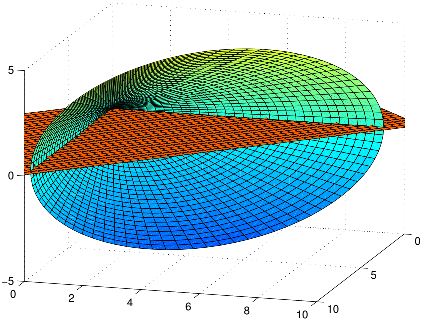
An Introduction to Conic and Semidefinite Programming with Applications

Christoph Helmberg
Technische Universität Chemnitz

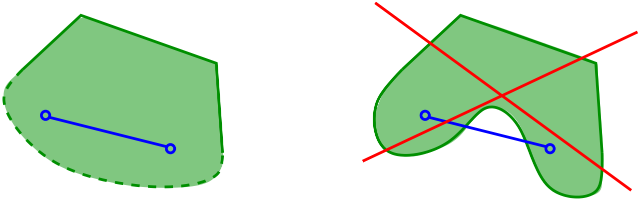


Edinburgh, October 19, 2018

Contents

Convex Sets

A set $C \subseteq \mathbb{R}^n$ is convex, if for all $x, y \in C$
the straight line segment $\{\alpha x + (1-\alpha)y : \alpha \in [0,1]\}$ lies in C.

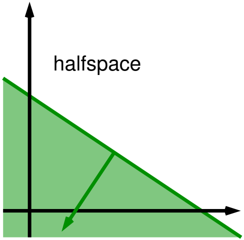


Convex Sets

A set $C \subseteq \mathbb{R}^n$ is convex, if for all $x, y \in C$
the straight line segment $\{\alpha x + (1-\alpha)y : \alpha \in [0,1]\}$ lies in C.

Examples:

- $\emptyset$, $\mathbb{R}^n$
- halfspaces: for given $z \in \mathbb{R}^n, \zeta \in \mathbb{R}$ the set $\{x \in \mathbb{R}^n : z^T x \geq \zeta\}$



Convex Sets

A set $C \subseteq \mathbb{R}^n$ is convex, if for all $x, y \in C$
the straight line segment $\{\alpha x + (1-\alpha)y : \alpha \in [0,1]\}$ lies in C.

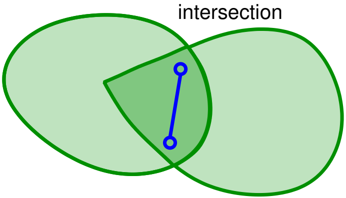
Examples:

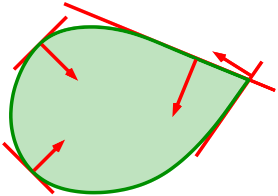
- $\emptyset$, $\mathbb{R}^n$
- halfspaces: for given $z \in \mathbb{R}^n, \zeta \in \mathbb{R}$ the set $\{x \in \mathbb{R}^n : z^T x \geq \zeta\}$

Note:

- the intersection of convex sets is convex
- any closed convex set is the intersection of the halfspaces containing it

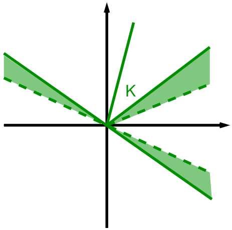
---





(Convex) Cones

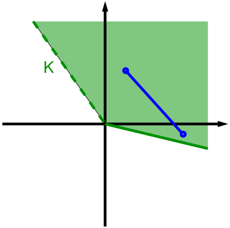
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.



(Convex) Cones

A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

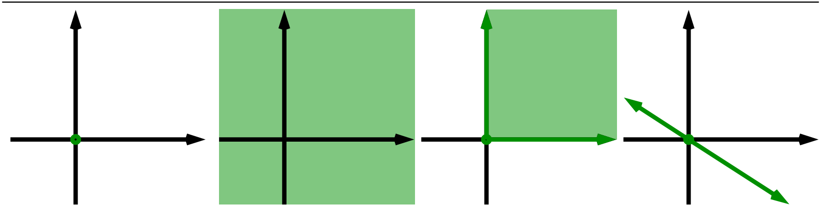


(Convex) Cones

A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.



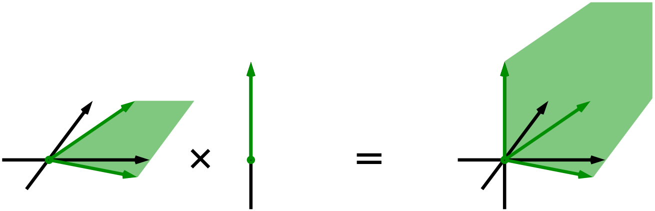
(Convex) Cones

A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x+y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.



(Convex) Cones

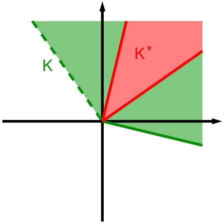
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

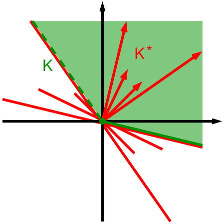
A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ is its dual cone.





(Convex) Cones

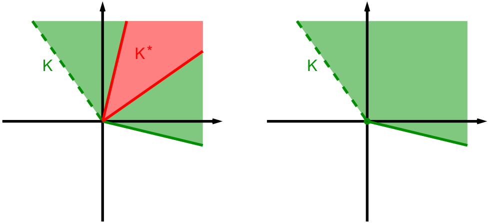
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x+y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \; \forall x \in K\}$ is its dual cone.



(Convex) Cones

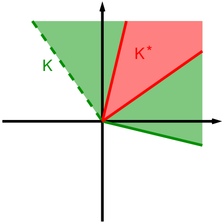
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

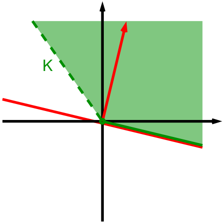
A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ is its dual cone.





(Convex) Cones

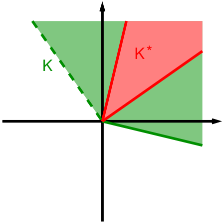
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

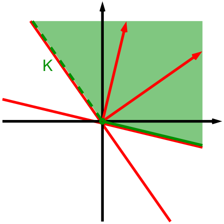
A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x+y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ is its dual cone.





(Convex) Cones

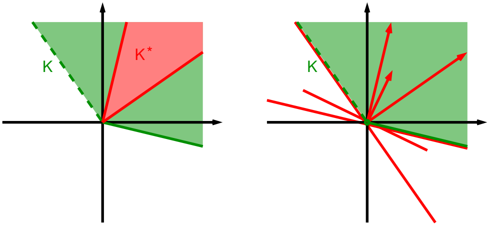
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ is its dual cone.



(Convex) Cones

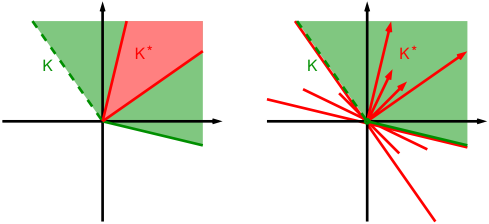
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \; \forall x \in K\}$ is its dual cone.



K^* is always closed!

(Convex) Cones

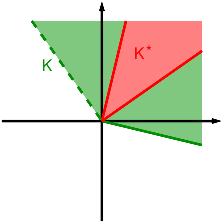
A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

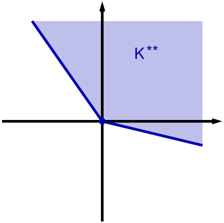
A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ is its dual cone.





$(K^*)^* \supseteq K$,
equality holds iff
K is a closed convex cone!

(Convex) Cones

A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

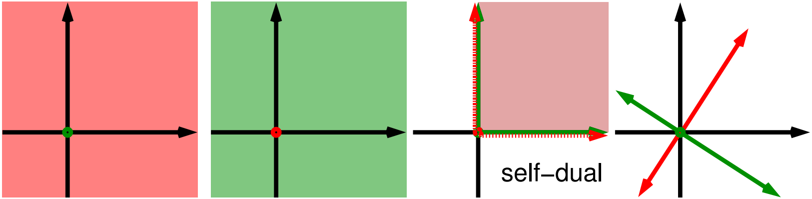
A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x + y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ is its dual cone.

Exs.: $\{0\}^* = \mathbb{R}^n$, $(\mathbb{R}^n)^* = \{0\}$, $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, $(K_1 \times K_2)^* = K_1^* \times K_2^*$.



(Convex) Cones

A set $K \subseteq \mathbb{R}^n$ is a cone, if for $x \in K$ also $\alpha x \in K$ for $\alpha \geq 0$.

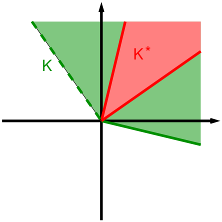
A set $K \subseteq \mathbb{R}^n$ is a convex cone, if $x, y \in K$ and $\alpha \geq 0 \Rightarrow \alpha(x+y) \in K$

Examples: $\emptyset$, $\{0\}$, $\mathbb{R}^n$, $\mathbb{R}^n_+$, linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$.

If $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ are cvx cones, so is $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

For a cvx cone $K \subseteq \mathbb{R}^n$, $K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ is its dual cone.

Exs.: $\{0\}^* = \mathbb{R}^n$, $(\mathbb{R}^n)^* = \{0\}$, $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, $(K_1 \times K_2)^* = K_1^* \times K_2^*$.



Important property for optimisation:

$$\inf_{x \in K} z^T x = \begin{cases} 0 & \Leftrightarrow z \in K^*, \\ -\infty & \textit{otherwise}. \end{cases}$$

Contents

Primal/Dual Linear Programs over Cones

Replace $\mathbb{R}^n_+$ within linear optimisation by a convex cone $K \subseteq \mathbb{R}^n$:

$$
\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array}
\quad \Leftrightarrow \quad
\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in \mathbb{R}^n_+ \end{array}
\quad \rightarrow \quad
\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array}
$$

Primal/Dual Linear Programs over Cones

Replace $\mathbb{R}^n_+$ within linear optimisation by a convex cone $K \subseteq \mathbb{R}^n$:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in \mathbb{R}^n_+ \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array}$$

Define the Lagrange function

$$L(x, y) := c^T x + y^T (b - Ax) \quad \text{for } (x, y) \in K \times \mathbb{R}^m.$$

For $y \in \mathbb{R}^m$ and $Ax = b$ we have $(b - Ax)^T y = 0$, hence

$$\text{for all } y \in \mathbb{R}^m: \quad \inf_{x \in K} L(x, y) \leq \inf\{c^T x \colon Ax = b, x \in K\}.$$

Primal/Dual Linear Programs over Cones

Replace $\mathbb{R}^n_+$ within linear optimisation by a convex cone $K \subseteq \mathbb{R}^n$:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in \mathbb{R}^n_+ \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array}$$

Define the Lagrange function

$$L(x,y) := c^T x + y^T(b - Ax) \quad \text{for } (x,y) \in K \times \mathbb{R}^m.$$

For $y \in \mathbb{R}^m$ and $Ax = b$ we have $(b - Ax)^T y = 0$, hence

$$\text{for all } y \in \mathbb{R}^m\colon \quad \inf_{x \in K} L(x,y) \leq \inf\{c^T x \colon Ax = b, x \in K\}.$$

The best lower bound (Lagrangian relaxation) is

$$\sup_{y \in \mathbb{R}^m} \inf_{x \in K} L(x,y) = \sup_{y \in \mathbb{R}^m} [b^T y + \inf_{x \in K} x^T(c - A^T y)]$$

Primal/Dual Linear Programs over Cones

Replace $\mathbb{R}^n_+$ within linear optimisation by a convex cone $K \subseteq \mathbb{R}^n$:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in \mathbb{R}^n_+ \end{array} \quad \rightarrow \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array}$$

Define the Lagrange function

$$L(x, y) := c^T x + y^T (b - Ax) \quad \text{for } (x, y) \in K \times \mathbb{R}^m.$$

For $y \in \mathbb{R}^m$ and $Ax = b$ we have $(b - Ax)^T y = 0$, hence

$$\text{for all } y \in \mathbb{R}^m\colon \quad \inf_{x \in K} L(x, y) \leq \inf\{c^T x\colon Ax = b, x \in K\}.$$

The best lower bound (Lagrangian relaxation) is

$$\sup_{y \in \mathbb{R}^m} \inf_{x \in K} L(x, y) = \sup_{y \in \mathbb{R}^m} [b^T y + \inf_{x \in K} x^T (c - A^T y)]$$

The inner inf is finite only for $z = c - A^T y \in K^*$ giving the dual program

$$\begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K^* \end{array}$$

Weak and Strong Duality

Let $K \subseteq \mathbb{R}^n$ be a closed convex cone.

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K^* \end{array}$$

Weak duality, i.e., $v(P) \geq v(D)$, always holds by construction.
Equality does NOT hold in general (see later examples)!

Weak and Strong Duality

Let $K \subseteq \mathbb{R}^n$ be a closed convex cone.

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \qquad\qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K^* \end{array}$$

Weak duality, i.e., $v(P) \geq v(D)$, always holds by construction.
Equality does NOT hold in general (see later examples)!

To ensure strong duality we need to require additional properties:
A primal feasible $\bar{x}$ is strictly feasible for (P) if $\bar{x}$ lies in the interior of K,

$$\exists \rho > 0\colon B_\rho(\bar{x}) := \{x \in \mathbb{R}^n \colon \|x - \bar{x}\| \leq \rho\} \subseteq K$$

If such an $\bar{x}$ exists, (P) is strictly feasible.

Weak and Strong Duality

Let $K \subseteq \mathbb{R}^n$ be a closed convex cone.

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K^* \end{array}$$

Weak duality, i.e., $v(P) \geq v(D)$, always holds by construction.
Equality does NOT hold in general (see later examples)!

To ensure strong duality we need to require additional properties:
A primal feasible $\bar{x}$ is strictly feasible for (P) if $\bar{x}$ lies in the interior of K,

$$\exists \rho > 0\colon B_\rho(\bar{x}) := \{x \in \mathbb{R}^n \colon \|x - \bar{x}\| \leq \rho\} \subseteq K$$

If such an $\bar{x}$ exists, (P) is strictly feasible.
A dual feasible $(\bar{y}, \bar{z})$ is strictly feasible for (D), if $\bar{z}$ lies in the interior of K^* ($\exists \rho > 0\colon B_\rho(\bar{z}) \subseteq K^*$). If such $(\bar{y}, \bar{z})$ exist, (D) is strictly feasible.

Weak and Strong Duality

Let $K \subseteq \mathbb{R}^n$ be a closed convex cone.

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K^* \end{array}$$

Weak duality, i.e., $v(P) \geq v(D)$, always holds by construction.
Equality does NOT hold in general (see later examples)!

To ensure strong duality we need to require additional properties:
A primal feasible $\bar{x}$ is strictly feasible for (P) if $\bar{x}$ lies in the interior of K,

$$\exists \rho > 0 \colon B_\rho(\bar{x}) := \{x \in \mathbb{R}^n \colon \|x - \bar{x}\| \leq \rho\} \subseteq K$$

If such an $\bar{x}$ exists, (P) is strictly feasible.
A dual feasible $(\bar{y}, \bar{z})$ is strictly feasible for (D), if $\bar{z}$ lies in the interior of K^* ($\exists \rho > 0 \colon B_\rho(\bar{z}) \subseteq K^*$). If such $(\bar{y}, \bar{z})$ exist, (D) is strictly feasible.

Theorem (Strong Duality)

If (P) is strictly feasible, the dual optimum $v(D)$ is attained.
If (D) is strictly feasible, the primal optimum $v(P)$ is attained.
In both cases there holds $v(P) = v(D)$.

Self-dual Cones

Here we mainly consider three special types of cones K:

- $K = \mathbb{R}^n_+$, the nonnegative orthant
- $K = \mathcal{Q}^n$, the second order/quadratic/Lorentz/ice cream cone
- $K = S^n_+$ the cone of positive semidefinite matrices

The detailed definitions of $\mathcal{Q}^n$ and S^n_+ will be given soon.

Self-dual Cones

Here we mainly consider three special types of cones K:

- $K = \mathbb{R}^n_+$, the nonnegative orthant
- $K = \mathcal{Q}^n$, the second order/quadratic/Lorentz/ice cream cone
- $K = S^n_+$ the cone of positive semidefinite matrices

The detailed definitions of $\mathcal{Q}^n$ and S^n_+ will be given soon.
The most important properties of these three are:

- They are self-dual, i.e., $K = K^*$. [+ homogeneous $\rightarrow$ symmetric]

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K \end{array}$$

Self-dual Cones

Here we mainly consider three special types of cones K:

- $K = \mathbb{R}^n_+$, the nonnegative orthant
- $K = \mathcal{Q}^n$, the second order/quadratic/Lorentz/ice cream cone
- $K = S^n_+$ the cone of positive semidefinite matrices

The detailed definitions of $\mathcal{Q}^n$ and S^n_+ will be given soon.
The most important properties of these three are:

- They are self-dual, i.e., $K = K^*$. [+ homogeneous $\rightarrow$ symmetric]

$$
(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \qquad\qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K \end{array}
$$

- Interior-point codes of good quality exist that allow the simultaneous use of all three, e.g. Mosek, SeDuMi and SDPT3.
- They allow to model and solve many important applications.

Self-dual Cones

Here we mainly consider three special types of cones K:

- $K = \mathbb{R}^n_+$, the nonnegative orthant
- $K = \mathcal{Q}^n$, the second order/quadratic/Lorentz/ice cream cone
- $K = S^n_+$ the cone of positive semidefinite matrices

The detailed definitions of $\mathcal{Q}^n$ and S^n_+ will be given soon.
The most important properties of these three are:

- They are self-dual, i.e., $K = K^*$. [+ homogeneous → symmetric]

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \in K \end{array} \qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \in K \end{array}$$

- Interior-point codes of good quality exist that allow the simultaneous use of all three, e.g. Mosek, SeDuMi and SDPT3.
- They allow to model and solve many important applications.

In applications K is typically composed of several subcones,

$$K = \mathbb{R}^n_+ \times \mathcal{Q}^{m_1} \times \cdots \times \mathcal{Q}^{m_k} \times S^{n_1}_+ \times \cdots \times S^{n_h}_+$$

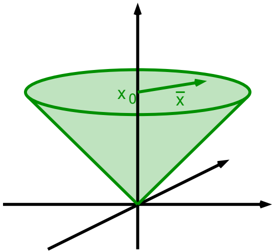
This will arise naturally and $K = K^*$ always holds for these combinations.

Contents

The Second Order Cone

The Second-Order-Cone (SOC)

$$\mathcal{Q}^n = \left\{ \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \in \mathbb{R}^{n+1} : x_0 \geq \|\bar{x}\| \right\}$$



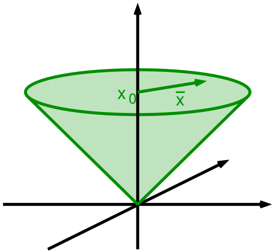
The Second Order Cone

The Second-Order-Cone (SOC)

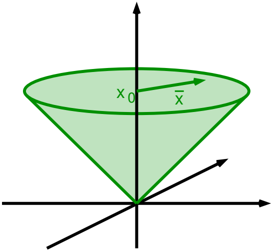
$$\mathcal{Q}^n = \left\{ \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \in \mathbb{R}^{n+1} \colon x_0 \geq \|\bar{x}\| \right\}$$

is a convex cone,
because for $x, y \in \mathcal{Q}^n$, $\alpha \geq 0$ we have
$\|\alpha(\bar{x} + \bar{y})\| \leq \alpha\|\bar{x}\| + \alpha\|\bar{y}\| \leq \alpha(x_0 + y_0)$.

$\mathcal{Q}^n$ is self-dual, $(\mathcal{Q}^n)^* = \mathcal{Q}^n$.



The Second Order Cone



The Second-Order-Cone (SOC)

$$\mathcal{Q}^n = \left\{ \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \in \mathbb{R}^{n+1} \colon x_0 \geq \|\bar{x}\| \right\}$$

is a convex cone,
because for $x, y \in \mathcal{Q}^n$, $\alpha \geq 0$ we have
$\|\alpha(\bar{x} + \bar{y})\| \leq \alpha\|\bar{x}\| + \alpha\|\bar{y}\| \leq \alpha(x_0 + y_0)$.

$\mathcal{Q}^n$ is self-dual, $(\mathcal{Q}^n)^* = \mathcal{Q}^n$.

Instead of $x \in \mathcal{Q}^n$ we often write $x \geq_{\mathcal{Q}} 0$. For $a, b \in \mathbb{R}^{n+1}$, $a \geq_{\mathcal{Q}} b$ is defined by $a - b \geq_{\mathcal{Q}} 0$, (or $a - b \in \mathcal{Q}^n$).

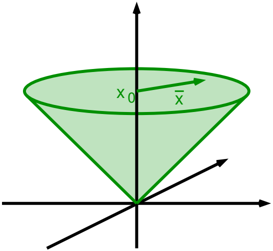
The Second Order Cone

The Second-Order-Cone (SOC)

$$\mathcal{Q}^n = \left\{ \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \in \mathbb{R}^{n+1} \colon x_0 \geq \|\bar{x}\| \right\}$$

is a convex cone,
because for $x, y \in \mathcal{Q}^n$, $\alpha \geq 0$ we have
$\|\alpha(\bar{x} + \bar{y})\| \leq \alpha\|\bar{x}\| + \alpha\|\bar{y}\| \leq \alpha(x_0 + y_0)$.

$\mathcal{Q}^n$ is self-dual, $(\mathcal{Q}^n)^* = \mathcal{Q}^n$.



Instead of $x \in \mathcal{Q}^n$ we often write $x \geq_{\mathcal{Q}} 0$. For $a, b \in \mathbb{R}^{n+1}$, $a \geq_{\mathcal{Q}} b$ is defined by $a - b \geq_{\mathcal{Q}} 0$, (or $a - b \in \mathcal{Q}^n$).

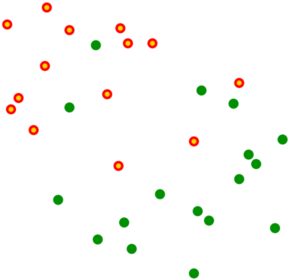
A linear program that only uses cones $\mathbb{R}^{n_i}_+$ and at least one $\mathcal{Q}^n$ is a second-order-cone program (SOCP in short).
An SOCP with just one $\mathcal{Q}^n$ reads

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq_{\mathcal{Q}} 0 \end{array} \qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & y \in \mathbb{R}^m, z \geq_{\mathcal{Q}} 0 \end{array}$$

[An SOCP with exactly one single SOC as here is solvable explicitly.]

Classification, Support-Vector

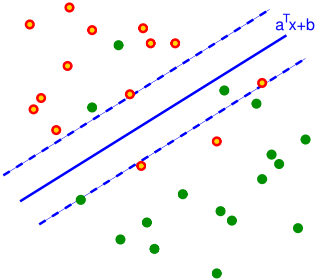
For data points in $\mathbb{R}^n$, that have or have not a certain property, we search for a hyperplane that separates the points according to this property as good as possible (goal: classify new points)



Classification, Support-Vector

For data points in $\mathbb{R}^n$, that have or have not a certain property, we search for a hyperplane that separates the points according to this property as good as possible (goal: classify new points)

Given two disjoint finite sets $G, R \subset \mathbb{R}^n$, find $a^T x + \beta$ (with variables a and β) with „preferably" $a^T x + \beta \geq 1$ for $x \in G$ and $a^T x + \beta \leq -1$ for $x \in R$.



Difficulties:

- For good separation $\|a\|$ should be small.
- What to do, if classification failures cannot be avoided?

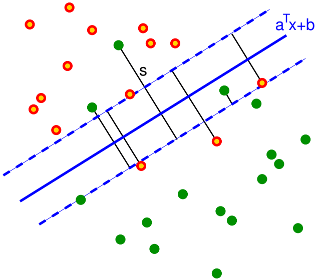
Classification, Support-Vector

For data points in $\mathbb{R}^n$, that have or have not a certain property, we search for a hyperplane that separates the points according to this property as good as possible (goal: classify new points)

Given two disjoint finite sets $G, R \subset \mathbb{R}^n$, find $a^T x + \beta$ (with variables a and β) with „preferably“ $a^T x + \beta \geq 1$ for $x \in G$ and $a^T x + \beta \leq -1$ for $x \in R$.

Difficulties:

- For good separation $\|a\|$ should be small.
- What to do, if classification failures cannot be avoided?



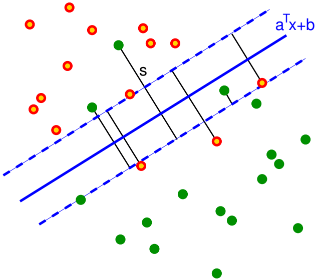
One approach: Minimise simultaneously $\|a\|$ and the sum of violations of the inequality constraints, scalarised by parameter $\gamma > 0$,

$$
\begin{array}{lll}
\min & \|a\| + \gamma \sum_{x \in G \cup R} s_x & \\
\text{s.t.} & x^T a - \beta \geq 1 - s_x & x \in G \\
& x^T a - \beta \leq s_x - 1 & x \in R \\
& a \in \mathbb{R}^n, \beta \in \mathbb{R}, s \in \mathbb{R}_+^{G \cup R} &
\end{array}
$$

Classification, Support-Vector

For data points in $\mathbb{R}^n$, that have or have not a certain property, we search for a hyperplane that separates the points according to this property as good as possible (goal: classify new points)

Given two disjoint finite sets $G, R \subset \mathbb{R}^n$, find $a^T x + \beta$ (with variables a and β) with „preferably" $a^T x + \beta \geq 1$ for $x \in G$ and $a^T x + \beta \leq -1$ for $x \in R$.



Difficulties:

- For good separation $\|a\|$ should be small.
- What to do, if classification failures cannot be avoided?

One approach: Minimise simultaneously $\|a\|$ and the sum of violations of the inequality constraints, scalarised by parameter $\gamma > 0$,

$$\begin{array}{lll} \min & \|a\| + \gamma \sum_{x \in G \cup R} s_x & \\ \text{s.t.} & x^T a - \beta \geq 1 - s_x & x \in G \\ & x^T a - \beta \leq s_x - 1 & x \in R \\ & a \in \mathbb{R}^n, \beta \in \mathbb{R}, s \in \mathbb{R}_+^{G \cup R} & \end{array} \quad \to \quad \begin{array}{lll} \min & a_0 + \gamma\, \mathbf{1}^T s & \\ \text{s.t.} & x^T a - \beta \geq 1 - s_x & x \in G \\ & x^T a - \beta \leq s_x - 1 & x \in R \\ & \begin{bmatrix} a_0 \\ a \end{bmatrix} \geq_{\mathcal{Q}} 0, \beta \in \mathbb{R}, s \geq 0 & \end{array}$$

The Markowitz Model

In the Markowitz model of portfolio optimisation, a given budget is to be invested with given expected profit so that risk is minimised.

The Markowitz Model

In the Markowitz model of portfolio optimisation, a given budget is to be invested with given expected profit so that risk is minimised.

$x \in \mathbb{R}^n_+$ with $\mathbf{1}^T x = 1$ represents the fraction of the budget invested into stock $1, \ldots, n$. The profit g per investment is a random variable with expectation $\bar{g} \in \mathbb{R}^n$ and covariance matrix $G \in S^n_+$ ($n \times n$, positive semidefinite). Let $s \in \mathbb{R}$ be a given profit threshold. As a risk measure the Markowitz model uses $x^T Gx$. [better measures exist]

$$\begin{array}{ll} \min & x^T G x \\ \text{s.t.} & \bar{g}^T x \geq s \\ & \mathbf{1}^T x = 1 \\ & x \in \mathbb{R}^n_+ \end{array}$$

The Markowitz Model

In the Markowitz model of portfolio optimisation, a given budget is to be invested with given expected profit so that risk is minimised.

$x \in \mathbb{R}^n_+$ with $\mathbf{1}^T x = 1$ represents the fraction of the budget invested into stock $1, \ldots, n$. The profit g per investment is a random variable with expectation $\bar{g} \in \mathbb{R}^n$ and covariance matrix $G \in S^n_+$ ($n \times n$, positive semidefinite). Let $s \in \mathbb{R}$ be a given profit threshold. As a risk measure the Markowitz model uses $x^T G x$. [better measures exist]

$$\begin{array}{ll} \min & x^T G x \\ \text{s.t.} & \bar{g}^T x \geq s \\ & \mathbf{1}^T x = 1 \\ & x \in \mathbb{R}^n_+ \end{array}$$

Because G is positive semidefinite this is a convex quadratic problem. [The two criteria profit against risk are now implemented by a constraint on one of the criteria.]

How to model this as an SOCP?

Quadratic Constraints with SOCP

Let $Q \in S_+^n$ be positive semidefinite, $q \in \mathbb{R}^n$, $\delta \in \mathbb{R}$. The convex quadratic constraint

$$x^T Qx + q^T x + \delta \leq 0$$

may be represented as an SOCP-constraint by (factor $Q = LL^T$)

$$\left\| \begin{array}{c} L^T x \\ \frac{1+(q^T x+\delta)}{2} \end{array} \right\| \leq \frac{1-(q^T x + \delta)}{2}$$

(proof: square both sides).

Quadratic Constraints with SOCP

Let $Q \in S_+^n$ be positive semidefinite, $q \in \mathbb{R}^n$, $\delta \in \mathbb{R}$. The convex quadratic constraint

$$x^T Q x + q^T x + \delta \leq 0$$

may be represented as an SOCP-constraint by (factor $Q = LL^T$)

$$\left\| \begin{array}{c} L^T x \\ \frac{1+(q^T x+\delta)}{2} \end{array} \right\| \leq \frac{1-(q^T x + \delta)}{2}$$

(proof: square both sides). $\rightarrow$ constrained least squares problems!

Quadratic Constraints with SOCP

Let $Q \in S^n_+$ be positive semidefinite, $q \in \mathbb{R}^n$, $\delta \in \mathbb{R}$. The convex quadratic constraint

$$x^T Qx + q^T x + \delta \leq 0$$

may be represented as an SOCP-constraint by (factor $Q = LL^T$)

$$\left\| \begin{array}{c} L^T x \\ \frac{1+(q^T x+\delta)}{2} \end{array} \right\| \leq \frac{1-(q^T x + \delta)}{2}$$

(proof: square both sides). $\rightarrow$ constrained least squares problems!

For the Markowitz model just use $x_0 \geq \|L^T x\|$ with $G = LL^T$, then

$$\begin{array}{ll} \min & x_0 \\ \text{s.t.} & \bar{x} = L^T x \\ & \bar{g}^T x \geq s \\ & \mathbf{1}^T x = 1 \\ & \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \geq_{\mathcal{Q}} 0, x \geq 0 \end{array}$$

Probabilistic Constraints, Chance Constraint

Assume profit g is distributed normally with mean $\bar{g}$ and variance G. In addition to $\bar{g}^T x \geq s$ we now also require with probability at least $\eta \in (0, 1)$ that the profit is above a threshold value $\underline{s} < s$,

$$\mathbb{P}(g^T x \geq \underline{s}) \geq \eta$$

Probabilistic Constraints, Chance Constraint

Assume profit g is distributed normally with mean $\bar{g}$ and variance G. In addition to $\bar{g}^T x \geq s$ we now also require with probability at least $\eta \in (0, 1)$ that the profit is above a threshold value $\underline{s} < s$,

$$\mathbb{P}(g^T x \geq \underline{s}) \geq \eta \qquad \rightarrow \qquad \begin{array}{ll} \min & x^T G x \\ \text{s.t.} & \bar{g}^T x \geq s \\ & \mathbb{P}(g^T x \geq \underline{s}) \geq \eta \\ & \mathbf{1}^T x = 1 \\ & x \in \mathbb{R}^n_+ \end{array}$$

This is modelled using a technique of robust optimisation:
$g^T x \geq \underline{s}$ is interpreted as an inequality with uncertain coefficients.

Linear Constraints with Uncertain Coefficients

If the coefficients of inequality $a^T x \le b$ are only known to lie inside the ellipsoid $a \in \{\bar{a} + Hu \colon \|u\| \le 1\}$ for given $H \in S^n_+$ (pos. semidef.) and if x has to satisfy this inequality for all such a, this requires

$$\max_{\|u\|=1} \bar{a}^T x + u^T Hx = \bar{a}^T x + \|Hx\| \le b$$

Linear Constraints with Uncertain Coefficients

If the coefficients of inequality $a^T x \le b$ are only known to lie inside the ellipsoid $a \in \{\bar{a} + Hu\colon \|u\| \le 1\}$ for given $H \in S^n_+$ (pos. semidef.) and if x has to satisfy this inequality for all such a, this requires

$$\max_{\|u\|=1} \bar{a}^T x + u^T Hx = \bar{a}^T x + \|Hx\| \le b$$

The latter inequality may be represented via the SOC constraint

$$\begin{array}{rcl} \xi_0 & = & b - \bar{a}^T x \\ \xi & = & Hx \\ \begin{bmatrix} \xi_0 \\ \xi \end{bmatrix} & \ge_{\mathcal{Q}} & 0 \end{array}$$

Linear Constraints with Uncertain Coefficients

If the coefficients of inequality $a^T x \le b$ are only known to lie inside the ellipsoid $a \in \{\bar{a} + Hu \colon \|u\| \le 1\}$ for given $H \in S^n_+$ (pos. semidef.) and if x has to satisfy this inequality for all such a, this requires

$$\max_{\|u\|=1} \bar{a}^T x + u^T Hx = \bar{a}^T x + \|Hx\| \le b$$

The latter inequality may be represented via the SOC constraint

$$\begin{aligned} \xi_0 &= b - \bar{a}^T x \\ \xi &= Hx \\ \begin{bmatrix} \xi_0 \\ \xi \end{bmatrix} &\ge_{\mathcal{Q}} 0 \end{aligned}$$

For its probabilistic interpretation let g be distributed normally around $\bar{g}$ with covariance matrix $G = H^2$ and suppose $g^T x \ge \underline{s}$ needs to be satisfied with probability $0 < \eta < 1$. Then $\mathbb{P}(g^T x \ge \underline{s}) \ge \eta$ corresponds to the constraint $-\bar{g}^T x + \Phi^{-1}(\eta)\|Hx\| \le -\underline{s}$. [$\Phi$ ... normal distribution]

Contents

Positive Semidefinite Matrices

A symmetric matrix $A \in S^n := \{A \in \mathbb{R}^{n \times n} \colon A = A^T\}$ is positive semidefinite, if $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; we write $A \in S^n_+$ or $A \succeq 0$.

It is positive definite ($A \in S^n_{++}$, $A \succ 0$), if $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
[For $A \succeq 0$ ($\succ 0$) and $J \subseteq \{1, \ldots, n\}$ we have $A_{J,J} \succeq 0$ ($\succ 0$).]

Positive Semidefinite Matrices

A symmetric matrix $A \in S^n := \{A \in \mathbb{R}^{n\times n} \colon A = A^T\}$ is positive semidefinite, if $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; we write $A \in S^n_+$ or $A \succeq 0$.

It is positive definite ($A \in S^n_{++}$, $A \succ 0$), if $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
[For $A \succeq 0$ ($\succ 0$) and $J \subseteq \{1, \dots, n\}$ we have $A_{J,J} \succeq 0$ ($\succ 0$).]

$\lambda \in \mathbb{R}$ is an eigenvalue and $v \in \mathbb{R}^n \setminus \{0\}$ an eigenvector of A, if $Av = \lambda v$.
For each $A \in S^n$ there exist an eigenvalue decomposition $A = P \Lambda P^T$ with real $\Lambda = \mathrm{Diag}(\lambda_1, \dots, \lambda_n)$ and orthogonal $P \in \mathbb{R}^{n\times n}$ (i.e., $P^T P = I$).
For $P = [v_1, \dots, v_n]$ we get $A = P \Lambda P^T = \sum_{i=1}^n \lambda_i v_i v_i^T$.

Positive Semidefinite Matrices

A symmetric matrix $A \in S^n := \{A \in \mathbb{R}^{n\times n} \colon A = A^T\}$ is positive semidefinite, if $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; we write $A \in S^n_+$ or $A \succeq 0$.

It is positive definite ($A \in S^n_{++}$, $A \succ 0$), if $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
[For $A \succeq 0$ ($\succ 0$) and $J \subseteq \{1,\dots,n\}$ we have $A_{J,J} \succeq 0$ ($\succ 0$).]

$\lambda \in \mathbb{R}$ is an eigenvalue and $v \in \mathbb{R}^n \setminus \{0\}$ an eigenvector of A, if $Av = \lambda v$.
For each $A \in S^n$ there exist an eigenvalue decomposition $A = P\Lambda P^T$ with real $\Lambda = \operatorname{Diag}(\lambda_1,\dots,\lambda_n)$ and orthogonal $P \in \mathbb{R}^{n\times n}$ (i.e., $P^T P = I$).
For $P = [v_1,\dots,v_n]$ we get $A = P\Lambda P^T = \sum_{i=1}^n \lambda_i v_i v_i^T$.

For $A, B \in S^n$ we use the Frobenius inner product

$$\langle A, B\rangle := \sum_{1\leq i,j\leq n} A_{ij} B_{ij} \qquad [= \operatorname{vec}(A)^T \operatorname{vec}(B),\ \text{frequently } A \bullet B]$$

Positive Semidefinite Matrices

A symmetric matrix $A \in S^n := \{A \in \mathbb{R}^{n\times n} : A = A^T\}$ is positive semidefinite, if $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; we write $A \in S^n_+$ or $A \succeq 0$.

It is positive definite ($A \in S^n_{++}$, $A \succ 0$), if $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
[For $A \succeq 0$ ($\succ 0$) and $J \subseteq \{1, \dots, n\}$ we have $A_{J,J} \succeq 0$ ($\succ 0$).]

$\lambda \in \mathbb{R}$ is an eigenvalue and $v \in \mathbb{R}^n \setminus \{0\}$ an eigenvector of A, if $Av = \lambda v$.
For each $A \in S^n$ there exist an eigenvalue decomposition $A = P\Lambda P^T$ with real $\Lambda = \mathrm{Diag}(\lambda_1, \dots, \lambda_n)$ and orthogonal $P \in \mathbb{R}^{n\times n}$ (i.e., $P^T P = I$).
For $P = [v_1, \dots, v_n]$ we get $A = P\Lambda P^T = \sum_{i=1}^n \lambda_i v_i v_i^T$.

For $A, B \in S^n$ we use the Frobenius inner product

$$\langle A, B\rangle := \sum_{1\leq i,j\leq n} A_{ij} B_{ij} \qquad [= \mathrm{vec}(A)^T \mathrm{vec}(B),\ \text{frequently } A \bullet B]$$

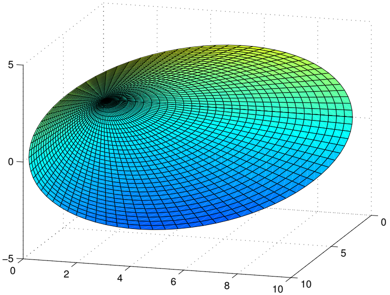
Theorem

For $A \in S^n$ the following are equivalent:

- $A \succeq 0$,
- $\lambda_i(A) \geq 0$, $i = 1, \dots, n$, [$\Rightarrow \det(A) \geq 0$]
- $A = C^T C$ for some $C \in \mathbb{R}^{k\times n}$, [there holds: $\mathrm{rank}(A) = \mathrm{rank}(C)$]
- $\langle A, B\rangle \geq 0 \quad \forall B \succeq 0$.

The Cone of Positive Semidefinite Matrices

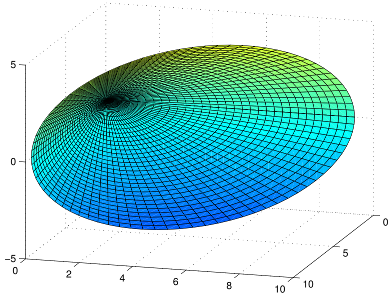
The positive semidefinite matrices S_+^n form a convex cone, because for $X, Y \in S_+^n$, $\alpha \geq 0$ and all $v \in \mathbb{R}^n$
$v^T(\alpha(X+Y))v = \alpha(v^T X v + v^T Y v) \geq 0.$



The Cone of Positive Semidefinite Matrices

The positive semidefinite matrices S_+^n form a convex cone, because for $X, Y \in S_+^n$, $\alpha \geq 0$ and all $v \in \mathbb{R}^n$
$v^T(\alpha(X+Y))v = \alpha(v^T X v + v^T Y v) \geq 0$.

$A \in S_+^n \Leftrightarrow \langle A, B \rangle \geq 0 \; \forall B \succeq 0$ implies S_+^n is self-dual, $(S_+^n)^* = S_+^n$.

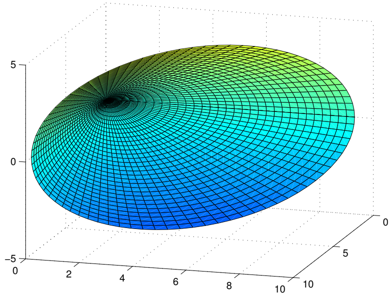


The Cone of Positive Semidefinite Matrices

The positive semidefinite matrices S_+^n form a convex cone, because for $X, Y \in S_+^n$, $\alpha \geq 0$ and all $v \in \mathbb{R}^n$
$v^T(\alpha(X+Y))v = \alpha(v^T X v + v^T Y v) \geq 0.$

$A \in S_+^n \Leftrightarrow \langle A, B \rangle \geq 0 \; \forall B \succeq 0$ implies S_+^n is self-dual, $(S_+^n)^* = S_+^n$.

right image: $S_+^2 = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$

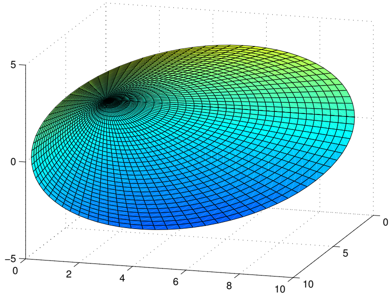


The Cone of Positive Semidefinite Matrices

The positive semidefinite matrices S^n_+ form a convex cone, because
for $X, Y \in S^n_+$, $\alpha \geq 0$ and all $v \in \mathbb{R}^n$
$v^T(\alpha(X+Y))v = \alpha(v^T X v + v^T Y v) \geq 0.$

$A \in S^n_+ \Leftrightarrow \langle A, B \rangle \geq 0 \ \forall B \succeq 0$ implies
S^n_+ is self-dual, $(S^n_+)^* = S^n_+$.

right image: $S^2_+ = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$



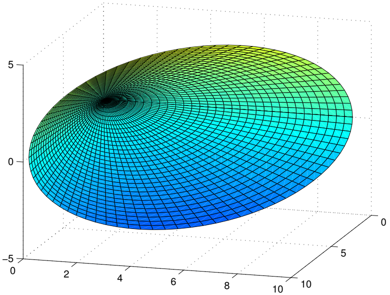
- If $R \in \mathbb{R}^{n \times n}$ is regular (=invertible), then $X \succeq 0 \quad \Leftrightarrow \quad R^T X R \succeq 0.$

The Cone of Positive Semidefinite Matrices

The positive semidefinite matrices S_+^n form a convex cone, because for $X, Y \in S_+^n$, $\alpha \geq 0$ and all $v \in \mathbb{R}^n$
$v^T(\alpha(X+Y))v = \alpha(v^TXv + v^TYv) \geq 0.$

$A \in S_+^n \Leftrightarrow \langle A, B\rangle \geq 0 \ \forall B \succeq 0$ implies S_+^n is self-dual, $(S_+^n)^* = S_+^n$.

right image: $S_+^2 = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$



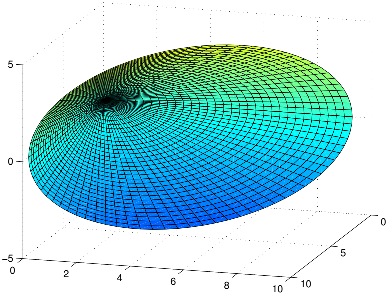
- If $R \in \mathbb{R}^{n\times n}$ is regular (=invertible), then $X \succeq 0 \quad \Leftrightarrow \quad R^TXR \succeq 0.$
- For $A - B \succeq 0$ we also write $A \succeq B$.

The Cone of Positive Semidefinite Matrices

The positive semidefinite matrices S^n_+ form a convex cone, because for $X, Y \in S^n_+$, $\alpha \geq 0$ and all $v \in \mathbb{R}^n$
$v^T(\alpha(X+Y))v = \alpha(v^TXv + v^TYv) \geq 0.$

$A \in S^n_+ \Leftrightarrow \langle A, B\rangle \geq 0 \; \forall B \succeq 0$ implies S^n_+ is self-dual, $(S^n_+)^* = S^n_+$.

right image: $S^2_+ = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$



- If $R \in \mathbb{R}^{n\times n}$ is regular (=invertible), then $\quad X \succeq 0 \quad \Leftrightarrow \quad R^TXR \succeq 0.$
- For $A - B \succeq 0$ we also write $A \succeq B$.

Frequently used in formulating applications as semidefinite programs:

Theorem (Schur Complement)

For $A \in S^m_{++}$, $C \in S^n$ and $B \in \mathbb{R}^{m\times n}$ there holds

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \quad (\textit{resp.} \succ 0) \quad \Longleftrightarrow \quad C \succeq B^TA^{-1}B \quad (\textit{resp.} \succ 0)$$

LP $\leftrightarrow$ Semidefinite Programs (SDP)

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad\qquad \begin{aligned} \min \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & X \succeq 0 \end{aligned}$$

$$x \in \mathbb{R}^n_+ \qquad\qquad X \in S^n_+$$

$$c^T x = \sum_i c_i x_i \qquad\qquad \langle C, X \rangle = \sum_{i,j} C_{ij} X_{ij}$$

$$Ax = \begin{pmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{pmatrix} \qquad\qquad \mathcal{A}X = \begin{pmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{pmatrix}$$

$$A^T y = \sum_i a_i y_i \qquad\qquad \mathcal{A}^T y = \sum_i A_i y_i$$

$$\begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & A^T y + z = c \\ & y \in \mathbb{R}^m, z \geq 0 \end{aligned} \qquad\qquad \begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & \mathcal{A}^T y + Z = C \\ & y \in \mathbb{R}^m, Z \succeq 0 \end{aligned}$$

Semidefinite Programs (SDP) in Normal Form

$$
(P) \quad \begin{array}{ll} \min & \langle C, X \rangle \\ \text{s.t.} & \mathcal{A}X = b \\ & X \succeq 0 \end{array}
\qquad\qquad
(D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & \mathcal{A}^T y + Z = C \\ & y \in \mathbb{R}^n, Z \succeq 0 \end{array}
$$

If one of both is strictly feasible there holds $v(P) = v(D)$.

Semidefinite Programs (SDP) in Normal Form

$$
(P)\quad \begin{array}{ll} \min & \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}X = b \\ & X \succeq 0 \end{array}
\qquad\qquad
(D)\quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & \mathcal{A}^T y + Z = C \\ & y \in \mathbb{R}^n, Z \succeq 0 \end{array}
$$

If one of both is strictly feasible there holds $v(P) = v(D)$.

In applications several cones $X_i \succeq 0$ may appear, for theory one suffices:

$$
X_1 \succeq 0,\ X_2 \succeq 0,\ \ldots,\ X_k \succeq 0 \quad \Leftrightarrow \quad \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & X_k \end{bmatrix} \succeq 0.
$$

$\Rightarrow$ Semidefinite Optimisation contains Linear Optimisation ($X_i \in S^1_+$).

Semidefinite Programs (SDP) in Normal Form

$$(P) \quad \begin{array}{ll} \min & \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}X = b \\ & X \succeq 0 \end{array} \qquad (D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & \mathcal{A}^T y + Z = C \\ & y \in \mathbb{R}^n, Z \succeq 0 \end{array}$$

If one of both is strictly feasible there holds $v(P) = v(D)$.

In applications several cones $X_i \succeq 0$ may appear, for theory one suffices:

$$X_1 \succeq 0,\ X_2 \succeq 0,\ \ldots,\ X_k \succeq 0 \quad \Leftrightarrow \quad \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & X_k \end{bmatrix} \succeq 0.$$

$\Rightarrow$ Semidefinite Optimisation contains Linear Optimisation ($X_i \in S^1_+$).

Semidefinite Optimisation also allows to formulate SOC-constraints:

$$\begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \geq_{\mathcal{Q}} 0 \quad \overset{x_0 \geq 0}{\Leftrightarrow} \quad x_0 \geq \frac{1}{x_0} \bar{x}^T I \bar{x} \quad \overset{\text{Schur}}{\Leftrightarrow} \quad \begin{bmatrix} x_0 & \bar{x}^T \\ \bar{x} & x_0 I \end{bmatrix} \succeq 0.$$

[for $x_0 = 0$ this is checked directly]

Illustration: $X \in S^2_+$ intersected with $\langle A, X \rangle = \beta$

$X = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$

$\Rightarrow x \geq 0,\ y \geq 0,\ xy - z^2 \geq 0$

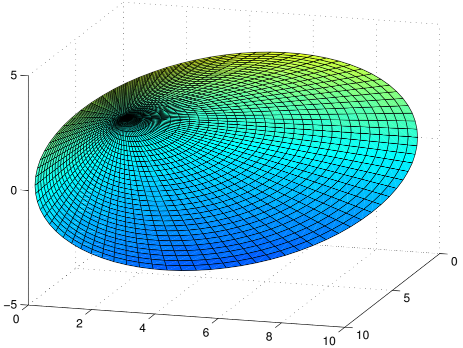


Illustration: $X \in S^2_+$ intersected with $\langle A, X \rangle = \beta$

$$X = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$$

$$\Rightarrow x \geq 0,\ y \geq 0,\ xy - z^2 \geq 0$$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \beta = 0 \to z = 0$$

$$\Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2_+, \text{ like LP}$$

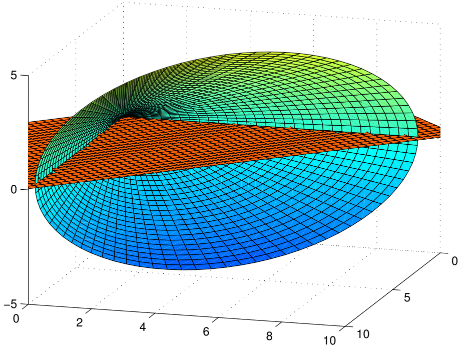


Illustration: $X \in S_+^2$ intersected with $\langle A, X\rangle = \beta$

$X = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$

$\Rightarrow x \geq 0,\ y \geq 0,\ xy - z^2 \geq 0$

$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \beta = 0 \to z = 0$

$\Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}_+^2$, like LP

$A \succ 0, \beta > 0$

$\to$ "general elliptic case"

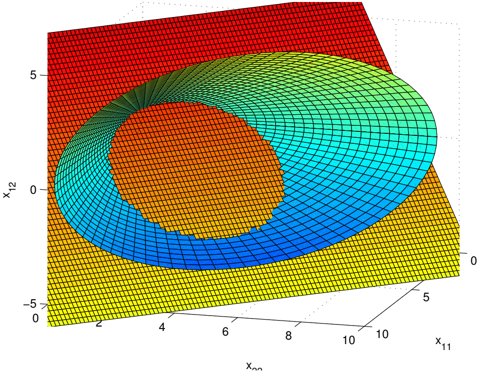


Illustration: $X \in S^2_+$ intersected with $\langle A, X \rangle = \beta$

$$X = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$$

$\Rightarrow x \geq 0,\ y \geq 0,\ xy - z^2 \geq 0$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \beta = 0 \rightarrow z = 0$$

$\Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2_+$, like LP

$A \succ 0, \beta > 0$

$\rightarrow$ "general elliptic case"

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \beta < 0 \rightarrow z = \tfrac{1}{2}\beta$$

$\Leftrightarrow xy \geq \frac{1}{4}\beta^2$, hyperbola

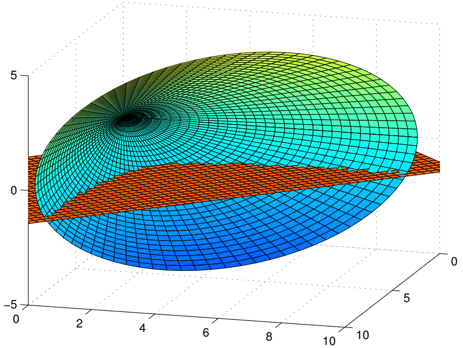


Illustration: $X \in S^2_+$ intersected with $\langle A, X \rangle = \beta$

$X = \begin{bmatrix} x & z \\ z & y \end{bmatrix} \succeq 0.$

$\Rightarrow x \geq 0,\ y \geq 0,\ xy - z^2 \geq 0$

$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \beta = 0 \rightarrow z = 0$

$\Leftrightarrow \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2_+$, like LP

$A \succ 0, \beta > 0$

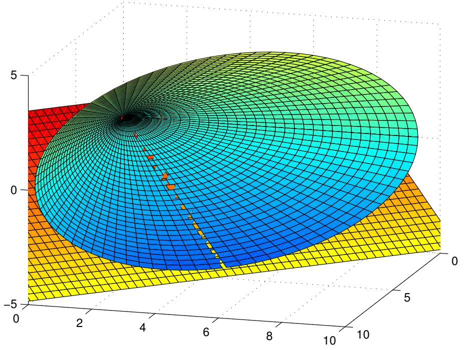
$\rightarrow$ "general elliptic case"

$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \beta < 0 \rightarrow z = \frac{1}{2}\beta$

$\Leftrightarrow xy \geq \frac{1}{4}\beta^2$, hyperbola

$A = vv^T, \beta = 0 \rightarrow v$ Evec to $\lambda_1 = 0$

boundary points, numerically difficult!



Linear Matrix Inequalities (LMI)

A constraint of the form

$$y_1 A_1 + y_2 A_2 + \cdots + y_m A_m \preceq C$$

with $A_i, C \in S^n$ is a Linear Matrix Inequality.

Feasible $y \in \mathbb{R}^m$ are SDP-representable, $\{y \in \mathbb{R}^m \colon \mathcal{A}^T y + Z = C, Z \succeq 0\}$.

Linear Matrix Inequalities (LMI)

A constraint of the form

$$y_1 A_1 + y_2 A_2 + \cdots + y_m A_m \preceq C$$

with $A_i, C \in S^n$ is a Linear Matrix Inequality.

Feasible $y \in \mathbb{R}^m$ are SDP-representable, $\{y \in \mathbb{R}^m \colon \mathcal{A}^T y + Z = C, Z \succeq 0\}$.

Ex.: the Lyapunov inequality requires for fixed $P \in \mathbb{R}^{n \times n}$

$$P^T X + XP \prec 0, \quad X \succ 0.$$

In LMI-representation write $y = [x_{11}, x_{12}, \ldots, x_{1n}, x_{22}, x_{23}, \ldots, x_{nn}]^T$, but it is cumbersome/useless to list the A_i for this constraint. It is better to exploit the structure directly within SDP.

Linear Matrix Inequalities (LMI)

A constraint of the form

$$y_1 A_1 + y_2 A_2 + \cdots + y_m A_m \preceq C$$

with $A_i, C \in S^n$ is a Linear Matrix Inequality.

Feasible $y \in \mathbb{R}^m$ are SDP-representable, $\{y \in \mathbb{R}^m : \mathcal{A}^T y + Z = C, Z \succeq 0\}$.

Ex.: the Lyapunov inequality requires for fixed $P \in \mathbb{R}^{n \times n}$

$$P^T X + XP \prec 0, \quad X \succ 0.$$

In LMI-representation write $y = [x_{11}, x_{12}, \ldots, x_{1n}, x_{22}, x_{23}, \ldots, x_{nn}]^T$, but it is cumbersome/useless to list the A_i for this constraint. It is better to exploit the structure directly within SDP.

For recognising LMIs it suffices to ensure that all matrices depend linearly on the corresponding variables:

the matrix multiplication $P^T X$ (XP resp.) is linear in X.

Applications of Semidefinite Optimisation

- optimal control
- eigenvalue optimisation
- experiment design in statistics
- combinatorial optimisation
- global optimisation over polynomials
- moment problems in probability theory
- signal processing
- robust truss topology design
- free material design
- robust optimisation
- optimisation (trust-region subproblems, quadratic relaxations)

Robust Stability of Dynamical Systems

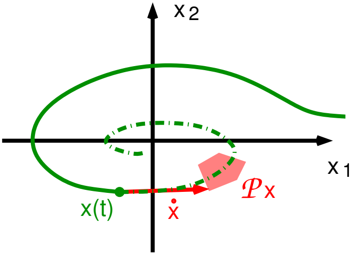
Given a (homogenous linear) dynamical system with uncertain data,

$$(DS) \quad \dot{x} = P(t)x(t) \quad \text{with } P(t) \in \mathcal{P} := \text{conv}\{P_1, \ldots, P_k\} \subset \mathbb{R}^{n\times n},$$

where
- $x(t)$ . . . state of the system at time t.
- $\dot{x} := \frac{d}{dt}x(t)$ . . . (infinitesimal) change of $x(\cdot)$
- $P(t)$ . . . uncertain transition matrix at time t,

(DS) is stable if $x(t) \to 0$ for $t \to \infty$ and arbitrary $P(t) \in \mathcal{P}$.

[In optimal control, $\mathcal{P}$ would comprise the possible effects of imperfect implementations of the control. Does it do its job anyways even with tiny mistakes?]



Robust Stability of Dynamical Systems

Given a (homogenous linear) dynamical system with uncertain data,

$$(DS) \quad \dot{x} = P(t)x(t) \quad \text{with } P(t) \in \mathcal{P} := \text{conv}\{P_1, \ldots, P_k\} \subset \mathbb{R}^{n\times n},$$

where
- $x(t)$ . . . state of the system at time t.
- $\dot{x} := \frac{d}{dt}x(t)$ . . . (infinitesimal) change of $x(\cdot)$
- $P(t)$ . . . uncertain transition matrix at time t,

(DS) is stable if $x(t) \to 0$ for $t \to \infty$ and arbitrary $P(t) \in \mathcal{P}$.

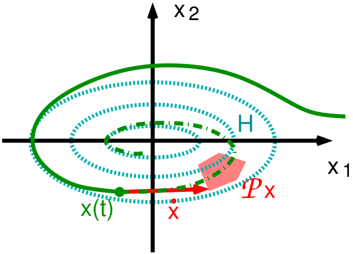
---

sufficient condition: there is a norm

$$\|x\|_H := \sqrt{x^T H x} \quad \text{with } H \succ 0$$

so that $\frac{d}{dt}\|x(t)\|_H^2 < 0$ for all trajectories

(the system is quadratically stable, $x^T Hx$ a quadratic Lyapunov Function).



Robust Lyapunov Stability by SDP

$$(DS) \quad \dot{x} = P(t)x(t) \quad \text{with } P(t) \in \mathcal{P} := \text{conv}\{P_1, \ldots, P_k\} \subset \mathbb{R}^{n \times n}$$

Find $H \succ 0$ with $\frac{d}{dt}\|x(t)\|_H^2 < 0$.

Robust Lyapunov Stability by SDP

$$(DS) \quad \dot{x} = P(t)x(t) \quad \text{with } P(t) \in \mathcal{P} := \text{conv}\{P_1, \ldots, P_k\} \subset \mathbb{R}^{n \times n}$$

Find $H \succ 0$ with $\frac{d}{dt}\|x(t)\|_H^2 < 0$.

$$\frac{d}{dt}\|x(t)\|_H^2 \;=\; \frac{d}{dt}x^T H x \;=\; \dot{x}^T H x + x^T H \dot{x} \;=\; x^T(P(t)^T H + HP(t))x$$

Robust Lyapunov Stability by SDP

$$(DS) \quad \dot{x} = P(t)x(t) \quad \text{with } P(t) \in \mathcal{P} := \text{conv}\{P_1, \ldots, P_k\} \subset \mathbb{R}^{n \times n}$$

Find $H \succ 0$ with $\frac{d}{dt}\|x(t)\|_H^2 < 0$.

$$\frac{d}{dt}\|x(t)\|_H^2 \;=\; \frac{d}{dt}x^T Hx \;=\; \dot{x}^T Hx + x^T H\dot{x} \;=\; x^T(P(t)^T H + HP(t))x$$

If $A := P^T H + HP \prec 0$ (negative definite), then $v^T Av < 0 \; \forall v \in \mathbb{R}^n \setminus \{0\}$.

Robust Lyapunov Stability by SDP

$$(DS) \quad \dot{x} = P(t)x(t) \quad \text{with } P(t) \in \mathcal{P} := \text{conv}\{P_1, \ldots, P_k\} \subset \mathbb{R}^{n \times n}$$

Find $H \succ 0$ with $\frac{d}{dt}\|x(t)\|_H^2 < 0$.

$$\frac{d}{dt}\|x(t)\|_H^2 \;=\; \frac{d}{dt}x^T Hx \;=\; \dot{x}^T Hx + x^T H\dot{x} \;=\; x^T(P(t)^T H + HP(t))x$$

If $A := P^T H + HP \prec 0$ (negative definite), then $v^T Av < 0 \; \forall v \in \mathbb{R}^n \setminus \{0\}$.
$\Rightarrow$ the system is quadratically stable, if

$$H \succ 0, \qquad P_i^T H + HP_i \prec 0 \quad \text{for } i = 1, \ldots, k$$

has feasible solutions, because for any such H also each convex combination $P \in \mathcal{P}$ satisfies the condition $P^T H + HP \prec 0$.

Robust Lyapunov Stability by SDP

$$(DS) \quad \dot{x} = P(t)x(t) \quad \text{with } P(t) \in \mathcal{P} := \text{conv}\{P_1, \dots, P_k\} \subset \mathbb{R}^{n\times n}$$

Find $H \succ 0$ with $\frac{d}{dt}\|x(t)\|_H^2 < 0$.

$$\frac{d}{dt}\|x(t)\|_H^2 \;=\; \frac{d}{dt}x^T Hx \;=\; \dot{x}^T Hx + x^T H\dot{x} \;=\; x^T(P(t)^T H + HP(t))x$$

If $A := P^T H + HP \prec 0$ (negative definite), then $v^T Av < 0\ \forall v \in \mathbb{R}^n \setminus \{0\}$.
$\Rightarrow$ the system is quadratically stable, if

$$H \succ 0, \qquad P_i^T H + HP_i \prec 0 \quad \text{for } i = 1, \dots, k$$

has feasible solutions, because for any such H also each convex combination $P \in \mathcal{P}$ satisfies the condition $P^T H + HP \prec 0$.

Search for H by eigenvalue optimisation:

$$\max\ \lambda \text{ s.t. } H \succeq \lambda I, \quad P_i^T H + HP_i \preceq -\lambda I \quad \text{for } i = 1, \dots, k.$$

SDP and Eigenvalue Optimisation

For $A \in S^n$ let $\lambda_{\min}(A) := \lambda_1(A) \leq \cdots \leq \lambda_n(A) =: \lambda_{\max}(A)$.
There holds $\lambda_i(A + y_0 I) = \lambda_i(A) + y_0$ for $i = 1, \ldots, n$ and $y_0 \in \mathbb{R}$.

SDP and Eigenvalue Optimisation

For $A \in S^n$ let $\lambda_{\min}(A) := \lambda_1(A) \leq \cdots \leq \lambda_n(A) =: \lambda_{\max}(A)$.
There holds $\lambda_i(A + y_0 I) = \lambda_i(A) + y_0$ for $i = 1, \ldots, n$ and $y_0 \in \mathbb{R}$.

In optimal control stability of a system is ensured for control parameters $y \in \mathbb{R}^m$ if the control dependent system matrix $A(y)$ satisfies $\lambda_{\max}(A(y)) < 0$.

For affine $A(y)$, e.g. $A(y) := C - \sum_{i=1}^m y_i A_i$ with $C, A_i \in S^n$, this leads to

$$\min_{y \in \mathbb{R}^m} \lambda_{\max}(C - \mathcal{A}^T y)$$

SDP and Eigenvalue Optimisation

For $A \in S^n$ let $\lambda_{\min}(A) := \lambda_1(A) \leq \cdots \leq \lambda_n(A) =: \lambda_{\max}(A)$.
There holds $\lambda_i(A + y_0 I) = \lambda_i(A) + y_0$ for $i = 1, \dots, n$ and $y_0 \in \mathbb{R}$.

In optimal control stability of a system is ensured for control parameters $y \in \mathbb{R}^m$ if the control dependent system matrix $A(y)$ satisfies $\lambda_{\max}(A(y)) < 0$.

For affine $A(y)$, e.g. $A(y) := C - \sum_{i=1}^m y_i A_i$ with $C, A_i \in S^n$, this leads to

$$\min_{y \in \mathbb{R}^m} \lambda_{\max}(C - \mathcal{A}^T y)$$

To model this as SDP: $\lambda_{\max}(A) = -\lambda_{\min}(-A)$ and

$$y_0 \geq \lambda_{\max}(C - \mathcal{A}^T y) \;\Leftrightarrow\; y_0 + \lambda_{\min}(\mathcal{A}^T y - C) \geq 0 \;\Leftrightarrow\; \lambda_{\min}(y_0 I + \mathcal{A}^T y - C) \geq 0$$

Because $Z \succeq 0 \;\Leftrightarrow\; \lambda_{\min}(Z) \geq 0$ we have

$$\min_{y \in \mathbb{R}^m} \lambda_{\max}(C - \mathcal{A}^T y) \quad \Leftrightarrow \quad \begin{array}{ll} \min & y_0 \\ \text{s.t.} & Z = y_0 I + \mathcal{A}^T y - C \\ & y \in \mathbb{R}^m, Z \succeq 0 \end{array}$$

Design of Experiments

In order to estimate the value of some parameter vector $\xi \in \mathbb{R}^p$,
a set $\mathcal{R} = \{r_i \in \mathbb{R}^p \colon i = 1, \ldots, n\}$ of possible experiments are available.
Each execution of experiment i delivers a measured value
$r_i^T \xi + \rho_i$ with independent $(\mu = 0, \sigma^2 = 1)$ normally distributed error ρ_i.

Design of Experiments

In order to estimate the value of some parameter vector $\xi \in \mathbb{R}^p$, a set $\mathcal{R} = \{r_i \in \mathbb{R}^p \colon i = 1, \ldots, n\}$ of possible experiments are available. Each execution of experiment i delivers a measured value $r_i^T \xi + \rho_i$ with independent $(\mu = 0, \sigma^2 = 1)$ normally distributed error ρ_i.

If m experiments $a_j \in \mathcal{R}$ (repetitions are allowed) are performed resulting in measurements $\eta_j = a_j^T \xi + \rho_j$, the maximum-likelihood estimate (for $\text{rank}[a_1, \ldots, a_m] = p$) yields an estimated

$$\hat{\xi} = G \sum_{j=1}^{m} \eta_j a_j \qquad \text{with} \quad G = \left(\sum_{j=1}^{m} a_j a_j^T \right)^{-1},$$

whose error distribution has expectation 0 and covariance matrix G.

Design of Experiments

In order to estimate the value of some parameter vector $\xi \in \mathbb{R}^p$,
a set $\mathcal{R} = \{r_i \in \mathbb{R}^p \colon i = 1, \dots, n\}$ of possible experiments are available.
Each execution of experiment i delivers a measured value
$r_i^T \xi + \rho_i$ with independent $(\mu = 0, \sigma^2 = 1)$ normally distributed error ρ_i.

If m experiments $a_j \in \mathcal{R}$ (repetitions are allowed) are performed resulting in measurements $\eta_j = a_j^T \xi + \rho_j$, the maximum-likelihood estimate (for $\operatorname{rank}[a_1, \dots, a_m] = p$) yields an estimated

$$\hat{\xi} = G \sum_{j=1}^{m} \eta_j a_j \qquad \text{with} \quad G = \left(\sum_{j=1}^{m} a_j a_j^T \right)^{-1},$$

whose error distribution has expectation 0 and covariance matrix G.

Let G and G' be two covariance matrices of this kind and suppose $G \preceq G'$, then the experiments of G are better, because variance of the estimation error is smaller.
$\rightarrow$ Find a minimal (w.r.t. $\preceq$) element of

$$\left\{ G = \left(\sum_{i=1}^{n} m_i r_i r_i^T \right)^{-1} \colon m_i \in \mathbb{N}_0, \sum_i m_i = m \right\}.$$

Relaxations

Rather than selecting m experiments, determine their relative contribution,

$$\left\{ G = \left(\sum_{i=1}^{n} \alpha_i r_i r_i^T \right)^{-1} : \mathbf{1}^T \alpha = 1, \alpha \geq 0 \right\}.$$

Relaxations

Rather than selecting m experiments, determine their relative contribution,

$$\left\{ G = \left(\sum_{i=1}^{n} \alpha_i r_i r_i^T \right)^{-1} : \mathbf{1}^T \alpha = 1, \alpha \geq 0 \right\}.$$

There are several approaches for finding a $\preceq$-minimal G. For this, interpret G as a "confidence ellipsoid" with semi axes of length $\lambda_j(G)$,

$$\mathcal{E} = \{\zeta \colon (\zeta - \hat{\xi})^T G^{-1} (\zeta - \hat{\xi}) \leq \beta\}.$$

D-**optimal design**: minimise the volume of the confidence ellipsoid
E-**optimal design**: minimise the longest semi axis
A-**optimal design**: minimise the sum of the semi axes

Relaxations

Rather than selecting m experiments, determine their relative contribution,

$$\left\{ G = \left(\sum_{i=1}^{n} \alpha_i r_i r_i^T \right)^{-1} : \mathbf{1}^T\alpha = 1, \alpha \geq 0 \right\}.$$

There are several approaches for finding a $\preceq$-minimal G. For this, interpret G as a "confidence ellipsoid" with semi axes of length $\lambda_j(G)$,

$$\mathcal{E} = \{\zeta : (\zeta - \hat{\xi})^T G^{-1} (\zeta - \hat{\xi}) \leq \beta\}.$$

D-**optimal design**: minimise the volume of the confidence ellipsoid
E-**optimal design**: minimise the longest semi axis
A-**optimal design**: minimise the sum of the semi axes

D-**optimal design:** the volume is proportional to $\det G = \prod \lambda_j(G)$.
Because $\det(G^{-1}) = \det(G)^{-1} \Leftrightarrow$ maximise the determinant of G^{-1},

$$\begin{array}{ll} \min & -\log\det X \\ \text{s.t.} & X = \sum_{i=1}^{n} \alpha_i r_i r_i^T \\ & \mathbf{1}^T\alpha = 1 \\ & \alpha \geq 0, [X \succ 0] \end{array}$$

Relaxations

Rather than selecting m experiments, determine their relative contribution,

$$\left\{ G = \left(\sum_{i=1}^{n} \alpha_i r_i r_i^T \right)^{-1} : \mathbf{1}^T \alpha = 1, \alpha \geq 0 \right\}.$$

There are several approaches for finding a $\preceq$-minimal G. For this, interpret G as a "confidence ellipsoid" with semi axes of length $\lambda_j(G)$,

$$\mathcal{E} = \{\zeta \colon (\zeta - \hat{\xi})^T G^{-1} (\zeta - \hat{\xi}) \leq \beta\}.$$

D-**optimal design**: minimise the volume of the confidence ellipsoid
E-**optimal design**: minimise the longest semi axis
A-**optimal design**: minimise the sum of the semi axes

E-**optimal design:** the longest semi axis is $\lambda_{\max}(G)$.
Because $\lambda_{\min}(G^{-1}) = \lambda_{\max}(G)^{-1} \Leftrightarrow$ maximise $\lambda_{\min}(G^{-1})$,

$$\begin{array}{ll} \max & -\lambda \\ \text{s.t.} & \sum_{i=1}^{n} \alpha_i r_i r_i^T \succeq \lambda I \\ & \mathbf{1}^T \alpha = 1 \\ & \alpha \geq 0, \lambda \in \mathbb{R} \end{array}$$

Relaxations

Rather than selecting m experiments, determine their relative contribution,

$$\left\{G = \left(\sum_{i=1}^{n} \alpha_i r_i r_i^T\right)^{-1} : \mathbf{1}^T \alpha = 1, \alpha \geq 0\right\}.$$

There are several approaches for finding a $\preceq$-minimal G. For this, interpret G as a "confidence ellipsoid" with semi axes of length $\lambda_j(G)$,

$$\mathcal{E} = \{\zeta \colon (\zeta - \hat{\xi})^T G^{-1} (\zeta - \hat{\xi}) \leq \beta\}.$$

D-**optimal design**: minimise the volume of the confidence ellipsoid
E-**optimal design**: minimise the longest semi axis
A-**optimal design**: minimise the sum of the semi axes

A-**optimal Design:** $\sum_{j=1}^p \lambda_j(G) = \sum_{j=1}^p G_{jj} = \sum_{j=1}^p e_j^T G e_j$.
For each j represent the inequality $u_j \succeq e_j^T G e_j$ by its Schur complement:

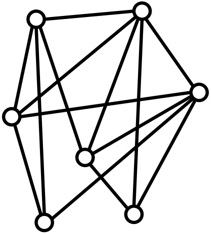
$$\begin{aligned} \min \quad & \mathbf{1}^T u \\ \text{s.t.} \quad & \begin{bmatrix} \sum_{i=1}^n \alpha_i r_i r_i^T & e_j \\ e_j^T & u_j \end{bmatrix} \succeq 0, \quad j = 1, \ldots, p \\ & \mathbf{1}^T \alpha = 1, \alpha \geq 0, u \in \mathbb{R}^p \end{aligned}$$

Graph Partition: Max-Cut

Graph Partition: Max-Cut

Given: graph $G = (V, E)$, $V = \{1, \dots, n\}$,
$E \subseteq \{ij \colon i, j \in V, i < j\}$, edge weights a_{ij}

Find: $S \subset V$ with maximum weight cut
$\delta(S) := \{ij \in E \colon i \in S, j \in V \setminus S\}$

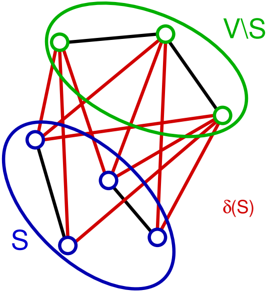


$$\text{(MC)} \qquad \max_{S \subseteq V} \sum_{ij \in \delta(S)} a_{ij} \qquad [NP\text{-compl.}]$$

Graph Partition: Max-Cut

Given: graph $G = (V, E)$, $V = \{1, \ldots, n\}$,
$E \subseteq \{ij : i, j \in V, i < j\}$, edge weights a_{ij}

Find: $S \subset V$ with maximum weight cut
$\delta(S) := \{ij \in E : i \in S, j \in V \setminus S\}$



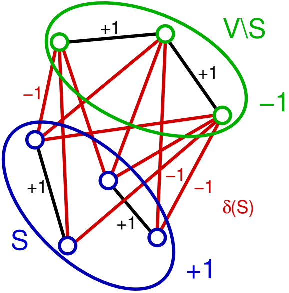
$$\text{(MC)} \qquad \max_{S \subseteq V} \sum_{ij \in \delta(S)} a_{ij} \qquad [NP\text{-compl.}]$$

$$\text{then } x_i x_j = \begin{cases} -1 & ij \in \delta(S) \\ 1 & \text{otherwise} \end{cases}, \quad \text{or} \quad \frac{1 - x_i x_j}{2} = \begin{cases} 1 & ij \in \delta(S) \\ 0 & \text{otherwise} \end{cases}$$

Graph Partition: Max-Cut

Given: graph $G = (V, E)$, $V = \{1, \dots, n\}$, $E \subseteq \{ij : i, j \in V, i < j\}$, edge weights a_{ij}

Find: $S \subset V$ with maximum weight cut $\delta(S) := \{ij \in E : i \in S, j \in V \setminus S\}$



$$\text{(MC)} \qquad \max_{S \subseteq V} \sum_{ij \in \delta(S)} a_{ij} \qquad [NP\text{-compl.}]$$

Model: represent the partition by

$$x \in \{-1, 1\}^n \quad \text{with} \quad x_i = \begin{cases} 1 & i \in S \\ -1 & i \in V \setminus S \end{cases}$$

$$\text{then } x_i x_j = \begin{cases} -1 & ij \in \delta(S) \\ 1 & \text{otherwise} \end{cases}, \quad \text{or} \quad \frac{1 - x_i x_j}{2} = \begin{cases} 1 & ij \in \delta(S) \\ 0 & \text{otherwise} \end{cases}$$

$$\max_{S \subseteq V} \sum_{ij \in \delta(S)} a_{ij} = \max_{x \in \{-1,1\}^n} \sum_{ij \in E} a_{ij} \frac{1 - x_i x_j}{2} \qquad \to \qquad \max_{x \in \{-1,1\}^n} x^T C x$$

[$C \in S^n$: $C_{ii} = \frac{1}{4} \sum_{j:\, ij \in E} a_{ij}$ (for $i \in V$), $C_{ij} = -\frac{1}{4} a_{ij}$ (for $ij \in E$), 0 otherw.]

Equivalent to quadratic 0-1 optimisation!

Semidefinite Max-Cut Relaxation

Observe: $x^T C x = \langle Cx, x \rangle = \left\langle C, xx^T \right\rangle$

Properties of $xx^T = [x_i x_j]$ for $x \in \{-1, 1\}^n$:

- $x_i^2 = 1 \quad \Rightarrow \quad \text{diag}(xx^T) = \mathbf{1}$
- xx^T is positive semidefinite, $xx^T \succeq 0$
- $\text{rank}(xx^T) = 1$

Semidefinite Max-Cut Relaxation

Observe: $x^T Cx = \langle Cx, x\rangle = \left\langle C, xx^T \right\rangle$

Properties of $xx^T = [x_i x_j]$ for $x \in \{-1,1\}^n$:

- $x_i^2 = 1 \quad \Rightarrow \quad \text{diag}(xx^T) = \mathbf{1}$
- xx^T is positive semidefinite, $xx^T \succeq 0$
- $\text{rank}(xx^T) = 1$

Relaxation idea: replace xx^T by a positive semidefinite matrix X.

$$\max_{x\in\{-1,1\}^n} x^T Cx \;\le\; \begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \text{diag}(X) = \mathbf{1} \\ & X \succeq 0 \\ & [\text{rank}(X) = 1] \end{array}$$

[with rank 1 ⇔ (MC), *NP*-compl.]

Semidefinite Max-Cut Relaxation

Observe: $x^T Cx = \langle Cx, x\rangle = \left\langle C, xx^T \right\rangle$

Properties of $xx^T = [x_i x_j]$ for $x \in \{-1, 1\}^n$:

- $x_i^2 = 1 \quad \Rightarrow \quad \text{diag}(xx^T) = \mathbf{1}$
- xx^T is positive semidefinite, $xx^T \succeq 0$
- $\text{rank}(xx^T) = 1$

Relaxation idea: replace xx^T by a positive semidefinite matrix X.

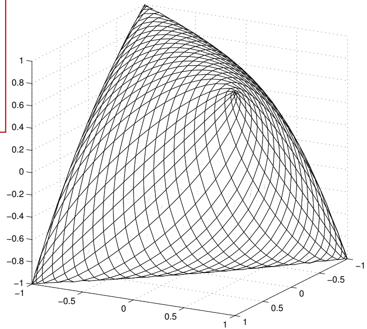
$$\max_{x\in\{-1,1\}^n} x^T Cx \;\le\; \begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \text{diag}(X) = \mathbf{1} \\ & X \succeq 0 \\ & [\text{rank}(X) = 1] \end{array}$$

[with rank 1 $\Leftrightarrow$ (MC), *NP*-compl.]

Illustration for $n = 3$:

boundary described by

$$\det \begin{bmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{bmatrix} = 0.$$



Semidefinite Max-Cut Relaxation

Observe: $x^T Cx = \langle Cx, x\rangle = \left\langle C, xx^T\right\rangle$

Properties of $xx^T = [x_i x_j]$ for $x \in \{-1,1\}^n$:

- $x_i^2 = 1 \quad \Rightarrow \quad \text{diag}(xx^T) = \mathbf{1}$
- xx^T is positive semidefinite, $xx^T \succeq 0$
- $\text{rank}(xx^T) = 1$

Relaxation idea: replace xx^T by a positive semidefinite matrix X.

$$\max_{x\in\{-1,1\}^n} x^T Cx \;\leq\; \begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \text{diag}(X) = \mathbf{1} \\ & X \succeq 0 \\ & [\text{rank}(X) = 1] \end{array}$$

[with rank 1 ⇔ (MC), *NP*-compl.]

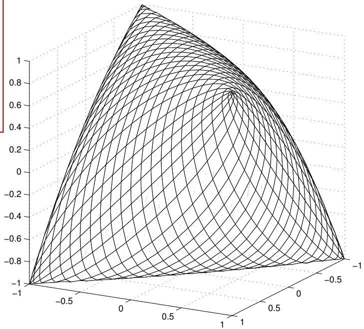


Illustration for $n = 3$:
boundary described by

$$\det \begin{bmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{bmatrix} = 0.$$

[Approx.-alg. of GW95: factorise X, use randomised hyperplane rounding]

Moment Matrices and Optimisation Over Polynomials

[Lasserre]

Polynomial $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ of degree $2m$. Find

$$p_* := \min_{x \in \mathbb{R}^n} p(x)$$

Moment Matrices and Optimisation Over Polynomials

[Lasserre]

Polynomial $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ of degree $2m$. Find

$$p_* := \min_{x \in \mathbb{R}^n} p(x) = \min_{\mu \in \mathcal{P}(\mathbb{R}^n)} \int p(x) \mu(dx)$$

Moment Matrices and Optimisation Over Polynomials

[Lasserre]

Polynomial $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ of degree $2m$. Find

$$p_* := \min_{x \in \mathbb{R}^n} p(x) = \min_{\mu \in \mathcal{P}(\mathbb{R}^n)} \int p(x)\mu(dx) \to \sum_{\alpha=(\alpha_1,\ldots,\alpha_n)} p_\alpha \underbrace{\int x_1^{\alpha_1} \cdots x_n^{\alpha_n} \mu(dx)}_{=: \, y_\alpha}$$

Moment Matrices and Optimisation Over Polynomials

[Lasserre]

Polynomial $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ of degree $2m$. Find

$$p_* := \min_{x \in \mathbb{R}^n} p(x) = \min_{\mu \in \mathcal{P}(\mathbb{R}^n)} \int p(x)\mu(dx) \rightarrow \sum_{\alpha=(\alpha_1,\dots,\alpha_n)} p_\alpha \underbrace{\int x_1^{\alpha_1} \cdots x_n^{\alpha_n} \mu(dx)}_{=: \, y_\alpha}$$

For moment vectors $y = (y_\alpha)$ of prob. distributions the moment matrix

$$[M_m(y)]_{\alpha\beta} = [y_{\alpha+\beta}], \qquad \text{e.g. } M_1(y) = \begin{bmatrix} 1 & y_{(1,0)} & y_{(0,1)} \\ y_{(1,0)} & y_{(2,0)} & y_{(1,1)} \\ y_{(0,1)} & y_{(1,1)} & y_{(0,2)} \end{bmatrix}$$

has to be positive semidefinite (necessary, not sufficient).

Moment Matrices and Optimisation Over Polynomials

[Lasserre]

Polynomial $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ of degree $2m$. Find

$$p_* := \min_{x \in \mathbb{R}^n} p(x) = \min_{\mu \in \mathcal{P}(\mathbb{R}^n)} \int p(x)\mu(dx) \rightarrow \sum_{\alpha=(\alpha_1,\dots,\alpha_n)} p_\alpha \underbrace{\int x_1^{\alpha_1} \cdots x_n^{\alpha_n} \mu(dx)}_{=:\, y_\alpha}$$

For moment vectors $y = (y_\alpha)$ of prob. distributions the moment matrix

$$[M_m(y)]_{\alpha\beta} = [y_{\alpha+\beta}], \qquad \text{e.g. } M_1(y) = \begin{bmatrix} 1 & y_{(1,0)} & y_{(0,1)} \\ y_{(1,0)} & y_{(2,0)} & y_{(1,1)} \\ y_{(0,1)} & y_{(1,1)} & y_{(0,2)} \end{bmatrix}$$

has to be positive semidefinite (necessary, not sufficient).

$[p^T M p = \sum_{\alpha,\beta} p_\alpha p_\beta \int x^\alpha x^\beta \mu(dx) = \int p(x)^2 \mu(dx) \geq 0]$

Moment Matrices and Optimisation Over Polynomials

[Lasserre]

Polynomial $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ of degree $2m$. Find

$$p_* := \min_{x \in \mathbb{R}^n} p(x) = \min_{\mu \in \mathcal{P}(\mathbb{R}^n)} \int p(x)\mu(dx) \to \sum_{\alpha=(\alpha_1,\ldots,\alpha_n)} p_\alpha \underbrace{\int x_1^{\alpha_1} \cdots x_n^{\alpha_n} \mu(dx)}_{=: y_\alpha}$$

For moment vectors $y = (y_\alpha)$ of prob. distributions the moment matrix

$$[M_m(y)]_{\alpha\beta} = [y_{\alpha+\beta}], \qquad \text{e.g. } M_1(y) = \begin{bmatrix} 1 & y_{(1,0)} & y_{(0,1)} \\ y_{(1,0)} & y_{(2,0)} & y_{(1,1)} \\ y_{(0,1)} & y_{(1,1)} & y_{(0,2)} \end{bmatrix}$$

has to be positive semidefinite (necessary, not sufficient).

$[p^T M p = \sum_{\alpha,\beta} p_\alpha p_\beta \int x^\alpha x^\beta \mu(dx) = \int p(x)^2 \mu(dx) \geq 0]$

Relax minimising over prob. distributions to

$$\begin{array}{ll} \min & \sum p_\alpha y_\alpha \\ \text{s.t.} & M_m(y) \succeq 0. \end{array}$$

Moment Matrices and Optimisation Over Polynomials

[Lasserre]

Polynomial $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ of degree $2m$. Find

$$p_* := \min_{x \in \mathbb{R}^n} p(x) = \min_{\mu \in \mathcal{P}(\mathbb{R}^n)} \int p(x)\mu(dx) \to \sum_{\alpha=(\alpha_1,\ldots,\alpha_n)} p_\alpha \underbrace{\int x_1^{\alpha_1} \cdots x_n^{\alpha_n} \mu(dx)}_{=:\, y_\alpha}$$

For moment vectors $y = (y_\alpha)$ of prob. distributions the moment matrix

$$[M_m(y)]_{\alpha\beta} = [y_{\alpha+\beta}], \qquad \text{e.g. } M_1(y) = \begin{bmatrix} 1 & y_{(1,0)} & y_{(0,1)} \\ y_{(1,0)} & y_{(2,0)} & y_{(1,1)} \\ y_{(0,1)} & y_{(1,1)} & y_{(0,2)} \end{bmatrix}$$

has to be positive semidefinite (necessary, not sufficient).

$[p^T M p = \sum_{\alpha,\beta} p_\alpha p_\beta \int x^\alpha x^\beta \mu(dx) = \int p(x)^2 \mu(dx) \geq 0]$

Relax minimising over prob. distributions to

$$\begin{array}{ll} \min & \sum p_\alpha y_\alpha \\ \text{s.t.} & M_m(y) \succeq 0. \end{array}$$

exact $\Leftrightarrow$ $p(x) - p_*$ is a sum of squares of polynomials (SOS).

Sum of Squares Decomposition

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ polynomial $p(x) - \rho$ of deg. $2m$ is nonnegative.

Sum of Squares Decomposition

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ polynomial $p(x) - \rho$ of deg. $2m$ is nonnegative.
If $p(x) - \rho$ is a SOS-polynomial then $\rho \leq \min_x p(x)$. Check if

$$p(x) - \rho = \sum_{\alpha=(\alpha_1,\ldots,\alpha_n)} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} - \rho \qquad = \qquad \sum f_i(x)^2$$

for polynomials f_i of degree $\leq m$.

Sum of Squares Decomposition

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ polynomial $p(x) - \rho$ of deg. $2m$ is nonnegative.
If $p(x) - \rho$ is a SOS-polynomial then $\rho \leq \min_x p(x)$. Check if

$$p(x) - \rho = \sum_{\alpha=(\alpha_1,\ldots,\alpha_n)} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} - \rho \qquad = \qquad \sum f_i(x)^2$$

for polynomials f_i of degree $\leq m$.
Put $z = (1, x_1, x_2, \ldots, x_n, x_1x_2, x_1x_3, \ldots, x_n^m)$ and write f_i as

$$f_i(x) = a_i^T z, \quad \text{then} \quad \sum f_i(x)^2 = z^T A A^T z \quad \text{with} \quad H = AA^T \succeq 0.$$

Sum of Squares Decomposition

$\rho \le \min_x p(x) \quad \Leftrightarrow \quad$ polynomial $p(x) - \rho$ of deg. $2m$ is nonnegative.
If $p(x) - \rho$ is a SOS-polynomial then $\rho \le \min_x p(x)$. Check if

$$p(x) - \rho = \sum_{\alpha=(\alpha_1,\dots,\alpha_n)} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} - \rho \qquad = \qquad \sum f_i(x)^2$$

for polynomials f_i of degree $\le m$.
Put $z = (1, x_1, x_2, \dots, x_n, x_1x_2, x_1x_3, \dots, x_n^m)$ and write f_i as

$$f_i(x) = a_i^T z, \quad \text{then} \quad \sum f_i(x)^2 = z^T A A^T z \quad \text{with} \quad H = AA^T \succeq 0.$$

By comparing coefficients such a representation exists iff

$$\exists H \succeq 0: \qquad \sum_{\beta+\gamma=\alpha} H_{\beta,\gamma} = p_\alpha \quad \text{for all monomials } \alpha$$

Sum of Squares Decomposition

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ polynomial $p(x) - \rho$ of deg. $2m$ is nonnegative.
If $p(x) - \rho$ is a SOS-polynomial then $\rho \leq \min_x p(x)$. Check if

$$p(x) - \rho = \sum_{\alpha=(\alpha_1,\dots,\alpha_n)} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} - \rho \qquad = \qquad \sum f_i(x)^2$$

for polynomials f_i of degree $\leq m$.
Put $z = (1, x_1, x_2, \dots, x_n, x_1 x_2, x_1 x_3, \dots, x_n^m)$ and write f_i as

$$f_i(x) = a_i^T z, \quad \text{then} \quad \sum f_i(x)^2 = z^T A A^T z \quad \text{with} \quad H = AA^T \succeq 0.$$

By comparing coefficients such a representation exists iff

$$\exists H \succeq 0 \colon \qquad \sum_{\beta+\gamma=\alpha} H_{\beta,\gamma} = p_\alpha \quad \text{for all monomials } \alpha$$

Try to find $\min_x p(x) = p_*$ via maximising p_0,

$$\begin{aligned} \min\ & \sum p_\alpha y_\alpha \\ \text{s.t.}\ & \begin{bmatrix} 1 & y_{(1,0)} & y_{(0,1)} \\ y_{(1,0)} & y_{(2,0)} & y_{(1,1)} \\ y_{(0,1)} & y_{(1,1)} & y_{(0,2)} \end{bmatrix} = B_0 + \sum_{\alpha \neq 0} B_\alpha y_\alpha \succeq 0 \end{aligned}$$

$$\begin{aligned} \max \quad & \langle B_0, H \rangle \ [= -H_{0,0} = -p_0] \\ \text{s.t.} \quad & \langle B_\alpha, H \rangle = p_\alpha \quad \alpha \neq 0 \\ & H \succeq 0. \end{aligned}$$

If $p(x) - p_*$ is not SOS ...

one may approximate p_* via hierarchies of semidefinite relaxations to arbitrary precision under compactness assumptions. [very expensive]

If $p(x) - p_*$ is not SOS ...

one may approximate p_* via hierarchies of semidefinite relaxations to arbitrary precision under compactness assumptions. [very expensive]

The approach is extendable to sets restricted by polynomial (in)equalities
$\rightarrow$ GloptiPoly

If $p(x) - p_*$ is not SOS ...

one may approximate p_* via hierarchies of semidefinite relaxations to arbitrary precision under compactness assumptions. [very expensive]

The approach is extendable to sets restricted by polynomial (in)equalities $\rightarrow$ GloptiPoly

---

For equality restrictions $x_i^2 = x_i$ ($x \in \{0,1\}^n$) the initial relaxation is equivalent to the semidefinite Max-Cut relaxation.

If $p(x) - p_*$ is not SOS ...

one may approximate p_* via hierarchies of semidefinite relaxations to arbitrary precision under compactness assumptions. [very expensive]

The approach is extendable to sets restricted by polynomial (in)equalities $\rightarrow$ GloptiPoly

---

For equality restrictions $x_i^2 = x_i$ ($x \in \{0,1\}^n$) the initial relaxation is equivalent to the semidefinite Max-Cut relaxation.

The complex hermitian case provided excellent results for optimal power flow [LavaeiLow2012]

Contents

Duality Gap Example

$$
\begin{array}{ll}
\min & x_{12} \\
\text{s.t.} & \begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1 + x_{12} \end{bmatrix} \succeq 0
\end{array}
$$

Duality Gap Example

$$\begin{array}{ll} \min & x_{12} \\ \text{s.t.} & \begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1 + x_{12} \end{bmatrix} \succeq 0 \end{array}$$

corresponding coefficient matrices:

$$C = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{array}{ll} \max & \langle C, X \rangle \\ & \langle A_1, X \rangle = 1 \\ & \langle A_2, X \rangle = 0 \\ & \langle A_3, X \rangle = 0 \\ & \langle A_4, X \rangle = 0 \\ & X \succeq 0 \end{array}$$

Duality Gap Example

$$\min \quad x_{12} \quad \text{s.t.} \begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1+x_{12} \end{bmatrix} \succeq 0$$

$$\max \quad y_1 \quad \text{s.t.} \ Z = \begin{bmatrix} -y_2 & \frac{1+y_1}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \succeq 0$$

corresponding coefficient matrices:

$$C = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ & \langle A_1, X \rangle = 1 \\ & \langle A_2, X \rangle = 0 \\ & \langle A_3, X \rangle = 0 \\ & \langle A_4, X \rangle = 0 \\ & X \succeq 0 \end{aligned}$$

Duality Gap Example

$$\begin{array}{ll} \min & x_{12} \\ \text{s.t.} & \begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1+x_{12} \end{bmatrix} \succeq 0 \end{array} \qquad \begin{array}{ll} \max & y_1 \\ \text{s.t.} & Z = \begin{bmatrix} -y_2 & \frac{1+y_1}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \succeq 0 \end{array}$$

corresponding coefficient matrices:

$$C = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{array}{ll} \max & \langle C, X \rangle \\ & \langle A_1, X \rangle = 1 \\ & \langle A_2, X \rangle = 0 \\ & \langle A_3, X \rangle = 0 \\ & \langle A_4, X \rangle = 0 \\ & X \succeq 0 \end{array}$$

$x_{11} = 0 \quad \Rightarrow \quad x_{12} = 0,$ primal optimal value is 0.

Duality Gap Example

$$\begin{array}{ll} \min & x_{12} \\ \text{s.t.} & \begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1+x_{12} \end{bmatrix} \succeq 0 \end{array} \qquad \begin{array}{ll} \max & y_1 \\ \text{s.t.} & Z = \begin{bmatrix} -y_2 & \frac{1+y_1}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \succeq 0 \end{array}$$

corresponding coefficient matrices:

$$C = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{array}{ll} \max & \langle C, X \rangle \\ & \langle A_1, X \rangle = 1 \\ & \langle A_2, X \rangle = 0 \\ & \langle A_3, X \rangle = 0 \\ & \langle A_4, X \rangle = 0 \\ & X \succeq 0 \end{array}$$

$x_{11} = 0 \quad \Rightarrow \quad x_{12} = 0,$ primal optimal value is 0.

$z_{22} = 0 \quad \Rightarrow \quad \frac{1+y_1}{2} = 0,$ dual optimal value is -1.

Difficulty: Primal Problem is Unstable

$$\min \quad x_{12}$$
$$\text{s.t.} \begin{bmatrix} \varepsilon & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1+x_{12} \end{bmatrix} \succeq 0$$

$$\max \quad y_1 + \varepsilon y_2$$
$$\text{s.t.} \quad Z = \begin{bmatrix} -y_2 & \frac{1+y_1}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \succeq 0$$

corresponding coefficient matrices:

$$C = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

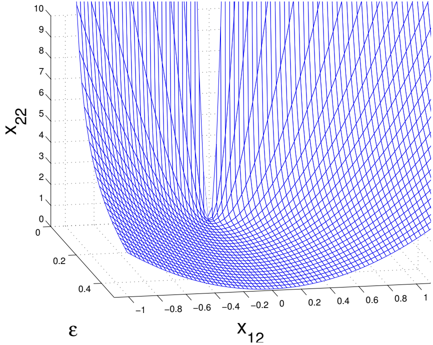
$$A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ & \langle A_1, X \rangle = 1 \\ & \langle A_2, X \rangle = \varepsilon \\ & \langle A_3, X \rangle = 0 \\ & \langle A_4, X \rangle = 0 \\ & X \succeq 0 \end{aligned}$$

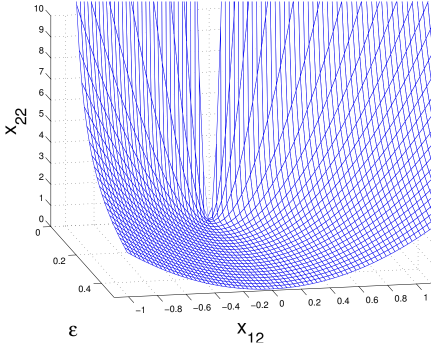
$x_{33} \geq 0 \;\Rightarrow\; x_{12} \geq -1,\; x_{22} \geq \frac{x_{12}^2}{\varepsilon}$, primal optimal value is -1.

$z_{22} = 0 \;\Rightarrow\; \frac{1+y_1}{2} = 0,\; y_2 = 0$, dual optimal value is -1.

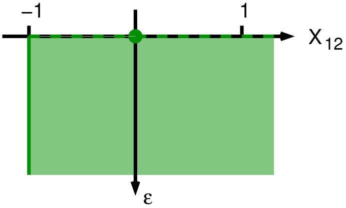
$$\begin{bmatrix} \varepsilon & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1 + x_{12} \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad \begin{cases} \varepsilon > 0 & x_{12} \geq -1 \quad x_{22} \geq \frac{x_{12}^2}{\varepsilon} \\ \varepsilon = 0 & x_{12} = 0 \quad\quad x_{22} \geq 0 \end{cases}$$



$$\begin{bmatrix} \varepsilon & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1 + x_{12} \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad \begin{cases} \varepsilon > 0 & x_{12} \geq -1 \quad x_{22} \geq \frac{x_{12}^2}{\varepsilon} \\ \varepsilon = 0 & x_{12} = 0 \quad\;\; x_{22} \geq 0 \end{cases}$$

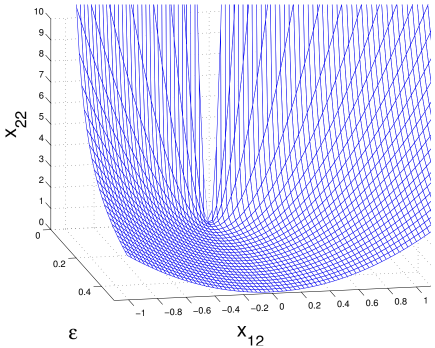


projection to (ε, x_{12})-plane:

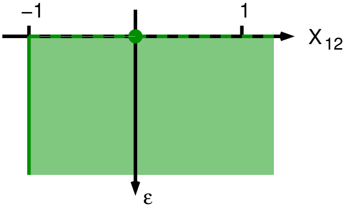


For $\varepsilon > 0$ all $x_{12} \in [-1, -\infty)$,
for $\varepsilon = 0$ only $x_{12} \in \{0\}$
feasible!

$$\begin{bmatrix} \varepsilon & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1 + x_{12} \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad \begin{cases} \varepsilon > 0 & x_{12} \geq -1 \quad x_{22} \geq \frac{x_{12}^2}{\varepsilon} \\ \varepsilon = 0 & x_{12} = 0 \quad\;\;\, x_{22} \geq 0 \end{cases}$$



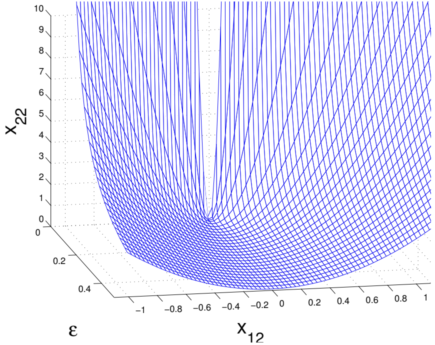
projection to (ε, x_{12})-plane:



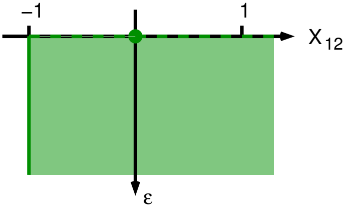
For $\varepsilon > 0$ all $x_{12} \in [-1, -\infty)$, for $\varepsilon = 0$ only $x_{12} \in \{0\}$ feasible!

mathematical reason: the set $\left\{ \begin{bmatrix} \langle C, X \rangle \\ \mathcal{A}X \end{bmatrix} : X \succeq 0 \right\}$ is not closed.

$$\begin{bmatrix} \varepsilon & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1 + x_{12} \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad \begin{cases} \varepsilon > 0 & x_{12} \geq -1 \quad x_{22} \geq \frac{x_{12}^2}{\varepsilon} \\ \varepsilon = 0 & x_{12} = 0 \quad x_{22} \geq 0 \end{cases}$$



projection to (ε, x_{12})-plane:



For $\varepsilon > 0$ all $x_{12} \in [-1, -\infty)$, for $\varepsilon = 0$ only $x_{12} \in \{0\}$ feasible!

mathematical reason: the set $\left\{ \begin{bmatrix} \langle C, X \rangle \\ \mathcal{A}X \end{bmatrix} : X \succeq 0 \right\}$ is not closed.

To avoid this, require strictly feasible points or apply facial reduction.

Facial Structure and Facial Reduction

The faces of S^n_+ are: $\emptyset$, $\{\mathbf{0}\}$ and

for each r-dim. linear subspace $\mathcal{L}$ of $\mathbb{R}^n$ represented by some basis $P \in \mathbb{R}^{n\times r}$,

$F_{\mathcal{L}} = \{X = PUP^T : U \in S^r_+\}$

[Barker and Carlson 1975]

Facial Structure and Facial Reduction

The faces of S^n_+ are: $\emptyset$, $\{\mathbf{0}\}$ and

for each r-dim. linear subspace $\mathcal{L}$ of $\mathbb{R}^n$
represented by some basis $P \in \mathbb{R}^{n\times r}$,

$F_{\mathcal{L}} = \{X = PUP^T : U \in S^r_+\}$

[Barker and Carlson 1975]

- $\dim F_{\mathcal{L}} = \binom{r+1}{2}$
- minimal generating system: $S^n_+ = \text{cone}\,\{vv^T : \|v\| = 1\}$

Facial Structure and Facial Reduction

The faces of S^n_+ are: $\emptyset$, $\{\mathbf{0}\}$ and

for each r-dim. linear subspace $\mathcal{L}$ of $\mathbb{R}^n$
represented by some basis $P \in \mathbb{R}^{n\times r}$,

$F_{\mathcal{L}} = \{X = PUP^T : U \in S^r_+\}$

[Barker and Carlson 1975]

- $\dim F_{\mathcal{L}} = \binom{r+1}{2}$
- minimal generating system: $S^n_+ = \text{cone}\{vv^T : \|v\| = 1\}$

Feasible $\begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1+x_{12} \end{bmatrix} \succeq 0$ live on the face $F_{\mathcal{L}}$ for $P = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$,
the dual $Z \in F^*_{\mathcal{L}}$ requires positive semidefiniteness on this subspace

$$\begin{aligned} \min \quad & x_{12} \\ \text{s.t.} \quad & \begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1+x_{12} \end{bmatrix} \in F_{\mathcal{L}} \end{aligned} \qquad \begin{aligned} \max \quad & y_1 \\ \text{s.t.} \quad & P^TZP = \begin{bmatrix} 0 & -y_4 \\ -y_4 & -y_1 \end{bmatrix} \succeq 0 \end{aligned}$$

Facial reduction ensures primal strict feasibility $\rightarrow$ both optima 0

Complexity of SDP

The feasibility problem in the real number model is in $NP \cap Co\text{-}NP$, in the Turing- or bit-model the status is still open!

Complexity of SDP

The feasibility problem in the real number model is in $NP \cap Co\text{-}NP$, in the Turing- or bit-model the status is still open!

Example: [Ramana1997]

$$\begin{aligned} \min \quad & x_m \\ \text{s.t.} \quad & (x_1 - 4) \succeq 0, \begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix} \succeq 0, \begin{bmatrix} 1 & x_2 \\ x_2 & x_3 \end{bmatrix} \succeq 0, \ldots, \begin{bmatrix} 1 & x_{m-1} \\ x_{m-1} & x_m \end{bmatrix} \succeq 0. \end{aligned}$$

Complexity of SDP

The feasibility problem in the real number model is in $NP \cap Co\text{-}NP$, in the Turing- or bit-model the status is still open!

Example: [Ramana1997]

$$\begin{aligned} \min \quad & x_m \\ \text{s.t.} \quad & (x_1 - 4) \succeq 0, \begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix} \succeq 0, \begin{bmatrix} 1 & x_2 \\ x_2 & x_3 \end{bmatrix} \succeq 0, \ldots, \begin{bmatrix} 1 & x_{m-1} \\ x_{m-1} & x_m \end{bmatrix} \succeq 0. \end{aligned}$$

$$\begin{aligned} \Rightarrow x_1 &\geq 2^2, \\ x_2 &\geq x_1^2 \geq (2^2)^2 = 2^{(2^2)}, \\ x_3 &\geq x_2^2 \geq 2^{(2^3)}, \\ &\vdots \\ x_m &\geq x_{m-1}^2 \geq 2^{(2^m)} \end{aligned}$$

doubly exponential values!

Complexity of SDP

The feasibility problem in the real number model is in $NP \cap Co\text{-}NP$, in the Turing- or bit-model the status is still open!

Example: [Ramana1997]

$$\begin{array}{ll} \min & x_m \\ \text{s.t.} & (x_1 - 4) \succeq 0, \begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix} \succeq 0, \begin{bmatrix} 1 & x_2 \\ x_2 & x_3 \end{bmatrix} \succeq 0, \dots, \begin{bmatrix} 1 & x_{m-1} \\ x_{m-1} & x_m \end{bmatrix} \succeq 0. \end{array}$$

$$\begin{aligned} \Rightarrow x_1 &\geq 2^2, \\ x_2 &\geq x_1^2 \geq (2^2)^2 = 2^{(2^2)}, \\ x_3 &\geq x_2^2 \geq 2^{(2^3)}, \\ &\vdots \\ x_m &\geq x_{m-1}^2 \geq 2^{(2^m)} \end{aligned}$$

doubly exponential values!

---

polynomial for ε-solutions in bounded regions (ellipsoid method) [Grötschel Lovász Schrijver 1988]

Contents

Solution Methods

- interior-point methods ("polynomial")
  codes: SDPT3, Sedumi, SDPA, CSDP, Mosek, ...
- penalty methods
  code: Pennon
- spectral bundle method ($f(y) := \lambda_{\max}(C - \mathcal{A}^T y) + b^T y$)
  code: ConicBundle
- quadratic reformulations (replace $0 \preceq X = LL^T$)
  code: SDPLR

Interior-Point Methods

based on the barrier idea take from *nonlinear optimisation*:

start in the interior of the feasible set,

avoid leaving it via a barrier-function.

Interior-Point Methods

based on the barrier idea take from *nonlinear optimisation*:
start in the interior of the feasible set,
avoid leaving it via a barrier-function.

Dual problem

$$(D) \qquad \begin{array}{ll} \min & b^T y \\ \text{s.t.} & Z = \mathcal{A}^T y - C \succeq 0 \end{array}$$

Interior-Point Methods

based on the barrier idea take from *nonlinear optimisation*:

start in the interior of the feasible set,

avoid leaving it via a barrier-function.

Dual problem

$$(\mathrm{D}) \qquad \begin{array}{ll} \min & b^T y \\ \text{s.t.} & Z = \mathcal{A}^T y - C \succeq 0 \end{array}$$

barrier-function for $Z \in S_n^+$: $-\log\det Z$

- $\det Z = \prod \lambda_i(Z)$ is > 0 in the interior of S_n^+ and 0 on the boundary
- $\log\det Z = \sum \log \lambda_i(Z)$
- if some $\lambda_i(Z) \to 0$ then $-\log\det Z \to \infty$
- $-\log\det Z$ is smooth and strictly convex on the interior of S_n^+

[strongly self concordant]

Interior-Point Methods

based on the barrier idea take from *nonlinear optimisation*:
start in the interior of the feasible set,
avoid leaving it via a barrier-function.

Dual problem

$$\text{(D)} \qquad \begin{array}{ll} \min & b^T y \\ \text{s.t.} & Z = \mathcal{A}^T y - C \succeq 0 \end{array}$$

barrier-function for $Z \in S_n^+$: $\quad -\log\det Z$

- $\det Z = \prod \lambda_i(Z)$ is > 0 in the interior of S_n^+ and 0 on the boundary
- $\log\det Z = \sum \log \lambda_i(Z)$
- if some $\lambda_i(Z) \to 0$ then $\quad -\log\det Z \to \infty$
- $-\log\det Z$ is smooth and strictly convex on the interior of S_n^+
[strongly self concordant]

SUMT, Fiacco and McCormick 1968: solve a sequence of barrier-problems

$$\min_y \; b^T y - \mu \log\det(\underbrace{\mathcal{A}^T y - C}_{=Z})$$

by Newton's method for $\mu > 0$, $\mu \to 0$.

Minimise $f(y) = b^T y - \mu \log\det(\mathcal{A}^T y - C)$ by Newton:

1. first order necessary (here sufficient) conditions

$$\nabla f(y) = 0$$

2. determine step Δy so that the linearisation in the current point y_c,

$$\nabla f(y_c) + \nabla^2 f(y_c)^T \Delta y = 0,$$

   satisfies the conditions for $y_c + \Delta y$.
   [= minimises the quadratic model of f]

3. damped Newton step: $y_+ = y_c + \alpha\Delta y$ with $\alpha \in (0, 1]$
   so that y_+ is at least feasible.

Minimise $f(y) = b^T y - \mu \log\det(\mathcal{A}^T y - C)$ by Newton:

1. first order necessary (here sufficient) conditions

$$\nabla f(y) = 0$$

2. determine step Δy so that the linearisation in the current point y_c,

$$\nabla f(y_c) + \nabla^2 f(y_c)^T \Delta y = 0,$$

satisfies the conditions for $y_c + \Delta y$.

[= minimises the quadratic model of f]

3. damped Newton step: $y_+ = y_c + \alpha\Delta y$ with $\alpha \in (0, 1]$
so that y_+ is at least feasible.

Assumptions:

(A) $\mathcal{A}$ has full row rank. [w.l.o.g.]

(S) There exist primal and dual strictly feasible solutions.

Minimise $f(y) = b^T y - \mu \log\det(\mathcal{A}^T y - C)$ by Newton:

1. Sufficient first order optimality conditions:

$\nabla f(y) = 0$, use $\nabla_Z \log\det Z = Z^{-1}$

$$\boxed{b - \mathcal{A}[\mu(\mathcal{A}^T y - C)^{-1}] = 0}$$

yields methods of Jarre 1993, Nesterov and Nemirovskii 1994

Minimise $f(y) = b^T y - \mu \log\det(\mathcal{A}^T y - C)$ by Newton:

1. Sufficient first order optimality conditions:

$\nabla f(y) = 0$, use $\nabla_Z \log\det Z = Z^{-1}$

$$\boxed{b - \mathcal{A}[\mu(\mathcal{A}^T y - C)^{-1}] = 0}$$

yields methods of Jarre 1993, Nesterov and Nemirovskii 1994

Primal-dual approach: $Z = \mathcal{A}^T y - C$, $X = \mu Z^{-1}$

$$\boxed{\begin{array}{ll} b - \mathcal{A}X = 0 & \\ Z = \mathcal{A}^T y - C & \\ XZ = \mu I & X, Z \succ 0 \end{array}}$$

[primal feasibility]
[dale feasibility]
[perturbed complementarity]

Minimise $\quad f(y) = b^T y - \mu \log\det(\mathcal{A}^T y - C) \quad$ by Newton:

1. Sufficient first order optimality conditions:

$\nabla f(y) = 0$, use $\quad \nabla_Z \log\det Z = Z^{-1}$

$$\boxed{b - \mathcal{A}[\mu(\mathcal{A}^T y - C)^{-1}] = 0}$$

yields methods of Jarre 1993, Nesterov and Nemirovskii 1994

Primal-dual approach: $Z = \mathcal{A}^T y - C$, $X = \mu Z^{-1}$

$$\boxed{\begin{array}{ll} b - \mathcal{A}X = 0 & \\ Z = \mathcal{A}^T y - C & \\ XZ = \mu I & X, Z \succ 0 \end{array}}$$

[primal feasibility]
[dale feasibility]
[perturbed complementarity]

- Each $\mu > 0$ has a unique solution (X_μ, y_μ, Z_μ) [requires (S) and (A)]
- (y_μ, Z_μ) is the optimal solution of the dual barrier problem
- X_μ is the optimal solution of the primal barrier-problem
- The curve $\{(X_\mu, y_\mu, Z_\mu) \colon \mu > 0\}$ is the *central path*.

Minimise $f(y) = b^T y - \mu \log\det(\mathcal{A}^T y - C)$ by Newton:

1. Sufficient first order optimality conditions:

$\nabla f(y) = 0$, use $\nabla_Z \log\det Z = Z^{-1}$

$$\boxed{b - \mathcal{A}[\mu(\mathcal{A}^T y - C)^{-1}] = 0}$$

yields methods of Jarre 1993, Nesterov and Nemirovskii 1994

Primal-dual approach: $Z = \mathcal{A}^T y - C$, $X = \mu Z^{-1}$

$$\begin{array}{ll} b - \mathcal{A}X = 0 & \\ Z = \mathcal{A}^T y - C & \\ XZ = \mu I & X, Z \succ 0 \end{array}$$

[primal feasibility]
[dale feasibility]
[perturbed complementarity]

- Each $\mu > 0$ has a unique solution (X_μ, y_μ, Z_μ) [requires (S) and (A)]
- (y_μ, Z_μ) is the optimal solution of the dual barrier problem
- X_μ is the optimal solution of the primal barrier-problem
- The curve $\{(X_\mu, y_\mu, Z_\mu) \colon \mu > 0\}$ is the *central path*.

[Maximising the determinant ($\mu = 1$) is relevant on its own!]

2. Primal-dual Linearisation

$$
\begin{array}{rc}
\mathrm{I} & b - \mathcal{A}(X + \Delta X) = 0 \\
\mathrm{II} & Z + \Delta Z = \mathcal{A}^T(y + \Delta y) - C \\
\mathrm{III} & XZ + X\Delta Z + \Delta X Z = \mu I
\end{array}
$$

Solution $(\Delta X, \Delta y, \Delta Z)$ is the step direction/Newton step

Difficulty: ΔZ is symmetric [II] but i.g. ΔX is not [III]

2. Primal-dual Linearisation

$$\begin{array}{ll} \text{I} & b - \mathcal{A}(X + \Delta X) = 0 \\ \text{II} & Z + \Delta Z = \mathcal{A}^T(y + \Delta y) - C \\ \text{III} & XZ + X\Delta Z + \Delta X Z = \mu I \end{array}$$

Solution $(\Delta X, \Delta y, \Delta Z)$ is the step direction/Newton step

Difficulty: ΔZ is symmetric [II] but i.g. ΔX is not [III]
Suggestions:

(a) HRVW96/KSH97/M97: use the symmetric part,

$$\frac{1}{2}(\Delta X + \Delta X^T)$$

(b) NT97: scale III by the matrix $W = X^{\frac{1}{2}}(X^{\frac{1}{2}} Z X^{\frac{1}{2}})^{-\frac{1}{2}} X^{\frac{1}{2}}$,

$$W^{-1} \Delta X W^{-1} + \Delta Z = \mu X^{-1} - Z$$

(c) AHO98: symmetrise III directly,

$$X\Delta Z + \Delta Z X + \Delta X Z + Z \Delta X = 2\mu I - XZ - ZX$$

Several others exist, all differ slightly.(Todd 1999).

Algor. Scheme for Primal-Dual Interior-Point Methods

Input: $\mathcal{A}$, b, C, starting point (X^0, y^0, Z^0) with $X^0 \succ 0$ and $Z^0 \succ 0$

1. Choose $\mu = \sigma \frac{\langle X,Z \rangle}{n}$ with $\sigma \in (0, 1]$.
2. Compute $(\Delta X, \Delta y, \Delta Z)$.
3. Line search: determine $\alpha \in (0, 1]$ with $X + \alpha \Delta X \succ 0$ and $Z + \alpha \Delta Z \succ 0$.
4. Put $(X, y, Z) := (X + \alpha \Delta X, y + \alpha \Delta y, Z + \alpha \Delta Z)$.
5. If $\|\mathcal{A}X - b\|$, $\|\mathcal{A}^T y + Z - C\|$ and $\langle X, Z \rangle$ "sufficiently small", **stop**, otherwise **goto** 1.

Algor. Scheme for Primal-Dual Interior-Point Methods

Input: $\mathcal{A}$, b, C, starting point (X^0, y^0, Z^0) with $X^0 \succ 0$ and $Z^0 \succ 0$

1. Choose $\mu = \sigma \frac{\langle X, Z \rangle}{n}$ with $\sigma \in (0, 1]$.
2. Compute $(\Delta X, \Delta y, \Delta Z)$.
3. Line search: determine $\alpha \in (0, 1]$ with $X + \alpha \Delta X \succ 0$ and $Z + \alpha \Delta Z \succ 0$.
4. Put $(X, y, Z) := (X + \alpha \Delta X, y + \alpha \Delta y, Z + \alpha \Delta Z)$.
5. If $\|\mathcal{A}X - b\|$, $\|\mathcal{A}^T y + Z - C\|$ and $\langle X, Z \rangle$ "sufficiently small", **stop**, otherwise **goto** 1.

Theorem (Kojima, Shindoh and Hara 1997)

(X^0, y^0, Z^0) *feasible and "centred". Choose* $\sigma = 1 - \frac{0.35}{\sqrt{n}}$ *and* $\alpha = 1$, *then each step is feasible and* $\langle X, Z \rangle < \varepsilon$ *in*

$$\boxed{O\left(\sqrt{n} \log \frac{\langle X^0, Z^0 \rangle}{\varepsilon}\right)} \text{ iterations.}$$

[$O(n^2)$ Variable!]

Work per Iteration

Computation of the HRVW/KSH/M step:

$$\Delta Z = \mathcal{A}^T(y + \Delta y) - C - Z$$
$$\Delta X = \mu Z^{-1} - X - X \Delta Z Z^{-1}$$

solve $\qquad \mathcal{A}(X \mathcal{A}^T(\Delta y) Z^{-1}) = M \Delta y = \ldots$

Work per Iteration

Computation of the HRVW/KSH/M step:

$$\Delta Z = \mathcal{A}^T(y + \Delta y) - C - Z$$
$$\Delta X = \mu Z^{-1} - X - X \Delta Z Z^{-1}$$

solve $\quad \mathcal{A}(X \mathcal{A}^T(\Delta y) Z^{-1}) = M \Delta y = \dots$

with $\quad \boxed{M_{ij} = \operatorname{tr} X A_i Z^{-1} A_j} \qquad [\operatorname{tr} B = \sum B_{ii}]$

X and Z^{-1} are in general dense.

$\Rightarrow$ M is a dense positive definite matrix of order m.
Cholesky factorisation needs $m^3/3$ flops and $O(m^2)$ memory.

Work per Iteration

Computation of the HRVW/KSH/M step:

$$\Delta Z = \mathcal{A}^T(y + \Delta y) - C - Z$$
$$\Delta X = \mu Z^{-1} - X - X \Delta Z Z^{-1}$$

solve $\mathcal{A}(X\mathcal{A}^T(\Delta y)Z^{-1}) = M\Delta y = \dots$

with $\boxed{M_{ij} = \operatorname{tr} X A_i Z^{-1} A_j}$ $\quad [\operatorname{tr} B = \sum B_{ii}]$

X and Z^{-1} are in general dense.

$\Rightarrow$ M is a dense positive definite matrix of order m.
Cholesky factorisation needs $m^3/3$ flops and $O(m^2)$ memory.

Line search: requires one to three Cholesky factorisations for dense X (Z is frequently sparse) with $n^3/3$ flops and $O(n^2)$ storage.

Work per Iteration

Computation of the HRVW/KSH/M step:

$$\Delta Z = \mathcal{A}^T(y + \Delta y) - C - Z$$
$$\Delta X = \mu Z^{-1} - X - X \Delta Z Z^{-1}$$

solve $\mathcal{A}(X \mathcal{A}^T(\Delta y) Z^{-1}) = M \Delta y = \dots$

with $\boxed{M_{ij} = \operatorname{tr} X A_i Z^{-1} A_j}$ $[\operatorname{tr} B = \sum B_{ii}]$

X and Z^{-1} are in general dense.

$\Rightarrow$ M is a dense positive definite matrix of order m.
Cholesky factorisation needs $m^3/3$ flops and $O(m^2)$ memory.

Line search: requires one to three Cholesky factorisations for dense X (Z is frequently sparse) with $n^3/3$ flops and $O(n^2)$ storage.

In practice: split matrices into blocks exploiting semidefinite completion techniques, exploit structure induced by these blocks, solve the big system without M via iterative methods, ...

Some General References

- Miguel F. Anjos and Jean B. Lasserre (eds.), *Handbook of semidefinite, conic and polynomial optimization*, International Series in Operations Research & Management Science, vol. 166, Springer, 2012.
- Aharon Ben-Tal and Arkadi Nemirovski, *Lectures on modern convex optimization. Analysis, algorithms, and engineering applications*, MPS/ SIAM Series on Optimization, SIAM, Philadelphia, 2001.
- Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004, Reprinted 2007 with corrections.
- Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe (eds.), *Handbook of semidefinite programming*, International Series in Operations Research and Management Science, vol. 27, Kluwer Academic Publishers, Boston/Dordrecht/London, 2000.