DATTORRO

# Convex Optimization & Euclidean Distance Geometry

Mεβoo

Dattorro

# convex optimization & euclidean distance geometry

Meboo

# Convex Optimization
# &
# Euclidean Distance Geometry

Jon Dattorro

*$\mathcal{M}\varepsilon\beta oo$ Publishing*

Meboo Publishing USA
345 Stanford Shopping Ctr.
Suite 410
Palo Alto, California 94304-1413

Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* $\mathcal{M}\varepsilon\beta oo$, 2005.

ISBN 0976401304 (English)    ISBN 9781847280640 (International)

Version 2007.09.17 available in print as conceived in color.

cybersearch:

I. convex optimization
II. semidefinite programming
III. rank constraint
IV. convex geometry
V. distance matrix
VI. convex cones

programs by Matlab

typesetting by WinEdt

with donations from **SIAM** and **AMS**.

*for Jennie Columba*

◊

◊ *Antonio*

◊

& *Sze Wan*

$$\mathbb{EDM} = \mathbb{S}_h \cap \left(\mathbb{S}_c^\perp - \mathbb{S}_+\right)$$

# 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 [57]

Convex Analysis is the calculus of inequalities while Convex Optimization is its application. Analysis is inherently the domain of the mathematician while Optimization belongs to the engineer.



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

There is a great race under way to determine which important problems can be posed in a convex setting. 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 the convex equivalent differently; hence, one solution is likely different from the other for the same problem. Any presumption of only one right or correct solution becomes nebulous. Study of equivalence, sameness, and uniqueness therefore pervade study of Optimization.

Tremendous benefit accrues when an optimization problem can be transformed to its convex equivalent, primarily because any locally optimal solution is then guaranteed globally optimal. Solving a nonlinear system, for example, by instead solving an equivalent convex optimization problem is therefore highly preferable.[0.1] Yet it can be difficult for the engineer to apply theory without an understanding of Analysis.

These pages comprise my journal over a seven year period bridging gaps between engineer and mathematician; they constitute a translation, unification, and cohering of about two hundred papers, books, and reports from several different fields of mathematics and engineering. 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. Consequently there is much cross-referencing and background material provided in the text, footnotes, and appendices so as to be self-contained and to provide understanding of fundamental concepts.

−Jon Dattorro
Stanford, California
2007

[0.1] That is what motivates a convex optimization known as *geometric programming* [46, p.188] [45] which has driven great advances in the electronic circuit design industry. [27, §4.7] [181] [291] [294] [67] [130] [138] [139] [140] [141] [142] [143] [193] [194] [202]

# Convex Optimization
# &
# Euclidean Distance Geometry

# List of Figures

# List of Tables



 
Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry*, Meboo Publishing USA, 2005.

# 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. [302] [240]

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 follow by epiphany that many problems, presumed nonconvex, can be so transformed. [8] [9] [44] [63] [100] [102] [204] [221] [229] [271] [272] [299] [302] [27, §4.3, p.316-322]



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

Figure 1: *Orion nebula.* (astrophotography by Massimo Robberto)

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 [72].

It is a common misconception to presume that some desired point conformation cannot be recovered in the absence of complete interpoint distance information. We might, for example, want to realize a constellation given only interstellar distance (or, equivalently, distance from Earth and relative angular measurement; the Earth as vertex to two stars). 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 can have a shape (three points can form a triangle and its interior, for example, four points a tetrahedron), then

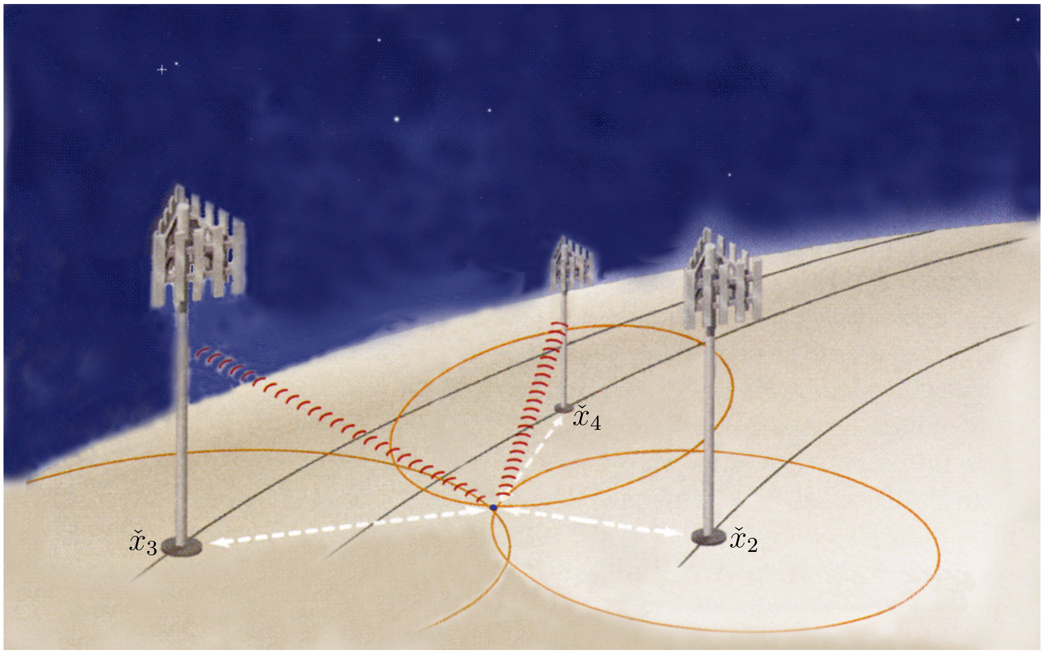



Figure 2: Application of trilateration (§5.4.2.2.4) is localization (determining position) of a radio signal source in 2 dimensions; more commonly known by radio engineers as the process "triangulation". In this scenario, anchors $\check{x}_2\,, \check{x}_3\,, \check{x}_4$ are illustrated as fixed antennae. [154] The radio signal source (a sensor $\bullet$ $x_1$) 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. So & Ye proved trilateration is expressible as a semidefinite program; hence, a convex optimization problem. [239]

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 fixed points of reference are also provided, it becomes possible to determine absolute position or location; *e.g.*, Figure **2**.

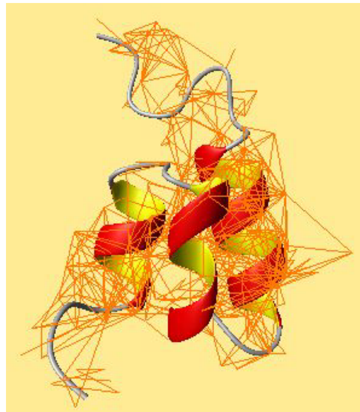


Figure 3: [137] [80] Distance data collected via nuclear magnetic resonance (NMR) helped render this 3-dimensional depiction of a protein molecule. *At the beginning of the 1980s, Kurt Wüthrich* [Nobel laureate], *developed an idea about how NMR could be extended to cover biological molecules such as 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.* [207] [286] [132]

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 **3**. [60] [265] [132] [286] [95] Many disciplines have already benefitted and simplified consequent to this theory; *e.g.*, distance-based *pattern recognition* (Figure **4**), *localization* in wireless sensor networks by measurement of intersensor distance along channels of communication, [35, §5] [297] [33] *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.

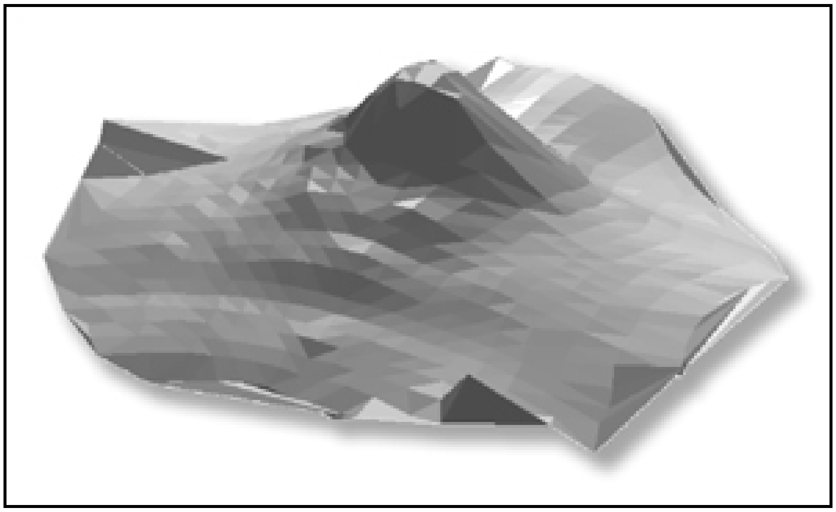


Figure 4: This coarsely discretized triangulated algorithmically flattened human face (made by Kimmel & the Bronsteins [165]) represents a stage in machine recognition of human identity; called *face recognition.* Distance geometry is applied to determine discriminating-features.

Euclidean distance geometry together with convex optimization have also found application in *artificial intelligence*:

- to *machine learning* by discerning naturally occurring manifolds in:
  - Euclidean bodies (Figure **5**, §6.4.0.0.1),
  - Fourier spectra of kindred utterances [156],
  - and image sequences [281],
- to *robotics*; *e.g.*, automatic control of vehicles maneuvering in formation. (Figure **8**)

## by chapter

We study the pervasive convex Euclidean bodies and their various representations. In particular, we make convex polyhedra, cones, and dual cones more visceral through illustration in **chapter 2**, **Convex geometry**, and we study the geometric relation of polyhedral cones to nonorthogonal

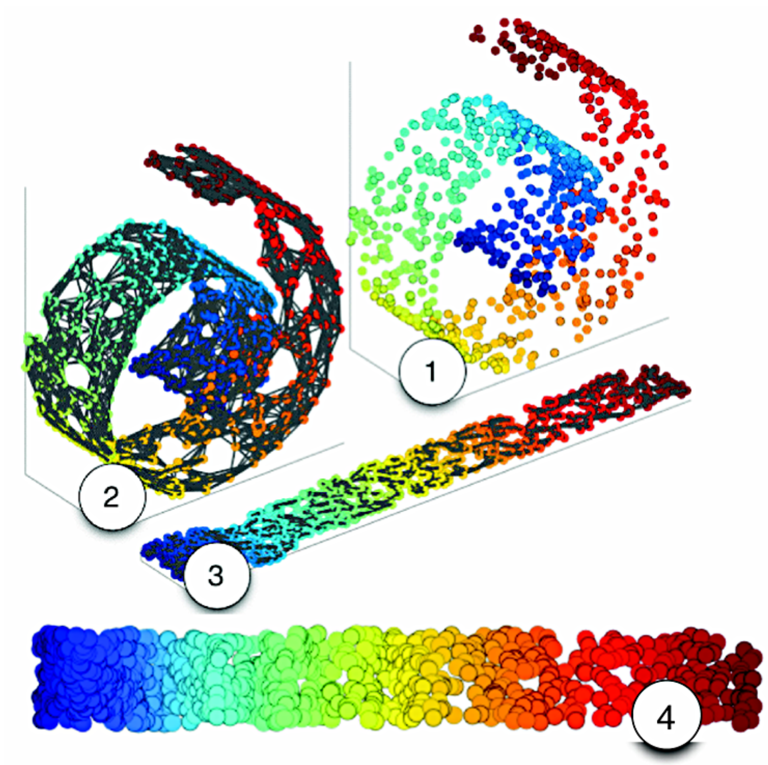



Figure 5: *Swiss roll* from Weinberger & Saul [281]. *The problem of manifold learning, illustrated for $N = 800$ data points sampled from a "Swiss roll" ①. A discretized manifold is revealed by connecting each data point and its $k = 6$ nearest neighbors ②. An unsupervised learning algorithm unfolds the Swiss roll while preserving the local geometry of nearby data points ③. Finally, the data points are projected onto the two dimensional subspace that maximizes their variance, yielding a faithful embedding of the original manifold ④.*

bases (biorthogonal expansion). 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* [44] can be explained by *generalized inequalities* with respect to convex cones and their duals. The conic analogue to linear independence, called *conic independence*, is introduced as a new tool in the study of cones; the logical next step in the progression: linear, affine, conic.

Any convex optimization problem can be visualized geometrically. Desire to visualize in high dimension is deeply embedded in the mathematical psyche. 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. 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 2.9.2.6.3) There is a closed form for projection on those convex subsets.

- The positive semidefinite cone is a circular cone in low dimension, while *Geršgorin discs* specify inscription of a polyhedral cone into that positive semidefinite cone. (Figure **36**)

**Chapter 3**, **Geometry of convex functions**, observes analogies 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 cookbook for the most useful of convex functions and optimization problems, methods are drawn from matrix calculus for determining convexity and discerning geometry.

**Semidefinite programming** is reviewed in **chapter 4** with particular attention to optimality conditions of prototypical primal and dual conic programs, their interplay, and the *perturbation method* of rank reduction of optimal solutions (extant but not well-known). Positive definite Farkas'


Figure 6: About five thousand points along the borders constituting the United States were used to create an exhaustive matrix of interpoint distance for each and every pair of points in the ordered set (a list); called a *Euclidean distance matrix.* From that noiseless distance information, it is easy to reconstruct the map exactly via the Schoenberg criterion (728). (§5.13.1.0.1, *confer* Figure **91**) 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 (just as it is for all conventional map preparation).

lemma is derived, and we also show how to determine if a feasible set belongs exclusively to a positive semidefinite cone boundary. A three-dimensional polyhedral analogue for the positive semidefinite cone of $3\times 3$ symmetric matrices is introduced. This analogue is a new tool for visualizing coexistence of low- and high-rank optimal solutions in 6 dimensions. We find a minimum-cardinality Boolean solution to an instance of $Ax=b$:

$$\begin{array}{cl}\underset{x}{\text{minimize}} & \|x\|_0\\ \text{subject to} & Ax=b\\ & x_i\in\{0,1\}\ ,\quad i=1\ldots n\end{array}\tag{576}$$

The *sensor-network localization* problem is solved in any dimension in this chapter. We introduce the method of *convex iteration* for constraining rank in the form $\operatorname{rank} G\leq\rho$ for some semidefinite programs in $G$.

Figure 7: These bees construct a honeycomb by solving a convex optimization problem. (§5.4.2.2.3) The most dense packing of identical spheres about a central sphere in two dimensions is 6. Packed sphere centers describe a regular lattice.

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 the Euclidean metric; thereby, observing existence of an infinity of properties of the Euclidean metric beyond the 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 infinitum*) 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 **6**. We also generate a distorted but recognizable isotonic map using only comparative distance information (only ordinal distance data). We demonstrate an elegant method for including dihedral (or *torsion*) angle constraints into a molecular conformation problem. More geometrical problems solvable via EDMs are presented with the best methods for posing them, EDM problems are posed as convex optimizations, and we show how to recover exact relative position given incomplete noiseless interpoint distance information.

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

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{ \begin{array}{c} -V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0 \\ D \in \mathbb{S}_h^N \end{array} \right. \tag{728}$$

Our proof relies on fundamental geometry; assuming, any EDM must correspond to a list of points contained in some polyhedron (possibly at

Figure 8: Robotic vehicles in concert can move larger objects, guard a perimeter, perform surveillance, find land mines, or localize a plume of gas, liquid, or radio waves. [94]

its vertices) and *vice versa.* It is known, but not obvious, this *Schoenberg criterion* implies nonnegativity of the EDM entries; proved here.

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 $\mathbb{EDM}^N$;

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{\begin{array}{l} \lambda\left(\left[\begin{array}{cc} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{array}\right]\right) \in \left[\begin{array}{c} \mathbb{R}_+^N \\ \mathbb{R}_- \end{array}\right] \cap \partial\mathcal{H} \\ D \in \mathbb{S}_h^N \end{array}\right. \tag{932}$$

We will see: spectral cones are not unique.

In **chapter 6**, **EDM cone**, we explain the 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 (728), relating the EDM cone and a positive semidefinite cone, is revealed to be a discretized membership relation (dual generalized inequalities, a new Farkas'-like lemma) between the EDM cone and its ordinary dual, $\mathbb{EDM}^{N^*}$. A matrix criterion for membership to the dual EDM cone is derived that is simpler than the Schoenberg criterion:

$$D^* \in \mathbb{EDM}^{N^*} \Leftrightarrow \delta(D^*\mathbf{1}) - D^* \succeq 0 \qquad (1080)$$

We derive a new concise equality of the EDM cone to two subspaces and a positive semidefinite cone;

$$\mathbb{EDM}^N = \mathbb{S}_h^N \cap \left(\mathbb{S}_c^{N\perp} - \mathbb{S}_+^N\right) \qquad (1074)$$

In **chapter 7**, **Proximity problems**, we 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 $H$:

$$\begin{array}{cll}\underset{D}{\text{minimize}} & \|-V(D-H)V\|_{\mathrm F}^2 \\ \text{subject to} & \operatorname{rank} VDV \leq \rho \\ & D \in \mathbb{EDM}^N\end{array} \qquad \begin{array}{cll}\underset{\sqrt[\circ]{D}}{\text{minimize}} & \|\sqrt[\circ]{D}-H\|_{\mathrm F}^2 \\ \text{subject to} & \operatorname{rank} VDV \leq \rho \\ & \sqrt[\circ]{D} \in \sqrt{\mathbb{EDM}^N}\end{array}$$

$$\begin{array}{cll}\underset{D}{\text{minimize}} & \|D-H\|_{\mathrm F}^2 \\ \text{subject to} & \operatorname{rank} VDV \leq \rho \\ & D \in \mathbb{EDM}^N\end{array} \qquad \begin{array}{cll}\underset{\sqrt[\circ]{D}}{\text{minimize}} & \|-V(\sqrt[\circ]{D}-H)V\|_{\mathrm F}^2 \\ \text{subject to} & \operatorname{rank} VDV \leq \rho \\ & \sqrt[\circ]{D} \in \sqrt{\mathbb{EDM}^N}\end{array} \qquad (1116)$$

We apply the new convex iteration method for constraining rank. Known heuristics for solving the problems when compounded with rank minimization are also explained. We offer a new geometrical proof of a famous result discovered by Eckart & Young in 1936 [85] regarding Euclidean projection of a point on that generally nonconvex subset of the positive semidefinite cone boundary comprising all positive semidefinite matrices having rank not exceeding a prescribed bound $\rho$. We explain how this problem is transformed to a convex optimization for any rank $\rho$.

## novelty

p.120 *Conic independence* is introduced as a natural extension to linear and affine independence; a new tool in convex analysis most useful for manipulation of cones.

p.148 Arcane theorems of alternative *generalized inequality* are simply derived from cone *membership relations*; generalizations of *Farkas' lemma* translated to the geometry of convex cones.

p.229 We present an arguably good 3-dimensional polyhedral analogue, to the isomorphically 6-dimensional positive semidefinite cone, as an aid to understand semidefinite programming.

p.256 We show how to constrain rank in the form $\operatorname{rank} G \leq \rho$ for some semidefinite programs.

p.307, p.311 *Kissing-number of sphere packing* (first solved by Isaac Newton) and *trilateration* or *localization* are shown to be convex optimization problems.

p.322 We show how to elegantly include *torsion* or *dihedral* angle constraints into the *molecular conformation problem.*

p.353 Geometrical proof: *Schoenberg criterion* for a Euclidean distance matrix.

p.371 We experimentally demonstrate a conjecture of Borg & Groenen by reconstructing a map of the USA using only ordinal (comparative) distance information.

p.6, p.389 There is an 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.

p.433 The *Schoenberg criterion* for a Euclidean distance matrix is revealed to be a discretized *membership relation* (or *dual generalized inequalities*) between the EDM cone and its dual.

## appendices

Provided 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**),
- a 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**).

# 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 [20, p.vii]

There is relatively less published pertaining to matrix-valued convex sets and functions. [158] [151, §6.6] [218] As convex geometry and linear algebra are inextricably bonded, we provide much background material on linear algebra (especially in the appendices) although a reader is assumed comfortable with [249], [251], [150], or any other intermediate-level text. The essential references to convex analysis are [148] [230]. The reader is referred to [247] [20] [280] [30] [46] [227] [268] for a comprehensive treatment of convexity.



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

## 2.1 Convex set

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

$$\mu Y + (1-\mu)Z \in \mathcal{C} \tag{1}$$

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 Euclidean real vector space $\mathbb{R}^n$ or $\mathbb{R}^{m\times n}$ (matrices), then (1) represents the closed line segment joining them. All line segments are thereby convex sets. Apparent from the definition, a convex set is a connected set. [189, §3.4, §3.5] [30, p.2]

An *ellipsoid* centered at $x=a$ (Figure **10**, p.38), given matrix $C \in \mathbb{R}^{m\times n}$

$$\{x \in \mathbb{R}^n \mid \|C(x-a)\|^2 = (x-a)^T C^T C(x-a) \le 1\} \tag{2}$$

is a good icon for a convex set.

### 2.1.1 subspace

A nonempty subset of Euclidean real vector space $\mathbb{R}^n$ is called a *subspace* (formally defined in §2.5) if every vector[2.1] of the form $\alpha x + \beta y$, for $\alpha, \beta \in \mathbb{R}$, is in the subset whenever vectors $x$ and $y$ are. [182, §2.3] A subspace is a convex set containing the *origin* $\mathbf{0}$, by definition. [230, p.4] Any subspace is therefore open in the sense that it contains no boundary, but closed in the sense [189, §2]

$$\mathbb{R}^n + \mathbb{R}^n = \mathbb{R}^n \tag{3}$$

It is not difficult to show

$$\mathbb{R}^n = -\mathbb{R}^n \tag{4}$$

as is true for any subspace $\mathcal{R}$, because $x \in \mathbb{R}^n \Leftrightarrow -x \in \mathbb{R}^n$.

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.*,[2.2] any line through the origin in two-dimensional Euclidean space $\mathbb{R}^2$. The vector space $\mathbb{R}^n$ is itself a conventional subspace, inclusively, [166, §2.1] although not proper.

[2.1] A *vector* is assumed, throughout, to be a column vector.
[2.2] We substitute the abbreviation *e.g.* in place of the Latin *exempli gratia*.

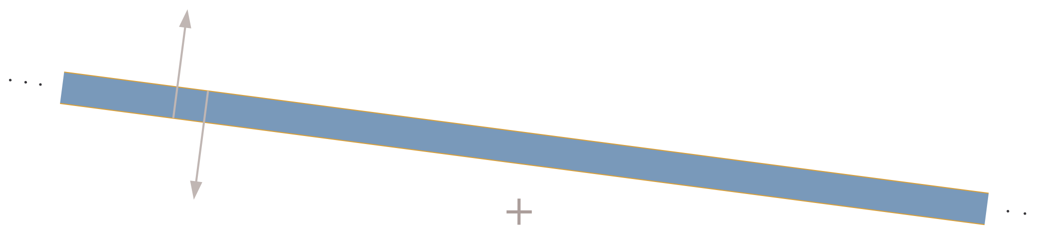


Figure 9: A *slab* is a convex Euclidean body infinite in extent but not affine. Illustrated in $\mathbb{R}^2$, it may be constructed by intersecting two opposing halfspaces whose bounding hyperplanes are parallel but not coincident. Because number of halfspaces used in its construction is finite, slab is a polyhedron. (Cartesian axes and vector inward-normal to each halfspace-boundary are drawn for reference.)

### 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 $\zeta \in \mathbb{R}^N$

$$\Gamma_1\,\zeta_1 + \cdots + \Gamma_{N-1}\,\zeta_{N-1} + \Gamma_N\,\zeta_N = \mathbf{0} \tag{5}$$

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

#### 2.1.2.1 preservation

Linear independence can be preserved under linear transformation. Given matrix $Y \triangleq [\,y_1 \;\; y_2 \;\cdots\; y_N\,] \in \mathbb{R}^{N\times N}$, consider the mapping

$$T(\Gamma) : \mathbb{R}^{n\times N} \rightarrow \mathbb{R}^{n\times N} \;\triangleq\; \Gamma\, Y \tag{6}$$

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 exists no nontrivial solution $\zeta \in \mathbb{R}^N$ to

$$\Gamma y_1\,\zeta_i + \cdots + \Gamma y_{N-1}\,\zeta_{N-1} + \Gamma y_N\,\zeta_N = \mathbf{0} \tag{7}$$

By factoring $\Gamma$, we see that 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 $\mathbb{R}^2$. 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.*,

$$\mathbb{R}^n_+ \triangleq \{x \in \mathbb{R}^n \mid x_i \geq 0 \;\; \forall\, i\} \tag{8}$$

The nonpositive orthant $\mathbb{R}^n_-$ or $\mathbb{R}^{n\times n}_-$ (analogue to quadrant III) denotes negative and 0 entries. Orthant convexity[2.3] 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. An affine set is convex and open so contains no boundary: *e.g.*, $\emptyset$, point, line, plane, *hyperplane* (§2.4.2), subspace, *etcetera.* For some *parallel*[2.4] subspace $\mathcal{M}$ and any point $x \in \mathcal{A}$

$$\begin{aligned} \mathcal{A} \text{ is affine } \;\Leftrightarrow\;\; \mathcal{A} &= x + \mathcal{M} \\ &= \{y \mid y - x \in \mathcal{M}\} \end{aligned} \tag{9}$$

The intersection of an arbitrary collection of affine sets remains affine. The *affine hull* of a set $\mathcal{C} \subseteq \mathbb{R}^n$ (§2.3.1) is the smallest affine set containing it.

### 2.1.5 dimension

*Dimension* of an arbitrary set $\mathcal{S}$ is the dimension of its affine hull; [280, p.14]

$$\dim \mathcal{S} \triangleq \dim \operatorname{aff} \mathcal{S} \;=\; \dim \operatorname{aff}(\mathcal{S} - s)\;, \quad s \in \mathcal{S} \tag{10}$$

the same as dimension of the subspace parallel to that affine set $\operatorname{aff} \mathcal{S}$ when nonempty. Hence dimension (of a set) is synonymous with *affine dimension.* [148, A.2.1]

---

[2.3] All orthants are self-dual simplicial cones. (§2.13.5.1, §2.12.3.1.1)

[2.4] Two affine sets are said to be parallel when one is a translation of the other. [230, p.4]

### 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. Thus the term *relative* is the conventional fix to this ambiguous terminology:[2.5]

- An ordinary flat sheet of paper is an example of a nonempty convex set in $\mathbb{R}^3$ having empty interior but relatively nonempty interior.

#### 2.1.6.1 relative interior

We distinguish interior from *relative interior* throughout. [247] [280] [268] The relative interior $\operatorname{rel\,int}\mathcal{C}$ of a convex set $\mathcal{C}\subseteq\mathbb{R}^n$ is its interior relative to its affine hull.[2.6] Thus defined, it is common (though confusing) for $\operatorname{int}\mathcal{C}$ 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 ($\dim\operatorname{aff}\mathcal{C} < n$ , *e.g.*, were $\mathcal{C}$ a flat piece of paper in $\mathbb{R}^3$) or in the exception when $\mathcal{C}$ is a single point; [189, §2.2.1]

$$\operatorname{rel\,int}\{x\} \triangleq \operatorname{aff}\{x\} = \{x\}\ , \qquad \operatorname{int}\{x\} = \emptyset\ , \qquad x\in\mathbb{R}^n \tag{11}$$

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

$$\overline{\operatorname{rel\,int}\mathcal{C}} = \overline{\mathcal{C}} \tag{12}$$

If $\mathcal{C}$ is convex then $\operatorname{rel\,int}\mathcal{C}$ and $\overline{\mathcal{C}}$ are convex, [148, p.24] and it is always possible to pass to a smaller ambient Euclidean space where a nonempty set acquires an interior. [20, §II.2.3].

Given the intersection of convex set $\mathcal{C}$ with an affine set $\mathcal{A}$

$$\operatorname{rel\,int}(\mathcal{C}\cap\mathcal{A}) = \operatorname{rel\,int}(\mathcal{C})\cap\mathcal{A} \tag{13}$$

If $\mathcal{C}$ has nonempty interior, then $\operatorname{rel\,int}\mathcal{C} = \operatorname{int}\mathcal{C}$ .

---

[2.5] 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.

[2.6] Likewise for *relative boundary* (§2.6.1.3), although *relative closure* is superfluous. [148, §A.2.1]

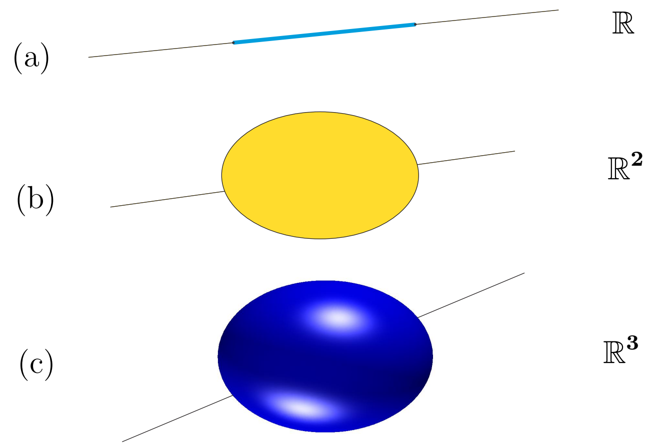



Figure 10: **(a)** Ellipsoid in $\mathbb{R}$ is a line segment whose boundary comprises two points. Intersection of line with ellipsoid in $\mathbb{R}$ , **(b)** in $\mathbb{R}^2$, **(c)** in $\mathbb{R}^3$. Each ellipsoid illustrated has entire boundary constituted by zero-dimensional faces; in fact, 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 classical boundary

(*confer* §2.6.1.3) *Boundary* of a set $\mathcal{C}$ is the closure of $\mathcal{C}$ less its interior presumed nonempty; [41, §1.1]

$$\partial\mathcal{C} = \overline{\mathcal{C}} \setminus \operatorname{int}\mathcal{C} \tag{14}$$

which follows from the fact

$$\overline{\operatorname{int}\mathcal{C}} = \overline{\mathcal{C}} \tag{15}$$

assuming nonempty interior.[2.7] One implication is: an open set has a boundary defined although not contained in the set.

[2.7]Otherwise, for $x \in \mathbb{R}^n$ as in (11), [189, §2.1, §2.3]

$$\overline{\operatorname{int}\{x\}} = \overline{\emptyset} = \emptyset$$

the empty set is both open and closed.

#### 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. [20, §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 **10** 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 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. For example: The ellipsoids in Figure **10** have boundaries composed only of zero-dimensional faces. The two-dimensional slab in Figure **9** is a *polyhedron* having one-dimensional faces making its boundary. The three-dimensional bounded polyhedron in Figure **12** has zero-, one-, and two-dimensional polygonal faces constituting its boundary.

**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). The 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 is known in closed form (§7.1.2). Given, for example, $\lambda\in\operatorname{int}\mathbb{R}^3_+$ and *diagonalization* (§A.5.2) of exterior point

$$X = Q\Lambda Q^T\in\mathbb{S}^3,\quad \Lambda\triangleq\begin{bmatrix}\lambda_1 & & \mathbf{0}\\ & \lambda_2 & \\ \mathbf{0} & & -\lambda_3\end{bmatrix}\tag{16}$$

where $Q \in \mathbb{R}^{\mathbf{3}\times\mathbf{3}}$ is an orthogonal matrix, then the projection on $\mathbb{S}_+^{\mathbf{3}}$ in $\mathbb{R}^{\mathbf{6}}$ is

$$PX = Q\begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \lambda_2 & \\ \mathbf{0} & & 0 \end{bmatrix} Q^T \in \mathbb{S}_+^{\mathbf{3}} \tag{17}$$

This positive semidefinite matrix $PX$ nearest $X$ thus has rank 2, found by discarding all negative eigenvalues. 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* $\mathbb{S}_+^{\mathbf{3}}$ 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,

$$X + (PX - X)t\big|_{t=1+\varepsilon} = Q(\Lambda + (P\Lambda - \Lambda)(1+\varepsilon))Q^T = Q\begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \lambda_2 & \\ \mathbf{0} & & \varepsilon\lambda_3 \end{bmatrix} Q^T \in \operatorname{int} \mathbb{S}_+^{\mathbf{3}} \tag{18}$$

$\square$

#### 2.1.7.2 Tangential line intersection with boundary

A higher-dimensional boundary $\partial\mathcal{C}$ of a convex Euclidean body $\mathcal{C}$ 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 $\mathcal{C}$ [161, §15.3]) through a single point relatively interior to a three-dimensional face on $\partial\mathcal{C}$. Let's understand why by inductive reasoning.

Figure **11**(a) 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 **11**(b) 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).

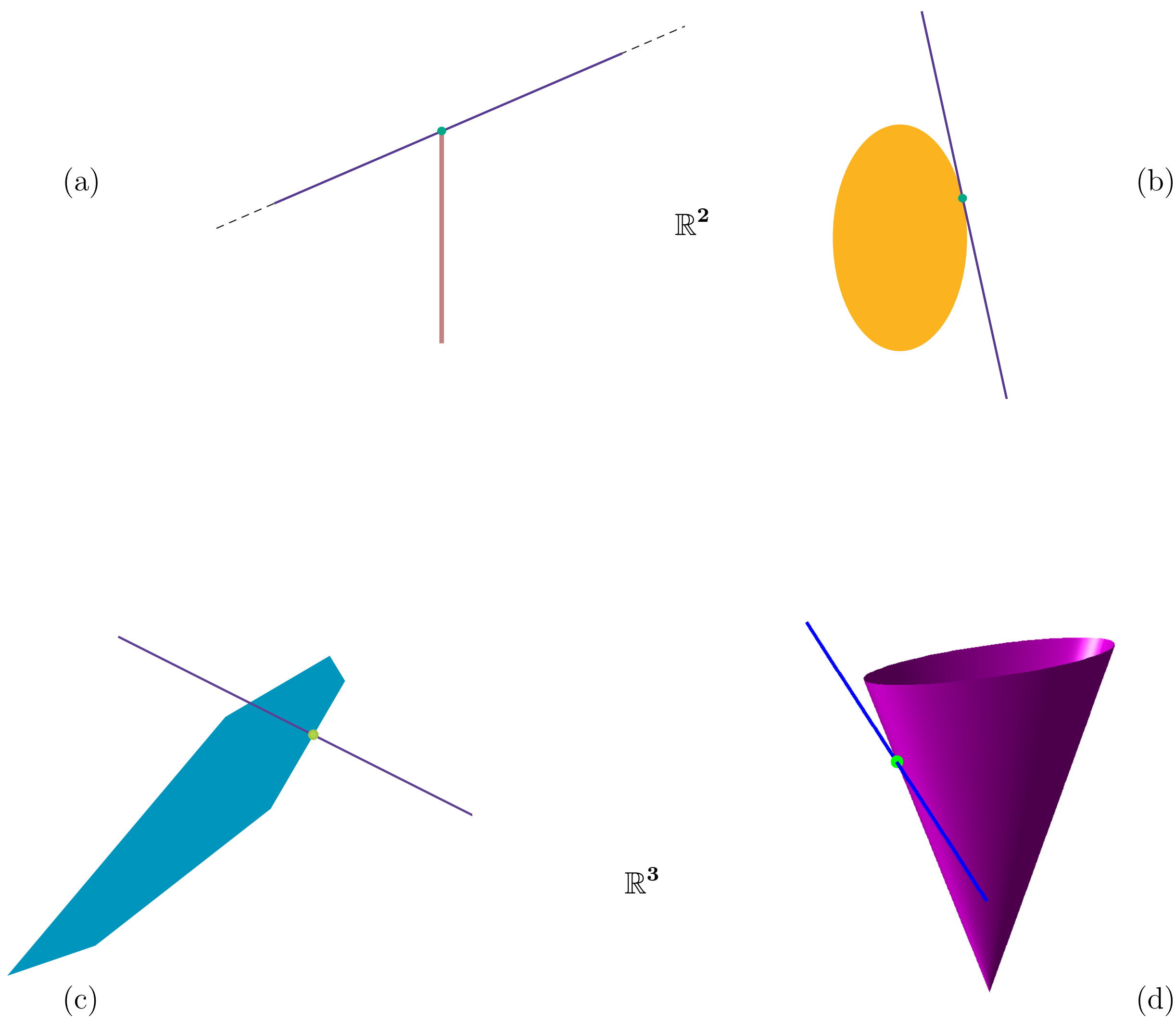



Figure 11: Line tangential **(a)** **(b)** to relative interior of a zero-dimensional face in $\mathbb{R}^2$, **(c)** **(d)** to relative interior of a one-dimensional face in $\mathbb{R}^3$.

Now let's move to an ambient space of three dimensions. Figure **11**(c) 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 **11**(d) illustrates a solid circular pyramid (upside-down) whose one-dimensional faces are line-segments emanating from its pointed end (its *vertex*). This pyramid's boundary is constituted solely by these one-dimensional line-segments. 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 pyramid.

From these few examples, way 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 **12** 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 $\operatorname{int}\mathcal{P}$) 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 $\operatorname{rel}\operatorname{int}\mathcal{P}$ as claimed, requires a line existing in at least five dimensions.

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 (§5.4.2.2.4).

### 2.1.8 intersection, sum, difference, product

**2.1.8.0.1 Theorem.** *Intersection.* [46, §2.3.1] [230, §2]
The intersection of an arbitrary collection of convex sets is convex. ⋄

This theorem in converse is implicitly false in so far as a convex set can be formed by the intersection of sets that are not.

*Vector sum* of two convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$

$$\mathcal{C}_1+\mathcal{C}_2 = \{x+y \mid x \in \mathcal{C}_1\ ,\ y \in \mathcal{C}_2\} \tag{19}$$

is convex.

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

$$\mathcal{C}_1-\mathcal{C}_2 = \{x-y \mid x \in \mathcal{C}_1\ ,\ y \in \mathcal{C}_2\} \tag{20}$$

which is convex. Applying this definition to nonempty convex set $\mathcal{C}_1$ , its self-difference $\mathcal{C}_1-\mathcal{C}_1$ is generally nonempty, nontrivial, and convex; *e.g.*, for any *convex cone* $\mathcal{K}$ , (§2.7.2) the set $\mathcal{K}-\mathcal{K}$ constitutes its affine hull. [230, p.15]

*Cartesian product* of convex sets

$$\mathcal{C}_1\times\mathcal{C}_2 = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \mid x \in \mathcal{C}_1\ ,\ \ y \in \mathcal{C}_2\right\} = \begin{bmatrix} \mathcal{C}_1 \\ \mathcal{C}_2 \end{bmatrix} \tag{21}$$

remains convex. The converse also holds; *id est*, a Cartesian product is convex iff each set is. [148, p.23]

Convex results are also obtained for scaling $\kappa\mathcal{C}$ of a convex set $\mathcal{C}$ , rotation/reflection $Q\mathcal{C}$ , or translation $\mathcal{C}+\alpha$ ; all similarly defined.

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

$$T(\mathcal{C}) \triangleq \{T(x) \mid x\in\mathcal{C}\} \tag{22}$$

Given linear operator $T$ , it therefore follows from (19),

$$\begin{aligned} T(\mathcal{C}_1+\mathcal{C}_2) &= \{T(x+y) \mid x\in\mathcal{C}_1\ ,\ y\in\mathcal{C}_2\} \\ &= \{T(x)+\ T(y) \mid x\in\mathcal{C}_1\ ,\ y\in\mathcal{C}_2\} \\ &= T(\mathcal{C}_1)+\ T(\mathcal{C}_2) \end{aligned} \tag{23}$$

### 2.1.9 inverse image

While *epigraph* and *sublevel sets* (§3.1.7) of a convex function must be convex, it generally holds that image and inverse image of a convex function are not. Although there are many examples to the contrary, the most prominent are the affine functions:

**2.1.9.0.1 Theorem.** *Image, Inverse image.* [230, §3] [46, §2.3.2]
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 (§3.1.6)
$$f(\mathcal{C}) = \{f(X) \mid X\in\mathcal{C}\} \subseteq \mathbb{R}^{m\times n} \tag{24}$$
is convex.

- The inverse image[2.8] of a convex set $\mathcal{F}$,
$$f^{-1}(\mathcal{F}) = \{X \mid f(X)\in\mathcal{F}\} \subseteq \mathbb{R}^{p\times k} \tag{25}$$
a single or many-valued mapping, under any affine function $f$ is convex.

⋄

In particular, any affine transformation of an affine set remains affine. [230, p.8] Ellipsoids are invariant to any [*sic*] affine transformation.

Each converse of this two-part theorem is generally false; *id est*, given $f$ affine, a convex image $f(\mathcal{C})$ does not imply that set $\mathcal{C}$ is convex, and neither does a convex inverse image $f^{-1}(\mathcal{F})$ imply set $\mathcal{F}$ is convex. A counter-example is easy to visualize when the affine function is an orthogonal projector [249] [182]:

**2.1.9.0.2 Corollary.** *Projection on subspace.* [230, §3][2.9]
Orthogonal projection of a convex set on a subspace 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.

---

[2.8] See §2.9.1.0.2 for an example.

[2.9] The corollary holds more generally for projection on hyperplanes (§2.4.2); [280, §6.6] hence, for projection on affine subsets (§2.3.1, nonempty intersections of hyperplanes). Orthogonal projection on affine subsets is reviewed in §E.4.0.0.1.

## 2.2 Vectorized-matrix inner product

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

$$\langle y\,,z\rangle \triangleq y^Tz \tag{26}$$

We prefer those angle brackets to connote a geometric rather than algebraic perspective. Two vectors are *orthogonal* (*perpendicular*) to one another if and only if their inner product vanishes;

$$y\perp z \;\Leftrightarrow\; \langle y\,,z\rangle = 0 \tag{27}$$

A vector inner-product defines a *norm*

$$\|y\|_2 \triangleq \sqrt{y^Ty}\;, \qquad \|y\|_2 = 0 \;\Leftrightarrow\; y = 0 \tag{28}$$

When orthogonal vectors each have unit norm, then they are *orthonormal.* For linear operation $A$ on a vector, represented by a real matrix, the *adjoint operation* $A^T$ is transposition and defined for matrix $A$ by [166, §3.10]

$$\langle y\,,A^Tz\rangle \triangleq \langle Ay\,,z\rangle \tag{29}$$

The 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 [166, App.A1.2]) transformation *vectorization.* For example, the vectorization of $Y=[\,y_1\;y_2\cdots\,y_k\,]\in\mathbb{R}^{p\times k}$ [116] [245] is

$$\operatorname{vec}Y \triangleq \begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_k\end{bmatrix}\in\mathbb{R}^{pk} \tag{30}$$

Then the vectorized-matrix inner-product is trace of matrix inner-product; for $Z\in\mathbb{R}^{p\times k}$, [46, §2.6.1] [148, §0.3.1] [290, §8] [275, §2.2]

$$\langle Y\,,\,Z\rangle \;\triangleq\; \operatorname{tr}(Y^TZ) \;=\; \operatorname{vec}(Y)^T\operatorname{vec}Z \tag{31}$$

where (§A.1.1)

$$\operatorname{tr}(Y^TZ) = \operatorname{tr}(ZY^T) = \operatorname{tr}(YZ^T) = \operatorname{tr}(Z^TY) = \mathbf{1}^T(Y\circ Z)\mathbf{1} \tag{32}$$

and where $\circ$ denotes the Hadamard product[2.10] of matrices [150] [110, §1.1.4]. The adjoint operation $A^T$ on a matrix can therefore be defined in like manner:

$$\langle Y\,,\,A^TZ\rangle \triangleq \langle AY\,,\,Z\rangle \tag{33}$$

For example, take any element $\mathcal{C}_1$ from a matrix-valued set in $\mathbb{R}^{p\times k}$, and consider any particular dimensionally compatible real vectors $v$ and $w$. Then vector inner-product of $\mathcal{C}_1$ with $vw^T$ is

$$\langle vw^T,\,\mathcal{C}_1\rangle \;=\; v^T\mathcal{C}_1 w \;=\; \operatorname{tr}(wv^T\mathcal{C}_1) \;=\; \mathbf{1}^T\big((vw^T)\circ\mathcal{C}_1\big)\,\mathbf{1} \tag{34}$$

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

$$\mathcal{Y} \triangleq v^T\mathcal{C}w \;=\; \langle vw^T,\,\mathcal{C}\rangle \subseteq \mathbb{R} \tag{35}$$

is convex. This result is a consequence of the *image theorem.* Yet it is easy to show directly that convex combination of elements from $\mathcal{Y}$ remains an element of $\mathcal{Y}$.[2.11] □

More generally, $vw^T$ in (35) may be replaced with any particular matrix $Z\in\mathbb{R}^{p\times k}$ while convexity of the 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 $U^T\mathcal{C}W$ is convex by the *image theorem* because it is a linear mapping of $\mathcal{C}$.

---

[2.10] The Hadamard product is a simple entrywise product of corresponding entries from two matrices of like size; *id est*, not necessarily square.

[2.11] 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 (35) that are two elements of $\mathcal{Y}$ by definition. Now make a convex combination of those inner products; *videlicet*, for $0\leq\mu\leq1$

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

The two sides are equivalent by linearity of inner product. The right-hand side remains a vector inner-product of $vw^T$ 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. ♦

## 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. The image and inverse image under the transformation operator are then called isomorphic vector spaces. △

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 also be connected by a line segment. Two Euclidean bodies may be considered isomorphic of there exists an isomorphism of their corresponding ambient spaces. [276, §I.1]

When $Z = Y \in \mathbb{R}^{p\times k}$ in (31), *Frobenius' norm* is resultant from vector inner-product; (*confer* (1467))

$$\begin{aligned}\|Y\|_{\mathrm F}^2 \;&=\; \|\operatorname{vec} Y\|_2^2 \;=\; \langle Y\,,\,Y\rangle \;=\; \operatorname{tr}(Y^TY)\\ &=\; \sum_{i,j} Y_{ij}^2 \;=\; \sum_i \lambda(Y^TY)_i \;=\; \sum_i \sigma(Y)_i^2\end{aligned} \tag{36}$$

where $\lambda(Y^TY)_i$ is the $i^{\text{th}}$ eigenvalue of $Y^TY$, and $\sigma(Y)_i$ the $i^{\text{th}}$ singular value of $Y$. Were $Y$ a *normal matrix* (§A.5.2), then $\sigma(Y) = |\lambda(Y)|$ [301, §8.1] thus

$$\|Y\|_{\mathrm F}^2 \;=\; \sum_i \lambda(Y)_i^2 \;=\; \|\lambda(Y)\|_2^2 \tag{37}$$

The converse (37) $\Rightarrow$ normal matrix $Y$ also holds. [150, §2.5.4]

Because the metrics are equivalent

$$\|\operatorname{vec} X - \operatorname{vec} Y\|_2 = \|X - Y\|_{\mathrm F} \tag{38}$$

and because vectorization (30) is a linear bijective map, then vector space $\mathbb{R}^{p\times k}$ is isometrically isomorphic with vector space $\mathbb{R}^{pk}$ in the Euclidean sense and vec is an isometric isomorphism on $\mathbb{R}^{p\times k}$.[2.12] Because of this Euclidean structure, all the known results from convex analysis in Euclidean space $\mathbb{R}^n$ carry over directly to the space of real matrices $\mathbb{R}^{p\times k}$.

[2.12] Given matrix $A$, its range $\mathcal{R}(A)$ (§2.5) is isometrically isomorphic with its vectorized range $\operatorname{vec}\mathcal{R}(A)$ but not with $\mathcal{R}(\operatorname{vec} A)$.

**2.2.1.0.2 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 $x, y \in \operatorname{dom} T$

$$\|Tx - Ty\| = \|x - y\| \tag{39}$$

Then the isometric isomorphism $T$ is a *bijective isometry.* △

*Unitary linear operator* $Q : \mathbb{R}^n \to \mathbb{R}^n$ representing orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ (§B.5), for example, is an isometric isomorphism. Yet isometric operator $T : \mathbb{R}^2 \to \mathbb{R}^3$ representing $T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ on $\mathbb{R}^2$ is *injective* but not a *surjective* map [166, §1.6] to $\mathbb{R}^3$.

The Frobenius norm is *orthogonally invariant*; meaning, for $X, Y \in \mathbb{R}^{p\times k}$ and dimensionally compatible *orthonormal matrix*[2.13] $U$ and orthogonal matrix $Q$

$$\|U(X-Y)Q\|_{\mathrm F} = \|X-Y\|_{\mathrm F} \tag{40}$$

### 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;

$$\mathbb{S}^M \triangleq \left\{A \in \mathbb{R}^{M\times M} \mid A = A^T\right\} \subseteq \mathbb{R}^{M\times M} \tag{41}$$

This subspace comprising symmetric matrices $\mathbb{S}^M$ 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* [249] [182] of $\mathbb{S}^M$ is

$$\mathbb{S}^{M\perp} \triangleq \left\{A \in \mathbb{R}^{M\times M} \mid A = -A^T\right\} \subset \mathbb{R}^{M\times M} \tag{42}$$

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

$$\mathbb{S}^M \oplus \mathbb{S}^{M\perp} = \mathbb{R}^{M\times M} \tag{43}$$

where unique vector sum $\oplus$ is defined on page 676. △

[2.13] Any matrix $U$ whose columns are orthonormal with respect to each other ($U^T U = I$); these include the orthogonal matrices.

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

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T) \tag{44}$$

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

$$\operatorname{tr}\big((A^T + A)(A - A^T)\big) = 0 \tag{45}$$

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

$$\mathbb{S}^{M\perp} \triangleq \left\{ \frac{1}{2}(A - A^T) \mid A\in\mathbb{R}^{M\times M} \right\} \subset \mathbb{R}^{M\times M} \tag{46}$$

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

#### 2.2.2.1 Isomorphism on symmetric matrix subspace

When a matrix is symmetric in $\mathbb{S}^M$, we may still employ the vectorization transformation (30) 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 $Y = [Y_{ij}] \in \mathbb{S}^M$ we introduce the symmetric vectorization

$$\operatorname{svec} Y \triangleq \begin{bmatrix} Y_{11} \\ \sqrt{2}Y_{12} \\ Y_{22} \\ \sqrt{2}Y_{13} \\ \sqrt{2}Y_{23} \\ Y_{33} \\ \vdots \\ Y_{MM} \end{bmatrix} \in \mathbb{R}^{M(M+1)/2} \tag{47}$$

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

$$\langle Y\,,\,Z\rangle \;\triangleq\; \operatorname{tr}(Y^TZ) \;=\; \operatorname{vec}(Y)^T\operatorname{vec} Z \;=\; \operatorname{svec}(Y)^T\operatorname{svec} Z \tag{48}$$

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

$$\|\operatorname{svec} X - \operatorname{svec} Y\|_2 = \|X - Y\|_{\mathrm{F}} \tag{49}$$

and because symmetric vectorization (47) is a linear bijective mapping, then svec is an isometric isomorphism on the symmetric matrix subspace. In other words, $\mathbb{S}^M$ is isometrically isomorphic with $\mathbb{R}^{M(M+1)/2}$ in the Euclidean sense under transformation svec .

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

$$\{E_{ij} \in \mathbb{S}^M\} = \left\{\begin{array}{ll} e_i e_i^T\,, & i = j = 1\ldots M \\ \frac{1}{\sqrt{2}}\Big(e_i e_j^T + e_j e_i^T\Big)\,, & 1 \le i < j \le M \end{array}\right\} \tag{50}$$

where $M(M+1)/2$ *standard basis matrices* $E_{ij}$ are formed from the *standard basis vectors* $e_i \in \mathbb{R}^M$. Thus we have a basic *orthogonal expansion* for $Y \in \mathbb{S}^M$

$$Y = \sum_{j=1}^{M}\sum_{i=1}^{j}\langle E_{ij}\,, Y\rangle\, E_{ij} \tag{51}$$

whose coefficients

$$\langle E_{ij}\,, Y\rangle \;\;=\;\; \begin{cases} Y_{ii}\,, & i = 1\ldots M \\ Y_{ij}\sqrt{2}\,, & 1 \le i < j \le M \end{cases} \tag{52}$$

correspond to entries of the symmetric vectorization (47).

### 2.2.3 Symmetric hollow subspace

**2.2.3.0.1 Definition.** *Hollow subspaces.* [266]
Define a subspace of $\mathbb{R}^{M\times M}$: the convex set of all (real) symmetric $M\times M$ matrices having $\mathbf{0}$ main diagonal;

$$\mathbb{R}_h^{M\times M} \triangleq \left\{A\in\mathbb{R}^{M\times M} \mid A=A^T,\;\; \delta(A)=\mathbf{0}\right\} \subset \mathbb{R}^{M\times M} \tag{53}$$

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

$$\delta(A) \in \mathbb{R}^M \tag{1220}$$

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 $\Lambda\in\mathbb{S}^N$, operator $\delta(\delta(\Lambda))$ returns $\Lambda$ itself;

$$\delta^2(\Lambda) \equiv \delta(\delta(\Lambda)) \triangleq \Lambda \tag{1222}$$

The subspace $\mathbb{R}_h^{M\times M}$ (53) comprising (real) symmetric hollow matrices is isomorphic with subspace $\mathbb{R}^{M(M-1)/2}$. The orthogonal complement of $\mathbb{R}_h^{M\times M}$ is

$$\mathbb{R}_h^{M\times M\perp} \triangleq \left\{A\in\mathbb{R}^{M\times M} \mid A=-A^T+2\delta^2(A)\right\} \subseteq \mathbb{R}^{M\times M} \tag{54}$$

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

$$\mathbb{R}_h^{M\times M} \oplus \mathbb{R}_h^{M\times M\perp} = \mathbb{R}^{M\times M} \tag{55}$$

Yet defined instead as a proper subspace of $\mathbb{S}^M$

$$\mathbb{S}_h^M \triangleq \left\{A\in\mathbb{S}^M \mid \delta(A) = \mathbf{0}\right\} \subset \mathbb{S}^M \tag{56}$$

the orthogonal complement $\mathbb{S}_h^{M\perp}$ of $\mathbb{S}_h^M$ in ambient $\mathbb{S}^M$

$$\mathbb{S}_h^{M\perp} \triangleq \left\{A\in\mathbb{S}^M \mid A=\delta^2(A)\right\} \subseteq \mathbb{S}^M \tag{57}$$

is simply the subspace of diagonal matrices; *id est*,

$$\mathbb{S}_h^M \oplus \mathbb{S}_h^{M\perp} = \mathbb{S}^M \tag{58}$$

△

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

$$A = \left(\frac{1}{2}(A+A^T) - \delta^2(A)\right) \;+\; \left(\frac{1}{2}(A-A^T) + \delta^2(A)\right) \tag{59}$$

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

$$\operatorname{tr}\left(\left(\frac{1}{2}(A+A^T) - \delta^2(A)\right)\left(\frac{1}{2}(A-A^T) + \delta^2(A)\right)\right) = 0 \tag{60}$$

In the ambient space of real matrices, the antisymmetric antihollow subspace is described

$$\mathbb{S}_h^{M\perp} \triangleq \left\{\frac{1}{2}(A-A^T) + \delta^2(A) \mid A\in\mathbb{R}^{M\times M}\right\} \subseteq \mathbb{R}^{M\times M} \tag{61}$$

because any matrix in $\mathbb{S}_h^M$ is orthogonal to any matrix in $\mathbb{S}_h^{M\perp}$. Yet in the ambient space of symmetric matrices $\mathbb{S}^M$, the antihollow subspace is nontrivial;

$$\mathbb{S}_h^{M\perp} \triangleq \left\{\delta^2(A) \mid A\in\mathbb{S}^M\right\} \;=\; \left\{\delta(u) \mid u\in\mathbb{R}^M\right\} \subseteq \mathbb{S}^M \tag{62}$$

In anticipation of their utility with Euclidean distance matrices (EDMs) in §5, for symmetric hollow matrices we introduce the linear bijective vectorization dvec that is the natural analogue to symmetric matrix vectorization svec (47): for $Y = [Y_{ij}]\in\mathbb{S}_h^M$

$$\operatorname{dvec} Y \;\triangleq\; \sqrt{2}\begin{bmatrix} Y_{12} \\ Y_{13} \\ Y_{23} \\ Y_{14} \\ Y_{24} \\ Y_{34} \\ \vdots \\ Y_{M-1,M} \end{bmatrix} \in \mathbb{R}^{M(M-1)/2} \tag{63}$$

Like svec (47), dvec is an isometric isomorphism on the symmetric hollow subspace.

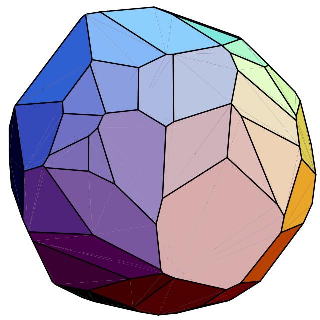


Figure 12: Convex hull of a random list of points in $\mathbb{R}^3$. Some points from that generating list reside in the interior of this convex polyhedron. [282, *Convex Polyhedron*] (Avis-Fukuda-Mizukoshi)

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

$$\{E_{ij}\in\mathbb{S}_h^M\} \;=\; \left\{\frac{1}{\sqrt{2}}\left(e_i e_j^T + e_j e_i^T\right),\quad 1\le i<j\le M\right\} \tag{64}$$

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* on page 484 characterizes eigenvalues of symmetric hollow matrices.

## 2.3 Hulls

### 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), 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. [230, §1] [148, §A.2.1]

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

$$X = [\,x_1 \;\cdots\; x_N\,] \in \mathbb{R}^{n\times N} \tag{65}$$

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 $X$ ;

$$r \triangleq \dim \operatorname{aff} X \tag{66}$$

Affine dimension $r$ is a lower bound sometimes called *embedding dimension.* [266] [134] That affine set $\mathcal{A}$ in which those points are embedded is unique and called the *affine hull* [46, §2.1.2] [247, §2.1];

$$\begin{array}{rlcl} \mathcal{A} \triangleq & \operatorname{aff}\{x_\ell \in \mathbb{R}^n,\ \ell=1 \dots N\} & = & \operatorname{aff} X \\ = & x_1 + \mathcal{R}\{x_\ell - x_1\ ,\ \ell=2 \dots N\} & = & \{Xa \mid a^T\mathbf{1} = 1\} \ \subseteq \ \mathbb{R}^n \end{array} \tag{67}$$

parallel to subspace

$$\mathcal{R}\{x_\ell - x_1\ ,\ \ell=2 \dots N\} \ = \ \mathcal{R}(X - x_1\mathbf{1}^T) \ \subseteq \ \mathbb{R}^n \tag{68}$$

where

$$\mathcal{R}(A) \ = \ \{Ax \mid \forall\, x\} \qquad (120)$$

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

$$\operatorname{aff} \mathcal{C} = x + \operatorname{aff}(\mathcal{C} - x) \tag{69}$$

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

**2.3.1.0.1 Definition.** *Affine subset.*
We analogize *affine subset* to subspace,[2.14] defining it to be any nonempty affine set (§2.1.4). △

$$\operatorname{aff} \emptyset \triangleq \emptyset \tag{70}$$

The affine hull of a point $x$ is that point itself;

$$\operatorname{aff}\{x\} = \{x\} \tag{71}$$

The affine hull of two distinct points is the unique line through them. (Figure **13**) 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; (§2.9)

$$\operatorname{aff} \mathbb{S}^m_+ = \mathbb{S}^m \tag{72}$$

[2.14] The popular term *affine subspace* is an oxymoron.

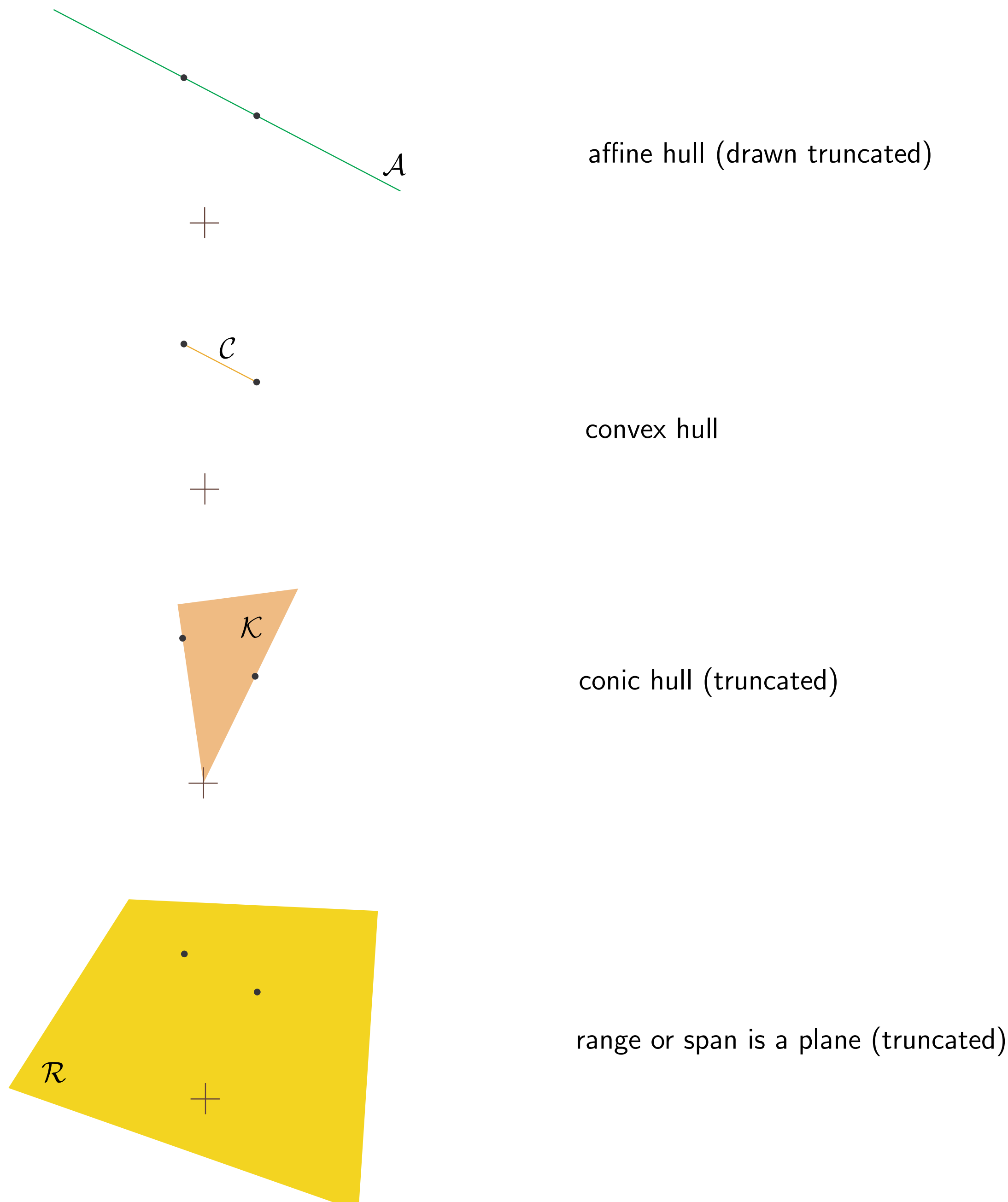



Figure 13: Given two points in Euclidean vector space of any dimension, their various hulls are illustrated. Each hull is a subset of range; generally, $\mathcal{A}\,,\mathcal{C}\,,\mathcal{K} \subseteq \mathcal{R}\,.$ (Cartesian axes drawn for reference.)

**2.3.1.0.2 Example.** *Affine hull of rank-1 correlation matrices.* [160]
The set of all $m\times m$ *rank-1 correlation matrices* is defined by all the binary vectors $y$ in $\mathbb{R}^m$ (*confer* §5.9.1.0.1)

$$\{yy^T\in\mathbb{S}_+^m \mid \delta(yy^T)=\mathbf{1}\} \tag{73}$$

Affine hull of the rank-1 correlation matrices is equal to the set of normalized symmetric matrices; *id est*,

$$\text{aff}\{yy^T\in\mathbb{S}_+^m \mid \delta(yy^T)=\mathbf{1}\} \;=\; \{A\in\mathbb{S}^m \mid \delta(A)=\mathbf{1}\} \tag{74}$$

□

**2.3.1.0.3 Exercise.** *Affine hull of correlation matrices.*
Prove (74) via definition of affine hull. Find the convex hull instead. ▼

#### 2.3.1.1 Comparison with respect to $\mathbb{R}_+^N$ and $\mathbb{S}_+^M$

The notation $a\succeq 0$ means vector $a$ belongs to the nonnegative orthant $\mathbb{R}_+^N$ , whereas $a\succeq b$ 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 $b$ necessarily belongs to that orthant. In particular, $a\succeq b \;\Leftrightarrow\; a_i\succeq b_i \;\forall\, i$ . (320)

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 $a^Ty\geq b$ . 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* [148, §A.1.4] [46, §2.1.4] [230] of any *bounded*[2.15] list (or set) of $N$ points $X\in\mathbb{R}^{n\times N}$ forms a unique *convex polyhedron* (§2.12.0.0.1) whose vertices constitute some subset of that list;

$$\mathcal{P} \;\triangleq\; \text{conv}\,\{x_\ell\;,\;\ell=1\ldots N\} \;=\; \text{conv}\,X \;=\; \{Xa \mid a^T\mathbf{1}=1,\; a\succeq 0\}\subseteq\mathbb{R}^n \tag{75}$$

[2.15] A set in $\mathbb{R}^n$ is bounded if and only if it can be contained in a Euclidean ball having finite radius. [77, §2.2] (*confer* §5.7.3.0.1)

The union of relative interior and *relative boundary* (§2.6.1.3) of the polyhedron comprise the convex hull $\mathcal{P}$, the smallest closed convex set that contains the list $X$; *e.g.*, Figure **12**. Given $\mathcal{P}$, the *generating list* $\{x_\ell\}$ is not unique.

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

$$\operatorname{conv}\mathcal{C} \;\subseteq\; \operatorname{aff}\mathcal{C} \;=\; \operatorname{aff}\overline{\mathcal{C}} \;=\; \overline{\operatorname{aff}\mathcal{C}} \;=\; \operatorname{aff}\operatorname{conv}\mathcal{C} \tag{76}$$

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

$$\mathcal{C} = \operatorname{conv}\partial\mathcal{C} \tag{77}$$

$$\operatorname{conv}\emptyset \triangleq \emptyset \tag{78}$$

**2.3.2.0.1 Example.** *Hull of outer product.* [212] [9, §4.1] [216, §3] [175, §2.4] Convex hull of the set comprising outer product of orthonormal matrices has equivalent expression: for $1\leq k\leq N$ (§2.9.0.1)

$$\operatorname{conv}\left\{UU^T \mid U\in\mathbb{R}^{N\times k},\; U^TU=I\right\} = \left\{A\in\mathbb{S}^N \mid I\succeq A\succeq 0\,,\; \langle I\,,\,A\rangle=k\right\} \subset \mathbb{S}_+^N \tag{79}$$

This important convex body we call *Fantope* (after mathematician Ky Fan). In case $k=1$, there is slight simplification: ((1403), Example 2.9.2.4.1)

$$\operatorname{conv}\left\{UU^T \mid U\in\mathbb{R}^{N},\; U^TU=I\right\} = \left\{A\in\mathbb{S}^N \mid A\succeq 0\,,\; \langle I\,,\,A\rangle=1\right\} \tag{80}$$

In case $k=N$, the Fantope is identity matrix $I$. More generally, the set

$$\{UU^T \mid U\in\mathbb{R}^{N\times k},\; U^TU=I\} \tag{81}$$

comprises the extreme points (§2.6.0.0.1) of its convex hull. By (1268), each and every extreme point $UU^T$ has only $k$ nonzero eigenvalues $\lambda$ and they all equal 1; *id est*, $\lambda(UU^T)_{1:k}=\lambda(U^TU)=\mathbf{1}$. So the Frobenius norm of each and every extreme point equals the same constant

$$\|UU^T\|_{\mathrm{F}}^2 = k \tag{82}$$

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*

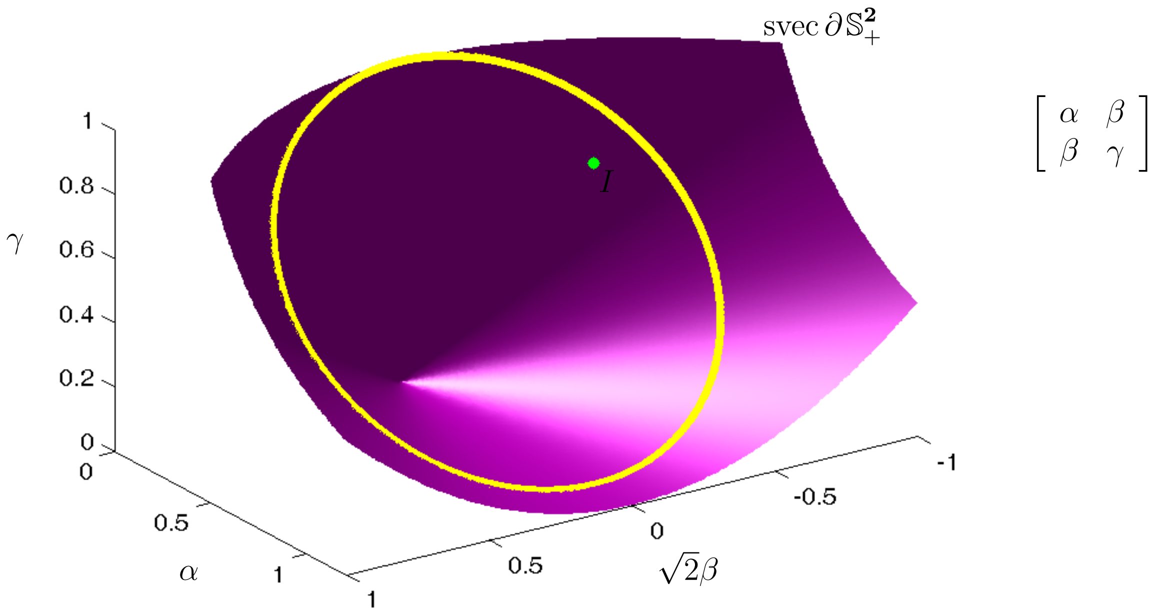



Figure 14: Two Fantopes. Circle, (radius $\frac{1}{\sqrt{2}}$) shown here on boundary of positive semidefinite cone $\mathbb{S}^2_+$ in isometrically isomorphic $\mathbb{R}^3$ from Figure **31**, comprises boundary of a Fantope (79) in this dimension $(k=1,\ N=2)$. Lone point illustrated is identity matrix $I$ and that Fantope corresponding to $k=2,\ N=2$. (View is from inside PSD cone looking toward origin.)

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 $\mathbf{0}$) through the identity matrix $I$ in isomorphic vector space $\mathbb{R}^{N(N+1)/2}$ (§2.2.2.1).

Figure **14** illustrates extreme points (81) 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. (§2.9.2.5) □

### 2.3.3 Conic hull

In terms of a finite-length point list (or set) arranged columnar in $X\in\mathbb{R}^{n\times N}$ (65), its conic hull is expressed

$$\mathcal{K} \triangleq \operatorname{cone}\{x_\ell\,,\ \ell=1\ldots N\} \;=\; \operatorname{cone}X \;=\; \{Xa \mid a\succeq 0\} \;\subseteq\; \mathbb{R}^n \tag{83}$$

*id est*, every nonnegative combination of points from the list. The conic hull of any finite-length list forms a *polyhedral cone* [148, §A.4.3] (§2.12.1.0.1; *e.g.*, Figure **15**); the smallest closed convex cone that contains the list.

By convention, the aberration [247, §2.1]

$$\operatorname{cone}\emptyset \triangleq \{\mathbf{0}\} \tag{84}$$

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

$$\operatorname{conv}\mathcal{C} \subseteq \operatorname{cone}\mathcal{C} \tag{85}$$

### 2.3.4 Vertex-description

The conditions in (67), (75), and (83) 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.*

## 2.4 Halfspace, Hyperplane

A two-dimensional affine subset is called a *plane.* An $(n-1)$-dimensional affine subset in $\mathbb{R}^n$ is called a *hyperplane.* [230] [148] Every hyperplane partially bounds a halfspace (which is convex but not affine).

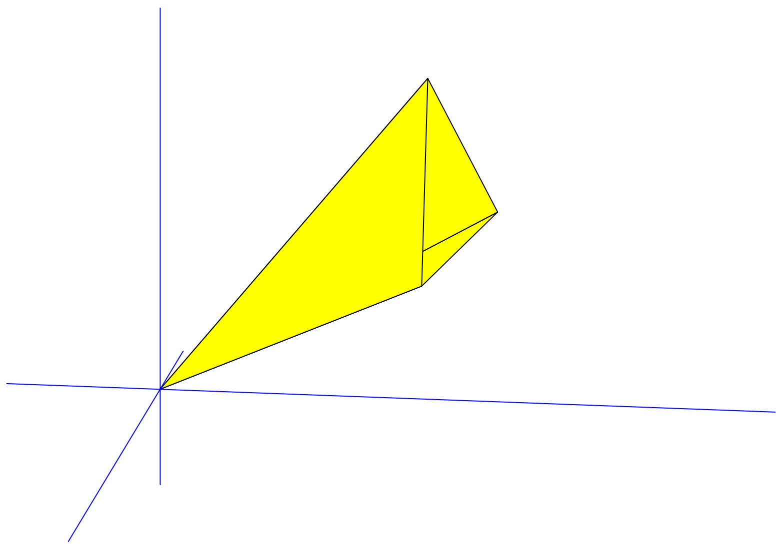


Figure 15: A simplicial cone (§2.12.3.1.1) in $\mathbb{R}^{\mathbf{3}}$ whose boundary is drawn truncated; constructed using $A \in \mathbb{R}^{\mathbf{3}\times\mathbf{3}}$ and $C = \mathbf{0}$ in (246). By the most fundamental definition of a cone (§2.7.1), entire boundary can be constructed from an 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.*

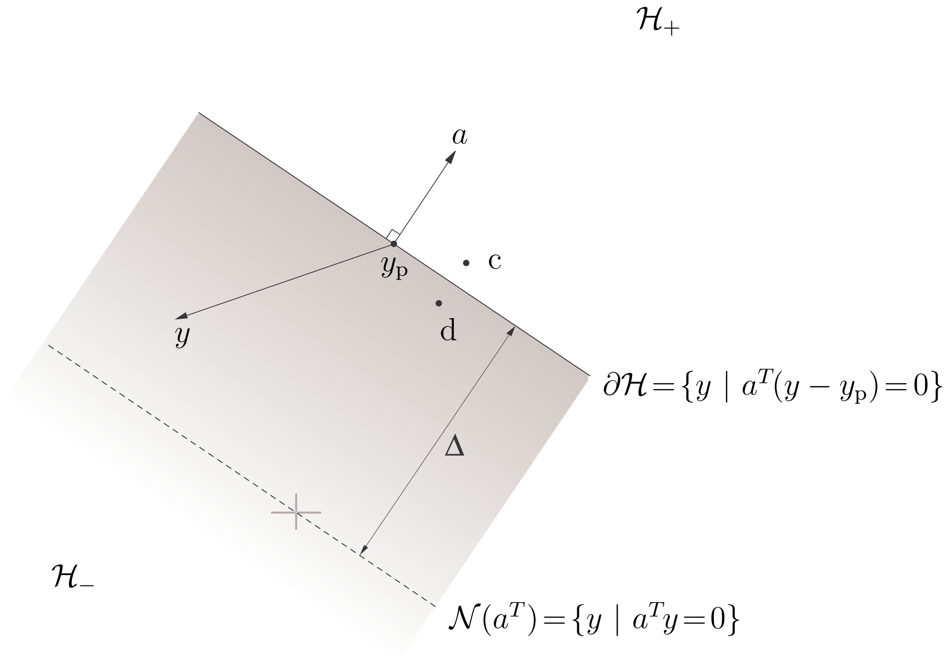



Figure 16: Hyperplane illustrated $\partial\mathcal{H}$ is a line partially bounding halfspaces $\mathcal{H}_- = \{y \mid a^T(y - y_{\rm p}) \leq 0\}$ and $\mathcal{H}_+ = \{y \mid a^T(y - y_{\rm p}) \geq 0\}$ in $\mathbb{R}^{\mathbf{2}}$. Shaded is a rectangular piece of semi-infinite $\mathcal{H}_-$ 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 $a^Ty_{\rm p} > 0$. Hyperplane, halfspace, and nullspace are each drawn truncated. Points c and d are equidistant from hyperplane, and vector $\mathrm{c} - \mathrm{d}$ 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 into two halfspaces 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

$$\mathcal{H}_- = \{y \mid a^T y \leq b\} = \{y \mid a^T(y - y_{\rm p}) \leq 0\} \subset \mathbb{R}^n \tag{86}$$
$$\mathcal{H}_+ = \{y \mid a^T y \geq b\} = \{y \mid a^T(y - y_{\rm p}) \geq 0\} \subset \mathbb{R}^n \tag{87}$$

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 $y - y_{\rm p}$ that makes an obtuse angle with normal $a$, vector $y$ will lie in the halfspace $\mathcal{H}_-$ on one side (shaded in Figure **16**) 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 **16**,

$$\mathcal{H}_- = \{y \mid \|y - {\rm d}\| \leq \|y - {\rm c}\|\} \tag{88}$$

This representation, in terms of proximity, is resolved with the more conventional representation of a halfspace (86) by squaring both sides of the inequality in (88);

$$\mathcal{H}_- = \left\{y \mid ({\rm c} - {\rm d})^T y \leq \frac{\|{\rm c}\|^2 - \|{\rm d}\|^2}{2}\right\} = \left\{y \mid ({\rm c} - {\rm d})^T\left(y - \frac{{\rm c} + {\rm d}}{2}\right) \leq 0\right\} \tag{89}$$

#### 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* [182, §5.12] [16, §1] [88, §I.1.2] which guarantees any closed convex set to be an intersection of halfspaces.

**2.4.1.1.1 Theorem.** *Halfspaces.* [46, §2.3.1] [230, §18] [148, §A.4.2(b)] [30, §2.4] A closed convex set in $\mathbb{R}^n$ is equivalent to the intersection of all halfspaces that contain it. ⋄

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

$$\bigcap_i \mathcal{H}_{i-} \;=\; \{y \mid A^T y \preceq b\} \;=\; \{y \mid A^T(y-y_{\rm p}) \preceq 0\} \tag{90}$$

$$\bigcap_i \mathcal{H}_{i+} \;=\; \{y \mid A^T y \succeq b\} \;=\; \{y \mid A^T(y-y_{\rm p}) \succeq 0\} \tag{91}$$

where $b$ is now a vector, and the $i^{\text{th}}$ 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.*

### 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.

$$\dim \partial\mathcal{H} = n-1 \tag{92}$$

so a hyperplane is a point in $\mathbb{R}$, a line in $\mathbb{R}^{\mathbf{2}}$, a plane in $\mathbb{R}^{\mathbf{3}}$, and so on. Every hyperplane can be described as the intersection of complementary halfspaces; [230, §19]

$$\partial\mathcal{H} \;=\; \mathcal{H}_- \cap \mathcal{H}_+ \;=\; \{y \mid a^T y \le b \;,\; a^T y \ge b\} \;=\; \{y \mid a^T y = b\} \tag{93}$$

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 **16** and Figure **17** for $\mathbb{R}^{\mathbf{2}}$)

$$\partial\mathcal{H} \stackrel{\Delta}{=} \{y \mid a^T y = b\} = \{y \mid a^T(y-y_{\rm p}) = 0\} = \{Z\xi + y_{\rm p} \mid \xi \in \mathbb{R}^{n-1}\} \subset \mathbb{R}^n \tag{94}$$

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

$$y = Z\xi + y_{\rm p} \tag{95}$$

where the columns of $Z \in \mathbb{R}^{n\times n-1}$ constitute a basis for the nullspace $\mathcal{N}(a^T) = \{x \in \mathbb{R}^n \mid a^T x = \mathbf{0}\}$.[2.16]

[2.16] We will later find this expression for $y$ in terms of nullspace of $a^T$ (more generally, of matrix $A^T$ (122)) to be a useful device for eliminating affine equality constraints, much as we did here.

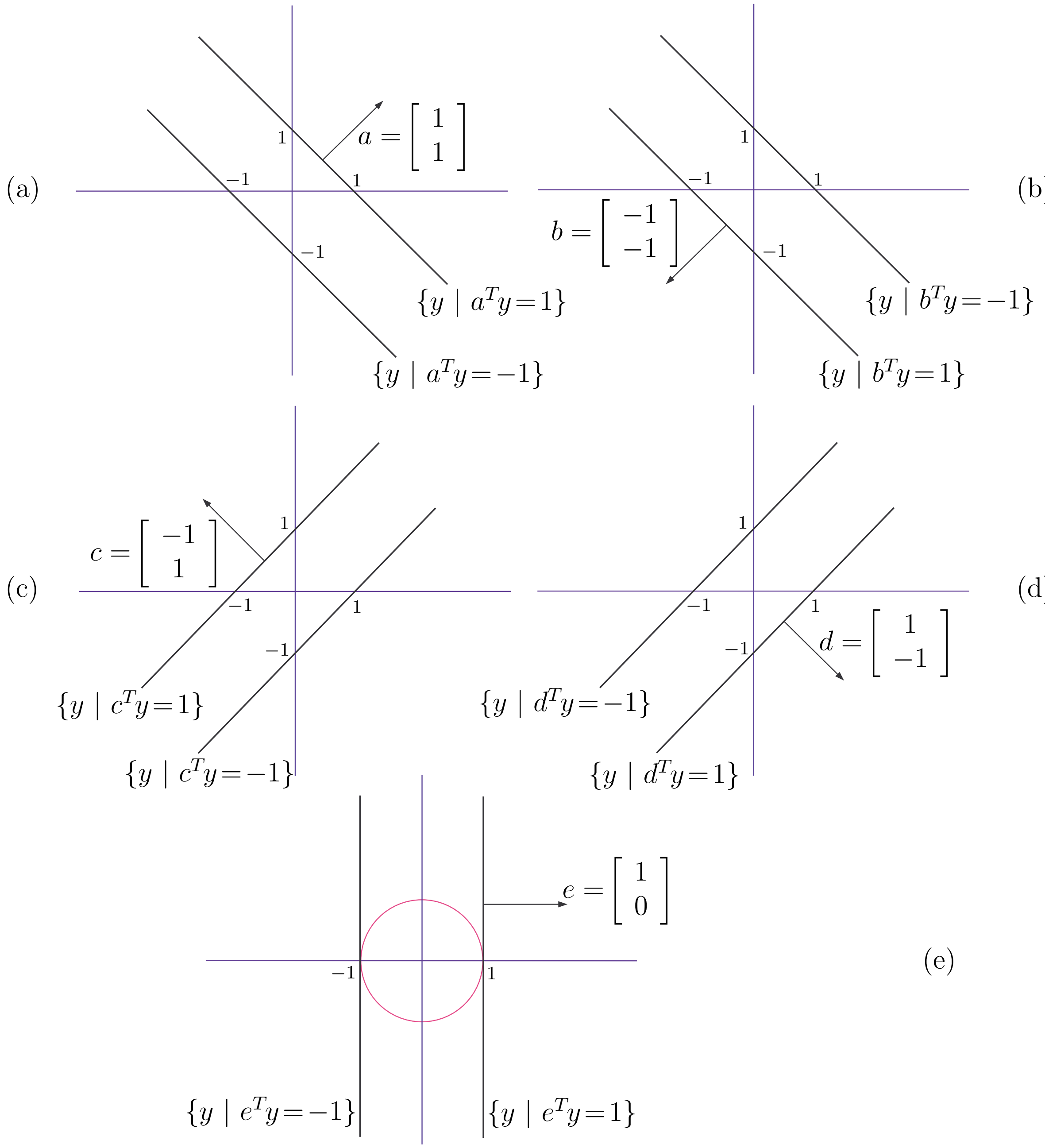



Figure 17: **(a)-(d)** Hyperplanes in $\mathbb{R}^2$ (truncated). Movement in normal direction increases vector inner-product. This visual concept is exploited to attain analytical solution of linear programs; *e.g.*, Example 2.4.2.6.2, Example 3.1.6.0.1, [46, exer.4.8-exer.4.20]. Each graph is also interpretable as a contour plot of a real affine function of two variables as in Figure **55**. **(e)** Ratio $|b|/\|a\|$ from $\{x \mid a^T x = b\}$ represents radius of hypersphere about $\mathbf{0}$ supported by hyperplane whose normal is $a$.

Conversely, given any point $y_{\rm p}$ in $\mathbb{R}^n$, the unique hyperplane containing it having normal $a$ is the affine set $\partial\mathcal{H}$ (94) where $b$ equals $a^Ty_{\rm p}$ and where a basis for $\mathcal{N}(a^T)$ is arranged in $Z$ columnar. Hyperplane dimension is apparent from the dimensions 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 $\{x\in\mathbb{R}^{\mathbf{2}} \mid x^Ty=1\}$. Suppose $z=\frac{1}{2}y$. On the same plot, draw the hyperplane $\{x\in\mathbb{R}^{\mathbf{2}} \mid x^Tz=1\}$. Now suppose $z=2y$, 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 $\partial\mathcal{H}$ (§E.5.0.0.5), equal to $a\Delta/\|a\|$ if the origin is known *a priori* to belong to halfspace $\mathcal{H}_-$ (Figure **16**), or equal to $-a\Delta/\|a\|$ if the origin belongs to halfspace $\mathcal{H}_+$; *id est*, when $\mathcal{H}_-\ni\mathbf{0}$

$$\partial\mathcal{H} = \left\{y \mid a^T(y-a\Delta/\|a\|)=0\right\} = \left\{y \mid a^Ty=\|a\|\Delta\right\} \tag{96}$$

or when $\mathcal{H}_+\ni\mathbf{0}$

$$\partial\mathcal{H} = \left\{y \mid a^T(y+a\Delta/\|a\|)=0\right\} = \left\{y \mid a^Ty=-\|a\|\Delta\right\} \tag{97}$$

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 $b=\langle A\,,Y_{\rm p}\rangle\in\mathbb{R}$,

$$\begin{aligned}\mathcal{H}_- &= \{Y\in\mathbb{R}^{mn} \mid \langle A\,,Y\rangle\leq b\} = \{Y\in\mathbb{R}^{mn} \mid \langle A\,,Y-Y_{\rm p}\rangle\leq 0\} &(98)\\ \mathcal{H}_+ &= \{Y\in\mathbb{R}^{mn} \mid \langle A\,,Y\rangle\geq b\} = \{Y\in\mathbb{R}^{mn} \mid \langle A\,,Y-Y_{\rm p}\rangle\geq 0\} &(99)\end{aligned}$$

Recall vector inner-product from §2.2, $\langle A\,,Y\rangle=\mathrm{tr}(A^TY)$.

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

$$\partial\mathcal{H} = \{Y \mid \langle A\,, Y\rangle = b\} = \{Y \mid \langle A\,, Y - Y_{\rm p}\rangle = 0\} \subset \mathbb{R}^{mn} \tag{100}$$

Vector $a$ from Figure **16** is normal to the hyperplane illustrated. Likewise, nonzero vectorized matrix $A$ is normal to hyperplane $\partial\mathcal{H}$ ;

$$A \perp \partial\mathcal{H} \text{ in } \mathbb{R}^{mn} \tag{101}$$

#### 2.4.2.2 Vertex-description of hyperplane

Any hyperplane in $\mathbb{R}^n$ may be described as the affine hull of a *minimal set* of points $\{x_\ell \in \mathbb{R}^n,\ \ell = 1 \ldots n\}$ arranged columnar in a matrix $X \in \mathbb{R}^{n\times n}$ (65):

$$\begin{aligned}
\partial\mathcal{H} &= \operatorname{aff}\{x_\ell \in \mathbb{R}^n,\ \ell = 1 \ldots n\}\ , && \dim \operatorname{aff}\{x_\ell\ \forall\,\ell\} = n-1\\
&= \operatorname{aff} X\ , && \dim \operatorname{aff} X = n-1\\
&= x_1 + \mathcal{R}\{x_\ell - x_1\ ,\ \ell = 2 \ldots n\}\ , && \dim \mathcal{R}\{x_\ell - x_1\ ,\ \ell = 2 \ldots n\} = n-1\\
&= x_1 + \mathcal{R}(X - x_1\mathbf{1}^T)\ , && \dim \mathcal{R}(X - x_1\mathbf{1}^T) = n-1
\end{aligned} \tag{102}$$

where

$$\mathcal{R}(A) = \{Ax \mid \forall\, x\} \qquad (120)$$

#### 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 descriptive 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 $\zeta \in \mathbb{R}^N \ni \zeta^T\mathbf{1} = 1,\ \zeta_k = 0$ (*confer* §2.1.2)

$$x_i\,\zeta_i + \cdots + x_j\,\zeta_j - x_k = \mathbf{0}\,, \qquad i \neq \cdots \neq j \neq k = 1 \ldots N \tag{103}$$

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

$$\text{l.i.} \Rightarrow \text{a.i.} \tag{104}$$

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. [153, §3] (Figure **18**) This is equivalent to the property that the columns of $\begin{bmatrix} X \\ \mathbf{1}^T \end{bmatrix}$ (for $X \in \mathbb{R}^{n\times N}$ as in (65)) form a linearly independent set. [148, §A.1.3]

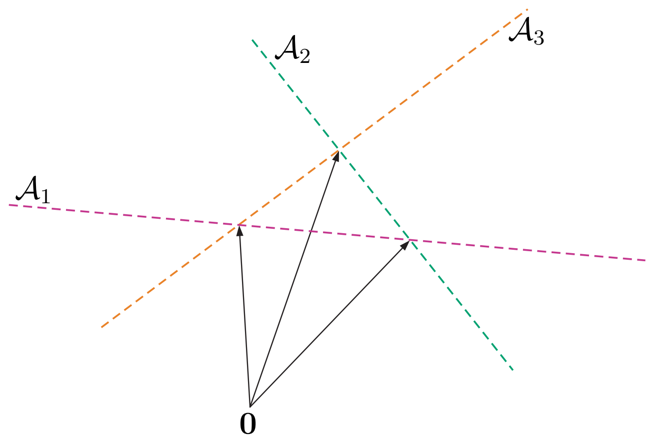



Figure 18: Any one particular point of three points illustrated does not belong to affine hull $\mathcal{A}_i$ $(i \in 1,2,3,$ each drawn truncated$)$ of points remaining. Three corresponding vectors in $\mathbb{R}^{\mathbf{2}}$ are, therefore, affinely independent (but neither linearly or conically independent).

#### 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 $X \in \mathbb{R}^{n\times N}$ (65) holds an affinely independent set in its columns. Consider a transformation

$$T(X) : \mathbb{R}^{n\times N} \to \mathbb{R}^{n\times N} \triangleq XY \tag{105}$$

where the given matrix $Y \triangleq [\, y_1 \;\; y_2 \;\cdots\; y_N \,] \in \mathbb{R}^{N\times N}$ is represented by linear operator $T$. Affine independence of $\{Xy_i \in \mathbb{R}^n ,\; i=1 \ldots N\}$ demands $($by definition (103)$)$ there exists no solution $\zeta \in \mathbb{R}^N \;\ni\; \zeta^T\mathbf{1}=1,\; \zeta_k = 0,$ to

$$Xy_i\,\zeta_i + \cdots + Xy_j\,\zeta_j - Xy_k = \mathbf{0}\,, \qquad i \neq \cdots \neq j \neq k = 1 \ldots N \tag{106}$$

By factoring $X$, 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

$$\mathcal{N}(A) = \{x \mid Ax = \mathbf{0}\} \qquad (121)$$

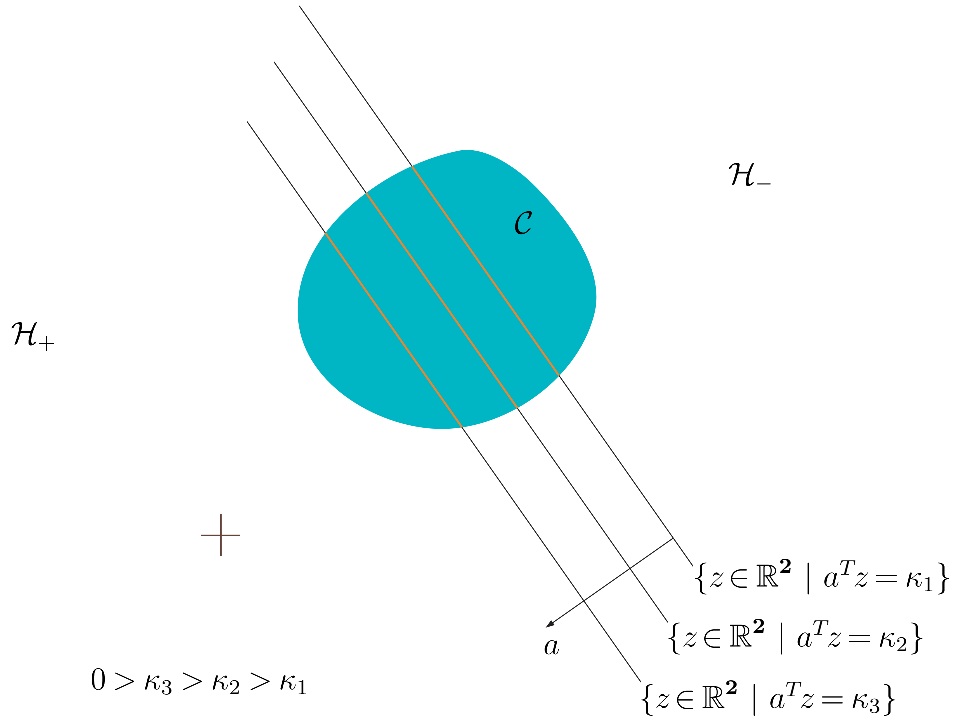



Figure 19: Each shaded line segment $\{z \in \mathcal{C} \mid a^T z = \kappa_i\}$ belonging to set $\mathcal{C} \subset \mathbb{R}^2$ shows intersection with hyperplane parametrized by scalar $\kappa_i$ ; each shows a (linear) contour in vector $z$ of equal inner product with normal vector $a$ . Cartesian axes drawn for reference. (*confer* Figure **55**)

#### 2.4.2.5 affine maps

Affine transformations preserve affine hulls. Given any affine mapping $T$ of vector spaces and some arbitrary set $\mathcal{C}$ [230, p.8]

$$\text{aff}(T\mathcal{C}) = T(\text{aff}\,\mathcal{C}) \tag{107}$$

#### 2.4.2.6 PRINCIPLE 2: Supporting hyperplane

The second most fundamental principle of convex geometry also follows from the *geometric Hahn-Banach theorem* [182, §5.12] [16, §1] that guarantees existence of at least one hyperplane in $\mathbb{R}^n$ supporting a convex set (having

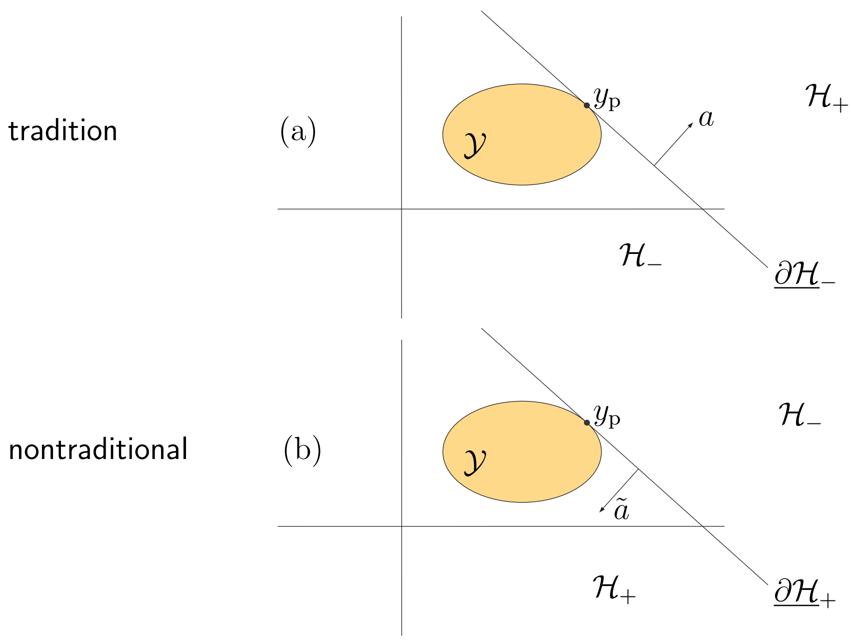



Figure 20: **(a)** Hyperplane $\underline{\partial\mathcal{H}_-}$ (108) supporting closed set $\mathcal{Y}\in\mathbb{R}^2$. Vector $a$ is inward-normal to hyperplane with respect to halfspace $\mathcal{H}_+$, but outward-normal with respect to set $\mathcal{Y}$. A supporting hyperplane can be considered the limit of an increasing sequence in the normal-direction like that in Figure **19**. **(b)** Hyperplane $\underline{\partial\mathcal{H}_+}$ nontraditionally supporting $\mathcal{Y}$. Vector $\tilde{a}$ is inward-normal to hyperplane now with respect to both halfspace $\mathcal{H}_+$ and set $\mathcal{Y}$. Tradition [148] [230] recognizes only positive normal polarity in support function $\sigma_{\mathcal{Y}}$ as in (108); *id est*, normal $a$, figure (a). But both interpretations of supporting hyperplane are useful.

nonempty interior)[2.17] at each point on its boundary.

**2.4.2.6.1 Definition.** *Supporting hyperplane* $\underline{\partial\mathcal{H}}$.
The partial boundary $\partial\mathcal{H}$ of a closed halfspace that contains arbitrary set $\mathcal{Y}$ is called a supporting hyperplane $\underline{\partial\mathcal{H}}$ to $\mathcal{Y}$ when the hyperplane contains at least one point of $\overline{\mathcal{Y}}$. [230, §11] Specifically, given normal $a \neq \mathbf{0}$ (belonging to $\mathcal{H}_+$ by definition), the supporting hyperplane to $\mathcal{Y}$ at $y_{\mathrm{p}} \in \partial\mathcal{Y}$ [*sic*] is

$$\begin{aligned}\underline{\partial\mathcal{H}}_- &= \left\{y \mid a^T(y-y_{\mathrm{p}})=0\,,\quad y_{\mathrm{p}}\in\overline{\mathcal{Y}}\,,\quad a^T(z-y_{\mathrm{p}})\leq 0\quad \forall\, z\in\overline{\mathcal{Y}}\right\}\\ &= \left\{y \mid a^Ty=\sup\{a^Tz \mid z\in\mathcal{Y}\}\right\}\end{aligned} \tag{108}$$

where normal $a$ and set $\mathcal{Y}$ reside in opposite halfspaces. (Figure 20(a)) Real function

$$\sigma_{\mathcal{Y}}(a) \triangleq \sup\{a^Tz \mid z\in\mathcal{Y}\} \qquad (458)$$

is called the *support function* for $\mathcal{Y}$.

An equivalent but nontraditional representation[2.18] for a supporting hyperplane is obtained by reversing polarity of normal $a$; (1459)

$$\begin{aligned}\underline{\partial\mathcal{H}}_+ &= \left\{y \mid \tilde{a}^T(y-y_{\mathrm{p}})=0\,,\quad y_{\mathrm{p}}\in\overline{\mathcal{Y}}\,,\quad \tilde{a}^T(z-y_{\mathrm{p}})\geq 0\quad \forall\, z\in\overline{\mathcal{Y}}\right\}\\ &= \left\{y \mid \tilde{a}^Ty=-\inf\{\tilde{a}^Tz \mid z\in\mathcal{Y}\}=\sup\{-\tilde{a}^Tz \mid z\in\mathcal{Y}\}\right\}\end{aligned} \tag{109}$$

where normal $\tilde{a}$ and set $\mathcal{Y}$ now both reside in $\mathcal{H}_+$. (Figure 20(b))

When a supporting hyperplane contains only a single point of $\overline{\mathcal{Y}}$, that hyperplane is termed *strictly supporting* (and termed *tangent* to $\mathcal{Y}$ if the supporting hyperplane is unique there [230, §18, p.169]). △

A closed convex set $\mathcal{C}\subset\mathbb{R}^n$, for example, can be expressed as the intersection of all halfspaces partially bounded by hyperplanes supporting it; *videlicet*, [182, p.135]

$$\overline{\mathcal{C}} = \bigcap_{a\in\mathbb{R}^n}\left\{y \mid a^Ty\leq\sigma_{\mathcal{C}}(a)\right\} \tag{110}$$

by the *halfspaces theorem* (§2.4.1.1.1).

[2.17] It is conventional to speak of a hyperplane supporting set $\mathcal{C}$ but not containing $\mathcal{C}$; called *nontrivial support*. [230, p.100] Hyperplanes in support of lower-dimensional bodies are admitted.

[2.18] useful for constructing the dual cone; *e.g.*, Figure 42(b). Tradition recognizes the polar cone; which is the negative of the dual cone.

There is no geometric difference[2.19] between supporting hyperplane $\underline{\partial\mathcal{H}}_+$ or $\underline{\partial\mathcal{H}}_-$ or $\underline{\partial\mathcal{H}}$ and an ordinary hyperplane $\partial\mathcal{H}$ coincident with them.

**2.4.2.6.2 Example.** *Minimization on the unit cube.*
Consider minimization of a linear function on a hypercube, given vector $c$

$$\begin{array}{cl}\underset{x}{\text{minimize}} & c^T x\\ \text{subject to} & -\mathbf{1} \preceq x \preceq \mathbf{1}\end{array} \tag{111}$$

This convex optimization problem is called a *linear program* because the objective of minimization is linear and the constraints describe a polyhedron (intersection of a finite number of halfspaces and hyperplanes). Applying graphical concepts from Figure **17**, Figure **19**, and Figure **20**, an optimal solution can be shown to be $x^\star = -\operatorname{sgn}(c)$ but is not necessarily unique. Because a solution always exists at a hypercube vertex (§2.6.1.0.1) regardless of the value of nonzero vector $c$ [64], mathematicians see this geometry as a means to *relax* a discrete problem (whose desired solution is integer, *confer* Example 4.2.3.0.2). $\square$

**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 we instead minimize absolute value in (111). Are the following programs equivalent for some arbitrary real convex set $\mathcal{C}$ ? (*confer* (433))

$$\begin{array}{cl}\underset{x\in\mathbb{R}}{\text{minimize}} & |x|\\ \text{subject to} & -1 \leq x \leq 1\\ & x \in \mathcal{C}\end{array} \quad\equiv\quad \begin{array}{cl}\underset{x_+\,,\,x_-}{\text{minimize}} & x_+ + x_-\\ \text{subject to} & 1 \geq x_- \geq 0\\ & 1 \geq x_+ \geq 0\\ & x_+ - x_- \in \mathcal{C}\end{array} \tag{112}$$

Many optimization problems of interest and some older methods of solution require nonnegative variables. The method illustrated below splits a variable into its nonnegative and negative parts; $x = x_+ - x_-$ (extensible

[2.19] If vector-normal polarity is unimportant, we may instead signify a supporting hyperplane by $\underline{\partial\mathcal{H}}$ .

to vectors). Under what conditions on vector $a$ and scalar $b$ is an optimal solution $x^\star$ negative infinity?

$$\begin{array}{cl}\underset{x_+\in\mathbb{R}\;,\;x_-\in\mathbb{R}}{\text{minimize}} & x_+ - x_-\\ \text{subject to} & x_- \geq 0\\ & x_+ \geq 0\\ & a^T\begin{bmatrix} x_+ \\ x_- \end{bmatrix} = b\end{array} \tag{113}$$

Minimization of the *objective function*[2.20] entails maximization of $x_-$ .

▼

#### 2.4.2.7 PRINCIPLE 3: Separating hyperplane

The third most fundamental principle of convex geometry again follows from the *geometric Hahn-Banach theorem* [182, §5.12] [16, §1] [88, §I.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 (§2.6.1.3), then there exists a separating hyperplane $\underline{\partial\mathcal{H}}$ 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. [46, §2.5.1]

2) A *strictly separating hyperplane* $\partial\mathcal{H}$ intersects the closure of neither set; its existence is guaranteed when the intersection of the closures is empty and at least one set is bounded. [148, §A.4.1]

### 2.4.3 Angle between hyperspaces

Given halfspace descriptions, the dihedral angle between hyperplanes and 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

$$\measuredangle(\partial\mathcal{H}_a\,,\partial\mathcal{H}_b) \triangleq \arccos\left(\frac{\langle a\;,\;b\rangle}{\|a\|\;\|b\|}\right) \text{ radians} \tag{114}$$

[2.20] The objective is the function that is argument to minimization or maximization.

## 2.5 Subspace representations

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

The fundamental vector subspaces associated with a matrix $A\in\mathbb{R}^{m\times n}$ [249, §3.1] are ordinarily related

$$\mathcal{R}(A^T)\perp\mathcal{N}(A)\,,\qquad \mathcal{N}(A^T)\perp\mathcal{R}(A) \tag{115}$$

$$\mathcal{R}(A^T)\oplus\mathcal{N}(A)=\mathbb{R}^n\,,\qquad \mathcal{N}(A^T)\oplus\mathcal{R}(A)=\mathbb{R}^m \tag{116}$$

and of dimension:

$$\dim\mathcal{R}(A^T)\;=\;\dim\mathcal{R}(A)\;=\;\operatorname{rank}A\;\leq\;\min\{m\,,n\} \tag{117}$$

with complementarity

$$\dim\mathcal{N}(A)=n-\operatorname{rank}A\;,\qquad \dim\mathcal{N}(A^T)=m-\operatorname{rank}A \tag{118}$$

From these four fundamental subspaces, the rowspace and range identify one form of subspace description (range form or vertex-description (§2.3.4))

$$\mathcal{R}(A^T)\triangleq\operatorname{span}A^T=\{A^Ty\mid y\in\mathbb{R}^m\}=\{x\in\mathbb{R}^n\mid A^Ty=x\,,\;\; y\in\mathcal{R}(A)\} \tag{119}$$

$$\mathcal{R}(A)\triangleq\operatorname{span}A\;=\;\{Ax\mid x\in\mathbb{R}^n\}\;=\;\{y\in\mathbb{R}^m\mid Ax=y\,,\;\; x\in\mathcal{R}(A^T)\} \tag{120}$$

while the nullspaces identify the second common form (nullspace form or halfspace-description (93))

$$\mathcal{N}(A)\triangleq\{x\in\mathbb{R}^n\mid Ax=\mathbf{0}\} \tag{121}$$

$$\mathcal{N}(A^T)\triangleq\{y\in\mathbb{R}^m\mid A^Ty=\mathbf{0}\} \tag{122}$$

Range forms (119) (120) are realized as the respective span of the column vectors in matrices $A^T$ and $A$, whereas nullspace form (121) or (122) is the solution set to a linear equation similar to hyperplane definition (94). 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 (121), each row of $A$ is normal to a hyperplane while each row of $A^T$ is a normal for (122).

### 2.5.1 Subspace or affine subset ...

Any particular vector subspace $\mathcal{R}_\mathrm{p}$ can be described as $\mathcal{N}(A)$ the nullspace of some matrix $A$ or as $\mathcal{R}(B)$ the range of some matrix $B$.

More generally, we have the choice of expressing an $n-m$-dimensional affine subset in $\mathbb{R}^n$ 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* (rank $= \min\{m\,,n\}$) matrix

$$A \triangleq \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m\times n} \tag{123}$$

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

$$\mathcal{A} \triangleq \{x \in \mathbb{R}^n \mid Ax = b\} = \bigcap_{i=1}^{m} \left\{x \mid a_i^T x = b_i\right\} \tag{124}$$

a halfspace-description. (93)

For example: The intersection of any two independent[2.21] 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.

For $n > k$

$$\mathcal{A} \cap \mathbb{R}^k = \{x \in \mathbb{R}^n \mid Ax = b\} \cap \mathbb{R}^k = \bigcap_{i=1}^{m} \left\{x \in \mathbb{R}^k \mid a_i(1\!:\!k)^T x = b_i\right\} \tag{125}$$

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

[2.21] Hyperplanes are said to be *independent* iff the normals defining them are linearly independent.

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

Alternatively, we may compute a basis for the nullspace of matrix $A$ in (123) and then equivalently express the affine subset as its range plus an offset: Define

$$Z \triangleq \operatorname{basis} \mathcal{N}(A) \in \mathbb{R}^{n\times n-m} \tag{126}$$

so $AZ = \mathbf{0}$. Then we have the vertex-description,

$$\mathcal{A} = \{x \in \mathbb{R}^n \mid Ax = b\} = \left\{Z\xi + x_{\mathrm{p}} \mid \xi \in \mathbb{R}^{n-m}\right\} \subseteq \mathbb{R}^n \tag{127}$$

the offset span of $n-m$ column vectors, where $x_{\mathrm{p}}$ is any particular solution to $Ax = b$.

**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 $A_1 \in \mathbb{R}^{\mathbf{2}\times\mathbf{4}}$ ($m=2$, $n=4$) as in (124):

$$\mathcal{A}_1 \triangleq \left\{x \in \mathbb{R}^{\mathbf{4}} \;\middle|\; \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} x = b_1 \right\} \tag{128}$$

The nullspace of $A_1$ is two dimensional (from $Z$ in (127)), so $\mathcal{A}_1$ represents a plane in four dimensions. Similarly define a second plane in terms of $A_2 \in \mathbb{R}^{\mathbf{2}\times\mathbf{4}}$:

$$\mathcal{A}_2 \triangleq \left\{x \in \mathbb{R}^{\mathbf{4}} \;\middle|\; \begin{bmatrix} a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} x = b_2 \right\} \tag{129}$$

If the two planes are independent (meaning any line in one is linearly independent of any line from the other), they will intersect at a point because then $\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ is invertible;

$$\mathcal{A}_1 \cap \mathcal{A}_2 = \left\{x \in \mathbb{R}^{\mathbf{4}} \;\middle|\; \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right\} \tag{130}$$

□

### 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

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}\left(\begin{bmatrix} A \\ B \end{bmatrix}\right) \triangleq \left\{x \in \mathbb{R}^n \mid \begin{bmatrix} A \\ B \end{bmatrix} x = \mathbf{0}\right\} \tag{131}$$

the 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 $\mathcal{N}(BZ)$. Then [110, §12.4.2]

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{R}(ZW) \tag{132}$$

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 [18, §6], or in the circumstance $A$ and $B$ are two linearly independent dyads (§B.1.1), then

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}(A+B)\,, \quad \left\{\begin{array}{l} A, B \in \mathbb{S}_+^M \\ \quad\textbf{or} \\ A+B = u_1 v_1^T + u_2 v_2^T \quad \text{(l.i.)} \end{array}\right. \tag{133}$$

## 2.6 Extreme, Exposed

**2.6.0.0.1 Definition.** *Extreme point.*
An extreme point $x_\varepsilon$ of a convex set $\mathcal{C}$ is a point, belonging to its closure $\overline{\mathcal{C}}$ [30, §3.3], that is not expressible as a convex combination of points in $\overline{\mathcal{C}}$ distinct from $x_\varepsilon$ ; *id est*, for $x_\varepsilon \in \overline{\mathcal{C}}$ and all $x_1, x_2 \in \overline{\mathcal{C}} \backslash x_\varepsilon$

$$\mu\, x_1 + (1-\mu) x_2 \neq x_\varepsilon\,, \quad \mu \in [0,1] \tag{134}$$

$\triangle$

In other words, $x_\varepsilon$ is an extreme point of $\mathcal{C}$ if and only if $x_\varepsilon$ is not a point relatively interior to any line segment in $\overline{\mathcal{C}}$ . [268, §2.10]

Borwein & Lewis offer: [41, §4.1.6] An extreme point of a convex set $\mathcal{C}$ is a point $x_\varepsilon$ in $\overline{\mathcal{C}}$ whose *relative complement* $\overline{\mathcal{C}}\setminus x_\varepsilon$ is convex.

The set consisting of a single point $\mathcal{C}=\{x_\varepsilon\}$ is itself an extreme point.

**2.6.0.0.2 Theorem.** *Extreme existence.* [230, §18.5.3] [20, §II.3.5]
A nonempty closed convex set containing no lines has at least one extreme point. ⋄

**2.6.0.0.3 Definition.** *Face, edge.* [148, §A.2.3]

- A *face* $\mathcal{F}$ of convex set $\mathcal{C}$ is a convex subset $\mathcal{F}\subseteq\overline{\mathcal{C}}$ such that every closed line segment $\overline{x_1x_2}$ in $\overline{\mathcal{C}}$ , having a relatively interior point ($x\in\operatorname{rel\,int}\overline{x_1x_2}$) in $\mathcal{F}$ , has both endpoints in $\mathcal{F}$ . The zero-dimensional faces of $\mathcal{C}$ constitute its extreme points. The empty set and $\overline{\mathcal{C}}$ itself are conventional faces of $\mathcal{C}$ . [230, §18]

- All faces $\mathcal{F}$ are extreme sets by definition; *id est*, for $\mathcal{F}\subseteq\overline{\mathcal{C}}$ and all $x_1, x_2\in\overline{\mathcal{C}}\setminus\mathcal{F}$
$$\mu\, x_1 + (1-\mu)x_2 \notin \mathcal{F}\ , \quad \mu\in[0,1] \tag{135}$$

- A one-dimensional face of a convex set is called an *edge*. △

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

$$\dim\mathcal{F} = \sup_\rho \dim\{x_2 - x_1\ ,\ x_3 - x_1\ ,\ \ldots\ ,\ x_\rho - x_1 \mid x_i\in\mathcal{F}\ ,\ i=1\ldots\rho\} \tag{136}$$

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

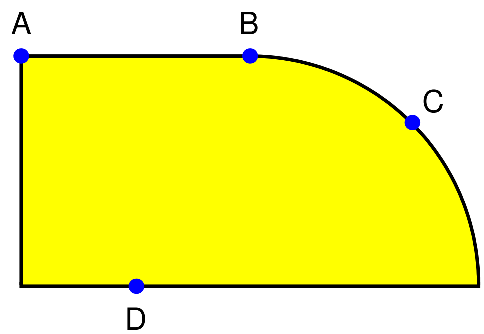



Figure 21: Closed convex set in $\mathbb{R}^2$. Point A is exposed hence extreme; a classical vertex. Point B is extreme but not an exposed point. Point C is exposed and extreme; zero-dimensional exposure makes it a vertex. Point D is neither an exposed or extreme point although it belongs to a one-dimensional exposed face. [148, §A.2.4] [247, §3.6] Closed face $\overline{\mathrm{AB}}$ 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.

## 2.6.1 Exposure

**2.6.1.0.1 Definition.** *Exposed face, exposed point, vertex, facet.* [148, §A.2.3, A.2.4]

- $\mathcal{F}$ is an *exposed face* of an $n$-dimensional convex set $\mathcal{C}$ iff there is a supporting hyperplane $\underline{\partial\mathcal{H}}$ to $\overline{\mathcal{C}}$ such that

$$\mathcal{F} = \overline{\mathcal{C}} \cap \underline{\partial\mathcal{H}} \tag{137}$$

  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.

- A *facet* is an $(n-1)$-dimensional exposed face of an $n$-dimensional convex set $\mathcal{C}$ ; in one-to-one correspondence with the $(n-1)$-dimensional faces.[2.22]

- $\overline{\{\text{exposed points}\}} = \{\text{extreme points}\}$
  $\{\text{exposed faces}\} \subseteq \{\text{faces}\}$ △

#### 2.6.1.1 Density of exposed points

For any closed convex set $\mathcal{C}$ , its exposed points constitute a *dense* subset of its extreme points; [230, §18] [252] [247, §3.6, p.115] dense in the sense [282] that closure of that subset yields the set of extreme points.

For the convex set illustrated in Figure **21**, point B cannot be exposed because it relatively bounds both the facet $\overline{\mathrm{AB}}$ 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 .

[2.22] This coincidence occurs simply because the facet's dimension is the same as the dimension of the supporting hyperplane exposing it.

#### 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. [230, §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 $\mathcal{F}_3$ . [164, def.115/6, p.358] 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 **12**, 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.

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

$$\mathcal{F}(\mathcal{C} \ni G) \tag{138}$$

*videlicet*, $\mathcal{C} \supseteq \mathcal{F}(\mathcal{C} \ni G) \ni G$ . An affine set has no faces except itself and the empty set. The smallest face that contains $G$ of the intersection of convex set $\mathcal{C}$ with an affine set $\mathcal{A}$ [175, §2.4]

$$\mathcal{F}\big((\mathcal{C} \cap \mathcal{A}) \ni G\big) = \mathcal{F}(\mathcal{C} \ni G) \cap \mathcal{A} \tag{139}$$

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

#### 2.6.1.3 Boundary

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

$$\partial\mathcal{C} = \overline{\mathcal{C}} \setminus \operatorname{int}\mathcal{C} \qquad (14)$$

More suitable for the study of convex sets is the relative boundary; defined [148, §A.2.1.2]

$$\operatorname{rel}\partial\mathcal{C} = \overline{\mathcal{C}} \setminus \operatorname{rel}\operatorname{int}\mathcal{C} \tag{140}$$

the boundary relative to the affine hull of $\mathcal{C}$ , conventionally equivalent to:

**2.6.1.3.1 Definition.** *Conventional boundary of convex set.* [148, §C.3.1] The relative boundary $\partial\mathcal{C}$ of a nonempty convex set $\mathcal{C}$ is the union of all the exposed faces of $\overline{\mathcal{C}}$ . △

Equivalence of this definition to (140) comes about because it is conventionally presumed that any supporting hyperplane, central to the definition of exposure, does not contain $\mathcal{C}$ . [230, p.100]

Any face $\mathcal{F}$ of convex set $\mathcal{C}$ (that is not $\mathcal{C}$ itself) belongs to $\operatorname{rel}\partial\mathcal{C}$ . (§2.8.2.1) In the exception when $\mathcal{C}$ is a single point $\{x\}$ , (11)

$$\operatorname{rel}\partial\{x\} \;=\; \overline{\{x\}}\backslash\{x\} \;=\; \emptyset\;, \qquad x\in\mathbb{R}^n \tag{141}$$

A bounded convex polyhedron (§2.12.0.0.1) having nonempty interior, for example, in $\mathbb{R}$ has a boundary constructed from two points, in $\mathbb{R}^{\mathbf{2}}$ from at least three line segments, in $\mathbb{R}^{\mathbf{3}}$ from convex polygons, while a convex *polychoron* (a bounded polyhedron in $\mathbb{R}^{\mathbf{4}}$ [282]) has a boundary constructed from three-dimensional convex polyhedra.

By Definition 2.6.1.3.1, an affine set has no relative boundary.

## 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 (§E), whereas projection on a closed convex cone can be determined from projection instead on its *algebraic complement* (§2.13, §E.9.2.1); called the *polar cone.*

**2.7.0.0.1 Definition.** *Ray.*
The one-dimensional set

$$\{\zeta\Gamma + B \mid \zeta \geq 0\,,\;\Gamma \neq \mathbf{0}\} \subset \mathbb{R}^n \tag{142}$$

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

The conventional boundary of a single ray, base $\mathbf{0}$, 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.

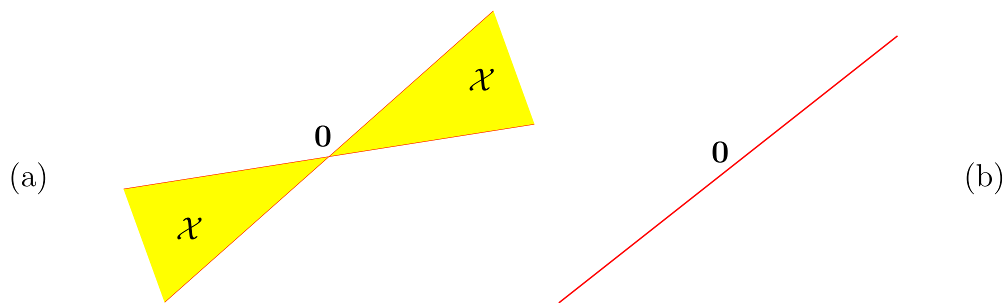



Figure 22: **(a)** Two-dimensional nonconvex cone drawn truncated. Boundary of this cone is itself a cone. Each polar half is itself a convex cone. **(b)** This convex cone **(**drawn truncated**)** is a line through the origin in any dimension. It has no relative boundary, while its relative interior comprises entire line.

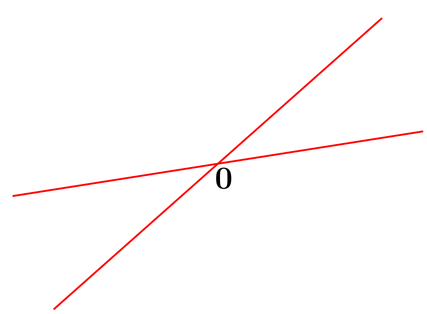



Figure 23: This nonconvex cone in $\mathbb{R}^2$ is a pair of lines through the origin. [182, §2.4]

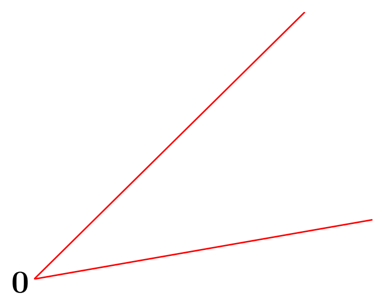



Figure 24: Boundary of a convex cone in $\mathbb{R}^2$ is a nonconvex cone; a pair of rays emanating from the origin.

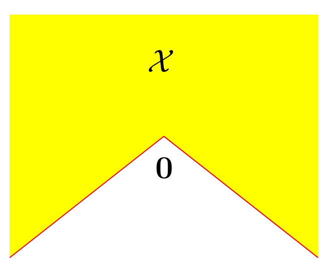



Figure 25: Nonconvex cone $\mathcal{X}$ drawn truncated in $\mathbb{R}^2$. Boundary is also a cone. [182, §2.4] Cone exterior is convex cone.

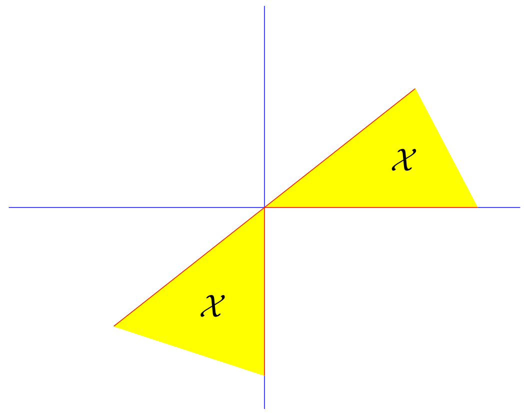



Figure 26: Truncated nonconvex cone $\mathcal{X} = \{x \in \mathbb{R}^{\mathbf{2}} \mid x_1 \geq x_2\ ,\ x_1 x_2 \geq 0\}$. Boundary is also a cone. [182, §2.4] Cartesian axes drawn for reference. Each half (about the origin) is itself a convex cone.

### 2.7.1 Cone defined

A set $\mathcal{X}$ is called, simply, *cone* if and only if

$$\Gamma \in \mathcal{X} \;\Rightarrow\; \zeta\Gamma \in \overline{\mathcal{X}} \;\text{ for all }\; \zeta \geq 0 \tag{143}$$

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.*, $\mathcal{X} = \left\{x \in \mathbb{R}^{\mathbf{2}} \mid x_1 x_2 \geq 0\right\}$ which is not convex. [280, §2.5] Similar examples are shown in Figure **22** and Figure **26**.

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 and are unbounded, excepting the simplest cone $\{\mathbf{0}\}$. The empty set $\emptyset$ is not a cone, but its conic hull is;

$$\operatorname{cone}\emptyset \triangleq \{\mathbf{0}\} \tag{84}$$

### 2.7.2 Convex cone

We call the set $\mathcal{K} \subseteq \mathbb{R}^M$ a *convex cone* iff

$$\Gamma_1, \Gamma_2 \in \mathcal{K} \;\Rightarrow\; \zeta\Gamma_1 + \xi\Gamma_2 \in \overline{\mathcal{K}} \text{ for all } \zeta, \xi \geq 0 \tag{144}$$

Apparent from this definition, $\zeta\Gamma_1 \in \overline{\mathcal{K}}$ and $\xi\Gamma_2 \in \overline{\mathcal{K}}$ for all $\zeta, \xi \geq 0$. The set $\mathcal{K}$ is convex since, for any particular $\zeta, \xi \geq 0$

$$\mu\zeta\Gamma_1 + (1-\mu)\xi\Gamma_2 \in \overline{\mathcal{K}} \quad \forall \mu \in [0,1] \tag{145}$$

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

Obviously,

$$\{\mathcal{X}\} \supset \{\mathcal{K}\} \tag{146}$$

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). The interior of a convex cone is possibly empty.

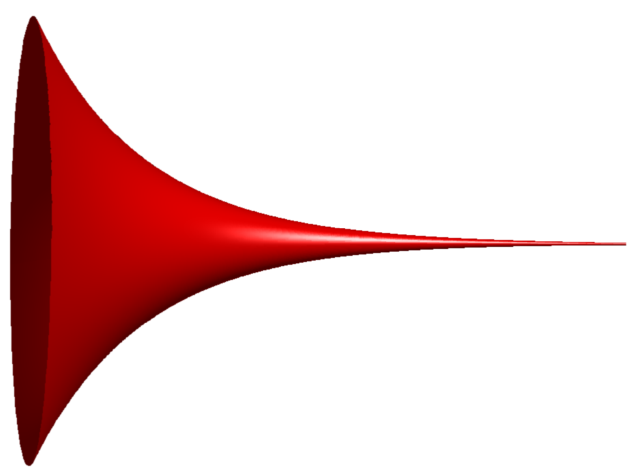


Figure 27: Not a cone; ironically, the three-dimensional *flared horn* (with or without its interior) resembling the mathematical symbol $\succ$ denoting cone membership and partial order.

Familiar examples of convex cones include an unbounded *ice-cream cone* united with its interior (a.k.a: *second-order cone*, *quadratic cone*, *circular cone* (§2.9.2.5.1), *Lorentz cone* (*confer* Figure **34**) [46, exmps.2.3 & 2.25]),

$$\mathcal{K}_\ell = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_\ell \leq t \right\} , \qquad \ell = 2 \tag{147}$$

and any 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 $\mathbb{S}_+^M$ (160), the cone of Euclidean distance matrices $\mathbb{EDM}^N$ (711) (§6), any subspace, and Euclidean vector space $\mathbb{R}^n$.

#### 2.7.2.1 cone invariance

(*confer* Figures: **15**, **22**, **23**, **24**, **25**, **26**, **27**, **29**, **31**, **38**, **41**, **44**, **46**, **47**, **48**, **49**, **50**, **51**, **52**, **95**, **108**, **130**) More Euclidean bodies are cones, it seems, than are not. This class of convex body, the convex cone, is invariant to nonnegative scaling, vector summation, affine and inverse affine transformation, Cartesian product, and intersection, [230, p.22] but is not invariant to projection; *e.g.*, Figure **33**.

**2.7.2.1.1 Theorem.** *Cone intersection.* [230, §2, §19]
The intersection of an arbitrary collection of convex cones is a convex cone. Intersection of an arbitrary collection of closed convex cones is a closed convex cone. [189, §2.3] Intersection of a finite number of polyhedral cones (§2.12.1.0.1, Figure **38** p.123) is polyhedral. ⋄

The property *pointedness* is associated with a convex cone.

**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 $\Gamma \in \overline{\mathcal{K}}$ such that $-\Gamma \in \overline{\mathcal{K}}$. [46, §2.4.1] If the origin is an extreme point of $\overline{\mathcal{K}}$ or, equivalently, if

$$\overline{\mathcal{K}} \cap -\overline{\mathcal{K}} = \{\mathbf{0}\} \tag{148}$$

then $\mathcal{K}$ is pointed, and *vice versa.* [247, §2.10] △

Thus the simplest and only bounded [280, p.75] convex cone $\mathcal{K} = \{\mathbf{0}\} \subseteq \mathbb{R}^n$ is pointed by convention, but has empty interior. Its relative boundary is the empty set (141) while its relative interior is the point itself (11). The pointed convex cone that is a halfline emanating from the origin in $\mathbb{R}^n$ has the origin as relative boundary while its relative interior is the halfline itself, excluding the origin.

**2.7.2.1.3 Theorem.** *Pointed cones.* [41, §3.3.15, exer.20]
A closed convex cone $\mathcal{K} \subset \mathbb{R}^n$ is pointed if and only if there exists a normal $\alpha$ such that the set

$$\mathcal{C} \triangleq \{x \in \mathcal{K} \mid \langle x\,, \alpha \rangle = 1\} \tag{149}$$

is closed, bounded, and $\mathcal{K} = \operatorname{cone} \mathcal{C}$. Equivalently, if and only if there exists a vector $\beta$ and positive scalar $\epsilon$ such that

$$\langle x\,, \beta \rangle \geq \epsilon \|x\| \quad \forall\, x \in \mathcal{K} \tag{150}$$

is $\mathcal{K}$ pointed. $\diamond$

If closed convex cone $\mathcal{K}$ is not pointed, then it has no extreme point. Yet a pointed closed convex cone has only one extreme point; it resides at the origin. [30, §3.3]

From the *cone intersection theorem* it follows that an intersection of convex cones is pointed if at least one of the cones is.

#### 2.7.2.2 Pointed closed convex cone and partial order

A pointed closed convex cone $\mathcal{K}$ induces *partial order* [282] on $\mathbb{R}^n$ or $\mathbb{R}^{m \times n}$, [18, §1] [242, p.7] respectively defined by vector or matrix inequality;

$$x \underset{\mathcal{K}}{\preceq} z \quad \Leftrightarrow \quad z - x \in \mathcal{K} \tag{151}$$

$$x \underset{\mathcal{K}}{\prec} z \quad \Leftrightarrow \quad z - x \in \operatorname{rel\,int} \mathcal{K} \tag{152}$$

Neither $x$ or $z$ is necessarily a member of $\mathcal{K}$ for these relations to hold. Only when $\mathcal{K}$ is the nonnegative orthant do these inequalities reduce to ordinary entrywise comparison. (§2.13.4.2.3) Inclusive of that special case, we ascribe

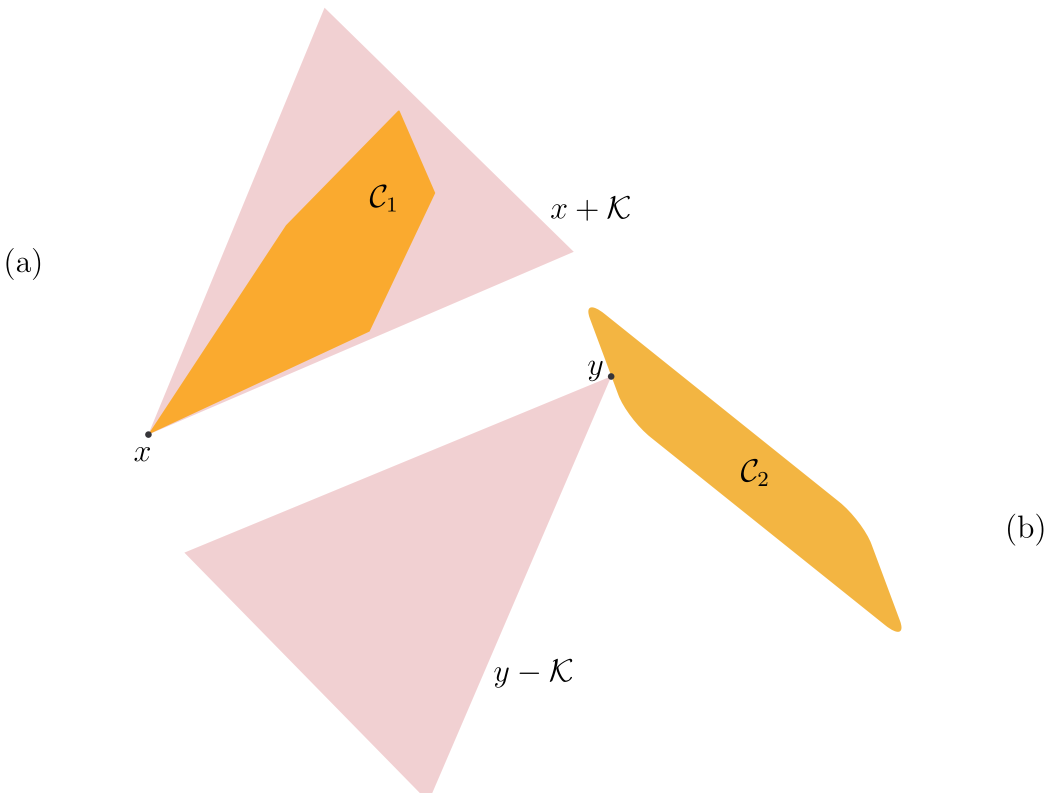



Figure 28: **(a)** Point $x$ is the minimum element of set $\mathcal{C}_1$ with respect to cone $\mathcal{K}$ because cone translated to $x \in \mathcal{C}_1$ contains set. **(b)** Point $y$ is a minimal element of set $\mathcal{C}_2$ with respect to cone $\mathcal{K}$ because negative cone translated to $y \in \mathcal{C}_2$ contains only $y$. (Cones drawn truncated in $\mathbb{R}^2$.)

nomenclature *generalized inequality* to comparison with respect to a pointed closed convex cone.

The visceral mechanics of actually comparing points when the cone $\mathcal{K}$ is not an orthant is well illustrated in the example of Figure **49** which relies on the equivalent membership-interpretation in definition (151) or (152). *Comparable points* and the *minimum element* of some vector- or matrix-valued partially ordered set are thus well defined, so decreasing 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 $\mathcal{C}\subseteq x+\mathcal{K}$ .[2.23]

Further properties of partial ordering with respect to pointed closed convex cone $\mathcal{K}$ are:

$$\begin{array}{rll}
\text{reflexivity} & (x\preceq x) & \\
\text{antisymmetry} & (x\preceq z\,,\ z\preceq x \Rightarrow x=z) & \\
\text{transitivity} & (x\preceq y\,,\ y\preceq z \Rightarrow x\preceq z), & (x\preceq y\,,\ y\prec z \Rightarrow x\prec z) \\
\text{homogeneity} & (x\preceq y\,,\ \lambda\geq 0 \Rightarrow \lambda x\preceq \lambda z), & (x\prec y\,,\ \lambda> 0 \Rightarrow \lambda x\prec \lambda z) \\
\text{additivity} & (x\preceq z\,,\ u\preceq v \Rightarrow x+u\preceq z+v), & (x\prec z\,,\ u\preceq v \Rightarrow x+u\prec z+v)
\end{array}$$

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 $\mathcal{K}$ if and only if $(x-\mathcal{K})\cap\mathcal{C}=x$ . (Figure **28**) No uniqueness is implied here, although implicit is the assumption: $\dim\mathcal{K}\geq\dim\operatorname{aff}\mathcal{C}$ .

**2.7.2.2.1 Definition.** *Proper cone*: [46, §2.4.1] a cone that is

- pointed
- closed
- convex
- has nonempty interior (is full-dimensional). △

[2.23] Borwein & Lewis [41, §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 proper cone remains proper under injective linear transformation. [62, §5.1]

Examples of proper cones are the positive semidefinite cone $\mathbb{S}_+^M$ in the ambient space of symmetric matrices (§2.9), the nonnegative real line $\mathbb{R}_+$ in vector space $\mathbb{R}$ , or any orthant in $\mathbb{R}^n$.

## 2.8 Cone boundary

Every hyperplane supporting a convex cone contains the origin. [148, §A.4.2] Because any supporting hyperplane to a convex cone must therefore itself be a cone, then from the *cone intersection theorem* it follows:

**2.8.0.0.1 Lemma.** *Cone faces.* [20, §II.8]
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 $\Gamma$ lies on the boundary $\partial\mathcal{K}$ of proper cone $\mathcal{K}$ in $\mathbb{R}^n$. Then it follows that the ray $\{\zeta\Gamma \mid \zeta \geq 0\}$ also belongs to $\partial\mathcal{K}$. ⋄

**Proof.** By virtue of its propriety, a proper cone guarantees the existence of a strictly supporting hyperplane at the origin. [230, Cor.11.7.3][2.24] Hence the origin belongs to the boundary of $\mathcal{K}$ because it is the zero-dimensional exposed face. The origin belongs to the ray through $\Gamma$, and the ray belongs to $\mathcal{K}$ by definition (143). By the *cone faces lemma*, each and every nonempty exposed face must include the origin. Hence the closed line segment $\overline{\mathbf{0}\Gamma}$ must lie in an exposed face of $\mathcal{K}$ because both endpoints do by Definition 2.6.1.3.1. That means there exists a supporting hyperplane $\partial\underline{\mathcal{H}}$ to $\mathcal{K}$ containing $\overline{\mathbf{0}\Gamma}$. So the ray through $\Gamma$ belongs both to $\mathcal{K}$ and to $\partial\underline{\mathcal{H}}$. $\partial\underline{\mathcal{H}}$ must therefore expose a face of $\mathcal{K}$ that contains the ray; *id est*,

$$\{\zeta\Gamma \mid \zeta \geq 0\} \subseteq \mathcal{K} \cap \partial\underline{\mathcal{H}} \subset \partial\mathcal{K} \tag{153}$$

♦

[2.24] Rockafellar's corollary yields a supporting hyperplane at the origin to any convex cone in $\mathbb{R}^n$ not equal to $\mathbb{R}^n$.

Proper cone $\{\mathbf{0}\}$ in $\mathbb{R}^0$ has no boundary (140) because (11)

$$\operatorname{rel\,int}\{\mathbf{0}\} = \{\mathbf{0}\} \tag{154}$$

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 $\mathcal{K}\subset\mathbb{R}^n$, being analogous to extreme point. [230, §18, p.162][2.25] An extreme direction $\Gamma_\varepsilon$ of pointed $\mathcal{K}$ is a vector corresponding to an edge that is a ray emanating from the origin.[2.26] Nonzero direction $\Gamma_\varepsilon$ in pointed $\mathcal{K}$ is extreme if and only if

$$\zeta_1\Gamma_1+\zeta_2\Gamma_2 \neq \Gamma_\varepsilon \quad \forall\,\zeta_1\,,\zeta_2 \geq 0\,, \quad \forall\,\Gamma_1\,,\Gamma_2 \in \mathcal{K}\setminus\{\zeta\Gamma_\varepsilon\in\mathcal{K} \mid \zeta\geq 0\} \tag{155}$$

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 any ray directions in the cone distinct from it.

An extreme ray is a one-dimensional face of $\mathcal{K}$. By (85), 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*[2.27] 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

[2.25] We diverge from Rockafellar's extreme direction: "extreme point at infinity".

[2.26] 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 ($\mathcal{K}^*$ in Figure **29**).

[2.27] A planar fragment; in this context, a planar cone.

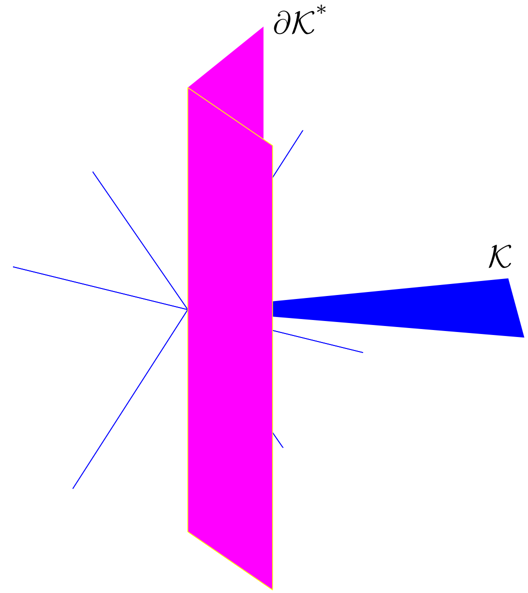



Figure 29: $\mathcal{K}$ is a pointed polyhedral cone having empty interior in $\mathbb{R}^3$ (drawn truncated and in a plane parallel to the floor upon which you stand). $\mathcal{K}^*$ is a *wedge* whose truncated boundary is illustrated (drawn perpendicular to the floor). In this particular instance, $\mathcal{K} \subset \operatorname{int} \mathcal{K}^*$ (excepting the origin). Cartesian coordinate axes drawn for reference.

the same extreme direction are therefore interpreted to be identical extreme directions.[2.28]

The extreme directions of the polyhedral cone in Figure **15** (page 60), for example, correspond to its three edges.

The extreme directions of the positive semidefinite cone (§2.9) comprise the infinite set of all symmetric rank-one matrices. [18, §6] [145, §III] It is sometimes prudent to instead consider the less infinite but complete normalized set, for $M>0$ (*confer* (193))

$$\{zz^T\in\mathbb{S}^M \mid \|z\|=1\} \tag{156}$$

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 $\mathcal{K}=\{\mathbf{0}\}$ 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. [18, §1]

Conversely, pointed closed convex cone $\mathcal{K}$ is equivalent to the convex hull of its vertex and all its extreme directions. [230, §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.* [247, §3.6] [230, §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.*, Example 2.9.2.4.1.

### 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.

[2.28] Like vectors, an extreme direction can be identified by the Cartesian point at the vector's head with respect to the origin.

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 $\mathcal{R}(A)$.

**2.8.1.2.1 Example.** *Application of extremes theorem.*
Given an extreme point at the origin and $N$ extreme rays, denoting the $i^{\text{th}}$ extreme direction by $\Gamma_i \in \mathbb{R}^n$, then the convex hull is (75)

$$\begin{aligned}\mathcal{P} &= \left\{[\mathbf{0}\;\;\Gamma_1\;\;\Gamma_2\cdots\Gamma_N]\,a\,\zeta \mid a^T\mathbf{1} = 1,\; a \succeq 0,\; \zeta \geq 0\right\}\\ &= \left\{[\Gamma_1\;\;\Gamma_2\cdots\Gamma_N]\,a\,\zeta \mid a^T\mathbf{1} \leq 1,\; a \succeq 0,\; \zeta \geq 0\right\}\\ &= \left\{[\Gamma_1\;\;\Gamma_2\cdots\Gamma_N]\,b \mid b \succeq 0\right\} \subset \mathbb{R}^n\end{aligned} \tag{157}$$

a closed convex set that is simply a conic hull like (83). □

### 2.8.2 Exposed direction

**2.8.2.0.1 Definition.** *Exposed point & direction of pointed convex cone.* [230, §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:

$$\overline{\{\text{exposed directions}\}} = \{\text{extreme directions}\} \tag{158}$$

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 $\mathbf{0}$.

The pointed closed convex cone $\mathbb{EDM}^{\mathbf{2}}$, for example, is a ray in isomorphic subspace $\mathbb{R}$ whose relative boundary (§2.6.1.3.1) is the origin. The conventionally exposed directions of $\mathbb{EDM}^{\mathbf{2}}$ 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.* [230, §18.7] (*confer* §2.8.1.1.1)
Any closed convex set $\mathcal{C}$ containing no lines (and whose dimension is at least 2) can be expressed as the closure of the convex hull of its exposed points and exposed rays. ⋄

From Theorem 2.8.1.1.1,

$$\left.\begin{aligned}
\operatorname{rel}\partial\mathcal{C} &= \overline{\mathcal{C}} \setminus \operatorname{rel}\operatorname{int}\mathcal{C} && (140)\\
&= \overline{\operatorname{conv}\{\text{exposed points and exposed rays}\}} \setminus \operatorname{rel}\operatorname{int}\mathcal{C}\\
&= \operatorname{conv}\{\text{extreme points and extreme rays}\} \setminus \operatorname{rel}\operatorname{int}\mathcal{C}
\end{aligned}\right\} \tag{159}$$

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 the relative interior by definition.

The relationship between extreme sets and the relative boundary actually goes deeper: Any face $\mathcal{F}$ of convex set $\mathcal{C}$ (that is not $\mathcal{C}$ itself) belongs to $\operatorname{rel}\partial\mathcal{C}$, so $\dim\mathcal{F} < \dim\mathcal{C}$. [230, §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*

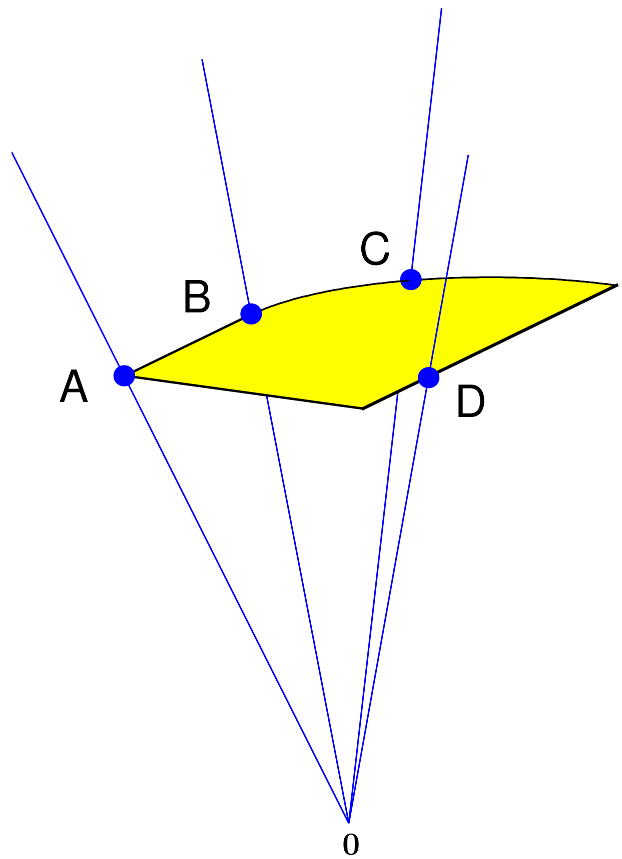



Figure 30: Properties of extreme points carry over to extreme directions. [230, §18] Four rays (drawn truncated) on boundary of conic hull of two-dimensional closed convex set from Figure 21 lifted to $\mathbb{R}^3$. Ray through point A is exposed hence extreme. Extreme direction B on cone boundary is 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.

*boundary definition* (§2.6.1.3.1); 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 $\mathbf{0}$, on the boundary of a convex cone is not necessarily an exposed or extreme direction. For the polyhedral cone illustrated in Figure **15**, 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 **21**, 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 **30**.

## 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 [20, 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*:

$$\begin{aligned}\mathbb{S}_+^M &\triangleq \left\{A\in\mathbb{S}^M \mid A\succeq 0\right\}\\ &= \left\{A\in\mathbb{S}^M \mid y^TAy\geq 0\ \ \forall\,\|y\|=1\right\}\\ &= \bigcap_{\|y\|=1}\left\{A\in\mathbb{S}^M \mid \langle yy^T,\,A\rangle\geq 0\right\}\end{aligned}\tag{160}$$

formed by the intersection of an infinite number of halfspaces (§2.4.1.1) in vectorized variable $A$ ,[2.29] 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 $\mathbb{S}^M$.

[2.29] infinite in number when $M>1$. Because $y^TA\,y=y^TA^Ty$ , matrix $A$ is almost always assumed symmetric. (§A.2.1)

The positive definite (full-rank) matrices comprise the cone interior,[2.30]

$$\begin{aligned}\operatorname{int}\mathbb{S}_+^M &= \left\{A\in\mathbb{S}^M \mid A\succ 0\right\}\\ &= \left\{A\in\mathbb{S}^M \mid y^TAy>0\ \ \forall\,\|y\|=1\right\}\\ &= \{A\in\mathbb{S}_+^M \mid \operatorname{rank}A=M\}\end{aligned}\tag{161}$$

while all singular positive semidefinite matrices (having at least one 0 eigenvalue) reside on the cone boundary (Figure **31**); (§A.7.5)

$$\begin{aligned}\partial\mathbb{S}_+^M &= \left\{A\in\mathbb{S}^M \mid \min\{\lambda(A)_i\,,\ i=1\ldots M\}=0\right\}\\ &= \left\{A\in\mathbb{S}_+^M \mid \langle yy^T,\,A\rangle=0\ \text{ for some } \|y\|=1\right\}\\ &= \left\{A\in\mathbb{S}_+^M \mid \operatorname{rank}A<M\right\}\end{aligned}\tag{162}$$

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

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

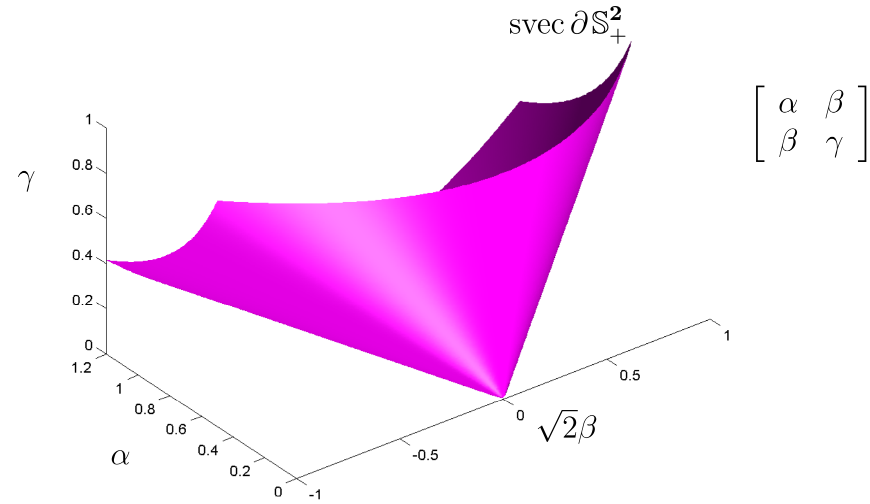
#### 2.9.0.1 Membership

Observe the notation $A\succeq 0$; meaning,[2.31] matrix $A$ is symmetric and belongs to the positive semidefinite cone in the subspace of symmetric matrices, whereas $A\succ 0$ denotes membership to that cone's interior. (§2.13.2) This notation further implies that coordinates [*sic*] for orthogonal expansion of a positive (semi)definite matrix must be its (nonnegative) positive eigenvalues (§2.13.7.1.1, §E.6.4.1.1) when expanded in its *eigenmatrices* (§A.5.1).

Generalizing comparison on the real line, the notation $A\succeq B$ denotes comparison with respect to the positive semidefinite cone; (§A.3.1) *id est*, $A\succeq B \;\Leftrightarrow\; A-B\in\mathbb{S}_+^M$ but neither matrix $A$ or $B$ necessarily belongs to the positive semidefinite cone. Yet, (1286) $A\succeq B\,,\ B\succeq 0 \;\Rightarrow\; A\succeq 0$; *id est*, $A\in\mathbb{S}_+^M$.

[2.30] The remaining inequalities in (160) also become strict for membership to the cone interior.

[2.31] 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 $a^Ty\geq b$, while $a\succeq b$ denotes comparison of vectors on $\mathbb{R}^M$ with respect to the nonnegative orthant $\mathbb{R}_+^M$ (§2.3.1.1).




Minimal set of generators are the extreme directions: $\operatorname{svec}\{yy^T \mid y \in \mathbb{R}^M\}$

Figure 31: Truncated boundary of PSD cone in $\mathbb{S}^{\mathbf{2}}$ plotted in isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$ via svec (47); courtesy, Alexandre W. d'Aspremont. (Plotted is 0-contour of smallest eigenvalue (162). Lightest shading is closest. Darkest shading is furthest and inside shell.) Entire boundary can be constructed from an aggregate of rays (§2.7.0.0.1) emanating exclusively from the origin, $\left\{\kappa^2[\, z_1^2 \;\; \sqrt{2}z_1z_2 \;\; z_2^2\,]^T \mid \kappa \in \mathbb{R}\right\}$. In this dimension the cone is circular (§2.9.2.5) while each and every ray on boundary corresponds to an extreme direction, but such is not the case in any higher dimension (*confer* Figure **15**). PSD cone geometry is not as simple in higher dimensions [20, §II.12], although for real matrices it is self-dual (321) in ambient space of symmetric matrices. [145, §II] PSD cone has no two-dimensional faces in any dimension, and its only extreme point resides at the origin.

**2.9.0.1.1 Example.** *Equality constraints in semidefinite program* (546). Employing properties of partial ordering (§2.7.2.2) for the pointed closed convex positive semidefinite cone, it is easy to show, given $A+S=C$

$$S \succeq 0 \;\Leftrightarrow\; A \preceq C \tag{163}$$

□

### 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 (144); [150, §7.1] *videlicet*, for all $\zeta_1, \zeta_2 \geq 0$ and each and every $A_1, A_2 \in \mathbb{S}^M$

$$\zeta_1 A_1 + \zeta_2 A_2 \succeq 0 \;\Leftarrow\; A_1 \succeq 0\,,\; A_2 \succeq 0 \tag{164}$$

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

$$A \succeq 0 \Leftrightarrow x^T A\, x \geq 0 \;\text{ for each and every } \|x\| = 1 \tag{165}$$

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

$$\zeta_1\, x^T A_1 x + \zeta_2\, x^T A_2 x \geq 0 \quad \text{for each and every normalized } x \in \mathbb{R}^M \tag{166}$$

The convex cone $\mathbb{S}_+^M$ 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 semi-infinite in expanse in $\mathbb{R}^3$, having boundary illustrated in Figure **31**. When $M=3$ the PSD cone is six-dimensional, and so on.

**2.9.1.0.1 Example.** *Sets from maps of positive semidefinite cone.* The set

$$\mathcal{C} = \{X \in \mathbb{S}^n \times x \in \mathbb{R}^n \mid X \succeq xx^T\} \tag{167}$$

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

$$X - xx^T \succeq 0 \;\;\Leftrightarrow\;\; f(X\,,\,x) \triangleq \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \tag{168}$$

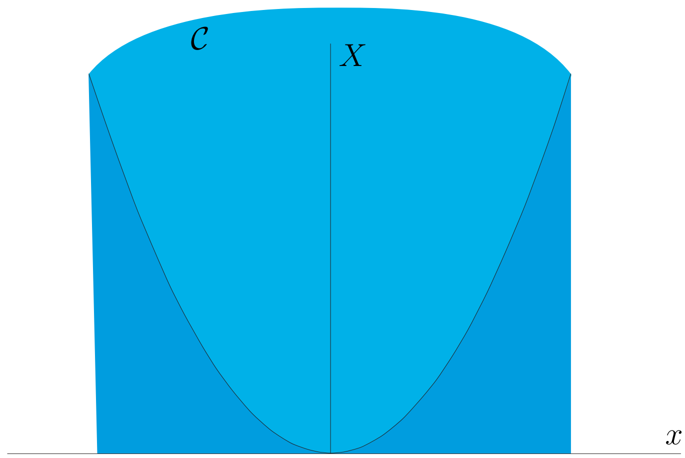



Figure 32: Convex set $\mathcal{C} = \{X \in \mathbb{S} \times x \in \mathbb{R} \mid X \succeq xx^T\}$ drawn truncated.

*e.g.*, Figure **32**. Set $\mathcal{C}$ is the inverse image (§2.1.9.0.1) of $\mathbb{S}_+^{n+1}$ under the affine mapping $f$. The set $\{X \in \mathbb{S}^n \times x \in \mathbb{R}^n \mid X \preceq xx^T\}$ is not convex, in contrast, having no Schur-form. Yet for fixed $x = x_{\rm p}$, the set

$$\{X \in \mathbb{S}^n \mid X \preceq x_{\rm p} x_{\rm p}^T\} \tag{169}$$

is simply the negative semidefinite cone shifted to $x_{\rm p} x_{\rm p}^T$. □

**2.9.1.0.2 Example.** *Inverse image of positive semidefinite cone.*
Now consider finding the set of all matrices $X \in \mathbb{S}^N$ satisfying

$$AX + B \succeq 0 \tag{170}$$

given $A, B \in \mathbb{S}^N$. Define the set

$$\mathcal{X} \triangleq \{X \mid AX + B \succeq 0\} \subseteq \mathbb{S}^N \tag{171}$$

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 (30) which is easier to visualize:

$$\operatorname{vec} g(X) = \operatorname{vec}(AX) + \operatorname{vec} B = (I \otimes A) \operatorname{vec} X + \operatorname{vec} B \tag{172}$$

where

$$I \otimes A \triangleq Q\Lambda Q^T \in \mathbb{S}^{N^2} \tag{173}$$

is a block-diagonal matrix formed by Kronecker product (§A.1 no.21, §D.1.2.1). Assign

$$\begin{aligned} x &\triangleq \operatorname{vec} X \in \mathbb{R}^{N^2} \\ b &\triangleq \operatorname{vec} B \in \mathbb{R}^{N^2} \end{aligned} \tag{174}$$

then make the equivalent problem: Find

$$\operatorname{vec} \mathcal{X} = \{x \in \mathbb{R}^{N^2} \mid (I \otimes A)x + b \in \mathcal{K}\} \tag{175}$$

where

$$\mathcal{K} \triangleq \operatorname{vec} \mathbb{S}_+^N \tag{176}$$

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 (173),

$$\begin{aligned} \operatorname{vec} \mathcal{X} &= \{x \mid \Lambda Q^T x \in Q^T(\mathcal{K} - b)\} \\ &= \{x \mid \Phi Q^T x \in \Lambda^\dagger Q^T(\mathcal{K} - b)\} \subseteq \mathbb{R}^{N^2} \end{aligned} \tag{177}$$

where $^\dagger$ denotes matrix *pseudoinverse* (§E) and

$$\Phi \triangleq \Lambda^\dagger \Lambda \tag{178}$$

is a diagonal projection matrix whose entries are either 1 or 0 (§E.3). We have the complementary sum

$$\Phi Q^T x + (I - \Phi)Q^T x = Q^T x \tag{179}$$

So, adding $(I - \Phi)Q^T x$ to both sides of the membership within (177) admits

$$\begin{aligned} \operatorname{vec} \mathcal{X} &= \{x \in \mathbb{R}^{N^2} \mid Q^T x \in \Lambda^\dagger Q^T(\mathcal{K} - b) + (I - \Phi)Q^T x\} \\ &= \{x \mid Q^T x \in \Phi\big(\Lambda^\dagger Q^T(\mathcal{K} - b)\big) \oplus (I - \Phi)\mathbb{R}^{N^2}\} \\ &= \{x \in Q\Lambda^\dagger Q^T(\mathcal{K} - b) \oplus Q(I - \Phi)\mathbb{R}^{N^2}\} \\ &= (I \otimes A)^\dagger(\mathcal{K} - b) \,\oplus\, \mathcal{N}(I \otimes A) \end{aligned} \tag{180}$$

where we used the facts: linear function $Q^T x$ in $x$ on $\mathbb{R}^{N^2}$ is a *bijection*, and $\Phi\Lambda^\dagger = \Lambda^\dagger$.

$$\operatorname{vec} \mathcal{X} = (I \otimes A)^\dagger \operatorname{vec}(\mathbb{S}_+^N - B) \,\oplus\, \mathcal{N}(I \otimes A) \tag{181}$$

In words, set $\operatorname{vec}\mathcal{X}$ 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 $\operatorname{vec}\mathcal{X}=(I\otimes A)^{-1}\operatorname{vec}(\mathbb{S}_+^N - B)$ which is the expected result.

□

## 2.9.2 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 $Y$; [249, §6.3]

$$A = YY^T \in \mathbb{S}_+^M\ , \quad \operatorname{rank} A = \rho\ , \quad Y\in\mathbb{R}^{M\times\rho} \tag{182}$$

Then the boundary of the positive semidefinite cone may be expressed

$$\partial\mathbb{S}_+^M = \left\{A\in\mathbb{S}_+^M \mid \operatorname{rank} A < M\right\} = \left\{YY^T \mid Y\in\mathbb{R}^{M\times M-1}\right\} \tag{183}$$

Because the boundary of any convex body is obtained with closure of its relative interior (§2.1.7, §2.6.1.3), from (161) we must also have

$$\begin{aligned}\mathbb{S}_+^M = \overline{\left\{A\in\mathbb{S}_+^M \mid \operatorname{rank} A = M\right\}} &= \overline{\left\{YY^T \mid Y\in\mathbb{R}^{M\times M},\ \operatorname{rank} Y = M\right\}}\\ &= \left\{YY^T \mid Y\in\mathbb{R}^{M\times M}\right\}\end{aligned} \tag{184}$$

### 2.9.2.1 rank $\rho$ subset of the positive semidefinite cone

For the same reason (closure), this applies more generally; for $0\leq\rho\leq M$

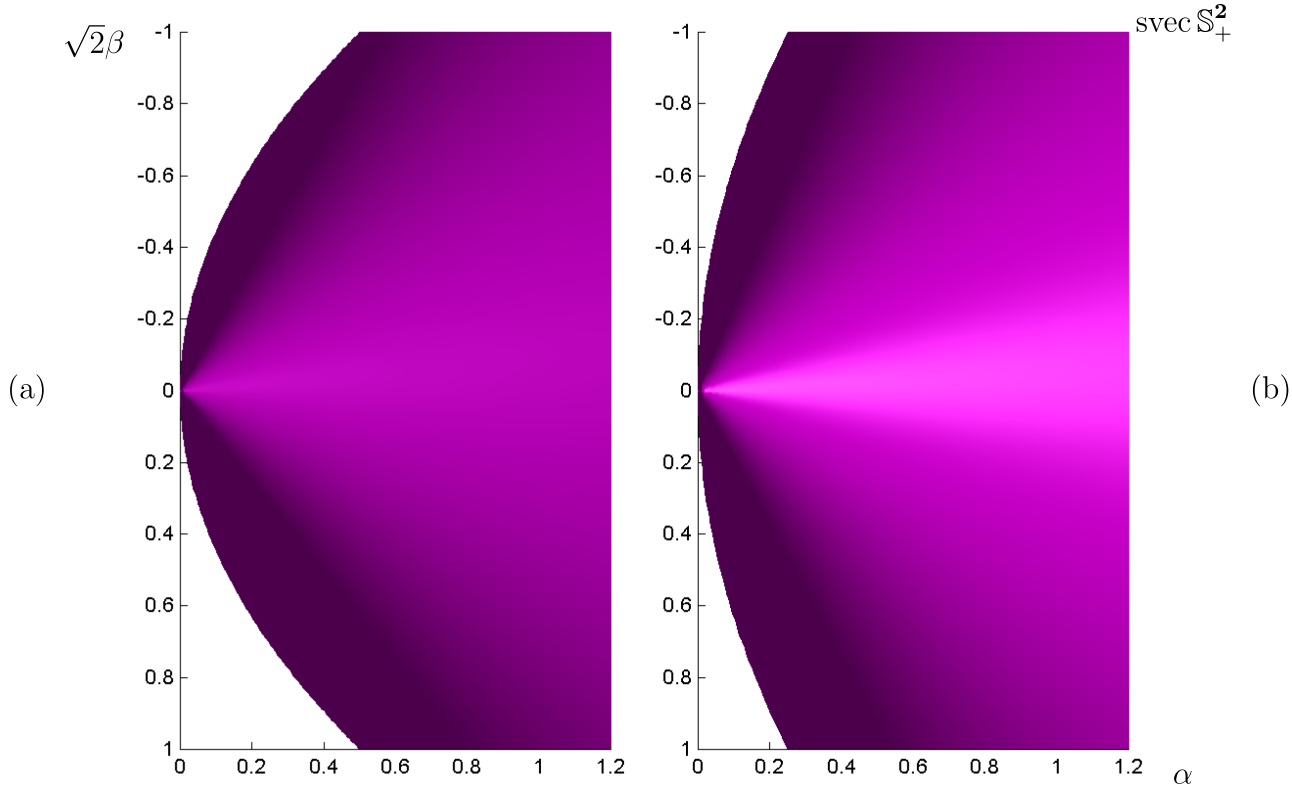
$$\overline{\left\{A\in\mathbb{S}_+^M \mid \operatorname{rank} A = \rho\right\}} = \left\{A\in\mathbb{S}_+^M \mid \operatorname{rank} A \leq \rho\right\} \tag{185}$$

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 (185). ▼

For example,

$$\partial\mathbb{S}_+^M = \overline{\left\{A\in\mathbb{S}_+^M \mid \operatorname{rank} A = M-1\right\}} = \left\{A\in\mathbb{S}_+^M \mid \operatorname{rank} A \leq M-1\right\} \tag{186}$$




$$\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$$

Figure 33: **(a)** Projection of the PSD cone $\mathbb{S}^{\mathbf{2}}_+$ , truncated above $\gamma = 1$, on $\alpha\beta$-plane in isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$. View is from above with respect to Figure 31. **(b)** Truncated above $\gamma = 2$. From these plots we may infer, for example, the line $\left\{[\,0 \quad 1/\sqrt{2} \quad \gamma\,]^T \mid \gamma \in \mathbb{R}\right\}$ intercepts the PSD cone at some large value of $\gamma$; in fact, $\gamma = \infty$.

In $\mathbb{S}^{\mathbf{2}}$, each and every ray on the boundary of the positive semidefinite cone in isomorphic $\mathbb{R}^{\mathbf{3}}$ corresponds to a symmetric rank-1 matrix (Figure **31**), 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 (185) is left unclosed as in

$$\mathcal{M}(\rho) \triangleq \left\{A \in \mathbb{S}_+^M \mid \text{rank}\, A = \rho\right\} \tag{187}$$

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 $B \in \mathcal{M}(\rho)$ [135, §5, prop.1.1]

$$\mathcal{R}_{\mathcal{M}}(B) \triangleq \{XB + BX^T \mid X \in \mathbb{R}^{M\times M}\} \subseteq \mathbb{S}^M \tag{188}$$

has dimension

$$\dim \text{svec}\, \mathcal{R}_{\mathcal{M}}(B) \;=\; \rho\left(M - \frac{\rho-1}{2}\right) \;=\; \rho(M-\rho) + \frac{\rho(\rho+1)}{2} \tag{189}$$

Tangent subspace $\mathcal{R}_{\mathcal{M}}$ contains no member of the positive semidefinite cone $\mathbb{S}_+^M$ whose rank exceeds $\rho$.

A good example of such a tangent subspace is given in §E.7.2.0.2 by (1768); $\mathcal{R}_{\mathcal{M}}(\mathbf{1}\mathbf{1}^T) = \mathbb{S}_c^{M\perp}$, orthogonal complement to the *geometric center subspace.* (Figure **105**, p.424)

#### 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. [179, §6] [158, §2.3.4] Because each and every face of the positive semidefinite cone contains the origin (§2.8.0.0.1), each face belongs to a subspace of the same dimension.

Given positive semidefinite matrix $A \in \mathbb{S}_+^M$, define $\mathcal{F}\big(\mathbb{S}_+^M \ni A\big)$ (138) as the smallest face that contains $A$ of the positive semidefinite cone $\mathbb{S}_+^M$. Then $A$, having ordered diagonalization $Q\Lambda Q^T$ (§A.5.2), is relatively interior to [20, §II.12] [77, §31.5.3] [175, §2.4]

$$\begin{aligned}\mathcal{F}\big(\mathbb{S}_+^M \ni A\big) &= \{X \in \mathbb{S}_+^M \mid \mathcal{N}(X) \supseteq \mathcal{N}(A)\}\\ &= \{X \in \mathbb{S}_+^M \mid \langle Q(I - \Lambda\Lambda^\dagger)Q^T,\, X\rangle = 0\}\\ &\simeq \mathbb{S}_+^{\text{rank}\, A}\end{aligned} \tag{190}$$

which is isomorphic with the convex cone $\mathbb{S}_+^{\operatorname{rank} A}$. Thus dimension of the smallest face containing given matrix $A$ is

$$\dim \mathcal{F}\big(\mathbb{S}_+^M \ni A\big) = \operatorname{rank}(A)\big(\operatorname{rank}(A)+1\big)/2 \tag{191}$$

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.

#### 2.9.2.3.1 Table: Rank $k$ *versus* dimension of $\mathbb{S}_+^3$ faces

| | $k$ | $\dim \mathcal{F}(\mathbb{S}_+^3 \ni \text{rank-}k \text{ matrix})$ |
|---|---|---|
| | 0 | 0 |
| boundary | 1 | 1 |
| | 2 | 3 |
| $\overline{\text{interior}}$ | 3 | 6 |

For the positive semidefinite cone $\mathbb{S}_+^2$ in isometrically isomorphic $\mathbb{R}^3$ depicted in Figure **31**, for example, rank-2 matrices belong to the interior of the face having dimension 3 (the entire closed cone), while rank-1 matrices belong to the relative interior of a face having dimension 1 (the boundary constitutes all the one-dimensional faces, in this dimension, which are rays emanating from the origin), and the only rank-0 matrix is the point at the origin (the zero-dimensional face).

Any *simultaneously diagonalizable* positive semidefinite rank-$k$ matrices belong to the same face (190). That observation leads to the following hyperplane characterization of PSD cone faces: Any rank-$k<M$ positive semidefinite matrix $A$ belongs to a face, of the positive semidefinite cone, described by an intersection with a hyperplane: for $A=Q\Lambda Q^T$ and $0\leq k<M$

$$\begin{aligned}
\mathcal{F}\big(\mathbb{S}_+^M \ni A \ni \operatorname{rank}(A)=k\big) &= \{X\in\mathbb{S}_+^M \mid \langle Q(I-\Lambda\Lambda^\dagger)Q^T,\, X\rangle = 0\} \\
&= \left\{X\in\mathbb{S}_+^M \;\middle|\; \left\langle Q\left(I-\begin{bmatrix} I\in\mathbb{S}^k & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}\end{bmatrix}\right)Q^T,\, X\right\rangle = 0\right\} \\
&= \mathbb{S}_+^M \cap \underline{\partial\mathcal{H}}_+
\end{aligned} \tag{192}$$

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 $\operatorname{svec}\big(Q(:,k+1:M)Q(:,k+1:M)^T\big)$ to a supporting hyperplane $\underline{\partial\mathcal{H}_+}$ (containing the origin) exposing (§2.11) a face of the positive semidefinite cone containing only rank-$k$ matrices.

#### 2.9.2.4 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 $M$; *id est*, because of the *cone faces lemma* (§2.8.0.0.1) and the direct correspondence of exposed faces to faces of $\mathbb{S}_+^M$, 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 $M>0$

$$\{yy^T\in\mathbb{S}^M \mid y\in\mathbb{R}^M\} \subset \partial\mathbb{S}_+^M \tag{193}$$

this superset (*confer* (156)) of extreme directions for the positive semidefinite cone is, generally, a subset of the boundary. For two-dimensional matrices, (Figure **31**)

$$\{yy^T\in\mathbb{S}^{\mathbf{2}} \mid y\in\mathbb{R}^{\mathbf{2}}\} = \partial\mathbb{S}_+^{\mathbf{2}} \tag{194}$$

while for one-dimensional matrices, in exception, (§2.7)

$$\{yy^T\in\mathbb{S} \mid y\neq\mathbf{0}\} = \operatorname{int}\mathbb{S}_+ \tag{195}$$

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; *videlicet*,[2.32]

$$\measuredangle(yy^T,\, I) = \arccos\frac{\langle yy^T,\, I\rangle}{\|yy^T\|_{\mathrm F}\,\|I\|_{\mathrm F}} = \arccos\left(\frac{1}{\sqrt{M}}\right) \quad \forall\, y\in\mathbb{R}^M \tag{196}$$

[2.32] Analogy with respect to the EDM cone is considered by Hayden & Wells *et alii* [134, p.162] where it is found: angle is not constant. The extreme directions of the EDM cone can be found in §6.5.3.1 while the cone axis is $-E=\mathbf{1}\mathbf{1}^T-I$ (902).

**2.9.2.4.1 Example.** *Positive semidefinite matrix from extreme directions.* Diagonalizability (§A.5) of symmetric matrices yields the following results:

Any symmetric positive semidefinite matrix (1253) can be written in the form

$$A = \sum_i \lambda_i z_i z_i^T = \hat{A}\hat{A}^T = \sum_i \hat{a}_i \hat{a}_i^T \succeq 0\,, \qquad \lambda \succeq 0 \tag{197}$$

a conic combination of linearly independent extreme directions $\big(\hat{a}_i \hat{a}_i^T$ or $z_i z_i^T$ where $\|z_i\| = 1\big)$, where $\lambda$ is a vector of eigenvalues.

If we limit consideration to all symmetric positive semidefinite matrices bounded such that $\operatorname{tr} A = 1$

$$\mathcal{C} \triangleq \{A \succeq 0 \mid \operatorname{tr} A = 1\} \tag{198}$$

then any matrix $A$ from that set may be expressed as a convex combination of linearly independent extreme directions;

$$A = \sum_i \lambda_i z_i z_i^T \in \mathcal{C}\;, \qquad \mathbf{1}^T \lambda = 1\,, \quad \lambda \succeq 0 \tag{199}$$

Implications are:

1. set $\mathcal{C}$ is convex, (it is an intersection of PSD cone with hyperplane)
2. because the set of eigenvalues corresponding to a given square matrix $A$ is unique, no single eigenvalue can exceed $1$; *id est*, $I \succeq A$.

Set $\mathcal{C}$ is an instance of Fantope (80). □

#### 2.9.2.5 Positive semidefinite cone is generally not circular

Extreme angle equation (196) 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:

**2.9.2.5.1 Definition.** *Circular cone*:[2.33]
a pointed closed convex cone having hyperspherical sections orthogonal to its *axis of revolution* about which the cone is invariant to rotation. △

[2.33] A circular cone is assumed convex throughout, although not so by other authors. We also assume a *right* circular cone.

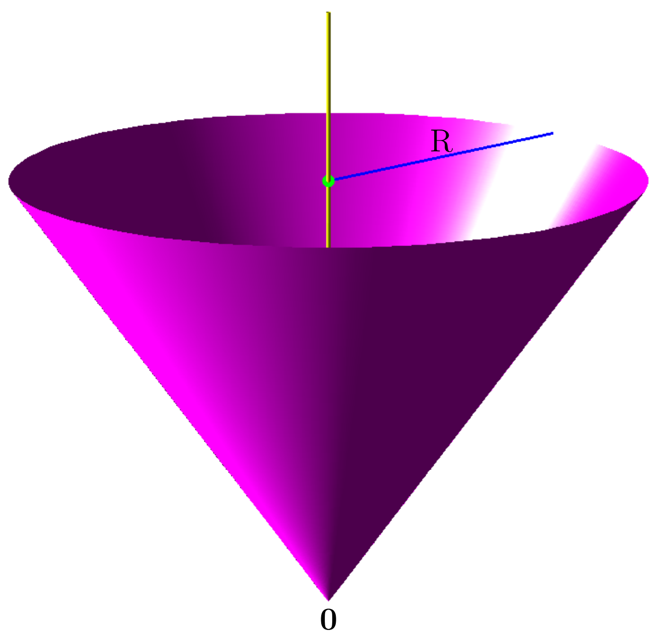



Figure 34: This circular cone continues upward infinitely. Axis of revolution is illustrated as vertical line segment through origin. R is the radius, the distance measured from any extreme direction to axis of revolution. Were this a Lorentz cone, any plane slice containing the axis of revolution would make a right angle.

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 **34**) A prominent example of a circular cone in convex analysis is the Lorentz cone (147). 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 $\mathbf{0}$) through the identity matrix $I$. Consider the set of normalized extreme directions of the positive semidefinite cone: for some arbitrary positive constant $a\in\mathbb{R}_+$

$$\{yy^T\in\mathbb{S}^M \mid \|y\| = \sqrt{a}\} \subset \partial\mathbb{S}_+^M \tag{200}$$

The distance from each extreme direction to the axis of revolution is the radius

$$\mathrm{R} \triangleq \inf_c \|yy^T - cI\|_{\mathrm{F}} = a\sqrt{1-\frac{1}{M}} \tag{201}$$

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 2.9.2.4.1, the convex hull (excluding the vertex at the origin) of the normalized extreme directions is a conic section

$$\mathcal{C} \triangleq \operatorname{conv}\{yy^T \mid y\in\mathbb{R}^M,\; y^Ty=a\} = \mathbb{S}_+^M \cap \{A\in\mathbb{S}^M \mid \langle I\,,\,A\rangle = a\} \tag{202}$$

orthogonal to the identity matrix $I$;

$$\langle \mathcal{C}-\frac{a}{M}I\;,\;I\rangle = \operatorname{tr}(\mathcal{C}-\frac{a}{M}I) = 0 \tag{203}$$

Although the positive semidefinite cone possesses some characteristics of a circular cone, we can prove it is not by demonstrating a 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 **35**, [288, §1-7]

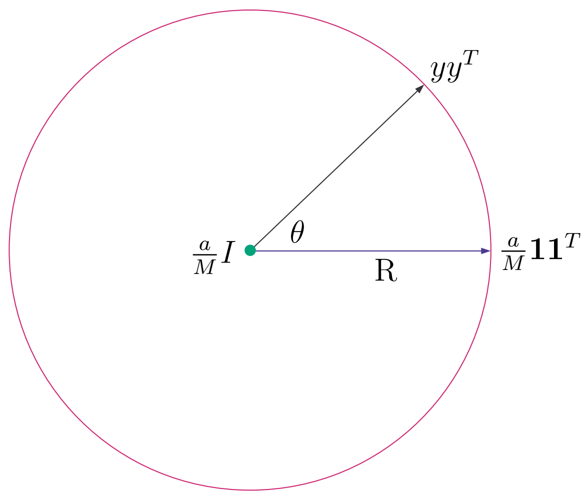



Figure 35: Illustrated is a section, perpendicular to axis of revolution, of circular cone from Figure **34**. Radius R is distance from any extreme direction to axis at $\frac{a}{M}I$. Vector $\frac{a}{M}\mathbf{1}\mathbf{1}^T$ is an arbitrary reference by which to measure angle $\theta$.

$$\cos\theta = \frac{\left\langle \frac{a}{M}\mathbf{1}\mathbf{1}^T - \frac{a}{M}I \ , \ yy^T - \frac{a}{M}I \right\rangle}{a^2(1-\frac{1}{M})} \tag{204}$$

Solving for vector $y$ we get

$$a\big(1 + (M-1)\cos\theta\big) = (\mathbf{1}^T y)^2 \tag{205}$$

Because this does not have real solution for every matrix dimension $M$ and for all $0 \leq \theta \leq 2\pi$, then we can conclude that the positive semidefinite cone might be circular but only in matrix dimensions 1 and 2 .[2.34] ♦

Because of a shortage of extreme directions, conic section (202) cannot be hyperspherical by the *extremes theorem* (§2.8.1.1.1).

[2.34] In fact, the positive semidefinite cone is circular in matrix dimensions 1 and 2 while it is a rotation of the Lorentz cone in matrix dimension 2.

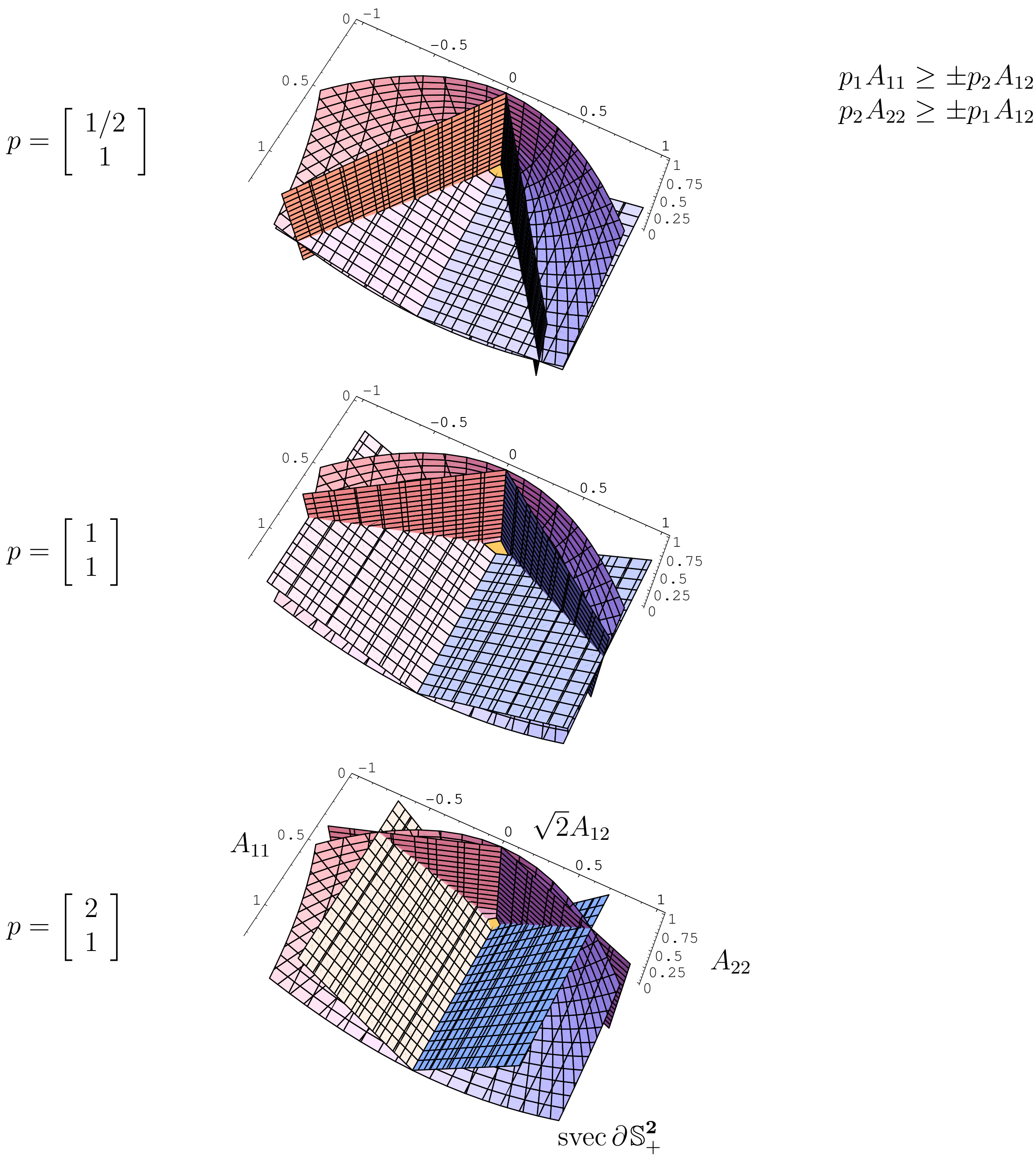



Figure 36: Polyhedral proper cone $\mathcal{K}$ , created by intersection of halfspaces, inscribes PSD cone in isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$ as predicted by *Geršgorin discs theorem* for $A = [A_{ij}] \in \mathbb{S}^{\mathbf{2}}$. Hyperplanes supporting $\mathcal{K}$ intersect along boundary of PSD cone. Four extreme directions of $\mathcal{K}$ coincide with extreme directions of PSD cone.

**2.9.2.5.2 Example.** *PSD cone inscription in three dimensions.*

**Theorem.** *Geršgorin discs.* [150, §6.1] [274]
For $p \in \mathbb{R}_+^m$ given $A = [A_{ij}] \in \mathbb{S}^m$, then all eigenvalues of $A$ belong to the union of $m$ closed intervals on the real line;

$$\lambda(A) \in \bigcup_{i=1}^{m} \left\{ \xi \in \mathbb{R} \;\middle|\; |\xi - A_{ii}| \leq \varrho_i \triangleq \frac{1}{p_i} \sum_{\substack{j=1 \\ j \neq i}}^{m} p_j |A_{ij}| \right\} = \bigcup_{i=1}^{m} [A_{ii} - \varrho_i \;,\; A_{ii} + \varrho_i] \tag{206}$$

*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

$$A_{ii} \geq \varrho_i \tag{207}$$

Suppose

$$m = 2 \tag{208}$$

so $A \in \mathbb{S}^{\mathbf{2}}$. Vectorizing $A$ as in (47), $\operatorname{svec} A$ belongs to isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$. Then we have $m2^{m-1} = 4$ inequalities, in the matrix entries $A_{ij}$ with Geršgorin parameters $p = [p_i] \in \mathbb{R}_+^{\mathbf{2}}$,

$$\begin{aligned} p_1 A_{11} &\geq \pm p_2 A_{12} \\ p_2 A_{22} &\geq \pm p_1 A_{12} \end{aligned} \tag{209}$$

which describe an intersection of four halfspaces in $\mathbb{R}^{m(m+1)/2}$. That intersection creates the polyhedral proper cone $\mathcal{K}$ (§2.12.1) whose construction is illustrated in Figure **36**. Drawn truncated is the boundary of the positive semidefinite cone $\operatorname{svec} \mathbb{S}_+^{\mathbf{2}}$ and the bounding hyperplanes supporting $\mathcal{K}$.

Created by means of Geršgorin discs, $\mathcal{K}$ always belongs to the positive semidefinite cone for any nonnegative value of $p \in \mathbb{R}_+^m$. Hence any point in $\mathcal{K}$ corresponds to some positive semidefinite matrix $A$. Only the extreme directions of $\mathcal{K}$ intersect the positive semidefinite cone boundary in this dimension; the four extreme directions of $\mathcal{K}$ are extreme directions of the positive semidefinite cone. As $p_1/p_2$ increases in value from 0, two extreme directions of $\mathcal{K}$ sweep the entire boundary of this positive semidefinite cone. Because the entire positive semidefinite cone can be swept by $\mathcal{K}$ , the system of linear inequalities

$$Y^T\text{svec}\, A \triangleq \begin{bmatrix} p_1 & \pm p_2/\sqrt{2} & 0 \\ 0 & \pm p_1/\sqrt{2} & p_2 \end{bmatrix} \text{svec}\, A \succeq 0 \tag{210}$$

when made dynamic can replace a semidefinite constraint $A \succeq 0$ ; *id est*, for

$$\mathcal{K} = \{z \mid Y^T z \succeq 0\} \subset \text{svec}\, \mathbb{S}^m_+ \tag{211}$$

given $p$ where $Y \in \mathbb{R}^{m(m+1)/2 \times m2^{m-1}}$

$$\text{svec}\, A \in \mathcal{K} \;\Rightarrow\; A \in \mathbb{S}^m_+ \tag{212}$$

but

$$\exists p \;\ni\; Y^T\text{svec}\, A \succeq 0 \;\Leftrightarrow\; A \succeq 0 \tag{213}$$

In other words, *diagonal dominance* [150, p.349, §7.2.3]

$$A_{ii} \geq \sum_{\substack{j=1 \\ j \neq i}}^{m} |A_{ij}| \;, \quad \forall\, i = 1 \ldots m \tag{214}$$

is only a sufficient condition for membership to the PSD cone; but by dynamic weighting $p$ in this dimension, it was made necessary and sufficient. □

In higher dimension ($m > 2$), the boundary of the positive semidefinite cone is no longer constituted completely by its extreme directions (symmetric rank-one matrices); the geometry becomes complicated. How all the extreme directions can be swept by an inscribed polyhedral cone,[2.35] similarly to the foregoing example, remains an open question.

**2.9.2.5.3 Exercise.** *Dual inscription.*
Find dual polyhedral proper cone $\mathcal{K}^*$ from Figure 36. ▼

[2.35] It is not necessary to sweep the entire boundary in higher dimension.

#### 2.9.2.6 Boundary constituents of the positive semidefinite cone

**2.9.2.6.1 Lemma.** *Sum of positive semidefinite matrices.*
For $A, B \in \mathbb{S}_+^M$

$$\operatorname{rank}(A+B) = \operatorname{rank}(\mu A + (1-\mu)B) \tag{215}$$

over the open interval $(0,1)$ of $\mu$ . ◇

**Proof.** Any positive semidefinite matrix belonging to the PSD cone has an eigen decomposition that is a positively scaled sum of linearly independent symmetric dyads. By the *linearly independent dyads definition* in §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 $A+B$ . ♦

**2.9.2.6.2 Example.** *Rank function quasiconcavity.* (*confer* §3.3)
For $A, B \in \mathbb{R}^{m\times n}$ [150, §0.4]

$$\operatorname{rank} A + \operatorname{rank} B \;\geq\; \operatorname{rank}(A+B) \tag{216}$$

that follows from the fact [249, §3.6]

$$\dim \mathcal{R}(A) + \dim \mathcal{R}(B) \;=\; \dim \mathcal{R}(A+B) + \dim\big(\mathcal{R}(A) \cap \mathcal{R}(B)\big) \tag{217}$$

For $A, B \in \mathbb{S}_+^M$ [46, §3.4.2]

$$\operatorname{rank} A + \operatorname{rank} B \;\geq\; \operatorname{rank}(A+B) \;\geq\; \min\{\operatorname{rank} A\,,\ \operatorname{rank} B\} \tag{218}$$

that follows from the fact

$$\mathcal{N}(A+B) = \mathcal{N}(A) \cap \mathcal{N}(B)\;, \qquad A, B \in \mathbb{S}_+^M \qquad (133)$$

Rank is a *quasiconcave* function on $\mathbb{S}_+^M$ because the right-hand inequality in (218) has the concave form (539); *videlicet*, Lemma 2.9.2.6.1. □

From this example we see, unlike convex functions, *quasiconvex* functions are not necessarily continuous. (§3.3) We also glean:

**2.9.2.6.3 Theorem.** *Convex subsets of positive semidefinite cone.*
The subsets of the positive semidefinite cone $\mathbb{S}_+^M$ , for $0 \leq \rho \leq M$

$$\mathbb{S}_+^M(\rho) \triangleq \{X \in \mathbb{S}_+^M \mid \operatorname{rank} X \geq \rho\} \tag{219}$$

are pointed convex cones, but not closed unless $\rho = 0$ ; *id est*, $\mathbb{S}_+^M(0) = \mathbb{S}_+^M$ .
$\diamond$

**Proof.** Given $\rho$ , a subset $\mathbb{S}_+^M(\rho)$ is convex if and only if convex combination of any two members has rank at least $\rho$ . That is confirmed applying identity (215) from Lemma 2.9.2.6.1 to (218); *id est*, for $A, B \in \mathbb{S}_+^M(\rho)$ on the closed interval $\mu \in [0,1]$

$$\operatorname{rank}(\mu A + (1-\mu)B) \;\geq\; \min\{\operatorname{rank} A\,,\ \operatorname{rank} B\} \tag{220}$$

It can similarly be shown, almost identically to proof of the lemma, any conic combination of $A, B$ in subset $\mathbb{S}_+^M(\rho)$ remains a member; *id est*, $\forall\, \zeta\,, \xi \geq 0$

$$\operatorname{rank}(\zeta A + \xi B) \;\geq\; \min\{\operatorname{rank}(\zeta A)\,,\ \operatorname{rank}(\xi B)\} \tag{221}$$

Therefore, $\mathbb{S}_+^M(\rho)$ is a convex cone. $\blacklozenge$

Another proof of convexity can be made by projection arguments:

### 2.9.2.7 Projection on $\mathbb{S}_+^M(\rho)$

Because these cones $\mathbb{S}_+^M(\rho)$ indexed by $\rho$ (219) are convex, projection on them is straightforward. Given a symmetric matrix $H$ having diagonalization $H \triangleq Q\Lambda Q^T \in \mathbb{S}^M$ (§A.5.2) with eigenvalues $\Lambda$ arranged in nonincreasing order, then its *Euclidean projection* (minimum-distance projection) on $\mathbb{S}_+^M(\rho)$

$$P_{\mathbb{S}_+^M(\rho)} H = Q\,\Upsilon^\star Q^T \tag{222}$$

corresponds to a map of its eigenvalues:

$$\Upsilon_{ii}^\star = \begin{cases} \max\{\epsilon\,,\ \Lambda_{ii}\}\ , & i = 1 \ldots \rho \\ \max\{0\,,\ \Lambda_{ii}\}\ , & i = \rho+1 \ldots M \end{cases} \tag{223}$$

where $\epsilon$ is positive but arbitrarily close to 0.

**2.9.2.7.1 Exercise.** *Projection on open convex cones.*
Prove (223) using Theorem E.9.2.0.1. ▼

Because each $H\in\mathbb{S}^M$ has unique projection on $\mathbb{S}_+^M(\rho)$ (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 (223) to the well-known result regarding Euclidean projection on a rank $\rho$ subset of the positive semidefinite cone (§2.9.2.1)

$$\mathbb{S}_+^M\setminus\mathbb{S}_+^M(\rho+1) \;=\; \{X\in\mathbb{S}_+^M \mid \operatorname{rank}X\leq\rho\} \tag{224}$$

$$P_{\mathbb{S}_+^M\setminus\mathbb{S}_+^M(\rho+1)}H = Q\Upsilon^\star Q^T \tag{225}$$

As proved in §7.1.4, this projection of $H$ corresponds to the eigenvalue map

$$\Upsilon^\star_{ii} = \left\{\begin{array}{ll}\max\{0\,,\,\Lambda_{ii}\}\;, & i=1\ldots\rho\\ 0\;, & i=\rho+1\ldots M\end{array}\right. \tag{1147}$$

Together these two results (223) and (1147) mean: A higher-rank solution to projection on the positive semidefinite cone lies arbitrarily close to any given lower-rank projection, but not *vice versa.* Were the number of nonnegative eigenvalues in $\Lambda$ known *a priori* not to exceed $\rho$, then these two different projections would produce identical results in the limit $\epsilon\to0$.

### 2.9.2.8 Uniting constituents

Interior of the PSD cone $\operatorname{int}\mathbb{S}_+^M$ is convex by Theorem 2.9.2.6.3, 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 $0<\rho<M$

$$\{Y\in\mathbb{S}_+^M \mid \operatorname{rank}Y=\rho\} \tag{226}$$

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)

$$\partial\mathbb{S}_+^M = \bigcup_{\rho=0}^{M-1}\{Y\in\mathbb{S}_+^M \mid \operatorname{rank}Y=\rho\} \tag{227}$$

The composite sequence, the cone interior in union with each successive constituent, remains convex at each step; *id est*, for $0 \leq k \leq M$

$$\bigcup_{\rho=k}^{M} \{Y \in \mathbb{S}_+^M \mid \operatorname{rank} Y = \rho\} \tag{228}$$

is convex for each $k$ by Theorem 2.9.2.6.3.

#### 2.9.2.9 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.

**2.9.2.9.1 Exercise.** *Difference* $A-B$ .
What about the difference of matrices $A,B$ belonging to the positive semidefinite cone? Show:

- The difference of any two points on the boundary belongs to the boundary or exterior.
- The 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.** (Barvinok) *Affine intersection with PSD cone.* [20, §II.13] [21, §2.2] Consider finding a matrix $X \in \mathbb{S}^N$ satisfying

$$X \succeq 0\ , \qquad \langle A_j\ ,\ X\rangle = b_j\ , \quad j = 1 \ldots m \tag{229}$$

given nonzero linearly independent $A_j \in \mathbb{S}^N$ and real $b_j$. Define the affine subset

$$\mathcal{A} \triangleq \{X \mid \langle A_j\,,\ X\rangle = b_j\ ,\ j=1\ldots m\} \subseteq \mathbb{S}^N \tag{230}$$

If the intersection $\mathcal{A} \cap \mathbb{S}^N_+$ is nonempty, then there exists a matrix $X \in \mathcal{A} \cap \mathbb{S}^N_+$ such that given a number of equalities $m$

$$\operatorname{rank} X\,(\operatorname{rank} X + 1)/2 \leq m \tag{231}$$

whence the upper bound[2.36]

$$\operatorname{rank} X \leq \left\lfloor \frac{\sqrt{8m+1}-1}{2} \right\rfloor \tag{232}$$

or given desired rank instead, equivalently,

$$m < (\operatorname{rank} X + 1)(\operatorname{rank} X + 2)/2 \tag{233}$$

An extreme point of $\mathcal{A} \cap \mathbb{S}^N_+$ satisfies (232) and (233). (*confer* §4.1.1.2) 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; [175, §2.4] *id est*, iff (138)

$$\begin{aligned}&\dim \mathcal{F}\big((\mathcal{A} \cap \mathbb{S}^N_+) \ni X\big) \qquad (234)\\ =\;& \operatorname{rank}(X)\big(\operatorname{rank}(X)+1\big)/2 - \operatorname{rank}\big[\operatorname{svec} RA_1R^T\ \ \operatorname{svec} RA_2R^T \cdots \operatorname{svec} RA_mR^T\,\big]\end{aligned}$$

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

$$m = (\rho+1)(\rho+2)/2 \tag{235}$$

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

$$\operatorname{rank} X \leq \rho \tag{236}$$

This represents a *tightening* of the upper bound; a reduction by exactly 1 of the bound provided by (232) given the same specified number $m$ (235) of equalities; *id est*,

$$\operatorname{rank} X \leq \frac{\sqrt{8m+1}-1}{2} - 1 \tag{237}$$

◇

[2.36] §4.1.1.2 contains an intuitive explanation. This bound is itself limited above, of course, by $N$; a tight limit corresponding to an interior point of $\mathbb{S}^N_+$.

When the intersection $\mathcal{A} \cap \mathbb{S}^N_+$ is known *a priori* to consist only of a single point, then Barvinok's proposition provides the greatest upper bound on its rank not exceeding $N$. The intersection can be a single nonzero point only if the number of linearly independent hyperplanes $m$ constituting $\mathcal{A}$ satisfies[2.37]

$$N(N+1)/2 - 1 \le m \le N(N+1)/2 \tag{238}$$

## 2.10 Conic independence (c.i.)

In contrast to extreme direction, the property *conically independent direction* is more generally applicable, inclusive of all closed convex cones (not only pointed closed convex cones). Similar to the definition for linear independence, arbitrary given directions $\{\Gamma_i \in \mathbb{R}^n,\ i=1 \ldots N\}$ are *conically independent* if and only if, for all $\zeta \in \mathbb{R}^N_+$

$$\Gamma_i\,\zeta_i + \cdots + \Gamma_j\,\zeta_j - \Gamma_\ell\,\zeta_\ell = \mathbf{0}\,, \qquad i \neq \cdots \neq j \neq \ell = 1 \ldots N \tag{239}$$

has only the trivial solution $\zeta = \mathbf{0}$; in words, iff no direction from the given set can be expressed as a conic combination of those remaining. (Figure **37**, for example. A Matlab implementation of test (239) is given in §F.2.) It is evident that linear independence (l.i.) of $N$ directions implies their conic independence;

- l.i. $\Rightarrow$ c.i.

Arranging any set of generators for a particular convex cone in a matrix columnar,

$$X \triangleq [\,\Gamma_1 \;\; \Gamma_2 \;\; \cdots \;\; \Gamma_N\,] \in \mathbb{R}^{n \times N} \tag{240}$$

then the relationship l.i. $\Rightarrow$ c.i. suggests: the number of l.i. generators in the columns of $X$ cannot exceed the number of c.i. generators. Denoting by $\mathbf{k}$ the number of conically independent generators contained in $X$, we have the most fundamental rank inequality for convex cones

$$\dim \operatorname{aff} \mathcal{K} \;=\; \dim \operatorname{aff}[\,\mathbf{0} \;\; X\,] \;=\; \operatorname{rank} X \le \mathbf{k} \le N \tag{241}$$

Whereas $N$ directions in $n$ dimensions can no longer be linearly independent once $N$ exceeds $n$, conic independence remains possible:

[2.37] For $N>1$, $N(N+1)/2-1$ independent hyperplanes in $\mathbb{R}^{N(N+1)/2}$ can make a line

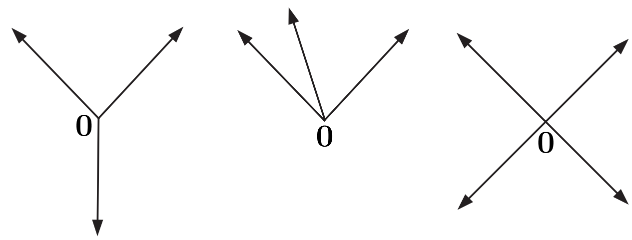



Figure 37: Vectors in $\mathbb{R}^2$: **(a)** affinely and conically independent, **(b)** affinely independent but not conically independent, **(c)** conically independent but not affinely independent. None of the examples exhibits linear independence. (In general, a.i. $\not\Leftrightarrow$ c.i.)

#### 2.10.0.0.1 Table: Maximum number of c.i. directions

| $n$ | $\sup \mathbf{k}$ (pointed) | $\sup \mathbf{k}$ (not pointed) |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 1 | 2 |
| 2 | 2 | 4 |
| 3 | $\infty$ | $\infty$ |
| $\vdots$ | $\vdots$ | $\vdots$ |

Assuming veracity of this table, there is an apparent vastness between two and three dimensions. The finite numbers of conically independent directions indicate:

- 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. [20, p.vii]

From this table it is also evident that dimension of Euclidean space cannot exceed the number of conically independent directions possible;

- $n \leq \sup \mathbf{k}$

tangent to $\operatorname{svec} \partial \mathbb{S}_+^N$ at a point because all one-dimensional faces of $\mathbb{S}_+^N$ are exposed. Because a pointed convex cone has only one vertex, the origin, there can be no intersection of $\operatorname{svec} \partial \mathbb{S}_+^N$ with any higher-dimensional affine subset $\mathcal{A}$ that will make a nonzero point.

**2.10.0.0.2 Exercise.** *Conically independent columns and rows.*
We suspect the number of conically independent columns (rows) of $X$ to be the same for $X^{\dagger T}$, where $^\dagger$ denotes matrix pseudoinverse (§E). Prove whether it holds that the columns (rows) of $X$ are c.i. $\Leftrightarrow$ the columns (rows) of $X^{\dagger T}$ are c.i. ▼

## 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 $X \in \mathbb{R}^{n\times N}$ (240) holds a conically independent set columnar. Consider the transformation

$$T(X) : \mathbb{R}^{n\times N} \to \mathbb{R}^{n\times N} \triangleq XY \tag{242}$$

where the given matrix $Y \triangleq [\, y_1 \;\; y_2 \; \cdots \; y_N \,] \in \mathbb{R}^{N\times N}$ is represented by linear operator $T$. Conic independence of $\{Xy_i \in \mathbb{R}^n,\; i=1\ldots N\}$ demands, by definition (239),

$$Xy_i\,\zeta_i + \cdots + Xy_j\,\zeta_j - Xy_\ell\,\zeta_\ell = \mathbf{0}\,, \qquad i \neq \cdots \neq j \neq \ell = 1\ldots N \tag{243}$$

have no nontrivial solution $\zeta \in \mathbb{R}^N_+$. That is ensured by conic independence of $\{y_i \in \mathbb{R}^N\}$ and by $\mathcal{R}(Y) \cap \mathcal{N}(X) = \mathbf{0}$; seen by factoring $X$.

### 2.10.1.1 linear maps of cones

[18, §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 closed or has nonempty interior in $\mathcal{R}$, then so is $T(\mathcal{K})$ in $\mathcal{M}$.

If $T$ is a linear bijection, then $x \preceq y \;\Leftrightarrow\; T(x) \preceq T(y)$. Further, if $\mathcal{F}$ is a face of $\mathcal{K}$, then $T(\mathcal{F})$ is a face of $T(\mathcal{K})$.

## 2.10.2 Pointed closed convex $\mathcal{K}$ & conic independence

The following bullets can be derived from definitions (155) and (239) 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 **15** on page 60 with Figure **38**(a))

(a)

(b)

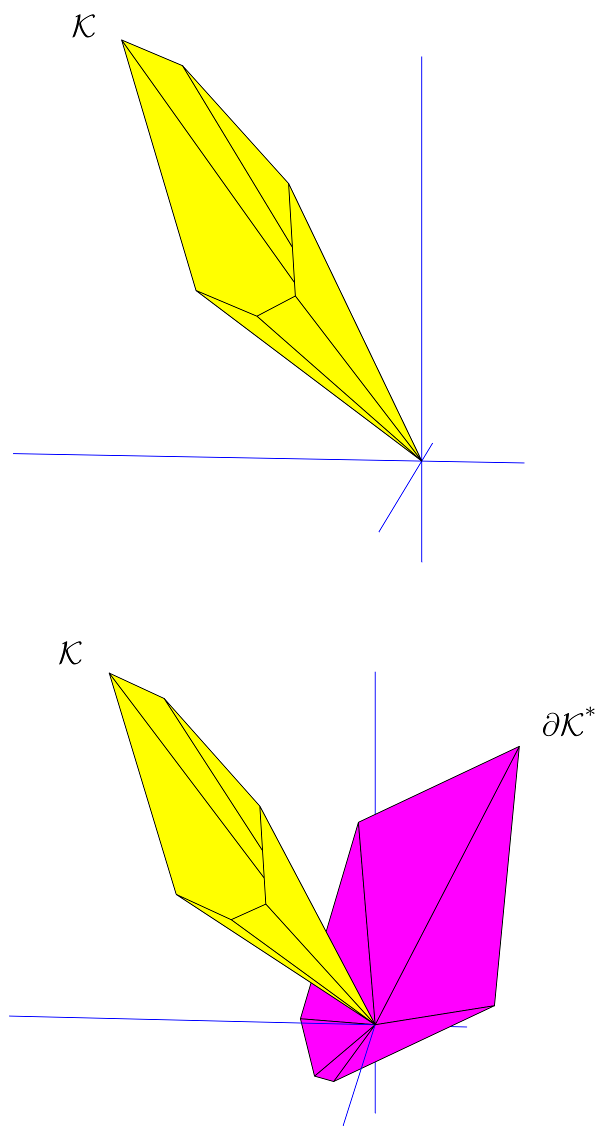



Figure 38: **(a)** A pointed polyhedral cone (drawn truncated) in $\mathbb{R}^{\mathbf{3}}$ having six facets. The extreme directions, corresponding to six edges emanating from the origin, are generators for this cone; not linearly independent but they must be conically independent. **(b)** The boundary 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.

- $\{$extreme directions$\} \Rightarrow \{$c.i.$\}$

Conversely, when a conically independent set of directions from pointed closed convex cone $\mathcal{K}$ is known *a priori* to comprise generators, then all directions from that set must be extreme directions of the cone;

- $\{$extreme directions$\} \Leftrightarrow \{$c.i. generators of pointed closed convex $\mathcal{K}\}$

Barker & Carlson [18, §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*,[2.38] for a pointed closed convex cone.

Any collection of $n$ or fewer extreme directions from pointed closed convex cone $\mathcal{K} \subset \mathbb{R}^n$ must be linearly independent;

- $\{\leq n$ extreme directions in $\mathbb{R}^n\} \Rightarrow \{$l.i.$\}$

Conversely, because l.i. $\Rightarrow$ c.i.,

- $\{$extreme directions$\} \Leftarrow \{$l.i. generators of pointed closed convex $\mathcal{K}\}$

**2.10.2.0.1 Example.** *Vertex-description of halfspace $\mathcal{H}$ about origin.*
From $n+1$ points in $\mathbb{R}^n$ we can make a vertex-description of a convex cone that is a halfspace $\mathcal{H}$, where $\{x_\ell \in \mathbb{R}^n,\ \ell = 1 \ldots n\}$ constitutes a minimal set of generators for a hyperplane $\partial\mathcal{H}$ through the origin. An example is illustrated in Figure 39. By demanding the augmented set $\{x_\ell \in \mathbb{R}^n,\ \ell = 1 \ldots n+1\}$ be affinely independent (we want $x_{n+1}$ not parallel to $\partial\mathcal{H}$), then

$$\begin{aligned}
\mathcal{H} &= \bigcup_{\zeta \geq 0} (\zeta\, x_{n+1} + \partial\mathcal{H}) \\
&= \{\zeta\, x_{n+1} + \operatorname{cone}\{x_\ell \in \mathbb{R}^n,\ \ell = 1 \ldots n\} \mid \zeta \geq 0\} \qquad (244) \\
&= \operatorname{cone}\{x_\ell \in \mathbb{R}^n,\ \ell = 1 \ldots n+1\}
\end{aligned}$$

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 $\mathcal{H}$ which has no extreme elements. □

[2.38] 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 (case $n = 2$).

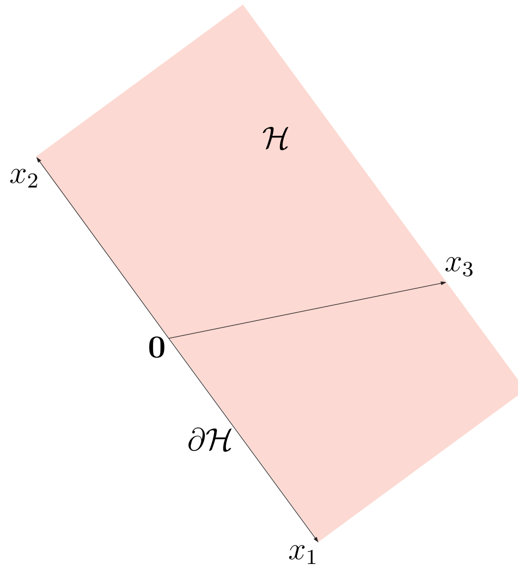



Figure 39: Minimal set of generators $X = [\,x_1\ x_2\ x_3\,] \in \mathbb{R}^{\mathbf{2\times 3}}$ for halfspace about origin.

### 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 polyhedral set in $\mathbb{R}^n$ having nonempty interior, distinction between the terms *extreme* and *exposed* vanishes [247, §2.4] [77, §2.2] for faces of all dimensions except $n$ ; their meanings become equivalent as we saw in Figure **12** (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 *vice versa*; *e.g.*, Figure **15**.

Lewis [179, §6] [158, §2.3.4] claims nonempty extreme proper subsets and the exposed subsets coincide for $\mathbb{S}^n_+$ ; *id est*, each and every face of the positive semidefinite cone, whose dimension is less than the dimension of the cone, is exposed. A more general discussion of cones having this property can be found in [257]; *e.g.*, the Lorentz cone (147) [17, §II.A].

## 2.12 Convex polyhedra

Every polyhedron, such as the convex hull (75) of a bounded list $X$ , can be expressed as the solution set of a finite system of linear equalities and inequalities, and *vice versa.* [77, §2.2]

**2.12.0.0.1 Definition.** *Convex polyhedra, halfspace-description.*
[46, §2.2.4] A convex polyhedron is the intersection of a finite number of halfspaces and hyperplanes;

$$\mathcal{P} = \{y \mid Ay \succeq b\,,\ Cy = d\} \subseteq \mathbb{R}^n \tag{245}$$

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 having possibly empty interior; *e.g.*, Figure **12**. Convex polyhedra[2.39] are finite-dimensional comprising all affine sets (§2.3.1), polyhedral cones, line segments, rays, halfspaces, convex polygons, *solids* [164, def.104/6, p.343], polychora, *polytopes*,[2.40] *etcetera.*

[2.39] We consider only convex polyhedra throughout, but acknowledge the existence of concave polyhedra. [282, *Kepler-Poinsot Solid*]

[2.40] Some authors distinguish bounded polyhedra via the designation *polytope.* [77, §2.2]

It follows from definition (245) by exposure that each face of a convex polyhedron is a convex polyhedron.

The projection of any polyhedron on a subspace remains a polyhedron. More generally, the image of a polyhedron under any linear transformation is a polyhedron. [20, §I.9]

When $b$ and $d$ in (245) are $\mathbf{0}$, 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* [230, §19]. 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.

**2.12.1.0.1 Definition.** *Polyhedral cone, halfspace-description.*[2.41]
(*confer* (252)) A polyhedral cone is the intersection of a finite number of halfspaces and hyperplanes about the origin;

$$\begin{aligned}
\mathcal{K} &= \{y \mid Ay \succeq 0\,,\ Cy = \mathbf{0}\} \ \subseteq\ \mathbb{R}^n && \text{(a)}\\
&= \{y \mid Ay \succeq 0\,,\ Cy \succeq 0\,,\ Cy \preceq 0\} && \text{(b)}\\
&= \left\{y \mid \begin{bmatrix} A \\ C \\ -C \end{bmatrix} y \succeq 0\right\} && \text{(c)}
\end{aligned} \tag{246}$$

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, possibly has empty interior, and only a finite number of generators (§2.8.1.2), and *vice versa.* (Minkowski/Weyl) [247, §2.8] [230, thm.19.1]

[2.41] Rockafellar [230, §19] proposes affine sets be handled via complementary pairs of affine inequalities; *e.g.*, $Cy \succeq d$ and $Cy \preceq d$.

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 $\mathbb{R}$, polyhedral flavors of the (proper) Lorentz cone (*confer* (147))

$$\mathcal{K}_\ell = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_\ell \leq t \right\} , \qquad \ell = 1 \text{ or } \infty \tag{247}$$

any subspace, and $\mathbb{R}^n$. More examples are illustrated in Figure **38** and Figure **15**.

### 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. [164, def.115/6, p.358] In Figure **12**, each vertex of the polyhedron is located at the intersection of three or more facets, and every edge belongs to precisely two facets [20, §VI.1, p.252]. In Figure **15**, 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. Not all convex polyhedra are bounded, evidently, neither can they all be described by the convex hull of a bounded set of points as we defined it in (75). Hence we propose a universal vertex-description of polyhedra in terms of that same finite-length list $X$ (65):

**2.12.2.0.1 Definition.** *Convex polyhedra, vertex-description.*
(*confer* §2.8.1.1.1) Denote the truncated $a$-vector,

$$a_{i:\ell} = \begin{bmatrix} a_i \\ \vdots \\ a_\ell \end{bmatrix} \tag{248}$$

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

$$\mathcal{P} = \left\{ Xa \mid a_{1:k}^T\mathbf{1} = 1\,,\; a_{m:N} \succeq 0\,,\; \{1\ldots k\} \cup \{m\ldots N\} = \{1\ldots N\} \right\} \tag{249}$$

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. $\triangle$

Coefficient indices in (249) may or may not be overlapping, but all the coefficients are assumed constrained. From (67), (75), and (83), we summarize how the coefficient conditions may be applied;

$$\left.\begin{array}{lcl}\text{affine sets} & \longrightarrow & a_{1:k}^T\mathbf{1}=1 \\ \text{polyhedral cones} & \longrightarrow & a_{m:N}\succeq 0\end{array}\right\} \longleftarrow \text{convex hull } (m\leq k) \tag{250}$$

It is always possible to describe a convex hull in the region of overlapping indices because, for $1\leq m\leq k\leq N$

$$\{a_{m:k} \mid a_{m:k}^T\mathbf{1}=1,\ a_{m:k}\succeq 0\} \subseteq \{a_{m:k} \mid a_{1:k}^T\mathbf{1}=1,\ a_{m:N}\succeq 0\} \tag{251}$$

Members of a generating list are not necessarily vertices of the corresponding polyhedron; certainly true for (75) and (249), some subset of list members reside in the polyhedron's relative interior. Conversely, when boundedness (75) applies, the convex hull of the vertices is a polyhedron identical to the 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 (240), then that cone is described setting $m=1$ and $k=0$ in vertex-description (249):

$$\mathcal{K} = \operatorname{cone}(X) = \{Xa \mid a\succeq 0\} \subseteq \mathbb{R}^n \tag{252}$$

a conic hull, like (83), of $N$ generators.

This vertex description is extensible to an infinite number of generators; which follows from the *extremes theorem* (§2.8.1.1.1) and Example 2.8.1.2.1.

#### 2.12.2.2 Pointedness

[247, §2.10] Assuming all generators constituting the columns of $X\in\mathbb{R}^{n\times N}$ are nonzero, polyhedral cone $\mathcal{K}$ is pointed (§2.7.2.1.2) if and only if there is no nonzero $a\succeq 0$ that solves $Xa=\mathbf{0}$; *id est*, iff $\mathcal{N}(X)\cap\mathbb{R}_+^N=\mathbf{0}$. **(**If $\operatorname{rank} X=n$, then the *dual cone* $\mathcal{K}^*$ is pointed. (268)**)**

A polyhedral proper cone in $\mathbb{R}^n$ must have at least $n$ linearly independent generators, or be the intersection of at least $n$ halfspaces whose partial boundaries have normals that are linearly independent. Otherwise, the cone

$$\mathcal{S} = \{s \mid s \succeq 0\,,\ \mathbf{1}^T s \le 1\}$$

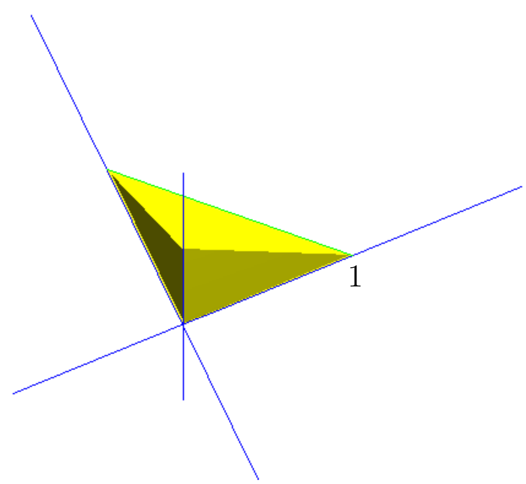



Figure 40: Unit simplex $\mathcal{S}$ in $\mathbb{R}^{\mathbf{3}}$ is a unique solid tetrahedron, but is not *regular.*

will contain at least one line and there can be no vertex; *id est*, the cone cannot otherwise be pointed.

For any pointed polyhedral cone, there is a one-to-one correspondence of one-dimensional faces with extreme directions.

Examples of pointed closed convex cones $\mathcal{K}$ are not limited to polyhedral cones: the origin, any $\mathbf{0}$-based ray in a subspace, any two-dimensional V-shaped cone in a subspace, the Lorentz (ice-cream) cone and its polyhedral flavors, the cone of Euclidean distance matrices $\mathbb{EDM}^N$ in $\mathbb{S}_h^N$, the proper cones: $\mathbb{S}_+^M$ in ambient $\mathbb{S}^M$, any orthant in $\mathbb{R}^n$ or $\mathbb{R}^{m\times n}$; *e.g.*, the nonnegative real line $\mathbb{R}_+$ in vector space $\mathbb{R}$.

### 2.12.3 Unit simplex

A peculiar convex subset of the nonnegative orthant having halfspace-description

$$\mathcal{S} \triangleq \{s \mid s \succeq 0\,,\ \mathbf{1}^T s \leq 1\} \subseteq \mathbb{R}^n_+ \tag{253}$$

is a unique bounded convex polyhedron called *unit simplex* (Figure **40**) having nonempty interior, $n+1$ vertices, and dimension [46, §2.2.4]

$$\dim \mathcal{S} = n \tag{254}$$

The origin supplies one vertex while heads of the *standard basis* [150] [249] $\{e_i\,,\ i=1\ldots n\}$ in $\mathbb{R}^n$ constitute those remaining;[2.42] thus its vertex-description:

$$\begin{aligned} \mathcal{S} &= \operatorname{conv}\{\mathbf{0},\ \{e_i\,,\ i=1\ldots n\}\} \\ &= \left\{[\,\mathbf{0}\ \ e_1\ \ e_2\ \cdots\ e_n\,]\,a \mid a^T\mathbf{1} = 1\,,\ a \succeq 0\right\} \end{aligned} \tag{255}$$

#### 2.12.3.1 Simplex

The unit simplex comes from a class of general polyhedra called *simplex*, having vertex-description: [64] [230] [280] [77]

$$\operatorname{conv}\{x_\ell \in \mathbb{R}^n\} \mid \ell = 1\ldots k+1\,,\quad \dim\operatorname{aff}\{x_\ell\} = k\,,\quad n \geq k \tag{256}$$

So defined, a simplex is a closed bounded convex set having possibly empty interior. Examples of simplices, by increasing affine dimension, are: a point, any line segment, any triangle and its relative interior, a general tetrahedron, polychoron, and so on.

**2.12.3.1.1 Definition.** *Simplicial cone.*
A polyhedral proper (§2.7.2.2.1) cone $\mathcal{K}$ in $\mathbb{R}^n$ is called *simplicial* iff $\mathcal{K}$ has exactly $n$ extreme directions; [17, §II.A] equivalently, iff proper $\mathcal{K}$ has exactly $n$ linearly independent generators contained in any given set of generators. △

There are an infinite variety of simplicial cones in $\mathbb{R}^n$; *e.g.*, Figure **15**, Figure **41**, Figure **50**. Any orthant is simplicial, as is any rotation thereof.

[2.42] 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 **40**), in $\mathbb{R}^4$ it is the convex hull of a pentatope [282], and so on.

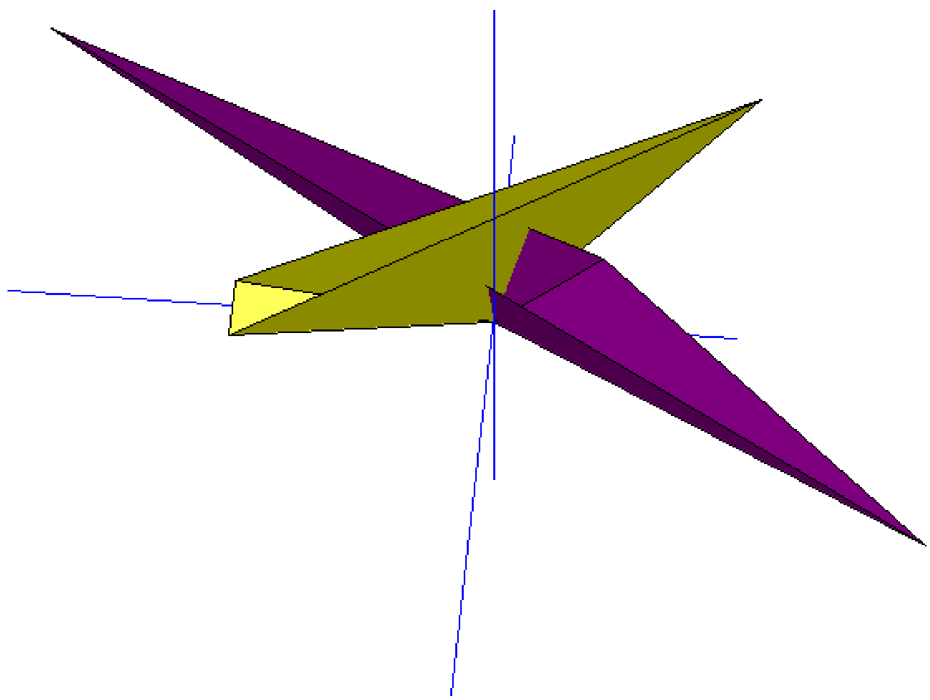
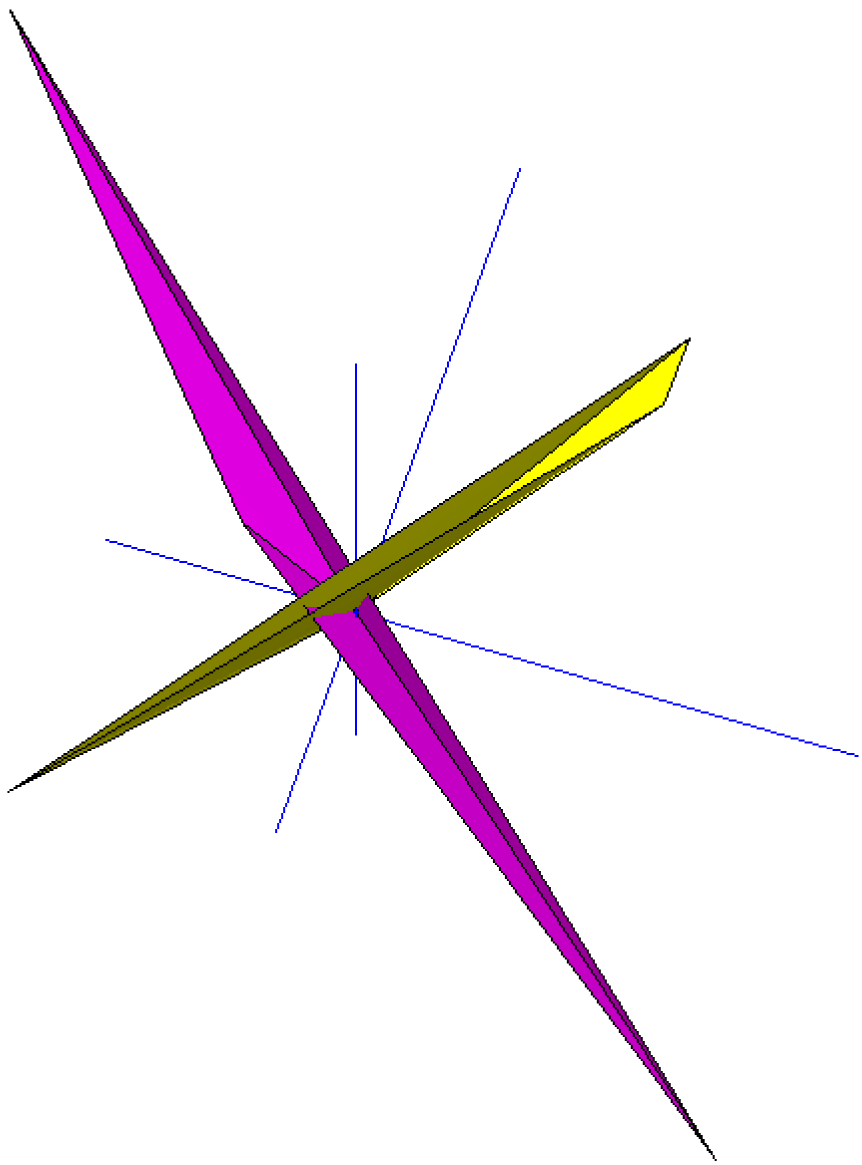


Figure 41: Two views of a simplicial cone and its dual in $\mathbb{R}^3$ (second view on next page). Semi-infinite boundary of each cone is truncated for illustration. Cartesian axes are drawn for reference.



### 2.12.4 Converting between descriptions

Conversion between halfspace-descriptions (245) (246) and equivalent vertex-descriptions (75) (249) is nontrivial, in general, [13] [77, §2.2] but the conversion is easy for simplices. [46, §2.2] Nonetheless, we tacitly assume the two descriptions to be equivalent. [230, §19, thm.19.1] We explore conversions in §2.13.4 and §2.13.9:

## 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. [299] [204, §4.5] For unique minimum-distance projection on a closed convex cone $\mathcal{K}$, the negative dual cone $-\mathcal{K}^*$ plays the role that orthogonal complement plays for subspace projection.[2.43] (§E.9.2.1) Indeed, $-\mathcal{K}^*$ is the *algebraic complement* in $\mathbb{R}^n$;

$$\mathcal{K} \boxplus -\mathcal{K}^* = \mathbb{R}^n \tag{257}$$

where $\boxplus$ 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 or nonnegative orthant. Generalized inequality $\succeq_{\mathcal{K}}$ is a formalized means to determine membership to any pointed closed convex cone (§2.7.2.2) whereas *biorthogonal expansion* is, fundamentally, an expression of coordinates in a pointed conic system. When cone $\mathcal{K}$ is the nonnegative orthant, then these three concepts come into alignment with the Cartesian prototype; biorthogonal expansion becomes orthogonal expansion.

[2.43] Namely, projection on a subspace is ascertainable from its projection on the orthogonal complement.

### 2.13.1 Dual cone

For any set $\mathcal{K}$ (convex or not), the dual cone [46, §2.6.1] [73, §4.2]

$$\mathcal{K}^* \triangleq \left\{ y \in \mathbb{R}^n \mid \langle y\,,\, x\rangle \geq 0 \text{ for all } x \in \mathcal{K} \right\} \tag{258}$$

is a unique cone[2.44] that is always closed and convex because it is an intersection of halfspaces (*halfspaces theorem* (§2.4.1.1.1)) whose partial boundaries each contain the origin, each halfspace having inward-normal $x$ belonging to $\mathcal{K}$; *e.g.*, Figure **42**(a).

When cone $\mathcal{K}$ is convex, there is a second and equivalent construction: Dual cone $\mathcal{K}^*$ is the union of each and every vector $y$ inward-normal to a hyperplane supporting $\mathcal{K}$ or bounding a halfspace containing $\mathcal{K}$; *e.g.*, Figure **42**(b). When $\mathcal{K}$ is represented by a halfspace-description such as (246), for example, where

$$A \triangleq \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m\times n}\,, \qquad C \triangleq \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix} \in \mathbb{R}^{p\times n} \tag{259}$$

then the dual cone can be represented as the conic hull

$$\mathcal{K}^* = \text{cone}\{a_1\,, \ldots,\, a_m\,,\, \pm c_1\,, \ldots,\, \pm c_p\} \tag{260}$$

a vertex-description, because each and every conic combination of normals from the halfspace-description of $\mathcal{K}$ yields another inward-normal to a hyperplane supporting or bounding a halfspace containing $\mathcal{K}$.

$\mathcal{K}^*$ can also be constructed pointwise using projection theory from §E.9.2: for $P_{\mathcal{K}}x$ the Euclidean projection of point $x$ on closed convex cone $\mathcal{K}$

$$-\mathcal{K}^* = \{x - P_{\mathcal{K}}x \mid x \in \mathbb{R}^n\} = \{x \in \mathbb{R}^n \mid P_{\mathcal{K}}x = \mathbf{0}\} \tag{1793}$$

**2.13.1.0.1 Exercise.** *Manual dual cone construction.*
Perhaps the most instructive graphical method of dual cone construction is cut-and-try. Find the dual of each polyhedral cone from Figure **43** by using dual cone equation (258). ▼

[2.44] The dual cone is the negative *polar cone* defined by many authors; $\mathcal{K}^* = -\mathcal{K}^\circ$. [148, §A.3.2] [230, §14] [29] [20] [247, §2.7]

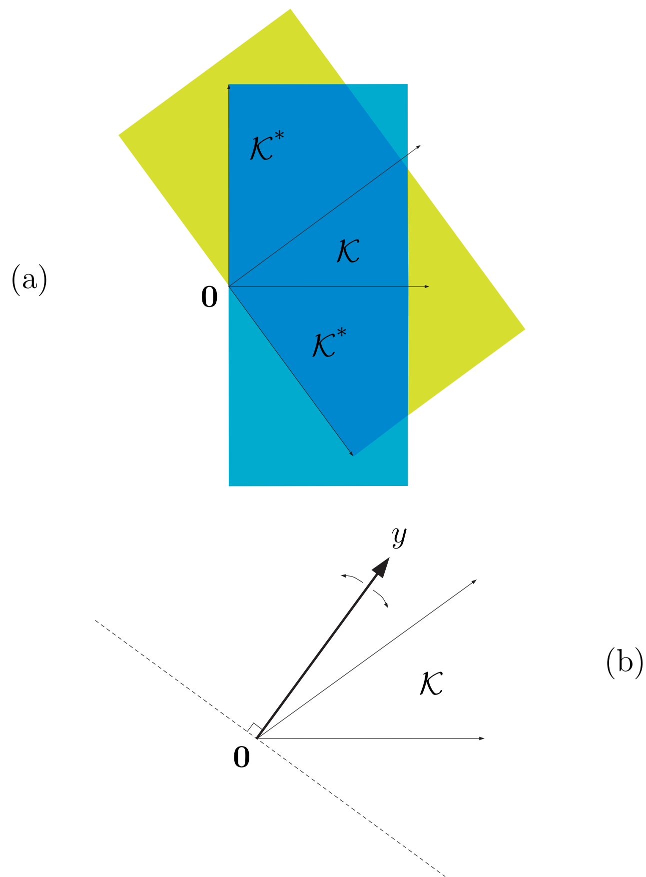
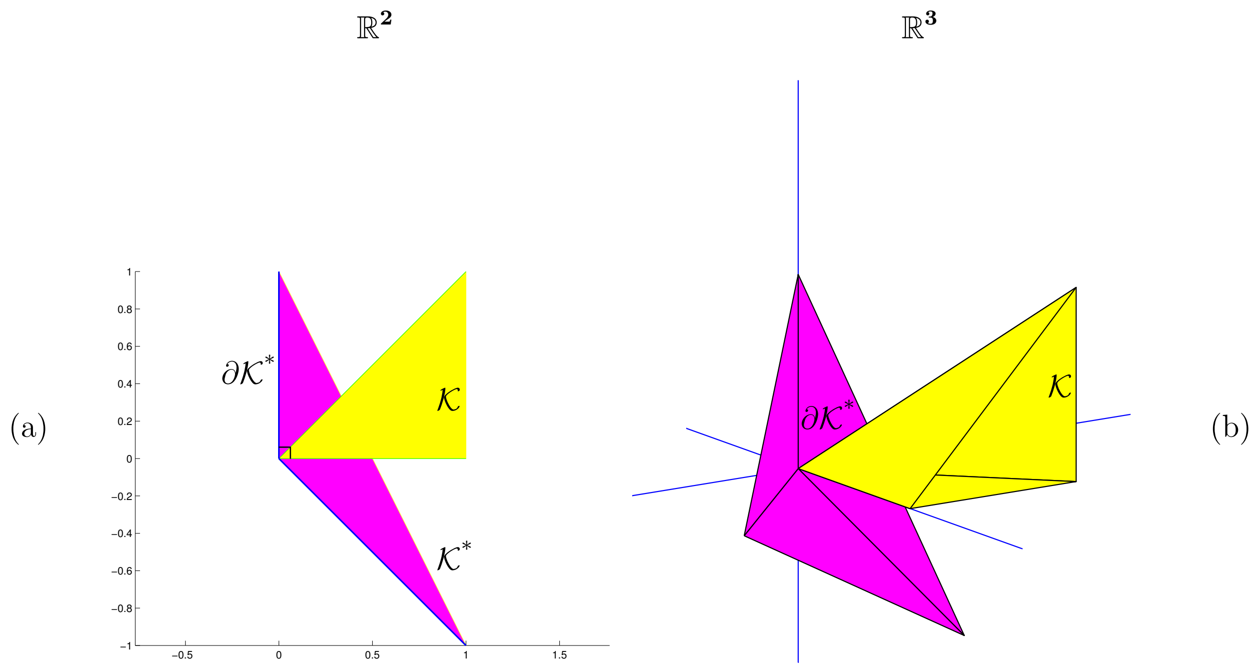



Figure 42: Two equivalent constructions of dual cone $\mathcal{K}^*$ in $\mathbb{R}^2$ : **(a)** Showing construction by intersection of halfspaces about $\mathbf{0}$ (drawn truncated). Only those two halfspaces whose bounding hyperplanes have inward-normal corresponding to an extreme direction of this pointed closed convex cone $\mathcal{K} \subset \mathbb{R}^2$ need be drawn; by (319). **(b)** Suggesting construction by union of inward-normals $y$ to each and every hyperplane $\underline{\partial\mathcal{H}}_+$ 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).




$$x \in \mathcal{K} \;\Leftrightarrow\; \langle y\,,\,x\rangle \geq 0 \;\text{ for all }\; y \in \mathcal{G}(\mathcal{K}^*) \tag{317}$$

Figure 43: Dual cone construction by right angle. Each extreme direction of a 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 $\partial$ 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 **44** would be reached in limit. **(b)** Same polyhedral cone and its dual continued into three dimensions. (*confer* Figure **50**)

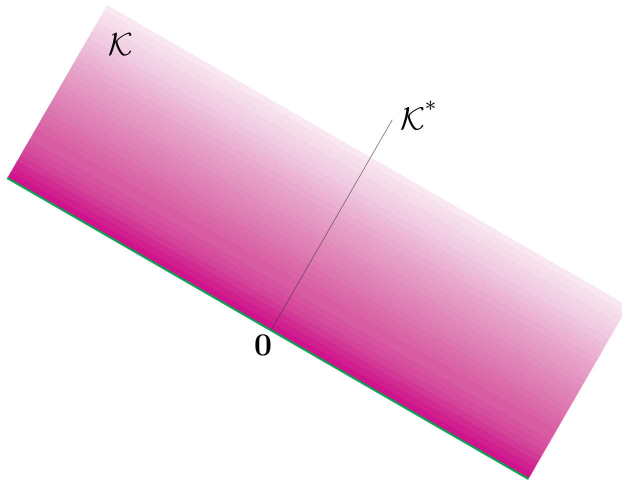



Figure 44: $\mathcal{K}$ is a halfspace about the origin in $\mathbb{R}^2$. $\mathcal{K}^*$ is a ray base $\mathbf{0}$, hence has empty interior in $\mathbb{R}^2$; so $\mathcal{K}$ cannot be pointed. (Both convex cones appear truncated.)

**2.13.1.0.2 Exercise.** *Dual cone definitions.*
What is $\{x\in\mathbb{R}^n \mid x^Tz\geq 0 \;\; \forall\, z\in\mathbb{R}^n\}$ ?
What is $\{x\in\mathbb{R}^n \mid x^Tz\geq 1 \;\; \forall\, z\in\mathbb{R}^n\}$ ?
What is $\{x\in\mathbb{R}^n \mid x^Tz\geq 1 \;\; \forall\, z\in\mathbb{R}^n_+\}$ ? ▼

As defined, dual cone $\mathcal{K}^*$ exists even when the affine hull of the original cone is a proper subspace; *id est*, even when the original cone has empty interior. Rockafellar formulates the dimension of $\mathcal{K}$ and $\mathcal{K}^*$. [230, §14][2.45]

To further motivate our understanding of the dual cone, consider the ease with which convergence can be observed in the following optimization problem (p):

**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*

[2.45] His monumental work *Convex Analysis* has not one figure or illustration. See [20, §II.16] for a good illustration of Rockafellar's *recession cone* [30].

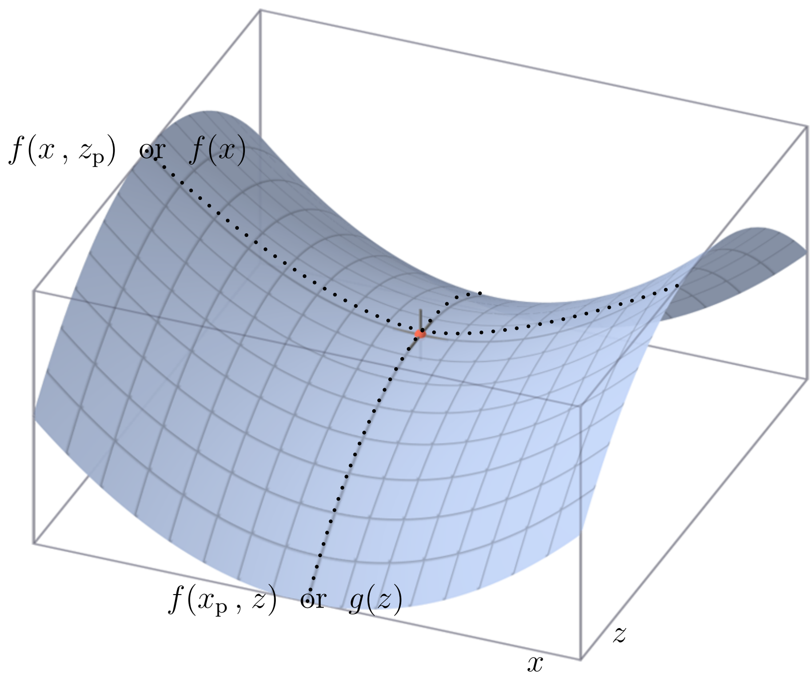



Figure 45: Although objective functions from conic problems (263p) and (263d) are linear, this is a mnemonic icon for primal and dual problems. When problems are *strong duals*, duality gap is 0 ; meaning, functions $f(x)$ and $g(z)$ (dotted) kiss at saddle value, as depicted at center. Otherwise, dual functions never meet $\big(f(x) > g(z)\big)$ by (261). Drawing by http://en.wikipedia.org/wiki/User:Kieff

*the primal program. Third, the maximum value achieved by the dual problem is often equal to the minimum of the primal.* [222, §2.1.3] Essentially, duality theory concerns representation of a given optimization problem as half a *minimax problem.* [230, §36] [46, §5.4] Given any real function $f(x\,,z)$

$$\underset{x}{\text{minimize}}\,\underset{z}{\text{maximize}}\, f(x\,,z) \geq \underset{z}{\text{maximize}}\,\underset{x}{\text{minimize}}\, f(x\,,z) \tag{261}$$

always holds. When

$$\underset{x}{\text{minimize}}\,\underset{z}{\text{maximize}}\, f(x\,,z) = \underset{z}{\text{maximize}}\,\underset{x}{\text{minimize}}\, f(x\,,z) \tag{262}$$

we have *strong duality* and then a *saddle value* [104] exists. (Figure **45**) [227, p.3] Consider primal conic problem (p) and its corresponding dual problem (d): [217, §3.3.1] [175, §2.1] given vectors $\alpha\,,\beta$ and matrix constant $C$

$$\text{(p)}\quad \begin{array}{ll} \underset{x}{\text{minimize}} & \alpha^T x \\ \text{subject to} & x \in \mathcal{K} \\ & Cx = \beta \end{array} \qquad \begin{array}{ll} \underset{y\,,\,z}{\text{maximize}} & \beta^T z \\ \text{subject to} & y \in \mathcal{K}^* \\ & C^T z + y = \alpha \end{array} \quad \text{(d)} \tag{263}$$

Observe the dual problem is also conic, and its objective function value never exceeds that of the primal;

$$\begin{aligned} \alpha^T x &\geq \beta^T z \\ x^T(C^T z + y) &\geq (Cx)^T z \\ x^T y &\geq 0 \end{aligned} \tag{264}$$

which holds by definition (258). Under the sufficient condition: (263p) is a *convex problem* and satisfies *Slater's condition*,[2.46] then 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* (primal−dual objective difference) is 0. Then (p) and (d) are together equivalent to the minimax problem

$$\begin{array}{lll} \underset{x,y,z}{\text{minimize}} & \alpha^T x - \beta^T z & \\ \text{subject to} & x \in \mathcal{K}\,, & y \in \mathcal{K}^* \\ & Cx = \beta\,, & C^T z + y = \alpha \end{array} \qquad \text{(p)}-\text{(d)} \tag{265}$$

---

[2.46] A convex problem, essentially, has convex objective function optimized over a convex set. (§4) In this context, (p) is convex if $\mathcal{K}$ is a convex cone. Slater's condition is satisfied whenever any primal strictly feasible point exists. (p.235) whose optimal objective always has the saddle value 0 (regardless of the particular convex cone $\mathcal{K}$ and other problem parameters). [269, §3.2] Thus determination of convergence for either primal or dual problem is facilitated.

□

#### 2.13.1.1 Key properties of dual cone

- For any cone, $(-\mathcal{K})^* = -\mathcal{K}^*$
- For any cones $\mathcal{K}_1$ and $\mathcal{K}_2$ , $\mathcal{K}_1 \subseteq \mathcal{K}_2 \Rightarrow \mathcal{K}_1^* \supseteq \mathcal{K}_2^*$ [247, §2.7]
- (Cartesian product) For closed convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$ , their Cartesian product $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2$ is a closed convex cone, and

$$\mathcal{K}^* = \mathcal{K}_1^* \times \mathcal{K}_2^* \tag{266}$$

- (conjugation) [230, §14] [73, §4.5] When $\mathcal{K}$ is any convex cone, the dual of the dual cone is the closure of the original cone; $\mathcal{K}^{**} = \overline{\mathcal{K}}$ . Because $\mathcal{K}^{***} = \mathcal{K}^*$

$$\mathcal{K}^* = (\overline{\mathcal{K}})^* \tag{267}$$

  When $\mathcal{K}$ is closed and convex, then the dual of the dual cone is the original cone; $\mathcal{K}^{**} = \mathcal{K}$ .

- If any cone $\mathcal{K}$ has nonempty interior, then $\mathcal{K}^*$ is pointed;

$$\mathcal{K} \text{ nonempty interior } \Rightarrow \mathcal{K}^* \text{ pointed} \tag{268}$$

  Conversely, if the closure of any convex cone $\mathcal{K}$ is pointed, then $\mathcal{K}^*$ has nonempty interior;

$$\overline{\mathcal{K}} \text{ pointed } \Rightarrow \mathcal{K}^* \text{ nonempty interior} \tag{269}$$

  Given that a cone $\mathcal{K} \subset \mathbb{R}^n$ is closed and convex, $\mathcal{K}$ is pointed if and only if $\mathcal{K}^* - \mathcal{K}^* = \mathbb{R}^n$ ; *id est*, iff $\mathcal{K}^*$ has nonempty interior. [41, §3.3, exer.20]

- (vector sum) [230, thm.3.8] For convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$

$$\mathcal{K}_1 + \mathcal{K}_2 = \text{conv}(\mathcal{K}_1 \cup \mathcal{K}_2) \tag{270}$$

- (dual vector-sum) [230, §16.4.2] [73, §4.6] For convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$

$$\mathcal{K}_1^* \cap \mathcal{K}_2^* \;=\; (\mathcal{K}_1 + \mathcal{K}_2)^* \;=\; (\mathcal{K}_1 \cup \mathcal{K}_2)^* \tag{271}$$

  (closure of vector sum of duals)[2.47] For closed convex cones $\mathcal{K}_1$ and $\mathcal{K}_2$

$$(\mathcal{K}_1 \cap \mathcal{K}_2)^* \;=\; \overline{\mathcal{K}_1^* + \mathcal{K}_2^*} \;=\; \overline{\operatorname{conv}(\mathcal{K}_1^* \cup \mathcal{K}_2^*)} \tag{272}$$

  where closure becomes superfluous under the condition $\mathcal{K}_1 \cap \operatorname{int} \mathcal{K}_2 \neq \emptyset$ [41, §3.3, exer.16, §4.1, exer.7].

- (Krein-Rutman) For closed convex cones $\mathcal{K}_1 \subseteq \mathbb{R}^m$ and $\mathcal{K}_2 \subseteq \mathbb{R}^n$ and any linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, then provided $\operatorname{int} \mathcal{K}_1 \cap A\mathcal{K}_2 \neq \emptyset$ [41, §3.3.13, *confer* §4.1, exer.9]

$$(A^{-1}\mathcal{K}_1 \cap \mathcal{K}_2)^* \;=\; A^T\mathcal{K}_1^* + \mathcal{K}_2^* \tag{273}$$

  where the dual of cone $\mathcal{K}_1$ is with respect to its ambient space $\mathbb{R}^m$ and the dual of cone $\mathcal{K}_2$ is with respect to $\mathbb{R}^n$, where $A^{-1}\mathcal{K}_1$ denotes the inverse image (§2.1.9.0.1) of $\mathcal{K}_1$ under mapping $A$, and where $A^T$ denotes the adjoint operation.

- $\mathcal{K}$ is proper if and only if $\mathcal{K}^*$ is proper.

- $\mathcal{K}$ is polyhedral if and only if $\mathcal{K}^*$ is polyhedral. [247, §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 polyhedral proper cones (Figure **50**, Figure **41**), but not the converse.

- $\mathcal{K} \boxplus -\mathcal{K}^* = \mathbb{R}^n \Leftrightarrow \mathcal{K}$ is closed and convex. (1792) (p.676)

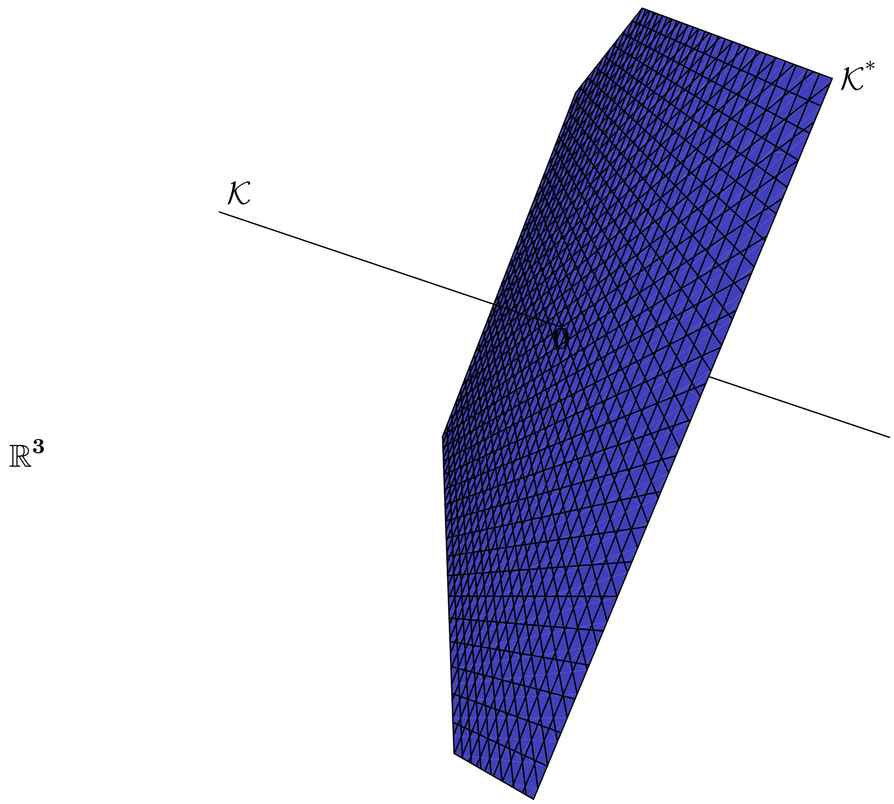
- Any direction in a proper cone $\mathcal{K}$ is normal to a hyperplane separating $\mathcal{K}$ from $-\mathcal{K}^*$.

---

[2.47] These parallel analogous results for subspaces $\mathcal{R}_1, \mathcal{R}_2 \subseteq \mathbb{R}^n$; [73, §4.6]

$$\begin{aligned}(\mathcal{R}_1 + \mathcal{R}_2)^\perp &= \mathcal{R}_1^\perp \cap \mathcal{R}_2^\perp \\ (\mathcal{R}_1 \cap \mathcal{R}_2)^\perp &= \overline{\mathcal{R}_1^\perp + \mathcal{R}_2^\perp}\end{aligned}$$

$\mathcal{R}^{\perp\perp} = \mathcal{R}$ for any subspace $\mathcal{R}$.



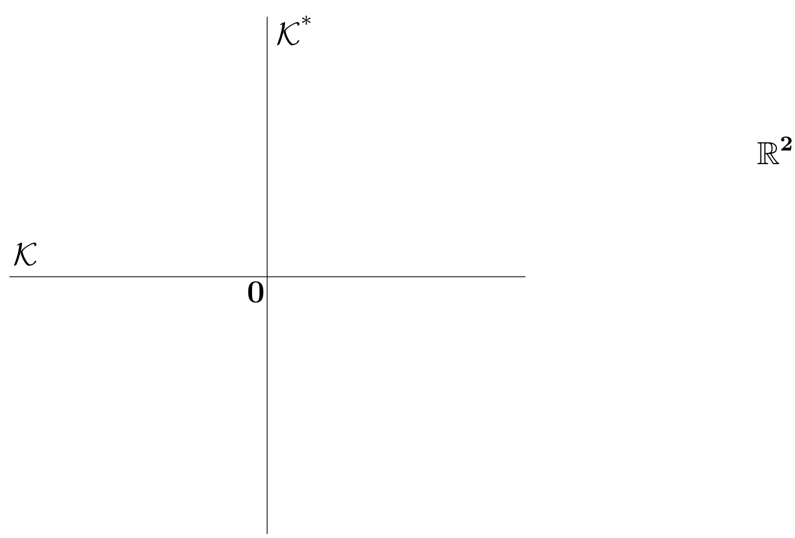




Figure 46: When convex cone $\mathcal{K}$ is any one Cartesian axis, its dual cone is the convex hull of all axes remaining. In $\mathbb{R}^3$, dual cone $\mathcal{K}^*$ (drawn tiled and truncated) is a hyperplane through origin; its normal belongs to line $\mathcal{K}$. In $\mathbb{R}^2$, dual cone $\mathcal{K}^*$ is a line through origin while convex cone $\mathcal{K}$ is that line orthogonal.

#### 2.13.1.2 Examples of dual cone

When $\mathcal{K}$ is $\mathbb{R}^n$, $\mathcal{K}^*$ is the point at the origin, and *vice versa.*

When $\mathcal{K}$ is a subspace, $\mathcal{K}^*$ is its orthogonal complement, and *vice versa.* (§E.9.2.1, Figure **46**)

When cone $\mathcal{K}$ is a halfspace in $\mathbb{R}^n$ with $n>0$ (Figure **44** for example), the dual cone $\mathcal{K}^*$ is a ray (base $\mathbf{0}$) belonging to that halfspace but orthogonal to its bounding hyperplane (that contains the origin), and *vice versa.*

When convex cone $\mathcal{K}$ is a closed halfplane in $\mathbb{R}^3$ (Figure **47**), it is neither pointed or of nonempty interior; hence, the dual cone $\mathcal{K}^*$ can be neither of nonempty interior or pointed.

When $\mathcal{K}$ is any particular orthant in $\mathbb{R}^n$, the dual cone is identical; *id est*, $\mathcal{K}=\mathcal{K}^*$.

When $\mathcal{K}$ is any quadrant in subspace $\mathbb{R}^2$, $\mathcal{K}^*$ is a *wedge*-shaped polyhedral cone in $\mathbb{R}^3$; *e.g.*, for $\mathcal{K}$ equal to quadrant I, $\mathcal{K}^*=\begin{bmatrix}\mathbb{R}^2_+\\ \mathbb{R}\end{bmatrix}$.

When $\mathcal{K}$ is a polyhedral flavor of the Lorentz cone $\mathcal{K}_\ell$ (247), the dual is the polyhedral proper cone $\mathcal{K}_q$: for $\ell=1$ or $\infty$

$$\mathcal{K}_q \;=\; \mathcal{K}_\ell^* \;=\; \left\{\begin{bmatrix} x\\ t\end{bmatrix}\in\mathbb{R}^n\times\mathbb{R} \mid \|x\|_q\leq t\right\} \tag{274}$$

where $\|x\|_q$ is the *dual norm* determined via solution to $1/\ell+1/q=1$.

### 2.13.2 Abstractions of *Farkas' lemma*

**2.13.2.0.1 Corollary.** *Generalized inequality and membership relation.* [148, §A.4.2] Let $\mathcal{K}$ be any closed convex cone and $\mathcal{K}^*$ its dual, and let $x$ and $y$ belong to a vector space $\mathbb{R}^n$. Then

$$y\in\mathcal{K}^* \;\Leftrightarrow\; \langle y\,,\,x\rangle\geq 0 \;\text{ for all }\; x\in\mathcal{K} \tag{275}$$

which is, merely, a statement of fact by definition of dual cone (258). By closure we have conjugation:

$$x\in\mathcal{K} \;\Leftrightarrow\; \langle y\,,\,x\rangle\geq 0 \;\text{ for all }\; y\in\mathcal{K}^* \tag{276}$$

which may be regarded as a simple translation of the *Farkas lemma* [89] as in [230, §22] to the language of convex cones, and a generalization of the well-known Cartesian fact

$$x\succeq 0 \;\Leftrightarrow\; \langle y\,,\,x\rangle\geq 0 \;\text{ for all }\; y\succeq 0 \tag{277}$$

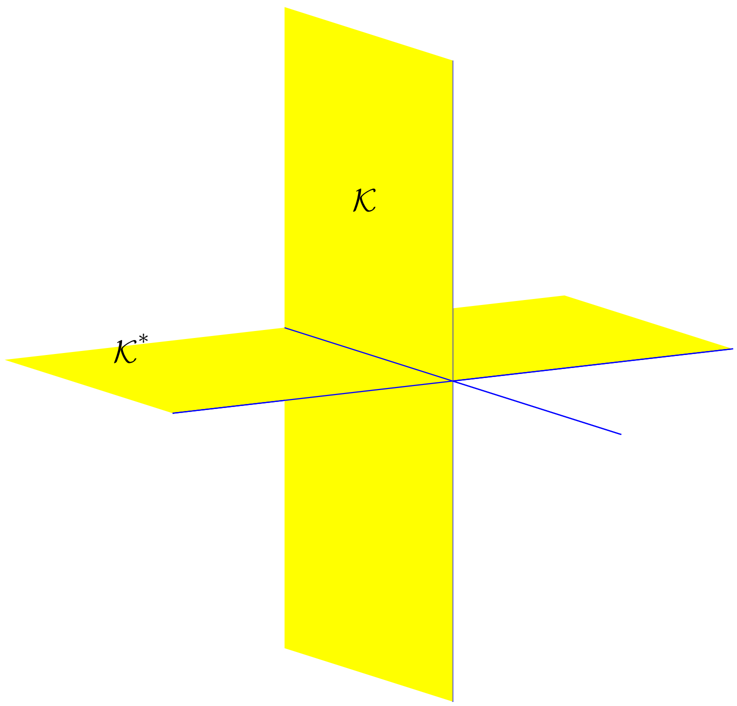



Figure 47: $\mathcal{K}$ and $\mathcal{K}^*$ are halfplanes in $\mathbb{R}^3$; blades. Both semi-infinite convex cones appear truncated. Each cone is like $\mathcal{K}$ in Figure **44**, but embedded in a two-dimensional subspace of $\mathbb{R}^3$. Cartesian coordinate axes drawn for reference.

for which implicitly $\mathcal{K}=\mathcal{K}^*=\mathbb{R}^n_+$ the nonnegative orthant.

*Membership relation* (276) is often written instead as *dual generalized inequalities*, when $\mathcal{K}$ and $\mathcal{K}^*$ are pointed closed convex cones,

$$x \underset{\mathcal{K}}{\succeq} 0 \;\Leftrightarrow\; \langle y\,,x\rangle \geq 0 \;\text{ for all }\; y \underset{\mathcal{K}^*}{\succeq} 0 \tag{278}$$

meaning, coordinates for biorthogonal expansion of $x$ (§2.13.8) [275] must be nonnegative when $x$ belongs to $\mathcal{K}$. By conjugation [230, thm.14.1]

$$y \underset{\mathcal{K}^*}{\succeq} 0 \;\Leftrightarrow\; \langle y\,,x\rangle \geq 0 \;\text{ for all }\; x \underset{\mathcal{K}}{\succeq} 0 \tag{279}$$

◇

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:

$$\begin{array}{c} x\in\mathcal{K} \;\Leftrightarrow\; \langle y\,,x\rangle \geq 0 \;\text{ for all }\; y\in\mathcal{K}^* \qquad (276) \\ \Leftrightarrow \\ x\in\mathcal{K} \;\Leftrightarrow\; \langle y\,,x\rangle \geq 0 \;\text{ for all }\; y\in\mathcal{K}^*,\;\; \|y\|=1 \end{array} \tag{280}$$

**2.13.2.0.2 Exercise.** *Test of dual generalized inequalities.*
Test Corollary 2.13.2.0.1 and (280) graphically on the two-dimensional polyhedral cone and its dual in Figure **43**. ▼

When pointed closed convex cone $\mathcal{K}$ is implicit from context: (*confer* §2.7.2.2)

$$\begin{array}{c} x \succeq 0 \;\Leftrightarrow\; x\in\mathcal{K} \\ x \succ 0 \;\Leftrightarrow\; x\in\operatorname{rel\,int}\mathcal{K} \end{array} \tag{281}$$

Strict inequality $x \succ 0$ means coordinates for biorthogonal expansion of $x$ must be positive when $x$ belongs to $\operatorname{rel\,int}\mathcal{K}$. Strict membership relations are useful; *e.g.*, for any proper cone $\mathcal{K}$ and its dual $\mathcal{K}^*$

$$x\in\operatorname{int}\mathcal{K} \;\Leftrightarrow\; \langle y\,,x\rangle > 0 \;\text{ for all }\; y\in\mathcal{K}^*,\;\; y\neq\mathbf{0} \tag{282}$$

$$x\in\mathcal{K}\,,\;\; x\neq\mathbf{0} \;\Leftrightarrow\; \langle y\,,x\rangle > 0 \;\text{ for all }\; y\in\operatorname{int}\mathcal{K}^* \tag{283}$$

By conjugation, we also have the dual relations:

$$y \in \operatorname{int} \mathcal{K}^* \Leftrightarrow \langle y\,,\, x\rangle > 0 \text{ for all } x \in \mathcal{K}\,,\; x \neq \mathbf{0} \tag{284}$$

$$y \in \mathcal{K}^*,\; y \neq \mathbf{0} \Leftrightarrow \langle y\,,\, x\rangle > 0 \text{ for all } x \in \operatorname{int} \mathcal{K} \tag{285}$$

Boundary-membership relations for proper cones are also useful:

$$x \in \partial\mathcal{K} \Leftrightarrow \exists\, y \ni \langle y\,,\, x\rangle = 0\,,\; y \in \mathcal{K}^*,\; y \neq \mathbf{0}\,,\; x \in \mathcal{K} \tag{286}$$

$$y \in \partial\mathcal{K}^* \Leftrightarrow \exists\, x \ni \langle y\,,\, x\rangle = 0\,,\; x \in \mathcal{K}\,,\; x \neq \mathbf{0}\,,\; y \in \mathcal{K}^* \tag{287}$$

**2.13.2.0.3 Example.** *Linear inequality.* [254, §4]
(*confer* §2.13.5.1.1) Consider a given matrix $A$ and closed convex cone $\mathcal{K}$. By membership relation we have

$$\begin{aligned} Ay \in \mathcal{K}^* &\Leftrightarrow x^TAy \geq 0 \;\; \forall\, x \in \mathcal{K} \\ &\Leftrightarrow y^Tz \geq 0 \;\; \forall\, z \in \{A^Tx \mid x \in \mathcal{K}\} \\ &\Leftrightarrow y \in \{A^Tx \mid x \in \mathcal{K}\}^* \end{aligned} \tag{288}$$

This implies

$$\{y \mid Ay \in \mathcal{K}^*\} = \{A^Tx \mid x \in \mathcal{K}\}^* \tag{289}$$

If we regard $A$ as a linear operator, then $A^T$ is its adjoint. When, for example, $\mathcal{K}$ is the *self-dual* nonnegative orthant, (§2.13.5.1) then

$$\{y \mid Ay \succeq 0\} = \{A^Tx \mid x \succeq 0\}^* \tag{290}$$

□

### 2.13.2.1 Null certificate, Theorem of the alternative

If in particular $x_{\rm p} \notin \mathcal{K}$ a closed convex cone, then the construction in Figure **42**(b) suggests there exists a supporting hyperplane (having inward-normal belonging to dual cone $\mathcal{K}^*$) separating $x_{\rm p}$ from $\mathcal{K}$; indeed, (276)

$$x_{\rm p} \notin \mathcal{K} \Leftrightarrow \exists\, y \in \mathcal{K}^* \ni \langle y\,,\, x_{\rm p}\rangle < 0 \tag{291}$$

The existence of any one such $y$ is a certificate of null membership. From a different perspective,

$$\begin{gathered} x_{\rm p} \in \mathcal{K} \\ \text{or in the alternative} \\ \exists\, y \in \mathcal{K}^* \ni \langle y\,,\, x_{\rm p}\rangle < 0 \end{gathered} \tag{292}$$

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 (277) (with respect to the nonnegative orthant $\mathbb{R}^m_+$),

$$y \succeq 0 \;\Leftrightarrow\; x^Ty \geq 0 \;\text{ for all }\; x \succeq 0 \tag{293}$$

Given $A \in \mathbb{R}^{n\times m}$, we make vector substitution $A^Ty \leftarrow y$

$$A^Ty \succeq 0 \;\Leftrightarrow\; x^TA^Ty \geq 0 \;\text{ for all }\; x \succeq 0 \tag{294}$$

Introducing a new vector by calculating $b \triangleq Ax$ we get

$$A^Ty \succeq 0 \quad\Leftrightarrow\quad b^Ty \geq 0\,,\quad b = Ax \;\text{ for all }\; x \succeq 0 \tag{295}$$

By complementing sense of the scalar inequality:

$$\begin{array}{c} A^Ty \succeq 0 \\ \text{or in the alternative} \\ b^Ty < 0\,,\quad \exists\, b = Ax\,,\quad x \succeq 0 \end{array} \tag{296}$$

If one system has a solution, then the other does not; define a convex cone $\mathcal{K} \triangleq \{y \mid A^Ty \succeq 0\}$, then $y \in \mathcal{K}$ or in the alternative $y \notin \mathcal{K}$. Scalar inequality $b^Ty < 0$ can be moved to the other side of the alternative, but that requires some explanation: From the results of Example 2.13.2.0.3, the dual cone is $\mathcal{K}^* = \{Ax \mid x \succeq 0\}$. We have

$$y \in \mathcal{K} \quad\Leftrightarrow\quad b^Ty \geq 0 \;\text{ for all }\; b \in \mathcal{K}^* \tag{297}$$

$$A^Ty \succeq 0 \quad\Leftrightarrow\quad b^Ty \geq 0 \;\text{ for all }\; b \in \{Ax \mid x \succeq 0\} \tag{298}$$

Given some $b$ vector and $y \in \mathcal{K}$, then $b^Ty < 0$ can only mean $b \notin \mathcal{K}^*$. An alternative system is therefore simply $b \in \mathcal{K}^*$: [148, p.59] (Farkas/Tucker)

$$\begin{array}{c} A^Ty \succeq 0\,,\quad b^Ty < 0 \\ \text{or in the alternative} \\ b = Ax\,,\quad x \succeq 0 \end{array} \tag{299}$$

For another example, from membership relation (275) with affine transformation of dual variable we may write: Given $A\in\mathbb{R}^{n\times m}$ and $b\in\mathbb{R}^n$

$$b - Ay \in \mathcal{K}^* \;\Leftrightarrow\; x^T(b-Ay)\geq 0 \qquad \forall\, x\in\mathcal{K} \tag{300}$$

$$A^Tx=\mathbf{0}\,,\quad b - Ay \in \mathcal{K}^* \;\Rightarrow\; x^Tb\geq 0 \qquad \forall\, x\in\mathcal{K} \tag{301}$$

From membership relation (300), conversely, suppose we allow any $y\in\mathbb{R}^m$. Then because $-x^TA\,y$ is unbounded below, $x^T(b-Ay)\geq 0$ implies $A^Tx=\mathbf{0}$ : for $y\in\mathbb{R}^m$

$$A^Tx=\mathbf{0}\,,\quad b - Ay \in \mathcal{K}^* \;\Leftarrow\; x^T(b-Ay)\geq 0 \qquad \forall\, x\in\mathcal{K} \tag{302}$$

*In toto*,

$$b - Ay \in \mathcal{K}^* \;\Leftrightarrow\; x^Tb\geq 0\,,\;\; A^Tx=\mathbf{0} \quad \forall\, x\in\mathcal{K} \tag{303}$$

Vector $x$ belongs to cone $\mathcal{K}$ but is also constrained to lie in a subspace of $\mathbb{R}^n$ specified by an intersection of hyperplanes through the origin $\{x\in\mathbb{R}^n \mid A^Tx=\mathbf{0}\}$. From this, alternative systems of generalized inequality with respect to pointed closed convex cones $\mathcal{K}$ and $\mathcal{K}^*$

$$\begin{array}{c} Ay \underset{\mathcal{K}^*}{\preceq} b \\ \\ \text{or in the alternative} \\ \\ x^Tb<0\,,\quad A^Tx=\mathbf{0}\,,\quad x \underset{\mathcal{K}}{\succeq} 0 \end{array} \tag{304}$$

derived from (303) 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.[2.48] [230, p.201]

By invoking a strict membership relation between proper cones (282), we can construct a more exotic alternative strengthened by demand for an interior point;

---

[2.48]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: $x\in\mathbb{S}^{\mathbf{2}}$, $y\in\mathbb{R}$ , $A=\begin{bmatrix}1&0\\0&0\end{bmatrix}$, and $b=\begin{bmatrix}0&1\\1&0\end{bmatrix}$ where $x^Tb$ means $\langle x\,,\,b\rangle$ . A better strategy than disallowing solutions at $\pm\infty$ is to demand an interior point as in (306) or Lemma 4.2.1.1.2. Then question of simultaneous infeasibility is moot.

$$b - Ay \underset{\mathcal{K}^*}{\succ} 0 \quad \Leftrightarrow \quad x^T b > 0\,, \quad A^T x = \mathbf{0} \quad \forall\, x \underset{\mathcal{K}}{\succeq} 0\,, \quad x \neq \mathbf{0} \tag{305}$$

From this, alternative systems of generalized inequality [46, pages:50,54,262]

$$\begin{array}{c} Ay \underset{\mathcal{K}^*}{\prec} b \\ \\ \text{or in the alternative} \\ \\ x^T b \leq 0\,, \quad A^T x = \mathbf{0}\,, \quad x \underset{\mathcal{K}}{\succeq} 0\,, \quad x \neq \mathbf{0} \end{array} \tag{306}$$

derived from (305) by taking the complementary sense of the inequality in $x^T b$.

And from this, alternative systems with respect to the nonnegative orthant attributed to Gordan in 1873: [111] [41, §2.2] substituting $A \leftarrow A^T$ and setting $b = \mathbf{0}$

$$\begin{array}{c} A^T y \prec 0 \\ \\ \text{or in the alternative} \\ \\ Ax = \mathbf{0}\,, \quad x \succeq 0\,, \quad \|x\|_1 = 1 \end{array} \tag{307}$$

□

### 2.13.3 Optimality condition

The general first-order necessary and sufficient condition for optimality of solution $x^\star$ to a minimization problem ((263p) for example) with real differentiable convex objective function $f(x) : \mathbb{R}^n \to \mathbb{R}$ is [229, §3] (*confer* §2.13.10.1) (Figure **53**)

$$\nabla f(x^\star)^T (x - x^\star) \geq 0 \quad \forall\, x \in \mathcal{C}\,, \quad x^\star \in \mathcal{C} \tag{308}$$

where $\mathcal{C}$ is the *feasible set*, a convex set of all variable values satisfying the problem constraints, and where $\nabla f(x^\star)$ is the *gradient* of $f$ (§3.1.8) with respect to $x$ evaluated at $x^\star$.

**2.13.3.0.1 Example.** *Equality constrained problem.*
Given a real differentiable convex function $f(x): \mathbb{R}^n \to \mathbb{R}$ defined on domain $\mathbb{R}^n$, a fat full-rank matrix $C \in \mathbb{R}^{p\times n}$, and vector $d \in \mathbb{R}^p$, the convex optimization problem

$$\begin{array}{cl}\underset{x}{\text{minimize}} & f(x)\\ \text{subject to} & Cx = d\end{array} \tag{309}$$

is characterized by the well-known necessary and sufficient optimality condition [46, §4.2.3]

$$\nabla f(x^\star) + C^T\nu = \mathbf{0} \tag{310}$$

where $\nu \in \mathbb{R}^p$ is the eminent *Lagrange multiplier.* [228] Feasible solution $x^\star$ is optimal, in other words, if and only if $\nabla f(x^\star)$ belongs to $\mathcal{R}(C^T)$. Via membership relation, we now derive this particular condition from the general first-order condition for optimality (308):

In this case, the feasible set is

$$\mathcal{C} \triangleq \{x \in \mathbb{R}^n \mid Cx = d\} \;=\; \{Z\xi + x_{\mathrm{p}} \mid \xi \in \mathbb{R}^{n-\operatorname{rank} C}\} \tag{311}$$

where $Z \in \mathbb{R}^{n\times n-\operatorname{rank} C}$ holds basis $\mathcal{N}(C)$ columnar, and $x_{\mathrm{p}}$ is any particular solution to $Cx = d$. Since $x^\star \in \mathcal{C}$, we arbitrarily choose $x_{\mathrm{p}} = x^\star$ which yields the equivalent optimality condition

$$\nabla f(x^\star)^T Z\xi \ge 0 \quad \forall\, \xi \in \mathbb{R}^{n-\operatorname{rank} C} \tag{312}$$

But this is simply half of a membership relation, and the cone dual to $\mathbb{R}^{n-\operatorname{rank} C}$ is the origin in $\mathbb{R}^{n-\operatorname{rank} C}$. We must therefore have

$$Z^T\nabla f(x^\star) = \mathbf{0} \;\Leftrightarrow\; \nabla f(x^\star)^T Z\xi \ge 0 \quad \forall\, \xi \in \mathbb{R}^{n-\operatorname{rank} C} \tag{313}$$

meaning, $\nabla f(x^\star)$ must be orthogonal to $\mathcal{N}(C)$. This condition

$$Z^T\nabla f(x^\star) = \mathbf{0}\,, \qquad x^\star \in \mathcal{C} \tag{314}$$

is necessary and sufficient for optimality of $x^\star$. □

### 2.13.4 Discretization of membership relation

#### 2.13.4.1 Dual halfspace-description

Halfspace-description of the dual cone is equally simple (and extensible to an infinite number of generators) as vertex-description (252) for the corresponding closed convex cone: By definition (258), for $X\in\mathbb{R}^{n\times N}$ as in (240), (*confer* (246))

$$\begin{aligned}\mathcal{K}^* &= \left\{y\in\mathbb{R}^n \mid z^Ty\geq 0 \text{ for all } z\in\mathcal{K}\right\}\\ &= \left\{y\in\mathbb{R}^n \mid z^Ty\geq 0 \text{ for all } z=Xa\,,\ a\succeq 0\right\}\\ &= \left\{y\in\mathbb{R}^n \mid a^TX^Ty\geq 0\,,\ a\succeq 0\right\}\\ &= \left\{y\in\mathbb{R}^n \mid X^Ty\succeq 0\right\}\end{aligned} \tag{315}$$

that follows from the *generalized inequality and membership corollary* (277). 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 $\mathcal{K}$ is known to be closed and convex, then the converse must also hold; *id est*, given any set of generators for $\mathcal{K}^*$ arranged columnar in $Y$ , then the consequent vertex-description of the dual cone connotes a halfspace-description for $\mathcal{K}$: [247, §2.8]

$$\mathcal{K}^* = \{Ya \mid a\succeq 0\} \ \Leftrightarrow\ \mathcal{K}^{**}=\mathcal{K}=\left\{z \mid Y^Tz\succeq 0\right\} \tag{316}$$

#### 2.13.4.2 First dual-cone formula

From these two results (315) and (316) we deduce a general principle:

- From any given vertex-description of a convex cone $\mathcal{K}$ , a halfspace-description of the dual cone $\mathcal{K}^*$ is immediate by matrix transposition; conversely, from any given halfspace-description, a dual vertex-description is immediate.

Various other converses are just a little trickier. (§2.13.9)

We deduce further: For any polyhedral cone $\mathcal{K}$ , the dual cone $\mathcal{K}^*$ is also polyhedral and $\mathcal{K}^{**}=\mathcal{K}$ . [247, §2.8]

The *generalized inequality and membership corollary* is discretized in the following theorem [18, §1][2.49] that follows directly from (315) and (316):

[2.49] Barker & Carlson state the theorem only for the pointed closed convex case.

**2.13.4.2.1 Theorem.** *Discrete membership.* (*confer* §2.13.2.0.1)
Given any set of generators (§2.8.1.2) denoted by $\mathcal{G}(\mathcal{K})$ for closed convex cone $\mathcal{K}\subseteq\mathbb{R}^n$ and any set of generators denoted $\mathcal{G}(\mathcal{K}^*)$ for its dual, let $x$ and $y$ belong to vector space $\mathbb{R}^n$. Then discretization of the *generalized inequality and membership corollary* is necessary and sufficient for certifying membership:

$$x \in \mathcal{K} \;\Leftrightarrow\; \langle \gamma^*, x\rangle \geq 0 \;\text{ for all }\; \gamma^* \in \mathcal{G}(\mathcal{K}^*) \tag{317}$$

$$y \in \mathcal{K}^* \;\Leftrightarrow\; \langle \gamma\,, y\rangle \geq 0 \;\text{ for all }\; \gamma \in \mathcal{G}(\mathcal{K}) \tag{318}$$

◇

**2.13.4.2.2 Exercise.** *Test of discretized dual generalized inequalities.*
Test Theorem 2.13.4.2.1 on Figure **43**(a) using the extreme directions as generators. ▼

From the *discrete membership theorem* we may further deduce a more surgical description of dual cone that prescribes only a finite number of halfspaces for its construction when polyhedral: (Figure **42**(a))

$$\mathcal{K}^* = \{y \in \mathbb{R}^n \mid \langle \gamma\,, y\rangle \geq 0 \;\text{ for all }\; \gamma \in \mathcal{G}(\mathcal{K})\} \tag{319}$$

**2.13.4.2.3 Exercise.** *Comparison with respect to orthant.*
When comparison is with respect to the nonnegative orthant $\mathcal{K}=\mathbb{R}^n_+$ , then from the *discrete membership theorem* it directly follows:

$$x \preceq z \;\Leftrightarrow\; x_i \leq z_i \;\; \forall\, i \tag{320}$$

Generate simple counterexamples demonstrating that this equivalence with entrywise inequality holds only when the underlying cone inducing partial order is the nonnegative orthant. ▼

### 2.13.5 Dual PSD cone and generalized inequality

The *dual positive semidefinite cone* $\mathcal{K}^*$ is confined to $\mathbb{S}^M$ by convention;

$$\mathbb{S}_+^{M\,*} \triangleq \{Y \in \mathbb{S}^M \mid \langle Y\,,\, X\rangle \geq 0 \;\text{ for all }\; X \in \mathbb{S}_+^M\} \;=\; \mathbb{S}_+^M \tag{321}$$

The positive semidefinite cone is self-dual in the ambient space of symmetric matrices [46, exmp.2.24] [28] [145, §II]; $\mathcal{K}=\mathcal{K}^*$.

Dual generalized inequalities with respect to the positive semidefinite cone in the ambient space of symmetric matrices can therefore be simply stated: (Fejér)

$$X \succeq 0 \;\Leftrightarrow\; \operatorname{tr}(Y^TX) \geq 0 \;\text{ for all }\; Y \succeq 0 \tag{322}$$

Membership to this cone can be determined in the isometrically isomorphic Euclidean space $\mathbb{R}^{M^2}$ via (31). (§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 (322):

$$X \succeq 0 \;\Leftrightarrow\; \langle yy^T,\, X\rangle \geq 0 \;\;\forall\, yy^T(\succeq 0) \tag{1245}$$

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 (§2.9.2.4).

#### 2.13.5.1 self-dual cones

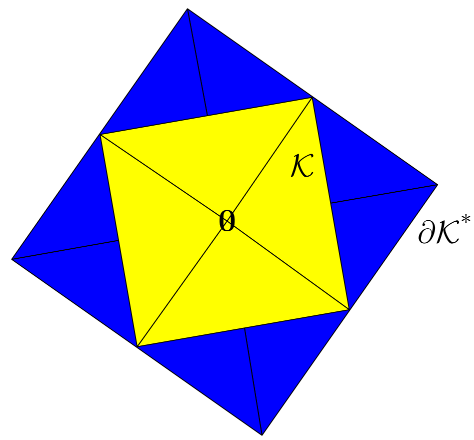
From (110) (a consequence of the *halfspaces theorem* (§2.4.1.1.1)), where the only finite value of the support function for a convex cone is 0 [148, §C.2.3.1], or from discretized definition (319) of the dual cone we get a rather self-evident characterization of self-duality:

$$\mathcal{K} = \mathcal{K}^* \quad\Leftrightarrow\quad \mathcal{K} = \bigcap_{\gamma\in\mathcal{G}(\mathcal{K})} \left\{y \mid \gamma^Ty \geq 0\right\} \tag{323}$$

In words: Cone $\mathcal{K}$ is *self-dual* iff its own extreme directions are inward-normals to a (minimal) set of hyperplanes bounding halfspaces whose intersection constructs it. This means each extreme direction of $\mathcal{K}$ is normal to a hyperplane exposing one of its own faces; a necessary but insufficient condition for self-duality (Figure **48**, for example).

Self-dual cones are of necessarily nonempty interior [25, §I] and invariant to rotation about the origin. Their most prominent representatives are the orthants, the positive semidefinite cone $\mathbb{S}^M_+$ in the ambient space of symmetric matrices (321), and the Lorentz cone (147) [17, §II.A] [46, exmp.2.25]. In three dimensions, a plane containing the axis of revolution of a self-dual cone (and the origin) will produce a *slice* whose boundary makes a right angle.

(a)




(b)

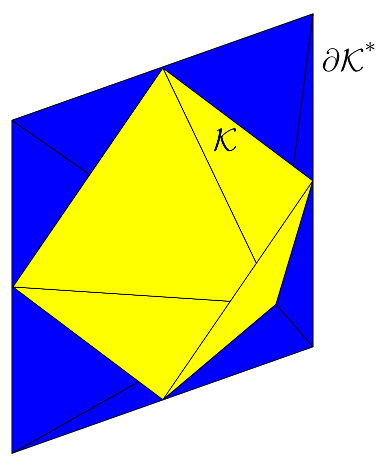



Figure 48: Two (truncated) views of a polyhedral cone $\mathcal{K}$ and its dual in $\mathbb{R}^3$. Each of four extreme directions from $\mathcal{K}$ belongs to a face of dual cone $\mathcal{K}^*$. Shrouded (inside) cone $\mathcal{K}$ is symmetrical about its axis of revolution. Each pair of diametrically opposed extreme directions from $\mathcal{K}$ makes a right angle. An orthant (or any rotation thereof; a simplicial cone) is not the only self-dual polyhedral cone in three or more dimensions; [17, §2.A.21] *e.g.*, consider an equilateral with five extreme directions. In fact, every self-dual polyhedral cone in $\mathbb{R}^3$ has an odd number of extreme directions. [19, thm.3]

**2.13.5.1.1 Example.** *Linear matrix inequality.* (*confer* §2.13.2.0.3)
Consider a peculiar vertex-description for a closed convex cone defined over the positive semidefinite cone (instead of the nonnegative orthant as in definition (83)): for $X\in\mathbb{S}^n$ given $A_j\in\mathbb{S}^n$, $j=1\ldots m$

$$\begin{aligned}\mathcal{K} &= \left\{\begin{bmatrix}\langle A_1\,,\,X\rangle\\ \vdots\\ \langle A_m\,,\,X\rangle\end{bmatrix} \mid X\succeq 0\right\}\subseteq\mathbb{R}^m\\ &= \left\{\begin{bmatrix}\operatorname{svec}(A_1)^T\\ \vdots\\ \operatorname{svec}(A_m)^T\end{bmatrix}\operatorname{svec}X \mid X\succeq 0\right\}\\ &\triangleq \{A\operatorname{svec}X \mid X\succeq 0\}\end{aligned}\tag{324}$$

where $A\in\mathbb{R}^{m\times n(n+1)/2}$, and where symmetric vectorization svec is defined in (47). $\mathcal{K}$ is indeed a convex cone because by (144)

$$A\operatorname{svec}X_{\mathrm{p}_1}\,,\,A\operatorname{svec}X_{\mathrm{p}_2}\in\mathcal{K}\;\Rightarrow\;A(\zeta\operatorname{svec}X_{\mathrm{p}_1}+\xi\operatorname{svec}X_{\mathrm{p}_2})\in\mathcal{K}\;\text{ for all }\;\zeta\,,\xi\geq 0\tag{325}$$

since a nonnegatively weighted sum of positive semidefinite matrices must be positive semidefinite. (§A.3.1.0.2) Although matrix $A$ is finite-dimensional, $\mathcal{K}$ is generally not a polyhedral cone (unless $m$ equals 1 or 2) because $X\in\mathbb{S}^n_+$. Provided the $A_j$ matrices are linearly independent, then

$$\operatorname{rel\,int}\mathcal{K}=\operatorname{int}\mathcal{K}\tag{326}$$

meaning, the cone interior is nonempty implying the dual cone is pointed by (268).

If matrix $A$ has no nullspace, on the other hand, then (by §2.10.1.1 and Definition 2.2.1.0.1) $A\operatorname{svec}X$ is an isomorphism in $X$ between the positive semidefinite cone and $\mathcal{R}(A)$. In that case, convex cone $\mathcal{K}$ has relative interior

$$\operatorname{rel\,int}\mathcal{K}=\{A\operatorname{svec}X \mid X\succ 0\}\tag{327}$$

and boundary

$$\operatorname{rel}\partial\mathcal{K}=\{A\operatorname{svec}X \mid X\succeq 0\,,\;X\nsucc 0\}\tag{328}$$

Now consider the (closed convex) dual cone:

$$\begin{aligned}\mathcal{K}^* &= \{y \mid \langle A \operatorname{svec} X\,,\; y\rangle \geq 0 \;\text{ for all }\; X \succeq 0\} \subseteq \mathbb{R}^m \\ &= \left\{y \mid \langle \operatorname{svec} X\,,\; A^T y\rangle \geq 0 \;\text{ for all }\; X \succeq 0\right\} \\ &= \left\{y \mid \; \operatorname{svec}^{-1}(A^T y) \succeq 0\right\}\end{aligned} \tag{329}$$

that follows from (322) and leads to an equally peculiar halfspace-description

$$\mathcal{K}^* = \{y \in \mathbb{R}^m \mid \sum_{j=1}^{m} y_j A_j \succeq 0\} \tag{330}$$

The summation inequality with respect to the positive semidefinite cone is known as a *linear matrix inequality.* [44] [102] [192] [272]

When the $A_j$ matrices are linearly independent, function $g(y) \triangleq \sum y_j A_j$ on $\mathbb{R}^m$ is a linear bijection. The inverse image of the positive semidefinite cone under $g(y)$ must therefore have dimension $m$. In that circumstance, the dual cone interior is nonempty

$$\operatorname{int} \mathcal{K}^* = \{y \in \mathbb{R}^m \mid \sum_{j=1}^{m} y_j A_j \succ 0\} \tag{331}$$

having boundary

$$\partial \mathcal{K}^* = \{y \in \mathbb{R}^m \mid \sum_{j=1}^{m} y_j A_j \succeq 0\,,\;\; \sum_{j=1}^{m} y_j A_j \nsucc 0\} \tag{332}$$

□

### 2.13.6 Dual of pointed polyhedral cone

In a subspace of $\mathbb{R}^n$, now we consider a pointed polyhedral cone $\mathcal{K}$ given in terms of its extreme directions $\Gamma_i$ arranged columnar in $X$;

$$X = [\, \Gamma_1 \;\; \Gamma_2 \;\cdots\; \Gamma_N \,] \in \mathbb{R}^{n \times N} \tag{240}$$

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.* (*encore*) (*confer* (252) (157)) Given pointed polyhedral cone $\mathcal{K}$ in a subspace of $\mathbb{R}^n$, denoting its $i^{\text{th}}$ extreme direction by $\Gamma_i \in \mathbb{R}^n$ arranged in a matrix $X$ as in (240), then that cone may be described: (75) (*confer* (253))

$$\begin{aligned} \mathcal{K} &= \left\{ [\,\mathbf{0}\ \ X\,]\, a\,\zeta \mid a^T\mathbf{1} = 1,\ a \succeq 0,\ \zeta \geq 0 \right\} \\ &= \left\{ X a\,\zeta \mid a^T\mathbf{1} \leq 1,\ a \succeq 0,\ \zeta \geq 0 \right\} \\ &= \left\{ X b \mid b \succeq 0 \right\} \subseteq \mathbb{R}^n \end{aligned} \tag{333}$$

that is simply a conic hull (like (83)) of a finite number $N$ of directions. △

Whenever cone $\mathcal{K}$ is pointed closed and convex (not only polyhedral), then dual cone $\mathcal{K}^*$ has a halfspace-description in terms of the extreme directions $\Gamma_i$ of $\mathcal{K}$:

$$\mathcal{K}^* = \left\{ y \mid \gamma^T y \geq 0 \ \text{ for all } \ \gamma \in \{\Gamma_i\ ,\ i=1 \ldots N\} \subseteq \operatorname{rel} \partial\mathcal{K} \right\} \tag{334}$$

because when $\{\Gamma_i\}$ constitutes any set of generators for $\mathcal{K}$, the discretization result in §2.13.4.1 allows relaxation of the requirement $\forall\, x \in \mathcal{K}$ in (258) to $\forall\, \gamma \in \{\Gamma_i\}$ directly.[2.50] That dual cone so defined is unique, identical to (258), polyhedral whenever the number of generators $N$ is finite

$$\mathcal{K}^* = \left\{ y \mid X^T y \succeq 0 \right\} \subseteq \mathbb{R}^n \tag{315}$$

and has nonempty interior because $\mathcal{K}$ is assumed pointed (but $\mathcal{K}^*$ is not necessarily pointed unless $\mathcal{K}$ has nonempty interior (§2.13.1.1)).

### 2.13.6.1 Facet normal & extreme direction

We see from (315) that the conically independent generators of cone $\mathcal{K}$ (namely, the extreme directions of pointed closed convex cone $\mathcal{K}$ constituting the columns of $X$) each define an inward-normal to a hyperplane supporting $\mathcal{K}^*$ (§2.4.2.6.1) and exposing a dual facet when $N$ is finite. Were $\mathcal{K}^*$ pointed and finitely generated, then by conjugation the dual 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 **43** or Figure **48**, for example.

[2.50] The extreme directions of $\mathcal{K}$ constitute a minimal set of generators.

We may conclude the extreme directions of polyhedral proper $\mathcal{K}$ are respectively orthogonal to the facets of $\mathcal{K}^*$; likewise, the extreme directions of polyhedral proper $\mathcal{K}^*$ 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 49 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*, [275]

$$\begin{aligned}\Gamma_4^T\Gamma_1 = \Gamma_3^T\Gamma_2 = 0\\ \Gamma_3^T\Gamma_1 \neq 0\,, \quad \Gamma_4^T\Gamma_2 \neq 0\end{aligned} \tag{335}$$

Biorthogonal expansion of $x\in\mathcal{K}$ is then

$$x \;=\; \Gamma_1\,\frac{\Gamma_3^T x}{\Gamma_3^T\Gamma_1} \;+\; \Gamma_2\,\frac{\Gamma_4^T x}{\Gamma_4^T\Gamma_2} \tag{336}$$

where $\Gamma_3^T x/(\Gamma_3^T\Gamma_1)$ is the nonnegative coefficient of nonorthogonal projection (§E.6.1) of $x$ on $\Gamma_1$ in the direction orthogonal to $\Gamma_3$, and where $\Gamma_4^T x/(\Gamma_4^T\Gamma_2)$ is the nonnegative coefficient of nonorthogonal projection of $x$ on $\Gamma_2$ in the direction orthogonal to $\Gamma_4$; they are coordinates in this nonorthogonal system. Those coefficients must be nonnegative $x\succeq_{\mathcal{K}} 0$ because $x\in\mathcal{K}$ (281) and $\mathcal{K}$ is simplicial.

If we ascribe the extreme directions of $\mathcal{K}$ to the columns of a matrix

$$X \triangleq [\,\Gamma_1 \;\; \Gamma_2\,] \tag{337}$$

then we find that the pseudoinverse transpose matrix

$$X^{\dagger T} = \left[\Gamma_3\frac{1}{\Gamma_3^T\Gamma_1} \quad \Gamma_4\frac{1}{\Gamma_4^T\Gamma_2}\right] \tag{338}$$

holds the extreme directions of the dual cone. Therefore,

$$x = XX^\dagger x \qquad\qquad (344)$$

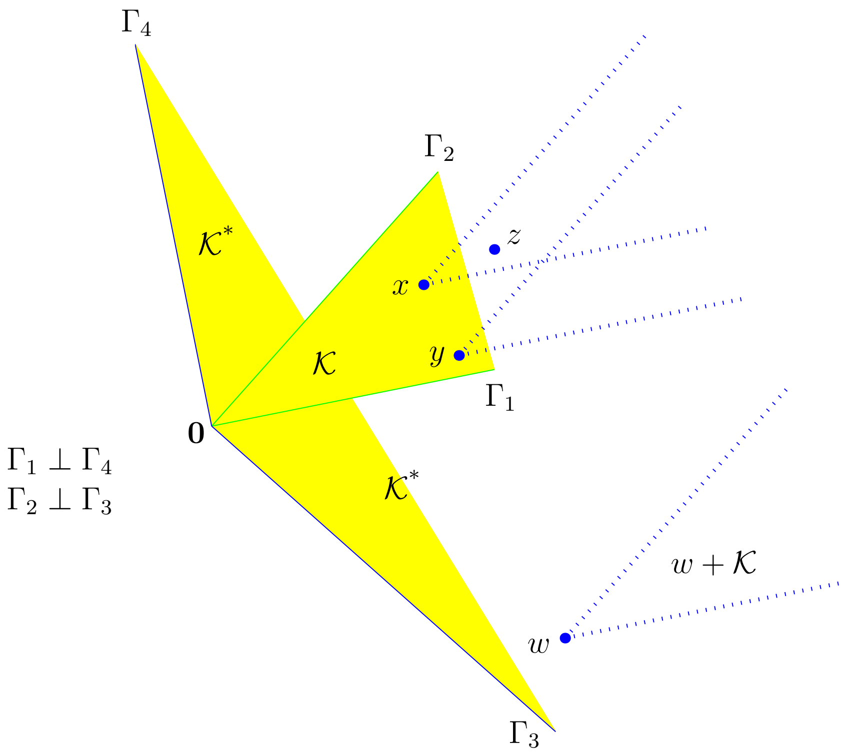



Figure 49: (*confer* Figure **119**) Simplicial cone $\mathcal{K}$ in $\mathbb{R}^2$ and its dual $\mathcal{K}^*$ drawn truncated. Conically independent generators $\Gamma_1$ and $\Gamma_2$ constitute extreme directions of $\mathcal{K}$ while $\Gamma_3$ and $\Gamma_4$ constitute extreme directions of $\mathcal{K}^*$. Dotted ray-pairs bound translated cones $\mathcal{K}$. Point $x$ is comparable to point $z$ (and *vice versa*) but not to $y$; $z \succeq x \Leftrightarrow z - x \in \mathcal{K} \Leftrightarrow z - x \succeq_{\mathcal{K}} 0$ iff $\exists$ nonnegative coordinates for biorthogonal expansion of $z - x$. Point $y$ is not comparable to $z$ because $z$ does not belong to $y \pm \mathcal{K}$. Flipping a translated cone is quite helpful for visualization: $x \preceq z \Leftrightarrow x \in z - \mathcal{K} \Leftrightarrow x - z \preceq_{\mathcal{K}} 0$. Points need not belong to $\mathcal{K}$ to be comparable; *e.g.*, all points greater than $w$ belong to $w + \mathcal{K}$.

is the biorthogonal expansion (336) (§E.0.1), and the biorthogonality condition (335) can be expressed succinctly (§E.1.1)[2.51]

$$X^\dagger X = I \tag{345}$$

Expansion $w = XX^\dagger w$ for any $w \in \mathbb{R}^2$ is unique if and only if the extreme directions of $\mathcal{K}$ are linearly independent; *id est*, iff $X$ has no nullspace.

□

#### 2.13.7.1 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 $\mathcal{K}$ (§B.1.1.1); generators because every $x \in \mathcal{K}$ is their conic combination. From §2.10.2 we know that means $\Gamma_1$ and $\Gamma_2$ must be extreme directions of $\mathcal{K}$.

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 having empty interior in §2.13.8.

**2.13.7.1.1 Example.** *Expansions implied by diagonalization.* (*confer* §6.5.3.1.1) When matrix $X \in \mathbb{R}^{M\times M}$ is *diagonalizable* (§A.5),

$$X = S\Lambda S^{-1} = [\, s_1 \cdots s_M \,]\, \Lambda \begin{bmatrix} w_1^T \\ \vdots \\ w_M^T \end{bmatrix} = \sum_{i=1}^M \lambda_i\, s_i w_i^T \tag{1339}$$

coordinates for biorthogonal expansion are its eigenvalues $\lambda_i$ (contained in diagonal matrix $\Lambda$) when expanded in $S$;

$$X = SS^{-1}X = [\, s_1 \cdots s_M \,] \begin{bmatrix} w_1^T X \\ \vdots \\ w_M^T X \end{bmatrix} = \sum_{i=1}^M \lambda_i\, s_i w_i^T \tag{339}$$

Coordinate value depend upon the geometric relationship of $X$ to its linearly independent eigenmatrices $s_i w_i^T$. (§A.5.1, §B.1.1)

[2.51] Possibly confusing is the fact that formula $XX^\dagger x$ is simultaneously the orthogonal projection of $x$ on $\mathcal{R}(X)$ (1678), and a sum of nonorthogonal projections of $x \in \mathcal{R}(X)$ on the range of each and every column of full-rank $X$ 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 $\mathcal{K}$ in a subspace of $\mathbb{R}^{M\times M}$.

When $S$ is real and $X$ belongs to that polyhedral cone $\mathcal{K}$, for example, then coordinates of expansion (the eigenvalues $\lambda_i$) must be nonnegative.

When $X = Q\Lambda Q^T$ is symmetric, coordinates for biorthogonal expansion are its eigenvalues when expanded in $Q$; *id est*, for $X\in\mathbb{S}^M$

$$X \;=\; QQ^TX \;=\; \sum_{i=1}^{M} q_i\, q_i^T X \;=\; \sum_{i=1}^{M} \lambda_i\, q_i q_i^T \;\in\; \mathbb{S}^M \tag{340}$$

becomes an orthogonal expansion with *orthonormality condition* $Q^TQ = I$ where $\lambda_i$ is the $i^{\text{th}}$ eigenvalue of $X$, $q_i$ is the corresponding $i^{\text{th}}$ eigenvector arranged columnar in orthogonal matrix

$$Q = [\, q_1 \;\; q_2 \;\; \cdots \;\; q_M \,] \in \mathbb{R}^{M\times M} \tag{341}$$

and where eigenmatrix $q_i q_i^T$ is an extreme direction of some pointed polyhedral cone $\mathcal{K}\subset\mathbb{S}^M$ and an extreme direction of the positive semidefinite cone $\mathbb{S}_+^M$.

- Orthogonal expansion is a special case of biorthogonal expansion of $X\in\operatorname{aff}\mathcal{K}$ occurring when polyhedral cone $\mathcal{K}$ is any rotation about the origin of an orthant belonging to a subspace.

Similarly, when $X = Q\Lambda Q^T$ belongs to the positive semidefinite cone in the subspace of symmetric matrices, coordinates for orthogonal expansion must be its nonnegative eigenvalues (1253) when expanded in $Q$; *id est*, for $X\in\mathbb{S}_+^M$

$$X \;=\; QQ^TX \;=\; \sum_{i=1}^{M} q_i\, q_i^T X \;=\; \sum_{i=1}^{M} \lambda_i\, q_i q_i^T \;\in\; \mathbb{S}_+^M \tag{342}$$

where $\lambda_i \geq 0$ is the $i^{\text{th}}$ eigenvalue of $X$. This means $X$ simultaneously belongs to the positive semidefinite cone and to the pointed polyhedral cone $\mathcal{K}$ formed by the conic hull of its eigenmatrices. □

**2.13.7.1.2 Example.** *Expansion respecting nonpositive orthant.*
Suppose $x \in \mathcal{K}$ any orthant in $\mathbb{R}^n$.[2.52] 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 $\mathcal{K} = \mathbb{R}^n_-$. Then the biorthogonal expansion becomes an orthogonal expansion

$$x = XX^T x = \sum_{i=1}^{n} -e_i(-e_i^T x) = \sum_{i=1}^{n} -e_i|e_i^T x| \in \mathbb{R}^n_- \tag{343}$$

and the coordinates of expansion are nonnegative. For this orthant $\mathcal{K}$ 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 $\mathcal{K}$.

Of course, this expansion $x = XX^T x$ 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 $\mathcal{K}$ are assumed to constitute the basis while those of the dual cone $\mathcal{K}^*$ determine coordinates.

Unique biorthogonal expansion with respect to $\mathcal{K}$ depends upon existence of its linearly independent extreme directions: Polyhedral cone $\mathcal{K}$ 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 $\mathcal{K}$ having possibly empty interior; *id est*, we consider a basis spanning a subspace. Then we need only observe that section of dual cone $\mathcal{K}^*$ in the affine hull of $\mathcal{K}$ because, by *expansion* of $x$, membership $x \in \operatorname{aff} \mathcal{K}$ is implicit and because any breach of the ordinary dual cone into ambient space becomes irrelevant (§2.13.9.3). *Biorthogonal expansion*

$$x = XX^\dagger x \in \operatorname{aff} \mathcal{K} = \operatorname{aff} \operatorname{cone}(X) \tag{344}$$

is expressed in the extreme directions $\{\Gamma_i\}$ of $\mathcal{K}$ arranged columnar in

$$X = [\,\Gamma_1 \;\; \Gamma_2 \;\; \cdots \;\; \Gamma_N\,] \in \mathbb{R}^{n \times N} \tag{240}$$

[2.52] An orthant is simplicial and self-dual.

under assumption of *biorthogonality*

$$X^\dagger X = I \tag{345}$$

where $^\dagger$ denotes matrix pseudoinverse (§E). We therefore seek, in this section, a vertex-description for $\mathcal{K}^* \cap \operatorname{aff} \mathcal{K}$ in terms of linearly independent dual generators $\{\Gamma_i^*\} \subset \operatorname{aff} \mathcal{K}$ in the same finite quantity[2.53] as the extreme directions $\{\Gamma_i\}$ of

$$\mathcal{K} = \operatorname{cone}(X) = \{Xa \mid a \succeq 0\} \subseteq \mathbb{R}^n \tag{252}$$

We assume the quantity of extreme directions $N$ does not exceed the dimension $n$ of ambient vector space because, otherwise, the expansion could not be unique; *id est*, assume $N$ linearly independent extreme directions hence $N \leq n$ ($X$ *skinny*[2.54]-or-square full-rank). In other words, fat full-rank matrix $X$ is prohibited by uniqueness because of the 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 $x \in \mathcal{K}$

Suppose $x$ belongs to $\mathcal{K} \subseteq \mathbb{R}^n$. Then $x = Xa$ for some $a \succeq 0$. Vector $a$ is unique only when $\{\Gamma_i\}$ is a linearly independent set.[2.55] Vector $a \in \mathbb{R}^N$ can take the form $a = Bx$ if $\mathcal{R}(B) = \mathbb{R}^N$. Then we require $Xa = XBx = x$ and $Bx = BXa = a$. The pseudoinverse $B = X^\dagger \in \mathbb{R}^{N \times n}$ (§E) is suitable when $X$ is skinny-or-square and full-rank. In that case $\operatorname{rank} X = N$, and for all $c \succeq 0$ and $i = 1 \ldots N$

$$a \succeq 0 \;\Leftrightarrow\; X^\dagger X a \succeq 0 \;\Leftrightarrow\; a^T X^T X^{\dagger T} c \geq 0 \;\Leftrightarrow\; \Gamma_i^T X^{\dagger T} c \geq 0 \tag{346}$$

The penultimate inequality follows from the *generalized inequality and membership corollary*, while the last inequality is a consequence of that

[2.53] When $\mathcal{K}$ is contained in a proper subspace of $\mathbb{R}^n$, the ordinary dual cone $\mathcal{K}^*$ will have more generators in any minimal set than $\mathcal{K}$ has extreme directions.

[2.54] "Skinny" meaning thin; more rows than columns.

[2.55] Conic independence alone (§2.10) is insufficient to guarantee uniqueness.

corollary's discretization (§2.13.4.2.1).[2.56] From (346) and (334) we deduce

$$\mathcal{K}^* \cap \operatorname{aff} \mathcal{K} \;=\; \operatorname{cone}(X^{\dagger T}) \;=\; \{X^{\dagger T} c \mid c \succeq 0\} \subseteq \mathbb{R}^n \tag{347}$$

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 (268), we know $\mathcal{K}^* \cap \operatorname{aff} \mathcal{K}$ must be pointed if $\operatorname{rel\,int} \mathcal{K}$ is logically assumed nonempty with respect to $\operatorname{aff} \mathcal{K}$.

Conversely, suppose full-rank skinny-or-square matrix

$$X^{\dagger T} \triangleq \begin{bmatrix} \Gamma_1^* & \Gamma_2^* & \cdots & \Gamma_N^* \end{bmatrix} \in \mathbb{R}^{n \times N} \tag{348}$$

comprises the extreme directions $\{\Gamma_i^*\} \subset \operatorname{aff} \mathcal{K}$ of the dual-cone intersection with the affine hull of $\mathcal{K}$.[2.57] From the *discrete membership theorem* and (272) we get a partial dual to (334); *id est*, assuming $x \in \operatorname{aff} \operatorname{cone} X$

$$x \in \mathcal{K} \;\Leftrightarrow\; \gamma^{*T} x \geq 0 \;\text{ for all }\; \gamma^* \in \left\{\Gamma_i^*,\; i = 1 \ldots N\right\} \subset \partial \mathcal{K}^* \cap \operatorname{aff} \mathcal{K} \tag{349}$$

$$\Leftrightarrow\; X^{\dagger} x \succeq 0 \tag{350}$$

that leads to a partial halfspace-description,

$$\mathcal{K} \;=\; \left\{x \in \operatorname{aff} \operatorname{cone} X \mid X^{\dagger} x \succeq 0\right\} \tag{351}$$

For $\gamma^* = X^{\dagger T} e_i$, any $x = Xa$, and for all $i$ we have $e_i^T X^{\dagger} X a = e_i^T a \geq 0$ only when $a \succeq 0$. Hence $x \in \mathcal{K}$.

---

[2.56]

$$\begin{aligned} a \succeq 0 \;\Leftrightarrow\; a^T X^T X^{\dagger T} c \geq 0 \quad & \forall\, (c \succeq 0 \;\Leftrightarrow\; a^T X^T X^{\dagger T} c \geq 0 \quad \forall\, a \succeq 0) \\ & \forall\, (c \succeq 0 \;\Leftrightarrow\; \Gamma_i^T X^{\dagger T} c \geq 0 \quad\;\; \forall\, i\,) \qquad \blacklozenge \end{aligned}$$

Intuitively, any nonnegative vector $a$ is a conic combination of the standard basis $\{e_i \in \mathbb{R}^N\}$; $a \succeq 0 \Leftrightarrow a_i e_i \succeq 0$ for all $i$. The last inequality in (346) 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; $a = e_i \succeq 0$. Theoretically, because $c \succeq 0$ defines a pointed polyhedral cone (in fact, the nonnegative orthant in $\mathbb{R}^N$), we can take (346) one step further by discretizing $c$:

$$a \succeq 0 \;\Leftrightarrow\; \Gamma_i^T \Gamma_j^* \geq 0 \text{ for } i, j = 1 \ldots N \;\Leftrightarrow\; X^{\dagger} X \geq \mathbf{0}$$

In words, $X^{\dagger} X$ must be a matrix whose entries are each nonnegative.

[2.57] When closed convex cone $\mathcal{K}$ has empty interior, $\mathcal{K}^*$ has no extreme directions.

When $X$ is full-rank, then the unique biorthogonal expansion of $x \in \mathcal{K}$ becomes (344)

$$x \;=\; XX^{\dagger}x \;=\; \sum_{i=1}^{N} \Gamma_i \, \Gamma_i^{*T} x \tag{352}$$

whose *coordinates* $\Gamma_i^{*T} x$ must be nonnegative because $\mathcal{K}$ 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 $\mathcal{K}^* \cap \operatorname{aff} \mathcal{K}$ ; hence, the columns constitute its extreme directions. (§2.10) That section of the dual cone is itself a polyhedral cone (by (246) or the *cone intersection theorem*, §2.7.2.1.1) having the same number of extreme directions as $\mathcal{K}$ .

#### 2.13.8.2 $x \in \operatorname{aff} \mathcal{K}$

The extreme directions of $\mathcal{K}$ and $\mathcal{K}^* \cap \operatorname{aff} \mathcal{K}$ 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, precisely, a biorthogonality condition, [275, §2.2.4] [150] implying each set of extreme directions is linearly independent. (§B.1.1.1)

The biorthogonal expansion therefore applies more broadly; meaning, for any $x \in \operatorname{aff} \mathcal{K}$ , vector $x$ can be uniquely expressed $x = Xb$ where $b \in \mathbb{R}^N$ because $\operatorname{aff} \mathcal{K}$ contains the origin. Thus, for any such $x \in \mathcal{R}(X)$ (*confer* §E.1.1), biorthogonal expansion (352) becomes $x = XX^{\dagger}Xb = Xb$ .

### 2.13.9 Formulae, algorithm finding dual cone

#### 2.13.9.1 Pointed $\mathcal{K}$ , dual, $X$ skinny-or-square full-rank

We wish to derive expressions for a convex cone and its ordinary dual under the general assumptions: pointed polyhedral $\mathcal{K}$ denoted by its linearly independent extreme directions arranged columnar in matrix $X$ such that

$$\operatorname{rank}(X \in \mathbb{R}^{n \times N}) \;=\; N \;\triangleq\; \dim \operatorname{aff} \mathcal{K} \;\leq\; n \tag{353}$$

The vertex-description is given:

$$\mathcal{K} = \{Xa \mid a \succeq 0\} \subseteq \mathbb{R}^n \tag{354}$$

from which a halfspace-description for the dual cone follows directly:

$$\mathcal{K}^* = \{y \in \mathbb{R}^n \mid X^T y \succeq 0\} \tag{355}$$

By defining a matrix

$$X^\perp \triangleq \text{basis}\,\mathcal{N}(X^T) \tag{356}$$

(a columnar basis for the orthogonal complement of $\mathcal{R}(X)$), we can say

$$\text{aff cone}\, X \;=\; \text{aff}\,\mathcal{K} \;=\; \{x \mid X^{\perp T} x = \mathbf{0}\} \tag{357}$$

meaning $\mathcal{K}$ lies in a subspace, perhaps $\mathbb{R}^n$. Thus we have a halfspace-description

$$\mathcal{K} = \{x \in \mathbb{R}^n \mid X^\dagger x \succeq 0\,,\;\; X^{\perp T} x = \mathbf{0}\} \tag{358}$$

and from (272), a vertex-description[2.58]

$$\mathcal{K}^* = \{\, [\, X^{\dagger T} \;\; X^\perp \;\; -X^\perp \,] b \mid b \succeq 0 \,\} \subseteq \mathbb{R}^n \tag{359}$$

These results are summarized for a pointed polyhedral cone, having linearly independent generators, and its ordinary dual:

| **Cone Table 1** | $\mathcal{K}$ | $\mathcal{K}^*$ |
|---|---|---|
| vertex-description | $X$ | $X^{\dagger T},\ \pm X^\perp$ |
| halfspace-description | $X^\dagger,\ X^{\perp T}$ | $X^T$ |

#### 2.13.9.2 Simplicial case

When a convex cone is simplicial (§2.12.3), Cone Table **1** simplifies because then $\text{aff cone}\, X = \mathbb{R}^n$: For square $X$ and assuming simplicial $\mathcal{K}$ such that

$$\text{rank}(X \in \mathbb{R}^{n\times N}) \;=\; N \;\triangleq\; \dim\text{aff}\,\mathcal{K} \;=\; n \tag{360}$$

we have

| **Cone Table S** | $\mathcal{K}$ | $\mathcal{K}^*$ |
|---|---|---|
| vertex-description | $X$ | $X^{\dagger T}$ |
| halfspace-description | $X^\dagger$ | $X^T$ |

[2.58]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) when only three would suffice.

For example, vertex-description (359) simplifies to

$$\mathcal{K}^* = \{X^{\dagger T} b \mid b \succeq 0\} \subset \mathbb{R}^n \tag{361}$$

Now, because $\dim \mathcal{R}(X) = \dim \mathcal{R}(X^{\dagger T})$ , (§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 (258) of the ordinary dual cone $\mathcal{K}^*$ in ambient vector space $\mathcal{R}$ that its determination instead in subspace $\mathcal{M} \subseteq \mathcal{R}$ is identical to its intersection with $\mathcal{M}$ ; *id est*, assuming closed convex cone $\mathcal{K} \subseteq \mathcal{M}$ and $\mathcal{K}^* \subseteq \mathcal{R}$

$$(\mathcal{K}^* \text{ were ambient } \mathcal{M}) \equiv (\mathcal{K}^* \text{ in ambient } \mathcal{R}) \cap \mathcal{M} \tag{362}$$

because

$$\big\{y \in \mathcal{M} \mid \langle y\,,\, x\rangle \geq 0 \text{ for all } x \in \mathcal{K}\big\} = \big\{y \in \mathcal{R} \mid \langle y\,,\, x\rangle \geq 0 \text{ for all } x \in \mathcal{K}\big\} \cap \mathcal{M} \tag{363}$$

From this, a constrained membership relation for the ordinary dual cone $\mathcal{K}^* \subseteq \mathcal{R}$ , assuming $x, y \in \mathcal{M}$ and closed convex cone $\mathcal{K} \subseteq \mathcal{M}$

$$y \in \mathcal{K}^* \cap \mathcal{M} \;\Leftrightarrow\; \langle y\,,\, x\rangle \geq 0 \text{ for all } x \in \mathcal{K} \tag{364}$$

By closure in subspace $\mathcal{M}$ we have conjugation (§2.13.1.1):

$$x \in \mathcal{K} \;\Leftrightarrow\; \langle y\,,\, x\rangle \geq 0 \text{ for all } y \in \mathcal{K}^* \cap \mathcal{M} \tag{365}$$

This means membership determination in subspace $\mathcal{M}$ requires knowledge of the dual cone only in $\mathcal{M}$. For sake of completeness, for proper cone $\mathcal{K}$ with respect to subspace $\mathcal{M}$ (*confer* (282))

$$x \in \operatorname{int} \mathcal{K} \;\Leftrightarrow\; \langle y\,,\, x\rangle > 0 \text{ for all } y \in \mathcal{K}^* \cap \mathcal{M}\,,\; y \neq \mathbf{0} \tag{366}$$

$$x \in \mathcal{K}\,,\; x \neq \mathbf{0} \;\Leftrightarrow\; \langle y\,,\, x\rangle > 0 \text{ for all } y \in \operatorname{int} \mathcal{K}^* \cap \mathcal{M} \tag{367}$$

(By conjugation, we also have the dual relations.) Yet when $\mathcal{M}$ equals $\operatorname{aff} \mathcal{K}$ for $\mathcal{K}$ a closed convex cone

$$x \in \operatorname{rel\,int} \mathcal{K} \;\Leftrightarrow\; \langle y\,,\, x\rangle > 0 \text{ for all } y \in \mathcal{K}^* \cap \operatorname{aff} \mathcal{K}\,,\; y \neq \mathbf{0} \tag{368}$$

$$x \in \mathcal{K}\,,\; x \neq \mathbf{0} \;\Leftrightarrow\; \langle y\,,\, x\rangle > 0 \text{ for all } y \in \operatorname{rel\,int}(\mathcal{K}^* \cap \operatorname{aff} \mathcal{K}) \tag{369}$$

#### 2.13.9.4 Subspace $\mathcal{M} = \operatorname{aff} \mathcal{K}$

Assume now a subspace $\mathcal{M}$ 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

$$\operatorname{rank}(X \in \mathbb{R}^{n\times N}) \;=\; N \;\triangleq\; \dim \operatorname{aff} \mathcal{K} \;\leq\; n \tag{353}$$

We want expressions for the convex cone and its dual in subspace $\mathcal{M} = \operatorname{aff} \mathcal{K}$:

| **Cone Table A** | $\mathcal{K}$ | $\mathcal{K}^* \cap \operatorname{aff} \mathcal{K}$ |
|---|---|---|
| vertex-description | $X$ | $X^{\dagger T}$ |
| halfspace-description | $X^{\dagger}\,,\; X^{\perp T}$ | $X^{T},\; X^{\perp T}$ |

When $\dim \operatorname{aff} \mathcal{K} = n$, 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.[2.59]

**2.13.9.4.1 Example.** *Monotone nonnegative cone.*
[46, exer.2.33] [264, §2] Simplicial cone (§2.12.3.1.1) $\mathcal{K}_{\mathcal{M}+}$ is the cone of all nonnegative vectors having their entries sorted in nonincreasing order:

$$\begin{aligned}\mathcal{K}_{\mathcal{M}+} &\triangleq \{x \mid x_1 \geq x_2 \geq \cdots \geq x_n \geq 0\} \subseteq \mathbb{R}^n_+ \\ &= \{x \mid (e_i - e_{i+1})^T x \geq 0,\;\; i = 1 \ldots n{-}1,\;\; e_n^T x \geq 0\} \\ &= \{x \mid X^{\dagger} x \succeq 0\}\end{aligned} \tag{370}$$

a halfspace-description where $e_i$ is the $i^{\text{th}}$ standard basis vector, and where

$$X^{\dagger T} \triangleq [\, e_1 - e_2 \quad e_2 - e_3 \quad \cdots \quad e_n \,] \in \mathbb{R}^{n\times n} \tag{371}$$

(With $X^{\dagger}$ in hand, we might concisely scribe the remaining vertex and halfspace-descriptions from the tables for $\mathcal{K}_{\mathcal{M}+}$ and its dual. Instead we use dual generalized inequalities in their derivation.) For any vectors $x$ and $y$, simple algebra demands

$$\begin{aligned} x^T y \;=\; \sum_{i=1}^{n} x_i y_i \;=\; &(x_1 - x_2) y_1 + (x_2 - x_3)(y_1 + y_2) + (x_3 - x_4)(y_1 + y_2 + y_3) + \cdots \\ &+ (x_{n-1} - x_n)(y_1 + \cdots + y_{n-1}) + x_n (y_1 + \cdots + y_n) \end{aligned} \tag{372}$$

[2.59] 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 $\mathbb{R}^{N\times N}$.

(a)

(b)

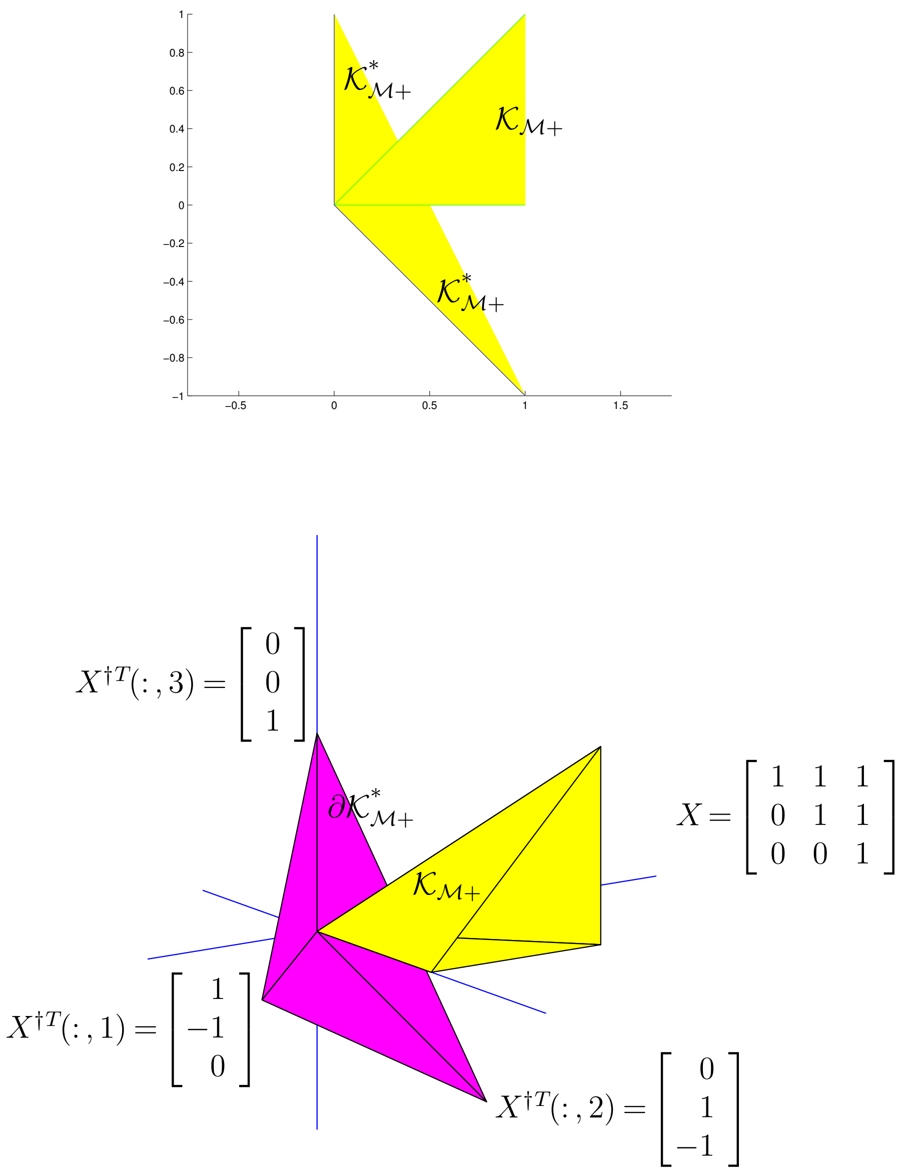



Figure 50: Simplicial cones. **(a)** Monotone nonnegative cone $\mathcal{K}_{\mathcal{M}+}$ and its dual $\mathcal{K}^*_{\mathcal{M}+}$ (drawn truncated) in $\mathbb{R}^{\mathbf{2}}$. **(b)** Monotone nonnegative cone and boundary of its dual (both drawn truncated) in $\mathbb{R}^{\mathbf{3}}$. Extreme directions of $\mathcal{K}^*_{\mathcal{M}+}$ are indicated.

Because $x_i - x_{i+1} \geq 0 \;\forall\, i$ by assumption whenever $x \in \mathcal{K}_{\mathcal{M}+}$ , we can employ dual generalized inequalities (279) with respect to the self-dual nonnegative orthant $\mathbb{R}^n_+$ to find the halfspace-description of the dual cone $\mathcal{K}^*_{\mathcal{M}+}$ . We can say $x^T y \geq 0$ for all $X^\dagger x \succeq 0$ [*sic*] if and only if

$$y_1 \geq 0\,, \quad y_1 + y_2 \geq 0\,, \quad \ldots\,, \quad y_1 + y_2 + \cdots + y_n \geq 0 \tag{373}$$

*id est*,

$$x^T y \geq 0 \;\; \forall\, X^\dagger x \succeq 0 \quad \Leftrightarrow \quad X^T y \succeq 0 \tag{374}$$

where

$$X = [\, e_1 \quad e_1 + e_2 \quad e_1 + e_2 + e_3 \quad \cdots \quad \mathbf{1}\,] \in \mathbb{R}^{n \times n} \tag{375}$$

Because $X^\dagger x \succeq 0$ connotes membership of $x$ to pointed $\mathcal{K}_{\mathcal{M}+}$ , then by (258) the dual cone we seek comprises all $y$ for which (374) holds; thus its halfspace-description

$$\mathcal{K}^*_{\mathcal{M}+} = \{y \underset{\mathcal{K}^*_{\mathcal{M}+}}{\succeq} 0\} = \{y \mid \textstyle\sum_{i=1}^k y_i \geq 0\,,\; k = 1 \ldots n\} = \{y \mid X^T y \succeq 0\} \subset \mathbb{R}^n \tag{376}$$

The monotone nonnegative cone and its dual are simplicial, illustrated for two Euclidean spaces in Figure **50**.

From §2.13.6.1, the extreme directions of proper $\mathcal{K}_{\mathcal{M}+}$ are respectively orthogonal to the facets of $\mathcal{K}^*_{\mathcal{M}+}$ . Because $\mathcal{K}^*_{\mathcal{M}+}$ is simplicial, the inward-normals to its facets constitute the linearly independent rows of $X^T$ by (376). Hence the vertex-description for $\mathcal{K}_{\mathcal{M}+}$ employs the columns of $X$ in agreement with Cone Table **S** because $X^\dagger = X^{-1}$. Likewise, the extreme directions of proper $\mathcal{K}^*_{\mathcal{M}+}$ are respectively orthogonal to the facets of $\mathcal{K}_{\mathcal{M}+}$ whose inward-normals are contained in the rows of $X^\dagger$ by (370). So the vertex-description for $\mathcal{K}^*_{\mathcal{M}+}$ employs the columns of $X^{\dagger T}$. □

**2.13.9.4.2 Example.** *Monotone cone.*
(Figure **51**, Figure **52**) Of nonempty interior but not pointed, the monotone cone is polyhedral and defined by the halfspace-description

$$\mathcal{K}_{\mathcal{M}} \triangleq \{x \in \mathbb{R}^n \mid x_1 \geq x_2 \geq \cdots \geq x_n\} = \{x \in \mathbb{R}^n \mid X^{*T} x \succeq 0\} \tag{377}$$

Its dual is therefore pointed but of empty interior, having vertex-description

$$\mathcal{K}^*_{\mathcal{M}} = \{\, X^* b \triangleq [\, e_1 - e_2 \;\; e_2 - e_3 \;\; \cdots \;\; e_{n-1} - e_n \,]\, b \mid b \succeq 0 \,\} \subset \mathbb{R}^n \tag{378}$$

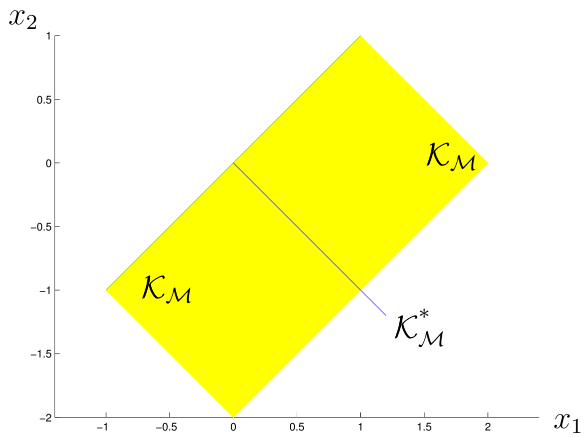



Figure 51: Monotone cone $\mathcal{K}_{\mathcal{M}}$ and its dual $\mathcal{K}_{\mathcal{M}}^*$ (drawn truncated) in $\mathbb{R}^2$.

where the columns of $X^*$ comprise the extreme directions of $\mathcal{K}_{\mathcal{M}}^*$. Because $\mathcal{K}_{\mathcal{M}}^*$ is pointed and satisfies

$$\operatorname{rank}(X^* \in \mathbb{R}^{n\times N}) \;=\; N \;\triangleq\; \dim\operatorname{aff}\mathcal{K}^* \;\leq\; n \tag{379}$$

where $N = n-1$, and because $\mathcal{K}_{\mathcal{M}}$ is closed and convex, we may adapt Cone Table **1** as follows:

| **Cone Table 1*** | $\mathcal{K}^*$ | $\mathcal{K}^{**} = \mathcal{K}$ |
|---|---|---|
| vertex-description | $X^*$ | $X^{*\dagger T},\ \pm X^{*\perp}$ |
| halfspace-description | $X^{*\dagger},\ X^{*\perp T}$ | $X^{*T}$ |

The vertex-description for $\mathcal{K}_{\mathcal{M}}$ is therefore

$$\mathcal{K}_{\mathcal{M}} = \{[\,X^{*\dagger T} \quad X^{*\perp} \quad -X^{*\perp}\,]a \mid a \succeq 0\} \subset \mathbb{R}^n \tag{380}$$

where $X^{*\perp} = \mathbf{1}$ and

$$X^{*\dagger} = \frac{1}{n}\begin{bmatrix} n-1 & -1 & -1 & \cdots & -1 & -1 & -1 \\ n-2 & n-2 & -2 & \ddots & \cdots & -2 & -2 \\ \vdots & n-3 & n-3 & \ddots & -(n-4) & \vdots & -3 \\ 3 & \vdots & n-4 & \ddots & -(n-3) & -(n-3) & \vdots \\ 2 & 2 & \cdots & \ddots & 2 & -(n-2) & -(n-2) \\ 1 & 1 & 1 & \cdots & 1 & 1 & -(n-1) \end{bmatrix} \in \mathbb{R}^{n-1\times n} \tag{381}$$

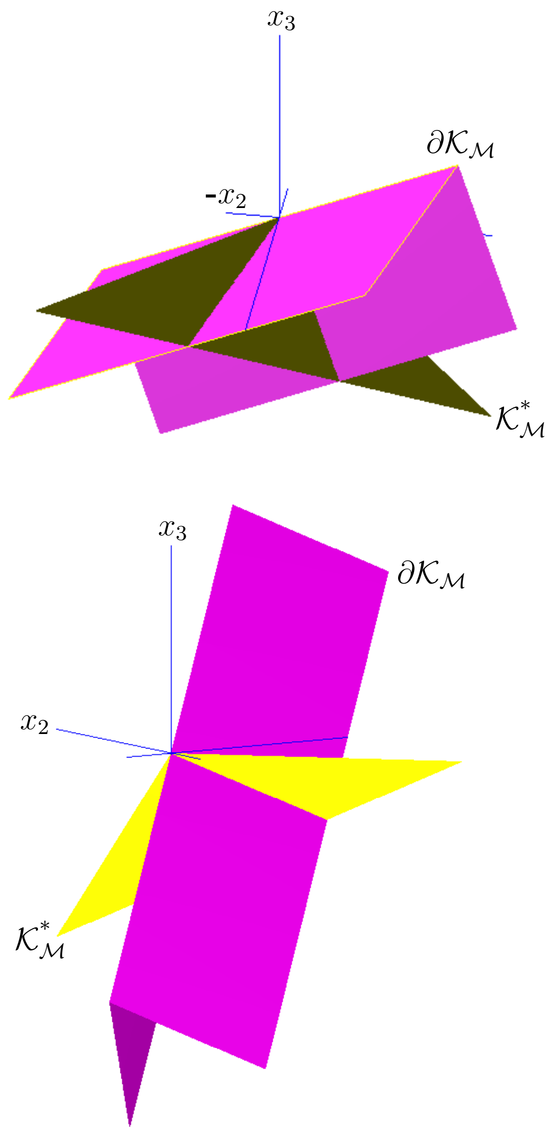



Figure 52: Two views of monotone cone $\mathcal{K}_\mathcal{M}$ and its dual $\mathcal{K}^*_\mathcal{M}$ (drawn truncated) in $\mathbb{R}^3$. Monotone cone is not pointed. Dual monotone cone has empty interior. Cartesian coordinate axes are drawn for reference.

while

$$\mathcal{K}_{\mathcal{M}}^* = \{y \in \mathbb{R}^n \mid X^{*\dagger} y \succeq 0\,,\; X^{*\perp T} y = \mathbf{0}\} \tag{382}$$

is the dual monotone cone halfspace-description. □

#### 2.13.9.5 More pointed cone descriptions with equality condition

Consider pointed polyhedral cone $\mathcal{K}$ having a linearly independent set of generators and whose subspace membership is explicit; *id est*, we are given the ordinary halfspace-description

$$\mathcal{K} = \{x \mid Ax \succeq 0\,,\; Cx = \mathbf{0}\} \subseteq \mathbb{R}^n \qquad \text{(246a)}$$

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 $\xi$ :

$$\mathcal{K} = \{Z\xi \in \mathbb{R}^n \mid AZ\xi \succeq 0\} \tag{383}$$

where $\mathcal{R}(Z \in \mathbb{R}^{n\times n-\operatorname{rank} C}) \triangleq \mathcal{N}(C)$. Assuming (353) is satisfied

$$\operatorname{rank} X \triangleq \operatorname{rank}\big((AZ)^\dagger \in \mathbb{R}^{n-\operatorname{rank} C\times m}\big) = m - \ell = \dim\operatorname{aff}\mathcal{K} \le n - \operatorname{rank} C \tag{384}$$

where $\ell$ is the number of conically dependent rows in $AZ$ (§2.10) that must be removed to make $\hat{A}Z$ before the cone tables become applicable.[2.60] Then the results collected in the cone tables admit the assignment $\hat{X} \triangleq (\hat{A}Z)^\dagger \in \mathbb{R}^{n-\operatorname{rank} C\times m-\ell}$, where $\hat{A} \in \mathbb{R}^{m-\ell\times n}$, followed with linear transformation by $Z$. So we get the vertex-description, for $(\hat{A}Z)^\dagger$ skinny-or-square full-rank,

$$\mathcal{K} = \{Z(\hat{A}Z)^\dagger b \mid b \succeq 0\} \tag{385}$$

From this and (315) we get a halfspace-description of the dual cone

$$\mathcal{K}^* = \{y \in \mathbb{R}^n \mid (Z^T\hat{A}^T)^\dagger Z^T y \succeq 0\} \tag{386}$$

[2.60] When the conically dependent rows are removed, the rows remaining must be linearly independent for the cone tables to apply.

From this and Cone Table **1** (p.167) we get a vertex-description, (1642)

$$\mathcal{K}^* = \{[\, Z^{\dagger T}(\hat{A}Z)^T \quad C^T \quad -C^T\,]c \mid c \succeq 0\} \tag{387}$$

Yet because

$$\mathcal{K} = \{x \mid Ax \succeq 0\} \cap \{x \mid Cx = \mathbf{0}\} \tag{388}$$

then, by (272), we get an equivalent vertex-description for the dual cone

$$\begin{aligned}\mathcal{K}^* &= \overline{\{x \mid Ax \succeq 0\}^* + \{x \mid Cx = \mathbf{0}\}^*} \\ &= \{[\, A^T \quad C^T \quad -C^T\,]b \mid b \succeq 0\}\end{aligned} \tag{389}$$

from which the conically dependent columns may, of course, be removed.

### 2.13.10 Dual cone-translate

First-order optimality condition (308) inspires a dual-cone variant: For any set $\mathcal{K}$, the negative dual of its translation by any $a \in \mathbb{R}^n$ is

$$\begin{aligned}-(\mathcal{K}-a)^* &\triangleq \left\{y \in \mathbb{R}^n \mid \langle y\,,\, x-a\rangle \leq 0 \;\text{ for all }\; x \in \mathcal{K}\right\} \\ &= \left\{y \in \mathbb{R}^n \mid \langle y\,,\, x\rangle \leq 0 \;\text{ for all }\; x \in \mathcal{K}-a\right\}\end{aligned} \tag{390}$$

a closed convex cone called the *normal cone* to $\mathcal{K}$ at point $a$. (§E.10.3.2.1) From this, a new membership relation like (276) for closed convex cone $\mathcal{K}$:

$$y \in -(\mathcal{K}-a)^* \;\Leftrightarrow\; \langle y\,,\, x-a\rangle \leq 0 \;\text{ for all }\; x \in \mathcal{K} \tag{391}$$

#### 2.13.10.1 first-order optimality condition - restatement

The general first-order necessary and sufficient condition for optimality of solution $x^\star$ to a minimization problem with real differentiable convex objective function $f(x): \mathbb{R}^n \to \mathbb{R}$ over convex feasible set $\mathcal{C}$ is [229, §3] (*confer* (308))

$$-\nabla f(x^\star) \in -(\mathcal{C}-x^\star)^*\,, \qquad x^\star \in \mathcal{C} \tag{392}$$

*id est*, the negative gradient (§3.1.8) belongs to the normal cone at $x^\star$ as in Figure **53**.

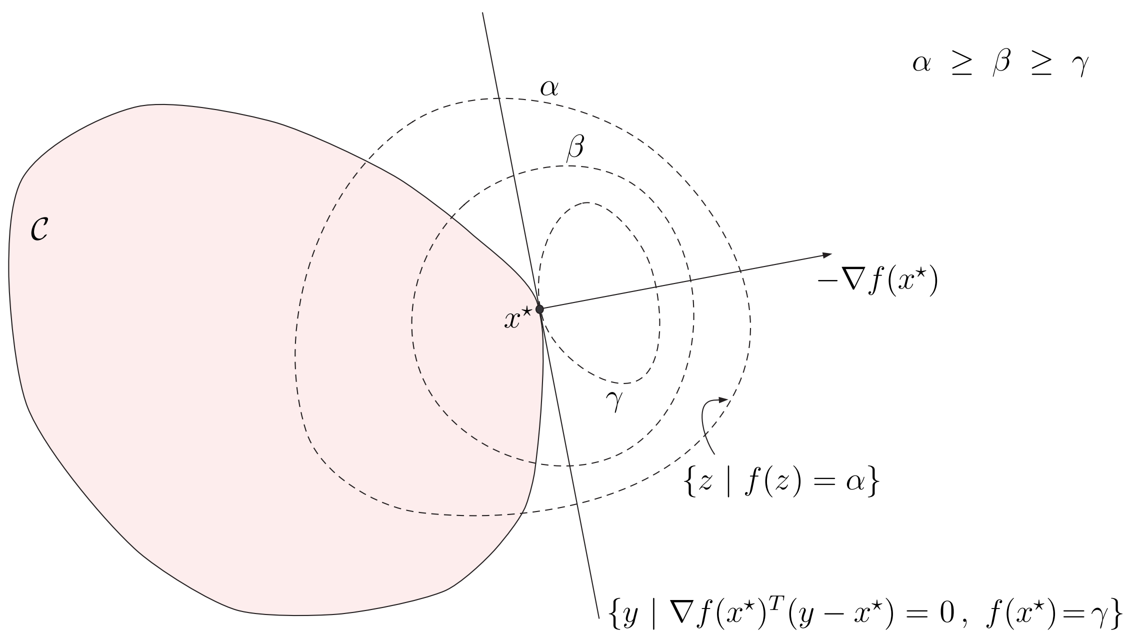



Figure 53: Shown is a plausible contour plot in $\mathbb{R}^2$ of some arbitrary differentiable real convex function $f(x)$ at selected levels $\alpha$, $\beta$, and $\gamma$; *id est*, contours of equal level $f$ (level sets) drawn (dashed) in function's domain. Function is minimized over convex set $\mathcal{C}$ at point $x^\star$ iff negative gradient $-\nabla f(x^\star)$ belongs to normal cone to $\mathcal{C}$ there. In circumstance depicted, normal cone is a ray whose direction is coincident with negative gradient. From results in §3.1.9 (p.211), $\nabla f(x^\star)$ is normal to the $\gamma$-sublevel set by Definition E.9.1.0.1.

**2.13.10.1.1 Example.** *Normal cone to orthant.*
Consider proper cone $\mathcal{K}=\mathbb{R}^n_+$ , the self-dual nonnegative orthant in $\mathbb{R}^n$. The normal cone to $\mathbb{R}^n_+$ at $a\in\mathcal{K}$ is (1860)

$$\mathcal{K}^{\perp}_{\mathbb{R}^n_+}(a\in\mathbb{R}^n_+) = -(\mathbb{R}^n_+ - a)^* = -\mathbb{R}^n_+ \cap a^{\perp} \;, \quad a\in\mathbb{R}^n_+ \tag{393}$$

where $-\mathbb{R}^n_+ = -\mathcal{K}^*$ is the algebraic complement of $\mathbb{R}^n_+$ , and $a^{\perp}$ is the orthogonal complement of point $a$ . This means: When point $a$ is interior to $\mathbb{R}^n_+$ , the normal cone is the origin. If $n_p$ represents the number of nonzero entries in point $a \in \partial\mathbb{R}^n_+$ , then $\dim(-\mathbb{R}^n_+ \cap a^{\perp}) = n - n_p$ and there is a complementary relationship between the nonzero entries in point $a$ and the nonzero entries in any vector $x\in -\mathbb{R}^n_+ \cap a^{\perp}$. □

**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 $\mathbb{R}^n$ ;

$$\begin{array}{cl}\underset{x}{\text{minimize}} & f(x)\\ \text{subject to} & x \in \mathcal{K}\end{array} \tag{394}$$

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 fat full-rank matrix $C\in\mathbb{R}^{p\times n}$ ; *id est*, we are given the halfspace-description

$$\mathcal{K} \;=\; \{x \mid Ax \succeq 0\,,\; Cx = \mathbf{0}\} \subseteq \mathbb{R}^n \tag{246a}$$

where $A\in\mathbb{R}^{m\times n}$. The vertex-description of this cone, assuming $(\hat{A}Z)^{\dagger}$ skinny-or-square full-rank, is

$$\mathcal{K} = \{Z(\hat{A}Z)^{\dagger} b \mid b \succeq 0\} \tag{385}$$

where $\hat{A}\in\mathbb{R}^{m-\ell\times n}$, $\ell$ is the number of conically dependent rows in $AZ$ (§2.10) that must be removed, and $Z\in\mathbb{R}^{n\times n-\operatorname{rank} C}$ holds basis $\mathcal{N}(C)$ columnar.

From optimality condition (308),

$$\nabla f(x^\star)^T \big(Z(\hat{A}Z)^\dagger b - x^\star\big) \geq 0 \quad \forall\, b \succeq 0 \tag{395}$$

$$-\nabla f(x^\star)^T Z(\hat{A}Z)^\dagger (b - b^\star) \leq 0 \quad \forall\, b \succeq 0 \tag{396}$$

because

$$x^\star \triangleq Z(\hat{A}Z)^\dagger b^\star \in \mathcal{K} \tag{397}$$

From membership relation (391) and Example 2.13.10.1.1

$$\begin{array}{c} \langle -(Z^T\hat{A}^T)^\dagger Z^T \nabla f(x^\star)\,,\, b - b^\star\rangle \leq 0 \;\text{ for all }\; b \in \mathbb{R}_+^{m-\ell} \\ \Leftrightarrow \\ -(Z^T\hat{A}^T)^\dagger Z^T \nabla f(x^\star) \in -\mathbb{R}_+^{m-\ell} \cap b^{\star\perp} \end{array} \tag{398}$$

Then the equivalent necessary and sufficient conditions for optimality of the conic program (394) with pointed polyhedral feasible set $\mathcal{K}$ are: (*confer* (314))

$$(Z^T\hat{A}^T)^\dagger Z^T \nabla f(x^\star) \underset{\mathbb{R}_+^{m-\ell}}{\succeq} 0\,, \qquad b^\star \underset{\mathbb{R}_+^{m-\ell}}{\succeq} 0\,, \qquad \nabla f(x^\star)^T Z(\hat{A}Z)^\dagger b^\star = 0 \tag{399}$$

When $\mathcal{K} = \mathbb{R}_+^n$ , in particular, then $C = \mathbf{0}$, $A = Z = I \in \mathbb{S}^n$ ; *id est*,

$$\begin{array}{cl} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \underset{\mathbb{R}_+^n}{\succeq} 0 \end{array} \tag{400}$$

The necessary and sufficient conditions become (*confer* [46, §4.2.3])

$$\nabla f(x^\star) \underset{\mathbb{R}_+^n}{\succeq} 0\,, \qquad x^\star \underset{\mathbb{R}_+^n}{\succeq} 0\,, \qquad \nabla f(x^\star)^T x^\star = 0 \tag{401}$$

□

**2.13.10.1.3 Example.** *Linear complementarity.* [200] [233]
Given matrix $A \in \mathbb{R}^{n\times n}$ and vector $q \in \mathbb{R}^n$, the *complementarity problem* is a *feasibility problem*:

$$\begin{array}{rl} \text{find} & w\,,\, z \\ \text{subject to} & w \succeq 0 \\ & z \succeq 0 \\ & w^T z = 0 \\ & w = q + Az \end{array} \tag{402}$$

Volumes have been written about this problem, most notably by Cottle [59]. The problem is not convex if both vectors $w$ and $z$ are variable. But if one of them is fixed, then the problem becomes convex with a very simple geometric interpretation: Define the affine subset

$$\mathcal{A} \triangleq \{y \in \mathbb{R}^n \mid Ay = w - q\} \tag{403}$$

For $w^T z$ 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 feasible solutions:

$$z \in -\mathcal{K}^{\perp}_{\mathbb{R}^n_+}(w \in \mathbb{R}^n_+) \cap \mathcal{A} \;=\; \mathbb{R}^n_+ \cap w^{\perp} \cap \mathcal{A} \tag{404}$$

where $\mathcal{K}^{\perp}_{\mathbb{R}^n_+}(w)$ is the normal cone to $\mathbb{R}^n_+$ at $w$ (393). If this intersection is nonempty, then the problem is solvable. □

### 2.13.11 Proper nonsimplicial $\mathcal{K}$, dual, $X$ fat full-rank

Assume we are given a set of $N$ conically independent generators[2.61] (§2.10) of an arbitrary polyhedral proper cone $\mathcal{K}$ in $\mathbb{R}^n$ arranged columnar in $X \in \mathbb{R}^{n \times N}$ such that $N > n$ (fat) and $\operatorname{rank} X = n$. Having found formula (361) to determine the dual of a simplicial cone, the easiest way to find a vertex-description of the proper dual cone $\mathcal{K}^*$ is to first decompose $\mathcal{K}$ into simplicial parts $\mathcal{K}_i$ so that $\mathcal{K} = \bigcup \mathcal{K}_i$ .[2.62] Each component simplicial cone in $\mathcal{K}$ 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 $\mathcal{K}$. Finding the dual of $\mathcal{K}$ amounts to finding the dual of each simplicial part:

[2.61] We can always remove conically dependent columns from $X$ to construct $\mathcal{K}$ or to determine $\mathcal{K}^*$. (§F.2)

[2.62] That proposition presupposes, of course, that we know how to perform simplicial decomposition efficiently; also called "triangulation". [226] [123, §3.1] [124, §3.1] Existence of multiple simplicial parts means expansion of $x \in \mathcal{K}$ like (352) can no longer be unique because $N$ the number of extreme directions in $\mathcal{K}$ exceeds $n$ the dimension of the space.

**2.13.11.0.1 Theorem.** *Dual cone intersection.* [247, §2.7]
Suppose proper cone $\mathcal{K} \subset \mathbb{R}^n$ equals the union of $M$ simplicial cones $\mathcal{K}_i$ whose extreme directions all coincide with those of $\mathcal{K}$. Then proper dual cone $\mathcal{K}^*$ is the intersection of $M$ dual simplicial cones $\mathcal{K}_i^*$ ; *id est*,

$$\mathcal{K} = \bigcup_{i=1}^{M} \mathcal{K}_i \quad \Rightarrow \quad \mathcal{K}^* = \bigcap_{i=1}^{M} \mathcal{K}_i^* \tag{405}$$

◇

**Proof.** For $X_i \in \mathbb{R}^{n \times n}$, a complete matrix of linearly independent extreme directions (p.124) arranged columnar, corresponding simplicial $\mathcal{K}_i$ (§2.12.3.1.1) has vertex-description

$$\mathcal{K}_i = \{X_i\, c \mid c \succeq 0\} \tag{406}$$

Now suppose,

$$\mathcal{K} \;=\; \bigcup_{i=1}^{M} \mathcal{K}_i \;=\; \bigcup_{i=1}^{M} \{X_i\, c \mid c \succeq 0\} \tag{407}$$

The union of all $\mathcal{K}_i$ can be equivalently expressed

$$\mathcal{K} \;=\; \left\{ [\, X_1 \; X_2 \; \cdots \; X_M \,] \begin{bmatrix} a \\ b \\ \vdots \\ c \end{bmatrix} \;\mid\; a, b, \ldots, c \succeq 0 \right\} \tag{408}$$

Because extreme directions of the simplices $\mathcal{K}_i$ are extreme directions of $\mathcal{K}$ by assumption, then by the *extremes theorem* (§2.8.1.1.1),

$$\mathcal{K} \;=\; \{\, [\, X_1 \; X_2 \; \cdots \; X_M \,]\, d \mid d \succeq 0 \,\} \tag{409}$$

Defining $X \triangleq [\, X_1 \; X_2 \; \cdots \; X_M \,]$ (with any redundant [*sic*] columns optionally removed from $X$), then $\mathcal{K}^*$ can be expressed, (315) (Cone Table **S**, p.167)

$$\mathcal{K}^* \;=\; \{y \mid X^T y \succeq 0\} \;=\; \bigcap_{i=1}^{M} \{y \mid X_i^T y \succeq 0\} \;=\; \bigcap_{i=1}^{M} \mathcal{K}_i^* \tag{410}$$

◆

To find the extreme directions of the dual cone, first we observe that some facets of each simplicial part $\mathcal{K}_i$ are common to facets of $\mathcal{K}$ by assumption, and the union of all those common facets comprises the set of all facets of $\mathcal{K}$ by design. For any particular polyhedral proper cone $\mathcal{K}$, the extreme directions of dual cone $\mathcal{K}^*$ are respectively orthogonal to the facets of $\mathcal{K}$. (§2.13.6.1) Then the extreme directions of the dual cone can be found among the inward-normals to facets of the component simplicial cones $\mathcal{K}_i$ ; those normals are extreme directions of the dual simplicial cones $\mathcal{K}_i^*$. From the theorem and Cone Table **S** (p.167),

$$\mathcal{K}^* = \bigcap_{i=1}^{M} \mathcal{K}_i^* = \bigcap_{i=1}^{M} \{X_i^{\dagger T} c \mid c \succeq 0\} \tag{411}$$

The set of extreme directions $\{\Gamma_i^*\}$ for proper dual cone $\mathcal{K}^*$ is therefore constituted by the conically independent generators, from the columns of all the dual simplicial matrices $\{X_i^{\dagger T}\}$, that do not violate discrete definition (315) of $\mathcal{K}^*$;

$$\left\{\Gamma_1^*,\, \Gamma_2^* \ldots \Gamma_N^*\right\} = \text{c.i.}\left\{X_i^{\dagger T}(:,j)\,,\; i=1\ldots M\,,\; j=1\ldots n \mid X_i^{\dagger}(j,:)\Gamma_\ell \geq 0,\; \ell=1\ldots N\right\} \tag{412}$$

where c.i. denotes selection of only the conically independent vectors from the argument set, argument $(:,j)$ denotes the $j^{\text{th}}$ column while $(j,:)$ denotes the $j^{\text{th}}$ row, and $\{\Gamma_\ell\}$ constitutes the extreme directions of $\mathcal{K}$. Figure **38**(b) (p.123) shows a cone and its dual found via this formulation.

**2.13.11.0.2 Example.** *Dual of* $\mathcal{K}$ *nonsimplicial in subspace* aff $\mathcal{K}$.
Given conically independent generators for pointed closed convex cone $\mathcal{K}$ in $\mathbb{R}^4$ arranged columnar in

$$X = [\,\Gamma_1 \;\; \Gamma_2 \;\; \Gamma_3 \;\; \Gamma_4\,] = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & -1 \end{bmatrix} \tag{413}$$

having $\dim \operatorname{aff} \mathcal{K} = \operatorname{rank} X = 3$, then performing the most inefficient simplicial decomposition in $\operatorname{aff} \mathcal{K}$ we find

$$\begin{array}{c}X_1 = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \\ X_3 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}, \quad X_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}\end{array} \tag{414}$$

The corresponding dual simplicial cones in $\operatorname{aff} \mathcal{K}$ have generators respectively columnar in

$$\begin{array}{c}4X_1^{\dagger T} = \begin{bmatrix} 2 & 1 & 1 \\ -2 & 1 & 1 \\ 2 & -3 & 1 \\ -2 & 1 & -3 \end{bmatrix}, \quad 4X_2^{\dagger T} = \begin{bmatrix} 1 & 2 & 1 \\ -3 & 2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & -3 \end{bmatrix} \\ 4X_3^{\dagger T} = \begin{bmatrix} 3 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & -2 & 3 \\ -1 & -2 & -1 \end{bmatrix}, \quad 4X_4^{\dagger T} = \begin{bmatrix} 3 & -1 & 2 \\ -1 & 3 & -2 \\ -1 & -1 & 2 \\ -1 & -1 & -2 \end{bmatrix}\end{array} \tag{415}$$

Applying (412) we get

$$\begin{bmatrix} \Gamma_1^* & \Gamma_2^* & \Gamma_3^* & \Gamma_4^* \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & -1 & -2 \\ 1 & -2 & -1 & 2 \\ -3 & -2 & -1 & -2 \end{bmatrix} \tag{416}$$

whose rank is 3, and is the known result;[2.63] the conically independent generators for that pointed section of the dual cone $\mathcal{K}^*$ in $\operatorname{aff} \mathcal{K}$; *id est*, $\mathcal{K}^* \cap \operatorname{aff} \mathcal{K}$. □

[2.63] These calculations proceed so as to be consistent with [78, §6]; as if the ambient vector space were the proper subspace $\operatorname{aff} \mathcal{K}$ 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 the dimension of the ordinary dual cone to be 3; it is, in fact, 4.

# 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.*

−Stephen Boyd & Lieven Vandenberghe [46, §3.1.7]

We limit our treatment of *multidimensional functions*[3.1] to finite-dimensional Euclidean space. Then the icon for the one-dimensional (real) *convex function* is bowl-shaped (Figure **59**), 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* in the literature. Despite the iconic imagery, the reader is reminded that the set of all convex, concave, quasiconvex, and quasiconcave functions contains the *monotonic functions* [151] [158, §2.3.5]; *e.g.*, [46, §3.6, exer.3.46].

[3.1] 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.



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

# 3.1 Convex function

## 3.1.1 real and vector-valued function

Vector-valued function

$$f(X) : \mathbb{R}^{p\times k} \to \mathbb{R}^M = \begin{bmatrix} f_1(X) \\ \vdots \\ f_M(X) \end{bmatrix} \tag{417}$$

assigns each $X$ in its domain $\operatorname{dom} f$ (a subset of ambient vector space $\mathbb{R}^{p\times k}$) to a specific *element* [189, 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 $Y, Z \in \operatorname{dom} f$ and $\alpha, \beta \in \mathbb{R}$

$$f(\alpha Y + \beta Z) = \alpha f(Y) + \beta f(Z) \tag{418}$$

A vector-valued function $f(X) : \mathbb{R}^{p\times k} \to \mathbb{R}^M$ is convex in $X$ if and only if $\operatorname{dom} f$ is a convex set and, for each and every $Y, Z \in \operatorname{dom} f$ and $0 \le \mu \le 1$

$$f(\mu Y + (1-\mu)Z) \underset{\mathbb{R}^M_+}{\preceq} \mu f(Y) + (1-\mu) f(Z) \tag{419}$$

As defined, continuity is implied but not differentiability. Reversing the sense of the inequality flips this definition to concavity. Linear functions are, apparently, simultaneously convex and concave.

Vector-valued functions are most often compared (151) as in (419) with respect to the $M$-dimensional self-dual nonnegative orthant $\mathbb{R}^M_+$, a proper cone.[3.2] In this case, the test prescribed by (419) 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. [46, §3, §4] This conclusion follows from theory of dual generalized inequalities (§2.13.2.0.1) which asserts

[3.2] The definition of convexity can be broadened to other (not necessarily proper) cones; referred to in the literature as *K-convexity*. [218]

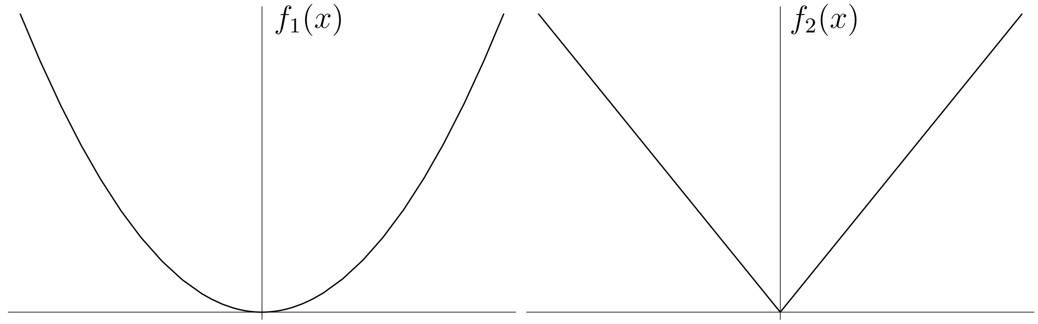



Figure 54: Each convex real function has a unique minimizer $x^\star$ but, for $x \in \mathbb{R}$, $f_1(x) = x^2$ is strictly convex whereas $f_2(x) = \sqrt{x^2} = |x|$ is not. Strict convexity of a real function is therefore only a sufficient condition for minimizer uniqueness.

$$f \text{ convex} \;\Leftrightarrow\; w^T\! f \text{ convex } \;\forall\, w \in \mathcal{G}(\mathbb{R}^M_+) \tag{420}$$

shown by substitution of the defining inequality (419). Discretization (§2.13.4.2.1) allows relaxation of the semi-infinite number of conditions $\forall\, w \succeq 0$ to:

$$\forall\, w \in \mathcal{G}(\mathbb{R}^M_+) = \{e_i\,,\; i = 1 \ldots M\} \tag{421}$$

(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; *id est*, the test for convexity of a vector-valued function is a comparison on $\mathbb{R}$ of each entry.

**3.1.1.0.1 Exercise.** *Cone of convex functions.*
Prove that relation (420) implies: the set of all vector-valued convex functions in $\mathbb{R}^M$ is a convex cone. Indeed, any nonnegatively weighted sum of (strictly) convex functions remains (strictly) convex.[3.3] Interior to the cone are the strictly convex functions. ▼

[3.3] The strict case excludes the cone's point at the origin.

### 3.1.2 strict convexity

When $f(X)$ instead satisfies, for each and every distinct $Y$ and $Z$ in $\operatorname{dom} f$ and all $0<\mu<1$

$$f(\mu Y+(1-\mu)Z) \underset{\mathbb{R}_+^M}{\prec} \mu f(Y)+(1-\mu)f(Z) \tag{422}$$

then we shall say $f$ is a *strictly convex function.*

Similarly to (420)

$$f \text{ strictly convex } \Leftrightarrow\; w^T f \text{ strictly convex } \;\forall\, w\in\mathcal{G}(\mathbb{R}_+^M) \tag{423}$$

discretization allows relaxation of the semi-infinite number of conditions $\forall\, w\succeq 0\,,\; w\neq\mathbf{0}$ (282) to a finite number (421). More tests for strict convexity are given in §3.1.8.1.2, §3.1.11, and §3.2.3.0.2.

Any convex real function $f(X)$ has unique minimum value over any convex subset of its domain. Yet solution to some convex optimization problem is, in general, not unique; *e.g.*, given a minimization of a convex real function over some abstracted convex set $\mathcal{C}$

$$\begin{array}{cl}\underset{X}{\text{minimize}} & f(X)\\ \text{subject to} & X\in\mathcal{C}\end{array} \tag{424}$$

any *optimal solution* $X^\star$ comes from a convex set of optimal solutions

$$X^\star\in\;\{X \mid f(X)=\inf_{Y\in\mathcal{C}}\, f(Y)\,\} \tag{425}$$

But a strictly convex real function has a unique minimizer $X^\star$; *id est*, for the optimal solution set in (425) to be a single point, it is sufficient (Figure **54**) that $f(X)$ be a strictly convex real[3.4] function and set $\mathcal{C}$ convex.

Quadratic real functions $x^T P x + q^T x + r$ characterized by a symmetric positive definite matrix $P$ are strictly convex in $x$. The vector 2-norm square $\|x\|^2$ (Euclidean norm square) and Frobenius norm square $\|X\|_\mathrm{F}^2$, for example, are strictly convex functions of their respective argument (each norm is convex but not strictly convex). Figure **54**(a) illustrates a strictly convex real function.

[3.4] 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 value of the objective. (§2.7.2.2) Study of multidimensional objective functions is called *multicriteria optimization* [242] or *vector optimization.*

### 3.1.3 norm functions, absolute value

$$\begin{array}{rcll}\|x\|_1 & = & \underset{t\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T t \\ & & \text{subject to} & -t \preceq x \preceq t\end{array} \tag{426}$$

where $|x| = t^\star$.

$$\begin{array}{rcll}\|x\|_2 & = & \underset{t\in\mathbb{R}}{\text{minimize}} & t \\ & & \text{subject to} & \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0\end{array} \tag{427}$$

where $\|x\| = t^\star$.

$$\begin{array}{rcll}\|x\|_\infty & = & \underset{t\in\mathbb{R}}{\text{minimize}} & t \\ & & \text{subject to} & -t\mathbf{1} \preceq x \preceq t\mathbf{1}\end{array} \tag{428}$$

where $\max\{|x_i|\ ,\ i=1\ldots n\} = t^\star$.

$$\begin{array}{rcll}\|x\|_1 & = & \underset{\alpha\in\mathbb{R}^n,\ \beta\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T(\alpha+\beta) \\ & & \text{subject to} & \alpha\,,\beta \succeq 0 \\ & & & x = \alpha - \beta\end{array} \tag{429}$$

where $|x| = \alpha^\star + \beta^\star$ because of complementarity $\alpha^{\star T}\beta^\star = 0$.

Optimal solution is norm dependent. [46, p.297] Given set $\mathcal{C}$

$$\begin{array}{ll}\underset{x\in\mathbb{R}^n}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & x\in\mathcal{C}\end{array} \equiv \begin{array}{ll}\underset{x\in\mathbb{R}^n,\ t\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T t \\ \text{subject to} & -t \preceq x \preceq t \\ & x\in\mathcal{C}\end{array} \tag{430}$$

$$\begin{array}{ll}\underset{x\in\mathbb{R}^n}{\text{minimize}} & \|x\|_2 \\ \text{subject to} & x\in\mathcal{C}\end{array} \equiv \begin{array}{ll}\underset{x\in\mathbb{R}^n,\ t\in\mathbb{R}}{\text{minimize}} & t \\ \text{subject to} & \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0 \\ & x\in\mathcal{C}\end{array} \tag{431}$$

$$\begin{array}{ll}\underset{x\in\mathbb{R}^n}{\text{minimize}} & \|x\|_\infty \\ \text{subject to} & x\in\mathcal{C}\end{array} \equiv \begin{array}{ll}\underset{x\in\mathbb{R}^n,\ t\in\mathbb{R}}{\text{minimize}} & t \\ \text{subject to} & -t\mathbf{1} \preceq x \preceq t\mathbf{1} \\ & x\in\mathcal{C}\end{array} \tag{432}$$

In $\mathbb{R}^n$ these norms represent: $\|x\|_1$ is length measured along a grid, $\|x\|_2$ is Euclidean length, $\|x\|_\infty$ is maximum |coordinate|.

(Ye)

$$\begin{array}{cl}\underset{x\in\mathbb{R}^n}{\text{minimize}} & \|x\|_1\\ \text{subject to} & x\in\mathcal{C}\end{array} \quad\equiv\quad \begin{array}{cl}\underset{\alpha\in\mathbb{R}^n,\ \beta\in\mathbb{R}^n,\ x\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T(\alpha+\beta)\\ \text{subject to} & \alpha\,,\beta\succeq 0\\ & x=\alpha-\beta\\ & x\in\mathcal{C}\end{array} \tag{433}$$

All these problems are convex when set $\mathcal{C}$ is.

#### 3.1.3.1 k smallest/largest entries

Sum of the $k$ smallest entries of $x\in\mathbb{R}^n$ is the optimal objective value from: for $1\leq k\leq n$

$$\sum_{i=n-k+1}^{n}\pi(x)_i \;=\; \begin{array}{cl}\underset{y\in\mathbb{R}^n}{\text{minimize}} & x^Ty\\ \text{subject to} & 0\preceq y\preceq\mathbf{1}\\ & \mathbf{1}^Ty=k\end{array} \quad\text{or}\quad \sum_{i=n-k+1}^{n}\pi(x)_i \;=\; \begin{array}{cl}\underset{z\in\mathbb{R}^n,\ t\in\mathbb{R}}{\text{maximize}} & k\,t+\mathbf{1}^Tz\\ \text{subject to} & x\succeq t\,\mathbf{1}+z\\ & z\preceq 0\end{array} \tag{434}$$

which are dual programs, where $\pi(x)_1=\max\{x_i\;,\;i=1\ldots n\}$ where $\pi$ is the nonlinear permutation operator sorting its vector argument into nonincreasing order.

Sum of the $k$ largest entries of $x\in\mathbb{R}^n$ is the optimal objective value from: [46, exer.5.19]

$$\sum_{i=1}^{k}\pi(x)_i \;=\; \begin{array}{cl}\underset{y\in\mathbb{R}^n}{\text{maximize}} & x^Ty\\ \text{subject to} & 0\preceq y\preceq\mathbf{1}\\ & \mathbf{1}^Ty=k\end{array} \quad\text{or}\quad \sum_{i=1}^{k}\pi(x)_i \;=\; \begin{array}{cl}\underset{z\in\mathbb{R}^n,\ t\in\mathbb{R}}{\text{minimize}} & k\,t+\mathbf{1}^Tz\\ \text{subject to} & x\preceq t\,\mathbf{1}+z\\ & z\succeq 0\end{array} \tag{435}$$

which are dual programs.

Let $\Pi\,x$ be a permutation of entries $x_i$ such that their absolute value becomes arranged in nonincreasing order: $|\Pi\,x|_1\geq|\Pi\,x|_2\geq\cdots\geq|\Pi\,x|_n$ . By properties of vector norm, [166, p.59] [110, p.52] sum of the $k$ largest entries of $|x|\in\mathbb{R}^n$ is a norm:

$$\|x\|_{\substack{n\\k}} \;\stackrel{\Delta}{=}\; \sum_{i=1}^{k}|\Pi\,x|_i \;=\; \begin{array}{cl}\underset{z\in\mathbb{R}^n,\ t\in\mathbb{R}}{\text{minimize}} & k\,t+\mathbf{1}^Tz\\ \text{subject to} & -t\,\mathbf{1}-z\;\preceq\;x\;\preceq\;t\,\mathbf{1}+z\\ & z\succeq 0\end{array} \tag{436}$$

where the norm subscript derives from a binomial coefficient $\binom{n}{k}$, and

$$\begin{aligned} \|x\|_{\substack{n\\n}} &\triangleq \|x\|_1 \\ \|x\|_{\substack{n\\1}} &\triangleq \|x\|_\infty \end{aligned} \tag{437}$$

Finding $k$ largest absolute entries of an $n$-length vector $x$ is expressible as supremum of $2^k n!/(k!(n-k)!)$ linear functions of $x$. [46, exer.6.3(e)]

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \|x\|_{\substack{n\\k}} \\ \text{subject to} & x\in\mathcal{C} \end{array} \equiv \begin{array}{ll} \underset{z\in\mathbb{R}^n,\ t\in\mathbb{R}\ ,\ x\in\mathbb{R}^n}{\text{minimize}} & k\,t + \mathbf{1}^T z \\ \text{subject to} & -t\,\mathbf{1} - z \preceq x \preceq t\,\mathbf{1} + z \\ & z \succeq 0 \\ & x\in\mathcal{C} \end{array} \tag{438}$$

#### 3.1.3.2 clipping

Clipping negative vector entries is accomplished:

$$\begin{array}{rcll} \|x_+\|_1 & = & \underset{t\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T t \\ & & \text{subject to} & x \preceq t \\ & & & 0 \preceq t \end{array} \tag{439}$$

where, for $x = [x_i] \in \mathbb{R}^n$

$$x_+ = t^\star = \left[\begin{cases} x_i\,, & x_i \geq 0 \\ 0\,, & x_i < 0 \end{cases},\quad i=1\ldots n\right] \tag{440}$$

(clipping)

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \|x_+\|_1 \\ \text{subject to} & x\in\mathcal{C} \end{array} \equiv \begin{array}{ll} \underset{x\in\mathbb{R}^n,\ t\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T t \\ \text{subject to} & x \preceq t \\ & 0 \preceq t \\ & x\in\mathcal{C} \end{array} \tag{441}$$

### 3.1.4 inverted

We wish to implement objectives of the form $x^{-1}$. Suppose we have a $2\times2$ matrix

$$T \triangleq \begin{bmatrix} x & z \\ z & y \end{bmatrix} \in \mathbb{R}^{\mathbf{2}} \tag{442}$$

which is positive semidefinite by (1314) when

$$T \succeq 0 \quad \Leftrightarrow \quad x > 0 \;\; \textbf{and} \;\; xy \geq z^2 \tag{443}$$

This means we may formulate convex problems, having inverted variables, as semidefinite programs; *e.g.*,

$$\begin{array}{cl} \underset{x\in\mathbb{R}}{\text{minimize}} & x^{-1} \\ \text{subject to} & x > 0 \\ & x \in \mathcal{C} \end{array} \quad \equiv \quad \begin{array}{cl} \underset{x\,,\,y\,\in\,\mathbb{R}}{\text{minimize}} & y \\ \text{subject to} & \begin{bmatrix} x & 1 \\ 1 & y \end{bmatrix} \succeq 0 \\ & x \in \mathcal{C} \end{array} \tag{444}$$

or

$$x > 0\,, \quad y \geq \frac{1}{x} \quad \Leftrightarrow \quad \begin{bmatrix} x & 1 \\ 1 & y \end{bmatrix} \succeq 0 \tag{445}$$

(inverted) For vector $x = [x_i\,,\; i=1\ldots n] \in \mathbb{R}^n$

$$\begin{array}{cl} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \sum\limits_{i=1}^{n} x_i^{-1} \\ \text{subject to} & x \succ 0 \\ & x \in \mathcal{C} \end{array} \quad \equiv \quad \begin{array}{cl} \underset{x\in\mathbb{R}^n,\; y\in\mathbb{R}}{\text{minimize}} & y \\ \text{subject to} & \begin{bmatrix} x_i & \sqrt{n} \\ \sqrt{n} & y \end{bmatrix} \succeq 0\,, \quad i=1\ldots n \\ & x \in \mathcal{C} \end{array} \tag{446}$$

or

$$x \succ 0\,, \quad y \geq \operatorname{tr}\big(\delta(x)^{-1}\big) \quad \Leftrightarrow \quad \begin{bmatrix} x_i & \sqrt{n} \\ \sqrt{n} & y \end{bmatrix} \succeq 0\,, \quad i=1\ldots n \tag{447}$$

### 3.1.5 fractional power

[100] 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:

$$\alpha \triangleq \frac{k}{2^q} \tag{448}$$

where $k \in \{0,\,1,\,2 \ldots 2^q - 1\}$. 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 $y = [y_i\,,\ i = 0 \ldots q]$ with $y_0 = 1$.

#### 3.1.5.1 positive

Then we have the equivalent semidefinite program for maximizing a concave function $x^\alpha$, for quantized $0 \leq \alpha < 1$

$$\begin{array}{cl}\underset{x\in\mathbb{R}}{\text{maximize}} & x^\alpha \\ \text{subject to} & x > 0 \\ & x \in \mathcal{C}\end{array} \quad\equiv\quad \begin{array}{cl}\underset{x\in\mathbb{R}\,,\ y\in\mathbb{R}^{q+1}}{\text{maximize}} & y_q \\ \text{subject to} & \begin{bmatrix} y_{i-1} & y_i \\ y_i & x^{b_i}\end{bmatrix} \succeq 0\,, \quad i = 1 \ldots q \\ & x \in \mathcal{C}\end{array} \tag{449}$$

where nonnegativity of $y_q$ is enforced by maximization; *id est*,

$$x > 0\,, \quad y_q \leq x^\alpha \quad\Leftrightarrow\quad \begin{bmatrix} y_{i-1} & y_i \\ y_i & x^{b_i}\end{bmatrix} \succeq 0\,, \quad i = 1 \ldots q \tag{450}$$

#### 3.1.5.2 negative

Is it also desirable implement an objective of the form $x^{-\alpha}$ for positive $\alpha$. The technique is nearly the same as before: for quantized $0 \leq \alpha < 1$

$$\begin{array}{cl}\underset{x\in\mathbb{R}}{\text{minimize}} & x^{-\alpha} \\ \text{subject to} & x > 0 \\ & x \in \mathcal{C}\end{array} \quad\equiv\quad \begin{array}{cl}\underset{x\,,\,z\in\mathbb{R}\,,\ y\in\mathbb{R}^{q+1}}{\text{minimize}} & z \\ \text{subject to} & \begin{bmatrix} y_{i-1} & y_i \\ y_i & x^{b_i}\end{bmatrix} \succeq 0\,, \quad i = 1 \ldots q \\ & \begin{bmatrix} z & 1 \\ 1 & y_q\end{bmatrix} \succeq 0 \\ & x \in \mathcal{C}\end{array} \tag{451}$$

or

$$x > 0\,, \quad z \geq x^{-\alpha} \quad\Leftrightarrow\quad \begin{array}{l}\begin{bmatrix} y_{i-1} & y_i \\ y_i & x^{b_i}\end{bmatrix} \succeq 0\,, \quad i = 1 \ldots q \\ \begin{bmatrix} z & 1 \\ 1 & y_q\end{bmatrix} \succeq 0\end{array} \tag{452}$$

#### 3.1.5.3 positive inverse

Now define vector $t=[t_i\,,\ i=0\ldots q]$ with $t_0=1$. To implement an objective $x^{1/\alpha}$ for quantized $0\leq\alpha<1$ as in (448)

$$\begin{array}{cl}\underset{x\in\mathbb{R}}{\text{minimize}} & x^{1/\alpha}\\ \text{subject to} & x\geq 0\\ & x\in\mathcal{C}\end{array}\quad\equiv\quad\begin{array}{cl}\underset{y\in\mathbb{R}\;,\;t\in\mathbb{R}^{q+1}}{\text{minimize}} & y\\ \text{subject to} & \begin{bmatrix} t_{i-1} & t_i\\ t_i & y^{b_i}\end{bmatrix}\succeq 0\;,\quad i=1\ldots q\\ & x=t_q\geq 0\\ & x\in\mathcal{C}\end{array}\tag{453}$$

or

$$x\geq 0\,,\quad y\geq x^{1/\alpha}\quad\Leftrightarrow\quad\begin{array}{l}\begin{bmatrix} t_{i-1} & t_i\\ t_i & y^{b_i}\end{bmatrix}\succeq 0\;,\quad i=1\ldots q\\ x=t_q\geq 0\end{array}\tag{454}$$

### 3.1.6 affine function

A function $f(X)$ is *affine* when it is continuous and has the dimensionally extensible form (*confer* §2.9.1.0.2)

$$f(X)=AX+B\tag{455}$$

When $B=\mathbf{0}$ then $f(X)$ is a *linear function.* All affine functions are simultaneously convex and concave.

The real affine function in Figure **55** illustrates hyperplanes in its domain constituting contours of equal function-value (*level sets* $\{z\mid f(z)=\kappa\}$).

Variegated multidimensional affine functions are recognized by the existence of no multivariate terms in argument entries and no polynomial terms in argument entries of degree higher than 1; *id est*, entries of the function are characterized only by linear combinations of the argument entries plus constants.

For $X\in\mathbb{S}^M$ and matrices $A\,,B\,,Q\,,R$ of any compatible dimensions, for example, the expression $XAX$ is not affine in $X$ whereas

$$g(X)=\begin{bmatrix} R & B^TX\\ XB & Q+A^TX+XA\end{bmatrix}\tag{456}$$

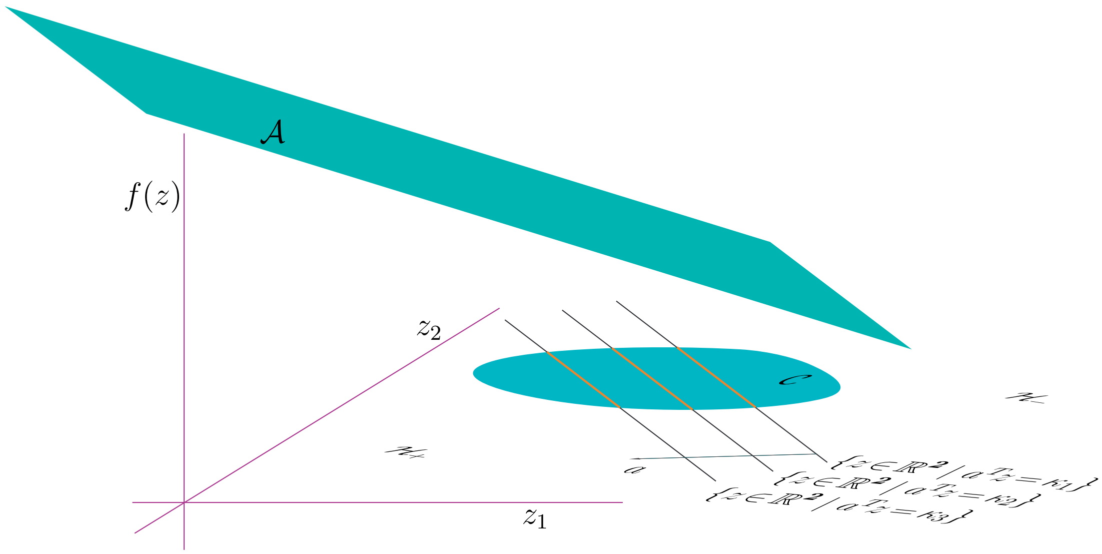



Figure 55: Cartesian axes in $\mathbb{R}^3$ and three hyperplanes intersecting convex set $\mathcal{C}\subset\mathbb{R}^2$ reproduced from Figure **19**. Plotted with third dimension is affine set $\mathcal{A}=f(\mathbb{R}^2)$ a plane. Sequence of hyperplanes, w.r.t domain $\mathbb{R}^2$ of an affine function $f(z)=a^Tz+b:\mathbb{R}^2\to\mathbb{R}$ , is increasing in direction of gradient $a$ (§3.1.8.0.3) because affine function increases in normal direction (Figure **17**).

is an affine multidimensional function. Such a function is typical in engineering control. [296, §2.2][3.5] [44] [102]

**3.1.6.0.1 Example.** *Linear objective.*
Consider minimization of a real affine function $f(z)=a^Tz+b$ over the convex feasible set $\mathcal{C}$ in its domain $\mathbb{R}^2$ illustrated in Figure **55**. Since vector $b$ is fixed, the problem posed is the same as the convex optimization

$$\begin{array}{cl}\underset{z}{\text{minimize}} & a^Tz\\ \text{subject to} & z\in\mathcal{C}\end{array}\tag{457}$$

[3.5] The interpretation from this citation of $\{X\in\mathbb{S}^M \mid g(X)\succeq 0\}$ 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. (*confer* §4.2.3.0.1) 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 $\mathcal{C}$ 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 **55**. Visualization in the function domain is easier because of lower dimension and because level sets 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 (457) graphically by finding that hyperplane intersecting feasible set $\mathcal{C}$ furthest right (in the direction of negative gradient $-a$ (§3.1.8)). $\square$

When a differentiable convex objective function $f$ is nonlinear, the negative gradient $-\nabla f$ is a viable search direction (replacing $-a$ in (457)). (§2.13.10.1, Figure **53**) [104] Then the nonlinear objective function can be replaced with a dynamic linear objective; linear as in (457).

**3.1.6.0.2 Example.** *Support function.* [46, §3.2]
For arbitrary set $\mathcal{Y} \subseteq \mathbb{R}^n$, its *support function* $\sigma_{\mathcal{Y}}(a) : \mathbb{R}^n \to \mathbb{R}$ is defined

$$\sigma_{\mathcal{Y}}(a) \triangleq \sup_{z \in \mathcal{Y}} a^T z \tag{458}$$

whose range contains $\pm\infty$ [182, p.135] [148, §C.2.3.1]. For each $z \in \mathcal{Y}$, $a^T z$ is a linear function of vector $a$. Because $\sigma_{\mathcal{Y}}(a)$ is the pointwise supremum of linear functions, it is convex in $a$. (Figure **56**) Application of the support function is illustrated in Figure **20**(a) for one particular normal $a$. $\square$

### 3.1.7 epigraph, sublevel set

It is well established that a continuous real function is convex if and only if its *epigraph* makes a convex set. [148] [230] [268] [280] [182] Thereby, piecewise-continuous convex functions are admitted. Epigraph is the connection between convex sets and convex functions. Its generalization to a vector-valued function $f(X) : \mathbb{R}^{p \times k} \to \mathbb{R}^M$ is straightforward: [218]

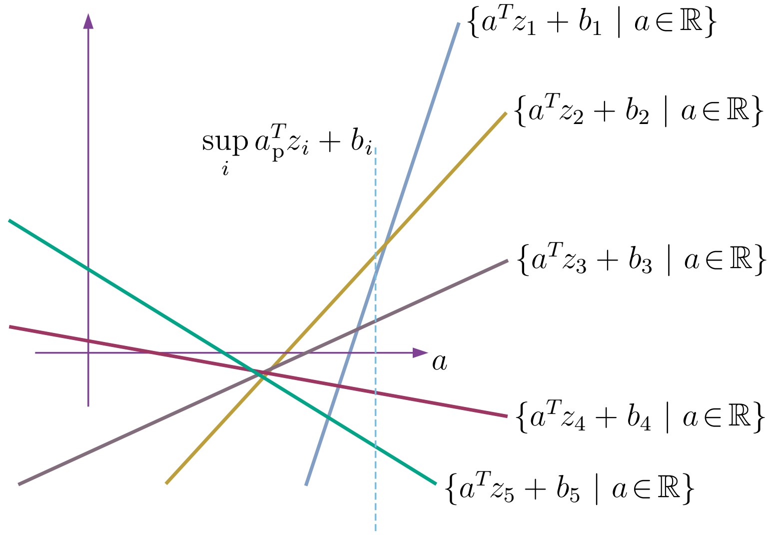



Figure 56: Pointwise supremum of convex functions remains a convex function. Illustrated is a supremum of affine functions in variable $a$ evaluated for a particular argument $a_{\mathrm{p}}$. Topmost affine function is supremum for each value of $a$.

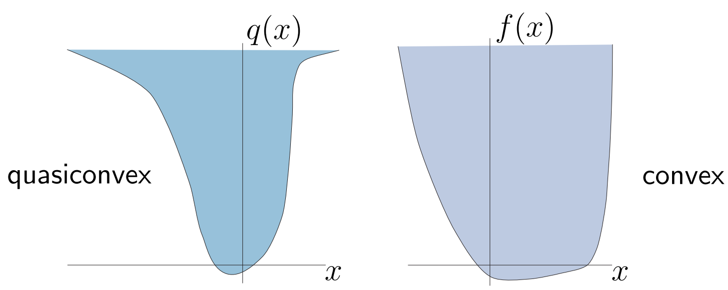



Figure 57: Quasiconvex function $q$ epigraph is not necessarily convex, but convex function $f$ epigraph is convex in any dimension. Sublevel sets are necessarily convex for either.

$$\operatorname{epi} f \triangleq \{(X\,,\ t)\in\mathbb{R}^{p\times k}\times\mathbb{R}^M \mid X\in\operatorname{dom} f\ ,\ \ f(X)\underset{\mathbb{R}^M_+}{\preceq} t\,\} \tag{459}$$

*id est*,

$$f \text{ convex } \Leftrightarrow \operatorname{epi} f \text{ convex} \tag{460}$$

Necessity is proven: [46, exer.3.60] Given any $(X\,,u)\,,(Y\,,v)\in\operatorname{epi} f\,,$ we must show for all $\mu\in[0,1]$ that $\mu(X\,,u)+(1-\mu)(Y\,,v)\in\operatorname{epi} f\,;$ *id est*, we must show

$$f(\mu X+(1-\mu)Y)\underset{\mathbb{R}^M_+}{\preceq}\mu\,u+(1-\mu)v \tag{461}$$

Yet this holds by definition because $f(\mu X+(1-\mu)Y)\preceq\mu f(X)+(1-\mu)f(Y)$. The converse also holds. ♦

**3.1.7.0.1 Exercise.** *Epigraph sufficiency.*
Prove that converse: Given any $(X\,,u)\,,(Y\,,v)\in\operatorname{epi} f\,,$ if for all $\mu\in[0,1]$ $\mu(X\,,u)+(1-\mu)(Y\,,v)\in\operatorname{epi} f$ holds, then $f$ must be convex. ▼

*Sublevel sets* of a real convex function are convex. Likewise, corresponding to each and every $\nu\in\mathbb{R}^M$

$$\mathcal{L}_\nu f \triangleq \{X\in\operatorname{dom} f \mid f(X)\underset{\mathbb{R}^M_+}{\preceq}\nu\,\}\subseteq\mathbb{R}^{p\times k} \tag{462}$$

sublevel sets of a vector-valued convex function are convex. As for real functions, the converse does not hold. (Figure **57**)

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,

$$f(\mu\,X+(1-\mu)Y)\underset{\mathbb{R}^M_+}{\preceq}\mu f(X)+(1-\mu)f(Y)\underset{\mathbb{R}^M_+}{\preceq}\nu \tag{463}$$

♦

When an epigraph (459) is artificially bounded above, $t \preceq \nu$, then the corresponding sublevel set can be regarded as an orthogonal projection of the epigraph on the function domain.

Sense of the inequality is reversed in (459), for concave functions, and we use instead the nomenclature *hypograph.* Sense of the inequality in (462) is reversed, similarly, with each convex set then called *superlevel set.*

**3.1.7.0.2 Example.** *Matrix pseudofractional function.*
Consider a real function of two variables on $\operatorname{dom} f = \mathbb{S}^n_+ \times \mathcal{R}(A)$

$$f(A\,,x) : \mathbb{S}^n \times \mathbb{R}^n \to \mathbb{R} \;=\; x^T A^\dagger x \tag{464}$$

This function is convex simultaneously in both variables when variable matrix $A$ belongs to the entire positive semidefinite cone $\mathbb{S}^n_+$ 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 (1311) from §A.4: for $t \in \mathbb{R}$

$$\begin{aligned} G(A\,,z\,,t) = \begin{bmatrix} A & z \\ z^T & t \end{bmatrix} &\succeq 0 \\ &\Leftrightarrow \\ z^T(I - AA^\dagger) &= \mathbf{0} \\ t - z^T A^\dagger z &\geq 0 \\ A &\succeq 0 \end{aligned} \tag{465}$$

Inverse image of the positive semidefinite cone $\mathbb{S}^{n+1}_+$ under affine mapping $G(A\,,z\,,t)$ is convex by Theorem 2.1.9.0.1. Of the equivalent conditions for positive semidefiniteness of $G$, the first is an equality demanding vector $z$ belong to $\mathcal{R}(A)$. Function $f(A\,,z) = z^T A^\dagger z$ is convex on $\mathbb{S}^n_+ \times \mathcal{R}(A)$ because the Cartesian product constituting its epigraph

$$\operatorname{epi} f(A\,,z) = \left\{(A\,,z\,,t) \mid A \succeq 0\,,\; z \in \mathcal{R}(A)\,,\; z^T A^\dagger z \leq t\right\} = G^{-1}\big(\mathbb{S}^{n+1}_+\big) \tag{466}$$

is convex. $\square$

**3.1.7.0.3 Exercise.** *Matrix product function.*
Continue Example 3.1.7.0.2 by introducing vector variable $x$ and making the substitution $z \leftarrow Ax$. Because of matrix symmetry (§E), for all $x \in \mathbb{R}^n$

$$f\big(A\,,z(x)\big) \;=\; z^T A^\dagger z \;=\; x^T A^T A^\dagger A\, x \;=\; x^T A\, x \;=\; f(A\,,x) \tag{467}$$

whose epigraph is

$$\operatorname{epi} f(A\,,x) \;=\; \big\{(A\,,x\,,t) \mid A\succeq 0\,,\ x^TA\,x\leq t\big\} \tag{468}$$

Provide two simple explanations why $f(A\,,x)=x^TA\,x$ is not a function convex simultaneously in positive semidefinite matrix $A$ and vector $x$ on $\operatorname{dom} f=\mathbb{S}_+^n\times\mathbb{R}^n$. ▼

**3.1.7.0.4 Example.** *Matrix fractional function.* (*confer* §3.2.2.0.1)
Continuing Example 3.1.7.0.2, now consider a real function of two variables on $\operatorname{dom} f=\mathbb{S}_+^n\times\mathbb{R}^n$ for small positive constant $\epsilon$ (*confer* (1639))

$$f(A\,,x)=\epsilon\,x^T(A+\epsilon\,I)^{-1}x \tag{469}$$

where the inverse always exists by (1255). This function is convex simultaneously in both variables over the entire positive semidefinite cone $\mathbb{S}_+^n$ and all $x\in\mathbb{R}^n$: Consider Schur-form (1314) from §A.4: for $t\in\mathbb{R}$

$$\begin{aligned} G(A\,,z\,,t)=\begin{bmatrix} A+\epsilon\,I & z\\ z^T & \epsilon^{-1}\,t\end{bmatrix}&\succeq 0\\ \Leftrightarrow&\\ t-\epsilon\,z^T(A+\epsilon\,I)^{-1}z&\geq 0\\ A+\epsilon\,I&\succ 0\end{aligned} \tag{470}$$

Inverse image of the positive semidefinite cone $\mathbb{S}_+^{n+1}$ under affine mapping $G(A\,,z\,,t)$ is convex by Theorem 2.1.9.0.1. Function $f(A\,,z)$ is convex on $\mathbb{S}_+^n\times\mathbb{R}^n$ because its epigraph is that inverse image:

$$\operatorname{epi} f(A\,,z)=\big\{(A\,,z\,,t)\mid A+\epsilon\,I\succ 0\,,\ \epsilon\,z^T(A+\epsilon\,I)^{-1}z\leq t\big\}=G^{-1}\big(\mathbb{S}_+^{n+1}\big) \tag{471}$$

□

### 3.1.7.1 matrix fractional projector function

Consider nonlinear function $f$ having orthogonal projector $W$ as argument:

$$f(W,\,x)=\epsilon\,x^T(W+\epsilon\,I)^{-1}x \tag{472}$$

Projection matrix $W$ has property $W^\dagger=W^T=W\succeq 0$ (1683). 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 (1639) for any $\epsilon > 0$ and idempotent symmetric $W$, $\epsilon(W + \epsilon I)^{-1} = I - (1+\epsilon)^{-1}W$ from which

$$f(W, x) \;=\; \epsilon\, x^T(W + \epsilon I)^{-1}x \;=\; x^T\big(I - (1+\epsilon)^{-1}W\big)\,x \tag{473}$$

Therefore

$$\lim_{\epsilon\to 0^+} f(W, x) \;=\; \lim_{\epsilon\to 0^+} \epsilon\, x^T(W + \epsilon I)^{-1}x \;=\; x^T(I - W)x \tag{474}$$

where $I - W$ is also an orthogonal projector (§E.2).

We learned from Example 3.1.7.0.4 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 $W$)

$$f(W, x) = x^T\big(I - (1+\epsilon)^{-1}W\big)\,x \tag{475}$$

cannot be convex simultaneously in both variables on either the positive semidefinite or symmetric projection matrices.

Suppose we allow $\operatorname{dom} f$ to constitute the entire positive semidefinite cone but constrain $W$ to a Fantope (79); *e.g.*, for convex set $\mathcal{C}$ and $0 < k < n$ as in

$$\begin{array}{cl}
\underset{x\in\mathbb{R}^n,\; W\in\mathbb{S}^n}{\text{minimize}} & \epsilon\, x^T(W + \epsilon I)^{-1}x\\
\text{subject to} & 0 \preceq W \preceq I\\
& \operatorname{tr} W = k\\
& x \in \mathcal{C}
\end{array} \tag{476}$$

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 (473), and then iterate solution of convex problem

$$\begin{array}{cl}\underset{x\in\mathbb{R}^n}{\text{minimize}} & x^T(I-(1+\epsilon)^{-1}W)x\\ \text{subject to} & x\in\mathcal{C}\end{array}\tag{477}$$

with convex problem

$$\text{(a)}\quad\begin{array}{cl}\underset{W\in\mathbb{S}^n}{\text{minimize}} & x^{\star T}(I-(1+\epsilon)^{-1}W)x^\star\\ \text{subject to} & 0\preceq W\preceq I\\ & \operatorname{tr}W=k\end{array}\quad\equiv\quad\begin{array}{cl}\underset{W\in\mathbb{S}^n}{\text{maximize}} & x^{\star T}Wx^\star\\ \text{subject to} & 0\preceq W\preceq I\\ & \operatorname{tr}W=k\end{array}\tag{478}$$

until convergence, where $x^\star$ represents an optimal solution of (477) from any particular iteration. The idea is to optimally solve for the partitioned variables which are later combined to solve the original problem (476). 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.[3.6]

But partitioning alone does not guarantee a projector. To make orthogonal projector $W$ a certainty, we must invoke a known analytical optimal solution to problem (478): Diagonalize optimal solution from problem (477) $x^\star x^{\star T}\stackrel{\Delta}{=}Q\Lambda Q^T$ (§A.5.2) and set $U^\star=Q(:,1\!:\!k)\in\mathbb{R}^{n\times k}$ *per* (1480c);

$$W=U^\star U^{\star T}=\frac{x^\star x^{\star T}}{\|x^\star\|^2}+Q(:,2\!:\!k)Q(:,2\!:\!k)^T\tag{479}$$

Then optimal solution $(x^\star,\,U^\star)$ to problem (476) is found, for small $\epsilon$, by iterating solution to problem (477) with optimal (projector) solution (479) to convex problem (478).

**Proof.** Optimal vector $x^\star$ is orthogonal to the last $n-1$ columns of orthogonal matrix $Q$, so

$$f^\star_{(477)}=\|x^\star\|^2(1-(1+\epsilon)^{-1})\tag{480}$$

after each iteration. Convergence of $f^\star_{(477)}$ is proven with the observation that iteration (477) (478a) is a nonincreasing sequence that is bounded below by 0. Any bounded monotonic sequence in $\mathbb{R}$ is convergent. [189, §1.2]

[3.6] A convex problem has convex feasible set, and the objective *surface* has one and only one global minimum.

[30, §1.1] Expression (479) for optimal projector $W$ holds at each iteration, therefore $\|x^\star\|^2(1-(1+\epsilon)^{-1})$ must also represent the optimal objective value $f^\star_{(477)}$ at convergence.

Because the objective $f_{(476)}$ from problem (476) is also bounded below by 0 on the same domain, this convergent optimal objective value $f^\star_{(477)}$ (for positive $\epsilon$ arbitrarily close to 0) is necessarily optimal for (476); *id est*,

$$f^\star_{(477)} \geq f^\star_{(476)} \geq 0 \tag{481}$$

by (1462), and

$$\lim_{\epsilon\to 0^+} f^\star_{(477)} = 0 \tag{482}$$

Since optimal $(x^\star,\, U^\star)$ from problem (477) is feasible to problem (476), and because their objectives are equivalent for projectors by (473), then converged $(x^\star,\, U^\star)$ must also be optimal to (476) in the limit. Because problem (476) is convex, this represents a globally optimal solution. ♦

#### 3.1.7.2 Semidefinite program via Schur

Schur complement (1311) 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 [133] (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 $L<M$

$$\begin{array}{cl}\underset{X\in\,\mathbb{S}^M}{\text{minimize}} & \|A-X\|_{\mathrm F}^2\\ \text{subject to} & S^TXS\succeq 0\end{array} \tag{483}$$

Variable $X$ is constrained to be positive semidefinite, but only on a subspace determined by $S$. First we write the epigraph form:

$$\begin{array}{cl}\underset{X\in\,\mathbb{S}^M,\;t\in\mathbb{R}}{\text{minimize}} & t\\ \text{subject to} & \|A-X\|_{\mathrm F}^2\leq t\\ & S^TXS\succeq 0\end{array} \tag{484}$$

Next we use the Schur complement [204, §6.4.3] [181] and matrix vectorization (§2.2):

$$\begin{array}{cl}\underset{X\in\mathbb{S}^M,\ t\in\mathbb{R}}{\text{minimize}} & t\\ \text{subject to} & \left[\begin{array}{cc} tI & \operatorname{vec}(A-X)\\ \operatorname{vec}(A-X)^T & 1\end{array}\right]\succeq 0\\ & S^TXS\succeq 0\end{array}\tag{485}$$

This semidefinite program is an epigraph form in disguise, equivalent to (483); it demonstrates how a quadratic objective or constraint can be converted to a semidefinite constraint.

Were problem (483) instead equivalently expressed without the square

$$\begin{array}{cl}\underset{X\in\mathbb{S}^M}{\text{minimize}} & \|A-X\|_{\mathrm{F}}\\ \text{subject to} & S^TXS\succeq 0\end{array}\tag{486}$$

then we get a subtle variation:

$$\begin{array}{cl}\underset{X\in\mathbb{S}^M,\ t\in\mathbb{R}}{\text{minimize}} & t\\ \text{subject to} & \|A-X\|_{\mathrm{F}}\leq t\\ & S^TXS\succeq 0\end{array}\tag{487}$$

that leads to an equivalent for (486)

$$\begin{array}{cl}\underset{X\in\mathbb{S}^M,\ t\in\mathbb{R}}{\text{minimize}} & t\\ \text{subject to} & \left[\begin{array}{cc} tI & \operatorname{vec}(A-X)\\ \operatorname{vec}(A-X)^T & t\end{array}\right]\succeq 0\\ & S^TXS\succeq 0\end{array}\tag{488}$$

**3.1.7.2.1 Example.** *Schur anomaly.*
Consider a problem abstract in the convex constraint, given symmetric matrix $A$

$$\begin{array}{cl}\underset{X\in\mathbb{S}^M}{\text{minimize}} & \|X\|_{\mathrm{F}}^2-\|A-X\|_{\mathrm{F}}^2\\ \text{subject to} & X\in\mathcal{C}\end{array}\tag{489}$$

the minimization of a difference of two quadratic functions each convex in matrix $X$. Observe equality

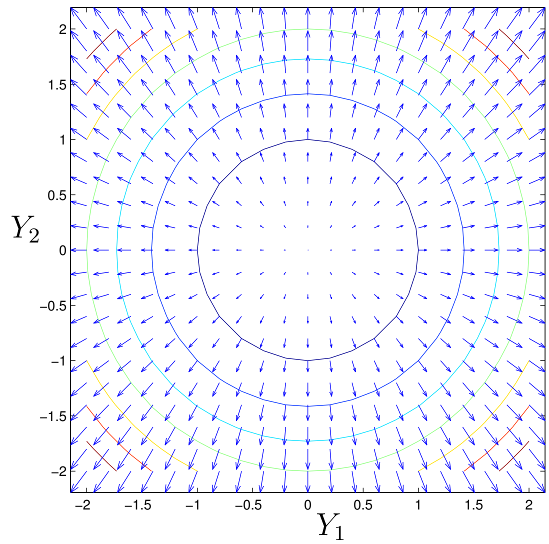



Figure 58: Gradient in $\mathbb{R}^2$ evaluated on grid over some open disc in domain of convex quadratic bowl $f(Y)=Y^TY:\mathbb{R}^2\to\mathbb{R}$ illustrated in Figure **59**. Circular contours are level sets; each defined by a constant function-value.

$$\|X\|_F^2 - \|A-X\|_F^2 = 2\operatorname{tr}(XA) - \|A\|_F^2 \tag{490}$$

So problem (489) is equivalent to the convex optimization

$$\begin{array}{cl}\underset{X\in\mathbb{S}^M}{\text{minimize}} & \operatorname{tr}(XA)\\ \text{subject to} & X\in\mathcal{C}\end{array} \tag{491}$$

But this problem (489) does not have Schur-form

$$\begin{array}{cl}\underset{X\in\mathbb{S}^M,\ \alpha\ ,\ t}{\text{minimize}} & t-\alpha\\ \text{subject to} & X\in\mathcal{C}\\ & \|X\|_F^2\le t\\ & \|A-X\|_F^2\ge\alpha\end{array} \tag{492}$$

because the constraint in $\alpha$ is nonconvex. (§2.9.1.0.1) $\square$

### 3.1.8 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 $f(x):\mathbb{R}^K \to \mathbb{R}$ with respect to its vector domain is defined

$$\nabla f(x) \triangleq \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_K} \end{bmatrix} \in \mathbb{R}^K \tag{1536}$$

while the second-order gradient of the twice differentiable real function with respect to its vector domain is traditionally called the *Hessian*;[3.7]

$$\nabla^2 f(x) \triangleq \begin{bmatrix} \frac{\partial^2 f(x)}{\partial^2 x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_K} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial^2 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_K \partial x_1} & \frac{\partial^2 f(x)}{\partial x_K \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial^2 x_K} \end{bmatrix} \in \mathbb{S}^K \tag{1537}$$

The gradient can be interpreted as a vector pointing in the direction of greatest change. [161, §15.6] The gradient can also be interpreted as that vector normal to a level set; *e.g.*, Figure **60**, Figure **53**.

For the quadratic bowl in Figure **59**, the gradient maps to $\mathbb{R}^{\mathbf{2}}$; illustrated in Figure **58**. 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 $\nabla f = \mathbf{0}$ is a necessary condition for its unconstrained minimization [104, §3.2]:

[3.7] *Jacobian* is the Hessian transpose, so commonly confused in matrix calculus.

**3.1.8.0.1 Example.** *Projection on a rank-1 subset.*
For $A \in \mathbb{S}^N$ 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 $\mathbb{S}^N_+$: Defining $\lambda_1 \triangleq \max_i\{\lambda(A)_i\}$ and corresponding eigenvector $v_1$

$$\begin{aligned}\underset{x}{\text{minimize}}\ \|xx^T - A\|_F^2 &= \underset{x}{\text{minimize}}\ \operatorname{tr}\big(xx^T(x^Tx) - 2Axx^T + A^TA\big)\\ &= \begin{cases}\|\lambda(A)\|^2\ , & \lambda_1 \leq 0\\ \|\lambda(A)\|^2 - \lambda_1^2\ , & \lambda_1 > 0\end{cases}\end{aligned} \tag{1475}$$

$$\arg\ \underset{x}{\text{minimize}}\ \|xx^T - A\|_F^2 = \begin{cases}\mathbf{0}\ , & \lambda_1 \leq 0\\ v_1\sqrt{\lambda_1}\ , & \lambda_1 > 0\end{cases} \tag{1476}$$

From (1563) and §D.2.1, the gradient of $\|xx^T - A\|_F^2$ is

$$\nabla_x\big((x^Tx)^2 - 2x^TA\,x\big) = 4(x^Tx)x - 4Ax \tag{493}$$

Setting the gradient to $\mathbf{0}$

$$Ax = x(x^Tx) \tag{494}$$

is necessary for optimal solution. Replace vector $x$ with a normalized eigenvector $v_i$ of $A \in \mathbb{S}^N$, corresponding to a positive eigenvalue $\lambda_i$, scaled by square root of that eigenvalue. Then (494) is satisfied

$$x \leftarrow v_i\sqrt{\lambda_i} \;\Rightarrow\; Av_i = v_i\lambda_i \tag{495}$$

$xx^T = \lambda_i v_i v_i^T$ 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 $x = \mathbf{0}$ yields the minimum. □

- For any differentiable multidimensional convex function, zero gradient $\nabla f = \mathbf{0}$ is a necessary and sufficient condition for its unconstrained minimization [46, §5.5.3]:

**3.1.8.0.2 Example.** *Pseudoinverse.*
The pseudoinverse matrix is the unique solution to an unconstrained convex optimization problem [110, §5.5.4]: given $A \in \mathbb{R}^{m\times n}$

$$\underset{X\in\mathbb{R}^{n\times m}}{\text{minimize}}\ \|XA - I\|_F^2 \tag{496}$$

where

$$\|XA - I\|_F^2 = \operatorname{tr}\big(A^TX^TXA - XA - A^TX^T + I\big) \tag{497}$$

whose gradient (§D.2.3)

$$\nabla_X \|XA - I\|_F^2 \;=\; 2\big(XAA^T - A^T\big) = \mathbf{0} \tag{498}$$

vanishes when

$$XAA^T = A^T \tag{499}$$

When $A$ is fat full-rank, then $AA^T$ is invertible, $X^\star = A^T(AA^T)^{-1}$ is the pseudoinverse $A^\dagger$, and $AA^\dagger = I$. Otherwise, we can make $AA^T$ invertible by adding a positively scaled identity, for any $A \in \mathbb{R}^{m\times n}$

$$X = A^T(AA^T + t\,I)^{-1} \tag{500}$$

Invertibility is guaranteed for any finite positive value of $t$ by (1255). Then matrix $X$ becomes the pseudoinverse $X \to A^\dagger \overset{\Delta}{=} X^\star$ in the limit $t \to 0^+$. Minimizing instead $\|AX - I\|_F^2$ yields the second flavor in (1638). □

**3.1.8.0.3 Example.** *Hyperplane, line, described by affine function.*
Consider the real affine function of vector variable,

$$f(x) : \mathbb{R}^p \to \mathbb{R} \;=\; a^Tx + b \tag{501}$$

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* [148, §B.1.2] hyperplane $\partial\mathcal{H}$ in the space $\mathbb{R}^p \times \mathbb{R}$ for any particular vector $a$ (*confer* §2.4.2);

$$\partial\mathcal{H} = \left\{ \begin{bmatrix} x \\ a^Tx + b \end{bmatrix} \mid x \in \mathbb{R}^p \right\} \subset \mathbb{R}^p \times \mathbb{R} \tag{502}$$

having nonzero normal

$$\eta = \begin{bmatrix} a \\ -1 \end{bmatrix} \in \mathbb{R}^p \times \mathbb{R} \tag{503}$$

This equivalence to a hyperplane holds only for real functions.[3.8] The epigraph of the real affine function $f(x)$ is therefore a halfspace in $\left[\begin{array}{c}\mathbb{R}^p\\ \mathbb{R}\end{array}\right]$, so we have:

The real affine function is to convex functions
as
the hyperplane is to convex sets.

Similarly, the matrix-valued affine function of real variable $x$, for any particular matrix $A\in\mathbb{R}^{M\times N}$,

$$h(x) : \mathbb{R}\to\mathbb{R}^{M\times N} = Ax + B \tag{504}$$

describes a line in $\mathbb{R}^{M\times N}$ in direction $A$

$$\{Ax + B \mid x\in\mathbb{R}\} \subseteq \mathbb{R}^{M\times N} \tag{505}$$

and describes a line in $\mathbb{R}\times\mathbb{R}^{M\times N}$

$$\left\{\left[\begin{array}{c}x\\ Ax+B\end{array}\right] \mid x\in\mathbb{R}\right\} \subset \mathbb{R}\times\mathbb{R}^{M\times N} \tag{506}$$

whose slope with respect to $x$ is $A$. □

[3.8]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

$$\left\{\left[\begin{array}{c}x\\ Ax+b\end{array}\right] \mid x\in\mathbb{R}^p\right\} \subset \mathbb{R}^p\times\mathbb{R}^M$$

is perpendicular to

$$\eta \triangleq \left[\begin{array}{c}\nabla f(x)\\ -I\end{array}\right] \in \mathbb{R}^{p\times M}\times\mathbb{R}^{M\times M}$$

because

$$\eta^T\left(\left[\begin{array}{c}x\\ Ax+b\end{array}\right] - \left[\begin{array}{c}0\\ b\end{array}\right]\right) = 0 \quad \forall\, x\in\mathbb{R}^p$$

Yet $\eta$ is a vector (in $\mathbb{R}^p\times\mathbb{R}^M$) only when $M=1$. ♦

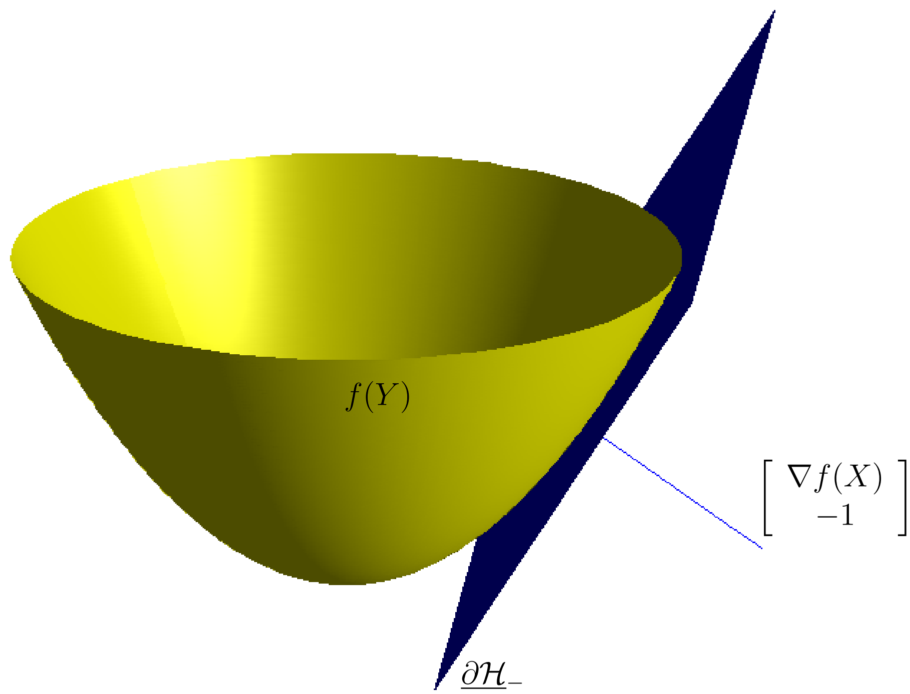



Figure 59: When a real function $f$ is differentiable at each point in its open domain, there is an intuitive geometric interpretation of function convexity in terms of its gradient $\nabla f$ and its epigraph: Drawn is a convex quadratic bowl in $\mathbb{R}^2 \times \mathbb{R}$ (*confer* Figure **117**, p.563); $f(Y) = Y^T Y : \mathbb{R}^2 \to \mathbb{R}$ *versus* $Y$ on some open disc in $\mathbb{R}^2$. Supporting hyperplane $\underline{\partial\mathcal{H}}_- \in \mathbb{R}^2 \times \mathbb{R}$ (which is tangent, 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 such a hyperplane containing $[\,Y^T \;\; f(Y)\,]^T$ and supporting the epigraph.

#### 3.1.8.1 monotonic function

A real differentiable function $f$ of real argument is called *monotonic* when its first derivative (not necessarily continuous) maintains sign over the function domain.

**3.1.8.1.1 Definition.** *Monotonicity.*
Multidimensional function $f$ is monotonic when $\operatorname{sgn}\langle f(Y)-f(X)\,,\,Y-X\rangle$ is invariant (ignoring 0) to all $X,Y\in\operatorname{dom} f$. Nonnegative (nonpositive) sign denotes nonnegative (nonpositive) monotonicity. △

When argument $X$ and $f(X)$ are dimensionally incompatible, the one having smaller dimension is padded with $\mathbf{1}$ to complete the test. It is necessary and sufficient for each entry $f_i$ from this monotonicity definition to be monotonic with the same sign.

A convex function can be characterized by a similar kind of nonnegative monotonicity of its gradient:

**3.1.8.1.2 Theorem.** *Gradient monotonicity.* [148, §B.4.1.4] [41, §3.1, exer.20] Given $f(X):\mathbb{R}^{p\times k}\to\mathbb{R}$ a real differentiable function with matrix argument on open convex domain, the condition

$$\langle\nabla f(Y)-\nabla f(X)\,,\;Y-X\rangle\geq 0 \quad\text{for each and every } X,Y\in\operatorname{dom} f \tag{507}$$

is necessary and sufficient for convexity of $f$. Strict inequality and *caveat* distinct $Y,X$ provide necessary and sufficient conditions for strict convexity. ⋄

**3.1.8.1.3 Example.** *Composition of functions.* [46, §3.2.4] [148, §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

$$f(x)=g\big(h(x)\big)\,,\qquad \operatorname{dom} f=\{x\in\operatorname{dom} h \mid h(x)\in\operatorname{dom} g\} \tag{508}$$

is convex when

- g is convex nonnegatively monotonic **and** h is convex
- g is convex nonpositively monotonic **and** h is concave

and composite function $f$ is concave when

- g is concave nonnegatively monotonic **and** h is concave
- g is concave nonpositively monotonic **and** h is convex

where $\infty$ $(-\infty)$ is assigned to convex (concave) $g$ when evaluated outside its domain. When functions are differentiable, these rules are consequent to (1564). Convexity (concavity) of any $g$ is preserved when $h$ is affine. $\square$

### 3.1.9 first-order convexity condition, real function

Discretization of $w\succeq 0$ in (420) invites refocus to the real-valued function:

**3.1.9.0.1 Theorem.** *Necessary and sufficient convexity condition.* [46, §3.1.3] [88, §I.5.2] [299, §1.2.3] [30, §1.2] [247, §4.2] [229, §3] For real differentiable function $f(X):\mathbb{R}^{p\times k}\to\mathbb{R}$ with matrix argument on open convex domain, the condition (*confer* §D.1.7)

$$f(Y) \geq f(X) + \langle\nabla f(X)\;,\;Y-X\rangle \quad \text{for each and every } X,Y\in\operatorname{dom} f \tag{509}$$

is necessary and sufficient for convexity of $f$. $\diamond$

When $f(X):\mathbb{R}^p\to\mathbb{R}$ is a real differentiable convex function with vector argument on open convex domain, there is simplification of the first-order condition (509); for each and every $X,Y\in\operatorname{dom} f$

$$f(Y) \geq f(X) + \nabla f(X)^T(Y-X) \tag{510}$$

From this we can find a unique [280, §5.5.4] nonvertical [148, §B.1.2] hyperplane $\underline{\partial\mathcal{H}_-}$ (§2.4), expressed in terms of the function gradient, supporting $\operatorname{epi} f$ at $\begin{bmatrix} X \\ f(X) \end{bmatrix}$: *videlicet*, defining $f(Y\notin\operatorname{dom} f)\triangleq\infty$ [46, §3.1.7]

$$\begin{bmatrix} Y \\ t \end{bmatrix} \in \operatorname{epi} f \;\Leftrightarrow\; t \geq f(Y) \;\Rightarrow\; \begin{bmatrix} \nabla f(X)^T & -1 \end{bmatrix}\left(\begin{bmatrix} Y \\ t \end{bmatrix} - \begin{bmatrix} X \\ f(X) \end{bmatrix}\right) \leq 0 \tag{511}$$

This means, for each and every point $X$ in the domain of a real convex function $f(X)$, there exists a hyperplane $\underline{\partial\mathcal{H}}_-$ in $\mathbb{R}^p \times \mathbb{R}$ having normal $\begin{bmatrix} \nabla f(X) \\ -1 \end{bmatrix}$ supporting the function epigraph at $\begin{bmatrix} X \\ f(X) \end{bmatrix} \in \underline{\partial\mathcal{H}}_-$

$$\underline{\partial\mathcal{H}}_- = \left\{ \begin{bmatrix} Y \\ t \end{bmatrix} \in \begin{bmatrix} \mathbb{R}^p \\ \mathbb{R} \end{bmatrix} \;\middle|\; \begin{bmatrix} \nabla f(X)^T & -1 \end{bmatrix}\left(\begin{bmatrix} Y \\ t \end{bmatrix} - \begin{bmatrix} X \\ f(X) \end{bmatrix}\right) = 0 \right\} \tag{512}$$

One such supporting hyperplane (*confer* Figure **20**(a)) is illustrated in Figure **59** for a convex quadratic.

From (510) we deduce, for each and every $X, Y \in \operatorname{dom} f$

$$\nabla f(X)^T (Y - X) \geq 0 \;\Rightarrow\; f(Y) \geq f(X) \tag{513}$$

meaning, the gradient at $X$ identifies a supporting hyperplane there in $\mathbb{R}^p$

$$\{Y \in \mathbb{R}^p \mid \nabla f(X)^T (Y - X) = 0\} \tag{514}$$

to the convex sublevel sets of convex function $f$ (*confer* (462))

$$\mathcal{L}_{f(X)} f \triangleq \{Y \in \operatorname{dom} f \mid f(Y) \leq f(X)\} \subseteq \mathbb{R}^p \tag{515}$$

illustrated for an arbitrary real convex function in Figure **60**.

### 3.1.10 first-order convexity condition, vector function

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 $\operatorname{dom} f$ is open, convex, and for each and every $X, Y \in \operatorname{dom} f$

$$f(Y) \underset{\mathbb{R}^M_+}{\succeq} f(X) + \overset{\to Y-X}{df(X)} \;=\; f(X) + \left.\frac{d}{dt}\right|_{t=0} f(X + t\,(Y - X)) \tag{516}$$

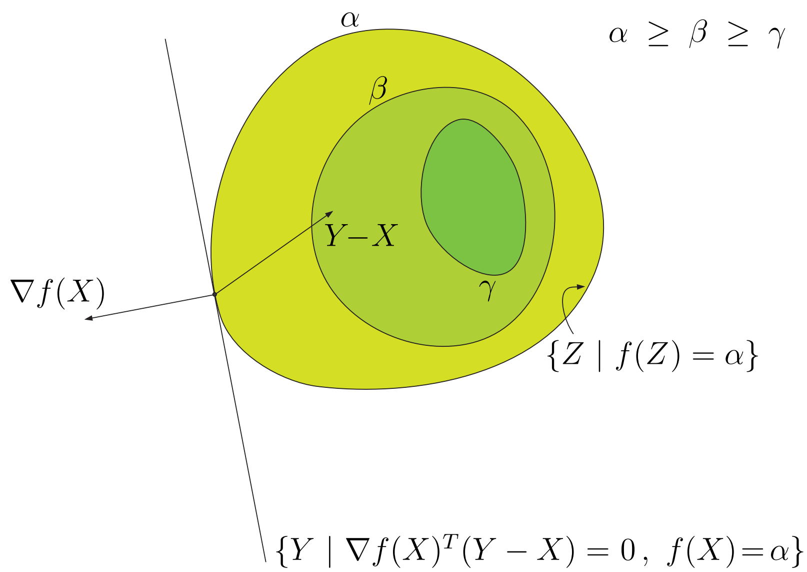



Figure 60: Shown is a plausible contour plot in $\mathbb{R}^2$ of some arbitrary real convex function $f(Z)$ at selected levels $\alpha$, $\beta$, and $\gamma$; contours of equal level $f$ (level sets) drawn in the function's domain. A convex function has convex sublevel sets $\mathcal{L}_{f(X)}f$ (515). [230, §4.6] The sublevel set whose boundary is the level set at $\alpha$, for instance, comprises all the shaded regions. For any particular convex function, the family comprising all its sublevel sets is nested. [148, p.75] Were the sublevel sets not convex, we may certainly conclude the corresponding function is neither convex. Contour plots of real affine functions are illustrated in Figure **17** and Figure **55**.

where $\overset{\to Y-X}{df(X)}$ is the *directional derivative*[3.9] [161] [250] 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),

$$f(Y)-f(X)-\overset{\to Y-X}{df(X)}\underset{\mathbb{R}_+^M}{\succeq}0 \;\Leftrightarrow\; \left\langle f(Y)-f(X)-\overset{\to Y-X}{df(X)}\;,\;w\right\rangle\geq 0\quad\forall\,w\underset{\mathbb{R}_+^M}{\succeq}0 \tag{517}$$

where

$$\overset{\to Y-X}{df(X)}=\left[\begin{array}{c}\operatorname{tr}\big(\nabla f_1(X)^T(Y-X)\big)\\ \operatorname{tr}\big(\nabla f_2(X)^T(Y-X)\big)\\ \vdots\\ \operatorname{tr}\big(\nabla f_M(X)^T(Y-X)\big)\end{array}\right]\in\mathbb{R}^M \tag{518}$$

Necessary and sufficient discretization (420) allows relaxation of the semi-infinite number of conditions $w\succeq 0$ instead to $w\in\{e_i\,,\;i=1\ldots M\}$ the extreme directions of the nonnegative orthant. Each extreme direction picks out a real entry $f_i$ and $\overset{\to Y-X}{df(X)}_i$ from vector-valued function $f$ and its directional derivative $\overset{\to Y-X}{df(X)}$, then Theorem 3.1.9.0.1 applies.

The vector-valued function case (516) is therefore a straightforward application of the first-order convexity condition for real functions to each entry of the vector-valued function.

### 3.1.11 second-order convexity condition

Again, by discretization (420), 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 $f(X):\mathbb{R}^p\to\mathbb{R}^M$, a twice differentiable vector-valued function with vector argument on open convex domain,

$$\nabla^2 f_i(X)\underset{\mathbb{S}_+^p}{\succeq}0\quad\forall\,X\in\operatorname{dom}f\;,\quad i=1\ldots M \tag{519}$$

[3.9] 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 the scope of the Taylor series (§D.1.7). The right-hand side of the inequality (516) is the first-order Taylor series expansion of $f$ about $X$.

is a necessary and sufficient condition for convexity of $f$. Obviously, when $M=1$, this convexity condition also serves for a real function. Intuitively, condition (519) precludes points of inflection, as in Figure **61** on page 220.

Strict inequality is a sufficient condition for strict convexity, but that is nothing new; *videlicet*, the 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 is reliably true.

**3.1.11.0.1 Exercise.** *Real fractional function.* (*confer* §3.1.4, §3.1.7.0.4) Prove that real function $f(x,y)=x/y$ is not convex on the nonnegative orthant. Also exhibit this in a plot of the function. **(**In fact, $f$ is *quasilinear* (p.222) on $\{y>0\}$.**)** ▼

#### 3.1.11.1 second-order ⇒ first-order condition

For a twice-differentiable real function $f_i(X):\mathbb{R}^p\to\mathbb{R}$ having open domain, a consequence of the *mean value theorem* from calculus allows compression of its complete Taylor series expansion about $X\in\operatorname{dom} f_i$ (§D.1.7) to three terms: On some open interval of $\|Y\|$ so each and every line segment $[X,Y]$ belongs to $\operatorname{dom} f_i$, there exists an $\alpha\in[0,1]$ such that [299, §1.2.3] [30, §1.1.4]

$$f_i(Y) = f_i(X) + \nabla f_i(X)^T(Y-X) + \frac{1}{2}(Y-X)^T\nabla^2 f_i\big(\alpha X+(1-\alpha)Y\big)(Y-X) \tag{520}$$

The first-order condition for convexity (510) follows directly from this and the second-order condition (519).

## 3.2 Matrix-valued convex 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 (521) with respect to the positive semidefinite cone $\mathbb{S}_+^M$ in the ambient space of symmetric matrices.[3.10]

[3.10]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

**3.2.0.0.1 Definition.** *Convex matrix-valued function*:
**1)** *Matrix-definition.*
A function $g(X) : \mathbb{R}^{p\times k} \rightarrow \mathbb{S}^M$ is convex in $X$ iff $\operatorname{dom} g$ is a convex set and, for each and every $Y, Z \in \operatorname{dom} g$ and all $0 \leq \mu \leq 1$ [158, §2.3.7]

$$g(\mu Y + (1-\mu)Z) \underset{\mathbb{S}^M_+}{\preceq} \mu\, g(Y) + (1-\mu)g(Z) \tag{521}$$

Reversing the sense of the inequality flips this definition to concavity. Strict convexity is defined less a stroke of the pen in (521) similarly to (422).
**2)** *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 (521). By dual generalized inequalities we have the equivalent but more broad criterion, (§2.13.5)

$$g \text{ convex } \Leftrightarrow \langle W\,,\ g\rangle \text{ convex } \quad \forall\, W \underset{\mathbb{S}^M_+}{\succeq} 0 \tag{522}$$

Strict convexity on both sides requires *caveat* $W \neq \mathbf{0}$. Because the set of all extreme directions for the positive semidefinite cone (§2.9.2.4) 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. △

### 3.2.1 first-order convexity condition, matrix function

From the *scalar-definition* we have, for differentiable matrix-valued function $g$ and for each and every real vector $w$ of unit norm $\|w\| = 1$,

$$w^T g(Y) w \geq w^T g(X) w \,+\, w^T \overset{\rightarrow Y-X}{dg(X)}\, w \tag{523}$$

that follows immediately from the first-order condition (509) for convexity of a real function because

$$w^T \overset{\rightarrow Y-X}{dg(X)}\, w \;=\; \left\langle \nabla_X\, w^T g(X) w \;,\; Y - X \right\rangle \tag{524}$$

respect to the nonnegative orthant $\mathbb{R}^{m\times k}_+$. Symmetric convex functions share the same benefits as symmetric matrices. Horn & Johnson [150, §7.7] liken symmetric matrices to real numbers, and (symmetric) positive definite matrices to positive real numbers.

where $\overset{\rightarrow Y-X}{dg}(X)$ is the directional derivative (§D.1.4) of function $g$ at $X$ in direction $Y-X$. By discretized dual generalized inequalities, (§2.13.5)

$$g(Y)-g(X)-\overset{\rightarrow Y-X}{dg}(X) \underset{\mathbb{S}_+^M}{\succeq} 0 \;\Leftrightarrow\; \left\langle g(Y)-g(X)-\overset{\rightarrow Y-X}{dg}(X)\;,\; ww^T\right\rangle \geq 0 \;\;\forall\, ww^T(\underset{\mathbb{S}_+^M}{\succeq} 0) \tag{525}$$

For each and every $X,Y\in \operatorname{dom} g$ (*confer* (516))

$$g(Y) \underset{\mathbb{S}_+^M}{\succeq} g(X) + \overset{\rightarrow Y-X}{dg}(X) \tag{526}$$

must therefore be necessary and sufficient for convexity of a matrix-valued function of matrix variable on open convex domain.

### 3.2.2 epigraph of matrix-valued function, sublevel sets

We generalize the epigraph to a continuous matrix-valued function $g(X) : \mathbb{R}^{p\times k}\to\mathbb{S}^M$ :

$$\operatorname{epi} g \triangleq \{(X\,,\,T)\in\mathbb{R}^{p\times k}\times\mathbb{S}^M \mid X\in\operatorname{dom} g\;,\;\; g(X) \underset{\mathbb{S}_+^M}{\preceq} T\,\} \tag{527}$$

from which it follows

$$g \text{ convex } \Leftrightarrow \operatorname{epi} g \text{ convex} \tag{528}$$

Proof of necessity is similar to that in §3.1.7 on page 196.

Sublevel sets of a matrix-valued convex function corresponding to each and every $S\in\mathbb{S}^M$ (*confer* (462))

$$\mathcal{L}_S\, g \triangleq \{X\in\operatorname{dom} g \mid g(X) \underset{\mathbb{S}_+^M}{\preceq} S\,\} \subseteq \mathbb{R}^{p\times k} \tag{529}$$

are convex. There is no converse.

**3.2.2.0.1 Example.** *Matrix fractional function* redux.
Generalizing Example 3.1.7.0.4 consider a matrix-valued function of two variables on $\operatorname{dom} g = \mathbb{S}^N_+ \times \mathbb{R}^{n\times N}$ for small positive constant $\epsilon$ (*confer* (1639))

$$g(A\,,X) = \epsilon\, X(A+\epsilon\, I)^{-1}X^T \tag{530}$$

where the inverse always exists by (1255). This function is convex simultaneously in both variables over the entire positive semidefinite cone $\mathbb{S}^N_+$ and all $X\in\mathbb{R}^{n\times N}$: Consider Schur-form (1314) from §A.4: for $T\in\mathbb{S}^n$

$$\begin{array}{c} G(A\,,X\,,T) = \begin{bmatrix} A+\epsilon\, I & X^T \\ X & \epsilon^{-1}\,T \end{bmatrix} \succeq 0 \\ \Leftrightarrow \\ T-\epsilon\, X(A+\epsilon\, I)^{-1}X^T \succeq 0 \\ A+\epsilon\, I \succ 0 \end{array} \tag{531}$$

By Theorem 2.1.9.0.1, inverse image of the positive semidefinite cone $\mathbb{S}^{N+n}_+$ under affine mapping $G(A\,,X\,,T)$ is convex. Function $g(A\,,X)$ is convex on $\mathbb{S}^N_+\times\mathbb{R}^{n\times N}$ because its epigraph is that inverse image:

$$\operatorname{epi} g(A\,,X) = \left\{(A\,,X\,,T) \mid A+\epsilon\, I\succ 0\,,\ \epsilon\, X(A+\epsilon\, I)^{-1}X^T \preceq T\right\} = G^{-1}\left(\mathbb{S}^{N+n}_+\right) \tag{532}$$

□

## 3.2.3 second-order condition, matrix function

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 $g(X) : \mathbb{R}^{p\times k}\to\mathbb{S}^M$ and particular $X\,,Y\in\mathbb{R}^{p\times k}$ not necessarily in that function's domain, then we say a line $\{X+t\,Y \mid t\in\mathbb{R}\}$ passes through $\operatorname{dom} g$ when $X+t\,Y\in\operatorname{dom} g$ over some interval of $t\in\mathbb{R}$.

**3.2.3.0.1 Theorem.** *Line theorem.* [46, §3.1.1]
Matrix-valued function $g(X) : \mathbb{R}^{p\times k}\to\mathbb{S}^M$ 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.

**3.2.3.0.2 Definition.** *Differentiable convex matrix-valued function.*
Matrix-valued function $g(X):\mathbb{R}^{p\times k}\to\mathbb{S}^M$ is convex in $X$ iff $\operatorname{dom} g$ is an open convex set, and its second derivative $g''(X+t\,Y):\mathbb{R}\to\mathbb{S}^M$ is positive semidefinite on each point of intersection along every line $\{X+t\,Y \mid t\in\mathbb{R}\}$ that intersects $\operatorname{dom} g$; *id est*, iff for each and every $X, Y\in\mathbb{R}^{p\times k}$ such that $X+t\,Y\in\operatorname{dom} g$ over some open interval of $t\in\mathbb{R}$

$$\frac{d^2}{dt^2}\,g(X+t\,Y)\underset{\mathbb{S}^M_+}{\succeq}0 \tag{533}$$

Similarly, if

$$\frac{d^2}{dt^2}\,g(X+t\,Y)\underset{\mathbb{S}^M_+}{\succ}0 \tag{534}$$

then $g$ is strictly convex; the converse is generally false. [46, §3.1.4][3.11] △

**3.2.3.0.3 Example.** *Matrix inverse.* (*confer* §3.1.5)
The matrix-valued function $X^\mu$ is convex on $\operatorname{int}\mathbb{S}^M_+$ for $-1\leq\mu\leq 0$ or $1\leq\mu\leq 2$ and concave for $0\leq\mu\leq 1$. [46, §3.6.2] In particular, the function $g(X)=X^{-1}$ is convex on $\operatorname{int}\mathbb{S}^M_+$. For each and every $Y\in\mathbb{S}^M$ (§D.2.1, §A.3.1.0.5)

$$\frac{d^2}{dt^2}\,g(X+t\,Y)\;=\;2(X+t\,Y)^{-1}\,Y(X+t\,Y)^{-1}\,Y(X+t\,Y)^{-1}\underset{\mathbb{S}^M_+}{\succeq}0 \tag{535}$$

on some open interval of $t\in\mathbb{R}$ such that $X+t\,Y\succ 0$. Hence, $g(X)$ is convex in $X$. This result is extensible;[3.12] $\operatorname{tr}X^{-1}$ is convex on that same domain. [150, §7.6, prob.2] [41, §3.1, exer.25] □

**3.2.3.0.4 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 $t\in\mathbb{R}$ (§D.2.1)

[3.11] Quadratic forms constitute a notable exception where the strict-case converse is reliably true.
[3.12] $d/dt\operatorname{tr}g(X+t\,Y)=\operatorname{tr}d/dt\,g(X+t\,Y)$. [151, p.491]

$$\frac{d^2}{dt^2}\, g(X+t\,Y) \;=\; \frac{d^2}{dt^2}(X+t\,Y)^2 \;=\; 2Y^2 \tag{536}$$

which is positive semidefinite when $Y$ is symmetric because then $Y^2 = Y^T Y$ (1261).[3.13]

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.574), and strictly convex when $A$ is positive definite and $X$ is skinny-or-square full-rank (Corollary A.3.1.0.5). $\square$

**3.2.3.0.5 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* [118] 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

$$\frac{d^2}{dt^2} e^{X+t\,Y} \;=\; Y e^{X+t\,Y} Y \underset{\mathbb{S}_+^M}{\succeq} 0\;, \qquad (XY)^T \;=\; XY \tag{537}$$

because all circulant matrices are *commutative* and, for symmetric matrices, $XY = YX \Leftrightarrow (XY)^T = XY$ (1279). Given symmetric argument, the matrix exponential always resides interior to the cone of positive semidefinite matrices in the symmetric matrix subspace; $e^A \succ 0 \;\; \forall A \in \mathbb{S}^M$ (1636). Then for any matrix $Y$ of compatible dimension, $Y^T e^A Y$ is positive semidefinite. (§A.3.1.0.5)

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. [183, §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 [46, §3.1.5]. $\square$

There are, of course, multifarious methods to determine function convexity, [46] [30] [88] each of them efficient when appropriate.

[3.13] By (1280) in §A.3.1, changing the domain instead to all symmetric and nonsymmetric positive semidefinite matrices, for example, will not produce a convex function.

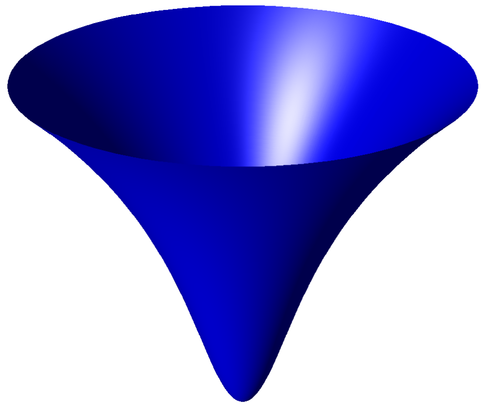


Figure 61: Iconic unimodal differentiable quasiconvex function of two variables graphed in $\mathbb{R}^{\mathbf{2}} \times \mathbb{R}$ on some open disc in $\mathbb{R}^{\mathbf{2}}$. Note reversal of *curvature* in direction of gradient.

**3.2.3.0.6 Exercise.** *log det function.*
Show by two different methods: $\log\det X$ is concave on the interior of the positive semidefinite cone. ▼

## 3.3 Quasiconvex

Quasiconvex functions [46, §3.4] [148] [247] [280] [178, §2] are useful in practical problem solving 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 **61**.

**3.3.0.0.1 Definition.** *Quasiconvex function.*
$f(X) : \mathbb{R}^{p\times k} \to \mathbb{R}$ is a quasiconvex function of matrix $X$ iff $\operatorname{dom} f$ is a convex set and for each and every $Y, Z \in \operatorname{dom} f$, $0 \leq \mu \leq 1$

$$f(\mu Y + (1-\mu)Z) \;\leq\; \max\{f(Y)\,,\; f(Z)\} \tag{538}$$

A quasiconcave function is determined:

$$f(\mu Y+(1-\mu)Z)\;\geq\;\min\{f(Y)\,,\;f(Z)\}\tag{539}$$

△

Unlike convex functions, quasiconvex functions are not necessarily continuous; *e.g.*, quasiconcave $\operatorname{rank}(X)$ on $\mathbb{S}_+^M$ (§2.9.2.6.2) and $\operatorname{card}(x)$ on $\mathbb{R}_+^M$. Although insufficient for convex functions, convexity of each and every sublevel set serves as a definition of quasiconvexity:

**3.3.0.0.2 Definition.** *Quasiconvex multidimensional function.*
Scalar-, vector-, or matrix-valued function $g(X):\mathbb{R}^{p\times k}\to\mathbb{S}^M$ is a quasiconvex function of matrix $X$ iff $\operatorname{dom} g$ is a convex set and the sublevel set corresponding to each and every $S\in\mathbb{S}^M$

$$\mathcal{L}_S\, g=\{X\in\operatorname{dom} g \mid g(X)\preceq S\}\;\subseteq\;\mathbb{R}^{p\times k}\tag{529}$$

is convex. Vectors are compared with respect to the nonnegative orthant $\mathbb{R}_+^M$ while matrices are with respect to the positive semidefinite cone $\mathbb{S}_+^M$.

Convexity of the *superlevel set* corresponding to each and every $S\in\mathbb{S}^M$, likewise

$$\mathcal{L}^S g=\{X\in\operatorname{dom} g \mid g(X)\succeq S\}\;\subseteq\;\mathbb{R}^{p\times k}\tag{540}$$

is necessary and sufficient for quasiconcavity. △

**3.3.0.0.3 Exercise.** *Nonconvexity of matrix product.*
Consider the real function on a positive definite domain

$$f(X)=\operatorname{tr}(X_1X_2)\;,\qquad \operatorname{dom} f\triangleq\begin{bmatrix}\operatorname{rel\,int}\mathbb{S}_+^N & \mathbf{0}\\ \mathbf{0} & \operatorname{rel\,int}\mathbb{S}_+^N\end{bmatrix}\tag{541}$$

where

$$X\triangleq\begin{bmatrix}X_1 & \mathbf{0}\\ \mathbf{0} & X_2\end{bmatrix}\underset{\mathbb{S}_+^{2N}}{\succ}0\tag{542}$$

with superlevel sets

$$\begin{aligned}\mathcal{L}^s f&=\{X\in\operatorname{dom} f \mid f(X)\geq s\}\\ &=\{X\in\operatorname{dom} f \mid \langle X_1\,,\,X_2\rangle\geq s\}\end{aligned}\tag{543}$$

Prove: $f(X)$ is not quasiconcave except when $N=1$, nor is it quasiconvex unless $X_1 = X_2$. ▼

When a function is simultaneously quasiconvex and quasiconcave, it is called *quasilinear.* Quasilinear functions are completely determined by convex level sets. One-dimensional function $f(x)=x^3$ and vector-valued signum function $\operatorname{sgn}(x)$ for example, are quasilinear. Any monotonic function is quasilinear.[3.14]

## 3.4 Salient properties of convex and quasiconvex functions

1. 
   - A convex (or concave) function is assumed continuous but not necessarily differentiable on the relative interior of its domain. [230, §10]
   - A quasiconvex (or quasiconcave) function is not necessarily a continuous function.

2. convexity ⇒ quasiconvexity ⇔ convex sublevel sets
   concavity ⇒ quasiconcavity ⇔ convex superlevel sets
   monotonicity ⇒ quasilinearity ⇔ convex level sets

3. 
   - (*homogeneity*) Convexity, concavity, quasiconvexity, and quasiconcavity are invariant to nonnegative scaling of function.
   - $g$ convex $\Leftrightarrow$ $-g$ concave
     $g$ quasiconvex $\Leftrightarrow$ $-g$ quasiconcave

4. The *line theorem* (§3.2.3.0.1) translates identically to quasiconvexity (quasiconcavity). [46, §3.4.2]

5. (*affine transformation of argument*) Composition $g(h(X))$ of a convex (concave) function $g$ with any affine function $h : \mathbb{R}^{m\times n} \to \mathbb{R}^{p\times k}$ remains convex (concave) in $X \in \mathbb{R}^{m\times n}$, where $h(\mathbb{R}^{m\times n}) \cap \operatorname{dom} g \neq \emptyset$. [148, §B.2.1] Likewise for the quasiconvex (quasiconcave) functions $g$.

[3.14] *e.g.*, a monotonic concave function is therefore quasiconvex, but it is best to avoid this confusion of terms.

6. 
- A nonnegatively weighted sum of (strictly) convex (concave) functions remains (strictly) convex (concave). (§3.1.1.0.1)
  Pointwise supremum (infimum) of convex (concave) functions remains convex (concave). (Figure **56**) [230, §5]
- A nonnegatively weighted maximum (minimum) of quasiconvex (quasiconcave) functions remains quasiconvex (quasiconcave).
  Pointwise supremum (infimum) of quasiconvex (quasiconcave) functions remains quasiconvex (quasiconcave).

# Chapter 4

# Semidefinite programming

> *Prior to 1984,*[4.1] *linear and nonlinear programming, 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) [98]

Given some application of convex analysis, it may at first seem puzzling why the 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. (§C) If a problem can be expressed in *convex form*, rather, then there exist computer programs providing efficient numerical global solution. [253] [117] [291] [292] [293] [299]

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: *A fundamental property of convex optimization problems is that any locally optimal point is also (globally) optimal.* [46, §4.2.2] [229, §1] Given convex real objective function $g$ and convex feasible set $\mathcal{C} \subseteq \operatorname{dom} g$, which is the set of all variable values satisfying the problem constraints, we

[4.1] nascence of interior-point methods of solution [273] [289], have the generic convex optimization problem

$$\begin{array}{ll}\underset{X}{\text{minimize}} & g(X)\\ \text{subject to} & X\in\mathcal{C}\end{array}\tag{544}$$

where constraints are abstract here in the membership of variable $X$ to feasible set $\mathcal{C}$. 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 (necessarily convex), as opposed to the larger set of all convex equality constraint functions having convex level sets, make convex optimization tractable.

Similarly, the problem

$$\begin{array}{ll}\underset{X}{\text{maximize}} & g(X)\\ \text{subject to} & X\in\mathcal{C}\end{array}\tag{545}$$

is convex were $g$ a real concave function. As conversion to convex form is not always possible, there is much ongoing research to determine which problem classes have convex expression or relaxation. [27] [44] [102] [204] [260] [100]

## 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.*
>
> −*SIAM News*, 2002. [79, p.9]

Consider a prototypical *conic problem* (p) and its dual (d): [217, §3.3.1] [175, §2.1]

$$\text{(p)}\quad\begin{array}{ll}\underset{x}{\text{minimize}} & c^Tx\\ \text{subject to} & x\in\mathcal{K}\\ & Ax=b\end{array}\qquad\qquad\begin{array}{ll}\underset{y,s}{\text{maximize}} & b^Ty\\ \text{subject to} & s\in\mathcal{K}^*\\ & A^Ty+s=c\end{array}\quad\text{(d)}\tag{263}$$

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* [64]. More generally, each optimization problem is convex when $\mathcal{K}$ is a closed convex cone. Unlike the optimal objective value, a solution to each problem is not necessarily unique; in other words, the optimal solution set $\{x^\star\}$ or $\{y^\star, s^\star\}$ is convex and may comprise more than a single point although the corresponding optimal objective value is unique when the feasible set is nonempty.

When $\mathcal{K}$ is the self-dual cone of positive semidefinite matrices in the subspace of symmetric matrices, then each conic problem is called a *semidefinite program* (SDP); [204, §6.4] primal problem (P) having matrix variable $X\in\mathbb{S}^n$ while corresponding dual (D) has matrix *slack variable* $S\in\mathbb{S}^n$ and vector variable $y=[y_i]\in\mathbb{R}^m$: [8] [9, §2] [299, §1.3.8]

$$\text{(P)}\quad \begin{array}{ll}\underset{X\in\mathbb{S}^n}{\text{minimize}} & \langle C\,,\,X\rangle\\ \text{subject to} & X\succeq 0\\ & A\,\text{svec}\,X=b\end{array} \qquad\qquad \begin{array}{ll}\underset{y\in\mathbb{R}^m,\,S\in\mathbb{S}^n}{\text{maximize}} & \langle b\,,\,y\rangle\\ \text{subject to} & S\succeq 0\\ & \text{svec}^{-1}(A^Ty)+S=C\end{array}\quad\text{(D)} \tag{546}$$

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

$$A \triangleq \begin{bmatrix}\text{svec}(A_1)^T\\ \vdots\\ \text{svec}(A_m)^T\end{bmatrix}\in\mathbb{R}^{m\times n(n+1)/2} \tag{547}$$

where $A_i\in\mathbb{S}^n$, $i=1\ldots m$, are given. Thus

$$\begin{aligned} A\,\text{svec}\,X &= \begin{bmatrix}\langle A_1\,,\,X\rangle\\ \vdots\\ \langle A_m\,,\,X\rangle\end{bmatrix}\\ \text{svec}^{-1}(A^Ty) &= \sum_{i=1}^m y_iA_i \end{aligned} \tag{548}$$

The vector inner-product for matrices is defined in the Euclidean/Frobenius sense in the isomorphic vector space $\mathbb{R}^{n(n+1)/2}$; *id est*,

$$\langle C\,,\,X\rangle \triangleq \text{tr}(C^TX) = \text{svec}(C)^T\text{svec}\,X \tag{31}$$

where $\operatorname{svec} X$ defined by (47) denotes symmetric vectorization.

Semidefinite programming has emerged recently to prominence primarily because it admits a new class of problem previously unsolvable by convex optimization techniques, [44] secondarily because it theoretically subsumes other convex techniques such as linear, quadratic, and *second-order cone programming.* Determination of the Riemann mapping function from complex analysis [211] [24, §8, §13], for example, can be posed as a semidefinite program.

### 4.1.1 Maximal complementarity

It has been shown that contemporary *interior-point methods* (developed *circa* 1990 [102]) [46, §11] [290] [214] [204] [9] [98] for numerical solution of semidefinite programs can converge to a solution of *maximal complementarity*; [126, §5] [298] [185] [109] not a vertex-solution but a solution of highest cardinality or rank among all optimal solutions.[4.2] [299, §2.5.3]

#### 4.1.1.1 Reduced-rank solution

A simple rank reduction algorithm for construction of a primal optimal solution $X^\star$ to (546P) 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; [20, §II.13.1] specifically, it asserts an extreme point (§2.6.0.0.1) of the primal feasible set $\mathcal{A} \cap \mathbb{S}_+^n$ satisfies upper bound

$$\operatorname{rank} X \leq \left\lfloor \frac{\sqrt{8m+1}-1}{2} \right\rfloor \tag{232}$$

where, given $A \in \mathbb{R}^{m\times n(n+1)/2}$ and $b \in \mathbb{R}^m$

$$\mathcal{A} \triangleq \{X \in \mathbb{S}^n \mid A \operatorname{svec} X = b\} \tag{549}$$

is the affine subset from primal problem (546P).

[4.2] This characteristic might be regarded as a disadvantage to this method of numerical solution, but this behavior is not certain and depends on solver implementation.

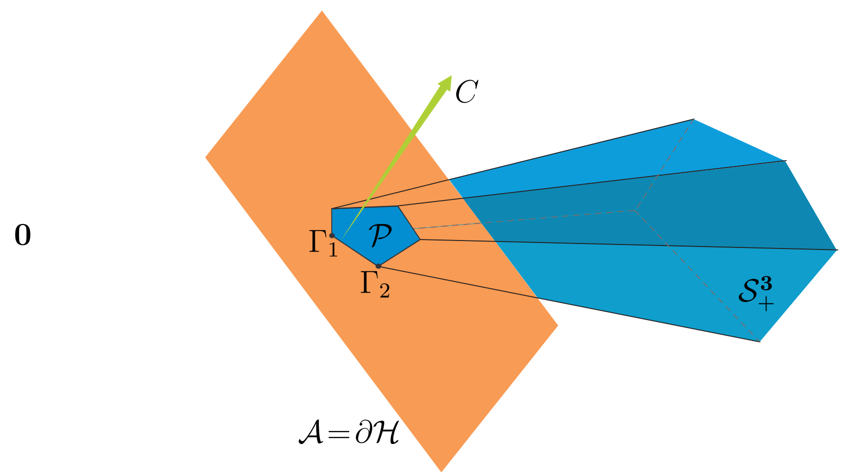



Figure 62: Visualizing positive semidefinite cone in high dimension: Proper polyhedral cone $\mathcal{S}_+^{\mathbf{3}} \subset \mathbb{R}^{\mathbf{3}}$ representing positive semidefinite cone $\mathbb{S}_+^{\mathbf{3}} \subset \mathbb{S}^{\mathbf{3}}$; analogizing its intersection with hyperplane $\mathbb{S}_+^{\mathbf{3}} \cap \partial\mathcal{H}$. Number of facets is arbitrary (analogy is not inspired by eigen decomposition). The rank-0 positive semidefinite matrix corresponds to the origin in $\mathbb{R}^{\mathbf{3}}$, rank-1 positive semidefinite matrices correspond to the edges of the polyhedral cone, rank-2 to the facet relative interiors, and rank-3 to the polyhedral cone interior. Vertices $\Gamma_1$ and $\Gamma_2$ are extreme points of polyhedron $\mathcal{P} = \partial\mathcal{H} \cap \mathcal{S}_+^{\mathbf{3}}$, and extreme directions of $\mathcal{S}_+^{\mathbf{3}}$. 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 **17**)

#### 4.1.1.2 Coexistence of low- and high-rank solutions; analogy

That low-rank and high-rank optimal solutions $\{X^\star\}$ of (546P) coexist may be grasped with the following analogy: We compare a proper polyhedral cone $\mathcal{S}^3_+$ in $\mathbb{R}^3$ (illustrated in Figure 62) to the positive semidefinite cone $\mathbb{S}^3_+$ in isometrically isomorphic $\mathbb{R}^6$, difficult to visualize. The analogy is good:

- $\operatorname{int}\mathbb{S}^3_+$ is constituted by rank-3 matrices
  $\operatorname{int}\mathcal{S}^3_+$ has three dimensions

- boundary $\partial\mathbb{S}^3_+$ contains rank-0, rank-1, and rank-2 matrices
  boundary $\partial\mathcal{S}^3_+$ 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 $\mathbb{S}^3_+$ and $\mathcal{S}^3_+$. The set of all rank-1 symmetric matrices in this dimension
$$\left\{G\in\mathbb{S}^3_+ \mid \operatorname{rank}G=1\right\} \tag{550}$$
  is not a connected set.

- *In any SDP feasibility problem, an SDP feasible solution with the lowest rank must be an extreme point of the feasible set. Thus, there must exist an SDP objective function such that this lowest-rank feasible solution uniquely optimizes it.* −Ye, 2006

- Rank of a sum of members $F+G$ in Lemma 2.9.2.6.1 and location of a difference $F-G$ in §2.9.2.9.1 similarly hold for $\mathbb{S}^3_+$ and $\mathcal{S}^3_+$.

- 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 $\partial\mathbb{S}^3_+$ to $\operatorname{int}\mathbb{S}^3_+$ is infinitesimal (§2.1.7.1.1)
  distance from any point in $\partial\mathcal{S}^3_+$ to $\operatorname{int}\mathcal{S}^3_+$ is infinitesimal

- faces of $\mathbb{S}^3_+$ correspond to faces of $\mathcal{S}^3_+$ (*confer* Table 2.9.2.3.1)

| | $k$ | $\dim \mathcal{F}(\mathcal{S}^3_+)$ | $\dim \mathcal{F}(\mathbb{S}^3_+)$ | $\dim \mathcal{F}(\mathbb{S}^3_+ \ni \text{rank-}k \text{ matrix})$ |
|---|---|---|---|---|
| | 0 | 0 | 0 | 0 |
| boundary | 1 | 1 | 1 | 1 |
| | 2 | 2 | 3 | 3 |
| interior | 3 | 3 | 6 | 6 |

Integer $k$ indexes $k$-dimensional faces $\mathcal{F}$ of $\mathcal{S}^3_+$. Positive semidefinite cone $\mathbb{S}^3_+$ has four kinds of faces, including cone itself ($k=3$, boundary + interior), whose dimensions in isometrically isomorphic $\mathbb{R}^6$ are listed under $\dim \mathcal{F}(\mathbb{S}^3_+)$. Smallest face $\mathcal{F}(\mathbb{S}^3_+ \ni \text{rank-}k \text{ matrix})$ that contains a rank-$k$ positive semidefinite matrix has dimension $k(k+1)/2$ by (191).

- For $\mathcal{A}$ equal to intersection of $m$ hyperplanes having independent normals, and for $X \in \mathcal{S}^3_+ \cap \mathcal{A}$, we have $\operatorname{rank} X \leq m$; the analogue to (232).

**Proof.** With reference to Figure **62**: 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 $X \in \partial\mathcal{H} \cap \mathcal{S}^3_+$ there exists an $X$ such that $\operatorname{rank} X \leq 1$. 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 $\operatorname{rank} X \leq 2$.

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 $\mathcal{S}^3_+$ is also the greatest upper bound: $\operatorname{rank} X \leq 3$. ♦

**4.1.1.2.1 Example.** *Optimization on* $\mathcal{A} \cap \mathbb{S}^3_+$ .
Consider minimization of the real linear function $\langle C\,,\,X\rangle$ on

$$\mathcal{P} \triangleq \mathcal{A} \cap \mathbb{S}^3_+ \tag{551}$$

a polyhedral feasible set;

$$\begin{array}{rll} f_0^\star \triangleq & \underset{X}{\text{minimize}} & \langle C\,,\,X\rangle \\ & \text{subject to} & X \in \mathcal{A} \cap \mathbb{S}^3_+ \end{array} \tag{552}$$

As illustrated for particular vector $C$ and hyperplane $\mathcal{A} = \partial\mathcal{H}$ in Figure **62**, this linear function is minimized (*confer* Figure **17**) on any $X$ belonging to the face of $\mathcal{P}$ containing extreme points $\{\Gamma_1\,,\,\Gamma_2\}$ and all the rank-2 matrices in between; *id est*, on any $X$ belonging to the face of $\mathcal{P}$

$$\mathcal{F}(\mathcal{P}) = \{X \mid \langle C\,,\,X\rangle = f_0^\star\} \cap \mathcal{A} \cap \mathbb{S}^3_+ \tag{553}$$

exposed by the hyperplane $\{X \mid \langle C\,,\,X\rangle = f_0^\star\}$. In other words, the set of all optimal points $X^\star$ is a face of $\mathcal{P}$

$$\{X^\star\} \;=\; \mathcal{F}(\mathcal{P}) \;=\; \overline{\Gamma_1\Gamma_2} \tag{554}$$

comprising rank-1 and rank-2 positive semidefinite matrices. Rank 1 is the upper bound on existence in the feasible set $\mathcal{P}$ for this case $m = 1$ hyperplane constituting $\mathcal{A}$. The rank-1 matrices $\Gamma_1$ and $\Gamma_2$ in face $\mathcal{F}(\mathcal{P})$ are extreme points of that face and (by transitivity (§2.6.1.2)) extreme points of the intersection $\mathcal{P}$ as well. As predicted by analogy to Barvinok's Proposition 2.9.3.0.1, the upper bound on rank of $X$ existent in the feasible set $\mathcal{P}$ is satisfied by an extreme point. The upper bound on rank of an optimal solution $X^\star$ existent in $\mathcal{F}(\mathcal{P})$ is thereby also satisfied by an extreme point of $\mathcal{P}$ precisely because $\{X^\star\}$ constitutes $\mathcal{F}(\mathcal{P})$ ;[4.3] in particular,

$$\{X^\star \in \mathcal{P} \mid \operatorname{rank} X^\star \leq 1\} \;=\; \{\Gamma_1\,,\,\Gamma_2\} \;\subseteq\; \mathcal{F}(\mathcal{P}) \tag{555}$$

As all linear functions on a polyhedron are minimized on a face, [64] [184] [203] [206] by analogy we so demonstrate coexistence of optimal solutions $X^\star$ of (546P) having assorted rank. □

[4.3] and every face contains a subset of the extreme points of $\mathcal{P}$ by the *extreme existence theorem* (§2.6.0.0.2). This means: because the affine subset $\mathcal{A}$ and hyperplane $\{X \mid \langle C\,,\,X\rangle = f_0^\star\}$ must intersect a whole face of $\mathcal{P}$, calculation of an upper bound on rank of $X^\star$ ignores counting the hyperplane when determining $m$ in (232).

#### 4.1.1.3 Previous work

Barvinok showed [21, §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 $X^\star$ to (546P) (substituting $\tilde{C}$) is an extreme point of $\mathcal{A} \cap \mathbb{S}_+^n$ and satisfies upper bound (232).[4.4] Given arbitrary positive definite $C$, this means nothing inherently guarantees that an optimal solution $X^\star$ to problem (546P) satisfies (232); certainly nothing given any symmetric matrix $C$, as the problem is posed. This can be proved by example:

**4.1.1.3.1 Example.** (Ye) *Maximal Complementarity.*
Assume dimension $n$ to be an even positive number. Then the particular instance of problem (546P),

$$\begin{array}{cl}\underset{X\in\mathbb{S}^n}{\text{minimize}} & \left\langle \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & 2I \end{bmatrix} , X \right\rangle \\ \text{subject to} & X \succeq 0 \\ & \langle I , X \rangle = n \end{array} \tag{556}$$

has optimal solution

$$X^\star = \begin{bmatrix} 2I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{S}^n \tag{557}$$

with an equal number of twos and zeros along the main diagonal. Indeed, optimal solution (557) is a terminal solution along the *central path* taken by the interior-point method as implemented in [299, §2.5.3]; it is also a solution of highest rank among all optimal solutions to (556). Clearly, rank of this primal optimal solution exceeds by far a rank-1 solution predicted by upper bound (232). □

#### 4.1.1.4 Later developments

This rational example (556) indicates the need for a more generally applicable and simple algorithm to identify an optimal solution $X^\star$ satisfying Barvinok's Proposition 2.9.3.0.1. We will review such an algorithm in §4.3, but first we provide more background.

[4.4] Further, the set of all such $\tilde{C}$ in that neighborhood is open and dense.

## 4.2 Framework

### 4.2.1 Feasible sets

Denote by $\mathcal{C}$ and $\mathcal{C}^*$ the convex sets of primal and dual points respectively satisfying the primal and dual constraints in (546), each assumed nonempty;

$$\begin{aligned}
\mathcal{C} &= \left\{ X \in \mathbb{S}^n_+ \mid \begin{bmatrix} \langle A_1 \,,\, X \rangle \\ \vdots \\ \langle A_m \,,\, X \rangle \end{bmatrix} = b \right\} = \mathcal{A} \cap \mathbb{S}^n_+ \\
\mathcal{C}^* &= \left\{ S \in \mathbb{S}^n_+ \,,\; y = [y_i] \in \mathbb{R}^m \mid \sum_{i=1}^m y_i A_i + S = C \right\}
\end{aligned} \tag{558}$$

These are the *primal feasible set* and *dual feasible set* in domain intersection of the respective constraint functions. 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 $\mathbb{R}^{n(n+1)/2}$. The affine subset has dimension $n(n+1)/2 - m$ when the $A_i$ are linearly independent. Dual feasible set $\mathcal{C}^*$ is the Cartesian product of the positive semidefinite cone with its inverse image (§2.1.9.0.1) under affine transformation $C - \sum y_i A_i$ .[4.5] Both sets are closed and convex and the objective functions on a Euclidean vector space are linear, hence (546P) and (546D) are convex optimization problems.

#### 4.2.1.1 $\mathcal{A} \cap \mathbb{S}^n_+$ emptiness determination via Farkas' lemma

**4.2.1.1.1 Lemma.** *Semidefinite Farkas' lemma.*
Given an arbitrary set $\{A_i \in \mathbb{S}^n,\ i = 1 \ldots m\}$ and a vector $b = [b_i] \in \mathbb{R}^m$, define the affine subset

$$\mathcal{A} = \{X \in \mathbb{S}^n \mid \langle A_i \,,\, X \rangle = b_i \,,\; i = 1 \ldots m\} \tag{549}$$

Primal feasible set $\mathcal{A} \cap \mathbb{S}^n_+$ 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 $\sum_{i=1}^m y_i A_i \succeq 0$.

[4.5] The inequality $C - \sum y_i A_i \succeq 0$ follows directly from (546D) (§2.9.0.1.1) and is known as a *linear matrix inequality.* (§2.13.5.1.1) Because $\sum y_i A_i \preceq C$, matrix $S$ is known as a slack variable (a term borrowed from linear programming [64]) since its inclusion raises this inequality to equality.

Equivalently, primal feasible set $\mathcal{A}\cap\mathbb{S}_+^n$ is nonempty if and only if $y^Tb\succeq 0$ holds for each and every norm-1 vector $\|y\|=1$ such that $\sum_{i=1}^m y_iA_i\succeq 0$. ⋄

*Semidefinite Farkas' lemma* follows directly from a membership relation (§2.13.2.0.1) and the closed convex cones from *linear matrix inequality example* 2.13.5.1.1; given convex cone $\mathcal{K}$ and its dual

$$\mathcal{K}=\{A\operatorname{svec}X\mid X\succeq 0\} \qquad (324)$$

$$\mathcal{K}^*=\{y\mid \sum_{j=1}^m y_jA_j\succeq 0\} \qquad (330)$$

where

$$A=\begin{bmatrix}\operatorname{svec}(A_1)^T\\ \vdots\\ \operatorname{svec}(A_m)^T\end{bmatrix}\in\mathbb{R}^{m\times n(n+1)/2} \qquad (547)$$

then we have membership relation

$$b\in\mathcal{K}\;\Leftrightarrow\;\langle y\,,b\rangle\geq 0\;\;\forall\,y\in\mathcal{K}^* \qquad (276)$$

and equivalents

$$b\in\mathcal{K}\quad\Leftrightarrow\quad\exists\,X\succeq 0\;\ni\;A\operatorname{svec}X=b\quad\Leftrightarrow\quad\mathcal{A}\cap\mathbb{S}_+^n\neq\emptyset \qquad (559)$$

$$b\in\mathcal{K}\quad\Leftrightarrow\quad\langle y\,,b\rangle\geq 0\;\;\forall\,y\in\mathcal{K}^*\quad\Leftrightarrow\quad\mathcal{A}\cap\mathbb{S}_+^n\neq\emptyset \qquad (560)$$

*Semidefinite Farkas' lemma* provides the conditions required for a set of hyperplanes to have a nonempty intersection $\mathcal{A}\cap\mathbb{S}_+^n$ with the positive semidefinite cone. While the lemma as stated is correct, Ye points out [299, §1.3.8] that a positive definite version of this lemma is required for semidefinite programming because any feasible point in the relative interior $\mathcal{A}\cap\operatorname{int}\mathbb{S}_+^n$ is required by *Slater's condition*[4.6] to achieve 0 duality gap (primal−dual objective difference §4.2.3, Figure **45**). In our circumstance, assuming a nonempty intersection, a positive definite lemma is required to insure a point of intersection closest to the origin is not at infinity;

[4.6]Slater's sufficient condition is satisfied whenever any primal *strictly feasible* point exists; *id est*, any point feasible with the affine equality (or affine inequality) constraint functions and relatively interior to convex cone $\mathcal{K}$. If cone $\mathcal{K}$ is polyhedral, then Slater's condition is satisfied when any feasible point exists relatively interior to $\mathcal{K}$ or on its relative boundary. [46, §5.2.3] [29, p.325] *e.g.*, Figure **33**. Then given $A \in \mathbb{R}^{m\times n(n+1)/2}$ having rank $m$, we wish to detect existence of a nonempty relative interior of the primal feasible set;[4.7]

$$b \in \operatorname{int}\mathcal{K} \quad \Leftrightarrow \quad \langle y\,, b\rangle > 0 \;\; \forall\, y \in \mathcal{K}^*,\;\; y \neq \mathbf{0} \quad \Leftrightarrow \quad \mathcal{A} \cap \operatorname{int}\mathbb{S}^n_+ \neq \emptyset \tag{561}$$

A positive definite Farkas' lemma can easily be constructed from this membership relation (282) and these proper convex cones $\mathcal{K}$ (324) and $\mathcal{K}^*$ (330):

**4.2.1.1.2 Lemma.** *Positive definite Farkas' lemma.*
Given a linearly independent set $\{A_i \in \mathbb{S}^n,\; i=1\ldots m\}$ and a vector $b=[b_i]\in\mathbb{R}^m$, define the affine subset

$$\mathcal{A} = \{X \in \mathbb{S}^n \mid \langle A_i\;,\; X\rangle = b_i\;,\;\; i=1\ldots m\} \tag{549}$$

Primal feasible set relative interior $\mathcal{A}\cap\operatorname{int}\mathbb{S}^n_+$ is nonempty if and only if $y^Tb>0$ holds for each and every vector $y=[y_i]\neq\mathbf{0}$ such that $\sum\limits_{i=1}^m y_iA_i \succeq 0$.

Equivalently, primal feasible set relative interior $\mathcal{A}\cap\operatorname{int}\mathbb{S}^n_+$ is nonempty if and only if $y^Tb>0$ holds for each and every norm-1 vector $\|y\|=1$ such that $\sum\limits_{i=1}^m y_iA_i \succeq 0$. ⋄

**4.2.1.1.3 Example.** *"New" Farkas' lemma.*
In 1995, Lasserre [167, §III] presented an example originally offered by Ben-Israel in 1969 [26, p.378] as evidence of failure in *semidefinite Farkas' Lemma* 4.2.1.1.1:

$$A \triangleq \begin{bmatrix} \operatorname{svec}(A_1)^T \\ \operatorname{svec}(A_2)^T \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{562}$$

The intersection $\mathcal{A}\cap\mathbb{S}^n_+$ is practically empty because the solution set

$$\{X \succeq 0 \mid A\operatorname{svec}X = b\} \;=\; \left\{ \begin{bmatrix} \alpha & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \succeq 0 \mid \alpha \in \mathbb{R} \right\} \tag{563}$$

[4.7] Detection of $\mathcal{A}\cap\operatorname{int}\mathbb{S}^n_+$ by examining $\mathcal{K}$ interior is a trick need not be lost.

is positive semidefinite only asymptotically $(\alpha\to\infty)$. Yet the dual system $\sum_{i=1}^m y_i A_i \succeq 0 \Rightarrow y^T b \geq 0$ indicates nonempty intersection; *videlicet*, for $\|y\|=1$

$$y_1\begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + y_2\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \succeq 0 \quad\Leftrightarrow\quad y = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad\Rightarrow\quad y^T b = 0 \tag{564}$$

On the other hand, *positive definite Farkas' Lemma* 4.2.1.1.2 shows $\mathcal{A}\cap \operatorname{int}\mathbb{S}_+^n$ is empty; what we need to know for semidefinite programming.

Based on Ben-Israel's example, Lasserre suggested addition of another condition to *semidefinite Farkas' Lemma* 4.2.1.1.1 to make a "new" lemma. Ye recommends *positive definite Farkas' Lemma* 4.2.1.1.2 instead; which is simpler and obviates Lasserre's proposed additional condition. □

#### 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*, for example, we get

$$\begin{gathered} \mathcal{A}\cap \operatorname{int}\mathbb{S}_+^n \neq \emptyset \\ \text{or in the alternative} \\ y^T b \leq 0\,,\quad \sum_{i=1}^m y_i A_i \succeq 0\,,\quad y \neq \mathbf{0} \end{gathered} \tag{565}$$

Any single vector $y$ satisfying the alternative certifies $\mathcal{A}\cap \operatorname{int}\mathbb{S}_+^n$ is empty. Such a vector can be found as a solution to another semidefinite program: for linearly independent set $\{A_i\in\mathbb{S}^n,\ i=1\ldots m\}$

$$\begin{array}{cl} \underset{y}{\text{minimize}} & y^T b \\ \text{subject to} & \sum_{i=1}^m y_i A_i \succeq 0 \\ & \|y\|^2 \leq 1 \end{array} \tag{566}$$

If an optimal vector $y^\star \neq \mathbf{0}$ can be found such that $y^{\star T} b \leq 0$, then relative interior of the primal feasible set $\mathcal{A}\cap \operatorname{int}\mathbb{S}_+^n$ from (558) is empty.

#### 4.2.1.3 Boundary-membership criterion

(*confer* (560) (561)) From boundary-membership relation (286) for proper cones and from *linear matrix inequality* cones $\mathcal{K}$ (324) and $\mathcal{K}^*$ (330)

$$b \in \partial\mathcal{K} \quad \Leftrightarrow \quad \exists\, y \ni \langle y\,,\,b\rangle = 0\,,\;\; y \in \mathcal{K}^*,\;\; y \neq \mathbf{0}\,,\;\; b \in \mathcal{K} \quad \Leftrightarrow \quad \partial\mathbb{S}^n_+ \supset \mathcal{A} \cap \mathbb{S}^n_+ \neq \emptyset \tag{567}$$

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: assuming $b \in \mathcal{K}$ (559) given linearly independent set $\{A_i \in \mathbb{S}^n,\; i=1\ldots m\}$ [4.8]

$$\begin{array}{rl} \text{find} & y \neq \mathbf{0} \\ \text{subject to} & y^T b = 0 \\ & \sum\limits_{i=1}^m y_i A_i \succeq 0 \end{array} \tag{568}$$

Any such feasible vector $y \neq \mathbf{0}$ certifies that affine subset $\mathcal{A}$ (549) intersects the positive semidefinite cone $\mathbb{S}^n_+$ only on its boundary; in other words, nonempty feasible set $\mathcal{A} \cap \mathbb{S}^n_+$ belongs to the positive semidefinite cone boundary $\partial\mathbb{S}^n_+$ .

### 4.2.2 Duals

The dual objective function evaluated at any feasible point represents a lower bound on the primal optimal objective value. We can see this by direct substitution: Assume the feasible sets $\mathcal{A} \cap \mathbb{S}^n_+$ and $\mathcal{C}^*$ are nonempty. Then it is always true:

$$\begin{aligned} \langle C\;,\;X\rangle &\geq \langle b\,,\,y\rangle \\ \left\langle \sum_i y_i A_i + S\;,\;X\right\rangle &\geq [\,\langle A_1\;,\;X\rangle \;\cdots\; \langle A_m\;,\;X\rangle\,]\;y \\ \langle S\;,\;X\rangle &\geq 0 \end{aligned} \tag{569}$$

The converse also follows because

$$X \succeq 0\,,\;\; S \succeq 0 \quad \Rightarrow \quad \langle S\,,X\rangle \geq 0 \tag{1285}$$

[4.8] 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 $X$ .

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, [299, §1.3.8] and can be used to detect convergence in any primal/dual numerical method of solution.

### 4.2.3 Optimality conditions

When any primal feasible point exists relatively interior to $\mathcal{A} \cap \mathbb{S}^n_+$ in $\mathbb{S}^n$, or when any dual feasible point exists relatively interior to $\mathcal{C}^*$ in $\mathbb{S}^n \times \mathbb{R}^m$, then by Slater's sufficient condition these two problems (546P) and (546D) become *strong duals.* In other words, the primal optimal objective value becomes equivalent to the dual optimal objective value: there is no duality gap (Figure **45**); *id est,* if $\exists\, X \in \mathcal{A} \cap \operatorname{int} \mathbb{S}^n_+$ or $\exists\, S, y \in \operatorname{rel\,int} \mathcal{C}^*$ then

$$\begin{aligned} \langle C\,,\, X^\star\rangle &= \langle b\,, y^\star\rangle \\ \left\langle \sum_i y_i^\star A_i + S^\star\,,\, X^\star\right\rangle &= [\,\langle A_1\,,\, X^\star\rangle \;\cdots\; \langle A_m\,,\, X^\star\rangle\,]\; y^\star \\ \langle S^\star,\, X^\star\rangle &= 0 \end{aligned} \tag{570}$$

where $S^\star$, $y^\star$ denote a dual optimal solution.[4.9] We summarize this:

**4.2.3.0.1 Corollary.** *Optimality and strong duality.* [269, §3.1] [299, §1.3.8] For semidefinite programs (546P) and (546D), assume primal and dual feasible sets $\mathcal{A} \cap \mathbb{S}^n_+ \subset \mathbb{S}^n$ and $\mathcal{C}^* \subset \mathbb{S}^n \times \mathbb{R}^m$ (558) are nonempty. Then

- $X^\star$ is optimal for (P)
- $S^\star$, $y^\star$ are optimal for (D)
- the duality gap $\langle C, X^\star\rangle - \langle b\,, y^\star\rangle$ is $0$

if and only if

**i)** $\exists\, X \in \mathcal{A} \cap \operatorname{int} \mathbb{S}^n_+$ **or** $\exists\, S\,,\ y \in \operatorname{rel\,int} \mathcal{C}^*$

**and**

**ii)** $\langle S^\star,\, X^\star\rangle = 0$ ⋄

[4.9] Optimality condition $\langle S^\star,\, X^\star\rangle = 0$ is called a *complementary slackness condition*, in keeping with the tradition of linear programming, [64] that forbids dual inequalities in (546) to simultaneously hold strictly. [229, §4]

For symmetric positive semidefinite matrices, requirement **ii** is equivalent to the *complementarity* (§A.7.4)

$$\langle S^\star,\, X^\star\rangle = 0 \quad\Leftrightarrow\quad S^\star X^\star = X^\star S^\star = \mathbf{0} \tag{571}$$

Commutativity of diagonalizable matrices is a necessary and sufficient condition [150, §1.3.12] for these two optimal symmetric matrices to be simultaneously diagonalizable. Therefore

$$\operatorname{rank} X^\star + \operatorname{rank} S^\star \leq\; n \tag{572}$$

**Proof.** To see that, the product of symmetric optimal matrices $X^\star, S^\star\in\mathbb{S}^n$ must itself be symmetric because of commutativity. (1279) The symmetric product has diagonalization [9, cor.2.11]

$$S^\star X^\star = X^\star S^\star = Q\Lambda_{S^\star}\Lambda_{X^\star}Q^T = \mathbf{0} \quad\Leftrightarrow\quad \Lambda_{X^\star}\Lambda_{S^\star} = \mathbf{0} \tag{573}$$

where $Q$ is an orthogonal matrix. The product of the nonnegative diagonal $\Lambda$ matrices can be $\mathbf{0}$ if their main diagonal zeros are complementary or coincide. Due only to symmetry, $\operatorname{rank} X^\star = \operatorname{rank}\Lambda_{X^\star}$ and $\operatorname{rank} S^\star = \operatorname{rank}\Lambda_{S^\star}$ for these optimal primal and dual solutions. (1264) So, because of the complementarity, the total number of nonzero diagonal entries from both $\Lambda$ cannot exceed $n$. ♦

When equality is attained in (572)

$$\operatorname{rank} X^\star + \operatorname{rank} S^\star =\; n \tag{574}$$

there are no coinciding main diagonal zeros in $\Lambda_{X^\star}\Lambda_{S^\star}$, and so we have what is called *strict complementarity*.[4.10] Logically it follows that a necessary and sufficient condition for strict complementarity of an optimal primal and dual solution is

$$X^\star + S^\star \succ 0 \tag{575}$$

The beauty of Corollary 4.2.3.0.1 is its conjugacy; *id est*, one can solve either the primal or dual problem 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 belongs to the hyperplane $\{X \mid \langle S^\star,\, X\rangle = 0\}$.

[4.10] distinct from maximal complementarity (§4.1.1).

**4.2.3.0.2 Example.** *Minimum cardinality Boolean.* [63] [27, §4.3.4] [260] (*confer* Example 4.4.3.0.1) Consider finding a *minimum cardinality* Boolean solution $x$ to the classic linear algebra problem $Ax=b$ given noiseless data $A\in\mathbb{R}^{m\times n}$ and $b\in\mathbb{R}^m$;

$$\begin{array}{cl}\underset{x}{\text{minimize}} & \|x\|_0\\ \text{subject to} & Ax=b\\ & x_i\in\{0,1\}\;, \quad i=1\ldots n\end{array} \tag{576}$$

where $\|x\|_0$ denotes cardinality of vector $x$ (`a.k.a`, 0-*norm*; not a convex function).

A minimum 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. [238] [157] [120] [121] 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 (576) instead the 1-norm or 2-norm; *id est*, the two problems

$$(576)\quad \begin{array}{cl}\underset{x}{\text{minimize}} & \|x\|_0\\ \text{subject to} & Ax=b\\ & x_i\in\{0,1\}\;, \quad i=1\ldots n\end{array} \quad=\quad \begin{array}{cl}\underset{x}{\text{minimize}} & \|x\|_1\\ \text{subject to} & Ax=b\\ & x_i\in\{0,1\}\;, \quad i=1\ldots n\end{array} \tag{577}$$

are the same. The Boolean constraint makes the 1-norm problem nonconvex.

Given data[4.11]

$$A=\begin{bmatrix}-1 & 1 & 8 & 1 & 1 & 0\\ -3 & 2 & 8 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2}-\frac{1}{3}\\ -9 & 4 & 8 & \frac{1}{4} & \frac{1}{9} & \frac{1}{4}-\frac{1}{9}\end{bmatrix}, \qquad b=\begin{bmatrix}1\\ \frac{1}{2}\\ \frac{1}{4}\end{bmatrix} \tag{578}$$

the obvious and desired solution to the problem posed,

$$x^\star=e_4\in\mathbb{R}^{\mathbf{6}} \tag{579}$$

has norm $\|x^\star\|_2=1$ and minimum cardinality; the minimum number of nonzero entries in vector $x$. The MATLAB backslash command `x=A\b` ,

[4.11] This particular matrix $A$ is full-rank having three-dimensional nullspace (but the columns are not conically independent).

for example, finds

$$x_{\mathrm{M}} = \begin{bmatrix} \frac{2}{128} \\ 0 \\ \frac{5}{128} \\ 0 \\ \frac{90}{128} \\ 0 \end{bmatrix} \tag{580}$$

having norm $\|x_{\mathrm{M}}\|_2 = 0.7044$. Coincidentally, $x_{\mathrm{M}}$ is a 1-norm solution; *id est*, an optimal solution to

$$\begin{array}{cl} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array} \tag{581}$$

The pseudoinverse solution (rounded)

$$x_{\mathrm{P}} = A^{\dagger} b = \begin{bmatrix} -0.0456 \\ -0.1881 \\ 0.0623 \\ 0.2668 \\ 0.3770 \\ -0.1102 \end{bmatrix} \tag{582}$$

has least norm $\|x_{\mathrm{P}}\|_2 = 0.5165$; *id est*, the optimal solution to

$$\begin{array}{cl} \underset{x}{\text{minimize}} & \|x\|_2 \\ \text{subject to} & Ax = b \end{array} \tag{583}$$

Certainly, none of the traditional methods provide $x^\star = e_4$ (579).

We can reformulate this minimum cardinality Boolean problem (576) as a semidefinite program: First transform the variable

$$x \triangleq (\hat{x} + \mathbf{1})\tfrac{1}{2} \tag{584}$$

so $\hat{x}_i \in \{-1, 1\}$; equivalently,

$$\begin{array}{cl} \underset{\hat{x}}{\text{minimize}} & \|(\hat{x} + \mathbf{1})\frac{1}{2}\|_0 \\ \text{subject to} & A(\hat{x} + \mathbf{1})\frac{1}{2} = b \\ & \delta(\hat{x}\hat{x}^T) = \mathbf{1} \end{array} \tag{585}$$

where $\delta$ is the main-diagonal linear operator (§A.1). By assigning (§B.1)

$$G = \begin{bmatrix} \hat{x} \\ 1 \end{bmatrix} [\,\hat{x}^T \;\; 1\,] = \begin{bmatrix} X & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \triangleq \begin{bmatrix} \hat{x}\hat{x}^T & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \in \mathbb{S}^{n+1} \tag{586}$$

problem (585) becomes equivalent to: (Theorem A.3.1.0.7)

$$\begin{array}{cl} \underset{X\in\,\mathbb{S}^n,\; \hat{x}\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T\hat{x} \\ \text{subject to} & A(\hat{x}+\mathbf{1})\frac{1}{2} = b \\ & G = \begin{bmatrix} X & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \\ & \delta(X) = \mathbf{1} \\ & (G \succeq 0) \\ & \operatorname{rank} G = 1 \end{array} \tag{587}$$

where solution is confined to rank-1 vertices of the elliptope in $\mathbb{S}^{n+1}$ (§5.9.1.0.1) by the rank constraint, the positive semidefiniteness, and the equality constraints $\delta(X)=\mathbf{1}$. The rank constraint makes this problem nonconvex; by removing it[4.12] we get the semidefinite program

$$\begin{array}{cl} \underset{X\in\,\mathbb{S}^n,\; \hat{x}\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T\hat{x} \\ \text{subject to} & A(\hat{x}+\mathbf{1})\frac{1}{2} = b \\ & G = \begin{bmatrix} X & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \succeq 0 \\ & \delta(X) = \mathbf{1} \end{array} \tag{588}$$

whose optimal solution $x^\star$ (584) is identical to that of minimum cardinality Boolean problem (576) if and only if $\operatorname{rank} G^\star = 1$. Hope[4.13] of acquiring a rank-1 solution is not ill-founded because $2^n$ elliptope vertices have rank 1, and we are minimizing an affine function on a subset of the elliptope (Figure **87**) containing rank-1 vertices; *id est*, by assumption that the feasible set of minimum cardinality Boolean problem (576) is nonempty,

[4.12] Relaxed problem (588) can also be derived via Lagrange duality; it is a dual of a dual program [*sic*] to (587). [227] [46, §5, exer.5.39] [285, §IV] [101, §11.3.4] The relaxed problem must therefore be convex having a larger feasible set; its optimal objective value represents a generally *loose* lower bound (1462) on the optimal objective of problem (587).

[4.13] A more deterministic approach to constraining rank and cardinality is developed in §4.4.3.0.8.

a desired solution resides on the elliptope relative boundary at a rank-1 vertex.[4.14]

For the data given in (578), our semidefinite program solver (accurate to approximately `1E-8`)[4.15] finds optimal solution to (588)

$$\operatorname{round}(G^\star) = \begin{bmatrix} 1 & 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 & -1 & -1 & 1 \end{bmatrix} \tag{589}$$

near a rank-1 vertex of the elliptope in $\mathbb{S}^{n+1}$; its sorted eigenvalues,

$$\lambda(G^\star) = \begin{bmatrix} 6.99999977799099 \\ 0.00000022687241 \\ 0.00000002250296 \\ 0.00000000262974 \\ -0.00000000999738 \\ -0.00000000999875 \\ -0.00000001000000 \end{bmatrix} \tag{590}$$

The 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 minimum cardinality Boolean solution by truncating all smaller eigenvalues. By so doing we find, indeed,

$$x^\star = \operatorname{round}\left( \begin{bmatrix} 0.00000000127947 \\ 0.00000000527369 \\ 0.00000000181001 \\ 0.99999997469044 \\ 0.00000001408950 \\ 0.00000000482903 \end{bmatrix} \right) = e_4 \tag{591}$$

the desired result (579). □

---

[4.14] Confinement to the elliptope can be regarded as a kind of normalization akin to matrix $A$ column normalization suggested in [82] and explored in Example 4.2.3.0.3.

[4.15] A typically ignored limitation of interior-point methods of solution is their relative accuracy of only about `1E-8` on a machine using 64-bit (*double precision*) floating-point arithmetic; *id est*, optimal solution cannot be more accurate than square root of machine epsilon (`2.2204E-16`).

**4.2.3.0.3 Example.** *Optimization on elliptope* versus 1-*norm polyhedron for minimum cardinality Boolean Example 4.2.3.0.2.* A minimum cardinality problem is typically formulated via, what is by now, the standard practice [82] of column normalization applied to a 1-norm problem surrogate like (581). Suppose we define a diagonal matrix

$$\Lambda \triangleq \begin{bmatrix} \|A(:,1)\|_2 & & & \mathbf{0} \\ & \|A(:,2)\|_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \|A(:,6)\|_2 \end{bmatrix} \in \mathbb{S}^6 \tag{592}$$

used to normalize the columns (assumed nonzero) of given noiseless data matrix $A$. Then approximate the minimum cardinality Boolean problem

$$\begin{array}{cl} \underset{x}{\text{minimize}} & \|x\|_0 \\ \text{subject to} & Ax = b \\ & x_i \in \{0,1\}\,, \quad i=1\ldots n \end{array} \tag{576}$$

as

$$\begin{array}{cl} \underset{\tilde{y}}{\text{minimize}} & \|\tilde{y}\|_1 \\ \text{subject to} & A\Lambda^{-1}\tilde{y} = b \\ & \mathbf{1} \succeq \Lambda^{-1}\tilde{y} \succeq 0 \end{array} \tag{593}$$

where optimal solution

$$y^\star = \text{round}(\Lambda^{-1}\tilde{y}^\star) \tag{594}$$

The inequality in (593) relaxes Boolean constraint $y_i \in \{0,1\}$ from (576); serving to bound any solution $y^\star$ to a unit cube whose vertices are binary numbers. Convex problem (593) is justified by the *convex envelope*

$$\text{cenv}\,\|x\|_0 \ \text{ on } \ \{x\in\mathbb{R}^n \mid \|x\|_\infty \leq \kappa\} = \frac{1}{\kappa}\|x\|_1 \tag{1171}$$

Donoho concurs with this particular formulation equivalently expressible as a linear program via (430).

Approximation (593) is therefore equivalent to minimization of an affine function on a bounded polyhedron, whereas semidefinite program

$$\begin{array}{cl} \underset{X\in\mathbb{S}^n,\ \hat{x}\in\mathbb{R}^n}{\text{minimize}} & \mathbf{1}^T\hat{x} \\ \text{subject to} & A(\hat{x}+\mathbf{1})\frac{1}{2} = b \\ & G = \begin{bmatrix} X & \hat{x} \\ \hat{x}^T & 1 \end{bmatrix} \succeq 0 \\ & \delta(X) = \mathbf{1} \end{array} \tag{588}$$

minimizes an affine function on the elliptope intersected by hyperplanes. Although the same Boolean solution is obtained from this approximation (593) as compared with semidefinite program (588) when given that particular data from Example 4.2.3.0.2, Singer confides a counter-example: Instead, given data

$$A = \begin{bmatrix} 1 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{595}$$

then solving approximation (593) yields

$$y^\star = \operatorname{round}\left(\begin{bmatrix} 1 - \frac{1}{\sqrt{2}} \\ 1 - \frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{596}$$

(infeasible, with or without rounding, with respect to original problem (576)) whereas solving semidefinite program (588) produces

$$\operatorname{round}(G^\star) = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix} \tag{597}$$

with sorted eigenvalues

$$\lambda(G^\star) = \begin{bmatrix} 3.99999965057264 \\ 0.00000035942736 \\ -0.00000000000000 \\ -0.00000001000000 \end{bmatrix} \tag{598}$$

Truncating all but the largest eigenvalue, we obtain (*confer* $y^\star$) (584)

$$x^\star = \operatorname{round}\left(\begin{bmatrix} 0.99999999625299 \\ 0.99999999625299 \\ 0.00000001434518 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \tag{599}$$

the desired minimum cardinality Boolean result.

We leave pending a general performance assessment of standard-practice approximation (593) as compared with our proposed semidefinite program (588). □

# 4.3 Rank reduction

> *. . . it is not clear generally how to predict* rank $X^\star$ *or* rank $S^\star$ *before solving the SDP problem.*
>
> −Farid Alizadeh (1995) [9, p.22]

The premise of rank reduction in semidefinite programming is: an optimal solution found does not satisfy Barvinok's upper bound (232) on rank. The particular numerical algorithm solving a semidefinite program may have instead returned a high-rank optimal solution (§4.1.1; *e.g.*, (557)) when a lower-rank optimal solution was expected.

## 4.3.1 Posit a perturbation of $X^\star$

Recall from §4.1.1.1, there is an extreme point of $\mathcal{A} \cap \mathbb{S}^n_+$ (549) satisfying upper bound (232) on rank. [21, §2.2] It is therefore sufficient to locate an extreme point of the intersection whose primal objective value (546P) is optimal:[4.16] [77, §31.5.3] [175, §2.4] [5, §3] [215]

Consider again the affine subset

$$\mathcal{A} = \{X \in \mathbb{S}^n \mid A \operatorname{svec} X = b\} \tag{549}$$

where for $A_i \in \mathbb{S}^n$

$$A \triangleq \begin{bmatrix} \operatorname{svec}(A_1)^T \\ \vdots \\ \operatorname{svec}(A_m)^T \end{bmatrix} \in \mathbb{R}^{m \times n(n+1)/2} \tag{547}$$

Given any optimal solution $X^\star$ to

$$\begin{array}{cl} \underset{X \in \mathbb{S}^n}{\text{minimize}} & \langle C\,,\,X\rangle \\ \text{subject to} & X \in \mathcal{A} \cap \mathbb{S}^n_+ \end{array} \qquad (546)(\text{P})$$

[4.16] *There is no known construction for Barvinok's tighter result* (237). −Monique Laurent whose rank does not satisfy upper bound (232), we posit existence of a set of perturbations

$$\{t_j B_j \mid t_j \in \mathbb{R} \,,\; B_j \in \mathbb{S}^n,\; j=1\ldots n\} \tag{600}$$

such that, for some $0 \leq i \leq n$ and scalars $\{t_j \,,\; j=1\ldots i\}$,

$$X^\star + \sum_{j=1}^{i} t_j B_j \tag{601}$$

becomes an extreme point of $\mathcal{A} \cap \mathbb{S}_+^n$ and remains an optimal solution of (546P). Membership of (601) to affine subset $\mathcal{A}$ is secured for the $i^{\text{th}}$ perturbation by demanding

$$\langle B_i \,,\, A_j \rangle = 0 \,, \quad j=1\ldots m \tag{602}$$

while membership to the positive semidefinite cone $\mathbb{S}_+^n$ is insured by small perturbation (611). In this manner feasibility is insured. 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.* (§F.4)

`initialize:` $B_i = \mathbf{0} \;\; \forall i$

`for` iteration `i=1...n`

{

1. `compute` a nonzero perturbation matrix $B_i$ of $X^\star + \sum_{j=1}^{i-1} t_j B_j$

2. `maximize` $t_i$

   `subject to` $X^\star + \sum_{j=1}^{i} t_j B_j \in \mathbb{S}_+^n$

} ¶

A rank-reduced optimal solution is then

$$X^\star \leftarrow X^\star + \sum_{j=1}^{i} t_j B_j \tag{603}$$

### 4.3.2 Perturbation form

The perturbations are independent of constants $C \in \mathbb{S}^n$ and $b \in \mathbb{R}^m$ in primal and dual programs (546). Numerical accuracy of any rank-reduced result, found by perturbation of an initial optimal solution $X^\star$, is therefore quite dependent upon initial accuracy of $X^\star$.

**4.3.2.0.1 Definition.** *Matrix step function.* (*confer* §A.6.5.0.1)
Define the signum-like quasiconcave real function $\psi : \mathbb{S}^n \to \mathbb{R}$

$$\psi(Z) \triangleq \begin{cases} 1\,, & Z \succeq 0 \\ -1\,, & \text{otherwise} \end{cases} \tag{604}$$

The value $-1$ is taken for indefinite or nonzero negative semidefinite argument. $\triangle$

Deza & Laurent [77, §31.5.3] prove: every perturbation matrix $B_i$ , $i=1 \ldots n$ , is of the form

$$B_i = -\psi(Z_i) R_i Z_i R_i^T \in \mathbb{S}^n \tag{605}$$

where

$$X^\star \triangleq R_1 R_1^T\,, \qquad X^\star + \sum_{j=1}^{i-1} t_j B_j \triangleq R_i R_i^T \in \mathbb{S}^n \tag{606}$$

where the $t_j$ are scalars and $R_i \in \mathbb{R}^{n\times\rho}$ is full-rank and skinny where

$$\rho \triangleq \operatorname{rank}\left(X^\star + \sum_{j=1}^{i-1} t_j B_j\right) \tag{607}$$

and where matrix $Z_i \in \mathbb{S}^\rho$ is found at each iteration $i$ by solving a very simple feasibility problem: [4.17]

$$\begin{array}{rl}\text{find} & Z_i \in \mathbb{S}^\rho \\ \text{subject to} & \langle Z_i\,,\; R_i^T A_j R_i\rangle = 0\;, \qquad j = 1 \ldots m\end{array} \tag{608}$$

Were there a sparsity pattern common to each member of the set $\{R_i^T A_j R_i \in \mathbb{S}^\rho,\; j = 1 \ldots m\}$, 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 $i$

$$X^\star + \sum_{j=1}^{i-1} t_j B_j \;+\; t_i B_i \;=\; R_i\big(I - t_i\,\psi(Z_i) Z_i\big) R_i^T \tag{609}$$

By fact (1253), therefore

$$X^\star + \sum_{j=1}^{i-1} t_j B_j \;+\; t_i B_i \succeq 0 \;\Leftrightarrow\; \mathbf{1} - t_i\,\psi(Z_i)\lambda(Z_i) \succeq 0 \tag{610}$$

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 (609) and locates a new point on the boundary $\partial(\mathcal{A} \cap \mathbb{S}_+^n)$ .[4.18] Maximization of $t_i$ thereby has closed form;

---

[4.17] 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 (546P) while rank $\rho$ decreases with each iteration $i$ . Otherwise, we might iterate indefinitely.

[4.18] This holds because rank of a positive semidefinite matrix in $\mathbb{S}^n$ is diminished below $n$ by the number of its 0 eigenvalues (1264), and because a positive semidefinite matrix having one or more 0 eigenvalues corresponds to a point on the PSD cone boundary (162). 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 (609).

$$(t_i^\star)^{-1} = \max\left\{\psi(Z_i)\lambda(Z_i)_j\ ,\ j=1\ldots\rho\right\} \tag{611}$$

When $Z_i$ is indefinite, the direction of perturbation (determined by $\psi(Z_i)$) is arbitrary. We may take an early exit from the Procedure were $Z_i$ to become $\mathbf{0}$ or were

$$\operatorname{rank}\left[\operatorname{svec} R_i^T A_1 R_i \quad \operatorname{svec} R_i^T A_2 R_i \quad \cdots \quad \operatorname{svec} R_i^T A_m R_i\right] = \rho(\rho+1)/2 \tag{612}$$

which characterizes the rank $\rho$ of any [*sic*] extreme point in $\mathcal{A}\cap\mathbb{S}_+^n$. [175, §2.4]

**Proof.** Assuming the form of every perturbation matrix is indeed (605), then by (608)

$$\operatorname{svec} Z_i \perp \left[\operatorname{svec}(R_i^T A_1 R_i) \quad \operatorname{svec}(R_i^T A_2 R_i) \quad \cdots \quad \operatorname{svec}(R_i^T A_m R_i)\right] \tag{613}$$

By orthogonal complement we have

$$\begin{aligned} &\operatorname{rank}\left[\operatorname{svec}(R_i^T A_1 R_i) \ \cdots \ \operatorname{svec}(R_i^T A_m R_i)\right]^\perp \\ +\ &\operatorname{rank}\left[\operatorname{svec}(R_i^T A_1 R_i) \ \cdots \ \operatorname{svec}(R_i^T A_m R_i)\right] = \rho(\rho+1)/2 \end{aligned} \tag{614}$$

When $Z_i$ can only be $\mathbf{0}$, then the perturbation is null because an extreme point has been found; thus

$$\left[\operatorname{svec}(R_i^T A_1 R_i) \ \cdots \ \operatorname{svec}(R_i^T A_m R_i)\right]^\perp = \mathbf{0} \tag{615}$$

from which the stated result (612) directly follows. ♦

### 4.3.3 Optimality of perturbed $X^\star$

We show that the optimal objective value is unaltered by perturbation (605); *id est*,

$$\langle C\,,\;X^\star + \sum_{j=1}^{i} t_j B_j\rangle \;=\; \langle C\,,\;X^\star\rangle \tag{616}$$

**Proof.** From Corollary 4.2.3.0.1 we have the necessary and sufficient relationship between optimal primal and dual solutions under the assumption of existence of a relatively interior feasible point:

$$S^\star X^\star = S^\star R_1 R_1^T \;=\; X^\star S^\star \;=\; R_1 R_1^T S^\star \;=\; \mathbf{0} \tag{617}$$

This means $\mathcal{R}(R_1) \subseteq \mathcal{N}(S^\star)$ and $\mathcal{R}(S^\star) \subseteq \mathcal{N}(R_1^T)$. From (606) and (609) we get the sequence:

$$\begin{aligned}
&X^\star = R_1 R_1^T\\
&X^\star + t_1 B_1 = R_2 R_2^T = R_1\big(I - t_1\psi(Z_1)Z_1\big)R_1^T\\
&X^\star + t_1 B_1 + t_2 B_2 = R_3 R_3^T = R_2\big(I - t_2\psi(Z_2)Z_2\big)R_2^T = R_1\big(I - t_1\psi(Z_1)Z_1\big)\big(I - t_2\psi(Z_2)Z_2\big)R_1^T\\
&\vdots\\
&X^\star + \sum_{j=1}^{i} t_j B_j = R_1\left(\prod_{j=1}^{i}\big(I - t_j\psi(Z_j)Z_j\big)\right)R_1^T
\end{aligned} \tag{618}$$

Substituting $C = \text{svec}^{-1}(A^T y^\star) + S^\star$ from (546),

$$\begin{aligned}
\langle C\,,\;X^\star + \sum_{j=1}^{i} t_j B_j\rangle &= \left\langle \text{svec}^{-1}(A^T y^\star) + S^\star\,,\;R_1\left(\prod_{j=1}^{i}\big(I - t_j\psi(Z_j)Z_j\big)\right)R_1^T\right\rangle\\
&= \left\langle \sum_{k=1}^{m} y_k^\star A_k\,,\;X^\star + \sum_{j=1}^{i} t_j B_j\right\rangle\\
&= \left\langle \sum_{k=1}^{m} y_k^\star A_k + S^\star\,,\;X^\star\right\rangle \;=\; \langle C\,,\;X^\star\rangle
\end{aligned} \tag{619}$$

because $\langle B_i\,,A_j\rangle = 0 \;\; \forall\, i\,,j$ by design (602). ♦

**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 (232): Assume a given vector $b \in \mathbb{R}^m$ belongs to the conic hull of the columns of a given matrix $A \in \mathbb{R}^{m\times n}$;

$$A = \begin{bmatrix} -1 & 1 & 8 & 1 & 1 \\ -3 & 2 & 8 & \frac{1}{2} & \frac{1}{3} \\ -9 & 4 & 8 & \frac{1}{4} & \frac{1}{9} \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \frac{1}{4} \end{bmatrix} \tag{620}$$

Consider the convex optimization problem

$$\begin{array}{cl} \underset{X\in\mathbb{S}^5}{\text{minimize}} & \operatorname{tr} X \\ \text{subject to} & X \succeq 0 \\ & A\delta(X) = b \end{array} \tag{621}$$

that minimizes the 1-norm of the main diagonal; *id est*, problem (621) is the same as

$$\begin{array}{cl} \underset{X\in\mathbb{S}^5}{\text{minimize}} & \|\delta(X)\|_1 \\ \text{subject to} & X \succeq 0 \\ & A\delta(X) = b \end{array} \tag{622}$$

that finds a solution to $A\delta(X) = b$. Rank-3 solution $X^\star = \delta(x_{\mathrm{M}})$ is optimal, where (*confer* (580))

$$x_{\mathrm{M}} = \begin{bmatrix} \frac{2}{128} \\ 0 \\ \frac{5}{128} \\ 0 \\ \frac{90}{128} \end{bmatrix} \tag{623}$$

Yet upper bound (232) predicts existence of at most a

$$\text{rank-}\left(\left\lfloor \frac{\sqrt{8m+1}-1}{2} \right\rfloor = 2\right) \tag{624}$$

feasible solution from $m = 3$ equality constraints. To find a lower rank $\rho$ optimal solution to (621) (barring combinatorics), we invoke Procedure 4.3.1.0.1:

`Initialize:`

$C = I\,,\;\; \rho = 3,\;\; A_j \stackrel{\Delta}{=} \delta(A(j,:))\,,\; j = 1,2,3,\;\; X^\star = \delta(x_{\rm M})\,,\;\; m = 3,\;\; n = 5.$

{

Iteration `i=1`:

**`Step 1:`** $R_1 = \begin{bmatrix} \sqrt{\frac{2}{128}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sqrt{\frac{5}{128}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{\frac{90}{128}} \end{bmatrix}.$

$$\begin{array}{rl} \text{find} & Z_1 \in \mathbb{S}^3 \\ \text{subject to} & \langle Z_1\,,\, R_1^T A_j R_1 \rangle = 0\,, \qquad j = 1,2,3 \end{array} \tag{625}$$

A nonzero randomly selected matrix $Z_1$ having $\mathbf{0}$ main diagonal is feasible and yields a nonzero perturbation matrix. Choose, arbitrarily,

$$Z_1 = \mathbf{1}\mathbf{1}^T - I \in \mathbb{S}^3 \tag{626}$$

then (rounding)

$$B_1 = \begin{bmatrix} 0 & 0 & 0.0247 & 0 & 0.1048 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0247 & 0 & 0 & 0 & 0.1657 \\ 0 & 0 & 0 & 0 & 0 \\ 0.1048 & 0 & 0.1657 & 0 & 0 \end{bmatrix} \tag{627}$$

**`Step 2:`** $t_1^\star = 1$ because $\lambda(Z_1) = [\,-1 \quad -1 \quad 2\,]^T$. So,

$$X^\star \leftarrow \delta(x_{\rm M}) + B_1 = \begin{bmatrix} \frac{2}{128} & 0 & 0.0247 & 0 & 0.1048 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0247 & 0 & \frac{5}{128} & 0 & 0.1657 \\ 0 & 0 & 0 & 0 & 0 \\ 0.1048 & 0 & 0.1657 & 0 & \frac{90}{128} \end{bmatrix} \tag{628}$$

has rank $\rho \leftarrow 1$ 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* (§4.1.1.3.1). 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. ▼

## 4.3.4 thoughts regarding rank reduction

Because the *rank reduction procedure* is guaranteed only to produce another optimal solution conforming to Barvinok's upper bound (232), 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 251, and the enormity of choices for $Z_i$ (608) 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 (546))[4.19] has inequality constraints of the form

$$\alpha_i^T \operatorname{svec} X \preceq \beta_i\ , \quad i = 1 \ldots k \tag{629}$$

where the $\beta_i$ are scalars. One expedient way to handle this circumstance is to convert the inequality constraints to equality constraints by introducing a slack variable $\gamma$; *id est*,

$$\alpha_i^T \operatorname{svec} X + \gamma_i = \beta_i\ , \quad i = 1 \ldots k\ , \qquad \gamma \succeq 0 \tag{630}$$

thereby converting the problem to prototypical form.

Alternatively, we say the $i^{\text{th}}$ inequality constraint is *active* when it is met with equality; *id est*, when for particular $i$ in (629), $\alpha_i^T \operatorname{svec} X^\star = \beta_i$ . An optimal high-rank solution $X^\star$ 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

[4.19] Contemporary numerical packages for solving semidefinite programs can solve a wider range of problem than our conic prototype (546). Generally, they do so by transforming a given problem into some prototypical form by introducing new constraints and variables. [9] [293] We are momentarily considering a departure from the primal prototype that augments the constraint set with affine inequalities.

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 $\hat{\mathcal{A}}$ substituting for (549). Then we proceed with rank reduction of $X^\star$ as though the semidefinite program were in prototypical form (546P).

## 4.4 Rank-constrained semidefinite program

Here we introduce a technique for finding low-rank optimal solutions to semidefinite programs of a more general form:

### 4.4.1 rank constraint by convex iteration

Given a feasibility problem of the form

$$\begin{array}{cl} \text{find} & G \in \mathbb{S}_+^N \\ \text{subject to} & G \in \mathcal{C} \\ & \text{rank}\, G \leq n \end{array} \tag{631}$$

where $\mathcal{C}$ is a convex set presumed to contain positive semidefinite matrices of rank $n$ or less, we instead solve the convex problem

$$\begin{array}{cl} \underset{G \in \mathbb{S}^N}{\text{minimize}} & \langle G\,,\,W\rangle \\ \text{subject to} & G \in \mathcal{C} \\ & G \succeq 0 \end{array} \tag{632}$$

where direction matrix $W$ is an optimal solution to semidefinite program

$$\sum_{i=n+1}^{N} \lambda(G^\star)_i \;=\; \begin{array}{cl} \underset{W \in \mathbb{S}^N}{\text{minimize}} & \langle G^\star,\, W\rangle \\ \text{subject to} & 0 \preceq W \preceq I \\ & \text{tr}\, W = N - n \end{array} \qquad (1480\text{a})$$

whose feasible set is a Fantope (§2.3.2.0.1), and where $G^\star$ is an optimal solution to problem (632) given some iterate $W$. The idea is to iterate solution of (632) and (1480a) until convergence, as defined in §4.4.1.1. [4.20]

[4.20] The proposed iteration is not an alternating projection. (*confer* Figure **120**)

Optimal matrix $W^\star$ is defined as any direction matrix yielding optimal solution $G^\star$ of rank $n$ or less to then convex equivalent (632) of feasibility problem (631); *id est*, any direction matrix for which the last $N-n$ eigenvalues $\lambda$ of $G^\star$ are zero: (p.541)

$$\sum_{i=n+1}^{N} \lambda(G^\star)_i \;=\; \langle G^\star,\, W^\star\rangle \;\stackrel{\Delta}{=}\; 0 \tag{633}$$

We emphasize that convex problem (632) is not a relaxation of the rank-constrained feasibility problem (631); at convergence, *convex iteration* (632) (1480a) makes it instead an *equivalent problem.*[4.21]

We make no assumption regarding uniqueness of direction matrix $W$. The feasible set of direction matrices in (1480a) is the convex hull of outer product of all rank-$(N-n)$ orthonormal matrices; *videlicet*,

$$\operatorname{conv}\big\{UU^T \mid U\in\mathbb{R}^{N\times N-n},\; U^TU=I\big\} = \big\{A\in\mathbb{S}^N \mid I\succeq A\succeq 0\,,\; \langle I\,,\,A\,\rangle = N-n\big\} \tag{79}$$

Set $\{UU^T \mid U\in\mathbb{R}^{N\times N-n},\; U^TU=I\}$ comprises the extreme points of this Fantope (79).

#### 4.4.1.1 convergence

We study convergence to ascertain conditions under which a direction matrix will reveal a feasible $G$ matrix of rank $n$ or less in semidefinite program (632). Denote by $W^\star$ a particular optimal direction matrix from semidefinite program (1480a) such that (633) holds. Then we define *global convergence* of the iteration (632) (1480a) to correspond with this vanishing vector inner-product (633) of optimal solutions.

Because this iterative technique for constraining rank is not a projection method, it can find a rank-$n$ solution $G^\star$ ((633) will be satisfied) only if at least one exists in the feasible set of program (632).

**4.4.1.1.1 Proof.** Suppose $\langle G^\star,\, W\,\rangle=\phi$ is satisfied for some nonnegative constant $\phi$ after any particular iteration (632) (1480a) of the two minimization problems. Once a particular value of $\phi$ is achieved,

[4.21] Terminology *equivalent problem* meaning, optimal solution to one problem can be derived from optimal solution to another. Terminology *same problem* means: optimal solution set for one problem is identical to the optimal solution set of another (without transformation).

it can never be exceeded by subsequent iterations because existence of feasible $G$ and $W$ having that vector inner-product $\phi$ 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. [189, §1.2] [30, §1.1] *Local convergence* to some $\phi$ is thereby established.

When a rank-$n$ feasible solution to (632) exists, it remains pending to show under what conditions $\langle G^\star, W^\star\rangle = 0$ (633) is achieved by iterative solution of semidefinite programs (632) and (1480a). Then pair $(G^\star, W^\star)$ becomes a fixed-point of iteration. ♦

A nonexistent feasible rank-$n$ solution would mean failure to converge by definition (633) but, as proved, the convex iteration always converges locally if not globally. Now, an application:

**4.4.1.1.2 Example.** *Sensor-Network Localization and Wireless Location.* Heuristic solution proposed by Carter & Jin to a sensor-network localization problem appeared in a reputable journal [51][4.22] despite the heavy reliance on heuristics, limitation to two Euclidean dimensions, and misapplication of semidefinite programming (SDP). A large network is partitioned into smaller subnetworks (as small as one sensor) 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 in so far 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 misapplied. As partitions are selected based on "rule sets" (heuristics, not geographics), they also term the partitioning *adaptive*. But there is no adaptation once a partition is determined; hence, another misapplication of an exacting technical term.

One can reasonably argue that semidefinite programming methods are unnecessary for localization of large sensor networks. In the past, these nonlinear localization problems were solved algebraically and computed by

[4.22] Despite the fact that his name appears as fourth author, Ye had no involvement in writing this cited paper nor did he contribute to its content. The paper constitutes Jin's dissertation for University of Toronto although her name appears as second author.

least squares solution to hyperbolic equations; called *multilateration.*[4.23] Indeed, practical contemporary numerical methods for global positioning by satellite (GPS) do not rely on semidefinite programming.

The beauty of semidefinite programming as relates to localization lies in convex expression of classical multilateration: So & Ye showed [239] 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 is also a question logically consequent to their reliance on complicated and extensive heuristics for partitioning a large network and for solving a partition whose intersensor measurement data is inadequate for localization by distance geometry. While partitions range in size between 2 and 10 sensors, 5 sensors optimally, heuristics provided are only for 2 spatial dimensions (no higher-dimensional algorithm is proposed). For these small numbers it remains unclarified as to precisely what advantage is gained over traditional least squares by solving many little semidefinite programs.

Partitioning of large sensor networks is a logical alternative to rapid growth of SDP computational complexity 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. [39, §2.2] (§5.13.2) 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. When heuristics fail, generally they are repaired by adding more heuristics. Tenuous is any presumption, for example, that distance measurement errors have distribution characterized by circular contours of equal probability about an unknown sensor-location. 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

[4.23] 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.

orientation and eccentricity as one recedes from a sensor. Each unknown sensor must therefore instead be bound to its own particular range of distance, primarily determined by the terrain.[4.24] The nonconvex problem we must instead solve is:

$$\begin{array}{cl}\underset{i\,,\,j\,\in\,\mathcal{I}}{\text{find}} & \{x_i\,,\ x_j\}\\ \text{subject to} & \underline{d_{ij}} \leq \|x_i - x_j\|^2 \leq \overline{d_{ij}}\end{array}\tag{634}$$

where $x_i$ represents sensor location, and where $\underline{d_{ij}}$ and $\overline{d_{ij}}$ respectively represent lower and upper bounds on measured distance from $i^{\text{th}}$ to $j^{\text{th}}$ sensor (or from sensor to anchor). Figure **67** 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 the network is partitioned 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. [12, §1] Biswas posed localization as an optimization problem minimizing a distance measure. [35] [33] Intuitively, minimization of any distance measure yields compacted solutions; (*confer* §6.4.0.0.1) 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: [32] [34].

The sensor-network localization problem is considered difficult. [12, §2] Rank constraints in optimization are considered more difficult. In what

[4.24] A distinct contour map corresponding to each anchor is required in practice.

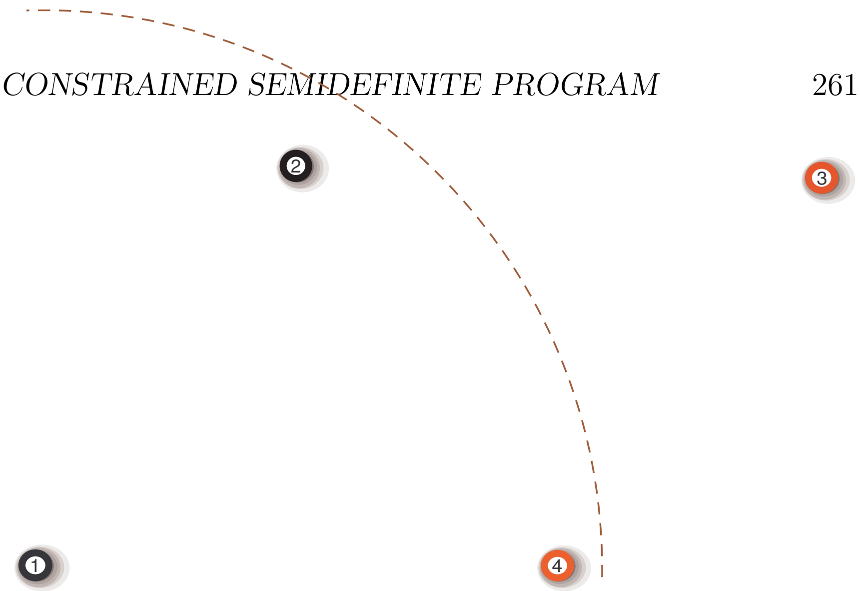



Figure 63: 2-lattice in $\mathbb{R}^2$, hand-drawn. Nodes 3 and 4 are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc.

follows, we present the localization problem as a semidefinite program (equivalent to (634)) having an explicit rank constraint which controls Euclidean 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.

**proposed standardized test**

Jin proposes an academic test in real Euclidean two-dimensional 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 **63** through Figure **66** and their particular connectivity represented by matrices (635) through (638) respectively.

$$\begin{matrix} 0 & \bullet & ? & \bullet \\ \bullet & 0 & \bullet & \bullet \\ ? & \bullet & 0 & \circ \\ \bullet & \bullet & \circ & 0 \end{matrix} \tag{635}$$

Matrix entries *dot* $\bullet$ indicate measurable distance between nodes while unknown distance is denoted by ? (*question mark*). Matrix entries *hollow dot* $\circ$ represent known distance between anchors (to very high

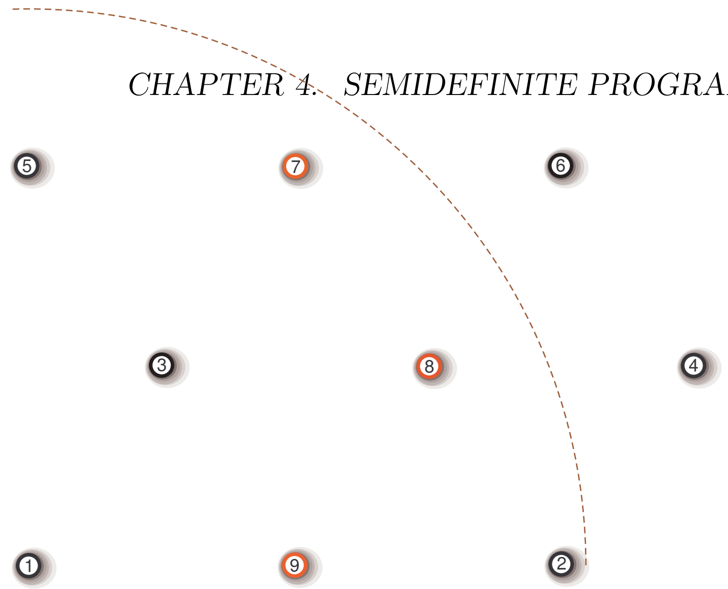



Figure 64: 3-lattice in $\mathbb{R}^2$, hand-drawn. Nodes 7, 8, and 9 are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc.

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.

$$\begin{array}{ccccccccc}
0 & \bullet & \bullet & ? & \bullet & ? & ? & \bullet & \bullet \\
\bullet & 0 & \bullet & \bullet & ? & \bullet & ? & \bullet & \bullet \\
\bullet & \bullet & 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
? & \bullet & \bullet & 0 & ? & \bullet & \bullet & \bullet & \bullet \\
\bullet & ? & \bullet & ? & 0 & \bullet & \bullet & \bullet & \bullet \\
? & \bullet & \bullet & \bullet & \bullet & 0 & \bullet & \bullet & \bullet \\
? & ? & \bullet & \bullet & \bullet & \bullet & 0 & \circ & \circ \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \circ & 0 & \circ \\
\bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \circ & \circ & 0
\end{array} \tag{636}$$

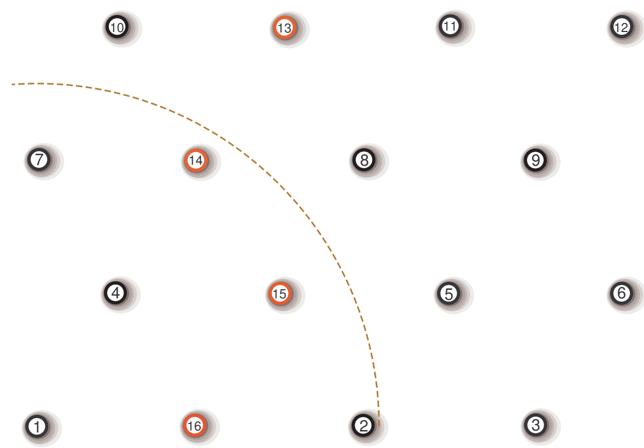



Figure 65: 4-lattice in $\mathbb{R}^2$, hand-drawn. Nodes 13, 14, 15, and 16 are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc.

$$
\begin{array}{cccccccccccccccc}
0 & ? & ? & \bullet & ? & ? & \bullet & ? & ? & ? & ? & ? & ? & ? & \bullet & \bullet \\
? & 0 & \bullet & \bullet & \bullet & \bullet & ? & \bullet & ? & ? & ? & ? & ? & \bullet & \bullet & \bullet \\
? & \bullet & 0 & ? & \bullet & \bullet & ? & ? & \bullet & ? & ? & ? & ? & ? & \bullet & \bullet \\
\bullet & \bullet & ? & 0 & \bullet & ? & \bullet & \bullet & ? & \bullet & ? & ? & \bullet & \bullet & \bullet & \bullet \\
? & \bullet & \bullet & \bullet & 0 & \bullet & ? & \bullet & \bullet & ? & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
? & \bullet & \bullet & ? & \bullet & 0 & ? & \bullet & \bullet & ? & \bullet & \bullet & ? & ? & ? & ? \\
\bullet & ? & ? & \bullet & ? & ? & 0 & ? & ? & \bullet & ? & ? & \bullet & \bullet & \bullet & \bullet \\
? & \bullet & ? & \bullet & \bullet & \bullet & ? & 0 & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\
? & ? & \bullet & ? & \bullet & \bullet & ? & \bullet & 0 & ? & \bullet & \bullet & \bullet & ? & \bullet & ? \\
? & ? & ? & \bullet & ? & ? & \bullet & \bullet & ? & 0 & \bullet & ? & \bullet & \bullet & \bullet & ? \\
? & ? & ? & ? & \bullet & \bullet & ? & \bullet & \bullet & \bullet & 0 & \bullet & \bullet & \bullet & \bullet & ? \\
? & ? & ? & ? & \bullet & \bullet & ? & \bullet & \bullet & ? & \bullet & 0 & ? & ? & ? & ? \\
? & ? & ? & \bullet & \bullet & ? & \bullet & \bullet & \bullet & \bullet & \bullet & ? & 0 & \circ & \circ & \circ \\
? & \bullet & ? & \bullet & \bullet & ? & \bullet & \bullet & ? & \bullet & \bullet & ? & \circ & 0 & \circ & \circ \\
\bullet & \bullet & \bullet & \bullet & \bullet & ? & \bullet & \bullet & \bullet & \bullet & \bullet & ? & \circ & \circ & 0 & \circ \\
\bullet & \bullet & \bullet & \bullet & \bullet & ? & \bullet & \bullet & ? & ? & ? & ? & \circ & \circ & \circ & 0
\end{array}
\tag{637}
$$

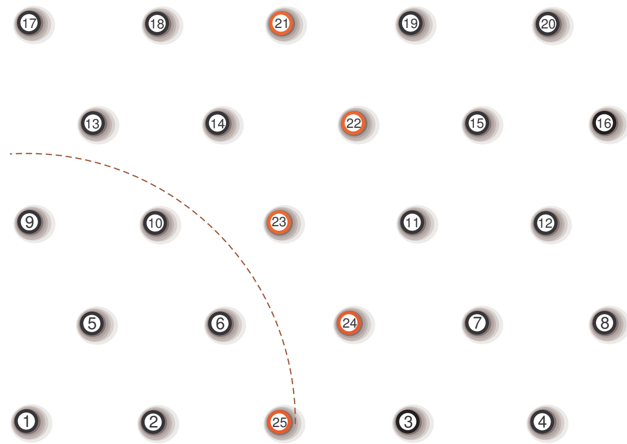



Figure 66: 5-lattice in $\mathbb{R}^{\mathbf{2}}$. Nodes 21 through 25 are anchors.

$$
\begin{array}{ccccccccccccccccccccccccc}
0 & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? \\
\bullet & 0 & ? & ? & \bullet & \bullet & ? & ? & ? & \bullet & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & \bullet \\
? & ? & 0 & \bullet & ? & \bullet & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & \bullet & \bullet \\
? & ? & \bullet & 0 & ? & ? & \bullet & \bullet & ? & ? & ? & \bullet & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & ? \\
\bullet & \bullet & ? & ? & 0 & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & ? & ? & ? & ? & ? & \bullet & ? & \bullet \\
\bullet & \bullet & \bullet & ? & \bullet & 0 & \bullet & ? & \bullet & \bullet & \bullet & ? & ? & \bullet & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & \bullet & \bullet \\
? & ? & \bullet & \bullet & ? & \bullet & 0 & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & ? & ? & ? & ? & \bullet & \bullet & \bullet \\
? & ? & \bullet & \bullet & ? & ? & \bullet & 0 & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & ? & ? & ? & ? & ? & \bullet & ? \\
\bullet & ? & ? & ? & \bullet & \bullet & ? & ? & 0 & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & ? & ? & ? & ? & ? \\
? & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & 0 & \bullet & ? & \bullet & \bullet & ? & ? & ? & \bullet & ? & ? & \bullet & \bullet & \bullet & \bullet & \bullet \\
? & ? & \bullet & ? & ? & \bullet & \bullet & \bullet & ? & \bullet & 0 & \bullet & ? & \bullet & \bullet & \bullet & ? & ? & \bullet & ? & ? & \bullet & \bullet & \bullet & \bullet \\
? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & 0 & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & \bullet & \bullet & \bullet & ? \\
? & ? & ? & ? & \bullet & ? & ? & ? & \bullet & \bullet & ? & ? & 0 & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & ? \\
? & ? & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & \bullet & ? & \bullet & 0 & \bullet & ? & \bullet & \bullet & \bullet & ? & \bullet & \bullet & \bullet & \bullet & ? \\
? & ? & ? & ? & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & \bullet & 0 & \bullet & ? & ? & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & ? \\
? & ? & ? & ? & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & 0 & ? & ? & \bullet & \bullet & ? & \bullet & ? & ? & ? \\
? & ? & ? & ? & \bullet & ? & ? & ? & \bullet & ? & ? & ? & \bullet & \bullet & ? & ? & 0 & \bullet & ? & ? & \bullet & ? & ? & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & 0 & \bullet & ? & \bullet & \bullet & \bullet & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & \bullet & ? & \bullet & \bullet & \bullet & ? & \bullet & 0 & \bullet & \bullet & \bullet & \bullet & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & ? & ? & \bullet & \bullet & ? & ? & \bullet & 0 & \bullet & \bullet & ? & ? & ? \\
? & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & ? & ? & \bullet & \bullet & \bullet & ? & \bullet & \bullet & \bullet & \bullet & 0 & \circ & \circ & \circ & \circ \\
? & ? & ? & ? & ? & ? & ? & ? & ? & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & ? & \bullet & \bullet & \bullet & \circ & 0 & \circ & \circ & \circ \\
? & ? & \bullet & ? & \bullet & \bullet & \bullet & ? & ? & \bullet & \bullet & \bullet & ? & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & \circ & \circ & 0 & \circ & \circ \\
? & \bullet & \bullet & \bullet & ? & \bullet & \bullet & \bullet & ? & \bullet & \bullet & \bullet & ? & \bullet & \bullet & ? & ? & ? & ? & ? & \circ & \circ & \circ & 0 & \circ \\
? & \bullet & \bullet & ? & \bullet & \bullet & \bullet & ? & ? & \bullet & \bullet & ? & ? & ? & ? & ? & ? & ? & ? & ? & \circ & \circ & \circ & \circ & 0
\end{array}
\tag{638}
$$

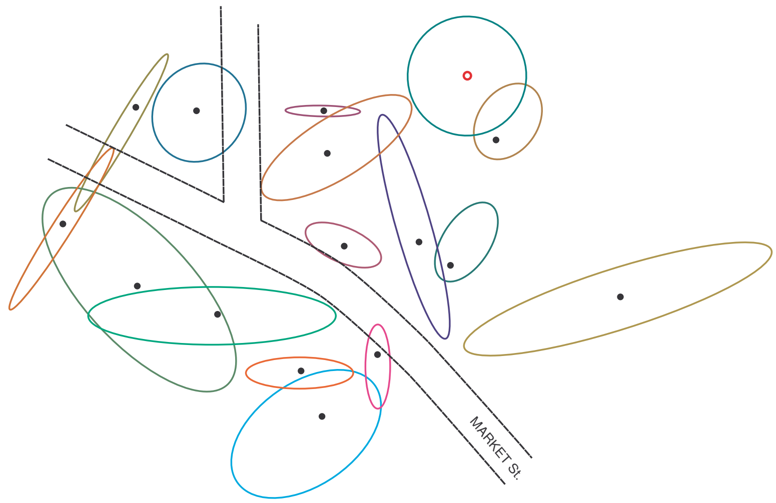



Figure 67: Uncertainty ellipsoid in $\mathbb{R}^{\mathbf{2}}$ for each of 15 sensors • located within three city blocks in downtown San Francisco. Data by Polaris Wireless. [241]

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

$$X = [\, x_1 \ \cdots \ x_N \,] \in \mathbb{R}^{n\times N} \qquad (65)$$

where $N$ is regarded as cardinality of list $X$. Positive semidefinite matrix $X^TX$, formed from inner product of the list, is a Gram matrix; [182, §3.6]

$$G \triangleq X^TX = \begin{bmatrix} \|x_1\|^2 & x_1^Tx_2 & x_1^Tx_3 & \cdots & x_1^Tx_N \\ x_2^Tx_1 & \|x_2\|^2 & x_2^Tx_3 & \cdots & x_2^Tx_N \\ x_3^Tx_1 & x_3^Tx_2 & \|x_3\|^2 & \ddots & x_3^Tx_N \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ x_N^Tx_1 & x_N^Tx_2 & x_N^Tx_3 & \cdots & \|x_N\|^2 \end{bmatrix} \in \mathbb{S}_+^N \qquad (718)$$

where $\mathbb{S}_+^N$ is the convex cone of $N\times N$ positive semidefinite matrices in the real symmetric matrix subspace $\mathbb{S}^N$.

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 $i^{\text{th}}$

to $j^{\text{th}}$ 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 **67**.

Our mathematical treatment of anchors and sensors is not dichotomized.[4.25] A known sensor position to high accuracy $\check{x}_i$ is an anchor. Then the sensor-network localization problem (634) can be expressed equivalently: Given a number of anchors $m$, and $\mathcal{I}$ a set of indices (corresponding to all existing distance measurements $\bullet$), for $0<n<N$

$$\begin{array}{cll}
\underset{G\in\mathbb{S}^N,\ X\in\mathbb{R}^{n\times N}}{\text{minimize}} & \operatorname{tr} Z & \\
\text{subject to} & \underline{d_{ij}} \leq \langle G\,,\,(e_i-e_j)(e_i-e_j)^T\rangle \leq \overline{d_{ij}} & \forall(i,j)\in\mathcal{I} \\
& \langle G\,,\,e_ie_i^T\rangle \qquad\qquad\qquad\qquad = \|\check{x}_i\|^2\,, & i=N-m+1\ldots N \\
& \langle G\,,\,(e_ie_j^T+e_je_i^T)/2\rangle \qquad\quad = \check{x}_i^T\check{x}_j\,, & i<j\,,\quad \forall\, i,j\in\{N-m+1\ldots N\} \\
& X(:,N-m+1:N) \qquad\qquad\;\; = [\,\check{x}_{N-m+1}\;\cdots\;\check{x}_N\,] & \\
& Z=\begin{bmatrix} I & X \\ X^T & G\end{bmatrix} \qquad\qquad\;\; \succeq 0 & \\
& \operatorname{rank} Z \qquad\qquad\qquad\qquad\;\; = n &
\end{array} \tag{639}$$

where $e_i$ is the $i^{\text{th}}$ member of the standard basis for $\mathbb{R}^N$. Distance-square

$$d_{ij} \;=\; \|x_i-x_j\|_2^2 \;\triangleq\; \langle x_i-x_j\;,\;x_i-x_j\rangle \tag{705}$$

is related to Gram matrix entries $G\triangleq[g_{ij}]$ by a vector inner-product

$$\begin{aligned} d_{ij} &= g_{ii}+g_{jj}-2g_{ij} \\ &= \langle G\,,\,(e_i-e_j)(e_i-e_j)^T\rangle \;\triangleq\; \operatorname{tr}(G^T(e_i-e_j)(e_i-e_j)^T) \end{aligned} \tag{720}$$

hence the scalar inequalities. The objective function $\operatorname{tr} Z$ is a heuristic whose sole purpose is to represent the convex envelope of $\operatorname{rank} Z$. (§7.2.2.1.1) By Schur complement (§A.4) any feasible $G$ and $X$ provide a comparison with respect to the positive semidefinite cone

$$G \succeq X^TX \tag{753}$$

[4.25] *Wireless location* problem thus stated identically; difference being: fewer sensors.

which is a convex relaxation of the desired equality constraint

$$\begin{bmatrix} I & X \\ X^T & G \end{bmatrix} = \begin{bmatrix} I \\ X^T \end{bmatrix} [\, I \;\; X\,] \qquad (754)$$

The rank constraint insures this equality holds thus restricting solution to $\mathbb{R}^n$.

**convex equivalent problem statement**

Problem statement (639) 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 $0 < n < N$

$$\begin{array}{cll}
\underset{G\in\mathbb{S}^N,\; X\in\mathbb{R}^{n\times N}}{\text{minimize}} & \langle Z\,,\, W\rangle & \\
\text{subject to} & \underline{d_{ij}} \leq \langle G\,,\,(e_i-e_j)(e_i-e_j)^T\rangle \leq \overline{d_{ij}} & \forall(i,j)\in\mathcal{I} \\
& \langle G\,,\, e_ie_i^T\rangle \qquad\qquad\qquad\quad = \|\check{x}_i\|^2\,, & i=N-m+1\ldots N \\
& \langle G\,,\,(e_ie_j^T+e_je_i^T)/2\rangle \qquad\; = \check{x}_i^T\check{x}_j\,, & i<j\,,\quad \forall\, i,j\in\{N-m+1\ldots N\} \\
& X(:\,,N-m+1\!:\!N) \qquad\qquad\; = [\,\check{x}_{N-m+1}\;\cdots\;\check{x}_N\,] & \\
& Z = \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} \qquad\qquad\;\; \succeq 0 &
\end{array} \qquad (640)$$

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 *slab* (an intersection of two parallel but opposing halfspaces, Figure **9**). In this convex optimization problem (640), a semidefinite program, we substitute a vector inner-product objective function for trace from nonconvex problem (639);

$$\langle Z\,,\, I\rangle \;=\; \operatorname{tr} Z \;\leftarrow\; \langle Z\,,\, W\rangle \qquad (641)$$

a generalization of the known trace heuristic [91] for minimizing convex envelope of rank, where $W\in\mathbb{S}_+^{N+n}$ is constant with respect to (640). Matrix $W$ is normal to a hyperplane in $\mathbb{S}^{N+n}$ minimized over a convex feasible set specified by the constraints in (640). Matrix $W$ is chosen so $-W$ points in the direction of a feasible rank-$n$ Gram matrix. Thus the purpose of vector inner-product objective (641) is to locate a feasible rank-$n$ Gram matrix that is presumed existent on the boundary of positive semidefinite cone $\mathbb{S}_+^N$.

**direction matrix $W$**

Denote by $Z^\star$ an optimal composite matrix from semidefinite program (640). Then for $Z^\star \in \mathbb{S}^{N+n}$ whose eigenvalues $\lambda(Z^\star) \in \mathbb{R}^{N+n}$ are arranged in nonincreasing order, (Fan)

$$\begin{array}{rcll}\sum_{i=n+1}^{N+n} \lambda(Z^\star)_i &=& \underset{W\in\,\mathbb{S}^{N+n}}{\text{minimize}} & \langle Z^\star,\, W\rangle \\ && \text{subject to} & 0 \preceq W \preceq I \\ &&& \operatorname{tr} W = N\end{array} \tag{1480a}$$

whose optimal solution is known in closed form. This eigenvalue sum is zero when $Z^\star$ has rank $n$ or less.

Foreknowledge of optimal $Z^\star$, to make possible this search for $W$, implies recursion; *id est*, semidefinite program (640) is solved for $Z^\star$ initializing $W = I$ or $W = \mathbf{0}$. Once found, $Z^\star$ becomes constant in semidefinite program (1480a) where a new normal direction $W$ is found as its optimal solution. Then the cycle (640) (1480a) iterates until convergence. When $\operatorname{rank} Z^\star = n$, solution via this convex iteration is optimal for sensor-network localization problem (634) and its equivalent (639).

**numerical solution**

In all examples to follow, number of anchors

$$m = \sqrt{N} \tag{642}$$

equals square root of cardinality $N$ of list $X$. Indices set $\mathcal{I}$ identifying all existing distance measurements $\bullet$ is ascertained from connectivity matrix (635), (636), (637), or (638). We solve iteration (640) (1480a) in dimension $n = 2$ for each respective example illustrated in Figure **63** through Figure **66**.

In presence of negligible noise, actual position is reliably localized for every standardized example; noteworthy in so far as each example represents an incomplete graph. This means the set of all solutions having lowest rank is a single point, to within a rigid transformation.

To make the examples interesting and consistent with previous work, we randomize each range of distance-square that bounds $\langle G\,,\,(e_i - e_j)(e_i - e_j)^T\rangle$ in (640); *id est*, for each and every $(i\,,j) \in \mathcal{I}$

$$\begin{aligned}\overline{d_{ij}} &= d_{ij}(1+\sqrt{3}\,\texttt{rand(1)}\,\eta)^2 \\ \underline{d_{ij}} &= d_{ij}(1-\sqrt{3}\,\texttt{rand(1)}\,\eta)^2\end{aligned} \tag{643}$$

where $\eta = 0.1$ is a constant noise factor, `rand(1)` is the MATLAB function providing one sample of uniformly distributed noise in the interval $[0,1]$, and $d_{ij}$ is actual distance-square from $i^{\text{th}}$ to $j^{\text{th}}$ sensor. Because of the separate function calls `rand(1)`, each range of distance-square $[\,\underline{d_{ij}}\,,\,\overline{d_{ij}}\,]$ is not necessarily centered on actual distance-square $d_{ij}$. The factor $\sqrt{3}$ provides unit variance on the stochastic range.

Figure **68** through Figure **71** each illustrate one realization of numerical solution to the standardized lattice problems posed by Figure **63** through Figure **66** respectively. Exact localization is impossible because of measurement noise. Certainly, by inspection of their published graphical data, our new results are competitive with those of Carter & Jin. 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 $\eta$; which is all we intended to demonstrate.

**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.

The legacy of Carter & Jin [51] is a sobering demonstration of the need for more efficient methods for solution of semidefinite programs, while that of So & Ye [239] is the bonding of distance geometry to semidefinite programming. Elegance of our semidefinite problem statement (640) for a sensor-network localization problem in any dimension should provide some impetus to focus more research on computational intensity. Higher speed and greater accuracy from a simplex-like solver is what is required. □

We numerically tested the foregoing technique for constraining rank on a wide range of problems including localization of randomized positions, *stress* (§7.2.2.7.1), *ball packing* (§5.4.2.2.3), and cardinality problems. We have had some success introducing the direction vector inner-product (641) as a *regularization* term (Pareto optimization) whose purpose is to constrain rank, affine dimension, or cardinality:

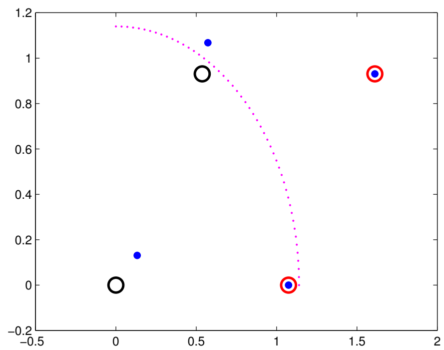



Figure 68: Typical solution for 2-lattice in Figure **63** with noise factor $\eta = 0.1$. Two red rightmost nodes are anchors; two remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius = 1.14. Actual sensor indicated by target ○ while its localization is indicated by bullet •. Rank-2 solution found in 1 iteration (640) (1480a) subject to reflection error.

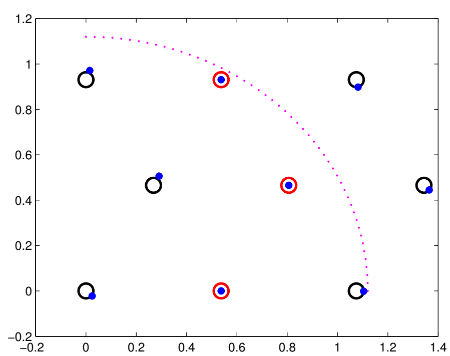



Figure 69: Typical solution for 3-lattice in Figure **64** with noise factor $\eta = 0.1$. Three red vertical middle nodes are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius = 1.12. Actual sensor indicated by target ○ while its localization is indicated by bullet •. Rank-2 solution found in 2 iterations (640) (1480a).

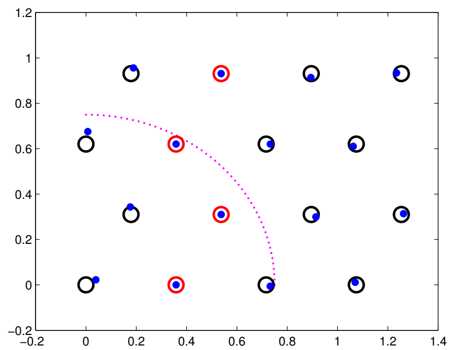


Figure 70: Typical solution for 4-lattice in Figure **65** with noise factor $\eta = 0.1$. Four red vertical middle-left nodes are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius $= 0.75$. Actual sensor indicated by target $\bigcirc$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 7 iterations (640) (1480a).

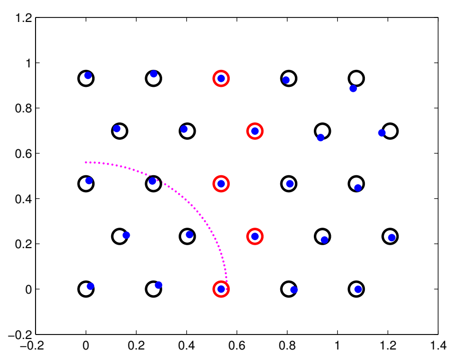


Figure 71: Typical solution for 5-lattice in Figure **66** with noise factor $\eta = 0.1$. Five red vertical middle nodes are anchors; remaining nodes are sensors. Radio range of sensor 1 indicated by arc; radius $= 0.56$. Actual sensor indicated by target $\bigcirc$ while its localization is indicated by bullet $\bullet$. Rank-2 solution found in 3 iterations (640) (1480a).

### 4.4.2 cardinality

Our goal is to reliably constrain rank in a semidefinite program. There is a direct analogy to linear programming that is simpler to present but equally hard to solve. In Optimization, that analogy is known as the *cardinality problem.* If we can solve the cardinality problem, then solution to the rank-constraint problem follows; and *vice versa.*

Consider a feasibility problem equivalent to the classical problem from linear algebra $Ax=b$, but with an upper bound $k$ on cardinality $\|x\|_0$ of a nonnegative solution $x$: for vector $b\in\mathcal{R}(A)$

$$\begin{array}{rl}\text{find} & x\in\mathbb{R}^n\\ \text{subject to} & Ax=b\\ & x\succeq 0\\ & \|x\|_0\leq k\end{array}\tag{644}$$

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 $\mathbb{S}^n_+$. [46, §3.4.2]

We propose that cardinality-constrained feasibility problem (644) is equivalently expressed with convex constraints:

$$\begin{array}{rl}\underset{x\in\mathbb{R}^n,\ y\in\mathbb{R}^n}{\text{minimize}} & x^Ty\\ \text{subject to} & Ax=b\\ & x\succeq 0\\ & 0\preceq y\preceq\mathbf{1}\\ & y^T\mathbf{1}=n-k\end{array}\tag{645}$$

whose *bilinear* objective function $x^Ty$ is quasiconcave only when $n=1$. This simple-looking problem (645) is very hard to solve, yet is not hard to understand. Because the sets feasible to $x$ and $y$ are not interdependent, we can separate the problem half in variable $y$:

$$\begin{array}{rl}\underset{y\in\mathbb{R}^n}{\text{minimize}} & x^Ty\\ \text{subject to} & 0\preceq y\preceq\mathbf{1}\\ & y^T\mathbf{1}=n-k\end{array}\tag{434}$$

This linear program sums the $n-k$ smallest entries from vector $x$. In context of problem (645), we want $n-k$ entries of $x$ to sum to zero; *id est*, we want a globally optimal objective $x^{\star T}y^{\star}=0$ to vanish. Because all entries in $x$ must be nonnegative, then $n-k$ entries are themselves zero whenever their sum is, and then cardinality of $x\in\mathbb{R}^n$ is at most $k$.

Ideally, one wants to solve (645) directly, but contemporary techniques for doing so are computationally intensive.[4.26] Nevertheless, solving (645) should not be ruled out, assuming an efficient method is discovered or transformation to a convex equivalent can be found.

One efficient way to solve (645) is by transforming it to a sequence of convex problems:

$$\begin{array}{cl}\underset{x\in\mathbb{R}^n}{\text{minimize}} & x^T y\\ \text{subject to} & Ax=b\\ & x\succeq 0\end{array} \tag{646}$$

$$\begin{array}{cl}\underset{y\in\mathbb{R}^n}{\text{minimize}} & x^T y\\ \text{subject to} & 0\preceq y\preceq \mathbf{1}\\ & y^T\mathbf{1}=n-k\end{array} \tag{434}$$

This sequence is iterated until $x^T y$ vanishes; *id est*, until desired cardinality is achieved. This technique works often and, for some problem classes (beyond $Ax=b$), it works all the time; meaning, optimal solution to problem (645) can often be found by this convex iteration. But examples can be found that make the iteration stall at a solution not of desired cardinality. Heuristics for breaking out of a stall can be implemented with some success:

## 4.4.3 more cardinality and rank constraint examples

**4.4.3.0.1 Example.** *Sparsest solution to* $Ax=b$.
Given data, from Example 4.2.3.0.2,

$$A=\begin{bmatrix}-1 & 1 & 8 & 1 & 1 & 0\\ -3 & 2 & 8 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2}-\frac{1}{3}\\ -9 & 4 & 8 & \frac{1}{4} & \frac{1}{9} & \frac{1}{4}-\frac{1}{9}\end{bmatrix},\qquad b=\begin{bmatrix}1\\ \frac{1}{2}\\ \frac{1}{4}\end{bmatrix} \tag{578}$$

[4.26] *e.g.*, *branch and bound* method.

the sparsest solution to the classical linear equation $Ax=b$ is $x=e_4\in\mathbb{R}^{\mathbf{6}}$ (*confer* (591)). And given data, from Example 4.2.3.0.3,

$$A = \begin{bmatrix} 1 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{595}$$

the most sparse solution is $x=[\,0\;\;0\;\;\sqrt{2}\,]^T\in\mathbb{R}^{\mathbf{3}}$ (*confer* (596)). Given random data, in Matlab notation,

$$\texttt{A}=\texttt{randn(m,n)}, \quad \texttt{index}=\texttt{round}((\texttt{n}-1)*\texttt{rand}(1))+1, \quad \texttt{b}=\texttt{A}(:,\texttt{index}) \tag{647}$$

where $\texttt{m}$ and $\texttt{n}$ are selected arbitrarily, the sparsest solution is $x=e_{\texttt{index}}\in\mathbb{R}^{\texttt{n}}$ from the standard basis. Although these sparsest solutions are recoverable by inspection, we seek to discern them instead by convex iteration; namely, by iterating problem sequence (646) (434). From the numerical data given, cardinality $\|x\|_0=1$ is expected. Iteration continues until $x^Ty$ vanishes (to within some numerical precision); *id est*, until desired cardinality is achieved.

All three examples return a correct cardinality-1 solution to within machine precision in few iterations, but are occasionally subject to stall. Stalls are remedied by reinitializing $y$ to a random state. □

Stalling is not an inevitable behavior. Convex iteration succeeds, for some types of problem, all the time:

**4.4.3.0.2 Example.** *Projection on ellipsoid boundary.* [38] [99, §5.1] [180, §2] This problem is exceptionally easy to solve by convex iteration: Consider classical linear equation $Ax=b$ but with a constraint on norm of solution $x$, given matrices $C$, $A$, and vector $b\in\mathcal{R}(A)$

$$\begin{array}{cl} \text{find} & x\in\mathbb{R}^N \\ \text{subject to} & Ax=b \\ & \|Cx\|=1 \end{array} \tag{648}$$

The set $\{x \mid \|Cx\|=1\}$ describes an ellipsoid boundary. This is a nonconvex problem because solution is constrained to that boundary. Assign

$$G = \begin{bmatrix} Cx \\ 1 \end{bmatrix} [\,x^TC^T\;\;1\,] = \begin{bmatrix} X & Cx \\ x^TC^T & 1 \end{bmatrix} \triangleq \begin{bmatrix} Cxx^TC^T & Cx \\ x^TC^T & 1 \end{bmatrix} \in \mathbb{S}^{N+1} \tag{649}$$

Any rank-1 solution must have this form. (§B.1.0.2) Ellipsoidally constrained feasibility problem (648) is equivalent to:

$$\begin{array}{cl}\underset{X\in\mathbb{S}^N}{\text{find}} & x\in\mathbb{R}^N\\ \text{subject to} & Ax=b\\ & G=\begin{bmatrix} X & Cx\\ x^TC^T & 1\end{bmatrix}\\ & (G\succeq 0)\\ & \operatorname{rank}G=1\\ & \operatorname{tr}X=1\end{array}\tag{650}$$

This is transformed to an equivalent convex problem by moving the rank constraint to the objective: We iterate solution of

$$\begin{array}{cl}\underset{X\in\mathbb{S}^N,\;x\in\mathbb{R}^N}{\text{minimize}} & \langle G\,,\,Y\rangle\\ \text{subject to} & Ax=b\\ & G=\begin{bmatrix} X & Cx\\ x^TC^T & 1\end{bmatrix}\succeq 0\\ & \operatorname{tr}X=1\end{array}\tag{651}$$

with

$$\begin{array}{cl}\underset{Y\in\mathbb{S}^{N+1}}{\text{minimize}} & \langle G^\star,\,Y\rangle\\ \text{subject to} & 0\preceq Y\preceq I\\ & \operatorname{tr}Y=N\end{array}\tag{652}$$

Direction matrix $Y\in\mathbb{S}^{N+1}$, initially $\mathbf{0}$, controls rank. (1480a) Taking singular value decomposition $G^\star=U\Sigma Q^T\in\mathbb{R}^{N+1}$, (§A.6) then a new direction matrix $Y=U(:\,,2\!:\!N\!+\!1)U(:\,,2\!:\!N\!+\!1)^T$ optimally solves (652) at each iteration. An optimal solution to (648) is thereby found in a few iterations, making convex problem (651) 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 random element:

$$Y=-U(:\,,2\!:\!N\!+\!1)\big(U(:\,,2\!:\!N\!+\!1)^T+\texttt{randn(N,1)}U(:\,,1)^T\big)\tag{653}$$

This heuristic is quite effective for this problem.

When $b \notin \mathcal{R}(A)$ then problem (648) must be restated as a projection:

$$\begin{array}{cl}\underset{x\in\mathbb{R}^N}{\text{minimize}} & \|Ax-b\| \\ \text{subject to} & \|Cx\|=1\end{array} \tag{654}$$

This is a projection of point $b$ on an ellipsoid boundary because any affine transformation of an ellipsoid remains an ellipsoid. Problem (651) in turn becomes

$$\begin{array}{cl}\underset{X\in\mathbb{S}^N,\ x\in\mathbb{R}^N}{\text{minimize}} & \langle G\,,\,Y\rangle \,+\, \|Ax-b\| \\ \text{subject to} & G=\begin{bmatrix} X & Cx \\ x^TC^T & 1\end{bmatrix}\succeq 0 \\ & \operatorname{tr}X=1\end{array} \tag{655}$$

We iterate this with calculation of direction matrix $Y$ as before until a rank-1 $G$ matrix is found. □

**4.4.3.0.3 Example.** *Tractable polynomial constraint.*
The ability to handle rank constraints makes polynomial constraints (generally nonconvex) transformable to convex constraints. *All optimization problems having polynomial objective and polynomial constraints can be reformulated as a semidefinite program with a rank*-1 *constraint.* [210] Suppose we require

$$3+2x-xy\leq 0 \tag{656}$$

Assign

$$G=\begin{bmatrix} x \\ y \\ 1\end{bmatrix}\begin{bmatrix} x & y & 1\end{bmatrix}=\begin{bmatrix} X & z \\ z^T & 1\end{bmatrix}\triangleq\begin{bmatrix} x^2 & xy & x \\ xy & y^2 & y \\ x & y & 1\end{bmatrix}\in\mathbb{S}^3 \tag{657}$$

The polynomial constraint (656) is equivalent to the constraint set (§B.1.0.2)

$$\begin{array}{l}\operatorname{tr}(GA)\leq 0 \\ G=\begin{bmatrix} X & z \\ z^T & 1\end{bmatrix} \\ (G\succeq 0) \\ \operatorname{rank}G=1\end{array} \tag{658}$$

in symmetric variable matrix $X \in \mathbb{S}^{\mathbf{2}}$ and variable vector $z \in \mathbb{R}^{\mathbf{2}}$ where

$$A = \begin{bmatrix} 0 & -\frac{1}{2} & 1 \\ -\frac{1}{2} & 0 & 0 \\ 1 & 0 & 3 \end{bmatrix} \tag{659}$$

Then the method of convex iteration from §4.4.1 is applied to implement the rank constraint. $\square$

**4.4.3.0.4 Example.** *Procrustes problem.* [38]
Example 4.4.3.0.2 is extensible. An orthonormal matrix $Q \in \mathbb{R}^{n\times p}$ is completely characterized by $Q^TQ = I$. Consider the particular case $Q = [\,x \;\; y\,] \in \mathbb{R}^{n\times\mathbf{2}}$ as variable to a Procrustes problem (§C.3): given $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times\mathbf{2}}$

$$\begin{array}{cl} \underset{Q\in\mathbb{R}^{n\times\mathbf{2}}}{\text{minimize}} & \|AQ - B\|_{\mathrm{F}} \\ \text{subject to} & Q^TQ = I \end{array} \tag{660}$$

which is nonconvex. By vectorizing matrix $Q$ we can make the assignment:

$$G = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} [\,x^T \;\; y^T \;\; 1\,] = \begin{bmatrix} X & Z & x \\ Z^T & Y & y \\ x^T & y^T & 1 \end{bmatrix} \triangleq \begin{bmatrix} xx^T & xy^T & x \\ yx^T & yy^T & y \\ x^T & y^T & 1 \end{bmatrix} \in \mathbb{S}^{2n+1} \tag{661}$$

Now Procrustes problem (660) can be equivalently restated:

$$\begin{array}{cl} \underset{X\,,\,Y\,,\,Z\,,\,x\,,\,y}{\text{minimize}} & \|A[\,x \;\; y\,] - B\|_{\mathrm{F}} \\ \text{subject to} & G = \begin{bmatrix} X & Z & x \\ Z^T & Y & y \\ x^T & y^T & 1 \end{bmatrix} \\ & (G \succeq 0) \\ & \operatorname{rank} G = 1 \\ & \operatorname{tr} X = 1 \\ & \operatorname{tr} Y = 1 \\ & \operatorname{tr} Z = 0 \end{array} \tag{662}$$

To solve this, we form the convex problem sequence:

$$\begin{array}{cl}\underset{X\,,\,Y\,,\,Z\,,\,x\,,\,y}{\text{minimize}} & \|A[\,x \;\; y\,]-B\|_{\mathrm F} \,+\, \langle G\,,\,W\rangle \\ \text{subject to} & G=\begin{bmatrix} X & Z & x \\ Z^T & Y & y \\ x^T & y^T & 1\end{bmatrix}\succeq 0 \\ & \operatorname{tr} X=1 \\ & \operatorname{tr} Y=1 \\ & \operatorname{tr} Z=0 \end{array} \tag{663}$$

and

$$\begin{array}{cl}\underset{W\in\,\mathbb{S}^{2n+1}}{\text{minimize}} & \langle G^\star,\,W\rangle \\ \text{subject to} & 0\preceq W\preceq I \\ & \operatorname{tr} W=2n \end{array} \tag{664}$$

whose optimal solution $W$, initially $\mathbf{0}$, is known in closed form (page 541). These two problems are iterated until convergence and a rank-1 $G$ matrix is found. Optimal $Q^\star$ equals $[\,x^\star \;\; y^\star\,]$.

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 $Q$. □

**4.4.3.0.5 Example.** *Boolean vector feasible to* $Ax\preceq b$. (*confer* §4.2.3.0.2) 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 $b\in\mathbb{R}^M$

$$\begin{array}{cl}\text{find} & x\in\mathbb{R}^N \\ \text{subject to} & Ax\preceq b \\ & \delta(xx^T)=\mathbf{1} \end{array} \tag{665}$$

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 $x$:

$$G=\begin{bmatrix} x \\ 1\end{bmatrix}[\,x^T \;\; 1\,]=\begin{bmatrix} X & x \\ x^T & 1\end{bmatrix}\triangleq\begin{bmatrix} xx^T & x \\ x^T & 1\end{bmatrix}\in\mathbb{S}^{N+1} \tag{666}$$

Then design an equivalent semidefinite feasibility problem to find a Boolean solution to $Ax \preceq b$:

$$\begin{array}{cl}\underset{X\in\mathbb{S}^N}{\text{find}} & x\in\mathbb{R}^N\\ \text{subject to} & Ax\preceq b\\ & G=\left[\begin{array}{cc}X & x\\ x^T & 1\end{array}\right]\\ & \operatorname{rank} G = 1\\ & (G\succeq 0)\\ & \delta(X)=\mathbf{1}\end{array}\tag{667}$$

where $x_i^\star \in \{-1,1\}$, $i=1\ldots N$. The two variables $X$ and $x$ are made dependent via their assignment to rank-1 matrix $G$. By (1404), an optimal rank-1 matrix $G^\star$ must take the form (666).

As before, we regularize the rank constraint by introducing a direction matrix $Y$ into the objective:

$$\begin{array}{cl}\underset{X\in\mathbb{S}^N,\ x\in\mathbb{R}^N}{\text{minimize}} & \langle G\,,\,Y\rangle\\ \text{subject to} & Ax\preceq b\\ & G=\left[\begin{array}{cc}X & x\\ x^T & 1\end{array}\right]\succeq 0\\ & \delta(X)=\mathbf{1}\end{array}\tag{668}$$

Solution of this semidefinite program is iterated with calculation of the direction matrix $Y$ from semidefinite program (652). At convergence, in the sense (633), convex problem (668) becomes equivalent to nonconvex Boolean problem (665).

By (1480a), direction matrix $Y$ can be an orthogonal projector having closed-form expression. Given randomized data $A$ and $b$ for a large problem, we find that stalling becomes likely (convergence of the iteration to a positive fixed point $\langle G^\star, Y\rangle$). To overcome this behavior, we introduce a heuristic into the implementation in §F.6 that momentarily reverses direction of search ($\approx -Y$) upon stall detection. We find that rate of convergence can be sped significantly by detecting stalls early. $\square$

**4.4.3.0.6 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:

$$x\|y\| = y \tag{669}$$

*id est*, $\|x\|=1$ and $x$ points in the same direction as $y$; *e.g.*,

$$\begin{array}{cl}\underset{x\,,\,y}{\text{minimize}} & f(x\,,\,y)\\ \text{subject to} & (x\,,\,y)\in\mathcal{C}\\ & x\|y\| = y\end{array} \tag{670}$$

where $f$ is some convex function and $\mathcal{C}$ is some convex set. We can realize the nonconvex equality by constraining rank and adding a regularization term to the objective. Make the assignment:

$$G = \begin{bmatrix} x\\ y\\ 1\end{bmatrix}[\,x^T \;\; y^T \;\; 1\,] = \begin{bmatrix} X & Z & x\\ Z & Y & y\\ x^T & y^T & 1\end{bmatrix} \triangleq \begin{bmatrix} xx^T & xy^T & x\\ yx^T & yy^T & y\\ x^T & y^T & 1\end{bmatrix} \in \mathbb{S}^{2N+1} \tag{671}$$

where $X\,,\,Y\in\mathbb{S}^N$, also $Z\in\mathbb{S}^N$ $[sic]$. Any rank-1 solution must take the form of (671). (§B.1) The problem statement equivalent to (670) is then written

$$\begin{array}{cl}\underset{X\,,\,Y\,,\,Z\,,\,x\,,\,y}{\text{minimize}} & f(x\,,\,y)\;+\;\|X-Y\|_{\mathrm{F}}\\ \text{subject to} & (x\,,\,y)\in\mathcal{C}\\ & G = \begin{bmatrix} X & Z & x\\ Z & Y & y\\ x^T & y^T & 1\end{bmatrix}\\ & \operatorname{rank} G = 1\\ & (G\succeq 0)\\ & \operatorname{tr}(X) = 1\\ & \delta(Z)\succeq 0\end{array} \tag{672}$$

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 $\|X-Y\|_{\mathrm{F}}$ then makes $x$ equal to $y$ to within a real scalar. (§C.2.0.0.1) To make this program solvable by convex iteration, as explained before, we move the rank constraint to the objective

$$\begin{array}{cl}\underset{X,Y,Z,x,y}{\text{minimize}} & f(x\,,\,y)\;+\;\|X-Y\|_{\mathrm F}\;+\;\langle G\,,\,Y\rangle\\ \text{subject to} & (x\,,\,y)\in\mathcal{C}\\ & G=\begin{bmatrix} X & Z & x\\ Z & Y & y\\ x^T & y^T & 1\end{bmatrix}\succeq 0\\ & \operatorname{tr}(X)=1\\ & \delta(Z)\succeq 0\end{array}\tag{673}$$

by introducing a direction matrix $Y$ found from (1480a)

$$\begin{array}{cl}\underset{Y\in\mathbb{S}^{2N+1}}{\text{minimize}} & \langle G^\star,\,Y\rangle\\ \text{subject to} & 0\preceq Y\preceq I\\ & \operatorname{tr}Y=2N\end{array}\tag{674}$$

whose optimal solution has closed form. Iteration (673) (674) terminates when $\operatorname{rank}G=1$ and regularization $\langle G\,,\,Y\rangle$ vanishes to within some numerical tolerance in (673); typically, in two iterations. If function $f$ competes too much with the regularization, positively weighting each regularization term will become required. At convergence, problem (673) becomes a convex equivalent to the original nonconvex problem (670). □

**4.4.3.0.7 Example.** FAST MAX CUT. [77]

> *Let $\Gamma$ be an $n$-node graph, and let the arcs $(i\,,\,j)$ of the graph be associated with [ ] weights $a_{ij}$. The problem is to find a cut of the largest possible weight, i.e., to partition the set of nodes into two parts $\mathcal{S},\mathcal{S}'$ in such a way that the total weight of all arcs linking $\mathcal{S}$ and $\mathcal{S}'$ (i.e., with one incident node in $\mathcal{S}$ and the other one in $\mathcal{S}'$) is as large as possible.* [27, §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 [83], quantum computing [295].

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.[4.27] MAX CUT is stated:

$$\begin{array}{cl}\underset{x}{\text{maximize}} & \sum\limits_{1\le i<j\le n} a_{ij}(1-x_i x_j)\frac{1}{2}\\ \text{subject to} & \delta(xx^T)=\mathbf{1}\end{array} \tag{675}$$

where $[a_{ij}]$ are real arc weights, and binary vector $x=[x_i]\in\mathbb{R}^n$ corresponds to the $n$ nodes; specifically,

$$\begin{array}{lcl}\text{node } i\in\mathcal{S} & \Leftrightarrow & x_i = \;\;\,1\\ \text{node } i\in\mathcal{S}' & \Leftrightarrow & x_i = -1\end{array} \tag{676}$$

If nodes $i$ and $j$ have the same binary value $x_i$ and $x_j$, then they belong to the same partition and contribute nothing to the cut. Arc $(i\,,j)$ traverses the cut, otherwise, adding its weight $a_{ij}$ to the cut.

MAX CUT statement (675) is the same as, for $A=[a_{ij}]\in\mathbb{S}^n$

$$\begin{array}{cl}\underset{x}{\text{maximize}} & \frac{1}{4}\langle\mathbf{1}\mathbf{1}^T - xx^T,\,A\rangle\\ \text{subject to} & \delta(xx^T)=\mathbf{1}\end{array} \tag{677}$$

Because of Boolean assumption $\delta(xx^T)=\mathbf{1}$

$$\langle\mathbf{1}\mathbf{1}^T - xx^T,\,A\rangle \;=\; \langle xx^T,\,\delta(A\mathbf{1})-A\rangle \tag{678}$$

so problem (677) is the same as

$$\begin{array}{cl}\underset{x}{\text{maximize}} & \frac{1}{4}\langle xx^T,\,\delta(A\mathbf{1})-A\rangle\\ \text{subject to} & \delta(xx^T)=\mathbf{1}\end{array} \tag{679}$$

This MAX CUT problem is combinatorial (nonconvex).

[4.27] We term our solution to MAX CUT *fast* because we sacrifice a little accuracy to achieve speed; *id est*, only about two or three convex iterations, achieved by heavily weighting a rank regularization term.

Because an estimate of upper bound to MAX CUT is needed to ascertain convergence when vector $x$ has large cardinality, we digress to derive the dual problem because it is instructive: Directly from (679), the *Lagrangian* is [46, §5.1.5] (1223)

$$\begin{aligned}\mathcal{L}(x\,,\,\nu) &= \tfrac{1}{4}\langle xx^T,\,\delta(A\mathbf{1})-A\rangle \,+\, \langle \nu\,,\,\delta(xx^T)-\mathbf{1}\rangle \\ &= \tfrac{1}{4}\langle xx^T,\,\delta(A\mathbf{1})-A\rangle \,+\, \langle \delta(\nu)\,,\,xx^T\rangle \,-\, \langle \nu\,,\,\mathbf{1}\rangle \\ &= \tfrac{1}{4}\langle xx^T,\,\delta(A\mathbf{1}+4\nu)-A\rangle \,-\, \langle \nu\,,\,\mathbf{1}\rangle\end{aligned} \tag{680}$$

where quadratic $x^T(\delta(A\mathbf{1}+4\nu)-A)x$ has supremum 0 if $\delta(A\mathbf{1}+4\nu)-A$ is negative semidefinite, and has supremum $\infty$ otherwise. The finite supremum of dual function

$$g(\nu) \,=\, \sup_x \mathcal{L}(x\,,\,\nu) \,=\, \begin{cases} -\langle \nu\,,\,\mathbf{1}\rangle\,, & A-\delta(A\mathbf{1}+4\nu)\succeq 0 \\ \infty & \text{otherwise}\end{cases} \tag{681}$$

is chosen to be the objective of minimization to dual convex problem

$$\begin{array}{ll}\underset{\nu}{\text{minimize}} & -\nu^T\mathbf{1} \\ \text{subject to} & A-\delta(A\mathbf{1}+4\nu)\succeq 0\end{array} \tag{682}$$

whose optimal value provides a least upper bound to MAX CUT, but is not tight (duality gap is nonzero). [108] In fact, we find that the bound's variance is relatively large for this problem; thus ending our digression.[4.28]

To transform MAX CUT to its convex equivalent, first define

$$X = xx^T \in \mathbb{S}^n \qquad (687)$$

then MAX CUT (679) becomes

$$\begin{array}{ll}\underset{X\in\,\mathbb{S}^n}{\text{maximize}} & \tfrac{1}{4}\langle X\,,\,\delta(A\mathbf{1})-A\rangle \\ \text{subject to} & \delta(X)=\mathbf{1} \\ & (X\succeq 0) \\ & \operatorname{rank} X = 1\end{array} \tag{683}$$

[4.28] Taking the dual of (682) would provide (683) but without the rank constraint. [101]

whose rank constraint can be regularized as in

$$\begin{array}{ll}\underset{X\in\mathbb{S}^n}{\text{maximize}} & \frac{1}{4}\langle X\,,\,\delta(A\mathbf{1})-A\rangle - w\langle X\,,\,W\rangle\\ \text{subject to} & \delta(X)=\mathbf{1}\\ & X\succeq 0\end{array} \tag{684}$$

where $w\approx 1000$ is a nonnegative fixed weight, and $W$ is a direction matrix determined from

$$\sum_{i=2}^{n}\lambda(X^\star)_i \;=\; \begin{array}{ll}\underset{W\in\mathbb{S}^n}{\text{minimize}} & \langle X^\star,\,W\rangle\\ \text{subject to} & 0\preceq W\preceq I\\ & \operatorname{tr}W=n-1\end{array} \qquad (1480\text{a})$$

whose optimal solution is known in closed form. These two problems (684) and (1480a) are iterated until convergence as defined on page 257.

Because convex problem statement (684) is so elegant, it is numerically solvable for large binary vectors within reasonable time.[4.29] 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 (679)[4.30] for a tight upper bound. Search-time limits binary vector lengths to 24 bits (about five days cpu time). Accuracy obtained, 98%, is independent of binary vector length (12, 13, 20, 24) when averaged over more than 231 problem instances including planar, randomized, and toroidal graphs.[4.31] A MATLAB program is provided in §F.7. That same accuracy is presumed maintained when binary vector length is further increased. □

---

[4.29] We solved for a length-250 binary vector in only a few minutes and convex iterations on a Dell Precision model M90.

[4.30] 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.

[4.31] Existence of a polynomial-time approximation to MAX CUT with accuracy better than 94.11% would refute proof of NP-hardness, which some researchers believe to be highly unlikely. [131, thm.8.2]

**4.4.3.0.8 Example.** *Cardinality/sparsity problem.*
d'Aspremont *et alii* [65] propose approximating a positive semidefinite matrix $A\in\mathbb{S}_+^N$ by a rank-one matrix having a constraint on cardinality $c$ : for $0<c<N$

$$\begin{array}{cl}\underset{z}{\text{minimize}} & \|A-zz^T\|_{\mathrm F}\\ \text{subject to} & \operatorname{card} z\leq c\end{array}\tag{685}$$

which, they explain, is a hard problem equivalent to

$$\begin{array}{cl}\underset{x}{\text{maximize}} & x^TA\,x\\ \text{subject to} & \|x\|=1\\ & \operatorname{card} x\leq c\end{array}\tag{686}$$

where $z\triangleq\sqrt{\lambda}\,x$ and where optimal solution $x^\star$ is a *principal eigenvector* (1474) (§A.5) of $A$ and $\lambda=x^{\star T}\!A\,x^\star$ is the *principal eigenvalue* 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

$$X\triangleq xx^T\in\mathbb{S}^N\tag{687}$$

whose desired rank is 1, and whose desired diagonal cardinality

$$\operatorname{card}\delta(X)\equiv\operatorname{card} x\tag{688}$$

is equivalent to cardinality $c$ of vector $x$ . Then we can transform cardinality problem (686) to an equivalent problem in new variable $X$ :[4.32]

$$\begin{array}{cl}\underset{X\in\mathbb{S}^N}{\text{maximize}} & \langle X\,,\,A\rangle\\ \text{subject to} & \langle X\,,\,I\rangle=1\\ & (X\succeq 0)\\ & \operatorname{rank} X=1\\ & \operatorname{card}\delta(X)\leq c\end{array}\tag{689}$$

We transform problem (689) to an equivalent convex problem by introducing two direction matrices: $W$ to achieve desired cardinality

[4.32] 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)

$\operatorname{card}\delta(X)$, and direction matrix $Y$ to find an approximating rank-one matrix $X$:

$$\begin{array}{cl}\underset{X\in\mathbb{S}^N}{\text{maximize}} & \langle X\,,\,A-w_1Y\rangle-w_2\langle\delta(X)\,,\,\delta(W)\rangle\\ \text{subject to} & \langle X\,,\,I\rangle=1\\ & X\succeq 0\end{array}\tag{690}$$

where $w_1$ and $w_2$ are positive scalars respectively weighting $\operatorname{tr}(XY)$ 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

$$\begin{array}{cl}\underset{Y\in\mathbb{S}^N}{\text{minimize}} & \langle X^\star,\,Y\rangle\\ \text{subject to} & 0\preceq Y\preceq I\\ & \operatorname{tr}Y=N-1\end{array}\tag{691}$$

and where diagonal direction matrix $W\in\mathbb{S}^N$ optimally solves linear program

$$\begin{array}{cl}\underset{W=\delta^2(W)}{\text{minimize}} & \langle\delta(X^\star)\,,\,\delta(W)\rangle\\ \text{subject to} & 0\preceq\delta(W)\preceq\mathbf{1}\\ & \operatorname{tr}W=N-c\end{array}\tag{692}$$

both direction matrix programs being derived from (1480a) whose analytical solution is known. We emphasize (*confer* p.257): because this iteration (690) (691) (692) (initial $Y,W=\mathbf{0}$) is not a projection method, success relies on existence of matrices in the feasible set of (690) having desired rank and diagonal cardinality. In particular, the feasible set of convex problem (690) is a Fantope (80) 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 (690) is neither a relaxation of cardinality problem (686); instead, problem (690) is a convex equivalent to (686) at convergence of iteration (690) (691) (692). Because the feasible set of convex problem (690) contains all normalized rank-one matrices of any desired diagonal cardinality, a constraint too low or high in cardinality will not prevent solution. An optimal solution, whose diagonal cardinality is equal to cardinality of a principal eigenvector of matrix $A$, will produce the lowest residual Frobenius norm (to within machine precision and other noise processes) in the original problem statement (685). $\square$

Figure 72: Massachusetts Institute of Technology (MIT) logo, including its white boundary, may be interpreted as a rank-5 matrix. (Stanford University logo rank is much higher;) This constitutes *Scene Y* observed by the one-pixel camera in Figure **73** for Example 4.4.3.0.9.

**4.4.3.0.9 Example.** *Compressed sensing.* [223]
*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 [81]

Lossy data compression techniques are popular, but it is also well known that losses become quite perceptible with signal processing that goes beyond mere playback of a compressed signal. Spatial or audible frequencies masked by a simultaneity become perceptible with significant post-filtering of the compressed signal, for example. Further, there can be no universally acceptable and unique metric of perception for gauging exactly how much data can be tossed. For these reasons, there will always be need for raw (uncompressed) data.

In this example we throw out only so much information as to leave perfect reconstruction within reach. Specifically, the MIT logo in Figure **72** is perfectly reconstructed from 700 time-sequential samples $\{y_i\}$ acquired by the one-pixel camera illustrated in Figure **73**. The MIT-logo image in this example effectively impinges a $46\times81$ array micromirror DMD.

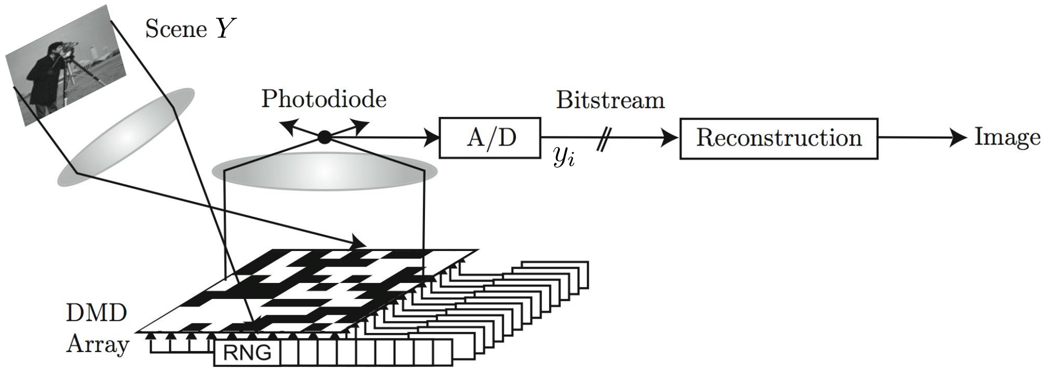



Figure 73: One-pixel camera from [279]. *Compressive imaging (CI) camera block diagram. Incident lightfield (corresponding to the desired image $Y$) is reflected off a digital micromirror device (DMD) array whose mirror orientations are modulated in the pseudorandom pattern supplied by the random number generators (RNG). Each different mirror pattern produces a voltage at the single photodiode that corresponds to one measurement $y_i$ .*

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 $y=[y_i]\in\mathbb{R}^{\mathbf{700}}$ from this one-pixel camera. That number represents only 19% of total information from the micromirrors.[4.33]

[4.33] This number is considered difficult to achieve judging from results reported in [223, §6]. If a minimal basis for the MIT logo were instead constructed, only five rows or columns worth of data are independent. This means a lower bound on achievable compression is about 230 samples; which corresponds to 6% of the original information.

Our approach to reconstruction is to look for low-rank solution to an *underdetermined* system:

$$\begin{array}{rl}\text{find} & X\\ \text{subject to} & A\operatorname{vec}X = y\\ & \operatorname{rank}X \leq 5\end{array} \tag{693}$$

where $\operatorname{vec}X$ is vectorized matrix $X\in\mathbb{R}^{\mathbf{46}\times\mathbf{81}}$ (stacked columns). 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 $i^{\text{th}}$ sample $y_i$ equals $A(i\,,:)\operatorname{vec}Y$ ; *id est*, $y = A\operatorname{vec}Y$ . Perfect reconstruction means optimal solution $X^\star$ equals scene $Y\in\mathbb{R}^{\mathbf{46}\times\mathbf{81}}$ to within machine precision.

Because variable matrix $X$ is generally not square or positive semidefinite, we constrain its rank by rewriting the problem equivalently

$$\begin{array}{rl}\text{find} & X\\ \text{subject to} & A\operatorname{vec}X = y\\ & \operatorname{rank}\left[\begin{array}{cc}W_1 & X\\ X^T & W_2\end{array}\right] \leq 5\end{array} \tag{694}$$

This rank constraint on the composite matrix insures $\operatorname{rank}X\leq 5$ 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

$$\begin{array}{rl}\underset{W_1\,,\,W_2\,,\,X}{\text{minimize}} & \left[\begin{array}{cc}W_1 & X\\ X^T & W_2\end{array}\right] Z\\ \text{subject to} & A\operatorname{vec}X = y\\ & \left[\begin{array}{cc}W_1 & X\\ X^T & W_2\end{array}\right] \succeq 0\end{array} \tag{695}$$

with semidefinite program

$$\begin{array}{rl}\underset{Z}{\text{minimize}} & \left[\begin{array}{cc}W_1 & X\\ X^T & W_2\end{array}\right]^\star Z\\ \text{subject to} & 0 \preceq Z \preceq I\\ & \operatorname{tr}Z = 46 + 81 - 5\end{array} \tag{696}$$

(whose solution has closed form, p.541) until a rank-5 composite matrix is found. With 1000 samples $\{y_i\}$ , convergence occurs in two iterations;

700 samples require more than ten, but reconstruction remains perfect. Reconstruction is independent of pseudorandom sequence parameters; *e.g.*, binary sequences succeed with the same efficiency as Gaussian or uniformly distributed sequences. □

### 4.4.4 rank-constraint conclusion

We find that this *direction matrix* idea works well and quite independently of desired rank or affine dimension.

There exists a common thread through all these Examples: that being, convex iteration with a direction matrix as normal to a linear regularization. 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 5.4.2.2.3 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 this chapter require an initial value of $\mathbf{0}$. The *sparsest solution* Example 4.4.3.0.1 wants a direction matrix corresponding to a rank-2 search when a rank-1 solution is desired. Finding a feasible Boolean vector in Example 4.4.3.0.5 requires a procedure to detect stalls, when other problems have no such requirement; and so on.

Nevertheless, this idea of direction matrix is good because of its simplicity: When one is confronted with a problem otherwise convex if not for a rank constraint, then that constraint becomes a linear regularization term in the objective. Some work remains in refining initial value of the direction matrix in the regularization because poor initialization of the convex iteration can lead to an erroneous result.

# Chapter 5

# Euclidean Distance Matrix

*These results were obtained by Schoenberg (1935), a surprisingly late date for such a fundamental property of Euclidean geometry.*

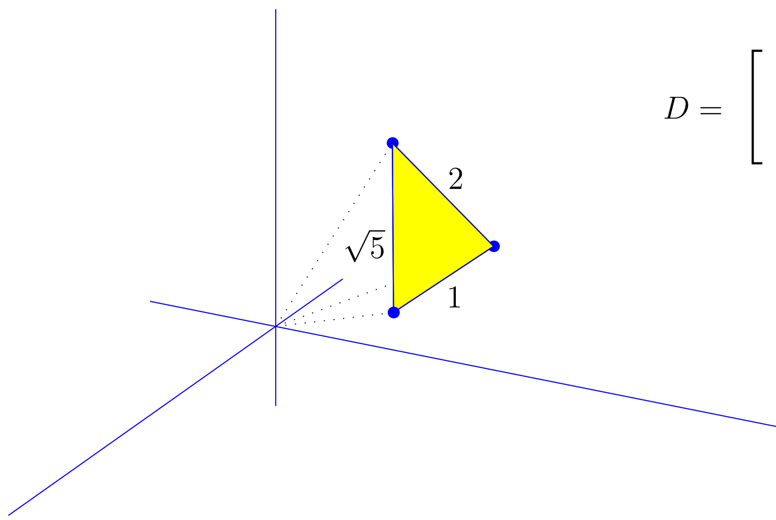
−John Clifford Gower [112, §3]

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) [69] 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.*, $\sqrt[\circ]{D}\in\mathbb{R}^{N\times N}$, 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.



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.




$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ D = & \left[\begin{array}{ccc} 0 & 1 & 5 \\ 1 & 0 & 4 \\ 5 & 4 & 0 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

Figure 74: Convex hull of three points ($N = 3$) is shaded in $\mathbb{R}^3$ ($n = 3$). Dotted lines are imagined vectors to points.

## 5.1 EDM

Euclidean space $\mathbb{R}^n$ is a finite-dimensional real vector space having an inner product defined on it, hence a metric as well. [166, §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:

$$d_{ij} \;=\; \|x_i - x_j\|_2^2 \;\triangleq\; \langle x_i - x_j\ ,\ x_i - x_j\rangle \tag{697}$$

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 $N=3$:

$$D \;=\; [d_{ij}] \;=\; \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix} = \begin{bmatrix} 0 & d_{12} & d_{13} \\ d_{12} & 0 & d_{23} \\ d_{13} & d_{23} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 5 \\ 1 & 0 & 4 \\ 5 & 4 & 0 \end{bmatrix} \tag{698}$$

Matrix $D$ has $N^2$ entries but only $N(N-1)/2$ pieces of information. In Figure 74 are shown three points in $\mathbb{R}^3$ that can be arranged in a list to correspond to $D$ in (698). Such a list is not unique because any rotation, reflection, or translation (§5.5) of the points in Figure **74** would produce the same EDM $D$.

## 5.2 First metric properties

For $i,j=1\ldots N$, the Euclidean distance between points $x_i$ and $x_j$ must satisfy the requirements imposed by any metric space: [166, §1.1] [189, §1.7]

1. $\sqrt{d_{ij}} \geq 0\,,\quad i \neq j$ — nonnegativity
2. $\sqrt{d_{ij}} = 0\,,\quad i = j$ — self-distance
3. $\sqrt{d_{ij}} = \sqrt{d_{ji}}$ — symmetry
4. $\sqrt{d_{ij}} \;\leq\; \sqrt{d_{ik}} + \sqrt{d_{kj}}\;,\quad i\neq j\neq k$ — triangle inequality

where $\sqrt{d_{ij}}$ is the Euclidean metric in $\mathbb{R}^n$ (§5.4). Then all entries of an EDM must be in concord with these Euclidean metric properties: specifically, each entry must be nonnegative,[5.2] the main diagonal must be $\mathbf{0}$,[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 ∃ 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. [112, §2] Yet any list of points or the vertices of any bounded convex polyhedron must conform to the properties.

---

[5.2] Implicit from the terminology, $\sqrt{d_{ij}} \geq 0 \Leftrightarrow d_{ij} \geq 0$ is always assumed.

[5.3] What we call self-distance, Marsden calls *nondegeneracy.* [189, §1.6] Kreyszig calls these first metric properties *axioms of the metric*; [166, p.4] Blumenthal refers to them as *postulates.* [37, p.15]

**5.3.0.0.1 Example.** *Triangle.*
Consider the EDM in (698), but missing one of its entries:

$$D = \begin{bmatrix} 0 & 1 & d_{13} \\ 1 & 0 & 4 \\ d_{31} & 4 & 0 \end{bmatrix} \tag{699}$$

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 $\mathbf{0}$, while property 3 makes $\sqrt{d_{31}} = \sqrt{d_{13}}$ . The fourth property tells us

$$1 \leq \sqrt{d_{13}} \leq 3 \tag{700}$$

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. □

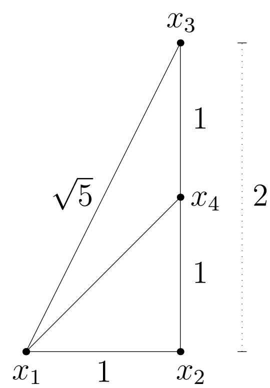
**5.3.0.0.2 Example.** *Small completion problem*, I.
Now consider the polyhedron in Figure **75**(b) formed from an unknown list $\{x_1, x_2, x_3, x_4\}$. The corresponding EDM less one critical piece of information, $d_{14}$ , is given by

$$D = \begin{bmatrix} 0 & 1 & 5 & d_{14} \\ 1 & 0 & 4 & 1 \\ 5 & 4 & 0 & 1 \\ d_{14} & 1 & 1 & 0 \end{bmatrix} \tag{701}$$

From metric property 4 we may write a few inequalities for the two triangles common to $d_{14}$ ; we find

$$\sqrt{5} - 1 \;\leq\; \sqrt{d_{14}} \;\leq\; 2 \tag{702}$$

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. □



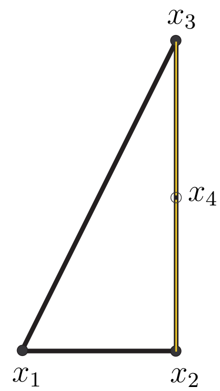




Figure 75: **(a)** Complete dimensionless EDM graph. **(b)** Emphasizing obscured segments $\overline{x_2 x_4}$ , $\overline{x_4 x_3}$ , and $\overline{x_2 x_3}$ , now only five $(2N-3)$ absolute distances are specified. EDM so represented is incomplete, missing $d_{14}$ as in (701), yet the isometric reconstruction (§5.4.2.2.5) 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.

We will return to this simple Example 5.3.0.0.2 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.

---

[5.4]The term *tight* with reference to an inequality means equality is achievable.

### 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

$$\begin{array}{c} \cos(\theta_{ik\ell}+\theta_{\ell kj}) \;\leq\; \cos\theta_{ikj} \;\leq\; \cos(\theta_{ik\ell}-\theta_{\ell kj}) \\ 0\leq \theta_{ik\ell}\,,\theta_{\ell kj}\,,\theta_{ikj}\leq \pi \end{array} \tag{703}$$

where $\theta_{ikj}=\theta_{jki}$ is the angle between vectors at vertex $x_k$ (770) (Figure **76**), 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 (1041) (728) (733)

$$\left.\begin{array}{r} -z^TDz \geq 0 \\ \mathbf{1}^Tz = 0 \\ (\forall\, \|z\| = 1) \\ D \in \mathbb{S}_h^N \end{array}\right. \;\Leftrightarrow\; D \in \mathbb{EDM}^N \tag{704}$$

where the convex cone of Euclidean distance matrices $\mathbb{EDM}^N \subseteq \mathbb{S}_h^N$ belongs to the subspace of symmetric hollow[5.5] matrices (§2.2.3.0.1). Having found equivalent matrix criteria, we will see there is a bridge from bounded convex polyhedra to EDMs in §5.9 .[5.6]

[5.5] $\mathbf{0}$ 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).

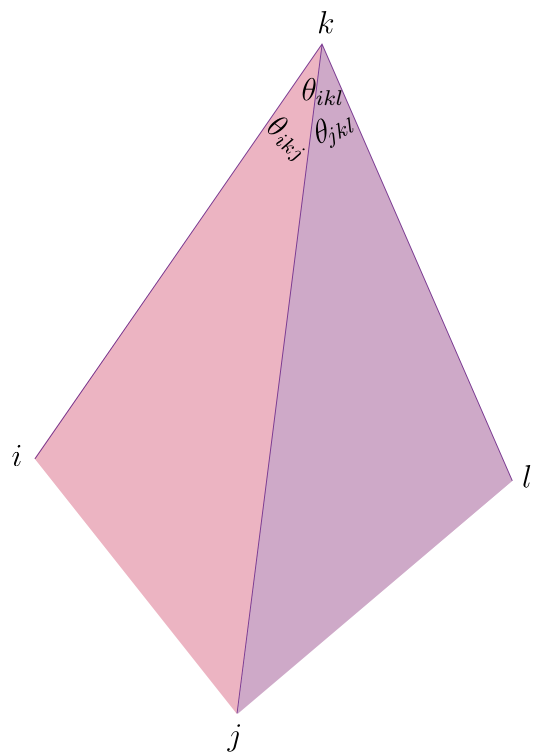



Figure 76: Nomenclature for fifth Euclidean metric property. Each angle $\theta$ is made by a vector pair at vertex $k$ while $i, j, k, l$ index four points. The fifth property is necessary for realization of four or more points reckoned by three angles in any dimension.

Now we develop some invaluable concepts, moving toward a link of the Euclidean metric properties to matrix criteria.

## 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

$$X = [\, x_1 \ \cdots \ x_N \,] \in \mathbb{R}^{n\times N} \qquad (65)$$

where $N$ is regarded as *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.23)

$$\begin{aligned} d_{ij} \;&=\; \|x_i - x_j\|^2 \;=\; (x_i - x_j)^T (x_i - x_j) \;=\; \|x_i\|^2 + \|x_j\|^2 - 2x_i^T x_j \\ &=\; \begin{bmatrix} x_i^T & x_j^T \end{bmatrix} \begin{bmatrix} I & -I \\ -I & I \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix} \\ &=\; \operatorname{vec}(X)^T (\Phi_{ij} \otimes I) \operatorname{vec} X \;=\; \langle \Phi_{ij} \,,\; X^T X \rangle \end{aligned} \qquad (705)$$

where

$$\operatorname{vec} X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^{nN} \qquad (706)$$

and where $\Phi_{ij} \otimes I$ has $I \in \mathbb{S}^n$ in its $ii^{\text{th}}$ and $jj^{\text{th}}$ block of entries while $-I \in \mathbb{S}^n$ fills its $ij^{\text{th}}$ and $ji^{\text{th}}$ block; *id est*,

$$\begin{aligned} \Phi_{ij} &\triangleq \delta\big((e_i e_j^T + e_j e_i^T)\mathbf{1}\big) - (e_i e_j^T + e_j e_i^T) \in \mathbb{S}_+^N \\ &= e_i e_i^T + e_j e_j^T - e_i e_j^T - e_j e_i^T \\ &= (e_i - e_j)(e_i - e_j)^T \end{aligned} \qquad (707)$$

where $\{e_i \in \mathbb{R}^N,\ i = 1 \ldots N\}$ is the set of standard basis vectors, and $\otimes$ signifies the Kronecker product (§D.1.2.1). Thus each entry $d_{ij}$ is a convex quadratic function [46, §3, §4] of $\operatorname{vec} X$ (30). [230, §6]

The collection of all Euclidean distance matrices $\mathbb{EDM}^N$ is a convex subset of $\mathbb{R}_+^{N\times N}$ called the *EDM cone* (§6, Figure **109**, p.442);

$$\mathbf{0} \in \mathbb{EDM}^N \subseteq \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N} \subset \mathbb{S}^N \tag{708}$$

An EDM $D$ must be expressible as a function of some list $X$; *id est*, it must have the form

$$\mathbf{D}(X) \triangleq \delta(X^TX)\mathbf{1}^T + \mathbf{1}\delta(X^TX)^T - 2X^TX \in \mathbb{EDM}^N \tag{709}$$
$$= [\operatorname{vec}(X)^T(\Phi_{ij} \otimes I)\operatorname{vec} X\ ,\quad i,j=1\ldots N] \tag{710}$$

Function $\mathbf{D}(X)$ will make an EDM given any $X\in\mathbb{R}^{n\times N}$, conversely, but $\mathbf{D}(X)$ is not a convex function of $X$ (§5.4.1). Now the EDM cone may be described:

$$\mathbb{EDM}^N = \left\{\mathbf{D}(X) \mid X\in\mathbb{R}^{N-1\times N}\right\} \tag{711}$$

Expression $\mathbf{D}(X)$ 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 (705), so holds for any $D$ expressible in the form $\mathbf{D}(X)$ in (709).

When we say $D$ is an EDM, reading from (709), it implicitly means the main diagonal must be $\mathbf{0}$ (property 2, self-distance) and $D$ must be symmetric (property 3); $\delta(D)=\mathbf{0}$ and $D^T=D$ or, equivalently, $D\in\mathbb{S}_h^N$ are necessary matrix criteria.

#### 5.4.0.1 homogeneity

Function $\mathbf{D}(X)$ is homogeneous in the sense, for $\zeta\in\mathbb{R}$

$$\sqrt[\circ]{\mathbf{D}(\zeta X)} = |\zeta|\sqrt[\circ]{\mathbf{D}(X)} \tag{712}$$

where the positive square root is entrywise.

Any nonnegatively scaled EDM remains an EDM; *id est*, the matrix class EDM is invariant to nonnegative scaling ($\alpha\,\mathbf{D}(X)$ for $\alpha\geq 0$) because all EDMs of dimension $N$ constitute a convex cone $\mathbb{EDM}^N$ (§6, Figure **96**).

### 5.4.1 $-V_{\mathcal{N}}^T\mathbf{D}(X)V_{\mathcal{N}}$ convexity

We saw that EDM entries $d_{ij}\left(\begin{bmatrix} x_i \\ x_j \end{bmatrix}\right)$ are convex quadratic functions. Yet $-\mathbf{D}(X)$ (709) is not a quasiconvex function of matrix $X\in\mathbb{R}^{n\times N}$ because the second directional derivative (§3.3)

$$-\frac{d^2}{dt^2}\bigg|_{t=0}\mathbf{D}(X+t\,Y) = 2\big(-\delta(Y^TY)\mathbf{1}^T - \mathbf{1}\delta(Y^TY)^T + 2\,Y^TY\big) \tag{713}$$

is indefinite for any $Y\in\mathbb{R}^{n\times N}$ since its main diagonal is $\mathbf{0}$. [110, §4.2.8] [150, §7.1, prob.2] Hence $-\mathbf{D}(X)$ can neither be convex in $X$.

The outcome is different when instead we consider

$$-V_{\mathcal{N}}^T\mathbf{D}(X)V_{\mathcal{N}} = 2V_{\mathcal{N}}^TX^TXV_{\mathcal{N}} \tag{714}$$

where we introduce the full-rank skinny *Schoenberg auxiliary matrix* (§B.4.2)

$$V_{\mathcal{N}} \triangleq \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & & & \mathbf{0} \\ & 1 & & \\ & & \ddots & \\ \mathbf{0} & & & 1 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} -\mathbf{1}^T \\ I \end{bmatrix} \in\mathbb{R}^{N\times N-1} \tag{715}$$

$(\mathcal{N}(V_{\mathcal{N}})=\mathbf{0})$ having range

$$\mathcal{R}(V_{\mathcal{N}}) = \mathcal{N}(\mathbf{1}^T)\;, \quad V_{\mathcal{N}}^T\mathbf{1} = \mathbf{0} \tag{716}$$

Matrix-valued function (714) meets the criterion for convexity in §3.2.3.0.2 over its domain that is all of $\mathbb{R}^{n\times N}$; *videlicet*, for any $Y\in\mathbb{R}^{n\times N}$

$$-\frac{d^2}{dt^2}V_{\mathcal{N}}^T\mathbf{D}(X+\,t\,Y)V_{\mathcal{N}} \;=\; 4V_{\mathcal{N}}^TY^TYV_{\mathcal{N}} \succeq 0 \tag{717}$$

Quadratic matrix-valued function $-V_{\mathcal{N}}^T\mathbf{D}(X)V_{\mathcal{N}}$ is therefore convex in $X$ achieving its minimum, with respect to a positive semidefinite cone (§2.7.2.2), at $X=\mathbf{0}$. When the penultimate number of points exceeds the dimension of the space $n<N-1$, strict convexity of the quadratic (714) becomes impossible because (717) could not then be positive definite.

### 5.4.2 Gram-form EDM definition

Positive semidefinite matrix $X^TX$ in (709), formed from inner product of the list, is known as a *Gram matrix*; [182, §3.6]

$$G \triangleq X^TX = \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix} [\, x_1 \cdots x_N \,] = \begin{bmatrix} \|x_1\|^2 & x_1^Tx_2 & x_1^Tx_3 & \cdots & x_1^Tx_N \\ x_2^Tx_1 & \|x_2\|^2 & x_2^Tx_3 & \cdots & x_2^Tx_N \\ x_3^Tx_1 & x_3^Tx_2 & \|x_3\|^2 & \ddots & x_3^Tx_N \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ x_N^Tx_1 & x_N^Tx_2 & x_N^Tx_3 & \cdots & \|x_N\|^2 \end{bmatrix} \in \mathbb{S}_+^N$$

$$= \delta\left(\begin{bmatrix} \|x_1\| \\ \|x_2\| \\ \vdots \\ \|x_N\| \end{bmatrix}\right) \begin{bmatrix} 1 & \cos\psi_{12} & \cos\psi_{13} & \cdots & \cos\psi_{1N} \\ \cos\psi_{12} & 1 & \cos\psi_{23} & \cdots & \cos\psi_{2N} \\ \cos\psi_{13} & \cos\psi_{23} & 1 & \ddots & \cos\psi_{3N} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \cos\psi_{1N} & \cos\psi_{2N} & \cos\psi_{3N} & \cdots & 1 \end{bmatrix} \delta\left(\begin{bmatrix} \|x_1\| \\ \|x_2\| \\ \vdots \\ \|x_N\| \end{bmatrix}\right)$$

$$\triangleq \sqrt{\delta^2(G)}\, \Psi \sqrt{\delta^2(G)} \tag{718}$$

where $\psi_{ij}$ (738) 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 $G$ ; [46, §8.3]

$$G \succeq 0 \Leftarrow \Psi \succeq 0 \tag{719}$$

When $\delta^2(G)$ is nonsingular, then $G \succeq 0 \Leftrightarrow \Psi \succeq 0$. (§A.3.1.0.5)

Distance-square $d_{ij}$ (705) is related to Gram matrix entries $G^T = G \triangleq [g_{ij}]$

$$\begin{aligned} d_{ij} &= g_{ii} + g_{jj} - 2g_{ij} \\ &= \langle \Phi_{ij} \,,\, G \rangle \end{aligned} \tag{720}$$

where $\Phi_{ij}$ is defined in (707). Hence the linear EDM definition

$$\left.\begin{aligned} \mathbf{D}(G) &\triangleq \delta(G)\mathbf{1}^T + \mathbf{1}\delta(G)^T - 2G \;\in \mathbb{EDM}^N \\ &= [\langle \Phi_{ij} \,,\, G \rangle \,,\; i,j=1 \ldots N] \end{aligned}\right\} \quad\Leftarrow\quad G \succeq 0 \tag{721}$$

The EDM cone may be described, (*confer* (798)(804))

$$\mathbb{EDM}^N = \left\{ \mathbf{D}(G) \mid G \in \mathbb{S}_+^N \right\} \tag{722}$$

#### 5.4.2.1 First point at origin

Assume the first point $x_1$ in an unknown list $X$ resides at the origin;

$$Xe_1 = \mathbf{0} \;\Leftrightarrow\; Ge_1 = \mathbf{0} \tag{723}$$

Consider the symmetric translation $(I - \mathbf{1}e_1^T)\mathbf{D}(G)(I - e_1\mathbf{1}^T)$ that shifts the first row and column of $\mathbf{D}(G)$ to the origin; setting Gram-form EDM operator $\mathbf{D}(G) = D$ for convenience,

$$-\big(D - (De_1\mathbf{1}^T + \mathbf{1}e_1^T D) + \mathbf{1}e_1^T De_1\mathbf{1}^T\big)\tfrac{1}{2} = G - (Ge_1\mathbf{1}^T + \mathbf{1}e_1^T G) + \mathbf{1}e_1^T Ge_1\mathbf{1}^T \tag{724}$$

where

$$e_1 \triangleq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{725}$$

is the first vector from the standard basis. Then it follows for $D \in \mathbb{S}_h^N$

$$\begin{aligned} G &= -\big(D - (De_1\mathbf{1}^T + \mathbf{1}e_1^T D)\big)\tfrac{1}{2}\;, \qquad && x_1 = \mathbf{0} \\ &= -\begin{bmatrix} \mathbf{0} & \sqrt{2}V_{\mathcal{N}} \end{bmatrix}^T D \begin{bmatrix} \mathbf{0} & \sqrt{2}V_{\mathcal{N}} \end{bmatrix}\tfrac{1}{2} \\ &= \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_{\mathcal{N}}^T DV_{\mathcal{N}} \end{bmatrix} \\ V_{\mathcal{N}}^T GV_{\mathcal{N}} &= -V_{\mathcal{N}}^T DV_{\mathcal{N}}\tfrac{1}{2} && \forall\, X \end{aligned} \tag{726}$$

where

$$I - e_1\mathbf{1}^T = \begin{bmatrix} \mathbf{0} & \sqrt{2}V_{\mathcal{N}} \end{bmatrix} \tag{727}$$

is a projector nonorthogonally projecting (§E.1) on

$$\begin{aligned} \mathbb{S}_1^N &= \{G \in \mathbb{S}^N \mid Ge_1 = \mathbf{0}\} \\ &= \Big\{\begin{bmatrix} \mathbf{0} & \sqrt{2}V_{\mathcal{N}} \end{bmatrix}^T Y \begin{bmatrix} \mathbf{0} & \sqrt{2}V_{\mathcal{N}} \end{bmatrix} \mid Y \in \mathbb{S}^N\Big\} \end{aligned} \tag{1770}$$

in the Euclidean sense. From (726) we get sufficiency of the first matrix criterion for an EDM proved by Schoenberg in 1935; [234][5.7]

[5.7]From (716) we know $\mathcal{R}(V_{\mathcal{N}}) = \mathcal{N}(\mathbf{1}^T)$, so (728) is the same as (704). In fact, any matrix $V$ in place of $V_{\mathcal{N}}$ will satisfy (728) whenever $\mathcal{R}(V) = \mathcal{R}(V_{\mathcal{N}}) = \mathcal{N}(\mathbf{1}^T)$. But $V_{\mathcal{N}}$ is the matrix implicit in Schoenberg's seminal exposition.

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{ \begin{array}{c} -V_{\mathcal{N}}^T D V_{\mathcal{N}} \in \mathbb{S}_+^{N-1} \\ D \in \mathbb{S}_h^N \end{array} \right. \tag{728}$$

We provide a rigorous complete more geometric proof of this *Schoenberg criterion* in §5.9.1.0.3.

By substituting $G = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_{\mathcal{N}}^T D V_{\mathcal{N}} \end{bmatrix}$ (726) into $\mathbf{D}(G)$ (721), assuming $x_1 = \mathbf{0}$

$$D = \begin{bmatrix} 0 \\ \delta\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big) \end{bmatrix} \mathbf{1}^T + \mathbf{1} \begin{bmatrix} 0 & \delta\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big)^T \end{bmatrix} - 2 \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_{\mathcal{N}}^T D V_{\mathcal{N}} \end{bmatrix} \tag{729}$$

We provide details of this bijection 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;

$$X\mathbf{1} = \mathbf{0} \;\Leftrightarrow\; G\mathbf{1} = \mathbf{0} \tag{730}$$

Now consider the calculation $(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)\mathbf{D}(G)(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T)$, a geometric centering or projection operation. (§E.7.2.0.2) Setting $\mathbf{D}(G) = D$ for convenience as in §5.4.2.1,

$$\begin{aligned} G &= -\big(D - \tfrac{1}{N}(D\mathbf{1}\mathbf{1}^T + \mathbf{1}\mathbf{1}^T D) + \tfrac{1}{N^2}\mathbf{1}\mathbf{1}^T D \mathbf{1}\mathbf{1}^T\big)\tfrac{1}{2}\,, \quad & X\mathbf{1} = \mathbf{0} \\ &= -VDV\tfrac{1}{2} \\ VGV &= -VDV\tfrac{1}{2} & \forall X \end{aligned} \tag{731}$$

where more properties of the auxiliary (geometric centering, projection) matrix

$$V \triangleq I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{S}^N \tag{732}$$

are found in §B.4. From (731) and the assumption $D \in \mathbb{S}_h^N$ we get sufficiency of the more popular form of Schoenberg's criterion:

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{\begin{array}{c} -VDV \in \mathbb{S}_+^N \\ D \in \mathbb{S}_h^N \end{array}\right. \tag{733}$$

Of particular utility when $D\in\mathbb{EDM}^N$ is the fact, (§B.4.2 no.20) (705)

$$\begin{aligned}
\operatorname{tr}\!\left(-VDV\tfrac{1}{2}\right) &= \tfrac{1}{2N}\sum_{i,j} d_{ij} & &= \tfrac{1}{2N}\operatorname{vec}(X)^T\left(\sum_{i,j}\Phi_{ij}\otimes I\right)\operatorname{vec}X \\
&= \operatorname{tr}(VGV)\;, \quad G\succeq 0 & & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (734) \\
&= \operatorname{tr}G & &= \sum_{\ell=1}^N \|x_\ell\|^2 \;=\; \|X\|_{\mathrm F}^2 \;, \qquad X\mathbf{1}=\mathbf{0}
\end{aligned}$$

where $\sum\Phi_{ij}\in\mathbb{S}_+^N$ (707), therefore convex in $\operatorname{vec}X$. 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 $G=-VDV\frac{1}{2}$ (731) into $\mathbf{D}(G)$ (721), assuming $X\mathbf{1}=\mathbf{0}$ (*confer* §5.6.1)

$$D = \delta\!\left(-VDV\tfrac{1}{2}\right)\mathbf{1}^T + \mathbf{1}\delta\!\left(-VDV\tfrac{1}{2}\right)^T - 2\!\left(-VDV\tfrac{1}{2}\right) \tag{735}$$

These relationships will allow combination of distance and Gram constraints in any optimization problem we may pose:

- Constraining all main diagonal entries of a Gram matrix to 1, for example, is equivalent to the constraint that all points lie on a hypersphere (§5.9.1.0.2) of radius 1 centered at the origin. This is equivalent to the EDM constraint: $D\mathbf{1}=2N\mathbf{1}$. [61, p.116] Any further constraint on that Gram matrix then applies only to interpoint angle $\Psi$.

- More generally, interpoint angle $\Psi$ can be constrained by fixing all the individual point lengths $\delta(G)^{1/2}$; then

$$\Psi = -\frac{1}{2}\delta^2(G)^{-1/2}VDV\delta^2(G)^{-1/2} \tag{736}$$

**5.4.2.2.1 Example.** *List member constraints via Gram matrix.*
Capitalizing on identity (731) relating Gram and EDM $D$ matrices, a constraint set such as

$$\left.\begin{array}{rcl}\operatorname{tr}\!\left(-\tfrac{1}{2}VDVe_ie_i^T\right) &=& \|x_i\|^2\\ \operatorname{tr}\!\left(-\tfrac{1}{2}VDV(e_ie_j^T+e_je_i^T)\tfrac{1}{2}\right) &=& x_i^Tx_j\\ \operatorname{tr}\!\left(-\tfrac{1}{2}VDVe_je_j^T\right) &=& \|x_j\|^2\end{array}\right\} \tag{737}$$

relates list member $x_i$ to $x_j$ to within an isometry through inner-product identity [288, §1-7]

$$\cos\psi_{ij} = \frac{x_i^Tx_j}{\|x_i\|\,\|x_j\|} \tag{738}$$

For $M$ list members, there are a total of $M(M+1)/2$ such constraints.
□

Consider the academic problem of finding a Gram matrix subject to constraints on each and every entry of the corresponding EDM:

$$\begin{array}{cl}\underset{D\in\mathbb{S}_h^N}{\text{find}} & -VDV\tfrac{1}{2}\in\mathbb{S}^N\\ \text{subject to} & \left\langle D\,,\,(e_ie_j^T+e_je_i^T)\tfrac{1}{2}\right\rangle = \check{d}_{ij}\;, \quad i,j\!=\!1\ldots N\,, \quad i<j\\ & -VDV\succeq 0\end{array} \tag{739}$$

where the $\check{d}_{ij}$ are given nonnegative constants. EDM $D$ can, of course, be replaced with the equivalent Gram-form (721). Requiring only the self-adjointness property (1223) of the main-diagonal linear operator $\delta$ we get, for $A\in\mathbb{S}^N$

$$\langle D\,,\,A\rangle \;=\; \left\langle\delta(G)\mathbf{1}^T+\mathbf{1}\delta(G)^T-2G\,,\,A\right\rangle \;=\; \langle G\,,\,\delta(A\mathbf{1})-A\rangle\,2 \tag{740}$$

Then the problem equivalent to (739) becomes, for $G\in\mathbb{S}_c^N \Leftrightarrow G\mathbf{1}\!=\!\mathbf{0}$

$$\begin{array}{cl}\underset{G\in\mathbb{S}_c^N}{\text{find}} & G\in\mathbb{S}^N\\ \text{subject to} & \left\langle G\,,\,\delta\big((e_ie_j^T+e_je_i^T)\mathbf{1}\big)-(e_ie_j^T+e_je_i^T)\right\rangle = \check{d}_{ij}\;, \quad i,j\!=\!1\ldots N\,, \quad i<j\\ & G\succeq 0\end{array} \tag{741}$$

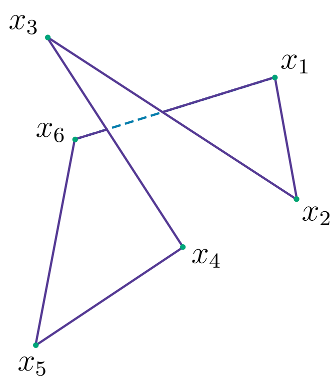



Figure 77: Arbitrary hexagon in $\mathbb{R}^3$ whose vertices are labelled clockwise.

Barvinok's Proposition 2.9.3.0.1 predicts existence for either formulation (739) or (741) such that implicit equality constraints induced by subspace membership are ignored

$$\operatorname{rank} G\ ,\ \operatorname{rank} VDV \leq \left\lfloor \frac{\sqrt{8(N(N-1)/2)+1}-1}{2} \right\rfloor = N-1 \tag{742}$$

because, in each case, the Gram matrix is confined to a face of positive semidefinite cone $\mathbb{S}_+^N$ isomorphic with $\mathbb{S}_+^{N-1}$ (§6.7.1). (§E.7.2.0.2) This bound is tight (§5.7.1.1) and is the greatest upper bound.[5.8]

**5.4.2.2.2 Example.** *Hexagon.*
Barvinok [22, §2.6] poses a problem in *geometric realizability* of an arbitrary hexagon (Figure **77**) having:

1. prescribed (one-dimensional) face-lengths $l$
2. prescribed angles $\varphi$ between the three pairs of opposing faces
3. a constraint on the sum of norm-square of each and every vertex $x$

[5.8] $-VDV|_{N\leftarrow 1} = 0$ (§B.4.1) ten affine equality constraints in all on a Gram matrix $G\in\mathbb{S}^{\mathbf{6}}$ (731). Let's realize this as a convex feasibility problem (with constraints written in the same order) also assuming $\mathbf{0}$ geometric center (730):

$$\begin{array}{cll}\underset{D\in\mathbb{S}_h^{\mathbf{6}}}{\text{find}} & -VDV\frac{1}{2}\in\mathbb{S}^{\mathbf{6}} & \\ \text{subject to} & \operatorname{tr}\big(D(e_ie_j^T+e_je_i^T)\frac{1}{2}\big) = l_{ij}^2\;, & j-1=(i=1\ldots 6)\bmod 6\\ & \operatorname{tr}\big(-\frac{1}{2}VDV(A_i+A_i^T)\frac{1}{2}\big) = \cos\varphi_i\;, & i=1,2,3\\ & \operatorname{tr}(-\frac{1}{2}VDV)=1 & \\ & -VDV\succeq 0 & \end{array}\tag{743}$$

where, for $A_i\in\mathbb{R}^{\mathbf{6}\times\mathbf{6}}$ (738)

$$\begin{aligned}A_1&=(e_1-e_6)(e_3-e_4)^T/(l_{61}l_{34})\\ A_2&=(e_2-e_1)(e_4-e_5)^T/(l_{12}l_{45})\\ A_3&=(e_3-e_2)(e_5-e_6)^T/(l_{23}l_{56})\end{aligned}\tag{744}$$

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 (740). We show how to numerically solve such a problem by *alternating projection* in §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 $\operatorname{tr}(-\frac{1}{2}VDV)=1$ (734). $\square$

**5.4.2.2.3 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? [302] The 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 **7**)

The question was presented in three dimensions to Isaac Newton in the context of celestial mechanics, and became controversy with David Gregory on the campus of Cambridge University in 1694. Newton correctly identified

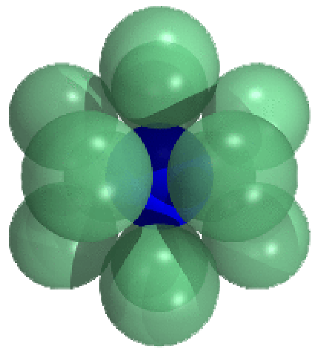


Figure 78: Sphere-packing illustration from [282, *kissing number*]. Translucent balls illustrated all have the same diameter.

the kissing number as 12 (Figure **78**) while Gregory argued for 13. Their dispute was finally resolved in 1953 by Schütte & van der Waerden. [221] In 2003, Oleg Musin tightened the upper bound on kissing number $K$ in four dimensions from 25 to $K=24$ by refining a method by Philippe Delsarte from 1973 providing an infinite number [14] of linear inequalities necessary for converting a rank-constrained semidefinite program[5.9] to a linear program.[5.10] [201]

There are no proofs known for kissing number in higher dimension excepting dimensions eight and twenty four.

Translating this problem to an *EDM graph* realization (Figure **75**, Figure **79**) 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 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, and assume a total of $N$

[5.9] whose feasible set belongs to that subset of an elliptope (§5.9.1.0.1) bounded above by some desired rank.

[5.10] Simplex-method solvers for linear programs produce numerically better results than contemporary log-barrier (interior-point method) solvers for semidefinite programs by about 7 orders of magnitude.

spheres:

$$\begin{array}{cll}\underset{D\in\mathbb{S}^N}{\text{minimize}} & -\operatorname{tr} WV_{\mathcal{N}}^TDV_{\mathcal{N}} & \\ \text{subject to} & D_{1j}=1\,, & j=2\ldots N \\ & D_{ij}\geq 1\,, & 2\leq i<j=3\ldots N \\ & D\in\mathbb{EDM}^N & \end{array} \tag{745}$$

Then the kissing number

$$K = N_{\max} - 1 \tag{746}$$

is found from the maximal number of spheres $N$ that solve this semidefinite program in a given affine dimension $r$. Matrix $W$ can be interpreted as the direction of search through the positive semidefinite cone $\mathbb{S}_+^{N-1}$ for a rank-$r$ optimal solution $-V_{\mathcal{N}}^TD^\star V_{\mathcal{N}}$; it is constant, in this program, determined by a method disclosed in §4.4.1. In two dimensions,

$$W = \begin{bmatrix} 4 & 1 & 2 & -1 & -1 & 1 \\ 1 & 4 & -1 & -1 & 2 & 1 \\ 2 & -1 & 4 & 1 & 1 & -1 \\ -1 & -1 & 1 & 4 & 1 & 2 \\ -1 & 2 & 1 & 1 & 4 & -1 \\ 1 & 1 & -1 & 2 & -1 & 4 \end{bmatrix}\frac{1}{6} \tag{747}$$

In three dimensions,

$$W = \begin{bmatrix} 9 & 1 & -2 & -1 & 3 & -1 & -1 & 1 & 2 & 1 & -2 & 1 \\ 1 & 9 & 3 & -1 & -1 & 1 & 1 & -2 & 1 & 2 & -1 & -1 \\ -2 & 3 & 9 & 1 & 2 & -1 & -1 & 2 & -1 & -1 & 1 & 2 \\ -1 & -1 & 1 & 9 & 1 & -1 & 1 & -1 & 3 & 2 & -1 & 1 \\ 3 & -1 & 2 & 1 & 9 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & 9 & 2 & -1 & 2 & -1 & 2 & 3 \\ -1 & 1 & -1 & 1 & 1 & 2 & 9 & 3 & -1 & 1 & -2 & -1 \\ 1 & -2 & 2 & -1 & -1 & -1 & 3 & 9 & 2 & -1 & 1 & 1 \\ 2 & 1 & -1 & 3 & -1 & 2 & -1 & 2 & 9 & -1 & 1 & -1 \\ 1 & 2 & -1 & 2 & -1 & -1 & 1 & -1 & -1 & 9 & 3 & 1 \\ -2 & -1 & 1 & -1 & 1 & 2 & -2 & 1 & 1 & 3 & 9 & -1 \\ 1 & -1 & 2 & 1 & -1 & 3 & -1 & 1 & -1 & 1 & -1 & 9 \end{bmatrix}\frac{1}{12} \tag{748}$$

A four-dimensional solution has rational direction matrix as well. Here is an optimal point list[5.11] in MATLAB output format:

[5.11] An optimal five-dimensional point list is known: *The answer was known at least 175*

```
Columns 1 through 6
X =     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
Columns 7 through 12
 -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
Columns 13 through 18
  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
Columns 19 through 25
 -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
```

This particular optimal solution was found by solving a problem sequence in increasing number of spheres. Numerical problems begin to arise with matrices of this cardinality due to interior-point methods of solution.

---

*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* $\mathrm{A}=\sqrt{2}$ *; then* $p(1)=(\mathrm{A},\mathrm{A},0,0,0)$, $p(2)=(-\mathrm{A},\mathrm{A},0,0,0)$, $p(3)=(\mathrm{A},-\mathrm{A},0,0,0),\ldots$ $p(40)=(0,0,0,-\mathrm{A},-\mathrm{A})$ *;* 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

By eliminating some equality constraints for this particular problem, matrix variable dimension can be reduced. From §5.8.3 we have

$$-V_{\mathcal{N}}^T D V_{\mathcal{N}} = \mathbf{1}\mathbf{1}^T - [\,\mathbf{0} \quad I\,]\, D \begin{bmatrix} \mathbf{0}^T \\ I \end{bmatrix} \frac{1}{2} \tag{749}$$

(which does not generally hold) where identity matrix $I \in \mathbb{S}^{N-1}$ has one less dimension than EDM $D$. By defining an EDM principal submatrix

$$\hat{D} \triangleq [\,\mathbf{0} \quad I\,]\, D \begin{bmatrix} \mathbf{0}^T \\ I \end{bmatrix} \in \mathbb{S}_h^{N-1} \tag{750}$$

we get a convex problem equivalent to (745)

$$\begin{array}{ll} \underset{\hat{D}\in\mathbb{S}^{N-1}}{\text{minimize}} & -\operatorname{tr}(W\hat{D}) \\ \text{subject to} & \hat{D}_{ij} \geq 1\;, \qquad\quad 1 \leq i < j = 2 \ldots N-1 \\ & \mathbf{1}\mathbf{1}^T - \hat{D}\frac{1}{2} \succeq 0 \\ & \delta(\hat{D}) = \mathbf{0} \end{array} \tag{751}$$

Any feasible matrix $\mathbf{1}\mathbf{1}^T - \hat{D}\frac{1}{2}$ belongs to an *elliptope* (§5.9.1.0.1). $\square$

This next example shows how finding the common point of intersection for three circles in a plane, a nonlinear problem, has convex expression.

**5.4.2.2.4 Example.** *So & Ye trilateration in wireless sensor network.* Given three known absolute point positions in $\mathbb{R}^{\mathbf{2}}$ (three *anchors* $\check{x}_2\,, \check{x}_3\,, \check{x}_4$) and only one unknown point (one *sensor* $x_1 \in \mathbb{R}^{\mathbf{2}}$), the sensor's absolute position is determined from its noiseless measured distance-square $\check{d}_{i1}$ to each of three anchors (Figure **2**, Figure **79**(a)). This trilateration can be expressed as a convex optimization problem in terms of list $X \triangleq [\,x_1 \;\; \check{x}_2 \;\; \check{x}_3 \;\; \check{x}_4\,] \in \mathbb{R}^{\mathbf{2}\times\mathbf{4}}$ and Gram matrix $G \in \mathbb{S}^{\mathbf{4}}$ (718):

$$\begin{array}{llll} \underset{G\in\mathbb{S}^{\mathbf{4}},\; X\in\mathbb{R}^{\mathbf{2}\times\mathbf{4}}}{\text{minimize}} & \operatorname{tr} G & & \\ \text{subject to} & \operatorname{tr}(G\,\Phi_{i1}) & = \check{d}_{i1}\;, & i = 2,3,4 \\ & \operatorname{tr}\!\left(G e_i e_i^T\right) & = \|\check{x}_i\|^2\;, & i = 2,3,4 \\ & \operatorname{tr}(G(e_i e_j^T + e_j e_i^T)/2) & = \check{x}_i^T \check{x}_j\;, & 2 \leq i < j = 3,4 \\ & X(:\,,2\!:\!4) & = [\,\check{x}_2 \;\; \check{x}_3 \;\; \check{x}_4\,] & \\ & \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} & \succeq 0 & \end{array} \tag{752}$$

(a) (b)

(c) (d)

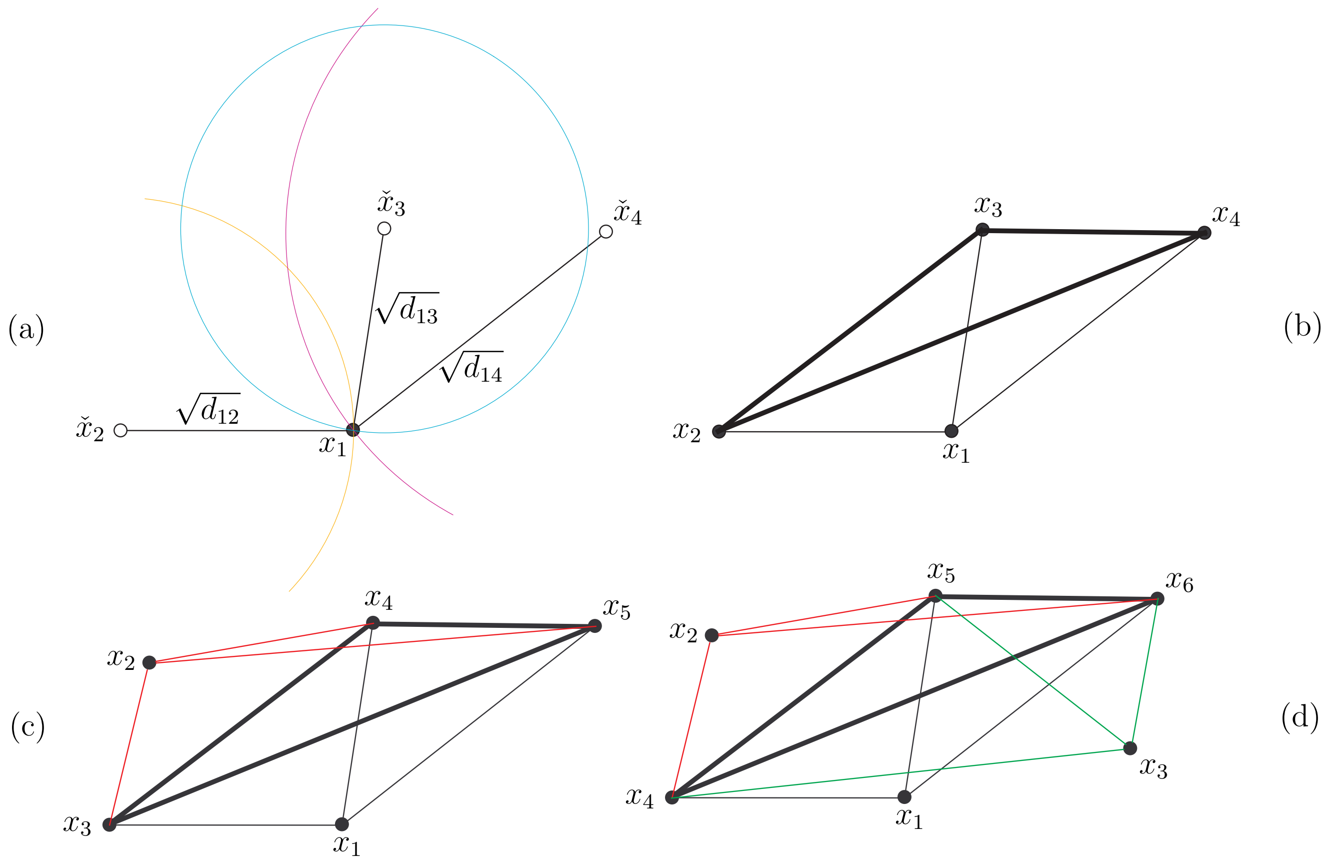



Figure 79: **(a)** Given three distances indicated with absolute point positions $\check{x}_2, \check{x}_3, \check{x}_4$ known and noncollinear, absolute position of $x_1$ in $\mathbb{R}^2$ can be precisely and uniquely determined by *trilateration*; solution to a system of nonlinear equations. Dimensionless 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 $\operatorname{rank} V_\mathcal{N}^T D V_\mathcal{N}$ 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 $\mathbf{D}$ (§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 (§5.4.2.2.5) of six points in $\mathbb{R}^2$ is certain.

where

$$\Phi_{ij} = (e_i - e_j)(e_i - e_j)^T \in \mathbb{S}^N_+ \qquad (707)$$

and where the constraint on distance-square $\check{d}_{i1}$ is equivalently written as a constraint on the Gram matrix via (720). There are 9 independent affine equality constraints on that Gram matrix while the sensor is constrained only by dimensioning to lie in $\mathbb{R}^{\mathbf{2}}$. Although $\operatorname{tr} G$ the objective of minimization[5.12] insures a solution on the boundary of positive semidefinite cone $\mathbb{S}^{\mathbf{4}}_+$, we claim that the set of feasible Gram matrices forms a line (§2.5.1.1) in isomorphic $\mathbb{R}^{\mathbf{10}}$ tangent (§2.1.7.2) to the positive semidefinite cone boundary. (*confer* §4.2.1.3)

By *Schur complement* (§A.4, §2.9.1.0.1) any feasible $G$ and $X$ provide

$$G \succeq X^T X \tag{753}$$

which is a convex *relaxation* of the desired (nonconvex) equality constraint

$$\begin{bmatrix} I & X \\ X^T & G \end{bmatrix} = \begin{bmatrix} I \\ X^T \end{bmatrix} [\, I \;\; X \,] \tag{754}$$

expected positive semidefinite rank-2 under noiseless conditions. But by (1292), the relaxation admits

$$(3 \geq) \operatorname{rank} G \geq \operatorname{rank} X \tag{755}$$

(a third dimension corresponding to an intersection of three spheres (not circles) were there noise). If rank

$$\operatorname{rank} \begin{bmatrix} I & X^\star \\ X^{\star T} & G^\star \end{bmatrix} = 2 \tag{756}$$

of an optimal solution equals 2, then $G^\star = X^{\star T} X^\star$ by Theorem A.4.0.0.4.

As posed, this *localization* problem does not require affinely independent (Figure **18**, 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 $x_1^\star = X^\star(:,1)$ is unique under noiseless measurement. [239] □

---

[5.12] Trace ($\operatorname{tr} G = \langle I\,,\, G\rangle$) 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 (753).

This preceding transformation of trilateration to a semidefinite program works all the time ((756) holds) despite relaxation (753) because the optimal solution set is a unique point.

**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 $G$

$$\begin{aligned}
\operatorname{tr} G &= \|x_1\|^2 + \|\check{x}_2\|^2 + \|\check{x}_3\|^2 + \|\check{x}_4\|^2 \\
\operatorname{tr}(G\Phi_{i1}) &= \check{d}_{i1}\ , && i=2,3,4 \\
\operatorname{tr}\big(Ge_ie_i^T\big) &= \|\check{x}_i\|^2\ , && i=2,3,4 \\
\operatorname{tr}(G(e_ie_j^T+e_je_i^T)/2) &= \check{x}_i^T\check{x}_j\ , && 2\leq i<j=3,4
\end{aligned} \tag{757}$$

which is invertible:

$$\operatorname{svec} G = \begin{bmatrix} \operatorname{svec}(I)^T \\ \operatorname{svec}(\Phi_{21})^T \\ \operatorname{svec}(\Phi_{31})^T \\ \operatorname{svec}(\Phi_{41})^T \\ \operatorname{svec}(e_2e_2^T)^T \\ \operatorname{svec}(e_3e_3^T)^T \\ \operatorname{svec}(e_4e_4^T)^T \\ \operatorname{svec}\big((e_2e_3^T+e_3e_2^T)/2\big)^T \\ \operatorname{svec}\big((e_2e_4^T+e_4e_2^T)/2\big)^T \\ \operatorname{svec}\big((e_3e_4^T+e_4e_3^T)/2\big)^T \end{bmatrix}^{-1} \begin{bmatrix} \|x_1\|^2 + \|\check{x}_2\|^2 + \|\check{x}_3\|^2 + \|\check{x}_4\|^2 \\ \check{d}_{21} \\ \check{d}_{31} \\ \check{d}_{41} \\ \|\check{x}_2\|^2 \\ \|\check{x}_3\|^2 \\ \|\check{x}_4\|^2 \\ \check{x}_2^T\check{x}_3 \\ \check{x}_2^T\check{x}_4 \\ \check{x}_3^T\check{x}_4 \end{bmatrix} \tag{758}$$

That line in the ambient space $\mathbb{S}^4$ of $G$ is traced by the right-hand side. One must show this line to be tangential (§2.1.7.2) to $\mathbb{S}^4_+$ 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 (but still containing the noiseless solution).

**5.4.2.2.5 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, or 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 **79** helps dispel any notion of distance data proliferation in *low affine dimension* $(r<N-2)$.[5.13] Huang, Liang, and Pardalos [153, §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 connectivity and point position. Alfakih shows how to ascertain uniqueness over all affine dimensions via *Gale matrix.* [7] [2] [3] Figure **75**(b) (page 295, from *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 **79**(b) will not yield a unique reconstruction when any one of the distances is left unspecified.

The list represented by the particular *dimensionless EDM graph* in Figure **80**, having only $2N-3=9$ absolute distances specified, has only one realization in $\mathbb{R}^2$ but has more realizations in higher dimensions. For sake of reference, we provide the complete corresponding EDM:

$$D=\begin{bmatrix} 0 & 50641 & 56129 & 8245 & 18457 & 26645\\ 50641 & 0 & 49300 & 25994 & 8810 & 20612\\ 56129 & 49300 & 0 & 24202 & 31330 & 9160\\ 8245 & 25994 & 24202 & 0 & 4680 & 5290\\ 18457 & 8810 & 31330 & 4680 & 0 & 6658\\ 26645 & 20612 & 9160 & 5290 & 6658 & 0 \end{bmatrix} \tag{759}$$

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 *ad hoc* method for recovery of the lowest-rank optimal solution under noiseless conditions is introduced:

[5.13] When 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. [136, §2.2]

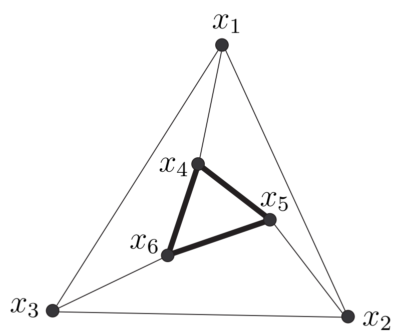



Figure 80: Incomplete EDM corresponding to this dimensionless EDM graph provides unique isometric reconstruction in $\mathbb{R}^2$. (drawn freehand, no symmetry intended)

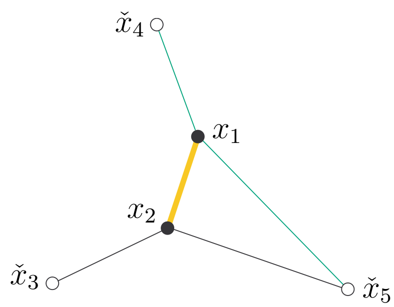



Figure 81: Two sensors $\bullet$ and three anchors $\circ$ in $\mathbb{R}^2$. (Ye) Connecting line-segments denote known absolute distances. Incomplete EDM corresponding to this dimensionless EDM graph provides unique isometric reconstruction in $\mathbb{R}^2$.

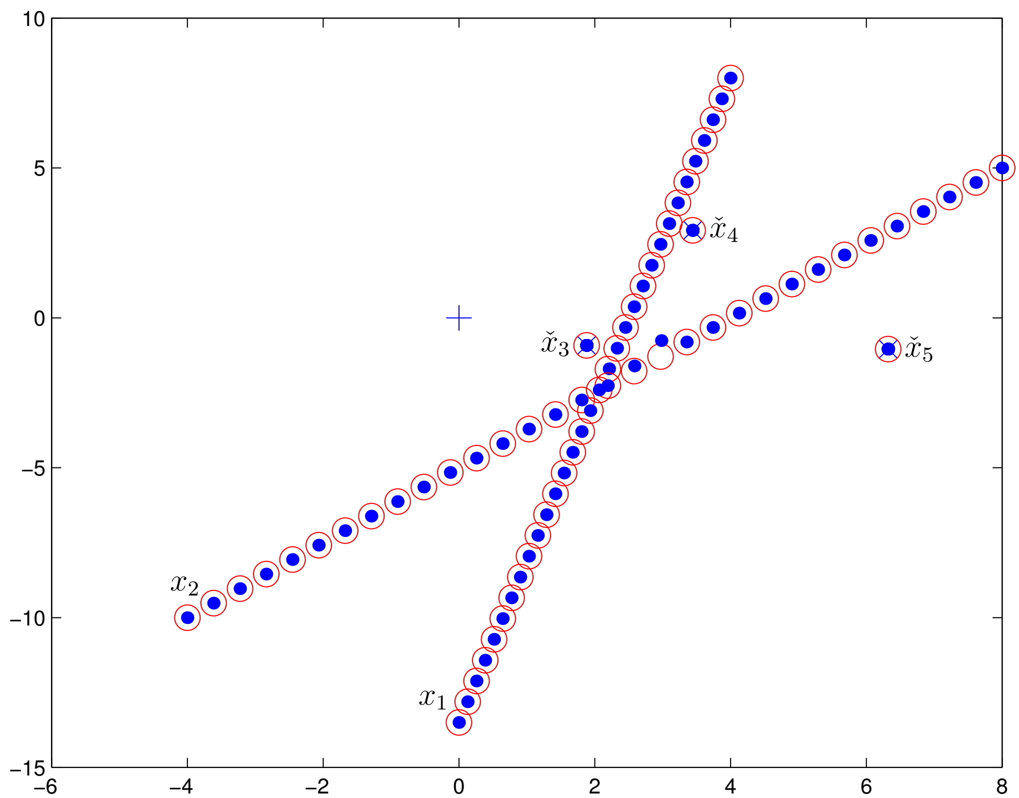



Figure 82: Given in red $\bigcirc$ are two discrete linear trajectories of sensors $x_1$ and $x_2$ in $\mathbb{R}^2$ localized by algorithm (760) as indicated by blue bullets $\bullet$. Anchors $\check{x}_3$ , $\check{x}_4$ , $\check{x}_5$ corresponding to Figure **81** are indicated by $\otimes$. When targets $\bigcirc$ and bullets $\bullet$ 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.

**5.4.2.2.6 Example.** *Tandem trilateration in wireless sensor network.* Given three known absolute point-positions in $\mathbb{R}^2$ (three anchors $\check{x}_3\,,\,\check{x}_4\,,\,\check{x}_5$) and two unknown sensors $x_1\,,\,x_2\in\mathbb{R}^2$, the sensors' absolute positions are determinable from their noiseless distances-square (as indicated in Figure **81**) assuming the anchors exhibit no rotational or reflective symmetry in their affine hull (§5.5.2). This example differs from Example 5.4.2.2.4 in so far 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 $X \triangleq [\,x_1\;\; x_2\;\; \check{x}_3\;\; \check{x}_4\;\; \check{x}_5\,]\in\mathbb{R}^{2\times 5}$ and Gram matrix $G\in\mathbb{S}^5$ (718) via relaxation (753):

$$\begin{array}{cll}
\underset{G\in\mathbb{S}^5,\;X\in\mathbb{R}^{2\times 5}}{\text{minimize}} & \operatorname{tr}G & \\
\text{subject to} & \operatorname{tr}(G\Phi_{i1}) = \check{d}_{i1}\;, & i=2,4,5\\
& \operatorname{tr}(G\Phi_{i2}) = \check{d}_{i2}\;, & i=3,5\\
& \operatorname{tr}\big(G e_i e_i^T\big) = \|\check{x}_i\|^2\;, & i=3,4,5\\
& \operatorname{tr}(G(e_i e_j^T + e_j e_i^T)/2) = \check{x}_i^T\check{x}_j\;, & 3\le i<j=4,5\\
& X(:\,,3\!:\!5) = [\,\check{x}_3\;\; \check{x}_4\;\; \check{x}_5\,] & \\
& \begin{bmatrix} I & X\\ X^T & G\end{bmatrix} \succeq 0 &
\end{array}\tag{760}$$

where

$$\Phi_{ij} = (e_i - e_j)(e_i - e_j)^T \in \mathbb{S}_+^N \tag{707}$$

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[5.14]. (By virtue of their dimensioning, the sensors are already constrained to $\mathbb{R}^2$ the affine hull of the anchors.)

Exhibited in Figure **82** are two mistakes in solution $X^\star(:\,,1\!:\!2)$ due to a rank-3 optimal Gram matrix $G^\star$. The trace objective is a heuristic minimizing convex envelope of quasiconcave function[5.15] $\operatorname{rank}G$. (§2.9.2.6.2, §7.2.2.1) A rank-2 optimal Gram matrix can be found and the errors

[5.14] the presumably nonempty convex set of all points $G$ and $X$ satisfying the constraints.
[5.15] 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. [264, §4]

corrected by choosing a different normal for the linear objective function, now implicitly the identity matrix $I$; *id est*,

$$\operatorname{tr} G = \langle G\,,\, I\rangle \;\leftarrow\; \langle G\,,\, \delta(u)\rangle \tag{761}$$

where vector $u \in \mathbb{R}^{\mathbf{5}}$ 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 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.[5.16] □

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:

**5.4.2.2.7 Example.** *Wireless location in a cellular telephone network.* Utilizing measurements of distance, time of flight, angle of arrival, or signal power, *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 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

[5.16] The $\log\det$ rank-heuristic from §7.2.2.4 does not work here because it chooses the wrong normal. Rank reduction (§4.1.1.1) is unsuccessful here because Barvinok's upper bound (§2.9.3.0.1) on rank of $G^\star$ is 4.

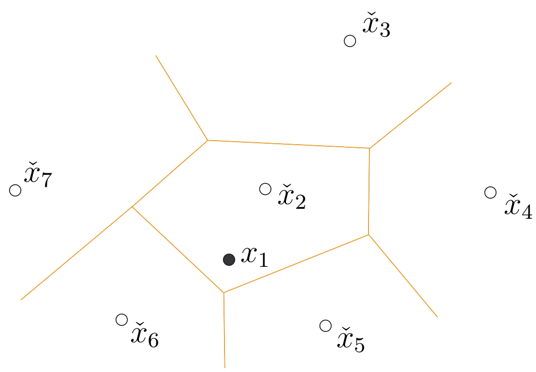



Figure 83: Regions of coverage by base stations $\circ$ in a cellular telephone network. The term *cellular* arises from packing of regions best covered by neighboring base stations. Illustrated is a pentagonal *cell* best covered by base station $\check{x}_2$. Like a Voronoi diagram, cell geometry depends 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 $\bullet$ ; practically, that number is closer to 6.

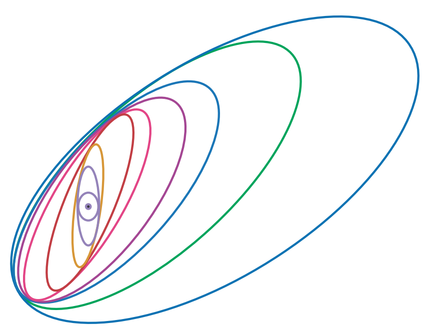


Figure 84: Some fitted contours of equal signal power in $\mathbb{R}^2$ transmitted from a commercial cellular telephone $\bullet$ over about 1 mile suburban terrain outside San Francisco in 2005. Data courtesy Polaris Wireless. [241]

by translated ellipsoids of graduated orientation and eccentricity as in Figure **84**.

Depicted in Figure **83** is one cell phone $x_1$ whose signal power is automatically and repeatedly measured by 6 base stations $\circ$ nearby.[5.17] Those signal power measurements are transmitted from that cell phone to base station $\check{x}_2$ who decides whether to transfer (*hand-off* or *hand-over*) responsibility for that call should the user roam outside its cell.[5.18]

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.[5.19] 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: $\overline{d_{i1}}$ is the upper bound on distance-square from the cell phone to $i^{\text{th}}$ base station, while $\underline{d_{i1}}$ is the lower bound. These bounds become the input data. Each measurement range is presumed different from the others.

Then convex problem (752) takes the form:

$$\begin{array}{cllll}
\underset{G\in\mathbb{S}^7,\ X\in\mathbb{R}^{2\times7}}{\text{minimize}} & \operatorname{tr} G & & & \\
\text{subject to} & \underline{d_{i1}} \leq \quad \operatorname{tr}(G\,\Phi_{i1}) & \leq & \overline{d_{i1}}\ , & i=2\ldots7 \\
& \operatorname{tr}\big(G e_i e_i^T\big) & = & \|\check{x}_i\|^2\ , & i=2\ldots7 \\
& \operatorname{tr}(G(e_i e_j^T + e_j e_i^T)/2) & = & \check{x}_i^T\check{x}_j\ , & 2\leq i<j=3\ldots7 \\
& X(:,2\!:\!7) & = & [\,\check{x}_2\ \ \check{x}_3\ \ \check{x}_4\ \ \check{x}_5\ \ \check{x}_6\ \ \check{x}_7\,] & \\
& \begin{bmatrix} I & X \\ X^T & G \end{bmatrix} & \succeq & 0 &
\end{array} \tag{762}$$

where

$$\Phi_{ij} = (e_i - e_j)(e_i - e_j)^T \in \mathbb{S}_+^N \qquad (707)$$

This semidefinite program realizes the wireless location problem illustrated

[5.17] Cell phone signal power is typically encoded logarithmically with 1-decibel increment and 64-decibel dynamic range.

[5.18] Because distance to base station is quite difficult to infer from signal power measurements in an urban environment, localization of a particular cell phone $\bullet$ by distance geometry would be far easier were the whole cellular system instead conceived so cell phone $x_1$ also transmits (to base station $\check{x}_2$) its signal power as received by all other cell phones within range.

[5.19] In Example 4.4.1.1.2, we explore how this convex optimization algorithm fares in the face of measurement noise.

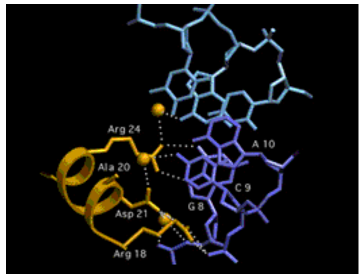



Figure 85: Example of molecular conformation. [80]

in Figure **83**. Location $X^\star(:,1)$ is taken as solution, although measurement noise will often cause $\operatorname{rank} G^\star$ to exceed 2. Randomized search for a rank-2 optimal solution is not so easy here as in Example 5.4.2.2.6. We introduce a method in §4.4 for enforcing the stronger rank-constraint (756). To formulate this same problem in three dimensions, point list $X$ is simply redimensioned in the semidefinite program. □

**5.4.2.2.8 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 **3**, Figure **85**. From this technique, distance, angle, and dihedral angle data 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.* [60, §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

$$X = [\, x_1 \;\cdots\; x_N \,] \in \mathbb{R}^{\mathbf{3}\times N} \tag{65}$$

and introduce a matrix $\aleph$ holding normals $\eta$ to planes respecting dihedral angles $\varphi$ :

$$\aleph \triangleq [\, \eta_1 \;\cdots\; \eta_M \,] \in \mathbb{R}^{\mathbf{3}\times M} \tag{763}$$

As in the other examples, we preferentially work with Gram matrices $G$ because of the bridge they provide between other variables; we define

$$\begin{bmatrix} G_\aleph & Z \\ Z^T & G_X \end{bmatrix} \triangleq \begin{bmatrix} \aleph^T\aleph & \aleph^T X \\ X^T\aleph & X^T X \end{bmatrix} = \begin{bmatrix} \aleph^T \\ X^T \end{bmatrix} [\, \aleph \;\; X \,] \in \mathbb{R}^{N+M\times N+M} \tag{764}$$

whose rank is 3 by assumption. So our problem's variables are the two Gram matrices and the matrix $Z$ of inner products. Then measurements of distance-square can be expressed as linear constraints on $G_X$ as in (762), dihedral angle $\varphi$ measurements can be expressed as linear constraints on $G_\aleph$ by (738), and the normal-vector conditions can be expressed by vanishing linear constraints on inner-product matrix $Z$. Consider three points $x$ labelled $1\,,2\,,3$ in the $\ell^{\text{th}}$ plane and its corresponding normal $\eta_\ell$. Then we may have, for example, the independent constraints

$$\begin{aligned} \eta_\ell^T(x_1 - x_2) = 0 \\ \eta_\ell^T(x_2 - x_3) = 0 \end{aligned} \tag{765}$$

expressible in terms of constant symmetric matrices $A$ ;

$$\begin{aligned} \langle Z\,,\, A_{\ell 12}\rangle = 0 \\ \langle Z\,,\, A_{\ell 23}\rangle = 0 \end{aligned} \tag{766}$$

NMR data is noisy, so measurements are described by given $\overline{\text{upper}}$ and $\underline{\text{lower}}$ bounds although normals $\eta$ can be constrained to be exactly unit length;

$$\delta(G_\aleph) = \mathbf{1} \tag{767}$$

Then we can express the molecular conformation problem: for $0 \leq \varphi \leq \pi$ and constant symmetric matrices $B$

$$\begin{array}{cll}
\underset{G_\aleph\in\mathbb{S}^M,\; G_X\in\mathbb{S}^N,\; Z\in\mathbb{R}^{M\times N}}{\text{find}} & G_X & \\
\text{subject to} & \underline{d_i} \quad \leq \quad \operatorname{tr}(G_X\,\Phi_i) \;\leq \overline{d_i}\;, & \forall\, i\in\mathcal{I}_1 \\
& \underline{\cos\varphi_j} \leq \quad \operatorname{tr}(G_\aleph\, B_j) \;\leq \overline{\cos\varphi_j}\;, & \forall\, j\in\mathcal{I}_2 \\
& \langle Z\,,\,A_k\rangle \qquad\qquad\qquad = 0\;, & \forall\, k\in\mathcal{I}_3 \\
& \delta(G_\aleph) \qquad\qquad\qquad\;\; = \mathbf{1} & \\
& \begin{bmatrix} G_\aleph & Z \\ Z^T & G_X \end{bmatrix} \qquad\qquad \succeq 0 & \\
& \operatorname{rank}\begin{bmatrix} G_\aleph & Z \\ Z^T & G_X \end{bmatrix} \qquad = 3 &
\end{array} \tag{768}$$

Ignoring the rank constraint tends to force inner-product matrix $Z$ to zero. What binds these three variables is the rank constraint; we show how to satisfy it in §4.4. □

### 5.4.3 Inner-product form EDM definition

> [p.20] *We might, for example, realize a constellation given only interstellar distance (or, equivalently, distance from Earth and relative angular measurement; the Earth as vertex to two stars).*

Equivalent to (705) is [288, §1-7] [249, §3.2]

$$\begin{aligned}
d_{ij} \;&=\; d_{ik} + d_{kj} - 2\sqrt{d_{ik}d_{kj}}\cos\theta_{ikj} \\
&=\; \begin{bmatrix}\sqrt{d_{ik}} & \sqrt{d_{kj}}\end{bmatrix}\begin{bmatrix} 1 & -e^{\imath\theta_{ikj}} \\ -e^{-\imath\theta_{ikj}} & 1\end{bmatrix}\begin{bmatrix}\sqrt{d_{ik}} \\ \sqrt{d_{kj}}\end{bmatrix}
\end{aligned} \tag{769}$$

called the *law of cosines*, where $\imath \triangleq \sqrt{-1}$, $i, k, j$ are positive integers, and $\theta_{ikj}$ is the angle at vertex $x_k$ formed by vectors $x_i - x_k$ and $x_j - x_k$ ;

$$\cos\theta_{ikj} \;=\; \frac{\frac{1}{2}(d_{ik} + d_{kj} - d_{ij})}{\sqrt{d_{ik}d_{kj}}} \;=\; \frac{(x_i - x_k)^T(x_j - x_k)}{\|x_i - x_k\|\,\|x_j - x_k\|} \tag{770}$$

where the numerator forms an inner product of vectors. Distance-square $d_{ij}\left(\begin{bmatrix}\sqrt{d_{ik}}\\ \sqrt{d_{kj}}\end{bmatrix}\right)$ is a convex quadratic function[5.20] on $\mathbb{R}^2_+$ whereas $d_{ij}(\theta_{ikj})$ is quasiconvex (§3.3) minimized over domain $-\pi \leq \theta_{ikj} \leq \pi$ by $\theta^\star_{ikj} = 0$, we get the *Pythagorean theorem* when $\theta_{ikj} = \pm\pi/2$, and $d_{ij}(\theta_{ikj})$ is maximized when $\theta^\star_{ikj} = \pm\pi$ ;

$$\begin{array}{ll} d_{ij} = \left(\sqrt{d_{ik}} + \sqrt{d_{kj}}\right)^2, & \theta_{ikj} = \pm\pi \\ d_{ij} = d_{ik} + d_{kj}\ , & \theta_{ikj} = \pm\frac{\pi}{2} \\ d_{ij} = \left(\sqrt{d_{ik}} - \sqrt{d_{kj}}\right)^2, & \theta_{ikj} = \ \ 0 \end{array} \tag{771}$$

so

$$|\sqrt{d_{ik}} - \sqrt{d_{kj}}\,| \ \leq\ \sqrt{d_{ij}}\ \leq\ \sqrt{d_{ik}} + \sqrt{d_{kj}} \tag{772}$$

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 (769) for $k=1$: By defining

$$\Theta^T\Theta \triangleq \begin{bmatrix} d_{12} & \sqrt{d_{12}d_{13}}\cos\theta_{213} & \sqrt{d_{12}d_{14}}\cos\theta_{214} & \cdots & \sqrt{d_{12}d_{1N}}\cos\theta_{21N} \\ \sqrt{d_{12}d_{13}}\cos\theta_{213} & d_{13} & \sqrt{d_{13}d_{14}}\cos\theta_{314} & \cdots & \sqrt{d_{13}d_{1N}}\cos\theta_{31N} \\ \sqrt{d_{12}d_{14}}\cos\theta_{214} & \sqrt{d_{13}d_{14}}\cos\theta_{314} & d_{14} & \ddots & \sqrt{d_{14}d_{1N}}\cos\theta_{41N} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \sqrt{d_{12}d_{1N}}\cos\theta_{21N} & \sqrt{d_{13}d_{1N}}\cos\theta_{31N} & \sqrt{d_{14}d_{1N}}\cos\theta_{41N} & \cdots & d_{1N} \end{bmatrix} \in \mathbb{S}^{N-1} \tag{773}$$

then any EDM may be expressed

$$\begin{aligned} \mathbf{D}(\Theta) &\triangleq \begin{bmatrix} 0 \\ \delta(\Theta^T\Theta) \end{bmatrix}\mathbf{1}^T + \mathbf{1}\begin{bmatrix} 0 & \delta(\Theta^T\Theta)^T \end{bmatrix} - 2\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \Theta^T\Theta \end{bmatrix} \ \in \mathbb{EDM}^N \\ &= \begin{bmatrix} 0 & \delta(\Theta^T\Theta)^T \\ \delta(\Theta^T\Theta) & \delta(\Theta^T\Theta)\mathbf{1}^T + \mathbf{1}\delta(\Theta^T\Theta)^T - 2\Theta^T\Theta \end{bmatrix} \end{aligned} \tag{774}$$

$$\mathbb{EDM}^N = \left\{\mathbf{D}(\Theta) \mid \Theta \in \mathbb{R}^{N-1\times N-1}\right\} \tag{775}$$

for which all Euclidean metric properties hold. Entries of $\Theta^T\Theta$ result from vector inner-products as in (770); *id est*,

---

**5.20** $\begin{bmatrix} 1 & -e^{\imath\theta_{ikj}} \\ -e^{-\imath\theta_{ikj}} & 1 \end{bmatrix} \succeq 0$, having eigenvalues $\{0,2\}$. Minimum is attained for $\begin{bmatrix}\sqrt{d_{ik}}\\ \sqrt{d_{kj}}\end{bmatrix} = \begin{cases} \mu\mathbf{1}\,, & \mu \geq 0\,,\ \theta_{ikj} = 0 \\ \mathbf{0}\,, & -\pi \leq \theta_{ikj} \leq \pi\ ,\ \theta_{ikj} \neq 0 \end{cases}$ . (§D.2.1, [46, exmp.4.5])

$$\Theta = [\, x_2 - x_1 \quad x_3 - x_1 \quad \cdots \quad x_N - x_1 \,] = X\sqrt{2}V_{\mathcal{N}} \in \mathbb{R}^{n\times N-1} \tag{776}$$

Inner product $\Theta^T\Theta$ is obviously related to a Gram matrix (718),

$$G = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \Theta^T\Theta \end{bmatrix}, \qquad x_1 = \mathbf{0} \tag{777}$$

For $D = \mathbf{D}(\Theta)$ and no condition on the list $X$ (*confer* (726) (731))

$$\Theta^T\Theta = -V_{\mathcal{N}}^T D V_{\mathcal{N}} \;\in \mathbb{R}^{N-1\times N-1} \tag{778}$$

#### 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 $\mathbf{D}$. Here is another one:

Like $\mathbf{D}(X)$ (709), $\mathbf{D}(\Theta)$ 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 (712). Scrutinizing $\Theta^T\Theta$ (773) 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 $\theta_{i1j}$ to fill $\Theta^T\Theta$ will not necessarily make an EDM; inner product (773) must be positive semidefinite.

$$\Theta^T\Theta = \sqrt{\delta(d)}\,\Omega\sqrt{\delta(d)} \triangleq$$

$$\begin{bmatrix} \sqrt{d_{12}} & & & \mathbf{0} \\ & \sqrt{d_{13}} & & \\ & & \ddots & \\ \mathbf{0} & & & \sqrt{d_{1N}} \end{bmatrix} \begin{bmatrix} 1 & \cos\theta_{213} & \cdots & \cos\theta_{21N} \\ \cos\theta_{213} & 1 & \ddots & \cos\theta_{31N} \\ \vdots & \ddots & \ddots & \vdots \\ \cos\theta_{21N} & \cos\theta_{31N} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sqrt{d_{12}} & & & \mathbf{0} \\ & \sqrt{d_{13}} & & \\ & & \ddots & \\ \mathbf{0} & & & \sqrt{d_{1N}} \end{bmatrix} \tag{779}$$

Expression $\mathbf{D}(\Theta)$ defines an EDM for any positive semidefinite *relative-angle matrix*

$$\Omega = [\cos\theta_{i1j}\;,\;\; i,j = 2\ldots N] \in \mathbb{S}^{N-1} \tag{780}$$

and any nonnegative distance vector

$$d = [d_{1j}\;,\;\; j = 2\ldots N] = \delta(\Theta^T\Theta) \in \mathbb{R}^{N-1} \tag{781}$$

because (§A.3.1.0.5)

$$\Omega \succeq 0 \;\Rightarrow\; \Theta^T\Theta \succeq 0 \tag{782}$$

Decomposition (779) and the *relative-angle matrix inequality* $\Omega \succeq 0$ lead to a different expression of an inner-product form EDM definition (774)

$$\begin{aligned}\mathbf{D}(\Omega\,,d) &\triangleq \begin{bmatrix}0\\ d\end{bmatrix}\mathbf{1}^T + \mathbf{1}\begin{bmatrix}0 & d^T\end{bmatrix} - 2\sqrt{\delta\left(\begin{bmatrix}0\\ d\end{bmatrix}\right)}\begin{bmatrix}0 & \mathbf{0}^T\\ \mathbf{0} & \Omega\end{bmatrix}\sqrt{\delta\left(\begin{bmatrix}0\\ d\end{bmatrix}\right)}\\ &= \begin{bmatrix}0 & d^T\\ d & d\mathbf{1}^T + \mathbf{1}d^T - 2\sqrt{\delta(d)}\,\Omega\sqrt{\delta(d)}\end{bmatrix} \in \mathbb{EDM}^N\end{aligned} \tag{783}$$

and another expression of the EDM cone:

$$\mathbb{EDM}^N = \left\{\mathbf{D}(\Omega\,,d) \mid \Omega \succeq 0\,,\ \sqrt{\delta(d)} \succeq 0\right\} \tag{784}$$

In the particular circumstance $x_1 = \mathbf{0}$, we can relate interpoint angle matrix $\Psi$ from the Gram decomposition in (718) to relative-angle matrix $\Omega$ in (779). Thus,

$$\Psi \equiv \begin{bmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \Omega\end{bmatrix}, \qquad x_1 = \mathbf{0} \tag{785}$$

#### 5.4.3.2 Inner-product form $-V_{\mathcal{N}}^T\mathbf{D}(\Theta)V_{\mathcal{N}}$ convexity

On page 325 we saw that each EDM entry $d_{ij}$ is a convex quadratic function of $\begin{bmatrix}\sqrt{d_{ik}}\\ \sqrt{d_{kj}}\end{bmatrix}$ and a quasiconvex function of $\theta_{ikj}$. Here the situation for inner-product form EDM operator $\mathbf{D}(\Theta)$ (774) is identical to that in §5.4.1 for list-form $\mathbf{D}(X)$; $-\mathbf{D}(\Theta)$ is not a quasiconvex function of $\Theta$ by the same reasoning, and from (778)

$$-V_{\mathcal{N}}^T\mathbf{D}(\Theta)V_{\mathcal{N}} = \Theta^T\Theta \tag{786}$$

is a convex quadratic function of $\Theta$ on domain $\mathbb{R}^{n\times N-1}$ achieving its minimum at $\Theta = \mathbf{0}$.

#### 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[5.21] with respect to EDM $D$ .

## 5.5 Invariance

When $D$ is an EDM, there exist an infinite number of corresponding $N$-point lists $X$ (65) 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 (705). Knowing that translation $\alpha$ in advance, we may remove it from the list constituting the columns of $X$ by subtracting $\alpha\mathbf{1}^T$. Then it stands to reason by list-form definition (709) of an EDM, for any translation $\alpha\in\mathbb{R}^n$

$$\mathbf{D}(X-\alpha\mathbf{1}^T)=\mathbf{D}(X) \tag{787}$$

In words, interpoint distances are unaffected by offset; EDM $D$ is *translation invariant.* When $\alpha=x_1$ in particular,

$$[\,x_2-x_1 \quad x_3-x_1 \quad \cdots \quad x_N-x_1\,]=X\sqrt{2}V_{\mathcal{N}}\in\mathbb{R}^{n\times N-1} \tag{776}$$

and so

$$\mathbf{D}(X-x_1\mathbf{1}^T) \;=\; \mathbf{D}(X-Xe_1\mathbf{1}^T) \;=\; \mathbf{D}\Big(X\Big[\,\mathbf{0} \quad \sqrt{2}V_{\mathcal{N}}\Big]\Big) \;=\; \mathbf{D}(X) \tag{788}$$

[5.21] The reason for the amount $O(N^2)$ information is because of the relative measurements. The use of a fixed reference in the measurement of angles and distances would reduce the 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 $x_2-x_1$ , $\theta_{i1j}$ becomes equivalent to $\sum\limits_{k=i}^{j-1}\theta_{k,1,k+1}\leq 2\pi$ and the amount of information is reduced to $2N-3$; rather, $O(N)$.

**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; [266] [113]

$$\alpha \;=\; \alpha_c \;\triangleq\; X b_c \;\triangleq\; \frac{1}{N} X\mathbf{1} \in \mathcal{P} \subseteq \mathcal{A} \tag{789}$$

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. [161] The geometric center always lies in the convex hull $\mathcal{P}$ of the list; *id est*, $\alpha_c \!\in\! \mathcal{P}$ because $b_c^T\mathbf{1}\!=\!1$ and $b_c \!\succeq\! 0$ .[5.22] Subtracting the geometric center from every list member,

$$X - \alpha_c\mathbf{1}^T \;=\; X - \frac{1}{N}X\mathbf{1}\mathbf{1}^T \;=\; X(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T) \;=\; XV \in \mathbb{R}^{n\times N} \tag{790}$$

So we have (*confer* (709))

$$\mathbf{D}(X) = \mathbf{D}(XV) = \delta(V^TX^TXV)\mathbf{1}^T + \mathbf{1}\delta(V^TX^TXV)^T - 2V^TX^TXV \in \mathbb{EDM}^N \tag{791}$$

$\square$

#### 5.5.1.1 Gram-form invariance

Following from (791) and the linear Gram-form EDM operator (721):

$$\mathbf{D}(G) = \mathbf{D}(VGV) = \delta(VGV)\mathbf{1}^T + \mathbf{1}\delta(VGV)^T - 2VGV \in \mathbb{EDM}^N \tag{792}$$

The Gram-form consequently exhibits invariance to translation by a *doublet* (§B.2) $u\mathbf{1}^T + \mathbf{1}u^T$ ;

$$\mathbf{D}(G) = \mathbf{D}\big(G - (u\mathbf{1}^T + \mathbf{1}u^T)\big) \tag{793}$$

because, for any $u \!\in\! \mathbb{R}^N$, $\mathbf{D}(u\mathbf{1}^T + \mathbf{1}u^T) \!=\! \mathbf{0}$. The collection of all such doublets forms the nullspace (802) to the operator; the *translation-invariant subspace* $\mathbb{S}_c^{N\perp}$ (1768) of the symmetric matrices $\mathbb{S}^N$. This means matrix $G$ can belong to an expanse more broad than a positive semidefinite cone; *id est*, $G \!\in\! \mathbb{S}_+^N - \mathbb{S}_c^{N\perp}$. So explains Gram matrix sufficiency in EDM definition (721).

[5.22] Any $b$ from $\alpha = Xb$ chosen such that $b^T\mathbf{1} = 1$, more generally, makes an auxiliary $V$-matrix. (§B.4.5)

### 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 $Q(X-\alpha\mathbf{1}^T)$ where $Q$ is an orthogonal matrix (§B.5).

We rightfully expect

$$\mathbf{D}\big(Q(X-\alpha\mathbf{1}^T)\big) = \mathbf{D}(QX-\beta\mathbf{1}^T) = \mathbf{D}(QX) = \mathbf{D}(X) \tag{794}$$

Because list-form $\mathbf{D}(X)$ 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 (794) follows directly from (709).

The class of premultiplying matrices for which interpoint distances are unaffected is a little more broad than orthogonal matrices. Looking at EDM definition (709), it appears that any matrix $Q_{\rm p}$ such that

$$X^TQ_{\rm p}^TQ_{\rm p}\,X = X^TX \tag{795}$$

will have the property

$$\mathbf{D}(Q_{\rm p}X) = \mathbf{D}(X) \tag{796}$$

An example is skinny $Q_{\rm p}\in\mathbb{R}^{m\times n}$ $(m>n)$ having orthonormal columns. We call such a matrix *orthonormal.*

#### 5.5.2.1 Inner-product form invariance

Likewise, $\mathbf{D}(\Theta)$ (774) is rotation/reflection invariant;

$$\mathbf{D}(Q_{\rm p}\Theta) = \mathbf{D}(Q\Theta) = \mathbf{D}(\Theta) \tag{797}$$

so (795) and (796) 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 in so far as every realization of a corresponding list $X$ is *congruent*:

## 5.6 Injectivity of D & unique reconstruction

Injectivity implies uniqueness of isometric reconstruction; hence, we endeavor to demonstrate it.

EDM operators list-form $\mathbf{D}(X)$ (709), Gram-form $\mathbf{D}(G)$ (721), and inner-product form $\mathbf{D}(\Theta)$ (774) are many-to-one surjections (§5.5) onto the same range; the EDM cone (§6): (*confer* (722) (804))

$$\begin{aligned}\mathbb{EDM}^N &= \left\{\mathbf{D}(X) : \mathbb{R}^{N-1\times N} \rightarrow \mathbb{S}_h^N \mid X \in \mathbb{R}^{N-1\times N}\right\}\\ &= \left\{\mathbf{D}(G) : \mathbb{S}^N \rightarrow \mathbb{S}_h^N \mid G \in \mathbb{S}_+^N - \mathbb{S}_c^{N\perp}\right\}\\ &= \left\{\mathbf{D}(\Theta) : \mathbb{R}^{N-1\times N-1} \rightarrow \mathbb{S}_h^N \mid \Theta \in \mathbb{R}^{N-1\times N-1}\right\}\end{aligned} \tag{798}$$

where (§5.5.1.1)

$$\mathbb{S}_c^{N\perp} = \{u\mathbf{1}^T + \mathbf{1}u^T \mid u\in\mathbb{R}^N\} \subseteq \mathbb{S}^N \tag{1768}$$

### 5.6.1 Gram-form bijectivity

Because linear Gram-form EDM operator

$$\mathbf{D}(G) = \delta(G)\mathbf{1}^T + \mathbf{1}\delta(G)^T - 2G \tag{721}$$

has no nullspace [58, §A.1] on the geometric center subspace[5.23] (§E.7.2.0.2)

$$\begin{aligned}\mathbb{S}_c^N &\triangleq \{G\in\mathbb{S}^N \mid G\mathbf{1} = \mathbf{0}\} && (1766)\\ &= \{G\in\mathbb{S}^N \mid \mathcal{N}(G) \supseteq \mathbf{1}\} = \{G\in\mathbb{S}^N \mid \mathcal{R}(G) \subseteq \mathcal{N}(\mathbf{1}^T)\}\\ &= \{VYV \mid Y\in\mathbb{S}^N\} \subset \mathbb{S}^N && (1767)\\ &\equiv \{V_{\mathcal{N}}AV_{\mathcal{N}}^T \mid A\in\mathbb{S}^{N-1}\}\end{aligned} \tag{799}$$

then $\mathbf{D}(G)$ on that subspace is injective.

[5.23]The equivalence $\equiv$ in (799) follows from the fact: Given $B = VYV = V_{\mathcal{N}}AV_{\mathcal{N}}^T \in \mathbb{S}_c^N$ with only matrix $A\in\mathbb{S}^{N-1}$ unknown, then $V_{\mathcal{N}}^{\dagger}BV_{\mathcal{N}}^{\dagger T} = A$ or $V_{\mathcal{N}}^{\dagger}YV_{\mathcal{N}}^{\dagger T} = A$.

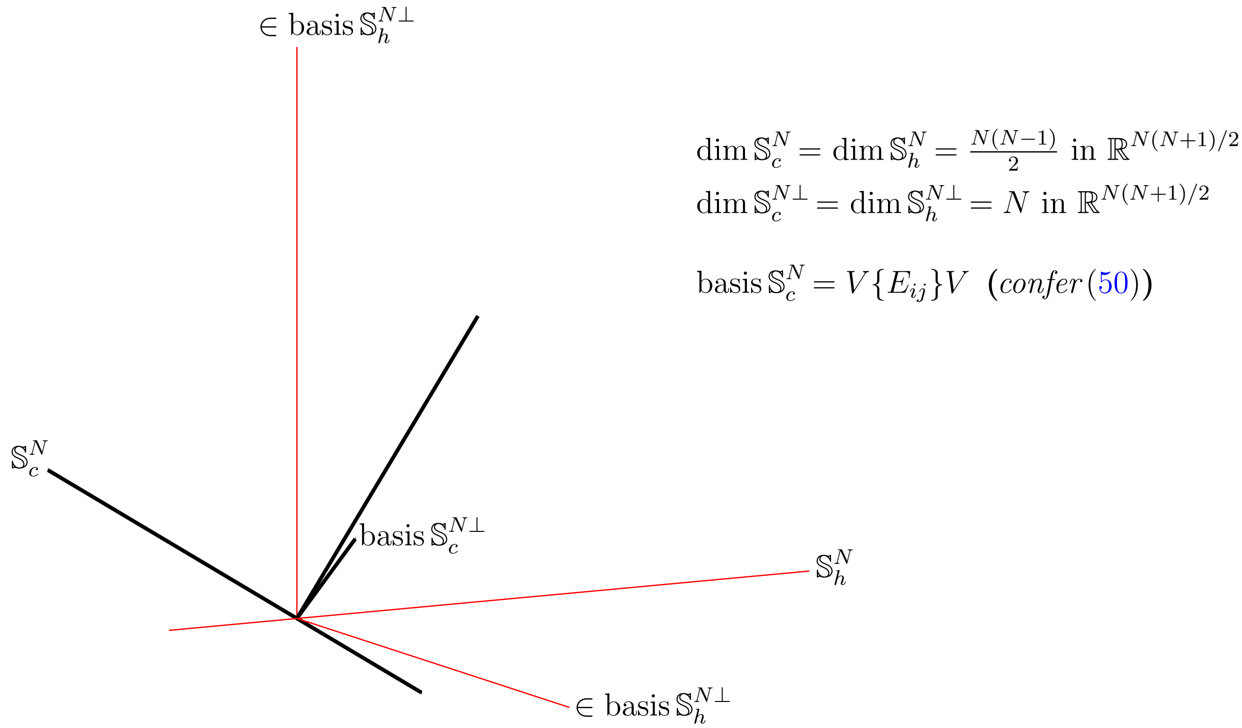



Figure 86: Orthogonal complements in $\mathbb{S}^N$ abstractly oriented in isometrically isomorphic $\mathbb{R}^{N(N+1)/2}$. Case $N=2$ accurately illustrated in $\mathbb{R}^3$. Orthogonal projection of basis for $\mathbb{S}_h^{N\perp}$ on $\mathbb{S}_c^{N\perp}$ yields another basis for $\mathbb{S}_c^{N\perp}$. (Basis vectors for $\mathbb{S}_c^{N\perp}$ are illustrated lying in a plane orthogonal to $\mathbb{S}_c^N$ in this dimension. Basis vectors for each $\perp$ space outnumber those for its respective orthogonal complement; such is not the case in higher dimension.)

To prove injectivity of $\mathbf{D}(G)$ on $\mathbb{S}_c^N$: Any matrix $Y\in\mathbb{S}^N$ can be decomposed into orthogonal components in $\mathbb{S}^N$;

$$Y = VYV + (Y - VYV) \tag{800}$$

where $VYV\in\mathbb{S}_c^N$ and $Y-VYV\in\mathbb{S}_c^{N\perp}$ (1768). Because of translation invariance (§5.5.1.1) and linearity, $\mathbf{D}(Y-VYV)=\mathbf{0}$ hence $\mathcal{N}(\mathbf{D})\supseteq\mathbb{S}_c^{N\perp}$. It remains only to show

$$\mathbf{D}(VYV) = \mathbf{0} \;\Leftrightarrow\; VYV = \mathbf{0} \tag{801}$$

$\big(\Leftrightarrow Y=u\mathbf{1}^T+\mathbf{1}u^T$ for some $u\in\mathbb{R}^N\big)$. $\mathbf{D}(VYV)$ will vanish whenever $2VYV = \delta(VYV)\mathbf{1}^T + \mathbf{1}\delta(VYV)^T$. But this implies $\mathcal{R}(\mathbf{1})$ (§B.2) were a subset of $\mathcal{R}(VYV)$, which is contradictory. Thus we have

$$\mathcal{N}(\mathbf{D}) \;=\; \{Y \mid \mathbf{D}(Y)=\mathbf{0}\} \;=\; \{Y \mid VYV=\mathbf{0}\} \;=\; \mathbb{S}_c^{N\perp} \tag{802}$$

♦

Since $G\mathbf{1}=\mathbf{0}\Leftrightarrow X\mathbf{1}=\mathbf{0}$ (730) simply means list $X$ is geometrically centered at the origin, and because the Gram-form EDM operator $\mathbf{D}$ is translation invariant and $\mathcal{N}(\mathbf{D})$ is the translation-invariant subspace $\mathbb{S}_c^{N\perp}$, then EDM definition $\mathbf{D}(G)$ (798) on[5.24] (*confer* §6.6.1, §6.7.1, §A.7.4.1)

$$\mathbb{S}_c^N\cap\mathbb{S}_+^N \;=\; \{VYV\succeq 0 \mid Y\in\mathbb{S}^N\} \;\equiv\; \{V_{\mathcal{N}}AV_{\mathcal{N}}^T \mid A\in\mathbb{S}_+^{N-1}\}\subset\mathbb{S}^N \tag{803}$$

must be surjective onto $\mathbb{EDM}^N$; (*confer* (722))

$$\mathbb{EDM}^N = \big\{\mathbf{D}(G) \mid G\in\mathbb{S}_c^N\cap\mathbb{S}_+^N\big\} \tag{804}$$

#### 5.6.1.1 Gram-form operator D inversion

Define the linear *geometric centering operator* $\mathbf{V}$; (*confer* (731))

$$\mathbf{V}(D) : \mathbb{S}^N\to\mathbb{S}^N \;\triangleq\; -V D V\tfrac{1}{2} \tag{805}$$

[61, §4.3][5.25] This orthogonal projector $\mathbf{V}$ has no nullspace on

$$\mathbb{S}_h^N = \operatorname{aff}\mathbb{EDM}^N \qquad (1055)$$

[5.24] Equivalence $\equiv$ in (803) follows from the fact: Given $B=VYV=V_{\mathcal{N}}AV_{\mathcal{N}}^T\in\mathbb{S}_+^N$ with only matrix $A$ unknown, then $V_{\mathcal{N}}^{\dagger}BV_{\mathcal{N}}^{\dagger T}=A$ and $A\in\mathbb{S}_+^{N-1}$ must be positive semidefinite by positive semidefiniteness of $B$ and Corollary A.3.1.0.5.

[5.25] Critchley cites Torgerson (1958) [262, ch.11, §2] for a history and derivation of (805).

because the projection of $-D/2$ on $\mathbb{S}_c^N$ (1766) can be $\mathbf{0}$ if and only if $D\in\mathbb{S}_c^{N\perp}$ ; but $\mathbb{S}_c^{N\perp}\cap\mathbb{S}_h^N=\mathbf{0}$ (Figure **86**). Projector $\mathbf{V}$ on $\mathbb{S}_h^N$ is therefore injective hence invertible. Further, $-V\mathbb{S}_h^N V/2$ is equivalent to the geometric center subspace $\mathbb{S}_c^N$ in the ambient space of symmetric matrices; a surjection,

$$\mathbb{S}_c^N \;=\; \mathbf{V}(\mathbb{S}^N) \;=\; \mathbf{V}\big(\mathbb{S}_h^N\oplus\mathbb{S}_h^{N\perp}\big) \;=\; \mathbf{V}\big(\mathbb{S}_h^N\big) \tag{806}$$

because (62)

$$\mathbf{V}\big(\mathbb{S}_h^N\big) \;\supseteq\; \mathbf{V}\big(\mathbb{S}_h^{N\perp}\big) \;=\; \mathbf{V}\big(\delta^2(\mathbb{S}^N)\big) \tag{807}$$

Because $\mathbf{D}(G)$ on $\mathbb{S}_c^N$ is injective, and $\operatorname{aff}\mathbf{D}\big(\mathbf{V}(\mathbb{EDM}^N)\big)=\mathbf{D}\big(\mathbf{V}(\operatorname{aff}\mathbb{EDM}^N)\big)$ by property (107) of the affine hull, we find for $D\in\mathbb{S}_h^N$ **(***confer* (735)**)**

$$\mathbf{D}(-VDV\tfrac{1}{2})=\delta(-VDV\tfrac{1}{2})\mathbf{1}^T+\mathbf{1}\delta(-VDV\tfrac{1}{2})^T-2(-VDV\tfrac{1}{2}) \tag{808}$$

*id est*,

$$D=\mathbf{D}\big(\mathbf{V}(D)\big) \tag{809}$$

$$-VDV=\mathbf{V}\big(\mathbf{D}(-VDV)\big) \tag{810}$$

or

$$\mathbb{S}_h^N=\mathbf{D}\big(\mathbf{V}(\mathbb{S}_h^N)\big) \tag{811}$$

$$-V\mathbb{S}_h^N V=\mathbf{V}\big(\mathbf{D}(-V\mathbb{S}_h^N V)\big) \tag{812}$$

These operators $\mathbf{V}$ and $\mathbf{D}$ are mutual inverses.

The Gram-form $\mathbf{D}\big(\mathbb{S}_c^N\big)$ (721) is equivalent to $\mathbb{S}_h^N$ ;

$$\mathbf{D}\big(\mathbb{S}_c^N\big) \;=\; \mathbf{D}\big(\mathbf{V}(\mathbb{S}_h^N\oplus\mathbb{S}_h^{N\perp})\big) \;=\; \mathbb{S}_h^N+\mathbf{D}\big(\mathbf{V}(\mathbb{S}_h^{N\perp})\big) \;=\; \mathbb{S}_h^N \tag{813}$$

because $\mathbb{S}_h^N\supseteq\mathbf{D}\big(\mathbf{V}(\mathbb{S}_h^{N\perp})\big)$. In summary, for the Gram-form we have the isomorphisms [62, §2] [61, p.76, p.107] [5, §2.1][5.26] [4, §2] [6, §18.2.1] [1, §2.1]

$$\mathbb{S}_h^N=\mathbf{D}(\mathbb{S}_c^N) \tag{814}$$

$$\mathbb{S}_c^N=\mathbf{V}(\mathbb{S}_h^N) \tag{815}$$

and from the bijectivity results in §5.6.1,

$$\mathbb{EDM}^N=\mathbf{D}(\mathbb{S}_c^N\cap\mathbb{S}_+^N) \tag{816}$$

$$\mathbb{S}_c^N\cap\mathbb{S}_+^N=\mathbf{V}(\mathbb{EDM}^N) \tag{817}$$

[5.26] In [5, p.6, line 20], delete sentence: *Since $G$ is also ... not a singleton set.* [5, p.10, line 11] $x_3=2$ (not 1).

### 5.6.2 Inner-product form bijectivity

The Gram-form EDM operator $\mathbf{D}(G) = \delta(G)\mathbf{1}^T + \mathbf{1}\delta(G)^T - 2G$ (721) 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

$$\begin{aligned}\mathbb{S}_1^N &\triangleq \{G \in \mathbb{S}^N \mid Ge_1 = \mathbf{0}\} \\ &= \left\{ \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} Y \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} \mid Y \in \mathbb{S}^N \right\}\end{aligned} \tag{1770}$$

because it obviously has no nullspace there. Since $Ge_1 = \mathbf{0} \Leftrightarrow Xe_1 = \mathbf{0}$ (723) means the first point in the list $X$ resides at the origin, then $\mathbf{D}(G)$ on $\mathbb{S}_1^N \cap \mathbb{S}_+^N$ must be surjective onto $\mathbb{EDM}^N$.

Substituting $\Theta^T\Theta \leftarrow -V_{\mathcal{N}}^T D V_{\mathcal{N}}$ (786) into inner-product form EDM definition $\mathbf{D}(\Theta)$ (774), it may be further decomposed: (*confer* (729))

$$\mathbf{D}(D) = \begin{bmatrix} 0 \\ \delta\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big) \end{bmatrix} \mathbf{1}^T + \mathbf{1} \begin{bmatrix} 0 & \delta\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big)^T \end{bmatrix} - 2 \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_{\mathcal{N}}^T D V_{\mathcal{N}} \end{bmatrix} \tag{818}$$

This linear operator $\mathbf{D}$ 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 $\mathbb{S}_h^N = \operatorname{aff} \mathbb{EDM}^N$, $\mathbf{D}(D)$ becomes an injective map onto that same subspace. Proof follows directly from the fact: linear $\mathbf{D}$ has no nullspace [58, §A.1] on $\mathbb{S}_h^N = \operatorname{aff} \mathbf{D}(\mathbb{EDM}^N) = \mathbf{D}(\operatorname{aff} \mathbb{EDM}^N)$ (107).

#### 5.6.2.1 Inversion of $\mathbf{D}\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big)$

Injectivity of $\mathbf{D}(D)$ suggests inversion of (*confer* (726))

$$\mathbf{V}_{\mathcal{N}}(D) : \mathbb{S}^N \to \mathbb{S}^{N-1} \triangleq -V_{\mathcal{N}}^T D V_{\mathcal{N}} \tag{819}$$

a linear surjective[5.27] mapping onto $\mathbb{S}^{N-1}$ having nullspace[5.28] $\mathbb{S}_c^{N\perp}$ ;

$$\mathbf{V}_{\mathcal{N}}(\mathbb{S}_h^N) = \mathbb{S}^{N-1} \tag{820}$$

---

[5.27] Surjectivity of $\mathbf{V}_{\mathcal{N}}(D)$ is demonstrated via the Gram-form EDM operator $\mathbf{D}(G)$: Since $\mathbb{S}_h^N = \mathbf{D}(\mathbb{S}_c^N)$ (813), then for any $Y \in \mathbb{S}^{N-1}$, $-V_{\mathcal{N}}^T \mathbf{D}(V_{\mathcal{N}}^{\dagger T} Y V_{\mathcal{N}}^{\dagger}/2) V_{\mathcal{N}} = Y$ .

[5.28] $\mathcal{N}(\mathbf{V}_{\mathcal{N}}) \supseteq \mathbb{S}_c^{N\perp}$ is apparent. There exists a linear mapping

$$T(\mathbf{V}_{\mathcal{N}}(D)) \triangleq V_{\mathcal{N}}^{\dagger T} \mathbf{V}_{\mathcal{N}}(D) V_{\mathcal{N}}^{\dagger} = -VDV\tfrac{1}{2} = \mathbf{V}(D)$$

injective on domain $\mathbb{S}_h^N$ because $\mathbb{S}_c^{N\perp} \cap \mathbb{S}_h^N = \mathbf{0}$. Revising the argument of this inner-product form (818), we get another flavor

$$\mathbf{D}\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big) = \begin{bmatrix} 0 \\ \delta\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big) \end{bmatrix} \mathbf{1}^T + \mathbf{1}\begin{bmatrix} 0 & \delta\big(-V_{\mathcal{N}}^T D V_{\mathcal{N}}\big)^T \end{bmatrix} - 2\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_{\mathcal{N}}^T D V_{\mathcal{N}} \end{bmatrix} \tag{821}$$

and we obtain mutual inversion of operators $\mathbf{V}_{\mathcal{N}}$ and $\mathbf{D}$, for $D \in \mathbb{S}_h^N$

$$D = \mathbf{D}\big(\mathbf{V}_{\mathcal{N}}(D)\big) \tag{822}$$

$$-V_{\mathcal{N}}^T D V_{\mathcal{N}} = \mathbf{V}_{\mathcal{N}}\big(\mathbf{D}(-V_{\mathcal{N}}^T D V_{\mathcal{N}})\big) \tag{823}$$

or

$$\mathbb{S}_h^N = \mathbf{D}\big(\mathbf{V}_{\mathcal{N}}(\mathbb{S}_h^N)\big) \tag{824}$$

$$-V_{\mathcal{N}}^T \mathbb{S}_h^N V_{\mathcal{N}} = \mathbf{V}_{\mathcal{N}}\big(\mathbf{D}(-V_{\mathcal{N}}^T \mathbb{S}_h^N V_{\mathcal{N}})\big) \tag{825}$$

Substituting $\Theta^T\Theta \leftarrow \Phi$ into inner-product form EDM definition (774), any EDM may be expressed by the new flavor

$$\begin{array}{c} \mathbf{D}(\Phi) \triangleq \begin{bmatrix} 0 \\ \delta(\Phi) \end{bmatrix} \mathbf{1}^T + \mathbf{1}\begin{bmatrix} 0 & \delta(\Phi)^T \end{bmatrix} - 2\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \Phi \end{bmatrix} \in \mathbb{EDM}^N \\ \Leftrightarrow \\ \Phi \succeq 0 \end{array} \tag{826}$$

where this $\mathbf{D}$ is a linear surjective operator onto $\mathbb{EDM}^N$ by definition, injective because it has no nullspace on domain $\mathbb{S}_+^{N-1}$. More broadly, $\operatorname{aff} \mathbf{D}(\mathbb{S}_+^{N-1}) = \mathbf{D}(\operatorname{aff} \mathbb{S}_+^{N-1})$ (107),

$$\begin{aligned} \mathbb{S}_h^N &= \mathbf{D}(\mathbb{S}^{N-1}) \\ \mathbb{S}^{N-1} &= \mathbf{V}_{\mathcal{N}}(\mathbb{S}_h^N) \end{aligned} \tag{827}$$

demonstrably isomorphisms, and by bijectivity of this inner-product form:

$$\mathbb{EDM}^N = \mathbf{D}(\mathbb{S}_+^{N-1}) \tag{828}$$

$$\mathbb{S}_+^{N-1} = \mathbf{V}_{\mathcal{N}}(\mathbb{EDM}^N) \tag{829}$$

---

such that

$$\mathcal{N}(T(\mathbf{V}_{\mathcal{N}})) = \mathcal{N}(\mathbf{V}) \supseteq \mathcal{N}(\mathbf{V}_{\mathcal{N}}) \supseteq \mathbb{S}_c^{N\perp} = \mathcal{N}(\mathbf{V})$$

where the equality $\mathbb{S}_c^{N\perp} = \mathcal{N}(\mathbf{V})$ is known (§E.7.2.0.2). ♦

## 5.7 Embedding in affine hull

The affine hull $\mathcal{A}$ (67) of a point list $\{x_\ell\}$ (arranged columnar in $X \in \mathbb{R}^{n\times N}$ (65)) is identical to the affine hull of that polyhedron $\mathcal{P}$ (75) formed from all convex combinations of the $x_\ell$ ; [46, §2] [230, §17]

$$\mathcal{A} \;=\; \operatorname{aff} X \;=\; \operatorname{aff} \mathcal{P} \tag{830}$$

Comparing hull definitions (67) and (75), it becomes obvious that the $x_\ell$ and their convex hull $\mathcal{P}$ are embedded in their unique affine hull $\mathcal{A}$ ;

$$\mathcal{A} \supseteq \mathcal{P} \supseteq \{x_\ell\} \tag{831}$$

Recall: affine dimension $r$ is a lower bound on embedding, equal to dimension of the subspace parallel to that nonempty affine set $\mathcal{A}$ in which the points are embedded. (§2.3.1) We define dimension of the convex hull $\mathcal{P}$ to be the same as dimension $r$ of the affine hull $\mathcal{A}$ [230, §2], but $r$ is not necessarily equal to the rank of $X$ (850).

For the particular example illustrated in Figure **74**, $\mathcal{P}$ is the triangle plus its relative interior while its three vertices constitute the entire list $X$ . The affine hull $\mathcal{A}$ is the unique plane that contains the triangle, so $r=2$ in that example while the rank of $X$ is 3. Were there only two points in Figure **74**, then the affine hull would instead be the unique line passing through them; $r$ would become 1 while the 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 $\mathcal{P}$ ; subtracting any $\alpha \in \mathcal{A}$ in the affine hull from every list member will work,

$$X - \alpha\mathbf{1}^T \tag{832}$$

translating $\mathcal{A}$ to the origin:[5.29]

$$\mathcal{A} - \alpha \;=\; \operatorname{aff}(X - \alpha\mathbf{1}^T) \;=\; \operatorname{aff}(X) - \alpha \tag{833}$$

$$\mathcal{P} - \alpha \;=\; \operatorname{conv}(X - \alpha\mathbf{1}^T) \;=\; \operatorname{conv}(X) - \alpha \tag{834}$$

[5.29] The manipulation of hull functions aff and conv follows from their definitions.

Because (830) and (831) translate,

$$\mathbb{R}^n \supseteq \mathcal{A}-\alpha = \operatorname{aff}(X-\alpha\mathbf{1}^T) = \operatorname{aff}(\mathcal{P}-\alpha) \supseteq \mathcal{P}-\alpha \supseteq \{x_\ell-\alpha\} \tag{835}$$

where from the previous relations it is easily shown

$$\operatorname{aff}(\mathcal{P}-\alpha) = \operatorname{aff}(\mathcal{P})-\alpha \tag{836}$$

Translating $\mathcal{A}$ neither changes its dimension or the dimension of the embedded polyhedron $\mathcal{P}$; (66)

$$r \triangleq \dim\mathcal{A} = \dim(\mathcal{A}-\alpha) \triangleq \dim(\mathcal{P}-\alpha) = \dim\mathcal{P} \tag{837}$$

For any $\alpha\in\mathbb{R}^n$, (833)-(837) remain true. [230, p.4, p.12] Yet when $\alpha\in\mathcal{A}$, the affine set $\mathcal{A}-\alpha$ becomes a unique subspace of $\mathbb{R}^n$ in which the $\{x_\ell-\alpha\}$ and their convex hull $\mathcal{P}-\alpha$ are embedded (835), 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 $\mathcal{A}$ and their convex hull $\mathcal{P}$ and, in particular, $x_1\in\mathcal{P}\subseteq\mathcal{A}$ to the origin in $\mathbb{R}^n$; *videlicet*,

$$X-x_1\mathbf{1}^T = X-Xe_1\mathbf{1}^T = X(I-e_1\mathbf{1}^T) = X\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix} \in\mathbb{R}^{n\times N} \tag{838}$$

where $V_{\mathcal{N}}$ is defined in (715), and $e_1$ in (725). Applying (835) to (838),

$$\mathbb{R}^n \supseteq \mathcal{R}(XV_{\mathcal{N}}) = \mathcal{A}-x_1 = \operatorname{aff}(X-x_1\mathbf{1}^T) = \operatorname{aff}(\mathcal{P}-x_1) \supseteq \mathcal{P}-x_1 \ni \mathbf{0} \tag{839}$$

where $XV_{\mathcal{N}}\in\mathbb{R}^{n\times N-1}$. Hence

$$r = \dim\mathcal{R}(XV_{\mathcal{N}}) \tag{840}$$

□

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

$$r = \dim\mathcal{R}(XV) \tag{841}$$

For any matrix whose range is $\mathcal{R}(V)=\mathcal{N}(\mathbf{1}^T)$ we get the same result; *e.g.*,

$$r = \dim \mathcal{R}(XV_{\mathcal{N}}^{\dagger T}) \tag{842}$$

because

$$\mathcal{R}(XV) = \{Xz \mid z\in\mathcal{N}(\mathbf{1}^T)\} \tag{843}$$

and $\mathcal{R}(V)=\mathcal{R}(V_{\mathcal{N}})=\mathcal{R}(V_{\mathcal{N}}^{\dagger T})$ (§E). These auxiliary matrices (§B.4.2) are more closely related;

$$V = V_{\mathcal{N}}V_{\mathcal{N}}^{\dagger} \qquad (1440)$$

#### 5.7.1.1 Affine dimension $r$ *versus* rank

Now, suppose $D$ is an EDM as defined by

$$\mathbf{D}(X) \triangleq \delta(X^TX)\mathbf{1}^T + \mathbf{1}\delta(X^TX)^T - 2X^TX \in \mathbb{EDM}^N \qquad (709)$$

and we premultiply by $-V_{\mathcal{N}}^T$ and postmultiply by $V_{\mathcal{N}}$. Then because $V_{\mathcal{N}}^T\mathbf{1}=\mathbf{0}$ (716), it is always true that

$$-V_{\mathcal{N}}^TDV_{\mathcal{N}} = 2V_{\mathcal{N}}^TX^TXV_{\mathcal{N}} = 2V_{\mathcal{N}}^TGV_{\mathcal{N}} \in \mathbb{S}^{N-1} \tag{844}$$

where $G$ is a Gram matrix. Similarly pre- and postmultiplying by $V$ (*confer* (731))

$$-VDV = 2VX^TXV = 2VGV \in \mathbb{S}^N \tag{845}$$

always holds because $V\mathbf{1}=\mathbf{0}$ (1430). Likewise, multiplying inner-product form EDM definition (774), it always holds:

$$-V_{\mathcal{N}}^TDV_{\mathcal{N}} = \Theta^T\Theta \in \mathbb{S}^{N-1} \qquad (778)$$

For any matrix $A$, $\operatorname{rank} A^TA = \operatorname{rank} A = \operatorname{rank} A^T$. [150, §0.4][5.30] So, by (843), affine dimension

$$\begin{aligned} r &= \operatorname{rank} XV = \operatorname{rank} XV_{\mathcal{N}} = \operatorname{rank} XV_{\mathcal{N}}^{\dagger T} = \operatorname{rank}\Theta \\ &= \operatorname{rank} VDV = \operatorname{rank} VGV = \operatorname{rank} V_{\mathcal{N}}^TDV_{\mathcal{N}} = \operatorname{rank} V_{\mathcal{N}}^TGV_{\mathcal{N}} \end{aligned} \tag{846}$$

[5.30] For $A\in\mathbb{R}^{m\times n}$, $\mathcal{N}(A^TA)=\mathcal{N}(A)$. [249, §3.3]

By *conservation of dimension*, (§A.7.3.0.1)

$$r+\dim\mathcal{N}(V_{\mathcal{N}}^TDV_{\mathcal{N}}) \;=\; N-1 \tag{847}$$

$$r+\dim\mathcal{N}(VDV) \;=\; N \tag{848}$$

For $D\in\mathbb{EDM}^N$

$$-V_{\mathcal{N}}^TDV_{\mathcal{N}}\succ 0 \;\Leftrightarrow\; r=N-1 \tag{849}$$

but $-VDV\nsucc 0$. The general fact[5.31] (*confer* (742))

$$r\leq\min\{n\,,\;N-1\} \tag{850}$$

is evident from (838) but can be visualized in the example illustrated in Figure **74**. 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: for list $X\in\mathbb{R}^{n\times N}$ and Gram matrix $G\in\mathbb{S}_+^N$

$$\begin{array}{rl}
r \;\triangleq & \dim(\mathcal{P}-\alpha) \;=\; \dim\mathcal{P} \;=\; \dim\operatorname{conv}X\\
= & \dim(\mathcal{A}-\alpha) \;=\; \dim\mathcal{A} \;=\; \dim\operatorname{aff}X\\
= & \operatorname{rank}(X-x_1\mathbf{1}^T) \;=\; \operatorname{rank}(X-\alpha_c\mathbf{1}^T)\\
= & \operatorname{rank}\Theta \quad (776)\\
= & \operatorname{rank}XV_{\mathcal{N}} \;=\; \operatorname{rank}XV \;=\; \operatorname{rank}XV_{\mathcal{N}}^{\dagger T}\\
= & \operatorname{rank}X\,,\quad Xe_1=\mathbf{0}\;\;\textbf{or}\;\;X\mathbf{1}=\mathbf{0}\\
= & \operatorname{rank}V_{\mathcal{N}}^TGV_{\mathcal{N}} \;=\; \operatorname{rank}VGV \;=\; \operatorname{rank}V_{\mathcal{N}}^{\dagger}GV_{\mathcal{N}}\\
= & \operatorname{rank}G\,,\quad Ge_1=\mathbf{0}\;(726)\;\;\textbf{or}\;\;G\mathbf{1}=\mathbf{0}\;(731)\\
= & \left.\begin{array}{l}\operatorname{rank}V_{\mathcal{N}}^TDV_{\mathcal{N}} \;=\; \operatorname{rank}VDV \;=\; \operatorname{rank}V_{\mathcal{N}}^{\dagger}DV_{\mathcal{N}} \;=\; \operatorname{rank}V_{\mathcal{N}}(V_{\mathcal{N}}^TDV_{\mathcal{N}})V_{\mathcal{N}}^T\\
= \operatorname{rank}\Lambda \quad (935)\\
= N-1-\dim\mathcal{N}\left(\begin{bmatrix}0 & \mathbf{1}^T\\ \mathbf{1} & -D\end{bmatrix}\right) = \operatorname{rank}\begin{bmatrix}0 & \mathbf{1}^T\\ \mathbf{1} & -D\end{bmatrix}-2 \quad (858)\end{array}\right\}\; D\in\mathbb{EDM}^N
\end{array} \tag{851}$$

[5.31] $\operatorname{rank}X\leq\min\{n\,,\;N\}$

### 5.7.3 Eigenvalues of $-VDV$ *versus* $-V_{\mathcal{N}}^{\dagger}DV_{\mathcal{N}}$

Suppose for $D\in\mathbb{EDM}^N$ we are given eigenvectors $v_i\in\mathbb{R}^N$ of $-VDV$ and corresponding eigenvalues $\lambda\in\mathbb{R}^N$ so that

$$-VDVv_i = \lambda_i v_i \;, \quad i=1\ldots N \tag{852}$$

From these we can determine the eigenvectors and eigenvalues of $-V_{\mathcal{N}}^{\dagger}DV_{\mathcal{N}}$ : Define

$$\nu_i \triangleq V_{\mathcal{N}}^{\dagger} v_i \;, \quad \lambda_i \neq 0 \tag{853}$$

Then we have:

$$-VDV_{\mathcal{N}}V_{\mathcal{N}}^{\dagger} v_i = \lambda_i v_i \tag{854}$$

$$-V_{\mathcal{N}}^{\dagger}VDV_{\mathcal{N}}\,\nu_i = \lambda_i V_{\mathcal{N}}^{\dagger} v_i \tag{855}$$

$$-V_{\mathcal{N}}^{\dagger}DV_{\mathcal{N}}\,\nu_i = \lambda_i \nu_i \tag{856}$$

the eigenvectors of $-V_{\mathcal{N}}^{\dagger}DV_{\mathcal{N}}$ are given by (853) while its corresponding nonzero eigenvalues are identical to those of $-VDV$ although $-V_{\mathcal{N}}^{\dagger}DV_{\mathcal{N}}$ is not necessarily positive semidefinite. In contrast, $-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}$ is positive semidefinite but its nonzero eigenvalues are generally different.

**5.7.3.0.1 Theorem.** *EDM rank versus affine dimension* $r$ . [113, §3] [133, §3] [112, §3] For $D\in\mathbb{EDM}^N$ (*confer* (1010))

1. $r=\operatorname{rank}(D)-1 \;\Leftrightarrow\; \mathbf{1}^TD^{\dagger}\mathbf{1}\neq 0$
   Points constituting a list $X$ generating the polyhedron corresponding to $D$ lie on the relative boundary of an $r$-dimensional *circumhypersphere* having

$$\begin{aligned} \text{diameter} &= \sqrt{2}\left(\mathbf{1}^TD^{\dagger}\mathbf{1}\right)^{-1/2} \\ \text{circumcenter} &= \tfrac{XD^{\dagger}\mathbf{1}}{\mathbf{1}^TD^{\dagger}\mathbf{1}} \end{aligned} \tag{857}$$

2. $r=\operatorname{rank}(D)-2 \;\Leftrightarrow\; \mathbf{1}^TD^{\dagger}\mathbf{1}=0$
   There can be no circumhypersphere whose relative boundary contains a generating list for the corresponding polyhedron.

3. In *Cayley-Menger form* [77, §6.2] [60, §3.3] [37, §40] (§5.11.2),

$$r \;=\; N-1-\dim\mathcal{N}\!\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix}\right) \;=\; \operatorname{rank}\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix} - 2 \tag{858}$$

   Circumhyperspheres exist for $r<\operatorname{rank}(D)-2$. [261, §7] ⋄

For all practical purposes, (850)

$$\max\{0\,,\ \operatorname{rank}(D)-2\}\ \leq\ r\ \leq\ \min\{n\,,\ N-1\} \tag{859}$$

# 5.8 Euclidean metric *versus* matrix criteria

## 5.8.1 Nonnegativity property 1

When $D=[d_{ij}]$ is an EDM (709), then it is apparent from (844)

$$2V_{\mathcal N}^TX^TXV_{\mathcal N}\ =\ -V_{\mathcal N}^TDV_{\mathcal N}\succeq 0 \tag{860}$$

because for any matrix $A$, $A^TA\succeq 0$ .[5.32] We claim nonnegativity of the $d_{ij}$ is enforced primarily by the matrix inequality (860); *id est*,

$$\left.\begin{array}{r}-V_{\mathcal N}^TDV_{\mathcal N}\succeq 0\\ D\in\mathbb S_h^N\end{array}\right\}\Rightarrow\ d_{ij}\geq 0\,,\ i\neq j \tag{861}$$

(The matrix inequality to enforce strict positivity differs by a stroke of the pen. (864))

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. [110, §4.2] Given $D\in\mathbb S_h^N$

$$-V_{\mathcal N}^TDV_{\mathcal N}=$$

$$\begin{bmatrix} d_{12} & \frac12(d_{12}+d_{13}-d_{23}) & \frac12(d_{1,i+1}+d_{1,j+1}-d_{i+1,j+1}) & \cdots & \frac12(d_{12}+d_{1N}-d_{2N})\\ \frac12(d_{12}+d_{13}-d_{23}) & d_{13} & \frac12(d_{1,i+1}+d_{1,j+1}-d_{i+1,j+1}) & \cdots & \frac12(d_{13}+d_{1N}-d_{3N})\\ \frac12(d_{1,j+1}+d_{1,i+1}-d_{j+1,i+1}) & \frac12(d_{1,j+1}+d_{1,i+1}-d_{j+1,i+1}) & d_{1,i+1} & \ddots & \frac12(d_{14}+d_{1N}-d_{4N})\\ \vdots & \vdots & \ddots & \ddots & \vdots\\ \frac12(d_{12}+d_{1N}-d_{2N}) & \frac12(d_{13}+d_{1N}-d_{3N}) & \frac12(d_{14}+d_{1N}-d_{4N}) & \cdots & d_{1N}\end{bmatrix}$$

$$=\tfrac12(\mathbf 1D_{1,2:N}+D_{2:N,1}\mathbf 1^T-D_{2:N,2:N})\in\mathbb S^{N-1} \tag{862}$$

[5.32] For $A\in\mathbb R^{m\times n}$, $A^TA\succeq 0\ \Leftrightarrow\ y^TA^TAy=\|Ay\|^2\geq 0$ for all $\|y\|=1$. When $A$ is full-rank skinny-or-square, $A^TA\succ 0$.

where row,column indices $i,j\in\{1\ldots N-1\}$. [234] It follows:

$$\left.\begin{array}{r}-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\succeq 0\\ D\in\mathbb{S}_{h}^{N}\end{array}\right\}\Rightarrow\;\delta(-V_{\mathcal{N}}^{T}DV_{\mathcal{N}})=\begin{bmatrix}d_{12}\\ d_{13}\\ \vdots\\ d_{1N}\end{bmatrix}\succeq 0 \tag{863}$$

Multiplication of $V_{\mathcal{N}}$ by any *permutation matrix* $\Xi$ has null effect on its range and nullspace. In other words, any permutation of the rows or columns of $V_{\mathcal{N}}$ produces a basis for $\mathcal{N}(\mathbf{1}^T)$; *id est*, $\mathcal{R}(\Xi_{\mathrm{r}}V_{\mathcal{N}})=\mathcal{R}(V_{\mathcal{N}}\Xi_{\mathrm{c}})=\mathcal{R}(V_{\mathcal{N}})=\mathcal{N}(\mathbf{1}^T)$. Hence, $-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\succeq 0 \;\Leftrightarrow\; -V_{\mathcal{N}}^{T}\,\Xi_{\mathrm{r}}^{T}D\Xi_{\mathrm{r}}V_{\mathcal{N}}\succeq 0 \quad (\Leftrightarrow\; -\Xi_{\mathrm{c}}^{T}V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\Xi_{\mathrm{c}}\succeq 0)$. Various permutation matrices[5.33] will sift the remaining $d_{ij}$ similarly to (863) thereby proving their nonnegativity. Hence $-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\succeq 0$ 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 (860). (§6.6.0.0.1)

#### 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;

$$\left.\begin{array}{r}-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\succ 0\\ D\in\mathbb{S}_{h}^{N}\end{array}\right\}\Rightarrow\; d_{ij}>0\,,\quad i\neq j \tag{864}$$

[5.33] The rule of thumb is: If $\Xi_{\mathrm{r}}(i\,,1)=1$, then $\delta(-V_{\mathcal{N}}^{T}\Xi_{\mathrm{r}}^{T}D\Xi_{\mathrm{r}}V_{\mathcal{N}})\in\mathbb{R}^{N-1}$ is some permutation of the $i^{\text{th}}$ row or column of $D$ excepting the 0 entry from the main diagonal.

### 5.8.2 Triangle inequality property 4

In light of Kreyszig's observation [166, §1.1, prob.15] that properties 2 through 4 of the Euclidean metric (§5.2) together imply property 1, the nonnegativity criterion (861) suggests that the matrix inequality $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$ might somehow take on the role of triangle inequality; *id est*,

$$\left.\begin{array}{r}\delta(D)=\mathbf{0}\\ D^T=D\\ -V_{\mathcal{N}}^T D V_{\mathcal{N}}\succeq 0\end{array}\right\}\Rightarrow \sqrt{d_{ij}}\;\leq\;\sqrt{d_{ik}}+\sqrt{d_{kj}}\;,\quad i\neq j\neq k \tag{865}$$

We now show that is indeed the case: Let $T$ be the *leading principal submatrix* in $\mathbb{S}^2$ of $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ (upper left $2\times 2$ submatrix from (862));

$$T \triangleq \begin{bmatrix} d_{12} & \frac{1}{2}(d_{12}+d_{13}-d_{23}) \\ \frac{1}{2}(d_{12}+d_{13}-d_{23}) & d_{13} \end{bmatrix} \tag{866}$$

Submatrix $T$ must be positive (semi)definite whenever $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ is. (§A.3.1.0.4, §5.8.3) Now we have,

$$\begin{array}{l} -V_{\mathcal{N}}^T D V_{\mathcal{N}}\succeq 0 \;\Rightarrow\; T\succeq 0 \;\Leftrightarrow\; \lambda_1\geq\lambda_2\geq 0\\ -V_{\mathcal{N}}^T D V_{\mathcal{N}}\succ 0 \;\Rightarrow\; T\succ 0 \;\Leftrightarrow\; \lambda_1>\lambda_2>0\end{array} \tag{867}$$

where $\lambda_1$ and $\lambda_2$ are the eigenvalues of $T$, real due only to symmetry of $T$:

$$\begin{array}{l}\lambda_1=\frac{1}{2}\left(d_{12}+d_{13}+\sqrt{d_{23}^2-2(d_{12}+d_{13})d_{23}+2(d_{12}^2+d_{13}^2)}\right)\in\mathbb{R}\\ \lambda_2=\frac{1}{2}\left(d_{12}+d_{13}-\sqrt{d_{23}^2-2(d_{12}+d_{13})d_{23}+2(d_{12}^2+d_{13}^2)}\right)\in\mathbb{R}\end{array} \tag{868}$$

Nonnegativity of eigenvalue $\lambda_1$ is guaranteed by only nonnegativity of the $d_{ij}$ which in turn is guaranteed by matrix inequality (861). Inequality between the eigenvalues in (867) follows from only realness of the $d_{ij}$. Since $\lambda_1$ always equals or exceeds $\lambda_2$, conditions for the 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; *videlicet*,

$$\begin{array}{rcl} T \succeq 0 \quad \Leftrightarrow \quad \det T = \lambda_1 \lambda_2 \geq 0 \;, \quad d_{12}\,, d_{13} \geq 0 & & \text{(c)} \\ \Leftrightarrow & & \\ \lambda_2 \geq 0 & & \text{(b)} \\ \Leftrightarrow & & \\ |\sqrt{d_{12}} - \sqrt{d_{23}}\,| \;\leq\; \sqrt{d_{13}} \;\leq\; \sqrt{d_{12}} + \sqrt{d_{23}} & & \text{(a)} \end{array} \tag{869}$$

Triangle inequality (869a) (*confer* (772) (881)), in terms of three rooted entries from $D$, is equivalent to metric property 4

$$\begin{array}{l} \sqrt{d_{13}} \leq \sqrt{d_{12}} + \sqrt{d_{23}} \\ \sqrt{d_{23}} \leq \sqrt{d_{12}} + \sqrt{d_{13}} \\ \sqrt{d_{12}} \leq \sqrt{d_{13}} + \sqrt{d_{23}} \end{array} \tag{870}$$

for the corresponding points $x_1, x_2, x_3$ from some length-$N$ list.[5.34]

#### 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 (772) 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, (869a) that came from $T$ (866). Yet we claim if $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$ then all triangle inequalities will be satisfied simultaneously;

$$|\sqrt{d_{ik}} - \sqrt{d_{kj}}\,| \;\leq\; \sqrt{d_{ij}} \;\leq\; \sqrt{d_{ik}} + \sqrt{d_{kj}} \;, \quad i<k<j \tag{871}$$

(There are no more.) To verify our claim, we must prove the matrix inequality $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$ to be a sufficient test of all the triangle inequalities; more efficient, we mention, for larger $N$ :

[5.34] Accounting for symmetry property 3, the fourth metric property demands three inequalities be satisfied *per* one of type (869a). The first of those inequalities in (870) is self evident from (869a), while the two remaining follow from the left-hand side of (869a) and the fact for scalars, $|a| \leq b \Leftrightarrow a \leq b$ and $-a \leq b$.

**5.8.2.1.1 Shore.** The columns of $\Xi_{\mathrm{r}} V_{\mathcal{N}} \Xi_{\mathrm{c}}$ hold a basis for $\mathcal{N}(\mathbf{1}^T)$ when $\Xi_{\mathrm{r}}$ and $\Xi_{\mathrm{c}}$ are permutation matrices. In other words, any permutation of the rows or columns of $V_{\mathcal{N}}$ leaves its range and nullspace unchanged; *id est*, $\mathcal{R}(\Xi_{\mathrm{r}} V_{\mathcal{N}} \Xi_{\mathrm{c}}) = \mathcal{R}(V_{\mathcal{N}}) = \mathcal{N}(\mathbf{1}^T)$ (716). Hence, two distinct matrix inequalities can be equivalent tests of the positive semidefiniteness of $D$ on $\mathcal{R}(V_{\mathcal{N}})$; *id est*, $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0 \Leftrightarrow -(\Xi_{\mathrm{r}} V_{\mathcal{N}} \Xi_{\mathrm{c}})^T D (\Xi_{\mathrm{r}} V_{\mathcal{N}} \Xi_{\mathrm{c}}) \succeq 0$. By properly choosing permutation matrices,[5.35] the leading principal submatrix $T_\Xi \in \mathbb{S}^2$ of $-(\Xi_{\mathrm{r}} V_{\mathcal{N}} \Xi_{\mathrm{c}})^T D (\Xi_{\mathrm{r}} V_{\mathcal{N}} \Xi_{\mathrm{c}})$ may be loaded with the entries of $D$ needed to test any particular triangle inequality (similarly to (862)-(869)). Because all the triangle inequalities can be individually tested using a test equivalent to the lone matrix inequality $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$, it logically follows that the lone matrix inequality tests all those triangle inequalities simultaneously. We conclude that $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$ is a sufficient test for the fourth property of the Euclidean metric, triangle inequality. ♦

#### 5.8.2.2 Strict triangle inequality

Without exception, all the inequalities in (869) and (870) can be made strict while their corresponding implications remain true. The then strict inequality (869a) or (870) may be interpreted as a *strict triangle inequality* under which collinear arrangement of points is not allowed. [164, §24/6, p.322] Hence by similar reasoning, $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succ 0$ is a sufficient test of all the strict triangle inequalities; *id est*,

$$\left.\begin{array}{r}\delta(D) = \mathbf{0}\\ D^T = D\\ -V_{\mathcal{N}}^T D V_{\mathcal{N}} \succ 0\end{array}\right\} \Rightarrow \sqrt{d_{ij}} < \sqrt{d_{ik}} + \sqrt{d_{kj}}\;, \quad i \neq j \neq k \tag{872}$$

### 5.8.3 $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ nesting

From (866) observe that $T = -V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N \leftarrow 3}$. In fact, for $D \in \mathbb{EDM}^N$, the leading principal submatrices of $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ form a nested sequence (by inclusion) whose members are individually positive semidefinite [110] [150] [249] and have the same form as $T$; *videlicet*,[5.36]

[5.35] 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 $\Xi_{\mathrm{r}}(i,1) = \Xi_{\mathrm{r}}(k,2) = \Xi_{\mathrm{r}}(j,3) = 1$ and $\Xi_{\mathrm{c}} = I$.

[5.36] $-VDV|_{N \leftarrow 1} = 0 \in \mathbb{S}_+^0$ (§B.4.1)

$$
\begin{aligned}
-V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow 1} &= [\,\emptyset\,] && \text{(o)}\\
-V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow 2} &= [d_{12}] \in \mathbb{S}_+ && \text{(a)}\\
-V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow 3} &= \begin{bmatrix} d_{12} & \frac{1}{2}(d_{12}+d_{13}-d_{23}) \\ \frac{1}{2}(d_{12}+d_{13}-d_{23}) & d_{13} \end{bmatrix} \;=\; T \in \mathbb{S}_+^{\mathbf{2}} && \text{(b)}\\
-V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow 4} &= \begin{bmatrix} d_{12} & \frac{1}{2}(d_{12}+d_{13}-d_{23}) & \frac{1}{2}(d_{12}+d_{14}-d_{24}) \\ \frac{1}{2}(d_{12}+d_{13}-d_{23}) & d_{13} & \frac{1}{2}(d_{13}+d_{14}-d_{34}) \\ \frac{1}{2}(d_{12}+d_{14}-d_{24}) & \frac{1}{2}(d_{13}+d_{14}-d_{34}) & d_{14} \end{bmatrix} && \text{(c)}\\
&\;\;\vdots\\
-V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow i} &= \begin{bmatrix} -V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow i-1} & \nu(i) \\ \nu(i)^T & d_{1i} \end{bmatrix} \in \mathbb{S}_+^{i-1} && \text{(d)}\\
&\;\;\vdots\\
-V_{\mathcal{N}}^T D V_{\mathcal{N}} &= \begin{bmatrix} -V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow N-1} & \nu(N) \\ \nu(N)^T & d_{1N} \end{bmatrix} \in \mathbb{S}_+^{N-1} && \text{(e)}
\end{aligned}
\tag{873}
$$

where

$$
\nu(i) \triangleq \frac{1}{2}\begin{bmatrix} d_{12}+d_{1i}-d_{2i} \\ d_{13}+d_{1i}-d_{3i} \\ \vdots \\ d_{1,i-1}+d_{1i}-d_{i-1,i} \end{bmatrix} \in \mathbb{R}^{i-2}, \quad i>2 \tag{874}
$$

Hence, the leading principal submatrices of EDM $D$ must also be EDMs.[5.37]

Bordered symmetric matrices in the form (873d) are known to have *intertwined* [249, §6.4] (or *interlaced* [150, §4.3] [246, §IV.4.1]) eigenvalues; (*confer* §5.11.1) that means, for the particular submatrices (873a) and (873b),

[5.37] In fact, each and every principal submatrix of an EDM $D$ is another EDM. [171, §4.1]

$$\lambda_2 \le d_{12} \le \lambda_1 \tag{875}$$

where $d_{12}$ is the eigenvalue of submatrix (873a) and $\lambda_1, \lambda_2$ are the eigenvalues of $T$ (873b) (866). Intertwining in (875) predicts that should $d_{12}$ become 0, then $\lambda_2$ must go to $0$.[5.38] The eigenvalues are similarly intertwined for submatrices (873b) and (873c);

$$\gamma_3 \le \lambda_2 \le \gamma_2 \le \lambda_1 \le \gamma_1 \tag{876}$$

where $\gamma_1, \gamma_2, \gamma_3$ are the eigenvalues of submatrix (873c). Intertwining likewise predicts that should $\lambda_2$ become 0 (a possibility revealed in §5.8.3.1), then $\gamma_3$ must go to 0. Combining results so far for $N=2,3,4$: (875) (876)

$$\gamma_3 \le \lambda_2 \le d_{12} \le \lambda_1 \le \gamma_1 \tag{877}$$

The preceding logic extends by induction through the remaining members of the sequence (873).

#### 5.8.3.1 Tightening the triangle inequality

Now we apply Schur complement from §A.4 to tighten the triangle inequality from (865) in case: cardinality $N=4$. We find that the gains by doing so are modest. From (873) we identify:

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \triangleq -V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow 4} \tag{878}$$

$$A \triangleq T = -V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow 3} \tag{879}$$

both positive semidefinite by assumption, where $B=\nu(4)$ (874), and $C=d_{14}$. Using nonstrict $CC^\dagger$-form (1311), $C\succeq 0$ by assumption (§5.8.1) and $CC^\dagger = I$. So by the *positive semidefinite ordering of eigenvalues theorem* (§A.3.1.0.1),

$$-V_{\mathcal{N}}^T D V_{\mathcal{N}}|_{N\leftarrow 4} \succeq 0 \;\Leftrightarrow\; T \succeq d_{14}^{-1}\nu(4)\nu(4)^T \;\Rightarrow\; \left\{\begin{array}{l} \lambda_1 \ge d_{14}^{-1}\|\nu(4)\|^2 \\ \lambda_2 \ge 0 \end{array}\right. \tag{880}$$

where $\{d_{14}^{-1}\|\nu(4)\|^2,\, 0\}$ are the eigenvalues of $d_{14}^{-1}\nu(4)\nu(4)^T$ while $\lambda_1, \lambda_2$ are the eigenvalues of $T$.

[5.38] If $d_{12}$ were 0, eigenvalue $\lambda_2$ becomes 0 (868) because $d_{13}$ must then be equal to $d_{23}$; *id est*, $d_{12}=0 \;\Leftrightarrow\; x_1 = x_2$. (§5.4)

**5.8.3.1.1 Example.** *Small completion problem*, II.
Applying the inequality for $\lambda_1$ in (880) to the *small completion problem* on page 294 Figure **75**, the lower bound on $\sqrt{d_{14}}$ (1.236 in (702)) is tightened to 1.289. The correct value of $\sqrt{d_{14}}$ to three significant figures is 1.414.

□

### 5.8.4 Affine dimension reduction in two dimensions

(*confer* §5.14.4) The leading principal $2\times 2$ submatrix $T$ of $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ has largest eigenvalue $\lambda_1$ (868) which is a convex function of $D$.[5.39] $\lambda_1$ 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 $\lambda_2$ is a concave function of $D$ that becomes 0 only at the upper and lower bounds of inequality (869a) and its equivalent forms: (*confer* (871))

$$\begin{array}{rcccll}
|\sqrt{d_{12}}-\sqrt{d_{23}}\,| & \leq & \sqrt{d_{13}} & \leq & \sqrt{d_{12}}+\sqrt{d_{23}} & \quad\text{(a)}\\
 & & \Leftrightarrow & & & \\
|\sqrt{d_{12}}-\sqrt{d_{13}}\,| & \leq & \sqrt{d_{23}} & \leq & \sqrt{d_{12}}+\sqrt{d_{13}} & \quad\text{(b)}\\
 & & \Leftrightarrow & & & \\
|\sqrt{d_{13}}-\sqrt{d_{23}}\,| & \leq & \sqrt{d_{12}} & \leq & \sqrt{d_{13}}+\sqrt{d_{23}} & \quad\text{(c)}
\end{array} \tag{881}$$

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 $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$, hence applicable to other triangles.

---

[5.39] The largest eigenvalue of any symmetric matrix is always a convex function of its entries, while the smallest eigenvalue is always concave. [46, exmp.3.10] In our particular case, say $\underline{d} \triangleq \begin{bmatrix} d_{12}\\ d_{13}\\ d_{23}\end{bmatrix} \in \mathbb{R}^{\mathbf{3}}$. Then the Hessian (1537) $\nabla^2\lambda_1(\underline{d})\succeq 0$ certifies convexity whereas $\nabla^2\lambda_2(\underline{d})\preceq 0$ certifies concavity. Each Hessian has rank equal to 1. The respective gradients $\nabla\lambda_1(\underline{d})$ and $\nabla\lambda_2(\underline{d})$ are nowhere $\mathbf{0}$.

## 5.9 Bridge: Convex polyhedra to EDMs

The criteria for the existence of an EDM include, by definition (709) (774), 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: for $i\,,j\,,k\in\{1\ldots N\}$ (*confer* §5.2)

$$\begin{array}{l}\sqrt{d_{ij}}\geq 0\,,\quad i\neq j\\ \sqrt{d_{ij}}=0\,,\quad i=j\\ \sqrt{d_{ij}}=\sqrt{d_{ji}}\\ \sqrt{d_{ij}}\;\leq\;\sqrt{d_{ik}}+\sqrt{d_{kj}}\;,\quad i\neq j\neq k\end{array}\quad\Leftarrow\quad\begin{array}{r}-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\succeq 0\\ \delta(D)=\mathbf{0}\\ D^{T}=D\end{array}\tag{882}$$

all implied by $D\in\mathbb{EDM}^N$. 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 *ab initio* the necessary and sufficient matrix criteria that will subsume all the Euclidean metric properties and any further requirements[5.40] for all $N>1$ (§5.8.3); *id est*,

$$\left.\begin{array}{r}-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\succeq 0\\ D\in\mathbb{S}_h^N\end{array}\right\}\;\Leftrightarrow\;D\in\mathbb{EDM}^N\tag{728}$$

or for EDM definition (783),

$$\left.\begin{array}{r}\Omega\succeq 0\\ \sqrt{\delta(d)}\succeq 0\end{array}\right\}\;\Leftrightarrow\;D=\mathbf{D}(\Omega\,,d)\in\mathbb{EDM}^N\tag{883}$$

[5.40] In 1935, Schoenberg [234, (1)] first extolled matrix product $-V_{\mathcal{N}}^{T}DV_{\mathcal{N}}$ (862) (predicated on symmetry and self-distance) specifically incorporating $V_{\mathcal{N}}$, albeit algebraically. He showed: nonnegativity $-y^{T}V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\,y\geq 0$, for all $y\in\mathbb{R}^{N-1}$, is necessary and sufficient for $D$ to be an EDM. Gower [112, §3] remarks how surprising it is that such a fundamental property of Euclidean geometry was obtained so late.

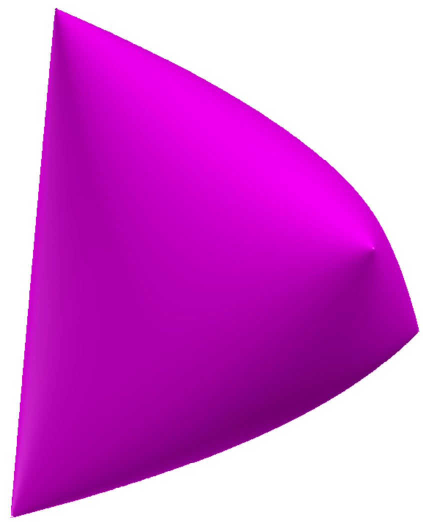


Figure 87: Elliptope $\mathcal{E}^{\mathbf{3}}$ in isometrically isomorphic $\mathbb{R}^{\mathbf{6}}$ (projected on $\mathbb{R}^{\mathbf{3}}$) is a convex body that appears to possess some kind of symmetry in this dimension; it resembles a malformed pillow in the shape of a bulging tetrahedron. Elliptope relative boundary is not *smooth* and comprises all set members (884) having at least one 0 eigenvalue. [174, §2.1] This elliptope has an infinity of vertices, but there are only four vertices corresponding to a rank-1 matrix. Those $yy^T$, evident in the illustration, have binary vector $y \in \mathbb{R}^{\mathbf{3}}$ with entries in $\{\pm 1\}$.

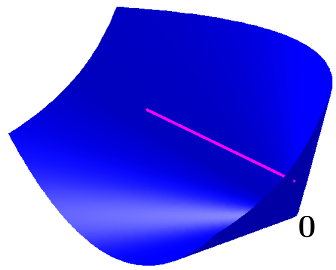



Figure 88: Elliptope $\mathcal{E}^{\mathbf{2}}$ in isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$ is a line segment illustrated interior to positive semidefinite cone $\mathbb{S}^{\mathbf{2}}_+$ (Figure **31**).

### 5.9.1 Geometric arguments

**5.9.1.0.1 Definition.** *Elliptope*: [174] [171, §2.3] [77, §31.5] a unique bounded immutable convex Euclidean body in $\mathbb{S}^n$; intersection of positive semidefinite cone $\mathbb{S}^n_+$ with that set of $n$ hyperplanes defined by unity main diagonal;

$$\mathcal{E}^n \triangleq \mathbb{S}^n_+ \cap \{\Phi \in \mathbb{S}^n \mid \delta(\Phi) = \mathbf{1}\} \tag{884}$$

`a.k.a`, the set of all *correlation matrices* of dimension

$$\dim \mathcal{E}^n = n(n-1)/2 \quad \text{in} \quad \mathbb{R}^{n(n+1)/2} \tag{885}$$

An elliptope $\mathcal{E}^n$ is not a polyhedron, in general, but has some polyhedral faces and an infinity of vertices.[5.41] Of those, $2^{n-1}$ vertices are extreme directions $yy^T$ of the positive semidefinite cone where the entries of vector $y \in \mathbb{R}^n$ each belong to $\{\pm 1\}$ while the vector exercises every combination. Each of the remaining vertices has rank belonging to the set $\{k > 0 \mid k(k+1)/2 \leq n\}$. Each and every face of an elliptope is exposed. △

[5.41] 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 **21**, for example.

The elliptope for dimension $n=2$ is a line segment in isometrically isomorphic $\mathbb{R}^{n(n+1)/2}$ (Figure **88**). Obviously, $\text{cone}(\mathcal{E}^n) \neq \mathbb{S}^n_+$. The elliptope for dimension $n=3$ is realized in Figure **87**.

**5.9.1.0.2 Lemma.** *Hypersphere.* [15, §4] (*confer* bullet p.304)
Matrix $A=[A_{ij}] \in \mathbb{S}^N$ belongs to the elliptope in $\mathbb{S}^N$ iff there exist $N$ points $p$ on the boundary of a hypersphere having radius 1 in $\mathbb{R}^{\text{rank}\,A}$ such that

$$\|p_i - p_j\| = \sqrt{2}\sqrt{1-A_{ij}}\ , \quad i\,,j=1\ldots N \tag{886}$$

$\diamond$

There is a similar theorem for Euclidean distance matrices:

We derive matrix criteria for $D$ to be an EDM, validating (728) using simple geometry; distance to the polyhedron formed by the convex hull of a list of points (65) in Euclidean space $\mathbb{R}^n$.

**5.9.1.0.3 EDM assertion.**
$D$ is a Euclidean distance matrix if and only if $D\in\mathbb{S}^N_h$ and distances-square from the origin

$$\{\|p(y)\|^2 = -y^T V_{\mathcal{N}}^T D V_{\mathcal{N}}\, y \mid y \in \mathcal{S}-\beta\} \tag{887}$$

correspond to points $p$ in some bounded convex polyhedron

$$\mathcal{P}-\alpha \;=\; \{p(y) \mid y \in \mathcal{S}-\beta\} \tag{888}$$

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 the 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 $V_{\mathcal{N}}$, the unit simplex (253) in $\mathbb{R}^{N-1}$ has an equivalent representation:

$$\mathcal{S} \;=\; \{s\in\mathbb{R}^{N-1} \mid \sqrt{2}V_{\mathcal{N}}\, s \succeq -e_1\} \tag{889}$$

where $e_1$ is as in (725). Incidental to the *EDM assertion*, shifting the unit-simplex domain in $\mathbb{R}^{N-1}$ translates the polyhedron $\mathcal{P}$ in $\mathbb{R}^n$. Indeed, there is a map from vertices of the unit simplex to members of the list generating $\mathcal{P}$ ;

$$\begin{array}{rcl} p \quad : \quad \mathbb{R}^{N-1} & \to & \mathbb{R}^n \\ p\left(\left\{\begin{array}{c} -\beta \\ e_1-\beta \\ e_2-\beta \\ \vdots \\ e_{N-1}-\beta \end{array}\right\}\right) & = & \left\{\begin{array}{c} x_1-\alpha \\ x_2-\alpha \\ x_3-\alpha \\ \vdots \\ x_N-\alpha \end{array}\right\} \end{array} \tag{890}$$

**5.9.1.0.4 Proof.** *EDM assertion.*
**(⇒)** We demonstrate that if $D$ is an EDM, then each distance-square $\|p(y)\|^2$ described by (887) corresponds to a point $p$ in some embedded polyhedron $\mathcal{P}-\alpha$. Assume $D$ is indeed an EDM; *id est*, $D$ can be made from some list $X$ of $N$ unknown points in Euclidean space $\mathbb{R}^n$; $D=\mathbf{D}(X)$ for $X\in\mathbb{R}^{n\times N}$ as in (709). Since $D$ is translation invariant (§5.5.1), we may shift the affine hull $\mathcal{A}$ of those unknown points to the origin as in (832). Then take any point $p$ in their convex hull (75);

$$\mathcal{P}-\alpha \;=\; \{p=(X-Xb\,\mathbf{1}^T)a \mid a^T\mathbf{1}=1,\; a\succeq 0\} \tag{891}$$

where $\alpha=Xb\in\mathcal{A} \;\Leftrightarrow\; b^T\mathbf{1}=1$. Solutions to $a^T\mathbf{1}=1$ are:[5.42]

$$a\in\left\{e_1+\sqrt{2}V_{\mathcal{N}}\,s \mid s\in\mathbb{R}^{N-1}\right\} \tag{892}$$

where $e_1$ is as in (725). Similarly, $b=e_1+\sqrt{2}V_{\mathcal{N}}\,\beta$.

$$\begin{aligned} \mathcal{P}-\alpha \;&=\; \{p=X(I-(e_1+\sqrt{2}V_{\mathcal{N}}\beta)\mathbf{1}^T)(e_1+\sqrt{2}V_{\mathcal{N}}\,s) \mid \sqrt{2}V_{\mathcal{N}}\,s\succeq -e_1\} \\ &=\; \{p=X\sqrt{2}V_{\mathcal{N}}(s-\beta) \mid \sqrt{2}V_{\mathcal{N}}\,s\succeq -e_1\} \end{aligned} \tag{893}$$

that describes the domain of $p(s)$ as the unit simplex

$$\mathcal{S}=\{s \mid \sqrt{2}V_{\mathcal{N}}\,s\succeq -e_1\}\subset\mathbb{R}_+^{N-1} \tag{889}$$

[5.42] Since $\mathcal{R}(V_{\mathcal{N}})=\mathcal{N}(\mathbf{1}^T)$ and $\mathcal{N}(\mathbf{1}^T)\perp\mathcal{R}(\mathbf{1})$, then over all $s\in\mathbb{R}^{N-1}$, $V_{\mathcal{N}}\,s$ is a hyperplane through the origin orthogonal to $\mathbf{1}$. Thus the solutions $\{a\}$ constitute a hyperplane orthogonal to the vector $\mathbf{1}$, and offset from the origin in $\mathbb{R}^N$ by any particular solution; in this case, $a=e_1$.

Making the substitution $s-\beta \leftarrow y$

$$\mathcal{P}-\alpha \;=\; \{p = X\sqrt{2}V_{\mathcal{N}}\, y \mid y\in\mathcal{S}-\beta\} \tag{894}$$

Point $p$ belongs to a convex polyhedron $\mathcal{P}-\alpha$ embedded in an $r$-dimensional subspace of $\mathbb{R}^n$ because the convex hull of any list forms a polyhedron, and because the translated affine hull $\mathcal{A}-\alpha$ contains the translated polyhedron $\mathcal{P}-\alpha$ (835) and the origin (when $\alpha\in\mathcal{A}$), and because $\mathcal{A}$ has dimension $r$ by definition (837). Now, any distance-square from the origin to the polyhedron $\mathcal{P}-\alpha$ can be formulated

$$\{p^Tp \;=\; \|p\|^2 \;=\; 2y^TV_{\mathcal{N}}^TX^TXV_{\mathcal{N}}\, y \mid y\in\mathcal{S}-\beta\} \tag{895}$$

Applying (844) to (895) we get (887).

**(⇐)** To validate the *EDM assertion* in the reverse direction, we prove: If each distance-square $\|p(y)\|^2$ (887) on the shifted unit-simplex $\mathcal{S}-\beta\subset\mathbb{R}^{N-1}$ corresponds to a point $p(y)$ in some embedded polyhedron $\mathcal{P}-\alpha$, then $D$ is an EDM. The $r$-dimensional subspace $\mathcal{A}-\alpha\subseteq\mathbb{R}^n$ is spanned by

$$p(\mathcal{S}-\beta) = \mathcal{P}-\alpha \tag{896}$$

because $\mathcal{A}-\alpha = \operatorname{aff}(\mathcal{P}-\alpha)\supseteq\mathcal{P}-\alpha$ (835). So, outside the domain $\mathcal{S}-\beta$ of linear surjection $p(y)$, the simplex complement $\backslash\mathcal{S}-\beta\subset\mathbb{R}^{N-1}$ must contain the domain of the distance-square $\|p(y)\|^2 = p(y)^Tp(y)$ to remaining points in the subspace $\mathcal{A}-\alpha$; *id est*, to the polyhedron's relative exterior $\backslash\mathcal{P}-\alpha$. For $\|p(y)\|^2$ to be nonnegative on the entire subspace $\mathcal{A}-\alpha$, $-V_{\mathcal{N}}^TDV_{\mathcal{N}}$ must be positive semidefinite and is assumed symmetric;[5.43]

$$-V_{\mathcal{N}}^TDV_{\mathcal{N}} \;\triangleq\; \Theta_{\mathrm{p}}^T\Theta_{\mathrm{p}} \tag{897}$$

where[5.44] $\Theta_{\mathrm{p}}\in\mathbb{R}^{m\times N-1}$ for some $m\geq r$. Because $p(\mathcal{S}-\beta)$ 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 (887) are

$$\{p^Tp \mid p\in\mathcal{P}-\alpha\} \;=\; \{p^Tp = y^T\Theta_{\mathrm{p}}^T\Theta_{\mathrm{p}}\, y \mid y\in\mathcal{S}-\beta\} \tag{898}$$

---

[5.43] The antisymmetric part $\big(-V_{\mathcal{N}}^TDV_{\mathcal{N}} - (-V_{\mathcal{N}}^TDV_{\mathcal{N}})^T\big)/2$ is annihilated by $\|p(y)\|^2$. By the same reasoning, any positive (semi)definite matrix $A$ is generally assumed symmetric because only the symmetric part $(A+A^T)/2$ survives the test $y^TAy\geq 0$. [150, §7.1]

[5.44] $A^T = A\succeq 0 \;\Leftrightarrow\; A = R^TR$ for some real matrix $R$. [249, §6.3]

Because $p \in \mathcal{P} - \alpha$ may be found by factoring (898), the list $\Theta_{\mathrm{p}}$ is found by factoring (897). A unique EDM can be made from that list using inner-product form definition $\mathbf{D}(\Theta)|_{\Theta=\Theta_{\mathrm{p}}}$ (774). That EDM will be identical to $D$ if $\delta(D) = \mathbf{0}$, by injectivity of $\mathbf{D}$ (818). ♦

### 5.9.2 Necessity and sufficiency

From (860) we learned that matrix inequality $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$ 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 (728), (883), and (704) are therefore necessary and sufficient and subsume all the Euclidean metric properties and further requirements.

### 5.9.3 Example revisited

Now we apply the necessary and sufficient EDM criteria (728) 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 **75** where $N = 4$, distance-square $d_{14}$ is ascertainable from the matrix inequality $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$. Because all distances in (701) are known except $\sqrt{d_{14}}$ , we may simply calculate the smallest eigenvalue of $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ over a range of $d_{14}$ as in Figure **89**. We observe a unique value of $d_{14}$ satisfying (728) where the abscissa 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 (871), 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) □

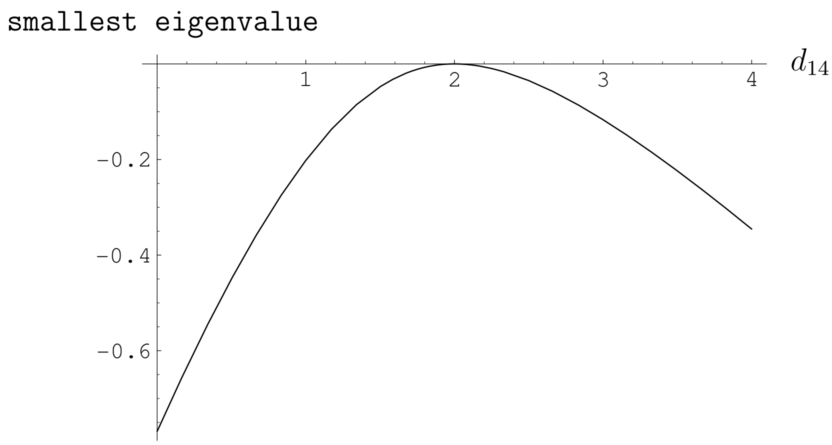



Figure 89: Smallest eigenvalue of $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ makes it a PSD matrix for only one value of $d_{14}$ : 2.

(a)

(b)

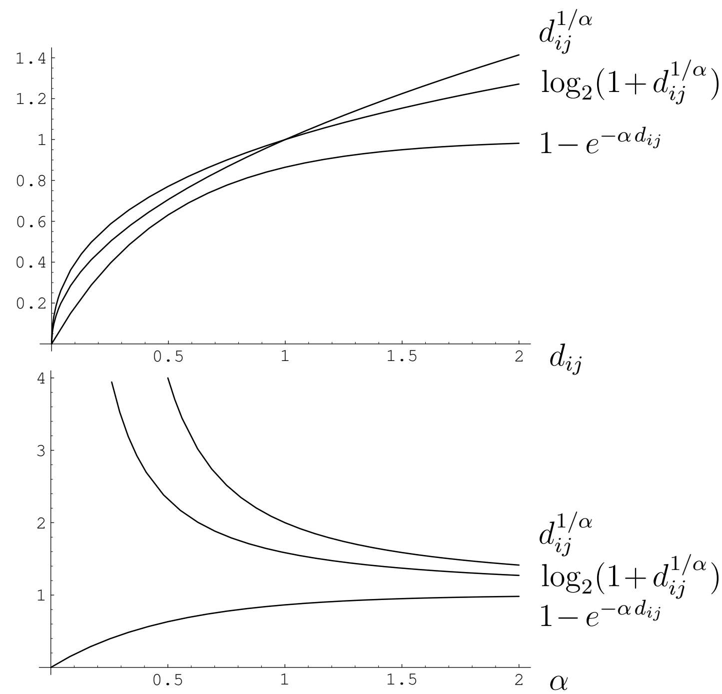



Figure 90: Some entrywise EDM compositions: **(a)** $\alpha = 2$. Concave nondecreasing in $d_{ij}$. **(b)** Trajectory convergence in $\alpha$ for $d_{ij} = 2$.

## 5.10 EDM-entry composition

Laurent [171, §2.3] applies results from Schoenberg (1938) [235] to show certain nonlinear compositions of individual EDM entries yield EDMs; in particular,

$$\begin{aligned} D \in \mathbb{EDM}^N \;&\Leftrightarrow\; [1 - e^{-\alpha\, d_{ij}}] \in \mathbb{EDM}^N \quad \forall\, \alpha > 0 \\ &\Leftrightarrow\; [e^{-\alpha\, d_{ij}}] \in \mathcal{E}^N \qquad\qquad\quad \forall\, \alpha > 0 \end{aligned} \tag{899}$$

where $D = [d_{ij}]$ and $\mathcal{E}^N$ is the elliptope (884).

**5.10.0.0.1 Proof.** (Laurent, 2003) [235] (*confer* [166])

***Lemma 2.1.*** from *A Tour d'Horizon ... on Completion Problems.* [171] The following assertions are equivalent: 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),

(i) $D \in \mathbb{EDM}^N$

(ii) $e^{-\alpha D} \triangleq [e^{-\alpha d_{ij}}] \in \mathcal{E}^N$ for all $\alpha > 0$

(iii) $\mathbf{1}\mathbf{1}^T - e^{-\alpha D} \triangleq [1 - e^{-\alpha d_{ij}}] \in \mathbb{EDM}^N$ for all $\alpha > 0$ $\diamond$

1) Equivalence of Lemma 2.1 (i) (ii) is stated in Schoenberg's Theorem 1 [235, p.527].

2) (ii) $\Rightarrow$ (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 $d = (d_{ij})_{i,j=0,1\ldots N}$ be a distance space on $N+1$ points (*i.e.*, symmetric hollow matrix of order $N+1$) and let $p = (p_{ij})_{i,j=1\ldots N}$ 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) $d_{0i} = p_{ii}\;,\quad d_{ij} = p_{ii} + p_{jj} - 2p_{ij}$ 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 [235]).

(ii) $\Rightarrow$ (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) $\Rightarrow$ (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) $\Rightarrow$ (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 [235, p.526]). ♦

Schoenberg's results [235, §6, thm.5] (*confer* [166, p.108-109]) also suggest certain finite positive roots of EDM entries produce EDMs; specifically,

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; [d_{ij}^{1/\alpha}] \in \mathbb{EDM}^N \quad \forall\, \alpha > 1 \tag{900}$$

The special case $\alpha = 2$ is of interest because it corresponds to absolute distance; *e.g.*,

$$D \in \mathbb{EDM}^N \Rightarrow \sqrt[\circ]{D} \in \mathbb{EDM}^N \tag{901}$$

Assuming that points constituting a corresponding list $X$ are distinct (864), then it follows: for $D \in \mathbb{S}_h^N$

$$\lim_{\alpha\to\infty}[d_{ij}^{1/\alpha}] \;=\; \lim_{\alpha\to\infty}[1 - e^{-\alpha\, d_{ij}}] \;=\; -E \;\triangleq\; \mathbf{1}\mathbf{1}^T - I \tag{902}$$

Negative elementary matrix $-E$ (§B.3) is relatively interior to the EDM cone (§6.6) and terminal to the respective trajectories (899) and (900) as functions of $\alpha$. Both trajectories are confined to the EDM cone; in engineering terms, the EDM cone is an *invariant set* [232] to either trajectory. Further, if $D$ is not an EDM but for some particular $\alpha_{\rm p}$ it becomes an EDM, then for all greater values of $\alpha$ it remains an EDM.

These preliminary findings lead one to speculate whether any concave nondecreasing composition of individual EDM entries $d_{ij}$ on $\mathbb{R}_+$ will produce another EDM; *e.g.*, empirical evidence suggests that given EDM $D$, for each fixed $\alpha \geq 1$ [*sic*] the composition $[\log_2(1+d_{ij}^{1/\alpha})]$ is also an EDM. Figure **90** illustrates that composition's concavity in $d_{ij}$ together with functions from (899) and (900).

### 5.10.1 EDM by elliptope

For some $\kappa \in \mathbb{R}_+$ and $C \in \mathbb{S}_+^N$ in the elliptope $\mathcal{E}^N$ (§5.9.1.0.1), Alfakih asserts any given EDM $D$ is expressible [7] [77, §31.5]

$$D = \kappa(\mathbf{1}\mathbf{1}^T - C) \in \mathbb{EDM}^N \tag{903}$$

This expression exhibits nonlinear combination of variables $\kappa$ and $C$. We therefore propose a different expression requiring redefinition of the elliptope (884) by parametrization;

$$\mathcal{E}_t^n \triangleq \mathbb{S}_+^n \cap \{\Phi \in \mathbb{S}^n \mid \delta(\Phi) = t\mathbf{1}\} \tag{904}$$

where, of course, $\mathcal{E}^n = \mathcal{E}_1^n$. Then any given EDM $D$ is expressible

$$D = t\mathbf{1}\mathbf{1}^T - \mathfrak{E} \in \mathbb{EDM}^N \tag{905}$$

which is linear in variables $t \in \mathbb{R}_+$ and $\mathfrak{E} \in \mathcal{E}_t^N$.

## 5.11 EDM indefiniteness

By the known result in §A.7.2 regarding a 0-valued entry on the main diagonal of a symmetric positive semidefinite matrix, there can be no positive nor negative semidefinite EDM except the $\mathbf{0}$ matrix because $\mathbb{EDM}^N \subseteq \mathbb{S}_h^N$ (708) and

$$\mathbb{S}_h^N \cap \mathbb{S}_+^N = \mathbf{0} \tag{906}$$

the origin. So when $D \in \mathbb{EDM}^N$, there can be no factorization $D = A^TA$ nor $-D = A^TA$. [249, §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: [249, §6.3]

$$-W^TDW \;=\; -\begin{bmatrix} V_{\mathcal N}^T \\ \mathbf{1}^T \end{bmatrix} D\,[\,V_{\mathcal N} \quad \mathbf{1}\,] \;=\; -\begin{bmatrix} V_{\mathcal N}^TDV_{\mathcal N} & V_{\mathcal N}^TD\mathbf{1} \\ \mathbf{1}^TDV_{\mathcal N} & \mathbf{1}^TD\mathbf{1} \end{bmatrix} \in \mathbb{S}^N \tag{907}$$

Because

$$W \triangleq [\,V_{\mathcal N} \quad \mathbf{1}\,] \in \mathbb{R}^{N\times N} \tag{908}$$

is full-rank, then (1316)

$$\text{inertia}(-D) = \text{inertia}\big(-W^TDW\big) \tag{909}$$

the congruence (907) 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 $-V_{\mathcal N}^TDV_{\mathcal N}$ and denote eigenvalues of the congruence $-W^TDW$ 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* [150, §4.3] [249, §6.4] [246, §IV.4.1]

$$\zeta_N \le \gamma_{N-1} \le \zeta_{N-1} \le \gamma_{N-2} \le \cdots \le \gamma_2 \le \zeta_2 \le \gamma_1 \le \zeta_1 \tag{910}$$

When $D\in\mathbb{EDM}^N$, then $\gamma_i \ge 0\ \forall\, i$ (1253) because $-V_{\mathcal N}^TDV_{\mathcal N}\succeq 0$ 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;[5.45] [77, §2.4.5]

[5.45] All the entries of the corresponding eigenvector must have the same sign with respect to each other [61, p.116] because that eigenvector is the *Perron vector* corresponding to the spectral radius; [150, §8.2.6] the predominant characteristic of negative [*sic*] matrices.

$$D \in \mathbb{EDM}^N \Rightarrow \left\{ \begin{array}{l} \lambda(-D)_i \;\geq\; 0\,, \quad i=1\ldots N-1 \\ \left( \sum\limits_{i=1}^{N} \lambda(-D)_i = 0 \right) \\ D \in \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N} \end{array} \right. \tag{911}$$

where the $\lambda(-D)_i$ are nonincreasingly ordered eigenvalues of $-D$ whose sum must be 0 only because $\operatorname{tr} D = 0$ [249, §5.1]. The eigenvalue summation condition, therefore, can be considered redundant. Even so, all these conditions are insufficient to determine whether some given $H \in \mathbb{S}_h^N$ is an EDM, as shown by counter-example.[5.46]

**5.11.1.0.1 Exercise.** *Spectral inequality.*
Prove whether it holds: for $D = [d_{ij}] \in \mathbb{EDM}^N$

$$\lambda(-D)_1 \geq d_{ij} \geq \lambda(-D)_{N-1} \qquad \forall\, i \neq j \tag{912}$$

▼

Terminology: a *spectral cone* is a convex cone containing all *eigenspectra* 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.

### 5.11.2 Spectral cones

Denoting the eigenvalues of Cayley-Menger matrix $\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix} \in \mathbb{S}^{N+1}$ by

$$\lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix}\right) \in \mathbb{R}^{N+1} \tag{913}$$

we have the Cayley-Menger form (§5.7.3.0.1) of necessary and sufficient conditions for $D \in \mathbb{EDM}^N$ from the literature: [133, §3][5.47] [53, §3] [77, §6.2]

[5.46] When $N=3$, for example, the symmetric hollow matrix

$$H = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 5 \\ 1 & 5 & 0 \end{bmatrix} \in \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N}$$

is not an EDM, although $\lambda(-H) = [5 \;\; 0.3723 \;\; -5.3723]^T$ conforms to (911).

[5.47] Recall: for $D \in \mathbb{S}_h^N$, $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0$ subsumes nonnegativity property 1 (§5.8.1).

(*confer* (728) (704))

$$D \in \mathbb{EDM}^N \Leftrightarrow \left\{ \begin{array}{l} \lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix}\right)_i \geq 0\ , \quad i = 1 \ldots N \\ D \in \mathbb{S}_h^N \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{c} -V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0 \\ D \in \mathbb{S}_h^N \end{array} \right. \tag{914}$$

These conditions say the Cayley-Menger form has one and only one negative eigenvalue. When $D$ is an EDM, eigenvalues $\lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix}\right)$ belong to that particular orthant in $\mathbb{R}^{N+1}$ having the $N+1^{\text{th}}$ coordinate as sole negative coordinate[5.48]:

$$\begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} = \operatorname{cone}\{e_1\ ,\ e_2\ ,\ \cdots\ e_N\ ,\ -e_{N+1}\} \tag{915}$$

#### 5.11.2.1 Cayley-Menger *versus* Schoenberg

Connection to the Schoenberg criterion (728) is made when the Cayley-Menger form is further partitioned:

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} \mathbf{1}^T \\ -D_{1,2:N} \end{bmatrix} \\ [\,\mathbf{1} \quad -D_{2:N,1}\,] & -D_{2:N,2:N} \end{bmatrix} \tag{916}$$

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; [15, §1] [159, §3] *id est*, (*confer* (862))

$$\begin{array}{c} D \in \mathbb{EDM}^N \\ \Leftrightarrow \\ -D_{2:N,2:N} - [\,\mathbf{1} \quad -D_{2:N,1}\,] \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1}^T \\ -D_{1,2:N} \end{bmatrix} = -2V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0 \\ \textbf{and} \\ D \in \mathbb{S}_h^N \end{array} \tag{917}$$

[5.48] Empirically, all except one entry of the corresponding eigenvector have the same sign with respect to each other.

Positive semidefiniteness of that Schur complement insures nonnegativity ($D\in\mathbb{R}_+^{N\times N}$, §5.8.1), whereas *complementary inertia* (1318) insures that lone negative eigenvalue of the Cayley-Menger form.

Now we apply results from chapter 2 with regard to polyhedral cones and their duals.

#### 5.11.2.2 Ordered eigenspectra

Conditions (914) specify eigenvalue membership to the smallest pointed polyhedral *spectral cone* for $\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbb{EDM}^N \end{bmatrix}$:

$$\begin{aligned}\mathcal{K}_\lambda &\triangleq \{\zeta\in\mathbb{R}^{N+1} \mid \zeta_1\geq\zeta_2\geq\cdots\geq\zeta_N\geq 0\geq\zeta_{N+1}\ ,\ \mathbf{1}^T\zeta=0\}\\ &= \mathcal{K}_\mathcal{M}\cap\begin{bmatrix}\mathbb{R}_+^N\\ \mathbb{R}_-\end{bmatrix}\cap\partial\mathcal{H}\\ &= \lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbb{EDM}^N \end{bmatrix}\right)\end{aligned}\tag{918}$$

where

$$\partial\mathcal{H}\triangleq\{\zeta\in\mathbb{R}^{N+1} \mid \mathbf{1}^T\zeta=0\}\tag{919}$$

is a hyperplane through the origin, and $\mathcal{K}_\mathcal{M}$ is the monotone cone (§2.13.9.4.2, implying ordered eigenspectra) which has nonempty interior but is not pointed;

$$\mathcal{K}_\mathcal{M}=\{\zeta\in\mathbb{R}^{N+1} \mid \zeta_1\geq\zeta_2\geq\cdots\geq\zeta_{N+1}\}\tag{377}$$

$$\mathcal{K}_\mathcal{M}^*=\{\,[\,e_1-e_2\quad e_2-e_3\quad\cdots\quad e_N-e_{N+1}\,]\,a \mid a\succeq 0\,\}\subset\mathbb{R}^{N+1}\tag{378}$$

So because of the hyperplane,

$$\dim\operatorname{aff}\mathcal{K}_\lambda\ =\ \dim\partial\mathcal{H}\ =\ N\tag{920}$$

indicating $\mathcal{K}_\lambda$ has empty interior. Defining

$$A\triangleq\begin{bmatrix} e_1^T-e_2^T\\ e_2^T-e_3^T\\ \vdots\\ e_N^T-e_{N+1}^T\end{bmatrix}\in\mathbb{R}^{N\times N+1}\,,\qquad B\triangleq\begin{bmatrix} e_1^T\\ e_2^T\\ \vdots\\ e_N^T\\ -e_{N+1}^T\end{bmatrix}\in\mathbb{R}^{N+1\times N+1}\tag{921}$$

we have the halfspace-description:

$$\mathcal{K}_\lambda = \{\zeta \in \mathbb{R}^{N+1} \mid A\zeta \succeq 0\,,\ B\zeta \succeq 0\,,\ \mathbf{1}^T\zeta = 0\} \tag{922}$$

From this and (385) we get a vertex-description for a pointed spectral cone having empty interior:

$$\mathcal{K}_\lambda = \left\{ V_{\mathcal{N}} \left( \begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} V_{\mathcal{N}} \right)^{\dagger} b \mid b \succeq 0 \right\} \tag{923}$$

where $V_{\mathcal{N}} \in \mathbb{R}^{N+1\times N}$, and where [*sic*]

$$\hat{B} = e_N^T \in \mathbb{R}^{1\times N+1} \tag{924}$$

and

$$\hat{A} = \begin{bmatrix} e_1^T - e_2^T \\ e_2^T - e_3^T \\ \vdots \\ e_{N-1}^T - e_N^T \end{bmatrix} \in \mathbb{R}^{N-1\times N+1} \tag{925}$$

hold those rows of $A$ and $B$ corresponding to conically independent rows in $\begin{bmatrix} A \\ B \end{bmatrix} V_{\mathcal{N}}$ .

Conditions (914) can be equivalently restated in terms of a spectral cone for Euclidean distance matrices:

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{ \begin{array}{l} \lambda\left( \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix} \right) \in \mathcal{K}_{\mathcal{M}} \cap \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} \cap \partial\mathcal{H} \\ D \in \mathbb{S}_h^N \end{array} \right. \tag{926}$$

Vertex-description of the dual spectral cone is, (272)

$$\begin{aligned} \mathcal{K}_\lambda^* &= \overline{\mathcal{K}_{\mathcal{M}}^* + \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix}^* + \partial\mathcal{H}^*} \subseteq \mathbb{R}^{N+1} \\ &= \left\{ \begin{bmatrix} A^T & B^T & \mathbf{1} & -\mathbf{1} \end{bmatrix} b \mid b \succeq 0 \right\} = \left\{ \begin{bmatrix} \hat{A}^T & \hat{B}^T & \mathbf{1} & -\mathbf{1} \end{bmatrix} a \mid a \succeq 0 \right\} \end{aligned} \tag{927}$$

From (923) and (386) we get a halfspace-description:

$$\mathcal{K}_\lambda^* = \{y \in \mathbb{R}^{N+1} \mid (V_{\mathcal{N}}^T [\,\hat{A}^T \;\; \hat{B}^T\,])^\dagger V_{\mathcal{N}}^T y \succeq 0\} \tag{928}$$

This polyhedral dual spectral cone $\mathcal{K}_\lambda^*$ is closed, convex, has nonempty interior because $\mathcal{K}_\lambda$ is pointed, but is not pointed because $\mathcal{K}_\lambda$ has empty interior.

#### 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.[5.49] Then things simplify: Conditions (914) now specify eigenvalue membership to the spectral cone

$$\begin{aligned}\lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbb{EDM}^N \end{bmatrix}\right) &= \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} \cap \partial\mathcal{H} \\ &= \{\zeta \in \mathbb{R}^{N+1} \mid B\zeta \succeq 0\,,\; \mathbf{1}^T\zeta = 0\}\end{aligned} \tag{929}$$

where $B$ is defined in (921), and $\partial\mathcal{H}$ in (919). From (385) we get a vertex-description for a pointed spectral cone having empty interior:

$$\begin{aligned}\lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbb{EDM}^N \end{bmatrix}\right) &= \left\{V_{\mathcal{N}}(\tilde{B}V_{\mathcal{N}})^\dagger\, b \mid b \succeq 0\right\} \\ &= \left\{\begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix} b \mid b \succeq 0\right\}\end{aligned} \tag{930}$$

where $V_{\mathcal{N}} \in \mathbb{R}^{N+1\times N}$ and

$$\tilde{B} \triangleq \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_N^T \end{bmatrix} \in \mathbb{R}^{N\times N+1} \tag{931}$$

holds only those rows of $B$ corresponding to conically independent rows in $BV_{\mathcal{N}}$.

[5.49] Eigenspectra ordering (represented by a cone having monotone description such as (918)) becomes benign in (1138), 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 presorted eigenvalues, (914) can be equivalently restated

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{ \begin{array}{l} \lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix}\right) \in \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} \cap \partial\mathcal{H} \\ D \in \mathbb{S}_h^N \end{array} \right. \tag{932}$$

Vertex-description of the dual spectral cone is, (272)

$$\begin{aligned} \lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbb{EDM}^N \end{bmatrix}\right)^* &= \begin{bmatrix} \mathbb{R}_+^N \\ \mathbb{R}_- \end{bmatrix} + \partial\mathcal{H}^* \subseteq \mathbb{R}^{N+1} \\ &= \left\{ \begin{bmatrix} B^T & \mathbf{1} & -\mathbf{1} \end{bmatrix} b \mid b \succeq 0 \right\} = \left\{ \begin{bmatrix} \tilde{B}^T & \mathbf{1} & -\mathbf{1} \end{bmatrix} a \mid a \succeq 0 \right\} \end{aligned} \tag{933}$$

From (386) we get a halfspace-description:

$$\begin{aligned} \lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbb{EDM}^N \end{bmatrix}\right)^* &= \{y \in \mathbb{R}^{N+1} \mid (V_{\mathcal{N}}^T \tilde{B}^T)^\dagger V_{\mathcal{N}}^T y \succeq 0\} \\ &= \{y \in \mathbb{R}^{N+1} \mid [\, I \;\; -\mathbf{1}\,]\, y \succeq 0\} \end{aligned} \tag{934}$$

This polyhedral dual spectral cone is closed, convex, has nonempty interior 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 self-dual. 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 self-dual positive semidefinite cone is also self-dual.

## 5.12 List reconstruction

The traditional term *multidimensional scaling*[5.50] [188] [71] [264] [69] [191] [61] refers to any reconstruction of a list $X\in\mathbb{R}^{n\times N}$ in Euclidean space from interpoint distance information, possibly incomplete (§6.4), ordinal (§5.13.2), or specified perhaps only by bounding-constraints (§5.4.2.2.7) [265]. 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* [115] which analyzes the correlation matrix (§5.9.1.0.1); [39, §22] `a.k.a`, *Karhunen−Loéve transform* in the digital signal processing literature. Isometric reconstruction (§5.5.3) of point list $X$ is best performed by eigen decomposition of a Gram matrix; for then, numerical errors of factorization are easily spotted in the eigenvalues.

We now consider how rotation/reflection and translation invariance factor into a reconstruction.

### 5.12.1 $x_1$ at the origin. $V_{\mathcal{N}}$

At the stage of reconstruction, we have $D\in\mathbb{EDM}^N$ and we wish to find a generating list (§2.3.2) for $\mathcal{P}-\alpha$ by factoring positive semidefinite $-V_{\mathcal{N}}^TDV_{\mathcal{N}}$ (897) as suggested in §5.9.1.0.4. One way to factor $-V_{\mathcal{N}}^TDV_{\mathcal{N}}$ is via diagonalization of symmetric matrices; [249, §5.6] [150] (§A.5.2, §A.3)

$$-V_{\mathcal{N}}^TDV_{\mathcal{N}} \triangleq Q\Lambda Q^T \tag{935}$$

$$Q\Lambda Q^T \succeq 0 \;\Leftrightarrow\; \Lambda\succeq 0 \tag{936}$$

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 (844);

$$-V_{\mathcal{N}}^TDV_{\mathcal{N}} \;=\; 2V_{\mathcal{N}}^TX^TXV_{\mathcal{N}} \;\triangleq\; Q\sqrt{\Lambda}\,Q_{\rm p}^TQ_{\rm p}\sqrt{\Lambda}\,Q^T \tag{937}$$

[5.50] Scaling [262] *means making a scale,* i.e., *a numerical representation of qualitative data. If the scale is multidimensional, it's* multidimensional scaling. −Jan de Leeuw
When the metric is Euclidean distance, then reconstruction is termed *metric* multidimensional scaling.

where $\sqrt{\Lambda}\, Q_{\mathrm p}^T Q_{\mathrm p} \sqrt{\Lambda} \triangleq \Lambda = \sqrt{\Lambda}\sqrt{\Lambda}$ and where $Q_{\mathrm p} \in \mathbb{R}^{n\times N-1}$ is unknown as is its dimension $n$. Rotation/reflection is accounted for by $Q_{\mathrm p}$ yet only its first $r$ columns are necessarily orthonormal.[5.51] Assuming membership to the unit simplex $y \in \mathcal{S}$ (894), then point $p = X\sqrt{2}V_{\mathcal N}\, y = Q_{\mathrm p}\sqrt{\Lambda}\, Q^T y$ in $\mathbb{R}^n$ belongs to the translated polyhedron

$$\mathcal{P} - x_1 \tag{938}$$

whose generating list constitutes the columns of (838)

$$\begin{aligned}\begin{bmatrix}\mathbf{0} & X\sqrt{2}V_{\mathcal N}\end{bmatrix} &= \begin{bmatrix}\mathbf{0} & Q_{\mathrm p}\sqrt{\Lambda}\, Q^T\end{bmatrix} \in \mathbb{R}^{n\times N} \\ &= \begin{bmatrix}\mathbf{0} & x_2 - x_1 & x_3 - x_1 & \cdots & x_N - x_1\end{bmatrix}\end{aligned} \tag{939}$$

The scaled auxiliary matrix $V_{\mathcal N}$ represents that translation. A simple choice for $Q_{\mathrm p}$ has $n$ set to $N-1$; *id est*, $Q_{\mathrm p} = I$. Ideally, each member of the generating list has at most $r$ nonzero entries; $r$ being, affine dimension

$$\operatorname{rank} V_{\mathcal N}^T D V_{\mathcal N} = \operatorname{rank} Q_{\mathrm p}\sqrt{\Lambda}\, Q^T = \operatorname{rank}\Lambda = r \tag{940}$$

Each member then has at least $N-1-r$ zeros in its higher-dimensional coordinates because $r \le N-1$. (850) To truncate those zeros, choose $n$ equal to affine dimension which is the smallest $n$ possible because $XV_{\mathcal N}$ has rank $r \le n$ (846).[5.52] In that case, the simplest choice for $Q_{\mathrm p}$ is $[\, I \quad \mathbf{0}\,]$ having dimensions $r \times N-1$.

We may wish to verify the list (939) found from the diagonalization of $-V_{\mathcal N}^T D V_{\mathcal N}$. Because of rotation/reflection and translation invariance (§5.5), EDM $D$ can be uniquely made from that list by calculating: (709)

[5.51] Recall $r$ signifies affine dimension. $Q_{\mathrm p}$ is not necessarily an orthogonal matrix. $Q_{\mathrm p}$ is constrained such that only its first $r$ columns are necessarily orthonormal because there are only $r$ nonzero eigenvalues in $\Lambda$ when $-V_{\mathcal N}^T D V_{\mathcal N}$ has rank $r$ (§5.7.1.1). Remaining columns of $Q_{\mathrm p}$ are arbitrary.

[5.52] If we write $Q^T = \begin{bmatrix} q_1^T \\ \vdots \\ q_{N-1}^T \end{bmatrix}$ as rowwise eigenvectors, $\Lambda = \begin{bmatrix} \lambda_1 & & & & \mathbf{0} \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \\ \mathbf{0} & & & & 0 \end{bmatrix}$ in terms of eigenvalues, and $Q_{\mathrm p} = \begin{bmatrix} q_{\mathrm{p}_1} \cdots q_{\mathrm{p}_{N-1}} \end{bmatrix}$ as column vectors, then $Q_{\mathrm p}\sqrt{\Lambda}\, Q^T = \sum_{i=1}^{r} \sqrt{\lambda_i}\, q_{\mathrm{p}_i} q_i^T$ is a sum of $r$ linearly independent rank-one matrices (§B.1.1). Hence the summation has rank $r$.

$$\mathbf{D}(X) \;=\; \mathbf{D}(X[\,\mathbf{0}\quad \sqrt{2}V_{\mathcal{N}}\,]) \;=\; \mathbf{D}(Q_{\rm p}[\,\mathbf{0}\quad \sqrt{\Lambda}\,Q^T\,]) \;=\; \mathbf{D}([\,\mathbf{0}\quad \sqrt{\Lambda}\,Q^T\,]) \tag{941}$$

This suggests a way to find EDM $D$ given $-V_{\mathcal{N}}^TDV_{\mathcal{N}}$ ; (*confer* (822))

$$D = \begin{bmatrix} 0 \\ \delta\big(-V_{\mathcal{N}}^TDV_{\mathcal{N}}\big) \end{bmatrix}\mathbf{1}^T + \mathbf{1}\begin{bmatrix} 0 & \delta\big(-V_{\mathcal{N}}^TDV_{\mathcal{N}}\big)^T \end{bmatrix} - 2\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & -V_{\mathcal{N}}^TDV_{\mathcal{N}} \end{bmatrix} \tag{729}$$

### 5.12.2 0 geometric center. $V$

Alternatively, we may perform reconstruction using instead the auxiliary matrix $V$ (§B.4.1), corresponding to the polyhedron

$$\mathcal{P} - \alpha_c \tag{942}$$

whose geometric center has been translated to the origin. Redimensioning diagonalization factors $Q\,,\Lambda\in\mathbb{R}^{N\times N}$ and unknown $Q_{\rm p}\in\mathbb{R}^{n\times N}$, (845)

$$-VDV \;=\; 2VX^TXV \;\triangleq\; Q\sqrt{\Lambda}\,Q_{\rm p}^TQ_{\rm p}\sqrt{\Lambda}\,Q^T \;\triangleq\; Q\Lambda Q^T \tag{943}$$

where the geometrically centered generating list constitutes (*confer* (939))

$$\begin{aligned} XV \;&=\; \tfrac{1}{\sqrt{2}}\,Q_{\rm p}\sqrt{\Lambda}\,Q^T \in \mathbb{R}^{n\times N} \\ &=\; [\,x_1 - \tfrac{1}{N}X\mathbf{1}\quad x_2 - \tfrac{1}{N}X\mathbf{1}\quad x_3 - \tfrac{1}{N}X\mathbf{1}\quad\cdots\quad x_N - \tfrac{1}{N}X\mathbf{1}\,] \end{aligned} \tag{944}$$

where $\alpha_c = \frac{1}{N}X\mathbf{1}$. (§5.5.1.0.1) The simplest choice for $Q_{\rm p}$ is $[\,I\quad \mathbf{0}\,]\in\mathbb{R}^{r\times N}$.

Now EDM $D$ can be uniquely made from the list found, by calculating: (709)

$$\mathbf{D}(X) \;=\; \mathbf{D}(XV) \;=\; \mathbf{D}(\frac{1}{\sqrt{2}}\,Q_{\rm p}\sqrt{\Lambda}\,Q^T) \;=\; \mathbf{D}(\sqrt{\Lambda}\,Q^T)\frac{1}{2} \tag{945}$$

This EDM is, of course, identical to (941). Similarly to (729), from $-VDV$ we can find EDM $D$ ; (*confer* (809))

$$D = \delta(-VDV\tfrac{1}{2})\mathbf{1}^T + \mathbf{1}\delta(-VDV\tfrac{1}{2})^T - 2(-VDV\tfrac{1}{2}) \tag{735}$$

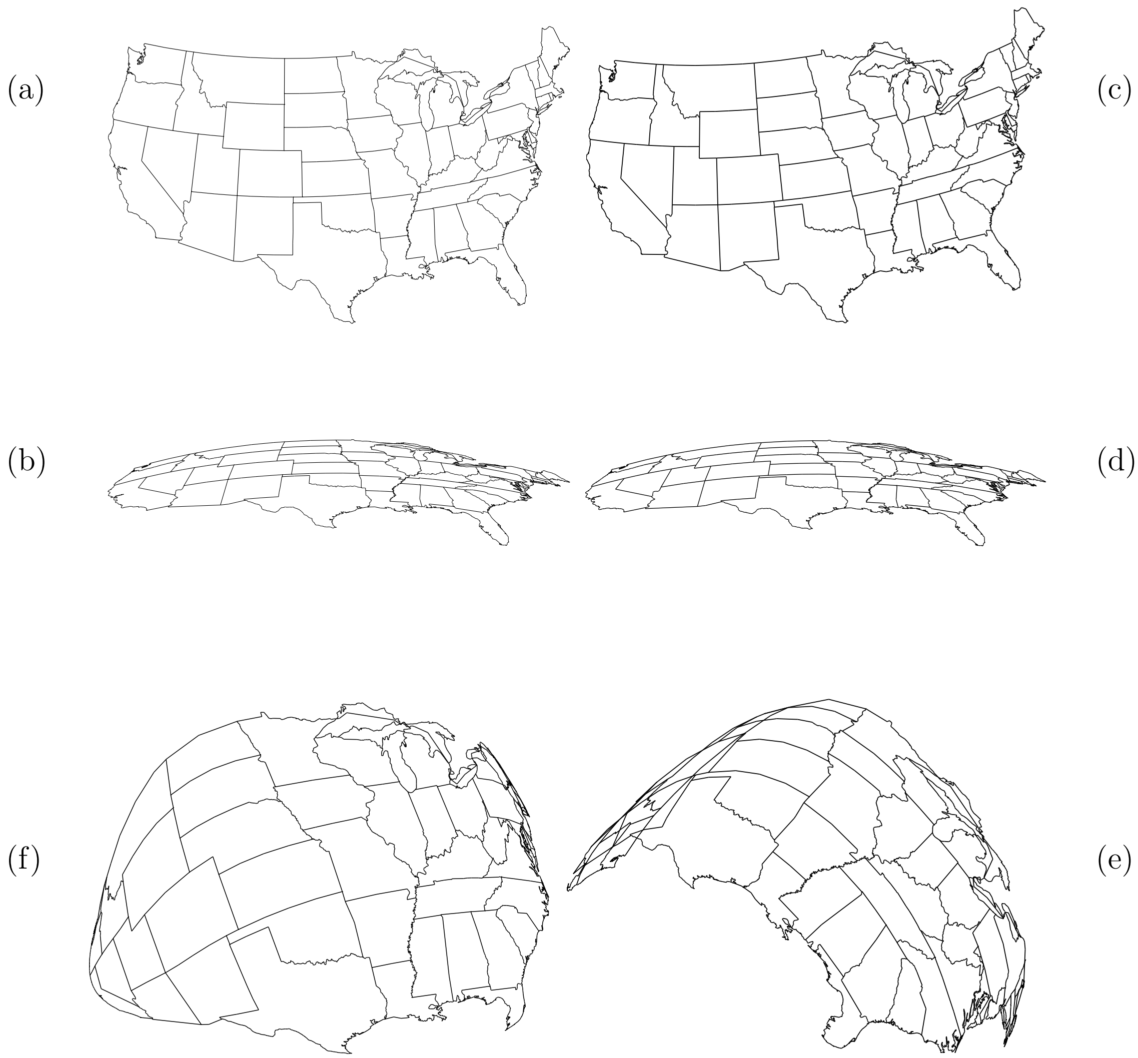



Figure 91: Map of United States of America showing some state boundaries and the Great Lakes. All plots made using 5020 connected points. Any difference in scale in (a) through (d) is an artifact of plotting routine.
**(a)** shows original map made from decimated (latitude, longitude) data.
**(b)** Original map data rotated (freehand) to highlight curvature of Earth.
**(c)** Map isometrically reconstructed from the EDM.
**(d)** Same reconstructed map illustrating curvature.
**(e)(f)** Two views of one isotonic reconstruction; problem (954) with no sort constraint $\Pi\,\underline{d}$ (and no hidden line removal).

# 5.13 Reconstruction examples

## 5.13.1 Isometric reconstruction

**5.13.1.0.1 Example.** *Map of the USA.*
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 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 the MATLAB Mapping Toolbox; the conversion to Cartesian coordinates $(x\,,y\,,z)$ via:

$$\begin{aligned}\phi &\triangleq \pi/2 - \text{latitude}\\ \theta &\triangleq \text{longitude}\\ x &= \sin(\phi)\cos(\theta)\\ y &= \sin(\phi)\sin(\theta)\\ z &= \cos(\phi)\end{aligned} \tag{946}$$

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 (937) are shown in Figure **91**(a)-(d). The MATLAB code is in §F.3.1. The eigenvalues computed for (935) are

$$\lambda(-V_{\mathcal{N}}^T D V_{\mathcal{N}}) = [199.8 \;\; 152.3 \;\; 2.465 \;\; 0 \;\; 0 \;\; 0 \;\; \cdots \;]^T \tag{947}$$

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, the EDM $D$ for three points is unknown:

$$D = [d_{ij}] = \begin{bmatrix} 0 & d_{12} & d_{13} \\ d_{12} & 0 & d_{23} \\ d_{13} & d_{23} & 0 \end{bmatrix} \in \mathbb{S}_h^{\mathbf{3}} \tag{698}$$

but the comparative data is available:

$$d_{13} \geq d_{23} \geq d_{12} \tag{948}$$

With the vectorization $\underline{d} = [d_{12}\ d_{13}\ d_{23}]^T \in \mathbb{R}^{\mathbf{3}}$, we express the comparative distance relationship as the nonincreasing sorting

$$\Pi\,\underline{d} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} d_{12} \\ d_{13} \\ d_{23} \end{bmatrix} = \begin{bmatrix} d_{13} \\ d_{23} \\ d_{12} \end{bmatrix} \in \mathcal{K}_{\mathcal{M}+} \tag{949}$$

where $\Pi$ is a given permutation matrix expressing the known sorting action on the entries of unknown EDM $D$, and $\mathcal{K}_{\mathcal{M}+}$ is the monotone nonnegative cone (§2.13.9.4.1)

$$\mathcal{K}_{\mathcal{M}+} \triangleq \{z \mid z_1 \geq z_2 \geq \cdots \geq z_{N(N-1)/2} \geq 0\} \subseteq \mathbb{R}_+^{N(N-1)/2} \tag{370}$$

where $N(N-1)/2 = 3$ for the present example. From the sorted vectorization (949) we create the *sort-index matrix*

$$O = \begin{bmatrix} 0 & 1^2 & 3^2 \\ 1^2 & 0 & 2^2 \\ 3^2 & 2^2 & 0 \end{bmatrix} \in \mathbb{S}_h^{\mathbf{3}} \cap \mathbb{R}_+^{\mathbf{3}\times\mathbf{3}} \tag{950}$$

generally defined

$$O_{ij} \triangleq k^2 \mid d_{ij} = \left(\Xi\,\Pi\,\underline{d}\right)_k\ , \qquad j \neq i \tag{951}$$

where $\Xi$ is a permutation matrix (1507) 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 (955).

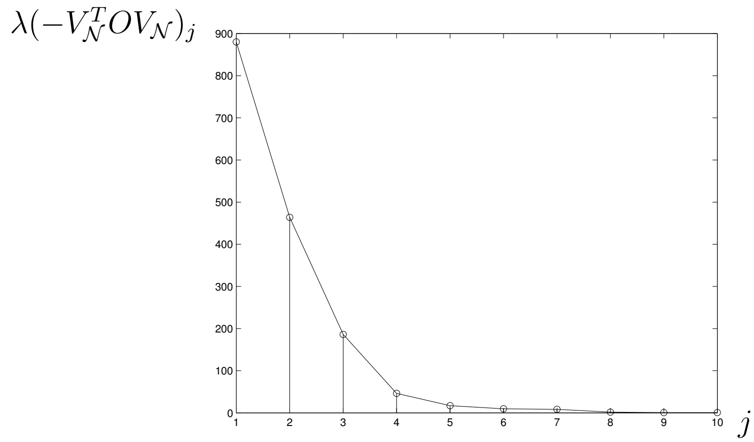



Figure 92: Largest ten eigenvalues of $-V_{\mathcal{N}}^T O V_{\mathcal{N}}$ for map of USA, sorted by nonincreasing value. In the code (§F.3.2), we normalize $O$ by $(N(N-1)/2)^2$.

Any process of reconstruction that leaves comparative distance information intact is called *ordinal multidimensional scaling* or *isotonic reconstruction.* Beyond rotation, reflection, and translation error, (§5.5) list reconstruction by isotonic reconstruction is subject to error in absolute scale (*dilation*) and distance ratio. Yet Borg & Groenen argue: [39, §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 better.

#### 5.13.2.1 Isotonic map of the USA

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 (947), 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 **92**:

$$\lambda(-V_{\mathcal{N}}^T O V_{\mathcal{N}}) = [880.1\;\; 463.9\;\; 186.1\;\; 46.20\;\; 17.12\;\; 9.625\;\; 8.257\;\; 1.701\;\; 0.7128\;\; 0.6460\;\; \cdots\;]^T \tag{952}$$

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 (a problem explored more in §7) while maintaining the known comparative distance relationship; *e.g.*, given permutation matrix $\Pi$ expressing the known sorting action on the entries $\underline{d}$ of unknown $D \in \mathbb{S}_h^N$, (63)

$$\underline{d} \triangleq \frac{1}{\sqrt{2}} \operatorname{dvec} D = \begin{bmatrix} d_{12} \\ d_{13} \\ d_{23} \\ d_{14} \\ d_{24} \\ d_{34} \\ \vdots \\ d_{N-1,N} \end{bmatrix} \in \mathbb{R}^{N(N-1)/2} \tag{953}$$

we can make the sort-index matrix $O$ input to the optimization problem

$$\begin{array}{cl} \underset{D}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-O)V_{\mathcal{N}}\|_{\mathrm{F}} \\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} \leq 3 \\ & \Pi\, \underline{d} \in \mathcal{K}_{\mathcal{M}+} \\ & D \in \mathbb{EDM}^N \end{array} \tag{954}$$

that finds the EDM $D$ $\big($corresponding to affine dimension not exceeding 3 in isomorphic $\operatorname{dvec} \mathbb{EDM}^N \cap \Pi^T \mathcal{K}_{\mathcal{M}+}\big)$ closest to $O$ in the sense of Schoenberg (728).

Analytical solution to this problem, ignoring the sort constraint $\Pi\, \underline{d} \in \mathcal{K}_{\mathcal{M}+}$, is known [264]: we get the convex optimization [*sic*] (§7.1)

$$\begin{array}{cl} \underset{D}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-O)V_{\mathcal{N}}\|_{\mathrm{F}} \\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} \leq 3 \\ & D \in \mathbb{EDM}^N \end{array} \tag{955}$$

Only the three largest nonnegative eigenvalues in (952) need be retained to make list (939); the rest are discarded. The reconstruction from EDM $D$ found in this manner is plotted in Figure **91**(e)(f) from which it becomes obvious that inclusion of the sort constraint is necessary for isotonic reconstruction.

That sort constraint demands: any optimal solution $D^\star$ must possess the known comparative distance relationship that produces the original ordinal distance data $O$ (951). 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 (955). 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 (954) into two problems we know how to solve and then alternate their solution until convergence:

$$\begin{array}{cll}\underset{D}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-O)V_{\mathcal{N}}\|_{\mathrm{F}} & \\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} \leq 3 & \quad\text{(a)}\\ & D\in\mathbb{EDM}^N & \\[1em] \underset{\sigma}{\text{minimize}} & \|\sigma-\Pi\,\underline{d}\| & \quad\text{(b)}\\ \text{subject to} & \sigma\in\mathcal{K}_{\mathcal{M}+} & \end{array} \tag{956}$$

where the sort-index matrix $O$ $\big($a given constant in (a)$\big)$ becomes an implicit vectorized variable $\underline{o}_i$ solving the $i^{\text{th}}$ instance of (956b)

$$\underline{o}_i \triangleq \Pi^T\sigma^\star = \frac{1}{\sqrt{2}}\operatorname{dvec} O_i \in \mathbb{R}^{N(N-1)/2}\,, \qquad i\in\{1,2,3\ldots\} \tag{957}$$

As mentioned in discussion of relaxed problem (955), a closed-form solution to problem (956a) exists. Only the first iteration of (956a) sees the original sort-index matrix $O$ whose entries are nonnegative whole numbers; *id est*, $O_0 = O\in\mathbb{S}_h^N\cap\mathbb{R}_+^{N\times N}$ (951). Subsequent iterations $i$ take the previous solution of (956b) as input

$$O_i = \operatorname{dvec}^{-1}(\sqrt{2}\,\underline{o}_i)\in\mathbb{S}^N \tag{958}$$

real successors to the sort-index matrix $O$.

New problem (956b) finds the unique minimum-distance projection of $\Pi\,\underline{d}$ on the monotone nonnegative cone $\mathcal{K}_{\mathcal{M}+}$. By defining

$$Y^{\dagger T} \triangleq [\,e_1-e_2 \quad e_2-e_3 \quad e_3-e_4 \quad \cdots \quad e_m\,]\in\mathbb{R}^{m\times m} \tag{371}$$

where $m \overset{\Delta}{=} N(N-1)/2$, we may rewrite (956b) as an equivalent *quadratic program*; a convex optimization problem [46, §4] in terms of the halfspace-description of $\mathcal{K}_{\mathcal{M}+}$ :

$$\begin{array}{cl}\underset{\sigma}{\text{minimize}} & (\sigma - \Pi\, \underline{d})^T(\sigma - \Pi\, \underline{d}) \\ \text{subject to} & Y^\dagger \sigma \succeq 0\end{array} \tag{959}$$

This quadratic program can be converted to a semidefinite program via Schur-form (§3.1.7.2); we get the equivalent problem

$$\begin{array}{cl}\underset{t\in\mathbb{R}\,,\,\sigma}{\text{minimize}} & t \\ \text{subject to} & \left[\begin{array}{cc} t\,I & \sigma - \Pi\,\underline{d} \\ (\sigma - \Pi\,\underline{d})^T & 1 \end{array}\right] \succeq 0 \\ & Y^\dagger \sigma \succeq 0\end{array} \tag{960}$$

#### 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 (956a) is equivalent to a projection of nonincreasingly ordered eigenvalues on a subset of the nonnegative orthant:

$$\begin{array}{cl}\underset{D}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-O)V_{\mathcal{N}}\|_{\mathrm{F}} \\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} \leq 3 \\ & D \in \mathbb{EDM}^N\end{array} \quad\equiv\quad \begin{array}{cl}\underset{\Upsilon}{\text{minimize}} & \|\Upsilon - \Lambda\|_{\mathrm{F}} \\ \text{subject to} & \delta(\Upsilon) \in \left[\begin{array}{c}\mathbb{R}_+^{\mathbf{3}} \\ \mathbf{0}\end{array}\right]\end{array} \tag{961}$$

where $-V_{\mathcal{N}}^T D V_{\mathcal{N}} \overset{\Delta}{=} U\Upsilon U^T \in \mathbb{S}^{N-1}$ and $-V_{\mathcal{N}}^T O V_{\mathcal{N}} \overset{\Delta}{=} Q\Lambda Q^T \in \mathbb{S}^{N-1}$ are ordered diagonalizations (§A.5). It so happens: optimal orthogonal $U^\star$ always equals $Q$ given. Linear operator $T(A) = U^{\star T}AU^\star$, acting on square matrix $A$, is a bijective isometry because the Frobenius norm is orthogonally invariant (40). This isometric isomorphism $T$ thus maps a nonconvex problem to a convex one that preserves distance.

Secondly, the second half (956b) of the *alternation* takes place in a different vector space; $\mathbb{S}_h^N$ (*versus* $\mathbb{S}^{N-1}$). From §5.6 we know these two vector spaces are related by an isomorphism, $\mathbb{S}^{N-1}=\mathbf{V}_{\mathcal{N}}(\mathbb{S}_h^N)$ (827), but not by an isometry.

We have, therefore, no guarantee from theory of alternating projection that the alternation (956) converges to a point, in the set of all EDMs corresponding to affine dimension not in excess of 3, belonging to $\operatorname{dvec}\mathbb{EDM}^N\cap\Pi^T\mathcal{K}_{\mathcal{M}+}$ .

#### 5.13.2.4 Interlude

We have not implemented the second half (959) of alternation (956) for USA map data because memory-demands exceed the capability of our 32-bit laptop 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 (948) is available. Given known set of indices $\mathcal{I}$

$$\begin{array}{ll}\underset{D}{\text{minimize}} & \operatorname{rank} VDV\\ \text{subject to} & d_{ij}\leq d_{kl}\leq d_{mn} \qquad \forall(i,j,k,l,m,n)\in\mathcal{I}\\ & D\in\mathbb{EDM}^N\end{array}\tag{962}$$

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 (955). 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 (962) is practically always different from the optimal solution set for its counterpart, problem (954).

## 5.14 Fifth property of Euclidean metric

We continue now with the question raised in §5.3 regarding the 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 $D\in\mathbb{EDM}^N$ 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 $N=3$, $-V_{\mathcal{N}}^T D V_{\mathcal{N}}\succeq 0$ (867) and $D\in\mathbb{S}_h^{\mathbf{3}}$ are necessary and sufficient conditions for $D$ to be an EDM. By (869), 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 $i,j\in\{1,2,3\}$

$$\begin{array}{l}\sqrt{d_{ij}}\geq 0\,,\quad i\neq j\\ \sqrt{d_{ij}}=0\,,\quad i=j\\ \sqrt{d_{ij}}=\sqrt{d_{ji}}\\ \sqrt{d_{ij}}\;\leq\;\sqrt{d_{ik}}+\sqrt{d_{kj}}\;,\quad i\neq j\neq k\end{array}\quad\Leftrightarrow\quad\begin{array}{r}-V_{\mathcal{N}}^T D V_{\mathcal{N}}\succeq 0\\ D\in\mathbb{S}_h^{\mathbf{3}}\end{array}\quad\Leftrightarrow\quad D\in\mathbb{EDM}^{\mathbf{3}}\tag{963}$$

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 (869a) was revealed within the leading principal $2\times 2$ submatrix of $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ when positive semidefinite. Our choice of the leading principal submatrix was arbitrary; actually, a unique triangle inequality like (772) corresponds to any one of the $(N-1)!/\big(2!(N-1-2)!\big)$ principal $2\times 2$ submatrices.[5.53] Assuming $D\in\mathbb{S}_h^{\mathbf{4}}$ and $-V_{\mathcal{N}}^T D V_{\mathcal{N}}\in\mathbb{S}^{\mathbf{3}}$, then by the *positive (semi)definite principal submatrices*

[5.53] There are fewer principal $2\times 2$ submatrices in $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ 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) *theorem* (§A.3.1.0.4) it is sufficient to prove all $d_{ij}$ are nonnegative, all triangle inequalities are satisfied, and $\det(-V_{\mathcal{N}}^{T} D V_{\mathcal{N}})$ is nonnegative. When $N=4$, in other words, that nonnegative determinant becomes the fifth and last Euclidean metric requirement for $D\in\mathbb{EDM}^N$. We now endeavor to ascribe geometric meaning to it.

#### 5.14.2.1 Nonnegative determinant

By (778) when $D\in\mathbb{EDM}^{\mathbf{4}}$, $-V_{\mathcal{N}}^{T} D V_{\mathcal{N}}$ is equal to inner product (773),

$$\Theta^T\Theta = \begin{bmatrix} d_{12} & \sqrt{d_{12}d_{13}}\cos\theta_{213} & \sqrt{d_{12}d_{14}}\cos\theta_{214} \\ \sqrt{d_{12}d_{13}}\cos\theta_{213} & d_{13} & \sqrt{d_{13}d_{14}}\cos\theta_{314} \\ \sqrt{d_{12}d_{14}}\cos\theta_{214} & \sqrt{d_{13}d_{14}}\cos\theta_{314} & d_{14} \end{bmatrix} \tag{964}$$

Because Euclidean space is an inner-product space, the more concise inner-product form of the determinant is admitted;

$$\det(\Theta^T\Theta) = -d_{12}d_{13}d_{14}\big(\cos(\theta_{213})^2+\cos(\theta_{214})^2+\cos(\theta_{314})^2 - 2\cos\theta_{213}\cos\theta_{214}\cos\theta_{314} - 1\big) \tag{965}$$

The determinant is nonnegative if and only if

$$\begin{array}{c} \cos\theta_{214}\cos\theta_{314} - \sqrt{\sin(\theta_{214})^2\sin(\theta_{314})^2} \;\leq\; \cos\theta_{213} \;\leq\; \cos\theta_{214}\cos\theta_{314} + \sqrt{\sin(\theta_{214})^2\sin(\theta_{314})^2} \\ \Leftrightarrow \\ \cos\theta_{213}\cos\theta_{314} - \sqrt{\sin(\theta_{213})^2\sin(\theta_{314})^2} \;\leq\; \cos\theta_{214} \;\leq\; \cos\theta_{213}\cos\theta_{314} + \sqrt{\sin(\theta_{213})^2\sin(\theta_{314})^2} \\ \Leftrightarrow \\ \cos\theta_{213}\cos\theta_{214} - \sqrt{\sin(\theta_{213})^2\sin(\theta_{214})^2} \;\leq\; \cos\theta_{314} \;\leq\; \cos\theta_{213}\cos\theta_{214} + \sqrt{\sin(\theta_{213})^2\sin(\theta_{214})^2} \end{array} \tag{966}$$

which simplifies, for $0\leq\theta_{i1\ell}\,,\theta_{\ell 1j}\,,\theta_{i1j}\leq\pi$ and all $i\neq j\neq\ell\in\{2,3,4\}$, to

$$\cos(\theta_{i1\ell}+\theta_{\ell 1j}) \;\leq\; \cos\theta_{i1j} \;\leq\; \cos(\theta_{i1\ell}-\theta_{\ell 1j}) \tag{967}$$

Analogously to triangle inequality (881), the determinant is 0 upon equality on either side of (967) which is tight. Inequality (967) can be equivalently written linearly as a "triangle inequality", but between relative angles [301, §1.4];

$$\begin{aligned} |\theta_{i1\ell}-\theta_{\ell 1j}| \;\leq\; \theta_{i1j} \;&\leq\; \theta_{i1\ell}+\theta_{\ell 1j} \\ \theta_{i1\ell}+\theta_{\ell 1j}+\theta_{i1j} \;&\leq\; 2\pi \\ 0\leq\theta_{i1\ell}\,,\theta_{\ell 1j}\,,\theta_{i1j} &\leq\pi \end{aligned} \tag{968}$$

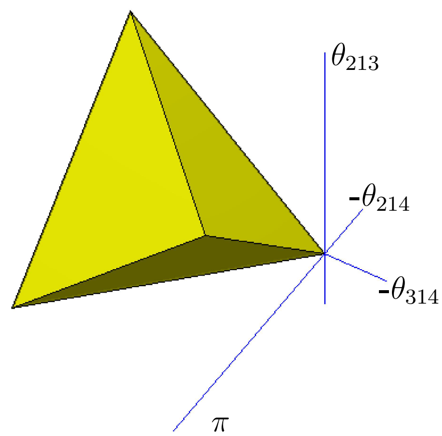



Figure 93: The *relative-angle inequality tetrahedron* (969) bounding $\mathbb{EDM}^4$ is regular; drawn in entirety. Each angle $\theta$ (770) must belong to this solid to be realizable.

Generalizing this:

**5.14.2.1.1 Fifth property of Euclidean metric - restatement.** *Relative-angle inequality.* (*confer* §5.3.1.0.1) [36] [37, p.17, p.107] [171, §3.1] 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

$$\begin{array}{cr} |\theta_{ik\ell} - \theta_{\ell kj}| \;\leq\; \theta_{ikj} \;\leq\; \theta_{ik\ell} + \theta_{\ell kj} & \text{(a)} \\ \theta_{ik\ell} + \theta_{\ell kj} + \theta_{ikj} \leq 2\pi & \text{(b)} \\ 0 \leq \theta_{ik\ell}, \theta_{\ell kj}, \theta_{ikj} \leq \pi & \text{(c)} \end{array} \tag{969}$$

where $\theta_{ikj} = \theta_{jki}$ is the angle between vectors at vertex $x_k$ (as defined in (770) and illustrated in Figure **76**), must be satisfied at each point $x_k$ regardless of affine dimension. $\diamond$

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 (969). 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)

$$\begin{array}{l}\text{Four Euclidean metric properties (§5.2).}\\ \text{Angle } \theta \text{ inequality (703) or (969).}\end{array} \;\Leftrightarrow\; \begin{array}{c}-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0\\ D \in \mathbb{S}_h^{\mathbf{4}}\end{array} \;\Leftrightarrow\; D = \mathbf{D}(\Theta) \in \mathbb{EDM}^{\mathbf{4}} \tag{970}$$

The relative-angle inequality, for this case, is illustrated in Figure **93**.

#### 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 [*sic*] in $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ . 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 (883) is necessary and sufficient to certify realizability; (§5.4.3.1)

$$\begin{array}{l}\text{Euclidean metric property 1 (§5.2).}\\ \text{Angle matrix inequality } \Omega \succeq 0 \text{ (779).}\end{array} \;\Leftrightarrow\; \begin{array}{c}-V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0\\ D \in \mathbb{S}_h^N\end{array} \;\Leftrightarrow\; D = \mathbf{D}(\Omega\,, d) \in \mathbb{EDM}^N \tag{971}$$

Like matrix criteria (704), (728), and (883), 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 $N=4$, the three-dimensional analogue to triangle & distance is tetrahedron & facet-area, while in the case $N=5$ the four-dimensional analogue is polychoron & facet-volume, *ad infinitum.* For $N$ points, $N+1$ metric properties are required.

#### 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-squared: $\mathrm{c}_1, \mathrm{c}_2, \mathrm{c}_3, \mathrm{c}_4$. Then analogous to metric property 4, we may write a tight[5.54] area inequality for the facets

$$\sqrt{\mathrm{c}_i} \leq \sqrt{\mathrm{c}_j} + \sqrt{\mathrm{c}_k} + \sqrt{\mathrm{c}_\ell}\ , \quad i \neq j \neq k \neq \ell \in \{1,2,3,4\} \tag{972}$$

which is a generalized "triangle" inequality [166, §1.1] that follows from

$$\sqrt{\mathrm{c}_i} = \sqrt{\mathrm{c}_j}\cos\varphi_{ij} + \sqrt{\mathrm{c}_k}\cos\varphi_{ik} + \sqrt{\mathrm{c}_\ell}\cos\varphi_{i\ell} \tag{973}$$

[177] [282, *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-squared of the $i^{\text{th}}$ triangular facet has a convenient formula in terms of $D_i \in \mathbb{EDM}^{N-1}$ the EDM corresponding to that particular facet: From the *Cayley-Menger determinant*[5.55] for simplices, [282] [87] [112, §4] [60, §3.3] the $i^{\text{th}}$ facet area-squared for $i \in \{1 \ldots N\}$ is (§A.4.1)

[5.54] The upper bound is met when all angles in (973) are simultaneously $0$; that occurs, for example, if one point is relatively interior to the convex hull of the three remaining.

[5.55] whose foremost characteristic is: the determinant vanishes if and only if affine dimension does not equal penultimate cardinality; *id est,* $\det\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix} = 0 \Leftrightarrow r < N-1$ where $D$ is any EDM (§5.7.3.0.1). Otherwise, the determinant is negative.

$$
\begin{aligned}
\mathrm{c}_i &= \frac{-1}{2^{N-2}(N-2)!^2}\,\det\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D_i \end{bmatrix} && (974)\\
&= \frac{(-1)^N}{2^{N-2}(N-2)!^2}\,\det D_i\left(\mathbf{1}^T D_i^{-1}\mathbf{1}\right) && (975)\\
&= \frac{(-1)^N}{2^{N-2}(N-2)!^2}\,\mathbf{1}^T\mathrm{cof}(D_i)^T\mathbf{1} && (976)
\end{aligned}
$$

where $D_i$ is the $i^{\text{th}}$ principal $N-1\times N-1$ submatrix[5.56] of $D\in\mathbb{EDM}^N$, and $\mathrm{cof}(D_i)$ is the $N-1\times N-1$ matrix of *cofactors* [249, §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 (972) are conditions necessary for $D$ to be an EDM. Prove their sufficiency in conjunction with the remaining three Euclidean metric properties. ▼

#### 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.[5.57] The polychoron has five $(N!/(4!(N-4)!))$ facets, each of them a general tetrahedron whose volume-squared $\mathrm{c}_i$ is calculated using the same formula; (974) where $D$ is the EDM corresponding to the polychoron, and $D_i$ is the EDM corresponding to the $i^{\text{th}}$ facet (the principal $4\times 4$ submatrix of $D\in\mathbb{EDM}^N$ corresponding to the $i^{\text{th}}$ tetrahedron). The analogue to triangle & distance is now polychoron & facet-volume. We could then write another generalized "triangle" inequality like (972) but in terms of facet volume; [287, §IV]

$$\sqrt{\mathrm{c}_i} \;\leq\; \sqrt{\mathrm{c}_j} + \sqrt{\mathrm{c}_k} + \sqrt{\mathrm{c}_\ell} + \sqrt{\mathrm{c}_m}\;, \quad i\neq j\neq k\neq \ell\neq m\in\{1\ldots 5\} \qquad (977)$$

[5.56] Every principal submatrix of an EDM remains an EDM. [171, §4.1]

[5.57] The simplest polychoron is called a pentatope [282]; a regular simplex hence convex. (A *pentahedron* is a three-dimensional body having five vertices.)

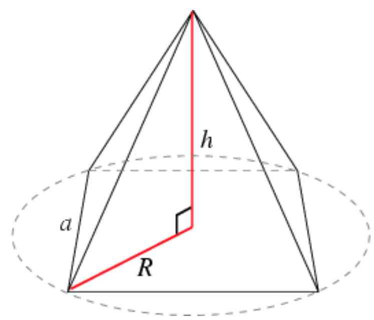



Figure 94: Length of one-dimensional face $a$ equals height $h=a=1$ of this nonsimplicial pyramid in $\mathbb{R}^3$ with square base inscribed in a circle of radius $R$ centered at the origin. [282, *Pyramid*]

**5.14.3.2.1 Exercise.** *Sufficiency for an EDM of five points.*
For $N=5$, triangle (distance) inequality (§5.2), area inequality (972), and volume inequality (977) 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. [299, §2.1] [208, p.173] [168] [169] [123] [124] Volume is a concept germane to $\mathbb{R}^3$; in higher dimensions it is called *content.* Applying the *EDM assertion* (§5.9.1.0.3) and a result given in [46, §8.3.1], a general nonempty simplex (§2.12.3) in $\mathbb{R}^{N-1}$ corresponding to an EDM $D\in\mathbb{S}_h^N$ has content

$$\sqrt{\mathrm{c}} = \mathrm{content}(\mathcal{S})\sqrt{\det(-V_{\mathcal{N}}^T D V_{\mathcal{N}})} \tag{978}$$

where the content-squared of the unit simplex $\mathcal{S}\subset\mathbb{R}^{N-1}$ is proportional to its Cayley-Menger determinant;

$$\mathrm{content}(\mathcal{S})^2 = \frac{-1}{2^{N-1}(N-1)!^2}\det\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbf{D}([\,\mathbf{0}\;\; e_1\;\; e_2\;\cdots\; e_{N-1}\,]) \end{bmatrix} \tag{979}$$

where $e_i\in\mathbb{R}^{N-1}$ and the EDM operator used is $\mathbf{D}(X)$ (709).

**5.14.3.3.1 Example.** *Pyramid.*
A formula for volume of a pyramid is known;[5.58] it is $\frac{1}{3}$ the product of its base area with its height. [164] The pyramid in Figure **94** has volume $\frac{1}{3}$. To find its volume using EDMs, we must first decompose the pyramid into simplicial parts. 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,

$$X = \begin{bmatrix} 1/2 & 1/2 & -1/2 & 0 \\ 1/2 & -1/2 & -1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{n\times N} \tag{980}$$

where $N = n+1$ for any nonempty simplex in $\mathbb{R}^n$. The volume of this simplex is half that of the entire pyramid; *id est*, $\sqrt{\mathrm{c}} = \frac{1}{6}$ found by evaluating (978). □

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. [249, §5.1] When $N = 4$ and $\det(\Theta^T\Theta)$ is 0, that means one or more eigenvalues of $\Theta^T\Theta \in \mathbb{R}^{\mathbf{3\times 3}}$ are 0. The determinant will go to 0 whenever equality is attained on either side of (703), (969a), or (969b), meaning that a tetrahedron has collapsed to a lower affine dimension; *id est*, $r = \operatorname{rank}\Theta^T\Theta = \operatorname{rank}\Theta$ 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 (703) or (969a) or (969b).

[5.58] Pyramid volume is independent of the paramount vertex position as long as its height remains constant.

#### 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 **75** where $N=4$, distance-square $d_{14}$ is ascertainable from the fifth Euclidean metric property. Because all distances in (701) are known except $\sqrt{d_{14}}$ , then $\cos\theta_{123}=0$ and $\theta_{324}=0$ result from identity (770). Applying (703),

$$\begin{array}{rcccl} \cos(\theta_{123}+\theta_{324}) & \leq & \cos\theta_{124} & \leq & \cos(\theta_{123}-\theta_{324}) \\ 0 & \leq & \cos\theta_{124} & \leq & 0 \end{array} \tag{981}$$

It follows again from (770) that $d_{14}$ can only be 2. As explained in this subsection, affine dimension $r$ cannot exceed $N-2$ because equality is attained in (981). □

# Chapter 6

# EDM cone

*For $N>3$, the cone of EDMs is no longer a circular cone and the geometry becomes complicated...*

−Hayden, Wells, Liu, & Tarazaga (1991) [134, §3]

In the subspace of symmetric matrices $\mathbb{S}^N$, we know the convex cone of Euclidean distance matrices $\mathbb{EDM}^N$ (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 nor negative semidefinite EDM. (906) [171]

$$\mathbb{EDM}^N \cap \mathbb{S}_+^N = \mathbf{0} \tag{982}$$

Even so, the two convex cones can be related. In §6.8.1 we prove the new equality

$$\mathbb{EDM}^N = \mathbb{S}_h^N \cap \left(\mathbb{S}_c^{N\perp} - \mathbb{S}_+^N\right) \qquad (1074)$$



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

a resemblance to EDM definition (709) where

$$\mathbb{S}_h^N \triangleq \left\{A \in \mathbb{S}^N \mid \delta(A) = \mathbf{0}\right\} \tag{56}$$

is the symmetric hollow subspace (§2.2.3) and where

$$\mathbb{S}_c^{N\perp} = \{u\mathbf{1}^T + \mathbf{1}u^T \mid u\in\mathbb{R}^N\} \tag{1768}$$

is the orthogonal complement of the geometric center subspace (§E.7.2.0.2)

$$\mathbb{S}_c^N \triangleq \{Y\in\mathbb{S}^N \mid Y\mathbf{1} = \mathbf{0}\} \tag{1766}$$

### 6.0.1 gravity

Equality (1074) is equally important as the known isomorphisms (816) (817) (828) (829) relating the EDM cone $\mathbb{EDM}^N$ to an $N(N-1)/2$-dimensional face of $\mathbb{S}_+^N$ (§5.6.1.1), or to $\mathbb{S}_+^{N-1}$ (§5.6.2.1).[6.1] Those isomorphisms have never led to this equality (1074) relating the whole cones $\mathbb{EDM}^N$ and $\mathbb{S}_+^N$.

Equality (1074) is not obvious from the various EDM matrix definitions such as (709) or (1000) because inclusion must be proved algebraically in order to establish equality; $\mathbb{EDM}^N \supseteq \mathbb{S}_h^N \cap (\mathbb{S}_c^{N\perp} - \mathbb{S}_+^N)$. We will instead prove (1074) using purely geometric methods.

### 6.0.2 highlight

In §6.8.1.7 we show: the Schoenberg criterion for discriminating Euclidean distance matrices

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{\begin{array}{c} -V_{\mathcal{N}}^T D V_{\mathcal{N}} \in \mathbb{S}_+^{N-1} \\ D \in \mathbb{S}_h^N \end{array}\right. \tag{728}$$

is a discretized membership relation (§2.13.4) between the EDM cone and its ordinary dual.

[6.1] Because both positive semidefinite cones are frequently in play, dimension is notated.

## 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:

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{\begin{array}{c} -VDV \in \mathbb{S}_+^N \\ D \in \mathbb{S}_h^N \end{array}\right. \tag{733}$$

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 $\mathbb{EDM}^N$ because any pair of EDMs satisfies the definition of a convex cone (144); *videlicet*, for each and every $\zeta_1, \zeta_2 \geq 0$ (§A.3.1.0.2)

$$\begin{array}{c} \zeta_1 VD_1V + \zeta_2 VD_2V \succeq 0 \\ \zeta_1 D_1 + \zeta_2 D_2 \in \mathbb{S}_h^N \end{array} \;\Leftarrow\; \begin{array}{cc} VD_1V \succeq 0\,, & VD_2V \succeq 0 \\ D_1 \in \mathbb{S}_h^N\,, & D_2 \in \mathbb{S}_h^N \end{array} \tag{983}$$

and convex cones are invariant to inverse affine transformation [230, p.22].

**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 *EDM cone*: for $N > 0$

$$\begin{aligned} \mathbb{EDM}^N &\triangleq \left\{D \in \mathbb{S}_h^N \mid -VDV \in \mathbb{S}_+^N\right\} \\ &= \bigcap_{z \in \mathcal{N}(\mathbf{1}^T)} \left\{D \in \mathbb{S}^N \mid \langle zz^T, -D\rangle \geq 0\,,\;\; \delta(D) = \mathbf{0}\right\} \end{aligned} \tag{984}$$

The EDM cone in isomorphic $\mathbb{R}^{N(N+1)/2}$ [*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 $\mathbb{EDM}^N$ has empty interior with respect to $\mathbb{S}^N$ because it is confined to the symmetric hollow subspace $\mathbb{S}_h^N$. The EDM cone relative interior comprises

$$\begin{aligned} \operatorname{rel\,int} \mathbb{EDM}^N &= \bigcap_{z \in \mathcal{N}(\mathbf{1}^T)} \left\{D \in \mathbb{S}^N \mid \langle zz^T, -D\rangle > 0\,,\;\; \delta(D) = \mathbf{0}\right\} \\ &= \left\{D \in \mathbb{EDM}^N \mid \operatorname{rank}(VDV) = N-1\right\} \end{aligned} \tag{985}$$

while its relative boundary comprises

$$\begin{aligned} \operatorname{rel} \partial \mathbb{EDM}^N &= \left\{D \in \mathbb{EDM}^N \mid \langle zz^T, -D\rangle = 0 \;\text{ for some } z \in \mathcal{N}(\mathbf{1}^T)\right\} \\ &= \left\{D \in \mathbb{EDM}^N \mid \operatorname{rank}(VDV) < N-1\right\} \end{aligned} \tag{986}$$

△

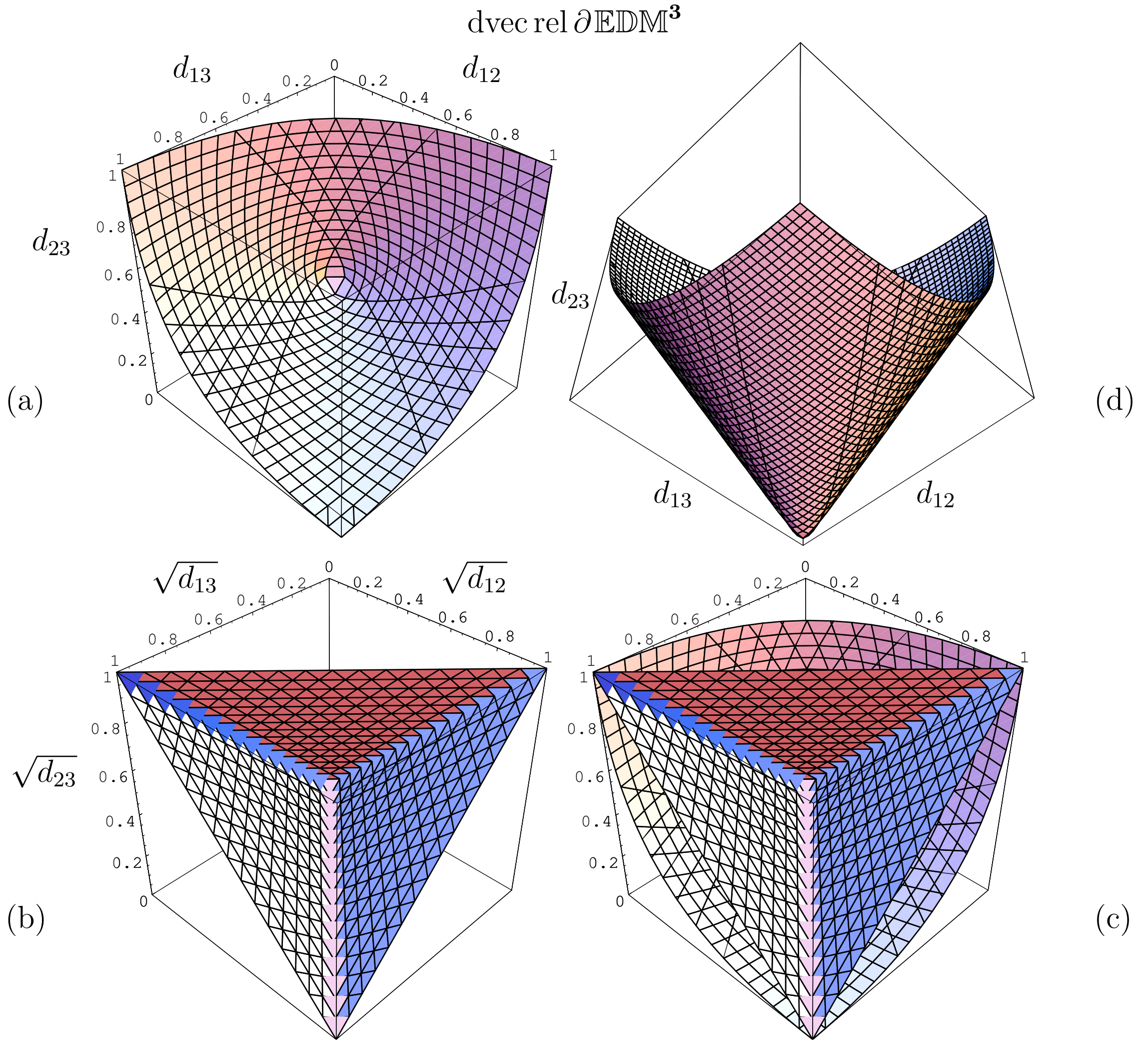



Figure 95: Relative boundary (tiled) of EDM cone $\mathbb{EDM}^3$ drawn truncated in isometrically isomorphic subspace $\mathbb{R}^3$. **(a)** EDM cone drawn in usual distance-square coordinates $d_{ij}$. View is from interior toward origin. Unlike positive semidefinite cone, EDM cone is not self-dual, 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 (870) whose partial boundaries each contain origin. Cone geometry becomes "complicated" (nonpolyhedral) in higher dimension. [134, §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 $\mathbf{0}$, pointed $\mathbb{EDM}^3$ is a circular cone having axis of revolution $\text{dvec}(-E) = \text{dvec}(\mathbf{1}\mathbf{1}^T - I)$ (902) (63). Rounded vertex is plot artifact.

This cone is more easily visualized in the isomorphic vector subspace $\mathbb{R}^{N(N-1)/2}$ corresponding to $\mathbb{S}_h^N$ :

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 **103**.

The EDM cone in the case $N=3$ is a circular cone in $\mathbb{R}^{\mathbf{3}}$ illustrated in Figure **95**(a)(d); rather, the set of all matrices

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} \\ d_{12} & 0 & d_{23} \\ d_{13} & d_{23} & 0 \end{bmatrix} \in \mathbb{EDM}^{\mathbf{3}} \tag{987}$$

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; (963). Thus triangle inequality property 4 describes three halfspaces (870) whose intersection makes a polyhedral cone in $\mathbb{R}^{\mathbf{3}}$ of realizable $\sqrt{d_{ij}}$ (absolute distance); an isomorphic subspace representation of the set of all EDMs $D$ in the natural coordinates

$$\sqrt[\circ]{D} \triangleq \begin{bmatrix} 0 & \sqrt{d_{12}} & \sqrt{d_{13}} \\ \sqrt{d_{12}} & 0 & \sqrt{d_{23}} \\ \sqrt{d_{13}} & \sqrt{d_{23}} & 0 \end{bmatrix} \tag{988}$$

illustrated in Figure **95**(b).

## 6.2 Polyhedral bounds

The convex cone of EDMs is nonpolyhedral in $d_{ij}$ for $N>2$; *e.g.*, Figure **95**(a). 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 polyhedral cone in Figure **95**(b); polyhedral because an intersection of three halfspaces in natural coordinates $\sqrt{d_{ij}}$ is provided by triangle inequalities (870). This polyhedral cone implicitly encompasses necessary and sufficient metric properties: nonnegativity, self-distance, symmetry, and triangle inequality.

(b) (c)

(a)

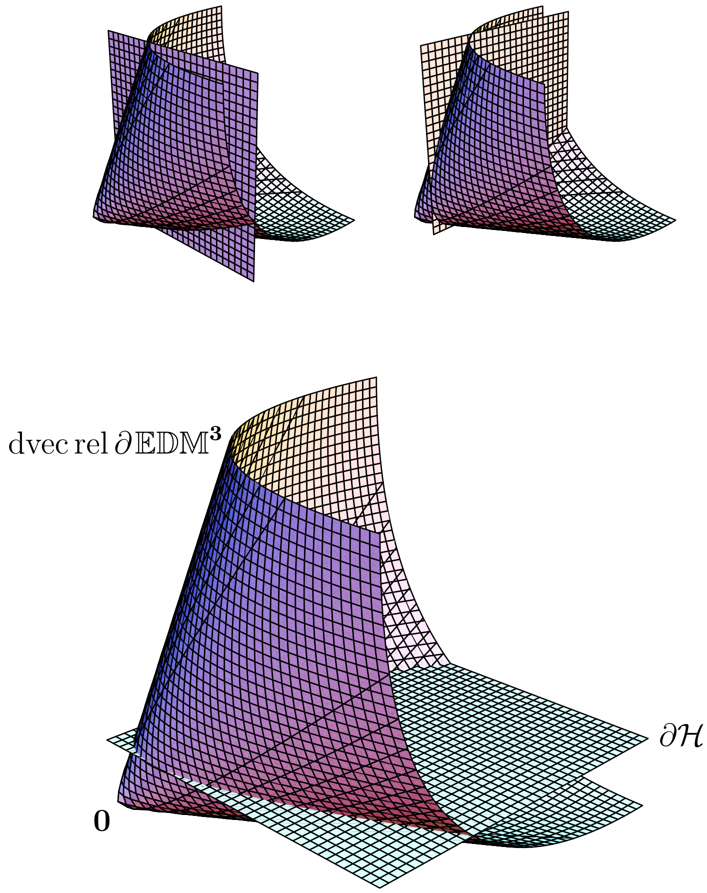



Figure 96: **(a)** In isometrically isomorphic subspace $\mathbb{R}^3$, intersection of $\mathbb{EDM}^3$ with hyperplane $\partial\mathcal{H}$ representing one fixed symmetric entry $d_{23}=\kappa$ (both drawn truncated, rounded vertex is artifact of plot). EDMs in this dimension corresponding to affine dimension 1 comprise relative boundary of EDM cone (§6.6). 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 $\operatorname{rel}\partial\mathbb{EDM}^3$ by Lemma 2.8.0.0.1. **(b)** $d_{13}=\kappa$. **(c)** $d_{12}=\kappa$.

$N=4$. Relative-angle inequality (969) together with four Euclidean metric properties are necessary and sufficient tests for realizability of tetrahedra. (970) Albeit relative angles $\theta_{ikj}$ (770) are nonlinear functions of the $d_{ij}$ , relative-angle inequality provides a regular tetrahedron in $\mathbb{R}^{\mathbf{3}}$ [*sic*] (Figure **93**) bounding angles $\theta_{ikj}$ at vertex $x_k$ consistently with $\mathbb{EDM}^{\mathbf{4}}$ .[6.2]

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 the molecular conformation problem (Figure **3**) or multidimensional scaling, [68] [267] 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 **95**(b)): [61, §4.5.2]

$$\sqrt{\mathbb{EDM}^N} \triangleq \{\sqrt[\circ]{D} \mid D\in\mathbb{EDM}^N\} \tag{989}$$

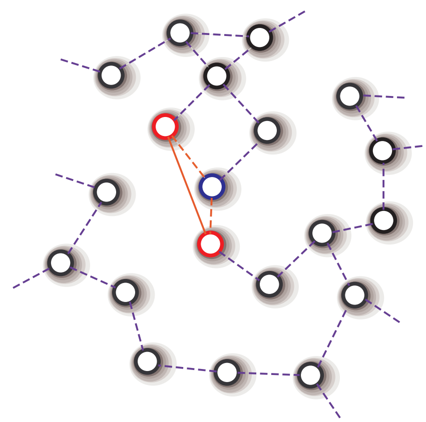
where $\sqrt[\circ]{D}$ is defined as in (988). It is a cone simply because any cone is completely constituted by rays emanating from the origin: (§2.7) Any given ray $\{\zeta\Gamma\in\mathbb{R}^{N(N-1)/2} \mid \zeta\geq 0\}$ remains a ray under entrywise square root: $\{\sqrt{\zeta\Gamma}\in\mathbb{R}^{N(N-1)/2} \mid \zeta\geq 0\}$. Because of how $\sqrt{\mathbb{EDM}^N}$ is defined, it is obvious that (*confer* §5.10)

$$D\in\mathbb{EDM}^N \;\Leftrightarrow\; \sqrt[\circ]{D}\in\sqrt{\mathbb{EDM}^N} \tag{990}$$

Were $\sqrt{\mathbb{EDM}^N}$ convex, then given $\sqrt[\circ]{D_1}\,,\sqrt[\circ]{D_2}\in\sqrt{\mathbb{EDM}^N}$ we would expect their conic combination $\sqrt[\circ]{D_1}+\sqrt[\circ]{D_2}$ to be a member of $\sqrt{\mathbb{EDM}^N}$. That is easily proven false by counter-example via (990), for then $(\sqrt[\circ]{D_1}+\sqrt[\circ]{D_2})\circ(\sqrt[\circ]{D_1}+\sqrt[\circ]{D_2})$ would need to be a member of $\mathbb{EDM}^N$.

Notwithstanding, in §7.2.1 we learn how to transform a nonconvex proximity problem in the natural coordinates $\sqrt{d_{ij}}$ to a convex optimization.

[6.2] Still, property-4 triangle inequalities (870) corresponding to each principal $3\times 3$ submatrix of $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ demand that the corresponding $\sqrt{d_{ij}}$ belong to a polyhedral cone like that in Figure **95**(b).

(a)  2 nearest neighbors

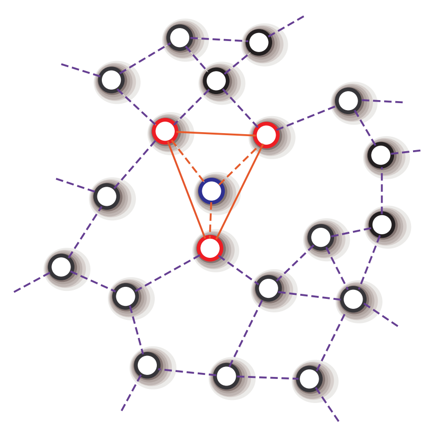
(b)  3 nearest neighbors

Figure 97: Neighborhood graph (dashed) with dimensionless EDM subgraph completion (solid) superimposed (but not covering dashed). Local view of a few dense samples ○ 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* [156, §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.

## 6.4 a geometry of completion

Intriguing is the question of whether the list in $X \in \mathbb{R}^{n\times N}$ (65) may be reconstructed given an incomplete noiseless EDM, and under what circumstances reconstruction is unique. [1] [2] [3] [4] [6] [15] [153] [159] [170] [171] [172]

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 $\mathbb{EDM}^N$ in isomorphic subspace $\mathbb{R}^{N(N-1)/2}$ with as many hyperplanes as there are fixed symmetric entries. (Depicted in Figure **96**(a) is an intersection of the EDM cone $\mathbb{EDM}^{\mathbf{3}}$ with a single hyperplane representing the set of all EDMs having one fixed symmetric entry.) 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* §2.9.2.6.2) that can be as small as a point. Indeed, Trosset remarks: [266, §1] *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.*

**6.4.0.0.1 Example.** *Diffusing, uncompacting, unfurling, unfolding.* [281] 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 self-intersections. [281, §2.2] The physical process is intuitively described as unfurling, unfolding, or *unraveling.* In particular instances, the process is a sort of flattening by stretching until taut (but not by crushing); *e.g.*, unfurling 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 **97** 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 **5**). 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 **97**.[6.3] 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 **98**(a). A corresponding Euclidean distance matrix $D=[d_{ij}\ ,\ i,j=1\ldots N]$ employing only 2 nearest neighbors is banded having the incomplete form

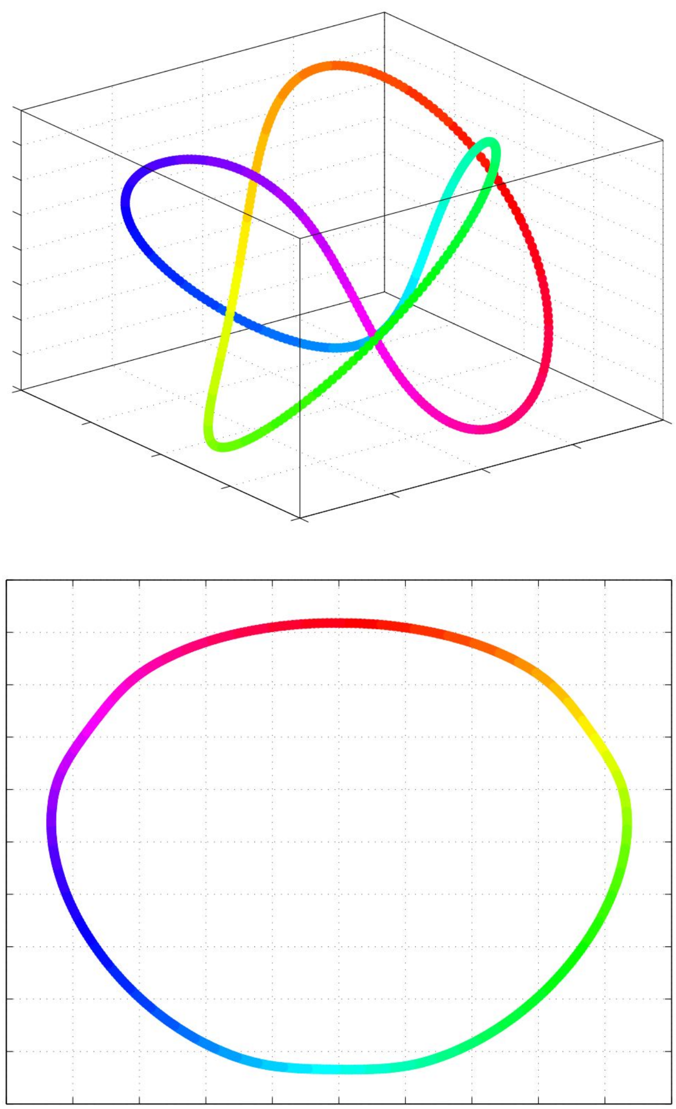
$$D \;=\; \begin{bmatrix} 0 & \check{d}_{12} & \check{d}_{13} & ? & \cdots & ? & \check{d}_{1,N-1} & \check{d}_{1N} \\ \check{d}_{12} & 0 & \check{d}_{23} & \check{d}_{24} & \ddots & ? & ? & \check{d}_{2N} \\ \check{d}_{13} & \check{d}_{23} & 0 & \check{d}_{34} & \ddots & ? & ? & ? \\ ? & \check{d}_{24} & \check{d}_{34} & 0 & \ddots & \ddots & ? & ? \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & ? \\ ? & ? & ? & \ddots & \ddots & 0 & \check{d}_{N-2,N-1} & \check{d}_{N-2,N} \\ \check{d}_{1,N-1} & ? & ? & ? & \ddots & \check{d}_{N-2,N-1} & 0 & \check{d}_{N-1,N} \\ \check{d}_{1N} & \check{d}_{2N} & ? & ? & ? & \check{d}_{N-2,N} & \check{d}_{N-1,N} & 0 \end{bmatrix} \tag{991}$$

where $\check{d}_{ij}$ denotes a given fixed distance-square. The unfurling algorithm can be expressed as an optimization problem; constrained distance-square maximization:

$$\begin{array}{cll} \underset{D}{\text{maximize}} & \langle -V\,,\,D\rangle & \\ \text{subject to} & \langle D\,,\,e_ie_j^T+e_je_i^T\rangle\frac{1}{2}=\check{d}_{ij} & \forall(i,j)\in\mathcal{I} \\ & \operatorname{rank}(VDV)=2 & \\ & D\in\mathbb{EDM}^N & \end{array} \tag{992}$$

[6.3] 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).

(a)



(b)

Figure 98: **(a)** *Trefoil knot* in $\mathbb{R}^3$ from Weinberger & Saul [281]. **(b)** Topological transformation algorithm employing 4 nearest neighbors and $N = 539$ samples reduces affine dimension of knot to $r = 2$. Choosing instead 2 nearest neighbors would make this embedding more circular.

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 (991), where $V \in \mathbb{R}^{N\times N}$ is the geometric centering matrix (§B.4.1), and where

$$\langle -V\,,\,D\rangle \;=\; \operatorname{tr}(-VDV) \;=\; 2\operatorname{tr}G \;=\; \frac{1}{N}\sum_{i,j} d_{ij} \qquad (734)$$

where $G$ is the Gram matrix producing $D$ assuming $G\mathbf{1}=\mathbf{0}$.

If we ignore the (rank) constraint on affine dimension, then problem (992) becomes convex, a corresponding solution $D^\star$ can be found, and a nearest rank-2 solution can be found by ordered eigen decomposition of $-VD^\star V$ followed by *spectral projection* (§7.1.3) on $\begin{bmatrix}\mathbb{R}^{\mathbf{2}}\\ \mathbf{0}\end{bmatrix} \subset \mathbb{R}^N$. 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 4 nearest neighbors is drawn in Figure **98**(b); a low-dimensional embedding of the trefoil knot.

This problem (992) can, of course, be written equivalently in terms of Gram matrix $G$, facilitated by (740); *videlicet*, for $\Phi_{ij}$ as in (707)

$$\begin{array}{cll}\underset{G\in\mathbb{S}_c^N}{\text{maximize}} & \langle I\,,\,G\rangle & \\ \text{subject to} & \langle G\,,\,\Phi_{ij}\rangle = \check{d}_{ij} & \forall(i,j)\in\mathcal{I} \\ & \operatorname{rank}G = 2 & \\ & G\succeq 0 & \end{array} \qquad (993)$$

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 5.4.2.2.4) Confining $G$ to the geometric center subspace suffers no loss of generality and serves no theoretical purpose; numerically, this implicit constraint $G\mathbf{1}=\mathbf{0}$ keeps $G$ independent of its translation-invariant subspace $\mathbb{S}_c^{N\perp}$ (§5.5.1.1, Figure **105**) so as not to become unbounded. $\square$

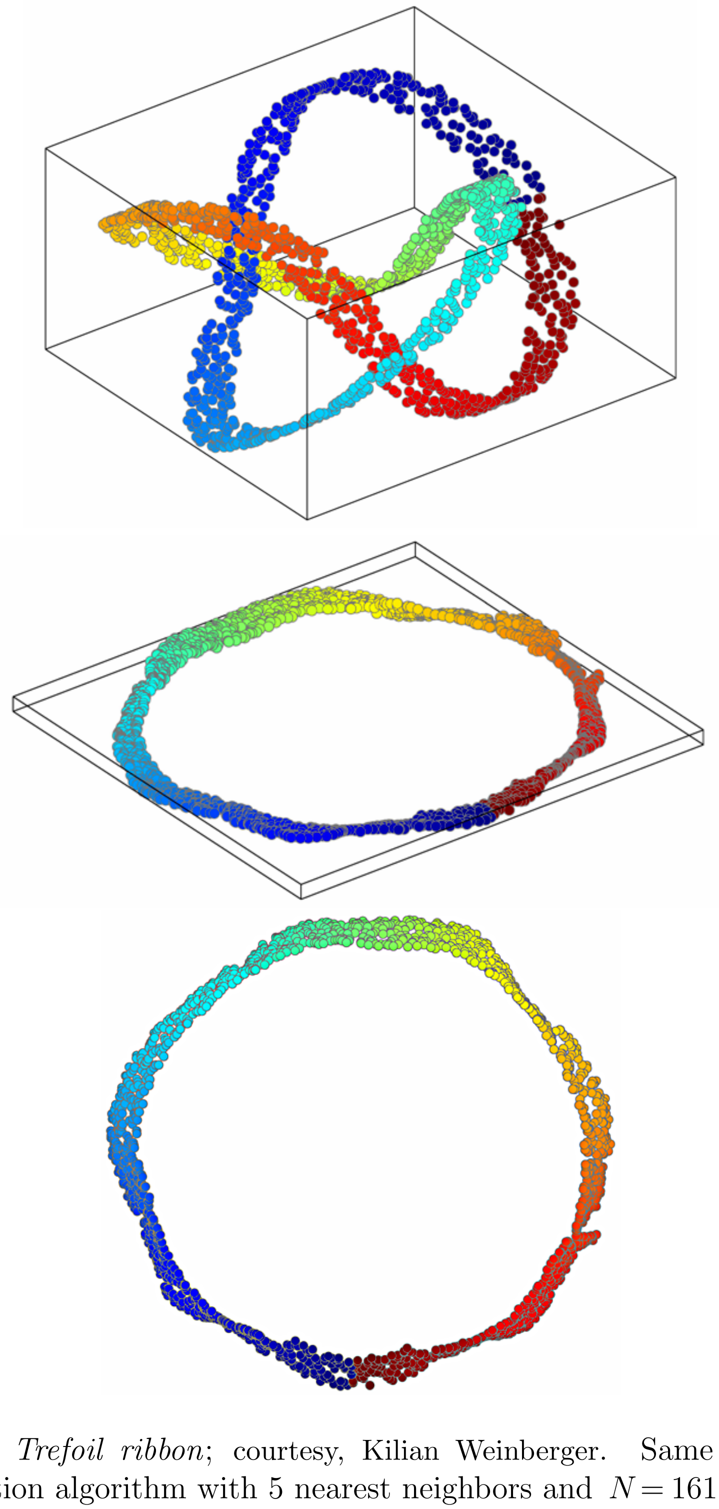


Figure 99: *Trefoil ribbon*; courtesy, Kilian Weinberger. Same topological transformation algorithm with 5 nearest neighbors and $N = 1617$ samples.

## 6.5 EDM definition in $\mathbf{1}\mathbf{1}^T$

Any EDM $D$ corresponding to affine dimension $r$ has representation

$$\mathbf{D}(V_{\mathcal{X}}) \triangleq \delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)\mathbf{1}^T + \mathbf{1}\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)^T - 2V_{\mathcal{X}}V_{\mathcal{X}}^T \in \mathbb{EDM}^N \tag{994}$$

where $\mathcal{R}(V_{\mathcal{X}} \in \mathbb{R}^{N\times r}) \subseteq \mathcal{N}(\mathbf{1}^T) = \mathbf{1}^\perp$,

$$V_{\mathcal{X}}^T V_{\mathcal{X}} = \delta^2(V_{\mathcal{X}}^T V_{\mathcal{X}}) \quad \textbf{and} \quad V_{\mathcal{X}} \text{ is full-rank with orthogonal columns.} \tag{995}$$

Equation (994) is simply the standard EDM definition (709) with a centered list $X$ as in (791); Gram matrix $X^TX$ has been replaced with the subcompact singular value decomposition (§A.6.2)[6.4]

$$V_{\mathcal{X}}V_{\mathcal{X}}^T \equiv V^T X^T X V \in \mathbb{S}_c^N \cap \mathbb{S}_+^N \tag{996}$$

This means: inner product $V_{\mathcal{X}}^T V_{\mathcal{X}}$ is an $r\times r$ diagonal matrix $\Sigma$ of nonzero singular values.

Vector $\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)$ may me decomposed into complementary parts by projecting it on orthogonal subspaces $\mathbf{1}^\perp$ and $\mathcal{R}(\mathbf{1})$: namely,

$$P_{\mathbf{1}^\perp}\big(\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)\big) = V\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \tag{997}$$

$$P_{\mathbf{1}}\big(\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)\big) = \frac{1}{N}\mathbf{1}\mathbf{1}^T\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \tag{998}$$

Of course

$$\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) = V\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) + \frac{1}{N}\mathbf{1}\mathbf{1}^T\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \tag{999}$$

by (732). Substituting this into EDM definition (994), we get the Hayden, Wells, Liu, & Tarazaga EDM formula [134, §2]

$$\mathbf{D}(V_{\mathcal{X}}\,,\, y) \triangleq y\mathbf{1}^T + \mathbf{1}y^T + \frac{\lambda}{N}\mathbf{1}\mathbf{1}^T - 2V_{\mathcal{X}}V_{\mathcal{X}}^T \in \mathbb{EDM}^N \tag{1000}$$

where

$$\lambda \triangleq 2\|V_{\mathcal{X}}\|_{\mathrm{F}}^2 = \mathbf{1}^T\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)2 \quad \textbf{and} \quad y \triangleq \delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) - \frac{\lambda}{2N}\mathbf{1} = V\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \tag{1001}$$

---

[6.4] Subcompact SVD: $V_{\mathcal{X}}V_{\mathcal{X}}^T \triangleq Q\sqrt{\Sigma}\sqrt{\Sigma}Q^T \equiv V^TX^TXV$. So $V_{\mathcal{X}}^T$ is not necessarily $XV$ (§5.5.1.0.1), although affine dimension $r = \operatorname{rank}(V_{\mathcal{X}}^T) = \operatorname{rank}(XV)$. (841)

and $y=\mathbf{0}$ if and only if $\mathbf{1}$ is an eigenvector of EDM $D$. Scalar $\lambda$ becomes an eigenvalue when corresponding eigenvector $\mathbf{1}$ exists.[6.5]

Then the particular dyad sum from (1000)

$$y\mathbf{1}^T + \mathbf{1}y^T + \frac{\lambda}{N}\mathbf{11}^T \in \mathbb{S}_c^{N\perp} \tag{1002}$$

must belong to the orthogonal complement of the geometric center subspace (p.612), whereas $V_\mathcal{X}V_\mathcal{X}^T \in \mathbb{S}_c^N \cap \mathbb{S}_+^N$ (996) 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

$$V_\mathcal{X}^T\mathbf{1} = \mathbf{0} \tag{1003}$$

it follows

$$\begin{aligned} D\mathbf{1} \;=\; \delta(V_\mathcal{X}V_\mathcal{X}^T)\mathbf{1}^T\mathbf{1} + \mathbf{1}\delta(V_\mathcal{X}V_\mathcal{X}^T)^T\mathbf{1} \;&=\; N\,\delta(V_\mathcal{X}V_\mathcal{X}^T) + \|V_\mathcal{X}\|_\mathrm{F}^2\mathbf{1} \\ &\Rightarrow \delta(V_\mathcal{X}V_\mathcal{X}^T) \propto \mathbf{1} \end{aligned} \tag{1004}$$

For some $\kappa \in \mathbb{R}_+$

$$\delta(V_\mathcal{X}V_\mathcal{X}^T)^T\mathbf{1} = N\kappa = \mathrm{tr}(V_\mathcal{X}^TV_\mathcal{X}) = \|V_\mathcal{X}\|_\mathrm{F}^2 \quad\Rightarrow\quad \delta(V_\mathcal{X}V_\mathcal{X}^T) = \frac{1}{N}\|V_\mathcal{X}\|_\mathrm{F}^2\mathbf{1} \tag{1005}$$

so $y=\mathbf{0}$.
($\Leftarrow$) Now suppose $\delta(V_\mathcal{X}V_\mathcal{X}^T) = \frac{\lambda}{2N}\mathbf{1}$; *id est*, $y=\mathbf{0}$. Then

$$D = \frac{\lambda}{N}\mathbf{11}^T - 2V_\mathcal{X}V_\mathcal{X}^T \in \mathbb{EDM}^N \tag{1006}$$

$\mathbf{1}$ is an eigenvector with corresponding eigenvalue $\lambda$. ♦

### 6.5.1 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 (994)[6.6] make a linearly independent set,

[6.5] *e.g.*, when $X=I$ in EDM definition (709).

[6.6] Identifying columns $V_\mathcal{X} \triangleq [\,v_1 \cdots v_r\,]$, then $V_\mathcal{X}V_\mathcal{X}^T = \sum_i v_iv_i^T$ is also a sum of dyads.

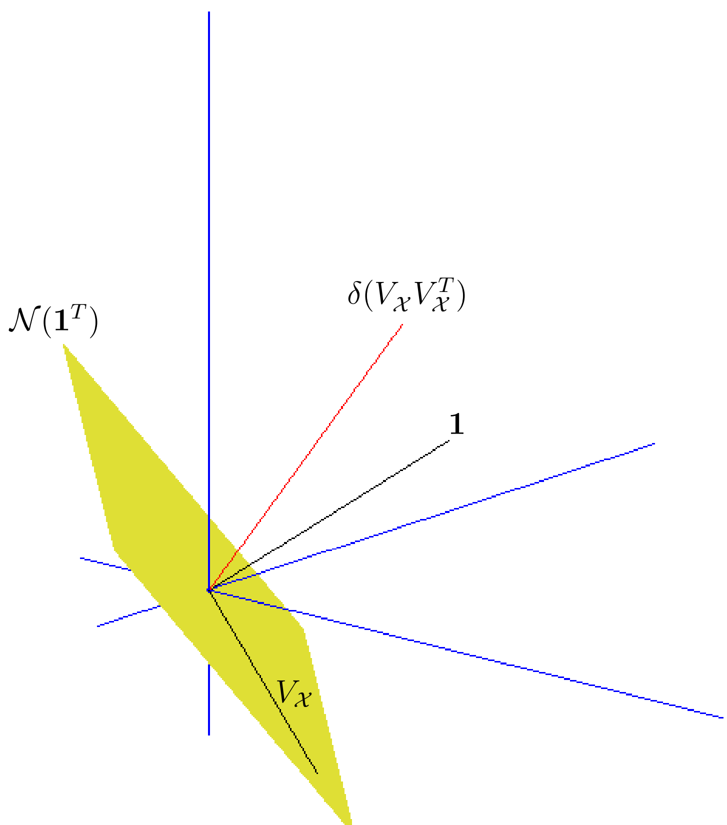



Figure 100: Example of $V_{\mathcal{X}}$ selection to make an EDM corresponding to cardinality $N=3$ and affine dimension $r=1$; $V_{\mathcal{X}}$ is a vector in nullspace $\mathcal{N}(\mathbf{1}^T) \subset \mathbb{R}^{\mathbf{3}}$. Nullspace of $\mathbf{1}^T$ is hyperplane in $\mathbb{R}^{\mathbf{3}}$ (drawn truncated) having normal $\mathbf{1}$. Vector $\delta(V_{\mathcal{X}} V_{\mathcal{X}}^T)$ may or may not be in plane spanned by $\{\mathbf{1}\,, V_{\mathcal{X}}\}$, but belongs to nonnegative orthant which is strictly supported by $\mathcal{N}(\mathbf{1}^T)$.

then the nontranspose halves constitute a basis for the range of EDM $D$. Saying this mathematically: For $D \in \mathbb{EDM}^N$

$$\begin{aligned}
\mathcal{R}(D) = \mathcal{R}\big([\,\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \quad \mathbf{1} \quad V_{\mathcal{X}}\,]\big) &\Leftarrow \operatorname{rank}\big([\,\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \quad \mathbf{1} \quad V_{\mathcal{X}}\,]\big) = 2 + r \\
\mathcal{R}(D) = \mathcal{R}\big([\,\mathbf{1} \quad V_{\mathcal{X}}\,]\big) &\Leftarrow \text{otherwise}
\end{aligned} \tag{1007}$$

To prove this, we need that condition under which the rank equality is satisfied: We know $\mathcal{R}(V_{\mathcal{X}}) \perp \mathbf{1}$, but what is the relative geometric orientation of $\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)$? $\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \succeq 0$ because $V_{\mathcal{X}}V_{\mathcal{X}}^T \succeq 0$, and $\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \propto \mathbf{1}$ remains possible (1004); this means $\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) \notin \mathcal{N}(\mathbf{1}^T)$ simply because it has no negative entries. (Figure **100**) If the projection of $\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)$ on $\mathcal{N}(\mathbf{1}^T)$ does not belong to $\mathcal{R}(V_{\mathcal{X}})$, then that is a necessary and sufficient condition for linear independence (l.i.) of $\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)$ with respect to $\mathcal{R}([\,\mathbf{1} \quad V_{\mathcal{X}}\,])$; *id est*,

$$\begin{aligned}
V\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) &\neq V_{\mathcal{X}}\,a \quad \text{for any } a \in \mathbb{R}^r \\
(I - \tfrac{1}{N}\mathbf{1}\mathbf{1}^T)\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) &\neq V_{\mathcal{X}}\,a \\
\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) - \tfrac{1}{N}\|V_{\mathcal{X}}\|_{\mathrm{F}}^2\mathbf{1} &\neq V_{\mathcal{X}}\,a \\
\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T) - \tfrac{\lambda}{2N}\mathbf{1} = y &\neq V_{\mathcal{X}}\,a \;\Leftrightarrow\; \{\mathbf{1}\,,\;\; \delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)\,,\;\; V_{\mathcal{X}}\} \text{ is l.i.}
\end{aligned} \tag{1008}$$

On the other hand when this condition is violated (when (1001) $y = V_{\mathcal{X}}\,a_{\mathrm{p}}$ for some particular $a \in \mathbb{R}^r$), then from (1000) we have

$$\begin{aligned}
\mathcal{R}\big(D = y\mathbf{1}^T + \mathbf{1}y^T + \tfrac{\lambda}{N}\mathbf{1}\mathbf{1}^T - 2V_{\mathcal{X}}V_{\mathcal{X}}^T\big) &= \mathcal{R}\big((V_{\mathcal{X}}\,a_{\mathrm{p}} + \tfrac{\lambda}{N}\mathbf{1})\mathbf{1}^T + (\mathbf{1}a_{\mathrm{p}}^T - 2V_{\mathcal{X}})V_{\mathcal{X}}^T\big) \\
&= \mathcal{R}\big([\,V_{\mathcal{X}}\,a_{\mathrm{p}} + \tfrac{\lambda}{N}\mathbf{1} \quad \mathbf{1}a_{\mathrm{p}}^T - 2V_{\mathcal{X}}\,]\big) \\
&= \mathcal{R}\big([\,\mathbf{1} \quad V_{\mathcal{X}}\,]\big)
\end{aligned} \tag{1009}$$

An example of such a violation is (1006) where, in particular, $a_{\mathrm{p}} = \mathbf{0}$. ♦

Then a statement parallel to (1007) is, for $D \in \mathbb{EDM}^N$ (Theorem 5.7.3.0.1)

$$\begin{aligned}
\operatorname{rank}(D) = r + 2 &\;\Leftrightarrow\; y \notin \mathcal{R}(V_{\mathcal{X}}) \;\; \big(\Leftrightarrow\; \mathbf{1}^T D^\dagger \mathbf{1} = 0\big) \\
\operatorname{rank}(D) = r + 1 &\;\Leftrightarrow\; y \in \mathcal{R}(V_{\mathcal{X}}) \;\; \big(\Leftrightarrow\; \mathbf{1}^T D^\dagger \mathbf{1} \neq 0\big)
\end{aligned} \tag{1010}$$

(a)

(b)

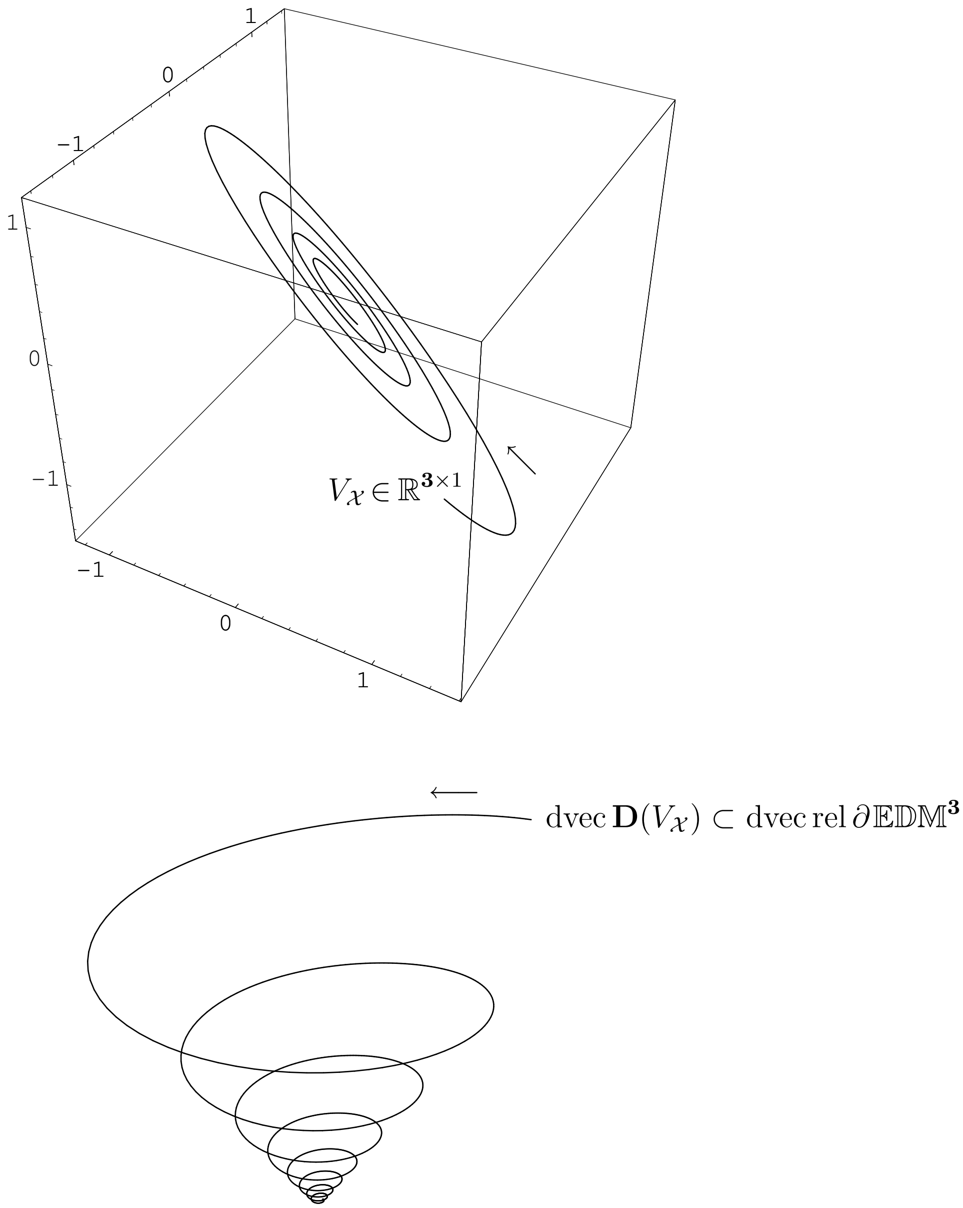



Figure 101: **(a)** Vector $V_{\mathcal{X}}$ from Figure 100 spirals in $\mathcal{N}(\mathbf{1}^T) \subset \mathbb{R}^3$ decaying toward origin. (Spiral is two-dimensional in vector space $\mathbb{R}^3$.) **(b)** Corresponding trajectory $\mathbf{D}(V_{\mathcal{X}})$ on EDM cone relative boundary creates a vortex also decaying toward origin. There are two complete orbits on EDM cone boundary about axis of revolution for every single revolution of $V_{\mathcal{X}}$ about origin. (Vortex is three-dimensional in isometrically isomorphic $\mathbb{R}^3$.)

### 6.5.2 Boundary constituents of EDM cone

Expression (994) has utility in forming the set of all EDMs corresponding to affine dimension $r$ :

$$\begin{aligned}&\left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)=r\right\}\\ &= \left\{\mathbf{D}(V_{\mathcal{X}}) \mid V_{\mathcal{X}}\in\mathbb{R}^{N\times r},\ \operatorname{rank}V_{\mathcal{X}}=r\ ,\ V_{\mathcal{X}}^TV_{\mathcal{X}}=\delta^2(V_{\mathcal{X}}^TV_{\mathcal{X}})\ ,\ \mathcal{R}(V_{\mathcal{X}})\subseteq\mathcal{N}(\mathbf{1}^T)\right\}\end{aligned} \tag{1011}$$

whereas $\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)\leq r\}$ is the closure of this same set;

$$\left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)\leq r\right\} = \overline{\left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)=r\right\}} \tag{1012}$$

For example,

$$\begin{aligned}\operatorname{rel}\partial\mathbb{EDM}^N &= \left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)<N-1\right\}\\ &= \bigcup_{r=0}^{N-2}\left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)=r\right\}\end{aligned} \tag{1013}$$

None of these are necessarily convex sets, although

$$\begin{aligned}\mathbb{EDM}^N &= \bigcup_{r=0}^{N-1}\left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)=r\right\}\\ &= \overline{\left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)=N-1\right\}}\\ \operatorname{rel\,int}\mathbb{EDM}^N &= \left\{D\in\mathbb{EDM}^N \mid \operatorname{rank}(VDV)=N-1\right\}\end{aligned} \tag{1014}$$

are pointed convex cones.

When cardinality $N=3$ and affine dimension $r=2$, for example, the relative interior $\operatorname{rel\,int}\mathbb{EDM}^{\mathbf{3}}$ is realized via (1011). (§6.6)

When $N=3$ and $r=1$, the relative boundary of the EDM cone $\operatorname{dvec}\operatorname{rel}\partial\mathbb{EDM}^{\mathbf{3}}$ is realized in isomorphic $\mathbb{R}^{\mathbf{3}}$ as in Figure **95**(d). This figure could be constructed via (1012) by spiraling vector $V_{\mathcal{X}}$ tightly about the origin in $\mathcal{N}(\mathbf{1}^T)$ ; as can be imagined with aid of Figure **100**. Vectors close to the origin in $\mathcal{N}(\mathbf{1}^T)$ are correspondingly close to the origin in $\mathbb{EDM}^N$. As vector $V_{\mathcal{X}}$ orbits the origin in $\mathcal{N}(\mathbf{1}^T)$, the corresponding EDM orbits the axis of revolution while remaining on the boundary of the circular cone $\operatorname{dvec}\operatorname{rel}\partial\mathbb{EDM}^{\mathbf{3}}$. (Figure **101**)

### 6.5.3 Faces of EDM cone

**6.5.3.0.1 Exercise.** *Isomorphic faces.*
Prove that in high cardinality $N$, any set of EDMs made via (1011) or (1012) with particular affine dimension $r$ is isomorphic with any set admitting the same affine dimension but made in lower cardinality. ▼

#### 6.5.3.1 Extreme direction of EDM cone

In particular, extreme directions (§2.8.1) of $\mathbb{EDM}^N$ 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 $\mathcal{N}(\mathbf{1}^T)$

$$\begin{aligned}\Gamma &\triangleq (z\circ z)\mathbf{1}^T + \mathbf{1}(z\circ z)^T - 2zz^T \in \mathbb{EDM}^N\\ &= \delta(zz^T)\mathbf{1}^T + \mathbf{1}\delta(zz^T)^T - 2zz^T\end{aligned} \tag{1015}$$

is an extreme direction corresponding to a one-dimensional face of the EDM cone $\mathbb{EDM}^N$ that is a ray in isomorphic subspace $\mathbb{R}^{N(N-1)/2}$.

Proving this would exercise the fundamental definition (155) of extreme direction. Here is a sketch: Any EDM may be represented

$$\mathbf{D}(V_{\mathcal{X}}) \triangleq \delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)\mathbf{1}^T + \mathbf{1}\delta(V_{\mathcal{X}}V_{\mathcal{X}}^T)^T - 2V_{\mathcal{X}}V_{\mathcal{X}}^T \in \mathbb{EDM}^N \tag{994}$$

where matrix $V_{\mathcal{X}}$ (995) has orthogonal columns. For the same reason (1272) that $zz^T$ is an extreme direction of the positive semidefinite cone (§2.9.2.4) for any particular nonzero vector $z$, there is no conic combination of distinct EDMs (each conically independent of $\Gamma$) equal to $\Gamma$. ■

**6.5.3.1.1 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 $D\in\mathbb{S}_h^N$ it holds

$$D = \delta\big(-VDV\tfrac{1}{2}\big)\mathbf{1}^T + \mathbf{1}\delta\big(-VDV\tfrac{1}{2}\big)^T - 2\big(-VDV\tfrac{1}{2}\big) \qquad (808)$$

By diagonalization $-VDV\frac{1}{2} \triangleq Q\Lambda Q^T \in \mathbb{S}_c^N$ (§A.5.2) we may write

$$\begin{aligned} D &= \delta\left(\sum_{i=1}^N \lambda_i\, q_i q_i^T\right)\mathbf{1}^T + \mathbf{1}\delta\left(\sum_{i=1}^N \lambda_i\, q_i q_i^T\right)^T - 2\sum_{i=1}^N \lambda_i\, q_i q_i^T \\ &= \sum_{i=1}^N \lambda_i \left(\delta(q_i q_i^T)\mathbf{1}^T + \mathbf{1}\delta(q_i q_i^T)^T - 2q_i q_i^T\right) \end{aligned} \qquad (1016)$$

where $q_i$ is the $i^{\text{th}}$ eigenvector of $-VDV\frac{1}{2}$ arranged columnar in orthogonal matrix

$$Q = [\, q_1 \;\; q_2 \;\; \cdots \;\; q_N \,] \in \mathbb{R}^{N\times N} \qquad (341)$$

and where $\{\delta(q_i q_i^T)\mathbf{1}^T + \mathbf{1}\delta(q_i q_i^T)^T - 2q_i q_i^T\,,\;\; i=1\ldots N\}$ are extreme directions of some pointed polyhedral cone $\mathcal{K}\subset\mathbb{S}_h^N$ and extreme directions of $\mathbb{EDM}^N$. Invertibility of (1016)

$$\begin{aligned} -VDV\tfrac{1}{2} &= -V\sum_{i=1}^N \lambda_i \left(\delta(q_i q_i^T)\mathbf{1}^T + \mathbf{1}\delta(q_i q_i^T)^T - 2q_i q_i^T\right) V\tfrac{1}{2} \\ &= \sum_{i=1}^N \lambda_i\, q_i q_i^T \end{aligned} \qquad (1017)$$

implies linear independence of those extreme directions. Then the biorthogonal expansion is expressed

$$\operatorname{dvec} D \;=\; YY^\dagger \operatorname{dvec} D \;=\; Y\,\lambda\big(-VDV\tfrac{1}{2}\big) \qquad (1018)$$

where

$$Y \triangleq \left[\operatorname{dvec}\big(\delta(q_i q_i^T)\mathbf{1}^T + \mathbf{1}\delta(q_i q_i^T)^T - 2q_i q_i^T\big)\,,\; i=1\ldots N\right] \in \mathbb{R}^{N(N-1)/2\times N} \qquad (1019)$$

When $D$ belongs to the EDM cone in the subspace of symmetric hollow matrices, unique coordinates $Y^\dagger \operatorname{dvec} D$ 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 $\operatorname{dvec}\mathcal{K} = \operatorname{cone}(Y)$. □

#### 6.5.3.2 Smallest face

Now suppose we are given a particular EDM $\mathbf{D}(V_{\mathcal{X}_\mathrm{p}})\in\mathbb{EDM}^N$ corresponding to affine dimension $r$ and parametrized by $V_{\mathcal{X}_\mathrm{p}}$ in (994). The EDM cone's smallest face that contains $\mathbf{D}(V_{\mathcal{X}_\mathrm{p}})$ is

$$\begin{aligned}&\mathcal{F}\big(\mathbb{EDM}^N\ni\mathbf{D}(V_{\mathcal{X}_\mathrm{p}})\big)\\ &=\big\{\mathbf{D}(V_{\mathcal{X}})\mid V_{\mathcal{X}}\in\mathbb{R}^{N\times r},\ \operatorname{rank}V_{\mathcal{X}}=r\ ,\ V_{\mathcal{X}}^TV_{\mathcal{X}}=\delta^2(V_{\mathcal{X}}^TV_{\mathcal{X}})\ ,\ \mathcal{R}(V_{\mathcal{X}})\subseteq\mathcal{R}(V_{\mathcal{X}_\mathrm{p}})\big\}\end{aligned} \tag{1020}$$

which is isomorphic[6.7] with the convex cone $\mathbb{EDM}^{r+1}$, hence of dimension

$$\dim\mathcal{F}\big(\mathbb{EDM}^N\ni\mathbf{D}(V_{\mathcal{X}_\mathrm{p}})\big)=(r+1)r/2 \tag{1021}$$

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 $\mathcal{R}(V_{\mathcal{X}_\mathrm{p}})$ is any two-dimensional subspace of three-dimensional $\mathcal{N}(\mathbf{1}^T)$ in $\mathbb{R}^4$, for example, then the corresponding face of $\mathbb{EDM}^4$ is isometrically isomorphic with: (1012)

$$\mathbb{EDM}^3=\{D\in\mathbb{EDM}^3\mid\operatorname{rank}(VDV)\leq2\}\simeq\mathcal{F}(\mathbb{EDM}^4\ni\mathbf{D}(V_{\mathcal{X}_\mathrm{p}})) \tag{1022}$$

Each two-dimensional subspace of $\mathcal{N}(\mathbf{1}^T)$ corresponds to another three-dimensional face.

Because each and every principal submatrix of an EDM in $\mathbb{EDM}^N$ (§5.14.3) is another EDM [171, §4.1], for example, then each principal submatrix belongs to a particular face of $\mathbb{EDM}^N$.

#### 6.5.3.3 Open question

This result (1021) is analogous to that for the positive semidefinite cone, although the question remains open whether all faces of $\mathbb{EDM}^N$ (whose dimension is less than the dimension of the cone) are exposed like they are for the positive semidefinite cone. (§2.9.2.3) [261]

[6.7] The fact that the smallest face is isomorphic with another (perhaps smaller) EDM cone is implicit in [134, §2].

## 6.6 Correspondence to PSD cone $\mathbb{S}_+^{N-1}$

Hayden & Wells *et alii* [134, §2] assert one-to-one correspondence of EDMs with positive semidefinite matrices in the symmetric subspace. Because $\operatorname{rank}(VDV) \leq N-1$ (§5.7.1.1), that positive semidefinite cone corresponding to the EDM cone can only be $\mathbb{S}_+^{N-1}$. [6, §18.2.1] To clearly demonstrate this correspondence, we invoke inner-product form EDM definition

$$\begin{array}{c}\mathbf{D}(\Phi) \triangleq \left[\begin{array}{c}0\\ \delta(\Phi)\end{array}\right]\mathbf{1}^T + \mathbf{1}\left[\begin{array}{cc}0 & \delta(\Phi)^T\end{array}\right] - 2\left[\begin{array}{cc}0 & \mathbf{0}^T\\ \mathbf{0} & \Phi\end{array}\right] \in \mathbb{EDM}^N\\ \Leftrightarrow\\ \Phi \succeq 0\end{array} \tag{826}$$

Then the EDM cone may be expressed

$$\mathbb{EDM}^N = \left\{\mathbf{D}(\Phi) \mid \Phi \in \mathbb{S}_+^{N-1}\right\} \tag{1023}$$

Hayden & Wells' assertion can therefore be equivalently stated in terms of an inner-product form EDM operator

$$\mathbf{D}(\mathbb{S}_+^{N-1}) = \mathbb{EDM}^N \tag{828}$$

$$\mathbf{V}_{\mathcal{N}}(\mathbb{EDM}^N) = \mathbb{S}_+^{N-1} \tag{829}$$

identity (829) holding because $\mathcal{R}(V_{\mathcal{N}}) = \mathcal{N}(\mathbf{1}^T)$ (716), linear functions $\mathbf{D}(\Phi)$ and $\mathbf{V}_{\mathcal{N}}(D) = -V_{\mathcal{N}}^T D V_{\mathcal{N}}$ (§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 $\mathcal{N}(\mathbf{1}^T)$ correspond to points relatively interior to the EDM cone.

$r < N-1$: Symmetric hollow matrices $-D$ positive semidefinite on $\mathcal{N}(\mathbf{1}^T)$, where $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ 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 $\mathcal{N}(\mathbf{1}^T)$ correspond to extreme directions (1015) of the EDM cone; *id est*, for some nonzero vector $u$ (§A.3.1.0.7)

$$\left.\begin{array}{c}\operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} = 1\\ D \in \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N}\end{array}\right\} \Leftrightarrow \begin{array}{c}D \in \mathbb{EDM}^N\\ D \text{ is an extreme direction}\end{array} \Leftrightarrow \left\{\begin{array}{c}-V_{\mathcal{N}}^T D V_{\mathcal{N}} \equiv uu^T\\ D \in \mathbb{S}_h^N\end{array}\right. \tag{1024}$$

**6.6.0.0.1 Proof.** Case $r=1$ is easily proved: From the nonnegativity development in §5.8.1, extreme direction (1015), and Schoenberg criterion (728), we need show only sufficiency; *id est*, prove

$$\left.\begin{array}{r}\operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} = 1\\ D\in\mathbb{S}_h^N\cap\mathbb{R}_+^{N\times N}\end{array}\right\}\;\Rightarrow\;\begin{array}{c}D\in\mathbb{EDM}^N\\ D \text{ is an extreme direction}\end{array}$$

Any symmetric matrix $D$ satisfying the rank condition must have the form, for $z,q\in\mathbb{R}^N$ and nonzero $z\in\mathcal{N}(\mathbf{1}^T)$,

$$D = \pm(\mathbf{1}q^T + q\mathbf{1}^T - 2zz^T) \tag{1025}$$

because (§5.6.2.1, *confer* §E.7.2.0.2)

$$\mathcal{N}(\mathbf{V}_{\mathcal{N}}(D)) = \{\mathbf{1}q^T + q\mathbf{1}^T \mid q\in\mathbb{R}^N\} \subseteq \mathbb{S}^N \tag{1026}$$

Hollowness demands $q=\delta(zz^T)$ while nonnegativity demands choice of positive sign in (1025). Matrix $D$ thus takes the form of an extreme direction (1015) of the EDM cone. ♦

The foregoing proof is not extensible in rank: An EDM with corresponding affine dimension $r$ has the general form, for $\{z_i\in\mathcal{N}(\mathbf{1}^T)\,,\; i=1\ldots r\}$ an independent set,

$$D = \mathbf{1}\delta\left(\sum_{i=1}^r z_i z_i^T\right)^T + \delta\left(\sum_{i=1}^r z_i z_i^T\right)\mathbf{1}^T - 2\sum_{i=1}^r z_i z_i^T \in \mathbb{EDM}^N \tag{1027}$$

The EDM so defined relies principally on the sum $\sum z_i z_i^T$ having positive summand coefficients ($\Leftrightarrow -V_{\mathcal{N}}^T D V_{\mathcal{N}}\succeq 0$)[6.8]. 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 $\mathcal{R}(V_{\mathcal{N}})$; *e.g.*, from page 362,

$$-V\begin{bmatrix}0&1&1\\1&0&5\\1&5&0\end{bmatrix}V\frac{1}{2} = z_1 z_1^T - z_2 z_2^T \tag{1028}$$

[6.8] ($\Leftarrow$) For $a_i\in\mathbb{R}^{N-1}$, let $z_i = V_{\mathcal{N}}^{\dagger T} a_i$.

**6.6.0.0.2 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 $\mathbb{EDM}^{\mathbf{3}}$ where $u=\begin{bmatrix}1\\2\end{bmatrix}$ in (1024). Because $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ has eigenvalues $\{0,5\}$, the ray whose direction is $D$ also lies on the relative boundary of $\mathbb{EDM}^{\mathbf{3}}$.

In exception, EDM $D=\kappa\begin{bmatrix}0&1\\1&0\end{bmatrix}$, for any particular $\kappa>0$, is an extreme direction of $\mathbb{EDM}^{\mathbf{2}}$ but $-V_{\mathcal{N}}^T D V_{\mathcal{N}}$ has only one eigenvalue: $\{\kappa\}$. Because $\mathbb{EDM}^{\mathbf{2}}$ is a ray whose relative boundary (§2.6.1.3.1) is the origin, this conventional boundary does not include $D$ which belongs to the relative interior in this dimension. (§2.7.0.0.1) □

### 6.6.1 Gram-form correspondence to $\mathbb{S}_+^{N-1}$

With respect to $\mathbf{D}(G)=\delta(G)\mathbf{1}^T+\mathbf{1}\delta(G)^T-2G$ (721) the linear Gram-form EDM operator, results in §5.6.1 provide [1, §2.6]

$$\mathbb{EDM}^N = \mathbf{D}\big(\mathbf{V}(\mathbb{EDM}^N)\big) \equiv \mathbf{D}\big(V_{\mathcal{N}}\mathbb{S}_+^{N-1}V_{\mathcal{N}}^T\big) \tag{1029}$$

$$V_{\mathcal{N}}\mathbb{S}_+^{N-1}V_{\mathcal{N}}^T \equiv \mathbf{V}\big(\mathbf{D}\big(V_{\mathcal{N}}\mathbb{S}_+^{N-1}V_{\mathcal{N}}^T\big)\big) = \mathbf{V}(\mathbb{EDM}^N) \overset{\Delta}{=} -V\,\mathbb{EDM}^N V\tfrac{1}{2} = \mathbb{S}_c^N\cap\mathbb{S}_+^N \tag{1030}$$

a one-to-one correspondence between $\mathbb{EDM}^N$ and $\mathbb{S}_+^{N-1}$.

### 6.6.2 EDM cone by elliptope

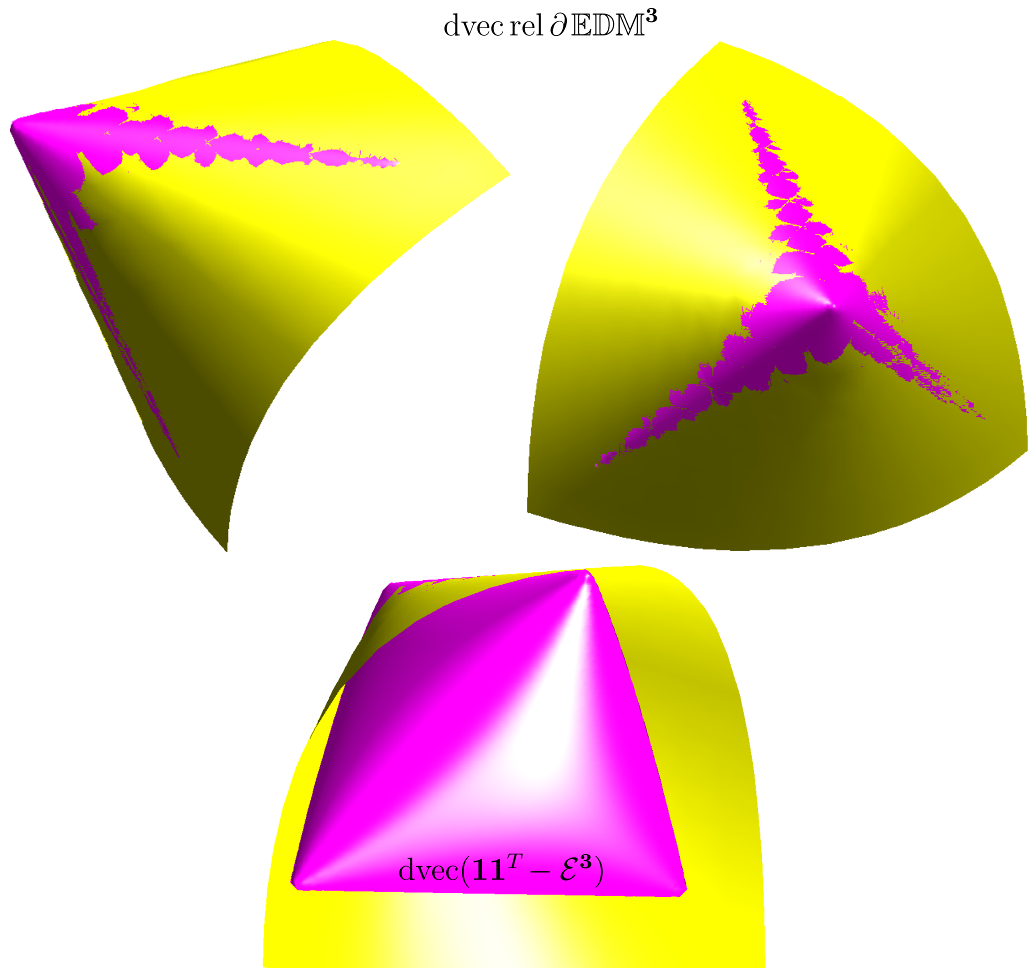
(*confer* §5.10.1) Defining the elliptope parametrized by scalar $t>0$

$$\mathcal{E}_t^N = \mathbb{S}_+^N \cap \{\Phi\in\mathbb{S}^N \mid \delta(\Phi)=t\mathbf{1}\} \tag{904}$$

then following Alfakih [7] we have

$$\mathbb{EDM}^N = \overline{\operatorname{cone}\{\mathbf{1}\mathbf{1}^T-\mathcal{E}_1^N\}} = \overline{\{t(\mathbf{1}\mathbf{1}^T-\mathcal{E}_1^N)\mid t\geq 0\}} \tag{1031}$$

Identification $\mathcal{E}^N=\mathcal{E}_1^N$ equates the standard elliptope (§5.9.1.0.1, Figure **87**) to our parametrized elliptope.




$$\mathbb{EDM}^N = \overline{\operatorname{cone}\{\mathbf{1 1}^T - \mathcal{E}^N\}} = \overline{\{t(\mathbf{1 1}^T - \mathcal{E}^N) \mid t \geq 0\}} \tag{1031}$$

Figure 102: Three views of translated negated elliptope $\mathbf{1 1}^T - \mathcal{E}_1^{\mathbf{3}}$ (*confer* Figure **87**) 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.

**6.6.2.0.1 Expository.** Define $\mathcal{T}_{\mathcal{E}}(\mathbf{1}\mathbf{1}^T)$ to be the *tangent cone* to the elliptope $\mathcal{E}$ at point $\mathbf{1}\mathbf{1}^T$; *id est*,

$$\mathcal{T}_{\mathcal{E}}(\mathbf{1}\mathbf{1}^T) \triangleq \overline{\{t(\mathcal{E} - \mathbf{1}\mathbf{1}^T) \mid t \geq 0\}} \tag{1032}$$

The normal cone $\mathcal{K}_{\mathcal{E}}^{\perp}(\mathbf{1}\mathbf{1}^T)$ to the elliptope at $\mathbf{1}\mathbf{1}^T$ is a closed convex cone defined (§E.10.3.2.1, Figure **130**)

$$\mathcal{K}_{\mathcal{E}}^{\perp}(\mathbf{1}\mathbf{1}^T) \triangleq \{B \mid \langle B\,,\, \Phi - \mathbf{1}\mathbf{1}^T\rangle \leq 0\,,\ \Phi \in \mathcal{E}\} \tag{1033}$$

The *polar cone* of any set $\mathcal{K}$ is the closed convex cone (*confer* (258))

$$\mathcal{K}^{\circ} \triangleq \{B \mid \langle B\,,\, A\rangle \leq 0\,,\ \text{for all } A \in \mathcal{K}\} \tag{1034}$$

The normal cone is well known to be the polar of the tangent cone,

$$\mathcal{K}_{\mathcal{E}}^{\perp}(\mathbf{1}\mathbf{1}^T) = \mathcal{T}_{\mathcal{E}}(\mathbf{1}\mathbf{1}^T)^{\circ} \tag{1035}$$

and *vice versa*; [148, §A.5.2.4]

$$\mathcal{K}_{\mathcal{E}}^{\perp}(\mathbf{1}\mathbf{1}^T)^{\circ} = \mathcal{T}_{\mathcal{E}}(\mathbf{1}\mathbf{1}^T) \tag{1036}$$

From Deza & Laurent [77, p.535] we have the EDM cone

$$\mathbb{EDM} = -\mathcal{T}_{\mathcal{E}}(\mathbf{1}\mathbf{1}^T) \tag{1037}$$

The polar EDM cone is also expressible in terms of the elliptope. From (1035) we have

$$\mathbb{EDM}^{\circ} = -\mathcal{K}_{\mathcal{E}}^{\perp}(\mathbf{1}\mathbf{1}^T) \tag{1038}$$

★

In §5.10.1 we proposed the expression for EDM $D$

$$D = t\mathbf{1}\mathbf{1}^T - \mathfrak{E} \in \mathbb{EDM}^N \tag{905}$$

where $t \in \mathbb{R}_+$ and $\mathfrak{E}$ belongs to the parametrized elliptope $\mathcal{E}_t^N$. We further propose, for any particular $t > 0$

$$\mathbb{EDM}^N = \overline{\operatorname{cone}\{t\mathbf{1}\mathbf{1}^T - \mathcal{E}_t^N\}} \tag{1039}$$

**Proof.** Pending.

Relationship of the translated negated elliptope with the EDM cone is illustrated in Figure **102**. We speculate

$$\mathbb{EDM}^N = \overline{\lim_{t\to\infty} t\mathbf{1}\mathbf{1}^T - \mathcal{E}_t^N} \tag{1040}$$

## 6.7 Vectorization & projection interpretation

In §E.7.2.0.2 we learn: $-VDV$ can be interpreted as orthogonal projection [4, §2] of vectorized $-D\in\mathbb{S}_h^N$ on the subspace of geometrically centered symmetric matrices

$$\begin{aligned}\mathbb{S}_c^N &= \{G\in\mathbb{S}^N \mid G\mathbf{1}=\mathbf{0}\} & (1766)\\ &= \{G\in\mathbb{S}^N \mid \mathcal{N}(G)\supseteq\mathbf{1}\} = \{G\in\mathbb{S}^N \mid \mathcal{R}(G)\subseteq\mathcal{N}(\mathbf{1}^T)\}\\ &= \{VYV \mid Y\in\mathbb{S}^N\}\subset\mathbb{S}^N & (1767)\\ &\equiv \{V_{\mathcal{N}}AV_{\mathcal{N}}^T \mid A\in\mathbb{S}^{N-1}\}\end{aligned}\tag{799}$$

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 (704),[6.9]

$$\left.\begin{array}{r}\langle zz^T, -D\rangle\geq 0 \quad \forall\, zz^T \mid \mathbf{1}\mathbf{1}^T zz^T=\mathbf{0}\\ D\in\mathbb{S}_h^N\end{array}\right\} \Leftrightarrow\; D\in\mathbb{EDM}^N \tag{1041}$$

this is equivalent, of course, to the Schoenberg criterion

$$\left.\begin{array}{r}-V_{\mathcal{N}}^T D V_{\mathcal{N}}\succeq 0\\ D\in\mathbb{S}_h^N\end{array}\right\} \Leftrightarrow\; D\in\mathbb{EDM}^N \tag{728}$$

because $\mathcal{N}(\mathbf{1}\mathbf{1}^T)=\mathcal{R}(V_{\mathcal{N}})$. When $D\in\mathbb{EDM}^N$, correspondence (1041) means $-z^TDz$ is proportional to a nonnegative coefficient of orthogonal projection (§E.6.4.2, Figure **104**) 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 $\mathbf{1}\mathbf{1}^T$ ; *id est*, on each and every member of

$$\begin{aligned}\mathcal{T} &\triangleq \left\{\operatorname{svec}(zz^T) \mid z\in\mathcal{N}(\mathbf{1}\mathbf{1}^T)=\mathcal{R}(V_{\mathcal{N}})\right\}\subset \operatorname{svec}\partial\mathbb{S}_+^N\\ &= \left\{\operatorname{svec}(V_{\mathcal{N}}\upsilon\upsilon^TV_{\mathcal{N}}^T) \mid \upsilon\in\mathbb{R}^{N-1}\right\}\end{aligned}\tag{1042}$$

whose dimension is

$$\dim\mathcal{T} = N(N-1)/2 \tag{1043}$$

[6.9] $\mathcal{N}(\mathbf{1}\mathbf{1}^T)=\mathcal{N}(\mathbf{1}^T)$ and $\mathcal{R}(zz^T)=\mathcal{R}(z)$

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) $\mathbb{S}_+^N$ , hence lie on its boundary. Yet only those dyads in $\mathcal{R}(V_{\mathcal{N}})$ are included in the test (1041), thus only a subset $\mathcal{T}$ of all vectorized extreme directions of $\mathbb{S}_+^N$ is observed.

In the particularly simple case $D \in \mathbb{EDM}^{\mathbf{2}} = \{D \in \mathbb{S}_h^{\mathbf{2}} \mid d_{12} \geq 0\}$ , for example, only one extreme direction of the PSD cone is involved:

$$zz^T = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{1044}$$

Any nonnegative scaling of vectorized $zz^T$ belongs to the set $\mathcal{T}$ illustrated in Figure **103** and Figure **104**.

### 6.7.1 Face of PSD cone $\mathbb{S}_+^N$ containing $V$

In any case, set $\mathcal{T}$ (1042) constitutes the vectorized extreme directions of an $N(N-1)/2$-dimensional face of the PSD cone $\mathbb{S}_+^N$ containing auxiliary matrix $V$ ; a face isomorphic with $\mathbb{S}_+^{N-1} = \mathbb{S}_+^{\operatorname{rank} V}$ (§2.9.2.3).

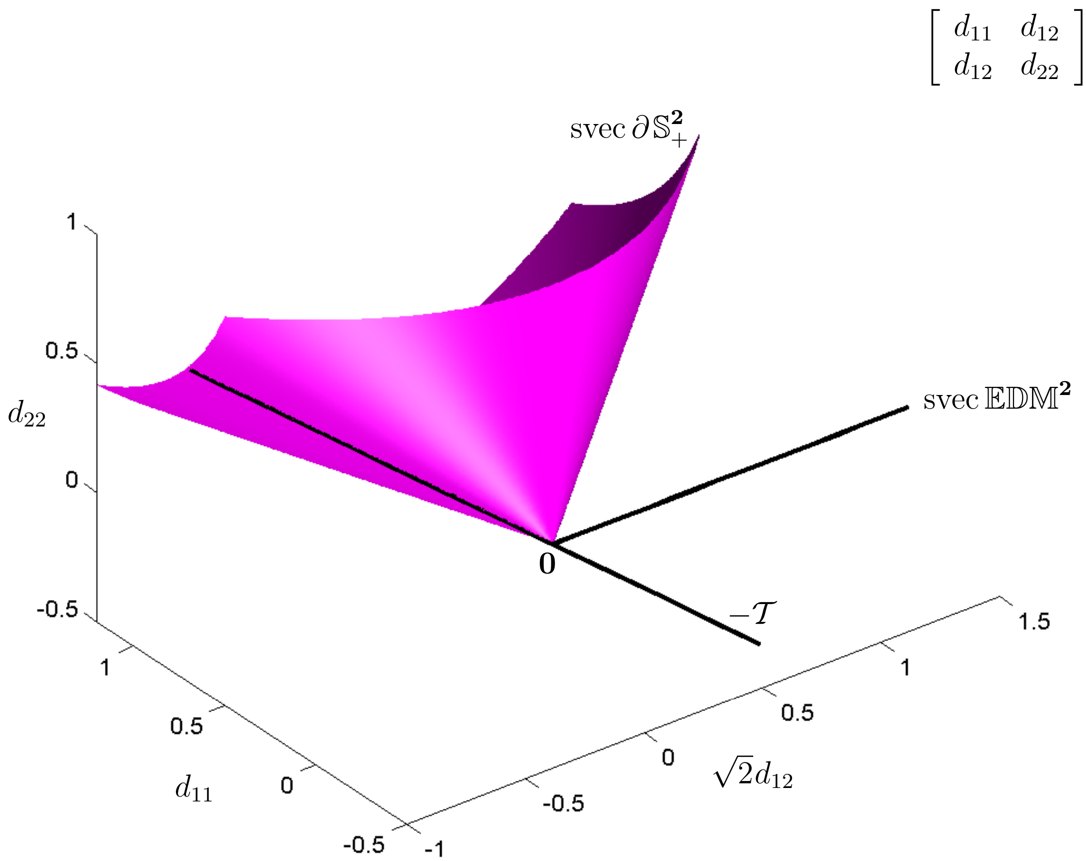
To show this, we must first find the smallest face that contains auxiliary matrix $V$ and then determine its extreme directions. From (190),

$$\begin{aligned} \mathcal{F}\big(\mathbb{S}_+^N \ni V\big) &= \{W \in \mathbb{S}_+^N \mid \mathcal{N}(W) \supseteq \mathcal{N}(V)\} = \{W \in \mathbb{S}_+^N \mid \mathcal{N}(W) \supseteq \mathbf{1}\} \\ &= \{VYV \succeq 0 \mid Y \in \mathbb{S}^N\} \equiv \{V_{\mathcal{N}} B V_{\mathcal{N}}^T \mid B \in \mathbb{S}_+^{N-1}\} \\ &\simeq \mathbb{S}_+^{\operatorname{rank} V} = -V_{\mathcal{N}}^T \, \mathbb{EDM}^N V_{\mathcal{N}} \end{aligned} \tag{1045}$$

where the equivalence $\equiv$ is from §5.6.1 while isomorphic equality $\simeq$ with transformed EDM cone is from (829). Projector $V$ belongs to $\mathcal{F}\big(\mathbb{S}_+^N \ni V\big)$ because $V_{\mathcal{N}} V_{\mathcal{N}}^\dagger V_{\mathcal{N}}^{\dagger T} V_{\mathcal{N}}^T = V$ . (§B.4.3) Each and every rank-one matrix belonging to this face is therefore of the form:

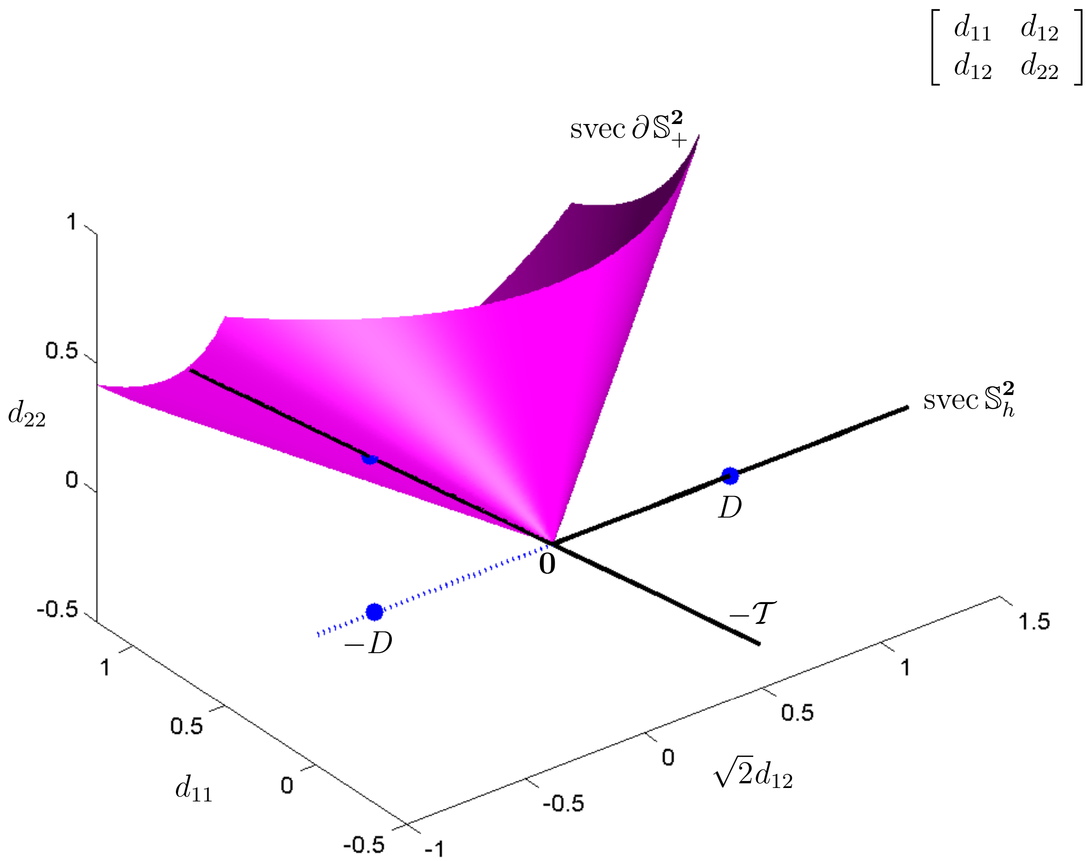
$$V_{\mathcal{N}} v v^T V_{\mathcal{N}}^T \mid v \in \mathbb{R}^{N-1} \tag{1046}$$

Because $\mathcal{F}\big(\mathbb{S}_+^N \ni V\big)$ is isomorphic with a positive semidefinite cone $\mathbb{S}_+^{N-1}$, then $\mathcal{T}$ constitutes the vectorized extreme directions of $\mathcal{F}$ , the origin constitutes the extreme points of $\mathcal{F}$ , and auxiliary matrix $V$ is some convex combination of those extreme points and directions by the *extremes theorem* (§2.8.1.1.1). ♦




$$\mathcal{T} \triangleq \left\{ \operatorname{svec}(zz^T) \mid z \in \mathcal{N}(\mathbf{1}\mathbf{1}^T) = \kappa \begin{bmatrix} 1 \\ -1 \end{bmatrix},\ \kappa \in \mathbb{R} \right\} \subset \operatorname{svec} \partial\mathbb{S}^2_+$$

Figure 103: Truncated boundary of positive semidefinite cone $\mathbb{S}^2_+$ in isometrically isomorphic $\mathbb{R}^3$ (via svec (47)) is, in this dimension, constituted solely by its extreme directions. Truncated cone of Euclidean distance matrices $\mathbb{EDM}^2$ in isometrically isomorphic subspace $\mathbb{R}$. Relative boundary of EDM cone is constituted solely by matrix $\mathbf{0}$. Halfline $\mathcal{T} = \{\kappa^2[1\ \ -\sqrt{2}\ \ \ 1]^T \mid \kappa \in \mathbb{R}\}$ on PSD cone boundary depicts that lone extreme ray (1044) on which orthogonal projection of $-D$ must be positive semidefinite if $D$ is to belong to $\mathbb{EDM}^2$. aff cone $\mathcal{T} = \operatorname{svec}\mathbb{S}^2_c$. (1049) Dual EDM cone is halfspace in $\mathbb{R}^3$ whose bounding hyperplane has inward-normal svec $\mathbb{EDM}^2$.




Projection of vectorized $-D$ on range of vectorized $zz^T$:

$$P_{\text{svec}\, zz^T}(\text{svec}(-D)) = \frac{\langle zz^T,\, -D\rangle}{\langle zz^T,\, zz^T\rangle}\, zz^T$$

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \left\{ \begin{array}{l} \langle zz^T, -D\rangle \geq 0 \quad \forall\, zz^T \mid \mathbf{1}\mathbf{1}^T zz^T = \mathbf{0} \\ D \in \mathbb{S}_h^N \end{array} \right. \tag{1041}$$

Figure 104: Given-matrix $D$ is assumed to belong to symmetric hollow subspace $\mathbb{S}_h^2$; a line in this dimension. Negating $D$, we find its polar along $\mathbb{S}_h^2$. Set $\mathcal{T}$ (1042) has only one ray member in this dimension; not orthogonal to $\mathbb{S}_h^2$. Orthogonal projection of $-D$ on $\mathcal{T}$ (indicated by half dot) has nonnegative projection coefficient. Matrix $D$ must therefore be an EDM.

In fact, the smallest face that contains auxiliary matrix $V$ of the PSD cone $\mathbb{S}_+^N$ is the intersection with the geometric center subspace (1766) (1767);

$$\begin{aligned}\mathcal{F}\big(\mathbb{S}_+^N \ni V\big) &= \operatorname{cone}\big\{V_{\mathcal{N}} \upsilon\upsilon^T V_{\mathcal{N}}^T \mid \upsilon \in \mathbb{R}^{N-1}\big\} \\ &= \mathbb{S}_c^N \cap \mathbb{S}_+^N\end{aligned} \tag{1047}$$

In isometrically isomorphic $\mathbb{R}^{N(N+1)/2}$

$$\operatorname{svec} \mathcal{F}\big(\mathbb{S}_+^N \ni V\big) = \operatorname{cone} \mathcal{T} \tag{1048}$$

related to $\mathbb{S}_c^N$ by

$$\operatorname{aff} \operatorname{cone} \mathcal{T} = \operatorname{svec} \mathbb{S}_c^N \tag{1049}$$

### 6.7.2 EDM criteria in $\mathbf{1}\mathbf{1}^T$

(*confer* §6.5) Laurent specifies an elliptope trajectory condition for EDM cone membership: [171, §2.3]

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; [1 - e^{-\alpha d_{ij}}] \in \mathbb{EDM}^N \quad \forall\, \alpha > 0 \qquad (899)$$

From the parametrized elliptope $\mathcal{E}_t^N$ in §6.6.2 we propose

$$D \in \mathbb{EDM}^N \;\Leftrightarrow\; \exists \left.\begin{array}{c} t \in \mathbb{R}_+ \\ \mathfrak{E} \in \mathcal{E}_t^N \end{array}\right\} \ni D = t\mathbf{1}\mathbf{1}^T - \mathfrak{E} \tag{1050}$$

Chabrillac & Crouzeix [53, §4] prove a different criterion they attribute to Finsler (1937) [97]. We apply it to EDMs: for $D \in \mathbb{S}_h^N$ (849)

$$\begin{array}{c} -V_{\mathcal{N}}^T D V_{\mathcal{N}} \succ 0 \;\Leftrightarrow\; \exists\, \kappa > 0 \;\ni\; -D + \kappa \mathbf{1}\mathbf{1}^T \succ 0 \\ \Leftrightarrow \\ D \in \mathbb{EDM}^N \text{ with corresponding affine dimension } r = N-1 \end{array} \tag{1051}$$

This *Finsler criterion* has geometric interpretation in terms of the vectorization & projection already discussed in connection with (1041). With reference to Figure **103**, the offset $\mathbf{1}\mathbf{1}^T$ is simply a direction orthogonal to $\mathcal{T}$ in isomorphic $\mathbb{R}^{\mathbf{3}}$. Intuitively, translation of $-D$ in direction $\mathbf{1}\mathbf{1}^T$ is like orthogonal projection on $\mathcal{T}$ in so far as similar information can be obtained.

When the Finsler criterion (1051) is applied despite lower affine dimension, the constant $\kappa$ can go to infinity making the test $-D+\kappa\mathbf{1}\mathbf{1}^T\succeq 0$ impractical for numerical computation. Chabrillac & Crouzeix invent a criterion for the semidefinite case, but is no more practical: for $D\in\mathbb{S}_h^N$

$$\begin{gathered} D\in\mathbb{EDM}^N \\ \Leftrightarrow \\ \exists\,\kappa_{\rm p}>0 \;\;\ni\;\; \forall\,\kappa\geq\kappa_{\rm p}\;,\;\; -D-\kappa\mathbf{1}\mathbf{1}^T\;[sic]\text{ has exactly one negative eigenvalue} \end{gathered} \tag{1052}$$

# 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 $\mathbb{EDM}^N$ is pointed, closed, convex, but has empty interior. 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):

$$\mathbb{EDM}^N = \bigcap_{\substack{z\in\mathcal{N}(\mathbf{1}^T)\\ i=1\ldots N}} \left\{D\in\mathbb{S}^N \mid \langle e_ie_i^T, D\rangle\geq 0\,,\;\langle e_ie_i^T, D\rangle\leq 0\,,\;\langle zz^T, -D\rangle\geq 0\right\} \tag{1053}$$

By definition (258), dual cone $\mathcal{K}^*$ comprises each and every vector inward-normal to a hyperplane supporting convex cone $\mathcal{K}$ (§2.4.2.6.1) or bounding a halfspace containing $\mathcal{K}$. 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 (1053):

$$\begin{aligned} \mathbb{EDM}^{N^*} &= \operatorname{cone}\{e_ie_i^T\,,\; -e_je_j^T \mid i\,,j=1\ldots N\} \;-\; \operatorname{cone}\{zz^T \mid \mathbf{1}\mathbf{1}^Tzz^T=\mathbf{0}\} \\ &= \{\sum_{i=1}^N\zeta_ie_ie_i^T - \sum_{j=1}^N\xi_je_je_j^T \mid \zeta_i\,,\xi_j\geq 0\} \;-\; \operatorname{cone}\{zz^T \mid \mathbf{1}\mathbf{1}^Tzz^T=\mathbf{0}\} \\ &= \{\delta(u)\mid u\in\mathbb{R}^N\} - \operatorname{cone}\big\{V_{\mathcal{N}}\upsilon\upsilon^TV_{\mathcal{N}}^T \mid \upsilon\in\mathbb{R}^{N-1},\;(\|v\|=1)\big\}\subset\mathbb{S}^N \\ &= \{\delta^2(Y) - V_{\mathcal{N}}\Psi V_{\mathcal{N}}^T \mid Y\in\mathbb{S}^N,\;\Psi\in\mathbb{S}_+^{N-1}\} \end{aligned} \tag{1054}$$

The EDM cone is not self-dual in ambient $\mathbb{S}^N$ because its affine hull belongs to a proper subspace

$$\operatorname{aff}\mathbb{EDM}^N = \mathbb{S}_h^N \tag{1055}$$

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 $V_{\mathcal{N}}$ is empty $[\,\emptyset\,]$, and dual cone $\mathbb{EDM}^*$ is the real line.

When $N=2$, the EDM cone is a nonnegative real line in isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$; there $\mathbb{S}_h^{\mathbf{2}}$ is a real line containing the EDM cone. Dual cone $\mathbb{EDM}^{\mathbf{2}^*}$ is the particular halfspace in $\mathbb{R}^{\mathbf{3}}$ whose boundary has inward-normal $\mathbb{EDM}^{\mathbf{2}}$. Diagonal matrices $\{\delta(u)\}$ in (1054) are represented by a hyperplane through the origin $\{\underline{d} \mid [0\;\;1\;\;0]\underline{d}=0\}$ while the term $\operatorname{cone}\{V_{\mathcal{N}}\upsilon\upsilon^T V_{\mathcal{N}}^T\}$ is represented by the halfline $\mathcal{T}$ in Figure **103** belonging to the positive semidefinite cone boundary. The dual EDM cone is formed by translating the hyperplane along the negative semidefinite halfline $-\mathcal{T}$; 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

$$\begin{aligned}
\mathcal{K}_1 &\triangleq \mathbb{S}_h^N\\
\mathcal{K}_2 &\triangleq \bigcap_{y\in\mathcal{N}(\mathbf{1}^T)} \left\{A\in\mathbb{S}^N \mid \langle yy^T, -A\rangle \geq 0\right\}\\
&= \left\{A\in\mathbb{S}^N \mid -z^T VAVz \geq 0 \quad \forall\, zz^T(\succeq 0)\right\}\\
&= \left\{A\in\mathbb{S}^N \mid -VAV \succeq 0\right\}
\end{aligned} \tag{1056}$$

so

$$\mathcal{K}_1\cap\mathcal{K}_2 = \mathbb{EDM}^N \tag{1057}$$

The dual cone $\mathcal{K}_1^* = \mathbb{S}_h^{N\perp} \subseteq \mathbb{S}^N$ (62) is the subspace of diagonal matrices. From (1054) via (272),

$$\mathcal{K}_2^* = -\operatorname{cone}\left\{V_{\mathcal{N}}\upsilon\upsilon^T V_{\mathcal{N}}^T \mid \upsilon\in\mathbb{R}^{N-1}\right\} \subset \mathbb{S}^N \tag{1058}$$

Gaffke & Mathar [99, §5.3] observe that projection on $\mathcal{K}_1$ and $\mathcal{K}_2$ have simple closed forms: Projection on subspace $\mathcal{K}_1$ is easily performed by symmetrization and zeroing the main diagonal or *vice versa*, while projection of $H\in\mathbb{S}^N$ on $\mathcal{K}_2$ is

$$P_{\mathcal{K}_2}H = H - P_{\mathbb{S}_+^N}(VHV) \tag{1059}$$

**Proof.** First, we observe membership of $H-P_{\mathbb{S}_+^N}(VHV)$ to $\mathcal{K}_2$ because

$$P_{\mathbb{S}_+^N}(VHV) - H \;=\; \Big(P_{\mathbb{S}_+^N}(VHV) - VHV\Big) \,+\, \big(VHV - H\big) \tag{1060}$$

The term $P_{\mathbb{S}_+^N}(VHV) - VHV$ necessarily belongs to the (dual) positive semidefinite cone by Theorem E.9.2.0.1. $V^2 = V$, hence

$$-V\Big(H - P_{\mathbb{S}_+^N}(VHV)\Big)V \succeq 0 \tag{1061}$$

by Corollary A.3.1.0.5.

Next, we require

$$\langle P_{\mathcal{K}_2}H - H \,,\, P_{\mathcal{K}_2}H\rangle = 0 \tag{1062}$$

Expanding,

$$\langle -P_{\mathbb{S}_+^N}(VHV) \,,\, H - P_{\mathbb{S}_+^N}(VHV)\rangle = 0 \tag{1063}$$

$$\langle P_{\mathbb{S}_+^N}(VHV) \,,\, (P_{\mathbb{S}_+^N}(VHV) - VHV) \,+\, (VHV - H)\rangle = 0 \tag{1064}$$

$$\langle P_{\mathbb{S}_+^N}(VHV) \,,\, (VHV - H)\rangle = 0 \tag{1065}$$

Product $VHV$ belongs to the geometric center subspace; (§E.7.2.0.2)

$$VHV \in \mathbb{S}_c^N \;=\; \{Y \in \mathbb{S}^N \mid \mathcal{N}(Y) \supseteq \mathbf{1}\} \tag{1066}$$

Diagonalize $VHV \triangleq Q\Lambda Q^T$ (§A.5) whose nullspace is spanned by the eigenvectors corresponding to 0 eigenvalues by Theorem A.7.3.0.1. Projection of $VHV$ on the PSD cone (§7.1) simply zeros negative eigenvalues in diagonal matrix $\Lambda$. Then

$$\mathcal{N}\big(P_{\mathbb{S}_+^N}(VHV)\big) \supseteq \mathcal{N}(VHV) \;\big(\supseteq \mathcal{N}(V)\,\big) \tag{1067}$$

from which it follows:

$$P_{\mathbb{S}_+^N}(VHV) \in \mathbb{S}_c^N \tag{1068}$$

so $P_{\mathbb{S}_+^N}(VHV) \perp (VHV - H)$ because $VHV - H \in \mathbb{S}_c^{N\perp}$.

Finally, we must have $P_{\mathcal{K}_2}H - H = -P_{\mathbb{S}_+^N}(VHV) \in \mathcal{K}_2^*$. From §6.7.1 we know dual cone $\mathcal{K}_2^* = -\mathcal{F}\big(\mathbb{S}_+^N \ni V\big)$ is the negative of the positive semidefinite cone's smallest face that contains auxiliary matrix $V$. Thus $P_{\mathbb{S}_+^N}(VHV) \in \mathcal{F}\big(\mathbb{S}_+^N \ni V\big) \Leftrightarrow \mathcal{N}\big(P_{\mathbb{S}_+^N}(VHV)\big) \supseteq \mathcal{N}(V)$ (§2.9.2.3) which was already established in (1067). ♦

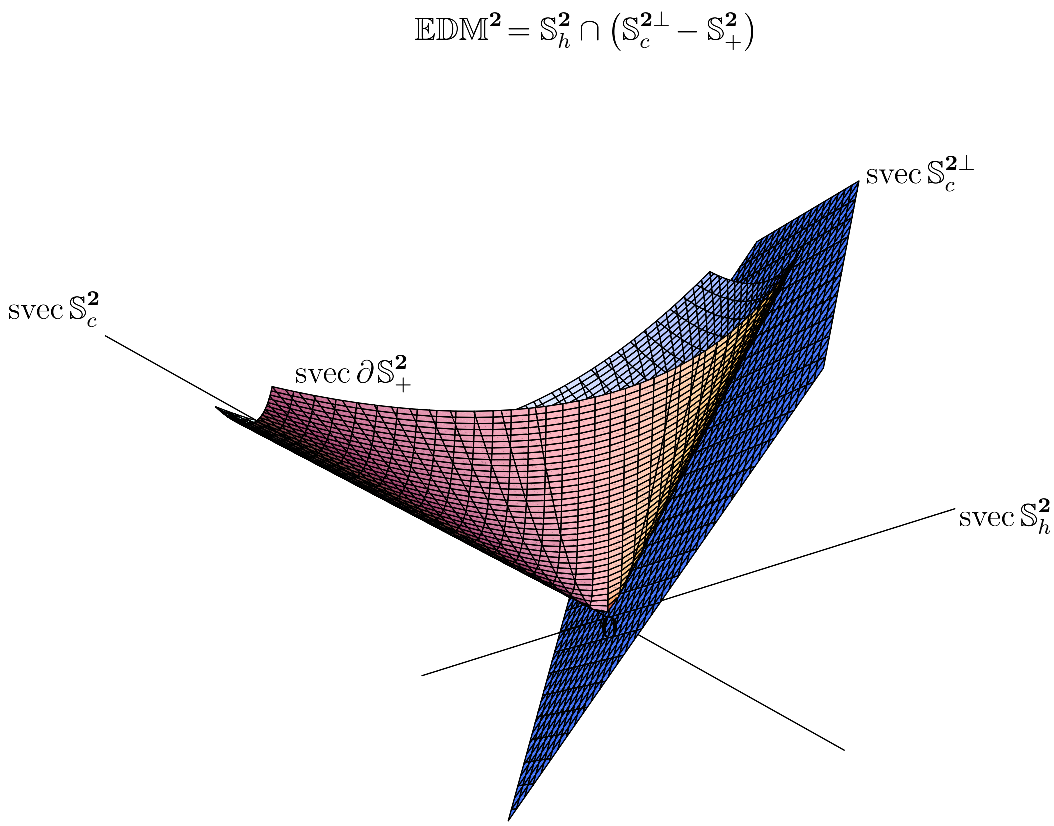



Figure 105: Orthogonal complement $\mathbb{S}_c^{\mathbf{2}\perp}$ (1768) (§B.2) of geometric center subspace (a plane in isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$; drawn is a tiled fragment) apparently supporting positive semidefinite cone. (Rounded vertex is artifact of plot.) Line $\text{svec}\,\mathbb{S}_c^{\mathbf{2}} = \text{aff cone}\,\mathcal{T}$ (1049) intersects $\text{svec}\,\partial\mathbb{S}_+^{\mathbf{2}}$, also drawn in Figure **103**; it runs along PSD cone boundary. (*confer* Figure **86**)

$$\mathbb{EDM}^2 = \mathbb{S}_h^2 \cap \left(\mathbb{S}_c^{2\perp} - \mathbb{S}_+^2\right)$$

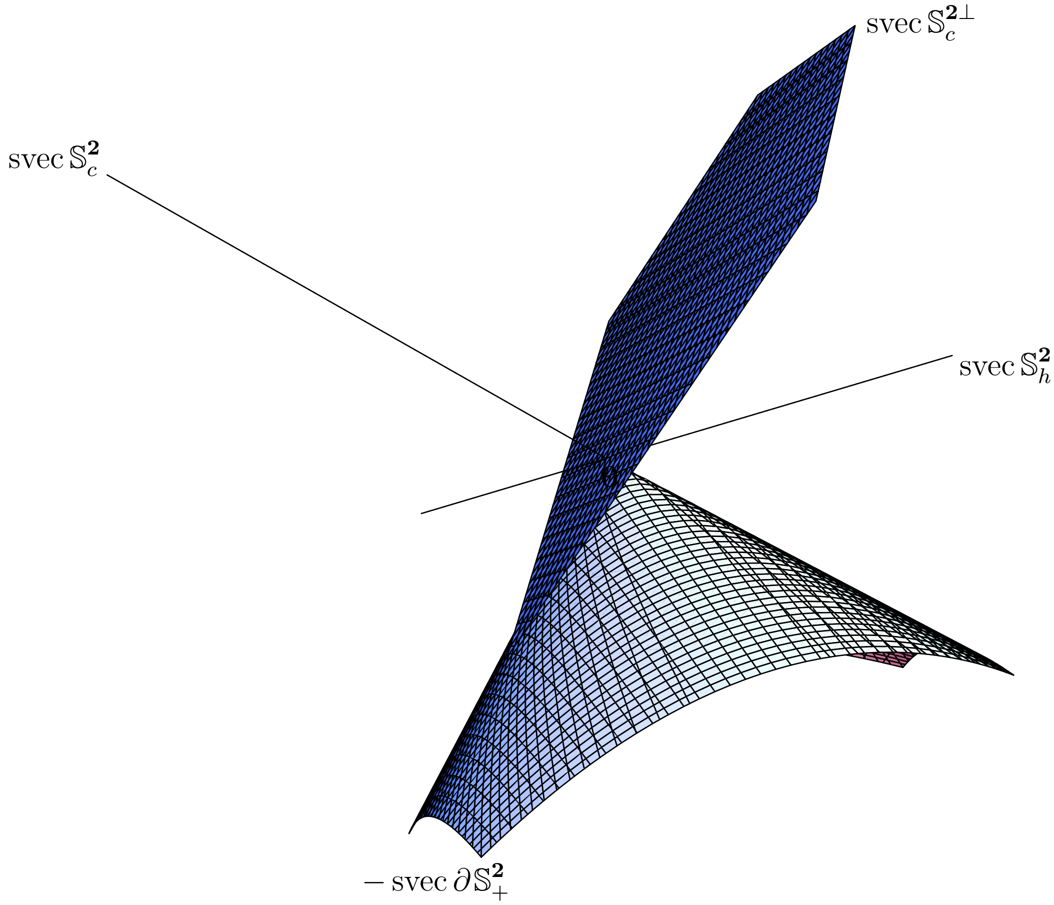



Figure 106: EDM cone construction in isometrically isomorphic $\mathbb{R}^3$ by adding polar PSD cone to svec $\mathbb{S}_c^{2\perp}$. Difference svec$\left(\mathbb{S}_c^{2\perp} - \mathbb{S}_+^2\right)$ is halfspace partially bounded by svec $\mathbb{S}_c^{2\perp}$. EDM cone is nonnegative halfline along svec $\mathbb{S}_h^2$ in this dimension.

From the results in §E.7.2.0.2, we know matrix product $VHV$ is the orthogonal projection of $H\in\mathbb{S}^N$ on the geometric center subspace $\mathbb{S}_c^N$. Thus the projection product

$$P_{\mathcal{K}_2}H = H - P_{\mathbb{S}_+^N}P_{\mathbb{S}_c^N}H \tag{1069}$$

**6.8.1.1.1 Lemma.** *Projection on PSD cone* $\cap$ *geometric center subspace.*

$$P_{\mathbb{S}_+^N\cap\,\mathbb{S}_c^N} \;=\; P_{\mathbb{S}_+^N}P_{\mathbb{S}_c^N} \tag{1070}$$

$\diamond$

**Proof.** For each and every $H\in\mathbb{S}^N$, projection of $P_{\mathbb{S}_c^N}H$ on the positive semidefinite cone remains in the geometric center subspace

$$\begin{aligned}
\mathbb{S}_c^N &= \{G\in\mathbb{S}^N \mid G\mathbf{1}=\mathbf{0}\} && (1766)\\
&= \{G\in\mathbb{S}^N \mid \mathcal{N}(G)\supseteq\mathbf{1}\} \;=\; \{G\in\mathbb{S}^N \mid \mathcal{R}(G)\subseteq\mathcal{N}(\mathbf{1}^T)\} && (799)\\
&= \{VYV \mid Y\in\mathbb{S}^N\}\subset\mathbb{S}^N && (1767)
\end{aligned}$$

That is because: eigenvectors of $P_{\mathbb{S}_c^N}H$ corresponding to 0 eigenvalues span its nullspace $\mathcal{N}(P_{\mathbb{S}_c^N}H)$. (§A.7.3.0.1) To project $P_{\mathbb{S}_c^N}H$ on the positive semidefinite cone, its negative eigenvalues are zeroed. (§7.1.2) The nullspace is thereby expanded while eigenvectors originally spanning $\mathcal{N}(P_{\mathbb{S}_c^N}H)$ 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 (1070) follows directly. $\blacklozenge$

From the lemma it follows

$$\{P_{\mathbb{S}_+^N}P_{\mathbb{S}_c^N}H \mid H\in\mathbb{S}^N\} \;=\; \{P_{\mathbb{S}_+^N\cap\,\mathbb{S}_c^N}H \mid H\in\mathbb{S}^N\} \tag{1071}$$

Then from (1793)

$$-\big(\mathbb{S}_c^N\cap\mathbb{S}_+^N\big)^* \;=\; \{H - P_{\mathbb{S}_+^N}P_{\mathbb{S}_c^N}H \mid H\in\mathbb{S}^N\} \tag{1072}$$

From (272) we get closure of a vector sum

$$\mathcal{K}_2 \;=\; -\big(\mathbb{S}_c^N\cap\mathbb{S}_+^N\big)^* \;=\; \mathbb{S}_c^{N\perp} - \mathbb{S}_+^N \tag{1073}$$

therefore the new equality

$$\mathbb{EDM}^N \;=\; \mathcal{K}_1\cap\mathcal{K}_2 \;=\; \mathbb{S}_h^N\cap\big(\mathbb{S}_c^{N\perp} - \mathbb{S}_+^N\big) \tag{1074}$$

whose veracity is intuitively evident, in hindsight, [61, p.109] from the most fundamental EDM definition (709). Formula (1074) is not a matrix criterion for membership to the EDM cone, and it is not an equivalence between EDM operators. 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. A realization of this construction in low dimension is illustrated in Figure **105** and Figure **106**.

The dual EDM cone follows directly from (1074) by standard properties of cones (§2.13.1.1):

$$\mathbb{EDM}^{N^*} = \overline{\mathcal{K}_1^* + \mathcal{K}_2^*} \;=\; \mathbb{S}_h^{N\perp} - \mathbb{S}_c^N \cap \mathbb{S}_+^N \tag{1075}$$

which bears strong resemblance to (1054).

#### 6.8.1.2 nonnegative orthant

That $\mathbb{EDM}^N$ is a proper subset of the nonnegative orthant is not obvious from (1074). We wish to verify

$$\mathbb{EDM}^N = \mathbb{S}_h^N \cap \left(\mathbb{S}_c^{N\perp} - \mathbb{S}_+^N\right) \subset \mathbb{R}_+^{N\times N} \tag{1076}$$

While there are many ways to prove this, it is sufficient to show that all entries of the extreme directions of $\mathbb{EDM}^N$ must be nonnegative; *id est*, for any particular nonzero vector $z = [z_i\,,\ i=1\ldots N] \in \mathcal{N}(\mathbf{1}^T)$ (§6.5.3.1),

$$\delta(zz^T)\mathbf{1}^T + \mathbf{1}\delta(zz^T)^T - 2zz^T \succeq \mathbf{0} \tag{1077}$$

where the inequality denotes entrywise comparison. The inequality holds because the $i\,,j^{\text{th}}$ 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

$$\delta(zz^T)\mathbf{1}^T + \mathbf{1}\delta(zz^T)^T \in \mathbb{S}_c^{N\perp} \tag{1078}$$

to the orthogonal complement (1768) of the geometric center subspace, while their difference is a member of the symmetric hollow subspace $\mathbb{S}_h^N$.

Here is an algebraic method provided by Trosset 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 $u$, $A = u\mathbf{1}^T + \mathbf{1}u^T = [A_{ij}] = [u_i + u_j]$ and $\delta(B) = \delta(A) = 2u$. Positive semidefiniteness of $B$ requires nonnegativity $A - B \succeq \mathbf{0}$ because

$$(e_i - e_j)^T B (e_i - e_j) \;=\; (B_{ii} - B_{ij}) - (B_{ji} - B_{jj}) \;=\; 2(u_i + u_j) - 2B_{ij} \geq 0 \tag{1079}$$

#### 6.8.1.3 Dual Euclidean distance matrix criterion

Conditions necessary for membership of a matrix $D^* \in \mathbb{S}^N$ to the dual EDM cone $\mathbb{EDM}^{N^*}$ may be derived from (1054): $D^* \in \mathbb{EDM}^{N^*} \Rightarrow D^* = \delta(y) - V_{\mathcal{N}} A V_{\mathcal{N}}^T$ for some vector $y$ and positive semidefinite matrix $A \in \mathbb{S}_+^{N-1}$. This in turn implies $\delta(D^*\mathbf{1}) = \delta(y)$. Then, for $D^* \in \mathbb{S}^N$

$$D^* \in \mathbb{EDM}^{N^*} \Leftrightarrow \delta(D^*\mathbf{1}) - D^* \succeq 0 \tag{1080}$$

where, for any symmetric matrix $D^*$

$$\delta(D^*\mathbf{1}) - D^* \in \mathbb{S}_c^N \tag{1081}$$

To show sufficiency of the matrix criterion in (1080), recall Gram-form EDM operator

$$\mathbf{D}(G) = \delta(G)\mathbf{1}^T + \mathbf{1}\delta(G)^T - 2G \tag{721}$$

where Gram matrix $G$ is positive semidefinite by definition, and recall the self-adjointness property of the main-diagonal linear operator $\delta$ (§A.1):

$$\langle D\,,\,D^*\rangle = \left\langle \delta(G)\mathbf{1}^T + \mathbf{1}\delta(G)^T - 2G\,,\,D^*\right\rangle = \langle G\,,\,\delta(D^*\mathbf{1}) - D^*\rangle\,2 \tag{740}$$

Assuming $\langle G\,,\,\delta(D^*\mathbf{1}) - D^*\rangle \geq 0$ (1285), then we have known membership relation (§2.13.2.0.1)

$$D^* \in \mathbb{EDM}^{N^*} \Leftrightarrow \langle D\,,\,D^*\rangle \geq 0 \quad \forall\, D \in \mathbb{EDM}^N \tag{1082}$$

♦

Elegance of this matrix criterion (1080) for membership to the dual EDM cone is the lack of any other assumptions except $D^*$ be symmetric. (Recall: Schoenberg criterion (728) for membership to the EDM cone requires membership to the symmetric hollow subspace.)

Linear Gram-form EDM operator (721) has adjoint, for $Y \in \mathbb{S}^N$

$$\mathbf{D}^T(Y) \triangleq (\delta(Y\mathbf{1}) - Y)\,2 \tag{1083}$$

Then we have: [61, p.111]

$$\mathbb{EDM}^{N^*} = \{Y \in \mathbb{S}^N \mid \delta(Y\mathbf{1}) - Y \succeq 0\} \tag{1084}$$

the dual EDM cone expressed in terms of the adjoint operator. A dual EDM cone determined this way is illustrated in Figure **108**.

**6.8.1.3.1 Exercise.** *Dual EDM spectral cone.*
Find a spectral cone as in §5.11.2 corresponding to $\mathbb{EDM}^{N^*}$. ▼

#### 6.8.1.4 Nonorthogonal components of dual EDM

Now we tie construct (1075) for the dual EDM cone together with the matrix criterion (1080) for dual EDM cone membership. For any $D^* \in \mathbb{S}^N$ it is obvious:

$$\delta(D^*\mathbf{1}) \in \mathbb{S}_h^{N\perp} \tag{1085}$$

any diagonal matrix belongs to the subspace of diagonal matrices (57). We know when $D^* \in \mathbb{EDM}^{N^*}$

$$\delta(D^*\mathbf{1}) - D^* \in \mathbb{S}_c^N \cap \mathbb{S}_+^N \tag{1086}$$

this adjoint expression (1083) belongs to that face (1047) of the positive semidefinite cone $\mathbb{S}_+^N$ in the geometric center subspace. Any nonzero dual EDM

$$D^* = \delta(D^*\mathbf{1}) - \big(\delta(D^*\mathbf{1}) - D^*\big) \in \mathbb{S}_h^{N\perp} \ominus \mathbb{S}_c^N \cap \mathbb{S}_+^N = \mathbb{EDM}^{N^*} \tag{1087}$$

can therefore be expressed as the difference of two linearly independent nonorthogonal components (Figure **86**, Figure **107**).

#### 6.8.1.5 Affine dimension complementarity

From §6.8.1.3 we have, for some $A \in \mathbb{S}_+^{N-1}$ (*confer* (1086))

$$\delta(D^*\mathbf{1}) - D^* = V_{\mathcal{N}} A V_{\mathcal{N}}^T \in \mathbb{S}_c^N \cap \mathbb{S}_+^N \tag{1088}$$

if and only if $D^*$ belongs to the dual EDM cone. Call $\operatorname{rank}(V_{\mathcal{N}} A V_{\mathcal{N}}^T)$ *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, $\mathbb{EDM}^{N^\circ} = -\mathbb{EDM}^{N^*}$, and its projection on the EDM cone; *id est*, the optimal solution of[6.10]

$$\begin{array}{cl} \underset{D^\circ \in\, \mathbb{S}^N}{\text{minimize}} & \|D^\circ - H\|_{\mathrm F} \\ \text{subject to} & D^\circ - \delta(D^\circ\mathbf{1}) \succeq 0 \end{array} \tag{1089}$$

[6.10] This dual projection can be solved quickly (without semidefinite programming) via

$$D^\circ = \delta(D^\circ\mathbf{1}) + \big(D^\circ - \delta(D^\circ\mathbf{1})\big) \in \mathbb{S}_h^{N\perp} \oplus \mathbb{S}_c^N \cap \mathbb{S}_+^N = \mathbb{EDM}^{N^\circ}$$

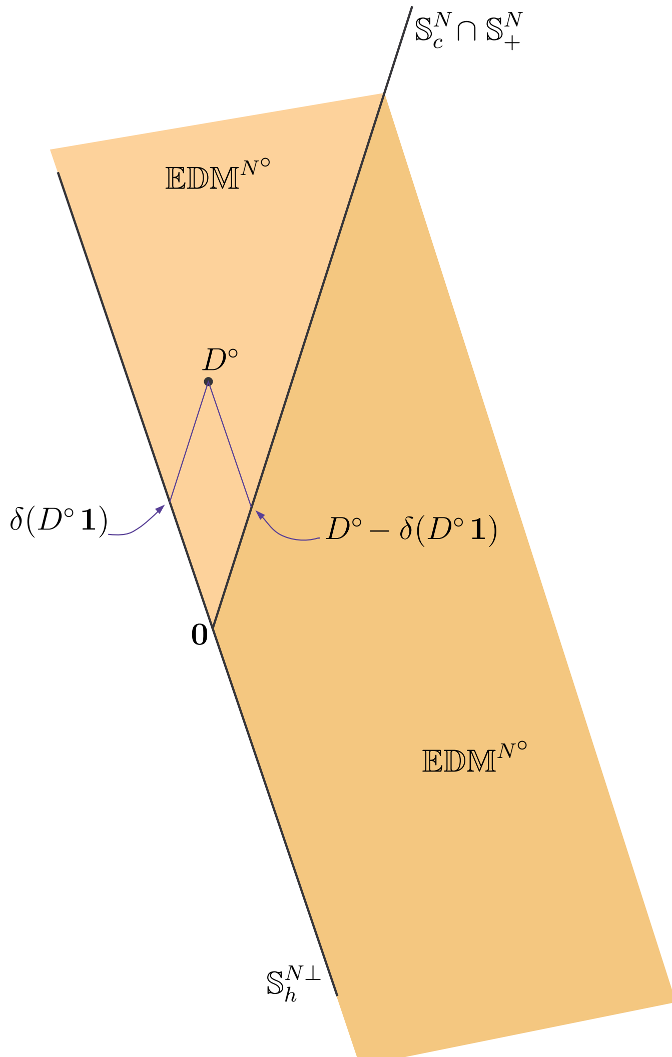



Figure 107: Hand-drawn abstraction of polar EDM cone (drawn truncated). Any member $D^\circ$ of polar EDM cone can be decomposed into two linearly independent nonorthogonal components: $\delta(D^\circ\mathbf{1})$ and $D^\circ - \delta(D^\circ\mathbf{1})$.

has dual affine dimension complementary to affine dimension corresponding to the optimal solution of

$$\begin{array}{cl}\underset{D\in\mathbb{S}_h^N}{\text{minimize}} & \|D-H\|_{\mathrm{F}}\\ \text{subject to} & -V_{\mathcal{N}}^TDV_{\mathcal{N}}\succeq 0\end{array}\tag{1090}$$

Precisely,

$$\operatorname{rank}(D^{\circ\star}-\delta(D^{\circ\star}\mathbf{1}))+\operatorname{rank}(V_{\mathcal{N}}^TD^{\star}V_{\mathcal{N}})=N-1\tag{1091}$$

and $\operatorname{rank}(D^{\circ\star}-\delta(D^{\circ\star}\mathbf{1}))\leq N-1$ because vector $\mathbf{1}$ is always in the nullspace of rank's argument. This is similar to the known result for projection on the self-dual positive semidefinite cone and its polar:

$$\operatorname{rank}P_{-\mathbb{S}_+^N}H\;+\;\operatorname{rank}P_{\mathbb{S}_+^N}H=N\tag{1092}$$

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 (1089) can be solved for $D^{\circ\star}$ by transforming to an equivalent Schur-form semidefinite program (§3.1.7.2). *Interior-point methods* for numerically solving semidefinite programs tend to produce high-rank solutions. (§4.1.1) Then $D^{\star}=H-D^{\circ\star}\in\mathbb{EDM}^N$ by Corollary E.9.2.2.1, and $D^{\star}$ 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 (1091) or (1092) (§7.1.4.1).

---

Lemma 6.8.1.1.1; rewriting,

$$\begin{array}{cl}\underset{D^{\circ}\in\mathbb{S}^N}{\text{minimize}} & \|(D^{\circ}-\delta(D^{\circ}\mathbf{1}))-(H-\delta(D^{\circ}\mathbf{1}))\|_{\mathrm{F}}\\ \text{subject to} & D^{\circ}-\delta(D^{\circ}\mathbf{1})\succeq 0\end{array}$$

which is the projection of affinely transformed optimal solution $H-\delta(D^{\circ\star}\mathbf{1})$ on $\mathbb{S}_c^N\cap\mathbb{S}_+^N$ ;

$$D^{\circ\star}-\delta(D^{\circ\star}\mathbf{1})=P_{\mathbb{S}_+^N}P_{\mathbb{S}_c^N}(H-\delta(D^{\circ\star}\mathbf{1}))$$

Foreknowledge of an optimal solution $D^{\circ\star}$ as argument to projection suggests recursion.

#### 6.8.1.6 EDM cone duality

In §5.6.1.1, via Gram-form EDM operator

$$\mathbf{D}(G) = \delta(G)\mathbf{1}^T + \mathbf{1}\delta(G)^T - 2G \in \mathbb{EDM}^N \quad \Leftarrow \quad G \succeq 0 \qquad (721)$$

we established clear connection between the EDM cone and that face (1047) of positive semidefinite cone $\mathbb{S}_+^N$ in the geometric center subspace:

$$\mathbb{EDM}^N = \mathbf{D}(\mathbb{S}_c^N \cap \mathbb{S}_+^N) \qquad (816)$$

$$\mathbf{V}(\mathbb{EDM}^N) = \mathbb{S}_c^N \cap \mathbb{S}_+^N \qquad (817)$$

where

$$\mathbf{V}(D) = -VDV\tfrac{1}{2} \qquad (805)$$

In §5.6.1 we established

$$\mathbb{S}_c^N \cap \mathbb{S}_+^N = V_{\mathcal{N}}\mathbb{S}_+^{N-1}V_{\mathcal{N}}^T \qquad (803)$$

Then from (1080), (1088), and (1054) we can deduce

$$\delta(\mathbb{EDM}^{N^*}\mathbf{1}) - \mathbb{EDM}^{N^*} = V_{\mathcal{N}}\mathbb{S}_+^{N-1}V_{\mathcal{N}}^T = \mathbb{S}_c^N \cap \mathbb{S}_+^N \qquad (1093)$$

which, by (816) and (817), means the EDM cone can be related to the dual EDM cone by an equality:

$$\mathbb{EDM}^N = \mathbf{D}\Big(\delta(\mathbb{EDM}^{N^*}\mathbf{1}) - \mathbb{EDM}^{N^*}\Big) \qquad (1094)$$

$$\mathbf{V}(\mathbb{EDM}^N) = \delta(\mathbb{EDM}^{N^*}\mathbf{1}) - \mathbb{EDM}^{N^*} \qquad (1095)$$

This means projection $-\mathbf{V}(\mathbb{EDM}^N)$ of the EDM cone on the geometric center subspace $\mathbb{S}_c^N$ (§E.7.2.0.2) is an affine transformation of the dual EDM cone: $\mathbb{EDM}^{N^*} - \delta(\mathbb{EDM}^{N^*}\mathbf{1})$. Secondarily, it means the EDM cone is not self-dual.

#### 6.8.1.7 Schoenberg criterion is discretized membership relation

We show the Schoenberg criterion

$$\left.\begin{array}{r}-V_{\mathcal{N}}^T D V_{\mathcal{N}} \in \mathbb{S}_+^{N-1} \\ D \in \mathbb{S}_h^N\end{array}\right\} \Leftrightarrow \quad D \in \mathbb{EDM}^N \tag{728}$$

to be a discretized membership relation (§2.13.4) between a closed convex cone $\mathcal{K}$ and its dual $\mathcal{K}^*$ like

$$\langle y\,,\, x\rangle \geq 0 \quad \text{for all} \quad y \in \mathcal{G}(\mathcal{K}^*) \quad \Leftrightarrow \quad x \in \mathcal{K} \tag{317}$$

where $\mathcal{G}(\mathcal{K}^*)$ is any set of generators whose conic hull constructs closed convex dual cone $\mathcal{K}^*$:

The Schoenberg criterion is the same as

$$\left.\begin{array}{r}\langle zz^T, -D\rangle \geq 0 \quad \forall\, zz^T \mid \mathbf{1}\mathbf{1}^T zz^T = \mathbf{0} \\ D \in \mathbb{S}_h^N\end{array}\right\} \Leftrightarrow \quad D \in \mathbb{EDM}^N \tag{1041}$$

which, by (1042), is the same as

$$\left.\begin{array}{r}\langle zz^T, -D\rangle \geq 0 \quad \forall\, zz^T \in \left\{V_{\mathcal{N}} \upsilon \upsilon^T V_{\mathcal{N}}^T \mid \upsilon \in \mathbb{R}^{N-1}\right\} \\ D \in \mathbb{S}_h^N\end{array}\right\} \Leftrightarrow \quad D \in \mathbb{EDM}^N \tag{1096}$$

where the $zz^T$ constitute a set of generators $\mathcal{G}$ for the positive semidefinite cone's smallest face $\mathcal{F}\big(\mathbb{S}_+^N \ni V\big)$ (§6.7.1) that contains auxiliary matrix $V$. From the aggregate in (1054) we get the ordinary membership relation, assuming only $D \in \mathbb{S}^N$ [148, p.58],

$$\langle D^*,\, D\rangle \geq 0 \quad \forall\, D^* \in \mathbb{EDM}^{N^*} \Leftrightarrow \quad D \in \mathbb{EDM}^N \tag{1097}$$

$$\langle D^*,\, D\rangle \geq 0 \quad \forall\, D^* \in \{\delta(u) \mid u \in \mathbb{R}^N\} - \operatorname{cone}\left\{V_{\mathcal{N}} \upsilon \upsilon^T V_{\mathcal{N}}^T \mid \upsilon \in \mathbb{R}^{N-1}\right\} \Leftrightarrow \quad D \in \mathbb{EDM}^N$$

Discretization yields (317):

$$\langle D^*,\, D\rangle \geq 0 \quad \forall\, D^* \in \{e_i e_i^T\,,\, -e_j e_j^T\,,\, -V_{\mathcal{N}} \upsilon \upsilon^T V_{\mathcal{N}}^T \mid i\,, j = 1 \ldots N\,,\ \upsilon \in \mathbb{R}^{N-1}\} \Leftrightarrow \quad D \in \mathbb{EDM}^N \tag{1098}$$

Because $\left\langle\{\delta(u) \mid u\in\mathbb{R}^N\}\,,\, D\right\rangle \geq 0 \;\Leftrightarrow\; D\in\mathbb{S}_h^N$ , we can restrict observation to the symmetric hollow subspace without loss of generality. Then for $D\in\mathbb{S}_h^N$

$$\langle D^*,\, D\rangle \geq 0 \quad \forall\, D^* \in \left\{-V_{\mathcal{N}}\upsilon\upsilon^T V_{\mathcal{N}}^T \mid \upsilon\in\mathbb{R}^{N-1}\right\} \;\Leftrightarrow\; D \in \mathbb{EDM}^N \tag{1099}$$

this discretized membership relation becomes (1096); 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 $\mathbb{S}_h^N$

When instead we consider the ambient space of symmetric hollow matrices (1055), then still we find the EDM cone is not self-dual for $N>2$. The simplest way to prove this is as follows:

Given a set of generators $\mathcal{G}=\{\Gamma\}$ (1015) for the pointed closed convex EDM cone, the *discrete 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:

$$\begin{aligned}
\mathbb{EDM}^{N^*}\cap\,\mathbb{S}_h^N &\triangleq \{D^*\in\mathbb{S}_h^N \mid \langle\Gamma\,,\, D^*\rangle \geq 0 \quad \forall\,\Gamma\in\mathcal{G}(\mathbb{EDM}^N)\} \\
&= \{D^*\in\mathbb{S}_h^N \mid \langle\delta(zz^T)\mathbf{1}^T + \mathbf{1}\delta(zz^T)^T - 2zz^T\,,\; D^*\rangle \geq 0 \;\; \forall\, z\in\mathcal{N}(\mathbf{1}^T)\} \\
&= \{D^*\in\mathbb{S}_h^N \mid \langle\mathbf{1}\delta(zz^T)^T - zz^T\,,\; D^*\rangle \geq 0 \;\; \forall\, z\in\mathcal{N}(\mathbf{1}^T)\}
\end{aligned} \tag{1100}$$

By comparison

$$\mathbb{EDM}^N = \{D \in \mathbb{S}_h^N \mid \langle -zz^T\,,\; D\rangle \geq 0 \;\; \forall\, z\in\mathcal{N}(\mathbf{1}^T)\} \tag{1101}$$

the term $\delta(zz^T)^T D^*\mathbf{1}$ foils any hope of self-duality in ambient $\mathbb{S}_h^N$ . ♦

To find the dual EDM cone in ambient $\mathbb{S}_h^N$ *per* §2.13.9.4 we prune the aggregate in (1054) describing the ordinary dual EDM cone, removing any member having nonzero main diagonal:

$$\begin{aligned}\mathbb{EDM}^{N^*}\cap\,\mathbb{S}_h^N \;=\; &\operatorname{cone}\{\delta^2(V_{\mathcal{N}}\upsilon\upsilon^T V_{\mathcal{N}}^T) - V_{\mathcal{N}}\upsilon\upsilon^T V_{\mathcal{N}}^T \mid \upsilon\in\mathbb{R}^{N-1}\}\\ =\; &\{\delta^2(V_{\mathcal{N}}\Psi V_{\mathcal{N}}^T) - V_{\mathcal{N}}\Psi V_{\mathcal{N}}^T \mid \Psi\in\mathbb{S}_+^{N-1}\}\end{aligned} \tag{1102}$$

When $N=1$, the EDM cone and its dual in ambient $\mathbb{S}_h$ each comprise the origin in isomorphic $\mathbb{R}^0$; thus, self-dual in this dimension. (*confer* (84))

When $N=2$, the EDM cone is the nonnegative real line in isomorphic $\mathbb{R}$. (Figure **103**) $\mathbb{EDM}^{\mathbf{2}^*}$ in $\mathbb{S}_h^{\mathbf{2}}$ is identical, thus self-dual in this dimension. This result is in agreement with (1100), verified directly: for all $\kappa\in\mathbb{R}$, $z=\kappa\begin{bmatrix}1\\-1\end{bmatrix}$ and $\delta(zz^T)=\kappa^2\begin{bmatrix}1\\1\end{bmatrix} \Rightarrow\; d_{12}^*\geq 0$.

The first case adverse to self-duality $N=3$ may be deduced from Figure **95**; the EDM cone is a circular cone in isomorphic $\mathbb{R}^{\mathbf{3}}$ corresponding to no rotation of the Lorentz cone (147) (the self-dual circular cone). Figure **108** illustrates the EDM cone and its dual in ambient $\mathbb{S}_h^{\mathbf{3}}$; no longer self-dual.

## 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.

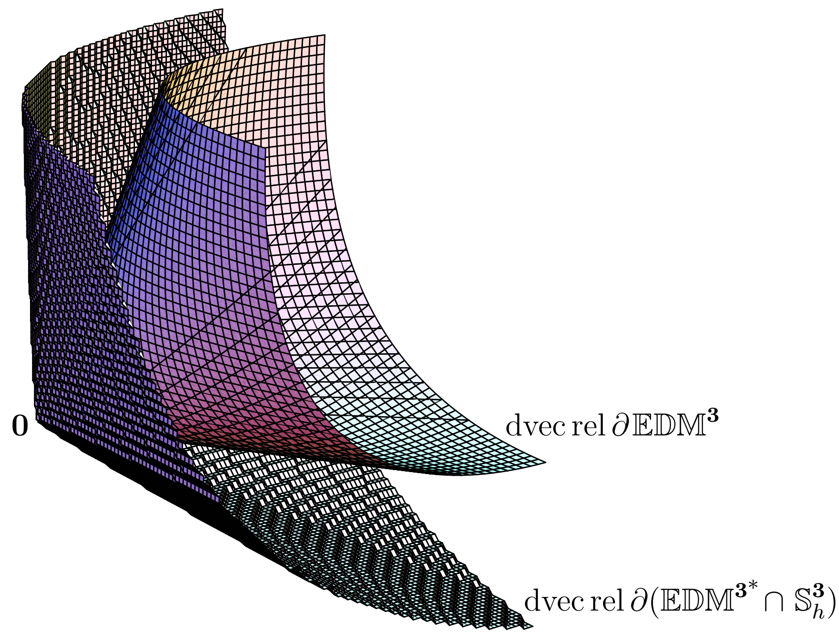
**6.9.0.0.1 Theorem.** *EDM alternative.* [113, §1]
Given $D\in\mathbb{S}_h^N$

$$D\in\mathbb{EDM}^N$$

or in the alternative

$$\exists\; z \;\text{ such that }\begin{cases}\mathbf{1}^T z = 1\\ Dz=\mathbf{0}\end{cases} \tag{1103}$$

In words, either $\mathcal{N}(D)$ intersects hyperplane $\{z \mid \mathbf{1}^T z=1\}$ or $D$ is an EDM; the alternatives are incompatible. ⋄




$$D^* \in \mathbb{EDM}^{N^*} \Leftrightarrow \ \delta(D^*\mathbf{1}) - D^* \succeq 0 \qquad (1080)$$

Figure 108: Ordinary dual EDM cone projected on $\mathbb{S}_h^{\mathbf{3}}$ shrouds $\mathbb{EDM}^{\mathbf{3}}$; drawn tiled in isometrically isomorphic $\mathbb{R}^{\mathbf{3}}$. (It so happens: intersection $\mathbb{EDM}^{N^*} \cap \mathbb{S}_h^N$ (§2.13.9.3) is identical to projection of dual EDM cone on $\mathbb{S}_h^N$.)

When $D$ is an EDM [190, §2]

$$\mathcal{N}(D) \subset \mathcal{N}(\mathbf{1}^T) = \{z \mid \mathbf{1}^T z = 0\} \tag{1104}$$

Because [113, §2] (§E.0.1)

$$\begin{aligned} DD^\dagger \mathbf{1} &= \mathbf{1} \\ \mathbf{1}^T D^\dagger D &= \mathbf{1}^T \end{aligned} \tag{1105}$$

then

$$\mathcal{R}(\mathbf{1}) \subset \mathcal{R}(D) \tag{1106}$$

## 6.10 postscript

We provided an equality (1074) relating the convex cone of Euclidean distance matrices to the convex cone of positive semidefinite matrices. Projection on the positive semidefinite cone constrained by an upper bound on rank is easy and well known; [85] simply a matter of truncating a list of eigenvalues. Projection on the positive semidefinite cone with such a rank constraint is, in fact, a convex optimization problem. (§7.1.4)

Yet, it remains difficult to project on the EDM cone under a constraint on rank or affine dimension. One thing we can do is invoke the Schoenberg criterion (728) and then project on a positive semidefinite cone under a constraint bounding affine dimension from above.

It is our hope that the equality herein relating EDM and PSD cones will become a step toward understanding projection on the EDM cone under a rank constraint.

# Chapter 7

# Proximity problems

*In summary, we find that the solution to problem* [(1116.3) p.445] *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) [134, §3]

A problem common to various sciences is to find the Euclidean distance matrix (EDM) $D \in \mathbb{EDM}^N$ 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 $\rho$.



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

### 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)

$$\mathcal{K} \triangleq \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N} \tag{1107}$$

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 possess 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).

#### 7.0.1.1 Order of imposition

Since convex cone $\mathcal{K}$ (1107) 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 $\mathcal{K}$ (that member of $\mathcal{K}$ 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 the intersection of any convex set $\mathcal{C}$ with any subspace. Consider the *proximity problem*[7.1] over convex feasible set $\mathbb{S}_h^N \cap \mathcal{C}$ given nonsymmetric nonhollow $H \in \mathbb{R}^{N\times N}$:

$$\begin{array}{cl}\underset{B\in\mathbb{S}_h^N}{\text{minimize}} & \|B-H\|_{\mathrm{F}}^2\\ \text{subject to} & B\in\mathcal{C}\end{array}\tag{1108}$$

a convex optimization problem. Because the symmetric hollow subspace is orthogonal to the antisymmetric antihollow subspace (§2.2.3), then for $B \in \mathbb{S}_h^N$

$$\operatorname{tr}\left(B^T\left(\frac{1}{2}(H-H^T)+\delta^2(H)\right)\right)=0\tag{1109}$$

so the objective function is equivalent to

$$\|B-H\|_{\mathrm{F}}^2 \;\equiv\; \left\|B-\left(\frac{1}{2}(H+H^T)-\delta^2(H)\right)\right\|_{\mathrm{F}}^2 + \;\left\|\frac{1}{2}(H-H^T)+\delta^2(H)\right\|_{\mathrm{F}}^2\tag{1110}$$

This means the antisymmetric antihollow part of given matrix $H$ would be ignored by minimization with respect to symmetric hollow variable $B$ under the Frobenius norm; *id est*, minimization proceeds as though given the symmetric hollow part of $H$.

This action of the Frobenius norm (1110) is effectively a Euclidean projection (minimum-distance projection) of $H$ on the symmetric hollow subspace $\mathbb{S}_h^N$ prior to minimization. Thus minimization proceeds inherently following the correct order for projection on $\mathbb{S}_h^N \cap \mathcal{C}$. Therefore we may either assume $H \in \mathbb{S}_h^N$, 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.

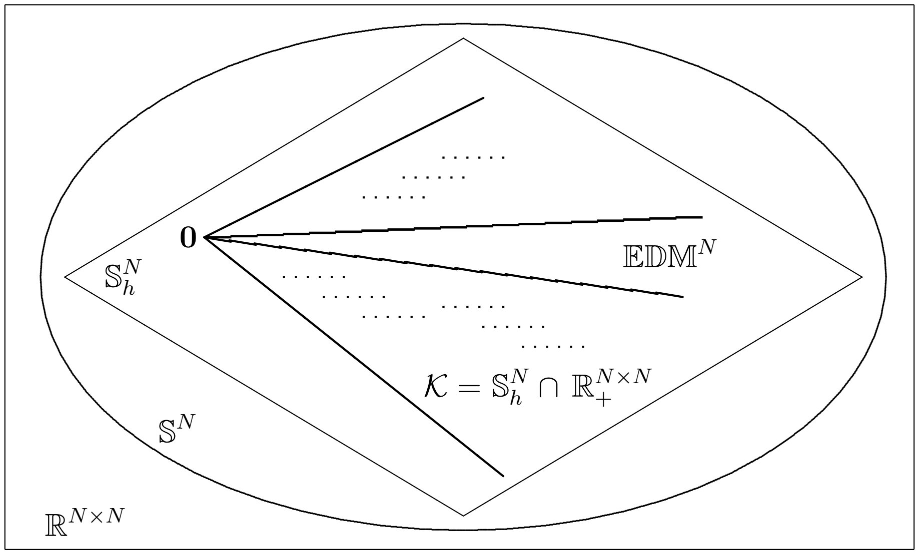



Figure 109: Pseudo-Venn diagram: The EDM cone belongs to the intersection of the symmetric hollow subspace with the nonnegative orthant; $\mathbb{EDM}^N \subseteq \mathcal{K}$ (708). $\mathbb{EDM}^N$ cannot exist outside $\mathbb{S}_h^N$ , but $\mathbb{R}_+^{N\times N}$ does.

#### 7.0.1.2 Egregious input error under nonnegativity demand

More pertinent to the optimization problems presented herein where

$$\mathcal{C} \triangleq \mathbb{EDM}^N \subseteq \mathcal{K} = \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N} \tag{1111}$$

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 $\mathbb{EDM}^N$ would be guaranteed incorrect (out of order).

Now comes a surprising fact: Even were we to correctly follow the order-of-projection rule and provide $H \in \mathcal{K}$ prior to optimization, then the ensuing projection on $\mathbb{EDM}^N$ will be incorrect whenever input $H$ has negative entries and some proximity problem demands nonnegative input $H$ .

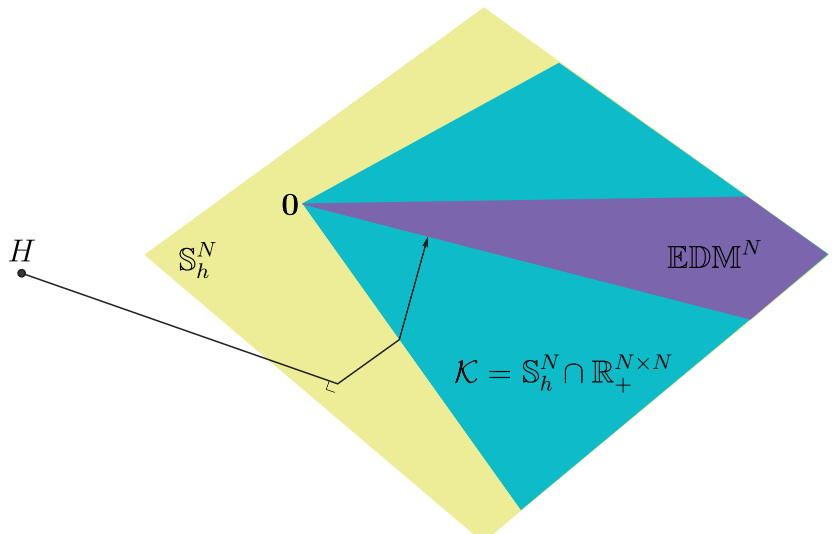



Figure 110: Pseudo-Venn diagram from Figure **109** showing elbow placed in path of projection of $H$ on $\mathbb{EDM}^N \subset \mathbb{S}_h^N$ by an optimization problem demanding nonnegative input matrix $H$. The first two line segments leading away from $H$ result from correct order-of-projection required to provide nonnegative $H$ prior to optimization. Were $H$ nonnegative, then its projection on $\mathbb{S}_h^N$ would instead belong to $\mathcal{K}$; making the elbow disappear. (*confer* Figure **121**)

This is best understood referring to Figure **109**: Suppose nonnegative input $H$ is demanded, and then the problem realization correctly projects its input first on $\mathbb{S}_h^N$ and then directly on $\mathcal{C} = \mathbb{EDM}^N$. 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 $\mathbb{S}_h^N$ first). Yet such an imposition prior to projection on $\mathbb{EDM}^N$ generally introduces an *elbow* into the path of projection (illustrated in Figure **110**) 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.

### 7.0.2 Lower bound

Most of the problems we encounter in this chapter have the general form:

$$\begin{array}{cl} \underset{B}{\text{minimize}} & \|B - A\|_\mathrm{F} \\ \text{subject to} & B \in \mathcal{C} \end{array} \tag{1112}$$

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 the Frobenius norm when squared is a strictly convex function of variable $B$ and because the optimal solution is the same regardless of the square (1466). When $\mathcal{C}$ is a subspace, then the direction of projection is orthogonal to $\mathcal{C}$.

Denoting by $A \stackrel{\Delta}{=} U_A \Sigma_A Q_A^T$ and $B \stackrel{\Delta}{=} U_B \Sigma_B Q_B^T$ 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;

$$\|\Sigma_B - \Sigma_A\|_\mathrm{F} \le \inf_{U_A,\, U_B,\, Q_A,\, Q_B} \|B - A\|_\mathrm{F} \tag{1113}$$

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 [246, §II.3]. [150, §7.4.51]

### 7.0.3 Problem approach

Problems traditionally posed in terms of point position $\{x_i\,,\ i=1\ldots N\}$, such as

$$\underset{\{x_i\}}{\text{minimize}} \sum_{i\,,\,j\,\in\,\mathcal{I}} (\|x_i - x_j\| - h_{ij})^2 \tag{1114}$$

or

$$\underset{\{x_i\}}{\text{minimize}} \sum_{i\,,\,j\,\in\,\mathcal{I}} (\|x_i - x_j\|^2 - h_{ij}^2)^2 \tag{1115}$$

(where $\mathcal{I}$ is an abstract set of indices and $h_{ij}$ is given data) are everywhere converted (in this book) 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 an infinitude of local minima in (1114) or (1115) [70].

### 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 (1116.1), (1116.2), and (1116.3): [258]

$$\begin{array}{cllcll}
 & \underset{D}{\text{minimize}} & \|-V(D-H)V\|_{\mathrm F}^2 & & \underset{\sqrt[\circ]{D}}{\text{minimize}} & \|\sqrt[\circ]{D}-H\|_{\mathrm F}^2 \\
(1) & \text{subject to} & \operatorname{rank} VDV \le \rho & & \text{subject to} & \operatorname{rank} VDV \le \rho \quad (2) \\
 & & D \in \mathbb{EDM}^N & & & \sqrt[\circ]{D} \in \sqrt{\mathbb{EDM}^N} \\
 & \underset{D}{\text{minimize}} & \|D-H\|_{\mathrm F}^2 & & \underset{\sqrt[\circ]{D}}{\text{minimize}} & \|-V(\sqrt[\circ]{D}-H)V\|_{\mathrm F}^2 \\
(3) & \text{subject to} & \operatorname{rank} VDV \le \rho & & \text{subject to} & \operatorname{rank} VDV \le \rho \quad (4) \\
 & & D \in \mathbb{EDM}^N & & & \sqrt[\circ]{D} \in \sqrt{\mathbb{EDM}^N}
\end{array} \tag{1116}$$

where we have made explicit an imposed upper bound $\rho$ on affine dimension

$$r \;=\; \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} \;=\; \operatorname{rank} V D V \tag{851}$$

that is benign when $\rho = N-1$, and where $D \overset{\Delta}{=} [d_{ij}]$ and $\sqrt[\circ]{D} \overset{\Delta}{=} [\sqrt{d_{ij}}\,]$. Problems (1116.2) and (1116.3) are Euclidean projections of a vectorized matrix $H$ on an EDM cone (§6.3), whereas problems (1116.1) and (1116.4) are Euclidean projections of a vectorized matrix $-VHV$ on a PSD cone. Problem (1116.4) is not posed in the literature because it has limited theoretical foundation.[7.2]

Analytical solution to (1116.1) 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 (1116.1) 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 (1114), problem (1116.2) is a variant of what is known in statistics literature as the *stress problem.* [39, p.34] [68] [267] Problems (1116.2) and (1116.3) are convex optimization problems in $D$ for the case $\rho = N-1$. Even with the rank constraint removed from (1116.2), we will see the convex problem remaining inherently minimizes affine dimension.

Generally speaking, each problem in (1116) produces a different result because there is no isometry relating them. Of the various auxiliary $V$-matrices (§B.4), the geometric centering matrix $V$ (732) appears in the literature most often although $V_{\mathcal{N}}$ (715) is the auxiliary matrix naturally consequent to Schoenberg's seminal exposition [234]. Substitution of any auxiliary matrix or its pseudoinverse into these problems produces another valid problem.

Substitution of $V_{\mathcal{N}}^T$ for left-hand $V$ in (1116.1), in particular, produces a different result because

$$\begin{array}{ll} \underset{D}{\text{minimize}} & \|-V(D-H)V\|_{\mathrm{F}}^2 \\ \text{subject to} & D \in \mathbb{EDM}^N \end{array} \tag{1117}$$

[7.2] $D \in \mathbb{EDM}^N \Rightarrow \sqrt[\circ]{D} \in \mathbb{EDM}^N,\; -V\sqrt[\circ]{D}\,V \in \mathbb{S}_+^N$ (§5.10) 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

$$\begin{array}{ll}\underset{D}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-H)V_{\mathcal{N}}\|_{\mathrm{F}}^2\\ \text{subject to} & D\in\mathbb{EDM}^N\end{array}\tag{1118}$$

attains Euclidean distance of vectorized $-V_{\mathcal{N}}^T H V_{\mathcal{N}}$ to the positive semidefinite cone in isometrically isomorphic subspace $\mathbb{R}^{N(N-1)/2}$; quite different projections[7.3] regardless of whether affine dimension is constrained. But substitution of auxiliary matrix $V_{\mathcal{W}}^T$ (§B.4.3) or $V_{\mathcal{N}}^\dagger$ yields the same result as (1116.1) because $V=V_{\mathcal{W}}V_{\mathcal{W}}^T=V_{\mathcal{N}}V_{\mathcal{N}}^\dagger$ ; *id est*,

$$\begin{aligned}\|-V(D-H)V\|_{\mathrm{F}}^2 &= \|-V_{\mathcal{W}}V_{\mathcal{W}}^T(D-H)V_{\mathcal{W}}V_{\mathcal{W}}^T\|_{\mathrm{F}}^2 = \|-V_{\mathcal{W}}^T(D-H)V_{\mathcal{W}}\|_{\mathrm{F}}^2\\ &= \|-V_{\mathcal{N}}V_{\mathcal{N}}^\dagger(D-H)V_{\mathcal{N}}V_{\mathcal{N}}^\dagger\|_{\mathrm{F}}^2 = \|-V_{\mathcal{N}}^\dagger(D-H)V_{\mathcal{N}}\|_{\mathrm{F}}^2\end{aligned}\tag{1119}$$

We see no compelling reason to prefer one particular auxiliary $V$-matrix over another. Each has its own coherent interpretations; *e.g.*, §5.4.2, §6.7. Neither can we say any particular problem formulation produces generally better results than another.

## 7.1 First prevalent problem: Projection on PSD cone

This first problem

$$\left.\begin{array}{ll}\underset{D}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-H)V_{\mathcal{N}}\|_{\mathrm{F}}^2\\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}}\leq\rho\\ & D\in\mathbb{EDM}^N\end{array}\right\}\quad\text{Problem 1}\tag{1120}$$

---

[7.3] The isomorphism $T(Y)=V_{\mathcal{N}}^{\dagger T}YV_{\mathcal{N}}^\dagger$ onto $\mathbb{S}_c^N=\{VXV \mid X\in\mathbb{S}^N\}$ relates the map in (1118) to that in (1117), but is not an isometry.

poses a Euclidean projection of $-V_{\mathcal{N}}^T H V_{\mathcal{N}}$ 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. (728) [234] 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 [188, thm.14.4.2] for all nonnegative $\rho \leq N-1$ .[7.4]

We assume only that the given measurement matrix $H$ is symmetric;[7.5]

$$H \in \mathbb{S}^N \tag{1121}$$

Arranging the eigenvalues $\lambda_i$ of $-V_{\mathcal{N}}^T H V_{\mathcal{N}}$ in nonincreasing order for all $i$ , $\lambda_i \geq \lambda_{i+1}$ with $v_i$ the corresponding $i^{\text{th}}$ eigenvector, then an optimal solution to Problem 1 is [264, §2]

$$-V_{\mathcal{N}}^T D^\star V_{\mathcal{N}} = \sum_{i=1}^{\rho} \max\{0,\ \lambda_i\}\, v_i v_i^T \tag{1122}$$

where

$$-V_{\mathcal{N}}^T H V_{\mathcal{N}} \triangleq \sum_{i=1}^{N-1} \lambda_i\, v_i v_i^T \in \mathbb{S}^{N-1} \tag{1123}$$

is an eigenvalue decomposition and

$$D^\star \in \mathbb{EDM}^N \tag{1124}$$

is an optimal Euclidean distance matrix.

In §7.1.4 we show how to transform Problem 1 to a convex optimization problem for any $\rho$ .

---

[7.4] being first pronounced in the context of multidimensional scaling by Mardia [187] in 1978 who attributes the generic result (§7.1.2) to Eckart & Young, 1936 [85].

[7.5] 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. (*confer* Figure **110**)

### 7.1.1 Closest-EDM Problem 1, convex case

**7.1.1.0.1 Proof.** *Solution* (1122), *convex case.*
When desired affine dimension is unconstrained, $\rho = N-1$, the rank function disappears from (1120) leaving a convex optimization problem; a simple unique minimum-distance projection on the positive semidefinite cone $\mathbb{S}_+^{N-1}$ : *videlicet*

$$\begin{array}{cl}\underset{D\in\mathbb{S}_h^N}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-H)V_{\mathcal{N}}\|_{\mathrm{F}}^2\\ \text{subject to} & -V_{\mathcal{N}}^T D V_{\mathcal{N}} \succeq 0\end{array} \tag{1125}$$

by (728). Because

$$\mathbb{S}^{N-1} = -V_{\mathcal{N}}^T \mathbb{S}_h^N V_{\mathcal{N}} \qquad (820)$$

then the necessary and sufficient conditions for projection in isometrically isomorphic $\mathbb{R}^{N(N-1)/2}$ on the self-dual (321) positive semidefinite cone $\mathbb{S}_+^{N-1}$ are:[7.6] (§E.9.2.0.1) (1375) (*confer* (1800))

$$\begin{array}{l}-V_{\mathcal{N}}^T D^\star V_{\mathcal{N}} \succeq 0\\ -V_{\mathcal{N}}^T D^\star V_{\mathcal{N}}\left(-V_{\mathcal{N}}^T D^\star V_{\mathcal{N}} + V_{\mathcal{N}}^T H V_{\mathcal{N}}\right) = \mathbf{0}\\ -V_{\mathcal{N}}^T D^\star V_{\mathcal{N}} + V_{\mathcal{N}}^T H V_{\mathcal{N}} \succeq 0\end{array} \tag{1126}$$

Symmetric $-V_{\mathcal{N}}^T H V_{\mathcal{N}}$ is diagonalizable hence decomposable in terms of its eigenvectors $v$ and eigenvalues $\lambda$ as in (1123). Therefore (*confer* (1122))

$$-V_{\mathcal{N}}^T D^\star V_{\mathcal{N}} = \sum_{i=1}^{N-1} \max\{0,\ \lambda_i\} v_i v_i^T \tag{1127}$$

satisfies (1126), optimally solving (1125). To see that, recall: these eigenvectors constitute an orthogonal set and

$$-V_{\mathcal{N}}^T D^\star V_{\mathcal{N}} + V_{\mathcal{N}}^T H V_{\mathcal{N}} = -\sum_{i=1}^{N-1} \min\{0,\ \lambda_i\} v_i v_i^T \tag{1128}$$

♦

[7.6] These conditions for projection on a convex cone are identical to the Karush-Kuhn-Tucker (KKT) optimality conditions for problem (1125).

### 7.1.2 Generic problem

Prior to determination of $D^\star$, analytical solution (1122) to Problem 1 is equivalent to solution of a generic rank-constrained projection problem; a 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 $\rho < N-1$):

$$\left.\begin{array}{cl}\underset{B\in\mathbb{S}^{N-1}}{\text{minimize}} & \|B-A\|_{\mathrm{F}}^2\\ \text{subject to} & \operatorname{rank} B\leq\rho\\ & B\succeq 0\end{array}\right\}\quad\text{Generic 1}\tag{1129}$$

whose optimal solution (Eckart & Young) [85]

$$B^\star \stackrel{\Delta}{=} -V_{\mathcal{N}}^T D^\star V_{\mathcal{N}} = \sum_{i=1}^{\rho}\max\{0,\,\lambda_i\}\,v_i v_i^T \in \mathbb{S}^{N-1}\tag{1122}$$

is well known given desired affine dimension $\rho$ and

$$A \stackrel{\Delta}{=} -V_{\mathcal{N}}^T H V_{\mathcal{N}} = \sum_{i=1}^{N-1}\lambda_i\,v_i v_i^T \in \mathbb{S}^{N-1}\tag{1123}$$

Once optimal $B^\star$ is found, the technique of §5.12 can be used to determine a uniquely corresponding optimal Euclidean distance matrix $D^\star$; a unique correspondence by injectivity arguments in §5.6.2.

#### 7.1.2.1 Projection on rank $\rho$ subset of PSD cone

Because Problem 1 is the same as

$$\begin{array}{cl}\underset{D\in\mathbb{S}_h^N}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-H)V_{\mathcal{N}}\|_{\mathrm{F}}^2\\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}}\leq\rho\\ & -V_{\mathcal{N}}^T D V_{\mathcal{N}}\succeq 0\end{array}\tag{1130}$$

and because (820) provides invertible mapping to the generic problem, then Problem 1 is truly a Euclidean projection of vectorized $-V_{\mathcal{N}}^T H V_{\mathcal{N}}$ 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 $\rho$ ;[7.7] called rank $\rho$ subset: (185) (219)

$$\mathbb{S}_+^{N-1} \setminus \mathbb{S}_+^{N-1}(\rho+1) \;=\; \{X \in \mathbb{S}_+^{N-1} \mid \operatorname{rank} X \leq \rho\} \tag{224}$$

### 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.*, (1142). In this section we develop a method of spectral projection for constraining rank of positive semidefinite matrices in a proximity problem like (1129). 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 $\pi(x) : \mathbb{R}^n \to \mathbb{R}^n$ 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 ($R$ , §B.5.4) nonincreasingly ordered ($\pi$) vector ($\delta$) of eigenvalues

$$\pi\big(\delta(R^T \Lambda R)\big) \tag{1131}$$

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). △

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.5.2). 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 (1122). Although solution to generic problem (1129) is known since 1936 [85], Trosset [264, §2] first observed its equivalence in 1997 to projection

[7.7] 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.

of an ordered vector of eigenvalues (in diagonal matrix $\Lambda$) on a subset of the monotone nonnegative cone (§2.13.9.4.1)

$$\mathcal{K}_{\mathcal{M}+} = \{\upsilon \mid \upsilon_1 \geq \upsilon_2 \geq \cdots \geq \upsilon_{N-1} \geq 0\} \subseteq \mathbb{R}^{N-1}_+ \tag{370}$$

Of interest, momentarily, is only the smallest convex subset of the monotone nonnegative cone $\mathcal{K}_{\mathcal{M}+}$ containing every nonincreasingly ordered eigenspectrum corresponding to a rank $\rho$ subset of the positive semidefinite cone $\mathbb{S}^{N-1}_+$; *id est*,

$$\mathcal{K}^{\rho}_{\mathcal{M}+} \triangleq \{\upsilon \in \mathbb{R}^{\rho} \mid \upsilon_1 \geq \upsilon_2 \geq \cdots \geq \upsilon_{\rho} \geq 0\} \subseteq \mathbb{R}^{\rho}_+ \tag{1132}$$

a pointed polyhedral cone, a $\rho$-dimensional convex subset of the monotone nonnegative cone $\mathcal{K}_{\mathcal{M}+} \subseteq \mathbb{R}^{N-1}_+$ having property, for $\lambda$ denoting eigenspectra,

$$\begin{bmatrix} \mathcal{K}^{\rho}_{\mathcal{M}+} \\ \mathbf{0} \end{bmatrix} = \pi(\lambda(\text{rank } \rho \text{ subset})) \subseteq \mathcal{K}^{N-1}_{\mathcal{M}+} \triangleq \mathcal{K}_{\mathcal{M}+} \tag{1133}$$

For each and every elemental eigenspectrum

$$\gamma \in \lambda(\text{rank } \rho \text{ subset}) \subseteq \mathbb{R}^{N-1}_+ \tag{1134}$$

of the rank $\rho$ subset (ordered or unordered in $\lambda$), there is a nonlinear surjection $\pi(\gamma)$ onto $\mathcal{K}^{\rho}_{\mathcal{M}+}$.

**7.1.3.0.2 Exercise.** *Smallest spectral cone.*
Prove that there is no convex subset of $\mathcal{K}_{\mathcal{M}+}$ smaller than $\mathcal{K}^{\rho}_{\mathcal{M}+}$ 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) [129, §X] [41, §1.2] Any vectors $\sigma$ and $\gamma$ in $\mathbb{R}^{N-1}$ satisfy a tight inequality

$$\pi(\sigma)^T\pi(\gamma) \;\geq\; \sigma^T\gamma \;\geq\; \pi(\sigma)^T\Xi\,\pi(\gamma) \tag{1135}$$

where $\Xi$ is the order-reversing permutation matrix defined in (1507), 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

$$\|\pi(\sigma) - \pi(\gamma)\| \leq \|\sigma - \gamma\| \tag{1136}$$

where nonlinear function $\pi(\gamma)$ sorts vector $\gamma$ into nonincreasing order thereby providing the least lower bound with respect to every possible sorting. $\diamond$

Given $\gamma \in \mathbb{R}^{N-1}$

$$\inf_{\sigma\in\mathbb{R}_+^{N-1}} \|\sigma-\gamma\| = \inf_{\sigma\in\mathbb{R}_+^{N-1}} \|\pi(\sigma)-\pi(\gamma)\| = \inf_{\sigma\in\mathbb{R}_+^{N-1}} \|\sigma-\pi(\gamma)\| = \inf_{\sigma\in\mathcal{K}_{\mathcal{M}+}} \|\sigma-\pi(\gamma)\| \tag{1137}$$

Yet for $\gamma$ representing an arbitrary vector of eigenvalues, because

$$\inf_{\sigma\in\left[\begin{smallmatrix}\mathbb{R}_+^\rho\\ \mathbf{0}\end{smallmatrix}\right]} \|\sigma - \gamma\|^2 \geq \inf_{\sigma\in\left[\begin{smallmatrix}\mathbb{R}_+^\rho\\ \mathbf{0}\end{smallmatrix}\right]} \|\sigma - \pi(\gamma)\|^2 = \inf_{\sigma\in\left[\begin{smallmatrix}\mathcal{K}_{\mathcal{M}+}^\rho\\ \mathbf{0}\end{smallmatrix}\right]} \|\sigma - \pi(\gamma)\|^2 \tag{1138}$$

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 $\pi(\gamma)_{1:\rho} \succeq \pi(\gamma_{1:\rho})$

$$\begin{aligned}
\inf_{\sigma\in\left[\begin{smallmatrix}\mathbb{R}_+^\rho\\ \mathbf{0}\end{smallmatrix}\right]} \|\sigma - \gamma\|^2 &= \gamma_{\rho+1:N-1}^T \gamma_{\rho+1:N-1} + \inf_{\sigma\in\mathbb{R}_+^{N-1}} \|\sigma_{1:\rho} - \gamma_{1:\rho}\|^2 \\
&= \gamma^T\gamma + \inf_{\sigma\in\mathbb{R}_+^{N-1}} \sigma_{1:\rho}^T\sigma_{1:\rho} - 2\sigma_{1:\rho}^T\gamma_{1:\rho} \\
&\geq \gamma^T\gamma + \inf_{\sigma\in\mathbb{R}_+^{N-1}} \sigma_{1:\rho}^T\sigma_{1:\rho} - 2\sigma_{1:\rho}^T\pi(\gamma)_{1:\rho} \\
\inf_{\sigma\in\left[\begin{smallmatrix}\mathbb{R}_+^\rho\\ \mathbf{0}\end{smallmatrix}\right]} \|\sigma - \gamma\|^2 &\geq \inf_{\sigma\in\left[\begin{smallmatrix}\mathbb{R}_+^\rho\\ \mathbf{0}\end{smallmatrix}\right]} \|\sigma - \pi(\gamma)\|^2
\end{aligned} \tag{1139}$$

$\blacklozenge$

#### 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 $\mathcal{K}_{\mathcal{M}+}^{\rho}$ of the monotone nonnegative cone become equivalent to unique spectral projection on a subset $\mathbb{R}_+^{\rho}$ 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 $\mathcal{K}_{\mathcal{M}+}^{\rho}$ of the monotone nonnegative cone in that same subspace.

### 7.1.4 Closest-EDM Problem 1, "nonconvex" case

Trosset's proof of solution (1122), for projection on a rank $\rho$ subset of the positive semidefinite cone $\mathbb{S}_+^{N-1}$, was algebraic in nature. [264, §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* (1122), *nonconvex case.*
As explained in §7.1.2, we may instead work with the more facile generic problem (1129). With diagonalization of unknown

$$B \triangleq U\Upsilon U^T \in \mathbb{S}^{N-1} \tag{1140}$$

given desired affine dimension $0 \le \rho \le N-1$ and diagonalizable

$$A \triangleq Q\Lambda Q^T \in \mathbb{S}^{N-1} \tag{1141}$$

having eigenvalues in $\Lambda$ arranged in nonincreasing order, by (40) the generic problem is equivalent to

$$\begin{array}{cl}\underset{B\in\mathbb{S}^{N-1}}{\text{minimize}} & \|B-A\|_{\mathrm{F}}^2\\ \text{subject to} & \operatorname{rank} B \le \rho\\ & B \succeq 0\end{array} \quad\equiv\quad \begin{array}{cl}\underset{R\,,\,\Upsilon}{\text{minimize}} & \|\Upsilon - R^T\Lambda R\|_{\mathrm{F}}^2\\ \text{subject to} & \operatorname{rank}\Upsilon \le \rho\\ & \Upsilon \succeq 0\\ & R^{-1} = R^T\end{array} \tag{1142}$$

where

$$R \overset{\Delta}{=} Q^T U \in \mathbb{R}^{N-1\times N-1} \tag{1143}$$

in $U$ on the set of orthogonal matrices is a linear bijection. We propose solving (1142) by instead solving the problem sequence:

$$\begin{array}{cll} \underset{\Upsilon}{\text{minimize}} & \|\Upsilon - R^T\Lambda R\|_{\mathrm F}^2 & \\ \text{subject to} & \operatorname{rank}\Upsilon \le \rho & \text{(a)} \\ & \Upsilon \succeq 0 & \\ & & \\ \underset{R}{\text{minimize}} & \|\Upsilon^\star - R^T\Lambda R\|_{\mathrm F}^2 & \text{(b)} \\ \text{subject to} & R^{-1} = R^T & \end{array} \tag{1144}$$

Problem (1144a) is equivalent to: (1) orthogonal projection of $R^T\Lambda R$ 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 $\begin{bmatrix}\mathbb{R}_+^\rho \\ \mathbf{0}\end{bmatrix}$. (§E.9.5) Projection on that $N-1$-dimensional subspace amounts to zeroing $R^T\Lambda R$ at all entries off the main diagonal; thus, the equivalent sequence leading with a spectral projection:

$$\begin{array}{cll} \underset{\Upsilon}{\text{minimize}} & \|\,\delta(\Upsilon) - \pi\big(\delta(R^T\Lambda R)\big)\,\|^2 & \\ \text{subject to} & \delta(\Upsilon) \in \begin{bmatrix}\mathbb{R}_+^\rho \\ \mathbf{0}\end{bmatrix} & \text{(a)} \\ & & \\ \underset{R}{\text{minimize}} & \|\Upsilon^\star - R^T\Lambda R\|_{\mathrm F}^2 & \text{(b)} \\ \text{subject to} & R^{-1} = R^T & \end{array} \tag{1145}$$

Because any permutation matrix is an orthogonal matrix, it is always feasible that $\delta(R^T\Lambda R)\in\mathbb{R}^{N-1}$ be arranged in nonincreasing order; hence, the permutation operator $\pi$ . Unique minimum-distance projection of vector $\pi\big(\delta(R^T\Lambda R)\big)$ on the $\rho$-dimensional subset $\begin{bmatrix}\mathbb{R}_+^\rho \\ \mathbf{0}\end{bmatrix}$ of nonnegative orthant

$\mathbb{R}_+^{N-1}$ requires: (§E.9.2.0.1)

$$\begin{aligned} \delta(\Upsilon^\star)_{\rho+1:N-1} &= \mathbf{0} \\ \delta(\Upsilon^\star) &\succeq 0 \\ \delta(\Upsilon^\star)^T\big(\delta(\Upsilon^\star) - \pi\big(\delta(R^T\Lambda R)\big)\big) &= 0 \\ \delta(\Upsilon^\star) - \pi\big(\delta(R^T\Lambda R)\big) &\succeq 0 \end{aligned} \tag{1146}$$

which are necessary and sufficient conditions. Any value $\Upsilon^\star$ satisfying conditions (1146) is optimal for (1145a). So

$$\delta(\Upsilon^\star)_i = \begin{cases} \max\left\{0\,,\ \pi\big(\delta(R^T\Lambda R)\big)_i\right\}, & i=1\ldots\rho \\ 0\,, & i=\rho+1\ldots N-1 \end{cases} \tag{1147}$$

specifies an optimal solution. The lower bound on the objective with respect to $R$ in (1145b) is tight; by (1113)

$$\|\,|\Upsilon^\star| - |\Lambda|\,\|_F \;\le\; \|\Upsilon^\star - R^T\Lambda R\|_F \tag{1148}$$

where $|\ |$ denotes absolute entry-value. For selection of $\Upsilon^\star$ as in (1147), this lower bound is attained when (*confer* §C.4.2.2)

$$R^\star = I \tag{1149}$$

which is the known solution. ◆

#### 7.1.4.1 Significance

Importance of this well-known [85] optimal solution (1122) 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 (224) of the positive semidefinite cone (generally, a nonconvex subset of its boundary) from either the exterior or interior of that cone.[7.8] By problem transformation to the spectral domain, projection on a rank $\rho$ subset becomes a convex optimization problem.

[7.8] Projection on the boundary from the interior of a convex Euclidean body is generally a nonconvex problem.

- This solution is closed-form and the only method known for enforcing a constraint on rank of an EDM in projection problems such as (1116).

- 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 $U^\star = Q$ , 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.[7.9]

- For the convex case problem, 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 .[7.10]

### 7.1.5 Problem 1 in spectral norm, convex case

When instead we pose the matrix 2-norm (*spectral norm*) in Problem 1 (1120) for the convex case $\rho = N-1$, then the new problem

$$\begin{array}{ll}\underset{D}{\text{minimize}} & \|-V_{\mathcal{N}}^T(D-H)V_{\mathcal{N}}\|_2\\ \text{subject to} & D\in\mathbb{EDM}^N\end{array}\tag{1150}$$

is convex although its solution is not necessarily unique;[7.11] giving rise to nonorthogonal projection (§E.1) on the positive semidefinite cone $\mathbb{S}_+^{N-1}$. Indeed, its solution set includes the Frobenius solution (1122) for the convex case whenever $-V_{\mathcal{N}}^T H V_{\mathcal{N}}$ is a normal matrix. [133, §1] [127] [46, §8.1.1] *Singular value problem* (1150) is equivalent to

$$\begin{array}{ll}\underset{\mu\,,\,D}{\text{minimize}} & \mu\\ \text{subject to} & -\mu I \preceq -V_{\mathcal{N}}^T(D-H)V_{\mathcal{N}} \preceq \mu I\\ & D\in\mathbb{EDM}^N\end{array}\tag{1151}$$

---

[7.9] 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.

[7.10] Uncertainty of uniqueness prevents the erroneous conclusion that a rank $\rho$ subset (185) were a convex body by the *Bunt-Motzkin theorem* (§E.9.0.0.1).

[7.11] For each and every $|t|\leq 2$, for example, $\begin{bmatrix}2&0\\0&t\end{bmatrix}$ has the same spectral-norm value.

where

$$\mu^\star = \max_i \left\{ \left|\lambda\big(-V_{\mathcal{N}}^T(D^\star - H)V_{\mathcal{N}}\big)_i\right| \ , \quad i=1\ldots N-1 \right\} \in \mathbb{R}_+ \tag{1152}$$

the minimized largest absolute eigenvalue (due to matrix symmetry).

For lack of a 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

$$\sqrt[\circ]{D} \triangleq [\sqrt{d_{ij}}\,] \in \mathcal{K} = \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N} \tag{1153}$$

be an unknown matrix of absolute distance; *id est*,

$$D = [d_{ij}] \triangleq \sqrt[\circ]{D} \circ \sqrt[\circ]{D} \in \mathbb{EDM}^N \tag{1154}$$

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 $\operatorname{rel}\partial\sqrt{\mathbb{EDM}^N}$: (§6.3, *confer* Figure **95**(b))

$$\left.\begin{array}{cl} \underset{\sqrt[\circ]{D}}{\text{minimize}} & \|\sqrt[\circ]{D} - H\|_{\mathrm{F}}^2 \\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^T D V_{\mathcal{N}} \leq \rho \\ & \sqrt[\circ]{D} \in \sqrt{\mathbb{EDM}^N} \end{array}\right\} \quad \text{Problem 2} \tag{1155}$$

where

$$\sqrt{\mathbb{EDM}^N} = \{\sqrt[\circ]{D} \mid D \in \mathbb{EDM}^N\} \qquad (989)$$

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

$$\|\sqrt[\circ]{D} - H\|_{\mathrm{F}}^2 = \sum_{i,j}(\sqrt{d_{ij}} - h_{ij})^2 \tag{1156}$$

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;

$$H = [h_{ij}] \in \mathbb{R}_+^{N\times N} \tag{1157}$$

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 $\mathcal{K} = \mathbb{S}_h^N \cap \mathbb{R}_+^{N\times N}$ (1107) prior to application of this proposed solution is incorrect.

### 7.2.1 Convex case

When $\rho = N-1$, the rank constraint vanishes and a convex problem emerges:[7.12]

$$\begin{array}{ll}\underset{\sqrt[\circ]{D}}{\text{minimize}} & \|\sqrt[\circ]{D} - H\|_{\mathrm{F}}^2 \\ \text{subject to} & \sqrt[\circ]{D} \in \sqrt{\mathbb{EDM}^N}\end{array} \quad\Leftrightarrow\quad \begin{array}{ll}\underset{D}{\text{minimize}} & \sum\limits_{i,j} d_{ij} - 2h_{ij}\sqrt{d_{ij}} \,+\, h_{ij}^2 \\ \text{subject to} & D \in \mathbb{EDM}^N\end{array} \tag{1158}$$

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 [46, §3.2.1] and because the feasible set is convex in $D$, we have a convex optimization problem:

$$\begin{array}{ll}\underset{D}{\text{minimize}} & \mathbf{1}^T(D - 2H\circ\sqrt[\circ]{D}\,)\mathbf{1} + \|H\|_{\mathrm{F}}^2 \\ \text{subject to} & D \in \mathbb{EDM}^N\end{array} \tag{1159}$$

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 $D^\star$ for this second prevalent problem depends upon nonnegativity of $H$ and a convex feasible set (§3.1.2).[7.13]

---

[7.12] still thought to be a nonconvex problem as late as 1997 [267] even though discovered convex by de Leeuw in 1993. [68] [39, §13.6] Yet using methods from §3, it can be easily ascertained: $\|\sqrt[\circ]{D} - H\|_{\mathrm{F}}$ is not convex in $D$.

[7.13] The transformed problem in variable $D$ no longer describes Euclidean projection on an EDM cone. Otherwise we might erroneously conclude $\sqrt{\mathbb{EDM}^N}$ were a convex body by the *Bunt-Motzkin theorem* (§E.9.0.0.1).

#### 7.2.1.1 Equivalent semidefinite program, Problem 2, convex case

Convex problem (1158) is numerically solvable for its global minimum using an interior-point method [46, §11] [299] [214] [204] [290] [9] [98]. We translate (1158) 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 [253] [117] [269] [27] [291] [292] [293].

Substituting a new matrix variable $Y \overset{\Delta}{=} [y_{ij}] \in \mathbb{R}_+^{N\times N}$

$$h_{ij}\sqrt{d_{ij}} \;\leftarrow\; y_{ij} \tag{1160}$$

Boyd proposes: problem (1158) is equivalent to the semidefinite program

$$\begin{array}{cl}\underset{D\,,\,Y}{\text{minimize}} & \displaystyle\sum_{i,j} d_{ij} - 2y_{ij} + h_{ij}^2\\ \text{subject to} & \begin{bmatrix} d_{ij} & y_{ij}\\ y_{ij} & h_{ij}^2\end{bmatrix} \succeq 0\;, \quad i,j=1\ldots N\\ & D \in \mathbb{EDM}^N\end{array} \tag{1161}$$

To see that, recall $d_{ij}\geq 0$ is implicit to $D\in\mathbb{EDM}^N$ (§5.8.1, (728)). So when $H\in\mathbb{R}_+^{N\times N}$ is nonnegative as assumed,

$$\begin{bmatrix} d_{ij} & y_{ij}\\ y_{ij} & h_{ij}^2\end{bmatrix} \succeq 0 \;\Leftrightarrow\; h_{ij}\sqrt{d_{ij}} \geq \sqrt{y_{ij}^2} \tag{1162}$$

Minimization of the objective function implies maximization of $y_{ij}$ that is bounded above. Hence nonnegativity of $y_{ij}$ is implicit to (1161) and, as desired, $y_{ij}\rightarrow h_{ij}\sqrt{d_{ij}}$ as optimization proceeds. ♦

If the given matrix $H$ is now assumed symmetric and nonnegative,

$$H = [h_{ij}] \in \mathbb{S}^N \cap \mathbb{R}_+^{N\times N} \tag{1163}$$

then $Y = H\circ\sqrt[\circ]{D}$ must belong to $\mathcal{K}=\mathbb{S}_h^N\cap\mathbb{R}_+^{N\times N}$ (1107). Because $Y\in\mathbb{S}_h^N$ (§B.4.2 no.20), then

$$\|\sqrt[\circ]{D} - H\|_{\mathrm{F}}^2 \;=\; \sum_{i,j} d_{ij} - 2y_{ij} + h_{ij}^2 \;=\; -N\operatorname{tr}\big(V(D-2\,Y)V\big) + \|H\|_{\mathrm{F}}^2 \tag{1164}$$

So convex problem (1161) is equivalent to the semidefinite program

$$\begin{array}{cl}\underset{D\,,\,Y}{\text{minimize}} & -\operatorname{tr}\big(V(D-2\,Y)V\big)\\ \text{subject to} & \left[\begin{array}{cc} d_{ij} & y_{ij}\\ y_{ij} & h_{ij}^2\end{array}\right]\succeq 0\,, \quad j>i=1\ldots N-1\\ & Y\in\mathbb{S}_h^N\\ & D\in\mathbb{EDM}^N\end{array} \tag{1165}$$

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 (1165) in Gram-form because Gram matrix $G$ is a bidirectional bridge between point list $X$ and distance matrix $D$; *e.g.*, Example 5.4.2.2.4, Example 6.4.0.0.1. 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 (1165) may be equivalently written via linear bijective (§5.6.1) EDM operator $\mathbf{D}(G)$ (721);

$$\begin{array}{cl}\underset{G\in\mathbb{S}_c^N\,,\;Y\in\mathbb{S}_h^N}{\text{minimize}} & -\operatorname{tr}\big(V(\mathbf{D}(G)-2\,Y)V\big)\\ \text{subject to} & \left[\begin{array}{cc} \langle\Phi_{ij}\,,\,G\rangle & y_{ij}\\ y_{ij} & h_{ij}^2\end{array}\right]\succeq 0\,, \quad j>i=1\ldots N-1\\ & G\succeq 0\end{array} \tag{1166}$$

where distance-square $D=[d_{ij}]\in\mathbb{S}_h^N$ (705) is related to Gram matrix entries $G=[g_{ij}]\in\mathbb{S}_c^N\cap\mathbb{S}_+^N$ by

$$\begin{aligned} d_{ij} &= g_{ii}+g_{jj}-2g_{ij}\\ &= \langle\Phi_{ij}\,,\,G\rangle\end{aligned} \tag{720}$$

where

$$\Phi_{ij}=(e_i-e_j)(e_i-e_j)^T\in\mathbb{S}_+^N \tag{707}$$

Confinement of $G$ to the geometric center subspace provides numerical stability and no loss of generality (*confer* (993)); implicit constraint $G\mathbf{1}=\mathbf{0}$ is otherwise unnecessary.

To include constraints on the list $X\in\mathbb{R}^{n\times N}$, we would first rewrite (1166)

$$\begin{array}{cl}\underset{G\in\mathbb{S}_c^N\,,\;Y\in\mathbb{S}_h^N\,,\;X\in\mathbb{R}^{n\times N}}{\text{minimize}} & -\operatorname{tr}(V(\mathbf{D}(G)-2Y)V)\\ \text{subject to} & \begin{bmatrix}\langle\Phi_{ij}\,,\,G\rangle & y_{ij}\\ y_{ij} & h_{ij}^2\end{bmatrix}\succeq 0\,,\quad j>i=1\ldots N-1\\ & \begin{bmatrix}I & X\\ X^T & G\end{bmatrix}\succeq 0\\ & X\in\mathcal{C}\end{array}\tag{1167}$$

and then add the constraints, realized here in abstract membership to some convex set $\mathcal{C}$. This problem realization includes a convex relaxation of the nonconvex constraint $G=X^TX$ and, if desired, more constraints on $G$ could be added. This technique is discussed in §5.4.2.2.4.

## 7.2.2 Minimization of affine dimension in Problem 2

When desired affine dimension $\rho$ is diminished, the rank function becomes reinserted into problem (1161) that is then rendered difficult to solve because the feasible set $\{D\,,\,Y\}$ loses convexity in $\mathbb{S}_h^N\times\mathbb{R}^{N\times N}$. Indeed, the rank function is quasiconcave (§3.3) on the positive semidefinite cone; (§2.9.2.6.2) *id est*, its sublevel sets are not convex.

### 7.2.2.1 Rank minimization heuristic

A remedy developed in [91] [192] [92] [90] introduces convex envelope (cenv) of the quasiconcave rank function: (Figure **111**)

**7.2.2.1.1 Definition.** *Convex envelope.* [147]
The convex envelope of a function $f:\mathcal{C}\to\mathbb{R}$ is defined as the largest convex function $g$ such that $g\leq f$ on convex domain $\mathcal{C}\subseteq\mathbb{R}^n$. [7.14] △

[7.14] Provided $f\not\equiv+\infty$ and there exists an affine function $h\leq f$ on $\mathbb{R}^n$, then the convex envelope is equal to the convex conjugate (the *Legendre-Fenchel transform*) of the convex conjugate of $f$; *id est*, the conjugate-conjugate function $f^{**}$. [148, §E.1]

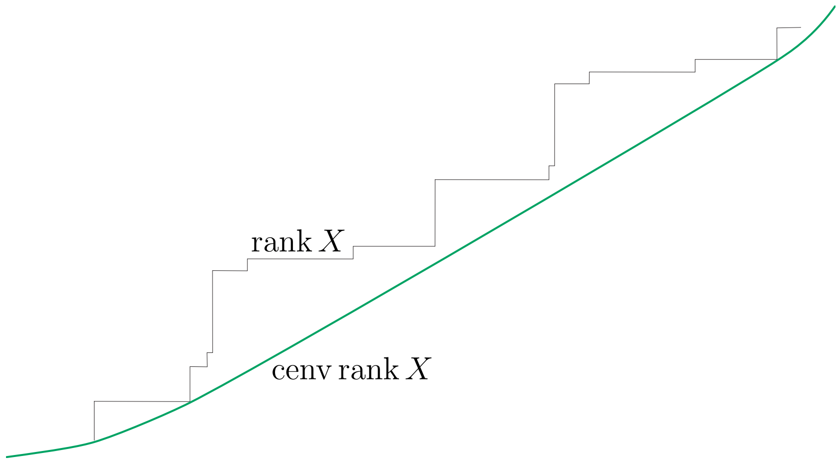



Figure 111: Abstraction of convex envelope of rank function. Rank is a quasiconcave function on the positive semidefinite cone, but its convex envelope is the smallest convex function enveloping it.

- [91] [90] Convex envelope of rank function: for $\sigma$ a singular value,

$$\operatorname{cenv}(\operatorname{rank} A) \quad \text{on} \quad \{A \in \mathbb{R}^{m\times n} \mid \|A\|_2 \leq \kappa\} \;=\; \frac{1}{\kappa}\sum_i \sigma(A)_i \tag{1168}$$

$$\operatorname{cenv}(\operatorname{rank} A) \quad \text{on} \quad \{A \in \mathbb{S}^n_+ \mid \|A\|_2 \leq \kappa\} \;=\; \frac{1}{\kappa}\operatorname{tr}(A) \tag{1169}$$

A properly scaled trace thus represents the best convex lower bound on rank for positive semidefinite matrices. The idea, then, is to substitute the convex envelope for rank of some variable $A \in \mathbb{S}^M_+$ (§A.6.5)

$$\operatorname{rank} A \;\leftarrow\; \operatorname{cenv}(\operatorname{rank} A) \;\;\propto\;\; \operatorname{tr} A \;=\; \sum_i \sigma(A)_i \;=\; \sum_i \lambda(A)_i \;=\; \|\lambda(A)\|_1 \tag{1170}$$

which is equivalent to the sum of all eigenvalues or singular values.

- [90] Convex envelope of the cardinality function is proportional to the 1-norm:

$$\operatorname{cenv}(\operatorname{card} x) \quad \text{on} \quad \{x \in \mathbb{R}^n \mid \|x\|_\infty \leq \kappa\} \;=\; \frac{1}{\kappa}\|x\|_1 \tag{1171}$$

#### 7.2.2.2 Applying trace rank-heuristic to Problem 2

Substituting rank envelope for rank function in Problem 2, for $D \in \mathbb{EDM}^N$ (*confer* (851))

$$\operatorname{cenv}\operatorname{rank}(-V_{\mathcal{N}}^T D V_{\mathcal{N}}) = \operatorname{cenv}\operatorname{rank}(-VDV) \;\propto\; -\operatorname{tr}(VDV) \tag{1172}$$

and for desired affine dimension $\rho \le N-1$ and nonnegative $H$ [*sic*] we get a convex optimization problem

$$\begin{array}{cl} \underset{D}{\text{minimize}} & \|\sqrt[\circ]{D} - H\|_{\mathrm{F}}^2 \\ \text{subject to} & -\operatorname{tr}(VDV) \le \kappa\,\rho \\ & D \in \mathbb{EDM}^N \end{array} \tag{1173}$$

where $\kappa \in \mathbb{R}_+$ is a constant determined by cut-and-try. The equivalent semidefinite program makes $\kappa$ variable: for nonnegative and symmetric $H$

$$\begin{array}{cl} \underset{D\,,\,Y\,,\,\kappa}{\text{minimize}} & \kappa\,\rho + 2\operatorname{tr}(VYV) \\ \text{subject to} & \begin{bmatrix} d_{ij} & y_{ij} \\ y_{ij} & h_{ij}^2 \end{bmatrix} \succeq 0\,, \quad j > i = 1 \ldots N-1 \\ & -\operatorname{tr}(VDV) \le \kappa\,\rho \\ & Y \in \mathbb{S}_h^N \\ & D \in \mathbb{EDM}^N \end{array} \tag{1174}$$

which is the same as (1165), 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 (1165) and it equivalents (all the way back to (1158)): When the given measurement matrix $H$ is nonnegative and symmetric, finding the closest EDM $D$ as in problem (1158), (1161), or (1165) 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 (1174), in other words, and their optimal solutions are unique because of the strictly convex objective function in (1158).

#### 7.2.2.3 Rank-heuristic insight

Minimization of affine dimension by use of this trace rank-heuristic (1172) tends to find the list configuration of least energy; rather, it tends to optimize compaction of the reconstruction by minimizing total distance. (734) It is best used where some physical equilibrium implies such an energy minimization; *e.g.*, [266, §5].

For this Problem 2, the trace rank-heuristic arose naturally in the objective in terms of $V$. We observe: $V$ (in contrast to $V_{\mathcal{N}}^T$) spreads energy over all available distances (§B.4.2 no.20, *confer* no.22) although the rank function itself is insensitive to choice of auxiliary matrix.

#### 7.2.2.4 Rank minimization heuristic beyond convex envelope

Fazel, Hindi, and Boyd [92] [295] [93] propose a rank heuristic more potent than trace (1170) for problems of rank minimization;

$$\operatorname{rank} Y \;\leftarrow\; \log\det(Y+\varepsilon I) \tag{1175}$$

the concave surrogate function $\log\det$ in place of quasiconcave $\operatorname{rank} Y$ (§2.9.2.6.2) 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 $Y_i$ ; *id est*, from the first-order Taylor series expansion about $Y_i$ on some open interval of $\|Y\|$ (§D.1.7)

$$\log\det(Y+\varepsilon I) \approx \log\det(Y_i+\varepsilon I) + \operatorname{tr}\big((Y_i+\varepsilon I)^{-1}(Y-Y_i)\big) \tag{1176}$$

we make the surrogate sequence of infima over bounded convex feasible set $\mathcal{C}$

$$\arg\inf_{Y\in\mathcal{C}} \operatorname{rank} Y \;\leftarrow\; \lim_{i\to\infty} Y_{i+1} \tag{1177}$$

where, for $i=0\ldots$

$$Y_{i+1} = \arg\inf_{Y\in\mathcal{C}} \operatorname{tr}\big((Y_i+\varepsilon I)^{-1}Y\big) \tag{1178}$$

Choosing $Y_0=I$, the first step becomes equivalent to finding the infimum of $\operatorname{tr} Y$; the trace rank-heuristic (1170). The intuition underlying (1178) 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

$$\begin{aligned} Y_i + \varepsilon I &\triangleq Q(\Lambda + \varepsilon I)Q^T & \text{(a)} \\ Y &\triangleq U\Upsilon U^T & \text{(b)} \end{aligned} \tag{1179}$$

into (1178). Then from (1481) we have,

$$\begin{aligned} \inf_{\Upsilon \in U^{\star T}\mathcal{C}U^{\star}} \delta((\Lambda + \varepsilon I)^{-1})^T \delta(\Upsilon) &= \inf_{\Upsilon \in U^T\mathcal{C}U} \; \inf_{R^T = R^{-1}} \operatorname{tr}\big((\Lambda + \varepsilon I)^{-1} R^T \Upsilon R\big) \\ &\leq \inf_{Y \in \mathcal{C}} \operatorname{tr}((Y_i + \varepsilon I)^{-1} Y) \end{aligned} \tag{1180}$$

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 the 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 (1174) becomes the problem sequence in $i$

$$\begin{array}{cl} \underset{D\,,\,Y\,,\,\kappa}{\text{minimize}} & \kappa\,\rho + 2\operatorname{tr}(VYV) \\ \text{subject to} & \begin{bmatrix} d_{jl} & y_{jl} \\ y_{jl} & h_{jl}^2 \end{bmatrix} \succeq 0\,, \quad l > j = 1 \ldots N-1 \\ & -\operatorname{tr}\big((-VD_iV + \varepsilon I)^{-1} VDV\big) \leq \kappa\,\rho \\ & Y \in \mathbb{S}_h^N \\ & D \in \mathbb{EDM}^N \end{array} \tag{1181}$$

where $D_{i+1} \triangleq D^\star \in \mathbb{EDM}^N$, and $D_0 \triangleq \mathbf{1}\mathbf{1}^T - I$.

#### 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 the eigenvalues of the sum are simply determined, $\lambda(Y_i + \varepsilon I) = \delta(\Lambda + \varepsilon I)$, we may certainly force selected weights to $\varepsilon^{-1}$ by manipulating diagonalization (1179a). 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 the *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 [210] based on alternating projection (§E.10) that does not solve the *ball packing* problem presented in §5.4.2.2.3, so it is not evaluated further herein. None of these exceed performance of the *convex iteration* method for constraining rank developed in §4.4:

**7.2.2.7.1 Example.** *Rank regularization enforcing affine dimension.*
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.

*Unidimensional scaling*, [70] 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 finding a list constrained to lie in one affine dimension. In terms of point list, the nonconvex problem is: given nonnegative symmetric matrix $H=[h_{ij}]\in\mathbb{S}^N\cap\mathbb{R}_+^{N\times N}$ (1163) whose entries $h_{ij}$ are all known, (1114)

$$\underset{\{x_i\in\mathbb{R}\}}{\text{minimize}}\ \sum_{i\,,\,j=1}^{N}(\|x_i-x_j\|-h_{ij})^2 \tag{1182}$$

called a *raw stress* problem [39, 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.*, [52].

As always, we first transform variables to distance-square $D\in\mathbb{S}_h^N$; so we begin with convex problem (1165) on page 461

$$\begin{array}{cl}\underset{D\,,\,Y}{\text{minimize}} & -\operatorname{tr}\big(V(D-2\,Y\,)V\big)\\ \text{subject to} & \left[\begin{array}{cc}d_{ij} & y_{ij}\\ y_{ij} & h_{ij}^2\end{array}\right]\succeq 0\,,\quad j>i=1\ldots N-1\\ & Y\in\mathbb{S}_h^N\\ & D\in\mathbb{EDM}^N\\ & \operatorname{rank}V_{\mathcal{N}}^TDV_{\mathcal{N}}=1\end{array} \tag{1183}$$

that becomes equivalent to (1182) by making explicit the constraint on affine dimension via rank. The iteration is formed by moving the dimensional constraint to the objective:

$$\begin{array}{cl}\underset{D\,,\,Y}{\text{minimize}} & -\langle V(D-2\,Y)V\;,\;I\rangle - w\langle V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\;,\;W\rangle\\ \text{subject to} & \left[\begin{array}{cc} d_{ij} & y_{ij}\\ y_{ij} & h_{ij}^2\end{array}\right]\succeq 0\;,\quad j>i=1\ldots N-1\\ & Y\in\mathbb{S}_h^N\\ & D\in\mathbb{EDM}^N\end{array}\tag{1184}$$

where $w$ ($\approx 10$) is a positive scalar just large enough to make $\langle V_{\mathcal{N}}^{T}DV_{\mathcal{N}}\;,\;W\rangle$ vanish to within some numerical precision, and where direction matrix $W$ is an optimal solution to semidefinite program (1480a)

$$\begin{array}{cl}\underset{W}{\text{minimize}} & -\langle V_{\mathcal{N}}^{T}D^{\star}V_{\mathcal{N}}\;,\;W\rangle\\ \text{subject to} & 0\preceq W\preceq I\\ & \operatorname{tr}W=N-1\end{array}\tag{1185}$$

known in closed form. Semidefinite programs (1184) and (1185) are iterated until convergence in the sense defined on page 257. This iteration is not a projection method. Convex problem (1184) is neither a relaxation of unidimensional scaling problem (1183); instead, problem (1184) is a convex equivalent to (1183) at convergence of the iteration.

Jan de Leeuw provided us with some test data

$$H=\left[\begin{array}{cccccc} 0.000000 & 5.235301 & 5.499274 & 6.404294 & 6.486829 & 6.263265\\ 5.235301 & 0.000000 & 3.208028 & 5.840931 & 3.559010 & 5.353489\\ 5.499274 & 3.208028 & 0.000000 & 5.679550 & 4.020339 & 5.239842\\ 6.404294 & 5.840931 & 5.679550 & 0.000000 & 4.862884 & 4.543120\\ 6.486829 & 3.559010 & 4.020339 & 4.862884 & 0.000000 & 4.618718\\ 6.263265 & 5.353489 & 5.239842 & 4.543120 & 4.618718 & 0.000000 \end{array}\right]\tag{1186}$$

and a globally optimal solution

$$\begin{aligned} X^{\star} &= [\,-4.981494 \quad -2.121026 \quad -1.038738 \quad 4.555130 \quad 0.764096 \quad 2.822032\,]\\ &= [\quad x_1^{\star} \qquad x_2^{\star} \qquad x_3^{\star} \qquad x_4^{\star} \qquad x_5^{\star} \qquad x_6^{\star}\quad ] \end{aligned}\tag{1187}$$

found by searching $6!$ local minima of (1182) [70]. By iterating convex problems (1184) and (1185) about twenty times (initial $W = \mathbf{0}$) we find the global infimum 98.12812 of stress problem (1182), and by (939) we find a corresponding one-dimensional point list that is a rigid transformation in $\mathbb{R}$ of $X^\star$.

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. $\square$

## 7.3 Third prevalent problem: Projection on EDM cone in $d_{ij}$

Reformulating Problem 2 (p.458) in terms of EDM $D$ changes the problem considerably:

$$\left.\begin{array}{cl}\underset{D}{\text{minimize}} & \|D-H\|_{\mathrm{F}}^2\\ \text{subject to} & \operatorname{rank} V_{\mathcal{N}}^{T} D V_{\mathcal{N}} \leq \rho\\ & D \in \mathbb{EDM}^N\end{array}\right\}\quad \text{Problem 3} \tag{1188}$$

This third prevalent proximity problem is a Euclidean projection of given matrix $H$ on a generally nonconvex subset ($\rho < N-1$) of $\partial\mathbb{EDM}^N$ the boundary of the convex cone of Euclidean distance matrices relative to subspace $\mathbb{S}_h^N$ (Figure 95(d)). Because coordinates of projection are distance-square and $H$ presumably now 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

$$\begin{array}{cl}\underset{D}{\text{minimize}} & \|D-H\|_{\mathrm{F}}^2\\ \text{subject to} & D \in \mathbb{EDM}^N\end{array} \tag{1189}$$

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

$$\|D-H\|_{\mathrm{F}}^2 \;=\; \sum_{i,j}(d_{ij}-h_{ij})^2 \tag{1190}$$

is a strictly convex quadratic in $D$ ;[7.15]

$$\begin{array}{cl}\underset{D}{\text{minimize}} & \sum_{i,j} d_{ij}^2 - 2h_{ij}\,d_{ij} \;+\; h_{ij}^2\\ \text{subject to} & D\in\mathbb{EDM}^N\end{array} \tag{1191}$$

Optimal solution $D^\star$ is therefore unique, as expected, for this simple projection on the EDM cone.

#### 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) [106] [99] [134, §1] We translate (1191) to an equivalent semidefinite program because we have a good solver:

Assume the given measurement matrix $H$ to be nonnegative and symmetric;[7.16]

$$H=[h_{ij}]\in\mathbb{S}^N\cap\mathbb{R}_+^{N\times N} \qquad (1163)$$

We then propose: Problem (1191) is equivalent to the semidefinite program, for

$$\partial \triangleq [d_{ij}^2] = D\circ D \tag{1192}$$

[7.15] For nonzero $Y\in\mathbb{S}_h^N$ and some open interval of $t\in\mathbb{R}$ (§3.2.3.0.2, §D.2.3)

$$\frac{d^2}{dt^2}\|(D+t\,Y)-H\|_{\mathrm{F}}^2 \;=\; 2\operatorname{tr}Y^TY>0 \qquad \blacklozenge$$

[7.16] If that $H$ given has negative entries, then the technique of solution presented here becomes invalid. Projection of $H$ on $\mathcal{K}$ (1107) prior to application of this proposed technique, as explained in §7.0.1, is incorrect.

a matrix of distance-square squared,

$$\begin{array}{cl}\underset{\partial\,,\,D}{\text{minimize}} & -\operatorname{tr}\big(V(\partial-2\,H\circ D)V\big)\\ \text{subject to} & \begin{bmatrix}\partial_{ij} & d_{ij}\\ d_{ij} & 1\end{bmatrix}\succeq 0\,,\quad j>i=1\ldots N-1\\ & D\in\mathbb{EDM}^N\\ & \partial\in\mathbb{S}_h^N\end{array}\tag{1193}$$

where

$$\begin{bmatrix}\partial_{ij} & d_{ij}\\ d_{ij} & 1\end{bmatrix}\succeq 0 \;\Leftrightarrow\; \partial_{ij}\geq d_{ij}^2\tag{1194}$$

Symmetry of input $H$ facilitates trace in the objective (§B.4.2 no.20), while its nonnegativity causes $\partial_{ij}\rightarrow d_{ij}^2$ as optimization proceeds.

**7.3.1.1.1 Example.** *Alternating projection on nearest EDM.*
By solving (1193) we confirm the result from an example given by Glunt, Hayden, *et alii* [106, §6] who found an analytical solution to convex optimization problem (1189) for particular cardinality $N=3$ by using the alternating projection method of von Neumann (§E.10):

$$H=\begin{bmatrix}0&1&1\\1&0&9\\1&9&0\end{bmatrix}\,,\qquad D^\star=\begin{bmatrix}0&\frac{19}{9}&\frac{19}{9}\\ \frac{19}{9}&0&\frac{76}{9}\\ \frac{19}{9}&\frac{76}{9}&0\end{bmatrix}\tag{1195}$$

The original problem (1189) of projecting $H$ on the EDM cone is transformed to an equivalent iterative sequence of projections on the two convex cones (1056) from §6.8.1.1. Using ordinary alternating projection, input $H$ goes to $D^\star$ with an accuracy of four decimal places in about 17 iterations. Affine dimension corresponding to this optimal solution is $r=1$.

Obviation 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 (1193) can be reformulated by moving the objective function in

$$\begin{array}{cl}\underset{D}{\text{minimize}} & \|D-H\|_{\mathrm F}^2\\ \text{subject to} & D\in\mathbb{EDM}^N\end{array}\tag{1189}$$

to the constraints. This makes an equivalent second-order cone program: for any measurement matrix $H$

$$\begin{array}{cl}\underset{t\in\mathbb{R}\;,\;D}{\text{minimize}} & t\\ \text{subject to} & \|D-H\|_{\mathrm{F}}^2\leq t\\ & D\in\mathbb{EDM}^N\end{array}\tag{1196}$$

We can transform this problem to an equivalent Schur-form semidefinite program; (§3.1.7.2)

$$\begin{array}{cl}\underset{t\in\mathbb{R}\;,\;D}{\text{minimize}} & t\\ \text{subject to} & \left[\begin{array}{cc}tI & \operatorname{vec}(D-H)\\ \operatorname{vec}(D-H)^T & 1\end{array}\right]\succeq 0\\ & D\in\mathbb{EDM}^N\end{array}\tag{1197}$$

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 (1193). (§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 $\mathbf{D}(G)$ (721);

$$\begin{array}{cl}\underset{G\in\mathbb{S}_c^N\;,\;t\in\mathbb{R}}{\text{minimize}} & t\\ \text{subject to} & \left[\begin{array}{cc}tI & \operatorname{vec}(\mathbf{D}(G)-H)\\ \operatorname{vec}(\mathbf{D}(G)-H)^T & 1\end{array}\right]\succeq 0\\ & G\succeq 0\end{array}\tag{1198}$$

To include constraints on the list $X\in\mathbb{R}^{n\times N}$, we would rewrite this:

$$\begin{array}{cl}\underset{G\in\mathbb{S}_c^N\;,\;t\in\mathbb{R}\;,\;X\in\mathbb{R}^{n\times N}}{\text{minimize}} & t\\ \text{subject to} & \left[\begin{array}{cc}tI & \operatorname{vec}(\mathbf{D}(G)-H)\\ \operatorname{vec}(\mathbf{D}(G)-H)^T & 1\end{array}\right]\succeq 0\\ & \left[\begin{array}{cc}I & X\\ X^T & G\end{array}\right]\succeq 0\\ & X\in\mathcal{C}\end{array}\tag{1199}$$

where $\mathcal{C}$ is some abstract convex set. This technique is discussed in §5.4.2.2.4.

#### 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 $\mathcal{K}$ is a convex cone and point $x$ exists exterior to the cone or on its boundary, distance to the nearest point $Px$ in $\mathcal{K}$ is found as the optimal value of the objective

$$\begin{array}{rll} \|x - Px\| = \underset{a}{\text{maximize}} & a^T x & \\ \text{subject to} & \|a\| \leq 1 & \quad (1784) \\ & a \in \mathcal{K}^\circ & \end{array}$$

where $\mathcal{K}^\circ$ is the polar cone.

Applying this result to (1189), we get a convex optimization for any given symmetric matrix $H$ exterior to or on the EDM cone boundary:

$$\begin{array}{ll} \underset{D}{\text{minimize}} & \|D - H\|_F^2 \\ \text{subject to} & D \in \mathbb{EDM}^N \end{array} \equiv \begin{array}{ll} \underset{A^\circ}{\text{maximize}} & \langle A^\circ,\, H\rangle \\ \text{subject to} & \|A^\circ\|_F \leq 1 \\ & A^\circ \in \mathbb{EDM}^{N^\circ} \end{array} \tag{1200}$$

Then from (1786) projection of $H$ on cone $\mathbb{EDM}^N$ is

$$D^\star = H - A^{\circ\star}\langle A^{\circ\star},\, H\rangle \tag{1201}$$

Critchley proposed, instead, projection on the polar EDM cone in his 1980 thesis [61, p.113]: In that circumstance, by projection on the algebraic complement (§E.9.2.2.1),

$$D^\star = A^\star\langle A^\star,\, H\rangle \tag{1202}$$

which is equal to (1201) when $A^\star$ solves

$$\begin{array}{ll} \underset{A}{\text{maximize}} & \langle A\,,\, H\rangle \\ \text{subject to} & \|A\|_F = 1 \\ & A \in \mathbb{EDM}^N \end{array} \tag{1203}$$

This projection of symmetric $H$ on polar cone $\mathbb{EDM}^{N^\circ}$ 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 the desired affine dimension $\rho$ is diminished, Problem 3 (1188) is difficult to solve [134, §3] because the feasible set in $\mathbb{R}^{N(N-1)/2}$ loses convexity. By substituting rank envelope (1172) into Problem 3, then for any given $H$ we get a convex problem

$$\begin{array}{cl}\underset{D}{\text{minimize}} & \|D-H\|_{\mathrm{F}}^2\\ \text{subject to} & -\operatorname{tr}(VDV)\leq\kappa\,\rho\\ & D\in\mathbb{EDM}^N\end{array}\tag{1204}$$

where $\kappa\in\mathbb{R}_+$ is a constant determined by cut-and-try. Given $\kappa$ , problem (1204) 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 (1204) does not move $\kappa$ into the variables as on page 464: for nonnegative symmetric input $H$ and distance-square squared variable $\partial$ as in (1192),

$$\begin{array}{cl}\underset{\partial\,,\,D}{\text{minimize}} & -\operatorname{tr}\big(V(\partial-2\,H\circ D)V\big)\\ \text{subject to} & \begin{bmatrix}\partial_{ij} & d_{ij}\\ d_{ij} & 1\end{bmatrix}\succeq 0\,,\quad j>i=1\ldots N-1\\ & -\operatorname{tr}(VDV)\leq\kappa\,\rho\\ & D\in\mathbb{EDM}^N\\ & \partial\in\mathbb{S}_h^N\end{array}\tag{1205}$$

That means we will not see equivalence of this cenv(rank)-minimization problem to the non−rank-constrained problems (1191) and (1193) like we saw for its counterpart (1174) 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 cenv(rank)-minimization as in (1205), 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 $-VGV$ or a variant to reduce affine dimension, because that particular method comes from projection on a positive semidefinite cone (1142); zeroing those eigenvalues here in Problem 3 would place an elbow in the projection path (*confer* Figure **110**) 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:

#### 7.3.3.1 Cayley-Menger form

We use Cayley-Menger composition of the Euclidean distance matrix to solve a problem that is the same as Problem 3 (1188): (§5.7.3.0.1)

$$\begin{array}{cl}\underset{D}{\text{minimize}} & \left\| \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix} - \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -H \end{bmatrix} \right\|_{\mathrm{F}}^2 \\ \text{subject to} & \operatorname{rank} \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D \end{bmatrix} \leq \rho + 2 \\ & D \in \mathbb{EDM}^N \end{array} \tag{1206}$$

a projection of $H$ on a generally nonconvex subset (when $\rho < N-1$) of the Euclidean distance matrix cone boundary $\operatorname{rel}\partial\mathbb{EDM}^N$; *id est*, projection from the EDM cone interior or exterior on a subset of its relative boundary (§6.6, (986)).

Rank of an optimal solution is intrinsically bounded above and below;

$$2 \leq \operatorname{rank} \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D^\star \end{bmatrix} \leq \rho + 2 \leq N+1 \tag{1207}$$

Our proposed strategy for low-rank solution is projection on that subset of a spectral cone $\lambda\left(\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -\mathbb{EDM}^N \end{bmatrix}\right)$ (§5.11.2.3) corresponding to affine dimension not in excess of that $\rho$ desired; *id est*, spectral projection on

$$\left[\begin{array}{l}\mathbb{R}_+^{\rho+1}\\ \mathbf{0}\\ \mathbb{R}_-\end{array}\right] \cap\, \partial\mathcal{H} \subset \mathbb{R}^{N+1} \tag{1208}$$

where

$$\partial\mathcal{H} = \{\lambda\in\mathbb{R}^{N+1} \mid \mathbf{1}^T\lambda = 0\} \qquad (919)$$

is a hyperplane through the origin. This pointed polyhedral cone (1208), to which membership subsumes the rank constraint, has empty interior.

Given desired affine dimension $0\leq\rho\leq N-1$ and diagonalization (§A.5) of unknown EDM $D$

$$\left[\begin{array}{cc}0 & \mathbf{1}^T\\ \mathbf{1} & -D\end{array}\right] \triangleq U\Upsilon U^T\in\mathbb{S}_h^{N+1} \tag{1209}$$

and given symmetric $H$ in diagonalization

$$\left[\begin{array}{cc}0 & \mathbf{1}^T\\ \mathbf{1} & -H\end{array}\right] \triangleq Q\Lambda Q^T\in\mathbb{S}^{N+1} \tag{1210}$$

having eigenvalues arranged in nonincreasing order, then by (932) problem (1206) is equivalent to

$$\begin{array}{cl}\underset{\Upsilon\,,\,R}{\text{minimize}} & \left\|\delta(\Upsilon) - \pi\big(\delta(R^T\Lambda R)\big)\right\|^2\\ \text{subject to} & \delta(\Upsilon) \in \left[\begin{array}{l}\mathbb{R}_+^{\rho+1}\\ \mathbf{0}\\ \mathbb{R}_-\end{array}\right] \cap\, \partial\mathcal{H}\\ & \delta(QR\Upsilon R^TQ^T) = \mathbf{0}\\ & R^{-1} = R^T\end{array} \tag{1211}$$

where $\pi$ is the permutation operator from §7.1.3 arranging its vector argument in nonincreasing order,[7.17] where

$$R \triangleq Q^TU\in\mathbb{R}^{N+1\times N+1} \tag{1212}$$

in $U$ on the set of orthogonal matrices is a bijection, and where $\partial\mathcal{H}$ insures one negative eigenvalue. Hollowness constraint $\delta(QR\Upsilon R^TQ^T) = \mathbf{0}$ makes problem (1211) difficult by making the two variables dependent.

[7.17] Recall, any permutation matrix is an orthogonal matrix.

Our plan is to instead divide problem (1211) into two and then alternate their solution:

$$\begin{array}{cll}\underset{\Upsilon}{\text{minimize}} & \left\|\delta(\Upsilon) - \pi\big(\delta(R^T\Lambda R)\big)\right\|^2 & \\ \text{subject to} & \delta(\Upsilon) \in \begin{bmatrix}\mathbb{R}_+^{\rho+1} \\ \mathbf{0} \\ \mathbb{R}_-\end{bmatrix} \cap \partial\mathcal{H} & \text{(a)}\end{array}$$

$$\begin{array}{cll}\underset{R}{\text{minimize}} & \|R\Upsilon R^T - \Lambda\|_{\mathrm{F}}^2 & \\ \text{subject to} & \delta(QR\Upsilon R^TQ^T) = \mathbf{0} & \text{(b)} \\ & R^{-1} = R^T & \end{array} \tag{1213}$$

We justify disappearance of the hollowness constraint in convex optimization problem (1213a): From the arguments in §7.1.3 with regard to $\pi$ the permutation operator, cone membership constraint $\delta(\Upsilon) \in \begin{bmatrix}\mathbb{R}_+^{\rho+1} \\ \mathbf{0} \\ \mathbb{R}_-\end{bmatrix} \cap \partial\mathcal{H}$ from (1213a) is equivalent to

$$\delta(\Upsilon) \in \begin{bmatrix}\mathbb{R}_+^{\rho+1} \\ \mathbf{0} \\ \mathbb{R}_-\end{bmatrix} \cap \partial\mathcal{H} \cap \mathcal{K}_{\mathcal{M}} \tag{1214}$$

where $\mathcal{K}_{\mathcal{M}}$ is the monotone cone (§2.13.9.4.2). Membership of $\delta(\Upsilon)$ to the *polyhedral cone of majorization* (Theorem A.1.2.0.1)

$$\mathcal{K}_{\lambda\delta}^* = \partial\mathcal{H} \cap \mathcal{K}_{\mathcal{M}+}^* \tag{1229}$$

where $\mathcal{K}_{\mathcal{M}+}^*$ is the dual monotone nonnegative cone (§2.13.9.4.1), is a condition (in absence of a hollowness constraint) that would insure existence of a symmetric hollow matrix $\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & -D\end{bmatrix}$. Curiously, intersection of this feasible superset $\begin{bmatrix}\mathbb{R}_+^{\rho+1} \\ \mathbf{0} \\ \mathbb{R}_-\end{bmatrix} \cap \partial\mathcal{H} \cap \mathcal{K}_{\mathcal{M}}$ from (1214) with the cone of majorization $\mathcal{K}_{\lambda\delta}^*$ is a benign operation; *id est*,

$$\partial\mathcal{H} \cap \mathcal{K}_{\mathcal{M}+}^* \cap \mathcal{K}_{\mathcal{M}} = \partial\mathcal{H} \cap \mathcal{K}_{\mathcal{M}} \tag{1215}$$

verifiable by observing conic dependencies (§2.10.3) among the aggregate of halfspace-description normals. The cone membership constraint in (1213a) therefore inherently insures existence of a symmetric hollow matrix. ♦

Optimization (1213b) 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 (1213b) by a method introduced in §4.4.3.0.4: Define $R = [\, r_1 \cdots r_{N+1} \,] \in \mathbb{R}^{N+1\times N+1}$ and make the assignment

$$\begin{aligned} G &= \begin{bmatrix} r_1 \\ \vdots \\ r_{N+1} \\ 1 \end{bmatrix} [\, r_1^T \cdots r_{N+1}^T \;\; 1 \,] \qquad \in \mathbb{S}^{(N+1)^2+1} \\ &= \begin{bmatrix} R_{11} & \cdots & R_{1,N+1} & r_1 \\ \vdots & \ddots & & \vdots \\ R_{1,N+1}^T & & R_{N+1,N+1} & r_{N+1} \\ r_1^T & \cdots & r_{N+1}^T & 1 \end{bmatrix} \triangleq \begin{bmatrix} r_1 r_1^T & \cdots & r_1 r_{N+1}^T & r_1 \\ \vdots & \ddots & & \vdots \\ r_{N+1} r_1^T & & r_{N+1} r_{N+1}^T & r_{N+1} \\ r_1^T & \cdots & r_{N+1}^T & 1 \end{bmatrix} \end{aligned} \tag{1216}$$

where $R_{ij} \triangleq r_i r_j^T \in \mathbb{R}^{N+1\times N+1}$. Then (1213b) is equivalently expressed:

$$\begin{array}{cll} \underset{R_{ij}\,,\, r_i}{\text{minimize}} & \left\| \sum_{i=1}^{N+1} \Upsilon_{ii} R_{ii} - \Lambda \right\|_{\mathrm F}^2 & \\ \text{subject to} & \operatorname{tr} R_{ii} = 1\,, & i=1\ldots N+1 \\ & \operatorname{tr} R_{ij} = 0\,, & i<j = 2\ldots N+1 \\ & G = \begin{bmatrix} R_{11} & \cdots & R_{1,N+1} & r_1 \\ \vdots & \ddots & & \vdots \\ R_{1,N+1}^T & & R_{N+1,N+1} & r_{N+1} \\ r_1^T & \cdots & r_{N+1}^T & 1 \end{bmatrix} (\succeq 0) & \\ & \delta\Big( Q \sum_{i=1}^{N+1} \Upsilon_{ii} R_{ii}\, Q^T \Big) = \mathbf{0} & \\ & \operatorname{rank} G = 1 & \end{array} \tag{1217}$$

The rank constraint is regularized by method of convex iteration developed in §4.4. Problem (1217) is partitioned into two convex problems:

$$\begin{array}{cll}\underset{R_{ij}\,,\,r_i}{\text{minimize}} & \left\|\sum_{i=1}^{N+1}\Upsilon_{ii}\,R_{ii}-\Lambda\right\|_{\mathrm{F}}^2+\;\langle G\,,\,W\rangle & \\ \text{subject to} & \operatorname{tr}R_{ii}=1\,, & i=1\ldots N+1 \\ & \operatorname{tr}R_{ij}=0\,, & i<j=2\ldots N+1 \\ & G=\begin{bmatrix} R_{11} & \cdots & R_{1,N+1} & r_1 \\ \vdots & \ddots & & \vdots \\ R_{1,N+1}^T & & R_{N+1,N+1} & r_{N+1} \\ r_1^T & \cdots & r_{N+1}^T & 1\end{bmatrix}\succeq 0 & \\ & \delta\!\left(Q\sum_{i=1}^{N+1}\Upsilon_{ii}\,R_{ii}\;Q^T\right)=\mathbf{0} & \end{array}\tag{1218}$$

and

$$\begin{array}{cl}\underset{W\in\,\mathbb{S}^{(N+1)^2+1}}{\text{minimize}} & \langle G^\star,\,W\rangle \\ \text{subject to} & 0\preceq W\preceq I \\ & \operatorname{tr}W=(N+1)^2\end{array}\tag{1219}$$

then iterated until a rank-1 $G$ matrix is found.

## 7.4 Conclusion

There has been little progress in spectral projection since the discovery by Eckart & Young in 1936 [85] of a formula for projection on a rank $\rho$ subset of the positive semidefinite cone (§2.9.2.1). 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 recourse is intentional misapplication of Eckart & Young's result by introducing spectral projection on a positive semidefinite cone into Problem 3 via $\mathbf{D}(G)$ (721), for example. [52] 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; *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 265. That redesign (*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 just like 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.

The importance and application of solving rank-constrained problems are enormous, a conclusion generally accepted *gratis* by the mathematics and engineering communities. Rank-constrained semidefinite programs arise in many vital feedback and control problems [122], optics [54], and communications [209] [186]. Rank-constrained problems also appear naturally in *combinatorial optimization.* (§4.4.3.0.7)

# Appendix A

# Linear algebra

## A.1 Main-diagonal $\delta$ operator, $\lambda$ , trace, vec

We introduce notation $\delta$ denoting the main-diagonal linear self-adjoint 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;

$$\delta(A) \in \mathbb{R}^N \tag{1220}$$

Operating on a vector $y\in\mathbb{R}^N$, $\delta$ naturally returns a diagonal matrix;

$$\delta(y) \in \mathbb{S}^N \tag{1221}$$

Operating recursively on a vector $\Lambda\in\mathbb{R}^N$ or diagonal matrix $\Lambda\in\mathbb{S}^N$, $\delta\big(\delta(\Lambda)\big)$ returns $\Lambda$ itself;

$$\delta^2(\Lambda) \equiv \delta\big(\delta(\Lambda)\big) \stackrel{\Delta}{=} \Lambda \tag{1222}$$

Defined in this manner, main-diagonal linear operator $\delta$ is *self-adjoint* [166, §3.10, §9.5-1];[A.1] *videlicet*, (§2.2)

$$\delta(A)^T y \;=\; \langle \delta(A)\,,\, y\rangle \;=\; \langle A \;,\; \delta(y)\rangle \;=\; \operatorname{tr}\big(A^T\delta(y)\big) \tag{1223}$$

---

[A.1]Linear operator $T:\mathbb{R}^{m\times n}\to\mathbb{R}^{M\times N}$ is self-adjoint when, for each and every $X_1\,,\,X_2\in\mathbb{R}^{m\times n}$

$$\langle T(X_1)\,,\,X_2\rangle = \langle X_1\,,\,T(X_2)\rangle$$

 
Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

### A.1.1 Identities

This $\delta$ notation is efficient and unambiguous as illustrated in the following examples where $A \circ B$ denotes Hadamard product [150] [110, §1.1.4] of matrices of like size, $\otimes$ the Kronecker product (§D.1.2.1), $y$ a vector, $X$ a matrix, $e_i$ the $i^{\text{th}}$ member of the standard basis for $\mathbb{R}^n$, $\mathbb{S}_h^N$ the symmetric hollow subspace, $\sigma$ a vector of (nonincreasingly) ordered singular values, and $\lambda(A)$ denotes a vector of nonincreasingly ordered eigenvalues of matrix $A$:

1. $\delta(A) = \delta(A^T)$
2. $\operatorname{tr}(A) = \operatorname{tr}(A^T) = \delta(A)^T \mathbf{1}$
3. $\langle I\,, A\rangle = \operatorname{tr} A$
4. $\delta(cA) = c\,\delta(A)\,,\quad c \in \mathbb{R}$
5. $\operatorname{tr}(c\sqrt{A^TA}\,) = c \operatorname{tr}\sqrt{A^TA} = c\,\mathbf{1}^T\sigma(A)\,,\quad c \in \mathbb{R}$
6. $\operatorname{tr}(cA) = c \operatorname{tr}(A) = c\,\mathbf{1}^T\lambda(A)\,,\quad c \in \mathbb{R}$
7. $\delta(A+B) = \delta(A) + \delta(B)$
8. $\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$
9. $\delta(AB) = (A \circ B^T)\mathbf{1} = (B^T \circ A)\mathbf{1}$
10. $\delta(AB)^T = \mathbf{1}^T(A^T \circ B) = \mathbf{1}^T(B \circ A^T)$
11. $\delta(uv^T) = \begin{bmatrix} u_1 v_1 \\ \vdots \\ u_N v_N \end{bmatrix} = u \circ v\,,\quad u\,,v \in \mathbb{R}^N$
12. $\operatorname{tr}(A^TB) = \operatorname{tr}(AB^T) = \operatorname{tr}(BA^T) = \operatorname{tr}(B^TA)$
    $= \mathbf{1}^T(A \circ B)\mathbf{1} = \mathbf{1}^T\delta(AB^T) = \delta(A^TB)^T\mathbf{1} = \delta(BA^T)^T\mathbf{1} = \delta(B^TA)^T\mathbf{1}$
13. $D = [d_{ij}] \in \mathbb{S}_h^N\,,\quad H = [h_{ij}] \in \mathbb{S}_h^N\,,\quad V = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{S}^N$ (*confer* §B.4.2 no.20)
    $N \operatorname{tr}(-V(D \circ H)V) = \operatorname{tr}(D^TH) = \mathbf{1}^T(D \circ H)\mathbf{1} = \operatorname{tr}\big(\mathbf{1}\mathbf{1}^T(D \circ H)\big) = \sum\limits_{i,j} d_{ij}\,h_{ij}$
14. $\operatorname{tr}(\Lambda A) = \delta(\Lambda)^T\delta(A)\,,\quad \delta^2(\Lambda) \stackrel{\Delta}{=} \Lambda \in \mathbb{S}^N$

15. $y^T B\,\delta(A) = \operatorname{tr}\big(B\,\delta(A)y^T\big) \;= \operatorname{tr}\big(\delta(B^T y)A\big) \;= \operatorname{tr}\big(A\,\delta(B^T y)\big)$
$= \delta(A)^T B^T y = \operatorname{tr}\big(y\,\delta(A)^T B^T\big) = \operatorname{tr}\big(A^T \delta(B^T y)\big) = \operatorname{tr}\big(\delta(B^T y)A^T\big)$

16. $\delta^2(A^T A) = \sum\limits_i e_i e_i^T A^T A e_i e_i^T$

17. $\delta\big(\delta(A)\mathbf{1}^T\big) \;=\; \delta\big(\mathbf{1}\,\delta(A)^T\big) \;=\; \delta(A)$

18. $\delta(A\mathbf{1})\mathbf{1} \;=\; \delta(A\mathbf{1}\mathbf{1}^T) \;=\; A\mathbf{1}\;, \qquad \delta(y)\mathbf{1} \;=\; \delta(y\mathbf{1}^T) \;=\; y$

19. $\delta(I\mathbf{1}) \;=\; \delta(\mathbf{1}) \;=\; I$

20. $\delta(e_i e_j^T \mathbf{1}) \;=\; \delta(e_i) \;=\; e_i e_i^T$

21. $\operatorname{vec}(AXB) \;=\; (B^T \otimes A)\operatorname{vec} X$

22. $\operatorname{vec}(BXA) \;=\; (A^T \otimes B)\operatorname{vec} X$

23. $\operatorname{tr}(AXBX^T) \;=\; \operatorname{vec}(X)^T \operatorname{vec}(AXB) \;=\; \operatorname{vec}(X)^T (B^T \otimes A)\operatorname{vec} X$ [116]

24. $\begin{aligned}\operatorname{tr}(AX^T BX) \;&=\; \operatorname{vec}(X)^T \operatorname{vec}(BXA) \;=\; \operatorname{vec}(X)^T (A^T \otimes B)\operatorname{vec} X \\ &=\; \delta\big(\operatorname{vec}(X)\operatorname{vec}(X)^T (A^T \otimes B)\big)^T \mathbf{1}\end{aligned}$

25. For $\zeta = [\zeta_i] \in \mathbb{R}^k$ and $x = [x_i] \in \mathbb{R}^k$, $\sum\limits_i \zeta_i / x_i = \zeta^T \delta(x)^{-1}\mathbf{1}$

26. For any permutation matrix $\Xi$ and dimensionally compatible vector $y$ or matrix $A$

$$\delta(\Xi\, y) \;=\; \Xi\,\delta(y)\,\Xi^T \tag{1224}$$
$$\delta(\Xi\, A\,\Xi^T) = \Xi\,\delta(A) \tag{1225}$$

So given any permutation matrix $\Xi$ and any dimensionally compatible matrix $B$ , for example,

$$\delta^2(B) = \Xi\,\delta^2(\Xi^T B\,\Xi)\Xi^T \tag{1226}$$

27. $\pi\big(\delta(A)\big) = \lambda(I \circ A)$ where $\pi$ is the presorting function

### A.1.2 Majorization

**A.1.2.0.1 Theorem.** (Schur) *Majorization.* [301, §7.4] [150, §4.3] [151, §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

$$\exists\, A \in \mathbb{S}^N \ \ni\ \lambda(A) = \lambda \ \textbf{ and } \ \delta(A) = \delta \quad \Leftarrow \quad \lambda - \delta \in \mathcal{K}_{\lambda\delta}^* \tag{1227}$$

and conversely

$$A \in \mathbb{S}^N \ \Rightarrow \ \lambda(A) - \delta(A) \in \mathcal{K}_{\lambda\delta}^* \tag{1228}$$

The difference belongs to the pointed polyhedral cone of majorization (empty interior, *confer* (271))

$$\mathcal{K}_{\lambda\delta}^* \triangleq \mathcal{K}_{\mathcal{M}+}^* \cap \{\zeta \mathbf{1} \mid \zeta \in \mathbb{R}\}^* \tag{1229}$$

where $\mathcal{K}_{\mathcal{M}+}^*$ is the dual monotone nonnegative cone (376), 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 $\mathcal{K}_{\lambda\delta}^*$ is naturally consequent to the definition of majorization; *id est*, vector $y \in \mathbb{R}^N$ *majorizes* vector $x$ if and only if

$$\sum_{i=1}^k x_i \le \sum_{i=1}^k y_i \quad \forall\, 1 \le k \le N \tag{1230}$$

and

$$\mathbf{1}^T x = \mathbf{1}^T y \tag{1231}$$

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

$$\exists\, A \in \mathbb{S}_h^N \ \ni\ \lambda(A) = \lambda \quad \Leftarrow \quad \lambda \in \mathcal{K}_{\lambda\delta}^* \tag{1232}$$

and conversely

$$A \in \mathbb{S}_h^N \ \Rightarrow \ \lambda(A) \in \mathcal{K}_{\lambda\delta}^* \tag{1233}$$

where $\mathcal{K}_{\lambda\delta}^*$ is defined in (1229). ⋄

## 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: [151, §1]

$$x^T A\, x \geq 0 \quad \text{for each and every } x \in \mathbb{R}^n \text{ such that } \|x\| = 1 \tag{1234}$$

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 $x = x_{\rm p}$, then it is nonnegative for any $\alpha x_{\rm p}$ 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$)$[A.2] are admitted to the set of positive semidefinite matrices (1237); an unnecessary prohibitive condition.

### A.2.1 Symmetry *versus* semidefiniteness

We call (1234) *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[A.3] 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*,

$$A = \frac{(A + A^T)}{2} + \frac{(A - A^T)}{2} \tag{44}$$

Because, for all real $A$, the antisymmetric part vanishes under real test,

$$x^T \frac{(A - A^T)}{2}\, x = 0 \tag{1235}$$

[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.

[A.3] Golub & Van Loan [110, §4.2.2], for example, admit nonsymmetric real matrices.

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 (1234). 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 (1234), that presumption is typically bolstered with compelling examples from the physical sciences where symmetric matrices occur within the mathematical exposition of natural phenomena.[A.4] [96, §52]

Perhaps a better explanation of this pervasive presumption of symmetry comes from Horn & Johnson [150, §7.1] whose perspective[A.5] is the complex matrix, thus necessitating the complex domain of test throughout their treatise. They explain, if $A\in\mathbb{C}^{n\times n}$

> *. . . and if $x^HAx$ is real for all $x\in\mathbb{C}^n$, then $A$ is Hermitian. Thus, the assumption that $A$ 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 $x^HAx$ comes from the Hermitian part $(A+A^H)/2$ of $A$;

$$\operatorname{Re}(x^HAx) = x^H\frac{A+A^H}{2}\,x \tag{1236}$$

rather,

$$x^HAx\in\mathbb{R} \;\Leftrightarrow\; A^H = A \tag{1237}$$

because the imaginary part of $x^HAx$ comes from the anti-Hermitian part $(A-A^H)/2$;

$$\operatorname{Im}(x^HAx) = x^H\frac{A-A^H}{2}\,x \tag{1238}$$

that vanishes for nonzero $x$ if and only if $A=A^H$. So the Hermitian symmetry assumption is unnecessary, according to Horn & Johnson, not

[A.4] Symmetric matrices are certainly pervasive in the our chosen subject as well.

[A.5] A totally complex perspective is not necessarily more advantageous. The positive semidefinite cone, for example, is not self-dual (§2.13.5) in the ambient space of Hermitian matrices. [145, §II] 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 (1234) 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 $x$ in $\mathbb{R}^n$ as in (1234), 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 $x$ 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.

**A.2.1.0.1 Example.** *Nonsymmetric positive definite product.*
Horn & Johnson assert and Zhang agrees:

> *If $A, B \in \mathbb{C}^{n\times n}$ are positive definite, then we know that the product $AB$ is positive definite if and only if $AB$ is Hermitian.* [150, §7.6, prob.10] [301, §6.2, §3.2]

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 (1234) that adopts a real domain of test.

$$A^T = A = \begin{bmatrix} 3 & 0 & -1 & 0 \\ 0 & 5 & 1 & 0 \\ -1 & 1 & 4 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}, \qquad \lambda(A) = \begin{bmatrix} 5.9 \\ 4.5 \\ 3.4 \\ 2.0 \end{bmatrix} \tag{1239}$$

$$B^T = B = \begin{bmatrix} 4 & 4 & -1 & -1 \\ 4 & 5 & 0 & 0 \\ -1 & 0 & 5 & 1 \\ -1 & 0 & 1 & 4 \end{bmatrix}, \qquad \lambda(B) = \begin{bmatrix} 8.8 \\ 5.5 \\ 3.3 \\ 0.24 \end{bmatrix} \tag{1240}$$

$$(AB)^T \neq AB = \begin{bmatrix} 13 & 12 & -8 & -4 \\ 19 & 25 & 5 & 1 \\ -5 & 1 & 22 & 9 \\ -5 & 0 & 9 & 17 \end{bmatrix}, \qquad \lambda(AB) = \begin{bmatrix} 36. \\ 29. \\ 10. \\ 0.72 \end{bmatrix} \tag{1241}$$

$$\tfrac{1}{2}(AB+(AB)^T) = \begin{bmatrix} 13 & 15.5 & -6.5 & -4.5 \\ 15.5 & 25 & 3 & 0.5 \\ -6.5 & 3 & 22 & 9 \\ -4.5 & 0.5 & 9 & 17 \end{bmatrix}, \quad \lambda\big(\tfrac{1}{2}(AB+(AB)^T)\big) = \begin{bmatrix} 36. \\ 30. \\ 10. \\ 0.014 \end{bmatrix} \tag{1242}$$

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 (1268) and Corollary A.3.1.0.5. Yet positive definiteness of the product $AB$ is certified instead by the nonnegative eigenvalues $\lambda\big(\frac{1}{2}(AB+(AB)^T)\big)$ in (1242) (§A.3.1.0.1) despite the fact $AB$ is not symmetric.[A.6] Horn & Johnson and Zhang resolve the anomaly by choosing to exclude nonsymmetric matrices and products; they do so by expanding the domain of test to $\mathbb{C}^n$. □

## A.3 Proper statements of positive semidefiniteness

Unlike Horn & Johnson and Zhang, we never adopt the complex domain of test in regard to 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 the corresponding statements for the complex case (found elsewhere, [150] [301]).

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 (1234); a few require proof as they depart from the standard texts, while those remaining are well established or obvious.

[A.6] It is a little more difficult to find a counter-example 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 real vector $x$ of unit norm, $\|x\| = 1$ ,[A.7] we have $x^T A\, x \geq 0$ (1234);

$$A \succeq 0 \;\Leftrightarrow\; \operatorname{tr}(xx^T A) \;=\; x^T A\, x \;\geq\; 0 \tag{1243}$$

Matrix $A \in \mathbb{S}^M$ is positive definite if and only if for each and every $\|x\| = 1$ we have $x^T A\, x > 0$ ;

$$A \succ 0 \;\Leftrightarrow\; \operatorname{tr}(xx^T A) \;=\; x^T A\, x \;>\; 0 \tag{1244}$$

$\diamond$

**Proof.** Statements (1243) and (1244) are each a particular instance of dual generalized inequalities (§2.13.2) with respect to the positive semidefinite cone; *videlicet*, [270]

$$\begin{array}{l} A \succeq 0 \;\Leftrightarrow\; \langle xx^T,\, A\rangle \geq 0 \;\;\forall\, xx^T(\succeq 0) \\ A \succ 0 \;\Leftrightarrow\; \langle xx^T,\, A\rangle > 0 \;\;\forall\, xx^T(\succeq 0)\,,\;\; xx^T \neq \mathbf{0} \end{array} \tag{1245}$$

Relations (1243) and (1244) remain true when $xx^T$ is replaced with "for each and every" $X \in \mathbb{S}_+^M$ [46, §2.6.1] (§2.13.5) of unit norm $\|X\| = 1$ as in

$$\begin{array}{l} A \succeq 0 \;\Leftrightarrow\; \operatorname{tr}(XA) \geq 0 \;\;\forall\, X \in \mathbb{S}_+^M \\ A \succ 0 \;\Leftrightarrow\; \operatorname{tr}(XA) > 0 \;\;\forall\, X \in \mathbb{S}_+^M\,,\quad X \neq \mathbf{0} \end{array} \tag{1246}$$

but this condition is far more than what is necessary. By the *discrete membership theorem* in §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 (1246) certifying membership to that cone. $\blacklozenge$

[A.7] The traditional condition requiring all $x \in \mathbb{R}^M$ for defining positive (semi)definiteness is actually far 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.3.1 Semidefiniteness, eigenvalues, nonsymmetric

When $A \in \mathbb{R}^{n\times n}$, let $\lambda\big(\frac{1}{2}(A+A^T)\big) \in \mathbb{R}^n$ 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] [200, §1.3.1] (*confer* §A.3.1.0.1)

$$x^T\! A\, x \geq 0 \quad \forall\, x \in \mathbb{R}^n \;\Leftrightarrow\; A + A^T \succeq 0 \;\Leftrightarrow\; \lambda(A + A^T) \succeq 0 \tag{1247}$$

- (§2.9.0.1)

$$A \succeq 0 \;\Rightarrow\; A^T = A \tag{1248}$$

$$A \succeq B \;\Leftrightarrow\; A - B \succeq 0 \;\not\Rightarrow\; A \succeq 0 \;\text{ or }\; B \succeq 0 \tag{1249}$$

$$x^T\! A\, x \geq 0 \;\; \forall\, x \;\not\Rightarrow\; A^T = A \tag{1250}$$

- Matrix symmetry is not intrinsic to positive semidefiniteness;

$$A^T = A\,, \quad \lambda(A) \succeq 0 \;\Rightarrow\; x^T\! A\, x \geq 0 \;\; \forall\, x \tag{1251}$$

$$\lambda(A) \succeq 0 \;\Leftarrow\; A^T = A\,, \quad x^T\! A\, x \geq 0 \;\; \forall\, x \tag{1252}$$

- If $A^T = A$ then

$$\lambda(A) \succeq 0 \;\Leftrightarrow\; A \succeq 0 \tag{1253}$$

  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.

$$\langle A\,,\, A\rangle = \langle \lambda(A)\,,\, \lambda(A)\rangle \tag{1254}$$

- 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 $A$

$$\lambda(\mu I + A) = \mu\mathbf{1} + \lambda(A) \tag{1255}$$

  **Proof:** $A = MJM^{-1}$ and $\mu I + A = M(\mu I + J)M^{-1}$ where $J$ is the *Jordan form* for $A$; [249, §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. ♦

[A.8] The symmetrization of $A$ is $(A + A^T)/2$. $\lambda\big(\frac{1}{2}(A+A^T)\big) = \lambda(A+A^T)/2$.

[A.9] Strang agrees [249, 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. (1238)

By similar reasoning,

$$\lambda(I+\mu A) = \mathbf{1} + \lambda(\mu A) \tag{1256}$$

For vector $\sigma(A)$ holding the singular values of any matrix $A$

$$\sigma(I+\mu A^T\!A) = \pi\big(|\mathbf{1}+\mu\,\sigma(A^T\!A)|\big) \tag{1257}$$
$$\sigma(\mu I+A^T\!A) = \pi\big(|\mu\mathbf{1}+\sigma(A^T\!A)|\big) \tag{1258}$$

where $\pi$ is the nonlinear permutation operator sorting its vector argument into nonincreasing order.

- For $A\in\mathbb{S}^M$ and each and every $\|w\|=1$ [150, §7.7, prob.9]

$$w^T\!Aw \le \mu \;\Leftrightarrow\; A \preceq \mu I \;\Leftrightarrow\; \lambda(A) \preceq \mu\mathbf{1} \tag{1259}$$

- [150, §2.5.4] (*confer* (36))

$$A \text{ is normal } \Leftrightarrow \|A\|_{\mathrm F}^2 = \lambda(A)^T\lambda(A) \tag{1260}$$

- For $A\in\mathbb{R}^{m\times n}$

$$A^T\!A \succeq 0\,, \qquad AA^T \succeq 0 \tag{1261}$$

because, for dimensionally compatible vector $x$, $x^T\!A^T\!Ax = \|Ax\|_2^2$ , $x^T\!AA^T\!x = \|A^Tx\|_2^2$ .

- For $A\in\mathbb{R}^{n\times n}$ and $c\in\mathbb{R}$

$$\operatorname{tr}(cA) \;=\; c\operatorname{tr}(A) \;=\; c\,\mathbf{1}^T\lambda(A) \qquad (\text{§A.1.1 no.6})$$

For $m$ a nonnegative integer, (1637)

$$\det(A^m) = \prod_{i=1}^n \lambda(A)_i^m \tag{1262}$$

$$\operatorname{tr}(A^m) = \sum_{i=1}^n \lambda(A)_i^m \tag{1263}$$

- For $A$ diagonalizable (§A.5), $A = S\Lambda S^{-1}$, (*confer* [249, p.255])

$$\operatorname{rank} A = \operatorname{rank} \delta\big(\lambda(A)\big) = \operatorname{rank} \Lambda \tag{1264}$$

meaning, rank is equal to the number of nonzero eigenvalues in vector

$$\lambda(A) \stackrel{\Delta}{=} \delta(\Lambda) \tag{1265}$$

by the 0 *eigenvalues theorem* (§A.7.3.0.1).

- (Fan) For $A, B \in \mathbb{S}^n$ [41, §1.2] (*confer* (1520))

$$\operatorname{tr}(AB) \leq \lambda(A)^T \lambda(B) \tag{1266}$$

with equality (Theobald) when $A$ and $B$ are simultaneously diagonalizable [150] with the same ordering of eigenvalues.

- For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$

$$\operatorname{tr}(AB) = \operatorname{tr}(BA) \tag{1267}$$

and $\eta$ eigenvalues of the product and commuted product are identical, including their multiplicity; [150, §1.3.20] *id est*,

$$\lambda(AB)_{1:\eta} = \lambda(BA)_{1:\eta}\,, \qquad \eta \stackrel{\Delta}{=} \min\{m\,, n\} \tag{1268}$$

Any eigenvalues remaining are zero. By the 0 *eigenvalues theorem* (§A.7.3.0.1),

$$\operatorname{rank}(AB) = \operatorname{rank}(BA)\,, \quad AB \text{ and } BA \text{ diagonalizable} \tag{1269}$$

- For any compatible matrices $A, B$ [150, §0.4]

$$\min\{\operatorname{rank} A\,, \operatorname{rank} B\} \geq \operatorname{rank}(AB) \tag{1270}$$

- For $A, B \in \mathbb{S}^n_+$ (218)

$$\operatorname{rank} A + \operatorname{rank} B \geq \operatorname{rank}(A+B) \geq \min\{\operatorname{rank} A\,, \operatorname{rank} B\} \geq \operatorname{rank}(AB) \tag{1271}$$

- For $A, B \in \mathbb{S}^n_+$ linearly independent (§B.1.1),

$$\operatorname{rank} A + \operatorname{rank} B = \operatorname{rank}(A+B) > \min\{\operatorname{rank} A\,, \operatorname{rank} B\} \geq \operatorname{rank}(AB) \tag{1272}$$

- Because $\mathcal{R}(A^TA) = \mathcal{R}(A^T)$ and $\mathcal{R}(AA^T) = \mathcal{R}(A)$, for any $A \in \mathbb{R}^{m\times n}$

$$\operatorname{rank}(AA^T) = \operatorname{rank}(A^TA) = \operatorname{rank} A \tag{1273}$$

- For $A \in \mathbb{R}^{m\times n}$ having no nullspace, and for any $B \in \mathbb{R}^{n\times k}$

$$\operatorname{rank}(AB) = \operatorname{rank}(B) \tag{1274}$$

**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 (1369). ♦

- For $A \in \mathbb{S}^n$ and any nonsingular matrix $Y$

$$\operatorname{inertia}(A) = \operatorname{inertia}(YAY^T) \tag{1275}$$

`a.k.a`, *Sylvester's law of inertia.* (1316) [77, §2.4.3]

- For $A, B \in \mathbb{R}^{n\times n}$ square, [150, §0.3.5]

$$\det(AB) = \det(BA) \tag{1276}$$
$$\det(AB) = \det A \det B \tag{1277}$$

Yet for $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$ [55, p.72]

$$\det(I + AB) = \det(I + BA) \tag{1278}$$

- For $A, B \in \mathbb{S}^n$, product $AB$ is symmetric if and only if $AB$ is commutative;

$$(AB)^T = AB \;\Leftrightarrow\; AB = BA \tag{1279}$$

**Proof.** (⇒) Suppose $AB = (AB)^T$. $(AB)^T = B^TA^T = BA$. $AB = (AB)^T \Rightarrow AB = BA$.
(⇐) Suppose $AB = BA$. $BA = B^TA^T = (AB)^T$. $AB = BA \Rightarrow AB = (AB)^T$. ♦

Commutativity alone is insufficient for symmetry of the product. [249, p.26]

- Diagonalizable matrices $A\,,B\in\mathbb{R}^{n\times n}$ commute if and only if they are *simultaneously diagonalizable.* [150, §1.3.12] A product of diagonal matrices is always commutative.

- For $A,B\in\mathbb{R}^{n\times n}$ and $AB = BA$

$$x^TA\,x \geq 0\,,\; x^TB\,x \geq 0 \;\forall\,x \;\Rightarrow\; \lambda(A+A^T)_i\;\lambda(B+B^T)_i \geq 0 \;\forall\,i \;\not\Leftrightarrow\; x^TAB\,x \geq 0 \;\forall\,x \tag{1280}$$

the negative result arising because of the schism between the product of eigenvalues $\lambda(A+A^T)_i\;\lambda(B+B^T)_i$ and the eigenvalues of the symmetrized matrix product $\lambda\big(AB+(AB)^T\big)_i$ . For example, $X^2$ is generally not positive semidefinite unless matrix $X$ is symmetric; then (1261) applies. Simply substituting symmetric matrices changes the outcome:

- For $A,B\in\mathbb{S}^n$ and $AB = BA$

$$A\succeq 0\,,\; B\succeq 0 \;\Rightarrow\; \lambda(AB)_i\!=\!\lambda(A)_i\,\lambda(B)_i\!\geq\! 0 \;\;\forall\,i \;\Leftrightarrow\; AB\succeq 0 \tag{1281}$$

Positive semidefiniteness of $A$ and $B$ is sufficient but not a necessary condition for positive semidefiniteness of the product $AB$ .

**Proof.** Because all symmetric matrices are diagonalizable, (§A.5.2) [249, §5.6] we have $A\!=\!S\Lambda S^T$ 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 $S$ , [150, §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 $i\!=\!1\ldots n$ . $AB\!=\!S\Lambda\Delta S^T$ is symmetric and has nonnegative eigenvalues contained in diagonal matrix $\Lambda\Delta$ by assumption; hence positive semidefinite by (1247). 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 $AB\!=\!S\Lambda\Delta S^T\!\succeq\!0$. Then $\Lambda\Delta\!\succeq\!0$ by (1247), and so all products $\lambda(A)_i\,\lambda(B)_i$ must be nonnegative; meaning, $\operatorname{sgn}(\lambda(A))\!=\!\operatorname{sgn}(\lambda(B))$. 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)

$$AB=BA\ \Rightarrow\ \lambda(AB)_i=\lambda(A)_i\,\lambda(B)_i\geq 0\ \ \forall\,i\ \ \Rightarrow\ AB\succeq 0 \tag{1282}$$

$$AB=BA\ \Rightarrow\ \lambda(AB)_i\geq 0\,,\ \ \lambda(A)_i\,\lambda(B)_i\geq 0\ \forall\,i\ \ \Leftrightarrow\ AB\succeq 0 \tag{1283}$$

- For $A,B\in\mathbb{S}^n$ [301, §6.2]

$$A\succeq 0\ \ \Rightarrow\ \ \operatorname{tr}A\geq 0 \tag{1284}$$

$$A\succeq 0\,,\ B\succeq 0\ \ \Rightarrow\ \ \operatorname{tr}A\ \operatorname{tr}B\ \geq\ \operatorname{tr}(AB)\geq 0 \tag{1285}$$

Because $A\succeq 0\,,\ B\succeq 0\ \Rightarrow\ \lambda(AB)=\lambda(\sqrt{A}B\sqrt{A})\succeq 0$ by (1268) and Corollary A.3.1.0.5, then we have $\operatorname{tr}(AB)\geq 0$.

$$A\succeq 0\ \ \Leftrightarrow\ \ \operatorname{tr}(AB)\geq 0\quad \forall\,B\succeq 0 \qquad\qquad (322)$$

- For $A,B,C\in\mathbb{S}^n$ (Löwner)

$$A\preceq B\,,\ \ B\preceq C\ \ \Rightarrow\ \ A\preceq C \tag{1286}$$

$$A\preceq B\ \Leftrightarrow\ A+C\preceq B+C \tag{1287}$$

$$A\preceq B\,,\ \ A\succeq B\ \ \Rightarrow\ \ A=B \tag{1288}$$

- For $A,B\in\mathbb{R}^{n\times n}$

$$x^TA\,x\geq x^TB\,x\ \ \forall\,x\ \ \Rightarrow\ \ \operatorname{tr}A\geq\operatorname{tr}B \tag{1289}$$

**Proof.** $x^TA\,x\geq x^TB\,x\ \forall\,x\ \ \Leftrightarrow\ \ \lambda\big((A-B)+(A-B)^T\big)/2\succeq 0\ \ \Rightarrow$ $\operatorname{tr}\big(A+A^T-(B+B^T)\big)/2=\operatorname{tr}(A-B)\geq 0$. There is no converse. ◆

- For $A,B\in\mathbb{S}^n$ [301, §6.2, prob.1] (Theorem A.3.1.0.4)

$$A\succeq B\ \Rightarrow\ \ \operatorname{tr}A\geq\operatorname{tr}B \tag{1290}$$

$$A\succeq B\ \Rightarrow\ \delta(A)\succeq\delta(B) \tag{1291}$$

There is no converse, and restriction to the positive semidefinite cone does not improve the situation. The all-strict versions hold. From [301, §6.2]

$$A\succeq B\succeq 0\ \Rightarrow\ \ \operatorname{rank}A\geq\operatorname{rank}B \tag{1292}$$

$$A\succeq B\succeq 0\ \Rightarrow\ \ \det A\geq\det B\geq 0 \tag{1293}$$

$$A\succ B\succeq 0\ \Rightarrow\ \ \det A>\det B\geq 0 \tag{1294}$$

- For $A,B\in\operatorname{int}\mathbb{S}_+^n$ [27, §4.2] [150, §7.7.4]

$$A\succeq B\ \Leftrightarrow\ A^{-1}\preceq B^{-1} \tag{1295}$$

- For $A, B \in \mathbb{S}^n$ [301, §6.2]

$$A \succeq B \succeq 0 \;\Rightarrow\; \sqrt{A} \succeq \sqrt{B} \tag{1296}$$

- For $A, B \in \mathbb{S}^n$ and $AB = BA$ [301, §6.2, prob.3]

$$A \succeq B \succeq 0 \;\Rightarrow\; A^k \succeq B^k\,, \quad k=1,2,\ldots \tag{1297}$$

**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 $\lambda\big(\frac{1}{2}(A+A^T)\big)\,, \lambda\big(\frac{1}{2}(B+B^T)\big) \in \mathbb{R}^M$. Then, [249, §6]

$$\begin{aligned} x^TA\,x \geq 0 \;\; \forall\,x \quad & \Leftrightarrow\; \lambda\big(A+A^T\big) \succeq 0 & (1298)\\ x^TA\,x > 0 \;\; \forall\,x \neq \mathbf{0} \; & \Leftrightarrow\; \lambda\big(A+A^T\big) \succ 0 & (1299)\end{aligned}$$

because $x^T(A-A^T)x=0$. (1235) Now arrange the entries of $\lambda\big(\frac{1}{2}(A+A^T)\big)$ and $\lambda\big(\frac{1}{2}(B+B^T)\big)$ in nonincreasing order so $\lambda\big(\frac{1}{2}(A+A^T)\big)_1$ holds the largest eigenvalue of symmetrized $A$ while $\lambda\big(\frac{1}{2}(B+B^T)\big)_1$ holds the largest eigenvalue of symmetrized $B$, and so on. Then [150, §7.7, prob.1, prob.9] for $\kappa \in \mathbb{R}$

$$\begin{aligned} x^TA\,x \geq x^TB\,x \;\; \forall\,x \;\; & \Rightarrow \;\; \lambda\big(A+A^T\big) \succeq \lambda\big(B+B^T\big) \\ x^TA\,x \geq x^TI\,x\,\kappa \;\; \forall\,x \;\; & \Leftrightarrow \;\; \lambda\big(\tfrac{1}{2}(A+A^T)\big) \succeq \kappa\mathbf{1} \end{aligned} \tag{1300}$$

Now let $A\,, B \in \mathbb{S}^M$ have diagonalizations $A=Q\Lambda Q^T$ and $B=U\Upsilon U^T$ with $\lambda(A)=\delta(\Lambda)$ and $\lambda(B)=\delta(\Upsilon)$ arranged in nonincreasing order. Then

$$\begin{aligned} A \succeq B \;&\Leftrightarrow\; \lambda(A-B) \succeq 0 & (1301)\\ A \succeq B \;&\Rightarrow\; \lambda(A) \succeq \lambda(B) & (1302)\\ A \succeq B \;&\not\Leftarrow\; \lambda(A) \succeq \lambda(B) & (1303)\\ S^TAS \succeq B \;&\Leftarrow\; \lambda(A) \succeq \lambda(B) & (1304)\end{aligned}$$

where $S=QU^T$. [301, §7.5] ⋄

**A.3.1.0.2 Theorem.** (Weyl) *Eigenvalues of sum.* [150, §4.3]
For $A\,,B\in\mathbb{R}^{M\times M}$, place the eigenvalues of each symmetrized matrix into the respective vectors $\lambda\big(\frac{1}{2}(A+A^T)\big)\,,\,\lambda\big(\frac{1}{2}(B+B^T)\big)\in\mathbb{R}^M$ in nonincreasing order so $\lambda\big(\frac{1}{2}(A+A^T)\big)_1$ holds the largest eigenvalue of symmetrized $A$ while $\lambda\big(\frac{1}{2}(B+B^T)\big)_1$ holds the largest eigenvalue of symmetrized $B$, and so on. Then, for any $k\in\{1\ldots M\}$

$$\lambda\big(A+A^T\big)_k+\lambda\big(B+B^T\big)_M\;\leq\;\lambda\big((A+A^T)+(B+B^T)\big)_k\;\leq\;\lambda\big(A+A^T\big)_k+\lambda\big(B+B^T\big)_1 \tag{1305}$$

$\diamond$

Weyl's theorem establishes positive semidefiniteness of a sum of positive semidefinite matrices. Because $\mathbb{S}_+^M$ is a convex cone (§2.9.0.0.1), then by (144)

$$A\,,B\succeq 0\;\Rightarrow\;\zeta A+\xi B\succeq 0\;\text{ for all }\;\zeta\,,\xi\geq 0 \tag{1306}$$

**A.3.1.0.3 Corollary.** *Eigenvalues of sum and difference.* [150, §4.3]
For $A\in\mathbb{S}^M$ and $B\in\mathbb{S}_+^M$, place the eigenvalues of each matrix into the 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 $k\in\{1\ldots M\}$

$$\lambda(A-B)_k\;\leq\;\lambda(A)_k\;\leq\;\lambda(A+B)_k \tag{1307}$$

$\diamond$

When $B$ is rank-one positive semidefinite, the eigenvalues interlace; *id est*, for $B=qq^T$

$$\lambda(A)_{k-1}\;\leq\;\lambda(A-qq^T)_k\;\leq\;\lambda(A)_k\;\leq\;\lambda(A+qq^T)_k\;\leq\;\lambda(A)_{k+1} \tag{1308}$$

**A.3.1.0.4 Theorem.** *Positive (semi)definite principal submatrices.*[A.10]

- $A \in \mathbb{S}^M$ 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. [200, §1.3.1]

**A.3.1.0.5 Corollary.** *Positive (semi)definite symmetric products.* [150, p.399]

- If $A \in \mathbb{S}^M$ is positive definite and any particular dimensionally compatible matrix $Z$ has no nullspace, then $Z^TAZ$ is positive definite.

- If matrix $A \in \mathbb{S}^M$ is positive (semi)definite then, for any matrix $Z$ of compatible dimension, $Z^TAZ$ is positive semidefinite.

- $A \in \mathbb{S}^M$ is positive (semi)definite if and only if there exists a nonsingular $Z$ such that $Z^TAZ$ 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 $Z^TAZ$ becomes positive definite.[A.11] ⋄

[A.10] A recursive condition for positive (semi)definiteness, this theorem is a synthesis of facts from [150, §7.2] [249, §6.3] (*confer* [200, §1.3.1]). Principal submatrices are formed by discarding any subset of rows and columns having the same indices. There are $M!/(1!(M-1)!)$ principal $1\times1$ submatrices, $M!/(2!(M-2)!)$ principal $2\times2$ submatrices, and so on, totaling $2^M-1$ principal submatrices including $A$ itself. By loading $y$ in $y^TAy$ with various patterns of ones and zeros, it follows that any principal submatrix must be positive (semi)definite whenever $A$ is.

[A.11] Using the interpretation in §E.6.4.3, this means coefficients of orthogonal projection of vectorized $A$ on a subset of extreme directions from $\mathbb{S}^M_+$ determined by $Z$ can be positive.

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 $\begin{bmatrix} Z^T \\ Y^T \end{bmatrix} A \,[\, Z \quad Y \,]$ is positive semidefinite. In other words, from the Corollary it follows: for dimensionally compatible $Z$

- $A \succeq 0 \;\Leftrightarrow\; Z^TAZ \succeq 0$ and $Z^T$ has a left inverse

Products such as $Z^\dagger Z$ and $ZZ^\dagger$ are symmetric and positive semidefinite although, given $A \succeq 0$, $Z^\dagger AZ$ and $ZAZ^\dagger$ are neither necessarily symmetric or positive semidefinite.

**A.3.1.0.6 Theorem.** *Symmetric projector semidefinite.* [17, §III] [18, §6] [163, p.55] For symmetric idempotent matrices $P$ and $R$

$$\begin{gathered} P, R \succeq 0 \\ P \succeq R \;\Leftrightarrow\; \mathcal{R}(P) \supseteq \mathcal{R}(R) \;\Leftrightarrow\; \mathcal{N}(P) \subseteq \mathcal{N}(R) \end{gathered} \tag{1309}$$

Projector $P$ is never positive definite [251, §6.5, prob.20] unless it is the identity matrix. ⋄

**A.3.1.0.7 Theorem.** *Symmetric positive semidefinite.*
Given real matrix $\Psi$ with $\operatorname{rank}\Psi = 1$

$$\Psi \succeq 0 \;\Leftrightarrow\; \Psi = uu^T \tag{1310}$$

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 $\Psi = uv^T$. (§B.1) Suppose $\Psi$ is symmetric; *id est*, $v = u$. For all $y \in \mathbb{R}^M$, $y^Tu\,u^Ty \geq 0$. Conversely, suppose $uv^T$ is positive semidefinite. We know that can hold if and only if $uv^T + vu^T \succeq 0 \;\Leftrightarrow\;$ for all normalized $y \in \mathbb{R}^M$, $2\,y^Tu\,v^Ty \geq 0$; 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 the *Schur-form* partitioned matrix $G$: Given $A^T = A$ and $C^T = C$, then [44]

$$\begin{array}{c} G = \left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right] \succeq 0 \\ \Leftrightarrow \;\; A \succeq 0\,, \quad B^T(I - AA^\dagger) = \mathbf{0}\,, \quad C - B^T A^\dagger B \succeq 0 \\ \Leftrightarrow \;\; C \succeq 0\,, \quad B(I - CC^\dagger) = \mathbf{0}\,, \quad A - B\,C^\dagger B^T \succeq 0 \end{array} \tag{1311}$$

where $A^\dagger$ denotes the Moore-Penrose (pseudo)inverse (§E). In the first instance, $I - AA^\dagger$ is a symmetric projection matrix orthogonally projecting on $\mathcal{N}(A^T)$. (1678) It is apparently required

$$\mathcal{R}(B) \perp \mathcal{N}(A^T) \tag{1312}$$

which precludes $A = \mathbf{0}$ 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, $I - CC^\dagger$ projects orthogonally on $\mathcal{N}(C^T)$. It is required

$$\mathcal{R}(B^T) \perp \mathcal{N}(C^T) \tag{1313}$$

which precludes $C = \mathbf{0}$ for $B$ nonzero. Again, $C \succ 0 \;\Rightarrow\; C^\dagger = C^{-1}$. So we get, for $A$ or $C$ nonsingular,

$$\begin{array}{c} G = \left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right] \succeq 0 \\ \Leftrightarrow \\ A \succ 0\,, \quad C - B^T A^{-1} B \succeq 0 \\ \textbf{or} \\ C \succ 0\,, \quad A - B\,C^{-1} B^T \succeq 0 \end{array} \tag{1314}$$

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,

$$\begin{array}{c} G = \left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right] \succ 0 \\ \Leftrightarrow \;\; A \succ 0\,, \quad C - B^T A^{-1} B \succ 0 \\ \Leftrightarrow \;\; C \succ 0\,, \quad A - BC^{-1} B^T \succ 0 \end{array} \tag{1315}$$

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 $A - BC^{-1}B^T$. [103, §4.8]

Origin of the term *Schur complement* is from complementary *inertia*: [77, §2.4.4] Define

$$\operatorname{inertia}\big(G\in\mathbb{S}^M\big) \triangleq \{p\,,z\,,n\} \tag{1316}$$

where $p\,,z\,,n$ respectively represent number of positive, zero, and negative eigenvalues of $G$; *id est*,

$$M = p+z+n \tag{1317}$$

Then, when $A$ is invertible,

$$\operatorname{inertia}(G) = \operatorname{inertia}(A) + \operatorname{inertia}(C - B^TA^{-1}B) \tag{1318}$$

and when $C$ is invertible,

$$\operatorname{inertia}(G) = \operatorname{inertia}(C) + \operatorname{inertia}(A - BC^{-1}B^T) \tag{1319}$$

When $A=C=\mathbf{0}$, denoting by $\sigma(B)\in\mathbb{R}^m_+$ the nonincreasingly ordered singular values of matrix $B\in\mathbb{R}^{m\times m}$, then we have the eigenvalues [41, §1.2, prob.17]

$$\lambda(G) \;=\; \lambda\left(\begin{bmatrix}\mathbf{0} & B\\ B^T & \mathbf{0}\end{bmatrix}\right) \;=\; \begin{bmatrix}\sigma(B)\\ -\Xi\,\sigma(B)\end{bmatrix} \tag{1320}$$

and

$$\operatorname{inertia}(G) = \operatorname{inertia}(B^TB) + \operatorname{inertia}(-B^TB) \tag{1321}$$

where $\Xi$ is the order-reversing permutation matrix defined in (1507).

**A.4.0.0.1 Example.** *Nonnegative polynomial.* [27, p.163]
Schur-form positive semidefiniteness is necessary and sufficient for quadratic polynomial nonnegativity; *videlicet*, for all compatible $x$

$$[\,x^T\;\;1\,]\begin{bmatrix}A & b\\ b^T & c\end{bmatrix}\begin{bmatrix}x\\ 1\end{bmatrix} \geq 0 \quad\Leftrightarrow\quad x^TA\,x + 2b^Tx + c \geq 0 \tag{1322}$$

$\square$

**A.4.0.0.2 Example.** *Sparse Schur conditions.*
Setting matrix $A$ to the identity simplifies the Schur conditions. One consequence relates the definiteness of three quantities:

$$\begin{bmatrix}I & B\\ B^T & C\end{bmatrix}\succeq 0 \quad\Leftrightarrow\quad C-B^TB\succeq 0 \quad\Leftrightarrow\quad \begin{bmatrix}I & \mathbf{0}\\ \mathbf{0}^T & C-B^TB\end{bmatrix}\succeq 0 \tag{1323}$$

$\square$

**A.4.0.0.3 Exercise.** *Eigenvalues $\lambda$ of sparse Schur-form.*
Prove: given $C-B^TB=\mathbf{0}$, for $B\in\mathbb{R}^{m\times n}$ and $C\in\mathbb{S}^n$

$$\lambda\left(\begin{bmatrix} I & B\\ B^T & C\end{bmatrix}\right)_i = \begin{cases} 1+\lambda(C)_i\,, & 1\le i\le n\\ 1\,, & n<i\le m\\ 0\,, & \text{otherwise}\end{cases} \tag{1324}$$

▼

**A.4.0.0.4 Theorem.** *Rank of partitioned matrices.*
When symmetric matrix $A$ is invertible and $C$ is symmetric,

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} A & B\\ B^T & C\end{bmatrix} &= \operatorname{rank}\begin{bmatrix} A & \mathbf{0}\\ \mathbf{0}^T & C-B^TA^{-1}B\end{bmatrix}\\ &= \operatorname{rank} A \,+\, \operatorname{rank}(C-B^TA^{-1}B)\end{aligned} \tag{1325}$$

equals rank of a block on the main diagonal plus rank of its Schur complement [301, §2.2, prob.7]. Similarly, when symmetric matrix $C$ is invertible and $A$ is symmetric,

$$\begin{aligned}\operatorname{rank}\begin{bmatrix} A & B\\ B^T & C\end{bmatrix} &= \operatorname{rank}\begin{bmatrix} A-BC^{-1}B^T & \mathbf{0}\\ \mathbf{0}^T & C\end{bmatrix}\\ &= \operatorname{rank}(A-BC^{-1}B^T) + \operatorname{rank} C\end{aligned} \tag{1326}$$

⋄

**Proof.** The first assertion (1325) holds if and only if [150, §0.4.6(c)]

$$\exists \text{ nonsingular } X, Y \;\ni\; X\begin{bmatrix} A & B\\ B^T & C\end{bmatrix}Y = \begin{bmatrix} A & \mathbf{0}\\ \mathbf{0}^T & C-B^TA^{-1}B\end{bmatrix} \tag{1327}$$

Let [150, §7.7.6]

$$Y \,=\, X^T = \begin{bmatrix} I & -A^{-1}B\\ \mathbf{0}^T & I\end{bmatrix} \tag{1328}$$

♦

**A.4.0.0.5 Lemma.** *Rank of Schur-form block.* [92] [90]
Matrix $B\in\mathbb{R}^{m\times n}$ has $\operatorname{rank} B\le\rho$ if and only if there exist matrices $A\in\mathbb{S}^m$ and $C\in\mathbb{S}^n$ such that

$$\operatorname{rank}\begin{bmatrix} A & \mathbf{0}\\ \mathbf{0}^T & C\end{bmatrix} \le 2\rho \qquad \textbf{and} \qquad G = \begin{bmatrix} A & B\\ B^T & C\end{bmatrix} \succeq 0 \tag{1329}$$

⋄

Schur-form positive semidefiniteness alone implies $\operatorname{rank} A \geq \operatorname{rank} B$ and $\operatorname{rank} C \geq \operatorname{rank} B$. But, even in absence of semidefiniteness, we must always have $\operatorname{rank} G \geq \operatorname{rank} A, \operatorname{rank} B, \operatorname{rank} C$ by fundamental linear algebra.

### A.4.1 Determinant

$$G = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \tag{1330}$$

We consider again a matrix $G$ partitioned like (1311), but not necessarily positive (semi)definite, where $A$ and $C$ are symmetric.

- When $A$ is invertible,

  $$\det G \;=\; \det A \;\det(C - B^TA^{-1}B) \tag{1331}$$

  When $C$ is invertible,

  $$\det G \;=\; \det C \;\det(A - BC^{-1}B^T) \tag{1332}$$

- When $B$ is full-rank and skinny, $C = \mathbf{0}$, and $A \succeq 0$, then [46, §10.1.1]

  $$\det G \neq 0 \;\Leftrightarrow\; A + BB^T \succ 0 \tag{1333}$$

  When $B$ is a (column) vector, then for all $C \in \mathbb{R}$ and all $A$ of dimension compatible with $G$

  $$\det G \;=\; \det(A)C - B^TA_{\mathrm{cof}}^TB \tag{1334}$$

  while for $C \neq 0$

  $$\det G \;=\; C\,\det(A - \frac{1}{C}BB^T) \tag{1335}$$

  where $A_{\mathrm{cof}}$ is the matrix of cofactors [249, §4] corresponding to $A$.

- When $B$ is full-rank and fat, $A = \mathbf{0}$, and $C \succeq 0$, then

  $$\det G \neq 0 \;\Leftrightarrow\; C + B^TB \succ 0 \tag{1336}$$

  When $B$ is a row vector, then for $A \neq 0$ and all $C$ of dimension compatible with $G$

  $$\det G \;=\; A\,\det(C - \frac{1}{A}B^TB) \tag{1337}$$

  while for all $A \in \mathbb{R}$

  $$\det G \;=\; \det(C)A - BC_{\mathrm{cof}}^TB^T \tag{1338}$$

  where $C_{\mathrm{cof}}$ is the matrix of cofactors corresponding to $C$.

## A.5 eigen decomposition

When a square matrix $X\in\mathbb{R}^{m\times m}$ is *diagonalizable*, [249, §5.6] then

$$X \;=\; S\Lambda S^{-1} = [\,s_1 \cdots s_m\,]\;\Lambda \begin{bmatrix} w_1^T \\ \vdots \\ w_m^T \end{bmatrix} = \sum_{i=1}^{m} \lambda_i\, s_i w_i^T \tag{1339}$$

where $s_i\in\mathbb{C}^m$ are linearly independent (right-)eigenvectors[A.12] constituting the columns of $S\in\mathbb{C}^{m\times m}$ defined by

$$XS = S\Lambda \tag{1340}$$

$w_i^T\in\mathbb{C}^m$ are linearly independent *left-eigenvectors* of $X$ constituting the rows of $S^{-1}$ defined by [150]

$$S^{-1}X = \Lambda S^{-1} \tag{1341}$$

and where $\{\lambda_i\in\mathbb{C}\}$ are eigenvalues (populating diagonal matrix $\Lambda\in\mathbb{C}^{m\times m}$) corresponding to both left and right eigenvectors; *id est*, $\lambda(X)=\lambda(X^T)$.

There is no connection between diagonalizability and invertibility of $X$. [249, §5.2] Diagonalizability is guaranteed by a full set of linearly independent eigenvectors, whereas invertibility is guaranteed by all nonzero eigenvalues.

$$\begin{array}{c} \text{distinct eigenvalues} \;\Rightarrow\; \text{l.i. eigenvectors} \;\Leftrightarrow\; \text{diagonalizable} \\ \text{not diagonalizable} \;\Rightarrow\; \text{repeated eigenvalue} \end{array} \tag{1342}$$

**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 $\lambda=\lambda^*$, $x^HAx=\lambda x^Hx=\lambda\|x\|^2=x^TAx^* \Rightarrow x=x^*$, where $x^H=x^{*T}$. The converse is equally simple. ♦

### A.5.0.1 Uniqueness

From the *fundamental theorem of algebra* 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.)

Uniqueness of eigenvectors, in contrast, disallows multiplicity of the same direction.

[A.12] Eigenvectors must, of course, be nonzero. The prefix *eigen* is from the German; in this context meaning, something akin to "characteristic". [246, p.14]

**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 (1339), for example, then $-S$ is certainly another matrix of eigenvectors decomposing $X$ with the same eigenvalues.

For any square matrix, the eigenvector corresponding to a distinct eigenvalue is unique; [246, p.220]

$$\text{distinct eigenvalues} \;\Rightarrow\; \text{eigenvectors unique} \tag{1343}$$

Eigenvectors corresponding to a repeated eigenvalue are not unique for a diagonalizable matrix;

$$\text{repeated eigenvalue} \;\Rightarrow\; \text{eigenvectors not unique} \tag{1344}$$

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[A.13] $\dim\mathcal{N}(X-\lambda I)$ 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,

$$\text{distinct eigenvalues} \;\Leftrightarrow\; \text{eigenvectors unique} \tag{1345}$$

### A.5.1 eigenmatrix

The (right-)eigenvectors $\{s_i\}$ are naturally orthogonal to the left-eigenvectors $\{w_i\}$ except, for $i=1\ldots m$, $w_i^T s_i = 1$; called a biorthogonality condition [275, §2.2.4] [150] 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

[A.13] For a diagonalizable matrix, *algebraic multiplicity* is the same as *geometric multiplicity.* [246, p.15]

$$s_i w_i^T s_i w_i^T = s_i w_i^T \tag{1346}$$

(whereas the dyads of singular value decomposition are not inherently projectors (*confer* (1350))).

The dyads of eigen decomposition can be termed *eigenmatrices* because

$$X\, s_i w_i^T = \lambda_i\, s_i w_i^T \tag{1347}$$

### A.5.2 Symmetric matrix diagonalization

The set of *normal matrices* is, precisely, that set of all real matrices having a complete orthonormal set of eigenvectors; [301, §8.1] [251, prob.10.2.31] *id est*, any matrix $X$ for which $XX^T = X^TX$ ; [110, §7.1.3] [246, p.3] *e.g.*, orthogonal and circulant matrices [118]. All normal matrices are diagonalizable. A symmetric matrix is a special normal matrix whose eigenvalues must be real and whose eigenvectors can be chosen to make a real orthonormal set; [251, §6.4] [249, p.315] *id est*, for $X \in \mathbb{S}^m$

$$X \;=\; S\Lambda S^T \;=\; [\, s_1 \cdots s_m \,]\; \Lambda \begin{bmatrix} s_1^T \\ \vdots \\ s_m^T \end{bmatrix} \;=\; \sum_{i=1}^{m} \lambda_i\, s_i s_i^T \tag{1348}$$

where $\delta^2(\Lambda) = \Lambda \in \mathbb{S}^m$ (§A.1) and $S^{-1} = S^T \in \mathbb{R}^{m\times m}$ (orthogonal matrix, §B.5) because of symmetry: $S\Lambda S^{-1} = S^{-T}\Lambda S^T$.

Because the 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.5.2.1 Positive semidefinite matrix square root

When $X \in \mathbb{S}_+^m$ , its unique positive semidefinite matrix square root is defined

$$\sqrt{X} \triangleq S\sqrt{\Lambda}\, S^T \in \mathbb{S}_+^m \tag{1349}$$

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

[110, §2.5.4] For any $A\in\mathbb{R}^{m\times n}$

$$A = U\Sigma Q^T = [\,u_1 \cdots u_\eta\,]\;\Sigma \begin{bmatrix} q_1^T \\ \vdots \\ q_\eta^T \end{bmatrix} = \sum_{i=1}^{\eta} \sigma_i\, u_i q_i^T \qquad U\in\mathbb{R}^{m\times\eta}\,,\quad \Sigma\in\mathbb{R}^{\eta\times\eta}\,,\quad Q\in\mathbb{R}^{n\times\eta} \tag{1350}$$

where $U$ and $Q$ are always skinny-or-square each having orthonormal columns, and where

$$\eta \triangleq \min\{m\,,\,n\} \tag{1351}$$

Square matrix $\Sigma$ is diagonal (§A.1.1)

$$\delta^2(\Sigma) = \Sigma\in\mathbb{R}^{\eta\times\eta} \tag{1352}$$

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 by[A.14]

$$\sigma(A)_i = \sigma(A^T)_i = \begin{cases} \sqrt{\lambda(A^TA)_i} = \sqrt{\lambda(AA^T)_i} = \lambda\Big(\sqrt{A^TA}\Big)_i = \lambda\Big(\sqrt{AA^T}\Big)_i > 0\,, & 1\leq i\leq\rho \\ 0\,, & \rho<i\leq\eta \end{cases} \tag{1353}$$

of which the last $\eta-\rho$ are $0$ ,[A.15] where

$$\rho \triangleq \operatorname{rank} A = \operatorname{rank}\Sigma \tag{1354}$$

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 (1350), where [110, §2.5.3]

$$\begin{array}{l} \mathcal{R}\{u_i \mid \sigma_i\neq 0\} = \mathcal{R}(A) \\ \mathcal{R}\{u_i \mid \sigma_i=0\} \subseteq \mathcal{N}(A^T) \\ \mathcal{R}\{q_i \mid \sigma_i\neq 0\} = \mathcal{R}(A^T) \\ \mathcal{R}\{q_i \mid \sigma_i=0\} \subseteq \mathcal{N}(A) \end{array} \tag{1355}$$

---

[A.14] When $A$ is normal, $\sigma(A)=|\lambda(A)|$. [301, §8.1]

[A.15] For $\eta=n$ , $\sigma(A)=\sqrt{\lambda(A^TA)}=\lambda\Big(\sqrt{A^TA}\Big)$ where $\lambda$ denotes eigenvalues. For $\eta=m$ , $\sigma(A)=\sqrt{\lambda(AA^T)}=\lambda\Big(\sqrt{AA^T}\Big)$.

### 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 $A\in\mathbb{R}^{m\times n}$

$$\rho \;=\; \operatorname{rank} A \;=\; \operatorname{rank}\Sigma \;=\; \max\{i\in\{1\ldots\eta\} \mid \sigma_i \neq 0\} \;\leq\; \eta \tag{1356}$$

- 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

$$\begin{gathered} A \;=\; U\Sigma Q^T \;=\; [\,u_1 \cdots u_\rho\,]\;\Sigma \begin{bmatrix} q_1^T \\ \vdots \\ q_\rho^T \end{bmatrix} \;=\; \sum_{i=1}^{\rho} \sigma_i\, u_i q_i^T \\ U\in\mathbb{R}^{m\times\rho}, \quad \Sigma\in\mathbb{R}^{\rho\times\rho}, \quad Q\in\mathbb{R}^{n\times\rho} \end{gathered} \tag{1357}$$

where the main diagonal of diagonal matrix $\Sigma$ has no 0 entries, and

$$\begin{aligned} \mathcal{R}\{u_i\} &= \mathcal{R}(A) \\ \mathcal{R}\{q_i\} &= \mathcal{R}(A^T) \end{aligned} \tag{1358}$$

### 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 (1355) provides what is called the *full singular value decomposition* of $A\in\mathbb{R}^{m\times n}$ [249, App.A]. Orthonormal matrices $U$ and $Q$ become orthogonal matrices (§B.5):

$$\begin{aligned} \mathcal{R}\{u_i \mid \sigma_i\neq 0\} &= \mathcal{R}(A) \\ \mathcal{R}\{u_i \mid \sigma_i = 0\} &= \mathcal{N}(A^T) \\ \mathcal{R}\{q_i \mid \sigma_i\neq 0\} &= \mathcal{R}(A^T) \\ \mathcal{R}\{q_i \mid \sigma_i = 0\} &= \mathcal{N}(A) \end{aligned} \tag{1359}$$

For any matrix $A$ having rank $\rho$ (= rank $\Sigma$)

$$A = U\Sigma Q^T = [\,u_1 \cdots u_m\,]\;\Sigma\begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} = \sum_{i=1}^{\eta}\sigma_i\, u_i q_i^T$$

$$= \big[\; m\times\rho \;\; \text{basis}\, \mathcal{R}(A) \qquad m\times m-\rho \;\; \text{basis}\, \mathcal{N}(A^T)\,\big] \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \\ & & \end{bmatrix} \begin{bmatrix} \big(n\times\rho \;\; \text{basis}\, \mathcal{R}(A^T)\big)^T \\ \big(n\times n-\rho \;\; \text{basis}\, \mathcal{N}(A)\big)^T \end{bmatrix}$$

$$U\in\mathbb{R}^{m\times m}, \quad \Sigma\in\mathbb{R}^{m\times n}, \quad Q\in\mathbb{R}^{n\times n} \tag{1360}$$

where upper limit of summation $\eta$ is defined in (1351). Matrix $\Sigma$ is no longer necessarily square, now padded with respect to (1352) 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 (1353).

*An important geometrical interpretation of SVD is given in Figure* ***112*** *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* ***112*** *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* `eigshow.m` (see [248]).

*A direct consequence of the geometric interpretation is that the largest singular value* $\sigma_1$ *measures the "magnitude" of* $A$ *(its* 2*-norm)*:

$$\|A\|_2 = \sup_{\|x\|_2=1}\|Ax\|_2 = \sigma_1 \tag{1361}$$

*This means that* $\|A\|_2$ *is the length of the longest principal semiaxis of the ellipse.*

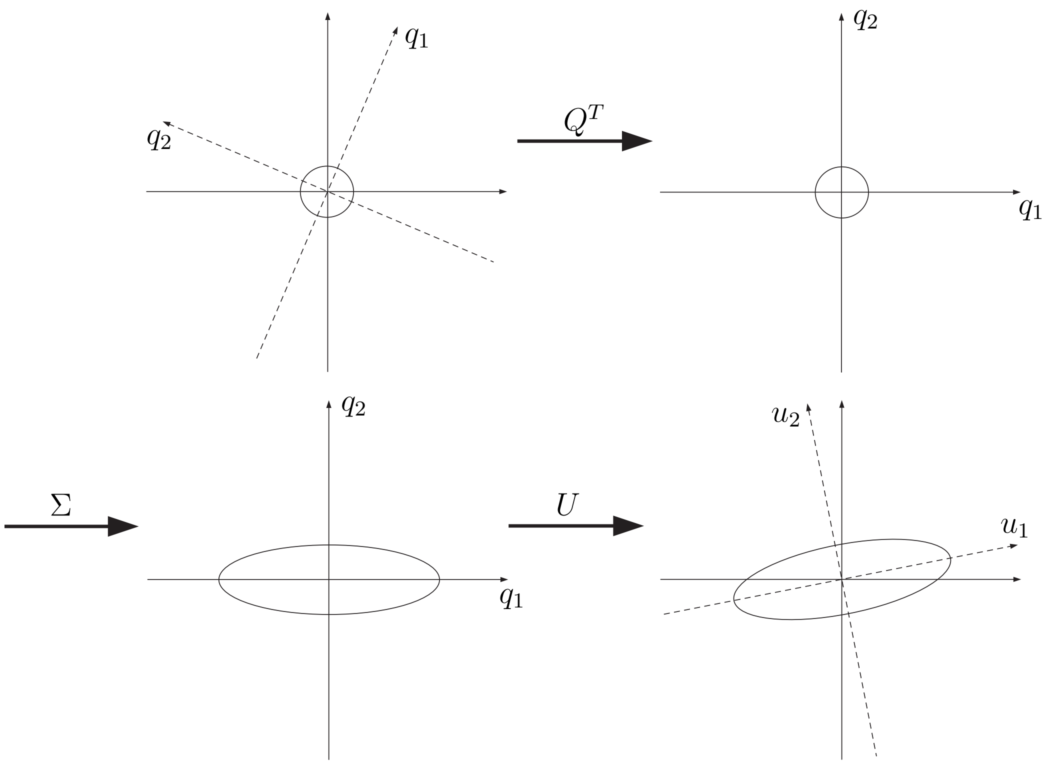



Figure 112: Geometrical interpretation of full SVD [199]: Image of circle $\{x \in \mathbb{R}^{\mathbf{2}} \mid \|x\|_2 = 1\}$ under matrix multiplication $Ax$ is, in general, an ellipse. For the example illustrated, $U \stackrel{\Delta}{=} [\, u_1 \;\; u_2 \,] \in \mathbb{R}^{\mathbf{2} \times \mathbf{2}}$, $Q \stackrel{\Delta}{=} [\, q_1 \;\; q_2 \,] \in \mathbb{R}^{\mathbf{2} \times \mathbf{2}}$.

*Expressions for* $U$ , $Q$ , *and* $\Sigma$ *follow readily from* (1360),

$$AA^TU = U\Sigma\Sigma^T \quad \text{and} \quad A^TAQ = Q\Sigma^T\Sigma \tag{1362}$$

*demonstrating that the columns of* $U$ *are the eigenvectors of* $AA^T$ *and the columns of* $Q$ *are the eigenvectors of* $A^TA$ *.* −Neil Muller *et alii* [199]

## A.6.4 Pseudoinverse by SVD

Matrix pseudoinverse (§E) is nearly synonymous with singular value decomposition because of the elegant expression, given $A = U\Sigma Q^T$

$$A^\dagger = Q\Sigma^{\dagger T}U^T \in \mathbb{R}^{n\times m} \tag{1363}$$

that applies to all three flavors of SVD, where $\Sigma^\dagger$ simply inverts nonzero entries of matrix $\Sigma$ .

Given symmetric matrix $A\in\mathbb{S}^n$ and its diagonalization $A = S\Lambda S^T$ (§A.5.2), its pseudoinverse simply inverts all nonzero eigenvalues;

$$A^\dagger = S\Lambda^\dagger S^T \tag{1364}$$

## A.6.5 SVD of symmetric matrices

**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:

$$\psi(a) \triangleq \left[\lim_{x_i\to a_i}\frac{x_i}{|x_i|} = \left\{\begin{array}{rl} 1\,, & a_i \geq 0 \\ -1\,, & a_i < 0 \end{array}\right. , \quad i=1\ldots n\right] \in \mathbb{R}^n \tag{1365}$$

△

Eigenvalue signs of a symmetric matrix having diagonalization $A = S\Lambda S^T$ (1348) can be absorbed either into real $U$ or real $Q$ from the full SVD; [263, p.34] (*confer* §C.4.2.1)

$$A \;=\; S\Lambda S^T \;=\; S\delta(\psi(\delta(\Lambda)))\,|\Lambda|S^T \triangleq U\,\Sigma Q^T \in \mathbb{S}^n \tag{1366}$$

or

$$A \;=\; S\Lambda S^T \;=\; S|\Lambda|\,\delta(\psi(\delta(\Lambda)))S^T \triangleq U\Sigma\,Q^T \in \mathbb{S}^n \tag{1367}$$

where $\Sigma = |\Lambda|$ , denoting entrywise absolute value of diagonal matrix $\Lambda$ .

## A.7 Zeros

### A.7.1 zero norm

For any given norm, by definition,

$$\|x\|_\ell = 0 \quad \Leftrightarrow \quad x = \mathbf{0} \tag{1368}$$

### 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 $A_{ij} + A_{ji} = 0 \;\; \forall j$ . [200, §1.3.1]

Any symmetric positive semidefinite matrix having a 0 entry on its main diagonal must be $\mathbf{0}$ along the entire row and column to which that 0 entry belongs. [110, §4.2.8] [150, §7.1, prob.2]

### 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 $A \in \mathbb{R}^{m\times n}$

$$\operatorname{rank}(A) + \dim \mathcal{N}(A) = n \tag{1369}$$

by conservation of dimension. [150, §0.4.4]

For any square matrix $A \in \mathbb{R}^{m\times m}$, the number of 0 eigenvalues is at least equal to $\dim \mathcal{N}(A)$

$$\dim \mathcal{N}(A) \;\leq\; \text{number of } 0 \text{ eigenvalues} \;\leq\; m \tag{1370}$$

while the eigenvectors corresponding to those 0 eigenvalues belong to $\mathcal{N}(A)$. [249, §5.1][A.16]

[A.16] We take as given the well-known fact that the number of 0 eigenvalues cannot be less than the dimension of the nullspace. We offer an example of the converse:

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

For diagonalizable matrix $A$ (§A.5), the number of $0$ eigenvalues is precisely $\dim\mathcal{N}(A)$ while the corresponding eigenvectors span $\mathcal{N}(A)$. The real and imaginary parts of the eigenvectors remaining span $\mathcal{R}(A)$.

(TRANSPOSE.)

Likewise, for any matrix $A\in\mathbb{R}^{m\times n}$

$$\operatorname{rank}(A^T) + \dim\mathcal{N}(A^T) = m \tag{1371}$$

For any square $A\in\mathbb{R}^{m\times m}$, the number of $0$ eigenvalues is at least equal to $\dim\mathcal{N}(A^T) = \dim\mathcal{N}(A)$ while the left-eigenvectors (eigenvectors of $A^T$) corresponding to those 0 eigenvalues belong to $\mathcal{N}(A^T)$.

For diagonalizable $A$, the number of $0$ eigenvalues is precisely $\dim\mathcal{N}(A^T)$ while the corresponding left-eigenvectors span $\mathcal{N}(A^T)$. The real and imaginary parts of the left-eigenvectors remaining span $\mathcal{R}(A^T)$. $\diamond$

**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 $m=\operatorname{rank}(A)$ eigenvalues are nonzero. [249, §5.1]

Suppose $\operatorname{rank}(A)<m$. Then $\dim\mathcal{N}(A) = m-\operatorname{rank}(A)$. Thus there is a set of $m-\operatorname{rank}(A)$ linearly independent vectors spanning $\mathcal{N}(A)$. Each of those can be an eigenvector associated with a $0$ eigenvalue because $A$ is diagonalizable $\Leftrightarrow$ $\exists$ $m$ linearly independent eigenvectors. [249, §5.2] Eigenvectors of a real matrix corresponding to 0 eigenvalues must be real.[A.17] Thus $A$ has at least $m-\operatorname{rank}(A)$ eigenvalues equal to 0.

Now suppose $A$ has more than $m-\operatorname{rank}(A)$ eigenvalues equal to 0. Then there are more than $m-\operatorname{rank}(A)$ linearly independent eigenvectors associated with 0 eigenvalues, and each of those eigenvectors must be in $\mathcal{N}(A)$. Thus there are more than $m-\operatorname{rank}(A)$ linearly independent vectors in $\mathcal{N}(A)$; a contradiction.

$\dim\mathcal{N}(A)=2$, $\lambda(A)=[\,0\;0\;0\;1\,]^T$; three eigenvectors in the nullspace but only two are independent. The right-hand side of (1370) is tight for nonzero matrices; *e.g.*, (§B.1) dyad $uv^T\in\mathbb{R}^{m\times m}$ has $m$ 0-eigenvalues when $u\in v^\perp$.

[A.17] Let $*$ denote complex conjugation. Suppose $A=A^*$ and $As_i=\mathbf{0}$. Then $s_i=s_i^*$ $\Rightarrow$ $As_i=As_i^*$ $\Rightarrow$ $As_i^*=\mathbf{0}$. Conversely, $As_i^*=\mathbf{0}$ $\Rightarrow$ $As_i=As_i^*$ $\Rightarrow$ $s_i=s_i^*$.

Therefore diagonalizable $A$ has $\operatorname{rank}(A)$ nonzero eigenvalues and exactly $m-\operatorname{rank}(A)$ 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 (1339), assuming 0 eigenvalues are ordered last,

$$A \;=\; \sum_{i=1}^{m} \lambda_i\, s_i w_i^T \;=\; \sum_{\substack{i=1\\ \lambda_i\neq 0}}^{k\,\leq\, m} \lambda_i\, s_i w_i^T \tag{1372}$$

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 $\operatorname{rank} A = k$, the number of nonzero 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 (1372), we get the partial summation

$$\begin{aligned} \lambda_i\, s_i w_i^T +\, \lambda_i^*\, s_i^* w_i^{*T} \;&=\; 2\operatorname{Re}(\lambda_i\, s_i w_i^T) \\ &=\; 2\big(\operatorname{Re} s_i\, \operatorname{Re}(\lambda_i\, w_i^T) - \operatorname{Im} s_i\, \operatorname{Im}(\lambda_i\, w_i^T)\big) \end{aligned} \tag{1373}$$

where[A.18] $\lambda_i^* \triangleq \lambda_{i+1}$ , $s_i^* \triangleq s_{i+1}$ , and $w_i^* \triangleq w_{i+1}$ . Then (1372) is equivalently written

$$A \;=\; 2 \sum_{\substack{i\\ \lambda\in\mathbb{C}\\ \lambda_i\neq 0}} \operatorname{Re} s_{2i}\, \operatorname{Re}(\lambda_{2i}\, w_{2i}^T) - \operatorname{Im} s_{2i}\, \operatorname{Im}(\lambda_{2i}\, w_{2i}^T) \;+\; \sum_{\substack{j\\ \lambda\in\mathbb{R}\\ \lambda_j\neq 0}} \lambda_j\, s_j w_j^T \tag{1374}$$

The summation (1374) shows: $A$ is a linear combination of real and imaginary parts of its right-eigenvectors corresponding to nonzero eigenvalues. The $k$ vectors $\{\operatorname{Re} s_i\in\mathbb{R}^m,\, \operatorname{Im} s_i\in\mathbb{R}^m \mid \lambda_i\neq 0,\, i\in\{1\ldots m\}\}$ must therefore span the range of diagonalizable matrix $A$ .

The argument is similar regarding the span of the left-eigenvectors. ♦

[A.18] The complex conjugate of $w$ is denoted $w^*$, while its conjugate transpose is denoted by $w^H = w^{*T}$.

### A.7.4 0 trace and matrix product

For $X, A \in \mathbb{S}_+^M$ [27, §2.6.1, exer.2.8] [269, §3.1]

$$\operatorname{tr}(XA) = 0 \;\Leftrightarrow\; XA = AX = \mathbf{0} \tag{1375}$$

**Proof.** ($\Leftarrow$) Suppose $XA = AX = \mathbf{0}$. Then $\operatorname{tr}(XA) = 0$ is obvious.
($\Rightarrow$) Suppose $\operatorname{tr}(XA) = 0$. $\operatorname{tr}(XA) = \operatorname{tr}(\sqrt{A}\,X\sqrt{A})$ 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 feasible positive semidefinite matrix, having all 0 eigenvalues, resides at the origin; (*confer* (1399)) *id est*,

$$\sqrt{A}\,X\sqrt{A} \;=\; \left(\sqrt{X}\sqrt{A}\right)^T \sqrt{X}\sqrt{A} \;=\; \mathbf{0} \tag{1376}$$

implying $\sqrt{X}\sqrt{A} = \mathbf{0}$ which in turn implies $\sqrt{X}(\sqrt{X}\sqrt{A})\sqrt{A} = XA = \mathbf{0}$. Arguing similarly yields $AX = \mathbf{0}$. ◆

Diagonalizable matrices $A$ and $X$ are simultaneously diagonalizable if and only if they are commutative under multiplication; [150, §1.3.12] *id est*, iff they share a complete set of eigenvectors.

#### A.7.4.1 an equivalence in nonisomorphic spaces

Identity (1375) leads to an unusual equivalence relating convex geometry to traditional linear algebra: The convex sets, given $A \succeq 0$

$$\{X \mid \langle X\,,\,A\rangle = 0\} \cap \{X \succeq 0\} \;\equiv\; \{X \mid \mathcal{N}(X) \supseteq \mathcal{R}(A)\} \cap \{X \succeq 0\} \tag{1377}$$

(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,

$$\begin{aligned}\mathbb{S}_c^M &= \{Y \in \mathbb{S}^M \mid Y\mathbf{1} = \mathbf{0}\}\\ &= \{Y \in \mathbb{S}^M \mid \mathcal{N}(Y) \supseteq \mathbf{1}\} \;=\; \{Y \in \mathbb{S}^M \mid \mathcal{R}(Y) \subseteq \mathcal{N}(\mathbf{1}^T)\}\end{aligned} \tag{1766}$$

from which we derive (*confer* (803))

$$\mathbb{S}_c^M \cap \mathbb{S}_+^M \;\equiv\; \{X \succeq 0 \mid \langle X\,,\,\mathbf{1}\mathbf{1}^T\rangle = 0\} \tag{1378}$$

### 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 $A+A^T\succeq 0$)

$$\{x \mid x^TAx=0\} \;=\; \mathcal{N}(A+A^T) \tag{1379}$$

because $\exists\, R \ni A{+}A^T{=}R^TR$ , $\|Rx\|{=}0 \Leftrightarrow Rx{=}\mathbf{0}$, and $\mathcal{N}(A{+}A^T){=}\mathcal{N}(R)$. Then given any particular vector $x_{\rm p}$ , $x_{\rm p}^TAx_{\rm p}{=}0 \;\Leftrightarrow\; x_{\rm p}\in\mathcal{N}(A+A^T)$. For any positive definite matrix $A$ (for $A+A^T\succ 0$)

$$\{x \mid x^TAx=0\} \;=\; \mathbf{0} \tag{1380}$$

Further, [301, §3.2, prob.5]

$$\{x \mid x^TAx=0\} = \mathbb{R}^M \;\Leftrightarrow\; A^T=-A \tag{1381}$$

while

$$\{x \mid x^HAx=0\} = \mathbb{C}^M \;\Leftrightarrow\; A=\mathbf{0} \tag{1382}$$

The positive semidefinite matrix

$$A=\begin{bmatrix}1&2\\0&1\end{bmatrix} \tag{1383}$$

for example, has no nullspace. Yet

$$\{x \mid x^TAx=0\} \;=\; \{x \mid \mathbf{1}^Tx=0\}\subset\mathbb{R}^{\mathbf{2}} \tag{1384}$$

which is the nullspace of the symmetrized matrix. Symmetric matrices are not spared from the excess; *videlicet*,

$$B=\begin{bmatrix}1&2\\2&1\end{bmatrix} \tag{1385}$$

has eigenvalues $\{-1,3\}$, no nullspace, but is zero definite on[A.19]

$$\mathcal{X}\triangleq\{x\in\mathbb{R}^{\mathbf{2}} \mid x_2=(-2\pm\sqrt{3})x_1\} \tag{1386}$$

---

[A.19] These two lines represent the limit in the union of two generally distinct hyperbolae; *id est*, for matrix $B$ and set $\mathcal{X}$ as defined

$$\lim_{\varepsilon\to0^+}\{x\in\mathbb{R}^{\mathbf{2}} \mid x^TBx=\varepsilon\} \;=\; \mathcal{X}$$

**A.7.5.0.1 Proposition.** (Sturm) *Dyad-decompositions.* [254, §5.2]
Let positive semidefinite matrix $X\in\mathbb{S}_+^M$ have rank $\rho$. Then given symmetric matrix $A\in\mathbb{S}^M$, $\langle A\,,X\rangle=0$ if and only if there exists a dyad-decomposition

$$X=\sum_{j=1}^{\rho} x_j x_j^T \tag{1387}$$

satisfying

$$\langle A\,,\,x_j x_j^T\rangle=0 \quad\text{for each and every } j\in\{1\ldots\rho\} \tag{1388}$$

$\diamond$

The dyad-decomposition of $X$ proposed is generally not that obtained from a standard diagonalization by eigen decomposition, unless $\rho=1$ or the given matrix $A$ is diagonalizable simultaneously (§A.7.4) with $X$. That means, elemental dyads in decomposition (1387) constitute a generally nonorthogonal set. Sturm & Zhang give a simple procedure for constructing the dyad-decomposition; (§F.5) 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 $x^Tv=0$;

$$\{x\mid x^Tuv^Tx=0\}\;=\;\{x\mid x^Tu=0\}\cup\{x\mid v^Tx=0\} \tag{1389}$$

*id est*, on $u^\perp\cup v^\perp$. Symmetrizing the dyad does not change the outcome:

$$\{x\mid x^T(uv^T+vu^T)x/2=0\}\;=\;\{x\mid x^Tu=0\}\cup\{x\mid v^Tx=0\} \tag{1390}$$

$\square$

# 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×w *need not equal* w×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.*

−Marie A. Vitulli [277]



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

## B.1 Rank-one matrix (dyad)

Any matrix formed from the unsigned outer product of two vectors,

$$\Psi = uv^T \in \mathbb{R}^{M\times N} \tag{1391}$$

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 $\Psi$. [150, prob.1.4.1] Product $-uv^T$ is a *negative dyad.* For matrix products $AB^T$, in general, we have

$$\mathcal{R}(AB^T) \subseteq \mathcal{R}(A)\;, \qquad \mathcal{N}(AB^T) \supseteq \mathcal{N}(B^T) \tag{1392}$$

with equality when $B=A$ [249, §3.3, §3.6][B.1] or respectively when $B$ is invertible and $\mathcal{N}(A)=\mathbf{0}$. Yet for all nonzero dyads we have

$$\mathcal{R}(uv^T) = \mathcal{R}(u)\;, \qquad \mathcal{N}(uv^T) = \mathcal{N}(v^T) \equiv v^\perp \tag{1393}$$

where $\dim v^\perp = N-1$.

It is obvious a dyad can be $\mathbf{0}$ only when $u$ or $v$ is $\mathbf{0}$;

$$\Psi = uv^T = \mathbf{0} \;\Leftrightarrow\; u=\mathbf{0} \;\;\textbf{or}\;\; v=\mathbf{0} \tag{1394}$$

The matrix 2-norm for $\Psi$ is equivalent to the Frobenius norm;

$$\|\Psi\|_2 \;=\; \|uv^T\|_{\mathrm F} \;=\; \|uv^T\|_2 \;=\; \|u\|\,\|v\| \tag{1395}$$

When $u$ and $v$ are normalized, the pseudoinverse is the transposed dyad. Otherwise,

$$\Psi^\dagger \;=\; (uv^T)^\dagger \;=\; \frac{vu^T}{\|u\|^2\,\|v\|^2} \tag{1396}$$

---

[B.1] **Proof.** $\mathcal{R}(AA^T) \subseteq \mathcal{R}(A)$ is obvious.

$$\begin{aligned}\mathcal{R}(AA^T) \;&=\; \{AA^Ty \mid y\in\mathbb{R}^m\}\\ &\supseteq\; \{AA^Ty \mid A^Ty\in\mathcal{R}(A^T)\} = \mathcal{R}(A) \;\text{ by (120)}\end{aligned} \qquad \blacklozenge$$

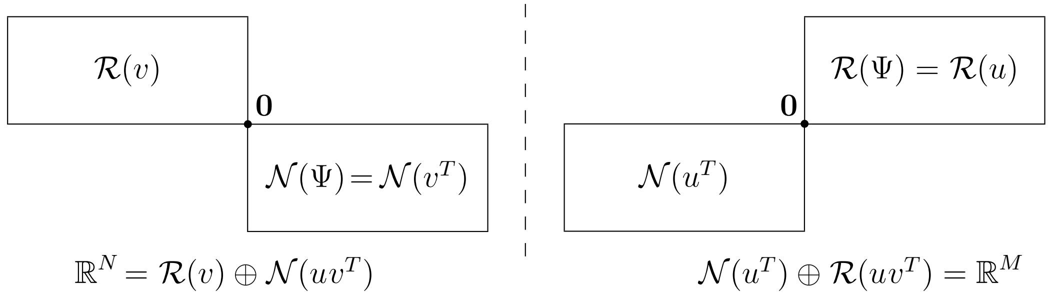



Figure 113: The four fundamental subspaces [251, §3.6] of any dyad $\Psi = uv^T \in \mathbb{R}^{M\times N}$. $\Psi(x) \triangleq uv^T x$ is a linear mapping from $\mathbb{R}^N$ to $\mathbb{R}^M$. The map from $\mathcal{R}(v)$ to $\mathcal{R}(u)$ is bijective. [249, §3.1]

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

$$\lambda = v^T u = \operatorname{tr}(uv^T) \in \mathbb{R} \tag{1397}$$

Determinant is a product of the eigenvalues; so, it is always true that

$$\det \Psi = \det(uv^T) = 0 \tag{1398}$$

When $\lambda = 1$, the square dyad is a nonorthogonal projector projecting on its range ($\Psi^2 = \Psi$, §E.6.2.1); a *projector dyad.* It is quite possible that $u \in v^\perp$ making the remaining eigenvalue instead 0;[B.2] $\lambda = 0$ 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

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \tag{1399}$$

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, $\dim u = M = \dim v = N$ :

[B.2] A dyad is not always diagonalizable (§A.5) because its eigenvectors are not necessarily independent.

**Proof.** Figure **113** shows the four fundamental subspaces for the dyad. Linear operator $\Psi : \mathbb{R}^N \to \mathbb{R}^M$ provides a map between vector spaces that remain distinct when $M = N$;

$$\begin{aligned} u &\in \mathcal{R}(uv^T) \\ u \in \mathcal{N}(uv^T) &\Leftrightarrow v^Tu = 0 \\ \mathcal{R}(uv^T) \cap \mathcal{N}(uv^T) &= \emptyset \end{aligned} \tag{1400}$$

♦

#### B.1.0.1 rank-one modification

If $A \in \mathbb{R}^{N\times N}$ is any nonsingular matrix and $1+v^TA^{-1}u \neq 0$, then [162, App.6] [301, §2.3, prob.16] [103, §4.11.2] (Sherman-Morrison)

$$(A+uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u} \tag{1401}$$

#### B.1.0.2 dyad symmetry

In the specific circumstance that $v = u$, then $uu^T \in \mathbb{R}^{N\times N}$ is symmetric, rank-one, and positive semidefinite having exactly $N-1$ 0-eigenvalues. In fact, (Theorem A.3.1.0.7)

$$uv^T \succeq 0 \;\Leftrightarrow\; v = u \tag{1402}$$

and the remaining eigenvalue is almost always positive;

$$\lambda \;=\; u^Tu \;=\; \operatorname{tr}(uu^T) > 0 \;\text{ unless }\; u = \mathbf{0} \tag{1403}$$

The matrix

$$\begin{bmatrix} \Psi & u \\ u^T & 1 \end{bmatrix} \tag{1404}$$

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 (1391) as encountered in diagonalization and singular value decomposition:

$$\mathcal{R}\left(\sum_{i=1}^k s_i w_i^T\right) = \sum_{i=1}^k \mathcal{R}\big(s_i w_i^T\big) = \sum_{i=1}^k \mathcal{R}(s_i) \;\Leftarrow\; w_i \;\forall\, i \text{ are l.i.} \tag{1405}$$

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 the vector sums are unique (p.676): for $\{w_i\}$ l.i. and $\{s_i\}$ l.i.

$$\mathcal{R}\left(\sum_{i=1}^{k} s_i w_i^T\right) = \mathcal{R}\big(s_1 w_1^T\big) \oplus \ldots \oplus \mathcal{R}\big(s_k w_k^T\big) = \mathcal{R}(s_1) \oplus \ldots \oplus \mathcal{R}(s_k) \quad (1406)$$

**B.1.1.0.1 Definition.** *Linearly independent dyads.* [155, p.29, thm.11] [256, p.2] The set of $k$ dyads

$$\left\{s_i w_i^T \mid i=1\ldots k\right\} \quad (1407)$$

where $s_i \in \mathbb{C}^M$ and $w_i \in \mathbb{C}^N$, is said to be linearly independent iff

$$\operatorname{rank}\left(SW^T \triangleq \sum_{i=1}^{k} s_i w_i^T\right) = k \quad (1408)$$

where $S \triangleq [s_1 \cdots s_k] \in \mathbb{C}^{M\times k}$ and $W \triangleq [w_1 \cdots w_k] \in \mathbb{C}^{N\times k}$. △

As defined, dyad independence does not preclude existence of a nullspace $\mathcal{N}(SW^T)$, nor does it imply $SW^T$ is full-rank. In absence of an assumption of independence, generally, $\operatorname{rank} SW^T \leq k$. 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. ⋄

**Proof.** Linear independence of $k$ dyads is identical to definition (1408).
(⇒) Suppose $\{s_i\}$ and $\{w_i\}$ are each linearly independent sets. Invoking Sylvester's rank inequality, [150, §0.4] [301, §2.4]

$$\operatorname{rank} S + \operatorname{rank} W - k \;\leq\; \operatorname{rank}(SW^T) \;\leq\; \min\{\operatorname{rank} S\,,\ \operatorname{rank} W\} \;\; (\leq\ k) \quad (1409)$$

Then $k \leq \operatorname{rank}(SW^T) \leq k$ that implies the dyads are independent.
(⇐) Conversely, suppose $\operatorname{rank}(SW^T) = k$. Then

$$k \leq \min\{\operatorname{rank} S\,,\ \operatorname{rank} W\} \leq k \quad (1410)$$

implying the vector sets are each independent. ♦

[B.3] Move of range $\mathcal{R}$ to inside the summation depends on linear independence of $\{w_i\}$.

#### B.1.1.1 Biorthogonality condition, Range and Nullspace of Sum

Dyads characterized by a biorthogonality condition $W^T S = I$ are independent; *id est*, for $S \in \mathbb{C}^{M\times k}$ and $W \in \mathbb{C}^{N\times k}$, if $W^T S = I$ then $\operatorname{rank}(SW^T) = k$ by the *linearly independent dyads theorem* because (*confer* §E.1.1)

$$W^T S = I \quad\Leftrightarrow\quad \operatorname{rank} S = \operatorname{rank} W = k \leq M = N \tag{1411}$$

To see that, we need only show: $\mathcal{N}(S) = \mathbf{0} \;\Leftrightarrow\; \exists\, B \;\ni\; BS = I$ .[B.4]
**(⇐)** Assume $BS = I$ . Then $\mathcal{N}(BS) = \mathbf{0} = \{x \mid BSx = \mathbf{0}\} \supseteq \mathcal{N}(S)$. (1392)
**(⇒)** If $\mathcal{N}(S) = \mathbf{0}$ then $S$ must be full-rank skinny-or-square.
$\therefore\; \exists\, A, B, C \;\ni\; \begin{bmatrix} B \\ C \end{bmatrix} [\, S \;\; A \,] = I$ (*id est*, $[\, S \;\; A \,]$ is invertible) $\Rightarrow\; BS = I$ .
Left inverse $B$ is given as $W^T$ here. Because of reciprocity with $S$ , it immediately follows: $\mathcal{N}(W) = \mathbf{0} \;\Leftrightarrow\; \exists\, S \;\ni\; S^T W = I$ . ◆

Dyads produced by diagonalization, for example, are independent because of their inherent biorthogonality. (§A.5.1) The converse is generally false; *id est*, linearly independent dyads are not necessarily biorthogonal.

**B.1.1.1.1 Theorem.** *Nullspace and range of dyad sum.*
Given a sum of dyads represented by $SW^T$ where $S \in \mathbb{C}^{M\times k}$ and $W \in \mathbb{C}^{N\times k}$

$$\begin{aligned} \mathcal{N}(SW^T) = \mathcal{N}(W^T) \quad&\Leftarrow\quad \exists\, B \;\ni\; BS = I \\ \mathcal{R}(SW^T) = \mathcal{R}(S) \quad&\Leftarrow\quad \exists\, Z \;\ni\; W^T Z = I \end{aligned} \tag{1412}$$

⋄

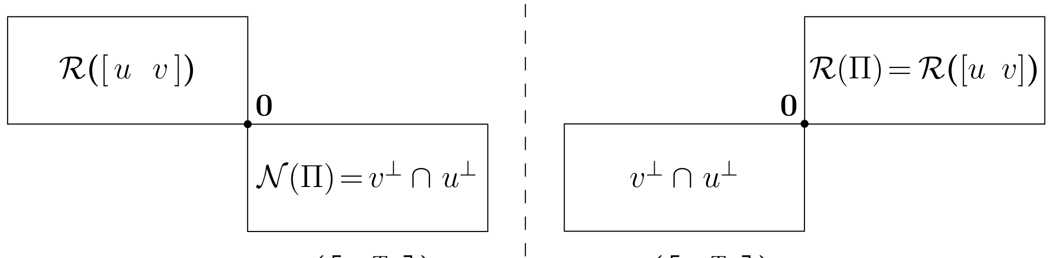
**Proof.** **(⇒)** $\mathcal{N}(SW^T) \supseteq \mathcal{N}(W^T)$ and $\mathcal{R}(SW^T) \subseteq \mathcal{R}(S)$ are obvious.
**(⇐)** Assume the existence of a left inverse $B \in \mathbb{R}^{k\times N}$ and a right inverse $Z \in \mathbb{R}^{N\times k}$ .[B.5]

$$\mathcal{N}(SW^T) \;=\; \{x \mid SW^T x = \mathbf{0}\} \;\subseteq\; \{x \mid BSW^T x = \mathbf{0}\} \;=\; \mathcal{N}(W^T) \tag{1413}$$

$$\mathcal{R}(SW^T) \;=\; \{SW^T x \mid x \in \mathbb{R}^N\} \;\supseteq\; \{SW^T Z y \mid Zy \in \mathbb{R}^N\} \;=\; \mathcal{R}(S) \tag{1414}$$

◆

[B.4] Left inverse is not unique, in general.
[B.5] By counter-example, the theorem's converse cannot be true; *e.g.*, $S = W = [\,\mathbf{1} \;\; \mathbf{0}\,]$ .




$$\mathbb{R}^N=\mathcal{R}([\,u\ \ v\,])\oplus\mathcal{N}\left(\left[\begin{array}{c}v^T\\u^T\end{array}\right]\right) \qquad\qquad \mathcal{N}\left(\left[\begin{array}{c}v^T\\u^T\end{array}\right]\right)\oplus\mathcal{R}([\,u\ \ v\,])=\mathbb{R}^N$$

Figure 114: Four fundamental subspaces [251, §3.6] of a doublet $\Pi=uv^T+vu^T\in\mathbb{S}^N$. $\Pi(x)=(uv^T+vu^T)x$ is a linear bijective mapping from $\mathcal{R}([\,u\ \ v\,])$ to $\mathcal{R}([\,u\ \ v\,])$.

## B.2 Doublet

Consider a sum of two linearly independent square dyads, one a transposition of the other:

$$\Pi \;=\; uv^T+vu^T \;=\; [\,u\ \ v\,]\left[\begin{array}{c}v^T\\u^T\end{array}\right] \;=\; SW^T\in\mathbb{S}^N \tag{1415}$$

where $u\,,v\in\mathbb{R}^N$. Like the dyad, a doublet can be $\mathbf{0}$ only when $u$ or $v$ is $\mathbf{0}$;

$$\Pi=uv^T+vu^T=\mathbf{0}\;\Leftrightarrow\;u=\mathbf{0}\;\;\textbf{or}\;\;v=\mathbf{0} \tag{1416}$$

By assumption of independence, a nonzero doublet has two nonzero eigenvalues

$$\lambda_1\triangleq u^Tv+\|uv^T\|\;, \qquad \lambda_2\triangleq u^Tv-\|uv^T\| \tag{1417}$$

where $\lambda_1>0>\lambda_2$ , with corresponding eigenvectors

$$x_1\triangleq\frac{u}{\|u\|}+\frac{v}{\|v\|}\;, \qquad x_2\triangleq\frac{u}{\|u\|}-\frac{v}{\|v\|} \tag{1418}$$

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.

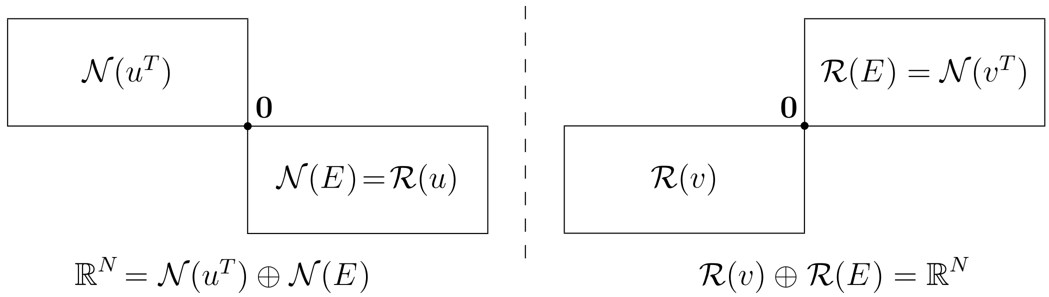



Figure 115: $v^T u = 1/\zeta$ . The four fundamental subspaces [251, §3.6] of elementary matrix $E$ as a linear mapping $E(x) = \left(I - \dfrac{uv^T}{v^T u}\right)x$ .

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

$$\mathcal{R}(\Pi) = \mathcal{R}([\,u \;\; v\,])\ , \qquad \mathcal{N}(\Pi) = v^\perp \cap u^\perp \tag{1419}$$

of respective dimension 2 and $N-2$.

## B.3 Elementary matrix

A matrix of the form

$$E = I - \zeta u v^T \in \mathbb{R}^{N\times N} \tag{1420}$$

where $\zeta \in \mathbb{R}$ is finite and $u, v \in \mathbb{R}^N$, is called an *elementary matrix* or a *rank-one modification of the identity.* [152] 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

$$\lambda = 1 - \zeta v^T u \tag{1421}$$

corresponds to eigenvector $u$ .[B.6] From [162, App.7.A.26] the determinant:

$$\det E \;=\; 1 - \operatorname{tr}\!\left(\zeta u v^T\right) \;=\; \lambda \tag{1422}$$

[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.

If $\lambda \neq 0$ then $E$ is invertible; [103]

$$E^{-1} = I + \frac{\zeta}{\lambda} u v^T \tag{1423}$$

Eigenvectors corresponding to 0 eigenvalues belong to $\mathcal{N}(E)$, and the number of 0 eigenvalues must be at least $\dim \mathcal{N}(E)$ 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 the 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 ($E^2 = E$, §E.1) and $\mathcal{N}(E) = \mathcal{R}(u)$; eigenvector $u$ spans the nullspace when it exists. By conservation of dimension, $\dim \mathcal{R}(E) = N - \dim \mathcal{N}(E)$. It is apparent from (1420) that $v^\perp \subseteq \mathcal{R}(E)$, but $\dim v^\perp = N - 1$. Hence $\mathcal{R}(E) \equiv v^\perp$ when the nullspace exists, and the remaining eigenvectors span it.

In summary, when a nontrivial nullspace of $E$ exists,

$$\mathcal{R}(E) = \mathcal{N}(v^T), \qquad \mathcal{N}(E) = \mathcal{R}(u), \qquad v^T u = 1/\zeta \tag{1424}$$

illustrated in Figure **115**, which is opposite to the assignment of subspaces for a dyad (Figure **113**). Otherwise, $\mathcal{R}(E) = \mathbb{R}^N$.

When $E = E^T$, the spectral norm is

$$\|E\|_2 = \max\{1\,, |\lambda|\} \tag{1425}$$

### B.3.1 Householder matrix

An elementary matrix is called a Householder matrix when it has the defining form, for nonzero vector $u$ [110, §5.1.2] [103, §4.10.1] [249, §7.3] [150, §2.2]

$$H = I - 2\frac{uu^T}{u^T u} \in \mathbb{S}^N \tag{1426}$$

which is a symmetric orthogonal (reflection) matrix $\big(H^{-1} = H^T = H$ (§B.5.2)$\big)$. 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

$$\det H = -1 \tag{1427}$$

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,

$$H^T = H\,, \qquad H^{-1} = H^T\,, \qquad \|H\|_2 = 1\,, \qquad H \nsucceq 0 \tag{1428}$$

For example, the *permutation matrix*

$$\Xi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{1429}$$

is a Householder matrix having $u = [\,0 \quad 1 \quad -1\,]^T/\sqrt{2}$. Not all permutation matrices are Householder matrices, although all permutation matrices are orthogonal matrices [249, §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.*, $\Xi = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ (1507) 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 $\alpha_c$ (§5.5.1.0.1) of some list $X$ to the origin. In place of $X - \alpha_c\mathbf{1}^T$ we may write $XV$ as in (790) where

$$V \;\triangleq\; I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{S}^N \tag{732}$$

is an elementary matrix called the *geometric centering matrix.*

Any elementary matrix in $\mathbb{R}^{N\times N}$ has $N-1$ eigenvalues equal to $1$. For the particular elementary matrix $V$, the $N^{\text{th}}$ 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

$$V\mathbf{1} = \mathbf{0} \tag{1430}$$

the nullspace $\mathcal{N}(V) = \mathcal{R}(\mathbf{1})$ is spanned by the eigenvector $\mathbf{1}$. The remaining eigenvectors span $\mathcal{R}(V) \equiv \mathbf{1}^\perp = \mathcal{N}(\mathbf{1}^T)$ that has dimension $N-1$.

Because

$$V^2 = V \tag{1431}$$

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 $\mathbb{R}^N$

$$V = I - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T \tag{1432}$$

The $\{0, 1\}$ eigenvalues also indicate diagonalizable $V$ is a projection matrix. [301, §4.1, thm.4.1] Symmetry of $V$ denotes orthogonal projection; from (1683),

$$V^T = V\ , \qquad V^\dagger = V\ , \qquad \|V\|_2 = 1\,, \qquad V \succeq 0 \tag{1433}$$

Matrix $V$ is also circulant [118].

**B.4.1.0.1 Example.** *Relationship of auxiliary to Householder matrix.*
Let $H \in \mathbb{S}^N$ be a Householder matrix (1426) defined by

$$u = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 + \sqrt{N} \end{bmatrix} \in \mathbb{R}^N \tag{1434}$$

Then we have [106, §2]

$$V = H \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} H \tag{1435}$$

Let $D \in \mathbb{S}^N_h$ and define

$$-HDH \triangleq -\begin{bmatrix} A & b \\ b^T & c \end{bmatrix} \tag{1436}$$

where $b$ is a vector. Then because $H$ is nonsingular (§A.3.1.0.5) [133, §3]

$$-VDV = -H \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} H \succeq 0 \ \Leftrightarrow\ -A \succeq 0 \tag{1437}$$

and affine dimension is $r = \operatorname{rank} A$ when $D$ is a Euclidean distance matrix.

□

### B.4.2 Schoenberg auxiliary matrix $V_{\mathcal{N}}$

1. $V_{\mathcal{N}} = \frac{1}{\sqrt{2}}\begin{bmatrix} -\mathbf{1}^T \\ I \end{bmatrix} \in \mathbb{R}^{N\times N-1}$
2. $V_{\mathcal{N}}^T\mathbf{1} = \mathbf{0}$
3. $I - e_1\mathbf{1}^T = \begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}$
4. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix} V_{\mathcal{N}} = V_{\mathcal{N}}$
5. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix} V = V$
6. $V\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}$
7. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}$
8. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}^\dagger = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} V$
9. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}^\dagger V = \begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}^\dagger$
10. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}^\dagger = V$
11. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}^\dagger\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix}$
12. $\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} = \begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix}$
13. $\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix}\begin{bmatrix}\mathbf{0} & \sqrt{2}V_{\mathcal{N}}\end{bmatrix} = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix}$
14. $[\,V_{\mathcal{N}} \quad \tfrac{1}{\sqrt{2}}\mathbf{1}\,]^{-1} = \begin{bmatrix} V_{\mathcal{N}}^\dagger \\ \frac{\sqrt{2}}{N}\mathbf{1}^T \end{bmatrix}$
15. $V_{\mathcal{N}}^\dagger = \sqrt{2}\begin{bmatrix}-\frac{1}{N}\mathbf{1} & I - \frac{1}{N}\mathbf{1}\mathbf{1}^T\end{bmatrix} \in \mathbb{R}^{N-1\times N}, \qquad \left(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{S}^{N-1}\right)$
16. $V_{\mathcal{N}}^\dagger\mathbf{1} = \mathbf{0}$
17. $V_{\mathcal{N}}^\dagger V_{\mathcal{N}} = I$

18. $V^T = V = V_{\mathcal{N}} V_{\mathcal{N}}^\dagger = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathbb{S}^N$

19. $-V_{\mathcal{N}}^\dagger(\mathbf{1}\mathbf{1}^T - I)V_{\mathcal{N}} = I\ , \qquad \left(\mathbf{1}\mathbf{1}^T - I \in \mathbb{EDM}^N\right)$

20. $D = [d_{ij}] \in \mathbb{S}_h^N$ (734)

    $\operatorname{tr}(-VDV) = \operatorname{tr}(-VD) = \operatorname{tr}(-V_{\mathcal{N}}^\dagger D V_{\mathcal{N}}) = \frac{1}{N}\mathbf{1}^T D\,\mathbf{1} = \frac{1}{N}\operatorname{tr}(\mathbf{1}\mathbf{1}^T D) = \frac{1}{N}\sum_{i,j} d_{ij}$

    Any elementary matrix $E \in \mathbb{S}^N$ of the particular form

$$E = k_1 I - k_2\,\mathbf{1}\mathbf{1}^T \tag{1438}$$

    where $k_1\,, k_2 \in \mathbb{R}$ ,[B.7] will make $\operatorname{tr}(-ED)$ proportional to $\sum d_{ij}$ .

21. $D = [d_{ij}] \in \mathbb{S}^N$

    $\operatorname{tr}(-VDV) = \frac{1}{N}\sum_{\substack{i,j\\ i\neq j}} d_{ij} - \frac{N-1}{N}\sum_i d_{ii} = \mathbf{1}^T D \mathbf{1}\frac{1}{N} - \operatorname{tr} D$

22. $D = [d_{ij}] \in \mathbb{S}_h^N$

    $\operatorname{tr}(-V_{\mathcal{N}}^T D V_{\mathcal{N}}) = \sum_j d_{1j}$

23. For $Y \in \mathbb{S}^N$

    $V\big(Y - \delta(Y\mathbf{1})\big)V = Y - \delta(Y\mathbf{1})$

### B.4.3 Orthonormal auxiliary matrix $V_{\mathcal{W}}$

The skinny matrix

$$V_{\mathcal{W}} \triangleq \begin{bmatrix} \frac{-1}{\sqrt{N}} & \frac{-1}{\sqrt{N}} & \cdots & \frac{-1}{\sqrt{N}} \\ 1 + \frac{-1}{N+\sqrt{N}} & \frac{-1}{N+\sqrt{N}} & \cdots & \frac{-1}{N+\sqrt{N}} \\ \frac{-1}{N+\sqrt{N}} & \ddots & \ddots & \frac{-1}{N+\sqrt{N}} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{-1}{N+\sqrt{N}} & \frac{-1}{N+\sqrt{N}} & \cdots & 1 + \frac{-1}{N+\sqrt{N}} \end{bmatrix} \in \mathbb{R}^{N\times N-1} \tag{1439}$$

[B.7] If $k_1$ is $1-\rho$ while $k_2$ equals $-\rho \in \mathbb{R}$ , then all eigenvalues of $E$ for $-1/(N-1) < \rho < 1$ are guaranteed positive and therefore $E$ is guaranteed positive definite. [225]

has $\mathcal{R}(V_{\mathcal{W}})=\mathcal{N}(\mathbf{1}^T)$ and orthonormal columns. [4] We defined three auxiliary $V$-matrices: $V$ , $V_{\mathcal{N}}$ (715), and $V_{\mathcal{W}}$ sharing some attributes listed in Table **B.4.4**. For example, $V$ can be expressed

$$V = V_{\mathcal{W}} V_{\mathcal{W}}^T = V_{\mathcal{N}} V_{\mathcal{N}}^\dagger \tag{1440}$$

but $V_{\mathcal{W}}^T V_{\mathcal{W}} = I$ means $V$ is an orthogonal projector (1680) and

$$V_{\mathcal{W}}^\dagger = V_{\mathcal{W}}^T\,, \qquad \|V_{\mathcal{W}}\|_2 = 1\,, \qquad V_{\mathcal{W}}^T \mathbf{1} = \mathbf{0} \tag{1441}$$

### B.4.4 Auxiliary $V$-matrix Table

| | $\dim V$ | $\operatorname{rank} V$ | $\mathcal{R}(V)$ | $\mathcal{N}(V^T)$ | $V^TV$ | $VV^T$ | $VV^\dagger$ |
|---|---|---|---|---|---|---|---|
| $V$ | $N\times N$ | $N-1$ | $\mathcal{N}(\mathbf{1}^T)$ | $\mathcal{R}(\mathbf{1})$ | $V$ | $V$ | $V$ |
| $V_{\mathcal{N}}$ | $N\times(N-1)$ | $N-1$ | $\mathcal{N}(\mathbf{1}^T)$ | $\mathcal{R}(\mathbf{1})$ | $\frac{1}{2}(I+\mathbf{1}\mathbf{1}^T)$ | $\frac{1}{2}\begin{bmatrix} N-1 & -\mathbf{1}^T \\ -\mathbf{1} & I \end{bmatrix}$ | $V$ |
| $V_{\mathcal{W}}$ | $N\times(N-1)$ | $N-1$ | $\mathcal{N}(\mathbf{1}^T)$ | $\mathcal{R}(\mathbf{1})$ | $I$ | $V$ | $V$ |

### B.4.5 More auxiliary matrices

Mathar shows [190, §2] that any elementary matrix (§B.3) of the form

$$V_{\mathcal{M}} = I - b\mathbf{1}^T \in \mathbb{R}^{N\times N} \tag{1442}$$

such that $b^T\mathbf{1}=1$ (*confer* [112, §2]), is an auxiliary $V$-matrix having

$$\begin{aligned} \mathcal{R}(V_{\mathcal{M}}^T) &= \mathcal{N}(b^T)\,, & \mathcal{R}(V_{\mathcal{M}}) &= \mathcal{N}(\mathbf{1}^T) \\ \mathcal{N}(V_{\mathcal{M}}) &= \mathcal{R}(b)\,, & \mathcal{N}(V_{\mathcal{M}}^T) &= \mathcal{R}(\mathbf{1}) \end{aligned} \tag{1443}$$

Given $X\in\mathbb{R}^{n\times N}$, the choice $b=\frac{1}{N}\mathbf{1}$ $(V_{\mathcal{M}}=V)$ minimizes $\|X(I-b\mathbf{1}^T)\|_{\mathrm{F}}$ . [114, §3.2.1]

# B.5 Orthogonal matrix

## B.5.1 Vector rotation

The property $Q^{-1} = Q^T$ completely defines an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ employed to effect vector rotation; [249, §2.6, §3.4] [251, §6.5] [150, §2.1] for $x \in \mathbb{R}^n$

$$\|Qx\| = \|x\| \tag{1444}$$

The orthogonal matrix $Q$ is a normal matrix further characterized:

$$Q^{-1} = Q^T, \qquad \|Q\|_2 = 1 \tag{1445}$$

Applying characterization (1445) to $Q^T$ we see it too is an orthogonal matrix. Hence the rows and columns of $Q$ respectively form an orthonormal set.

All permutation matrices $\Xi$ , for example, are orthogonal matrices. The largest magnitude entry of any orthogonal matrix is 1; for each and every $j \in 1 \ldots n$

$$\begin{array}{c} \|Q(j\,,:)\|_\infty \leq 1 \\ \|Q(:\,,j)\|_\infty \leq 1 \end{array} \tag{1446}$$

Each and every eigenvalue of a (real) orthogonal matrix has magnitude 1

$$\lambda(Q) \in \mathbb{C}^n, \qquad |\lambda(Q)| = \mathbf{1} \tag{1447}$$

while only the identity matrix can be simultaneously positive definite and orthogonal.

A *unitary matrix* is a complex generalization of the orthogonal matrix. The conjugate transpose defines it: $U^{-1} = U^H$. An orthogonal matrix is simply a real unitary matrix.

## B.5.2 Reflection

A matrix for pointwise reflection is defined by imposing symmetry upon the orthogonal matrix; *id est*, a reflection matrix is completely defined by $Q^{-1} = Q^T = Q$. The reflection matrix is an orthogonal matrix, characterized:

$$Q^T = Q\,, \qquad Q^{-1} = Q^T, \qquad \|Q\|_2 = 1 \tag{1448}$$

The Householder matrix (§B.3.1) is an example of a symmetric orthogonal (reflection) matrix.

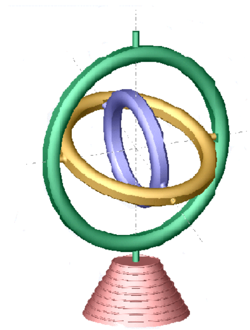


Figure 116: *Gimbal*: a mechanism imparting three degrees of dimensional freedom to a Euclidean body suspended at the device's center. Each ring is free to rotate about one axis. (Courtesy of The MathWorks Inc.)

Reflection matrices have eigenvalues equal to $\pm 1$ and so $\det Q = \pm 1$. 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 [152, §1, prob.44]

$$Q = I - 2P \tag{1449}$$

Yet $P$ is, generally, neither orthogonal or invertible. (§E.3.2)

$$\lambda(Q) \in \mathbb{R}^n\,, \qquad |\lambda(Q)| = \mathbf{1} \tag{1450}$$

Reflection is with respect to $\mathcal{R}(P)^\perp$. 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.5.3 Rotation of range and rowspace

Given orthogonal matrix $Q$, column vectors of a matrix $X$ are simultaneously rotated by the product $QX$. In three dimensions ($X \in \mathbb{R}^{\mathbf{3} \times N}$), the precise meaning of rotation is best illustrated in Figure **116** where the gimbal aids visualization of rotation achievable about the origin.

**B.5.3.0.1 Example.** *One axis of revolution.*
Partition an $n+1$-dimensional Euclidean space $\mathbb{R}^{n+1} \triangleq \begin{bmatrix} \mathbb{R}^n \\ \mathbb{R} \end{bmatrix}$ and define an $n$-dimensional subspace

$$\mathcal{R} \triangleq \{\lambda \in \mathbb{R}^{n+1} \mid \mathbf{1}^T\lambda = 0\} \tag{1451}$$

(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)

$$\measuredangle(\mathbb{R}^n,\, \mathcal{R}) \;=\; \arccos\left(\frac{\langle e_{n+1}\;,\;\mathbf{1}\rangle}{\|e_{n+1}\|\;\|\mathbf{1}\|}\right) \;=\; \arccos\left(\frac{1}{\sqrt{n+1}}\right) \text{ radians} \tag{1452}$$

The vertex-description of the nonnegative orthant in $\mathbb{R}^{n+1}$ is

$$\{[\,e_1\;\; e_2\; \cdots\; e_{n+1}\,]\,a \mid a \succeq 0\} \;=\; \{a \succeq 0\} \;=\; \mathbb{R}^{n+1}_+ \subset \mathbb{R}^{n+1} \tag{1453}$$

Consider rotation of these vertices via orthogonal matrix

$$Q \triangleq [\,\mathbf{1}\tfrac{1}{\sqrt{n+1}} \quad \Xi V_{\mathcal{W}}\,]\Xi \in \mathbb{R}^{n+1\times n+1} \tag{1454}$$

where permutation matrix $\Xi \in \mathbb{S}^{n+1}$ is defined in (1507), and $V_{\mathcal{W}} \in \mathbb{R}^{n+1\times n}$ 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 $\mathcal{R}^\perp = \mathbf{1}$. The rotated standard basis vectors remaining are orthonormal spanning $\mathcal{R}$. □

Another interpretation of product $QX$ is rotation/reflection of $\mathcal{R}(X)$. Rotation of $X$ as in $QXQ^T$ is the simultaneous rotation/reflection of range and rowspace.[B.8]

**Proof.** Any matrix can be expressed as a singular value decomposition $X = U\Sigma W^T$ (1350) where $\delta^2(\Sigma) = \Sigma$, $\mathcal{R}(U) \supseteq \mathcal{R}(X)$, and $\mathcal{R}(W) \supseteq \mathcal{R}(X^T)$. ♦

[B.8] The product $Q^TAQ$ can be regarded as a coordinate transformation; *e.g.*, given linear map $y = Ax : \mathbb{R}^n \to \mathbb{R}^n$ and orthogonal $Q$, the transformation $Qy = AQx$ is a rotation/reflection of the range and rowspace (119) of matrix $A$ where $Qy \in \mathcal{R}(A)$ and $Qx \in \mathcal{R}(A^T)$ (120).

### B.5.4 Matrix rotation

Orthogonal matrices are also employed to rotate/reflect like vectors other matrices: [*sic*] [110, §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 $\mathbb{R}^{n^2}$ because the Frobenius norm is orthogonally invariant (§2.2.1);

$$\|Q^TA\|_F = \sqrt{\operatorname{tr}(A^TQQ^TA)} = \|A\|_F \tag{1455}$$

(likewise for $AQ$). Were $A$ symmetric, such a rotation would depart from $\mathbb{S}^n$. One remedy is to instead form the product $Q^TAQ$ because

$$\|Q^TAQ\|_F = \sqrt{\operatorname{tr}(Q^TA^TQQ^TAQ)} = \|A\|_F \tag{1456}$$

Matrix $A$ is *orthogonally equivalent* to $B$ if $B=S^TAS$ for some orthogonal matrix $S$. Every square matrix, for example, is orthogonally equivalent to a matrix having equal entries along the main diagonal. [150, §2.2, prob.3]

#### B.5.4.1 bijection

Any product of orthogonal matrices $AQ$ remains orthogonal. Given any other dimensionally compatible orthogonal matrix $U$, the mapping $g(A)=U^TAQ$ is a linear bijection on the domain of orthogonal matrices. [175, §2.1]

# Appendix C

# Some analytical optimal results

## C.1 properties of infima

- $$\begin{aligned} \inf \emptyset &\triangleq \quad \infty \\ \sup \emptyset &\triangleq -\infty \end{aligned} \qquad \begin{aligned}(1457)\\(1458)\end{aligned}$$

- Given $f(x) : \mathcal{X} \to \mathbb{R}$ defined on arbitrary set $\mathcal{X}$ [148, §0.1.2]

$$\begin{aligned} \inf_{x \in \mathcal{X}} f(x) &= -\sup_{x \in \mathcal{X}} -f(x) \\ \sup_{x \in \mathcal{X}} f(x) &= -\inf_{x \in \mathcal{X}} -f(x) \end{aligned} \tag{1459}$$

$$\begin{aligned} \arg \inf_{x \in \mathcal{X}} f(x) &= \arg \sup_{x \in \mathcal{X}} -f(x) \\ \arg \sup_{x \in \mathcal{X}} f(x) &= \arg \inf_{x \in \mathcal{X}} -f(x) \end{aligned} \tag{1460}$$

- Given $f(x) : \mathcal{X} \to \mathbb{R}$ and $g(x) : \mathcal{X} \to \mathbb{R}$ defined on arbitrary set $\mathcal{X}$ [148, §0.1.2]

$$\inf_{x \in \mathcal{X}} \big(f(x) + g(x)\big) \geq \inf_{x \in \mathcal{X}} f(x) + \inf_{x \in \mathcal{X}} g(x) \tag{1461}$$

 
Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry*, Meboo Publishing USA, 2005.

- Given $f(x) : \mathcal{X} \cup \mathcal{Y} \rightarrow \mathbb{R}$ and arbitrary sets $\mathcal{X}$ and $\mathcal{Y}$ [148, §0.1.2]

$$\mathcal{X} \subset \mathcal{Y} \Rightarrow \inf_{x\in\mathcal{X}} f(x) \geq \inf_{x\in\mathcal{Y}} f(x) \tag{1462}$$

$$\inf_{x\in\mathcal{X}\cup\mathcal{Y}} f(x) = \min\{\inf_{x\in\mathcal{X}} f(x)\,,\ \inf_{x\in\mathcal{Y}} f(x)\} \tag{1463}$$

$$\inf_{x\in\mathcal{X}\cap\mathcal{Y}} f(x) \geq \max\{\inf_{x\in\mathcal{X}} f(x)\,,\ \inf_{x\in\mathcal{Y}} f(x)\} \tag{1464}$$

- Over some convex set $\mathcal{C}$ given vector constant $y$ or matrix constant $Y$

$$\arg \inf_{x\in\mathcal{C}} \|x-y\|_2 \;=\; \arg \inf_{x\in\mathcal{C}} \|x-y\|_2^2 \tag{1465}$$

$$\arg \inf_{X\in\mathcal{C}} \|X-Y\|_{\mathrm{F}} \;=\; \arg \inf_{X\in\mathcal{C}} \|X-Y\|_{\mathrm{F}}^2 \tag{1466}$$

## C.2 diagonal, trace, singular and eigen values

- For $A\in\mathbb{R}^{m\times n}$ and $\sigma(A)$ denoting its singular values, [46, §A.1.6] [91, §1] (*confer* (36))

$$\begin{aligned}
\sum_i \sigma(A)_i \;=\; \operatorname{tr}\sqrt{A^TA} \;=\; \sup_{\|X\|_2\leq 1} \operatorname{tr}(X^TA) \;=\; & \underset{X\in\mathbb{R}^{m\times n}}{\text{maximize}} \quad \operatorname{tr}(X^TA) \\
& \text{subject to} \quad \begin{bmatrix} I & X \\ X^T & I \end{bmatrix} \succeq 0 \\
=\; \tfrac{1}{2}\, & \underset{X\in\mathbb{S}^m,\; Y\in\mathbb{S}^n}{\text{minimize}} \quad \operatorname{tr} X + \operatorname{tr} Y \\
& \text{subject to} \quad \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \succeq 0
\end{aligned} \tag{1467}$$

- For $X\in\mathbb{S}^m$, $Y\in\mathbb{S}^n$, $A\in\mathcal{C}\subseteq\mathbb{R}^{m\times n}$ for set $\mathcal{C}$ convex, and $\sigma(A)$ denoting the singular values of $A$ [91, §3]

$$\begin{array}{cl}\underset{A}{\text{minimize}} & \sum_i \sigma(A)_i\\ \text{subject to} & A\in\mathcal{C}\end{array} \quad\equiv\quad \begin{array}{cl}\frac{1}{2}\underset{A\,,\,X\,,\,Y}{\text{minimize}} & \operatorname{tr}X+\operatorname{tr}Y\\ \text{subject to} & \begin{bmatrix} X & A\\ A^T & Y\end{bmatrix}\succeq 0\\ & A\in\mathcal{C}\end{array} \tag{1468}$$

- For $A\in\mathbb{S}_+^N$ and $\beta\in\mathbb{R}$

$$\beta\operatorname{tr}A \;=\; \begin{array}{cl}\underset{X\in\,\mathbb{S}^N}{\text{maximize}} & \operatorname{tr}(XA)\\ \text{subject to} & X\preceq\beta I\end{array} \tag{1469}$$

But the following statement is numerically stable, preventing an unbounded solution in direction of a 0 eigenvalue:

$$\begin{array}{cl}\underset{X\in\,\mathbb{S}^N}{\text{maximize}} & \operatorname{sgn}(\beta)\operatorname{tr}(XA)\\ \text{subject to} & X\preceq\ \ |\beta|\,I\\ & X\succeq-|\beta|\,I\end{array} \tag{1470}$$

where $\beta\operatorname{tr}A=\operatorname{tr}(X^\star A)$. If $\beta\geq 0$, then $X\succeq-|\beta|I \leftarrow X\succeq 0$.

- For $A\in\mathbb{S}^N$ having eigenvalues $\lambda(A)\in\mathbb{R}^N$, its smallest and largest eigenvalue is respectively [9, §4.1] [31, §I.6.15] [150, §4.2] [175, §2.1]

$$\min_i\{\lambda(A)_i\} \;=\; \inf_{\|x\|=1} x^TA\,x \;=\; \begin{array}{cl}\underset{X\in\,\mathbb{S}_+^N}{\text{minimize}} & \operatorname{tr}(XA)\\ \text{subject to} & \operatorname{tr}X=1\end{array} \;=\; \begin{array}{cl}\underset{t\in\mathbb{R}}{\text{maximize}} & t\\ \text{subject to} & A\succeq t\,I\end{array} \tag{1471}$$

$$\max_i\{\lambda(A)_i\} \;=\; \sup_{\|x\|=1} x^TA\,x \;=\; \begin{array}{cl}\underset{X\in\,\mathbb{S}_+^N}{\text{maximize}} & \operatorname{tr}(XA)\\ \text{subject to} & \operatorname{tr}X=1\end{array} \;=\; \begin{array}{cl}\underset{t\in\mathbb{R}}{\text{minimize}} & t\\ \text{subject to} & A\preceq t\,I\end{array} \tag{1472}$$

The smallest eigenvalue of any symmetric matrix is always a concave function of its entries, while the largest eigenvalue is always convex. [46, exmp.3.10] For $v_1$ a normalized eigenvector of $A$ corresponding to the largest eigenvalue, and $v_N$ a normalized eigenvector corresponding to the smallest eigenvalue,

$$v_N = \arg \inf_{\|x\|=1} x^T A\, x \tag{1473}$$

$$v_1 = \arg \sup_{\|x\|=1} x^T A\, x \tag{1474}$$

- For $A \in \mathbb{S}^N$ 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) of the boundary of positive semidefinite cone $\mathbb{S}^N_+$: Defining $\lambda_1 \triangleq \max\limits_i \{\lambda(A)_i\}$ and corresponding eigenvector $v_1$

$$\begin{aligned}\underset{x}{\text{minimize}}\ \|xx^T - A\|_F^2 &= \underset{x}{\text{minimize}}\ \operatorname{tr}\big(xx^T(x^Tx) - 2Axx^T + A^TA\big)\\ &= \begin{cases} \|\lambda(A)\|^2\,, & \lambda_1 \leq 0\\ \|\lambda(A)\|^2 - \lambda_1^2\,, & \lambda_1 > 0\end{cases}\end{aligned} \tag{1475}$$

$$\arg \underset{x}{\text{minimize}}\ \|xx^T - A\|_F^2 = \begin{cases} \mathbf{0}\,, & \lambda_1 \leq 0\\ v_1\sqrt{\lambda_1}\,, & \lambda_1 > 0\end{cases} \tag{1476}$$

**Proof.** This is simply the Eckart & Young solution from §7.1.2:

$$x^\star x^{\star T} = \begin{cases} \mathbf{0}\,, & \lambda_1 \leq 0\\ \lambda_1 v_1 v_1^T\,, & \lambda_1 > 0\end{cases} \tag{1477}$$

Given nonincreasingly ordered diagonalization $A = Q\Lambda Q^T$ where $\Lambda = \delta\big(\lambda(A)\big)$ (§A.5), then (1475) has minimum value

$$\underset{x}{\text{minimize}}\ \|xx^T - A\|_F^2 = \begin{cases} \|Q\Lambda Q^T\|_F^2 = \|\delta(\Lambda)\|^2\,, & \lambda_1 \leq 0\\ \left\| Q\left(\begin{bmatrix}\lambda_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{bmatrix} - \Lambda\right)Q^T\right\|_F^2 = \left\|\begin{bmatrix}\lambda_1\\ 0\\ \vdots\\ 0\end{bmatrix} - \delta(\Lambda)\right\|^2, & \lambda_1 > 0\end{cases} \tag{1478}$$

♦

**C.2.0.0.1 Exercise.** *Rank-1 approximation.*
Given symmetric matrix $A \in \mathbb{S}^N$, prove:

$$v_1 = \begin{array}{cl} \arg \underset{x}{\text{minimize}} & \|xx^T - A\|_F^2 \\ \text{subject to} & \|x\| = 1 \end{array} \tag{1479}$$

where $v_1$ is a normalized eigenvector of $A$ corresponding to its largest eigenvalue. ▼

- (Fan) For $B \in \mathbb{S}^N$ whose eigenvalues $\lambda(B) \in \mathbb{R}^N$ are arranged in nonincreasing order, and for $1 \leq k \leq N$ [9, §4.1] [158] [150, §4.3.18] [269, §2] [175, §2.1]

$$\begin{aligned}
\sum_{i=N-k+1}^{N} \lambda(B)_i = \inf_{\substack{U \in \mathbb{R}^{N \times k} \\ U^T U = I}} \operatorname{tr}(UU^T B) &= \begin{array}{cl} \underset{X \in \mathbb{S}_+^N}{\text{minimize}} & \operatorname{tr}(XB) \\ \text{subject to} & X \preceq I \\ & \operatorname{tr} X = k \end{array} && \text{(a)} \\
&= \begin{array}{cl} \underset{\mu \in \mathbb{R}\,,\; Z \in \mathbb{S}_+^N}{\text{maximize}} & (k - N)\mu + \operatorname{tr}(B - Z) \\ \text{subject to} & \mu I + Z \succeq B \end{array} && \text{(b)} \\
\sum_{i=1}^{k} \lambda(B)_i = \sup_{\substack{U \in \mathbb{R}^{N \times k} \\ U^T U = I}} \operatorname{tr}(UU^T B) &= \begin{array}{cl} \underset{X \in \mathbb{S}_+^N}{\text{maximize}} & \operatorname{tr}(XB) \\ \text{subject to} & X \preceq I \\ & \operatorname{tr} X = k \end{array} && \text{(c)} \\
&= \begin{array}{cl} \underset{\mu \in \mathbb{R}\,,\; Z \in \mathbb{S}_+^N}{\text{minimize}} & k\mu + \operatorname{tr} Z \\ \text{subject to} & \mu I + Z \succeq B \end{array} && \text{(d)}
\end{aligned} \tag{1480}$$

Given ordered diagonalization $B = Q\Lambda Q^T$, (§A.5.2) then optimal $U$ for the infimum is $U^\star = Q(:\,, N - k + 1 : N) \in \mathbb{R}^{N \times k}$ whereas $U^\star = Q(:\,, 1 : k) \in \mathbb{R}^{N \times k}$ for the supremum. In both cases, $X^\star = U^\star U^{\star T}$. Optimization (a) searches the convex hull of the outer product $UU^T$ of all $N \times k$ orthonormal matrices. (§2.3.2.0.1)

- For $B\in\mathbb{S}^N$ whose eigenvalues $\lambda(B)\in\mathbb{R}^N$ are arranged in nonincreasing order, and for diagonal matrix $\Upsilon\in\mathbb{S}^k$ whose diagonal entries are arranged in nonincreasing order where $1\leq k\leq N$, we utilize the main-diagonal $\delta$ operator's self-adjointness property (1223): [10, §4.2]

$$\begin{aligned}
\sum_{i=1}^k \Upsilon_{ii}\,\lambda(B)_{N-i+1} &= \inf_{\substack{U\in\mathbb{R}^{N\times k}\\ U^TU=I}} \operatorname{tr}(\Upsilon\, U^TBU) = \inf_{\substack{U\in\mathbb{R}^{N\times k}\\ U^TU=I}} \delta(\Upsilon)^T\delta(U^TBU)\\
&= \begin{array}{cll}\underset{V_i\in\mathbb{S}^N}{\text{minimize}} & \operatorname{tr}\left(B\sum_{i=1}^k(\Upsilon_{ii}-\Upsilon_{i+1,i+1})V_i\right) & \\ \text{subject to} & \operatorname{tr}V_i=i\,, & i=1\ldots k\\ & I\succeq V_i\succeq 0\,, & i=1\ldots k\end{array}
\end{aligned}\tag{1481}$$

where $\Upsilon_{k+1,k+1}\triangleq 0$. We speculate,

$$\sum_{i=1}^k \Upsilon_{ii}\,\lambda(B)_i = \sup_{\substack{U\in\mathbb{R}^{N\times k}\\ U^TU=I}} \operatorname{tr}(\Upsilon\, U^TBU) = \sup_{\substack{U\in\mathbb{R}^{N\times k}\\ U^TU=I}} \delta(\Upsilon)^T\delta(U^TBU)\tag{1482}$$

Alizadeh shows: [9, §4.2]

$$\begin{aligned}
\sum_{i=1}^k \Upsilon_{ii}\,\lambda(B)_i &= \begin{array}{cll}\underset{\mu\in\mathbb{R}^k,\, Z_i\in\mathbb{S}^N}{\text{minimize}} & \sum_{i=1}^k i\mu_i+\operatorname{tr}Z_i & \\ \text{subject to} & \mu_iI+Z_i-(\Upsilon_{ii}-\Upsilon_{i+1,i+1})B\succeq 0\,, & i=1\ldots k\\ & Z_i\succeq 0\,, & i=1\ldots k\end{array}\\
&= \begin{array}{cll}\underset{V_i\in\mathbb{S}^N}{\text{maximize}} & \operatorname{tr}\left(B\sum_{i=1}^k(\Upsilon_{ii}-\Upsilon_{i+1,i+1})V_i\right) & \\ \text{subject to} & \operatorname{tr}V_i=i\,, & i=1\ldots k\\ & I\succeq V_i\succeq 0\,, & i=1\ldots k\end{array}
\end{aligned}\tag{1483}$$

where $\Upsilon_{k+1,k+1}\triangleq 0$.

- The largest eigenvalue magnitude $\mu$ of $A\in\mathbb{S}^N$

$$\max_i\{\,|\lambda(A)_i|\,\} = \begin{array}{cl}\underset{\mu\in\mathbb{R}}{\text{minimize}} & \mu\\ \text{subject to} & -\mu I\preceq A\preceq \mu I\end{array}\tag{1484}$$

is minimized over convex set $\mathcal{C}$ by semidefinite program: (*confer* §7.1.5)

$$\begin{array}{ll}\underset{A}{\text{minimize}} & \|A\|_2 \\ \text{subject to} & A\in\mathcal{C}\end{array} \quad\equiv\quad \begin{array}{ll}\underset{\mu\,,\,A}{\text{minimize}} & \mu \\ \text{subject to} & -\mu I \preceq A \preceq \mu I \\ & A\in\mathcal{C}\end{array} \tag{1485}$$

*id est*,

$$\mu^\star \triangleq \max_i \{\, |\lambda(A^\star)_i| \;,\; i=1\ldots N \} \in \mathbb{R}_+ \tag{1486}$$

- 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: $|\Pi\,\lambda(B)|_1 \geq |\Pi\,\lambda(B)|_2 \geq \cdots \geq |\Pi\,\lambda(B)|_N$ . Then, for $1\leq k\leq N$ [9, §4.3][C.1]

$$\sum_{i=1}^k |\Pi\,\lambda(B)|_i \;=\; \begin{array}{ll}\underset{\mu\in\mathbb{R}\;,\;Z\in\mathbb{S}_+^N}{\text{minimize}} & k\mu + \operatorname{tr} Z \\ \text{subject to} & \mu I + Z + B \succeq 0 \\ & \mu I + Z - B \succeq 0\end{array} \;=\; \begin{array}{ll}\underset{V,W\in\mathbb{S}_+^N}{\text{maximize}} & \langle B\,,\,V-W\rangle \\ \text{subject to} & I\succeq V\,,W \\ & \operatorname{tr}(V+W)=k\end{array} \tag{1487}$$

For diagonal matrix $\Upsilon\in\mathbb{S}^k$ whose diagonal entries are arranged in nonincreasing order where $1\leq k\leq N$

$$\begin{aligned}\sum_{i=1}^k \Upsilon_{ii}|\Pi\,\lambda(B)|_i \;&=\; \begin{array}{lll}\underset{\mu\in\mathbb{R}^k\,,\;Z_i\in\mathbb{S}^N}{\text{minimize}} & \sum\limits_{i=1}^k i\mu_i + \operatorname{tr} Z_i & \\ \text{subject to} & \mu_i I + Z_i + (\Upsilon_{ii}-\Upsilon_{i+1,i+1})B \succeq 0\,, & i=1\ldots k \\ & \mu_i I + Z_i - (\Upsilon_{ii}-\Upsilon_{i+1,i+1})B \succeq 0\,, & i=1\ldots k \\ & Z_i \succeq 0\,, & i=1\ldots k\end{array} \\ &=\; \begin{array}{lll}\underset{V_i\,,W_i\in\mathbb{S}^N}{\text{maximize}} & \operatorname{tr}\Big(B\sum\limits_{i=1}^k(\Upsilon_{ii}-\Upsilon_{i+1,i+1})(V_i-W_i)\Big) & \\ \text{subject to} & \operatorname{tr}(V_i+W_i)=i\;, & i=1\ldots k \\ & I\succeq V_i\succeq 0\,, & i=1\ldots k \\ & I\succeq W_i\succeq 0\,, & i=1\ldots k\end{array}\end{aligned} \tag{1488}$$

where $\Upsilon_{k+1,k+1}\triangleq 0$.

[C.1] We eliminate a redundant positive semidefinite variable from Alizadeh's minimization. There exist typographical errors in [216, (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 $A = W_A \Upsilon W_A^T$ and $B = W_B \Lambda W_B^T$ [149] [175, §2.1]

$$\lambda(A)^T \lambda(B) = \sup_{\substack{U \in \mathbb{R}^{N \times N} \\ U^T U = I}} \operatorname{tr}(A^T U^T B U) \geq \operatorname{tr}(A^T B) \qquad (1506)$$

(*confer* (1511)) where optimal $U$ is

$$U^\star = W_B W_A^T \in \mathbb{R}^{N \times N} \qquad (1503)$$

We can push that upper bound higher using a result in §C.4.2.1:

$$|\lambda(A)|^T |\lambda(B)| = \sup_{\substack{U \in \mathbb{C}^{N \times N} \\ U^H U = I}} \operatorname{Re} \operatorname{tr}(A^T U^H B U) \qquad (1489)$$

For step function $\psi$ as defined in (1365), optimal $U$ becomes

$$U^\star = W_B \sqrt{\delta(\psi(\delta(\Lambda)))}^H \sqrt{\delta(\psi(\delta(\Upsilon)))}\, W_A^T \in \mathbb{C}^{N \times N} \qquad (1490)$$

## C.3 Orthogonal Procrustes problem

Given matrices $A, B \in \mathbb{R}^{n \times N}$, their product having full singular value decomposition (§A.6.3)

$$AB^T \triangleq U \Sigma Q^T \in \mathbb{R}^{n \times n} \qquad (1491)$$

then an optimal solution $R^\star$ to the orthogonal Procrustes problem

$$\begin{array}{ll} \underset{R}{\text{minimize}} & \|A - R^T B\|_{\mathrm{F}} \\ \text{subject to} & R^T = R^{-1} \end{array} \qquad (1492)$$

maximizes $\operatorname{tr}(A^T R^T B)$ over the nonconvex manifold of orthogonal matrices: [150, §7.4.8]

$$R^\star = Q U^T \in \mathbb{R}^{n \times n} \qquad (1493)$$

A necessary and sufficient condition for optimality

$$AB^T R^\star \succeq 0 \qquad (1494)$$

holds whenever $R^\star$ is an orthogonal matrix. [114, §4]

Solution to problem (1492) can reveal rotation/reflection (§5.5.2, §B.5) of one list in the columns of $A$ with respect to another list $B$. Solution is unique if $\operatorname{rank} BV_{\mathcal{N}} = n$. [77, §2.4.1] The optimal value for objective of minimization is

$$\begin{aligned}\operatorname{tr}\left(A^TA + B^TB - 2AB^TR^\star\right) &= \operatorname{tr}(A^TA) + \operatorname{tr}(B^TB) - 2\operatorname{tr}(U\Sigma U^T)\\ &= \|A\|_F^2 + \|B\|_F^2 - 2\delta(\Sigma)^T\mathbf{1}\end{aligned} \tag{1495}$$

while the optimal value for corresponding trace maximization is

$$\sup_{R^T=R^{-1}} \operatorname{tr}(A^TR^TB) = \operatorname{tr}(A^TR^{\star T}B) = \delta(\Sigma)^T\mathbf{1} \geq \operatorname{tr}(A^TB) \tag{1496}$$

The same optimal solution $R^\star$ solves

$$\begin{array}{cl}\underset{R}{\text{maximize}} & \|A + R^TB\|_F\\ \text{subject to} & R^T = R^{-1}\end{array} \tag{1497}$$

### C.3.1 Effect of translation

Consider the impact of dc offset in known lists $A, B \in \mathbb{R}^{n\times N}$ on problem (1492). 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

$$\begin{array}{cl}\underset{R}{\text{minimize}} & \|AV - R^TBV\|_F\\ \text{subject to} & R^T = R^{-1}\end{array} \tag{1498}$$

where $V \in \mathbb{S}^N$ is the geometric centering matrix (§B.4.1). Now we define the full singular value decomposition

$$AVB^T \triangleq U\Sigma Q^T \in \mathbb{R}^{n\times n} \tag{1499}$$

and an optimal rotation matrix

$$R^\star = QU^T \in \mathbb{R}^{n\times n} \qquad (1493)$$

The desired result is an optimally rotated offset list

$$R^{\star T}BV + A(I - V) \approx A \tag{1500}$$

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 $R^{\star T}B = A$ or $B\mathbf{1} = A\mathbf{1} = \mathbf{0}$, this result (1500) reduces to $R^{\star T}B \approx A$.

#### C.3.1.1 Translation of extended list

Suppose an optimal rotation matrix $R^\star\in\mathbb{R}^{n\times n}$ 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 (1500) makes $\mathbf{1}^T$ of compatible dimension;

$$R^{\star T}[\,C-B\mathbf{1}\mathbf{1}^T\tfrac{1}{N}\quad BV\,]\;+\;A\mathbf{1}\mathbf{1}^T\tfrac{1}{N} \tag{1501}$$

*id est*, $C-B\mathbf{1}\mathbf{1}^T\frac{1}{N}\in\mathbb{R}^{n\times M}$ and $A\mathbf{1}\mathbf{1}^T\frac{1}{N}\in\mathbb{R}^{n\times M+N}$.

## C.4 Two-sided orthogonal Procrustes

#### C.4.0.1 Minimization

Given symmetric $A\,,B\in\mathbb{S}^N$, each having diagonalization (§A.5.2)

$$A\triangleq Q_A\Lambda_A Q_A^T\;,\qquad B\triangleq Q_B\Lambda_B Q_B^T \tag{1502}$$

where eigenvalues are arranged in their respective diagonal matrix $\Lambda$ in nonincreasing order, then an optimal solution [86]

$$R^\star = Q_B Q_A^T\in\mathbb{R}^{N\times N} \tag{1503}$$

to the two-sided orthogonal Procrustes problem

$$\begin{array}{cl}\underset{R}{\text{minimize}} & \|A-R^TBR\|_\mathrm{F}\\ \text{subject to} & R^T=R^{-1}\end{array}\;=\;\begin{array}{cl}\underset{R}{\text{minimize}} & \operatorname{tr}\big(A^TA-2A^TR^TBR+B^TB\big)\\ \text{subject to} & R^T=R^{-1}\end{array} \tag{1504}$$

maximizes $\operatorname{tr}(A^TR^TBR)$ over the nonconvex manifold of orthogonal matrices. Optimal product $R^{\star T}BR^\star$ has the eigenvectors of $A$ but the eigenvalues of $B$. [114, §7.5.1] The optimal value for the objective of minimization is, by (40)

$$\|Q_A\Lambda_AQ_A^T-R^{\star T}Q_B\Lambda_BQ_B^TR^\star\|_\mathrm{F}\;=\;\|Q_A(\Lambda_A-\Lambda_B)Q_A^T\|_\mathrm{F}\;=\;\|\Lambda_A-\Lambda_B\|_\mathrm{F} \tag{1505}$$

while the corresponding trace maximization has optimal value

$$\sup_{R^T=R^{-1}}\operatorname{tr}(A^TR^TBR)\;=\;\operatorname{tr}(A^TR^{\star T}BR^\star)\;=\;\operatorname{tr}(\Lambda_A\Lambda_B)\;\geq\;\operatorname{tr}(A^TB) \tag{1506}$$

#### C.4.0.2 Maximization

Any permutation matrix is an orthogonal matrix. Defining a row- and column-swapping permutation matrix (a reflection matrix, B.5.2)

$$\Xi \;=\; \Xi^T \;=\; \begin{bmatrix} \mathbf{0} & & & 1 \\ & & \cdot & \\ & \cdot & & \\ & 1 & & \\ 1 & & & \mathbf{0} \end{bmatrix} \tag{1507}$$

then an optimal solution $R^\star$ to the maximization problem [*sic*]

$$\begin{array}{cl} \underset{R}{\text{maximize}} & \|A - R^T B R\|_{\mathrm F} \\ \text{subject to} & R^T = R^{-1} \end{array} \tag{1508}$$

minimizes $\operatorname{tr}(A^T R^T B R)$ : [149] [175, §2.1]

$$R^\star = Q_B \Xi Q_A^T \in \mathbb{R}^{N\times N} \tag{1509}$$

The optimal value for the objective of maximization is

$$\begin{aligned} \|Q_A\Lambda_A Q_A^T - R^{\star T} Q_B \Lambda_B Q_B^T R^\star\|_{\mathrm F} \;&=\; \|Q_A\Lambda_A Q_A^T - Q_A \Xi^T \Lambda_B \Xi Q_A^T\|_{\mathrm F} \\ &=\; \|\Lambda_A - \Xi\Lambda_B\Xi\|_{\mathrm F} \end{aligned} \tag{1510}$$

while the corresponding trace minimization has optimal value

$$\inf_{R^T=R^{-1}} \operatorname{tr}(A^T R^T B R) \;=\; \operatorname{tr}(A^T R^{\star T} B R^\star) \;=\; \operatorname{tr}(\Lambda_A \Xi \Lambda_B \Xi) \tag{1511}$$

### C.4.1 Procrustes' relation to linear programming

Although these two-sided Procrustes problems are nonconvex, a connection with linear programming [64] was discovered by Anstreicher & Wolkowicz [10, §3] [175, §2.1]: Given $A\,, B \in \mathbb{S}^N$, this semidefinite program in $S$ and $T$

$$\begin{array}{clcll} \underset{R}{\text{minimize}} & \operatorname{tr}(A^T R^T B R) & = & \underset{S\,,\,T\in\mathbb{S}^N}{\text{maximize}} & \operatorname{tr}(S+T) \\ \text{subject to} & R^T = R^{-1} & & \text{subject to} & A^T\otimes B - I\otimes S - T\otimes I \succeq 0 \end{array} \tag{1512}$$

(where $\otimes$ signifies Kronecker product (§D.1.2.1)) has optimal objective value (1511). These two problems are strong duals (§2.13.1.0.3). Given ordered diagonalizations (1502), make the observation:

$$\inf_R \operatorname{tr}(A^T R^T B R) = \inf_{\hat{R}} \operatorname{tr}(\Lambda_A \hat{R}^T \Lambda_B \hat{R}) \tag{1513}$$

because $\hat{R} \triangleq Q_B^T R Q_A$ 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;

$$\begin{array}{cl}\underset{S,T\in\mathbb{S}^N}{\text{maximize}} & \mathbf{1}^T\delta(S+T)\\ \text{subject to} & \delta(\Lambda_A\otimes(\Xi\Lambda_B\Xi) - I\otimes S - T\otimes I)\succeq 0\end{array} \tag{1514}$$

a linear program in $\delta(S)$ and $\delta(T)$ having the same optimal objective value as the semidefinite program (1512).

We relate their results to Procrustes problem (1504) by manipulating signs (1459) and permuting eigenvalues:

$$\begin{array}{cl}\underset{R}{\text{maximize}} & \operatorname{tr}(A^T R^T B R)\\ \text{subject to} & R^T = R^{-1}\end{array} = \begin{array}{cl}\underset{S\,,\,T\in\mathbb{S}^N}{\text{minimize}} & \mathbf{1}^T\delta(S+T)\\ \text{subject to} & \delta(I\otimes S + T\otimes I - \Lambda_A\otimes\Lambda_B)\succeq 0\end{array}$$

$$= \begin{array}{cl}\underset{S\,,\,T\in\mathbb{S}^N}{\text{minimize}} & \operatorname{tr}(S+T)\\ \text{subject to} & I\otimes S + T\otimes I - A^T\otimes B\succeq 0\end{array} \tag{1515}$$

This formulation has optimal objective value identical to that in (1506).

### 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 (1506) a little further: Given real matrices $A\,,B$ each having full singular value decomposition (§A.6.3)

$$A \triangleq U_A \Sigma_A Q_A^T \in \mathbb{R}^{m\times n}\,, \qquad B \triangleq U_B \Sigma_B Q_B^T \in \mathbb{R}^{m\times n} \tag{1516}$$

then a well-known optimal solution $R^\star$, $S^\star$ to the problem

$$\begin{array}{cl}\underset{R\,,\,S}{\text{minimize}} & \|A - SBR\|_{\mathrm F}\\ \text{subject to} & R^H = R^{-1}\\ & S^H = S^{-1}\end{array} \tag{1517}$$

maximizes $\operatorname{Re}\operatorname{tr}(A^TSBR)$ : [236] [213] [38] [144] optimal orthogonal matrices

$$S^\star = U_A U_B^H \in \mathbb{R}^{m\times m}\;, \qquad R^\star = Q_B Q_A^H \in \mathbb{R}^{n\times n} \tag{1518}$$

[*sic*] are not necessarily unique [150, §7.4.13] because the feasible set is not convex. The optimal value for the objective of minimization is, by (40)

$$\|U_A\Sigma_A Q_A^H - S^\star U_B\Sigma_B Q_B^H R^\star\|_{\mathrm F} \;=\; \|U_A(\Sigma_A-\Sigma_B)Q_A^H\|_{\mathrm F} \;=\; \|\Sigma_A-\Sigma_B\|_{\mathrm F} \tag{1519}$$

while the corresponding trace maximization has optimal value [31, §III.6.12]

$$\sup_{\substack{R^H=R^{-1}\\ S^H=S^{-1}}} |\operatorname{tr}(A^TSBR)| \;=\; \sup_{\substack{R^H=R^{-1}\\ S^H=S^{-1}}} \operatorname{Re}\operatorname{tr}(A^TSBR) \;=\; \operatorname{Re}\operatorname{tr}(A^TS^\star BR^\star) \;=\; \operatorname{tr}(\Sigma_A^T\Sigma_B) \;\geq\; \operatorname{tr}(A^TB) \tag{1520}$$

for which it is necessary

$$A^TS^\star BR^\star \succeq 0\;, \qquad BR^\star A^TS^\star \succeq 0 \tag{1521}$$

The lower bound on inner product of singular values in (1520) 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 (§B.5.1), the upper bound on trace (1506) is thereby raised: Suppose we are given diagonalizations for (real) symmetric $A\,,B$ (§A.5)

$$A = W_A\Upsilon W_A^T \in \mathbb{S}^n\,, \qquad \delta(\Upsilon)\in\mathcal{K}_{\mathcal M} \tag{1522}$$

$$B = W_B\Lambda W_B^T \in \mathbb{S}^n\,, \qquad \delta(\Lambda)\in\mathcal{K}_{\mathcal M} \tag{1523}$$

having their respective eigenvalues in diagonal matrices $\Upsilon\,,\Lambda\in\mathbb{S}^n$ arranged in nonincreasing order (membership to the monotone cone $\mathcal{K}_{\mathcal M}$ (377)). Then by splitting eigenvalue signs, we invent a symmetric SVD-like decomposition

$$A \triangleq U_A\Sigma_A Q_A^H \in \mathbb{S}^n\;, \qquad B \triangleq U_B\Sigma_B Q_B^H \in \mathbb{S}^n \tag{1524}$$

where $U_A\,, U_B\,, Q_A\,, Q_B \in \mathbb{C}^{n\times n}$ are unitary matrices defined by (*confer* §A.6.5)

$$U_A \triangleq W_A\sqrt{\delta(\psi(\delta(\Upsilon)))}\;, \quad Q_A \triangleq W_A\sqrt{\delta(\psi(\delta(\Upsilon)))}^H, \quad \Sigma_A = |\Upsilon| \tag{1525}$$

$$U_B \triangleq W_B\sqrt{\delta(\psi(\delta(\Lambda)))}\;, \quad Q_B \triangleq W_B\sqrt{\delta(\psi(\delta(\Lambda)))}^H, \quad \Sigma_B = |\Lambda| \tag{1526}$$

where step function $\psi$ is defined in (1365). In this circumstance,

$$S^\star = U_A U_B^H = R^{\star T} \in \mathbb{C}^{n\times n} \tag{1527}$$

optimal matrices (1518) now unitary are related by transposition. The optimal value of objective (1519) is

$$\|U_A \Sigma_A Q_A^H - S^\star U_B \Sigma_B Q_B^H R^\star\|_{\mathrm F} = \|\,|\Upsilon| - |\Lambda|\,\|_{\mathrm F} \tag{1528}$$

while the corresponding optimal value of trace maximization (1520) is

$$\sup_{\substack{R^H = R^{-1}\\ S^H = S^{-1}}} \operatorname{Re}\operatorname{tr}(A^T S B R) = \operatorname{tr}\big(|\Upsilon|\,|\Lambda|\big) \tag{1529}$$

#### C.4.2.2 Diagonal matrices

Now suppose $A$ and $B$ are diagonal matrices

$$A = \Upsilon = \delta^2(\Upsilon) \in \mathbb{S}^n\,, \quad \delta(\Upsilon) \in \mathcal{K}_{\mathcal M} \tag{1530}$$

$$B = \Lambda = \delta^2(\Lambda) \in \mathbb{S}^n\,, \quad \delta(\Lambda) \in \mathcal{K}_{\mathcal M} \tag{1531}$$

both having their respective main diagonal entries arranged in nonincreasing order:

$$\begin{array}{cl}\underset{R\,,\,S}{\text{minimize}} & \|\Upsilon - S\Lambda R\|_{\mathrm F}\\ \text{subject to} & R^H = R^{-1}\\ & S^H = S^{-1}\end{array} \tag{1532}$$

Then we have a symmetric decomposition from unitary matrices as in (1524) where

$$U_A \triangleq \sqrt{\delta(\psi(\delta(\Upsilon)))}\,, \quad Q_A \triangleq \sqrt{\delta(\psi(\delta(\Upsilon)))}^{\,H}, \quad \Sigma_A = |\Upsilon| \tag{1533}$$

$$U_B \triangleq \sqrt{\delta(\psi(\delta(\Lambda)))}\,, \quad Q_B \triangleq \sqrt{\delta(\psi(\delta(\Lambda)))}^{\,H}, \quad \Sigma_B = |\Lambda| \tag{1534}$$

Procrustes solution (1518) again sees the transposition relationship

$$S^\star = U_A U_B^H = R^{\star T} \in \mathbb{C}^{n\times n} \qquad (1527)$$

but both optimal unitary matrices are now themselves diagonal. So,

$$S^\star \Lambda R^\star = \delta(\psi(\delta(\Upsilon)))\Lambda\delta(\psi(\delta(\Lambda))) = \delta(\psi(\delta(\Upsilon)))|\Lambda| \tag{1535}$$

# Appendix D

# Matrix calculus

*From too much study, and from extreme passion, cometh madnesse.*

−Isaac Newton [105, §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 domain is defined

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_K} \end{bmatrix} \in \mathbb{R}^K \tag{1536}$$

while the second-order gradient of the twice differentiable real function with respect to its vector domain is traditionally called the *Hessian*;

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial^2 x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_K} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial^2 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_K \partial x_1} & \frac{\partial^2 f(x)}{\partial x_K \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial^2 x_K} \end{bmatrix} \in \mathbb{S}^K \tag{1537}$$



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

The gradient of vector-valued function $v(x) : \mathbb{R} \to \mathbb{R}^N$ on real domain is a row-vector

$$\nabla v(x) \triangleq \begin{bmatrix} \frac{\partial v_1(x)}{\partial x} & \frac{\partial v_2(x)}{\partial x} & \cdots & \frac{\partial v_N(x)}{\partial x} \end{bmatrix} \in \mathbb{R}^N \tag{1538}$$

while the second-order gradient is

$$\nabla^2 v(x) \triangleq \begin{bmatrix} \frac{\partial^2 v_1(x)}{\partial x^2} & \frac{\partial^2 v_2(x)}{\partial x^2} & \cdots & \frac{\partial^2 v_N(x)}{\partial x^2} \end{bmatrix} \in \mathbb{R}^N \tag{1539}$$

Gradient of vector-valued function $h(x) : \mathbb{R}^K \to \mathbb{R}^N$ on vector domain is

$$\begin{aligned} \nabla h(x) &\triangleq \begin{bmatrix} \frac{\partial h_1(x)}{\partial x_1} & \frac{\partial h_2(x)}{\partial x_1} & \cdots & \frac{\partial h_N(x)}{\partial x_1} \\ \frac{\partial h_1(x)}{\partial x_2} & \frac{\partial h_2(x)}{\partial x_2} & \cdots & \frac{\partial h_N(x)}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial h_1(x)}{\partial x_K} & \frac{\partial h_2(x)}{\partial x_K} & \cdots & \frac{\partial h_N(x)}{\partial x_K} \end{bmatrix} \\ &= [\, \nabla h_1(x) \;\; \nabla h_2(x) \;\; \cdots \;\; \nabla h_N(x) \,] \in \mathbb{R}^{K \times N} \end{aligned} \tag{1540}$$

while the second-order gradient has a three-dimensional representation dubbed *cubix* ;[D.1]

$$\begin{aligned} \nabla^2 h(x) &\triangleq \begin{bmatrix} \nabla \frac{\partial h_1(x)}{\partial x_1} & \nabla \frac{\partial h_2(x)}{\partial x_1} & \cdots & \nabla \frac{\partial h_N(x)}{\partial x_1} \\ \nabla \frac{\partial h_1(x)}{\partial x_2} & \nabla \frac{\partial h_2(x)}{\partial x_2} & \cdots & \nabla \frac{\partial h_N(x)}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \nabla \frac{\partial h_1(x)}{\partial x_K} & \nabla \frac{\partial h_2(x)}{\partial x_K} & \cdots & \nabla \frac{\partial h_N(x)}{\partial x_K} \end{bmatrix} \\ &= [\, \nabla^2 h_1(x) \;\; \nabla^2 h_2(x) \;\; \cdots \;\; \nabla^2 h_N(x) \,] \in \mathbb{R}^{K \times N \times K} \end{aligned} \tag{1541}$$

where the gradient of each real entry is with respect to vector $x$ as in (1536).

[D.1] The word *matrix* comes from the Latin for *womb* ; related to the prefix *matri-* derived from *mater* meaning *mother*.

The gradient of real function $g(X) : \mathbb{R}^{K\times L} \rightarrow \mathbb{R}$ on matrix domain is

$$\nabla g(X) \triangleq \begin{bmatrix} \frac{\partial g(X)}{\partial X_{11}} & \frac{\partial g(X)}{\partial X_{12}} & \cdots & \frac{\partial g(X)}{\partial X_{1L}} \\ \frac{\partial g(X)}{\partial X_{21}} & \frac{\partial g(X)}{\partial X_{22}} & \cdots & \frac{\partial g(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g(X)}{\partial X_{K1}} & \frac{\partial g(X)}{\partial X_{K2}} & \cdots & \frac{\partial g(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K\times L}$$
$$= \begin{bmatrix} \nabla_{X(:,1)}\, g(X) & & & \\ & \nabla_{X(:,2)}\, g(X) & & \\ & & \ddots & \\ & & & \nabla_{X(:,L)}\, g(X) \end{bmatrix} \in \mathbb{R}^{K\times 1\times L} \tag{1542}$$

where the gradient $\nabla_{X(:,i)}$ is with respect to the $i^{\text{th}}$ column of $X$. The strange appearance of (1542) in $\mathbb{R}^{K\times 1\times L}$ is meant to suggest a third dimension perpendicular to the page (not a diagonal matrix). The second-order gradient has representation

$$\nabla^2 g(X) \triangleq \begin{bmatrix} \nabla\frac{\partial g(X)}{\partial X_{11}} & \nabla\frac{\partial g(X)}{\partial X_{12}} & \cdots & \nabla\frac{\partial g(X)}{\partial X_{1L}} \\ \nabla\frac{\partial g(X)}{\partial X_{21}} & \nabla\frac{\partial g(X)}{\partial X_{22}} & \cdots & \nabla\frac{\partial g(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \nabla\frac{\partial g(X)}{\partial X_{K1}} & \nabla\frac{\partial g(X)}{\partial X_{K2}} & \cdots & \nabla\frac{\partial g(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K\times L\times K\times L}$$
$$= \begin{bmatrix} \nabla\nabla_{X(:,1)}\, g(X) & & & \\ & \nabla\nabla_{X(:,2)}\, g(X) & & \\ & & \ddots & \\ & & & \nabla\nabla_{X(:,L)}\, g(X) \end{bmatrix} \in \mathbb{R}^{K\times 1\times L\times K\times L} \tag{1543}$$

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

$$\nabla g(X) \triangleq \begin{bmatrix} \nabla_{X(:,1)}\, g_1(X) & \nabla_{X(:,1)}\, g_2(X) & \cdots & \nabla_{X(:,1)}\, g_N(X) \\ \nabla_{X(:,2)}\, g_1(X) & \nabla_{X(:,2)}\, g_2(X) & \cdots & \nabla_{X(:,2)}\, g_N(X) \\ \ddots & \ddots & & \ddots \\ \nabla_{X(:,L)}\, g_1(X) & \nabla_{X(:,L)}\, g_2(X) & \cdots & \nabla_{X(:,L)}\, g_N(X) \end{bmatrix}$$

$$= [\, \nabla g_1(X) \;\; \nabla g_2(X) \;\; \cdots \;\; \nabla g_N(X)\,] \in \mathbb{R}^{K\times N\times L} \tag{1544}$$

while the second-order gradient has a five-dimensional representation;

$$\nabla^2 g(X) \triangleq \begin{bmatrix} \nabla\nabla_{X(:,1)}\, g_1(X) & \nabla\nabla_{X(:,1)}\, g_2(X) & \cdots & \nabla\nabla_{X(:,1)}\, g_N(X) \\ \nabla\nabla_{X(:,2)}\, g_1(X) & \nabla\nabla_{X(:,2)}\, g_2(X) & \cdots & \nabla\nabla_{X(:,2)}\, g_N(X) \\ \ddots & \ddots & & \ddots \\ \nabla\nabla_{X(:,L)}\, g_1(X) & \nabla\nabla_{X(:,L)}\, g_2(X) & \cdots & \nabla\nabla_{X(:,L)}\, g_N(X) \end{bmatrix}$$

$$= [\, \nabla^2 g_1(X) \;\; \nabla^2 g_2(X) \;\; \cdots \;\; \nabla^2 g_N(X)\,] \in \mathbb{R}^{K\times N\times L\times K\times L} \tag{1545}$$

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 *quartix*

$$\nabla g(X) \triangleq \begin{bmatrix} \nabla g_{11}(X) & \nabla g_{12}(X) & \cdots & \nabla g_{1N}(X) \\ \nabla g_{21}(X) & \nabla g_{22}(X) & \cdots & \nabla g_{2N}(X) \\ \vdots & \vdots & & \vdots \\ \nabla g_{M1}(X) & \nabla g_{M2}(X) & \cdots & \nabla g_{MN}(X) \end{bmatrix} \in \mathbb{R}^{M\times N\times K\times L} \tag{1546}$$

while the second-order gradient has six-dimensional representation

$$\nabla^2 g(X) \triangleq \begin{bmatrix} \nabla^2 g_{11}(X) & \nabla^2 g_{12}(X) & \cdots & \nabla^2 g_{1N}(X) \\ \nabla^2 g_{21}(X) & \nabla^2 g_{22}(X) & \cdots & \nabla^2 g_{2N}(X) \\ \vdots & \vdots & & \vdots \\ \nabla^2 g_{M1}(X) & \nabla^2 g_{M2}(X) & \cdots & \nabla^2 g_{MN}(X) \end{bmatrix} \in \mathbb{R}^{M\times N\times K\times L\times K\times L} \tag{1547}$$

and so on.

### D.1.2 Product rules for matrix-functions

Given dimensionally compatible matrix-valued functions of matrix variable $f(X)$ and $g(X)$

$$\nabla_X\big(f(X)^T g(X)\big) = \nabla_X(f)\, g \,+\, \nabla_X(g)\, f \tag{1548}$$

while [39, §8.3] [237]

$$\nabla_X \operatorname{tr}\big(f(X)^T g(X)\big) = \nabla_X\Big(\operatorname{tr}\big(f(X)^T g(Z)\big) \,+\, \operatorname{tr}\big(g(X)\, f(Z)^T\big)\Big)\Big|_{Z\leftarrow X} \tag{1549}$$

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 $f(X) : \mathbb{R}^{\mathbf{2}\times\mathbf{2}} \to \mathbb{R}^{\mathbf{2}} = X^T a$ and $g(X) : \mathbb{R}^{\mathbf{2}\times\mathbf{2}} \to \mathbb{R}^{\mathbf{2}} = Xb$. We wish to find

$$\nabla_X\big(f(X)^T g(X)\big) = \nabla_X\, a^T X^2 b \tag{1550}$$

using the product rule. Formula (1548) calls for

$$\nabla_X\, a^T X^2 b = \nabla_X(X^T a)\, Xb \,+\, \nabla_X(Xb)\, X^T a \tag{1551}$$

Consider the first of the two terms:

$$\begin{aligned}\nabla_X(f)\, g &= \nabla_X(X^T a)\, Xb \\ &= \big[\, \nabla(X^T a)_1 \quad \nabla(X^T a)_2 \,\big]\, Xb\end{aligned} \tag{1552}$$

The gradient of $X^T a$ forms a cubix in $\mathbb{R}^{\mathbf{2}\times\mathbf{2}\times\mathbf{2}}$.

$$\nabla_X(X^T a)\, Xb = \left[\begin{array}{cccc} \frac{\partial(X^T a)_1}{\partial X_{11}} & & \frac{\partial(X^T a)_2}{\partial X_{11}} & \\ & \frac{\partial(X^T a)_1}{\partial X_{12}} & & \frac{\partial(X^T a)_2}{\partial X_{12}} \\ \frac{\partial(X^T a)_1}{\partial X_{21}} & & \frac{\partial(X^T a)_2}{\partial X_{21}} & \\ & \frac{\partial(X^T a)_1}{\partial X_{22}} & & \frac{\partial(X^T a)_2}{\partial X_{22}} \end{array}\right] \left[\begin{array}{c} (Xb)_1 \\ \\ (Xb)_2 \end{array}\right] \in \mathbb{R}^{\mathbf{2}\times 1\times\mathbf{2}} \tag{1553}$$

Because gradient of the product (1550) 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 the second dimension of the cubix (indicated by dotted line segments);

$$\begin{aligned}
\nabla_X(X^Ta)\,Xb &= \begin{bmatrix} a_1 & & 0 & \\ & 0 & & a_1 \\ a_2 & & 0 & \\ & 0 & & a_2 \end{bmatrix} \begin{bmatrix} b_1X_{11} + b_2X_{12} \\ b_1X_{21} + b_2X_{22} \end{bmatrix} \in \mathbb{R}^{\mathbf{2}\times 1\times \mathbf{2}} \\
&= \begin{bmatrix} a_1(b_1X_{11} + b_2X_{12}) & a_1(b_1X_{21} + b_2X_{22}) \\ a_2(b_1X_{11} + b_2X_{12}) & a_2(b_1X_{21} + b_2X_{22}) \end{bmatrix} \in \mathbb{R}^{\mathbf{2}\times\mathbf{2}} \\
&= ab^TX^T
\end{aligned} \tag{1554}$$

where the cubix appears as a complete $2\times 2\times 2$ matrix. In like manner for the second term $\nabla_X(g)\,f$

$$\begin{aligned}
\nabla_X(Xb)\,X^Ta &= \begin{bmatrix} b_1 & & 0 & \\ & b_2 & & 0 \\ 0 & & b_1 & \\ & 0 & & b_2 \end{bmatrix} \begin{bmatrix} X_{11}a_1 + X_{21}a_2 \\ X_{12}a_1 + X_{22}a_2 \end{bmatrix} \in \mathbb{R}^{\mathbf{2}\times 1\times \mathbf{2}} \\
&= X^Tab^T \in \mathbb{R}^{\mathbf{2}\times\mathbf{2}}
\end{aligned} \tag{1555}$$

The solution

$$\nabla_X\, a^TX^2b = ab^TX^T + X^Tab^T \tag{1556}$$

can be found from Table **D.2.1** or verified using (1549). □

#### D.1.2.1 Kronecker product

A partial remedy for venturing into *hyperdimensional* representations, such as the cubix or quartix, is to first vectorize matrices as in (30). This device gives rise to the Kronecker product of matrices $\otimes$; `a.k.a`, *direct product* or *tensor product*. Although it sees reversal in the literature, [245, §2.1] we adopt the definition: for $A\in\mathbb{R}^{m\times n}$ and $B\in\mathbb{R}^{p\times q}$

$$B\otimes A \triangleq \begin{bmatrix} B_{11}A & B_{12}A & \cdots & B_{1q}A \\ B_{21}A & B_{22}A & \cdots & B_{2q}A \\ \vdots & \vdots & & \vdots \\ B_{p1}A & B_{p2}A & \cdots & B_{pq}A \end{bmatrix} \in \mathbb{R}^{pm\times qn} \tag{1557}$$

One advantage to vectorization is existence of a traditional two-dimensional matrix representation for the second-order gradient of a real function with respect to a vectorized matrix. For example, from §A.1.1 no.23 (§D.2.1) for square $A\,,B\in\mathbb{R}^{n\times n}$ [116, §5.2] [10, §3]

$$\nabla^2_{\operatorname{vec} X}\operatorname{tr}(AXBX^T) \;=\; \nabla^2_{\operatorname{vec} X}\operatorname{vec}(X)^T(B^T\otimes A)\operatorname{vec} X \;=\; B\otimes A^T + B^T\otimes A \in\mathbb{R}^{n^2\times n^2} \tag{1558}$$

To disadvantage is a large new but known set of algebraic rules 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

$$W \triangleq \begin{bmatrix} w_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0}^T & & w_N \end{bmatrix} \in\mathbb{S}^N \tag{1559}$$

The conventional way of accomplishing that is to multiply $S$ by diagonal matrix $W$ on the right-hand side:[D.2]

$$SW = S\begin{bmatrix} w_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0}^T & & w_N \end{bmatrix} = \begin{bmatrix} S(:,1)w_1 & \cdots & S(:,N)w_N \end{bmatrix} \in\mathbb{R}^{M\times N} \tag{1560}$$

To reverse product order such that diagonal matrix $W$ instead appears to the left of $S$: (Sze Wan)

$$SW \;=\; (\delta(W)^T\otimes I)\begin{bmatrix} S(:,1) & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & S(:,2) & \ddots & \\ & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & & \mathbf{0} & S(:,N) \end{bmatrix} \in\mathbb{R}^{M\times N} \tag{1561}$$

where $I\in\mathbb{S}^M$. For any matrices of like size, $S,Y\in\mathbb{R}^{M\times N}$

$$S\circ Y = \begin{bmatrix} \delta(Y(:,1)) & \cdots & \delta(Y(:,N)) \end{bmatrix}\begin{bmatrix} S(:,1) & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & S(:,2) & \ddots & \\ & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & & \mathbf{0} & S(:,N) \end{bmatrix} \in\mathbb{R}^{M\times N} \tag{1562}$$

---

[D.2] Multiplying on the left by $W\in\mathbb{S}^M$ would instead weight the rows of $S$.

which converts a Hadamard product into a standard matrix product. Hadamard product can be extracted from within the Kronecker product. [150, p.475]

### D.1.3 Chain rules for composite matrix-functions

Given dimensionally compatible matrix-valued functions of matrix variable $f(X)$ and $g(X)$ [161, §15.7]

$$\nabla_X\, g\big(f(X)^T\big) = \nabla_X f^T\, \nabla_f\, g \tag{1563}$$

$$\nabla_X^2\, g\big(f(X)^T\big) \;=\; \nabla_X\big(\nabla_X f^T\, \nabla_f\, g\big) \;=\; \nabla_X^2 f\; \nabla_f\, g \;+\; \nabla_X f^T\, \nabla_f^2 g\; \nabla_X f \tag{1564}$$

#### D.1.3.1 Two arguments

$$\nabla_X\, g\big(f(X)^T,\, h(X)^T\big) = \nabla_X f^T\, \nabla_f\, g \;+\; \nabla_X h^T\, \nabla_h\, g \tag{1565}$$

**D.1.3.1.1 Example.** *Chain rule for two arguments.* [30, §1.1]

$$g\big(f(x)^T,\, h(x)^T\big) = (f(x) + h(x))^T A\, (f(x) + h(x)) \tag{1566}$$

$$f(x) = \begin{bmatrix} x_1 \\ \varepsilon x_2 \end{bmatrix}, \qquad h(x) = \begin{bmatrix} \varepsilon x_1 \\ x_2 \end{bmatrix} \tag{1567}$$

$$\nabla_x\, g\big(f(x)^T,\, h(x)^T\big) = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}(A + A^T)(f + h) + \begin{bmatrix} \varepsilon & 0 \\ 0 & 1 \end{bmatrix}(A + A^T)(f + h) \tag{1568}$$

$$\nabla_x\, g\big(f(x)^T,\, h(x)^T\big) = \begin{bmatrix} 1+\varepsilon & 0 \\ 0 & 1+\varepsilon \end{bmatrix}(A + A^T)\left(\begin{bmatrix} x_1 \\ \varepsilon x_2 \end{bmatrix} + \begin{bmatrix} \varepsilon x_1 \\ x_2 \end{bmatrix}\right) \tag{1569}$$

$$\lim_{\varepsilon \to 0} \nabla_x\, g\big(f(x)^T,\, h(x)^T\big) = (A + A^T)x \tag{1570}$$

from Table D.2.1. □

These 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 $\operatorname{dom} g$ which is an open set. We seek simple expressions for the first and second directional derivatives in direction $Y \in \mathbb{R}^{K\times L}$, $\overset{\rightarrow Y}{dg} \in \mathbb{R}^{M\times N}$ and $\overset{\rightarrow Y}{dg^2} \in \mathbb{R}^{M\times N}$ respectively.

Assuming that the limit exists, we may state the partial derivative of the $mn^{\text{th}}$ entry of $g$ with respect to the $kl^{\text{th}}$ entry of $X$;

$$\frac{\partial g_{mn}(X)}{\partial X_{kl}} = \lim_{\Delta t \to 0} \frac{g_{mn}(X + \Delta t\, e_k e_l^T) - g_{mn}(X)}{\Delta t} \in \mathbb{R} \tag{1571}$$

where $e_k$ is the $k^{\text{th}}$ standard basis vector in $\mathbb{R}^K$ while $e_l$ is the $l^{\text{th}}$ standard basis vector in $\mathbb{R}^L$. The total number of partial derivatives equals $KLMN$ while the gradient is defined in their terms; the $mn^{\text{th}}$ entry of the gradient is

$$\nabla g_{mn}(X) = \begin{bmatrix} \frac{\partial g_{mn}(X)}{\partial X_{11}} & \frac{\partial g_{mn}(X)}{\partial X_{12}} & \cdots & \frac{\partial g_{mn}(X)}{\partial X_{1L}} \\ \frac{\partial g_{mn}(X)}{\partial X_{21}} & \frac{\partial g_{mn}(X)}{\partial X_{22}} & \cdots & \frac{\partial g_{mn}(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_{mn}(X)}{\partial X_{K1}} & \frac{\partial g_{mn}(X)}{\partial X_{K2}} & \cdots & \frac{\partial g_{mn}(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K\times L} \tag{1572}$$

while the gradient is a quartix

$$\nabla g(X) = \begin{bmatrix} \nabla g_{11}(X) & \nabla g_{12}(X) & \cdots & \nabla g_{1N}(X) \\ \nabla g_{21}(X) & \nabla g_{22}(X) & \cdots & \nabla g_{2N}(X) \\ \vdots & \vdots & & \vdots \\ \nabla g_{M1}(X) & \nabla g_{M2}(X) & \cdots & \nabla g_{MN}(X) \end{bmatrix} \in \mathbb{R}^{M\times N\times K\times L} \tag{1573}$$

By simply rotating our perspective of the four-dimensional representation of gradient matrix, we find one of three useful transpositions of this quartix (connoted $T_1$):

$$\nabla g(X)^{T_1} = \begin{bmatrix} \frac{\partial g(X)}{\partial X_{11}} & \frac{\partial g(X)}{\partial X_{12}} & \cdots & \frac{\partial g(X)}{\partial X_{1L}} \\ \frac{\partial g(X)}{\partial X_{21}} & \frac{\partial g(X)}{\partial X_{22}} & \cdots & \frac{\partial g(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g(X)}{\partial X_{K1}} & \frac{\partial g(X)}{\partial X_{K2}} & \cdots & \frac{\partial g(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K\times L\times M\times N} \tag{1574}$$

When the limit for $\Delta t \in \mathbb{R}$ exists, it is easy to show by substitution of variables in (1571)

$$\frac{\partial g_{mn}(X)}{\partial X_{kl}} Y_{kl} \;=\; \lim_{\Delta t\to 0} \frac{g_{mn}(X+\Delta t\, Y_{kl}\, e_k e_l^T) - g_{mn}(X)}{\Delta t} \in \mathbb{R} \tag{1575}$$

which may be interpreted as the change in $g_{mn}$ at $X$ when the change in $X_{kl}$ is equal to $Y_{kl}$ , the $kl^{\text{th}}$ 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 $mn^{\text{th}}$ entry of the directional derivative is the corresponding total differential [161, §15.8]

$$\begin{aligned} dg_{mn}(X)|_{dX\to Y} \;&=\; \sum_{k,l} \frac{\partial g_{mn}(X)}{\partial X_{kl}} Y_{kl} \;=\; \operatorname{tr}\!\left(\nabla g_{mn}(X)^T Y\right) &(1576)\\ &=\; \sum_{k,l} \lim_{\Delta t\to 0} \frac{g_{mn}(X+\Delta t\, Y_{kl}\, e_k e_l^T) - g_{mn}(X)}{\Delta t} &(1577)\\ &=\; \lim_{\Delta t\to 0} \frac{g_{mn}(X+\Delta t\, Y) - g_{mn}(X)}{\Delta t} &(1578)\\ &=\; \left.\frac{d}{dt}\right|_{t=0} g_{mn}(X+t\,Y) &(1579) \end{aligned}$$

where $t\in\mathbb{R}$. Assuming finite $Y$, equation (1578) is called the *Gâteaux differential* [29, App.A.5] [148, §D.2.1] [268, §5.28] whose existence is implied by the existence of the *Fréchet differential*, the sum in (1576). [182, §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 $Y$.[D.3] Hence the directional derivative,

$$\begin{aligned}
\overset{\rightarrow Y}{dg}(X) &\triangleq \left.\begin{bmatrix} dg_{11}(X) & dg_{12}(X) & \cdots & dg_{1N}(X) \\ dg_{21}(X) & dg_{22}(X) & \cdots & dg_{2N}(X) \\ \vdots & \vdots & & \vdots \\ dg_{M1}(X) & dg_{M2}(X) & \cdots & dg_{MN}(X) \end{bmatrix}\right|_{dX\rightarrow Y} \in \mathbb{R}^{M\times N} \\
&= \begin{bmatrix} \operatorname{tr}\big(\nabla g_{11}(X)^T Y\big) & \operatorname{tr}\big(\nabla g_{12}(X)^T Y\big) & \cdots & \operatorname{tr}\big(\nabla g_{1N}(X)^T Y\big) \\ \operatorname{tr}\big(\nabla g_{21}(X)^T Y\big) & \operatorname{tr}\big(\nabla g_{22}(X)^T Y\big) & \cdots & \operatorname{tr}\big(\nabla g_{2N}(X)^T Y\big) \\ \vdots & \vdots & & \vdots \\ \operatorname{tr}\big(\nabla g_{M1}(X)^T Y\big) & \operatorname{tr}\big(\nabla g_{M2}(X)^T Y\big) & \cdots & \operatorname{tr}\big(\nabla g_{MN}(X)^T Y\big) \end{bmatrix} \\
&= \begin{bmatrix} \sum\limits_{k,l} \frac{\partial g_{11}(X)}{\partial X_{kl}} Y_{kl} & \sum\limits_{k,l} \frac{\partial g_{12}(X)}{\partial X_{kl}} Y_{kl} & \cdots & \sum\limits_{k,l} \frac{\partial g_{1N}(X)}{\partial X_{kl}} Y_{kl} \\ \sum\limits_{k,l} \frac{\partial g_{21}(X)}{\partial X_{kl}} Y_{kl} & \sum\limits_{k,l} \frac{\partial g_{22}(X)}{\partial X_{kl}} Y_{kl} & \cdots & \sum\limits_{k,l} \frac{\partial g_{2N}(X)}{\partial X_{kl}} Y_{kl} \\ \vdots & \vdots & & \vdots \\ \sum\limits_{k,l} \frac{\partial g_{M1}(X)}{\partial X_{kl}} Y_{kl} & \sum\limits_{k,l} \frac{\partial g_{M2}(X)}{\partial X_{kl}} Y_{kl} & \cdots & \sum\limits_{k,l} \frac{\partial g_{MN}(X)}{\partial X_{kl}} Y_{kl} \end{bmatrix}
\end{aligned} \tag{1580}$$

from which it follows

$$\overset{\rightarrow Y}{dg}(X) = \sum_{k,l} \frac{\partial g(X)}{\partial X_{kl}} Y_{kl} \tag{1581}$$

Yet for all $X \in \operatorname{dom} g$, any $Y \in \mathbb{R}^{K\times L}$, and some open interval of $t \in \mathbb{R}$

$$g(X + t\,Y) = g(X) \;+\; t\,\overset{\rightarrow Y}{dg}(X) \;+\; o(t^2) \tag{1582}$$

which is the first-order Taylor series expansion about $X$. [161, §18.4] [104, §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 (1579). 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 $X \in \operatorname{dom} g$ becomes

[D.3] Although $Y$ is a matrix, we may regard it as a vector in $\mathbb{R}^{KL}$.

$$\overset{\rightarrow Y}{dg}(X) = \left.\frac{d}{dt}\right|_{t=0} g(X+t\,Y) \in \mathbb{R}^{M\times N} \tag{1583}$$

[204, §2.1, §5.4.5] [27, §6.3.1] which is simplest. In case of a real function $g(X) : \mathbb{R}^{K\times L}\to\mathbb{R}$

$$\overset{\rightarrow Y}{dg}(X) = \operatorname{tr}\big(\nabla g(X)^T Y\big) \qquad (1605)$$

In case $g(X) : \mathbb{R}^{K}\to\mathbb{R}$

$$\overset{\rightarrow Y}{dg}(X) = \nabla g(X)^T Y \qquad (1608)$$

Unlike gradient, directional derivative does not expand dimension; directional derivative (1583) 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, $\overset{\rightarrow Y}{dg}(X) = g(Y)$. [182, §7.2]

#### D.1.4.1 Interpretation of directional derivative

In the case of any differentiable real function $f(X) : \mathbb{R}^{K\times L}\to\mathbb{R}$, the directional derivative of $f(X)$ at $X$ in any direction $Y$ yields the slope of $f$ along the line $\{X+t\,Y \mid t\in\mathbb{R}\}$ through its domain evaluated at $t=0$. For higher-dimensional functions, by (1580), this slope interpretation can be applied to each entry of the directional derivative.

Figure **117**, 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. (1608) 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. (*confer* §3.1.8)

**D.1.4.1.1 Theorem.** *Directional derivative optimality condition.*
[182, §7.4] Suppose $f(X) : \mathbb{R}^{K\times L}\to\mathbb{R}$ is minimized on convex set $\mathcal{C}\subseteq\mathbb{R}^{p\times k}$ by $X^\star$, and the directional derivative of $f$ exists there. Then for all $X\in\mathcal{C}$

$$\overset{\rightarrow X-X^\star}{df}(X) \geq 0 \tag{1584}$$

◇

$$\upsilon \triangleq \begin{bmatrix} \nabla_x f(\alpha) \\ \\ \frac{1}{2} \overset{\to\nabla_x f(\alpha)}{df(\alpha)} \end{bmatrix}$$

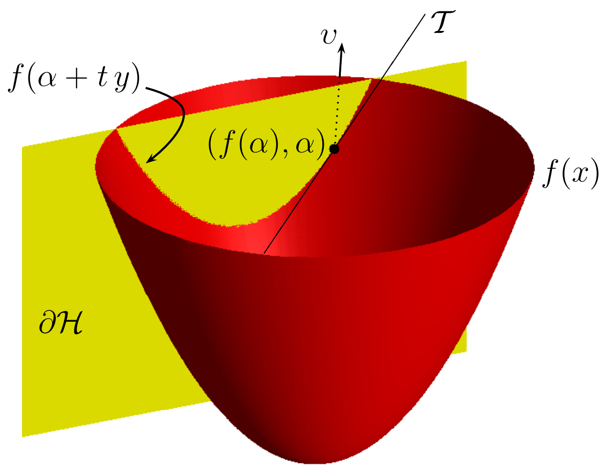



Figure 117: Convex quadratic bowl in $\mathbb{R}^{\mathbf{2}} \times \mathbb{R}$ ; $f(x) = x^T x : \mathbb{R}^{\mathbf{2}} \to \mathbb{R}$ *versus* $x$ on some open disc in $\mathbb{R}^{\mathbf{2}}$. Plane slice $\partial\mathcal{H}$ is perpendicular to function domain. Slice intersection with domain connotes bidirectional vector $y$ . Slope of tangent line $\mathcal{T}$ at point $(\alpha\,,\, f(\alpha))$ is value of directional derivative $\nabla_x f(\alpha)^T y$ (1608) at $\alpha$ in slice direction $y$ . Negative gradient $-\nabla_x f(x) \in \mathbb{R}^{\mathbf{2}}$ is direction of *steepest descent.* [283] [161, §15.6] [104] When vector $\upsilon \in \mathbb{R}^{\mathbf{3}}$ entry $\upsilon_3$ is half directional derivative in gradient direction at $\alpha$ and when $\begin{bmatrix} \upsilon_1 \\ \upsilon_2 \end{bmatrix} = \nabla_x f(\alpha)$ , then $-\upsilon$ points directly toward bowl bottom.

**D.1.4.1.2 Example.** *Simple bowl.*
Bowl function (Figure **117**)

$$f(x) : \mathbb{R}^K \rightarrow \mathbb{R} \stackrel{\Delta}{=} (x-a)^T(x-a) - b \tag{1585}$$

has function offset $-b \in \mathbb{R}$, axis of revolution at $x = a$, and positive definite Hessian (1537) everywhere in its domain (an open *hyperdisc* in $\mathbb{R}^K$); *id est*, strictly convex quadratic $f(x)$ has unique global minimum equal to $-b$ at $x = a$. A vector $-\upsilon$ based anywhere in $\operatorname{dom} f \times \mathbb{R}$ pointing toward the unique bowl-bottom is specified:

$$\upsilon \propto \begin{bmatrix} x-a \\ f(x)+b \end{bmatrix} \in \mathbb{R}^K \times \mathbb{R} \tag{1586}$$

Such a vector is

$$\upsilon = \begin{bmatrix} \nabla_x f(x) \\ \\ \frac{1}{2} \overset{\rightarrow \nabla_x f(x)}{df(x)} \end{bmatrix} \tag{1587}$$

since the gradient is

$$\nabla_x f(x) = 2(x-a) \tag{1588}$$

and the directional derivative in the direction of the gradient is (1608)

$$\overset{\rightarrow \nabla_x f(x)}{df(x)} = \nabla_x f(x)^T \nabla_x f(x) = 4(x-a)^T(x-a) = 4(f(x)+b) \tag{1589}$$

$\square$

### 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,

$$\nabla \frac{\partial g_{mn}(X)}{\partial X_{kl}} = \frac{\partial \nabla g_{mn}(X)}{\partial X_{kl}} = \begin{bmatrix} \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{11}} & \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{12}} & \cdots & \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{1L}} \\ \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{21}} & \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{22}} & \cdots & \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{K1}} & \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{K2}} & \cdots & \frac{\partial^2 g_{mn}(X)}{\partial X_{kl} \partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K \times L} \tag{1590}$$

$$\begin{aligned}\nabla^2 g_{mn}(X) &= \begin{bmatrix} \nabla\frac{\partial g_{mn}(X)}{\partial X_{11}} & \nabla\frac{\partial g_{mn}(X)}{\partial X_{12}} & \cdots & \nabla\frac{\partial g_{mn}(X)}{\partial X_{1L}} \\ \nabla\frac{\partial g_{mn}(X)}{\partial X_{21}} & \nabla\frac{\partial g_{mn}(X)}{\partial X_{22}} & \cdots & \nabla\frac{\partial g_{mn}(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \nabla\frac{\partial g_{mn}(X)}{\partial X_{K1}} & \nabla\frac{\partial g_{mn}(X)}{\partial X_{K2}} & \cdots & \nabla\frac{\partial g_{mn}(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K\times L\times K\times L} \\ &= \begin{bmatrix} \frac{\partial \nabla g_{mn}(X)}{\partial X_{11}} & \frac{\partial \nabla g_{mn}(X)}{\partial X_{12}} & \cdots & \frac{\partial \nabla g_{mn}(X)}{\partial X_{1L}} \\ \frac{\partial \nabla g_{mn}(X)}{\partial X_{21}} & \frac{\partial \nabla g_{mn}(X)}{\partial X_{22}} & \cdots & \frac{\partial \nabla g_{mn}(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \nabla g_{mn}(X)}{\partial X_{K1}} & \frac{\partial \nabla g_{mn}(X)}{\partial X_{K2}} & \cdots & \frac{\partial \nabla g_{mn}(X)}{\partial X_{KL}} \end{bmatrix}\end{aligned} \tag{1591}$$

Rotating our perspective, we get several views of the second-order gradient:

$$\nabla^2 g(X) = \begin{bmatrix} \nabla^2 g_{11}(X) & \nabla^2 g_{12}(X) & \cdots & \nabla^2 g_{1N}(X) \\ \nabla^2 g_{21}(X) & \nabla^2 g_{22}(X) & \cdots & \nabla^2 g_{2N}(X) \\ \vdots & \vdots & & \vdots \\ \nabla^2 g_{M1}(X) & \nabla^2 g_{M2}(X) & \cdots & \nabla^2 g_{MN}(X) \end{bmatrix} \in \mathbb{R}^{M\times N\times K\times L\times K\times L} \tag{1592}$$

$$\nabla^2 g(X)^{T_1} = \begin{bmatrix} \nabla\frac{\partial g(X)}{\partial X_{11}} & \nabla\frac{\partial g(X)}{\partial X_{12}} & \cdots & \nabla\frac{\partial g(X)}{\partial X_{1L}} \\ \nabla\frac{\partial g(X)}{\partial X_{21}} & \nabla\frac{\partial g(X)}{\partial X_{22}} & \cdots & \nabla\frac{\partial g(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \nabla\frac{\partial g(X)}{\partial X_{K1}} & \nabla\frac{\partial g(X)}{\partial X_{K2}} & \cdots & \nabla\frac{\partial g(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K\times L\times M\times N\times K\times L} \tag{1593}$$

$$\nabla^2 g(X)^{T_2} = \begin{bmatrix} \frac{\partial \nabla g(X)}{\partial X_{11}} & \frac{\partial \nabla g(X)}{\partial X_{12}} & \cdots & \frac{\partial \nabla g(X)}{\partial X_{1L}} \\ \frac{\partial \nabla g(X)}{\partial X_{21}} & \frac{\partial \nabla g(X)}{\partial X_{22}} & \cdots & \frac{\partial \nabla g(X)}{\partial X_{2L}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \nabla g(X)}{\partial X_{K1}} & \frac{\partial \nabla g(X)}{\partial X_{K2}} & \cdots & \frac{\partial \nabla g(X)}{\partial X_{KL}} \end{bmatrix} \in \mathbb{R}^{K\times L\times K\times L\times M\times N} \tag{1594}$$

Assuming the limits exist, we may state the partial derivative of the $mn^{\text{th}}$ entry of $g$ with respect to the $kl^{\text{th}}$ and $ij^{\text{th}}$ entries of $X$ ;

$$\frac{\partial^2 g_{mn}(X)}{\partial X_{kl}\,\partial X_{ij}} = \lim_{\Delta\tau,\Delta t\to 0} \frac{g_{mn}(X+\Delta t\, e_k e_l^T+\Delta\tau\, e_i e_j^T)-g_{mn}(X+\Delta t\, e_k e_l^T)-\big(g_{mn}(X+\Delta\tau\, e_i e_j^T)-g_{mn}(X)\big)}{\Delta\tau\,\Delta t} \tag{1595}$$

Differentiating (1575) and then scaling by $Y_{ij}$

$$\begin{aligned}\frac{\partial^2 g_{mn}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl} Y_{ij} &= \lim_{\Delta t\to 0} \frac{\partial g_{mn}(X+\Delta t\, Y_{kl}\, e_k e_l^T)-\partial g_{mn}(X)}{\partial X_{ij}\,\Delta t} Y_{ij} \qquad (1596) \\ &= \lim_{\Delta\tau,\Delta t\to 0} \frac{g_{mn}(X+\Delta t\, Y_{kl}\, e_k e_l^T+\Delta\tau\, Y_{ij}\, e_i e_j^T)-g_{mn}(X+\Delta t\, Y_{kl}\, e_k e_l^T)-\big(g_{mn}(X+\Delta\tau\, Y_{ij}\, e_i e_j^T)-g_{mn}(X)\big)}{\Delta\tau\,\Delta t}\end{aligned}$$

which can be proved by substitution of variables in (1595). The $mn^{\text{th}}$ second-order total differential due to any $Y\in\mathbb{R}^{K\times L}$ is

$$\begin{aligned} d^2 g_{mn}(X)|_{dX\to Y} &= \sum_{i,j}\sum_{k,l}\frac{\partial^2 g_{mn}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl} Y_{ij} = \operatorname{tr}\Big(\nabla_X \operatorname{tr}\big(\nabla g_{mn}(X)^T Y\big)^T Y\Big) && (1597)\\ &= \sum_{i,j}\lim_{\Delta t\to 0}\frac{\partial g_{mn}(X+\Delta t\,Y)-\partial g_{mn}(X)}{\partial X_{ij}\,\Delta t} Y_{ij} && (1598)\\ &= \lim_{\Delta t\to 0}\frac{g_{mn}(X+2\Delta t\,Y)-2g_{mn}(X+\Delta t\,Y)+g_{mn}(X)}{\Delta t^2} && (1599)\\ &= \left.\frac{d^2}{dt^2}\right|_{t=0} g_{mn}(X+t\,Y) && (1600)\end{aligned}$$

Hence the second directional derivative,

$$\overset{\to Y}{dg^2}(X) \triangleq \left.\begin{bmatrix} d^2 g_{11}(X) & d^2 g_{12}(X) & \cdots & d^2 g_{1N}(X)\\ d^2 g_{21}(X) & d^2 g_{22}(X) & \cdots & d^2 g_{2N}(X)\\ \vdots & \vdots & & \vdots\\ d^2 g_{M1}(X) & d^2 g_{M2}(X) & \cdots & d^2 g_{MN}(X)\end{bmatrix}\right|_{dX\to Y} \in\mathbb{R}^{M\times N}$$

$$= \begin{bmatrix} \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{11}(X)^T Y\big)^T Y\Big) & \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{12}(X)^T Y\big)^T Y\Big) & \cdots & \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{1N}(X)^T Y\big)^T Y\Big)\\ \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{21}(X)^T Y\big)^T Y\Big) & \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{22}(X)^T Y\big)^T Y\Big) & \cdots & \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{2N}(X)^T Y\big)^T Y\Big)\\ \vdots & \vdots & & \vdots\\ \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{M1}(X)^T Y\big)^T Y\Big) & \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{M2}(X)^T Y\big)^T Y\Big) & \cdots & \operatorname{tr}\Big(\nabla\operatorname{tr}\big(\nabla g_{MN}(X)^T Y\big)^T Y\Big)\end{bmatrix}$$

$$= \begin{bmatrix} \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{11}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij} & \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{12}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij} & \cdots & \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{1N}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij}\\ \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{21}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij} & \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{22}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij} & \cdots & \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{2N}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij}\\ \vdots & \vdots & & \vdots\\ \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{M1}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij} & \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{M2}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij} & \cdots & \sum\limits_{i,j}\sum\limits_{k,l}\frac{\partial^2 g_{MN}(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl}Y_{ij}\end{bmatrix} \tag{1601}$$

from which it follows

$$\overset{\rightarrow Y}{dg}{}^2(X) = \sum_{i,j}\sum_{k,l} \frac{\partial^2 g(X)}{\partial X_{kl}\,\partial X_{ij}} Y_{kl} Y_{ij} = \sum_{i,j} \frac{\partial}{\partial X_{ij}} \overset{\rightarrow Y}{dg}(X)\, Y_{ij} \tag{1602}$$

Yet for all $X \in \operatorname{dom} g$, any $Y \in \mathbb{R}^{K\times L}$, and some open interval of $t \in \mathbb{R}$

$$g(X + t\,Y) = g(X) \;+\; t\,\overset{\rightarrow Y}{dg}(X) \;+\; \frac{1}{2!} t^2\, \overset{\rightarrow Y}{dg}{}^2(X) \;+\; o(t^3) \tag{1603}$$

which is the second-order Taylor series expansion about $X$. [161, §18.4] [104, §2.3.4] Differentiating twice with respect to $t$ and subsequent $t$-zeroing isolates the third term of the expansion. Thus differentiating and zeroing $g(X + t\,Y)$ in $t$ is an operation equivalent to individually differentiating and zeroing every entry $g_{mn}(X + t\,Y)$ as in (1600). So the second directional derivative of $g(X) : \mathbb{R}^{K\times L} \rightarrow \mathbb{R}^{M\times N}$ becomes [204, §2.1, §5.4.5] [27, §6.3.1]

$$\overset{\rightarrow Y}{dg}{}^2(X) = \left.\frac{d^2}{dt^2}\right|_{t=0} g(X + t\,Y) \;\in \mathbb{R}^{M\times N} \tag{1604}$$

which is again simplest. (*confer* (1583)) 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 the directional derivatives are in $\mathbb{R}$:

$$\overset{\rightarrow Y}{dg}(X) = \operatorname{tr}\big(\nabla g(X)^T Y\big) \tag{1605}$$

$$\overset{\rightarrow Y}{dg}{}^2(X) = \operatorname{tr}\Big(\nabla_X \operatorname{tr}\big(\nabla g(X)^T Y\big)^T Y\Big) = \operatorname{tr}\Big(\nabla_X \overset{\rightarrow Y}{dg}(X)^T Y\Big) \tag{1606}$$

$$\overset{\rightarrow Y}{dg}{}^3(X) = \operatorname{tr}\bigg(\nabla_X \operatorname{tr}\Big(\nabla_X \operatorname{tr}\big(\nabla g(X)^T Y\big)^T Y\Big)^T Y\bigg) = \operatorname{tr}\Big(\nabla_X \overset{\rightarrow Y}{dg}{}^2(X)^T Y\Big) \tag{1607}$$

In the case $g(X) : \mathbb{R}^K \rightarrow \mathbb{R}$ has vector argument, they further simplify:

$$\overset{\rightarrow Y}{dg}(X) = \nabla g(X)^T Y \tag{1608}$$

$$\overset{\rightarrow Y}{dg^2}(X) = Y^T \nabla^2 g(X) Y \tag{1609}$$

$$\overset{\rightarrow Y}{dg^3}(X) = \nabla_X \big(Y^T \nabla^2 g(X) Y\big)^T Y \tag{1610}$$

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 (1582) and (1603). Assuming $g(X)$ has continuous first-, second-, and third-order gradients over the open set $\operatorname{dom} g$, then for $X \in \operatorname{dom} g$ and any $Y \in \mathbb{R}^{K \times L}$ the complete Taylor series on some open interval of $\mu \in \mathbb{R}$ is expressed

$$g(X+\mu Y) = g(X) + \mu \overset{\rightarrow Y}{dg}(X) + \frac{1}{2!}\mu^2 \overset{\rightarrow Y}{dg^2}(X) + \frac{1}{3!}\mu^3 \overset{\rightarrow Y}{dg^3}(X) + o(\mu^4) \tag{1611}$$

or on some open interval of $\|Y\|$

$$g(Y) = g(X) \;+\; \overset{\rightarrow Y-X}{dg}(X) \;+\; \frac{1}{2!}\overset{\rightarrow Y-X}{dg^2}(X) \;+\; \frac{1}{3!}\overset{\rightarrow Y-X}{dg^3}(X) \;+\; o(\|Y\|^4) \tag{1612}$$

which are third-order expansions about $X$. The *mean value theorem* from calculus is what insures finite order of the series. [161] [30, §1.1] [29, App.A.5] [148, §0.4]

**D.1.7.0.1 Exercise.** *log det.* (*confer* [46, p.644])
Find the first two terms of the Taylor series expansion (1612) for $\log\det X$.
▼

### D.1.8 Correspondence of gradient to derivative

From the foregoing expressions for directional derivative, we derive a relationship between the gradient with respect to matrix $X$ and the derivative with respect to real variable $t$ :

#### D.1.8.1 first-order

Removing from (1583) the evaluation at $t=0$ ,[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 $\operatorname{dom} g$

$$\overset{\to Y}{dg}(X+t\,Y) = \frac{d}{dt}g(X+t\,Y) \tag{1613}$$

In the general case $g(X):\mathbb{R}^{K\times L}\to\mathbb{R}^{M\times N}$, from (1576) and (1579) we find

$$\operatorname{tr}\big(\nabla_X\, g_{mn}(X+t\,Y)^T Y\big) = \frac{d}{dt}g_{mn}(X+t\,Y) \tag{1614}$$

which is valid at $t=0$, of course, when $X\in\operatorname{dom} g$. In the important case of a real function $g(X):\mathbb{R}^{K\times L}\to\mathbb{R}$ , from (1605) we have simply

$$\operatorname{tr}\big(\nabla_X\, g(X+t\,Y)^T Y\big) = \frac{d}{dt}g(X+t\,Y) \tag{1615}$$

When, additionally, $g(X):\mathbb{R}^{K}\to\mathbb{R}$ has vector argument,

$$\nabla_X\, g(X+t\,Y)^T Y = \frac{d}{dt}g(X+t\,Y) \tag{1616}$$

[D.4] Justified by replacing $X$ with $X+t\,Y$ in (1576)-(1578); beginning,

$$dg_{mn}(X+t\,Y)|_{dX\to Y} = \sum_{k\,,\,l}\frac{\partial g_{mn}(X+t\,Y)}{\partial X_{kl}}Y_{kl}$$

**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,

$$\operatorname{tr}\big(\nabla_X\, g(X+t\,Y)^T Y\big) \;=\; \operatorname{tr}\big(2ww^T(X^T + t\,Y^T)Y\big) \tag{1617}$$
$$=\; 2w^T(X^T Y + t\,Y^T Y)w \tag{1618}$$

Applying the equivalence (1615),

$$\frac{d}{dt}g(X+t\,Y) \;=\; \frac{d}{dt}w^T(X+t\,Y)^T(X+t\,Y)w \tag{1619}$$
$$=\; w^T\big(X^T Y + Y^T X + 2t\,Y^T Y\big)\,w \tag{1620}$$
$$=\; 2w^T(X^T Y + t\,Y^T Y)w \tag{1621}$$

which is the same as (1618); hence, equivalence is demonstrated.

It is easy to extract $\nabla g(X)$ from (1621) knowing only (1615):

$$\begin{aligned}
\operatorname{tr}\big(\nabla_X\, g(X+t\,Y)^T Y\big) &= 2w^T(X^T Y + t\,Y^T Y)w\\
&= 2\operatorname{tr}\big(ww^T(X^T + t\,Y^T)Y\big)\\
\operatorname{tr}\big(\nabla_X\, g(X)^T Y\big) &= 2\operatorname{tr}\big(ww^T X^T Y\big)\\
&\Leftrightarrow\\
\nabla_X\, g(X) &= 2Xww^T
\end{aligned} \tag{1622}$$

□

### D.1.8.2 second-order

Likewise removing the evaluation at $t=0$ from (1604),

$$\overset{\to Y}{dg}{}^2(X+t\,Y) = \frac{d^2}{dt^2}g(X+t\,Y) \tag{1623}$$

we can find a similar relationship between the second-order gradient and the second derivative: In the general case $g(X) : \mathbb{R}^{K\times L} \to \mathbb{R}^{M\times N}$ from (1597) and (1600),

$$\operatorname{tr}\Big(\nabla_X \operatorname{tr}\big(\nabla_X\, g_{mn}(X+t\,Y)^T Y\big)^T Y\Big) = \frac{d^2}{dt^2}g_{mn}(X+t\,Y) \tag{1624}$$

In the case of a real function $g(X) : \mathbb{R}^{K\times L} \to \mathbb{R}$ we have, of course,

$$\operatorname{tr}\Big(\nabla_X \operatorname{tr}\big(\nabla_X\, g(X+t\,Y)^T Y\big)^T Y\Big) = \frac{d^2}{dt^2}g(X+t\,Y) \tag{1625}$$

From (1609), the simpler case, where the real function $g(X) : \mathbb{R}^K \to \mathbb{R}$ has vector argument,

$$Y^T \nabla_X^2 g(X+t\,Y)Y = \frac{d^2}{dt^2} g(X+t\,Y) \tag{1626}$$

**D.1.8.2.1 Example.** *Second-order gradient.*
Given real function $g(X) = \log\det X$ having domain $\operatorname{int}\mathbb{S}_+^K$, we want to find $\nabla^2 g(X) \in \mathbb{R}^{K\times K\times K\times K}$. From the tables in §D.2,

$$h(X) \triangleq \nabla g(X) = X^{-1} \in \operatorname{int}\mathbb{S}_+^K \tag{1627}$$

so $\nabla^2 g(X) = \nabla h(X)$. By (1614) and (1582), for $Y \in \mathbb{S}^K$

$$\begin{aligned}
\operatorname{tr}\big(\nabla h_{mn}(X)^T Y\big) &= \left.\frac{d}{dt}\right|_{t=0} h_{mn}(X+t\,Y) & (1628)\\
&= \left(\left.\frac{d}{dt}\right|_{t=0} h(X+t\,Y)\right)_{mn} & (1629)\\
&= \left(\left.\frac{d}{dt}\right|_{t=0} (X+t\,Y)^{-1}\right)_{mn} & (1630)\\
&= -\big(X^{-1}YX^{-1}\big)_{mn} & (1631)
\end{aligned}$$

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 (32) of the trace function we find

$$\nabla^2 g(X)_{mnkl} = \operatorname{tr}\big(\nabla h_{mn}(X)^T e_k e_l^T\big) = \nabla h_{mn}(X)_{kl} = -\big(X^{-1}e_k e_l^T X^{-1}\big)_{mn} \tag{1632}$$

$$\nabla^2 g(X)_{kl} = \nabla h(X)_{kl} = -\big(X^{-1}e_k e_l^T X^{-1}\big) \in \mathbb{R}^{K\times K} \tag{1633}$$

$\square$

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

[116] [50]

- When proving results for symmetric matrices algebraically, it is critical to take gradients ignoring symmetry and to then substitute symmetric entries afterward.

- $a\,,b\in\mathbb{R}^n$, $x\,,y\in\mathbb{R}^k$, $A\,,B\in\mathbb{R}^{m\times n}$, $X\,,Y\in\mathbb{R}^{K\times L}$, $t\,,\mu\in\mathbb{R}$ , $i\,,j\,,k\,,\ell\,,K\,,L\,,m\,,n\,,M\,,N$ are integers, unless otherwise noted.

- $x^\mu$ means $\delta\big(\delta(x)^\mu\big)$ for $\mu\in\mathbb{R}$ ; *id est*, entrywise vector exponentiation. $\delta$ is the main-diagonal linear operator (1220). $x^0\triangleq\mathbf{1}$, $X^0\triangleq I$ if square.

- $\frac{d}{dx}\triangleq\begin{bmatrix}\frac{d}{dx_1}\\ \vdots\\ \frac{d}{dx_k}\end{bmatrix}$, $\overset{\rightarrow y}{dg}(x)$ , $\overset{\rightarrow y}{dg^2}(x)$ (directional derivatives §D.1), $\log x$ , $\operatorname{sgn} x$ , $x/y$ (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)

$$\frac{d}{dx}x^{-1}\triangleq\nabla_x\,\mathbf{1}^T\delta(x)^{-1}\mathbf{1} \tag{1634}$$

- For $A$ a scalar or matrix, we have the Taylor series [55, §3.6]

$$e^A\triangleq\sum_{k=0}^{\infty}\frac{1}{k!}A^k \tag{1635}$$

  Further, [249, §5.4]

$$e^A\succ 0 \qquad \forall\,A\in\mathbb{S}^m \tag{1636}$$

- For all square $A$ and integer $k$

$$\det{}^k A=\det A^k \tag{1637}$$

## D.2.1 algebraic

| | |
|---|---|
| $\nabla_x\, x = \nabla_x\, x^T = I \in \mathbb{R}^{k\times k}$ | $\nabla_X X = \nabla_X X^T \triangleq I \in \mathbb{R}^{K\times L\times K\times L}$ (identity) |
| $\nabla_x (Ax - b) = A^T$ | |
| $\nabla_x \left(x^TA - b^T\right) = A$ | |
| $\nabla_x (Ax-b)^T(Ax-b) = 2A^T(Ax-b)$ | |
| $\nabla_x^2 (Ax-b)^T(Ax-b) = 2A^TA$ | |
| $\nabla_x \left(x^TAx + 2x^TBy + y^TCy\right) = (A+A^T)x + 2By$ | |
| $\nabla_x^2 \left(x^TAx + 2x^TBy + y^TCy\right) = A + A^T$ | |
| | $\nabla_X\, a^TXb = \nabla_X\, b^TX^Ta = ab^T$ |
| | $\nabla_X\, a^TX^2b = X^Tab^T + ab^TX^T$ |
| | $\nabla_X\, a^TX^{-1}b = -X^{-T}ab^TX^{-T}$ |
| | $\nabla_X (X^{-1})_{kl} = \dfrac{\partial X^{-1}}{\partial X_{kl}} = -X^{-1}e_k e_l^T X^{-1}$, *confer* (1574)(1633) |
| $\nabla_x\, a^Tx^Txb = 2xa^Tb$ | $\nabla_X\, a^TX^TXb = X(ab^T + ba^T)$ |
| $\nabla_x\, a^Txx^Tb = (ab^T + ba^T)x$ | $\nabla_X\, a^TXX^Tb = (ab^T + ba^T)X$ |
| $\nabla_x\, a^Tx^Txa = 2xa^Ta$ | $\nabla_X\, a^TX^TXa = 2Xaa^T$ |
| $\nabla_x\, a^Txx^Ta = 2aa^Tx$ | $\nabla_X\, a^TXX^Ta = 2aa^TX$ |
| $\nabla_x\, a^Tyx^Tb = ba^Ty$ | $\nabla_X\, a^TYX^Tb = ba^TY$ |
| $\nabla_x\, a^Ty^Txb = yb^Ta$ | $\nabla_X\, a^TY^TXb = Yab^T$ |
| $\nabla_x\, a^Txy^Tb = ab^Ty$ | $\nabla_X\, a^TXY^Tb = ab^TY$ |
| $\nabla_x\, a^Tx^Tyb = ya^Tb$ | $\nabla_X\, a^TX^TYb = Yba^T$ |

**algebraic** continued

$$\frac{d}{dt}(X+tY) = Y$$

$$\frac{d}{dt}B^T(X+tY)^{-1}A = -B^T(X+tY)^{-1}Y(X+tY)^{-1}A$$
$$\frac{d}{dt}B^T(X+tY)^{-T}A = -B^T(X+tY)^{-T}Y^T(X+tY)^{-T}A$$
$$\frac{d}{dt}B^T(X+tY)^{\mu}A = \ldots, \quad -1 \leq \mu \leq 1, \quad X, Y \in \mathbb{S}_+^M$$

$$\frac{d^2}{dt^2}B^T(X+tY)^{-1}A = 2B^T(X+tY)^{-1}Y(X+tY)^{-1}Y(X+tY)^{-1}A$$

$$\frac{d}{dt}\big((X+tY)^TA(X+tY)\big) = Y^TAX + X^TAY + 2\,t\,Y^TAY$$

$$\frac{d^2}{dt^2}\big((X+tY)^TA(X+tY)\big) = 2\,Y^TAY$$

$$\frac{d}{dt}\big((X+tY)A(X+tY)\big) = YAX + XAY + 2\,t\,YAY$$

$$\frac{d^2}{dt^2}\big((X+tY)A(X+tY)\big) = 2\,YAY$$

## D.2.2 trace Kronecker

$$\nabla_{\mathrm{vec}\,X}\operatorname{tr}(AXBX^T) \;=\; \nabla_{\mathrm{vec}\,X}\operatorname{vec}(X)^T(B^T\otimes A)\operatorname{vec}X \;=\; (B\otimes A^T + B^T\otimes A)\operatorname{vec}X$$

$$\nabla^2_{\mathrm{vec}\,X}\operatorname{tr}(AXBX^T) \;=\; \nabla^2_{\mathrm{vec}\,X}\operatorname{vec}(X)^T(B^T\otimes A)\operatorname{vec}X \;=\; B\otimes A^T + B^T\otimes A$$

### D.2.3 trace

| | |
|---|---|
| $\nabla_x \, \mu \, x = \mu I$ | $\nabla_X \operatorname{tr} \mu X = \nabla_X \, \mu \operatorname{tr} X = \mu I$ |
| $\nabla_x \, \mathbf{1}^T \delta(x)^{-1} \mathbf{1} = \frac{d}{dx} x^{-1} = -x^{-2}$<br>$\nabla_x \, \mathbf{1}^T \delta(x)^{-1} y = -\delta(x)^{-2} y$ | $\nabla_X \operatorname{tr} X^{-1} = -X^{-2T}$<br>$\nabla_X \operatorname{tr}(X^{-1} Y) = \nabla_X \operatorname{tr}(Y X^{-1}) = -X^{-T} Y^T X^{-T}$ |
| $\frac{d}{dx} x^{\mu} = \mu x^{\mu - 1}$ | $\nabla_X \operatorname{tr} X^{\mu} = \mu X^{\mu - 1} \,, \qquad X \in \mathbb{S}^M$ |
| | $\nabla_X \operatorname{tr} X^j = j X^{(j-1)T}$ |
| $\nabla_x (b - a^T x)^{-1} = (b - a^T x)^{-2} a$ | $\nabla_X \operatorname{tr}\big((B - AX)^{-1}\big) = \big((B - AX)^{-2} A\big)^T$ |
| $\nabla_x (b - a^T x)^{\mu} = -\mu (b - a^T x)^{\mu - 1} a$ | |
| $\nabla_x \, x^T y = \nabla_x \, y^T x = y$ | $\nabla_X \operatorname{tr}(X^T Y) = \nabla_X \operatorname{tr}(Y X^T) = \nabla_X \operatorname{tr}(Y^T X) = \nabla_X \operatorname{tr}(X Y^T) = Y$ |
| | $\nabla_X \operatorname{tr}(A X B X^T) = \nabla_X \operatorname{tr}(X B X^T A) = A^T X B^T \; + \; A X B$<br>$\nabla_X \operatorname{tr}(A X B X) \;\; = \nabla_X \operatorname{tr}(X B X A) \;\; = A^T X^T B^T + B^T X^T A^T$ |
| | $\nabla_X \operatorname{tr}(A X A X A X) = \nabla_X \operatorname{tr}(X A X A X A) = 3(A X A X A)^T$ |
| | $\nabla_X \operatorname{tr}(Y X^k) = \nabla_X \operatorname{tr}(X^k Y) = \sum_{i=0}^{k-1} \big(X^i Y X^{k-1-i}\big)^T$ |
| | $\nabla_X \operatorname{tr}(Y^T X X^T Y) = \nabla_X \operatorname{tr}(X^T Y Y^T X) = 2\, Y Y^T X$<br>$\nabla_X \operatorname{tr}(Y^T X^T X Y) \; = \nabla_X \operatorname{tr}(X Y Y^T X^T) = 2 X Y Y^T$ |
| | $\nabla_X \operatorname{tr}\big((X + Y)^T (X + Y)\big) = 2(X + Y)$<br>$\nabla_X \operatorname{tr}((X + Y)(X + Y)) \;\; = 2(X + Y)^T$ |
| | $\nabla_X \operatorname{tr}(A^T X B) \;\; = \nabla_X \operatorname{tr}(X^T A B^T) \;\; = \qquad A B^T$<br>$\nabla_X \operatorname{tr}(A^T X^{-1} B) = \nabla_X \operatorname{tr}(X^{-T} A B^T) = -X^{-T} A B^T X^{-T}$ |
| | $\nabla_X \, a^T X b \;\; = \nabla_X \operatorname{tr}(b a^T X) \;\; = \nabla_X \operatorname{tr}(X b a^T) \;\; = a b^T$<br>$\nabla_X \, b^T X^T a \; = \nabla_X \operatorname{tr}(X^T a b^T) = \nabla_X \operatorname{tr}(a b^T X^T) = a b^T$<br>$\nabla_X \, a^T X^{-1} b \; = \nabla_X \operatorname{tr}(X^{-T} a b^T) = -X^{-T} a b^T X^{-T}$<br>$\nabla_X \, a^T X^{\mu} b \; = \ldots$ |

**trace** continued

$$\frac{d}{dt}\operatorname{tr} g(X+t\,Y) = \operatorname{tr}\frac{d}{dt}\, g(X+t\,Y) \qquad [151,\ \text{p.491}]$$

$$\frac{d}{dt}\operatorname{tr}(X+t\,Y) = \operatorname{tr} Y$$

$$\frac{d}{dt}\operatorname{tr}^j(X+t\,Y) = j\operatorname{tr}^{j-1}(X+t\,Y)\operatorname{tr} Y$$

$$\frac{d}{dt}\operatorname{tr}(X+t\,Y)^j = j\operatorname{tr}((X+t\,Y)^{j-1}\,Y) \qquad (\forall\, j)$$

$$\frac{d}{dt}\operatorname{tr}((X+t\,Y)Y) = \operatorname{tr} Y^2$$

$$\frac{d}{dt}\operatorname{tr}\big((X+t\,Y)^k\,Y\big) = \frac{d}{dt}\operatorname{tr}(Y(X+t\,Y)^k) = k\operatorname{tr}\big((X+t\,Y)^{k-1}\,Y^2\big)\ , \quad k\in\{0,1,2\}$$

$$\frac{d}{dt}\operatorname{tr}\big((X+t\,Y)^k\,Y\big) = \frac{d}{dt}\operatorname{tr}(Y(X+t\,Y)^k) = \operatorname{tr}\sum_{i=0}^{k-1}(X+t\,Y)^i\,Y(X+t\,Y)^{k-1-i}\,Y$$

$$\begin{aligned}
\frac{d}{dt}\operatorname{tr}((X+t\,Y)^{-1}\,Y) &= -\operatorname{tr}((X+t\,Y)^{-1}\,Y(X+t\,Y)^{-1}\,Y)\\
\frac{d}{dt}\operatorname{tr}\big(B^T(X+t\,Y)^{-1}A\big) &= -\operatorname{tr}\big(B^T(X+t\,Y)^{-1}\,Y(X+t\,Y)^{-1}A\big)\\
\frac{d}{dt}\operatorname{tr}\big(B^T(X+t\,Y)^{-T}A\big) &= -\operatorname{tr}\big(B^T(X+t\,Y)^{-T}\,Y^T(X+t\,Y)^{-T}A\big)\\
\frac{d}{dt}\operatorname{tr}\big(B^T(X+t\,Y)^{-k}A\big) &= \ldots, \quad k>0\\
\frac{d}{dt}\operatorname{tr}\big(B^T(X+t\,Y)^{\mu}A\big) &= \ldots, \quad -1\le\mu\le 1, \quad X,Y\in\mathbb{S}_+^M
\end{aligned}$$

$$\frac{d^2}{dt^2}\operatorname{tr}\big(B^T(X+t\,Y)^{-1}A\big) = 2\operatorname{tr}\big(B^T(X+t\,Y)^{-1}\,Y(X+t\,Y)^{-1}\,Y(X+t\,Y)^{-1}A\big)$$

$$\frac{d}{dt}\operatorname{tr}\big((X+t\,Y)^TA(X+t\,Y)\big) = \operatorname{tr}\big(Y^TAX + X^TAY + 2\,t\,Y^TAY\big)$$

$$\frac{d^2}{dt^2}\operatorname{tr}\big((X+t\,Y)^TA(X+t\,Y)\big) = 2\operatorname{tr}\big(Y^TAY\big)$$

$$\frac{d}{dt}\operatorname{tr}\big((X+t\,Y)A(X+t\,Y)\big) = \operatorname{tr}(YAX + XAY + 2\,t\,YAY)$$

$$\frac{d^2}{dt^2}\operatorname{tr}\big((X+t\,Y)A(X+t\,Y)\big) = 2\operatorname{tr}(YAY)$$

### D.2.4 log determinant

$x \succ 0$, $\det X > 0$ on some neighborhood of $X$, and $\det(X + tY) > 0$ on some open interval of $t$; otherwise, $\log(\;)$ would be discontinuous.

| | |
|---|---|
| $\frac{d}{dx}\log x = x^{-1}$ | $\nabla_X \log\det X = X^{-T}$ |
| | $\nabla_X^2 \log\det(X)_{kl} = \dfrac{\partial X^{-T}}{\partial X_{kl}} = -\left(X^{-1}e_k e_l^T X^{-1}\right)^T$, *confer* (1591)(1633) |
| $\frac{d}{dx}\log x^{-1} = -x^{-1}$ | $\nabla_X \log\det X^{-1} = -X^{-T}$ |
| $\frac{d}{dx}\log x^{\mu} = \mu x^{-1}$ | $\nabla_X \log\det^{\mu} X = \mu X^{-T}$ |
| | $\nabla_X \log\det X^{\mu} = \mu X^{-T}$ |
| | $\nabla_X \log\det X^k = \nabla_X \log\det^k X = kX^{-T}$ |
| | $\nabla_X \log\det^{\mu}(X + tY) = \mu(X + tY)^{-T}$ |
| $\nabla_x \log(a^Tx + b) = a\frac{1}{a^Tx+b}$ | $\nabla_X \log\det(AX + B) = A^T(AX + B)^{-T}$ |
| | $\nabla_X \log\det(I \pm A^TXA) = \pm A(I \pm A^TXA)^{-T}A^T$ |
| | $\nabla_X \log\det(X + tY)^k = \nabla_X \log\det^k(X + tY) = k(X + tY)^{-T}$ |
| | $\frac{d}{dt}\log\det(X + tY) = \operatorname{tr}\left((X + tY)^{-1}Y\right)$ |
| | $\frac{d^2}{dt^2}\log\det(X + tY) = -\operatorname{tr}\left((X + tY)^{-1}Y(X + tY)^{-1}Y\right)$ |
| | $\frac{d}{dt}\log\det(X + tY)^{-1} = -\operatorname{tr}\left((X + tY)^{-1}Y\right)$ |
| | $\frac{d^2}{dt^2}\log\det(X + tY)^{-1} = \operatorname{tr}\left((X + tY)^{-1}Y(X + tY)^{-1}Y\right)$ |
| | $\frac{d}{dt}\log\det\left(\delta(A(x + ty) + a)^2 + \mu I\right)$ $= \operatorname{tr}\left(\left(\delta(A(x + ty) + a)^2 + \mu I\right)^{-1} 2\delta(A(x + ty) + a)\delta(Ay)\right)$ |

### D.2.5 determinant

$$\nabla_X \det X = \nabla_X \det X^T = \det(X) X^{-T}$$

$$\nabla_X \det X^{-1} = -\det(X^{-1}) X^{-T} = -\det(X)^{-1} X^{-T}$$

$$\nabla_X \det^\mu X = \mu \det^\mu(X) X^{-T}$$

$$\nabla_X \det X^\mu = \mu \det(X^\mu) X^{-T}$$

$$\nabla_X \det X^k = k \det^{k-1}(X)\big(\operatorname{tr}(X) I - X^T\big)\ , \qquad X \in \mathbb{R}^{\mathbf{2}\times\mathbf{2}}$$

$$\nabla_X \det X^k = \nabla_X \det^k X = k \det(X^k) X^{-T} = k \det^k(X) X^{-T}$$

$$\nabla_X \det^\mu(X + t\,Y) = \mu \det^\mu(X + t\,Y)(X + t\,Y)^{-T}$$

$$\nabla_X \det(X + t\,Y)^k = \nabla_X \det^k(X + t\,Y) = k \det^k(X + t\,Y)(X + t\,Y)^{-T}$$

$$\tfrac{d}{dt} \det(X + t\,Y) = \det(X + t\,Y) \operatorname{tr}((X + t\,Y)^{-1} Y)$$

$$\tfrac{d^2}{dt^2} \det(X + t\,Y) = \det(X + t\,Y)\big(\operatorname{tr}^2((X + t\,Y)^{-1} Y) - \operatorname{tr}((X + t\,Y)^{-1} Y (X + t\,Y)^{-1} Y)\big)$$

$$\tfrac{d}{dt} \det(X + t\,Y)^{-1} = -\det(X + t\,Y)^{-1} \operatorname{tr}((X + t\,Y)^{-1} Y)$$

$$\tfrac{d^2}{dt^2} \det(X + t\,Y)^{-1} = \det(X + t\,Y)^{-1}\big(\operatorname{tr}^2((X + t\,Y)^{-1} Y) + \operatorname{tr}((X + t\,Y)^{-1} Y (X + t\,Y)^{-1} Y)\big)$$

$$\tfrac{d}{dt} \det^\mu(X + t\,Y) = \mu \det^\mu(X + t\,Y) \operatorname{tr}((X + t\,Y)^{-1} Y)$$

### D.2.6 logarithmic

Matrix logarithm.

$$\frac{d}{dt}\log(X+t\,Y)^{\mu} = \mu Y(X+t\,Y)^{-1} = \mu(X+t\,Y)^{-1}Y\,, \qquad XY = YX$$

$$\frac{d}{dt}\log(I-t\,Y)^{\mu} = -\mu Y(I-t\,Y)^{-1} = -\mu(I-t\,Y)^{-1}Y \qquad \text{[151, p.493]}$$

### D.2.7 exponential

Matrix exponential. [55, §3.6, §4.5] [249, §5.4]

$$\nabla_X e^{\operatorname{tr}(Y^TX)} = \nabla_X \det e^{Y^TX} = e^{\operatorname{tr}(Y^TX)}Y \qquad (\forall\, X, Y)$$

$$\nabla_X \operatorname{tr} e^{YX} = e^{Y^TX^T}Y^T = Y^T e^{X^TY^T}$$

log-sum-exp & geometric mean [46, p.74]...

$$\frac{d^j}{dt^j}e^{\operatorname{tr}(X+t\,Y)} = e^{\operatorname{tr}(X+t\,Y)}\operatorname{tr}^j(Y)$$

$$\frac{d}{dt}e^{tY} = e^{tY}Y = Ye^{tY}$$

$$\frac{d}{dt}e^{X+t\,Y} = e^{X+t\,Y}Y = Ye^{X+t\,Y}\,, \qquad XY = YX$$

$$\frac{d^2}{dt^2}e^{X+t\,Y} = e^{X+t\,Y}Y^2 = Ye^{X+t\,Y}Y = Y^2e^{X+t\,Y}\,, \qquad XY = YX$$

# Appendix E

# Projection

For any $A \in \mathbb{R}^{m\times n}$, the *pseudoinverse* [150, §7.3, prob.9] [182, §6.12, prob.19] [110, §5.5.4] [249, App.A]

$$A^\dagger \triangleq \lim_{t\to 0^+} (A^TA + t\,I)^{-1}A^T = \lim_{t\to 0^+} A^T(AA^T + t\,I)^{-1} \in \mathbb{R}^{n\times m} \tag{1638}$$

is a unique matrix solving $\underset{X}{\text{minimize}}\ \|AX - I\|_F^2$ . For any $t > 0$

$$I - A(A^TA + t\,I)^{-1}A^T = t(AA^T + t\,I)^{-1} \tag{1639}$$

Equivalently, pseudoinverse $A^\dagger$ is that unique matrix satisfying the *Moore-Penrose conditions*: [152, §1.3] [282]

| | |
|---|---|
| 1. $AA^\dagger A = A$ | 3. $(AA^\dagger)^T = AA^\dagger$ |
| 2. $A^\dagger AA^\dagger = A^\dagger$ | 4. $(A^\dagger A)^T = A^\dagger A$ |

which are necessary and sufficient to establish the pseudoinverse whose principal action is to injectively map $\mathcal{R}(A)$ onto $\mathcal{R}(A^T)$. Taken rowwise, these conditions are respectively necessary and sufficient for $AA^\dagger$ to be the orthogonal projector on $\mathcal{R}(A)$, and for $A^\dagger A$ to be the orthogonal projector on $\mathcal{R}(A^T)$.

Range and nullspace of the pseudoinverse [195] [246, §III.1, exer.1]

$$\mathcal{R}(A^\dagger) = \mathcal{R}(A^T)\,, \qquad \mathcal{R}(A^{\dagger T}) = \mathcal{R}(A) \tag{1640}$$

$$\mathcal{N}(A^\dagger) = \mathcal{N}(A^T)\,, \qquad \mathcal{N}(A^{\dagger T}) = \mathcal{N}(A) \tag{1641}$$

can be derived by singular value decomposition (§A.6).

The following relations reliably hold without qualification:

a. $A^{T\dagger} = A^{\dagger T}$
b. $A^{\dagger\dagger} = A$
c. $(AA^T)^\dagger = A^{\dagger T}A^\dagger$
d. $(A^TA)^\dagger = A^\dagger A^{\dagger T}$
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.* [119] [42] [176, exer.7.23]
For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$

$$(AB)^\dagger = B^\dagger A^\dagger \tag{1642}$$

if and only if

$$\mathcal{R}(A^TAB) \subseteq \mathcal{R}(B) \quad \textbf{and} \quad \mathcal{R}(BB^TA^T) \subseteq \mathcal{R}(A^T) \tag{1643}$$

◇

$U^T = U^\dagger$ for orthonormal (including the orthogonal) matrices $U$. So, for orthonormal matrices $U, Q$ and arbitrary $A$

$$(UAQ^T)^\dagger = QA^\dagger U^T \tag{1644}$$

**E.0.0.0.2 Exercise.** *Kronecker inverse.*
Prove:

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger \tag{1645}$$

▼

### E.0.1 Logical deductions

When $A$ is invertible, $A^\dagger = A^{-1}$; so $A^\dagger A = AA^\dagger = I$. Otherwise, for $A \in \mathbb{R}^{m\times n}$ [103, §5.3.3.1] [176, §7] [219]

g. $A^\dagger A = I$, $\quad A^\dagger = (A^TA)^{-1}A^T$, $\quad \operatorname{rank} A = n$
h. $AA^\dagger = I$, $\quad A^\dagger = A^T(AA^T)^{-1}$, $\quad \operatorname{rank} A = m$
i. $A^\dagger A\omega = \omega$, $\quad \omega \in \mathcal{R}(A^T)$
j. $AA^\dagger \upsilon = \upsilon$, $\quad \upsilon \in \mathcal{R}(A)$
k. $A^\dagger A = AA^\dagger$, $\quad A$ normal
l. $A^{k\dagger} = A^{\dagger k}$, $\quad A$ normal, $k$ an integer

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** $A^{\dagger T} = A A^\dagger A^{\dagger T}$

When $A$ is symmetric, $A^\dagger$ is symmetric and (§A.6)

$$A \succeq 0 \;\Leftrightarrow\; A^\dagger \succeq 0 \tag{1646}$$

**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.

For any shape of matrix $A$ of any rank, and given any vector $b$ that may or may not be in $\mathcal{R}(A)$, we wish to find a best Euclidean solution $x^\star$ to

$$Ax = b \tag{1647}$$

(more generally, $Ax \approx b$ given arbitrary matrices) by solving

$$\underset{x}{\text{minimize}} \;\; \|Ax - b\|^2 \tag{1648}$$

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)

$$A^T A\, x = A^T b \tag{1649}$$

*normal* because error vector $b - Ax$ is perpendicular to $\mathcal{R}(A)$; *id est*, $A^T(b - Ax) = \mathbf{0}$. Given any matrix $A$ and any vector $b$, the normal equation is solvable exactly; always so, because $\mathcal{R}(A^T A) = \mathcal{R}(A^T)$ and $A^T b \in \mathcal{R}(A^T)$.

When $A$ is skinny-or-square full-rank, normal equation (1649) can be solved exactly by inversion:

$$x^\star = (A^T A)^{-1} A^T b \;\equiv\; A^\dagger b \tag{1650}$$

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: (1638)

$$x^\star = \lim_{t\to 0^+}(A^TA + t\,I)^{-1}A^Tb \;=\; A^\dagger b \tag{1651}$$

invertible for any positive value of $t$ by (1255). The exact inversion (1650) and this pseudoinverse solution (1651) each solve

$$\lim_{t\to 0^+}\ \underset{x}{\text{minimize}}\ \|Ax-b\|^2 \,+\, t\,\|x\|^2 \tag{1652}$$

simultaneously providing least squares solution to (1648) and the classical *least norm* solution[E.1] [249, App.A.4] (*confer* §E.5.0.0.5)

$$\begin{array}{cl}\underset{x}{\text{arg minimize}} & \|x\|^2\\ \text{subject to} & Ax = AA^\dagger b\end{array} \tag{1653}$$

where $AA^\dagger b$ is the orthogonal projection of vector $b$ on $\mathcal{R}(A)$. $\square$

## E.1 Idempotent matrices

Projection matrices are square and defined by *idempotence*, $P^2 = P$ ; [249, §2.6] [152, §1.3] equivalent to the condition, $P$ be diagonalizable [150, §3.3, prob.3] with eigenvalues $\phi_i \in \{0,1\}$. [301, §4.1, thm.4.1] Idempotent matrices are not necessarily symmetric. The transpose of an idempotent matrix remains idempotent; $P^TP^T = P^T$. Solely excepting $P = I$ , all projection matrices are neither orthogonal (§B.5) or invertible. [249, §3.4] The collection of all projection matrices of particular dimension does not form a convex set.

Suppose we wish to project nonorthogonally (*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:

$$P = A(A^\dagger + BZ^T) \in \mathbb{R}^{m\times m} \tag{1654}$$

[E.1] This means: optimal solutions of lesser norm than the so-called *least norm* solution (1653) 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 (1652).

where $\mathcal{R}(P)=\mathcal{R}(A)$ ,[E.2] $B\in\mathbb{R}^{n\times k}$ for $k\in\{1\ldots m\}$ is otherwise arbitrary, and $Z\in\mathbb{R}^{m\times k}$ is any matrix whose range is in $\mathcal{N}(A^T)$ ; *id est*,

$$A^T Z = A^\dagger Z = \mathbf{0} \tag{1655}$$

Evidently, the collection of nonorthogonal projectors projecting on $\mathcal{R}(A)$ is an affine subset

$$\mathcal{P}_k = \left\{A(A^\dagger + BZ^T) \mid B\in\mathbb{R}^{n\times k}\right\} \tag{1656}$$

When matrix $A$ in (1654) is skinny full-rank ($A^\dagger A=I$) or has orthonormal columns ($A^TA=I$), either property leads to a biorthogonal characterization of nonorthogonal projection:

### E.1.1 Biorthogonal characterization

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 $Q^TU=I$ ; the *biorthogonal decomposition* of $P$ being (*confer* (1654))

$$P = UQ^T\,, \qquad Q^TU = I \tag{1657}$$

where[E.3] (§B.1.1.1)

$$\mathcal{R}(P)=\mathcal{R}(U)\ , \qquad \mathcal{N}(P)=\mathcal{N}(Q^T) \tag{1658}$$

and where generally (*confer* (1683))[E.4]

$$P^T \neq P\,, \qquad P^\dagger \neq P\,, \qquad \|P\|_2 \neq 1\,, \qquad P \nsucceq 0 \tag{1659}$$

and $P$ is not nonexpansive (1684).

---

[E.2] **Proof.** $\mathcal{R}(P)\subseteq\mathcal{R}(A)$ is obvious [249, §3.6]. By (119) and (120),

$$\begin{aligned}
\mathcal{R}(A^\dagger + BZ^T) &= \{(A^\dagger + BZ^T)y \mid y\in\mathbb{R}^m\}\\
&\supseteq \{(A^\dagger + BZ^T)y \mid y\in\mathcal{R}(A)\} = \mathcal{R}(A^T)\\
\mathcal{R}(P) &= \{A(A^\dagger + BZ^T)y \mid y\in\mathbb{R}^m\}\\
&\supseteq \{A(A^\dagger + BZ^T)y \mid (A^\dagger + BZ^T)y\in\mathcal{R}(A^T)\} = \mathcal{R}(A) \qquad \blacklozenge
\end{aligned}$$

[E.3] **Proof.** Obviously, $\mathcal{R}(P)\subseteq\mathcal{R}(U)$. Because $Q^TU=I$

$$\begin{aligned}
\mathcal{R}(P) &= \{UQ^Tx \mid x\in\mathbb{R}^m\}\\
&\supseteq \{UQ^TUy \mid y\in\mathbb{R}^k\} = \mathcal{R}(U) \qquad \blacklozenge
\end{aligned}$$

[E.4] Orthonormal decomposition (1680) (*confer* §E.3.4) is a special case of biorthogonal decomposition (1657) characterized by (1683). So, these characteristics (1659) are not necessary conditions for biorthogonality.

($\Leftarrow$) To verify assertion (1657) we observe: because idempotent matrices are diagonalizable (§A.5), [150, §3.3, prob.3] they must have the form (1339)

$$P \;=\; S\Phi S^{-1} \;=\; \sum_{i=1}^{m} \phi_i\, s_i w_i^T \;=\; \sum_{i=1}^{k\,\leq\, m} s_i w_i^T \tag{1660}$$

that is a sum of $k=\operatorname{rank} P$ independent projector dyads (idempotent dyads, §B.1.1, §E.6.2.1) where $\phi_i \in \{0,1\}$ are the eigenvalues of $P$ [301, §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.[E.5] Therefore

$$U \;\triangleq\; S(:\,,1\!:\!k) \;=\; \big[\, s_1 \cdots s_k \,\big] \in \mathbb{R}^{m\times k} \tag{1661}$$

is the full-rank matrix $S\in\mathbb{R}^{m\times m}$ having $m-k$ columns truncated (corresponding to 0 eigenvalues), while

$$Q^T \triangleq S^{-1}(1\!:\!k\,,:) = \begin{bmatrix} w_1^T \\ \vdots \\ w_k^T \end{bmatrix} \in \mathbb{R}^{k\times m} \tag{1662}$$

is matrix $S^{-1}$ having the corresponding $m-k$ rows truncated. By the 0 *eigenvalues theorem* (§A.7.3.0.1), $\mathcal{R}(U)=\mathcal{R}(P)$, $\mathcal{R}(Q)=\mathcal{R}(P^T)$, and

$$\begin{aligned} \mathcal{R}(P) \;&=\; \operatorname{span}\{s_i \mid \phi_i = 1 \;\; \forall\, i\} \\ \mathcal{N}(P) \;&=\; \operatorname{span}\{s_i \mid \phi_i = 0 \;\; \forall\, i\} \\ \mathcal{R}(P^T) \;&=\; \operatorname{span}\{w_i \mid \phi_i = 1 \;\; \forall\, i\} \\ \mathcal{N}(P^T) \;&=\; \operatorname{span}\{w_i \mid \phi_i = 0 \;\; \forall\, i\} \end{aligned} \tag{1663}$$

Thus biorthogonality $Q^T U = I$ is a necessary condition for idempotence, and so the collection of nonorthogonal projectors projecting on $\mathcal{R}(U)$ is the affine subset $\mathcal{P}_k = U\mathcal{Q}_k^T$ where $\mathcal{Q}_k = \{Q \mid Q^T U = I\,,\; Q\in\mathbb{R}^{m\times k}\}$.

($\Rightarrow$) Biorthogonality is a sufficient condition for idempotence;

$$P^2 \;=\; \sum_{i=1}^{k} s_i w_i^T \;\sum_{j=1}^{k} s_j w_j^T \;=\; P \tag{1664}$$

*id est*, if the cross-products are annihilated, then $P^2 = P$. ♦

[E.5] Eigenvectors of a real matrix corresponding to real eigenvalues must be real. (§A.5.0.0.1)

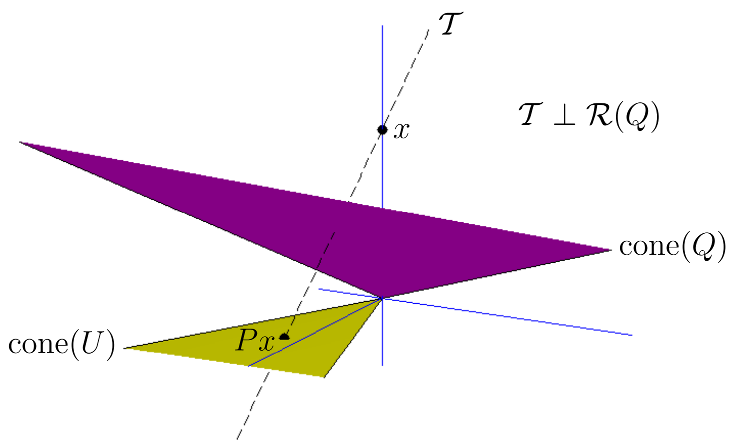



Figure 118: Nonorthogonal projection of $x \in \mathbb{R}^{\mathbf{3}}$ on $\mathcal{R}(U) = \mathbb{R}^{\mathbf{2}}$ under biorthogonality condition; *id est*, $Px = UQ^T x$ such that $Q^T U = I$. Any point along imaginary line $\mathcal{T}$ connecting $x$ to $Px$ will be projected nonorthogonally on $Px$ with respect to horizontal plane constituting $\mathbb{R}^{\mathbf{2}}$ 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 $\mathcal{R}(U)$).

Nonorthogonal projection of $x$ on $\mathcal{R}(P)$ has expression like a biorthogonal expansion,

$$Px \;=\; UQ^T x \;=\; \sum_{i=1}^{k} w_i^T x\, s_i \tag{1665}$$

When the domain is restricted to the range of $P$, say $x = U\xi$ for $\xi \in \mathbb{R}^k$, then $x = Px = UQ^T U\xi = U\xi$ 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. The direction of nonorthogonal projection is orthogonal to $\mathcal{R}(Q) \Leftrightarrow Q^T U = I$; *id est*, for $Px \in \mathcal{R}(U)$

$$Px - x \;\perp\; \mathcal{R}(Q) \text{ in } \mathbb{R}^m \tag{1666}$$

**E.1.1.0.1 Example.** *Illustration of nonorthogonal projector.*
Figure **118** shows $\operatorname{cone}(U)$, the conic hull of the columns of

$$U = \begin{bmatrix} 1 & 1 \\ -0.5 & 0.3 \\ 0 & 0 \end{bmatrix} \tag{1667}$$

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 (1654)

$$\begin{aligned} Q = U^{\dagger T} + ZB^T &= \begin{bmatrix} 0.3750 & 0.6250 \\ -1.2500 & 1.2500 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [\,0.5 \;\; 0.5\,] \\ &= \begin{bmatrix} 0.3750 & 0.6250 \\ -1.2500 & 1.2500 \\ 0.5000 & 0.5000 \end{bmatrix} \end{aligned} \tag{1668}$$

where $Z \in \mathcal{N}(U^T)$ and matrix $B$ is selected arbitrarily; *id est*, $Q^T U = I$ because $U$ is full-rank.

Direction of projection on $\mathcal{R}(U)$ is orthogonal to $\mathcal{R}(Q)$. Any point along line $\mathcal{T}$ in the figure, for example, will have the same projection. Were matrix $Z$ instead equal to $\mathbf{0}$, then $\operatorname{cone}(Q)$ would become the relative dual to $\operatorname{cone}(U)$ (sharing the same affine hull; §2.13.8, *confer* Figure **43**(a)) In that case, projection $Px = UU^{\dagger}x$ of $x$ on $\mathcal{R}(U)$ becomes orthogonal projection (and unique minimum-distance). □

### E.1.2 Idempotence summary

Nonorthogonal subspace-projector $P$ is a linear operator defined by idempotence or biorthogonal decomposition (1657), but characterized not by symmetry nor positive semidefiniteness nor nonexpansivity (1684).

## 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

$$I-P \;=\; S(I-\Phi)S^{-1} \;=\; \sum_{i=1}^{m}(1-\phi_i)s_i w_i^T \tag{1669}$$

where $(1-\phi_i)\in\{1,0\}$ are the eigenvalues of $I-P$ (1256) whose eigenvectors $s_i\,, w_i$ are identical to those of $P$ in (1660). A consequence of that complementary relationship of eigenvalues is the fact, [259, §2] [255, §2] for subspace projector $P=P^2\in\mathbb{R}^{m\times m}$

$$\begin{aligned}
\mathcal{R}(P) &= \operatorname{span}\{s_i \mid \phi_i=1 \;\;\forall\, i\} = \operatorname{span}\{s_i \mid (1-\phi_i)=0 \;\;\forall\, i\} = \mathcal{N}(I-P)\\
\mathcal{N}(P) &= \operatorname{span}\{s_i \mid \phi_i=0 \;\;\forall\, i\} = \operatorname{span}\{s_i \mid (1-\phi_i)=1 \;\;\forall\, i\} = \mathcal{R}(I-P)\\
\mathcal{R}(P^T) &= \operatorname{span}\{w_i \mid \phi_i=1 \;\;\forall\, i\} = \operatorname{span}\{w_i \mid (1-\phi_i)=0 \;\;\forall\, i\} = \mathcal{N}(I-P^T)\\
\mathcal{N}(P^T) &= \operatorname{span}\{w_i \mid \phi_i=0 \;\;\forall\, i\} = \operatorname{span}\{w_i \mid (1-\phi_i)=1 \;\;\forall\, i\} = \mathcal{R}(I-P^T)
\end{aligned} \tag{1670}$$

that is easy to see from (1660) and (1669). 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 $\mathcal{R}(P)$ [166, §3.3] is equivalent to $\mathcal{N}(P)$ ;[E.6] *id est*,

$$\mathcal{R}(P)\oplus\mathcal{N}(P) \;=\; \mathcal{R}(P^T)\oplus\mathcal{N}(P^T) \;=\; \mathcal{R}(P^T)\oplus\mathcal{N}(P) \;=\; \mathcal{R}(P)\oplus\mathcal{N}(P^T) \;=\; \mathbb{R}^m \tag{1671}$$

because $\mathcal{R}(P)\oplus\mathcal{R}(I-P)=\mathbb{R}^m$. For idempotent $P\in\mathbb{R}^{m\times m}$, consequently,

$$\operatorname{rank} P \;+\; \operatorname{rank}(I-P) = m \tag{1672}$$

[E.6] The same phenomenon occurs with symmetric (nonidempotent) matrices, for example. When the summands in $A\oplus B=\mathbb{R}^m$ are orthogonal vector spaces, the algebraic complement is the orthogonal complement.

**E.2.0.0.1 Theorem.** *Rank/Trace.* [301, §4.1, prob.9] (*confer* (1688))

$$\begin{gathered} P^2 = P \\ \Leftrightarrow \\ \operatorname{rank} P = \operatorname{tr} P \quad \textbf{and} \quad \operatorname{rank}(I-P) = \operatorname{tr}(I-P) \end{gathered} \tag{1673}$$

$\diamond$

### 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 $x \in \mathbb{R}^m$

$$Px \,+\, (I-P)x \,=\, x \tag{1674}$$

must hold where $\mathcal{R}(I-P) = \mathcal{N}(P)$ is the algebraic complement of $\mathcal{R}(P)$. The algebraic complement of closed convex cone $\mathcal{K}$, for example, is the negative dual cone $-\mathcal{K}^*$. (1792)

## 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 $\mathcal{R}(P)$ ; *id est*, for $P^2 = P \in \mathbb{S}^m$ and projection $Px \in \mathcal{R}(P)$

$$Px - x \;\perp\; \mathcal{R}(P) \text{ in } \mathbb{R}^m \tag{1675}$$

Perpendicularity is a necessary and sufficient condition for orthogonal projection on a subspace. [73, §4.9]

A condition equivalent to (1675) is: Norm of direction $x - Px$ is the infimum over all nonorthogonal projections of $x$ on $\mathcal{R}(P)$ ; [182, §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 (1654), and given $x \in \mathbb{R}^m$

$$\|x - Px\|_2 \;=\; \|x - AA^\dagger x\|_2 \;=\; \inf_{B \in \mathbb{R}^{n\times k}} \|x - A(A^\dagger + BZ^T)x\|_2 \tag{1676}$$

The infimum is attained for $\mathcal{R}(B) \subseteq \mathcal{N}(A)$ over any affine subset of nonorthogonal projectors (1656) indexed by $k$ .

Proof is straightforward: The vector 2-norm is a convex function. Setting gradient of the norm-square to $\mathbf{0}$, applying §D.2,

$$\begin{gathered}\left(A^TABZ^T - A^T(I-AA^\dagger)\right)xx^TA \;=\; \mathbf{0}\\ \Leftrightarrow\\ A^TABZ^Txx^TA = \mathbf{0}\end{gathered} \tag{1677}$$

because $A^T = A^TAA^\dagger$. 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 $\mathbb{R}^m$ orthogonally on $\mathcal{R}(A)$. Under either condition (1675) or (1676), 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 $A \in \mathbb{R}^{m\times n}$

$$\begin{array}{lcll} A^\dagger A & : & \mathbb{R}^n \text{ on } \mathcal{R}(A^\dagger A) & = \mathcal{R}(A^T)\\ AA^\dagger & : & \mathbb{R}^m \text{ on } \mathcal{R}(AA^\dagger) & = \mathcal{R}(A)\\ I - A^\dagger A & : & \mathbb{R}^n \text{ on } \mathcal{R}(I - A^\dagger A) & = \mathcal{N}(A)\\ I - AA^\dagger & : & \mathbb{R}^m \text{ on } \mathcal{R}(I - AA^\dagger) & = \mathcal{N}(A^T) \end{array} \tag{1678}$$

For completeness:[E.7] (1670)

$$\begin{array}{ll} \mathcal{N}(A^\dagger A) & = \mathcal{N}(A)\\ \mathcal{N}(AA^\dagger) & = \mathcal{N}(A^T)\\ \mathcal{N}(I - A^\dagger A) & = \mathcal{R}(A^T)\\ \mathcal{N}(I - AA^\dagger) & = \mathcal{R}(A) \end{array} \tag{1679}$$

[E.7] **Proof** is by singular value decomposition (§A.6.2): $\mathcal{N}(A^\dagger A) \subseteq \mathcal{N}(A)$ is obvious. Conversely, suppose $A^\dagger Ax = \mathbf{0}$. Then $x^TA^\dagger Ax = x^TQQ^Tx = \|Q^Tx\|^2 = 0$ where $A = U\Sigma Q^T$ is the subcompact singular value decomposition. Because $\mathcal{R}(Q) = \mathcal{R}(A^T)$, then $x \in \mathcal{N}(A)$ that implies $\mathcal{N}(A^\dagger A) \supseteq \mathcal{N}(A)$. ♦

### E.3.2 Orthogonal characterization

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)

$$P = QQ^T\,, \qquad Q^TQ = I \tag{1680}$$

where [249, §3.3] (1392)

$$\mathcal{R}(P) = \mathcal{R}(Q)\ , \qquad \mathcal{N}(P) = \mathcal{N}(Q^T) \tag{1681}$$

is an orthogonal projector projecting on $\mathcal{R}(Q)$ having, for $Px \in \mathcal{R}(Q)$ (*confer* (1666))

$$Px - x \ \perp\ \mathcal{R}(Q) \text{ in } \mathbb{R}^m \tag{1682}$$

From (1680), orthogonal projector $P$ is obviously positive semidefinite (§A.3.1.0.6); necessarily,

$$P^T = P\ , \qquad P^\dagger = P\ , \qquad \|P\|_2 = 1\,, \qquad P \succeq 0 \tag{1683}$$

and $\|Px\| = \|QQ^Tx\| = \|Q^Tx\|$ because $\|Qy\| = \|y\|\ \forall\, y \in \mathbb{R}^k$. All orthogonal projectors are therefore *nonexpansive* because

$$\sqrt{\langle Px\,,\,x\rangle}\ =\ \|Px\|\ =\ \|Q^Tx\|\ \leq\ \|x\| \quad \forall\, x \in \mathbb{R}^m \tag{1684}$$

the Bessel inequality, [73] [166] with equality when $x \in \mathcal{R}(Q)$.

From the diagonalization of idempotent matrices (1660) on page 586

$$P\ =\ S\Phi S^T\ =\ \sum_{i=1}^{m} \phi_i\, s_i s_i^T\ =\ \sum_{i=1}^{k\,\leq\, m} s_i s_i^T \tag{1685}$$

orthogonal projection of point $x$ on $\mathcal{R}(P)$ has expression like an orthogonal expansion [73, §4.10]

$$Px\ =\ QQ^Tx\ =\ \sum_{i=1}^{k} s_i^Tx\, s_i \tag{1686}$$

where

$$Q\ =\ S(:\,,1\!:\!k)\ =\ \left[\, s_1 \cdots s_k \,\right] \in \mathbb{R}^{m\times k} \tag{1687}$$

and where the $s_i$ [*sic*] are orthonormal eigenvectors of symmetric idempotent $P$. When the domain is restricted to the 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* (1673) (1260))

$$P^T = P\,,\quad P^2 = P$$
$$\Leftrightarrow$$
$$\operatorname{rank} P = \operatorname{tr} P = \|P\|_F^2 \quad \textbf{and} \quad \operatorname{rank}(I-P) = \operatorname{tr}(I-P) = \|I-P\|_F^2 \tag{1688}$$
$\diamond$

**Proof.** We take as given Theorem E.2.0.0.1 establishing idempotence. We have left only to show $\operatorname{tr} P = \|P\|_F^2 \Rightarrow P^T = P$, established in [301, §7.1]. ♦

### E.3.3 Summary, symmetric idempotent

In summary, orthogonal projector $P$ is a linear operator defined [148, §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 $Z = \mathbf{0}$ in the general nonorthogonal projector $A(A^\dagger + BZ^T)$ (1654), 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 (1680). [152, §1, prob.42]

To verify that assertion for the four fundamental subspaces (1678), we need only to express $A$ by subcompact singular value decomposition (§A.6.2): From pseudoinverse (1363) of $A = U\Sigma Q^T \in \mathbb{R}^{m\times n}$

$$\begin{array}{rlrl} AA^\dagger &= U\Sigma\Sigma^\dagger U^T = UU^T\,, & A^\dagger A &= Q\Sigma^\dagger\Sigma Q^T = QQ^T \\ I - AA^\dagger &= I - UU^T = U^\perp U^{\perp T}\,, & I - A^\dagger A &= I - QQ^T = Q^\perp Q^{\perp T} \end{array} \tag{1689}$$

where $U^\perp \in \mathbb{R}^{m\times m-\operatorname{rank} A}$ holds columnar an orthonormal basis for the orthogonal complement of $\mathcal{R}(U)$, and likewise for $Q^\perp \in \mathbb{R}^{n\times n-\operatorname{rank} A}$. Existence of an orthonormal decomposition is sufficient to establish idempotence and symmetry of an orthogonal projector (1680). ♦

### E.3.5 Unifying trait of all projectors: direction

Relation (1689) shows: orthogonal projectors simultaneously possess a biorthogonal decomposition (*confer* §E.1.1) (for example, $AA^\dagger$ for skinny-or-square $A$ full-rank) and an orthonormal decomposition ($UU^T$ whence $Px = UU^Tx$).

#### E.3.5.1 orthogonal projector, orthonormal decomposition

Consider orthogonal expansion of $x \in \mathcal{R}(U)$ :

$$x = UU^Tx = \sum_{i=1}^{n} u_i u_i^T x \tag{1690}$$

a sum of one-dimensional orthogonal projections (§E.6.3), where

$$U \triangleq [\, u_1 \cdots u_n \,] \qquad \text{and} \qquad U^TU = I \tag{1691}$$

and where the subspace projector has two expressions, (1689)

$$AA^\dagger \triangleq UU^T \tag{1692}$$

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 the remaining vectors constituting the columns of $U$ ;

$$\begin{aligned} u_j^T(u_j u_j^T x - x) &= 0 \\ u_j u_j^T x - x = u_j u_j^T x - UU^Tx &\in \mathcal{R}(\{u_i \mid i=1 \ldots n\,,\ i \neq j\}) \end{aligned} \tag{1693}$$

#### E.3.5.2 orthogonal projector, biorthogonal decomposition

We get a similar result for the biorthogonal expansion of $x \in \mathcal{R}(A)$. Define

$$A \triangleq [\, a_1 \;\; a_2 \;\cdots\; a_n \,] \in \mathbb{R}^{m\times n} \tag{1694}$$

and the rows of the pseudoinverse

$$A^\dagger \triangleq \begin{bmatrix} a_1^{*T} \\ a_2^{*T} \\ \vdots \\ a_n^{*T} \end{bmatrix} \in \mathbb{R}^{n\times m} \tag{1695}$$

under the biorthogonality condition $A^\dagger A = I$. In the biorthogonal expansion (§2.13.8)

$$x = AA^\dagger x = \sum_{i=1}^{n} a_i a_i^{*T} x \tag{1696}$$

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 $A$;

$$\begin{aligned} a_j^{*T}(a_j a_j^{*T} x - x) = 0 \\ a_j a_j^{*T} x - x = a_j a_j^{*T} x - AA^\dagger x \in \mathcal{R}(\{a_i \mid i = 1 \ldots n,\ i \neq j\}) \end{aligned} \tag{1697}$$

#### E.3.5.3 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 $P = UQ^T$ (1657) under biorthogonality condition $Q^T U = I$, in the biorthogonal expansion of $x \in \mathcal{R}(U)$

$$x = UQ^T x = \sum_{i=1}^{k} u_i q_i^T x \tag{1698}$$

where

$$\begin{aligned} U &\triangleq \left[\, u_1 \cdots u_k \,\right] \in \mathbb{R}^{m \times k} \\ Q^T &\triangleq \begin{bmatrix} q_1^T \\ \vdots \\ q_k^T \end{bmatrix} \in \mathbb{R}^{k \times m} \end{aligned} \tag{1699}$$

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 $u_i$ :

$$\begin{aligned} q_j^T(u_j q_j^T x - x) = 0 \\ u_j q_j^T x - x = u_j q_j^T x - UQ^T x \in \mathcal{R}(\{u_i \mid i = 1 \ldots k,\ i \neq j\}) \end{aligned} \tag{1700}$$

## E.4 Algebra of projection on affine subsets

Let $P_{\mathcal{A}}x$ denote projection of $x$ on affine subset $\mathcal{A} \triangleq \mathcal{R} + \alpha$ where $\mathcal{R}$ is a subspace and $\alpha \in \mathcal{A}$. Then, because $\mathcal{R}$ is parallel to $\mathcal{A}$, it holds:

$$\begin{aligned} P_{\mathcal{A}}x \;=\; P_{\mathcal{R}+\alpha}x \;&=\; (I - P_{\mathcal{R}})(\alpha) + P_{\mathcal{R}}x \\ &=\; P_{\mathcal{R}}(x - \alpha) + \alpha \end{aligned} \tag{1701}$$

Subspace projector $P_{\mathcal{R}}$ is a linear operator ($P_{\mathcal{A}}$ is not), and $P_{\mathcal{R}}(x+y) = P_{\mathcal{R}}x$ whenever $y \perp \mathcal{R}$ and $P_{\mathcal{R}}$ is an orthogonal projector.

**E.4.0.0.1 Theorem.** *Orthogonal projection on affine subset.* [73, §9.26]
Let $\mathcal{A} = \mathcal{R} + \alpha$ be an affine subset where $\alpha \in \mathcal{A}$, and let $\mathcal{R}^{\perp}$ be the orthogonal complement of subspace $\mathcal{R}$. Then $P_{\mathcal{A}}x$ is the orthogonal projection of $x \in \mathbb{R}^n$ on $\mathcal{A}$ if and only if

$$P_{\mathcal{A}}x \in \mathcal{A}\,, \quad \langle P_{\mathcal{A}}x - x\,,\, a - \alpha\rangle = 0 \quad \forall\, a \in \mathcal{A} \tag{1702}$$

or if and only if

$$P_{\mathcal{A}}x \in \mathcal{A}\,, \quad P_{\mathcal{A}}x - x \in \mathcal{R}^{\perp} \tag{1703}$$

⋄

## 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; $A \triangleq [\, a_1 \; a_2 \; \cdots \; a_n \,]$. Then orthogonal projection of vector $x \in \mathbb{R}^n$ on $\mathcal{R}(A)$ is a sum of one-dimensional orthogonal projections

$$Px \;=\; AA^{\dagger}x \;=\; A(A^TA)^{-1}A^Tx \;=\; AA^Tx \;=\; \sum_{i=1}^{n} a_i a_i^T x \tag{1704}$$

where each symmetric dyad $a_i a_i^T$ is an orthogonal projector projecting on $\mathcal{R}(a_i)$. (§E.6.3) Because $\|x - Px\|$ is minimized by orthogonal projection, $Px$ is considered to be the best approximation (in the Euclidean sense) to $x$ from the set $\mathcal{R}(A)$. [73, §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,

$$A = [\, a_1 \;\; a_2 \; \cdots \; a_n \,] \in \mathbb{R}^{m\times n} \qquad (1694)$$

and define the rows $a_i^{*T}$ of its pseudoinverse $A^\dagger$ as in (1695). Then orthogonal projection of vector $x\in\mathbb{R}^n$ on $\mathcal{R}(A)$ is a sum of one-dimensional nonorthogonal projections

$$Px \;=\; AA^\dagger x \;=\; \sum_{i=1}^{n} a_i a_i^{*T} x \qquad (1705)$$

where each nonsymmetric dyad $a_i a_i^{*T}$ is a nonorthogonal projector projecting on $\mathcal{R}(a_i)$, (§E.6.1) idempotent because of biorthogonality condition $A^\dagger A = I$.

The projection $Px$ 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 **119**.

Suppose matrix $A\in\mathbb{R}^{m\times n}$ holds a nonorthogonal basis for $\mathcal{R}(A)$ in its columns as in (1694), and the rows of pseudoinverse $A^\dagger$ are defined as in (1695). Assuming the most general biorthogonality condition $(A^\dagger + BZ^T)A = I$ with $BZ^T$ defined as for (1654), then biorthogonal expansion of vector $x$ is a sum of one-dimensional nonorthogonal projections; for $x\in\mathcal{R}(A)$

$$x \;=\; A(A^\dagger + BZ^T)x \;=\; AA^\dagger x \;=\; \sum_{i=1}^{n} a_i a_i^{*T} x \qquad (1706)$$

where each dyad $a_i a_i^{*T}$ is a nonorthogonal projector projecting on $\mathcal{R}(a_i)$. (§E.6.1) The extreme directions of $\mathcal{K} = \operatorname{cone}(A)$ are $\{a_1, \ldots, a_n\}$ the linearly independent columns of $A$ while the extreme directions $\{a_1^*, \ldots, a_n^*\}$

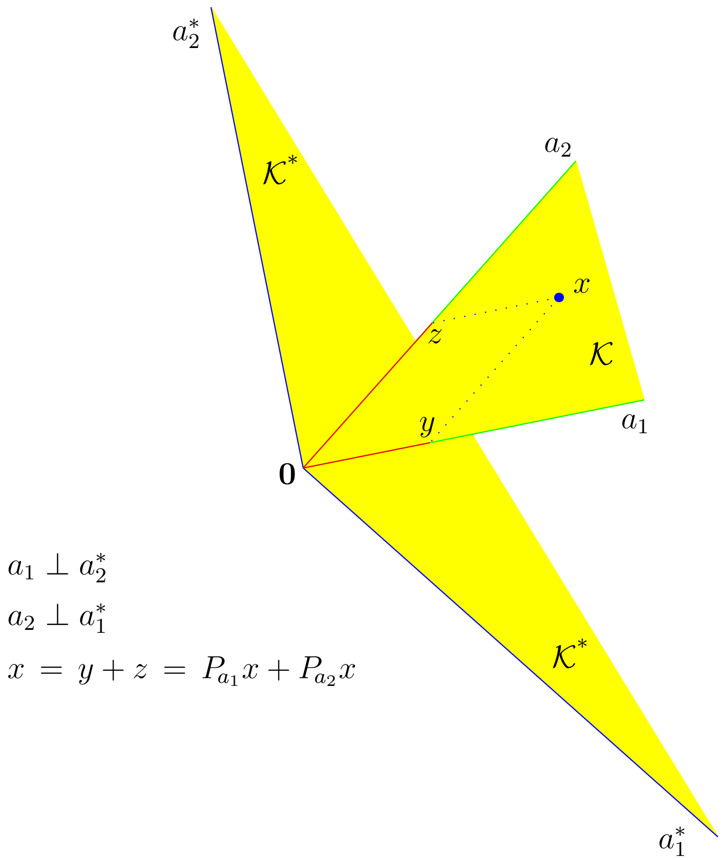



Figure 119: (*confer* Figure **49**) Biorthogonal expansion of point $x \in \operatorname{aff} \mathcal{K}$ is found by projecting $x$ nonorthogonally on range of extreme directions of polyhedral cone $\mathcal{K} \subset \mathbb{R}^{\mathbf{2}}$. Direction of projection on extreme direction $a_1$ is orthogonal to extreme direction $a_1^*$ of dual cone $\mathcal{K}^*$ and parallel to $a_2$ (§E.3.5); similarly, direction of projection 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.

of relative dual cone $\mathcal{K}^* \cap \operatorname{aff} \mathcal{K} = \operatorname{cone}(A^{\dagger T})$ (§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 $a_i a_i^{*T} x - x$ on $\mathcal{R}(a_i)$ is orthogonal to $a_i^*$.[E.8]

Because the extreme directions of this cone $\mathcal{K}$ are linearly independent, the 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

$$\mathcal{R}(A) \;=\; \operatorname{aff} \mathcal{K} \;=\; \mathcal{R}(a_1) \oplus \mathcal{R}(a_2) \oplus \ldots \oplus \mathcal{R}(a_n) \tag{1707}$$

$\square$

**E.5.0.0.4 Example.** *Nonorthogonal projection on elementary matrix.* Suppose $P_{\mathcal{Y}}$ 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

$$\mathcal{M} = \mathcal{R}(u) \oplus \mathcal{Y} \tag{1708}$$

Assuming $P_{\mathcal{M}} x = P_u x + P_{\mathcal{Y}} x$ holds, then it follows for vector $x \in \mathcal{M}$

$$P_u x = x - P_{\mathcal{Y}} x \;, \qquad P_{\mathcal{Y}} x = x - P_u x \tag{1709}$$

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 $A$

$$A \triangleq [\operatorname{basis} \mathcal{Y} \quad u\,] \in \mathbb{R}^{n+1} \tag{1710}$$

Then $P_{\mathcal{M}} = AA^{\dagger}$ fulfills the requirements, with $P_u = A(:\,, n+1)A^{\dagger}(n+1\,,:)$ and $P_{\mathcal{Y}} = A(:\,,1:n)A^{\dagger}(1:n\,,:)$. Observe, $P_{\mathcal{M}}$ is an orthogonal projector whereas $P_{\mathcal{Y}}$ and $P_u$ are nonorthogonal projectors.

Now suppose, for example, $P_{\mathcal{Y}}$ is an elementary matrix (§B.3); in particular,

$$P_{\mathcal{Y}} \;=\; I - e_1 \mathbf{1}^T \;=\; \begin{bmatrix} \mathbf{0} & \sqrt{2} V_{\mathcal{N}} \end{bmatrix} \in \mathbb{R}^{N \times N} \tag{1711}$$

[E.8] This remains true in high dimension although only a little more difficult to visualize in $\mathbb{R}^{\mathbf{3}}$; *confer*, Figure **50**.

where $\mathcal{Y}=\mathcal{N}(\mathbf{1}^T)$. We have $\mathcal{M}=\mathbb{R}^N$, $A=[\sqrt{2}V_{\mathcal{N}} \quad e_1]$, and $u=e_1$. Thus $P_u=e_1\mathbf{1}^T$ is a nonorthogonal projector projecting on $\mathcal{R}(u)$ in a direction parallel to a vector in $\mathcal{Y}$ (§E.3.5), and $P_{\mathcal{Y}}x=x-e_1\mathbf{1}^Tx$ is a nonorthogonal projection of $x$ on $\mathcal{Y}$ in a direction parallel to $u$. □

**E.5.0.0.5 Example.** *Projecting the origin on a hyperplane.*
(*confer* §2.4.2.0.2) Given the hyperplane representation having $b\in\mathbb{R}$ and nonzero normal $a\in\mathbb{R}^m$

$$\partial\mathcal{H}=\{y \mid a^Ty=b\}\subset\mathbb{R}^m \qquad (94)$$

orthogonal projection of the origin $P\mathbf{0}$ on that hyperplane is the unique optimal solution to a minimization problem: (1676)

$$\begin{aligned}\|P\mathbf{0}-\mathbf{0}\|_2 &= \inf_{y\in\partial\mathcal{H}} \|y-\mathbf{0}\|_2\\ &= \inf_{\xi\in\mathbb{R}^{m-1}} \|Z\xi+x\|_2\end{aligned} \qquad (1712)$$

where $x$ is any solution to $a^Ty=b$, 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

$$\xi^\star=-(Z^TZ)^{-1}Z^Tx \qquad (1713)$$

so

$$y^\star=\big(I-Z(Z^TZ)^{-1}Z^T\big)x \qquad (1714)$$

and from (1678)

$$P\mathbf{0} = y^\star = a(a^Ta)^{-1}a^Tx = \frac{a}{\|a\|}\frac{a^T}{\|a\|}x \triangleq AA^\dagger x = a\frac{b}{\|a\|^2} \qquad (1715)$$

In words, any point $x$ in the hyperplane $\partial\mathcal{H}$ projected on its normal $a$ (*confer* (1740)) yields that point $y^\star$ in the hyperplane closest to the origin.

□

**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

$$\mathcal{A} = \{y \mid Ay = b\} \subset \mathbb{R}^m \tag{1716}$$

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:

$$\begin{aligned} \|Px - x\|_2 &= \inf_{y\in\mathcal{A}} \|y - x\|_2 \\ &= \inf_{\xi\in\mathbb{R}^{n-\operatorname{rank} A}} \|Z\xi + y_{\mathrm{p}} - x\|_2 \end{aligned} \tag{1717}$$

where $y_{\mathrm{p}}$ is any solution to $Ay = b$, and where the columns of $Z \in \mathbb{R}^{n\times n-\operatorname{rank} A}$ constitute a basis for $\mathcal{N}(A)$ so that $y = Z\xi + y_{\mathrm{p}} \in \mathcal{A}$ for all $\xi \in \mathbb{R}^{n-\operatorname{rank} A}$.

The infimum is found by setting the gradient of the strictly convex norm-square to $\mathbf{0}$. The minimizing argument is

$$\xi^\star = -(Z^TZ)^{-1}Z^T(y_{\mathrm{p}} - x) \tag{1718}$$

so

$$y^\star = \left(I - Z(Z^TZ)^{-1}Z^T\right)(y_{\mathrm{p}} - x) + x \tag{1719}$$

and from (1678),

$$\begin{aligned} Px = y^\star &= x - A^\dagger(Ax - b) \\ &= (I - A^\dagger A)x + A^\dagger A y_{\mathrm{p}} \end{aligned} \tag{1720}$$

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 $X \in \mathbb{R}^{n\times N}$ (65); *id est*,

$$\mathcal{A} \overset{\Delta}{=} \operatorname{aff} X = \{Xa \mid a^T\mathbf{1} = 1\} \subseteq \mathbb{R}^n \tag{1721}$$

Here *minimal set* means $XV_{\mathcal{N}} = [\,x_2 - x_1 \;\; x_3 - x_1 \;\; \cdots \;\; x_N - x_1\,]/\sqrt{2}$ (776) is full-rank (§2.4.2.2) where $V_{\mathcal{N}} \in \mathbb{R}^{N\times N-1}$ is the Schoenberg auxiliary matrix

(§B.4.2). Then the orthogonal projection $Px$ of any point $x\in\mathbb{R}^n$ on $\mathcal{A}$ is the solution to a minimization problem:

$$\begin{aligned}\|Px-x\|_2 &= \inf_{a^T\mathbf{1}=1} \|Xa-x\|_2 \\ &= \inf_{\xi\in\mathbb{R}^{N-1}} \|X(V_{\mathcal{N}}\xi+a_{\mathrm{p}})-x\|_2\end{aligned} \tag{1722}$$

where $a_{\mathrm{p}}$ is any solution to $a^T\mathbf{1}=1$. We find the minimizing argument

$$\xi^\star = -(V_{\mathcal{N}}^T X^T X V_{\mathcal{N}})^{-1} V_{\mathcal{N}}^T X^T (Xa_{\mathrm{p}}-x) \tag{1723}$$

and so the orthogonal projection is [153, §3]

$$Px = Xa^\star = \big(I - XV_{\mathcal{N}}(XV_{\mathcal{N}})^\dagger\big)Xa_{\mathrm{p}} + XV_{\mathcal{N}}(XV_{\mathcal{N}})^\dagger x \tag{1724}$$

a projection of point $x$ on $\mathcal{R}(XV_{\mathcal{N}})$ then translated perpendicularly with respect to that range until it meets the affine subset $\mathcal{A}$. □

**E.5.0.0.8 Example.** *Projecting on hyperplane, halfspace, slab.*
Given the hyperplane representation having $b\in\mathbb{R}$ and nonzero normal $a\in\mathbb{R}^m$

$$\partial\mathcal{H} = \{y \mid a^Ty = b\} \subset \mathbb{R}^m \qquad (94)$$

the orthogonal projection of any point $x\in\mathbb{R}^m$ on that hyperplane is

$$Px = x - a(a^Ta)^{-1}(a^Tx-b) \tag{1725}$$

Orthogonal projection of $x$ on the halfspace parametrized by $b\in\mathbb{R}$ and nonzero normal $a\in\mathbb{R}^m$

$$\mathcal{H}_- = \{y \mid a^Ty \le b\} \subset \mathbb{R}^m \qquad (86)$$

is the point

$$Px = x - a(a^Ta)^{-1}\max\{0\,,\ a^Tx-b\} \tag{1726}$$

Orthogonal projection of $x$ on the convex slab (Figure 9), for $c<b$

$$\mathcal{B} \triangleq \{y \mid c \le a^Ty \le b\} \subset \mathbb{R}^m \tag{1727}$$

is the point [99, §5.1]

$$Px = x - a(a^Ta)^{-1}\big(\max\{0\,,\ a^Tx-b\} - \max\{0\,,\ c-a^Tx\}\big) \tag{1728}$$

□

# 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*,

$$\frac{\langle z,x\rangle}{\langle z,y\rangle}\,y \;=\; \frac{z^Tx}{z^Ty}\,y \;=\; \frac{yz^T}{z^Ty}\,x \;\triangleq\; P_0\,x \tag{1729}$$

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 projecting on $\mathcal{R}(y)$. The direction of nonorthogonal projection is orthogonal to $z$; *id est*,

$$P_0\,x - x \;\perp\; \mathcal{R}(P_0^T) \tag{1730}$$

## E.6.2 Nonorthogonal projection on vectorized matrix

Formula (1729) 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)

$$\frac{\langle Z\,,\,X\rangle}{\langle Z\,,\,Y\rangle}\,Y\;, \qquad \langle Z\,,\,Y\rangle \neq 0 \tag{1731}$$

where $\langle Z\,,\,X\rangle/\langle Z\,,\,Y\rangle$ is the coefficient of projection. The inequality accounts for the fact: projection on $\mathcal{R}(\operatorname{vec} Y)$ is in a direction orthogonal to $\operatorname{vec} Z$.

### E.6.2.1 Nonorthogonal projection on dyad

Now suppose we have nonorthogonal *projector dyad*

$$P_0 = \frac{yz^T}{z^Ty} \in \mathbb{R}^{m\times m} \tag{1732}$$

Analogous to (1729), for $X \in \mathbb{R}^{m\times m}$

$$P_0XP_0 = \frac{yz^T}{z^Ty}\,X\,\frac{yz^T}{z^Ty} \;=\; \frac{z^TXy}{(z^Ty)^2}\,yz^T \;=\; \frac{\langle zy^T,\,X\rangle}{\langle zy^T,\,yz^T\rangle}\,yz^T \tag{1733}$$

is a nonorthogonal projection of matrix $X$ on the range of vectorized dyad $P_0$ ; from which it follows:

$$P_0 X P_0 = \frac{z^T X y}{z^T y} \frac{y z^T}{z^T y} = \left\langle \frac{z y^T}{z^T y}, X \right\rangle \frac{y z^T}{z^T y} = \langle P_0^T, X \rangle P_0 = \frac{\langle P_0^T, X \rangle}{\langle P_0^T, P_0 \rangle} P_0 \tag{1734}$$

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 (1732), therefore,

$$\mathcal{R}(\operatorname{vec} P_0 X P_0) = \mathcal{R}(\operatorname{vec} P_0) \text{ in } \mathbb{R}^{m^2} \tag{1735}$$

and

$$P_0 X P_0 - X \perp P_0^T \text{ in } \mathbb{R}^{m^2} \tag{1736}$$

**E.6.2.1.1 Example.** *$\lambda$ as coefficients of nonorthogonal projection.*
Any diagonalization (§A.5)

$$X = S \Lambda S^{-1} = \sum_{i=1}^{m} \lambda_i \, s_i w_i^T \in \mathbb{R}^{m \times m} \qquad (1339)$$

may be expressed as a sum of one-dimensional nonorthogonal projections of $X$ , each on the range of a vectorized eigenmatrix $P_j \triangleq s_j w_j^T$ ;

$$\begin{aligned}
X &= \sum_{i,j=1}^{m} \langle (S e_i e_j^T S^{-1})^T, X \rangle \, S e_i e_j^T S^{-1} \\
&= \sum_{j=1}^{m} \langle (s_j w_j^T)^T, X \rangle \, s_j w_j^T \;+ \sum_{\substack{i,j=1 \\ j \neq i}}^{m} \langle (S e_i e_j^T S^{-1})^T, S \Lambda S^{-1} \rangle \, S e_i e_j^T S^{-1} \\
&= \sum_{j=1}^{m} \langle (s_j w_j^T)^T, X \rangle \, s_j w_j^T \\
&\triangleq \sum_{j=1}^{m} \langle P_j^T, X \rangle \, P_j = \sum_{j=1}^{m} s_j w_j^T X s_j w_j^T = \sum_{j=1}^{m} P_j X P_j \\
&= \sum_{j=1}^{m} \lambda_j \, s_j w_j^T
\end{aligned} \tag{1737}$$

This biorthogonal expansion of matrix $X$ is a sum of nonorthogonal projections because the term outside the projection coefficient $\langle \; \rangle$ 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 (1737),

$$\mathcal{R}(\operatorname{vec} P_j X P_j) \;=\; \mathcal{R}(\operatorname{vec} s_j w_j^T) \;=\; \mathcal{R}(\operatorname{vec} P_j) \text{ in } \mathbb{R}^{m^2} \tag{1738}$$

and

$$P_j X P_j - X \;\perp\; P_j^T \text{ in } \mathbb{R}^{m^2} \tag{1739}$$

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; [11, §3.3] [249, §3.2] [275, §2.2] [288, §1-8]

$$\frac{\langle y, x\rangle}{\langle y, y\rangle}\, y \;=\; \frac{y^T x}{y^T y}\, y \;=\; \frac{y y^T}{y^T y}\, x \;\triangleq\; P_1 x \tag{1740}$$

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

$$P_1 x - x \;\perp\; \mathcal{R}(P_1) \tag{1741}$$

because $P_1^T = P_1$.

### E.6.4 Orthogonal projection on a vectorized matrix

From (1740), 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)

$$\frac{\langle Y, X\rangle}{\langle Y, Y\rangle}\, Y \tag{1742}$$

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 $z\in\mathbb{R}^n$

$$\frac{\langle yz^T,\, X\rangle}{\langle yz^T,\, yz^T\rangle}\, yz^T \;=\; \frac{y^TXz}{y^Ty\; z^Tz}\, yz^T \tag{1743}$$

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 (1743):

$$\frac{y^TXz}{y^Ty\; z^Tz}\, yz^T \;=\; \frac{yy^T}{y^Ty}\, X\, \frac{zz^T}{z^Tz} \;\triangleq\; P_1XP_2 \tag{1744}$$

So for projector dyads, projection (1744) is the orthogonal projection in $\mathbb{R}^{mn}$ if and only if projectors $P_1$ and $P_2$ are symmetric;[E.9] 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 (1743) must be identical to the terms inside in three places.

When $P_1$ and $P_2$ are rank-one symmetric projectors as in (1744), (30)

$$\mathcal{R}(\operatorname{vec} P_1XP_2) = \mathcal{R}(\operatorname{vec} yz^T) \text{ in } \mathbb{R}^{mn} \tag{1745}$$

and

$$P_1XP_2 - X \;\perp\; yz^T \text{ in } \mathbb{R}^{mn} \tag{1746}$$

When $y=z$ then $P_1=P_2=P_2^T$ and

$$P_1XP_1 \;=\; \langle P_1\,,\, X\rangle\, P_1 \;=\; \frac{\langle P_1\,,\, X\rangle}{\langle P_1\,,\, P_1\rangle}\, P_1 \tag{1747}$$

---

[E.9] For diagonalizable $X\in\mathbb{R}^{m\times m}$ (§A.5), its orthogonal projection in isomorphic $\mathbb{R}^{m^2}$ on the range of vectorized $yz^T\in\mathbb{R}^{m\times m}$ becomes:

$$P_1XP_2 = \sum_{i=1}^{m} \lambda_i\, P_1\, s_i w_i^T P_2$$

When $\mathcal{R}(P_1)=\mathcal{R}(w_j)$ and $\mathcal{R}(P_2)=\mathcal{R}(s_j)$, the $j^{\text{th}}$ 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 normal [301, §3.2]. Yet when $\mathcal{R}(P_2)=\mathcal{R}(s_k)$, $k\neq j\in\{1\ldots m\}$, then $P_1XP_2=\mathbf{0}$.

meaning, $P_1XP_1$ 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.6.4.1.1 Example.** *Eigenvalues $\lambda$ as coefficients of orthogonal projection.* Let $\mathcal{C}$ represent any convex subset of subspace $\mathbb{S}^M$, and let $\mathcal{C}_1$ be any element of $\mathcal{C}$. Then $\mathcal{C}_1$ can be expressed as the orthogonal expansion

$$\mathcal{C}_1 = \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \geq i}}^{M} \langle E_{ij} \,, \mathcal{C}_1 \rangle \, E_{ij} \ \in \mathcal{C} \subset \mathbb{S}^M \tag{1748}$$

where $E_{ij} \in \mathbb{S}^M$ is a member of the standard orthonormal basis for $\mathbb{S}^M$ (50). This expansion is a sum of one-dimensional orthogonal projections of $\mathcal{C}_1$; each projection on the range of a vectorized standard basis matrix. Vector inner-product $\langle E_{ij} \,, \mathcal{C}_1 \rangle$ is the coefficient of projection of $\operatorname{svec} \mathcal{C}_1$ on $\mathcal{R}(\operatorname{svec} E_{ij})$.

When $\mathcal{C}_1$ is any member of a convex set $\mathcal{C}$ whose dimension is $L$, *Carathéodory's theorem* [77] [230] [148] [29] [30] guarantees that no more than $L+1$ affinely independent members from $\mathcal{C}$ are required to faithfully represent $\mathcal{C}_1$ by their linear combination.[E.10]

Dimension of $\mathbb{S}^M$ is $L = M(M+1)/2$ in isometrically isomorphic $\mathbb{R}^{M(M+1)/2}$. Yet because any symmetric matrix can be diagonalized, (§A.5.2) $\mathcal{C}_1 \in \mathbb{S}^M$ is a linear combination of its $M$ eigenmatrices $q_i q_i^T$ (§A.5.1) weighted by its eigenvalues $\lambda_i$;

$$\mathcal{C}_1 \,=\, Q\Lambda Q^T = \sum_{i=1}^{M} \lambda_i \, q_i q_i^T \tag{1749}$$

where $\Lambda \in \mathbb{S}^M$ is a diagonal matrix having $\delta(\Lambda)_i = \lambda_i$, and $Q = [\, q_1 \cdots q_M \,]$ is an orthogonal matrix in $\mathbb{R}^{M \times M}$ containing corresponding eigenvectors.

To derive eigen decomposition (1749) from expansion (1748), $M$ standard basis matrices $E_{ij}$ are rotated (§B.5) into alignment with the $M$ eigenmatrices

[E.10] Carathéodory's theorem guarantees existence of a biorthogonal expansion for any element in $\operatorname{aff} \mathcal{C}$ when $\mathcal{C}$ is any pointed closed convex cone.

$q_i q_i^T$ of $\mathcal{C}_1$ by applying a *similarity transformation*; [249, §5.6]

$$\{QE_{ij}Q^T\} = \left\{\begin{array}{ll} q_i q_i^T\,, & i=j=1\ldots M \\ \frac{1}{\sqrt{2}}\big(q_i q_j^T + q_j q_i^T\big)\,, & 1\leq i<j\leq M \end{array}\right\} \tag{1750}$$

which remains an orthonormal basis for $\mathbb{S}^M$. Then remarkably

$$\begin{aligned} \mathcal{C}_1 &= \sum_{\substack{i,j=1\\ j\geq i}}^{M} \langle QE_{ij}Q^T,\, \mathcal{C}_1\rangle\, QE_{ij}Q^T \\ &= \sum_{i=1}^{M} \langle q_i q_i^T,\, \mathcal{C}_1\rangle\, q_i q_i^T \;+\; \sum_{\substack{i,j=1\\ j>i}}^{M} \langle QE_{ij}Q^T,\, Q\Lambda Q^T\rangle\, QE_{ij}Q^T \\ &= \sum_{i=1}^{M} \langle q_i q_i^T,\, \mathcal{C}_1\rangle\, q_i q_i^T \\ &\triangleq \sum_{i=1}^{M} \langle P_i\,,\, \mathcal{C}_1\rangle\, P_i \;=\; \sum_{i=1}^{M} q_i q_i^T \mathcal{C}_1\, q_i q_i^T \;=\; \sum_{i=1}^{M} P_i\,\mathcal{C}_1 P_i \\ &= \sum_{i=1}^{M} \lambda_i\, q_i q_i^T \end{aligned} \tag{1751}$$

this orthogonal expansion becomes the diagonalization; still a sum of one-dimensional orthogonal projections. The eigenvalues

$$\lambda_i = \langle q_i q_i^T,\, \mathcal{C}_1\rangle \tag{1752}$$

are clearly coefficients of projection of $\mathcal{C}_1$ 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 $P_i$ is rank-one symmetric as in (1751),

$$\mathcal{R}(\operatorname{svec} P_i\,\mathcal{C}_1 P_i) \;=\; \mathcal{R}(\operatorname{svec} q_i q_i^T) \;=\; \mathcal{R}(\operatorname{svec} P_i) \text{ in } \mathbb{R}^{M(M+1)/2} \tag{1753}$$

and

$$P_i\,\mathcal{C}_1 P_i - \mathcal{C}_1 \;\perp\; P_i \text{ in } \mathbb{R}^{M(M+1)/2} \tag{1754}$$

□

#### E.6.4.2 Positive semidefiniteness test as orthogonal projection

For any given $X\in\mathbb{R}^{m\times m}$ the familiar quadratic construct $y^TXy\geq 0$, over broad domain, is a fundamental test for positive semidefiniteness.

(§A.2) It is a fact that $y^TXy$ is always proportional to a coefficient of orthogonal projection; letting $z$ in formula (1743) become $y\in\mathbb{R}^m$, then $P_2 = P_1 = yy^T/y^Ty = yy^T/\|yy^T\|_2$ (*confer* (1395)) and formula (1744) becomes

$$\frac{\langle yy^T,\, X\rangle}{\langle yy^T,\, yy^T\rangle}\, yy^T \;=\; \frac{y^TXy}{y^Ty}\frac{yy^T}{y^Ty} \;=\; \frac{yy^T}{y^Ty}\, X\, \frac{yy^T}{y^Ty} \;\triangleq\; P_1XP_1 \tag{1755}$$

By (1742), product $P_1XP_1$ is the one-dimensional orthogonal projection of $X$ in isomorphic $\mathbb{R}^{m^2}$ on the range of vectorized $P_1$ because, for $\operatorname{rank} P_1 = 1$ and $P_1^2 = P_1 \in \mathbb{S}^m$ (*confer* (1734))

$$P_1XP_1 \;=\; \frac{y^TXy}{y^Ty}\frac{yy^T}{y^Ty} \;=\; \left\langle \frac{yy^T}{y^Ty}\,,\, X\right\rangle \frac{yy^T}{y^Ty} \;=\; \langle P_1\,,\, X\rangle\, P_1 \;=\; \frac{\langle P_1\,,\, X\rangle}{\langle P_1\,,\, P_1\rangle}\, P_1 \tag{1756}$$

The coefficient of orthogonal projection $\langle P_1\,,\, X\rangle = y^TXy/(y^Ty)$ is also known as *Rayleigh's quotient*.[E.11] When $P_1$ is rank-one symmetric as in (1755),

$$\mathcal{R}(\operatorname{vec} P_1XP_1) \;=\; \mathcal{R}(\operatorname{vec} P_1) \text{ in } \mathbb{R}^{m^2} \tag{1757}$$

and

$$P_1XP_1 - X \;\perp\; P_1 \text{ in } \mathbb{R}^{m^2} \tag{1758}$$

---

[E.11] When $y$ becomes the $j^{\text{th}}$ eigenvector $s_j$ of diagonalizable $X$, for example, $\langle P_1\,,\, X\rangle$ becomes the $j^{\text{th}}$ eigenvalue: [145, §III]

$$\langle P_1\,,\, X\rangle\big|_{y=s_j} \;=\; \frac{s_j^T\left(\sum\limits_{i=1}^{m} \lambda_i\, s_i w_i^T\right)s_j}{s_j^Ts_j} \;=\; \lambda_j$$

Similarly for $y = w_j$, the $j^{\text{th}}$ left-eigenvector,

$$\langle P_1\,,\, X\rangle\big|_{y=w_j} \;=\; \frac{w_j^T\left(\sum\limits_{i=1}^{m} \lambda_i\, s_i w_i^T\right)w_j}{w_j^Tw_j} \;=\; \lambda_j$$

A quandary may arise regarding the potential annihilation of the antisymmetric part of $X$ when $s_j^TXs_j$ 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 $\operatorname{Re}(s_j^HXs_j) = s_j^H(X+X^T)s_j/2$ [150, §7.1] where $(X+X^T)/2$ is the symmetric part of $X$, and $s_j^H$ denotes the conjugate transpose.

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 $yy^T$ (§2.8.1) from the positive semidefinite cone in the ambient space of symmetric matrices.

#### E.6.4.3 $PXP \succeq 0$

In some circumstances, it may be desirable to limit the domain of test $y^TXy \geq 0$ for positive semidefiniteness; *e.g.*, $\|y\| = 1$. Another example of limiting domain-of-test is central to Euclidean distance geometry: For $\mathcal{R}(V) = \mathcal{N}(\mathbf{1}^T)$, the test $-VDV \succeq 0$ determines whether $D \in \mathbb{S}_h^N$ is a Euclidean distance matrix. The same test may be stated: For $D \in \mathbb{S}_h^N$ (and optionally $\|y\| = 1$)

$$D \in \mathbb{EDM}^N \Leftrightarrow -y^TDy = \langle yy^T, -D\rangle \geq 0 \quad \forall\, y \in \mathcal{R}(V) \tag{1759}$$

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 $\mathcal{R}(yy^T) = \mathcal{R}(y) \subseteq \mathcal{R}(V)$. (The validity of this result is independent of whether $V$ is itself a projection matrix.)

## E.7 on vectorized matrices of higher rank

### E.7.1 $PXP$ misinterpretation for higher-rank $P$

For a projection matrix $P$ of rank greater than 1, $PXP$ is generally not commensurate with $\frac{\langle P,X\rangle}{\langle P,P\rangle}P$ as is the case for projector dyads (1756). Yet for a symmetric idempotent matrix $P$ of any rank we are tempted to say " $PXP$ is the orthogonal projection of $X \in \mathbb{S}^m$ on $\mathcal{R}(\operatorname{vec} P)$ ". The fallacy is: $\operatorname{vec} PXP$ does not necessarily belong to the range of vectorized $P$ ; the most basic requirement for projection on $\mathcal{R}(\operatorname{vec} P)$.

### 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 (1654), and defining

$$P_1 \triangleq A_1A_1^\dagger \in \mathbb{S}^m\,, \qquad P_2 \triangleq A_2A_2^\dagger \in \mathbb{S}^p \tag{1760}$$

then, given compatible $X$

$$\|X - P_1 X P_2\|_F = \inf_{B_1,\,B_2 \in \mathbb{R}^{n\times k}} \|X - A_1(A_1^\dagger + B_1 Z_1^T)X(A_2^{\dagger T} + Z_2 B_2^T)A_2^T\|_F \tag{1761}$$

As for all subspace projectors, range of the projector is the subspace on which projection is made: $\{P_1 Y P_2 \mid Y \in \mathbb{R}^{m\times p}\}$. Altogether, for projectors $P_1$ and $P_2$ of any rank, this means projection $P_1 X P_2$ is unique minimum-distance, orthogonal

$$P_1 X P_2 - X \perp \{P_1 Y P_2 \mid Y \in \mathbb{R}^{m\times p}\} \text{ in } \mathbb{R}^{mp} \tag{1762}$$

and $P_1$ and $P_2$ must each be symmetric (*confer* (1744)) to attain the infimum.

**E.7.2.0.1 Proof.** *Minimum Frobenius norm* (1761).
Defining $P \triangleq A_1(A_1^\dagger + B_1 Z_1^T)$,

$$\begin{aligned}
&\inf_{B_1,\,B_2} \|X - A_1(A_1^\dagger + B_1 Z_1^T)X(A_2^{\dagger T} + Z_2 B_2^T)A_2^T\|_F^2 \\
&= \inf_{B_1,\,B_2} \|X - PX(A_2^{\dagger T} + Z_2 B_2^T)A_2^T\|_F^2 \\
&= \inf_{B_1,\,B_2} \operatorname{tr}\Big((X^T - A_2(A_2^\dagger + B_2 Z_2^T)X^T P^T)(X - PX(A_2^{\dagger T} + Z_2 B_2^T)A_2^T)\Big) \\
&= \inf_{B_1,\,B_2} \operatorname{tr}\Big(X^T X - X^T P X(A_2^{\dagger T} + Z_2 B_2^T)A_2^T - A_2(A_2^\dagger + B_2 Z_2^T)X^T P^T X \\
&\qquad + A_2(A_2^\dagger + B_2 Z_2^T)X^T P^T P X(A_2^{\dagger T} + Z_2 B_2^T)A_2^T\Big)
\end{aligned} \tag{1763}$$

Necessary conditions for a global minimum are $\nabla_{B_1} = \mathbf{0}$ and $\nabla_{B_2} = \mathbf{0}$. Terms not containing $B_2$ in (1763) will vanish from gradient $\nabla_{B_2}$; (§D.2.3)

$$\begin{aligned}
&\nabla_{B_2} \operatorname{tr}\Big(-X^T P X Z_2 B_2^T A_2^T - A_2 B_2 Z_2^T X^T P^T X + A_2 A_2^\dagger X^T P^T P X Z_2 B_2^T A_2^T \\
&\qquad + A_2 B_2 Z_2^T X^T P^T P X A_2^{\dagger T} A_2^T + A_2 B_2 Z_2^T X^T P^T P X Z_2 B_2^T A_2^T\Big) \\
&= -2A_2^T X^T P X Z_2 + 2A_2^T A_2 A_2^\dagger X^T P^T P X Z_2 + 2A_2^T A_2 B_2 Z_2^T X^T P^T P X Z_2 \\
&= A_2^T\Big(-X^T + A_2 A_2^\dagger X^T P^T + A_2 B_2 Z_2^T X^T P^T\Big) P X Z_2 \\
&= \mathbf{0} \qquad \Leftrightarrow \\
&\qquad \mathcal{R}(B_1) \subseteq \mathcal{N}(A_1) \quad \textbf{and} \quad \mathcal{R}(B_2) \subseteq \mathcal{N}(A_2)
\end{aligned} \tag{1764}$$

(or $Z_2 = \mathbf{0}$) because $A^T = A^T A A^\dagger$. Symmetry requirement (1760) is implicit. Were instead $P^T \triangleq (A_2^{\dagger T} + Z_2 B_2^T)A_2^T$ and the gradient with respect to $B_1$ observed, then similar results are obtained. The projector is unique.

Perpendicularity (1762) establishes uniqueness [73, §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 & $\mathcal{N}(\mathbf{V})$.
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 $\mathbb{R}^m$:

$$\begin{aligned}\mathbb{R}_c^{m\times n} &\triangleq \{Y\in\mathbb{R}^{m\times n} \mid Y\mathbf{1}=\mathbf{0}\}\\ &= \{Y\in\mathbb{R}^{m\times n} \mid \mathcal{N}(Y)\supseteq\mathbf{1}\} \;=\; \{Y\in\mathbb{R}^{m\times n} \mid \mathcal{R}(Y^T)\subseteq\mathcal{N}(\mathbf{1}^T)\}\\ &= \{XV \mid X\in\mathbb{R}^{m\times n}\}\subset\mathbb{R}^{m\times n}\end{aligned}\tag{1765}$$

the *nonsymmetric geometric center subspace.* Further suppose $V\in\mathbb{S}^n$ is a projection matrix having $\mathcal{N}(V)=\mathcal{R}(\mathbf{1})$ and $\mathcal{R}(V)=\mathcal{N}(\mathbf{1}^T)$. 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, $\mathcal{N}(XV)\supseteq\mathbf{1}$, and $\mathcal{R}(VX^T)\subseteq\mathcal{N}(\mathbf{1}^T)$.

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,

$$\begin{aligned}\mathbb{S}_c^n &\triangleq \{Y\in\mathbb{S}^n \mid Y\mathbf{1}=\mathbf{0}\}\\ &= \{Y\in\mathbb{S}^n \mid \mathcal{N}(Y)\supseteq\mathbf{1}\} \;=\; \{Y\in\mathbb{S}^n \mid \mathcal{R}(Y)\subseteq\mathcal{N}(\mathbf{1}^T)\}\end{aligned}\tag{1766}$$

the *geometric center subspace.* Further suppose $V\in\mathbb{S}^n$ is a projection matrix, the same as before. Then $VXV$ is the orthogonal projection of any $X\in\mathbb{S}^n$ on $\mathbb{S}_c^n$ in the Euclidean sense (1762) because $V$ is symmetric, $VXV\mathbf{1}=\mathbf{0}$, and $\mathcal{R}(VXV)\subseteq\mathcal{N}(\mathbf{1}^T)$. Two-sided projection is necessary only to remain in the ambient symmetric matrix subspace. Then

$$\mathbb{S}_c^n = \{VXV \mid X\in\mathbb{S}^n\}\subset\mathbb{S}^n\tag{1767}$$

has $\dim\mathbb{S}_c^n = n(n-1)/2$ in isomorphic $\mathbb{R}^{n(n+1)/2}$. We find its orthogonal complement as the aggregate of all negative directions of orthogonal projection on $\mathbb{S}_c^n$: the *translation-invariant subspace* (§5.5.1.1)

$$\begin{aligned}\mathbb{S}_c^{n\perp} &\triangleq \{X-VXV \mid X\in\mathbb{S}^n\}\subset\mathbb{S}^n\\ &= \{u\mathbf{1}^T+\mathbf{1}u^T \mid u\in\mathbb{R}^n\}\end{aligned}\tag{1768}$$

characterized by the doublet $u\mathbf{1}^T + \mathbf{1}u^T$ (§B.2).[E.12] Defining the geometric center mapping $\mathbf{V}(X) = -VXV\frac{1}{2}$ consistently with (805), then $\mathcal{N}(\mathbf{V}) = \mathcal{R}(I - \mathbf{V})$ on domain $\mathbb{S}^n$ analogously to vector projectors (§E.2); *id est*,

$$\mathcal{N}(\mathbf{V}) = \mathbb{S}_c^{n\perp} \tag{1769}$$

a subspace of $\mathbb{S}^n$ whose dimension is $\dim \mathbb{S}_c^{n\perp} = n$ in isomorphic $\mathbb{R}^{n(n+1)/2}$. Intuitively, operator $\mathbf{V}$ 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

$$\begin{aligned}
\mathbb{S}_1^n &\triangleq \{Y \in \mathbb{S}^n \mid Ye_1 = \mathbf{0}\} \\
&= \left\{\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} X \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} \mid X \in \mathbb{S}^n \right\} \\
&= \left\{\begin{bmatrix} \mathbf{0} & \sqrt{2}V_{\mathcal{N}} \end{bmatrix}^T Z \begin{bmatrix} \mathbf{0} & \sqrt{2}V_{\mathcal{N}} \end{bmatrix} \mid Z \in \mathbb{S}^N \right\}
\end{aligned} \tag{1770}$$

where $P = \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix}$ is an orthogonal projector. Then, similarly, $PXP$ is the orthogonal projection of any $X \in \mathbb{S}^n$ on $\mathbb{S}_1^n$ in the Euclidean sense (1762), and

$$\begin{aligned}
\mathbb{S}_1^{n\perp} &\triangleq \left\{\begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} X \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I \end{bmatrix} - X \mid X \in \mathbb{S}^n \right\} \subset \mathbb{S}^n \\
&= \left\{ue_1^T + e_1u^T \mid u \in \mathbb{R}^n\right\}
\end{aligned} \tag{1771}$$

Obviously, $\mathbb{S}_1^n \oplus \mathbb{S}_1^{n\perp} = \mathbb{S}^n$. □

---

[E.12]**Proof.**

$$\begin{aligned}
\{X - VXV \mid X \in \mathbb{S}^n\} &= \{X - (I - \tfrac{1}{n}\mathbf{1}\mathbf{1}^T)X(I - \mathbf{1}\mathbf{1}^T\tfrac{1}{n}) \mid X \in \mathbb{S}^n\} \\
&= \{\tfrac{1}{n}\mathbf{1}\mathbf{1}^TX + X\mathbf{1}\mathbf{1}^T\tfrac{1}{n} - \tfrac{1}{n}\mathbf{1}\mathbf{1}^TX\mathbf{1}\mathbf{1}^T\tfrac{1}{n} \mid X \in \mathbb{S}^n\}
\end{aligned}$$

Because $\{X\mathbf{1} \mid X \in \mathbb{S}^n\} = \mathbb{R}^n$,

$$\begin{aligned}
\{X - VXV \mid X \in \mathbb{S}^n\} &= \{\mathbf{1}\zeta^T + \zeta\mathbf{1}^T - \mathbf{1}\mathbf{1}^T(\mathbf{1}^T\zeta\tfrac{1}{n}) \mid \zeta \in \mathbb{R}^n\} \\
&= \{\mathbf{1}\zeta^T(I - \mathbf{1}\mathbf{1}^T\tfrac{1}{2n}) + (I - \tfrac{1}{2n}\mathbf{1}\mathbf{1}^T)\zeta\mathbf{1}^T \mid \zeta \in \mathbb{R}^n\}
\end{aligned}$$

where $I - \frac{1}{2n}\mathbf{1}\mathbf{1}^T$ is invertible. ♦

## 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 $X=U\Sigma Q^T\in\mathbb{R}^{m\times p}$, as in compact singular value decomposition (1350),

$$P_1XP_2^T = \sum_{i=1}^{\eta}\sigma_i\,P_1\,u_i q_i^T P_2^T = \sum_{i=1}^{\eta}\sigma_i\,P_1 u_i\,(P_2\,q_i)^T \tag{1772}$$

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 $P_2$ .[E.13]

## 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 $P_{\mathcal{C}}x$ of point $x$ on any closed set $\mathcal{C}\subseteq\mathbb{R}^n$ it is obvious:

$$\mathcal{C} = \{P_{\mathcal{C}}x \mid x\in\mathbb{R}^n\} \tag{1773}$$

If $\mathcal{C}\subseteq\mathbb{R}^n$ is a closed convex set, then for each and every $x\in\mathbb{R}^n$ there exists a unique point $Px$ belonging to $\mathcal{C}$ that is closest to $x$ in the Euclidean sense. Like (1676), unique projection $Px$ (or $P_{\mathcal{C}}x$) of a point $x$ on convex set $\mathcal{C}$ is that point in $\mathcal{C}$ closest to $x$ ; [182, §3.12]

$$\|x-Px\|_2 \;=\; \inf_{y\in\mathcal{C}}\|x-y\|_2 \tag{1774}$$

There exists a converse:

---

[E.13] When $P_1$ and $P_2$ are symmetric and $\mathcal{R}(P_1)=\mathcal{R}(u_j)$ and $\mathcal{R}(P_2)=\mathcal{R}(q_j)$ , then the $j^{\text{th}}$ dyad term from the singular value decomposition of $X$ is isolated by the projection. Yet if $\mathcal{R}(P_2)=\mathcal{R}(q_\ell)$ , $\ell\neq j\in\{1\ldots\eta\}$ , then $P_1XP_2=\mathbf{0}$.

**E.9.0.0.1 Theorem.** (Bunt-Motzkin) *Convex set if projections unique.* [280, §7.5] [146] If $\mathcal{C} \subseteq \mathbb{R}^n$ is a nonempty closed set and if for each and every $x$ in $\mathbb{R}^n$ there is a unique Euclidean projection $Px$ of $x$ on $\mathcal{C}$ belonging to $\mathcal{C}$, then $\mathcal{C}$ is convex. ⋄

Borwein & Lewis propose, for closed convex set $\mathcal{C}$ [41, §3.3, exer.12(d)]

$$\nabla \|x - Px\|_2^2 = 2(x - Px) \tag{1775}$$

for any point $x$ whereas, for $x \notin \mathcal{C}$

$$\nabla \|x - Px\|_2 = (x - Px)\,\|x - Px\|_2^{-1} \tag{1776}$$

**E.9.0.0.2 Exercise.** *Norm gradient.*
Prove (1775) and (1776). (Not proved in [41].) ▼

A well-known equivalent characterization of projection on a convex set is a generalization of the perpendicularity condition (1675) for projection on a subspace:

### E.9.1 Dual interpretation of projection on convex set

**E.9.1.0.1 Definition.** *Normal vector.* [230, p.15]
Vector $z$ is *normal* to convex set $\mathcal{C}$ at point $Px \in \mathcal{C}$ if

$$\langle z\,,\, y - Px\rangle \leq 0 \quad \forall\, y \in \mathcal{C} \tag{1777}$$

△

A convex set has a nonzero normal at each of its boundary points. [230, p.100] Hence, the *normal* or *dual* interpretation of projection:

**E.9.1.0.2 Theorem.** *Unique minimum-distance projection.* [148, §A.3.1] [182, §3.12] [73, §4.1] [56] (Figure **124**(b), p.632) Given a closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$, point $Px$ is the unique projection of a given point $x \in \mathbb{R}^n$ on $\mathcal{C}$ ($Px$ is that point in $\mathcal{C}$ nearest $x$) if and only if

$$Px \in \mathcal{C}\,, \qquad \langle x - Px\,,\, y - Px\rangle \leq 0 \quad \forall\, y \in \mathcal{C} \tag{1778}$$

⋄

As for subspace projection, operator $P$ is idempotent in the sense: for each and every $x\in\mathbb{R}^n$, $P(Px)=Px$. Yet operator $P$ is not linear;

- 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.*[E.14] [73, §4.3] Given closed convex set $\mathcal{C}\subseteq\mathbb{R}^n$, point $Px$ is the unique projection of a given point $x\in\mathbb{R}^n$ on $\mathcal{C}$ if and only if

$$Px\in\mathcal{C}\;, \qquad Px-x\;\in\;(\mathcal{C}-Px)^* \tag{1779}$$

In other words, $Px$ is that point in $\mathcal{C}$ nearest $x$ if and only if $Px-x$ belongs to that cone dual to translate $\mathcal{C}-Px$. ⋄

### E.9.1.1 Dual interpretation as optimization

Deutsch [76, thm.2.3] [75, §2] and Luenberger [182, p.134] carry forward Nirenberg's dual interpretation of projection [205] as solution to a maximization problem: Minimum distance from a point $x\in\mathbb{R}^n$ to a convex set $\mathcal{C}\subset\mathbb{R}^n$ 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 of a separating hyperplane (§2.4.2.7) presumes point $x$ lies on the boundary or exterior to set $\mathcal{C}$.

The optimal separating hyperplane is characterized by the fact it also supports $\mathcal{C}$. Any hyperplane supporting $\mathcal{C}$ (Figure **20**(a)) has form

$$\underline{\partial\mathcal{H}}_-=\left\{y\in\mathbb{R}^n \mid a^Ty=\sigma_{\mathcal{C}}(a)\right\} \tag{108}$$

where the support function is convex, defined

$$\sigma_{\mathcal{C}}(a)\triangleq\sup_{z\in\mathcal{C}}\,a^Tz \tag{458}$$

When point $x$ is finite and set $\mathcal{C}$ 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 $P_{\underline{\partial\mathcal{H}}_-}x$ of $x$ on any given supporting hyperplane $\underline{\partial\mathcal{H}}_-$ is

$$P_{\underline{\partial\mathcal{H}}_-}x\;=\;x-a(a^Ta)^{-1}\big(a^Tx-\sigma_{\mathcal{C}}(a)\big) \tag{1780}$$

[E.14] $-(\mathcal{C}-Px)^*$ is the normal cone to set $\mathcal{C}$ at point $Px$. (§E.10.3.2)

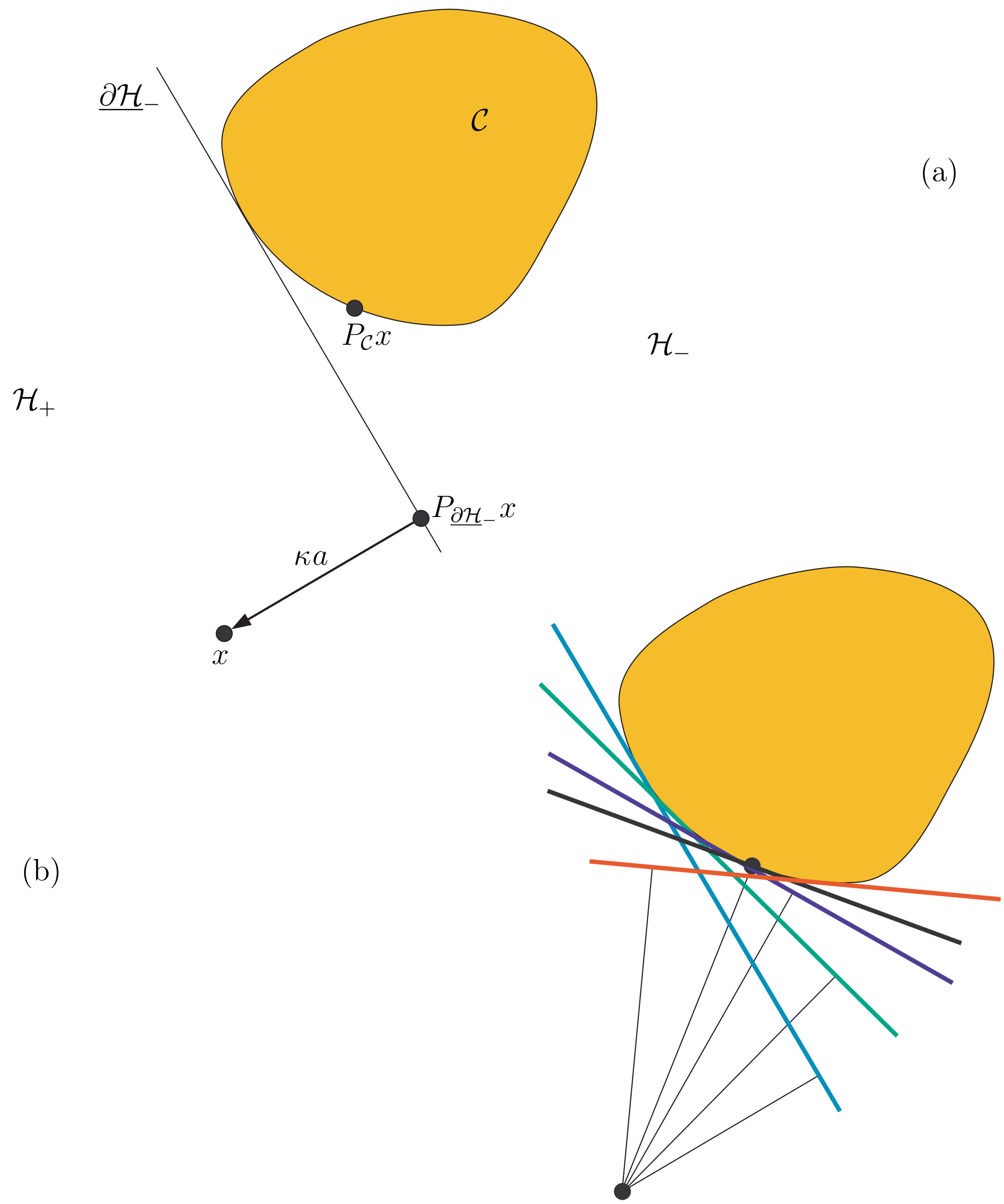



Figure 120: Dual interpretation of projection of point $x$ on convex set $\mathcal{C}$ in $\mathbb{R}^2$. **(a)** $\kappa = (a^Ta)^{-1}\big(a^Tx - \sigma_{\mathcal{C}}(a)\big)$ **(b)** Minimum distance from $x$ to $\mathcal{C}$ is found by maximizing distance to all hyperplanes supporting $\mathcal{C}$ and separating it from $x$. A convex problem for any convex set, distance of maximization is unique.

With reference to Figure 120, identifying

$$\mathcal{H}_+ = \{y \in \mathbb{R}^n \mid a^T y \geq \sigma_{\mathcal{C}}(a)\} \qquad (87)$$

then

$$\begin{aligned} \|x - P_{\mathcal{C}}x\| = \sup_{\underline{\partial\mathcal{H}}_- \mid x\in\mathcal{H}_+} \|x - P_{\underline{\partial\mathcal{H}}_-}x\| &= \sup_{a \mid x\in\mathcal{H}_+} \|a(a^Ta)^{-1}(a^Tx - \sigma_{\mathcal{C}}(a))\| \\ &= \sup_{a \mid x\in\mathcal{H}_+} \tfrac{1}{\|a\|}|a^Tx - \sigma_{\mathcal{C}}(a)| \end{aligned} \tag{1781}$$

which can be expressed as a convex optimization, for arbitrary positive constant $\tau$

$$\begin{array}{rll} \|x - P_{\mathcal{C}}x\| = \dfrac{1}{\tau}\underset{a}{\text{maximize}} & a^Tx - \sigma_{\mathcal{C}}(a) \\ \text{subject to} & \|a\| \leq \tau \end{array} \tag{1782}$$

The unique minimum-distance projection on convex set $\mathcal{C}$ is therefore

$$P_{\mathcal{C}}x = x - a^\star\left(a^{\star T}x - \sigma_{\mathcal{C}}(a^\star)\right)\frac{1}{\tau^2} \tag{1783}$$

where optimally $\|a^\star\| = \tau$ .

**E.9.1.1.1 Exercise.** *Dual projection technique on polyhedron.*
Test that projection paradigm from Figure 120 on any convex polyhedral set. ▼

### E.9.1.2 Dual interpretation of projection on cone

In the circumstance set $\mathcal{C}$ is a closed convex cone $\mathcal{K}$ and there exists a hyperplane separating given point $x$ from $\mathcal{K}$ , then optimal $\sigma_{\mathcal{K}}(a^\star)$ takes value 0 [148, §C.2.3.1]. So problem (1782) for projection of $x$ on $\mathcal{K}$ becomes

$$\begin{array}{rll} \|x - P_{\mathcal{K}}x\| = \dfrac{1}{\tau}\underset{a}{\text{maximize}} & a^Tx \\ \text{subject to} & \|a\| \leq \tau \\ & a \in \mathcal{K}^\circ \end{array} \tag{1784}$$

The norm inequality in (1784) can be handled by Schur complement (§3.1.7.2). Normals $a$ to all hyperplanes supporting $\mathcal{K}$ belong to the polar cone $\mathcal{K}^\circ = -\mathcal{K}^*$ by definition: (275)

$$a \in \mathcal{K}^\circ \;\Leftrightarrow\; \langle a\,, x\rangle \leq 0 \;\text{ for all }\; x \in \mathcal{K} \tag{1785}$$

Projection on cone $\mathcal{K}$ is

$$P_{\mathcal{K}}x = (I - \frac{1}{\tau^2}a^\star a^{\star T})x \tag{1786}$$

whereas projection on the polar cone $-\mathcal{K}^*$ is (§E.9.2.2.1)

$$P_{\mathcal{K}^\circ}x \;=\; x - P_{\mathcal{K}}x \;=\; \frac{1}{\tau^2}a^\star a^{\star T}x \tag{1787}$$

Negating vector $a$ , this maximization problem (1784) becomes a minimization (the same problem) and the polar cone becomes the dual cone:

$$\begin{array}{rll} \|x - P_{\mathcal{K}}x\| \;=\; -\dfrac{1}{\tau}\underset{a}{\text{minimize}} & a^T x \\ \text{subject to} & \|a\| \leq \tau \\ & a \in \mathcal{K}^* \end{array} \tag{1788}$$

### E.9.2 Projection on cone

When convex set $\mathcal{C}$ is a cone, there is a finer statement of optimality conditions:

**E.9.2.0.1 Theorem.** *Unique projection on cone.* [148, §A.3.2]
Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a closed convex cone, and $\mathcal{K}^*$ its dual (§2.13.1). Then $Px$ is the unique minimum-distance projection of $x \in \mathbb{R}^n$ on $\mathcal{K}$ if and only if

$$Px \in \mathcal{K}\,, \quad \langle Px - x\,,\, Px\rangle = 0\,, \quad Px - x \in \mathcal{K}^* \tag{1789}$$

$\diamond$

In words, $Px$ is the unique minimum-distance projection of $x$ on $\mathcal{K}$ if and only if

1) projection $Px$ lies in $\mathcal{K}$

2) direction $Px - x$ is orthogonal to the projection $Px$

3) direction $Px - x$ lies in the dual cone $\mathcal{K}^*$.

As the theorem is stated, it admits projection on $\mathcal{K}$ having empty interior; *id est*, on convex cones in a proper subspace of $\mathbb{R}^n$.

Projection on $\mathcal{K}$ of any point $x \in -\mathcal{K}^*$, belonging to the negative dual cone, is the origin. By (1789): the set of all points reaching the origin, when projecting on $\mathcal{K}$, constitutes the negative dual cone; `a.k.a`, the polar cone

$$\mathcal{K}^\circ = -\mathcal{K}^* = \{x \in \mathbb{R}^n \mid Px = \mathbf{0}\} \tag{1790}$$

#### E.9.2.1 Relation to subspace projection

Conditions 1 and 2 of the theorem are common with orthogonal projection on a subspace $\mathcal{R}(P)$: Condition 1 is the most basic requirement; namely, $Px \in \mathcal{R}(P)$, the projection belongs to the subspace. Invoking perpendicularity condition (1675), we recall the second requirement for projection on a subspace:

$$Px - x \perp \mathcal{R}(P) \quad \textbf{or} \quad Px - x \in \mathcal{R}(P)^\perp \tag{1791}$$

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 $-\mathcal{K}^*$ plays the role $\mathcal{R}(P)^\perp$ plays for subspace projection ($P_{\mathcal{R}}x = x - P_{\mathcal{R}^\perp}x$). Indeed, $-\mathcal{K}^*$ is the algebraic complement in the orthogonal vector sum (p.676) [197] [148, §A.3.2.5]

$$\mathcal{K} \boxplus -\mathcal{K}^* = \mathbb{R}^n \;\Leftrightarrow\; \text{cone } \mathcal{K} \text{ is closed and convex} \tag{1792}$$

Also, given unique minimum-distance projection $Px$ on $\mathcal{K}$ satisfying Theorem E.9.2.0.1, then by projection on the algebraic complement via $I - P$ in §E.2 we have

$$-\mathcal{K}^* = \{x - Px \mid x \in \mathbb{R}^n\} = \{x \in \mathbb{R}^n \mid Px = \mathbf{0}\} \tag{1793}$$

consequent to Moreau (1796). Recalling any subspace is a closed convex cone[E.15]

$$\mathcal{K} = \mathcal{R}(P) \;\Leftrightarrow\; -\mathcal{K}^* = \mathcal{R}(P)^\perp \tag{1794}$$

meaning, when a cone is a subspace $\mathcal{R}(P)$ then the dual cone becomes its orthogonal complement $\mathcal{R}(P)^\perp$. [46, §2.6.1] In this circumstance, condition 3 becomes coincident with condition 2.

The properties of projection on cones following in §E.9.2.2 further generalize to subspaces by: (4)

$$\mathcal{K} = \mathcal{R}(P) \;\Leftrightarrow\; -\mathcal{K} = \mathcal{R}(P) \tag{1795}$$

[E.15] but a proper subspace is not a proper cone (§2.7.2.2.1).

#### E.9.2.2 Salient properties: Projection $Px$ on closed convex cone $\mathcal{K}$

[148, §A.3.2] [73, §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. $P_{\mathcal{K}} x = \mathbf{0} \;\Leftrightarrow\; x \in -\mathcal{K}^*$
4. $P_{\mathcal{K}}(-x) = -P_{-\mathcal{K}} x$
5. (Jean-Jacques Moreau (1962)) [197]

$$\begin{array}{c} x = x_1 + x_2\;, \quad x_1 \in \mathcal{K}\;, \quad x_2 \in -\mathcal{K}^*, \quad x_1 \perp x_2 \\ \Leftrightarrow \\ x_1 = P_{\mathcal{K}} x\;, \quad x_2 = P_{-\mathcal{K}^*} x \end{array} \tag{1796}$$

6. $\mathcal{K} \;=\; \{x - P_{-\mathcal{K}^*} x \mid x \in \mathbb{R}^n\} \;=\; \{x \in \mathbb{R}^n \mid P_{-\mathcal{K}^*} x = \mathbf{0}\}$
7. $-\mathcal{K}^* \;=\; \{x - P_{\mathcal{K}} x \mid x \in \mathbb{R}^n\} \;=\; \{x \in \mathbb{R}^n \mid P_{\mathcal{K}} x = \mathbf{0}\}$ (1793)

**E.9.2.2.1 Corollary.** *$I - P$ for cones.* (*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. ⋄

**Proof.** Assume $x_1 = P_{\mathcal{K}} x$. Then by Theorem E.9.2.0.1 we have

$$x_1 \in \mathcal{K}\;, \qquad x_1 - x \perp x_1\;, \qquad x_1 - x \in \mathcal{K}^* \tag{1797}$$

Now assume $x - x_1 = P_{-\mathcal{K}^*} x$. Then we have

$$x - x_1 \in -\mathcal{K}^*, \qquad -x_1 \perp x - x_1\;, \qquad -x_1 \in -\mathcal{K} \tag{1798}$$

But these two assumptions are apparently identical. We must therefore have

$$x - P_{-\mathcal{K}^*} x \;=\; x_1 \;=\; P_{\mathcal{K}} x \tag{1799}$$

♦

**E.9.2.2.2 Corollary.** *Unique projection via dual or normal cone.* [73, §4.7] (§E.10.3.2, *confer* Theorem E.9.1.0.3) Given point $x \in \mathbb{R}^n$ and closed convex cone $\mathcal{K} \subseteq \mathbb{R}^n$, the following are equivalent statements:

1. point $Px$ is the unique minimum-distance projection of $x$ on $\mathcal{K}$
2. $Px \in \mathcal{K}\,, \qquad x - Px \in -(\mathcal{K} - Px)^* \;=\; -\mathcal{K}^* \cap (Px)^\perp$
3. $Px \in \mathcal{K}\,, \qquad \langle x - Px\,,\, Px\rangle = 0\,, \qquad \langle x - Px\,,\, y\rangle \leq 0 \;\;\forall\, y \in \mathcal{K}$

$\diamond$

**E.9.2.2.3 Example.** *Unique projection on nonnegative orthant.* (*confer* (1126)) From Theorem E.9.2.0.1, to project matrix $H \in \mathbb{R}^{m\times n}$ on the self-dual orthant (§2.13.5.1) of nonnegative matrices $\mathbb{R}^{m\times n}_+$ in isomorphic $\mathbb{R}^{mn}$, the necessary and sufficient conditions are:

$$\begin{array}{l} H^\star \geq \mathbf{0} \\ \operatorname{tr}\big((H^\star - H)^T H^\star\big) = 0 \\ H^\star - H \geq \mathbf{0} \end{array} \tag{1800}$$

where the inequalities denote entrywise comparison. The optimal solution $H^\star$ is simply $H$ having all its negative entries zeroed;

$$H^\star_{ij} = \max\{H_{ij}\,,\, 0\}\,, \quad i\,,j \in \{1 \ldots m\} \times \{1 \ldots n\} \tag{1801}$$

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 [73, §4.8]

$$H^\star_{ij} = \max\{H_{ij}\,,\, T_{ij}\} \tag{1802}$$

$\square$

**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:

$$\begin{array}{cl} \underset{y \in \mathbb{R}^N}{\text{minimize}} & \|x - Ay\|_2 \\ \text{subject to} & y \succeq 0 \\ & \|y\|_\infty \leq 1 \end{array} \tag{1803}$$

where the convex cone has vertex-description (§2.12.2.0.1), for $A \in \mathbb{R}^{n\times N}$

$$\mathcal{K} = \{Ay \mid y \succeq 0\} \tag{1804}$$

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). [66] The problem is equivalent to a Schur-form semidefinite program (§3.1.7.2)

$$\begin{array}{cl}\underset{y\in\mathbb{R}^N,\ t\in\mathbb{R}}{\text{minimize}} & t\\ \text{subject to} & \begin{bmatrix} tI & x - Ay\\ (x-Ay)^T & t\end{bmatrix} \succeq 0\\ & 0 \preceq y \preceq \mathbf{1}\end{array} \tag{1805}$$

□

### E.9.3 nonexpansivity

**E.9.3.0.1 Theorem.** *Nonexpansivity.* [125, §2] [73, §5.3]
When $\mathcal{C} \subset \mathbb{R}^n$ is an arbitrary closed convex set, projector $P$ projecting on $\mathcal{C}$ is nonexpansive in the sense: for any vectors $x, y \in \mathbb{R}^n$

$$\|Px - Py\| \leq \|x - y\| \tag{1806}$$

with equality when $x - Px = y - Py$ .[E.16] ◇

**Proof.** [40]

$$\begin{aligned}\|x-y\|^2 = \ &\|Px-Py\|^2 + \|(I-P)x - (I-P)y\|^2\\ &+ 2\langle x-Px\,,\, Px-Py\rangle + 2\langle y-Py\,,\, Py-Px\rangle\end{aligned} \tag{1807}$$

Nonnegativity of the last two terms follows directly from the *unique minimum-distance projection theorem* (§E.9.1.0.2). ♦

[E.16] This condition for equality corrects an error in [56] (where the norm is applied to each side of the condition given here) easily revealed by counter-example.

The foregoing proof reveals another flavor of nonexpansivity; for each and every $x, y \in \mathbb{R}^n$

$$\|Px - Py\|^2 + \|(I-P)x - (I-P)y\|^2 \leq \|x-y\|^2 \tag{1808}$$

Deutsch shows yet another: [73, §5.5]

$$\|Px - Py\|^2 \leq \langle x - y\ ,\ Px - Py\rangle \tag{1809}$$

### E.9.4 Easy projections

- Projecting any matrix $H \in \mathbb{R}^{n\times n}$ in the Euclidean/Frobenius sense orthogonally on the subspace of symmetric matrices $\mathbb{S}^n$ in isomorphic $\mathbb{R}^{n^2}$ amounts to taking the 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 the symmetric hollow subspace $\mathbb{S}^n_h$ in isomorphic $\mathbb{R}^{n^2}$ (§2.2.3.0.1), we may take the symmetric part and then zero all entries along the main diagonal, or *vice versa* (because this is projection on the intersection of two subspaces); *id est*, $(H+H^T)/2 - \delta^2(H)$ .

- To project a matrix on the nonnegative orthant $\mathbb{R}^{m\times n}_+$, simply clip all negative entries to 0. Likewise, projection on the nonpositive orthant $\mathbb{R}^{m\times n}_-$ sees all positive entries clipped to 0. Projection on other orthants is equally simple with appropriate clipping.

- Clipping in excess of $|1|$ each entry of a point $x \in \mathbb{R}^n$ is equivalent to unique minimum-distance projection of $x$ on the unit hypercube centered at the origin. (*confer* §E.10.3.2)

- Projecting on hyperplane, halfspace, slab: §E.5.0.0.8.

- Projection of $x \in \mathbb{R}^n$ on a (rectangular) hyperbox: [46, §8.1.1]

$$\mathcal{C} = \{y \in \mathbb{R}^n \mid l \preceq y \preceq u\,,\ l \prec u\} \tag{1810}$$

$$P(x)_k = \begin{cases} l_k\ , & x_k \leq l_k \\ x_k\ , & l_k \leq x_k \leq u_k \\ u_k\ , & x_k \geq u_k \end{cases} \tag{1811}$$

- Unique minimum-distance projection of $H \in \mathbb{S}^n$ on the positive semidefinite cone $\mathbb{S}^n_+$ in the Euclidean/Frobenius sense is accomplished by eigen decomposition (diagonalization) followed by clipping all negative eigenvalues to 0.

- Unique minimum-distance projection on the 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 the smallest nonnegative eigenvalues keeping only $\rho$ largest. (§7.1.2)

- Unique minimum-distance projection of $H \in \mathbb{R}^{m\times n}$ on the set of all $m\times n$ matrices of rank no greater than $k$ in the Euclidean/Frobenius sense is the singular value decomposition (§A.6) of $H$ having all singular values beyond the $k^{\text{th}}$ zeroed. [246, p.208] This is also a solution to the projection in the sense of spectral norm. [46, §8.1]

- Projection on $\mathcal{K}$ of any point $x \in -\mathcal{K}^*$, belonging to the polar cone, is equivalent to projection on the origin. (§E.9.2)

- Projection on Lorentz cone: [46, exer.8.3(c)]

- $P_{\mathbb{S}^N_+ \cap \mathbb{S}^N_c} = P_{\mathbb{S}^N_+} P_{\mathbb{S}^N_c}$ (1070)

- $P_{\mathbb{R}^{N\times N}_+ \cap \mathbb{S}^N_h} = P_{\mathbb{R}^{N\times N}_+} P_{\mathbb{S}^N_h}$ (§7.0.1.1)

- $P_{\mathbb{R}^{N\times N}_+ \cap \mathbb{S}^N} = P_{\mathbb{R}^{N\times N}_+} P_{\mathbb{S}^N}$ (§E.9.5)

**E.9.4.0.1 Exercise.** *Largest singular value.*
Find the unique minimum-distance projection on the set of all $m\times n$ matrices whose largest singular value does not exceed 1. ▼

Deutsch [75] provides an algorithm for projection on polyhedral cones.

Youla [300, §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 4.4.3.0.2, Figure **10**)

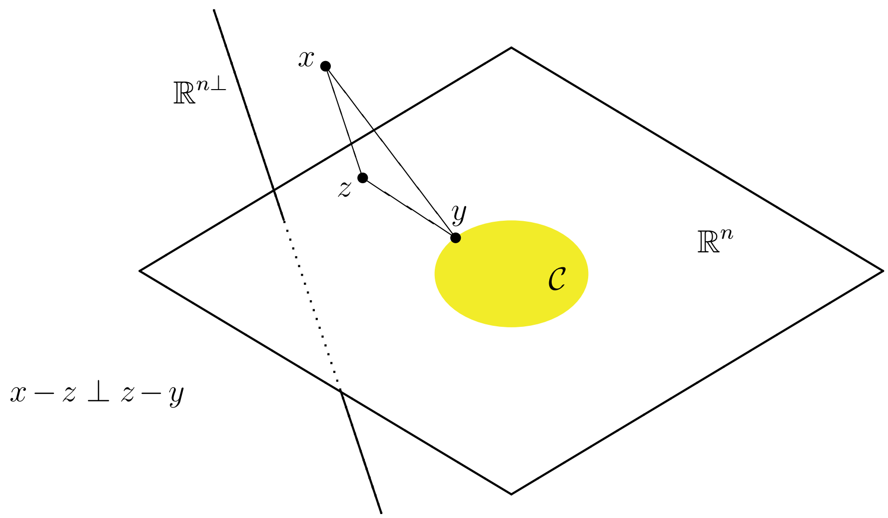



Figure 121: Closed convex set $\mathcal{C}$ belongs to subspace $\mathbb{R}^n$ (shown bounded in sketch and drawn without proper perspective). Point $y$ is unique minimum-distance projection of $x$ on $\mathcal{C}$; equivalent to product of orthogonal projection of $x$ on $\mathbb{R}^n$ and minimum-distance projection of result $z$ on $\mathcal{C}$.

### E.9.5 Projection on convex set in subspace

Suppose a convex set $\mathcal{C}$ 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 $\mathcal{C}$ can be accomplished by first projecting orthogonally on that subspace, and then projecting the result on $\mathcal{C}$; [73, §5.14] *id est*, the ordered product of two individual projections that is not commutable.

**Proof.** (⇐) To show that, suppose unique minimum-distance projection $P_{\mathcal{C}}x$ on $\mathcal{C} \subset \mathbb{R}^n$ is $y$ as illustrated in Figure **121**;

$$\|x-y\| \leq \|x-q\| \quad \forall\, q \in \mathcal{C} \tag{1812}$$

Further suppose $P_{\mathbb{R}^n}x$ equals $z$. By the *Pythagorean theorem*

$$\|x-y\|^2 = \|x-z\|^2 + \|z-y\|^2 \tag{1813}$$

because $x-z \perp z-y$. (1675) [182, §3.3] Then point $y = P_{\mathcal{C}}x$ is the same as $P_{\mathcal{C}}z$ because

$$\|z-y\|^2 \;=\; \|x-y\|^2 - \|x-z\|^2 \;\leq\; \|z-q\|^2 \;=\; \|x-q\|^2 - \|x-z\|^2 \quad \forall\, q \in \mathcal{C} \tag{1814}$$

which holds by assumption (1812).
(⇒) Now suppose $z = P_{\mathbb{R}^n}x$ and

$$\|z-y\| \leq \|z-q\| \quad \forall\, q \in \mathcal{C} \tag{1815}$$

meaning $y = P_{\mathcal{C}}z$. Then point $y$ is identical to $P_{\mathcal{C}}x$ because

$$\|x-y\|^2 \;=\; \|x-z\|^2 + \|z-y\|^2 \;\leq\; \|x-q\|^2 \;=\; \|x-z\|^2 + \|z-q\|^2 \quad \forall\, q \in \mathcal{C} \tag{1816}$$

by assumption (1815). ♦

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 $\mathcal{K} = \mathbb{S}^n \cap \mathbb{R}_+^{n\times n}$ in isomorphic $\mathbb{R}^{n^2}$ 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 $\mathcal{C}_k$ , 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 $\mathcal{C}_k$ , often easier. 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 it. Dependably finding that point is solved by an elegantly simple enhancement to the alternating projection technique: this *Dykstra algorithm* (1854) 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 $\mathcal{C}_k$. 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

Given two arbitrary convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$ and their respective minimum-distance projection operators $P_1$ and $P_2$ , if projectors commute for each and every $x\in\mathbb{R}^n$ then it is easy to show $P_1P_2x\in\mathcal{C}_1\cap\mathcal{C}_2$ and $P_2P_1x\in\mathcal{C}_1\cap\mathcal{C}_2$ . When projectors commute ($P_1P_2=P_2P_1$), a point in the intersection can be found in a finite number of steps; while commutativity is a sufficient condition, it is not necessary (§6.8.1.1.1 for example).

When $\mathcal{C}_1$ and $\mathcal{C}_2$ are subspaces, in particular, projectors $P_1$ and $P_2$ commute if and only if $P_1P_2=P_{\mathcal{C}_1\cap\mathcal{C}_2}$ or iff $P_2P_1=P_{\mathcal{C}_1\cap\mathcal{C}_2}$ or iff $P_1P_2$ is the orthogonal projection on a Euclidean subspace. [73, lem.9.2] Subspace projectors will commute, for example, when $P_1(\mathcal{C}_2)\subset\mathcal{C}_2$ or $P_2(\mathcal{C}_1)\subset\mathcal{C}_1$ or $\mathcal{C}_1\subset\mathcal{C}_2$ or $\mathcal{C}_2\subset\mathcal{C}_1$ or $\mathcal{C}_1\perp\mathcal{C}_2$ . When subspace projectors commute, this means we can find a point in the intersection of those subspaces in a finite

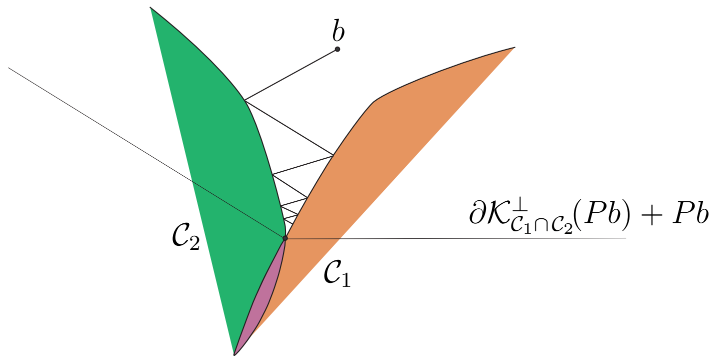



Figure 122: First several alternating projections (1827) in von Neumann-style projection of point $b$ converging on closest point $Pb$ in intersection of two closed convex sets in $\mathbb{R}^2$; $\mathcal{C}_1$ and $\mathcal{C}_2$ are partially drawn in vicinity of their intersection. The pointed normal cone $\mathcal{K}^{\perp}$ (1856) is translated to $Pb$, the unique minimum-distance projection of $b$ on intersection. For this particular example, it is possible to start anywhere in a large neighborhood of $b$ and still converge to $Pb$. The alternating projections are themselves robust with respect to some significant amount of noise because they belong to translated normal cone.

number of steps; we find, in fact, the closest point.

**E.10.0.1.1 Theorem.** *Kronecker projector.* [245, §2.7]
Given any projection matrices $P_1$ and $P_2$ (subspace projectors), then

$$P_1 \otimes P_2 \qquad \textbf{and} \qquad P_1 \otimes I \tag{1817}$$

are projection matrices. The product preserves symmetry if present. ⋄

### E.10.0.2 noncommutative projectors

Typically, one considers the method of alternating projection when projectors do not commute; *id est*, when $P_1P_2 \neq P_2P_1$.

The iconic example for noncommutative projectors illustrated in Figure **122** shows the iterates converging to the closest point in the intersection of two arbitrary convex sets. Yet simple examples like Figure **123** reveal that noncommutative alternating projection does not

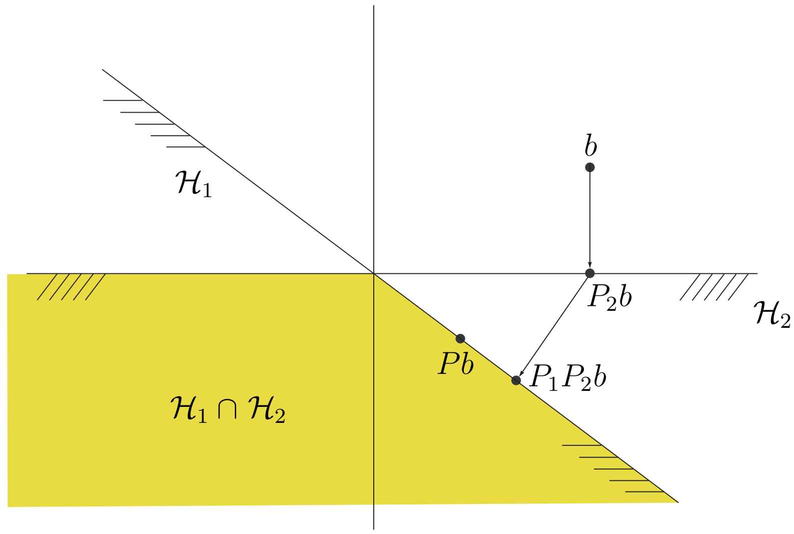



Figure 123: The sets $\{\mathcal{C}_k\}$ in this example comprise two halfspaces $\mathcal{H}_1$ and $\mathcal{H}_2$ . The von Neumann-style alternating projection in $\mathbb{R}^2$ quickly converges to $P_1P_2b$ (feasibility). The unique minimum-distance projection on the intersection is, of course, $Pb$ .

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 *successive projection* [128] [125] [48], *cyclic projection* [99] [191, §3.2], *successive approximation* [56], or *projection on convex sets* [243] [244, §6.4]. It is traced back to von Neumann (1933) [278] and later Wiener [284] 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 $\mathcal{R}_1$ and $\mathcal{R}_2$ are closed subspaces 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,

$$\lim_{i\to\infty}(P_1P_2)^i b = P_{\mathcal{R}_1\cap\mathcal{R}_2}b \tag{1818}$$

Deutsch [73, thm.9.8, thm.9.35] shows rate of convergence for subspaces to be *geometric* [299, §1.4.4]; bounded above by $\kappa^{2i+1}\|b\|$ , $i=0,1,2\ldots,$ where

$0 \leq \kappa < 1$:

$$\| (P_1 P_2)^i b - P_{\mathcal{R}_1 \cap \mathcal{R}_2} b \| \leq \kappa^{2i+1} \|b\| \tag{1819}$$

This means convergence can be slow when $\kappa$ is close to 1. The rate of convergence on intersecting halfspaces is also geometric. [74] [220]

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 **122** illustrates one application where convergence is reasonably geometric and the result is the unique minimum-distance projection. Figure **123**, in contrast, demonstrates convergence in one iteration to a *fixed point* (of the projection product)[E.17] in the intersection of two halfspaces; a.k.a, feasibility problem. It was Dykstra who in 1983 [84] (§E.10.3) first solved this projection problem.

#### 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 **124**(a)(b). Find a unique point of projection $P_1 b \in \mathcal{C}_1$ that attains the distance between any two closed convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$;

$$\|P_1 b - b\| = \operatorname{dist}(\mathcal{C}_1 , \mathcal{C}_2) \triangleq \inf_{z \in \mathcal{C}_2} \|P_1 z - z\| \tag{1820}$$

- **Feasibility.** Figure **124**(c), $\bigcap \mathcal{C}_k \neq \emptyset$. Given a number of indexed closed convex sets $\mathcal{C}_k \subset \mathbb{R}^n$, find any fixed point in their intersection by iterating $(i)$ a projection product starting from $b$;

$$\left( \prod_{i=1}^{\infty} \prod_k P_k \right) b \in \bigcap_k \mathcal{C}_k \tag{1821}$$

- **Optimization.** Figure **124**(c), $\bigcap \mathcal{C}_k \neq \emptyset$. Given a number of indexed closed convex sets $\mathcal{C}_k \subset \mathbb{R}^n$, uniquely project a given point $b$ on $\bigcap \mathcal{C}_k$;

$$\|Pb - b\| = \inf_{x \in \bigcap \mathcal{C}_k} \|x - b\| \tag{1822}$$

---

[E.17] A fixed point of a mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ is a point $x$ whose image is identical under the map; *id est*, $Tx = x$.

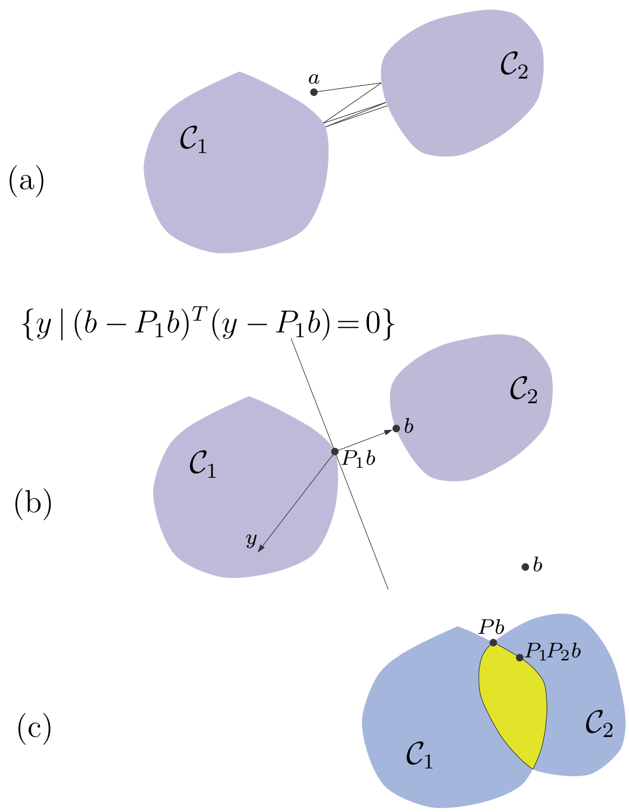



Figure 124:
**(a)** (**distance**) Intersection of two convex sets in $\mathbb{R}^{\mathbf{2}}$ is empty. Method of alternating projection would be applied to find that point in $\mathcal{C}_1$ nearest $\mathcal{C}_2$ .
**(b)** (**distance**) Given $b\in\mathcal{C}_2$ , then $P_1b\in\mathcal{C}_1$ is nearest $b$ iff $(y-P_1b)^T(b-P_1b)\leq 0 \;\forall\, y\in\mathcal{C}_1$ by the *unique minimum-distance projection theorem* (§E.9.1.0.2). When $P_1b$ attains the distance between the two sets, hyperplane $\{y \mid (b-P_1b)^T(y-P_1b)=0\}$ separates $\mathcal{C}_1$ from $\mathcal{C}_2$ . [46, §2.5.1]
**(c)** (**0 distance**) Intersection is nonempty.
(**optimization**) We may want the point $Pb$ in $\bigcap\mathcal{C}_k$ nearest point $b$ .
(**feasibility**) We may instead be satisfied with a fixed point of the projection product $P_1P_2b$ in $\bigcap\mathcal{C}_k$ .

## E.10.1 Distance and existence

Existence of a fixed point is established:

**E.10.1.0.1 Theorem.** *Distance.* [56]
Given any two closed convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$ in $\mathbb{R}^n$, then $P_1 b \in \mathcal{C}_1$ is a fixed point of the projection product $P_1 P_2$ if and only if $P_1 b$ is a point of $\mathcal{C}_1$ nearest $\mathcal{C}_2$. ◇

**Proof.** (⇒) Given fixed point $a = P_1 P_2 a \in \mathcal{C}_1$ with $b \triangleq P_2 a \in \mathcal{C}_2$ in tandem so that $a = P_1 b$, then by the *unique minimum-distance projection theorem* (§E.9.1.0.2)

$$\begin{array}{c}(b-a)^T(u-a) \leq 0 \quad \forall\, u \in \mathcal{C}_1 \\ (a-b)^T(v-b) \leq 0 \quad \forall\, v \in \mathcal{C}_2 \\ \Leftrightarrow \\ \|a-b\| \leq \|u-v\| \quad \forall\, u \in \mathcal{C}_1 \text{ and } \forall\, v \in \mathcal{C}_2 \end{array} \tag{1823}$$

by Schwarz inequality $\|\langle x\,,y\rangle\| \leq \|x\|\,\|y\|$ [166] [230].
(⇐) Suppose $a \in \mathcal{C}_1$ and $\|a - P_2 a\| \leq \|u - P_2 u\|\ \forall\, u \in \mathcal{C}_1$. Now suppose we choose $u = P_1 P_2 a$. Then

$$\|u - P_2 u\| = \|P_1 P_2 a - P_2 P_1 P_2 a\| \leq \|a - P_2 a\| \;\Leftrightarrow\; a = P_1 P_2 a \tag{1824}$$

Thus $a = P_1 b$ (with $b = P_2 a \in \mathcal{C}_2$) is a fixed point in $\mathcal{C}_1$ of the projection product $P_1 P_2$.[E.18] ♦

## E.10.2 Feasibility and convergence

The set of all fixed points of any nonexpansive mapping is a closed convex set. [107, lem.3.4] [23, §1] The projection product $P_1 P_2$ is nonexpansive by Theorem E.9.3.0.1 because, for any vectors $x\,, a \in \mathbb{R}^n$

$$\|P_1 P_2 x - P_1 P_2 a\| \;\leq\; \|P_2 x - P_2 a\| \;\leq\; \|x - a\| \tag{1825}$$

If the intersection of two closed convex sets $\mathcal{C}_1 \cap \mathcal{C}_2$ 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

[E.18] Point $b = P_2 a$ can be shown, similarly, to be a fixed point of the product $P_2 P_1$.

(a feasible point) 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 [23, §2] argue that any sequence monotonic in the sense of Fejér is convergent.[E.19]

**E.10.2.0.1 Definition.** *Fejér monotonicity.* [198]
Given closed convex set $\mathcal{C} \neq \emptyset$, then a *sequence* $x_i \in \mathbb{R}^n$, $i=0,1,2\ldots$, is monotonic in the sense of Fejér with respect to $\mathcal{C}$ iff

$$\|x_{i+1} - c\| \leq \|x_i - c\| \quad \text{for all } i \geq 0 \;\; \textbf{and} \;\; \text{each and every } c \in \mathcal{C} \tag{1826}$$

$\triangle$

Given $x_0 \triangleq b$, if we express each iteration of alternating projection by

$$x_{i+1} = P_1 P_2\, x_i\ , \quad i=0,1,2\ldots \tag{1827}$$

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

$$\|P_1 P_2\, x_i - a\| \leq \|x_i - a\| \quad \forall\, i \geq 0 \tag{1828}$$

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; [189, §1.2] [30, §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 $\mathcal{C}_1 \cap \mathcal{C}_2$ is nonempty, convergence is to some point there by the *distance theorem.* Otherwise, $x_i$ converges to a point in $\mathcal{C}_1$ of minimum distance to $\mathcal{C}_2$.

**E.10.2.0.2 Example.** *Hyperplane/orthant intersection.*
Find a feasible point (1821) belonging to the nonempty intersection of two convex sets: given $A \in \mathbb{R}^{m\times n}$, $\beta \in \mathcal{R}(A)$

$$\mathcal{C}_1 \cap \mathcal{C}_2 \;=\; \mathbb{R}^n_+ \cap \mathcal{A} \;=\; \{y \mid y \succeq 0\} \cap \{y \mid Ay = \beta\} \subset \mathbb{R}^n \tag{1829}$$

the nonnegative orthant with affine subset $\mathcal{A}$ an intersection of hyperplanes. Projection of an iterate $x_i \in \mathbb{R}^n$ on $\mathcal{A}$ is calculated

$$P_2\, x_i \;=\; x_i - A^T(AA^T)^{-1}(Ax_i - \beta) \qquad (1720)$$

[E.19] Other authors prove convergence by different means; *e.g.*, [125] [48].

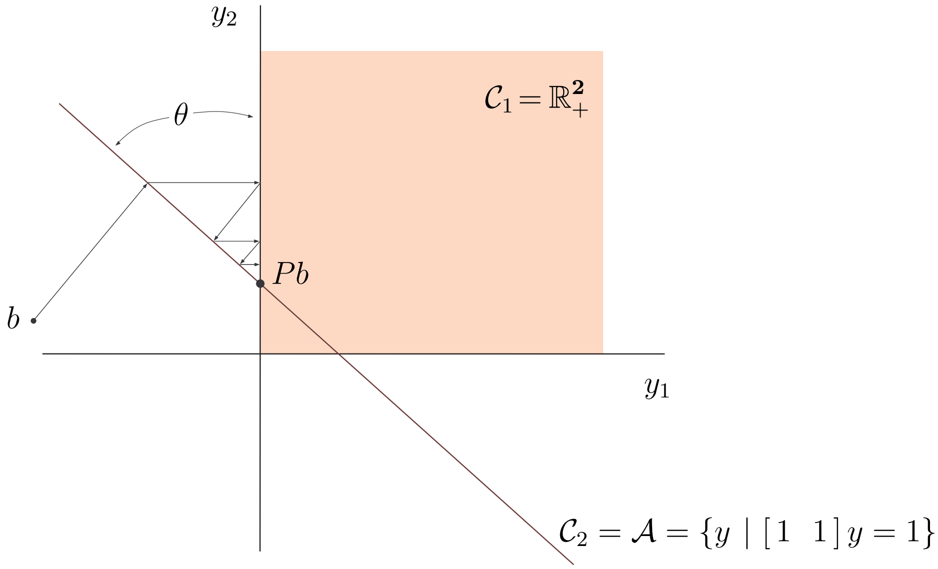



Figure 125: From Example E.10.2.0.2 in $\mathbb{R}^2$, showing von Neumann-style alternating projection to find feasible point belonging to intersection of nonnegative orthant with hyperplane. Point $Pb$ lies at intersection of hyperplane with ordinate axis. In this particular example, the feasible point found is coincidentally optimal. Rate of convergence depends upon angle $\theta$; as it becomes more acute, convergence slows. [125, §3]

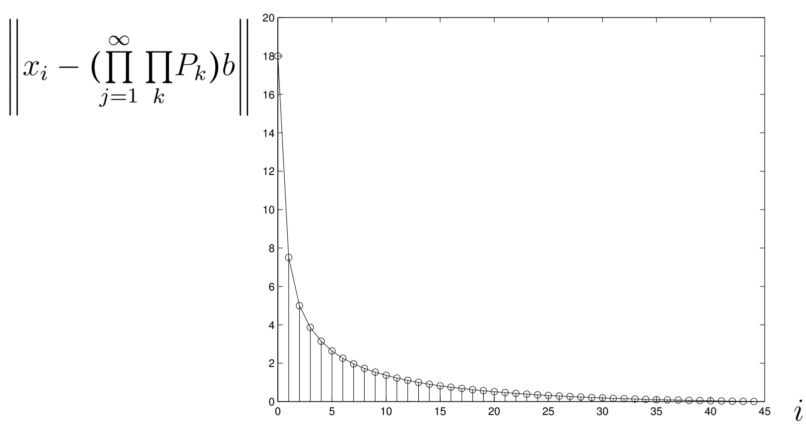



Figure 126: Geometric convergence of iterates in norm, for Example E.10.2.0.2 in $\mathbb{R}^{1000}$.

while, thereafter, projection of the result on the orthant is simply

$$x_{i+1} = P_1 P_2 x_i = \max\{\mathbf{0}, P_2 x_i\} \tag{1830}$$

where the maximum is entrywise (§E.9.2.2.3).

One realization of this problem in $\mathbb{R}^{\mathbf{2}}$ is illustrated in Figure **125**: For $A = [\,1 \quad 1\,]$, $\beta = 1$, and $x_0 = b = [\,-3 \quad 1/2\,]^T$, the iterates converge to the feasible point $Pb = [\,0 \quad 1\,]^T$.

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 $A \in \mathbb{R}^{\mathbf{400 \times 1000}}$, $\beta \in \mathbb{R}^{\mathbf{400}}$, and $b \in \mathbb{R}^{\mathbf{1000}}$, 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 **126**. □

This application of alternating projection to feasibility is extensible to any finite number of closed convex sets.

**E.10.2.0.3 Example.** *Under- and over-projection.* [43, §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 $\mathcal{C}$ and then project that same point $x_0$ on convex set $\mathcal{D}$. To the first iterate we assign $x_1 = (P_{\mathcal{C}}(x_0) + P_{\mathcal{D}}(x_0))\frac{1}{2}$. More generally,

$$x_{i+1} = (P_{\mathcal{C}}(x_i) + P_{\mathcal{D}}(x_i))\frac{1}{2}\,, \quad i = 0, 1, 2 \ldots \tag{1831}$$

Because the Cartesian product of convex sets remains convex, (§2.1.8) we can reformulate this problem.

Consider the convex set

$$\mathcal{S} \triangleq \begin{bmatrix} \mathcal{C} \\ \mathcal{D} \end{bmatrix} \tag{1832}$$

representing Cartesian product $\mathcal{C} \times \mathcal{D}$. Now, those two projections $P_{\mathcal{C}}$ and $P_{\mathcal{D}}$ are equivalent to one projection on the Cartesian product; *id est*,

$$P_{\mathcal{S}}\left(\begin{bmatrix} x_i \\ x_i \end{bmatrix}\right) = \begin{bmatrix} P_{\mathcal{C}}(x_i) \\ P_{\mathcal{D}}(x_i) \end{bmatrix} \tag{1833}$$

Define the subspace

$$\mathcal{R} \triangleq \left\{ v \in \begin{bmatrix} \mathbb{R}^n \\ \mathbb{R}^n \end{bmatrix} \;\middle|\; [\,I \quad -I\,]\, v = \mathbf{0} \right\} \tag{1834}$$

By the results in Example E.5.0.0.6

$$P_{\mathcal{RS}}\left(\begin{bmatrix} x_i \\ x_i \end{bmatrix}\right) = P_{\mathcal{R}}\left(\begin{bmatrix} P_{\mathcal{C}}(x_i) \\ P_{\mathcal{D}}(x_i) \end{bmatrix}\right) = \begin{bmatrix} P_{\mathcal{C}}(x_i) + P_{\mathcal{D}}(x_i) \\ P_{\mathcal{C}}(x_i) + P_{\mathcal{D}}(x_i) \end{bmatrix}\frac{1}{2} \tag{1835}$$

This means the proposed variation of alternating projection is equivalent to an alternation of projection on convex sets $\mathcal{S}$ and $\mathcal{R}$. If $\mathcal{S}$ and $\mathcal{R}$ intersect, these iterations will converge to a point in their intersection; hence, to a point in the intersection of $\mathcal{C}$ and $\mathcal{D}$.

We need not apply equal weighting to the projections, as supposed in (1831). In that case, definition of $\mathcal{R}$ 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 **126** employs a redundant sequence: The first sequence (indexed by $j$) estimates point $(\prod_{j=1}^{\infty}\prod_k P_k)b$ in the presumably nonempty intersection, then the quantity

$$\left\| x_i - \left(\prod_{j=1}^{\infty}\prod_k P_k\right) b \right\| \tag{1836}$$

in second sequence $x_i$ is observed *per* iteration $i$ for convergence. *A priori* knowledge of a feasible point (1821) is both impractical and antithetical. We need another measure:

Nonexpansivity implies

$$\left\| \left(\prod_{\ell} P_\ell\right) x_{k,i-1} - \left(\prod_{\ell} P_\ell\right) x_{ki} \right\| = \|x_{ki} - x_{k,i+1}\| \leq \|x_{k,i-1} - x_{ki}\| \tag{1837}$$

where

$$x_{ki} \triangleq P_k x_{k+1,i} \in \mathbb{R}^n \tag{1838}$$

represents unique minimum-distance projection of $x_{k+1,i}$ on convex set $k$ at iteration $i$. So a good convergence measure is the total monotonic sequence

$$\varepsilon_i \triangleq \sum_k \|x_{ki} - x_{k,i+1}\|\,, \qquad i = 0, 1, 2 \ldots \tag{1839}$$

where $\lim\limits_{i\to\infty} \varepsilon_i = 0$ 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

$$X \succeq 0\,, \qquad \langle A_j\;,\;X\rangle = b_j\;, \quad j=1\ldots m \tag{1840}$$

given nonzero $A_j \in \mathbb{S}^n$ and real $b_j$. Here we take $\mathcal{C}_1$ to be the positive semidefinite cone $\mathbb{S}^n_+$ while $\mathcal{C}_2$ is the affine subset of $\mathbb{S}^n$

$$\begin{aligned}\mathcal{C}_2 \;=\; \mathcal{A} \;&\triangleq\; \{X \mid \operatorname{tr}(A_j X) = b_j\;,\; j=1\ldots m\} \subseteq \mathbb{S}^n \\ &=\; \{X \mid \begin{bmatrix}\operatorname{svec}(A_1)^T \\ \vdots \\ \operatorname{svec}(A_m)^T\end{bmatrix} \operatorname{svec} X = b\} \\ &\triangleq\; \{X \mid A \operatorname{svec} X = b\}\end{aligned} \tag{1841}$$

where $b = [b_j] \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n(n+1)/2}$, and symmetric vectorization svec is defined by (47). Projection of iterate $X_i \in \mathbb{S}^n$ on $\mathcal{A}$ is: (§E.5.0.0.6)

$$P_2 \operatorname{svec} X_i = \operatorname{svec} X_i - A^\dagger (A \operatorname{svec} X_i - b) \tag{1842}$$

Euclidean distance from $X_i$ to $\mathcal{A}$ is therefore

$$\operatorname{dist}(X_i\;,\;\mathcal{A}) \;=\; \|X_i - P_2 X_i\|_{\mathrm{F}} \;=\; \|A^\dagger (A \operatorname{svec} X_i - b)\|_2 \tag{1843}$$

Projection of $P_2 X_i \triangleq \sum\limits_j \lambda_j\, q_j q_j^T$ on the positive semidefinite cone (§7.1.2) is found from its eigen decomposition (§A.5.2);

$$P_1 P_2 X_i = \sum_{j=1}^{n} \max\{0\;,\;\lambda_j\}\, q_j q_j^T \tag{1844}$$

Distance from $P_2 X_i$ to the positive semidefinite cone is therefore

$$\operatorname{dist}(P_2 X_i\;,\;\mathbb{S}^n_+) \;=\; \|P_2 X_i - P_1 P_2 X_i\|_{\mathrm{F}} \;=\; \sqrt{\sum_{j=1}^{n} \min\{0\,,\lambda_j\}^2} \tag{1845}$$

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 point feasible with (1840) exists, then there exists an $X \in \mathcal{A} \cap \mathbb{S}^n_+$ such that

$$\operatorname{rank} X \leq \left\lfloor \frac{\sqrt{8m+1}-1}{2} \right\rfloor \qquad (232)$$

□

**E.10.2.1.2 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 $\{E_{\ell q} \in \mathbb{S}^n\}$ (50)

$$\{A_j \in \mathbb{S}^n,\ j=1\ldots m\} \subseteq \{E_{\ell q}\} = \left\{ \begin{array}{ll} e_\ell e_\ell^T\,, & \ell = q = 1\ldots n \\ \frac{1}{\sqrt{2}}(e_\ell e_q^T + e_q e_\ell^T)\,, & 1 \leq \ell < q \leq n \end{array} \right\} \tag{1846}$$

and when the constants $b_j$ are set to constrained entries of variable $X \triangleq [X_{\ell q}] \in \mathbb{S}^n$

$$\{b_j\,,\ j=1\ldots m\} \subseteq \left\{ \begin{array}{ll} X_{\ell q}\,, & \ell = q = 1\ldots n \\ X_{\ell q}\sqrt{2}\,, & 1 \leq \ell < q \leq n \end{array} \right\} = \{\langle X, E_{\ell q}\rangle\} \tag{1847}$$

then the equality constraints in (1840) 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* (1842))

$$P_2 \operatorname{svec} X_i = \operatorname{svec} X_i - A^T(A \operatorname{svec} X_i - b) \tag{1848}$$

From this we can see that orthogonal projection is achieved simply by setting corresponding entries of $P_2 X_i$ to the known entries of $X$, while the remaining entries of $P_2 X_i$ are set to corresponding entries of the current iterate $X_i$.

Using this technique, we find a positive semidefinite completion for

$$\begin{bmatrix} 4 & 3 & ? & 2 \\ 3 & 4 & 3 & ? \\ ? & 3 & 4 & 3 \\ 2 & ? & 3 & 4 \end{bmatrix} \tag{1849}$$

Initializing the unknown entries to 0, they all converge geometrically to 1.5858 (rounded) after about 42 iterations.

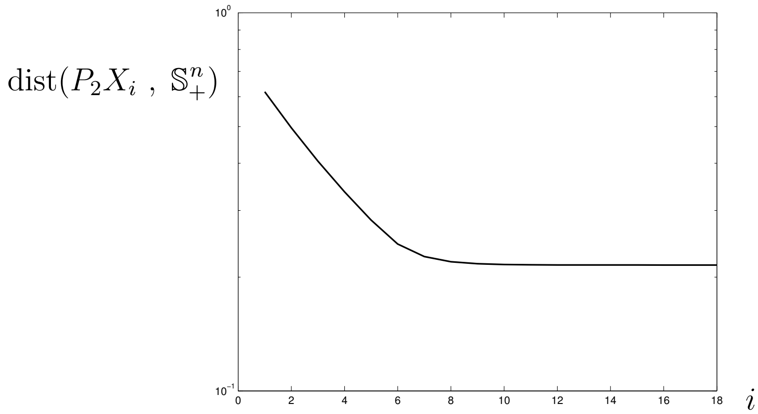



Figure 127: Distance (*confer* (1845)) between PSD cone and iterate (1848) in affine subset $\mathcal{A}$ (1841) for Laurent's completion problem; initially, decreasing geometrically.

Laurent gives a problem for which no positive semidefinite completion exists: [172]

$$\begin{bmatrix} 1 & 1 & ? & 0 \\ 1 & 1 & 1 & ? \\ ? & 1 & 1 & 1 \\ 0 & ? & 1 & 1 \end{bmatrix} \tag{1850}$$

Initializing unknowns to 0, by alternating projection we find the constrained matrix closest to the positive semidefinite cone,

$$\begin{bmatrix} 1 & 1 & 0.5454 & 0 \\ 1 & 1 & 1 & 0.5454 \\ 0.5454 & 1 & 1 & 1 \\ 0 & 0.5454 & 1 & 1 \end{bmatrix} \tag{1851}$$

and we find the positive semidefinite matrix closest to the affine subset $\mathcal{A}$ (1841):

$$\begin{bmatrix} 1.0521 & 0.9409 & 0.5454 & 0.0292 \\ 0.9409 & 1.0980 & 0.9451 & 0.5454 \\ 0.5454 & 0.9451 & 1.0980 & 0.9409 \\ 0.0292 & 0.5454 & 0.9409 & 1.0521 \end{bmatrix} \tag{1852}$$

These matrices (1851) and (1852) attain the Euclidean distance $\text{dist}(\mathcal{A}\,, \mathbb{S}^n_+)$. Convergence is illustrated in Figure **127**. $\square$

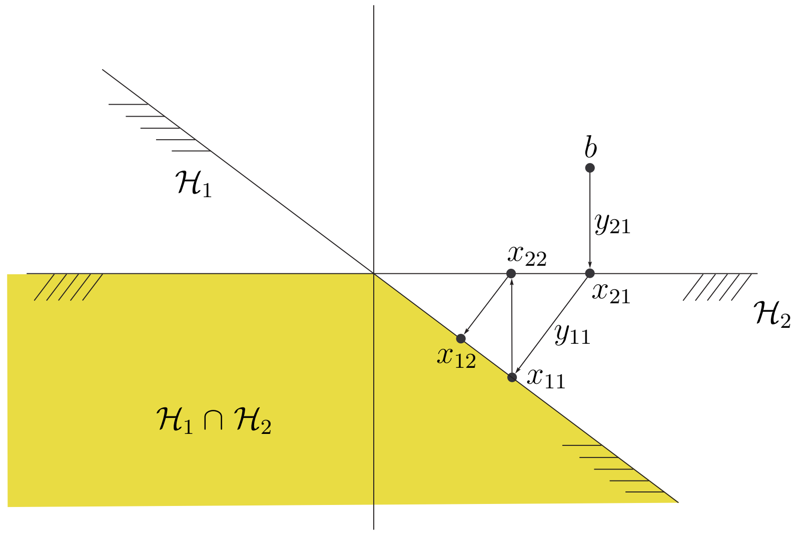



Figure 128: $\mathcal{H}_1$ and $\mathcal{H}_2$ are the same halfspaces as in Figure **123**. Dykstra's alternating projection algorithm generates the alternations $b\,, x_{21}\,, x_{11}\,, x_{22}\,, x_{12}\,, x_{12}\ldots,$. The path illustrated from $b$ to $x_{12}$ in $\mathbb{R}^2$ terminates at the desired result, $Pb$. The alternations are not so robust in presence of noise as for the example in Figure **122**.

### 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 [84] [47] who in 1983 provided an elegant algorithm that prevails today. In 1988, Han [128] rediscovered the algorithm and provided a primal−dual convergence proof. A synopsis of the history of alternating projection[E.20] can be found in [49] where it becomes apparent that Dykstra's work is seminal.

[E.20] For a synopsis of alternating projection applied to distance geometry, see [264, §3.1].

#### E.10.3.1 Dykstra's algorithm

Assume we are given some point $b\in\mathbb{R}^n$ and closed convex sets $\{\mathcal{C}_k\subset\mathbb{R}^n \mid k=1\ldots L\}$. 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 **128** and Figure **129**) associated with set $k$ at iteration $i$. Initialize

$$y_{k0} = 0 \quad \forall\, k=1\ldots L \qquad \text{and} \qquad x_{1,0} = b \tag{1853}$$

Denoting by $P_k t$ the unique minimum-distance projection of $t$ on $\mathcal{C}_k$ , and for convenience $x_{L+1,i} \triangleq x_{1,i-1}$ , calculation of the iterates $x_{1i}$ proceeds:[E.21]

$$\begin{array}{l}
\texttt{for } i=1,2,\ldots \texttt{ until convergence } \{ \\
\quad \texttt{for } k=L\ldots 1\ \{ \\
\qquad t = x_{k+1,i} - y_{k,i-1} \\
\qquad x_{ki} = P_k t \\
\qquad y_{ki} = P_k t - t \\
\quad \} \\
\}
\end{array} \tag{1854}$$

Assuming a nonempty intersection, then the iterates converge to the unique minimum-distance projection of point $b$ on that intersection; [73, §9.24]

$$Pb = \lim_{i\to\infty} x_{1i} \tag{1855}$$

In the case all the $\mathcal{C}_k$ are affine, then calculation of $y_{ki}$ is superfluous and the algorithm becomes identical to alternating projection. [73, §9.26] [99, §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.10.3.2 Normal cone

Glunt [106, §4] observes that the overall effect of Dykstra's iterative procedure is to drive $t$ toward the translated normal cone to $\bigcap \mathcal{C}_k$ at the solution $Pb$ (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 $\mathcal{C}_k$ touching $Pb$. The relative interior of the normal cone subtends these normal vectors.

[E.21] We reverse order of projection ($k=L\ldots 1$) in the algorithm for continuity of exposition.

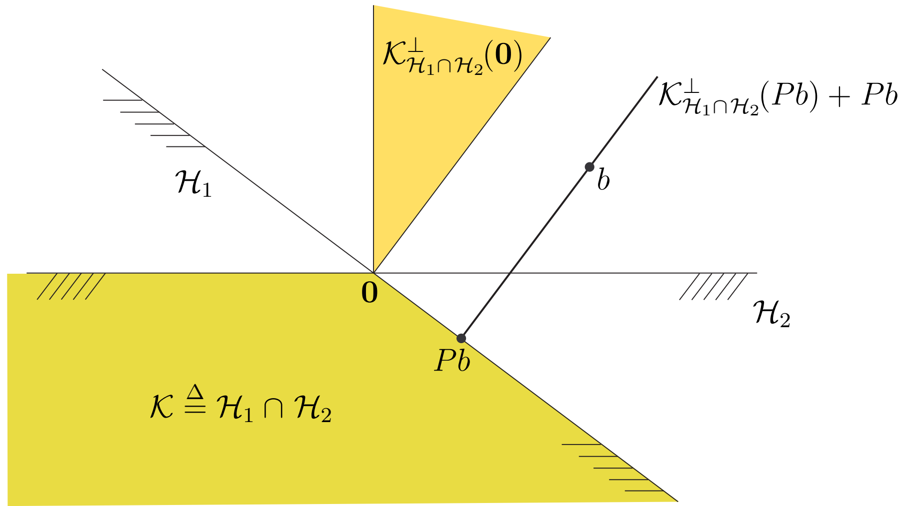



Figure 129: Two examples (truncated): Normal cone to $\mathcal{H}_1\cap\mathcal{H}_2$ at the origin, and at point $Pb$ on the boundary. $\mathcal{H}_1$ and $\mathcal{H}_2$ are the same halfspaces from Figure **128**. The normal cone at the origin $\mathcal{K}^{\perp}_{\mathcal{H}_1\cap\mathcal{H}_2}(\mathbf{0})$ is simply $-\mathcal{K}^*$.

**E.10.3.2.1 Definition.** *Normal cone.* [196] [30, p.261] [148, §A.5.2] [41, §2.1] [229, §3] The normal cone to any set $\mathcal{S}\subseteq\mathbb{R}^n$ at any particular point $a\in\mathbb{R}^n$ is defined as the closed cone

$$\mathcal{K}^{\perp}_{\mathcal{S}}(a)\triangleq\{z\in\mathbb{R}^n \mid z^T(y-a)\leq 0 \quad \forall\, y\in\mathcal{S}\} \;=\; -(\mathcal{S}-a)^* \tag{1856}$$

an intersection of halfspaces about the origin in $\mathbb{R}^n$ hence convex regardless of the convexity of $\mathcal{S}$ ; the negative dual cone to the translate $\mathcal{S}-a$ . $\triangle$

Examples of normal cone construction are illustrated in Figure **129**: The normal cone at the origin is the vector sum (§2.1.8) of two normal cones; [41, §3.3, exer.10] for $\mathcal{H}_1\cap\operatorname{int}\mathcal{H}_2\neq\emptyset$

$$\mathcal{K}^{\perp}_{\mathcal{H}_1\cap\mathcal{H}_2}(\mathbf{0})=\mathcal{K}^{\perp}_{\mathcal{H}_1}(\mathbf{0})+\mathcal{K}^{\perp}_{\mathcal{H}_2}(\mathbf{0}) \tag{1857}$$

This formula applies more generally to other points in the intersection.

The normal cone to any affine set $\mathcal{A}$ at $\alpha\in\mathcal{A}$, for example, is the orthogonal complement of $\mathcal{A}-\alpha$ . Projection of any point in the translated normal cone $\mathcal{K}^{\perp}_{\mathcal{C}}(a\in\mathcal{C})+a$ on convex set $\mathcal{C}$ is identical to $a$ ; in other words, point $a$ is that point in $\mathcal{C}$ closest to any point belonging to the translated normal cone $\mathcal{K}^{\perp}_{\mathcal{C}}(a)+a$ ; *e.g.*, Theorem E.4.0.0.1.

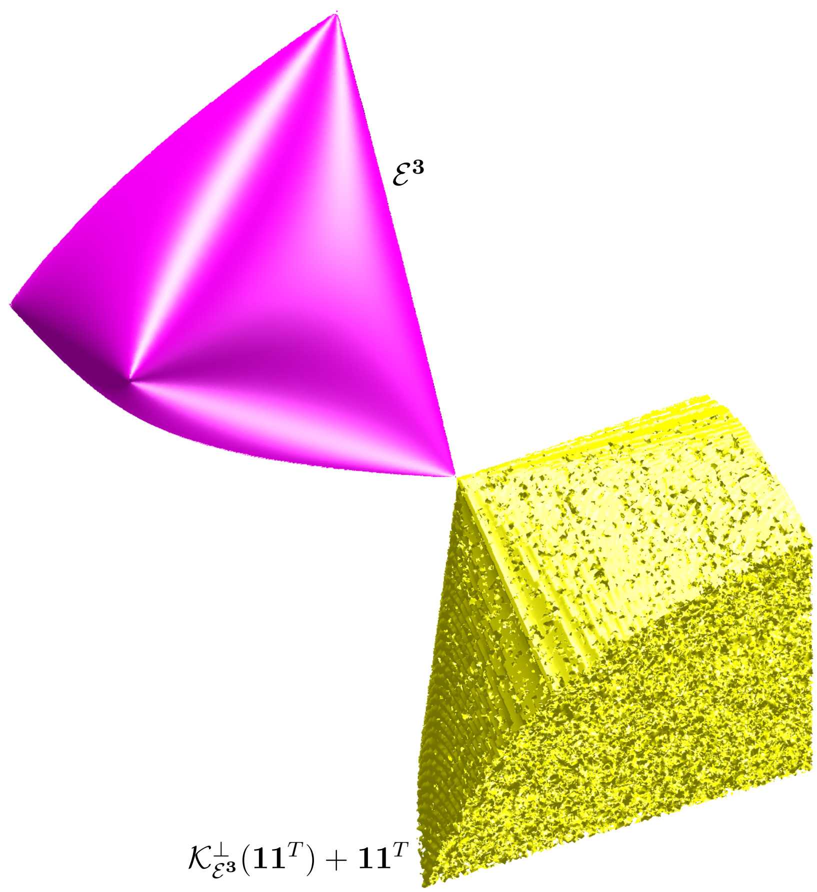
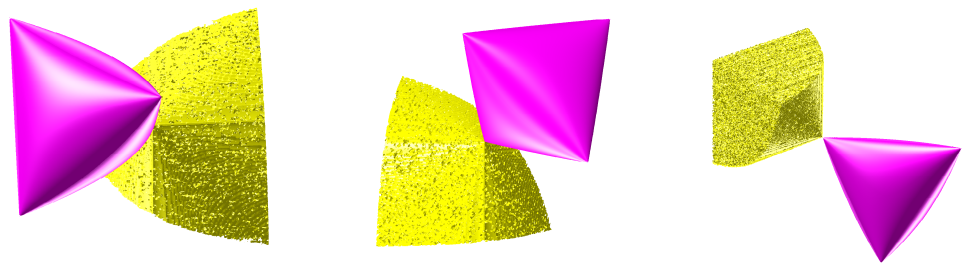



Figure 130: A few renderings (next page) of normal cone $\mathcal{K}_{\mathcal{E}^{\mathbf{3}}}^{\perp}$ to elliptope $\mathcal{E}^{\mathbf{3}}$ (Figure **87**) at point $\mathbf{1}\mathbf{1}^T$, projected on $\mathbb{R}^{\mathbf{3}}$. In [173, fig.2], normal cone is claimed circular in this dimension (severe numerical artifacts corrupt boundary and make interior corporeal, drawn truncated).



When set $\mathcal{S}$ is a convex cone $\mathcal{K}$, then the normal cone to $\mathcal{K}$ at the origin

$$\mathcal{K}_{\mathcal{K}}^{\perp}(\mathbf{0}) = -\mathcal{K}^* \tag{1858}$$

is the negative dual cone. Any point belonging to $-\mathcal{K}^*$, projected on $\mathcal{K}$, projects on the origin. More generally, [73, §4.5]

$$\mathcal{K}_{\mathcal{K}}^{\perp}(a) = -(\mathcal{K} - a)^* \tag{1859}$$

$$\mathcal{K}_{\mathcal{K}}^{\perp}(a \in \mathcal{K}) = -\mathcal{K}^* \cap a^{\perp} \tag{1860}$$

The normal cone to $\bigcap \mathcal{C}_k$ at $Pb$ in Figure **123** is the ray $\{\xi(b-Pb) \mid \xi \geq 0\}$ illustrated in Figure **129**. Applying Dykstra's algorithm to that example, convergence to the desired result is achieved in two iterations as illustrated in Figure **128**. Yet applying Dykstra's algorithm to the example in Figure **122** 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.

#### E.10.3.3 speculation

From these few examples we surmise, unique minimum-distance projection on *blunt* polyhedral cones having nonempty interior may be found by Dykstra's algorithm in few iterations.

# Appendix F

# MATLAB programs

Made by The MathWorks http://www.mathworks.com , MATLAB is a high level programming language and graphical user interface for linear algebra.

## F.1 isedm()

```
% Is real D a Euclidean Distance Matrix. -Jon Dattorro
%
% [Dclosest,X,isisnot,r] = isedm(D,tolerance,verbose,dimension,V)
%
% Returns: closest EDM in Schoenberg sense (default output),
%          a generating list X,
%          string 'is' or 'isnot' EDM,
%          actual affine dimension r of EDM output.
% Input: matrix D,
%        optional absolute numerical tolerance for EDM determination,
%        optional verbosity 'on' or 'off',
%        optional desired affine dim of generating list X output,
%        optional choice of 'Vn' auxiliary matrix (default) or 'V'.

function [Dclosest,X,isisnot,r] = isedm(D,tolerance_in,verbose,dim,V);

isisnot = 'is';
N = length(D);
```



Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

```
if nargin < 2 | isempty(tolerance_in)
   tolerance_in = eps;
end
tolerance = max(tolerance_in, eps*N*norm(D));
if nargin < 3 | isempty(verbose)
   verbose = 'on';
end
if nargin < 5 | isempty(V)
   use = 'Vn';
else
   use = 'V';
end

% is empty
if N < 1
   if strcmp(verbose,'on'), disp('Input D is empty.'), end
   X = [ ];
   Dclosest = [ ];
   isisnot = 'isnot';
   r = [ ];
   return
end
% is square
if size(D,1) ~= size(D,2)
   if strcmp(verbose,'on'), disp('An EDM must be square.'), end
   X = [ ];
   Dclosest = [ ];
   isisnot = 'isnot';
   r = [ ];
   return
end
% is real
if ~isreal(D)
   if strcmp(verbose,'on'), disp('Because an EDM is real,'), end
   isisnot = 'isnot';
   D = real(D);
end
```

```
% is nonnegative
if sum(sum(chop(D,tolerance) < 0))
   isisnot = 'isnot';
   if strcmp(verbose,'on'), disp('Because an EDM is nonnegative,'),end
end
% is symmetric
if sum(sum(abs(chop((D - D')/2,tolerance)) > 0))
   isisnot = 'isnot';
   if strcmp(verbose,'on'), disp('Because an EDM is symmetric,'), end
   D = (D + D')/2;  % only required condition
end
% has zero diagonal
if sum(abs(diag(chop(D,tolerance))) > 0)
   isisnot = 'isnot';
   if strcmp(verbose,'on')
      disp('Because an EDM has zero main diagonal,')
   end
end
% is EDM
if strcmp(use,'Vn')
   VDV = -Vn(N)'*D*Vn(N);
else
   VDV = -Vm(N)'*D*Vm(N);
end
[Evecs Evals] = signeig(VDV);
if ~isempty(find(chop(diag(Evals),...
            max(tolerance_in,eps*N*normest(VDV))) < 0))
   isisnot = 'isnot';
   if strcmp(verbose,'on'), disp('Because -VDV < 0,'), end
end
if strcmp(verbose,'on')
   if strcmp(isisnot,'isnot')
      disp('matrix input is not EDM.')
   elseif tolerance_in == eps
      disp('Matrix input is EDM to machine precision.')
   else
      disp('Matrix input is EDM to specified tolerance.')
   end
```

```
end

% find generating list
r = max(find(chop(diag(Evals),...
        max(tolerance_in,eps*N*normest(VDV))) > 0));
if isempty(r)
   r = 0;
end
if nargin < 4 | isempty(dim)
   dim = r;
else
   dim = round(dim);
end
t = r;
r = min(r,dim);
if r == 0
   X = zeros(1,N);
else
   if strcmp(use,'Vn')
      X = [zeros(r,1) diag(sqrt(diag(Evals(1:r,1:r))))*Evecs(:,1:r)'];
   else
      X = [diag(sqrt(diag(Evals(1:r,1:r))))*Evecs(:,1:r)']/sqrt(2);
   end
end
if strcmp(isisnot,'isnot') | dim < t
   Dclosest = Dx(X);
else
   Dclosest = D;
end
```

## F.1.1 Subroutines for isedm()

### F.1.1.1 chop()

```
% zeroing entries below specified absolute tolerance threshold
% -Jon Dattorro
function Y = chop(A,tolerance)

R = real(A);
I = imag(A);

if nargin == 1
   tolerance = max(size(A))*norm(A)*eps;
end
idR = find(abs(R) < tolerance);
idI = find(abs(I) < tolerance);

R(idR) = 0;
I(idI) = 0;

Y = R + i*I;
```

### F.1.1.2 Vn()

```
function y = Vn(N)

y = [-ones(1,N-1);
      eye(N-1)]/sqrt(2);
```

### F.1.1.3 Vm()

```
% returns EDM V matrix
function V = Vm(n)

V = [eye(n)-ones(n,n)/n];
```

#### F.1.1.4 signeig()

```
% Sorts signed real part of eigenvalues
% and applies sort to values and vectors.
% [Q, lam] = signeig(A)
% -Jon Dattorro

function [Q, lam] = signeig(A);

[q l] = eig(A);

lam = diag(l);
[junk id] = sort(real(lam));
id = id(length(id):-1:1);
lam = diag(lam(id));
Q = q(:,id);

if nargout < 2
   Q = diag(lam);
end
```

#### F.1.1.5 Dx()

```
% Make EDM from point list
function D = Dx(X)

[n,N] = size(X);
one = ones(N,1);

del = diag(X'*X);
D = del*one' + one*del' - 2*X'*X;
```

## F.2 conic independence, conici()

(§2.10) The recommended subroutine lp() (§F.2.1) is a linear program solver (*simplex method*) from MATLAB's *Optimization Toolbox* v2.0 (R11). Later releases of MATLAB replace lp() with linprog() (interior-point method) that we find quite inferior to lp() on an assortment of problems; indeed, inherent limitation of numerical precision to 1E-8 in linprog() causes failure in programs previously working with lp().

Given an arbitrary set of directions, this c.i. subroutine removes the conically dependent members. Yet a conically independent set returned is not necessarily unique. In that case, if desired, the set returned may be altered by reordering the set input.

```
% Test for c.i. of arbitrary directions in rows or columns of X.
% -Jon Dattorro

function [Xci, indep_str, how_many_depend] = conici(X,rowORcol,tol);

if nargin < 3
   tol=max(size(X))*eps*norm(X);
end
if nargin < 2 | strcmp(rowORcol,'col')
   rowORcol = 'col';
   Xin = X;
elseif strcmp(rowORcol,'row')
   Xin = X';
else
   disp('Invalid rowORcol input.')
   return
end
[n, N] = size(Xin);

indep_str = 'conically independent';
how_many_depend = 0;
if rank(Xin) == N
   Xci = X;
   return
end
```

```
count = 1;
new_N = N;
% remove zero rows or columns
for i=1:N
   if chop(Xin(:,count),tol)==0
      how_many_depend = how_many_depend + 1;
      indep_str = 'conically Dependent';
      Xin(:,count) = [ ];
      new_N = new_N - 1;
   else
      count = count + 1;
   end
end
% remove conic dependencies
count = 1;
newer_N = new_N;
for i=1:new_N
   if newer_N > 1
      A = [Xin(:,1:count-1) Xin(:,count+1:newer_N); -eye(newer_N-1)];
      b = [Xin(:,count); zeros(newer_N-1,1)];
      [a, lambda, how] = lp(zeros(newer_N-1,1),A,b,[ ],[ ],[ ],n,-1);
      if ~strcmp(how,'infeasible')
         how_many_depend = how_many_depend + 1;
         indep_str = 'conically Dependent';
         Xin(:,count) = [ ];
         newer_N = newer_N - 1;
      else
         count = count + 1;
      end
   end
end
if strcmp(rowORcol,'col')
   Xci = Xin;
else
   Xci = Xin';
end
```

### F.2.1 lp()

```
LP     Linear programming.
X=LP(f,A,b) solves the linear programming problem:

             min f'x    subject to:   Ax <= b
              x

X=LP(f,A,b,VLB,VUB) defines a set of lower and upper
bounds on the design variables, X, so that the solution is
always in the range VLB <= X <= VUB.

X=LP(f,A,b,VLB,VUB,X0) sets the initial starting point to X0.

X=LP(f,A,b,VLB,VUB,X0,N) indicates that the first N constraints
defined by A and b are equality constraints.

X=LP(f,A,b,VLB,VUB,X0,N,DISPLAY) controls the level of warning
messages displayed.  Warning messages can be turned off with
DISPLAY = -1.

[X,LAMBDA]=LP(f,A,b) returns the set of Lagrangian multipliers,
LAMBDA, at the solution.

[X,LAMBDA,HOW] = LP(f,A,b) also returns a string how that
indicates error conditions at the final iteration.

LP produces warning messages when the solution is either
unbounded or infeasible.
```

## F.3 Map of the USA

### F.3.1 EDM, mapusa()

(§5.13.1.0.1)

```
% Find map of USA using only distance information.
% -Jon Dattorro
% Reconstruction from EDM.
clear all;
close all;

load usalo; % From Matlab Mapping Toolbox
% http://www-ccs.ucsd.edu/matlab/toolbox/map/usalo.html

% To speed-up execution (decimate map data), make
% 'factor' bigger positive integer.
factor = 1;
Mg = 2*factor; % Relative decimation factors
Ms = factor;
Mu = 2*factor;

gtlakelat  = decimate(gtlakelat,Mg);
gtlakelon  = decimate(gtlakelon,Mg);
statelat   = decimate(statelat,Ms);
statelon   = decimate(statelon,Ms);
uslat      = decimate(uslat,Mu);
uslon      = decimate(uslon,Mu);

lat = [gtlakelat; statelat; uslat]*pi/180;
lon = [gtlakelon; statelon; uslon]*pi/180;
phi = pi/2 - lat;
theta = lon;
x = sin(phi).*cos(theta);
y = sin(phi).*sin(theta);
z = cos(phi);
```

```
% plot original data
plot3(x,y,z), axis equal, axis off

lengthNaN = length(lat);
id = find(isfinite(x));
X = [x(id)'; y(id)'; z(id)'];
N = length(X(1,:))

% Make the distance matrix
clear gtlakelat gtlakelon statelat statelon
clear factor x y z phi theta conus
clear uslat uslon Mg Ms Mu lat lon
D = diag(X'*X)*ones(1,N) + ones(N,1)*diag(X'*X)' - 2*X'*X;

% destroy input data
clear X

Vn = [-ones(1,N-1); speye(N-1)];
VDV = (-Vn'*D*Vn)/2;

clear D Vn
pack

[evec evals flag] = eigs(VDV, speye(size(VDV)), 10, 'LR');
if flag, disp('convergence problem'), return, end;
evals = real(diag(evals));

index = find(abs(evals) > eps*normest(VDV)*N);
n = sum(evals(index) > 0);
Xs = [zeros(n,1)  diag(sqrt(evals(index)))*evec(:,index)'];

warning off; Xsplot=zeros(3,lengthNaN)*(0/0); warning on;
Xsplot(:,id) = Xs;
figure(2)

% plot map found via EDM.
plot3(Xsplot(1,:), Xsplot(2,:), Xsplot(3,:))
axis equal, axis off
```

#### F.3.1.1 USA map input-data decimation, decimate()

```
function xd = decimate(x,m)
roll = 0;
rock = 1;
for i=1:length(x)
   if isnan(x(i))
      roll = 0;
      xd(rock) = x(i);
      rock=rock+1;
   else
      if ~mod(roll,m)
         xd(rock) = x(i);
         rock=rock+1;
      end
      roll=roll+1;
   end
end
xd = xd';
```

### F.3.2 EDM using ordinal data, omapusa()

(§5.13.2.1)

```
% Find map of USA using ordinal distance information.
% -Jon Dattorro
clear all;
close all;

load usalo; % From Matlab Mapping Toolbox
% http://www-ccs.ucsd.edu/matlab/toolbox/map/usalo.html

factor = 1;
Mg = 2*factor; % Relative decimation factors
Ms = factor;
Mu = 2*factor;

gtlakelat  = decimate(gtlakelat,Mg);
gtlakelon  = decimate(gtlakelon,Mg);
```

```
statelat   = decimate(statelat,Ms);
statelon   = decimate(statelon,Ms);
uslat      = decimate(uslat,Mu);
uslon      = decimate(uslon,Mu);

lat = [gtlakelat; statelat; uslat]*pi/180;
lon = [gtlakelon; statelon; uslon]*pi/180;
phi = pi/2 - lat;
theta = lon;
x = sin(phi).*cos(theta);
y = sin(phi).*sin(theta);
z = cos(phi);

% plot original data
plot3(x,y,z), axis equal, axis off

lengthNaN = length(lat);
id = find(isfinite(x));
X = [x(id)'; y(id)'; z(id)'];
N = length(X(1,:))

% Make the distance matrix
clear gtlakelat gtlakelon statelat
clear statelon state stateborder greatlakes
clear factor x y z phi theta conus
clear uslat uslon Mg Ms Mu lat lon
D = diag(X'*X)*ones(1,N) + ones(N,1)*diag(X'*X)' - 2*X'*X;

% ORDINAL MDS - vectorize D
count = 1;
M = (N*(N-1))/2;
f = zeros(M,1);
for j=2:N
   for i=1:j-1
         f(count) = D(i,j);
         count = count + 1;
   end
end
```

```
% sorted is f(idx)
[sorted idx] = sort(f);
clear D sorted X
f(idx)=((1:M).^2)/M^2;

% Create ordinal data matrix
O = zeros(N,N);
count = 1;
for j=2:N
   for i=1:j-1
         O(i,j) = f(count);
         O(j,i) = f(count);
         count = count+1;
   end
end

clear f idx

Vn = sparse([-ones(1,N-1); eye(N-1)]);
VOV = (-Vn'*O*Vn)/2;

clear O Vn
pack

[evec evals flag] = eigs(VOV, speye(size(VOV)), 10, 'LR');
if flag, disp('convergence problem'), return, end;
evals = real(diag(evals));

Xs = [zeros(3,1)  diag(sqrt(evals(1:3)))*evec(:,1:3)'];

warning off; Xsplot=zeros(3,lengthNaN)*(0/0); warning on;
Xsplot(:,id) = Xs;
figure(2)

% plot map found via Ordinal MDS.
plot3(Xsplot(1,:), Xsplot(2,:), Xsplot(3,:))
axis equal, axis off
```

## F.4 Rank reduction subroutine, RRf()

(§4.3.1.0.1)

```
% Rank Reduction function -Jon Dattorro
% Inputs are:
%   Xstar matrix,
%   affine equality constraint matrix A whose rows are in svec format.
%
% Tolerance scheme needs revision...

function X = RRf(Xstar,A);
rand('seed',23);
m = size(A,1);
n = size(Xstar,1);
if size(Xstar,1)~=size(Xstar,2)
   disp('Rank Reduction subroutine: Xstar not square'), pause
end
toler = norm(eig(Xstar))*size(Xstar,1)*1e-9;
if sum(chop(eig(Xstar),toler)<0) ~= 0
   disp('Rank Reduction subroutine: Xstar not PSD'), pause
end
X = Xstar;
for i=1:n
   [v,d]=signeig(X);
   d(find(d<0))=0;
   rho = rank(d);
   for l=1:rho
      R(:,l,i)=sqrt(d(l,l))*v(:,l);
   end
   % find Zi
   svectRAR=zeros(m,rho*(rho+1)/2);
   cumu=0;
   for j=1:m
      temp = R(:,1:rho,i)'*svectinv(A(j,:))*R(:,1:rho,i);
      svectRAR(j,:) = svect(temp)';
      cumu = cumu + abs(temp);
   end
```

```
   % try to find sparsity pattern for Z_i
   tolerance = norm(X,'fro')*size(X,1)*1e-9;
   Ztem = zeros(rho,rho);
   pattern = find(chop(cumu,tolerance)==0);
   if isempty(pattern) % if no sparsity, do random projection
      ranp = svect(2*(rand(rho,rho)-0.5));
      Z(1:rho,1:rho,i)...
       =svectinv((eye(rho*(rho+1)/2)-pinv(svectRAR)*svectRAR)*ranp);
   else
      disp('sparsity pattern found')
      Ztem(pattern)=1;
      Z(1:rho,1:rho,i) = Ztem;
   end
   phiZ = 1;
   toler = norm(eig(Z(1:rho,1:rho,i)))*rho*1e-9;
   if sum(chop(eig(Z(1:rho,1:rho,i)),toler)<0) ~= 0
      phiZ = -1;
   end
   B(:,:,i) = -phiZ*R(:,1:rho,i)*Z(1:rho,1:rho,i)*R(:,1:rho,i)';
   % calculate t_i^*
   t(i) = max(phiZ*eig(Z(1:rho,1:rho,i)))^-1;
   tolerance = norm(X,'fro')*size(X,1)*1e-6;
   if chop(Z(1:rho,1:rho,i),tolerance)==zeros(rho,rho)
      break
   else
      X = X + t(i)*B(:,:,i);
   end
end
```

### F.4.1 svect()

```
% Map from symmetric matrix to vector
% -Jon Dattorro

function y = svect(Y,N)

if nargin == 1
   N=size(Y,1);
end

y = zeros(N*(N+1)/2,1);
count = 1;
for j=1:N
   for i=1:j
      if i~=j
         y(count) = sqrt(2)*Y(i,j);
      else
         y(count) = Y(i,j);
      end
      count = count + 1;
   end
end
```

### F.4.2 svectinv()

```
% convert vector into symmetric matrix.  m is dim of matrix.
% -Jon Dattorro
function A = svectinv(y)

m = round((sqrt(8*length(y)+1)-1)/2);
if length(y) ~= m*(m+1)/2
   disp('dimension error in svectinv()');
   pause
end

A = zeros(m,m);
count = 1;
for j=1:m
   for i=1:m
      if i<=j
         if i==j
            A(i,i) = y(count);
         else
            A(i,j) = y(count)/sqrt(2);
            A(j,i) = A(i,j);
         end
         count = count+1;
      end
   end
end
```

## F.5 Sturm's procedure

This is a demonstration program that can easily be transformed to a subroutine for decomposing positive semidefinite matrix $X$. This procedure provides a nonorthogonal alternative (§A.7.5.0.1) to eigen decomposition. That particular decomposition obtained is dependent on choice of matrix $A$.

```
% Sturm procedure to find dyad-decomposition of X  -Jon Dattorro
clear all

N = 4;
r = 2;
X = 2*(rand(r,N)-0.5);
X = X'*X;

t = null(svect(X)');
A = svectinv(t(:,1));

% Suppose given matrix A is positive semidefinite
%[v,d] = signeig(X);
%d(1,1)=0; d(2,2)=0; d(3,3)=pi;
%A = v*d*v';

tol = 1e-8;
Y = X;
y = zeros(size(X,1),r);
rho = r;
for k=1:r
    [v,d] = signeig(Y);
    v = v*sqrt(chop(d,1e-14));
    viol = 0;
    j = [ ];
    for i=2:rho
        if chop((v(:,1)'*A*v(:,1))*(v(:,i)'*A*v(:,i)),tol) ~= 0
            viol = 1;
        end
        if (v(:,1)'*A*v(:,1))*(v(:,i)'*A*v(:,i)) < 0
            j = i;
```

```
            break
        end
    end
    if ~viol
        y(:,k) = v(:,1);
    else
     if isempty(j)
      disp('Sturm procedure taking default j'), j = 2; return
     end  % debug
     alpha = (-2*(v(:,1)'*A*v(:,j)) + sqrt((2*v(:,1)'*A*v(:,j)).^2 ...
      -4*(v(:,j)'*A*v(:,j))*(v(:,1)'*A*v(:,1)))))/(2*(v(:,j)'*A*v(:,j)));
     y(:,k) = (v(:,1) + alpha*v(:,j))/sqrt(1+alpha^2);
     if chop(y(:,k)'*A*y(:,k),tol) ~= 0
      alpha = (-2*(v(:,1)'*A*v(:,j)) - sqrt((2*v(:,1)'*A*v(:,j)).^2 ...
       -4*(v(:,j)'*A*v(:,j))*(v(:,1)'*A*v(:,1)))))/(2*(v(:,j)'*A*v(:,j)));
      y(:,k) = (v(:,1) + alpha*v(:,j))/sqrt(1+alpha^2);
      if chop(y(:,k)'*A*y(:,k),tol) ~= 0
       disp('Zero problem in Sturm!'), return
      end  % debug
     end
    end
    Y = Y - y(:,k)*y(:,k)';
    rho = rho - 1;
end
z = zeros(size(y));
e = zeros(N,N);
for i=1:r
    z(:,i) = y(:,i)/norm(y(:,i));
    e(i,i) = norm(y(:,i))^2;
end
lam = diag(e);
[junk id] = sort(real(lam));
id = id(length(id):-1:1);
z = [z(:,id(1:r)) null(z')]   % Sturm
e = diag(lam(id))
[v,d] = signeig(X)            % eigenvalue decomposition
X-z*e*z'
traceAX = trace(A*X)
```

## F.6 Convex iteration demonstration

We demonstrate implementation of a rank constraint in a semidefinite Boolean feasibility problem from §4.4.3.0.5. It requires `CVX`, [117] an intuitive MATLAB interface for interior-point method solvers.

There are a finite number $2^{N=50} \approx$ `1E15` of binary vectors $x$. The feasible set of semidefinite program (668) is the intersection of an elliptope with $M=10$ halfspaces in vectorized variable $G$. Size of the optimal rank-1 solution set is proportional to the positive factor scaling vector `b`. The smaller that optimal Boolean solution set, the harder this problem is to solve. That scale factor and initial states of random number generators, making matrix `A` and vector `b`, are selected to demonstrate Boolean solution to one instance in about 7 iterations (about 6 seconds), whereas a sequential binary search tests 25.7 million vectors (in one hour) before finding one Boolean solution feasible to nonconvex problem (665). (Other parameters can be selected to reverse these timings.)

```
% Discrete optimization problem demo.
% -Jon Dattorro, June 4, 2007
% Find  x\in{-1,1}^N  such that  Ax <= b
clear all;
format short g;
M = 10;
N = 50;
randn('state',0); rand('state',0);
A = randn(M,N);
b = rand(M,1)*5;

disp('Find binary solution by convex iteration:')
tic
Y = zeros(N+1);
count = 1;
traceGY = 1e15;
cvx_precision([1e-12, 1e-4]);
cvx_quiet(true);
```

```
while 1
    cvx_begin  % requires CVX Boyd
      variable X(N,N) symmetric;
      variable x(N,1);
      G = [X,  x;
           x', 1];
      minimize(trace(G*Y));
      diag(X) == 1;
      G == semidefinite(N+1);
      A*x <= b;
    cvx_end

    [v,d,q] = svd(G);
    Y = v(:,2:N+1)*v(:,2:N+1)';
    rankG = sum(diag(d) > max(diag(d))*1e-8)
    oldtrace = traceGY;
    traceGY = trace(G*Y)
    if rankG == 1
        break
    end

    digits = 1e3;
    if round((oldtrace - traceGY)*digits) == 0
        disp('STALLED');disp(' ');
        Y = -v(:,2:N+1)*(v(:,2:N+1)' + randn(N,1)*v(:,1)');
    end
    count = count + 1;
end
x
count
toc
disp('Ax <= b ,  x\in{-1,1}^N')
```

```
disp(' ');disp('Combinatorial search for a feasible binary solution:')
tic
for i=1:2^N
    binary = str2num(dec2bin(i-1)');
    binary(find(~binary)) = -1;
    y = [-ones(N-length(binary),1); binary];
    if sum(A*y <= b) == M
        disp('Feasible binary solution found.')
        y
        break
    end
end
toc
```

## F.7 FAST MAX CUT

We use the graph generator (C program) RUDY written by Giovanni Rinaldi [224] which can be found at http://convexoptimization.com/TOOLS/RUDY together with graph data. (§4.4.3.0.7)

```
% fast max cut, Jon Dattorro, July 2007, http://convexoptimization.com
clear all;
format short g; tic
fid = fopen('graphs12','r');
average = 0;
NN = 0;
s = fgets(fid);
cvx_precision([1e-12, 1e-4]);
cvx_quiet(true);
w = 1000;
while s ~= -1
    s = str2num(s);
    N = s(1);
    A = zeros(N);
    for i=1:s(2)
        s = str2num(fgets(fid));
        A(s(1),s(2)) = s(3);
        A(s(2),s(1)) = s(3);
    end
    Q = (diag(A*ones(N,1)) - A)/4;
    W = zeros(N);
    traceXW = 1e15;
    while 1
        cvx_begin                          % CVX Boyd
           variable X(N,N) symmetric;
           maximize(trace(Q*X) - w*trace(W*X));
           X == semidefinite(N);
           diag(X) == 1;
        cvx_end
        [v,d,q] = svd(X);
        W = v(:,2:N)*v(:,2:N)';
        rankX = sum(diag(d) > max(diag(d))*1e-8)
```

```
        oldtrace = traceXW;
        traceXW = trace(X*W)
        if (rankX == 1)
            break
        end
        if round((oldtrace - traceXW)*1e3) <= 0
            disp('STALLED');disp(' ')
            W = -v(:,2:N)*(v(:,2:N)' + randn(N-1,1)*v(:,1)');
        end
    end
    x = sqrt(d(1,1))*v(:,1)
    disp(' ');
    disp('Combinatorial search for optimal binary solution...')
    maxim = -1e15;
    ymax = zeros(N,1);
    for i=1:2^N
        binary = str2num(dec2bin(i-1)');
        binary(find(~binary)) = -1;
        y = [-ones(N-length(binary),1); binary];
        if y'*Q*y > maxim
            maxim = y'*Q*y;
            ymax = y;
        end
    end
    if (maxim == 0) && (abs(trace(Q*X)) <= 1e-8)
        optimality_ratio = 1
    elseif maxim <= 0
        optimality_ratio = maxim/trace(Q*X)
    else
        optimality_ratio = trace(Q*X)/maxim
    end
    ymax
    average = average + optimality_ratio;
    NN = NN + 1
    running_average = average/NN
    toc, disp(' ')
    s = fgets(fid);
end
```

# Appendix G

# Notation and a few definitions

$b$ vector, scalar, logical condition (italic $abcdefghijklmnopqrstuvwxyz$)

$b_i$ $i^{\text{th}}$ *entry of vector* $b=[b_i\,,\ i=1\ldots n]$ or $i^{\text{th}}$ $b$ *vector from a set or list* $\{b_j\,,\ j=1\ldots n\}$ or $i^{\text{th}}$ *iterate of vector* $b$

$b_{i:j}$ or $b(i\!:\!j)\,,$ truncated vector comprising $i^{\text{th}}$ through $j^{\text{th}}$ entry of vector $b$

$b_k(i\!:\!j)$ truncated vector comprising $i^{\text{th}}$ through $j^{\text{th}}$ entry of vector $b_k$

$b^T$ vector transpose

$b^H$ Hermitian (conjugate) transpose

$A^{-2T}$ matrix transpose of squared inverse

$A^{T_1}$ first of various transpositions of a cubix or quartix $A$

$A$ matrix, scalar, or logical condition (italic $ABCDEFGHIJKLMNOPQRSTUVWXYZ$)

*skinny* a skinny matrix; meaning, more rows than columns: $\left[\begin{array}{c} \\ \\ \\ \end{array}\right]$. *When there are more equations than unknowns, we say that the system* $Ax=b$ *is* overdetermined. [110, §5.3]

*fat* a fat matrix; meaning, more columns than rows: $\left[\qquad\qquad\right]$ *underdetermined*

 
Citation: Jon Dattorro, *Convex Optimization & Euclidean Distance Geometry,* Meboo Publishing USA, 2005.

| | |
|---|---|
| $\mathcal{A}$ | some set (calligraphic $\mathcal{ABCDEFGHIJKLMNOPQRSTUVWXYZ}$) |
| $\mathcal{F}(\mathcal{C} \ni A)$ | smallest face (138) that contains element $A$ of set $\mathcal{C}$ |
| $\mathcal{G}(\mathcal{K})$ | generators (§2.8.1.2) of set $\mathcal{K}$; any collection of points and directions whose hull constructs $\mathcal{K}$ |
| $A^{-1}$ | inverse of matrix $A$ |
| $A^{\dagger}$ | Moore-Penrose pseudoinverse of matrix $A$ |
| $\sqrt{\ }$ | positive square root |
| $A^{1/2}$ and $\sqrt{A}$ | $A^{1/2}$ is any matrix such that $A^{1/2}A^{1/2} = A$. For $A \in \mathbb{S}^n_+$, $\sqrt{A} \in \mathbb{S}^n_+$ is unique and $\sqrt{A}\sqrt{A} = A$. [41, §1.2] (§A.5.2.1) |
| $\sqrt[\circ]{D}$ | $\triangleq [\sqrt{d_{ij}}\,]$. (1153) *Hadamard positive square root:* $D = \sqrt[\circ]{D} \circ \sqrt[\circ]{D}$ . |
| $E$ | elementary matrix |
| $E_{ij}$ | member of standard orthonormal basis for symmetric (50) or symmetric hollow (64) matrices |
| $A_{ij}$ | $ij^{\text{th}}$ entry of matrix $A$ |
| $A_i$ | $i^{\text{th}}$ matrix from a set |
| $D_i$ | $i^{\text{th}}$ principal submatrix or $i^{\text{th}}$ iterate of $D$ |
| $A(:,i)$ | $i^{\text{th}}$ column of matrix $A$ [110, §1.1.8] |
| $A(j,:)$ | $j^{\text{th}}$ row of matrix $A$ |
| $A_{i:j,k:\ell}$ | or $A(i\!:\!j\,,\,k\!:\!\ell)$, submatrix taken from $i^{\text{th}}$ through $j^{\text{th}}$ row and $k^{\text{th}}$ through $\ell^{\text{th}}$ column |
| *e.g.* | *exempli gratia*, from the Latin meaning *for sake of example* |
| no. | *number*, from the Latin *numero* |
| a.i. | affinely independent |
| c.i. | conically independent |

| | |
|---|---|
| l.i. | linearly independent |
| w.r.t | *with respect to* |
| `a.k.a` | *also known as* |
| Re | real part |
| Im | imaginary part |
| $\imath$ or $\jmath$ | $\sqrt{-1}$ |
| $\subset \quad \supset \quad \cap \quad \cup$ | standard set theory, *subset, superset, intersection, union* |
| $\in$ | membership, *element belongs to*, or *element is a member of* |
| $\ni$ | membership, *contains* as in $\mathcal{C} \ni y$ ($\mathcal{C}$ contains element $y$) |
| ∍ | *such that* |
| $\exists$ | *there exists* |
| $\therefore$ | *therefore* |
| $\forall$ | *for all*, or *over all* |
| $\propto$ | *proportional to* |
| $\infty$ | infinity |
| $\equiv$ | *equivalent to* |
| $\triangleq$ | *defined equal to*, or *equal by definition* |
| $\approx$ | *approximately equal to* |
| $\simeq$ | *isomorphic to* or *with* |
| $\cong$ | *congruent to* or *with* |
| — | Hadamard quotient as in, for $x, y \in \mathbb{R}^n$, $\dfrac{x}{y} \triangleq [\, x_i/y_i \;,\; i=1 \ldots n \,] \in \mathbb{R}^n$ |
| $\circ$ | Hadamard product of matrices: $x \circ y \triangleq [\, x_i \, y_i \;,\; i=1 \ldots n \,] \in \mathbb{R}^n$ |

$\otimes$ Kronecker product of matrices (§D.1.2.1)

$\oplus$ vector sum of sets $\mathcal{X} = \mathcal{Y} \oplus \mathcal{Z}$ where every element $x \in \mathcal{X}$ has unique expression $x = y + z$ where $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$; [230, p.19] then the summands are algebraic complements. $\mathcal{X} = \mathcal{Y} \oplus \mathcal{Z} \Rightarrow \mathcal{X} = \mathcal{Y} + \mathcal{Z}$. Now assume $\mathcal{Y}$ and $\mathcal{Z}$ are nontrivial subspaces. $\mathcal{X} = \mathcal{Y} + \mathcal{Z} \Rightarrow \mathcal{X} = \mathcal{Y} \oplus \mathcal{Z} \Leftrightarrow \mathcal{Y} \cap \mathcal{Z} = \mathbf{0}$ [231, §1.2] [73, §5.8]. Each element from a vector sum ($+$) of subspaces has a unique representation ($\oplus$) when a basis from each subspace is linearly independent of bases from all the other subspaces.

$\ominus$ likewise, the vector difference of sets

$\boxplus$ orthogonal vector sum of sets $\mathcal{X} = \mathcal{Y} \boxplus \mathcal{Z}$ where every element $x \in \mathcal{X}$ has unique orthogonal expression $x = y + z$ where $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, and $y \perp z$. [247, p.51] $\mathcal{X} = \mathcal{Y} \boxplus \mathcal{Z} \Rightarrow \mathcal{X} = \mathcal{Y} + \mathcal{Z}$. If $\mathcal{Z} \subseteq \mathcal{Y}^\perp$ then $\mathcal{X} = \mathcal{Y} \boxplus \mathcal{Z} \Leftrightarrow \mathcal{X} = \mathcal{Y} \oplus \mathcal{Z}$. [73, §5.8] If $\mathcal{Z} = \mathcal{Y}^\perp$ then the summands are orthogonal complements.

$\pm$ *plus or minus*

$\perp$ as in $A \perp B$ meaning $A$ *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), $A \perp B \Leftrightarrow \langle A, B\rangle = 0$ $\Leftrightarrow \|A + B\|^2 = \|A\|^2 + \|B\|^2$

$\backslash$ as in $\backslash\mathcal{A}$ means logical *not* $\mathcal{A}$, or *relative complement of set* $\mathcal{A}$; *id est*, $\backslash\mathcal{A} = \{x \notin \mathcal{A}\}$; *e.g.*, $\mathcal{B}\backslash\mathcal{A} \triangleq \{x \in \mathcal{B} \mid x \notin \mathcal{A}\} \equiv \mathcal{B} \cap \backslash\mathcal{A}$

$\Rightarrow$ or $\Leftarrow$ sufficiency or necessity, *implies*; *e.g.*, $A \Rightarrow B \Leftrightarrow \backslash A \Leftarrow \backslash B$

$\Leftrightarrow$ *if and only if* (iff) or *corresponds to* or *necessary and sufficient* or *the same as*

*is* as in $A$ *is* $B$ means $A \Rightarrow B$; conventional usage of English language by mathematicians

$\nRightarrow$ or $\nLeftarrow$ *does not imply*

$\leftarrow$ *is replaced with*; substitution, assignment

$\rightarrow$ *goes to*, or *approaches*, or *maps to*

| | |
|---|---|
| $t \to 0^+$ | $t$ *goes to* 0 *from above*; meaning, *from the positive* [148, p.2] |
| : | as in $f:\mathbb{R}^n \to \mathbb{R}^m$ meaning $f$ *is a mapping*, or sequence of successive integers specified by bounds as in $i:j$ (if $j<i$ then sequence is descending) |
| $f:\mathcal{M}\to\mathcal{R}$ | meaning $f$ *is a mapping from ambient space* $\mathcal{M}$ *to ambient* $\mathcal{R}$ , not necessarily denoting either domain or range |
| $\vert$ | as in $f(x)\,\vert\, x\in\mathcal{C}$ means *with the condition(s)* or *such that* or *evaluated for*, or as in $\{f(x)\,\vert\, x\in\mathcal{C}\}$ means *evaluated for each and every* $x$ *belonging to set* $\mathcal{C}$ |
| $g\vert_{x_{\rm p}}$ | *expression* $g$ *evaluated at* $x_{\rm p}$ |
| $A\,,B$ | as in, for example, $A\,,B\in\mathbb{S}^N$ means $A\in\mathbb{S}^N$ and $B\in\mathbb{S}^N$ |
| $(A\,,B)$ | *open interval* between $A$ and $B$ in $\mathbb{R}$ , or *variable pair* perhaps of disparate dimension |
| $[A\,,B\,]$ | *closed interval* or *line segment* between $A$ and $B$ in $\mathbb{R}$ |
| ( ) | *hierarchal*, *parenthetical*, *optional* |
| { } | curly braces denote a set or list, *e.g.*, $\{Xa \mid a\succeq 0\}$ *the set of all* $Xa$ *for each and every* $a\succeq 0$ where membership of $a$ to some space is implicit, a *union* |
| $\langle\ \rangle$ | angle brackets denote vector inner-product (26) (31) |
| [ ] | matrix or vector, or quote insertion, or citation |
| $[d_{ij}]$ | matrix whose $ij^{\rm th}$ entry is $d_{ij}$ |
| $[x_i]$ | vector whose $i^{\rm th}$ entry is $x_i$ |
| $x_{\rm p}$ | particular value of $x$ |
| $x_0$ | particular instance of $x$ , or initial value of a sequence $x_i$ |
| $x_1$ | first entry of vector $x$ , or first element of a set or list $\{x_i\}$ |
| $x_\varepsilon$ | *extreme point* |

$x_+$ vector $x$ whose negative entries are replaced with $0$, or *clipped vector* $x$ or *nonnegative part of* $x$

$x^\star$ optimal value of variable $x$

$x^*$ *complex conjugate* or *dual variable*

$f^*$ *convex conjugate function*

$P_{\mathcal{C}}x$ or $Px$ projection of point $x$ on set $\mathcal{C}$, $P$ is operator or idempotent matrix

$P_k x$ projection of point $x$ on set $\mathcal{C}_k$ or on range of implicit vector

$\delta(A)$ (§A.1) *vector made from the main diagonal of* $A$ if $A$ is a matrix; otherwise, *diagonal matrix made from vector* $A$

$\delta^2(A)$ $\equiv \delta\big(\delta(A)\big)$. For vector or diagonal matrix $\Lambda$, $\delta^2(\Lambda) = \Lambda$

$\delta(A)^2$ $= \delta(A)\delta(A)$ where $A$ is a vector

$\lambda_i(X)$ $i^{\text{th}}$ entry of vector $\lambda$ is function of $X$

$\lambda(X)_i$ $i^{\text{th}}$ entry of vector-valued function of $X$

$\lambda(A)$ *vector of eigenvalues of matrix* $A$, (1265) typically arranged in nonincreasing order

$\sigma(A)$ *vector of singular values of matrix* $A$ (always arranged in nonincreasing order), or *support function*

$\Sigma$ 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 matrix

$\prod$ product

| | |
|---|---|
| $\psi(Z)$ | signum-like *step function* that returns a scalar for matrix argument (604), it returns a vector for vector argument (1365) |
| $D$ | symmetric hollow matrix of distance-square, or *Euclidean distance matrix* |
| $\mathbf{D}$ | Euclidean distance matrix operator |
| $\mathbf{D}^T(X)$ | adjoint operator |
| $\mathbf{D}(X)^T$ | transpose of $\mathbf{D}(X)$ |
| $\mathbf{D}^{-1}(X)$ | inverse operator |
| $\mathbf{D}(X)^{-1}$ | inverse of $\mathbf{D}(X)$ |
| $D^\star$ | optimal value of variable $D$ |
| $D^*$ | dual to variable $D$ |
| $D^\circ$ | polar variable $D$ |
| $\partial$ | *partial derivative* or *matrix of distance-square squared* or as in $\partial\mathcal{K}$ ; *boundary* of set $\mathcal{K}$ |
| $\partial y$ | partial differential of $y$ |
| $\sqrt{d_{ij}}$ | (absolute) distance scalar |
| $d_{ij}$ | distance-square scalar, EDM entry |
| $\mathbf{V}$ | geometric centering operator, $\mathbf{V}(D) = -VDV\frac{1}{2}$ |
| $\mathbf{V}_\mathcal{N}$ | $\mathbf{V}_\mathcal{N}(D) = -V_\mathcal{N}^T D V_\mathcal{N}$ |
| $V$ | $N\times N$ symmetric elementary, auxiliary, projector, geometric centering matrix, $\mathcal{R}(V) = \mathcal{N}(\mathbf{1}^T)$, $\mathcal{N}(V) = \mathcal{R}(\mathbf{1})$, $V^2 = V$ (§B.4.1) |
| $V_\mathcal{N}$ | $N\times N-1$ Schoenberg auxiliary matrix, $\mathcal{R}(V_\mathcal{N}) = \mathcal{N}(\mathbf{1}^T)$, $\mathcal{N}(V_\mathcal{N}^T) = \mathcal{R}(\mathbf{1})$ (§B.4.2) |
| $V_\mathcal{X}$ | $V_\mathcal{X} V_\mathcal{X}^T \equiv V^T X^T X V$ (996) |

| | |
|---|---|
| $X$ | point list (having cardinality $N$) arranged columnar in $\mathbb{R}^{n\times N}$, or set of generators, or extreme directions, or matrix variable |
| $G$ | Gram matrix $X^TX$ |
| $r$ | affine dimension |
| $n$ | Euclidean dimension of list $X$, or integer |
| $N$ | cardinality of list $X$, or integer |
| dom | function domain |
| *on* | *function* $f(x)$ *on* $\mathcal{A}$ means $\mathcal{A}$ is $\operatorname{dom} f$, or *projection of* $x$ *on* $\mathcal{A}$ means $\mathcal{A}$ is Euclidean body on which projection of $x$ is made |
| *onto* | *function* $f(x)$ *maps onto* $\mathcal{M}$ means $f$ over its domain is a surjection with respect to $\mathcal{M}$ |
| epi | function epigraph |
| span | as in $\operatorname{span} A = \mathcal{R}(A) = \{Ax \mid x\in\mathbb{R}^n\}$ when $A$ is a matrix |
| $\mathcal{R}(A)$ | the subspace: *range of* $A$, or $\operatorname{span}\operatorname{basis}\mathcal{R}(A)$; $\mathcal{R}(A)\perp\mathcal{N}(A^T)$ |
| $\operatorname{basis}\mathcal{R}(A)$ | *columnar basis for range of* $A$, or *a minimal set constituting generators for the vertex-description of* $\mathcal{R}(A)$, or *a linearly independent set of vectors spanning* $\mathcal{R}(A)$ |
| $\mathcal{N}(A)$ | the subspace: *nullspace of* $A$; $\mathcal{N}(A)\perp\mathcal{R}(A^T)$ |
| $\times$ | Cartesian product |
| $\left[\begin{array}{c}\mathbb{R}^m\\ \mathbb{R}^n\end{array}\right]$ | $\mathbb{R}^m\times\mathbb{R}^n = \mathbb{R}^{m+n}$ |
| $\mathbb{R}^{m\times n}$ | Euclidean vector space of $m$ by $n$ dimensional real matrices |
| $\mathbb{R}^n$ | Euclidean $n$-dimensional real vector space (nonnegative integer $n$) |
| $\mathbb{C}^n$ 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 comprising all (real) symmetric $n\times n$ matrices, the *symmetric matrix subspace* |
| $\mathbb{S}^{n\perp}$ | orthogonal complement of $\mathbb{S}^n$ in $\mathbb{R}^{n\times n}$, the antisymmetric matrices |
| $\mathbb{S}^n_+$ | convex cone comprising all (real) symmetric positive semidefinite $n\times n$ matrices, the *positive semidefinite cone* |
| $\operatorname{int}\mathbb{S}^n_+$ | interior of convex cone comprising all (real) symmetric positive semidefinite $n\times n$ matrices; *id est*, positive definite matrices |
| $\mathbb{S}^n_+(\rho)$ | convex set of all positive semidefinite $n\times n$ matrices whose rank equals or exceeds $\rho$ |
| $\mathbb{EDM}^N$ | cone of $N\times N$ Euclidean distance matrices in the symmetric hollow subspace |
| $\sqrt{\mathbb{EDM}^N}$ | nonconvex cone of $N\times N$ Euclidean absolute distance matrices in the symmetric hollow subspace |
| PSD | positive semidefinite |
| SDP | semidefinite program |
| SVD | singular value decomposition |
| EDM | Euclidean distance matrix |
| $\mathbb{S}^n_1$ | subspace comprising all symmetric $n\times n$ matrices having all zeros in first row and column (1770) |
| $\mathbb{S}^n_h$ | subspace comprising all symmetric hollow $n\times n$ matrices $($**0** main diagonal$)$, the *symmetric hollow subspace* (56) |
| $\mathbb{S}^{n\perp}_h$ | orthogonal complement of $\mathbb{S}^n_h$ in $\mathbb{S}^n$ (57), the set of all diagonal matrices |
| $\mathbb{S}^n_c$ | subspace comprising all geometrically centered symmetric $n\times n$ matrices; *geometric center subspace* $\mathbb{S}^N_c \triangleq \{Y\in\mathbb{S}^N \mid Y\mathbf{1}=\mathbf{0}\}$ (1766) |
| $\mathbb{S}^{n\perp}_c$ | orthogonal complement of $\mathbb{S}^n_c$ in $\mathbb{S}^n$ (1768) |

| | |
|---|---|
| $\mathbb{R}_c^{m\times n}$ | subspace comprising all geometrically centered $m\times n$ matrices |
| $X^\perp$ | basis $\mathcal{N}(X^T)$ |
| $x^\perp$ | $\mathcal{N}(x^T)\,,\ \ \{y \mid x^T y = 0\}$ |
| $\mathcal{R}(P)^\perp$ | $\mathcal{N}(P^T)$ |
| $\mathcal{R}^\perp$ | orthogonal complement of $\mathcal{R}\subseteq\mathbb{R}^n$; $\mathcal{R}^\perp \triangleq \{y\in\mathbb{R}^n \mid \langle x,y\rangle = 0 \ \ \forall\, x\in\mathcal{R}\}$ |
| $\mathcal{K}^\perp$ | normal cone |
| $\mathcal{K}$ | *cone* |
| $\mathcal{K}^*$ | *dual cone* |
| $\mathcal{K}^\circ$ | *polar cone*; $\mathcal{K}^* = -\mathcal{K}^\circ$ |
| $\mathcal{K}_{\mathcal{M}+}$ | monotone nonnegative cone |
| $\mathcal{K}_{\mathcal{M}}$ | 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 (86) to the hyperplane partially bounding it |
| $\mathcal{H}_+$ | halfspace described using an inward-normal (87) to the hyperplane partially bounding it |
| $\partial\mathcal{H}$ | hyperplane; *id est*, partial boundary of halfspace |
| $\underline{\partial\mathcal{H}}$ | supporting hyperplane |
| $\underline{\partial\mathcal{H}}_-$ | a supporting hyperplane having outward-normal with respect to set it supports |
| $\underline{\partial\mathcal{H}}_+$ | a supporting hyperplane having inward-normal with respect to set it supports |

| | |
|---|---|
| $\underline{d}$ | vector of distance-square |
| $\underline{d_{ij}}$ | lower bound on distance-square $d_{ij}$ |
| $\overline{d_{ij}}$ | upper bound on distance-square $d_{ij}$ |
| $\overline{\mathrm{AB}}$ | closed line segment between points A and B |
| $AB$ | matrix multiplication of $A$ and $B$ |
| $\overline{\mathcal{C}}$ | *closure of set* $\mathcal{C}$ |
| *decomposition* | *orthonormal* (1680) page 592, *biorthogonal* (1657) page 585 |
| *expansion* | *orthogonal* (1690) page 594, *biorthogonal* (344) page 163 |
| *vector* | *column vector* in $\mathbb{R}^n$ |
| *cubix* | member of $\mathbb{R}^{M\times N\times L}$ |
| *quartix* | member of $\mathbb{R}^{M\times N\times L\times K}$ |
| *feasible set* | most simply, *the set of all variable values satisfying all constraints of an optimization problem* |
| *solution set* | most simply, *the set of all optimal solutions to an optimization problem*; a subset of the feasible set and not necessarily a single point |
| *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 $\delta(A)$ means *taken from left to right and top to bottom* |
| *tight* | with reference to a bound means *a bound that can be met*,<br>with reference to an inequality means *equality is achievable* |
| $g'$ | first derivative of possibly multidimensional function with respect to real argument |
| $g''$ | second derivative with respect to real argument |
| $\overset{\rightarrow Y}{dg}$ | first directional derivative of possibly multidimensional function $g$ in direction $Y\in\mathbb{R}^{K\times L}$ (maintains dimensions of $g$) |

$\overset{\rightarrow Y}{dg^2}$ second directional derivative of $g$ in direction $Y$

$\nabla$ *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* $y$, $\nabla^2$ is *second-order gradient*

$\Delta$ distance scalar, or infinitesimal difference operator, or diagonal matrix

$\triangle_{ijk}$ triangle made by vertices $i$, $j$, and $k$

I Roman numeral

$I$ identity matrix

$\mathcal{I}$ *index set*, a discrete set of indices

$\emptyset$ *empty set*, an implicit member of every set

0 real zero

$\mathbf{0}$ *origin* or vector or matrix of zeros

$O$ *sort-index matrix*, or *order of magnitude* of information required, or *computational intensity*: $O(N)$ is first order, $O(N^2)$ is second, and so on

1 real one

$\mathbf{1}$ vector of ones

$e_i$ vector whose $i^{\text{th}}$ entry is 1 (otherwise 0), or $i^{\text{th}}$ *member of the standard basis for* $\mathbb{R}^n$

max *maximum* [148, §0.1.1] or *largest element of a totally ordered set*

$\underset{x}{\text{maximize}}$ *find the maximum of a function with respect to variable* $x$

arg *argument* of operator or function, or *variable of optimization*

$\sup \mathcal{X}$ *supremum* of totally ordered set $\mathcal{X}$, *least upper bound*, may or may not belong to set [148, §0.1.1]

| | |
|---|---|
| $\arg\sup f(x)$ | *argument* $x$ *at supremum of function* $f$ ; not necessarily unique or a member of function domain |
| subject to | specifies constraints to an optimization problem |
| min | *minimum* [148, §0.1.1] or *smallest element of a totally ordered set* |
| $\underset{x}{\text{minimize}}$ | *find the function minimum with respect to variable* $x$ |
| $\inf \mathcal{X}$ | *infimum* of totally ordered set $\mathcal{X}$ , *greatest lower bound*, may or may not belong to set [148, §0.1.1] |
| $\arg\inf f(x)$ | *argument* $x$ *at infimum of function* $f$ ; not necessarily unique or a member of function domain |
| iff | *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, an esoteric practice worthy of abolition because of ambiguity thus conferred |
| rel | relative |
| int | interior |
| lim | limit |
| sgn | signum function or *sign* |
| round | round to nearest integer |
| mod | modulus function |
| tr | matrix trace |
| rank | as in $\operatorname{rank} A$ , *rank of matrix* $A$ ; $\dim \mathcal{R}(A)$ |
| dim | dimension, $\dim \mathbb{R}^n = n$ , $\dim(x \in \mathbb{R}^n) = n$ , $\dim \mathcal{R}(x \in \mathbb{R}^n) = 1$, $\dim \mathcal{R}(A \in \mathbb{R}^{m\times n}) = \operatorname{rank}(A)$ |
| aff | affine hull |
| dim aff | affine dimension |

card — cardinality, $\operatorname{card} x \triangleq \|x\|_0$

conv — convex hull

cenv — convex envelope

cone — conic hull

content — 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

vec — vectorization of $m \times n$ matrix, Euclidean dimension $mn$ (30)

svec — vectorization of symmetric $n \times n$ matrix, Euclidean dimension $n(n+1)/2$ (47)

dvec — vectorization of symmetric hollow $n \times n$ matrix, Euclidean dimension $n(n-1)/2$ (63)

$\measuredangle(x\,,y)$ — *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 $\mathcal{K}$*, or *comparison to the origin with respect to some implicit pointed closed convex cone*, or (when $\mathcal{K} = \mathbb{S}^n_+$) *matrix $A$ belongs to the positive semidefinite cone of symmetric matrices* (§2.9.0.1), or (when $\mathcal{K} = \mathbb{R}^n_+$) *vector $A$ belongs to the nonnegative orthant* (each vector entry is nonnegative, §2.3.1.1)

$\succ$ — strict generalized inequality

$\not\succ$ — 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 $\alpha \in \mathbb{R}$, $\alpha \geq 0$

| | |
|---|---|
| $>$ | *greater than* |
| *positive* | for $\alpha \in \mathbb{R}$ , $\alpha > 0$ |
| $\lfloor\ \rfloor$ | floor function, $\lfloor x \rfloor$ is greatest integer not exceeding $x$ |
| $\|\ \|$ | entrywise absolute value of scalars, vectors, and matrices |
| det | matrix determinant |
| $\\|x\\|$ | vector 2-norm or *Euclidean norm* $\\|x\\|_2$ |
| $\\|x\\|_\ell$ | $= \sqrt[\ell]{\sum_{j=1}^{n} \|x_j\|^\ell}$ vector $\ell$-*norm* |
| $\\|x\\|_\infty$ | $= \max\{\|x_j\|\ \ \forall\, j\}$ *infinity-norm* |
| $\\|x\\|_2^2$ | $= x^T x = \langle x , x \rangle$ |
| $\\|x\\|_1$ | $= \mathbf{1}^T \|x\|$ 1-norm, *dual infinity-norm* |
| $\\|x\\|_0$ | 0-norm, *cardinality of vector* $x$ , $\operatorname{card} x \equiv \\|x\\|_0$ , $0^0 \triangleq 0$ |
| $\\|X\\|_2$ | $= \sup\limits_{\\|a\\|=1} \\|Xa\\|_2 = \sigma_1 = \sqrt{\lambda(X^T X)_1}$ matrix 2-norm (*spectral norm*), *largest singular value*, $\\|\delta(x)\\|_2 = \\|x\\|_\infty$ |
| $\\|X\\|$ | $= \\|X\\|_{\mathrm{F}}$ *Frobenius' matrix norm* |

# Bibliography


[1] Suliman Al-Homidan and Henry Wolkowicz. Approximate and exact completion problems for Euclidean distance matrices using semidefinite programming, April 2004.
orion.math.uwaterloo.ca/∼hwolkowi/henry/reports/edmapr04.ps

[2] Abdo Y. Alfakih. On the uniqueness of Euclidean distance matrix completions. *Linear Algebra and its Applications*, 370:1–14, 2003.

[3] Abdo Y. Alfakih. On the uniqueness of Euclidean distance matrix completions: the case of points in general position. *Linear Algebra and its Applications*, 397:265–277, 2005.

[4] Abdo Y. Alfakih, Amir Khandani, and Henry Wolkowicz. Solving Euclidean distance matrix completion problems via semidefinite programming. *Computational Optimization and Applications*, 12(1):13–30, January 1999.
http://citeseer.ist.psu.edu/alfakih97solving.html

[5] Abdo Y. Alfakih and Henry Wolkowicz. On the embeddability of weighted graphs in Euclidean spaces. Research Report CORR 98-12, Department of Combinatorics and Optimization, University of Waterloo, May 1998.
http://citeseer.ist.psu.edu/alfakih98embeddability.html
Erratum: p.334 herein.

[6] Abdo Y. Alfakih and Henry Wolkowicz. Matrix completion problems. In Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 18. Kluwer, 2000.

[7] Abdo Y. Alfakih and Henry Wolkowicz. Two theorems on Euclidean distance matrices and Gale transform. *Linear Algebra and its Applications*, 340:149–154, 2002.

[8] Farid Alizadeh. *Combinatorial Optimization with Interior Point Methods and Semi-Definite Matrices*. PhD thesis, University of Minnesota, Minneapolis, Computer Science Department, October 1991.

[9] Farid Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, February 1995.

[10] Kurt Anstreicher and Henry Wolkowicz. On Lagrangian relaxation of quadratic matrix constraints. *SIAM Journal on Matrix Analysis and Applications*, 22(1):41–55, 2000.

[11] Howard Anton. *Elementary Linear Algebra*. Wiley, second edition, 1977.

[12] James Aspnes, David Goldenberg, and Yang Richard Yang. On the computational complexity of sensor network localization. In *Proceedings of the First International Workshop on Algorithmic Aspects of Wireless Sensor Networks: ALGOSENSORS 2004*, volume 3121 of *Lecture Notes in Computer Science*, pages 32–44, Turku, Finland, July 2004. Springer-Verlag.
cs-www.cs.yale.edu/homes/aspnes/localization-abstract.html

[13] D. Avis and K. Fukuda. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete and Computational Geometry*, 8:295–313, 1992.

[14] Christine Bachoc and Frank Vallentin. New upper bounds for kissing numbers from semidefinite programming. *ArXiv.org*, October 2006.
http://arxiv.org/abs/math/0608426

[15] Mihály Bakonyi and Charles R. Johnson. The Euclidean distance matrix completion problem. *SIAM Journal on Matrix Analysis and Applications*, 16(2):646–654, April 1995.

[16] Keith Ball. An elementary introduction to modern convex geometry. In Silvio Levy, editor, *Flavors of Geometry*, volume 31, chapter 1, pages 1–58. MSRI Publications, 1997.
www.msri.org/publications/books/Book31/files/ball.pdf

[17] George Phillip Barker. Theory of cones. *Linear Algebra and its Applications*, 39:263–291, 1981.

[18] George Phillip Barker and David Carlson. Cones of diagonally dominant matrices. *Pacific Journal of Mathematics*, 57(1):15–32, 1975.

[19] George Phillip Barker and James Foran. Self-dual cones in Euclidean spaces. *Linear Algebra and its Applications*, 13:147–155, 1976.

[20] Alexander Barvinok. *A Course in Convexity*. American Mathematical Society, 2002.

[21] Alexander I. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discrete & Computational Geometry*, 13(2):189–202, 1995.

[22] Alexander I. Barvinok. A remark on the rank of positive semidefinite matrices subject to affine constraints. *Discrete & Computational Geometry*, 25(1):23–31, 2001.
http://citeseer.ist.psu.edu/304448.html

[23] Heinz H. Bauschke and Jonathan M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, September 1996.

[24] Steven R. Bell. *The Cauchy Transform, Potential Theory, and Conformal Mapping*. CRC Press, 1992.

[25] Jean Bellissard and Bruno Iochum. Homogeneous and facially homogeneous self-dual cones. *Linear Algebra and its Applications*, 19:1–16, 1978.

[26] Adi Ben-Israel. Linear equations and inequalities on finite dimensional, real or complex, vector spaces: A unified theory. *Journal of Mathematical Analysis and Applications*, 27:367–389, 1969.

[27] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001.

[28] Abraham Berman. *Cones, Matrices, and Mathematical Programming*, volume 79 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1973.

[29] Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, second edition, 1999.

[30] Dimitri P. Bertsekas, Angelia Nedić, and Asuman E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[31] Rajendra Bhatia. *Matrix Analysis*. Springer-Verlag, 1997.

[32] Pratik Biswas, Tzu-Chen Liang, Kim-Chuan Toh, Yinyu Ye, and Ta-Chung Wang. Semidefinite programming approaches for sensor network localization with noisy distance measurements. *IEEE Transactions on Automation Science and Engineering*, 3(4):360–371, October 2006.

[33] Pratik Biswas, Tzu-Chen Liang, Ta-Chung Wang, and Yinyu Ye. Semidefinite programming based algorithms for sensor network localization, 2005.
http://www.stanford.edu/∼yyye/combined_rev3.pdf

[34] Pratik Biswas, Kim-Chuan Toh, and Yinyu Ye. A distributed SDP approach for large-scale noisy anchor-free graph realization with applications to molecular conformation, February 2007.
www.convexoptimization.com/TOOLS/distmolecule.pdf

[35] Pratik Biswas and Yinyu Ye. Semidefinite programming for ad hoc wireless sensor network localization, September 2003.
http://www.stanford.edu/∼yyye/adhocn2.pdf

[36] Leonard M. Blumenthal. On the four-point property. *Bulletin of the American Mathematical Society*, 39:423–426, 1933.

[37] Leonard M. Blumenthal. *Theory and Applications of Distance Geometry*. Oxford University Press, 1953.

[38] A. W. Bojanczyk and A. Lutoborski. The Procrustes problem for orthogonal Stiefel matrices. *SIAM Journal on Scientific Computing*, 21(4):1291–1304, February 1998. Date is electronic publication: http://citeseer.ist.psu.edu/500185.html

[39] Ingwer Borg and Patrick Groenen. *Modern Multidimensional Scaling*. Springer-Verlag, 1997.

[40] Jonathan M. Borwein and Heinz Bauschke.
Projection algorithms and monotone operators, 1998.
http://oldweb.cecm.sfu.ca/personal/jborwein/projections4.pdf

[41] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. Springer-Verlag, 2000.

[42] Richard Bouldin. The pseudo-inverse of a product. *SIAM Journal on Applied Mathematics*, 24(4):489–495, June 1973.
http://www.convexoptimization.com/TOOLS/Pseudoinverse.pdf

[43] Stephen Boyd and Jon Dattorro. Alternating projections, 2003.
http://www.stanford.edu/class/ee392o/alt_proj.pdf

[44] Stephen Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.

[45] Stephen Boyd, Seung-Jean Kim, Lieven Vandenberghe, and Arash Hassibi. A tutorial on geometric programming. *Optimization and Engineering*, 8(1):1389–4420, March 2007.
http://www.stanford.edu/∼boyd/reports/gp_tutorial.pdf

[46] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
http://www.stanford.edu/∼boyd/cvxbook

[47] James P. Boyle and Richard L. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In R. Dykstra, T. Robertson, and F. T. Wright, editors, *Advances in Order Restricted Statistical Inference*, pages 28–47. Springer-Verlag, 1986.

[48] Lev M. Brègman. The method of successive projection for finding a common point of convex sets. *Soviet Mathematics*, 162(3):487–490, 1965. AMS translation of Doklady Akademii Nauk SSSR, 6:688-692.

[49] Lev M. Brègman, Yair Censor, Simeon Reich, and Yael Zepkowitz-Malachi. Finding the projection of a point onto the intersection of convex sets via projections onto halfspaces, 2003.
http://www.optimization-online.org/DB_FILE/2003/06/669.pdf

[50] Mike Brookes. Matrix reference manual: Matrix calculus, 2002.
http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html

[51] Michael Carter, Holly Jin, Michael Saunders, and Yinyu Ye. SPASELOC: An adaptive subproblem algorithm for scalable wireless sensor network localization. *SIAM Journal on Optimization*, 17(4):1102–1128, 2006.
www.convexoptimization.com/TOOLS/JinSIAM.pdf
Erratum: p.258 herein.

[52] Lawrence Cayton and Sanjoy Dasgupta. Robust Euclidean embedding. In *Proceedings of the* 23$^{\text{rd}}$ *International Conference on Machine Learning - ICML 2006*, Pittsburgh, Pennsylvania USA, 2006.
icml2006.org/icml_documents/camera-ready/022_Robust_Euclidean_Emb.pdf

[53] Yves Chabrillac and Jean-Pierre Crouzeix. Definiteness and semidefiniteness of quadratic forms revisited. *Linear Algebra and its Applications*, 63:283–292, 1984.

[54] Manmohan K. Chandraker, Sameer Agarwal, Fredrik Kahl, David Nistér, and David J. Kriegman. Autocalibration via rank-constrained estimation of the absolute quadric. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2007.
http://vision.ucsd.edu/∼manu/cvpr07_quadric.pdf

[55] Chi-Tsong Chen. *Linear System Theory and Design*. Oxford University Press, 1999.

[56] Ward Cheney and Allen A. Goldstein. Proximity maps for convex sets. *Proceedings of the American Mathematical Society*, 10:448–450, 1959.

[57] Steven Chu. Autobiography from *Les Prix Nobel*, 1997.
nobelprize.org/physics/laureates/1997/chu-autobio.html

[58] John B. Conway. *A Course in Functional Analysis*. Springer-Verlag, second edition, 1990.

[59] Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. *The Linear Complementarity Problem*. Academic Press, 1992.

[60] G. M. Crippen and T. F. Havel. *Distance Geometry and Molecular Conformation*. Wiley, 1988.

[61] Frank Critchley. *Multidimensional scaling: a critical examination and some new proposals*. PhD thesis, University of Oxford, Nuffield College, 1980.

[62] Frank Critchley. On certain linear mappings between inner-product and squared-distance matrices. *Linear Algebra and its Applications*, 105:91–107, 1988.

[63] Joachim Dahl, Bernard H. Fleury, and Lieven Vandenberghe. Approximate maximum-likelihood estimation using semidefinite programming. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume VI, pages 721–724, April 2003.
www.ee.ucla.edu/faculty/papers/vandenberghe_ICASSP03_apr03.pdf

[64] George B. Dantzig. *Linear Programming and Extensions*.
Princeton University Press, 1998.

[65] Alexandre d'Aspremont, Laurent El Ghaoui, Michael I. Jordan, and Gert R. G. Lanckriet. A direct formulation for sparse PCA using semidefinite programming. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 41–48. MIT Press, Cambridge, MA, 2005.
http://books.nips.cc/papers/files/nips17/NIPS2004_0645.pdf

[66] Jon Dattorro. Constrained least squares fit of a filter bank to an arbitrary magnitude frequency response, 1991.
http://www.stanford.edu/∼dattorro/Hearing.htm

[67] Joel Dawson, Stephen Boyd, Mar Hershenson, and Thomas Lee. Optimal allocation of local feedback in multistage amplifiers via geometric programming. *IEEE Circuits and Systems I: Fundamental Theory and Applications*, 2001.

[68] Jan de Leeuw. Fitting distances by least squares. UCLA Statistics Series Technical Report No. 130, Interdivisional Program in Statistics, UCLA, Los Angeles, CA, 1993.
http://citeseer.ist.psu.edu/deleeuw93fitting.html

[69] Jan de Leeuw. Multidimensional scaling. In *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 2001.
http://preprints.stat.ucla.edu/274/274.pdf

[70] Jan de Leeuw. Unidimensional scaling. In Brian S. Everitt and David C. Howell, editors, *Encyclopedia of Statistics in Behavioral Science*, volume 4, pages 2095–2097. Wiley, 2005.
http://repositories.cdlib.org/uclastat/papers/2004032701

[71] Jan de Leeuw and Willem Heiser. Theory of multidimensional scaling. In P. R. Krishnaiah and L. N. Kanal, editors, *Handbook of Statistics*, volume 2, chapter 13, pages 285–316. North-Holland Publishing, Amsterdam, 1982.

[72] Erik D. Demaine, Francisco Gomez-Martin, Henk Meijer, David Rappaport, Perouz Taslakian, Godfried T. Toussaint, Terry Winograd, and David R. Wood. The distance geometry of music, 2007.
http://arxiv.org/abs/0705.4085

[73] Frank Deutsch. *Best Approximation in Inner Product Spaces*. Springer-Verlag, 2001.

[74] Frank Deutsch and Hein Hundal. The rate of convergence of Dykstra's cyclic projections algorithm: The polyhedral case. *Numerical Functional Analysis and Optimization*, 15:537–565, 1994.

[75] Frank Deutsch, John H. McCabe, and George M. Phillips. Some algorithms for computing best approximations from convex cones. *SIAM Journal on Numerical Analysis*, 12(3):390–403, June 1975.

[76] Frank R. Deutsch and Peter H. Maserick. Applications of the Hahn-Banach theorem in approximation theory. *SIAM Review*, 9(3):516–530, July 1967.

[77] Michel Marie Deza and Monique Laurent. *Geometry of Cuts and Metrics*. Springer-Verlag, 1997.

[78] Carolyn Pillers Dobler. A matrix approach to finding a set of generators and finding the polar (dual) of a class of polyhedral cones. *SIAM Journal on Matrix Analysis and Applications*, 15(3):796–803, July 1994. Erratum: p.182 herein.

[79] Elizabeth D. Dolan, Robert Fourer, Jorge J. Moré, and Todd S. Munson. Optimization on the NEOS server. *SIAM News*, 35(6):4,8,9, August 2002.

[80] Bruce Randall Donald. 3-D structure in chemistry and molecular biology, 1998.
http://www.cs.dartmouth.edu/∼brd/Teaching/Bio

[81] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.
www-stat.stanford.edu/∼donoho/Reports/2004/CompressedSensing091604.pdf

[82] David L. Donoho, Michael Elad, and Vladimir Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise, February 2004.
www-stat.stanford.edu/∼donoho/Reports/2004/StableSparse-Donoho-etal.pdf

[83] Miguel Nuno Ferreira Fialho dos Anjos. *New Convex Relaxations for the Maximum Cut and VLSI Layout Problems*. PhD thesis, University of Waterloo, Ontario, Canada, Department of Combinatorics and Optimization, 2001.
cheetah.vlsi.uwaterloo.ca/∼anjos/MFAnjosPhDThesis.pdf

[84] Richard L. Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.

[85] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, September 1936.
http://www.stanford.edu/~dattorro/eckart&young.1936.pdf

[86] Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[87] John J. Edgell. Graphics calculator applications on 4-D constructs, 1996.
http://archives.math.utk.edu/ICTCM/EP-9/C47/pdf/paper.pdf

[88] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems*. SIAM, 1999.

[89] Julius Farkas. Theorie der einfachen Ungleichungen. *Journal für die reine und angewandte Mathematik*, 124:1–27, 1902.
dz-srv1.sub.uni-goettingen.de/sub/digbib/loader?ht=VIEW&did=D261364

[90] Maryam Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, Department of Electrical Engineering, March 2002.
http://www.cds.caltech.edu/~maryam/thesis-final.pdf

[91] Maryam Fazel, Haitham Hindi, and Stephen P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, volume 6, pages 4734–4739. American Automatic Control Council (AACC), June 2001.
http://www.cds.caltech.edu/~maryam/nucnorm.html

[92] Maryam Fazel, Haitham Hindi, and Stephen P. Boyd. Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In *Proceedings of the American Control Conference*. American Automatic Control Council (AACC), June 2003.
http://www.cds.caltech.edu/~maryam/acc03_final.pdf

[93] Maryam Fazel, Haitham Hindi, and Stephen P. Boyd. Rank minimization and applications in system theory. In *Proceedings of the American Control Conference*. American Automatic Control Council

(AACC), June 2004.
http://www.cds.caltech.edu/∼maryam/acc04-tutorial.pdf

[94] J. T. Feddema, R. H. Byrne, J. J. Harrington, D. M. Kilman, C. L. Lewis, R. D. Robinett, B. P. Van Leeuwen, and J. G. Young. Advanced mobile networking, sensing, and controls. Sandia Report SAND2005-1661, Sandia National Laboratories, Albuquerque, New Mexico, March 2005.
www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/2005/051661.pdf

[95] César Fernández, Thomas Szyperski, Thierry Bruyère, Paul Ramage, Egon Mösinger, and Kurt Wüthrich. NMR solution structure of the pathogenesis-related protein P14a. *Journal of Molecular Biology*, 266:576–593, 1997.

[96] Richard Phillips Feynman, Robert B. Leighton, and Matthew L. Sands. *The Feynman Lectures on Physics: Commemorative Issue*, volume I. Addison-Wesley, 1989.

[97] Paul Finsler. Über das Vorkommen definiter und semidefiniter Formen in Scharen quadratischer Formen. *Commentarii Mathematici Helvetici*, 9:188–192, 1937.

[98] Anders Forsgren, Philip E. Gill, and Margaret H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 44(4):525–597, 2002.

[99] Norbert Gaffke and Rudolf Mathar. A cyclic projection algorithm via duality. *Metrika*, 36:29–54, 1989.

[100] Jérôme Galtier. Semi-definite programming as a simple extension to linear programming: convex optimization with queueing, equity and other telecom functionals. In *3ème Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel)*. INRIA, France, 2001.
www-sop.inria.fr/mascotte/Jerome.Galtier/misc/Galtier01b.pdf

[101] Laurent El Ghaoui. EE 227A: Convex Optimization and Applications, Lecture 11 − October 3. University of California, Berkeley, Fall 2006. Scribe: Nikhil Shetty.
www.convexoptimization.com/TOOLS/Ghaoui.pdf

[102] Laurent El Ghaoui and Silviu-Iulian Niculescu, editors. *Advances in Linear Matrix Inequality Methods in Control*. SIAM, 2000.

[103] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Numerical Linear Algebra and Optimization*, volume 1. Addison-Wesley, 1991.

[104] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization*. Academic Press, 1999.

[105] James Gleik. *Isaac Newton*. Pantheon Books, 2003.

[106] W. Glunt, Tom L. Hayden, S. Hong, and J. Wells. An alternating projection algorithm for computing the nearest Euclidean distance matrix. *SIAM Journal on Matrix Analysis and Applications*, 11(4):589–600, 1990.

[107] K. Goebel and W. A. Kirk. *Topics in Metric Fixed Point Theory*. Cambridge University Press, 1990.

[108] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the Association for Computing Machinery*, 42(6):1115–1145, November 1995.
http://www-math.mit.edu/∼goemans/maxcut-jacm.pdf

[109] D. Goldfarb and K. Scheinberg. Interior point trajectories in semidefinite programming. *SIAM Journal on Optimization*, 8(4):871–886, 1998.

[110] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins, third edition, 1996.

[111] P. Gordan. Ueber die auflösung linearer gleichungen mit reellen coefficienten. *Mathematische Annalen*, 6:23–28, 1873.

[112] John Clifford Gower. Euclidean distance geometry. *The Mathematical Scientist*, 7:1–14, 1982.

[113] John Clifford Gower. Properties of Euclidean and non-Euclidean distance matrices. *Linear Algebra and its Applications*, 67:81–97, 1985.

[114] John Clifford Gower and Garmt B. Dijksterhuis. *Procrustes Problems*. Oxford University Press, 2004.

[115] John Clifford Gower and David J. Hand. *Biplots*. Chapman & Hall, 1996.

[116] Alexander Graham. *Kronecker Products and Matrix Calculus: with Applications*. Wiley, 1981.

[117] Michael Grant, Stephen Boyd, and Yinyu Ye.
`cvx`: Matlab software for disciplined convex programming, 2007.
http://www.stanford.edu/∼boyd/cvx

[118] Robert M. Gray. Toeplitz and circulant matrices: A review, 2002.
http://www-ee.stanford.edu/∼gray/toeplitz.pdf

[119] T. N. E. Greville. Note on the generalized inverse of a matrix product. *SIAM Review*, 8:518–521, 1966.

[120] Rémi Gribonval and Morten Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure, October 2003.
http://www.math.aau.dk/research/reports/R-2003-16.pdf

[121] Rémi Gribonval and Morten Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):1320–1325, December 2003.
http://www.math.aau.dk/∼mnielsen/papers.htm

[122] Karolos M. Grigoriadis and Eric B. Beran. Alternating projection algorithms for linear matrix inequalities problems with rank constraints. In Laurent El Ghaoui and Silviu-Iulian Niculescu, editors, *Advances in Linear Matrix Inequality Methods in Control*, chapter 13, pages 251–267. SIAM, 2000.

[123] Peter Gritzmann and Victor Klee. On the complexity of some basic problems in computational convexity: II. Volume and mixed volumes. Technical Report TR:94-31, DIMACS, Rutgers University, 1994.
ftp://dimacs.rutgers.edu/pub/dimacs/TechnicalReports/TechReports/1994/94-31.ps

[124] Peter Gritzmann and Victor Klee. On the complexity of some basic problems in computational convexity: II. Volume and mixed volumes. In T. Bisztriczky, P. McMullen, R. Schneider, and A. Ivić Weiss, editors, *Polytopes: Abstract, Convex and Computational*, pages 373–466. Kluwer Academic Publishers, 1994.

[125] L. G. Gubin, B. T. Polyak, and E. V. Raik. The method of projections for finding the common point of convex sets. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 7(6):1–24, 1967.

[126] Osman Güler and Yinyu Ye. Convergence behavior of interior-point algorithms. *Mathematical Programming*, 60(2):215–228, 1993.

[127] P. R. Halmos. Positive approximants of operators. *Indiana University Mathematics Journal*, 21:951–960, 1972.

[128] Shih-Ping Han. A successive projection method. *Mathematical Programming*, 40:1–14, 1988.

[129] Godfrey H. Hardy, John E. Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, second edition, 1952.

[130] Arash Hassibi and Mar Hershenson. Automated optimal design of switched-capacitor filters. *Design Automation and Test in Europe Conference*, 2001.

[131] Johan Håstad. Some optimal inapproximability results, 1999. http://citeseer.ist.psu.edu/280448.html

[132] Timothy F. Havel and Kurt Wüthrich. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *Journal of Molecular Biology*, 182:281–294, 1985.

[133] Tom L. Hayden and Jim Wells. Approximation by matrices positive semidefinite on a subspace. *Linear Algebra and its Applications*, 109:115–130, 1988.

[134] Tom L. Hayden, Jim Wells, Wei-Min Liu, and Pablo Tarazaga. The cone of distance matrices. *Linear Algebra and its Applications*, 144:153–169, 1991.

[135] Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. Springer-Verlag, 1994.

[136] Bruce Hendrickson. Conditions for unique graph realizations. *SIAM Journal on Computing*, 21(1):65–84, February 1992.

[137] T. Herrmann, Peter Güntert, and Kurt Wüthrich. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology*, 319(1):209–227, May 2002.

[138] Mar Hershenson. Design of pipeline analog-to-digital converters via geometric programming. *International Conference on Computer Aided Design - ICCAD*, 2002.

[139] Mar Hershenson. Efficient description of the design space of analog circuits. $40^{\text{th}}$ *Design Automation Conference*, 2003.

[140] Mar Hershenson, Stephen Boyd, and Thomas Lee. Optimal design of a CMOS OpAmp via geometric programming. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2001.

[141] Mar Hershenson, Dave Colleran, Arash Hassibi, and Navraj Nandra. Synthesizable full custom mixed-signal IP. *Electronics Design Automation Consortium - EDA*, 2002.

[142] Mar Hershenson, Sunderarajan S. Mohan, Stephen Boyd, and Thomas Lee. Optimization of inductor circuits via geometric programming, 1999.

[143] Mar Hershenson and Xiling Shen. A new analog design flow. *Cadence*, 2002.

[144] Nick Higham. Matrix Procrustes problems, 1995.
http://www.ma.man.ac.uk/∼higham/talks
Lecture notes.

[145] Richard D. Hill and Steven R. Waters. On the cone of positive semidefinite matrices. *Linear Algebra and its Applications*, 90:81–88, 1987.

[146] Jean-Baptiste Hiriart-Urruty. Ensembles de Tchebychev *vs.* ensembles convexes: l'état de la situation vu via l'analyse convexe non lisse. *Annales des Sciences Mathématiques du Québec*, 22(1):47–62, 1998.

[147] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*. Springer-Verlag, second edition, 1996.

[148] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of Convex Analysis*. Springer-Verlag, 2001.

[149] Alan J. Hoffman and Helmut W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20:37–40, 1953.

[150] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1987.

[151] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1994.

[152] Alston S. Householder. *The Theory of Matrices in Numerical Analysis*. Dover, 1975.

[153] Hong-Xuan Huang, Zhi-An Liang, and Panos M. Pardalos. Some properties for the Euclidean distance matrix and positive semidefinite matrix completion problems. *Journal of Global Optimization*, 25(1):3–21, 2003.

[154] 5W Infographic. Wireless 911. *Technology Review*, 107(5):78–79, June 2004. http://www.technologyreview.com

[155] Nathan Jacobson. *Lectures in Abstract Algebra, vol. II - Linear Algebra*. Van Nostrand, 1953.

[156] Viren Jain and Lawrence K. Saul. Exploratory analysis and visualization of speech and music by locally linear embedding. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 984–987, 2004. http://www.cis.upenn.edu/∼lsaul/papers/lle_icassp04.pdf

[157] Joakim Jaldén, Cristoff Martin, and Björn Ottersten. Semidefinite programming for detection in linear systems − Optimality conditions and space-time decoding. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume VI, April 2003.
http://www.s3.kth.se/signal/reports/03/IR-S3-SB-0309.pdf

[158] Florian Jarre. Convex analysis on symmetric matrices. In Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 2. Kluwer, 2000.

[159] Charles R. Johnson and Pablo Tarazaga. Connections between the real positive semidefinite and distance matrix completion problems. *Linear Algebra and its Applications*, 223/224:375–391, 1995.

[160] Charles R. Johnson and Pablo Tarazaga. Binary representation of normalized symmetric and correlation matrices. *Linear and Multilinear Algebra*, 52(5):359–366, 2004.

[161] George B. Thomas, Jr. *Calculus and Analytic Geometry*. Addison-Wesley, fourth edition, 1972.

[162] Thomas Kailath. *Linear Systems*. Prentice-Hall, 1980.

[163] Tosio Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, 1966.

[164] Paul J. Kelly and Norman E. Ladd. *Geometry*. Scott, Foresman and Company, 1965.

[165] Ron Kimmel. *Numerical Geometry of Images: Theory, Algorithms, and Applications*. Springer-Verlag, 2003.

[166] Erwin Kreyszig. *Introductory Functional Analysis with Applications*. Wiley, 1989.

[167] Jean B. Lasserre. A new Farkas lemma for positive semidefinite matrices. *IEEE Transactions on Automatic Control*, 40(6):1131–1133, June 1995.

[168] Jean B. Lasserre and Eduardo S. Zeron. A Laplace transform algorithm for the volume of a convex polytope. *Journal of the Association for Computing Machinery*, 48(6):1126–1140, November 2001.

[169] Jean B. Lasserre and Eduardo S. Zeron. A new algorithm for the volume of a convex polytope. *arXiv.org*, June 2001.
http://arxiv.org/abs/math/0106168

[170] Monique Laurent. A connection between positive semidefinite and Euclidean distance matrix completion problems. *Linear Algebra and its Applications*, 273:9–22, 1998.

[171] Monique Laurent. A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. In Panos M. Pardalos and Henry Wolkowicz, editors, *Topics in Semidefinite and Interior-Point Methods*, pages 51–76. American Mathematical Society, 1998.

[172] Monique Laurent. Matrix completion problems. In Christodoulos A. Floudas and Panos M. Pardalos, editors, *Encyclopedia of Optimization*, volume III (Interior - M), pages 221–229. Kluwer, 2001.
http://homepages.cwi.nl/∼monique/files/opt.ps

[173] Monique Laurent and Svatopluk Poljak. On a positive semidefinite relaxation of the cut polytope. *Linear Algebra and its Applications*, 223/224:439–461, 1995.

[174] Monique Laurent and Svatopluk Poljak. On the facial structure of the set of correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 17(3):530–547, July 1996.

[175] Monique Laurent and Franz Rendl. Semidefinite programming and integer programming. *Optimization Online*, 2002.
http://www.optimization-online.org/DB_HTML/2002/12/585.html

[176] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*. SIAM, 1995.

[177] Jung Rye Lee. The law of cosines in a tetrahedron. *Journal of the Korea Society of Mathematical Education Series B: The Pure and Applied Mathematics*, 4(1):1–6, 1997.

[178] Vladimir L. Levin. Quasi-convex functions and quasi-monotone operators. *Journal of Convex Analysis*, 2(1/2):167–172, 1995.

[179] Adrian S. Lewis. Eigenvalue-constrained faces. *Linear Algebra and its Applications*, 269:159–181, 1998.

[180] Anhua Lin. *Projection algorithms in nonlinear programming.* PhD thesis, Johns Hopkins University, 2003.

[181] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228, November 1998. Special Issue on Linear Algebra in Control, Signals and Image Processing. http://www.stanford.edu/∼boyd/socp.html

[182] David G. Luenberger. *Optimization by Vector Space Methods.* Wiley, 1969.

[183] David G. Luenberger. *Introduction to Dynamic Systems: Theory, Models, & Applications.* Wiley, 1979.

[184] David G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, second edition, 1989.

[185] Zhi-Quan Luo, Jos F. Sturm, and Shuzhong Zhang. Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. *SIAM Journal on Optimization*, 8(1):59–81, 1998.

[186] Zhi-Quan Luo and Wei Yu. An introduction to convex optimization for communications and signal processing. *IEEE Journal On Selected Areas In Communications*, 24(8):1426–1438, August 2006.

[187] K. V. Mardia. Some properties of classical multi-dimensional scaling. *Communications in Statistics: Theory and Methods*, A7(13):1233–1241, 1978.

[188] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis.* Academic Press, 1979.

[189] Jerrold E. Marsden and Michael J. Hoffman. *Elementary Classical Analysis.* Freeman, second edition, 1995.

[190] Rudolf Mathar. The best Euclidean fit to a given distance matrix in prescribed dimensions. *Linear Algebra and its Applications*, 67:1–6, 1985.

[191] Rudolf Mathar. *Multidimensionale Skalierung*. B. G. Teubner Stuttgart, 1997.

[192] Mehran Mesbahi and G. P. Papavassilopoulos. On the rank minimization problem over a positive semi-definite linear matrix inequality. *IEEE Transactions on Automatic Control*, 42(2):239–243, 1997.

[193] Sunderarajan S. Mohan, Mar Hershenson, Stephen Boyd, and Thomas Lee. Simple accurate expressions for planar spiral inductances. *IEEE Journal of Solid-State Circuits*, 1999.

[194] Sunderarajan S. Mohan, Mar Hershenson, Stephen Boyd, and Thomas Lee. Bandwidth extension in CMOS with optimized on-chip inductors. *IEEE Journal of Solid-State Circuits*, 2000.

[195] E. H. Moore. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, 26:394–395, 1920. Abstract.

[196] B. S. Mordukhovich. Maximum principle in the problem of time optimal response with nonsmooth constraints. *Journal of Applied Mathematics and Mechanics*, 40:960–969, 1976.

[197] Jean-Jacques Moreau. Décomposition orthogonale d'un espace Hilbertien selon deux cônes mutuellement polaires. *Comptes Rendus de l'Académie des Sciences, Paris*, 255:238–240, 1962.

[198] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.

[199] Neil Muller, Lourenço Magaia, and B. M. Herbst. Singular value decomposition, eigenfaces, and 3D reconstructions. *SIAM Review*, 46(3):518–545, September 2004.

[200] Katta G. Murty and Feng-Tien Yu. *Linear Complementarity, Linear and Nonlinear Programming*. Heldermann Verlag, Internet edition, 1988.
ioe.engin.umich.edu/people/fac/books/murty/linear_complementarity_webbook

[201] Oleg R. Musin. An extension of Delsarte's method. The kissing problem in three and four dimensions. *ArXiv.org*, December 2005. http://arxiv.org/abs/math.MG/0512649

[202] Navraj Nandra. Synthesizable analog IP. *IP Based Design Workshop*, 2002.

[203] Stephen G. Nash and Ariela Sofer. *Linear and Nonlinear Programming*. McGraw-Hill, 1996.

[204] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.

[205] L. Nirenberg. *Functional Analysis*. New York University, New York, 1961. Lectures given in 1960-1961, notes by Lesley Sibner.

[206] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer-Verlag, 1999.

[207] Royal Swedish Academy of Sciences. Nobel prize in chemistry, 2002. http://nobelprize.org/chemistry/laureates/2002/public.html

[208] C. S. Ogilvy. *Excursions in Geometry*. Dover, 1990. Citation: *Proceedings of the CUPM Geometry Conference*, Mathematical Association of America, No.16 (1967), p.21.

[209] Önder Filiz and Aylin Yener. Rank constrained temporal-spatial filters for CDMA systems with base station antenna arrays. In *Proceedings of the Johns Hopkins University Conference on Information Sciences and Systems*, March 2003. labs.ee.psu.edu/labs/wcan/Publications/filiz-yener_ciss03.pdf

[210] Robert Orsi, Uwe Helmke, and John B. Moore. A Newton-like method for solving rank constrained linear matrix inequalities, April 2006. users.rsise.anu.edu.au/∼robert/publications/OHM06_extended_version.pdf

[211] Brad Osgood. Notes on the Ahlfors mapping of a multiply connected domain, 2000. www-ee.stanford.edu/∼osgood/Ahlfors-Bergman-Szego.pdf

[212] M. L. Overton and R. S. Womersley. On the sum of the largest eigenvalues of a symmetric matrix. *SIAM Journal on Matrix Analysis and Applications*, 13:41–45, 1992.

[213] Pythagoras Papadimitriou. *Parallel Solution of SVD-Related Problems, With Applications*. PhD thesis, Department of Mathematics, University of Manchester, October 1993.

[214] Panos M. Pardalos and Henry Wolkowicz, editors. *Topics in Semidefinite and Interior-Point Methods*. American Mathematical Society, 1998.

[215] Gábor Pataki. Cone-LP's and semidefinite programs: Geometry and a simplex-type method. In William H. Cunningham, S. Thomas McCormick, and Maurice Queyranne, editors, *Integer Programming and Combinatorial Optimization, Proceedings of the 5th International IPCO Conference, Vancouver, British Columbia, Canada, June 3-5, 1996*, volume 1084 of *Lecture Notes in Computer Science*, pages 162–174. Springer-Verlag, 1996.

[216] Gábor Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of Operations Research*, 23(2):339–358, 1998.
http://citeseer.ist.psu.edu/pataki97rank.html
Erratum: p.543 herein.

[217] Gábor Pataki. The geometry of semidefinite programming. In Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe, editors, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, chapter 3. Kluwer, 2000.

[218] Teemu Pennanen and Jonathan Eckstein. Generalized Jacobians of vector-valued convex functions. Technical Report RRR 6-97, RUTCOR, Rutgers University, May 1997.
rutcor.rutgers.edu/pub/rrr/reports97/06.ps

[219] Roger Penrose. A generalized inverse for matrices. In *Proceedings of the Cambridge Philosophical Society*, volume 51, pages 406–413, 1955.

[220] Chris Perkins. A convergence analysis of Dykstra's algorithm for polyhedral sets. *SIAM Journal on Numerical Analysis*, 40(2):792–804, 2002.

[221] Florian Pfender and Günter M. Ziegler. Kissing numbers, sphere packings, and some unexpected proofs. *Notices of the American Mathematical Society*, 51(8):873–883, September 2004.
www.stanford.edu/∼dattorro/Pfender.pdf

[222] Benjamin Recht. *Convex Modeling with Priors*. PhD thesis, Massachusetts Institute of Technology, Media Arts and Sciences Department, 2006.
http://www.media.mit.edu/physics/publications/theses/06.06.Recht.pdf

[223] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *Optimization Online*, 2007.
http://www.optimization-online.org/DB_HTML/2007/06/1707.html

[224] Franz Rendl, Giovanni Rinaldi, and Angelika Wiegele. Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations, May 2007.
www.optimization-online.org/DB_HTML/2007/05/1659.html

[225] Alex Reznik. Problem 3.41, from Sergio Verdú. *Multiuser Detection*. Cambridge University Press, 1998.
www.ee.princeton.edu/∼verdu/mud/solutions/3/3.41.areznik.pdf , 2001.

[226] Sara Robinson. Hooked on meshing, researcher creates award-winning triangulation program. *SIAM News*, 36(9), November 2003.
http://www.siam.org/news/news.php?id=370

[227] R. Tyrrell Rockafellar. *Conjugate Duality and Optimization*. SIAM, 1974.

[228] R. Tyrrell Rockafellar. Lagrange multipliers in optimization. *SIAM-American Mathematical Society Proceedings*, 9:145–168, 1976.

[229] R. Tyrrell Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2):183–238, June 1993.

[230] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1997. First published in 1970.

[231] C. K. Rushforth. Signal restoration, functional analysis, and Fredholm integral equations of the first kind. In Henry Stark, editor, *Image Recovery: Theory and Application*, chapter 1, pages 1–27. Academic Press, 1987.

[232] Shankar Sastry. *Nonlinear Systems: Analysis, Stability, and Control*. Springer-Verlag, 1999.

[233] Uwe Schäfer. A linear complementarity problem with a P-matrix. *SIAM Review*, 46(2):189–201, June 2004.

[234] Isaac J. Schoenberg. Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert". *Annals of Mathematics*, 36(3):724–732, July 1935.
http://www.stanford.edu/∼dattorro/Schoenberg2.pdf

[235] Isaac J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44:522–536, 1938.
http://www.stanford.edu/∼dattorro/Schoenberg3.pdf

[236] Peter H. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

[237] Peter H. Schönemann, Tim Dorcey, and K. Kienapple. Subadditive concatenation in dissimilarity judgements. *Perception and Psychophysics*, 38:1–17, 1985.

[238] Joshua A. Singer. *Log-Penalized Linear Regression*. PhD thesis, Stanford University, Department of Electrical Engineering, June 2004.
www.stanford.edu/∼dattorro/Josh.pdf

[239] Anthony Man-Cho So and Yinyu Ye. Theory of semidefinite programming for sensor network localization, 2004.
http://www.stanford.edu/∼yyye/local-theory.pdf

[240] Anthony Man-Cho So and Yinyu Ye. A semidefinite programming approach to tensegrity theory and realizability of graphs, 2005.
http://www.stanford.edu/∼yyye/d-real-soda2.pdf

[241] Steve Spain. Polaris Wireless corp., 2005.
http://www.polariswireless.com
Personal communication.

[242] Wolfram Stadler, editor. *Multicriteria Optimization in Engineering and in the Sciences*. Springer-Verlag, 1988.

[243] Henry Stark, editor. *Image Recovery: Theory and Application*. Academic Press, 1987.

[244] Henry Stark. Polar, spiral, and generalized sampling and interpolation. In Robert J. Marks II, editor, *Advanced Topics in Shannon Sampling and Interpolation Theory*, chapter 6, pages 185–218. Springer-Verlag, 1993.

[245] Willi-Hans Steeb. *Matrix Calculus and Kronecker Product with Applications and C++ Programs*. World Scientific Publishing Co., 1997.

[246] Gilbert W. Stewart and Ji-guang Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

[247] Josef Stoer and Christoph Witzgall. *Convexity and Optimization in Finite Dimensions I*. Springer-Verlag, 1970.

[248] Gilbert Strang. Course 18.06: Linear algebra, 2004.
http://web.mit.edu/18.06/www

[249] Gilbert Strang. *Linear Algebra and its Applications*. Harcourt Brace, third edition, 1988.

[250] Gilbert Strang. *Calculus*. Wellesley-Cambridge Press, 1992.

[251] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, second edition, 1998.

[252] S. Straszewicz. Über exponierte Punkte abgeschlossener Punktmengen. *Fund. Math.*, 24:139–143, 1935.

[253] Jos F. Sturm. SeDuMi (self-dual minimization).
Software for optimization over symmetric cones, 2005.
http://sedumi.mcmaster.ca

[254] Jos F. Sturm and Shuzhong Zhang. On cones of nonnegative quadratic functions. *Optimization Online*, April 2001.
http://www.optimization-online.org/DB_HTML/2001/05/324.html

[255] George P. H. Styan. A review and some extensions of Takemura's generalizations of Cochran's theorem. Technical Report 56, Stanford University, Department of Statistics, September 1982.

[256] George P. H. Styan and Akimichi Takemura. Rank additivity and matrix polynomials. Technical Report 57, Stanford University, Department of Statistics, September 1982.

[257] Chen Han Sung and Bit-Shun Tam. A study of projectionally exposed cones. *Linear Algebra and its Applications*, 139:225–252, 1990.

[258] Yoshio Takane. On the relations among four methods of multidimensional scaling. *Behaviormetrika*, 4:29–43, 1977.
http://takane.brinkster.net/Yoshio/p008.pdf

[259] Akimichi Takemura. On generalizations of Cochran's theorem and projection matrices. Technical Report 44, Stanford University, Department of Statistics, August 1980.

[260] Peng Hui Tan and Lars K. Rasmussen. The application of semidefinite programming for detection in CDMA. *IEEE Journal on Selected Areas in Communications*, 19(8), August 2001.

[261] Pablo Tarazaga. Faces of the cone of Euclidean distance matrices: Characterizations, structure and induced geometry. *Linear Algebra and its Applications*, 408:1–13, 2005.

[262] Warren S. Torgerson. *Theory and Methods of Scaling*. Wiley, 1958.

[263] Lloyd N. Trefethen and David Bau, III. *Numerical Linear Algebra*. SIAM, 1997.

[264] Michael W. Trosset. Extensions of classical multidimensional scaling: Computational theory. www.math.wm.edu/~trosset/r.mds.html , 2001. Revision of technical report entitled "Computing distances between convex sets and subsets of the positive semidefinite matrices" first published in 1997.

[265] Michael W. Trosset. Applications of multidimensional scaling to molecular conformation. *Computing Science and Statistics*, 29:148–152, 1998.

[266] Michael W. Trosset. Distance matrix completion by numerical optimization. *Computational Optimization and Applications*, 17(1):11–22, October 2000.

[267] Michael W. Trosset and Rudolf Mathar. On the existence of nonglobal minimizers of the STRESS criterion for metric multidimensional scaling. In *Proceedings of the Statistical Computing Section*, pages 158–162. American Statistical Association, 1997.

[268] Jan van Tiel. *Convex Analysis, an Introductory Text*. Wiley, 1984.

[269] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

[270] Lieven Vandenberghe and Stephen Boyd. Connections between semi-infinite and semidefinite programming. In R. Reemtsen and J.-J. Rückmann, editors, *Semi-Infinite Programming*, chapter 8, pages 277–294. Kluwer Academic Publishers, 1998.

[271] Lieven Vandenberghe and Stephen Boyd. Applications of semidefinite programming. *Applied Numerical Mathematics*, 29(3):283–299, March 1999.

[272] Lieven Vandenberghe, Stephen Boyd, and Shao-Po Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, April 1998.

[273] Robert J. Vanderbei. Convex optimization: Interior-point methods and applications, 1999. www.sor.princeton.edu/∼rvdb/pdf/talks/pumath/talk.pdf

[274] Richard S. Varga. *Geršgorin and His Circles*. Springer-Verlag, 2004.

[275] Martin Vetterli and Jelena Kovačević. *Wavelets and Subband Coding*. Prentice Hall, 1995.

[276] È. B. Vinberg. The theory of convex homogeneous cones. *Transactions of the Moscow Mathematical Society*, 12:340–403, 1963. American Mathematical Society and London Mathematical Society joint translation, 1965.

[277] Marie A. Vitulli. A brief history of linear algebra and matrix theory, 2004.
darkwing.uoregon.edu/∼vitulli/441.sp04/LinAlgHistory.html

[278] John von Neumann. *Functional Operators, Volume II: The Geometry of Orthogonal Spaces*. Princeton University Press, 1950. Reprinted from mimeographed lecture notes first distributed in 1933.

[279] Michael B. Wakin, Jason N. Laska, Marco F. Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin F. Kelly, and Richard G. Baraniuk. An architecture for compressive imaging. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 1273–1276, October 2006.
http://www.dsp.rice.edu/cs/CSCam-ICIP06.pdf

[280] Roger Webster. *Convexity*. Oxford University Press, 1994.

[281] Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised learning of image manifolds by semidefinite programming. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 988–995, 2004.
http://www.cis.upenn.edu/∼lsaul/papers/sde_cvpr04.pdf

[282] Eric W. Weisstein. Mathworld – *A Wolfram Web Resource*, 2007.
http://mathworld.wolfram.com/search

[283] Bernard Widrow and Samuel D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, 1985.

[284] Norbert Wiener. On factorization of matrices. *Commentarii Mathematici Helvetici*, 29:97–111, 1955.

[285] Ami Wiesel, Yonina C. Eldar, and Shlomo Shamai (Shitz). Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels. *IEEE Signal Processing Letters*, 12(9):653–656, September 2005.

[286] Michael P. Williamson, Timothy F. Havel, and Kurt Wüthrich. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by $^1$H nuclear magnetic resonance and distance geometry. *Journal of Molecular Biology*, 182:295–315, 1985.

[287] Willie W. Wong. Cayley-Menger determinant and generalized N-dimensional Pythagorean theorem, November 2003.
http://www.princeton.edu/∼wwong/papers/gp-r.pdf
Application of Linear Algebra: Notes on Talk given to Princeton University Math Club.

[288] William Wooton, Edwin F. Beckenbach, and Frank J. Fleming. *Modern Analytic Geometry*. Houghton Mifflin, 1975.

[289] Margaret H. Wright. The interior-point revolution in optimization: History, recent developments, and lasting consequences. *Bulletin of the American Mathematical Society*, 42(1):39–56, January 2005.

[290] Stephen J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, 1997.

[291] Shao-Po Wu. *max-det Programming with Applications in Magnitude Filter Design*. A dissertation submitted to the department of Electrical Engineering, Stanford University, December 1997.

[292] Shao-Po Wu and Stephen Boyd. sdpsol: A parser/solver for semidefinite programming and determinant maximization problems with matrix structure. User's guide, 1996.
http://www.stanford.edu/∼boyd/SDPSOL.html

[293] Shao-Po Wu and Stephen Boyd. sdpsol: A parser/solver for semidefinite programs with matrix structure. In Laurent El Ghaoui and Silviu-Iulian Niculescu, editors, *Advances in Linear Matrix Inequality Methods in Control*, chapter 4, pages 79–91. SIAM, 2000.
http://www.stanford.edu/∼boyd/sdpsol.html

[294] Shao-Po Wu, Stephen Boyd, and Lieven Vandenberghe. FIR filter design via spectral factorization and convex optimization, 1997.
http://www.stanford.edu/∼boyd/reports/magdes.pdf

[295] Naoki Yamamoto and Maryam Fazel. A computational approach to quantum encoder design for purity optimization, 2006.
http://arxiv.org/abs/quant-ph/0606106

[296] David D. Yao, Shuzhong Zhang, and Xun Yu Zhou. Stochastic linear-quadratic control via primal-dual semidefinite programming. *SIAM Review*, 46(1):87–111, March 2004. Erratum: p.193 herein.

[297] Yinyu Ye. Semidefinite programming for Euclidean distance geometric optimization. Lecture notes, 2003.
http://www.stanford.edu/class/ee392o/EE392o-yinyu-ye.pdf

[298] Yinyu Ye. Convergence behavior of central paths for convex homogeneous self-dual cones, 1996.
http://www.stanford.edu/~yyye/yyye/ye.ps

[299] Yinyu Ye. *Interior Point Algorithms: Theory and Analysis*. Wiley, 1997.

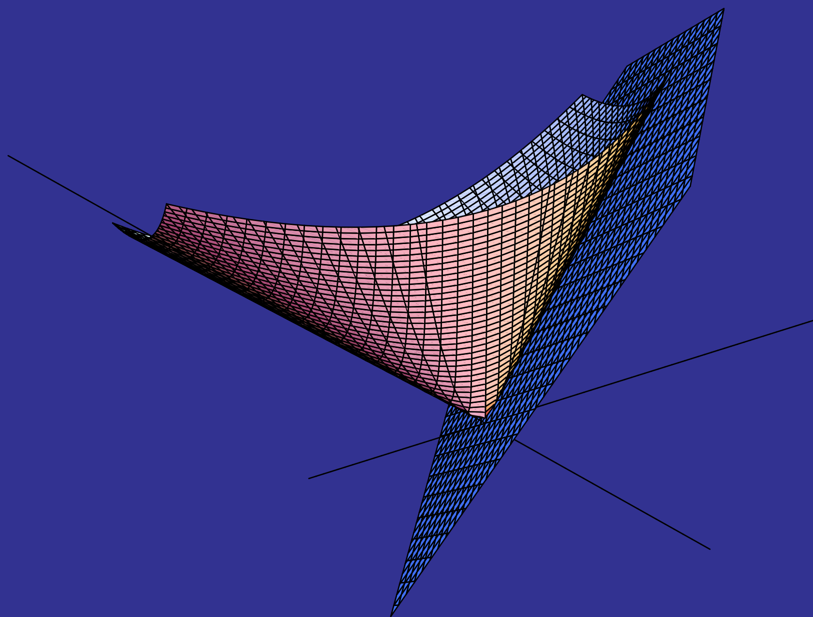
[300] D. C. Youla. Mathematical theory of image restoration by the method of convex projection. In Henry Stark, editor, *Image Recovery: Theory and Application*, chapter 2, pages 29–77. Academic Press, 1987.

[301] Fuzhen Zhang. *Matrix Theory: Basic Results and Techniques*. Springer-Verlag, 1999.

[302] Günter M. Ziegler. Kissing numbers: Surprises in dimension four. *Emissary*, pages 4–5, Spring 2004.
http://www.msri.org/communications/emissary

# Index


Felice Frankel
photography



Convex Optimization & Euclidean Distance Geometry, Dattorro

ISBN 0-9764013-0-4

9 780976 401308 59999

$\mathcal{M}\varepsilon\beta oo$ *Publishing* USA