^n_+) = -(\mathbb{R}^n_+ - a)^* = -\mathbb{R}^n_+ \cap a^\perp , \quad a \in \mathbb{R}^n_+ \] (452)

where \(-\mathbb{R}^n_+ = -K^*\) is the algebraic complement of \( \mathbb{R}^n_+ \), and \( a^\perp\) is the orthogonal complement to range of vector \( a \). This means: When point \( a \) is interior to \( \mathbb{R}^n_+ \), the normal cone is the origin. If \( n_p \) represents number of nonzero entries in vector \( a \in \partial \mathbb{R}^n_+ \), then \( \dim(-\mathbb{R}^n_+ \cap a^\perp) = n - n_p \) and there is a complementary relationship between the nonzero entries in vector \( a \) and the nonzero entries in any vector \( x \in -\mathbb{R}^n_+ \cap a^\perp \).
Figure 69: (confer Figure 80) Shown is a plausible contour plot in $\mathbb{R}^2$ of some arbitrary differentiable convex real function $f(x)$ at selected levels $\alpha$, $\beta$, and $\gamma$; id est, contours of equal level $f$ (level sets) drawn dashed in function’s domain. From results in §3.6.2 (p. 218), gradient $\nabla f(x^*)$ is normal to $\gamma$-sublevel set $L_\gamma f$ (560) by Definition E.9.1.0.1. From §2.13.10.1, function is minimized over convex set $C$ at point $x^*$ iff negative gradient $-\nabla f(x^*)$ belongs to normal cone to $C$ there. In circumstance depicted, normal cone is a ray whose direction is coincident with negative gradient. So, gradient is normal to a hyperplane supporting both $C$ and the $\gamma$-sublevel set.
2.13. DUAL CONE & GENERALIZED INEQUALITY

2.13.10.1.2 Example. Optimality conditions for conic problem.
Consider a convex optimization problem having real differentiable convex objective function $f(x) : \mathbb{R}^n \to \mathbb{R}$ defined on domain $\mathbb{R}^n$

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in \mathcal{K}
\end{align*} \quad (453)$$

Let’s first suppose that the feasible set is a pointed polyhedral cone $\mathcal{K}$ possessing a linearly independent set of generators and whose subspace membership is made explicit by full-rank matrix $C \in \mathbb{R}^{p \times n}$; id est, we are given the halfspace-description, for $A \in \mathbb{R}^{m \times n}$

$$\mathcal{K} = \{x \mid Ax \succeq 0, Cx = 0\} \subseteq \mathbb{R}^n \quad (286a)$$

(We’ll generalize to any convex cone $\mathcal{K}$ shortly.) Vertex-description of this cone, assuming $(\hat{A}Z)^\dagger$ skinny-or-square full-rank, is

$$\mathcal{K} = \{Z(\hat{A}Z)^\dagger b \mid b \succeq 0\} \quad (443)$$

where $\hat{A} \in \mathbb{R}^{m-\ell \times n}$, $\ell$ is the number of conically dependent rows in $AZ$ (§2.10) which must be removed, and $Z \in \mathbb{R}^{n \times \text{rank}_C}$ holds basis $\mathcal{N}(C)$ columnar.

From optimality condition (351),

$$\begin{align*}
\nabla f(x^\star)^T(Z(\hat{A}Z)^\dagger b - x^\star) & \geq 0 \quad \forall b \succeq 0 \quad (454) \\
-\nabla f(x^\star)^TZ(\hat{A}Z)^\dagger(b - b^\star) & \leq 0 \quad \forall b \succeq 0
\end{align*} \quad (455)$$

because

$$x^\star \triangleq Z(\hat{A}Z)^\dagger b^\star \in \mathcal{K} \quad (456)$$

From membership relation (449) and Example 2.13.10.1.1

$$\langle -(Z^T\hat{A}^T)^\dagger Z^T\nabla f(x^\star), b - b^\star \rangle \leq 0 \quad \text{for all } b \in \mathbb{R}^{m-\ell}_+$$

$$\Leftrightarrow - (Z^T\hat{A}^T)^\dagger Z^T\nabla f(x^\star) \in -\mathbb{R}^{m-\ell}_+ \cap b^\star\perp \quad (457)$$

Then equivalent necessary and sufficient conditions for optimality of conic problem (453) with feasible set $\mathcal{K}$ are: (confer (361))

$$\begin{align*}
(Z^T\hat{A}^T)^\dagger Z^T\nabla f(x^\star) & \succeq 0, \quad b^\star \succeq 0, \quad \nabla f(x^\star)^TZ(\hat{A}Z)^\dagger b^\star = 0 \quad (458)
\end{align*}$$

expressible, by (444),

$$\begin{align*}
\nabla f(x^\star) & \in \mathcal{K}^*, \quad x^\star \in \mathcal{K}, \quad \nabla f(x^\star)^Tx^\star = 0 \quad (459)
\end{align*}$$

This result (459) actually applies more generally to any convex cone $\mathcal{K}$ comprising the feasible set: Necessary and sufficient optimality conditions are in terms of objective gradient

$$- \nabla f(x^\star) \in -(\mathcal{K} - x^\star)^*, \quad x^\star \in \mathcal{K} \quad (451)$$
whose membership to normal cone, assuming only cone $\mathcal{K}$ convexity,

$$-(\mathcal{K} - x^*)^* = \mathcal{K}_{\mathcal{K}}^\perp(x^* \in \mathcal{K}) = -\mathcal{K}^* \cap x^\perp$$  \hspace{1cm} (2189)

equivalently expresses conditions (459).

When $\mathcal{K} = \mathbb{R}_+^n$, in particular, then $C = 0$, $A = Z = I \in \mathbb{S}^n$; id est,

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \succeq 0 \\
& \quad x^\perp \\
& \quad z \succeq 0, \quad \nabla f(x)^T z = 0
\end{align*}$$  \hspace{1cm} (460)

Necessary and sufficient optimality conditions become (confer [63, §4.2.3])

$$\begin{align*}
\nabla f(x^*) & \succeq 0, \quad x^* \succeq 0, \quad \nabla f(x^*)^T x^* = 0
\end{align*}$$  \hspace{1cm} (461)

equivalent to condition (329)\textsuperscript{2.81} (under nonzero gradient) for membership to the nonnegative orthant boundary $\partial \mathbb{R}_+^n$.

### 2.13.10.1.3 Example. Complementarity problem. [220]

A complementarity problem in nonlinear function $f$ is nonconvex

$$\begin{align*}
\text{find} & \quad z \in \mathcal{K} \\
\text{subject to} & \quad f(z) \in \mathcal{K}^* \\
& \quad \langle z, f(z) \rangle = 0
\end{align*}$$  \hspace{1cm} (462)

yet bears strong resemblance to (459) and to (2125) Moreau’s decomposition theorem on page 635 for projection $P$ on mutually polar cones $\mathcal{K}$ and $-\mathcal{K}^*$. Identify a sum of mutually orthogonal projections $x \triangleq z - f(z)$; in Moreau’s terms, $z = P_{\mathcal{K}}x$ and $-f(z) = P_{-\mathcal{K}^*}x$. Then $f(z) \in \mathcal{K}^*$ (§E.9.2.2 no.4) and $z$ is a solution to the complementarity problem iff it is a fixed point of

$$z = P_{\mathcal{K}}x = P_{\mathcal{K}}(z - f(z))$$  \hspace{1cm} (463)

Given that a solution exists, existence of a fixed point would be guaranteed by theory of contraction. [237, p.300] But because only nonexpansivity (Theorem E.9.3.0.1) is achievable by a projector, uniqueness cannot be assured. [213, p.155] Elegant proofs of equivalence between complementarity problem (462) and fixed point problem (463) are provided by Németh [391, Fixed point problems].

### 2.13.10.1.4 Example. Linear complementarity problem. [90] [282] [322]

Given matrix $B \in \mathbb{R}^{n \times n}$ and vector $q \in \mathbb{R}^n$, a prototypical complementarity problem on the nonnegative orthant $\mathcal{K} = \mathbb{R}_+^n$ is linear in $w = f(z)$:

$$\begin{align*}
\text{find} & \quad z \succeq 0 \\
\text{subject to} & \quad w \succeq 0 \\
& \quad w^T z = 0 \\
& \quad w = q + B z
\end{align*}$$  \hspace{1cm} (464)

\textsuperscript{2.81} and equivalent to well-known Karush-Kuhn-Tucker (KKT) optimality conditions [63, §5.5.3] because the dual variable becomes gradient $\nabla f(x)$. \hfill \square
This problem is not convex when both vectors $w$ and $z$ are variable. Notwithstanding, this linear complementarity problem can be solved by identifying $w \leftarrow \nabla f(z) = q + Bz$ then substituting that gradient into (462)

$$\begin{align*}
\text{find} & \quad z \in \mathcal{K} \\
\text{subject to} & \quad \nabla f(z) \in \mathcal{K}^* \\
& \quad \langle z, \nabla f(z) \rangle = 0
\end{align*}$$

(465)

which is simply a restatement of optimality conditions (459) for conic problem (453). Suitable $f(z)$ is the quadratic objective from convex problem

$$\begin{align*}
\text{minimize} & \quad \frac{1}{2} z^T B z + q^T z \\
\text{subject to} & \quad z \succeq 0
\end{align*}$$

(466)

which means $B \in \mathbb{S}^n_+$ should be (symmetric) positive semidefinite for solution of (464) by this method. Then (464) has solution iff (466) does. ☐

2.13.10.1.5 Exercise. **Optimality for equality constrained conic problem.**

Consider a conic optimization problem like (453) having real differentiable convex objective function $f(x) : \mathbb{R}^n \to \mathbb{R}$

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad Cx = d \\
& \quad x \in \mathcal{K}
\end{align*}$$

(467)

minimized over convex cone $\mathcal{K}$ but, this time, constrained to affine set $\mathcal{A} = \{x \mid Cx = d\}$. Show, by means of first-order optimality condition (351) or (451), that necessary and sufficient optimality conditions are: (confer (459))

$$\begin{align*}
x^* & \in \mathcal{K} \\
C x^* & = d \\

\nabla f(x^*) + C^T \nu^* & \in \mathcal{K}^* \\
\langle \nabla f(x^*) + C^T \nu^*, x^* \rangle & = 0
\end{align*}$$

(468)

where $\nu^*$ is any vector\textsuperscript{2.83} satisfying these conditions. ▽

\textsuperscript{2.82}But if one of them is fixed, then the problem becomes convex with a very simple geometric interpretation: Define the affine subset

$$\mathcal{A} \triangleq \{y \in \mathbb{R}^n \mid By = w - q\}$$

For $w^T x$ to vanish, there must be a complementary relationship between the nonzero entries of vectors $w$ and $z$; id est, $w_i z_i = 0 \forall i$. Given $w \succeq 0$, then $z$ belongs to the convex set of solutions:

$$z \in -\mathcal{K}_{\mathbb{R}^n_+}^+(w \in \mathbb{R}^n_+) \cap \mathcal{A} = \mathbb{R}^n_+ \cap w^+ \cap \mathcal{A}$$

where $\mathcal{K}_{\mathbb{R}^n_+}^+(w)$ is the normal cone to $\mathbb{R}^n_+$ at $w$ (452). If this intersection is nonempty, then the problem is solvable. \textsuperscript{2.83}an optimal dual variable, these optimality conditions are equivalent to the KKT conditions \cite[§5.5.3]{63}.
2.13.11 Proper nonsimplicial \( K \), dual, \( X \) fat full-rank

Since conically dependent columns can always be removed from \( X \) to construct \( K \) or to determine \( K^* \) \cite{384}, then assume we are given a set of \( N \) conically independent generators \( \{ E \} \) of an arbitrary proper polyhedral cone \( K \) in \( \mathbb{R}^n \) arranged columnar in \( X \in \mathbb{R}^{n \times N} \) such that \( N > n \) (fat) and rank \( X = n \). Having found formula (419) to determine the dual of a simplicial cone, the easiest way to find a vertex-description of proper dual cone \( K^* \) is to first decompose \( K \) into simplicial parts \( K_i \) so that \( K = \bigcup K_i \) \cite{2.84}. Each component simplicial cone in \( K \) corresponds to some subset of \( n \) linearly independent columns from \( X \). The key idea, here, is how the extreme directions of the simplicial parts must remain extreme directions of \( K \). Finding the dual of \( K \) amounts to finding the dual of each simplicial part:

**2.13.11.0.1 Theorem. Dual cone intersection.** \cite{340, §2.7}

Suppose proper cone \( K \subset \mathbb{R}^n \) equals the union of \( M \) simplicial cones \( K_i \) whose extreme directions all coincide with those of \( K \). Then proper dual cone \( K^* \) is the intersection of \( M \) dual simplicial cones \( K_i^* \); *id est*,

\[
K = \bigcup_{i=1}^{M} K_i \quad \Rightarrow \quad K^* = \bigcap_{i=1}^{M} K_i^*
\]

(469)

**Proof.** For \( X_i \in \mathbb{R}^{n \times n} \), a complete matrix of linearly independent extreme directions (p.125) arranged columnar, corresponding simplicial \( K_i \) (§2.12.3.1.1) has vertex-description

\[
K_i = \{ X_i c \mid c \succeq 0 \}
\]

(470)

Now suppose,

\[
K = \bigcup_{i=1}^{M} K_i = \bigcup_{i=1}^{M} \{ X_i c \mid c \succeq 0 \}
\]

(471)

The union of all \( K_i \) can be equivalently expressed

\[
K = \left\{ \begin{bmatrix} X_1 & X_2 & \cdots & X_M \end{bmatrix} \begin{bmatrix} a \\ b \\ \vdots \\ c \end{bmatrix} \mid a, b, \ldots c \succeq 0 \right\}
\]

(472)

Because extreme directions of the simplices \( K_i \) are extreme directions of \( K \) by assumption, \footnote{That proposition presupposes, of course, that we know how to perform simplicial decomposition efficiently; also called “triangulation”. \cite{314} \cite{183, §3.1} \cite{184, §3.1} Existence of multiple simplicial parts means expansion of \( x \in K \), like (410), can no longer be unique because number \( N \) of extreme directions in \( K \) exceeds dimension \( n \) of the space.}
2.13. DUAL CONE & GENERALIZED INEQUALITY

then
\[ K = \{ [X_1 \ X_2 \ \cdots \ X_M]d \mid d \geq 0 \} \]  \hspace{1cm} (473)

by the extremes theorem (§2.8.1.1). Defining \( X \triangleq [X_1 \ X_2 \ \cdots \ X_M] \) (with any redundant \textit{sic} columns optionally removed from \( X \)), then \( K^\ast \) can be expressed ((362), Cone Table S p.166)

\[ K^* = \{ y \mid X^T y \succeq 0 \} = \bigcap_{i=1}^{M} \{ y \mid X_i^T y \succeq 0 \} = \bigcap_{i=1}^{M} K_i^* \]  \hspace{1cm} (474)

To find the extreme directions of the dual cone, first we observe that some facets of each simplicial part \( K_i \) are common to facets of \( K \) by assumption, and the union of all those common facets comprises the set of all facets of \( K \) by design. For any particular proper polyhedral cone \( K \), the extreme directions of dual cone \( K^\ast \) are respectively orthogonal to the facets of \( K \). (§2.13.6.1) Then the extreme directions of the dual cone can be found among inward-normals to facets of the component simplicial cones \( K_i^\ast \); those normals are extreme directions of the dual simplicial cones \( K_i^\ast \). From the theorem and Cone Table S (p.166),

\[ K^* = \bigcap_{i=1}^{M} K_i^* = \bigcap_{i=1}^{M} \{ X_i^{\dagger T}c \mid c \geq 0 \} \]  \hspace{1cm} (475)

The set of extreme directions \( \{ \Gamma_i^\ast \} \) for proper dual cone \( K^\ast \) is therefore constituted by those conically independent generators, from the columns of all the dual simplicial matrices \( \{ X_i^{\dagger T} \} \), that do not violate discrete definition (362) of \( K^\ast \):

\[ \{ \Gamma_i^1, \Gamma_i^2 \ldots \Gamma_i^N \} = \text{c.i.} \{ X_i^{\dagger T}(\cdot, j) \mid i=1 \ldots M, \ j=1 \ldots n \mid X_i^{\dagger T}(j, \cdot) \Gamma_{\ell} \geq 0, \ \ell=1 \ldots N \} \]  \hspace{1cm} (476)

where \text{c.i.} denotes selection of only the conically independent vectors from the argument set, argument \((\cdot, j)\) denotes the \( j \)th column while \((j, \cdot)\) denotes the \( j \)th row, and \( \{ \Gamma_i^\ast \} \) constitutes the extreme directions of \( K \). Figure 52b (p.124) shows a cone and its dual found via this algorithm.

2.13.11.0.2 Example. Dual of \( K \) nonsimplicial in subspace \( \text{aff } K \).

Given conically independent generators for pointed closed convex cone \( K \) in \( \mathbb{R}^4 \) arranged columnar in

\[ X = [\Gamma_1 \ \Gamma_2 \ \Gamma_3 \ \Gamma_4] = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & -1 \end{bmatrix} \]  \hspace{1cm} (477)
having \( \dim \text{aff } \mathcal{K} = \text{rank } X = 3 \), \((281)\) then performing the most inefficient simplicial decomposition in \( \text{aff } \mathcal{K} \) we find

\[
X_1 = \begin{bmatrix}
1 & 1 & 0 \\
-1 & 0 & 1 \\
0 & -1 & 0 \\
0 & 0 & -1
\end{bmatrix}, \quad X_2 = \begin{bmatrix}
1 & 1 & 0 \\
-1 & 0 & 0 \\
0 & -1 & 1 \\
0 & 0 & -1
\end{bmatrix}
\]

\[
X_3 = \begin{bmatrix}
1 & 0 & 0 \\
-1 & 1 & 0 \\
0 & 0 & 1 \\
0 & -1 & -1
\end{bmatrix}, \quad X_4 = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
-1 & 1 & 0 \\
0 & 0 & -1
\end{bmatrix}
\]

The corresponding dual simplicial cones in \( \text{aff } \mathcal{K} \) have generators respectively columnar in

\[
4X_1^T = \begin{bmatrix}
2 & 1 & 1 \\
-2 & 1 & 1 \\
2 & -3 & 1 \\
-2 & 1 & -3
\end{bmatrix}, \quad 4X_2^T = \begin{bmatrix}
1 & 2 & 1 \\
-3 & 2 & 1 \\
1 & -2 & 1 \\
1 & -2 & -3
\end{bmatrix}
\]

\[
4X_3^T = \begin{bmatrix}
3 & 2 & -1 \\
-1 & 2 & -1 \\
-1 & -2 & 3 \\
-1 & -2 & -1
\end{bmatrix}, \quad 4X_4^T = \begin{bmatrix}
3 & -1 & 2 \\
-1 & 3 & -2 \\
-1 & -1 & 2 \\
-1 & -1 & -2
\end{bmatrix}
\]

Applying algorithm \((476)\) we get

\[
[\Gamma_1^* \quad \Gamma_2^* \quad \Gamma_3^* \quad \Gamma_4^*] = \frac{1}{4} \begin{bmatrix}
1 & 2 & 3 & 2 \\
1 & 2 & -1 & -2 \\
1 & -2 & -1 & 2 \\
-3 & -2 & -1 & -2
\end{bmatrix}
\]

whose rank is 3, and is the known result;\(^{2.85}\) a conically independent set of generators for that pointed section of the dual cone \( \mathcal{K}^* \) in \( \text{aff } \mathcal{K} \); \textit{id est}, \( \mathcal{K}^* \cap \text{aff } \mathcal{K} \).

\section*{2.13.11.0.3 Example. Dual of proper polyhedral \( K \) in \( \mathbb{R}^4 \).}

Given conically independent generators for a full-dimensional pointed closed convex cone \( \mathcal{K} \)

\[
X = [\Gamma_1 \quad \Gamma_2 \quad \Gamma_3 \quad \Gamma_4 \quad \Gamma_5] = \begin{bmatrix}
1 & 1 & 0 & 1 & 0 \\
-1 & 0 & 1 & 0 & 1 \\
0 & -1 & 0 & 1 & 0 \\
0 & 0 & -1 & -1 & 0
\end{bmatrix}
\]

\(^{2.85}\)These calculations proceed so as to be consistent with [121, §6]; as if the ambient vector space were proper subspace \( \text{aff } \mathcal{K} \) whose dimension is 3. In that ambient space, \( \mathcal{K} \) may be regarded as a proper cone. Yet that author (from the citation) erroneously states dimension of the ordinary dual cone to be 3; it is, in fact, 4.
Applying algorithm (2.86) we count $5!/(5 - 4)! 4! = 5$ component simplices. Applying algorithm (476), we find the six extreme directions of dual cone $K^*$ (with $\Gamma_2 = \Gamma_3^*$)

$$X^* = [\Gamma_1^* \Gamma_2^* \Gamma_3^* \Gamma_4^* \Gamma_5^* \Gamma_6^*] = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 1 \\ 1 & -1 & -1 & 1 & 0 & 0 \end{bmatrix} \quad (482)$$

which means, (§2.13.6.1) this proper polyhedral $K = \text{cone}(X)$ has six (three-dimensional) facets generated $\mathcal{G}$ by its \{extreme directions\}:

$$\mathcal{G} \left\{ \begin{array}{c} \mathcal{F}_1 \\ \mathcal{F}_2 \\ \mathcal{F}_3 \\ \mathcal{F}_4 \\ \mathcal{F}_5 \\ \mathcal{F}_6 \end{array} \right\} = \left\{ \begin{array}{c} \Gamma_1 \Gamma_2 \Gamma_3 \\ \Gamma_1 \Gamma_2 \Gamma_5 \\ \Gamma_1 \Gamma_3 \Gamma_4 \\ \Gamma_1 \Gamma_3 \Gamma_5 \\ \Gamma_2 \Gamma_3 \Gamma_4 \\ \Gamma_2 \Gamma_3 \Gamma_5 \end{array} \right\} \quad (483)$$

whereas dual proper polyhedral cone $K^*$ has only five:

$$\mathcal{G} \left\{ \begin{array}{c} \mathcal{F}_1^* \\ \mathcal{F}_2^* \\ \mathcal{F}_3^* \\ \mathcal{F}_4^* \\ \mathcal{F}_5^* \end{array} \right\} = \left\{ \begin{array}{c} \Gamma_1^* \Gamma_2^* \Gamma_3^* \Gamma_4^* \Gamma_5^* \Gamma_6^* \\ \Gamma_1^* \Gamma_2^* \Gamma_5^* \\ \Gamma_1^* \Gamma_3^* \Gamma_4^* \Gamma_5^* \Gamma_6^* \\ \Gamma_2^* \Gamma_3^* \Gamma_4^* \Gamma_5^* \Gamma_6^* \\ \Gamma_2^* \Gamma_3^* \Gamma_5^* \end{array} \right\} \quad (484)$$

Six two-dimensional cones, having generators respectively \{\Gamma_1^* \Gamma_3^*\} \{\Gamma_2^* \Gamma_4^*\} \{\Gamma_1^* \Gamma_5^*\} \{\Gamma_4^* \Gamma_6^*\} \{\Gamma_2^* \Gamma_5^*\} \{\Gamma_3^* \Gamma_6^*\}, are relatively interior to dual facets; so cannot be two-dimensional faces of $K^*$ (by Definition 2.6.0.0.3).

We can check this result (482) by reversing the process; we find $6!/(6 - 4)! 4! - 3 = 12$ component simplices in the dual cone. Applying algorithm (476) to those simplices returns a conically independent set of generators for $K$ equivalent to (481). \hfill \Box

2.13.11.0.4 Exercise. Reaching proper polyhedral cone interior.

Name two extreme directions $\Gamma_1$ of cone $K$ from Example 2.13.11.0.3 whose convex hull passes through that cone’s interior. Explain why. Are there two such extreme directions of dual cone $K^*$?

2.13.12 coordinates in proper nonsimplicial system

A natural question pertains to whether a theory of unique coordinates, like biorthogonal expansion w.r.t pointed closed convex $K$, is extensible to proper cones whose extreme directions number in excess of ambient spatial dimensionality.

2.86 There are no linearly dependent combinations of three or four extreme directions in the primal cone.

2.87 Three combinations of four dual extreme directions are linearly dependent; they belong to the dual facets. But there are no linearly dependent combinations of three dual extreme directions.
2.13.12.0.1 Theorem. Conic coordinates.

With respect to vector \( v \) in some finite-dimensional Euclidean space \( \mathbb{R}^n \), define a coordinate \( t^*_v \) of point \( x \) in full-dimensional pointed closed convex cone \( K \)

\[
t^*_v(x) \triangleq \sup\{ t \in \mathbb{R} \mid x - tv \in K \} \tag{485}
\]

Given points \( x \) and \( y \) in cone \( K \), if \( t^*_v(x) = t^*_v(y) \) for each and every extreme direction \( v \) of \( K \) then \( x = y \).

Conic coordinate definition (485) acquires its heritage from conditions (375) for generator membership to a smallest face. Coordinate \( t^*_v(x) = 0 \), for example, correspond to unbounded \( \mu \) in (375); indicating, extreme direction \( v \) cannot belong to the smallest face of cone \( K \) that contains \( x \).

2.13.12.0.2 Proof. Vector \( x - t^*v \) must belong to the cone boundary \( \partial K \) by definition (485). So there must exist a nonzero vector \( \lambda \) that is inward-normal to a hyperplane supporting cone \( K \) and containing \( x - t^*v \): \( \lambda \) is the inward-normal by definition (485). So there must exist a nonzero vector \( \lambda \) that is inward-normal to a hyperplane supporting cone \( K \) and containing \( x - t^*v \); \( \lambda \) is the inward-normal by definition (485)

\[
x - t^*v \in \partial K \iff \exists \lambda \neq 0 \in \mathbb{R} \wedge \langle \lambda, x - t^*v \rangle = 0, \quad \lambda \in K^*, \quad x - t^*v \in K \tag{329}
\]

where

\[
K^* = \{ w \in \mathbb{R}^n \mid \langle v, w \rangle \geq 0 \text{ for all } v \in G(K) \} \tag{368}
\]

is the full-dimensional pointed closed convex dual cone. The set \( G(K) \), of possibly infinite cardinality \( N \), comprises generators for cone \( K \); \( \lambda \) is the extreme direction which constitute a minimal generating set. If \( x - t^*v \) is nonzero, any such vector \( \lambda \) must belong to the dual cone boundary by conjugate boundary membership relation

\[
\lambda \in \partial K^* \iff \exists x - t^*v \neq 0 \in \mathbb{R} \wedge \langle \lambda, x - t^*v \rangle = 0, \quad x - t^*v \in K, \quad \lambda \in K^* \tag{330}
\]

where

\[
K = \{ z \in \mathbb{R}^n \mid \langle \lambda, z \rangle \geq 0 \text{ for all } \lambda \in G(K^*) \} \tag{367}
\]

This description of \( K \) means: cone \( K \) is an intersection of halfspaces whose inward-normals are generators of the dual cone. Each and every face of cone \( K \) (except the cone itself) belongs to a hyperplane supporting \( K \). Each and every vector \( x - t^*v \) on the cone boundary must therefore be orthogonal to an extreme direction constituting generators \( G(K^*) \) of the dual cone.

To the \( i \)th extreme direction \( v = \Gamma_i \in \mathbb{R}^n \) of cone \( K \), ascribe a coordinate \( t_i^*(x) \in \mathbb{R} \) of \( x \) from definition (485). On domain \( K \), the mapping

\[
t^*(x) = \begin{bmatrix} t_1^*(x) \\ \vdots \\ t_N^*(x) \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^N \tag{486}
\]

has no nontrivial nullspace. Because \( x - t^*v \) must belong to \( \partial K \) by definition, the mapping \( t^*(x) \) is equivalent to a convex problem (separable in index \( i \)) whose objective (by (329))
is tightly bounded below by 0:

\[ t^*(x) \equiv \arg \min_{t \in \mathbb{R}^N} \sum_{i=1}^{N} \Gamma^*_j(i) (x - t_i \Gamma_i) \]

subject to \( x - t_i \Gamma_i \in \mathcal{K}, \quad i = 1 \ldots N \) (487)

where index \( j \in I \) is dependent on \( i \) and where (by (367)) \( \lambda = \Gamma^*_j \in \mathbb{R}^n \) is an extreme direction of dual cone \( \mathcal{K}^* \) that is normal to a hyperplane supporting \( \mathcal{K} \) and containing \( x - t^*_i \Gamma_i \). Because extreme-direction cardinality \( N \) for cone \( \mathcal{K} \) is not necessarily the same as for dual cone \( \mathcal{K}^* \), index \( j \) must be judiciously selected from a set \( I \).

To prove injectivity when extreme-direction cardinality \( N > n \) exceeds spatial dimension, we need only show mapping \( t^*(x) \) to be invertible; [135, thm.9.2.3] id est, \( x \) is recoverable given \( t^*(x) \):

\[ x = \arg \min_{\tilde{x} \in \mathbb{R}^n} \sum_{i=1}^{N} \Gamma^*_j(i) (\tilde{x} - t^*_i \Gamma_i) \]

subject to \( \tilde{x} - t^*_i \Gamma_i \in \mathcal{K}, \quad i = 1 \ldots N \) (488)

The feasible set of this nonseparable convex problem is an intersection of translated full-dimensional pointed closed convex cones \( \bigcap_i \mathcal{K} + t^*_i \Gamma_i \). The objective function's linear part describes movement in normal-direction \( -\Gamma^*_j \) for each of \( N \) hyperplanes. The optimal point of hyperplane intersection is the unique solution \( x \) when \( \{ \Gamma^*_j \} \) comprises \( n \) linearly independent normals that come from the dual cone and make the objective vanish. Because the dual cone \( \mathcal{K}^* \) is full-dimensional, pointed, closed, and convex by assumption, there exist \( N \) extreme directions \( \{ \Gamma^*_j \} \) from \( \mathcal{K}^* \subset \mathbb{R}^n \) that span \( \mathbb{R}^n \). So we need simply choose \( N \) spanning dual extreme directions that make the optimal objective vanish. Because such dual extreme directions preexist by (329), \( t^*(x) \) is invertible.

Otherwise, in the case \( N \leq n \), \( t^*(x) \) holds coordinates for biorthogonal expansion. Reconstruction of \( x \) is therefore unique.

### 2.13.12.1 reconstruction from conic coordinates

The foregoing proof of the conic coordinates theorem is not constructive; it establishes existence of dual extreme directions \( \{ \Gamma^*_j \} \) that will reconstruct a point \( x \) from its coordinates \( t^*(x) \) via (488), but does not prescribe the index set \( I \). There are at least two computational methods for specifying \( \{ \Gamma^*_j(i) \} \): one is combinatorial but sure to succeed, the other is a geometric method that searches for a minimum of a nonconvex function.

We describe the latter:

Convex problem (P)

\[
(P) \quad \begin{array}{ll}
\text{maximize} & t \\
\text{subject to} & x - tv \in \mathcal{K}
\end{array} \quad \begin{array}{ll}
\text{minimize} & \lambda^T x \\
\text{subject to} & \lambda^Tv = 1 \quad (D)
\end{array} \quad (489)
\]

is equivalent to definition (485) whereas convex problem (D) is its dual;\(^{2.88}\) meaning, primal and dual optimal objectives are equal \( t^* = \lambda^* x \) assuming Slater’s condition (p.247)

\[^{2.88}\]Form a Lagrangian associated with primal problem (P):
CHAPTER 2. CONVEX GEOMETRY

is satisfied. Under this assumption of strong duality, \( \lambda^T (x - t^* v) = t^* (1 - \lambda^T v) = 0 \); which implies, the primal problem is equivalent to

\[
\begin{align*}
\text{minimize} & \quad \lambda^T (x - t^* v) \\
\text{subject to} & \quad x - t^* v \in \mathcal{K} \\
\end{align*}
\]

(490)

while the dual problem is equivalent to

\[
\begin{align*}
\text{minimize} & \quad \lambda^T (x - t^* v) \\
\text{subject to} & \quad \lambda^T v = 1 \\
& \quad \lambda \in \mathcal{K}^* \\
\end{align*}
\]

(491)

Instead given coordinates \( t^* (x) \) and a description of cone \( \mathcal{K} \), we propose inversion by alternating solution of respective primal and dual problems

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{N} \Gamma_i^T (x - t_i^* \Gamma_i) \\
\text{subject to} & \quad x - t_i^* \Gamma_i \in \mathcal{K}, \quad i = 1 \ldots N \\
\end{align*}
\]

(492)

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{N} \Gamma_i^* (x^* - t_i^* \Gamma_i) \\
\text{subject to} & \quad \Gamma_i^* T \Gamma_i = 1, \quad i = 1 \ldots N \\
& \quad \Gamma_i^* \in \mathcal{K}^*, \quad i = 1 \ldots N \\
\end{align*}
\]

(493)

where dual extreme directions \( \Gamma_i^* \) are initialized arbitrarily and ultimately ascertained by the alternation. Convex problems (492) and (493) are iterated until convergence which is guaranteed by virtue of a monotonically nonincreasing real sequence of objective values. Convergence can be fast. The mapping \( t^* (x) \) is uniquely inverted when the necessarily nonnegative objective vanishes; \id est, when \( \Gamma_i^* (x^* - t_i^* \Gamma_i) = 0 \) \( \forall i \). Here, a zero objective can occur only at the true solution \( x \). But this global optimality condition cannot be guaranteed by the alternation because the common objective function, when regarded in both primal \( x \) and dual \( \Gamma_i^* \) variables simultaneously, is generally neither quasiconvex or monotonic. (\S3.8.0.0.3)

Conversely, a nonzero objective at convergence is a certificate that inversion was not performed properly. A nonzero objective indicates that a global minimum of a multimodal objective function could not be found by this alternation. That is a flaw in this particular iterative algorithm for inversion; not in theory.2.89 A numerical remedy is to reinitialize the \( \Gamma_i^* \) to different values.

\[
\mathcal{L}(t, \lambda) = t + \lambda^T (x - t^* v) = \lambda^T x + t (1 - \lambda^T v), \quad \lambda \geq 0, \quad \lambda \in \mathcal{K}^* \\
\sup_{t} \mathcal{L}(t, \lambda) = \lambda^T x, \quad 1 - \lambda^T v = 0
\]

Dual variable (Lagrange multiplier [260, p.216]) \( \lambda \) generally has a nonnegative sense \( \geq \) for primal maximization with any cone membership constraint, whereas \( \lambda \) would have a nonpositive sense \( \leq \) were the primal instead a minimization problem with a cone membership constraint.

2.89 The Proof 2.13.12.0.2, that suitable dual extreme directions \( \{ \Gamma_i^* \} \) always exist, means that a global optimization algorithm would always find the zero objective of alternation (492) (493); hence, the unique inversion \( x \). But such an algorithm can be combinatorial.
Chapter 3

Geometry of convex functions

The link between convex sets and convex functions is via the epigraph: A function is convex if and only if its epigraph is a convex set.

— Werner Fenchel

We limit our treatment of multidimensional functions to finite-dimensional Euclidean space. Then an icon for a one-dimensional (real) convex function is bowl-shaped (Figure 79), whereas the concave icon is the inverted bowl; respectively characterized by a unique global minimum and maximum whose existence is assumed. Because of this simple relationship, usage of the term convexity is often implicitly inclusive of concavity. Despite iconic imagery, the reader is reminded that the set of all convex, concave, quasiconvex, and quasiconcave functions contains the monotonic functions [213] [225, §2.3.5].

3.1 Convex function

3.1.1 real and vector-valued function

Vector-valued function

\[ f(X) : \mathbb{R}^{p \times k} \rightarrow \mathbb{R}^M = \begin{bmatrix} f_1(X) \\ \vdots \\ f_M(X) \end{bmatrix} \] (494)

assigns each \( X \) in its domain \( \text{dom} f \) (a subset of ambient vector space \( \mathbb{R}^{p \times k} \)) to a specific element \([268, \text{p.3}]\) of its range (a subset of \( \mathbb{R}^M \)). Function \( f(X) \) is linear in \( X \) on its domain if and only if, for each and every \( Y, Z \in \text{dom} f \) and \( \alpha, \beta \in \mathbb{R} \)

\[ f(\alpha Y + \beta Z) = \alpha f(Y) + \beta f(Z) \] (495)

\footnote{Vector- or matrix-valued functions including the real functions. Appendix D, with its tables of first- and second-order gradients, is the practical adjunct to this chapter.}

Dattorro, Convex Optimization \& Euclidean Distance Geometry 2e, Meijo, v2015.04.11. 185
A vector-valued function \( f(X) : \mathbb{R}^{p \times k} \to \mathbb{R}^M \) is convex in \( X \) if and only if \( \text{dom} f \) is a convex set and, for each and every \( Y, Z \in \text{dom} f \) and \( 0 \leq \mu \leq 1 \)

\[
f(\mu Y + (1 - \mu)Z) \preceq_{\mathbb{R}^M} \mu f(Y) + (1 - \mu) f(Z)
\]

(496)

As defined, continuity is implied but not differentiability (not smoothness).\(^3\) Apparently some, but not all, nonlinear functions are convex. Reversing sense of the inequality flips this definition to concavity. Linear (and affine\(^3\)) functions attain equality in this definition. Linear functions are therefore simultaneously convex and concave.

Vector-valued functions are most often compared (as in (496)) with respect to the \( M \)-dimensional selfdual nonnegative orthant \( \mathbb{R}_+^M \), a proper cone.\(^3\) In this case, the test prescribed by (496) is simply a comparison on \( \mathbb{R} \) of each entry \( f_i \) of a vector-valued function \( f \). (§2.13.4.2.3) The vector-valued function case is therefore a straightforward generalization of conventional convexity theory for a real function. This conclusion follows from theory of dual generalized inequalities (§2.13.2.0.1) which asserts

\[
f \text{ convex w.r.t } \mathbb{R}_+^M \Leftrightarrow w^T f \text{ convex } \forall w \in G(\mathbb{R}_+^{M^*})
\]

(497)

shown by substitution of the defining inequality (496). Discretization allows relaxation (§2.13.4.2.1) of a semiinfinite number of conditions \( \{w \in \mathbb{R}_+^{M^*} \} \) to:

\[
\{w \in G(\mathbb{R}_+^{M^*}) \} \equiv \{e_i \in \mathbb{R}^M, i=1 \ldots M\}
\]

(498)

(the standard basis for \( \mathbb{R}^M \) and a minimal set of generators (§2.8.1.2) for \( \mathbb{R}_+^M \)) from which the stated conclusion follows; \textit{id est}, the test for convexity of a vector-valued function is a comparison on \( \mathbb{R} \) of each entry.

### 3.1.2 strict convexity

When \( f(X) \) instead satisfies, for each and every distinct \( Y \) and \( Z \) in \( \text{dom} f \) and all \( 0 < \mu < 1 \) on an open interval

\[
f(\mu Y + (1 - \mu)Z) \prec_{\mathbb{R}^M} \mu f(Y) + (1 - \mu) f(Z)
\]

(499)

then we shall say \( f \) is a \textit{strictly convex function}.

Similarly to (497)

\[
f \text{ strictly convex w.r.t } \mathbb{R}_+^M \Leftrightarrow w^T f \text{ strictly convex } \forall w \in G(\mathbb{R}_+^{M^*})
\]

(500)

discretization allows relaxation of the semiinfinite number of conditions \( \{w \in \mathbb{R}_+^{M^*}, w \neq 0\} \) (325) to a finite number (498). More tests for strict convexity are given in §3.6.1.0.2, §3.6.4, and §3.7.3.0.2.

\(^3\) While linear functions are not invariant to translation (offset), convex functions are.

\(^3\) Definition of convexity can be broadened to other (not necessarily proper) cones. Referred to in the literature as \( K \)-convexity, [305] \( \mathbb{R}_+^{M^*} \) (497) generalizes to \( K^* \).
3.1. CONVEX FUNCTION

Figure 70: Convex real functions here have a unique minimizer $x^*$. For $x \in \mathbb{R}$, $f_1(x) = x^2 = \|x\|_2^2$ is strictly convex whereas nondifferentiable function $f_2(x) = \sqrt{x^2} = |x| = \|x\|_2$ is convex but not strictly. Strict convexity of a real function is only a sufficient condition for minimizer uniqueness.

### 3.1.2.1 local/global minimum, uniqueness of solution

A local minimum of any convex real function is also its global minimum. In fact, any convex real function $f(X)$ has one minimum value over any convex subset of its domain, id est, given minimization of a convex real function over some convex feasible set

$$
\begin{align*}
\text{minimize} & \quad f(X) \\
\text{subject to} & \quad X \in \mathcal{C}
\end{align*}
$$

any optimal solution $X^*$ comes from a convex set of optimal solutions

$$
X^* \in \{X : f(X) = \inf_{Y \in \mathcal{C}} f(Y)\} \subseteq \mathcal{C}
$$

But a strictly convex real function has a unique minimizer $X^*$; id est, for the optimal solution set in (502) to be a single point, it is sufficient (Figure 70) that $f(X)$ be a strictly convex real function and set $\mathcal{C}$ convex. But strict convexity is not necessary for minimizer uniqueness: existence of any strictly supporting hyperplane to a convex function epigraph (p.185, §3.5) at its minimum over $\mathcal{C}$ is necessary and sufficient for uniqueness.

Quadratic real functions $x^T A x + b^T x + c$ are convex in $x$ iff $A \succeq 0$. (§3.6.4.0.1) Quadratics characterized by positive definite matrix $A \succ 0$ are strictly convex and vice versa. The vector 2-norm square $\|x\|^2$ (Euclidean norm square) and Frobenius' norm square $\|X\|_F^2$, for example, are strictly convex functions of their respective argument. (Each absolute norm is convex but not strictly.) Figure 70a illustrates a strictly convex real function.

---

3.5 It is more customary to consider only a real function for the objective of a convex optimization problem because vector- or matrix-valued functions can introduce ambiguity into the optimal objective value. (§2.7.2.2, §3.1.2.2) Study of multidimensional objective functions is called multicriteria- or multiobjective- or vector-optimization.
Figure 71: (confer Figure 42) Function range is convex for a convex problem.
(a) Point $f(X^*)$ is the unique minimum element of function range $\mathcal{R}f$.
(b) Point $f(X^*)$ is a minimal element of depicted range.
(Cartesian axes drawn for reference.)
3.1.2.2 minimum/minimal element, dual cone characterization

$f(X^*)$ is the minimum element of its range if and only if, for each and every $w \in \text{int } \mathbb{R}^M_+$, it is the unique minimizer of $w^Tf$. (Figure 71) [63, §2.6.3]

If $f(X^*)$ is a minimal element of its range, then there exists a nonzero $w \in \mathbb{R}^M_+$ such that $f(X^*)$ minimizes $w^Tf$. If $f(X^*)$ minimizes $w^Tf$ for some $w \in \text{int } \mathbb{R}^M_+$, conversely, then $f(X^*)$ is a minimal element of its range.

3.1.2.2.1 Exercise. Cone of convex functions.

Prove that relation (497) implies: The set of all convex vector-valued functions forms a convex cone in some space. Indeed, any nonnegatively weighted sum of convex functions remains convex. So trivial function $f = 0$ is convex. Relatively interior to each face of this cone are the strictly convex functions of corresponding dimension. How do convex real functions fit into this cone? Where are the affine functions? ▽

3.1.2.2.2 Example. Conic origins of Lagrangian.

The cone of convex functions, implied by membership relation (497), provides foundation for what is known as a Lagrangian function. [262, p.398] [291] Consider a conic optimization problem, for proper cone $K$ and affine subset $A$

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad g(x) \succeq_K 0 \\
& \quad h(x) \in A
\end{align*} \tag{503}$$

A Cartesian product of convex functions remains convex, so we may write

$$\begin{bmatrix} f \\ g \\ h \end{bmatrix} \text{ convex w.r.t } \begin{bmatrix} \mathbb{R}^M_+ \\ K^\circ \\ A \end{bmatrix} \iff \begin{bmatrix} w^T \\ \lambda^T \\ \nu^T \end{bmatrix} \begin{bmatrix} f \\ g \\ h \end{bmatrix} \text{ convex } \forall \begin{bmatrix} w \\ \lambda \\ \nu \end{bmatrix} \in \begin{bmatrix} \mathbb{R}^M_+ \\ K^\circ \\ A^\perp \end{bmatrix} \tag{504}$$

where $A^\perp$ is a normal cone to $A$. A normal cone to an affine subset is the orthogonal complement of its parallel subspace (§E.10.3.2.1).

Membership relation (504) holds because of equality for $h$ in convexity criterion (496) and because normal-cone membership relation (450), given point $a \in A$, becomes

$$h \in A \iff \langle \nu, h - a \rangle = 0 \quad \text{for all } \nu \in A^\perp \tag{505}$$

In other words: all affine functions are convex (with respect to any given proper cone), all convex functions are translation invariant, whereas any affine function must satisfy (505).

A real Lagrangian arises from the scalar term in (504); id est,

$$\mathcal{L} \triangleq \begin{bmatrix} w^T \\ \lambda^T \\ \nu^T \end{bmatrix} \begin{bmatrix} f \\ g \\ h \end{bmatrix} = w^Tf + \lambda^Tg + \nu^Th \tag{506}$$

}\footnote{Strict case excludes cone’s point at origin and zero weighting.}
3.2 Practical norm functions, absolute value

To some mathematicians, “all norms on \( \mathbb{R}^n \) are equivalent” [168, p.53]; meaning, ratios of different norms are bounded above and below by finite predeterminable constants. But to statisticians and engineers, all norms are certainly not created equal; as evidenced by the *compressed sensing* \((\text{sparsity})\) revolution, begun in 2004, whose focus is predominantly the argument of a 1-norm minimization.

A norm on \( \mathbb{R}^n \) is a convex function \( f: \mathbb{R}^n \to \mathbb{R} \) satisfying: for \( x, y \in \mathbb{R}^n, \alpha \in \mathbb{R} \):

1. \( f(x) \geq 0 \) \((f(x) = 0 \iff x = 0)\) (nonnegativity)
2. \( f(x + y) \leq f(x) + f(y) \) \(^{3.7}\) (triangle inequality)
3. \( f(\alpha x) = |\alpha| f(x) \) (nonnegative homogeneity)

Convexity follows by properties 2 and 3. Most useful are 1-, 2-, and \( \infty \)-norm:

\[
\|x\|_1 = \min_{t \in \mathbb{R}^n} t^T t \quad \text{subject to } -t \preceq x \preceq t \quad (507)
\]

where \( |x| = t^* \) (entrywise absolute value equals optimal \( t \)).\(^{3.8}\)

\[
\|x\|_2 = \min_{t \in \mathbb{R}} t \quad \text{subject to } \begin{bmatrix} tI & x \\ x^T & t^T \end{bmatrix} \succeq 0 \quad (508)
\]

where \( \|x\|_2 = \|x\| \triangleq \sqrt{x^T x} = t^* \).

\[
\|x\|_\infty = \min_{t \in \mathbb{R}} t \quad \text{subject to } -t \mathbf{1} \preceq x \preceq t \mathbf{1} \quad (509)
\]

where \( \max\{|x_i|, i = 1 \ldots n\} = t^* \) because \( \|x\|_\infty = \max\{|x_i|\} \leq t \iff |x| \preceq t \mathbf{1} \). Absolute value \( |x| \) inequality, in this sense, describes a norm ball.

\[
\|x\|_1 = \min_{\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^n} \mathbf{1}^T (\alpha + \beta) \quad \text{subject to } \begin{aligned}
\alpha &\geq 0 \\
\beta &\geq 0 \\
x &\preceq \alpha - \beta
\end{aligned} \quad (510)
\]

where \( |x| = \alpha^* + \beta^* \) because of complementarity \( \alpha^T \beta^* = 0 \) at optimality. \((507)\) \((509)\) \((510)\) represent linear programs, \((508)\) is a semidefinite program.

Over some convex set \( \mathcal{C} \), given vector constant \( y \) or matrix constant \( Y \)

\[
\arg \inf_{x \in \mathcal{C}} \|x - y\| = \arg \inf_{x \in \mathcal{C}} \|x - y\|^2 \quad (511)
\]

\(^{3.7}\|x + y\| \leq \|x\| + \|y\| \) for any norm, with equality iff \( x = \kappa y \) where \( \kappa \geq 0 \).

\(^{3.8}\) Vector \( \mathbf{1} \) may be replaced with any positive \([\text{sic}]\) vector to get absolute value, theoretically, although \( \mathbf{1} \) provides the 1-norm.
arg \( \inf_{X \in C} \|X - Y\| = \arg \inf_{X \in C} \|X - Y\|^2 \) (512)

are unconstrained convex problems for any convex norm and any affine transformation of variable. (But equality would not hold for, instead, a sum of norms; e.g., §5.4.2.2.4.)

Optimal solution is norm dependent: [63, p.297]

3.2. PRACTICAL NORM FUNCTIONS, ABSOLUTE VALUE

3.2.0.0.1 Example. Projecting the origin, on an affine subset, in 1-norm.

In (1978) we interpret least norm solution to linear system \( Ax = b \) as orthogonal projection of the origin \( \mathbf{0} \) on affine subset \( A = \{ x \in \mathbb{R}^n \mid Ax = b \} \) where \( A \in \mathbb{R}^{m \times n} \) is fat full-rank. Suppose, instead of the Euclidean metric, we use taxicab distance to do projection. Then the least 1-norm problem is stated, for \( b \in \mathcal{R}(A) \)

\[
\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{subject to} \quad \|x\|_1 = 1^T(\alpha + \beta) \\
\alpha, \beta \geq 0 \quad \text{subject to} \quad x = \alpha - \beta \\
x = \mathcal{A} \quad \text{subject to} \quad Ax = b
\]

These foregoing problems (507)-(516) are convex whenever set \( \mathcal{C} \) is. Their equivalence transformations make objectives smooth.
CHAPTER 3. GEOMETRY OF CONVEX FUNCTIONS

\[ \mathbf{A} = \{ \mathbf{x} \in \mathbb{R}^3 \mid A \mathbf{x} = \mathbf{b} \} \]

\[ \mathbf{B}_1 = \{ \mathbf{x} \in \mathbb{R}^3 \mid \| \mathbf{x} \|_1 \leq 1 \} \]

Figure 72: 1-norm ball \( \mathbf{B}_1 \) is convex hull of all cardinality-1 vectors of unit norm (its vertices). Ball boundary contains all points equidistant from origin in 1-norm. (Cartesian axes drawn for reference.) Plane \( \mathbf{A} \) is overhead (drawn truncated). If 1-norm ball is expanded until it kisses \( \mathbf{A} \) (intersects ball only at boundary), then distance (in 1-norm) from origin to \( \mathbf{A} \) is achieved. Euclidean ball would be spherical in this dimension. Only were \( \mathbf{A} \) parallel to two axes could there be a minimal cardinality least Euclidean norm solution. Yet 1-norm ball offers infinitely many, but not all, \( \mathbf{A} \)-orientations resulting in a minimal cardinality solution. (1-norm ball is an octahedron in this dimension while \( \infty \)-norm ball is a cube.)
3.2. PRACTICAL NORM FUNCTIONS, ABSOLUTE VALUE

\[ \begin{align*}
\text{minimize } & \|x\|_1 \\
\text{subject to } & Ax = b
\end{align*} \]  \hspace{1cm} \begin{align*}
\text{minimize } & \|x\|_1 \\
\text{subject to } & Ax = b, \quad x \geq 0
\end{align*}  

Figure 73: Exact recovery transition: Respectively signed [124, 126] or positive [131, 129, 130] solutions \( x \) to \( Ax = b \) with sparsity \( k \) below thick curve are recoverable. For Gaussian random matrix \( A \in \mathbb{R}^{m \times n} \), thick curve demarcates phase transition in ability to find sparsest solution \( x \) by linear programming. These results empirically reproduced in [39].

the distance in 1-norm from the origin to \( A \). For the example illustrated \((m=1, n=3)\), it appears that a vertex of the ball will be first to touch \( A \). 1-norm ball vertices in \( \mathbb{R}^3 \) represent nontrivial points of minimal cardinality 1, whereas edges represent cardinality 2, while relative interiors of facets represent maximal cardinality 3. By reorienting affine subset \( A \) so it were parallel to an edge or facet, it becomes evident as we expand or contract the ball that a kissing point is not necessarily unique.3.9

The 1-norm ball in \( \mathbb{R}^n \) has \( 2^n \) facets and \( 2^n \) vertices.3.10 For \( n > 0 \)

\[ x = \{ x \in \mathbb{R}^n \mid \|x\|_1 \leq 1 \} = \text{conv}\{\|x\|_1 = 1 \mid \text{card } x = 1\} = \text{conv}\{\pm e_i \in \mathbb{R}^n, \ i = 1 \ldots n\} \]

is a vertex-description of the unit 1-norm ball. Maximization of the 1-norm ball, until it kisses \( A \), is equivalent to minimization of the 1-norm ball until it no longer intersects \( A \). Then projection of the origin on affine subset \( A \) is

\[ \begin{align*}
\text{minimize } & \|x\|_1 \\
\text{subject to } & Ax = b \equiv \text{minimize } c \\
& x \in cB_1 \text{ subject to } Ax = b
\end{align*} \]

3.9 This is unlike the case for the Euclidean ball (1978) where minimum-distance projection on a convex set is unique (§E.9); all kissable faces of the Euclidean ball are single points (vertices).

3.10 The \( \infty \)-norm ball in \( \mathbb{R}^n \) has \( 2n \) facets and \( 2^n \) vertices.
Figure 74: Under 1-norm $f_2(x)$, histogram (hatched) of residual amplitudes $Ax - b$ exhibits predominant accumulation of zero-residuals. Nonnegatively constrained 1-norm $f_3(x)$ from (522) accumulates more zero-residuals than $f_2(x)$. Under norm $f_4(x)$ (not discussed), histogram would exhibit predominant accumulation of (nonzero) residuals at gradient discontinuities.

where

$$c\mathcal{B}_1 = \{ [I \in \mathbb{R}^{n \times n} -I \in \mathbb{R}^{n \times n}] a \mid a^T 1 = c, \ a \succeq 0 \}$$

which is the convex hull of 1-norm ball vertices. Then (519) is equivalent to

$$\begin{align*}
\text{minimize} & \quad c \\
\text{subject to} & \quad x = [I -I]a \\
& \quad a^T 1 = c \\
& \quad a \succeq 0 \\
& \quad Ax = b
\end{align*}$$

$$\begin{align*}
\text{minimize} & \quad \|a\|_1 \\
\text{subject to} & \quad [A -A]a = b \\
& \quad a \succeq 0
\end{align*}$$

where $x^* = [I -I]a^*$. (confer (516)) Significance of this result:

- (confer p.345) Any vector 1-norm minimization problem may have its variable replaced with a nonnegative variable of the same optimal cardinality but twice the length.

All other things being equal, nonnegative variables are easier to solve for sparse solutions. (Figure 73, Figure 74, Figure 105) The compressed sensing problem (517) becomes easier to interpret; e.g., for $A \in \mathbb{R}^{m \times n}$

$$\begin{align*}
\text{minimize} & \quad \|x\|_1 \\
\text{subject to} & \quad Ax = b \\
& \quad x \succeq 0
\end{align*}$$

$$\begin{align*}
\text{minimize} & \quad 1^T x \\
\text{subject to} & \quad Ax = b \\
& \quad x \succeq 0
\end{align*}$$

movement of a hyperplane (Figure 28, Figure 32) over a bounded polyhedron always has a vertex solution [97, p.158]. Or vector $b$ might lie on the relative boundary of a pointed polyhedral cone $K = \{ Ax \mid x \succeq 0 \}$. In the latter case, we find practical application of the smallest face $F$ containing $b$ from §2.13.4.3 to remove all columns of matrix $A$ not belonging to $F$; because those columns correspond to 0-entries in vector $x$ (and vice versa).
3.2. PRACTICAL NORM FUNCTIONS, ABSOLUTE VALUE

3.2.0.0.2 Exercise. Combinatorial optimization.
A device commonly employed to relax combinatorial problems is to arrange desirable solutions at vertices of bounded polyhedra; e.g., the permutation matrices of dimension $n$, which are factorial in number, are the extreme points of a polyhedron in the nonnegative orthant described by an intersection of $2n$ hyperplanes (§2.3.2.0.4). Minimizing a linear objective function over a bounded polyhedron is a convex problem (a linear program) that always has an optimal solution residing at a vertex.

What about minimizing other functions? Given some nonsingular matrix $A$, geometrically describe three circumstances under which there are likely to exist vertex solutions to

$$\min_{x \in \mathbb{R}^n} \|Ax\|_1 \quad \text{subject to} \quad x \in \mathcal{P}$$

optimized over some bounded polyhedron $\mathcal{P}$ .

3.2.1 $k$ smallest entries

Sum of the $k$ smallest entries of $x \in \mathbb{R}^n$ is the optimal objective value from: for $1 \leq k \leq n$

$$\sum_{i=n-k+1}^{n} \pi(x)_i = \min_{y \in \mathbb{R}^n} x^T y \quad \text{subject to} \quad 0 \preceq y \preceq 1 \quad 1^T y = k$$

which are dual linear programs, where $\pi(x)_1 = \max\{x_i, \ i=1 \ldots n\}$ where $\pi$ is a nonlinear permutation-operator sorting its vector argument into nonincreasing order. Finding $k$ smallest entries of an $n$-length vector $x$ is expressible as an infimum of $n!/ (k!(n-k)!)$ linear functions of $x$. The sum $\sum \pi(x)_i$ is therefore a concave function of $x$; in fact, monotonic (§3.6.1.0.1) in this instance.

3.2.2 $k$ largest entries

Sum of the $k$ largest entries of $x \in \mathbb{R}^n$ is the optimal objective value from: [63, exer.5.19]

$$\sum_{i=1}^{k} \pi(x)_i = \max_{y \in \mathbb{R}^n} x^T y \quad \text{subject to} \quad 0 \preceq y \preceq 1 \quad 1^T y = k$$

which are dual linear programs. Finding $k$ largest entries of an $n$-length vector $x$ is expressible as a supremum of $n!/ (k!(n-k)!)$ linear functions of $x$. (Figure 76) The summation is therefore a convex function (and monotonic in this instance, §3.6.1.0.1).

\[ \text{Hint: Suppose, for example, } \mathcal{P} \text{ belongs to an orthant and } A \text{ were orthogonal. Begin with } A=I \text{ and apply level sets of the objective, as in Figure 69 and Figure 72. Or rewrite the problem as a linear program like (513) and (515) but in a composite variable } \begin{bmatrix} x \\ t \end{bmatrix} \leftarrow y. \]
3.2.2.1 \( k \)-largest norm

Let \( \Pi x \) be a permutation of entries \( x_i \) such that their absolute value becomes arranged in nonincreasing order: \( |\Pi x|_1 \geq |\Pi x|_2 \geq \cdots \geq |\Pi x|_n \). Sum of the \( k \) largest entries of \( |x| \in \mathbb{R}^n \) is a norm, by properties of vector norm (§3.2), and is the optimal objective value of a linear program:

\[
\|x\|_n^k \triangleq \sum_{i=1}^k |\Pi x|_i = \sum_{i=1}^k \pi(|x|)_i = \min_{z \in \mathbb{R}^n, t \in \mathbb{R}} \quad k t + 1^T z
\]

subject to \( -t 1 - z \preceq x \preceq t 1 + z \)

\[
= \sup_{a \in \mathcal{I}} \left\{ a^T x \mid a_{ij} \in \{-1, 0, 1\} \right\} = \max_{y_1, y_2 \in \mathbb{R}^n} \quad (y_1 - y_2)^T x
\]

subject to \( 0 \preceq y_1 \preceq 1 \)

\[
0 \preceq y_2 \preceq 1
\]

\[
(y_1 + y_2)^T 1 = k
\]

where the norm subscript derives from a binomial coefficient \( \binom{n}{k} \), and

\[
\|x\|_{n_1} = \|x\|_1 \\
\|x\|_{n_2} = \|x\|_\infty \\
\|x\|_{n_k} = \|\pi(|x|)_{1,k}\|_1
\]

(526)

Sum of \( k \) largest absolute entries of an \( n \)-length vector \( x \) is expressible as a supremum of \( 2^k n! / (k! (n-k)! ) \) linear functions of \( x \); (Figure 76) hence, this norm is convex in \( x \).

[63, exer.6.3e]

3.2.2.1.1 Exercise. \( k \)-largest norm.

Make those \( \text{card } \mathcal{I} = 2^k n! / (k! (n-k)! ) \) linear functions explicit for \( \|x\|_2 \) and \( \|x\|_1 \) on \( \mathbb{R}^2 \) and \( \|x\|_3 \) on \( \mathbb{R}^3 \). Plot \( \|x\|_2 \) and \( \|x\|_1 \) in three dimensions. ▼

3.2.2.1.2 Exercise. Norm strict convexity.

Which vector norms \( \|x\|_n^k \), \( \|x\|_1 \), \( \|x\|_2 \), \( \|x\|_\infty \) become strictly convex when squared? Do they remain strictly convex when raised to third and higher whole powers? ▼

3.2.2.1.3 Example. Compressed sensing problem.

Conventionally posed as convex problem (517), we showed: the compressed sensing problem can always be posed equivalently with a nonnegative variable as in convex statement (522). The 1-norm predominantly appears in the literature because it is convex, it inherently minimizes cardinality under some technical conditions, [70] and because the desirable 0-norm is intractable.
3.2. PRACTICAL NORM FUNCTIONS, ABSOLUTE VALUE

Assuming a cardinality-$k$ solution exists, the compressed sensing problem may be written as a difference of two convex functions: for $A \in \mathbb{R}^{m \times n}$

$$\min_{x \in \mathbb{R}^n} \|x\|_1 - \|x\|_k \quad \text{subject to} \quad Ax = b, \quad x \geq 0$$

which is a nonconvex statement, a minimization of $n-k$ smallest entries of variable vector $x$, minimization of a concave function on $\mathbb{R}^n_+$ (§3.2.1 [318, §32]); but a statement of compressed sensing more precise than (522) because of its equivalence to 0-norm. $\|x\|_n$ is the convex $k$-largest norm of $x$ (monotonic on $\mathbb{R}^n_+$) while $\|x\|_0$ (quasiconcave on $\mathbb{R}^n_+$) expresses its cardinality. Global optimality occurs at a zero objective of minimization; id est, when the smallest $n-k$ entries of variable vector $x$ are zeroed. Under nonnegativity constraint, this compressed sensing problem (529) becomes the same as

$$\min_{z(\cdot), x \in \mathbb{R}^n} (1-z)^T x \quad \text{subject to} \quad Ax = b, \quad x \geq 0$$

where

$$1 = \nabla \|x\|_1 = \nabla 1^T x$$
$$z = \nabla \|x\|_k = \nabla z^T x \quad \text{subject to} \quad x \geq 0$$

where gradient of $k$-largest norm is an optimal solution to a convex problem:

$$\|x\|_k = \max_{y \in \mathbb{R}^n} y^T x \quad \text{subject to} \quad 0 \leq y \leq 1$$
$$\|x\|_k = \arg \max_{y \in \mathbb{R}^n} y^T x \quad \text{subject to} \quad 0 \leq y \leq 1$$

3.2.2.1.4 Exercise. $k$-largest norm gradient.
Prove (531). Is $\nabla \|x\|_k$ unique? Find $\nabla \|x\|_1$ and $\nabla \|x\|_k$ on $\mathbb{R}^n$. ▼

3.2.3 clipping

Zeroing negative vector entries under 1-norm is accomplished:

$$\|x_+\|_1 = \min_{t \in \mathbb{R}^n} 1^T t \quad \text{subject to} \quad x \leq t, \quad 0 \leq t$$

3.12 Hint: §D.2.1.
where, for \( x = [x_i, i = 1\ldots n] \in \mathbb{R}^n \)

\[
x_+ \triangleq t^* = \left[ \begin{array}{c} x_i, x_i \geq 0 \\ 0, x_i < 0 \end{array} \right], i = 1\ldots n = \frac{1}{2}(x + |x|)
\] (clipping)

\[
\minimize_{x \in \mathbb{R}^n} \|x_+\|_1 = \minimize_{x \in \mathbb{R}^n, t \in \mathbb{R}^n} 1^T t \\
\text{subject to} \ x \in C \\
\] (535)

### 3.3 Inverted functions and roots

A given function \( f \) is convex iff \( -f \) is concave. Both functions are loosely referred to as convex since \( -f \) is simply \( f \) inverted about the abscissa axis, and minimization of \( f \) is equivalent to maximization of \( -f \).

A given positive function \( f \) is convex iff \( 1/f \) is concave; \( f \) inverted about ordinate 1 is concave. Minimization of \( f \) is maximization of \( 1/f \).

We wish to implement objectives of the form \( x^{-1} \). Suppose we have a \( 2 \times 2 \) matrix

\[
T \triangleq \begin{bmatrix} x & z \\ z & y \end{bmatrix} \in \mathbb{R}^2
\] (536)

which is positive semidefinite by (1596) when

\[
T \succeq 0 \iff x > 0 \ \text{and} \ y \geq z^2
\] (537)

A polynomial constraint such as this is therefore called a conic constraint.\(^{3.13}\) This means we may formulate convex problems, having inverted variables, as semidefinite programs in Schur-form (§A.4); e.g.,

\[
\minimize_{x \in \mathbb{R}} x^{-1} = \minimize_{x, y \in \mathbb{R}} y \\
\text{subject to} \ x > 0 \\
x \in C
\] (538)

rather

\[
x > 0, \ y \geq \frac{1}{x} \iff \begin{bmatrix} x & 1 \\ 1 & y \end{bmatrix} \succeq 0
\] (539)

(inverted) For vector \( x = [x_i, i = 1\ldots n] \in \mathbb{R}^n \)

\[
\minimize_{x \in \mathbb{R}^n} \sum_{i=1}^n x_i^{-1} = \minimize_{x \in \mathbb{R}^n, y \in \mathbb{R}} y \\
\text{subject to} \ x \succ 0 \\
x \in C
\] (540)

\(^{3.13}\)In this dimension, the convex cone formed from the set of all values \{\( x, y, z \)\} satisfying constraint (537) is called a rotated quadratic or circular cone or positive semidefinite cone.
rather
\[ x > 0, \ y \geq \text{tr}(\delta(x)^{-1}) \iff \left[ \begin{array}{c} x_i \sqrt{n} \\ y \end{array} \right] \succeq 0, \ i = 1 \ldots n \] (541)

3.3.1 fractional power

[156] To implement an objective of the form \( x^\alpha \) for positive \( \alpha \), we quantize \( \alpha \) and work instead with that approximation. Choose nonnegative integer \( q \) for adequate quantization of \( \alpha \) like so:

\[ \alpha \triangleq \frac{k}{2^q}, \quad k \in \{0, 1, 2 \ldots 2^q-1\} \] (542)

Any \( k \) from that set may be written \( k = \sum_{i=1}^{q} b_i 2^{i-1} \) where \( b_i \in \{0,1\} \). Define vector \( y = [y_i, i=0 \ldots q] \in \mathbb{R}^{q+1} \) with \( y_0 = 1 \):

3.3.1.1 positive

Then we have the equivalent semidefinite program for maximizing a concave function \( x^\alpha \), for quantized \( 0 \leq \alpha < 1 \):

\[
\begin{align*}
\text{maximize} & \quad x^\alpha \\
\text{subject to} & \quad x > 0 \\
& \quad x \in C
\end{align*}
\]

\[
\equiv \quad \text{maximize} y_q \\
\text{subject to} & \quad \left[ \begin{array}{cc} y_{i-1} & y_i \\ y_i & x^{b_i} \end{array} \right] \succeq 0, \quad i = 1 \ldots q \\
& \quad x \in C
\]

where nonnegativity of \( y_q \) is enforced by maximization; id est,

\[ x > 0, \ y_q \leq x^\alpha \iff \left[ \begin{array}{cc} y_{i-1} & y_i \\ y_i & x^{b_i} \end{array} \right] \succeq 0, \quad i = 1 \ldots q \] (544)

3.3.1.1.1 Example. Square root. \( \alpha = \frac{1}{2} \). Choose \( q = 1 \) and \( k = 2^0 \).

\[
\begin{align*}
\text{maximize} & \quad \sqrt{x} \\
\text{subject to} & \quad x > 0 \\
& \quad x \in C
\end{align*}
\]

\[
\equiv \quad \text{maximize} y_1 \\
\text{subject to} & \quad \left[ \begin{array}{cc} y_0 = 1 & y_1 \\ y_1 & x \end{array} \right] \succeq 0 \\
& \quad x \in C
\]

where

\[ x > 0, \ y_1 \leq \sqrt{x} \iff \left[ \begin{array}{cc} 1 & y_1 \\ y_1 & x \end{array} \right] \succeq 0 \] (546)
3.3.1.2 negative

It is also desirable to implement an objective of the form \( x^{-\alpha} \) for positive \( \alpha \). The technique is nearly the same as before: for quantized \( 0 \leq \alpha < 1 \)

\[
\begin{align*}
\text{minimize} & \quad z \\
x \in \mathbb{R}^+ & \quad \text{subject to} \quad \begin{bmatrix} y_{i-1} & y_i \\ y_i & x^{b_i} \\ z & 1 \\ 1 & y_q \end{bmatrix} \succeq 0, \quad i = 1 \ldots q \\
x \in C
\end{align*}
\]

rather

\[
\begin{align*}
x > 0, \quad z \geq x^{-\alpha} \iff \left\{ \begin{array}{l}
y_{i-1} & y_i \\ y_i & x^{b_i} \\ z & 1 \\ 1 & y_q \end{array} \right\} \succeq 0, \quad i = 1 \ldots q \\
x \in C
\end{align*}
\]

3.3.1.3 positive inverted

Now define vector \( t = [t_i, i = 0 \ldots q] \in \mathbb{R}^{q+1} \) with \( t_0 = 1 \). To implement an objective \( x^{1/\alpha} \) for quantized \( 0 \leq \alpha < 1 \) as in (542)

\[
\begin{align*}
\text{minimize} & \quad y \\
x \in \mathbb{R}^+ & \quad \text{subject to} \quad \begin{bmatrix} t_{i-1} & t_i \\ t_i & y^{b_i} \\ z & 1 \\ 1 & y_q \end{bmatrix} \succeq 0, \quad i = 1 \ldots q \\
x = t_q > 0 & \quad x \in C
\end{align*}
\]

rather

\[
\begin{align*}
x > 0, \quad y \geq x^{1/\alpha} \iff \left\{ \begin{array}{l}
t_{i-1} & t_i \\ t_i & y^{b_i} \\ z & 1 \\ 1 & y_q \end{array} \right\} \succeq 0, \quad i = 1 \ldots q \\
x = t_q > 0
\end{align*}
\]

3.4 Affine function

A function \( f(X) \) is affine when it is continuous and has the dimensionally extensible form (confer §2.9.1.0.2)

\[
f(X) = AX + B
\]

All affine functions are simultaneously convex and concave. Both \(-AX + B\) and \(AX + B\), for example, are convex functions of \( X \). The linear functions constitute a proper subset of affine functions; e.g., when \( B = 0 \), function \( f(X) \) is linear.

Unlike other convex functions, affine function convexity holds with respect to any dimensionally compatible proper cone substituted into convexity definition (496). All affine functions satisfy a membership relation, for some normal cone, like (505). Affine
multidimensional functions are more easily recognized by existence of no multiplicati
e multivariate terms and no polynomial terms of degree higher than 1; *id est*, entries of the
function are characterized by only linear combinations of argument entries plus constants.

\[ A^T X A + B^T B \]  

is affine in \( X \), for example. Trace is an affine function; actually, a
real linear function expressible as inner product \( f(X) = \langle A, X \rangle \) with matrix \( A \) being the
Identity. The real affine function in Figure 75 illustrates hyperplanes, in its domain,
constituting contours of equal function-value (*level sets* (556)).

### 3.4.0.0.1 Example.  Engineering control.

For \( X \in S^M \) and matrices \( A, B, Q, R \) of any compatible dimensions, for example, the
expression \( XAX \) is not affine in \( X \) whereas

\[ g(X) = \begin{bmatrix} R & B^T X \\ X B & Q + A^T X + X A \end{bmatrix} \]  

(552)
is an affine multidimensional function. Such a function is typical in engineering control.

(\textit{confer} Figure 17) Any single- or many-valued inverse of an affine function is affine.

### 3.4.0.0.2 Example.  Linear objective.

Consider minimization of a real affine function \( f(z) = a^T z + b \) over the convex feasible set
\( \mathcal{C} \) in its domain \( \mathbb{R}^2 \) illustrated in Figure 75. Since scalar \( b \) is fixed, the problem posed is
the same as the convex optimization

\[
\begin{align*}
\text{minimize} & \quad a^T z \\
\text{subject to} & \quad z \in \mathcal{C}
\end{align*}
\]  

(553)
whose objective of minimization is a real linear function. Were convex set \( \mathcal{C} \) polyhedral
(§2.12), then this problem would be called a linear program. Were convex set \( \mathcal{C} \) an
intersection with a positive semidefinite cone, then this problem would be called a
semidefinite program.

There are two distinct ways to visualize this problem: one in the objective function’s
domain \( \mathbb{R}^2 \), the other including the ambient space of the objective function’s range as
in \( \begin{bmatrix} \mathbb{R}^2 \\ \mathbb{R} \end{bmatrix} \). Both visualizations are illustrated in Figure 75. Visualization in the function
domain is easier because of lower dimension and because
- level sets (556) of any affine function are affine. (§2.1.9)

In this circumstance, the level sets are parallel hyperplanes with respect to \( \mathbb{R}^2 \). One
solves optimization problem (553) graphically by finding that hyperplane intersecting
feasible set \( \mathcal{C} \) furthest right (in the direction of negative gradient \(-a\) (§3.6)).

---

3.14 The interpretation from this citation of \( \{ X \in S^M \mid g(X) \succeq 0 \} \) as “an intersection between a linear
subspace and the cone of positive semidefinite matrices” is incorrect. (See §2.9.1.0.2 for a similar example.)
The conditions they state under which strong duality holds for semidefinite programming are conservative.
(\textit{confer} §4.2.3.0.1)
Figure 75: (confer Figure 30) Three hyperplanes intersecting convex set $C \subset \mathbb{R}^2$ from Figure 30. Cartesian axes in $\mathbb{R}^3$: Plotted is affine subset $A = f(\mathbb{R}^2) \subset \mathbb{R}^2 \times \mathbb{R}$; a plane with third dimension. We say sequence of hyperplanes, w.r.t domain $\mathbb{R}^2$ of affine function $f(z) = a^Tz + b : \mathbb{R}^2 \to \mathbb{R}$, is increasing in normal direction (Figure 28) because affine function increases in direction of gradient $a$ (§3.6.0.0.3). Minimization of $a^Tz + b$ over $C$ is equivalent to minimization of $a^Tz$. 
3.4. AFFINE FUNCTION

When a differentiable convex objective function $f$ is nonlinear, the negative gradient $-\nabla f$ is a viable search direction (replacing $-a$ in (553)). (§2.13.10.1, Figure 69) \[160\] Then the nonlinear objective function can be replaced with a dynamic linear objective; linear as in (553).

**3.4.0.3 Example.** Support function. \[209, §C.2.1-§C.2.3.1\]
For arbitrary set $\mathcal{Y} \subseteq \mathbb{R}^n$, its support function $\sigma_{\mathcal{Y}}(a) : \mathbb{R}^n \to \mathbb{R}$ is defined

$$\sigma_{\mathcal{Y}}(a) \triangleq \sup_{z \in \mathcal{Y}} a^T z$$

(554)

a positively homogeneous function of direction $a$ whose range contains $\pm \infty$. [260, p.135] For each $z \in \mathcal{Y}$, $a^T z$ is a linear function of vector $a$. Because $\sigma_{\mathcal{Y}}(a)$ is a pointwise supremum of linear functions, it is convex in $a$ (Figure 76). Application of the support function is illustrated in Figure 31a for one particular normal $a$. Given nonempty closed bounded convex sets $\mathcal{Y}$ and $\mathcal{Z}$ in $\mathbb{R}^n$ and nonnegative scalars $\beta$ and $\gamma$ [403, p.234]

$$\sigma_{\beta \mathcal{Y} + \gamma \mathcal{Z}}(a) = \beta \sigma_{\mathcal{Y}}(a) + \gamma \sigma_{\mathcal{Z}}(a)$$

(555)

\[\Box\]

**3.4.0.4 Exercise.** Level sets.
Given a function $f$ and constant $\kappa$, its level sets are defined

$$\mathcal{L}_{\kappa}^f \triangleq \{ z \mid f(z) = \kappa \}$$

(556)

Give two distinct examples of convex function, that are not affine, having convex level sets. \[\nabla\]
3.4.0.5 Exercise. Epigraph intersection. (confer Figure 76) Draw three hyperplanes in $\mathbb{R}^3$ representing $\max(0, x) \triangleq \sup\{0, x_i \mid x \in \mathbb{R}^n\}$ in $\mathbb{R}^2 \times \mathbb{R}$ to see why maximum of nonnegative vector entries is a convex function of variable $x$. What is the normal to each hyperplane? (hint: p.216) Why is $\max(x)$ convex?

3.5 Epigraph, Sublevel set

It is well established that a continuous real function is convex if and only if its epigraph makes a convex set; [209] [318] [370] [403] [260] epigraph is the connection between convex sets and convex functions (p.185). Piecewise-continuous convex functions are admitted, thereby, and all invariant properties of convex sets carry over directly to convex functions. Generalization of epigraph to a vector-valued function $f(X) : \mathbb{R}^{p \times k} \rightarrow \mathbb{R}^M$ is straightforward: [305]

$$\text{epi } f \triangleq \{(X, t) \in \mathbb{R}^{p \times k} \times \mathbb{R}^M \mid X \in \text{dom } f, \ f(X) \leq t\}$$  \hspace{1cm} (557)

\text{id est,} \hspace{1cm} f \text{ convex} \iff \text{epi } f \text{ convex} \hspace{1cm} (558)

Necessity is proven: [63, exer.3.60] Given any $(X, u), (Y, v) \in \text{epi } f$, we must show for all $\mu \in [0, 1]$ that $\mu(X, u) + (1 - \mu)(Y, v) \in \text{epi } f$; \text{id est}, we must show

$$f(\mu X + (1 - \mu)Y) \leq \mu u + (1 - \mu)v$$  \hspace{1cm} (559)

Yet this holds by definition because $f(\mu X + (1 - \mu)Y) \leq \mu f(X) + (1 - \mu)f(Y)$. The converse also holds. ♦
3.5. EPIGRAPH, SUBLEVEL SET

3.5.0.0.1 Exercise. Epigraph sufficiency.
Prove that converse: Given any \((X, u), (Y, v) \in \text{epi } f\), if \(\mu(X, u) + (1-\mu)(Y, v) \in \text{epi } f\) holds for all \(\mu \in [0,1]\), then \(f\) must be convex.

Sublevel sets of a convex real function are convex. Likewise, corresponding to each and every \(\nu \in \mathbb{R}^M\)

\[
\mathcal{L}_\nu f \triangleq \{X \in \text{dom } f \mid f(X) \preceq \nu\} \subseteq \mathbb{R}^{p \times k}
\]

sublevel sets of a convex vector-valued function are convex. As for convex real functions, the converse does not hold. (Figure 77)

To prove necessity of convex sublevel sets: For any \(X, Y \in \mathcal{L}_\nu f\) we must show for each and every \(\mu \in [0,1]\) that \(\mu X + (1-\mu)Y \in \mathcal{L}_\nu f\). By definition,

\[
f(\mu X + (1-\mu)Y) \preceq \mu f(X) + (1-\mu)f(Y) \preceq \nu
\]

When an epigraph (557) is artificially bounded above, \(t \preceq \nu\), then the corresponding sublevel set can be regarded as an orthogonal projection of epigraph on the function domain.

Sense of the inequality is reversed in (557), for concave functions, and we use instead the nomenclature hypograph. Sense of the inequality in (560) is reversed, similarly, with each convex set then called superlevel set.

3.5.0.0.2 Example. Matrix pseudofractional function.
Consider a real function of two variables \(f(A, x) : S^n \times \mathbb{R}^n \to \mathbb{R} = x^T A^t x\) on \(\text{dom } f = S^n_+ \times \mathcal{R}(A)\). This function is convex simultaneously in both variables when variable matrix \(A\) belongs to the entire positive semidefinite cone \(S^n_+\) and variable vector \(x\) is confined to range \(\mathcal{R}(A)\) of matrix \(A\).

To explain this, we need only demonstrate that the function epigraph is convex. Consider Schur-form (1593) from §A.4: for \(t \in \mathbb{R}\)

\[
G(A, z, t) = \begin{bmatrix} A & z \\ z^T & t \end{bmatrix} \succeq 0
\]

\[
\Leftrightarrow
\begin{align*}
  z^T(I - AA^t) & = 0 \\
  t - z^T A^t z & \geq 0 \\
  A & \succeq 0
\end{align*}
\]

Inverse image of the positive semidefinite cone \(S^n_+\) under affine mapping \(G(A, z, t)\) is convex by Theorem 2.1.9.0.1. Of the equivalent conditions for positive semidefiniteness of
CHAPTER 3. GEOMETRY OF CONVEX FUNCTIONS

$G$, the first is an equality demanding vector $z$ belong to $\mathcal{R}(A)$. Function $f(A, z) = z^TA^\dagger z$ is convex on convex domain $S^n_+ \times \mathcal{R}(A)$ because the Cartesian product constituting its epigraph

$$\text{epi } f(A, z) = \{(A, z, t) \mid A \succeq 0, z \in \mathcal{R}(A), z^TA^\dagger z \leq t\} = G^{-1}(S^{n+1}_+)$$

(564)
is convex. □

3.5.0.0.3 Exercise. Matrix product function.

Continue Example 3.5.0.0.2 by introducing vector variable $x$ and making the substitution $z \leftarrow Ax$. Because of matrix symmetry (§E), for all $x \in \mathbb{R}^n$

$$f(A, z(x)) = z^TA^\dagger z = x^TA^\dagger Ax = x^TAx = f(A, x)$$

(565)

whose epigraph is

$$\text{epi } f(A, x) = \{(A, x, t) \mid A \succeq 0, x^TAx \leq t\}$$

(566)

Provide two simple explanations why $f(A, x) = x^TAx$ is not a function convex simultaneously in positive semidefinite matrix $A$ and vector $x$ on $\text{dom } f = S^n_+ \times \mathbb{R}^n$. ▼

3.5.0.0.4 Example. Matrix fractional function. (confer §3.7.2.0.1)

Continuing Example 3.5.0.0.2, now consider a real function of two variables on $\text{dom } f = S^n_+ \times \mathbb{R}^n$ for small positive constant $\epsilon$ (confer (1963))

$$f(A, x) = \epsilon x^T(A + \epsilon I)^{-1}x$$

(567)

where the inverse always exists by (1533). This function is convex simultaneously in both variables over the entire positive semidefinite cone $S^n_+$ and all $x \in \mathbb{R}^n$: Consider Schur-form (1596) from §A.4: for $t \in \mathbb{R}$

$$G(A, z, t) = \begin{bmatrix} A + \epsilon I & z \\ z^T & \epsilon^{-1}t \end{bmatrix} \succeq 0$$

$$t - \epsilon z^T(A + \epsilon I)^{-1}z \geq 0$$

$$A + \epsilon I \succ 0$$

(568)

Inverse image of the positive semidefinite cone $S^{n+1}_+$ under affine mapping $G(A, z, t)$ is convex by Theorem 2.1.9.0.1. Function $f(A, z)$ is convex on $S^n_+ \times \mathbb{R}^n$ because its epigraph is that inverse image:

$$\text{epi } f(A, z) = \{(A, z, t) \mid A + \epsilon I \succ 0, \epsilon z^T(A + \epsilon I)^{-1}z \leq t\} = G^{-1}(S^{n+1}_+)$$

(569)

□
3.5.1 matrix fractional projector

Consider nonlinear function $f$ having orthogonal projector $W$ as argument:

$$f(W, x) = \epsilon x^T(W + \epsilon I)^{-1}x$$  \hfill (570)

Projection matrix $W$ has property $W^T = W = W \succeq 0$. Any orthogonal projector can be decomposed into an outer product of orthonormal matrices $W = UU^T$ where $U^TU = I$ as explained in §E.3.2. From (1963) for any $\epsilon > 0$ and idempotent symmetric $W$, $\epsilon(W + \epsilon I)^{-1} = I - (1 + \epsilon)^{-1}W$ from which

$$f(W, x) = \epsilon x^T(W + \epsilon I)^{-1}x = x^T(I - (1 + \epsilon)^{-1}W)x$$  \hfill (571)

Therefore

$$\lim_{\epsilon \to 0^+} f(W, x) = \lim_{\epsilon \to 0^+} \epsilon x^T(W + \epsilon I)^{-1}x = x^T(I - W)x$$  \hfill (572)

where $I - W$ is also an orthogonal projector (§E.2).

We learned from Example 3.5.0.0.4 that $f(W, x) = \epsilon x^T(W + \epsilon I)^{-1}x$ is convex simultaneously in both variables over all $x \in \mathbb{R}^n$ when $W \in \mathbb{S}_+^n$ is confined to the entire positive semidefinite cone (including its boundary). It is now our goal to incorporate $f$ into an optimization problem such that an optimal solution returned always comprises a projection matrix $W$. The set of orthogonal projection matrices is a nonconvex subset of the positive semidefinite cone. So $f$ cannot be convex on the projection matrices, and its equivalent (for idempotent $W$)

$$f(W, x) = x^T(I - (1 + \epsilon)^{-1}W)x$$  \hfill (573)

cannot be convex simultaneously in both variables on either the positive semidefinite or symmetric projection matrices.

Suppose we allow dom$f$ to constitute the entire positive semidefinite cone but constrain $W$ to a Fantope (90); e.g., for convex set $C$ and $0 < k < n$ as in

$$\begin{aligned}
\text{minimize} & \quad \epsilon x^T(W + \epsilon I)^{-1}x \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \tr W = k \\
& \quad x \in C
\end{aligned}$$  \hfill (574)

Although this is a convex problem, there is no guarantee that optimal $W$ is a projection matrix because only extreme points of a Fantope are orthogonal projection matrices $UU^T$.

Let’s try partitioning the problem into two convex parts (one for $x$ and one for $W$), substitute equivalence (571), and then iterate solution of convex problem

$$\begin{aligned}
\text{minimize} & \quad x^T(I - (1 + \epsilon)^{-1}W)x \\
\text{subject to} & \quad x \in C
\end{aligned}$$  \hfill (575)

with convex problem

$$\begin{aligned}
\text{minimize} & \quad x^*^T(I - (1 + \epsilon)^{-1}W)x^* \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \tr W = k
\end{aligned} \quad \equiv \quad \begin{aligned}
\text{maximize} & \quad x^*^TWx^* \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \tr W = k
\end{aligned}$$  \hfill (576)
until convergence, where \( x^* \) represents an optimal solution of (575) from any particular iteration. The idea is to optimally solve for the partitioned variables which are later combined to solve the original problem (574). What makes this approach sound is that the constraints are separable, the partitioned feasible sets are not interdependent, and the fact that the original problem (though nonlinear) is convex simultaneously in both variables.\(^{3.15}\)

But partitioning alone does not guarantee a projector as solution. To make orthogonal projector \( W \) a certainty, we must invoke a known analytical optimal solution to problem (576): Diagonalize optimal solution from problem (575) \( x^*x^*^T \triangleq QAQ^T \) (§A.5.1) and set \( U^* = Q(:,1:k) \in \mathbb{R}^{n \times k} \) per (1792c);

\[
W = U^*U^*^T = \frac{x^*x^*^T}{\|x^*\|^2} + Q(:,2:k)Q(:,2:k)^T \tag{577}
\]

Then optimal solution \((x^*, U^*)\) to problem (574) is found, for small \( \epsilon \), by iterating solution to problem (575) with optimal (projector) solution (577) to convex problem (576).

**Proof.** Optimal vector \( x^* \) is orthogonal to the last \( n-1 \) columns of orthogonal matrix \( Q \), so

\[
f_{\text{opt}}(575) = \|x^*\|^2(1 - (1 + \epsilon)^{-1}) \tag{578}
\]

after each iteration. Convergence of \( f_{\text{opt}}(575) \) is proven with the observation that iteration (575) (576a) is a nonincreasing sequence that is bounded below by 0. Any bounded monotonic sequence in \( \mathbb{R} \) is convergent. \([268, \S 1.2] [43, \S 1.1] \) Expression (577) for optimal projector \( W \) holds at each iteration, therefore \( \|x^*\|^2(1 - (1 + \epsilon)^{-1}) \) must also represent the optimal objective value \( f_{\text{opt}}(575) \) at convergence.

Because the objective \( f_{\text{opt}}(574) \) from problem (574) is also bounded below by 0 on the same domain, this convergent optimal objective value \( f_{\text{opt}}(575) \) (for positive \( \epsilon \) arbitrarily close to 0) is necessarily optimal for (574); id est,

\[
f_{\text{opt}}(575) \geq f_{\text{opt}}(574) \geq 0 \tag{579}
\]

by (1771), and

\[
\lim_{\epsilon \to 0^+} f_{\text{opt}}(575) = 0 \tag{580}
\]

Since optimal \((x^*, U^*)\) from problem (575) is feasible to problem (574), and because their objectives are equivalent for projectors by (571), then converged \((x^*, U^*)\) must also be optimal to (574) in the limit. Because problem (574) is convex, this represents a globally optimal solution.

\(^{3.15}\)A convex problem has convex feasible set, and the objective surface has one and only one global minimum. \([313, \text{p.123}]\)
3.5.2 semidefinite program via Schur

Schur complement (1593) can be used to convert a projection problem to an optimization problem in epigraph form. Suppose, for example, we are presented with the constrained projection problem studied by Hayden & Wells in [194] (who provide analytical solution):

Given $A \in \mathbb{R}^{M \times M}$ and some full-rank matrix $S \in \mathbb{R}^{M \times L}$ with $L < M$,

\[
\begin{align*}
\text{minimize} & \quad \|A - X\|_F^2 \\
\text{subject to} & \quad S^T X S \succeq 0
\end{align*}
\] (581)

Variable $X$ is constrained to be positive semidefinite, but only on a subspace determined by $S$. First we write the epigraph form:

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \|A - X\|_F^2 \leq t \\
& \quad S^T X S \succeq 0
\end{align*}
\] (582)

Next we use Schur complement [288, §6.4.3] [258] and matrix vectorization (§2.2):

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \begin{bmatrix} t I & \text{vec}(A - X) \\ \text{vec}(A - X)^T & 1 \end{bmatrix} \succeq 0 \\
& \quad S^T X S \succeq 0
\end{align*}
\] (583)

This semidefinite program (§4) is an epigraph form in disguise, equivalent to (581); it demonstrates how a quadratic objective or constraint can be converted to a semidefinite constraint.

Were problem (581) instead equivalently expressed without the square

\[
\begin{align*}
\text{minimize} & \quad \|A - X\|_F \\
\text{subject to} & \quad S^T X S \succeq 0
\end{align*}
\] (584)

then we get a subtle variation:

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \|A - X\|_F \leq t \\
& \quad S^T X S \succeq 0
\end{align*}
\] (585)

that leads to an equivalent for (584) (and for (581) by (512))

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \begin{bmatrix} t I & \text{vec}(A - X) \\ \text{vec}(A - X)^T & t \end{bmatrix} \succeq 0 \\
& \quad S^T X S \succeq 0
\end{align*}
\] (586)
3.5.2.0.1 Example. Schur anomaly.
Consider a problem abstract in the convex constraint, given symmetric matrix \( A \):

\[
\begin{align*}
\text{minimize} & \quad \| X \|^2_F - \| A - X \|^2_F \\
\text{subject to} & \quad X \in C
\end{align*}
\tag{587}
\]

the minimization of a difference of two quadratic functions each convex in matrix \( X \).
Observe equality

\[
\| X \|^2_F - \| A - X \|^2_F = 2 \text{tr}(XA) - \| A \|^2_F
\tag{588}
\]

So problem (587) is equivalent to the convex optimization

\[
\begin{align*}
\text{minimize} & \quad \text{tr}(XA) \\
\text{subject to} & \quad X \in C
\end{align*}
\tag{589}
\]

But this problem (587) does not have Schur-form

\[
\begin{align*}
\text{minimize} & \quad t - \alpha \\
\text{subject to} & \quad X \in C \\
& \quad \| X \|^2_F \leq t \\
& \quad \| A - X \|^2_F \geq \alpha
\end{align*}
\tag{590}
\]

because the constraint in \( \alpha \) is nonconvex. (§2.9.1.0.1)

Matrix 2-norm (spectral norm) coincides with largest singular value.

\[
\| X \|_2 = \sup_{\| a \|=1} \| Xa \|_2 = \sigma(X)_1 = \sqrt{\lambda(X^TX)_1} = \min_{t \in \mathbb{R}} t
\tag{591}
\]

subject to \[
\begin{bmatrix}
tI & X \\
X^T & tI
\end{bmatrix} \succeq 0
\]

This supremum of a family of convex functions in \( X \) must be convex because it constitutes an intersection of epigraphs of convex functions.

3.6 Gradient

Gradient \( \nabla f \) of any differentiable multidimensional function \( f \) (formally defined in §D.1) maps each entry \( f_i \) to a space having the same dimension as the ambient space of its domain. Notation \( \nabla f \) is shorthand for gradient \( \nabla_x f(x) \) of \( f \) with respect to \( x \). \( \nabla f(y) \) can mean \( \nabla_y f(y) \) or gradient \( \nabla_x f(y) \) of \( f(x) \) with respect to \( x \) evaluated at \( y \); a distinction that should become clear from context.

Gradient of a differentiable real function \( f(x) : \mathbb{R}^K \rightarrow \mathbb{R} \), with respect to its vector argument, is uniquely defined

\[
\nabla f(x) = \left[ \begin{array}{c}
\frac{\partial f(x)}{\partial x_1} \\
\frac{\partial f(x)}{\partial x_2} \\
\vdots \\
\frac{\partial f(x)}{\partial x_K}
\end{array} \right] \in \mathbb{R}^K \tag{1852}
\]
3.6. GRADIENT

Figure 78: Gradient in $\mathbb{R}^2$ evaluated on grid over some open disc in domain of convex quadratic bowl $f(Y) = Y^TY : \mathbb{R}^2 \to \mathbb{R}$ illustrated in Figure 79. Circular contours are level sets; each defined by a constant function-value.

while the second-order gradient of the twice differentiable real function, with respect to its vector argument, is traditionally called the Hessian.\textsuperscript{3.16}

$$\nabla^2 f(x) = \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_K} \\
\frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_K} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f(x)}{\partial x_K \partial x_1} & \frac{\partial^2 f(x)}{\partial x_K \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_K^2}
\end{bmatrix} \in \mathbb{S}^K \quad (1853)$$

The gradient can be interpreted as a vector pointing in the direction of greatest change, \[229, \S15.6\] or polar to direction of steepest descent.\textsuperscript{3.17} [407] The gradient can also be interpreted as that vector normal to a sublevel set; e.g., Figure 80, Figure 69.

For the quadratic bowl in Figure 79, the gradient maps to $\mathbb{R}^2$; illustrated in Figure 78. For a one-dimensional function of real variable, for example, the gradient evaluated at any point in the function domain is just the slope (or derivative) of that function there. (confer §D.1.4.1)

- For any differentiable multidimensional function, zero gradient $\nabla f = 0$ is a condition necessary for its unconstrained minimization [160, §3.2]:

\textsuperscript{3.16} Jacobian is the Hessian transpose, so commonly confused in matrix calculus.

\textsuperscript{3.17} Newton’s direction $-\nabla^2 f(x)^{-1} \nabla f(x)$ is better for optimization of nonlinear functions well approximated locally by a quadratic function. [160, p.105]
3.6.0.0.1 Example. Projection on rank-1 subset.
For $A \in S^N$ having eigenvalues $\lambda(A) = [\lambda_i] \in \mathbb{R}^N$, consider the unconstrained nonconvex optimization that is a projection on the rank-1 subset (§2.9.2.1) of positive semidefinite cone $S^+_N$: Defining $\lambda_1 \triangleq \max_i \{\lambda(A)_i\}$ and corresponding eigenvector $v_1$

$$\begin{align*}
\text{minimize } & \left\| xx^T - A \right\|_F^2 \\
& \text{subject to } x \in \mathbb{R}^N
\end{align*}
$$

$$= \begin{cases} 
\|\lambda(A)\|_2^2, & \lambda_1 \leq 0 \\
\|\lambda(A)\|_2^2 - \lambda_1^2, & \lambda_1 > 0 
\end{cases} \quad (1787)
$$

$$\begin{align*}
\text{arg minimize } & \left\| xx^T - A \right\|_F^2 \\
& \text{subject to } x \in \mathbb{R}^N
\end{align*}
$$

From (1882) and §D.2.1, the gradient of $\| xx^T - A \|_F^2$ is

$$\nabla_x ((x^T x)^2 - 2xx^T A x) = 4(x^T x) x - 4Ax \quad (592)$$

Setting the gradient to 0

$$Ax = x(x^T x) \quad (593)$$

is necessary for optimal solution. Replace vector $x$ with a normalized eigenvector $v_i$ of $A \in S^N$, corresponding to a positive eigenvalue $\lambda_i$, scaled by square root of that eigenvalue. Then (593) is satisfied

$$x \leftarrow v_i \sqrt{\lambda_i} \Rightarrow Av_i = v_i \lambda_i \quad (594)$$

$xx^T = \lambda_i v_i v_i^T$ is a rank-1 matrix on the positive semidefinite cone boundary, and the minimum is achieved (§7.1.2) when $\lambda_i = \lambda_1$ is the largest positive eigenvalue of $A$. If $A$ has no positive eigenvalue, then $x = 0$ yields the minimum.

Differentiability is a prerequisite neither to convexity or to numerical solution of a convex optimization problem. The gradient provides a necessary and sufficient condition (351) (451) for optimality in the constrained case, nevertheless, as it does in the unconstrained case:

- For any differentiable multidimensional convex function, zero gradient $\nabla f = 0$ is a necessary and sufficient condition for its unconstrained minimization [63, §5.5.3]:

3.6.0.0.2 Example. Pseudoinverse.
The pseudoinverse matrix is the unique solution to an unconstrained convex optimization problem [168, §5.5.4]: given $A \in \mathbb{R}^{m \times n}$

$$\begin{align*}
\text{minimize } & \left\| XA - I \right\|_F^2 \\
& \text{subject to } X \in \mathbb{R}^{n \times m}
\end{align*}
$$

where

$$\left\| XA - I \right\|_F^2 = \text{tr} (A^T X^T X A - X A - A^T X^T + I) \quad (596)$$

whose gradient (§D.2.3)

$$\nabla_X \left\| XA - I \right\|_F^2 = 2(X A A^T - A^T) = 0 \quad (597)$$
vanishes when
\[ XAA^T = A^T \] (598)

When \( A \) is fat full-rank, then \( AAT \) is invertible, \( X^* = A^T(AA^T)^{-1} \) is the pseudoinverse \( A^\dagger \), and \( AA^\dagger = I \). Otherwise, we can make \( AA^T \) invertible by adding a positively scaled identity, for any \( A \in \mathbb{R}^{m \times n} \):
\[ X = A^T(AA^T + t I)^{-1} \] (599)
Invertibility is guaranteed for any finite positive value of \( t \) by (1533). Then matrix \( X \) becomes the pseudoinverse \( X \mapsto A^\dagger \approx X^* \) in the limit \( t \to 0^+ \). Minimizing instead \( \|AX - I\|_2^2 \) yields the second flavor in (1622).

3.6.0.0.3 Example. Hyperplane, line, described by affine function.
Consider the real affine function of vector variable, (confer Figure 75)
\[ f(x) : \mathbb{R}^p \to \mathbb{R} = a^T x + b \] (600)
whose domain is \( \mathbb{R}^p \) and whose gradient \( \nabla f(x) = a \) is a constant vector (independent of \( x \)). This function describes the real line \( \mathbb{R} \) (its range), and it describes a nonvertical \( [209, \S B.1.2] \) hyperplane \( \partial H \) in the space \( \mathbb{R}^p \times \mathbb{R} \) for any particular vector \( a \) (confer \( \S 2.4.2 \));
\[ \partial H = \left\{ \begin{bmatrix} x \\ a^T x + b \end{bmatrix} : x \in \mathbb{R}^p \right\} \subset \mathbb{R}^p \times \mathbb{R} \] (601)
having nonzero normal
\[ \eta = \begin{bmatrix} a \\ -1 \end{bmatrix} \in \mathbb{R}^p \times \mathbb{R} \] (602)
This equivalence to a hyperplane holds only for real functions.\(^{3.18}\) Epigraph of real affine function \( f(x) \) is therefore a halfspace in \( [\mathbb{R}^p; \mathbb{R}] \), so we have:

The real affine function is to convex functions as the hyperplane is to convex sets.

\(^{3.18}\) To prove that, consider a vector-valued affine function
\[ f(x) : \mathbb{R}^p \to \mathbb{R}^M = Ax + b \]
having gradient \( \nabla f(x) = A^T \in \mathbb{R}^{p \times M} \); The affine set
\[ \left\{ \begin{bmatrix} x \\ Ax + b \end{bmatrix} : x \in \mathbb{R}^p \right\} \subset \mathbb{R}^p \times \mathbb{R}^M \]
is perpendicular to
\[ \eta \triangleq \begin{bmatrix} \nabla f(x) \\ -I \end{bmatrix} \in \mathbb{R}^{p \times M} \times \mathbb{R}^{M \times M} \]
because
\[ \eta^T \left( \begin{bmatrix} x \\ Ax + b \end{bmatrix} - \begin{bmatrix} 0 \\ b \end{bmatrix} \right) = 0 \quad \forall x \in \mathbb{R}^p \]
Yet \( \eta \) is a vector (in \( \mathbb{R}^p \times \mathbb{R}^M \)) only when \( M = 1 \).
Similarly, the matrix-valued affine function of real variable \( x \), for any particular matrix \( A \in \mathbb{R}^{M \times N} \)

\[
h(x) : \mathbb{R} \rightarrow \mathbb{R}^{M \times N} = Ax + B
\]

describes a line in \( \mathbb{R}^{M \times N} \) in direction \( A \)

\[
\{ Ax + B \mid x \in \mathbb{R} \} \subseteq \mathbb{R}^{M \times N}
\]

and describes a line in \( \mathbb{R} \times \mathbb{R}^{M \times N} \)

\[
\left\{ \left[ \begin{array}{c} x \\ Ax + B \end{array} \right] \mid x \in \mathbb{R} \right\} \subseteq \mathbb{R} \times \mathbb{R}^{M \times N}
\]

whose slope with respect to \( x \) is \( A \).

### 3.6.1 monotonic function

A real function of real argument is called \emph{monotonic} when it is exclusively nonincreasing or nondecreasing over the whole of its domain. A real differentiable function of real argument is monotonic when its first derivative (not necessarily continuous) maintains sign over the function domain.

#### 3.6.1.0.1 Definition. Monotonicity.

In pointed closed convex cone \( \mathcal{K} \), multidimensional function \( f(X) \) is

\[
\text{nondecreasing monotonic when } \quad Y \succeq_{\mathcal{K}} X \Rightarrow f(Y) \succeq f(X) \\
\text{nonincreasing monotonic when } \quad Y \succeq_{\mathcal{K}} X \Rightarrow f(Y) \preceq f(X) \quad (606)
\]

\( \forall X, Y \in \text{dom } f \). Multidimensional function \( f(X) \) is

\[
\text{increasing monotonic when } \quad Y \succ_{\mathcal{K}} X \Rightarrow f(Y) \succ f(X) \\
\text{decreasing monotonic when } \quad Y \succ_{\mathcal{K}} X \Rightarrow f(Y) \prec f(X) \quad (607)
\]

These latter inequalities define \emph{strict monotonicity} when they hold over all \( X, Y \in \text{dom } f \).

\[\triangle\]

For monotonicity of vector-valued functions, \( f \) compared with respect to the nonnegative orthant, it is necessary and sufficient for each entry \( f_i \) to be monotonic in the same sense.

Any affine function is monotonic. In \( \mathcal{K} = \mathbb{S}_+^M \), for example, \( \text{tr}(Z^TX) \) is a nondecreasing monotonic function of matrix \( X \in \mathbb{S}^M \) when constant matrix \( Z \) is positive semidefinite; which follows from a result (377) of Fejér.

A convex function can be characterized by another kind of nondecreasing monotonicity of its gradient:
3.6. GRADIENT

3.6.1.0.2 Theorem. Gradient monotonicity. [209, §B.4.1.4] [57, §3.1 exer.20]
Given real differentiable function $f(X) : \mathbb{R}^{p \times k} \to \mathbb{R}$ with matrix argument on open convex domain, the condition

$$\langle \nabla f(Y) - \nabla f(X), Y - X \rangle \geq 0 \text{ for each and every } X, Y \in \text{dom } f$$  \hspace{1cm} (608)

is necessary and sufficient for convexity of $f$. Strict inequality and caveat $Y \neq X$ constitute necessary and sufficient conditions for strict convexity. ◁

3.6.1.0.3 Example. Composition of functions. [63, §3.2.4] [209, §B.2.1]
Monotonic functions play a vital role determining convexity of functions constructed by transformation. Given functions $g : \mathbb{R}^k \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^k$, their composition $f = g(h) : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) = g(h(x)) , \quad \text{dom } f = \{ x \in \text{dom } h \mid h(x) \in \text{dom } g \}$$  \hspace{1cm} (609)

is convex if

- $g$ is convex nondecreasing monotonic and $h$ is convex
- $g$ is convex nonincreasing monotonic and $h$ is concave

and composite function $f$ is concave if

- $g$ is concave nondecreasing monotonic and $h$ is concave
- $g$ is concave nonincreasing monotonic and $h$ is convex

where $\infty (-\infty)$ is assigned to convex (concave) $g$ when evaluated outside its domain. For differentiable functions, these rules are consequent to (1883).

- Convexity (concavity) of any $g$ is preserved when $h$ is affine. □

If $f$ and $g$ are nonnegative convex real functions, then $(f(x)^k + g(x)^k)^{1/k}$ is also convex for integer $k \geq 1$. [259, p.44] A squared norm is convex having the same minimum because a squaring operation is convex nondecreasing monotonic on the nonnegative real line.

3.6.1.0.4 Exercise. Order of composition.
Real function $f = x^{-2}$ is convex on $\mathbb{R}_+$ but not predicted so by results in Example 3.6.1.0.3 when $g = h(x)^{-1}$ and $h = x^2$. Explain this anomaly. ▼

The following result for product of real functions is extensible to inner product of multidimensional functions on real domain:
3.6.1.0.5 Exercise. Product and ratio of convex functions. [63, exer.3.32]
In general the product or ratio of two convex functions is not convex. [240] However, there are some results that apply to functions on \( \mathbb{R} \) [real domain]. Prove the following.\(^\text{3.19}\)

(a) If \( f \) and \( g \) are convex, both nondecreasing (or nonincreasing), and positive functions on an interval, then \( fg \) is convex.

(b) If \( f, g \) are concave, positive, with one nondecreasing and the other nonincreasing, then \( fg \) is concave.

(c) If \( f \) is convex, nondecreasing, and positive, and \( g \) is concave, nonincreasing, and positive, then \( f/g \) is convex. ▼

3.6.2 first-order convexity condition, real function
Discretization of \( w \succeq 0 \) in (497) invites refocus to the real-valued function:

3.6.2.0.1 Theorem. Necessary and sufficient convexity condition. [43, §1.2] [317, §3] [63, §3.1.3] [139, §1.5.2] [340, §4.2] [423, §1.2.3] For \( f(X) : \mathbb{R}^p \times \mathbb{R} \rightarrow \mathbb{R} \) a real differentiable function with matrix argument on open convex domain, the condition (confer §D.1.7)
\[
f(Y) \geq f(X) + \langle \nabla f(X), Y - X \rangle \quad \text{for each and every } X, Y \in \text{dom } f
\]
(610)
is necessary and sufficient for convexity of \( f \). Caveat \( Y \neq X \) and strict inequality again constitute necessary and sufficient conditions for strict convexity. [209, §B.4.1.1] ◊

When \( f(X) : \mathbb{R}^p \rightarrow \mathbb{R} \) is a real differentiable convex function with vector argument on open convex domain, there is simplification of the first-order condition (610); for each and every \( X, Y \in \text{dom } f \)
\[
f(Y) \geq f(X) + \nabla f(X)^T (Y - X) \quad \text{for each and every } X, Y \in \text{dom } f
\]
(611)
From this we can find \( \partial \mathcal{H}_- \) a unique [403, p.220-229] nonvertical [209, §B.1.2] hyperplane (§2.4), expressed in terms of function gradient, supporting epi \( f \) at \( \begin{bmatrix} X \\ f(X) \end{bmatrix} \): videlicet, defining \( f(Y \notin \text{dom } f) \triangleq \infty \) [63, §3.1.7]
\[
\begin{bmatrix} Y \\ t \end{bmatrix} \in \text{epi } f \Leftrightarrow t \geq f(Y) \Rightarrow \left[ \nabla f(X)^T - 1 \right] \begin{bmatrix} Y \\ t \end{bmatrix} - \begin{bmatrix} X \\ f(X) \end{bmatrix} \leq 0
\]
(612)
This means, for each and every point \( X \) in the domain of a convex real function \( f(X) \), there exists a hyperplane \( \partial \mathcal{H}_- \) in \( \mathbb{R}^p \times \mathbb{R} \) having normal \( \begin{bmatrix} \nabla f(X) \\ -1 \end{bmatrix} \) supporting the function epigraph at \( \begin{bmatrix} X \\ f(X) \end{bmatrix} \in \partial \mathcal{H}_- \)
\[
\partial \mathcal{H}_- = \left\{ \begin{bmatrix} Y \\ t \end{bmatrix} \in \mathbb{R}^p \times \mathbb{R} \left| \begin{bmatrix} \nabla f(X)^T \\ -1 \end{bmatrix} \begin{bmatrix} Y \\ t \end{bmatrix} - \begin{bmatrix} X \\ f(X) \end{bmatrix} \right| = 0 \right\}
\]
\(^{3.19}\)Hint: Prove §3.6.1.0.5a by verifying Jensen’s inequality ((496) at \( \mu = \frac{1}{2} \)).
3.6. GRADIENT

Figure 79: When a real function $f$ is differentiable at each point in its open domain, there is an intuitive geometric interpretation of function convexity in terms of its gradient $\nabla f$ and its epigraph: Drawn is a convex quadratic bowl in $\mathbb{R}^2 \times \mathbb{R}$ (confer Figure 172 p.584); $f(Y) = Y^T Y : \mathbb{R}^2 \to \mathbb{R}$ versus $Y$ on some open disc in $\mathbb{R}^2$. Unique strictly supporting hyperplane $\partial H \subseteq \mathbb{R}^2 \times \mathbb{R}$ (only partially drawn) and its normal vector $[\nabla f(X)^T - 1]^T$ at the particular point of support $[X^T \ f(X)]^T$ are illustrated. The interpretation: At each and every coordinate $Y$, there is a unique hyperplane containing $[Y^T \ f(Y)]^T$ and supporting the epigraph of convex differentiable $f$. 
CHAPTER 3. GEOMETRY OF CONVEX FUNCTIONS

Such a hyperplane is strictly supporting whenever a function is strictly convex. One such supporting hyperplane (confer Figure 31a) is illustrated in Figure 79 for a convex quadratic.

From (611) we deduce, for each and every \( X, Y \in \text{dom } f \) in the domain,

\[
\nabla f(X)^T(Y - X) \geq 0 \Rightarrow f(Y) \geq f(X)
\]

meaning, the gradient at \( X \) identifies a supporting hyperplane there in \( \mathbb{R}^p \)

\[
\{ Y \in \mathbb{R}^p \mid \nabla f(X)^T(Y - X) = 0 \}
\]

to the convex sublevel sets of convex function \( f \) (confer (560))

\[
\mathcal{L}_{f(X)} f \triangleq \{ Z \in \text{dom } f \mid f(Z) \leq f(X) \} \subseteq \mathbb{R}^p
\]

illustrated for an arbitrary convex real function in Figure 80 and Figure 69. That supporting hyperplane is unique for twice differentiable \( f \). [229, p.501]

3.6.3 first-order convexity condition, vector-valued \( f \)

Now consider the first-order necessary and sufficient condition for convexity of a vector-valued function: Differentiable function \( f(X) : \mathbb{R}^{p \times k} \to \mathbb{R}^M \) is convex if and only if dom \( f \) is open, convex, and for each and every \( X, Y \in \text{dom } f \)

\[
f(Y) \geq f(X) + \frac{\partial f}{\partial X}(X) \triangleq f(X) + \frac{d}{dt} \bigg|_{t=0} f(X + t(Y - X))
\]

where \( \frac{\partial f}{\partial X}(X) \) is the directional derivative \( \text{3.20} \) [229] [342] of \( f \) at \( X \) in direction \( Y - X \). This, of course, follows from the real-valued function case: by dual generalized inequalities (§2.13.2.0.1),

\[
f(Y) - f(X) - \frac{\partial f}{\partial X}(X) \geq 0 \iff \begin{pmatrix} f(Y) - f(X) - \frac{\partial f}{\partial X}(X) \end{pmatrix} \begin{pmatrix} w \end{pmatrix} \geq 0 \quad \forall w \succeq 0 \quad \mathbb{R}_+^M
\]

where

\[
\frac{\partial f}{\partial X}(X) = \begin{bmatrix} \text{tr}(\nabla f_1(X)^T(Y - X)) \\ \text{tr}(\nabla f_2(X)^T(Y - X)) \\ \vdots \\ \text{tr}(\nabla f_M(X)^T(Y - X)) \end{bmatrix} \in \mathbb{R}^M
\]

Necessary and sufficient discretization (497) allows relaxation of the semiinfinite number of conditions \( \{ w \geq 0 \} \) instead to \( \{ w \in \{ e_i \_i = 1 \ldots M \} \} \) the extreme directions of the selfdual nonnegative orthant. Each extreme direction picks out a real entry \( f_i \) and \( \frac{\partial f}{\partial X} \_i \) from the vector-valued function and its directional derivative, then Theorem 3.6.2.0.1 applies.

The vector-valued function case (617) is therefore a straightforward application of the first-order convexity condition for real functions to each entry of the vector-valued function.

\text{3.20} We extend the traditional definition of directional derivative in §D.1.4 so that direction may be indicated by a vector or a matrix, thereby broadening scope of the Taylor series (§D.1.7). The right-hand side of inequality (617) is the first-order Taylor series expansion of \( f \) about \( X \).
Figure 80: (confer Figure 69) Shown is a plausible contour plot in $\mathbb{R}^2$ of some arbitrary real differentiable convex function $f(Z)$ at selected levels $\alpha$, $\beta$, and $\gamma$; contours of equal level $f$ (level sets) drawn in the function’s domain. A convex function has convex sublevel sets $L_{f(X)}f$ (616). [318, §4.6] The sublevel set whose boundary is the level set at $\alpha$, for instance, comprises all shaded regions. For any particular convex function, the family comprising all its sublevel sets is nested. [209, p.75] Were sublevel sets not convex, we may certainly conclude the corresponding function is neither convex. Contour plots of real affine functions are illustrated in Figure 28 and Figure 75.
3.6.4 second-order convexity condition

Again, by discretization (497), we are obliged only to consider each individual entry $f_i$ of a vector-valued function $f$; id est, the real functions $\{f_i\}$.

For $f(X): \mathbb{R}^p \rightarrow \mathbb{R}^M$, a twice differentiable vector-valued function with vector argument on open convex domain,

$$\nabla^2 f_i(X) \succeq S^p_{+} \quad \forall X \in \text{dom } f, \quad i=1 \ldots M \quad (620)$$

is a necessary and sufficient condition for convexity of $f$. Obviously, when $M=1$, this convexity condition also serves for a real function. Condition (620) demands nonnegative curvature, intuitively, hence precluding points of inflection as in Figure 81 (p.226).

Strict inequality in (620) provides only a sufficient condition for strict convexity, but that is nothing new; videlicet, strictly convex real function $f_i(x) = x^4$ does not have positive second derivative at each and every $x \in \mathbb{R}$. Quadratic forms constitute a notable exception where the strict-case converse holds reliably.

3.6.4.0.1 Example. Convex quadratic.

Real quadratic multivariate polynomial in matrix $A$ and vector $b$

$$x^T A x + 2b^T x + c \quad (621)$$

is convex if and only if $A \succeq 0$. Proof follows by observing second order gradient: ($\S$D.2.1)

$$\nabla^2 x (x^T A x + 2b^T x + c) = A + A^T \quad (622)$$

Because $x^T(A + A^T)x = 2x^T A x$, matrix $A$ can be assumed symmetric. □

3.6.4.0.2 Exercise. Real fractional function. (confer $\S$3.3, $\S$3.5.0.4, $\S$3.8.2.0.1)

Prove that real function $f(x, y) = y/x$ is not convex on the first quadrant. Also exhibit this in a plot of the function. ($f$ is quasilinear (p.235) on $\{x > 0\}$ and nonmonotonic; even on the first quadrant.) ▼

3.6.4.0.3 Exercise. Stress function.

Define $|x-y| \triangleq \sqrt{(x-y)^2}$ and

$$X = [x_1 \cdots x_N] \in \mathbb{R}^{1 \times N} \quad (76)$$

Given symmetric nonnegative data $[h_{ij}] \in \mathbb{S}^N \cap \mathbb{R}^{N \times N}_{+}$, consider function

$$f(\text{vec } X) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (|x_i-x_j| - h_{ij})^2 \in \mathbb{R} \quad (1390)$$

Find a gradient and Hessian for $f$. Then explain why $f$ is not a convex function; id est, why doesn’t second-order condition (620) apply to the constant positive semidefinite Hessian matrix you found. For $N=6$ and $h_{ij}$ data from (1463), apply line theorem 3.7.3.0.1 to plot $f$ along some arbitrary lines through its domain. ▼
3.6.4.1 second-order ⇒ first-order condition

For a twice-differentiable real function \( f_i(X) : \mathbb{R}^p \rightarrow \mathbb{R} \) having open domain, a consequence of the \textit{mean value theorem} from calculus allows compression of its complete Taylor series expansion about \( X \in \text{dom} f_i \) (§D.1.7) to three terms: On some open interval of \( \|Y\|_2 \), so that each and every line segment \([X, Y]\) belongs to \( \text{dom} f_i \), there exists an \( \alpha \in [0, 1] \) such that

\[
f_i(Y) = f_i(X) + \nabla f_i(X)^T (Y - X) + \frac{1}{2} (Y - X)^T \nabla^2 f_i(\alpha X + (1 - \alpha)Y)(Y - X)
\]  

(623)

The first-order condition for convexity (611) follows directly from this and the second-order condition (620).

3.7 Convex matrix-valued function

We need different tools for matrix argument: We are primarily interested in continuous matrix-valued functions \( g(X) \). We choose symmetric \( g(X) \in \mathbb{S}^M \) because matrix-valued functions are most often compared (624) with respect to the positive semidefinite cone \( \mathbb{S}^M_+ \) in the ambient space of symmetric matrices.

3.7.0.0.1 Definition. \textit{Convex matrix-valued function}:

1) \textit{Matrix-definition}.
   A function \( g(X) : \mathbb{R}^{p \times k} \rightarrow \mathbb{S}^M \) is convex in \( X \) iff \( \text{dom} g \) is a convex set and, for each and every \( Y, Z \in \text{dom} g \) and all \( 0 \leq \mu \leq 1 \)

\[
g(\mu Y + (1 - \mu)Z) \preceq \mu g(Y) + (1 - \mu)g(Z)
\]

(624)

Reversing sense of the inequality flips this definition to concavity. Strict convexity is defined less a stroke of the pen in (624) similarly to (499).

2) \textit{Scalar-definition}.
   It follows that \( g(X) : \mathbb{R}^{p \times k} \rightarrow \mathbb{S}^M \) is convex in \( X \) iff \( w^T g(X)w : \mathbb{R}^{p \times k} \rightarrow \mathbb{R} \) is convex in \( X \) for each and every \( \|w\| = 1 \); shown by substituting the defining inequality (624). By dual generalized inequalities we have the equivalent but more broad criterion, (§2.13.5)

\[
g \text{ convex w.r.t } \mathbb{S}^M_+ \iff \langle W, g \rangle \text{ convex for each and every } W \succeq 0 \quad \mathbb{S}^M_+^* \]

(625)

Strict convexity on both sides requires \textit{caveat} \( W \neq 0 \). Because the set of all extreme directions for the selfdual positive semidefinite cone (§2.9.2.7) comprises a minimal set of generators for that cone, discretization (§2.13.4.2.1) allows replacement of matrix \( W \) with symmetric dyad \( ww^T \) as proposed.

Function symmetry is not a necessary requirement for convexity; indeed, for \( A \in \mathbb{R}^{m \times p} \) and \( B \in \mathbb{R}^{m \times k} \), \( g(X) = AX + B \) is a convex (affine) function in \( X \) on domain \( \mathbb{R}^{p \times k} \) with respect to the nonnegative orthant \( \mathbb{R}^{m \times k}_+ \). Symmetric convex functions share the same benefits as symmetric matrices. Horn & Johnson [212, §7.7] liken symmetric matrices to real numbers, and (symmetric) positive definite matrices to positive real numbers.
3.7.0.2 Example. Taxicab distance matrix.
Consider an $n$-dimensional vector space $\mathbb{R}^n$ with metric induced by the 1-norm. Then distance between points $x_1$ and $x_2$ is the norm of their difference: $\|x_1 - x_2\|_1$. Given a list of points arranged columnar in a matrix

$$X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \in \mathbb{R}^{n \times N}$$

(76)

then we could define a taxicab distance matrix

$$D_1(X) \triangleq (I \otimes 1^T_n) |\text{vec}(X)| 1^T - 1 \otimes X | \in S_0^N \cap \mathbb{R}^{N \times N}$$

$$= \begin{bmatrix} 0 & \|x_1 - x_2\|_1 & \|x_1 - x_3\|_1 & \cdots & \|x_1 - x_N\|_1 \\ \|x_1 - x_2\|_1 & 0 & \|x_2 - x_3\|_1 & \cdots & \|x_2 - x_N\|_1 \\ \|x_1 - x_3\|_1 & \|x_2 - x_3\|_1 & 0 & \cdots & \|x_3 - x_N\|_1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \|x_1 - x_N\|_1 & \|x_2 - x_N\|_1 & \|x_3 - x_N\|_1 & \cdots & 0 \end{bmatrix}$$

(626)

where $1_n$ is a vector of ones having dim $1_n = n$ and where $\otimes$ represents Kronecker product. This matrix-valued function is convex with respect to the nonnegative orthant since, for each and every $Y, Z \in \mathbb{R}^{N \times N}$ and all $0 \leq \mu \leq 1$

$$D_1(\mu Y + (1 - \mu)Z) \preceq \mu D_1(Y) + (1 - \mu)D_1(Z)$$

(627)

3.7.0.3 Exercise. 1-norm distance matrix.
The 1-norm is called taxicab distance because to go from one point to another in a city by car, road distance is a sum of grid lengths. Prove (627). □

3.7.1 first-order convexity condition, matrix-valued $f$

From the scalar-definition (§3.7.0.1) of a convex matrix-valued function, for differentiable function $g$ and for each and every real vector $w$ of unit norm $\|w\| = 1$, we have

$$w^T g(Y) w \geq w^T g(X) w + w^T \nabla_{-Y} g(X) w$$

(628)

that follows immediately from the first-order condition (610) for convexity of a real function because

$$w^T \nabla_{-Y} g(X) w = \langle \nabla_X w^T g(X) w, Y - X \rangle$$

(629)

where $\nabla_{-Y} g(X)$ is the directional derivative (§D.1.4) of function $g$ at $X$ in direction $Y - X$.

By discretized dual generalized inequalities, (§2.13.5)

$$g(Y) - g(X) - \nabla_{-Y} g(X) \succeq 0 \iff \left(g(Y) - g(X) - \nabla_{-Y} g(X), w w^T \right) \succeq 0 \quad \forall w w^T (\succeq 0)$$

(630)

For each and every $X, Y \in \text{dom} g$ (confer (617))
3.7. CONVEX MATRIX-VALUED FUNCTION

\[ g(Y) \succeq g(X) + \frac{Y - X}{S^M} dg(X) \]  

(631)

must therefore be necessary and sufficient for convexity of a matrix-valued function of matrix variable on open convex domain.

3.7.2 epigraph of matrix-valued function, sublevel sets

We generalize epigraph to a continuous matrix-valued function \( g(X) : \mathbb{R}^{p \times k} \rightarrow S^M \):

\[ \text{epi } g \triangleq \{(X, T) \in \mathbb{R}^{p \times k} \times S^M \mid X \in \text{dom } g, g(X) \preceq S^M + T \} \]  

(632)

from which it follows

\[ g \text{ convex } \iff \text{epi } g \text{ convex} \]  

(633)

Proof of necessity is similar to that in §3.5 on page 204.

Sublevel sets of a convex matrix-valued function corresponding to each and every \( S \in S^M \) (confer (560))

\[ \mathcal{L}_S g \triangleq \{X \in \text{dom } g \mid g(X) \preceq S \} \subseteq \mathbb{R}^{p \times k} \]  

(634)

are convex. There is no converse.

3.7.2.0.1 Example. Matrix fractional function redux.  
[35, p.155]

Generalizing Example 3.5.0.0.4 consider a matrix-valued function of two variables on \( \text{dom } g = S^N_+ \times \mathbb{R}^{n \times N} \) for small positive constant \( \epsilon \) (confer (1963))

\[ g(A, X) = \epsilon X(A + \epsilon I)^{-1}X^T \]  

(635)

where the inverse always exists by (1533). This function is convex simultaneously in both variables over the entire positive semidefinite cone \( S^N_+ \) and all \( X \in \mathbb{R}^{n \times N} \): Consider Schur-form (1596) from §A.4: for \( T \in S^n \)

\[ G(A, X, T) = \begin{bmatrix} A + \epsilon I & X^T \\ X & \epsilon^{-1} T \end{bmatrix} \succeq 0 \iff \\
T - \epsilon X(A + \epsilon I)^{-1}X^T \succeq 0 \\
A + \epsilon I \succ 0 \]  

(636)

By Theorem 2.1.9.0.1, inverse image of the positive semidefinite cone \( S^N_+ \) under affine mapping \( G(A, X, T) \) is convex. Function \( g(A, X) \) is convex on \( S^N_+ \times \mathbb{R}^{n \times N} \) because its epigraph is that inverse image:

\[ \text{epi } g(A, X) = \{(A, X, T) \mid A + \epsilon I \succ 0, \epsilon X(A + \epsilon I)^{-1}X^T \preceq T \} = G^{-1}(S^N_+) \]  

(637)
3.7.3 second-order convexity condition, matrix-valued $f$

The following line theorem is a potent tool for establishing convexity of a multidimensional function. To understand it, what is meant by line must first be solidified. Given a function $g(X) : \mathbb{R}^{p \times k} \rightarrow S_M^+$ and particular $X, Y \in \mathbb{R}^{p \times k}$ not necessarily in that function’s domain, then we say a line $\{X + tY \mid t \in \mathbb{R}\}$ (infinite in extent) passes through dom $g$ when $X + tY \in \text{dom} \, g$ over some interval of $t \in \mathbb{R}$.

3.7.3.0.1 Theorem. Line theorem. (confer [63, §3.1.1])

Multidimensional function $f(X) : \mathbb{R}^{p \times k} \rightarrow \mathbb{R}^M$ or $g(X) : \mathbb{R}^{p \times k} \rightarrow S_M^+$ is convex in $X$ if and only if it remains convex on the intersection of any line with its domain. ☐

Now we assume a twice differentiable function.

3.7.3.0.2 Definition. Differentiable convex matrix-valued function. Matrix-valued function $g(X) : \mathbb{R}^{p \times k} \rightarrow S_M^+$ is convex in $X$ iff dom $g$ is an open convex set, and its second derivative $g''(X + tY) : \mathbb{R} \rightarrow S_M^+$ is positive semidefinite on each point of intersection along every line $\{X + tY \mid t \in \mathbb{R}\}$ that intersects dom $g$; id est, iff for each and every $X, Y \in \mathbb{R}^{p \times k}$ such that $X + tY \in \text{dom} \, g$ over some open interval of $t \in \mathbb{R}$

$$\frac{d^2}{dt^2} g(X + tY) \succeq 0 \quad (638)$$

Similarly, if

$$\frac{d^2}{dt^2} g(X + tY) \succ 0 \quad (639)$$

then $g$ is strictly convex; the converse is generally false. [63, §3.1.4] 3.22 △

3.7.3.0.3 Example. Matrix inverse. (confer §3.3.1)

The matrix-valued function $X^\mu$ is convex on int $S_M^+$ for $-1 \leq \mu \leq 0$ or $1 \leq \mu \leq 2$ and concave for $0 \leq \mu \leq 1$. [63, §3.6.2] In particular, the function $g(X) = X^{-1}$ is convex on int $S_M^+$. For each and every $Y \in S_M^+$ (§D.2.1, §A.3.1.0.5)

$$\frac{d^2}{dt^2} g(X + tY) = 2(X + tY)^{-1}Y(X + tY)^{-1}Y(X + tY)^{-1} \succeq 0 \quad (640)$$

on some open interval of $t \in \mathbb{R}$ such that $X + tY \succ 0$. Hence, $g(X)$ is convex in $X$. This result is extensible: 3.23 $\text{tr} \, X^{-1}$ is convex on that same domain. [212, §7.6 prob.2] [57, §3.1 exer.25] ☐

\[3.22\] The strict-case converse is reliably true for quadratic forms.

\[3.23\] $d/dt \, \text{tr} \, g(X + tY) = \text{tr} \, d/dt \, g(X + tY)$. [213, p.491]
3.7. CONVEX MATRIX-VALUED FUNCTION

3.7.3.0.4 Example. Matrix squared.
Iconic real function \( f(x) = x^2 \) is strictly convex on \( \mathbb{R} \). The matrix-valued function \( g(X) = X^2 \) is convex on the domain of symmetric matrices; for \( X, Y \in \mathbb{S}^M \) and any open interval of \( t \in \mathbb{R} \):\(^{\text{D.2.1}}\)

\[
\frac{d^2}{dt^2} g(X + tY) = \frac{d^2}{dt^2} (X + tY)^2 = 2Y^2
\]
which is positive semidefinite when \( Y \) is symmetric because then \( Y^2 = Y^TY \).\(^{\text{E.2.1}}\)

A more appropriate matrix-valued counterpart for \( f \) is \( g(X) = X^T X \) which is a convex function on domain \( \{ X \in \mathbb{R}^{m \times n} \} \), and strictly convex whenever \( X \) is skinny-or-square full-rank. This matrix-valued function can be generalized to \( g(X) = X^T A X \) which is convex whenever matrix \( A \) is positive semidefinite (p.594), and strictly convex when \( A \) is positive definite and \( X \) is skinny-or-square full-rank (Corollary A.3.1.0.5). \( \Box \)

3.7.3.0.5 Exercise. Squared maps.
Give seven examples of distinct polyhedra \( \mathcal{P} \) for which the set \( \{ X^T X \mid X \in \mathcal{P} \} \subseteq \mathbb{S}^n_+ \)
were convex. Is this set convex, in general, for any polyhedron \( \mathcal{P} \)? (confer (1315)(1322)) Is the epigraph of function \( g(X) = X^T X \) convex for any polyhedral domain? \( \nabla \)

3.7.3.0.6 Exercise. Squared inverse.\(^{\text{confer \S 3.7.2.0.1}}\)
For positive scalar \( a \), real function \( f(x) = ax^{-2} \) is convex on the nonnegative real line. Given positive definite matrix constant \( A \), prove via line theorem that \( g(X) = \text{tr}((X^TA^{-1}X)^{-1}) \) is generally not convex unless \( X \succ 0 \).\(^{3.25}\) From this result, show how it follows via Definition 3.7.0.0.1-2 that \( h(X) = (X^TA^{-1}X)^{-1} \) is generally neither convex.

3.7.3.0.7 Example. Matrix exponential.
The matrix-valued function \( g(X) = e^X : \mathbb{S}^M \to \mathbb{S}^M \) is convex on the subspace of circulant \[178\] symmetric matrices. Applying the line theorem, for all \( t \in \mathbb{R} \) and circulant \( X, Y \in \mathbb{S}^M \), from Table D.2.7 we have

\[
\frac{d^2}{dt^2} e^{X + tY} = Y e^{X + tY} \gtrless 0, \quad (XY)^T = XY
\]
because all circulant matrices are commutative and, for symmetric matrices, \( XY = YX \iff (XY)^T = XY \). Given symmetric argument, the matrix exponential always resides interior to the cone of positive semidefinite matrices in the symmetric matrix subspace; \( e^A \succ 0 \forall A \in \mathbb{S}^M \). Then for any matrix \( Y \) of compatible dimension, \( Y^Te^AY \) is positive semidefinite. (\S A.3.1.0.5)

\(^{3.24}\)By (1557) in \S A.3.1, changing the domain instead to all symmetric and nonsymmetric positive semidefinite matrices, for example, will not produce a convex function. \(^{3.25}\)Hint: D.2.3.
The subspace of circulant symmetric matrices contains all diagonal matrices. The matrix exponential of any diagonal matrix $e^{\Lambda}$ exponentiates each individual entry on the main diagonal. [261, §5.3] So, changing the function domain to the subspace of real diagonal matrices reduces the matrix exponential to a vector-valued function in an isometrically isomorphic subspace $\mathbb{R}^M$; known convex (§3.1.1) from the real-valued function case [63, §3.1.5]. □

There are more methods for determining function convexity; [43] [63] [139] one can be more efficient than another depending on the function in hand.

3.7.3.0.8 Exercise. *log det function.*

Matrix determinant is neither a convex or concave function, in general, but its inverse is convex when domain is restricted to interior of a positive semidefinite cone. [35, p.149] Show by three different methods: On interior of the positive semidefinite cone, $\log \det X = -\log \det X^{-1}$ is concave.

3.8 Quasiconvex

Quasiconvex functions [209] [340] [403] [254, §2] are valuable, pragmatically, because they are unimodal (by definition when nonmonotonic): a global minimum is guaranteed to exist over any convex set in the function domain; e.g, Figure 81. Optimal solution to quasiconvex problems is by method of bisection (a.k.a binary search). [63, §4.2.5]
3.8. QUASICONVEX

3.8.0.0.1 Definition. Quasiconvex function. (confer (496))

\( f(X) : \mathbb{R}^{p \times k} \rightarrow \mathbb{R} \) is a quasiconvex function of matrix \( X \) iff \( \text{dom} \ f \) is a convex set and for each and every \( Y, Z \in \text{dom} \ f \) and \( 0 \leq \mu \leq 1 \)

\[
f(\mu Y + (1 - \mu)Z) \leq \max\{f(Y), f(Z)\}
\]

(644)

A quasiconcave function is similarly defined:

\[
f(\mu Y + (1 - \mu)Z) \geq \min\{f(Y), f(Z)\}
\]

(645)

Caveat. \( Y \neq Z \) and strict inequality on an open interval \( 0 < \mu < 1 \) constitute necessary and sufficient conditions for strict quasiconvexity. △

Unlike convex functions, quasiconvex functions are not necessarily continuous; \( e.g. \), quasiconcave \( \text{rank}(X) \) on \( \mathbb{S}_+^M \) (§2.9.2.9.2) and \( \text{card}(x) \) on \( \mathbb{R}_+^M \). Although insufficient for convex functions, convexity of each and every sublevel set serves as a definition of quasiconvexity:

3.8.0.0.2 Definition. Quasiconvex multidimensional function.

Scalar-, vector-, or matrix-valued function \( g(X) : \mathbb{R}^{p \times k} \rightarrow \mathbb{S}_+^M \) is a quasiconvex function of matrix \( X \) iff \( \text{dom} \ g \) is a convex set and its sublevel set

\[
\mathcal{L}_S g = \{ X \in \text{dom} \ g \mid g(X) \preceq S \} \subseteq \mathbb{R}^{p \times k}
\]

(corresponding to each and every \( S \in \mathbb{S}_+^M \)) is convex. Vectors are compared with respect to the nonnegative orthant \( \mathbb{R}_+^M \) while matrices are with respect to the positive semidefinite cone \( \mathbb{S}_+^M \).

Likewise, convexity of each and every superlevel set

\[
\mathcal{L}^S g = \{ X \in \text{dom} \ g \mid g(X) \succeq S \} \subseteq \mathbb{R}^{p \times k}
\]

is necessary and sufficient for quasiconcavity. △

3.8.0.0.3 Exercise. Nonconvexity of matrix product.

Consider real function \( f \) on a positive definite domain

\[
f(X) = \text{tr}(X_1 X_2), \quad X \triangleq \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \in \text{dom} \ f \triangleq \begin{bmatrix} \text{rel \ int} \mathbb{S}_+^N \\ \text{rel \ int} \mathbb{S}_+^N \end{bmatrix}
\]

(647)

with superlevel sets

\[
\mathcal{L}^s f = \{ X \in \text{dom} \ f \mid \langle X_1, X_2 \rangle \geq s \}
\]

(648)

Prove that \( f(X) \) is not quasiconcave except when \( N = 1 \), nor is it quasiconvex unless \( X_1 = X_2 \). ▼
Figure 82: Strictly monotonic quasiconcave real product function $xu$ is bowed (not affine) on the nonnegative orthants.
3.8 QUASICONVEX

3.8.1 bilinear

Bilinear (inner product) function\(^3\)\(^2\)\(^6\) of vectors \(x\) and \(u\) is quasiconcave and strictly monotonic on the nonnegative orthants \(\mathbb{R}_+^n \times \mathbb{R}_+^\eta\) only when dimension \(\eta\) equals 1. (Figure 82)

When variable \(x \leftarrow \beta\) has dimension 1 but \(u\) is a vector of arbitrary dimension \(\eta\), real function \(f(\beta, u) = \beta^T u = \beta \text{tr} \delta(u)\) is quasiconcave strictly monotonic on the nonnegative orthants \(\mathbb{R}_+ \times \mathbb{R}_+^\eta\). \(f(\beta, u)\) is quasiconcave strictly monotonic on \(\mathbb{R}_+ \times \mathbb{R}_+^\eta\) when \(1^T u \geq 0\).

3.8.1.0.1 Proof. Domain of function \(f(\beta, u) = \beta^T u, \ 1^T u \geq 0\) is a nonpointed polyhedral cone in \(\mathbb{R}_+^{\eta+1}\), its range a halfline \(\mathbb{R}_+\).

(quasiconcavity) Because this function spans an orthant, its 0-superlevel set is the deepest superlevel set and identical to its domain. Higher superlevel sets of the function, given some fixed nonzero scalar \(\zeta \geq 0\)

\[
\{ \beta,u \mid f(\beta, u) \geq \zeta, \beta > 0, 1^T u \geq 0 \} = \{ \beta,u \mid \beta^T 1 u \geq \zeta, \beta > 0, 1^T u \geq 0 \}
\]

are not polyhedral but they are convex because (\(\S A.4\))

\[
1^T u \geq \frac{\zeta}{\beta}, \quad \beta > 0 \iff \begin{bmatrix} \beta \\ \sqrt{\zeta} \\ \text{tr} \delta(u) \end{bmatrix} \succeq 0 \quad (651)
\]

and because inverse image of a positive semidefinite cone under affine transformation is convex by Theorem 2.1.9.0.1. Convex superlevel sets are necessary and sufficient for quasiconcavity by Definition 3.8.0.0.2.

(monotonicity) By Definition 3.6.1.0.1,

\[
f \text{ is increasing monotonic when } \begin{bmatrix} \beta \\ u \end{bmatrix} \succ \begin{bmatrix} \tau \\ z \end{bmatrix} \Rightarrow f(\beta, u) \succ f(\tau, z) \quad (652)
\]

for all \(\beta, u, \tau, z\) in the domain. Assuming \(\beta > \tau \geq 0\) and \(u \succ z\), it follows that \(1^T u > 1^T z\). (Exercise 2.13.7.0.2) Therefore \(\beta 1^T u > \tau 1^T z\) and so \(f(\beta, u) = \beta 1^T u\) is strictly monotonic.

3.8.1.0.2 Example. Arbitrary magnitude analog filter design.

Analog filter design means determination of Laplace transfer function coefficients to meet specified tolerances in magnitude and phase over given frequencies. The problem posed by this example is to find a stable minimum phase filter \(H(\omega)\), of given order \(\eta\), closest to specified samples of frequency-domain magnitude

\[
\{ g_i \mid \omega_i \in \Omega \} \quad (653)
\]

\(^3\)\(^2\)\(^6\)Convex envelope of bilinear functions is well known. [4]
We consider a recursive filter whose real causal impulse response has Laplace transform

\[ H(s) = \frac{1 + b_1 s + b_2 s^2 + \ldots + b_n s^n}{1 + a_1 s + a_2 s^2 + \ldots + a_n s^n} \]  

(654)

whose poles and zeros lie in the left halfplane \((\text{re} s = \sigma < 0)\) and whose gain is unity at DC \((s = 0)\). This transfer function is a ratio of polynomials in complex variable \(s\) defined

\[ s \triangleq \sigma + j\omega \]  

(655)

To reduce passive component sensitivity, physical implementation is facilitated by factoring transform \((654)\) into parallel or cascade second order sections which are needed to realize complex poles and zeros. Magnitude square of a second order transfer function evaluated along the \(j\omega\) axis (the Fourier domain) is

\[ |H(j\omega)|^2 = H(j\omega)H(-j\omega) \triangleq \frac{V(\omega)}{U(\omega)} = \frac{1 + v_1 \omega^2 + v_2 \omega^4 + \ldots + v_n \omega^{2n}}{1 + u_1 \omega^2 + u_2 \omega^4 + \ldots + u_n \omega^{2n}} \]  

(657)

A cascade implementation of second order sections can realize a high order unity gain filter. Magnitude square of \(\eta\)th order transfer function \(H\), evaluated along the \(j\omega\) axis, is

\[ |H(j\omega)|^2 = H(j\omega)H(-j\omega) \triangleq \frac{V(\omega)}{U(\omega)} = \frac{1 + v_1 \omega^2 + v_2 \omega^4 + \ldots + v_\eta \omega^{2\eta}}{1 + u_1 \omega^2 + u_2 \omega^4 + \ldots + u_\eta \omega^{2\eta}} \]  

(658)

where \((\text{confer §4.4.1.2.5})\)

\[
\begin{align*}
v_1 &= \tilde{v}_1 + \tilde{v}_3, & u_1 &= \tilde{u}_1 + \tilde{u}_3, \\
v_2 &= \tilde{v}_2 + \tilde{v}_4 + \tilde{v}_1 \tilde{v}_3, & u_2 &= \tilde{u}_2 + \tilde{u}_4 + \tilde{u}_1 \tilde{u}_3, \\
v_3 &= \tilde{v}_2 \tilde{v}_3 + \tilde{v}_1 \tilde{v}_4, & u_3 &= \tilde{u}_2 \tilde{u}_3 + \tilde{u}_1 \tilde{u}_4, \\
v_4 &= \tilde{v}_2 \tilde{v}_4, & u_4 &= \tilde{u}_2 \tilde{u}_4
\end{align*}
\]  

(659)

Odd \(\eta\) is implemented by cascading one first order section.

To guise filter design as an optimization problem, magnitude square filter coefficients

\[
v \triangleq \begin{bmatrix} v_1 \\
v_2 \\
\vdots \\
v_\eta \end{bmatrix} \in \mathbb{R}^\eta, \quad u \triangleq \begin{bmatrix} u_1 \\
u_2 \\
\vdots \\
u_\eta \end{bmatrix} \in \mathbb{R}^\eta
\]  

(660)
3.8. QUASICONVEX

become the variables. Given a set of frequencies \( \Omega \) at which finite \( g(\omega) \) is sampled, our notation reflects this role reversal:

\[
\frac{V_i}{U_i} = \frac{V_i(v)}{U_i(u)} \triangleq \frac{1 + v_1 \omega_i^2 + v_2 \omega_i^4 + \ldots + v_n \omega_i^{2n}}{1 + u_1 \omega_i^2 + u_2 \omega_i^4 + \ldots + u_n \omega_i^{2n}}, \quad \omega_i \in \Omega \quad (661)
\]

A better filter design may be obtained if stability and minimum phase requirements are ignored by the optimization. Then coefficients \( v, u \) need not be nonnegatively constrained. A least peak deviation filter design problem may be expressed

\[
\begin{align*}
\text{minimize} & \quad v, u, \beta \\
\text{subject to} & \quad \frac{1}{\beta} \leq \frac{U_i}{V_i} g_i^2 \leq \beta, \quad \omega_i \in \Omega \\
& \quad V_i \geq 0, \quad U_i \geq 0, \quad \omega_i \in \Omega
\end{align*}
\]

which is a nonconvex optimization problem because of a ratio in variables \( v \) & \( u \). Eliminating the ratio:

\[
\begin{align*}
\text{minimize} & \quad \beta \\
\text{subject to} & \quad U_i g_i^2 \leq \beta V_i, \quad \omega_i \in \Omega \\
& \quad V_i g_i^{-2} \leq \beta U_i, \quad \omega_i \in \Omega \\
& \quad V_i \geq 0, \quad U_i \geq 0, \quad \omega_i \in \Omega
\end{align*}
\]

This equivalent problem is likewise nonconvex because of new products in variables \( \beta \) & \( v \) and \( \beta \) & \( u \). The feasible set in variables \( v, u, \beta \) is also nonconvex.

But problem (663) can be solved by applying the fact that each product \( \beta V_i(v), \beta U_i(u) \) increases monotonically. We may then rely on convexity of the feasible set described by linear constraints for fixed \( \beta \). One strategy is to choose a sufficiently large positive \( \beta \) so that (663) is initially feasible; say \( \beta = \beta_\infty \). Then halve \( \beta \) until (663) becomes infeasible; say \( \beta = \beta_\infty \). At infeasibility, backtrack to a feasible \( \beta \). Toward a perfect fit, in the sense that all \( g_i \) are collocated, optimal \( \beta \) approaches 1. This proposed iteration represents a binary search (a.k.a bisection) for minimum \( \beta \), say \( \beta = \beta^\ast \), that is assumed to lie in the interval \( [\beta_\infty, \beta_\infty] \):

\[
\begin{align*}
\beta &= \beta_\infty \\
\beta_\infty &= 0 \\
\text{for } k=1,2,\ldots \text{ until convergence} \{ & \\
\quad \text{if (663) feasible} \{ & \\
\quad & \beta_\infty = \beta_{k-1} \\
\quad \text{else} \{ & \\
\quad & \beta^\ast = \beta_{k-1} \\
\quad \} & \\
\quad \} & \\
\beta &= \beta_k = (\beta_\infty + \beta_\infty)/2
\end{align*}
\]

\footnote{Poles and zeros of \( H(j\omega) \) possess conjugate symmetry because of its real coefficients. Poles and zeros of \( H(j\omega)H(-j\omega) \) possess both conjugate and real symmetry. Stability and minimum phase may be imposed postoptimization by picking only those poles and zeros, respectively from \( H(j\omega)H(-j\omega) \), that reside in the left half s-plane.}
Global convergence to any desired tolerance is certain, in absence of numerical error, characterized by an ever narrowing gap between $\beta_\infty$ and $\beta_0$. At convergence of (664), the nonnegative $\beta^*$ that minimizes (662) is found.

**Proof.** Problem (663) may be equivalently written

\[
\begin{align*}
\text{minimize} & \quad \beta \\
\text{subject to} & \quad U_i g_i^2 \leq r, \quad \omega_i \in \Omega \\
& \quad r \leq \beta V_i, \quad \omega_i \in \Omega \\
& \quad V_i g_i^{-2} \leq t, \quad \omega_i \in \Omega \\
& \quad t \leq \beta U_i, \quad \omega_i \in \Omega \\
& \quad V_i \geq 0, \quad U_i \geq 0, \quad \omega_i \in \Omega
\end{align*}
\] (665)

where $\beta, t, r \in \mathbb{R}_+$ are implicitly nonnegative. The feasible set is not convex because $t$ is variable in $t \leq \beta U_i$; this implicit union of superlevel sets is nonconvex. The same holds for $r \leq \beta V_i$. By Schur complement (§A.4), strictly monotonic quasiconcave functions $\beta V_i$ and $\beta U_i$ (§3.8.1.0.1) may be decomposed

\[
\begin{align*}
\beta V_i \geq r, \quad \beta > 0 & \iff \begin{bmatrix} V_i & \sqrt{r} & \beta \\
\sqrt{r} & \beta & 0 \\
0 & \sqrt{t} & U_i \end{bmatrix} \succeq 0 \\
\beta U_i \geq t, \quad \beta > 0 & \iff \begin{bmatrix} \beta & \sqrt{t} \\
\sqrt{t} & U_i \end{bmatrix} \succeq 0
\end{align*}
\] (666)

By Theorem A.3.1.0.4, problem (665) is thereby equivalent to

\[
\begin{align*}
\text{minimize} & \quad \begin{bmatrix} V(v) & \sqrt{r} & 0 \\
\sqrt{r} & \beta & \sqrt{t} \\
0 & \sqrt{t} & U(u) \end{bmatrix} \\
\text{subject to} & \quad U_i g_i^2 \leq r, \quad \omega_i \in \Omega \\
& \quad V_i g_i^{-2} \leq t, \quad \omega_i \in \Omega \\
& \quad \begin{bmatrix} V_i & \sqrt{r} \beta & 0 \\
\sqrt{r} & \beta \sqrt{t} & 0 \\
0 & \sqrt{t} & U_i \end{bmatrix} \succeq 0, \quad \omega_i \in \Omega
\end{align*}
\] (667)

This means minimizing $\beta$ is like simultaneously minimizing functions $\beta V(v)$ and $\beta U(u)$ over a Cartesian subspace that is the intersection of their domains; namely, over a line containing the $\beta$ axis in an increasing direction. Minimization on any line retains quasiconcavity [120], while minimization on any line in an increasing direction maintains strict monotonicity. But local minima cannot be precluded because minimizing a strictly monotonic function over a more general convex feasible set does not imply increasing direction (§3.6.1.0.1, Figure 82) and because, to begin with, the feasible set in variables $v, u, \beta, t, r$ is nonconvex due to $\sqrt{r}$ and $\sqrt{t}$.

One recourse is to perform bisection (664) on $\beta$ to find its globally optimal value $\beta^*$ in problem (663):

\[
\begin{align*}
\text{find} & \quad v, u \\
\text{subject to} & \quad U_i g_i^2 \leq \beta V_i, \quad \omega_i \in \Omega \\
& \quad V_i g_i^{-2} \leq \beta U_i, \quad \omega_i \in \Omega \\
& \quad V_i \geq 0, \quad U_i \geq 0, \quad \omega_i \in \Omega
\end{align*}
\] (668)
3.8. QUASICONVEX

At each iteration, the feasible set in variables $v, u$ is convex because all the inequalities are linear when $\beta$ is fixed. A necessary and sufficient condition for bisection to find $\beta^*$ requires feasibility for all $\beta \geq \beta^*$ and infeasibility for all $0 \leq \beta < \beta^*$. $\beta$ is bounded below by 0 but not bounded above (nor need it be). Feasibility is guaranteed by existence of a pointed polyhedral cone $K$ described by the linear inequalities as an intersection of halfspaces about the origin in $\mathbb{R}^2$; namely

$$K = \left\{ \left[ \begin{array}{c} V_i \\ U_i \end{array} \right] \in \mathbb{R}^2 \left| \begin{array}{c} \beta - g_i^2 \\ -g_i^{-2} \beta \end{array} \right\} \geq 0 \right\} \subseteq \mathbb{R}^2_+ \quad (669)$$

Cone $K$, the feasible set in dependent variables $V_i$ and $U_i$, ceases to exist for any $\beta < \beta^*$. This fact becomes evident by considering reach of the hyperplane normals which are linear in $\beta$: $[\beta - g_i^2]_T$ is confined to the fourth quadrant, $[-g_i^{-2} \beta]_T$ is confined to the second. As $\beta \to \infty$, the extreme directions of $K$ approach the Cartesian axes in the first quadrant. As $\beta \to 0$, extreme directions collapse inward so as to narrow the cone then empty the feasible set. When $\beta = \beta^* \geq 1$, $K$ becomes a ray emanating from the origin in the first quadrant. Hyperplane-normal linearity in $\beta$ and lack of an objective function here obviate the local-minima obstacle.

High powers of radian frequency, in magnitude square transfer function $|H(j\omega)|^2$ (657), demand excessive precision from floating-point numerics. Even some second order filter design optimization problems can cause the best linear program solvers to fail because double precision (64 bits, 52-bit mantissa) is inadequate. (Saunders provides a quadruple precision solver: 128-bit wordlength, 112-bit mantissa.) High powers of frequency are ameliorated simply by scaling specification (653):

$$\Omega' \triangleq \frac{1}{c} \Omega \quad (670)$$

Scaled frequency-domain magnitude specifications become

$$\{ g'_i \mid \omega'_i \in \Omega' \} = \{ g_i \mid \omega_i \in \Omega \} \quad (671)$$

where

$$\omega' \triangleq \frac{\omega}{c} \quad (672)$$

and $c > 1$ is a constant. A filter is designed as before except magnitude square transfer function (657) is replaced with

$$|H(j\omega')|^2 \triangleq \frac{V'(\omega')}{{U'}'(\omega')} = \frac{1 + v'_1 \omega'^2 + v'_2 \omega'^4 + \ldots + v'^{2n}_n \omega'^{2n}}{1 + u'_1 \omega'^2 + u'_2 \omega'^4 + \ldots + u'^{2n}_n \omega'^{2n}} \quad (673)$$

Optimization problem (663) is solved for primed magnitude square filter coefficients $v', u'$:

$$\frac{V'_i(u')}{U'_i(u')} = \frac{1 + v'_1 \omega'_i^2 + v'_2 \omega'_i^4 + \ldots + v'^{2n}_n \omega'^{2n}_i}{1 + u'_1 \omega'_i^2 + u'_2 \omega'_i^4 + \ldots + u'^{2n}_n \omega'^{2n}_i} , \quad \omega'_i \in \Omega' \quad (674)$$
Figure 83: Arbitrary magnitude analog filter design specifications represented by blue dots. Red circles represent fit.
3.9. **SALIENT PROPERTIES**

Optimal unprimed coefficients are scaled replicas of optimal primed:

\[
\mathbf{v}^* \triangleq \begin{bmatrix}
\frac{v_1^*}{c^2} \\
\frac{v_2^*}{c^4} \\
\vdots \\
\frac{v_n^*}{c^{2n}}
\end{bmatrix}, \quad \mathbf{u}^* \triangleq \begin{bmatrix}
\frac{u_1^*}{c^2} \\
\frac{u_2^*}{c^4} \\
\vdots \\
\frac{u_n^*}{c^{2n}}
\end{bmatrix}
\]  

(675)

The originally desired optimal fit is achieved as primed frequency expands:

\[
|H(j\omega)^*|^2 = \frac{1 + v_1^*(\omega'c)^2 + v_2^*(\omega'c)^4 + \ldots + v_n^*(\omega'c)^{2n}}{1 + u_1^*(\omega'c)^2 + u_2^*(\omega'c)^4 + \ldots + u_n^*(\omega'c)^{2n}}
\]  

(676)

A benefit, of this design methodology, is an unexpectedly low order \( \eta \) required to meet a given filter specification (653) to within reasonable tolerance \( \beta^* \); e.g.,

\[
\frac{1}{2\pi} \Omega = \{ 20 \ 30 \ 45 \ 60 \ 90 \ 125 \ 187 \ 250 \ 375 \ 500 \ 1000 \ 1250 \ 1500 \ 2000 \ 3000 \ 4000 \ 6000 \ 8000 \ 12000 \ 16000 \} \\
20 \log_{10} \{ g_i \} = \{ 0 \ 0 \ 0 \ 0 \ 0 \ 3 \ 6 \ 6 \ 9 \ 12 \ 12 \ 12 \ 6 \ 6 \ 6 \ 12 \ 21 \ 21 \ 21 \ 21 \ 21 \ 21 \}
\]  

(677)

illustrated in Figure 83.

\[ \square \]  

### 3.8.2 quasilinear

When a function is simultaneously quasiconvex and quasiconcave, it is called *quasilinear*. Quasilinear functions are completely determined by convex level sets. Multidimensional quasilinear functions are not necessarily monotonic; e.g., Exercise 3.6.4.0.2.

One-dimensional function \( f(x) = x^3 \) and vector-valued signum function \( \text{sgn}(x) \) are quasilinear, for example.

**3.8.2.0.1 Exercise.** *Quasiconcave product function.* (confer §3.6.4.0.2)

Show that vector-valued function \( \beta \mathbf{u} : \mathbb{R}_+ \times \mathbb{R}_+^\eta \to \mathbb{R}_+^\eta \) is quasiconcave and strictly monotonic but not quasilinear. \[ \nabla \]

### 3.9 Salient properties

**of convex and quasiconvex functions**

1. • A convex function is assumed continuous but not necessarily differentiable on the relative interior of its domain. [318, §10]

• A quasiconvex function is not necessarily a continuous function.

2. **convex epigraph ⇔ convexity ⇒ quasiconvexity ⇔ convex sublevel sets.**

**convex hypograph ⇔ concavity ⇒ quasiconcavity ⇔ convex superlevel.**

**quasilinearity ⇔ convex level sets.**
3. log-convex ⇒ convex ⇒ quasiconvex.  
converse ⇒ quasiconcave ⇔ log-concave ⇐ positive concave.  

4. Line Theorem 3.7.3.0.1 translates identically to quasiconvexity (quasiconcavity).  

5. • $g$ convex ⇔ $-g$ concave.  
$g$ quasiconvex ⇔ $-g$ quasiconcave.  
$g$ log-convex ⇔ $1/g$ log-concave.  
• (translation, homogeneity) Function convexity, concavity, quasiconvexity, and  
quasiconcavity are invariant to offset and nonnegative scaling.  

6. (affine transformation of argument) Composition $g(h(X))$ of convex (concave)  
function $g$ with any affine function $h : \mathbb{R}^m \times n \to \mathbb{R}^p \times k$ remains convex (concave) in  
$X \in \mathbb{R}^n \times n$, where $h(\mathbb{R}^m \times n) \cap \text{dom} \ g \neq \emptyset$. [209, §B.2.1] Likewise for the quasiconvex  
(quasiconcave) functions $g$.  

7. • – Nonnegatively weighted sum of convex (concave) functions remains convex  
(concave). (§3.1.2.2.1)  
– Nonnegatively weighted nonzero sum of strictly convex (concave) functions  
remains strictly convex (concave).  
– Nonnegatively weighted maximum (minimum) of convex $\text{3.29}$ (concave)  
functions remains convex (concave).  
– Pointwise supremum (infimum) of convex (concave) functions remains  
convex (concave). (Figure 76) [318, §5]  
• – Sum of quasiconvex functions not necessarily quasiconvex.  
– Nonnegatively weighted maximum (minimum) of quasiconvex  
(quasiconcave) functions remains quasiconvex (quasiconcave).  
– Pointwise supremum (infimum) of quasiconvex (quasiconcave) functions  
remains quasiconvex (quasiconcave).

---

3.28 Log-convex means: logarithm of function $f$ is convex on $\text{dom} \ f$.  
3.29 Supremum and maximum of convex functions are proven convex by intersection of epigraphs.
Chapter 4

Semidefinite programming

Prior to 1984, linear and nonlinear programming\textsuperscript{4,1} one a subset of the other, had evolved for the most part along unconnected paths, without even a common terminology. (The use of “programming” to mean “optimization” serves as a persistent reminder of these differences.)

– Forsgren, Gill, & Wright, 2002 [152]

Given some practical application of convex analysis, it may at first seem puzzling why a search for its solution ends abruptly with a formalized statement of the problem itself as a constrained optimization. The explanation is: typically we do not seek analytical solution because there are relatively few. (§3.5.2, §C) If a problem can be expressed in convex form, rather, then there exist computer programs providing efficient numerical global solution. [177] [416] [417] [415] [360] [346] The goal, then, becomes conversion of a given problem (perhaps a nonconvex or combinatorial problem statement) to an equivalent convex form or to an alternation of convex subproblems convergent to a solution of the original problem:

By the fundamental theorem of Convex Optimization, any locally optimal point (solution) of a convex problem is globally optimal. [63, §4.2.2] [317, §1] Given convex real objective function \( g \) and convex feasible set \( D \subseteq \text{dom} \ g \), which is the set of all variable values satisfying the problem constraints, we pose a generic convex optimization problem

\[
\begin{align*}
\text{minimize} & \quad g(X) \\
\text{subject to} & \quad X \in D
\end{align*}
\]

\textsuperscript{4,1}nascence of polynomial-time interior-point methods of solution [375] [413].

Linear programming \( \subseteq (\text{convex} \cap \text{nonlinear}) \) programming.
where constraints are abstract here in membership of variable $X$ to convex feasible set $\mathcal{D}$. Inequality constraint functions of a convex optimization problem are convex while equality constraint functions are conventionally affine, but not necessarily so. Affine equality constraint functions, as opposed to the superset of all convex equality constraint functions having convex level sets (§3.4.0.4), make convex optimization tractable.

Similarly, the problem

$$\begin{align*}
\text{maximize} & \quad g(X) \\
\text{subject to} & \quad X \in \mathcal{D}
\end{align*}$$

(679)

is called convex were $g$ a real concave function and feasible set $\mathcal{D}$ convex. As conversion to convex form is not always possible, there is much ongoing research to determine which problem classes have convex expression or relaxation. [35] [61] [159] [288] [356] [156]

4.1 Conic problem

Still, we are surprised to see the relatively small number of submissions to semidefinite programming (SDP) solvers, as this is an area of significant current interest to the optimization community. We speculate that semidefinite programming is simply experiencing the fate of most new areas: Users have yet to understand how to pose their problems as semidefinite programs, and the lack of support for SDP solvers in popular modelling languages likely discourages submissions.


(confer p.138) Consider a conic problem (p) and its dual (d): [304, §3.3.1] [249, §2.1] [250]

$$\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad x \in \mathcal{K} \\
Ax &= b
\end{align*}$$

(p) (301)

$$\begin{align*}
\text{maximize} & \quad b^T y \\
\text{subject to} & \quad s \in \mathcal{K}^* \\
A^T y + s &= c
\end{align*}$$

(d)

where $\mathcal{K}$ is a closed convex cone, $\mathcal{K}^*$ is its dual, matrix $A$ is fixed, and the remaining quantities are vectors.

When $\mathcal{K}$ is a polyhedral cone (§2.12.1), then each conic problem becomes a linear program; the selfdual nonnegative orthant providing the prototypical primal linear program and its dual. [97, §3-1] More generally, each optimization problem is convex when $\mathcal{K}$ is a closed convex cone. Solution to each convex problem is not necessarily unique; the optimal solution sets $\{x^*\}$ and $\{y^*, s^*\}$ are convex and may comprise more than a single point.

\footnote{Dantzig explains reasoning behind a nonnegativity constraint: \ldots negative quantities of activities are not possible. \ldots a negative number of cases cannot be shipped.}
4.1. CONIC PROBLEM

4.1.1 a semidefinite program

When \( K \) is the self-dual cone of positive semidefinite matrices \( \mathbb{S}_+^n \) in the subspace of symmetric matrices \( \mathbb{S}^n \), then each conic problem is called a semidefinite program (SDP); [288, §6.4] primal problem (P) having matrix variable \( X \in \mathbb{S}^n \) while corresponding dual (D) has slack variable \( S \in \mathbb{S}^n \) and vector variable \( y = [y_i] \in \mathbb{R}^m \): [11] [12, §2] [423, §1.3.8]

\[
\begin{align*}
\text{(P)} & \quad \text{minimize} \quad \langle C, X \rangle \\
& \quad \text{subject to} \quad X \succeq 0 \\
\end{align*}
\]

\[
\begin{align*}
\text{(D)} & \quad \text{maximize} \quad \langle b, y \rangle \\
& \quad \text{subject to} \quad S \succeq 0 \\
\end{align*}
\]

This is the prototypical semidefinite program and its dual, where matrix \( C \in \mathbb{S}^n \) and vector \( b \in \mathbb{R}^m \) are fixed as is

\[
A \svec X = b \quad \text{svec}^{-1}(A^T y) + S = C
\]

In a national planning problem of some size, one may easily run into several hundred variables and perhaps a hundred or more degrees of freedom. . . . It should always be remembered that any mathematical method and particularly methods in linear programming must be judged with reference to the type of computing machinery available. Our outlook may perhaps be changed when we get used to the super modern, high capacity electronic comptuer that will be available here from the middle of next year.

− Ragnar Frisch [154]
Figure 84: Venn diagram of programming hierarchy. Semidefinite program is a subset of convex program $\mathcal{PC}$. Semidefinite program subsumes other convex program classes excepting geometric program. Second-order cone program and quadratic program each subsume linear program. Nonconvex program $\mathcal{PC}$ comprises those for which convex equivalents have not yet been found.

in 1984, [285, §11] [413] [375] [288, p.3] but numerical performance of contemporary general-purpose semidefinite program solvers remains limited: Computational intensity for dense systems varies as $O(m^2n)$ ($m$ constraints $\ll n$ variables) based on interior-point methods that produce solutions no more relatively accurate than $1E-8$. There are no solvers capable of handling in excess of $n=100,000$ variables without significant, sometimes crippling, loss of precision or time.\(^4\,^3\) [36] [287, p.258] [69, p.3]

Nevertheless, semidefinite programming has recently emerged to prominence because it admits a new class of problem previously unsolvable by convex optimization techniques, [61] and because it theoretically subsumes other convex techniques: (Figure 84) linear programming and quadratic programming and second-order cone programming.\(^4\,^4\)

Determination of the Riemann mapping function from complex analysis [297] [30, §8, §13], for example, can be posed as a semidefinite program.

4.1.2 Maximal complementarity

It has been shown [423, §2.5.3] that contemporary interior-point methods [414] [300] [288] [12] [63, §11] [152] (developed circa 1990 [159]) for numerical solution of semidefinite

\(^4\,^3\)Heuristics are not ruled out by SIOPT; indeed I would suspect that most successful methods have (appropriately described) heuristics under the hood - my codes certainly do. . . . Of course, there are still questions relating to high-accuracy and speed, but for many applications a few digits of accuracy suffices and overnight runs for non-real-time delivery is acceptable.

—Nicholas I. M. Gould, Stanford alumnus, SIOPT Editor in Chief

\(^4\,^4\)Second-order cone programming was born in the 1990s; it is not posable as a quadratic program. [258]
4.1. CONIC PROBLEM

programs) can converge to a solution of maximal complementarity; \cite{186, §5} \cite{422} \cite{263} \cite{166} not a vertex solution but a solution of highest cardinality or rank among all optimal solutions.\textsuperscript{4.5}

This phenomenon can be explained by recognizing that interior-point methods generally find solutions relatively interior to a feasible set by design.\textsuperscript{4.6} Log barriers are designed to fail numerically at the feasible set boundary. So low-rank solutions, all on the boundary, are rendered more difficult to find as numerical error becomes more prevalent there.

4.1.2.1 Reduced-rank solution

A simple rank reduction algorithm, for construction of a primal optimal solution $X^*$ to (680P) satisfying an upper bound on rank governed by Proposition 2.9.3.0.1, is presented in §4.3. That proposition asserts existence of feasible solutions with an upper bound on their rank; \cite[§II.13.1]{27} specifically, it asserts an extreme point (§2.6.0.0.1) of primal feasible set $A \cap S^n_+$ satisfies upper bound

$$\text{rank } X \leq \left\lfloor \sqrt{\frac{8m+1}{2}} - 1 \right\rfloor$$

(272)

where, given $A \in \mathbb{R}^{m \times n(n+1)/2}$ and $b \in \mathbb{R}^m$,

$$A \triangleq \{ X \in S^n \mid A \text{svec } X = b \}$$

(683)

is the affine subset from primal problem (680P).

4.1.2.2 Coexistence of low- and high-rank solutions; analogy

That low-rank and high-rank optimal solutions $\{X^*\}$ of (680P) coexist may be grasped with the following analogy: We compare a proper polyhedral cone $S^3_+$ in $\mathbb{R}^3$ (illustrated in Figure 85) to the positive semidefinite cone $S^3_+$ in isometrically isomorphic $\mathbb{R}^6$, difficult to visualize. The analogy is good:

- int $S^3_+$ is constituted by rank-3 matrices.
- int $S^3_+$ has three dimensions.
- boundary $\partial S^3_+$ contains rank-0, rank-1, and rank-2 matrices.
- boundary $\partial S^3_+$ contains 0-, 1-, and 2-dimensional faces.
- the only rank-0 matrix resides in the vertex at the origin.
- Rank-1 matrices are in one-to-one correspondence with extreme directions of $S^3_+$ and $S^3_+$. The set of all rank-1 symmetric matrices in this dimension

$$\{ G \in S^3_+ \mid \text{rank } G = 1 \}$$

(684)

is not a connected set.

\textsuperscript{4.5}This characteristic might be regarded as a disadvantage to interior-point methods of numerical solution, but this behavior is not certain and depends on solver implementation.

\textsuperscript{4.6}Simplex methods, in contrast, find vertex solutions. \cite[§158]{97} \cite[p.2]{16}.
Figure 85: Visualizing positive semidefinite cone in high dimension: Proper polyhedral cone $S^3_+ \subset \mathbb{R}^3$ representing positive semidefinite cone $S^3_+ \subset S^3$; analogizing its intersection $S^3_+ \cap \partial \mathcal{H}$ with hyperplane. Number of facets is arbitrary (an analogy not inspired by eigenvalue decomposition). The rank-0 positive semidefinite matrix corresponds to origin in $\mathbb{R}^3$, rank-1 positive semidefinite matrices correspond to edges of polyhedral cone, rank-2 to facet relative interiors, and rank-3 to polyhedral cone interior. Vertices $\Gamma_1$ and $\Gamma_2$ are extreme points of polyhedron $\mathcal{P} = \partial \mathcal{H} \cap S^3_+$, and extreme directions of $S^3_+$. A given vector $C$ is normal to another hyperplane (not illustrated but independent w.r.t $\partial \mathcal{H}$) containing line segment $\overline{\Gamma_1 \Gamma_2}$ minimizing real linear function $\langle C, X \rangle$ on $\mathcal{P}$. (confer Figure 28, Figure 32)
4.1. CONIC PROBLEM

- Rank of a sum of members \( F + G \) in Lemma 2.9.2.9.1 and location of a difference \( F - G \) in §2.9.2.12.1 similarly hold for \( S^3_+ \) and \( S^3_{++} \).

- Euclidean distance from any particular rank-3 positive semidefinite matrix (in the cone interior) to the closest rank-2 positive semidefinite matrix (on the boundary) is generally less than the distance to the closest rank-1 positive semidefinite matrix. (§7.1.2)

- distance from any point in \( \partial S^3_+ \) to \( \text{int} S^3_+ \) is infinitesimal (§2.1.7.1.1).

- faces of \( S^3_+ \) correspond to faces of \( S^3_{++} \) (confer Table 2.9.2.3.1):

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \dim \mathcal{F}(S^3_+) )</th>
<th>( \dim \mathcal{F}(S^3_{++}) )</th>
<th>( \dim \mathcal{F}(S^3_{++} \ni \text{rank-k matrix}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>boundary</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>interior</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Integer \( k \) indexes \( k \)-dimensional faces \( \mathcal{F} \) of \( S^3_+ \). Positive semidefinite cone \( S^3_+ \) has four kinds of faces, including cone itself \( (k=3, \text{boundary } + \text{interior}) \), whose dimensions in isometrically isomorphic \( \mathbb{R}^6 \) are listed under \( \dim \mathcal{F}(S^3_+) \). Smallest face \( \mathcal{F}(S^3_+ \ni \text{rank-k matrix}) \) that contains a rank-\( k \) positive semidefinite matrix has dimension \( k(k+1)/2 \) by (222).

- For \( \mathcal{A} \) equal to intersection of \( m \) hyperplanes having linearly independent normals, and for \( X \in S^3_+ \cap \mathcal{A} \), we have \( \text{rank } X \leq m \); the analogue to (272).

**Proof.** With reference to Figure 85: Assume one \( (m=1) \) hyperplane \( \mathcal{A} = \partial \mathcal{H} \) intersects the polyhedral cone. Every intersecting plane contains at least one matrix having rank less than or equal to 1; id est, from all \( X \in \partial \mathcal{H} \cap S^2_+ \) there exists an \( X \) such that \( \text{rank } X \leq 1 \). Rank 1 is therefore an upper bound in this case.

Now visualize intersection of the polyhedral cone with two \( (m=2) \) hyperplanes having linearly independent normals. The hyperplane intersection \( \mathcal{A} \) makes a line. Every intersecting line contains at least one matrix having rank less than or equal to 2, providing an upper bound. In other words, there exists a positive semidefinite matrix \( X \) belonging to any line intersecting the polyhedral cone such that \( \text{rank } X \leq 2 \).

In the case of three independent intersecting hyperplanes \( (m=3) \), the hyperplane intersection \( \mathcal{A} \) makes a point that can reside anywhere in the polyhedral cone. The upper bound on a point in \( S^3_+ \) is also the greatest upper bound: \( \text{rank } X \leq 3 \). ♦
4.1.2.2.1 Example. Optimization over $A \cap S^3_+$.  
Consider minimization of the real linear function $\langle C, X \rangle$ over 
$$P \triangleq A \cap S^3_+$$ (685) 
an a polyhedral feasible set; 
$$f_0^* \triangleq \min_X \langle C, X \rangle \quad \text{subject to} \quad X \in A \cap S^3_+$$ (686) 
As illustrated for particular vector $C$ and hyperplane $A = \partial H$ in Figure 85, this linear function is minimized on any $X$ belonging to the face of $P$ containing extreme points $\{\Gamma_1, \Gamma_2\}$ and all the rank-2 matrices in between; id est, on any $X$ belonging to the face $F(P) = \{X \mid \langle C, X \rangle = f_0^*\}$ of $P$ exposed by the hyperplane $\{X \mid \langle C, X \rangle = f_0^*\}$. In other words, the set of all optimal points $X^*$ is a face of $P$ 
$$\{X^*\} = F(P) = \Gamma_1 \Gamma_2$$ (688) 
comprising rank-1 and rank-2 positive semidefinite matrices. Rank 1 is the upper bound on existence in the feasible set $P$ for this case $m = 1$ hyperplane constituting $A$. The rank-1 matrices $\Gamma_1$ and $\Gamma_2$ in face $F(P)$ are extreme points of that face and (by transitivity (§2.6.1.2)) extreme points of the intersection $P$ as well. As predicted by analogy to Barvinok’s Proposition 2.9.3.0.1, the upper bound on rank of $X^*$ existent in $F(P)$ is thereby also satisfied by an extreme point of $P$ precisely because $\{X^*\}$ constitutes $F(P)$:4.7 
$$\{X^* \in P \mid \text{rank } X^* \leq 1\} = \{\Gamma_1, \Gamma_2\} \subseteq F(P)$$ (689) 
As all linear functions on a polyhedron are minimized on a face, [97] [262] [284] [291] by analogy we so demonstrate coexistence of optimal solutions $X^*$ of (680P) having assorted rank.

4.1.2.3 Previous work

Barvinok showed, [25, §2.2] when given a positive definite matrix $C$ and an arbitrarily small neighborhood of $C$ comprising positive definite matrices, there exists a matrix $\tilde{C}$ from that neighborhood such that optimal solution $X^*$ to (680P) (substituting $\tilde{C}$) is an extreme point of $A \cap S^n_+$ and satisfies upper bound (272).4.8 Given arbitrary positive definite $C$, this means nothing inherently guarantees that an optimal solution $X^*$ to problem (680P) satisfies (272); certainly nothing given any symmetric matrix $C$, as the problem is posed. This can be proved by example:

4.7 and every face contains a subset of the extreme points of $P$ by the extreme existence theorem (§2.6.0.0.2). This means: because the affine subset $A$ and hyperplane $\{X \mid \langle C, X \rangle = f_0^*\}$ must intersect a whole face of $P$, calculation of an upper bound on rank of $X^*$ ignores counting the hyperplane when determining $m$ in (272).

4.8 Further, the set of all such $\tilde{C}$ in that neighborhood is open and dense.
4.1.2.3.1 **Example.** (Ye) **Maximal Complementarity.**

Assume dimension \( n \) to be an even positive number. Then the particular instance of problem \((680P)\),

\[
\begin{align*}
\text{minimize} & \quad \langle I \ 0 \ 2I \ , \ X \rangle \\
\text{subject to} & \quad X \succeq 0 \\
& \quad \langle I \ , \ X \rangle = n
\end{align*}
\]

has optimal solution

\[
X^* = \begin{bmatrix} 2I & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{S}^n
\]

with an equal number of twos and zeros along the main diagonal. Indeed, optimal solution \((691)\) is a terminal solution along the central path taken by the interior-point method as implemented in [423, §2.5.3]; it is also a solution of highest rank among all optimal solutions to \((690)\). Clearly, rank of this primal optimal solution exceeds by far a rank-1 solution predicted by upper bound \((272)\).

4.1.2.4 **Later developments**

This rational example \((690)\) indicates the need for a more generally applicable and simple algorithm to identify an optimal solution \(X^*\) satisfying Barvinok’s Proposition 2.9.3.0.1. We will review such an algorithm in §4.3, but first we provide more background.

4.2 **Framework**

4.2.1 **Feasible sets**

Denote by \( \mathcal{D} \) and \( \mathcal{D}^* \) the convex sets of primal and dual points respectively satisfying the primal and dual constraints in \((680)\), each assumed nonempty;

\[
\mathcal{D} = \left\{ X \in \mathbb{S}^n_+ \mid \begin{bmatrix} \langle A_1 , X \rangle \\ \vdots \\ \langle A_m , X \rangle \end{bmatrix} = b \right\} = \mathcal{A} \cap \mathbb{S}^n_+
\]

\[
\mathcal{D}^* = \left\{ S \in \mathbb{S}^n_+ , \ y = [y_i] \in \mathbb{R}^m \mid \sum_{i=1}^m y_i A_i + S = C \right\}
\]

These are the **primal feasible set** and **dual feasible set**. Geometrically, primal feasible \( \mathcal{A} \cap \mathbb{S}^n_+ \) represents an intersection of the positive semidefinite cone \( \mathbb{S}^n_+ \) with an affine subset \( \mathcal{A} \) of the subspace of symmetric matrices \( \mathbb{S}^n \) in isometrically isomorphic \( \mathbb{R}^{n(n+1)/2} \). \( \mathcal{A} \) has dimension \( n(n+1)/2 - m \) when the vectorized \( A_i \) are linearly independent. Dual feasible set \( \mathcal{D}^* \) is a Cartesian product of the positive semidefinite cone with its inverse image \((\S.1.9.0.1)\) under affine transformation \( C - \sum y_i A_i \). Both feasible sets are convex, and

\[
\sum y_i A_i \leq C \quad \text{follows directly from \((680D)(\S.9.0.1.1)\) and is known as a linear matrix inequality. \((\S.13.5.1.1)\) Because \( \sum y_i A_i \succeq C \), matrix \( S \) is known as a slack variable (a term borrowed from linear programming [97]) since its inclusion raises this inequality to equality.}
the objective functions are linear on a Euclidean vector space. Hence, (680P) and (680D) are convex optimization problems.

4.2.1.1 \( A \cap S_n^+ \) emptiness determination via Farkas’ lemma

4.2.1.1 Lemma. Semidefinite Farkas’ lemma.
Given set \( \{ A_i \in S^n, i = 1 \ldots m \} \), vector \( b = [b_i] \in \mathbb{R}^m \), and affine subset

\[
(683) \ A = \{ X \in S^n \mid \langle A_i, X \rangle = b_i, i = 1 \ldots m \} \supset \{ A \text{svec } X \mid X \succeq 0 \} \tag{379}
\]

is closed, then primal feasible set \( A \cap S_n^+ \) is nonempty if and only if \( y^T b \geq 0 \) holds for each and every vector \( y = [y_i] \in \mathbb{R}^m \) such that \( \sum_{i=1}^m y_i A_i \succeq 0 \).

Equivalently, primal feasible set \( A \cap S_n^+ \) is nonempty if and only if \( y^T b \geq 0 \) holds for each and every vector \( ||y|| = 1 \) such that \( \sum_{i=1}^m y_i A_i \succeq 0 \).

Semidefinite Farkas’ lemma provides necessary and sufficient conditions for a set of hyperplanes to have nonempty intersection \( A \cap S_n^+ \) with the positive semidefinite cone. Given

\[
A = \begin{bmatrix}
\text{svec}(A_1)^T \\
\vdots \\
\text{svec}(A_m)^T
\end{bmatrix} \in \mathbb{R}^{m \times (n+1)/2} \tag{681}
\]

semidefinite Farkas’ lemma assumes that a convex cone

\[
K = \{ A \text{svec } X \mid X \succeq 0 \} \tag{379}
\]

is closed per membership relation (319) from which the lemma springs: \[242, \S1\] \( K \) closure is attained when matrix \( A \) satisfies the cone closedness invariance corollary (p.157). Given closed convex cone \( K \) and its dual from Example 2.13.5.1.1

\[
K^* = \{ y \mid \sum_{j=1}^m y_j A_j \succeq 0 \} \tag{386}
\]

then we can apply membership relation

\[
b \in K \iff \langle y, b \rangle \geq 0 \ \forall \ y \in K^* \tag{319}
\]

to obtain the lemma

\[
b \in K \iff \exists X \succeq 0 \supset A \text{svec } X = b \iff A \cap S_n^+ \neq \emptyset \tag{693}
\]
\[
b \in K \iff \langle y, b \rangle \geq 0 \ \forall \ y \in K^* \iff A \cap S_n^+ \neq \emptyset \tag{694}
\]

The final equivalence synopsizes semidefinite Farkas’ lemma.

While the lemma is correct as stated, a positive definite version is required for semidefinite programming \[423, \S1.3.8\] because existence of a feasible solution in the cone
interior $\mathcal{A} \cap \text{int} \mathbb{S}^n$ is required by Slater’s condition\textsuperscript{4.10} to achieve 0 duality gap (optimal primal–dual objective difference §4.2.3, Figure 60). Geometrically, a positive definite lemma is required to insure that a point of intersection closest to the origin is not at infinity; e.g., Figure 47. Then given $A \in \mathbb{R}^{m \times n(n+1)/2}$ having rank $m$, we wish to detect existence of nonempty primal feasible set interior to the PSD cone;\textsuperscript{4.11} (382)

$$b \in \text{int} \mathcal{K} \iff \langle y, b \rangle > 0 \quad \forall y \in \mathcal{K}^*, \ y \neq 0 \quad \iff \mathcal{A} \cap \text{int} \mathbb{S}^n_+ \neq \emptyset \quad (695)$$

Positive definite Farkas’ lemma is made from proper cones, $\mathcal{K}$ (379) and $\mathcal{K}^*$ (386), and membership relation (325) for which $\mathcal{K}$ closedness is unnecessary:

4.2.1.1.2 Lemma. Positive definite Farkas’ lemma.
Given l.i. set \{ $A_i \in \mathbb{S}^n$, $i = 1 \ldots m$ \} and vector $b = [b_i] \in \mathbb{R}^m$, make affine set

$$\mathcal{A} = \{ X \in \mathbb{S}^n | \langle A_i, X \rangle = b_i, \ i = 1 \ldots m \} \quad (683)$$

Primal feasible cone interior $\mathcal{A} \cap \text{int} \mathbb{S}^n_+$ is nonempty if and only if $y^T b > 0$ holds for each and every vector $y = [y_i] \neq 0$ such that $\sum_{i=1}^m y_i A_i \succeq 0$.

Equivalently, primal feasible cone interior $\mathcal{A} \cap \text{int} \mathbb{S}^n_+$ is nonempty if and only if $y^T b > 0$ holds for each and every vector $\| y \| = 1 \Rightarrow \sum_{i=1}^m y_i A_i \succeq 0$. $\diamond$

4.2.1.1.3 Example. “New” Farkas’ lemma.
Lasserre [242, §III] presented an example in 1995, originally offered by Ben-Israel in 1969 [33, p.378], to support closedness in semidefinite Farkas’ Lemma 4.2.1.1.1:

$$A \triangleq \begin{bmatrix} \text{svec}(A_1)^T \\ \text{svec}(A_2)^T \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (696)$$

Intersection $\mathcal{A} \cap \mathbb{S}^n_+$ is practically empty because the solution set

$$\{ X \succeq 0 | A \text{ svec} \ X = b \} = \left\{ \left[ \begin{array}{c} \alpha \\ \frac{1}{\sqrt{2}} \end{array} \right] \succeq 0 | \alpha \in \mathbb{R} \right\} \quad (697)$$

is positive semidefinite only asymptotically ($\alpha \to \infty$). Yet $\sum_{i=1}^m y_i A_i \succeq 0 \Rightarrow y^T b \succeq 0$ the dual system erroneously indicates nonempty intersection because $\mathcal{K}$ (379) violates a closedness condition of the lemma; \textit{videlicet}, for $\| y \| = 1$

$$y_1 \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + y_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \succeq 0 \Leftrightarrow y = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow y^T b = 0 \quad (698)$$

\textsuperscript{4.10}Slater’s sufficient constraint qualification is satisfied whenever any primal or dual strictly feasible solution exists; \textit{id est}, any point satisfying the respective affine constraints and relatively interior to the convex cone. [340, p.66] [42, p.325] If the cone were polyhedral, then Slater’s constraint qualification is satisfied when any feasible solution exists (relatively interior to the cone or on its relative boundary).

\textsuperscript{4.11}Detection of $\mathcal{A} \cap \text{int} \mathbb{S}^n_+ \neq \emptyset$ by examining $\text{int} \mathcal{K}$ instead is a trick need not be lost.
On the other hand, positive definite Farkas’ Lemma 4.2.1.1.2 certifies that $\mathcal{A} \cap \text{int} \mathbb{S}^n_+$ is empty; what we need to know for semidefinite programming.

Lasserre suggested addition of another condition to semidefinite Farkas’ lemma (§4.2.1.1.1) to make a new lemma having no closedness condition. But positive definite Farkas’ lemma (§4.2.1.1.2) is simpler and obviates the additional condition proposed.

4.2.1.2 Theorem of the alternative for semidefinite programming

Because these Farkas’ lemmas follow from membership relations, we may construct alternative systems from them. Applying the method of §2.13.2.1.1, then from positive definite Farkas’ lemma we get

$$\mathcal{A} \cap \text{int} \mathbb{S}^n_+ \neq \emptyset$$

or in the alternative

$$y^T b \leq 0, \quad \sum_{i=1}^{m} y_i A_i \succeq 0, \quad y \neq 0$$

Any single vector $y$ satisfying the alternative certifies $\mathcal{A} \cap \text{int} \mathbb{S}^n_+$ is empty. Such a vector can be found as a solution to another semidefinite program: for linearly independent (vectorized) set $\{A_i \in \mathbb{S}^n, i=1\ldots m\}$

$$\begin{align*}
\text{minimize} & \quad y^T b \\
\text{subject to} & \quad \sum_{i=1}^{m} y_i A_i \succeq 0 \\
& \quad \|y\|^2 \leq 1
\end{align*}$$

(699)

If an optimal vector $y^* \neq 0$ can be found such that $y^{*T} b \leq 0$, then primal feasible cone interior $\mathcal{A} \cap \text{int} \mathbb{S}^n_+$ is empty.

4.2.1.3 Boundary-membership criterion

(confer (694)(695)) From boundary-membership relation (329), for proper cones $\mathcal{K}$ (379) and $\mathcal{K}^*$ (386) of linear matrix inequality,

$$b \in \partial \mathcal{K} \iff \exists y \neq 0 \text{ s.t. } \langle y, b \rangle = 0, \quad y \in \mathcal{K}^*, \quad b \in \mathcal{K} \iff \partial \mathbb{S}^n_+ \supset \mathcal{A} \cap \mathbb{S}^n_+ \neq \emptyset$$

(701)

Whether vector $b \in \partial \mathcal{K}$ belongs to cone $\mathcal{K}$ boundary, that is a determination we can indeed make: one that is certainly expressible as a feasibility problem: Given linearly independent set $\{A_i \in \mathbb{S}^n, i=1\ldots m\}$, for $b \in \mathcal{K}$ (693)

$$\begin{align*}
\text{find} & \quad y \neq 0 \\
\text{subject to} & \quad y^T b = 0 \\
& \quad \sum_{i=1}^{m} y_i A_i \succeq 0
\end{align*}$$

(702)

\footnote{From the results of Example 2.13.5.1.1, vector $b$ on the boundary of $\mathcal{K}$ cannot be detected simply by looking for 0 eigenvalues in matrix $X$. We do not consider a skinny-or-square matrix $A$ because then feasible set $\mathcal{A} \cap \mathbb{S}^n_+$ is at most a single point.}
Any such nonzero solution \( y \) certifies that affine subset \( A \) intersects the positive semidefinite cone \( \mathcal{S}_+^n \) only on its boundary; in other words, nonempty feasible set \( A \cap \mathcal{S}_+^n \) belongs to the positive semidefinite cone boundary \( \partial \mathcal{S}_+^n \).

### 4.2.2 Duals

The dual objective function from (680D) evaluated at any feasible solution represents a lower bound on the primal optimal objective value from (680P). We can see this by direct substitution: Assume the feasible sets \( A \cap \mathcal{S}_+^n \) and \( D^* \) are nonempty. Then it is always true:

\[
\langle C, X \rangle \geq \langle b, y \rangle \\
\left\langle \sum_i y_i A_i + S, X \right\rangle \geq \left[ \langle A_1, X \rangle \cdots \langle A_m, X \rangle \right] y \\
\langle S, X \rangle \geq 0
\]

The converse also follows because

\[
X \succeq 0, \ S \succeq 0 \Rightarrow \langle S, X \rangle \geq 0 \quad (1567)
\]

Optimal value of the dual objective thus represents the greatest lower bound on the primal. This fact is known as the weak duality theorem for semidefinite programming, [423, §1.3.8] and can be used to detect convergence in any primal/dual numerical method of solution.

#### 4.2.2.1 Dual problem statement is not unique

Even subtle but equivalent restatements of a primal convex problem can lead to vastly different statements of a corresponding dual problem. This phenomenon is of interest because a particular instantiation of dual problem might be easier to solve numerically or it might take one of few forms for which analytical solution is known.

Here is a canonical restatement of prototypical dual semidefinite program (680D), for example, equivalent by (194):

\[
\text{(D)} \quad \begin{array}{ll}
\text{maximize} & \langle b, y \rangle \\
\text{subject to} & S \succeq 0 \\
& \text{svec}^{-1}(A^T y) + S = C
\end{array}
\]

\[
\text{(680\tilde{D})} \quad \begin{array}{ll}
\text{maximize} & \langle b, y \rangle \\
\text{subject to} & \text{svec}^{-1}(A^T y) \preceq C
\end{array}
\]

Dual feasible cone interior in \( \text{int} \mathcal{S}_+^n \) (692) (682) thereby corresponds with canonical dual (\( \tilde{D} \)) feasible interior

\[
\text{rel int} \tilde{D}^* \triangleq \left\{ y \in \mathbb{R}^m \mid \sum_{i=1}^m y_i A_i \prec C \right\} \quad (704)
\]

#### 4.2.2.1 Exercise. Primal prototypical semidefinite program.

Derive prototypical primal (680P) from its canonical dual (680D); \( \text{id est} \), demonstrate that particular connectivity in Figure 86.
CHAPTER 4. SEMIDEFINITE PROGRAMMING

Figure 86: Connectivity indicates paths between particular primal and dual problems from Exercise 4.2.2.1.1. More generally, any path between primal problems $\mathbf{P}$ (and equivalent $\tilde{\mathbf{P}}$) and dual $\mathbf{D}$ (and equivalent $\tilde{\mathbf{D}}$) is possible: implying, any given path is not necessarily circuital; dual of a dual problem is not necessarily stated in precisely same manner as corresponding primal convex problem, in other words, although its solution set is equivalent to within some transformation.

4.2.3 Optimality conditions

When primal feasible cone interior $\mathcal{A} \cap \text{int} \mathbb{S}_+^n$ exists in $\mathbb{S}^n$ or when canonical dual feasible interior $\text{rel int} \tilde{\mathcal{D}}^*$ exists in $\mathbb{R}^m$, then these two problems $(680\mathbf{P}) (680\mathbf{D})$ become strong duals by Slater’s sufficient condition (p. 247). In other words, the primal optimal objective value becomes equal to the dual optimal objective value: there is no duality gap (Figure 60) and so determination of convergence is facilitated; id est, if $\exists X \in \mathcal{A} \cap \text{int} \mathbb{S}_+^n$ or $\exists y \in \text{rel int} \tilde{\mathcal{D}}^*$ then

$$\langle C, X^* \rangle = \langle b, y^* \rangle$$
$$\left\langle \sum_i y_i^* A_i + S^*, X^* \right\rangle = \left[ \langle A_1, X^* \rangle \cdots \langle A_m, X^* \rangle \right] y^*$$

(705)

where $S^*, y^*$ denote a dual optimal solution.\textsuperscript{4.13} We summarize this:

4.2.3.0.1 Corollary. Optimality and strong duality. [371, §3.1] [423, §1.3.8]

For semidefinite programs $(680\mathbf{P})$ and $(680\mathbf{D})$, assume primal and dual feasible sets $\mathcal{A} \cap \mathbb{S}_+^n \subset \mathbb{S}^n$ and $\mathcal{D}^* \subset \mathbb{S}^n \times \mathbb{R}^m (692)$ are nonempty. Then

- $X^*$ is optimal for $(680\mathbf{P})$
- $S^*, y^*$ are optimal for $(680\mathbf{D})$
- duality gap $\langle C, X^* \rangle - \langle b, y^* \rangle$ is 0

\textsuperscript{4.13}Optimality condition $\langle S^*, X^* \rangle = 0$ is called a \textit{complementary slackness condition}, in keeping with linear programming tradition [97], that forbids dual inequalities in $(680)$ to simultaneously hold strictly. [317, §4]
if and only if

\begin{align*}
\text{i) } & \exists X \in A \cap \text{int } \mathbb{S}_+^n \quad \text{or} \quad \exists y \in \text{rel int } \tilde{D}^* \\
\text{and} & \\
\text{ii) } & \langle S^*, X^* \rangle = 0
\end{align*}

For symmetric positive semidefinite matrices, requirement ii is equivalent to the complementarity (§A.7.4)

\[ \langle S^*, X^* \rangle = 0 \iff S^* X^* = X^* S^* = 0 \] (706)

Commutativity of diagonalizable matrices is a necessary and sufficient condition [212, §1.3.12] for these two optimal symmetric matrices to be simultaneously diagonalizable. Therefore

\[ \text{rank } X^* + \text{rank } S^* \leq n \] (707)

**Proof.** To see that, the product of symmetric optimal matrices $X^*, S^* \in \mathbb{S}_n^+$ must itself be symmetric because of commutativity. (1556) The symmetric product has diagonalization [12, cor.2.11]

\[ S^* X^* = X^* S^* = Q \Lambda_S \Lambda_X Q^T = 0 \iff \Lambda_X \Lambda_S = 0 \] (708)

where $Q$ is an orthogonal matrix. The product of the nonnegative diagonal $\Lambda$ matrices can be $0$ if their main diagonal zeros are complementary or coincide. Due only to symmetry, $\text{rank } X^* = \text{rank } \Lambda_X$ and $\text{rank } S^* = \text{rank } \Lambda_S$ for these optimal primal and dual solutions. (1542) So, because of the complementarity, the total number of nonzero diagonal entries from both $\Lambda$ cannot exceed $n$.

When equality is attained in (707)

\[ \text{rank } X^* + \text{rank } S^* = n \] (709)

there are no coinciding main diagonal zeros in $\Lambda_X \Lambda_S$, and so we have what is called **strict complementarity**.\(^{4,14}\) Logically it follows that a necessary and sufficient condition for strict complementarity of an optimal primal and dual solution is

\[ X^* + S^* \succ 0 \] (710)

4.2.3.1 **solving primal problem via dual**

The beauty of Corollary 4.2.3.0.1 is its conjugacy; id est, one can solve either the primal or dual problem in (680) and then find a solution to the other via the optimality conditions. When a dual optimal solution is known, for example, a primal optimal solution is any primal feasible solution in hyperplane \( \{ X \mid \langle S^*, X \rangle = 0 \} \).

\(^{4,14}\) distinct from maximal complementarity (§4.1.2).
4.2.3.1.1 Example. Minimal cardinality Boolean. 

Consider finding a minimal cardinality Boolean solution $x$ to the classic linear algebra problem $Ax = b$ given noiseless data $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$:

$$
\begin{align*}
\text{minimize} & \quad \|x\|_0 \\
\text{subject to} & \quad Ax = b \\
& \quad x_i \in \{0, 1\}, \quad i = 1 \ldots n
\end{align*}
$$

where $\|x\|_0$ denotes cardinality of vector $x$ (a.k.a 0-norm; not a convex function).

A minimal cardinality solution answers the question: “Which fewest linear combination of columns in $A$ constructs vector $b$?” Cardinality problems have extraordinarily wide appeal, arising in many fields of science and across many disciplines. Yet designing an efficient algorithm to optimize cardinality has proved difficult. In this example, we also constrain the variable to be Boolean. The Boolean constraint forces an identical solution were the norm in problem (711) instead the 1-norm or 2-norm; id est, the two problems

$$
\begin{align*}
\text{minimize} & \quad \|x\|_0 \\
\text{subject to} & \quad Ax = b \\
& \quad x_i \in \{0, 1\}, \quad i = 1 \ldots n \\
\end{align*}
$$

$$
\begin{align*}
\text{minimize} & \quad \|x\|_1 \\
\text{subject to} & \quad Ax = b \\
& \quad x_i \in \{0, 1\}, \quad i = 1 \ldots n
\end{align*}
$$

are the same. The Boolean constraint makes the 1-norm problem nonconvex.

Given data

$$
A = \begin{bmatrix}
-1 & 1 & 8 & 1 & 1 & 0 \\
-3 & 2 & 8 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\
-9 & 4 & 8 & \frac{1}{4} & \frac{1}{9} & \frac{1}{16}
\end{bmatrix}, \quad b = \begin{bmatrix}
1 \\
\frac{1}{2} \\
\frac{1}{4}
\end{bmatrix}
$$

the obvious and desired solution to the problem posed,

$$
x^* = e_4 \in \mathbb{R}^6
$$

has norm $\|x^*\|_2 = 1$ and minimal cardinality; the minimum number of nonzero entries in vector $x$. The MATLAB backslash command $x = A \backslash b$, for example, finds

$$
x_M = \begin{bmatrix}
\frac{2}{128} \\
0 \\
\frac{5}{128} \\
0 \\
\frac{90}{128} \\
0
\end{bmatrix}
$$

having norm $\|x_M\|_2 = 0.7044$. Coincidentally, $x_M$ is a 1-norm solution; id est, an optimal solution to

$$
\begin{align*}
\text{minimize} & \quad \|x\|_1 \\
\text{subject to} & \quad Ax = b
\end{align*}
$$
4.2. FRAMEWORK

The pseudoinverse solution (rounded)

\[ x_p = A^+ b = \begin{bmatrix} -0.0456 \\ -0.1881 \\ 0.0623 \\ 0.2668 \\ 0.3770 \\ -0.1102 \end{bmatrix} \]  \hspace{1cm} (716)

has least norm \( \|x_p\|_2 = 0.5165 \); \textit{id est}, the optimal solution to (E.0.1.0.1)

\[
\begin{align*}
\text{minimize} & \quad \|x\|_2 \\
\text{subject to} & \quad Ax = b
\end{align*}
\]  \hspace{1cm} (717)

Certainly none of the traditional methods provide \( x^* = e_4 \) (714) because, and in general, for \( Ax = b \)

\[
\|\operatorname{arg inf} \|x\|_2\|_2 \leq \|\operatorname{arg inf} \|x\|_1\|_2 \leq \|\operatorname{arg inf} \|x\|_0\|_2
\]  \hspace{1cm} (718)

We can reformulate this minimal cardinality Boolean problem (711) as a semidefinite program: First transform the variable

\[
x \triangleq (\hat{x} + 1) \frac{1}{2}
\]  \hspace{1cm} (719)

so \( \hat{x}_i \in \{-1, 1\} \); equivalently,

\[
\begin{align*}
\text{minimize} & \quad \|\hat{x} + 1\|_0 \\
\text{subject to} & \quad A(\hat{x} + 1) \frac{1}{2} = b \\
& \quad \delta(\hat{x}\hat{x}^T) = 1
\end{align*}
\]  \hspace{1cm} (720)

where \( \delta \) is the main-diagonal linear operator (§A.1). By assigning (§B.1)

\[
G = \begin{bmatrix} \hat{x} & 1 \\ \hat{x}^T & 1 \end{bmatrix} = \begin{bmatrix} X & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \triangleq \begin{bmatrix} \hat{x}\hat{x}^T & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \in \mathbb{S}^{n+1}
\]  \hspace{1cm} (721)

problem (720) becomes equivalent to: (Theorem A.3.1.0.7)

\[
\begin{align*}
\text{minimize}_{\hat{x} \in \mathbb{S}^n, \hat{x} \in \mathbb{R}^n} & \quad 1^T \hat{x} \\
\text{subject to} & \quad A(\hat{x} + 1) \frac{1}{2} = b \\
& \quad G = \begin{bmatrix} X & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} (\succeq 0) \\
& \quad \delta(X) = 1 \\
& \quad \text{rank} G = 1
\end{align*}
\]  \hspace{1cm} (722)

where solution is confined to rank-1 vertices of the \textit{elliptope} in \( \mathbb{S}^{n+1} \) (§5.9.1.0.1) by the rank constraint, the positive semidefiniteness, and the equality constraints \( \delta(X) = 1 \). The
rank constraint makes this problem nonconvex; by removing it we get the semidefinite program

$$\begin{align*}
\text{minimize} & \quad \mathbf{1}^T \hat{x} \\
\text{subject to} & \quad A(\hat{x} + 1)\frac{1}{2} = b \\
& \quad G = \begin{bmatrix}
X & \hat{x} \\
\hat{x}^T & 1
\end{bmatrix} \succeq 0 \\
& \quad \delta(X) = 1
\end{align*}$$

(723)

whose optimal solution \( x^\star (719) \) is identical to that of minimal cardinality Boolean problem (711) if and only if rank \( G^\star = 1 \).

Hope of acquiring a rank-1 solution is not ill-founded because \( 2^n \) ellipope vertices have rank 1, and we are minimizing an affine function on a subset of the ellipope (Figure 142) containing rank-1 vertices; \textit{id est}, by assumption that the feasible set of minimal cardinality Boolean problem (711) is nonempty, a desired solution resides on the ellipope relative boundary at a rank-1 vertex.

For that data given in (713), our semidefinite program solver \texttt{sdpsol} [416] [417] (accurate in solution to approximately \( 1E-8 \)) finds optimal solution to (723)

$$\text{round}(G^\star) = \begin{bmatrix}
1 & 1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 & 1 & 1 & -1 \\
-1 & -1 & -1 & 1 & -1 & -1 & 1 \\
1 & 1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 & 1 & 1 & -1 \\
-1 & -1 & -1 & 1 & -1 & -1 & 1
\end{bmatrix}$$

(724)

near a rank-1 vertex of the ellipope in \( \mathbb{S}^{n+1} \) (Theorem 5.9.1.0.2); its sorted eigenvalues,

$$\lambda(G^\star) = \begin{bmatrix}
6.99999977799099 \\
0.00000022687241 \\
0.00000002250296 \\
0.00000002250296 \\
-0.00000000999738 \\
-0.00000000999875 \\
-0.00000001000000
\end{bmatrix}$$

(725)

Negative eigenvalues are undoubtedly finite-precision effects. Because the largest eigenvalue predominates by many orders of magnitude, we can expect to find a good
approximation to a minimal cardinality Boolean solution by truncating all smaller eigenvalues. We find, indeed, the desired result (714)

\[ x^+ = \text{round} \begin{pmatrix}
0.00000000127947 \\
0.00000000527369 \\
0.00000000181001 \\
0.9999999460044 \\
0.00000001408950 \\
0.00000000482903
\end{pmatrix} = e_4 \quad (726) \]

These numerical results are solver dependent; insofar, not all SDP solvers will return a rank-1 vertex solution.

### 4.2.3.1.2 Example. **Optimization over elliptope versus 1-norm polyhedron for minimal cardinality Boolean Example 4.2.3.1.1.**

A minimal cardinality problem is typically formulated via, what is by now, a standard practice [127] [70, §3.2, §3.4] of column normalization applied to a 1-norm problem surrogate like (517). Suppose we define a diagonal matrix

\[ \Lambda \triangleq \begin{bmatrix}
\|A(:, 1)\|_2 & 0 \\
\|A(:, 2)\|_2 & \ddots \\
0 & \|A(:, 6)\|_2
\end{bmatrix} \in \mathbb{S}^6 \quad (727) \]

used to normalize the columns (assumed nonzero) of given noiseless data matrix \( A \). Then approximate the minimal cardinality Boolean problem

\[
\begin{align*}
\text{minimize} & \quad \|x\|_0 \\
\text{subject to} & \quad Ax = b \\
& \quad x_i \in \{0, 1\}, \quad i = 1 \ldots n \\
\end{align*}
\]

as

\[
\begin{align*}
\text{minimize} & \quad \|\tilde{y}\|_1 \\
\text{subject to} & \quad A\Lambda^{-1}\tilde{y} = b \\
& \quad 1 \succeq \Lambda^{-1}\tilde{y} \succeq 0
\end{align*}
\]

where optimal solution

\[ y^* = \text{round}(\Lambda^{-1}\tilde{y}^*) \quad (729) \]

The inequality in (728) relaxes Boolean constraint \( y_i \in \{0, 1\} \) from (711); bounding any solution \( y^* \) to a nonnegative unit hypercube whose vertices are binary numbers. Convex problem (728) is justified by the convex envelope

\[
\text{cenv} \|x\|_0 \text{ on } \{ x \in \mathbb{R}^n \mid \|x\|_\infty \leq \kappa \} = \frac{1}{\kappa} \|x\|_1 \quad (1448)
\]

Donoho concurs with this particular formulation, equivalently expressible as a linear program via (513).
Approximation (728) is therefore equivalent to minimization of an affine function (§3.2) on a bounded polyhedron, whereas semidefinite program

\[
\begin{align*}
\text{minimize} & \quad 1^T \hat{x} \\
\text{subject to} & \quad A(\hat{x} + 1)^{\frac{1}{2}} = b \\
& \quad G = \begin{bmatrix} X & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \succeq 0 \\
& \quad \delta(X) = 1
\end{align*}
\]

(723)

minimizes an affine function on an intersection of the elliptope with hyperplanes. Although the same Boolean solution is obtained from this approximation (728) as compared with semidefinite program (723), when given that particular data from Example 4.2.3.1.1, Singer confides a counterexample: Instead, given data

\[
A = \begin{bmatrix} 1 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]

(730)

then solving approximation (728) yields

\[
y^* = \text{round} \left( \begin{bmatrix} 1 - \frac{1}{\sqrt{2}} \\ 1 - \frac{1}{\sqrt{2}} \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

(731)

(infeasible, with or without rounding, with respect to original problem (711)) whereas solving semidefinite program (723) produces

\[
\text{round}(G^*) = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix}
\]

(732)

with sorted eigenvalues

\[
\lambda(G^*) = \begin{bmatrix} 3.99999965057264 \\ 0.00000035942736 \\ -0.00000000000000 \\ -0.0000000001000000 \end{bmatrix}
\]

(733)

Truncating all but the largest eigenvalue, from (719) we obtain (confer \(y^*\))

\[
x^* = \text{round} \left( \begin{bmatrix} 0.999999965299 \\ 0.9999999625299 \\ 0.00000001434518 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]

(734)

the desired minimal cardinality Boolean result.

\[\square\]

4.2.3.1.3 Exercise. Minimal cardinality Boolean art.
Assess general performance of standard-practice approximation (728) as compared with the proposed semidefinite program (723). ▼
4.3. RANK REDUCTION

4.2.3.1.4 Exercise. Conic independence.
Matrix $A$ from (713) is full-rank having three-dimensional nullspace. Find its four conically independent columns. (§2.10) To what part of proper cone $\mathcal{K} = \{Ax \mid x \succeq 0\}$ does vector $b$ belong? ▼

4.2.3.1.5 Exercise. Linear independence.
Show why fat matrix $A$, from compressed sensing problem (517) or (522), may be regarded full-rank without loss of generality. In other words: Is a minimal cardinality solution invariant to linear dependence of rows? ▼

4.3 Rank reduction

... it is not clear generally how to predict rank $X^*$ or rank $S^*$ before solving the SDP problem.

—Farid Alizadeh, 1995 [12, p.22]

The premise of rank reduction in semidefinite programming is: an optimal solution $X^*$ found does not satisfy Barvinok’s upper bound (272) on rank. The particular numerical algorithm solving a semidefinite program may have instead returned a high-rank optimal solution (§4.1.2; e.g., (691)) when a lower-rank optimal solution was expected. Rank reduction is a means to adjust rank of an optimal solution, returned by a solver, until it satisfies Barvinok’s upper bound.

4.3.1 Posit a perturbation of $X^*$

Recall from §4.1.2.1, there is an extreme point of $\mathcal{A} \cap S_n^+$ (683) satisfying upper bound (272) on rank. [25, §2.2] It is therefore sufficient to locate an extreme point of the intersection whose primal objective value (680P) is optimal.4.19 [119, §31.5.3] [249, §2.4] [250] [8, §3] [302]

Consider again affine subset

$$\mathcal{A} = \{X \in S^n \mid A \text{svec } X = b\}$$

where for $A_i \in S^n$

$$A \triangleq \begin{bmatrix}
\text{svec}(A_1)^T \\
\vdots \\
\text{svec}(A_m)^T
\end{bmatrix} \in \mathbb{R}^{m \times n(n+1)/2}$$

Given any optimal solution $X^*$ to

$$\begin{array}{ll}
\text{minimize} & \langle C, X \rangle \\
\text{subject to} & X \in \mathcal{A} \cap S_n^+
\end{array}$$

(680P)

4.19 There is no known construction for Barvinok’s tighter result (277). —Monique Laurent, 2004
whose rank does not satisfy upper bound (272), we posit existence of a set of perturbations
\[
\{ t_j B_j \mid t_j \in \mathbb{R}, B_j \in S^n, \ j = 1 \ldots n \}
\]
(735)
such that, for some \(0 \leq i \leq n\) and scalars \(\{ t_j, j = 1 \ldots i \}\),
\[
X^* + \sum_{j=1}^{i} t_j B_j
\]
(736)
becomes an extreme point of \(A \cap S^n_+\) and remains an optimal solution of (680P). Membership of (736) to affine subset \(A\) is secured for the \(i^\text{th}\) perturbation by demanding
\[
\langle B_i, A_j \rangle = 0, \quad j = 1 \ldots m
\]
(737)
while membership to the positive semidefinite cone \(S^n_+\) is insured by small perturbation (746). Feasibility of (736) is insured in this manner, optimality is proved in §4.3.3.

The following simple algorithm has very low computational intensity and locates an optimal extreme point, assuming a nontrivial solution:

4.3.1.0.1 Procedure. Rank reduction.
initialize: \(B_i = 0\ \forall i\)
for iteration \(i=1\ldots n\)
\{  
1. compute a nonzero perturbation matrix \(B_i\) of \(X^* + \sum_{j=1}^{i-1} t_j B_j\)
2. maximize \(t_i\)
   subject to \(X^* + \sum_{j=1}^{i} t_j B_j \in S^n_+\)
\}

A rank-reduced optimal solution is then
\[
X^* \leftarrow X^* + \sum_{j=1}^{i} t_j B_j
\]
(738)

4.3.2 Perturbation form
Perturbations of \(X^*\) are independent of constants \(C \in S^n\) and \(b \in \mathbb{R}^m\) in primal and dual problems (680). Numerical accuracy of any rank-reduced result, found by perturbation of an initial optimal solution \(X^*\), is therefore quite dependent upon initial accuracy of \(X^*\).

4.3.2.0.1 Definition. Matrix step function.
(confer §A.6.5.0.1)
Define the signum-like quasiconcave real function \(\psi : S^n \to \mathbb{R}\)
\[
\psi(Z) \triangleq \begin{cases} 1, & Z \succeq 0 \\ -1, & \text{otherwise} \end{cases}
\]
(739)
The value \(-1\) is taken for indefinite or nonzero negative semidefinite argument. \(\triangle\)
4.3. RANK REDUCTION

Deza & Laurent [119, §31.5.3] prove: every perturbation matrix \( B_i, i=1\ldots n \), is of the form
\[
B_i = -\psi(Z_i)R_iZ_iR_i^T \in \mathbb{S}^n
\]  
(740)

where
\[
X^+ \triangleq R_1R_1^T, \quad X^+ + \sum_{j=1}^{i-1} t_j B_j \triangleq R_iR_i^T \in \mathbb{S}^n
\]  
(741)

where the \( t_j \) are scalars and \( R_i \in \mathbb{R}^{n \times \rho} \) is full-rank and skinny where
\[
\rho \triangleq \text{rank} \left( X^+ + \sum_{j=1}^{i-1} t_j B_j \right)
\]  
(742)

and where matrix \( Z_i \in \mathbb{S}^\rho \) is found at each iteration \( i \) by solving a very simple feasibility problem:

\[
\begin{align*}
\text{find} & \quad Z_i \in \mathbb{S}^\rho \\
\text{subject to} & \quad \langle Z_i, R_i^TA_jR_i \rangle = 0, \quad j=1\ldots m
\end{align*}
\]  
(743)

Were there a sparsity pattern common to each member of set \( \{ R_i^TA_jR_i \in \mathbb{S}^\rho, j=1\ldots m \} \), then a good choice for \( Z_i \) has 1 in each entry corresponding to a 0 in the pattern; id est, a sparsity pattern complement. At iteration \( i \)
\[
X^+ + \sum_{j=1}^{i-1} t_j B_j + t_i B_i = R_i(I - t_i \psi(Z_i)Z_i)R_i^T
\]  
(744)

By fact (1532), therefore
\[
X^+ + \sum_{j=1}^{i-1} t_j B_j + t_i B_i \succeq 0 \iff 1 - t_i \psi(Z_i) \lambda(Z_i) \succeq 0
\]  
(745)

where \( \lambda(Z_i) \in \mathbb{R}^\rho \) denotes the eigenvalues of \( Z_i \).

Maximization of each \( t_i \) in step 2 of the Procedure reduces rank of (744) and locates a new point on the boundary \( \partial(A \cap \mathbb{S}^n_+) \). \( t_i \) has closed form; \( t_i^{*} = \max \{ \psi(Z_i) \lambda(Z_i) \} \)
\[
(746)
\]

A simple method of solution is closed-form projection of a random nonzero point on that proper subspace of isometrically isomorphic \( \mathbb{R}^{\rho(\rho+1)/2} \) specified by the constraints. (§E.5.0.0.6) Such a solution is nontrivial assuming the specified intersection of hyperplanes is not the origin; guaranteed by \( \rho(\rho+1)/2 > m \). Indeed, this geometric intuition about forming the perturbation is what bounds any solution’s rank from below; \( m \) is fixed by the number of equality constraints in (680P) while rank \( \rho \) decreases with each iteration \( i \). Otherwise, we might iterate indefinitely.

This holds because rank of a positive semidefinite matrix in \( \mathbb{S}^n \) is diminished below \( n \) by the number of its 0 eigenvalues (1542), and because a positive semidefinite matrix having one or more 0 eigenvalues corresponds to a point on the PSD cone boundary (193). Necessity and sufficiency are due to the facts: \( R_i \) can be completed to a nonsingular matrix (§A.3.1.0.5), and \( I - t_i \psi(Z_i)Z_i \) can be padded with zeros while maintaining equivalence in (744).
When $Z_i$ is indefinite, direction of perturbation (determined by $\psi(Z_i)$) is arbitrary. We may take an early exit from the Procedure were $Z_i$ to become 0 or were $\rho$ to become equal to 1 (assuming a nontrivial solution) or were

$$\text{rank}\left[\text{svec} R^T_i A_1 R_i \text{ svec} R^T_i A_2 R_i \cdots \text{svec} R^T_i A_m R_i\right] = \rho(\rho + 1)/2 \quad (747)$$

which characterizes rank $\rho$ of any [sic] extreme point in $\mathcal{A} \cap S^n_+$. [249, §2.4] [250]

**Proof.** Assuming the form of every perturbation matrix is indeed (740), then by (743)

$$\text{svec} Z_i \perp \left[\text{svec}(R^T_i A_1 R_i) \text{ svec}(R^T_i A_2 R_i) \cdots \text{svec}(R^T_i A_m R_i)\right] \quad (748)$$

By orthogonal complement we have

$$\text{rank}\left[\text{svec}(R^T_i A_1 R_i) \cdots \text{svec}(R^T_i A_m R_i)\right] \perp + \text{rank}\left[\text{svec}(R^T_i A_1 R_i) \cdots \text{svec}(R^T_i A_m R_i)\right] = \rho(\rho + 1)/2 \quad (749)$$

When $Z_i$ can only be 0, then the perturbation is null because an extreme point has been found; thus

$$\left[\text{svec}(R^T_i A_1 R_i) \cdots \text{svec}(R^T_i A_m R_i)\right] \perp = 0 \quad (750)$$

from which the stated result (747) directly follows. ♦

### 4.3.3 Optimality of perturbed $X^*$

We show that the optimal objective value is unaltered by perturbation (740); id est,

$$\langle C , X^* + \sum_{j=1}^i t_j B_j \rangle = \langle C , X^* \rangle \quad (751)$$

**Proof.** From Corollary 4.2.3.0.1 we have the necessary and sufficient relationship between optimal primal and dual solutions under assumption of nonempty primal feasible cone interior $\mathcal{A} \cap \text{int} S^n_+$:

$$S^*X^* = S^*R_1 R^T_1 = X^*S^* = R_1 R^T_1 S^* = 0 \quad (752)$$

This means $\mathcal{R}(R_1) \subseteq \mathcal{N}(S^*)$ and $\mathcal{R}(S^*) \subseteq \mathcal{N}(R^T_1)$. From (741) and (744) we get the sequence:

$$X^* = R_1 R^T_1$$

$$X^* + t_1 B_1 = R_2 R^T_2 = R_1(I - t_1 \psi(Z_1)Z_1)R^T_1$$

$$X^* + t_1 B_1 + t_2 B_2 = R_3 R^T_3 = R_2(I - t_2 \psi(Z_2)Z_2)R^T_2 = R_1(I - t_1 \psi(Z_1)Z_1)(I - t_2 \psi(Z_2)Z_2)R^T_1$$

$$\vdots$$

$$X^* + \sum_{j=1}^i t_j B_j = R_1 \left( \prod_{j=1}^i (I - t_j \psi(Z_j)Z_j) \right)R^T_1 \quad (753)$$
4.3. RANK REDUCTION

Substituting $C = \text{svec}^{-1}(A^T y^*) + S^*$ from (680),

$$
\langle C, X^* + \sum_{j=1}^t t_j B_j \rangle = \left\langle \text{svec}^{-1}(A^T y^*) + S^*, R_1 \left( \prod_{j=1}^t (I - t_j \psi(Z_j)) \right) R_1^T \right\rangle = \sum_{k=1}^m y_k^* A_k, X^* + \sum_{j=1}^t t_j B_j \right\rangle = \sum_{k=1}^m y_k^* A_k + S^*, X^* = \langle C, X^* \rangle
$$

(754)

because $\langle B_i, A_j \rangle = 0 \forall i, j$ by design (737).

4.3.3.0.1 Example. $A \delta(X) = b$.

This academic example demonstrates that a solution found by rank reduction can certainly have rank less than Barvinok’s upper bound (272): Assume a given vector $b$ belongs to the conic hull of columns of a given matrix $A$.

$$
A = \begin{bmatrix}
-1 & 1 & 8 & 1 & 1 \\
-3 & 2 & 8 & \frac{1}{2} & \frac{1}{2} \\
-9 & 4 & 8 & \frac{1}{4} & \frac{1}{4}
\end{bmatrix} \in \mathbb{R}^{m \times n}, \quad b = \begin{bmatrix}
1 \\
1 \\
2
\end{bmatrix} \in \mathbb{R}^n
$$

(755)

Consider the convex optimization problem

$$
\begin{align*}
\text{minimize} & \quad \text{tr} \, X \\
\text{subject to} & \quad X \succeq 0 \\
& \quad A \delta(X) = b
\end{align*}
$$

(756)

that minimizes the 1-norm of the main diagonal; id est, problem (756) is the same as

$$
\begin{align*}
\text{minimize} & \quad \| \delta(X) \|_1 \\
\text{subject to} & \quad X \succeq 0 \\
& \quad A \delta(X) = b
\end{align*}
$$

(757)

that finds a solution to $A \delta(X) = b$. Rank-3 solution $X^* = \delta(x_M)$ is optimal, where (confer (715))

$$
x_M = \begin{bmatrix}
\frac{2}{128} \\
0 \\
\frac{5}{128} \\
0 \\
\frac{90}{128}
\end{bmatrix}
$$

(758)

Yet upper bound (272) predicts existence of at most a

$$
\text{rank}(\left\lfloor \frac{\sqrt{8m} + 1 - 1}{2} \right\rfloor) = 2
$$

(759)
feasible solution from \( m = 3 \) equality constraints. To find a lower rank \( \rho \) optimal solution to (756) (barring combinatorics), we invoke Procedure 4.3.1.0.1:

Initialize:
\[
C = I, \quad \rho = 3, \quad A_j \triangleq \delta(A(j,:)), \ j = 1, 2, 3, \quad X^* = \delta(x_M), \quad m = 3, \ n = 5.
\]

\{ Iteration i=1: \\
\[
\text{Step 1: } R_1 = \begin{bmatrix}
\sqrt{\frac{2}{128}} & 0 & 0 \\
0 & \sqrt{\frac{5}{128}} & 0 \\
0 & 0 & \sqrt{\frac{90}{128}}
\end{bmatrix}.
\]

find \( Z_1 \in S^3 \)
subject to \( \langle Z_1, R_1^T A_j R_1 \rangle = 0, \ j = 1, 2, 3 \) (760)
A nonzero randomly selected matrix \( Z_1 \), having 0 main diagonal, is a solution yielding nonzero perturbation matrix \( B_1 \). Choose arbitrarily
\[
Z_1 = 11^T - I \in S^3
\] (761)
then (rounding)
\[
B_1 = \begin{bmatrix}
0 & 0 & 0.0247 & 0.1048 \\
0 & 0.0247 & 0 & 0.1657 \\
0.1048 & 0.1657 & 0 & 0
\end{bmatrix}
\]

\{ Step 2: \( t_1^* = 1 \) because \( \lambda(Z_1) = [-1 \ -1 \ 2]^T \). So,
\[
X^* \leftarrow \delta(x_M) + B_1 = \begin{bmatrix}
\sqrt{\frac{2}{128}} & 0 & 0.0247 & 0.1048 \\
0 & 0 & 0 & 0 \\
0.0247 & 0 & \frac{5}{128} & 0.1657 \\
0.1048 & 0.1657 & 0 & \frac{90}{128}
\end{bmatrix}
\]

has rank \( \rho = 1 \) and produces the same optimal objective value. \}

\[ \square \]

4.3.3.0.2 Exercise. Rank reduction of maximal complementarity.
Apply rank reduction Procedure 4.3.1.0.1 to the maximal complementarity example (§4.1.2.3.1). Demonstrate a rank-1 solution; which can certainly be found (by Barvinok’s Proposition 2.9.3.0.1) because there is only one equality constraint. ▼
4.4. Rank-constrained semidefinite program

4.3.4 thoughts regarding rank reduction

Because rank reduction Procedure 4.3.1.0.1 is guaranteed only to produce another optimal solution conforming to Barvinok’s upper bound (272), the Procedure will not necessarily produce solutions of arbitrarily low rank; but if they exist, the Procedure can. Arbitrariness of search direction when matrix $Z_i$ becomes indefinite, mentioned on page 260, and the enormity of choices for $Z_i$ (743) are liabilities for this algorithm.

4.3.4.1 inequality constraints

The question naturally arises: what to do when a semidefinite program (not in prototypical form (680))\footnote{Contemporary numerical packages for solving semidefinite programs can solve a range of problems wider than prototype (680). Generally, they do so by transforming a given problem into prototypical form by introducing new constraints and variables. [12] [417] We are momentarily considering a departure from the primal prototype that augments the constraint set with linear inequalities.} has linear inequality constraints of the form

$$\alpha_i^T \text{svec} X \preceq \beta_i, \quad i = 1 \ldots k$$

(764)

where $\{\beta_i\}$ are given scalars and $\{\alpha_i\}$ are given vectors. One expedient way to handle this circumstance is to convert the inequality constraints to equality constraints by introducing a slack variable $\gamma$; id est,

$$\alpha_i^T \text{svec} X + \gamma_i = \beta_i, \quad i = 1 \ldots k, \quad \gamma \succeq 0$$

(765)

thereby converting the problem to prototypical form.

Alternatively, we say the $i^{th}$ inequality constraint is active when it is met with equality; id est, when for particular $i$ in (764), $\alpha_i^T \text{svec} X^* = \beta_i$. An optimal high-rank solution $X^*$ is, of course, feasible (satisfying all the constraints). But for the purpose of rank reduction, inactive inequality constraints are ignored while active inequality constraints are interpreted as equality constraints. In other words, we take the union of active inequality constraints (as equalities) with equality constraints $A \text{svec} X = b$ to form a composite affine subset $\hat{A}$ substituting for (683). Then we proceed with rank reduction of $X^*$ as though the semidefinite program were in prototypical form (680P).

4.4 Rank-constrained semidefinite program

We generalize the trace heuristic (§7.2.2.1), for finding low-rank optimal solutions to SDPs of a more general form:

4.4.1 rank constraint by convex iteration

Consider a semidefinite feasibility problem of the form

$$\begin{align*}
\text{find} \quad & G \\
\text{subject to} \quad & G \in \mathbb{S}^n \\
& G \in \mathcal{C} \\
& G \succeq 0 \\
& \text{rank} G \leq n
\end{align*}$$

(766)
CHAPTER 4. SEMIDEFINITE PROGRAMMING

where $C$ is a convex set presumed to contain positive semidefinite matrices of rank $n$ or less; \textit{id est}, $C$ intersects the positive semidefinite cone boundary. We propose that this rank-constrained feasibility problem can be equivalently expressed as iteration of the convex problem sequence (767) and (1792a):

$$\begin{aligned}
\text{minimize} & \quad \langle G, W \rangle \\
\text{subject to} & \quad G \in C \\
& \quad G \succeq 0
\end{aligned} \quad (767)$$

where \textit{direction vector} \textsuperscript{4.23} $W$ is an optimal solution to semidefinite program, for $0 \leq n \leq N - 1$

$$\sum_{i=n+1}^{N} \lambda(G^*)_i = \minimize_{W \in \mathbb{S}^N} \langle G^*, W \rangle \quad (1792a)$$

whose feasible set is a Fantope (§2.3.2.0.1), and where $G^*$ is an optimal solution to problem (767) given some iterate $W$. The idea is to iterate solution of (767) and (1792a) until convergence as defined in §4.4.1.2: \textsuperscript{4.24} (confer (803))

$$\sum_{i=n+1}^{N} \lambda(G^*)_i = \langle G^*, W^* \rangle = \lambda(G^*)^T \lambda(W^*) \triangleq 0 \quad (768)$$

defines \textit{global convergence} of the iteration; a vanishing objective that is a certificate of global optimality but cannot be guaranteed. \textit{Optimal direction vector} $W^*$ is defined as any positive semidefinite matrix yielding optimal solution $G^*$ of rank $n$ or less to then convex equivalent (767) of feasibility problem (766):

$$\begin{aligned}
\text{find} & \quad G \\
\text{subject to} & \quad G \in C \\
& \quad G \succeq 0 \\
& \quad \text{rank} G \leq n
\end{aligned} \quad (766)$$

\textit{id est}, any direction vector for which the last $N - n$ nonincreasingly ordered eigenvalues $\lambda$ of $G^*$ are zero.

In any semidefinite feasibility problem, a solution of least rank must be an extreme point of the feasible set.\textsuperscript{4.25} This means there exists a hyperplane supporting the feasible set at that extreme point. Then there must exist a linear objective function such that this least-rank feasible solution optimizes the resultant semidefinite program.

\textsuperscript{4.23} Search \textit{direction} $W$ is a hyperplane-normal pointing opposite to direction of movement describing minimization of a real linear function $\langle G, W \rangle$ (p.67).

\textsuperscript{4.24} Proposed iteration is neither \textit{dual projection} (Figure 176) or \textit{alternating projection} (Figure 180). Sum of eigenvalues follows from results of Ky Fan (page 565). Inner product of eigenvalues follows from (1674) and properties of commutative matrix products (page 522).

\textsuperscript{4.25} which follows by \textit{extremes theorem} 2.8.1.1.1, by rank of a sum of positive semidefinite matrices (1548) (257), and by definition of extreme point (167) for which no convex combination can produce it: If a least rank solution were expressible as a convex combination of feasible points, then there could exist feasible matrices of lesser rank.
4.4. RANK-CONSTRAINED SEMIDEFINITE PROGRAM

We emphasize that convex problem (767) is not a relaxation of rank-constrained feasibility problem (766); at global convergence, convex iteration (767) (1792a) makes it instead an equivalent problem.

4.4.1.1 direction matrix interpretation

(confer §4.5.1.2) The feasible set of direction matrices in (1792a) is the convex hull of outer product of all rank-(N − n) orthonormal matrices; videlicet,

\[ \text{conv}\{UU^T \mid U \in \mathbb{R}^{N \times N-n}, \ U^TU = I\} = \left\{ A \in \mathbb{S}^N \mid I \succeq A \succeq 0, \ \langle I, A\rangle = N-n \right\} \quad (90) \]

This set (92), argument to conv\{\}, comprises the extreme points of this Fantope (90). An optimal solution \( W \) to (1792a), that is an extreme point, is known in closed form (p.565): Given ordered diagonalization \( G^* = QAQ^T \in \mathbb{S}^N_+ \) (§A.5.1), then direction matrix \( W = U^*U^{*T} \) is optimal and extreme where \( U^* = Q(:, n+1:N) \in \mathbb{R}^{N \times N-n} \). Eigenvalue \( \lambda(W) \) has 1 in each entry corresponding to the \( N-n \) smallest entries of \( \delta(\Lambda) \) and has 0 elsewhere. By (221) (223), polar direction \(-W\) can be regarded as pointing toward the set of all rank-\( n \) (or less) positive semidefinite matrices whose nullspace contains that of \( G^* \). For that particular closed-form solution \( W \), consequently, (confer (805))

\[ \sum_{i=n+1}^{N} \lambda(G^*)_i = \langle G^*, W \rangle = \lambda(G^*)^T \lambda(W) \geq 0 \quad (769) \]

This is the connection to cardinality minimization of vectors; \(^4\text{26} \) id est, eigenvalue \( \lambda \) cardinality (rank) is analogous to vector \( x \) cardinality via (805): for positive semidefinite \( X \)

\[ \sum_i \lambda(X)_i = \text{tr}X = \|X\|_2^* \iff \|x\|_1 \]
\[ \sqrt{\sum_i \lambda(X)_i^2} = \sqrt{\text{tr}(X^2)} = \|X\|_F \iff \|x\|_2 \]
\[ \max_i \{\lambda(X)_i\} \quad = \|X\|_\infty \iff \|x\|_\infty \quad (770) \]

So that this method, for constraining rank, will not be misconstrued under closed-form solution \( W \) to (1792a): Define (confer (221))

\[ \mathcal{S}_n \triangleq \{(I-W)G(I-W) \mid G \in \mathbb{S}^N\} = \{X \in \mathbb{S}^N \mid \mathcal{N}(X) \supseteq \mathcal{N}(G^*)\} \quad (771) \]

as the symmetric subspace of rank\( \leq n \) matrices whose nullspace contains \( \mathcal{N}(G^*) \). Then projection of \( G^* \) on \( \mathcal{S}_n \) is \( (I-W)G^*(I-W) \). (§E.7) Direction of projection is \(-WG^*W\). (Figure 87) \text{tr}(WG^*W)\) is a measure of proximity to \( \mathcal{S}_n \) because its orthogonal complement is \( \mathcal{S}_n^\perp = \{WGW \mid G \in \mathbb{S}^N\} \); the point being, convex iteration incorporating constrained \text{tr}(WGW) = \langle G, W \rangle \) minimization is not a projection method; certainly, not on these two subspaces.

\(^4\text{26} \) not trace minimization of a nonnegative diagonal matrix \( \delta(x) \) as in [142, g1] [311, g2]. To make rank-constrained problem (766) resemble cardinality problem (529), we could make \( \mathcal{C} \) an affine subset:

\[
\begin{align*}
\text{find} & \quad X \in \mathbb{S}^N \\
\text{subject to} & \quad A\text{svec}X = b \\
& \quad X \succeq 0 \\
& \quad \text{rank}X \leq n
\end{align*}
\]
Figure 87: (confer Figure 177) Projection of $G^*$ on subspace $\mathcal{S}_n$ of rank $\leq n$ matrices whose nullspace contains $\mathcal{N}(G^*)$. This direction $W$ is closed-form solution to (1792a).

Figure 88: (confer Figure 104) Trace heuristic can be interpreted as minimization of a hyperplane, with normal $I$, over positive semidefinite cone drawn here in isometrically isomorphic $\mathbb{R}^3$. Polar of direction vector $W = I$ points toward origin.
4.4. RANK-CONSTRAINED SEMIDEFINITE PROGRAM

Closed-form solution $W$ to problem (1792a), though efficient, comes with a caveat: there exist cases where this projection matrix solution $W$ does not provide the shortest route to an optimal rank-$n$ solution $G^*$; id est, direction $W$ is not unique. So we sometimes choose to solve (1792a) instead of employing a known closed-form solution.

When direction matrix $W = I$, as in the trace heuristic for example, then $-W$ points directly at the origin (the rank-0 PSD matrix, Figure 88). Vector inner-product of an optimization variable with direction matrix $W$ is therefore a generalization of the trace heuristic (§7.2.2.1) for rank minimization; $-W$ is instead trained toward the boundary of the positive semidefinite cone.

4.4.1.2 convergence

We study convergence to ascertain conditions under which a direction matrix will reveal a feasible solution $G$, of rank $n$ or less, to semidefinite program (767). Denote by $W^*$ a particular optimal direction matrix from semidefinite program (1792a) such that (768) holds (feasible rank $G \leq n$ found). Then we define global convergence of the iteration (767) (1792a) to correspond with this vanishing vector inner-product (768) of optimal solutions.

Because this iterative technique for constraining rank is not a projection method, it can find a rank-\(n\) solution $G^*$ (768) will be satisfied) only if at least one exists in the feasible set of program (767).

4.4.1.2.1 Proof. Suppose $\langle G^*, W \rangle = \tau$ is satisfied for some nonnegative constant $\tau$ after any particular iteration (767) (1792a) of the two minimization problems. Once a particular value of $\tau$ is achieved, it can never be exceeded by subsequent iterations because existence of feasible $G$ and $W$ having that vector inner-product $\tau$ has been established simultaneously in each problem. Because the infimum of vector inner-product of two positive semidefinite matrix variables is zero, the nonincreasing sequence of iterations is thus bounded below hence convergent because any bounded monotonic sequence in $\mathbb{R}$ is convergent. [268, §1.2] [43, §1.1] Local convergence to some nonnegative objective value $\tau$ is thereby established.

Local convergence, in this context, means convergence to a fixed point of possibly infeasible rank. Only local convergence can be established because objective $\langle G, W \rangle$, when instead regarded simultaneously in two variables $(G, W)$, is generally multimodal. (§3.8.0.0.3)

Local convergence, convergence to $\tau \neq 0$ and definition of a stall, never implies nonexistence of a rank-$n$ feasible solution to (767). A nonexistent rank-$n$ feasible solution would mean certain failure to converge globally by definition (768) (convergence to $\tau \neq 0$) but, as proved, convex iteration always converges locally if not globally.

When a rank-$n$ feasible solution to (767) exists, it remains an open problem to state conditions under which $\langle G^*, W^* \rangle = \tau = 0$ (768) is achieved by iterative solution of semidefinite programs (767) and (1792a). Then rank $G^* \leq n$ and pair $(G^*, W^*)$ becomes a globally optimal fixed point of iteration. There can be no proof of global convergence because of the implicit high-dimensional multimodal manifold in variables $(G, W)$. When stall occurs, direction vector $W$ can be manipulated to steer out; e.g, reversal of search direction as in Example 4.6.0.0.1, or reinitialization to a random
rank-$(N-n)$ matrix in the same positive semidefinite cone face (§2.9.2.3) demanded by the current iterate: given ordered diagonalization $G^* = Q\Lambda Q^T \in S^N$, then $W = U^* \Phi U^T$ where $U^* = Q(:, n+1:N) \in \mathbb{R}^{N \times N-n}$ and where eigenvalue vector $\lambda(W)_{1:N-n} = \lambda(\Phi)$ has nonnegative uniformly distributed random entries in $(0,1]$ by selection of $\Phi \in S^{N-n}$ while $\lambda(W)_{N-n+1:N} = 0$. Zero eigenvalues act as memory while randomness largely reduces likelihood of stall. When this direction works, rank and objective sequence $(G^*, W)$ with respect to iteration tend to be noisily monotonic.

4.4.1.2.2 Exercise. Completely positive semidefinite matrix.

Given rank-2 positive semidefinite matrix $G = \begin{bmatrix} 0.50 & 0.55 & 0.20 \\ 0.55 & 0.61 & 0.22 \\ 0.20 & 0.22 & 0.08 \end{bmatrix}$, find a positive factorization $G = X^T X$ (978) by solving

\[
\begin{align*}
\text{find} & \quad X \in \mathbb{R}^{2\times3} \\
\text{subject to} & \quad Z = \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} \succeq 0 \\
& \quad \text{rank} \, Z \leq 2
\end{align*}
\]

via convex iteration. \(\nabla\)

4.4.1.2.3 Exercise. Nonnegative matrix factorization.

Given rank-2 nonnegative matrix $X = \begin{bmatrix} 17 & 28 & 42 \\ 16 & 47 & 51 \\ 17 & 82 & 72 \end{bmatrix}$, find a nonnegative factorization $X = WH$ (773) by solving

\[
\begin{align*}
\text{find} & \quad W, H \in \mathbb{R}^{3\times2}, \mathbb{R}^{2\times3} \\
\text{subject to} & \quad Z = \begin{bmatrix} I & W^T & H \\ W & A & X \\ H^T & X^T & B \end{bmatrix} \succeq 0 \\
& \quad W \succeq 0 \\
& \quad H \succeq 0 \\
& \quad \text{rank} \, Z \leq 2
\end{align*}
\]

which follows from the fact, at optimality,

\[
Z^* = \begin{bmatrix} I & W^T & H \\ W & H^T \end{bmatrix}
\]

Use the known closed-form solution for a direction vector $Y$ to regulate rank by convex iteration; set $Z^* = Q\Lambda Q^T \in S^8$ to an ordered diagonalization and $U^* = Q(:, 3:8) \in \mathbb{R}^{8\times6}$, then $Y = U^* U^T$ (§4.4.1.1).
4.4. RANK-CONSTRAINED SEMIDEFINITE PROGRAM

Figure 89: Sensor-network localization in $\mathbb{R}^2$, illustrating connectivity and circular radio-range per sensor. Smaller dark grey regions each hold an anchor at their center; known fixed sensor positions. Sensor/anchor distance is measurable with negligible uncertainty for sensor within those grey regions. (Graphic by Geoff Merrett)

In summary, initialize $Y$ then iterate numerical solution of (convex) semidefinite program

$$
\begin{align*}
\text{minimize} \quad & \langle Z, Y \rangle \\
\text{subject to} \quad & Z = \begin{bmatrix} I & W^T & H \\ W & A & X \\ H^T & X^T & B \end{bmatrix} \succeq 0 \\
& W \geq 0 \\
& H \geq 0
\end{align*}
$$

(776)

with $Y = U^*U^{*T}$ until convergence (which is global and occurs in very few iterations for this instance).

Now, an application to optimal regulation of affine dimension:

4.4.1.2.4 Example. Sensor-Network Localization and Wireless Location. Heuristic solution to a sensor-network localization problem, proposed by Carter, Jin,
Saunders, & Ye in [74], is limited to two Euclidean dimensions and applies semidefinite programming (SDP) to little subproblems. There, a large network is partitioned into smaller subnetworks (as small as one sensor—a mobile point, whereabouts unknown) and then semidefinite programming and heuristics called SPASELOC are applied to localize each and every partition by two-dimensional distance geometry. Their partitioning procedure is one-pass, yet termed iterative; a term applicable only insofar as adjoining partitions can share localized sensors and anchors (absolute sensor positions known a priori). But there is no iteration on the entire network, hence the term “iterative” is perhaps inappropriate. As partitions are selected based on “rule sets” (heuristics, not geographics), they also term the partitioning adaptive. But no adaptation of a partition actually occurs once it has been determined.

One can reasonably argue that semidefinite programming methods are unnecessary for localization of small partitions of large sensor networks. In the past, these nonlinear localization problems were solved algebraically and computed by least squares solution to hyperbolic equations; called multilateration. Indeed, practical contemporary numerical methods for global positioning (GPS) by satellite do not rely on convex optimization.

Modern distance geometry is inextricably melded with semidefinite programming. The beauty of semidefinite programming, as relates to localization, lies in convex expression of classical multilateration: So & Ye showed that the problem of finding unique solution, to a noiseless nonlinear system describing the common point of intersection of hyperspheres in real Euclidean vector space, can be expressed as a semidefinite program via distance geometry.

But the need for SDP methods in Carter & Jin et alii is enigmatic for two more reasons: 1) guessing solution to a partition whose intersensor measurement data or connectivity is inadequate for localization by distance geometry, 2) reliance on complicated and extensive heuristics for partitioning a large network that could instead be efficiently solved whole by one semidefinite program. While partitions range in size between 2 and 10 sensors, 5 sensors optimally, heuristics provided are only for two spatial dimensions (no higher-dimensional heuristics are proposed). For these small numbers it remains unclarified as to precisely what advantage is gained over traditional least squares: it is difficult to determine what part of their noise performance is attributable to SDP and what part is attributable to their heuristic geometry.

Partitioning of large sensor networks is a compromise to rapid growth of SDP computational intensity with problem size. But when impact of noise on distance measurement is of most concern, one is averse to a partitioning scheme because noise-effects vary inversely with problem size. Since an individual partition’s solution is not iterated in Carter & Jin and is interdependent with adjoining partitions, we expect errors to propagate from one partition to the next; the ultimate partition solved, expected to suffer most.

Heuristics often fail on real-world data because of unanticipated circumstances.

---

4.27 The paper constitutes Jin’s dissertation for University of Toronto although her name appears as second author. Ye’s authorship is honorary.

4.28 Multilateration—literally, having many sides; shape of a geometric figure formed by nearly intersecting lines of position. In navigation systems, therefore: Obtaining a fix from multiple lines of position. Multilateration can be regarded as noisy trilateration.
When heuristics fail, generally they are repaired by adding more heuristics. Tenous is any presumption, for example, that distance measurement errors have distribution characterized by circular contours of equal probability about an unknown sensor-location. (Figure 89) That presumption effectively appears within Carter & Jin’s optimization problem statement as affine equality constraints relating unknowns to distance measurements that are corrupted by noise. Yet in most all urban environments, this measurement noise is more aptly characterized by ellipsoids of varying orientation and eccentricity as one recedes from a sensor. (Figure 138) Each unknown sensor must therefore instead be bound to its own particular range of distance, primarily determined by the terrain.4.29 The nonconvex problem we must instead solve is:

\[
\begin{align*}
\text{find} & \quad \{x_i, x_j\} \\
\text{subject to} & \quad d_{ij} \leq ||x_i - x_j||^2 \leq \bar{d}_{ij}
\end{align*}
\]

(777)

where \(x_i\) represents sensor location, and where \(d_{ij}\) and \(\bar{d}_{ij}\) respectively represent lower and upper bounds on measured distance-square from \(i^{th}\) to \(j^{th}\) sensor (or from sensor to anchor). Figure 94 illustrates contours of equal sensor-location uncertainty. By establishing these individual upper and lower bounds, orientation and eccentricity can effectively be incorporated into the problem statement.

Generally speaking, there can be no unique solution to the sensor-network localization problem because there is no unique formulation; that is the art of Optimization. Any optimal solution obtained depends on whether or how a network is partitioned, whether distance data is complete, presence of noise, and how the problem is formulated. When a particular formulation is a convex optimization problem, then the set of all optimal solutions forms a convex set containing the actual or true localization. Measurement noise precludes equality constraints representing distance. The optimal solution set is consequently expanded; necessitated by introduction of distance inequalities admitting more and higher-rank solutions. Even were the optimal solution set a single point, it is not necessarily the true localization because there is little hope of exact localization by any algorithm once significant noise is introduced.

Carter & Jin gauge performance of their heuristics to the SDP formulation of author Biswas whom they regard as vanguard to the art. [15, §1] Biswas posed localization as an optimization problem minimizing a distance measure. [48] [46] Intuitively, minimization of any distance measure yields compacted solutions; (confer §6.7.0.0.1) precisely the anomaly motivating Carter & Jin. Their two-dimensional heuristics outperformed Biswas’ localizations both in execution-time and proximity to the desired result. Perhaps, instead of heuristics, Biswas’ approach to localization can be improved: [45] [47].

The sensor-network localization problem is considered difficult. [15, §2] Rank constraints in optimization are considered more difficult. Control of affine dimension in Carter & Jin is suboptimal because of implicit projection on \(\mathbb{R}^2\). In what follows, we present the localization problem as a semidefinite program (equivalent to (777)) having an explicit rank constraint which controls affine dimension of an optimal solution. We show how to achieve that rank constraint only if the feasible set contains a matrix of desired rank. Our problem formulation is extensible to any spatial dimension.

4.29 A distinct contour map corresponding to each anchor is required in practice.
proposed standardized test

Jin proposes an academic test in two-dimensional real Euclidean space $\mathbb{R}^2$ that we adopt. In essence, this test is a localization of sensors and anchors arranged in a regular triangular lattice. Lattice connectivity is solely determined by sensor radio range; a connectivity graph is assumed incomplete. In the interest of test standardization, we propose adoption of a few small examples: Figure 90 through Figure 93 and their particular connectivity represented by matrices (778) through (781) respectively.

$$
\begin{array}{cccc}
0 & \bullet & ? & \bullet \\
\bullet & 0 & \bullet & \bullet \\
? & \bullet & 0 & \circ \\
\bullet & \bullet & \circ & 0
\end{array}
$$

Matrix entries $\bullet$ indicate measurable distance between nodes while unknown distance is denoted by $?$ (question mark). Matrix entries hollow dot $\circ$ represent known distance between anchors (to high accuracy) while zero distance is denoted 0. Because measured distances are quite unreliable in practice, our solution to the localization problem substitutes a distinct range of possible distance for each measurable distance; equality constraints exist only for anchors.

Anchors are chosen so as to increase difficulty for algorithms dependent on existence of sensors in their convex hull. The challenge is to find a solution in two dimensions close to the true sensor positions given incomplete noisy intersensor distance information.
Figure 91: 3-lattice in $\mathbb{R}^2$, hand-drawn. Nodes 7, 8, and 9 are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc.
Figure 92: 4-lattice in $\mathbb{R}^2$, hand-drawn. Nodes 13, 14, 15, and 16 are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc.
Figure 93: 5-lattice in $\mathbb{R}^2$. Nodes 21 through 25 are anchors.
**Problem Statement**

Ascribe points in a list \( \{x_\ell \in \mathbb{R}^n, \ \ell = 1 \ldots N \} \) to the columns of a matrix \( X \):

\[
X = [x_1 \cdots x_N] \in \mathbb{R}^{n \times N} \tag{76}
\]

where \( N \) is regarded as cardinality of list \( X \). Positive semidefinite matrix \( X^T X \), formed from inner product of the list, is a *Gram matrix*; \([260, \S 3.6]\)

\[
G = X^T X = \begin{bmatrix}
\|x_1\|^2 & x_1^T x_2 & x_1^T x_3 & \cdots & x_1^T x_N \\
x_2^T x_1 & \|x_2\|^2 & x_2^T x_3 & \cdots & x_2^T x_N \\
x_3^T x_1 & x_3^T x_2 & \|x_3\|^2 & \ddots & x_3^T x_N \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
x_N^T x_1 & x_N^T x_2 & x_N^T x_3 & \cdots & \|x_N\|^2
\end{bmatrix} \in S_N^+ \tag{978}
\]

where \( S_N^+ \) is the convex cone of \( N \times N \) positive semidefinite matrices in the symmetric matrix subspace \( S^N \).

Existence of noise precludes measured distance from the input data. We instead assign measured distance to a range estimate specified by individual upper and lower bounds: \( \overline{d}_{ij} \) is an upper bound on distance-square from \( i \)th to \( j \)th sensor, while \( \underline{d}_{ij} \) is a lower bound. These bounds become the input data. Each measurement range is presumed different from the others because of measurement uncertainty; *e.g.*, Figure 94.
Our mathematical treatment of anchors and sensors is not dichotomized.\(^{4.30}\) A sensor position that is known \textit{a priori} to high accuracy (with absolute certainty) \(\hat{x}_i\) is called an anchor. Then the sensor-network localization problem \((777)\) can be expressed equivalently: Given a number of anchors \(m\), and \(I\) a set of indices (corresponding to all measurable distances \(\bullet\)), for \(0 < n < N\)

\[
\begin{aligned}
\text{find } & X \\
\text{subject to } & \frac{d_{ij}}{\sqrt{2}} \leq \langle G, (e_i - e_j)(e_i - e_j)^T \rangle \leq \bar{d}_{ij} \quad \forall (i, j) \in I \\
& \langle G, e_i e_i^T \rangle = \|\hat{x}_i\|^2, \quad i = N - m + 1 \ldots N \\
& \langle G, (e_i e_j^T + e_j e_i^T)/2 \rangle = \hat{x}_i^T \hat{x}_j, \quad i < j, \quad \forall i, j \in \{N - m + 1 \ldots N\} \\
& X(:, N - m + 1:N) = [\hat{x}_{N-m+1} \ldots \hat{x}_N] \\
& Z = \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} \succeq 0 \\
& \text{rank } Z = n 
\end{aligned}
\]

(782)

where \(e_i\) is the \(i\)th member of the standard basis for \(\mathbb{R}^N\). Distance-square

\[
d_{ij} = \|x_i - x_j\|^2 = \langle x_i - x_j, x_i - x_j \rangle
\]

(965)
is related to Gram matrix entries \(G \triangleq [g_{ij}]\) by vector inner-product

\[
d_{ij} = g_{ii} + g_{jj} - 2g_{ij} = \langle G, (e_i - e_j)(e_i - e_j)^T \rangle = \text{tr}(G^T(e_i - e_j)(e_i - e_j)^T)
\]

(980)
hence the scalar inequalities. Each linear equality constraint in \(G \in \mathbb{S}^N\) represents a hyperplane in isometrically isomorphic Euclidean vector space \(\mathbb{R}^{N(N+1)/2}\), while each linear inequality pair represents a convex Euclidean body known as \textit{slab}.\(^{4.31}\) By Schur complement (\(\S \text{A.4}\)), any solution \((G, X)\) provides comparison with respect to the positive semidefinite cone

\[
G \succeq X^T X 
\]

(1018)

which is a convex relaxation of the desired equality constraint

\[
\begin{bmatrix} I & X \\ X^T & G \end{bmatrix} = \begin{bmatrix} I & X \\ X^T \end{bmatrix} \begin{bmatrix} I & X \end{bmatrix}
\]

(1019)

The rank constraint insures this equality holds, by Theorem \(\S \text{A.4.0.1.3}\), thus restricting solution to \(\mathbb{R}^n\). Assuming full-rank solution (list) \(X\)

\[
\text{rank } Z = \text{rank } G = \text{rank } X 
\]

(783)
convex equivalent problem statement

Problem statement (782) is nonconvex because of the rank constraint. We do not eliminate or ignore the rank constraint; rather, we find a convex way to enforce it: for $0 < n < N$

\[
\begin{align*}
\text{minimize} & \quad (Z, W) \\
\text{subject to} & \quad d_{ij} \leq \langle G, (e_i - e_j)(e_i - e_j)^T \rangle \leq \overline{d}_{ij} & \forall (i,j) \in \mathcal{I} \\
& \quad \langle G, (e_i e_j^T + e_j e_i^T) / 2 \rangle = \tilde{x}_i^T \tilde{x}_j & \forall i,j \in \{N - m + 1 \ldots N\} \\
& \quad X(:,N-m+1:N) = [\tilde{x}_{N-m+1} \ldots \tilde{x}_N] \\
& \quad Z = \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} \succeq 0 \quad (784)
\end{align*}
\]

Convex function $\text{tr} \ Z$ is a well-known heuristic whose sole purpose is to represent convex envelope of rank $Z$. (§7.2.2.1) In this convex optimization problem (784), a semidefinite program, we substitute a vector inner-product objective function for trace;

\[
\text{tr} \ Z = \langle Z, I \rangle - \langle Z, W \rangle \quad (785)
\]
a generalization of the trace heuristic for minimizing convex envelope of rank, where $W \in \mathbb{S}^{N+n}_+$ is constant with respect to (784). Matrix $W$ is normal to a hyperplane in $\mathbb{S}^{N+n}_+$ minimized over a convex feasible set specified by the constraints in (784). Matrix $W$ is chosen so $-W$ points in direction of rank-$n$ feasible solutions $G$. For properly chosen $W$, problem (784) becomes an equivalent to (782). Thus the purpose of vector inner-product objective (785) is to locate a rank-$n$ feasible Gram matrix assumed existent on the boundary of positive semidefinite cone $\mathbb{S}^{N}_+$, as explained beginning in §4.4.1; how to choose direction vector $W$ is explained there and in what follows:

direction matrix $W$

Denote by $Z^*$ an optimal composite matrix from semidefinite program (784). Then for $Z^* \in \mathbb{S}^{N+n}_+$ whose eigenvalues $\lambda(Z^*) \in \mathbb{R}^{N+n}$ are arranged in nonincreasing order, (Ky Fan)

\[
\sum_{i=n+1}^{N+n} \lambda(Z^*)_i = \minimize_{W \in \mathbb{S}^{N+n}_+} \langle Z^*, W \rangle \quad (1792a)
\]

\[
\text{subject to} \quad 0 \preceq W \preceq I
\]

\[
\text{tr} W = N
\]

which has an optimal solution that is known in closed form (p.565, §4.4.1.1). This eigenvalue sum is zero when $Z^*$ has rank $n$ or less.

Foreknowledge of optimal $Z^*$, to make possible this search for $W$, implies iteration; id est, semidefinite program (784) is solved for $Z^*$ initializing $W = I$ or $W = 0$. Once found, $Z^*$ becomes constant in semidefinite program (1792a) where a new normal direction $W$ is found as its optimal solution. Then this cycle (784) (1792a) iterates until convergence. When rank $Z^* = n$, solution via this convex iteration solves sensor-network localization problem (777) and its equivalent (782).
Figure 95: Typical solution for 2-lattice in Figure 90 with noise factor $\eta = 0.1$. Two red rightmost nodes are anchors; two remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius = 1.14. Actual sensor indicated by target $\bigcirc$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 1 iteration (784) (1792a) subject to reflection error.

Figure 96: Typical solution for 3-lattice in Figure 91 with noise factor $\eta = 0.1$. Three red vertical middle nodes are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius = 1.12. Actual sensor indicated by target $\bigcirc$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 2 iterations (784) (1792a).
Figure 97: Typical solution for 4-lattice in Figure 92 with noise factor $\eta = 0.1$. Four red vertical middle-left nodes are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius = 0.75. Actual sensor indicated by target $\circ$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 7 iterations (784) (1792a).

Figure 98: Typical solution for 5-lattice in Figure 93 with noise factor $\eta = 0.1$. Five red vertical middle nodes are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius = 0.56. Actual sensor indicated by target $\circ$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 3 iterations (784) (1792a).
4.4. RANK-CONSTRAINED SEMIDEFINITE PROGRAM

Figure 99: Typical solution for 10-lattice with noise factor $\eta = 0.1$ compares better than Carter & Jin [74, fig.4.2]. Ten red vertical middle nodes are anchors; the rest are sensors. Radio range of sensor 1 indicated by arc; radius = 0.25. Actual sensor indicated by target $\bigcirc$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 5 iterations (784) (1792a).

Figure 100: Typical localization of 100 randomized noiseless sensors ($\eta = 0$) is exact despite incomplete EDM. Ten red vertical middle nodes are anchors; remaining nodes are sensors. Radio range of sensor at origin indicated by arc; radius = 0.25. Actual sensor indicated by target $\bigcirc$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 3 iterations (784) (1792a).
Figure 101: Typical solution for 100 randomized sensors with noise factor $\eta = 0.1$; worst measured average sensor error $\approx 0.0044$ compares better than Carter & Jin's $0.0154$ computed in 0.71s [74, p.19]. Ten red vertical middle nodes are anchors; same as before. Remaining nodes are sensors. Interior anchor placement makes localization difficult. Radio range of sensor at origin indicated by arc; radius = 0.25. Actual sensor indicated by target $\circ$ while its localization is indicated by bullet $\bullet$. After 1 iteration rank $G = 92$, after 2 iterations rank $G = 4$. Rank-2 solution found in 3 iterations (784) (1792a). (Regular lattice in Figure 99 is actually harder to solve, requiring more iterations.) Runtime for SDPT3 [360] under cvx [177] is a few minutes on 2009 vintage laptop Core 2 Duo CPU (Intel T6400@2GHz, 800MHz FSB).
4.4. RANK-CONSTRAINED SEMIDEFINITE PROGRAM

In all examples to follow, number of anchors

\[ m = \sqrt{N} \]  

equals square root of cardinality \( N \) of list \( X \). Indices set \( I \) identifying all measurable distances \( \bullet \) is ascertained from connectivity matrix (778), (779), (780), or (781). We solve iteration (784) (1792a) in dimension \( n = 2 \) for each respective example illustrated in Figure 90 through Figure 93.

In presence of negligible noise, true position is reliably localized for every standardized example; noteworthy insofar as each example represents an incomplete graph. This implies that the set of all optimal solutions having least rank must be small.

To make the examples interesting and consistent with previous work, we randomize each range of distance-square that bounds \((G, (e_i - e_j)(e_i - e_j)^T)\) in (784); id est, for each and every \((i, j) \in I\)

\[
\begin{align*}
\overline{d_{ij}} &= d_{ij}(1 + \sqrt{3} \eta \chi_i)^2 \\
\underline{d_{ij}} &= d_{ij}(1 - \sqrt{3} \eta \chi_{i+1})^2
\end{align*}
\]

where \( \eta = 0.1 \) is a constant noise factor, \( \chi_i \) is the \( i \)th sample of a noise process realization uniformly distributed in the interval \((0, 1)\) like rand(1) from MATLAB, and \( d_{ij} \) is actual distance-square from \( i \)th to \( j \)th sensor. Because of distinct function calls to rand(), each range of distance-square \([\underline{d_{ij}}, \overline{d_{ij}}]\) is not necessarily centered on actual distance-square \( d_{ij} \). Unit stochastic variance is provided by factor \( \sqrt{3} \).

Figure 95 through Figure 98 each illustrate one realization of numerical solution to the standardized lattice problems posed by Figure 90 through Figure 93 respectively. Exact localization, by any method, is impossible because of measurement noise. Certainly, by inspection of their published graphical data, our results are better than those of Carter & Jin. (Figure 99, 100, 101) Obviously our solutions do not suffer from those compaction-type errors (clustering of localized sensors) exhibited by Biswas’ graphical results for the same noise factor \( \eta \).

localization example conclusion

Solution to this sensor-network localization problem became apparent by understanding geometry of optimization. Trace of a matrix, to a student of linear algebra, is perhaps a sum of eigenvalues. But to us, trace represents the normal \( I \) to some hyperplane in Euclidean vector space. (Figure 88)

Our solutions are globally optimal, requiring: 1) no centralized-gradient postprocessing heuristic refinement as in [45] because there is effectively no relaxation of (782) at global optimality, 2) no implicit postprojection on rank-2 positive semidefinite matrices induced by nonzero \( G - X^TX \) denoting suboptimality as occurs in [46] [47] [48] [74] [226] [234]; indeed, \( G^* = X^*X^* \) by convex iteration.

Numerical solution to noisy problems, containing sensor variables well in excess of 100, becomes difficult via the holistic semidefinite program we proposed. When problem size is within reach of contemporary general-purpose semidefinite program solvers, then
the convex iteration we presented inherently overcomes limitations of Carter & Jin with respect to both noise performance and ability to localize in any desired affine dimension.

The legacy of Carter, Jin, Saunders, & Ye [74] is a sobering demonstration of the need for more efficient methods for solution of semidefinite programs, while that of So & Ye [331] forever bonds distance geometry to semidefinite programming. Elegance of our semidefinite problem statement (784), for constraining affine dimension of sensor-network localization, should provide some impetus to focus more research on computational intensity of general-purpose semidefinite program solvers. An approach different from interior-point methods is required; higher speed and greater accuracy from a simplex-like solver is what is needed.

\[ 4.4.1.2.5 \text{ Example.} \quad \text{Nonnegative spectral factorization.} \]

Having found optimal real coefficient vectors \( v^*, u^* \) for a sixteenth order magnitude square transfer function, evaluated along the \( j\omega \) axis (p.229),

\[
|H(j\omega)|^2 = H(j\omega)H(-j\omega) = \frac{1 + v_1^2\omega^2 + v_2^2\omega^4 + \ldots + v_8^2\omega^{16}}{1 + u_1^2\omega^2 + u_2^2\omega^4 + \ldots + u_8^2\omega^{16}}
\]

we wish to find real coefficients \( b, a \) for corresponding Fourier transform

\[
H(j\omega) = \frac{1 + b_1j\omega + b_2(j\omega)^2 + \ldots + b_8(j\omega)^8}{1 + a_1j\omega + a_2(j\omega)^2 + \ldots + a_8(j\omega)^8}
\]

These coefficients \( b, a, v^*, u^* \) are related through simultaneous nonlinear algebraic equations:

\[
\begin{align*}
v_1^* &= b_1^2 - 2b_2, \\
v_2^* &= b_2^2 - 2b_1b_3 + 2b_4, \\
v_3^* &= b_3^2 - 2b_4b_5 + 2b_1b_5 - 2b_6, \\
v_4^* &= b_4^2 - 2b_5b_6 + 2b_2b_6, \\
v_5^* &= b_5^2 - 2b_6b_7 - 2b_5b_7, \\
v_6^* &= b_6^2 - 2b_5b_7, \\
v_7^* &= b_7^2 - 2b_5b_8, \\
v_8^* &= b_8^2
\end{align*}
\]

These equations are specified by a particular sum is specified by a

\[
G(b) \triangleq \begin{bmatrix} 1 & b \end{bmatrix} \begin{bmatrix} 1 & b^T \end{bmatrix} \in S^9
\]

(Matrix \( G(a) \) is similarly defined.) Observe that \( v^* \) in (788) is formed by summing antidiagonals of \( G(b) \) whose entries alternate sign. A particular sum is specified by a
4.4. RANK-CONSTRAINED SEMIDEFINITE PROGRAM

Figure 102: Nonnegative spectral factorization, high order bisection strategy. \( \eta = 8 \)th order Laplace transform corresponds to \( 2\eta = 16 \)th order magnitude square transfer function. Because numerator \( v \) and denominator \( u \) are factored separately, number of factorizations \( = 2(\log_2(\eta) - 1) \). In the text, double dots \( \bar{v}, \bar{u} \) connote first bifurcation (level 2). Triple dots \( \bar{v}, \bar{u} \) connote second bifurcations (level 3). Factors per level \( = 2^{\text{level} - 1} \).

predetermined symmetric constant matrix \( A_i \) (confer (57)) from a set \( \{ A_i \in \mathbb{S}^9, \ i = 1 \ldots 8 \} \). With

\[
A = \begin{bmatrix} \text{svec}(A_1)^T \\ \vdots \\ \text{svec}(A_8)^T \end{bmatrix} \in \mathbb{R}^{8 \times 9(9+1)/2} \quad (681)
\]
as previously defined in §4.1.1, all the sums (788) may be stated as two linear equalities \( A \text{svec}(G(b)) = v^* \) and \( A \text{svec}(G(a)) = u^* \). Then the problem of finding coefficients \( b \) may be stated as a feasibility problem\(^{4.32}\)

\[
\begin{align*}
\text{find} & \quad b \in \mathbb{R}^8 \\
\text{subject to} & \quad A \text{svec}(G) = v^* \\
& \quad \begin{bmatrix} 1 \\ b \end{bmatrix} = G(:,1) \\
& \quad b \succeq 0 \\
& \quad (G \succeq 0) \\
& \quad \text{rank} G = 1 
\end{align*} \quad (790)
\]

The rank-one constraint is handled by convex iteration, as explained in §4.4.1. Positive semidefiniteness is parenthetical here because, for rank-one matrices, symmetry is necessary and sufficient (§A.3.1.0.7).

\(^{4.32}\) separately from the similar optimization problem to find vector \( a \). Stability requires \( a \succeq 0 \) with additional constraints on \( a \). Minimum phase requires \( b \succeq 0 \) plus more constraints on \( b \) that are missing from problem statement (790). Both stability and minimum phase may be enforced, subsequent to spectral factorization, by negating positive real parts of poles and zeros respectively in order to move them into the left half (Laplace) \( s \)-plane with no impact to \( |H(j\omega)| \).
4.4.1.2.6 Example. Nonnegative spectral factorization II.

The purpose of spectral factorization, in electronics, is to facilitate high order filter implementation in the form of passive and active circuitry. Cascades of second order (Laplace) sections are preferred because component sensitivity becomes manageable and because needed complex poles and zeros cannot be obtained from a first order section.

Nonnegative spectral factorization on a magnitude square transfer function, evaluated along the \( j\omega \) axis, was performed in Example 4.4.1.2.5 to recover its corresponding Fourier transform.\(^{4,33}\) In this example, we nonnegatively decompose a high order magnitude square transfer function into a product of successively lower order magnitude square transfer functions. Once fourth order magnitude square functions are found, then corresponding second order Laplace transfer function coefficients are ascertained from (656) and then passive component values can be determined from those coefficients.

Our strategy, for an eighth order Laplace transfer function, is illustrated in Figure 102. We begin at the tree’s level 2 factorization. Nonnegative decomposition of a 16th order magnitude square transfer function into two 8th order functions

\[
\frac{1 + v_1^* \omega^2 + v_2^* \omega^4 + \cdots + v_8^* \omega^{16}}{1 + u_1^* \omega^2 + u_2^* \omega^4 + \cdots + u_8^* \omega^{16}} = \frac{1 + \tilde{v}_1 \omega^2 + \tilde{v}_2 \omega^4 + \tilde{v}_3 \omega^6 + \tilde{v}_4 \omega^8}{1 + \tilde{u}_1 \omega^2 + \tilde{u}_2 \omega^4 + \tilde{u}_3 \omega^6 + \tilde{u}_4 \omega^8} \quad (791)
\]

implies these simultaneous algebraic identifications with known real coefficient vectors \( v^*, u^* \):

\[
\begin{align*}
v_1^* &= \tilde{v}_1 + \tilde{v}_5, & u_1^* &= \tilde{u}_1 + \tilde{u}_5 \\
v_2^* &= \tilde{v}_2 + \tilde{v}_6 + \tilde{v}_1 \tilde{v}_5, & u_2^* &= \tilde{u}_2 + \tilde{u}_6 + \tilde{u}_1 \tilde{u}_5 \\
v_3^* &= \tilde{v}_3 + \tilde{v}_7 + \tilde{v}_1 \tilde{v}_6 + \tilde{v}_2 \tilde{v}_5, & u_3^* &= \tilde{u}_3 + \tilde{u}_7 + \tilde{u}_1 \tilde{u}_6 + \tilde{u}_2 \tilde{u}_5 \\
v_4^* &= \tilde{v}_4 \tilde{v}_5 + \tilde{v}_1 \tilde{v}_7 + \tilde{v}_2 \tilde{v}_6 + \tilde{v}_1 \tilde{v}_8, & u_4^* &= \tilde{u}_4 \tilde{u}_5 + \tilde{u}_1 \tilde{u}_7 + \tilde{u}_2 \tilde{u}_6 + \tilde{u}_1 \tilde{u}_8 \\
v_5^* &= \tilde{v}_4 \tilde{v}_6 + \tilde{v}_3 \tilde{v}_7 + \tilde{v}_2 \tilde{v}_8, & u_5^* &= \tilde{u}_4 \tilde{u}_6 + \tilde{u}_3 \tilde{u}_7 + \tilde{u}_2 \tilde{u}_8 \\
v_6^* &= \tilde{v}_4 \tilde{v}_7 + \tilde{v}_3 \tilde{v}_8, & u_6^* &= \tilde{u}_4 \tilde{u}_7 + \tilde{u}_3 \tilde{u}_8 \\
v_7^* &= \tilde{v}_4 \tilde{v}_8, & u_7^* &= \tilde{u}_4 \tilde{u}_8
\end{align*}
\]

Now define a rank-one matrix for the numerator

\[
G(\tilde{v}) \triangleq \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} \tilde{v}^T \end{bmatrix} \in \mathbb{S}^9 \quad (793)
\]

(Matrix \( G(\tilde{v}) \) is defined similarly for the denominator.) Terms in (792) are picked out of \( G(\tilde{v}) \) by a predetermined symmetric constant matrix \( \tilde{A}_i \) (confer (57)) from a set

When there are no poles on the \( j\omega \) axis, a Laplace transform can be recovered from a Fourier transform by substitution \( j\omega \rightarrow s \).
\{ \tilde{A}_i \in \mathbb{S}^9, i = 1 \ldots 8 \}. Populating rows of

\[
A = \begin{bmatrix}
\text{svec}(\tilde{A}_1)^T \\
\vdots \\
\text{svec}(\tilde{A}_8)^T 
\end{bmatrix} \in \mathbb{R}^{8 \times 9(9+1)/2} \tag{681}
\]

with vectorized \( \tilde{A}_i \) (as in \( \S 4.1.1 \)), sums (792) are succinctly represented by two linear equalities \( A \text{svec}G(\tilde{v}) = \tilde{v}^* \) and \( A \text{svec}G(\tilde{u}) = \tilde{u}^* \). Then this spectral factorization in \( \tilde{v} \) may be posed as a feasibility problem

\[
\begin{align*}
\text{find} & \quad \tilde{v} \in \mathbb{R}^8 \\
\text{subject to} & \quad A \text{svec}G = \tilde{v}^* \\
& \quad \begin{bmatrix} 1 & \tilde{v} \end{bmatrix} = G(:,1) \\
& \quad \tilde{v} \succeq 0 \\
& \quad (G \succeq 0) \\
& \quad \text{rank } G = 1 
\end{align*} \tag{794}
\]

Having found two 8th order square spectral factors in nonnegative \( \tilde{v}^* \) from (794), two pairs of 4th order level 3 factors remain to be found:

\[
\begin{align*}
1 + \tilde{u}_1^* \omega^2 + \tilde{v}_2^* \omega^4 + \tilde{v}_3^* \omega^6 + \tilde{v}_4^* \omega^8 & = 1 + \tilde{v}_1^* \omega^2 + \tilde{v}_2^* \omega^4 + \tilde{v}_3^* \omega^6 + \tilde{v}_4^* \omega^8 \\
1 + \tilde{u}_5^* \omega^2 + \tilde{v}_6^* \omega^4 + \tilde{v}_7^* \omega^6 + \tilde{v}_8^* \omega^8 & = 1 + \tilde{v}_5^* \omega^2 + \tilde{v}_6^* \omega^4 + \tilde{v}_7^* \omega^6 + \tilde{v}_8^* \omega^8 \\ 
1 + \tilde{u}_1^* \omega^2 + \tilde{u}_2^* \omega^4 + \tilde{u}_3^* \omega^6 + \tilde{u}_4^* \omega^8 & = 1 + \tilde{u}_1^* \omega^2 + \tilde{u}_2^* \omega^4 + \tilde{u}_3^* \omega^6 + \tilde{u}_4^* \omega^8 \\
1 + \tilde{u}_5^* \omega^2 + \tilde{u}_6^* \omega^4 + \tilde{u}_7^* \omega^6 + \tilde{u}_8^* \omega^8 & = 1 + \tilde{u}_5^* \omega^2 + \tilde{u}_6^* \omega^4 + \tilde{u}_7^* \omega^6 + \tilde{u}_8^* \omega^8 
\end{align*} \tag{795}
\]

\[
\begin{align*}
\tilde{v}_1^* & = \tilde{v}_1 + \tilde{v}_3 \\
\tilde{v}_2^* & = \tilde{v}_2 + \tilde{v}_4 + \tilde{v}_1 \tilde{v}_3 \\
\tilde{v}_3^* & = \tilde{v}_1 \tilde{v}_4 + \tilde{v}_2 \tilde{v}_3 \\
\tilde{v}_4^* & = \tilde{v}_2 \tilde{v}_4 \\
\tilde{v}_5^* & = \tilde{v}_5 + \tilde{v}_7 \\
\tilde{v}_6^* & = \tilde{v}_6 + \tilde{v}_8 + \tilde{v}_5 \tilde{v}_7 \\
\tilde{v}_7^* & = \tilde{v}_5 \tilde{v}_8 + \tilde{v}_6 \tilde{v}_7 \\
\tilde{v}_8^* & = \tilde{v}_6 \tilde{v}_8 \\
\tilde{u}_1^* & = \tilde{u}_1 + \tilde{u}_3 \\
\tilde{u}_2^* & = \tilde{u}_2 + \tilde{u}_4 + \tilde{u}_1 \tilde{u}_3 \\
\tilde{u}_3^* & = \tilde{u}_1 \tilde{u}_4 + \tilde{u}_2 \tilde{u}_3 \\
\tilde{u}_4^* & = \tilde{u}_2 \tilde{u}_4 \\
\tilde{u}_5^* & = \tilde{u}_5 + \tilde{u}_7 \\
\tilde{u}_6^* & = \tilde{u}_6 + \tilde{u}_8 + \tilde{u}_5 \tilde{u}_7 \\
\tilde{u}_7^* & = \tilde{u}_5 \tilde{u}_8 + \tilde{u}_6 \tilde{u}_7 \\
\tilde{u}_8^* & = \tilde{u}_6 \tilde{u}_8 
\end{align*} \tag{796}
\]

\[
G(\tilde{v}) \triangleq \begin{bmatrix} 1 & \tilde{v}_1 & \tilde{v}_2 & \tilde{v}_3 & \tilde{v}_4 & \tilde{v}_5 & \tilde{v}_6 & \tilde{v}_7 & \tilde{v}_8 \\
\tilde{v}_1 & \tilde{v}_1^2 & \tilde{v}_1 \tilde{v}_2 & \tilde{v}_1 \tilde{v}_3 & \tilde{v}_1 \tilde{v}_4 & \tilde{v}_1 \tilde{v}_5 & \tilde{v}_1 \tilde{v}_6 & \tilde{v}_1 \tilde{v}_7 & \tilde{v}_1 \tilde{v}_8 \\
\tilde{v}_2 & \tilde{v}_1 \tilde{v}_2 & \tilde{v}_2^2 & \tilde{v}_2 \tilde{v}_3 & \tilde{v}_2 \tilde{v}_4 & \tilde{v}_2 \tilde{v}_5 & \tilde{v}_2 \tilde{v}_6 & \tilde{v}_2 \tilde{v}_7 & \tilde{v}_2 \tilde{v}_8 \\
\tilde{v}_3 & \tilde{v}_1 \tilde{v}_3 & \tilde{v}_2 \tilde{v}_3 & \tilde{v}_3^2 & \tilde{v}_3 \tilde{v}_4 & \tilde{v}_3 \tilde{v}_5 & \tilde{v}_3 \tilde{v}_6 & \tilde{v}_3 \tilde{v}_7 & \tilde{v}_3 \tilde{v}_8 \\
\tilde{v}_4 & \tilde{v}_1 \tilde{v}_4 & \tilde{v}_2 \tilde{v}_4 & \tilde{v}_3 \tilde{v}_4 & \tilde{v}_4^2 & \tilde{v}_4 \tilde{v}_5 & \tilde{v}_4 \tilde{v}_6 & \tilde{v}_4 \tilde{v}_7 & \tilde{v}_4 \tilde{v}_8 \\
\tilde{v}_5 & \tilde{v}_1 \tilde{v}_5 & \tilde{v}_2 \tilde{v}_5 & \tilde{v}_3 \tilde{v}_5 & \tilde{v}_4 \tilde{v}_5 & \tilde{v}_5^2 & \tilde{v}_5 \tilde{v}_6 & \tilde{v}_5 \tilde{v}_7 & \tilde{v}_5 \tilde{v}_8 \\
\tilde{v}_6 & \tilde{v}_1 \tilde{v}_6 & \tilde{v}_2 \tilde{v}_6 & \tilde{v}_3 \tilde{v}_6 & \tilde{v}_4 \tilde{v}_6 & \tilde{v}_5 \tilde{v}_6 & \tilde{v}_6^2 & \tilde{v}_6 \tilde{v}_7 & \tilde{v}_6 \tilde{v}_8 \\
\tilde{v}_7 & \tilde{v}_1 \tilde{v}_7 & \tilde{v}_2 \tilde{v}_7 & \tilde{v}_3 \tilde{v}_7 & \tilde{v}_4 \tilde{v}_7 & \tilde{v}_5 \tilde{v}_7 & \tilde{v}_6 \tilde{v}_7 & \tilde{v}_7^2 & \tilde{v}_7 \tilde{v}_8 \\
\tilde{v}_8 & \tilde{v}_1 \tilde{v}_8 & \tilde{v}_2 \tilde{v}_8 & \tilde{v}_3 \tilde{v}_8 & \tilde{v}_4 \tilde{v}_8 & \tilde{v}_5 \tilde{v}_8 & \tilde{v}_6 \tilde{v}_8 & \tilde{v}_7 \tilde{v}_8 & \tilde{v}_8^2 
\end{bmatrix} \in \mathbb{S}^9 \tag{797}
\]
CHAPTER 4. SEMIDEFINITE PROGRAMMING

Figure 103: Regularization curve, parametrized by weight $w$ for real convex objective $f$ minimization (801) with rank constraint to $k$ by convex iteration.

Setting

\[
A = \begin{bmatrix}
\text{svec}(A_1)^T \\
\vdots \\
\text{svec}(A_8)^T
\end{bmatrix} \in \mathbb{R}^{8 \times (9+1)/2}
\] (681)

then all level 3 (Figure 102) nonnegative spectral factorization coefficients $\bar{v}$ are found at once by solving

\[
\begin{align*}
\text{find } & \bar{v} \in \mathbb{R}^8 \\
\text{subject to } & A \text{svec } G = \bar{v}^* \\
& \begin{bmatrix} 1 \\ \bar{v} \end{bmatrix} = G(:,1) \\
& \bar{v} \succeq 0 \\
& (G \succeq 0) \\
& \text{rank } G = 1
\end{align*}
\] (800)

The feasibility problem to find $\bar{u}$ is similar. All second order Laplace transfer function coefficients can be found via (656).

4.4.1.3 regularization

We test the convex iteration technique, for constraining rank, over a wide range of problems beyond localization of randomized positions (Figure 101); e.g., stress (§7.2.2.7.1), ball
packing (§5.4.2.6), and cardinality (§4.6). We have had some success introducing the direction matrix inner-product (785) as a regularization term

\[
\begin{align*}
\text{minimize} & \quad f(Z) + w(Z, W) \\
\text{subject to} & \quad Z \in C \\
& \quad Z \succeq 0
\end{align*}
\]

(801)

\[
\begin{align*}
\text{minimize} & \quad f(Z^*) + w(Z^*, W) \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \text{tr } W = N - n
\end{align*}
\]

(802)

whose purpose is to constrain rank, affine dimension, or cardinality:

The abstraction, that is Figure 103, is a synopsis; a broad generalization of accumulated empirical evidence: There exists a critical (smallest) weight \( w_c \) for which a minimal-rank constraint is met. Graphical discontinuity can subsequently exist when there is a range of greater \( w \) providing required rank \( k \) but not necessarily increasing a minimization objective function \( f \); e.g., §4.6.0.0.2. Positive scalar \( w \) is well chosen by cut-and-try.

## 4.5 Constraining cardinality

The convex iteration technique for constraining rank can be applied to cardinality problems. There are parallels in its development analogous to how prototypical semidefinite program (680) resembles linear program (301) on page 238 [414]:

### 4.5.1 nonnegative variable

Our goal has been to reliably constrain rank in a semidefinite program. There is a direct analogy to linear programming that is simpler to present but, perhaps, more difficult to solve. In Optimization, that analogy is known as the cardinality problem.

Consider a feasibility problem \( Ax = b \), but with an upper bound \( k \) on cardinality \( \|x\|_0 \) of a nonnegative solution \( x \): for \( A \in \mathbb{R}^{m \times n} \) and vector \( b \in \mathbb{R}^m \)

\[
\begin{align*}
\text{find} & \quad x \in \mathbb{R}^n \\
\text{subject to} & \quad Ax = b \\
& \quad x \succeq 0 \\
& \quad \|x\|_0 \leq k
\end{align*}
\]

(529)

where \( \|x\|_0 \leq k \) means vector \( x \) has at most \( k \) nonzero entries; such a vector is presumed existent in the feasible set. Nonnegativity constraint \( x \succeq 0 \) is analogous to positive semidefiniteness; the notation means vector \( x \) belongs to the nonnegative orthant \( \mathbb{R}^n_+ \). Cardinality is quasiconcave on \( \mathbb{R}^n_+ \) just as rank is quasiconcave on \( \mathbb{S}^n_+ \). [63, §3.4.2]

\[\text{multiobjective- or vector optimization. Proof of convergence for this convex iteration is identical to that in §4.4.1.2.1 because } f \text{ is a convex real function, hence bounded below, and } f(Z^*) \text{ is constant in (802).}\]

\[\text{Although it is a metric (§5.2), cardinality } \|x\|_0 \text{ cannot be a norm (§3.2) because it is not positively homogeneous.}\]
4.5.1.1  direction vector

We propose that cardinality-constrained feasibility problem (529) can be equivalently expressed as iteration of a sequence of two convex problems: for $0 \leq k \leq n - 1$

\[
\begin{align*}
\text{minimize} & \quad \langle x, y \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]  
(156)

\[
\begin{align*}
\sum_{i=k+1}^{n} \pi(x^*)_i = & \quad \text{minimize} \quad \langle x^*, y \rangle \\
\text{subject to} & \quad 0 \leq y \leq 1 \\
& \quad y^T 1 = n - k
\end{align*}
\]  
(524)

where $\pi$ is the (nonincreasing) presorting function. This sequence is iterated until $x^T y^*$ vanishes; \emph{id est}, until desired cardinality is achieved. But this \emph{global convergence} cannot always be guaranteed.\footnote{When it succeeds, a sequence may be regarded as a \emph{homotopy} to minimal 0-norm.}

Problem (524) is analogous to the rank constraint problem; (p. 264)

\[
\sum_{i=k+1}^{N} \lambda(G^*)_i = \text{minimize} \quad \langle G^*, W \rangle \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \text{tr} W = N - k
\]  
(1792a)

Linear program (524) sums the $n - k$ smallest entries from vector $x$. In context of problem (529), we want $n - k$ entries of $x$ to sum to zero; \emph{id est}, we want a globally optimal objective $x^T y^*$ to vanish: more generally, (confer (768))

\[
\sum_{i=k+1}^{n} \pi(|x^*_i|) = \langle |x^*| , y^* \rangle = |x^*|^T y^* \triangleq 0
\]  
(803)

defines \emph{global convergence} for the iteration. Then $n - k$ entries of $x^*$ are themselves zero whenever their absolute sum is, and cardinality of $x^* \in \mathbb{R}^n$ is at most $k$. \emph{Optimal direction vector} $y^*$ is defined as any nonnegative vector for which

\[
\begin{align*}
\text{find} & \quad x \in \mathbb{R}^n \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0 \\
& \quad \|x\|_0 \leq k
\end{align*}
\]  
(529)

\[
\begin{align*}
\text{minimize} & \quad \langle x , y^* \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]  
(156)

Existence of such a $y^*$, whose nonzero entries are complementary to those of $x^*$, is obvious assuming existence of a cardinality-$k$ solution $x^*$. 
4.5. CONSTRAINING CARDINALITY

Figure 104: (confer Figure 88) 1-norm heuristic for cardinality minimization can be interpreted as minimization of a hyperplane, $\partial H$ with normal $\mathbf{1}$, over nonnegative orthant drawn here in $\mathbb{R}^3$. Polar of direction vector $y = \mathbf{1}$ points toward origin.

### 4.5.1.2 Direction vector interpretation

(confer §4.4.1.1) Vector $y$ may be interpreted as a negative search direction; it points opposite to direction of movement of hyperplane $\{x \mid \langle x, y \rangle = \tau\}$ in a minimization of real linear function $\langle x, y \rangle$ over the feasible set in linear program (156). (p.67) Direction vector $y$ is not unique. The feasible set of direction vectors in (524) is the convex hull of all cardinality-$(n-k)$ one-vectors; videlicet,

\[
\text{conv}\{u \in \mathbb{R}^n \mid \text{card } u = n-k, \ u_i \in \{0,1\}\} = \{a \in \mathbb{R}^n \mid \mathbf{1} \succeq a \succeq 0, \ \langle \mathbf{1}, a \rangle = n-k\}
\]

This set, argument to conv{ }, comprises the extreme points of set (804) which is a nonnegative hypercube slice. An optimal solution $y$ to (524), that is an extreme point of its feasible set, is known in closed form: it has 1 in each entry corresponding to the $n-k$ smallest entries of $x^*$ and has 0 elsewhere. That particular polar direction $-y$ can be interpreted\(^{4.37}\) by Proposition 7.1.3.0.3 as pointing toward the nonnegative orthant in the Cartesian subspace, whose basis is a subset of the Cartesian axes, containing all cardinality $k$ (or less) vectors having the same ordering as $x^*$. Consequently, for that closed-form solution, (confer (769))

\[
\sum_{i=k+1}^{n} \pi(|x^*|_i) = \langle |x^*|, y \rangle = |x^*|^Ty \geq 0
\]

\(^{4.37}\)Convex iteration (156) (524) is not a projection method because there is no thresholding or discard of variable-vector $x$ entries. An optimal direction vector $y$ must always reside on the feasible set boundary in (524) page 290; id est, it is ill-advised to attempt simultaneous optimization of variables $x$ and $y$. 

\[
\partial H = \{x \mid \langle x, \mathbf{1} \rangle = \kappa\}
\]
When \( y = 1 \), as in 1-norm minimization for example, then polar direction \(-y\) points directly at the origin (the cardinality-0 nonnegative vector) as in Figure 104. We sometimes solve (524) instead of employing a known closed form because a direction vector is not unique. Setting direction vector \( y \) instead in accordance with an iterative inverse weighting scheme, called \textit{reweighting} [170], was described for the 1-norm by Huo [217, §4.11.3] in 1999.

### 4.5.1.3 convergence can mean stalling

Convex iteration (156) (524) always converges to a \textit{locally optimal solution}, a fixed point of possibly infeasible cardinality, by virtue of a monotonically nonincreasing real objective sequence. [268, §1.2] [43, §1.1] There can be no proof of global convergence, defined by (803). Constraining cardinality, solution to problem (529), can often be achieved but simple examples can be contrived that \textit{stall} at a fixed point of infeasible cardinality; at a positive objective value \( \langle x^*, y \rangle = \tau > 0 \). Direction vector \( y \) is then manipulated, as countermeasure, to steer out of local minima; \textit{e.g.}, complete randomization as in Example 4.5.1.5.1, or reinitialization to a random cardinality-\((n-k)\) vector in the same nonnegative orthant face demanded by the current iterate: \( y \) has nonnegative uniformly distributed random entries in \((0, 1]\) corresponding to the \( n-k \) smallest entries of \( x^* \) and has 0 elsewhere. Zero entries behave like memory or state while randomness greatly diminishes likelihood of a stall. When this particular heuristic is successful, cardinality and objective sequence \( \langle x^*, y \rangle \) \textit{versus} iteration are characterized by noisy monotonicity.

### 4.5.1.4 algebraic derivation of direction vector for convex iteration

In §3.2.2.1.3, the compressed sensing problem was precisely represented as a nonconvex difference of convex functions bounded below by 0

\[
\begin{align*}
\text{find} & \quad x \in \mathbb{R}^n \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0 \\
& \quad \|x\|_0 \leq k
\end{align*}
\]

\[
\equiv \quad \minimize_{x \in \mathbb{R}^n} \|x\|_1 - \|x\|_n \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\]

(529)

where convex \( k \)-largest norm \( \|x\|_n \) is monotonic on \( \mathbb{R}^n_+ \). There we showed how (529) is equivalently stated in terms of gradients

\[
\begin{align*}
\minimize_{x \in \mathbb{R}^n} & \quad \langle x, \nabla \|x\|_1 - \nabla \|x\|_n \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad x \geq 0
\end{align*}
\]

(806)

because

\[
\|x\|_1 = x^T \nabla \|x\|_1, \quad \|x\|_n = x^T \nabla \|x\|_n, \quad x \geq 0
\]

(807)

The objective function from (806) is a directional derivative (at \( x \) in direction \( x \), §D.1.6, \textit{confer} §D.1.4.1.1) of the objective function from (529) while the direction vector of convex iteration

\[
y = \nabla \|x\|_1 - \nabla \|x\|_n
\]

(808)
is an objective gradient where \( \nabla \|x\|_1 = \nabla 1^T x = 1 \) under nonnegativity and

\[
\nabla \|x\|_k = \nabla z^T x = \arg \max_{z \in \mathbb{R}^k} z^T x \quad \text{subject to} \quad 0 \preceq z \preceq 1, \quad z^T 1 = k \]

is not unique. Substituting \( 1 - z \leftarrow z \) the direction vector becomes

\[
y = 1 - \arg \max_{z \in \mathbb{R}^n} z^T x \quad \leftarrow \arg \min_{z \in \mathbb{R}^n} z^T x \quad \text{subject to} \quad 0 \preceq z \preceq 1, \quad z^T 1 = k \]

4.5.1.5 optimality conditions for compressed sensing

Now we see how global optimality conditions can be stated without reference to a dual problem: From conditions (468) for optimality of (529), it is necessary [63, §5.5.3] that

\[
x^* \succeq 0 \quad (1) \quad Ax^* = b \quad (2) \quad \nabla \|x^*\|_1 - \nabla \|x^*\|_n + A^T \nu^* \succeq 0 \quad (3) \quad (809)\]

\[
\langle \nabla \|x^*\|_1 - \nabla \|x^*\|_n + A^T \nu^*, x^* \rangle = 0 \quad (4g) \]

These conditions must hold at any optimal solution (locally or globally). By (807), the fourth condition is identical to

\[
\|x^*\|_1 - \|x^*\|_n + \nu^T A x^* = 0 \quad (4f) \quad (810)\]

Because a 1-norm

\[
\|x\|_1 = \|x\|_n + \|\pi(|x|)_{k+1:n}\|_1 \quad (811)\]

is separable into \( k \) largest and \( n - k \) smallest absolute entries,

\[
\|\pi(|x|)_{k+1:n}\|_1 = 0 \iff \|x\|_0 \leq k \quad (4g) \quad (812)\]

is a necessary condition for global optimality. By assumption, matrix \( A \) is fat and \( b \neq 0 \Rightarrow Ax^* \neq 0 \). This means \( \nu^* \in \mathcal{N}(A^T) \subset \mathbb{R}^m \), and \( \nu^* = 0 \) when \( A \) is full-rank. By definition, \( \nabla \|x\|_1 \succeq \nabla \|x\|_n \) always holds. Assuming existence of a cardinality-\( k \) solution, then only three of the four conditions are necessary and sufficient for global optimality of (529):

\[
x^* \succeq 0 \quad (1) \quad Ax^* = b \quad (2) \quad \|x^*\|_1 - \|x^*\|_n = 0 \quad (4g) \quad (813)\]

meaning, global optimality of a feasible solution to (529) is identified by a zero objective.
Figure 105: For Gaussian random matrix $A \in \mathbb{R}^{m \times n}$, graph illustrates Donoho/Tanner least lower bound on number of measurements $m$ below which recovery of $k$-sparse $n$-length signal $x$ by linear programming fails with overwhelming probability. Hard problems are below curve, but not the reverse; id est, failure above depends on proximity. Inequality demarcates approximation (dashed curve) empirically observed in [24]. Problems having nonnegativity constraint (dotted) are easier to solve. [129] [130]
4.5. CONSTRAINING CARDINALITY

4.5.1.5.1 Example. Sparsest solution to $Ax = b$.

(Confer Example 4.5.2.0.4) Problem (713) has sparsest solution not easily recoverable by least 1-norm; id est, not by compressed sensing because of proximity to a theoretical lower bound on number of measurements $m$ depicted in Figure 105: for $A \in \mathbb{R}^{m \times n}$

- Given data from Example 4.2.3.1.1, for $m = 3$, $n = 6$, $k = 1$

  
  \[
  A = \begin{bmatrix}
    -1 & 1 & 8 & 1 & 1 & 0 \\
    -3 & 2 & 8 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\
    -9 & 4 & 8 & \frac{1}{4} & \frac{1}{2} & \frac{1}{2}
  \end{bmatrix}, \quad b = \begin{bmatrix}
    1 \\
    \frac{1}{2} \\
    \frac{1}{4}
  \end{bmatrix}
  \]  

  (713)

  the sparsest solution to classical linear equation $Ax = b$ is $x = e_4 \in \mathbb{R}^6$ (confer (726)).

Although the sparsest solution is recoverable by inspection, we discern it instead by convex iteration; namely, by iterating problem sequence (156) (524) on page 290. From the numerical data given, cardinality $\|x\|_0 = 1$ is expected. Iteration continues until $x^T y$ vanishes (to within some numerical precision); id est, until desired cardinality is achieved. But this comes not without a stall.

Stalling, whose occurrence is sensitive to initial conditions of convex iteration, is a consequence of finding a local minimum of a multimodal objective $(x, y)$ when regarded as simultaneously variable in $x$ and $y$. (§3.8.0.0.3) Stalls are simply detected as fixed points $x$ of infeasible cardinality, sometimes remedied by reinitializing direction vector $y$ to a random positive state.

Bolstered by success in breaking out of a stall, we then apply convex iteration to 22,000 randomized problems:

- Given random data for $m = 3$, $n = 6$, $k = 1$, in MATLAB notation

  \[
  A = \text{randn}(3, 6), \quad \text{index} = \text{round}(5 \ast \text{rand}(1)) + 1, \quad b = \text{rand}(1) \ast A(:, \text{index})
  \]  

  (814)

  the sparsest solution $x \propto e_{\text{index}}$ is a scaled standard basis vector.

Without convex iteration or a nonnegativity constraint $x \succeq 0$, rate of failure for this minimal cardinality problem $Ax = b$ by 1-norm minimization of $x$ is 22%. That failure rate drops to 6% with a nonnegativity constraint. If we then engage convex iteration, detect stalls, and randomly reinitialize the direction vector, failure rate drops to 0% but the amount of computation is approximately doubled.

Stalling is not an inevitable behavior. For some problem types (beyond mere $Ax = b$), convex iteration succeeds nearly all the time. Here is a cardinality problem, with noise, whose statement is just a bit more intricate but easy to solve in a few convex iterations:

4.5.1.5.2 Example. Signal dropout.

(Confer §6.2) Signal dropout is an old problem; well studied from both an industrial and academic perspective. Essentially dropout means momentary loss or gap in a signal, while passing through some channel, caused by some man-made or natural phenomenon. The signal lost is assumed completely destroyed somehow. What remains within the time-gap is system or idle channel noise. The signal could be voice over Internet protocol (VoIP), for
example, audio data from a compact disc (CD) or video data from a digital video disc (DVD), a television transmission over cable or the airwaves, or a typically ravaged cell phone communication, etcetera.

Here we consider signal dropout in a discrete-time signal corrupted by additive white noise assumed uncorrelated to the signal. The linear channel is assumed to introduce no filtering. We create a discretized windowed signal for this example by positively combining $k$ randomly chosen vectors from a discrete cosine transform (DCT) basis denoted $\Psi \in \mathbb{R}^{n \times n}$. Frequency increases, in the Fourier sense, from DC toward Nyquist as column index of basis $\Psi$ increases. Otherwise, details of the basis are unimportant except for its orthogonality $\Psi^T = \Psi^{-1}$. Transmitted signal is denoted

$$s = \Psi z \in \mathbb{R}^n$$

whose upper bound on DCT basis coefficient cardinality $\text{card} \ z \leq k$ is assumed known, hence a critical assumption: transmitted signal $s$ is sparsely supported ($k < n$) on the DCT basis. It is further assumed that nonzero signal coefficients in vector $z$ place each chosen basis vector above the noise floor.

We also assume that the gap’s beginning and ending in time are precisely localized to within a sample; id est, index $\ell$ locates the last sample prior to the gap’s onset, while index $n - \ell + 1$ locates the first sample subsequent to the gap: for rectangularly windowed received signal $g$ possessing a time-gap loss and additive noise $\eta \in \mathbb{R}^n$

$$g = \begin{bmatrix} s_{1:\ell} + \eta_{1:\ell} \\ s_{n-\ell+1:n} + \eta_{n-\ell+1:n} \end{bmatrix} \in \mathbb{R}^n$$

The window is thereby centered on the gap and short enough so that the DCT spectrum of signal $s$ can be assumed static over the window’s duration $n$. Signal to noise ratio within this window is defined

$$\text{SNR} \triangleq 20 \log \frac{\left\| s_{1:\ell} \right\|}{\left\| \eta \right\|}$$

In absence of noise, knowing the signal DCT basis and having a good estimate of basis coefficient cardinality makes perfectly reconstructing gap-loss easy: it amounts to solving a linear system of equations and requires little or no optimization; with caveat, number of equations exceeds cardinality of signal representation (roughly $\ell \geq k$) with respect to DCT basis.

But addition of a significant amount of noise $\eta$ increases level of difficulty dramatically; a 1-norm based method of reducing cardinality, for example, almost always returns DCT basis coefficients numbering in excess of minimal cardinality. We speculate that is because signal cardinality $2\ell$ becomes the predominant cardinality. DCT basis coefficient cardinality is an explicit constraint to the optimization problem we shall pose: In presence of noise, constraints equating reconstructed signal $f$ to received signal $g$ are not possible.

---

4.38 This simplifies exposition, although it may be an unrealistic assumption in many applications.
Figure 106: (a) Signal dropout in signal $s$ corrupted by noise $\eta$ (SNR = 10dB, $g = s + \eta$). Flatline indicates duration of signal dropout. (b) Reconstructed signal $f$ (red) overlaid with corrupted signal $g$. 
Figure 107: (a) Error signal power (reconstruction $f$ less original noiseless signal $s$) is 36dB below $s$. (b) Original signal $s$ overlaid with reconstruction $f$ (red) from signal $g$ having dropout plus noise.
4.5. CONSTRAINING CARDINALITY

We can instead formulate the dropout recovery problem as a best approximation:

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & & \| f_{1:\ell} - g_{1:\ell} \\ & & f_{n-\ell+1:n} - g_{n-\ell+1:n} \| \\
\text{subject to} & & f = \Psi x \\
& & x \geq 0 \\
& & \text{card } x \leq k
\end{align*}
\]

(818)

We propose solving this nonconvex problem (818) by moving the cardinality constraint to the objective as a regularization term as explained in §4.5: \textit{id est}, by iteration of two convex problems until convergence:

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & & \langle x, y \rangle + \| f_{1:\ell} - g_{1:\ell} \\ & & f_{n-\ell+1:n} - g_{n-\ell+1:n} \| \\
\text{subject to} & & f = \Psi x \\
& & x \geq 0
\end{align*}
\]

(819)

and

\[
\begin{align*}
\min_{y \in \mathbb{R}^n} & & \langle x^*, y \rangle \\
\text{subject to} & & 0 \leq y \leq 1 \\
& & y^T \mathbf{1} = n - k
\end{align*}
\]

(524)

Signal cardinality \(2\ell \) is implicit to the problem statement. When number of samples in the dropout region exceeds half the window size, then that deficient cardinality of signal remaining becomes a source of degradation to reconstruction in presence of noise. Thus, by observation, we divine a reconstruction rule for this signal dropout problem to attain good noise suppression: \(\ell \) must exceed a maximum of cardinality bounds; \(2\ell \geq \max\{2k, n/2\}\).

Figure 106 and Figure 107 show one realization of this dropout problem. Original signal \(s\) is created by adding four \((k = 4)\) randomly selected DCT basis vectors, from \(\Psi\) \((n = 500\) in this example), whose amplitudes are randomly selected from a uniform distribution above the noise floor; in the interval \([10^{-10/20}, 1]\). Then a 240-sample dropout is realized \((\ell = 130)\) and Gaussian noise \(\eta\) added to make corrupted signal \(g\) (from which a best approximation \(f\) will be made) having 10dB signal to noise ratio (817). The time gap contains much noise, as apparent from Figure 106a. But in only a few iterations (819) (524), original signal \(s\) is recovered with relative error power 36dB down; illustrated in Figure 107. Correct cardinality is also recovered \((\text{card } x = \text{card } z)\) along with the basis vector indices used to make original signal \(s\). Approximation error is due to DCT basis coefficient estimate error. When this experiment is repeated 1000 times on noisy signals averaging 10dB SNR, the correct cardinality and indices are recovered 99% of the time with average relative error power 30dB down. Without noise, we get perfect reconstruction in one iteration. [398, MATLAB code]

4.5.1.6 Compressed sensing geometry with a nonnegative variable

It is well known that cardinality problem (529), on page 197, is easier to solve by linear programming when variable \(x\) is nonnegatively constrained than when not; \textit{e.g}, Figure 73, Figure 105. We postulate a simple geometrical explanation:
$A = \{ x \in \mathbb{R}^3 \mid Ax = b \}$

$cS = \{ x \mid x \succeq 0, \ 1^T x \leq c \}$

Figure 108: Simplex $S$ is convex hull of origin and all cardinality-1 nonnegative vectors of unit norm (its vertices). Line $A$, intersecting two-dimensional (cardinality-2) face $F$ of nonnegative simplex $cS$, emerges from $cS$ at a cardinality-1 vertex. $S$ equals nonnegative orthant $\mathbb{R}_+^3 \cap 1$-norm ball $B_1$ (Figure 72). Kissing point achieved when $\bullet$ (on edge) meets $A$ as simplex contracts (as scalar $c$ diminishes) under optimization (522).
4.5. CONSTRAINING CARDINALITY

Figure 72 illustrates 1-norm ball $B_1$ in $\mathbb{R}^3$ and affine subset $\mathcal{A}$ defined \{\(x \in \mathbb{R}^3 \mid Ax = b\). Prototypical compressed sensing problem, for $A \in \mathbb{R}^{m \times n}$

\[
\begin{align*}
\text{minimize} \quad & \|x\|_1 \\
\text{subject to} \quad & Ax = b
\end{align*}
\] (517)

is solved when the 1-norm ball $B_1$ kisses the affine subset.

If variable $x$ is constrained to the nonnegative orthant

\[
\begin{align*}
\text{minimize} \quad & \|x\|_1 \\
\text{subject to} \quad & Ax = b \\
& x \succeq 0
\end{align*}
\] (522)

then 1-norm ball $B_1$ becomes nonnegative simplex $S$ in Figure 108 where

\[cS = \{[I \in \mathbb{R}^{n \times n}, 0 \in \mathbb{R}^n]a \mid a^T 1 = c, a \succeq 0\} = \{x \mid x \succeq 0, \ 1^T x \leq c\}\] (820)

Nonnegative simplex $S$ is the convex hull of its vertices. All $n+1$ vertices of $S$ are constituted by standard basis vectors and the origin. In other words, all its nonzero extreme points are cardinality-1.

Affine subset $\mathcal{A}$ kisses nonnegative simplex $c^* S$ at optimality of (522). A kissing point is achieved at $x^*$ for optimal $c^*$ as $B_1$ or $S$ contracts. Whereas 1-norm ball $B_1$ has only six vertices in $\mathbb{R}^3$ corresponding to cardinality-1 solutions, simplex $S$ has three edges (along the Cartesian axes) containing an infinity of cardinality-1 solutions. And whereas $B_1$ has twelve edges containing cardinality-2 solutions, $S$ has three (out of total four) facets constituting cardinality-2 solutions. In other words, likelihood of a low-cardinality solution is higher by kissing nonnegative simplex $S$ (522) than by kissing 1-norm ball $B_1$ (517) because facial dimension (corresponding to given cardinality) is higher in $S$.

Empirically, this observation also holds in other Euclidean dimensions (Figure 73, Figure 105).

4.5.1.7 cardinality-1 compressed sensing problem always solvable

In the special case of cardinality-1 feasible solution to nonnegative compressed sensing problem (522), there is a geometrical interpretation that leads to an algorithm.

Figure 108 illustrates a cardinality-1 feasible solution to problem (522) in $\mathbb{R}^3$; a vertex solution. But first-octant $S$ of 1-norm ball $B_1$ does not kiss line $\mathcal{A}$; which would be an optimality condition. How can we perform optimization and make $\mathcal{A}$ intersect $S$ at a vertex? Assuming that nonnegative cardinality-1 solutions exist in the feasible set, it so happens:

4.5.1.7.1 Algorithm. Deprecation.

Columns of measurement matrix $A$, corresponding to high cardinality solution of (522)\(^{4.39}\) found by Simplex method \([97]\), may be deprecated and the problem solved

\(^{4.39}\)Because signed compressed sensing problem (517) can be equivalently expressed in a nonnegative variable, as we learned in Example 3.2.0.0.1 (p.194), and because a cardinality-1 constraint in (517) transforms to a cardinality-1 constraint in its nonnegative equivalent (521), then this cardinality-1 recursive reconstruction algorithm continues to hold for a signed variable as in (517).
again with those columns missing. Such columns are recursively removed from $A$ until a cardinality-1 solution is found.

This algorithm intimates that either a solution to problem (522) is cardinality-1 or column indices of $A$, corresponding to a higher cardinality solution, do not intersect that index corresponding to a cardinality-1 feasible solution.

When problem (522) is first solved, in the example of Figure 108, solution is cardinality-2 at a kissing point on that edge of simplex $cS$ indicated by $\bullet$. Imagining that the corresponding cardinality-2 face $F$ has collapsed, as a result of zeroing those two extreme points whose convex hull constructs that same edge $\bullet$ of $F$, then the simplex collapses to a line segment along the $y$ axis. When that line segment kisses $A$, then the cardinality-1 vertex solution illustrated has been found.\footnote{4.40}

4.5.1.7.2 Proof (pending). Deprecation algorithm 4.5.1.7.1.

We require proof that a cardinality-1 feasible solution to (522) cannot exist within a higher cardinality optimal solution found by Simplex method; for only then can corresponding columns of $A$ be eliminated without precluding cardinality-1 at optimality of the deprecated problem. Crucial is the Simplex method of solution because then an optimal solution is guaranteed to reside at a vertex of the feasible set. [97, p.158] [16, p.2]

Although it is more efficient (compared with our algorithm) to search over individual columns of matrix $A$ for a cardinality-1 solution known a priori to exist, tables are turned when cardinality exceeds 1:

4.5.2 cardinality-$k$ geometric presolver

This idea of deprecating columns has foundation in convex cone theory. (§2.13.4.3) Removing columns (and rows)\footnote{4.41} from $A \in \mathbb{R}^{m \times n}$ in a linear program like (522) (§3.2) is known in the industry as presolving;\footnote{4.42} the elimination of redundant constraints and identically zero variables prior to numerical solution. We offer a different and geometric presolver:

Two interpretations of the constraints from problem (522) are realized in Figure 109. Assuming that a cardinality-$k$ solution exists and matrix $A$ describes a pointed polyhedral cone $K = \{Ax \mid x \succeq 0\}$, as in Figure 109b, columns are removed from $A$ if they do not belong to the smallest face $F$ of $K$ containing vector $b$; those columns correspond to 0-entries in variable vector $x$ (and vice versa). Generators of that smallest face always hold a minimal cardinality solution, in other words, because a generator outside the smallest face (having positive coefficient) would violate the assumption that $b$ belongs to that face.

\footnote{4.40}A similar argument holds for any orientation of line $A$ and cardinality-1 point of emergence from simplex $cS$. This cardinality-1 reconstruction algorithm also holds more generally when affine subset $\mathcal{A}$ has any higher dimension $n-m$.

\footnote{4.41}Rows of matrix $A$ are removed based upon linear dependence. Assuming $b \in \mathbb{R}(A)$, corresponding entries of vector $b$ may also be removed without loss of generality.

\footnote{4.42}The conic technique proposed here can exceed performance of the best industrial presolvers in terms of number of columns removed, but not in execution time. This geometric presolver becomes attractive when a linear or integer program is not solvable by other means; perhaps because of sheer dimension.
4.5. CONSTRAINING CARDINALITY

Figure 109: Constraint interpretations: (a) Halfspace-description of feasible set in problem (522) is a polyhedron $P$ formed by intersection of nonnegative orthant $\mathbb{R}^n_+$ with hyperplanes $A$ prescribed by equality constraint. (Drawing by Pedro Sánchez.) (b) Vertex-description of constraints in problem (522): point $b$ belongs to polyhedral cone $K = \{Ax \mid x \succeq 0\}$. Number of extreme directions in $K$ may exceed dimensionality of ambient space.
Benefit accrues when vector \( b \) does not belong to relative interior of \( \mathcal{K} \); there would be no columns to remove were \( b \in \text{rel int} \mathcal{K} \) since the smallest face becomes cone \( \mathcal{K} \) itself (Example 4.5.2.0.4). Were \( b \) an extreme direction, at the other end of the spectrum, then the smallest face is an edge that is a ray containing \( b \); this geometrically describes a cardinality-1 case where all columns, save one, would be removed from \( A \).

When vector \( b \) resides in a face \( \mathcal{F} \) of \( \mathcal{K} \) that is not cone \( \mathcal{K} \) itself, benefit is realized as a reduction in computational intensity because the consequent equivalent problem has smaller dimension. Number of columns removed depends completely on geometry of a given problem; particularly, location of \( b \) within \( \mathcal{K} \). In the example of Figure 109b, interpreted literally in \( \mathbb{R}^3 \), all but two columns of \( A \) are discarded by our presolver when \( b \) belongs to facet \( \mathcal{F} \).

There are always exactly \( n \) linear feasibility problems to solve in order to discern generators of the smallest face of \( \mathcal{K} \) containing \( b \); the topic of §2.13.4.3.4.43

4.5.2.0.3 Exercise. Minimal cardinality generators.
Prove that generators of the smallest face \( \mathcal{F} \) of \( \mathcal{K} = \{Ax \mid x \succeq 0\} \) containing vector \( b \) always hold a minimal cardinality solution to \( Ax = b \). ▼

4.5.2.0.4 Example. Presolving for cardinality-2 solution to \( Ax = b \).
(confer Example 4.5.1.5.1) Again taking data from Example 4.2.3.1.1 (\( A \in \mathbb{R}^{m \times n} \), desired cardinality of \( x \) is \( k \)), for \( m = 3 \), \( n = 6 \), \( k = 2 \)

\[
A = \begin{bmatrix}
-1 & 1 & 8 & 1 & 1 & 0 \\
-3 & 2 & 8 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{3} \\
-9 & 4 & 8 & \frac{1}{3} & \frac{1}{9} & \frac{1}{2} - \frac{1}{9}
\end{bmatrix}, \quad b = \begin{bmatrix}
1 \\
\frac{1}{2} \\
\frac{1}{4}
\end{bmatrix}
\] (713)

proper cone \( \mathcal{K} = \{Ax \mid x \succeq 0\} \) is pointed as proven by method of §2.12.2.2. A cardinality-2 solution is known to exist; sum of the last two columns of matrix \( A \). Generators of the smallest face that contains vector \( b \), found by the method in Example 2.13.4.3.1, comprise the entire \( A \) matrix because \( b \in \text{int} \mathcal{K} \) (§2.13.4.2.4). So geometry of this particular problem does not permit number of generators to be reduced below \( n \) by discerning the smallest face.4.44 □

There is wondrous bonus to presolving when a constraint matrix is sparse. After columns are removed by theory of convex cones (finding the smallest face), some remaining rows may become \( 0^T \), identical to other rows, or nonnegative. When nonnegative rows appear in an equality constraint to \( 0 \), all nonnegative variables corresponding to nonnegative entries in those rows must vanish (§A.7.1); meaning, more columns may be removed. Once rows and columns have been removed from a constraint matrix, even more rows and columns may be removed by repeating the presolver procedure.

4.43 Comparison of computational intensity for this conic presolver to a brute force search would pit combinatorial complexity, a binomial coefficient \( \propto \binom{n}{k} \), against polynomial complexity; \( n \) linear feasibility problems plus numerical solution of the presolved problem.
4.44 But a canonical set of conically independent generators of \( \mathcal{K} \) comprise only the first two and last two columns of \( A \).
4.5.3 constraining cardinality of signed variable

Now consider a feasibility problem equivalent to the classical problem from linear algebra

\[ Ax = b \]

but with an upper bound \( k \) on cardinality \( \|x\|_0 \): for vector \( b \in \mathcal{R}(A) \)

\[
\begin{align*}
\text{find} & \quad x \in \mathbb{R}^n \\
\text{subject to} & \quad Ax = b \\
& \quad \|x\|_0 \leq k
\end{align*}
\]

(821)

where \( \|x\|_0 \leq k \) means vector \( x \) has at most \( k \) nonzero entries; such a vector is presumed existent in the feasible set. Convex iteration (\S 4.5.1) utilizes a nonnegative variable; so absolute value \( |x| \) is needed here. We propose that nonconvex problem (821) can be equivalently written as a sequence of convex problems that move the cardinality constraint to the objective:

\[
\begin{align*}
\text{minimize} & \quad \langle |x|, y \rangle \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

(822)

\[
\begin{align*}
\text{minimize} & \quad \langle t, y + \varepsilon \mathbf{1} \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad -t \preceq x \preceq t
\end{align*}
\]

(823)

where \( \varepsilon \) is a relatively small positive constant. This sequence is iterated until a direction vector \( y \) is found that makes \( |x| y^T \) vanish. The term \( \langle t, \varepsilon \mathbf{1} \rangle \) in (822) is necessary to determine absolute value \( |x^*| \) because vector \( y \) can have zero-valued entries. By initializing \( y \) to \( (1 - \varepsilon) \mathbf{1} \), the first iteration of problem (822) is a 1-norm problem (513); \textit{id est},

\[
\begin{align*}
\text{minimize} & \quad \langle t, \mathbf{1} \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad -t \preceq x \preceq t
\end{align*}
\]

(517)

Subsequent iterations of problem (822) engaging cardinality term \( \langle t, y \rangle \) can be interpreted as corrections to this 1-norm problem leading to a 0-norm solution; vector \( y \) can be interpreted as a direction of search.

4.5.3.1 local convergence

As before (\S 4.5.1.3), convex iteration (822) (524) always converges to a locally optimal solution; a fixed point of possibly infeasible cardinality.

4.5.3.2 simple variations on a signed variable

Several useful equivalents to linear programs (822) (524) are easily devised, but their geometrical interpretation is not as apparent: \textit{e.g}, equivalent in the limit \( \varepsilon \to 0^+ \)

\[
\begin{align*}
\text{minimize} & \quad \langle t, y \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad -t \preceq x \preceq t
\end{align*}
\]

(823)
CHAPTER 4. SEMIDEFINITE PROGRAMMING

\[ \begin{align*}
\text{minimize} & \quad \langle |x^*|, y \rangle \\
\text{subject to} & \quad 0 \preceq y \preceq 1 \\
& \quad y^T 1 = n - k
\end{align*} \]

We get another equivalent to linear programs (822) (524), in the limit, by interpreting problem (517) as infimum to a vertex-description of the 1-norm ball (Figure 72, Example 3.2.0.0.1, confer (516)):

\[ \begin{align*}
\text{minimize} & \quad \|x\|_1 \\
\text{subject to} & \quad Ax = b \equiv \text{minimize } \langle a, y \rangle \\
& \quad \text{subject to } [A - A]a = b \\
& \quad a \succeq 0
\end{align*} \]

\[ \begin{align*}
\text{minimize} & \quad \langle a^*, y \rangle \\
\text{subject to} & \quad 0 \preceq y \preceq 1 \\
& \quad y^T 1 = 2n - k
\end{align*} \]

where \( x^* = [I - I]a^* \); from which it may be construed that any vector 1-norm minimization problem has equivalent expression in a nonnegative variable.

4.6 Cardinality and rank constraint examples

4.6.0.0.1 Example. Projection on ellipsoid boundary. [54] [155, §5.1] [257, §2]
Consider classical linear equation \( Ax = b \) but with constraint on norm of solution \( x \), given matrices \( C \), fat \( A \), and vector \( b \in \mathbb{R}^\langle A \rangle \)

\[ \begin{align*}
\text{find} & \quad x \in \mathbb{R}^N \\
\text{subject to} & \quad Ax = b \\
& \quad \|Cx\| = 1
\end{align*} \]

The set \( \{ x \mid \|Cx\| = 1 \} \) (2) describes an ellipsoid boundary (Figure 14). This is a nonconvex problem because solution is constrained to that boundary. Assign

\[ \begin{align*}
G = \begin{bmatrix} Cx & X \\
1 & 1 \end{bmatrix} \begin{bmatrix} x^T C^T & X \\
x^T C \end{bmatrix} \Delta \begin{bmatrix} Cxx^T C^T & Cx \\
x^T C \end{bmatrix} \in \mathbb{S}^{N+1}
\end{align*} \]

Any rank-1 solution must have this form. (§B.1.0.2) Ellipsoidally constrained feasibility problem (825) is equivalent to:

\[ \begin{align*}
\text{find} & \quad x \in \mathbb{R}^N \\
\text{subject to} & \quad Ax = b \\
& \quad G = \begin{bmatrix} X & Cx \\
x^T C & 1 \end{bmatrix} \geq 0 \\
& \quad \text{rank } G = 1 \\
& \quad \text{tr } X = 1
\end{align*} \]
This is transformed to an equivalent convex problem by moving the rank constraint to the objective: We iterate solution of

$$\begin{align*}
\text{minimize} & \quad \langle G, Y \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad G = \begin{bmatrix} X & Cx \\ x^T C^T & 1 \end{bmatrix} \succeq 0 \\
& \quad \operatorname{tr} X = 1
\end{align*}$$

with

$$\begin{align*}
\text{minimize} & \quad \langle G^*, Y \rangle \\
\text{subject to} & \quad 0 \preceq Y \preceq I \\
& \quad \operatorname{tr} Y = N
\end{align*}$$

until convergence. Initially $0$, direction matrix $Y \in S^{N+1}$ regulates rank. (1792a) Singular value decomposition $G^* = U\Sigma Q^T \in S^{N+1}$ (§A.6) provides a new direction matrix $Y = U(:, 2:N+1)U(:, 2:N+1)^T$ that optimally solves (829) at each iteration. An optimal solution to (825) is thereby found in a few iterations, making convex problem (828) its equivalent.

It remains possible for the iteration to stall; were a rank-1 $G$ matrix not found. In that case, the current search direction is momentarily reversed with an added randomized element:

$$Y = -U(:, 2:N+1) * (U(:, 2:N+1)^T + \text{randn}(N, 1) * U(:, 1)^T)$$

in MATLAB notation. This heuristic is quite effective for problem (825) which is exceptionally easy to solve by convex iteration.

When $b \notin \mathcal{R}(A)$ then problem (825) must be restated as a projection:

$$\begin{align*}
\text{minimize} & \quad \|Ax - b\| \\
\text{subject to} & \quad \|Cx\| = 1
\end{align*}$$

This is a projection of point $b$ on an ellipsoid boundary because any affine transformation of an ellipsoid remains an ellipsoid. Problem (828) in turn becomes

$$\begin{align*}
\text{minimize} & \quad \langle G, Y \rangle + \|Ax - b\| \\
\text{subject to} & \quad G = \begin{bmatrix} X & Cx \\ x^T C^T & 1 \end{bmatrix} \succeq 0 \\
& \quad \operatorname{tr} X = 1
\end{align*}$$

We iterate this with calculation (829) of direction matrix $Y$ as before until a rank-1 $G$ matrix is found.

### 4.6.0.0.2 Example. Orthonormal Procrustes.

Example 4.6.0.1 is extensible. An orthonormal matrix $Q \in \mathbb{R}^{n \times p}$ is characterized $Q^T Q = I$. Consider the particular case $Q = [x \ y] \in \mathbb{R}^{n \times 2}$ as variable to a Procrustes problem (§C.3): given $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times 2}$

$$\begin{align*}
\text{minimize} & \quad \|AQ - B\|_F \\
\text{subject to} & \quad Q^T Q = I
\end{align*}$$
which is nonconvex. By vectorizing matrix $Q$ we can make the assignment:

$$G = \begin{bmatrix} x & y \\ 1 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} x^T & y^T & 1 \end{bmatrix} \end{bmatrix} \triangleq \begin{bmatrix} x^T, y^T & x \\ y^T & y \\ 1 \end{bmatrix} \in S^{2n+1} \quad (834)$$

Now orthonormal Procrustes problem (833) can be equivalently restated:

$$\begin{align*}
\text{minimize} & \quad \|A[x \ y] - B\|_F \\
\text{subject to} & \quad G = \begin{bmatrix} X & Z & x \\ Z^T & Y & y \\ x^T & y^T & 1 \end{bmatrix} \succeq 0 \\
& \quad \text{rank} \ G = 1 \\
& \quad \text{tr} \ X = 1 \\
& \quad \text{tr} \ Y = 1 \\
& \quad \text{tr} \ Z = 0 \\
\end{align*} \quad (835)$$

To solve this, we form the convex problem sequence:

$$\begin{align*}
\text{minimize} & \quad \|A[x \ y] - B\|_F + \langle G, W \rangle \\
\text{subject to} & \quad G = \begin{bmatrix} X & Z & x \\ Z^T & Y & y \\ x^T & y^T & 1 \end{bmatrix} \succeq 0 \\
& \quad \text{tr} \ X = 1 \\
& \quad \text{tr} \ Y = 1 \\
& \quad \text{tr} \ Z = 0 \\
\end{align*} \quad (836)$$

and

$$\begin{align*}
\text{minimize} & \quad \langle G^*, W \rangle \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \text{tr} \ W = 2n \\
\end{align*} \quad (837)$$

which has an optimal solution $W$ that is known in closed form (p.565). These two problems are iterated until convergence and a rank-1 $G$ matrix is found. A good initial value for direction matrix $W$ is $0$. Optimal $Q^*$ equals $[x^* \ y^*]$.

Numerically, this Procrustes problem is easy to solve; a solution seems always to be found in one or few iterations. This problem formulation is extensible, of course, to orthogonal (square) matrices $Q$.

4.6.0.0.3 Example. Combinatorial Procrustes problem.
In case $A, B \in \mathbb{R}^n$, when vector $A = \Xi B$ is known to be a permutation of vector $B$, solution to orthogonal Procrustes problem

$$\begin{align*}
\text{minimize} & \quad \|A - XB\|_F \\
\text{subject to} & \quad X^T = X^{-1} \\
\end{align*} \quad (1804)$$
is not necessarily a permutation matrix $\Xi$ even though an optimal objective value of 0 is found by the known analytical solution (§C.3). The simplest method of solution finds permutation matrix $X^* = \Xi$ simply by sorting vector $B$ with respect to $A$.

Instead of sorting, we design two different convex problems each of whose optimal solution is a permutation matrix: one design is based on rank constraint, the other on cardinality. Because permutation matrices are sparse by definition, we depart from a traditional Procrustes problem by instead demanding a vector 1-norm which is known to produce solutions more sparse than Frobenius’ norm.

There are two principal facts exploited by the first convex iteration design (§4.4.1) we propose. Permutation matrices $\Xi$ constitute:

1) the set of all nonnegative orthogonal matrices,
2) all points extreme to the polyhedron (100) of doubly stochastic matrices.

That means:

1) norm of each row and column is 1,

\[ \| \Xi(:,i) \| = 1, \quad \| \Xi(i,:) \| = 1, \quad i = 1 \ldots n \] (838)

2) sum of each nonnegative row and column is 1, (§2.3.2.0.4)

\[ \Xi^T 1 = 1, \quad \Xi 1 = 1, \quad \Xi \geq 0 \] (839)

**solution via rank constraint**

The idea is to individually constrain each column of variable matrix $X$ to have unity norm. Matrix $X$ must also belong to that polyhedron, (100) in the nonnegative orthant, implied by constraints (839); so each row-sum and column-sum of $X$ must also be unity. It is this combination of nonnegativity, sum, and sum square constraints that extracts the permutation matrices: (Figure 110) given nonzero vectors $A, B$

\[
\begin{align*}
\text{minimize} & \quad \|A - X B\|_1 + w \sum_{i=1}^{n} (G_i, W_i) \\
\text{subject to} & \quad G_i = \begin{bmatrix} G_i(1:n, 1:n) & X(:,i)^T \\ X(:,i) & 1 \end{bmatrix} \succeq 0, \quad i = 1 \ldots n \\
& \quad \text{tr} G_i = 2 \\
& \quad X^T 1 = 1 \\
& \quad X 1 = 1 \\
& \quad X \geq 0
\end{align*}
\] (840)

This fact would be superfluous were the objective of minimization linear, because the permutation matrices reside at the extreme points of a polyhedron (100) implied by (839). But as posed, only either rows or columns need be constrained to unit norm because matrix orthogonality implies transpose orthogonality. (§B.5.2) Absence of vanishing inner product constraints that help define orthogonality, like $\text{tr} Z = 0$ from Example 4.6.0.0.2, is a consequence of nonnegativity; *id est*, the only orthogonal matrices having exclusively nonnegative entries are permutations of the Identity.
where \( w \approx 10 \) positively weights the rank regularization term. Optimal solutions \( G_i^* \) are key to finding direction matrices \( W_i \) for the next iteration of semidefinite programs (840) (841):

\[
\begin{align*}
\text{minimize} & \quad \langle G_i^*, W_i \rangle \\
\text{subject to} & \quad 0 \preceq W_i \preceq I \\
& \quad \text{tr} W_i = n \quad i = 1 \ldots n
\end{align*}
\]

(841)

Direction matrices thus found lead toward rank-1 matrices \( G_i^* \) on subsequent iterations. Constraint on trace of \( G_i^* \) normalizes the \( i \)th column of \( X^* \) to unity because (confer p.375)

\[
G_i^* = \begin{bmatrix} X^*(\cdot, i) \\ 1 \end{bmatrix} \begin{bmatrix} X^*(\cdot, i)^T & 1 \end{bmatrix}
\]

(842)

at convergence. Binary-valued \( X^* \) column entries result from the further sum constraint \( X1 = 1 \). Columnar orthogonality is a consequence of the further transpose-sum constraint \( X^T1 = 1 \) in conjunction with nonnegativity constraint \( X \geq 0 \); but we leave proof of orthogonality as an exercise. The optimal objective value is 0 for both semidefinite programs when vectors \( A \) and \( B \) are related by permutation. In any case, optimal solution \( X^* \) becomes a permutation matrix \( \Xi \).

Because there are \( n \) direction matrices \( W_i \) to find, it can be advantageous to invoke a known closed-form solution for each from page 565. What makes this combinatorial problem more tractable are relatively small semidefinite constraints in (840). (confer (836))

When a permutation \( A \) of vector \( B \) exists, number of iterations can be as small as 1. But this combinatorial Procrustes problem can be made even more challenging when vector \( A \) has repeated entries.
solution via cardinality constraint

Now the idea is to force solution at a vertex of permutation polyhedron (100) by finding a solution of desired sparsity. Because permutation matrix \( X \) is \( n \)-sparse by assumption, this combinatorial Procrustes problem may instead be formulated as a compressed sensing problem with convex iteration on cardinality of vectorized \( X \) (§4.5.1): given nonzero vectors \( A, B \)

\[
\begin{align*}
\text{minimize} & \quad \| A - X B \|_1 + w \langle X, Y \rangle \\
\text{subject to} & \quad X^T 1 = 1 \\
& \quad X 1 = 1 \\
& \quad X \geq 0
\end{align*}
\]

(843)

where direction vector \( Y \) is an optimal solution to

\[
\begin{align*}
\text{minimize} & \quad \langle X^*, Y \rangle \\
\text{subject to} & \quad 0 \leq Y \leq 1 \\
& \quad 1^T Y 1 = n^2 - n
\end{align*}
\]

(524)

each a linear program. In this circumstance, use of closed-form solution for direction vector \( Y \) is discouraged. When vector \( A \) is a permutation of \( B \), both linear programs have objectives that converge to 0. When vectors \( A \) and \( B \) are permutations and no entries of \( A \) are repeated, optimal solution \( X^* \) can be found as soon as the first iteration.

In any case, \( X^* = \Xi \) is a permutation matrix. □

4.6.0.0.4 Exercise. Combinatorial Procrustes constraints.

Assume that the objective of semidefinite program (840) is 0 at optimality. Prove that the constraints in program (840) are necessary and sufficient to produce a permutation matrix as optimal solution. Alternatively and equivalently, prove those constraints necessary and sufficient to optimally produce a nonnegative orthogonal matrix. ▼

4.6.0.0.5 Example. Tractable polynomial constraint.

The set of all coefficients for which a multivariate polynomial were convex is generally difficult to determine. But the ability to handle rank constraints makes any nonconvex polynomial constraint transformable to a convex constraint. All optimization problems having polynomial objective and polynomial constraints can be reformulated as a semidefinite program with a rank-1 constraint. [296] Suppose we require

\[
3 + 2x - xy \leq 0
\]

(844)

Identify

\[
G = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \begin{bmatrix} x & y & 1 \end{bmatrix} = \begin{bmatrix} x^2 & xy & x \\ xy & y^2 & y \\ x & y & 1 \end{bmatrix} \in \mathbb{S}^3
\]

(845)

Then nonconvex polynomial constraint (844) is equivalent to constraint set

\[
\begin{align*}
\text{tr}(GA) & \leq 0 \\
G_{33} & = 1 \\
(G \succeq 0) \\
\text{rank } G & = 1
\end{align*}
\]

(846)
with direct correspondence to sense of trace inequality where \( G \) is assumed symmetric (§B.1.0.2) and

\[
A = \begin{bmatrix}
  0 & -\frac{1}{2} & 1 \\
  -\frac{1}{2} & 0 & 0 \\
  1 & 0 & 3
\end{bmatrix} \in \mathbb{S}^3
\]  

Then the method of convex iteration from §4.4.1 is applied to implement the rank constraint.

\[\Box\]

### 4.6.0.0.6 Exercise. Binary Pythagorean theorem.
The technique in Example 4.6.0.0.5 is extensible to any quadratic constraint; e.g., \( x^T A x + 2b^T x + c \leq 0 \), \( x^T A x + 2b^T x + c \geq 0 \), and \( x^T A x + 2b^T x + c = 0 \). Write a rank-constrained semidefinite program to solve (Figure 110)

\[
\begin{align*}
\{ & x + y = 1 \\
   & x^2 + y^2 = 1
\end{align*}
\]  

whose feasible set is not connected. Implement this system in cvx [177] by convex iteration.

\[\nabla\]

### 4.6.0.0.7 Example. High order polynomials.
Consider nonconvex problem from Canadian Mathematical Olympiad 1999:

\[
\begin{align*}
\text{find} & \quad x, \ y, \ z \\
\text{subject to} & \quad x^2 y + y^2 z + z^2 x = \frac{x^2}{3^3} \\
   & \quad x + y + z = 1 \\
   & \quad x, \ y, \ z \geq 0
\end{align*}
\]  

We wish to solve for, what is known to be, a tight upper bound \( \frac{x^2}{3^3} \) on the constrained polynomial \( x^2 y + y^2 z + z^2 x \) by transformation to a rank-constrained semidefinite program. First identify

\[
G = \begin{bmatrix} x^2 & y^2 & z^2 & x & y & z \\
   & x & y & z & x & 1
\end{bmatrix} = \begin{bmatrix} x^2 & xy & zx & x \\
x y & y^2 & yz & y \\
z x & zx & z^2 & z \\
x & y & z & 1
\end{bmatrix} \in \mathbb{S}^4
\]  

\[
X = \begin{bmatrix} x^2 & y^2 & z^2 & x & y & z & 1 \\
y^2 & y^2 & z^2 & x & y & z & 1 \\
z^2 & z^2 & z^2 & x & y & z & 1 \\
x & y & z & x & y & z & 1 \\
x y & y^2 & y z & x & y & z & 1 \\
x z & y z & x & y & z & x & y \\
x & y & z & x & y & z & 1
\end{bmatrix} \in \mathbb{S}^7
\]
then apply convex iteration (§4.4.1) to implement rank constraints:

\[
\begin{align*}
\text{find} & \quad A, C \in \mathbb{S}, b \\
\text{subject to} & \quad \text{tr}(X) = 2 \\
& \quad \begin{bmatrix} A & b \\ b^T & 1 \end{bmatrix} \succeq 0 \\
& \quad \begin{bmatrix} C & \delta(A) \\ \delta(A)^T & b^T \end{bmatrix} \succeq 0 \\
& \quad 1^T b = 1 \\
& \quad b \succeq 0 \\
& \quad \text{rank } G = 1 \\
& \quad \text{rank } X = 1
\end{align*}
\]

(852)

where

\[
E = \begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

(853)

[392, MATLAB code]. Positive semidefiniteness is optional only when rank-1 constraints are explicit by Theorem A.3.1.0.7. Optimal solution \((x, y, z) = (0, \frac{2}{3}, \frac{1}{3})\) to problem (849) is not unique.

4.6.0.0.8 Exercise. Motzkin polynomial.
Prove \(xy^2 + x^2y - 3xy + 1\) to be nonnegative on the nonnegative orthant.

4.6.0.0.9 Example. Boolean vector satisfying \(Ax \succeq b\).
Now we consider solution to a discrete problem whose only known analytical method of solution is combinatorial in complexity: given \(A \in \mathbb{R}^{M \times N}\) and \(b \in \mathbb{R}^M\)

\[
\begin{align*}
\text{find} & \quad x \in \mathbb{R}^N \\
\text{subject to} & \quad Ax \succeq b \\
& \quad \delta(x x^T) = 1
\end{align*}
\]

(854)

This nonconvex problem demands a Boolean solution \([x_i = \pm 1, \ i = 1 \ldots N]\).
Assign a rank-1 matrix of variables; symmetric variable matrix \(X\) and solution vector \(x\):

\[
G = \begin{bmatrix}
x \\
x^T
\end{bmatrix} \begin{bmatrix}
1 \\
x T
\end{bmatrix} = \begin{bmatrix}
x x^T & x \\
x T & 1
\end{bmatrix} \triangleq \begin{bmatrix}
x x^T & x \\
x T & 1
\end{bmatrix} \in \mathbb{S}^{N+1}
\]

(855)

Then design an equivalent semidefinite feasibility problem to find a Boolean solution to
$Ax \preceq b$:

\[
\begin{align*}
\text{find} & \quad x \in \mathbb{R}^N \\
\text{subject to} & \quad Ax \preceq b \\
& \quad G = \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} (\succeq 0) \\
& \quad \text{rank } G = 1 \\
& \quad \delta(X) = 1
\end{align*}
\]

where $x_i^* \in \{-1, 1\}$, $i = 1, \ldots, N$. The two variables $X$ and $x$ are made dependent via their assignment to rank-1 matrix $G$. By (1704), an optimal rank-1 matrix $G^*$ must take the form (855).

As before, we regularize the rank constraint by introducing a direction matrix $Y$ into the objective:

\[
\begin{align*}
\text{minimize} & \quad \langle G, Y \rangle \\
\text{subject to} & \quad Ax \preceq b \\
& \quad G = \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \\
& \quad \delta(X) = 1
\end{align*}
\]

Solution of this semidefinite program is iterated with calculation of the direction matrix $Y$ from semidefinite program (829). At convergence, in the sense (768), convex problem (857) becomes equivalent to nonconvex Boolean problem (854).

Direction matrix $Y$ can be an orthogonal projector having closed-form expression, by (1792a), although convex iteration is not a projection method. (§4.4.1.1) Given randomized data $A$ and $b$ for a large problem, we find that stalling becomes likely (convergence of the iteration to a positive objective $\langle G^*, Y \rangle$). To overcome this behavior, we introduce a heuristic into the implementation on $	extit{W\ddot{u}kminization}$ [382] that momentarily reverses direction of search (like (830)) upon stall detection. We find that rate of convergence can be sped significantly by detecting stalls early.

4.6.0.0.10 Example. Variable-vector normalization.

Suppose, within some convex optimization problem, we want vector variables $x, y \in \mathbb{R}^N$ constrained by a nonconvex equality:

\[
x \|y\| = y
\]

\textit{id est}, $\|x\| = 1$ and $x$ points in the same direction as $y \neq 0$: e.g,

\[
\begin{align*}
\text{minimize} & \quad f(x, y) \\
\text{subject to} & \quad (x, y) \in C \\
& \quad x \|y\| = y
\end{align*}
\]

where $f$ is some convex function and $C$ is some convex set. We can realize the nonconvex equality by constraining rank and adding a regularization term to the objective. Make the
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

assignment:

\[
G = \begin{bmatrix}
  x & y \\
  x^T & y^T & 1
\end{bmatrix} = \begin{bmatrix}
  X & Z & x \\
  Z & Y & y \\
  x^T & y^T & 1
\end{bmatrix} \triangleq \begin{bmatrix}
  xx^T & xy^T & x \\
  yx^T & yy^T & y \\
  x^T & y^T & 1
\end{bmatrix} \in \mathbb{S}^{2N+1}
\]

(860)

where \( X, Y \in \mathbb{S}^N \), also \( Z \in \mathbb{S}^N \) [sic]. Any rank-1 solution must take the form of (860).

(§B.1) The problem statement equivalent to (859) is then written

\[
\begin{aligned}
\text{minimize} & \quad f(x, y) + \|X - Y\|_F \\
\text{subject to} & \quad (x, y) \in C \\
& \quad G = \begin{bmatrix}
  X & Z & x \\
  Z & Y & y \\
  x^T & y^T & 1
\end{bmatrix} \succeq 0 \\
& \quad \text{rank} G = 1 \\
& \quad \text{tr}(X) = 1 \\
& \quad \delta(Z) \succeq 0
\end{aligned}
\]

(861)

The trace constraint on \( X \) normalizes vector \( x \) while the diagonal constraint on \( Z \) maintains sign between respective entries of \( x \) and \( y \). Regularization term \( \|X - Y\|_F \) then makes \( x \) equal to \( y \) to within a real scalar; (§C.2.0.0.2) in this case, a positive scalar.

To make this program solvable by convex iteration, as explained in Example 4.4.1.2.4 and other previous examples, we move the rank constraint to the objective

\[
\begin{aligned}
\text{minimize} & \quad f(x, y) + \|X - Y\|_F + \langle G, W \rangle \\
\text{subject to} & \quad (x, y) \in C \\
& \quad G = \begin{bmatrix}
  X & Z & x \\
  Z & Y & y \\
  x^T & y^T & 1
\end{bmatrix} \succeq 0 \\
& \quad \text{tr}(X) = 1 \\
& \quad \delta(Z) \succeq 0
\end{aligned}
\]

(862)

by introducing a direction matrix \( W \) found from (1792a):

\[
\begin{aligned}
\text{minimize} & \quad \langle G^*, W \rangle \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \text{tr} W = 2N
\end{aligned}
\]

(863)

This semidefinite program has an optimal solution that is known in closed form. Iteration (862) (863) terminates when \( \text{rank} G = 1 \) and linear regularization \( \langle G, W \rangle \) vanishes to within some numerical tolerance in (862); typically, in two iterations. If function \( f \) competes too much with the regularization, positively weighting each regularization term will become required. At convergence, problem (862) becomes a convex equivalent to the original nonconvex problem (859).
Figure 111: A cut partitions nodes \( \{i = 1 \ldots 16\} \) of this graph into \( S \) and \( S' \). Linear arcs have circled weights. The problem is to find a cut maximizing total weight of all arcs linking partitions made by the cut.
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

4.6.0.0.11 Example. FAST MAX CUT.

Let $\Gamma$ be an $n$-node graph, and let the arcs $(i, j)$ of the graph be associated withweights $a_{ij}$. The problem is to find a cut of the largest possible weight, i.e., to partition the set ofnodes into two parts $S, S'$ in such a way that the total weight of all arcs linking $S$ and $S'$ (i.e., with one incident node in $S$ and theother one in $S'$ [Figure 111]) is as large as possible. — [35, §4.3.3]

Literature on the max cut problem is vast because this problem has elegant primal and dual formulation, its solution is very difficult, and there exist many commercial applications; e.g., semiconductor design [132], quantum computing [419].

Our purpose here is to demonstrate how iteration of two simple convex problems can quickly converge to an optimal solution of the max cut problem with a 98% success rate, on average.\textsuperscript{4.46} MAX CUT is stated:

\[
\begin{align*}
\text{maximize} & \quad \sum_{1 \leq i < j \leq n} a_{ij} (1 - x_i x_j)^{\frac{1}{2}} \\
\text{subject to} & \quad \delta(xx^T) = 1
\end{align*}
\]

where $[a_{ij}]$ are real arc weights, and binary vector $x = [x_i] \in \mathbb{R}^n$ corresponds to the $n$nodes; specifically,

\[
\begin{align*}
\text{node } i \in S & \iff x_i = 1 \\
\text{node } i \in S' & \iff x_i = -1
\end{align*}
\]

If nodes $i$ and $j$ have the same binary value $x_i$ and $x_j$, then they belong to the samepartition and contribute nothing to the cut. Arc $(i, j)$ traverses the cut, otherwise, addingits weight $a_{ij}$ to the cut.

MAX CUT statement (864) is the same as, for $A = [a_{ij}] \in \mathbb{S}^n$

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{4} \langle 11^T - xx^T, A \rangle \\
\text{subject to} & \quad \delta(xx^T) = 1
\end{align*}
\]

Because of Boolean assumption $\delta(xx^T) = 1$

\[
\langle 11^T - xx^T, A \rangle = \langle xx^T, \delta(A1) - A \rangle
\]

so problem (866) is the same as

\[
\begin{align*}
\text{maximize} & \quad \frac{1}{4} \langle xx^T, \delta(A1) - A \rangle \\
\text{subject to} & \quad \delta(xx^T) = 1
\end{align*}
\]

This MAX CUT problem is combinatorial (nonconvex).

Because an estimate of upper bound to MAX CUT is needed to ascertain convergence when vector $x$ has large dimension, we digress to derive the dual problem: Directly from (868), its Lagrangian is [63, §5.1.5] (1500)

\[
\begin{align*}
\mathcal{L}(x, \nu) &= \frac{1}{4} \langle xx^T, \delta(A1) - A \rangle + \langle \nu, \delta(xx^T) - 1 \rangle \\
&= \frac{1}{4} \langle xx^T, \delta(A1) - A \rangle + \langle \delta(\nu), xx^T - \nu, 1 \rangle \\
&= \frac{1}{4} \langle xx^T, \delta(A1 + 4\nu) - A \rangle - \langle \nu, 1 \rangle
\end{align*}
\]

\textsuperscript{4.46} We term our solution to MAX CUT \textit{fast} because we sacrifice a little accuracy to achieve speed; \textit{id est}, only about two or three convex iterations, achieved by heavily weighting a rank regularization term.
where quadratic $x^T(\delta(A_1 + 4\nu) - A)x$ has supremum 0 if $\delta(A_1 + 4\nu) - A$ is negative semidefinite, and has supremum $\infty$ otherwise. The finite supremum of dual function

$$g(\nu) = \sup_x \mathcal{L}(x, \nu) = \begin{cases} -\langle \nu, 1 \rangle, & A - \delta(A_1 + 4\nu) \succeq 0 \\ \infty, & \text{otherwise} \end{cases}$$ (870)

is chosen to be the objective of minimization to dual (convex) problem

$$\begin{align*}
\text{minimize} & \quad -\nu^T 1 \\
\text{subject to} & \quad A - \delta(A_1 + 4\nu) \succeq 0
\end{align*}$$ (871)

whose optimal value provides a least upper bound to MAX CUT, but is not tight ($\frac{1}{2}\langle xx^T, \delta(A_1) - A \rangle < g(\nu)$, duality gap is nonzero). [165] In fact, we find that the bound’s variance with problem instance is too large to be useful for this problem; thus ending our digression.\footnote{Taking the dual of dual problem (871) would provide (872) but without the rank constraint. [158] Dual of a dual of even a convex primal problem is not necessarily the same primal problem; although, optimal solution of one can be obtained from the other.}

To transform MAX CUT to its convex equivalent, first define

$$X = xx^T \in \mathbb{S}^n$$ (876)

then MAX CUT (868) becomes

$$\begin{align*}
\text{maximize} & \quad \frac{1}{4}\langle X, \delta(A_1) - A \rangle \\
\text{subject to} & \quad \delta(X) = 1 \\
& \quad \langle X \succeq 0 \rangle \\
& \quad \text{rank} X = 1
\end{align*}$$ (872)

whose rank constraint can be regularized as in

$$\begin{align*}
\text{maximize} & \quad \frac{1}{4}\langle X, \delta(A_1) - A \rangle - w\langle X, W \rangle \\
\text{subject to} & \quad \delta(X) = 1 \\
& \quad X \succeq 0 \\
& \quad X \preceq 0
\end{align*}$$ (873)

where $w \approx 1000$ is a nonnegative fixed weight, and $W$ is a direction matrix determined from

$$\sum_{i=2}^n \lambda(X^*)_i = \underset{W \in \mathbb{S}^n}{\text{minimize}} \quad \langle X^*, W \rangle$$ (1792a)

subject to $0 \preceq W \preceq I$

$$\text{tr } W = n - 1$$

which has an optimal solution that is known in closed form. These two problems (873) and (1792a) are iterated until convergence as defined on page 264.
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

Because convex problem statement (873) is so elegant, it is numerically solvable for large binary vectors within reasonable time.\footnote{We solved for a length-250 binary vector in only a few minutes and convex iterations on a 2006 vintage laptop Core 2 CPU (Intel T7400@2.16GHz, 666MHz FSB).} To test our convex iterative method, we compare an optimal convex result to an actual solution of the MAX CUT problem found by performing a brute force combinatorial search of (868)\footnote{more computationally intensive than the proposed convex iteration by many orders of magnitude. Solving MAX CUT by searching over all binary vectors of length 100, for example, would occupy a contemporary supercomputer for a million years.} for a tight upper bound. Search-time limits binary vector lengths to 24 bits (about five days CPU time). 98% accuracy, actually obtained, is independent of binary vector length (12, 13, 20, 24) when averaged over more than 231 problem instances including planar, randomized, and toroidal graphs.\footnote{Existence of a polynomial-time approximation to MAX CUT with accuracy provably better than 94.11% would refute NP-hardness; which Håstad believes to be highly unlikely. [191, thm.8.2] [192]} When failure occurred, large and small errors were manifest. That same 98% average accuracy is presumed maintained when binary vector length is further increased. A MATLAB program is provided on Winitzimization [387].

4.6.0.0.12 Example. Cardinality/rank problem.

d’Aspremont, El Ghaoui, Jordan, & Lanckriet [98] propose approximating a positive semidefinite matrix $A \in \mathbb{S}^N_+$ by a rank-one matrix having constraint on cardinality $c$:

for $0 < c < N$

\[
\begin{align*}
\text{minimize} & \quad \|A - zz^T\|_F \\
\text{subject to} & \quad \text{card } z \leq c
\end{align*}
\] (874)

which, they explain, is a hard problem equivalent to

\[
\begin{align*}
\text{maximize} & \quad x^T A x \\
\text{subject to} & \quad \|x\| = 1 \\
& \quad \text{card } x \leq c
\end{align*}
\] (875)

where $z \triangleq \sqrt{\lambda} x$ and where optimal solution $x^*$ is a principal eigenvector (1786) (§A.5) of $A$ and $\lambda = x^* A x^*$ is the principal eigenvalue [168, p.331] when $c$ is true cardinality of that eigenvector. This is principal component analysis with a cardinality constraint which controls solution sparsity. Define the matrix variable

\[
X \triangleq xx^T \in \mathbb{S}^N
\] (876)

whose desired rank is 1, and whose desired diagonal cardinality

\[
\text{card } \delta(X) \equiv \text{card } x
\] (877)

is equivalent to cardinality $c$ of vector $x$. Then we can transform cardinality problem (875) to an equivalent problem in new variable $X$:\footnote{A semidefiniteness constraint $X \succeq 0$ is not required, theoretically, because positive semidefiniteness of a rank-1 matrix is enforced by symmetry. (Theorem A.3.1.0.7)}
maximize $\langle X, A \rangle$
subject to $\langle X, I \rangle = 1$
$X \succeq 0$
$\text{rank} \ X = 1$
$\text{card} \ \delta(X) \leq c$

We transform problem (878) to an equivalent convex problem by introducing two direction matrices into regularization terms: $W$ to achieve desired cardinality $\text{card} \ \delta(X)$, and $Y$ to find an approximating rank-one matrix $X$:

$$\begin{align*}
\text{maximize} \quad & \langle X, A - w_1 Y \rangle - w_2 \langle \delta(X), \delta(W) \rangle \\
\text{subject to} \quad & \langle X, I \rangle = 1 \\
& X \succeq 0
\end{align*}$$

where $w_1$ and $w_2$ are positive scalars respectively weighting $\text{tr}(XY)$ and $\delta(X)^T \delta(W)$ just enough to insure that they vanish to within some numerical precision, where direction matrix $Y$ is an optimal solution to semidefinite program

$$\begin{align*}
\text{minimize} \quad & \langle X^*, Y \rangle \\
\text{subject to} \quad & 0 \preceq Y \preceq I \\
& \text{tr} \ Y = N - 1
\end{align*}$$

and where diagonal direction matrix $W \in \mathbb{S}^N$ optimally solves linear program

$$\begin{align*}
\text{minimize} \quad & \langle \delta(X^*), \delta(W) \rangle \\
\text{subject to} \quad & 0 \preceq \delta(W) \preceq 1 \\
& \text{tr} \ W = N - c
\end{align*}$$

Both direction matrix programs are derived from (1792a) whose analytical solution is known but is not necessarily unique. We emphasize (confer p.264): because this iteration (879) (880) (881) (initial $Y, W = 0$) is not a projection method (§4.4.1.1), success relies on existence of matrices in the feasible set of (879) having desired rank and diagonal cardinality. In particular, the feasible set of convex problem (879) is a Fantope (91) whose extreme points constitute the set of all normalized rank-one matrices; among those are found rank-one matrices of any desired diagonal cardinality.

Convex problem (879) is neither a relaxation of cardinality problem (875); instead, problem (879) becomes a convex equivalent to (875) at global convergence of iteration (879) (880) (881). Because the feasible set of convex problem (879) contains all normalized (§B.1) symmetric rank-one matrices of every nonzero diagonal cardinality, a constraint too low or high in cardinality $c$ will not prevent solution. An optimal rank-one solution $X^*$, whose diagonal cardinality is equal to cardinality of a principal eigenvector of matrix $A$, will produce the least residual Frobenius norm (to within machine noise processes) in the original problem statement (874).
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

Figure 112: Shepp-Logan phantom from MATLAB image processing toolbox.

4.6.0.0.13 Example. Compressive sampling of a phantom.
In Summer 2004, Candès, Romberg, & Tao [72] and Donoho [125] released papers on perfect signal reconstruction from samples that stand in violation of Shannon’s classical sampling theorem. These defiant signals are assumed sparse inherently or under some sparsifying affine transformation. Essentially, they proposed sparse sampling theorems asserting average sample rate independent of signal bandwidth and less than Shannon’s rate.

Minimum sampling rate:
- of $\Omega$-bandlimited signal: $2\Omega$ ([295, §3.2] Shannon)
- of $k$-sparse length-$n$ signal: $k \log_2(1+n/k)$ (Figure 105 Candès/Donoho)

Certainly, much was already known about nonuniform or random sampling [37] [218] and about subsampling or multirate systems [94] [369]. Vetterli, Marziliano, & Blu [378] had congealed a theory of noiseless signal reconstruction, in May 2001, from samples that violate the Shannon rate. [397, Sampling Sparsity] They anticipated the sparsifying transform by recognizing: it is the innovation (onset) of functions constituting a (not necessarily bandlimited) signal that determines minimum sampling rate for perfect reconstruction. Average onset (sparsity), Vetterli et alii call, the rate of innovation. Vector inner-products that Candès/Donoho call samples or measurements, Vetterli calls projections. From those projections Vetterli demonstrates reconstruction (by digital signal processing and “root finding”) of a Dirac comb, the very same prototypical signal from which Candès probabilistically derives minimum sampling rate [Compressive Sampling and Frontiers in Signal Processing, University of Minnesota, June 6, 2007]. Combining their terminology, we paraphrase a sparse sampling theorem:

- Minimum sampling rate, asserted by Candès/Donoho, $\propto$ Vetterli’s rate of innovation (a.k.a: information rate, degrees of freedom [ibidem June 5, 2007]).

What distinguishes these researchers are their methods of reconstruction.
Properties of the 1-norm were also well understood by June 2004 finding applications in deconvolution of linear systems \cite{86}, constrained linear regression (Lasso) \cite{359} \cite{329}, and basis pursuit \cite{80} \cite{223}. But never before had there been a formalized and rigorous sense that perfect reconstruction were possible by convex optimization of 1-norm when information lost in a subsampling process became nonrecoverable by classical methods. Donoho named this discovery *compressed sensing* to describe a nonadaptive perfect reconstruction method by means of linear programming. By the time Candès’ and Donoho’s landmark papers were finally published by IEEE in 2006, compressed sensing was old news that had spawned intense research which still persists; notably, from prominent members of the *wavelet* community.

Reconstruction of the *Shepp-Logan phantom* (Figure 112), from a severely aliased image (Figure 114) obtained by Magnetic Resonance Imaging (MRI), was the *impetus* driving Candès’ quest for a sparse sampling theorem. He realized that line segments appearing in the aliased image were regions of high total variation. There is great motivation, in the medical community, to apply compressed sensing to MRI because it translates to reduced scan-time which brings great technological and physiological benefits. MRI is now about 35 years old, beginning in 1973 with Nobel laureate Paul Lauterbur from Stony Brook USA. There has been much progress in MRI and compressed sensing since 2004, but there have also been indications of 1-norm abandonment (indigenous to reconstruction by compressed sensing) in favor of criteria closer to 0-norm because of a correspondingly smaller number of measurements required to accurately reconstruct a sparse signal.\footnote{\textsuperscript{4.52}Efficient techniques continually emerge urging 1-norm criteria abandonment; \cite{83} \cite{368} \cite{367}, §IDE, e.g., five techniques for compressed sensing are compared in \cite{38} demonstrating that 1-norm performance limits for cardinality minimization can be reliably exceeded.}

5481 complex samples (22 radial lines, $\approx 256$ complex samples per) were required in June 2004 to reconstruct a noiseless $256 \times 256$-pixel Shepp-Logan phantom by 1-norm minimization of an image-gradient integral estimate called total variation; \textit{id est}, 8.4\% subsampling of 65536 data. \cite[§1.1]{72} \cite[§3.2]{71} It was soon discovered that reconstruction of the Shepp-Logan phantom were possible with only 2521 complex samples (10 radial lines, Figure 113); 3.8\% subsampled data input to a (nonconvex) $\frac{1}{2}$-norm total-variation minimization. \cite[§IIIA]{78} The closer to 0-norm, the fewer the samples required for perfect reconstruction.

Passage of a few years witnessed an algorithmic speedup and dramatic reduction in minimum number of samples required for perfect reconstruction of the noiseless Shepp-Logan phantom. But minimization of total variation is ideally suited to recovery of any piecewise-constant image, like a phantom, because gradient of such images is highly sparse by design.

There is no inherent characteristic of real-life MRI images that would make reasonable an expectation of sparse gradient. Sparsification of a discrete image-gradient tends to preserve edges. Then minimization of total variation seeks an image having fewest edges. There is no deeper theoretical foundation than that. When applied to human brain scan or angiogram, with as much as 20\% of $256 \times 256$ Fourier samples, we have observed\footnote{\textsuperscript{4.53}Experiments with real-life images were performed by Christine Law at Lucas Center for Imaging, Stanford University.} a 30dB

\[4.52\]\[4.53\]
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

image/reconstruction-error ratio\(^4.54\) barrier that seems impenetrable by the total-variation objective. Total-variation minimization has met with moderate success, in retrospect, only because some medical images are moderately piecewise-constant signals. One simply hopes a reconstruction, that is in some sense equal to a known subset of samples and whose gradient is most sparse, is that unique image we seek.\(^4.55\)

The total-variation objective, operating on an image, is expressible as norm of a linear transformation (900). It is natural to ask whether there exist other sparsifying transforms that might break the real-life 30dB barrier (any sampling pattern @20% 256\(\times\)256 data) in MRI. There has been much research into application of wavelets, discrete cosine transform (DCT), randomized orthogonal bases, splines, etcetera, but with suspiciously little focus on objective measures like image/error or illustration of difference images; the predominant basis of comparison instead being subjectively visual (Duensing & Huang, ISMRM Toronto 2008).\(^4.56\) Despite choice of transform, there seems yet to have been a breakthrough of the 30dB barrier. Application of compressed sensing to MRI, therefore, remains fertile in 2008 for continued research.

Lagrangian form of compressed sensing in imaging

We now repeat Candès’ image reconstruction experiment from 2004 which led to discovery of sparse sampling theorems. [72, §1.2] But we achieve perfect reconstruction with an algorithm based on vanishing gradient of a compressed sensing problem’s Lagrangian, which is computationally efficient. Our contraction method (p.328) is fast also because matrix multiplications are replaced by fast Fourier transforms and number of constraints is cut in half by sampling symmetrically. Convex iteration for cardinality minimization (§4.5) is incorporated which allows perfect reconstruction of a phantom at 4.1% subsampling rate; 50% Candès’ rate. By making neighboring-pixel selection adaptive, convex iteration reduces discrete image-gradient sparsity of the Shepp-Logan phantom to 1.9%; 33% lower than previously reported.

We demonstrate application of discrete image-gradient sparsification to the \(n \times n = 256 \times 256\) Shepp-Logan phantom, simulating idealized acquisition of MRI data by radial sampling in the Fourier domain (Figure 113).\(^4.57\) Define a Nyquist-centric discrete Fourier transform (DFT) matrix

\(^4.54\)Noise considered here is due only to the reconstruction process itself; id est, noise in excess of that produced by the best reconstruction of an image from a complete set of samples in the sense of Shannon.

At less than 30dB image/error, artifacts generally remain visible to the naked eye. We estimate about 50dB is required to eliminate noticeable distortion in a visual A/B comparison.

\(^4.55\)In vascular radiology, diagnoses are almost exclusively based on morphology of vessels and, in particular, presence of stenoses. There is a compelling argument for total-variation reconstruction of magnetic resonance angiogram because it helps isolate structures of particular interest.

\(^4.56\)I have never calculated the PSNR of these reconstructed images [of Barbara]. − Jean-Luc Starck

The sparsity of the image is the percentage of transform coefficients sufficient for diagnostic-quality reconstruction. Of course the term “diagnostic quality” is subjective. . . . I have yet to see an “objective” measure of image quality. Difference images, in my experience, definitely do not tell the whole story. Often I would show people some of my results and get mixed responses, but when I add artificial Gaussian noise to an image, often people say that it looks better. − Michael Lustig

\(^4.57\)k-space is conventional acquisition terminology indicating domain of the continuous raw data provided by an MRI machine. An image is reconstructed by inverse discrete Fourier transform of that data interpolated on a Cartesian grid in two dimensions.
a symmetric (nonHermitian) unitary matrix characterized

\[
F = F^T
\]

\[
F^{-1} = F^H
\]

Denoting an unknown image \( U \in \mathbb{R}^{n \times n} \), its two-dimensional discrete Fourier transform \( \mathcal{F} \) is

\[
\mathcal{F}(U) \triangleq FUF
\]

hence the inverse discrete transform

\[
U = F^H \mathcal{F}(U) F^H
\]

From §A.1.1 no.31 we have a vectorized two-dimensional DFT via Kronecker product \( \otimes \)

\[
\text{vec} \mathcal{F}(U) \triangleq (F \otimes F) \text{vec} U
\]

and from (885) its inverse [176, p.24]

\[
\text{vec} U = (F^H \otimes F^H)(F \otimes F) \text{vec} U = (F^H F \otimes F^H F) \text{vec} U
\]

Idealized radial sampling in the Fourier domain can be simulated by Hadamard product \( \circ \) with a binary mask \( \Phi \in \mathbb{R}^{n \times n} \) whose nonzero entries could, for example, correspond with the radial line segments in Figure 113. To make the mask Nyquist-centric, like DFT matrix \( F \), define a circulant \([178]\) symmetric permutation matrix

\[
\Theta \triangleq \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \in \mathbb{S}^n
\]

Then given subsampled Fourier domain (MRI k-space) measurements in incomplete \( K \in \mathbb{C}^{n \times n} \), we might constrain \( \mathcal{F}(U) \) thus:

\[
\Theta \Phi \Theta \circ FUF = K
\]

and in vector form, (42) (1880)

\[
\delta(\text{vec} \Theta \Phi \Theta)(F \otimes F) \text{vec} U = \text{vec} K
\]

Because measurements \( K \) are complex, there are actually twice the number of equality constraints as there are measurements.

\[^{4.58}\text{MATLAB} \text{ fftshift()}\]
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

Figure 113: MRI radial sampling pattern, in DC-centric Fourier domain, representing 4.1% (10 lines) subsampled data. Only half of these complex samples, in any halfspace about the origin in theory, need be acquired for a real image because of conjugate symmetry. Due to MRI machine imperfections, samples are generally taken over full extent of each radial line segment. MRI acquisition time is proportional to number of lines.

We can cut that number of constraints in half via vertical and horizontal mask \( \Phi \) symmetry which forces the imaginary inverse transform to 0: The inverse subsampled transform in matrix form is

\[
F^H (\Theta \Phi \Theta \circ F \mathcal{U} F) F^H = F^H K F^H
\]

and in vector form

\[
(F^H \otimes F^H) \delta(\text{vec } \Theta \Phi \Theta)(F \otimes F) \text{vec } \mathcal{U} = (F^H \otimes F^H) \text{vec } K
\]

later abbreviated

\[
P \text{vec } \mathcal{U} = f
\]

where

\[
P \triangleq (F^H \otimes F^H) \delta(\text{vec } \Theta \Phi \Theta)(F \otimes F) \in \mathbb{C}^{n^2 \times n^2}
\]

Because of idempotence \( P = P^2 \), \( P \) is a projection matrix. Because of its Hermitian symmetry [176, p.24]

\[
P = (F^H \otimes F^H) \delta(\text{vec } \Theta \Phi \Theta)(F \otimes F) = (F \otimes F)^H \delta(\text{vec } \Theta \Phi \Theta)(F^H \otimes F^H) = P^H
\]

\( P \) is an orthogonal projector. \( P \text{vec } \mathcal{U} \) is real when \( P \) is real; id est, when for positive even integer \( n \)

\[
\Phi = \begin{bmatrix}
\Phi_{11} & \Phi(1,2:n) \Xi \\
\Xi \Phi(2:n,1) & \Xi \Phi(2:n,2:n) \Xi
\end{bmatrix} \in \mathbb{R}^{n \times n}
\]

\( \Phi \) is a diagonalization of matrix \( P \) whose binary eigenvalues are \( \delta(\text{vec } \Theta \Phi \Theta) \) while the corresponding eigenvectors constitute the columns of unitary matrix \( F^H \otimes F^H \).
where $\Xi \in S^{n-1}$ is the order-reversing permutation matrix (1820). In words, this necessary and sufficient condition on $\Phi$ (for a real inverse subsampled transform [295, p.53]) demands vertical symmetry about row $\frac{n}{2} + 1$ and horizontal symmetry about column $\frac{n}{2} + 1$.

Define
\[
\Delta \triangleq \begin{bmatrix}
1 & 0 & & & & & 0 \\
-1 & 1 & 0 & & & & \\
& -1 & 1 & \ddots & & & \\
& & \ddots & \ddots & \ddots & & \\
& & & \ddots & 1 & 0 & \\
0^T & & & & -1 & 1
\end{bmatrix} \in \mathbb{R}^{n \times n}
\] (897)

Express an image-gradient estimate
\[
\nabla U \triangleq \begin{bmatrix}
U \Delta \\
U \Delta^T \\
\Delta U \\
\Delta^T U
\end{bmatrix} \in \mathbb{R}^{4n \times n}
\] (898)

that is a simple first-order difference of neighboring pixels (Figure 115) to the right, left,
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

above, and below.\footnote{There is significant improvement in reconstruction quality by augmentation of a nominally two-point discrete image-gradient estimate to four points per pixel by inclusion of two polar directions. Improvement is due to centering; symmetry of discrete differences about a central pixel. We find small improvement on real-life images, \( \approx 1 \text{dB} \) empirically, by further augmentation with diagonally adjacent pixel differences.}

\[ \text{vec } \nabla U = \begin{bmatrix} \Delta^T \otimes I \\ \Delta \otimes I \\ I \otimes \Delta \\ I \otimes \Delta^T \end{bmatrix} \text{vec } U = \begin{bmatrix} \Psi_1 \\ \Psi_1^T \\ \Psi_2 \\ \Psi_2^T \end{bmatrix} \text{vec } U \triangleq \Psi \text{vec } U \in \mathbb{R}^{4n^2} \quad (899) \]

where \( \Psi \in \mathbb{R}^{4n^2 \times n^2} \). A total-variation minimization for reconstructing MRI image \( U \), that is known suboptimal \([217] [73] \), may be concisely posed

\[ \begin{aligned} & \text{minimize } \| \Psi \text{vec } U \|_1 \\ & \text{subject to } P \text{vec } U = f \end{aligned} \quad (900) \]

where

\[ f = (F^H \otimes F^H) \text{vec } K \in \mathbb{C}^{n^2} \quad (901) \]

is the known inverse subsampled Fourier data (a vectorized aliased image, Figure 114), and where a norm of discrete image-gradient \( \nabla U \) is equivalently expressed as norm of a linear transformation \( \Psi \text{vec } U \).

Although this simple problem statement (900) is equivalent to a linear program (§3.2), its numerical solution is beyond the capability of even the most highly regarded of contemporary commercial solvers.\footnote{for images as small as 128 \times 128 pixels. Obstacle to numerical solution is not a computer resource: e.g., execution time, memory. The obstacle is, in fact, inadequate numerical precision. Even when all dependent equality constraints are manually removed, the best commercial solvers fail simply because computer numerics become nonsense; \textit{id est}, numerical errors enter significant digits and the algorithm exits prematurely, loops indefinitely, or produces an infeasible solution.}

Our only recourse is to recast the problem in Lagrangian form (§3.1.2.2) and write customized code to solve it:

\[ \begin{aligned} & \text{minimize } \langle |\Psi \text{vec } U|, y \rangle \\ & \text{subject to } P \text{vec } U = f \end{aligned} \quad (a) \]

\[ \begin{aligned} & \text{minimize } \langle |\Psi \text{vec } U|, y \rangle + \frac{1}{2}\lambda \|P \text{vec } U - f\|_2^2 \quad (b) \end{aligned} \]

where multiobjective parameter \( \lambda \in \mathbb{R}_+ \) is quite large (\( \lambda \approx 1 \text{E8} \)) so as to enforce the equality constraint: \( P \text{vec } U - f = 0 \Leftrightarrow \|P \text{vec } U - f\|_2 = 0 \) (§A.7.1). We introduce a direction vector \( y \in \mathbb{R}^{4n^2} \) as part of a convex iteration (§4.5.3) to overcome that known suboptimal minimization of discrete image-gradient cardinality: \textit{id est}, there exists a vector \( y^* \) with entries \( y^*_i \in \{0, 1\} \) such that

\[ \begin{aligned} & \text{minimize } \| \Psi \text{vec } U \|_0 \\ & \text{subject to } P \text{vec } U = f \end{aligned} \]

\[ \equiv \text{minimize } \langle |\Psi \text{vec } U|, y^* \rangle + \frac{1}{2}\lambda \|P \text{vec } U - f\|_2^2 \quad (903) \]

Existence of such a \( y^* \), complementary to an optimal vector \( \Psi \text{vec } U^* \), is obvious by definition of global optimality \( \langle |\Psi \text{vec } U^*|, y^* \rangle = 0 \) (803) under which a cardinality-c optimal objective \( \| \Psi \text{vec } U^* \|_0 \) is assumed to exist.
Figure 115: Neighboring-pixel stencil [368] for image-gradient estimation on Cartesian grid. Implementation selects adaptively from darkest four • about central. Continuous image-gradient from two pixels holds only in a limit. For discrete differences, better practical estimates are obtained when centered.

Because \((902b)\) is an unconstrained convex problem, a zero objective function gradient is necessary and sufficient for optimality (\$2.13.3\); \textit{id est}, \((\S D.2.1)\)

\[
\Psi^T \delta(y) \text{sgn}(\Psi \text{vec} U) + \lambda P \text{H}(P \text{vec} U - f) = 0
\]

(904)

Because of \(P\) idempotence and Hermitian symmetry and \(\text{sgn}(x) = x/|x|\), this is equivalent to

\[
\lim_{\varepsilon \to 0} \left( \Psi^T \delta(y) \delta(|\Psi \text{vec} U| + \varepsilon 1)^{-1} \Psi + \lambda P \right) \text{vec} U = \lambda P f
\]

(905)

where small positive constant \(\varepsilon \in \mathbb{R}_+\) has been introduced for invertibility. Speaking more analytically, introduction of \(\varepsilon\) serves to uniquely define the objective’s gradient everywhere in the function domain; \textit{id est}, it transforms absolute value in \((902b)\) from a function differentiable almost everywhere into a differentiable function. An example of such a transformation in one dimension is illustrated in Figure 116. When small enough for practical purposes\(^{4.63}\) \((\varepsilon \approx 10^{-3})\), we may ignore the limiting operation. Then the mapping, for \(0 \leq y \leq 1\)

\[
\text{vec} U_{t+1} = \left( \Psi^T \delta(y) \delta(|\Psi \text{vec} U| + \varepsilon 1)^{-1} \Psi + \lambda P \right)^{-1} \lambda P f
\]

(906)

is a \textit{contraction} in \(U_t\) that can be solved recursively in \(t\) for its unique \textit{fixed point}; \textit{id est}, until \(U_{t+1} \to U_t\). \([237, \text{p.300}]\) \([213, \text{p.155}]\) Calculating this inversion directly is not possible.\(^{4.63}\)

\(^4.63\)We are looking for at least 50dB image/error ratio from only 4.1% subsampled data (10 radial lines in k-space). With this setting of \(\varepsilon\), we actually attain in excess of 100dB from a simple MATLAB program in about a minute on a 2006 vintage laptop Core 2 CPU (Intel T7400@2.16GHz, 666MHz FSB). By trading execution time and treating discrete image-gradient cardinality as a known quantity for this phantom, over 160dB is achievable.
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

\[
\int_{-1}^{x} \frac{y}{|y| + \epsilon} dy
\]

Figure 116: Real absolute value function \( f_2(x) = |x| \) on \( x \in [-1, 1] \) from Figure 70b superimposed upon integral of its derivative at \( \epsilon = 0.05 \) which smooths objective function.

for large matrices on contemporary computers because of numerical precision, so instead we apply the conjugate gradient method of solution to

\[
(\Psi^T \delta(y) \delta(|\Psi \text{vec} U| + \epsilon 1)^{-1} \Psi + \lambda P) \text{vec} U_{t+1} = \lambda P f
\]

(907)

which is linear in \( U_{t+1} \) at each recursion in the MATLAB program [383].

Observe that \( P \) (894), in the equality constraint from problem (902a), is not a fat matrix. Although number of Fourier samples taken is equal to the number of nonzero entries in binary mask \( \Phi \), matrix \( P \) is square but never actually formed during computation. Rather, a two-dimensional fast Fourier transform of \( U \) is computed followed by masking with \( \Theta \Phi \Theta \) and then an inverse fast Fourier transform. This technique significantly reduces memory requirements and, together with contraction method of solution, is the principal reason for relatively fast computation.

convex iteration

By convex iteration we mean alternation of solution to (902a) and (908) until convergence. Direction vector \( y \) is initialized to \( 1 \) until the first fixed point is found; which means, the contraction recursion begins calculating a (1-norm) solution \( U^* \) to (900) via problem (902b). Once \( U^* \) is found, vector \( y \) is updated according to an estimate of discrete image-gradient cardinality \( c \): Sum of the \( 4n^2 - c \) smallest entries of \( |\Psi \text{vec} U^*| \in \mathbb{R}^{4n^2} \) is the optimal objective value from a linear program, for \( 0 \leq c \leq 4n^2 - 1 \) (524)

\[
\sum_{i=c+1}^{4n^2} \pi(|\Psi \text{vec} U^*|)_i = \text{minimize } \langle |\Psi \text{vec} U^*|, y \rangle \\
\text{subject to } 0 \leq y \leq 1 \\
y^T 1 = 4n^2 - c
\]

(908)

where \( \pi \) is the nonlinear permutation-operator sorting its vector argument into nonincreasing order. An optimal solution \( y \) to (908), that is an extreme point of its feasible

\[^4.64\text{Conjugate gradient method requires positive definiteness. [160, §4.8.3.2]}
\[^4.65\text{Fat is typical of compressed sensing problems; e.g., [71] [78].}\]
set, is known in closed form: it has 1 in each entry corresponding to the \(4n^2 - c\) smallest entries of \(|\Psi \text{vec} U^*|\) and has 0 elsewhere. – p.291 Updated image \(U^*\) is assigned to \(U_t\), the contraction is recomputed solving (902b), direction vector \(y\) is updated again, and so on until convergence which is guaranteed by virtue of a monotonically nonincreasing real sequence of objective values in (902a) and (908).

There are two features that distinguish problem formulation (902b) and our particular implementation of it [383, MATLAB code]:

1) An image-gradient estimate may engage any combination of four adjacent pixels. In other words, the algorithm is not locked into a four-point gradient estimate (Figure 115); number of points constituting an estimate is directly determined by direction vector \(y\). Indeed, we find only \(c = 5092\) zero entries in \(y^*\) for the Shepp-Logan phantom; meaning, discrete image-gradient sparsity is actually closer to 1.9% than the 3% reported elsewhere; e.g., [367, §II].

2) Numerical precision of the fixed point of contraction (906) (≈1E-2 for perfect reconstruction @−103dB error) is a parameter to the implementation; meaning, direction vector \(y\) is updated after contraction begins but prior to its culmination. Impact of this idiosyncrasy tends toward simultaneous optimization in variables \(U\) and \(y\) while insuring \(y\) settles on a boundary point of its feasible set (nonnegative hypercube slice) in (908) at every iteration; for only a boundary point can yield the sum of smallest entries in \(|\Psi \text{vec} U^*|\).

Perfect reconstruction of the Shepp-Logan phantom (at 103dB image/error) is achieved in a MATLAB minute with 4.1% subsampled data (2671 complex samples); well below an 11% least lower bound predicted by the sparse sampling theorem. Because reconstruction approaches optimal solution to a 0-norm problem, minimum number of Fourier-domain samples is bounded below by cardinality of discrete image-gradient at 1.9%.

\[\square\]

4.6.0.0.14 Exercise. Contraction operator. Determine conditions on \(\lambda\) and \(\epsilon\) under which \(\Psi^T \delta(y)\delta(|\Psi \text{vec} U_t| + \epsilon 1)^{-1} \Psi + \lambda P\) from (907) is positive definite and (906) is a contraction.

4.6.0.0.15 Example. Eternity II. A tessellation puzzle game, playable by children, commenced world-wide in July 2007: introduced in London by Christopher Walter Monckton, 3rd Viscount Monckton of Brenchley. Called Eternity II, its name derives from an estimate of time that would pass while trying all allowable tilings of puzzle pieces before obtaining a complete solution. By the end of 2008, a complete solution had not yet been found although a $10,000 USD prize was awarded for a high score 467 (out of \(480 = 2\sqrt{M}(\sqrt{M} - 1)\)) obtained by heuristic methods. No prize was awarded for 2009 and 2010. Game-rules state that a $2M prize

---

\(^4.66\)This adaptive gradient was not contrived. It is an artifact of the convex iteration method for minimal cardinality solution; in this case, cardinality minimization of a discrete image-gradient.

\(^4.67\)Simultaneous optimization of these two variables \(U\) and \(y\) should never be a pinnacle of aspiration; for then, optimal \(y\) might not attain a boundary point.

\(^4.68\)That score means all but a few of the 256 pieces had been placed successfully (including the mandatory piece). Although distance between 467 to 480 is relatively small, there is apparently vast distance to a solution because no complete solution was found in 2009.
Figure 117: *Eternity* II is a board game in the puzzle genre. (a) Shown are all of the 16 puzzle pieces (indexed as in the tableau alongside) from a scaled-down computerized demonstration-version on the TOMY website. Puzzle pieces are square and triangularly partitioned into four colors (with associated symbols). Pieces may be moved, removed, and rotated at random on a $4 \times 4$ board. (b) Illustrated is one complete solution to this puzzle whose solution is not unique. The piece, whose border is lightly outlined, was placed last in this realization. There is no mandatory piece placement as for the full game, except the grey board-boundary. Solution time for a human is typically on the order of a minute. (c) This puzzle has four colors, indexed 1 through 4; grey corresponds to 0.
would be awarded to the first person who completely solves the puzzle before December 31, 2010, but the prize remains unclaimed after the deadline.

The full game comprises $M = 256$ square pieces and a $16 \times 16$ gridded board (Figure 118) whose complete tessellation is considered NP-hard.\footnote{Even so, combinatorial-intensity brute-force backtracking methods can solve similar puzzles in minutes given $M = 196$ pieces on a $14 \times 14$ test board; as demonstrated by Yannick Kirschhoffer. There is a steep rise in level of difficulty going to a $15 \times 15$ board.} \cite{354} A player may tile, retile, and rotate pieces, indexed 1 through 256, in any order face-up on the square board. Pieces are immutable in the sense that each is characterized by 4 colors (and their uniquely associated symbols), one at each edge, which are not necessarily the same per piece or from piece to piece; \textit{id est}, different pieces may or may not have some edge-colors in common. There are $L = 22$ distinct edge-colors plus a solid grey. The object of the game is to completely tile the board with pieces whose touching edges have identical color. The boundary of the board must be colored grey.

\textbf{full-game rules}

1) Any puzzle piece may be rotated face-up in quadrature and placed or replaced on the square board.

2) Only one piece may occupy any particular cell on the board.

3) All adjacent pieces must match in color (and symbol) at their touching edges.

4) Solid grey edges must appear all along the board’s boundary.

5) One mandatory piece (numbered 139 in the full game) must have a predetermined orientation in a predetermined cell on the board.

6) The board must be tiled completely (\textit{covered}).

A scaled-down demonstration version of the game is illustrated in Figure 117. Differences between the full game (Figure 118) and scaled-down game are the number of edge-colors $L$ (22 \textit{versus} 4, ignoring solid grey), number of pieces $M$ (256 \textit{versus} 16), and a single mandatory piece placement interior to the board in the full game. The scaled-down game has four distinct edge-colors, plus a solid grey, whose coding is illustrated in Figure 117c.

- $L = 22$ distinct edge-colors and number of puzzle pieces $M = 256$ and board-dimension $\sqrt{M} \times \sqrt{M} = 16 \times 16$ for the full game.

- There are $L = 4$ distinct edge-colors and $M = 16$ pieces and dimension $\sqrt{M} \times \sqrt{M} = 4 \times 4$ for the scaled-down demonstration board.
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

4.70 Were edge-points ordered sequentially with piece number, then this EDM would have a block-diagonal structure of known entries.

4.71 Translation occurs when a piece moves on the board in Figure 118, rotation occurs when colors are aligned with a neighboring piece.

4.72 There exists at least one solution; their exact number is unknown although Monckton insists they number in thousands. Ignoring boundary constraints and the single mandatory piece placement in the full game, a loose upper bound on number of combinations is $M! 4^M = 256! 4^{256}$. That number gets loosened: $150638!/(256!(150638−256)!)$ after presolving Eternity II (930).

Figure 118: Eternity II full-game board ($16 \times 16$, not actual size). Grid facilitates piece placement within unit-square cell; one piece per cell.

Euclidean distance intractability

If each square puzzle piece were characterized by four points in quadrature, one point representing board coordinates and color per edge, then Euclidean distance geometry would be suitable for solving this puzzle. Since all interpoint distances per piece are known, this game may be regarded as a Euclidean distance matrix completion problem in EDM$^{4M}$. Because distance information provides for reconstruction of point position to within an isometry (§5.5), piece translation and rotation are isometric transformations that abide by rules of the game. Convex constraints can be devised to prevent puzzle-piece reflection and to quantize rotation such that piece-edges stay aligned with the board boundary. (§5.5.2.0.1)

But manipulating such a large EDM is too numerically difficult for contemporary general-purpose semidefinite program solvers which incorporate interior-point methods; indeed, they are hard-pressed to find a solution for variable matrices of dimension as small as 100. Our challenge, therefore, is to express this game’s rules as constraints in a convex and numerically tractable way so as to find one solution from a googol of possible combinations.
permutation polyhedron strategy
To each puzzle piece, from a set of \( M \) pieces \( \{ P_i, i=1\ldots M \} \), assign an index \( i \) representing a unique piece-number. Each square piece is characterized by four colors, in quadrature, corresponding to its four edges. Each color \( p_{ij} \in \mathbb{R}^L \) is represented by \( e_\ell \in \mathbb{R}^L \) an \( L \)-dimensional standard basis vector or \( 0 \) if grey. These four edge-colors are represented in a \( 4 \times L \)-dimensional matrix; one matrix per piece
\[
P_i \triangleq \begin{bmatrix} p_{1i} & p_{2i} & p_{3i} & p_{4i} \end{bmatrix}^T \in \mathbb{R}^{4 \times L}, \quad i = 1\ldots M \tag{909}
\]
In other words, each distinct nongrey color is assigned a unique corresponding index \( \ell \in \{ 1\ldots L \} \) identifying a standard basis vector \( e_\ell \in \mathbb{R}^L \) (Figure 117c) that becomes a vector \( p_{ij} \in \{ e_1\ldots e_L, 0 \} \subset \mathbb{R}^L \) constituting matrix \( P_i \) representing a particular piece. Rows \( \{ p_{ij}^T, j=1\ldots 4 \} \) of \( P_i \) are ordered counterclockwise as in Figure 119. Color data is given in Figure 120 for the demonstration board. Then matrix \( P_i \) describes the \( i^{th} \) piece, excepting its orientation and location on the board.

Our strategy to solve Eternity II is to first vectorize the board, with respect to the whole pieces, and then relax a very hard combinatorial problem: All pieces are initially placed, as in Figure 120, in order of their given index. Then the vectorized game-board has initial state, as in Figure 120, represented within a matrix
\[
P \triangleq \begin{bmatrix} P_1 & \vdots & P_M \end{bmatrix} \in \mathbb{R}^{4M \times L} \tag{910}
\]
Moving pieces about the square board all at once corresponds to permuting pieces \( P_i \) on the vectorized board represented by matrix \( P \), while rotating the \( i^{th} \) piece is equivalent to circularly shifting row indices of \( P_i \) (rowwise permutation). This permutation problem, as stated, is doubly combinatorial (\( M! 4^M \) combinations) because we must find a permutation of pieces \( (M!) \)
\[
\Xi \in \mathbb{R}^{M \times M} \tag{911}
\]
and a rotation \( \Pi_i \in \mathbb{R}^{4 \times 4} \) of each individual piece \( (4^M) \) that solve the puzzle;
\[
(\Xi \otimes I_4) \Pi P = (\Xi \otimes I_4) \begin{bmatrix} \Pi_1 P_1 & \vdots & \Pi_M P_M \end{bmatrix} \in \mathbb{R}^{4M \times L} \tag{912}
\]
### 4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

<table>
<thead>
<tr>
<th>$P_1$</th>
<th>$[e_3 \ 0 \ 0 \ e_1]^T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_2$</td>
<td>$[e_2 \ e_4 \ e_4 \ e_4]^T$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>$[e_2 \ e_1 \ 0 \ e_1]^T$</td>
</tr>
<tr>
<td>$P_4$</td>
<td>$[e_4 \ e_1 \ 0 \ e_1]^T$</td>
</tr>
<tr>
<td>$P_5$</td>
<td>$[0 \ 0 \ e_3 \ e_1]^T$</td>
</tr>
<tr>
<td>$P_6$</td>
<td>$[e_2 \ e_2 \ e_4 \ e_2]^T$</td>
</tr>
<tr>
<td>$P_7$</td>
<td>$[e_2 \ e_3 \ 0 \ e_3]^T$</td>
</tr>
<tr>
<td>$P_8$</td>
<td>$[e_4 \ e_3 \ 0 \ e_3]^T$</td>
</tr>
<tr>
<td>$P_9$</td>
<td>$[0 \ e_3 \ e_3 \ 0]^T$</td>
</tr>
<tr>
<td>$P_{10}$</td>
<td>$[e_2 \ e_2 \ e_4 \ e_4]^T$</td>
</tr>
<tr>
<td>$P_{11}$</td>
<td>$[e_2 \ e_3 \ 0 \ e_1]^T$</td>
</tr>
<tr>
<td>$P_{12}$</td>
<td>$[e_4 \ e_1 \ 0 \ e_3]^T$</td>
</tr>
<tr>
<td>$P_{13}$</td>
<td>$[0 \ e_1 \ e_1 \ 0]^T$</td>
</tr>
<tr>
<td>$P_{14}$</td>
<td>$[e_2 \ e_2 \ e_4 \ e_4]^T$</td>
</tr>
<tr>
<td>$P_{15}$</td>
<td>$[e_2 \ e_1 \ 0 \ e_3]^T$</td>
</tr>
<tr>
<td>$P_{16}$</td>
<td>$[e_4 \ e_3 \ 0 \ e_1]^T$</td>
</tr>
</tbody>
</table>

Figure 120: Vectorized demo-game board has $M = 16$ matrices in $\mathbb{R}^{4 \times L}$ describing initial state of game pieces; 4 colors per puzzle-piece (Figure 119), $L = 4$ colors total in game (Figure 117c). So that color difference measurement remains unweighted, standard basis vectors in $\mathbb{R}^L$ represent color.
where

\[ \Pi_{i} \in \{ \pi_{1}, \pi_{2}, \pi_{3}, \pi_{4} \} \triangleq \begin{cases} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, & \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{cases} \]

(913)

Initial game-board state \( P \) (910) corresponds to \( \Xi = I \) and \( \Pi_{i} = \pi_{1} = I \ \forall \ i \).

Circulant \([178]\) permutation matrices \( \{ \pi_{1}, \pi_{2}, \pi_{3}, \pi_{4} \} \subset \mathbb{R}^{4 \times 4} \) correspond to clockwise piece-rotations \( \{ 0^\circ, 90^\circ, 180^\circ, 270^\circ \} \).

Rules of the game dictate that adjacent pieces on the square board have colors that match at their touching edges as in Figure 117b. A complete match is therefore equivalent to demanding that a constraint, comprising numeric color differences between \( 2\sqrt{M}(\sqrt{M} - 1) \) touching edges, vanish. Because the vectorized board layout is fixed and its cells are loaded or reloaded with pieces during play, locations of adjacent edges in \( \mathbb{R}^{4M \times L} \) are known a priori. We need simply form differences between colors from adjacent edges of pieces loaded into those known locations (Figure 121). Each difference may be represented by a constant vector from a set \( \{ \Delta_{i} \in \mathbb{R}^{4M}, i = 1 \ldots 2\sqrt{M}(\sqrt{M} - 1) \} \). Defining sparse constant fat matrix

\[ \Delta \triangleq \begin{bmatrix} \Delta_{1}^T \\ \vdots \\ \Delta_{2\sqrt{M}(\sqrt{M} - 1)}^T \end{bmatrix} \in \mathbb{R}^{2\sqrt{M}(\sqrt{M} - 1) \times 4M} \]

(915)

whose entries belong to \( \{-1, 0, 1\} \), then the desired constraint is

\[ \Delta(\Xi \otimes I_{4})\Pi P = 0 \in \mathbb{R}^{2\sqrt{M}(\sqrt{M} - 1) \times L} \]

(916)

Boundary of the square board must be colored grey. This means there are \( 4\sqrt{M} \) boundary locations in \( \mathbb{R}^{4M \times L} \) that must have value \( 0^T \). These can all be lumped into one equality constraint

\[ \beta^T(\Xi \otimes I_{4})\Pi P 1 = 0 \]

(917)

where \( \beta \in \mathbb{R}^{4M} \) is a sparse constant vector having entries in \{0, 1\} complementary to the known \( 4\sqrt{M} \) zeros (Figure 121).

By combining variables:

\[ \Phi \triangleq (\Xi \otimes I_{4})\Pi \in \mathbb{R}^{4M \times 4M} \]

(918)

this square matrix becomes a structured permutation matrix replacing the product of permutation matrices. Partition the composite variable \( \Phi \) into blocks

\[ \Phi \triangleq \begin{bmatrix} \phi_{11} & \cdots & \phi_{1M} \\ \vdots & \ddots & \vdots \\ \phi_{M1} & \cdots & \phi_{MM} \end{bmatrix} \in \mathbb{R}^{4M \times 4M} \]

(919)

4.73 Piece adjacencies on the square board map linearly to the vectorized board, of course.
Figure 121: Initial piece state. ○ indicate boundary $\beta$, line segments indicate differences $\Delta (915)$. Entries are indices $\ell$ identifying standard basis vectors $e_\ell \in \mathbb{R}^L$ from Figure 120.
Figure 122: Sparsity pattern for composite permutation matrix \( \Phi^{\star} \in \mathbb{R}^{4M \times 4M} \) representing solution from Figure 117b. Each four-point cluster represents a circulant permutation matrix from (913). Any \( M=16 \)-piece solution may be verified on the TOMY website.

where \( \Phi_{ij}^{\star} \in \{0, 1\} \) because (913)

\[
\phi_{ij}^{\star} \in \{\pi_1, \pi_2, \pi_3, \pi_4, 0\} \subset \mathbb{R}^{4 \times 4}
\]  

(920)

An optimal composite permutation matrix \( \Phi^{\star} \) is represented pictorially in Figure 122. Now we ask what are necessary conditions on \( \Phi^{\star} \) at optimality:

- 4M-sparsity and nonnegativity.
- Each column has one 1. Each row has one 1.
- Entries along each and every diagonal of each and every 4\( \times \)4 block \( \phi_{ij}^{\star} \) are equal.
- Corner pair of 2\( \times \)2 submatrices on antidiagonal of each and every 4\( \times \)4 block \( \phi_{ij}^{\star} \) are equal.

We want an objective function whose global optimum, if attained, certifies that the puzzle has been solved. Then, in terms of this \( \Phi^{\star} \) partitioning (919), the Eternity II problem is a minimization of cardinality

\[
\text{minimize} \quad \| \text{vec} \Phi \|_0
\]

subject to

\[
\Delta \Phi P = 0
\]

\[
\beta^T \Phi P 1 = 0
\]

\[
\Phi 1 = 1
\]

\[
\Phi^T 1 = 1
\]

\[
(I \otimes R_{d}) \Phi (I \otimes R_{d}^T) = (I \otimes S_{d}) \Phi (I \otimes S_{d}^T)
\]

\[
(I \otimes R_{\phi}) \Phi (I \otimes S_{\phi}^T) = (I \otimes S_{\phi}) \Phi (I \otimes R_{\phi}^T)
\]

\[
(e_{121} \otimes I_4)^T \Phi (e_{130} \otimes I_4) = \pi_3
\]

\[
\Phi \geq 0
\]

(921)
which is convex in the constraints where \( e_{121}, e_{139} \in \mathbb{R}^M \) are members of the standard basis representing mandatory piece placement in the full game, and where

\[
R_d \triangleq \begin{bmatrix}
1 & 0 \\
1 & 0 \\
1 & 0
\end{bmatrix} \in \mathbb{R}^{3 \times 4}, \quad S_d \triangleq \begin{bmatrix}
0 & 1 \\
0 & 1 \\
0 & 0
\end{bmatrix} \in \mathbb{R}^{3 \times 4}
\]

\[
R_\phi \triangleq \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix} \in \mathbb{R}^{2 \times 4}, \quad S_\phi \triangleq \begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} \in \mathbb{R}^{2 \times 4}
\]

(922)

(923)

These matrices \( R \) and \( S \) enforce circulance. Mandatory piece placement in the full game requires the equality constraint in \( \pi_3 \). Constraints \( \Phi_1 = 1 \) and \( \Phi^T 1 = 1 \) confine nonnegative \( \Phi \) to the permutation polyhedron (100) in \( \mathbb{R}^{4M \times 4M} \). The feasible set of problem (921) is an intersection of the permutation polyhedron with a large number of hyperplanes. Any vertex in the permutation polyhedron, which is the convex hull of permutation matrices, has minimal cardinality. The intersection contains a vertex of the permutation polyhedron because a solution \( \Phi^* \) cannot otherwise be a permutation matrix; such a solution is known to exist, so it must also be a vertex of the intersection.

In the vectorized variable, problem (921) is equivalent to

\[
\begin{align*}
\text{minimize} & \quad \| \text{vec} \Phi \|_0 \\
\text{subject to} & \quad \Sigma \text{vec} \Phi = 0 \\
& \quad (P^T \otimes \Delta) \text{vec} \Phi = 0 \\
& \quad (P1 \otimes \beta)^T \text{vec} \Phi = 0 \\
& \quad (1_M \otimes I_{4M}) \text{vec} \Phi = 1_{4M} \\
& \quad (I_{4M} \otimes 1_M) \text{vec} \Phi = 1_{4M} \\
& \quad (I \otimes R_d \otimes I \otimes R_d - I \otimes S_d \otimes I \otimes S_d) \text{vec} \Phi = 0 \\
& \quad (I \otimes S_\phi \otimes I \otimes R_\phi - I \otimes R_\phi \otimes I \otimes S_\phi) \text{vec} \Phi = 0 \\
& \quad (e_{139} \otimes I_4 \otimes e_{121} \otimes I_4)^T \text{vec} \Phi = \text{vec} \pi_3 \\
& \quad \Phi \geq 0
\end{align*}
\]

(924)

This problem is abbreviated:

\[
\begin{align*}
\text{minimize} & \quad \| \text{vec} \Phi \|_0 \\
\text{subject to} & \quad E \text{vec} \Phi = \tau \\
& \quad \Phi \geq 0
\end{align*}
\]

(925)

where \( E \in \mathbb{R}^{2L \sqrt{M(M-1)} + 8M + 13M^2 + 17 \times 16M^2} \) is sparse and optimal objective value is \( 4M \); dimension of \( E \) is \( 864,593 \times 1,048,576 \). A compressed sensing paradigm \( [72] \) is not inherent here. To solve this by linear programming, a direction vector is introduced for cardinality minimization as in §4.5. It is critical, in this case, to add enough random noise to the

4.74 Meaning that piece numbered 139 by the game designer must be placed in cell 121 on the vectorized board (Figure 118) with orientation \( \pi_3 \) (p.336).

4.75 Since \( 0 \) is the trivial circulant matrix, application is democratic over all blocks \( \phi_{ij} \).

4.76 Vertex means zero-dimensional exposed face (§2.6.1.0.1); intersection with a strictly supporting hyperplane. There can be no further intersection with a feasible affine subset that would enlarge that face; id est, a vertex of the permutation polyhedron persists in the feasible set.
direction vector so as to insure a vertex solution \([97, \text{p.158}].\) For the demonstration game, in fact, choosing a direction vector randomly will find an optimal solution in only a few iterations.\(^4.77\) But for the full game, numerical errors prevent solution of (925); number of equality constraints 864,593 is too large.\(^4.78\) So again, we reformulate the problem:

**canonical Eternity II**

Because each block \(\phi_{ij}\) of \(\Phi\) is optimally circulant having only four degrees of freedom (920), we may take as variable every fourth column of \(\Phi\):

\[
\tilde{\Phi} \triangleq [\Phi(:,1) \ \Phi(:,5) \ \Phi(:,9) \ \cdots \ \Phi(:,4M-3)] \in \mathbb{R}^{4M \times M} \tag{926}
\]

where \(\tilde{\Phi}_{ij} \in \{0,1\}\). Then, for \(e_i \in \mathbb{R}^4\)

\[
\Phi = (\tilde{\Phi} \otimes e_i^T) + (I \otimes \pi_4)(\tilde{\Phi} \otimes e_2^T) + (I \otimes \pi_3)(\tilde{\Phi} \otimes e_3^T) + (I \otimes \pi_2)(\tilde{\Phi} \otimes e_4^T) \in \mathbb{R}^{4M \times 4M} \tag{927}
\]

From §A.1.1 no.31 and no.40

\[
\text{vec } \Phi = (I \otimes e_1 \otimes I_{4M} + I \otimes e_2 \otimes I \otimes \pi_4 + I \otimes e_3 \otimes I \otimes \pi_3 + I \otimes e_4 \otimes I \otimes \pi_2) \text{ vec } \tilde{\Phi} \\
\triangleq Y \text{ vec } \tilde{\Phi} \in \mathbb{R}^{16M^2} \tag{928}
\]

where \(Y \in \mathbb{R}^{16M^2 \times 4M^2}\). Permutation polyhedron (100) now demands that each consecutive quadruple of adjacent rows of \(\tilde{\Phi}\) sum to 1: \((I \otimes 1^T_M)\tilde{\Phi}=1\). Constraints in \(R\) and \(S\) (which are most numerous) may be dropped because circulance of \(\phi_{ij}\) is built into \(\tilde{\Phi}\)-reconstruction formula (927). Eternity II (924) is thus equivalently transformed

\[
\begin{align*}
\min_{\tilde{\Phi} \in \mathbb{R}^{4M \times M}} & \quad \| \text{vec } \tilde{\Phi} \|_0 \\
\text{subject to} & \quad (P^T \otimes \Delta)Y \text{ vec } \tilde{\Phi} = 0 \\
& \quad (P1 \otimes \beta^T)Y \text{ vec } \tilde{\Phi} = 0 \\
& \quad (1^T \otimes I \otimes 1^T_M) \text{ vec } \tilde{\Phi} = 1 \\
& \quad (I \otimes 1^T_M) \text{ vec } \tilde{\Phi} = 1 \\
& \quad (e_{139} \otimes e_1 \otimes e_{121} \otimes I_M)Y \text{ vec } \tilde{\Phi} = \pi_3 e_1 \\
& \quad \tilde{\Phi} \succeq 0
\end{align*} \tag{929}
\]

whose optimal objective value is \(M\) with \(\tilde{\Phi}^*\)-entries in \(\{0,1\}\) and where \(e_1 \in \mathbb{R}^4\) (§A.1.1 no.39) and \(e_{121}, e_{139} \in \mathbb{R}^M\). In abbreviation

\[
\begin{align*}
\min_{\tilde{\Phi} \in \mathbb{R}^{4M \times M}} & \quad \| \text{vec } \tilde{\Phi} \|_0 \\
\text{subject to} & \quad \tilde{E} \text{ vec } \tilde{\Phi} = \tilde{\tau} \\
& \quad \tilde{\Phi} \succeq 0
\end{align*} \tag{930}
\]

\(^4.77\)This can only mean: there are many optimal solutions. A simplex-method solver is required for numerical solution; interior-point methods will not work. A randomized direction vector also works for Clue Puzzles provided by the toy maker: similar 6\(\times\)6 and 6\(\times\)12 puzzles whose solution each provide a clue to solution of the full game. Even better is a nonnegative uniformly distributed randomized direction vector having \(4M\) entries (\(M\) entries, in case (930)), corresponding to the largest entries of \(\Phi^*\), zeroed.

\(^4.78\)Saunders’ program \texttt{lpsol} can reduce that number to 797,508 constraints by eliminating linearly dependent rows of matrix \(E\), but that reduction is not enough to overcome numerical issues with the best contemporary linear program solvers.
of reformulation (929), number of equality constraints is now 11,077; an order of magnitude fewer constraints than (925) from where sparse $\tilde{E} \in \mathbb{R}^{2L\sqrt{M}(\sqrt{M}-1)+2M\times 5\times 4M^2}$ replaces the $E$ matrix. Number of columns in matrix $\tilde{E}$ has been reduced from a million to 262,144; dimension of $\tilde{E}$ goes to 11,077 $\times$ 262,144. But this dimension remains out of reach of most highly regarded contemporary academic and commercial linear program solvers because of numerical failure; especially disappointing insofar as sparsity of $\tilde{E}$ is high with only 0.07% nonzero entries $\in \{-1, 0, 1, 2\}$; element $\{2\}$ arising only in the $\beta$ constraint which is soon to disappear after presolving.

Variable vec $\Phi$ itself is too large in dimension. Notice that the constraint in $\beta$, which zeroes the board at its edges, has all positive coefficients. The zero sum means that all vec $\tilde{\Phi}$ entries, corresponding to nonzero entries in row vector $(P_1 \otimes \beta)^T Y$, must be zero. For the full game, this means we may immediately eliminate 57,840 variables from 262,144. After zero-row and dependent row removal, dimension of $\tilde{E}$ goes to 10,054 $\times$ 204,304 with entries in $\{-1, 0, 1\}$.

### polyhedral cone theory

Eternity II problem (930) constraints are interpretable in the language of convex cones: The columns of matrix $\tilde{E}$ constitute a set of generators for a pointed polyhedral cone $K = \{ \tilde{E} \text{vec} \tilde{\Phi} | \tilde{\Phi} \geq 0 \}$. (§2.12.2.2) Even more intriguing is the observation: vector $\tilde{\tau}$ resides on that polyhedral cone’s boundary. (§2.13.4.2.4)

We may apply techniques from §2.13.4.3 to prune generators not belonging to the smallest face of that cone, to which $\tilde{\tau}$ belongs, because generators of that smallest face must hold a minimal cardinality solution. Matrix dimension is thereby reduced: The $i^{th}$ column $\tilde{E}(::, i)$ of matrix $\tilde{E}$ belongs to the smallest face $F$ of $K$ that contains $\tilde{\tau}$ if and only if

$$\begin{align*}
\text{find } & \tilde{\Phi}, \mu \\
\text{subject to } & \mu \tilde{\tau} - \tilde{E}(::, i) = \tilde{E} \text{vec} \tilde{\Phi} \\
& \text{vec} \tilde{\Phi} \succeq 0
\end{align*}$$

is feasible. By a transformation of Saunders, this linear feasibility problem is the same as

$$\begin{align*}
\text{find } & \tilde{\Phi}, \mu \\
\text{subject to } & \tilde{E} \text{vec} \tilde{\Phi} = \mu \tilde{\tau} \\
& \text{vec} \tilde{\Phi} \succeq 0 \\
& (\text{vec} \tilde{\Phi})_i \geq 1
\end{align*}$$

A minimal cardinality solution to Eternity II (930) implicitly constrains $\tilde{\Phi}^*$ to be binary. So this test (931) of membership to $F(K \ni \tilde{\tau})$ may be tightened to a test of $(\text{vec} \tilde{\Phi})_i = 1$;
id est, for \(i = 1 \ldots 4M^2\) distinct feasibility problems

\[
\begin{align*}
\text{find} & \quad \tilde{\Phi} \\
\text{subject to} & \quad \tilde{E} \text{vec} \tilde{\Phi} = \tilde{\tau} \\
& \quad \text{vec} \tilde{\Phi} \succeq 0 \\
& \quad (\text{vec} \tilde{\Phi})_i = 1
\end{align*}
\]  \tag{932}

whose feasible set is a proper subset of that in (931). Real variable \(\mu\) can be set to 1 because if it must not be, then feasible \((\text{vec} \tilde{\Phi})_i = 1\) could not be feasible to Eternity II (930). If infeasible here in (932), then the only choice remaining for \((\text{vec} \tilde{\Phi})_i = 0\); meaning, column \(\tilde{E}(\cdot, i)\) may be discarded after all columns have been tested. This tightened problem (932) therefore tells us two things when feasible: \(\tilde{E}(\cdot, i)\) belongs to the smallest face of \(\mathcal{K}\) that contains \(\tilde{\tau}\), and \((\text{vec} \tilde{\Phi})_i\) constitutes a nonzero vertex-coordinate of permutation polyhedron (100). After presolving via this conic pruning method (with subsequent zero-row and dependent row removal), dimension of \(\tilde{E}\) goes to 7,362 \(\times\) 150,638.

generators of smallest face are conically independent
Designate \(A \in \mathbb{R}^{7362 \times 150638} \triangleq \mathbb{R}^{m \times n}\) as matrix \(\tilde{E}\) after having discarded all generators not in the smallest face \(\mathcal{F}\) of cone \(\mathcal{K}\) that contains \(\tilde{\tau}\). The Eternity II problem (930) becomes

\[
\begin{align*}
\text{minimize} & \quad \|x\|_0 \\
\text{subject to} & \quad Ax = b \\
& \quad x \succeq 0
\end{align*}
\]  \tag{933}

To further prune all generators relatively interior to that smallest face, we may subsequently test for conic dependence as described in §2.10 (280): for \(i = 1 \ldots 150,638\)

\[
\begin{align*}
\text{find} & \quad x \\
\text{subject to} & \quad Ax = A(\cdot, i) \\
& \quad x \succeq 0 \\
& \quad x_i = 0
\end{align*}
\]  \tag{934}

where \(x\) is vec \(\tilde{\Phi}\) corresponding to columns of \(\tilde{E}\) not previously discarded by (932).\footnote{4.80} If feasible, then column \(A(\cdot, i)\) is a conically dependent generator of the smallest face and must be discarded from matrix \(A\) before proceeding with test of remaining columns. It turns out, for Eternity II: generators of the smallest face, previously found via (932), comprise a minimal set; id est, (934) is never feasible and so no column of \(A\) can be discarded (\(A\) remains unaltered).\footnote{4.81}

affinity for maximization
Designate vector \(b \in \mathbb{R}^m\) to be \(\tilde{\tau}\) after discarding all entries corresponding to dependent rows in \(\tilde{E}\); id est, \(b\) is \(\tilde{\tau}\) subsequent to presolving. Then Eternity II resembles Figure 32a

\footnote{4.80}Discarded entries in vec \(\tilde{\Phi}\) are optimally 0.\footnote{4.81}One cannot help but notice a binary selection of variable by tests (932) and (934): Geometrical test (932) (smallest face) checks feasibility of vector entry 1 while geometrical test (934) (conic independence) checks feasibility of 0. Changing 1 to 0 in (932) is always feasible for Eternity II.
4.6. CARDINALITY AND RANK CONSTRAINT EXAMPLES

(not (b)) because variable $x$ is implicitly bounded above by design; 1 ⪰ $x$ by confinement of $\Phi$ in (921) to the permutation polyhedron (100), for $i=1\ldots150,638$

$$1 = \max_x x_i \quad \text{subject to} \ Ax = b \quad x \geq 0 \quad (935)$$

Unity is always attainable, by (932). By (926) this means (§4.5.1.4)

$$M = \max_{y(x), x} (1 - y)^T x \quad \text{subject to} \ Ax = b \quad x \geq 0$$

$$\equiv \max_x \|x\|_M \quad \text{subject to} \ Ax = b \quad x \geq 0 \quad (936)$$

where

$$y = 1 - \nabla \|x\|_M$$

is a direction vector from the technique of convex iteration in §4.5.1.1 and $\|x\|_M$ is a $k$-largest norm (§3.2.2.1, $k=M$). When upper bound $M$ in (936) is met, solution $x$ will be optimal for Eternity II because it must then be a Boolean vector with minimal cardinality $M$.

Maximization of convex function $\|x\|_M$ (monotonic on $\mathbb{R}_+^n$) is not a convex problem, though the constraints are convex. [318, §32] This problem formulation is unusual, nevertheless, insofar as its geometrical visualization is quite clear. We therefore choose to work with a complementary direction vector $z$, in what follows, in predilection for a mental picture of convex function maximization.

**Direction vector is optimal solution at global convergence**

Instead of solving (936), which is difficult, we propose iterating a convex problem sequence: for $1 - y \leftarrow z$

$$\max_{x \in \mathbb{R}_+^n} z^T x \quad \text{subject to} \ Ax = b \quad x \geq 0 \quad (937)$$

$$\max_{z \in \mathbb{R}^n} z^T x^* \quad \text{subject to} \ 0 \leq z \leq 1 \quad z^T 1 = M \quad (525)$$

Variable $x$ is implicitly bounded above at unity by design of $A$. A globally optimal complementary direction vector $z^*$ will always exactly match an optimal solution $x^*$ for convex iteration of any problem formulated as maximization of a Boolean variable; here we have

$$z^T x^* \triangleq M \quad (938)$$
rumination

- Adding a few more clue pieces makes the problem harder in the sense that solution space is diminished; the target gets smaller.
- Because \( z^* = x^* \), Eternity II can instead be formulated equivalently as a convex-quadratic maximization:

\[
\begin{align*}
\text{maximize} & \quad x^T x \\
\text{subject to} & \quad Ax = b \\
\end{align*}
\]

a nonconvex problem but requiring no convex iteration. If it were possible to form a nullspace basis \( Z \), for \( A \) of about equal sparsity,\(^4.82\) such that

\[
x = Z \xi + x_p \quad (115)
\]

then the equivalent problem formulation

\[
\begin{align*}
\text{maximize} & \quad (Z \xi + x_p)^T (Z \xi + x_p) \\
\text{subject to} & \quad Z \xi + x_p \succeq 0
\end{align*}
\]

might invoke optimality conditions as obtained in \([207, \text{thm.}8]\)^4.83.

\[\Box\]

### 4.7 Constraining rank of indefinite matrices

Example 4.7.0.0.1, which follows, demonstrates that convex iteration is more generally applicable: to indefinite or nonsquare matrices \( X \in \mathbb{R}^{m \times n} \); not only to positive semidefinite matrices. Indeed,

\[
\begin{align*}
\text{find} & \quad X \\
\text{subject to} & \quad \text{rank } X \leq k
\end{align*}
\]

\[
\begin{align*}
\text{find} & \quad X, Y, Z \\
\text{subject to} & \quad \text{rank } X \leq k
\end{align*}
\]

\[
G = \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix}
\]

\[
\text{rank } G \leq k
\]

**Proof.** \( \text{rank } G \leq k \Rightarrow \text{rank } X \leq k \) because \( X \) is the projection of composite matrix \( G \) on subspace \( \mathbb{R}^{m \times n} \). For symmetric \( Y \) and \( Z \), any rank-k positive semidefinite composite matrix \( G \) can be factored into rank-k terms \( R \): \( G = R^T R \) where \( R \equiv [B \ C] \) and \( \text{rank } B, \text{rank } C \leq \text{rank } R \) and \( B \in \mathbb{R}^{k \times m} \) and \( C \in \mathbb{R}^{k \times n} \). Because \( Y \) and \( Z \) and \( X = B^T C \) are variable, \((1547)\) rank \( X \leq \text{rank } B, \text{rank } C \leq \text{rank } R = \text{rank } G \) is tight.

---

\(^4.82\)Define sparsity as ratio of number of nonzero entries to matrix-dimension product. For matrices, the average number of nonzeros per row or column is easier to understand and likely to be small for typical LP problems, independent of the dimensions. —Michael Saunders

\(^4.83\) ...the assumptions in Theorem 8 ask for the \( Q \) being positive definite (see the top of the page of Theorem 8). I must confess that I do not remember why. —Jean-Baptiste Hiriart-Urruty
4.7. CONSTRAINING RANK OF INDEFINITE MATRICES

So there must exist an optimal direction vector \( W^* \) such that

\[
\begin{align*}
\text{find} & \quad X, Y, Z \\
\text{subject to} & \quad X \in \mathcal{C} \\
& \quad G = \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \\
& \quad \text{rank} \ G \leq k \\
\text{minimize} & \quad \langle G, W^* \rangle \\
\text{subject to} & \quad X \in \mathcal{C} \\
& \quad G = \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \succeq 0
\end{align*}
\] (942)

Were \( W^* = I \), by (1778) the optimal resulting trace objective would be equivalent to the minimization of nuclear norm of \( X \) over \( \mathcal{C} \). This means:

- (confer p.194) The argument of any nuclear norm minimization problem may be replaced with a composite semidefinite variable of the same optimal rank but doubly dimensioned.

Then Figure 88 becomes an accurate geometrical description of a consequent composite semidefinite problem objective. But there are better direction vectors than Identity \( I \) which occurs only under special conditions:

4.7.0.0.1 Example. Compressed sensing, compressive sampling.

As our modern technology-driven civilization acquires and exploits ever-increasing amounts of data, everyone now knows that most of the data we acquire can be thrown away with almost no perceptual loss — witness the broad success of lossy compression formats for sounds, images, and specialized technical data. The phenomenon of ubiquitous compressibility raises very natural questions: Why go to so much effort to acquire all the data when most of what we get will be thrown away? Can’t we just directly measure the part that won’t end up being thrown away?

—David Donoho [125]

Lossy data compression techniques like JPEG are popular, but it is also well known that compression artifacts become quite perceptible with signal postprocessing that goes beyond mere playback of a compressed signal. [230] [255] Spatial or audio frequencies presumed masked by a simultaneity are not encoded, for example, so rendered imperceptible even with significant postfiltering (of the compressed signal) that is meant to reveal them; \textit{id est}, desirable artifacts are forever lost, so highly compressed data is not amenable to analysis and postprocessing; \textit{e.g.}, sound effects [101] [102] [104] or image enhancement (Adobe Photoshop).\textsuperscript{4.84} Further, there can be no universally acceptable unique metric of perception for gauging exactly how much data can be tossed. For these reasons, there will always be need for raw (noncompressed) data.

In this example we throw out only so much information as to leave perfect reconstruction within reach. Specifically, the MIT logo in Figure 123 is perfectly reconstructed from 700 time-sequential samples \( \{y_i\} \) acquired by the one-pixel camera illustrated in Figure 124. The MIT-logo image in this example impinges a 46×81

\textsuperscript{4.84} As simple a process as upward scaling of signal amplitude or image size will always introduce noise; even to a noncompressed signal. But scaling-noise is particularly noticeable in a JPEG-compressed image; \textit{e.g.}, text or any sharp edge.
array micromirror device. This mirror array is modulated by a pseudonoise source that independently positions all the individual mirrors. A single photodiode (one pixel) integrates incident light from all mirrors. After stabilizing the mirrors to a fixed but pseudorandom pattern, light so collected is then digitized into one sample $y_i$ by analog-to-digital (A/D) conversion. This sampling process is repeated with the micromirror array modulated to a new pseudorandom pattern.

The most important questions are: How many samples do we need for perfect reconstruction? Does that number of samples represent compression of the original data?

We claim that perfect reconstruction of the MIT logo can be reliably achieved with as few as 700 samples $y = [y_i] \in \mathbb{R}^{700}$ from this one-pixel camera. That number represents only 19% of information obtainable from 3726 micromirrors.\footnote{That number (700 samples) is difficult to achieve, as reported in [311, §6]. If a minimal basis for the MIT logo were instead constructed, only five rows or columns worth of data (from a $46 \times 81$ matrix) are linearly independent. This means a lower bound on achievable compression is about $5 \times 46 = 230$ samples plus 81 samples column encoding; which corresponds to 8% of the original information. (Figure 125)}

Our approach to reconstruction is to look for low-rank solution to an underdetermined system:

\[
\begin{align*}
\text{find} & \quad X \in \mathbb{R}^{46 \times 81} \\
\text{subject to} & \quad A \text{vec } X = y \\
\text{rank } X & \leq 5
\end{align*}
\] (943)

where vec $X$ is the vectorized (37) unknown image matrix. Each row of fat matrix $A$ is one realization of a pseudorandom pattern applied to the micromirrors. Since these patterns are deterministic (known), then the $i^{th}$ sample $y_i$ equals $A(i, :) \text{vec } Y$; id est, $y_i = A \text{vec } Y$.

Perfect reconstruction here means optimal solution $X^*$ equals scene $Y \in \mathbb{R}^{46 \times 81}$ to within machine precision.

Because variable matrix $X$ is generally not square or positive semidefinite, we constrain
Figure 124: One-pixel camera. Compressive imaging camera block diagram. Incident lightfield (corresponding to the desired image $Y$) is reflected off a digital micromirror device (DMD) array whose mirror orientations are modulated in the pseudorandom pattern supplied by the random number generators (RNG). Each different mirror pattern produces a voltage at the single photodiode that corresponds to one measurement $y_i$. \cite{355, 402}

Figure 125: Estimates of compression for various encoding methods:
1) linear interpolation (140 samples),
2) minimal columnar basis (311 samples),
3) convex iteration (700 samples) can achieve lower bound predicted by compressed sensing (670 samples, $n=46\times81$, $k=140$, Figure 105) whereas nuclear norm minimization alone does not \cite[§6]{311},
4) JPEG @100% quality (2588 samples),
5) no compression (3726 samples).
its rank by rewriting the problem equivalently

\[
\begin{align*}
\text{find} & & X \\
\text{subject to} & & \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \leq 5 \\
& & A \text{vec} \, X = y
\end{align*}
\]

This rank constraint on the composite (block) matrix insures \( \text{rank} \, X \leq 5 \) for any choice of dimensionally compatible matrices \( W_1 \) and \( W_2 \). But to solve this problem by convex iteration, we alternate solution of semidefinite program

\[
\begin{align*}
\text{minimize} & & \text{tr} \left( \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \right) \\
\text{subject to} & & A \text{vec} \, X = y \\
& & \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0
\end{align*}
\]

with semidefinite program

\[
\begin{align*}
\text{minimize} & & \text{tr} \left( \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \right) \\
\text{subject to} & & 0 \preceq Z \preceq I \\
& & \text{tr} \, Z = 46 + 81 - 5
\end{align*}
\]

(which has optimal solution known in closed form, p.565) until a rank-5 composite matrix is found.

With 1000 samples \( \{y_i\} \), convergence occurs in two iterations; 700 samples require more than ten iterations but reconstruction remains perfect. Iterating more admits taking of fewer samples. Reconstruction is independent of pseudorandom sequence parameters; e.g., binary sequences succeed with the same efficiency as Gaussian or uniformly distributed sequences.

4.7.1 rank-constraint midsummary

We find that this direction matrix idea works well and quite independently of desired rank or affine dimension. This idea of direction matrix is good principally because of its simplicity: When confronted with a problem otherwise convex if not for a rank or cardinality constraint, then that constraint becomes a linear regularization term in the objective.

There exists a common thread through all these Examples; that being, convex iteration with a direction matrix as normal to a linear regularization (a generalization of the well-known trace heuristic). But each problem type (per Example) possesses its own idiosyncrasies that slightly modify how a rank-constrained optimal solution is actually obtained: The ball packing problem in Chapter 5.4.2.2.6, for example, requires a problem sequence in a progressively larger number of balls to find a good initial value for the direction matrix, whereas many of the examples in the present chapter require an initial value of 0. Finding a Boolean solution in Example 4.6.0.0.9 requires a procedure to detect
stalls, while other problems have no such requirement. The combinatorial Procrustes problem in Example 4.6.0.0.3 allows use of a known closed-form solution for direction vector when solved via rank constraint, but not when solved via cardinality constraint. Some problems require a careful weighting of the regularization term, whereas other problems do not, and so on. It would be nice if there were a universally applicable method for constraining rank; one that is less susceptible to quirks of a particular problem type.

Poor initialization of the direction matrix from the regularization can lead to an erroneous result. We speculate one reason to be a simple dearth of optimal solutions of desired rank or cardinality; an unfortunate choice of initial search direction leading astray. Ease of solution by convex iteration occurs when optimal solutions abound. With this speculation in mind, we now propose a further generalization of convex iteration for constraining rank that attempts to ameliorate quirks and unify problem types:

### 4.8 Convex Iteration rank-1

We now develop a general method for constraining rank that first decomposes a given problem via standard diagonalization of matrices (§A.5). This method is motivated by observation (§4.4.1) that an optimal direction matrix can be simultaneously diagonalizable with an optimal variable matrix. This suggests minimization of an objective function directly in terms of eigenvalues. A second motivating observation is that variable orthogonal matrices seem easily found by convex iteration; e.g., Procrustes Example 4.6.0.0.2.

#### 4.8.1 rank-1 transformation

It turns out that this general method always requires solution to a rank-1 constrained problem regardless of desired rank \( \rho \) from the original problem. To demonstrate, we pose a semidefinite feasibility problem

\[
\begin{align*}
\text{find} & \quad X \in \mathbb{S}^n \\
\text{subject to} & \quad A \operatorname{svec} X = b \\
& \quad X \succeq 0 \\
& \quad \text{rank} X \leq \rho
\end{align*}
\]

(947)

given an upper bound \( 0 < \rho < n \) on rank, a vector \( b \in \mathbb{R}^m \), and typically fat full-rank

\[
A = \begin{bmatrix} \operatorname{svec}(A_1)^T \\ \vdots \\ \operatorname{svec}(A_m)^T \end{bmatrix} \in \mathbb{R}^{m \times n(n+1)/2}
\]

(681)

where \( A_i \in \mathbb{S}^n \), \( i = 1 \ldots m \). So, for symmetric matrix vectorization \( \operatorname{svec} \) as defined in (56),

\[
A \operatorname{svec} X = \begin{bmatrix} \operatorname{tr}(A_1 X) \\ \vdots \\ \operatorname{tr}(A_m X) \end{bmatrix}
\]

(682)

In Convex Optimization, an optimal solution generally comes from a convex set of optimal solutions; (§3.1.2.1) that set can be large.
This program (947) is a statement of the classical problem of finding a matrix \( X \) of maximum rank \( \rho \) in the intersection of the positive semidefinite cone with a given number \( m \) of hyperplanes in the subspace of symmetric matrices \( \mathbb{S}^n \). [27, §II.13] [25, §2.2] Such a matrix is presumed to exist.

To begin transformation of (947), express the nonincreasingly ordered diagonalization (§A.5.1) of positive semidefinite variable matrix

\[
X \triangleq QAQ^T = \sum_{i=1}^{n} \lambda_i Q_{ii} \in \mathbb{S}^n
\]  

which is a sum of rank-1 orthogonal projection matrices \( Q_{ii} \) weighted by eigenvalues \( \lambda_i \) where \( Q_{ij} \triangleq q_i q_j^T \in \mathbb{R}^{n \times n} \), \( Q = [q_1 \cdots q_n] \in \mathbb{R}^{n \times n} \), \( Q^T = Q^{-1} \), \( \Lambda_{ii} = \lambda_i \in \mathbb{R} \), and

\[
\Lambda = \begin{bmatrix}
\lambda_1 & 0 \\
\lambda_2 & \ddots \\
0 & \ddots \\
\end{bmatrix} \in \mathbb{S}^n
\]  

where \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \). Recall the fact:

\[
\Lambda \succeq 0 \iff X \succeq 0 \quad (1532)
\]

From orthogonal matrix \( Q \) in ordered diagonalization (948) of variable \( X \), take a matrix

\[
U \triangleq [u_1 \cdots u_\rho] \triangleq Q(:, 1: \rho) \sqrt{\Lambda(1: \rho, 1: \rho)} = [\sqrt{\lambda_1} q_1 \cdots \sqrt{\lambda_\rho} q_\rho] \in \mathbb{R}^{n \times \rho}
\]  

Then \( U \) has orthogonal but unnormalized columns;

\[
X = UU^T = \sum_{i=1}^{\rho} u_i u_i^T = \sum_{i=1}^{\rho} U_{ii} = \sum_{i=1}^{\rho} \lambda_i q_i q_i^T \in \mathbb{S}^n
\]  

Make an assignment

\[
Z = \begin{bmatrix}
[u_1] \\
\vdots \\
[u_\rho]
\end{bmatrix} \begin{bmatrix}
[u_1^T] \\
\vdots \\
[u_\rho^T]
\end{bmatrix} \in \mathbb{S}^{n \times \rho}
\]

\[
= \begin{bmatrix}
U_{11} & \cdots & U_{1\rho} \\
\vdots & \ddots & \vdots \\
U_{\rho1} & \cdots & U_{\rho\rho}
\end{bmatrix} \triangleq \begin{bmatrix}
[u_1 u_1^T] & \cdots & u_1 u_\rho^T \\
\vdots & \ddots & \vdots \\
u_\rho u_1^T & \cdots & u_\rho u_\rho^T
\end{bmatrix}
\]  

(952)
Then transformation of (947) to its rank-1 equivalent is:

\[
\begin{align*}
\text{find } & \quad X = \sum_{i=1}^{\rho} U_{ii} \\
\text{subject to } & \quad Z = \begin{bmatrix} U_{11} & \cdots & U_{1\rho} \\ \vdots & \ddots & \vdots \\ U_{1\rho}^T & \cdots & U_{\rho\rho} \end{bmatrix} \succeq 0 \\
& \quad A \text{ svec } \sum_{i=1}^{\rho} U_{ii} = b \\
& \quad \text{tr}(U_{ij}) = 0 \quad i < j = 2 \ldots \rho \\
& \quad \text{rank } Z = 1
\end{align*}
\]

(953)

Symmetry is necessary and sufficient for positive semidefiniteness of a rank-1 matrix. (§A.3.1.0.7) Matrix \(X\) is positive semidefinite whenever \(Z\) is. (§A.3.1.0.4, §A.3.1.0.2) This new problem always enforces a rank-1 constraint on matrix \(Z\); \textit{id est}, regardless of upper bound on rank \(\rho\) of variable matrix \(X\), this equivalent problem always poses a rank-1 constraint. We propose solving (953) by iteration of convex problem

\[
\begin{align*}
\text{minimize } & \quad \text{tr}(Z W) \\
\text{subject to } & \quad Z = \begin{bmatrix} U_{11} & \cdots & U_{1\rho} \\ \vdots & \ddots & \vdots \\ U_{1\rho}^T & \cdots & U_{\rho\rho} \end{bmatrix} \succeq 0 \\
& \quad A \text{ svec } \sum_{i=1}^{\rho} U_{ii} = b \\
& \quad \text{tr}(U_{ij}) = 0 \quad i < j = 2 \ldots \rho
\end{align*}
\]

(954)

with convex problem

\[
\begin{align*}
\text{minimize } & \quad \text{tr}(Z^* W) \\
\text{subject to } & \quad 0 \preceq W \preceq I \\
& \quad \text{tr}(W) = n\rho - 1
\end{align*}
\]

(955)

the latter providing direction of search \(W\) for a rank-1 matrix \(Z\) in (954). These convex problems (954) (955) are iterated until a rank-1 \(Z\) matrix is found (until the objective of (954) vanishes). Initial value of direction matrix \(W\) is the Identity. For subsequent iterations, an optimal solution to (955) has closed form (p.565).

Because of the nonconvex nature of a rank-constrained problem, there can be no proof of global convergence of this convex iteration. But this iteration always converges to a local minimum because the sequence of objective values is monotonic and nonincreasing; any monotonically nonincreasing real sequence converges. [268, §1.2] [43, §1.1] A rank \(\rho\) matrix \(X\) solving the original problem (947) is found when the objective in (954) converges to 0; a certificate of global optimality for the convex iteration. In practice, incidence of global convergence is quite high (99.99% [385]); failures being mostly attributable to numerical
accuracy. Upper bound $\rho$ on rank of positive semidefinite matrix $X$ is assured by rank-1 optimal matrix $Z$.

### 4.8.1.0.1 Example. Singular value decomposition by convex iteration. [167]

This diagonal decomposition technique (transformation to a rank-1 problem) is extensible to other problem types; e.g., [235, §III]. Rank-1 transformation makes singular value decomposition (SVD, §A.6) possible by convex iteration because orthogonality constraints may then be introduced. We learn that any uniqueness properties, the SVD of rank-$\rho$ matrix $X \triangleq USV^T \in \mathbb{R}^{m \times n}$ might enjoy, stem from demand for singular vector orthonormality.\textsuperscript{4.87}

Assignment $Z \in \mathbb{S}_{+}^{2mp+np+p+1}$ is key to finding the SVD of $X$ by convex optimization:

\[
\begin{align*}
\text{find} & \quad U, \delta(S), V \\
\text{subject to} & \quad Z = \begin{bmatrix}
1 & \text{vec}(H)^T & \text{vec}(U)^T & \delta(S)^T & \text{vec}(V)^T \\
\text{vec}H & J & \delta(S) & \text{vec}V \\
\delta(S) & \geq 0 \\
H = US \\
X = HV^T \\
HU^T & \text{symmetry} \\
U^TH & \text{perpendicularity} \\
\text{tr}(H(:,i)H(:,i)^T) = S(i,i)^2 & \quad i = 1 \ldots \rho \\
\text{tr}(H(:,i)U(:,i)^T) = S(i,i) & \quad i = 1 \ldots \rho \\
H & \text{orthogonality} \\
U, V & \text{orthonormality} \\
\text{rank } Z & = 1
\end{bmatrix} \geq 0
\end{align*}
\]

(956)

where variable matrix $J \in \mathbb{S}^{2mp+np+p+1}$ is a large partition of $Z$, where given rank-$\rho$ matrix $X \in \mathbb{R}^{m \times n}$ is subject to SVD in unknown orthonormal matrices $U \in \mathbb{R}^{m \times \rho}$ and $V \in \mathbb{R}^{n \times \rho}$ and unknown diagonal matrix of singular values $S \in \mathbb{R}^{\rho \times \rho}$, and where introduction of variable $H \triangleq US \in \mathbb{R}^{m \times \rho}$ makes identification of input $X = HV^T$ possible within partition $J$. Orthogonality constraints on columns of $H$, within $J$, and orthonormality constraints on columns of $U$ and $V$ are critical; \textit{videlicet}, $h \perp v \iff \text{tr}(hv^T) = 0$; $v^Tv = 1 \iff \text{tr}(vv^T) = 1$.

Symmetric matrix $Z$ is positive semidefinite rank-1 at optimality, regardless of $\rho$. That rank constraint is the only nonconvex constraint in (956); the only constraint that cannot be directly implemented in a convex manner per partition $J$. But the rank constraint is handled well by convex iteration. Matlab implementation of SVD by convex iteration is intricate although incidence of global convergence is 99.99% [399], barring numerical error.

\textsuperscript{4.87} Otherwise, there exist many similarly structured tripartite nonorthogonal matrix decompositions; in place of $\rho$ nonzero singular values, diagonal matrix $S$ would instead hold exactly $\rho$ coordinates; orthonormal columns in $U$ and $V$ would become merely linearly independent.
4.8. CONVEX ITERATION RANK-1

Figure 126: $W_1$, $W_2$, and $W_3$ represent the last three direction vectors in a sequence. $m_1$ represents the midpoint between direction vectors $W_1$ and $W_2$; $m_2$ is the midpoint of $W_2$ and $W_3$. Straight line passes through midpoints.

4.8.2 convex iteration accelerant

Convex iteration can be made to converge faster; sometimes, by orders of magnitude. The idea here is to determine whether the last three direction vectors are close to their fit to a straight line. When three direction vectors are close to a straight line, then the last direction vector may be replaced with its extrapolation along that line.

To reduce computation time, the fitted line is not a best fit. Instead, the midpoint between each pair of iteration-adjacent direction vectors is calculated (Figure 126). A straight line is uniquely defined by two midpoints in any dimension. The distance of each direction vector to the line is calculated, then those three distances summed. When a sum is small, three direction vectors are deemed close to the line determined by them.

What is meant above by close and small depends on the particular problem type at hand. For the parameters and normalized random data chosen for two Matlab realizations [385] [399] on $W_1$imization (corresponding to problems (953) and (956)), small is numerically defined to be 1 or less in the statement if $\text{straight} < 1$ whose purpose is to determine straightness of the last three direction vectors of convex iteration. The smaller the value of the normalized sum called straight, the closer the last three direction vectors are to a straight line. Variable straight is bounded below by 0 which indicates three direction vectors precisely on the line going through them.

If linear extrapolation goes too far, then the objective of convex iteration will increase or a solver may fail numerically. In either case, one must forget the last iteration and back up the linear extrapolation until the objective decreases. These techniques are illustrated by the Matlab programs.
Chapter 5

Euclidean Distance Matrix

These results [(988)] were obtained by Schoenberg (1935), a surprisingly late date for such a fundamental property of Euclidean geometry.

—John Clifford Gower [172, §3]

By itself, distance information between many points in Euclidean space is lacking. We might want to know more; such as, relative or absolute position or dimension of some hull. A question naturally arising in some fields (e.g., geodesy, economics, genetics, psychology, biochemistry, engineering) [109] asks what facts can be deduced given only distance information. What can we know about the underlying points that the distance information purports to describe? We also ask what it means when given distance information is incomplete; or suppose the distance information is not reliable, available, or specified only by certain tolerances (affine inequalities). These questions motivate a study of interpoint distance, well represented in any spatial dimension by a simple matrix from linear algebra.5.1 In what follows, we will answer some of these questions via Euclidean distance matrices.

5.1 e.g. $\sqrt{D} \in \mathbb{R}^{N \times N}$, a classical two-dimensional matrix representation of absolute interpoint distance because its entries (in ordered rows and columns) can be written neatly on a piece of paper. Matrix $D$ will be reserved throughout to hold distance-square.
CHAPTER 5. EUCLIDEAN DISTANCE MATRIX

Figure 127: Convex hull of three points \( N = 3 \) is shaded in \( \mathbb{R}^3 \). Dotted lines are imagined vectors to points whose affine dimension is 2.

5.1 EDM

Euclidean space \( \mathbb{R}^n \) is a finite-dimensional real vector space having an inner product defined on it, inducing a metric. [237, §3.1] A Euclidean distance matrix, an EDM in \( \mathbb{R}^{N \times N}_+ \), is an exhaustive table of distance-square \( d_{ij} \) between points taken by pair from a list of \( N \) points \( \{ x_\ell, \ell = 1 \ldots N \} \) in \( \mathbb{R}^n \); the squared metric, the measure of distance-square:

\[
d_{ij} = \| x_i - x_j \|^2_2 = (x_i - x_j, x_i - x_j) \tag{957}
\]

Each point is labelled ordinally, hence the row or column index of an EDM, \( i \) or \( j = 1 \ldots N \), individually addresses all the points in the list.

Consider the following example of an EDM for the case \( N = 3 \):

\[
D = [d_{ij}] = \begin{bmatrix}
d_{11} & d_{12} & d_{13} \\
d_{21} & d_{22} & d_{23} \\
d_{31} & d_{32} & d_{33}
\end{bmatrix}
= \begin{bmatrix}
0 & d_{12} & d_{13} \\
d_{12} & 0 & d_{23} \\
d_{13} & d_{23} & 0
\end{bmatrix}
= \begin{bmatrix}
0 & 1 & 5 \\
1 & 0 & 4 \\
5 & 4 & 0
\end{bmatrix} \tag{958}
\]

Matrix \( D \) has \( N^2 \) entries but only \( N(N-1)/2 \) pieces of information. In Figure 127 are shown three points in \( \mathbb{R}^3 \) that can be arranged in a list to correspond to \( D \) in (958). But such a list is not unique because any rotation, reflection, or translation (§5.5) of those points would produce the same EDM \( D \).
5.2  First metric properties

For $i,j = 1 \ldots N$, absolute distance between points $x_i$ and $x_j$ must satisfy the defining requirements imposed upon any metric space: \[\text{[237, §1.1] [268, §1.7]}\] namely, for Euclidean metric $\sqrt{d_{ij}}$ ($\S$5.4) in $\mathbb{R}^n$

1. $\sqrt{d_{ij}} \geq 0$, $i \neq j$  
   - nonnegativity
2. $\sqrt{d_{ij}} = 0 \iff x_i = x_j$  
   - selfdistance
3. $\sqrt{d_{ij}} = \sqrt{d_{ji}}$  
   - symmetry
4. $\sqrt{d_{ij}} \leq \sqrt{d_{ik}} + \sqrt{d_{kj}}$, $i \neq j \neq k$  
   - triangle inequality

Then all entries of an EDM must be in concord with these Euclidean metric properties: specifically, each entry must be nonnegative, \textsuperscript{5.2} the main diagonal must be 0, \textsuperscript{5.3} and an EDM must be symmetric. The fourth property provides upper and lower bounds for each entry. Property 4 is true more generally when there are no restrictions on indices $i,j,k$, but furnishes no new information.

5.3  $\exists$ fifth Euclidean metric property

The four properties of the Euclidean metric provide information insufficient to certify that a bounded convex polyhedron more complicated than a triangle has a Euclidean realization. \[\text{[172, §2]}\] Yet any list of points or the vertices of any bounded convex polyhedron must conform to the properties.

5.3.0.0.1 Example.  Triangle.

Consider the EDM in (958), but missing one of its entries:

$$D = \begin{bmatrix} 0 & 1 & d_{13} \\ 1 & 0 & 4 \\ d_{31} & 4 & 0 \end{bmatrix}$$

Can we determine unknown entries of $D$ by applying the metric properties? Property 1 demands $\sqrt{d_{13}}, \sqrt{d_{31}} \geq 0$, property 2 requires the main diagonal be 0, while property 3 makes $\sqrt{d_{31}} = \sqrt{d_{13}}$. The fourth property tells us

$$1 \leq \sqrt{d_{13}} \leq 3$$

Indeed, described over that closed interval $[1,3]$ is a family of triangular polyhedra whose angle at vertex $x_2$ varies from 0 to $\pi$ radians. So, yes we can determine the unknown entries of $D$, but they are not unique; nor should they be from the information given for this example. \[\square\]

\textsuperscript{5.2}Implicit from the terminology, $\sqrt{d_{ij}} \geq 0 \iff d_{ij} \geq 0$ is always assumed.

\textsuperscript{5.3}What we call selfdistance, Marsden calls nondegeneracy. \[\text{[268, §1.6]}\] Kreyszig calls these first metric properties axioms of the metric; \[\text{[237, p.4]}\] Blumenthal refers to them as postulates. \[\text{[52, p.15]}\]
CHAPTER 5. EUCLIDEAN DISTANCE MATRIX

Figure 128: (a) Complete dimensionless EDM graph. (b) Emphasizing obscured segments $x_2x_4$, $x_4x_3$, and $x_2x_3$, now only five $(2N - 3)$ absolute distances are specified. EDM so represented is incomplete, missing $d_{14}$ as in (961), yet the isometric reconstruction (§5.4.2.2.10) is unique as proved in §5.9.3.0.1 and §5.14.4.1.1. First four properties of Euclidean metric are not a recipe for reconstruction of this polyhedron.

5.3.0.0.2 Example. Small completion problem, I.
Now consider the polyhedron in Figure 128b formed from an unknown list $\{x_1, x_2, x_3, x_4\}$. The corresponding EDM less one critical piece of information, $d_{14}$, is given by

$$D = \begin{bmatrix} 0 & 1 & 5 & d_{14} \\ 1 & 0 & 4 & 1 \\ 5 & 4 & 0 & 1 \\ d_{14} & 1 & 1 & 0 \end{bmatrix}$$

(961)

From metric property 4 we may write a few inequalities for the two triangles common to $d_{14}$; we find

$$\sqrt{5} - 1 \leq \sqrt{d_{14}} \leq 2$$

(962)

We cannot further narrow those bounds on $\sqrt{d_{14}}$ using only the four metric properties (§5.8.3.1.1). Yet there is only one possible choice for $\sqrt{d_{14}}$ because points $x_2, x_3, x_4$ must be collinear. All other values of $\sqrt{d_{14}}$ in the interval $[\sqrt{5} - 1, 2]$ specify impossible distances in any dimension; *id est*, in this particular example the triangle inequality does not yield an interval for $\sqrt{d_{14}}$ over which a family of convex polyhedra can be reconstructed. □
5.3. \textit{Fifth Euclidean Metric Property} 359

We will return to this simple Example 5.3.0.2 to illustrate more elegant methods of solution in §5.8.3.1.1, §5.9.3.0.1, and §5.14.4.1.1. Until then, we can deduce some general principles from the foregoing examples:

- Unknown $d_{ij}$ of an EDM are not necessarily uniquely determinable.
- The triangle inequality does not produce necessarily tight bounds.\(^{5.4}\)
- Four Euclidean metric properties are insufficient for reconstruction.

5.3.1 Lookahead

There must exist at least one requirement more than the four properties of the Euclidean metric that makes them altogether necessary and sufficient to certify realizability of bounded convex polyhedra. Indeed, there are infinitely many more; there are precisely $N + 1$ necessary and sufficient Euclidean metric requirements for $N$ points constituting a generating list (§2.3.2). Here is the fifth requirement:

5.3.1.0.1 Fifth Euclidean metric property. Relative-angle inequality.

(confer §5.14.2.1.1) Augmenting the four fundamental properties of the Euclidean metric in $\mathbb{R}^n$, for all $i, j, \ell \neq k \in \{1 \ldots N\}$, $i < j < \ell$, and for $N \geq 4$ distinct points $\{x_k\}$, the inequalities

$$
\begin{align*}
\cos(\theta_{ik\ell} + \theta_{\ellkj}) & \leq \cos \theta_{ikj} \leq \cos(\theta_{ik\ell} - \theta_{\ellkj}) \\
0 & \leq \theta_{ik\ell}, \theta_{\ellkj}, \theta_{ikj} \leq \pi
\end{align*}
$$

(963)

where $\theta_{ikj} = \theta_{jki}$ represents angle between vectors at vertex $x_k$ (1035) (Figure 129), must be satisfied at each point $x_k$ regardless of affine dimension.$^\Diamond$

We will explore this in §5.14. One of our early goals is to determine matrix criteria that subsume all the Euclidean metric properties and any further requirements. Looking ahead, we will find (1314) (988) (992)

$$
\begin{align*}
-z^T D z & \geq 0 \\
1^T z & = 0 \\
(\forall \|z\| = 1) \\
D & \in \mathbb{S}_N^N
\end{align*}
$$

(64)

where the convex cone of Euclidean distance matrices $\mathbb{EDM}^N \subseteq \mathbb{S}_N^N$ belongs to the subspace of symmetric hollow matrices (§2.2.3.0.1). (Numerical test \texttt{isedm}(D) provided on \textit{Wolframization} [401].) Having found equivalent matrix criteria, we will see there is a bridge from bounded convex polyhedra to EDMs in §5.9.5.6.

Now we develop some invaluable concepts, moving toward a link of the Euclidean metric properties to matrix criteria.$^5$

\(^{5.4}\)The term tight with reference to an inequality means equality is achievable.

\(^{5.5}\) The main diagonal.

\(^{5.6}\) From an EDM, a generating list (§2.3.2, §2.12.2) for a polyhedron can be found (§5.12) correct to within a rotation, reflection, and translation (§5.5).
Figure 129: Fifth Euclidean metric property nomenclature. Each angle $\theta$ is made by a vector pair at vertex $k$ while $i, j, k, \ell$ index four points at the vertices of a generally irregular tetrahedron. The fifth property is necessary for realization of four or more points; a reckoning by three angles in any dimension. Together with the first four Euclidean metric properties, this fifth property is necessary and sufficient for realization of four points.
5.4 EDM definition

Ascribe points in a list \( \{ x_\ell \in \mathbb{R}^n, \ \ell = 1 \ldots N \} \) to the columns of a matrix

\[
X = [ x_1 \cdots x_N ] \in \mathbb{R}^{n \times N} \tag{76}
\]

where \( N \) is regarded as the cardinality of list \( X \). When matrix \( D = [ d_{ij} ] \) is an EDM, its entries must be related to those points constituting the list by the Euclidean distance-square: for \( i, j = 1 \ldots N \) (\[A.1.1\] no. 33)

\[
d_{ij} = \| x_i - x_j \|^2 = (x_i - x_j)^T (x_i - x_j) = \| x_i \|^2 + \| x_j \|^2 - 2x_i^T x_j
\]

\[
= [ x_i^T \ x_j^T ] \begin{bmatrix} I & -I \\ -I & I \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix} = \text{vec}(X)^T (\Phi_{ij} \otimes I) \text{vec} X = \langle \Phi_{ij}, X^T X \rangle
\]

where

\[
\text{vec} X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^{nN} \tag{966}
\]

and where \( \otimes \) signifies Kronecker product (\[D.1.2.1\]). \( \Phi_{ij} \otimes I \) is positive semidefinite (1564) having \( I \in \mathbb{S}^n \) in its \( ii \)th and \( jj \)th block of entries while \( -I \in \mathbb{S}^n \) fills its \( ij \)th and \( ji \)th block; i.e.,

\[
\Phi_{ij} \triangleq \delta((e_i e_j^T + e_j e_i^T)1) - (e_i e_j^T + e_j e_i^T) \in \mathbb{S}_+^N
\]

\[
= (e_i - e_j)(e_i - e_j)^T
\]

(967)

where \( \{ e_i \in \mathbb{R}^n, \ i = 1 \ldots N \} \) is the set of standard basis vectors. Thus each entry \( d_{ij} \) is a convex quadratic function (\[A.4.0.0.2\]) of \( \text{vec} X \) (37). [318, \[6\]

The collection of all Euclidean distance matrices \( \text{EDM}^N \) is a convex subset of \( \mathbb{R}^{N \times N} \) called the EDM cone (\[6\], Figure 164 p.481):

\[
0 \in \text{EDM}^N \subseteq \mathbb{S}_+^N \cap \mathbb{R}^{N \times N} \subset \mathbb{S}^N
\]

(968)

An EDM \( D \) must be expressible as a function of some list \( X \); i.e., it must have the form

\[
D(X) \triangleq \delta(X^T X)1^T + 1 \delta(X^T X)^T - 2X^T X \in \text{EDM}^N \tag{969}
\]

\[
= [ \text{vec}(X)^T (\Phi_{ij} \otimes I) \text{vec} X, \ i, j = 1 \ldots N ] \tag{970}
\]

Function \( D(X) \) will make an EDM given any \( X \in \mathbb{R}^{n \times N} \), conversely, but \( D(X) \) is not a convex function of \( X \) (\[5.4.1\]). Now the EDM cone may be described:

\[
\text{EDM}^N = \left\{ D(X) \mid X \in \mathbb{R}^{N-1 \times N} \right\}
\]

(971)
Expression $D(X)$ is a matrix definition of EDM and so conforms to the Euclidean metric properties:

Nonnegativity of EDM entries (property 1, §5.2) is obvious from the distance-square definition (965), so holds for any $D$ expressible in the form $D(X)$ in (969).

When we say $D$ is an EDM, reading from (969), it implicitly means the main diagonal must be 0 (property 2, selfdistance) and $D$ must be symmetric (property 3); $\delta(D) = 0$ and $D^T = D$ or, equivalently, $D \in \mathbb{S}_N^+$ are necessary matrix criteria.

5.4.0.1 homogeneity

Function $D(X)$ is homogeneous in the sense, for $\zeta \in \mathbb{R}$

$$\sqrt{\zeta} D(\zeta X) = |\zeta| \sqrt{D(X)}$$

(972)

where the positive square root is entrywise (^o).

Any nonnegatively scaled EDM remains an EDM; id est, the matrix class EDM is invariant to nonnegative scaling ($\alpha D(X)$ for $\alpha \geq 0$) because all EDMs of dimension $N$ constitute a convex cone $\text{EDM}_N^+$ (§6, Figure 156).

5.4.1 $-V_N^T D(X) V_N$ convexity

We saw that EDM entries $d_{ij} \left( \begin{array}{c} x_i \\ x_j \end{array} \right)$ are convex quadratic functions. Yet $-D(X)$ (969) is not a quasiconvex function of matrix $X \in \mathbb{R}^{n \times N}$ because the second directional derivative (§3.8)

$$-\frac{d^2}{dt^2} \bigg|_{t=0} D(X + tY) = 2(-\delta(Y^T Y)1^T - 1\delta(Y^T Y)^T + 2Y^T Y)$$

(973)

is indefinite for any $Y \in \mathbb{R}^{n \times N}$ since its main diagonal is 0. [168, §4.2.8] [212, §7.1 prob.2]

Hence $-D(X)$ can neither be convex in $X$.

The outcome is different when instead we consider

$$-V_N^T D(X) V_N = 2V_N^T X^T X V_N$$

(974)

where we introduce the full-rank skinny Schoenberg auxiliary matrix (§B.4.2)

$$V_N \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & 0 & \cdots \\ \vdots & \ddots & \ddots \\ 0 & \cdots & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1^T \\ I \end{bmatrix} \in \mathbb{R}^{N \times N-1}$$

(975)

($\mathcal{N}(V_N) = 0$) having range

$$\mathcal{R}(V_N) = \mathcal{N}(1^T), \quad V_N^T 1 = 0$$

(976)
5.4. EDM DEFINITION

Matrix-valued function (974) meets the criterion for convexity in §3.7.3.0.2 over its domain that is all of $\mathbb{R}^{n \times N}$; videlicet, for any $Y \in \mathbb{R}^{n \times N}$

$$\frac{d^2}{dt^2} V_N^T D(X + t Y)V_N = 4 V_N^T YTV_N \succeq 0 \quad (977)$$

Quadratic matrix-valued function $-V_N^T D(X)V_N$ is therefore convex in $X$ achieving its minimum, with respect to a positive semidefinite cone ($\mathbb{R}$ that is all of $\mathbb{R}$), positive definite. Matrix-valued function ($\mathbb{R}$; $\mathbb{R}$) becomes impossible because (977) could not then be positive definite.

5.4.2 Gram-form EDM definition

Positive semidefinite matrix $X^T X$ in (969), formed from inner product of list $X$, is known as a Gram matrix; [260, §3.6]

$$G \triangleq X^T X = \begin{bmatrix} x_1^T & \cdots & x_N^T \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix} = \begin{bmatrix} \|x_1\|^2 & x_1^T x_2 & \cdots & x_1^T x_N \\ x_2^T x_1 & \|x_2\|^2 & \cdots & x_2^T x_N \\ \vdots & \vdots & \ddots & \vdots \\ x_N^T x_1 & x_N^T x_2 & \cdots & \|x_N\|^2 \end{bmatrix} \in \mathbb{S}^N$$

$$= \delta \left( \begin{bmatrix} \|x_1\| \\ \|x_2\| \\ \vdots \\ \|x_N\| \end{bmatrix} \right) \begin{bmatrix} 1 & \cos \psi_{12} & \cdots & \cos \psi_{1N} \\ \cos \psi_{12} & 1 & \cdots & \cos \psi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \cos \psi_{1N} & \cos \psi_{2N} & \cdots & 1 \end{bmatrix} \delta \left( \begin{bmatrix} \|x_1\| \\ \|x_2\| \\ \vdots \\ \|x_N\| \end{bmatrix} \right) \quad (978)$$

$$\triangleq \sqrt{\delta^2(G)} \Psi \sqrt{\delta^2(G)}$$

where $\psi_{ij}$ (997) is angle between vectors $x_i$ and $x_j$, and where $\delta^2$ denotes a diagonal matrix in this case. Positive semidefiniteness of interpoint angle matrix $\Psi$ implies positive semidefiniteness of Gram matrix $G$;

$$G \succeq 0 \iff \Psi \succeq 0 \quad (979)$$

When $\delta^2(G)$ is nonsingular, then $G \succeq 0 \iff \Psi \succeq 0$. ($\S A.3.1.0.5$)

Distance-square $d_{ij}$ (965) is related to Gram matrix entries $G^T = G \triangleq [g_{ij}]$

$$d_{ij} = g_{ii} + g_{jj} - 2g_{ij} = \langle \Phi_{ij} , G \rangle \quad (980)$$

where $\Phi_{ij}$ is defined in (967). Hence the linear EDM definition

$$D(G) \triangleq \delta(G)1^T + \mathbf{1} \delta(G)^T - 2G \in \text{EDM}^N \quad \iff \quad G \succeq 0 \quad (981)$$
CHAPTER 5. EUCLIDEAN DISTANCE MATRIX

The EDM cone may be described, \(\text{(confer (1070)(1076))}\)

\[
\text{EDM}^N = \left\{ D(G) \mid G \in S_+^N \right\}
\]  

(982)

5.4.2.1 First point at origin

Assume the first point \(x_1\) in an unknown list \(X\) resides at the origin;

\[Xe_1 = 0 \iff Ge_1 = 0\]  

(983)

Consider the symmetric translation \((I - 1e_1^T)D(G)(I - e_1^T)\) that shifts the first row and column of \(D(G)\) to the origin; setting Gram-form EDM operator \(D(G) = D\) for convenience,

\[-(D - (De_1^T + e_1^TD)e_1^T)\frac{1}{2} = G - (Ge_1^T + e_1^TG) + e_1^TG e_1^T\]  

(984)

where

\[e_1 \triangleq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\]  

(985)

is the first vector from the standard basis. Then it follows for \(D \in S_+^N\)

\[G = -(D - (De_1^T + e_1^TD))\frac{1}{2}, \quad x_1 = 0\]

\[= -\begin{bmatrix} 0 & \sqrt{2}V_N^T \\ \sqrt{2}V_N & 0 \end{bmatrix}D \begin{bmatrix} 0 & \sqrt{2}V_N \\ \sqrt{2}V_N & 0 \end{bmatrix}\frac{1}{2}\]

\[= \begin{bmatrix} 0 & 0^T \\ 0 & -V_N^TDV_N \end{bmatrix}\]

(986)

\[V_N^TGV_N = -V_N^TDV_N\frac{1}{2} \quad \forall X\]

\[I - e_1^T = \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}\]  

(987)

is a projector nonorthogonally projecting \((\S E.1)\) on subspace

\[S_1^N = \left\{ G \in S^N \mid Ge_1 = 0 \right\}\]

\[= \left\{ [0 \ \sqrt{2}V_N]^T Y [0 \ \sqrt{2}V_N] \mid Y \in S^N \right\}\]  

(209)

in the Euclidean sense. From (986) we get sufficiency of the first matrix criterion for an EDM proved by Schoenberg in 1935; \([323]^{5.7}\)

\[D \in \text{EDM}^N \iff \left\{ \begin{array}{l} -V_N^TDV_N \in S_+^{N-1} \\ D \in S_+^N \end{array} \right\}\]  

(988)

\[^{5.7}\text{From (976) we know } \mathcal{R}(V_N) = \mathcal{N}(1^T), \text{ so (988) is the same as (964). In fact, any matrix } V \text{ in place of } V_N \text{ will satisfy (988) whenever } \mathcal{R}(V) = \mathcal{R}(V_N) = \mathcal{N}(1^T). \text{ But } V_N \text{ is the matrix implicit in Schoenberg’s seminal exposition.}\]
5.4. EDM DEFINITION

We provide a rigorous complete more geometric proof of this fundamental Schoenberg criterion in §5.9.1.0.4. [401, isedm(D)]

By substituting \( G = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_N^T D V_N \end{bmatrix} \) (986) into \( D(G) \) (981),

\[
D = \begin{bmatrix} 0 \\ \delta \mathbf{0}^T \end{bmatrix} \mathbf{1}^T + \begin{bmatrix} 0 & \delta \mathbf{0}^T \\ \mathbf{0} & -V_N^T D V_N \end{bmatrix} - 2 \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_N^T D V_N \end{bmatrix}
\]

(1090)

assuming \( x_1 = \mathbf{0} \). Details of this bijection are provided in §5.6.2.

5.4.2.2 0 geometric center

Assume the geometric center (§5.5.1.0.1) of an unknown list \( X \) is the origin;

\[
X_1 = \mathbf{0} \iff G_1 = \mathbf{0}
\]

(989)

Now consider the calculation \( (I - \frac{1}{N} \mathbf{1}^T \mathbf{1}) D(G)(I - \frac{1}{N} \mathbf{1}^T \mathbf{1}) \), a geometric centering or projection operation. (§E.7.2.0.2) Setting \( D(G) = D \) for convenience as in §5.4.2.1,

\[
G = -\frac{1}{N}(D - \frac{1}{N} \mathbf{1}^T \mathbf{1} D \mathbf{1}^T) + \frac{1}{N} \mathbf{1}^T \mathbf{1} D \mathbf{1}^T \frac{1}{2}, \quad X_1 = \mathbf{0}
\]

(990)

\[
V G V = -V D V \frac{1}{2} \quad \forall X
\]

where more properties of the auxiliary (geometric centering, projection) matrix

\[
V \triangleq I - \frac{1}{N} \mathbf{1}^T \mathbf{1} \in \mathbb{S}_+^N
\]

(991)

are found in §B.4. \( V G V \) may be regarded as a covariance matrix of means \( \mathbf{0} \). From (990) and the assumption \( D \in \mathbb{S}_+^N \) we get sufficiency of the more popular form of Schoenberg criterion:

\[
D \in \text{EDM}^N \iff \begin{cases} -V D V \in \mathbb{S}_+^N \\ D \in \mathbb{S}_+^N \end{cases}
\]

(992)

Of particular utility when \( D \in \text{EDM}^N \) is the fact, (§B.4.2 no.20) (965)

\[
\text{tr}(-V D V \frac{1}{2}) = \frac{1}{N} \sum_{i,j} d_{ij} = \frac{1}{N} \text{vec}(X)^T \left( \sum_{i,j} \Phi_{ij} \otimes I \right) \text{vec} X
\]

\[
= \text{tr}(V G V), \quad G \succeq 0
\]

(993)

\[
= \text{tr} G = \frac{1}{N} \sum_{t=1}^N ||x_t||^2 = ||X||^2_F, \quad X_1 = \mathbf{0}
\]

where \( \sum_{i,j} \Phi_{ij} \in \mathbb{S}_+^N \) (967), therefore convex in vec \( X \). We will find this trace useful as a heuristic to minimize affine dimension of an unknown list arranged columnar in \( X \) (§7.2.2), but it tends to facilitate reconstruction of a list configuration having least energy; id est, it compacts a reconstructed list by minimizing total norm-square of the vertices.

By substituting \( G = -V D V \frac{1}{2} \) (990) into \( D(G) \) (981), assuming \( X_1 = \mathbf{0} \)

\[
D = \delta(-V D V \frac{1}{2}) \mathbf{1}^T + \mathbf{1} \delta(-V D V \frac{1}{2})^T - 2(-V D V \frac{1}{2})
\]

(1080)

Details of this bijection can be found in §5.6.1.1.
5.4.2.2.1 Example. Hypersphere. These foregoing relationships allow combination of distance and Gram constraints in any optimization problem we might pose:

- Interpoint angle $\Psi$ can be constrained by fixing all the individual point lengths $\sqrt{\delta(G)}$; then
  \[ \Psi = -\sqrt{\delta(G)^{-1}}VDV_\frac{1}{2}\sqrt{\delta(G)^{-1}}, \quad X1 = 0 \]  
  (994)

- (conf err §5.9.1.0.3, (1179) (1323)) Constraining all main diagonal entries $g_{ii}$ of a Gram matrix to 1, for example, is equivalent to the constraint that all points lie on a hypersphere of radius 1 centered at the origin.

\[ D = 2(g_{11}11^T - G) \in \mathbb{EDM}^N \]  
(995)

Requiring 0 geometric center then becomes equivalent to the constraint: $D1 = 2N1$.

Because any point list may be constrained to lie on a hypersphere boundary whose affine dimension exceeds that of the list, a Gram matrix may always be constrained to have equal positive values along its main diagonal. (Laura Kranfer 1933 [323, §3]) This observation renewed interest in the elliptope (§5.9.1.0.1).

5.4.2.2.2 Example. List-member constraints via Gram matrix. Capitalizing on identity (990) relating Gram and EDM $D$ matrices, a constraint set such as

\[
\begin{align*}
\text{tr}(VDV(e_i e_j^T) &= ||x_i||^2 \\
\text{tr}(VDV) &= x_i^T x_j \\
\text{tr}(VDV) &= ||x_j||^2
\end{align*}
\]
(996)
relates list member $x_i$ to $x_j$ to within an isometry through inner-product identity [412, §1-7]

\[ \cos \psi_{ij} = \frac{x_i^T x_j}{||x_i|| ||x_j||} \]  
(977)
where $\psi_{ij}$ is angle between the two vectors as in (978). For $M$ list members, there total $M(M+1)/2$ such constraints. Angle constraints are incorporated in Example 5.4.2.2.5 and Example 5.4.2.2.13.

5.4.2.2.3 Example. Gram matrix as optimization problem. Consider the academic problem of finding a Gram matrix (990) subject to constraints on each and every entry of the corresponding EDM:

\[
\begin{align*}
\text{find } D &\in S_N^N \\
\text{subject to } \langle D, (e_i e_j^T + e_j e_i^T) \frac{1}{2} \rangle &= \delta_{ij}, \quad i, j = 1 \ldots N, \quad i \neq j \\
VDV &\succeq 0
\end{align*}
\]  
(998)
5.4. **EDM DEFINITION**

Figure 130: Rendering of *Fermat point* in acrylic on canvas by Suman Vaze. Three circles intersect at Fermat point of minimum total distance from three vertices of (and interior to) red/black/white triangle.

where the $d_{ij}$ are given nonnegative constants. EDM $D$ can, of course, be replaced with the equivalent Gram-form (981). Requiring only the selfadjointness property (1500) of the main-diagonal linear operator $\delta$ we get, for $A \in S^N$

$$\langle D, A \rangle = \langle \delta(G)1^T + 1 \delta(G)^T - 2G, A \rangle = 2 \langle G, \delta(A1) - A \rangle$$

(999)

Then the problem equivalent to (998) becomes, for $G \in S^N \Leftrightarrow G1 = 0$

$$\begin{align*}
\text{find} & \quad G \in S^N \\
\text{subject to} & \quad \langle G, \delta((e_i e_j^T + e_j e_i^T)1) - (e_i e_j^T + e_j e_i^T) \rangle = d_{ij}, \quad i, j = 1 \ldots N, \quad i < j \\
& \quad G \succeq 0
\end{align*}$$

(1000)

Barvinok’s Proposition 2.9.3.0.1 predicts existence for either formulation (998) or (1000) such that implicit equality constraints induced by subspace membership are ignored

$$\text{rank} G , \text{rank} VDV \leq \left\lfloor \frac{\sqrt{8(N(N-1)/2) + 1} - 1}{2} \right\rfloor = N - 1$$

(1001)

because, in each case, the Gram matrix is confined to a face of positive semidefinite cone $S^N$ isomorphic with $S^{N-1}_+$ (§6.6.1). (§E.7.2.0.2) This bound is tight (§5.7.1.1) and is the greatest upper bound.\footnote{§5.8 - $VDV|_{N-1} = 0$ (§B.4.1)}
5.4.2.2.4 Example. First duality. Kuhn reports that the first dual optimization problem\(^5\) to be recorded in the literature dates back to 1755.\(^389\) Perhaps more intriguing is the fact: this earliest instance of duality is a two-dimensional Euclidean distance geometry problem known as Fermat point (Figure 130) named after the French mathematician. Given \(N\) distinct points in the plane \(\{x_i \in \mathbb{R}^2, i = 1 \ldots N\}\), the Fermat point \(y\) is an optimal solution to

\[
\min_{y} \sum_{i=1}^{N} \|y - x_i\| \tag{1002}
\]

a convex minimization of total distance. The historically first dual problem formulation asks for the smallest equilateral triangle encompassing \((N = 3)\) three points \(x_i\). Another problem dual to (1002) (Kuhn 1967)

\[
\begin{align*}
\max_{\{z_i\}} \quad & \sum_{i=1}^{N} \langle z_i , x_i \rangle \\
\text{subject to} \quad & \sum_{i=1}^{N} z_i = 0 \\
& \|z_i\| \leq 1 \quad \forall i
\end{align*} \tag{1003}
\]

has interpretation as minimization of work required to balance potential energy in an \(N\)-way tug-of-war between equally matched opponents situated at \(\{x_i\}\).\(^406\)

It is not so straightforward to write the Fermat point problem (1002) equivalently in terms of a Gram matrix from this section. Squaring instead

\[
\min_{\alpha} \sum_{i=1}^{N} \|\alpha - x_i\|^2 = \min_{D \in \mathbb{S}^{N+1}} \langle -V , D \rangle \\
\text{subject to} \quad \langle D , e_i e_j^T + e_j e_i^T \rangle \frac{1}{2} = d_{ij} \quad \forall (i,j) \in \mathcal{I} \tag{1004}
\]

yields an inequivalent convex geometric centering problem whose equality constraints comprise EDM \(D\) main-diagonal zeros and known distances-square.\(^5\)\(^10\) Going the other way, a problem dual to total distance-square maximization (Example 6.7.0.0.1) is a penultimate minimum eigenvalue problem having application to PageRank calculation by search engines [241, §4]. [350]

Fermat function (1002) is empirically compared with (1004) in [63, §8.7.3], but for multiple unknowns in \(\mathbb{R}^2\), where propensity of (1002) for producing zero distance between unknowns is revealed. An optimal solution to (1002) gravitates toward gradient discontinuities (§D.2.1), as in Figure 74, whereas optimal solution to (1004) is less compact in the unknowns.\(^5\)\(^11\)

---

\(^5\)\(^9\) By dual problem is meant, in the strongest sense: the optimal objective achieved by a maximization problem, dual to a given minimization problem (related to each other by a Lagrangian function), is always equal to the optimal objective achieved by the minimization. (Figure 60 Example 2.13.1.0.3) A dual problem is always convex.

\(^5\)\(^10\) \(\alpha^*\) is geometric center of points \(x_i\) (1054). For three points, \(\mathcal{I} = \{1, 2, 3\}\); optimal affine dimension (§5.7) must be 2 because a third dimension can only increase total distance. Minimization of \(-V, D\) is a heuristic for rank minimization. (§7.2.2)

\(^5\)\(^11\) Optimal solution to (1002) has mechanical interpretation in terms of interconnecting springs with constant force when distance is nonzero; otherwise, 0 force. Problem (1004) is interpreted instead using linear springs.
5.4. EDM DEFINITION

369

Figure 131: Arbitrary hexagon in $\mathbb{R}^3$ whose vertices are labelled clockwise.

5.4.2.2.5 Example. Hexagon.

Barvinok \[26, \S 2.6\] poses a problem in geometric realizability of an arbitrary hexagon (Figure 131) having:

1. prescribed (one-dimensional) face-lengths $l$
2. prescribed angles $\phi$ between the three pairs of opposing faces
3. a constraint on the sum of norm-square of each and every vertex $x$

and where the first constraint on length-square $l_{ij}^2$ can be equivalently written as a constraint on the Gram matrix $-VDV\frac{1}{2}$ via (999). We show how to numerically solve such a problem by alternating projection in \S E.10.2.1.1. Barvinok’s Proposition 2.9.3.0.1 asserts existence of a list, corresponding to Gram matrix $G$ solving this feasibility problem, whose affine dimension (§5.7.1.1) does not exceed 3 because the convex feasible set is bounded by the third constraint $\text{tr}(-\frac{1}{2}VDV) = 1$ (993).
5.4.2.2.6 Example. Kissing number of sphere packing.

Two nonoverlapping Euclidean balls are said to kiss if they touch. An elementary geometrical problem can be posed: Given hyperspheres, each having the same diameter 1, how many hyperspheres can simultaneously kiss one central hypersphere? [426] Noncentral hyperspheres are allowed, but not required, to kiss.

As posed, the problem seeks the maximal number of spheres $K$ kissing a central sphere in a particular dimension. The total number of spheres is $N = K + 1$. In one dimension, the answer to this kissing problem is 2. In two dimensions, 6. (Figure 8)

The question was presented, in three dimensions, to Isaac Newton by David Gregory in the context of celestial mechanics. And so was born a controversy between the two scholars on the campus of Trinity College Cambridge in 1694. Newton correctly identified the kissing number as 12 (Figure 132) while Gregory argued for 13. Their dispute was finally resolved in 1953 by Schütte & van der Waerden. [309] In 2003, Oleg Musin tightened the upper bound on kissing number $K$ in four dimensions from 25 to $K = 24$ by refining a method of Philippe Delsarte from 1973. Delsarte’s method provides an infinite number [17] of linear inequalities necessary for converting a rank-constrained semidefinite program $^{5,12}$ to a linear program $^{5,13}$ [283].

There are no proofs known for kissing number in higher dimension excepting dimensions eight and twenty four. Interest persists [87] because sphere packing has found application to error correcting codes from the fields of communications and information theory; specifically to quantum computing. [95]

Translating this problem to an EDM graph realization (Figure 128, Figure 133) is suggested by Pfender & Ziegler. Imagine the centers of each sphere are connected by line segments. Then the distance between centers must obey simple criteria: Each sphere touching the central sphere has a line segment of length exactly 1 joining its center to

\[ \text{Figure 132: Sphere-packing illustration from [405, kissing number]. Translucent balls illustrated all have the same diameter.} \]
the central sphere’s center. All spheres, excepting the central sphere, must have centers separated by a distance of at least 1.

From this perspective, the kissing problem can be posed as a semidefinite program. Assign index 1 to the central sphere assuming a total of \( N \) spheres:

\[
\begin{align*}
\text{minimize} & \quad - \text{tr}(W V_N^T D V_N) \\
\text{subject to} & \quad D_{1j} = 1, \quad j = 2 \ldots N \\
& \quad D_{ij} \geq 1, \quad 2 \leq i < j = 3 \ldots N \\
& \quad D \in \text{EDM}^N
\end{align*}
\]

Then kissing number

\[
K = N_{\text{max}} - 1
\]

is found from the maximal number \( N \) of spheres that solve this semidefinite program in a given affine dimension \( r \) whose realization is assured by 0 optimal objective. Matrix \( W \) is constant, in this program, determined by a method disclosed in §4.4.1. Matrix \( W \in \mathbb{S}_+^N \) can be interpreted as direction of search through the positive semidefinite cone for a rank-\( r \) optimal solution

\[
-V_N^T P_{\mathcal{V}} V_N \in \mathbb{S}_+^{N-1}.
\]

In one dimension, optimal direction matrix \( W^* \) has rank \( K - r = 2 - 1 = 1 \):

\[
W^* = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]

(1009)

In two dimensions, optimal \( W^* \) has rank \( K - r = 6 - 2 = 4 \):

\[
W^* = \begin{bmatrix}
4 & 1 & 2 & -1 & -1 & 1 \\
1 & 4 & -1 & -1 & 2 & 1 \\
2 & -1 & 4 & 1 & 1 & -1 \\
-1 & -1 & 1 & 4 & 1 & 2 \\
-1 & 2 & 1 & 1 & 4 & -1 \\
1 & 1 & -1 & 2 & -1 & 4
\end{bmatrix}
\]

(1010)

In three dimensions, we leave it an exercise to find a rational optimal direction matrix \( W^* \) having rank \( K - r = 12 - 3 = 9 \). Here is a full-rank rational optimal direction matrix:

\[
W^* = \begin{bmatrix}
9 & 1 & -2 & -1 & 3 & -1 & -1 & 1 & 2 & 1 & -2 & 1 \\
1 & 9 & 3 & -1 & -1 & 1 & 1 & -2 & 1 & 2 & -1 & -1 \\
-2 & 3 & 9 & 1 & 2 & -1 & -1 & 2 & -1 & -1 & 1 & 2 \\
-1 & -1 & 1 & 9 & 1 & -1 & 1 & -1 & 3 & 2 & -1 & 1 \\
3 & -1 & 2 & 1 & 9 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\
-1 & 1 & -1 & -1 & 1 & 9 & 2 & -1 & 2 & -1 & 2 & 3 \\
-1 & 1 & -1 & 1 & 1 & 2 & 9 & 3 & -1 & 1 & -2 & -1 \\
1 & -2 & 2 & -1 & -1 & -1 & 3 & 9 & 2 & -1 & 1 & 1 \\
2 & 1 & -1 & 3 & -1 & 2 & -1 & 2 & 9 & -1 & 1 & -1 \\
1 & 2 & -1 & 2 & -1 & -1 & 1 & -1 & -1 & 9 & 3 & 1 \\
-2 & -1 & 1 & -1 & 1 & 2 & -2 & 1 & 1 & 3 & 9 & -1 \\
1 & -1 & 2 & 1 & -1 & 3 & -1 & 1 & -1 & 1 & -1 & 9
\end{bmatrix} \frac{1}{12}
\]

(1011)

A four-dimensional solution also has rational optimal direction matrix \( W^* \) with rank \( K - r = 24 - 4 = 20 \):
but these direction matrices are not unique and their precision not critical. Here is an optimal four-dimensional point list, $^{5.14}$ in MATLAB output format, reconstructed by a method in §5.12:

Columns 1 through 6
$$X = \begin{bmatrix} 0 & -0.1983 & -0.4584 & 0.1657 & 0.9399 & 0.7416 \\ 0 & 0.6863 & 0.2936 & 0.6239 & -0.2936 & 0.3927 \\ 0 & -0.4835 & 0.8146 & -0.6448 & -0.0611 & -0.4224 \\ 0 & 0.5059 & 0.2004 & -0.4093 & -0.1632 & 0.3427 \end{bmatrix}$$

Columns 7 through 12
$$\begin{bmatrix} -0.4815 & -0.9399 & -0.7416 & 0.1983 & 0.4584 & -0.2832 \\ 0 & 0.2936 & -0.3927 & -0.6863 & -0.2936 & -0.6863 \\ -0.8756 & -0.0611 & 0.4224 & 0.4835 & -0.8146 & -0.3922 \\ -0.0372 & 0.1632 & -0.3427 & -0.5059 & -0.2004 & -0.5431 \end{bmatrix}$$

Columns 13 through 18
$$\begin{bmatrix} 0.2832 & -0.2926 & -0.6473 & 0.0943 & 0.3640 & -0.3640 \\ 0.6863 & 0.9176 & -0.6239 & -0.2313 & -0.0624 & 0.0624 \\ 0.3922 & 0.1698 & -0.2309 & -0.6533 & -0.1613 & 0.1613 \\ 0.5431 & -0.2088 & 0.3721 & 0.7147 & -0.9152 & 0.9152 \end{bmatrix}$$

Columns 19 through 25
$$\begin{bmatrix} -0.0943 & 0.6473 & -0.1657 & 0.2926 & -0.5759 & 0.5759 & 0.4815 \\ 0.2313 & 0.6239 & -0.6239 & -0.9176 & 0.2313 & -0.2313 & 0 \\ 0.6533 & 0.2309 & 0.6448 & -0.1698 & -0.2224 & 0.2224 & 0.8756 \\ -0.7147 & -0.3721 & 0.4093 & 0.2088 & -0.7520 & 0.7520 & 0.0372 \end{bmatrix}$$

$^{5.14}$ An optimal five-dimensional point list is known: The answer was known at least 175 years ago. I believe Gauss knew it. Moreover, Korkine & Zolotarev proved in 1882 that $D_5$ is the densest lattice in five dimensions. So they proved that if a kissing arrangement in five dimensions can be extended to some lattice, then $k(5) = 40$. Of course, the conjecture in the general case also is: $k(5) = 40$. You would like to see coordinates? Easily. Let $A = \sqrt{2}$. Then $p(1) = (A, A, 0, 0, 0)$, $p(2) = (-A, A, 0, 0, 0)$, $p(3) = (A, -A, 0, 0, 0)$, $p(4) = (0, 0, 0, -A, -A)$; i.e., we are considering points with coordinates that have two A and three 0 with any choice of signs and any ordering of the coordinates; the same coordinates-expression in dimensions 3 and 4.

The first miracle happens in dimension 6. There are better packings than $D_6$ (Conjecture: $k(6) = 72$).

It's a real miracle how dense the packing is in eight dimensions ($E_8 = Korkine & Zolotarev$ packing that was discovered in 1880s) and especially in dimension 24, that is the so-called Leech lattice.

Actually, people in coding theory have conjectures on the kissing numbers for dimensions up to 32 (or even greater?). However, sometimes they found better lower bounds. I know that Ericson & Zinoviev a few years ago discovered (by hand, no computer) in dimensions 13 and 14 better kissing arrangements than were known before. – Oleg Musin
The $r$ nonzero optimal eigenvalues of $-V_N^T D^* V_N$ are equal; remaining eigenvalues are zero as per $-\text{tr}(W^* V_N^T D^* V_N) = 0$ (768). Numerical problems begin to arise with matrices of this size due to interior-point methods of solution to (1007). By eliminating some equality constraints from the kissing number problem, matrix size can be reduced: From §5.8.3 we have

$$-V_N^T D V_N = 11^T - [0 \ I] D \begin{bmatrix} 0^T \\ I \end{bmatrix} \frac{1}{2}$$  \hspace{1cm} (1013)

(which does not hold more generally) where Identity matrix $I \in \mathbb{S}^{N-1}$ has one less dimension than EDM $D$. By defining an EDM principal submatrix

$$\hat{D} \triangleq [0 \ I] D \begin{bmatrix} 0^T \\ I \end{bmatrix} \in \mathbb{S}^{N-1}_h$$  \hspace{1cm} (1014)

so that

$$-V_N^T D V_N = 11^T - \hat{D}^{\frac{1}{2}}$$  \hspace{1cm} (1015)

we get a convex problem equivalent to (1007)

$$\begin{align*}
\text{minimize} & \quad -\text{tr}(W \hat{D}) \\
\text{subject to} & \quad \hat{D}_{ij} \geq 1, \quad 1 \leq i < j \leq 2 \ldots K \\
& \quad 11^T - \hat{D}^{\frac{1}{2}} \succeq 0 \\
& \quad \delta(\hat{D}) = 0
\end{align*}$$  \hspace{1cm} (1016)

Any feasible solution $11^T - \hat{D}^{\frac{1}{2}}$ belongs to an elliptope (§5.9.1.0.1).

5.4.2.2.7 Exercise. Rational optimal kissing direction matrix $W^*$. Replace (1011) with a rational $W^*$ having rank $= K - r = 12 - 3 = 9$, main diagonal 9, and common denominator 12.

This next example shows how finding the common point of intersection for three circles in a plane, a nonlinear problem, has convex expression.

5.4.2.2.8 Example. Trilateration in wireless sensor network. [171] Given three known absolute point positions in $\mathbb{R}^2$ (three anchors $\hat{x}_2, \hat{x}_3, \hat{x}_4$) and only one unknown point (one sensor $x_1$), the sensor’s absolute position is determined from its noiseless measured distance-square $\hat{d}_{1i}$ to each of three anchors (Figure 3, Figure 133a). This trilateration can be expressed as a convex optimization problem in terms of list $X \triangleq [x_1 \ x_2 \ x_3 \ x_4] \in \mathbb{R}^{2\times 4}$ and Gram matrix $G \in \mathbb{S}^3$ (978):
CHAPTER 5. EUCLIDEAN DISTANCE MATRIX

Figure 133: (a) Given three distances indicated with absolute point positions $\bar{x}_2, \bar{x}_3, \bar{x}_4$ known and noncollinear, absolute position of $x_1$ in $\mathbb{R}^2$ can be precisely and uniquely determined by trilateration; solution to a system of nonlinear equations. Dimensionless EDM graphs (b) (c) (d) represent EDMs in various states of completion. Line segments represent known absolute distances and may cross without vertex at intersection. (b) Four-point list must always be embeddable in affine subset having dimension rank $V_N^T D V_N$ not exceeding 3. To determine relative position of $x_2, x_3, x_4$, triangle inequality is necessary and sufficient (§5.14.1). Knowing all distance information, then (by injectivity of $D$ (§5.6)) point position $x_1$ is uniquely determined to within an isometry in any dimension. (c) When fifth point is introduced, only distances to $x_3, x_4, x_5$ are required to determine relative position of $x_2$ in $\mathbb{R}^2$. Graph represents first instance of missing distance information; $\sqrt{d_{12}}$. (d) Three distances are absent ($\sqrt{d_{12}}, \sqrt{d_{13}}, \sqrt{d_{23}}$) from complete set of interpoint distances, yet unique isometric reconstruction (§5.4.2.2.10) of six points in $\mathbb{R}^2$ is certain.
5.4. EDM DEFINITION

\[
\begin{align*}
\text{minimize} & \quad \text{tr } G \\
\text{subject to} & \quad \text{tr}(G\Phi_{ij}) = \bar{d}_{i1}, \quad i = 2, 3, 4 \\
& \quad \text{tr}(Ge_i e_i^T) = \|\tilde{x}_i\|^2, \quad i = 2, 3, 4 \\
& \quad \text{tr}(Ge_i e_j^T + e_j e_i^T)/2 = \bar{x}_j^T \tilde{x}_j, \quad 2 \leq i < j = 3, 4 \\
X(:, 2:4) & = [\tilde{x}_2 \tilde{x}_3 \tilde{x}_4] \\
\end{align*}
\]

where

\[
\Phi_{ij} = (e_i - e_j)(e_i - e_j)^T \in S_+^N
\]

and where the constraint on distance-square \( \bar{d}_{i1} \) is equivalently written as a constraint on the Gram matrix via (980). There are 9 linearly independent affine equality constraints on that Gram matrix while the sensor is constrained, only by dimensioning, to lie in \( \mathbb{R}^2 \).

Although the objective \( \text{tr } G \) of minimization\(^{5,15} \) insures a solution on the boundary of positive semidefinite cone \( S_+^4 \), for this problem, we claim that the set of feasible Gram matrices forms a line (§2.5.1.1) in isomorphic \( \mathbb{R}^{10} \) tangent (§2.1.7.1.2) to the positive semidefinite cone boundary. (§5.4.2.2.9, confer §4.2.1.3)

By Schur complement (§A.4, §2.9.1.0.1), any feasible \( G \) and \( X \) provide

\[
G \succeq X^T X
\]

which is a convex relaxation of the desired (nonconvex) equality constraint

\[
\begin{bmatrix}
I & X \\
X^T & G
\end{bmatrix} \succeq \begin{bmatrix}
I & X
\end{bmatrix}
\]

expected positive semidefinite rank-2 under noiseless conditions. But, by (1573), the relaxation admits

\[
(3 \geq \text{rank } G \geq \text{rank } X)
\]

(a third dimension corresponding to an intersection of three spheres, not circles, were there noise). If rank of an optimal solution equals 2,

\[
\text{rank} \begin{bmatrix}
I & X^* \\
X^T & G^*
\end{bmatrix} = 2
\]

then \( G^* = X^* X^* \) by Theorem A.4.0.1.3.

As posed, this localization problem does not require affinely independent (Figure 29, three noncollinear) anchors. Assuming the anchors exhibit no rotational or reflective symmetry in their affine hull (§5.5.2) and assuming the sensor \( x_1 \) lies in that affine hull, then sensor position solution \( x_1^* = X^*(; , 1) \) is unique under noiseless measurement. [331]

\(^{5,15}\)Trace \( \text{tr } G = (I, G) \) minimization is a heuristic for rank minimization. (§7.2.2.1) It may be interpreted as squashing \( G \) which is bounded below by \( X^T X \) as in (1018); \( \text{id est, } G - X^T X \succeq 0 \Rightarrow \text{tr } G \succeq \text{tr } X^T X \) (1571). \( \delta(G - X^T X) = 0 \Leftrightarrow G = X^T X \) (§A.7.2) \( \Rightarrow \text{tr } G = \text{tr } X^T X \) which is a condition necessary for equality.
This preceding transformation of trilateration to a semidefinite program works all the time (\((1021\) holds) despite relaxation (1018) because the optimal solution set is a unique point.

5.4.2.2.9 Proof (sketch). Only the sensor location \(x_1\) is unknown. The objective function together with the equality constraints make a linear system of equations in Gram matrix variable \(G\):

\[
\begin{align*}
\text{tr } G &= \|x_1\|^2 + \|\tilde{x}_2\|^2 + \|\tilde{x}_3\|^2 + \|\tilde{x}_4\|^2 \\
\text{tr}(G \Phi_{i1}) &= \tilde{d}_{i1}, \quad i = 2, 3, 4 \\
\text{tr}(Ge_i e_i^T) &= \|\tilde{x}_i\|^2, \quad i = 2, 3, 4 \\
\text{tr}((e_i e_j^T + e_j e_i^T)/2) &= \tilde{x}_i^T \tilde{x}_j, \quad 2 \leq i < j = 3, 4
\end{align*}
\]

which is invertible:

\[
\text{svec } G = \begin{bmatrix}
\text{svec}(I) \; & \text{svec}(\Phi_{21}) \; & \text{svec}(\Phi_{31}) \; & \text{svec}(\Phi_{41}) \\
\text{svec}(e_2 e_2^T) \; & \text{svec}(e_3 e_3^T) \; & \text{svec}(e_4 e_4^T) \\
\text{svec}((e_2 e_3^T + e_3 e_2^T)/2) \; & \text{svec}((e_2 e_4^T + e_4 e_2^T)/2) \; & \text{svec}((e_3 e_4^T + e_4 e_3^T)/2)
\end{bmatrix}^{-1}
\begin{bmatrix}
\|x_1\|^2 + \|\tilde{x}_2\|^2 + \|\tilde{x}_3\|^2 + \|\tilde{x}_4\|^2 \\
\tilde{d}_{21} \\
\tilde{d}_{31} \\
\tilde{d}_{41} \\
\|\tilde{x}_2\|^2 \\
\|\tilde{x}_3\|^2 \\
\|\tilde{x}_4\|^2 \\
\tilde{x}_2^T \tilde{x}_3 \\
\tilde{x}_2^T \tilde{x}_4 \\
\tilde{x}_3^T \tilde{x}_4
\end{bmatrix}
\]

That line in the ambient space \(S^4\) of \(G\), claimed on page 375, is traced by \(\|x_1\|^2 \in \mathbb{R}\) on the right-hand side, as it turns out. One must show this line to be tangential (§2.1.7.1.2) to \(S^4\) in order to prove uniqueness. Tangency is possible for affine dimension 1 or 2 while its occurrence depends completely on the known measurement data. ■

But as soon as significant noise is introduced or whenever distance data is incomplete, such problems can remain convex although the set of all optimal solutions generally becomes a convex set bigger than a single point (and still containing the noiseless solution).

5.4.2.2.10 Definition. Isometric reconstruction. (confer §5.5.3) Isometric reconstruction from an EDM means building a list \(X\) correct to within a rotation, reflection, and translation; in other terms, reconstruction of relative position, correct to within an isometry, correct to within a rigid transformation. △

How much distance information is needed to uniquely localize a sensor (to recover actual relative position)? The narrative in Figure 133 helps dispel any notion of distance data proliferation in low affine dimension \((r < N - 2)\).\(^{5.16}\) Huang, Liang,

---

5.16When affine dimension \(r\) reaches \(N - 2\), then all distances-square in the EDM must be known for unique isometric reconstruction in \(\mathbb{R}^r\); going the other way, when \(r = 1\) then the condition that the dimensionless EDM graph be connected is necessary and sufficient. [197, §2.2]
and Pardalos [215, §4.2] claim \(O(2N)\) distances is a least lower bound (independent of affine dimension \(r\)) for unique isometric reconstruction; achievable under certain noiseless conditions on graph connectivity and point position. Alfakih shows how to ascertain uniqueness over all affine dimensions via \textit{Gale matrix}. [10] [5] [6] Figure 128b (p.358, from \textit{small completion problem} Example 5.3.0.0.2) is an example in \(\mathbb{R}^2\) requiring only \(2N - 3 = 5\) known symmetric entries for unique isometric reconstruction, although the four-point example in Figure 133b will not yield a unique reconstruction when any one of the distances is left unspecified.

The list represented by the particular dimensionless \textit{EDM graph} in Figure 134, having only \(2N - 3 = 9\) absolute distances specified, has only one realization in \(\mathbb{R}^2\) but has more realizations in higher dimensions. Unique \(r\)-dimensional isometric reconstruction by semidefinite relaxation like (1018) occurs iff realization in \(\mathbb{R}^r\) is unique and there exist no nontrivial higher-dimensional realizations. [331] For sake of reference, we provide the complete corresponding EDM:

\[
D = \begin{bmatrix}
0 & 50641 & 56129 & 8245 & 18457 & 26645 \\
50641 & 0 & 49300 & 25994 & 8810 & 20612 \\
56129 & 49300 & 0 & 24202 & 31330 & 9160 \\
8245 & 25994 & 24202 & 0 & 4680 & 5290 \\
18457 & 8810 & 31330 & 4680 & 0 & 6658 \\
26645 & 20612 & 9160 & 5290 & 6658 & 0 \\
\end{bmatrix}
\] (1024)

We consider paucity of distance information in this next example which shows it is possible to recover exact relative position given incomplete noiseless distance information. An \textit{ad hoc} method for recovery of the least-rank optimal solution under noiseless conditions is introduced:

### 5.4.2.2.11 Example. \textit{Tandem trilateration in wireless sensor network.}

Given three known absolute point-positions in \(\mathbb{R}^2\) (three anchors \(\hat{x}_3, \hat{x}_4, \hat{x}_5\)), two unknown sensors \(x_1, x_2 \in \mathbb{R}^2\) have absolute position determinable from their noiseless distances-square (as indicated in Figure 135) assuming the anchors exhibit no rotational or reflective symmetry in their affine hull (§5.5.2). This example differs from Example 5.4.2.2.8 insofar as trilateration of each sensor is now in terms of one unknown position: the other sensor. We express this localization as a convex optimization problem (a semidefinite program, §4.1) in terms of list \(X \triangleq [x_1, x_2, \hat{x}_3, \hat{x}_4, \hat{x}_5] \in \mathbb{R}^{2 \times 5}\) and Gram matrix \(G \in \mathbb{S}^5\) (978) via relaxation (1018):

\[
\begin{align*}
\text{minimize} & \quad \operatorname{tr} G \\
\text{subject to} & \quad \operatorname{tr}(G \Phi_{i1}) = \hat{d}_{i1}, \quad i = 2, 4, 5 \\
& \quad \operatorname{tr}(G \Phi_{i2}) = \hat{d}_{i2}, \quad i = 3, 5 \\
& \quad \operatorname{tr}(G e_i e_i^T) = \|\hat{x}_i\|^2, \quad i = 3, 4, 5 \\
& \quad \operatorname{tr}(G (e_i e_j^T + e_j e_i^T)/2) = \hat{x}_i^T \hat{x}_j, \quad 3 \leq i < j \leq 4, 5 \\
& \quad X(:, 3:5) = [\hat{x}_3 \ \hat{x}_4 \ \hat{x}_5] \\
& \quad \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} \succeq 0
\end{align*}
\] (1025)
CHAPTER 5. EUCLIDEAN DISTANCE MATRIX

Figure 134: (confer (1024)) Incomplete EDM corresponding to this dimensionless EDM graph (drawn freehand; no symmetry intended) provides unique isometric reconstruction in $\mathbb{R}^2$.

Figure 135: (Ye) Two sensors ● and three anchors ○ in $\mathbb{R}^2$. Connecting line-segments denote known absolute distances. Incomplete EDM corresponding to this dimensionless EDM graph provides unique isometric reconstruction in $\mathbb{R}^2$. 
Figure 136: Given in red ⊙ are two discrete linear trajectories of sensors $x_1$ and $x_2$ in $\mathbb{R}^2$ localized by algorithm (1025) as indicated by blue bullets •. Anchors $\hat{x}_3$, $\hat{x}_4$, $\hat{x}_5$, corresponding to Figure 135, are indicated by ⊗. When targets ⊙ and bullets • coincide under these noiseless conditions, localization is successful. On this run, two visible localization errors are due to rank-3 Gram optimal solutions. These errors can be corrected by choosing a different normal in objective of minimization.
where

\[ \Phi_{ij} = (e_i - e_j)(e_i - e_j)^T \in S^N_+ \] (967)

This problem realization is fragile because of the unknown distances between sensors and anchors. Yet there is no more information we may include beyond the 11 independent equality constraints on the Gram matrix (nonredundant constraints not antithetical) to reduce the feasible set.\(^5\)

Exhibited in Figure 136 are two mistakes in solution \( X^*(:, 1:2) \) due to a rank-3 optimal Gram matrix \( G^* \). The trace objective is a heuristic minimizing convex envelope of quasiconcave function \( \text{rank} \ G \). (§2.9.2.9.2, §7.2.2.1) A rank-2 optimal Gram matrix can be found and the errors corrected by choosing a different normal for the linear objective function, now implicitly the Identity matrix \( I \); \( \text{id est} \),

\[ \text{tr} G = \langle G, I \rangle \leftarrow \langle G, \delta(u) \rangle \] (1026)

where vector \( u \in \mathbb{R}^5 \) is randomly selected. A random search for a good normal \( \delta(u) \) in only a few iterations is quite easy and effective because: the problem is small, an optimal solution is known \( a \text{ priori} \) to exist in two dimensions, a good normal direction is not necessarily unique, and (we speculate) because the feasible affine-subset slices the positive semidefinite cone thinly in the Euclidean sense.\(^5\)

We explore ramifications of noise and incomplete data throughout; their individual effect being to expand the optimal solution set, introducing more solutions and higher-rank solutions. Hence our focus shifts in §4.4 to discovery of a reliable means for diminishing the optimal solution set by introduction of a rank constraint.

Now we illustrate how a problem in distance geometry can be solved without equality constraints representing measured distance; instead, we have only upper and lower bounds on distances measured:

5.4.2.2.12 Example. **Wireless location in a cellular telephone network.**

Utilizing measurements of distance, time of flight, angle of arrival, or signal power in the context of wireless telephony, *multilateration* is the process of localizing (determining absolute position of) a radio signal source \( \bullet \) by inferring geometry relative to multiple fixed *base stations* \( \circ \) whose locations are known.

We consider localization of a cellular telephone by distance geometry, so we assume distance to any particular base station can be inferred from received signal power. On a large open flat expanse of terrain, signal-power measurement corresponds well with inverse distance. But it is not uncommon for measurement of signal power to suffer

\(^5\)By virtue of their dimensioning, the sensors are already constrained to \( \mathbb{R}^2 \) the affine hull of the anchors.

\(^5\)Projection on that nonconvex subset of all \( N \times N \)-dimensional positive semidefinite matrices, in an affine subset, whose rank does not exceed 2 is a problem considered difficult to solve. [365, §4]

\(^5\)The log det rank-heuristic from §7.2.2.4 does not work here because it chooses the wrong normal. Rank reduction (§4.1.2.1) is unsuccessful here because Barvinok’s upper bound (§2.9.3.0.1) on rank of \( G^* \) is 4.
Figure 137: Regions of coverage by base stations ○ in a cellular telephone network. The term *cellular* arises from packing of regions best covered by neighboring base stations. Illustrated is a pentagonal cell best covered by base station $\hat{x}_2$. Like a Voronoi diagram, cell geometry depends on base-station arrangement. In some US urban environments, it is not unusual to find base stations spaced approximately 1 mile apart. There can be as many as 20 base-station antennae capable of receiving signal from any given cell phone ●; practically, that number is closer to 6.

Figure 138: Some fitted contours of equal signal power in $\mathbb{R}^2$ transmitted from a commercial cellular telephone ● over about 1 mile suburban terrain outside San Francisco in 2005. (Data by courtesy of Polaris Wireless.)
20 decibels in loss caused by factors such as *multipath* interference (signal reflections), mountainous terrain, man-made structures, turning one’s head, or rolling the windows up in an automobile. Consequently, contours of equal signal power are no longer circular; their geometry is irregular and would more aptly be approximated by translated ellipsoids of graduated orientation and eccentricity as in Figure 138.

Depicted in Figure 137 is one cell phone $x_1$ whose signal power is automatically and repeatedly measured by 6 base stations nearby. Those signal power measurements are transmitted from that cell phone to base station $\hat{x}_2$ who decides whether to transfer (hand-off or hand-over) responsibility for that call should the user roam outside its cell.

Due to noise, at least one distance measurement more than the minimum number of measurements is required for reliable localization in practice; 3 measurements are minimum in two dimensions, 4 in three. Existence of noise precludes measured distance from the input data. We instead assign measured distance to a range estimate specified by individual upper and lower bounds: $d_{i1}$ is the upper bound on distance-square from the cell phone to $i$th base station, while $d_{i1}$ is the lower bound. These bounds become the input data. Each measurement range is presumed different from the others.

Then convex problem (1017) takes the form:

$$\begin{align*}
\text{minimize} & \quad \text{tr} G \\
\text{subject to} & \quad d_{i1} \leq \text{tr}(G \Phi_{i1}) \leq d_{i1}, & i = 2 \ldots 7 \\
& \quad \text{tr}(G(e_i e_i^T)) = \|\hat{x}_i\|^2, & i = 2 \ldots 7 \\
& \quad \text{tr}(G(e_i e_j^T + e_j e_i^T)/2) = \hat{x}_i^T \hat{x}_j, & 2 \leq i < j = 3 \ldots 7 \\
& \quad \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} \succeq 0
\end{align*}$$

where

$$\Phi_{ij} = (e_i - e_j)(e_i - e_j)^T \in S_+^N$$

This semidefinite program realizes the wireless location problem illustrated in Figure 137. Location $X^*(\cdot, 1)$ is taken as solution, although measurement noise will often cause rank $G^*$ to exceed 2. Randomized search for a rank-2 optimal solution is not so easy here as in Example 5.4.2.2.11. We introduce a method in §4.4 for enforcing the stronger rank-constraint (1021). To formulate this same problem in three dimensions, point list $X$ is simply redimensioned in the semidefinite program.

---

5.20 Cell phone signal power is typically encoded logarithmically with 1-decibel increment and 64-decibel dynamic range.

5.21 Because distance to base station is quite difficult to infer from signal power measurements in an urban environment, localization of a particular cell phone by distance geometry would be far easier were the whole cellular system instead conceived so cell phone $x_1$ also transmits (to base station $\hat{x}_2$) its signal power as received by all other cell phones within range.

5.22 In Example 4.4.1.2.4, we explore how this convex optimization algorithm fares in the face of measurement noise.
5.4. EDM DEFINITION

5.4.2.2.13 Example. (Biswas, Nigam, Ye)  
*Molecular Conformation.*
The subatomic measurement technique called *nuclear magnetic resonance spectroscopy* (NMR) is employed to ascertain physical conformation of molecules; e.g., Figure 4, Figure 139. From this technique, distance, angle, and dihedral angle measurements can be obtained. Dihedral angles arise consequent to a phenomenon where atom subsets are physically constrained to Euclidean planes.

> *In the* rigid covalent geometry approximation, the bond lengths and angles are treated as completely fixed, so that a given spatial structure can be described very compactly indeed by a list of torsion angles alone... These are the dihedral angles between the planes spanned by the two consecutive triples in a chain of four covalently bonded atoms.*

—G. M. Crippen & T. F. Havel, 1988 [91, §1.1]

Crippen & Havel recommend working exclusively with distance data because they consider angle data to be mathematically cumbersome. The present example shows instead how inclusion of dihedral angle data into a problem statement can be made elegant and convex.

As before, ascribe position information to the matrix

$$X = [x_1 \cdots x_N] \in \mathbb{R}^{3 \times N} \quad (76)$$

and introduce a matrix \( \mathcal{N} \) holding normals \( \eta \) to planes respecting dihedral angles \( \varphi \):

$$\mathcal{N} \triangleq [\eta_1 \cdots \eta_M] \in \mathbb{R}^{3 \times M} \quad (1028)$$

As in the other examples, we preferentially work with Gram matrices \( G \) because of the bridge they provide between other variables; we define

Figure 139: A depiction of molecular conformation. [123]
whose rank is 3 by assumption. So our problem's variables are the two Gram matrices $G_X$ and $G_R$ and matrix $Z = \mathbb{R}^{T}X$ of cross products. Then measurements of distance-square $d$ can be expressed as linear constraints on $G_X$ as in (1027), dihedral angle $\varphi$ measurements can be expressed as linear constraints on $G_R$ by (997), and normal-vector $\eta$ conditions can be expressed by vanishing linear constraints on cross-product matrix $Z$: Consider three points $x$ labelled $1, 2, 3$ assumed to lie in the $\ell$th plane whose normal is $\eta_\ell$. There might occur, for example, the independent constraints

$$
\eta_\ell^T(x_1 - x_2) = 0 \\
\eta_\ell^T(x_2 - x_3) = 0
$$

which are expressible in terms of constant matrices $A_k \in \mathbb{R}^{M \times N}$;

$$
\langle Z, A_{12} \rangle = 0 \\
\langle Z, A_{23} \rangle = 0
$$

Although normals $\eta$ can be constrained exactly to unit length,

$$
\delta(G_R) = 1
$$

NMR data is noisy; so measurements are given as upper and lower bounds. Given bounds on dihedral angles respecting $0 \leq \varphi_j \leq \pi$ and bounds on distances $d_i$ and given constant matrices $A_k$ (1031) and symmetric matrices $\Phi_i$ (967) and $B_j$ per (997), then a molecular conformation problem can be expressed:

$$
\begin{align*}
\text{find} & & \quad G_X \\
\text{subject to} & & \quad \frac{d_i}{2} \leq \text{tr}(G_X \Phi_i) \leq \bar{d}_i & \forall i \in I_1 \\
& & \cos \varphi_j \leq \text{tr}(G_R B_j) \leq \cos \bar{\varphi}_j & \forall j \in I_2 \\
& & \langle Z, A_k \rangle = 0 & \forall k \in I_3 \\
& & \quad G_X1 = 0 \\
& & \quad \delta(G_R) = 1 \\
& & \quad \begin{bmatrix} G_R & Z \\ Z^T & G_X \end{bmatrix} \succeq 0 \\
& & \quad \text{rank} \begin{bmatrix} G_R & Z \\ Z^T & G_X \end{bmatrix} = 3
\end{align*}
$$

where $G_X1 = 0$ provides a geometrically centered list $X$ (§5.4.2.2). Ignoring the rank constraint would tend to force cross-product matrix $Z$ to zero. What binds these three variables is the rank constraint; we show how to satisfy it in §4.4. □
5.4. EDM DEFINITION

5.4.3 Inner-product form EDM definition

We might, for example, want to realize a constellation given only interstellar
distance (or, equivalently, parsecs from our Sun and relative angular
measurement; the Sun as vertex to two distant stars); called stellar
cartography...

Equivalent to (965) is [412, §1-7] [341, §3.2]

\[
d_{ij} = d_{ik} + d_{kj} - 2 \sqrt{d_{ik}d_{kj}} \cos \theta_{ikj}
\]

\[
= \left[ \sqrt{d_{ik}} \quad \sqrt{d_{kj}} \right] \left[ \begin{array}{cc}
1 & -e^{i\theta_{ikj}} \\
-e^{-i\theta_{ikj}} & 1
\end{array} \right] \left[ \begin{array}{c}
\sqrt{d_{ik}} \\
\sqrt{d_{kj}}
\end{array} \right]
\]

(1034)
called law of cosines where \( i \equiv \sqrt{-1}, \ i, j, k \) are positive integers, and \( \theta_{ikj} \) is the angle
at vertex \( x_k \) formed by vectors \( x_i - x_k \) and \( x_j - x_k \);

\[
cos \theta_{ikj} = \frac{1}{2}(d_{ik} + d_{kj} - d_{ij}) \quad = \frac{(x_i - x_k)^T(x_j - x_k)}{\|x_i - x_k\| \|x_j - x_k\|}
\]

(1035)
where the numerator forms an inner product of vectors. Distance-square \( d_{ij} \)
is a convex quadratic function\(^{5.23}\) on \( \mathbb{R}^2 \) whereas \( d_{ij}(\theta_{ikj}) \) is quasiconvex (§3.8)
minimized over domain \( \{-\pi \leq \theta_{ikj} \leq \pi \} \) by \( d^*_{ikj} = 0 \), we get the Pythagorean theorem when
\( \theta_{ikj} = \pm \pi/2 \), and \( d_{ij}(\theta_{ikj}) \) is maximized when \( \theta^*_{ikj} = \pm \pi \);

\[
\begin{align*}
d_{ij} &= (\sqrt{d_{ik}} + \sqrt{d_{kj}})^2, & \theta_{ikj} &= \pm \pi \\
d_{ij} &= d_{ik} + d_{kj}, & \theta_{ikj} &= \pm \frac{\pi}{2} \\
d_{ij} &= (\sqrt{d_{ik}} - \sqrt{d_{kj}})^2, & \theta_{ikj} &= 0
\end{align*}
\]

(1036)
so

\[
|\sqrt{d_{ik}} - \sqrt{d_{kj}}| \leq \sqrt{d_{ij}} \leq \sqrt{d_{ik}} + \sqrt{d_{kj}}
\]

(1037)
Hence the triangle inequality, Euclidean metric property 4, holds for any EDM \( D \).

We may construct an inner-product form of the EDM definition for matrices by
evaluating (1034) for \( k = 1 \): By defining

\[
\Theta^T \Theta \triangleq \begin{bmatrix}
  d_{12} & \sqrt{d_{12}d_{13}} \cos \theta_{123} & \sqrt{d_{12}d_{14}} \cos \theta_{124} & \cdots & \sqrt{d_{12}d_{1N}} \cos \theta_{12N} \\
  \sqrt{d_{12}d_{13}} \cos \theta_{213} & d_{13} & \sqrt{d_{13}d_{14}} \cos \theta_{214} & \cdots & \sqrt{d_{13}d_{1N}} \cos \theta_{21N} \\
  \sqrt{d_{12}d_{14}} \cos \theta_{124} & \sqrt{d_{13}d_{14}} \cos \theta_{214} & d_{14} & \cdots & \sqrt{d_{14}d_{1N}} \cos \theta_{21N} \\
  \vdots & \vdots & \vdots & \ddots & \ddots \\
  \sqrt{d_{12}d_{1N}} \cos \theta_{12N} & \sqrt{d_{13}d_{1N}} \cos \theta_{21N} & \sqrt{d_{14}d_{1N}} \cos \theta_{21N} & \cdots & d_{1N}
\end{bmatrix} \in \mathbb{S}^{N-1}
\]

(1038)
\(^{5.23}\) \[ \begin{bmatrix}
  1 & -e^{i\theta_{ikj}} \\
  -e^{-i\theta_{ikj}} & 1
\end{bmatrix} \succeq 0, \text{ having eigenvalues } \{0, 2\}. \]

Minimum is attained for \( \Theta^T \Theta \) when

\[
\begin{align*}
\Theta^T \Theta &= \begin{bmatrix}
m_1 & \mu \\
\mu e^{-i\theta} & \mu
\end{bmatrix}, & m_1 & \geq 0, & \mu & \geq 0, & \theta_{ikj} & = 0 \\
\Theta^T \Theta &= \begin{bmatrix} 0 & \mu \\
\mu e^{-i\theta} & 0
\end{bmatrix}, & -\pi & \leq \theta_{ikj} & \leq \pi, & \theta_{ikj} & \neq 0
\end{align*}
\]

(§D.2.1, [63, exmp.4.5]).
then any EDM may be expressed
\[ D(\Theta) \triangleq \begin{bmatrix} 0 & \delta(\Theta^T \Theta) \\ \delta(\Theta^T \Theta) & 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ \delta(\Theta^T \Theta) \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ \Theta^T \Theta \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} \in \text{EDM} \]
(1039)
for which all Euclidean metric properties hold. Entries of \( \Theta^T \Theta \) result from vector inner-products as in (1035); id est, \( \Theta = [x_2 - x_1, x_3 - x_1, \ldots, x_N - x_1] = X\sqrt{\mathcal{V}} \in \mathbb{R}^{n \times N} \)
(1041)
Inner product \( \Theta^T \Theta \) is obviously related to a Gram matrix (978),
\[ G = \begin{bmatrix} 0 & 0 \\ 0 & \Theta^T \Theta \end{bmatrix}, \quad x_1 = 0 \]
(1042)
For \( D = D(\Theta) \) and no condition on the list \( X \) (confer (986) (990))
\[ \Theta^T \Theta = -\mathcal{V}_X^T \mathcal{V} \in \mathbb{R}^{N-1 \times N-1} \]
(1043)

5.4.3.1 Relative-angle form

The inner-product form EDM definition is not a unique definition of Euclidean distance matrix; there are approximately five flavors distinguished by their argument to operator \( D \). Here is another one:

Like \( D(X) \) (969), \( D(\Theta) \) will make an EDM given any \( \Theta \in \mathbb{R}^{n \times N-1} \), it is neither a convex function of \( \Theta \) (§5.4.3.2), and it is homogeneous in the sense (972). Scrutinizing \( \Theta^T \Theta \) (1038) we find that because of the arbitrary choice \( k = 1 \), distances therein are all with respect to point \( x_1 \). Similarly, relative angles in \( \Theta^T \Theta \) are between all vector pairs having vertex \( x_1 \). Yet picking arbitrary \( \theta_{ij} \) to fill \( \Theta^T \Theta \) will not necessarily make an EDM; inner product (1038) must be positive semidefinite.

\[ \Theta^T \Theta = \sqrt{\delta(d)} \Omega \sqrt{\delta(d)} \triangleq \]
\[ \begin{bmatrix} \sqrt{d_{12}} \\ \sqrt{d_{13}} \\ \vdots \\ \sqrt{d_{1N}} \end{bmatrix} \begin{bmatrix} 1 & \cos \theta_{213} & \cdots & \cos \theta_{21N} \\ \cos \theta_{213} & 1 & \cdots & \cos \theta_{31N} \\ \vdots & \vdots & \ddots & \vdots \\ \cos \theta_{21N} & \cos \theta_{31N} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sqrt{d_{12}} \\ \sqrt{d_{13}} \\ \vdots \\ \sqrt{d_{1N}} \end{bmatrix} \]
(1044)
Expression \( D(\Theta) \) defines an EDM for any positive semidefinite relative-angle matrix
\[ \Omega = [\cos \theta_{ij}, \ i, j = 2 \ldots N] \in \mathbb{S}^{N-1} \]
(1045)
and any nonnegative distance vector
\[ d = [d_{ij}, j = 2 \ldots N] = \delta(\Theta^T \Theta) \in \mathbb{R}^{N-1} \]  
(1046)
because (\S A.3.1.0.5)
\[ \Omega \succeq 0 \Rightarrow \Theta^T \Theta \succeq 0 \]  
(1047)
Decomposition (1044) and the relative-angle matrix inequality \[ \Omega \succeq 0 \] lead to a different expression of an inner-product form EDM definition (1039)
\[
\begin{align*}
D(\Omega, d) & \triangleq \begin{bmatrix} 0 \\ d \end{bmatrix}^T \mathbf{1} + 1 \begin{bmatrix} 0 & d^T \end{bmatrix} - 2 \sqrt{\delta \left( \begin{bmatrix} 0 \\ d \end{bmatrix} \right) \Omega \sqrt{\delta(d)}} \\
& = \begin{bmatrix} 0 & d^T \end{bmatrix} \mathbf{1} + 1 d^T - 2 \sqrt{\delta(d)} \Omega \sqrt{\delta(d)}} \end{align*}
\]  
(1048)
and another expression of the EDM cone:
\[
\text{EDM}^N = \left\{ D(\Omega, d) \mid \Omega \succeq 0, \sqrt{\delta(d)} \succeq 0 \right\} 
\]  
(1049)
In the particular circumstance \( x_1 = 0 \), we can relate interpoint angle matrix \( \Psi \) from the Gram decomposition in (978) to relative-angle matrix \( \Omega \) in (1044). Thus,
\[
\Psi = \begin{bmatrix} 1 & 0^T \\ 0 & \Omega \end{bmatrix}, \quad x_1 = 0
\]  
(1050)
5.4.3.2 Inner-product form \(-V_N^T D(\Theta) V_N^T\) convexity
On page 385 we saw that each EDM entry \( d_{ij} \) is a convex quadratic function of \[ \frac{\sqrt{d_{ik}}}{\sqrt{d_{kj}}} \] and a quasiconvex function of \( \theta_{ikj} \). Here the situation for inner-product form EDM operator \( D(\Theta) \) (1039) is identical to that in \( \S 5.4.1 \) for list-form \( D(X) \); \( -D(\Theta) \) is not a quasiconvex function of \( \Theta \) by the same reasoning, and from (1043)
\[
-\n\]  
(1051)
is a convex quadratic function of \( \Theta \) on domain \( \mathbb{R}^{n \times N-1} \) achieving its minimum at \( \Theta = 0 \).
5.4.3.3 Inner-product form, discussion
We deduce that knowledge of interpoint distance is equivalent to knowledge of distance and angle from the perspective of one point, \( x_1 \) in our chosen case. The total amount of information \( N(N-1)/2 \) in \( \Theta^T \Theta \) is unchanged with respect to EDM \( D \).
\[ \text{5.24} \] The reason for amount \( \mathcal{O}(N^2) \) information is because of the relative measurements. Use of a fixed reference in measurement of angles and distances would reduce required information but is antithetical. In the particular case \( n = 2 \), for example, ordering all points \( x_\ell \) (in a length-\( N \) list) by increasing angle of vector \( x_\ell - x_1 \) with respect to \( x_2 - x_1 \), \( \theta_{1ij} \) becomes equivalent to \( \sum_{k=1}^{j-1} \theta_{k,1,k+1} \leq 2\pi \) and the amount of information is reduced to \( 2N-3 \); rather, \( \mathcal{O}(N) \).
5.5 Invariance

When \( D \) is an EDM, there exist an infinite number of corresponding \( N \)-point lists \( X \) (76) in Euclidean space. All those lists are related by isometric transformation: rotation, reflection, and translation (offset or shift).

5.5.1 Translation

Any translation common among all the points \( x_\ell \) in a list will be cancelled in the formation of each \( d_{ij} \). Proof follows directly from (965). Knowing that translation \( \alpha \) in advance, we may remove it from the list constituting the columns of \( X \) by subtracting \( \alpha 1^T \). Then it stands to reason by list-form definition (969) of an EDM, for any translation \( \alpha \in \mathbb{R}^n \)

\[
D(X - \alpha 1^T) = D(X)
\]

In words, interpoint distances are unaffected by offset; EDM \( D \) is translation invariant. When \( \alpha = x_1 \) in particular,

\[
\begin{bmatrix}
x_2-x_1 & x_3-x_1 & \cdots & x_N-x_1
\end{bmatrix} = X\sqrt{2}V_N \in \mathbb{R}^{n \times N-1}
\]

and so

\[
D(X - x_1 1^T) = D(X - x_1 1^T) = D(X \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}) = D(X)
\]

5.5.1.0.1 Example. Translating geometric center to origin.

We might choose to shift the geometric center \( \alpha_c \) of an \( N \)-point list \( \{x_\ell\} \) (arranged columnar in \( X \)) to the origin; \( [364] [173] \)

\[
\alpha = \alpha_c \triangleq Xb_c \triangleq X1 \frac{1}{N} \in \mathcal{P} \subseteq \mathcal{A}
\]

where \( \mathcal{A} \) represents the list’s affine hull. If we were to associate a point-mass \( m_\ell \) with each of the points \( x_\ell \) in the list, then their center of mass (or gravity) would be \((\sum x_\ell m_\ell) / \sum m_\ell \). The geometric center is the same as the center of mass under the assumption of uniform mass density across points. \([229]\) The geometric center always lies in the convex hull \( \mathcal{P} \) of the list; \( \text{id est,} \ \alpha_c \in \mathcal{P} \) because \( b_c^T 1 = 1 \) and \( b_c \geq 0 \). \(5.25\)

Subtracting the geometric center from every list member,

\[
X - \alpha_c 1^T = X - \frac{1}{N}X11^T = X(I - \frac{1}{N}11^T) = XV \in \mathbb{R}^{n \times N}
\]

where \( V \) is the geometric centering matrix (991). So we have (confer (969))

\[
D(X) = D(XV) = \delta(V^TX^TXV)1^T + 1\delta(V^TX^TXV)^T - 2V^TX^TXV \in \text{EDM}^N
\]

\(5.25\) Any \( b \) from \( \alpha = Xb \) chosen such that \( b^T 1 = 1 \), more generally, makes an auxiliary \( V \)-matrix. (§B.4.5)
5.5. INVARIANCE

5.5.1.1 Gram-form invariance

Following from (1056) and the linear Gram-form EDM operator (981):

\[ D(G) = D(V G V) = \delta(V G V)1^T + \delta(V G V)^T - 2V G V \in \text{EDM}^N \]  

(1057)

The Gram-form consequently exhibits invariance to translation by a doublet \( u1^T + 1u^T \) (§B.2)

\[ D(G) = D(G - (u1^T + 1u^T)) \]  

(1058)

because, for any \( u \in \mathbb{R}^N \), \( D(u1^T + 1u^T) = 0 \). The collection of all such doublets forms the nullspace (1074) to the operator; the translation-invariant subspace \( S_N^\perp \) (2097) of the symmetric matrices \( S_N \). This means matrix \( G \) is not unique and can belong to an expanse more broad than a positive semidefinite cone; \textit{id est}, \( G \in S_N^+ - S_N^\perp \). So explains Gram matrix sufficiency in EDM definition (981).

5.5.2 Rotation/Reflection

Rotation of the list \( X \in \mathbb{R}^{n \times N} \) about some arbitrary point \( \alpha \in \mathbb{R}^n \), or reflection through some affine subset containing \( \alpha \), can be accomplished via \( Q(X - \alpha 1^T) \) where \( Q \) is an orthogonal matrix (§B.5).

We rightfully expect

\[ D(Q(X - \alpha 1^T)) = D(QX - \beta 1^T) = D(QX) = D(X) \]  

(1059)

Because list-form \( D(X) \) is translation invariant, we may safely ignore offset and consider only the impact of matrices that premultiply \( X \). Interpoint distances are unaffected by rotation or reflection; we say, EDM \( D \) is rotation/reflection invariant. Proof follows from the fact, \( Q^T = Q^{-1} \Rightarrow X^TQ^TQX = X^TX \). So (1059) follows directly from (969).

The class of premultiplying matrices for which interpoint distances are unaffected is a little more broad than orthogonal matrices. Looking at EDM definition (969), it appears that any matrix \( Q_p \) such that

\[ X^TQ_p^TQ_pX = X^TX \]  

(1060)

will have the property

\[ D(Q_pX) = D(X) \]  

(1061)

An example is skinny \( Q_p \in \mathbb{R}^{m \times n} (m > n) \) having orthonormal columns; an orthonormal matrix.

5.5.2.0.1 Example. Reflection prevention and quadrature rotation.

Consider the EDM graph in Figure 140b representing known distance between vertices (Figure 140a) of a tilted-square diamond in \( \mathbb{R}^2 \). Suppose some geometrical optimization problem were posed where isometric transformation is allowed excepting reflection, and where rotation must be quantized so that only \textit{quadrature} rotations are allowed; only multiples of \( \pi/2 \).
Figure 140: (a) Four points in quadrature in two dimensions about their geometric center. (b) Complete EDM graph of diamond-shaped vertices. (c) Quadrature rotation of Euclidean body in $\mathbb{R}^2$ first requires shroud: the smallest Cartesian square containing it.

In two dimensions, a counterclockwise rotation of any vector about the origin by angle $\theta$ is prescribed by the orthogonal matrix

$$Q = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

(1062)

whereas reflection of any point through a hyperplane containing the origin

$$\partial \mathcal{H} = \left\{ x \in \mathbb{R}^2 \left| \left[ \begin{array}{c} \cos \theta \\ \sin \theta \end{array} \right]^T x = 0 \right. \right\}$$

(1063)

is accomplished via multiplication with symmetric orthogonal matrix (§B.5.3)

$$R = \begin{bmatrix} \sin^2(\theta) - \cos^2(\theta) & -2\sin(\theta)\cos(\theta) \\ -2\sin(\theta)\cos(\theta) & \cos^2(\theta) - \sin^2(\theta) \end{bmatrix}$$

(1064)

Rotation matrix $Q$ is characterized by identical diagonal entries and by antidiagonal entries equal but opposite in sign, whereas reflection matrix $R$ is characterized in the reverse sense.

Assign the diamond vertices $\{ x_\ell \in \mathbb{R}^2, \ell = 1 \ldots 4 \}$ to columns of a matrix

$$X = [ x_1 \ x_2 \ x_3 \ x_4 ] \in \mathbb{R}^{2 \times 4}$$

(76)

Our scheme to prevent reflection enforces a rotation matrix characteristic upon the coordinates of adjacent points themselves: First shift the geometric center of $X$ to the origin; for geometric centering matrix $V \in \mathbb{S}^4$ (§5.5.1.0.1), define

$$Y \triangleq XV \in \mathbb{R}^{2 \times 4}$$

(1065)

To maintain relative quadrature between points (Figure 140a) and to prevent reflection, it is sufficient that all interpoint distances be specified and that adjacencies $Y(:, 1:2),$
5.6. **INJECTIVITY OF D & UNIQUE RECONSTRUCTION**

Let $Y(:, 2:3)$ and $Y(:, 3:4)$ be proportional to $2 \times 2$ rotation matrices; any clockwise rotation would ascribe a reflection matrix characteristic. Counterclockwise rotation is thereby enforced by constraining equality among diagonal and antidiagonal entries as prescribed by (1062):

$$Y(:, 1:3) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} Y(:, 2:4)$$ (1066)

Quadrature quantization of rotation can be regarded as a constraint on tilt of the smallest Cartesian square containing the diamond as in Figure 140c. Our scheme to quantize rotation requires that all square vertices be described by vectors whose entries are nonnegative when the square is translated anywhere interior to the nonnegative orthant. We capture the four square vertices as columns of a product $YC$ where

$$C = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$ (1067)

Then, assuming a unit-square shroud, the affine constraint

$$YC + \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} 1^T \geq 0$$ (1068)

quantizes rotation, as desired.

5.5.2.1 **Inner-product form invariance**

Likewise, $D(\Theta)$ (1039) is rotation/reflection invariant;

$$D(Q_p \Theta) = D(Q \Theta) = D(\Theta)$$ (1069)

so (1060) and (1061) similarly apply.

5.5.3 **Invariance conclusion**

In the making of an EDM, absolute rotation, reflection, and translation information is lost. Given an EDM, reconstruction of point position (§5.12, the list $X$) can be guaranteed correct only in affine dimension $r$ and relative position. Given a noiseless complete EDM, this isometric reconstruction is unique insofar as every realization of a corresponding list $X$ is congruent:

5.6 **Injectivity of $D$ & unique reconstruction**

Injectivity implies uniqueness of isometric reconstruction (§5.4.2.2.10); hence, we endeavor to demonstrate it.
Figure 141: Orthogonal complements in $\mathbb{S}^N$ abstractly oriented in isometrically isomorphic $\mathbb{R}^{N(N+1)/2}$. Case $N = 2$ accurately illustrated in $\mathbb{R}^3$. Orthogonal projection of basis for $\mathbb{S}^N_{c \perp}$ on $\mathbb{S}^N_{h \perp}$ yields another basis for $\mathbb{S}^N_{c \perp}$. (Basis vectors for $\mathbb{S}^N_{c \perp}$ are illustrated lying in a plane orthogonal to $\mathbb{S}^N_{c}$ in this dimension. Basis vectors for each $\perp$ space outnumber those for its respective orthogonal complement; such is not the case in higher dimension.)

EDM operators list-form $\mathbf{D}(X)$ (969), Gram-form $\mathbf{D}(G)$ (981), and inner-product form $\mathbf{D}(\Theta)$ (1039) are many-to-one surjections (§5.5) onto the same range; the EDM cone (§6): (confer (982) (1076))

\[
\text{EDM}^N = \left\{ \mathbf{D}(X) : \mathbb{R}^{N-1 \times N} \to \mathbb{S}^N_h \mid X \in \mathbb{R}^{N-1 \times N} \right\} = \left\{ \mathbf{D}(G) : \mathbb{S}^N \to \mathbb{S}^N_h \mid G \in \mathbb{S}^N - \mathbb{S}^N_{c \perp} \right\} = \left\{ \mathbf{D}(\Theta) : \mathbb{R}^{N-1 \times N-1} \to \mathbb{S}^N_h \mid \Theta \in \mathbb{R}^{N-1 \times N-1} \right\} \tag{1070}
\]

where (§5.5.1.1)

\[
\mathbb{S}^N_{c \perp} = \{ u\mathbf{1}^T + \mathbf{1}u^T \mid u \in \mathbb{R}^N \} \subseteq \mathbb{S}^N \tag{2097}
\]
5.6. **Injectivity of \(D\) & Unique Reconstruction**

### 5.6.1 Gram-form bijectivity

Because linear Gram-form EDM operator

\[
D(G) = \delta(G) 1^T + 1 \delta(G)^T - 2G
\]

has no nullspace [88, §A.1] on the geometric center subspace \(^{5.27}\) \((\text{§E.7.2.0.2})\)

\[
\mathbb{S}_c^N \triangleq \{ G \in \mathbb{S}^N \mid G1 = 0 \}
\]

\[
= \{ G \in \mathbb{S}^N \mid \mathcal{N}(G) \supseteq 1 \} = \{ G \in \mathbb{S}^N \mid \mathcal{R}(G) \subseteq \mathcal{N}(1^T) \}
\]

\[
= \{ VYV \mid Y \in \mathbb{S}^N \} \subset \mathbb{S}^N
\]

\[
\equiv \{ V_Y^T A V_Y \mid A \in \mathbb{S}^{N-1} \}
\]

then \(D(G)\) on that subspace is injective.

To prove injectivity of \(D(G)\) on \(\mathbb{S}_c^N\): Any matrix \(Y \in \mathbb{S}^N\) can be decomposed into orthogonal components in \(\mathbb{S}_c^N\):

\[
Y = VYV + (Y - VYV)
\]

where \(VYV \in \mathbb{S}_c^N\) and \(Y - VYV \in \mathbb{S}_c^{N\perp}\) \(^{2097}\). Because of translation invariance (§5.5.1.1) and linearity, \(D(Y - VYV) = 0\) hence \(\mathcal{N}(D) \supseteq \mathbb{S}_c^{N\perp}\). It remains only to show

\[
D(VYV) = 0 \iff VYV = 0
\]

\((\iff Y = u1^T + 1u^T \text{ for some } u \in \mathbb{R}^N)\). \(D(VYV)\) will vanish whenever \(2VYV = \delta(VYV)1^T + 1 \delta(VYV)^T\). But this implies \(\mathcal{R}(1)\) (§B.2) were a subset of \(\mathcal{R}(VYV)\), which is contradictory. Thus we have

\[
\mathcal{N}(D) = \{ Y \mid D(Y) = 0 \} = \{ Y \mid VYV = 0 \} = \mathbb{S}_c^{N\perp}
\]

\[^{2096}\]

Since \(G1 = 0 \iff X1 = 0\) \(^{989}\) simply means list \(X\) is geometrically centered at the origin, and because the Gram-form EDM operator \(D\) is translation invariant and \(\mathcal{N}(D)\) is the translation-invariant subspace \(\mathbb{S}_c^{N\perp}\), then EDM definition \(D(G)\) \(^{1070}\) on \(^{5.28}\) \((\text{confer §6.5.1, §6.6.1, §A.7.4.0.1})\)

\[
\mathbb{S}_c^N \cap \mathbb{S}_+^N = \{ VYV \geq 0 \mid Y \in \mathbb{S}^N \} = \{ V_Y^T A V_Y^T \mid A \in \mathbb{S}_+^{N-1} \} \subset \mathbb{S}^N
\]

must be surjective onto \(\text{EDM}^N\): \((\text{confer(982)})\)

\[
\text{EDM}^N = \left\{ D(G) \mid G \in \mathbb{S}_c^N \cap \mathbb{S}_+^N \right\}
\]

\(^{5.27}\): Equivalence \(\equiv\) in (1071) follows from the fact: Given \(B = VYV = V_Y^T A V_Y \in \mathbb{S}_c^N\) with only matrix \(A \in \mathbb{S}_+^{N-1}\) unknown, then \(V_Y^T B V_Y = A\) or \(V_Y^T Y V_Y = A\).

\(^{5.28}\): Equivalence \(\equiv\) in (1075) follows from the fact: Given \(B = VYV = V_Y^T A V_Y \in \mathbb{S}_+^N\) with only matrix \(A\) unknown, then \(V_Y^T B V_Y = A\) and \(A \in \mathbb{S}_+^{N-1}\) must be positive semidefinite by positive semidefiniteness of \(B\) and Corollary A.3.1.0.5.
5.6.1.1 Gram-form operator $D$ inversion

Define the linear geometric centering operator $V$; (confer (990))

$V(D): S^N \rightarrow S^N \triangleq -VDV_2^{-1}$

(1077)

This orthogonal projector $V$ has no nullspace on $S^N_h = \text{aff EDM}(N)$ because the projection of $-D/2$ on $S^N_c$ (2095) can be 0 if and only if $D \in S^N \perp_c$; but $S^N \perp_c \cap S^N_h = 0$ (Figure 141). Projector $V$ on $S^N_h$ is therefore injective hence uniquely invertible. Further, $-V S^N_h V/2$ is equivalent to the geometric center subspace $S^N_c$ in the ambient space of symmetric matrices; a surjection,

$S^N_c = V(S^N) = V(S^N_h \oplus S^N_{h \perp}) = V(S^N_h)$

(1078)

Because $D(G)$ on $S^N_c$ is injective, and $\text{aff } D(V(EDM(N))) = D(V(\text{aff EDM}(N)))$ by property (127) of the affine hull, we find for $D \in S^N_h$

$D(-VDV_2^{-1}) = \delta(-VDV_2^{-1})1^T + 1\delta(-VDV_2^{-1})^T - 2(-VDV_2^{-1})$

id est,

$D = D(V(D))$

(1081)

$-VDV = V(D(-VDV))$

(1082)

or

$S^N_h = D(V(S^N_h))$

(1083)

$-VS^N_h V = V(D(-VS^N_h V))$

(1084)

These operators $V$ and $D$ are mutual inverses.

The Gram-form $D(S^N_c)$ (981) is equivalent to $S^N_h$:

$D(S^N_c) = D(V(S^N_h \oplus S^N_{h \perp})) = S^N_h + D(V(S^N_{h \perp})) = S^N_h$

(1085)

because $S^N_h \supseteq D(V(S^N_{h \perp}))$. In summary, for the Gram-form we have the isomorphisms

$S^N_h = D(S^N_c)$

(1086)

$S^c = V(S^N_h)$

(1087)
and from bijectivity results in §5.6.1,
\[ \mathrm{EDM}^N_D = D(S_c^N \cap S_+^N) \]  
(1088)
\[ S_c^N \cap S_+^N = V(\mathrm{EDM}^N) \]  
(1089)

5.6.2 Inner-product form bijectivity
The Gram-form EDM operator \( D(G) = \delta(G)1^T + 1\delta(G)^T - 2G \) (981) is an injective map, for example, on the domain that is the subspace of symmetric matrices having all zeros in the first row and column
\[ S_1^N = \{ G \in S^N \mid Ge_1 = 0 \} \]
\[ = \left\{ \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} Y \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} \mid Y \in S^N \right\} \]  
(1009)
because it obviously has no nullspace there. Since \( Ge_1 = 0 \iff Xe_1 = 0 \) (983) means the first point in the list \( X \) resides at the origin, then \( D(G) \) on \( S_1^N \cap S_+^N \) must be surjective onto \( \mathrm{EDM}^N \).

Substituting \( \Theta^T \Theta \leftarrow -V_N^T D V_N \) (1051) into inner-product form EDM definition \( D(\Theta) \) (1039), it may be further decomposed:
\[ D(D) = \begin{bmatrix} 0 \\ \delta(-V_N^T D V_N) \end{bmatrix} 1^T + 1 \begin{bmatrix} 0 & \delta(-V_N^T D V_N)^T \end{bmatrix} - 2 \begin{bmatrix} 0 & 0^T \\ 0 & -V_N^T D V_N \end{bmatrix} \]  
(1089)
This linear operator \( D \) is another flavor of inner-product form and an injective map of the EDM cone onto itself. Yet when its domain is instead the entire symmetric hollow subspace \( S_h^N = \text{aff EDM}^N, \) \( D(D) \) becomes an injective map onto that same subspace. Proof follows directly from the fact: linear \( D \) has no nullspace [88, §A.1] on \( S_h^N = \text{aff EDM}^N) = D(\text{aff EDM}^N) \) (127).

5.6.2.1 Inversion of \( D(-V_N^T D V_N) \)
Injectivity of \( D(D) \) suggests inversion of \( \text{(confers (986))} \)
\[ V_N(D) : S^N \to S^{N-1} \triangleq -V_N^T D V_N \]  
(1091)
a linear surjective\(^{5.31} \) mapping onto \( S^{N-1} \) having nullspace\(^{5.32} \) \( S_c^{N \perp} \);
\[ V_N(S_h^N) = S^{N-1} \]  
(1092)
\(^{5.31} \) Surjectivity of \( V_N(D) \) is demonstrated via the Gram-form EDM operator \( D(G) : \) Since \( S_h^N = D(S_c^N) \) (1085), then for any \( Y \in S^{N-1}, -V_N^T D(V_N^T Y V_N^T / 2) V_N = Y. \)
\(^{5.32} \) \( N(V_N) \supseteq S_c^{N \perp} \) is apparent. There exists a linear mapping
\[ T(V_N(D)) \triangleq V_N^T V_N(D) V_N^T = -VDV_N^T = V(D) \]
such that
\[ N(T(V_N)) = N(V) \supseteq N(V_N) \supseteq S_c^{N \perp} = N(V) \]
where the equality \( S_c^{N \perp} = N(V) \) is known (§E.7.2.0.2).
injective on domain $S_h^N$ because $S_c^{N,⊥} \cap S_h^N = \emptyset$. Revising the argument of this inner-product form (1090), we get another flavor

$$D(-V_N^T D V_N) = \begin{bmatrix} 0 & \delta(-V_N^T D V_N) \\ 0 & 0 \end{bmatrix}^T + 1 \begin{bmatrix} 0 & \delta(-V_N^T D V_N)^T \\ 0 & 0 \end{bmatrix} - 2 \begin{bmatrix} 0 & 0^T \\ 0 & V_N^T D V_N \end{bmatrix}$$

and we obtain mutual inversion of operators $V_N$ and $D$, for $D \in S_h^N$

$$D = D(V_N(D))$$

$$-V_N^T D V_N = V_N(D(-V_N^T D V_N))$$

or

$$S_h^N = D(V_N(S_h^N))$$

$$-V_N^T S_h^N V_N = V_N(D(-V_N^T S_h^N V_N))$$

Substituting $\Theta^T \Theta \leftarrow \Phi$ into inner-product form EDM definition (1039), any EDM may be expressed by the new flavor

$$D(\Phi) \triangleq \begin{bmatrix} 0 & \delta(\Phi) \\ \delta(\Phi)^T & 0 \end{bmatrix}^T + 1 \begin{bmatrix} 0 & \delta(\Phi)^T \\ 0 & 0 \end{bmatrix} - 2 \begin{bmatrix} 0 & 0^T \\ 0 & \Phi \end{bmatrix} \in \text{EDM}^N$$

$$\Leftrightarrow \Phi \succeq 0$$

where this $D$ is a linear surjective operator onto $\text{EDM}^N$ by definition, injective because it has no nullspace on domain $S_{+}^{N-1}$. More broadly, $\text{aff} D(S_{+}^{N-1}) = D(\text{aff} S_{+}^{N-1})$ (127),

$$S_h^N = D(S_{+}^{N-1})$$

$$S_{+}^{N-1} = V_N(S_h^N)$$

demonstrably isomorphisms, and by bijectivity of this inner-product form:

$$\text{EDM}^N = D(S_{+}^{N-1})$$

$$S_{+}^{N-1} = V_N(\text{EDM}^N)$$

5.7 Embedding in affine hull

The affine hull $A$ (78) of a point list $\{x_\ell\}$ (arranged columnar in $X \in \mathbb{R}^{n \times N}$ (76)) is identical to the affine hull of that polyhedron $P$ (86) formed from all convex combinations of the $x_\ell$; [63, §2] [318, §17]

$$A = \text{aff} X = \text{aff} P$$

Comparing hull definitions (78) and (86), it becomes obvious that the $x_\ell$ and their convex hull $P$ are embedded in their unique affine hull $A$;

$$A \supseteq P \supseteq \{x_\ell\}$$
5.7. EMBEDDING IN AFFINE HULL

Recall: affine dimension $r$ is a lower bound on embedding, equal to dimension of the subspace parallel to that nonempty affine set $A$ in which the points are embedded. (§2.3.1) We define dimension of the convex hull $P$ to be the same as dimension $r$ of the affine hull $A$ [318, §2], but $r$ is not necessarily equal to rank of $X$ (1122).

For the particular example illustrated in Figure 127, $P$ is the triangle in union with its relative interior while its three vertices constitute the entire list $X$. Affine hull $A$ is the unique plane that contains the triangle, so affine dimension $r = 2$ in that example while rank of $X$ is 3. Were there only two points in Figure 127, then the affine hull would instead be the unique line passing through them; $r$ would become 1 while rank would then be 2.

5.7.1 Determining affine dimension

Knowledge of affine dimension $r$ becomes important because we lose any absolute offset common to all the generating $x_\ell$ in $\mathbb{R}^n$ when reconstructing convex polyhedra given only distance information. (§5.5.1) To calculate $r$, we first remove any offset that serves to increase dimensionality of the subspace required to contain polyhedron $P$; subtracting any $\alpha \in A$ in the affine hull from every list member will work,

\[ X - \alpha 1^T \] (1104)

translating $A$ to the origin: 5.33

\[ A - \alpha = \text{aff}(X - \alpha 1^T) = \text{aff}(X) - \alpha \] (1105)
\[ P - \alpha = \text{conv}(X - \alpha 1^T) = \text{conv}(X) - \alpha \] (1106)

Because (1102) and (1103) translate,

\[ \mathbb{R}^n \supseteq A - \alpha = \text{aff}(X - \alpha 1^T) = \text{aff}(P - \alpha) \supseteq P - \alpha \supseteq \{x_\ell - \alpha\} \] (1107)

where from the previous relations it is easily shown

\[ \text{aff}(P - \alpha) = \text{aff}(P) - \alpha \] (1108)

Translating $A$ neither changes its dimension or the dimension of the embedded polyhedron $P$; (77)

\[ r \triangleq \dim A = \dim(A - \alpha) \triangleq \dim(P - \alpha) = \dim P \] (1109)

For any $\alpha \in \mathbb{R}^n$, (1105)-(1109) remain true. [318, p.4, p.12] Yet when $\alpha \in A$, the affine set $A - \alpha$ becomes a unique subspace of $\mathbb{R}^n$ in which the $\{x_\ell - \alpha\}$ and their convex hull $P - \alpha$ are embedded (1107), and whose dimension is more easily calculated.

5.7.1.0.1 Example. Translating first list-member to origin.

Subtracting the first member $\alpha \triangleq x_1$ from every list member will translate their affine hull $A$ and their convex hull $P$ and, in particular, $x_1 \in P \subseteq A$ to the origin in $\mathbb{R}^n$; videlicet,

\[ X - x_1 1^T = X - Xe_1 1^T = X(I - e_1 1^T) = X\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} \in \mathbb{R}^{n \times N} \] (1110)
where \( V_N \) is defined in (975), and \( e_1 \) in (985). Applying (1107) to (1110),

\[ \mathbb{R}^n \supseteq \mathcal{R}(XV_N) = A - x_1 = \text{aff}(X - x_1^1) = \text{aff}(\mathcal{P} - x_1) \supseteq \mathcal{P} - x_1 \ni 0 \]  

(1111)

where \( XV_N \in \mathbb{R}^{n \times N - 1} \). Hence

\[ r = \dim \mathcal{R}(XV_N) \]  

(1112)

Since shifting the geometric center to the origin (§5.5.1.0.1) translates the affine hull to the origin as well, then it must also be true

\[ r = \dim \mathcal{R}(XV) \]  

(1113)

For any matrix whose range is \( \mathcal{R}(V) = \mathcal{N}(1^T) \) we get the same result; e.g.,

\[ r = \dim \mathcal{R}(XV_N^{1T}) \]  

(1114)

because

\[ \mathcal{R}(XV) = \{Xz \mid z \in \mathcal{N}(1^T)\} \]  

(1115)

and \( \mathcal{R}(V) = \mathcal{R}(V_N) = \mathcal{R}(V_N^{1T}) \) (§E). These auxiliary matrices (§B.4.2) are more closely related:

\[ V = V_N V_N^1 \]  

(1740)

### 5.7.1.1 Affine dimension \( r \) versus rank

Now, suppose \( D \) is an EDM as defined by

\[ D(X) = \delta(X^TX)1^T + 1\delta(X^TX)^T - 2X^TX \in \text{EDM}^N \]  

(969)

and we premultiply by \(-V_N^T\) and postmultiply by \(V_N\). Then because \( V_N^T 1 = 0 \) (976), it is always true that

\[ -V_N^T DV_N = 2V_N^X X^TXV_N = 2V_N^T GV_N \in \mathbb{S}^{N-1} \]  

(1116)

where \( G \) is a Gram matrix. Similarly pre- and postmultiplying by \( V \) (confer (990))

\[ -VDV = 2VX^TXV = 2GV \in \mathbb{S}^N \]  

(1117)

always holds because \( V1 = 0 \) (1730). Likewise, multiplying inner-product form EDM definition (1039), it always holds:

\[ -V_N^TDV_N = \Theta^T \Theta \in \mathbb{S}^{N-1} \]  

(1043)

For any matrix \( A \), \( \text{rank} A^T A = \text{rank} A = \text{rank} A^T \). \[212, \S0.4\] \[5.34\] So, by (1115), affine dimension

\[ r = \text{rank} XV = \text{rank} XV_N = \text{rank} XV_N^{1T} = \text{rank} \Theta \]  

\[ = \text{rank} VDV = \text{rank} VGV = \text{rank} V_N^T DV_N = \text{rank} V_N^T GV_N \]  

(1118)

\[ ^{5.34}\text{For } A \in \mathbb{R}^{m \times n}, \mathcal{N}(A^T A) = \mathcal{N}(A). \ [341, \S3.3] \]
By conservation of dimension, (§A.7.3.0.1)

\[ r + \dim \mathcal{N}(V_N^T D V_N) = N - 1 \quad (1119) \]
\[ r + \dim \mathcal{N}(D V) = N \quad (1120) \]

For \( D \in \text{EDM}^N \)

\[ -V_N^T D V_N > 0 \iff r = N - 1 \quad (1121) \]

but \(-D V \not\succ 0\). The general fact\(^{5.35}\) (confer (1001))

\[ r \leq \min\{n, N - 1\} \quad (1122) \]

is evident from (1110) but can be visualized in the example illustrated in Figure 127. There we imagine a vector from the origin to each point in the list. Those three vectors are linearly independent in \( \mathbb{R}^3 \), but affine dimension \( r \) is 2 because the three points lie in a plane. When that plane is translated to the origin, it becomes the only subspace of dimension \( r = 2 \) that can contain the translated triangular polyhedron.

### 5.7.2 Précis

We collect expressions for affine dimension \( r \): for list \( X \in \mathbb{R}^{n \times N} \) and Gram matrix \( G \in \mathbb{S}_+^N \)

\[ r \triangleq \dim(\mathcal{P} - \alpha) = \dim \mathcal{P} = \dim \text{conv} X \]
\[ = \dim(\mathcal{A} - \alpha) = \dim \mathcal{A} = \dim \text{aff} X \]
\[ = \text{rank}(X - x_1 1^T) = \text{rank}(X - \alpha^T 1) \]
\[ = \text{rank} \Theta \quad (1041) \]
\[ = \text{rank} X V_N = \text{rank} XV = \text{rank} X V_N^T \]
\[ = \text{rank} X, \quad X e_1 = 0 \text{ or } X 1 = 0 \]
\[ = \text{rank} V_N^T G V_N = \text{rank} V G V = \text{rank} V_N^T G V_N \]
\[ = \text{rank} G, \quad G e_1 = 0 \quad (986) \text{ or } G 1 = 0 \quad (990) \]
\[ = \text{rank} V_N^T D V_N = \text{rank} D V = \text{rank} V_N^T D V_N = \text{rank} V_N (V_N^T D V_N) V_N^T \]
\[ = \text{rank} \Lambda \quad (1209) \]
\[ = N - 1 - \dim \mathcal{N} \left( \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \right) = \text{rank} \left[ \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} - 2 \right] \quad (1130) \quad \{ D \in \text{EDM}^N \} \]

### 5.7.3 Eigenvalues of \(-D V \) versus \(-V_N^T D V_N \)

Suppose for \( D \in \text{EDM}^N \) we are given eigenvectors \( v_i \in \mathbb{R}^N \) of \(-D V \) and corresponding eigenvalues \( \lambda \in \mathbb{R}^N \) so that

\[ -D V v_i = \lambda_i v_i, \quad i = 1 \ldots N \quad (1124) \]

From these we can determine the eigenvectors and eigenvalues of \(-V_N^T D V_N\): Define

\[ v_i \triangleq V_N^T v_i, \quad \lambda_i \neq 0 \quad (1125) \]

\(^{5.35}\) rank \( X \leq \min\{n, N\} \)
Then we have:

\[ -VDV_N \nu_i = \lambda_i \nu_i \]  
\[ V_N^\dagger VDV_N \nu_i = \lambda_i V_N^\dagger \nu_i \]  
\[ V_N^\dagger DV_N \nu_i = \lambda_i \nu_i \]  

the eigenvectors of \(-V_N^\dagger DV_N\) are given by (1125) while its corresponding nonzero eigenvalues are identical to those of \(-VDV\) although \(-V_N^\dagger DV_N\) is not necessarily positive semidefinite. In contrast, \(-V_N^T DV_N\) is positive semidefinite but its nonzero eigenvalues are generally different.

5.7.3.0.1 Theorem. EDM rank versus affine dimension \(r\). [173, §3] [194, §3] [172, §3] For \(D \in \mathbb{EDM}^N\) (confer (1283))

1. \(r = \text{rank}(D) - 1 \iff 1^T D^\dagger 1 \neq 0\)
Points constituting a list \(X\) generating the polyhedron corresponding to \(D\) lie on the relative boundary of an \(r\)-dimensional circumhypersphere having
\[
\text{diameter} = \sqrt{2} (1^T D^\dagger 1)^{-1/2} \\
\text{circumcenter} = \frac{X D^\dagger 1}{1^T D^\dagger 1}
\]  

2. \(r = \text{rank}(D) - 2 \iff 1^T D^\dagger 1 = 0\)
There can be no circumhypersphere whose relative boundary contains a generating list for the corresponding polyhedron.

3. In Cayley-Menger form [119, §6.2] [91, §3.3] [52, §40] (§5.11.2),
\[
r = N - 1 - \dim N \left( \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \right) = \text{rank} \left( \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \right) - 2
\]  
Circumhyperspheres exist for \(r < \text{rank}(D) - 2\). [358, §7]

For all practical purposes, (1122)
\[
\max\{0, \text{rank}(D) - 2\} \leq r \leq \min\{n, N - 1\}
\]  

5.8 Euclidean metric versus matrix criteria

5.8.1 Nonnegativity property 1

When \(D = [d_{ij}]\) is an EDM (969), then it is apparent from (1116)
\[
2V_N^T X^T X V_N = -V_N^T DV_N \geq 0
\]
because for any matrix \( A \), \( A^T A \succeq 0 \).\(^{5.36}\) We claim nonnegativity of the \( d_{ij} \) is enforced primarily by the matrix inequality (1132); \textit{id est},

\[
-V_N^T D V_N \succeq 0 \\
\quad D \in \mathbb{S}_h^N
\] 

(1133)

(The matrix inequality to enforce strict positivity differs by a stroke of the pen. (1136))

We now support our claim: If any matrix \( A \in \mathbb{R}^{m \times m} \) is positive semidefinite, then its main diagonal \( \delta(A) \in \mathbb{R}^m \) must have all nonnegative entries. [168, §4.2]

Given \( D \in \mathbb{S}_h^N \)

\[-V_N^T D V_N =
\begin{bmatrix}
  d_{12} & \frac{1}{2}(d_{12} + d_{13} - d_{23}) & \frac{1}{2}(d_{1,1+1} + d_{1,1+1} - d_{1,j+1}) & \cdots & \frac{1}{2}(d_{1,1+1} + d_{1N} - d_{2N}) \\
  \frac{1}{2}(d_{12} + d_{13} - d_{23}) & d_{13} & \frac{1}{2}(d_{1,1+1} + d_{1,1+1} - d_{1,j+1}) & \cdots & \frac{1}{2}(d_{1,1+1} + d_{1N} - d_{3N}) \\
  \frac{1}{2}(d_{1,1+1} + d_{1,1+1} - d_{1,j+1}) & \frac{1}{2}(d_{1,1+1} + d_{1,1+1} - d_{1,j+1}) & d_{1,1+1} & \cdots & \frac{1}{2}(d_{1,1+1} + d_{1N} - d_{4N}) \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  \frac{1}{2}(d_{1,1+1} + d_{1N} - d_{2N}) & \frac{1}{2}(d_{1,1+1} + d_{1N} - d_{3N}) & \frac{1}{2}(d_{1,1+1} + d_{1N} - d_{4N}) & \cdots & d_{1N}
\end{bmatrix}
\]

\[= \frac{1}{2}(1D_{1,2:N} + d_{2:N,1}^T - D_{2:N,2:N}) \in \mathbb{S}_h^{N-1} \] (1134)

where row, column indices \( i,j \in \{1 \ldots N-1\} \). [323] It follows:

\[-V_N^T D V_N \succeq 0 \\
\quad D \in \mathbb{S}_h^N
\] 

\[
\Rightarrow \delta(-V_N^T D V_N) = \begin{bmatrix}
  d_{12} \\
  d_{13} \\
  \vdots \\
  d_{1N}
\end{bmatrix} \succeq 0 \quad (1135)
\]

Multiplication of \( V_N \) by any permutation matrix \( \Xi \) has null effect on its range and nullspace. In other words, any permutation of the rows or columns of \( V_N \) produces a basis for \( \mathcal{N}(1^T) \); \textit{id est}, \( \mathcal{R}(\Xi V_N) = \mathcal{R}(V_N \Xi) = \mathcal{R}(V_N) = \mathcal{N}(1^T) \).

Hence, \( -V_N^T D V_N \succeq 0 \iff -V_N^T \Xi^T D \Xi V_N \succeq 0 \) \( \iff -\Xi^T V_N^T D V_N \Xi \succeq 0 \).

Various permutation matrices\(^5.37\) will sift the remaining \( d_{ij} \) similarly to (1135) thereby proving their nonnegativity. Hence \( -V_N^T D V_N \succeq 0 \) is a sufficient test for the first property (§5.2) of the Euclidean metric, nonnegativity. ♦

When affine dimension \( r \) equals 1, in particular, nonnegativity symmetry and hollowness become necessary and sufficient criteria satisfying matrix inequality (1132). (§6.5.0.0.1)

\(^{5.36}\) For \( A \in \mathbb{R}^{m \times n} \), \( A^T A \succeq 0 \iff \| y^T A \|^2 \geq 0 \) for all \( \| y \| = 1 \). When \( A \) is full-rank skinny-or-square, \( A^T A \succeq 0 \).

\(^{5.37}\) The rule of thumb is: If \( \Xi(i,1) = 1 \), then \( \delta(-V_N^T \Xi^T D V_N) \in \mathbb{R}^{N-1} \) is some permutation of the \( i \)th row or column of \( D \) excepting the 0 entry from the main diagonal.
5.8.1.1 Strict positivity

Should we require the points in $\mathbb{R}^n$ to be distinct, then entries of $D$ off the main diagonal must be strictly positive $\{d_{ij} > 0, \ i \neq j\}$ and only those entries along the main diagonal of $D$ are 0. By similar argument, the strict matrix inequality is a sufficient test for strict positivity of Euclidean distance-square:

$\begin{align*}
-V_N^T D V_N \succeq 0 \\
D \in S_n^+ 
\end{align*}$  \quad \Rightarrow \quad d_{ij} > 0, \ i \neq j \quad (1136)

5.8.2 Triangle inequality property 4

In light of Kreyszig’s observation [237, §1.1 prob.15] that properties 2 through 4 of the Euclidean metric (§5.2) together imply nonnegativity property 1,

$$2\sqrt{d_{jk}} = \sqrt{d_{jk}} + \sqrt{d_{kj}} \geq \sqrt{d_{jj}} = 0, \ j \neq k \quad (1137)$$

nonnegativity criterion (1133) suggests that matrix inequality $-V_N^T D V_N \succeq 0$ might somehow take on the role of triangle inequality; 

$$\begin{align*}
\delta(D) = 0 \\
D^T = D \\
-V_N^T D V_N \succeq 0
\end{align*} \quad \Rightarrow \quad \sqrt{d_{ij}} \leq \sqrt{d_{ik}} + \sqrt{d_{kj}}, \ i \neq j \neq k \quad (1138)$$

We now show that is indeed the case: Let $T$ be the leading principal submatrix in $S^2$ of $-V_N^T D V_N$ (upper left $2 \times 2$ submatrix from (1134));

$$T \triangleq \begin{bmatrix}
d_{12} & \frac{1}{2}(d_{12} + d_{13} - d_{23}) \\
\frac{1}{2}(d_{12} + d_{13} - d_{23}) & d_{13}
\end{bmatrix} \quad (1139)$$

Submatrix $T$ must be positive (semi)definite whenever $-V_N^T D V_N$ is. (§A.3.1.0.4, §5.8.3)

Now we have,

$$-V_N^T D V_N \succeq 0 \Rightarrow T \succeq 0 \iff \lambda_1 \geq \lambda_2 \geq 0$$

$$-V_N^T D V_N \succ 0 \Rightarrow T \succ 0 \iff \lambda_1 > \lambda_2 > 0 \quad (1140)$$

where $\lambda_1$ and $\lambda_2$ are the eigenvalues of $T$, real due only to symmetry of $T$:

$$\begin{align*}
\lambda_1 &= \frac{1}{2} \left( d_{12} + d_{13} + \sqrt{d_{23}^2 - 2(d_{12} + d_{13})d_{23} + 2(d_{12}^2 + d_{13}^2)} \right) \\
\lambda_2 &= \frac{1}{2} \left( d_{12} + d_{13} - \sqrt{d_{23}^2 - 2(d_{12} + d_{13})d_{23} + 2(d_{12}^2 + d_{13}^2)} \right)
\end{align*} \in \mathbb{R} \quad (1141)$$

Nonnegativity of eigenvalue $\lambda_1$ is guaranteed by only nonnegativity of the $d_{ij}$ which in turn is guaranteed by matrix inequality (1133). Inequality between the eigenvalues in (1140) follows from only realness of the $d_{ij}$. Since $\lambda_1$ always equals or exceeds $\lambda_2$, conditions for positive (semi)definiteness of submatrix $T$ can be completely determined by examining $\lambda_2$ the smaller of its two eigenvalues. A triangle inequality is made apparent when we express $T$ eigenvalue nonnegativity in terms of $D$ matrix entries; 

$$T \succeq 0 \iff \det T = \lambda_1 \lambda_2 \geq 0, \ d_{12}, d_{13} \geq 0 \quad (c)$$

$$\iff \lambda_2 \geq 0 \quad (b) \quad (1142)$$

$$|\sqrt{d_{12}} - \sqrt{d_{23}}| \leq \sqrt{d_{13}} \leq \sqrt{d_{12}} + \sqrt{d_{23}} \quad (a)$$
Triangle inequality (1142a) (confer (1037) (1154)), in terms of three rooted entries from $D$, is equivalent to metric property 4

\[
\begin{align*}
\sqrt{d_{13}} & \leq \sqrt{d_{12}} + \sqrt{d_{23}} \\
\sqrt{d_{23}} & \leq \sqrt{d_{12}} + \sqrt{d_{13}} \\
\sqrt{d_{12}} & \leq \sqrt{d_{13}} + \sqrt{d_{23}}
\end{align*}
\]  

(1143)

for the corresponding points $x_1, x_2, x_3$ from some length-$N$ list.\(^{5.38}\)

### 5.8.2.1 Comment

Given $D$ whose dimension $N$ equals or exceeds 3, there are $N!/(3!(N-3)!)$ distinct triangle inequalities in total like (1037) that must be satisfied, of which each $d_{ij}$ is involved in $N-2$, and each point $x_i$ is in $(N-1)!/(2!(N-1-2)!)$). We have so far revealed only one of those triangle inequalities; namely, (1142a) that came from $T$ (1139). Yet we claim if $-V_N^T D V_N \succeq 0$ then all triangle inequalities will be satisfied simultaneously;

\[
|\sqrt{d_{ik}} - \sqrt{d_{kj}}| \leq \sqrt{d_{ij}} \leq \sqrt{d_{ik}} + \sqrt{d_{kj}}, \quad i < k < j
\]

(1144)

(There are no more.) To verify our claim, we must prove the matrix inequality $-V_N^T D V_N \succeq 0$ to be a sufficient test of all the triangle inequalities; more efficient, we mention, for larger $N$:

#### 5.8.2.1.1 Shore

The columns of $\Xi V_N \Xi_c$ hold a basis for $\mathcal{N}(1^T)$ when $\Xi_r$ and $\Xi_c$ are permutation matrices. In other words, any permutation of the rows or columns of $V_N$ leaves its range and nullspace unchanged; id est, $R(\Xi_r V_N \Xi_c) = R(V_N) = \mathcal{N}(1^T)$ (976). Hence, two distinct triangle inequalities can be equivalent tests of the positive semidefiniteness of $D$ on $R(V_N)$; id est, $-V_N^T D V_N \succeq 0 \Leftrightarrow -(\Xi_r V_N \Xi_c)^T D (\Xi_r V_N \Xi_c)^T \succeq 0$. By properly choosing permutation matrices,\(^{5.39}\) the leading principal submatrix $T_2 \in S^2$ of $-(\Xi_r V_N \Xi_c)^T D (\Xi_r V_N \Xi_c)$ may be loaded with the entries of $D$ needed to test any particular triangle inequality (similarly to (1134)-(1142)). Because all the triangle inequalities can be individually tested using a test equivalent to the lone matrix inequality $-V_N^T D V_N \succeq 0$, it logically follows that the lone matrix inequality tests all those triangle inequalities simultaneously. We conclude that $-V_N^T D V_N \succeq 0$ is a sufficient test for the fourth property of the Euclidean metric, triangle inequality.\(\top\)

#### 5.8.2.2 Strict triangle inequality

Without exception, all the inequalities in (1142) and (1143) can be made strict while their corresponding implications remain true. The then strict inequality (1142a) or (1143) may be interpreted as a strict triangle inequality under which collinear arrangement of points is

\(^{5.38}\)Accounting for symmetry property 3, the fourth metric property demands three inequalities be satisfied per one of type (1142a). The first of those inequalities in (1143) is self evident from (1142a), while the two remaining follow from the left-hand side of (1142a) and the fact for scalars, $|a| \leq b \Leftrightarrow a \leq b$ and $-a \leq b$.

\(^{5.39}\)To individually test triangle inequality $|\sqrt{d_{ik}} - \sqrt{d_{kj}}| \leq \sqrt{d_{ij}} \leq \sqrt{d_{ik}} + \sqrt{d_{kj}}$ for particular $i, k, j$, set $\Xi_r(i, 1) = \Xi_r(k, 2) = \Xi_r(j, 3) = 1$ and $\Xi_c = I$.\(\top\)
not allowed. [233, §24/6, p.322] Hence by similar reasoning, \(-V_X^T D V_N > 0\) is a sufficient test of all the strict triangle inequalities; \textit{id est},

\[
\begin{align*}
\delta(D) &= 0 \\
D^T &= D \\
- V_X^T D V_N > 0
\end{align*}
\]

\Rightarrow \sqrt{d_{ij}} < \sqrt{d_{ik}} + \sqrt{d_{kj}}, \quad i \neq j \neq k \quad (1145)

5.8.3 \(-V_X^T D V_N\) nesting

From (1139) observe that \(T = -V_X^T D V_N|_{|N-1}\). In fact, for \(D \in \text{EDM}_N\), the leading principal submatrices of \(-V_X^T D V_N\) form a nested sequence (by inclusion) whose members are individually positive semidefinite [168] [212] [341] and have the same form as \(T\); \textit{videlicet},\(^5\)

\[-V_X^T D V_N|_{|N-1} = [\emptyset] \quad (o)\]

\[-V_X^T D V_N|_{|N-2} = [d_{12}] \in S_+ \quad (a)\]

\[-V_X^T D V_N|_{|N-3} = \begin{bmatrix}
d_{12} & \frac{1}{2}(d_{12} + d_{13} - d_{23}) \\
\frac{1}{2}(d_{12} + d_{13} - d_{23}) & d_{13}
\end{bmatrix} = T \in S_+^2 \quad (b)\]

\[-V_X^T D V_N|_{|N-4} = \begin{bmatrix}
d_{12} & \frac{1}{2}(d_{12} + d_{13} - d_{23}) & \frac{1}{2}(d_{12} + d_{14} - d_{24}) \\
\frac{1}{2}(d_{12} + d_{13} - d_{23}) & d_{13} & \frac{1}{2}(d_{13} + d_{14} - d_{34}) \\
\frac{1}{2}(d_{12} + d_{14} - d_{24}) & \frac{1}{2}(d_{13} + d_{14} - d_{34}) & d_{14}
\end{bmatrix} \quad (c)\]

\[
\vdots
\]

\[-V_X^T D V_N|_{|N-i} = \begin{bmatrix}
- V_X^T D V_N|_{|N-i-1} & \nu(i) \\
\nu(i)^T & d_{1i}
\end{bmatrix} \in S_{+}^{i-1} \quad (d)\]

\[
\vdots
\]

\[-V_X^T D V_N = \begin{bmatrix}
- V_X^T D V_N|_{|N-N-1} & \nu(N) \\
\nu(N)^T & d_{1N}
\end{bmatrix} \in S_{+}^{N-1} \quad (e)\]

where

\[
\nu(i) \triangleq \frac{1}{2} \begin{bmatrix}
d_{12} + d_{1i} - d_{2i} \\
d_{13} + d_{1i} - d_{3i} \\
\vdots \\
d_{1,i-1} + d_{1i} - d_{i-1,i}
\end{bmatrix} \in \mathbb{R}^{i-2}, \quad i > 2 \quad (1147)
\]

Hence, the leading principal submatrices of EDM \(D\) must also be EDMs.\(^5\)\(^4\)

\(^5\)\(^4\) \(-V_D V|_{|N-1} = 0 \in S_+^0\) (§B.4.1)

\(^5\)\(^4\) In fact, each and every principal submatrix of an EDM \(D\) is another EDM. [245, §4.1]
Bordered symmetric matrices in the form (1146) are known to have intertwined \[\text{eigenvalues}; (\text{confer } \S 5.11.1)\] that means, for the particular submatrices (1146a) and (1146b),

\[\lambda_2 \leq d_{12} \leq \lambda_1 \tag{1148}\]

where \(d_{12}\) is the eigenvalue of submatrix (1146a) and \(\lambda_1, \lambda_2\) are the eigenvalues of \(T\) (1146b). Intertwining in (1148) predicts that should \(d_{12}\) become 0, then \(\lambda_2\) must go to 0. The eigenvalues are similarly intertwined for submatrices (1146b) and (1146c);

\[\gamma_3 \leq \lambda_2 \leq \gamma_2 \leq \lambda_1 \leq \gamma_1 \tag{1149}\]

where \(\gamma_1, \gamma_2, \gamma_3\) are the eigenvalues of submatrix (1146c). Intertwining likewise predicts that should \(\lambda_2\) become 0 (a possibility revealed in \(\S 5.8.3.1\)), then \(\gamma_3\) must go to 0.

The preceding logic extends by induction through the remaining members of the sequence (1146).

\subsection*{5.8.3.1 Tightening the triangle inequality}

Now we apply Schur complement from \(\S A.4\) to tighten the triangle inequality from (1138) in case: cardinality \(N = 4\). We find that the gains by doing so are modest. From (1146) we identify:

\[
\begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} \triangleq -V_N^T D V_N|_{N=4} \\
A \triangleq T = -V_N^T D V_N|_{N=3}
\]

both positive semidefinite by assumption, where \(B = \nu(4)\) (1147), and \(C = d_{14}\). Using nonstrict \(CC^\dagger\)-form (1593), \(C \succeq 0\) by assumption (\(\S 5.8.1\)) and \(CC^\dagger = I\). So by the positive semidefinite ordering of eigenvalues theorem (\(\S A.3.1.0.1\)),

\[-V_N^T D V_N|_{N=4} \succeq 0 \iff T \succeq d_{14}^{-1} \nu(4) \nu(4)^T \Rightarrow \begin{cases}
\lambda_1 \geq d_{14}^{-1} \|
u(4)\|^2 \\
\lambda_2 \geq 0
\end{cases}\]

where \(\{d_{14}^{-1} \|
u(4)\|^2, 0\}\) are the eigenvalues of \(d_{14}^{-1} \nu(4) \nu(4)^T\) while \(\lambda_1, \lambda_2\) are the eigenvalues of \(T\).

\subsection*{5.8.3.1.1 Example. Small completion problem, II.}

Applying the inequality for \(\lambda_1\) in (1153) to the small completion problem on page 358 Figure 128, the lower bound on \(\sqrt{d_{14}}\) (1.236 in (962)) is tightened to 1.289. The correct value of \(\sqrt{d_{14}}\) to three significant figures is 1.414.

\footnote{If \(d_{12}\) were 0, eigenvalue \(\lambda_2\) becomes 0 (1141) because \(d_{13}\) must then be equal to \(d_{23}\); \textit{id est}, \(d_{12} = 0 \iff x_1 = x_2\). \(\S 5.4\)}
5.8.4 Affine dimension reduction in two dimensions

(Confer §5.14.4) The leading principal $2 \times 2$ submatrix $T$ of $-V_N^T D V_N$ has largest eigenvalue $\lambda_1$ (1141) which is a convex function of $D$. $\lambda_1$ can never be 0 unless $d_{12} = d_{13} = d_{23} = 0$. Eigenvalue $\lambda_1$ can never be negative while the $d_{ij}$ are nonnegative.

The remaining eigenvalue $\lambda_2$ (1141) is a concave function of $D$ that becomes 0 only at the upper and lower bounds of triangle inequality (1142a) and its equivalent forms:

\[
\begin{align*}
|\sqrt{d_{12}} - \sqrt{d_{23}}| &\leq \sqrt{d_{13}} \leq \sqrt{d_{12}} + \sqrt{d_{23}} \\
|\sqrt{d_{12}} - \sqrt{d_{13}}| &\leq \sqrt{d_{23}} \leq \sqrt{d_{12}} + \sqrt{d_{13}} \\
|\sqrt{d_{13}} - \sqrt{d_{23}}| &\leq \sqrt{d_{12}} \leq \sqrt{d_{13}} + \sqrt{d_{23}}
\end{align*}
\] (1154)

In between those bounds, $\lambda_2$ is strictly positive; otherwise, it would be negative but prevented by the condition $T \succeq 0$.

When $\lambda_2$ becomes 0, it means triangle $\triangle_{123}$ has collapsed to a line segment; a potential reduction in affine dimension $r$. The same logic is valid for any particular principal $2 \times 2$ submatrix of $-V_N^T D V_N$, hence applicable to other triangles.

5.9 Bridge: Convex polyhedra to EDMs

The criteria for the existence of an EDM include, by definition (969) (1039), the properties imposed upon its entries $d_{ij}$ by the Euclidean metric. From §5.8.1 and §5.8.2, we know there is a relationship of matrix criteria to those properties. Here is a snapshot of what we are sure:

\[
\begin{align*}
\sqrt{d_{ij}} &\geq 0, \ i \neq j \\
\sqrt{d_{ij}} &\leq \sqrt{d_{ik}} + \sqrt{d_{kj}}, \ i \neq j \neq k \\
\end{align*}
\]

all implied by $D \in \text{EDM}^N$. In words, these four Euclidean metric properties are necessary conditions for $D$ to be a distance matrix. At the moment, we have no converse. As of concern in §5.3, we have yet to establish metric requirements beyond the four Euclidean metric properties that would allow $D$ to be certified an EDM or might facilitate polyhedron or list reconstruction from an incomplete EDM. We deal with this problem in §5.14. Our present goal is to establish \textit{ab initio} the necessary and sufficient matrix criteria that will

---

**Footnotes:**

- The largest eigenvalue of any symmetric matrix is always a convex function of its entries, while the smallest eigenvalue is always concave. [63, exmp.3.10] In our particular case, say $d \triangleq \begin{bmatrix} d_{12} \\ d_{13} \\ d_{23} \end{bmatrix} \in \mathbb{R}^3$. Then the Hessian (1853) $\nabla^2 \lambda_1(d) \succeq 0$ certifies convexity whereas $\nabla^2 \lambda_2(d) \preceq 0$ certifies concavity. Each Hessian has rank equal to 1. The respective gradients $\nabla \lambda_1(d)$ and $\nabla \lambda_2(d)$ are nowhere 0 and can be uniquely defined.
subsume all the Euclidean metric properties and any further requirements\(^5\text{.}4^4\) for all \(N > 1\) (§5.8.3); \(\text{id est}\),

\[
-DV_N^TDV_N \geq 0 \quad \text{for all} \quad D \in S_N^+ \quad \iff D \in \text{EDM}^N
\]  

(988)

or for EDM definition (1048),

\[
\sqrt{\delta(d)} \geq 0 \quad \iff D = D(\Omega, d) \in \text{EDM}^N
\]  

(1156)

5.9.1 Geometric arguments

5.9.1.0.1 Definition. Elliptope:

[248] [245, §2.3] [119, §31.5]
a unique bounded immutable convex Euclidean body in \(S^n\); intersection of positive semidefinite cone \(S_+^n\) with that set of \(n\) hyperplanes defined by unity main diagonal;

\[
\mathcal{E}^n \triangleq S_+^n \cap \{ \Phi \in S^n \mid \delta(\Phi) = 1 \}
\]  

(1157)

\(\text{a.k.a the set of all correlation matrices of dimension}\)

\[
\dim \mathcal{E}^n = n(n-1)/2 \quad \text{in } \mathbb{R}^{n(n+1)/2}
\]  

(1158)

An elliptope \(\mathcal{E}^n\) is not a polyhedron, in general, but has some polyhedral faces and an infinity of vertices.\(^5\text{.}4^5\) Of those, \(2^{n-1}\) vertices (some extreme points of the elliptope) are extreme directions \(yy^T\) of the positive semidefinite cone, where the entries of vector \(y \in \mathbb{R}^n\) belong to \(\{\pm 1\}\) and exercise every combination. Each of the remaining vertices has rank, greater than one, belonging to the set \(\{k > 0 \mid k(k+1)/2 \leq n\}\). Each and every face of an elliptope is exposed. \(\triangle\)

In fact, any positive semidefinite matrix whose entries belong to \(\{\pm 1\}\) is a rank-one correlation matrix; and \textit{vice versa}:\(^5\text{.}4^6\)

5.9.1.0.2 Theorem. Elliptope vertices rank-one.

[132, §2.1.1]

For \(Y \in \mathbb{S}^n\), \(y \in \mathbb{R}^n\), and all \(i, j \in \{1 \ldots n\}\) (confer §2.3.1.0.1)

\[
Y \succeq 0, \quad Y_{ij} \in \{\pm 1\} \quad \iff Y = yy^T, \quad y_i \in \{\pm 1\}
\]  

(1159)

\(\Diamond\)

\(^5\text{.}4^4\)In 1935, Schoenberg [323, (1)] first extolled matrix product \(-V_N^TDV_N\) (1134) (predicated on symmetry and selfdistance) specifically incorporating \(V_N\), albeit algebraically. He showed: nonnegativity \(-y^TV_N^TDV_Ny \geq 0\), for all \(y \in \mathbb{R}^{n-1}\), is necessary and sufficient for \(D\) to be an EDM. Gower [172, §3] remarks how surprising it is that such a fundamental property of Euclidean geometry was obtained so late.

\(^5\text{.}4^5\)Laurent defines vertex distinctly from the sense herein (§2.6.1.0.1); she defines vertex as a point with full-dimensional (nonempty interior) normal cone (§E.10.3.2.1). Her definition excludes point C in Figure 34, for example.

\(^5\text{.}4^6\)As there are few equivalent conditions for rank constraints, this device is rather important for relaxing integer, combinatorial, or Boolean problems.
Figure 142: Elliptope $E^3$ in isometrically isomorphic $\mathbb{R}^6$ (projected on $\mathbb{R}^3$) is a convex body that appears to possess some kind of symmetry in this dimension; it resembles a malformed pillow in the shape of a bulging tetrahedron. Elliptope relative boundary is not smooth and comprises all set members (1157) having at least one 0 eigenvalue. [248, §2.1] This elliptope has an infinity of vertices, but there are only four vertices corresponding to a rank-1 matrix. Those $yy^\top$, evident in the illustration, have binary vector $y \in \mathbb{R}^3$ with entries in $\{\pm 1\}$. 
5.9. BRIDGE: CONVEX POLYHEDRA TO EDMS

Figure 143: Elliptope $\mathcal{E}^2$ in isometrically isomorphic $\mathbb{R}^3$ is a line segment illustrated interior to positive semidefinite cone $\mathbb{S}^2_+$ (Figure 45).

The elliptope for dimension $n=2$ is a line segment in isometrically isomorphic $\mathbb{R}^{n(n+1)/2}$ (Figure 143). Obviously, $\text{cone}(\mathcal{E}^n) \neq \mathbb{S}^n_+$. The elliptope for dimension $n=3$ is realized in Figure 142.

5.9.1.0.3 Lemma. Hypersphere. (confer bullet p.366) [18, §4]
Matrix $\Psi = [\Psi_{ij}] \in \mathbb{S}^N$ belongs to the elliptope in $\mathbb{S}^N$ iff there exist $N$ points $p$ on the boundary of a hypersphere in $\mathbb{R}^{\text{rank } \Psi}$ having radius 1 such that

$$\|p_i - p_j\|^2 = 2(1 - \Psi_{ij}), \quad i, j = 1 \ldots N$$

(1160)

There is a similar theorem for Euclidean distance matrices:
We derive matrix criteria for $D$ to be an EDM, validating (988) using simple geometry; distance to the polyhedron formed by the convex hull of a list of points (76) in Euclidean space $\mathbb{R}^n$.

5.9.1.0.4 EDM assertion.
$D$ is a Euclidean distance matrix if and only if $D \in \mathbb{S}^N_+$ and distances-square from the origin

$$\{\|p(y)\|^2 = -y^T V_N^T D V_N y \mid y \in \mathcal{S} - \beta\}$$

(1161)
correspond to points $p$ in some bounded convex polyhedron

$$\mathcal{P} - \alpha = \{p(y) \mid y \in \mathcal{S} - \beta\}$$

(1162)

having $N$ or fewer vertices embedded in an $r$-dimensional subspace $\mathcal{A} - \alpha$ of $\mathbb{R}^n$, where $\alpha \in \mathcal{A} = \text{aff } \mathcal{P}$ and where domain of linear surjection $p(y)$ is the unit simplex $\mathcal{S} \subset \mathbb{R}^{N-1}_+$ shifted such that its vertex at the origin is translated to $-\beta$ in $\mathbb{R}^{N-1}$. When $\beta = 0$, then $\alpha = x_1$. 

$\diamond$
In terms of $V_N$, the unit simplex (292) in $\mathbb{R}^{N-1}$ has an equivalent representation:

$$S = \{ s \in \mathbb{R}^{N-1} \mid \sqrt{2}V_N s \geq -e_1 \}$$  \hspace{1cm} (1163)

where $e_1$ is as in (985). Incidental to the EDM assertion, shifting the unit-simplex domain in $\mathbb{R}^{N-1}$ translates the polyhedron $P$ in $\mathbb{R}^n$. Indeed, there is a map from vertices of the unit simplex to members of the list generating $P$;

$$p : \mathbb{R}^{N-1} \rightarrow \mathbb{R}^n$$

$$p \begin{pmatrix} \ -\beta & \ e_1 - \beta & \ e_2 - \beta & \ \vdots & \ e_{N-1} - \beta \end{pmatrix} = \begin{pmatrix} x_1 - \alpha \\ x_2 - \alpha \\ x_3 - \alpha \\ \vdots \\ x_N - \alpha \end{pmatrix}$$  \hspace{1cm} (1164)

5.9.1.0.5 Proof. EDM assertion.

($\Rightarrow$) We demonstrate that if $D$ is an EDM, then each distance-square $\|p(y)\|^2$ described by (1161) corresponds to a point $p$ in some embedded polyhedron $P - \alpha$. Assume $D$ is indeed an EDM; id est, $D$ can be made from some list $X$ of $N$ unknown points in Euclidean space $\mathbb{R}^n$; $D = D(X)$ for $X \in \mathbb{R}^{n \times N}$ as in (969). Since $D$ is translation invariant ($\S\S 5.5.1$), we may shift the affine hull $A$ of those unknown points to the origin as in (1104). Then take any point $p$ in their convex hull (86);

$$P - \alpha = \{ p = (X - Xb)1^T a \mid a^T 1 = 1, \ a \succeq 0 \}$$  \hspace{1cm} (1165)

where $\alpha = Xb \in A \Leftrightarrow b^T 1 = 1$. Solutions to $a^T 1 = 1$ are:

$$a \in \left\{ e_1 + \sqrt{2}V_N s \mid s \in \mathbb{R}^{N-1} \right\}$$  \hspace{1cm} (1166)

where $e_1$ is as in (985). Similarly, $b = e_1 + \sqrt{2}V_N \beta$.

$$P - \alpha = \{ p = X(I - (e_1 + \sqrt{2}V_N \beta)1^T)(e_1 + \sqrt{2}V_N s) \mid \sqrt{2}V_N s \geq -e_1 \}$$

$$= \{ p = X\sqrt{2}V_N (s - \beta) \mid \sqrt{2}V_N s \geq -e_1 \}$$  \hspace{1cm} (1167)

that describes the domain of $p(s)$ as the unit simplex

$$S = \{ s \mid \sqrt{2}V_N s \geq -e_1 \} \subset \mathbb{R}^{N-1}_+$$  \hspace{1cm} (1163)

Making the substitution $s - \beta \leftarrow y$

$$P - \alpha = \{ p = X\sqrt{2}V_N y \mid y \in S - \beta \}$$  \hspace{1cm} (1168)

Point $p$ belongs to a convex polyhedron $P - \alpha$ embedded in an $r$-dimensional subspace of $\mathbb{R}^n$ because the convex hull of any list forms a polyhedron, and because the translated

\textsuperscript{5.47}Since $R(V_N) = N(1^T)$ and $N(1^T)^T \perp R(1^T)$, then over all $s \in \mathbb{R}^{N-1}$, $V_N s$ is a hyperplane through the origin orthogonal to $1$. Thus the solutions $\{ a \}$ constitute a hyperplane orthogonal to the vector $1$, and offset from the origin in $\mathbb{R}^N$ by any particular solution; in this case, $a = e_1$.\textsuperscript{5.47}
affine hull $A - \alpha$ contains the translated polyhedron $P - \alpha$ (1107) and the origin (when $\alpha \in A$), and because $A$ has dimension $r$ by definition (1109). Now, any distance-square from the origin to the polyhedron $P - \alpha$ can be formulated

\[ \{ p^T p = \|p\|^2 = 2y^T V_N^T X^T X V_N y \mid y \in S - \beta \} \quad (1169) \]

Applying (1116) to (1169) we get (1161).

$(\Leftarrow)$ To validate the EDM assertion in the reverse direction, we prove: If each distance-square $\|p(y)\|^2$ (1161) on the shifted unit-simplex $S - \beta \subset \mathbb{R}^{N-1}$ corresponds to a point $p(y)$ in some embedded polyhedron $P - \alpha$, then $D$ is an EDM. The $r$-dimensional subspace $A - \alpha \subseteq \mathbb{R}^n$ is spanned by

\[ p(S - \beta) = P - \alpha \quad (1170) \]

because $A - \alpha = \text{aff}(P - \alpha) \supseteq P - \alpha$ (1107). So, outside domain $S - \beta$ of linear surjection $p(y)$, simplex complement $\setminus S - \beta \subset \mathbb{R}^{N-1}$ must contain domain of the distance-square $\|p(y)\|^2 = p(y)^T p(y)$ to remaining points in subspace $A - \alpha$; id est, to the polyhedron’s relative exterior $\setminus P - \alpha$. For $\|p(y)\|^2$ to be nonnegative on the entire subspace $A - \alpha$, $-V_N^T D V_N$ must be positive semidefinite and is assumed symmetric:

\[ -V_N^T D V_N \triangleq \Theta_p^T \Theta_p \quad (1171) \]

where $\Theta_p \in \mathbb{R}^{m \times N-1}$ for some $m \geq r$. Because $p(S - \beta)$ is a convex polyhedron, it is necessarily a set of linear combinations of points from some length-$N$ list because every convex polyhedron having $N$ or fewer vertices can be generated that way (§2.12.2). Equivalent to (1161) are

\[ \{ p^T p \mid p \in P - \alpha \} = \{ p^T p = y^T \Theta_p^T \Theta_p y \mid y \in S - \beta \} \quad (1172) \]

Because $p \in P - \alpha$ may be found by factoring (1172), the list $\Theta_p$ is found by factoring (1171). A unique EDM can be made from that list using inner-product form definition $D(\Theta)|_{\Theta = \Theta_p}$ (1039). That EDM will be identical to $D$ if $\delta(D) = 0$, by injectivity of $D$ (1090).

\[ \downarrow \]

5.9.2 Necessity and sufficiency

From (1132) we learned that matrix inequality $-V_N^T D V_N \succeq 0$ is a necessary test for $D$ to be an EDM. In §5.9.1, the connection between convex polyhedra and EDMs was pronounced by the EDM assertion; the matrix inequality together with $D \in \mathbb{S}_h^N$ became a sufficient test when the EDM assertion demanded that every bounded convex polyhedron have a corresponding EDM. For all $N>1$ (§5.8.3), the matrix criteria for the existence of an EDM in (988), (1156), and (964) are therefore necessary and sufficient and subsume all the Euclidean metric properties and further requirements.

\[ ^{5.49} \text{The antisymmetric part } (-V_N^T D V_N - (-V_N^T D V_N)^T)/2 \text{ is annihilated by } \|p(y)\|^2. \text{ By the same reasoning, any positive (semi)definite matrix } A \text{ is generally assumed symmetric because only the symmetric part } (A + A^T)/2 \text{ survives the test } y^T A y \geq 0. \]  

\[ ^{5.48} A^T = A \succeq 0 \iff A = R^T R \text{ for some real matrix } R. \]  

\[ ^{341} \text{§6.3] } \]
5.9.3 Example revisited

Now we apply the necessary and sufficient EDM criteria (988) to an earlier problem.

5.9.3.0.1 Example. Small completion problem, III. (confer §5.8.3.1.1)

Continuing Example 5.3.0.0.2 pertaining to Figure 128 where $N = 4$, distance-square $d_{14}$ is ascertainable from the matrix inequality $-V_N^T D V_N \succeq 0$. Because all distances in (961) are known except $\sqrt{d_{14}}$, we may simply calculate the smallest eigenvalue of $-V_N^T D V_N$ over a range of $d_{14}$ as in Figure 144. We observe a unique value of $d_{14}$ satisfying (988) where the abscissa axis is tangent to the hypograph of the smallest eigenvalue. Since the smallest eigenvalue of a symmetric matrix is known to be a concave function (§5.8.4), we calculate its second partial derivative with respect to $d_{14}$ evaluated at 2 and find $-1/3$. We conclude there are no other satisfying values of $d_{14}$. Further, that value of $d_{14}$ does not meet an upper or lower bound of a triangle inequality like (1144), so neither does it cause the collapse of any triangle. Because the smallest eigenvalue is 0, affine dimension $r$ of any point list corresponding to $D$ cannot exceed $N - 2$. (§5.7.1.1) □

5.10 EDM-entry composition

Laurent [245, §2.3] applies results from Schoenberg, 1938 [324], to show certain nonlinear compositions of individual EDM entries yield EDMs; in particular,

$$D \in \text{EDM}_N \iff [1 - e^{-\alpha d_{ij}}] \in \text{EDM}_N \quad \forall \alpha > 0 \quad (a)$$

$$\iff [e^{-\alpha d_{ij}}] \in \mathcal{E}^N \quad \forall \alpha > 0 \quad (b)$$

where $D = [d_{ij}]$ and $\mathcal{E}^N$ is the elliptope (1157).

5.10.0.0.1 Proof. (Monique Laurent, 2003) [324] (confer [237])

Lemma 2.1. from A Tour d’Horizon . . . on Completion Problems. [245]

For $D = [d_{ij}, i, j = 1 \ldots N] \in \mathbb{S}_h^N$ and $\mathcal{E}^N$ the elliptope in $\mathbb{S}^N$ (§5.9.1.0.1), the following assertions are equivalent:

(i) $D \in \text{EDM}_N$

(ii) $e^{-\alpha D} \triangleq [e^{-\alpha d_{ij}}] \in \mathcal{E}^N$ for all $\alpha > 0$

(iii) $11^T - e^{-\alpha D} \triangleq [1 - e^{-\alpha d_{ij}}] \in \text{EDM}_N$ for all $\alpha > 0$ ◇

1) Equivalence of Lemma 2.1 (i) (ii) is stated in Schoenberg’s Theorem 1 [324, p.527].
5.10. EDM-ENTRY COMPOSITION

Figure 144: Smallest eigenvalue of $-V_N^T D V_N$ makes it a PSD matrix for only one value of $d_{14}$: 2.

Figure 145: Some entrywise EDM compositions: (a) $\alpha = 2$. Concave nondecreasing in $d_{ij}$. (b) Trajectory convergence in $\alpha$ for $d_{ij} = 2$. 
2) (ii) ⇒ (iii) can be seen from the statement in the beginning of section 3, saying that a distance space embeds in $L_2$ iff some associated matrix is PSD. We reformulate it:

Let $d = (d_{ij})_{i,j=0,1,...,N}$ be a distance space on $N+1$ points (i.e., symmetric hollow matrix of order $N+1$) and let $p = (p_{ij})_{i,j=1,...,N}$ be the symmetric matrix of order $N$ related by:

(A) $2p_{ij} = d_{0i} + d_{0j} - d_{ij}$ for $i, j = 1, \ldots, N$ or equivalently

(B) $d_{0i} = p_{ii}$, $d_{ij} = p_{ii} + p_{jj} - 2p_{ij}$ for $i, j = 1, \ldots, N$

Then $d$ embeds in $L_2$ iff $p$ is a positive semidefinite matrix iff $d$ is of negative type (second half page 525/top of page 526 in [324]).

For the implication from (ii) to (iii), set: $p = e^{-\alpha d}$ and define $d'$ from $p$ using (B) above. Then $d'$ is a distance space on $N+1$ points that embeds in $L_2$. Thus its subspace of $N$ points also embeds in $L_2$ and is precisely $1 - e^{-\alpha d}$.

Note that (iii) ⇒ (ii) cannot be read immediately from this argument since (iii) involves the subdistance of $d'$ on $N$ points (and not the full $d'$ on $N+1$ points).

3) Show (iii) ⇒ (i) by using the series expansion of the function $1 - e^{-\alpha d}$: the constant term cancels, $\alpha$ factors out; there remains a summation of $d$ plus a multiple of $\alpha$. Letting $\alpha$ go to 0 gives the result.

This is not explicitly written in Schoenberg, but he also uses such an argument; expansion of the exponential function then $\alpha \to 0$ (first proof on [324, p.526]).

Schoenberg’s results [324, §6 thm.5] (confer [237, p.108-109]) also suggest certain finite positive roots of EDM entries produce EDMs; specifically,

$$D \in \text{EDM}^N \iff [d_{ij}^{1/\alpha}] \in \text{EDM}^N \forall \alpha > 1$$  (1174)

The special case $\alpha = 2$ is of interest because it corresponds to absolute distance; e.g.,

$$D \in \text{EDM}^N \Rightarrow \sqrt{D} \in \text{EDM}^N$$  (1175)

Assuming that points constituting a corresponding list $X$ are distinct (1136), then it follows: for $D \in S^N_+$

$$\lim_{\alpha \to \infty} [d_{ij}^{1/\alpha}] = \lim_{\alpha \to \infty} [1 - e^{-\alpha d_{ij}}] = -E \equiv 11^T - I$$  (1176)

Negative elementary matrix $-E$ (§B.3) is: relatively interior to the EDM cone (§6.5), on its axis, and terminal to respective trajectories (1173a) and (1174) as functions of $\alpha$. Both trajectories are confined to the EDM cone; in engineering terms, the EDM cone is an invariant set [321] to either trajectory. Further, if $D$ is not an EDM but for some particular $\alpha_p$ it becomes an EDM, then for all greater values of $\alpha$ it remains an EDM.

5.10.0.0.2 Exercise. Concave nondecreasing EDM-entry composition.
Given EDM $D = [d_{ij}]$, empirical evidence suggests that the composition $[\log_2(1 + d_{ij}^{1/\alpha})]$ is also an EDM for each fixed $\alpha \geq 1$ [sic]. Its concavity in $d_{ij}$ is illustrated in Figure 145 together with functions from (1173a) and (1174). Prove whether it holds more generally: Any concave nondecreasing composition of individual EDM entries $d_{ij}$ on $\mathbb{R}_+$ produces another EDM.
5.11. EDM INDEFINENESS

5.10.0.0.3 Exercise. Taxicab distance matrix as EDM.
Determine whether taxicab distance matrices \((D_1(X)\) in Example 3.7.0.0.2) are all numerically equivalent to EDMs. Explain why or why not.

5.10.1 EDM by elliptope
(\textit{confer (995)}) For some \(\kappa \in \mathbb{R}_+\) and \(C \in \mathbb{S}^N_+\) in elliptope \(E^N\) (§5.9.1.0.1), Alfakih asserts:
any given EDM \(D\) is expressible \([10] [119, \S 31.5]\)
\[
D = \kappa (11^T - C) \in \text{EDM}^N
\]  
(1177)
This expression exhibits nonlinear combination of variables \(\kappa\) and \(C\). We therefore propose a different expression requiring redefinition of the elliptope (1157) by scalar parametrization;
\[
E^N_t \triangleq \mathbb{S}^N_+ \cap \{\Phi \in \mathbb{S}^N \mid \delta(\Phi) = t1\}
\]  
(1178)
where, of course, \(E^n = E^N_t\). Then any given EDM \(D\) is expressible
\[
D = t11^T - \mathcal{E} \in \text{EDM}^N
\]  
(1179)
which is linear in variables \(t \in \mathbb{R}_+\) and \(\mathcal{E} \in E^N_t\).

5.11 EDM indefiniteness
By known result (§A.7.2) regarding a 0-valued entry on the main diagonal of a symmetric positive semidefinite matrix, there can be no positive or negative semidefinite EDM except the 0 matrix because \(\text{EDM}^N \subseteq \mathbb{S}^N_+\) (968) and
\[
\mathbb{S}^N_+ \cap \mathbb{S}^N_- = 0
\]  
(1180)
the origin. So when \(D \in \text{EDM}^N\), there can be no factorization \(D = A^T A\) or \(-D = A^T A\). [341, \S 6.3] Hence eigenvalues of an EDM are neither all nonnegative or all nonpositive; an EDM is indefinite and possibly invertible.

5.11.1 EDM eigenvalues, congruence transformation
For any symmetric \(-D\), we can characterize its eigenvalues by congruence transformation:
[341, \S 6.3]
\[
-W^T DW = - \begin{bmatrix} V_N^T & 1^T \end{bmatrix} D [V_N \ 1] = - \begin{bmatrix} V_N^T D V_N & V_N^T D 1 \\ 1^T D V_N & 1^T D 1 \end{bmatrix} \in \mathbb{S}^N
\]  
(1181)
Because
\[
W \triangleq [V_N \ 1] \in \mathbb{R}^{N \times N}
\]  
(1182)
is full-rank, then (1598)
i\text{inertia}(-D) = \text{inertia}(-W^T DW)
(1183)
the congruence (1181) has the same number of positive, zero, and negative eigenvalues as \(-D\). Further, if we denote by \(\{\gamma_i, \, i = 1 \ldots N-1\}\) the eigenvalues of \(-V_N^T D V_N\) and denote eigenvalues of the congruence \(-W^T D W\) by \(\{\zeta_i, \, i = 1 \ldots N\}\) and if we arrange each respective set of eigenvalues in nonincreasing order, then by theory of interlacing eigenvalues for bordered symmetric matrices [212, §4.3] [341, §6.4] [338, §IV.4.1]

\[
\begin{align*}
\zeta_N &\le \gamma_{N-1} \le \zeta_{N-1} \le \gamma_{N-2} \le \cdots \le \gamma_2 \le \zeta_2 \le \gamma_1 \le \zeta_1 
\end{align*}
\]

When \(D \in \text{EDM}^N\), then \(\gamma_i \ge 0 \, \forall \, i \) (1532) because \(-V_N^T D V_N \succeq 0\) as we know. That means the congruence must have \(N-1\) nonnegative eigenvalues; \(\zeta_i \ge 0, \, i = 1 \ldots N-1\). The remaining eigenvalue \(\zeta_N\) cannot be nonnegative because then \(-D\) would be positive semidefinite, an impossibility; so \(\zeta_N < 0\). By congruence, nontrivial \(-D\) must therefore have exactly one negative eigenvalue;\(^{5.50}\) [119, §2.4.5]

\[
D \in \text{EDM}^N \Rightarrow \begin{cases} 
\lambda(-D)_i \ge 0, \, i = 1 \ldots N-1 \\
\sum_{i=1}^{N} \lambda(-D)_i = 0 \\
D \in S_h^N \cap R^{N \times N}_+
\end{cases}
\]

(1185)

where the \(\lambda(-D)_i\) are nonincreasingly ordered eigenvalues of \(-D\) whose sum must be 0 only because \(\text{tr} \, D = 0\) [341, §5.1]. The eigenvalue summation condition, therefore, can be considered redundant. Even so, all these conditions are insufficient to determine whether some given \(H \in S_h^N\) is an EDM, as shown by counterexample.\(^{5.51}\)

**5.11.1.0.1 Exercise. Spectral inequality.**

Prove whether it holds: for \(D = [d_{ij}] \in \text{EDM}^N\)

\[
\lambda(-D)_1 \ge d_{ij} \ge \lambda(-D)_{N-1} \quad \forall \, i \neq j
\]

(1186)

Terminology: a *spectral cone* is a convex cone containing all eigenspectra [237, p.365] [338, p.26] corresponding to some set of matrices. Any positive semidefinite matrix, for example, possesses a vector of nonnegative eigenvalues corresponding to an eigenspectrum contained in a spectral cone that is a nonnegative orthant.

\(^{5.50}\)All entries of the corresponding eigenvector must have the same sign, with respect to each other, [92, p.116] because that eigenvector is the *Perron vector* corresponding to *spectral radius*; [212, §8.3.1] the predominant characteristic of square nonnegative matrices. Unlike positive semidefinite matrices, nonnegative matrices are guaranteed only to have at least one nonnegative eigenvalue.

\(^{5.51}\)When \(N = 3\), for example, the symmetric hollow matrix

\[
H = \begin{bmatrix}
0 & 1 & 1 \\
1 & 0 & 5 \\
1 & 5 & 0
\end{bmatrix} \in S_h^N \cap R^{N \times N}_+
\]

is not an EDM, although \(\lambda(-H) = [5 \quad 0.3723 \quad -5.3723]^T\) conforms to (1185).
5.11. EDM INDEFINITENESS

5.11.2 Spectral cones

Denoting the eigenvalues of Cayley-Menger matrix \( \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \in \mathbb{S}^{N+1} \) by

\[
\lambda \left( \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \right) \in \mathbb{R}^{N+1}
\]

(1187)

we have the Cayley-Menger form (§5.7.3.0.1) of necessary and sufficient conditions for \( D \in \text{EDM}^N \) from the literature: [194, §3] [76, §3] [119, §6.2] (confer (988) (964))

\[
D \in \text{EDM}^N \iff \left\{ \lambda \left( \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \right) \geq 0, \quad i = 1 \ldots N \right\} \iff \left\{ -V_N^T D V_N \succeq 0 \right\} \quad D \in \mathbb{S}_h^N
\]

(1188)

These conditions say the Cayley-Menger form has one and only one negative eigenvalue. When \( D \) is an EDM, eigenvalues \( \lambda \left( \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} \right) \) belong to that particular orthant in \( \mathbb{R}^{N+1} \) having the \( N+1 \)th coordinate as sole negative coordinate\( ^{5.53} \):

\[
\begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} = \text{cone} \{ e_1, e_2, \ldots, e_N, -e_{N+1} \}
\]

(1189)

5.11.2.1 Cayley-Menger versus Schoenberg

Connection to the Schoenberg criterion (988) is made when the Cayley-Menger form is further partitioned:

\[
\begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1^T \\ -D_{1,2:N} \end{bmatrix} \begin{bmatrix} -D_{1,2:N} \\ -D_{2,2:N} \end{bmatrix}
\]

(1190)

Matrix \( D \in \mathbb{S}_h^N \) is an EDM if and only if the Schur complement of \( \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \) (§A.4) in this partition is positive semidefinite; [18, §1] [227, §3] id est, (confer (1134))

\[
D \iff -D_{2,2:N} - \begin{bmatrix} 1 & -D_{2,2:N} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1^T \\ -D_{1,2:N} \end{bmatrix} = -2V_N^T D V_N \succeq 0
\]

and

\[
D \in \mathbb{S}_h^N
\]

(1191)

Positive semidefiniteness of that Schur complement insures nonnegativity (\( D \in \mathbb{R}_+^{N \times N} \), §5.8.1), whereas complementary inertia (1600) insures existence of that lone negative eigenvalue of the Cayley-Menger form.

Now we apply results from chapter 2 with regard to polyhedral cones and their duals.

\(^{5.52}\)Recall: for \( D \in \mathbb{S}_h^N \), \(-V_N^T D V_N \succeq 0\) subsumes nonnegativity property 1 (§5.8.1).

\(^{5.53}\)Empirically, all except one entry of the corresponding eigenvector have the same sign with respect to each other.
5.11.2.2 Ordered eigenspectra

Conditions (1188) specify eigenvalue membership to the smallest pointed polyhedral spectral cone for \[
\begin{bmatrix}
0 & 1^T \\
1 & -\text{EDM}_N
\end{bmatrix}
\]:

\[
\mathcal{K}_\lambda \triangleq \{ \zeta \in \mathbb{R}^{N+1} \mid \zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_N \geq 0 \geq \zeta_{N+1}, \ 1^T \zeta = 0 \}
\]

\[= \mathcal{K}_M \cap \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} \cap \partial \mathcal{H} \]

\[= \lambda \left( \begin{bmatrix} 0 & 1^T \\ 1 & -\text{EDM}_N \end{bmatrix} \right) \tag{1192}\]

where

\[
\partial \mathcal{H} \triangleq \{ \zeta \in \mathbb{R}^{N+1} \mid 1^T \zeta = 0 \} \tag{1193}
\]

is a hyperplane through the origin, and \(\mathcal{K}_M\) is the monotone cone (§2.13.9.4.3, implying ordered eigenspectra) which is full-dimensional but is not pointed;

\[
\mathcal{K}_M = \{ \zeta \in \mathbb{R}^{N+1} \mid \zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_N \geq 0 \} \tag{435}
\]

\[
\mathcal{K}_M^* = \{ [e_1 - e_2, e_2 - e_3, \cdots, e_N - e_{N+1}]a \mid a \succeq 0 \} \subset \mathbb{R}^{N+1} \tag{436}
\]

So because of the hyperplane,

\[
\dim \text{aff} \mathcal{K}_\lambda = \dim \partial \mathcal{H} = N \tag{1194}
\]

indicating \(\mathcal{K}_\lambda\) is not full-dimensional. Defining

\[
A \triangleq \begin{bmatrix}
e_1^T - e_2^T \\
e_2^T - e_3^T \\
\vdots \\
e_N^T - e_{N+1}^T
\end{bmatrix} \in \mathbb{R}^{N \times N+1}, \quad B \triangleq \begin{bmatrix}
e_1^T \\
e_2^T \\
\vdots \\
e_N^T \\
-e_{N+1}^T
\end{bmatrix} \in \mathbb{R}^{N+1 \times N+1} \tag{1195}
\]

we have the halfspace-description:

\[
\mathcal{K}_\lambda = \{ \zeta \in \mathbb{R}^{N+1} \mid A \zeta \succeq 0, \ B \zeta \succeq 0, \ 1^T \zeta = 0 \} \tag{1196}
\]

From this and (443) we get a vertex-description for a pointed spectral cone that is not full-dimensional:

\[
\mathcal{K}_\lambda = \left\{ V_\mathcal{N} \left( \begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} \right) V_\mathcal{N}^\dagger b \mid b \succeq 0 \right\} \tag{1197}
\]

where \(V_\mathcal{N} \in \mathbb{R}^{N+1 \times N}\), and where \([\text{sic}]\)

\[
\hat{B} = e_N^T \in \mathbb{R}^{1 \times N+1} \tag{1198}
\]
and

\[
\hat{A} = \begin{bmatrix}
\epsilon^T_1 - \epsilon^T_2 \\
\epsilon^T_2 - \epsilon^T_3 \\
\vdots \\
\epsilon^T_{N-1} - \epsilon^T_N
\end{bmatrix} \in \mathbb{R}^{N-1 \times N+1} \quad (1199)
\]

hold those rows of \( A \) and \( B \) corresponding to conically independent rows (§2.10) in \( [A B] V_N \).

Conditions (1188) can be equivalently restated in terms of a spectral cone for Euclidean distance matrices:

\[
D \in \text{EDM}^N \Leftrightarrow \begin{cases}
\lambda \left( \begin{bmatrix} 0 & 1^T \end{bmatrix} - D \right) \in \mathcal{K}_M \cap \left[ \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} \right] \cap \partial \mathcal{H} \\
D \in \mathbb{S}_h^N
\end{cases} \quad (1200)
\]

Vertex-description of the dual spectral cone is, (313)

\[
\mathcal{K}_\lambda^* = \mathcal{K}_M^* + \left[ \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} \right]^* + \partial \mathcal{H}^* \subseteq \mathbb{R}^{N+1}
\]

\[
= \left\{ \begin{bmatrix} A^T & B^T & 1 \\ -1 \\ 1 \end{bmatrix} b \mid b \succeq 0 \right\} = \left\{ \begin{bmatrix} \hat{A}^T & \hat{B}^T & 1 \\ -1 \end{bmatrix} a \mid a \succeq 0 \right\} \quad (1201)
\]

From (1197) and (444) we get a halfspace-description:

\[
\mathcal{K}_\lambda^* = \{ y \in \mathbb{R}^{N+1} \mid (V_N^T [\hat{A}^T \hat{B}^T])^T V_N^T y \geq 0 \} \quad (1202)
\]

This polyhedral dual spectral cone \( \mathcal{K}_\lambda^* \) is closed, convex, full-dimensional because \( \mathcal{K}_\lambda \) is pointed, but is not pointed because \( \mathcal{K}_\lambda \) is not full-dimensional.

### 5.11.2.3 Unordered eigenspectra

Spectral cones are not unique; eigenspectra ordering can be rendered benign within a cone by presorting a vector of eigenvalues into nonincreasing order.\(^{5.54}\) Then things simplify: Conditions (1188) now specify eigenvalue membership to the spectral cone

\[
\lambda \left( \begin{bmatrix} 0 & 1^T \end{bmatrix} - \text{EDM}^N \right) = \left[ \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} \right] \cap \partial \mathcal{H} \quad (1203)
\]

where \( B \) is defined in (1195), and \( \partial \mathcal{H} \) in (1193). From (443) we get a vertex-description

\[^{5.54}\text{Eigenspectra ordering (represented by a cone having monotone description such as (1192)) becomes benign in (1414), for example, where projection of a given presorted vector on the nonnegative orthant in a subspace is equivalent to its projection on the monotone nonnegative cone in that same subspace; equivalence is a consequence of presorting.}\]
for a pointed spectral cone not full-dimensional:

\[
\lambda\left(\begin{bmatrix} 0 & 1^T \\ 1 & -\text{EDM}^N \end{bmatrix}\right) = \left\{ V_N(\tilde{B}V_N)^\dagger b \mid b \succeq 0 \right\} \\
= \left\{ \begin{bmatrix} I \\ -1^T \end{bmatrix} b \mid b \succeq 0 \right\}
\]

(1204)

where \( V_N \in \mathbb{R}^{N+1 \times N} \) and

\[
\tilde{B} \triangleq \begin{bmatrix} e_1^T \\ e_2^T \\ \cdots \\ e_N^T \end{bmatrix} \in \mathbb{R}^{N \times N+1}
\]

(1205)

holds only those rows of \( B \) corresponding to conically independent rows in \( BV_N \).

For presorted eigenvalues, (1188) can be equivalently restated

\[
D \in \text{EDM}^N \iff \left\{ \lambda\left(\begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix}\right) \in \begin{bmatrix} \mathbb{R}^N_+ \\ \mathbb{R}^- \end{bmatrix} \cap \partial\mathcal{H} \\
D \in S_N^h \right\}
\]

(1206)

Vertex-description of the dual spectral cone is, (313)

\[
\lambda\left(\begin{bmatrix} 0 & 1^T \\ 1 & -\text{EDM}^N \end{bmatrix}\right)^* = \begin{bmatrix} \mathbb{R}^N_+ \\ \mathbb{R}^- \end{bmatrix} + \partial\mathcal{H}^* \subseteq \mathbb{R}^{N+1}
\]

= \{ \begin{bmatrix} B^T & 1 \\ -1 \end{bmatrix} b \mid b \succeq 0 \} = \{ \begin{bmatrix} \tilde{B}^T & 1 \\ -1 \end{bmatrix} a \mid a \succeq 0 \}
\]

(1207)

From (444) we get a halfspace-description:

\[
\lambda\left(\begin{bmatrix} 0 & 1^T \\ 1 & -\text{EDM}^N \end{bmatrix}\right)^* = \{ y \in \mathbb{R}^{N+1} \mid (V_N^T \tilde{B}^T)^\dagger V_N^Ty \geq 0 \}
\]

= \{ y \in \mathbb{R}^{N+1} \mid \begin{bmatrix} I & -1 \end{bmatrix} y \geq 0 \}
\]

(1208)

This polyhedral dual spectral cone is closed, convex, full-dimensional but is not pointed. (Notice that any nonincreasingly ordered eigenspectrum belongs to this dual spectral cone.)

5.11.2.4 Dual cone versus dual spectral cone

An open question regards the relationship of convex cones and their duals to the corresponding spectral cones and their duals. A positive semidefinite cone, for example, is selfdual. Both the nonnegative orthant and the monotone nonnegative cone are spectral cones for it. When we consider the nonnegative orthant, then that spectral cone for the selfdual positive semidefinite cone is also selfdual.
5.12 List reconstruction

The term **metric multidimensional scaling**\(^{5.55}\)\cite{[267]}\cite{[111]}\cite{[365]}\cite{[109]}\cite{[270]}\cite{[92]} refers to any reconstruction of a list \(X \in \mathbb{R}^{n \times N}\) in Euclidean space from interpoint distance information, possibly incomplete (§6.7), ordinal (§5.13.2), or specified perhaps only by bounding-constraints (§5.4.2.12)\cite{[363]}. Techniques for reconstruction are essentially methods for optimally embedding an unknown list of points, corresponding to given Euclidean distance data, in an affine subset of desired or minimum dimension. The oldest known precursor is called **principal component analysis**\cite{[175]} which analyzes the correlation matrix (§5.9.1.0.1);\cite{[55], [22]} a.k.a, **Karhunen-Loève transform** in the digital signal processing literature.

Isometric reconstruction (§5.5.3) of point list \(X\) is best performed by eigenvalue decomposition of a Gram matrix; for then, numerical errors of factorization are easily spotted in the eigenvalues: Now we consider how rotation/reflection and translation invariance factor into a reconstruction.

### 5.12.1 \(x_1\) at the origin. \(V_N\)

At the stage of reconstruction, we have \(D \in \text{EDM}^N\) and wish to find a generating list (§2.3.2) for polyhedron \(\mathcal{P} - \alpha\) by factoring Gram matrix \(-V_N^T D V_N\)\cite{[986]} as in (1171). One way to factor \(-V_N^T D V_N\) is via diagonalization of symmetric matrices;\cite{[314]],[5.6]\cite{[212]} (§A.5.1, §A.3)

\[
-V_N^T D V_N \triangleq QAQ^T \quad (1209)
\]

\[
QAQ^T \succeq 0 \iff \Lambda \succeq 0 \quad (1210)
\]

where \(Q \in \mathbb{R}^{N-1 \times N-1}\) is an orthogonal matrix containing eigenvectors while \(\Lambda \in \mathbb{S}^{N-1}\) is a diagonal matrix containing corresponding nonnegative eigenvalues ordered by nonincreasing value. From the diagonalization, identify the list using (1116):

\[
-V_N^T D V_N = 2V_N^T X^T X V_N \triangleq Q\sqrt{\Lambda} Q_p^T \sqrt{\Lambda} Q_p^T \quad (1211)
\]

where \(\sqrt{\Lambda} Q_p^T Q_p \sqrt{\Lambda} \triangleq \Lambda = \sqrt{\Lambda} \sqrt{\Lambda}\) and where \(Q_p \in \mathbb{R}^{n \times N-1}\) is unknown as is its dimension \(n\). Rotation/reflection is accounted for by \(Q_p\) yet only its first \(r\) columns are necessarily orthonormal\(^{5.56}\). Assuming membership to the unit simplex \(y \in \mathcal{S}\)\cite{[1168]}, then point \(p = X \sqrt{2} V_N y = Q_p \sqrt{\Lambda} Q^T y\) in \(\mathbb{R}^n\) belongs to the translated polyhedron

\[
\mathcal{P} - x_1 \quad (1212)
\]

\(^{5.55}\)Scaling\cite{[361]} means making a scale, i.e., a numerical representation of qualitative data. If the scale is multidimensional, it’s multidimensional scaling.

\(^{5.56}\)Recall \(r\) signifies affine dimension. \(Q_p\) is not necessarily an orthogonal matrix. \(Q_p\) is constrained such that only its first \(r\) columns are necessarily orthonormal because there are only \(r\) nonzero eigenvalues in \(\Lambda\) when \(-V_N^T D V_N\) has rank \(r\) (§5.7.1.1). Remaining columns of \(Q_p\) are arbitrary.
whose generating list constitutes the columns of (1110)

\[
\begin{bmatrix} 0 & X \sqrt{2} V_N \end{bmatrix} = \begin{bmatrix} 0 & Q_p \sqrt{\Lambda} Q_T \end{bmatrix} \in \mathbb{R}^{n \times N} \\
= \begin{bmatrix} 0 & x_2 - x_1 & x_3 - x_1 & \cdots & x_N - x_1 \end{bmatrix}
\]

(1213)

The scaled auxiliary matrix \( V_N \) represents that translation. A simple choice for \( Q_p \) has \( n \) set to \( N - 1 \); \( \text{id est, } Q_p = I \). Ideally, each member of the generating list has at most \( r \) nonzero entries; \( r \) being, affine dimension

\[
\text{rank } V_N^{T} DV_N = \text{rank } Q_p \sqrt{\Lambda} Q_T = \text{rank } \Lambda = r
\]

(1214)

Each member then has at least \( N - 1 - r \) zeros in its higher-dimensional coordinates because \( r \leq N - 1 \). (1122) To truncate those zeros, choose \( n \) equal to affine dimension which is the smallest \( n \) possible because \( XV_N \) has rank \( r \leq n \) (1118). In that case, the simplest choice for \( Q_p \) is \( [1 \ 0] \) having dimension \( r \times N - 1 \).

We may wish to verify the list (1213) found from the diagonalization of \( -V_N^{T} DV_N \). Because of rotation/reflection and translation invariance (§5.5), EDM \( D \) can be uniquely made from that list by calculating: (969)

\[
D(X) = D(X[0 \ \sqrt{2} V_N]) = D(Q_p[0 \ \sqrt{\Lambda} Q_T]) = D([0 \ \sqrt{\Lambda} Q^T])
\]

(1215)

This suggests a way to find EDM \( D \) given \( -V_N^{T} DV_N \). (confer (1094))

\[
D = \begin{bmatrix} 0 & \delta(-V_N^{T} DV_N) \end{bmatrix}^{T} + 1 \begin{bmatrix} 0 & \delta(-V_N^{T} DV_N)^{T} \end{bmatrix} - 2 \begin{bmatrix} 0 & 0^{T} \\ 0 & -V_N^{T} DV_N \end{bmatrix}
\]

(1090)

5.12.2 geometric center. \( V \)

Alternatively we may perform reconstruction using auxiliary matrix \( V \) (§B.4.1) and Gram matrix \( -VDV_{1/2} \) (990) instead; to find a generating list for polyhedron

\[
\mathcal{P} - \alpha_c
\]

(1216)

whose geometric center has been translated to the origin. Redimensioning diagonalization factors \( Q, \Lambda \in \mathbb{R}^{N \times N} \) and unknown \( Q_p \in \mathbb{R}^{n \times N} \), (1117)

\[
-VDV = 2VX^{T} XV = Q\sqrt{\Lambda} Q_p T Q_p \sqrt{\Lambda} Q_T \triangleq QAQ^T
\]

(1217)

where the geometrically centered generating list constitutes (confer (1213))

\[
XV = \frac{1}{\sqrt{2}} Q_p \sqrt{\Lambda} Q^T \in \mathbb{R}^{n \times N} \\
= [x_1 - \frac{1}{N} X1 \ x_2 - \frac{1}{N} X1 \ x_3 - \frac{1}{N} X1 \ \cdots \ x_N - \frac{1}{N} X1]
\]

(1218)

5.57 If we write \( Q^T = \begin{bmatrix} q_1^T \\ \vdots \\ q_{N-1}^T \end{bmatrix} \) as rowwise eigenvectors, \( \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ \vdots & \lambda_r & 0 \\ 0 & \vdots & 0 \end{bmatrix} \) in terms of eigenvalues, and \( Q_p = [q_{p1} \ \cdots \ q_{p,N-1}] \) as column vectors, then \( Q_p \sqrt{\Lambda} Q^T = \sum_{i=1}^{r} \sqrt{\lambda_i} q_{pi} q_{i}^{T} \) is a sum of \( r \) linearly independent rank-one matrices (§B.1.1). Hence the summation has rank \( r \).
where $\alpha_c = \frac{1}{N} X 1$. (§5.5.1.0.1) The simplest choice for $Q_p$ is $[I \ 0] \in \mathbb{R}^{r \times N}$.

Now EDM $D$ can be uniquely made from the list found: (969)

$$D(X) = D(X^T) = D\left(\frac{1}{\sqrt{2}} Q_p \sqrt{\Lambda} Q^T\right) = D\left(\sqrt{\Lambda} Q^T\right) \frac{1}{2}$$

(1219)

This EDM is, of course, identical to (1215). Similarly to (1090), from $-VDV$ we can find EDM $D$ (confer (1081))

$$D = \delta(-VDV\frac{1}{2})1^T + 1\delta(-VDV\frac{1}{2})^T - 2(-VDV\frac{1}{2})$$

(1080)

5.13 Reconstruction examples

5.13.1 Isometric reconstruction

5.13.1.0.1 Example. Cartography.

The most fundamental application of EDMs is to reconstruct relative point position given only interpoint distance information. Drawing a map of the United States is a good illustration of isometric reconstruction (§5.4.2.2.10) from complete distance data. We obtained latitude and longitude information for the coast, border, states, and Great Lakes from the usalo atlas data file within MATLAB Mapping Toolbox; conversion to Cartesian coordinates $(x, y, z)$ via:

$$\phi \triangleq \pi/2 - \text{latitude}$$

$$\theta \triangleq \text{longitude}$$

$$x = \sin(\phi) \cos(\theta)$$

$$y = \sin(\phi) \sin(\theta)$$

$$z = \cos(\phi)$$

We used 64% of the available map data to calculate EDM $D$ from $N = 5020$ points. The original (decimated) data and its isometric reconstruction via (1211) are shown in Figure 146a-d. [393, MATLAB code] The eigenvalues computed for (1209) are

$$\lambda(-V_N^T D V_N) = [199.8 \ 152.3 \ 2.465 \ 0 \ 0 \ 0 \ 0 \ 0 \ \cdots] \!^T$$

(1221)

The 0 eigenvalues have absolute numerical error on the order of $2E-13$; meaning, the EDM data indicates three dimensions ($r = 3$) are required for reconstruction to nearly machine precision.

5.13.2 Isotonic reconstruction

Sometimes only comparative information about distance is known (Earth is closer to the Moon than it is to the Sun). Suppose, for example, EDM $D$ for three points is unknown:

$$D = [d_{ij}] = \begin{bmatrix}
0 & d_{12} & d_{13} \\
d_{12} & 0 & d_{23} \\
d_{13} & d_{23} & 0
\end{bmatrix} \in S^3_3$$

(958)
Figure 146: Map of United States of America showing some state boundaries and the Great Lakes. All plots made by connecting 5020 points. Any difference in scale in (a) through (d) is artifact of plotting routine.

(a) Shows original map made from decimated (latitude, longitude) data.
(b) Original map data rotated (freehand) to highlight curvature of Earth.
(c) Map isometrically reconstructed from an EDM (from distance only).
(d) Same reconstructed map illustrating curvature.
(e)(f) Two views of one isotonic reconstruction (from comparative distance);
    problem (1229) with no sort constraint $\Pi g$ (and no hidden line removal).
but comparative distance data is available:

\[ d_{13} \geq d_{23} \geq d_{12} \quad (1222) \]

With vectorization \( \mathbf{d} = [d_{12} \ d_{13} \ d_{23}]^T \in \mathbb{R}^3 \), we express the comparative data as the nonincreasing sorting

\[
\Pi \mathbf{d} = \begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
d_{12} \\
d_{13} \\
d_{23}
\end{bmatrix} = \begin{bmatrix}
d_{13} \\
d_{23} \\
d_{12}
\end{bmatrix} \in \mathcal{K}_{\mathcal{M}+} \quad (1223)
\]

where \( \Pi \) is a given permutation matrix expressing known sorting action on the entries of unknown EDM \( D \), and \( \mathcal{K}_{\mathcal{M}+} \) is the monotone nonnegative cone (§2.13.9.4.2)

\[ \mathcal{K}_{\mathcal{M}+} = \{ z \mid z_1 \geq z_2 \geq \cdots \geq z_{N(N-1)/2} \geq 0 \} \subseteq \mathbb{R}_+^{N(N-1)/2} \quad (428) \]

where \( N(N-1)/2 = 3 \) for the present example. From sorted vectorization (1223) we create the sort-index matrix

\[
O = \begin{bmatrix}
0 & 1^2 & 3^2 \\
1^2 & 0 & 2^2 \\
3^2 & 2^2 & 0
\end{bmatrix} \in \mathbb{S}_h^3 \cap \mathbb{R}_+^{3 \times 3} \quad (1224)
\]

generally defined

\[
O_{ij} \triangleq k^2 \mid d_{ij} = (\Xi \Pi \mathbf{d})_k, \quad j \neq i \quad (1225)
\]

where \( \Xi \) is a permutation matrix (1820) completely reversing order of vector entries.

Replacing EDM data with indices-square of a nonincreasing sorting like this is, of course, a heuristic we invented and may be regarded as a nonlinear introduction of much noise into the Euclidean distance matrix. For large data sets, this heuristic makes an otherwise intense problem computationally tractable; we see an example in relaxed problem (1230).

Any process of reconstruction that leaves comparative distance information intact is called \textit{ordinal multidimensional scaling} or \textit{isotonic reconstruction}. Beyond rotation, reflection, and translation error, (§5.5) list reconstruction by isotonic reconstruction is subject to error in absolute scale (\textit{dilation}) and distance ratio. Yet Borg & Groenen argue: [55, §2.2] reconstruction from complete comparative distance information for a large number of points is as highly constrained as reconstruction from an EDM; the larger the number, the smaller the optimal solution set; whereas,

\[
\text{isotonic solution set } \supseteq \text{ isometric solution set} \quad (1226)
\]

### 5.13.2.1 Isotonic cartography

To test Borg & Groenen’s conjecture, suppose we make a complete sort-index matrix \( O \in \mathbb{S}_h^N \cap \mathbb{R}_+^{N \times N} \) for the map of the USA and then substitute \( O \) in place of EDM \( D \) in the reconstruction process of §5.12. Whereas EDM \( D \) returned only three significant eigenvalues (1221), the sort-index matrix \( O \) is generally not an EDM (certainly not an
EDM with corresponding affine dimension 3) so returns many more. The eigenvalues, calculated with absolute numerical error approximately $5E-7$, are plotted in Figure 147:

$$\lambda(-V_N^T O V_N) = [880.1 \ 463.9 \ 186.1 \ 46.20 \ 17.12 \ 9.625 \ 8.257 \ 1.701 \ 0.7128 \ 0.6460 \ \cdots]^T$$

The extra eigenvalues indicate that affine dimension corresponding to an EDM near $O$ is likely to exceed 3. To realize the map, we must simultaneously reduce that dimensionality and find an EDM $D$ closest to $O$ in some sense\(^{5.58}\) while maintaining the known comparative distance relationship. For example: given permutation matrix $\Pi$ expressing the known sorting action like $$(1223)$$ on entries

$$d \triangleq \frac{1}{\sqrt{2}} \text{dvec } D = \begin{bmatrix} d_{12} \\ d_{13} \\ d_{23} \\ d_{14} \\ d_{24} \\ d_{34} \\ \vdots \\ d_{N-1,N} \end{bmatrix} \in \mathbb{R}^{N(N-1)/2}$$

of unknown $D \in \mathbb{S}^N_+$, we can make sort-index matrix $O$ input to the optimization problem

\[
\begin{align*}
\text{minimize} \quad & \| -V_N^T (D - O)V_N \|_F \\
\text{subject to} \quad & \text{rank } V_N^T D V_N \leq 3 \\
& \Pi d \in \mathcal{K}_{M^+} \\
& D \in \text{EDM}^N
\end{align*}
\]

\(^{5.58}\) a problem explored more in §7.
that finds the EDM $D$ (corresponding to affine dimension not exceeding 3 in isomorphic dvec $\mathbb{EDM}^N \cap \Pi^T \mathcal{K}_{\mathcal{M}^+}$) closest to $O$ in the sense of Schoenberg (988).

Analytical solution to this problem, ignoring the sort constraint $\Pi d \in \mathcal{K}_{\mathcal{M}^+}$, is known [365]: we get the convex optimization [sic] (§7.1)

$$
\text{minimize} \quad \| -V_N^T(D - O)V_N \|_F \\
\text{subject to} \quad \text{rank} V_N^T DV_N \leq 3 \\
D \in \mathbb{EDM}^N
$$

Only the three largest nonnegative eigenvalues in (1227) need be retained to make list (1213); the rest are discarded. The reconstruction from EDM $D$ found in this manner is plotted in Figure 146-e-f. (In the MATLAB code on Wuzeimization [386], matrix $O$ is normalized by $(N(N-1)/2)^2$.) From these plots it becomes obvious that inclusion of the sort constraint is necessary for isotonic reconstruction.

That sort constraint demands: any optimal solution $D^*$ must possess the known comparative distance relationship that produces the original ordinal distance data $O$ (1225). Ignoring the sort constraint, apparently, violates it. Yet even more remarkable is how much the map reconstructed using only ordinal data still resembles the original map of the USA after suffering the many violations produced by solving relaxed problem (1230). This suggests the simple reconstruction techniques of §5.12 are robust to a significant amount of noise.

5.13.2.2 Isotonic solution with sort constraint

Because problems involving rank are generally difficult, we will partition (1229) into two problems we know how to solve and then alternate their solution until convergence:

$$
\begin{align*}
\text{minimize} & \quad \| -V_N^T(D - O)V_N \|_F \\
\text{subject to} & \quad \text{rank} V_N^T DV_N \leq 3 \\
D & \in \mathbb{EDM}^N \\
\end{align*}
$$

where sort-index matrix $O$ (a given constant in (a)) becomes an implicit vector variable $O_i$ solving the $i^{th}$ instance of (1231b)

$$
\frac{1}{\sqrt{2}} \text{dvec} O_i = o_i \triangleq \Pi^T \sigma^* \in \mathbb{R}^{N(N-1)/2}, \quad i \in \{1, 2, 3 \ldots\}
$$

As mentioned in discussion of relaxed problem (1230), a closed-form solution to problem (1231a) exists. Only the first iteration of (1231a) sees the original sort-index matrix $O$ whose entries are nonnegative whole numbers; id est, $O_0 = O \in S_h^N \cap \mathbb{R}^{N \times N}$ (1225). Subsequent iterations $i$ take the previous solution of (1231b) as input

$$
O_i = \text{dvec}^{-1}(\sqrt{2} o_i) \in S^N
$$
real successors, estimating distance-square not order, to the sort-index matrix $O$.

New convex problem (1231b) finds the unique minimum-distance projection of $\Pi d$ on the monotone nonnegative cone $K_{M^+}$. By defining

$$Y^\dagger = \begin{bmatrix} e_1 - e_2 & e_2 - e_3 & e_3 - e_4 & \cdots & e_m \end{bmatrix} \in \mathbb{R}^{m \times m}$$

(429)

where $m \triangleq N(N-1)/2$, we may rewrite (1231b) as an equivalent quadratic program; a convex problem in terms of the halfspace-description of $K_{M^+}$:

$$\begin{array}{l}
\text{minimize} & (\sigma - \Pi d)^T (\sigma - \Pi d) \\
\text{subject to} & Y^\dagger \sigma \succeq 0
\end{array}$$

(1234)

This quadratic program can be converted to a semidefinite program via Schur-form (§3.5.2); we get the equivalent problem

$$\begin{array}{l}
\text{minimize} & t \\
\text{subject to} & \begin{bmatrix} tI & \sigma - \Pi d \\ (\sigma - \Pi d)^T & 1 \end{bmatrix} \succeq 0 \\
& Y^\dagger \sigma \succeq 0
\end{array}$$

(1235)

5.13.2.3 Convergence

In §E.10 we discuss convergence of alternating projection on intersecting convex sets in a Euclidean vector space; convergence to a point in their intersection. Here the situation is different for two reasons:

Firstly, sets of positive semidefinite matrices having an upper bound on rank are generally not convex. Yet in §7.1.4.0.1 we prove that (1231a) is equivalent to a projection of nonincreasingly ordered eigenvalues on a subset of the nonnegative orthant:

$$\begin{array}{l}
\text{minimize} & \| D - (O \cdot V_N) \|_F \\
\text{subject to} & \text{rank} \left( V_N D V_N \right) \leq 3 = \delta(\Upsilon) \\
& D \in \text{EDM}_N
\end{array} \equiv \begin{array}{l}
\text{minimize} & \| Y - \Lambda \|_F \\
\text{subject to} & \delta(\Upsilon) \in \mathbb{R}^3_+ \end{array}$$

(1236)

where $-V_N^T D V_N \triangleq U\Upsilon U^T \in S^{N-1}$ and $-V_N^T O V_N \triangleq Q\Lambda Q^T \in S^{N-1}$ are ordered diagonalizations (§A.5). It so happens: optimal orthogonal $U^*$ always equals $Q$ given. Linear operator $T(A) = U^* A U^*$, acting on square matrix $A$, is an isometry because Frobenius’ norm is orthogonally invariant (48). This isometric isomorphism $T$ thus maps a nonconvex problem to a convex one that preserves distance.

Secondly, the second half (1231b) of the alternation takes place in a different vector space; $S^N_h$ (versus $S^{N-1}$). From §5.6 we know these two vector spaces are related by an isomorphism, $S^{N-1} = V_N(S^N_h)$ (1099), but not by an isometry.

We have, therefore, no guarantee from theory of alternating projection that the alternation (1231) converges to a point, in the set of all EDMs corresponding to affine dimension not in excess of 3, belonging to $\text{dvec EDM}_N \cap \Pi^T K_{M^+}$. 

5.14. **FIFTH PROPERTY OF EUCLIDEAN METRIC**

### 5.13.2.4 Interlude

Map reconstruction from comparative distance data, isotonic reconstruction, would also prove invaluable to stellar cartography where absolute interstellar distance is difficult to acquire. But we have not yet implemented the second half (1234) of alternation (1231) for USA map data because memory-demands exceed capability of our computer.

#### 5.13.2.4.1 Exercise. Convergence of isotonic solution by alternation.

Empirically demonstrate convergence, discussed in §5.13.2.3, on a smaller data set. ▼

It would be remiss not to mention another method of solution to this isotonic reconstruction problem: Once again we assume only comparative distance data like (1222) is available. Given known set of indices $\mathcal{I}$

\[
\begin{align*}
\text{minimize} & \quad \text{rank} VDV \\
\text{subject to} & \quad d_{ij} \leq d_{kl} \leq d_{mn} \quad \forall (i, j, k, l, m, n) \in \mathcal{I} \\
& \quad D \in \text{EDM}^N
\end{align*}
\]

(1237)

this problem minimizes affine dimension while finding an EDM whose entries satisfy known comparative relationships. Suitable rank heuristics are discussed in §4.4.1 and §7.2.2 that will transform this to a convex optimization problem.

Using contemporary computers, even with a rank heuristic in place of the objective function, this problem formulation is more difficult to compute than the relaxed counterpart problem (1230). That is because there exist efficient algorithms to compute a selected few eigenvalues and eigenvectors from a very large matrix. Regardless, it is important to recognize: the optimal solution set for this problem (1237) is practically always different from the optimal solution set for its counterpart, problem (1229).

### 5.14 Fifth property of Euclidean metric

We continue now with the question raised in §5.3 regarding necessity for at least one requirement more than the four properties of the Euclidean metric (§5.2) to certify realizability of a bounded convex polyhedron or to reconstruct a generating list for it from incomplete distance information. There we saw that the four Euclidean metric properties are necessary for $D \in \text{EDM}^N$ in the case $N = 3$, but become insufficient when cardinality $N$ exceeds 3 (regardless of affine dimension).

#### 5.14.1 Recapitulate

In the particular case $N = 3$, $-V_N^T D V_N \succeq 0$ (1140) and $D \in S_h^3$ are necessary and sufficient conditions for $D$ to be an EDM. By (1142), triangle inequality is then the only Euclidean condition bounding the necessarily nonnegative $d_{ij}$; and those bounds are tight. That means the first four properties of the Euclidean metric are necessary and sufficient
conditions for $D$ to be an EDM in the case $N = 3$; for $i, j \in \{1, 2, 3\}$

\[
\begin{align*}
\sqrt{d_{ij}} & \geq 0, \quad i \neq j \\
\sqrt{d_{ij}} & = 0, \quad i = j \\
\sqrt{d_{ij}} & = \sqrt{d_{jk}} + \sqrt{d_{kj}}, \quad i \neq j \neq k
\end{align*}
\]

\[\Leftrightarrow -V_N^T D V_N \succeq 0 \quad D \in \mathbb{S}_N^4 \quad \Leftarrow D \in \text{EDM}^3 \quad (1238)\]

Yet those four properties become insufficient when $N > 3$.

### 5.14.2 Derivation of the fifth

Correspondence between the triangle inequality and the EDM was developed in §5.8.2 where a triangle inequality (1142a) was revealed within the leading principal $2 \times 2$ submatrix of $-V_N^T D V_N$ when positive semidefinite. Our choice of the leading principal submatrix was arbitrary; actually, a unique triangle inequality like (1037) corresponds to any one of the $(N - 1)!/(2!(N - 1 - 2)!)$ principal $2 \times 2$ submatrices.\(^{5.59}\) Assuming $D \in \mathbb{S}_N^4$ and $-V_N^T D V_N \in \mathbb{S}_N^3$, then by the **positive (semi)definite principal submatrices theorem** (§A.3.1.0.4) it is sufficient to prove: all $d_{ij}$ are nonnegative, all triangle inequalities are satisfied, and $\det(-V_N^T D V_N)$ is nonnegative. When $N = 4$, in other words, that nonnegative determinant becomes the fifth and last Euclidean metric requirement for $D \in \text{EDM}^N$. We now endeavor to ascribe geometric meaning to it.

#### 5.14.2.1 Nonnegative determinant

By (1043) when $D \in \text{EDM}^4$, $-V_N^T D V_N$ is equal to inner product (1038),

\[
\Theta^T \Theta = \begin{bmatrix}
d_{12} & \sqrt{d_{12}d_{13}} \cos \theta_{213} & \sqrt{d_{12}d_{14}} \cos \theta_{214} \\
\sqrt{d_{12}d_{13}} \cos \theta_{213} & d_{13} & \sqrt{d_{13}d_{14}} \cos \theta_{314} \\
\sqrt{d_{12}d_{14}} \cos \theta_{214} & \sqrt{d_{13}d_{14}} \cos \theta_{314} & d_{14}
\end{bmatrix} \quad (1239)
\]

Because Euclidean space is an inner-product space, the more concise inner-product form of the determinant is admitted;

\[
\det(\Theta^T \Theta) = -d_{12}d_{13}d_{14}(\cos(\theta_{213})^2 + \cos(\theta_{214})^2 + \cos(\theta_{314})^2 - 2 \cos \theta_{213} \cos \theta_{214} \cos \theta_{314} - 1) \quad (1240)
\]

The determinant is nonnegative if and only if

\[
\begin{align*}
\cos \theta_{213} \cos \theta_{314} - \sqrt{\sin(\theta_{213})^2 \sin(\theta_{314})^2} & \leq \cos \theta_{213} \leq \cos \theta_{214} \cos \theta_{314} + \sqrt{\sin(\theta_{214})^2 \sin(\theta_{314})^2} \\
\cos \theta_{213} \cos \theta_{314} - \sqrt{\sin(\theta_{213})^2 \sin(\theta_{314})^2} & \leq \cos \theta_{214} \leq \cos \theta_{213} \cos \theta_{314} + \sqrt{\sin(\theta_{213})^2 \sin(\theta_{314})^2} \\
\cos \theta_{213} \cos \theta_{214} - \sqrt{\sin(\theta_{213})^2 \sin(\theta_{214})^2} & \leq \cos \theta_{314} \leq \cos \theta_{213} \cos \theta_{214} + \sqrt{\sin(\theta_{213})^2 \sin(\theta_{214})^2}
\end{align*}
\]

\(^{5.59}\)There are fewer principal $2 \times 2$ submatrices in $-V_N^T D V_N$ than there are triangles made by four or more points because there are $N!/(3!(N - 3)!)$ triangles made by point triples. The triangles corresponding to those submatrices all have vertex $x_1$. (confer §5.8.2.1)
5.14. FIFTH PROPERTY OF EUCLIDEAN METRIC

Figure 148: The relative-angle inequality tetrahedron (1244) bounding EDM$^4$ is regular; drawn in entirety. Each angle $\theta$ (1035) must belong to this solid to be realizable.

which simplifies, for $0 \leq \theta_{i\ell}, \theta_{\ell j}, \theta_{ij} \leq \pi$ and all $i \neq j \neq \ell \in \{2, 3, 4\}$, to

$$\cos(\theta_{i\ell} + \theta_{\ell j}) \leq \cos \theta_{ij} \leq \cos(\theta_{i\ell} - \theta_{\ell j}) \quad (1242)$$

Analogously to triangle inequality (1154), the determinant is 0 upon equality on either side of (1242) which is tight. Inequality (1242) can be equivalently written linearly as a triangle inequality between relative angles 

$$|\theta_{i\ell} - \theta_{ij}| \leq \theta_{ij} \leq \theta_{i\ell} + \theta_{\ell j}$$

$$\theta_{i\ell} + \theta_{\ell j} + \theta_{ij} \leq 2\pi$$

$$0 \leq \theta_{i\ell}, \theta_{ij}, \theta_{\ell j} \leq \pi$$

(1243)

Generalizing this:

5.14.2.1.1 Fifth property of Euclidean metric - restatement.

Relative-angle inequality.  \[51\] [52, p.17, p.107] [245, §3.1] (confer §5.3.1.0.1) Augmenting the four fundamental Euclidean metric properties in $\mathbb{R}^n$, for all $i, j, \ell \neq k \in \{1 \ldots N\}$, $i < j < \ell$, and for $N \geq 4$ distinct points $\{x_k\}$, the inequalities

$$|\theta_{ikt} - \theta_{tkj}| \leq \theta_{ikt} \leq \theta_{ikt} + \theta_{tkj} \quad (a)$$

$$\theta_{ikt} + \theta_{tkj} + \theta_{ijk} \leq 2\pi$$

$$0 \leq \theta_{ikt}, \theta_{tkj}, \theta_{ijk} \leq \pi$$

(1244)

where $\theta_{ijk} = \theta_{jki}$ is the angle between vectors at vertex $x_k$ (as defined in (1035) and illustrated in Figure 129), must be satisfied at each point $x_k$ regardless of affine dimension.
Because point labelling is arbitrary, this fifth Euclidean metric requirement must apply to each of the \( N \) points as though each were in turn labelled \( x_1 \); hence the new index \( k \) in (1244). Just as the triangle inequality is the ultimate test for realizability of only three points, the relative-angle inequality is the ultimate test for only four. For four distinct points, the triangle inequality remains a necessary although penultimate test; (§5.4.3)

Four Euclidean metric properties (§5.2).
Angle \( \theta \) inequality \((963)\) or \((1244)\).

\[
\implies -V_N^T D V_N \succeq 0 \quad D \in \mathbb{S}_h^4 \quad \iff \quad D = D(\Theta) \in \text{EDM}^4 \quad (1245)
\]

The relative-angle inequality, for this case, is illustrated in Figure 148.

5.14.2.2 Beyond the fifth metric property

When cardinality \( N \) exceeds 4, the first four properties of the Euclidean metric and the relative-angle inequality together become insufficient conditions for realizability. In other words, the four Euclidean metric properties and relative-angle inequality remain necessary but become a sufficient test only for positive semidefiniteness of all the principal \( 3 \times 3 \) submatrices \([\text{sic}]\) in \(-V_N^T D V_N\). Relative-angle inequality can be considered the ultimate test only for realizability at each vertex \( x_k \) of each and every purported tetrahedron constituting a hyperdimensional body.

When \( N = 5 \) in particular, relative-angle inequality becomes the penultimate Euclidean metric requirement while nonnegativity of then unwieldy \( \det(\Theta^T \Theta) \) corresponds (by the positive (semi)definite principal submatrices theorem in §A.3.1.0.4) to the sixth and last Euclidean metric requirement. Together these six tests become necessary and sufficient, and so on.

Yet for all values of \( N \), only assuming nonnegative \( d_{ij} \), relative-angle matrix inequality in (1156) is necessary and sufficient to certify realizability; (§5.4.3.1)

Euclidean metric property 1 (§5.2).
Angle matrix inequality \( \Omega \succeq 0 \) \((1044)\).

\[
\implies -V_N^T D V_N \succeq 0 \quad D \in \mathbb{S}_h^N \quad \iff \quad D = D(\Omega, d) \in \text{EDM}^N \quad (1246)
\]

Like matrix criteria \((964), (988),\) and \((1156)\), the relative-angle matrix inequality and nonnegativity property subsume all the Euclidean metric properties and further requirements.

5.14.3 Path not followed

As a means to test for realizability of four or more points, an intuitively appealing way to augment the four Euclidean metric properties is to recognize generalizations of the triangle inequality: In the case of cardinality \( N = 4 \), the three-dimensional analogue to triangle & distance is tetrahedron & facet-area, while in case \( N = 5 \) the four-dimensional analogue is polychoron & facet-volume, \textit{ad infinitum}. For \( N \) points, \( N + 1 \) metric properties are required.
5.14. FIFTH PROPERTY OF EUCLIDEAN METRIC

5.14.3.1 $N = 4$

Each of the four facets of a general tetrahedron is a triangle and its relative interior. Suppose we identify each facet of the tetrahedron by its area-square: $c_1, c_2, c_3, c_4$. Then analogous to metric property 4, we may write a tight area inequality for the facets

$$\sqrt{c_i} \leq \sqrt{c_j} + \sqrt{c_k} + \sqrt{c_\ell}, \quad i \neq j, k \neq \ell \in \{1, 2, 3, 4\}$$

(1247)

which is a generalized “triangle” inequality [237, §1.1] that follows from

$$\sqrt{c_i} = \sqrt{c_j} \cos \varphi_{ij} + \sqrt{c_k} \cos \varphi_{ik} + \sqrt{c_\ell} \cos \varphi_{i\ell}$$

(1248)

[252] [405, Law of Cosines] where $\varphi_{ij}$ is the dihedral angle at the common edge between triangular facets $i$ and $j$.

If $D$ is the EDM corresponding to the whole tetrahedron, then area-square of the $i$th triangular facet has a convenient formula in terms of $D_i \in \text{EDM}^{N-1}$ the EDM corresponding to that particular facet: From the Cayley-Menger determinant for simplices, [405] [138] [172, §4] [91, §3.3] the $i$th facet area-square for $i \in \{1 \ldots N\}$ is (§A.4.1)

$$c_i = \frac{-1}{2^{N-2}(N-2)!^2} \det \begin{bmatrix} 0 & 1^T \\ 1 & -D_i \end{bmatrix}$$

(1249)

$$= \frac{(-1)^N}{2^{N-2}(N-2)!^2} \det D_i \left(1^T D_i^{-1} 1\right)$$

(1250)

$$= \frac{(-1)^N}{2^{N-2}(N-2)!^2} 1^T \text{cof}(D_i) 1$$

(1251)

where $D_i$ is the $i$th principal $N-1 \times N-1$ submatrix of $D \in \text{EDM}^N$, and $\text{cof}(D_i)$ is the $N-1 \times N-1$ matrix of cofactors [341, §4] corresponding to $D_i$.

The number of principal 3 $\times$ 3 submatrices in $D$ is, of course, equal to the number of triangular facets in the tetrahedron; four ($N!/(3!(N-3)!)$) when $N = 4$.

5.14.3.1.1 Exercise. Sufficiency conditions for an EDM of four points. Triangle inequality (property 4) and area inequality (1247) are conditions necessary for $D$ to be an EDM. Prove their sufficiency in conjunction with the remaining three Euclidean metric properties.

---

5.60 The upper bound is met when all angles in (1248) are simultaneously 0; that occurs, for example, if one point is relatively interior to the convex hull of the three remaining.

5.61 whose foremost characteristic is: the determinant vanishes if and only if affine dimension does not equal penultimate cardinality; id est, $\det \begin{bmatrix} 0 & 1^T \\ 1 & -D \end{bmatrix} = 0 \Leftrightarrow r < N-1$ where $D$ is any EDM (§5.7.3.0.1).

Otherwise, the determinant is negative.

5.62 Every principal submatrix of an EDM remains an EDM. [245, §4.1]
CHAPTER 5. EUCLIDEAN DISTANCE MATRIX

5.14.3.2 \( N = 5 \)

Moving to the next level, we might encounter a Euclidean body called polychoron: a bounded polyhedron in four dimensions.\(^{5.63}\) Our polychoron has five (\( N!/(4!(N-4)!) \)) facets, each of them a general tetrahedron whose volume-square \( c_i \) is calculated using the same formula: (1249) where \( D \) is the EDM corresponding to the polychoron, and \( D_i \) is the EDM corresponding to the \( i^{th} \) facet (the principal \( 4 \times 4 \) submatrix of \( D \in \mathbb{ EDM}^N \) corresponding to the \( i^{th} \) tetrahedron). The analogue to triangle & distance is now polychoron & facet-volume. We could then write another generalized “triangle” inequality like (1247) but in terms of facet volume; \(^{5.63}\) [1249, §IV]

\[
\sqrt{c_i} \leq \sqrt{c_j} + \sqrt{c_k} + \sqrt{c_l} + \sqrt{c_m}, \quad i \neq j \neq k \neq \ell \neq m \in \{1 \ldots 5\} \quad (1252)
\]

5.14.3.2.1 Exercise. Sufficiency for an EDM of five points.

For \( N = 5 \), triangle (distance) inequality (§5.2), area inequality (1247), and volume inequality (1252) are conditions necessary for \( D \) to be an EDM. Prove their sufficiency. ▼

5.14.3.3 Volume of simplices

There is no known formula for the volume of a bounded general convex polyhedron expressed either by halfspace or vertex-description. \(^{[423, \S 2.1]}\) \(^{[293, \text{p.173}]}\) \(^{[243]}\) \(^{[183]}\) \(^{[184]}\) Volume is a concept germane to \( \mathbb{R}^3 \); in higher dimensions it is called content. Applying the EDM assertion (§5.9.1.0.4) and a result from \(^{[63, \text{p.407}]}\), a general nonempty simplex (\( \S 2.12.3 \)) in \( \mathbb{R}^N \) corresponding to an EDM \( D \in \mathbb{ EDM}^N \) has content

\[
\sqrt{\mathcal{C}} = \text{content}(S) \sqrt{\det(-V^T_N DV_N)} \quad (1253)
\]

where content-square of the unit simplex \( S \subset \mathbb{R}^{N-1} \) is proportional to its Cayley-Menger determinant;

\[
\text{content}(S)^2 = \frac{-1}{2^{N-1}(N-1)!} \det \begin{bmatrix} 0 & 1^T \\ 1 & -D([0 \ e_1 \ e_2 \ \cdots \ e_{N-1}]) \end{bmatrix} \quad (1254)
\]

where \( e_i \in \mathbb{R}^{N-1} \) and the EDM operator used is \( D(X) \) (969).

5.14.3.3.1 Example. Pyramid.

A formula for volume of a pyramid is known;\(^{5.64}\) it is \( \frac{1}{3} \) the product of its base area with its height. \(^{[233]}\) The pyramid in Figure 149 has volume \( \frac{1}{3} \). To find its volume using EDMs, we must first decompose the pyramid into

---

\(^{5.63}\) The simplest polychoron is called a pentatope \(^{[405]}\); a regular simplex hence convex. (A pentahedron is a three-dimensional body having five vertices.)

\(^{5.64}\) Pyramid volume is independent of the paramount vertex position as long as its height remains constant.
5.14. **FIFTH PROPERTY OF EUCLIDEAN METRIC**

Figure 149: Length of one-dimensional face \( a \) equals height \( h=a=1 \) of this convex nonsimplicial pyramid in \( \mathbb{R}^3 \) with square base inscribed in a circle of radius \( R \) centered at the origin. [405, *Pyramid*]

Slicing it in half along the plane containing the line segments corresponding to radius \( R \) and height \( h \) we find the vertices of one simplex,

\[
X = \begin{bmatrix}
1/2 & 1/2 & -1/2 & 0 \\
1/2 & -1/2 & -1/2 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix} \in \mathbb{R}^{n \times N} \tag{1255}
\]

where \( N = n + 1 \) for any nonempty simplex in \( \mathbb{R}^n \). The volume of this simplex is half that of the entire pyramid; \( \text{id est, } \sqrt{c} = \frac{1}{6} \) found by evaluating (1253).

With that, we conclude digression of path.

### 5.14.4 Affine dimension reduction in three dimensions

(*confer §5.8.4*) The determinant of any \( M \times M \) matrix is equal to the product of its \( M \) eigenvalues. [341, §5.1] When \( N = 4 \) and \( \det(\Theta^T\Theta) = 0 \), that means one or more eigenvalues of \( \Theta^T\Theta \in \mathbb{R}^{3 \times 3} \) are 0. The determinant will go to 0 whenever equality is attained on either side of (963), (1244a), or (1244b), meaning that a tetrahedron has collapsed to a lower affine dimension; \( \text{id est, } r = \text{rank } \Theta^T\Theta = \text{rank } \Theta \) is reduced below \( N-1 \) exactly by the number of 0 eigenvalues (§5.7.1.1).

In solving completion problems of any size \( N \) where one or more entries of an EDM are unknown, therefore, dimension \( r \) of the affine hull required to contain the unknown points is potentially reduced by selecting distances to attain equality in (963) or (1244a) or (1244b).
5.14.4.1 Exemplum redux

We now apply the fifth Euclidean metric property to an earlier problem:

5.14.4.1.1 Example. Small completion problem, IV. (confer §5.9.3.0.1)
Returning again to Example 5.3.0.0.2 that pertains to Figure 128 where \( N = 4 \),
distance-square \( d_{14} \) is ascertainable from the fifth Euclidean metric property. Because
all distances in (961) are known except \( \sqrt{d_{14}} \), then \( \cos \theta_{123} = 0 \) and \( \theta_{324} = 0 \) result from
identity (1035). Applying (963),

\[
\cos(\theta_{123} + \theta_{324}) \leq \cos \theta_{124} \leq \cos(\theta_{123} - \theta_{324})
\]

\( 0 \leq \cos \theta_{124} \leq 0 \) \( (1256) \)

It follows again from (1035) that \( d_{14} \) can only be 2. As explained in this subsection, affine
dimension \( r \) cannot exceed \( N - 2 \) because equality is attained in (1256). \( \square \)
Chapter 6

Cone of distance matrices

For $N > 3$, the cone of EDMs is no longer a circular cone and the geometry becomes complicated...

– Hayden, Wells, Liu, & Tarazaga, 1991 [195, §3]

In the subspace of symmetric matrices $\mathbb{S}^N$, we know that the convex cone of Euclidean distance matrices $\text{EDM}^N$ (the EDM cone) does not intersect the positive semidefinite cone $\mathbb{S}^N_+$ (PSD cone) except at the origin, their only vertex; there can be no positive or negative semidefinite EDM. (1180) [245]

$$\text{EDM}^N \cap \mathbb{S}^N_+ = \mathbf{0} \quad (1257)$$

Even so, the two convex cones can be related. In §6.8.1 we prove the equality

$$\text{EDM}^N = \mathbb{S}^N_h \cap \left( \mathbb{S}^{N\perp}_c - \mathbb{S}^N_+ \right) \quad (1350)$$

a resemblance to EDM definition (969) where

$$\mathbb{S}^N_h = \left\{ A \in \mathbb{S}^N \mid \delta(A) = \mathbf{0} \right\} \quad (66)$$

is the symmetric hollow subspace (§2.2.3) and where

$$\mathbb{S}^{N\perp}_c = \left\{ u\mathbf{1}^T + \mathbf{1}u^T \mid u \in \mathbb{R}^N \right\} \quad (2097)$$

is the orthogonal complement of the geometric center subspace (§E.7.2.0.2)

$$\mathbb{S}^N_c = \left\{ Y \in \mathbb{S}^N \mid Y\mathbf{1} = \mathbf{0} \right\} \quad (2095)$$
6.0.1 gravity
Equality (1350) is equally important as the known isomorphisms (1088) (1089) (1100) (1101) relating the EDM cone $\text{EDM}_N$ to positive semidefinite cone $\mathbb{S}^N_{+}$ (§5.6.2.1) or to an $N(N-1)/2$-dimensional face of $\mathbb{S}^N_{+}$ (§5.6.1.1). But those isomorphisms have never led to this equality relating whole cones $\text{EDM}_N$ and $\mathbb{S}^N_{+}$.

Equality (1350) is not obvious from the various EDM definitions such as (969) or (1273) because inclusion must be proved algebraically in order to establish equality; $\text{EDM}_N \supseteq \mathbb{S}^N_{h} \cap \left(\mathbb{S}^N_{c} - \mathbb{S}^N_{+}\right)$. We will instead prove (1350) using purely geometric methods.

6.0.2 highlight
In §6.8.1.7 we show: the Schoenberg criterion for discriminating Euclidean distance matrices

$$D \in \text{EDM}_N \iff \begin{cases} -V_N^T D V_N \in \mathbb{S}^{N-1}_{+} \\ D \in \mathbb{S}^N_{h} \end{cases}$$

is a discretized membership relation (§2.13.4, dual generalized inequalities) between the EDM cone and its ordinary dual.

6.1 Defining EDM cone
We invoke a popular matrix criterion to illustrate correspondence between the EDM and PSD cones belonging to the ambient space of symmetric matrices:

$$D \in \text{EDM}_N \iff \begin{cases} -V D V \in \mathbb{S}^N_{+} \\ D \in \mathbb{S}^N_{h} \end{cases}$$

where $V \in \mathbb{S}^N$ is the geometric centering matrix (§B.4). The set of all EDMs of dimension $N \times N$ forms a closed convex cone $\text{EDM}_N$ because any pair of EDMs satisfies the definition of a convex cone (175); \textit{videlicet}, for each and every $\zeta_1, \zeta_2 \geq 0$ (§A.3.1.0.2)

$$\zeta_1 V D_1 V + \zeta_2 V D_2 V \succeq 0 \iff V D_1 V \succeq 0 , \quad V D_2 V \succeq 0$$

and convex cones are invariant to inverse linear transformation [318, p.22].

\footnote{Because both positive semidefinite cones are frequently in play, dimension is explicit.}
6.1. DEFINING EDM CONE

Figure 150: Relative boundary (tiled) of EDM cone $\mathbb{EDM}^3$ drawn truncated in isometrically isomorphic subspace $\mathbb{R}^3$. (a) EDM cone drawn in usual distance-square coordinates $d_{ij}$. View is from interior toward origin. Unlike positive semidefinite cone, EDM cone is not selfdual; neither is it proper in ambient symmetric subspace (dual EDM cone for this example belongs to isomorphic $\mathbb{R}^6$). (b) Drawn in its natural coordinates $\sqrt{d_{ij}}$ (absolute distance), cone remains convex (confer §5.10); intersection of three halfspaces (1143) whose partial boundaries each contain origin. Cone geometry becomes nonconvex (nonpolyhedral) in higher dimension. (§6.3) (c) Two coordinate systems artificially superimposed. Coordinate transformation from $d_{ij}$ to $\sqrt{d_{ij}}$ appears a topological contraction. (d) Sitting on its vertex $0$, pointed $\mathbb{EDM}^3$ is a circular cone having axis of revolution $dvec(-E) = dvec(11^T - I)$ (1176) (73). (Rounded vertex is plot artifact.)
6.1.0.0.1 Definition. Cone of Euclidean distance matrices.

In the subspace of symmetric matrices, the set of all Euclidean distance matrices forms a unique immutable pointed closed convex cone called the \( \text{EDM cone} \): for \( N > 0 \)

\[
\text{EDM}^N \triangleq \left\{ D \in S^N_+ \mid -VDV \in S^N_+ \right\}
\]

(1259)

The EDM cone in isomorphic \( \mathbb{R}^{N(N+1)/2} \) \( \operatorname{sic} \) is the intersection of an infinite number (when \( N > 2 \)) of halfspaces about the origin and a finite number of hyperplanes through the origin in vectorized variable \( D = [d_{ij}] \). Hence \( \text{EDM}^N \) is not full-dimensional with respect to \( S^N \) because it is confined to the symmetric hollow subspace \( S^N_h \). The EDM cone relative interior comprises

\[
\text{rel int EDM}^N = \bigcap_{z \in \mathbb{N}(1^T)} \left\{ D \in S^N \mid \langle zz^T, -D \rangle > 0, \; \delta(D) = 0 \right\}
\]

(1260)

while its relative boundary comprises

\[
\text{rel } \partial \text{EDM}^N = \left\{ D \in \text{EDM}^N \mid \langle zz^T, -D \rangle = 0 \text{ for some } z \in \mathbb{N}(1^T) \right\}
\]

(1261)

\( \triangle \)

This cone is more easily visualized in the isomorphic vector subspace \( \mathbb{R}^{N(N-1)/2} \) corresponding to \( S^N_h \):

In the case \( N = 1 \) point, the EDM cone is the origin in \( \mathbb{R}^0 \).

In the case \( N = 2 \), the EDM cone is the nonnegative real line in \( \mathbb{R} \); a halfline in a subspace of the realization in Figure 154.

The EDM cone in the case \( N = 3 \) is a circular cone in \( \mathbb{R}^3 \) illustrated in Figure 150(a)(d); rather, the set of all matrices

\[
D = \begin{bmatrix} 0 & d_{12} & d_{13} \\ d_{12} & 0 & d_{23} \\ d_{13} & d_{23} & 0 \end{bmatrix} \in \text{EDM}^3
\]

(1262)

makes a circular cone in this dimension. In this case, the first four Euclidean metric properties are necessary and sufficient tests to certify realizability of triangles; (1238).

Thus triangle inequality property 4 describes three halfspaces (1143) whose intersection makes a polyhedral cone in \( \mathbb{R}^3 \) of realizable \( \sqrt{d_{ij}} \) (absolute distance); an isomorphic subspace representation of the set of all EDMs \( D \) in the natural coordinates

\[
\sqrt{D} \triangleq \begin{bmatrix} 0 & \sqrt{d_{12}} & \sqrt{d_{13}} \\ \sqrt{d_{12}} & 0 & \sqrt{d_{23}} \\ \sqrt{d_{13}} & \sqrt{d_{23}} & 0 \end{bmatrix}
\]

(1263)

illustrated in Figure 150b.
6.2 Polyhedral bounds

The convex cone of EDMs is nonpolyhedral in $d_{ij}$ for $N > 2$; e.g., Figure 150a. Still we found necessary and sufficient bounding polyhedral relations consistent with EDM cones for cardinality $N = 1, 2, 3, 4$:

$N = 3$. Transforming distance-square coordinates $d_{ij}$ by taking their positive square root provides the polyhedral cone in Figure 150b; polyhedral because an intersection of three halfspaces in natural coordinates $\sqrt{d_{ij}}$ is provided by triangle inequalities (1143). This polyhedral cone implicitly encompasses necessary and sufficient metric properties: nonnegativity, selfdistance, symmetry, and triangle inequality.

$N = 4$. Relative-angle inequality (1244) together with four Euclidean metric properties are necessary and sufficient tests for realizability of tetrahedra. (1245) Albeit relative angles $\theta_{ikj}$ (1035) are nonlinear functions of the $d_{ij}$, relative-angle inequality provides a regular tetrahedron in $\mathbb{R}^3$ [sic] (Figure 148) bounding angles $\theta_{ikj}$ at vertex $x_k$ consistently with EDM.

Yet were we to employ the procedure outlined in §5.14.3 for making generalized triangle inequalities, then we would find all the necessary and sufficient $d_{ij}$-transformations for generating bounding polyhedra consistent with EDMs of any higher dimension ($N > 3$).

6.3 $\sqrt{\text{EDM}}$ cone is not convex

For some applications, like a molecular conformation problem (Figure 4, Figure 139) or multidimensional scaling [108] [366], absolute distance $\sqrt{d_{ij}}$ is the preferred variable. Taking square root of the entries in all EDMs $D$ of dimension $N$, we get another cone but not a convex cone when $N > 3$ (Figure 150b): [92, §4.5.2]

$$\sqrt{\text{EDM}}^N \triangleq \{ \sqrt{D} \mid D \in \text{EDM}^N \}$$

(1264)

where $\sqrt{D}$ is defined like (1263). It is a cone simply because any cone is completely constituted by rays emanating from the origin: (§2.7) Any given ray $\{ \zeta^\Gamma \in \mathbb{R}^{N(N-1)/2} \mid \zeta \geq 0 \}$ remains a ray under entrywise square root: $\{ \sqrt{\zeta^\Gamma} \in \mathbb{R}^{N(N-1)/2} \mid \zeta \geq 0 \}$. It is already established that

$$D \in \text{EDM}^N \Rightarrow \sqrt{D} \in \text{EDM}^N \quad (1175)$$

But because of how $\sqrt{\text{EDM}}^N$ is defined, it is obvious that (confer §5.10)

$$D \in \text{EDM}^N \iff \sqrt{D} \in \sqrt{\text{EDM}}^N \quad (1265)$$

Were $\sqrt{\text{EDM}}^N$ convex, then given $\sqrt{D_1}, \sqrt{D_2} \in \sqrt{\text{EDM}}^N$ we would expect their conic combination $\sqrt{D_1} + \sqrt{D_2}$ to be a member of $\sqrt{\text{EDM}}^N$. That is easily proven false by

---

6.2 Still, property-4 triangle inequalities (1143) corresponding to each principal $3 \times 3$ submatrix of $-V^T_NDV_N$ demand that the corresponding $\sqrt{d_{ij}}$ belong to a polyhedral cone like that in Figure 150b.
counterexample via (1265), for then \((\sqrt{D_1} + \sqrt{D_2}) \circ (\sqrt{D_1} + \sqrt{D_2})\) would need to be a member of \(\text{EDM}^N\).

Notwithstanding,
\[
\sqrt{\text{EDM}^N} \subseteq \text{EDM}^N
\]  
(1266)

by (1175) (Figure 150), and we learn how to transform a nonconvex \textit{proximity problem} in the natural coordinates \(\sqrt{d_{ij}}\) to a convex optimization in §7.2.1.

### 6.4 EDM definition in \(11^T\)

Any EDM \(D\) corresponding to affine dimension \(r\) has representation
\[
D(V_X) \triangleq \delta(V_X V_X^T) 1^T + 1 \delta(V_X V_X^T)^T - 2 V_X V_X^T \in \text{EDM}^N
\]  
(1267)

where \(\mathcal{R}(V_X \in \mathbb{R}^{N \times r}) \subseteq \mathcal{N}(1^T) = 1^T\)
\[
V_X^T V_X = \delta^2(V_X V_X^T) \quad \text{and} \quad V_X \text{ is full-rank with orthogonal columns.}
\]  
(1268)

Equation (1267) is simply the standard EDM definition (969) with a centered list \(X\) as in (1056); Gram matrix \(X^T X\) has been replaced with the subcompact singular value decomposition (§A.6.2)\(^6,3\)
\[
V_X^T V_X = V_X^T X^T X V_X \in S_+^N \cap S_-^N
\]  
(1269)

This means: inner product \(V_X^T V_X\) is an \(r \times r\) diagonal matrix \(\Sigma\) of nonzero singular values.

Vector \(\delta(V_X V_X^T)\) may me decomposed into complementary parts by projecting it on orthogonal subspaces \(1^\perp\) and \(\mathcal{R}(1)\): namely,
\[
P_{1^\perp}(\delta(V_X V_X^T)) = V \delta(V_X V_X^T)
\]  
(1270)
\[
P_1(\delta(V_X V_X^T)) = \frac{1}{N} 11^T \delta(V_X V_X^T)
\]  
(1271)

Of course
\[
\delta(V_X V_X^T) = V \delta(V_X V_X^T) + \frac{1}{N} 11^T \delta(V_X V_X^T)
\]  
(1272)

by (991). Substituting this into EDM definition (1267), we get the Hayden, Wells, Liu, & Tarazaga EDM formula [195, §2]
\[
D(V_X, y) \triangleq y 1^T + 1 y^T + \frac{\lambda}{N} 11^T - 2 V_X V_X^T \in \text{EDM}^N
\]  
(1273)

where
\[
\lambda \triangleq 2 \| V_X \|^2_F = 1^T \delta(V_X V_X^T) 2 \quad \text{and} \quad y \triangleq \delta(V_X V_X^T) - \frac{\lambda}{2N} 1 = V \delta(V_X V_X^T)
\]  
(1274)

\(^6,3\)Subcompact SVD: \(V_X V_X^T \triangleq Q \sqrt{\Sigma} Q^T \equiv V^T X^T X V\). So \(V_X^T\) is not necessarily \(XV\) (§5.5.1.0.1), although affine dimension \(r = \text{rank}(V_X^T) = \text{rank}(XV)\). (1113)
and \( y=0 \) if and only if \( \mathbf{1} \) is an eigenvector of EDM \( D \). Scalar \( \lambda \) becomes an eigenvalue when corresponding eigenvector \( \mathbf{1} \) exists.\(^{6.4} \)

Then the particular dyad sum from (1273)

\[
y \mathbf{1}^T + \mathbf{1} y^T + \frac{\lambda}{N} \mathbf{1}^T \mathbf{1}^T \in \mathbb{S}_{c}^{N,l} \tag{1275}
\]

must belong to the orthogonal complement of the geometric center subspace (p.628), whereas \( V_X V_X^T \in \mathbb{S}_c^N \cap \mathbb{S}_+^N \) (1269) belongs to the positive semidefinite cone in the geometric center subspace.

**Proof.** We validate eigenvector \( \mathbf{1} \) and eigenvalue \( \lambda \).

\((\Rightarrow) \) Suppose \( \mathbf{1} \) is an eigenvector of EDM \( D \). Then because

\[
V_X^T \mathbf{1} = 0 \tag{1276}
\]

it follows

\[
D \mathbf{1} = \delta(V_X V_X^T) \mathbf{1}^T \mathbf{1} + \mathbf{1} \delta(V_X V_X^T) \mathbf{1} = N \delta(V_X V_X^T) + \|V_X\|^2_F \mathbf{1} \\
\Rightarrow \delta(V_X V_X^T) \propto \mathbf{1} \tag{1277}
\]

For some \( \kappa \in \mathbb{R}_+ \)

\[
\delta(V_X V_X^T) \mathbf{1} = N \kappa = \text{tr}(V_X^T V_X) = \|V_X\|^2_F \Rightarrow \delta(V_X V_X^T) = \frac{1}{N} \|V_X\|^2_F \mathbf{1} \tag{1278}
\]

so \( y=0 \).

\((\Leftarrow) \) Now suppose \( \delta(V_X V_X^T) = \frac{\lambda}{2N} \mathbf{1} \); \textit{id est}, \( y=0 \). Then

\[
D = \frac{\lambda}{N} \mathbf{1}^T - 2V_X V_X^T \in \text{EDM}^N \tag{1279}
\]

\( \mathbf{1} \) is an eigenvector with corresponding eigenvalue \( \lambda \). \( \diamond \)

### 6.4.1 Range of EDM \( D \)

From §B.1.1 pertaining to linear independence of dyad sums: If the transpose halves of all the dyads in the sum (1267)\(^{6.5} \) make a linearly independent set, then the nontranspose halves constitute a basis for the range of EDM \( D \). Saying this mathematically: For \( D \in \text{EDM}^N \)

\[
\mathcal{R}(D) = \mathcal{R}(\left[ \delta(V_X V_X^T) \quad \mathbf{1} \quad V_X \right]) \Leftarrow \text{rank}(\left[ \delta(V_X V_X^T) \quad \mathbf{1} \quad V_X \right]) = 2 + r \\
\mathcal{R}(D) = \mathcal{R}(\left[ \mathbf{1} \quad V_X \right]) \Leftarrow \text{otherwise} \tag{1280}
\]

\(^{6.4} \text{e.g., when } X = I \text{ in EDM definition (909).} \)

\(^{6.5} \text{Identifying columns } V_X \equiv [v_1 \cdots v_r] \text{, then } V_X V_X^T = \sum_i v_i v_i^T \text{ is also a sum of dyads.} \)
Figure 151: Example of $V_X$ selection to make an EDM corresponding to cardinality $N = 3$ and affine dimension $r = 1$; $V_X$ is a vector in nullspace $\mathcal{N}(1^T) \subset \mathbb{R}^3$. Nullspace of $1^T$ is hyperplane in $\mathbb{R}^3$ (drawn truncated) having normal $1$. Vector $\delta(V_X V_X^T)$ may or may not be in plane spanned by $\{1, V_X\}$, but belongs to nonnegative orthant which is strictly supported by $\mathcal{N}(1^T)$. 
6.4. EDM DEFINITION IN 11T

6.4.1.0.1 Proof. We need that condition under which the rank equality is satisfied: We know \( R(V_X) \perp 1 \), but what is the relative geometric orientation of \( \delta(V_XV_X^T) \)? \( \delta(V_XV_X^T) \geq 0 \) because \( V_XV_X^T \geq 0 \), and \( \delta(V_XV_X^T) \propto 1 \) remains possible (1277); this means \( \delta(V_XV_X^T) \notin \mathcal{N}(1^T) \) simply because it has no negative entries. (Figure 151) If the projection of \( \delta(V_XV_X^T) \) on \( \mathcal{N}(1^T) \) does not belong to \( R(V_X) \), then that is a necessary and sufficient condition for linear independence (l.i.) of \( \delta(V_XV_X^T) \) with respect to \( R([1 \; V_X]) \); id est,

\[
V\delta(V_XV_X^T) \neq V_X a \quad \text{for any } a \in \mathbb{R}^r
\]

\[
(J - \frac{1}{N} \mathbf{11}^T)\delta(V_XV_X^T) + \frac{1}{N} \mathbf{1} - V_X a
\]

\[
\delta(V_XV_X^T) - \frac{1}{N}\mathbf{1} = y \neq V_X a \iff \{1, \; \delta(V_XV_X^T), \; V_X\} \text{ is l.i.}
\]

When this condition is violated (when (1274) \( y = V_X a_p \) for some particular \( a \in \mathbb{R}^r \), on the other hand, then from (1273) we have

\[
\mathcal{R}(D = y1^T + 1y^T + \frac{1}{N} \mathbf{11}^T - 2V_XV_X^T) = \mathcal{R}((V_Xa_p + \frac{1}{N}\mathbf{1})1^T + (1a_p^T - 2V_X)1^T)
\]

\[
= \mathcal{R}([V_Xa_p + \frac{1}{N}\mathbf{1} \; 1a_p^T - 2V_X])
\]

An example of such a violation is (1279) where, in particular, \( a_p = 0 \).

Then a statement parallel to (1280) is, for \( D \in \text{EDM}^N \) (Theorem 5.7.3.0.1)

\[
\text{rank}(D) = r + 2 \iff y \notin \mathcal{R}(V_X) \quad (\iff 1^T D^\dagger 1 = 0)
\]

\[
\text{rank}(D) = r + 1 \iff y \in \mathcal{R}(V_X) \quad (\iff 1^T D^\dagger 1 \neq 0)
\]

6.4.2 Boundary constituents of EDM cone

Expression (1267) has utility in forming the set of all EDMs corresponding to affine dimension \( r \):

\[
\{D \in \text{EDM}^N \mid \text{rank}(VDV) = r\}
\]

\[
= \{D(V_X) \mid V_X \in \mathbb{R}^{N \times r}, \text{ rank } V_X = r, \; V_X^T V_X = \delta(V_XV_X^T), \; \mathcal{R}(V_X) \subseteq \mathcal{N}(1^T)\}
\]

whereas \( \{D \in \text{EDM}^N \mid \text{rank}(VDV) \leq r\} \) is the closure of this same set;

\[
\{D \in \text{EDM}^N \mid \text{rank}(VDV) \leq r\} = \left\{D \in \text{EDM}^N \mid \text{rank}(VDV) = r\right\}
\]

For example,

\[
\text{rel } \partial \text{EDM}^N = \left\{D \in \text{EDM}^N \mid \text{rank}(VDV) < N - 1\right\}
\]

\[
= \bigcup_{r=0}^{N-2} \left\{D \in \text{EDM}^N \mid \text{rank}(VDV) = r\right\}
\]
Figure 152: (a) Vector $V_X$ from Figure 151 spirals in $\mathcal{N}(1^T) \subset \mathbb{R}^3$ decaying toward origin. (Spiral is two-dimensional in vector space $\mathbb{R}^3$.)  (b) Corresponding trajectory $D(V_X)$ on EDM cone relative boundary creates a vortex also decaying toward origin. There are two complete orbits on EDM cone boundary about axis of revolution for every single revolution of $V_X$ about origin. (Vortex is three-dimensional in isometrically isomorphic $\mathbb{R}^3$.)
None of these are necessarily convex sets, although

\[
\text{EDM}^N = \bigcup_{r=0}^{N-1} \left\{ D \in \text{EDM}^N \mid \text{rank}(VDV) = r \right\}
\]

= \left\{ D \in \text{EDM}^N \mid \text{rank}(VDV) = N - 1 \right\} \quad (1287)

\[
\text{rel int EDM}^N = \left\{ D \in \text{EDM}^N \mid \text{rank}(VDV) = N - 1 \right\}
\]

are pointed convex cones.

When cardinality \( N = 3 \) and affine dimension \( r = 2 \), for example, the relative interior \( \text{rel int EDM}^3 \) is realized via \((1284)\). \(\text{rel int EDM}^3\)

When \( N = 3 \) and \( r = 1 \), the relative boundary of the EDM cone \( \text{dvec rel \partial EDM}^3 \) is realized in isomorphic \( \mathbb{R}^3 \) as in Figure 150d. This figure could be constructed via \((1285)\) by spiraling vector \( V_\chi \) tightly about the origin in \( \mathcal{N}(1^T) \); as can be imagined with aid of Figure 151. Vectors close to the origin in \( \mathcal{N}(1^T) \) are correspondingly close to the origin in \( \text{EDM}^N \). As vector \( V_\chi \) orbits the origin in \( \mathcal{N}(1^T) \), the corresponding EDM orbits the axis of revolution while remaining on the boundary of the circular cone \( \text{dvec rel \partial EDM}^3 \). (Figure 152)

6.4.3 Faces of EDM cone

Like the positive semidefinite cone, EDM cone faces are EDM cones.

6.4.3.0.1 Exercise. Isomorphic faces.

Prove that in high cardinality \( N \), any set of EDMs made via \((1284)\) or \((1285)\) with particular affine dimension \( r \) is isomorphic with any set admitting the same affine dimension but made in lower cardinality.

\[
\dim \mathcal{F} \left( \text{EDM}^N \ni D(V_\chi) \right) = (r + 1)r/2 \quad (1289)
\]

\text{6.6} The fact that the smallest face is isomorphic with another EDM cone (perhaps smaller than \( \text{EDM}^N \)) is implicit in [195, \S2].
in isomorphic $\mathbb{R}^{N(N-1)/2}$. Not all dimensions are represented; e.g., the EDM cone has no two-dimensional faces.

When cardinality $N=4$ and affine dimension $r=2$ so that $\mathcal{R}(V_{\mathcal{X}})$ is any two-dimensional subspace of three-dimensional $\mathcal{N}(1^T)$ in $\mathbb{R}^4$, for example, then the corresponding face of $\text{EDM}^4$ is isometrically isomorphic with: (1285)

$$\text{EDM}^3 = \{ D \in \text{EDM}^3 \mid \text{rank}(VDV) \leq 2 \} \simeq \mathcal{F}(\text{EDM}^4 \ni D(V_{\mathcal{X}})) \quad (1290)$$

Each two-dimensional subspace of $\mathcal{N}(1^T)$ corresponds to another three-dimensional face.

Because each and every principal submatrix of an EDM in $\text{EDM}^N$ ([5.14.3] is another EDM [245, §4.1], for example, then each principal submatrix belongs to a particular face of $\text{EDM}^N$.

### 6.4.3.2 extreme directions of EDM cone

In particular, extreme directions ([2.8.1] of $\text{EDM}^N$ correspond to affine dimension $r=1$ and are simply represented: for any particular cardinality $N \geq 2$ ([2.8.2] and each and every nonzero vector $z$ in $\mathcal{N}(1^T)$

$$\Gamma \triangleq (z \circ z)1^T + 1(z \circ z)^T - 2zz^T \in \text{EDM}^N$$

$$= \delta(zz^T)1^T + 11^T \delta(zz^T) - 2zz^T \quad (1291)$$

is an extreme direction corresponding to a one-dimensional face of the EDM cone $\text{EDM}^N$ that is a ray in isomorphic subspace $\mathbb{R}^{N(N-1)/2}$.

Proving this would exercise the fundamental definition (186) of extreme direction. Here is a sketch: Any EDM may be represented

$$D(V_{\mathcal{X}}) = \delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)1^T + 11^T \delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) - 2V_{\mathcal{X}}V_{\mathcal{X}}^T \in \text{EDM}^N \quad (1267)$$

where matrix $V_{\mathcal{X}}$ (1268) has orthogonal columns. For the same reason (1549) that $zz^T$ is an extreme direction of the positive semidefinite cone ([2.9.2.7) for any particular nonzero vector $z$, there is no conic combination of distinct EDMs (each conically independent of $\Gamma$ ([2.10]) equal to $\Gamma$. $\blacksquare$

### 6.4.3.2.1 Example. Biorthogonal expansion of an EDM. (confer §2.13.7.1.1)

When matrix $D$ belongs to the EDM cone, nonnegative coordinates for biorthogonal expansion are the eigenvalues $\lambda \in \mathbb{R}^N$ of $-VDV_\frac{1}{2}$: For any $D \in \mathbb{S}_+^N$ it holds

$$D = \delta(-VDV_\frac{1}{2})1^T + 11^T \delta(-VDV_\frac{1}{2}) - 2(-VDV_\frac{1}{2}) \quad (1080)$$

By diagonalization $-VDV_\frac{1}{2} \triangleq QAQ^T \in \mathbb{S}^N_+$ ([A.5.1]) we may write

$$D = \delta\left( \sum_{i=1}^{N} \lambda_i q_i q_i^T \right)1^T + 11^T \delta\left( \sum_{i=1}^{N} \lambda_i q_i q_i^T \right) - 2\sum_{i=1}^{N} \lambda_i q_i q_i^T$$

$$= \sum_{i=1}^{N} \lambda_i \left( \delta(q_i q_i^T)1^T + 11^T \delta(q_i q_i^T) - 2q_i q_i^T \right) \quad (1292)$$
where \( q_i \) is the \( i \)th eigenvector of \(-VDV \frac{1}{2}\) arranged columnar in orthogonal matrix

\[
Q = [q_1 \ q_2 \ \cdots \ q_N] \in \mathbb{R}^{N \times N}
\] (399)

and where \( \{\delta(q_i q_i^T)1^T + 1\delta(q_i q_i^T)^T - 2q_i q_i^T, \ i = 1 \ldots N\} \) are extreme directions of some pointed polyhedral cone \( K \subset S^N_+ \) and extreme directions of \( \text{EDM}^N \). Invertibility of (1292)

\[
-VDV \frac{1}{2} = -V \sum_{i=1}^{N} \lambda_i (\delta(q_i q_i^T)1^T + 1\delta(q_i q_i^T)^T - 2q_i q_i^T) V \frac{1}{2}
\] (1293)

implies linear independence of those extreme directions. Then biorthogonal expansion is expressed

\[
dvec D = YY^\dagger dvec D = Y \lambda (-VDV \frac{1}{2})
\] (1294)

where

\[
Y \triangleq [dvec(\delta(q_i q_i^T)1^T + 1\delta(q_i q_i^T)^T - 2q_i q_i^T), \ i = 1 \ldots N] \in \mathbb{R}^{N(N-1)/2 \times N}
\] (1295)

When \( D \) belongs to the EDM cone in the subspace of symmetric hollow matrices, unique coordinates \( Y^\dagger dvec D \) for this biorthogonal expansion must be the nonnegative eigenvalues \( \lambda \) of \(-VDV \frac{1}{2}\). This means \( D \) simultaneously belongs to the EDM cone and to the pointed polyhedral cone \( dvec K = \text{cone}(Y) \).

### 6.4.3.3 open question

Result (1289) is analogous to that for the positive semidefinite cone (222), although the question remains open whether all faces of \( \text{EDM}^N \) (whose dimension is less than dimension of the cone) are exposed like they are for the positive semidefinite cone.\(^{6,7}\) (§2.9.2.3) [358]

### 6.5 Correspondence to PSD cone \( S^{N-1}_+ \)

Hayden, Wells, Liu, & Tarazaga [195, §2] assert one-to-one correspondence of EDMs with positive semidefinite matrices in the symmetric subspace. Because \( \text{rank}(VDV) \leq N - 1 \) (§5.7.1.1), that PSD cone corresponding to the EDM cone can only be \( S^{N-1}_+ \). [9, §18.2.1]

To clearly demonstrate this correspondence, we invoke inner-product form EDM definition

\[
D(\Phi) = \begin{bmatrix} 0 & \delta(\Phi) \end{bmatrix} 1^T + 1 \begin{bmatrix} 0 & \delta(\Phi)^T \end{bmatrix} - 2 \begin{bmatrix} 0 & 0^T \Phi \end{bmatrix} \in \text{EDM}^N
\]

\(
\Phi \succeq 0
\)

(1098)

Then the EDM cone may be expressed

\[
\text{EDM}^N = \left\{ D(\Phi) \mid \Phi \in S^{N-1}_+ \right\}
\] (1296)

\(^6,7\)Elementary example of face not exposed is given by the closed convex set in Figure 34 and in Figure 44.
Hayden & Wells’ assertion can therefore be equivalently stated in terms of an inner-product form EDM operator

\[ D(S^N_{+}^{-1}) = \text{EDM}^N \]  \hspace{1cm} (1100)

\[ V_N' (\text{EDM}^N) = S^N_{+}^{-1} \]  \hspace{1cm} (1101)

identity (1101) holding because \( R(V_N) = \mathcal{N}(1^T) \) (976), linear functions \( D(\Phi) \) and \( V_N(D) = -V_N^T DV_N' \) (§5.6.2.1) being mutually inverse.

In terms of affine dimension \( r \), Hayden & Wells claim particular correspondence between PSD and EDM cones:

\( r = N - 1 \): Symmetric hollow matrices \(-D\) positive definite on \( \mathcal{N}(1^T) \) correspond to points relatively interior to the EDM cone.

\( r < N - 1 \): Symmetric hollow matrices \(-D\) positive semidefinite on \( \mathcal{N}(1^T) \), where \(-V_N^T DV_N'\) has at least one 0 eigenvalue, correspond to points on the relative boundary of the EDM cone.

\( r = 1 \): Symmetric hollow nonnegative matrices rank-one on \( \mathcal{N}(1^T) \) correspond to extreme directions (1291) of the EDM cone; \( \text{id est} \), for some nonzero vector \( u \) (¶A.3.1.0.7)

\[ \text{rank} V_N^T DV_N' = 1 \]
\[ D \in S_h^N \cap \mathbb{R}_+^{N \times N} \] \( \Leftrightarrow \) \( D \in \text{EDM}^N \) \( \Leftrightarrow \) \( -V_N^T DV_N' \equiv uu^T \) \( D \in S_h^N \) \hspace{1cm} (1297)

6.5.0.0.1 Proof. Case \( r = 1 \) is easily proved: From the nonnegativity development in §5.6.1, extreme direction (1291), and Schoenberg criterion (988), we need show only sufficiency; \( \text{id est} \), prove

\[ \text{rank} V_N^T DV_N' = 1 \]
\[ D \in S_h^N \cap \mathbb{R}_+^{N \times N} \] \( \Rightarrow \) \( D \in \text{EDM}^N \) \Rightarrow \( D \) is an extreme direction

Any symmetric matrix \( D \) satisfying the rank condition must have the form, for \( z, q \in \mathbb{R}^N \) and nonzero \( z \in \mathcal{N}(1^T) \),

\[ D = \pm (1q^T + q1^T - 2zz^T) \] \hspace{1cm} (1298)

because (§5.6.2.1, confer ¶E.7.2.0.2)

\[ \mathcal{N}(V_N(D)) = \{ 1q^T + q1^T | q \in \mathbb{R}^N \} \subseteq S^N \] \hspace{1cm} (1299)

Hollowness demands \( q = \delta(zz^T) \) while nonnegativity demands choice of positive sign in (1298). Matrix \( D \) thus takes the form of an extreme direction (1291) of the EDM cone.

The foregoing proof is not extensible in rank: An EDM with corresponding affine dimension \( r \) has the general form, for \( \{ z_i \in \mathcal{N}(1^T), i = 1 \ldots r \} \) an independent set,

\[ D = 1\delta \left( \sum_{i=1}^r z_i z_i^T \right) + \delta \left( \sum_{i=1}^r z_i z_i^T \right) 1^T - 2 \sum_{i=1}^r z_i z_i^T \in \text{EDM}^N \] \hspace{1cm} (1300)
6.5. CORRESPONDENCE TO PSD CONE $S_{++}^{N-1}$

The EDM so defined relies principally on the sum $\sum z_i z_i^T$ having positive summand coefficients ($\leftrightarrow -V_N^T D V_N \succeq 0$). Then it is easy to find a sum incorporating negative coefficients while meeting rank, nonnegativity, and symmetric hollowness conditions but not positive semidefiniteness on subspace $R(V_N)$; e.g., from page 416,

$$-V \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 5 \\ 1 & 5 & 0 \end{bmatrix} V_2^1 = z_1 z_1^T - z_2 z_2^T$$

(1301)

6.5.0.0.2 Example. Extreme rays versus rays on the boundary.

The EDM $D = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ 4 & 1 & 0 \end{bmatrix}$ is an extreme direction of $\text{EDM}^3$ where $u = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ in (1297).

Because $-V_N^T D V_N$ has eigenvalues $\{0, 5\}$, the ray whose direction is $D$ also lies on the relative boundary of $\text{EDM}^3$.

In exception, EDM $D = \kappa \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, for any particular $\kappa > 0$, is an extreme direction of $\text{EDM}^2$ but $-V_N^T D V_N$ has only one eigenvalue: $\{\kappa\}$. Because $\text{EDM}^2$ is a ray whose relative boundary (§2.6.1.4.1) is the origin, this conventional boundary does not include $D$ which belongs to the relative interior in this dimension. (§2.7.0.0.1) \(\square\)

6.5.1 Gram-form correspondence to $S_{++}^{N-1}$

With respect to $D(G) = \delta(G) 1^T + 1 \delta(G)^T - 2G$ (981) the linear Gram-form EDM operator, results in §5.6.1 provide \[3, \S2.6\]

$$\text{EDM}^N = D\left(V(\text{EDM}^N)\right) = D\left(V_N S_{++}^{N-1} V_N^T\right)$$

(1302)

$$V_N S_{++}^{N-1} V_N^T \equiv V\left(D\left(V_N S_{++}^{N-1} V_N^T\right)\right) = V(\text{EDM}^N) \triangleq -V \text{EDM}^N V_2^1 = S_{c}^N \cap S_{+}^N$$

(1303)

a one-to-one correspondence between $\text{EDM}^N$ and $S_{++}^{N-1}$.

6.5.2 EDM cone by ellipope

Having defined the ellipope parametrized by scalar $t > 0$

$$E^N_t = S_{++}^N \cap \{\Phi \in S^N \mid \delta(\Phi) = t 1\}$$

(1178)

then following Alfakih \[10\] we have

$$\text{EDM}^N = \text{cone}\{11^T - E^N_t\} = \{t(11^T - E^N_t) \mid t \geq 0\}$$

(1304)

Identification $E^N = E^N_t$ equates the standard ellipope (§5.9.1.0.1, Figure 142) to our parametrized ellipope.

6.8 (⇒) For $a_i \in \mathbb{R}^{N-1}$, let $z_i = V_N^T a_i$. 

\[ \text{dvec rel } \partial \text{EDM}^3 \]

\[ \text{EDM}^N = \text{cone}\{\mathbf{11}^T - \mathcal{E}^N\} = \{t(\mathbf{11}^T - \mathcal{E}^N) \mid t \geq 0\} \] (1304)

Figure 153: Three views of translated negated elliptope \( \mathbf{11}^T - \mathcal{E}^3 \) (confer Figure 142) shrouded by truncated EDM cone. Fractal on EDM cone relative boundary is numerical artifact belonging to intersection with elliptope relative boundary. The fractal is trying to convey existence of a neighborhood about the origin where the translated elliptope boundary and EDM cone boundary intersect.
6.5. CORRESPONDENCE TO PSD CONE $S^{N-1}$

6.5.2.0.1 Expository. Normal cone, tangent cone, elllipse.
Define $T_\mathcal{E}(\mathbf{1}^T)$ to be the tangent cone to the elllipse $\mathcal{E}$ at point $\mathbf{1}^T$; id est,

$$T_\mathcal{E}(\mathbf{1}^T) \triangleq \{ t(\mathcal{E} - \mathbf{1}^T) \mid t \geq 0 \}$$  \hfill (1305)

The normal cone $K_\perp E(\mathbf{1}^T)$ to the elllipse at $\mathbf{1}^T$ is a closed convex cone defined ($\S$E.10.3.2.1, Figure 188)

$$K_\perp E(\mathbf{1}^T) \triangleq \{ B \mid \langle B, \Phi - \mathbf{1}^T \rangle \leq 0, \Phi \in \mathcal{E} \}$$  \hfill (1306)

The polar cone of any set $\mathcal{K}$ is the closed convex cone (confer (296))

$$\mathcal{K}^\circ \triangleq \{ B \mid \langle B, A \rangle \leq 0, \text{ for all } A \in \mathcal{K} \}$$  \hfill (1307)

The normal cone is well known to be the polar of the tangent cone,

$$K_\perp E(\mathbf{1}^T) = T_\mathcal{E}(\mathbf{1}^T)^\circ$$ \hfill (1308)

and vice versa; [209, §A.5.2.4]

$$K_\perp E(\mathbf{1}^T)^\circ = T_\mathcal{E}(\mathbf{1}^T)$$  \hfill (1309)

From Deza & Laurent [119, p.535] we have the EDM cone

$$\text{EDM} = -T_\mathcal{E}(\mathbf{1}^T)$$  \hfill (1310)

The polar EDM cone is also expressible in terms of the elllipse. From (1308) we have

$$\text{EDM}^\circ = -K_\perp E(\mathbf{1}^T)$$  \hfill (1311)

In $\S$5.10.1 we proposed the expression for EDM $D$

$$D = t\mathbf{1}^T - \mathcal{E} \in \text{EDM}^N$$ \hfill (1179)

where $t \in \mathbb{R}_+$ and $\mathcal{E}$ belongs to the parametrized elllipse $\mathcal{E}^N$. We further propose, for any particular $t > 0$

$$\text{EDM}^N = \text{cone}\{t\mathbf{1}^T - \mathcal{E}^N_t \}$$ \hfill (1312)

Proof (pending). \hfill ■

6.5.2.0.2 Exercise. EDM cone from elllipse.
Relationship of the translated negated elllipse with the EDM cone is illustrated in Figure 153. Prove whether it holds that

$$\text{EDM}^N = \lim_{t \to \infty} t\mathbf{1}^T - \mathcal{E}^N_t$$ \hfill (1313)

▼
6.6 Vectorization & projection interpretation

In §E.7.2.0.2 we learn: $-VDV$ can be interpreted as orthogonal projection \[7\] of vectorized $-D \in \mathbb{S}_h^N$ on the subspace of geometrically centered symmetric matrices

$$\mathbb{S}_c^N = \{G \in \mathbb{S}_h^N \mid G1 = 0\}$$
$$= \{G \in \mathbb{S}_h^N \mid \mathcal{N}(G) \supseteq 1\} = \{G \in \mathbb{S}_h^N \mid \mathcal{R}(G) \subseteq \mathcal{N}(1^T)\}$$
$$= \{VYV^T \mid Y \in \mathbb{S}_h^N\} \subset \mathbb{S}_h^N$$
$$\equiv \{V_NAV_N^T \mid A \in \mathbb{S}_h^{N-1}\}$$

(1071)

because elementary auxiliary matrix $V$ is an orthogonal projector (§B.4.1). Yet there is another useful projection interpretation:

Revising the fundamental matrix criterion for membership to the EDM cone (964), 6.9

$$\langle zz^T, -D \rangle \geq 0 \quad \forall \ z \in \mathbb{R}^N$$
$$\implies D \in \mathcal{R}(V_N)$$

(988)

this is equivalent, of course, to the Schoenberg criterion

$$-V_N^T D V_N \succeq 0 \quad \forall \ z \in \mathbb{R}^N$$

(1314)

$$\iff D \in \mathcal{R}(V_N)$$

because $\mathcal{N}(11^T) = \mathcal{R}(V_N)$. When $D \in \mathcal{R}(V_N)$, correspondence (1314) means $-z^TDz$ is proportional to a nonnegative coefficient of orthogonal projection (§E.6.4.2, Figure 155) of $-D$ in isometrically isomorphic $\mathbb{R}^{N(N+1)/2}$ on the range of each and every vectorized (§2.2.2.1) symmetric dyad (§B.1) in the nullspace of $11^T$; id est, on each and every member of

$$T \triangleq \{svec(zz^T) \mid z \in \mathcal{R}(11^T) = \mathcal{R}(V_N)\} \subset \text{svec } \mathbb{S}_h^N$$

$$= \{svec(V_N \psi \psi^T V_N^T) \mid \psi \in \mathbb{R}^{N-1}\}$$

(1315)

whose dimension is

$$\dim T = N(N - 1)/2$$

(1316)

The set of all symmetric dyads $\{zz^T \mid z \in \mathbb{R}^N\}$ constitute the extreme directions of the positive semidefinite cone (§2.8.1, §2.9) $\mathbb{S}_h^N$, hence lie on its boundary. Yet only those dyads in $\mathcal{R}(V_N)$ are included in the test (1314), thus only a subset $T$ of all vectorized extreme directions of $\mathbb{S}_h^N$ is observed.

In the particularly simple case $D = \mathcal{R}(11^T) = \mathcal{N}(11^T)$, for example, only one extreme direction of the PSD cone is involved:

$$zz^T = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

(1317)

Any nonnegative scaling of vectorized $zz^T$ belongs to the set $T$ illustrated in Figure 154 and Figure 155.

$^6.9 \mathcal{N}(11^T) = \mathcal{N}(1^T)$ and $\mathcal{R}(zz^T) = \mathcal{R}(z)$
6.6. VECTORIZATION & PROJECTION INTERPRETATION

\[
\begin{bmatrix}
d_{11} & d_{12} \\
d_{12} & d_{22}
\end{bmatrix}
\]

Figure 154: Truncated boundary of positive semidefinite cone \( S^2_+ \) in isometrically isomorphic \( \mathbb{R}^3 \) (via \( \text{svec} \)) is, in this dimension, constituted solely by its extreme directions. Truncated cone of Euclidean distance matrices \( \text{EDM}^2 \) in isometrically isomorphic subspace \( \mathbb{R} \). Relative boundary of EDM cone is constituted solely by matrix 0. Halfline \( T = \{ \kappa^2 [1 \ -\sqrt{2} \ 1]^T \ | \ \kappa \in \mathbb{R} \} \) on PSD cone boundary depicts that lone extreme ray (1317) on which orthogonal projection of \(-D\) must be positive semidefinite if \( D \) is to belong to \( \text{EDM}^2 \). aff cone \( \mathcal{T} = \text{svec} S^2_+ \). (1322) Dual EDM cone is halfspace in \( \mathbb{R}^3 \) whose bounding hyperplane has inward-normal \( \text{svec} \text{EDM}^2 \).
\[
\begin{bmatrix}
    d_{11} & d_{12} \\
    d_{12} & d_{22}
\end{bmatrix}
\]

\[svec D \partial S_2^+\]

\[\langle zz^T, -D \rangle \]

\[\langle zz^T, zz^T \rangle zz^T\]

\[zz^T\text{ is projection of vectorized } -D \text{ on range of vectorized } zz^T.\]

\[D \in EDM_N \Leftrightarrow \left\{ \begin{array}{l}
    \langle zz^T, -D \rangle \geq 0 \quad \forall zz^T \\
    D \in S_h^N
\end{array} \right. \quad (1314)\]

Figure 155: Given-matrix \( D \) is assumed to belong to symmetric hollow subspace \( S_h^2 \); a line in this dimension. Negative \( D \) is found along \( S_h^2 \). Set \( T \) (1315) has only one ray member in this dimension; not orthogonal to \( S_h^2 \). Orthogonal projection of \(-D\) on \( T \) (indicated by half dot) has nonnegative projection coefficient. Matrix \( D \) must therefore be an EDM.
6.6. VECTORIZATION & PROJECTION INTERPRETATION

6.6.1 Face of PSD cone $S^N_+$ containing $V$

In any case, set $T$ (1315) constitutes the vectorized extreme directions of an $N(N-1)/2$-dimensional face of the PSD cone $S^N_+$ containing auxiliary matrix $V$; a face isomorphic with $S^N_+ = S^\operatorname{rank} V$ (§2.9.2.3).

To show this, we must first find the smallest face that contains auxiliary matrix $V$ and then determine its extreme directions. From (221),

$$F(S^N_+ \ni V) = \{ W \in S^N_+ | N(W) \supseteq N(V) \} = \{ W \in S^N_+ | N(W) \supseteq 1 \}$$

$$= \{ VYV \geq 0 | Y \in S^N \} \equiv \{ V_N B V_N^T | B \in S^N_{-1} \}$$

$$\simeq S^\operatorname{rank} V = -V_N^T \operatorname{EDM}^N V_N \tag{1318}$$

where the equivalence $\equiv$ is from §5.6.1 while isomorphic equality $\simeq$ with transformed EDM cone is from (1101). Projector $V$ belongs to $F(S^N_+ \ni V)$ because $V_N V_N^T V_N^T = V$ (§B.4.3). Each and every rank-one matrix belonging to this face is therefore of the form:

$$V_N^T V_N^T v v^T \ | \ v \in \mathbb{R}^{N-1} \tag{1319}$$

Because $F(S^N_+ \ni V)$ is isomorphic with a positive semidefinite cone $S^N_{-1}$, then $T$ constitutes the vectorized extreme directions of $F$, the origin constitutes the extreme points of $F$, and auxiliary matrix $V$ is some convex combination of those extreme points and directions by the extremes theorem (§2.8.1.1.1).

In fact the smallest face, that contains auxiliary matrix $V$, of the PSD cone $S^N_+$ is the intersection with the geometric center subspace (2095) (2096);

$$F(S^N_+ \ni V) = \text{cone}\left\{ V_N^T v v^T \ | \ v \in \mathbb{R}^{N-1} \right\}$$

$$= S^N_+ \cap S^N_+ \tag{1320}$$

$$\equiv \{ X \succeq 0 | \langle X, 1 1^T \rangle = 0 \} \quad \text{ (1677)}$$

In isometrically isomorphic $\mathbb{R}^{N(N+1)/2}$

$$\text{svec } F(S^N_+ \ni V) = \text{cone } T \tag{1321}$$

related to $S^N_c$ by

$$\text{aff cone } T = \text{svec } S^N_c \tag{1322}$$

6.6.2 EDM criteria in $11^T$

(see §6.4, (995)) Laurent specifies an elliptope trajectory condition for EDM cone membership: [245, §2.3]

$$D \in \operatorname{EDM}^N \iff [1 - e^{-\alpha d_{ij}}] \in \operatorname{EDM}^N \forall \alpha > 0 \quad \text{(1173a)}$$
From the parametrized elliptope $\mathcal{E}_t^N$ in §6.5.2 and §5.10.1 we propose

$$D \in \text{EDM}^N \iff \exists t \in \mathbb{R}_+ \in \mathcal{E}_t^N \ni D = t11^T - \mathcal{E}$$

(1323)

Chabrillac & Crouzeix [76, §4] prove a different criterion they attribute to Finsler, 1937 [151]. We apply it to EDMs: for $D \in \mathbb{S}_h^N$ (1121)

$$-V_N^T D V_N \succ 0 \iff \exists \kappa > 0 \ni -D + \kappa 11^T \succ 0$$

$$\iff D \in \text{EDM}^N \text{ with corresponding affine dimension } r = N - 1$$

(1324)

This Finsler criterion has geometric interpretation in terms of the vectorization & projection already discussed in connection with (1314). With reference to Figure 154, the offset $11^T$ is simply a direction orthogonal to $T$ in isomorphic $\mathbb{R}^3$. Intuitively, translation of $-D$ in direction $11^T$ is like orthogonal projection on $T$ insofar as similar information can be obtained.

When the Finsler criterion (1324) is applied despite lower affine dimension, the constant $\kappa$ can go to infinity making the test $-D + \kappa 11^T \succeq 0$ impractical for numerical computation. Chabrillac & Crouzeix invent a criterion for the semidefinite case, but is no more practical: for $D \in \mathbb{S}_h^N$

$$D \in \text{EDM}^N$$

$$\exists \kappa > 0 \ni \forall \kappa \geq \kappa_p , -D - \kappa 11^T \text{ [sic] has exactly one negative eigenvalue}$$

(1325)

### 6.7 A geometry of completion

*It is not known how to proceed if one wishes to restrict the dimension of the Euclidean space in which the configuration of points may be constructed.*

---

Michael W. Trosset, 2000 [364, §1]

Given an incomplete noiseless EDM, intriguing is the question of whether a list in $X \in \mathbb{R}^{n \times N}$ (76) may be reconstructed and under what circumstances reconstruction is unique. [3] [5] [6] [7] [9] [18] [69] [215] [227] [244] [245] [246]

If one or more entries of a particular EDM are fixed, then geometric interpretation of the feasible set of completions is the intersection of the EDM cone $\text{EDM}^N$ in isomorphic subspace $\mathbb{R}^{N(N-1)/2}$ with as many hyperplanes as there are fixed symmetric entries.6.10

Assuming a nonempty intersection, then the number of completions is generally infinite, and those corresponding to particular affine dimension $r < N - 1$ belong to some generally nonconvex subset of that intersection (confer §2.9.2.9.2) that can be as small as a point.

6.10 Depicted in Figure 156a is an intersection of the EDM cone $\text{EDM}^3$ with a single hyperplane representing the set of all EDMs having one fixed symmetric entry. This representation is practical because it is easily combined with or replaced by other convex constraints; e.g. slab inequalities in (782) that admit bounding of noise processes.
Figure 156: (a) In isometrically isomorphic subspace $\mathbb{R}^3$, intersection of EDM$^3$ with hyperplane $\partial H$ representing one fixed symmetric entry $d_{23} = \kappa$ (both drawn truncated, rounded vertex is artifact of plot). EDMs in this dimension corresponding to affine dimension 1 comprise relative boundary of EDM cone (§6.5). Since intersection illustrated includes a nontrivial subset of cone’s relative boundary, then it is apparent there exist infinitely many EDM completions corresponding to affine dimension 1. In this dimension it is impossible to represent a unique nonzero completion corresponding to affine dimension 1, for example, using a single hyperplane because any hyperplane supporting relative boundary at a particular point $\Gamma$ contains an entire ray $\{\zeta \Gamma \mid \zeta \geq 0\}$ belonging to rel $\partial$EDM$^3$ by Lemma 2.8.0.0.1. (b) $d_{13} = \kappa$. (c) $d_{12} = \kappa$. 
CHAPTER 6. CONE OF DISTANCE MATRICES

Figure 157: One dimensionless EDM subgraph completion (solid) superimposed on (but not obscuring) neighborhood graph (dashed). Local view of a few dense samples \( \bigcirc \) from relative interior of some arbitrary Euclidean manifold whose affine dimension appears two-dimensional in this neighborhood. All line segments measure absolute distance. Dashed line segments help visually locate nearest neighbors; suggesting, best number of nearest neighbors can be greater than value of embedding dimension after topological transformation (confer [222, §2]). Solid line segments represent completion of EDM subgraph from available distance data for an arbitrarily chosen sample and its nearest neighbors. Each distance from EDM subgraph becomes distance-square in corresponding EDM submatrix.
6.7. A GEOMETRY OF COMPLETION

6.7.0.0.1 Example. Maximum variance unfolding.

A process minimizing affine dimension (§2.1.5) of certain kinds of Euclidean manifold by topological transformation can be posed as a completion problem (confer §E.10.2.1.2). Weinberger & Saul, who originated the technique, specify an applicable manifold in three dimensions by analogy to an ordinary sheet of paper (confer §2.1.6); imagine, we find it deformed from flatness in some way introducing neither holes, tears, or selfintersections. The physical process is intuitively described as unfolding, unfolding, diffusing, decompacting, or unraveling. In particular instances, the process is a sort of flattening by stretching until taut (but not by crushing); e.g., unfolding a three-dimensional Euclidean body resembling a billowy national flag reduces that manifold’s affine dimension to \( r = 2 \).

Data input to the proposed process originates from distances between neighboring relatively dense samples of a given manifold. Figure 157 realizes a densely sampled neighborhood; called, neighborhood graph. Essentially, the algorithmic process preserves local isometry between nearest neighbors allowing distant neighbors to excurse expansively by “maximizing variance” (Figure 6). The common number of nearest neighbors to each sample is a data-dependent algorithmic parameter whose minimum value connects the graph. The dimensionless EDM subgraph between each sample and its nearest neighbors is completed from available data and included as input; one such EDM subgraph completion is drawn superimposed upon the neighborhood graph in Figure 157. The consequent dimensionless EDM graph comprising all the subgraphs is incomplete, in general, because the neighbor number is relatively small; incomplete even though it is a superset of the neighborhood graph. Remaining distances (those not graphed at all) are squared then made variables within the algorithm; it is this variability that admits unfurling.

To demonstrate, consider untying the trefoil knot drawn in Figure 158a. A corresponding EDM \( D = [d_{ij}, i, j = 1 \ldots N] \) employing only two nearest neighbors is banded having the incomplete form

\[
D = \begin{bmatrix}
0 & \hat{d}_{12} & \hat{d}_{13} & ? & \cdots & ? & \hat{d}_{1,N-1} & \hat{d}_{1,N} \\
\hat{d}_{12} & 0 & \hat{d}_{23} & \hat{d}_{24} & \ddots & ? & ? & \hat{d}_{2N} \\
\hat{d}_{13} & \hat{d}_{23} & 0 & \hat{d}_{34} & \ddots & ? & ? & ? \\
? & \hat{d}_{24} & \hat{d}_{34} & 0 & \ddots & \ddots & ? & ? \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & ? \\
? & ? & ? & \ddots & \ddots & 0 & \hat{d}_{N-2,N-1} & \hat{d}_{N-2,N} \\
\hat{d}_{1,N-1} & ? & ? & \cdots & \hat{d}_{N-2,N-1} & 0 & \hat{d}_{N-1,N} \\
\hat{d}_{1,N} & \hat{d}_{2N} & ? & ? & \hat{d}_{N-2,N} & \hat{d}_{N-1,N} & 0
\end{bmatrix}
\]  

(1326)

where \( \hat{d}_{ij} \) denotes a given fixed distance-square. The unfurling algorithm can be expressed

[6.11] Local reconstruction of point position, from the EDM submatrix corresponding to a complete dimensionless EDM subgraph, is unique to within an isometry (§5.6, §5.12).
Figure 158: (a) Trefoil knot in $\mathbb{R}^3$ from Weinberger & Saul [404]. (b) Topological transformation algorithm employing four nearest neighbors and $N = 539$ samples reduces affine dimension of knot to $r = 2$. Choosing instead two nearest neighbors would make this embedding more circular.
6.7. A GEOMETRY OF COMPLETION

as an optimization problem; constrained total distance-square maximization:

\[
\begin{align*}
\text{maximize} & \quad \langle -V, D \rangle \\
\text{subject to} & \quad \langle D, e_i e_j^T + e_j e_i^T \rangle \frac{1}{2} = \tilde{d}_{ij} \quad \forall (i, j) \in \mathcal{I} \\
\text{rank}(VDV) & = 2 \\
D & \in \text{EDM}^N
\end{align*}
\]

where \( e_i \in \mathbb{R}^N \) is the \( i \text{th} \) member of the standard basis, where set \( \mathcal{I} \) indexes the given distance-square data like that in (1326), where \( V \in \mathbb{R}^{N \times N} \) is the geometric centering matrix (§B.4.1), and where

\[
\langle -V, D \rangle = \text{tr}(-VDV) = 2 \text{tr} G = \frac{1}{N} \sum_{i,j} d_{ij}
\]

where \( G \) is the Gram matrix producing \( D \) assuming \( G1 = 0 \).

If the (rank) constraint on affine dimension is ignored, then problem (1327) becomes convex, a corresponding solution \( D^* \) can be found, and a nearest rank-2 solution is then had by ordered eigenvalue decomposition of \(-VD^*V\) followed by spectral projection (§7.1.3) on \([\mathbb{R}^2_+ \subset \mathbb{R}^N\). This two-step process is necessarily suboptimal. Yet because the decomposition for the trefoil knot reveals only two dominant eigenvalues, the spectral projection is nearly benign. Such a reconstruction of point position (§5.12) utilizing four nearest neighbors is drawn in Figure 158b; a low-dimensional embedding of the trefoil knot.

This problem (1327) can, of course, be written equivalently in terms of Gram matrix \( G \), facilitated by (999); \textit{videlicet}, for \( \Phi_{ij} \) as in (967)

\[
\begin{align*}
\text{maximize} & \quad \langle I, G \rangle \\
\text{subject to} & \quad \langle G, \Phi_{ij} \rangle = \tilde{d}_{ij} \quad \forall (i, j) \in \mathcal{I} \\
\text{rank} G & = 2 \\
G & \succeq 0
\end{align*}
\]

The advantage to converting EDM to Gram is: Gram matrix \( G \) is a bridge between point list \( X \) and EDM \( D \); constraints on any or all of these three variables may now be introduced. (Example 5.4.2.2.8) Confinement to the geometric center subspace \( S_c^N \) (implicit constraint \( G1 = 0 \)) keeps \( G \) independent of its translation-invariant subspace \( S_c^{N\perp} \) (§5.5.1.1, Figure 160) so as not to become numerically unbounded.

A problem dual to \textit{maximum variance unfolding problem} (1328) (less the Gram rank constraint) has been called the \textit{fastest mixing Markov process}. That dual has simple interpretations in graph and circuit theory and in mechanical and thermal systems, explored in [350], and has direct application to quick calculation of \textit{PageRank} by search engines [241, §4]. Optimal Gram rank turns out to be tightly bounded above by minimum multiplicity of the second smallest eigenvalue of a dual optimal variable.
Figure 159: Trefoil ribbon (Kilian Weinberger). Same topological transformation algorithm as in Figure 158b with five nearest neighbors and $N = 1617$ samples.
6.8 Dual EDM cone

6.8.1 Ambient $\mathbb{S}^N$

We consider finding the ordinary dual EDM cone in ambient space $\mathbb{S}^N$ where $\mathbb{E}^N$ is pointed, closed, convex, but not full-dimensional. The set of all EDMs in $\mathbb{S}^N$ is a closed convex cone because it is the intersection of halfspaces about the origin in vectorized variable $D$ (§2.4.1.1.1, §2.7.2):

$$\mathbb{E}^N = \bigcap_{z \in \mathbb{S}^N(1^T)} \left\{ D \in \mathbb{S}^N \mid \langle e_i e_i^T, D \rangle \geq 0, \langle e_i e_j^T, D \rangle \leq 0, \langle z z^T, -D \rangle \geq 0 \right\} \quad (1329)$$

By definition (296), dual cone $K^*$ comprises each and every vector inward-normal to a hyperplane supporting convex cone $K$. The dual EDM cone in the ambient space of symmetric matrices is therefore expressible as the aggregate of every conic combination of inward-normals from (1329):

$$\mathbb{E}^N^* = \text{cone}\{e_i e_i^T, -e_j e_j^T \mid i, j = 1 \ldots N\} - \text{cone}\{zz^T \mid 11^T zz^T = 0\}$$
$$= \{ \sum_{i=1}^N \zeta_i e_i e_i^T - \sum_{j=1}^N \xi_j e_j e_j^T \mid \zeta_i, \xi_j \geq 0 \} - \text{cone}\{zz^T \mid 11^T zz^T = 0\} \quad (1330)$$

The EDM cone is not selfdual in ambient $\mathbb{S}^N$ because its affine hull belongs to a proper subspace $\text{aff } \mathbb{E}^N = \mathbb{S}^N_h$ (1331)

The ordinary dual EDM cone cannot, therefore, be pointed. (§2.13.1.1)

When $N = 1$, the EDM cone is the point at the origin in $\mathbb{R}$. Auxiliary matrix $V_N$ is empty $\{0\}$, and dual cone $\mathbb{E}^N^*$ is the real line.

When $N = 2$, the EDM cone is a nonnegative real line in isometrically isomorphic $\mathbb{R}^3$; there $\mathbb{S}^2$ is a real line containing the EDM cone. Dual cone $\mathbb{E}^2^*$ is the particular halfspace in $\mathbb{R}^3$ whose boundary has inward-normal $\mathbb{E}^2$. Diagonal matrices $\{\delta(u)\}$ in (1330) are represented by a hyperplane through the origin $\{d \mid \{0 1 0\}d = 0\}$ while the term $\text{cone}\{V_N^T v v^T V_N^T \}$ is represented by the halfline $T$ in Figure 154 belonging to the positive semidefinite cone boundary. The dual EDM cone is formed by translating the hyperplane along the negative semidefinite halfline $-T$; the union of each and every translation. (confer §2.10.2.0.1)

When cardinality $N$ exceeds 2, the dual EDM cone can no longer be polyhedral simply because the EDM cone cannot. (§2.13.1.1)
6.8.1.1 EDM cone and its dual in ambient $\mathbb{S}^N$

Consider the two convex cones

\[ K_1 \triangleq \mathbb{S}_h^N \]
\[ K_2 \triangleq \bigcap_{y \in \mathcal{N}(1^T)} \left\{ A \in \mathbb{S}^N \mid \langle yy^T, -A \rangle \geq 0 \right\} \]
\[ = \left\{ A \in \mathbb{S}^N \mid -z^T V A V z \geq 0 \quad \forall z \in \mathbb{R}^N \right\} \]
\[ = \left\{ A \in \mathbb{S}^N \mid -V A V \succeq 0 \right\} \]

so

\[ K_1 \cap K_2 = \text{EDM}^N \]  \hfill (1333)

Dual cone $K_1^* = \mathbb{S}_h^N \subseteq \mathbb{S}^N$ \hfill (72)

is the subspace of diagonal matrices. From (1330) via (313),

\[ K_2^* = \text{cone} \left\{ V_N u u^T V_N^T \mid u \in \mathbb{R}^{N-1} \right\} \subseteq \mathbb{S}^N \]  \hfill (1334)

Gaffke & Mathar \cite[§5.3]{155} observe that projection on $K_1$ and $K_2$ have simple closed forms: Projection on subspace $K_1$ is easily performed by symmetrization and zeroing the main diagonal or vice versa, while projection of $H \in \mathbb{S}^N$ on $K_2$ is

\[ P_{K_2} H = H - P_{\mathbb{S}_+^N}(V H V) \]  \hfill (1335)

**Proof.** First, we observe membership of $H - P_{\mathbb{S}_+^N}(V H V)$ to $K_2$ because

\[ P_{\mathbb{S}_+^N}(V H V) \quad - H = \left( P_{\mathbb{S}_+^N}(V H V) - V H V \right) + \left( V H V - H \right) \]  \hfill (1336)

The term $P_{\mathbb{S}_+^N}(V H V) - V H V$ necessarily belongs to the (dual) positive semidefinite cone by Theorem E.9.2.0.1. $V^2 = V$, hence

\[ -V \left( H - P_{\mathbb{S}_+^N}(V H V) \right) V \succeq 0 \]  \hfill (1337)

by Corollary A.3.1.0.5.

Next, we require

\[ \langle P_{K_2} H - H, P_{K_2} H \rangle = 0 \]  \hfill (1338)

Expanding,

\[ \langle -P_{\mathbb{S}_+^N}(V H V), H - P_{\mathbb{S}_+^N}(V H V) \rangle = 0 \]  \hfill (1339)

\[ \langle P_{\mathbb{S}_+^N}(V H V), (P_{\mathbb{S}_+^N}(V H V) - V H V) + (V H V - H) \rangle = 0 \]  \hfill (1340)

\[ \langle P_{\mathbb{S}_+^N}(V H V), (V H V - H) \rangle = 0 \]  \hfill (1341)

Product $V H V$ belongs to the geometric center subspace; \hfill (§E.7.2.0.2)

\[ V H V \in \mathbb{S}_c^N = \{ Y \in \mathbb{S}^N \mid \mathcal{N}(Y) \supseteq 1 \} \]  \hfill (1342)
Diagonalize $V H V \triangleq Q A Q^T$ (§A.5) whose nullspace is spanned by the eigenvectors corresponding to 0 eigenvalues by Theorem A.7.3.0.1. Projection of $V H V$ on the PSD cone (§7.1) simply zeroes negative eigenvalues in diagonal matrix $\Lambda$. Then

$$\mathcal{N}(P_{S^N_+}(V H V)) \supseteq \mathcal{N}(V H V) \supseteq \mathcal{N}(V) \quad (1343)$$

from which it follows:

$$P_{S^N_+}(V H V) \in S^N_c \quad (1344)$$

so $P_{S^N_+}(V H V) \perp (V H V - H)$ because $V H V - H \in S^N_{c^\perp}$.

Finally, we must have $P_{K^Z_2} H - H = -P_{S^N_+}(V H V) \in K^*_2$. Dual cone $K^*_2 = -\mathcal{F}(S^N_{c^\perp} \ni \mathcal{V})$ is the negative of the positive semidefinite cone’s smallest face that contains auxiliary matrix $V$. (§6.6.1) Thus $P_{S^N_+}(V H V) \in \mathcal{F}(S^N_{c^\perp} \ni \mathcal{V}) \iff \mathcal{N}(P_{S^N_+}(V H V)) \supseteq \mathcal{N}(V)$ (§2.9.2.3) which was already established in (1343). ♦

From results in §E.7.2.0.2, we know matrix product $V H V$ is the orthogonal projection of $H \in S^N$ on the geometric center subspace $S^N_c$. Thus the projection product

$$P_{K^Z_2} H = H - P_{S^N_+} P_{S^N_0} H \quad (1345)$$

6.8.1.1.1 Lemma. Projection on PSD cone $\cap$ geometric center subspace.

$$P_{S^N_+ \cap S^N_c} = P_{S^N_+} P_{S^N_c} \quad (1346)$$

Proof. For each and every $H \in S^N$, projection of $P_{S^N_0} H$ on the positive semidefinite cone remains in the geometric center subspace

$$S^N_c = \{ G \in S^N \mid G 1 = 0 \}
= \{ G \in S^N \mid \mathcal{N}(G) \supseteq 1 \}
= \{ G \in S^N \mid \mathcal{R}(G) \subseteq \mathcal{N}(1^T) \}
= \{ V Y V \mid Y \in S^N \} \subset S^N \quad (1071)$$

That is because: eigenvectors of $P_{S^N_0} H$ corresponding to 0 eigenvalues span its nullspace $\mathcal{N}(P_{S^N_0} H)$. (§A.7.3.0.1) To project $P_{S^N_0} H$ on the positive semidefinite cone, its negative eigenvalues are zeroed. (§7.1.2) The nullspace is thereby expanded while eigenvectors originally spanning $\mathcal{N}(P_{S^N_0} H)$ remain intact. Because the geometric center subspace is invariant to projection on the PSD cone, then the rule for projection on a convex set in a subspace governs (§E.9.5, projectors do not commute) and statement (1346) follows directly. ♦

From the lemma it follows

$$\{ P_{S^N_+} P_{S^N_0} H \mid H \in S^N \} = \{ P_{S^N_+ \cap S^N_c} H \mid H \in S^N \} \quad (1347)$$

Then from (2122)

$$-\left( S^N_c \cap S^N_+ \right)^* = \{ H - P_{S^N_+} P_{S^N_0} H \mid H \in S^N \} \quad (1348)$$
\[
\text{EDM}^2 = S_h^2 \cap (S_c^{2\perp} - S_+^2)
\]

Figure 160: A plane in isometrically isomorphic \(\mathbb{R}^3\), orthogonal complement \(S_c^{2\perp}\) (§B.2) of geometric center subspace (tiled fragment drawn) apparently supports PSD cone (rounded vertex is plot artifact). Line \(\text{svec} S_c^2 = \text{aff} \ T\) (1322), intersecting \(\text{svec} \partial S_+^2\) and drawn in Figure 154, runs along PSD cone boundary. (confer Figure 141)
$$EDM^2 = S^2_h \cap (S^2_{c} - svec \left( S^2_+ - S^2_2 \right))$$

Figure 161: EDM cone construction in isometrically isomorphic $\mathbb{R}^3$ by adding polar PSD cone to $svec S^2_{c}$. Difference $svec \left( S^2_{c} - S^2_+ \right)$ is halfspace partially bounded by $svec S^2_{c}$. EDM cone is nonnegative halfline along $svec S^2_2$ in this dimension.
From (313) we get closure of a vector sum
\[ K_2 = -\left(\mathbb{S}_c^N \cap \mathbb{S}_+^N\right) = \mathbb{S}_c^{N \perp} - \mathbb{S}_+^N \] (1349)
therefore the equality [105]
\[ \text{EDM}^N = K_1 \cap K_2 = \mathbb{S}_h^N \cap \left(\mathbb{S}_c^{N \perp} - \mathbb{S}_+^N\right) \] (1350)
whose veracity is intuitively evident, in hindsight, [92, p.109] from the most fundamental EDM definition (969).\(^6\)\(^{12}\) A realization of this construction in low dimension is illustrated in Figure 160 and Figure 161.

The dual EDM cone follows directly from (1350) by standard properties of cones (§2.13.1.1):
\[ \text{EDM}^{N^*} = K_1^* + K_2^* = \mathbb{S}_h^{N \perp} - \mathbb{S}_c^N \cap \mathbb{S}_+^N \] (1351)
which bears strong resemblance to (1330).

### 6.8.1.2 nonnegative orthant contains EDM\(^N\)

That \( \text{EDM}^N \) is a proper subset of the nonnegative orthant is not obvious from (1350). We wish to verify
\[ \text{EDM}^N = \mathbb{S}_h^N \cap \left(\mathbb{S}_c^{N \perp} - \mathbb{S}_+^N\right) \subset \mathbb{R}^{N \times N}_+ \] (1352)

While there are many ways to prove this, it is sufficient to show that all entries of the extreme directions of \( \text{EDM}^N \) must be nonnegative; \emph{id est}, for any particular nonzero vector \( z = [z_i, i=1 \ldots N] \in \mathcal{N}(1^T) \) (§6.4.3.2),
\[ \delta(zz^T)1^T + 1\delta(zz^T)^T - 2zz^T \geq 0 \] (1353)
where the inequality denotes entrywise comparison. The inequality holds because the \( i, j \)th entry of an extreme direction is squared: \((z_i - z_j)^2\).

We observe that the dyad \( 2zz^T \in \mathbb{S}_+^N \) belongs to the positive semidefinite cone, the doublet
\[ \delta(zz^T)1^T + 1\delta(zz^T)^T \in \mathbb{S}_c^{N \perp} \] (1354)
to the orthogonal complement (2097) of the geometric center subspace, while their difference is a member of the symmetric hollow subspace \( \mathbb{S}_h^N \). \(\diamondsuit\)

Here is an algebraic method to prove nonnegativity: Suppose we are given \( A \in \mathbb{S}_c^{N \perp} \) and \( B = [B_{ij}] \in \mathbb{S}_+^N \) and \( A - B \in \mathbb{S}_h^N \). Then we have, for some vector \( u \), \( A = u1^T + 1u^T = [A_{ij}] = [u_i + u_j] \) and \( \delta(B) = \delta(A) = 2u \). Positive semidefiniteness of \( B \) requires nonnegativity \( A - B \geq 0 \) because
\[ (e_i - e_j)^TB(e_i - e_j) = (B_{ii} - B_{ij}) - (B_{ji} - B_{jj}) = 2(u_i + u_j) - 2B_{ij} \geq 0 \] (1355) \(\diamondsuit\)

---

\(^6\)\(^{12}\)Formula (1350) is not a matrix criterion for membership to the EDM cone, it is not an EDM definition, and it is not an equivalence between EDM operators or an isomorphism. Rather, it is a recipe for constructing the EDM cone whole from large Euclidean bodies: the positive semidefinite cone, orthogonal complement of the geometric center subspace, and symmetric hollow subspace.
6.8. DUAL EDM CONE

6.8.1.3 Dual Euclidean distance matrix criterion

Conditions necessary for membership of a matrix $D^* \in S^N$ to the dual EDM cone $\text{EDM}_N^*$ may be derived from (1330): $D^* \in \text{EDM}_N^* \Rightarrow D^* = \delta(y) - V_N A V_N^T$ for some vector $y$ and positive semidefinite matrix $A \in S_+^{N-1}$. This in turn implies $\delta(D^* 1) = \delta(y)$. Then, for $D^* \in S^N$

$$D^* \in \text{EDM}_N^* \iff \delta(D^* 1) - D^* \succeq 0 \quad (1356)$$

where, for any symmetric matrix $D^*$

$$\delta(D^* 1) - D^* \in S_c^N \quad (1357)$$

To show sufficiency of the matrix criterion in (1356), recall Gram-form EDM operator

$$D(G) = \delta(G) 1^T + 1 \delta(G)^T - 2G \quad (981)$$

where Gram matrix $G$ is positive semidefinite by definition, and recall the selfadjointness property of the main-diagonal linear operator $\delta$ (§A.1):

$$\langle D, D^* \rangle = \langle \delta(G) 1^T + 1 \delta(G)^T - 2G, D^* \rangle = \langle G, \delta(D^* 1) - D^* \rangle \geq 0 \quad (999)$$

Assuming $\langle G, \delta(D^* 1) - D^* \rangle \geq 0$ (1567), then we have known membership relation (§2.13.2.0.1)

$$D^* \in \text{EDM}_N^* \iff \langle D, D^* \rangle \geq 0 \quad \forall D \in \text{EDM}_N \quad (1358)$$

Elegance of this matrix criterion (1356) for membership to the dual EDM cone derives from lack of any other assumptions except that $D^*$ be symmetric. Linear Gram-form EDM operator $D(Y)$ (981) has adjoint, for $Y \in S^N$

$$D^T(Y) \triangleq (\delta(Y 1) - Y) \frac{1}{2} \quad (1359)$$

Then from (1358) and (982) we have: [92, p.111]

$$\text{EDM}_N^* = \{ D^* \in S^N \mid \langle D, D^* \rangle \geq 0 \quad \forall D \in \text{EDM}_N \}
= \{ D^* \in S^N \mid \langle D(G), D^* \rangle \geq 0 \quad \forall G \in S^N_+ \}
= \{ D^* \in S^N \mid \langle G, D^T(D^*) \rangle \geq 0 \quad \forall G \in S^N_+ \}
= \{ D^* \in S^N \mid \delta(D^* 1) - D^* \succeq 0 \} \quad (1360)$$

the dual EDM cone expressed in terms of the adjoint operator. A dual EDM cone determined this way is illustrated in Figure 163.

6.8.1.3.1 Exercise. Dual EDM spectral cone.

Find a spectral cone as in §5.11.2 corresponding to $\text{EDM}_N^*$. ▼

---

6.13 Recall: Schoenberg criterion (988) for membership to the EDM cone requires membership to the symmetric hollow subspace.
$D^\circ = \delta(D^\circ \mathbf{1}) + (D^\circ - \delta(D^\circ \mathbf{1})) \in S_h^{N^\perp} \oplus S_c^N \cap S_+^N = \text{EDM}^{N^\circ}$

Figure 162: Hand-drawn abstraction of polar EDM cone $\text{EDM}^{N^\circ}$ (drawn truncated). Any member $D^\circ$ of polar EDM cone can be decomposed into two linearly independent nonorthogonal components: $\delta(D^\circ \mathbf{1})$ and $D^\circ - \delta(D^\circ \mathbf{1})$. 
6.8. DUAL EDM CONE

6.8.1.4 Nonorthogonal components of dual EDM

Now we tie construct (1351) for the dual EDM cone together with the matrix criterion (1356) for dual EDM cone membership. For any $D^* \in \mathbb{S}_N^+$ it is obvious:

$$\delta(D^* 1) \in \mathbb{S}_N^{\perp}$$  \hspace{1cm} (1361)

any diagonal matrix belongs to the subspace of diagonal matrices (67). We know when $D^* \in \text{EDM}^{N^*}$

$$\delta(D^* 1) - D^* \in \mathbb{S}_c^N \cap \mathbb{S}_+^N$$  \hspace{1cm} (1362)

this adjoint expression (1359) belongs to that face (1320) of the positive semidefinite cone $\mathbb{S}_N^+$ in the geometric center subspace. Any nonzero dual EDM

$$D^* = \delta(D^* 1) - (\delta(D^* 1) - D^*) \in \mathbb{S}_c^N \cap \mathbb{S}_+^N = \text{EDM}^{N^*}$$  \hspace{1cm} (1363)

can therefore be expressed as the difference of two linearly independent (when vectorized) nonorthogonal components (Figure 141, Figure 162).

6.8.1.5 Affine dimension complementarity

From §6.8.1.3 we have, for some $A \in \mathbb{S}_N^{-1}$ (confer (1362))

$$\delta(D^* 1) - D^* = V_N^T A V_N \in \mathbb{S}_c^N \cap \mathbb{S}_+^N$$  \hspace{1cm} (1364)

if and only if $D^*$ belongs to the dual EDM cone. Call $\text{rank}(V_N^T A V_N)$ dual affine dimension. Empirically, we find a complementary relationship in affine dimension between the projection of some arbitrary symmetric matrix $H$ on the polar EDM cone, $\mathbb{E} \text{DM}^{N^o} = - \mathbb{E} \text{DM}^{N^*}$, and its projection on the EDM cone; id est, the optimal solution of

\[
\begin{align*}
\text{minimize} & \quad \|D^o - H\|_F \\
\text{subject to} & \quad D^o - \delta(D^o 1) \succeq 0
\end{align*}
\]  \hspace{1cm} (1365)

has dual affine dimension complementary to affine dimension corresponding to the optimal solution of

\[
\begin{align*}
\text{minimize} & \quad \|D - H\|_F \\
\text{subject to} & \quad -V_N^T D V_N \succeq 0
\end{align*}
\]  \hspace{1cm} (1366)

Precisely,

$$\text{rank}(D^{o*} - \delta(D^{o*} 1)) + \text{rank}(V_N^T D^o V_N) = N - 1$$  \hspace{1cm} (1367)

This polar projection can be solved quickly (without semidefinite programming) via Lemma 6.8.1.1.1; rewriting,

\[
\begin{align*}
\text{minimize} & \quad \|D^o - \delta(D^o 1) - (H - \delta(D^o 1))\|_F \\
\text{subject to} & \quad D^o - \delta(D^o 1) \succeq 0
\end{align*}
\]

which is the projection of affinely transformed optimal solution $H - \delta(D^o 1)$ on $\mathbb{S}_c^N \cap \mathbb{S}_+^N$;

$$D^{o*} - \delta(D^{o*} 1) = P_{\mathbb{S}_c^N} P_{\mathbb{S}_+^N} (H - \delta(D^o 1))$$

Foreknowledge of an optimal solution $D^{o*}$ as argument to projection suggests recursion.
and \( \text{rank}(D^{\circ \ast} - \delta(D^{\ast \ast} \mathbf{1})) \leq N - 1 \) because vector \( \mathbf{1} \) is always in the nullspace of rank’s argument. This is similar to the known result for projection on the selfdual positive semidefinite cone and its polar:

\[
\text{rank} P_{S_+^N} H + \text{rank} P_{\mathbb{S}_+} H = N \quad (1368)
\]

When low affine dimension is a desirable result of projection on the EDM cone, projection on the polar EDM cone should be performed instead. Convex polar problem (1365) can be solved for \( D^{\circ \ast} \) by transforming to an equivalent Schur-form semidefinite program (§3.5.2). Interior-point methods for numerically solving semidefinite programs tend to produce high-rank solutions. (§4.1.2) Then \( D^{\ast} = H - D^{\circ \ast} \in \text{EDM}^N \) by Corollary E.9.2.2.1, and \( D^{\ast} \) will tend to have low affine dimension. This approach breaks when attempting projection on a cone subset discriminated by affine dimension or rank, because then we have no complementarity relation like (1367) or (1368) (§7.1.4.1).

### 6.8.1.6 EDM cone is not selfdual

In §5.6.1, via Gram-form EDM operator

\[
D(G) = \delta(G)1^T + 1\delta(G)^T - 2G \in \text{EDM}^N \quad \Leftrightarrow \quad G \succeq 0 \quad (981)
\]

we established clear connection between the EDM cone and that face (1320) of positive semidefinite cone \( S_+^N \) in the geometric center subspace:

\[
\begin{align*}
\text{EDM}^N &= D(S_c^N \cap S_+^N) \quad (1088) \\
\mathbf{V}(\text{EDM}^N) &= S_c^N \cap S_+^N \quad (1089)
\end{align*}
\]

where

\[
\mathbf{V}(D) = -VDV_1^T \quad (1077)
\]

In §5.6.1 we established

\[
S_c^N \cap S_+^N = V_N S_+^{N-1} V_N^T \quad (1075)
\]

Then from (1356), (1364), and (1330) we can deduce

\[
\delta(\text{EDM}^{N^*}) - \text{EDM}^{N^*} = V_N S_+^{N-1} V_N^T = S_c^N \cap S_+^N \quad (1369)
\]

which, by (1088) and (1089), means the EDM cone can be related to the dual EDM cone by an equality:

\[
\begin{align*}
\text{EDM}^N &= D \left( \delta(\text{EDM}^{N^*}) - \text{EDM}^{N^*} \right) \quad (1370) \\
\mathbf{V}(\text{EDM}^N) &= \delta(\text{EDM}^{N^*} \mathbf{1}) - \text{EDM}^{N^*} \quad (1371)
\end{align*}
\]

This means projection \( -\mathbf{V}(\text{EDM}^N) \) of the EDM cone on the geometric center subspace \( S_c^N \) (§E.7.2.0.2) is a linear transformation of the dual EDM cone: \( \text{EDM}^{N^*} - \delta(\text{EDM}^{N^*} \mathbf{1}) \). Secondarily, it means the EDM cone is not selfdual in \( S^N \).
6.8 DUAL EDM CONE

6.8.1.7 Schoenberg criterion is discretized membership relation

We show the Schoenberg criterion

\[-V_N^T D V_N \in S_{N-1}^N \]

\[D \in S_h^N \iff D \in \text{EDM}_N^N \quad (988)\]

to be a discretized membership relation (§2.13.4) between a closed convex cone \( K \) and its dual \( K^* \) like

\[\langle y, x \rangle \geq 0 \quad \forall \ y \in G(K^*) \iff x \in K \quad (365)\]

where \( G(K^*) \) is any set of generators whose conic hull constructs closed convex dual cone \( K^* \):

The Schoenberg criterion is the same as

\[\langle zz^T, -D \rangle \geq 0 \quad \forall \ zz^T \in \left\{ V_N^T V_N^T \mid v \in \mathbb{R}^{N-1} \right\} \]

\[D \in S_h^N \iff D \in \text{EDM}_N^N \quad (1372)\]

where the \( zz^T \) constitute a set of generators \( G \) for the positive semidefinite cone’s smallest face \( F\left(S_+^N \supset V\right) \) (§6.6.1) that contains auxiliary matrix \( V \). From the aggregate in (1330) we get the ordinary membership relation, assuming only \( D \in S^N \) [209, p.58]

\[\langle D^*, D \rangle \geq 0 \quad \forall \ D^* \in \text{EDM}_N^N \iff D \in \text{EDM}_N^N \quad (1373)\]

\[\langle D^*, D \rangle \geq 0 \quad \forall \ D^* \in \left\{ \delta(u) \mid u \in \mathbb{R}^N \right\} - \text{cone}\left\{ V_N^T V_N^T \mid v \in \mathbb{R}^{N-1} \right\} \iff D \in \text{EDM}_N^N \quad (1374)\]

Discretization (365) yields:

\[\langle D^*, D \rangle \geq 0 \quad \forall \ D^* \in \left\{ e_i e_i^T, -e_j e_j^T, -V_N^T v v^T V_N^T \mid i, j = 1 \ldots N, \ v \in \mathbb{R}^{N-1} \right\} \iff D \in \text{EDM}_N^N \quad (1375)\]

Because \( \left\{ \delta(u) \mid u \in \mathbb{R}^N \right\} \), \( D \geq 0 \iff D \in S_h^N \), we can restrict observation to the symmetric hollow subspace without loss of generality. Then for \( D \in S^N \)

\[\langle D^*, D \rangle \geq 0 \quad \forall \ D^* \in \left\{ -V_N^T v v^T V_N^T \mid v \in \mathbb{R}^{N-1} \right\} \iff D \in \text{EDM}_N^N \quad (1376)\]

this discretized membership relation becomes (1372); identical to the Schoenberg criterion.

Hitherto a correspondence between the EDM cone and a face of a PSD cone, the Schoenberg criterion is now accurately interpreted as a discretized membership relation between the EDM cone and its ordinary dual.
6.8.2 Ambient $S^N_h$

When instead we consider the ambient space of symmetric hollow matrices (1331), then still we find the EDM cone is not selfdual for $N > 2$. The simplest way to prove this is as follows:

Given a set of generators $G = \{ \Gamma \}$ (1291) for the pointed closed convex EDM cone, the discretized membership theorem in §2.13.4.2.1 asserts that members of the dual EDM cone in the ambient space of symmetric hollow matrices can be discerned via discretized membership relation:

$$\text{EDM}^N \cap S^N_h \equiv \{ D^* \in S^N_h \mid \langle \Gamma, D^* \rangle \geq 0 \ \forall \Gamma \in G(\text{EDM}^N) \}$$

$$= \{ D^* \in S^N_h \mid \langle \delta(zz^T)1^T + 1\delta(zz^T)^T - 2zz^T, D^* \rangle \geq 0 \ \forall z \in N(1^T) \}$$

By comparison

$$\text{EDM}^N = \{ D \in S^N_h \mid \langle -zz^T, D \rangle \geq 0 \ \forall z \in N(1^T) \}$$

the term $\delta(zz^T)^T D^* 1$ foils any hope of selfdualness in ambient $S^N_h$. ♦

To find the dual EDM cone in ambient $S^N_h$ per §2.13.9.4 we prune the aggregate in (1330) describing the ordinary dual EDM cone, removing any member having nonzero main diagonal:

$$\text{EDM}^N \cap S^N_h = \text{cone}\left\{ \delta^2(V_N^T V_N^T V_N^T - V_N^T V_N^T V_N^T) \mid v \in \mathbb{R}^{N-1} \right\}$$

$$= \{ \delta^2(V_N^T \Psi V_N^T - V_N^T \Psi V_N^T) \mid \Psi \in S^{N-1}_+ \}$$

When $N = 1$, the EDM cone and its dual in ambient $S_h$ each comprise the origin in isomorphic $\mathbb{R}^0$; thus, selfdual in this dimension. (confer (104))

When $N = 2$, the EDM cone is the nonnegative real line in isomorphic $\mathbb{R}$. (Figure 154) EDM$^2$ in $S^2_h$ is identical, thus selfdual in this dimension. This result is in agreement with (1376), verified directly: for all $\kappa \in \mathbb{R}$, $z = \kappa \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $d_{12}^* \geq 0$.

The first case adverse to selfdualness $N = 3$ may be deduced from Figure 150; the EDM cone is a circular cone in isomorphic $\mathbb{R}^3$ corresponding to no rotation of Lorentz cone (178) (the selfdual circular cone). Figure 163 illustrates the EDM cone and its dual in ambient $S^3_h$; no longer selfdual.

6.8.2.0.1 Exercise. Positive semidefinite cone from EDM cone.
What, if any, is the inversion of semidefinite and distance cone equality (1350)? That is to say, can $S_+^N$ be expressed only in terms of EDM, $S_h^N$, and $S_c^N$?

6.8.2.0.2 Exercise. Rank complementarity for EDM cone.
Prove (1367).
\[ D^* \in \text{EDM}^N \Leftrightarrow \delta(D^* 1) - D^* \succeq 0 \quad (1356) \]

Figure 163: Ordinary dual EDM cone projected on \( S^3_h \) shrouds \( \text{EDM}^3 \); drawn tiled in isometrically isomorphic \( \mathbb{R}^3 \). (It so happens: intersection \( \text{EDM}^N \cap S^N_h \) (§2.13.9.3) is identical to projection of dual EDM cone on \( S^N_h \).)
6.9 Theorem of the alternative

In §2.13.2.1.1 we showed how alternative systems of generalized inequality can be derived from closed convex cones and their duals. This section is, therefore, a fitting postscript to the discussion of the dual EDM cone.

6.9.0.0.3 Theorem. EDM alternative. [173, §1]

Given \( D \in \mathbb{S}_h^N \)

\[
D \in \text{EDM}^N
\]

or in the alternative

\[
\exists z \text{ such that } \begin{cases} 
1^T z = 1 \\
D z = 0 
\end{cases}
\]

(1379)

In words, either \( \mathcal{N}(D) \) intersects hyperplane \( \{ z | 1^T z = 1 \} \) or \( D \) is an EDM; the alternatives are incompatible.

When \( D \) is an EDM [269, §2]

\[
\mathcal{N}(D) \subset \mathcal{N}(1^T) = \{ z | 1^T z = 0 \}
\]

(1380)

Because [173, §2] (§E.0.1)

\[
DD^\dagger 1 = 1 \\
1^T D^\dagger 1 = 1^T
\]

(1381)

then

\[
\mathcal{R}(1) \subset \mathcal{R}(D)
\]

(1382)

6.10 Postscript

We provided an equality (1350) relating the convex cone of Euclidean distance matrices to the convex cone of positive semidefinite matrices. Projection on a positive semidefinite cone, constrained by an upper bound on rank, is easy and well known; [136] simply, a matter of truncating a list of eigenvalues. Projection on a positive semidefinite cone with such a rank constraint is, in fact, a convex optimization problem. (§7.1.4)

In the past, it was difficult to project on the EDM cone under a constraint on rank or affine dimension. A surrogate method was to invoke the Schoenberg criterion (988) and then project on a positive semidefinite cone under a rank constraint bounding affine dimension from above. But a solution acquired that way is necessarily suboptimal.

In §7.3.3 we present a method for projecting directly on the EDM cone under a constraint on rank or affine dimension.
Chapter 7

Proximity problems

In the “extremely large-scale case” (N of order of tens and hundreds of thousands), [iteration cost O(N^3)] rules out all advanced convex optimization techniques, including all known polynomial time algorithms.

—Arkadi Nemirovski, 2004

A problem common to various sciences is to find the Euclidean distance matrix (EDM) $D \in \text{EDM}^N$ closest in some sense to a given complete matrix of measurements $H$ under a constraint on affine dimension $0 \leq r \leq N-1$ (§2.3.1, §5.7.1.1); rather, $r$ is bounded above by desired affine dimension $\rho$.

### 7.0.1 Measurement matrix $H$

Ideally, we want a given matrix of measurements $H \in \mathbb{R}^{N \times N}$ to conform with the first three Euclidean metric properties (§5.2); to belong to the intersection of the orthant of nonnegative matrices $\mathbb{R}^{N \times N}_+$ with the symmetric hollow subspace $\mathbb{S}_h^N$ (§2.2.3.0.1). Geometrically, we want $H$ to belong to the polyhedral cone (§2.12.1.0.1)

$$\mathcal{K} \triangleq \mathbb{S}_h^N \cap \mathbb{R}^{N \times N}_+$$

Yet in practice, $H$ can possess significant measurement uncertainty (noise).

Sometimes realization of an optimization problem demands that its input, the given matrix $H$, possess some particular characteristics; perhaps symmetry and hollowness or nonnegativity. When that $H$ given does not have the desired properties, then we must impose them upon $H$ prior to optimization:

- When measurement matrix $H$ is not symmetric or hollow, taking its symmetric hollow part is equivalent to orthogonal projection on the symmetric hollow subspace $\mathbb{S}_h^N$.
- When measurements of distance in $H$ are negative, zeroing negative entries effects unique minimum-distance projection on the orthant of nonnegative matrices $\mathbb{R}^{N \times N}_+$ in isomorphic $\mathbb{R}^{N^2}$ (§E.9.2.2.3).
CHAPTER 7. PROXIMITY PROBLEMS

7.0.1.1 Order of imposition

Since convex cone $K$ is the intersection of an orthant with a subspace, we want to project on that subset of the orthant belonging to the subspace; on the nonnegative orthant in the symmetric hollow subspace that is, in fact, the intersection. For that reason alone, unique minimum-distance projection of $H$ on $K$ (that member of $K$ closest to $H$ in isomorphic $\mathbb{R}^{N^2}$ in the Euclidean sense) can be attained by first taking its symmetric hollow part, and only then clipping negative entries of the result to 0; *id est*, there is only one correct order of projection, in general, on an orthant intersecting a subspace:

- project on the subspace, then project the result on the orthant in that subspace. (*confer §E.9.5*)

In contrast, order of projection on an intersection of subspaces is arbitrary.

That order of projection rule applies more generally, of course, to intersection of any convex set $C$ with any subspace. Consider the *proximity problem* over convex feasible set $S^N_h \cap C$ given nonsymmetric nonhollow $H \in \mathbb{R}^{N \times N}$:

$$\begin{align*}
\text{minimize} & \quad \|B - H\|_F^2 \\
\text{subject to} & \quad B \in C
\end{align*}$$

(1384)

a convex optimization problem. Because the symmetric hollow subspace $S^N_h$ is orthogonal to the antisymmetric antihollow subspace $\mathbb{R}^{N \times N \perp}$ (*§2.2.3*), then for $B \in S^N_h$

$$\text{tr} \left( B^T \left( \frac{1}{2} (H - H^T) + \delta^2(H) \right) \right) = 0$$

(1385)

so the objective function is equivalent to

$$\|B - H\|_F^2 \equiv \left\| B - \left( \frac{1}{2} (H + H^T) - \delta^2(H) \right) \right\|_F^2 + \left\| \frac{1}{2} (H - H^T) + \delta^2(H) \right\|_F^2$$

(1386)

This means the antisymmetric antihollow part of given matrix $H$ would be ignored by minimization with respect to symmetric hollow variable $B$ under Frobenius’ norm; *id est*, minimization proceeds as though given the symmetric hollow part of $H$.

This action of Frobenius’ norm (1386) is effectively a Euclidean projection (minimum-distance projection) of $H$ on the symmetric hollow subspace $S^N_h$ prior to minimization. Thus minimization proceeds inherently following the correct order for projection on $S^N_h \cap C$. Therefore we may either assume $H \in S^N_h$, or take its symmetric hollow part prior to optimization.

---

7.1 There are two equivalent interpretations of projection (*§E.9*): one finds a set normal, the other, minimum distance between a point and a set. Here we realize the latter view.
7.0.1.2 Flagrant input error under nonnegativity demand

More pertinent to the optimization problems presented herein where

\[ C \triangleq \text{EDM}^N \subseteq \mathcal{K} = S^N_h \cap \mathbb{R}^N_{+ \times N} \]  \hspace{1cm} (1387)

then should some particular realization of a proximity problem demand input \( H \) be nonnegative, and were we only to zero negative entries of a nonsymmetric nonhollow input \( H \) prior to optimization, then the ensuing projection on \( \text{EDM}^N \) would be guaranteed incorrect (out of order).

Now comes a surprising fact: Even were we to correctly follow the order of projection rule so as to provide \( H \in \mathcal{K} \) prior to optimization, then the ensuing projection on \( \text{EDM}^N \) will be incorrect whenever input \( H \) has negative entries and some proximity problem demands nonnegative input \( H \).

This is best understood referring to Figure 164: Suppose nonnegative input \( H \) is demanded, and then the problem realization correctly projects its input first on \( S^N_h \) and then directly on \( C = \text{EDM}^N \). That demand for nonnegativity effectively requires imposition of \( \mathcal{K} \) on input \( H \) prior to optimization so as to obtain correct order of projection (on \( S^N_h \) first). Yet such an imposition prior to projection on \( \text{EDM}^N \) generally introduces an elbow into the path of projection (illustrated in Figure 165) caused by the technique itself; that being, a particular proximity problem realization requiring nonnegative input.

Any procedure for imposition of nonnegativity on input \( H \) can only be incorrect in this circumstance. There is no resolution unless input \( H \) is guaranteed nonnegative with no tinkering. Otherwise, we have no choice but to employ a different problem realization; one not demanding nonnegative input.
Figure 165: Pseudo-Venn diagram from Figure 164 showing elbow placed in path of projection of $H$ on $\text{EDM}^N \subset S_h^N$ by an optimization problem demanding nonnegative input matrix $H$. The first two line segments, leading away from $H$, result from correct order of projection required to provide nonnegative $H$ prior to optimization. Were $H$ nonnegative, then its projection on $S_h^N$ would instead belong to $\mathcal{K}$; making the elbow disappear. (confer Figure 179)
7.0.2 Lower bound

Most of the problems we encounter in this chapter have the general form:

\[
\begin{aligned}
\text{minimize} & \quad \| B - A \|_F \\
\text{subject to} & \quad B \in \mathcal{C}
\end{aligned}
\]  

(1388)

where \( A \in \mathbb{R}^{m \times n} \) is given data. This particular objective denotes Euclidean projection (§E) of vectorized matrix \( A \) on the set \( \mathcal{C} \) which may or may not be convex. When \( \mathcal{C} \) is convex, then the projection is unique minimum-distance because Frobenius’ norm when squared is a strictly convex function of variable \( B \) and because the optimal solution is the same regardless of the square (512). When \( \mathcal{C} \) is a subspace, then the direction of projection is orthogonal to \( \mathcal{C} \).

Denoting by \( A \triangleq U_A \Sigma_A Q_A^T \) and \( B \triangleq U_B \Sigma_B Q_B^T \) their full singular value decompositions (whose singular values are always nonincreasingly ordered (§A.6)), there exists a tight lower bound on the objective over the manifold of orthogonal matrices:

\[
\| \Sigma_B - \Sigma_A \|_F \leq \inf_{U_A, U_B, Q_A, Q_B} \| B - A \|_F
\]  

(1389)

This least lower bound holds more generally for any orthogonally invariant norm on \( \mathbb{R}^{m \times n} \) (§2.2.1) including the Frobenius and spectral norm [338, §II.3]. [212, §7.4.51]

7.0.3 Problem approach

Problems traditionally posed in terms of point position \( \{ x_i \in \mathbb{R}^n, \ i = 1 \ldots N \} \) such as

\[
\text{minimize} \sum_{i, j \in \mathcal{I}} (\| x_i - x_j \| - h_{ij})^2
\]  

(1390)

or

\[
\text{minimize} \sum_{i, j \in \mathcal{I}} (\| x_i - x_j \|^2 - h_{ij})^2
\]  

(1391)

(where \( \mathcal{I} \) is an abstract set of indices and \( h_{ij} \) is given data) are everywhere converted herein to the distance-square variable \( D \) or to Gram matrix \( G \); the Gram matrix acting as bridge between position and distance. (That conversion is performed regardless of whether known data is complete.) Then the techniques of chapter 5 or chapter 6 are applied to find relative or absolute position. This approach is taken because we prefer introduction of rank constraints into convex problems rather than searching a googol of local minima in nonconvex problems like (1390) [110] (§3.6.4.0.3, §7.2.2.7.1) or (1391).

7.0.4 Three prevalent proximity problems

There are three statements of the closest-EDM problem prevalent in the literature, the multiplicity due primarily to choice of projection on the EDM versus positive semidefinite (PSD) cone and vacillation between the distance-square variable \( d_{ij} \) versus absolute
distance $\sqrt{d_{ij}}$. In their most fundamental form, the three prevalent proximity problems are (1392.1), (1392.2), and (1392.3): [352] for $D \equiv [d_{ij}]$ and $\sqrt{D} \equiv [\sqrt{d_{ij}}]$

(1) minimize $\| -V(D - H)V\|_F^2$ subject to $\operatorname{rank} VDV \leq \rho$ $D \in \operatorname{EDM}^N$

(2) minimize $\| \sqrt{D} - H\|_F^2$ subject to $\operatorname{rank} VDV \leq \rho$ $\sqrt{D} \in \sqrt{\operatorname{EDM}^N}$

(3) minimize $\| D - H\|_F^2$ subject to $\operatorname{rank} VDV \leq \rho$ $D \in \operatorname{EDM}^N$

(4) minimize $\| -V(\sqrt{D} - H)V\|_F^2$ subject to $\operatorname{rank} VDV \leq \rho$ $\sqrt{D} \in \sqrt{\operatorname{EDM}^N}$

where we have made explicit an imposed upper bound $\rho$ on affine dimension

$$r = \operatorname{rank} V_N^T D V_N = \operatorname{rank} VDV$$

that is benign when $\rho = N-1$ or $H$ were realizable with $r \leq \rho$. Problems (1392.2) and (1392.3) are Euclidean projections of a vectorized matrix $H$ on an EDM cone (§6.3), whereas problems (1392.1) and (1392.4) are Euclidean projections of a vectorized matrix $-VH$ on a PSD cone.7.2 Problem (1392.4) is not posed in the literature because it has limited theoretical foundation.7.3

Analytical solution to (1392.1) is known in closed form for any bound $\rho$ although, as the problem is stated, it is a convex optimization only in the case $\rho = N-1$. We show in §7.1.4 how (1392.1) becomes a convex optimization problem for any $\rho$ when transformed to the spectral domain. When expressed as a function of point list in a matrix $X$ as in (1390), problem (1392.2) becomes a variant of what is known in statistics literature as a stress problem. [55, p.34] [108] [366] Problems (1392.2) and (1392.3) are convex optimization problems in $D$ for the case $\rho = N-1$ wherein (1392.3) becomes equivalent to (1391). Even with the rank constraint removed from (1392.2), we will see that the convex problem remaining inherently minimizes affine dimension.

Generally speaking, each problem in (1392) produces a different result because there is no isometry relating them. Of the various auxiliary $V$-matrices (§B.4), the geometric centering matrix $V$ (991) appears in the literature most often although $V_N$ (975) is the auxiliary matrix naturally consequent to Schoenberg’s seminal exposition [323]. Substitution of any auxiliary matrix or its pseudoinverse into these problems produces another valid problem.

Substitution of $V_N^T$ for left-hand $V$ in (1392.1), in particular, produces a different result because

$$\minimize_{D} \| -V(D - H)V\|_F^2$$

subject to $D \in \operatorname{EDM}^N$ (1393)

---

7.2 Because $-VH$ is orthogonal projection of $-H$ on the geometric center subspace $S_N^+$ (§E.7.2.0.2), problems (1392.1) and (1392.4) may be interpreted as oblique (nonminimum distance) projections of $-H$ on a positive semidefinite cone.

7.3 $D \in \operatorname{EDM}^N \Rightarrow \sqrt{D} \in \sqrt{\operatorname{EDM}^N}$, $-V\sqrt{D}V \in S_N^+$ (§5.10)
7.1. **FIRST PREVALENT PROBLEM:**

finds $D$ to attain Euclidean distance of vectorized $-VHV$ to the positive semidefinite cone in ambient isometrically isomorphic $\mathbb{R}^{N(N+1)/2}$, whereas

\[
\begin{align*}
\text{minimize} & \quad \| -V_N^T (D - H)V_N \|_F^2 \\
\text{subject to} & \quad D \in \text{EDM}^N
\end{align*}
\]

attains Euclidean distance of vectorized $-V_N^T HV_N$ to the positive semidefinite cone in isometrically isomorphic subspace $\mathbb{R}^{N(N-1)/2}$; quite different projections\(^7.4\) regardless of whether affine dimension is constrained. But substitution of auxiliary matrix $V_W^T$ (§B.4.3) or $V_N^T$ yields the same result as (1392.1) because $V = V_W V_W^T = V_N V_N^T$; \emph{id est},

\[
\begin{align*}
\| -V(D - H)V \|_F^2 &= \| -V_W^T V_W V_W^T (D - H)V_W \|_F^2 = \| -V_W^T (D - H)V_W \|_F^2 \\
&= \| -V_N^T (D - H)V_N \|_F^2
\end{align*}
\] (1395)

We see no compelling reason to prefer one particular auxiliary $V$-matrix over another. Each has its own coherent interpretations; \textit{e.g.} §5.4.2, §6.6, §B.4.5. Neither can we say any particular problem formulation produces generally better results than another.\(^7.5\)

7.1 **First prevalent problem:**

**Projection on PSD cone**

This first problem

\[
\begin{align*}
\text{minimize} & \quad \| -V_N^T (D - H)V_N \|_F^2 \\
\text{subject to} & \quad \text{rank} V_N^T D V_N \leq \rho \\
& \quad D \in \text{EDM}^N
\end{align*}
\]

poses a Euclidean projection of $-V_N^T HV_N$ in subspace $\mathbb{S}^{N-1}$ on a generally nonconvex subset (when $\rho < N - 1$) of the positive semidefinite cone boundary $\partial \mathbb{S}^{N-1}_+$ whose elemental matrices have rank no greater than desired affine dimension $\rho$ (§5.7.1.1). Problem 1 finds the closest EDM $D$ in the sense of Schoenberg. (1988) [323] As it is stated, this optimization problem is convex only when desired affine dimension is largest $\rho = N - 1$ although its analytical solution is known [267, thm.14.4.2] for all nonnegative $\rho \leq N - 1$.\(^7.6\)

We assume only that the given measurement matrix $H$ is symmetric;\(^7.7\)

\[
H \in \mathbb{S}^N
\] (1397)

\(^7.4\) The isomorphism $T(Y) = V_N^T Y V_N^T$ onto $\mathbb{S}^N = \{ V^T X V \mid X \in \mathbb{S}^N \}$ relates the map in (1394) to that in (1393), but is not an isometry.

\(^7.5\) All four problem formulations (1392) produce identical results when affine dimension $r$, implicit to a realizable measurement matrix $H$, does not exceed desired affine dimension $\rho$; because, the optimal objective value will vanish ($\| \cdot \| = 0$).

\(^7.6\) Being first pronounced in the context of multidimensional scaling by Mardia [266] in 1978 who attributes the generic result (§7.1.2) to Eckart & Young, 1936 [136].

\(^7.7\) Projection in Problem 1 is on a rank $\rho$ subset of the positive semidefinite cone $\mathbb{S}^{N-1}_+$ (§2.9.2.1) in the subspace of symmetric matrices $\mathbb{S}^{N-1}$. It is wrong here to zero the main diagonal of given $H$ because first projecting $H$ on the symmetric hollow subspace places an elbow in the path of projection in Problem 1. (Figure 165)
Arranging the eigenvalues $\lambda_i$ of $-V_N^T H V_N$ in nonincreasing order for all $i$, $\lambda_i \geq \lambda_{i+1}$ with $v_i$ the corresponding $i^{th}$ eigenvector, then an optimal solution to Problem 1 is

$$-V_N^T D^* V_N = \sum_{i=1}^{\rho} \max \{0, \lambda_i\} v_i v_i^T$$

(1398)

where

$$-V_N^T H V_N \triangleq \sum_{i=1}^{N-1} \lambda_i v_i v_i^T \in S_N^{-1}$$

(1399)

is an eigenvalue decomposition and

$$D^* \in \text{EDM}^N$$

(1400)

is an optimal Euclidean distance matrix.

In §7.1.4 we show how to transform Problem 1 to a convex optimization problem for any $\rho$.

7.1.1 Closest-EDM Problem 1, convex case

7.1.1.0.1 Proof. Solution (1398), convex case.

When desired affine dimension is unconstrained, $\rho = N - 1$, the rank function disappears from (1396) leaving a convex optimization problem; a simple unique minimum-distance projection on the positive semidefinite cone $S_+^{N-1}$: \textit{videlicet}

$$\min_{D \in S_N^N} \| -V_N^T (D - H) V_N \|_F^2$$

subject to $-V_N^T D V_N \succeq 0$

(1401)

by (988). Because

$$S^{N-1} = -V_N^T S_h^N V_N$$

(1092)

then the necessary and sufficient conditions for projection in isometrically isomorphic $\mathbb{R}^{N(N-1)/2}$ on the selfdual (376) positive semidefinite cone $S_+^{N-1}$ are:7.8 (§E.9.2.0.1) (1674) (confer (2129))

$$-V_N^T D^* V_N \succeq 0$$
$$-V_N^T D^* V_N (-V_N^T D^* V_N + V_N^T H V_N) = 0$$
$$-V_N^T D^* V_N + V_N^T H V_N \succeq 0$$

(1402)

Symmetric $-V_N^T H V_N$ is diagonalizable hence decomposable in terms of its eigenvectors $v$ and eigenvalues $\lambda$ as in (1399). Therefore (confer (1398))

$$-V_N^T D^* V_N = \sum_{i=1}^{N-1} \max \{0, \lambda_i\} v_i v_i^T$$

(1403)

7.8 The Karush-Kuhn-Tucker (KKT) optimality conditions [291, p.328] [63, §5.5.3] for problem (1401) are identical to these conditions for projection on a convex cone.
satisfies (1402), optimally solving (1401). To see that, recall: these eigenvectors constitute an orthogonal set and

\[-V_N^T D^* V_N + V_N^T H V_N = - \sum_{i=1}^{N-1} \min\{0, \lambda_i\} v_i v_i^T \]  

(1404)

\[7.1.2 \text{ generic problem}\]

Prior to determination of $D^*$, analytical solution (1398) to Problem 1 is equivalent to solution of a generic rank-constrained projection problem: Given desired affine dimension $\rho$ and

\[ A \triangleq -V_N^T H V_N = \sum_{i=1}^{N-1} \lambda_i v_i v_i^T \in \mathbb{S}^{N-1} \]  

(1399)

Euclidean projection on a rank $\rho$ subset of a PSD cone (on a generally nonconvex subset of the PSD cone boundary $\partial \mathbb{S}^{N-1}_+$ when $\rho < N-1$)

\[
\begin{align*}
\text{minimize} & \quad ||B - A||_F^2 \\
\text{subject to} & \quad \text{rank } B \leq \rho \\
& \quad B \succeq 0
\end{align*}
\]

(1405)

has well known optimal solution (Eckart & Young) [136]

\[ B^* \triangleq -V_N^T D^* V_N = \sum_{i=1}^{\rho} \max\{0, \lambda_i\} v_i v_i^T \in \mathbb{S}^{N-1} \]  

(1398)

Once optimal $B^*$ is found, the technique of §5.12 can be used to determine a uniquely corresponding optimal Euclidean distance matrix $D^*$; a unique correspondence by injectivity arguments in §5.6.2.

\[7.1.2.1 \text{ Projection on rank } \rho \text{ subset of PSD cone}\]

Because Problem 1 is the same as

\[
\begin{align*}
\text{minimize} & \quad ||-V_N^T (D - H)V_N||_F^2 \\
\text{subject to} & \quad \text{rank } V_N^T D V_N \leq \rho \\
& \quad -V_N^T D V_N \succeq 0
\end{align*}
\]

(1406)

and because (1092) provides invertible mapping to the generic problem, then Problem 1 is truly a Euclidean projection of vectorized $-V_N^T H V_N$ on that generally nonconvex subset of symmetric matrices (belonging to the positive semidefinite cone $\mathbb{S}^{N-1}_+$) having rank no greater than desired affine dimension $\rho$; called rank $\rho$ subset: (260)

\[ \mathbb{S}^{N-1}_+ \setminus \mathbb{S}^{N-1}_+(ho + 1) = \{ X \in \mathbb{S}^{N-1}_+ \mid \text{rank } X \leq \rho \} \]  

(216)

\[\text{7.9 a} \]

Recall: affine dimension is a lower bound on embedding (§2.3.1), equal to dimension of the smallest affine set in which points from a list $X$ corresponding to an EDM $D$ can be embedded.
7.1.3 Choice of spectral cone

Spectral projection substitutes projection on a polyhedral cone, containing a complete set of eigenspectra, in place of projection on a convex set of diagonalizable matrices; e.g. (1418). In this section we develop a method of spectral projection for constraining rank of positive semidefinite matrices in a proximity problem like (1405). We will see why an orthant turns out to be the best choice of spectral cone, and why presorting is critical.

Define a nonlinear permutation-operator \( \pi(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n \) that sorts its vector argument \( x \) into nonincreasing order.

**7.1.3.0.1 Definition. Spectral projection.**

Let \( R \) be an orthogonal matrix and \( \Lambda \) a nonincreasingly ordered diagonal matrix of eigenvalues. Spectral projection means unique minimum-distance projection of a rotated \((R, \pi)\) vector \( \delta \) of eigenvalues \( \pi(R^T \Lambda R) \)

\[
\pi(\delta(R^T \Lambda R))
\]

(1407)
on a polyhedral cone containing all eigenspectra corresponding to a rank \( \rho \) subset of a positive semidefinite cone (§2.9.2.1) or the EDM cone (in Cayley-Menger form, §5.11.2.3).

\[ \square \]

In the simplest and most common case, projection on a positive semidefinite cone, orthogonal matrix \( R \) equals \( I \) (§7.1.4.0.1) and diagonal matrix \( \Lambda \) is ordered during diagonalization (§A.5.1). Then spectral projection simply means projection of \( \delta(\Lambda) \) on a subset of the nonnegative orthant, as we shall now ascertain:

It is curious how nonconvex Problem 1 has such a simple analytical solution (1398). Although solution to generic problem (1405) is well known since 1936 \[136\], its equivalence was observed in 1997 \[365, \S 2\] to projection of an ordered vector of eigenvalues (in diagonal matrix \( \Lambda \)) on a subset of the monotone nonnegative cone (§2.13.9.4.2)

\[ K_M^+ = \{ v \mid v_1 \geq v_2 \geq \cdots \geq v_{N-1} \geq 0 \} \subseteq \mathbb{R}^{N-1}_+ \quad (428) \]

Of interest, momentarily, is only the smallest convex subset of the monotone nonnegative cone \( K_M^+ \) containing every nonincreasingly ordered eigenspectrum corresponding to a rank \( \rho \) subset of the positive semidefinite cone \( \mathbb{S}_+^{N-1} \); id est,

\[
K_{M+}^\rho = \{ v \in \mathbb{R}^\rho \mid v_1 \geq v_2 \geq \cdots \geq v_\rho \geq 0 \} \subseteq \mathbb{R}^\rho_+ \quad (1408)
\]
a pointed polyhedral cone, a \( \rho \)-dimensional convex subset of the monotone nonnegative cone \( K_{M+} \) having property, for \( \lambda \) denoting eigenspectra,

\[
\begin{bmatrix}
K_{M+}^\rho \\
0
\end{bmatrix} = \pi(\lambda(\text{rank } \rho \text{ subset})) \subseteq K_{M+}^{N-1} \triangleq K_M^+ \quad (1409)
\]

For each and every elemental eigenspectrum

\[
\gamma \in \lambda(\text{rank } \rho \text{ subset}) \subseteq \mathbb{R}^{N-1}_+
\]

(1410) of the rank \( \rho \) subset (ordered or unordered in \( \lambda \)), there is a nonlinear surjection \( \pi(\gamma) \) onto \( K_{M+}^\rho \).
7.1. FIRST PREVALENT PROBLEM:

7.1.3.0.2 Exercise. Smallest spectral cone.  
Prove that there is no convex subset of $K_{M+}$ smaller than $K_{M+}^\rho$ containing every ordered eigenspectrum corresponding to the rank $\rho$ subset of a positive semidefinite cone (§2.9.2.1).

7.1.3.0.3 Proposition. (Hardy-Littlewood-Pólya) Inequalities. [189, §X] [57, §1.2]  
Any vectors $\sigma$ and $\gamma$ in $\mathbb{R}^{N-1}$ satisfy a tight inequality
\[
\pi(\sigma)^T \pi(\gamma) \geq \sigma^T \gamma \geq \pi(\sigma)^T \Xi \pi(\gamma) \tag{1411}
\]
where $\Xi$ is the order-reversing permutation matrix defined in (1820), and permutator $\pi(\gamma)$ is a nonlinear function that sorts vector $\gamma$ into nonincreasing order thereby providing the greatest upper bound and least lower bound with respect to every possible sorting. ♦

7.1.3.0.4 Corollary. Monotone nonnegative sort.  
Any given vectors $\sigma, \gamma \in \mathbb{R}^{N-1}$ satisfy a tight Euclidean distance inequality
\[
\|\pi(\sigma) - \pi(\gamma)\| \leq \|\sigma - \gamma\| \tag{1412}
\]
where nonlinear function $\pi(\gamma)$ sorts vector $\gamma$ into nonincreasing order thereby providing the least lower bound with respect to every possible sorting. ♦

Given $\gamma \in \mathbb{R}^{N-1}$
\[
\inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma - \gamma\| = \inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\pi(\sigma) - \pi(\gamma)\| = \inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma - \pi(\gamma)\| = \inf_{\sigma \in K_{M+}} \|\sigma - \pi(\gamma)\| \tag{1413}
\]
Yet for $\gamma$ representing an arbitrary vector of eigenvalues, because
\[
\inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma - \gamma\|^2 \geq \inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma - \pi(\gamma)\|^2 = \inf_{\sigma \in K_{M+}} \|\sigma - \pi(\gamma)\|^2 \tag{1414}
\]
then projection of $\gamma$ on the eigenspectra of a rank $\rho$ subset can be tightened simply by presorting $\gamma$ into nonincreasing order.

Proof. Simply because $\pi(\gamma)_{1:\rho} \geq \pi(\gamma_{1:\rho})$
\[
\inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma - \gamma\|^2 = \gamma_{\rho+1:N-1}^T \gamma_{\rho+1:N-1} + \inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma_{1:\rho} - \gamma_{1:\rho}\|^2
\]
\[
= \gamma_{\rho+1:N-1}^T \gamma_{\rho+1:N-1} + \inf_{\sigma \in \mathbb{R}^{N-1}_+} \sigma_{1:\rho}^T \sigma_{1:\rho} - 2\gamma_{1:\rho}^T \gamma_{1:\rho}
\]
\[
\geq \gamma_{\rho+1:N-1}^T \gamma_{\rho+1:N-1} + \inf_{\sigma \in \mathbb{R}^{N-1}_+} \sigma_{1:\rho}^T \sigma_{1:\rho} - 2\gamma_{1:\rho}^T \pi(\gamma)_{1:\rho} \tag{1415}
\]
\[
\inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma - \gamma\|^2 \geq \inf_{\sigma \in \mathbb{R}^{N-1}_+} \|\sigma - \pi(\gamma)\|^2
\]
♦
7.1.3.1 Orthant is best spectral cone for Problem 1

This means unique minimum-distance projection of $\gamma$ on the nearest spectral member of the rank $\rho$ subset is tantamount to presorting $\gamma$ into nonincreasing order. Only then does unique spectral projection on a subset $K_{M+}^\rho$ of the monotone nonnegative cone become equivalent to unique spectral projection on a subset $R_{+}^\rho$ of the nonnegative orthant (which is simpler); in other words, unique minimum-distance projection of sorted $\gamma$ on the nonnegative orthant in a $\rho$-dimensional subspace of $\mathbb{R}^N$ is indistinguishable from its projection on the subset $K_{M+}^\rho$ of the monotone nonnegative cone in that same subspace.

7.1.4 Closest-EDM Problem 1, “nonconvex” case

Proof of solution (1398), for projection on a rank $\rho$ subset of the positive semidefinite cone $S_{+}^{N-1}$, can be algebraic in nature. [365, §2] Here we derive that known result but instead using a more geometric argument via spectral projection on a polyhedral cone (subsuming the proof in §7.1.1). In so doing, we demonstrate how nonconvex Problem 1 is transformed to a convex optimization:

7.1.4.0.1 Proof. Solution (1398), nonconvex case.

As explained in §7.1.2, we may instead work with the more facile generic problem (1405). With diagonalization of unknown

$$B \triangleq U \Upsilon U^T \in S_{+}^{N-1}$$

(1416)

given desired affine dimension $0 \leq \rho \leq N - 1$ and diagonalizable

$$A \triangleq Q \Lambda Q^T \in S_{+}^{N-1}$$

(1417)

having eigenvalues in $\Lambda$ arranged in nonincreasing order, by (48) the generic problem is equivalent to

$$\begin{align*}
\min_{B \in S_{+}^{N-1}} & \quad \|B - A\|_F^2 \\
\text{subject to} & \quad \text{rank } B \leq \rho \\
& \quad B \succeq 0
\end{align*} \equiv \begin{align*}
\min_{\Upsilon, R} & \quad \|\Upsilon - R^T \Lambda R\|_F^2 \\
\text{subject to} & \quad \text{rank } \Upsilon \leq \rho \\
& \quad \Upsilon \succeq 0 \\
& \quad R^T = R^{-1}
\end{align*}$$

(1418)

where

$$R \triangleq Q^T U \in \mathbb{R}^{N-1 \times N-1}$$

(1419)

in $U$ on the set of orthogonal matrices is a bijection. We propose solving (1418) by instead solving the problem sequence:

$$\begin{align*}
\min_{\Upsilon} & \quad \|\Upsilon - R^T \Lambda R\|_F^2 \\
\text{subject to} & \quad \text{rank } \Upsilon \leq \rho \\
& \quad \Upsilon \succeq 0
\end{align*} \equiv \begin{align*}
\min_{R} & \quad \|\Upsilon^* - R^T \Lambda R\|_F^2 \\
\text{subject to} & \quad R^{-1} = R^T
\end{align*}$$

(1420)

Problem (1420a) is equivalent to:
7.1. FIRST PREVALENT PROBLEM:

(1) orthogonal projection of $R^TAR$ on an $N - 1$-dimensional subspace of isometrically isomorphic $\mathbb{R}^{N(N-1)/2}$ containing $\delta(\Upsilon) \in \mathbb{R}^{N-1}$

(2) nonincreasingly ordering the result,

(3) unique minimum-distance projection of the ordered result on $\begin{bmatrix} \mathbb{R}_+^\rho \\ 0 \end{bmatrix}$ (§E.9.5).

Projection on that $N - 1$-dimensional subspace amounts to zeroing $R^TAR$ at all entries off the main diagonal; thus, the equivalent sequence leading with a spectral projection:

$$\begin{align*}
\text{minimize} & \quad \| \delta(\Upsilon) - \pi(\delta(R^TAR)) \|^2 \\
\text{subject to} & \quad \delta(\Upsilon) \in \begin{bmatrix} \mathbb{R}_+^\rho \\ 0 \end{bmatrix} \\
\end{align*}$$

(a)\hspace{2cm}(1421)

$$\begin{align*}
\text{minimize} & \quad \| \Upsilon^* - R^TAR \|^2_F \\
\text{subject to} & \quad R^{-1} = R^T \\
\end{align*}$$

(b)

Because any permutation matrix is an orthogonal matrix, $\delta(R^TAR) \in \mathbb{R}^{N-1}$ can always be arranged in nonincreasing order without loss of generality; hence, permutation operator $\pi$. Unique minimum-distance projection of vector $\pi(\delta(R^TAR))$ on the $\rho$-dimensional subset $\begin{bmatrix} \mathbb{R}_+^\rho \\ 0 \end{bmatrix}$ of nonnegative orthant $\mathbb{R}_+^{N-1}$ requires: (§E.9.2.0.1)

$$\begin{align*}
\delta(\Upsilon^*)_{\rho+i:N-1} = 0 \\
\delta(\Upsilon^*) \succeq 0 \\
\delta(\Upsilon^*)^T(\delta(\Upsilon^*) - \pi(\delta(R^TAR))) = 0 \\
\delta(\Upsilon^*) - \pi(\delta(R^TAR)) \succeq 0
\end{align*}$$

(1422)

which are necessary and sufficient conditions. Any value $\Upsilon^*$ satisfying conditions (1422) is optimal for (1421a). So

$$\delta(\Upsilon^*)_i = \begin{cases} 
\max \left\{ 0, \pi(\delta(R^TAR))_i \right\}, & i = 1 \ldots \rho \\
0, & i = \rho + 1 \ldots N - 1 
\end{cases}$$

specifies an optimal solution. The lower bound on the objective with respect to $R$ in (1421b) is tight: by (1389)

$$\| |\Upsilon^*| - |A| \|_F \leq \| \Upsilon^* - R^TAR \|_F$$

(1424)

where $|$ denotes absolute entry-value. For selection of $\Upsilon^*$ as in (1423), this lower bound is attained when (confer §C.4.2.2)

$$R^* = I$$

(1425)

which is the known solution.

♦
7.1.4.1 Significance

Importance of this well-known [136] optimal solution (1398) for projection on a rank $\rho$ subset of a positive semidefinite cone should not be dismissed:

- Problem 1, as stated, is generally nonconvex. This analytical solution at once encompasses projection on a rank $\rho$ subset (216) of the positive semidefinite cone (generally, a nonconvex subset of its boundary) from either the exterior or interior of that cone.\textsuperscript{7.10} By problem transformation to the spectral domain, projection on a rank $\rho$ subset becomes a convex optimization problem.

- This solution is closed form.

- This solution is equivalent to projection on a polyhedral cone in the spectral domain (spectral projection, projection on a spectral cone §7.1.3.0.1); a necessary and sufficient condition (§A.3.1) for membership of a symmetric matrix to a rank $\rho$ subset of a positive semidefinite cone (§2.9.2.1).

- Because $U^* = Q$, a minimum-distance projection on a rank $\rho$ subset of the positive semidefinite cone is a positive semidefinite matrix orthogonal (in the Euclidean sense) to direction of projection.\textsuperscript{7.11}

- For the convex case problem (1401), this solution is always unique. Otherwise, distinct eigenvalues (multiplicity 1) in $\Lambda$ guarantees uniqueness of this solution by the reasoning in §A.5.0.1.\textsuperscript{7.12}

7.1.5 Problem 1 in spectral norm, convex case

When instead we pose the matrix 2-norm (spectral norm) in Problem 1 (1396) for the convex case $\rho = N - 1$, then the new problem

$$\text{minimize}_{D} \| -V_N^T(D - H)V_N\|_2$$

subject to $D \in \mathbb{EDM}^N$ (1426)

is convex although its solution is not necessarily unique;\textsuperscript{7.13} giving rise to nonorthogonal projection (§E.1) on the positive semidefinite cone $\mathbb{S}_+^{N-1}$. Indeed, its solution set includes the Frobenius solution (1398) for the convex case whenever $-V_N^T H V_N$ is a normal matrix. [194, §1] [187] [63, §8.1.1] Proximity problem (1426) is equivalent to

$$\text{minimize}_{\mu, D} \mu$$

subject to $-\mu I \leq -V_N^T(D - H)V_N \leq \mu I$

$D \in \mathbb{EDM}^N$ (1427)

\textsuperscript{7.10}Projection on the boundary from the interior of a convex Euclidean body is generally a nonconvex problem.

\textsuperscript{7.11}But Theorem E.9.2.0.1 for unique projection on a closed convex cone does not apply here because the direction of projection is not necessarily a member of the dual PSD cone. This occurs, for example, whenever positive eigenvalues are truncated.

\textsuperscript{7.12}Uncertainty of uniqueness prevents the erroneous conclusion that a rank $\rho$ subset (216) were a convex body by the Bunt-Motzkin theorem (§E.9.0.0.1).

\textsuperscript{7.13}For each and every $|t| \leq 2$, for example, \[ \begin{bmatrix} 2 & 0 \\ 0 & t \end{bmatrix} \] has the same spectral-norm value.
7.2. SECOND PREVALENT PROBLEM:  

by (1796) where

$$\mu^* = \max_i \left\{ \left| \lambda \left( -V_N^T (D^* - H) V_N \right) \right| \right\}, \ i = 1 \ldots N - 1 \in \mathbb{R}_+$$  \hspace{1cm} (1428)

the minimized largest absolute eigenvalue (due to matrix symmetry).

For lack of unique solution here, we prefer the Frobenius rather than spectral norm.

7.2 Second prevalent problem:

Projection on EDM cone in $\sqrt{d_{ij}}$

Let

$$\sqrt{D} \triangleq [\sqrt{d_{ij}}] \in \mathcal{K} = S_h^N \cap \mathbb{R}_+^{N \times N}$$  \hspace{1cm} (1429)

be an unknown matrix of absolute distance; \textit{id est},

$$D = [d_{ij}] \triangleq \sqrt{D} \circ \sqrt{D} \in \text{EDM}^N$$  \hspace{1cm} (1430)

where $\circ$ denotes Hadamard product. The second prevalent proximity problem is a Euclidean projection (in the natural coordinates $\sqrt{d_{ij}}$) of matrix $H$ on a nonconvex subset of the boundary of the nonconvex cone of Euclidean absolute-distance matrices rel $\partial \sqrt{\text{EDM}^N}$: (§6.3, confer Figure 150b)

$$\begin{align*}
\text{minimize} & \quad \| \sqrt{D} - H \|_F^2 \\
\text{subject to} & \quad \text{rank } V_N^T D V_N \leq \rho \\
& \quad \sqrt{D} \in \sqrt{\text{EDM}^N}
\end{align*}$$  \hspace{1cm} (1431)

where

$$\sqrt{\text{EDM}^N} = \{ \sqrt{D} \mid D \in \text{EDM}^N \}$$  \hspace{1cm} (1264)

This statement of the second proximity problem is considered difficult to solve because of the constraint on desired affine dimension $\rho$ (§5.7.2) and because the objective function

$$\| \sqrt{D} - H \|_F^2 = \sum_{i,j} (\sqrt{d_{ij}} - h_{ij})^2$$  \hspace{1cm} (1432)

is expressed in the natural coordinates; projection on a doubly nonconvex set.

Our solution to this second problem prevalent in the literature requires measurement matrix $H$ to be nonnegative;

$$H = [h_{ij}] \in \mathbb{R}_+^{N \times N}$$  \hspace{1cm} (1433)

If the $H$ matrix given has negative entries, then the technique of solution presented here becomes invalid. As explained in §7.0.1, projection of $H$ on $\mathcal{K} = S_h^N \cap \mathbb{R}_+^{N \times N}$ (1383) prior to application of this proposed solution is incorrect.
CHAPTER 7. PROXIMITY PROBLEMS

7.2.1 Convex case

When \( \rho = N - 1 \), the rank constraint vanishes and a convex problem that is equivalent to (1390) emerges:\(^7\)

\[
\begin{align*}
\text{minimize} & \quad \| \sqrt{D} - H \|_F^2 \\
\text{subject to} & \quad \sqrt{D} \in \sqrt{\text{EDM}_N} \quad \Leftrightarrow \quad \text{minimize} & \quad \sum_{i,j} d_{ij} - 2h_{ij} \sqrt{d_{ij}} + h_{ij}^2 \\
& \quad D \in \text{EDM}_N
\end{align*}
\]

(1434)

For any fixed \( i \) and \( j \), the argument of summation is a convex function of \( d_{ij} \) because (for nonnegative constant \( h_{ij} \)) the negative square root is convex in nonnegative \( d_{ij} \) and because \( d_{ij} + h_{ij}^2 \) is affine (convex). Because the sum of any number of convex functions in \( D \) remains convex [63, §3.2.1] and because the feasible set is convex in \( D \), we have a convex optimization problem:

\[
\begin{align*}
\text{minimize} & \quad 1^T(D - 2H \circ \sqrt{D})1 + \| H \|_F^2 \\
\text{subject to} & \quad D \in \text{EDM}_N
\end{align*}
\]

(1435)

The objective function being a sum of strictly convex functions is, moreover, strictly convex in \( D \) on the nonnegative orthant. Existence of a unique solution \( D^* \) for this second prevalent problem depends upon nonnegativity of \( H \) and a convex feasible set (§3.1.2).\(^7\)

7.2.1.1 Equivalent semidefinite program, Problem 2, convex case

Convex problem (1434) is numerically solvable for its global minimum using an interior-point method [423] [300] [288] [414] [12] [152]. We translate (1434) to an equivalent semidefinite program (SDP) for a pedagogical reason made clear in §7.2.2.2 and because there exist readily available computer programs for numerical solution [177] [416] [417] [371] [35] [415] [360] [346].

Substituting a new matrix variable \( Y \equiv [y_{ij}] \in \mathbb{R}_+^{N \times N} \)

\[
h_{ij} \sqrt{d_{ij}} \leftarrow y_{ij}
\]

(1436)

Boyd proposes: problem (1434) is equivalent to the semidefinite program

\[
\begin{align*}
\text{minimize} & \quad \sum_{i,j} d_{ij} - 2y_{ij} + h_{ij}^2 \\
\text{subject to} & \quad \begin{bmatrix} d_{ij} & y_{ij} \\ y_{ij} & h_{ij}^2 \end{bmatrix} \succeq 0, \quad i,j = 1 \ldots N \\
& \quad D \in \text{EDM}_N
\end{align*}
\]

(1437)

\(^7\)\(^7\)\(^7\) still thought to be a nonconvex problem as late as 1997 [366] even though discovered convex by de Leew in 1993. [108] [55, §13.6] Yet using methods from §3, it can be easily ascertained: \( \| \sqrt{D} - H \|_F \) is not convex in \( D \).

\(^7\)\(^7\) The transformed problem in variable \( D \) no longer describes Euclidean projection on an EDM cone. Otherwise we might erroneously conclude \( \sqrt{\text{EDM}_N} \) were a convex body by the Bunt-Motzkin theorem (§E.9.0.0.1).
7.2. SECOND PREVALENT PROBLEM:

To see that, recall: \( d_{ij} \geq 0 \) is implicit to \( D \in \text{EDM}^N \) (§5.8.1, (988)). So when \( H \in \mathbb{R}^{N \times N}_+ \) is nonnegative, as assumed,

\[
\begin{bmatrix}
  d_{ij} & y_{ij} & h_{ij}^2 \\
  y_{ij} & & \\
  h_{ij}^2 & & 
\end{bmatrix} \succeq 0 \iff h_{ij} \sqrt{d_{ij}} \geq \sqrt{y_{ij}^2}
\]

(1438)

Minimization of the objective function implies maximization of \( y_{ij} \) that is bounded above. Hence nonnegativity of \( y_{ij} \) is implicit to (1437) and, as desired, \( y_{ij} \to h_{ij} \sqrt{d_{ij}} \) as optimization proceeds.

If the given matrix \( H \) is now assumed symmetric and nonnegative, \( H = [h_{ij}] \in \mathbb{S}^N \cap \mathbb{R}^{N \times N}_+ \) (1439), then \( Y = H \circ \sqrt{D} \) must belong to \( K = \mathbb{S}^N_h \cap \mathbb{R}^{N \times N}_+ \) (1383). Because \( Y \in \mathbb{S}^N_h \) (§B.4.2 no.20), then

\[
\| \sqrt{D} - H \|_F^2 = \sum_{i,j} d_{ij} - 2y_{ij} + h_{ij}^2 = -N \text{tr}(V(D - 2Y)V) + \|H\|_F^2
\]

(1440)

So convex problem (1437) is equivalent to the semidefinite program

\[
\begin{align*}
\text{minimize} & \quad -\text{tr}(V(D - 2Y)V) \\
\text{subject to} & \quad \begin{bmatrix} d_{ij} & y_{ij} & h_{ij}^2 \\ y_{ij} & & \\ h_{ij}^2 & & \end{bmatrix} \succeq 0, \quad N \geq j > i = 1 \ldots N - 1 \\
Y & \in \mathbb{S}^N_h \\
D & \in \text{EDM}^N
\end{align*}
\]

(1441)

where the constants \( h_{ij}^2 \) and \( N \) have been dropped arbitrarily from the objective.

7.2.1.2 Gram-form semidefinite program, Problem 2, convex case

There is great advantage to expressing problem statement (1441) in Gram-form because Gram matrix \( G \) is a bidirectional bridge between point list \( X \) and distance matrix \( D \); e.g., §5.4.2.2.8, §6.7.0.0.1. This way, problem convexity can be maintained while simultaneously constraining point list \( X \), Gram matrix \( G \), and distance matrix \( D \) at our discretion.

Convex problem (1441) may be equivalently written via linear bijective (§5.6.1) EDM operator \( D(G) \) (981):

\[
\begin{align*}
\text{minimize} & \quad -\text{tr}(V(D(G) - 2Y)V) \\
\text{subject to} & \quad \begin{bmatrix} \langle \Phi_{ij} \ , \ G \rangle & y_{ij} & h_{ij}^2 \\ y_{ij} & & \\ h_{ij}^2 & & \end{bmatrix} \succeq 0, \quad N \geq j > i = 1 \ldots N - 1 \\
G & \succeq 0
\end{align*}
\]

(1442)
CHAPTER 7. PROXIMITY PROBLEMS

where distance-square \( D = [d_{ij}] \in S_N^h \) (965) is related to \( G = [g_{ij}] \in S_N^c \cap S_N^+ \) Gram matrix entries by

\[
d_{ij} = g_{ii} + g_{jj} - 2g_{ij} = \langle \Phi_{ij}, G \rangle \quad (980)
\]

where

\[
\Phi_{ij} = (e_i - e_j)(e_i - e_j)^T \in S_N^+ \quad (967)
\]

Confinement of \( G \) to the geometric center subspace provides numerical stability and no loss of generality (confer (1328)); implicit constraint \( G1 = 0 \) is otherwise unnecessary.

To include constraints on the list \( X \in \mathbb{R}^{n \times N} \), we would first rewrite (1442)

\[
\begin{align*}
\text{minimize} & \quad -\text{tr}(V(D(G) - 2Y)V) \\
\text{subject to} & \quad \begin{bmatrix}
\langle \Phi_{ij}, G \rangle & y_{ij} \\
y_{ij} & h_{ij}^2
\end{bmatrix} \succeq 0, \quad N \geq j > i = 1 \ldots N - 1 \\
& \quad \begin{bmatrix}
I & X \\
X^T & G
\end{bmatrix} \succeq 0 \\
& \quad X \in \mathbb{C}
\end{align*}
\]

and then add the constraints, realized here in abstract membership to some convex set \( \mathbb{C} \). This problem realization includes a convex relaxation of the nonconvex constraint \( G = X^T X \) and, if desired, more constraints on \( G \) could be added. This technique is discussed in §5.4.2.2.8.

7.2.2 Minimization of affine dimension in Problem 2

When desired affine dimension \( \rho \) is diminished, the rank function becomes reinserted into problem (1437) that is then rendered difficult to solve because feasible set \( \{D, Y\} \) loses convexity in \( S_N^h \times \mathbb{R}^{N \times N} \). Indeed, the rank function is quasiconcave (§3.8) on the positive semidefinite cone; (§2.9.2.9.2) id est, its sublevel sets are not convex.

7.2.2.1 Rank minimization heuristic

A remedy developed in [272] [142] [143] [141] introduces convex envelope of the quasiconcave rank function: (Figure 166)

7.2.2.1.1 Definition. Convex envelope.

Convex envelope \( \text{cenv} f \) of a function \( f: \mathbb{C} \to \mathbb{R} \) is defined to be the largest convex function \( g \) such that \( g \leq f \) on convex domain \( \mathbb{C} \subseteq \mathbb{R}^n \). 7.16

\[\Delta\]

\[\text{Provided } f \neq +\infty \text{ and there exists an affine function } h \leq f \text{ on } \mathbb{R}^n, \text{ then the convex envelope is equal to the convex conjugate (the Legendre-Fenchel transform) of the convex conjugate of } f; \text{ id est, the conjugate-conjugate function } f^{**}. \quad \text{[209, §I.E.1]}\]
Figure 166: Abstraction of convex envelope of rank function. Rank is a quasiconcave function on positive semidefinite cone, but its convex envelope is the largest convex function whose epigraph contains it. Vertical bar labelled $g$ measures a trace/rank gap; \textit{id est}, rank found always exceeds estimate; large decline in trace required here for only small decrease in rank.

- \[142\] \[141\] Convex envelope of rank function: for $\sigma_i$ a singular value, (1648)

\[
\text{cenv}(\text{rank } A) \text{ on } \{ A \in \mathbb{R}^{m \times n} \mid \| A \|_2 \leq \kappa \} = \frac{1}{\kappa} 1^T \sigma(A) = \frac{1}{\kappa} \text{tr}(A^T A)
\]  

(1444)

\[
\text{cenv}(\text{rank } A) \text{ on } \{ A \text{ normal } \mid \| A \|_2 \leq \kappa \} = \frac{1}{\kappa} \| \lambda(A) \|_1 = \frac{1}{\kappa} \text{tr}(A^T A)
\]  

(1445)

\[
\text{cenv}(\text{rank } A) \text{ on } \{ A \in \mathbb{S}^n_+ \mid \| A \|_2 \leq \kappa \} = \frac{1}{\kappa} 1^T \lambda(A) = \frac{1}{\kappa} \text{tr}(A)
\]  

(1446)

A properly scaled trace thus represents the best convex lower bound on rank for positive semidefinite matrices. The idea, then, is to substitute convex envelope for rank of some variable $A \in \mathbb{S}^n_+$ (§A.6.5)

\[
\text{rank } A \leftarrow \text{cenv}(\text{rank } A) \propto \text{tr } A = \sum_i \sigma(A)_i = \sum_i \lambda(A)_i
\]  

(1447)

which is equivalent to the sum of all eigenvalues or singular values.

- \[141\] Convex envelope of the cardinality function is proportional to the 1-norm:

\[
\text{cenv}(\text{card } x) \text{ on } \{ x \in \mathbb{R}^n \mid \| x \|_\infty \leq \kappa \} = \frac{1}{\kappa} \| x \|_1
\]  

(1448)

\[
\text{cenv}(\text{card } x) \text{ on } \{ x \in \mathbb{R}^n_+ \mid \| x \|_\infty \leq \kappa \} = \frac{1}{\kappa} 1^T x
\]  

(1449)
7.2.2.2 Applying trace rank-heuristic to Problem 2

Substituting rank envelope for rank function in Problem 2, for \( D \in \text{EDM}^N \) (confer (1123))

\[
\text{cenv rank}(V_A^T D V_N) = \text{cenv rank}(V D V) \propto -\text{tr}(V D V) \quad (1450)
\]

and for desired affine dimension \( \rho \leq N - 1 \) and nonnegative \( H \) [sic] we get a convex optimization problem

\[
\begin{align*}
\text{minimize} & \quad \|\sqrt{D} - H\|_F^2 \\
\text{subject to} & \quad -\text{tr}(V D V) \leq \kappa \rho \\
D & \in \text{EDM}^N
\end{align*}
\quad (1451)
\]

where \( \kappa \in \mathbb{R}_+ \) is a constant determined by cut-and-try. The equivalent semidefinite program makes \( \kappa \) variable: for nonnegative and symmetric \( H \)

\[
\begin{align*}
\text{minimize} & \quad \kappa \rho + 2 \text{tr}(V Y V) \\
\text{subject to} & \quad \begin{bmatrix}
  d_{ij} & y_{ij} \\
  y_{ij} & h_{ij}^2
\end{bmatrix} \succeq 0, \quad N \geq j > i = 1 \ldots N - 1 \\
& \quad -\text{tr}(V D V) \leq \kappa \rho \\
& \quad Y \in S_h^N \\
& \quad D \in \text{EDM}^N
\end{align*}
\quad (1452)
\]

which is the same as (1441), the problem with no explicit constraint on affine dimension. As the present problem is stated, the desired affine dimension \( \rho \) yields to the variable scale factor \( \kappa \); \( \rho \) is effectively ignored.

Yet this result is an illuminant for problem (1441) and it equivalents (all the way back to (1434)): When the given measurement matrix \( H \) is nonnegative and symmetric, finding the closest EDM \( D \) as in problem (1434), (1437), or (1441) implicitly entails minimization of affine dimension (confer §5.8.4, §5.14.4). Those non-rank-constrained problems are each inherently equivalent to cenv(rank)-minimization problem (1452), in other words, and their optimal solutions are unique because of the strictly convex objective function in (1434).

7.2.2.3 Rank-heuristic insight

Minimization of affine dimension by use of this trace rank-heuristic (1450) tends to find a list configuration of least energy; rather, it tends to optimize compaction of the reconstruction by minimizing total distance. (993) It is best used where some physical equilibrium implies such an energy minimization; e.g., [364, §5].

For this Problem 2, the trace rank-heuristic arose naturally in the objective in terms of \( V \). We observe: \( V \) (in contrast to \( V_A^T \)) spreads energy over all available distances (§B.4.2 no.20, contrast no.22) although the rank function itself is insensitive to choice of auxiliary matrix.

Trace rank-heuristic (1446) is useless when a main diagonal is constrained to be constant. Such would be the case were optimization over an ellipope (§5.4.2.2.1), or when the diagonal represents a Boolean vector; e.g., §4.2.3.1.1, §4.6.0.0.9.
7.2. SECOND PREVALENT PROBLEM:

7.2.2.4 Rank minimization heuristic beyond convex envelope

Fazel, Hindi, & Boyd [143] [419] [144] propose a rank heuristic more potent than trace (1447) for problems of rank minimization;

\[
\text{rank } Y \leftarrow \log \text{det}(Y + \varepsilon I) \tag{1453}
\]

the concave surrogate function \( \log \text{det} \) in place of quasiconcave rank \( Y \) (§2.9.2.9.2) when \( Y \in \mathbb{S}^n_+ \) is variable and where \( \varepsilon \) is a small positive constant. They propose minimization of the surrogate by substituting a sequence comprising infima of a linearized surrogate about the current estimate \( Y_i \); \text{id est}, from the first-order Taylor series expansion about \( Y_i \) on some open interval of \( \|Y\|_2 \) (§D.1.7)

\[
\log \text{det}(Y + \varepsilon I) \approx \log \text{det}(Y_i + \varepsilon I) + \text{tr}((Y_i + \varepsilon I)^{-1}(Y - Y_i)) \tag{1454}
\]

we make the surrogate sequence of infima over bounded convex feasible set \( C \)

\[
\text{arg inf}_{Y \in C} \text{rank } Y \leftarrow \lim_{i \to \infty} Y_{i+1} \tag{1455}
\]

where, for \( i = 0 \ldots \)

\[
Y_{i+1} = \text{arg inf}_{Y \in C} \text{tr}((Y_i + \varepsilon I)^{-1}Y) \tag{1456}
\]

a matrix analogue to the reweighting scheme disclosed in [217, §4.11.3]. Choosing \( Y_0 = I \), the first step becomes equivalent to finding the infimum of \( \text{tr } Y \); the trace rank-heuristic (1447). The intuition underlying (1456) is the new term in the argument of trace; specifically, \( (Y_i + \varepsilon I)^{-1} \) weights \( Y \) so that relatively small eigenvalues of \( Y \) found by the infimum are made even smaller.

To see that, substitute the nonincreasingly ordered diagonalizations

\[
Y_i + \varepsilon I \triangleq Q(\Lambda + \varepsilon I)Q^T \quad (a) \\
Y \triangleq U\Upsilon U^T \quad (b) \tag{1457}
\]

into (1456). Then from (1793) we have,

\[
\inf_{\Upsilon \in U^\star C U^\star} \delta((\Lambda + \varepsilon I)^{-1})^T \delta(Y) = \inf_{\Upsilon \in U^\star C U^\star} \inf_{R^T R = R^{-1}} \text{tr}((\Lambda + \varepsilon I)^{-1} R^T \Upsilon R) \\
\leq \inf_{R \in C} \text{tr}((Y_i + \varepsilon I)^{-1} Y) \tag{1458}
\]

where \( R \triangleq Q^T U \) in \( U \) on the set of orthogonal matrices is a bijection. The role of \( \varepsilon \) is, therefore, to limit maximum weight; the smallest entry on the main diagonal of \( \Upsilon \) gets the largest weight.
7.2.2.5 Applying log det rank-heuristic to Problem 2

When the log det rank-heuristic is inserted into Problem 2, problem (1452) becomes the problem sequence in

\[
\begin{align*}
\text{minimize} & \quad \kappa \rho + 2 \text{tr}(VYV) \\
\text{subject to} & \quad \begin{bmatrix} d_{ji} & y_{ji} & h_{ji}^2 \end{bmatrix} \succeq 0, \quad l > j = 1 \ldots N - 1 \\
& \quad - \text{tr}((-VD_iV + \varepsilon I)^{-1}VDV) \leq \kappa \rho \\
& \quad Y \in S^N_+ \\
& \quad D \in \text{EDM}_N
\end{align*}
\]

where \( D_{i+1} \triangleq D^* \in \text{EDM}_N \) and \( D_0 \triangleq 11^T - I \).

7.2.2.6 Tightening this log det rank-heuristic

Like the trace method, this log det technique for constraining rank offers no provision for meeting a predetermined upper bound \( \rho \). Yet since eigenvalues are simply determined, \( \lambda(Y_i + \varepsilon I) = \delta(\Lambda + \varepsilon I) \), we may certainly force selected weights to \( \varepsilon^{-1} \) by manipulating diagonalization (1457a). Empirically we find this sometimes leads to better results, although affine dimension of a solution cannot be guaranteed.

7.2.2.7 Cumulative summary of rank heuristics

We have studied a perturbation method of rank reduction in §4.3 as well as the trace heuristic (convex envelope method §7.2.2.1.1) and log det heuristic in §7.2.2.4. There is another good contemporary method called LMIRank [296] based on alternating projection (§E.10).\textsuperscript{7,17}

7.2.2.7.1 Example. Unidimensional scaling.

We apply the convex iteration method from §4.4.1 to numerically solve an instance of Problem 2; a method empirically superior to the foregoing convex envelope and log det heuristics for rank regularization and enforcing affine dimension.

**Unidimensional scaling.** [110] a historically practical application of multidimensional scaling (§5.12), entails solution of an optimization problem having local minima whose multiplicity varies as the factorial of point-list cardinality; geometrically, it means reconstructing a list constrained to lie in one affine dimension. In terms of point list, the nonconvex problem is: given nonnegative symmetric matrix \( H = [h_{ij}] \in G^N \cap \mathbb{R}_+^{N \times N} \) (1439) whose entries \( h_{ij} \) are all known,

\[
\text{minimize}_{\{x_i \in \mathbb{R}\}} \sum_{i, j = 1}^{N} (|x_i - x_j| - h_{ij})^2
\]

(1390)
called a raw stress problem [55, p.34] which has an implicit constraint on dimensional embedding of points \( \{x_i \in \mathbb{R}, \ i = 1 \ldots N\} \). This problem has proven NP-hard; e.g., [75].\textsuperscript{7,17} that does not solve the ball packing problem presented in §5.4.2.2.6.
7.2. SECOND PREVALENT PROBLEM:

As always, we first transform variables to distance-square $D \in S^N_h$; so begin with convex problem (1441) on page 495

$$\begin{align*}
\text{minimize} & \quad -\text{tr}(V(D - 2Y)V) \\
\text{subject to} & \quad \begin{bmatrix} d_{ij} & y_{ij} & h_{ij}^2 \end{bmatrix} \succeq 0, \quad N \geq j > i = 1 \ldots N - 1 \\
& \quad Y \in S^N_h \\
& \quad D \in \text{EDM}^N \\
& \quad \text{rank} V^T_NDV_N = 1
\end{align*} \tag{1460}$$

that becomes equivalent to (1390) by making explicit the constraint on affine dimension via rank. The iteration is formed by moving the dimensional constraint to the objective:

$$\begin{align*}
\text{minimize} & \quad -\langle V(D - 2Y)V, I \rangle - w\langle V^T_NDV_N, W \rangle \\
\text{subject to} & \quad \begin{bmatrix} d_{ij} & y_{ij} & h_{ij}^2 \end{bmatrix} \succeq 0, \quad N \geq j > i = 1 \ldots N - 1 \\
& \quad Y \in S^N_h \\
& \quad D \in \text{EDM}^N
\end{align*} \tag{1461}$$

where $w \approx 10$ is a positive scalar just large enough to make $\langle V^T_NDV_N, W \rangle$ vanish to within some numerical precision, and where direction matrix $W$ is an optimal solution to semidefinite program (1792a)

$$\begin{align*}
\text{minimize} & \quad -\langle V^T_NDV_N, W \rangle \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \text{tr} W = N - 1
\end{align*} \tag{1462}$$

one of which is known in closed form. Semidefinite programs (1461) and (1462) are iterated until convergence in the sense defined on page 264. This iteration is not a projection method. (§4.4.1.1) Convex problem (1461) is neither a relaxation of unidimensional scaling problem (1460); instead, problem (1461) is a convex equivalent to (1460) at convergence of the iteration.

Jan de Leeuw provided us with some test data

$$H = \begin{bmatrix}
0.000000 & 5.235301 & 5.499274 & 6.404294 & 6.486829 & 6.263265 \\
5.235301 & 0.000000 & 3.208028 & 5.840931 & 3.559010 & 5.353489 \\
5.499274 & 3.208028 & 0.000000 & 5.679550 & 4.020339 & 5.239842 \\
6.404294 & 5.840931 & 5.679550 & 0.000000 & 4.862884 & 4.543120 \\
6.486829 & 3.559010 & 4.020339 & 4.862884 & 0.000000 & 4.618718 \\
6.263265 & 5.353489 & 5.239842 & 4.543120 & 4.618718 & 0.000000
\end{bmatrix} \tag{1463}$$

and a globally optimal solution

$$X^* = \begin{bmatrix}
-4.981494 & -2.121026 & -1.038738 & 4.555130 & 0.764096 & 2.822032 \\
x_1^* & x_2^* & x_3^* & x_4^* & x_5^* & x_6^*
\end{bmatrix} \tag{1464}$$
found by searching 6! local minima of (1390) [110]. By iterating convex problems (1461) and (1462) about twenty times (initial $W = 0$) we find a global infimum 98.12812 of stress problem (1390), and by (1213) we find a corresponding one-dimensional point list that is a rigid transformation in $\mathbb{R}$ of $X^*$.

Here we found the infimum to accuracy of the given data, but that ceases to hold as problem size increases. Because of machine numerical precision and an interior-point method of solution, we speculate, accuracy degrades quickly as problem size increases beyond this.

7.3 Third prevalent problem: Projection on EDM cone in $d_{ij}$

In summary, we find that the solution to problem [(1392.3) p.484] is difficult and depends on the dimension of the space as the geometry of the cone of EDMs becomes more complex.

-- Hayden, Wells, Liu, & Tarazaga, 1991 [195, §3]

Reformulating Problem 2 (p.493), in terms of EDM $D$, changes it considerably:

$$\begin{align*}
\text{minimize} & \quad \|D - H\|_F^2 \\
\text{subject to} & \quad \text{rank} V_N^T D V_N \leq \rho \\
& \quad D \in \text{EDM}_N
\end{align*} \tag{1465}$$

This third prevalent proximity problem is a Euclidean projection of given matrix $H$ on a generally nonconvex subset ($\rho < N - 1$) of $\partial \text{EDM}_N$ the boundary of the convex cone of Euclidean distance matrices relative to subspace $S_h^N$ (Figure 150d). Because coordinates of projection are distance-square and $H$ now presumably holds distance-square measurements, numerical solution to Problem 3 is generally different than that of Problem 2.

For the moment, we need make no assumptions regarding measurement matrix $H$.

7.3.1 Convex case

$$\begin{align*}
\text{minimize} & \quad \|D - H\|_F^2 \\
\text{subject to} & \quad D \in \text{EDM}_N
\end{align*} \tag{1466}$$

When the rank constraint disappears (for $\rho = N - 1$), this third problem becomes obviously convex because the feasible set is then the entire EDM cone and because the objective function

$$\|D - H\|_F^2 = \sum_{i,j} (d_{ij} - h_{ij})^2 \tag{1467}$$
7.3. **THIRD PREVALENT PROBLEM:**

is a strictly convex quadratic in $D$:

$$
\begin{align*}
\text{minimize} \quad & \sum_{i,j} d_{ij}^2 - 2h_{ij} d_{ij} + h_{ij}^2 \\
\text{subject to} \quad & D \in \text{EDM}^N
\end{align*}
$$

(1468)

Optimal solution $D^*$ is therefore unique, as expected, for this simple projection on the EDM cone equivalent to (1391).

### 7.3.1.1 Equivalent semidefinite program, Problem 3, convex case

In the past, this convex problem was solved numerically by means of alternating projection. (Example 7.3.1.1.1) [163] [155] [195, §1] We translate (1468) to an equivalent semidefinite program because we have a good solver:

Assume the given measurement matrix $H$ to be nonnegative and symmetric:

$$
H = [h_{ij}] \in S^N \cap \mathbb{R}_{+}^{N \times N} \quad (1439)
$$

We then propose: Problem (1468) is equivalent to the semidefinite program, for

$$
\partial \triangleq [d_{ij}^2] = D \circ D \quad (1469)
$$

a matrix of distance-square squared,

$$
\begin{align*}
\text{minimize} \quad & -\text{tr}(V(\partial - 2H \circ D)V) \\
\text{subject to} \quad & \begin{bmatrix} \partial_{ij} & d_{ij} \\ d_{ij} & 1 \end{bmatrix} \succeq 0, \quad N \geq j > i = 1 \ldots N - 1 \\
& D \in \text{EDM}^N \\
& \partial \in S_N^h
\end{align*}
$$

(1470)

where

$$
\begin{bmatrix} \partial_{ij} & d_{ij} \\ d_{ij} & 1 \end{bmatrix} \succeq 0 \iff \partial_{ij} \geq d_{ij}^2
$$

(1471)

Symmetry of input $H$ facilitates trace in the objective (§B.4.2 no. 20), while its nonnegativity causes $\partial_{ij} \to d_{ij}^2$ as optimization proceeds.

### 7.3.1.1.1 Example. Alternating projection on nearest EDM.

By solving (1470) we confirm the result from an example given by Glunt, Hayden, Hong, & Wells [163, §6] who found analytical solution to convex optimization problem (1466) for nonzero $Y \in S_N^h$ and some open interval of $t \in \mathbb{R}$ (§3.7.3.0.2, §D.2.3)

$$
\frac{d^2}{dt^2} \| (D + tY) - H \|^2_F = 2 \text{tr} Y^T Y > 0
$$

If that $H$ given has negative entries, then the technique of solution presented here becomes invalid. Projection of $H$ on $\mathcal{K}$ (1383) prior to application of this proposed technique, as explained in §7.0.1, is incorrect.
particular cardinality $N = 3$ by using the alternating projection method of von Neumann (§E.10):

$$H = \begin{bmatrix}
0 & 1 & 1 \\
1 & 0 & 9 \\
1 & 9 & 0
\end{bmatrix}, \quad D^* = \begin{bmatrix}
\frac{19}{9} & \frac{10}{9} & \frac{10}{9} \\
\frac{10}{9} & 0 & \frac{76}{9} \\
\frac{10}{9} & \frac{76}{9} & 0
\end{bmatrix}$$

(1472)

The problem (1466), of projecting $H$ on the EDM cone, is transformed to an equivalent iterative sequence of projections on the two convex cones (1332) from §6.8.1.1. Utilizing projector (1335) in an ordinary alternating projection, input $H$ goes to $D^*$ with an accuracy of four decimal places in about 17 iterations. Affine dimension corresponding to this optimal solution is $r = 1$.

Obviating of semidefinite programming's computational expense is the principal advantage of this alternating projection technique.

7.3.1.2 Schur-form semidefinite program, Problem 3 convex case

Semidefinite program (1470) can be reformulated by moving the objective function in

$$\begin{align*}
\text{minimize} & \quad \|D - H\|_F^2 \\
\text{subject to} & \quad D \in \text{EDM}^N
\end{align*}$$

(1466)

to the constraints. This makes an equivalent epigraph form of the problem: for any measurement matrix $H$

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \|D - H\|_F^2 \leq t \\
& \quad D \in \text{EDM}^N
\end{align*}$$

(1473)

We can transform this problem to an equivalent Schur-form semidefinite program; (§3.5.2)

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \begin{bmatrix}
    tI & \text{vec}(D - H) \\
    \text{vec}(D - H)^T & 1
\end{bmatrix} \succeq 0 \\
& \quad D \in \text{EDM}^N
\end{align*}$$

(1474)

characterized by great sparsity and structure. The advantage of this SDP is lack of conditions on input $H$; e.g., negative entries would invalidate any solution provided by (1470). (§7.0.1.2)

7.3.1.3 Gram-form semidefinite program, Problem 3 convex case

Further, this problem statement may be equivalently written in terms of a Gram matrix via linear bijective (§5.6.1) EDM operator $D(G)$ (981);

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \begin{bmatrix}
    tI & \text{vec}(D(G) - H) \\
    \text{vec}(D(G) - H)^T & 1
\end{bmatrix} \succeq 0 \\
& \quad G \succeq 0
\end{align*}$$

(1475)
7.3. THIRD PREVALENT PROBLEM:

To include constraints on the list \( X \in \mathbb{R}^{n \times N} \), we would rewrite this:

\[
\begin{align*}
\text{minimize} & \quad G \in \mathbb{S}_N, t \in \mathbb{R}, X \in \mathbb{R}^{n \times N} \\
\text{subject to} & \quad \begin{bmatrix}
  tI & \text{vec}(D(G) - H) \\
  \text{vec}(D(G) - H)^T & 1 \\
  I & X \\
  X^T & G
\end{bmatrix} \succeq 0 \\
& \quad X \in C
\end{align*}
\]

(1476)

where \( C \) is some abstract convex set. This technique is discussed in §5.4.2.2.8.

### 7.3.1.4 Dual interpretation, projection on EDM cone

From §E.9.1.1 we learn that projection on a convex set has a dual form. In the circumstance \( K \) is a convex cone and point \( x \) exists exterior to the cone or on its boundary, distance to the nearest point \( Px \) in \( K \) is found as the optimal value of the objective

\[
\|x - Px\| = \max_a a^T x \\
\text{subject to} \quad \|a\| \leq 1
\]

(2113)

where \( K^\circ \) is the polar cone.

Applying this result to (1466), we get a convex optimization for any given symmetric matrix \( H \) exterior to or on the EDM cone boundary:

\[
\begin{align*}
\text{minimize} & \quad \|D - H\|_F \\
\text{subject to} & \quad D \in \text{EDM}^N \quad \max_{A^\circ} \langle A^\circ, H \rangle \\
& \quad \|A^\circ\|_F \leq 1 \\
& \quad A^\circ \in \text{EDM}^{N^\circ}
\end{align*}
\]

(1477)

Then, from (2115), projection of \( H \) on cone \( \text{EDM}^N \) is

\[
D^* = H - A^{\circ*} \langle A^{\circ*}, H \rangle
\]

(1478)

Critchley proposed, instead, projection on the polar EDM cone in his 1980 thesis [92, p.113]: In that circumstance, by projection on the algebraic complement (§E.9.2.2.1),

\[
D^* = A^* \langle A^*, H \rangle
\]

(1479)

which is equal to (1478) when \( A^* \) solves

\[
\begin{align*}
\text{maximize} & \quad \langle A^*, H \rangle \\
\text{subject to} & \quad \|A\|_F = 1 \\
& \quad A \in \text{EDM}^N
\end{align*}
\]

(1480)

This projection of symmetric \( H \) on polar cone \( \text{EDM}^{N^\circ} \) can be made a convex problem, of course, by relaxing the equality constraint (\( \|A\|_F \leq 1 \)).
7.3.2 Minimization of affine dimension in Problem 3

When desired affine dimension \( \rho \) is diminished, Problem 3 (1465) is difficult to solve [195, §3] because the feasible set in \( \mathbb{R}^{N(N-1)/2} \) loses convexity. By substituting rank envelope (1450) into Problem 3, then for any given \( H \) we get a convex problem

\[
\begin{align*}
\minimize_D & \quad \|D - H\|_F^2 \\
\text{subject to} & \quad -\text{tr}(VDV) \leq \kappa \rho \\
D & \in \text{EDM}^N
\end{align*}
\]

(1481)

where \( \kappa \in \mathbb{R}_+ \) is a constant determined by cut-and-try. Given \( \kappa \), problem (1481) is a convex optimization having unique solution in any desired affine dimension \( \rho \); an approximation to Euclidean projection on that nonconvex subset of the EDM cone containing EDMs with corresponding affine dimension no greater than \( \rho \).

The SDP equivalent to (1481) does not move \( \kappa \) into the variables as on page 498: for nonnegative symmetric input \( H \) and distance-square squared variable \( \partial \) as in (1469),

\[
\begin{align*}
\minimize_{\partial, D} & \quad -\text{tr}(V(\partial - 2H \circ D)V) \\
\text{subject to} & \quad \frac{\partial_{ij}}{d_{ij}} \geq 0, \quad N \geq j > i = 1 \ldots N-1 \\
& \quad -\text{tr}(VDV) \leq \kappa \rho \\
D & \in \text{EDM}^N \\
\partial & \in \mathbb{S}^N
\end{align*}
\]

(1482)

That means we will not see equivalence of this \text{cenv(rank)}-minimization problem to the non-rank-constrained problems (1468) and (1470) like we saw for its counterpart (1452) in Problem 2.

Another approach to affine dimension minimization is to project instead on the polar EDM cone; discussed in §6.8.1.5.

7.3.3 Constrained affine dimension, Problem 3

When one desires affine dimension diminished further below what can be achieved via \text{cenv(rank)}-minimization as in (1482), spectral projection can be considered a natural means in light of its successful application to projection on a rank \( \rho \) subset of the positive semidefinite cone in §7.1.4.

Yet it is wrong here to zero eigenvalues of \(-VDV\) or \(-GV\) or a variant to reduce affine dimension, because that particular method comes from projection on a positive semidefinite cone (1418); zeroing those eigenvalues here in Problem 3 would place an elbow in the projection path (Figure 165) thereby producing a result that is necessarily suboptimal. Problem 3 is instead a projection on the EDM cone whose associated spectral cone is considerably different. (§5.11.2.3) Proper choice of spectral cone is demanded by diagonalization of that variable argument to the objective:
7.3. THIRD PREVALENT PROBLEM:

7.3.3.1 Cayley-Menger form

We use Cayley-Menger composition of the Euclidean distance matrix to solve a problem that is the same as Problem 3 (1465): (§5.7.3.0.1)

\[
\begin{align*}
\text{minimize} & \quad \| \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbf{D} \end{bmatrix} - \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbf{H} \end{bmatrix} \|_F^2 \\
\text{subject to} & \quad \text{rank} \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbf{D} \end{bmatrix} \leq \rho + 2
\end{align*}
\]

(1483)

\(D \in \text{EDM}^N\)

a projection of \(H\) on a generally nonconvex subset (when \(\rho < N - 1\)) of the Euclidean distance matrix cone boundary rel \(\partial \text{EDM}^N\); id est, projection from the EDM cone interior or exterior on a subset of its relative boundary (§6.5, (1261)).

Rank of an optimal solution is intrinsically bounded above and below;

\[2 \leq \text{rank} \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbf{D}^* \end{bmatrix} \leq \rho + 2 \leq N + 1\]

(1484)

Our proposed strategy for low-rank solution is projection on that subset of a spectral cone \(\lambda \left( \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\text{EDM}^N \end{bmatrix} \right)\) (§5.11.2.3) corresponding to affine dimension not in excess of that \(\rho\) desired; id est, spectral projection on

\[
\begin{bmatrix} \mathbb{R}_{\rho + 1}^+ \\ 0 \\ \mathbb{R}^- \end{bmatrix} \cap \partial \mathcal{H} \subset \mathbb{R}^{N+1}
\]

(1485)

where

\[
\partial \mathcal{H} = \{ \lambda \in \mathbb{R}^{N+1} \mid \mathbf{1}^T \lambda = 0 \}
\]

(1193)

is a hyperplane through the origin. This pointed polyhedral cone (1485), to which membership subsumes the rank constraint, is not full-dimensional.

Given desired affine dimension \(0 \leq \rho \leq N - 1\) and diagonalization (§A.5) of unknown EDM \(\mathbf{D}\)

\[
\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbf{D} \end{bmatrix} \triangleq \mathbf{U} \Sigma \mathbf{U}^T \in \mathbb{S}^{N+1}
\]

(1486)

and given symmetric \(\mathbf{H}\) in diagonalization

\[
\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbf{H} \end{bmatrix} \triangleq \mathbf{Q} \Lambda \mathbf{Q}^T \in \mathbb{S}^{N+1}
\]

(1487)

having eigenvalues arranged in nonincreasing order, then by (1206) problem (1483) is equivalent to

\[
\begin{align*}
\text{minimize} & \quad \| \delta(\mathbf{Y}) - \pi(\delta(R^T \mathbf{A} \mathbf{R})) \|_2^2 \\
\text{subject to} & \quad \delta(\mathbf{Y}) \in \begin{bmatrix} \mathbb{R}^{\rho + 1}_+ \\ \mathbf{0} \\ \mathbb{R}_- \end{bmatrix} \cap \partial \mathcal{H} \\
& \quad \delta(QR \mathbf{Y} R^T Q^T) = \mathbf{0} \\
& \quad R^{-1} = R^T
\end{align*}
\]

(1488)
where \( \pi \) is the permutation operator from \( \S 7.1.3 \) arranging its vector argument in nonincreasing order.\(^{7.20}\) where

\[
R \triangleq Q^T U \in \mathbb{R}^{N+1 \times N+1}
\]  

(1489)

in \( U \) on the set of orthogonal matrices is a bijection, and where \( \partial \mathcal{H} \) insures one negative eigenvalue. Hollowness constraint \( \delta(QRYR^TQ^T) = 0 \) makes problem (1488) difficult by making the two variables dependent.

Our plan is to instead divide problem (1488) into two and then iterate their solution:

\[
\begin{align*}
\text{minimize} \quad & \|\delta(Y) - \pi(\delta(R^T \Lambda R))\|^2 \\
\text{subject to} \quad & \delta(Y) \in \left[ \begin{array}{c} \mathbb{R}_+^{\rho+1} \\ 0 \\ \mathbb{R}_- \end{array} \right] \cap \partial \mathcal{H} \quad (a) \\
\end{align*}
\]

(1490)

\[
\begin{align*}
\text{minimize} \quad & \|RY^* R^T - \Lambda\|^2_F \\
\text{subject to} \quad & \delta(QRY^* R^TQ^T) = 0 \\
& R^{-1} = R^T \\
\end{align*}
\]

Proof. We justify disappearance of the hollowness constraint in convex optimization problem (1490a): From the arguments in \( \S 7.1.3 \) with regard to \( \pi \) the permutation operator, cone membership constraint \( \delta(Y) \in \left[ \begin{array}{c} \mathbb{R}_+^{\rho+1} \\ 0 \\ \mathbb{R}_- \end{array} \right] \cap \partial \mathcal{H} \) from (1490a) is equivalent to

\[
\delta(Y) \in \left[ \begin{array}{c} \mathbb{R}_+^{\rho+1} \\ 0 \\ \mathbb{R}_- \end{array} \right] \cap \partial \mathcal{H} \cap K_M 
\]

(1491)

where \( K_M \) is the monotone cone (\( \S 2.13.9.4.3 \)). Membership of \( \delta(Y) \) to the polyhedral cone of majorization (Theorem A.1.2.0.1)

\[
K_{\lambda \delta}^* = \partial \mathcal{H} \cap K_{M+}^* 
\]

(1508)

where \( K_{M+}^* \) is the dual monotone nonnegative cone (\( \S 2.13.9.4.2 \)), is a condition (in absence of a hollowness constraint) that would insure existence of a symmetric hollow matrix

\[
\left[ \begin{array}{cc} 0 & 1^T \\ 1 & -D \end{array} \right].
\]

Curiously, intersection of this feasible superset \( \left[ \begin{array}{c} \mathbb{R}_+^{\rho+1} \\ 0 \\ \mathbb{R}_- \end{array} \right] \cap \partial \mathcal{H} \cap K_M \) from (1491) with the cone of majorization \( K_{\lambda \delta}^* \) is a benign operation; \( \text{id est} \),

\[
\partial \mathcal{H} \cap K_{M+}^* \cap K_M = \partial \mathcal{H} \cap K_M 
\]

(1492)

verifiable by observing conic dependencies (\( \S 2.10.3 \)) among the aggregate of halfspace-description normals. The cone membership constraint in (1490a) therefore inherently insures existence of a symmetric hollow matrix.

\(^{7.20}\)Recall, any permutation matrix is an orthogonal matrix.
7.3. THIRD PREVALENT PROBLEM:

Optimization (1490b) would be a Procrustes problem (§C.4) were it not for the hollowness constraint. It is, instead, a minimization over the intersection of the nonconvex manifold of orthogonal matrices with another nonconvex set in variable \( R \) specified by the hollowness constraint. We solve problem (1490b) by a method introduced in §4.6.0.0.2: Define \( R = [r_1 \cdots r_{N+1}] \in \mathbb{R}^{N+1 \times N+1} \) and make the assignment

\[
G = \begin{bmatrix}
  r_1 & \cdots & r_{N+1} \\
  \vdots & & \vdots \\
  r_{N+1} & \cdots & 1 \\
\end{bmatrix} \in \mathbb{S}^{(N+1)^2+1}
\]

subject to \( r_i r_j = \delta_{jj} \), \( i < j \).

\[
= \begin{bmatrix}
  R_{11} & \cdots & r_1 \\
  \vdots & & \vdots \\
  R_{N+1,1} & \cdots & r_{N+1} \\
\end{bmatrix} \triangleq \begin{bmatrix}
  r_1 r_1^T & \cdots & r_1 r_{N+1}^T \\
  \vdots & \ddots & \vdots \\
  r_{N+1} r_1^T & \cdots & r_{N+1} r_{N+1}^T \\
\end{bmatrix} (\geq 0)
\]

where \( R_{ij} \triangleq r_i r_j^T \in \mathbb{R}^{N+1 \times N+1} \) and \( \Upsilon ii \in \mathbb{R} \). Since \( R \Upsilon^* R^T = \sum_{i=1}^{N+1} \Upsilon ii R_{ii} \) then problem (1490b) is equivalently expressed:

\[
\begin{align*}
\text{minimize} & \quad \left\| \sum_{i=1}^{N+1} \Upsilon ii R_{ii} - \Lambda \right\|_F^2 \\
\text{subject to} & \quad \text{tr} R_{ii} = 1, \quad i = 1 \ldots N+1 \\
& \quad \text{tr} R_{ij} = 0, \quad i < j = 2 \ldots N+1 \\
& \quad G = \begin{bmatrix}
  R_{11} & \cdots & r_1 \\
  \vdots & \ddots & \vdots \\
  R_{N+1,1} & \cdots & r_{N+1} \\
\end{bmatrix} (\geq 0) \\
& \quad \delta \left( Q \sum_{i=1}^{N+1} \Upsilon ii R_{ii} Q^T \right) = 0 \\
& \quad \text{rank} G = 1
\end{align*}
\]

The rank constraint is regularized by method of convex iteration developed in §4.4. Problem (1494) is partitioned into two convex problems:

\[
\begin{align*}
\text{minimize} & \quad \left\| \sum_{i=1}^{N+1} \Upsilon ii R_{ii} - \Lambda \right\|_F^2 + \langle G, W \rangle \\
\text{subject to} & \quad \text{tr} R_{ii} = 1, \quad i = 1 \ldots N+1 \\
& \quad \text{tr} R_{ij} = 0, \quad i < j = 2 \ldots N+1 \\
& \quad G = \begin{bmatrix}
  R_{11} & \cdots & r_1 \\
  \vdots & \ddots & \vdots \\
  R_{N+1,1} & \cdots & r_{N+1} \\
\end{bmatrix} (\geq 0) \\
& \quad \delta \left( Q \sum_{i=1}^{N+1} \Upsilon ii R_{ii} Q^T \right) = 0 \\
\end{align*}
\]

and

\[
\begin{align*}
\text{minimize} & \quad \langle G^*, W \rangle \\
\text{subject to} & \quad 0 \preceq W \preceq I \\
& \quad \text{tr} W = (N + 1)^2
\end{align*}
\]
then iterated with convex problem (1490a) until a rank-1 \( G \) matrix is found and the objective of (1490a) is minimized. An optimal solution to (1496) is known in closed form (p.565). The hollowness constraint in (1495) may cause numerical infeasibility; in that case, it can be moved to the objective within an added weighted norm. Conditions for termination of the iteration would then comprise a vanishing norm of hollowness.

7.4 Conclusion

The importance and application of solving rank- or cardinality-constrained problems are enormous, a conclusion generally accepted \textit{gratis} by the mathematics and engineering communities. Rank-constrained semidefinite programs arise in many vital feedback and control problems [182], optics [77] (Figure 124), and communications [149] [264] (Figure 106). For example, one might be interested in the minimal order dynamic output feedback which stabilizes a given linear time invariant plant (this problem is considered to be among the most important open problems in control). [273] Rank and cardinality constraints also arise naturally in combinatorial optimization (§4.6.0.0.11, Figure 117), and find application to face recognition (Figure 5), cartography (Figure 146), video surveillance, latent semantic indexing [241], sparse or low-rank matrix completion for preference models and collaborative filtering, multidimensional scaling or principal component analysis (§5.12), medical imaging (Figure 112), digital filter design with time domain constraints [390], molecular conformation (Figure 139), sensor-network localization and wireless location (Figure 89), etcetera.

There has been little progress in spectral projection since the discovery by Eckart & Young in 1936 [136] leading to a formula for projection on a rank \( \rho \) subset of a positive semidefinite cone (§2.9.2.1). [153] The only closed-form spectral method presently available for solving proximity problems, having a constraint on rank, is based on their discovery (Problem 1, §7.1, §5.13).

- One popular recourse is intentional misapplication of Eckart & Young’s result by introducing spectral projection on a positive semidefinite cone into Problem 3 via \( D(G) \) (981), for example. [75] Since Problem 3 instead demands spectral projection on the EDM cone, any solution acquired that way is necessarily suboptimal.

- A second recourse is problem redesign: A presupposition to all proximity problems in this chapter is that matrix \( H \) is given. We considered \( H \) having various properties such as nonnegativity, symmetry, hollowness, or lack thereof. It was assumed that if \( H \) did not already belong to the EDM cone, then we wanted an EDM closest to \( H \) in some sense; \textit{id est}, input-matrix \( H \) was assumed corrupted somehow. For practical problems, it withstands reason that such a proximity problem could instead be reformulated so that some or all entries of \( H \) were unknown but bounded above and below by known limits; the norm objective is thereby eliminated as in the development beginning on page 276. That particular redesign (\textit{the art}, p.8), in terms of the Gram-matrix bridge between point-list \( X \) and EDM \( D \), at once encompasses proximity and completion problems.

- A third recourse is to apply the method of convex iteration as we did in §7.2.2.7.1. This technique is applicable to any semidefinite problem requiring a rank constraint; it places a regularization term in the objective that enforces the rank constraint.
Appendix A

Linear algebra

A.1 Main-diagonal $\delta$ operator, $\lambda$, $\text{tr}$, $\text{vec}$

We introduce notation $\delta$ denoting the main-diagonal linear selfadjoint operator. When linear function $\delta$ operates on a square matrix $A \in \mathbb{R}^{N \times N}$, $\delta(A)$ returns a vector composed of all the entries from the main diagonal in the natural order;

$$\delta(A) \in \mathbb{R}^N$$ (1497)

Operating on a vector $y \in \mathbb{R}^N$, $\delta$ naturally returns a diagonal matrix;

$$\delta(y) \in \mathbb{S}^N$$ (1498)

Operating recursively on a vector $\Lambda \in \mathbb{R}^N$ or diagonal matrix $\Lambda \in \mathbb{S}^N$, $\delta(\delta(\Lambda))$ returns $\Lambda$ itself;

$$\delta^2(\Lambda) \equiv \delta(\delta(\Lambda)) \triangleq \Lambda$$ (1499)

Defined in this manner, [237, §3.10, §9.5-1]$^\text{A.1}$ main-diagonal linear operator $\delta$ is selfadjoint; videlicet, (§2.2)

$$\delta(A)^T y = \langle \delta(A) , y \rangle = \langle A , \delta(y) \rangle = \text{tr}(A^T \delta(y))$$ (1500)

A.1.1 Identities

This $\delta$ notation is efficient and unambiguous as illustrated in the following examples where: $A \circ B$ denotes Hadamard product [212] [168, §1.1.4] of matrices of like size, $\otimes$ Kronecker product [176] (§D.1.2.1), $y$ a vector, $X$ a matrix, $e_i$ the $i$th member of the standard basis for $\mathbb{R}^n$, $\mathbb{S}_h^N$ the symmetric hollow subspace, $\sigma(A)$ a vector of (nonincreasingly) ordered singular values of matrix $A$, and $\lambda(A)$ a vector of nonincreasingly ordered eigenvalues:

$^\text{A.1}$Linear operator $T : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{M \times N}$ is selfadjoint when, $\forall X_1, X_2 \in \mathbb{R}^{m \times n}$

$$\langle T(X_1) , X_2 \rangle = \langle X_1 , T(X_2) \rangle$$
1. \(\delta(A) = \delta(A^T)\)

2. \(\text{tr}(A) = \text{tr}(A^T) = \delta(A)^T 1 = (I, A)\)

3. \(\delta(cA) = c \delta(A)\) \(c \in \mathbb{R}\)

4. \(\text{tr}(cA) = c \text{tr}(A) = c 1^T \lambda(A)\) \(c \in \mathbb{R}\)

5. \(\text{vec}(cA) = c \text{vec}(A)\) \(c \in \mathbb{R}\)

6. \(A \circ cB = cA \circ B\) \(c \in \mathbb{R}\)

7. \(A \otimes cB = cA \otimes B\) \(c \in \mathbb{R}\)

8. \(\delta(A + B) = \delta(A) + \delta(B)\)

9. \(\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)\)

10. \(\text{vec}(A + B) = \text{vec}(A) + \text{vec}(B)\)

11. \((A + B) \circ C = A \circ C + B \circ C\)
    \[A \circ (B + C) = A \circ B + A \circ C\]

12. \((A + B) \otimes C = A \otimes C + B \otimes C\)
    \[A \otimes (B + C) = A \otimes B + A \otimes C\]

13. \(\text{sgn}(c) \lambda(|c|A) = c \lambda(A)\) \(c \in \mathbb{R}\)

14. \(\text{sgn}(c) \sigma(|c|A) = c \sigma(A)\) \(c \in \mathbb{R}\)

15. \(\text{tr}(c \sqrt{A^T A}) = c \text{tr}\sqrt{A^T A} = c 1^T \sigma(A)\) \(c \in \mathbb{R}\)

16. \(\pi(\delta(A)) = \lambda(I \circ A)\) where \(\pi\) is the presorting function.

17. \(\delta(AB) = (A \circ B^T) 1 = (B^T \circ A) 1\)

18. \(\delta(AB)^T = 1^T (A^T \circ B) = 1^T (B \circ A^T)\)

19. \(\delta(uv^T) = \begin{bmatrix} u_1 v_1 \\ \vdots \\ u_N v_N \end{bmatrix} = u \circ v, \ u, v \in \mathbb{R}^N\)

20. \(\text{tr}(A^T B) = \text{tr}(AB^T) = \text{tr}(BA^T) = \text{tr}(B^TA)\)
    \[= 1^T (A \circ B) 1 = 1^T \delta(AB^T) = \delta(A^T B^T) 1 = \delta(BA^T) 1 = \delta(B^T A^T) 1\]

21. \(D = [d_{ij}] \in \mathbb{S}^N_h, \ H = [h_{ij}] \in \mathbb{S}_h^N, \ V = I - \frac{1}{N} 11^T \in \mathbb{S}^N\) \(\text{(confer §B.4.2 no. 20)}\)
    \[N \text{tr}(-V (D \circ H)V) = \text{tr}(D^T H) = 1^T (D \circ H) 1 = \text{tr}(11^T (D \circ H)) = \sum_{i,j} d_{ij} h_{ij}\]

22. \(\text{tr}(\Lambda A) = \delta(\Lambda)^T \delta(A), \ \delta(\Lambda) \triangleq \Lambda \in \mathbb{S}^N\)

23. \(y^T B \delta(A) = \text{tr}(B \delta(A)y^T) = \text{tr}(\delta(B^T y) A) = \text{tr}(A \delta(B^T y))\)
    \[= \delta(A)^T B^T y = \text{tr}(y \delta(A)^T B^T) = \text{tr}(A^T \delta(B^T y)) = \text{tr}(\delta(B^T y) A^T)\)
A.1. MAIN-DIAGONAL δ OPERATOR, λ, TR, VEC

24. \( \delta^2(A^T A) = \sum_i e_i e_i^T A^T A e_i e_i^T \)
25. \( \delta(\lambda_1^T) = \delta(1 \lambda_1^T) = \delta(\lambda_1) \)
26. \( \delta( \lambda_1 1^T ) = A \lambda_1 , \quad \delta(y 1^T ) = y \)
27. \( \delta(1) = \delta(1) = I \)
28. \( \delta(e_i e_j^T 1) = \delta(e_i) = e_i e_i^T \)
29. For \( \zeta = [\zeta_i] \in \mathbb{R}^k \) and \( x = [x_i] \in \mathbb{R}^k \), \( \sum_i \zeta_i / x_i = \zeta^T \delta(x)^{-1} 1 \)
30. \( \text{vec}(A \circ B) = \text{vec}(A) \circ \text{vec}(B) = \delta(\text{vec}(A)) \text{vec}(B) = \text{vec}(B) \circ \text{vec}(A) = \delta(\text{vec}(B)) \text{vec}(A) \quad (42)(1880) \)
31. \( \text{vec}(AXB) = (B^T \circ A) \text{vec} X \quad (\text{not } \mathbb{H}) \)
32. \( \text{vec}(BXA) = (A^T \circ B) \text{vec} X \)
33. \( \text{tr}(AXBX^T) = \text{vec}(X)^T \text{vec}(AXB) = \text{vec}(X)^T (B^T \circ A) \text{vec} X \quad [176] \)
34. \( \text{tr}(AX^T BX) = \text{vec}(X)^T \text{vec}(BXA) = \text{vec}(X)^T (A^T \circ B) \text{vec} X \quad \delta(\text{vec}(X)) \text{vec}(X)^T (A^T \circ B) \)
35. For any permutation matrix \( \Xi \) and dimensionally compatible vector \( y \) or matrix \( A \)

\[
\delta(\Xi y) = \Xi \delta(y) \Xi^T \quad (1501)
\]
\[
\delta(\Xi A \Xi^T) = \Xi \delta(A) \quad (1502)
\]

So given any permutation matrix \( \Xi \) and any dimensionally compatible matrix \( B \), for example,

\[
\delta^2(B) = \Xi \delta^2(\Xi^T B \Xi) \Xi^T \quad (1503)
\]
36. \( A \circ 1 = 1 \circ A = A \)
37. \( A \circ (B \circ C) = (A \circ B) \circ C \)
38. \( (A \circ B)(C \circ D) = AC \circ BD \)
39. For \( A \) a vector, \( (A \circ B) = (A \circ I) B \)
40. For \( B \) a row vector, \( (A \circ B) = A(I \circ B) \)
41. \( (A \circ B)^T = A^T \circ B^T \)
42. \( (A \circ B)^{-1} = A^{-1} \circ B^{-1} \)
43. \( \text{tr}(A \circ B) = \text{tr} A \text{ tr} B \)
44. For \( A \in \mathbb{R}^{m \times m} \), \( B \in \mathbb{R}^{n \times n} \), \( \det(A \circ B) = \det^n(A) \det^m(B) \)
45. There exist permutation matrices $\Xi_1$ and $\Xi_2$ such that \[ A \otimes B = \Xi_1(B \otimes A)\Xi_2 \tag{1504} \]

46. For eigenvalues $\lambda(A) \in \mathbb{C}^n$ and eigenvectors $v(A) \in \mathbb{C}^{n \times n}$ such that $A = v\delta(\lambda)v^{-1} \in \mathbb{R}^{n \times n}$

\[ \lambda(A \otimes B) = \lambda(A) \otimes \lambda(B), \quad v(A \otimes B) = v(A) \otimes v(B) \tag{1505} \]

47. Given analytic function $[85] f : \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$, then $f(I \otimes A) = I \otimes f(A)$ and $f(A \otimes I) = f(A) \otimes I$ \[176, p.28]\]

A.1.2 Majorization

A.1.2.0.1 Theorem. (Schur) Majorization. \[425, \S 7.4] \[212, \S 4.3] \[213, \S 5.5]\]

Let $\lambda \in \mathbb{R}^N$ denote a given vector of eigenvalues and let $\delta \in \mathbb{R}^N$ denote a given vector of main diagonal entries, both arranged in nonincreasing order. Then

\[ \exists A \in S^N \ni \lambda(A) = \lambda \text{ and } \delta(A) = \delta \iff \lambda - \delta \in K_{\lambda\delta}^* \tag{1506} \]

and conversely

\[ A \in S^N \Rightarrow \lambda(A) - \delta(A) \in K_{\lambda\delta}^* \tag{1507} \]

The difference belongs to the pointed polyhedral cone of majorization (not a full-dimensional cone, confer (312))

\[ K_{\lambda\delta}^* \triangleq K_{\lambda+}^* \cap \{ \zeta \mathbf{1} \mid \zeta \in \mathbb{R} \}^* \tag{1508} \]

where $K_{\lambda+}^*$ is the dual monotone nonnegative cone (434), and where the dual of the line is a hyperplane; $\partial \mathcal{H} = \{ \zeta \mathbf{1} \mid \zeta \in \mathbb{R} \}^* = \mathbf{1}^\perp$.

Majorization cone $K_{\lambda\delta}^*$ is naturally consequent to the definition of majorization: \textit{id est}, vector $y \in \mathbb{R}^N$majorizes vector $x$ if and only if

\[ \sum_{i=1}^k x_i \leq \sum_{i=1}^k y_i \quad \forall 1 \leq k \leq N \tag{1509} \]

and

\[ \mathbf{1}^T x = \mathbf{1}^T y \tag{1510} \]

Under these circumstances, rather, vector $x$ is majorized by vector $y$.

In the particular circumstance $\delta(A) = \mathbf{0}$ we get:

A.1.2.0.2 Corollary. Symmetric hollow majorization. Let $\lambda \in \mathbb{R}^N$ denote a given vector of eigenvalues arranged in nonincreasing order. Then

\[ \exists A \in S^N_h \ni \lambda(A) = \lambda \iff \lambda \in K_{\lambda\delta}^* \tag{1511} \]

and conversely

\[ A \in S^N_h \Rightarrow \lambda(A) \in K_{\lambda\delta}^* \tag{1512} \]

where $K_{\lambda\delta}^*$ is defined in (1508).
A.2. SEMIDEFINITENESS: DOMAIN OF TEST

The most fundamental necessary, sufficient, and definitive test for positive semidefiniteness of matrix \( A \in \mathbb{R}^{n \times n} \) is: [213, §1]

\[
x^T A x \geq 0 \quad \text{for each and every } x \in \mathbb{R}^n \text{ such that } \|x\| = 1 \quad (1513)
\]

Traditionally, authors demand evaluation over broader domain; namely, over all \( x \in \mathbb{R}^n \) which is sufficient but unnecessary. Indeed, that standard textbook requirement is far over-reaching because if \( x^T A x \) is nonnegative for particular \( x = x_p \), then it is nonnegative for any \( \alpha x_p \) where \( \alpha \in \mathbb{R} \). Thus, only normalized \( x \) in \( \mathbb{R}^n \) need be evaluated.

Many authors add the further requirement that the domain be complex; the broadest domain. By so doing, only Hermitian matrices (\( A^H = A \) where superscript \( H \) denotes conjugate transpose) are admitted to the set of positive semidefinite matrices (1516); an unnecessary prohibitive condition.

A.2.1 Symmetry versus semidefiniteness

We call (1513) the most fundamental test of positive semidefiniteness. Yet some authors instead say, for real \( A \) and complex domain \( \{ x \in \mathbb{C}^n \} \), the complex test \( x^H A x \geq 0 \) is most fundamental. That complex broadening of the domain of test causes nonsymmetric real matrices to be excluded from the set of positive semidefinite matrices. Yet admitting nonsymmetric real matrices or not is a matter of preference unless that complex test is adopted, as we shall now explain.

Any real square matrix \( A \) has a representation in terms of its symmetric and antisymmetric parts; id est,

\[
A = \frac{(A + A^T)}{2} + \frac{(A - A^T)}{2} \quad (53)
\]

Because, for all real \( A \), the antisymmetric part vanishes under real test,

\[
x^T \frac{(A - A^T)}{2} x = 0 \quad (1514)
\]

only the symmetric part of \( A \), \( (A + A^T)/2 \), has a role determining positive semidefiniteness. Hence the oft-made presumption that only symmetric matrices may be positive semidefinite is, of course, erroneous under (1513). Because eigenvalue-signs of a symmetric matrix translate unequivocally to its semidefiniteness, the eigenvalues that determine semidefiniteness are always those of the symmetrized matrix. (§A.3) For that reason, and because symmetric (or Hermitian) matrices must have real eigenvalues, the convention adopted in the literature is that semidefinite matrices are synonymous with symmetric semidefinite matrices. Certainly misleading under (1513), that presumption is typically bolstered with compelling examples from the physical sciences where symmetric matrices occur within the mathematical exposition of natural phenomena.

\[\text{A.2}\] Hermitian symmetry is the complex analogue; the real part of a Hermitian matrix is symmetric while its imaginary part is antisymmetric. A Hermitian matrix has real eigenvalues and real main diagonal.

\[\text{A.3}\] Golub & Van Loan [168, §4.2.2], for example, admit nonsymmetric real matrices.

\[\text{A.4}\] Symmetric matrices are certainly pervasive in the our chosen subject as well.
Perhaps a better explanation of this pervasive presumption of symmetry comes from Horn & Johnson [212, §7.1] whose perspective\textsuperscript{A.5} is the complex matrix, thus necessitating the complex domain of test throughout their treatise. They explain, if $A \in \mathbb{C}^{n \times n}$

$$\ldots \text{and if } x^H Ax \text{ is real for all } x \in \mathbb{C}^n, \text{ then } A \text{ is Hermitian. Thus, the assumption that } A \text{ is Hermitian is not necessary in the definition of positive definiteness. It is customary, however.}$$

Their comment is best explained by noting, the real part of $x^H Ax$ comes from the Hermitian part $(A + A^H)/2$ of $A$;

$$\text{re}(x^H Ax) = x^H \frac{A + A^H}{2} x$$  \hfill (1515)$$

rather,

$$x^H Ax \in \mathbb{R} \iff A^H = A$$  \hfill (1516)$$

because the imaginary part of $x^H Ax$ comes from the anti-Hermitian part $(A - A^H)/2$;

$$\text{im}(x^H Ax) = x^H \frac{A - A^H}{2} x$$  \hfill (1517)$$

that vanishes for nonzero $x$ if and only if $A = A^H$. So the Hermitian symmetry assumption is unnecessary, according to Horn & Johnson, not because nonHermitian matrices could be regarded positive semidefinite, rather because nonHermitian (includes nonsymmetric real) matrices are not comparable on the real line under $x^H Ax$. Yet that complex edifice is dismantled in the test of real matrices (1513) because the domain of test is no longer necessarily complex; meaning, $x^T Ax$ will certainly always be real, regardless of symmetry, and so real $A$ will always be comparable.

In summary, if we limit the domain of test to all $x$ in $\mathbb{R}^n$ as in (1513), then nonsymmetric real matrices are admitted to the realm of semidefinite matrices because they become comparable on the real line. One important exception occurs for rank-one matrices $\Psi = uv^T$ where $u$ and $v$ are real vectors: $\Psi$ is positive semidefinite if and only if $\Psi = uu^T$. (§A.3.1.0.7)

We might choose to expand the domain of test to all $x$ in $\mathbb{C}^n$ so that only symmetric matrices would be comparable. The alternative to expanding domain of test is to assume all matrices of interest to be symmetric; that is commonly done, hence the synonymous relationship with semidefinite matrices.

\textbf{A.2.1.0.1 Example. Nonsymmetric positive definite product.} Horn & Johnson assert and Zhang agrees:

\textit{If } $A, B \in \mathbb{C}^{n \times n}$ \textit{are positive definite, then we know that the product } $AB$ \textit{is positive definite if and only if } $AB$ \textit{is Hermitian.} [212, §7.6 prob.10] [425, §6.2, §3.2]

\textsuperscript{A.5} A totally complex perspective is not necessarily more advantageous. The positive semidefinite cone, for example, is not selfdual (§2.13.5) in the ambient space of Hermitian matrices. [205, §II]
A.3. PROPER STATEMENTS

Implicitly in their statement, \( A \) and \( B \) are assumed individually Hermitian and the domain of test is assumed complex.

We prove that assertion to be false for real matrices under (1513) that adopts a real domain of test.

\[
A^T = A = \begin{bmatrix}
3 & 0 & -1 & 0 \\
0 & 5 & 1 & 0 \\
-1 & 1 & 4 & 1 \\
0 & 0 & 1 & 4
\end{bmatrix}, \quad \lambda(A) = \begin{bmatrix}
5.9 \\
4.5 \\
3.4 \\
2.0
\end{bmatrix} \tag{1518}
\]

\[
B^T = B = \begin{bmatrix}
4 & 4 & -1 & -1 \\
4 & 5 & 0 & 0 \\
-1 & 0 & 5 & 1 \\
-1 & 0 & 1 & 4
\end{bmatrix}, \quad \lambda(B) = \begin{bmatrix}
8.8 \\
5.5 \\
3.3 \\
0.24
\end{bmatrix} \tag{1519}
\]

\[
(AB)^T \neq AB = \begin{bmatrix}
13 & 12 & -8 & -4 \\
19 & 25 & 5 & 1 \\
-5 & 1 & 22 & 9 \\
-5 & 0 & 9 & 17
\end{bmatrix}, \quad \lambda(AB) = \begin{bmatrix}
36. \\
29. \\
10. \\
0.72
\end{bmatrix} \tag{1520}
\]

\[
\frac{1}{2}(AB + (AB)^T) = \begin{bmatrix}
13 & 15.5 & -6.5 & -4.5 \\
15.5 & 25 & 3 & 0.5 \\
-6.5 & 3 & 22 & 9 \\
-4.5 & 0.5 & 9 & 17
\end{bmatrix}, \quad \lambda\left(\frac{1}{2}(AB + (AB)^T)\right) = \begin{bmatrix}
36. \\
30. \\
10. \\
0.014
\end{bmatrix} \tag{1521}
\]

Whenever \( A \in \mathbb{S}_n^+ \) and \( B \in \mathbb{S}_n^+ \), then \( \lambda(AB) = \lambda(\sqrt{A}B\sqrt{A}) \) will always be a nonnegative vector by (1545) and Corollary A.3.1.0.5. Yet positive definiteness of product \( AB \) is certified instead by the nonnegative eigenvalues \( \lambda(\frac{1}{2}(AB + (AB)^T)) \) in (1521) (§A.3.1.0.1) despite the fact that \( AB \) is not symmetric.\(^\text{A.6}\)

Horn & Johnson and Zhang resolve this anomaly by choosing to exclude nonsymmetric matrices and products; they do so by expanding the domain of test to \( \mathbb{C}^n \).

\[\Box\]

A.3 Proper statements of positive semidefiniteness

Unlike Horn & Johnson and Zhang, we never adopt a complex domain of test with real matrices. So motivated is our consideration of proper statements of positive semidefiniteness under real domain of test. This restriction, ironically, complicates the facts when compared to corresponding statements for the complex case (found elsewhere [212] [425]).

We state several fundamental facts regarding positive semidefiniteness of real matrix \( A \) and the product \( AB \) and sum \( A + B \) of real matrices under fundamental real test (1513); a few require proof as they depart from the standard texts, while those remaining are well established or obvious.

\[^{\text{A.6R}}\] is a little more difficult to find a counterexample in \( \mathbb{R}^{2 \times 2} \) or \( \mathbb{R}^{3 \times 3} \); which may have served to advance any confusion.
A.3.0.0.1 Theorem. Positive (semi)definite matrix.

\( A \in \mathbb{S}^M \) is positive semidefinite if and only if for each and every vector \( x \in \mathbb{R}^M \) of unit norm, \( \|x\| = 1 \), \(^{A.7}\) we have \( x^T A x \geq 0 \) \(^{(1513)}\);

\[ A \succeq 0 \iff \text{tr}(xx^T A) \geq 0 \quad \forall \, xx^T \] \(^{(1522)}\)

Matrix \( A \in \mathbb{S}^M \) is positive definite if and only if for each and every \( \|x\| = 1 \) we have \( x^T A x > 0 \);

\[ A \succ 0 \iff \text{tr}(xx^T A) > 0 \quad \forall \, xx^T, \, xx^T \neq 0 \] \(^{(1523)}\)

\[ \text{Proof.} \quad \text{Statements } (1522) \text{ and } (1523) \text{ are each a particular instance of dual generalized inequalities } (\text{\S}2.13.2) \text{ with respect to the positive semidefinite cone; \textit{videlicet}, [372]} \]

\[ A \succeq 0 \iff \langle xx^T, A \rangle \geq 0 \quad \forall \, xx^T \geq 0 \]

\[ A \succ 0 \iff \langle xx^T, A \rangle > 0 \quad \forall \, xx^T, \, xx^T \neq 0 \] \(^{(1524)}\)

This says: positive semidefinite matrix \( A \) must belong to the normal side of every hyperplane whose normal is an extreme direction of the positive semidefinite cone. Relations \( (1522) \) and \( (1523) \) remain true when \( xx^T \) is replaced with “for each and every” positive semidefinite matrix \( X \in \mathbb{S}_+^M \) \(^{(\text{\S}2.13.5)}\) of unit norm, \( \|X\| = 1 \), as in

\[ A \succeq 0 \iff \text{tr}(XA) \geq 0 \quad \forall \, X \in \mathbb{S}_+^M \]

\[ A \succ 0 \iff \text{tr}(XA) > 0 \quad \forall \, X \in \mathbb{S}_+^M, \, X \neq 0 \] \(^{(1525)}\)

But that condition is more than what is necessary. By the \textit{discretized membership theorem} in \( \text{\S}2.13.4.2.1 \), the extreme directions \( xx^T \) of the positive semidefinite cone constitute a minimal set of generators necessary and sufficient for discretization of dual generalized inequalities \( (1525) \) certifying membership to that cone. \( \blacklozenge \)

A.3.1 Semidefiniteness, eigenvalues, nonsymmetric

When \( A \in \mathbb{R}^{n \times n} \) let \( \lambda(A + A^T) \in \mathbb{R}^n \) denote eigenvalues of the symmetrized matrix \(^{A.8}\) arranged in nonincreasing order.

- By positive semidefiniteness of \( A \in \mathbb{R}^{n \times n} \) we mean, \(^{A.9}\)
  \([282, \text{\S}1.3.1] \) \( (\text{confer } \text{\S}A.3.1.0.1) \)

\[ x^T A x \geq 0 \quad \forall \, x \in \mathbb{R}^n \iff A + A^T \succeq 0 \iff \lambda(A + A^T) \geq 0 \] \(^{(1526)}\)

\(^{A.7}\) The traditional condition requiring all \( x \in \mathbb{R}^M \) for defining positive (semi)definiteness is actually more than what is necessary. The set of norm-1 vectors is necessary and sufficient to establish positive semidefiniteness; actually, any particular norm and any nonzero norm-constant will work.

\(^{A.8}\) The symmetrization of \( A \) is \( (A + A^T)/2 \). \( \lambda(A + A^T) = \lambda(A + A^T)/2 \).

\(^{A.9}\) Strang agrees [341, p.334] it is not \( \lambda(A) \) that requires observation. Yet he is mistaken by proposing the Hermitian part alone \( x^H(A + A^H)x \) be tested, because the anti-Hermitian part does not vanish under complex test unless \( A \) is Hermitian. \( (1517) \)
A.3. PROPER STATEMENTS

• (§2.9.0.1)

\[
A \succeq 0 \Rightarrow A^T = A
\]  
(1527)

\[
A \succeq B \iff A - B \succeq 0 \iff A \succeq 0 \text{ or } B \succeq 0
\]  
(1528)

\[
x^T A x \geq 0 \ \forall x \Rightarrow A^T = A
\]  
(1529)

Matrix symmetry is not intrinsic to positive semidefiniteness;

\[
A^T = A, \ \lambda(A) \succeq 0 \Rightarrow x^T A x \geq 0 \ \forall x
\]  
(1530)

\[
\lambda(A) \succeq 0 \iff A^T = A, \ x^T A x \geq 0 \ \forall x
\]  
(1531)

• If \(A^T = A\) then

\[
\lambda(A) \succeq 0 \iff A \succeq 0
\]  
(1532)

meaning, matrix \(A\) belongs to the positive semidefinite cone in the subspace of symmetric matrices if and only if its eigenvalues belong to the nonnegative orthant.

\[
\langle A, A \rangle = \langle \lambda(A), \lambda(A) \rangle
\]  
(45)

• For \(\mu \in \mathbb{R}, \ A \in \mathbb{R}^{n \times n}\), and vector \(\lambda(A) \in \mathbb{C}^n\) holding the ordered eigenvalues of \(A\)

\[
\lambda(\mu I + A) = \mu 1 + \lambda(A)
\]  
(1533)

**Proof.** \(A = M J M^{-1}\) and \(\mu I + A = M(\mu I + J)M^{-1}\) where \(J\) is the Jordan form for \(A\); [341, §5.6, App.B] id est, \(\delta(J) = \lambda(A)\), so \(\lambda(\mu I + A) = \delta(\mu I + J)\) because \(\mu I + J\) is also a Jordan form. ♦

By similar reasoning,

\[
\lambda(I + \mu A) = 1 + \lambda(\mu A)
\]  
(1534)

For vector \(\sigma(A)\) holding the singular values of any matrix \(A\)

\[
\sigma(I + \mu A^T A) = \pi(1 + \mu \sigma(A^T A))
\]  
(1535)

\[
\sigma(\mu I + A^T A) = \pi(1 + \sigma(A^T A))
\]  
(1536)

where \(\pi\) is the nonlinear permutation-operator sorting its vector argument into nonincreasing order.

• For \(A \in \mathbb{S}^M\) and each and every \(\|w\| = 1\) [212, §7.7 prob.9]

\[
w^T A w \leq \mu \iff A \preceq \mu I \iff \lambda(A) \preceq \mu 1
\]  
(1537)

• [212, §2.5.4] (*confer (44)*)

\(A\) is normal matrix \(\iff \|A\|^2_F = \lambda(A)^T \lambda(A)\)

(1538)
For $A \in \mathbb{R}^{m \times n}$
\[ A^T A \succeq 0, \quad AA^T \succeq 0 \] (1539)
because, for dimensionally compatible vector $x$,
\[ x^T A^T A x = \| Ax \|^2, \quad x^T A A^T x = \| A^T x \|^2. \]

For $A \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}$
\[ \text{tr}(cA) = c \text{tr}(A) = c \mathbf{1}^T \lambda(A) \] (§A.1.1 no. 4)

For $m$ a nonnegative integer, (1961)
\[ \det(A^m) = \prod_{i=1}^{n} \lambda(A)_i^m \] (1540)
\[ \text{tr}(A^m) = \sum_{i=1}^{n} \lambda(A)_i^m \] (1541)

For $A$ diagonalizable (§A.5), $A = S \Lambda S^{-1}$, (confer [341, p.255])
\[ \text{rank } A = \text{rank } \delta(\lambda(A)) = \text{rank } \Lambda \] (1542)
meaning, rank is equal to the number of nonzero eigenvalues in vector
\[ \lambda(A) \triangleq \delta(\Lambda) \] (1543)
by the 0 eigenvalues theorem (§A.7.3.0.1).

(Ky Fan) For $A, B \in S^n$ [57, §1.2] (confer (1833))
\[ \text{tr}(AB) \leq \lambda(A)^T \lambda(B) \] (1819)
with equality (Theobald) when $A$ and $B$ are simultaneously diagonalizable [212] with the same ordering of eigenvalues.

For $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$
\[ \text{tr}(AB) = \text{tr}(BA) \] (1544)
and $\eta$ eigenvalues of the product and commuted product are identical, including their multiplicity; [212, §1.3.20] id est,
\[ \lambda(AB)_{1:\eta} = \lambda(BA)_{1:\eta}, \quad \eta \triangleq \min\{m, n\} \] (1545)
Any eigenvalues remaining are zero. By the 0 eigenvalues theorem (§A.7.3.0.1),
\[ \text{rank}(AB) = \text{rank}(BA), \quad AB \text{ and } BA \text{ diagonalizable} \] (1546)

For any compatible matrices $A, B$ [212, §0.4]
\[ \min\{\text{rank } A, \text{rank } B\} \geq \text{rank}(AB) \] (1547)
A.3. PROPER STATEMENTS

- For $A, B \in S^n_+$

\[
\text{rank } A + \text{rank } B \geq \text{rank}(A + B) \geq \min\{\text{rank } A, \text{rank } B\} \geq \text{rank}(AB) \tag{1548}
\]

- For linearly independent matrices $A, B \in S^n_+$

\[
\text{rank } A + \text{rank } B = \text{rank}(A + B) > \min\{\text{rank } A, \text{rank } B\} \geq \text{rank}(AB) \tag{1549}
\]

- Because $\mathcal{R}(A^T A) = \mathcal{R}(A^T)$ and $\mathcal{R}(AA^T) = \mathcal{R}(A)$ (p. 546), for any $A \in \mathbb{R}^{m \times n}$

\[
\text{rank}(AA^T) = \text{rank}(A^T A) = \text{rank } A = \text{rank } A^T \tag{1550}
\]

- For $A \in \mathbb{R}^{m \times n}$ having no nullspace, and for any $B \in \mathbb{R}^{n \times k}$

\[
\text{rank}(AB) = \text{rank}(B) \tag{1551}
\]

**Proof.** For any compatible matrix $C$, $\mathcal{N}(CAB) \supseteq \mathcal{N}(AB) \supseteq \mathcal{N}(B)$ is obvious. By assumption $\exists A^\dagger \ni A^\dagger A = I$. Let $C = A^\dagger$, then $\mathcal{N}(AB) = \mathcal{N}(B)$ and the stated result follows by conservation of dimension (1667).

- For $A \in S^n$ and any nonsingular matrix $Y$

\[
\text{inertia}(A) = \text{inertia}(YAY^T) \tag{1552}
\]

a.k.a. Sylvester’s law of inertia. (1598) [119, §2.4.3]

- For $A, B \in \mathbb{R}^{n \times n}$ square, [212, §0.3.5]

\[
\det(AB) = \det(BA) \tag{1553}
\]

\[
\det(AB) = \det A \det B \tag{1554}
\]

Yet for $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$ [79, p.72]

\[
\det(I + AB) = \det(I + BA) \tag{1555}
\]

- For $A, B \in S^n$, product $AB$ is symmetric iff $AB$ is commutative;

\[
(AB)^T = AB \iff AB = BA \tag{1556}
\]

**Proof.** $(\Rightarrow)$ Suppose $AB = (AB)^T$.

$(AB)^T = B^T A^T = BA$. $AB = (AB)^T \Rightarrow AB = BA$.

$(\Leftarrow)$ Suppose $AB = BA$.

$BA = B^T A^T = (AB)^T$. $AB = BA \Rightarrow AB = (AB)^T$.

Commutativity alone is insufficient for product symmetry. [341, p.26] Matrix symmetry alone is insufficient for product symmetry.
• Diagonalizable matrices $A, B \in \mathbb{R}^{n \times n}$ commute if and only if they are simultaneously diagonalizable. [212, §1.3.12] A product of diagonal matrices is always commutative.

• For $A, B \in \mathbb{R}^{n \times n}$ and $AB = BA$

$$x^TAx \geq 0, \ x^TBx \geq 0 \ \forall \ x \ \Rightarrow \ \lambda(A + A^T), \ \lambda(B + B^T) \geq 0 \ \forall \ i \ \Leftrightarrow \ x^TAXBx \geq 0 \ \forall \ x \ (1557)$$

the negative result arising because of the schism between the product of eigenvalues $\lambda(A + A^T), \ \lambda(B + B^T)$, and the eigenvalues of the symmetrized matrix product $\lambda(AB + (AB)^T)$. For example, $X^2$ is generally not positive semidefinite unless matrix $X$ is symmetric; then (1537) applies. Simply substituting symmetric matrices changes the outcome:

• For $A, B \in \mathbb{S}^n$ and $AB = BA$

$$A \succeq 0, \ B \succeq 0 \ \Rightarrow \ \lambda(AB)_i = \lambda(A)_i \lambda(B)_i \geq 0 \ \forall \ i \ \Leftrightarrow \ AB \succeq 0 \ (1558)$$

Positive semidefiniteness of commutative $A$ and $B$ is sufficient but not necessary for positive semidefiniteness of product $AB$.

**Proof.** Because all symmetric matrices are diagonalizable, (§A.5.1) [341, §5.6] we have $A = \Lambda \Delta S^T$ and $B = T \Delta T^T$, where $\Lambda$ and $\Delta$ are real diagonal matrices while $S$ and $T$ are orthogonal matrices. Because $(AB)^T = AB$, then $T$ must equal $S$, [212, §1.3] and the eigenvalues of $A$ are ordered in the same way as those of $B$; id est, $\lambda(A)_i = \delta(\Lambda)_i$ and $\lambda(B)_i = \delta(\Delta)_i$ correspond to the same eigenvector.

$(\Rightarrow)$ Assume $\lambda(A)_i \lambda(B)_i \geq 0$ for $i = 1 \ldots n$. $AB = \Lambda \Delta S^T$ is symmetric and has nonnegative eigenvalues contained in diagonal matrix $\Lambda \Delta$ by assumption; hence positive semidefinite by (1526). Now assume $A, B \succeq 0$. That, of course, implies $\lambda(A)_i \lambda(B)_i \geq 0$ for all $i$ because all the individual eigenvalues are nonnegative.

$(\Leftarrow)$ Suppose $AB = \Lambda \Delta S^T \succeq 0$. Then $\Lambda \Delta \succeq 0$ by (1526), and so all products $\lambda(A)_i \lambda(B)_i$ must be nonnegative; meaning, $\text{sgn}(\lambda(A)) = \text{sgn}(\lambda(B))$. We may, therefore, conclude nothing about the semidefiniteness of $A$ and $B$.

• For $A, B \in \mathbb{S}^n$ and $A \succeq 0, \ B \succeq 0$ (Example A.2.1.0.1)

$$AB = BA \Rightarrow \lambda(AB)_i = \lambda(A)_i \lambda(B)_i \geq 0 \ \forall \ i \ \Rightarrow \ AB \succeq 0 \ (1559)$$

$$AB = BA \Rightarrow \lambda(AB)_i \geq 0, \ \lambda(A)_i \lambda(B)_i \geq 0 \ \forall \ i \ \Leftrightarrow \ AB \succeq 0 \ (1560)$$

• For $A, B \in \mathbb{S}^n$ [212, §7.7 prob.3] [213, §4.2.13, §5.2.1]

$$A \succeq 0, \ B \succeq 0 \ \Rightarrow \ A \otimes B \succeq 0 \ (1561)$$

$$A \succeq 0, \ B \succeq 0 \ \Rightarrow \ A \circ B \succeq 0 \ (1562)$$

$$A \succ 0, \ B \succ 0 \ \Rightarrow \ A \otimes B \succ 0 \ (1563)$$

$$A \succ 0, \ B \succ 0 \ \Rightarrow \ A \circ B \succ 0 \ (1564)$$

where Kronecker and Hadamard products are symmetric.
A.3. PROPER STATEMENTS

- For $A, B \in \mathbb{S}^n$, (1532) $A \succeq 0 \iff \lambda(A) \succeq 0$ yet
  
  
  
  $A \succeq 0 \Rightarrow \delta(A) \succeq 0$     
  (1565)
  $A \succeq 0 \Rightarrow \text{tr} A \succeq 0$     
  (1566)
  
  $A \succeq 0, \ B \succeq 0 \Rightarrow \text{tr} A \text{tr} B \succeq \text{tr}(AB) \succeq 0$     
  (1567)

[425, §6.2] Because $A \succeq 0, \ B \succeq 0 \Rightarrow \lambda(AB) = \lambda(\sqrt{A}B\sqrt{A}) \succeq 0$ by (1545) and Corollary A.3.1.0.5, then we have $\text{tr}(AB) \succeq 0$.

\[
A \succeq 0 \iff \text{tr}(AB) \succeq 0 \ \forall \ B \succeq 0 \quad (377)
\]

- For $A, B, C \in \mathbb{S}^n$ (Löwner)
  
  $A \preceq B, \ B \leq C \Rightarrow A \leq C$     
  (transitivity)
  $A \preceq B \iff A + C \preceq B + C$     
  (additivity)
  $A \preceq B, \ A \succeq B \Rightarrow A = B$     
  (antisymmetry)
  $A \leq A$     
  (reflexivity)
  $A \preceq B, \ B \prec C \Rightarrow A \prec C$     
  (strict transitivity)
  $A \prec B \iff A + C \prec B + C$     
  (strict additivity)     
  (1569)

- For $A, B \in \mathbb{R}^{n \times n}$
  
  $x^T A x \geq x^T B x \ \forall \ x \Rightarrow \text{tr} A \geq \text{tr} B$     
  (1570)

Proof. $x^T A x \geq x^T B x \ \forall \ x \iff \lambda((A - B) + (A - B)^T)/2 \geq 0 \Rightarrow \text{tr}(A + A^T - (B + B^T))/2 = \text{tr}(A - B) \geq 0$. There is no converse. ♦

- For $A, B \in \mathbb{S}^n$ [425, §6.2] (Theorem A.3.1.0.4)
  
  $A \succeq B \Rightarrow \text{tr} A \geq \text{tr} B$     
  (1571)
  $A \succeq B \Rightarrow \delta(A) \succeq \delta(B)$     
  (1572)

There is no converse, and restriction to the positive semidefinite cone does not improve the situation. All-strict versions hold.

  $A \succeq B \geq 0 \Rightarrow \text{rank} A \geq \text{rank} B$     
  (1573)
  $A \succeq B \geq 0 \Rightarrow \det A \geq \det B \geq 0$     
  (1574)
  $A \succ B \geq 0 \Rightarrow \det A > \det B \geq 0$     
  (1575)

- For $A, B \in \text{int} \mathbb{S}^n_+$ [35, §4.2] [212, §7.7.4]
  
  $A \succeq B \iff A^{-1} \succeq B^{-1}, \ A \succ 0 \iff A^{-1} \succ 0$     
  (1576)

- For $A, B \in \mathbb{S}^n$ [425, §6.2] (1577)
  
  $A \succeq B \geq 0 \Rightarrow \sqrt{A} \succeq \sqrt{B}$
  $A \succeq 0 \iff A^{1/2} \succeq 0$

- For $A, B \in \mathbb{S}^n$ and $AB = BA$ [425, §6.2 prob.3]
  
  $A \succeq B \geq 0 \Rightarrow A^k \succeq B^k, \ k = 1, 2, \ldots$     
  (1578)
APPENDIX A. LINEAR ALGEBRA

A.3.1.0.1 Theorem. Positive semidefinite ordering of eigenvalues.
For $A, B \in \mathbb{R}^{M \times M}$, place the eigenvalues of each symmetrized matrix into the respective vectors $\lambda\left(\frac{1}{2}(A + A^T)\right), \lambda\left(\frac{1}{2}(B + B^T)\right) \in \mathbb{R}^M$. Then [341, §6]

$$
x^T A x \geq 0 \ \forall x \iff \lambda(A + A^T) \geq 0 \quad (1579)
$$

$$
x^T A x > 0 \ \forall x \neq 0 \iff \lambda(A + A^T) > 0 \quad (1580)
$$

because $x^T (A - A^T) x = 0$. (1514) Now arrange the entries of $\lambda\left(\frac{1}{2}(A + A^T)\right)$ and $\lambda\left(\frac{1}{2}(B + B^T)\right)$ in nonincreasing order so $\lambda\left(\frac{1}{2}(A + A^T)\right)$ holds the largest eigenvalue of symmetrized $A$ while $\lambda\left(\frac{1}{2}(B + B^T)\right)$ holds the largest eigenvalue of symmetrized $B$, and so on. Then [212, §7.7 prob.1 prob.9] for $\kappa \in \mathbb{R}$

$$
x^T A x \geq x^T B x \ \forall x \ \Rightarrow \ \lambda(A + A^T) \geq \lambda(B + B^T) \quad (1581)
$$

$$
x^T A x \geq x^T I x \kappa \ \forall x \ \iff \ \lambda\left(\frac{1}{2}(A + A^T)\right) \geq \kappa 1
$$

Now let $A, B \in \mathbb{S}^M$ have diagonalizations $A = Q\Lambda Q^T$ and $B = U\Upsilon U^T$ with $\lambda(A) = \delta(A)$ and $\lambda(B) = \delta(T)$ arranged in nonincreasing order. Then

$$
A \succeq B \iff \lambda(A - B) \geq 0 \
A \succeq B \Rightarrow \lambda(A) \geq \lambda(B) \
A \succeq B \iff \lambda(A) \geq \lambda(B) \
S^T A S \succeq B \iff \lambda(A) \geq \lambda(B) 
$$

where $S = QU^T$. [425, §7.5]

A.3.1.0.2 Theorem. (Weyl) Eigenvalues of sum. [212, §4.3.1]
For $A, B \in \mathbb{R}^{M \times M}$, place the eigenvalues of each symmetrized matrix into the respective vectors $\lambda\left(\frac{1}{2}(A + A^T)\right), \lambda\left(\frac{1}{2}(B + B^T)\right) \in \mathbb{R}^M$ in nonincreasing order so $\lambda\left(\frac{1}{2}(A + A^T)\right)$ holds the largest eigenvalue of symmetrized $A$ while $\lambda\left(\frac{1}{2}(B + B^T)\right)$ holds the largest eigenvalue of symmetrized $B$, and so on. Then, for any $k \in \{1 \ldots M\}$

$$
\lambda(A + A^T)_k + \lambda(B + B^T)_M \leq \lambda((A + A^T) + (B + B^T))_k \leq \lambda(A + A^T)_k + \lambda(B + B^T)_1 \quad (1586)
$$

Weyl’s theorem establishes: concavity of the smallest $\lambda_M$ and convexity of the largest eigenvalue $\lambda_1$ of a symmetric matrix, via (496), and positive semidefiniteness of a sum of positive semidefinite matrices; for $A, B \in \mathbb{S}_+^M$

$$
\lambda(A)_k + \lambda(B)_M \leq \lambda(A + B)_k \leq \lambda(A)_k + \lambda(B)_1 \quad (1587)
$$

Because $\mathbb{S}_+^M$ is a convex cone (§2.9.0.0.1), then by (175)

$$
A, B \succeq 0 \ \Rightarrow \ \zeta A + \xi B \succeq 0 \ \text{for all} \ \zeta, \xi \geq 0 \quad (1588)
$$
A.3. PROPER STATEMENTS

A.3.1.0.3 Corollary. Eigenvalues of sum and difference. \[212, \S 4.3\]
For \(A \in \mathbb{S}^M\) and \(B \in \mathbb{S}^M_+\), place the eigenvalues of each matrix into respective vectors \(\lambda(A), \lambda(B) \in \mathbb{R}^M\) in nonincreasing order so \(\lambda(A)_1\) holds the largest eigenvalue of \(A\) while \(\lambda(B)_1\) holds the largest eigenvalue of \(B\), and so on. Then, for any \(k \in \{1 \ldots M\}\)
\[
\lambda(A - B)_k \leq \lambda(A)_k \leq \lambda(A + B)_k
\]
(1589)

When \(B\) is rank-one positive semidefinite, the eigenvalues interlace; id est, for \(B = qq^T\)
\[
\lambda(A)_{k-1} \leq \lambda(A - qq^T)_k \leq \lambda(A)_k \leq \lambda(A + qq^T)_k \leq \lambda(A)_{k+1}
\]
(1590)

A.3.1.0.4 Theorem. Positive (semi)definite principal submatrices. \[212, \S 7.2\]
- \(A \in \mathbb{S}^M\) is positive semidefinite if and only if all \(M\) principal submatrices of dimension \(M - 1\) are positive semidefinite and \(det A\) is nonnegative.
- \(A \in \mathbb{S}^M\) is positive definite if and only if any one principal submatrix of dimension \(M - 1\) is positive definite and \(det A\) is positive.

If any one principal submatrix of dimension \(M - 1\) is not positive definite, conversely, then \(A\) can neither be. Regardless of symmetry, if \(A \in \mathbb{R}^{M \times M}\) is positive (semi)definite, then the determinant of each and every principal submatrix is (nonnegative) positive. \[282, \S 1.3.1\]

A.3.1.0.5 Corollary. Positive (semi)definite symmetric products. \[212, \text{p.399}\]
- If \(A \in \mathbb{S}^M\) is positive definite and any particular dimensionally compatible matrix \(Z\) has no nullspace, then \(Z^T AZ\) is positive definite.
- If matrix \(A \in \mathbb{S}^M\) is positive (semi)definite then, for any matrix \(Z\) of compatible dimension, \(Z^T AZ\) is positive semidefinite.
- \(A \in \mathbb{S}^M\) is positive (semi)definite if and only if there exists a nonsingular \(Z\) such that \(Z^T AZ\) is positive (semi)definite.
- If \(A \in \mathbb{S}^M\) is positive semidefinite and singular it remains possible, for some skinny \(Z \in \mathbb{R}^{M \times N}\) with \(N < M\), that \(Z^T AZ\) becomes positive definite. \[212, \text{\S 6.3}\]

We can deduce from these, given nonsingular matrix \(Z\) and any particular dimensionally compatible \(Y\): matrix \(A \in \mathbb{S}^M\) is positive semidefinite if and only if
\[
\begin{bmatrix}
Z^T \\
Y^T
\end{bmatrix}
\begin{bmatrix}
Z & Y
\end{bmatrix}
\]
is positive semidefinite. In other words, from the Corollary it follows: for dimensionally compatible \(Z\)

\[\text{A.10: A recursive condition for positive (semi)definiteness, this theorem is a synthesis of facts from } [212, \S 7.2] [341, \S 6.3] \text{ (confen } [282, \S 1.3.1]).\]

\[\text{A.11: Using the interpretation in } \text{\S } 6.4.3, \text{ this means coefficients of orthogonal projection of vectorized } A \text{ on a subset of extreme directions from } \mathbb{S}^M_+ \text{ determined by } Z \text{ can be positive.}\]
• $A \succeq 0 \iff Z^T A Z \succeq 0$ and $Z^T$ has a left inverse.

Products such as $Z^T Z$ and $ZZ^T$ are symmetric and positive semidefinite although, given $A \succeq 0$, $Z^T A Z$ and $Z A Z^T$ are neither necessarily symmetric or positive semidefinite.

**A.3.1.0.6 Theorem.** Symmetric projector semidefinite. [21, §III] [22, §6] [232, p.55]

For symmetric idempotent matrices $P$ and $R$

$$P, R \succeq 0$$

$$P \succeq R \iff \mathcal{R}(P) \supseteq \mathcal{R}(R) \iff \mathcal{N}(P) \subseteq \mathcal{N}(R)$$

Projector $P$ is never positive definite [343, §6.5 prob.20] unless it is the Identity matrix.

**A.3.1.0.7 Theorem.** Symmetric positive semidefinite. [212, p.400]

Given real matrix $\Psi$ with rank $\Psi = 1$

$$\Psi \succeq 0 \iff \Psi = uu^T$$

where $u$ is some real vector; id est, symmetry is necessary and sufficient for positive semidefiniteness of a rank-1 matrix.

**Proof.** Any rank-one matrix must have the form $\Psi = u v^T$. (§B.1) Suppose $\Psi$ is symmetric; id est, $v = u$. For all $y \in \mathbb{R}^M$, $y^T u u^T y \geq 0$. Conversely, suppose $u v^T$ is positive semidefinite. We know that can hold if and only if $u v^T + v u^T \succeq 0 \iff$ for all normalized $y \in \mathbb{R}^M$, $2 y^T u v^T y \geq 0$; but that is possible only if $v = u$.

The same does not hold true for matrices of higher rank, as Example A.2.1.0.1 shows.

### A.4 Schur complement

Consider *Schur-form* partitioned matrix $G$: Given $A^T = A$ and $C^T = C$, then [61]

$$G = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$$

$$\iff A \succeq 0, \quad B^T (I - A A^T) = 0, \quad C - B^T A^T B \succeq 0$$

$$\iff C \succeq 0, \quad B (I - C C^T) = 0, \quad A - B C^T B^T \succeq 0$$

(1593)

where $A^\dagger$ denotes the Moore-Penrose (pseudo)inverse (§E). In the first instance, $I - A A^\dagger$ is a symmetric projection matrix orthogonally projecting on $\mathcal{N}(A^T)$. (2004) It is apparently required

$$\mathcal{R}(B) \perp \mathcal{N}(A^T)$$

(1594)

which precludes $A = 0$ when $B$ is any nonzero matrix. Note that $A \succ 0 \Rightarrow A^\dagger = A^{-1}$; thereby, the projection matrix vanishes. Likewise, in the second instance, $I - C C^\dagger$ projects orthogonally on $\mathcal{N}(C^T)$. It is required

$$\mathcal{R}(B^T) \perp \mathcal{N}(C^T)$$

(1595)
which precludes \( C = 0 \) for \( B \) nonzero. Again, \( C > 0 \) \( \Rightarrow \) \( C^+ = C^{-1} \). So we get, for \( A \) or \( C \) nonsingular,

\[
G = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0
\]

\[
\iff A > 0, \ C - B^T A^{-1} B \succeq 0
\]

or

\[
C > 0, \ A - B C^{-1} B^T \succeq 0
\]

When \( A \) is full-rank then, for all \( B \) of compatible dimension, \( \mathcal{R}(B) \) is in \( \mathcal{R}(A) \). Likewise, when \( C \) is full-rank, \( \mathcal{R}(B^T) \) is in \( \mathcal{R}(C) \). Thus the flavor, for \( A \) and \( C \) nonsingular,

\[
G = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} > 0
\]

\[
\iff A > 0, \ C - B^T A^{-1} B > 0
\]

\[
\iff C > 0, \ A - B C^{-1} B^T > 0
\]

where \( C - B^T A^{-1} B \) is called the Schur complement of \( A \) in \( G \), while the Schur complement of \( C \) in \( G \) is \( A - B C^{-1} B^T \). \cite{161, §4.8}

Origin of the term Schur complement is from complementary inertia: \cite{119, §2.4.4} Define

\[
\text{inertia}(G \in \mathbb{S}^M) \triangleq \{p, z, n\}
\]

where \( p, z, n \) respectively represent number of positive, zero, and negative eigenvalues of \( G \); id est,

\[
M = p + z + n
\]

Then, when \( A \) is invertible,

\[
\text{inertia}(G) = \text{inertia}(A) + \text{inertia}(C - B^T A^{-1} B)
\]

and when \( C \) is invertible,

\[
\text{inertia}(G) = \text{inertia}(C) + \text{inertia}(A - B C^{-1} B^T)
\]

\[A.4.0.0.1\] Example. Equipartition inertia. \cite[§1.2 exer.17]{57}

When \( A = C = 0 \), denoting nonincreasingly ordered singular values of matrix \( B \in \mathbb{R}^{m \times m} \) by \( \sigma(B) \in \mathbb{R}^+_m \), then we have eigenvalues

\[
\lambda(G) = \lambda \left( \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \right) = \begin{bmatrix} \sigma(B) \\ -\Xi \sigma(B) \end{bmatrix}
\]

and

\[
\text{inertia}(G) = \text{inertia}(B^T B) + \text{inertia}(-B^T B)
\]

where \( \Xi \) is the order-reversing permutation matrix defined in (1820). \( \square \)
**APPENDIX A. LINEAR ALGEBRA**

### A.4.0.0.2 Example. Nonnegative polynomial.

[35, p.163]

Quadratic multivariate polynomial $x^TAx + 2b^Tx + c$ is a convex function of vector $x$ if and only if $A \succeq 0$, but sublevel set $\{ x \mid x^TAx + 2b^Tx + c \leq 0 \}$ is convex if $A \succeq 0$ yet not *vice versa*. Schur-form positive semidefiniteness is sufficient for polynomial convexity but necessary and sufficient for nonnegativity:

$$
\begin{bmatrix}
A & b \\
b^T & c
\end{bmatrix} \succeq 0 \iff 
\begin{bmatrix}
x^T & 1
\end{bmatrix}
\begin{bmatrix}
A & b \\
b^T & c
\end{bmatrix}
\begin{bmatrix}
x \\
1
\end{bmatrix} \succeq 0 \iff x^TAx + 2b^Tx + c \geq 0 \quad (1604)
$$

All is extensible to univariate polynomials; e.g., $x \triangleq [t^n \ t^{n-1} \ t^{n-2} \ldots t]^T$.

### A.4.0.0.3 Example. Schur-form fractional function trace minimization.

From (1566),

$$
\begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} \succeq 0 \implies \text{tr}(A + C) \geq 0
$$

$$
\downarrow
$$

$$
\begin{bmatrix}
A & 0 \\
0^T & C - B^TA^{-1}B
\end{bmatrix} \succeq 0 \implies \text{tr}(C - B^TA^{-1}B) \geq 0 \quad (1605)
$$

$$
\downarrow
$$

$$
\begin{bmatrix}
A - B C^{-1}B^T & 0 \\
0^T & C
\end{bmatrix} \succeq 0 \implies \text{tr}(A - B C^{-1}B^T) \geq 0
$$

Since $\text{tr}(C - B^TA^{-1}B) \geq 0 \iff \text{tr} C \geq \text{tr}(B^TA^{-1}B) \geq 0$ for example, then minimization of $\text{tr} C$ is necessary and sufficient for minimization of $\text{tr}(C - B^TA^{-1}B)$ when both are under constraint $\begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} \succeq 0$.

### A.4.0.1 Schur-form nullspace basis

From (1593),

$$
G = \begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} \succeq 0
$$

$$
\downarrow
$$

$$
\begin{bmatrix}
A & 0 \\
0^T & C - B^TA^TA
\end{bmatrix} \succeq 0 \quad \text{and} \quad B^T(I - AA^T) = 0 \quad (1606)
$$

$$
\downarrow
$$

$$
\begin{bmatrix}
A - B C^TA^TB & 0 \\
0^T & C
\end{bmatrix} \succeq 0 \quad \text{and} \quad B(I - CC^T) = 0
$$

These facts plus Moore-Penrose condition (§E.0.1) provide a partial basis:

$$
basis \mathcal{N} \left( \begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} \right) \succeq \begin{bmatrix}
I - AA^T & 0 \\
0^T & I - CC^T
\end{bmatrix} \quad (1607)
$$
A.4. SCHUR COMPLEMENT

A.4.0.1.1 Example. Sparse Schur conditions.
Setting matrix $A$ to the Identity simplifies the Schur conditions. One consequence relates definiteness of three quantities:

\[
\begin{bmatrix}
I & B \\
B^T & C
\end{bmatrix} \succeq 0 \iff C - B^T B \succeq 0 \iff \begin{bmatrix}
I & 0 \\
0 & C - B^T B
\end{bmatrix} \succeq 0
\]

(1608)

\[\Box\]

A.4.0.1.2 Exercise. Eigenvalues $\lambda$ of sparse Schur-form.
Prove: given $C - B^T B = 0$, for $B \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{S}^n$

\[
\lambda\left(\begin{bmatrix}
I \\
B^T \\
C
\end{bmatrix}\right)_i = \begin{cases}
1 + \lambda(C)_i, & 1 \leq i \leq n \\
1, & n < i \leq m \\
0, & \text{otherwise}
\end{cases}
\]

(1609)

\[\blacklozenge\]

A.4.0.1.3 Theorem. Rank of partitioned matrices. \cite[§2.2 prob.7]{425}
When symmetric matrix $A$ is invertible and $C$ is symmetric,

\[
\text{rank}\begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} = \text{rank}\begin{bmatrix}
A & 0 \\
0 & C - B^T A^{-1} B
\end{bmatrix} = \text{rank} A + \text{rank}(C - B^T A^{-1} B)
\]

(1610)

equals rank of main diagonal block $A$ plus rank of its Schur complement. Similarly, when symmetric matrix $C$ is invertible and $A$ is symmetric,

\[
\text{rank}\begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} = \text{rank}\begin{bmatrix}
A - B C^{-1} B^T & 0 \\
0 & C
\end{bmatrix} = \text{rank}(A - B C^{-1} B^T) + \text{rank} C
\]

(1611)

\[\Diamond\]

Proof. The first assertion (1610) holds if and only if \cite[§0.4.6c]{212} \[212, §7.7.6\]

\[
\exists \text{nonsingular } X, Y \ni X\begin{bmatrix}
A & B \\
B^T & C
\end{bmatrix} Y = \begin{bmatrix}
A & 0 \\
0 & C - B^T A^{-1} B
\end{bmatrix}
\]

(1612)

Let \cite[§7.7.6]{212}

\[
Y = X^T = \begin{bmatrix}
I & -A^{-1} B \\
0^T & I
\end{bmatrix}
\]

(1613)

\[\blacklozenge\]

From Corollary A.3.1.0.3, eigenvalues are related by

\[
0 \leq \lambda(C - B^T A^{-1} B) \leq \lambda(C)
\]

(1614)

\[
0 \leq \lambda(A - B C^{-1} B^T) \leq \lambda(A)
\]

(1615)
which means
\[
\text{rank}(C - B^T A^{-1} B) \leq \text{rank} C \\
\text{rank}(A - B C^{-1} B^T) \leq \text{rank} A
\] (1616)
\[
\text{rank}(A - B C^{-1} B^T) \leq \text{rank} A
\] (1617)

Therefore
\[
\text{rank} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \leq \text{rank} A + \text{rank} C
\] (1618)

**A.4.0.1.4 Lemma.** Rank of Schur-form block. \([143] \ [141]\)

Matrix \(B \in \mathbb{R}^{m \times n}\) has \(\text{rank} B \leq \rho\) if and only if there exist matrices \(A \in S^m\) and \(C \in S^n\) such that
\[
\text{rank} \begin{bmatrix} A & 0 \\ 0^T & C \end{bmatrix} \leq 2 \rho \quad \text{and} \quad G = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \quad (1619)
\]

Schur-form positive semidefiniteness alone implies \(\text{rank} A \geq \text{rank} B\) and \(\text{rank} C \geq \text{rank} B\). But, even in absence of semidefiniteness, we must always have \(\text{rank} G \geq \text{rank} A, \text{rank} B, \text{rank} C\) by fundamental linear algebra.

**A.4.1 Determinant**
\[
G = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \quad (1620)
\]

We consider again a matrix \(G\) partitioned like (1593), but not necessarily positive (semi)definite, where \(A\) and \(C\) are symmetric.

- When \(A\) is invertible,
  \[
  \det G = \det A \ \det(C - B^T A^{-1} B) \quad (1621)
  \]
  When \(C\) is invertible,
  \[
  \det G = \det C \ \det(A - B C^{-1} B^T) \quad (1622)
  \]

- When \(B\) is full-rank and skinny, \(C = 0\), and \(A \succeq 0\), then [63, §10.1.1]
  \[
  \det G \neq 0 \iff A + BB^T \succ 0 \quad (1623)
  \]
  When \(B\) is a (column) vector, then for all \(C \in \mathbb{R}\) and all \(A\) of dimension compatible with \(G\)
  \[
  \det G = \det(A)C - B^T A_{\text{cof}} B \quad (1624)
  \]
  while for \(C \neq 0\)
  \[
  \det G = C \det(A - \frac{1}{C} BB^T) \quad (1625)
  \]
  where \(A_{\text{cof}}\) is the matrix of cofactors [341, §4] corresponding to \(A\).
When $B$ is full-rank and fat, $A = 0$, and $C \succeq 0$, then
\[ \det G \neq 0 \iff C + B^T B > 0 \] (1626)
When $B$ is a row-vector, then for $A \neq 0$ and all $C$ of dimension compatible with $G$
\[ \det G = A \det(C - \frac{1}{A} B^T B) \] (1627)
while for all $A \in \mathbb{R}$
\[ \det G = \det(C)A - BC_{\text{cof}}B^T \] (1628)
where $C_{\text{cof}}$ is the matrix of cofactors corresponding to $C$.

A.5 Eigenvalue decomposition

All square matrices have associated eigenvalues $\lambda$ and eigenvectors; if not square, $Ax = \lambda_i x$ becomes impossible dimensionally. Eigenvectors must be nonzero. Prefix eigen is from the German; in this context meaning, something akin to “characteristic”. [338, p.14]

When a square matrix $X \in \mathbb{R}^{m \times m}$ is diagonalizable, [341, §5.6] then
\[ X = SAS^{-1} = [s_1 \cdots s_m] \Lambda \begin{bmatrix} w_{1}^T \\ \vdots \\ w_{m}^T \end{bmatrix} = \sum_{i=1}^{m} \lambda_i s_i w_{i}^T \] (1629)
where $\{s_i \in \mathcal{N}(X - \lambda_i I) \subseteq \mathbb{C}^m\}$ are l.i. (right-)eigenvectors constituting the columns of $S \in \mathbb{C}^{m \times m}$ defined by
\[ XS = SA \quad \text{rather} \quad Xs_i \triangleq \lambda_i s_i, \quad i = 1 \ldots m \] (1630)
$\{w_i \in \mathcal{N}(X^T - \lambda_i I) \subseteq \mathbb{C}^m\}$ are linearly independent left-eigenvectors of $X$ (eigenvectors of $X^T$) constituting the rows of $S^{-1}$ defined by [212]
\[ S^{-1}X = \Lambda S^{-1} \quad \text{rather} \quad w_i^T X \triangleq \lambda_i w_i^T, \quad i = 1 \ldots m \] (1631)
and where $\{\lambda_i \in \mathbb{C}\}$ are eigenvalues (1543)
\[ \delta(\lambda(X)) = \Lambda \in \mathbb{C}^{m \times m} \] (1632)
corresponding to both left and right eigenvectors; id est, $\lambda(X) = \lambda(X^T)$.

There is no connection between diagonalizability and invertibility of $X$. [341, §5.2] Diagonalizability is guaranteed by a full set of linearly independent eigenvectors, whereas invertibility is guaranteed by all nonzero eigenvalues.

\[ \text{distinct eigenvalues } \Rightarrow \text{l.i. eigenvectors } \Leftrightarrow \text{diagonalizable} \]
\[ \text{not diagonalizable } \Rightarrow \text{repeated eigenvalue} \] (1633)

\[
\begin{pmatrix}
1 & 0 & -1 \\
-1 & 1 & 0 \\
3 & -1 & -2
\end{pmatrix}
\]
is not diagonalizable, for example, having three 0-eigenvalues which are hard to compute with accuracy better than $1\text{E}-6$. (Yates, D’Errico)
A.5.0.0.1 Theorem. Real eigenvector.
Eigenvectors of a real matrix corresponding to real eigenvalues must be real. ♦

Proof. $Ax = \lambda x$. Given $\lambda = \lambda^*$, $x^H A x = \lambda ||x||^2 = x^* A x^* \Rightarrow x = x^*$, where $x^H = x^T$. The converse is equally simple. ♦

A.5.0.1 Uniqueness

From the fundamental theorem of algebra, [229] which guarantees existence of zeros for a given polynomial, it follows: eigenvalues, including their multiplicity, for a given square matrix are unique; meaning, there is no other set of eigenvalues for that matrix. (Conversely, many different matrices may share the same unique set of eigenvalues; e.g., for any $X$, $\lambda(X) = \lambda(X^T)$.)

Uniqueness of eigenvectors, in contrast, disallows multiplicity of the same direction:

A.5.0.1.1 Definition. Unique eigenvectors.
When eigenvectors are unique, we mean: unique to within a real nonzero scaling, and their directions are distinct. △

If $S$ is a matrix of eigenvectors of $X$ as in (1629), for example, then $-S$ is certainly another matrix of eigenvectors decomposing $X$ with the same eigenvalues. Although directions are distinct, eigenvectors $-S$ are equivalent to eigenvectors $S$ by Definition A.5.0.1.1.

For any square matrix, the eigenvector corresponding to a distinct eigenvalue is unique; [338, p.220]

\[ \text{distinct eigenvalues } \Rightarrow \text{eigenvectors unique} \quad (1634) \]

Eigenvectors corresponding to a repeated eigenvalue are not unique for a diagonalizable matrix;

\[ \text{repeated eigenvalue } \Rightarrow \text{eigenvectors not unique} \quad (1635) \]

Proof follows from the observation: any linear combination of distinct eigenvectors of diagonalizable $X$, corresponding to a particular eigenvalue, produces another eigenvector. For eigenvalue $\lambda$ whose multiplicity\footnote{A matrix is diagonalizable iff \textit{algebraic multiplicity} (number of occurrences of same eigenvalue) equals \textit{geometric multiplicity} $\dim \mathcal{N}(X-\lambda I) = m - \text{rank}(X-\lambda I)$ [338, p.15] (number of Jordan blocks w.r.t. $\lambda$ or number of corresponding l.i. eigenvectors).} exceeds 1, in other words, any choice of independent vectors from $\mathcal{N}(X-\lambda I)$ (of the same multiplicity) constitutes eigenvectors corresponding to $\lambda$. ♦

Caveat diagonalizability insures linear independence which implies existence of distinct eigenvectors. We may conclude, for diagonalizable matrices,

\[ \text{distinct eigenvalues } \Leftrightarrow \text{eigenvectors unique} \quad (1636) \]
A.5. EIGENVALUE DECOMPOSITION

A.5.0.2 Invertible matrix

When diagonalizable matrix \( X \in \mathbb{R}^{m \times m} \) is nonsingular (no zero eigenvalues), then it has an inverse obtained simply by inverting eigenvalues in (1629):

\[
X^{-1} = S \Lambda^{-1} S^{-1}
\]  

(1637)

A.5.0.3 eigenmatrix

The (right-)eigenvectors \( \{s_i\} \) (1629) are naturally orthogonal \( w_i^T s_j = 0 \) to left-eigenvectors \( \{w_i\} \) except, for \( i = 1, \ldots, m \), \( w_i^T s_i = 1 \); called a biorthogonality condition [377, §2.2.4] [212] because neither set of left or right eigenvectors is necessarily an orthogonal set. Consequently, each dyad from a diagonalization is an independent (§B.1.1) nonorthogonal projector because

\[
s_i w_i^T s_i w_i = s_i w_i^T
\]  

(1638)

whereas the dyads of singular value decomposition are not inherently projectors (confer (1645)).

Dyads of eigenvalue decomposition can be termed eigenmatrices because

\[
X s_i w_i^T = \lambda_i s_i w_i^T
\]  

(1639)

Sum of the eigenmatrices is the Identity;

\[
\sum_{i=1}^{m} s_i w_i^T = I
\]  

(1640)

A.5.1 Symmetric matrix diagonalization

The set of normal matrices is, precisely, that set of all real matrices having a complete orthonormal set of eigenvectors; [425, §8.1] [343, prob.10.2.31] id est, any matrix \( X \) for which \( XX^T = X^T X \); [168, §7.1.3] [338, p.3] e.g., symmetric, orthogonal, and circulant matrices [178]. All normal matrices are diagonalizable.

A symmetric matrix is a special normal matrix whose eigenvalues \( \Lambda \) must be real \( A.13 \) and whose eigenvectors \( S \) can be chosen to make a real orthonormal set; [343, §6.4] [341, p.315] id est, for \( X \in S^m \)

\[
X = SAS^T = \begin{bmatrix} s_1 & \cdots & s_m \end{bmatrix} \Lambda \begin{bmatrix} s_1^T \\ \vdots \\ s_m^T \end{bmatrix} = \sum_{i=1}^{m} \lambda_i s_i s_i^T
\]  

(1641)

where \( \delta^2(\Lambda) = \Lambda \in S^m \) (§A.1) and \( S^{-1} = S^T \in \mathbb{R}^{m \times m} \) (orthogonal matrix, §B.5.2) because of symmetry: \( SAS^{-1} = S^{-1}AS^T \). By 0 eigenvalues theorem A.7.3.0.1,

\[
\begin{align*}
\mathcal{R}\{s_i \mid \lambda_i \neq 0\} &= \mathcal{R}(A) = \mathcal{R}(A^T) \\
\mathcal{R}\{s_i \mid \lambda_i = 0\} &= \mathcal{N}(A^T) = \mathcal{N}(A)
\end{align*}
\]  

(1642)

\[A.13\] Proof. Suppose \( \lambda_i \) is an eigenvalue corresponding to eigenvector \( s_i \) of real \( A = A^T \). Then \( s_i^H A s_i = s_i^T A s_i^* \) (by transposition) \( \Rightarrow s_i^T \lambda_i s_i = s_i^T \lambda_i^* s_i^* \) because \( (A s_i)^* = (\lambda_i s_i)^* \) by assumption. So we have \( \lambda_i ||s_i||^2 = \lambda_i^* ||s_i||^2 \). There is no converse. \( \blacksquare \)
A.5.1.1 Diagonal matrix diagonalization

Because arrangement of eigenvectors and their corresponding eigenvalues is arbitrary, we almost always arrange eigenvalues in nonincreasing order as is the convention for singular value decomposition. Then to diagonalize a symmetric matrix that is already a diagonal matrix, orthogonal matrix $S$ becomes a permutation matrix.

A.5.1.2 Invertible symmetric matrix

When symmetric matrix $X \in S^m$ is nonsingular (invertible), then its inverse (obtained by inverting eigenvalues in (1641)) is also symmetric:

$$X^{-1} = S\Lambda^{-1}S^T \in S^m$$ (1643)

A.5.1.3 Positive semidefinite matrix square root

When $X \in S^m_+$, its unique positive semidefinite matrix square root is defined

$$\sqrt{X} \equiv S\sqrt{\Lambda}S^T \in S^m_+$$ (1644)

where the square root of nonnegative diagonal matrix $\sqrt{\Lambda}$ is taken entrywise and positive. Then $X = \sqrt{X}\sqrt{X}$.

A.6 Singular value decomposition, SVD

A.6.1 Compact SVD

[168, §2.5.4] For any $A \in \mathbb{R}^{m \times n}$

$$A = U\Sigma Q^T = [u_1 \cdots u_\eta] \Sigma \begin{bmatrix} q_1^T \\ \vdots \\ q_\eta^T \end{bmatrix} = \sum_{i=1}^\eta \sigma_i u_i q_i^T$$ (1645)

$$U \in \mathbb{R}^{m \times \eta}, \quad \Sigma \in \mathbb{R}^{\eta \times \eta}, \quad Q \in \mathbb{R}^{n \times \eta}$$

$$U^TU = I, \quad Q^TQ = I$$

where $U$ and $Q$ are always skinny-or-square real, each having orthonormal columns, and where

$$\eta \triangleq \min\{m, n\}$$ (1646)

Square matrix $\Sigma$ is diagonal ($\S A.1.1$)

$$\delta^2(\Sigma) = \Sigma \in \mathbb{R}^{\eta \times \eta}$$ (1647)
holding the singular values \( \{\sigma_i \in \mathbb{R}\} \) of \( A \) which are always arranged in nonincreasing order by convention and are related to eigenvalues \( \lambda \) by

\[
\sigma(A)_i = \sigma(A^T)_i = \begin{cases} 
\sqrt{\lambda(A^T A)_i} = \sqrt{\lambda(A A^T)}_i = \lambda(A^T A)_i > 0, & 1 \leq i \leq \rho \\
0, & \rho < i \leq \eta \end{cases}
\]

of which the last \( \eta - \rho \) are 0 ,\( ^{14} \) where

\[
\rho \triangleq \text{rank } A = \text{rank } \Sigma \quad (1649)
\]

A point sometimes lost: Any real matrix may be decomposed in terms of its real singular values \( \sigma(A) \in \mathbb{R}^\eta \) and real matrices \( U \) and \( Q \) as in (1645), where [168, §2.5.3]

\[
\mathcal{R}\{u_i \mid \sigma_i \neq 0\} = \mathcal{R}(A) \\
\mathcal{R}\{u_i \mid \sigma_i = 0\} \subseteq \mathcal{N}(A^T) \\
\mathcal{R}\{q_i \mid \sigma_i \neq 0\} = \mathcal{R}(A^T) \\
\mathcal{R}\{q_i \mid \sigma_i = 0\} \subseteq \mathcal{N}(A) 
\]

### A.6.2 Subcompact SVD

Some authors allow only nonzero singular values. In that case the compact decomposition can be made smaller; it can be redimensioned in terms of rank \( \rho \) because, for any \( A \in \mathbb{R}^{m \times n} \)

\[
\rho = \text{rank } A = \text{rank } \Sigma = \max \{i \in \{1 \ldots \eta\} \mid \sigma_i \neq 0\} \leq \eta \quad (1651)
\]

- There are \( \eta \) singular values. For any flavor SVD, rank is equivalent to the number of nonzero singular values on the main diagonal of \( \Sigma \).

Now

\[
A = U \Sigma Q^T = \begin{bmatrix} u_1 \cdots u_\rho \end{bmatrix} \Sigma \begin{bmatrix} q^T_1 \\ \vdots \\ q^T_\rho \end{bmatrix} = \sum_{i=1}^{\rho} \sigma_i u_i q^T_i
\]

\[
U \in \mathbb{R}^{m \times \rho}, \quad \Sigma \in \mathbb{R}^{\rho \times \rho}, \quad Q \in \mathbb{R}^{n \times \rho} \\
U^T U = I, \quad Q^T Q = I
\]

where the main diagonal of diagonal matrix \( \Sigma \) has no 0 entries, and

\[
\mathcal{R}\{u_i\} = \mathcal{R}(A) \\
\mathcal{R}\{q_i\} = \mathcal{R}(A^T)
\]

\(^{14}\)When matrix \( A \) is normal, \( \sigma(A) = |\lambda(A)| \). [425, §8.1]

\(^{15}\)For \( \eta = n \), \( \sigma(A) = \sqrt{\lambda(A^T A)} = \lambda \left( \sqrt{A^T A} \right) \).

For \( \eta = m \), \( \sigma(A) = \sqrt{\lambda(A A^T)} = \lambda \left( \sqrt{A A^T} \right) \).
A.6.3 Full SVD

Another common and useful expression of the SVD makes \( U \) and \( Q \) square; making the decomposition larger than compact SVD. Completing the nullspace bases in \( U \) and \( Q \) from (1650) provides what is called the full singular value decomposition of \( A \in \mathbb{R}^{m\times n} \) [341, App.A]. Orthonormal real matrices \( U \) and \( Q \) become orthogonal matrices (§B.5):

\[
\mathcal{R}\{u_i | \sigma_i \neq 0\} = \mathcal{R}(A) \\
\mathcal{R}\{u_i | \sigma_i = 0\} = \mathcal{N}(A^T) \\
\mathcal{R}\{q_i | \sigma_i \neq 0\} = \mathcal{R}(A^T) \\
\mathcal{R}\{q_i | \sigma_i = 0\} = \mathcal{N}(A)
\]  

(1654)

For any matrix \( A \) having rank \( \rho \) (= rank \( \Sigma \))

\[
A = U\Sigma Q^T = [u_1 \cdots u_m] \Sigma \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} = \sum_{i=1}^{n} \sigma_i u_i q_i^T
\]

\[
= \begin{bmatrix} m \times \rho \text{ basis } \mathcal{R}(A) & m \times m-\rho \text{ basis } \mathcal{N}(A^T) \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_\rho \end{bmatrix} \begin{bmatrix} (m \times \rho \text{ basis } \mathcal{R}(A^T))^T \\ (m \times m-\rho \text{ basis } \mathcal{N}(A))^T \end{bmatrix}
\]

\[
U \in \mathbb{R}^{m \times m}, \quad \Sigma \in \mathbb{R}^{m \times n}, \quad Q \in \mathbb{R}^{n \times n} \\
U^T = U^{-1}, \quad Q^T = Q^{-1}
\]  

(1655)

where upper limit of summation \( \eta \) is defined in (1646). Matrix \( \Sigma \) is no longer necessarily square, now padded with respect to (1647) by \( m-\eta \) zero rows or \( n-\eta \) zero columns; the nonincreasingly ordered (possibly 0) singular values appear along its main diagonal as for compact SVD (1648).

An important geometrical interpretation of SVD is given in Figure 167 for \( m = n = 2 \):

The image of the unit sphere under any \( m \times n \) matrix multiplication is an ellipse. Considering the three factors of the SVD separately, note that \( Q^T \) is a pure rotation of the circle. Figure 167 shows how the axes \( q_1 \) and \( q_2 \) are first rotated by \( Q^T \) to coincide with the coordinate axes. Second, the circle is stretched by \( \Sigma \) in the directions of the coordinate axes to form an ellipse. The third step rotates the ellipse by \( U \) into its final position. Note how \( q_1 \) and \( q_2 \) are rotated to end up as \( u_1 \) and \( u_2 \), the principal axes of the final ellipse.

A direct calculation shows that \( Aq_j = \sigma_j u_j \). Thus \( q_j \) is first rotated to coincide with the \( j^\text{th} \) coordinate axis, stretched by a factor \( \sigma_j \), and then rotated to point in the direction of \( u_j \). All of this is beautifully illustrated for \( 2 \times 2 \) matrices by the MATLAB code \texttt{eigshow.m} (see [344]).

A direct consequence of the geometric interpretation is that the largest singular value \( \sigma_1 \) measures the “magnitude” of \( A \) (its 2-norm):

\[
\|A\|_2 = \sup_{\|x\|_2 = 1} \|Ax\|_2 = \sigma_1
\]  

(1656)

This means that \( \|A\|_2 \) is the length of the longest principal semiaxis of the ellipse.
A.6. SINGULAR VALUE DECOMPOSITION, SVD

\[ A = U\Sigma Q^T = [u_1 \cdots u_m] \Sigma \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} = \sum_{i=1}^{n} \sigma_i u_i q_i^T \]

Figure 167: Full SVD geometrical interpretation [281]: Image of circle \( \{x \in \mathbb{R}^2 \mid \|x\|_2 = 1\} \) under matrix multiplication \( Ax \) is, in general, an ellipse. For the example illustrated, 
\( U \triangleq [u_1 \ u_2] \in \mathbb{R}^{2 \times 2}, \ Q \triangleq [q_1 \ q_2] \in \mathbb{R}^{2 \times 2}. \)
Expressions for \( U \), \( Q \), and \( \Sigma \) follow readily from (1655),
\[
AA^T U = U \Sigma \Sigma^T \quad \text{and} \quad A^T A Q = Q \Sigma \Sigma^T \quad \text{(1657)}
\]
demonstrating that the columns of \( U \) are the eigenvectors of \( AA^T \) and the columns of \( Q \) are the eigenvectors of \( A^T A \).

– Muller, Magaia, & Herbst [281]

### A.6.4 Pseudoinverse by SVD

Matrix pseudoinverse (§E) is nearly synonymous with singular value decomposition because of the elegant expression, given \( A = U \Sigma Q^T \in \mathbb{R}^{m \times n} \)
\[
A^\dagger = Q \Sigma^\dagger U^T \in \mathbb{R}^{n \times m} \quad \text{(1658)}
\]

that applies to all three flavors of SVD, where \( \Sigma^\dagger \) simply inverts nonzero entries of matrix \( \Sigma \).

Given symmetric matrix \( A \in \mathbb{S} \) and its diagonalization \( A = S \Lambda S^T \) (§A.5.1), its pseudoinverse simply inverts all nonzero eigenvalues:
\[
A^\dagger = S \Lambda^\dagger S^T \in \mathbb{S} \quad \text{(1659)}
\]

### A.6.5 SVD of symmetric matrices

From (1648) and (1644) for \( A = A^T \)
\[
\sigma(A)_i = \begin{cases} \sqrt{\lambda(A^2)_i} = \lambda(\sqrt{A^2})_i = |\lambda(A)_i| > 0, & 1 \leq i \leq \rho \\ 0, & \rho < i \leq \eta \end{cases} \quad \text{(1660)}
\]

#### A.6.5.0.1 Definition. Step function. (confer §4.3.2.0.1)
Define the signum-like quasilinear function \( \psi : \mathbb{R}^n \to \mathbb{R}^n \) that takes value 1 corresponding to a 0-valued entry in its argument:
\[
\psi(a) \triangleq \left[ \lim_{x_i \to a_i} \frac{x_i}{|x_i|} \right] = \begin{cases} 1, & a_i \geq 0 \\ -1, & a_i < 0 \end{cases}, \quad i = 1 \ldots n \in \mathbb{R}^n \quad \text{(1661)}
\]

Eigenvalue signs of a symmetric matrix having diagonalization \( A = S \Lambda S^T \) (1641) can be absorbed either into real \( U \) or real \( Q \) from the full SVD; [362, p.34] (confer §C.4.2.1)
\[
A = S \Lambda S^T = S \delta(\psi(\delta(\Lambda))) |\Lambda| S^T \triangleq U \Sigma Q^T \in \mathbb{S}^n \quad \text{(1662)}
\]
or
\[
A = S \Lambda S^T = S |\Lambda| \delta(\psi(\delta(\Lambda))) S^T \triangleq U \Sigma Q^T \in \mathbb{S}^n \quad \text{(1663)}
\]

where matrix of singular values \( \Sigma = |\Lambda| \) denotes entrywise absolute value of diagonal eigenvalue matrix \( \Lambda \).
A.7. Zeros

A.7.1 norm zero

For any given norm, by definition,

\[ \|x\|_\ell = 0 \iff x = 0 \quad (1664) \]

Consequently, a generally nonconvex constraint in \( x \) like \( \|Ax - b\| = \kappa \) becomes convex when \( \kappa = 0 \).

A.7.2 0 entry

If a positive semidefinite matrix \( A = [A_{ij}] \in \mathbb{R}^{n \times n} \) has a 0 entry \( A_{ii} \) on its main diagonal, then \( A_{ij} + A_{ji} = 0 \ \forall j \). [282, §1.3.1]

Any symmetric positive semidefinite matrix having a 0 entry on its main diagonal must be 0 along the entire row and column to which that 0 entry belongs. [168, §4.2.8] [212, §7.1 prob.2] From which it follows: for \( A \in \mathbb{S}_+^n \)

\[ \delta(A) = 0 \iff A = 0 \quad (1665) \]

\[ \text{tr}(A) = 0 \iff A = 0 \quad (1666) \]

A.7.3 0 eigenvalues theorem

This theorem is simple, powerful, and widely applicable:

A.7.3.0.1 Theorem. **Number of 0 eigenvalues.**

For any matrix \( A \in \mathbb{R}^{m \times n} \)

\[ \text{rank}(A) + \dim \mathcal{N}(A) = n \quad (1667) \]

by conservation of dimension. [212, §0.4.4]

For any square matrix \( A \in \mathbb{R}^{m \times m} \), number of 0 eigenvalues is at least equal to \( \dim \mathcal{N}(A) \)

\[ \dim \mathcal{N}(A) \leq \text{number of 0 eigenvalues} \leq m \quad (1668) \]

while all eigenvectors corresponding to those 0 eigenvalues belong to \( \mathcal{N}(A) \). [341, §5.1]^{A.16}

\(^{A.16}\)We take as given the well-known fact that the number of 0 eigenvalues cannot be less than dimension of the nullspace. We offer an example of the converse:

\[ A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \]

\( \dim \mathcal{N}(A) = 2 \), \( \lambda(A) = [0 \ 0 \ 0 \ 1]^T \); three eigenvectors in the nullspace but only two are independent. The right-hand side of (1668) is tight for nonzero matrices; e.g., (§B.1) dyad \( uv^T \in \mathbb{R}^{m \times m} \) has \( m \) 0-eigenvalues when \( u \in v^\perp \).
For diagonalizable matrix $A$ (§A.5), the number of 0 eigenvalues is precisely $\dim \mathcal{N}(A)$ while the corresponding eigenvectors span $\mathcal{N}(A)$. Real and imaginary parts of the eigenvectors remaining span $\mathcal{R}(A)$.

(TRANSPOSE.)

Likewise, for any matrix $A \in \mathbb{R}^{m \times n}$

$$\text{rank}(A^T) + \dim \mathcal{N}(A^T) = m \quad (1669)$$

For any square $A \in \mathbb{R}^{m \times m}$, number of 0 eigenvalues is at least equal to $\dim \mathcal{N}(A^T) = \dim \mathcal{N}(A)$ while all left-eigenvectors (eigenvectors of $A^T$) corresponding to those 0 eigenvalues belong to $\mathcal{N}(A^T)$.

For diagonalizable $A$, number of 0 eigenvalues is precisely $\dim \mathcal{N}(A^T)$ while the corresponding left-eigenvectors span $\mathcal{N}(A^T)$. Real and imaginary parts of the left-eigenvectors remaining span $\mathcal{R}(A^T)$.

**Proof.** First we show, for a diagonalizable matrix, the number of 0 eigenvalues is precisely the dimension of its nullspace while the eigenvectors corresponding to those 0 eigenvalues span the nullspace:

Any diagonalizable matrix $A \in \mathbb{R}^{m \times m}$ must possess a complete set of linearly independent eigenvectors. If $A$ is full-rank (invertible), then all $m = \text{rank}(A)$ eigenvalues are nonzero. [341, §5.1]

Suppose $\text{rank}(A) < m$. Then $\dim \mathcal{N}(A) = m - \text{rank}(A)$. Thus there is a set of $m - \text{rank}(A)$ linearly independent vectors spanning $\mathcal{N}(A)$. Each of those can be an eigenvector associated with a 0 eigenvalue because $A$ is diagonalizable $\iff \exists m$ linearly independent eigenvectors. [341, §5.2] Eigenvectors of a real matrix corresponding to 0 eigenvalues must be real.\[A.17] Thus $A$ has at least $m - \text{rank}(A)$ eigenvalues equal to 0.

Now suppose $A$ has more than $m - \text{rank}(A)$ eigenvalues equal to 0. Then there are more than $m - \text{rank}(A)$ linearly independent eigenvectors associated with 0 eigenvalues, and each of those eigenvectors must be in $\mathcal{N}(A)$. Thus there are more than $m - \text{rank}(A)$ linearly independent vectors in $\mathcal{N}(A)$; a contradiction.

Diagonalizable $A$ therefore has $\text{rank}(A)$ nonzero eigenvalues and exactly $m - \text{rank}(A)$ eigenvalues equal to 0 whose corresponding eigenvectors span $\mathcal{N}(A)$.

By similar argument, the left-eigenvectors corresponding to 0 eigenvalues span $\mathcal{N}(A^T)$.

Next we show when $A$ is diagonalizable, the real and imaginary parts of its eigenvectors (corresponding to nonzero eigenvalues) span $\mathcal{R}(A)$:

The (right-)eigenvectors of a diagonalizable matrix $A \in \mathbb{R}^{m \times m}$ are linearly independent if and only if the left-eigenvectors are. So, matrix $A$ has a representation in terms of its right- and left-eigenvectors; from the diagonalization (1629), assuming 0 eigenvalues are ordered last,

$$A = \sum_{i=1}^{m} \lambda_i \ s_i \ w_i^T = \sum_{i=1}^{k} \lambda_i \ s_i \ w_i^T \quad (1670)$$

From the linearly independent dyads theorem (§B.1.1.0.2), the dyads $\{s_i w_i^T\}$ must be independent because each set of eigenvectors are; hence $\text{rank} A = k$, the number of nonzero eigenvalues.

\[A.17\] Proof. Let $^*$ denote complex conjugation. Suppose $A = A^*$ and $As_i = 0$. Then $s_i = s_i^* \Rightarrow As_i = As_i^* \Rightarrow As_i^* = 0$. Conversely, $As_i^* = 0 \Rightarrow As_i = As_i^* \Rightarrow s_i = s_i^*$. ♦
eigenvalues. Complex eigenvectors and eigenvalues are common for real matrices, and must come in complex conjugate pairs for the summation to remain real. Assume that conjugate pairs of eigenvalues appear in sequence. Given any particular conjugate pair from \( (1670) \), we get the partial summation

\[
\lambda_i s_i w_i^T + \lambda_i^* s_i^* w_i^T = 2 \text{re}(\lambda_i s_i w_i^T) = 2(\text{re} s_i \text{re}(\lambda_i w_i^T) - \text{im} s_i \text{im}(\lambda_i w_i^T))
\]  

(1671)

where \( \lambda_i^* \triangleq \lambda_{i+1} \), \( s_i^* \triangleq s_{i+1} \), and \( w_i^* \triangleq w_{i+1} \). Then \( (1670) \) is equivalently written

\[
A = 2 \sum_{\lambda_i \in \mathbb{C}, \lambda_i \neq 0} \text{re} s_{2i} \text{re}(\lambda_{2i} w_{2i}^T) - \text{im} s_{2i} \text{im}(\lambda_{2i} w_{2i}^T) + \sum_{\lambda_j \in \mathbb{R}, \lambda_j \neq 0} \lambda_j s_j w_j^T
\]  

(1672)

The summation \( (1672) \) shows: \( A \) is a linear combination of real and imaginary parts of its (right-)eigenvectors corresponding to nonzero eigenvalues. The \( k \) vectors \( \{ \text{re} s_i \in \mathbb{R}^m, \text{im} s_i \in \mathbb{R}^m \mid \lambda_i \neq 0, i \in \{1 \ldots m\} \} \) must therefore span the range of diagonalizable matrix \( A \).

The argument is similar regarding span of the left-eigenvectors. ♦

A.7.4 0 trace and matrix product

For \( X, A \in \mathbb{R}^{M \times N} \) (39)

\[
\text{tr}(X^T A) = 0 \iff X \circ A = A \circ X = 0
\]  

(1673)

For \( X, A \in \mathbb{S}_{+}^M \) [35, §2.6.1 exer.2.8] [371, §3.1]

\[
\text{tr}(X A) = 0 \iff X A = A X = 0
\]  

(1674)

**Proof.** (\( \iff \)) Suppose \( X A = A X = 0 \). Then \( \text{tr}(X A) = 0 \) is obvious.

(\( \Rightarrow \)) Suppose \( \text{tr}(X A) = 0 \). \( \text{tr}(X A) = \text{tr}(\sqrt{A} X \sqrt{A}) \) whose argument is positive semidefinite by Corollary A.3.1.0.5. Trace of any square matrix is equivalent to the sum of its eigenvalues. Eigenvalues of a positive semidefinite matrix can total 0 if and only if each and every nonnegative eigenvalue is 0. The only positive semidefinite matrix, having all 0 eigenvalues, resides at the origin; \( (\text{confer (1698)}) \) id est,

\[
\sqrt{A} X \sqrt{A} = \left(\sqrt{X \sqrt{A}}\right)^T \sqrt{X \sqrt{A}} = 0
\]  

(1675)

implying \( \sqrt{X \sqrt{A}} = 0 \) which in turn implies \( \sqrt{X}(\sqrt{X \sqrt{A}})\sqrt{A} = X A = 0 \). Arguing similarly yields \( A X = 0 \). ♦

Diagonalizable matrices \( A \) and \( X \) are *simultaneously diagonalizable* if and only if they are commutative under multiplication; [212, §1.3.12] id est, iff they share a complete set of eigenvectors.

\[ \begin{align*}
A.18 & \text{Complex conjugate of } w \text{ is denoted } w^*. \text{ Conjugate transpose is denoted } w^H = w^*^T.
\end{align*} \]
A.7.4.0.1 Example. An equivalence in nonisomorphic spaces.

Identity (1674) leads to an unusual equivalence relating convex geometry to traditional linear algebra: The convex sets, given $A \succeq 0$
\[
\{ X \mid \langle X, A \rangle = 0 \} \cap \{ X \geq 0 \} \equiv \{ X \mid N(X) \supseteq R(A) \} \cap \{ X \geq 0 \} \quad (1676)
\]
(one expressed in terms of a hyperplane, the other in terms of nullspace and range) are equivalent only when symmetric matrix $A$ is positive semidefinite.

We might apply this equivalence to the geometric center subspace, for example, \[ S^M_c = \{ Y \in S^M \mid Y 1 = 0 \} \]
\[ = \{ Y \in S^M \mid \mathcal{N}(Y) \supseteq 1 \} = \{ Y \in S^M \mid \mathcal{R}(Y) \subseteq \mathcal{N}(1^T) \} \quad (2095) \]
from which we derive (confer (1075))
\[ S^M_c \cap S^M_+ \equiv \{ X \geq 0 \mid \langle X, 11^T \rangle = 0 \} \quad (1677) \]
\[ \square \]

A.7.5 Zero definite

The domain over which an arbitrary real matrix $A$ is zero definite can exceed its left and right nullspaces. For any positive semidefinite matrix $A \in \mathbb{R}^{M \times M}$ (for $A + A^T \succeq 0$)
\[
\{ x \mid x^T A x = 0 \} = \mathcal{N}(A + A^T) \quad (1678)
\]
because \( \exists R \ni A + A^T = R^T R \), \( \| R x \| = 0 \iff R x = 0 \), and \( \mathcal{N}(A + A^T) = \mathcal{N}(R) \).

Then given any particular vector $x_p$, \( x_p^T A x_p = 0 \iff x_p \in \mathcal{N}(A + A^T) \). For any positive definite matrix $A$ (for $A + A^T > 0$)
\[
\{ x \mid x^T A x = 0 \} = 0 \quad (1679)
\]
Further, [425, §3.2 prob.5]
\[
\{ x \mid x^T A x = 0 \} = \mathbb{R}^M \iff A^T = -A \quad (1680)
\]
while
\[
\{ x \mid x^H A x = 0 \} = \mathbb{C}^M \iff A = 0 \quad (1681)
\]

The positive semidefinite matrix
\[
A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad (1682)
\]
for example, has no nullspace. Yet
\[
\{ x \mid x^T A x = 0 \} = \{ x \mid 1^T x = 0 \} \subset \mathbb{R}^2 \quad (1683)
\]
which is the nullspace of the symmetrized matrix. Symmetric matrices are not spared from the excess; \textit{videlicet},
\[
B = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad (1684)
\]
has eigenvalues \{-1, 3\}, no nullspace, but is zero definite on

\[ X \triangleq \{ x \in \mathbb{R}^2 \mid x_2 = (−2 \pm \sqrt{3})x_1 \} \]  

(A.19)

A.7.5.0.1 Proposition. (Sturm/Zhang) **Dyad-decompositions.** [347, §5.2]

Let positive semidefinite matrix \( X \in \mathbb{S}^n_+ \) have rank \( \rho \). Then, given symmetric matrix \( A \in \mathbb{S}^n \), \( \langle A, X \rangle = 0 \) if and only if there exists a dyad-decomposition

\[ X = \sum_{j=1}^{\rho} x_j x_j^T \]  

(1686)

satisfying

\[ \langle A, x_j x_j^T \rangle = 0 \text{ for each and every } j \in \{1 \ldots \rho\} \]  

(1687)

⋄

The dyad-decomposition of \( X \) proposed is generally not that obtained from a standard diagonalization by eigenvalue decomposition, unless \( \rho = 1 \) or the given matrix \( A \) is simultaneously diagonalizable (§A.7.4) with \( X \). That means, elemental dyads in decomposition (1686) constitute a generally nonorthogonal set. Sturm & Zhang give a simple procedure for constructing the dyad-decomposition [400] where matrix \( A \) may be regarded as a parameter.

A.7.5.0.2 Example. **Dyad.**

The dyad \( uv^T \in \mathbb{R}^{M \times M} \) (§B.1) is zero definite on all \( x \) for which either \( x^Tu = 0 \) or \( x^Tv = 0 \);

\[ \{ x \mid x^T uv^T x = 0 \} = \{ x \mid x^Tu = 0 \} \cup \{ x \mid v^T x = 0 \} \]  

(1688)

id est, on \( u^\perp \cup v^\perp \). Symmetrizing the dyad does not change the outcome:

\[ \{ x \mid x^T(uv^T + vu^T)x/2 = 0 \} = \{ x \mid x^Tu = 0 \} \cup \{ x \mid v^T x = 0 \} \]  

(1689)

\[ \square \]

---

\( \text{A}^{19} \) These two lines represent the limit in the union of two generally distinct hyperbolae; id est, for matrix \( B \) and set \( X \) as defined

\[ \lim_{\varepsilon \to 0^+} \{ x \in \mathbb{R}^2 \mid x^TBx = \varepsilon \} = X \]
Appendix B

Simple matrices

Mathematicians also attempted to develop algebra of vectors but there was no natural definition of the product of two vectors that held in arbitrary dimensions. The first vector algebra that involved a noncommutative vector product (that is, \( v \times w \) need not equal \( w \times v \)) was proposed by Hermann Grassmann in his book Ausdehnungslehre (1844). Grassmann’s text also introduced the product of a column matrix and a row matrix, which resulted in what is now called a simple or a rank-one matrix. In the late 19th century the American mathematical physicist Willard Gibbs published his famous treatise on vector analysis. In that treatise Gibbs represented general matrices, which he called dyadics, as sums of simple matrices, which Gibbs called dyads. Later the physicist P. A. M. Dirac introduced the term “bra-ket” for what we now call the scalar product of a “bra” (row) vector times a “ket” (column) vector and the term “ket-bra” for the product of a ket times a bra, resulting in what we now call a simple matrix, as above. Our convention of identifying column matrices and vectors was introduced by physicists in the 20th century.

—Suddhendu Biswas [49, p.2]

B.1 Rank-one matrix (dyad)

Any matrix formed from the unsigned outer product of two vectors,

\[ \Psi = uv^T \in \mathbb{R}^{M \times N} \]  \hspace{1cm} (1690)

where \( u \in \mathbb{R}^M \) and \( v \in \mathbb{R}^N \), is rank-one and called a dyad. Conversely, any rank-one matrix must have the form \( \Psi \). [212, prob.1.4.1] Product \(-uv^T\) is a negative dyad. For matrix products \( AB^T\), in general, we have

\[ \mathcal{R}(AB^T) \subseteq \mathcal{R}(A) \quad \text{and} \quad \mathcal{N}(AB^T) \supseteq \mathcal{N}(B^T) \]  \hspace{1cm} (1691)

Dattorro, Convex Optimization † Euclidean Distance Geometry 2e, Meijo, v2015.04.11. 545
APPENDIX B. SIMPLE MATRICES

\[
\begin{align*}
\mathcal{R}(v) & \\
0 & \quad \mathcal{R}(\Psi) = \mathcal{R}(u) \\
\mathcal{N}(\Psi) = \mathcal{N}(v^T) & \\
\mathcal{N}(u^T) & \quad \mathcal{N}(uv^T) = \mathcal{R}(u^T) \oplus \mathcal{R}(uv^T) = \mathbb{R}^M
\end{align*}
\]

\( \mathbb{R}^N = \mathcal{R}(v) \oplus \mathcal{N}(uv^T) \)

Figure 168: The four fundamental subspaces [343, §3.6] for any dyad \( \Psi = uv^T \in \mathbb{R}^{M \times N} \): \( \mathcal{R}(v) \perp \mathcal{N}(\Psi) \perp \mathcal{N}(u^T) \perp \mathcal{R}(\Psi) \). \( \Psi(x) \triangleq uv^T x \) is a linear mapping from \( \mathbb{R}^N \) to \( \mathbb{R}^M \). Map from \( \mathcal{R}(v) \) to \( \mathcal{R}(u) \) is bijective. [341, §3.1]

with equality when \( B = A \) [341, §3.3, §3.6]\(^{B.1} \) or respectively when \( B \) is invertible and \( \mathcal{N}(A) = 0 \). Yet for all nonzero dyads we have

\[
\mathcal{R}(uv^T) = \mathcal{R}(u), \quad \mathcal{N}(uv^T) = \mathcal{N}(v^T) = v^perp
\]

where \( \dim v^perp = N-1 \).

It is obvious a dyad can be \( 0 \) only when \( u \) or \( v \) is \( 0 \):

\[
\Psi = uv^T = 0 \iff u = 0 \text{ or } v = 0
\]

The matrix 2-norm for \( \Psi \) is equivalent to Frobenius’ norm;

\[
\| \Psi \|_2 = \sigma_1 = \| uv^T \|_F = \| uv^T \|_2 = \| u \| \| v \|
\]

When \( u \) and \( v \) are normalized, the pseudoinverse is the transposed dyad. Otherwise,

\[
\Psi^\dagger = (uv^T)^\dagger = \frac{vu^T}{\| u \|^2 \| v \|^2}
\]

When dyad \( uv^T \in \mathbb{R}^{N \times N} \) is square, \( uv^T \) has at least \( N-1 \) 0-eigenvalues and corresponding eigenvectors spanning \( v^perp \). The remaining eigenvector \( u \) spans the range of \( uv^T \) with corresponding eigenvalue

\[
\lambda = v^Tu = \text{tr}(uv^T) \in \mathbb{R}
\]

Determinant is a product of the eigenvalues; so, it is always true that

\[
\det \Psi = \det(uv^T) = 0
\]

\(^{B.1}\)Proof. \( \mathcal{R}(AA^T) \subseteq \mathcal{R}(A) \) is obvious.

\[
\mathcal{R}(AA^T) = \{ AA^T y \mid y \in \mathbb{R}^m \} \\
\supseteq \{ AA^T y \mid A^T y \in \mathcal{R}(A^T) \} = \mathcal{R}(A) \text{ by (142)}
\]
When $\lambda = 1$, the square dyad is a nonorthogonal projector projecting on its range ($\Psi^2 = \Psi$, §E.6): a projector dyad. It is quite possible that $u \in v^\perp$ making the remaining eigenvalue instead 0; $\lambda = 0$ together with the first $N-1$ 0-eigenvalues; id est, it is possible $uv^T$ were nonzero while all its eigenvalues are 0. The matrix
\[
\begin{bmatrix}
  1 & 1 \\
-1 & -1
\end{bmatrix}
\begin{bmatrix}
  1 & 1 \\
-1 & -1
\end{bmatrix}
\]
(1698)
for example, has two 0-eigenvalues. In other words, eigenvector $u$ may simultaneously be a member of the nullspace and range of the dyad. The explanation is, simply, because $u$ and $v$ share the same dimension, $\dim u = M = \dim v = N$:

**Proof.** Figure 168 sees the four fundamental subspaces for the dyad. Linear operator $\Psi : \mathbb{R}^N \rightarrow \mathbb{R}^M$ provides a map between vector spaces that remain distinct when $M = N$;
\[
\begin{align*}
 u &\in \mathcal{R}(uv^T) \\
u &\in \mathcal{N}(uv^T) \iff v^Tu = 0 \\
\mathcal{R}(uv^T) \cap \mathcal{N}(uv^T) &\neq \emptyset
\end{align*}
\] (1699)

**B.1.0.1 rank-one modification**

For $A \in \mathbb{R}^{M \times N}$, $x \in \mathbb{R}^N$, $y \in \mathbb{R}^M$, and $y^TAx \neq 0$ [216, §2.1]
\[
\operatorname{rank}\left(A - \frac{Ax^TA}{y^TAx}\right) = \operatorname{rank}(A) - 1
\] (1700)

Given nonsingular matrix $A \in \mathbb{R}^{N \times N}$, $1 + v^TA^{-1}u \neq 0$, [161, §4.11.2] [231, App.6] [425, §2.3 prob.16] (Sherman-Morrison-Woodbury)
\[
(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}
\] (1701)
which is unstable numerically, by Saunders’ reckoning: “We know how to update matrix factorizations reliably (even if the input or output matrix is singular) but, in general, there’s no stable way to update a matrix inverse.”

**B.1.0.2 dyad symmetry**

In the specific circumstance that $v = u$, then $uv^T \in \mathbb{R}^{N \times N}$ is symmetric, rank-one, and positive semidefinite having exactly $N-1$ 0-eigenvalues. In fact, (Theorem A.3.1.0.7)
\[
\lambda = u^Tu = \operatorname{tr}(uu^T) > 0 \quad \text{unless} \quad u = 0
\] (1703)

A dyad is not always diagonalizable (§A.5) because its eigenvectors are not necessarily independent.
When $\lambda = 1$, the dyad becomes an orthogonal projector. Matrix

$$
\begin{bmatrix}
\Psi & u \\
u^T & 1
\end{bmatrix}
$$

(1704)

for example, is rank-1 positive semidefinite if and only if $\Psi = uu^T$.

### B.1.1 Dyad independence

Now we consider a sum of dyads like (1690) as encountered in diagonalization and singular value decomposition:

$$
\mathcal{R}\left(\sum_{i=1}^{k} s_i w_i^T\right) = \sum_{i=1}^{k} \mathcal{R}(s_i w_i^T) = \sum_{i=1}^{k} \mathcal{R}(s_i) \iff w_i \forall i \text{ are l.i.} \quad (1705)
$$

range of summation is the vector sum of ranges. \(^{B.3}\) (Theorem B.1.1.1.1) Under the assumption the dyads are linearly independent (l.i.), then vector sums are unique (p.662): for $\{w_i\}$ l.i. and $\{s_i\}$ li.

$$
\mathcal{R}\left(\sum_{i=1}^{k} s_i w_i^T\right) = \mathcal{R}(s_1 w_1^T) \oplus \ldots \oplus \mathcal{R}(s_k w_k^T) = \mathcal{R}(s_1) \oplus \ldots \oplus \mathcal{R}(s_k) \quad (1706)
$$

**B.1.1.0.1 Definition.** Linearly independent dyads. \([221, \text{p.29 thm.11}] [349, \text{p.2}]\)

The set of $k$ dyads

$$
\{s_i w_i^T \mid i = 1 \ldots k\}
$$

(1707)

where $s_i \in \mathbb{C}^M$ and $w_i \in \mathbb{C}^N$, is said to be linearly independent iff

$$
\text{rank } \left( S W^T \triangleq \sum_{i=1}^{k} s_i w_i^T \right) = k \quad (1708)
$$

where $S \triangleq [s_1 \ldots s_k] \in \mathbb{C}^{M \times k}$ and $W \triangleq [w_1 \ldots w_k] \in \mathbb{C}^{N \times k}$.

\(\triangle\)

Dyad independence does not preclude existence of a nullspace $\mathcal{N}(SW^T)$, as defined, nor does it imply $SW^T$ were full-rank. In absence of assumption of independence, generally, $\text{rank } SW^T \leq k$. Conversely, any rank-$k$ matrix can be written in the form $SW^T$ by singular value decomposition. (§A.6)

**B.1.1.0.2 Theorem.** Linearly independent (l.i.) dyads.

Vectors $\{s_i \in \mathbb{C}^M, \ i = 1 \ldots k\}$ are l.i. and vectors $\{w_i \in \mathbb{C}^N, \ i = 1 \ldots k\}$ are l.i. if and only if dyads $\{s_i w_i^T \in \mathbb{C}^{M \times N}, \ i = 1 \ldots k\}$ are l.i.

\(\diamond\)

\(^{B.3}\)Move of range $\mathcal{R}$ to inside summation admitted by linear independence of $\{w_i\}$. 

B.1. RANK-ONE MATRIX (DYAD)

Proof. Linear independence of \( k \) dyads is identical to definition (1708).

\((\Rightarrow)\) Suppose \( \{s_i\} \) and \( \{w_i\} \) are each linearly independent sets. Invoking Sylvester’s rank inequality, [212, §0.4] [425, §2.4]

\[
\text{rank } S + \text{rank } W - k \leq \text{rank}(SW^T) \leq \min\{\text{rank } S, \text{rank } W\} \quad (\leq k)
\]

Then \( k \leq \text{rank}(SW^T) \leq k \), which implies the dyads are independent.

\((\Leftarrow)\) Conversely, suppose \( \text{rank}(SW^T) = k \). Then

\[
k \leq \min\{\text{rank } S, \text{rank } W\} \leq k
\]

implying the vector sets are each independent. ♦

B.1.1.1 Biorthogonality condition, Range and Nullspace of Sum

Dyads characterized by biorthogonality condition \( W^TS = I \) are independent; \textit{id est}, for \( S \in \mathbb{C}^{M \times k} \) and \( W \in \mathbb{C}^{N \times k} \), if \( W^TS = I \) then \( \text{rank}(SW^T) = k \) by the \textit{linearly independent dyads theorem} because (confer §E.1.1)

\[
W^TS = I \Rightarrow \text{rank } S = \text{rank } W = k \leq M = N
\]

To see that, we need only show: \( \mathcal{N}(S) = 0 \iff \exists B \ni BS = I \). \(^{B.4}\)

\((\Leftarrow)\) Assume \( BS = I \). Then \( \mathcal{N}(BS) = 0 = \{x | BSx = 0\} \supseteq \mathcal{N}(S) \). (1691)

\((\Rightarrow)\) If \( \mathcal{N}(S) = 0 \) then \( S \) must be full-rank skinny-or-square.

\[
\therefore \exists A, B, C \ni \begin{bmatrix} B & C \end{bmatrix} \begin{bmatrix} S \ A \end{bmatrix} = I \quad (\text{id est, } \begin{bmatrix} S \ A \end{bmatrix} \text{ is invertible}) \Rightarrow BS = I.
\]

Left inverse \( B \) is given as \( W^T \) here. Because of reciprocity with \( S \), it immediately follows: \( \mathcal{N}(W) = 0 \iff \exists S \ni S^TW = I \). ♦

Dyads produced by diagonalization, for example, are independent because of their inherent biorthogonality. (§A.5.0.3) The converse is generally false; \textit{id est}, linearly independent dyads are not necessarily biorthogonal.

B.1.1.1.1 Theorem. \textit{Nullspace and range of dyad sum.}

Given a sum of dyads represented by \( SW^T \) where \( S \in \mathbb{C}^{M \times k} \) and \( W \in \mathbb{C}^{N \times k} \),

\[
\mathcal{N}(SW^T) = \mathcal{N}(W^T) \iff \exists B \ni BS = I \\
\mathcal{R}(SW^T) = \mathcal{R}(S) \iff \exists Z \ni W^TZ = I
\]

Proof. \((\Rightarrow)\) \( \mathcal{N}(SW^T) \supseteq \mathcal{N}(W^T) \) and \( \mathcal{R}(SW^T) \subseteq \mathcal{R}(S) \) are obvious.

\((\Leftarrow)\) Assume the existence of a left inverse \( B \in \mathbb{R}^{k \times N} \) and a right inverse \( Z \in \mathbb{R}^{N \times k} \). \(^{B.5}\)

\[
\mathcal{N}(SW^T) = \{x | SW^Tx = 0\} \supseteq \{x | BSW^Tx = 0\} = \mathcal{N}(W^T) \\
\mathcal{R}(SW^T) = \{SW^Tx | x \in \mathbb{R}^{N}\} \supseteq \{SW^TZY | ZY \in \mathbb{R}^{N}\} = \mathcal{R}(S)
\]

\(^{B.4}\)Left inverse is not unique, in general.

\(^{B.5}\)By counterexample, the theorem’s converse cannot be true; \textit{e.g.}, \( S = W = \begin{bmatrix} 1 & 0 \end{bmatrix} \).
APPENDIX B. SIMPLE MATRICES

B.2 Doublet

Consider a sum of two linearly independent square dyads, one a transposition of the other:

\[ \Pi = uv^T + vu^T = \begin{bmatrix} u & v \\ v^T & u^T \end{bmatrix} = SW^T \in \mathbb{S}^N \]  (1715)

where \( u, v \in \mathbb{R}^N \). Like the dyad, a doublet can be 0 only when \( u \) or \( v \) is 0; 

\[ \Pi = uv^T + vu^T = 0 \iff u = 0 \text{ or } v = 0 \]  (1716)

By assumption of independence, a nonzero doublet has two nonzero eigenvalues 

\[ \lambda_1 \triangleq u^Tv + \|uv^T\| , \quad \lambda_2 \triangleq u^Tv - \|uv^T\| \]  (1717)

where \( \lambda_1 > 0 > \lambda_2 \), with corresponding eigenvectors 

\[ x_1 \triangleq \frac{u}{\|u\|} + \frac{v}{\|v\|} , \quad x_2 \triangleq \frac{u}{\|u\|} - \frac{v}{\|v\|} \]  (1718)

spanning the doublet range. Eigenvalue \( \lambda_1 \) cannot be 0 unless \( u \) and \( v \) have opposing directions, but that is antithetical since then the dyads would no longer be independent. Eigenvalue \( \lambda_2 \) is 0 if and only if \( u \) and \( v \) share the same direction, again antithetical. Generally we have \( \lambda_1 > 0 \) and \( \lambda_2 < 0 \), so \( \Pi \) is indefinite.

By the nullspace and range of dyad sum theorem, doublet \( \Pi \) has \( N - 2 \) zero-eigenvalues remaining and corresponding eigenvectors spanning \( \mathcal{N}\left( \begin{bmatrix} v^T \\ u^T \end{bmatrix} \right) \). We therefore have

\[ \mathcal{R}(\Pi) = \mathcal{R}(\begin{bmatrix} u & v \end{bmatrix}) , \quad \mathcal{N}(\Pi) = v^\perp \cap u^\perp \]  (1719)

of respective dimension 2 and \( N - 2 \).

B.3 Elementary matrix

A matrix of the form

\[ E = I - \zeta uv^T \in \mathbb{R}^{N \times N} \]  (1720)
B.3. ELEMENTARY MATRIX

\[ N(u^T) \quad 0 \quad 0 \quad 0 \quad R(E) = N(v^T) \]
\[ N(E) = R(u) \quad R(v) \quad \mathbb{R}^N = N(u^T) \oplus N(E) \quad R(v) \oplus R(E) = \mathbb{R}^N \]

Figure 170: \( v^T u = 1/\zeta \) . The four fundamental subspaces [343, §3.6] for elementary matrix \( E \) as a linear mapping \( E(x) = \left( I - \frac{uv^T}{v^T u} \right) x \).

where \( \zeta \in \mathbb{R} \) is finite and \( u, v \in \mathbb{R}^N \), is called an elementary matrix or rank-one modification of the Identity. [214] Any elementary matrix in \( \mathbb{R}^{N \times N} \) has \( N - 1 \) eigenvalues equal to 1 corresponding to real eigenvectors that span \( v^\perp \). The remaining eigenvalue

\[ \lambda = 1 - \zeta v^T u \quad (1721) \]

corresponds to eigenvector \( u \). \(^{\text{B.6}}\) From [231, App.7.A.26] the determinant:

\[ \det E = 1 - \text{tr} \left( \zeta uv^T \right) = \lambda \quad (1722) \]

If \( \lambda \neq 0 \) then \( E \) is invertible; [161] (confer §B.1.0.1)

\[ E^{-1} = I + \frac{\zeta}{\lambda} uv^T \quad (1723) \]

Eigenvectors corresponding to 0 eigenvalues belong to \( N(E) \), and the number of 0 eigenvalues must be at least \( \dim N(E) \) which, here, can be at most one. (§A.7.3.0.1) The nullspace exists, therefore, when \( \lambda = 0 \); id est, when \( v^T u = 1/\zeta \); rather, whenever \( u \) belongs to hyperplane \( \{ z \in \mathbb{R}^N \mid v^T z = 1/\zeta \} \). Then (when \( \lambda = 0 \)) elementary matrix \( E \) is a nonorthogonal projector projecting on its range \( (E^2 = E, \ \S E.1) \) and \( N(E) = R(u) \); eigenvector \( u \) spans the nullspace when it exists. By conservation of dimension, \( \dim R(E) = N - \dim N(E) \). It is apparent from (1720) that \( v^\perp \subseteq R(E) \), but \( \dim v^\perp = N - 1 \). Hence \( R(E) = v^\perp \) when the nullspace exists, and the remaining eigenvectors span it.

In summary, when a nontrivial nullspace of \( E \) exists,

\[ R(E) = N(v^T), \quad N(E) = R(u), \quad v^T u = 1/\zeta \quad (1724) \]

illustrated in Figure 170, which is opposite to the assignment of subspaces for a dyad (Figure 168). Otherwise, \( R(E) = \mathbb{R}^N \).

When \( E = E^T \), the spectral norm is

\[ \| E \|_2 = \max \{ 1, |\lambda| \} \quad (1725) \]

\(^{\text{B.6}}\) Elementary matrix \( E \) is not always diagonalizable because eigenvector \( u \) need not be independent of the others; id est, \( u \in v^\perp \) is possible.
B.3.1 Householder matrix

An elementary matrix is called a Householder matrix when it has the defining form, for nonzero vector $u$ \[ H = I - 2 \frac{uu^T}{u^Tu} \in \mathbb{S}^N \] (1726)

which is a symmetric orthogonal (reflection) matrix ($H^{-1} = H^T = H$ (§B.5.3)). Vector $u$ is normal to an $N-1$-dimensional subspace $u^\perp$ through which this particular $H$ effects pointwise reflection; e.g., $Hu^\perp = u^\perp$ while $Hu = -u$.

Matrix $H$ has $N-1$ orthonormal eigenvectors spanning that reflecting subspace $u^\perp$ with corresponding eigenvalues equal to 1. The remaining eigenvector $u$ has corresponding eigenvalue $-1$; so

\[ \det H = -1 \] (1727)

Due to symmetry of $H$, the matrix 2-norm (the spectral norm) is equal to the largest eigenvalue-magnitude. A Householder matrix is thus characterized,

\[ H^T = H, \quad H^{-1} = H^T, \quad \|H\|_2 = 1, \quad H \geq 0 \] (1728)

For example, the permutation matrix

\[ \Xi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \] (1729)

is a Householder matrix having $u = [0 \ 1 \ -1] / \sqrt{2}$. Not all permutation matrices are Householder matrices, although all permutation matrices are orthogonal matrices (§B.5.2, $\Xi^T \Xi = I$) [341, §3.4] because they are made by permuting rows and columns of the Identity matrix. Neither are all symmetric permutation matrices Householder matrices; e.g., \[ \Xi = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \] (1820) is not a Householder matrix.

B.4 Auxiliary $V$-matrices

B.4.1 Auxiliary projector matrix $V$

It is convenient to define a matrix $V$ that arises naturally as a consequence of translating the geometric center $\alpha_e$ (§5.1.0.1) of some list $X$ to the origin. In place of $X - \alpha_e 1^T$ we may write $XV$ as in (1055) where

\[ V = I - \frac{1}{N} 11^T \in \mathbb{S}^N \] (991)

is an elementary matrix called the geometric centering matrix.
B.4. AUXILIARY V-MATRICES

Any elementary matrix in $\mathbb{R}^{N \times N}$ has $N - 1$ eigenvalues equal to 1. For the particular elementary matrix $V$, the $N^{th}$ eigenvalue equals 0. The number of 0 eigenvalues must equal $\dim \mathcal{N}(V) = 1$, by the 0 eigenvalues theorem (§A.7.3.0.1), because $V = V^T$ is diagonalizable. Because

$$V \mathbf{1} = \mathbf{0}$$

(1730)

the nullspace $\mathcal{N}(V) = \mathbb{R}(\mathbf{1})$ is spanned by the eigenvector $\mathbf{1}$. The remaining eigenvectors span $\mathbb{R}(V) \equiv \mathbb{R}(\mathbf{1}^T)$ that has dimension $N - 1$.

Because

$$V^2 = V$$

(1731)

and $V^T = V$, elementary matrix $V$ is also a projection matrix (§E.3) projecting orthogonally on its range $\mathcal{N}(\mathbf{1}^T)$ which is a hyperplane containing the origin in $\mathbb{R}^N$

$$V = I - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$$

(1732)

The $\{0,1\}$ eigenvalues also indicate diagonalizable $V$ is a projection matrix. [425, §4.1 thm.4.1] Symmetry of $V$ denotes orthogonal projection; from (2010),

$$V^T = V, \quad V^\dagger = V, \quad \|V\|_2 = 1, \quad V \succeq 0$$

(1733)

Matrix $V$ is also circulant [178].

B.4.1.0.1 Example. Relationship of Auxiliary to Householder matrix.

Let $H \in \mathbb{S}^N$ be a Householder matrix (1726) defined by

$$u = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^N$$

(1734)

Then we have [163, §2]

$$V = H \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0} \end{bmatrix} H$$

(1735)

Let $D \in \mathbb{S}^N_h$ and define

$$-HDH \triangleq -\begin{bmatrix} A & b \\ b^T & c \end{bmatrix}$$

(1736)

where $b$ is a vector. Then because $H$ is nonsingular (§A.3.1.0.5) [194, §3]

$$-VDV = -H \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0} \end{bmatrix} H \succeq 0 \iff -A \succeq 0$$

(1737)

and affine dimension is $r = \text{rank} A$ when $D$ is a Euclidean distance matrix. \hfill \Box
**B.4.2 Schoenberg auxiliary matrix $V_N$**

1. $V_N = \frac{1}{\sqrt{2}} \begin{bmatrix} -1^T \\ I \end{bmatrix} \in \mathbb{R}^{N \times N-1}$

2. $V_N^T \mathbf{1} = 0$

3. $I - e_1^T = \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}$

4. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} V_N = V_N$

5. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} V = V$

6. $V \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}$

7. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}$

8. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}^\dagger = \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} V$

9. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}^\dagger V = \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}^\dagger$

10. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}^\dagger = V$

11. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}^\dagger \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} = \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix}$

12. $\begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix}$

13. $\begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 & \sqrt{2}V_N \end{bmatrix} = \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix}$

14. $[V_N \sqrt{2} \mathbf{1}]^{-1} = \begin{bmatrix} V_N^\dagger \\ \sqrt{2} \mathbf{1}^T \end{bmatrix}$

15. $V_N^\dagger = \sqrt{2} \left[ -\frac{1}{N} \mathbf{1} \quad I - \frac{1}{N} \mathbf{1}^T \right] \in \mathbb{R}^{N-1 \times N}, \quad \left( I - \frac{1}{N} \mathbf{1}^T \in \mathbb{S}^{N-1} \right)$

16. $V_N^\dagger \mathbf{1} = 0$

17. $V_N^\dagger V_N = I$

18. $V^T = V = V_N V_N^\dagger = I - \frac{1}{N} \mathbf{1}^T \in \mathbb{S}^N$

19. $-V_N^\dagger (\mathbf{1}^T - I) V_N = I, \quad \left( \mathbf{1}^T - I \in \text{EDM}^N \right)$
B.4. AUXILIARY V-MATRICES

20. \(D = [d_{ij}] \in S_h^N\) \hfill (993)

\[\operatorname{tr}(-VDV) = \operatorname{tr}(-VD) = \operatorname{tr}(-V_N^T D V_N) = \frac{1}{N} 1^T D 1 = \frac{1}{N} \operatorname{tr}(11^T D) = \frac{1}{N} \sum_{i,j} d_{ij}\]

Any elementary matrix \(E \in S^N\) of the particular form

\[
E = k_1 I - k_2 11^T
\]

(1738)

where \(k_1, k_2 \in \mathbb{R}\), \(^7\) will make \(\operatorname{tr}(-ED)\) proportional to \(\sum d_{ij}\).

21. \(D = [d_{ij}] \in S^N\)

\[\operatorname{tr}(-VDV) = \frac{1}{N} \sum_{i,j} d_{ij} - \frac{N - 1}{N} \sum_i d_{ii} = \frac{1}{N} 1^T D 1 - \operatorname{tr} D\]

22. \(D = [d_{ij}] \in S_h^N\)

\[\operatorname{tr}(-V_N^T D V_N) = \sum_j d_{ij}\]

23. For \(Y \in S^N\)

\[V (Y - \delta(Y1)) V = Y - \delta(Y1)\]

B.4.3 Orthonormal auxiliary matrix \(V_W\)

The skinny matrix

\[V_W \equiv \begin{bmatrix}
\frac{-1}{N+\sqrt{N}} & \frac{-1}{N} & \ldots & \frac{-1}{N}\\
1 + \frac{-1}{N+\sqrt{N}} & \frac{-1}{N+\sqrt{N}} & \ldots & \frac{-1}{N+\sqrt{N}} \\
\frac{-1}{N+\sqrt{N}} & \ldots & \ldots & \frac{-1}{N+\sqrt{N}} \\
\vdots & \ddots & \ddots & \vdots \\
\frac{-1}{N+\sqrt{N}} & \frac{-1}{N+\sqrt{N}} & \ldots & 1 + \frac{-1}{N+\sqrt{N}}
\end{bmatrix} \in \mathbb{R}^{N \times N-1}\]

(1739)

has \(R(V_W) = N(1^T)\) and orthonormal columns. \(^7\) We defined three auxiliary \(V\)-matrices: \(V\), \(V_N\) \hfill (975), and \(V_W\) sharing some attributes listed in Table B.4.4. For example, \(V\) can be expressed

\[
V = V_W V_W^T = V_N V_N^T
\]

(1740)

but \(V_W V_W = I\) means \(V\) is an orthogonal projector \hfill (2007)\) and

\[
V_W^T = V_W^T, \quad \|V_W\|_2 = 1, \quad V_W^T 1 = 0
\]

(1741)

\(^7\) If \(k_1 = 1 - \rho\) while \(k_2 = -\rho \in \mathbb{R}\), then all eigenvalues of \(E\) for \(-1/(N-1) < \rho < 1\) are guaranteed positive and therefore \(E\) is guaranteed positive definite. \hfill [312]
### APPENDIX B. SIMPLE MATRICES

#### B.4.4 Auxiliary \( V \)-matrix Table

<table>
<thead>
<tr>
<th>( V )</th>
<th>dim ( V )</th>
<th>rank ( V )</th>
<th>( \mathcal{R}(V) )</th>
<th>( \mathcal{N}(V^T) )</th>
<th>( V^T )</th>
<th>( VV^T )</th>
<th>( VV^+ )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V )</td>
<td>( N \times N )</td>
<td>( N-1 )</td>
<td>( \mathcal{N}(1^T) )</td>
<td>( \mathcal{R}(1) )</td>
<td>( V )</td>
<td>( V )</td>
<td>( V )</td>
</tr>
<tr>
<td>( V_{NC} )</td>
<td>( N \times (N-1) )</td>
<td>( N-1 )</td>
<td>( \mathcal{N}(1^T) )</td>
<td>( \mathcal{R}(1) )</td>
<td>( \frac{1}{2}(I+11^T) )</td>
<td>( \frac{1}{2} \left[ \begin{array}{cc} N-1 &amp; -1^T \ -1 &amp; I \end{array} \right] )</td>
<td>( V )</td>
</tr>
<tr>
<td>( V_{W} )</td>
<td>( N \times (N-1) )</td>
<td>( N-1 )</td>
<td>( \mathcal{N}(1^T) )</td>
<td>( \mathcal{R}(1) )</td>
<td>( I )</td>
<td>( V )</td>
<td>( V )</td>
</tr>
</tbody>
</table>

#### B.4.5 More auxiliary matrices

Mathar shows [269, §2] that any elementary matrix (§B.3) of the form

\[
V_S = I - b1^T \in \mathbb{R}^{N \times N}
\]  

(1742)

such that \( b^T1 = 1 \) (confer [172, §2]), is an auxiliary \( V \)-matrix having

\[
\begin{align*}
\mathcal{R}(V_S^T) &= \mathcal{N}(b^T), & \mathcal{R}(V_S) &= \mathcal{N}(1^T) \\
\mathcal{N}(V_S) &= \mathcal{R}(b), & \mathcal{N}(V_S^T) &= \mathcal{R}(1)
\end{align*}
\]

(1743)

Given \( X \in \mathbb{R}^{n \times N} \), the choice \( b = \frac{1}{N}1 \) (\( V_S = V \)) minimizes \( \|X(I - b1^T)\|_F \). [174, §3.2.1]

#### B.5 Orthomatrices

##### B.5.1 Orthonormal matrix

Property \( Q^TQ = I \) completely defines orthonormal matrix \( Q \in \mathbb{R}^{n \times k} \) \((k \leq n)\); a skinny-or-square full-rank matrix characterized by nonexpansivity (2011)

\[
\|Qx\|_2 \leq \|x\|_2 \quad \forall x \in \mathbb{R}^n, \quad \|Qy\|_2 = \|y\|_2 \quad \forall y \in \mathbb{R}^k
\]

(1744)

and preservation of vector inner-product

\[
\langle Qy, Qz \rangle = \langle y, z \rangle
\]

(1745)

##### B.5.2 Orthogonal matrix \& vector rotation

An orthogonal matrix is a square orthonormal matrix. Property \( Q^{-1} = Q^T \) completely defines orthogonal matrix \( Q \in \mathbb{R}^{n \times n} \) employed to effect vector rotation; [341, §2.6, §3.4] [343, §6.5] [212, §2.1] for any \( x \in \mathbb{R}^n \)

\[
\|Qx\|_2 = \|x\|_2
\]

(1746)

In other words, the 2-norm is orthogonally invariant. Any antisymmetric matrix constructs an orthogonal matrix; \( id est \), for \( A = -A^T \)

\[
Q = (I + A)^{-1}(I - A)
\]

(1747)
A *unitary matrix* is a complex generalization of orthogonal matrix; conjugate transpose defines it: $U^{-1} = U^H$. An orthogonal matrix is simply a real unitary matrix.\(^{B.8}\)

Orthogonal matrix $Q$ is a normal matrix further characterized by spectral norm:

$$Q^{-1} = QT, \quad \|Q\|_2 = 1$$

(1748)

Applying characterization (1748) to $Q^T$ we see it too is an orthogonal matrix. Hence the rows and columns of $Q$ respectively form an orthonormal set. Normalcy guarantees diagonalization (§A.5.1) so, for $Q \triangleq SAS^H$

$$SA^{-1}S^H = S^*AS^T = S^*A^*S^H, \quad \|\delta(\Lambda)\|_\infty = 1$$

(1749)

characterizes an orthogonal matrix in terms of eigenvalues and eigenvectors.

All permutation matrices $\Xi$, for example, are nonnegative orthogonal matrices; and *vice versa*. Product or Kronecker product of any permutation matrices remains a permutator. Any product of permutation matrix with orthogonal matrix remains orthogonal. In fact, any product of orthogonal matrices $AQ$ remains orthogonal by definition. Given any other dimensionally compatible orthogonal matrix $U$, the mapping $g(A) = U^TAQ$ is a bijection on the domain of orthogonal matrices (a nonconvex manifold of dimension $\frac{1}{2}n(n-1)$ [54]). [249, §2.1] [250]

The Householder matrix (§B.3.1) is an example of symmetric orthogonal (reflection) matrix.

B.5.3 Reflection

A matrix for pointwise reflection is defined by imposing symmetry upon the orthogonal matrix; *id est*, a reflection matrix is completely defined by $Q^{-1} = QT = Q$. The reflection matrix is a symmetric orthogonal matrix, and *vice versa*, characterized:

$$Q^T = Q, \quad Q^{-1} = QT, \quad \|Q\|_2 = 1$$

(1752)

The Householder matrix (§B.3.1) is an example of symmetric orthogonal (reflection) matrix.

Reflection matrices have eigenvalues equal to ±1 and so $\det Q = \pm1$. It is natural to expect a relationship between reflection and projection matrices because all projection matrices have eigenvalues belonging to $\{0,1\}$. In fact, any reflection matrix $Q$ is related to some orthogonal projector $P$ by [214, §1 prob.44]

$$Q = I - 2P$$

(1753)

\(^{B.8}\)Orthogonal and unitary matrices are called *unitary linear operators.*
Yet $P$ is, generally, neither orthogonal or invertible. (§E.3.2)

$$\lambda(Q) \in \mathbb{R}^n, \quad |\lambda(Q)| = 1$$  \hspace{1cm} (1754)

Reflection is with respect to $R(P)^\perp$. Matrix $2P - I$ represents antireflection.

Every orthogonal matrix can be expressed as the product of a rotation and a reflection. The collection of all orthogonal matrices of particular dimension does not form a convex set.

### B.5.4 Rotation of range and rowspace

Given orthogonal matrix $Q$, column vectors of a matrix $X$ are simultaneously rotated about the origin via product $QX$. In three dimensions ($X \in \mathbb{R}^{3 \times N}$), the precise meaning of rotation is best illustrated in Figure 171 where the gimbal aids visualization of what is achievable; mathematically, (§5.5.2.0.1)

$$Q = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$  \hspace{1cm} (1755)

#### B.5.4.0.1 Example. One axis of revolution.

Partition $n + 1$-dimensional Euclidean space $\mathbb{R}^{n+1} \triangleq \begin{bmatrix} \mathbb{R}^n \\ \mathbb{R} \end{bmatrix}$ and define an $n$-dimensional subspace

$$\mathcal{R} \triangleq \{ \lambda \in \mathbb{R}^{n+1} \mid \mathbf{1}^T \lambda = 0 \}$$  \hspace{1cm} (1756)

(a hyperplane through the origin). We want an orthogonal matrix that rotates a list in the columns of matrix $X \in \mathbb{R}^{n+1 \times N}$ through the dihedral angle between $\mathbb{R}^n$ and $\mathcal{R}$ (§2.4.3)

$$\angle(\mathbb{R}^n, \mathcal{R}) = \arccos \left( \frac{\langle e_{n+1}, \mathbf{1} \rangle}{\| e_{n+1} \| \| \mathbf{1} \|} \right) = \arccos \left( \frac{1}{\sqrt{n+1}} \right) \text{ radians}$$  \hspace{1cm} (1757)
The vertex-description of the nonnegative orthant in $\mathbb{R}^{n+1}$ is
\[
\{[e_1 \ e_2 \ \cdots \ e_{n+1}]a \mid a \succeq 0\} = \{a \succeq 0\} = \mathbb{R}^{n+1}_+ \subset \mathbb{R}^{n+1}
\]
(1758)

Consider rotation of these vertices via orthogonal matrix
\[
Q \triangleq 1 - \frac{1}{\sqrt{n+1}} \Xi V \Xi^T \in \mathbb{R}^{n+1 \times n+1}
\]
(1759)
where permutation matrix $\Xi \in S^{n+1}_+$ is defined in (1820), and $V \in \mathbb{R}^{n+1 \times n}$ is the orthonormal auxiliary matrix defined in B.4.3. This particular orthogonal matrix is selected because it rotates any point in subspace $\mathbb{R}^n$ about one axis of revolution onto $\mathcal{R}$; e.g., rotation $Qe_{n+1}$ aligns the last standard basis vector with subspace normal $\mathcal{R}^\perp = 1$.

The rotated standard basis vectors remaining are orthonormal spanning $\mathbb{R}$.

Another interpretation of product $QX$ is rotation/reflection of $\mathcal{R}(X)$. Rotation of $X$ as in $QXQ^T$ is a simultaneous rotation/reflection of range and rowspace.\textsuperscript{B.9}

\textbf{Proof.} Any matrix can be expressed as a singular value decomposition $X = U\Sigma W^T$ (1645) where $\delta^2(\Sigma) = \Sigma$, $\mathcal{R}(U) \supseteq \mathcal{R}(X)$, and $\mathcal{R}(W) \supseteq \mathcal{R}(X^T)$.

\section*{B.5.5 Matrix rotation}

Orthogonal matrices are also employed to rotate/reflect other matrices like vectors: [168, §12.4.1] Given orthogonal matrix $Q$, the product $Q^TA$ will rotate $A \in \mathbb{R}^{n \times n}$ in the Euclidean sense in $\mathbb{R}^n$ because Frobenius’ norm is orthogonally invariant (§2.2.1):
\[
\|Q^T A\|_F = \sqrt{\text{tr}(A^T Q Q^T A)} = \|A\|_F
\]
(likewise for $AQ$). Were $A$ symmetric, such a rotation would depart from $S^n$. One remedy is to instead form product $Q^T A Q$ because
\[
\|Q^T A Q\|_F = \sqrt{\text{tr}(Q^T A^T Q Q^T A Q)} = \|A\|_F
\]
(1761)
By §A.1.1 no.31,
\[
\text{vec} Q^T A Q = (Q \otimes Q)^T \text{vec} A
\]
(1762)
which is a rotation of the vectorized $A$ matrix because Kronecker product of any orthogonal matrices remains orthogonal; e.g, by §A.1.1 no.38,
\[
(Q \otimes Q)^T (Q \otimes Q) = I
\]
(1763)

Matrix $A$ is \textit{orthogonally equivalent} to $B$ if $B = S^T A S$ for some orthogonal matrix $S$. Every square matrix, for example, is orthogonally equivalent to a matrix having equal entries along the main diagonal. [212, §2.2, prob.3]

The product $Q^T A Q$ can be regarded as a coordinate transformation; e.g., given linear map $y = Ax : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and orthogonal $Q$, the transformation $Qy = AQx$ is a rotation/reflection of range and rowspace (141) of matrix $A$ where $Qy \in \mathcal{R}(A)$ and $Qx \in \mathcal{R}(A^T)$ (142).
Appendix C

Some analytical optimal results

People have been working on Optimization since the ancient Greeks [Zenodorus, circa 200BC] learned that a string encloses the most area when it is formed into the shape of a circle.

—Roman Polyak

We speculate that optimization problems possessing analytical solution have convex transformation or constructive global optimality conditions, perhaps yet unknown; e.g., §7.1.4, (1792), §C.3.0.1.

C.1 Properties of infima

- \[
\inf \emptyset \triangleq -\infty \\
\sup \emptyset \triangleq \infty
\]  

- Given \( f(x) : \mathcal{X} \rightarrow \mathbb{R} \) defined on arbitrary set \( \mathcal{X} \) [209, §0.1.2]

\[
\begin{align*}
\inf_{x \in \mathcal{X}} f(x) &= -\sup_{x \in \mathcal{X}} -f(x) \\
\sup_{x \in \mathcal{X}} f(x) &= -\inf_{x \in \mathcal{X}} -f(x)
\end{align*}
\]  

\[
\begin{align*}
\arg\inf_{x \in \mathcal{X}} f(x) &= \arg\sup_{x \in \mathcal{X}} -f(x) \\
\arg\sup_{x \in \mathcal{X}} f(x) &= \arg\inf_{x \in \mathcal{X}} -f(x)
\end{align*}
\]  

- Given scalar \( \kappa \) and \( f(x) : \mathcal{X} \rightarrow \mathbb{R} \) and \( g(x) : \mathcal{X} \rightarrow \mathbb{R} \) defined on arbitrary set \( \mathcal{X} \) [209, §0.1.2]

\[
\begin{align*}
\inf_{x \in \mathcal{X}} (\kappa + f(x)) &= \kappa + \inf_{x \in \mathcal{X}} f(x) \\
\arg\inf_{x \in \mathcal{X}} (\kappa + f(x)) &= \arg\inf_{x \in \mathcal{X}} f(x)
\end{align*}
\]  

Dattorro, Convex Optimization † Euclidean Distance Geometry 2e, Meijoo, v2015.04.11. 561
\[
\inf_{x \in X} \kappa f(x) = \kappa \inf_{x \in X} f(x) \quad \text{arg inf}_{x \in X} \kappa f(x) = \arg \inf_{x \in X} f(x) \quad \kappa > 0 \quad (1768)
\]

\[
\inf_{x \in X} (f(x) + g(x)) \geq \inf_{x \in X} f(x) + \inf_{x \in X} g(x) \quad (1769)
\]

\begin{itemize}
  \item Given \( f(x) : X \to \mathbb{R} \) defined on arbitrary set \( X \)
    \[\arg \inf_{x \in X} |f(x)| = \arg \inf_{x \in X} f(x)^2 \quad (1770)\]
  \item Given \( f(x) : X \cup Y \to \mathbb{R} \) and arbitrary sets \( X \) and \( Y \) [209, §0.1.2]
    \[\inf_{x \in X \cup Y} f(x) = \min \{ \inf_{x \in X} f(x), \inf_{x \in Y} f(x) \} \quad (1772)\]
    \[\inf_{x \in X \cap Y} f(x) = \max \{ \inf_{x \in X} f(x), \inf_{x \in Y} f(x) \} \quad (1773)\]
\end{itemize}

\section*{C.2 Trace, singular and eigen values}

\begin{itemize}
  \item For \( A \in \mathbb{R}^{m \times n} \) and \( \sigma(A) \) denoting its singular values, the nuclear (Ky Fan) norm \( \| A \|^* \) of matrix \( A \) (confer (44), (1648), [213, p.200]) is
    \[\sum \sigma(A) = \text{tr} \sqrt{A^T A} = \| A \|^* = \sup_{\| X \|_2 \leq 1} \text{tr}(X^T A) = \maximize_{X \in \mathbb{R}^{m \times n}} \text{tr}(X^T A) \quad (1774)\]
    \[
    \begin{bmatrix}
      I & X \\
      X^T & I
    \end{bmatrix} \succeq 0 \\
    X \in \mathbb{S}^m, \ Y \in \mathbb{S}^n
    \]
    \[= \frac{1}{2} \minimize_{X \in \mathbb{S}^m, \ Y \in \mathbb{S}^n} \text{tr} X + \text{tr} Y \quad \text{subject to} \quad \begin{bmatrix}
      X & A \\
      A^T & Y
    \end{bmatrix} \succeq 0
    \]

\end{itemize}

This nuclear norm is convex\( ^{C.1} \) and dual to the spectral norm. [213, p.214] [63, §A.1.6] Given singular value decomposition \( A = S \Sigma Q^T \in \mathbb{R}^{m \times n} \) (A.6), then \( X^* = SQ^T \in \mathbb{R}^{m \times n} \) is an optimal solution to maximization (confer §2.3.2.0.5) while \( X^* = S \Sigma S^T \in \mathbb{S}^m \) and \( Y^* = Q \Sigma Q^T \in \mathbb{S}^n \) is an optimal solution to minimization [142]. Srebro [333] asserts

\[\sum \sigma(A) = \frac{1}{2} \minimize_{U, V} \| U \|_F^2 + \| V \|_F^2 \quad \text{subject to} \quad A = UV^T \quad (1775)\]

\[\minimize_{U, V} \| U \|_F \| V \|_F \quad \text{subject to} \quad A = UV^T \]

\(^{C.1}\) discernible as envelope of the rank function (1444) or as supremum of functions linear in \( A \) (Figure 76).
C.2. TRACE, SINGULAR AND EIGEN VALUES

- For $A \in \mathbb{R}^{m \times n}$ and $\sigma(A)_1$ connoting spectral norm,

$$\sigma(A)_1 = \sqrt{\lambda(A^T A)_1} = \|A\|_2 = \sup_{\|x\|_2 = 1} \|Ax\|_2 = \min_{t \in \mathbb{R}} \text{subject to } \begin{bmatrix} t I & A \\ A^T & t I \end{bmatrix} \succeq 0$$

(591)

By confining dyad $uv^T$ to the unit nuclear norm ball (94),

$$\sigma(A)_1 = \|A\|_2 = \sup_{\|u\|_1 = 1, \|v\|_1 = 1} \maximize_{Z \in \mathbb{S}_+^m, X \in \mathbb{S}_+^n, Y \in \mathbb{S}_+^n} \text{tr}(Z^T A)$$

subject to $\maximize_{X \in \mathbb{S}_+^m, Z \in \mathbb{S}_+^n} \text{tr} \left[ X A A^T Y \right] \succeq 0$

(1776)

with corresponding left and right singular vectors (optimal) $u^*$ and $v^*$. Applying (1783) to a result of Lanczos [167, p.207],

$$\sigma(A)_1 = \|A\|_2 = \sup_{\|u\|_1 = 1, \|v\|_1 = 1} \maximize_{Z \in \mathbb{S}_+^m, X \in \mathbb{S}_+^n, Y \in \mathbb{S}_+^n} \text{tr}(Z^T A)$$

subject to $\maximize_{X \in \mathbb{S}_+^m, Z \in \mathbb{S}_+^n} \text{tr} \left[ X A A^T Y \right] \succeq 0$

(1777)

whose corresponding left and right singular vectors are $\sqrt{2} u^*$ and $\sqrt{2} v^*$.

C.2.0.0.1 Exercise. Optimal matrix factorization.

Prove (1775), C.2

- For $X \in \mathbb{S}_+^m, Y \in \mathbb{S}_+^n, A \in \mathbb{C} \subseteq \mathbb{R}^{m \times n}$ for set $\mathcal{C}$ convex, and $\sigma(A)$ denoting the singular values of $A$ [142, §3],

$$\minimize_{A} \sum_i \sigma(A)_i = \frac{1}{2} \minimize_{A, X, Y} \text{tr} X + \text{tr} Y$$

subject to $A \in \mathcal{C}$

(1778)

For feasible set $\mathcal{C}$ equal to the unit nuclear norm ball (94),

$$\text{find } A \text{ subject to } A \in \{ Z \in \mathbb{R}^{m \times n} \mid \sum_i \sigma(Z)_i \leq 1 \} = \text{find } A \text{ subject to } \text{tr} X + \text{tr} Y \leq 2$$

(1779)

$C.2$ Hint: Write $A = \Sigma Q^T \in \mathbb{R}^{m \times n}$ and

$$\begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} = \begin{bmatrix} U & V \end{bmatrix} \begin{bmatrix} U^T & V^T \end{bmatrix} \succeq 0$$

Show $U^* = S \sqrt{\Sigma} \in \mathbb{R}^{m \times \min(m, n)}$ and $V^* = Q \sqrt{\Sigma} \in \mathbb{R}^{n \times \min(m, n)}$, hence $\|U^*\|_F^2 = \|V^*\|_F^2$. 


APPENDIX C. SOME ANALYTICAL OPTIMAL RESULTS

For $A \in S^+_N$ and $\beta \in \mathbb{R}$

$$\beta \operatorname{tr} A = \maximize_{X \in S^N} \operatorname{tr}(XA)$$
subject to $X \preceq \beta I$ \hfill (1780)

But the following statement is numerically stable, preventing an unbounded solution in direction of a 0 eigenvalue:

$$\maximize_{X \in S^N} \operatorname{sgn}(\beta) \operatorname{tr}(XA)$$
subject to $X \preceq |\beta| I$ \hfill (1781)

where $\beta \operatorname{tr} A = \operatorname{tr}(X^*A)$. If $\beta \geq 0$, then $(X \preceq -|\beta| I) \iff (X \succeq 0)$.

For symmetric $A \in S^N$, its smallest and largest eigenvalue in $\lambda(A) \in \mathbb{R}^N$ are respectively \((12, \S 4.1)[44, \S 1.6.15][212, \S 4.2][249, \S 2.1][250]\)

$$\min_i \{\lambda(A)_i\} = \inf_{\|x\| = 1} x^T A x = \minimize_{X \in S^N_+} \operatorname{tr}(XA) = \maximize_{t \in \mathbb{R}} t$$
subject to $X = 1$ subject to $A \succeq t I$ \hfill (1782)

$$\max_i \{\lambda(A)_i\} = \sup_{\|x\| = 1} x^T A x = \maximize_{X \in S^N_+} \operatorname{tr}(XA) = \minimize_{t \in \mathbb{R}} t$$
subject to $X = 1$ subject to $A \preceq t I$ \hfill (1783)

whereas

$$\lambda_N I \preceq A \preceq \lambda_1 I \hfill (1784)$$

The largest eigenvalue $\lambda_1$ is always convex in $A \in S^N$ because, given particular $x$, $x^T A x$ is linear in matrix $A$; supremum of a family of linear functions is convex, as illustrated in Figure 76. So for $A, B \in S^N$, $\lambda_1(A + B) \leq \lambda_1(A) + \lambda_1(B)$. \hfill (1586)

Similarly, the smallest eigenvalue $\lambda_N$ of any symmetric matrix is a concave function of its entries; $\lambda_N(A + B) \geq \lambda_N(A) + \lambda_N(B)$. \hfill (1586)

For $v_1$ a normalized eigenvector of $A$ corresponding to the largest eigenvalue, and $v_N$ a normalized eigenvector corresponding to the smallest eigenvalue,

$$v_N = \arg \inf_{\|x\| = 1} x^T A x \hfill (1785)$$
$$v_1 = \arg \sup_{\|x\| = 1} x^T A x \hfill (1786)$$

For $A \in S^N$ having eigenvalues $\lambda(A) \in \mathbb{R}^N$, consider the unconstrained nonconvex optimization that is a projection on the rank-1 subset \((2.9.2.1, \S 3.6.0.0.1)\) of the boundary of positive semidefinite cone $S^N_+$: Defining $\lambda_1 \triangleq \max_i \{\lambda(A)_i\}$ and corresponding eigenvector $v_1$.

$$\minimize_x \|xx^T - A\|_F^2 = \minimize_x \operatorname{tr}(xx^T(xx^T) - 2Ax^T + A^TA)$$
$$= \left\{ \begin{array}{ll}
\|\lambda(A)\|^2, & \lambda_1 \leq 0 \\
\|\lambda(A)\|^2 - \lambda_1^2, & \lambda_1 > 0
\end{array} \right. \hfill (1787)$$
C.2. TRACE, SINGULAR AND EIGEN VALUES

\[
\arg \min_x \|xx^T - A\|_F^2 = \begin{cases} 
0, & \lambda_1 \leq 0 \\
v_1 \sqrt{\lambda_1}, & \lambda_1 > 0
\end{cases} \quad (1788)
\]

**Proof.** This is simply the Eckart & Young solution from §7.1.2:

\[
x^*x^{*T} = \begin{cases} 
0, & \lambda_1 \leq 0 \\
\lambda_1 v_1 v_1^T, & \lambda_1 > 0
\end{cases} \quad (1789)
\]

Given nonincreasingly ordered diagonalization \( A = Q\Lambda Q^T \) where \( \Lambda = \delta(\lambda(A)) \) (§A.5), then (1787) has minimum value

\[
\min_x \|xx^T - A\|_F^2 = \begin{cases} 
\|Q\Lambda Q^T\|_F^2 = \|\delta(\Lambda)\|^2, & \lambda_1 \leq 0 \\
\|Q \left( \begin{array}{ccc}
\lambda_1 & 0 & \\
0 & \ddots & \\
0 & & 0
\end{array} \right) - \Lambda \right) Q^T \|_F^2 = \left\| \begin{array}{c}
\lambda_1 \\
0 \\
\ddots \\
0
\end{array} \right\|_F^2, & \lambda_1 > 0
\end{cases} \quad (1790)
\]

C.2.0.0.2 **Exercise.** *Rank-1 approximation.*

Given symmetric matrix \( A \in \mathbb{S}^N \), prove:

\[
v_1 = \arg \min_{x} \|xx^T - A\|_F^2 \quad \text{subject to} \quad \|x\| = 1 \quad (1791)
\]

where \( v_1 \) is a normalized eigenvector of \( A \) corresponding to its largest eigenvalue.

What is the objective's optimal value?

• (Ky Fan, 1949) For eigenvalues \( \lambda(B) \in \mathbb{R}^N \) of \( B \in \mathbb{S}^N \) arranged in nonincreasing order, and for \( 1 \leq k \leq N \) [12, §4.1] [225] [212, §4.3.18] [371, §2] [249, §2.1] [250]

\[
\sum_{i=N-k+1}^N \lambda(B)_i = \inf_{U \in \mathbb{R}^{N \times k}} \text{tr}(UU^T B) = \min_{X \in \mathbb{S}^N_+} \text{tr}(XB) \quad \text{subject to} \quad X \preceq I \quad (a) \\
\text{tr} X = k \\
= \max_{\mu \in \mathbb{R}, Z \in \mathbb{S}^N_+} (k-N) \mu + \text{tr}(B - Z) \quad (b)
\]

\[
\sum_{i=1}^k \lambda(B)_i = \sup_{U \in \mathbb{R}^{N \times k}} \text{tr}(UU^T B) = \max_{X \in \mathbb{S}^N_+} \text{tr}(XB) \quad \text{subject to} \quad X \preceq I \quad (c) \\
\text{tr} X = k \\
= \min_{\mu \in \mathbb{R}, Z \in \mathbb{S}^N_+} k \mu + \text{tr} Z \quad (d)
\]

\[
\text{subject to} \quad \mu I + Z \succeq B
\]
Given ordered diagonalization $B = QΛQ^T$, (§A.5.1) then an optimal $U$ for the infimum is $U^* = Q(:, N-k+1:N) ∈ ℝ^{N×k}$ whereas $U^* = Q(:, 1:k) ∈ ℝ^{N×k}$ for the supremum is more reliably computed. In both cases, $X^* = U^*U^T$. Optimization (a) searches the convex hull of outer product $UU^T$ of all $N×k$ orthonormal matrices. (§2.3.2.0.1)

- For $B ∈ S^N$ whose eigenvalues $λ(B) ∈ ℝ^N$ are arranged in nonincreasing order, and for diagonal matrix $Υ ∈ S^k$ whose diagonal entries are arranged in nonincreasing order where $1 ≤ k ≤ N$, we utilize the main-diagonal $δ$ operator’s selfadjointness property (1500): [13, §4.2]

$$\sum_{i=1}^{k} Υ_{ii} λ(B)_{N-i+1} = \inf_{U ∈ ℝ^{N×k}} \text{tr}(ΥU^TBU) = \inf_{U ∈ ℝ^{N×k}} \delta(Υ)^Tδ(U^TBU)$$

(1793)

$$= \minimize_{V_i ∈ S^N} \text{tr} \left( B \sum_{i=1}^{k} (Υ_{ii} − Υ_{i+1,i+1})V_i \right)$$

subject to

$$\begin{align*}
\text{tr} V_i &= i, & i = 1 \ldots k \\
I &⪰ V_i & i = 1 \ldots k
\end{align*}$$

where $Υ_{k+1,k+1} ≜ 0$. We speculate,

$$\sum_{i=1}^{k} Υ_{ii} λ(B)_i = \sup_{U ∈ ℝ^{N×k}} \text{tr}(ΥU^TBU) = \sup_{U ∈ ℝ^{N×k}} \delta(Υ)^Tδ(U^TBU)$$

(1794)

Alizadeh shows: [12, §4.2]

$$\sum_{i=1}^{k} Υ_{ii} λ(B)_i = \minimize_{µ ∈ ℝ^k, Z_i ∈ S^N} \text{subject to} \sum_{i=1}^{k} iµ_i + \text{tr} Z_i$$

$$\begin{align*}
µ_i + Z_i − (Υ_{ii} − Υ_{i+1,i+1})B &⪰ 0, & i = 1 \ldots k \\
Z_i &⪰ 0, & i = 1 \ldots k
\end{align*}$$

$$= \maximize_{V_i ∈ S^N} \text{subject to} \text{tr} V_i = i, & i = 1 \ldots k$$

$$\begin{align*}
I &⪰ V_i & i = 1 \ldots k
\end{align*}$$

(1795)

where $Υ_{k+1,k+1} ≜ 0$.

- The largest eigenvalue magnitude $µ$ of $A ∈ S^N$

$$\max_i \{ |λ(A)_i| \} = \minimize_{µ ∈ ℝ} µ$$

subject to

$$-µI ≤ A ≤ µI$$

(1796)
is minimized over convex set $C$ by semidefinite program: (confer §7.1.5)

$$\begin{align*}
\text{minimize} & \quad \|A\|_2 \\
\text{subject to} & \quad A \in C
\end{align*}$$

by semidefinite program: (confer §7.1.5)

$$\begin{align*}
\text{minimize} & \quad \mu, A \\
\text{subject to} & \quad -\mu I \preceq A \preceq \mu I
\end{align*}$$

id est,

$$\mu^* \triangleq \max_i \{|\lambda(A^*)|, \ i = 1 \ldots N\} \in \mathbb{R}_+$$

- For $B \in \mathbb{S}^N$ whose eigenvalues $\lambda(B) \in \mathbb{R}^N$ are arranged in nonincreasing order, let $\Pi \lambda(B)$ be a permutation of eigenvalues $\lambda(B)$ such that their absolute value becomes arranged in nonincreasing order: $|\Pi \lambda(B)|_i \geq |\Pi \lambda(B)|_2 \geq \cdots \geq |\Pi \lambda(B)|_N$. Then, for $1 \leq k \leq N$ [12, §4.3]

$$\sum_{i=1}^k |\Pi \lambda(B)|_i = \minimize_{\mu \in \mathbb{R}, Z \in \mathbb{S}^N_+} k\mu + \text{tr} Z \quad \text{subject to} \quad \mu I + Z + B \succeq 0 \quad \text{subject to} \quad I \succeq V, W \quad \text{tr}(V + W) = k$$

For diagonal matrix $\Upsilon \in \mathbb{S}^k$ whose diagonal entries are arranged in nonincreasing order where $1 \leq k \leq N$

$$\sum_{i=1}^k \Upsilon_{ii} |\Pi \lambda(B)|_i = \minimize_{\mu \in \mathbb{R}^k, Z_i \in \mathbb{S}^N} \sum_{i=1}^k i\mu_i + \text{tr} Z_i \quad \text{subject to} \quad \mu_i I + Z_i + (\Upsilon_{ii} - \Upsilon_{i+1,i+1})B \succeq 0, \quad i = 1 \ldots k \\
\mu_i I + Z_i - (\Upsilon_{ii} - \Upsilon_{i+1,i+1})B \succeq 0, \quad i = 1 \ldots k \\
Z_i \succeq 0, \quad i = 1 \ldots k$$

$$= \maximize_{V_i, W_i \in \mathbb{S}^k} \text{tr} \left( B \sum_{i=1}^k (\Upsilon_{ii} - \Upsilon_{i+1,i+1})(V_i - W_i) \right) \quad \text{subject to} \quad \text{tr}(V_i + W_i) = i, \quad i = 1 \ldots k \\
I \succeq V_i \succeq 0, \quad i = 1 \ldots k \\
I \succeq W_i \succeq 0, \quad i = 1 \ldots k$$

where $\Upsilon_{k+1,k+1} \triangleq 0$.

**C.2.0.0.3 Exercise.** *Weighted sum of largest eigenvalues.*

Prove (1794).

---

*[C.3]* We eliminate a redundant positive semidefinite variable from Alizadeh’s minimization. There exist typographical errors in [303, (6.49) (6.55)] for this minimization.
For \( A, B \in \mathbb{S}^{N} \) whose eigenvalues \( \lambda(A), \lambda(B) \in \mathbb{R}^{N} \) are respectively arranged in nonincreasing order, and for nonincreasingly ordered diagonalizations \( A = W_{A} \Upsilon_{A} W_{A}^{T} \) and \( B = W_{B} A W_{B}^{T} \) \([210][249, \S 2.1][250]\)
\[
\lambda(A)^{T} \lambda(B) = \sup_{U} \text{tr}(A^{T}U^{T}BU) \geq \text{tr}(A^{T}B) \quad (1819)
\]

\((\text{confer } (1824))\) where optimal \( U \) is
\[
U^{*} = W_{B} W_{A}^{T} \in \mathbb{R}^{N \times N} \quad (1816)
\]

We can push that upper bound higher using a result in \(\S C.4.2.1:\)
\[
|\lambda(A)|^{T} |\lambda(B)| = \sup_{U \in \mathbb{C}^{N \times N}} \text{re tr}(A^{T}U^{H}BU) \quad (1801)
\]

For step function \( \psi \) as defined in \((1661)\), optimal \( U \) becomes
\[
U^{*} = W_{B} \sqrt{\delta(\psi(\delta(\Lambda)))}^{H} \sqrt{\delta(\psi(\delta(\Upsilon)))} W_{A}^{T} \in \mathbb{C}^{N \times N} \quad (1802)
\]

### C.3 Orthogonal Procrustes problem

Given matrices \( A, B \in \mathbb{R}^{n \times N} \), their product having full singular value decomposition \(\S A.6.3\)
\[
AB^{T} \triangleq U \Sigma Q^{T} \in \mathbb{R}^{n \times n} \quad (1803)
\]
then an optimal solution \( R^{*} \) to the orthogonal Procrustes problem
\[
\begin{align*}
\text{minimize} & \quad \|A - R^{T}B\|_{F} \\
\text{subject to} & \quad R^{T} = R^{-1}
\end{align*} \quad (1804)
\]

maximizes \( \text{tr}(A^{T}R^{T}B) \) over the nonconvex manifold of orthogonal matrices: \([212, \S 7.4.8]\)
\[
R^{*} = QU^{T} \in \mathbb{R}^{n \times n} \quad (1805)
\]

A necessary and sufficient condition for optimality
\[
AB^{T}R^{*} \succeq 0 \quad (1806)
\]
holds whenever \( R^{*} \) is an orthogonal matrix. \([174, \S 4]\)

Optimal solution \( R^{*} \) can reveal rotation/reflection \((\S 5.5.2, \S B.5)\) of one list in the columns of matrix \( A \) with respect to another list in \( B \). Solution is unique if \( \text{rank } BV_{N} = n \). \([119, \S 2.4.1]\) In the case that \( A \) is a vector and permutation of \( B \), solution \( R^{*} \) is not necessarily a permutation matrix \((\S 4.6.0.0.3)\) although the optimal objective will be zero. More generally, the optimal value for objective of minimization is
\[
\begin{align*}
\text{tr}(A^{T}A + B^{T}B - 2AB^{T}R^{*}) &= \text{tr}(A^{T}A) + \text{tr}(B^{T}B) - 2 \text{tr}(U \Sigma U^{T}) \\
&= \|A\|_{F}^{2} + \|B\|_{F}^{2} - 2 \delta(\Sigma)^{T} 1
\end{align*} \quad (1807)
\]
while the optimal value for corresponding trace maximization is
\[
\sup_{R^T = R^{-1}} \text{tr}(A^TR^TB) = \text{tr}(A^TR^TB) = \delta(\Sigma)^T I \geq \text{tr}(A^TB) \tag{1808}
\]

The same optimal solution \( R^\star \) solves
\[
\maximize_{R} \|A + R^TB\|_F \quad \text{subject to} \quad R^T = R^{-1} \tag{1809}
\]

### C.3.0.1 Procrustes relaxation

By replacing its feasible set with (Example 2.3.2.0.5) the convex hull of orthogonal matrices, we relax Procrustes problem (1804) to a convex problem

\[
\minimize_{R} \|A - R^TB\|_F^2 = \text{tr}(A^TA + B^TB) - 2 \maximize_{R} \text{tr}(A^TR^TB) \]
\[
\text{subject to} \quad R^T = R^{-1} \quad \text{subject to} \quad \begin{bmatrix} I & R \\ R^T & I \end{bmatrix} \succeq 0 \tag{1810}
\]

whose adjusted objective must always equal Procrustes'.

### C.3.1 Effect of translation

Consider the impact on problem (1804) of DC offset in known lists \( A, B \in \mathbb{R}^{n \times N} \). Rotation of \( B \) there is with respect to the origin, so better results may be obtained if offset is first accounted. Because the geometric centers of the lists \( AV \) and \( BV \) are the origin, instead we solve

\[
\minimize_{R} \|AV - R^TBV\|_F \quad \text{subject to} \quad R^T = R^{-1} \tag{1811}
\]

where \( V \in \mathbb{S}^N \) is the geometric centering matrix (§B.4.1). Now we define the full singular value decomposition

\[
AVB^T \triangleq U\Sigma Q^T \in \mathbb{R}^{n \times n} \tag{1812}
\]

and an optimal rotation matrix

\[
R^\star = QU^T \in \mathbb{R}^{n \times n} \tag{1805}
\]

The desired result is an optimally rotated offset list

\[
R^\star BV + A(I - V) \approx A \tag{1813}
\]

which most closely matches the list in \( A \). Equality is attained when the lists are precisely related by a rotation/reflection and an offset. When \( R^\star B = A \) or \( B1 = A1 = 0 \), this result (1813) reduces to \( R^\star B \approx A \).

\[\text{C.4} \quad \text{(because orthogonal matrices are the extreme points of this hull) and whose optimal numerical solution (SDPT3 [360]) [177] is reliably observed to be orthogonal for } n \leq N.\]
C.3.1.1 Translation of extended list

Suppose an optimal rotation matrix $R^* \in \mathbb{R}^{n \times n}$ were derived as before from matrix $B \in \mathbb{R}^{n \times N}$, but $B$ is part of a larger list in the columns of $[C \ B] \in \mathbb{R}^{n \times M+N}$ where $C \in \mathbb{R}^{n \times M}$. In that event, we wish to apply the rotation/reflection and translation to the larger list. The expression supplanting the approximation in (1813) makes $1^T$ of compatible dimension:

$$R^*\left[ C - B \mathbf{1} \mathbf{1}^T_N \right] B V + A \mathbf{1} \mathbf{1}^T_N$$

id est, $C - B \mathbf{1} \mathbf{1}^T_N \in \mathbb{R}^{n \times M}$ and $A \mathbf{1} \mathbf{1}^T_N \in \mathbb{R}^{n \times M+N}$.

C.4 Two-sided orthogonal Procrustes

C.4.0.1 Minimization

Given symmetric $A, B \in \mathbb{S}^N$, each having diagonalization (§A.5.1)

$$A \triangleq Q_A \Lambda_A Q_A^T, \quad B \triangleq Q_B \Lambda_B Q_B^T$$

where eigenvalues are arranged in their respective diagonal matrix $\Lambda$ in nonincreasing order, then an optimal solution [137]

$$R^* = Q_B Q_A^T \in \mathbb{R}^{N \times N}$$

(1816) to the two-sided orthogonal Procrustes problem

$$\text{minimize} \quad \| A - R^T B R \|_F + \text{subject to} \quad R^T = R^{-1}$$

(1817)

maximizes $\text{tr}(A^T R^T B R)$ over the nonconvex manifold of orthogonal matrices. Optimal product $R^* R^*$ has the eigenvectors of $A$ but the eigenvalues of $B$. [174, §7.5.1] The optimal value for the objective of minimization is, by (48)

$$\|Q_A \Lambda_A Q_A^T - R^T Q_B \Lambda_B Q_B^T R^*\|_F = \|Q_A (\Lambda_A - \Lambda_B) Q_A^T\|_F = \|\Lambda_A - \Lambda_B\|_F$$

(1818)

while the corresponding trace maximization has optimal value

$$\sup_{R^T = R^{-1}} \text{tr}(A^T R^T B R) = \text{tr}(A^T R^* R^* B^*) = \text{tr}(\Lambda_A \Lambda_B) \geq \text{tr}(A^T B)$$

(1819)

The lower bound on inner product of eigenvalues is due to Fan (p.520).

C.4.0.2 Maximization

Any permutation matrix is an orthogonal matrix. Defining a row- and column-swapping permutation matrix (a reflection matrix, §B.5.3)

$$\Xi = \Xi^T = \begin{bmatrix} 0 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 0 \end{bmatrix}$$

(1820)
then an optimal solution $R^*$ to the maximization problem [sic]
\[
\max_R \| A - R^T B R \|_F \\
\text{subject to } R^T = R^{-1}
\]
minimizes $\text{tr}(A^T R^T B R)$: [210] [249, §2.1] [250]
\[
R^* = Q_B \Xi Q_A^T \in \mathbb{R}^{N \times N}
\]
The optimal value for the objective of maximization is
\[
\| Q_A \Lambda_A Q_A^T - R^* \Lambda_B Q_B^T \|_F = \| Q_A \Lambda_A Q_A^T - Q_A \Xi^T \Lambda_B \Xi Q_A^T \|_F
\]
while the corresponding trace minimization has optimal value
\[
\inf_{R^T = R^{-1}} \text{tr}(A^T R^T B R) = \text{tr}(A^T R^* T B R^*) = \text{tr}(\Lambda_A \Xi \Lambda_B \Xi)
\]

C.4.1 Procrustes’ relation to linear programming

Although these two-sided Procrustes problems are nonconvex, there is a connection with linear programming [13, §3] [249, §2.1] [250]: Given $A, B \in \mathbb{S}^N$, this semidefinite program in $S$ and $T$
\[
\min_{S, T \in \mathbb{S}^N} \text{tr}(A^T R^T B R) = \max_{S, T \in \mathbb{S}^N} \text{tr}(S + T) \\
\text{subject to } R^T = R^{-1} \quad \text{subject to } A^T \otimes B - I \otimes S - T \otimes I \succeq 0
\]
where $\otimes$ signifies Kronecker product (§D.1.2.1) has optimal objective value (1824).

These two problems in (1825) are strong duals (§2.13.1.0.3). Given ordered diagonalizations (1815), make the observation:
\[
\inf_{R^T = R^{-1}} \text{tr}(A^T R^T B R) = \inf_{\hat{R}} \text{tr}(\hat{A} \hat{R}^T \hat{B} \hat{R})
\]

because $\hat{R} = Q_B R Q_A$ on the set of orthogonal matrices (which includes the permutation matrices) is a bijection. This means, basically, diagonal matrices of eigenvalues $\Lambda_A$ and $\Lambda_B$ may be substituted for $A$ and $B$, so only the main diagonals of $S$ and $T$ come into play;
\[
\max_{S, T \in \mathbb{S}^N} 1^T \delta(S + T) \\
\text{subject to } \delta(\Lambda_A \otimes (\Xi \Lambda_B \Xi) - I \otimes S - T \otimes I) \succeq 0
\]
a linear program in $\delta(S)$ and $\delta(T)$ having the same optimal objective value as the semidefinite program (1825).

We relate their results to Procrustes problem (1817) by manipulating signs (1765) and permuting eigenvalues:
\[
\max_R \text{tr}(A^T R^T B R) = \min_{S, T \in \mathbb{S}^N} 1^T \delta(S + T) \\
\text{subject to } R^T = R^{-1} \quad \text{subject to } \delta(I \otimes S + T \otimes I - \Lambda_A \otimes \Lambda_B) \succeq 0
\]
\[
= \min_{S, T \in \mathbb{S}^N} \text{tr}(S + T) \\
\text{subject to } I \otimes S + T \otimes I - A^T \otimes B \succeq 0
\]
This formulation has optimal objective value identical to that in (1819).

C.4.2 Two-sided orthogonal Procrustes via SVD

By making left- and right-side orthogonal matrices independent, we can push the upper bound on trace (1819) a little further: Given real matrices $A, B$ each having full singular value decomposition ($\S A.6.3$)

$$A \triangleq U_A \Sigma_A Q_A^T \in \mathbb{R}^{m \times n}, \quad B \triangleq U_B \Sigma_B Q_B^T \in \mathbb{R}^{m \times n}$$  \hspace{1cm} (1829)

then a well-known optimal solution $R^*, S^*$ to the problem

$$\begin{align*}
\text{minimize} & \quad \| A - SBR \|_F \\
\text{subject to} & \quad R^H = R^{-1} \\
& \quad S^H = S^{-1}
\end{align*}$$  \hspace{1cm} (1830)

maximizes $\text{re tr}(A^T SBR)$: \cite{325, 299, 54, 204} optimal orthogonal matrices

$$S^* = U_A^U_B^H H \in \mathbb{R}^{m \times m}, \quad R^* = Q_B Q_A^H A \in \mathbb{R}^{n \times n}$$  \hspace{1cm} (1831)

[sic] are not necessarily unique \cite{212, §7.4.13} because the feasible set is not convex. The optimal value for the objective of minimization is, by (48)

$$\| U_A \Sigma_A Q_A^H - S^* U_B \Sigma_B Q_B^H R^* \|_F = \| U_A (\Sigma_A - \Sigma_B) \|_F = \| \Sigma_A - \Sigma_B \|_F$$  \hspace{1cm} (1832)

while the corresponding trace maximization has optimal value \cite{44, §III.6.12}

$$\sup_{R^H = R^{-1}, S^H = S^{-1}} \text{tr}(A^T SBR) = \text{re tr}(A^T S^* B R^*) = \text{tr}(\Sigma_A^T \Sigma_B) \geq \text{tr}(A^T B)$$  \hspace{1cm} (1833)

for which it is necessary

$$A^T S^* B R^* \succeq 0, \quad B R^* A^T S^* \succeq 0$$  \hspace{1cm} (1834)

The lower bound on inner product of singular values in (1833) is due to von Neumann. Equality is attained if $U_A^H U_B = I$ and $Q_B^H Q_A = I$.

C.4.2.1 Symmetric matrices

Now optimizing over the complex manifold of unitary matrices ($\S B.5.2$), the upper bound on trace (1819) is thereby raised: Suppose we are given diagonalizations for (real) symmetric $A, B$ ($\S A.5$)

$$A = W_A \Upsilon_A W_A^T \in \mathbb{S}^n, \quad \delta(\Upsilon) \in \mathcal{K}_M$$  \hspace{1cm} (1835)

$$B = W_B \Lambda_B W_B^T \in \mathbb{S}^n, \quad \delta(\Lambda) \in \mathcal{K}_M$$  \hspace{1cm} (1836)

having their respective eigenvalues in diagonal matrices $\Upsilon, \Lambda \in \mathbb{S}^n$ arranged in nonincreasing order (membership to the monotone cone $\mathcal{K}_M$ (435)). Then by splitting eigenvalue signs, we invent a symmetric SVD-like decomposition

$$A \triangleq U_A \Sigma_A Q_A^H \in \mathbb{S}^n, \quad B \triangleq U_B \Sigma_B Q_B^H \in \mathbb{S}^n$$  \hspace{1cm} (1837)
C.4. TWO-SIDED ORTHOGONAL PROCRUSTES

where $U_A, U_B, Q_A, Q_B \in \mathbb{C}^{n \times n}$ are unitary matrices defined by (confer §A.6.5)

$$U_A \triangleq W_A \sqrt{\delta(\psi(\delta(\Upsilon)))}, \quad Q_A \triangleq W_A \sqrt{\delta(\psi(\delta(\Upsilon)))}^H, \quad \Sigma_A = |\Upsilon| \quad (1838)$$

$$U_B \triangleq W_B \sqrt{\delta(\psi(\delta(\Lambda)))}, \quad Q_B \triangleq W_B \sqrt{\delta(\psi(\delta(\Lambda)))}^H, \quad \Sigma_B = |\Lambda| \quad (1839)$$

where step function $\psi$ is defined in (1661). In this circumstance,

$$S^* = U_A U_B^H = R^* T \in \mathbb{C}^{n \times n} \quad (1840)$$

optimal matrices (1831) now unitary are related by transposition. The optimal value of objective (1832) is

$$\|U_A \Sigma_A Q_A^H - S^* U_B \Sigma_B Q_B^H R^*\|_F = \| |\Upsilon| - |\Lambda| \|_F \quad (1841)$$

while the corresponding optimal value of trace maximization (1833) is

$$\underset{R^H = R^{-1}, S^H = S^{-1}}{\text{sup \ re \ tr}(A^T S B R)} = \text{tr}(|\Upsilon| |\Lambda|) \quad (1842)$$

C.4.2.2 Diagonal matrices

Now suppose $A$ and $B$ are diagonal matrices

$$A = \Upsilon = \delta^2(\Upsilon) \in \mathbb{S}^n, \quad \delta(\Upsilon) \in \mathcal{K}_M \quad (1843)$$

$$B = \Lambda = \delta^2(\Lambda) \in \mathbb{S}^n, \quad \delta(\Lambda) \in \mathcal{K}_M \quad (1844)$$

both having their respective main diagonal entries arranged in nonincreasing order:

$$\begin{align*}
\text{minimize} & \quad \|\Upsilon - S \Lambda R\|_F \\
\text{subject to} & \quad R^H = R^{-1} \\
& \quad S^H = S^{-1}
\end{align*} \quad (1845)$$

Then we have a symmetric decomposition from unitary matrices as in (1837) where

$$U_A \triangleq \sqrt{\delta(\psi(\delta(\Upsilon)))}, \quad Q_A \triangleq \sqrt{\delta(\psi(\delta(\Upsilon)))}^H, \quad \Sigma_A = |\Upsilon| \quad (1846)$$

$$U_B \triangleq \sqrt{\delta(\psi(\delta(\Lambda)))}, \quad Q_B \triangleq \sqrt{\delta(\psi(\delta(\Lambda)))}^H, \quad \Sigma_B = |\Lambda| \quad (1847)$$

Procrustes solution (1831) again sees the transposition relationship

$$S^* = U_A U_B^H = R^* T \in \mathbb{C}^{n \times n} \quad (1840)$$

but both optimal unitary matrices are now themselves diagonal. So,

$$S^* \Lambda R^* = \delta(\psi(\delta(\Upsilon))) \Lambda \delta(\psi(\delta(\Lambda))) = \delta(\psi(\delta(\Upsilon))) |\Lambda| \quad (1848)$$
C.5 Nonconvex quadratics

[357, §2] [332, §2] Given $A \in \mathbb{S}^n$, $b \in \mathbb{R}^n$, and $\rho > 0$

$$\begin{align*}
\text{minimize} & \quad \frac{1}{2} x^T A x + b^T x \\
\text{subject to} & \quad \|x\| \leq \rho
\end{align*}$$

(1849)

is a nonconvex problem for symmetric $A$ unless $A \succeq 0$. But necessary and sufficient global optimality conditions are known for any symmetric $A$: vector $x^*$ solves (1849) iff $\exists$ Lagrange multiplier $\lambda^* \geq 0$

\begin{align*}
\text{i) } & \quad (A + \lambda^* I)x^* = -b \\
\text{ii) } & \quad \lambda^*(\|x^*\| - \rho) = 0, \quad \|x^*\| \leq \rho \\
\text{iii) } & \quad A + \lambda^* I \succeq 0
\end{align*}

(1850)

$\lambda^*$ is unique.\footnote{The first two are necessary KKT conditions [63, §5.5.3] while condition iii governs passage to nonconvex global optimality.} $x^*$ is unique if $A + \lambda^* I \succ 0$. Equality constrained problem

$$\begin{align*}
\text{minimize} & \quad \frac{1}{2} x^T A x + b^T x \\
\text{subject to} & \quad \|x\| = \rho
\end{align*}$$

(1851)

is nonconvex for any symmetric $A$ matrix. $x^*$ solves (1851) iff $\exists \lambda^* \in \mathbb{R}$ satisfying the associated conditions. $\lambda^*$ and $x^*$ are unique as before.

Hiriart-Urruty disclosed global optimality conditions in 1998 [207] for maximizing a convexly constrained convex quadratic; a nonconvex problem [318, §32].
Appendix D

Matrix calculus

From too much study, and from extreme passion, cometh madnesse.

− Isaac Newton [162, §5]

D.1 Directional derivative, Taylor series

D.1.1 Gradients

Gradient of a differentiable real function $f(x) : \mathbb{R}^K \to \mathbb{R}$ with respect to its vector argument is defined uniquely in terms of partial derivatives

$$
\nabla f(x) \triangleq \begin{bmatrix}
\frac{\partial f(x)}{\partial x_1} \\
\frac{\partial f(x)}{\partial x_2} \\
\vdots \\
\frac{\partial f(x)}{\partial x_K}
\end{bmatrix} \in \mathbb{R}^K
$$

(1852)

while the second-order gradient of the twice differentiable real function with respect to its vector argument is traditionally called the Hessian;

$$
\nabla^2 f(x) \triangleq \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_K} \\
\frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_K} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f(x)}{\partial x_K \partial x_1} & \frac{\partial^2 f(x)}{\partial x_K \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_K^2}
\end{bmatrix} \in \mathbb{S}^K
$$

(1853)

The gradient of vector-valued function $v(x) : \mathbb{R} \to \mathbb{R}^N$ on real domain is a row-vector

$$
\nabla v(x) \triangleq \begin{bmatrix}
\frac{\partial v_1(x)}{\partial x} \\
\frac{\partial v_2(x)}{\partial x} \\
\vdots \\
\frac{\partial v_N(x)}{\partial x}
\end{bmatrix} \in \mathbb{R}^N
$$

(1854)
while the second-order gradient is

$$\nabla^2 v(x) \triangleq \begin{bmatrix} \frac{\partial^2 v_1(x)}{\partial x_1^2} & \frac{\partial^2 v_1(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 v_1(x)}{\partial x_1 \partial x_N} \\ \frac{\partial^2 v_2(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 v_2(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 v_2(x)}{\partial x_2 \partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 v_N(x)}{\partial x_N \partial x_1} & \frac{\partial^2 v_N(x)}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 v_N(x)}{\partial x_N^2} \end{bmatrix} \in \mathbb{R}^N$$

(1855)

Gradient of vector-valued function $h(x) : \mathbb{R}^K \to \mathbb{R}^N$ on vector domain is

$$\nabla h(x) \triangleq \begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \frac{\partial h_2(x)}{\partial x_1} & \cdots & \frac{\partial h_N(x)}{\partial x_1} \\ \frac{\partial h_1(x)}{\partial x_2} & \frac{\partial h_2(x)}{\partial x_2} & \cdots & \frac{\partial h_N(x)}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_1(x)}{\partial x_K} & \frac{\partial h_2(x)}{\partial x_K} & \cdots & \frac{\partial h_N(x)}{\partial x_K} \end{bmatrix}$$

(1856)

$$= [\nabla h_1(x) \ \nabla h_2(x) \ \cdots \ \nabla h_N(x)] \in \mathbb{R}^{K \times N}$$

while the second-order gradient has a three-dimensional written representation dubbed *cubix* \(^{D.1}\)

$$\nabla^2 h(x) \triangleq \begin{bmatrix} \nabla \frac{\partial h_1(x)}{\partial x_1} & \nabla \frac{\partial h_2(x)}{\partial x_1} & \cdots & \nabla \frac{\partial h_N(x)}{\partial x_1} \\ \nabla \frac{\partial h_1(x)}{\partial x_2} & \nabla \frac{\partial h_2(x)}{\partial x_2} & \cdots & \nabla \frac{\partial h_N(x)}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \nabla \frac{\partial h_1(x)}{\partial x_K} & \nabla \frac{\partial h_2(x)}{\partial x_K} & \cdots & \nabla \frac{\partial h_N(x)}{\partial x_K} \end{bmatrix}$$

(1857)

$$= [\nabla^2 h_1(x) \ \nabla^2 h_2(x) \ \cdots \ \nabla^2 h_N(x)] \in \mathbb{R}^{K \times N \times K}$$

where the gradient of each real entry is with respect to vector $x$ as in (1852).

The gradient of real function $g(X) : \mathbb{R}^{K \times L} \to \mathbb{R}$ on matrix domain is

$$\nabla g(X) \triangleq \begin{bmatrix} \frac{\partial g(X)}{\partial X_{11}} & \frac{\partial g(X)}{\partial X_{12}} & \cdots & \frac{\partial g(X)}{\partial X_{1L}} \\ \frac{\partial g(X)}{\partial X_{21}} & \frac{\partial g(X)}{\partial X_{22}} & \cdots & \frac{\partial g(X)}{\partial X_{2L}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g(X)}{\partial X_{K1}} & \frac{\partial g(X)}{\partial X_{K2}} & \cdots & \frac{\partial g(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K \times L}$$

(1858)

$$= [\nabla_{X(:,1)} g(X) \ \nabla_{X(:,2)} g(X) \ \cdots \ \nabla_{X(:,L)} g(X)] \in \mathbb{R}^{K \times 1 \times L}$$

where gradient $\nabla_{X(:,i)}$ is with respect to the $i^{th}$ column of $X$. The strange appearance of (1858) in $\mathbb{R}^{K \times 1 \times L}$ is meant to suggest a third dimension perpendicular to the page (not

\(^{D.1}\)The word *matrix* comes from the Latin for *womb*; related to the prefix *matri-* derived from *mater* meaning *mother.*
D.1. DIRECTIONAL DERIVATIVE, TAYLOR SERIES

a diagonal matrix). The second-order gradient has representation

\[
\nabla^2 g(X) \triangleq \begin{bmatrix}
\frac{\partial g(X)}{\partial X_{11}} & \frac{\partial g(X)}{\partial X_{12}} & \cdots & \frac{\partial g(X)}{\partial X_{1L}} \\
\frac{\partial g(X)}{\partial X_{21}} & \frac{\partial g(X)}{\partial X_{22}} & \cdots & \frac{\partial g(X)}{\partial X_{2L}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial g(X)}{\partial X_{K1}} & \frac{\partial g(X)}{\partial X_{K2}} & \cdots & \frac{\partial g(X)}{\partial X_{KL}}
\end{bmatrix} \in \mathbb{R}^{K \times L \times K \times L}
\]

(1859)

\[
\begin{align*}
\nabla \nabla_{X(:,1)} g(X) &= \nabla \nabla_{X(:,2)} g(X) \\
& \quad \cdots \\
& \quad \nabla \nabla_{X(:,L)} g(X)
\end{align*}
\]

\[\in \mathbb{R}^{K \times 1 \times L \times K \times L}\]

where the gradient \(\nabla\) is with respect to matrix \(X\).

Gradient of vector-valued function \(g(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R}^N\) on matrix domain is a cubix

\[
\nabla g(X) \triangleq \begin{bmatrix}
\nabla_{X(:,1)} g_1(X) & \nabla_{X(:,1)} g_2(X) & \cdots & \nabla_{X(:,1)} g_N(X) \\
\nabla_{X(:,2)} g_1(X) & \nabla_{X(:,2)} g_2(X) & \cdots & \nabla_{X(:,2)} g_N(X) \\
\vdots & \vdots & \ddots & \vdots \\
\nabla_{X(:,L)} g_1(X) & \nabla_{X(:,L)} g_2(X) & \cdots & \nabla_{X(:,L)} g_N(X)
\end{bmatrix} \in \mathbb{R}^{K \times N \times L}
\]

(1860)

while the second-order gradient has a five-dimensional representation:

\[
\nabla^2 g(X) \triangleq \begin{bmatrix}
\nabla \nabla_{X(:,1)} g_1(X) & \nabla \nabla_{X(:,1)} g_2(X) & \cdots & \nabla \nabla_{X(:,1)} g_N(X) \\
\nabla \nabla_{X(:,2)} g_1(X) & \nabla \nabla_{X(:,2)} g_2(X) & \cdots & \nabla \nabla_{X(:,2)} g_N(X) \\
\vdots & \vdots & \ddots & \vdots \\
\nabla \nabla_{X(:,L)} g_1(X) & \nabla \nabla_{X(:,L)} g_2(X) & \cdots & \nabla \nabla_{X(:,L)} g_N(X)
\end{bmatrix} \in \mathbb{R}^{K \times N \times L \times K \times L}
\]

(1861)

\[
= [\nabla^2 g_1(X) \nabla^2 g_2(X) \cdots \nabla^2 g_N(X)] \in \mathbb{R}^{K \times N \times L \times K \times L}
\]

The gradient of matrix-valued function \(g(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R}^{M \times N}\) on matrix domain has a four-dimensional representation called \textit{quartix} (fourth-order tensor)

\[
\nabla g(X) \triangleq \begin{bmatrix}
\nabla g_{11}(X) & \nabla g_{12}(X) & \cdots & \nabla g_{1N}(X) \\
\nabla g_{21}(X) & \nabla g_{22}(X) & \cdots & \nabla g_{2N}(X) \\
\vdots & \vdots & \ddots & \vdots \\
\nabla g_{M1}(X) & \nabla g_{M2}(X) & \cdots & \nabla g_{MN}(X)
\end{bmatrix} \in \mathbb{R}^{M \times N \times K \times L}
\]

(1862)
while the second-order gradient has a six-dimensional representation

\[
\nabla^2 g(X) \triangleq \begin{bmatrix}
\nabla^2 g_{11}(X) & \nabla^2 g_{12}(X) & \cdots & \nabla^2 g_{1N}(X) \\
\nabla^2 g_{21}(X) & \nabla^2 g_{22}(X) & \cdots & \nabla^2 g_{2N}(X) \\
\vdots & \vdots & \ddots & \vdots \\
\nabla^2 g_{M1}(X) & \nabla^2 g_{M2}(X) & \cdots & \nabla^2 g_{MN}(X)
\end{bmatrix} \in \mathbb{R}^{M \times N \times K \times L \times K \times L} \quad (1863)
\]

and so on.

### D.1.2 Product rules for matrix-functions

Given dimensionally compatible matrix-valued functions of matrix variable \( f(X) \) and \( g(X) \)

\[
\nabla_X (f(X)^T g(X)) = \nabla_X (f) g + \nabla_X (g) f
\]

while [55, §8.3] [326]

\[
\nabla_X \text{tr} (f(X)^T g(X)) = \nabla_X \left( \text{tr} (f(X)^T g(Z)) + \text{tr} (g(X) f(Z)^T) \right) \bigg|_{Z=X} \quad (1865)
\]

These expressions implicitly apply as well to scalar-, vector-, or matrix-valued functions of scalar, vector, or matrix arguments.

#### D.1.2.0.1 Example. Cubix.

Suppose \( f(X) : \mathbb{R}^{2 \times 2} \to \mathbb{R}^2 = X^T a \) and \( g(X) : \mathbb{R}^{2 \times 2} \to \mathbb{R}^2 = X b \). We wish to find

\[
\nabla_X (f(X)^T g(X)) = \nabla_X a^T X^2 b
\]

using the product rule. Formula (1864) calls for

\[
\nabla_X a^T X^2 b = \nabla_X (X^T a) X b + \nabla_X (X b) X^T a
\]

Consider the first of the two terms:

\[
\nabla_X (f) g = \nabla_X (X^T a) X b = \begin{bmatrix} \nabla (X^T a)_1 & \nabla (X^T a)_2 \end{bmatrix} X b
\]

The gradient of \( X^T a \) forms a cubix in \( \mathbb{R}^{2 \times 2 \times 2} \); a.k.a, third-order tensor.

\[
\nabla_X (X^T a) X b = \begin{bmatrix}
\frac{\partial (X^T a)_1}{\partial X_{11}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{11}} \\
\frac{\partial (X^T a)_1}{\partial X_{12}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{12}} \\
\frac{\partial (X^T a)_1}{\partial X_{21}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{21}} \\
\frac{\partial (X^T a)_1}{\partial X_{22}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{22}}
\end{bmatrix} \in \mathbb{R}^{2 \times 1 \times 2}
\]

\[
\begin{bmatrix}
(Xb)_1 \\
(Xb)_2
\end{bmatrix}
\]

\[
\begin{bmatrix}
\frac{\partial (X^T a)_1}{\partial X_{11}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{11}} \\
\frac{\partial (X^T a)_1}{\partial X_{12}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{12}} \\
\frac{\partial (X^T a)_1}{\partial X_{21}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{21}} \\
\frac{\partial (X^T a)_1}{\partial X_{22}} & \cdots & \frac{\partial (X^T a)_2}{\partial X_{22}}
\end{bmatrix} \in \mathbb{R}^{2 \times 1 \times 2}
\]

\[
\begin{bmatrix}
(Xb)_1 \\
(Xb)_2
\end{bmatrix}
\]
Because gradient of the product \((1866)\) requires total change with respect to change in each entry of matrix \(X\), the \(Xb\) vector must make an inner product with each vector in that second dimension of the cubix indicated by dotted line segments:

\[
\nabla_X (X^T a) X b = \begin{bmatrix} a_1 & 0 \\ 0 & a_1 \\ a_2 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} b_1 X_{11} + b_2 X_{12} \\ b_1 X_{21} + b_2 X_{22} \end{bmatrix} \in \mathbb{R}^{2 \times 1 \times 2}
\]

\[
= \begin{bmatrix} a_1 (b_1 X_{11} + b_2 X_{12}) \\ a_2 (b_1 X_{11} + b_2 X_{12}) \\ a_1 (b_1 X_{21} + b_2 X_{22}) \\ a_2 (b_1 X_{21} + b_2 X_{22}) \end{bmatrix} \in \mathbb{R}^{2 \times 2}
\]

\[
= ab^T X^T \tag{1870}
\]

where the cubix appears as a complete \(2 \times 2 \times 2\) matrix. In like manner for the second term \(\nabla_X (g) f\)

\[
\nabla_X (X b) X^T a = \begin{bmatrix} b_1 & 0 \\ b_2 & 0 \\ 0 & b_1 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} X_{11} a_1 + X_{21} a_2 \\ X_{12} a_1 + X_{22} a_2 \end{bmatrix} \in \mathbb{R}^{2 \times 1 \times 2}
\]

\[
= X^T a b^T \in \mathbb{R}^{2 \times 2} \tag{1871}
\]

The solution

\[
\nabla_X a^T X^2 b = ab^T X^T + X^T a b^T \tag{1872}
\]

can be found from Table D.2.1 or verified using \((1865)\). \(\Box\)

### D.1.2.1 Kronecker product

A partial remedy for venturing into hyperdimensional matrix representations, such as the cubix or quartix, is to first vectorize matrices as in \((37)\). This device gives rise to the Kronecker product of matrices \(\otimes\): a.k.a, tensor product. Although it sees reversal in the literature, [337, §2.1] we adopt the definition: for \(A \in \mathbb{R}^{m \times n}\) and \(B \in \mathbb{R}^{p \times q}\)

\[
B \otimes A \triangleq \begin{bmatrix} B_{11} A & B_{12} A & \cdots & B_{1q} A \\ B_{21} A & B_{22} A & \cdots & B_{2q} A \\ \vdots & \vdots & \ddots & \vdots \\ B_{p1} A & B_{p2} A & \cdots & B_{pq} A \end{bmatrix} \in \mathbb{R}^{pm \times qn} \tag{1873}
\]

for which \(A \otimes 1 = 1 \otimes A = A\) (real unity acts like Identity).

One advantage to vectorization is existence of the traditional two-dimensional matrix representation (second-order tensor) for the second-order gradient of a real function with respect to a vectorized matrix. For example, from §A.1.1 no.33 (§D.2.1) for square \(A, B \in \mathbb{R}^{n \times n}\) [176, §5.2] [13, §3]

\[
\nabla^2_{\text{vec} X} \text{tr}(AXBX^T) = \nabla^2_{\text{vec} X} \text{vec}(X^T)(B^T \otimes A) \text{vec} X = B \otimes A^T + B^T \otimes A \in \mathbb{R}^{n^2 \times n^2} \tag{1874}
\]
APPENDIX D. MATRIX CALCULUS

To disadvantage is a large new but known set of algebraic rules (§A.1.1) and the fact that its mere use does not generally guarantee two-dimensional matrix representation of gradients.

Another application of the Kronecker product is to reverse order of appearance in a matrix product: Suppose we wish to weight the columns of a matrix $S \in \mathbb{R}^{M \times N}$, for example, by respective entries $w_i$ from the main diagonal in

$$
W \triangleq \begin{bmatrix} w_1 & 0 & \cdots & 0 \\
0 & \ddots & & \\
\vdots & & \ddots & \\
0 & & & w_N \\
\end{bmatrix} \in \mathbb{S}^N 
$$

(1875)

A conventional means for accomplishing column weighting is to multiply $S$ by diagonal matrix $W$ on the right-hand side:

$$
SW = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\
0 & \ddots & & \\
\vdots & & \ddots & \\
0 & & & w_N \\
\end{bmatrix} \begin{bmatrix} S(:,1) & 0 & \cdots & 0 \\
0 & S(:,2) & & \\
\vdots & & \ddots & \\
0 & & & S(:,N) \\
\end{bmatrix} \in \mathbb{R}^{M \times N} 
$$

(1876)

To reverse product order such that diagonal matrix $W$ instead appears to the left of $S$: for $I \in \mathbb{S}^M$ (Law)

$$
SW = (\delta(W)^T \otimes I) \begin{bmatrix} S(:,1) & 0 & \cdots & 0 \\
0 & S(:,2) & & \\
\vdots & & \ddots & \\
0 & & & S(:,N) \\
\end{bmatrix} \in \mathbb{R}^{M \times N} 
$$

(1877)

To instead weight the rows of $S$ via diagonal matrix $W \in \mathbb{S}^M$, for $I \in \mathbb{S}^N$

$$
WS = \begin{bmatrix} S(1,:) & 0 & \cdots & 0 \\
0 & S(2,:) & & \\
\vdots & & \ddots & \\
0 & & & S(M,:) \\
\end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & 0 \\
\delta(W) & \otimes I \end{bmatrix} \in \mathbb{R}^{M \times N} 
$$

(1878)

For any matrices of like size, $S, Y \in \mathbb{R}^{M \times N}$

$$
S \circ Y = [\delta(Y(:,1)) \cdots \delta(Y(:,N))] \begin{bmatrix} S(:,1) & 0 & \cdots & 0 \\
0 & S(:,2) & & \\
\vdots & & \ddots & \\
0 & & & S(:,N) \\
\end{bmatrix} \in \mathbb{R}^{M \times N} 
$$

(1879)

which converts a Hadamard product into a standard matrix product. In the special case that $S = s$ and $Y = y$ are vectors in $\mathbb{R}^M$

$$
s \circ y = \delta(s)y 
$$

(1880)

$$
s^T \otimes y = ys^T 
$$

(1881)

$$
s \otimes y^T = sy^T 
$$

(1881)
D.1. Chain rules for composite matrix-functions

Given dimensionally compatible matrix-valued functions of matrix variable \( f(X) \) and \( g(X) \) \[229, \S 15.7\]

\[
\nabla_X g(f(X)^T) = \nabla_X f^T \nabla f \quad (1882)
\]

\[
\nabla^2_X g(f(X)^T) = \nabla_X (\nabla_X f^T \nabla f) = \nabla^2_X f \nabla f + \nabla_X f^T \nabla^2 f \nabla f \quad (1883)
\]

D.1.3.1 Two arguments

\[
\nabla_X g(f(X)^T, h(X)^T) = \nabla_X f^T \nabla f \nabla g + \nabla_X h^T \nabla h \nabla g \quad (1884)
\]

D.1.3.1.1 Example. Chain rule for two arguments. \[43, \S 1.1\]

\[
g(f(x)^T, h(x)^T) = (f(x) + h(x))^T A (f(x) + h(x)) \quad (1885)
\]

\[
f(x) = \begin{bmatrix} x_1 \\ \varepsilon x_2 \end{bmatrix}, \quad h(x) = \begin{bmatrix} \varepsilon x_1 \\ x_2 \end{bmatrix} \quad (1886)
\]

\[
\nabla_x g(f(x)^T, h(x)^T) = \begin{bmatrix} 1 & 0 & \varepsilon \\ 0 & 1 + \varepsilon \end{bmatrix} (A + A^T)(f + h) \quad (1887)
\]

\[
\nabla_x g(f(x)^T, h(x)^T) = \begin{bmatrix} 1 + \varepsilon & 0 \\ 0 & 1 + \varepsilon \end{bmatrix} (A + A^T) \left( \begin{bmatrix} x_1 \\ \varepsilon x_2 \end{bmatrix} + \begin{bmatrix} \varepsilon x_1 \\ x_2 \end{bmatrix} \right) \quad (1888)
\]

\[
\lim_{\varepsilon \to 0} \nabla_x g(f(x)^T, h(x)^T) = (A + A^T)x \quad (1889)
\]

from Table D.2.1.

These foregoing formulae remain correct when gradient produces hyperdimensional representation:

D.1.4 First directional derivative

Assume that a differentiable function \( g(X) : \mathbb{R}^{K \times L} \to \mathbb{R}^{M \times N} \) has continuous first- and second-order gradients \( \nabla g \) and \( \nabla^2 g \) over \( \text{dom} \ g \) which is an open set. We seek simple expressions for the first and second directional derivatives in direction \( Y \in \mathbb{R}^{K \times L} : \nabla Y \in \mathbb{R}^{M \times N} \) and \( \nabla^2 Y \in \mathbb{R}^{M \times N} \).

Assuming that the limit exists, we may state the partial derivative of the \( mn \)th entry of \( g \) with respect to the \( kl \)th entry of \( X \):

\[
\frac{\partial g_{mn}(X)}{\partial X_{kl}} = \lim_{\Delta t \to 0} \frac{g_{mn}(X + \Delta t e_k e_l^T) - g_{mn}(X)}{\Delta t} \in \mathbb{R} \quad (1890)
\]
where $e_k$ is the $k$th standard basis vector in $\mathbb{R}^K$ while $e_l$ is the $l$th standard basis vector in $\mathbb{R}^L$. The total number of partial derivatives equals $KLMN$ while the gradient is defined in their terms; the $mn$th entry of the gradient is

$$\nabla g_{mn}(X) = \begin{bmatrix} \frac{\partial g_{mn}(X)}{\partial X_{11}} & \frac{\partial g_{mn}(X)}{\partial X_{12}} & \cdots & \frac{\partial g_{mn}(X)}{\partial X_{1L}} \\ \frac{\partial g_{mn}(X)}{\partial X_{21}} & \frac{\partial g_{mn}(X)}{\partial X_{22}} & \cdots & \frac{\partial g_{mn}(X)}{\partial X_{2L}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_{mn}(X)}{\partial X_{K1}} & \frac{\partial g_{mn}(X)}{\partial X_{K2}} & \cdots & \frac{\partial g_{mn}(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K \times L} \quad (1891)$$

while the gradient is a quartix

$$\nabla g(X) = \begin{bmatrix} \nabla g_{11}(X) & \nabla g_{12}(X) & \cdots & \nabla g_{1N}(X) \\ \nabla g_{21}(X) & \nabla g_{22}(X) & \cdots & \nabla g_{2N}(X) \\ \vdots & \vdots & \ddots & \vdots \\ \nabla g_{M1}(X) & \nabla g_{M2}(X) & \cdots & \nabla g_{MN}(X) \end{bmatrix} \in \mathbb{R}^{M \times N \times K \times L} \quad (1892)$$

By simply rotating our perspective of a four-dimensional representation of gradient matrix, we find one of three useful transpositions of this quartix (connoted $T_1$):

$$\nabla g(X)^{T_1} = \begin{bmatrix} \frac{\partial g(X)}{\partial X_{11}} & \frac{\partial g(X)}{\partial X_{12}} & \cdots & \frac{\partial g(X)}{\partial X_{1L}} \\ \frac{\partial g(X)}{\partial X_{21}} & \frac{\partial g(X)}{\partial X_{22}} & \cdots & \frac{\partial g(X)}{\partial X_{2L}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g(X)}{\partial X_{K1}} & \frac{\partial g(X)}{\partial X_{K2}} & \cdots & \frac{\partial g(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K \times L \times M \times N} \quad (1893)$$

When the limit for $\Delta t \in \mathbb{R}$ exists, it is easy to show by substitution of variables in (1890)

$$\frac{\partial g_{mn}(X)}{\partial X_{kl}} Y_{kl} = \lim_{\Delta t \to 0} \frac{g_{mn}(X + \Delta t Y_{kl} e_k e_l^T) - g_{mn}(X)}{\Delta t} \in \mathbb{R} \quad (1894)$$

which may be interpreted as the change in $g_{mn}$ at $X$ when the change in $X_{kl}$ is equal to $Y_{kl}$, the $kl$th entry of any $Y \in \mathbb{R}^{K \times L}$. Because the total change in $g_{mn}(X)$ due to $Y$ is the sum of change with respect to each and every $X_{kl}$, the $mn$th entry of the directional derivative is the corresponding total differential [229, §15.8]

$$dg_{mn}(X)|_{dX \to Y} = \sum_{k,l} \frac{\partial g_{mn}(X)}{\partial X_{kl}} Y_{kl} = \text{tr}(\nabla g_{mn}(X)^T Y) \quad (1895)$$

$$= \sum_{k,l} \lim_{\Delta t \to 0} \frac{g_{mn}(X + \Delta t Y_{kl} e_k e_l^T) - g_{mn}(X)}{\Delta t} \quad (1896)$$

$$= \lim_{\Delta t \to 0} \frac{g_{mn}(X + \Delta t Y) - g_{mn}(X)}{\Delta t} \quad (1897)$$

$$= \frac{d}{dt} \bigg|_{t=0} g_{mn}(X + t Y) \quad (1898)$$
where \( t \in \mathbb{R} \). Assuming finite \( Y \), equation (1897) is called the \textit{Gâteaux differential} \cite{42, App.A.5} \cite{209, §D.2.1} \cite{370, §5.28} whose existence is implied by existence of the \textit{Fréchet differential} (the sum in (1895)). \cite{260, §7.2} Each may be understood as the change in \( g_{mn} \) at \( X \) when the change in \( X \) is equal in magnitude and direction to \( Y \). \( \text{D.2} \)

Hence the directional derivative,

\[
\neg Y \frac{d g(X)}{dX} = \begin{bmatrix}
\frac{\partial g_1(X)}{\partial X_{1,1}} & \frac{\partial g_{12}(X)}{\partial X_{1,2}} & \cdots & \frac{\partial g_{1N}(X)}{\partial X_{1,N}} \\
\frac{\partial g_{21}(X)}{\partial X_{2,1}} & \frac{\partial g_{22}(X)}{\partial X_{2,2}} & \cdots & \frac{\partial g_{2N}(X)}{\partial X_{2,N}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial g_{M1}(X)}{\partial X_{M,1}} & \frac{\partial g_{M2}(X)}{\partial X_{M,2}} & \cdots & \frac{\partial g_{MN}(X)}{\partial X_{M,N}}
\end{bmatrix}
\in \mathbb{R}^{M \times N}
\]

\[
= \begin{bmatrix}
\text{tr}(\nabla g_1(X)^T Y) & \text{tr}(\nabla g_{12}(X)^T Y) & \cdots & \text{tr}(\nabla g_{1N}(X)^T Y) \\
\text{tr}(\nabla g_{21}(X)^T Y) & \text{tr}(\nabla g_{22}(X)^T Y) & \cdots & \text{tr}(\nabla g_{2N}(X)^T Y) \\
\vdots & \vdots & \ddots & \vdots \\
\text{tr}(\nabla g_{M1}(X)^T Y) & \text{tr}(\nabla g_{M2}(X)^T Y) & \cdots & \text{tr}(\nabla g_{MN}(X)^T Y)
\end{bmatrix}
\quad (1899)
\]

from which it follows

\[
\neg Y d g(X) = \sum_{k,l} \frac{\partial g(X)}{\partial X_{kl}} Y_{kl} \quad (1900)
\]

Yet for all \( X \in \text{dom} \, g \), any \( Y \in \mathbb{R}^{K \times L} \), and some open interval of \( t \in \mathbb{R} \)

\[
g(X + t Y) = g(X) + t \neg Y d g(X) + o(t^2) \quad (1901)
\]

which is the first-order Taylor series expansion about \( X \). \cite{229, §18.4} \cite{160, §2.3.4} Differentiation with respect to \( t \) and subsequent \( t \)-zeroing isolates the second term of expansion. Thus differentiating and zeroing \( g(X + t Y) \) in \( t \) is an operation equivalent to individually differentiating and zeroing every entry \( g_{mn}(X + t Y) \) as in (1898). So the directional derivative of \( g(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R}^{M \times N} \) in any direction \( Y \in \mathbb{R}^{K \times L} \) evaluated at \( X \in \text{dom} \, g \) becomes

\[
\neg Y d g(X) = \left. \frac{d}{dt} \right|_{t=0} g(X + t Y) \in \mathbb{R}^{M \times N} \quad (1902)
\]

\cite{288, §2.1, §5.4.5} \cite{35, §6.3.1} which is simplest. In case of a real function \( g(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R} \)

\[
\neg Y d g(X) = \text{tr}(\nabla g(X)^T Y) \quad (1924)
\]

\( \text{D.2} \) Although \( Y \) is a matrix, we may regard it as a vector in \( \mathbb{R}^{KL} \).
Figure 172: Strictly convex quadratic bowl in \( \mathbb{R}^2 \times \mathbb{R} \); \( f(x) = x^T x : \mathbb{R}^2 \to \mathbb{R} \) versus \( x \) on some open disc in \( \mathbb{R}^2 \). Plane slice \( \partial \mathcal{H} \) is perpendicular to function domain. Slice intersection with domain connotes bi-directional vector \( y \). Slope of tangent line \( \mathcal{T} \) at point \( (\alpha, f(\alpha)) \) is value of directional derivative \( \nabla_x f(\alpha)^T y \) at \( \alpha \) in slice direction \( y \). Negative gradient \( -\nabla_x f(x) \in \mathbb{R}^2 \) is direction of steepest descent. [407] [229, §15.6] [160] When vector \( v \in \mathbb{R}^3 \) entry \( v_3 \) is half directional derivative in gradient direction at \( \alpha \) and when \( \begin{bmatrix} v_1 \\ v_2 \\ \frac{1}{2} df(\alpha) \end{bmatrix} = \nabla_x f(\alpha) \), then \( -v \) points directly toward bowl bottom.

In case \( g(X) : \mathbb{R}^K \to \mathbb{R} \)

\[
\overrightarrow{Y} \overset{\Delta}{=} \nabla g(X)^T Y 
\]  
(1927)

Unlike gradient, directional derivative does not expand dimension; directional derivative (1902) retains the dimensions of \( g \). The derivative with respect to \( t \) makes the directional derivative resemble ordinary calculus (§D.2); e.g., when \( g(X) \) is linear, \( \overrightarrow{Y} \overset{\Delta}{=} g(Y) \). [260, §7.2]

D.1.4.1 Interpretation of directional derivative

In the case of any differentiable real function \( g(X) : \mathbb{R}^{K \times L} \to \mathbb{R} \), the directional derivative of \( g(X) \) at \( X \) in any direction \( Y \) yields the slope of \( g \) along the line \( \{ X + t Y \mid t \in \mathbb{R} \} \) through its domain evaluated at \( t = 0 \). For higher-dimensional functions, by (1899), this slope interpretation can be applied to each entry of the directional derivative.

Figure 172, for example, shows a plane slice of a real convex bowl-shaped function \( f(x) \) along a line \( \{ \alpha + t y \mid t \in \mathbb{R} \} \) through its domain. The slice reveals a one-dimensional real function of \( t : f(\alpha + t y) \). The directional derivative at \( x = \alpha \) in direction \( y \) is the slope of \( f(\alpha + t y) \) with respect to \( t \) at \( t = 0 \). In the case of a real function having vector argument \( h(X) : \mathbb{R}^K \to \mathbb{R} \), its directional derivative in the normalized direction
of its gradient is the gradient magnitude. \(^{(1927)}\) For a real function of real variable, the directional derivative evaluated at any point in the function domain is just the slope of that function there scaled by the real direction. \(\text{confer \S 3.6}\)

Directional derivative generalizes our one-dimensional notion of derivative to a multidimensional domain. When direction \(Y\) coincides with a member of the standard Cartesian basis \(e_k e_l^T\) \(^{(60)}\), then a single partial derivative \(\partial g(X)/\partial X_{kl}\) is obtained from directional derivative \(^{(1900)}\); such is each entry of gradient \(\nabla g(X)\) in equalities \(^{(1924)}\) and \(^{(1927)}\), for example.

**D.1.4.1.1 Theorem.** Directional derivative optimality condition. \([260, \S 7.4]\)

Suppose \(f(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R}\) is minimized on convex set \(C \subseteq \mathbb{R}^{K \times L}\) by \(X^*\), and the directional derivative of \(f\) exists there. Then for all \(X \in C\)

\[
-\langle X - X^* \rangle d f(X) \geq 0 \tag{1903}
\]

\(\diamondsuit\)

**D.1.4.1.2 Example.** Simple bowl.

Bowl function (Figure 172)

\[
f(x) : \mathbb{R}^K \rightarrow \mathbb{R} \triangleq (x - a)^T(x - a) - b \tag{1904}
\]

has function offset \(-b \in \mathbb{R}\), axis of revolution at \(x = a\), and positive definite Hessian \(^{(1853)}\) everywhere in its domain (an open hyperdisc in \(\mathbb{R}^K\)); \(\text{id est}\), strictly convex quadratic \(f(x)\) has unique global minimum equal to \(-b\) at \(x = a\). A vector \(-\upsilon\) based anywhere in \(\text{dom } f \times \mathbb{R}\) pointing toward the unique bowl-bottom is specified:

\[
\upsilon \propto \begin{bmatrix} x - a \\ f(x) + b \end{bmatrix} \in \mathbb{R}^K \times \mathbb{R} \tag{1905}
\]

Such a vector is

\[
\upsilon = \begin{bmatrix} \nabla_x f(x) \\ -\nabla_x f(x) \\ \frac{1}{2} d f(x) \end{bmatrix} \tag{1906}
\]

since the gradient is

\[
\nabla_x f(x) = 2(x - a) \tag{1907}
\]

and the directional derivative in direction of the gradient is \(^{(1927)}\)

\[
-\nabla_x f(x) d f(x) = \nabla_x f(x)^T \nabla_x f(x) = 4(x - a)^T(x - a) = 4(f(x) + b) \tag{1908}
\]

\(\square\)
D.1.5 Second directional derivative

By similar argument, it so happens: the second directional derivative is equally simple. Given \( g(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R}^{M \times N} \) on open domain,

\[
\nabla \frac{\partial g_{mn}(X)}{\partial X_{kl}} = \frac{\partial \nabla g_{mn}(X)}{\partial X_{kl}} = \begin{bmatrix}
\frac{\partial^2 g_{mn}(X)}{\partial X_{k1} \partial X_{l1}} & \frac{\partial^2 g_{mn}(X)}{\partial X_{k1} \partial X_{l2}} & \cdots & \frac{\partial^2 g_{mn}(X)}{\partial X_{k1} \partial X_{lL}} \\
\frac{\partial^2 g_{mn}(X)}{\partial X_{k2} \partial X_{l1}} & \frac{\partial^2 g_{mn}(X)}{\partial X_{k2} \partial X_{l2}} & \cdots & \frac{\partial^2 g_{mn}(X)}{\partial X_{k2} \partial X_{lL}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 g_{mn}(X)}{\partial X_{kK} \partial X_{l1}} & \frac{\partial^2 g_{mn}(X)}{\partial X_{kK} \partial X_{l2}} & \cdots & \frac{\partial^2 g_{mn}(X)}{\partial X_{kK} \partial X_{lL}}
\end{bmatrix} \in \mathbb{R}^{K \times L} 
\]  

(1090)

\[
\nabla^2 g_{mn}(X) = \begin{bmatrix}
\nabla \frac{\partial g_{mn}(X)}{\partial X_{11}} & \nabla \frac{\partial g_{mn}(X)}{\partial X_{12}} & \cdots & \nabla \frac{\partial g_{mn}(X)}{\partial X_{1L}} \\
\nabla \frac{\partial g_{mn}(X)}{\partial X_{21}} & \nabla \frac{\partial g_{mn}(X)}{\partial X_{22}} & \cdots & \nabla \frac{\partial g_{mn}(X)}{\partial X_{2L}} \\
\vdots & \vdots & \ddots & \vdots \\
\nabla \frac{\partial g_{mn}(X)}{\partial X_{K1}} & \nabla \frac{\partial g_{mn}(X)}{\partial X_{K2}} & \cdots & \nabla \frac{\partial g_{mn}(X)}{\partial X_{KL}}
\end{bmatrix} \in \mathbb{R}^{K \times L \times K \times L} 
\]  

(1010)

Rotating our perspective, we get several views of the second-order gradient:

\[
\nabla^2 g(X) = \begin{bmatrix}
\nabla^2 g_{11}(X) & \nabla^2 g_{12}(X) & \cdots & \nabla^2 g_{1N}(X) \\
\nabla^2 g_{21}(X) & \nabla^2 g_{22}(X) & \cdots & \nabla^2 g_{2N}(X) \\
\vdots & \vdots & \ddots & \vdots \\
\nabla^2 g_{M1}(X) & \nabla^2 g_{M2}(X) & \cdots & \nabla^2 g_{MN}(X)
\end{bmatrix} \in \mathbb{R}^{M \times N \times K \times L \times K \times L} 
\]  

(1111)

\[
\nabla^2 g(X)^T_1 = \begin{bmatrix}
\nabla \frac{\partial g(X)}{\partial X_{11}} & \nabla \frac{\partial g(X)}{\partial X_{12}} & \cdots & \nabla \frac{\partial g(X)}{\partial X_{1L}} \\
\nabla \frac{\partial g(X)}{\partial X_{21}} & \nabla \frac{\partial g(X)}{\partial X_{22}} & \cdots & \nabla \frac{\partial g(X)}{\partial X_{2L}} \\
\vdots & \vdots & \ddots & \vdots \\
\nabla \frac{\partial g(X)}{\partial X_{K1}} & \nabla \frac{\partial g(X)}{\partial X_{K2}} & \cdots & \nabla \frac{\partial g(X)}{\partial X_{KL}}
\end{bmatrix} \in \mathbb{R}^{K \times L \times M \times N \times K \times L} 
\]  

(1112)

\[
\nabla^2 g(X)^T_2 = \begin{bmatrix}
\frac{\partial^2 g(X)}{\partial X_{11} \partial X_{11}} & \frac{\partial^2 g(X)}{\partial X_{11} \partial X_{12}} & \cdots & \frac{\partial^2 g(X)}{\partial X_{11} \partial X_{1L}} \\
\frac{\partial^2 g(X)}{\partial X_{12} \partial X_{11}} & \frac{\partial^2 g(X)}{\partial X_{12} \partial X_{12}} & \cdots & \frac{\partial^2 g(X)}{\partial X_{12} \partial X_{1L}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 g(X)}{\partial X_{KL} \partial X_{11}} & \frac{\partial^2 g(X)}{\partial X_{KL} \partial X_{12}} & \cdots & \frac{\partial^2 g(X)}{\partial X_{KL} \partial X_{1L}}
\end{bmatrix} \in \mathbb{R}^{K \times L \times K \times L \times M \times N} 
\]  

(1113)
D.1. DIRECTIONAL DERIVATIVE, TAYLOR SERIES

Assuming the limits exist, we may state the partial derivative of the \( mn \)-th entry of \( g \) with respect to the \( kl \)-th and \( ij \)-th entries of \( X \):

\[
\frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{ij}} = \lim_{\Delta t \to 0} \frac{g_{mn}(X + \Delta t e_k e_l + \Delta \tau e_i e_j) - g_{mn}(X + \Delta \tau e_i e_j) - (g_{mn}(X + \Delta \tau e_i e_j) - g_{mn}(X))}{\Delta \tau \Delta t}
\]  

(1914)

Differentiating (1994) and then scaling by \( Y_{ij} \)

\[
\frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} = \lim_{\Delta t \to 0} \frac{\partial g_{mn}(X + \Delta t Y_{kl} e_k e_l + \Delta \tau Y_{ij} e_i e_j) - \partial g_{mn}(X) Y_{ij}}{\partial X_{ij} \Delta t}
\]

\[
= \lim_{\Delta t \to 0} \frac{g_{mn}(X + 2 \Delta t Y) - 2g_{mn}(X + \Delta t Y) + g_{mn}(X)}{\Delta t^2}
\]

(1915)

which can be proved by substitution of variables in (1914). The \( mn \)-th second-order total differential due to any \( Y \in \mathbb{R}^{K \times L} \) is

\[
d^2 g_{mn}(X) \mid_{dX \rightarrow Y} = \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} = \text{tr} \left( \nabla_X \text{tr} \left( \nabla g_{mn}(X)^T Y \right)^T Y \right)
\]

(1916)

(1917)

(1918)

(1919)

Hence the second directional derivative,

\[
\frac{-Y}{dY^2} \Delta Y \triangleq \left[ \begin{array}{cccc}
    d^2 g_{11}(X) & d^2 g_{12}(X) & \cdots & d^2 g_{1N}(X) \\
    d^2 g_{21}(X) & d^2 g_{22}(X) & \cdots & d^2 g_{2N}(X) \\
    \vdots & \vdots & \ddots & \vdots \\
    d^2 g_{M1}(X) & d^2 g_{M2}(X) & \cdots & d^2 g_{MN}(X)
\end{array} \right] \in \mathbb{R}^{M \times N}
\]

\[
= \left[ \begin{array}{c}
    \text{tr} \left( \nabla \text{tr} \left( \nabla g_{11}(X)^T Y \right)^T Y \right) \\
    \text{tr} \left( \nabla \text{tr} \left( \nabla g_{12}(X)^T Y \right)^T Y \right) & \cdots & \text{tr} \left( \nabla \text{tr} \left( \nabla g_{1N}(X)^T Y \right)^T Y \right) \\
    \text{tr} \left( \nabla \text{tr} \left( \nabla g_{21}(X)^T Y \right)^T Y \right) & \cdots & \text{tr} \left( \nabla \text{tr} \left( \nabla g_{2N}(X)^T Y \right)^T Y \right) \\
    \vdots & \vdots & \ddots & \vdots \\
    \text{tr} \left( \nabla \text{tr} \left( \nabla g_{M1}(X)^T Y \right)^T Y \right) & \cdots & \text{tr} \left( \nabla \text{tr} \left( \nabla g_{MN}(X)^T Y \right)^T Y \right)
\end{array} \right]
\]

(1920)

\[
= \left[ \begin{array}{cccc}
    \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{11}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} & \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{12}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} & \cdots & \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{1N}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} \\
    \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{21}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} & \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{22}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} & \cdots & \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{2N}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} \\
    \vdots & \vdots & \ddots & \vdots \\
    \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{M1}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} & \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{M2}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} & \cdots & \sum_{i,j} \sum_{k,l} \frac{\partial^2 g_{MN}(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij}
\end{array} \right]
\]
from which it follows
\[
-^Ydg^2(X) = \sum_{i,j} \sum_{k,l} \frac{\partial^2 g(X)}{\partial X_{kl} \partial X_{ij}} Y_{kl} Y_{ij} = \sum_{i,j} \frac{\partial}{\partial X_{ij}} -^Ydg(X) Y_{ij}
\]  
(1921)

Yet for all \(X \in \text{dom } g\), any \(Y \in \mathbb{R}^{K \times L}\), and some open interval of \(t \in \mathbb{R}\)

\[
g(X + tY) = g(X) + t \cdot -^Ydg(X) + \frac{1}{2!} t^2 -^Ydg^2(X) + o(t^3)
\]  
(1922)

which is the second-order Taylor series expansion about \(X\). \[\text{[229, §18.4] [160, §2.3.4]}

Differentiating twice with respect to \(t\) and subsequent \(t\)-zeroing isolates the third term of the expansion. Thus differentiating and zeroing \(g(X + tY)\) in \(t\) is an operation equivalent to individually differentiating and zeroing every entry \(g_{mn}(X + tY)\) as in (1919). So the second directional derivative of \(g(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R}^{M \times N}\) becomes \[\text{[288, §2.1, §5.4.5] [35, §6.3.1]}

\[
-^Ydg^2(X) = \left. \frac{d^2}{dt^2} \right|_{t=0} g(X + tY) \in \mathbb{R}^{M \times N}
\]  
(1923)

which is again simplest. (confer (1902)) Directional derivative retains the dimensions of \(g\).

D.1.6 directional derivative expressions

In the case of a real function \(g(X) : \mathbb{R}^{K \times L} \rightarrow \mathbb{R}\), all its directional derivatives are in \(\mathbb{R}\):

\[
-^Ydg(X) = \text{tr}(\nabla g(X)^T Y)
\]  
(1924)

\[
-^Ydg^2(X) = \text{tr}(\nabla_X \text{tr}(\nabla g(X)^T Y)^T Y) = \text{tr}(\nabla_X -^Ydg(X)^T Y)
\]  
(1925)

\[
-^Ydg^3(X) = \text{tr}(\nabla_X \text{tr}(\nabla_X \text{tr}(\nabla g(X)^T Y)^T Y)^T Y) = \text{tr}(\nabla_X -^Ydg^2(X)^T Y)
\]  
(1926)

In the case \(g(X) : \mathbb{R}^{K} \rightarrow \mathbb{R}\) has vector argument, they further simplify:

\[
-^Ydg(X) = \nabla g(X)^T Y
\]  
(1927)

\[
-^Ydg^2(X) = Y^T \nabla^2 g(X) Y
\]  
(1928)

\[
-^Ydg^3(X) = \nabla_X (Y^T \nabla^2 g(X) Y)^T Y
\]  
(1929)

and so on.
D.1.7 Taylor series

Series expansions of the differentiable matrix-valued function \( g(X) \), of matrix argument, were given earlier in (1901) and (1922). Assuming \( g(X) \) has continuous first-, second-, and third-order gradients over the open set \( \text{dom } g \), then for \( X \in \text{dom } g \) and any \( Y \in \mathbb{R}^{K \times L} \), the complete Taylor series is expressed on some open interval of \( \mu \in \mathbb{R} \)

\[
g(X + \mu Y) = g(X) + \frac{-Y}{\mu} dg(X) + \frac{1}{2!} \frac{-Y}{\mu^2} dg^2(X) + \frac{1}{3!} \frac{-Y}{\mu^3} dg^3(X) + o(\mu^4) \tag{1930}
\]
or on some open interval of \( \|Y\|_2 \)

\[
g(Y) = g(X) + \frac{-Y}{\mu} dg(X) + \frac{1}{2!} \frac{-Y}{\mu^2} dg^2(X) + \frac{1}{3!} \frac{-Y}{\mu^3} dg^3(X) + o(\|Y\|^4) \tag{1931}
\]

which are third-order expansions about \( X \). The mean value theorem from calculus is what insures finite order of the series. [229] [43, §1.1] [42, App.A.5] [209, §0.4] These somewhat unbelievable formulae imply that a function can be determined over the whole of its domain by knowing its value and all its directional derivatives at a single point \( X \).

D.1.7.0.1 Example. Inverse-matrix function. Say \( g(Y) = Y^{-1} \). From the table on page 594,

\[
\frac{-Y}{\mu} dg(X) = \frac{d}{dt} \bigg|_{t=0} g(X + tY) = -X^{-1}YX^{-1} \tag{1932}
\]

\[
\frac{-Y}{\mu^2} dg^2(X) = \frac{d^2}{dt^2} \bigg|_{t=0} g(X + tY) = 2X^{-1}YX^{-1}YX^{-1} \tag{1933}
\]

\[
\frac{-Y}{\mu^3} dg^3(X) = \frac{d^3}{dt^3} \bigg|_{t=0} g(X + tY) = -6X^{-1}YX^{-1}YX^{-1}YX^{-1} \tag{1934}
\]

Let’s find the Taylor series expansion of \( g \) about \( X = I \): Since \( g(I) = I \), for \( \|Y\|_2 < 1 \) \( (\mu = 1 \text{ in } (1930)) \)

\[
g(I + Y) = (I + Y)^{-1} = I - Y + Y^2 - Y^3 + \ldots \tag{1935}
\]

If \( Y \) is small, \( (I + Y)^{-1} \approx I - Y \).\(^3\) Now we find Taylor series expansion about \( X \):

\[
g(X + Y) = (X + Y)^{-1} = X^{-1} - X^{-1}YX^{-1} + 2X^{-1}YX^{-1}YX^{-1} - \ldots \tag{1936}
\]

If \( Y \) is small, \( (X + Y)^{-1} \approx X^{-1} - X^{-1}YX^{-1} \). \( \square \)

D.1.7.0.2 Exercise. \( \log \det \). \((\text{confer } [63, \text{p.644}])\)

Find the first three terms of a Taylor series expansion for \( \log \det Y \). Specify an open interval over which the expansion holds in vicinity of \( X \). \( \nabla \)

\(^3\) Had we instead set \( g(Y) = (I + Y)^{-1} \), then the equivalent expansion would have been about \( X = 0 \).
D.1.8 Correspondence of gradient to derivative

From the foregoing expressions for directional derivative, we derive a relationship between gradient with respect to matrix \( X \) and derivative with respect to real variable \( t \):

D.1.8.1 first-order

Removing evaluation at \( t = 0 \) from \((1902)\),\(^D.4\) we find an expression for the directional derivative of \( g(X) \) in direction \( Y \) evaluated anywhere along a line \( \{X + t Y \mid t \in \mathbb{R}\} \) intersecting \( \text{dom } g \):

\[
\frac{d}{dt} g(X + t Y) = \frac{d}{dt} g(X + t Y) \tag{1937}
\]

In the general case \( g(X) : \mathbb{R}^{K \times L} \to \mathbb{R}^{M \times N} \), from \((1895)\) and \((1898)\) we find

\[
\text{tr} \left( \nabla_X g_{mn} (X + t Y)^T Y \right) = \frac{d}{dt} g_{mn}(X + t Y) \tag{1938}
\]

which is valid at \( t = 0 \), of course, when \( X \in \text{dom } g \). In the important case of a real function \( g(X) : \mathbb{R}^{K \times L} \to \mathbb{R} \), from \((1924)\) we have simply

\[
\text{tr} \left( \nabla_X g(X + t Y)^T Y \right) = \frac{d}{dt} g(X + t Y) \tag{1939}
\]

When, additionally, \( g(X) : \mathbb{R}^K \to \mathbb{R} \) has vector argument,

\[
\nabla_X g(X + t Y)^T Y = \frac{d}{dt} g(X + t Y) \tag{1940}
\]

D.1.8.1.1 Example. Gradient.

\( g(X) = w^T X^T X w, \ X \in \mathbb{R}^{K \times L}, \ w \in \mathbb{R}^L \). Using the tables in §D.2,

\[
\text{tr} \left( \nabla_X g(X + t Y)^T Y \right) = \text{tr} \left( 2w w^T (X^T + t Y^T) Y \right) = 2w^T (X^T Y + t Y^T Y) w \tag{1941}
\]

Applying equivalence \((1939)\),

\[
\frac{d}{dt} g(X + t Y) = \frac{d}{dt} w^T (X + t Y)^T (X + t Y) w = w^T (X^T Y + Y^T X + 2t Y^T Y) w = 2w^T (X^T Y + t Y^T Y) w \tag{1942}
\]

which is the same as \((1942)\). Hence, the equivalence is demonstrated.

\(^D.4\)Justified by replacing \( X \) with \( X + t Y \) in \((1895)\)-(1897); beginning,

\[
d_{g_{mn}}(X + t Y)|_{X \to Y} = \sum_{k,l} \frac{\partial g_{mn}(X + t Y)}{\partial X_{kl}} Y_{kl}
\]
D.1. DIRECTIONAL DERIVATIVE, TAYLOR SERIES

It is easy to extract $\nabla g(X)$ from (1945) knowing only (1939):

\[
\text{tr}(\nabla_X g(X + tY)^T Y) = 2w^T(X^TY + tY^TY)w \\
\text{tr}(\nabla_X g(X)^T Y) = 2\text{tr}(ww^TX^TY) \\
\nabla_X g(X) = 2Xww^T
\]

\[\text{(1946)}\]

D.1.8.2 second-order

Likewise removing the evaluation at $t = 0$ from (1923),

\[
-\frac{d}{dt}^2 g(X + tY) = \frac{d^2}{dt^2} g(X + tY)
\]

we can find a similar relationship between second-order gradient and second derivative: In the general case $g(X) : \mathbb{R}^{K \times L} \to \mathbb{R}^{M \times N}$ from (1916) and (1919),

\[
\text{tr}\left(\nabla_X \text{tr}(\nabla_X g_{mn}(X + tY)^T Y)^TY\right) = \frac{d^2}{dt^2} g_{mn}(X + tY)
\]

We have, of course,

\[
\text{tr}\left(\nabla_X \text{tr}(\nabla_X g(X + tY)^T Y)^T Y\right) = \frac{d^2}{dt^2} g(X + tY)
\]

In the case of a real function $g(X) : \mathbb{R}^{K \times L} \to \mathbb{R}$ we have, of course,

\[
\text{tr}\left(\nabla_X \text{tr}(\nabla_X g(X + tY)^T Y)^T Y\right) = \frac{d^2}{dt^2} g(X + tY)
\]

From (1928), the simpler case, where real function $g(X) : \mathbb{R}^K \to \mathbb{R}$ has vector argument,

\[
Y^T \nabla_X^2 g(X + tY) = \frac{d^2}{dt^2} g(X + tY)
\]

D.1.8.2.1 Example. Second-order gradient.

We want to find $\nabla^2 g(X) \in \mathbb{R}^{K \times K \times K \times K}$ given real function $g(X) = \log \det X$ having domain int $S^K_+$. From the tables in §D.2,

\[
h(X) \triangleq \nabla g(X) = X^{-1} \in \text{int} S^K_+
\]

so $\nabla^2 g(X) = \nabla h(X)$. By (1938) and (1901), for $Y \in S^K$

\[
\text{tr}(\nabla h_{mn}(X)^T Y) = \frac{d}{dt} \bigg|_{t=0} h_{mn}(X + tY)
\]

\[
= \left( \frac{d}{dt} \bigg|_{t=0} h(X + tY) \right)_{mn}
\]

\[
= \left( \frac{d}{dt} \bigg|_{t=0} (X + tY)^{-1} \right)_{mn}
\]

\[
= - (X^{-1} Y X^{-1})_{mn}
\]

\[\text{(1952)-(1955)}\]
Setting $Y$ to a member of \{ $e_k e_l^T \in \mathbb{R}^{K \times K} \mid k, l = 1 \ldots K$ \}, and employing a property (39) of the trace function we find

$$\nabla^2 g(X)_{mnkl} = \text{tr}(\nabla h_{mn}(X)^T e_k e_l^T) = \nabla h_{mn}(X)_{kl} = -(X^{-1} e_k e_l^T X^{-1})_{mn}$$  \hspace{1cm} (1956)

$$\nabla^2 g(X)_{kl} = \nabla h(X)_{kl} = -(X^{-1} e_k e_l^T X^{-1}) \in \mathbb{R}^{K \times K}$$  \hspace{1cm} (1957)

From all these first- and second-order expressions, we may generate new ones by evaluating both sides at arbitrary $t$ (in some open interval) but only after the differentiation.

### D.2 Tables of gradients and derivatives

- Results may be numerically proven by Romberg extrapolation. [114] When proving results for symmetric matrices algebraically, it is critical to take gradients ignoring symmetry and to then substitute symmetric entries afterward. [176] [67]

- $a, b \in \mathbb{R}^n, \ x, y \in \mathbb{R}^k, \ A, B \in \mathbb{R}^{m \times n}, \ X, Y \in \mathbb{R}^{K \times L}, \ t, \mu \in \mathbb{R}, \ i, j, k, \ell, K, L, m, n, M, N$ are integers, unless otherwise noted.

- $x^\mu$ means $\delta(\delta(x)^\mu)$ for $\mu \in \mathbb{R}$; id est, entrywise vector exponentiation. $\delta$ is the main-diagonal linear operator (1497). $x^0 \triangleq 1, \ X^0 \triangleq I$ if square.

- $\frac{d}{dx} \begin{bmatrix} \frac{d}{dx} y_1 \\ \vdots \\ \frac{d}{dx} y_k \end{bmatrix} = -y, \ -yg(x), \ -y^2g(x)$ (directional derivatives §D.1), $\log x$, $e^x$, $|x|$, $\text{sgn } x$, $x/y$ (Hadamard quotient), $\sqrt{x}$ (entrywise square root), etcetera, are maps $f : \mathbb{R}^k \to \mathbb{R}^k$ that maintain dimension; e.g, (§A.1.1)

$$\frac{d}{dx} x^{-1} \triangleq \nabla_x I^T \delta(x)^{-1}$$  \hspace{1cm} (1958)

- For $A$ a scalar or square matrix, we have the Taylor series [79, §3.6]

$$e^A \triangleq \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$  \hspace{1cm} (1959)

Further, [341, §5.4]

$$e^A \succ 0 \quad \forall \ A \in \mathbb{S}^m$$  \hspace{1cm} (1960)

- For all square $A$ and integer $k$

$$\det^k A = \det A^k$$  \hspace{1cm} (1961)
### D.2. Tables of Gradients and Derivatives

#### D.2.1 algebraic

<table>
<thead>
<tr>
<th>Expression</th>
<th>Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nabla_x x = \nabla_x x^T = I \in \mathbb{R}^{k \times k}$</td>
<td>$\nabla X X = \nabla X X^T \triangleq I \in \mathbb{R}^{K \times L \times K \times L}$ (Identity)</td>
</tr>
<tr>
<td>$\nabla_x (Ax - b) = A^T$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x (x^T A - b^T) = A$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x (Ax - b)^T(Ax - b) = 2A^T(Ax - b)$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x^2 (Ax - b)^T(Ax - b) = 2A^T A$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x |Ax - b| = A^T(Ax - b)/|Ax - b|$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x z^T</td>
<td>Ax - b</td>
</tr>
<tr>
<td>$\nabla_x 1^T f(|Ax - b|) = A^T \delta \left( \left. \frac{df(y)}{dy} \right</td>
<td>_{y = |Ax - b|} \right) \text{sgn}(Ax - b)$</td>
</tr>
<tr>
<td>$\nabla_x (x^T A x + 2x^T B y + y^T C y) = (A + A^T) x + 2B y$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x (x + y)^T A (x + y) = (A + A^T)(x + y)$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x^2 (x^T A x + 2x^T B y + y^T C y) = A + A^T$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T x b = \nabla_x b^T X^T a = ab^T$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T X^2 b = X^T a b^T + ab^T X^T$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T X^{-1} b = -X^{-T} a b^T X^{-T}$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x (X^{-1})<em>{kl} = \partial X^{-1} \over \partial X^{-1}</em>{kl} = -X^{-1} c_k e_l^T X^{-1}$, (1893) (1957)</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T X^T X b = X(a b^T + b a^T)$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T X X^T b = (a b^T + b a^T) X$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T X X a = 2X a a^T$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T X X a = 2 a a^T x$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T Y X^T b = ba^T Y$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T Y X^T b = Y a b^T$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T Y X^T b = ab^T Y$</td>
<td></td>
</tr>
<tr>
<td>$\nabla_x a^T X^T Y b = Y a b^T$</td>
<td></td>
</tr>
</tbody>
</table>
algebraic continued

\[
\frac{d}{dt}(X + tY) = Y
\]
\[
\frac{d}{dt}B^T(X + tY)^{-1}A = -B^T(X + tY)^{-1}Y(X + tY)^{-1}A
\]
\[
\frac{d}{dt}B^T(X + tY)^{-T}A = -B^T(X + tY)^{-T}Y^T(X + tY)^{-T}A
\]
\[
\frac{d}{dt}B^T(X + tY)^{\mu}A = \ldots, \quad -1 \leq \mu \leq 1, \quad X, Y \in S^+_n
\]
\[
\frac{d^2}{dt^2}B^T(X + tY)^{-1}A = 2B^T(X + tY)^{-1}Y(X + tY)^{-1}Y(X + tY)^{-1}A
\]
\[
\frac{d^3}{dt^3}B^T(X + tY)^{-1}A = 6B^T(X + tY)^{-1}Y(X + tY)^{-1}Y(X + tY)^{-1}Y(X + tY)^{-1}A
\]
\[
\frac{d}{dt}((X + tY)^TA(X + tY)) = Y^TAX + X^TAY + 2tY^TAY
\]
\[
\frac{d^2}{dt^2}((X + tY)^TA(X + tY)) = 2Y^TAY
\]
\[
\frac{d^3}{dt^3}((X + tY)^TA(X + tY))^{-1} = -((X + tY)^TA(X + tY))^{-1}(Y^TAX + X^TAY + 2tY^TAY)((X + tY)^TA(X + tY))^{-1}
\]
\[
\frac{d^4}{dt^4}(X + tY)^A(X + tY)) = YAX + XAY + 2tYAY
\]
\[
\frac{d^2}{dt^2}((X + tY)^A(X + tY)) = 2YAY
\]

D.2.1.0.1 Exercise. Expand these tables.
Provide unfinished table entries indicated by \ldots throughout §D.2.

D.2.1.0.2 Exercise. \textit{log}. \hspace{1cm} (§D.1.7)
Find the first four terms of the Taylor series expansion for \( \log x \) about \( x = 1 \). Prove that \( \log x \leq x - 1 \); alternatively, plot the supporting hyperplane to the hypograph of \( \log x \) at \[
\begin{bmatrix}
x \\
\log x
\end{bmatrix} = \begin{bmatrix}
1 \\
0
\end{bmatrix}.
\]

D.2.2 trace Kronecker

\[
\nabla_{vecX} \text{tr}(AXBX^T) = \nabla_{vecX} \text{vec}(X)^T(B^T \otimes A) \text{vec} X = (B \otimes A^T + B^T \otimes A) \text{vec} X
\]
\[
\nabla_{vecX}^2 \text{tr}(AXBX^T) = \nabla_{vecX}^2 \text{vec}(X)^T(B^T \otimes A) \text{vec} X = B \otimes A^T + B^T \otimes A
\]
### D.2.3 trace

<table>
<thead>
<tr>
<th>Expression</th>
<th>Derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nabla_x \mu x = \mu I )</td>
<td>( \nabla_X \mu X = \nabla_X \mu \ ) ( X = \mu I )</td>
</tr>
<tr>
<td>( \nabla_x 1^T \delta(x)!^{-1} = \frac{d}{dx} !^{-1} = -x^{-2} )</td>
<td>( \nabla_X \text{tr}X^{-1} = -X^{-2} )</td>
</tr>
<tr>
<td>( \nabla_x 1^T \delta(x)!^{-1}y = -\delta(x)^{-2}y )</td>
<td>( \nabla_X \text{tr}(X^{-1}Y) = \nabla_X \text{tr}(YX^{-1}) = -X^{-T}YTX^{-T} )</td>
</tr>
<tr>
<td>( \frac{d}{dx} x^\mu = \mu x^{\mu - 1} )</td>
<td>( \nabla_X \text{tr}X^\mu = \mu X^{\mu - 1}, \quad X \in \mathbb{S}^M )</td>
</tr>
<tr>
<td>( \nabla_x (b - a^Tx)^{-1} = (b - a^Tx)^{-2}a )</td>
<td>( \nabla_X \text{tr}((B - AX)^{-1}) = ((B - AX)^{-2}A)^T )</td>
</tr>
<tr>
<td>( \nabla_x (b - a^Tx)^\mu = -\mu (b - a^Tx)^{\mu - 1}a )</td>
<td>( \nabla_X \text{tr}(X^TY) = \nabla_X \text{tr}(YXT) = \nabla_X \text{tr}(XY^T) = Y )</td>
</tr>
<tr>
<td>( \nabla_x xy = \nabla_x y^Tx = y )</td>
<td>( \nabla_X \text{tr}(AXBX^T) = \nabla_X \text{tr}(XBXA) = A^TB + AB )</td>
</tr>
<tr>
<td>( \nabla_X \text{tr}(AXBX^T) = \nabla_X \text{tr}(XBX^T)A = A^TXB^T + AXB )</td>
<td>( \nabla_X \text{tr}(AXBX) = \nabla_X \text{tr}(XBXA) = A^TXX^TB + B^TX^TA )</td>
</tr>
<tr>
<td>( \nabla_X \text{tr}(AXAXAX) = \nabla_X \text{tr}(XAXAXA) = 3(AXAXA)^T )</td>
<td>( \nabla_X \text{tr}(YX^k) = \nabla_X \text{tr}(X^kY) = \sum_{i=0}^{k-1} (X^iYX^{k-1-i})^T )</td>
</tr>
<tr>
<td>( \nabla_X \text{tr}(Y^TX^TY^T) = \nabla_X \text{tr}(X^TY^TXY) = 2YY^TX )</td>
<td>( \nabla_X \text{tr}(Y^TX^XY) = \nabla_X \text{tr}(XY^TXY^T) = 2XY^Y^T )</td>
</tr>
<tr>
<td>( \nabla_X \text{tr}((X + Y)^T(X + Y)) = 2(X + Y)^2 = \nabla_X |X + Y|^2_| )</td>
<td>( \nabla_X \text{tr}((X + Y)(X + Y)) = 2(X + Y)^T )</td>
</tr>
<tr>
<td>( \nabla_X \text{tr}(A^TXB) = \nabla_X \text{tr}(X^TAB^T) = AB^T )</td>
<td>( \nabla_X \text{tr}(A^TX^{-1}B) = \nabla_X \text{tr}(X^{-T}AB^T) = -X^{-T}AB^TX^{-T} )</td>
</tr>
<tr>
<td>( \nabla_X a^TXb = \nabla_X \text{tr}(ba^TX) = \nabla_X \text{tr}(Xba^T) = ab^T )</td>
<td>( \nabla_X b^TX^Ta = \nabla_X \text{tr}(X^Tba^T) = \nabla_X \text{tr}(ab^TX^T) = ab^T )</td>
</tr>
<tr>
<td>( \nabla_X a^TX^{-1}b = \nabla_X \text{tr}(X^{-T}ab^T) = -X^{-T}ab^TX^{-T} )</td>
<td>( \nabla_X a^TX^\mu b = ... )</td>
</tr>
</tbody>
</table>
trace continued

\[
\frac{d}{dt} \text{tr} g(X + tY) = \text{tr} \frac{d}{dt} g(X + tY) \tag{213, p.491}
\]

\[
\frac{d}{dt} \text{tr}(X + tY) = \text{tr} Y
\]

\[
\frac{d}{dt} \text{tr}^j(X + tY) = j \text{tr}^{j-1}(X + tY) \text{tr} Y
\]

\[
\frac{d}{dt} \text{tr}(X + tY)^j = j \text{tr}((X + tY)^{j-1}Y) \quad (\forall j)
\]

\[
\frac{d}{dt} \text{tr}((X + tY)Y) = \text{tr} Y^2
\]

\[
\frac{d}{dt} \text{tr}((X + tY)^kY) = \frac{d}{dt} \text{tr}(Y(X + tY)^k) = k \text{tr}((X + tY)^{k-1}Y^2) , \quad k \in \{0, 1, 2\}
\]

\[
\frac{d}{dt} \text{tr}((X + tY)^kY) = \frac{d}{dt} \text{tr}(Y(X + tY)^k) = k \sum_{i=0}^{k-1} (X + tY)^iY(X + tY)^{k-1-i} Y
\]

\[
\frac{d}{dt} \text{tr}((X + tY)^{-1}Y) = - \text{tr}((X + tY)^{-1}Y(X + tY)^{-1}Y)
\]

\[
\frac{d}{dt} \text{tr}(B^T(X + tY)^{-1}A) = - \text{tr}(B^T(X + tY)^{-1}Y(X + tY)^{-1}A)
\]

\[
\frac{d}{dt} \text{tr}(B^T(X + tY)^{-T}A) = - \text{tr}(B^T(X + tY)^{-T}Y^T(X + tY)^{-T}A)
\]

\[
\frac{d}{dt} \text{tr}(B^T(X + tY)^{-k}A) = \ldots , \quad k > 0
\]

\[
\frac{d}{dt} \text{tr}(B^T(X + tY)^{-\mu}A) = \ldots , \quad -1 \leq \mu \leq 1 , \quad X, Y \in \mathbb{S}^M
\]

\[
\frac{d^2}{dt^2} \text{tr}(B^T(X + tY)^{-1}A) = 2 \text{tr}(B^T(X + tY)^{-1}Y(X + tY)^{-1}Y(X + tY)^{-1}Y)
\]

\[
\frac{d}{dt} \text{tr}((X + tY)^T A(X + tY)) = \text{tr}(Y^TAX + X^TAY + 2t Y^TAY)
\]

\[
\frac{d^2}{dt^2} \text{tr}((X + tY)^T A(X + tY)) = 2 \text{tr}(Y^TAY)
\]

\[
\frac{d}{dt} \text{tr}(((X + tY)^T A(X + tY))^{-1})
\]

\[
= - \text{tr}(((X + tY)^T A(X + tY))^{-1}Y^TAX + X^TAY + 2t Y^TAY)((X + tY)^T A(X + tY))^{-1})
\]

\[
\frac{d}{dt} \text{tr}((X + tY)A(X + tY)) = \text{tr}(YAX + XAY + 2t YAY)
\]

\[
\frac{d^2}{dt^2} \text{tr}((X + tY)A(X + tY)) = 2 \text{tr}(YAY)
\]
### D.2.4 logarithmic determinant

$x>0$, det $X > 0$ on some neighborhood of $X$, and det($X + tY$) > $0$ on some open interval of $t$; otherwise, log( ) would be discontinuous. [85, p.75]

<table>
<thead>
<tr>
<th>$\frac{d}{dx} \log x$</th>
<th>$= x^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nabla_X \log \det X$</td>
<td>$= X^{-T}$</td>
</tr>
<tr>
<td>$\nabla_X^2 \log \det(X)_{kl}$</td>
<td>$= \frac{\partial X_{kl}}{\partial X_{kl}} = -(X^{-1}e_k e_l^T X^{-1})^T$</td>
</tr>
<tr>
<td>$\nabla_X \log \det X^{-1}$</td>
<td>$= -X^{-T}$</td>
</tr>
<tr>
<td>$\nabla_X \log \det \mu X$</td>
<td>$= \mu X^{-T}$</td>
</tr>
<tr>
<td>$\nabla_X \log \det \mu$</td>
<td>$= \mu X^{-T}$</td>
</tr>
<tr>
<td>$\nabla_X \log \det k X$</td>
<td>$= kX^{-T}$</td>
</tr>
<tr>
<td>$\nabla_X \log \det \mu (X + t Y)$</td>
<td>$= \mu (X + t Y)^{-T}$</td>
</tr>
<tr>
<td>$\nabla_X \log \det (AX + B)$</td>
<td>$= A^T (AX + B)^{-T}$</td>
</tr>
<tr>
<td>$\nabla_X \log \det (I \pm A^T X A)$</td>
<td>$= \pm A (I \pm A^T X A)^{-T} A^T$</td>
</tr>
<tr>
<td>$\nabla_X \log \det (X + t Y)^k$</td>
<td>$= \nabla_X \log \det^k (X + t Y) = k (X + t Y)^{-T}$</td>
</tr>
<tr>
<td>$\frac{d}{dt} \log \det (X + t Y)$</td>
<td>$= \text{tr} ((X + t Y)^{-1} Y)$</td>
</tr>
<tr>
<td>$\frac{d^2}{dt^2} \log \det (X + t Y)$</td>
<td>$= - \text{tr} ((X + t Y)^{-1} Y (X + t Y)^{-1} Y)$</td>
</tr>
<tr>
<td>$\frac{d}{dt} \log \det (X + t Y)^{-1}$</td>
<td>$= - \text{tr} ((X + t Y)^{-1} Y)$</td>
</tr>
<tr>
<td>$\frac{d^2}{dt^2} \log \det (X + t Y)^{-1}$</td>
<td>$= \text{tr} ((X + t Y)^{-1} Y (X + t Y)^{-1} Y)$</td>
</tr>
<tr>
<td>$\frac{d}{dt} \log \det (\delta (A(x + t y) + a)^2 + \mu I)$</td>
<td>$= \text{tr} \left( (\delta (A(x + t y) + a)^2 + \mu I)^{-1} 2 \delta (A(x + t y) + a) \delta (Ay) \right)$</td>
</tr>
</tbody>
</table>
D.2.5 determinant

\[
\nabla_X \det X = \nabla_X \det X^T = \det(X)X^{-T} \\
\nabla_X \det X^{-1} = -\det(X^{-1})X^{-T} = -\det(X)^{-1}X^{-T} \\
\nabla_X \det^\mu X = \mu \det^\mu(X)X^{-T} \\
\nabla_X \det X^\mu = \mu \det(X^\mu)X^{-T} \\
\nabla_X \det X^k = k \det^{k-1}(X)(\text{tr}(X)I - X^T), \quad X \in \mathbb{R}^{2\times2} \\
\nabla_X \det X^k = \nabla_X \det^k X = k \det(X^k)X^{-T} = k \det^k(X)X^{-T} \\
\nabla_X \det^\mu(X + t Y) = \mu \det^\mu(X + t Y)(X + t Y)^{-T} \\
\nabla_X \det(X + t Y)^k = \nabla_X \det^k(X + t Y) = k \det^k(X + t Y)(X + t Y)^{-T} \\
\[ \frac{d}{dt} \det(X + t Y) = \det(X + t Y) \text{tr}((X + t Y)^{-1}Y) \\
\frac{d^2}{dt^2} \det(X + t Y) = \det(X + t Y)(\text{tr}^2((X + t Y)^{-1}Y) - \text{tr}((X + t Y)^{-1}Y(X + t Y)^{-1}Y)) \\
\frac{d}{dt} \det(X + t Y)^{-1} = -\det(X + t Y)^{-1} \text{tr}((X + t Y)^{-1}Y) \\
\frac{d^2}{dt^2} \det(X + t Y)^{-1} = \det(X + t Y)^{-1}(\text{tr}^2((X + t Y)^{-1}Y) + \text{tr}((X + t Y)^{-1}Y(X + t Y)^{-1}Y)) \\
\frac{d}{dt} \det^\mu(X + t Y) = \mu \det^\mu(X + t Y) \text{tr}((X + t Y)^{-1}Y) \\
\]

D.2.6 logarithmic

Matrix logarithm.

\[
\frac{d}{dt} \log(X + t Y)^\mu = \mu Y(X + t Y)^{-1} = \mu(X + t Y)^{-1}Y, \quad XY = YX \\
\frac{d}{dt} \log(I - t Y)^\mu = -\mu Y(I - t Y)^{-1} = -\mu(I - t Y)^{-1}Y \quad [213, \text{ p.493}] \\
\]
D.2.7 exponential

Matrix exponential. [79, §3.6, §4.5] [341, §5.4]

\[ \nabla_X e^{(Y^TX)} = \nabla_X \det e^{Y^TX} = e^{\text{tr}(Y^TX)}Y \quad (\forall X, Y) \]

\[ \nabla_X \text{tr} e^{YX} = e^{Y^TX^TY^T} = Y^T e^{X^TY} \]

\[ \nabla_x 1^T e^{Ax} = A^T e^{Ax} \]

\[ \nabla_x 1^T e^{[Ax]} = A^T \delta(\text{sgn}(Ax)) e^{[Ax]} \quad (Ax)_i \neq 0 \]

\[ \nabla_x \log(1^T e^x) = \frac{1}{1^T e^x} e^x \]

\[ \nabla_x^2 \log(1^T e^x) = \frac{1}{1^T e^x} \left( \delta(e^x) - \frac{1}{1^T e^x} e^x e^x^T \right) \]

\[ \nabla_x \prod_{i=1}^k x_i^{\frac{1}{k}} = \frac{1}{k} \left( \prod_{i=1}^k x_i^{\frac{1}{k}} \right) \frac{1}{x} \]

\[ \nabla_x^2 \prod_{i=1}^k x_i^{\frac{1}{k}} = -\frac{1}{k} \left( \prod_{i=1}^k x_i^{\frac{1}{k}} \right) \left( \delta(x)^{-2} - \frac{1}{k} (1/x)(1/x)^T \right) \]

\[ \frac{d}{dt} e^{tY} = e^{tY} Y = Ye^{tY} \]

\[ \frac{d}{dt} e^{X+Y} = e^{X+Y} Y = Ye^{X+Y}, \quad XY = YX \]

\[ \frac{d^2}{dt^2} e^{X+Y} = e^{X+Y} Y^2 = Ye^{X+Y} Y = Y^2 e^{X+Y}, \quad XY = YX \]

\[ \frac{d^2}{dt^2} e^{\text{tr}(X+Y)} = e^{\text{tr}(X+Y)} \text{tr}(Y) \]
Appendix E

Projection

Rob Reiner’s “A Few Good Men” is one of those movies that tells you what it’s going to do, does it, and then tells you what it did. It doesn’t think the audience is very bright.

—Roger Ebert, 1992

For any $A \in \mathbb{R}^{m \times n}$, the pseudoinverse $[212, \S 7.3 \text{ prob.9}] \ [260, \S 6.12 \text{ prob.19}] \ [168, \S 5.5.4]$ $[341, \text{ App.A}]$

$$A^\dagger \triangleq \lim_{t \to 0^+} (A^\top A + t I)^{-1} A^\top = \lim_{t \to 0^+} A^\top t (AA^\top + t I)^{-1} \in \mathbb{R}^{n \times m}$$ (1962)

is a unique matrix from the optimal solution set to minimize $\|AX - I\|_F^2$ (§3.6.0.0.2). For any $t > 0$

$$I - A(A^\top A + t I)^{-1} A^\top = t (AA^\top + t I)^{-1}$$ (1963)

Equivalently, pseudoinverse $A^\dagger$ is that unique matrix satisfying the Moore-Penrose conditions: $[214, \S 1.3] \ [405]$

1. $AA^\dagger A = A$
2. $A^\dagger AA^\dagger = A^\dagger$
3. $(AA^\dagger)^\top = AA^\dagger$
4. $(A^\dagger A)^\top = A^\dagger A$

which are necessary and sufficient to establish the pseudoinverse whose principal action is to injectively map $\mathcal{R}(A)$ onto $\mathcal{R}(A^\top)$ (Figure 173). Conditions 1 and 3 are necessary and sufficient for $AA^\dagger$ to be the orthogonal projector on $\mathcal{R}(A)$, while conditions 2 and 4 hold iff $A^\dagger A$ is the orthogonal projector on $\mathcal{R}(A^\top)$.

Range and nullspace of the pseudoinverse $[277] \ [338, \S III.1 \text{ exer.1}]$

$$\mathcal{R}(A^\dagger) = \mathcal{R}(A^\top), \quad \mathcal{R}(A^\top A^\dagger) = \mathcal{R}(A)$$ (1964)

$$\mathcal{N}(A^\dagger) = \mathcal{N}(A^\top), \quad \mathcal{N}(A^\top A^\dagger) = \mathcal{N}(A)$$ (1965)

can be derived by singular value decomposition (§A.6).

Dattorro, Convex Optimization † Euclidean Distance Geometry 2e, Meijo, v2015.04.11. 601
The following relations reliably hold without qualification:

a. \( A^T \dagger = A^\dagger T \)

b. \( A^\dagger \dagger = A \)

c. \( (AA^T)^\dagger = A^T A^\dagger \)

d. \( (A^T A)^\dagger = A^T A^\dagger \)

e. \( (AA^\dagger)^\dagger = AA^\dagger \)

f. \( (A^\dagger A)^\dagger = A^\dagger A \)

Yet for arbitrary \( A, B \) it is generally true that \( (AB)^\dagger \neq B^\dagger A^\dagger \):

**E.0.0.0.1 Theorem.** Pseudoinverse of product.

For \( A \in \mathbb{R}^{m \times n} \) and \( B \in \mathbb{R}^{n \times k} \)

\[
(AB)^\dagger = B^\dagger A^\dagger \tag{1966}
\]

if and only if

\[
\mathcal{R}(A^T AB) \subseteq \mathcal{R}(B) \quad \text{and} \quad \mathcal{R}(BB^T A^T) \subseteq \mathcal{R}(A^T) \tag{1967}
\]

Pseudoinverse of normalized vector \( u \) is the vector transpose. Otherwise,

\[
u^\dagger = \frac{u^T}{\|u\|^2} \tag{1968}
\]

\( U^\dagger = U^T \) for orthonormal (including the orthogonal) matrices \( U \). So, for orthonormal matrices \( U, Q \) and arbitrary \( A \)

\[
(UAQ^T)^\dagger = QA^T U^T \tag{1969}
\]
E.0.0.0.2  Exercise.  **Kronecker inverse.**
Prove:

\[(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger\]  \hspace{1em} (1970)

\[\nabla\]

E.0.1  **Logical deductions**

When \(A\) is invertible, \(A^\dagger = A^{-1}\); so \(A^\dagger A = AA^\dagger = I\).

Otherwise, for \(A \in \mathbb{R}^{m \times n}\) \[161, \S 5.3.3.1\] \[251, \S 7\] \[306\]

- g.  \(A^\dagger A = I\), \(A^\dagger = (A^T A)^{-1} A^T\), \(\text{rank } A = n\)
- h.  \(AA^\dagger = I\), \(A^\dagger = A^T (AA^T)^{-1}\), \(\text{rank } A = m\)
- i.  \(A^\dagger \omega = \omega\), \(\omega \in \mathcal{R}(A^T)\)
- j.  \(AA^\dagger v = v\), \(v \in \mathcal{R}(A)\)
- k.  \(A^\dagger A = AA^\dagger\)
- l.  \(A^k \dagger = A^{k^\dagger}\), \(k\) an integer, \(A\) normal

Equivalent to the corresponding Moore-Penrose condition:

1.  \(A^T = A^T A A^\dagger\) or \(A^T = A^\dagger A A^T\)
2.  \(A^{\dagger T} = A^{\dagger T} A^\dagger A\) or \(A^{\dagger T} = AA^\dagger A^{\dagger T}\)

When \(A\) is symmetric, \(A^\dagger\) is symmetric and (§A.6)

\[A \succeq 0 \iff A^\dagger \succeq 0\]  \hspace{1em} (1971)

E.0.1.0.1  **Example.**  **Solution to classical linear equation** \(Ax = b\).

In §2.5.1.1, the solution set to matrix equation \(Ax = b\) was represented as an intersection of hyperplanes. Regardless of rank of \(A\) or its shape (fat or skinny), interpretation as a hyperplane intersection describing a possibly empty affine set generally holds. If matrix \(A\) is rank deficient or fat, there is an infinity of solutions \(x\) when \(b \in \mathcal{R}(A)\). A unique solution occurs when the hyperplanes intersect at a single point.

Given arbitrary matrix \(A\) (of any rank and dimension) and vector \(b\) not necessarily in \(\mathcal{R}(A)\), we wish to find a best solution \(x^*\) to

\[Ax \approx b\]  \hspace{1em} (1972)

in a Euclidean sense by solving an algebraic expression for orthogonal projection of \(b\) on \(\mathcal{R}(A)\)

\[\text{minimize } \|Ax - b\|^2\]  \hspace{1em} (1973)

Necessary and sufficient condition for optimal solution to this unconstrained optimization is the so-called **normal equation** that results from zeroing the convex objective’s gradient: (§D.2.1)

\[A^T A x = A^T b\]  \hspace{1em} (1974)

*normal* because error vector \(b - Ax\) is perpendicular to \(\mathcal{R}(A)\); *id est*, \(A^T (b - Ax) = 0\).

Given any matrix \(A\) and any vector \(b\), the normal equation is solvable exactly; always so, because \(\mathcal{R}(A^\dagger A) = \mathcal{R}(A^T)\) and \(A^T b \in \mathcal{R}(A^T)\).
Given particular $\tilde{x} \in \mathcal{R}(A^T)$ solving (1974), then (Figure 173) it is necessarily unique and $\tilde{x} = x^* = A^1b$. When $A$ is skinny- or-square full-rank, normal equation (1974) can be solved exactly by inversion:

$$x^* = (A^TA)^{-1} A^T b \equiv A^1 b$$  \hspace{1cm} (1975)

For matrix $A$ of arbitrary rank and shape, on the other hand, $A^TA$ might not be invertible. Yet the normal equation can always be solved exactly by: (1962)

$$x^* = \lim_{t \to 0^+} (A^TA + t I)^{-1} A^T b = A^1b$$ \hspace{1cm} (1976)

invertible for any positive value of $t$ by (1533). The exact inversion (1975) and this pseudoinverse solution (1976) each solve a limited regularization

$$\lim_{t \to 0^+} \minimize_x \|Ax - b\|^2 + t \|x\|^2 \equiv \lim_{t \to 0^+} \minimize_x \left\| \begin{bmatrix} A \\ \sqrt{t}I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|^2$$ \hspace{1cm} (1977)

simultaneously providing least squares solution to (1973) and the classical least norm solution$^E.1$ [341, App.A.4] [50] (confer §3.2.0.1)

$$\begin{array}{l}
\arg \minimize_x \|x\|^2 \\
\text{subject to } Ax = AA^1b 
\end{array}$$ \hspace{1cm} (1978)

where $AA^1b$ is the orthogonal projection of vector $b$ on $\mathcal{R}(A)$. Least norm solution can be interpreted as orthogonal projection of the origin $\mathbf{0}$ on affine subset $\mathcal{A} = \{x \mid Ax = AA^1b\}$; (§E.5.0.0.5, §E.5.0.0.6)

$$\begin{array}{l}
\arg \minimize_x \|x - \mathbf{0}\|^2 \\
\text{subject to } x \in \mathcal{A} 
\end{array}$$ \hspace{1cm} (1979)

equivalently, maximization of the Euclidean ball until it kisses $\mathcal{A}$; rather, $\arg \text{dist}(\mathbf{0}, \mathcal{A})$.

\end{proof}

\section{E.1 Idempotent matrices}

Projection matrices are square and defined by \textit{idempotence}, $P^2 = P$; [341, §2.6] [214, §1.3] equivalent to the condition, $P$ be diagonalizable [212, §3.3 prob.3] with eigenvalues $\phi_i \in \{0, 1\}$. [425, §4.1 thm.4.1] Idempotent matrices are not necessarily symmetric. The transpose of an idempotent matrix remains idempotent; $P^TP = P^T$. Solely excepting $P = I$, all projection matrices are neither orthogonal (§B.5) or invertible. [341, §3.4] The collection of all projection matrices of particular dimension does not form a convex set.

Suppose we wish to project nonorthogonally (\textit{obliquely}) on the range of any particular matrix $A \in \mathbb{R}^{m \times n}$. All idempotent matrices projecting nonorthogonally on $\mathcal{R}(A)$ may be expressed:

$$P = A(A^\dagger + BZ^T) \in \mathbb{R}^{m \times m}$$  \hspace{1cm} (1980)

\begin{proof}
This means: optimal solutions of lesser norm than the so-called \textit{least norm} solution (1978) can be obtained (at expense of approximation $Ax \approx b$ hence, of perpendicularity) by ignoring the limiting operation and introducing finite positive values of $t$ into (1977).

$^E.1$
E.1. IDEMPOTENT MATRICES

where $\mathcal{R}(P) = \mathcal{R}(A)$, $B \in \mathbb{R}^{n \times k}$ for positive integer $k$ is arbitrary, and $Z \in \mathbb{R}^{m \times k}$ is any matrix whose range is in $\mathcal{N}(A^T)$; id est,

$$A^T Z = A^T \mathbf{1} = 0$$  \hspace{1cm} (1981)

Evidently, the collection of nonorthogonal projectors projecting on $\mathcal{R}(A)$ is an affine subset

$$\mathcal{P}_k = \left\{ A(A^T + BZ^T) \mid B \in \mathbb{R}^{n \times k} \right\}$$ \hspace{1cm} (1982)

When matrix $A$ in (1980) is skinny full-rank ($A^T A = I$) or has orthonormal columns ($A^T A = I$), either property leads to a biorthogonal characterization of nonorthogonal projection:

**E.1.1 Biorthogonal characterization of projector**

Any nonorthogonal projector $P^2 = P \in \mathbb{R}^{m \times m}$ projecting on nontrivial $\mathcal{R}(U)$ can be defined by a biorthogonality condition $Q^T U = I$; the biorthogonal decomposition of $P$ being (confer (1980))

$$P = UQ^T, \quad Q^T U = I$$ \hspace{1cm} (1983)

where (§B.1.1.1)

$$\mathcal{R}(P) = \mathcal{R}(U), \quad \mathcal{N}(P) = \mathcal{N}(Q^T)$$ \hspace{1cm} (1984)

and where generally (confer (2010))

$$P^T \neq P, \quad P^T \neq P, \quad \|P\|_2 \neq 1, \quad P \neq 0$$ \hspace{1cm} (1985)

and $P$ is not nonexpansive (2011) (2190).  

(\Leftarrow) To verify assertion (1983) we observe: because idempotent matrices are diagonalizable (§A.5), [212, §3.3 prob.3] they must have the form (1629)

$$P = S \Phi S^{-1} = \sum_{i=1}^{m} \phi_i s_i w_i^T = \sum_{i=1}^{k} s_i w_i^T$$ \hspace{1cm} (1986)

**E.2 Proof.** $\mathcal{R}(P) \subseteq \mathcal{R}(A)$ is obvious [341, §3.6]. By (141) and (142),

$$\mathcal{R}(A^T + BZ^T) = \left\{ (A^T + BZ^T) y \mid y \in \mathbb{R}^m \right\}$$

$$\supseteq \left\{ (A^T + BZ^T) y \mid y \in \mathcal{R}(A) \right\} = \mathcal{R}(A^T)$$

$$\mathcal{R}(P) = \left\{ A(A^T + BZ^T) y \mid y \in \mathbb{R}^m \right\}$$

$$\supseteq \left\{ A(A^T + BZ^T) y \mid (A^T + BZ^T) y \in \mathcal{R}(A^T) \right\} = \mathcal{R}(A)$$

**E.3 A ← U, A^T + BZ^T ← Q^T**

**E.4 Proof.** Obviously, $\mathcal{R}(P) \subseteq \mathcal{R}(U)$. Because $Q^T U = I$

$$\mathcal{R}(P) = \left\{ UQ^T x \mid x \in \mathbb{R}^m \right\}$$

$$\supseteq \left\{ UQ^T U y \mid y \in \mathbb{R}^k \right\} = \mathcal{R}(U)$$

**E.5** Orthonormal decomposition (2007) (confer §E.3.4) is a special case of biorthogonal decomposition (1983) characterized by (2010). So, these characteristics (1985) are not necessary conditions for biorthogonality.
that is a sum of \( k = \text{rank} P \) independent projector dyads (idempotent dyads, §B.1.1, §E.6.2.1) where \( \phi_i \in \{0, 1\} \) are the eigenvalues of \( P \) [425, §4.1 thm.4.1] in diagonal matrix \( \Phi \in \mathbb{R}^{m \times m} \) arranged in nonincreasing order, and where \( s_i, w_i \in \mathbb{R}^m \) are the right- and left-eigenvectors of \( P \), respectively, which are independent and real.\( ^\text{E.6} \) Therefore

\[
U \triangleq S(:, 1:k) = [s_1 \cdots s_k] \in \mathbb{R}^{m \times k}
\]

is the full-rank matrix \( S \in \mathbb{R}^{m \times m} \) having \( m - k \) columns (corresponding to 0 eigenvalues) truncated, while

\[
Q^T \triangleq S^{-1}(1:k,:) = \begin{bmatrix} w_1^T \\ \vdots \\ w_k^T \end{bmatrix} \in \mathbb{R}^{k \times m}
\]

is matrix \( S^{-1} \) having the corresponding \( m - k \) rows truncated. By the 0 eigenvalues theorem (§A.7.3.0.1), \( \mathcal{R}(U) = \mathcal{R}(P), \mathcal{N}(Q) = \mathcal{R}(P^T) \), and

\[
\begin{align*}
\mathcal{R}(P) &= \text{span} \{ s_i \mid \phi_i = 1 \ \forall i \} \\
\mathcal{N}(P) &= \text{span} \{ s_i \mid \phi_i = 0 \ \forall i \} \\
\mathcal{R}(P^T) &= \text{span} \{ w_i \mid \phi_i = 1 \ \forall i \} \\
\mathcal{N}(P^T) &= \text{span} \{ w_i \mid \phi_i = 0 \ \forall i \}
\end{align*}
\]

Thus biorthogonality \( Q^TU = I \) is a necessary condition for idempotence, and so the collection of nonorthogonal projectors projecting on \( \mathcal{R}(U) \) is the affine subset \( \mathcal{P}_k = UQ_k^T \) where \( Q_k = \{ Q \mid Q^TU = I, Q \in \mathbb{R}^{m \times k} \} \).

(\( \Rightarrow \)) Biorthogonality is a sufficient condition for idempotence;

\[
P^2 = \sum_{i=1}^{k} s_i w_i^T \sum_{j=1}^{k} s_j w_j^T = P
\]

\( \text{id est, if the cross-products are annihilated, then } P^2 = P. \)

Nonorthogonal projection of \( x \) on \( \mathcal{R}(P) \) has expression like a biorthogonal expansion,

\[
P x = U Q^T x = \sum_{i=1}^{k} w_i^T x s_i
\]

When the domain is restricted to range of \( P \), say \( x = U \xi \) for \( \xi \in \mathbb{R}^k \), then \( x = P x = U Q^T U \xi = U \xi \) and expansion is unique because the eigenvectors are linearly independent. Otherwise, any component of \( x \) in \( \mathcal{N}(P) = \mathcal{N}(Q^T) \) will be annihilated.

Direction of nonorthogonal projection is orthogonal to \( \mathcal{R}(Q) \) if \( \triangledown (Q^TU = I) \); \( \text{id est, for } P x \in \mathcal{R}(U) \) (confer (2001))

\[
P x - x \perp \mathcal{R}(Q) \text{ in } \mathbb{R}^m
\]

\( ^\text{E.6} \) Eigenvectors of a real matrix corresponding to real eigenvalues must be real. (§A.5.0.0.1)
Figure 174: Nonorthogonal projection of $x \in \mathbb{R}^3$ on $\mathcal{R}(U) = \mathbb{R}^2$ under biorthogonality condition; *id est*, $P \mathbf{x} = U Q^T \mathbf{x}$ such that $Q^T U = I$. Any point along imaginary line $\mathcal{T}$ connecting $x$ to $P \mathbf{x}$ will be projected nonorthogonally on $P \mathbf{x}$ with respect to horizontal plane constituting $\mathbb{R}^2 = \text{aff cone}(U)$ in this example. Extreme directions of cone($U$) correspond to two columns of $U$; likewise for cone($Q$). For purpose of illustration, we truncate each conic hull by truncating coefficients of conic combination at unity. Conic hull cone($Q$) is headed upward at an angle, out of plane of page. Nonorthogonal projection would fail were $\mathcal{N}(Q^T)$ in $\mathcal{R}(U)$ (were $\mathcal{T}$ parallel to a line in $\mathcal{R}(U)$).
E.1.1.0.1 Example. Illustration of nonorthogonal projector.

Figure 174 shows cone($U$), conic hull of the columns of

$$U = \begin{bmatrix} 1 & 1 \\ -0.5 & 0.3 \\ 0 & 0 \end{bmatrix}$$  \hspace{1cm} (1993)

from nonorthogonal projector $P = UQ^T$. Matrix $U$ has a limitless number of left inverses because $\mathcal{N}(U^T)$ is nontrivial. Similarly depicted is left inverse $Q^T$ from (1980)

$$Q = U^T + ZB^T = \begin{bmatrix} 0.3750 & 0.6250 \\ -1.2500 & 1.2500 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$  \hspace{1cm} (1994)

where $Z \in \mathcal{N}(U^T)$ and matrix $B$ is selected arbitrarily; id est, $Q^TU = I$ because $U$ is full-rank.

Direction of projection on $\mathcal{R}(U)$ is orthogonal to $\mathcal{R}(Q)$. Any point along line $T$ in the figure, for example, will have the same projection. Were matrix $Z$ instead equal to 0, then cone($Q$) would become the relative dual to cone($U$) (sharing the same affine hull; §2.13.8, confer Figure 58a). In that case, projection $Px = UU^Tx$ of $x$ on $\mathcal{R}(U)$ becomes orthogonal projection (and unique minimum-distance). ⊡

E.1.2 Idempotence summary

Nonorthogonal subspace-projector $P$ is a (convex) linear operator defined by idempotence or biorthogonal decomposition (1983), but characterized not by symmetry or positive semidefiniteness or nonexpansivity (2011).

E.2 $I - P$, Projection on algebraic complement

It follows from the diagonalizability of idempotent matrices that $I - P$ must also be a projection matrix because it too is idempotent, and because it may be expressed

$$I - P = S(I - \Phi)S^{-1} = \sum_{i=1}^{m} (1 - \phi_i) s_i w_i^T$$  \hspace{1cm} (1995)

where $(1 - \phi_i) \in \{1, 0\}$ are the eigenvalues of $I - P$ (1534) whose eigenvectors $s_i, w_i$ are identical to those of $P$ in (1986). A consequence of that complementary relationship of eigenvalues is the fact, [353, §2] [348, §2] for subspace projector $P = P^2 \in \mathbb{R}^{m \times m}$

$$\mathcal{R}(P) = \text{span} \{ s_i \mid \phi_i = 1 \ \forall i \} = \text{span} \{ s_i \mid (1 - \phi_i) = 0 \ \forall i \} = \mathcal{N}(I - P)$$

$$\mathcal{N}(P) = \text{span} \{ s_i \mid \phi_i = 0 \ \forall i \} = \text{span} \{ s_i \mid (1 - \phi_i) = 1 \ \forall i \} = \mathcal{R}(I - P)$$

$$\mathcal{R}(P^T) = \text{span} \{ w_i \mid \phi_i = 1 \ \forall i \} = \text{span} \{ w_i \mid (1 - \phi_i) = 0 \ \forall i \} = \mathcal{N}(I - P^T)$$

$$\mathcal{N}(P^T) = \text{span} \{ w_i \mid \phi_i = 0 \ \forall i \} = \text{span} \{ w_i \mid (1 - \phi_i) = 1 \ \forall i \} = \mathcal{R}(I - P^T)$$  \hspace{1cm} (1996)
that is easy to see from (1986) and (1995). Idempotent $I - P$ therefore projects vectors on its range: $\mathcal{N}(P)$. Because all eigenvectors of a real idempotent matrix are real and independent, the algebraic complement of $\mathcal{R}(P)$ [237, §3.3] is equivalent to $\mathcal{N}(P)$. 

$$\mathcal{R}(P) \oplus \mathcal{N}(P) = \mathcal{R}(P^T) \oplus \mathcal{N}(P^T) = \mathcal{R}(P^T) \oplus \mathcal{N}(P) = \mathcal{R}(P) \oplus \mathcal{N}(P^T) = \mathbb{R}^m$$

because $\mathcal{R}(P) \oplus \mathcal{R}(I - P) = \mathbb{R}^m$. For idempotent $P \in \mathbb{R}^{m \times m}$, consequently,

$$\text{rank } P + \text{rank}(I - P) = m \quad (1998)$$

E.2.0.0.1 Theorem. Rank/Trace. \[425, \S 4.1 \text{prob.9} \] (confer (2015))

$$P^2 = P \iff \text{rank } P = \text{tr } P \quad \text{and} \quad \text{rank}(I - P) = \text{tr}(I - P) \quad (1999)$$

E.2.1 Universal projector characteristic

Although projection is not necessarily orthogonal and $\mathcal{R}(P) \not\perp \mathcal{R}(I - P)$ in general, still for any projector $P$ and any $x \in \mathbb{R}^m$

$$Px + (I - P)x = x \quad (2000)$$

must hold where $\mathcal{R}(I - P) = \mathcal{N}(P)$ is the algebraic complement of $\mathcal{R}(P)$. Algebraic complement of closed convex cone $\mathcal{K}$, for example, is the negative dual cone $-\mathcal{K}^*$. \[(2121)\] (Figure 178)

E.3 Symmetric idempotent matrices

When idempotent matrix $P$ is symmetric, $P$ is an orthogonal projector. In other words, the direction of projection of point $x \in \mathbb{R}^m$ on subspace $\mathcal{R}(P)$ is orthogonal to $\mathcal{R}(P)$; \textit{id est}, for $P^2 = P \in \mathbb{S}^m$ and projection $Px \in \mathcal{R}(P)$

$$Px - x \perp \mathcal{R}(P) \text{ in } \mathbb{R}^m \quad (2001)$$

(confer (1992)) Perpendicularity is a necessary and sufficient condition for orthogonal projection on a subspace. \[115, \S 4.9\]

A condition equivalent to (2001) is: Norm of direction $x - Px$ is the infimum over all nonorthogonal projections of $x$ on $\mathcal{R}(P)$; \[260, \S 3.3\] for $P^2 = P \in \mathbb{S}^m$, $\mathcal{R}(P) = \mathcal{R}(A)$, matrices $A, B, Z$ and positive integer $k$ as defined for (1980), and given $x \in \mathbb{R}^m$

$$\|x - Px\|_2 = \|x - AA^T x\|_2 = \inf_{B \in \mathbb{R}^{n\times k}} \|x - A(A^T + BZ^T)x\|_2 = \text{dist}(x, \mathcal{R}(P)) \quad (2002)$$

The infimum is attained for $\mathcal{R}(B) \subseteq \mathcal{N}(A)$ over any affine subset of nonorthogonal projectors (1982) indexed by $k$.

\[E.7\] The same phenomenon occurs with symmetric (nonidempotent) matrices, for example. When summands in $A \oplus B = \mathbb{R}^m$ are orthogonal vector spaces, the algebraic complement is the orthogonal complement.
APPENDIX E. PROJECTION

Proof is straightforward: The vector 2-norm is a convex function. Setting gradient of the norm-square to 0, applying §D.2,

\[
(A^T A B Z^T - A^T (I - AA^\dagger)) x x^T A = 0
\]

(2003)

\[
A^T A B Z^T x x^T A = 0
\]

because \( A^T = A^T A A^\dagger \). Projector \( P = AA^\dagger \) is therefore unique; the minimum-distance projector is the orthogonal projector, and vice versa. ♦

We get \( P = AA^\dagger \) so this projection matrix must be symmetric. Then for any matrix \( A \in \mathbb{R}^{m \times n} \), symmetric idempotent \( P \) projects a given vector \( x \) in \( \mathbb{R}^n \) orthogonally on \( \mathcal{R}(A) \). Under either condition (2001) or (2002), the projection \( Px \) is unique minimum-distance; for subspaces, perpendicularity and minimum-distance conditions are equivalent.

E.3.1 Four subspaces

We summarize the orthogonal projectors projecting on the four fundamental subspaces: for \( A \in \mathbb{R}^{m \times n} \)

\[
\begin{align*}
A^\dagger A & : \mathbb{R}^n \text{ on } (\mathcal{R}(A^\dagger A) = \mathcal{R}(A^T)) \\
AA^\dagger & : \mathbb{R}^m \text{ on } (\mathcal{R}(AA^\dagger) = \mathcal{R}(A)) \\
I - A^\dagger A & : \mathbb{R}^n \text{ on } (\mathcal{R}(I - A^\dagger A) = \mathcal{N}(A)) \\
I - AA^\dagger & : \mathbb{R}^m \text{ on } (\mathcal{R}(I - AA^\dagger) = \mathcal{N}(A^T))
\end{align*}
\]

(2004)

Given a known subspace, matrix \( A \) is neither unique or necessarily full-rank. Despite that, a basis for each fundamental subspace comprises the linearly independent column vectors from its associated symmetric projection matrix:

\[
\begin{align*}
\text{basis } \mathcal{R}(A^T) & \subseteq A^\dagger A \subseteq \mathcal{R}(A^T) \\
\text{basis } \mathcal{R}(A) & \subseteq AA^\dagger \subseteq \mathcal{R}(A) \\
\text{basis } \mathcal{N}(A) & \subseteq I - A^\dagger A \subseteq \mathcal{N}(A) \\
\text{basis } \mathcal{N}(A^T) & \subseteq I - AA^\dagger \subseteq \mathcal{N}(A^T)
\end{align*}
\]

(2005)

For completeness,\( \text{E.8 (1996)} \)

\[
\begin{align*}
\mathcal{N}(A^\dagger A) & = \mathcal{N}(A) \\
\mathcal{N}(AA^\dagger) & = \mathcal{N}(A^T) \\
\mathcal{N}(I - A^\dagger A) & = \mathcal{R}(A^T) \\
\mathcal{N}(I - AA^\dagger) & = \mathcal{R}(A)
\end{align*}
\]

(2006)

\( \text{E.8 Proof} \) is by singular value decomposition (§A.6.2); \( \mathcal{N}(A^\dagger A) \subseteq \mathcal{N}(A) \) is obvious. Conversely, suppose \( A^\dagger A x = 0 \). Then \( x^\dagger A^\dagger A x = x^\dagger Q Q^\dagger x = \|Q^\dagger x\|^2 = 0 \) where \( A = U \Sigma Q^\dagger \) is the subcompact singular value decomposition. Because \( \mathcal{R}(Q) = \mathcal{R}(A^T) \), then \( x \in \mathcal{N}(A) \) which implies \( \mathcal{N}(A^\dagger A) \supseteq \mathcal{N}(A) \). \( \therefore \mathcal{N}(A^\dagger A) = \mathcal{N}(A) \). ♦
E.3. SYMMETRIC IDEMPOTENT MATRICES

E.3.2 Orthogonal characterization of projector

Any symmetric projector $P^2 = P \in \mathbb{S}^m$ projecting on nontrivial $\mathcal{R}(Q)$ can be defined by the orthonormality condition $Q^TQ = I$. When skinny matrix $Q \in \mathbb{R}^{m \times k}$ has orthonormal columns, then $Q^\dagger = Q^T$ by the Moore-Penrose conditions. Hence, any $P$ having an orthonormal decomposition (§E.3.4)

$$P = QQ^T, \quad Q^TQ = I$$

(2007)

where [341, §3.3] (1691)

$$\mathcal{R}(P) = \mathcal{R}(Q), \quad \mathcal{N}(P) = \mathcal{N}(Q^T)$$

(2008)

is an orthogonal projector projecting on $\mathcal{R}(Q)$ having, for $P \in \mathcal{R}(Q)$ (confer (1992))

$$Px - x \perp \mathcal{R}(Q) \text{ in } \mathbb{R}^m$$

(2009)

From (2007), orthogonal projector $P$ is obviously positive semidefinite (§A.3.1.0.6); necessarily, (confer (1985))

$$P^T = P, \quad P^\dagger = P, \quad \|P\|_2 = 1, \quad P \succeq 0$$

(2010)

and $\|Px\| = \|QQ^Tx\| = \|Q^Tx\|$ because $\|Qy\| = \|y\| \forall y \in \mathbb{R}^k$. All orthogonal projectors are therefore nonexpansive because

$$\sqrt{(Px, x)} = \|Px\| = \|Q^Tx\| \leq \|x\| \forall x \in \mathbb{R}^m$$

(2011)

the Bessel inequality, [115] [237] with equality when $x \in \mathcal{R}(Q)$.

From the diagonalization of idempotent matrices (1986) on page 605

$$P = S\Phi S^T = \sum_{i=1}^{m} \phi_i s_i s_i^T = \sum_{i=1}^{k \leq m} s_i s_i^T$$

(2012)

orthogonal projection of point $x$ on $\mathcal{R}(P)$ has expression like an orthogonal expansion [115, §4.10]

$$Px = QQ^Tx = \sum_{i=1}^{k} s_i^Tx s_i$$

(2013)

where

$$Q = S(:, 1: k) = [s_1 \cdots s_k] \in \mathbb{R}^{m \times k}$$

(2014)

and where the $s_i$ [sic] are orthonormal eigenvectors of symmetric idempotent $P$. When the domain is restricted to range of $P$, say $x = Q\xi$ for $\xi \in \mathbb{R}^k$, then $x = Px = QQ^TQ\xi = Q\xi$ and expansion is unique. Otherwise, any component of $x$ in $\mathcal{N}(Q^T)$ will be annihilated.
E.3.2.0.1 Theorem. Symmetric rank/trace. (confer (1999) (1538))

\[ P^T = P, \quad P^2 = P \]
\[ \iff \quad \text{rank} P = \text{tr} P = \|P\|_F^2 \quad \text{and} \quad \text{rank}(I - P) = \text{tr}(I - P) = \|I - P\|_F^2 \quad (2015) \]

**Proof.** We take, as given, Theorem E.2.0.0.1 establishing idempotence. We have left only to show \( \text{tr} P = \|P\|_F^2 \Rightarrow P^T = P \), established in [425, §7.1].

E.3.3 Summary, symmetric idempotent

(confer §E.1.2) Orthogonal projector \( P \) is a (convex) linear operator defined [209, §A.3.1] by idempotence and symmetry, and characterized by positive semidefiniteness and nonexpansivity. The algebraic complement (§E.2) to \( \mathcal{R}(P) \) becomes the orthogonal complement \( \mathcal{R}(I - P) \); id est, \( \mathcal{R}(P) \perp \mathcal{R}(I - P) \).

E.3.4 Orthonormal decomposition

When \( Z = 0 \) in the general nonorthogonal projector \( A(A^\dagger + BZ^T) \) (1980), an orthogonal projector results (for any matrix \( A \)) characterized principally by idempotence and symmetry. Any real orthogonal projector may, in fact, be represented by an orthonormal decomposition such as (2007). [214, §1 prob.42]

To verify that assertion for the four fundamental subspaces (2004), we need only to express \( A \) by subcompact singular value decomposition (§A.6.2): From pseudoinverse (1658) of \( A = U\Sigma Q^T \in \mathbb{R}^{m \times n} \)

\[ AA^\dagger = U\Sigma \Sigma^\dagger U^T = UU^T, \quad A^\dagger A = Q\Sigma^\dagger \Sigma Q^T = QQ^T \]
\[ I - AA^\dagger = I - UU^T = U\perp U^\perp T, \quad I - A^\dagger A = I - QQ^T = Q\perp Q^\perp T \quad (2016) \]

where \( U\perp \in \mathbb{R}^{m \times m - \text{rank} A} \) holds columnar an orthonormal basis for the orthogonal complement of \( \mathcal{R}(U) \), and likewise for \( Q\perp \in \mathbb{R}^{n \times n - \text{rank} A} \). Existence of an orthonormal decomposition is sufficient to establish idempotence and symmetry of an orthogonal projector (2007).

E.3.5 Unifying trait of all projectors: direction

Whereas nonorthogonal projectors possess only a biorthogonal decomposition (§E.1.1), relation (2016) shows: orthogonal projectors simultaneously possess a biorthogonal decomposition \( (AA^\dagger \text{ whence } P_x = AA^\dagger x) \) and an orthonormal decomposition \( (UU^T \text{ whence } P_x = UU^T x) \). Orthogonal projection of a point is unique but its expansion is not; e.g, \( A \) can have dependent columns.
E.3. SYMMETRIC IDEMPOTENT MATRICES

E.3.5.1 orthogonal projector, orthonormal decomposition

Consider orthogonal expansion of \( x \in \mathbb{R}^n \):

\[
x = UU^T x = \sum_{i=1}^{n} u_i u_i^T x
\]

(2017)

a sum of one-dimensional orthogonal projections (§E.6.3) where

\[
U \triangleq [u_1 \cdots u_n] \quad \text{and} \quad U^T U = I
\]

(2018)

and where the subspace projector has two expressions: (2016)

\[
AA^\dagger \triangleq UU^T
\]

(2019)

where \( A \in \mathbb{R}^{m \times n} \) has rank \( n \). The direction of projection of \( x \) on \( u_j \) for some \( j \in \{1 \ldots n\} \), for example, is orthogonal to \( u_j \) but parallel to a vector in the span of all remaining vectors constituting the columns of \( U \);

\[
u_j^T(u_j u_j^T x - x) = 0
\]

\[
u_j u_j^T x - x = u_j u_j^T x - UU^T x \in \mathcal{R}(\{u_i \mid i=1 \ldots n, i \neq j\})
\]

(2020)

E.3.5.2 orthogonal projector, biorthogonal decomposition

We get a similar result for biorthogonal expansion of \( x \in \mathcal{R}(A) \). Define

\[
A \triangleq [a_1 \ a_2 \ \cdots \ a_n] \in \mathbb{R}^{m \times n}
\]

(2021)

and rows of the pseudoinverse\(^{E.9}\)

\[
A^\dagger \triangleq \begin{bmatrix}
a_1^*^T \\
a_2^*^T \\
\vdots \\
a_n^*^T
\end{bmatrix} \in \mathbb{R}^{n \times m}
\]

(2022)

under biorthogonality condition \( A^\dagger A = I \). In biorthogonal expansion (§2.13.8)

\[
x = AA^\dagger x = \sum_{i=1}^{n} a_i a_i^*^T x
\]

(2023)

the direction of projection of \( x \) on \( a_j \) for some particular \( j \in \{1 \ldots n\} \), for example, is orthogonal to \( a_j^* \) and parallel to a vector in the span of all the remaining vectors constituting the columns of \( A \);

\[
a_j^*^T(a_j a_j^*^T x - x) = 0
\]

\[
a_j a_j^*^T x - x = a_j a_j^*^T x - AA^\dagger x \in \mathcal{R}(\{a_i \mid i=1 \ldots n, i \neq j\})
\]

(2024)

\(^{E.9}\)Notation * in this context connotes extreme direction of a dual cone; e.g., (406) or Example E.5.0.0.3.
nonorthogonal projector, biorthogonal decomposition

Because the result in §E.3.5.2 is independent of matrix symmetry \( AA^\dagger = (AA^\dagger)^T \), we must get the same result for any nonorthogonal projector characterized by a biorthogonality condition; namely, for nonorthogonal projector \( P = UQ^T \) (1983) under biorthogonality condition \( Q^TU = I \), in biorthogonal expansion of \( x \in \mathbb{R}(U) \)

\[
x = UQ^Tx = \sum_{i=1}^{k} u_i q_i^T x
\]

where

\[
U \triangleq [u_1 \cdots u_k] \in \mathbb{R}^{m \times k}
\]

\[
Q^T \triangleq \begin{bmatrix} q_1^T \\ \vdots \\ q_k^T \end{bmatrix} \in \mathbb{R}^{k \times m}
\]

the direction of projection of \( x \) on \( u_j \) is orthogonal to \( q_j \) and parallel to a vector in the span of the remaining \( u_i \):

\[
q_j^T(u_j q_j^T x - x) = 0 \\
u_j q_j^T x - x = u_j q_j^T x - UQ^T x \in R(\{u_i | i=1 \ldots k, i \neq j\})
\]

E.4 Algebra of projection on affine subsets

Let \( P_Ax \) denote orthogonal or nonorthogonal projection of \( x \) on affine subset \( A \triangleq \mathbb{R} + \alpha \) where \( \mathbb{R} \) is a subspace and \( \alpha \in A \). Let \( P_Rx \) denote projection of \( x \) on \( R \) in the same direction. Then, because \( R \) is parallel to \( A \), it holds:

\[
P_Ax = P_{R+\alpha}x = P_Rx + (I - P_R)\alpha = P_R(x - \alpha) + \alpha
\]

Orthogonal or nonorthogonal subspace-projector \( P_R \) is a linear operator (\( P_A \) is not). \( P_R(x + y) = P_Rx \) whenever \( y \perp R \) and \( P_R \) is an orthogonal projector.

E.4.0.1 Theorem. Orthogonal projection on affine subset. [[115, §9.26]]

Let \( A = \mathbb{R} + \alpha \) be an affine subset where \( \alpha \in A \), and let \( R^\perp \) be the orthogonal complement of subspace \( R \). Then \( P_Ax \) is the orthogonal projection of \( x \) on \( A \) if and only if

\[
P_Ax \in A, \quad \langle P_Ax - x , a - \alpha \rangle = 0 \quad \forall a \in A
\]

or if and only if

\[
P_Ax \in A, \quad P_Ax - x \in R^\perp
\]
E.4.0.0.2 Example. Intersection of affine subsets.
When designing an optimization problem, intersection of sets is easy to express when the individual sets themselves are. Suppose \( A = \{ x \mid Ax = b \} \) and \( C = \{ x \mid Cx = d \} \) denote two affine subsets. Then membership to their intersection
\[
A \cap C = \left\{ x \mid \begin{bmatrix} A \\ C \end{bmatrix} x = \begin{bmatrix} b \\ d \end{bmatrix} \right\}
\] (2031)
is realized as solution to simultaneous equations. In 1937, Kaczmarz instead proposed alternating projection (§E.10) on hyperplanes (whose normals constitute the rows of matrices \( A \) and \( C \)) as a means to overcome numerical instability when solving very large systems. The intersection may then be described as fixed points of projection on affine subsets, assuming \( A \cap C \neq \emptyset \)
\[
A \cap C = \{ x \mid P_C P_A x = x \}
\] (2032)
an affine system of equations, by (2028), extensible to \( M \) intersecting hyperplanes (§E.5.0.0.8):
\[
A \cap C = \{ x \mid P_M \cdots P_1 x = x \}
\] (2033)

E.5 Projection examples

E.5.0.0.1 Example. Orthogonal projection on orthogonal basis.
Orthogonal projection on a subspace can instead be accomplished by orthogonally projecting on the individual members of an orthogonal basis for that subspace. Suppose, for example, matrix \( A \in \mathbb{R}^{m \times n} \) holds an orthonormal basis for \( \mathcal{R}(A) \) in its columns; \( A = [a_1 \ a_2 \ \cdots \ a_n] \). Then orthogonal projection of vector \( x \in \mathbb{R}^n \) on \( \mathcal{R}(A) \) is a sum of one-dimensional orthogonal projections
\[
P x = AA^\dagger x = (A^T A)^{-1} A^T x = AA^T x = \sum_{i=1}^{n} a_i a_i^T x
\] (2034)
where each symmetric dyad \( a_i a_i^T \) is an orthogonal projector projecting on \( \mathcal{R}(a_i) \). (§E.6.3) Because \( \| x - P x \| \) is minimized by orthogonal projection, \( P x \) is considered to be the best approximation (in the Euclidean sense) to \( x \) from the set \( \mathcal{R}(A) \). [115, §4.9]

E.5.0.0.2 Example. Orthogonal projection on span of nonorthogonal basis.
Orthogonal projection on a subspace can also be accomplished by projecting nonorthogonally on the individual members of any nonorthogonal basis for that subspace. This interpretation is in fact the principal application of the pseudoinverse we discussed. Now suppose matrix \( A \) holds a nonorthogonal basis for \( \mathcal{R}(A) \) in its columns,
\[
A = [a_1 \ a_2 \ \cdots \ a_n] \in \mathbb{R}^{m \times n}
\] (2021)
and define the rows $a_i^* \mathbf{T}$ of its pseudoinverse $A^\dagger$ as in (2022). Then orthogonal projection of vector $x \in \mathbb{R}^n$ on $\mathcal{R}(A)$ is a sum of one-dimensional nonorthogonal projections

$$P_x = A A^\dagger x = \sum_{i=1}^{n} a_i a_i^* \mathbf{T} x$$

(2035)

where each nonsymmetric dyad $a_i a_i^* \mathbf{T}$ is a nonorthogonal projector projecting on $\mathcal{R}(a_i)$, (§E.6.1) idempotent because of biorthogonality condition $A^\dagger A = I$.

The projection $P_x$ is regarded as the best approximation to $x$ from the set $\mathcal{R}(A)$, as it was in Example E.5.0.0.1.

E.5.0.0.3 Example. Biorthogonal expansion as nonorthogonal projection.

Biorthogonal expansion can be viewed as a sum of components, each a nonorthogonal projection on the range of an extreme direction of a pointed polyhedral cone $\mathcal{K}$; e.g., Figure 175.

Suppose matrix $A \in \mathbb{R}^{m \times n}$ holds a nonorthogonal basis for $\mathcal{R}(A)$ in its columns as in (2021), and the rows of pseudoinverse $A^\dagger$ are defined as in (2022). Assuming the most general biorthogonality condition $(A^\dagger + BZ^T)A = I$ with $BZ^T$ defined as for (1980), then biorthogonal expansion of vector $x$ is a sum of one-dimensional nonorthogonal projections; for $x \in \mathcal{R}(A)$

$$x = A(A^\dagger + BZ^T)x = AA^\dagger x = \sum_{i=1}^{n} a_i a_i^* \mathbf{T} x$$

(2036)

where each dyad $a_i a_i^* \mathbf{T}$ is a nonorthogonal projector projecting on $\mathcal{R}(a_i)$. (§E.6.1) The extreme directions of $\mathcal{K} = \text{cone}(A)$ are $\{a_1 \ldots a_n\}$ the linearly independent columns of $A$, while the extreme directions $\{a_1^* \ldots a_n^*\}$ of relative dual cone $\mathcal{K}^* \cap \text{aff} \mathcal{K} = \text{cone}(A^T \mathbf{T})$ (§2.13.9.4) correspond to the linearly independent (§B.1.1.1) rows of $A^\dagger$. Directions of nonorthogonal projection are determined by the pseudoinverse; id est, direction of projection $a_i a_i^* \mathbf{T} x - x$ on $\mathcal{R}(a_i)$ is orthogonal to $a_i^*$.

Because the extreme directions of this cone $\mathcal{K}$ are linearly independent, component projections are unique in the sense:

- There is only one linear combination of extreme directions of $\mathcal{K}$ that yields a particular point $x \in \mathcal{R}(A)$ whenever

$$\mathcal{R}(A) = \text{aff} \mathcal{K} = \mathcal{R}(a_1) \oplus \mathcal{R}(a_2) \oplus \ldots \oplus \mathcal{R}(a_n)$$

(2037)

E.5.0.0.4 Example. Nonorthogonal projection on elementary matrix.

Suppose $P_Y$ is a linear nonorthogonal projector projecting on subspace $\mathcal{Y}$, and suppose the range of a vector $u$ is linearly independent of $\mathcal{Y}$; id est, for some other subspace $\mathcal{M}$ containing $\mathcal{Y}$ suppose

$$\mathcal{M} = \mathcal{R}(u) \oplus \mathcal{Y}$$

(2038)

This remains true in high dimension although only a little more difficult to visualize in $\mathbb{R}^3$; confer, Figure 66.
Figure 175: (confer Figure 65) Biorthogonal expansion of point $x \in \text{aff} \mathcal{K}$ is found by projecting $x$ nonorthogonally on extreme directions of polyhedral cone $\mathcal{K} \subset \mathbb{R}^2$. (Dotted lines of projection bound this translated negated cone.) Direction of projection $P_{a_1}$ on extreme direction $a_1$ is orthogonal to extreme direction $a_1^*$ of dual cone $\mathcal{K}^*$ and parallel to $a_2$ (§E.3.5); similarly, direction of projection $P_{a_2}$ on $a_2$ is orthogonal to $a_2^*$ and parallel to $a_1$. Point $x$ is sum of nonorthogonal projections: $x$ on $\mathcal{R}(a_1)$ and $x$ on $\mathcal{R}(a_2)$. Expansion is unique because extreme directions of $\mathcal{K}$ are linearly independent. Were $a_1$ orthogonal to $a_2$, then $\mathcal{K}$ would be identical to $\mathcal{K}^*$ and nonorthogonal projections would become orthogonal.
Assuming $P_M x = P_u x + P_Y x$ holds, then it follows for vector $x \in M$
\[
P_u x = x - P_Y x, \quad P_Y x = x - P_u x
\]
(2039)
nonorthogonal projection of $x$ on $\mathcal{R}(u)$ can be determined from nonorthogonal projection of $x$ on $\mathcal{Y}$, and vice versa.

Such a scenario is realizable were there some arbitrary basis for $\mathcal{Y}$ populating a full-rank skinny-or-square matrix $A$
\[
A \triangleq [\text{basis } \mathcal{Y} \quad u] \in \mathbb{R}^{N \times n+1}
\]
(2040)
With $P_u = A(:,n+1)A^\dagger(n+1,:)$ and $P_Y = A(:,1:n)A^\dagger(1:n,:)$, then $P_M = AA^\dagger$ fulfills the requirements. Observe, $P_M$ is an orthogonal projector whereas $P_Y$ and $P_u$ are nonorthogonal projectors.

Now suppose, for example, $P_Y$ is an elementary matrix (§B.3); in particular,
\[
P_Y = I - e_1 1^T = \begin{bmatrix} 0 \sqrt{2} V_N \end{bmatrix} \in \mathbb{R}^{N \times N}
\]
(2041)
where $\mathcal{Y} = \mathcal{N}(1^T)$. We have $\mathcal{M} = \mathbb{R}^N$, $A = [\sqrt{2} V_N \quad e_1]$, and $u = e_1$. Thus $P_u = e_1 1^T$ is a nonorthogonal projector projecting on $\mathcal{R}(u)$ in a direction parallel to a vector in $\mathcal{Y}$ (§E.3.5), and $P_Y x = x - e_1 1^T x$ is a nonorthogonal projection of $x$ on $\mathcal{Y}$ in a direction parallel to $u$.

\section*{E.5.0.0.5 Example. Projecting origin on a hyperplane. \hspace{1cm} (confer §2.4.2.0.2)}
Given the hyperplane representation having $b \in \mathbb{R}$ and nonzero normal $a \in \mathbb{R}^m$
\[
\partial \mathcal{H} = \{ y \mid a^T y = b \} \subset \mathbb{R}^m
\]
(114)
orthogonal projection of the origin $P \mathbf{0}$ on that hyperplane is the unique optimal solution to a minimization problem: \hspace{0.5cm} (2002)
\[
\| P \mathbf{0} - \mathbf{0} \|_2 = \inf_{y \in \partial \mathcal{H}} \| y - \mathbf{0} \|_2
\]
\[
= \inf_{\xi \in \mathbb{R}^{m-1}} \| Z \xi + x \|_2
\]
(2042)
where $x$ is any solution to $a^T y = b$, and where the columns of $Z \in \mathbb{R}^{m \times m-1}$ constitute a basis for $\mathcal{N}(a^T)$ so that $y = Z \xi + x \in \partial \mathcal{H}$ for all $\xi \in \mathbb{R}^{m-1}$.

The infimum can be found by setting the gradient (with respect to $\xi$) of the strictly convex norm-square to $\mathbf{0}$. We find the minimizing argument
\[
\xi^* = -(Z^T Z)^{-1} Z^T x
\]
(2043)
so \hspace{0.5cm} (404)
\[
y^* = (I - Z(Z^T Z)^{-1} Z^T)x
\]
(2044)
and from (2004)
\[
P \mathbf{0} = y^* = a (a^T a)^{-1} a^T x = \frac{a}{\| a \|^2} a^T x \triangleq A^\dagger A x = a \frac{b}{\| a \|^2}
\]
(2045)
In words, any point $x$ in the hyperplane $\partial \mathcal{H}$ projected on its normal $a$ \hspace{0.5cm} (confer (2069)) yields that point $y^*$ in the hyperplane closest to the origin. \hspace{0.5cm} $\Box$
E.5.0.0.6 Example. Projection on affine subset.
The technique of Example E.5.0.0.5 is extensible. Given an intersection of hyperplanes
\[ \mathcal{A} = \{ y \mid Ay = b \} \subset \mathbb{R}^m \] (2046)
where each row of \( A \in \mathbb{R}^{m \times n} \) is nonzero and \( b \in \mathcal{R}(A) \), then the orthogonal projection \( Px \) of any point \( x \in \mathbb{R}^n \) on \( \mathcal{A} \) is the solution to a minimization problem:
\[ \| Px - x \|_2 = \inf_{y \in \mathcal{A}} \| y - x \|_2 = \inf_{\xi \in \mathbb{R}^{n-\text{rank} A}} \| Z \xi + y_p - x \|_2 \] (2047)
where \( y_p \) is any solution to \( Ay = b \), and where the columns of \( Z \in \mathbb{R}^{n \times n-\text{rank} A} \) constitute
a basis for \( \mathcal{N}(A) \) so that \( y = Z \xi + y_p \in \mathcal{A} \) for all \( \xi \in \mathbb{R}^{n-\text{rank} A} \). When rank of full-rank
\( A \) is \( n - 1 \), then \( \mathcal{A} \) describes a line; when rank \( A = 1 \), then \( \mathcal{A} \) describes a hyperplane.
The infimum is found by setting the gradient of the strictly convex norm-square to 0.
The minimizing argument is
\[ \xi^* = -(Z^T Z)^{-1} Z^T (y_p - x) \] (2048)
so
\[ y^* = (I - Z(Z^T Z)^{-1} Z^T)(y_p - x) + x \] (2049)
and from (2004),
\[ Px = y^* = x - A^\dagger(Ax - b) = (I - A^\dagger A)x + A^\dagger A y_p \] (2050)
which is a projection of \( x \) on \( \mathcal{N}(A) \) then translated perpendicularly with respect to the nullspace until it meets the affine subset \( \mathcal{A} \).

E.5.0.0.7 Example. Projection on affine subset, vertex-description.
Suppose now we instead describe the affine subset \( \mathcal{A} \) in terms of some given minimal set
of generators arranged columnar in \( X \in \mathbb{R}^{n \times N} \) (76); id est,
\[ \mathcal{A} = \text{aff } X = \{ Xa \mid a^T 1 = 1 \} \subset \mathbb{R}^n \] (78)
Here minimal set means \( XV_N = [x_2 - x_1 \ x_3 - x_1 \ \cdots \ x_N - x_1] / \sqrt{2} \) (1041) is full-rank
(§2.4.2.2) where \( V_N \in \mathbb{R}^{N \times N-1} \) is the Schoenberg auxiliary matrix (§B.4.2). For \( N = 2 \)
affinely independent generators, \( \mathcal{A} \) represents a line; for \( N = n \) affinely independent generators, \( \mathcal{A} \) represents a hyperplane. Then the orthogonal projection \( Px \) of any point \( x \in \mathbb{R}^n \) on \( \mathcal{A} \) is the solution to a minimization problem:
\[ \| Px - x \|_2 = \inf_{a^T 1 = 1} \| Xa - x \|_2 = \inf_{\xi \in \mathbb{R}^{N-1}} \| X(V_N \xi + a_p) - x \|_2 \] (2051)
where \( a_p \) is any solution to \( a^T 1 = 1 \). We find the minimizing argument
\[ \xi^* = -(V_N^T XV_N)^{-1} V_N^T X^T (Xa_p - x) \] (2052)
and so the orthogonal projection is [215, §3]
\[
P(x) = Xa^* = (I - XV_N(XV_N)^\dagger)Xa + XV_N(XV_N)^\dagger x
\]  
(2053)
a projection of point \( x \) on \( \mathcal{R}(XV_N) \) then translated perpendicularly with respect to that range until it meets the affine subset \( A \).

**E.5.0.0.8 Example.** *Projecting on hyperplane, halfspace, slab.*

Given hyperplane representation having \( b \in \mathbb{R} \) and nonzero normal \( a \in \mathbb{R}^m \)
\[
\partial \mathcal{H} = \{ y \mid a^Ty = b \} \subset \mathbb{R}^m \quad (114)
\]
orthogonal projection of any point \( x \in \mathbb{R}^m \) on that hyperplane [216, §3.1] is unique:
\[
P(x) = x - a(a^Ta)^{-1}(a^Tx - b)
\]  
(2054)
Orthogonal projection of \( x \) on the halfspace parametrized by \( b \in \mathbb{R} \) and nonzero normal \( a \in \mathbb{R}^m \)
\[
\mathcal{H}_- = \{ y \mid a^Ty \leq b \} \subset \mathbb{R}^m \quad (106)
\]
is the point
\[
P(x) = x - a(a^Ta)^{-1}\max\{0, \ a^Tx - b\}
\]  
(2055)
Orthogonal projection of \( x \) on the convex slab (Figure 12), for \( c < b \)
\[
\{ y \mid c \leq a^Ty \leq b \} \subset \mathbb{R}^m
\]  
(2056)
is the point [155, §5.1]
\[
P(x) = x - a(a^Ta)^{-1}\left(\max\{0, \ a^Tx - b\} - \max\{0, \ c - a^Tx\}\right)
\]  
(2057)

**E.6 Vectorization interpretation, projection on a matrix**

**E.6.1 Nonorthogonal projection on a vector**

Nonorthogonal projection of vector \( x \) on the range of vector \( y \) is accomplished using a normalized dyad \( P_0 \) (§B.1); *videlicet,*
\[
\langle z, x \rangle y = \frac{z^Tx}{z^Ty} y = y\frac{z^Ty}{z^Tx} x \triangleq P_0 x
\]  
(2058)
where \( \langle z, x \rangle / \langle z, y \rangle \) is the coefficient of projection on \( y \). Because \( P_0^2 = P_0 \) and \( \mathcal{R}(P_0) = \mathcal{R}(y) \), rank-one matrix \( P_0 \) is a nonorthogonal projector dyad projecting on \( \mathcal{R}(y) \). Direction of nonorthogonal projection is orthogonal to \( z \); *id est,*
\[
P_0 x - x \perp \mathcal{R}(P_0^T)
\]  
(2059)
E.6. VECTORIZATION INTERPRETATION, 621

E.6.2 Nonorthogonal projection on vectorized matrix

Formula (2058) is extensible. Given $X, Y, Z \in \mathbb{R}^{m \times n}$, we have the one-dimensional nonorthogonal projection of $X$ in isomorphic $\mathbb{R}^{mn}$ on the range of vectorized $Y$: (§2.2)

$$\frac{\langle Z, X \rangle}{\langle Z, Y \rangle} Y, \quad \langle Z, Y \rangle \neq 0$$

(2060)

where $\langle Z, X \rangle / \langle Z, Y \rangle$ is the coefficient of projection. The inequality accounts for the fact: projection on $\mathcal{R}(\text{vec} Y)$ is in a direction orthogonal to vec $Z$.

E.6.2.1 Nonorthogonal projection on dyad

Now suppose we have nonorthogonal projector dyad

$$P_0 = \frac{yz^T}{z^T y} \in \mathbb{R}^{m \times m}$$

(2061)

Analogous to (2058), for $X \in \mathbb{R}^{m \times m}$

$$P_0 X P_0 = \frac{yz^T}{z^T y} X \frac{yz^T}{z^T y} = \frac{z^T X y}{(z^T y)^2} yz^T = \frac{\langle z y^T, X \rangle}{\langle z y^T, y z^T \rangle} yz^T$$

(2062)

is a nonorthogonal projection of matrix $X$ on the range of vectorized dyad $P_0$; from which it follows:

$$P_0 X P_0 = \frac{z^T X y}{z^T y} \frac{yz^T}{z^T y} = \langle z y^T, X \rangle \frac{yz^T}{z^T y} = \langle P_0^T, X \rangle P_0 = \frac{\langle P_0^T, X \rangle}{\langle P_0^T, P_0 \rangle} P_0$$

(2063)

Yet this relationship between matrix product and vector inner-product only holds for a dyad projector. When nonsymmetric projector $P_0$ is rank-one as in (2061), therefore,

$$\mathcal{R}(\text{vec} P_0 X P_0) = \mathcal{R}(\text{vec} P_0) \text{ in } \mathbb{R}^{m^2}$$

(2064)

and

$$P_0 X P_0 - X \perp P_0^T \text{ in } \mathbb{R}^{m^2}$$

(2065)

E.6.2.1.1 Example. $\lambda$ as coefficients of nonorthogonal projection.

Any diagonalization (§A.5)

$$X = S \Lambda S^{-1} = \sum_{i=1}^{m} \lambda_i s_i w_i^T \in \mathbb{R}^{m \times m}$$

(1629)
may be expressed as a sum of one-dimensional nonorthogonal projections of \( X \), each on the range of a vectorized eigenmatrix \( P_j \equiv s_j w_j^T \):

\[
X = \sum_{i,j=1}^{m} \langle (S e_i e_j^T S^{-1})^T, X \rangle S e_i e_j^T S^{-1} = \sum_{j=1}^{m} \langle (s_j w_j^T)^T, X \rangle s_j w_j^T + \sum_{i,j=1}^{m} \langle (S e_i e_j^T S^{-1})^T, S a S^{-1} \rangle S e_i e_j^T S^{-1}
\]

\[
= \sum_{j=1}^{m} \langle (s_j w_j^T)^T, X \rangle s_j w_j^T \quad \text{(2066)}
\]

\[
= \sum_{j=1}^{m} (P_j^T, X) P_j = \sum_{j=1}^{m} s_j w_j^T X s_j w_j^T = \sum_{j=1}^{m} P_j XP_j
\]

\[
= \sum_{j=1}^{m} \lambda_j s_j w_j^T
\]

This biorthogonal expansion of matrix \( X \) is a sum of nonorthogonal projections because the term outside the projection coefficient \( \langle \quad \rangle \) is not identical to the inside-term. (§E.6.4) The eigenvalues \( \lambda_j \) are coefficients of nonorthogonal projection of \( X \), while the remaining \( M(M-1)/2 \) coefficients (for \( i \neq j \)) are zeroed by projection. When \( P_j \) is rank-one as in (2066),

\[
\mathcal{R}(\text{vec} P_j XP_j) = \mathcal{R}(\text{vec} s_j w_j^T) = \mathcal{R}(\text{vec} P_j) \text{ in } \mathbb{R}^{m^2} \quad (2067)
\]

and

\[
P_j XP_j - X \perp P_j^T \text{ in } \mathbb{R}^{m^2} \quad (2068)
\]

Were matrix \( X \) symmetric, then its eigenmatrices would also be. So the one-dimensional projections would become orthogonal. (§E.6.4.1.1) \( \square \)

### E.6.3 Orthogonal projection on a vector

The formula for orthogonal projection of vector \( x \) on the range of vector \( y \) (one-dimensional projection) is basic analytic geometry; [14, §3.3] [341, §3.2] [377, §2.2] [412, §1-8]

\[
\langle y, x \rangle y = \frac{y^T x}{y^T y} y = \frac{yy^T}{y^T y} x \triangleq P_1 x \quad (2069)
\]

where \( \langle y, x \rangle/\langle y, y \rangle \) is the coefficient of projection on \( \mathcal{R}(y) \). An equivalent description is: Vector \( P_1 x \) is the orthogonal projection of vector \( x \) on \( \mathcal{R}(P_1) = \mathcal{R}(y) \). Rank-one matrix \( P_1 \) is a projection matrix because \( P_1^2 = P_1 \). The direction of projection is orthogonal

\[
P_1 x - x \perp \mathcal{R}(P_1) \quad (2070)
\]

because \( P_1^T = P_1 \).

### E.6.4 Orthogonal projection on a vectorized matrix

From (2069), given instead \( X, Y \in \mathbb{R}^{m \times n} \), we have the one-dimensional orthogonal projection of matrix \( X \) in isomorphic \( \mathbb{R}^{mn} \) on the range of vectorized \( Y \): (§2.2)

\[
\frac{\langle Y, X \rangle}{\langle Y, Y \rangle} Y \quad (2071)
\]
E.6. VECTORIZATION INTERPRETATION,

where \( \langle Y, X \rangle / \langle Y, Y \rangle \) is the coefficient of projection.

For orthogonal projection, the term outside the vector inner-products \( \langle \; \rangle \) must be identical to the terms inside in three places.

E.6.4.1 Orthogonal projection on dyad

There is opportunity for insight when \( Y \) is a dyad \( yz^T \) (§B.1): Instead given \( X \in \mathbb{R}^{m \times n} \), \( y \in \mathbb{R}^m \), and \( z \in \mathbb{R}^n \)

\[
\frac{\langle yz^T, X \rangle}{\langle yz^T, yz^T \rangle} yz^T = \frac{y^T X z}{y^T y z^T} yz^T \tag{2072}
\]

is the one-dimensional orthogonal projection of \( X \) in isomorphic \( \mathbb{R}^{mn} \) on the range of vectorized \( yz^T \). To reveal the obscured symmetric projection matrices \( P_1 \) and \( P_2 \) we rewrite (2072):

\[
\frac{y^T X z}{y^T y z^T} yz^T = \frac{y y^T}{y^T y} X \frac{z z^T}{z^T z} \triangleq P_1 XP_2 \tag{2073}
\]

So for projector dyads, projection (2073) is the orthogonal projection in \( \mathbb{R}^{mn} \) if and only if projectors \( P_1 \) and \( P_2 \) are symmetric; \( E.11 \) in other words,

- for orthogonal projection on the range of a vectorized dyad \( yz^T \), the term outside the vector inner-products \( \langle \; \rangle \) in (2072) must be identical to the terms inside in three places.

When \( P_1 \) and \( P_2 \) are rank-one symmetric projectors as in (2073), (37)

\[
\mathcal{R}(\text{vec } P_1 XP_2) = \mathcal{R}(\text{vec } yz^T) \text{ in } \mathbb{R}^{mn} \tag{2074}
\]

and

\[
P_1 XP_2 - X \perp yz^T \text{ in } \mathbb{R}^{mn} \tag{2075}
\]

When \( y = z \) then \( P_1 = P_2 = P_2^T \) and

\[
P_1 XP_1 = \langle P_1, X \rangle P_1 = \frac{\langle P_1, X \rangle}{\langle P_1, P_1 \rangle} P_1 \tag{2076}
\]

meaning, \( P_1 XP_1 \) is equivalent to orthogonal projection of matrix \( X \) on the range of vectorized projector dyad \( P_1 \). Yet this relationship between matrix product and vector inner-product does not hold for general symmetric projector matrices.

\( E.11 \) For diagonalizable \( X \in \mathbb{R}^{m \times m} \) (§A.5), its orthogonal projection in isomorphic \( \mathbb{R}^{mn} \) on the range of vectorized \( yz^T \in \mathbb{R}^{mn} \) becomes:

\[
P_1 XP_2 = \sum_{i=1}^{m} \lambda_i P_1 s_i w_i^T P_2
\]

When \( \mathcal{R}(P_1) = \mathcal{R}(w_j) \) and \( \mathcal{R}(P_2) = \mathcal{R}(s_j) \), the \( j \)th dyad term from the diagonalization is isolated but only, in general, to within a scale factor because neither set of left or right eigenvectors is necessarily orthonormal unless \( X \) is a normal matrix [425, §3.2]. Yet when \( \mathcal{R}(P_2) = \mathcal{R}(s_k), k \neq j \in \{1 \ldots m\} \), then \( P_1 XP_2 = 0 \).
E.6.4.1.1 Example. Eigenvalues λ as coefficients of orthogonal projection.

Let C represent any convex subset of subspace $S^M$, and let $C_1$ be any element of $C$. Then $C_1$ can be expressed as the orthogonal expansion

$$C_1 = \sum_{i=1}^{M} \sum_{j=1 \atop j \geq i}^{M} \langle E_{ij}, C_1 \rangle E_{ij} \in C \subset S^M$$

(2077)

where $E_{ij} \in S^M$ is a member of the standard orthonormal basis for $S^M$ (59). This expansion is a sum of one-dimensional orthogonal projections of $C_1$; each projection on the range of a vectorized standard basis matrix. Vector inner-product $\langle E_{ij}, C_1 \rangle$ is the coefficient of projection of svec$C_1$ on $R(svec E_{ij})$.

When $C_1$ is any member of a convex set $C$ whose dimension is $L$, Carathéodory’s theorem [119] [318] [209] [42] [43] guarantees that no more than $L+1$ affinely independent members from $C$ are required to faithfully represent $C_1$ by their linear combination. E.12

Dimension of $S^M$ is $L = M(M+1)/2$ in isometrically isomorphic $\mathbb{R}^{M(M+1)/2}$. Yet because any symmetric matrix can be diagonalized, (§A.5.1) $C_1 \in S^M$ is a linear combination of its $M$ eigenmatrices $q_i q_i^T$ (§A.5.0.3) weighted by its eigenvalues $\lambda_i$;

$$C_1 = Q \Lambda Q^T = \sum_{i=1}^{M} \lambda_i q_i q_i^T$$

(2078)

where $\Lambda \in S^M$ is a diagonal matrix having $\delta(A)_i = \lambda_i$, and $Q = [q_1 \ldots q_M]$ is an orthogonal matrix in $\mathbb{R}^{M \times M}$ containing corresponding eigenvectors.

To derive eigenvalue decomposition (2078) from expansion (2077), $M$ standard basis matrices $E_{ij}$ are rotated (§B.5) into alignment with the $M$ eigenmatrices $q_i q_i^T$ of $C_1$ by applying a similarity transformation; [341, §5.6]

$$\{QE_{ij}Q^T\} = \begin{cases} q_i q_i^T, & i = j = 1 \ldots M \\ \frac{1}{\sqrt{2}}(q_i q_j^T + q_j q_i^T), & 1 \leq i < j \leq M \end{cases}$$

(2079)

which remains an orthonormal basis for $S^M$. Then remarkably

$$C_1 = \sum_{i=1}^{M} \sum_{j=1 \atop j \geq i}^{M} \langle QE_{ij}Q^T, C_1 \rangle QE_{ij}Q^T$$

$$= \sum_{i=1}^{M} \langle q_i q_i^T, C_1 \rangle q_i q_i^T + \sum_{i, j=1}^{M} \langle QE_{ij}Q^T, \Lambda Q Q^T \rangle QE_{ij}Q^T$$

$$= \sum_{i=1}^{M} \langle q_i q_i^T, C_1 \rangle q_i q_i^T$$

(2080)

$$\Delta \sum_{i=1}^{M} \langle P_i, C_1 \rangle P_i = \sum_{i=1}^{M} q_i q_i^T C_1 q_i q_i^T = \sum_{i=1}^{M} P_i C_1 P_i$$

$$= \sum_{i=1}^{M} \lambda_i q_i q_i^T$$

E.12 Carathéodory’s theorem guarantees existence of a biorthogonal expansion for any element in aff $C$ when $C$ is any pointed closed convex cone.
E.6. VECTORIZATION INTERPRETATION, 625

this orthogonal expansion becomes the diagonalization; still a sum of one-dimensional orthogonal projections. The eigenvalues

\[ \lambda_i = \langle q_i q_i^T, C_1 \rangle \]  

(2081)

are clearly coefficients of projection of \( C_1 \) on the range of each vectorized eigenmatrix. (confer §E.6.2.1.1) The remaining \( M(M-1)/2 \) coefficients \( (i \neq j) \) are zeroed by projection. When \( P_1 \) is rank-one symmetric as in (2080),

\[ \mathcal{R}(\text{svec} \ P_1 C_1 P_1) = \mathcal{R}(\text{svec} \ q_i q_i^T) = \mathcal{R}(\text{svec} \ P_1) \text{ in } \mathbb{R}^{M(M+1)/2} \]  

(2082)

and

\[ P_1 C_1 P_1 - C_1 \perp P_1 \text{ in } \mathbb{R}^{M(M+1)/2} \]  

(2083)

\[ \square \]

E.6.4.2 Positive semidefiniteness test as orthogonal projection

For any given \( X \in \mathbb{R}^{m \times m} \) the familiar quadratic construct \( y^T X y \geq 0 \), over broad domain, is a fundamental test for positive semidefiniteness. (§A.2) It is a fact that \( y^T X y \) is always proportional to a coefficient of orthogonal projection; letting \( z \) in formula (2072) become \( y \in \mathbb{R}^m \), then \( P_2 = P_1 = y y^T / y^T y = y y^T / \| y y^T \|_2 \) (confer (1694)) and formula (2073) becomes

\[ \frac{\langle y y^T, X \rangle}{\langle y y^T, y y^T \rangle} y y^T = \frac{y y^T}{y^T y} = \frac{y y^T}{y^T y} \frac{X}{y^T y} \triangleq P_1 X P_1 \]  

(2084)

By (2071), product \( P_1 X P_1 \) is the one-dimensional orthogonal projection of \( X \) in isomorphic \( \mathbb{R}^{m^2} \) on the range of vectorized \( P_1 \) because, for rank \( P_1 = 1 \) and \( P_1^2 = P_1 \in \mathbb{S}^m \) (confer (2063))

\[ P_1 X P_1 = \frac{y y^T}{y^T y} = \frac{\langle y y^T, X \rangle}{\langle y y^T, y y^T \rangle} = \frac{P_1}{P_1, P_1} \]  

(2085)

The coefficient of orthogonal projection \( \langle P_1, X \rangle = y^T X y / (y^T y) \) is also known as Rayleigh’s quotient. E.13 When \( P_1 \) is rank-one symmetric as in (2084),

\[ \mathcal{R}(\text{vec} \ P_1 X P_1) = \mathcal{R}(\text{vec} \ P_1) \text{ in } \mathbb{R}^{m^2} \]  

(2086)

E.13 When \( y \) becomes the \( j \)th eigenvector \( s_j \) of diagonalizable \( X \), for example, \( \langle P_1, X \rangle \) becomes the \( j \)th eigenvalue: [205, §III]

\[ \langle P_1, X \rangle_{y = s_j} = \frac{s_j^T \left( \sum_{i=1}^m \lambda_i s_i w_i^T \right) s_j}{s_j s_j^T} = \lambda_j \]

Similarly for \( y = w_j \), the \( j \)th left-eigenvector,

\[ \langle P_1, X \rangle_{y = w_j} = \frac{w_j^T \left( \sum_{i=1}^m \lambda_i s_i w_i^T \right) w_j}{w_j w_j^T} = \lambda_j \]

A quandary may arise regarding the potential annihilation of the antisymmetric part of \( X \) when \( s_j^T X s_j \) is formed. Were annihilation to occur, it would imply the eigenvalue thus found came instead from the symmetric part of \( X \). The quandary is resolved recognizing that diagonalization of real \( X \) admits complex eigenvectors; hence, annihilation could only come about by forming \( \text{re}(s_j^H X s_j) = s_j^H (X + X^T) s_j / 2 \) [212, §7.1] where \( (X + X^T) / 2 \) is the symmetric part of \( X \), and \( s_j^H \) denotes conjugate transpose.
APPENDIX E. PROJECTION

and

\[ P_1 X P_1 - X \perp P_1 \text{ in } \mathbb{R}^{m^2} \]  \hspace{1cm} (2087)

The test for positive semidefiniteness, then, is a test for nonnegativity of the coefficient of orthogonal projection of \( X \) on the range of each and every vectorized extreme direction \( yy^T \) (\S2.8.1) from the positive semidefinite cone in the ambient space of symmetric matrices.

E.6.4.3 \( P XP \succeq 0 \)

In some circumstances, it may be desirable to limit the domain of test \( y^T X y \geq 0 \) for positive semidefiniteness; e.g., \( \{ \|y\| = 1 \} \). Another example of limiting domain-of-test is central to Euclidean distance geometry: For \( \mathcal{R}(V) = \mathcal{N}(1^T) \), the test \( -VDV \succeq 0 \) determines whether \( D \in \mathbb{S}_N^+ \) is a Euclidean distance matrix. The same test may be stated:

\[ D \in \text{EDM}^N \equiv -y^T D y = \langle yy^T, -D \rangle \geq 0 \quad \forall \ y \in \mathcal{R}(V) \]  \hspace{1cm} (2088)

The test \( -VDV \succeq 0 \) is therefore equivalent to a test for nonnegativity of the coefficient of orthogonal projection of \( -D \) on the range of each and every vectorized extreme direction \( yy^T \) from the positive semidefinite cone \( \mathbb{S}_N^+ \) such that \( \mathcal{R}(yy^T) = \mathcal{R}(y) \subseteq \mathcal{R}(V) \). (Validity of this result is independent of whether \( V \) is itself a projection matrix.)

E.7 Projection on matrix subspaces

E.7.1 \( P XP \) misinterpretation for higher-rank \( P \)

For a projection matrix \( P \) of rank greater than 1, \( P XP \) is generally not commensurate with \( \langle P, X \rangle = P^T X P \) as is the case for projector dyads (2085). Yet for a symmetric idempotent matrix \( P \) of any rank we are tempted to say “\( P XP \) is the orthogonal projection of \( X \in \mathbb{S}^m \) on \( \mathcal{R}(\text{vec } P) \)”. The fallacy is: \( \text{vec } P XP \) does not necessarily belong to the range of vectorized \( P \); the most basic requirement for projection on \( \mathcal{R}(\text{vec } P) \).

E.7.2 Orthogonal projection on matrix subspaces

With \( A_1 \in \mathbb{R}^{m \times n}, B_1 \in \mathbb{R}^{n \times k}, Z_1 \in \mathbb{R}^{m \times k}, A_2 \in \mathbb{R}^{p \times n}, B_2 \in \mathbb{R}^{n \times k}, Z_2 \in \mathbb{R}^{p \times k} \) as defined for nonorthogonal projector (1980), and defining

\[ P_1 \triangleq A_1 A_1^T \in \mathbb{S}^m, \quad P_2 \triangleq A_2 A_2^T \in \mathbb{S}^p \]  \hspace{1cm} (2089)

then, given compatible \( X \)

\[ \|X - P_1 X P_2\|_F = \inf_{B_1, B_2 \in \mathbb{R}^{n \times k}} \|X - A_1 (A_1^T + B_1 Z_1^T) X (A_2^T + Z_2 B_2^T) A_2\|_F \]  \hspace{1cm} (2090)

As for all subspace projectors, range of the projector is the subspace on which projection is made: \( \{ P_1 Y P_2 \mid Y \in \mathbb{R}^{m \times p} \} \). For projectors \( P_1 \) and \( P_2 \) of any rank, altogether, this means projection \( P_1 X P_2 \) is unique minimum-distance, orthogonal

\[ P_1 X P_2 - X \perp \{ P_1 Y P_2 \mid Y \in \mathbb{R}^{m \times p} \} \text{ in } \mathbb{R}^{mp} \]  \hspace{1cm} (2091)

and \( P_1 \) and \( P_2 \) must each be symmetric (confer (2073)) to attain the infimum.
E.7. PROJECTION ON MATRIX SUBSPACES

627


Defining \( P \triangleq A_1(A_1^T + B_1Z_1^T) \),

\[
\begin{align*}
\inf_{B_1, B_2} \|X - A_1(A_1^T + B_1Z_1^T)X(A_2^T + Z_2B_2^T)A_2^T\|^2_F \\
= \inf_{B_1, B_2} \|X - PX(A_1^T + Z_2B_2^T)A_2^T\|^2_F \\
= \inf_{B_1, B_2} \text{tr}\left((X^T - A_2(A_1^T + B_2Z_2^T)X^TP^T)(X - PX(A_1^T + Z_2B_2^T)A_2^T)\right) \\
= \inf_{B_1, B_2} \text{tr}\left(X^TPX(A_1^T + B_2Z_2^T)A_2^T - A_2(A_1^T + B_2Z_2^T)X^TP^TX \right. \\
\left. + A_2(A_1^T + B_2Z_2^T)X^TP^PX(A_2^T + Z_2B_2^T)A_2^T\right) \\
= \inf_{B_1, B_2} \text{tr}\left(-X^TPXZ_2B_2^T(A_1^T + B_2Z_2^T)A_2^T - A_2B_2Z_2^TX^TP^TXZ_2B_2^T(A_1^T + B_2Z_2^T)A_2^T \right. \\
\left. + A_2B_2Z_2^TX^TP^PXZ_2B_2^T(A_1^T + B_2Z_2^T)A_2^T \right) \\
= \inf_{B_1, B_2} \text{tr}\left(-2A_2A_2^TX^TPXZ_2 + 2A_2A_2^TX^TP^PXZ_2 + 2A_2A_2^TX^TP^PXZ_2 \right) \\
= \inf_{B_1, B_2} \text{tr}\left(-2A_2A_2^TX^TPXZ_2 + 2A_2A_2^TX^TP^PXZ_2 \right) \\
= 0 \\
\end{align*}
\]

(2092)

(2093)

(2094)

(2095)

Necessary conditions for a global minimum are \( \nabla_{B_1} = 0 \) and \( \nabla_{B_2} = 0 \). Terms not containing \( B_2 \) in (2092) will vanish from gradient \( \nabla_{B_2} \); (§D.2.3)

\[
\nabla_{B_2}\text{tr}\left(-X^TPXZ_2B_2^T(A_1^T + B_2Z_2^T)A_2^T - A_2B_2Z_2^TX^TP^TXZ_2B_2^T(A_1^T + B_2Z_2^T)A_2^T \right) \\
= \text{tr}\left(-2A_2A_2^TX^TPXZ_2 + 2A_2A_2^TX^TP^PXZ_2 \right) \\
= 0 \\
\Rightarrow \quad \mathcal{R}(B_1) \subseteq \mathcal{N}(A_1) \text{ and } \mathcal{R}(B_2) \subseteq \mathcal{N}(A_2) \\
\]

(or \( Z_2 = 0 \)) because \( A^T = A^TAA^T \). Symmetry requirement (2089) is implicit. Were instead \( P^T \triangleq (A_1^T + Z_2B_2^T)A_2^T \) and the gradient with respect to \( B_1 \) observed, then similar results are obtained. The projector is unique. Perpendicularity (2091) establishes uniqueness [115, §4.9] of projection \( P_1XP_2 \) on a matrix subspace. The minimum-distance projector is the orthogonal projector, and vice versa.

E.7.2.0.2 Example. \( PXP \) redux \& \( \mathcal{N}(V) \).

Suppose we define a subspace of \( m \times n \) matrices, each elemental matrix having columns constituting a list whose geometric center (§5.5.1.0.1) is the origin in \( \mathbb{R}^m \):

\[
\mathbb{R}_{c}^{m \times n} \triangleq \{ Y \in \mathbb{R}^{m \times n} \mid Y1 = 0 \} \\
= \{ Y \in \mathbb{R}^{m \times n} \mid \mathcal{N}(Y) \supseteq 1 \} = \{ Y \in \mathbb{R}^{m \times n} \mid \mathcal{R}(Y^T) \subseteq \mathcal{N}(1^T) \} \\
= \{ XV \mid X \in \mathbb{R}^{m \times n} \} \subseteq \mathbb{R}^{m \times n} \\
\]

the nonsymmetric geometric center subspace. Further suppose \( V \in \mathbb{S}^n \) is a projection matrix having \( \mathcal{N}(V) = \mathcal{R}(1) \) and \( \mathcal{R}(V) = \mathcal{N}(1^T) \). Then linear mapping \( T(X) = XV \) is the orthogonal projection of any \( X \in \mathbb{R}^{m \times n} \) on \( \mathbb{R}_{c}^{m \times n} \) in the Euclidean (vectorization) sense because \( V \) is symmetric, \( \mathcal{N}(XV) \supseteq 1 \), and \( \mathcal{R}(VX^T) \subseteq \mathcal{N}(1^T) \).

Now suppose we define a subspace of symmetric \( n \times n \) matrices each of whose columns constitute a list having the origin in \( \mathbb{R}^n \) as geometric center,

\[
\mathbb{S}^n_{c} \triangleq \{ Y \in \mathbb{S}^n \mid Y1 = 0 \} \\
= \{ Y \in \mathbb{S}^n \mid \mathcal{N}(Y) \supseteq 1 \} = \{ Y \in \mathbb{S}^n \mid \mathcal{R}(Y) \subseteq \mathcal{N}(1^T) \} \\
\]

(2095)
the geometric center subspace. Further suppose \( V \in S^n \) is a projection matrix, the same as before. Then \( V X V \) is the orthogonal projection of any \( X \in S^n \) on \( S^n \) in the Euclidean sense (2091) because \( V \) is symmetric, \( V X V = 0 \), and \( R(VXV) \subseteq N(1^T) \). Two-sided projection is necessary only to remain in the ambient symmetric matrix subspace. Then
\[
S^n_c = \{ VXV \mid X \in S^n \} \subset S^n
\]
has \( \dim S^n_c = n(n-1)/2 \) in isomorphic \( \mathbb{R}^{n(n+1)/2} \). We find its orthogonal complement as the aggregate of all negative directions of orthogonal projection on \( S^n_c \): the translation-invariant subspace (§5.5.1.1)
\[
S^n_{c\perp} \triangleq \{ X - VXX \mid X \in S^n \} \subset S^n
= \{ u1^T + 1u^T \mid u \in \mathbb{R}^n \}
\]
characterized by doublet \( u1^T + 1u^T \) (§B.2). Defining geometric center mapping
\[
V(X) = -VXX\frac{1}{2}
\]
consistently with (1077), then \( N(V) = R(I - V) \) on domain \( S^n \) analogously to vector projectors (§E.2); id est,
\[
N(V) = S^n_{c\perp}
\]
a subspace of \( S^n \) whose dimension is \( \dim S^n_{c\perp} = n \) in isomorphic \( \mathbb{R}^{n(n+1)/2} \). Intuitively, operator \( V \) is an orthogonal projector; any argument duplicitously in its range is a fixed point. So, this symmetric operator’s nullspace must be orthogonal to its range.

Now compare the subspace of symmetric matrices having all zeros in the first row and column
\[
S^n_1 \triangleq \{ Y \in S^n \mid Ye_1 = 0 \}
\]
\[
= \left\{ \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} X \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} \mid X \in S^n \right\}
\]
\[
= \left\{ \begin{bmatrix} 0 & \sqrt{2}V \end{bmatrix}^T Z \begin{bmatrix} 0 & \sqrt{2}V \end{bmatrix} \mid Z \in S^n \right\}
\]
where \( P = \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} \) is an orthogonal projector. Then, similarly, \( PXP \) is the orthogonal projection of any \( X \in S^n \) on \( S^n_1 \) in the Euclidean sense (2091), and
\[
S^n_{1\perp} \triangleq \left\{ \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} X \begin{bmatrix} 0 & 0^T \\ 0 & I \end{bmatrix} - X \mid X \in S^n \right\} \subset S^n
\]
\[
= \{ u1^T + e_1u^T \mid u \in \mathbb{R}^n \}
\]
Obviously, \( S^n_1 \oplus S^n_{1\perp} = S^n \).

**E.14 Proof.**
\[
\{ X - VXV \mid X \in S^n \} = \{ X - (I - \frac{1}{n}11^T)X(I - 11^T\frac{1}{n}) \mid X \in S^n \}
= \{ \frac{1}{n}11^TX + 11^T\frac{1}{n} - \frac{1}{n}11^TX11^T\frac{1}{n} \mid X \in S^n \}
\]
Because \( \{ X1 \mid X \in S^n \} = \mathbb{R}^n \),
\[
\{ X - VXV \mid X \in S^n \} = \{ \zeta^T + \zeta1^T - 11^T(11^T\zeta\frac{1}{n}) \mid \zeta \in \mathbb{R}^n \}
= \{ \zeta^T(I - 11^T\frac{1}{2n}) + (I - \frac{1}{n}11^T)\zeta1^T \mid \zeta \in \mathbb{R}^n \}
\]
where \( I - \frac{1}{n}11^T \) is invertible.

\( \boxdot \)
E.8 Range/Rowspace interpretation

For idempotent matrices $P_1$ and $P_2$ of any rank, $P_1XP_2^T$ is a projection of $\mathcal{R}(X)$ on $\mathcal{R}(P_1)$ and a projection of $\mathcal{R}(X^T)$ on $\mathcal{R}(P_2)$: For any given $X = U\Sigma Q^T = \sum_{i=1}^{\eta} \sigma_i u_i q_i^T \in \mathbb{R}^{m \times p}$, as in compact SVD (1645),

$$P_1XP_2^T = \sum_{i=1}^{\eta} \sigma_i u_i q_i^T P_2^T = \sum_{i=1}^{\eta} \sigma_i P_1 u_i (P_2 q_i)^T$$  \hspace{1cm} (2101)

where $\eta \triangleq \min\{m, p\}$. Recall $u_i \in \mathcal{R}(X)$ and $q_i \in \mathcal{R}(X^T)$ when the corresponding singular value $\sigma_i$ is nonzero. (§A.6.1) So $P_1$ projects $u_i$ on $\mathcal{R}(P_1)$ while $P_2$ projects $q_i$ on $\mathcal{R}(P_2)$; id est, the range and rowspace of any $X$ are respectively projected on the ranges of $P_1$ and $P_2$.  \hspace{1cm} E.15

E.9 Projection on convex set

Thus far we have discussed only projection on subspaces. Now we generalize, considering projection on arbitrary convex sets in Euclidean space; convex because projection is, then, unique minimum-distance and a convex optimization problem:

For projection $P_C x$ of point $x$ on any closed set $C \subseteq \mathbb{R}^n$ it is obvious:

$$C \equiv \{ P_C x \mid x \in \mathbb{R}^n \} = \{ x \in \mathbb{R}^n \mid P_C x = x \}$$  \hspace{1cm} (2102)

where $P_C$ is a projection operator that is convex when $C$ is convex. \hspace{1cm} [63, p.88]

If $C \subseteq \mathbb{R}^n$ is a closed convex set, then for each and every $x \in \mathbb{R}^n$ there exists a unique point $P_C x$ belonging to $C$ that is closest to $x$ in the Euclidean sense. Like (2002), unique projection $P x$ (or $P_C x$) of a point $x$ on convex set $C$ is that point in $C$ closest to $x$; \hspace{1cm} [260, §3.12]

$$\|x - P x\|_2 = \inf_{y \in C} \|x - y\|_2 = \text{dist}(x, C)$$  \hspace{1cm} (2103)

There exists a converse (in finite-dimensional Euclidean space):

E.9.0.0.1 Theorem. (Bunt–Motzkin) Convex set if projections unique.  \hspace{1cm} [403, §7.5]  \hspace{1cm} [206] If $C \subseteq \mathbb{R}^n$ is a nonempty closed set and if for each and every $x$ in $\mathbb{R}^n$ there is a unique Euclidean projection $P x$ of $x$ on $C$ belonging to $C$, then $C$ is convex.  \hspace{1cm} \diamond

Borwein & Lewis propose, for closed convex set $C$ \hspace{1cm} [57, §3.3 exer.12d]

$$\nabla \|x - P x\|_2^2 = 2(x - P x)$$  \hspace{1cm} (2104)

for any point $x$ whereas, for $x \notin C$

$$\nabla \|x - P x\|_2 = (x - P x) \|x - P x\|_2^{-1}$$  \hspace{1cm} (2105)

\hspace{1cm} E.15 When $P_1$ and $P_2$ are symmetric and $\mathcal{R}(P_1) = \mathcal{R}(u_i)$ and $\mathcal{R}(P_2) = \mathcal{R}(q_i)$, then the $j^{th}$ dyad term from the singular value decomposition of $X$ is isolated by the projection. Yet if $\mathcal{R}(P_2) = \mathcal{R}(q_i)$, $\ell \neq j \in \{1 \ldots \eta\}$, then $P_1XP_2 = 0$.  \hspace{1cm}
E.9.0.0.2 Exercise. **Norm gradient.**
Prove (2104) and (2105). (Not proved in [57].) ▼

A well-known equivalent characterization of projection on a convex set is a generalization of the perpendicularly condition (2001) for projection on a subspace:

E.9.1 Dual interpretation of projection on convex set

E.9.1.0.1 Definition. **Normal vector.**
Vector $z$ is normal to convex set $\mathcal{C}$ at point $P_x \in \mathcal{C}$ if

$$\langle z, y - P_x \rangle \leq 0 \quad \forall y \in \mathcal{C} \quad (2106)$$

A convex set has at least one nonzero normal at each of its boundary points. [318, p.100] (Figure 69) Hence, the normal or dual interpretation of projection:

E.9.1.0.2 Theorem. **Unique minimum-distance projection.** [209, §A.3.1] [260, §3.12] [115, §4.1] [81] (Figure 182b p.645) Given a closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$, point $P_x$ is the unique projection of a given point $x \in \mathbb{R}^n$ on $\mathcal{C}$ ($P_x$ is that point in $\mathcal{C}$ nearest $x$) if and only if

$$P_x \in \mathcal{C}, \quad \langle x - P_x, y - P_x \rangle \leq 0 \quad \forall y \in \mathcal{C} \quad (2107)$$

As for subspace projection, convex operator $P$ is idempotent in the sense: for each and every $x \in \mathbb{R}^n$, $P(Px) = Px$. Yet operator $P$ is nonlinear;

- Projector $P$ is a linear operator if and only if convex set $\mathcal{C}$ (on which projection is made) is a subspace. (§E.4)

E.9.1.0.3 Theorem. **Unique projection via normal cone.**

Given closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$, point $P_x$ is the unique projection of a given point $x \in \mathbb{R}^n$ on $\mathcal{C}$ if and only if

$$P_x \in \mathcal{C}, \quad P_x - x \in (\mathcal{C} - P_x)^* \quad (2108)$$

In other words, $P_x$ is that point in $\mathcal{C}$ nearest $x$ if and only if $P_x - x$ belongs to that cone dual to translate $\mathcal{C} - P_x$. □

E.9.1.1 Dual interpretation as optimization

Deutsch [117, thm.2.3] [118, §2] and Luenberger [260, p.134] carry forward Nirenberg’s dual interpretation of projection [290] as solution to a maximization problem: Minimum distance from a point $x \in \mathbb{R}^n$ to a convex set $\mathcal{C} \subseteq \mathbb{R}^n$ can be found by maximizing distance from $x$ to hyperplane $\partial \mathcal{H}$ over the set of all hyperplanes separating $x$ from $\mathcal{C}$. Existence

---

\[ -(\mathcal{C} - P_x)^* \] is the normal cone to set $\mathcal{C}$ at point $P_x$. (§E.10.3.2)
Figure 176: Dual interpretation of projection of point $x$ on convex set $C$ in $\mathbb{R}^2$. 
(a) $\kappa = (a^T a)^{-1} (a^T x - \sigma_C(a))$. 
(b) Minimum distance from $x$ to $C$ is found by maximizing distance to all hyperplanes supporting $C$ and separating it from $x$. A convex problem for any convex set, distance of maximization is unique.
of a separating hyperplane (§2.4.2.7) presumes point \( x \) lies on the boundary or exterior to set \( C \).

The optimal separating hyperplane is characterized by the fact it also supports \( C \). Any hyperplane supporting \( C \) (Figure 31a) has form
\[
\partial H = \{ y \in \mathbb{R}^n \mid a^T y = \sigma_C(a) \} \tag{129}
\]
where the support function is convex, defined
\[
\sigma_C(a) = \sup_{z \in C} a^T z \tag{554}
\]
When point \( x \) is finite and set \( C \) contains finite points, under this projection interpretation, if the supporting hyperplane is a separating hyperplane then the support function is finite. From Example E.5.0.0.8, projection \( P_{\partial H} x \) of \( x \) on any given supporting hyperplane \( \partial H \) is
\[
P_{\partial H} x = x - a(a^T a)^{-1}(a^T x - \sigma_C(a)) \tag{2109}
\]
With reference to Figure 176, identifying \( H_+ = \{ y \in \mathbb{R}^n \mid a^T y \geq \sigma_C(a) \} \) then
\[
\| x - P_C x \| = \sup_{\partial H_+ \mid x \in H_+} \| x - P_{\partial H} x \| = \sup_{a \mid x \in H_+} \| a(a^T a)^{-1}(a^T x - \sigma_C(a)) \|
\]
\[
= \sup_{a \mid x \in H_+} \frac{1}{\| a \|} | a^T x - \sigma_C(a) | \tag{2110}
\]
which can be expressed as a convex optimization, for arbitrary positive constant \( \tau \)
\[
\| x - P_C x \| = \frac{1}{\tau} \maximize_a \ a^T x - \sigma_C(a) \tag{2111}
\]
subject to \( \| a \| \leq \tau \)
The unique minimum-distance projection on convex set \( C \) is therefore
\[
P_C x = x - a^* (a^* x - \sigma_C(a^*)) \frac{1}{\tau^2} \tag{2112}
\]
where optimally \( \| a^* \| = \tau \).

**Exercise.** Dual projection technique on polyhedron.
Test that projection paradigm from Figure 176 on any convex polyhedral set.

**Dual interpretation of projection on cone**

In the circumstance set \( C \) is a closed convex cone \( K \) and there exists a hyperplane separating given point \( x \) from \( K \), then optimal \( \sigma_K(a^*) \) takes value 0 [200, §C.2.3.1]. So problem (2111) for projection of \( x \) on \( K \) becomes
\[
\| x - P_K x \| = \frac{1}{\tau} \maximize_a \ a^T x \tag{2113}
\]
subject to \( \| a \| \leq \tau \)
\[
a \in K^o
\]
E.9. PROJECTION ON CONVEX SET

The norm inequality in (2113) can be handled by Schur complement (§3.5.2). Normals \(a\) to all hyperplanes supporting \(K\) belong to the polar cone \(K^* = K^\circ\) by definition: (318)

\[
a \in K^\circ \iff \langle a, x \rangle \leq 0 \text{ for all } x \in K
\]

(2114)

Projection on cone \(K\) is

\[
P_Kx = (I - \frac{1}{\tau} a^* a^{**}T)x
\]

(2115)

whereas projection on the polar cone \(-K^*\) is (§E.9.2.2.1)

\[
P_{K^*}x = x - P_Kx = \frac{1}{\tau^2} a^* a^{**}T x
\]

(2116)

Negating vector \(a\), this maximization problem (2113) becomes a minimization (the same problem) and the polar cone becomes the dual cone:

\[
\|x - P_Kx\| = \frac{1}{\tau} \text{ minimize } a^T x \text{ subject to } \|a\| \leq \tau
\]

(2117)

E.9.2 Projection on cone

When convex set \(C\) is a cone, there is a finer statement of optimality conditions:

E.9.2.0.1 Theorem. Unique projection on cone. [209, §A.3.2]

Let \(K \subseteq \mathbb{R}^n\) be a closed convex cone, and \(K^*\) its dual (§2.13.1). Then \(P_x\) is the unique minimum-distance projection of \(x \in \mathbb{R}^n\) on \(K\) if and only if

\[
P_x \in K, \quad \langle P_x - x, P_x \rangle = 0, \quad P_x - x \in K^*
\]

(2118)

\[\Box\]

In words, \(P_x\) is the unique minimum-distance projection of \(x\) on \(K\) if and only if

1) projection \(P_x\) lies in \(K\)

2) direction \(P_x - x\) is orthogonal to the projection \(P_x\)

3) direction \(P_x - x\) lies in the dual cone \(K^*\).

As the theorem is stated, it admits projection on \(K\) not full-dimensional; id est, on closed convex cones in a proper subspace of \(\mathbb{R}^n\).

Projection on \(K\) of any point \(x \in -K^*\), belonging to the negative dual cone, is the origin. By (2118): the set of all points reaching the origin, when projecting on \(K\), constitutes the negative dual cone: a.k.a., the polar cone

\[
K^\circ = -K^* = \{x \in \mathbb{R}^n \mid P_x = 0\}
\]

(2119)
Figure 177: (confer Figure 87, Figure 178) Given orthogonal projection $(I - P)x$ of $x$ on orthogonal complement $\mathcal{R}(P)^\perp$, projection on $\mathcal{R}(P)$ is immediate: $x - (I - P)x$.

### E.9.2.1 Relationship to subspace projection

Conditions 1 and 2 of the theorem are common with orthogonal projection on a subspace $\mathcal{R}(P)$:

Condition 1 corresponds to the most basic requirement; namely, the projection $P \in \mathcal{R}(P)$ belongs to the subspace (confer (2102))

$$\mathcal{K} = \{ P \in \mathcal{R}(P) \} \triangleq \mathcal{R}(P) \quad (2120)$$

Recall the perpendicularity requirement for projection on a subspace:

$$P \perp \mathcal{R}(P) \quad \text{or} \quad P = \mathcal{R}(P)^\perp$$

which corresponds to condition 2.

Yet condition 3 is a generalization of subspace projection; *id est*, for unique minimum-distance projection on a closed convex cone $\mathcal{K}$, polar cone $-\mathcal{K}^*$ (Figure 178) plays the role $\mathcal{R}(P)^\perp$ plays for subspace projection (Figure 177: $P_{\mathcal{R}(P)^\perp} = x - P_{\mathcal{R}(P)^\perp}$). Indeed, $-\mathcal{K}^*$ is the algebraic complement in the orthogonal vector sum (p.662) [279, §A.3.2.5]

$$\mathcal{K} \oplus -\mathcal{K}^* = \mathbb{R}^n \iff \text{cone } \mathcal{K} \text{ is closed and convex} \quad (2121)$$

Given unique minimum-distance projection $P \in \mathcal{K}$ satisfying the theorem (§E.9.2.0.1), then by projection on the algebraic complement via $I - P$ in §E.2 we have

$$-\mathcal{K}^* = \{ x - P \in \mathbb{R}^n \} = \{ x \in \mathbb{R}^n \mid P \in \mathcal{N}(P) \} = \mathcal{N}(P) \quad (2122)$$
E.9. PROJECTION ON CONVEX SET

Consequent to Moreau (2125). Converse (2122)(2120) ⇒ (2121) holds as well. Recalling that any subspace is a closed convex cone

\[ \mathcal{K} = \mathcal{R}(P) \iff -\mathcal{K}^* = \mathcal{R}(P)^\perp \]  

(2123)

meaning, when a cone is a subspace \( \mathcal{R}(P) \), then the dual cone becomes its orthogonal complement \( \mathcal{R}(P)^\perp \). In this circumstance, condition 3 becomes coincident with condition 2.

Properties of projection on cones, following in §E.9.2.2, further generalize to subspaces by: (4)

\[ \mathcal{K} = \mathcal{R}(P) \iff -\mathcal{K} = \mathcal{R}(P) \]  

(2124)

E.9.2.2 Salient properties: Projection \( P_x \) on closed convex cone \( \mathcal{K} \)

[209, §A.3.2] [115, §5.6] For \( x, x_1, x_2 \in \mathbb{R}^n \)

1. \( P_\mathcal{K}(\alpha x) = \alpha P_\mathcal{K}x \quad \forall \alpha \geq 0 \) (nonnegative homogeneity)

2. \( \|P_\mathcal{K}x\| \leq \|x\| \)

3. \( P_\mathcal{K}x = 0 \iff x \in -\mathcal{K}^* \)

4. \( P_\mathcal{K}(-x) = -P_{-\mathcal{K}}x \)

5. (Jean-Jacques Moreau, 1962) [279]

\[ x = x_1 + x_2, \quad x_1 \in \mathcal{K}, \quad x_2 \in -\mathcal{K}^*, \quad x_1 \perp x_2 \]

\[ \iff \]

\[ x_1 = P_{\mathcal{K}}x, \quad x_2 = P_{-\mathcal{K}}x \]  

(2125)

\[ \mathcal{K} = \{x - P_{-\mathcal{K}}x \mid x \in \mathbb{R}^n\} = \{x \in \mathbb{R}^n \mid P_{-\mathcal{K}}x = 0\} \]

6. \( -\mathcal{K}^* = \{x - P_\mathcal{K}x \mid x \in \mathbb{R}^n\} = \{x \in \mathbb{R}^n \mid P_\mathcal{K}x = 0\} \)  

(2122)

E.9.2.2.1 Corollary. \( I-P \) for cones. \( \text{(confer §E.2)} \)

Denote by \( \mathcal{K} \subseteq \mathbb{R}^n \) a closed convex cone, and call \( \mathcal{K}^* \) its dual. Then \( x - P_{-\mathcal{K}}x \) is the unique minimum-distance projection of \( x \in \mathbb{R}^n \) on \( \mathcal{K} \) if and only if \( P_{-\mathcal{K}}x \) is the unique minimum-distance projection of \( x \) on \( -\mathcal{K}^* \) the polar cone.

E.17 but a proper subspace is not a proper cone (§2.7.2.2.1).
Given minimum-distance projection \( (I-P)x \) of \( x \) on negative dual cone \( K^\circ \), projection on \( K \) is immediate: \( x - (I-P)x = Px \).

**Proof.** Assume \( x_1 = P_K x \). Then by Theorem E.9.2.0.1 we have

\[
x_1 \in K, \quad x_1 - x \perp x_1, \quad x_1 - x \in K^* \quad (2126)
\]

Now assume \( x - x_1 = P_{-K^*} x \). Then we have

\[
x - x_1 \in -K^*, \quad -x_1 \perp x - x_1, \quad -x_1 \in -K \quad (2127)
\]

But these two assumptions are apparently identical. We must therefore have

\[
x - P_{-K^*} x = x_1 = P_K x \quad (2128)
\]

\[\diamondsuit\]

**E.9.2.2.2 Corollary.** Unique projection via dual or normal cone. [115, §4.7] (§E.10.3.2, confer Theorem E.9.1.0.3) Given point \( x \in \mathbb{R}^n \) and closed convex cone \( K \subseteq \mathbb{R}^n \), the following are equivalent statements:

1. point \( Px \) is the unique minimum-distance projection of \( x \) on \( K \)
2. \( Px \in K, \quad x - Px \in -(K - Px)^* = -K^* \cap (Px)^\perp \)
3. \( Px \in K, \quad \langle x - Px, Px \rangle = 0, \quad \langle x - Px, y \rangle \leq 0 \ \forall y \in K \quad \diamondsuit\)
E.9.2.2.3 Example. Unique projection on nonnegative orthant. (confer 1402)

To project matrix \( H \in \mathbb{R}^{m \times n} \) on the selfdual orthant (§2.13.5.1) of nonnegative matrices \( \mathbb{R}_+^{m \times n} \) in isomorphic \( \mathbb{R}^{mn} \), from Theorem E.9.2.0.1, necessary and sufficient conditions are:

\[
\begin{align*}
H^* & \succeq 0 \\
\text{tr} \left( (H^* - H)^T H^* \right) & = 0 \\
H^* - H & \succeq 0
\end{align*}
\]  

(2129)

where the inequalities denote entrywise comparison. The optimal solution \( H^* \) is simply \( H \) having all its negative entries zeroed:

\[
H^*_{ij} = \max \{ H_{ij}, 0 \}, \quad i, j \in \{1 \ldots m\} \times \{1 \ldots n\}
\]  

(2130)

Now suppose the nonnegative orthant is translated by \( T \in \mathbb{R}^{m \times n} \); id est, consider \( \mathbb{R}_+^{m \times n} + T \). Then projection on the translated orthant is [115, §4.8]

\[
H^*_{ij} = \max \{ H_{ij}, T_{ij} \}
\]  

(2131)

\[
\square
\]

E.9.2.2.4 Example. Unique projection on truncated convex cone.

Consider the problem of projecting a point \( x \) on a closed convex cone that is artificially bounded; really, a bounded convex polyhedron having a vertex at the origin:

\[
\begin{align*}
\text{minimize} & \quad \| x - Ay \|_2 \\
\text{subject to} & \quad y \succeq 0 \\
& \quad \| y \|_\infty \leq 1
\end{align*}
\]  

(2132)

where the convex cone has vertex-description (§2.12.2.0.1), for \( A \in \mathbb{R}^{n \times N} \)

\[
\mathcal{K} = \{ Ay \ | \ y \succeq 0 \}
\]  

(2133)

and where \( \| y \|_\infty \leq 1 \) is the artificial bound. This is a convex optimization problem having no known closed-form solution, in general. It arises, for example, in the fitting of hearing aids designed around a programmable graphic equalizer (a filter bank whose only adjustable parameters are gain per band each bounded above by unity). [99] [100] The problem is equivalent to a Schur-form semidefinite program (§3.5.2)

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad \begin{bmatrix} tI & x - Ay \\ (x - Ay)^T & t \end{bmatrix} \succeq 0 \\
& \quad 0 \leq y \leq 1
\end{align*}
\]  

(2134)

\[
\square
\]
E.9.3 nonexpansivity

**E.9.3.0.1 Theorem. Nonexpansivity.** [185, §2] [115, §5.3] When $C \subset \mathbb{R}^n$ is an arbitrary closed convex set, projector $P$ projecting on $C$ is nonexpansive in the sense: for any vectors $x, y \in \mathbb{R}^n$

$$\|Px - Py\| \leq \|x - y\|$$

with equality when $x - Px = y - Py$.

**Proof.** [56]

$$\|x - y\|^2 = \|Px - Py\|^2 + \|(I - P)x - (I - P)y\|^2 + 2\langle x - Px, Px - Py \rangle + 2\langle y - Py, Py - Px \rangle$$

Nonnegativity of the last two terms follows directly from the unique minimum-distance projection theorem (§E.9.1.0.2).

The foregoing proof reveals another flavor of nonexpansivity: for each and every $x, y \in \mathbb{R}^n$

$$\|Px - Py\|^2 + \|(I - P)x - (I - P)y\|^2 \leq \|x - y\|^2$$

Deutsch shows yet another: [115, §5.5]

$$\|Px - Py\|^2 \leq \langle x - y, Px - Py \rangle$$

E.9.4 Easy projections

- To project any matrix $H \in \mathbb{R}^{n \times n}$ orthogonally in Euclidean/Frobenius sense on subspace of symmetric matrices $S_n$ in isomorphic $\mathbb{R}^{n^2}$, take symmetric part of $H$; (§2.2.2.0.1) *id est*, $(H + H^T)/2$ is the projection.

- To project any $H \in \mathbb{R}^{n \times n}$ orthogonally on symmetric hollow subspace $S_n^h$ in isomorphic $\mathbb{R}^{n^2}$ (§2.2.3.0.1, §7.0.1), take symmetric part then zero all entries along main diagonal or *vice versa* (because this is projection on intersection of two subspaces); *id est*, $(H + H^T)/2 - \delta(H)$.

- To project a matrix on nonnegative orthant $\mathbb{R}_+^{m \times n}$, simply clip all negative entries to 0. Likewise, projection on nonpositive orthant $\mathbb{R}_-^{m \times n}$ sees all positive entries clipped to 0. Projection on other orthants is equally simple with appropriate clipping.

- Projecting on hyperplane, halfspace, slab: §E.5.0.0.8.

- Projection of $y \in \mathbb{R}^n$ on Euclidean ball $B = \{x \in \mathbb{R}^n \mid \|x - a\| \leq c\}$: for $y \neq a$, $P_B y = \frac{c}{\|y - a\|}(y - a) + a$.

---

This condition for equality corrects an error in [81] (where the norm is applied to each side of the condition given here) easily revealed by counterexample.
• Clipping in excess of \(|1|\) each entry of point \(x \in \mathbb{R}^n\) is equivalent to unique minimum-distance projection of \(x\) on a hypercube centered at the origin. (confer §E.10.3.2)

• Projection of \(x \in \mathbb{R}^n\) on a (rectangular) hyperbox: [63, §8.1.1]

\[
C = \{ y \in \mathbb{R}^n \mid l \leq y \leq u, \ l < u \}
\]

\[
P(x)_{k=0 \ldots n} = \begin{cases} 
  l_k, & x_k \leq l_k \\
  x_k, & l_k \leq x_k \leq u_k \\
  u_k, & x_k \geq u_k 
\end{cases}
\]

• Orthogonal projection of \(x\) on a Cartesian subspace, whose basis is some given subset of the Cartesian axes, zeroes entries corresponding to the remaining (complementary) axes.

• Projection of \(x\) on set of all cardinality-\(k\) vectors \(\{ y \mid \text{card } y \leq k \}\) keeps \(k\) entries of greatest magnitude and clips to 0 those remaining.

• Unique minimum-distance projection of \(H \in \mathbb{S}^n\) on positive semidefinite cone \(\mathbb{S}^n_+\) in Euclidean/Frobenius sense is accomplished by eigenvalue decomposition (diagonalization) followed by clipping all negative eigenvalues to 0.

• Unique minimum-distance projection on generally nonconvex subset of all matrices belonging to \(\mathbb{S}^n_+\) having rank not exceeding \(\rho\) (§2.9.2.1) is accomplished by clipping all negative eigenvalues to 0 and zeroing smallest nonnegative eigenvalues keeping only \(\rho\) largest. (§7.1.2)

• Unique minimum-distance projection, in Euclidean/Frobenius sense, of \(H \in \mathbb{R}^{m \times n}\) on generally nonconvex subset of all \(m \times n\) matrices of rank no greater than \(k\) is singular value decomposition (§A.6) of \(H\) having all singular values beyond \(k^{\text{th}}\) zeroed. This is also a solution to projection in sense of spectral norm. [338, p.79, p.208]

• Projection of a real vector on the monotone nonnegative cone is identical to its projection on the monotone cone followed by clipping all negative entries of the result to 0. [286, §5]

• Projection on monotone nonnegative cone \(\mathcal{K}_{\mathcal{M}+} \subset \mathbb{R}^n\) in less than one cycle (in sense of alternating projections §E.10): [388].

• Fast projection on a simplicial cone: [395].

• Projection on closed convex cone \(\mathcal{K}\) of any point \(x \in -\mathcal{K}^*\), belonging to polar cone, is equivalent to projection on origin. (§E.9.2)

\[
P_{\mathbb{S}^n_+ \cap \mathbb{S}_c^n} = P_{\mathbb{S}^n_+} P_{\mathbb{S}_c^n} \tag{1346}
\]

\[
P_{\mathbb{R}^{n \times n}_+ \cap \mathbb{S}_c^k} = P_{\mathbb{R}^{n \times n}_+} P_{\mathbb{S}_c^k} \tag{7.0.1.1}
\]

\[
P_{\mathbb{R}^{n \times n}_+ \cap \mathbb{S}^n} = P_{\mathbb{R}^{n \times n}_+} P_{\mathbb{S}^n} \tag{E.9.5}
\]
APPENDIX E. PROJECTION

Figure 179: Closed convex set $C$ belongs to subspace $\mathbb{R}^n$ (shown bounded in sketch and drawn without proper perspective). Point $y$ is unique minimum-distance projection of $x$ on $C$; equivalent to product of orthogonal projection of $x$ on $\mathbb{R}^n$ and minimum-distance projection of result $z$ on $C$.

E.9.4.0.1 Exercise. Projection on spectral norm ball.
Find the unique minimum-distance projection on the convex set of all $m \times n$ matrices whose largest singular value does not exceed 1; id est, on \( \{ X \in \mathbb{R}^{m \times n} \mid \|X\|_2 \leq 1 \} \) the spectral norm ball (§2.3.2.0.5).

E.9.4.1 notes
Projection on Lorentz (second-order) cone: [63, exer.8.3c].
Deutsch [118] provides an algorithm for projection on polyhedral cones.
Youla [424, §2.5] lists eleven “useful projections”, of square-integrable uni- and bivariate real functions on various convex sets, in closed form.
Unique minimum-distance projection on an ellipsoid: Example 4.6.0.0.1.
Numerical algorithms for projection on the nonnegative simplex and 1-norm ball are in [133].

E.9.5 Projection on convex set in subspace
Suppose a convex set $C$ is contained in some subspace $\mathbb{R}^n$. Then unique minimum-distance projection of any point in $\mathbb{R}^n \oplus \mathbb{R}^{n\perp}$ on $C$ can be accomplished by first projecting orthogonally on that subspace, and then projecting the result on $C$; [115, §5.14] id est, the ordered product of two individual projections that is not commutable.
Proof. To show that, suppose unique minimum-distance projection $P_C x$ on $C \subset \mathbb{R}^n$ is $y$ as illustrated in Figure 179:

$$\|x - y\| \leq \|x - q\| \quad \forall q \in C \quad (2141)$$

Further suppose $P_{\mathbb{R}^n} x$ equals $z$. By the Pythagorean theorem

$$\|x - y\|^2 = \|x - z\|^2 + \|z - y\|^2 \quad (2142)$$

because $x - z \perp z - y$. (2001) [260, §3.3] Then point $y = P_C x$ is the same as $P_C z$ because

$$\|z - y\|^2 = \|x - y\|^2 - \|x - z\|^2 \leq \|z - q\|^2 = \|x - q\|^2 - \|x - z\|^2 \quad \forall q \in C \quad (2143)$$

which holds by assumption (2141).

$(\Rightarrow)$ Now suppose $z = P_{\mathbb{R}^n} x$ and

$$\|z - y\| \leq \|z - q\| \quad \forall q \in C \quad (2144)$$

meaning $y = P_C z$. Then point $y$ is identical to $P_C x$ because

$$\|x - y\|^2 = \|x - z\|^2 + \|z - y\|^2 \leq \|x - q\|^2 = \|x - z\|^2 + \|z - q\|^2 \quad \forall q \in C \quad (2145)$$

by assumption (2144).

This proof is extensible via translation argument. (§E.4) Unique minimum-distance projection on a convex set contained in an affine subset is, therefore, similarly accomplished.

Projecting matrix $H \in \mathbb{R}^{n \times n}$ on convex cone $K = \mathbb{S}^n \cap \mathbb{R}^{n \times n}$ in isomorphic $\mathbb{R}^n$ can be accomplished, for example, by first projecting on $\mathbb{S}^n$ and only then projecting the result on $\mathbb{R}^{n \times n}$ (confer §7.0.1). This is because that projection product is equivalent to projection on the subset of the nonnegative orthant in the symmetric matrix subspace.

E.10 Alternating projection

Alternating projection is an iterative technique for finding a point in the intersection of a number of arbitrary closed convex sets $C_k$, or for finding the distance between two nonintersecting closed convex sets. Because it can sometimes be difficult or inefficient to compute the intersection or express it analytically, one naturally asks whether it is possible to instead project (unique minimum-distance) alternately on the individual $C_k$; often easier and what motivates adoption of this technique. Once a cycle of alternating projections (an iteration) is complete, we then iterate (repeat the cycle) until convergence.

If the intersection of two closed convex sets is empty, then by convergence we mean the iterate (the result after a cycle of alternating projections) settles to a point of minimum distance separating the sets.

While alternating projection can find the point in the nonempty intersection closest to a given point $b$, it does not necessarily find the closest point. Finding that closest point is made dependable by an elegantly simple enhancement via correction to the alternating
projection technique: this Dykstra algorithm \((\text{2184})\) for projection on the intersection is one of the most beautiful projection algorithms ever discovered. It is accurately interpreted as the discovery of what alternating projection originally sought to accomplish: unique minimum-distance projection on the nonempty intersection of a number of arbitrary closed convex sets \(C_k\). Alternating projection is, in fact, a special case of the Dykstra algorithm whose discussion we defer until §E.10.3.

E.10.0.1 commutative projectors

A product of projection operators is generally not another projector. Given two arbitrary convex sets \(C_1\) and \(C_2\) and their respective minimum-distance projection operators \(P_1\) and \(P_2\): If projectors commute (\(P_1P_2 = P_2P_1\)) for each and every \(x \in \mathbb{R}^n\), then it is easy to show \(P_1P_2x \in C_1 \cap C_2\) and \(P_2P_1x \in C_1 \cap C_2\). When projectors commute, their product is a projector; some point in the intersection can be found in a finite number of steps. While commutativity is a sufficient condition, it is not necessary; e.g., §6.8.1.1.1.

When \(C_1\) and \(C_2\) are subspaces, in particular, projectors \(P_1\) and \(P_2\) commute if and only if \(P_1P_2 = P_2P_1 \subseteq \) or iff \(P_2P_1 = P_2 \subseteq \) or iff \(P_1P_2 = P_1 \subseteq \). When subspace projectors commute, this means we can find a point from the intersection of those subspaces in a finite number of steps; in fact, the closest point. Orthogonal projection on orthogonal subspaces (or intersecting orthogonal affine subsets), in particular, can be performed in any order to find the closest point in their intersection in a number of steps = number of subspaces (or affine subsets).

\[ E.10.0.1.1 \text{ Theorem. Kronecker projector:} \] \([337, \S2.7]\)

Given any projection matrices \(P_1\) and \(P_2\) (subspace projectors), then

\[ P_1 \otimes P_2 \quad \text{and} \quad P_1 \otimes I \quad \text{(2146)} \]

are projection matrices. The product preserves symmetry if present.

\[ \Box \]

E.10.0.2 noncommutative projectors

Typically, one considers the method of alternating projection when projectors do not commute; \textit{id est}, when \(P_1P_2 \neq P_2P_1\).

The iconic example for noncommutative projectors illustrated in Figure \(180\) shows the iterates converging to the closest point in the intersection of two arbitrary convex sets. Yet simple examples like Figure \(181\) reveal that noncommutative alternating projection does not always yield the closest point, although we shall show it always yields some point in the intersection or a point that attains the distance between two convex sets.

Alternating projection is also known as \textit{successive projection} \([188]\) \([185]\) \([65]\), \textit{cyclic projection} \([155]\) \([270, \S3.2]\), \textit{successive approximation} \([81]\), or \textit{projection on convex sets} \([335]\) \([336, \S6.4]\). It is traced back to von Neumann, 1933 \([380]\), and later Wiener \([408]\) who showed that higher iterates of a product of two orthogonal projections on subspaces converge at each point in the ambient space to the unique minimum-distance projection on the intersection of the two subspaces. More precisely, if \(R_1\) and \(R_2\) are closed subspaces
Figure 180: First several alternating projections, in von Neumann-style projection (2157) of point \( b \), converging on closest point \( P_b \) in intersection of two closed convex sets in \( \mathbb{R}^2 \): \( C_1 \) and \( C_2 \) partially drawn in vicinity of their intersection. For this particular example, it is possible to start anywhere in a large neighborhood of \( b \) and still converge to \( P_b \). Pointed normal cone \( K_{\perp} \) at \( P_b \) (448) is translated to \( P_b \) the unique minimum-distance projection of \( b \) on intersection. Alternating projections are themselves robust with respect to significant noise because they belong to this translated normal cone.

of a Euclidean space and \( P_1 \) and \( P_2 \) respectively denote orthogonal projection on \( \mathcal{R}_1 \) and \( \mathcal{R}_2 \), then for each vector \( b \) in that space,

\[
\lim_{i \to \infty} (P_1 P_2)^i b = P_{\mathcal{R}_1 \cap \mathcal{R}_2} b
\]

(2147)

Deutsch [115, thm.9.8, thm.9.35] shows rate of convergence for subspaces to be geometric [423, §1.4.4]; bounded above by \( \kappa^{2i+1} \| b \| \), \( i = 0, 1, 2 \ldots \), where \( 0 \leq \kappa < 1 \):

\[
\| (P_1 P_2)^i b - P_{\mathcal{R}_1 \cap \mathcal{R}_2} b \| \leq \kappa^{2i+1} \| b \|
\]

(2148)

This means convergence can be slow when \( \kappa \) is close to 1. Rate of convergence on intersecting halfspaces is also geometric. [116] [307]

This von Neumann sense of alternating projection may be applied to convex sets that are not subspaces, although convergence is not necessarily to the unique minimum-distance projection on the intersection. Figure 180 illustrates one application where convergence is reasonably geometric and the result is the unique minimum-distance projection. Figure 181, in contrast, demonstrates convergence in one iteration to a fixed point (of the projection product)\( ^{E.19} \) in the intersection of two halfspaces; a.k.a, feasibility problem.

It was Dykstra who in 1983 [134] (§E.10.3) first solved this projection problem.

\( ^{E.19} \) Fixed point of mapping \( T : \mathbb{R}^n \to \mathbb{R}^n \) is a point \( x \) whose image is identical under the map; id est, \( Tx = x \).
Figure 181: The sets \( \{C_k\} \) in this example comprise two halfspaces \( H_1 \) and \( H_2 \). The von Neumann-style alternating projection in \( \mathbb{R}^2 \) quickly converges to \( P_1P_2b \) (feasibility). Unique minimum-distance projection on the intersection is, of course, \( Pb \).

**E.10.0.3 the bullets**

Alternating projection has, therefore, various meaning dependent on the application or field of study; it may be interpreted to be: a distance problem, a feasibility problem (von Neumann), or a projection problem (Dykstra):

- **Distance.** Figure 182a-b. Find a unique point of projection \( P_1b \in C_1 \) that attains the distance between any two closed convex sets \( C_1 \) and \( C_2 \):

\[
\|P_1b - b\| = \text{dist}(C_1, C_2) \triangleq \inf_{z \in C_2} \|P_1z - z\| \tag{2149}
\]

- **Feasibility.** Figure 182c, \( \bigcap C_k \neq \emptyset \). Given a number \( L \) of indexed closed convex sets \( C_k \subset \mathbb{R}^n \), find any fixed point in their intersection by iterating (i) a projection product starting from \( b \):

\[
\left( \prod_{i=1}^{\infty} \prod_{k=1}^{L} P_k \right) b \in \bigcap_{k=1}^{L} C_k \tag{2150}
\]

- **Optimization.** Figure 182c, \( \bigcap C_k \neq \emptyset \). Given a number of indexed closed convex sets \( C_k \subset \mathbb{R}^n \), uniquely project a given point \( b \) on \( \bigcap C_k \):

\[
\| Pb - b \| = \inf_{x \in \bigcap C_k} \| x - b \| \tag{2151}
\]
Figure 182:
(a) **(distance)** Intersection of two convex sets in $\mathbb{R}^2$ is empty. Method of alternating projection would be applied to find that point in $C_1$ nearest $C_2$.

(b) **(distance)** Given $b \in C_2$, by projection theorem E.9.1.0.2, point $\bullet P_1 b \in C_1$ is nearest $b$ iff $(b - P_1 b)^T(y - P_1 b) \leq 0 \forall y \in C_1$. When $P_1 b$ attains distance between the two sets, hyperplane $\{y | (b - P_1 b)^T(y - P_1 b) = 0\}$ separates $C_1$ from $C_2$. [63, §2.5.1]

(c) **(0 distance)** Intersection is nonempty; distance between sets equals 0.

**(feasibility)** We may just want a fixed point of projection product $P_1 P_2 b$ in $\bigcap C_k$.

**(optimization)** Or we may want that point $P b$ in $\bigcap C_k$ nearest point $b$. 
E.10.1 Distance and existence

Existence of a fixed point is established:

**E.10.1.0.1 Theorem.** Distance. [81]

Given any two closed convex sets $C_1$ and $C_2$ in $\mathbb{R}^n$, then $P_1b \in C_1$ is a fixed point of projection product $P_1P_2$ if and only if $P_1b$ is a point of $C_1$ nearest $C_2$.

\[ \begin{array}{c}
(b - a)^T(u - a) \leq 0 \quad \forall u \in C_1 \\
(a - b)^T(v - b) \leq 0 \quad \forall v \in C_2 \\
\|a - b\| \leq \|u - v\| \quad \forall u \in C_1 \text{ and } \forall v \in C_2
\end{array} \]  

(2152)

by Cauchy-Schwarz inequality [318]

\[ \langle x, y \rangle \leq \|x\| \|y\| \]  

(2153)

(with equality iff $x = \kappa y$ where $\kappa \in \mathbb{R}$ (33) [237, p.137]).

Proof. ($\Rightarrow$) Given fixed point $a = P_1P_2a \in C_1$ with $b \notin P_2a \in C_2$ in tandem so that $a = P_1b$, then by the unique minimum-distance projection theorem (§E.9.1.0.2)

\[ \begin{array}{c}
(b - a)^T(u - a) \leq 0 \quad \forall u \in C_1 \\
(a - b)^T(v - a) \leq 0 \quad \forall v \in C_2 \\
\|a - b\| \leq \|u - v\| \quad \forall u \in C_1 \text{ and } \forall v \in C_2
\end{array} \]

Thus $a = P_1b$ (with $b = P_2a \in C_2$) is a fixed point in $C_1$ of the projection product $P_1P_2$. E.20

E.10.2 Feasibility and convergence

The set of all fixed points of any nonexpansive mapping is a closed convex set. [164, lem.3.4] [29, §1] The projection product $P_1P_2$ is nonexpansive by Theorem E.9.3.0.1 because, for any vectors $x, a \in \mathbb{R}^n$

\[ \|P_1P_2x - P_1P_2a\| \leq \|P_2x - P_2a\| \leq \|x - a\| \]  

(2155)

If the intersection of two closed convex sets $C_1 \cap C_2$ is empty, then the iterates converge to a point of minimum distance, a fixed point of the projection product. Otherwise, convergence is to some fixed point in their intersection (a feasible solution) whose existence is guaranteed by virtue of the fact that each and every point in the convex intersection is in one-to-one correspondence with fixed points of the nonexpansive projection product.

Bauschke & Borwein [29, §2] argue that any sequence monotonic in the sense of Fejér is convergent. E.21

---

E.20 point $b = P_2a$ can be shown, similarly, to be a fixed point of product $P_2P_1$.

E.21 Other authors prove convergence by different means; e.g. [185] [65].
E.10. ALTERNATING PROJECTION

E.10.2.0.1 Definition. Fejér monotonicity.

Given closed convex set $C \neq \emptyset$, then a sequence $x_i \in \mathbb{R}^n$, $i = 0, 1, 2 \ldots$, is monotonically Fejér with respect to $C$ if

$$
\|x_{i+1} - c\| \leq \|x_i - c\| \quad \text{for all } i \geq 0 \quad \text{and each and every } c \in C
$$

(2156)

Given $x_0 \triangleq b$, if we express each iteration of alternating projection by

$$
x_{i+1} = P_1 P_2 x_i, \quad i = 0, 1, 2 \ldots
$$

(2157)

and define any fixed point $a = P_1 P_2 a$, then sequence $x_i$ is Fejér monotone with respect to fixed point $a$ because

$$
\|P_1 P_2 x_i - a\| \leq \|x_i - a\| \quad \forall i \geq 0
$$

(2158)

by nonexpansivity. The nonincreasing sequence $\|P_1 P_2 x_i - a\|$ is bounded below hence convergent because any bounded monotonic sequence in $\mathbb{R}$ is convergent; [268, §1.2] [43, §1.1] $P_1 P_2 x_{i+1} = P_1 P_2 x_i = x_{i+1}$. Sequence $x_i$ therefore converges to some fixed point. If the intersection $C_1 \cap C_2$ is nonempty, convergence is to some point there by the distance theorem. Otherwise, $x_i$ converges to a point in $C_1$ of minimum distance to $C_2$.

E.10.2.0.2 Example. Hyperplane/orthant intersection.

Find a feasible solution (2150) belonging to the nonempty intersection of two convex sets: given $A \in \mathbb{R}^{m \times n}$, $\beta \in \mathbb{R}^m$, the nonnegative orthant intersecting affine subset $A$ (an intersection of hyperplanes). Projection of an iterate $x_i \in \mathbb{R}^n$ on $A$ is calculated

$$
P_2 x_i = x_i - A^T (AA^T)^{-1} (Ax_i - \beta)
$$

(2050)

while, thereafter, projection of the result on the orthant is simply

$$
x_{i+1} = P_1 P_2 x_i = \max \{0, P_2 x_i\}
$$

(2160)

where the maximum is entrywise (§E.9.2.2.3).

One realization of this problem in $\mathbb{R}^2$ is illustrated in Figure 183: For $A = [1 \ 1]$, $\beta = 1$, and $x_0 = b = [-3 \ 1/2]^T$, iterates converge to a feasible solution $Pb = [0 \ 1]^T$.

To give a more palpable sense of convergence in higher dimension, we do this example again but now we compute an alternating projection for the case $A \in \mathbb{R}^{400 \times 1000}$, $\beta \in \mathbb{R}^{400}$, and $b \in \mathbb{R}^{1000}$, all of whose entries are independently and randomly set to a uniformly distributed real number in the interval $[-1, 1]$. Convergence is illustrated in Figure 184.

This application of alternating projection to feasibility is extensible to any finite number of closed convex sets.
Figure 183: From Example E.10.2.0.2 in $\mathbb{R}^2$ showing von Neumann-style alternating projection to find feasible solution belonging to intersection of nonnegative orthant with hyperplane $\mathcal{A}$. Point $Pb$ lies at intersection of hyperplane with ordinate axis. In this particular example, feasible solution found is coincidentally optimal. Rate of convergence depends upon angle $\theta$; as it becomes more acute, convergence slows. [185, §3]

Figure 184: Example E.10.2.0.2 in $\mathbb{R}^{1000}$; geometric convergence of iterates in norm.
E.10.2.0.3 Example. Under- and over-projection. [60, §3]
Consider the following variation of alternating projection: We begin with some point \( x_0 \in \mathbb{R}^n \) then project that point on convex set \( C \) and then project that same point \( x_0 \) on convex set \( D \). To the first iterate we assign \( x_1 = \frac{1}{2}(P_C(x_0) + P_D(x_0)) \). More generally,
\[
x_{i+1} = \frac{1}{2}(P_C(x_i) + P_D(x_i)), \quad i = 0, 1, 2 \ldots
\]  
(2161)
Because the Cartesian product of convex sets remains convex, (§2.1.8) we can reformulate this problem.
Consider the convex set
\[
S \triangleq \left[ \begin{array}{c} C \\ D \end{array} \right]
\]  
(2162)
representing Cartesian product \( C \times D \). Now, those two projections \( P_C \) and \( P_D \) are equivalent to one projection on the Cartesian product; id est,
\[
P_S \left( \begin{bmatrix} x_1 \\ x_1 \end{bmatrix} \right) = \begin{bmatrix} P_C(x_i) \\ P_D(x_i) \end{bmatrix}
\]  
(2163)
Define the subspace
\[
R \triangleq \left\{ v \in \mathbb{R}^n \mathbb{R}^n \left| [I - I] v = 0 \right. \right\}
\]  
(2164)
By the results in Example E.5.0.0.6
\[
P_{RS} \left( \begin{bmatrix} x_i \\ x_i \end{bmatrix} \right) = P_R \left( \begin{bmatrix} P_C(x_i) \\ P_D(x_i) \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} P_C(x_i) + P_D(x_i) \\ P_C(x_i) + P_D(x_i) \end{bmatrix}
\]  
(2165)
This means the proposed variation of alternating projection is equivalent to an alternation of projection on convex sets \( S \) and \( R \). If \( S \) and \( R \) intersect, these iterations will converge to a point in their intersection; hence, to a point in the intersection of \( C \) and \( D \).
We need not apply equal weighting to the projections, as supposed in (2161). In that case, definition of \( R \) would change accordingly.

E.10.2.1 Relative measure of convergence
Inspired by Fejér monotonicity, the alternating projection algorithm from the example of convergence illustrated by Figure 184 employs a redundant sequence: The first sequence (indexed by \( j \)) estimates point \( \left( \prod_{j=1}^{L} \prod_{k=1}^{L} P_k \right) b \) in the presumably nonempty intersection of \( L \) convex sets, then the quantity
\[
\left\| x_i - \left( \prod_{j=1}^{L} \prod_{k=1}^{L} P_k \right) b \right\|
\]  
(2166)
in second sequence \( x_i \) is observed per iteration \( i \) for convergence. A priori knowledge of a feasible solution (2150) is both impractical and antithetical. We need another measure:
Nonexpansivity implies
\[
\left\| \left( \prod_{\ell=1}^{L} P_{\ell} \right) x_{k,i-1} - \left( \prod_{\ell=1}^{L} P_{\ell} \right) x_{ki} \right\| = \left\| x_{ki} - x_{k,i+1} \right\| \leq \left\| x_{k,i-1} - x_{ki} \right\| \tag{2167}
\]
where
\[
x_{ki} \triangleq P_{k} x_{k+1,i} \in \mathbb{R}^{n}, \quad x_{L+1,i} \triangleq x_{1,i-1} \tag{2168}
\]
x_{ki} represents unique minimum-distance projection of x_{k+1,i} on convex set k at iteration i. So a good convergence measure is total monotonic sequence
\[
\varepsilon_{i} \triangleq \sum_{k=1}^{L} \left\| x_{ki} - x_{k,i+1} \right\|, \quad i = 0, 1, 2 \ldots \tag{2169}
\]
where \( \lim_{i \to \infty} \varepsilon_{i} = 0 \) whether or not the intersection is nonempty.

**E.10.2.1.1 Example.** Affine subset \( \cap \) positive semidefinite cone.

Consider the problem of finding \( X \in \mathbb{S}^{n} \) that satisfies
\[
X \succeq 0, \quad \langle A_{j}, X \rangle = b_{j}, \quad j = 1 \ldots m \tag{2170}
\]
given nonzero \( A_{j} \in \mathbb{S}^{n} \) and real \( b_{j} \). Here we take \( C_{1} \) to be the positive semidefinite cone \( \mathbb{S}^{n}_{+} \) while \( C_{2} \) is the affine subset of \( \mathbb{S}^{n} \)
\[
C_{2} = \mathcal{A} \triangleq \{ X \mid \text{tr}(A_{j}X) = b_{j}, \ j = 1 \ldots m \} \subseteq \mathbb{S}^{n}
\]
\[
= \{ X \mid \begin{bmatrix} \text{svec}(A_{1})^{T} \\ \vdots \\ \text{svec}(A_{m})^{T} \end{bmatrix} \text{svec} X = b \} \tag{2171}
\]
\[
\triangleq \{ X \mid A \text{svec} X = b \}
\]
where \( b = [b_{j}] \in \mathbb{R}^{m}, \ A \in \mathbb{R}^{m \times n(n+1)/2} \), and symmetric vectorization \( \text{svec} \) is defined by (56). Projection of iterate \( X_{i} \in \mathbb{S}^{n} \) on \( \mathcal{A} \) is: (§E.5.0.0.6)
\[
P_{2} \text{svec} X_{i} = \text{svec} X_{i} - A^{\dagger}(A \text{svec} X_{i} - b) \tag{2172}
\]
Euclidean distance from \( X_{i} \) to \( \mathcal{A} \) is therefore
\[
dist(X_{i}, \mathcal{A}) = \| X_{i} - P_{2} X_{i} \|_{F} = \| A^{\dagger}(A \text{svec} X_{i} - b) \|_{2} \tag{2173}
\]
Projection of \( P_{2} X_{i} \triangleq \sum_{j} \lambda_{j} q_{j} q_{j}^{T} \) on the positive semidefinite cone (§7.1.2) is found from its eigenvalue decomposition (§A.5.1);
\[
P_{1} P_{2} X_{i} = \sum_{j=1}^{n} \max \{ 0, \lambda_{j} \} q_{j} q_{j}^{T} \tag{2174}
\]
Distance from \( P_2X_i \) to the positive semidefinite cone is therefore

\[
\text{dist}(P_2X_i, \mathbb{S}_+^n) = \|P_2X_i - P_1P_2X_i\|_F = \sqrt{\sum_{j=1}^{n} (\min\{0, \lambda_j\})^2}
\]  

(2175)

When the intersection is empty \( \mathcal{A} \cap \mathbb{S}_+^n = \emptyset \), the iterates converge to that positive semidefinite matrix closest to \( \mathcal{A} \) in the Euclidean sense. Otherwise, convergence is to some point in the nonempty intersection.

Barvinok (§2.9.3.0.1) shows that if a solution to (2170) exists, then there exists an \( X \in \mathcal{A} \cap \mathbb{S}_+^n \) such that

\[
\text{rank } X \leq \left\lfloor \frac{\sqrt{8m + 1} - 1}{2} \right\rfloor
\]  

(272)

\[\square\]

**Example.** *Semidefinite matrix completion.*

Continuing Example E.10.2.1.1: When \( m \leq n(n + 1)/2 \) and the \( A_j \) matrices are distinct members of the standard orthonormal basis \( \{E_{\ell q} \in \mathbb{S}^n\} \) (59)

\[
\{A_j \in \mathbb{S}^n, j=1\ldots m \} \subseteq \{E_{\ell q}\} = \left\{ \begin{array}{ll} e_\ell e_\ell^T, & \ell = q = 1\ldots n \\ \frac{1}{\sqrt{2}}(e_\ell e_q^T + e_q e_\ell^T), & 1 \leq \ell < q \leq n \end{array} \right\}
\]  

(2176)

and when the constants \( b_j \) are set to constrained entries of variable \( X \triangleq [X_{\ell q}] \in \mathbb{S}^n \)

\[
\{b_j, j=1\ldots m \} \subseteq \left\{ \begin{array}{ll} X_{\ell q}, & \ell = q = 1\ldots n \\ X_{\ell q}/\sqrt{2}, & 1 \leq \ell < q \leq n \end{array} \right\} = \{(X, E_{\ell q})\}
\]  

(2177)

then the equality constraints in (2170) fix individual entries of \( X \in \mathbb{S}^n \). Thus the feasibility problem becomes a *positive semidefinite matrix completion problem.* Projection of iterate \( X_i \in \mathbb{S}^n \) on \( \mathcal{A} \) simplifies to (confer (2172))

\[
P_2 \text{svec} X_i = \text{svec} X_i - A^T(A \text{svec} X_i - b)
\]  

(2178)

From this we can see that orthogonal projection is achieved simply by setting corresponding entries of \( P_2X_i \) to the known entries of \( X \), while the entries of \( P_2X_i \) remaining are set to corresponding entries of the current iterate \( X_i \).

Using this technique, we find a positive semidefinite completion for

\[
\begin{bmatrix}
4 & 3 & ? & 2 \\
3 & 4 & 3 & ? \\
? & 3 & 4 & 3 \\
2 & ? & 3 & 4
\end{bmatrix}
\]  

(2179)

Initializing the unknown entries to 0, they all converge geometrically to 1.5858 (rounded) after about 42 iterations.
Laurent gives a problem for which no positive semidefinite completion exists: [246]

\[
\begin{bmatrix}
1 & 1 & ? & 0 \\
1 & 1 & 1 & ? \\
? & 1 & 1 & 1 \\
0 & ? & 1 & 1
\end{bmatrix}
\]  

(2180)

Initializing unknowns to 0, by alternating projection we find the constrained matrix closest to the positive semidefinite cone,

\[
\begin{bmatrix}
1 & 1 & 0.5454 & 0 \\
1 & 1 & 1 & 0.5454 \\
0.5454 & 1 & 1 & 1 \\
0 & 0.5454 & 1 & 1
\end{bmatrix}
\]  

(2181)

and we find the positive semidefinite matrix closest to affine subset \( \mathcal{A} (2171) \):

\[
\begin{bmatrix}
1.0521 & 0.9409 & 0.5454 & 0.0292 \\
0.9409 & 1.0980 & 0.9451 & 0.5454 \\
0.5454 & 0.9451 & 1.0980 & 0.9409 \\
0.0292 & 0.5454 & 0.9409 & 1.0521
\end{bmatrix}
\]  

(2182)

These matrices (2181) and (2182) attain the Euclidean distance \( \text{dist}(\mathcal{A}, S^n) \). Convergence is illustrated in Figure 185.
Figure 186: \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) are the same halfspaces as in Figure 181. Dykstra’s alternating projection algorithm generates the alternations \( b, x_{21}, x_{11}, x_{22}, x_{12}, \ldots, x_{12} \). The path illustrated from \( b \) to \( x_{12} \) in \( \mathbb{R}^2 \) terminates at the desired result: \( Pb \) in Figure 181. The \( \{y_{ki}\} \) correspond to the first two difference vectors drawn (in the first iteration \( i = 1 \)), then oscillate between zero and a negative vector thereafter. These alternations are not so robust in presence of noise as for the example in Figure 180.

### E.10.3 Optimization and projection

Unique projection on the nonempty intersection of arbitrary convex sets, to find the closest point therein, is a convex optimization problem. The first successful application of alternating projection to this problem is attributed to Dykstra [134] [64] who in 1983 provided an elegant algorithm that prevails today. In 1988, Han [188] rediscovered the algorithm and provided a primal–dual convergence proof. A synopsis of the history of alternating projection\(^{\text{E.22}}\) can be found in [66] where it becomes apparent that Dykstra’s work is seminal.

#### E.10.3.1 Dykstra’s algorithm

Assume we are given some point \( b \in \mathbb{R}^n \) and closed convex sets \( \{C_k \subset \mathbb{R}^n \mid k = 1 \ldots L \} \). Let \( x_{ki} \in \mathbb{R}^n \) and \( y_{ki} \in \mathbb{R}^n \) respectively denote a primal and dual vector (whose meaning can be deduced from Figure 186 and Figure 187) associated with set \( k \) at iteration \( i \). Initialize

\[
y_{k0} = 0 \quad \forall k = 1 \ldots L \quad \text{and} \quad x_{1,0} = b
\]  

\(^{\text{E.22}}\) For a synopsis of alternating projection applied to distance geometry, see [365, §3.1].
Figure 187: Two examples (truncated): normal cone to $\mathcal{H}_1 \cap \mathcal{H}_2$ at the origin and at boundary point $Pb$. $\mathcal{H}_1$ and $\mathcal{H}_2$ are the same halfspaces from Figure 186. Normal cone at origin $K_{\mathcal{H}_1 \cap \mathcal{H}_2}(0)$ is simply $-K^\circ$.

Denoting by $P_k t$ the unique minimum-distance projection of $t$ on $C_k$, and for convenience $x_{L+1,i} = x_{1,i-1}$ (2168), iterate $x_{1,i}$ calculation proceeds:

\[
\text{for } i = 1, 2, \ldots \text{ until convergence } \{ \\
\quad \text{for } k = L \ldots 1 \{ \\
\quad \quad t = x_{k+1,i} - y_{k,i-1} \\
\quad \quad x_{ki} = P_k t \\
\quad \quad y_{ki} = P_k t - t \\
\quad \} \\
\}
\]  

(2184)

Assuming a nonempty intersection, then iterates converge to the unique minimum-distance projection of point $b$ on that intersection; [115, §9.24]

\[
Pb = \lim_{i \to \infty} x_{1,i}
\]  

(2185)

In the case that all $C_k$ are affine, then calculation of $y_{ki}$ is superfluous and the algorithm becomes identical to alternating projection. [115, §9.26] [155, §1] Dykstra’s algorithm is so simple, elegant, and represents such a tiny increment in computational intensity over alternating projection, it is nearly always arguably cost effective.

E.23 We reverse order of projection ($k = L \ldots 1$) in the algorithm for continuity of exposition.
E.10. ALTERNATING PROJECTION

E.10.3.2 Normal cone

Glunt [163, §4] observes that the overall effect of Dykstra’s iterative procedure is to drive \( t \) toward the translated normal cone to \( \bigcap C_k \) at the solution \( P b \) (translated to \( Pb \)). The normal cone gets its name from its graphical construction; which is, loosely speaking, to draw the outward-normals at \( Pb \) (Definition E.9.1.0.1) to all the convex sets \( C_k \) touching \( Pb \). Relative interior of the normal cone subtends these normal vectors.

E.10.3.2.1 Definition. Normal cone. \[ 278 \, 43, \, p.261 \, A.5.2 \, (\text{2.13.10}) \, 57, \, \text{EN} \, 2.1 \, 317, \, \text{EN} \, 3 \, 318, \, p.15 \]

The normal cone to any set \( S \subseteq \mathbb{R}^n \) at any particular point \( a \in \mathbb{R}^n \) is defined as the closed cone

\[
K^\perp_S(a) = \{ z \in \mathbb{R}^n \mid z^T (y-a) \leq 0 \ \forall y \in S \} = -(S-a)^* \tag{448}
\]

an intersection of halfspaces about the origin in \( \mathbb{R}^n \), hence convex regardless of convexity of \( S \); it is the negative dual cone to translate \( S - a \); the set of all vectors normal to \( S \) at \( a \) (§E.9.1.0.1).

Examples of normal cone construction are illustrated in Figure 69, Figure 187, Figure 188, and Figure 189. Normal cone at \( 0 \) in Figure 187 is the vector sum (§2.1.8) of two normal cones; [57, §3.3 exer.10] for \( H_1 \cap H_2 \neq \emptyset \)

\[
K_{H_1 \cap H_2}(0) = K_{H_1}(0) + K_{H_2}(0) \tag{2186}
\]

This formula applies more generally to other points in the intersection.

The normal cone to any affine set \( A \) at \( \alpha \in A \), for example, is the orthogonal complement of \( A - \alpha \). When \( A = 0 \), \( K^\perp_A(0) = A^\perp \) is \( \mathbb{R}^n \) the ambient space of \( A \).

Projection of any point in the translated normal cone \( K^\perp_C(a \in C) + a \) on convex set \( C \) is identical to \( a \) ; in other words, point \( a \) is that point in \( C \) closest to any point belonging to the translated normal cone \( K^\perp_C(a) + a \); e.g., Theorem E.4.0.0.1. The normal cone to convex cone \( K \) at the origin

\[
K^\perp_K(0) = -K^* \tag{2187}
\]

is the negative dual cone. Any point belonging to \( -K^* \), projected on \( K \), projects on the origin. More generally, [115, §4.5]

\[
K^\perp_K(a) = -(K - a)^* \tag{2188}
\]

\[
K^\perp_K(a \in K) = -K^* \cap a^\perp \tag{2189}
\]

The normal cone to \( \bigcap C_k \) at \( Pb \) in Figure 181 is ray \( \{ \xi(b - Pb) \mid \xi \geq 0 \} \) illustrated in Figure 187. Applying Dykstra’s algorithm to that example, convergence to the desired result is achieved in two iterations as illustrated in Figure 186. Yet applying Dykstra’s algorithm to the example in Figure 180 does not improve rate of convergence, unfortunately, because the given point \( b \) and all the alternating projections already belong to the translated normal cone at the vertex of intersection.
Figure 188: A few renderings (see next page) of normal cone $\mathcal{K}_{\mathcal{E}_3}^\perp$ to ellipope $\mathcal{E}_3$ (Figure 142), at point $11^T$, projected on $\mathbb{R}^3$. In [247, fig.2], normal cone is claimed circular in this dimension. (Numerical artifacts corrupt boundary and make truncated relative interior appear corporeal.)
E.10. ALTERNATING PROJECTION

Figure 189: Rough sketch of normal cone to set $\mathcal{C} \subset \mathbb{R}^2$ as $\mathcal{C}$ wanders toward infinity. A point at which a normal cone is determined, here the origin, need not belong to the set. Normal cone to $\mathcal{C}^1$ is a ray. But as $\mathcal{C}$ moves outward, normal cone approaches a halfspace.
E.10.3.3 speculation

Dykstra’s algorithm always converges at least as quickly as classical alternating projection, never slower [115], and it succeeds where alternating projection fails. Rate of convergence is wholly dependent on particular geometry of a given problem. From these few examples we surmise, unique minimum-distance projection on blunt (not sharp or acute, informally) full-dimensional polyhedral cones may be found by Dykstra’s algorithm in few iterations. But total number of alternating projections, constituting those iterations, can never be less than number of convex sets.
Appendix F

Notation, Definitions, Glossary

\( b \) (italic \( abcdefghijklmnopqrstuvwxyz \))
- column vector, scalar, or logical condition

\( b_i \)
- \( i^{th} \) entry of vector \( b = [b_i, i = 1 \ldots n] \) or \( i^{th} \) iterate of vector \( b \)

\( b_{i:j} \) or \( b(i:j) \), truncated vector comprising \( i^{th} \) through \( j^{th} \) entry of vector \( b \)

\( b_{k(i:j)} \) or \( b_{i:j,k} \), truncated vector comprising \( i^{th} \) through \( j^{th} \) entry of vector \( b_k \)

\( b^T \)
- vector transpose or row vector

\( b^H \)
- complex conjugate transpose \( b^{*^T} \)

\( A \)
- matrix, scalar, or logical condition (italic \( ABCDEFGHIJKLMNOPQRSTUVWXYZ \))

\( A^T \)
- Matrix transpose \([A_{ij}] \leftarrow [A_{ji}] \) is a linear operator. Regarding \( A \) as a linear operator, \( A^T \) is its adjoint.

\( A^{-T} \)
- matrix transpose of inverse; and \textit{vice versa} \( (A^{-1})^T = (A^T)^{-1} \)

\( A^{T_1} \)
- first of various transpositions of a cubix or quartix \( A \) (p.582, p.586)

\textit{skinny}
- a skinny matrix; meaning, more rows than columns: \[
\begin{bmatrix}
\end{bmatrix}.
\]

\textit{When there are more equations than unknowns, we say that the system} \( Ax = b \) \textit{is overdetermined} [168, §5.3]

\textit{fat}
- a fat matrix; meaning, more columns than rows: \[
\begin{bmatrix}
\end{bmatrix}.
\]

\textit{underdetermined}

\( A \)
- some set (calligraphic \( ABCDEFGHIJKLMNOPQRSTUVWXYZ \))

\textit{Dattorro, Convex Optimization} \textit{† Euclidean Distance Geometry} 2e, \textit{Meijo}, v2015.04.11. 659
\(\mathcal{F}(C \ni A)\) smallest face (171) that contains element \(A\) of set \(C\)

\(\mathcal{F}\) discrete Fourier transform (884)

\(\mathcal{G}(K)\) generators (§2.13.4.2.1) of set \(K\); any collection of points and directions whose hull constructs \(K\)

\(L^\nu\) level set (556)

\(L^\nu\) sublevel set (560)

\(L^\nu\) superlevel set (646)

\(A^{-1}\) inverse of matrix \(A\)

\(A^\dagger\) Moore-Penrose pseudoinverse of matrix \(A\)

\(\sqrt{\cdot}\) positive square root

\(\sqrt{x}\) entrywise positive square root of vector \(x\)

\(\sqrt{\cdot}\) positive \(\ell\)th root

\(A^{1/2}\) and \(\sqrt{A}\) \(A^{1/2}\) is any matrix such that \(A^{1/2}A^{1/2} = A\).

For \(A \in S^n_+\), \(\sqrt{A} \in S^n_+\) is unique and \(\sqrt{A}\sqrt{A} = A\). [57, §1.2] (§A.5.1.3)

\(\sqrt{D}\) Hadamard positive square root: \(D = \sqrt{D} \circ \sqrt{D}\)

\(E\) Euler Fraktur \(\mathcal{E}\) (506)

\(E\) Lagrangian (506)

\(E\) member of elliptope \(E_t\) (1178) parametrized by scalar \(t\)

\(E\) elliptope (1157)

\(E\) elementary matrix

\(E_{ij}\) member of standard orthonormal basis for symmetric (59) or symmetric hollow (75) matrices

\(A_{ij}\) or \(A(i, j)\), \(ij\)th entry of matrix \(A\) = \[
\begin{bmatrix}
1 & \cdot & \cdot \\
\cdot & 2 & \cdot \\
\cdot & \cdot & 3
\end{bmatrix}
\]
or rank-one matrix \(a_ia_j^T\) (§4.8)

\(A(i, j)\) \(A\) is a function of \(i\) and \(j\)

\(A_i\) \(i\)th matrix from a set or \(i\)th principal submatrix or \(i\)th iterate of \(A\)

\(A(i, :)\) \(i\)th row of matrix \(A\)

\(A(:, j)\) \(j\)th column of matrix \(A\) [168, §1.1.8]
$A_{i:j,k:ℓ}$ or $A(i:j,k:ℓ)$, submatrix taken from $i^{\text{th}}$ through $j^{\text{th}}$ row and $k^{\text{th}}$ through $ℓ^{\text{th}}$ column

*id est* from the Latin meaning *that is*

*e.g.* *exempli gratia*, from the Latin meaning *for sake of example*

*videlicet* from the Latin meaning *it is permitted to see*

*no.* *number*, from the Latin *numero*

*a.i.* *affinely independent* (§2.4.2.3)

*c.i.* *conically independent* (§2.10)

*l.i.* *linearly independent*

*w.r.t.* *with respect to*

*a.k.a.* *also known as*

*re* *real part*

*im* *imaginary part*

*i* or *j* $\sqrt{-1}$

$	ext{⊂} \; \text{⊂} \; \cap \; \cup$ standard set theory, *subset, superset, intersection, union*

*∈* membership, *element belongs to, or element is a member of*

*$∋$ membership, *contains as in* $C ∋ y$ (C contains element $y$)

*$϶$ *such that*

*$∃$ *there exists*

*$∴$ *therefore*

*$∀$ *for all, or over all*

*$&$ (ampersand) *and*

*$\&$ (ampersand italic) *and*

*$∝$ *proportional to*

*$∞$ *infinity*

*$≡$ *equivalent to*

*$≜$ *defined equal to, equal by definition*

*$≈$ *approximately equal to*
\(\simeq\) isomorphic to or with
\(\cong\) congruent to or with

Hadamard quotient as in, for \(x, y \in \mathbb{R}^n\),
\[
\frac{x}{y} \triangleq [x_i/y_i, \; i = 1 \ldots n] \in \mathbb{R}^n
\]

Hadamard product of matrices: \(x \circ y \triangleq [x_i y_i, \; i = 1 \ldots n] \in \mathbb{R}^n\)

Kronecker product of matrices (§D.1.2.1)

\(\oplus\) vector sum of sets \(X = Y \oplus Z\) where every element \(x \in X\) has unique expression \(x = y + z\) where \(y \in Y\), \(z \in Z\); \([318, \text{p.} 19]\) then summands are \textit{algebraic complements}. \(X = Y \oplus Z \Rightarrow X = Y + Z\). Now assume \(Y\) and \(Z\) are nontrivial subspaces. \(X = Y + Z \Rightarrow X = Y \oplus Z \iff Y \cap Z = \{0\}\) \([319, \text{§1.2}]\) \([115, \text{§5.8}]\). Each element from a vector sum (+) of subspaces has unique expression (\(\oplus\)) when a basis from each subspace is linearly independent of bases from all the other subspaces.

\(\ominus\) likewise, the vector difference of sets

\(\boxdot\) orthogonal vector sum of sets \(X = Y \boxdot Z\) where every element \(x \in X\) has unique orthogonal expression \(x = y + z\) where \(y \perp Z\), \(z \in Z\), and \(y \perp z\). \([340, \text{p.} 51]\) \(X = Y \boxdot Z \Rightarrow X = Y + Z\). If \(Z \subseteq Y \perp\) then \(X = Y \boxdot Z \iff X = Y \oplus Z\). \([115, \text{§5.8}]\) If \(Z = Y \perp\) then summands are \textit{orthogonal complements}.

\(\pm\) \textit{plus or minus} or \textit{plus and minus}

\(\perp\) as in \(A \perp B\) meaning \textit{A is orthogonal to B (and vice versa)}, where \(A\) and \(B\) are sets, vectors, or matrices. When \(A\) and \(B\) are vectors (or matrices under Frobenius’ norm), \(A \perp B \iff \langle A, B \rangle = 0 \iff \|A + B\|^2 = \|A\|^2 + \|B\|^2\)

\(\neg\) as in \(\neg A\), or \textit{relative complement of set} \(A\); \textit{id est}, \(\neg A = \{x \not\in A\}\); \textit{e.g}, \(B \neg A \triangleq \{x \in B \mid x \not\in A\} \equiv B \cap \neg A\)

\(\Rightarrow\) or \(\Leftarrow\) \textit{sufficient or necessary, implies, or is implied by}; \textit{e.g},
\begin{align*}
A \text{ is sufficient: } & A \Rightarrow B, & A \text{ is necessary: } & A \Leftarrow B, \\
A \Rightarrow B & \iff \neg A \Leftarrow \neg B, & A \Leftarrow B & \iff \neg A \Rightarrow \neg B,
\end{align*}

\(\iff\) \textit{if and only if (iff) or corresponds with or necessary and sufficient or logical equivalence}

\(\Rightarrow\) as in \(A \Rightarrow B\) means \(A \Rightarrow B\); conventional usage of English language imposed by logicians

\(\nRightarrow\) or \(\nLeftarrow\) \textit{insufficient or unnecessary, does not imply, or is not implied by}; \textit{e.g},
\begin{align*}
A \nRightarrow B & \iff \neg A \nLeftarrow \neg B, & A \nLeftarrow B & \iff \neg A \nRightarrow \neg B.
\end{align*}

\(\leftarrow\) \textit{is replaced with}; substitution, assignment

\(\rightarrow\) \textit{goes to}, or \textit{approaches}, or \textit{maps to}
\( t \to 0^+ \) \( t \) goes to \( 0 \) from above; meaning, from the positive \([209, \text{p.2}]\)

... as in \( 1 \cdots 1 \) and \( [s_1 \cdots s_N] \) meaning \textit{continuation}; respectively, ones in a row and a matrix whose columns are \( s_i \) for \( i=1 \ldots N \)

... as in \( i=1 \ldots N \) meaning, \( i \) is a \textit{sequence} of successive integers beginning with \( 1 \) and ending with \( N \); \textit{id est}, \( 1 \ldots N = 1:N \)

\( \vdots \) \( \ddots \) \( \ldots \)

\( \vdots \) as in \( f : \mathbb{R}^n \to \mathbb{R}^m \) meaning \( f \) \textit{is a mapping}, or \textit{sequence} of successive integers specified by bounds as in \( i : j = i \ldots j \) (if \( j < i \) then sequence is descending)

\( f : \mathcal{M} \to \mathcal{R} \) meaning \( f \) \textit{is a mapping from ambient space} \( \mathcal{M} \) \textit{to ambient} \( \mathcal{R} \), not necessarily denoting either domain or range

\( g \big|_{x_p} \) \textit{expression} \( g \) \textit{evaluated at} \( x_p \)

\( A, B \) as in, for example, \( A, B \in \mathbb{S}^N \) means \( A \in \mathbb{S}^N \) and \( B \in \mathbb{S}^N \)

\( (A, B) \) \textit{open interval} between \( A \) and \( B \) in \( \mathbb{R} \), or \textit{variable pair} perhaps of disparate dimension

\( [A, B] \) \textit{closed interval} or \textit{line segment} between \( A \) and \( B \) in \( \mathbb{R} \)

( ) \textit{hierarchal, parenthetical, optional}

\{ \} \textit{curly braces} denote a \textit{set} or \textit{list}, \textit{e.g.} \( \{X a \mid a \geq 0\} \) \textit{the set comprising} \( X a \) \textit{evaluated for each and every} \( a \geq 0 \) \textit{where membership of} \( a \) \textit{to some space is implicit}, a \textit{union}

\( \langle \rangle \) \textit{angle brackets} denote \textit{vector inner-product} \((33)\) \((38)\)

[ ] \textit{matrix} or \textit{vector}, or \textit{quote insertion}, or \textit{citation}

\( [d_{ij}] \) \textit{matrix} whose \textit{ij}\textsuperscript{th} \textit{entry} is \( d_{ij} \)

\( [x_i] \) \textit{vector} whose \textit{i} \textit{th} \textit{entry} is \( x_i \)

\( x_p \) \textit{particular value} of \( x \)

\( x_0 \) \textit{particular instance} of \( x \), or \textit{initial value} of a sequence \( x_i \)

\( x_1 \) \textit{first entry} of \textit{vector} \( x \), or \textit{first element} of a set or list \( \{x_i\} \)

\( x_e \) \textit{extreme point}

\( x_+ \) \textit{vector} \( x \) \textit{whose negative entries are replaced with} \( 0 \); \( x_+ = \frac{1}{2}(x + |x|) \) \((534)\) or \textit{clipped vector} \( x \) or \textit{nonnegative part} of \( x \)

\( x_- \) \( x_- \triangleq \frac{1}{2}(x - |x|) \) or \textit{nonpositive part} of \( x = x_+ + x_- \)
\( \hat{x} \) known data

\( x^* \) optimal value of variable \( x \). optimal \( \Rightarrow \) feasible

\( x^* \) complex conjugate or dual variable or extreme direction of dual cone

\( f^* \) convex conjugate function \( f^*(s) = \sup\{\langle s, x \rangle - f(x) \mid x \in \text{dom } f \} \)

\( P_C x \) or \( P x \) projection of point \( x \) on set \( C \), \( P \) is operator or idempotent matrix

\( P_k x \) projection of point \( x \) on set \( C_k \) or on range of implicit vector

\( \delta(A) \) (a.k.a \( \text{diag}(A) \), §A.1) vector made from main diagonal of \( A \) if \( A \) is a matrix; otherwise, diagonal matrix made from vector \( A \)

\( \delta^2(A) \equiv \delta(\delta(A)) \). For vector or diagonal matrix \( \Lambda \), \( \delta^2(\Lambda) = \Lambda \)

\( \delta(A)^2 = \delta(A)\delta(A) \) where \( A \) is a vector

\( \lambda_i(X) \) \( i^{th} \) entry of vector \( \lambda \) is function of \( X \)

\( \lambda(X)_i \) \( i^{th} \) entry of vector-valued function of \( X \)

\( \lambda(A) \) vector of eigenvalues of matrix \( A \), (1543) typically arranged in nonincreasing order

\( \sigma(A) \) vector of singular values of matrix \( A \) (always arranged in nonincreasing order), or support function in direction \( A \)

\( \Sigma \) diagonal matrix of singular values, not necessarily square

\( \sum \) sum

\( \pi(\gamma) \) nonlinear permutation operator (or presorting function) arranges vector \( \gamma \) into nonincreasing order (§7.1.3)

\( \Xi \) permutation matrix

\( \Pi \) doublet or permutation operator or matrix

\( \prod \) product

\( \psi(Z) \) signum-like step function that returns a scalar for matrix argument (739), it returns a vector for vector argument (1661)

\( D \) symmetric hollow matrix of distance-square or Euclidean distance matrix

\( D \) Euclidean distance matrix operator

\( D^T(X) \) adjoint operator

\( D(X)^T \) transpose of \( D(X) \)

\( D^{-1}(X) \) inverse operator

\( D(X)^{-1} \) inverse of \( D(X) \)
\( D^* \) optimal value of variable \( D \)

\( D^\ast \) dual to variable \( D \)

\( D^\circ \) polar variable \( D \)

\( \partial \) partial derivative or partial differential or matrix of distance-square squared \((1469)\) or boundary of set \( K \) as in \( \partial K \) \((17)\) \((24)\)

\( \sqrt{d_{ij}} \) (absolute) distance scalar

\( d_{ij} \) distance-square scalar, EDM entry

\( V \) geometric centering operator, \( V(D) = -VDV^{\frac{1}{2}} \) \((1077)\)

\( V_N \) \( V_N(D) = -V_N^TDV_N \) \((1091)\)

\( V \) \( N \times N \) symmetric elementary, auxiliary, projector, geometric centering matrix, \( R(V) = N(1^T), N(V) = R(1) \), \( V^2 = V \) \((\text{§B.4.1})\)

\( V_N \) \( N \times N - 1 \) Schoenberg auxiliary matrix

\( R(V_N) = N(1^T), N(V_N) = R(1) \) \((\text{§B.4.2})\)

\( V_X \) \( V_X V_X^T \equiv V^T X^T X V \) \((1269)\)

\( X \) point list \((76)\) having cardinality \( N \) arranged columnar in \( \mathbb{R}^{n \times N} \), or set of generators, or extreme directions, or matrix variable

\( G \) Gram matrix \( X^T X \) \((978)\)

\( r \) affine dimension

\( k \) number of conically independent generators

\( k \) raw-data domain of Magnetic Resonance Imaging machine as in \textit{k-space}

\( n \) Euclidean (ambient spatial) dimension of list \( X \in \mathbb{R}^{n \times N} \), or integer

\( N \) cardinality of list \( X \in \mathbb{R}^{n \times N} \), or integer

\( \text{in} \) function \( f \) in \( x \) means \( x \) as argument to \( f \)

or \( x \) in \( C \) means element \( x \) is a member of set \( C \)

\( \text{on} \) function \( f(x) \) on \( A \) means \( A \) is \( \text{dom} \) \( f \)

or \( \text{relation} \leq \text{ on} \ A \) means \( A \) is set whose elements are subject to \( \leq \)

or \( \text{projection of} \ x \ \text{on} \ A \) means \( A \) is body on which projection is made

or \( \text{operating on vector} \) identifies argument type to \( f \) as “vector”

\( \text{onto} \) function \( f(x) \) maps onto \( M \) means \( f \) over its domain is a surjection with respect to \( M \)

\( \text{over} \) function \( f(x) \) over \( C \) means \( f \) evaluated at each and every element of set \( C \)
**APPENDIX F. NOTATION, DEFINITIONS, GLOSSARY**

**one-to-one** injective map or unique correspondence between sets

epi function epigraph

dom function domain

$\mathcal{R}f$ function range

$\mathcal{R}(A)$ the subspace: range of $A$, or span basis $\mathcal{R}(A)$; $\mathcal{R}(A) \perp \mathcal{N}(A^T)$

span as in $\text{span } A = \mathcal{R}(A) = \{ Ax \mid x \in \mathbb{R}^n \}$ when $A$ is a matrix

basis $\mathcal{R}(A)$ overcomplete columnar basis for range of $A$ or minimal set constituting generators for vertex-description of $\mathcal{R}(A)$ or linearly independent set of vectors spanning $\mathcal{R}(A)$

$\mathcal{N}(A)$ the subspace: nullspace of $A$; $\mathcal{N}(A) \perp \mathcal{R}(A^T)$

$\mathbb{R}^n$ Euclidean $n$-dimensional real vector space (nonnegative integer $n$).

$\mathbb{R}^0 = 0$, $\mathbb{R} = \mathbb{R}^1$ or vector space of unspecified dimension. [237] [412]

$\mathbb{R}^{m\times n}$ Euclidean vector space of $m$ by $n$ dimensional real matrices

$\times$ Cartesian product. $\mathbb{R}^{m\times n - m} \triangleq \mathbb{R}^{m\times (n - m)}$. $\mathcal{K}_1 \times \mathcal{K}_2 = \begin{bmatrix} \mathcal{K}_1 \\ \mathcal{K}_2 \end{bmatrix}$

$[ \begin{bmatrix} \mathbb{R}^m \\ \mathbb{R}^n \end{bmatrix} ]$ $\mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$

$\mathbb{C}^n$ or $\mathbb{C}^{n\times n}$ Euclidean complex vector space of respective dimension $n$ and $n \times n$

$\mathbb{R}^n_+$ or $\mathbb{R}^{n\times n}_+$ nonnegative orthant in Euclidean vector space of respective dimension $n$ and $n \times n$

$\mathbb{R}^n_-$ or $\mathbb{R}^{n\times n}_-$ nonpositive orthant in Euclidean vector space of respective dimension $n$ and $n \times n$

$\mathbb{S}^n$ subspace of real symmetric $n \times n$ matrices; the symmetric matrix subspace. $\mathbb{S}^0 = 0$, $\mathbb{S} = \mathbb{S}^1$ or symmetric subspace of unspecified dimension.

$\mathbb{S}^{n\perp}$ orthogonal complement of $\mathbb{S}^n$ in $\mathbb{R}^{n\times n}$, the antisymmetric matrices (51)

$\mathbb{S}^n_+$ convex cone comprising all (real) symmetric positive semidefinite $n \times n$ matrices, the positive semidefinite cone

$\text{int } \mathbb{S}^n_+$ interior of convex cone comprising all (real) symmetric positive semidefinite $n \times n$ matrices; *id est*, positive definite matrices

$\mathbb{S}^n_+(\rho) = \{ X \in \mathbb{S}^n_+ \mid \text{rank } X \geq \rho \}$ (260) convex set of all positive semidefinite $n \times n$ symmetric matrices whose rank equals or exceeds $\rho$

$\text{EDM}^N$ cone of $N \times N$ Euclidean distance matrices in the symmetric hollow subspace

$\sqrt{\text{EDM}}^N$ nonconvex cone of $N \times N$ Euclidean absolute distance matrices in the symmetric hollow subspace (§6.3)
CPU  central processing unit
dB  decibel
DC  direct current (0Hz)
Hz  hertz (cycles per second), MHz means *megahertz*, GHz *gigahertz*
EDM  Euclidean distance matrix
PSD  positive semidefinite
SDP  semidefinite program
SVD  singular value decomposition
SNR  signal to noise ratio

\[ S^n_1 \] subspace comprising all symmetric \( n \times n \) matrices having all zeros in first row and column (2099)
\[ S^n_h \] subspace comprising all symmetric hollow \( n \times n \) matrices (0 main diagonal), the *symmetric hollow subspace* (66)
\[ S^n_{h} \] orthogonal complement of \( S^n_h \) in \( S^n \) (67), the set of all diagonal matrices
\[ S^n_c \] subspace comprising all geometrically centered symmetric \( n \times n \) matrices; *geometric center subspace* \( S^n_c = \{ Y \in S^n \mid Y1 = 0 \} \) (2095)
\[ S^n_{c} \] orthogonal complement of \( S^n_c \) in \( S^n \) (2097)
\[ \mathbb{R}_{c}^{m \times n} \] subspace comprising all geometrically centered \( m \times n \) matrices
\[ X^\perp \] basis \( \mathcal{N}(X^T) \) (§2.13.9, §E.3.4)
\[ x^\perp \] \( \mathcal{N}(x^T) \); \( \{ y \in \mathbb{R}^n \mid x^Ty = 0 \} \) (§2.13.10.1.1)
\[ \mathcal{R}(P)^\perp \] \( \mathcal{N}(P^T) \) (fundamental subspace relations (137))
\[ \mathcal{N}(P)^\perp \] \( \mathcal{R}(P^T) \)
\[ \mathcal{R}^\perp = \{ y \in \mathbb{R}^n \mid \langle x, y \rangle = 0 \ \forall \ x \in \mathcal{R} \} \] (372), *Orthogonal complement of \( \mathcal{R} \) in \( \mathbb{R}^n \) when \( \mathcal{R} \) is a subspace*
\[ K^\perp \] normal cone (448)
\[ A^\perp \] normal cone to affine subset \( A \) (§3.1.2.2.2)
\[ K \] cone
\[ K^* \] dual cone \( -K^\circ \)
\[ K^\circ \] polar cone \( -K^* \), or *angular degree* as in 360°
\[ K_\mathcal{M}^+ \] monotone nonnegative cone
\( \mathcal{K}_M \) monotone cone
\( \mathcal{K}_\lambda \) spectral cone
\( \mathcal{K}_{\lambda\delta}^* \) cone of majorization
\( \mathcal{H} \) halfspace
\( \mathcal{H}_- \) halfspace described using an outward-normal (106) to the hyperplane partially bounding it
\( \mathcal{H}_+ \) halfspace described using an inward-normal (107) to the hyperplane partially bounding it
\( \partial \mathcal{H} \) hyperplane; \textit{id est}, partial boundary of halfspace
\( \partial \mathcal{H} \) supporting hyperplane
\( \partial \mathcal{H}_- \) a supporting hyperplane having outward-normal with respect to set it supports
\( \partial \mathcal{H}_+ \) a supporting hyperplane having inward-normal with respect to set it supports
\( d \) vector of distance-square
\( d_{ij} \) lower bound on distance-square \( d_{ij} \)
\( \overline{d}_{ij} \) upper bound on distance-square \( d_{ij} \)
\( \overline{AB} \) closed line segment between points A and B
\( AB \) matrix multiplication of A and B
\( \mathcal{C} \) closure of set \( \mathcal{C} \)

decomposition \( \text{orthonormal (2007, p.611), biorthogonal (1983, p.605)} \)

expansion \( \text{orthogonal (2017, p.613), biorthogonal (402, p.163)} \)

vector \( \text{column vector in } \mathbb{R}^n \)

entry \( \text{scalar element or real variable constituting a vector or matrix} \)

cubix \( \text{member of } \mathbb{R}^{M\times N\times L} \)

quartix \( \text{member of } \mathbb{R}^{M\times N\times L\times K} \)
feasible set \( \text{most simply, the set of all variable values satisfying all constraints of an optimization problem} \)

solution set \( \text{most simply, the set of all optimal solutions to an optimization problem; a subset of the feasible set and not necessarily a single point} \)

optimal \( \text{as in optimal solution, means a solution to an optimization problem. An optimal solution is not necessarily unique, but there is no better solution. optimal } \Rightarrow \text{ feasible} \)
feasible as in feasible solution, means satisfies the ("subject to") constraints of an optimization problem, may or may not be optimal

same as in same problem, means optimal solution set for one problem is identical to optimal solution set of another (without transformation)

equivalent as in equivalent problem, means optimal solution to one problem can be derived from optimal solution to another via suitable transformation

convex as in convex problem, essentially means a convex objective function optimized over a convex set (§4)

objective the three objectives of Optimization are minimize, maximize, and find

program Semidefinite program is any convex minimization, maximization, or feasibility problem constraining a variable to a subset of a positive semidefinite cone. Prototypical semidefinite program conventionally means: a semidefinite program having linear objective, affine equality constraints, but no inequality constraints except for cone membership. (§4.1.1) Linear program is any feasibility problem, or minimization or maximization of a linear objective, constraining the variable to some polyhedron (§2.13.1.0.3)

natural order with reference to stacking columns in a vectorization means a vector made from superposing column 1 on top of column 2 then superposing the result on column 3 and so on; as in a vector made from entries of the main diagonal δ(A) means taken from left to right and top to bottom

partial order relation \( \preceq \) is a partial order, on a set, if it possesses reflexivity, antisymmetry, and transitivity (§2.7.2.2)

operator mapping to a vector space (a multidimensional function)

projector short for projection operator; not necessarily minimum-distance or represented by a matrix

sparsity ratio of number of nonzero entries to matrix-dimension product

tight with reference to a bound means a bound that can be met, with reference to an inequality means equality is achievable

\( \hat{v} \) coefficient vector for two spectral factors, Figure 102 level 2

\( \bar{v} \) coefficient vector corresponding to four spectral factors, Figure 102 level 3

\( \overline{v} \) vector containing numerator or denominator coefficients of eight spectral factors; level 4 in a bifurcation tree like Figure 102

\( g' \) first derivative of possibly multidimensional function with respect to real argument or primed variable closely related to unprimed

\( g'' \) second derivative with respect to real argument
first directional derivative of possibly multidimensional function \( g \) in direction \( Y \in \mathbb{R}^{K \times L} \) (maintains dimensions of \( g \))

second directional derivative of \( g \) in direction \( Y \)

gradient from calculus, \( \nabla f \) is shorthand for \( \nabla_x f(x) \). \( \nabla f(y) \) means \( \nabla_y f(y) \) or gradient \( \nabla_x f(y) \) of \( f(x) \) with respect to \( x \) evaluated at \( y \)

second-order gradient

distance scalar (Figure 27), or first-order difference matrix (897), or infinitesimal difference operator (§D.1.4)

triangle made by vertices \( i \), \( j \), and \( k \)

Roman numeral one or capital i

Identity operator or matrix \( I = \delta^2(I) \), \( \delta(I) = 1 \). Variant: \( I_m \overset{\Delta}{=} I \in \mathbb{S}^m \)

index set, a discrete set of indices

empty set, an implicit member of every set

real zero

origin or vector or matrix of zeros

sort-index matrix

order of magnitude information required, or computational intensity: \( O(N) \) is first order, \( O(N^2) \) is second, and so on

real one

vector of ones. Variant: \( 1_m \overset{\Delta}{=} 1 \in \mathbb{R}^m \)

vector whose \( i \)th entry is 1 (otherwise 0), \( i \)th member of the standard basis for \( \mathbb{R}^n \) (60)

maximum [209, §0.1.1] or largest element of a totally ordered set

find maximum of objective function w.r.t independent variables \( x \).

Subscript \( x \leftarrow x \in \mathcal{C} \) may hold implicit constraints if context clear; e.g., semidefiniteness

argument of operator or function, or variable of optimization

supremum of totally ordered set \( \mathcal{X} \), least upper bound, may or may not belong to set [209, §0.1.1]; e.g., range \( \mathcal{X} \) of real function

argument \( x \) at supremum of function \( f \); not necessarily unique or a member of function domain
subject to specifies constraints of an optimization problem; generally, inequalities and affine equalities. Subject to implies: anything not an independent variable is constant, an assignment, or substitution

\[
\begin{align*}
\min & \quad \text{minimum [209, §0.1.1] or smallest element of a totally ordered set} \\
\text{minimize} & \quad \text{find objective function minimum w.r.t independent variables } x. \\
\text{find} & \quad \text{find any feasible solution, specified by the ("subject to") constraints, w.r.t independent variables } x. \\
\text{arg inf } f(x) & \quad \text{argument } x \text{ at infimum of function } f; \text{ not necessarily unique or a member of function domain} \\
\text{iff} & \quad \text{if and only if, necessary and sufficient; also the meaning indiscriminately attached to appearance of the word "if" in the statement of a mathematical definition, [135, p.106] [268, p.4] an esoteric practice worthy of abolition because of ambiguity thus conferred} \\
\text{rel} & \quad \text{relative} \\
\text{int} & \quad \text{interior} \\
\text{lim} & \quad \text{limit} \\
\text{sgn} & \quad \text{signum function or sign} \\
\text{round} & \quad \text{round to nearest integer} \\
\text{mod} & \quad \text{modulus function} \\
\text{tr} & \quad \text{matrix trace} \\
\text{rank} & \quad \text{as in rank } A, \text{ rank of matrix } A; \dim \mathcal{R}(A) \\
\text{dim} & \quad \text{dimension, } \dim \mathbb{R}^n = n, \dim(x \in \mathbb{R}^n) = n, \dim \mathcal{R}(x \in \mathbb{R}^n) = 1, \dim \mathcal{R}(A \in \mathbb{R}^{m \times n}) = \text{rank}(A) \\
\text{aff} & \quad \text{affine hull} \\
\text{dim aff} & \quad \text{affine dimension} \\
\text{card} & \quad \text{cardinality, number of nonzero entries } \text{card } x \triangleq \|x\|_0 \text{ or } N \text{ is cardinality of list } X \in \mathbb{R}^{n \times N} \text{ (p.276)}
\end{align*}
\]
conv  convex hull (§2.3.2)
cone  conic hull (§2.3.3)
cenv  convex envelope (§7.2.2.1)
content of high-dimensional bounded polyhedron, volume in 3 dimensions, area in 2, and so on
cof  matrix of cofactors corresponding to matrix argument
dist  distance between point or set arguments; e.g., dist(x, B)
vec  columnar vectorization of m \times n matrix, Euclidean dimension mn (37)
svec  columnar vectorization of symmetric n \times n matrix, Euclidean dimension n(n + 1)/2 (56)
dvec  columnar vectorization of symmetric hollow n \times n matrix, Euclidean dimension n(n - 1)/2 (73)
\langle x, y \rangle  angle between vectors x and y, or dihedral angle between affine subsets
\succeq  generalized inequality; e.g., A \succeq 0 means:
  • vector or matrix A must be expressible in a biorthogonal expansion having nonnegative coordinates with respect to extreme directions of some implicit pointed closed convex cone K (§2.13.2.0.1, §2.13.7.1.1),
  • or comparison to the origin with respect to some implicit pointed closed convex cone (2.7.2.2),
  • or (when K = S^n_+) matrix A belongs to the positive semidefinite cone of symmetric matrices (nonnegative eigenvalues, §2.9.0.1),
  • or (when K = R^n_+) vector A belongs to the nonnegative orthant (each vector entry is nonnegative, §2.3.1.1)
\succeq  as in x \succeq z means x - z \in K (182)
\kappa  as in x \succeq \kappa
\succ  strict generalized inequality, membership to cone interior; A \succ 0 means:
  • vector or matrix A must be expressible in a biorthogonal expansion having positive coordinates with respect to extreme directions of some implicit pointed closed convex cone K (§2.13.2.0.1, §2.13.7.1.1),
  • or comparison to the origin with respect to the interior of some implicit pointed closed convex cone (2.7.2.2),
  • or (when K = S^n_+) matrix A belongs to the interior of the positive semidefinite cone of symmetric matrices (positive eigenvalues, §2.9.0.1),
  • or (when K = R^n_+) vector A belongs to the interior of the nonnegative orthant (each vector entry is positive, §2.3.1.1)
not positive definite
\[ \geq \] scalar inequality, greater than or equal to; comparison of scalars, or entrywise comparison of vectors or matrices with respect to \( \mathbb{R}_+ \)

nonnegative for \( \alpha \in \mathbb{R}^n \), \( \alpha \geq 0 \); id est, nonnegative entries when with respect to nonnegative orthant, represents vector on boundary of or interior to pointed closed convex cone \( K \)

positive for \( \alpha \in \mathbb{R}^n \), \( \alpha > 0 \); id est, positive (nonzero) entries when with respect to nonnegative orthant, no vector on boundary of pointed closed convex cone \( K \)

\( \lfloor x \rfloor \) floor function, \( \lfloor x \rfloor \) is greatest integer not exceeding \( x \)

\( \| x \| \) entrywise absolute value of scalars, vectors, and matrices

log natural (or Napierian) logarithm

\( \text{det} \) matrix determinant

\( \| x \| = \sqrt{\sum_{j=1}^{n} |x_j|^2} \) Euclidean norm or vector 2-norm \( \| x \|_2 \) (§3.2)

\( \| x \|_2^2 = x^T x = \langle x, x \rangle \) Euclidean norm square (§3.1.2.1)

\( \| x \|_\ell = \ell \sqrt{\sum_{j=1}^{n} |x_j|\ell} \) vector \( \ell \)-norm for \( \ell \geq 1 \) (convex)

\[ \triangleq \sum_{j=1}^{n} |x_j|\ell \] vector \( \ell \)-norm for \( 0 \leq \ell < 1 \) (violates §3.2 no. 3)

\( \| x \|_0 = 1^T |x|^0 \) (00 \( \triangleq \) 0) “0-norm” or cardinality of vector \( x \) (card \( x \)) (§4.5.1)

\( \| x \|_1 = 1^T |x| \) 1-norm, dual infinity-norm (§3.2)

\( \| x \|_\infty = \max\{|x_j| \forall j\} \) infinity-norm (§3.2)

\( \| x \|_n_k = \sum_{i=1}^{k} \pi(|x|)_i \) \( k \)-largest norm (§3.2.2.1)

\( \| X \|_2 = \sup_{\| a \|=1} \| Xa \|_2 = \sigma_1 = \sqrt{\lambda(X^TX)} \) matrix 2-norm or spectral norm, largest singular value [168, p. 56]. For \( x \) a vector: \( \| \delta(x) \|_2 = \| x \|_\infty . \)

\[ \| Xa \|_2 \leq \| X \|_2 \| a \|_2 \] (2190)

\( \| X \|_2^* = 1^T \sigma(X) \) nuclear norm, dual spectral norm (§C.2)

\[ \| X \| = \sqrt{\sum_{i,j} X_{ij}^2} \] Frobenius’ matrix norm \( \| X \|_F \) (§2.2.1)
Bibliography


http://www.acm.caltech.edu/~emmanuel/papers/ExactRecovery.pdf

http://arxiv.org/abs/0711.1612v1


http://cseweb.ucsd.edu/~lcayton/robEmb.pdf


http://vision.ucsd.edu/~mana/cvpr07.quadric.pdf


nobelprize.org/nobel_prizes/physics/laureates/1997/chu-autobio.html


http://ccrma.stanford.edu/~dattorro/PhiLS.pdf

http://ccrma.stanford.edu/~dattorro/Hearing.htm


http://www.convexoptimization.com/TOOLS/DattorroLAA.pdf

http://www.stanford.edu/~boyd/papers/fdbk_alloc.html


http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.7472


BIBLIOGRAPHY


http://wcan.ee.psu.edu/papers/filiz-yener_ciss03.pdf


http://www.convexoptimization.com/TOOLS/Frisch57.pdf


http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.8.7378


http://www.convexoptimization.com/TOOLS/Ghaoui.pdf


http://www.stanford.edu/~boyd/papers/inductor_opt.html


http://www.convexoptimization.com/TOOLS/Jean_Baptiste.pdf


http://www.convexoptimization.com/TOOLS/AHorn.pdf


http://www.convexoptimization.com/TOOLS/pardalos.pdf

http://www.convexoptimization.com/TOOLS/hubert.pdf

www.convexoptimization.com/TOOLS/ReweightingFrom1999XiaomingHuo.pdf


http://www.technologyreview.com/communications/13643


http://www.convexoptimization.com/TOOLS/permu.pdf
http://www.convexoptimization.com/TOOLS/Mesbahi.pdf
http://www.stanford.edu/~boyd/papers/inductanceexpressions.html
http://www.stanford.edu/~boyd/papers/bandwidth_ext.html
ioe.engin.umich.edu/people/fac/books/murty/linear_complementarity_webbook
http://arxiv.org/abs/math.MG/0512649


http://www.convexoptimization.com/TOOLS/Marks.pdf


www-ee.stanford.edu/~osgood/Ahlfors-Bergman-Szego.pdf


http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.3831
Erratum: p.567 herein.


http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.22.3362


http://www.convexoptimization.com/TOOLS/YeSo.pdf


http://www.convexoptimization.com/TOOLS/soren.pdf


http://web.mit.edu/18.06/www
http://www.convexoptimization.com/TOOLS/Straszewicz.pdf

http://sedumi.ie.lehigh.edu

www.optimization-online.org/DB_HTML/2001/05/324.html


http://stanford.edu/~boyd/papers/fmmp.html


http://takane.brinkster.net/Yoshio/p008.pdf


http://www.convexoptimization.com/TOOLS/csCamera-SPIE-dec05.pdf


convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

[386] Wikimization. EDM using ordinal data. 
convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

convexoptimization.com/wikimization/index.php/Projection_on_Polyhedral_Cone

http://www.convexoptimization.com/wikimization/index.php/Fermat_point

convexoptimization.com/wikimization/index.php/Filter_design_by_convex_iteration

convexoptimization.com/wikimization/index.php/Complementarity_problem

convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

[393] Wikimization. Map of the USA. 
convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

convexoptimization.com/wikimization/index.php/Projection_on_Polyhedral_Cone

convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

convexoptimization.com/wikimization/index.php/Sampling_Sparsity

convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

[399] Wikimization. Singular value decomposition (SVD) by rank-1 transformation. 
convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization

[401] Wikimization. isedm(). 
convexoptimization.com/wikimization/index.php/Matlab_for_Convex_Optimization


http://mathworld.wolfram.com

http://www.convexoptimization.com/TOOLS/White.pdf


Index

∅, see empty set
0-norm, 193, 197, 252, 289, 294, 295, 300–303, 319, 327, 673
Boolean, 252
1-norm, 190–197, 265, 300, 305, 306, 497, 673
ball, see ball
Boolean, 252
distance matrix, 222, 415
gradient, 197
heuristic, 291, 292, 497
nonnegative, 193, 194, 294, 300–304
2-norm, 46, 187, 190, 191, 306, 593, 673
ball, see ball
Boolean, 252
matrix, 63, 210, 265, 483, 492, 536, 546, 551, 552, 563, 639, 673
constraint, 62
dual, 562, 673
inequality, 673
Schur-form, 210, 563
∞-norm, 190, 191, 196, 265, 673
ball, see ball
dual, 673
k-largest norm, 196, 292, 293, 343, 673
gradient, 197
ℓ-norm, 673
≻, 57, 88, 90, 101, 672
≥, 57, 90, 101, 672
≥, 57, 673
Boolean, 252
nonnegative, 78, 79, 193, 238, 294, 299–304
V, 365, 552
V_N, 362, 554
δ, 511, 523, 664
e, 52, 670
ι, 385, 661
μ, 323, 661
λ, 511, 520, 523, 664
θ, 38, 42, 575
π, 195, 196, 290, 488, 491, 508, 664
ψ, 258, 538
σ, 511, 512, 664
∇, 660
√, 199, 660, 665
×, 43, 140, 649, 666
tr, 520, 523, 671
− A −, 35
accuracy, see precision
active set, 263
adaptive, 323, 328, 330
additivity, 92, 101, 523
adjacency, 48
adjoint
operator, 46, 141, 471, 659, 664
self, 46, 367, 471, 511
affine, 35, 200
algebra, 36, 37, 55, 71, 84, 614
boundary, 35, 43
combination, 61, 63, 70, 129
dimension, 22, 36, 54, 397–399, 406, 473
complementary, 473
cone, 122
low, 376
minimization, 496, 506
Précis, 399
reduction, 435
spectral projection, 506
face, 84
hull, see hull
independence, 64, 69, 70, 122, 125
preservation, 70
subset, 70, 77, 78
interior, 37
intersection, 615
map, see transformation
membership relation, 189
normal cone, 655
set, 35, 36, 54
subset, 36, 63
hyperplane intersection, 76
independence, 70, 77, 78
nullspace basis span, 77
parallel, 36
projector, 605, 614
transformation, see transformation
algebra
affine, 36, 37, 55, 71, 84, 614
arg, 190, 293, 561, 562
cone, 141
face, 84
function
  linear, 185
  support, 203
fundamental theorem, 532
  linear, 75
interior, 37, 43, 156
linear, 75, 511
nullspace, 545
optimization, 190, 561, 562
projection, 614
subspace, 34, 75, 76, 545
algebraic
  complement, 85, 86, 134, 173, 609, 612, 662
  projection on, 608, 634
  multiplicity, 532
Alizadeh, 257, 566
alternating projection, see projection
alternation, 428
alternative, see theorem
ambient, see vector, space
anchor, 269, 270, 373, 377
angle, 45, 65, 75
  acute, 66, 648
  alternating projection, 648
brackets, 46
constraint, 366, 369, 383
dihedral, 26, 75, 383, 433, 558
EDM cone, 111
halfspace, 65, 75
hyperplane, 75
inequality, 359, 431
matrix, 387
interpoint, 363, 366
obtuse, 66
positive semidefinite cone, 111
  relative, 385, 386
  right, 113, 137, 155, 156
torsion, 26, 383
antidiagonal, 338, 390
antisymmetry, 90, 128, 523, 669
arg, 670
algebra, 190, 293, 561, 562
artificial intelligence, 22
asymmetry, 33
audio, 21, 22, 31, 229, 295, 296, 345, 637
Avis, 54
axioms of the metric, 357
axis of revolution, 112, 113, 155, 558, 585

  – B –, 191
ball
  1-norm, 191–194
  nonnegative, 300, 301
  2-norm, 63, 640
Euclidean, 36, 192, 193, 370, 604, 638
  smallest, 57
∞-norm, 190, 192, 193
nuclear norm, 59, 61
packing, 370
polyhedral, 192, 193
spectral norm, 63, 640
Barker & Carlson, 125
Barvinok, 33, 99, 120, 123, 125, 244, 245, 257,
  261–263, 367, 369, 380, 651
base, 85
station, 380
basis, 52, 163, 666
discrete cosine transform, 296
nonorthogonal, 163, 615
  projection on, 615
nullspace, 77, 81, 528, 610
orthogonal, 52, 54, 80, 612
  projection on, 615
pursuit, 322
standard
  matrix, 52, 54, 624, 651, 660
  vector, 52, 131, 304, 670
Bauschke, 646
bees, 29, 370
Bellman & Fan, 239
Ben-Israel, 148, 247
Bessel, 611
bijective, 46, 48, 50, 103, 158
Gram form, 393, 395
  inner-product form, 395, 396
invertible, 48, 602
isometry, 48, 428
linear, 47, 48, 52, 53, 123
orthogonal, 48, 428, 490, 557, 571
binary search, 226, 231
biorthogonal
  decomposition, 605, 612, 613
  expansion, 134, 145, 159, 163, 606, 613
  EDM, 448
  projection, 612, 616, 617
  unique, 161, 163, 164, 449, 549, 612, 616,
  617
biorthogonality, 161, 163
  condition, 159, 533, 549
Birkhoff, 62
bisector, 226, 231
Biswas
  Pratik, 271, 283, 383
  Sudhendru, 545
Blu, 321
Boolean, see problem
Borg & Groenen, 26, 425
Borwein, 82, 92, 629, 646
bound
  lower, 483
  polyhedral, 441
INDEX

boundary, 36, 38, 42, 57, 85
-point, 83
extreme, 97
of affine, 35, 43
of cone, 93, 152, 157, 158
EDM, 440, 445, 451
membership, 145, 152, 158, 248
PSD, 37, 99, 105, 117–120, 248, 455
ray, 93
of halfspace, 66
of hypercube slice, 330
of open set, 38
of point, 43
of ray, 85
relative, 36, 38, 42, 57, 85
bounded, 57, 75
bowl, 584, 585
box, 74, 639
bridge, 383, 406, 463, 483, 495
Bronstein, 26
Bunt, 119, 629

-C-, 33
calculus
matrix, 575
of inequalities, 7
of variations, 149
Carathéodory’s theorem, 145, 624
cardinality, 289, 361, 671, 673
1-norm heuristic, 291, 292, 497
-one, 192, 193, 300–302
constraint, see constraint
convex envelope, see convex geometry, 192, 194, 301
Boolean, 252
nonnegative, 193, 194, 196, 289–295, 299, 300, 302–304
rank connection, 265
rank one, 319
rank, full, 257
quasiconcavity, 197, 289
regularization, 299, 320
Carlson, 125
Cartesian
axes, 35, 56, 71, 87, 94, 132, 142, 291, 639
cone, see orthant
coordinates, 423
product, 43, 245, 649, 666
cone, 89, 90, 140
function, 189
subspace, 291, 639
Cauchy-Schwarz, 78, 646
Cayley-Menger
determinant, 433, 434
form, 400, 417, 488, 507
cellular, 381
telephone network, 380
center
geometric, 365, 368, 388, 422
gavity, 388
mass, 388
central path, 245
certificate, 184, 248, 249, 264, 338, 351
null, 146, 153
Chabrillac & Crouzeix, 458
chain rule, 581
two argument, 581
Chu, 7
clipping, 197, 480, 638, 639, 663
closed, see set
coefficient
binomial, 196, 304
projection, 159, 454, 620, 625
cofactor, 433
compaction, 271, 283, 365, 368, 461, 498
comparable, 92, 160
complement
algebraic, 85, 86, 134, 173, 609, 612, 662
projector, 609
halfspace, 63
orthogonal, 50, 75, 265, 609, 612, 662, 667
cone, dual, 134, 142, 143, 635
cone, normal, 655
vector, 667
relative, 82, 662
complementarity, 173, 190, 290
linear, 176
maximal, 240, 245, 262
problem, 176
linear, 176
semidefinite, 241, 251, 264
strict, 251
complementary
affine dimension, 473
dimension, 75
eigenvalues, 608
halfspace, 68
inertia, 417, 527
slackness, 250
subspace, see complement
formula, 294, 321
nonnegative, 193, 194, 294, 300–304
Procrustes, 311
prototypical, 301
compressive sampling, see compressed computational complexity, 240
concave, 185, 186, 198
condition
biorthogonality, 159, 533, 549
projection on matrices, 639
projection on PSD cone, 212, 319, 487, 492, 564
Schur-form, 63
singular value, 63
sort, 427
tractable, 238
content, 434, 672
contour plot, 174, 219
contraction, 63, 176, 328, 330
convergence, 267, 428, 641
geometric, 643
global, 264, 267, 290, 351
local, 267, 292, 295, 305, 351
measure, 254
convex, 7, 21, 33, 185–187
combination, 33, 63, 129
evernvelope, 496, 497
bilinear, 229
cardinality, 255, 497
rank, 278, 380, 496, 497
equivalent, 278
form, 8, 237
gometry, 33
fundamental, 66, 71, 75, 355, 365
hull, see hull
iteration
accelerant, 353
cardinality, 289–295, 305
cardinality one, 300–302
convergence, 264, 267, 290, 292, 305, 351
rank, 263, 265–267, 500
rank one, 307, 349, 350, 500
rank, fat, 344
rank, indefinite, 344
stall, 267, 292, 295, 307, 314, 349
log, 236
optimization, 7, 21, 187, 237, 373
art, 8, 256, 271, 510
fundamental, 237
solution set, 187, 349
polyhedron, see polyhedron
problem, 138, 149, 175, 177, 191, 201, 208, 210, 240, 246, 492, 629, 669
definition, 237
gometry, 21, 24
nonlinear, 237
statement as solution, 8, 237
tractable, 238
program, see problem
quadratic, see function
strictly, see function, convex
convexity, 185, 198
condition
first order, 216, 218, 221, 222
second order, 220, 224
eigenvalue, 406, 524, 564
K-, 186
log, 236
norm, 63, 187, 190–192, 210, 215, 562, 673
projector, 608, 612, 629, 630
strict, see function, convex
coordinates
conic, 134, 163, 165, 182–184
inversion, 183
CPU, 667
Crippen & Havel, 383
crippling, 240
Critchley, 394, 505
criterion
EDM, 454, 457
dual, 471
Finsler, 458
metric versus matrix, 400
Schoenberg, 27–29, 364, 365, 417, 438, 450, 454, 471, 475, 478, 626
Crouzeix, 458
cubix, 576, 578, 668
curvature, 220, 226
cylinder, 60
– D –, 364
d’Aspremont, 100, 319
Dantzig, 62, 238, 239
dB, 667
DC, 667
decomposition, 668
biorthogonal, 605, 612, 613
completely PSD, 89, 268
dyad, 543
eigenvalue, 531–534
factorization
nonnegative, 268
positive, 268
spectral, 284–286
orthonormal, 611–613
singular value, see singular value
decomposition, 668
bioorthogonal, 605, 612, 613
completely PSD, 89, 268
dyad, 543
eigenvalue, 531–534
factorization
nonnegative, 268
positive, 268
spectral, 284–286
orthonormal, 611–613
singular value, see singular value
decomposition, 668
definite, see matrix
deLeeuw, 421, 494, 501
delsarte, 370
dense, 84
derivative, 211
directional, 218, 581, 584, 585, 588–590
dimension, 584, 592
derivative, 585
directional, 586
derivative, 590
dimensional, 575, 665
table, 592
description
description
conversion, 131
halfspace-, 64, 66
of affine, 77
do of affine, 77
of dual cone, 150
of polyhedral cone, 128
vertex, 63, 64, 96
of affine, 77
of dual cone, 165, 178
of halfspace, 125, 126
of hyperplane, 69
of polyhedral cone, 63, 129, 146, 158
of polyhedron, 57, 129
projection on affine, 619
determinant, 220, 430, 520, 530, 592, 598
Cayley-Menger, 433, 434
inequality, 523
Kronecker, 513
product, 521
Deutsch, 630, 638, 640, 643
Deza, 259, 453
diag(), see δ
diagonal, 511, 523, 664
δ, 511, 523, 664
binary, 606
dominance, 117
inequality, 523
matrix, 531
commutative, 522
diagonalization, 534
nonnegative, 523
positive semidefinite, 523
values
eigen, 531
singular, 534
zero, 539
diagonalizable, 161, 531–533
simultaneously, 107, 251, 520, 522, 541
diagonalization, 40, 80, 531–534
diagonal matrix, 534
expansion by, 161
symmetric, 533
diamond, 390
differentiable, see function
differential
Frechet, 583
partial, 665
diffusion, 461
dilation, 425
dimension, 36
affine, see affine
complementary, 75
conservation, 75, 399, 539, 540, 551
domain, 49
embedding, 55
Euclidean, 665
invariance, 49
nullspace, 75
positive semidefinite cone
face, 106
Précis, 399
range, 49
rank, 122, 398, 399, 671
direction, 85, 95, 581
matrix, 264–267, 278, 320, 371, 501
existence, 264
Identity, 266, 345
interpretation, 265
Newton, 211
projection, 609, 612, 621, 622
nonorthogonal, 606–608, 616, 620
parallel, 613, 614, 617, 618
steepest descent, 211, 584
vector, 264–267, 290–292
existence, 264
Identity, 266, 345
interpretation, 291, 292, 497
optimality condition, 264, 290
unity, 291
disc, 59
discrete Fourier transform, 48, 323
discretization, 150, 186, 220, 475
dist, 644, 672
distance, 644, 645
1-norm, 191, 192, 222
absolute, 357, 414
comparative, see distance, ordinal
Euclidean, 361
geometry, 21, 380
matrix, see EDM
maximization, 631
minimization, 609, 626
ordinal, 423–429
origin to affine, 68, 191, 192, 604, 618
nonnegative, 193, 194, 300
property, 357
self, 357
taxicab, see distance, 1-norm
Donoho, 345
doublet, see matrix
dual
affine dimension, 473
dual, 140, 146, 151, 250, 254, 318
feasible set, 245
norm, 143, 562, 673
projection, 630, 631
on cone, 85, 134, 473, 505, 632–636
on convex set, 630, 631
on EDM cone, 473, 474, 505, 506
optimization, 630
strong, 138, 139, 250, 571
variable, 135, 176, 177, 184, 664
duality, 135, 368
gap, 138, 139, 247, 250, 318
strong, 138, 184, 201, 250
weak, 249
Duensing, 323
dvec, 53, 426, 672
dyad, see matrix
Dykstra, 643, 644, 654
algorithm, 642, 653, 655, 658

− E − , 407
Ebert, 601
Eckart & Young, 30, 485, 487, 510, 565
distance, 82
√EDM, 439, 441, 493
EDM, 28, 355, 356, 361, 478, 667
closest, 486
composition, 412–414, 439, 441
construction, 444
criterion, 454, 457
dual, 471
definition, 361, 442
Gram form, 363, 451
inner-product form, 385
interpoint angle, 363
relative-angle form, 386
dual, 471
exponential entry, 412
dual, 471, 473

EDM, 28, 355, 356, 361, 478, 667
closest, 486
composition, 412–414, 439, 441
construction, 444
criterion, 454, 457
dual, 471
definition, 361, 442
Gram form, 363, 451
inner-product form, 385
interpoint angle, 363
relative-angle form, 386
dual, 471, 473
exponential entry, 412
dual, 471, 473

ellipsoid, 34, 39, 40, 306
invariance, 44
equality constraint, 149, 177
affine, 35
closed, 38
cone, 88
face, 82
hull, 55, 57, 63
open, 38
entry, 35, 186, 659, 663–665, 668
largest, 195
smallest, 195
epigraph, 185, 187, 204, 223, 235
form, 209, 504
intersection, 203, 204, 210, 236
nonconvex, 204

empty
interior, 36, 37
set, 35–37, 43, 561
affine, 35
closed, 38
cone, 88
face, 82
hull, 55, 57, 63
open, 38
entry, 35, 186, 659, 663–665, 668
largest, 195
smallest, 195
epigraph, 185, 187, 204, 223, 235
form, 209, 504
intersection, 203, 204, 210, 236
nonconvex, 204

equality
constraint, 149, 177
affine, 238
EDM cone, 29, 437, 470
dual, 474
normal equation, 603
PSD cone, 29, 437, 470
equivalent, 669
Ericson & Zinoviev, 372
INDEX

errata, 180, 201, 270, 394, 567, 638
Eternity II, 330

extreme point, 339

Euclidean
ball, 36, 192, 193, 370, 604, 638

smallest, 57
distance, 356, 361
geometry, 21, 374, 380, 626
matrix, see EDM
metric property, 357, 402
fifth, 357, 359, 360, 429, 431, 436
norm, see 2-norm
projection, 118
space, 33, 34, 356, 666
ambient, 34, 465, 476

exclusive mutually, 148

expansion, 163, 668
biorthogonal, 134, 145, 159, 163, 606, 613
EDM, 448
projection, 612, 616, 617
unique, 161, 163, 164, 449, 549, 612, 616, 617

implied by diagonalization, 161
orthogonal, 52, 134, 162, 611–613
w.r.t orthant, 162

exponential, see matrix

exposed, 81, 82

closed, 449
direction, 96
extreme, 83, 90, 126
face, 82, 127
point, 82, 83, 96
density, 84

extreme, 81
boundary, 97
direction, 79, 93, 98, 123
distinct, 95
dual cone, 124, 159, 163, 165, 167, 179, 180
EDM cone, 97, 448, 470
one-dimensional, 95, 97
origin, 95
orthant, 79
pointed cone, 90, 95, 130, 138
PSD cone, 95, 100, 110, 111
shared, 181
supporting hyperplane, 159
unique, 95

exposed, 83, 90, 126
point, 62, 81–84, 244
cone, 90, 121
Fantope, 265
feasible set, 264
hypercube set, 291
ray, 95, 451
sum, 95

\[ \mathcal{F} \] , 84

face, 38, 82

affine, 84
algebra, 84
cone, see cone
exposed, see exposed
intersection, 153
isomorphic, 106, 367, 447
of face, 84
polyhedron, 126, 127
recognition, 26
smallest, 84, 106–109, 153, 154, see cone
transitivity, 84

facet, 84, 159, 181

factorization, see decomposition
Fan, 59, 239, 264, 278, 520, 562, 565

Fantope, 58, 59, 111, 207, 264, 265

extreme point, 265

Farkas’ lemma, 143, 147, 148
positive definite, 247
not, 248
positive semidefinite, 246

fat, see matrix

feasible, 669

Fejér, 154, 214, 646, 647, 649

Fenchel, 25, 185, 496

Fermat point, 367, 368

Fiacco & McCormick, 239

filter, 286, 637
design, 229, 510

find, 609, 671

finitely generated, 127

Finsler, 458

floor, 673

Forsgren, Gill, & Wright, 237

Fourier transform, discrete, 48, 323

Frankel, Felice, 727

Frisch, 239

Frobenius, 47, see norm

Fukuda, 54

calculus, 36

-dimensional, 36, 92
rank, 76, 257

function, 185, see operator

affine, 44, 189, 200–202, 213

supremum, 203, 204, 562

bilinear, 229

composition, 198, 215, 581

affine, 215, 236

concave, 185, 186, 198, 499, 564

conjugate, 490, 664

continuity, 186, 204, 221, 235

convex, 185, 186, 204, 362, 387
difference, 197, 210, 292

invariance, 186, 204, 236

nonlinear, 186, 203

INDEX 705
simultaneously, 206, 207, 223
strictly, 186, 187, 189, 215–218, 220, 224, 225, 494, 498, 503, 584, 585
sum, 189, 215, 236
supremum, 203, 210, 236
trivial, 189
differentiable, 186, 203, 216, 218, 224
non, 186, 187, 191, 212, 235, 328, 329
distance
1-norm, 222
Euclidean, 361
fractional, 206, 216, 220, 223, 528
inverted, 198, 592, 599
power, 199, 599
projector, 207
root, 198, 523, 534, 660
inverted, see function, fractional
invertible, 45, 48, 49
linear, see operator
matrix, 185, 187, 221
chain rule, 581
convexity, 222, 224
differential, 222
epigraph, 223
gradient, 577
line theorem, 224
product, 578
monotonic, 185, 195, 197, 214, 229, 235, 292
strictly, 214
multidimensional, 185, 221, 669
affine, 189, 201
convexity, 222, 224
gradient, 210
line theorem, 224
monotonic, 214
quasiconvex, 227
negative, 198
nonconvex, 204, 220, 483, 500
norm, 190, see norm
objective, see objective
presorting, 195, 290, 488, 491, 508, 664
product, 206, 216, 227–229, 235
projection, see projector
pseudofractional, 205
quadratic, 187, 209, 217, 220, 225, 385, 528, 625
convex, strictly, 187, 220, 224, 225, 503, 584, 585
distance, 361
maximization, 574
nonconvex, 312, 574
nonnegative, 528
quasiconcave, 227, 229, 235, 258
cardinality, 197, 289
rank, 117
strictly, 227
quasiconvex, 118, 204, 226, 227, 229
continuity, 227, 235
strictly, 227
quasilinear, 220, 235, 538
quotient, see function, fractional
ratio, see function, fractional
smooth, 186, 191
sorting, 195, 290, 488, 491, 508, 664
step, 538
matrix, 258
stress, 220, 484, 500
support, 73, 203
algebra, 203
vector, 185, 187, 664
convexity, 218
epigraph, 204
gradient, 575, 577
line theorem, 224
montonicity, 214
sublevel, 205
fundamental
convex
geometry, 66, 71, 75, 355, 365
optimization, 237
metric property, 357, 402
semidefiniteness test, 154, 515, 625
subspace, 45, 75, 76, 546, 550, 551, 602, 610
projector, 610
vectorized, 80

− G — , 151
Gaffke & Mathar, 466
Gâteaux differential, 583
generators, 55, 63, 96, 151
c.i., 125
face, smallest, 153, 154, 304
hull, 129
affine, 55, 69, 619
conic, 63, 125
corner, 57
l.i., 125
minimal set, 57, 69, 125, 342, 666
affine, 619
extreme, 96, 158
halfspace, 125, 126
hyperplane, 126
orthant, 186
PSD cone, 100, 154, 221

geometric
center, 365, 368, 388, 422
operator, 394
subspace, 106, 389, 392, 393, 437, 463, 467–469, 496, 542, 628
Hahn-Banach theorem, 66, 71, 75
mean, 599
multiplicity, 532
Geršgorin, 116
Gill, 237
hyperdisc, 585
hyperplane, 63, 65–69, 213
    angle, 75
    hypersphere radius, 67
    independence, 77
intersection, 76, 153
    convex set, 71, 459
movement, 67, 79, 80, 194
normal, 68
nullspace, 65, 68
parallel, 35
separating, 75, 141, 146, 147
strictly, 75
supporting, 71–73, 174, 216–219
    cone, 134, 136, 153, 159
    cone, EDM, 465
    cone, PSD, 106, 107
exposed face, 82
polarity, 73
strictly, 73, 90, 216, 217
trivially, 73, 154
unique, 216–218
vertex-description, 69
hypersphere, 59, 277, 366, 409
    circum-, 400
    packing, 370
radius, 67
hypograph, 205, 235, 594
Hz, 667

− I −, 670
idempotent, see matrix
iff, 67
image, 44
    inverse, 44, 45, 156
    cone, 102, 141, 157, 205, 206, 223, 229, 245
magnetic resonance, 322
in, 665
independence
    affine, 64, 69, 70, 122, 125
    preservation, 70
    subset, 70, 77, 78
com, 23, 64, 121–126, 257, 342
versus dimension, 122
dual cone, 167
    extreme direction, 125
    preservation, 123
    unique, 122, 126, 163, 166
hyperplane, 77
linear, 34, 64, 122, 125, 257, 302, 548
    matrix, 521
    of subspace, 34, 662
    preservation, 34
inequality
    angle, 359, 431
    matrix, 387
Bessel, 611
Cauchy-Schwarz, 78, 646
determinant, 523
diagonal, 523
eigenvalue, 520, 570
generalized, 92, 134, 143
    dual, 143
dual PSD, 154
Hardy-Littlewood-Pólya, 489
Jensen, 216
linear, 22, 33, 66, 146
    matrix, 156, 157, 245, 248
log, 594
norm, 253, 673
    triangle, 190
rank, 523
singular value, 572
spectral, 416
sum, 161
trace, 523
triangle, 357, 358, 385, 402, 403, 405, 406, 412,
    429, 431–433, 440, 441
    norm, 190
    strict, 403
    unique, 430
    vector, 190
variation, 149
inertia, 415, 527
    complementary, 417, 527
Sylvester’s law, 521
infimum, see minimization, 561, 609, 626
    of concave functions, 236
    of quasiconcave functions, 236
    of Rayleigh quotient, 564
    of totally ordered set, 671
inflection, 220
injective, 48, 49, 92, 183, 391, 666
    Gram form, 389, 393, 395
inner-product form, 395, 396
invertible, 48, 601, 602
non, 49, 50
nullspace, 48
innovation, rate, 321
interior, 36, 38, 42, 85
    -point, 36
        method, 128, 237, 240, 241, 254, 284, 333,
        340, 370, 373, 474, 502
algebra, 37, 43, 156
empty, 36, 37
of affine, 37
of cone, 156, 158
    EDM, 440, 445
    membership, 145
    positive semidefinite, 99, 247
of halfspace, 43
INDEX 709

of point, 37, 38
of ray, 85
relative, 36, 37, 42, 85
transformation, 156
intersection, 43
affine, 615
cone, 89, 141
pointed, 90
epigraph, 203, 204, 210, 236
face, 153
halfspace, 66, 140, 154
hyperplane, 76, 153
convex set, 71, 459
line with boundary, 38
nullspace, 78
planes, 77
positive semidefinite cone
affine, 120, 248
green inner-product form, 542
line, 375
subspace, 78
tangential, 40

invariance
closure, 37, 127, 157
cone, 89, 90, 123, 141, 438
pointed, 92, 130
convexity, 43
dimension, 49
elipsoid, 44
function
convex, 186, 204, 236
Gram form, 389
inner-product form, 391
isometric, 22, 376, 388, 391, 421, 422
optimization, 190, 202, 265, 561, 562, 669
orthogonal, 49, 428, 483, 556, 559
reflection, 389
rotation, 112, 389
scaling, 236, 362, 561
set, 414
translation, 388, 389, 393, 410
function, 186, 236, 561
subspace, 463, 628

inverse
image, see image
matrix, see matrix
minimization, see function, fractional
nullspace, 45, 48
pseudo, see matrix, pseudoinverse
inversion
conic coordinates, 183
Gram form, 394
invertible
injective, 48, 601, 602
operator, 45, 48, 49, 156
is, 662
isedm(), 359, 365

isometry, 48, 428, 484
isomorphic, 47, 50, 52, 662
face, 106, 367, 447
isometrically, 40, 48, 103, 154
non, 542
isomorphism, 47, 157, 394, 396, 428, 438
isometric, 48, 52, 54, 428
symmetric hollow subspace, 54
symmetric matrix subspace, 51
isotonic, 425
iterate, 641, 650, 651, 654
iteration
alternating projection, 504, 641, 647, 655
convex, see convex

– j –, 323
Jacobian, 211
Jobs, 9
Johnson, 221, 516, 517
Jordan, 319

– K –, 88
Kaczmarz, 615
Karhunen-Loève transform, 421
Karmarkar, 239
Kimmel, 26
Kirschhofer, 332
kissing, 139, 191, 192, 300–302, 604
number, 370
problem, 370
KKT conditions, 176, 177, 486, 574
Klanfer, 366
Klee, 95
Korkine & Zolotarev, 372
Krein-Rutman, 141
Kronecker product, 103, 511, 513, 571, 579, 580, 594
eigen, 514
inverse, 324
of vectors, 580
orthogonal, 559
positive semidefinite, 361, 522
projector, 642
pseudoinverse, 603
trace, 513
vectorization, 513
Kuhn, 368

– L –, 203
Lagrange multiplier, 150, 574
sign, 184
Lagrangian, 183, 189, 317, 327, 368
Lanckriet, 319
Lanczos, 563
Laplace transfer function, 229, 285
INDEX

Lasso, 322
lattice, 272–275, 279–281
regular, 29, 272
Laurent, 257, 259, 407, 412, 453, 457, 652
Lauterbur, 322
Law, 322, 580
law of cosines, 385
least
  1-norm, 191, 192, 252, 301
    nonnegative, 300, 301
    energy, 365, 498
    norm, 192, 253, 604
    squares, 270, 604
Legendre-Fenchel transform, 496
Lemaréchal, 21, 73
Lewis, 82, 92, 629
Liang, 376
line, 34, 35, 55, 213, 214, 224
tangential, 41
linear
algebra, 75, 511
bijection, see bijective
complementarity, 176
function, see operator
independence, 34, 64, 122, 125, 257, 302, 348
  matrix, 521
  of subspace, 34, 662
  preservation, 34
inequality, 33, 66, 146
  matrix, 156, 157, 245, 248
injective, see injective
operator, 43, 52, 185, 200, 201, 363, 394, 395,
  428, 450, 511, 628
  projector, 49, 608, 612, 614, 630
program, 74, 78–80, 140, 190, 201, 237–240,
  370, 669
dual, 140, 238
prototypical, 140, 238
regression, 322
surjective, see surjective
transformation, see transformation
list, 21, 28, 54, 663, 665
Littlewood, 489
Liu, 437, 442, 449, 502
localization, 22, 375
sensor network, 269–283
standardized test, 272
unique, 24, 270, 271, 374, 376
wireless, 269, 277, 380
log, 594, 598
barrier, 8, 370
-convex, 236
det, 226, 380, 499, 589, 597
Lowner, 101, 523
Luenberger, 630
Lustig, 323

M, 1–2, 170
machine learning, 23
Magaia, 538
majorization, 514
cone, 514
  symmetric hollow, 514
manifold, 23, 27, 62, 106, 460, 461, 483, 509, 557,
  568
map, see transformation
  isotonic, 424, 425
USA, 28, 423–429
Mardia, 485
Markov process, 463
Marziliano, 321
mater, 576
Mathar, 466, 556
Matlab, 31
backslash, 252
cartography, 423, 427
conic independence, 121
convex iteration
  accelerant, 353
  stall, 307
fast max cut, 319
MRI, 321, 324, 328–330
noise, 283
polynomial feasibility, 313
signal dropout, 299
SVD, 352, 536
matrix, 576
  angle
    interpoint, 363, 366
    relative, 386
  antisymmetric, 51, 515, 556, 666
    antihollow, 53, 480
auxiliary, 365, 552, 556
Householder, 553
orthonormal, 555
projector, 552
Schoenberg, 362, 554, 619, 665
table, 556
binary, see matrix, Boolean
Boolean, 254, 407
bordered, 405, 416
calculus, 575
circulant, 225, 533, 553
  permutation, 324, 336, 338–340
  symmetric, 324
commutative, see product
completion, see problem
correlation, 55, 407
covariance, 365
decomposition, see decomposition
determinant, see determinant
diagonal, see diagonal
diagonalizable, see diagonalizable
INDEX

direction, 264–267, 278, 320, 371, 501
existence, 264
Identity, 266, 345
interpretation, 265
distance
1-norm, 222, 415
Euclidean, see EDM
doublet, 389, 550, 628
nullspace, 550
range, 550
dyad, 543, 545
-decomposition, 543
hull, 59
independence, 80, 161, 422, 443, 548, 549
negative, 545
nullspace, 546
projection on, 621, 623
projector, 547, 548, 606, 620, 621
pseudo-inverse, 546
range, 546
sum, 531, 534, 549
symmetric, 110, 525, 526, 547
EDM, see EDM
elementary, 414, 550, 552
nullspace, 551
range, 551
exponential, 225, 599
factorization, see decomposition
fat, 50, 76, 344, 659
Fourier, 48, 323
full-rank, 76
Gale, 377
geometric centering, 365, 388, 438, 463, 552
Gram, 276, 363, 373
constraint, 366
unique, 389, 393
Hermitian, 515
hollow, 51–54
Householder, 552, 557
auxiliary, 553
idempotent, 608
nonsymmetric, 604
range, 605
symmetric, 609, 612
Identity, 670
direction vector, 266, 345
Fantope, 58, 59
orthogonal, 62, 309, 552, 557
permutation, 62, 309, 552
positive definite, 62, 111, 112, 526, 557
projection, 526
indefinite, 344, 415
inverse, 224, 523, 533, 547, 589, 659
injective, 602
minimize, see function, fractional
symmetric, 534
Jordan form, 519, 532
measurement, 479
nonexpansive, 556
nonnegative, 416, 450, 470, 479, 637
nonnegative versus, 416
-positive semidefinite
symmetric, 557
nonsingular, 533, 534
normal, 48, 519, 533, 535
normalization, 255
nullspace, see nullspace
orthogonal, 48, 309, 390, 533, 536, 556–559
manifold, 557
product, see product
symmetric, 557
orthonormal, 49, 59, 62, 265, 307, 389, 534,
536, 555, 566, 611
manifold, 62
nonexpansive, 556
pseudo-inverse, 602
partitioned, 526–530, see Schur
permutation, 62, 309, 401, 514, 534, 557
circulant, 324, 336, 338–340
extreme point, 62, 339
positive semidefinite, 62
product, see product
symmetric, 324, 552, 570
positive definite, 101
inverse, 523
positive real numbers, 221
positive factorization, 268
positive semidefinite, 99–121, 154, 416,
515–531
completely, 89, 268
difference, 120
domain of test, 515, 625
from extreme directions, 111
Hadamard, 522
Kronecker, 361, 522
nonnegative versus, 416
nonsymmetric, 518
permutation, 62
projection, 526, 611
pseudo-inverse, 603
rank, see rank
square root, 523, 534, 660
sum, eigenvalues, 524
sum, nullspace, 80
sum, rank, see rank
symmetry versus, 515
zero entry, 539
product, see product
projection, 103, 553, 611
Kronecker, 642
nonorthogonal, 604
orthogonal, 609
positive semidefinite, 526, 611
product, see product
pseudoinverse, 103, 163, 212, 253, 526, 601–603
Kronecker, 603
nullspace, 601, 602
of dyad, 546
of vector, 602
orthonormal, 602
positive semidefinite, 603
product, see product
range, 601, 602
SVD, 538
symmetric, 538, 603
transpose, 159
unique, 601
quotient, Hadamard, 592, 599, 662
range, see range
rank, see rank
reflection, 389, 390, 557, 570
rotation, 389, 390, 556, 558, 569
of range, 558
rotation of, 559
Schur-form, see Schur
simple, 545
skiny, 49, 163, 659
sort index, 425
square root, 660, see matrix, positive
squared, 225
Stiefel, see matrix, orthonormal
stochastic, 62, 309
sum
    eigenvalues, 524, 525
    nullspace, 80
    rank, see rank
symmetric, 46, 50–54, 515, 533
    Hermitian, 515
    inverse, 534
    pseudoinverse, 538, 603
    real numbers, 221
    subspace of, 50
symmetrized, 515
trace, see trace
transpose, 659
    conjugate, 515, 557, 659
    unitary, 557
symmetric, 324, 325
zero definite, 542, 543
max cut problem, 317
maximization, 203, 204, 342, 570, see supremum
maximum variance unfolding, 461, 463
McCormick, 239
membership, 57, 90, 92, 101, 672
boundary, 158
interior, 158
relation, 143
    affine, 189
    boundary, 145, 152, 248
discretized, 150, 151, 154, 475
EDM, 475
interior, 145, 247
normal cone, 173
orthant, 57, 152
subspace, 166
Menger, 400, 417, 433, 434, 488, 507
metric, 48, 222, 356
    postulate, 357
    property, 357, 402
    fifth, 357, 359, 360, 429, 431, 436
    space, 357
minimal
cardinality, see minimization
element, 91, 92, 188, 189
rank, see minimization
set, see set
minimization, 149, 173, 271, 561, 585
affine function, 202
    Boolean, 252
    nonnegative, 193, 194, 196, 289–295, 299,
        300, 302–304
    rank connection, 265
    rank one, 319
    rank, full, 257
fractional, inverted, see function
inverted, see function, fractional
norm
    1-, 191, see problem
    2-, 190, 367, 368
    ∞-, 191
    Frobenius, 191, 209, 212, 483, 627
    nuclear, 345, 497, 563
of concave functions, 236
of distance, 609, 626
of quasiconcave functions, 236
of Rayleigh quotient, 564
on hypercube, 74
over totally ordered set, 671
rank, 263, 264, 307, 346, 349, 500, 506
    cardinality connection, 265
    cardinality constrained, 319
    fat, 344
    indefinite, 344
    trace, 265–267, 345, 375, 496–498, 528, 563
minimizer, 187, 189
minimum
element, 91, 189
    unique, 92, 188
    global, 185, 187
    unique, 187, 585
Minkowski, 128
MIT, 346
Mizukoshi, 54
molecular conformation, 22, 25, 29, 383
Monckton, 330, 333
### INDEX

monotonic, 208, 214, 267, 292, 647, 650
Fejér, 647
function, 185, 195, 197, 214, 229, 235, 292
strictly, 214
gradient, 215
noisily, 268, 292
Moore-Penrose
conditions, 601
inverse, see matrix, pseudoinverse
Moreau, 176, 635
Motzkin, 119, 148, 313, 629
MRI, 321, 322, 326
Muller, 538
multidimensional
function, 185, 221, 669
affine, 189, 201
convexity, 222, 224
gradient, 210
line theorem, 224
monotonic, 214
quasiconvex, 227
objective, 91, 92, 187, 188, 289, 327
scaling, 22, 421
ordinal, 425
multilateration, 270, 380
multiobjective optimization, 91, 92, 187–189, 289, 327
multipath, 382
multiplicity, 520, 532
Musin, 370, 372

– N –, 76

neighborhood graph, 460, 461
neighbors
nearest, 27, 460–464
Németh, 149, 176
Nemirovski, 479
nested
sequence, 404
sublevel set, 219
Newton, 370, 575
direction, 211
Nigam, 383
Nirenberg, 630
node, 272–275, 279–282, 316, 317
nondegeneracy, 357
nonexpansive, 176, 556, 605, 611, 638
nonnegative, 400, 673

constraint, 62, 78, 79, 176, 190, 191, 193, 194, 238, 306
factorization, 268
part, 197, 480, 638, 639, 663
polynomial, 92, 528
nonnegativity, 190, 357
nonorthogonal
basis, 163, 615
projection on, 615
projection, see projection
nonvertical, 213
norm, 187, 190, 191, 194, 289, 673
0, 539
0-, see 0-norm
1-, see 1-norm
2-, see 2-norm
∞-, see ∞-norm
k-largest, 196, 292, 293, 343, 673
gradient, 197
\( \ell_p \)-, 673
ball, see ball
constraint, 306, 314, 539
spectral, 62
convexity, 63, 187, 190–192, 210, 215, 562, 673
dual, 143, 562, 673
dyad, 546
equivalence, 190
Euclidean, see 2-norm
Frobenius, 48, 187, 191, 265, 483, 673
minimization, 191, 209, 212, 483, 627
Schur-form, 209
gradient, 197, 593, 595, 610, 618, 619, 627, 629, 630
inequality, 253, 673
triangle, 190
Ky Fan, 562
least, 192, 253, 604
nuclear, 60, 265, 345, 497, 562, 563, 673
ball, 59–61, 563
orthogonally invariant, 49, 428, 483, 556, 559
property, 190
regularization, 315, 604
spectral, 63, 210, 265, 483, 492, 536, 546, 551, 552, 563, 639, 673
ball, 63, 640
constraint, 62
dual, 562, 673
inequality, 673
Schur-form, 210, 563
square, 187, 595, 618, 619, 627, 673
es. square root, 190, 215, 483, 610

normal, 68, 174, 630
cone, see cone
equation, 603
facet, 159
inward, 66, 465
outward, 66
vector, 68, 630, 655

Notation, 659
NP-hard, 500
nuclear
magnetic resonance, 383
norm, 60, 265, 345, 497, 562, 563, 673
ball, 59–61, 563
nullspace, 68, 76–81, 553, 628
INDEX

algebra, 545
basis, 77, 81, 528, 610
dimension, 75
doublet, 550
dyad, 546
elementary matrix, 551
form, 75
hyperplane, 65, 68
injective, 48
intersection, 78
inverse, 45, 48
orthogonal complement, 75
product, 521, 545, 549, 610
projector, 610
pseudoinverse, 601, 602
set, 153
sum, 80
trick, 68
vectorized, 81
numerical precision, see precision

− O −, 670
objective, 74, 669
convex, 237
strictly, 494, 498
function, 74, 138, 149, 173, 175, 669
linear, 201, 278, 375
multidimensional, 91, 92, 187, 188, 289, 327
nonlinear, 203
polynomial, 311
quadratic, 209
nonconvex, 574
real, 187
value, 249, 260
offset, see invariance, translation
on, 665
one-to-one, 666, see injective
only if, 662
onto, 665, see surjective
open, see set
operator, 669
adjoint, 46, 141, 471, 659, 664
self, 46, 367, 471, 511
affine, see function
injective, see injective
invertible, 45, 48, 49, 156
linear, 43, 52, 185, 200, 201, 363, 394, 395, 428, 450, 511, 628
projector, 49, 608, 612, 614, 630
nonlinear, 186, 203
nullspace, 389, 393–395, 628
permutation, 195, 290, 488, 491, 508, 664
projection, see projector
surjective, see surjective
unitary, 48, 557
optimal, 7, 21, 668

optimality
condition, 250
conic problem, 138, 175, 177
direction vector, 264, 290
directional derivative, 585
dual, 135
first order, 149, 173–177
KKT, 176, 177, 486, 574
linear program, 138
semidefinite program, 250
Slater, 138, 184, 247, 250
unconstrained, 149, 211, 212, 627
constraint
conic, 175, 177
equality, 149, 177
inequality, 176
nonnegative, 176
global, 7, 21, 187, 208, 237
convergence, 264, 267, 290, 351
local, 7, 21, 187, 237
convergence, 267, 292, 295, 305, 351
perturbation, 260
optimization, 21, 237
algebra, 190, 561, 562
conic, see problem
convex, 7, 21, 187, 237, 373
art, 8, 256, 271, 510
fundamental, 237
solution set, 187, 349
invariance, 190, 202, 265, 561, 562, 669
multiobjective, 91, 92, 187–189, 289, 327
tractable, 238
vector, 91, 92, 187–189, 289, 327
order
natural, 46, 511, 669
nonincreasing, 489, 534, 535, 664
of projection, 480–482, 640, 654
partial, 57, 88, 90–92, 101, 152, 159, 160, 523, 669, 672
generalized inequality, 90, 92, 101
orthant, 152
property, 90, 92
total, 90, 91, 134, 670, 671
transitivity, 90, 101, 523, 669
origin, 33, 34
cone, 63, 88, 90, 93, 95, 123
projection of, 191, 192, 604, 618
projection on, 633, 639
translation to, 397
Orion nebula, 23
orthant, 35, 131, 134, 143, 152, 162, 173
dual, 156
extreme direction, 79
nonnegative, 146, 470
orthogonal, 46, 609, 626
basis, 52, 54, 80, 612
projection on, 615
complement, 50, 75, 265, 609, 612, 662, 667
cone, dual, 134, 142, 143, 635
cone, normal, 655
vector, 667
condition, 609
constraint, 350, 384, see Procrustes equivalence, 559
expansion, 52, 134, 162, 611–613
invariance, 49, 428, 483, 556, 559
matrix, see matrix
projection, see projection set, 153
sum, 662
orthonormal, 46
condition, 162, 611
constraint, 307
derecomposition, 611–613
matrix, see matrix
over, 665
overdetermined, 659

PageRank, 463
Pardalos, 377
pattern recognition, 22
Penrose conditions, 601
pentahedron, 434
pentatope, 131, 434
permutation, see matrix
constraint, see constraint
perpendicular, see orthogonal perturbation, 257, 258
optimality, 260
Pfender & Ziegler, 370
phantom, 322
plane, 63
segment, 95
point
boundary, 83
of hypercube slice, 330
exposed, see exposed
feasible, see solution
fixed, 267, 328, 628, 643–647
inflection, 220
minimal, see minimal element
Pólya, 489
Polyak, 561
polychoron, 43, 127, 131, 432, 434
polyhedron, 38, 54, 74, 127, 406
bounded, 57, 62, 129, 194
face, 120, 127
halfspace-description, 127
norm ball, 193
permutation, 62, 63, 309, 311, 339
stochastic, 62, 309, 311, 339
transformation
linear, 123, 127, 146
unbounded, 35, 38, 128, 129
vertex, 62, 128, 193
vertex-description, 57, 129

polynomial
constraint, 198, 311, 312
convex, 220
Motzkin, 313
nonconvex, 311
nonnegative, 92, 528
objective, 311
quadratic, see function
polytope, 127
positive, 673
completely, 268
factorization, 268
semidefinite, see matrix
strictly, 402
power, see function, fractional
Précis, 399
precision
antisymmetry, 128
barrier, 241
eigenvalue, 531
inequalities complementary, 128
numerical, 351, 352
perturbation, 258
quadruple, 233
solver, 8, 233, 254, 284, 327, 333, 340, 341, 370, 373, 479, 502
variables, 240
presolver, 302–304, 333, 341, 342
primal
feasible set, 241, 245
problem, 138, 239
via dual, 140, 251
principal
component analysis, 319, 421
eigenvalue, 319
eigenvector, 319, 320
submatrix, 108, 373, 430, 433, 448, 525
leading, 108, 402, 404
principle
halfspace-description, 66
separating hyperplane, 75
supporting hyperplane, 71
problem
1-norm, 191–193, 195, 252, 305, 306
nonnegative, 193, 194, 291, 292, 294, 300–304, 497
ball packing, 370
Boolean, 62, 74, 252, 313, 338, 407
cardinality, see cardinality
INDEX

complementarity, 176
linear, 176
genometry, 458
semidefinite, 651, 652
compressed sensing, see compressed concave
conic, 138, 175, 177, 189, 238
dual, 138, 238
convex, 138, 149, 175, 177, 191, 201, 208, 210, 240, 246, 492, 629, 669
definition, 237
gometry, 21, 24
nonlinear, 237
statement as solution, 8, 237
tractable, 238
dual, see dual
epigraph form, 209, 504
equivalent, 190, 202, 265, 561, 562, 669
feasibility, 121, 130, 152, 153, 669, 671
semidefinite, 249, 263, 264, 349
fractional, inverted, see function
inverted, see function, fractional
kissing, 370
max cut, 317
minimax, 57, 135, 140
nonconvex, 483, 500
nonlinear, 373
convex, 237
open, 28, 117, 267, 420, 449, 510
primal, see primal
Procrustes, see Procrustes
proximity, 479, 480, 483, 485, 493, 502
EDM, nonconvex, 490
in spectral norm, 492
rank heuristic, 498, 500
SDP, Gram form, 495
semidefinite program, 487, 494, 503
quadratic
nonconvex, 574
rank, see rank
same, 669
sphere packing, 370
stress, 220, 484, 500
tractable, 238
procedure
rank reduction, 258
Procrustes
combinatorial, 308
compressed sensing, 311
diagonal, 573
linear program, 571
maximization, 570
orthogonal, 568
two sided, 570, 572
orthonormal, 307
permutation, 308
symmetric, 572
translation, 569
unique solution, 568
product, 227, 580
Cartesian, 43, 245, 649, 666
cone, 89, 90, 140
function, 189
commutative, 225, 251, 520–522, 541, 642
determinant, 521
eigenvalue, 520, 522, 568
–of, 522
fractional, see function
function, 206, 216, 227–229, 235
gradient, 578
Hadamard, 46, 493, 511, 513, 580, 662
of vectors, 580
positive semidefinite, 522
inner
matrix, 363
vector, 45, 65, 229, 239, 321, 385, 556, 621, 663
vector, positive semidefinite, 154
vectorized matrix, 45–48, 519
Kronecker, 103, 511, 513, 571, 579, 580, 594
eigen, 514
inverse, 324
of vectors, 580
orthogonal, 559
permutation, 557
positive semidefinite, 361, 522
projector, 642
pseudo inverse, 603
trace, 513
vectorization, 513
nullspace, 521, 545, 549, 610
orthogonal, 557
outer, 58, 59, 61, 62, 265, 566
positive semidefinite, 105
vector, see matrix, dyad
permutation, 557
positive definite, 516
nonsymmetric, 516
positive semidefinite, 522, 525
matrix outer, 105
nonsymmetric, 522
vector inner, 154
zero, 541
projector, 467, 621, 642
pseudo fractional, see function
pseudo inverse, 602
Kronecker, 603
quasiconcave, 229
range, 521, 545, 549, 610
INDEX

rank, 520, 521
symmetric, 521
tensor, 579
trace, see trace
zero, 541

program, 74, 237
cross, see problem
gometric, 7, 240
integer, 62, 74, 302, 338, 407
linear, 74, 78–80, 140, 190, 201, 237–240, 370, 669
dual, 140, 238
prototypical, 140, 238
nonlinear
convex, 237
quadratic, 240, 428
second-order cone, 240
semidefinite, 140, 190, 201, 237–240, 310, 669
dual, 140, 239
equality constraint, 101
Gram form, 504
prototypical, 140, 239, 249, 263, 669
Schur-form, 198, 209, 210, 428, 494, 504, 563, 637
projection, 48, 601
$I - P$, 608, 634–636
algebra, 614
alternating, 369, 641, 643
angle, 648
convergence, 646, 648, 649, 652
distance, 644, 645
Dykstra algorithm, 642, 653
feasibility, 644–646
iteration, 641, 655
on affine ∩ PSD cone, 650
on EDM cone, 503
on halfspaces, 644
on orthant ∩ hyperplane, 647, 648
on subspace/affine, orthogonal, 642
optimization, 644, 645, 653
over/under, 649
biorthogonal expansion, 612, 616, 617
coefficient, 159, 454, 620, 625
cyclic, 642
direction, 609, 612, 621, 622
nonorthogonal, 606–608, 616, 620
parallel, 613, 614, 617, 618
dual, see dual
easy, 638
Euclidean, 118, 480, 483, 485, 487, 493, 502, 629
matrix, see matrix
minimum-distance, 149, 608, 610, 626, 629
nonorthogonal, 364, 484, 492, 604, 607, 608, 617, 620
eigenvalue coefficients, 621
of origin, 191, 192
oblique, see projection, nonorthogonal
of convex set, 45
of hypercube, 47
of matrix, 626, 627
of origin, 191, 192, 604, 618
of polyhedron, 127
of PSD cone, 104
on 1-norm ball, 640
on affine, 45, 614, 619
hyperplane, 618, 620
of origin, 191, 192, 604, 618
orthogonal, 642
vertex-description, 619
on basis
nonorthogonal, 615
orthogonal, 615
on box, 639
on cardinality $k$, 639
on complement, 608, 634
on cone, 633, 635, 639
don, see dual
Lorentz (second-order), 640
monotone nonnegative, 419, 428, 488–490, 639
polar, see dual, projection, on cone
polyhedral, 640
simplicial, 639
truncated, 637
on cone, EDM, 493, 494, 502, 506, 507
don, see dual
on cone, PSD, 118, 485–487, 639
cardinality constrained, 319
geometric center subspace, 467
rank constrained, 212, 487, 492, 564
rank one, 319
on convex set, 149, 629, 630
in affine subset, 640
in subspace, 640
on convex sets, 642
on dyad, 621, 623
on ellipsoid, 306
on Euclidean ball, 638
on function domain, 205
on geometric center subspace, 467, 628
on halfspace, 620
on hyperbox, 639
on hypercube, 639
on hyperplane, 618, 620, see projection, on affine
on hypersphere, 638
on intersection, 640
on line, 619
on matrices
rank constrained, 639
on matrix, 620–622
on origin, 633, 639
on orthant, 489, 637, 638
INDEX

in subspace, 489
on range, 603
on rowspace, 50
on simplex, 640
on slab, 620
on spectral norm ball, 640
on subspace, 45, 610
Cartesian, 639
elementary, 616
hollow symmetric, 485, 638
matrix, 626
orthogonal, 642
orthogonal complement, 85, 134, 634
polyhedron, 127
symmetric matrices, 638
on vector
cardinality k, 639
nonorthogonal, 620
orthogonal, 622
one-dimensional, 622
order of, 480–482, 640, 654
orthogonal, 603, 609, 622
eigenvalue coefficients, 624
semidefiniteness test as, 625
range/rowspace, 629
spectral, 463, 488, 490
norm, 492, 639
unique, 490
successive, 642
two sided, 626, 627, 629
unbral, 45
unique, 608, 610, 626, 629, 630
vectorization, 454, 627
projector, 45, 610, 630, 669
affine subset, 614
auxiliary matrix, 552
characteristic, 608, 609, 612
nonorthogonal, 605
orthogonal, 611
commutative, 642
non, 467, 640, 642
complement, 609
convexity, 608, 612, 629, 630
direction, see projection
dyad, 547, 548, 606, 620, 621
fractional function, 207
Kronecker, 642
linear operator, 49, 608, 612, 614, 630
nonexpansive, 176, 605, 611, 638
nonorthogonal, 364, 608, 614, 620
affine subset, 605
nullspace, 610
orthogonal, 610–613
product, 467, 621, 642
range, 610
rank/trace, 609, 612
semidefinite, 526, 611
subspace, fundamental, 610
unique, 610, 627
prototypical
complementarity problem, 176
compressed sensing, 301
coordinate system, 134
linear program, 140, 238
SDP, 140, 239, 249, 263, 669
signal, 321
PSD, 667
pseudoinverse, see matrix
pyramid, 434, 435
– Q ~, 556
quadrant, 35
quadratic, see function
quadrature, 334, 389
quartix, 577, 668
quasi-, see function
quotient
Hadamard, 592, 599, 662
Rayleigh, 625
– R ~, 76
range, 56, 76, 626, 666
dimension, 49
doublet, 550
dyad, 546
elementary matrix, 551
form, 75
idempotent, 605
orthogonal complement, 75
product, 521, 545, 549, 610
projector, 610
pseudoinverse, 601, 602
rotation, 558
rank, 520, 521, 535, 536, 671
constraint, see constraint
convex envelope, see convex
convex iteration, see convex
diagonalizable, 520
dimension, 122, 398, 399, 671
affine, 400, 500
full, 76, 257
heuristic, 498, 500
indefinite, 344
inequality, 523
log det, 499, 500
minimization, 263, 264, 307, 346, 349, 500, 506
cardinality connection, 265
cardinality constrained, 319
fat, 344
indefinite, 344
<table>
<thead>
<tr>
<th>Index Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>One, 526, 545</td>
<td></td>
</tr>
<tr>
<td>Boolean, 254, 407</td>
<td></td>
</tr>
<tr>
<td>Convex iteration, 307, 349, 350, 509</td>
<td></td>
</tr>
<tr>
<td>Hull, see hull, convex modification, 547, 551</td>
<td></td>
</tr>
<tr>
<td>PSD, 110, 525, 526, 547</td>
<td></td>
</tr>
<tr>
<td>Subset, 212</td>
<td></td>
</tr>
<tr>
<td>Symmetric, 110, 525, 526, 547</td>
<td></td>
</tr>
<tr>
<td>Transformation, 349, 350</td>
<td></td>
</tr>
<tr>
<td>Update, see rank, one, modification</td>
<td></td>
</tr>
<tr>
<td>Partitioned matrix, 529</td>
<td></td>
</tr>
<tr>
<td>Positive semidefinite cone, 118, 241</td>
<td></td>
</tr>
<tr>
<td>Face, 106–109</td>
<td></td>
</tr>
<tr>
<td>Précis, 399</td>
<td></td>
</tr>
<tr>
<td>Product, 520, 521</td>
<td></td>
</tr>
<tr>
<td>Projection matrix, 609, 612</td>
<td></td>
</tr>
<tr>
<td>Quasiconcavity, 117</td>
<td></td>
</tr>
<tr>
<td>Reduction, 241, 257, 258, 261</td>
<td></td>
</tr>
<tr>
<td>Procedure, 258</td>
<td></td>
</tr>
<tr>
<td>Regularization, 288, 289, 307, 315, 348, 349, 500, 509</td>
<td></td>
</tr>
<tr>
<td>Schur-form, 529, 530</td>
<td></td>
</tr>
<tr>
<td>Sum, 117</td>
<td></td>
</tr>
<tr>
<td>Positive semidefinite, 117, 521</td>
<td></td>
</tr>
<tr>
<td>Trace heuristic, 265–267, 345, 496–498, 563</td>
<td></td>
</tr>
<tr>
<td>Zero, 107, 120, 241, 242, 267</td>
<td></td>
</tr>
<tr>
<td>Ray, 85</td>
<td></td>
</tr>
<tr>
<td>Boundary of, 85</td>
<td></td>
</tr>
<tr>
<td>Cone, 85, 89, 90, 123, 130</td>
<td></td>
</tr>
<tr>
<td>Boundary, 93</td>
<td></td>
</tr>
<tr>
<td>EDM, 97</td>
<td></td>
</tr>
<tr>
<td>Positive semidefinite, 95</td>
<td></td>
</tr>
<tr>
<td>Extreme, 95, 451</td>
<td></td>
</tr>
<tr>
<td>Interior, 85</td>
<td></td>
</tr>
<tr>
<td>Rayleigh quotient, 625</td>
<td></td>
</tr>
<tr>
<td>Realizable, 21, 369, 484, 485</td>
<td></td>
</tr>
<tr>
<td>Reconstruction, 421</td>
<td></td>
</tr>
<tr>
<td>Isometric, 376, 423</td>
<td></td>
</tr>
<tr>
<td>Isotonic, 423–429</td>
<td></td>
</tr>
<tr>
<td>List, 421</td>
<td></td>
</tr>
<tr>
<td>Unique, 358, 377, 378, 391</td>
<td></td>
</tr>
<tr>
<td>Recursion, 328, 473, 511, 525</td>
<td></td>
</tr>
<tr>
<td>Reflection, 389, 390, 570</td>
<td></td>
</tr>
<tr>
<td>Invariance, 389</td>
<td></td>
</tr>
<tr>
<td>Prevention, 389</td>
<td></td>
</tr>
<tr>
<td>Vector, 557</td>
<td></td>
</tr>
<tr>
<td>Reflexivity, 90, 523, 669</td>
<td></td>
</tr>
<tr>
<td>Regular</td>
<td></td>
</tr>
<tr>
<td>Lattice, 29, 272</td>
<td></td>
</tr>
<tr>
<td>Simplex, 434</td>
<td></td>
</tr>
<tr>
<td>Tetrahedron, 431, 441</td>
<td></td>
</tr>
<tr>
<td>Regularization</td>
<td></td>
</tr>
<tr>
<td>Cardinality, 299, 320</td>
<td></td>
</tr>
<tr>
<td>Norm, 315, 604</td>
<td></td>
</tr>
<tr>
<td>Rank, 288, 289, 307, 315, 348, 349, 500, 509</td>
<td></td>
</tr>
<tr>
<td>Relaxation, 62, 74, 195, 254, 255, 375, 407, 427</td>
<td></td>
</tr>
<tr>
<td>Reweighting, 292, 499</td>
<td></td>
</tr>
<tr>
<td>Riemann, 240</td>
<td></td>
</tr>
<tr>
<td>Robberto, 23</td>
<td></td>
</tr>
<tr>
<td>Robotics, 23, 30</td>
<td></td>
</tr>
<tr>
<td>Rockafellar, 73, 93, 128, 135</td>
<td></td>
</tr>
<tr>
<td>Root, see function, fractional rotation, 389, 558, 569</td>
<td></td>
</tr>
<tr>
<td>Invariance, 389</td>
<td></td>
</tr>
<tr>
<td>Of matrix, 559</td>
<td></td>
</tr>
<tr>
<td>Of range, 558</td>
<td></td>
</tr>
<tr>
<td>Vector, 114, 390, 556</td>
<td></td>
</tr>
<tr>
<td>Round, 671</td>
<td></td>
</tr>
<tr>
<td>Rutman, 141</td>
<td></td>
</tr>
<tr>
<td>$\mathcal{S}$, 131</td>
<td></td>
</tr>
<tr>
<td>Saddle, 139</td>
<td></td>
</tr>
<tr>
<td>Value, 138, 139, 239</td>
<td></td>
</tr>
<tr>
<td>Same, 669</td>
<td></td>
</tr>
<tr>
<td>Saul, 27, 461, 462</td>
<td></td>
</tr>
<tr>
<td>Saunders, 233, 340, 341, 344, 547</td>
<td></td>
</tr>
<tr>
<td>Scaling, 421</td>
<td></td>
</tr>
<tr>
<td>Frequency, 233</td>
<td></td>
</tr>
<tr>
<td>Invariance, 236, 362, 561</td>
<td></td>
</tr>
<tr>
<td>Multidimensional, 22, 421</td>
<td></td>
</tr>
<tr>
<td>Ordinal, 420</td>
<td></td>
</tr>
<tr>
<td>Unidimensional, 500</td>
<td></td>
</tr>
<tr>
<td>Schütte &amp; van der Waerden, 370</td>
<td></td>
</tr>
<tr>
<td>Schoenberg, 27, 355, 364, 407, 412, 414, 427, 484, 485</td>
<td></td>
</tr>
<tr>
<td>Auxiliary matrix, 362, 554, 619, 665</td>
<td></td>
</tr>
<tr>
<td>Criterion, 27–29, 364, 365, 417, 438, 450, 454, 471, 475, 478, 626</td>
<td></td>
</tr>
<tr>
<td>Schur, 514</td>
<td></td>
</tr>
<tr>
<td>-form, 526–530</td>
<td></td>
</tr>
<tr>
<td>Anomaly, 210</td>
<td></td>
</tr>
<tr>
<td>Constraint, 63</td>
<td></td>
</tr>
<tr>
<td>Convex set, 102</td>
<td></td>
</tr>
<tr>
<td>Norm, Frobenius, 209</td>
<td></td>
</tr>
<tr>
<td>Norm, spectral, 210, 563</td>
<td></td>
</tr>
<tr>
<td>Nullspace, 528</td>
<td></td>
</tr>
<tr>
<td>Rank, 529, 530</td>
<td></td>
</tr>
<tr>
<td>Semidefinite program, 198, 209, 210, 428, 494, 504, 563, 637</td>
<td></td>
</tr>
<tr>
<td>Sparse, 529</td>
<td></td>
</tr>
<tr>
<td>Complement, 232, 277, 375, 526, 527, 529</td>
<td></td>
</tr>
<tr>
<td>Schwarz, 78, 646</td>
<td></td>
</tr>
<tr>
<td>SDP, 667</td>
<td></td>
</tr>
<tr>
<td>Semidefinite</td>
<td></td>
</tr>
<tr>
<td>Positive, see matrix</td>
<td></td>
</tr>
<tr>
<td>Program, see program</td>
<td></td>
</tr>
<tr>
<td>Sensor, 24, 269, 270, 272–275, 279–282</td>
<td></td>
</tr>
<tr>
<td>Network localization, 269–283, 373, 377</td>
<td></td>
</tr>
<tr>
<td>Sequence, 647</td>
<td></td>
</tr>
<tr>
<td>Nested, 404</td>
<td></td>
</tr>
<tr>
<td>Set</td>
<td></td>
</tr>
<tr>
<td>Active, 263</td>
<td></td>
</tr>
<tr>
<td>Cartesian product, 43</td>
<td></td>
</tr>
</tbody>
</table>
closed, 34, 37, 38, 65, 66, 156
cone, 99, 105, 123, 127, 128, 140, 141, 146, 156, 157, 246, 247, 440
exposed, 449
polyhedron, 123, 127
connected, 33
cvx, 33
invariance, 43
projection on, 149, 629, 630
difference, 43, 141
empty, 35–37, 43, 561
affine, 35
closed, 38
cone, 88
face, 82
hull, 55, 57, 63
open, 38
feasible, 7, 42, 62, 149, 173, 187, 237, 668
dual, 245
extreme point, 264
primal, 241, 245
intersection, see intersection
invariance, 414
level, 174, 201, 203, 211, 219, 235, 238
affine, 201
minimal, 57, 69, 125, 342, 666
affine, 619
extreme, 96, 158
halfspace, 125, 126
hyperplane, 126
orthant, 186
positive semidefinite cone, 100, 154, 221
nonconvex, 37, 85–88
nullspace, 153
open, 34–38, 145
and closed, 34, 38
orthogonal, 153
projection of, 45
Schur-form, 102
smooth, 408
solution, 187, 349, 668
sublevel, 174, 204, 211, 218, 223, 227
cvx, 205, 219, 528
nested, 219
quasisconvex, 227
sum, 43, 662
superlevel, 205, 227
quasisconcave, 227
union, 43, 85, 87, 109
Shannon, 321
shape, 22
shell, 277
Sherman-Morrison-Woodbury, 547
shift, see invariance, translation
shroud, 390, 391, 452
cone, 155, 477
SIAM, 238

sigma delta, 21, 22
signal
dropout, 295
processing, digital, 295, 321, 345, 421
simplex, 130, 131
area, 433
content, 434
method, 239, 241, 284, 302, 340, 370
nonnegative, 300–302, 640
regular, 434
unit, 130, 131
volume, 434
singular value, 535, 562
σ, 511, 512, 664
constraint, 63
decomposition, 534–538, 562
compact, 534
cvx iteration, 352
diagonal, 573
full, 536
geometrical, 537
pseudoinverse, 538
real, 535
subcompact, 535
symmetric, 538, 572
eigenvalue, 535, 538
inequality, 572
inverse, 538
largest, 63, 210, 563, 640, 673
normal matrix, 48, 535
sum of, 48, 59, 562, 563
symmetric matrix, 538
triangle inequality, 61
SIOPT, 240
skinny, 163, 659
slab, 35, 277, 620
slack
complementary, 250
variable, 239, 245, 263
Slater, 138, 184, 247, 250
slice, 113, 156, 380, 584
slope, 211
SNR, 667
solid, 127, 130, 431
solution, 237
analytical, 209, 237, 561
feasible, 74, 149, 247, 644–648, 669
dual, 250
strictly, 247, 250
global, 237
local, 237
numerical, 283
optimal, 74, 78–80, 187, 668
problem statement as, 8, 237
set, 187, 349, 668
trivial, 34
unique, 42, 187, 375, 376, 492, 494, 498, 506
<table>
<thead>
<tr>
<th>INDEX</th>
<th>721</th>
</tr>
</thead>
<tbody>
<tr>
<td>vertex, 74, 194, 195, 241, 254, 255, 301, 302, 311, 340, 370</td>
<td>orthogonal complement, 106, 389, 463, 628</td>
</tr>
<tr>
<td>sort function, 195, 290, 488, 491, 508, 664</td>
<td>hollow, 52, 392, 469</td>
</tr>
<tr>
<td>index matrix, 425</td>
<td>independence, 34, 662</td>
</tr>
<tr>
<td>largest entries, 195</td>
<td>intersection, 78</td>
</tr>
<tr>
<td>monotone nonnegative, 489</td>
<td>hyperplane, 76</td>
</tr>
<tr>
<td>smallest entries, 195</td>
<td>nullspace basis span, 77</td>
</tr>
<tr>
<td>span, 56, 666</td>
<td>parallel, 36</td>
</tr>
<tr>
<td>gradient, 322</td>
<td>representation, 75</td>
</tr>
<tr>
<td>nonnegative, 193, 294, 300–304</td>
<td>symmetric hollow, 52</td>
</tr>
<tr>
<td>spectral spectahedron, 59</td>
<td>tangent, 106</td>
</tr>
<tr>
<td>spectral cone, 416–418, 488</td>
<td>translation invariant, 389, 393, 463, 628</td>
</tr>
<tr>
<td>dual, 420</td>
<td>vectorized, 80, 81</td>
</tr>
<tr>
<td>orthant, 490</td>
<td>successive approximation, 642</td>
</tr>
<tr>
<td>factorization, 284–286</td>
<td>projection, 642</td>
</tr>
<tr>
<td>inequality, 416</td>
<td>sum, 43</td>
</tr>
<tr>
<td>norm, 63, 210, 265, 483, 492, 536, 546, 551, 552, 563, 639, 673</td>
<td>of eigenvalues, 48, 520</td>
</tr>
<tr>
<td>ball, 63, 640</td>
<td>of extremes, 95</td>
</tr>
<tr>
<td>constraint, 62</td>
<td>of functions, 189, 215, 236</td>
</tr>
<tr>
<td>dual, 562, 673</td>
<td>of matrices</td>
</tr>
<tr>
<td>inequality, 673</td>
<td>eigenvalues, 524, 525</td>
</tr>
<tr>
<td>Schur-form, 210, 563</td>
<td>nullspace, 80</td>
</tr>
<tr>
<td>projection, see projection</td>
<td>rank, 117, 521</td>
</tr>
<tr>
<td>sphere packing, 370</td>
<td>of singular values, 48, 59, 562, 563</td>
</tr>
<tr>
<td>square root, 199, 660, 665, see matrix</td>
<td>vector, 43, 548, 662</td>
</tr>
<tr>
<td>squared, see matrix</td>
<td>orthogonal, 662</td>
</tr>
<tr>
<td>Srebro, 61, 562</td>
<td>unique, 51, 134, 548, 662</td>
</tr>
<tr>
<td>Starck, 321</td>
<td>supremum, 57, 236, 342, 561–563, 570, 670</td>
</tr>
<tr>
<td>steepest descent, 211, 584</td>
<td>of affine functions, 203, 204, 562</td>
</tr>
<tr>
<td>Stiemke, 148</td>
<td>of convex functions, 203, 210, 236</td>
</tr>
<tr>
<td>strict complementarity, 251</td>
<td>of distance, 631</td>
</tr>
<tr>
<td>feasibility, 247, 250</td>
<td>of linear functions, 564</td>
</tr>
<tr>
<td>positivity, 402</td>
<td>of Rayleigh quotient, 564</td>
</tr>
<tr>
<td>triangle inequality, 403</td>
<td>supporting hyperplane, 71–73</td>
</tr>
<tr>
<td>Sturm, 543</td>
<td>surjective, 49, 392, 393, 395, 488, 665</td>
</tr>
<tr>
<td>subject to, 671</td>
<td>linear, 395, 396, 409, 411</td>
</tr>
<tr>
<td>submatrix principal, 108, 373, 430, 433, 448, 525</td>
<td>SVD, 667</td>
</tr>
<tr>
<td>leading, 108, 402, 404</td>
<td>svect, 51, 239, 672</td>
</tr>
<tr>
<td>subspace, 34</td>
<td>Swiss roll, 27</td>
</tr>
<tr>
<td>algebra, 34, 75, 76, 545</td>
<td>symmetry, 357</td>
</tr>
<tr>
<td>antihollow, 53</td>
<td>tangent</td>
</tr>
<tr>
<td>antisymmetric, 53, 480</td>
<td>line, 40, 375</td>
</tr>
<tr>
<td>antisymmetric, 51</td>
<td>subspace, 106</td>
</tr>
<tr>
<td>Cartesian, 291, 639</td>
<td>tangential, 40, 73, 376</td>
</tr>
<tr>
<td>complementary, see complement</td>
<td>Tanner, 294</td>
</tr>
<tr>
<td>fundamental, 45, 75, 76, 546, 550, 551, 602, 610</td>
<td>Tarazaga, 437, 442, 449, 502</td>
</tr>
<tr>
<td>projector, 610</td>
<td>Taylor series, 583, 588, 589</td>
</tr>
<tr>
<td>geometric center, 392, 393, 437, 463, 467–469, 496, 542, 628</td>
<td>tensor, 577–579</td>
</tr>
<tr>
<td>– T –, 659</td>
<td>tesseract, 47</td>
</tr>
<tr>
<td>tangent line, 40, 375</td>
<td>tetrahedron, 130, 131, 300, 432</td>
</tr>
<tr>
<td>subspace, 106</td>
<td>angle inequality, 431</td>
</tr>
<tr>
<td>tangential, 40, 73, 376</td>
<td>regular, 431, 441</td>
</tr>
<tr>
<td>Tanner, 294</td>
<td>Theobald, 520</td>
</tr>
</tbody>
</table>
theorem
  0 eigenvalues, 539
alternating projection
  distance, 646
alternative, 146–148, 153
EDM, 478
  semidefinite, 248
weak, 148
Barvinok, 120
Bunt-Motzkin, 629
Carathéodory, 145, 624
compressed sensing, 294, 321
cone
  faces, 93
  intersection, 89
conic coordinates, 182
covexity condition, 216
decomposition
dyad, 543
directional derivative, 585
discretized membership, 151
dual cone intersection, 178
duality, weak, 249
EDM, 409
eigenvalue
  of difference, 525
  of sum, 524
  order, 524
  zero, 539
elliptope vertices, 407
exposed, 97
extreme existence, 82
extremes, 95, 96
Farkas' lemma, 143, 147
  not positive definite, 248
  positive definite, 247
  positive semidefinite, 246
fundamental
  algebra, 532
  algebra, linear, 75
  convex optimization, 237
generalized inequality and membership, 143
Geršgorin discs, 115, 116
gradient monotonicity, 215
Hahn-Banach, 66, 71, 75
halfspaces, 66
Hardy-Littlewood-Pólya, 489
hypersphere, 409
inequalities, 489
intersection, 43
inverse image, 44, 47
  closedness, 156
Klee, 95
line, 224, 236
linearly independent dyads, 548
majorization, 514
mean value, 589
Minkowski, 128
monotone nonnegative sort, 489
Morse decomposition, 176, 635
Motzkin, 629
  transposition, 148
nonexpansivity, 638
pointed cones, 90
positive semidefinite, 518
  cone subsets, 118
  matrix sum, 117
  principal submatrix, 525
  symmetric, 526
projection
  algebraic complement, 635
  minimum-distance, 630
  on affine, 614
  on cone, 633
  on convex set, 630
  on PSD geometric intersection, 467
  on subspace, 45
  via dual cone, 636
  via normal cone, 630
projector
  rank/rank, 609, 612
  semidefinite, 526
proper-cone boundary, 93
Pythagorean, 312, 385, 641
range of dyad sum, 549
rank
  affine dimension, 400
  partitioned matrix, 529
  Schur-form, 529, 530
  /trace, 609, 612
  real eigenvector, 532
  sparse sampling, 294, 321
  sparsity, 193
  Tour, 412
  Weyl, 128
eigenvalue, 524
tight, 359, 669
Torgerson, 394
trace, 201, 511, 512, 562, 564, 595
tr, 529, 523, 671
commutative, 520
eigenvalues, 520
heuristic, 265–267, 345, 496–498, 563
inequality, 523
Kronecker, 513
maximization, 463
minimization, 265–267, 345, 375, 496–498, 528, 563
nonnegative, 523
of product, 46, 520, 523
positive semidefinite, 523
product, 513, 523
projection matrix, 609, 612
/rank gap, 497
zero, 539, 541
trajectory, 414, 457
sensor, 379
transformation
affine, 44, 71, 156, 215, 236
positive semidefinite cone, 102
bijective, see bijective
Fourier, discrete, 48, 323
injective, see injective
invertible, 45, 48, 49
linear, 34, 45, 70, 123, 602
cone, 89, 92, 123, 127, 146, 157, 158
cone, dual, 141, 146, 474
inverse, 45, 438, 602
polyhedron, 123, 127, 146
rank one, 349, 350
rigid, 22, 376
similarity, 624
surjective, see surjective
transitivity
face, 84
order, 90, 101, 523, 669
trefoil, 461, 462, 464
triangle, 357
inequality, 357, 358, 385, 402, 403, 405, 406,
412, 429, 431–433, 440, 441
norm, 190
strict, 403
unique, 430
vector, 190
triangulation, 24
trilateration, 42, 374
tandem, 377
Tucker, 147, 176, 486

− U − , 59
unbounded below, 74, 78, 79, 147
underdetermined, 346, 659
unfolding, 461, 463
unfurling, 461
unimodal, 226
unique
cone, 99, 134, 440
dual, 134, 158
direction
extreme, 95
EDM, 411, 422, 423, 487
eigen
value, 111, 532
vector, 532
expansion
biorthogonal, 161, 163, 164, 449, 549, 612,
616, 617
hull
affine, 55
convex, 57
hyperplane
supporting, 216–218
independence
conic, 122, 126, 163, 166
inequality
triangle, 430
localization, 24, 270, 271, 374, 376
matrix
Gram, 389, 393
pseudoinverse, 601
minimum, 187, 585
element, 92, 188
projection, 608, 610, 626, 629, 630
spectral, 490
projector, 610, 627
reconstruction, 358, 377, 378, 391
solution, 42, 187, 375, 376, 492, 494, 498, 506
Procrustes, 568
sum
vector, 51, 134, 548, 662
value
eigen, 111, 532
unraveling, 461, 462, 464
untieing, 462, 464

− V − , 365
value
absolute, 190, 220
eigen, see eigen
unique, 111, 532
objective, 249, 260
Van Loan, 515
variable
dual, 135, 176, 177, 184, 664
matrix, 69
nonnegative, see constraint
slack, 239, 245, 263
variation
calculus, 149
total, 326
vec, 46, 324, 512, 513, 672
vector, 34, 668
binary, 55, 255, 313, 408
direction, 264–267, 290–292
existence, 264
Identity, 266, 345
interpretation, 291, 292, 497
optimality condition, 264, 290
unity, 291
dual, 653
normal, 68, 630, 655
optimization, 91, 92, 187–189, 289, 327
parallel, 613, 614, 617, 618
Perron, 416
primal, 653
product
Hadamard, 580
inner, 45–48, 65, 229, 239, 321, 385, 519, 556, 621, 663
inner, positive semidefinite, 154
Kronecker, 580
outer, see matrix, dyad
pseudoinverse, 602
quadrature, 389
reflection, 557
rotation, 114, 390, 556
space, 33, 34, 666
ambient, 34, 465, 476
Euclidean, 356
vectorization, 46, see vec svec & dvec
projection, 454
symmetric, 51
hollow, 53
Venn
EDM, 481, 482
programs, 240
sets, 129
vertex, 39, 42, 62, 82, 83, 193
cone, 90
description, 63, 64, 96
of affine, 77
of dual cone, 165, 178
of halfspace, 125, 126
of hyperplane, 69
of polyhedral cone, 63, 129, 146, 158
of polyhedron, 57, 129
projection on affine, 619
elliptope, 254, 407, 408
polyhedron, 62, 128, 193
solution, 74, 194, 195, 241, 254, 255, 301, 302, 311, 340, 370
Vetterli, 321
von Neumann, 504, 572, 642–644, 648
Voronoi diagram, 381
vortex, 446
Ye, 148, 245, 378, 383
Youla, 640
Young, 30, 485, 487, 510, 565
– Z –, 68
Zenodorus, 561
zero, 539
definite, 542, 543
diagonal, 539
entry, 539
gradient, 149, 211, 212, 627
matrix product, 541
norm-, 539
rank, 107, 120, 241, 242, 267
trace, 539, 541
Zhang
Fuzhen, 516, 517
Shuzhong, 543
Ziegler, 370

– W –, 264
wavelet, 322, 323
wedge, 93, 94, 143
Weinberger & Saul, 27, 461, 462, 464
Wells, 209, 437, 442, 449, 450, 502, 503
Weyl, 128, 524
Wiener, 642
wireless location, 22, 269, 277, 380
womb, 576
Wright, 237
Wüthrich, 25

– X –, 665

– Y –, 224