18. Symmetric cones

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

This lecture: theoretical properties of the following cones

- nonnegative orthant

\[ \mathbb{R}^+_p = \{ x \in \mathbb{R}^p \mid x_k \geq 0, \ k = 1, \ldots, p \} \]

- second-order cone

\[ Q^p = \{ (x_0, x_1) \in \mathbb{R} \times \mathbb{R}^{p-1} \mid \|x_1\|_2 \leq x_0 \} \]

- positive semidefinite cone

\[ S^p = \{ x \in \mathbb{R}^{p(p+1)/2} \mid \text{mat}(x) \succeq 0 \} \]

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

\[
    x \circ y = \frac{1}{\sqrt{2}} \begin{bmatrix}
        x^T y \\
        x_0y_1 + y_0x_1
    \end{bmatrix}
\]

- positive semidefinite cone: symmetrized matrix product

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

\[
\{ x^2 = x \circ x \mid x \in \mathbb{R}^n \}
\]

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
- $3 \times 3$ positive semidefinite matrices with octonion entries
- Cartesian products of these ‘primitive’ symmetric cones (such as $\mathbb{R}_+^P$)

Practical implication

can focus on $Q^p$, $S^p$ 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 $R_+^p$, $Q^p$, $S^p$ 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 $e$ that satisfies $e \circ x = x \circ e = x$ for all $x$

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

Note we use the same symbol $e$ for the identity element for each product.

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

$$\theta = p \quad (K = \mathbb{R}_+^p), \quad \theta = 2 \quad (K = \mathbb{Q}^p), \quad \theta = p \quad (K = \mathbb{S}^p)$$

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

with each symmetric cone/product we associate a ‘spectral’ decomposition

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

\( \lambda_i \) are the eigenvalues of \( x \); the eigenvectors \( q_i \) satisfy

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

- 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 \((K = S^p)\)

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

\[
\text{mat}(x) = \sum_{i=1}^{p} \lambda_i v_i v_i^T, \quad q_i = \text{vec}(v_i v_i^T)
\]

Second-order cone \((K = Q^p)\)

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

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

\(y = x_1/\|x_1\|_2\) 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

\[
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \sum_{i=1}^{\theta_1} \lambda_{1i} \begin{bmatrix} q_{1i} \\ 0 \end{bmatrix} + \sum_{i=1}^{\theta_2} \lambda_{2i} \begin{bmatrix} 0 \\ q_{2i} \end{bmatrix}
\]

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

**Nonnegative orthant** \((K = \mathbb{R}_+^p)\)

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

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

- $e^T q_i = (\sum_j q_j)^T q_i = q_i^T q_i = 1$

- $e^T x = \sum_{i=1}^{\theta} \lambda_i$

- $x \in K$ if and only $\lambda_i \geq 0$ for $i = 1, \ldots, \theta$

- $x \in \text{int } K$ if and only $\lambda_i > 0$ for $i = 1, \ldots, \theta$
Trace, determinant, and norm

\[ \text{tr } x = \sum_{i=1}^{\theta} \lambda_i, \quad \text{det } x = \prod_{i=1}^{\theta} \lambda_i, \quad \| x \|_F = (\sum_{i=1}^{\theta} \lambda_i^2)^{1/2} \]

- positive semidefinite cone \((K = S^p)\)
  \[ \text{tr } x = \text{tr}(\text{mat}(x)), \quad \text{det } x = \text{det}(\text{mat}(x)), \quad \| x \|_F = \| \text{mat}(x) \|_F \]

- second-order cone \((K = Q^p)\)
  \[ \text{tr } x = \sqrt{2}x_0, \quad \text{det } x = \frac{1}{2}(x_0^2 - x_1^T x_1), \quad \| x \|_F = \| x \|_2 \]

- nonnegative orthant \((K = \mathbb{R}_+^p)\)
  \[ \text{tr } x = \sum_{i=1}^{p} x_i, \quad \text{det } x = \prod_{i=1}^{p} x_i, \quad \| x \|_F = \| x \|_2 \]
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 \( \lambda_i^\alpha \) 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 \( x \circ x^{-1} = x^{-1} \circ x = e \)
Expressions for inverse

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

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

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

- **second-order cone ($K = Q^p$)**

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

- **semidefinite cone ($K = S^p$)**

  $$x^{-1} = \text{vec} \left( \text{mat}(x)^{-1} \right)$$
Expressions for square root

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

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

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

- second-order cone ($K = Q^p$)

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

- semidefinite cone ($K = S^p$)

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

- 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

\[
L(x)y = \frac{1}{2}(XY + YX), \quad X = \text{mat}(x), \quad Y = \text{mat}(y)
\]
Quadratic representation

the quadratic representation of $x$ is the matrix

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

(terminology is motivated by the property $P(x)e = x^2$)

- nonnegative orthant

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

- second-order cone

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

- positive semidefinite cone

$$P(x)y = \text{vec}(XYX) \quad \text{where} \quad X = \text{mat}(x), \; Y = \text{mat}(y)$$
Some useful properties

Powers

- $P(x)^\alpha = P(x^\alpha)$ if $x^\alpha$ 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.,

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

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

• if $x$ is invertible, then

$$P(x)K = K, \quad P(x) \text{ int } K = \text{ int } K$$

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 $P(x)y$

$$P(P(x)y) = P(x)P(y)P(x)$$

hence also $\det (P(x)y) = (\det x)^2 \det y$
Distance to cone boundary

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

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

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

\[
x + \alpha y \succeq 0 \iff \alpha \sigma_x(y) \leq 1
\]

Proof:

• from the definition of \( P \) on page 18-17: \( P(x^{1/2})e = x \)

• therefore, with \( v = P(x^{-1/2})y \)

\[
x + \alpha y \succeq 0 \iff P(x^{1/2})(e + \alpha v) \succeq 0
\]

\[
\iff e + \alpha v \succeq 0
\]

\[
\iff 1 + \alpha \lambda_i(v) \geq 0
\]
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

\[ P(x)e = x^2, \quad P(P(x)y) = P(x)P(y)P(x) \]

- 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

\[
\begin{align*}
P(w)z & = P(z^{-1/2})P(u^{1/2})e \\
& = P(z^{-1/2})u \\
& = s
\end{align*}
\]
Scaling point for nonnegative orthant

Scaling point

\[ w = \left( \sqrt{\frac{s_1}{z_1}}, \sqrt{\frac{s_2}{z_2}}, \ldots, \sqrt{\frac{s_p}{z_p}} \right) \]

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: \( w = \text{vec}(RR^T) \)

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

\[
R^T \text{mat}(z) R = R^{-1} \text{mat}(s) R^{-T} = \Sigma
\]

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

\[
\text{mat}(s) = L_1 L_1^T, \quad \text{mat}(z) = L_2 L_2^T, \quad L_2^T L_1 = U \Sigma V^T
\]

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

Scaling transformation: a congruence transformation

\[
P(w)y = RR^T \text{mat}(y)RR^T
\]
Outline

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

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

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

$\phi$ is logarithmically homogeneous with degree $\theta$ (i.e., the rank of $K$)

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

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

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

\( \phi \) is a convex function

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

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

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

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

hence restriction of \( \phi \) to arbitrary line is convex
Gradient and Hessian

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

\[
\nabla \phi(x) = -x^{-1}, \quad \nabla^2 \phi(x) = P(x)^{-1} = P(x^{-1})
\]

Proof: continues from last page

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

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

the expression for the Hessian follows from page 18-18
Dikin ellipsoid theorem for symmetric cones

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

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

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

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

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

$$\sum_{i=1}^{\theta} \lambda_i^2 \leq 1 \implies \min_i \lambda_i \geq -1$$

therefore $x + y \in \mathcal{E}_x$ implies $\sigma_x(y) \leq 1$ and $x + y \succeq 0$
Generalized convexity of inverse

for all $w \succeq 0$, the function

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

is convex

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

$$g(\alpha) = w^T (x + \alpha y)^{-1} = w^T \left( P(x^{-1/2})(e + \alpha v) \right)^{-1}$$

$$= w^T P(x^{1/2})(e + \alpha v)^{-1}$$

$$= \sum_{i-1}^{\theta} \frac{w^T P(x^{1/2})q_i}{1 + \alpha \lambda_i}$$

- $\lambda_i, q_i$ 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 $y$,

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

• 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 $\phi$ is a $\theta$-normal barrier
Proof: choose any $\sigma > \sigma_x(y)$; by definition of $\sigma_x$ 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

\[
-\sigma \left. \frac{d}{dt} \nabla^2 \phi(tx) \right|_{t=1} = 2\sigma \nabla^2 \phi(x)
\]

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

\[
\nabla g(u) = \left. \frac{d}{d\alpha} \nabla \phi(u + \alpha w) \right|_{\alpha=0}, \quad \nabla^2 g(u) = \left. \frac{d}{d\alpha} \nabla^2 \phi(u + \alpha w) \right|_{\alpha=0}
\]

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