18. Symmetric cones

- definition
- spectral decomposition
- quadratic representation
- log-det barrier

Introduction

This lecture: theoretical properties of the following cones

- nonnegative orthant

$$\mathbf{R}^p_+ = \{x \in \mathbf{R}^p \mid x_k \geq 0,\ k = 1, \ldots, p\}$$

- second-order cone

$$\mathcal{Q}^p = \{(x_0, x_1) \in \mathbf{R} \times \mathbf{R}^{p-1} \mid \|x_1\|_2 \leq x_0\}$$

- positive semidefinite cone

$$\mathcal{S}^p = \{x \in \mathbf{R}^{p(p+1)/2} \mid \mathrm{mat}(x) \succeq 0\}$$

these cones are not only self-dual, but **symmetric** (also known as *self-scaled*)

Outline

- **definition**
- spectral decomposition
- quadratic representation
- log-det barrier

Cones of squares

the three basic cones can be expressed as cones of squares

$$x^2 = x \circ x$$

for appropriately defined vector products $x \circ y$

- nonnegative orthant: componentwise product $x \circ y = \mathbf{diag}(x)y$
- second-order cone: the product of $x = (x_0, x_1)$ and $y = (y_0, y_1)$ is

$$x \circ y = \frac{1}{\sqrt{2}} \left[\begin{array}{c} x^T y \\ x_0 y_1 + y_0 x_1 \end{array} \right]$$

- positive semidefinite cone: symmetrized matrix product

$$x \circ y = \frac{1}{2} \mathrm{vec}(XY + YX) \quad \text{with } X = \mathrm{mat}(x),\ Y = \mathrm{mat}(Y)$$

Symmetric cones

the vector products satisfy the following properties

1. $x \circ y$ is bilinear (linear in x for fixed y and vice-versa)
2. $x \circ y = y \circ x$
3. $x^2 \circ (y \circ x) = (x^2 \circ y) \circ x$
4. $x^T(y \circ z) = (x \circ y)^T z$

except for the componentwise product, the products are not associative:

$$x \circ (y \circ z) \neq (x \circ y) \circ z \quad \text{in general}$$

Definition: a cone is *symmetric* if it is the cone of squares

$$\{x^2 = x \circ x \mid x \in \mathbf{R}^n\}$$

for a product $x \circ y$ that satisfies these four properties

Classification

- symmetric cones are studied in the theory of Euclidean Jordan algebras
- all possible symmetric cones have been characterized

List of symmetric cones

- the second-order cone
- the p.s.d. cone of Hermitian matrices with real, complex, or quaternion entries
- 3×3 positive semidefinite matrices with octonion entries
- Cartesian products of these 'primitive' symmetric cones (such as $\mathbf{R}^p_+$)

Practical implication

can focus on $\mathcal{Q}^p$, $\mathcal{S}^p$ and study these cones using elementary linear algebra

Outline

- definition
- **spectral decomposition**
- quadratic representation
- log-det barrier

Vector product

with each symmetric cone K we associate a bilinear vector product

- for $\mathbf{R}^p_+$, $\mathcal{Q}^p$, $\mathcal{S}^p$ we use the products on page 18-3
- for a cone $K = K_1 \times \cdots \times K_N$, with K_i of one of the three basic types,

$$(x_1, \ldots, x_N) \circ (y_1, \ldots, y_N) = (x_1 \circ y_1, \ldots, x_N \circ y_N)$$

we refer to the product associated with the cone K as 'the product for K'

Identity element

Identity element: the element $\mathbf{e}$ that satisfies $\mathbf{e} \circ x = x \circ \mathbf{e} = x$ for all x

- product for $\mathbf{R}^p_+$: $\mathbf{e} = \mathbf{1} = (1, 1, \ldots, 1)$
- product for $\mathcal{Q}^p$: $\mathbf{e} = (\sqrt{2}, 0, \ldots, 0)$
- product for $\mathcal{S}^p$: $\mathbf{e} = \mathrm{vec}(I)$
- product for $K_1 \times \cdots \times K_N$: the product of the N identity elements

note we use the same symbol $\mathbf{e}$ for the identity element for each product

Rank of the cone: $\theta = \mathbf{e}^T\mathbf{e}$ is called the *rank* of K

$$\theta = p \quad (K = \mathbf{R}^p_+), \qquad \theta = 2 \quad (K = \mathcal{Q}^p), \qquad \theta = p \quad (K = \mathcal{S}^p)$$

and $\theta = \sum_{i=1}^N \theta_i$ if $K = K_1 \times \cdots \times K_N$ and θ_i is the rank of K_i

Spectral decomposition

with each symmetric cone/product we associate a 'spectral' decomposition

$$x = \sum_{i=1}^{\theta} \lambda_i q_i$$

λ_i are the eigenvalues of x; the eigenvectors q_i satisfy

$$q_i^2 = q_i, \qquad q_i \circ q_j = 0 \quad (i \neq j), \qquad \sum_{i=1}^{\theta} q_i = \mathbf{e}$$

- theory can be developed from properties of the vector product on page 18-4
- we will define the decomposition by enumerating the symmetric cones

Spectral decomposition for primitive cones

Positive semidefinite cone ($K = \mathcal{S}^p$)

spectral decomposition of x follows from eigendecomposition of $\mathrm{mat}(x)$:

$$\mathrm{mat}(x) = \sum_{i=1}^{p} \lambda_i v_i v_i^T, \qquad q_i = \mathrm{vec}(v_i v_i^T)$$

Second-order cone ($K = \mathcal{Q}^p$)

spectral decomposition of $x = (x_0, x_1) \in \mathbf{R} \times \mathbf{R}^{p-1}$ is

$$\lambda_i = \frac{x_0 \pm \|x_1\|_2}{\sqrt{2}}, \qquad q_i = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ \pm y \end{bmatrix}, \qquad i = 1, 2$$

$y = x_1/\|x_1\|_2$ if $x_1 \neq 0$, and y is an arbitrary unit-norm vector otherwise

Spectral decomposition for composite cones

Product cone ($K = K_1 \times \cdots \times K_N$)

- spectral decomposition follows from decomposition of different blocks
- example ($K = K_1 \times K_2$): decomposition of $x = (x_1, x_2)$ is

$$\left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \sum_{i=1}^{\theta_1} \lambda_{1i} \left[\begin{array}{c} q_{1i} \\ 0 \end{array}\right] + \sum_{i=1}^{\theta_2} \lambda_{2i} \left[\begin{array}{c} 0 \\ q_{2i} \end{array}\right]$$

where $x_j = \sum_{i=1}^{\theta_j} \lambda_{ji} q_{ji}$ is the spectral decomposition of x_j, $j = 1, 2$

Nonnegative orthant ($K = \mathbf{R}^p_+$)

$$\lambda_i = x_i, \qquad q_i = e_i \quad (i\text{th unit vector}), \qquad i = 1, \ldots, n$$

Some properties

- the eigenvectors are normalized ($\|q_i\|_2 = 1$) and nonnegative ($q_i \in K$)
- eigenvectors are orthogonal: $q_i^T q_j = 0$ for $i \neq j$

  can be verified from the definitions, or from the properties on page 18-4

$$q_i^T q_j = q_i^T (q_j \circ q_j) = (q_i \circ q_j)^T q_j = 0$$

- $\mathbf{e}^T q_i = (\sum_j q_j)^T q_i = q_i^T q_i = 1$
- $\mathbf{e}^T x = \sum_{i=1}^{\theta} \lambda_i$
- $x \in K$ if and only $\lambda_i \geq 0$ for $i = 1, \ldots, \theta$
- $x \in \operatorname{int} K$ if and only $\lambda_i > 0$ for $i = 1, \ldots, \theta$

Trace, determinant, and norm

$$\operatorname{tr} x = \sum_{i=1}^{\theta} \lambda_i, \qquad \det x = \prod_{i=1}^{\theta} \lambda_i, \qquad \|x\|_F = (\sum_{i=1}^{\theta} \lambda_i^2)^{1/2}$$

- positive semidefinite cone ($K = \mathcal{S}^p$)

$$\operatorname{tr} x = \operatorname{tr}(\operatorname{mat}(x)), \quad \det x = \det(\operatorname{mat}(x)), \quad \|x\|_F = \|\operatorname{mat}(x)\|_F$$

- second-order cone ($K = \mathcal{Q}^p$)

$$\operatorname{tr} x = \sqrt{2} x_0, \qquad \det x = \frac{1}{2}(x_0^2 - x_1^T x_1), \qquad \|x\|_F = \|x\|_2$$

- nonnegative orthant ($K = \mathbf{R}_+^p$)

$$\operatorname{tr} x = \sum_{i=1}^{p} x_i, \qquad \det x = \prod_{i=1}^{p} x_i, \qquad \|x\|_F = \|x\|_2$$

Powers and inverse

powers of x can be defined in terms of the spectral decomposition

$$x^\alpha = \sum_i \lambda_i^\alpha q_i$$

(exists if λ_i^α is defined for all i)

- $x^\alpha \circ x^\beta = x^{\alpha+\beta}$

$$x^\alpha \circ x^\beta = \left(\sum_{i=1}^{\theta} \lambda_i^\alpha q_i\right) \circ \left(\sum_{i=1}^{\theta} \lambda_i^\beta q_i\right) = \sum_{i=1}^{\theta} \lambda_i^{\alpha+\beta} q_i = x^{\alpha+\beta}$$

- x is invertible if all $\lambda_i \neq 0$ (*i.e.*, $\det x \neq 0$)

  inverse $x^{-1} = \sum_{i=1}^{\theta} \lambda_i^{-1} q_i$ satisfies $x \circ x^{-1} = x^{-1} \circ x = \mathbf{e}$

Expressions for inverse

for invertible x (*i.e.*, $\lambda_i \neq 0$ for $i = 1, \ldots, \theta$)

- nonnegative orthant ($K = \mathbf{R}^p_+$)

$$x^{-1} = \left(\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_p}\right)$$

- second-order cone ($K = \mathcal{Q}^p$)

$$x^{-1} = \frac{2}{x^T J x} J x, \qquad J = \begin{bmatrix} 1 & 0 \\ 0 & -I \end{bmatrix}$$

- semidefinite cone ($K = \mathcal{S}^p$)

$$x^{-1} = \mathrm{vec}\left(\mathrm{mat}(x)^{-1}\right)$$

Expressions for square root

for nonnegative x (*i.e.*, $\lambda_i \geq 0$ for $i = 1, \ldots, \theta$)

- nonnegative orthant ($K = \mathbf{R}_+^p$)

$$x^{1/2} = \left(\sqrt{x_1}, \sqrt{x_2}, \ldots, \sqrt{x_p}\right)$$

- second-order cone ($K = \mathcal{Q}^p$)

$$x^{1/2} = \frac{1}{2^{1/4}\left(x_0 + \sqrt{x^T J x}\right)^{1/2}} \begin{bmatrix} x_0 + \sqrt{x^T J x} \\ x_1 \end{bmatrix}$$

- semidefinite cone ($K = \mathcal{S}^p$)

$$x^{1/2} = \text{vec}\left(\text{mat}(x)^{1/2}\right)$$

Outline

- definition
- spectral decomposition
- **quadratic representation**
- log-det barrier

Matrix representation of product

since the product is bilinear, it can be expressed as

$$x \circ y = L(x)y$$

$L(x)$ is a symmetric matrix, linearly dependent on x

- nonnegative orthant: $L(x) = \mathbf{diag}(x)$
- second-order cone

$$L(x) = \frac{1}{\sqrt{2}} \begin{bmatrix} x_0 & x_1^T \\ x_1 & x_0 I \end{bmatrix}$$

- semidefinite cone: the matrix defined by

$$L(x)y = \frac{1}{2}(XY + YX), \qquad X = \mathrm{mat}(x), \quad Y = \mathrm{mat}(y)$$

Quadratic representation

the quadratic representation of x is the matrix

$$P(x) = 2L(x)^2 - L(x^2)$$

(terminology is motivated by the property $P(x)\mathbf{e} = x^2$)

- nonnegative orthant

$$P(x) = \mathbf{diag}(x)^2$$

- second-order cone

$$P(x) = xx^T - \frac{x^T J x}{2} J, \qquad J = \begin{bmatrix} 1 & 0 \\ 0 & -I \end{bmatrix}$$

- positive semidefinite cone

$$P(x)y = \mathrm{vec}(XYX) \quad \text{where } X = \mathrm{mat}(x),\, Y = \mathrm{mat}(y)$$

Some useful properties

Powers

- $P(x)^\alpha = P(x^\alpha)$ if x^α exists
- $P(x)x^{-1} = x$ if x is invertible

Derivative of x^{-1}: for x invertible, $P(x)^{-1}$ is the derivative of $-x^{-1}$, *i.e.*,

$$\frac{d}{d\alpha}(x + \alpha y)^{-1}\Big|_{\alpha=0} = -P(x)^{-1}y$$

Scaling with $P(x)$

affine transformations $y \to P(x)y$ have important properties

- if x is invertible, then

$$P(x)K = K, \qquad P(x) \operatorname{int} K = \operatorname{int} K$$

  multiplication with $P(x)$ preserves the conic inequalities

- if x and y are invertible, then $P(x)y$ is invertible with inverse

$$(P(x)y)^{-1} = P(x^{-1})y^{-1} = P(x)^{-1}y^{-1}$$

- quadratic representation of $P(x)y$

$$P\left(P(x)y\right) = P(x)P(y)P(x)$$

  hence also $\det\left(P(x)y\right) = (\det x)^2 \det y$

Distance to cone boundary

for $x \succ 0$ define $\sigma_x(y) = 0$ if $y \succeq 0$ and

$$\sigma_x(y) = -\lambda_{\min}\left(P(x^{-1/2})y\right) \text{ otherwise}$$

$\sigma_x(y)$ characterizes distance of x to the boundary of K in the direction y:

$$x + \alpha y \succeq 0 \qquad \Longleftrightarrow \qquad \alpha\sigma_x(y) \le 1$$

Proof:

- from the definition of P on page 18-17: $P(x^{1/2})\mathbf{e} = x$
- therefore, with $v = P(x^{-1/2})y$

$$\begin{aligned} x + \alpha y \succeq 0 &\iff P(x^{1/2})(\mathbf{e} + \alpha v) \succeq 0 \\ &\iff \mathbf{e} + \alpha v \succeq 0 \\ &\iff 1 + \alpha\lambda_i(v) \ge 0 \end{aligned}$$

Scaling point

for a pair $s, z \succ 0$, the point

$$w = P(z^{-1/2}) \left(P(z^{1/2})s\right)^{1/2}$$

satisfies $w \succ 0$ and $s = P(w)z$

- the linear transformation $P(w)$ preserves the cone and maps z to s
- equivalently, $v = w^{1/2}$ defines a scaling $P(v) = P(w)^{1/2}$ that satisfies

$$P(v)^{-1}s = P(v)z$$

Proof: we use the properties

$$P(x)\mathbf{e} = x^2, \qquad P(P(x)y) = P(x)P(y)P(x)$$

- if we define $u = P(z^{1/2})s$, we can write $P(w)$ as

$$P(w) = P(z^{-1/2})P(u^{1/2})P(z^{-1/2})$$

- therefore

$$\begin{aligned} P(w)z &= P(z^{-1/2})P(u^{1/2})\mathbf{e} \\ &= P(z^{-1/2})u \\ &= s \end{aligned}$$

Scaling point for nonnegative orthant

Scaling point

$$w = \left(\sqrt{s_1/z_1},\ \sqrt{s_2/z_2},\ \ldots,\ \sqrt{s_p/z_p}\right)$$

Scaling transformation: a positive diagonal scaling

$$P(w) = \begin{bmatrix} s_1/z_1 & 0 & \cdots & 0 \\ 0 & s_2/z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p/z_p \end{bmatrix}$$

Scaling point for positive semidefinite cone

Scaling point: $w = \mathrm{vec}(RR^T)$

- R simultaneously diagonalizes $\mathrm{mat}(z)$ and $\mathrm{mat}(s)^{-1}$:

$$R^T \mathrm{mat}(z) R = R^{-1} \mathrm{mat}(s) R^{-T} = \Sigma$$

- can be computed from two Cholesky factorizations and an SVD: if

$$\mathrm{mat}(s) = L_1 L_1^T, \qquad \mathrm{mat}(z) = L_2 L_2^T, \qquad L_2^T L_1 = U \Sigma V^T$$

  then $R = L_1 V \Sigma^{-1/2} = L_2 U \Sigma^{1/2}$

Scaling transformation: a congruence transformation

$$P(w)y = RR^T \mathrm{mat}(y) RR^T$$

Outline

- definition
- spectral decomposition
- quadratic representation
- **log-det barrier**

Log-det barrier

Definition: the log-det barrier of a symmetric cone K is

$$\phi(x) = -\log\det x = -\sum_{i=1}^{\theta} \log \lambda_i, \qquad \operatorname{dom}\phi = \operatorname{int} K$$

ϕ is logarithmically homogeneous with degree θ (*i.e.*, the rank of K)

- nonnegative orthant ($K = \mathbf{R}^p_+$): $\phi(x) = -\sum_{i=1}^{p} \log x_i$
- second-order cone ($K = \mathcal{Q}^p$):

$$\phi(x) = -\log(x_0^2 - x_1^T x_1) + \log 2$$

- semidefinite cone ($K = \mathcal{S}^p$): $\phi(x) = -\log\det(\operatorname{mat} x)$
- composition $K = K_1 \times \cdots \times K_N$: sum of the log-det barriers

Convexity

ϕ is a convex function

Proof: consider arbitrary y and let λ_i be the eigenvalues of $v = P(x^{-1/2})y$

the restriction $g(\alpha) = \phi(x + \alpha y)$ of ϕ to the line $x + \alpha y$ is

$$\begin{aligned}
g(\alpha) &= -\log\det(x + \alpha y) \\
&= -\log\det\left(P(x^{1/2})(\mathbf{e} + \alpha v)\right) \\
&= -\log\det x - \log\det(\mathbf{e} + \alpha v) \\
&= -\log\det x - \sum_{i=1}^{\theta} \log(1 + \alpha\lambda_i)
\end{aligned}$$

(line 3 follows from page 18-18 and page 18-21)

hence restriction of ϕ to arbitrary line is convex

Gradient and Hessian

the gradient and Hessian of ϕ at a point $x \succ 0$ are

$$\nabla \phi(x) = -x^{-1}, \qquad \nabla^2 \phi(x) = P(x)^{-1} = P(x^{-1})$$

Proof: continues from last page

$$\nabla \phi(x)^T y = g'(0) = -\sum_{i=1}^{\theta} \lambda_i = -\operatorname{tr}(P(x^{-1/2})y) = -\mathbf{e}^T P(x^{-1/2})y$$

since this holds for all y, $\nabla \phi(x) = -P(x^{-1/2})\mathbf{e} = -x^{-1}$

the expression for the Hessian follows from page 18-18

Dikin ellipsoid theorem for symmetric cones

recall the definition of the Dikin ellipsoid at $x \succ 0$:

$$\mathcal{E}_x = \{x + y \mid y^T \nabla^2 \phi(x) y \le 1\} = \{x + y \mid y^T P(x)^{-1} y \le 1\}$$

- $x + y \in \mathcal{E}_x$ if and only if the eigenvalues λ_i of $P(x^{-1/2})y$ satisfy

$$\sum_{i=1}^{\theta} \lambda_i^2 \le 1$$

- for symmetric cones the Dikin ellipsoid theorem $\mathcal{E}_x \subseteq K$ follows from

$$\sum_{i=1}^{\theta} \lambda_i^2 \le 1 \quad \Longrightarrow \quad \min_i \lambda_i \ge -1$$

  therefore $x + y \in \mathcal{E}_x$ implies $\sigma_x(y) \le 1$ and $x + y \succeq 0$

Generalized convexity of inverse

for all $w \succeq 0$, the function

$$f(x) = w^T x^{-1}, \qquad \mathrm{dom}\, f = \mathrm{int}\, K$$

is convex

Proof: restriction $g(\alpha) = f(x + \alpha y)$ of f to a line is

$$\begin{aligned} g(\alpha) = w^T (x + \alpha y)^{-1} &= w^T \left(P(x^{-1/2})(\mathbf{e} + \alpha v) \right)^{-1} \\ &= w^T P(x^{1/2})(\mathbf{e} + \alpha v)^{-1} \\ &= \sum_{i-1}^{\theta} \frac{w^T P(x^{1/2}) q_i}{1 + \alpha \lambda_i} \end{aligned}$$

- λ_i, q_i are eigenvalues and eigenvectors of $v = P(x^{-1/2})y$
- $g(\alpha)$ is convex because $P(x^{1/2})K = K$; therefore $w^T P(x^{1/2}) q_i \geq 0$

Self-concordance

for all $x \succ 0$ and all y,

$$\frac{d}{d\alpha}\nabla^2\phi(x+\alpha y)\Big|_{\alpha=0} \preceq 2\sigma_x(y)\nabla^2\phi(x)$$

- from page 18-20, with $v = P(x^{-1/2})y$ and $\lambda_i = \lambda_i(v)$

$$\sigma_x(y) = -\min\{0, \lambda_{\min}\} \leq (\sum_{i=1}^{\theta}\lambda_i^2)^{1/2} = \|y\|_x$$

  hence, inequality implies self-concordance inequality on page 16-2

- this shows that the log-det barrier ϕ is a θ-normal barrier

Proof: choose any $\sigma > \sigma_x(y)$; by definition of σ_x we have $\sigma x + y \succeq 0$

$$\frac{d}{d\alpha}\nabla^2\phi(x+\alpha y) = \frac{d}{d\alpha}\nabla^2\phi(x-\alpha\sigma x) + \frac{d}{d\alpha}\nabla^2\phi(x+\alpha(\sigma x+y))$$

- from page 16-24, the first term on the r.h.s. (evaluated at $\alpha = 0$) is

$$-\sigma\,\frac{d}{dt}\nabla^2\phi(tx)\Big|_{t=1} = 2\sigma\nabla^2\phi(x)$$

- 2nd term is $\nabla^2 g(x)$ where $g(u) = w^T\nabla\phi(u)$ and $w = \sigma x + y$:

$$\nabla g(u) = \frac{d}{d\alpha}\nabla\phi(u+\alpha w)\Big|_{\alpha=0}, \qquad \nabla^2 g(u) = \frac{d}{d\alpha}\nabla^2\phi(u+\alpha w)\Big|_{\alpha=0}$$

- $\nabla^2 g(u) \preceq 0$ because from page 18-29, $g(u) = -w^T u^{-1}$ is concave

References

- J. Faraut and A. Korányi, *Analysis on Symmetric Cones* (1994).
- Yu. E. Nesterov and M. J. Todd, *Self-scaled barriers and interior-point methods for convex programming*, Mathematics of Operations Research (1997).
- L. Faybusovich, *Euclidean Jordan algebras and interior-point algorithms*, Positivity (1997).
- J. F. Sturm, *Similarity and other spectral relations for symmetric cones*, Linear Algebra and Its Applications (2000).
- S. H. Schmieta and F. Alizadeh, *Extension of primal-dual interior point algorithms to symmetric cones*, Math. Prog. (2003).
- F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Math. Prog. (2003).