Semidefinite Optimierung mit Anwendungen

Christoph Helmberg (TU Chemnitz)

- Semidefinite Optimierung
- Lösungsverfahren
- Anwendungen

Inhalt

Der Kegel der positiv semidefiniten Matrizen

$X \in S^n$ positiv semidefinit $(X \succeq 0, \in S^n_+) :\Longleftrightarrow v^T X v \geq 0 \quad \forall v \in \mathbb{R}^n$

$X \in S^n$ positiv definit $(X \succ 0, \in S^n_{++}) :\Longleftrightarrow v^T X v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$

Der Kegel der positiv semidefiniten Matrizen

$X \in S^n$ positiv semidefinit ($X \succeq 0, \in S_+^n$) $:\Longleftrightarrow v^T X v \geq 0 \quad \forall v \in \mathbb{R}^n$

$X \in S^n$ positiv definit ($X \succ 0, \in S_{++}^n$) $:\Longleftrightarrow v^T X v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$

$X \succeq 0 \Longleftrightarrow$

- $X = \sum_{i=1}^n \lambda_i(X) v_i v_i^T$ mit $\lambda_i \geq 0$.
- $X = V^T V$
- $\langle X, A \rangle \geq 0 \qquad \forall A \succeq 0$

Der Kegel der positiv semidefiniten Matrizen

$X \in S^n$ positiv semidefinit ($X \succeq 0, \in S^n_+$) $:\Longleftrightarrow v^T X v \geq 0 \quad \forall v \in \mathbb{R}^n$

$X \in S^n$ positiv definit ($X \succ 0, \in S^n_{++}$) $:\Longleftrightarrow v^T X v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$

$X \succeq 0 \Longleftrightarrow$

- $X = \sum_{i=1}^n \lambda_i(X) v_i v_i^T$ mit $\lambda_i \geq 0$.
- $X = V^T V$
- $\langle X, A \rangle \geq 0 \qquad \forall A \succeq 0$

$X \succ 0 \Longleftrightarrow$

$\det(X_J) > 0 \; \forall J \subseteq \{1, \ldots, n\}$

Der Kegel der positiv semidefiniten Matrizen

$X \in S^n$ positiv semidefinit $(X \succeq 0, \in S^n_+) :\Longleftrightarrow v^T X v \geq 0 \quad \forall v \in \mathbb{R}^n$

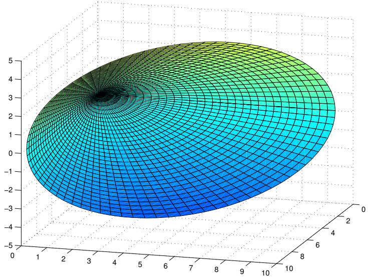
$X \in S^n$ positiv definit $(X \succ 0, \in S^n_{++}) :\Longleftrightarrow v^T X v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$

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

$X \succeq 0 \Longleftrightarrow$

- $X = \sum_{i=1}^n \lambda_i(X) v_i v_i^T$ mit $\lambda_i \geq 0$.
- $X = V^T V$
- $\langle X, A \rangle \geq 0 \qquad \forall A \succeq 0$

$X \succ 0 \Longleftrightarrow$
$\det(X_J) > 0 \ \forall J \subseteq \{1, \ldots, n\}$

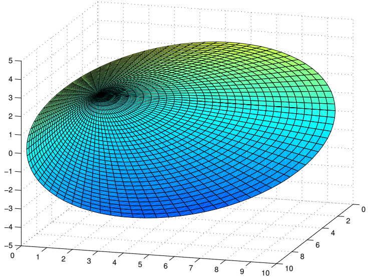


Der Kegel der positiv semidefiniten Matrizen

$X \in S^n$ positiv semidefinit ($X \succeq 0, \in S^n_+$) $:\Longleftrightarrow v^T X v \geq 0 \quad \forall v \in \mathbb{R}^n$

$X \in S^n$ positiv definit ($X \succ 0, \in S^n_{++}$) $:\Longleftrightarrow v^T X v > 0 \quad \forall v \in \mathbb{R}^n \setminus \{0\}$

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



$X \succeq 0 \Longleftrightarrow$

- $X = \sum_{i=1}^n \lambda_i(X) v_i v_i^T$ mit $\lambda_i \geq 0$.
- $X = V^T V$
- $\langle X, A \rangle \geq 0 \qquad \forall A \succeq 0$

$X \succ 0 \Longleftrightarrow$

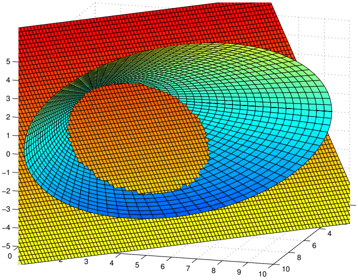
$\det(X_J) > 0 \; \forall J \subseteq \{1, \ldots, n\}$

$$A, B \in S^n_+, \; \alpha \geq 0 \quad \Rightarrow \quad \alpha(A + B) \in S^n_+$$

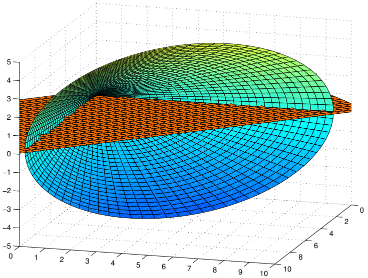
S^n_+ ist konvex, abg., selbstdual, spitz ($A \succeq B :\Leftrightarrow A - B \succeq 0$), aber nicht polyedrisch.

Beispiele: Schnitt mit einer Hyperebene

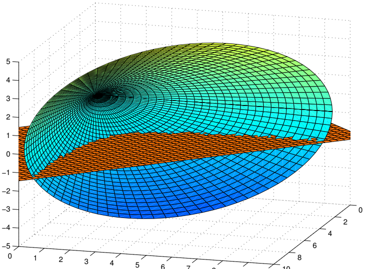
Normalfall



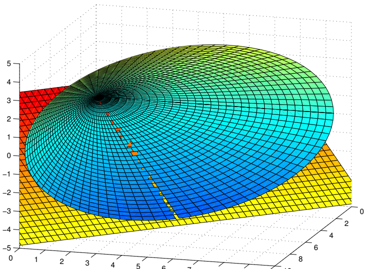
Nichtnegativer Orthant



Hyperbel (offene Projektion!)



Seitenfläche ($\equiv \lambda v v^T$)



Lineare Optimierung ↔ Semidefinite Optimierung

$$\begin{aligned} \max \quad & \langle c, x\rangle \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$	$$\begin{aligned} \max \quad & \langle C, X\rangle \\ \text{s.t.} \quad & \mathcal{A}X = b \\ & X \succeq 0 \end{aligned}$$
$x \in \mathbb{R}^n_+$ nichtneg. Orthant (polyedrisch)	$X \in \mathcal{S}^n_+$ pos. semidef. Matrizen (nicht polyedrisch)
$\langle c, x\rangle = \sum_i c_i x_i$	$\langle C, X\rangle = \sum_{i,j} C_{ij} X_{ij}$
$Ax = \begin{pmatrix} \langle a_1, x\rangle \\ \vdots \\ \langle a_m, x\rangle \end{pmatrix}$	$\mathcal{A}X = \begin{pmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{pmatrix}$
$A^T y = \sum_i a_i y_i$	$\mathcal{A}^T y = \sum_i A_i y_i$
$$\begin{aligned} \min \quad & \langle b, y\rangle \\ \text{s.t.} \quad & A^T y - z = c \\ & z \geq 0 \end{aligned}$$	$$\begin{aligned} \min \quad & \langle b, y\rangle \\ \text{s.t.} \quad & \mathcal{A}^T y - Z = C \\ & Z \succeq 0 \end{aligned}$$

Beispiel

$$\begin{array}{rl} \max & \langle C, X \rangle \\ \text{s.t.} & \langle I, X \rangle = 1 \\ & X \succeq 0 \end{array}$$

$$\begin{array}{rl} \min & y \\ \text{s.t.} & Z = yI - C \succeq 0 \end{array}$$

Beispiel

$$\begin{array}{rl} \max & \langle C, X \rangle \\ \text{s.t.} & \langle I, X \rangle = 1 \\ & X \succeq 0 \end{array} \qquad \begin{array}{rl} \min & y \\ \text{s.t.} & Z = yI - C \succeq 0 \end{array}$$

$$\{X \succeq 0 : \langle I, X \rangle = 1\} \;\; = \;\; \mathrm{conv}\left\{vv^T : \left\langle I, vv^T \right\rangle = v^T v = 1\right\}$$

$$\text{und} \qquad \max_{\|v\|^2=1} \left\langle C, vv^T \right\rangle \;\; = \;\; \max_{\|v\|=1} v^T C v \;\; = \;\; \lambda_{\max}(C)$$

Beispiel

$$\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \langle I, X\rangle = 1 \\ & X \succeq 0 \end{array} \qquad\qquad \begin{array}{rl} \min & y \\ \text{s.t.} & Z = yI - C \succeq 0 \end{array}$$

$$\{X \succeq 0 : \langle I, X\rangle = 1\} \quad = \quad \mathrm{conv}\left\{vv^T : \left\langle I, vv^T\right\rangle = v^Tv = 1\right\}$$

$$\text{und} \qquad \max_{\|v\|^2=1} \left\langle C, vv^T\right\rangle \quad = \quad \max_{\|v\|=1} v^TCv \quad = \quad \lambda_{\max}(C)$$

Menge primaler Optimallösungen:

$$\begin{array}{rll} & \mathrm{conv}\left\{vv^T : \left\langle I, vv^T\right\rangle = 1, v^TCv = \lambda_{\max}(C)\right\} & [v = Pu] \\ = & \mathrm{conv}\left\{Puu^TP^T : \left\langle I, uu^T\right\rangle = 1\right\} & \\ = & \left\{PUP^T : \langle I, U\rangle = 1, U \succeq 0\right\} & \end{array}$$

Spalten von P bilden orthonormale Basis des Eigenraums von $\lambda_{\max}(C)$.

Beispiel

$$\begin{array}{rl} \max & \langle C, X\rangle \\ \text{s.t.} & \langle I, X\rangle = 1 \\ & X \succeq 0 \end{array} \qquad\qquad \begin{array}{rl} \min & y \\ \text{s.t.} & Z = yI - C \succeq 0 \end{array}$$

$$\{X \succeq 0 : \langle I, X\rangle = 1\} \quad = \quad \mathrm{conv}\left\{vv^T : \left\langle I, vv^T\right\rangle = v^Tv = 1\right\}$$

$$\text{und} \qquad \max_{\|v\|^2=1} \left\langle C, vv^T\right\rangle \quad = \quad \max_{\|v\|=1}\; v^TCv \quad = \quad \lambda_{\max}(C)$$

Menge primaler Optimallösungen:

$$\begin{array}{rl} & \mathrm{conv}\left\{vv^T : \left\langle I, vv^T\right\rangle = 1, v^TCv = \lambda_{\max}(C)\right\} \qquad [v = Pu] \\ = & \mathrm{conv}\left\{Puu^TP^T : \left\langle I, uu^T\right\rangle = 1\right\} \\ = & \left\{PUP^T : \langle I, U\rangle = 1, U \succeq 0\right\} \end{array}$$

Spalten von P bilden orthonormale Basis des Eigenraums von $\lambda_{\max}(C)$.

---

dual: $\min\ \lambda$ s.t. $\lambda I - C \succeq 0 \quad \Rightarrow \quad$ optimales $\lambda = \lambda_{\max}(C)$

Die Seitenstruktur des positiv semidefiniten Kegels

Die Seiten von $\mathcal{S}_+^n$ sind: $\emptyset$, $\{\mathbf{0}\}$ [Barker und Carlson 1975]

und, für jeden r-dim. linearen Unterraum $\mathcal{L}$ von $\mathbb{R}^n$
mit einer beliebigen Basis $P \in \mathbb{R}^{n \times r}$,

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

Die Seitenstruktur des positiv semidefiniten Kegels

Die Seiten von $\mathcal{S}^n_+$ sind: $\emptyset$, $\{\mathbf{0}\}$ [Barker und Carlson 1975]

und, für jeden r-dim. linearen Unterraum $\mathcal{L}$ von $\mathbb{R}^n$ mit einer beliebigen Basis $P \in \mathbb{R}^{n \times r}$,

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

Die Seitenstruktur des positiv semidefiniten Kegels

Die Seiten von $\mathcal{S}_+^n$ sind: $\emptyset$, $\{\mathbf{0}\}$ [Barker und Carlson 1975]

und, für jeden r-dim. linearen Unterraum $\mathcal{L}$ von $\mathbb{R}^n$
mit einer beliebigen Basis $P \in \mathbb{R}^{n \times r}$,

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

- $\dim F_{\mathcal{L}} = \binom{r+1}{2}$ [wächst in Sprüngen]
- minimales Erzeugendensystem: $\mathcal{S}_+^n = \operatorname{cone}\{vv^T : \|v\| = 1\}$

Beispiel mit Dualitätslücke [Vandenberghe und Boyd 1996]

$$\min x_{12}$$

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

$$\max y_1$$

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

Beispiel mit Dualitätslücke [Vandenberghe und Boyd 1996]

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

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

Entsprechende Matrizen:

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

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

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

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

Beispiel mit Dualitätslücke [Vandenberghe und Boyd 1996]

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

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

Entsprechende Matrizen:

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

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

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

Beispiel mit Dualitätslücke [Vandenberghe und Boyd 1996]

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

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

Entsprechende Matrizen:

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

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

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

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

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

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

Beispiel mit Dualitätslücke [Vandenberghe und Boyd 1996]

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

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

Entsprechende Matrizen:

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

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

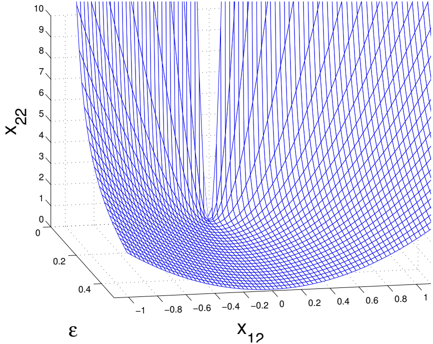
$$A_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

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

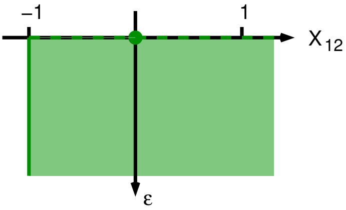
$x_{33} \geq 0 \;\Rightarrow\; x_{12} \geq -1,\; x_{22} \geq x_{12}^2/\varepsilon \geq \frac{1}{\varepsilon}$ primale Optimallösung ist -1.

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

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



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!

Problem: die Menge $\left\{ \begin{pmatrix} \langle C, X \rangle \\ \mathcal{A}X \end{pmatrix} : X \succeq 0 \right\}$ ist nicht abgeschlossen.

Starke Dualität

Annahme:

(S) Es existieren primal und dual streng zulässige Lösungen
$\mathcal{A}(X) = b, X \succ 0$ und $Z = \mathcal{A}^T y - C \succ 0$. [Slater-Punkte]

Satz (Starke Dualität)

Gilt (S), dann werden primale und duale Optimallösung angenommen und die Optimalwerte sind gleich.

Starke Dualität

Annahme:

(S) Es existieren primal und dual streng zulässige Lösungen
$\mathcal{A}(X) = b, X \succ 0$ und $Z = \mathcal{A}^T y - C \succ 0.$ [Slater-Punkte]

Satz (Starke Dualität)

Gilt (S), dann werden primale und duale Optimallösung angenommen und die Optimalwerte sind gleich.

Für primal und dual zulässige Punkte X und (y, Z) ergibt $\langle Z, X\rangle$ die Differenz der Zielfunktionswerte:

$$b^T y - \langle C, X\rangle = \langle AX, y\rangle - \left\langle \mathcal{A}^T y - Z, X \right\rangle = \langle Z, X\rangle \geq 0$$

Starke Dualität

Annahme:

(S) Es existieren primal und dual streng zulässige Lösungen
$\mathcal{A}(X) = b, X \succ 0$ und $Z = \mathcal{A}^T y - C \succ 0$. [Slater-Punkte]

Satz (Starke Dualität)

Gilt (S), dann werden primale und duale Optimallösung angenommen und die Optimalwerte sind gleich.

Für primal und dual zulässige Punkte X und (y, Z) ergibt $\langle Z, X\rangle$ die Differenz der Zielfunktionswerte:

$$b^T y - \langle C, X\rangle = \langle AX, y\rangle - \left\langle \mathcal{A}^T y - Z, X \right\rangle = \langle Z, X\rangle \geq 0$$

Hat das primale keinen inneren Punkt, kennt man aber die minimale Seite des Kegels, die die zulässige Menge enthält, erhält man ein stark duales Problem durch Projektion auf diese Seite.

Mehrere semidefinite Variable

$$
\begin{array}{ll}
\min & \sum_{i=1}^{k} \langle C_i, X_i \rangle \\
\text{s.t.} & \sum_{i=1}^{k} \mathcal{A}_i X_i = b \\
& X_1 \in \mathcal{S}_{n_1}^{+}, \dots, X_k \in \mathcal{S}_{n_k}^{+}
\end{array}
\qquad
\begin{array}{ll}
\max & \langle b, y \rangle \\
\text{s.t.} & \mathcal{A}_i^T y + Z_i = C_i \quad i = 1, \dots, k \\
& y \in \mathbb{R}^m, Z_1 \in \mathcal{S}_{n_1}^{+}, \dots, Z_k \in \mathcal{S}_{n_k}^{+}
\end{array}
$$

Mehrere semidefinite Variable

$$\begin{array}{ll} \min & \sum_{i=1}^k \langle C_i, X_i \rangle \\ \text{s.t.} & \sum_{i=1}^k \mathcal{A}_i X_i = b \\ & X_1 \in \mathcal{S}_{n_1}^+, \ldots, X_k \in \mathcal{S}_{n_k}^+ \end{array} \qquad \begin{array}{ll} \max & \langle b, y \rangle \\ \text{s.t.} & \mathcal{A}_i^T y + Z_i = C_i \quad i = 1, \ldots, k \\ & y \in \mathbb{R}^m, Z_1 \in \mathcal{S}_{n_1}^+, \ldots, Z_k \in \mathcal{S}_{n_k}^+ \end{array}$$

Für die Theorie unerheblich, weil

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

$\Rightarrow$ SDO enthält Lineare Optimierung

Mehrere semidefinite Variable

$$\begin{array}{ll} \min & \sum_{i=1}^k \langle C_i, X_i \rangle \\ \text{s.t.} & \sum_{i=1}^k \mathcal{A}_i X_i = b \\ & X_1 \in \mathcal{S}_{n_1}^+, \ldots, X_k \in \mathcal{S}_{n_k}^+ \end{array} \qquad \begin{array}{ll} \max & \langle b, y \rangle \\ \text{s.t.} & \mathcal{A}_i^T y + Z_i = C_i \quad i = 1, \ldots, k \\ & y \in \mathbb{R}^m, Z_1 \in \mathcal{S}_{n_1}^+, \ldots, Z_k \in \mathcal{S}_{n_k}^+ \end{array}$$

Für die Theorie unerheblich, weil

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

$\Rightarrow$ SDO enthält Lineare Optimierung

Eine Bedingung der Form

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

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

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

Mehrere semidefinite Variable

$$
\begin{array}{llll}
\min & \sum_{i=1}^k \langle C_i, X_i \rangle & \max & \langle b, y \rangle \\
\text{s.t.} & \sum_{i=1}^k \mathcal{A}_i X_i = b & \text{s.t.} & \mathcal{A}_i^T y + Z_i = C_i \quad i = 1, \dots, k \\
& X_1 \in \mathcal{S}_{n_1}^+, \dots, X_k \in \mathcal{S}_{n_k}^+ & & y \in \mathbb{R}^m, Z_1 \in \mathcal{S}_{n_1}^+, \dots, Z_k \in \mathcal{S}_{n_k}^+
\end{array}
$$

Für die Theorie unerheblich, weil

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

$\Rightarrow$ SDO enthält Lineare Optimierung

Eine Bedingung der Form

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

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

Zulässige $y \in \mathbb{R}^m$ sind SD-darstellbar, $\{y \in \mathbb{R}^m : \mathcal{A}^T y + Z = C, Z \succeq 0\}$.
Bsp: Lyapunov Ungleichung $\quad P^T X + XP \prec 0, \quad X \succ 0.$

Nützliches Modellierungselement: Schur-Komplement

Satz (Schur-Komplement)

Für $A \in \mathcal{S}^m_{++}$, $C \in \mathcal{S}^n_+$ *und* $B \in \mathbb{R}^{m \times n}$ *gilt*

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

Nützliches Modellierungselement: Schur-Komplement

Satz (Schur-Komplement)

Für $A \in \mathcal{S}^m_{++}$, $C \in \mathcal{S}^n_+$ und $B \in \mathbb{R}^{m\times n}$ gilt

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

Bsp: Die Second-Order-Cone Bedingung:

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

[für $x_0 = 0$ direkt nachprüfen]

Nützliches Modellierungselement: Schur-Komplement

Satz (Schur-Komplement)

Für $A \in \mathcal{S}^m_{++}$, $C \in \mathcal{S}^n_+$ *und* $B \in \mathbb{R}^{m\times n}$ *gilt*

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

Bsp: Die Second-Order-Cone Bedingung:

$$\|\bar{x}\| \le x_0 \quad \overset{x_0 \ge 0}{\Leftrightarrow} \quad x_0 \ge \frac{1}{x_0}\bar{x}^T I \bar{x} \quad \overset{\text{Schur}}{\Leftrightarrow} \quad \begin{bmatrix} x_0 & \bar{x}^T \\ \bar{x} & x_0 I \end{bmatrix} \succeq 0.$$

[für $x_0 = 0$ direkt nachprüfen]

• SDO enthält SOCO und konvex quadratisch restringierte konvex quadratische Optimierung, da für $Q = C^T C$

$$x^T Q x \le q^T x + c \qquad \Longleftrightarrow \qquad \begin{bmatrix} I & Cx \\ x^T C^T & q^T x + c \end{bmatrix} \succeq 0.$$

Komplexität von SDO

Im reellen Modell in $NP \cap Co\text{-}NP$,
im Turing- oder Bit-Modell ist der Status offen!

Komplexität von SDO

Im reellen Modell in $NP \cap Co\text{-}NP$,
im Turing- oder Bit-Modell ist der Status offen!

Beispiel: [Ramana1997]

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

Komplexität von SDO

Im reellen Modell in $NP \cap Co\text{-}NP$,
im Turing- oder Bit-Modell ist der Status offen!

Beispiel: [Ramana1997]

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

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

doppelt exponentieller Wert!

Komplexität von SDO

Im reellen Modell in $NP \cap \mathit{Co}\text{-}NP$,
im Turing- oder Bit-Modell ist der Status offen!

Beispiel: [Ramana1997]

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

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

doppelt exponentieller Wert!

Polynomial für ε-Lösungen in beschränktem Gebiet
(Ellipsoid-Methode). [Grötschel Lovász Schrijver 1988]

Inhalt

Lösungsverfahren

- Innere-Punkte-Verfahren ("polynomial")
  Codes: SDPT3, Sedumi, SDPA, CSDP
- Penalty Verfahren
  Code: Pennon
- Spektrales Bündelverfahren ($f(y) := \lambda_{\text{max}}(C - \mathcal{A}^T y) + b^T y$)
  Code: ConicBundle
- Quadratische Reformulierung (ersetze $0 \preceq X = LL^T$)
  Code: SDPLR

Innere-Punkte-Verfahren

Ansatz aus der *Nichtlinearen Optimierung*:

> *Starte im Inneren der zulässigen Menge und verhindere das Verlassen derselben durch eine Barriere-Funktion.*

Innere-Punkte-Verfahren

Ansatz aus der *Nichtlinearen Optimierung*:

Starte im Inneren der zulässigen Menge und verhindere das Verlassen derselben durch eine Barriere-Funktion.

Duales Problem

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

Innere-Punkte-Verfahren

Ansatz aus der *Nichtlinearen Optimierung*:

Starte im Inneren der zulässigen Menge und verhindere das Verlassen derselben durch eine Barriere-Funktion.

Duales Problem

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

$$\boxed{\text{Barriere-Funktion für } Z \in \mathcal{S}_n^+: \quad -\log\det Z}$$

- $\det Z = \prod \lambda_i(Z)$ ist > 0 im Inneren von $\mathcal{S}_n^+$ und 0 am Rand
- $\log\det Z = \sum \log \lambda_i(Z)$
- Falls ein $\lambda_i(Z) \to 0$, dann $\quad -\log\det Z \to \infty$
- $-\log\det Z$ ist glatt und streng konvex im Inneren von S_n^+

Innere-Punkte-Verfahren

Ansatz aus der *Nichtlinearen Optimierung*:

> *Starte im Inneren der zulässigen Menge und verhindere das Verlassen derselben durch eine Barriere-Funktion.*

Duales Problem

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

$$\boxed{\text{Barriere-Funktion für } Z \in \mathcal{S}_n^+: \quad -\log\det Z}$$

- $\det Z = \prod \lambda_i(Z)$ ist > 0 im Inneren von $\mathcal{S}_n^+$ und 0 am Rand
- $\log\det Z = \sum \log \lambda_i(Z)$
- Falls ein $\lambda_i(Z) \to 0$, dann $\quad -\log\det Z \to \infty$
- $-\log\det Z$ ist glatt und streng konvex im Inneren von S_n^+

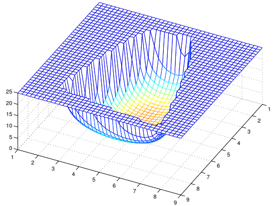
SUMT, Fiacco und McCormick 1968: Löse Folge von Barriere-Problemen

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

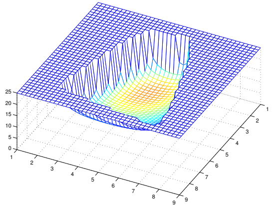
mittels Newton-Verfahren für $\mu > 0$, $\mu \to 0$.

Beispiel: Barriere-Problem für $\mu \to 0$

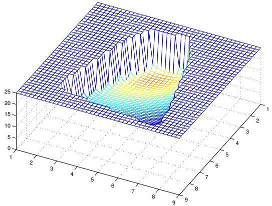
$\mu = 4$



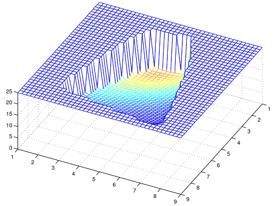
$\mu = 2$



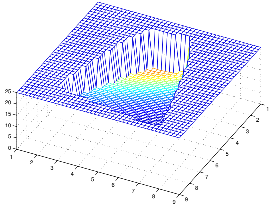
$\mu = 1$



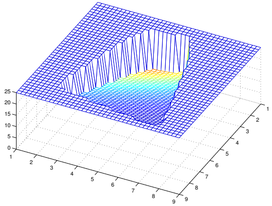
$\mu = 0.5$



$\mu = 0.25$



$\mu = 0.1$



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

1. Bestimme System der (hinr.) Bedingungen erster Ordnung

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

2. Bestimme Schritt Δy, sodass die Linearisierung im derzeitigen Punkt y_c,

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

das System der Bedingungen für $y_c + \Delta y$ erfüllt.
[= Löse das quadratische Modell von f optimal]

3. gedämpfter Newton-Schritt: $y_+ = y_c + \alpha \Delta y$ mit $\alpha \in (0, 1]$, sodass y_+ zumindest zulässig ist.

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

1. Bestimme System der (hinr.) Bedingungen erster Ordnung
$$\nabla f(y) = 0$$
2. Bestimme Schritt Δy, sodass die Linearisierung im derzeitigen Punkt y_c,
$$\nabla f(y_c) + \nabla^2 f(y_c)^T \Delta y = 0,$$
das System der Bedingungen für $y_c + \Delta y$ erfüllt.
[= Löse das quadratische Modell von f optimal]
3. gedämpfter Newton-Schritt: $y_+ = y_c + \alpha \Delta y$ mit $\alpha \in (0, 1]$, sodass y_+ zumindest zulässig ist.

Annahmen:

(A) $\mathcal{A}$ hat vollen Zeilenrang. [unwichtig]

(S) Es existieren primal und dual streng zulässige Lösungen.

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

1. Hinreichende Bedingung erster Ordnung:

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

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

liefert Verfahren von Jarre 1993, Nesterov und Nemirovskii 1994

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

1. Hinreichende Bedingung erster Ordnung:

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

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

liefert Verfahren von Jarre 1993, Nesterov und Nemirovskii 1994

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

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

[primale Zulässigkeit]
[duale Zulässigkeit]
[perturbierte Komplementarität]

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

1. Hinreichende Bedingung erster Ordnung:

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

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

liefert Verfahren von Jarre 1993, Nesterov und Nemirovskii 1994

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

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

[primale Zulässigkeit]
[duale Zulässigkeit]
[perturbierte Komplementarität]

- Für jedes $\mu > 0$ eindeutige Lösung (X_μ, y_μ, Z_μ) [erfordert (S) und (A)]
- (y_μ, Z_μ) ist die Optimallösung des dualen Barriere-Problems
- X_μ ist die Optimallösung des primalen Barriere-Problems
- Die Kurve der (X_μ, y_μ, Z_μ) für $\mu > 0$ heißt *zentraler Pfad*.

2. Primal-duale Linearisierung

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

Lösung $(\Delta X, \Delta y, \Delta Z)$ ergibt die Schritt-Richtung

Problem: ΔZ symmetrisch [II], aber nicht ΔX [III]

2. Primal-duale Linearisierung

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

Lösung $(\Delta X, \Delta y, \Delta Z)$ ergibt die Schritt-Richtung

Problem: ΔZ symmetrisch [II], aber nicht ΔX [III]
Vorschläge:

(a) HRVW96/KSH97/M97: Verwende symmetrischen Anteil,

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

(b) NT97: Skaliere III mit der Matrix $W = X^{\frac{1}{2}}(X^{\frac{1}{2}}ZX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}}$,

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

(C) AHO98: Symmetrisiere III direkt,

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

Es gibt viele weitere, alle unterscheiden sich leicht (Todd 1999).

Algorithmisches Schema primal-dualer Innere-Punkte-Verfahren

Input: $\mathcal{A}$, b, C, Startpunkt (X^0, y^0, Z^0) mit $X^0 \succ 0$ und $Z^0 \succ 0$

1. Wähle $\mu = \sigma \frac{\langle X, Z \rangle}{n}$ mit $\sigma \in (0, 1]$.
2. Berechne $(\Delta X, \Delta y, \Delta Z)$.
3. Line search: Bestimme $\alpha \in (0, 1]$ mit $X + \alpha \Delta X \succ 0$ und $Z + \alpha \Delta Z \succ 0$.
4. Setze $(X, y, Z) := (X + \alpha \Delta X, y + \alpha \Delta y, Z + \alpha \Delta Z)$.
5. Falls $\|\mathcal{A}X - b\|$, $\left\|\mathcal{A}^T y + Z - C\right\|_{\mathrm{F}}$ und $\langle X, Z \rangle$ „klein genug“, **stop**, sonst **goto** 1.

Algorithmisches Schema primal-dualer Innere-Punkte-Verfahren

Input: $\mathcal{A}$, b, C, Startpunkt (X^0, y^0, Z^0) mit $X^0 \succ 0$ und $Z^0 \succ 0$

1. Wähle $\mu = \sigma \frac{\langle X, Z \rangle}{n}$ mit $\sigma \in (0, 1]$.
2. Berechne $(\Delta X, \Delta y, \Delta Z)$.
3. Line search: Bestimme $\alpha \in (0, 1]$ mit $X + \alpha \Delta X \succ 0$ und $Z + \alpha \Delta Z \succ 0$.
4. Setze $(X, y, Z) := (X + \alpha \Delta X, y + \alpha \Delta y, Z + \alpha \Delta Z)$.
5. Falls $\|\mathcal{A}X - b\|$, $\left\|\mathcal{A}^T y + Z - C\right\|_{\mathrm{F}}$ und $\langle X, Z \rangle$ „klein genug", **stop**, sonst **goto** 1.

Satz (Kojima, Shindoh und Hara 1997)

(X^0, y^0, Z^0) zulässig und „zentriert". Wähle $\sigma = 1 - \frac{0.35}{\sqrt{n}}$ und $\alpha = 1$, dann ist jeder Schritt zulässig und $\langle X, Z \rangle < \varepsilon$ in

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

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

Aufwand pro Iteration

Berechnung der HRVW/KSH/M Schritt-Richtung:

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

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

Aufwand pro Iteration

Berechnung der HRVW/KSH/M Schritt-Richtung:

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

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

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

X und Z^{-1} sind im Allgemeinen dicht besetzt.

$\Rightarrow$ M ist eine dicht besetzte positiv definite Matrix der Ordnung m.
Cholesky-Faktorisierung benötigt $m^3/3$ Flops und $O(m^2)$ Speicher.

Aufwand pro Iteration

Berechnung der HRVW/KSH/M Schritt-Richtung:

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

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

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

X und Z^{-1} sind im Allgemeinen dicht besetzt.

$\Rightarrow$ M ist eine dicht besetzte positiv definite Matrix der Ordnung m.
Cholesky-Faktorisierung benötigt $m^3/3$ Flops und $O(m^2)$ Speicher.

Line search:

erfordert eine bis drei Cholesky-Faktorisierungen für dicht besetztes X (Z ist oft dünn besetzt) mit je $n^3/3$ Flops und $O(n^2)$ Speicher .

Inhalt

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)

Robuste Optimalsteuerung [BEFB 1994]

$$\text{(PLDI)} \quad \dot{x} = A(t)x \quad \text{mit } A(t) \in \operatorname{conv}\{A_1, \ldots, A_k\} \subset M_n.$$

- $x(t)$ … Zustand des Systems zur Zeit t.
- $A(t)$ … unsichere Übergangsmatrix zur Zeit t.
- Eine (PLDI) heißt *stabil*, wenn $x(t) \to 0$ für $t \to \infty$
  bei beliebiger Realisierung $A(t)$ aus $\operatorname{conv}\{\}$.

Robuste Optimalsteuerung [BEFB 1994]

$$\text{(PLDI)} \quad \dot{x} = A(t)x \quad \text{mit } A(t) \in \operatorname{conv}\{A_1, \ldots, A_k\} \subset M_n.$$

- $x(t)$ ... Zustand des Systems zur Zeit t.
- $A(t)$ ... unsichere Übergangsmatrix zur Zeit t.
- Eine (PLDI) heißt *stabil*, wenn $x(t) \to 0$ für $t \to \infty$ bei beliebiger Realisierung $A(t)$ aus $\operatorname{conv}\{\}$.

Hinreichend: Es gibt eine Norm

$$\|x\|_H = \sqrt{x^T H x} \text{ mit } H \succ 0, \text{ so dass } \frac{d}{dt}\|x(t)\|_H^2 < 0$$

auf allen Trajektorien (System ist *quadratisch stabil*, $x^T Hx$ heißt *quadratische Lyapunov Funktion*).

Robuste Optimalsteuerung [BEFB 1994]

$$\text{(PLDI)} \quad \dot{x} = A(t)x \quad \text{mit } A(t) \in \operatorname{conv}\{A_1, \ldots, A_k\} \subset M_n.$$

- $x(t)$ . . . Zustand des Systems zur Zeit t.
- $A(t)$ . . . unsichere Übergangsmatrix zur Zeit t.
- Eine (PLDI) heißt *stabil*, wenn $x(t) \to 0$ für $t \to \infty$ bei beliebiger Realisierung $A(t)$ aus $\operatorname{conv}\{\}$.

Hinreichend: Es gibt eine Norm

$$\|x\|_H = \sqrt{x^T H x} \text{ mit } H \succ 0, \text{ so dass } \frac{d}{dt}\|x(t)\|_H^2 < 0$$

auf allen Trajektorien (System ist *quadratisch stabil*, $x^T H x$ heißt *quadratische Lyapunov Funktion*).

$$\frac{d}{dt} x^T H x = x^T (A(t)^T H + H A(t)) x$$

$\Rightarrow$ quadratisch stabil falls es ein H gibt mit

$$H \succ 0, \qquad A_i^T H + H A_i \prec 0 \quad \text{für } i = 1, \ldots, k.$$

Robuste Optimalsteuerung [BEFB 1994]

$$\text{(PLDI)} \quad \dot{x} = A(t)x \quad \text{mit } A(t) \in \operatorname{conv}\{A_1, \ldots, A_k\} \subset M_n.$$

- $x(t)$ ... Zustand des Systems zur Zeit t.
- $A(t)$ ... unsichere Übergangsmatrix zur Zeit t.
- Eine (PLDI) heißt *stabil*, wenn $x(t) \to 0$ für $t \to \infty$ bei beliebiger Realisierung $A(t)$ aus $\operatorname{conv}\{\}$.

Hinreichend: Es gibt eine Norm

$$\|x\|_H = \sqrt{x^T H x} \text{ mit } H \succ 0, \text{ so dass } \frac{d}{dt}\|x(t)\|_H^2 < 0$$

auf allen Trajektorien (System ist *quadratisch stabil*, $x^T H x$ heißt *quadratische Lyapunov Funktion*).

$$\frac{d}{dt} x^T H x = x^T (A(t)^T H + H A(t)) x$$

$\Rightarrow$ quadratisch stabil falls es ein H gibt mit

$$H \succ 0, \qquad A_i^T H + H A_i \prec 0 \quad \text{für } i = 1, \ldots, k.$$

Test direkt mittels Innere-Punkte-Verfahren oder als Eigenwert-Problem

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

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 $\mathcal{R} = \{r_i \in \mathbb{R}^p : i = 1, \ldots, n\}$ 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 $\mathrm{Rang}[a_1, \ldots, a_m] = n$ ein geschätztes

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

dessen Fehlerverteilung Erwartungswert 0 und Kovarianzmatrix G hat.

Entwurf von Experimenten

Um die Werte eines Parametervektors $\xi \in \mathbb{R}^p$ zu schätzen, stehen $\mathcal{R} = \{r_i \in \mathbb{R}^p : i = 1, \ldots, n\}$ 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 $\mathrm{Rang}[a_1, \ldots, a_m] = n$ ein geschätztes

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

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.

$\rightarrow$ Finde die bzgl. $\preceq$ minimalen Elemente von

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

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},

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

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 $\lambda_{\min}(G^{-1})$,

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

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:

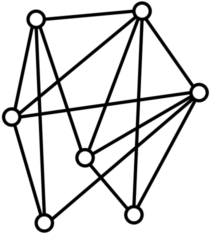
$$\begin{array}{ll} \min & \mathbf{1}^T u \\ \text{s.t.} & \begin{bmatrix} \sum_{i=1}^{n} \alpha_i r_i r_i^T & e_j \\ e_j^T & u_j \end{bmatrix} \succeq 0, \quad j = 1, \ldots, p \\ & \mathbf{1}^T \alpha = 1, \alpha \geq 0, u \in \mathbb{R}^p \end{array}$$

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

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

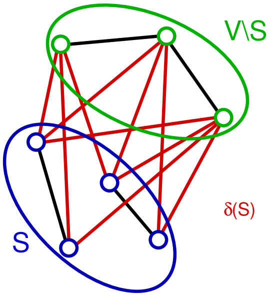


Graphenpartition: Max-Cut

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

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

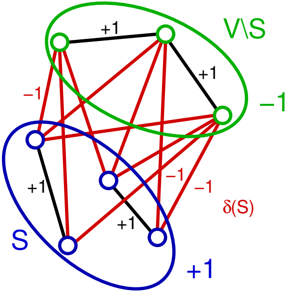


Graphenpartition: Max-Cut

Gegeben: Graph $G = (V, E)$, $V = \{1, \ldots, n\}$,
$E \subseteq \{ij : i, j \in V, i < j\}$, Kantengewichte a_{ij}

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

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



Modellierung: Repräsentiere die Partition durch

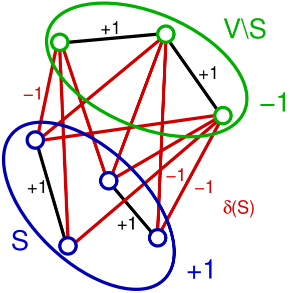
$$x \in \{-1, 1\}^n \quad \text{mit} \quad x_i = \begin{cases} 1 & i \in S \\ -1 & i \in V \setminus S \end{cases}$$

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

Graphenpartition: Max-Cut

Gegeben: Graph $G = (V, E)$, $V = \{1, \ldots, n\}$,
$E \subseteq \{ij : i, j \in V, i < j\}$, Kantengewichte a_{ij}

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



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

Modellierung: Repräsentiere die Partition durch

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

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

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

[$C \in \mathcal{S}^n$: $C_{ii} = \frac{1}{4} \sum_{j : ij \in E} a_{ij}$ (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: $x^T Cx = \langle Cx, x \rangle = \left\langle C, xx^T \right\rangle$

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

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

Semidefinite Max-Cut Relaxation

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

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

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

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

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

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

Semidefinite Max-Cut Relaxation

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

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

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

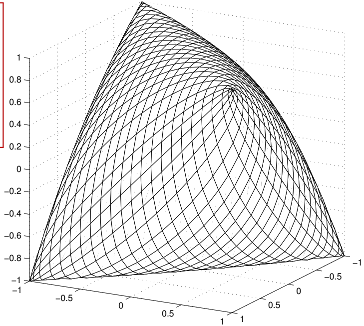
Relaxationsidee: Ersetze xx^T durch eine positiv semidefinite Matrix X.

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

[mit Rang 1 $\Leftrightarrow$ (MC), $\mathcal{NP}$-vollst.]

Illustration für $n = 3$: Den Rand beschreibt

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



Semidefinite Max-Cut Relaxation

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

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

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

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

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

[mit Rang 1 $\Leftrightarrow$ (MC), $\mathcal{NP}$-vollst.]

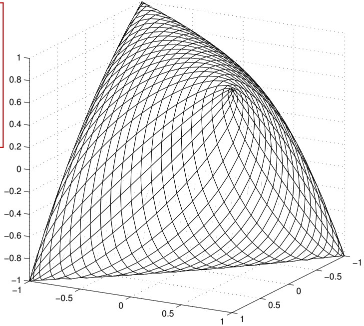


Illustration für $n = 3$: Den Rand beschreibt

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

[GW95]: Randomisiertes Runden von $X = R^T R$ ergibt 0.878 Approximation

Momentenmatrizen und Polynomoptimierung [Lasserre]

Polynom $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ vom Grad $2m$. Finde

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

Momentenmatrizen und Polynomoptimierung [Lasserre]

Polynom $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ vom Grad $2m$. Finde

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

Momentenmatrizen und Polynomoptimierung [Lasserre]

Polynom $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ vom Grad $2m$. Finde

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

Momentenmatrizen und Polynomoptimierung [Lasserre]

Polynom $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ vom Grad $2m$. Finde

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

Ist $y = (y_\alpha)$ Momentenvektor einer W-Verteilung, muss die Momentenmatrix

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

positive semidefinite sein (notwendig, nicht hinreichend).

Momentenmatrizen und Polynomoptimierung [Lasserre]

Polynom $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ vom Grad $2m$. Finde

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

Ist $y = (y_\alpha)$ Momentenvektor einer W-Verteilung, muss die Momentenmatrix

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

positive semidefinite sein (notwendig, nicht hinreichend).

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

Momentenmatrizen und Polynomoptimierung [Lasserre]

Polynom $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} \; > -\infty$ vom Grad $2m$. Finde

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

Ist $y = (y_\alpha)$ Momentenvektor einer W-Verteilung, muss die Momentenmatrix

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

positive semidefinite sein (notwendig, nicht hinreichend).

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

Relaxiere Minimierung über W-Verteilungen zu

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

Momentenmatrizen und Polynomoptimierung [Lasserre]

Polynom $p(x) = \sum_{\alpha \in \mathbb{N}^n} p_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n} > -\infty$ vom Grad $2m$. Finde

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

Ist $y = (y_\alpha)$ Momentenvektor einer W-Verteilung, muss die Momentenmatrix

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

positive semidefinite sein (notwendig, nicht hinreichend).

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

Relaxiere Minimierung über W-Verteilungen zu

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

exakt $\Leftrightarrow$ $p(x) - p_*$ ist Summe von Quadraten von Polynomen (SOS).

Sum of Squares-Zerlegung

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ Polynom $p(x) - \rho$ vom Grad $2m$ ist nichtnegativ.

Sum of Squares-Zerlegung

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ Polynom $p(x) - \rho$ vom Grad $2m$ ist nichtnegativ.
Ist $p(x) - \rho$ ein SOS-Polynom, gilt sicher $\rho \leq \min_x p(x)$. Prüfe, ob

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

für Polynome f_i mit Grad $\leq m$.

Sum of Squares-Zerlegung

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ Polynom $p(x) - \rho$ vom Grad $2m$ ist nichtnegativ.
Ist $p(x) - \rho$ ein SOS-Polynom, gilt sicher $\rho \leq \min_x p(x)$. Prüfe, ob

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

für Polynome f_i mit Grad $\leq m$.
Setze $z = (1, x_1, x_2, \dots, x_n, x_1x_2, x_1x_3, \dots, x_n^m)$, und schreibe f_i als

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

Sum of Squares-Zerlegung

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ Polynom $p(x) - \rho$ vom Grad $2m$ ist nichtnegativ.
Ist $p(x) - \rho$ ein SOS-Polynom, gilt sicher $\rho \leq \min_x p(x)$. Prüfe, ob

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

für Polynome f_i mit Grad $\leq m$.
Setze $z = (1, x_1, x_2, \ldots, x_n, x_1x_2, x_1x_3, \ldots, x_n^m)$, und schreibe f_i als

$$f_i(x) = a_i^T z, \quad \text{dann ist} \quad \sum f_i(x)^2 = z^T AA^T z \text{ mit } H = AA^T \succeq 0.$$

Koeffizientenvergleich liefert: So eine Darstellung existiert g.d.w.

$$\exists H \succeq 0 : \qquad \sum_{\beta+\gamma=\alpha} H_{\beta,\gamma} = p_\alpha \quad \text{für alle Monome } \alpha$$

Sum of Squares-Zerlegung

$\rho \leq \min_x p(x) \quad \Leftrightarrow \quad$ Polynom $p(x) - \rho$ vom Grad $2m$ ist nichtnegativ.
Ist $p(x) - \rho$ ein SOS-Polynom, gilt sicher $\rho \leq \min_x p(x)$. Prüfe, ob

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

für Polynome f_i mit Grad $\leq m$.
Setze $z = (1, x_1, x_2, \ldots, x_n, x_1x_2, x_1x_3, \ldots, x_n^m)$, und schreibe f_i als

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

Koeffizientenvergleich liefert: So eine Darstellung existiert g.d.w.

$$\exists H \succeq 0 : \qquad \sum_{\beta+\gamma=\alpha} H_{\beta,\gamma} = p_\alpha \quad \text{für alle Monome } \alpha$$

Versuche nun $\min_x p(x) = p_*$ über Maximierung von p_0 zu bestimmen:

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

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

Falls $p(x) - p_*$ nicht SOS ist ...

kann man über Kompaktheitsannahmen und Hierarchien von semidefiniten Relaxationen p_* beliebig annähern.

Der Ansatz ist auf durch Polynomungleichungen beschränkte Mengen erweiterbar $\rightarrow$ GloptiPoly

Für Polynomgleichungen $x_i^2 = x_i$ $(x \in \{0, 1\}^n)$ ist die Anfangsrelaxation äquivalent zur Max-Cut-Relaxation.

Danke für Ihre Aufmerksamkeit!