

Inhaltsübersicht für heute:

Semidefinite Optimierung

SDP Anwendung: Robuste Stabilität dynamischer Systeme

SDP Anwendung: Entwurf von Experimenten

(SDP Anwendung: Graphenpartition)

(SDP Anwendung: geometrische Einbettungen)

Inhaltsübersicht für heute:

Semidefinite Optimierung

SDP Anwendung: Robuste Stabilität dynamischer Systeme

SDP Anwendung: Entwurf von Experimenten

(SDP Anwendung: Graphenpartition)

(SDP Anwendung: geometrische Einbettungen)

Positiv semidefinite Matrizen

Eine symmetrische Matrix $A \in S^n := \{A \in \mathbb{R}^{n \times n} : A = A^T\}$ heißt **positiv semidefinit**, falls $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; wir schreiben $A \in S_+^n$ oder $A \succeq 0$.

Sie heißt **positiv definit** ($A \in S_{++}^n$, $A \succ 0$), falls $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
[Für $A \succeq 0$ ($\succ 0$) und $J \subseteq \{1, \dots, n\}$ ist $A_{J,J} \succeq 0$ ($\succ 0$).]

Positiv semidefinite Matrizen

Eine symmetrische Matrix $A \in S^n := \{A \in \mathbb{R}^{n \times n} : A = A^T\}$ heißt **positiv semidefinit**, falls $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; wir schreiben $A \in S_+^n$ oder $A \succeq 0$.

Sie heißt **positiv definit** ($A \in S_{++}^n$, $A \succ 0$), falls $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
 [Für $A \succeq 0$ ($\succ 0$) und $J \subseteq \{1, \dots, n\}$ ist $A_{J,J} \succeq 0$ ($\succ 0$).]

$\lambda \in \mathbb{R}$ heißt Eigenwert und $v \in \mathbb{R}^n \setminus \{0\}$ Eigenvektor von A , falls $Av = \lambda v$.
 Für jedes $A \in S^n$ gibt es eine Eigenwertzerlegung $A = P \Lambda P^T$ mit reellem $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$ und $P \in \mathbb{R}^{n \times n}$ orthogonal (d.h., $P^T P = I$).
 Für $P = [v_1, \dots, v_n]$ ist $A = P \Lambda P^T = \sum_{i=1}^n \lambda_i v_i v_i^T$.

Positiv semidefinite Matrizen

Eine symmetrische Matrix $A \in S^n := \{A \in \mathbb{R}^{n \times n} : A = A^T\}$ heißt **positiv semidefinit**, falls $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; wir schreiben $A \in S_+^n$ oder $A \succeq 0$.

Sie heißt **positiv definit** ($A \in S_{++}^n$, $A \succ 0$), falls $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
 [Für $A \succeq 0$ ($\succ 0$) und $J \subseteq \{1, \dots, n\}$ ist $A_{J,J} \succeq 0$ ($\succ 0$).]

$\lambda \in \mathbb{R}$ heißt Eigenwert und $v \in \mathbb{R}^n \setminus \{0\}$ Eigenvektor von A , falls $Av = \lambda v$.
 Für jedes $A \in S^n$ gibt es eine Eigenwertzerlegung $A = P \Lambda P^T$ mit reellem $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$ und $P \in \mathbb{R}^{n \times n}$ orthogonal (d.h., $P^T P = I$).
 Für $P = [v_1, \dots, v_n]$ ist $A = P \Lambda P^T = \sum_{i=1}^n \lambda_i v_i v_i^T$.

Für $A, B \in S^n$ verwenden wir als inneres Produkt

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

Positiv semidefinite Matrizen

Eine symmetrische Matrix $A \in S^n := \{A \in \mathbb{R}^{n \times n} : A = A^T\}$ heißt **positiv semidefinit**, falls $v^T A v \geq 0 \quad \forall v \in \mathbb{R}^n$; wir schreiben $A \in S^n_+$ oder $A \succeq 0$.

Sie heißt **positiv definit** ($A \in S^n_{++}$, $A \succ 0$), falls $v^T A v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.
 [Für $A \succeq 0$ ($\succ 0$) und $J \subseteq \{1, \dots, n\}$ ist $A_{J,J} \succeq 0$ ($\succ 0$).]

$\lambda \in \mathbb{R}$ heißt Eigenwert und $v \in \mathbb{R}^n \setminus \{0\}$ Eigenvektor von A , falls $Av = \lambda v$.
 Für jedes $A \in S^n$ gibt es eine Eigenwertzerlegung $A = P \Lambda P^T$ mit reellem $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n)$ und $P \in \mathbb{R}^{n \times n}$ orthogonal (d.h., $P^T P = I$).
 Für $P = [v_1, \dots, v_n]$ ist $A = P \Lambda P^T = \sum_{i=1}^n \lambda_i v_i v_i^T$.

Für $A, B \in S^n$ verwenden wir als inneres Produkt

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

Satz

Für $A \in S^n$ sind folgende Bedingungen äquivalent:

- $A \succeq 0$,
- $\lambda_i(A) \geq 0, i = 1, \dots, n$,
- $A = C^T C$ für ein $C \in \mathbb{R}^{k \times n}$,
- $\langle A, B \rangle \geq 0 \quad \forall B \succeq 0$.

$$[\Rightarrow \det(A) \geq 0]$$

$$[\text{Es gilt: } \text{Rang}(A) = \text{Rang}(C)]$$

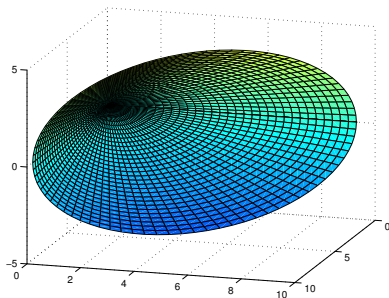
Der Kegel der positiv semidefiniten Matrizen

Die positiv semidefiniten Matrizen S_+^n bilden einen konvexen Kegel, denn für $X, Y \in S_+^n$, $\alpha \geq 0$ ist $\forall v \in \mathbb{R}^n$

$$v^T(\alpha(X + Y))v = \alpha(v^T X v + v^T Y v) \geq 0.$$

Aus $A \in S_+^n \Leftrightarrow \langle A, B \rangle \geq 0 \forall B \succeq 0$ folgt:
 S_+^n ist **selbstdual**, $(S_+^n)^* = S_+^n$.

Bild rechts: $S_+^2 = \left[\begin{array}{cc} x & z \\ z & y \end{array} \right] \succeq 0$.



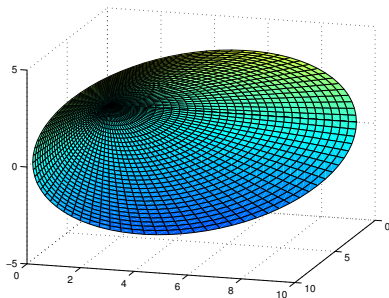
Der Kegel der positiv semidefiniten Matrizen

Die positiv semidefiniten Matrizen S_+^n bilden einen konvexen Kegel, denn für $X, Y \in S_+^n$, $\alpha \geq 0$ ist $\forall v \in \mathbb{R}^n$

$$v^T(\alpha(X + Y))v = \alpha(v^T X v + v^T Y v) \geq 0.$$

Aus $A \in S_+^n \Leftrightarrow \langle A, B \rangle \geq 0 \forall B \succeq 0$ folgt:
 S_+^n ist **selbstdual**, $(S_+^n)^* = S_+^n$.

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



• Ist $B \in \mathbb{R}^{n \times n}$ regulär (=invertierbar), gilt $X \succeq 0 \Leftrightarrow B^T X B \succeq 0$.

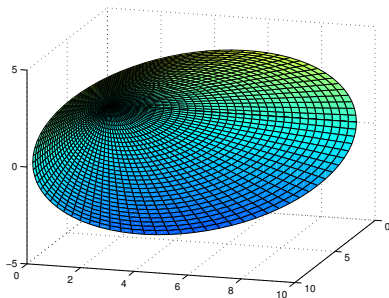
Der Kegel der positiv semidefiniten Matrizen

Die positiv semidefiniten Matrizen S_+^n bilden einen konvexen Kegel, denn für $X, Y \in S_+^n$, $\alpha \geq 0$ ist $\forall v \in \mathbb{R}^n$

$$v^T(\alpha(X + Y))v = \alpha(v^T X v + v^T Y v) \geq 0.$$

Aus $A \in S_+^n \Leftrightarrow \langle A, B \rangle \geq 0 \forall B \succeq 0$ folgt:
 S_+^n ist **selbstdual**, $(S_+^n)^* = S_+^n$.

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



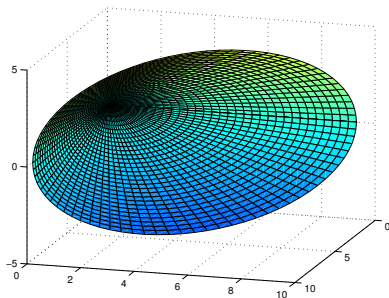
- Ist $B \in \mathbb{R}^{n \times n}$ regulär (=invertierbar), gilt $X \succeq 0 \Leftrightarrow B^T X B \succeq 0$.
- Für $A - B \succeq 0$ schreiben wir auch $A \succeq B$.

Der Kegel der positiv semidefiniten Matrizen

Die positiv semidefiniten Matrizen S_+^n bilden einen konvexen Kegel, denn für $X, Y \in S_+^n$, $\alpha \geq 0$ ist $\forall v \in \mathbb{R}^n$
 $v^T(\alpha(X + Y))v = \alpha(v^T X v + v^T Y v) \geq 0$.

Aus $A \in S_+^n \Leftrightarrow \langle A, B \rangle \geq 0 \forall B \succeq 0$ folgt:
 S_+^n ist **selbstdual**, $(S_+^n)^* = S_+^n$.

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



- Ist $B \in \mathbb{R}^{n \times n}$ regulär (=invertierbar), gilt $X \succeq 0 \Leftrightarrow B^T X B \succeq 0$.
- Für $A - B \succeq 0$ schreiben wir auch $A \succeq B$.

In der Modellierung semidefiniter Programme besonders nützlich:

Satz (Schur-Komplement)

Für $A \in S_{++}^m$, $C \in S_+^n$ und $B \in \mathbb{R}^{m \times n}$ gilt

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \quad (\text{bzw. } \succ 0) \iff C \succeq B^T A^{-1} B \quad (\text{bzw. } \succ 0)$$

LP \leftrightarrow Semidefinites Programm

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & X \succeq 0 \end{aligned}$$

$$\begin{aligned} x &\in \mathbb{R}_+^n \\ c^T x &= \sum_i c_i x_i \\ Ax &= \begin{pmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{pmatrix} \\ A^T y &= \sum_i a_i y_i \end{aligned}$$

$$\begin{aligned} X &\in S_+^n \\ \langle C, X \rangle &= \sum_{i,j} C_{ij} X_{ij} \\ \mathcal{A}X &= \begin{pmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{pmatrix} \\ \mathcal{A}^T y &= \sum_i A_i y_i \end{aligned}$$

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

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

Semidefinites Programm (SDP) in Normalform

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

Ist eines der beiden streng zulässig, gilt $v(P) = v(D)$.

Semidefinites Programm (SDP) in Normalform

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

Ist eines der beiden streng zulässig, gilt $v(P) = v(D)$.

In der Praxis gibt es oft mehrere $X_i \succeq 0$, für die Theorie reicht eine:

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

\Rightarrow Semidefinite Optimierung enthält Lineare Optimierung ($X_i \in S_+^1$).

Semidefinites Programm (SDP) in Normalform

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

Ist eines der beiden streng zulässig, gilt $v(P) = v(D)$.

In der Praxis gibt es oft mehrere $X_i \succeq 0$, für die Theorie reicht eine:

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

\Rightarrow Semidefinite Optimierung enthält Lineare Optimierung ($X_i \in S_+^1$).

Mit semidefiniter Optimierung sind auch SOC-Bedingungen darstellbar:

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

[für $x_0 = 0$ direkt nachprüfen]

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

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

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

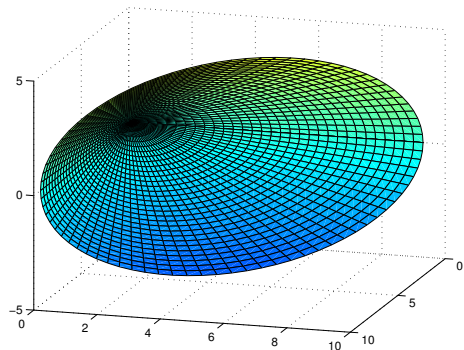


Illustration: $X \in S_+^2$ geschnitten mit $\langle A, X \rangle = \beta$

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

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

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

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

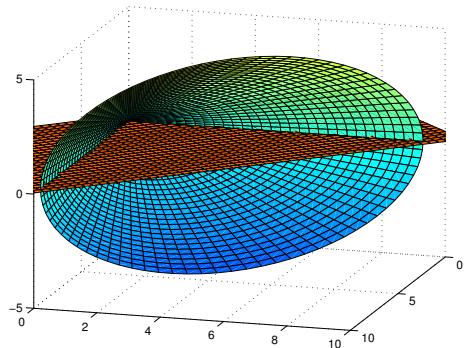


Illustration: $X \in S_+^2$ geschnitten mit $\langle A, X \rangle = \beta$

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

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

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

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

$$A \succ 0, \beta > 0$$

→ „Normalfall Ellipse“

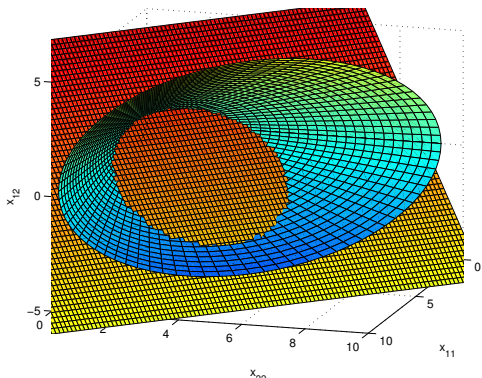


Illustration: $X \in S_+^2$ geschnitten mit $\langle A, X \rangle = \beta$

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

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

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

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

$$A \succ 0, \beta > 0$$

→ „Normalfall Ellipse“

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

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

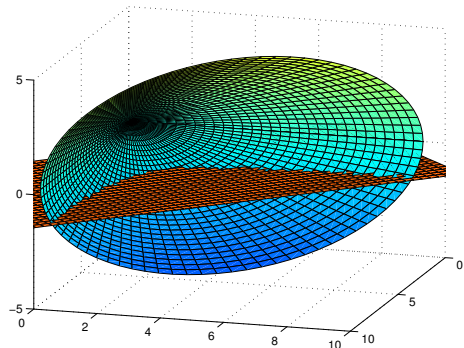


Illustration: $X \in S_+^2$ geschnitten mit $\langle A, X \rangle = \beta$

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

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

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

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

$$A \succ 0, \beta > 0$$

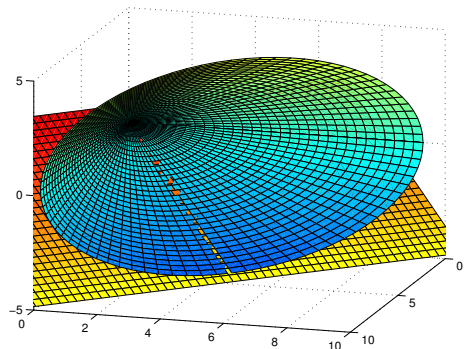
→ „Normalfall Ellipse“

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

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

$$A = vv^T, \beta = 0 \rightarrow v \text{ EV zu } \lambda_1 = 0$$

Nur Randpunkte, numerisch schwer!



Beispiel mit Dualitätslücke

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

Entsprechende Matrizen:

$$\begin{array}{l} C = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ A_1 = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

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

$x_{11} = 0 \Rightarrow x_{12} = 0$, primale Optimallösung ist 0.

$x_{22} = 0 \Rightarrow \frac{1+y_1}{2} = 0$, duale Optimallösung ist -1 .

Problem: Primales Problem instabil

$$\min \quad x_{12}$$

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

$$\max \quad y_1 + \varepsilon y_2$$

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

Entsprechende Matrizen:

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

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

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

$$A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

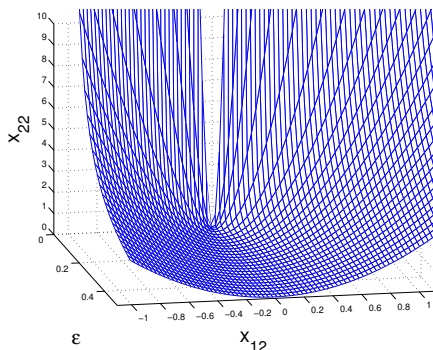
$$A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

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

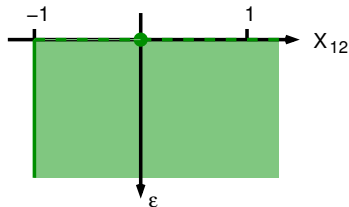
$$x_{33} \geq 0 \Rightarrow x_{12} \geq -1, x_{22} \geq \frac{x_{12}^2}{\varepsilon}, \quad \text{primale Optimallösung ist } -1.$$

$$z_{22} = 0 \Rightarrow \frac{1+y_1}{2} = 0, y_2 = 0, \quad \text{duale Optimallösung ist } -1.$$

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



Projektion auf (ε, x_{12}) -Ebene:



Für $\varepsilon > 0$ ist $x_{12} \in [-1, -\infty)$,
für $\varepsilon = 0$ nur noch $x_{12} \in \{0\}$
zulässig!

Mathem. Grund: die Menge $\left\{ \begin{bmatrix} \langle C, X \rangle \\ AX \end{bmatrix} : X \succeq 0 \right\}$ ist nicht abgeschlossen.

SDP und Eigenwertoptimierung

Für $A \in S^n$ bezeichne $\lambda_{\min}(A) := \lambda_1(A) \leq \dots \leq \lambda_n(A) =: \lambda_{\max}(A)$. Es gilt $\lambda_i(A + y_0 I) = \lambda_i(A) + y_0$ für $i = 1, \dots, n$ und $y_0 \in \mathbb{R}$.

SDP und Eigenwertoptimierung

Für $A \in S^n$ bezeichne $\lambda_{\min}(A) := \lambda_1(A) \leq \dots \leq \lambda_n(A) =: \lambda_{\max}(A)$. Es gilt $\lambda_i(A + y_0 I) = \lambda_i(A) + y_0$ für $i = 1, \dots, n$ und $y_0 \in \mathbb{R}$.

In der Optimalsteuerung ist die Stabilität eines Systems gewährleistet, wenn man für die von Steuerparametern $y \in \mathbb{R}^m$ abhängige Systemmatrix $A(y)$ nachweisen kann, dass $\lambda_{\max}(A(y)) < 0$.

Ist $A(y)$ affin, etwa $A(y) := C - \sum_{i=1}^m y_i A_i$ mit $C, A_i \in S^n$, führt das auf

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

SDP und Eigenwertoptimierung

Für $A \in S^n$ bezeichne $\lambda_{\min}(A) := \lambda_1(A) \leq \dots \leq \lambda_n(A) =: \lambda_{\max}(A)$. Es gilt $\lambda_i(A + y_0 I) = \lambda_i(A) + y_0$ für $i = 1, \dots, n$ und $y_0 \in \mathbb{R}$.

In der Optimalsteuerung ist die Stabilität eines Systems gewährleistet, wenn man für die von Steuerparametern $y \in \mathbb{R}^m$ abhängige Systemmatrix $A(y)$ nachweisen kann, dass $\lambda_{\max}(A(y)) < 0$.

Ist $A(y)$ affin, etwa $A(y) := C - \sum_{i=1}^m y_i A_i$ mit $C, A_i \in S^n$, führt das auf

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

Zur Modellierung als SDP: $\lambda_{\max}(A) = -\lambda_{\min}(-A)$ und

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

Wegen $Z \succeq 0 \Leftrightarrow \lambda_{\min}(Z) \geq 0$ gilt daher

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

Lineare Matrix Ungleichungen (LMI)

Eine Bedingung der Form

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

mit $A_i, C \in S^n$ heißt **lineare Matrix Ungleichung** (Linear Matrix Inequality).

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

Lineare Matrix Ungleichungen (LMI)

Eine Bedingung der Form

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

mit $A_i, C \in S^n$ heißt **lineare Matrix Ungleichung** (Linear Matrix Inequality).

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

Bsp: Die **Lyapunov Ungleichung** fordert für festes $P = [p_1, \dots, p_n] \in \mathbb{R}^{n \times n}$

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

In LMI-Darstellung wäre $y = [x_{11}, x_{12}, \dots, x_{1n}, x_{22}, x_{23}, \dots, x_{nn}]^T$, aber es wäre umständlich/sinnlos, die A_i für diese Ungleichungen anzugeben, die Struktur lässt sich im SDP besser direkt nutzen.

Lineare Matrix Ungleichungen (LMI)

Eine Bedingung der Form

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

mit $A_i, C \in S^n$ heißt **lineare Matrix Ungleichung** (Linear Matrix Inequality).

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

Bsp: Die **Lyapunov Ungleichung** fordert für festes $P = [p_1, \dots, p_n] \in \mathbb{R}^{n \times n}$

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

In LMI-Darstellung wäre $y = [x_{11}, x_{12}, \dots, x_{1n}, x_{22}, x_{23}, \dots, x_{nn}]^T$, aber es wäre umständlich/sinnlos, die A_i für diese Ungleichungen anzugeben, die Struktur lässt sich im SDP besser direkt nutzen.

Um LMIs zu erkennen, reicht es festzustellen, dass die Matrizen linear von den jeweiligen Variablen abhängen:

Die Matrix-Multiplikation $P^T X$ (bzw. XP) ist linear in X .

Wie erzwingt man positive Definitheit?

Die semidefinite Barrierefunktion – $-\log \det X$

Wegen $\det X = \prod_{k=1}^n \lambda_k(X)$ ist

$$-\log \det X = -\log \prod_{k=1}^n \lambda_k(X) = -\sum_{k=1}^n \log \lambda_k(X)$$

eine Barrierefunktion für $X \succeq 0$ ($\Leftrightarrow \lambda_i(X) \geq 0$) und nur für $X \succ 0$ ($\Leftrightarrow \lambda_i(X) > 0$) erklärt. [\[vergl. Barriere – \$\sum \log x_i\$ für \$x \geq 0\$ in LP\]](#)

Innere-Punkte-Verfahren für SDP nutzen diese Barrierefunktion. [\[wie LP!\]](#)

Die semidefinite Barrierefunktion – $-\log \det X$

Wegen $\det X = \prod_{k=1}^n \lambda_k(X)$ ist

$$-\log \det X = -\log \prod_{k=1}^n \lambda_k(X) = -\sum_{k=1}^n \log \lambda_k(X)$$

eine Barrierefunktion für $X \succeq 0$ ($\Leftrightarrow \lambda_i(X) \geq 0$) und nur für $X \succ 0$ ($\Leftrightarrow \lambda_i(X) > 0$) erklärt. [\[vergl. Barriere – \$\sum \log x_i\$ für \$x \geq 0\$ in LP\]](#)

Innere-Punkte-Verfahren für SDP nutzen diese Barrierefunktion. [\[wie LP!\]](#)

$-\log \det X$ wird auch genutzt, um streng zulässige Lösungen zu finden, z.B.

$$\begin{aligned} \min \quad & -\log \det X - \log \det Z \\ \text{s.t.} \quad & Z = P^T X + P X \\ & X \succ 0, Z \succ 0 \end{aligned}$$

→ Innere-Punkte-Verfahren für konstanten Barriereparameter $\mu = 1$.

Das Maximieren der Determinante hat zahlreiche Anwendungen, insbesondere im Entwurf von Experimenten (*experiment design*)

Anwendungen der Semidefiniten Optimierung

- Optimalsteuerung und Kontrolltheorie
- Signalverarbeitung
- Kombinatorische Optimierung
- Globale Optimierung über Polynomen
- Robuster Entwurf von Stabkonstruktionen (truss topology design)
- Entwurf von Materialien (free material design)
- Robuste Optimierung
- Momenten-Probleme in der Wahrscheinlichkeitstheorie
- Entwurf von Experimenten in der Statistik
- Eigenwert-Optimierung
- Optimierung (trust-region Bestimmung, quadratische Relaxationen)

Inhaltsübersicht für heute:

Semidefinite Optimierung

SDP Anwendung: Robuste Stabilität dynamischer Systeme

SDP Anwendung: Entwurf von Experimenten

(SDP Anwendung: Graphenpartition)

(SDP Anwendung: geometrische Einbettungen)

Robuste Stabilität dynamischer Systeme

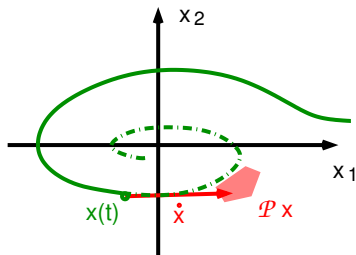
In einem (homogenen linearen) dynamischen System mit unsicheren Daten,

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

- beschreibt
- $x(t)$... Zustand des Systems zur Zeit t .
 - $\dot{x} := \frac{d}{dt}x(t)$... (infinitesimale) Veränderung von $x(\cdot)$
 - $P(t)$... unsichere Übergangsmatrix zur Zeit t .

(DS) heißt **stabil**, wenn $x(t) \rightarrow 0$ für $t \rightarrow \infty$ und beliebige $P(t) \in \mathcal{P}$.

[In der Regelungstechnik würde \mathcal{P} die möglichen Auswirkungen der Regelung/Steuerung umfassen. Man möchte wissen, ob diese auch bei unsauberer Realisierung in der Praxis den Zweck erfüllt.]



Robuste Stabilität dynamischer Systeme

In einem (homogenen linearen) dynamischen System mit unsicheren Daten,

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

beschreibt

- $x(t)$... Zustand des Systems zur Zeit t .
- $\dot{x} := \frac{d}{dt}x(t)$... (infinitesimale) Veränderung von $x(\cdot)$
- $P(t)$... unsichere Übergangsmatrix zur Zeit t .

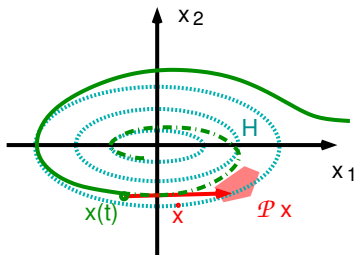
(DS) heißt **stabil**, wenn $x(t) \rightarrow 0$ für $t \rightarrow \infty$ und beliebige $P(t) \in \mathcal{P}$.

Hinreichend: Es gibt eine Norm

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

mit $\frac{d}{dt} \|x(t)\|_H^2 < 0$ auf allen Trajektorien

(das System heißt dann **quadratisch stabil**,
 $x^T H x$ **quadratische Lyapunov Funktion**).



Robuste Lyapunov Stabilität über SDP

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

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

Robuste Lyapunov Stabilität über SDP

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

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

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

Robuste Lyapunov Stabilität über SDP

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

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

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

Falls $A := P^T H + H P \prec 0$ (negativ definit), gilt $v^T A v < 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$.

Robuste Lyapunov Stabilität über SDP

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

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

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

Falls $A := P^T H + H P \prec 0$ (negativ definit), gilt $v^T A v < 0 \forall v \in \mathbb{R}^n \setminus \{0\}$.
 \Rightarrow Das System ist quadratisch stabil, falls

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

zulässig lösbar ist, denn für so ein H erfüllt auch jede Konvexkombination $P \in \mathcal{P}$ die Bedingung $P^T H + H P \prec 0$.

Robuste Lyapunov Stabilität über SDP

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

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

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

Falls $A := P^T H + H P \prec 0$ (negativ definit), gilt $v^T A v < 0 \forall v \in \mathbb{R}^n \setminus \{0\}$.
 \Rightarrow Das System ist quadratisch stabil, falls

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

zulässig lösbar ist, denn für so ein H erfüllt auch jede Konvexkombination $P \in \mathcal{P}$ die Bedingung $P^T H + H P \prec 0$.

Suche H über Determinanten-Maximierung oder Eigenwert-Optimierung:

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

Inhaltsübersicht für heute:

Semidefinite Optimierung

SDP Anwendung: Robuste Stabilität dynamischer Systeme

SDP Anwendung: Entwurf von Experimenten

(SDP Anwendung: Graphenpartition)

(SDP Anwendung: geometrische Einbettungen)

Entwurf von Experimenten

Um die Werte eines Parametervektors $\xi \in \mathbb{R}^p$ zu schätzen, stehen $\mathcal{R} = \{r_i \in \mathbb{R}^p : i = 1, \dots, n\}$ mögliche Experimente zur Verfügung. Experiment i liefert pro Durchführung einen Messwert $r_i^T \xi + \rho_i$ mit unabhängig ($\mu = 0, \sigma^2 = 1$)-normalverteiltem Messfehler ρ_i .

Entwurf von Experimenten

Um die Werte eines Parametervektors $\xi \in \mathbb{R}^p$ zu schätzen, stehen $\mathcal{R} = \{r_i \in \mathbb{R}^p : i = 1, \dots, n\}$ mögliche Experimente zur Verfügung. Experiment i liefert pro Durchführung einen Messwert $r_i^T \xi + \rho_i$ mit unabhängig $(\mu = 0, \sigma^2 = 1)$ -normalverteiltem Messfehler ρ_i .

Werden m Experimente $a_j \in \mathcal{R}$ (Wiederholungen sind erlaubt) mit Ergebnissen $\eta_j = a_j^T \xi + \rho_j$ durchgeführt, ergibt der Maximum-Likelihood-Schätzer bei $\text{Rang}[a_1, \dots, a_m] = n$ ein geschätztes

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

dessen Fehlerverteilung Erwartungswert 0 und Kovarianzmatrix G hat.

Entwurf von Experimenten

Um die Werte eines Parametervektors $\xi \in \mathbb{R}^p$ zu schätzen, stehen $\mathcal{R} = \{r_i \in \mathbb{R}^p : i = 1, \dots, n\}$ mögliche Experimente zur Verfügung. Experiment i liefert pro Durchführung einen Messwert $r_i^T \xi + \rho_i$ mit unabhängig ($\mu = 0, \sigma^2 = 1$)-normalverteiltem Messfehler ρ_i .

Werden m Experimente $a_j \in \mathcal{R}$ (Wiederholungen sind erlaubt) mit Ergebnissen $\eta_j = a_j^T \xi + \rho_j$ durchgeführt, ergibt der Maximum-Likelihood-Schätzer bei $\text{Rang}[a_1, \dots, a_m] = n$ ein geschätztes

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

dessen Fehlerverteilung Erwartungswert 0 und Kovarianzmatrix G hat.

Sind G und G' zwei Kovarianzmatrizen dieser Art und gilt $G \preceq G'$, dann ist die zu G gehörende Experimentfolge besser, weil die Varianz des Schätzfehlers kleiner ist.

→ Finde die bzgl. \preceq minimalen Elemente von

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

Relaxationen

Statt m Experimente ganzzahlig zu wählen, bestimmt man relative Anteile,

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

Relaxationen

Statt m Experimente ganzzahlig zu wählen, bestimmt man relative Anteile,

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

Es gibt mehrere Ansätze, ein bzgl. \preceq minimales G zu finden. Interpretiere G dazu als ein „Konfidenzellipsoid“ mit Halbachsenlängen $\lambda_j(G)$,

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

D -optimales Design: Minimiere das Volumen des Konfidenzellipsoids.

E -optimales Design: Minimiere die längste Halbachse.

A -optimales Design: Minimiere die Summe der Halbachsen.

Relaxationen

Statt m Experimente ganzzahlig zu wählen, bestimmt man relative Anteile,

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

Es gibt mehrere Ansätze, ein bzgl. \preceq minimales G zu finden. Interpretiere G dazu als ein „Konfidenzellipsoid“ mit Halbachsenlängen $\lambda_j(G)$,

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

D-optimales Design: Minimiere das Volumen des Konfidenzellipsoids.

E-optimales Design: Minimiere die längste Halbachse.

A-optimales Design: Minimiere die Summe der Halbachsen.

D-optimales Design. Das Volumen ist zu $\det G = \prod \lambda_j(G)$ proportional. Wegen $\det(G^{-1}) = \det(G)^{-1} \Leftrightarrow$ maximiere die Determinante von G^{-1} ,

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

Relaxationen

Statt m Experimente ganzzahlig zu wählen, bestimmt man relative Anteile,

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

Es gibt mehrere Ansätze, ein bzgl. \preceq minimales G zu finden. Interpretiere G dazu als ein „Konfidenzellipsoid“ mit Halbachsenlängen $\lambda_j(G)$,

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

D-optimales Design: Minimiere das Volumen des Konfidenzellipsoids.

E-optimales Design: Minimiere die längste Halbachse.

A-optimales Design: Minimiere die Summe der Halbachsen.

E-optimales Design. Die längste Halbachse ist $\lambda_{\max}(G)$.

Wegen $\lambda_{\min}(G^{-1}) = \lambda_{\max}(G)^{-1} \Leftrightarrow$ maximiere $\lambda_{\min}(G^{-1})$,

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

Relaxationen

Statt m Experimente ganzzahlig zu wählen, bestimmt man relative Anteile,

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

Es gibt mehrere Ansätze, ein bzgl. \preceq minimales G zu finden. Interpretiere G dazu als ein „Konfidenzellipsoid“ mit Halbachsenlängen $\lambda_j(G)$,

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

D-optimales Design: Minimiere das Volumen des Konfidenzellipsoids.

E-optimales Design: Minimiere die längste Halbachse.

A-optimales Design: Minimiere die Summe der Halbachsen.

A-optimales Design. $\sum_{j=1}^p \lambda_j(G) = \sum_{j=1}^p G_{jj} = \sum_{j=1}^p e_j^T G e_j$.

Für jedes j ist die Unglg. $u_j \succeq e_j^T G e_j$ über Schur-Komplement darstellbar:

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

Inhaltsübersicht für heute:

Semidefinite Optimierung

SDP Anwendung: Robuste Stabilität dynamischer Systeme

SDP Anwendung: Entwurf von Experimenten

(SDP Anwendung: Graphenpartition)

(SDP Anwendung: geometrische Einbettungen)

Graphenpartition: Max-Cut

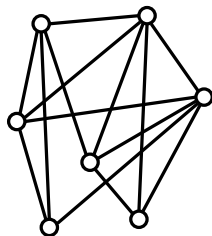
Gegeben: Graph $G = (V, E)$, $V = \{1, \dots, n\}$,

$E \subseteq \{ij : i, j \in V, i < j\}$, Kantengewichte a_{ij}

Gesucht: $S \subset V$ mit gewichtsmaximalem

Schnitt $\delta(S) := \{ij \in E : i \in S, j \in V \setminus S\}$

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



Graphenpartition: Max-Cut

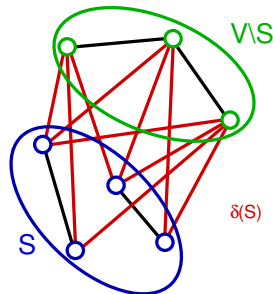
Gegeben: Graph $G = (V, E)$, $V = \{1, \dots, n\}$,

$E \subseteq \{ij : i, j \in V, i < j\}$, Kantengewichte a_{ij}

Gesucht: $S \subset V$ mit gewichtsmaximalem

Schnitt $\delta(S) := \{ij \in E : i \in S, j \in V \setminus S\}$

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



Graphenpartition: Max-Cut

Gegeben: Graph $G = (V, E)$, $V = \{1, \dots, n\}$,

$E \subseteq \{ij : i, j \in V, i < j\}$, Kantengewichte a_{ij}

Gesucht: $S \subset V$ mit gewichtsmaximalem

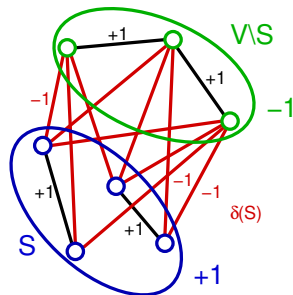
Schnitt $\delta(S) := \{ij \in E : i \in S, j \in V \setminus S\}$

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

Modellierung: Repräsentiere die Partition durch

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

$$\text{Dann ist } x_i x_j = \begin{cases} -1 & ij \in \delta(S) \\ 1 & \text{sonst} \end{cases}, \quad \text{bzw.} \quad \frac{1 - x_i x_j}{2} = \begin{cases} 1 & ij \in \delta(S) \\ 0 & \text{sonst.} \end{cases}$$



Graphenpartition: Max-Cut

Gegeben: Graph $G = (V, E)$, $V = \{1, \dots, n\}$,

$E \subseteq \{ij : i, j \in V, i < j\}$, Kantengewichte a_{ij}

Gesucht: $S \subset V$ mit gewichtsmaximalem

Schnitt $\delta(S) := \{ij \in E : i \in S, j \in V \setminus S\}$

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

Modellierung: Repräsentiere die Partition durch

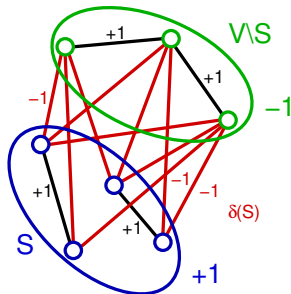
$$x \in \{-1, 1\}^n \quad \text{mit} \quad x_i = \begin{cases} 1 & i \in S \\ -1 & i \in V \setminus S \end{cases}$$

$$\text{Dann ist } x_i x_j = \begin{cases} -1 & ij \in \delta(S) \\ 1 & \text{sonst} \end{cases}, \quad \text{bzw.} \quad \frac{1 - x_i x_j}{2} = \begin{cases} 1 & ij \in \delta(S) \\ 0 & \text{sonst.} \end{cases}$$

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

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

Äquivalent zu quadratischer 0-1 Optimierung!



Semidefinite Max-Cut Relaxation

Beachte: $x^T C x = \langle Cx, x \rangle = \langle C, xx^T \rangle$

Eigenschaften von $xx^T = [x_i x_j]$ für $x \in \{-1, 1\}^n$:

- $x_i^2 = 1 \Rightarrow \text{diag}(xx^T) = \mathbf{1}$
- xx^T ist positiv semidefinit, $xx^T \succeq 0$
- $\text{Rang}(xx^T) = 1$

Semidefinite Max-Cut Relaxation

Beachte: $x^T C x = \langle C x, x \rangle = \langle C, x x^T \rangle$

Eigenschaften von $x x^T = [x_i x_j]$ für $x \in \{-1, 1\}^n$:

- $x_i^2 = 1 \Rightarrow \text{diag}(x x^T) = \mathbf{1}$
- $x x^T$ ist positiv semidefinit, $x x^T \succeq 0$
- $\text{Rang}(x x^T) = 1$

Relaxationsidee: Ersetze $x x^T$ durch eine positiv semidefinite Matrix X .

$$\max_{x \in \{-1, 1\}^n} x^T C x \leq$$

$$\begin{array}{ll} \max & \langle C, X \rangle \\ \text{s.t.} & \text{diag}(X) = \mathbf{1} \\ & X \succeq 0 \\ & [\text{Rang}(X) = 1] \end{array}$$

[mit Rang 1 \Leftrightarrow (MC), NP-vollst.]

Semidefinite Max-Cut Relaxation

Beachte: $x^T C x = \langle C x, x \rangle = \langle C, x x^T \rangle$

Eigenschaften von $x x^T = [x_i x_j]$ für $x \in \{-1, 1\}^n$:

- $x_i^2 = 1 \Rightarrow \text{diag}(x x^T) = \mathbf{1}$
- $x x^T$ ist positiv semidefinit, $x x^T \succeq 0$
- $\text{Rang}(x x^T) = 1$

Relaxationsidee: Ersetze $x x^T$ durch eine positiv semidefinite Matrix X .

$$\max_{x \in \{-1, 1\}^n} x^T C x \leq$$

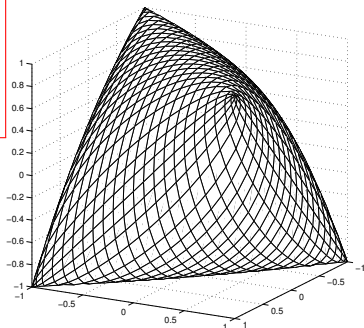
$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \text{diag}(X) = \mathbf{1} \\ & X \succeq 0 \\ & [\text{Rang}(X) = 1] \end{aligned}$$

[mit Rang 1 \Leftrightarrow (MC), NP-vollst.]

Illustration für $n = 3$:

Rand wird beschrieben von

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



Geometrische Interpretation und Rundung

Interpretiere OL X^* als **Gram-Matrix** $X^* = R^T R$ von $R = [r_1, \dots, r_n] \in \mathbb{R}^{k \times n}$,

$$X_{ij}^* = r_i^T r_j = \|r_i\| \|r_j\| \cos \angle(r_1, r_2)$$

Geometrische Interpretation und Rundung

Interpretiere OL X^* als **Gram-Matrix** $X^* = R^T R$ von $R = [r_1, \dots, r_n] \in \mathbb{R}^{k \times n}$,

$$X_{ij}^* = r_i^T r_j = \|r_i\| \|r_j\| \cos \angle(r_1, r_2) \quad [= \cos \angle(r_1, r_2).]$$

$\|r_i\| = 1$ wegen $\text{diag}(X^*) = \mathbf{1}$. Knoten i entspricht Vektor/Punkt $r_i \in \mathbb{R}^k$,

$$\langle C, X^* \rangle = \sum_{ij \in E} a_{ij} \frac{1 - X_{ij}^*}{2} = \sum_{ij \in E} a_{ij} \frac{1 - r_i^T r_j}{2}$$

Geometrische Interpretation und Rundung

Interpretiere OL X^* als **Gram-Matrix** $X^* = R^T R$ von $R = [r_1, \dots, r_n] \in \mathbb{R}^{k \times n}$,

$$X_{ij}^* = r_i^T r_j = \|r_i\| \|r_j\| \cos \angle(r_1, r_2) \quad [= \cos \angle(r_1, r_2).]$$

$\|r_i\| = 1$ wegen $\text{diag}(X^*) = \mathbf{1}$. Knoten i entspricht Vektor/Punkt $r_i \in \mathbb{R}^k$,

$$\langle C, X^* \rangle = \sum_{ij \in E} a_{ij} \frac{1 - X_{ij}^*}{2} = \sum_{ij \in E} a_{ij} \frac{1 - r_i^T r_j}{2}$$

Ist $r_i^T r_j$ nahe bei -1 (großer Winkel $\angle(r_1, r_2)$), sollte man i und j trennen.

Geht das für alle gleichzeitig?

Geometrische Interpretation und Rundung

Interpretiere OL X^* als **Gram-Matrix** $X^* = R^T R$ von $R = [r_1, \dots, r_n] \in \mathbb{R}^{k \times n}$,

$$X_{ij}^* = r_i^T r_j = \|r_i\| \|r_j\| \cos \angle(r_1, r_2) \quad [= \cos \angle(r_1, r_2).]$$

$\|r_i\| = 1$ wegen $\text{diag}(X^*) = \mathbf{1}$. Knoten i entspricht Vektor/Punkt $r_i \in \mathbb{R}^k$,

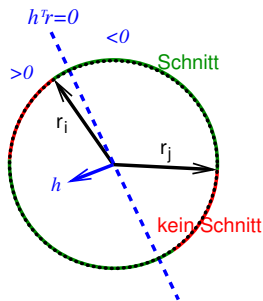
$$\langle C, X^* \rangle = \sum_{ij \in E} a_{ij} \frac{1 - X_{ij}^*}{2} = \sum_{ij \in E} a_{ij} \frac{1 - r_i^T r_j}{2}$$

Ist $r_i^T r_j$ nahe bei -1 (großer Winkel $\angle(r_1, r_2)$), sollte man i und j trennen.

Geht das für alle gleichzeitig? **„Zufälliges Runden mit Hyperebenen“**

Runde zu $x \in \{-1, 1\}^n$ über einen normalv. Zufallsvektor $h \in \mathbb{R}^k$ durch

$$x_i = \begin{cases} 1 & h^T r_i \geq 0 \\ -1 & \text{sonst} \end{cases} \rightarrow H = \sum_{ij \in E} a_{ij} \frac{1 - x_i x_j}{2}$$



Geometrische Interpretation und Rundung

Interpretiere OL X^* als **Gram-Matrix** $X^* = R^T R$ von $R = [r_1, \dots, r_n] \in \mathbb{R}^{k \times n}$,

$$X_{ij}^* = r_i^T r_j = \|r_i\| \|r_j\| \cos \angle(r_1, r_2) \quad [= \cos \angle(r_1, r_2).]$$

$\|r_i\| = 1$ wegen $\text{diag}(X^*) = \mathbf{1}$. Knoten i entspricht Vektor/Punkt $r_i \in \mathbb{R}^k$,

$$\langle C, X^* \rangle = \sum_{ij \in E} a_{ij} \frac{1 - X_{ij}^*}{2} = \sum_{ij \in E} a_{ij} \frac{1 - r_i^T r_j}{2}$$

Ist $r_i^T r_j$ nahe bei -1 (großer Winkel $\angle(r_1, r_2)$), sollte man i und j trennen.

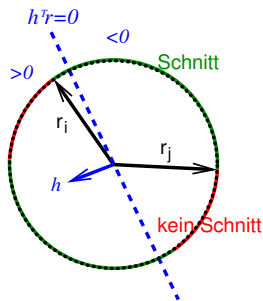
Geht das für alle gleichzeitig? „**Zufälliges Runden mit Hyperebenen**“

Runde zu $x \in \{-1, 1\}^n$ über einen normalv. Zufallsvektor $h \in \mathbb{R}^k$ durch

$$x_i = \begin{cases} 1 & h^T r_i \geq 0 \\ -1 & \text{sonst} \end{cases} \rightarrow H = \sum_{ij \in E} a_{ij} \frac{1 - x_i x_j}{2}$$

$\mathbb{E}(H)$ errechnet sich pro $ij \in E$ durch Projektion von h auf die $\{r_i, r_j\}$ -Ebene:

$$\mathbb{P}(x_i x_j = -1) = \arccos(r_i^T r_j) / \pi$$



Geometrische Interpretation und Rundung

Interpretiere OL X^* als **Gram-Matrix** $X^* = R^T R$ von $R = [r_1, \dots, r_n] \in \mathbb{R}^{k \times n}$,

$$X_{ij}^* = r_i^T r_j = \|r_i\| \|r_j\| \cos \angle(r_1, r_2) \quad [= \cos \angle(r_1, r_2).]$$

$\|r_i\| = 1$ wegen $\text{diag}(X^*) = \mathbf{1}$. Knoten i entspricht Vektor/Punkt $r_i \in \mathbb{R}^k$,

$$\langle C, X^* \rangle = \sum_{ij \in E} a_{ij} \frac{1 - X_{ij}^*}{2} = \sum_{ij \in E} a_{ij} \frac{1 - r_i^T r_j}{2}$$

Ist $r_i^T r_j$ nahe bei -1 (großer Winkel $\angle(r_1, r_2)$), sollte man i und j trennen.

Geht das für alle gleichzeitig? **„Zufälliges Runden mit Hyperebenen“**

Runde zu $x \in \{-1, 1\}^n$ über einen normalv. Zufallsvektor $h \in \mathbb{R}^k$ durch

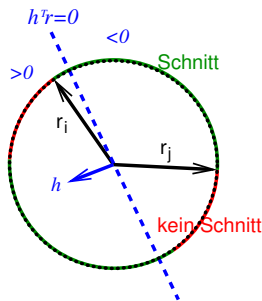
$$x_i = \begin{cases} 1 & h^T r_i \geq 0 \\ -1 & \text{sonst} \end{cases} \rightarrow H = \sum_{ij \in E} a_{ij} \frac{1 - x_i x_j}{2}$$

$\mathbb{E}(H)$ errechnet sich pro $ij \in E$ durch Projektion von h auf die $\{r_i, r_j\}$ -Ebene:

$$\mathbb{P}(x_i x_j = -1) = \arccos(r_i^T r_j) / \pi$$

Es gilt $\arccos(t) / \pi \geq 0.878 \frac{1-t}{2}$

und wenn $a_{ij} \geq 0 \Rightarrow \mathbb{E}(H) \geq 0.878 \langle C, X_* \rangle$



„0.878 Approximationsalgorithmus“ von Goemans und Williamson

Inhaltsübersicht für heute:

Semidefinite Optimierung

SDP Anwendung: Robuste Stabilität dynamischer Systeme

SDP Anwendung: Entwurf von Experimenten

(SDP Anwendung: Graphenpartition)

(SDP Anwendung: geometrische Einbettungen)

SDP, Gram-Matrix und geometrische Einbettung

Sind n Punkte $r_i \in \mathbb{R}^k$ gegeben und ist $R = [r_1, r_2, \dots, r_n]$, dann ist die **Gram-Matrix** $X = R^T R \succeq 0$ positiv semidefinit mit Rang k , und erfüllt

$$\|r_i - r_j\|^2 = r_i^T r_i - 2r_i^T r_j + r_j^T r_j = X_{ii} - 2X_{ij} + X_{jj} = \langle E_{ij}, X \rangle.$$

$E_{ij} \in \mathbb{R}^{n \times n}$, $E_{ii} = E_{jj} = 1$, $E_{ij} = E_{ji} = -1$, und 0 sonst: $E_{ij} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{matrix} i \\ j \end{matrix}$

SDP, Gram-Matrix und geometrische Einbettung

Sind n Punkte $r_i \in \mathbb{R}^k$ gegeben und ist $R = [r_1, r_2, \dots, r_n]$, dann ist die **Gram-Matrix** $X = R^T R \succeq 0$ positiv semidefinit mit Rang k , und erfüllt

$$\|r_i - r_j\|^2 = r_i^T r_i - 2r_i^T r_j + r_j^T r_j = X_{ii} - 2X_{ij} + X_{jj} = \langle E_{ij}, X \rangle.$$

$E_{ij} \in \mathbb{R}^{n \times n}$, $E_{ii} = E_{jj} = 1$, $E_{ij} = E_{ji} = -1$, und 0 sonst: $E_{ij} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{matrix} i \\ j \end{matrix}$

Ist umgekehrt die Lage der n Punkte nicht bekannt, sondern nur einige Distanzen $d_{ij} \in \mathbb{R}_+$ zwischen Paaren $ij \in E \subseteq \{ij : 1 \leq i < j \leq n\}$, sucht man eine niedrigdimensionale Einbettung der Punkte mit diesen Distanzen:

$$\begin{aligned} \min \quad & \text{Rang}(X) \\ \text{s.t.} \quad & \langle E_{ij}, X \rangle = d_{ij} \quad ij \in E \\ & X \succeq 0 \end{aligned}$$

SDP, Gram-Matrix und geometrische Einbettung

Sind n Punkte $r_i \in \mathbb{R}^k$ gegeben und ist $R = [r_1, r_2, \dots, r_n]$, dann ist die **Gram-Matrix** $X = R^T R \succeq 0$ positiv semidefinit mit Rang k , und erfüllt

$$X_{ij} = r_i^T r_j,$$

$$\|r_i - r_j\|^2 = r_i^T r_i - 2r_i^T r_j + r_j^T r_j = X_{ii} - 2X_{ij} + X_{jj} = \langle E_{ij}, X \rangle.$$

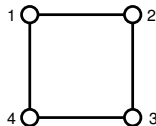
$E_{ij} \in \mathbb{R}^{n \times n}$, $E_{ii} = E_{jj} = 1$, $E_{ij} = E_{ji} = -1$, und 0 sonst: $E_{ij} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} i$

Ist umgekehrt die Lage der n Punkte nicht bekannt, sondern nur einige Distanzen $d_{ij} \in \mathbb{R}_+$ zwischen Paaren $ij \in E \subseteq \{ij : 1 \leq i < j \leq n\}$, sucht man eine niedrigdimensionale Einbettung der Punkte mit diesen Distanzen:

$$\begin{aligned} \min \quad & \text{Rang}(X) \\ \text{s.t.} \quad & \langle E_{ij}, X \rangle = d_{ij} \quad ij \in E \\ & X \succeq 0 \end{aligned}$$

Zwei Schwierigkeiten:

- Das geht gar nicht für beliebige d_{ij} !
Bsp: Kanten-Abstand im Kreis der Länge 4:
 $n = 4$, $d_{12} = d_{14} = d_{23} = d_{34} = 1$, $d_{13} = d_{24} = 2$
- Rang-Minimierung ist nicht konvex und *NP*-schwer.



Einbettung mit kleiner Verzerrung

Seien für n Punkte alle paarweisen Distanzen $d_{ij} > 0$ gegeben, und die Dreiecksungleichung $d_{ij} \leq d_{ih} + d_{hj}$ sei für alle i, j, h erfüllt. Eine Einbettung $\rho : \{1, \dots, n\} \rightarrow \mathbb{R}^k$ hat **Verzerrung (distortion)** $D > 0$ falls

$$d_{ij} \leq \|\rho(i) - \rho(j)\| \leq Dd_{ij} \quad \forall 1 \leq i < j \leq n$$

Einbettung mit kleiner Verzerrung

Seien für n Punkte alle paarweisen Distanzen $d_{ij} > 0$ gegeben, und die Dreiecksungleichung $d_{ij} \leq d_{ih} + d_{hj}$ sei für alle i, j, h erfüllt. Eine Einbettung $\rho : \{1, \dots, n\} \rightarrow \mathbb{R}^k$ hat **Verzerrung (distortion)** $D > 0$ falls

$$d_{ij} \leq \|\rho(i) - \rho(j)\| \leq Dd_{ij} \quad \forall 1 \leq i < j \leq n$$

Eine Einbettung geringster Verzerrung in \mathbb{R}^n ist per SDP konstruierbar:

$$\begin{array}{ll} \min & \delta \\ \text{s.t.} & d_{ij} \leq \langle E_{ij}, X \rangle \leq \delta d_{ij} \quad 1 \leq i < j \leq n \\ & X \succeq 0, \delta \in \mathbb{R} \quad [\delta = D^2] \end{array}$$

Faktorisiere X^* zum Beispiel über die Eigenwertzerlegung:

$$X^* = P\Lambda^*P^T = P(\Lambda^*)^{\frac{1}{2}} \underbrace{(\Lambda^*)^{\frac{1}{2}}P^T}_{=:R} = R^T R$$

Der Rang von R ist mit geringem Verlust in D auf $\sim \log n$ reduzierbar.