Nichtlineare Optimierung

(mit Schwerpunkt: semidefinite Optimierung)

Mathematisches Programm:

gegeben: Funktionen $f_0, f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$
Konstanten $b_1, \dots, b_m \in \mathbb{R}$

gesucht:

$$\begin{aligned} &\text{minimiere} && f_0(x) \\ &\text{so dass} && x \in \mathbb{R}^n \\ & && f_i(x) \le b_i, \quad i = 1, \dots, m. \end{aligned} \qquad \text{(MP)}$$

f_0: Zielfkt; $f_1, \dots, f_m$: Nebenbedingungsfkt.
x: Optimierungsvariable, $x \in \mathbb{R}^n$ mit $f_i(x) \le b_i$: gültige Lösung; x gültige Lösung mit $f_0(x) \le f_0(x')$ für alle gültigen Lösungen x': optimale Lösung

Lineares Programm (LP): Alle Fkt. $f_0, \dots, f_m$ sind linear.

- Theorie und Praxis von LPs ist sehr gut verstanden (→ Vorlesung OR).

<u>Nichtlineares Programm (NLP)</u>: Eine der Fkt. $f_0, \dots, f_m$ ist nicht linear

– Hoffnungslos! Viel zu allgemein.

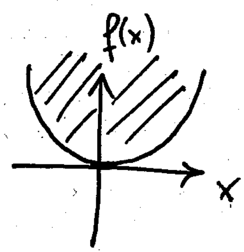
Spezieller (aber immer noch zu allg.): <u>Konvexe Programme</u>

Alle $f_0, \dots, f_m$ sind konvexe Funktionen

<u>Def.</u>: Fkt. $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ heißt <u>konvex</u>, falls ihr Epigraph

$$\mathrm{epi}(f) = \{ (x, \alpha) \in \mathbb{R}^{n+1} : \alpha \geq f(x) \}$$

eine konvexe Menge ist.

<u>Bsp.</u>: $f(x) = x^2$, $\mathrm{epi}(f) =$ 

<u>Ungleichung von Jensen</u>: Ist $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ konvex, dann

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$$

für alle $t \in [0,1]$, $x, y \in \mathbb{R}^n$.

<u>Bew.</u>: Es ist $\tilde{x} = (x, f(x)) \in \mathrm{epi}(f)$, $\tilde{y} = (y, f(y)) \in \mathrm{epi}(f)$.

Also $t\tilde{x} + (1-t)\tilde{y} \in \mathrm{epi}(f)$.

D.h.

$$\big(tx + (1-t)y,\ t f(x) + (1-t) f(y)\big) \in \operatorname{epi}(f),$$

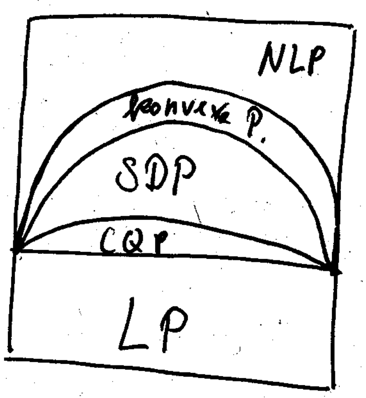
also

$$f(tx + (1-t)y) \le t f(x) + (1-t) f(y). \qquad \boxtimes$$

Fundamentale Eigenschaften konvexer Programme:

- Die Menge der gültigen Lösungen ist konvex.
  [folgt unmittelbar aus der Ungleichung von Jensen]
- Jedes lokale Minimum ist eine optimale Lösung
  [→ Übung].
- Immer noch zu allgemein: Die meisten NLPs können konvexifiziert werden [→ Ende der Vorlesung].

Übersicht:



SDP = semidefinite Programme

CQP = konische quadratische Programme („conic quadratic p.")

Kapitel I Konische Optimierung

§ 1 Konvexe Kegel

Def. 1 a) $K \subseteq \mathbb{R}^n$, $K \neq \emptyset$ heißt (konvexer) Kegel, falls $\forall \alpha, \beta \in \mathbb{R}_{\geq 0}$ $\forall x, y \in K$: $\alpha x + \beta y \in K$.

b) Ein konvexer Kegel K heißt spitz, falls aus $x, -x \in K$ stets $x = 0$ folgt.

c) Ein konvexer Kegel K heißt ordentlich, falls er spitz, abgeschlossen und volldimensional ist.

Ordentliche Kegel definieren Positivitätsbereiche und eine partielle Ordnung auf $\mathbb{R}^n$:

$$x \geq_K y \iff x - y \in K$$

- Falls $K = [0, \infty) \subseteq \mathbb{R}$, dann ist $x \geq_K y$ nichts anderes als $x \geq y$.
- Es kann passieren, dass $x, y \in \mathbb{R}^n$ nicht vergleichbar sind, d.h. es ist weder $x \geq_K y$ noch $y \geq_K x$. [$\rightarrow$ partielle Ordnung]

Proposition 2 Sei $K \subseteq \mathbb{R}^n$ ein spitzer, konvexer Kegel.
Dann gilt $\forall\, x, y, z, x', y' \in \mathbb{R}^n$ $\alpha \in \mathbb{R}_{\geq 0}$:

a) Reflexivität $\quad x \geq_K x$

b) Antisymmetrie $\quad x \geq_K y,\ y \geq_K x \Rightarrow x = y$

c) Transitivität $\quad x \geq_K y,\ y \geq_K z \Rightarrow x \geq_K z$

d) Homogenität $\quad x \geq_K y \Rightarrow \alpha x \geq_K \alpha y$

e) Additivität $\quad x \geq_K y,\ x' \geq_K y' \Rightarrow x + x' \geq_K y + y'$.

Bew.: z.B. b) Aus $x \geq_K y$, $y \geq_K x$ folgt $x - y,\ -(x-y) \in K$.
Da K spitz, muss $x - y = 0$ sein. $\boxtimes$

Falls K voll-dimensional ist, dann ist sein Inneres („interior")

$$\operatorname{int} K = \{x \in \mathbb{R}^n : \exists\, \varepsilon > 0 : B(x, \varepsilon) \subseteq K\} \neq \emptyset,$$

wobei $B(x, \varepsilon) = \{y \in \mathbb{R}^n : \|y - x\| \leq \varepsilon\}$. Dann kann man die strikte Ungleichung $\quad x >_K y \iff x - y \in \operatorname{int} K$ def.

Falls K abgeschlossen ist, ist folgender Grenzübergang erlaubt:

$$\lim_{i \to \infty} x_i = x,\quad \lim_{i \to \infty} y_i = y,\quad x_i \geq_K y_i \implies x \geq y.$$

§ 2 Konstruktionen und Beispiele

Konstruktionen

<u>Def. 1</u> a) Sei $A \subseteq \mathbb{R}^n$ eine Menge. Die <u>konische Hülle</u> von A ist

$$\operatorname{cone} A = \left\{ y : N \in \mathbb{N},\ a_1, \ldots, a_N \in A,\ \alpha_1, \ldots, \alpha_N \in \mathbb{R}_{\geq 0} : y = \sum_{i=1}^{N} \alpha_i a_i \right\}$$

b) Falls $|A| < \infty$, dann heißt $\operatorname{cone} A$ <u>endlich erzeugt</u>.

<u>Satz 2</u> Sei $A \subseteq \mathbb{R}^n$.

a) Es gilt $\operatorname{cone} A = \bigcap\limits_{\substack{K \supseteq A \\ K \text{ konvexer Kegel}}} K$.

b) (Minkowski - Weyl)

Falls $A = \{a_1, \ldots, a_N\}$, dann gibt es eine Matrix $B \in \mathbb{R}^{M \times n}$ mit

$$\operatorname{cone} A = \{ x \in \mathbb{R}^n : Bx \leq 0 \}.$$

<u>Bew.</u>: $\longrightarrow$ Vorlesung OR.

Def. 3 Seien $K_1 \subseteq \mathbb{R}^{n_1}$, $K_2 \subseteq \mathbb{R}^{n_2}$ konvexe Kegel. Ihr <u>direktes Produkt</u> ist

$$K_1 \times K_2 = \{(x_1, x_2) : x_1 \in K_1, x_2 \in K_2\} \subseteq \mathbb{R}^{n_1+n_2}.$$

Offensichtlich: $K_1 \times K_2$ ist konvexer Kegel.

Def. 4 Sei $K \subseteq \mathbb{R}^n$ ein konvexer Kegel. Dann ist sein <u>dualer Kegel</u> definiert als

$$K^* = \{y \in \mathbb{R}^n : x^T y \geq 0 \text{ für alle } x \in K\}.$$

Satz 5 Sei $K \subseteq \mathbb{R}^n$ ein konvexer Kegel. Dann:

a) $K \subseteq (K^*)^*$

b) Falls K abgeschlossen, dann $K = (K^*)^*$.

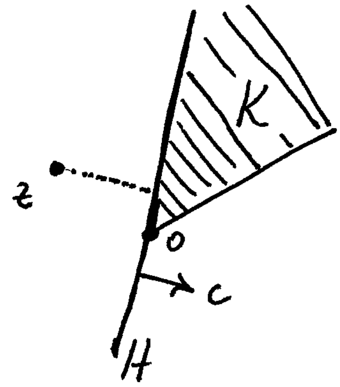
Bew.: a) klar

für b) benötigt man

Satz 6 Sei K ein konvexer, abgeschlossener Kegel, sei $z \notin K$. Dann gibt es ein $c \in \mathbb{R}^n$ mit

$$c^T z < 0 \quad \text{und} \quad c^T x \geq 0 \quad \text{für alle } x \in K.$$

D.h. die Hyperebene $H = \{ y : c^T y = 0 \}$ ist Trennhyperebene von $\{z\}$ und K. Sie trennt z strikt und ist Stützhyperebene an K.



<u>Bew.:</u> Satz 5 b)

Zu zeigen $K \supseteq (K^*)^*$.

Sei $z \in \mathbb{R}^n \setminus K$. Dann gibt es nach Satz 6 ein $c \in \mathbb{R}^n$ mit

$$c^T z < 0 \quad \text{und} \quad c^T x \geq 0 \quad \text{für alle } x \in K.$$

D.h. $c \in K^*$ und $z \notin (K^*)^*$.

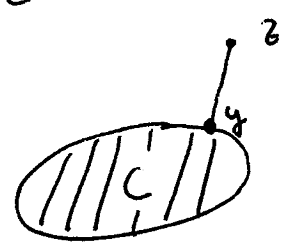
<u>Zum Beweis von Satz 6:</u>

<u>Wdh OR:</u> Metrische Projektion

Sei $C \subseteq \mathbb{R}^n$ abgeschlossen und konvex, $C \neq \emptyset$. Sei $z \notin C$.

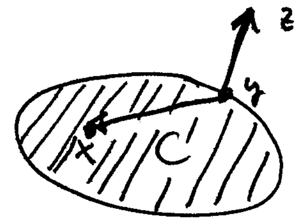
Dann $\exists !\ y \in C$ mit $\|y - z\| = \inf\limits_{x \in C} \|x - z\|$.

<u>Notation:</u> $\Pi_C(z) = y$



Zusatz: $(z-y)^T(x-y) \le 0$ für alle $x \in C$.

Bew.: Sei $\alpha \in [0,1]$. Dann ist



$$\begin{aligned}\|z-y\|^2 &\le \|z-((1-\alpha)y+\alpha x)\|^2 \\ &= \|z-y+\alpha(y-x)\|^2 \\ &= \|z-y\|^2 + 2\alpha(z-y)^T(y-x) + \alpha^2\|y-x\|^2\end{aligned}$$

Demnach für $\alpha \in (0,1]$:

$$(z-y)^T(x-y) \le \frac{\alpha}{2}\|y-x\|^2$$

und die Beh. folgt, da die rechte Seite beliebig klein ≥ 0 werden kann.

Bew.: (Satz 6)

Setze $c = y - z$.

Beh.: $c^T y = 0$

Bew.: Ang. $c^T y \ne 0$.

1. Fall: $c^T y > 0$.

Dann ist $(1-\varepsilon)y \in K$ für kleines $\varepsilon > 0$ und

$$\begin{aligned}\|(1-\varepsilon)y - z\|^2 &= \|y - z - \varepsilon y\|^2 \\ &= \|y-z\|^2 - 2\varepsilon(y-z)^T y + \varepsilon^2\|y\|^2\end{aligned}$$

Für $\varepsilon < \frac{2\, c^T y}{\|y\|^2}$ ist

$$-2\varepsilon (y-z)^T y + \varepsilon^2 \|y\|^2 < 0,$$

was ein Widerspruch zur Minimalität von $\|y-z\|^2$ ist.

2. Fall $c^T y < 0$

genauso: mit $(1+\varepsilon) y \in K$ und $\varepsilon \in \left(0, \frac{-2\, c^T y}{\|y\|^2}\right)$.

Nun ist

$$0 < c^T c = c^T (y-z) = c^T y - c^T z = -c^T z,$$

also $0 > c^T z$.

Andererseits für $x \in K$ nach dem Zusatz

$$0 \geq (z-y)^T (x-y) = -c^T (x-y) = -c^T x,$$

also $c^T x \geq 0$.

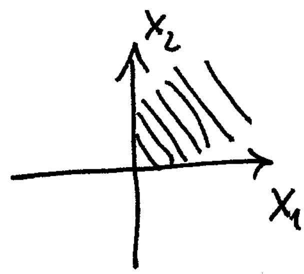
Beispiele

Bsp. 7 Sei $U \subseteq \mathbb{R}^n$ ein Untervektorraum. Dann ist U ein konvexer Kegel.

Dualer Kegel: $U^* = U^\perp$.

Bsp. 8 Nichtnegativer Orthant

$$\mathbb{R}^n_{\geq 0} = \{x \in \mathbb{R}^n : x_1 \geq 0, \dots, x_n \geq 0\}$$
$$= \mathbb{R}_{\geq 0} \times \dots \times \mathbb{R}_{\geq 0}$$
$$= \operatorname{cone}\{e_1, \dots, e_n\}.$$



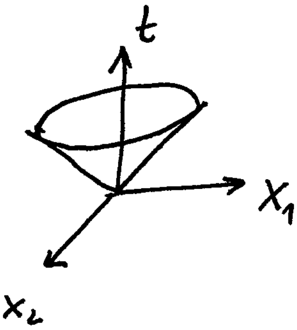
Dualer Kegel: $(\mathbb{R}^n_{\geq 0})^* = \mathbb{R}^n_{\geq 0}$.

[Allgemein gilt: $(K_1 \times K_2)^* = K_1^* \times K_2^*$
Bew.: leichte Übung].

Bsp. 9 Lorentz-Kegel („ice cream cone")

$$\mathcal{L}^{n+1} = \{(x,t) \in \mathbb{R}^{n+1} : \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \leq t\}$$

ist ein ordentlicher konvexer Kegel



Dualer Kegel:
$$(\mathcal{L}^{n+1})^* = \mathcal{L}^{n+1}.$$

$\longrightarrow$ Aufgabe 1.4.

Bsp. 10 Kegel der positiv semidefiniten Matrizen

$$S^n = \{ X \in \mathbb{R}^{n \times n} : X \text{ symmetrisch, d.h. } X_{ij} = X_{ji} \}.$$

$\frac{n(n+1)}{2}$ - dimensionaler Vektorraum der symmetrischen Matrizen.

Skalarprodukt:
$$\langle X, Y \rangle = Tr(Y^T X) = \sum_{i=1}^{n} \sum_{j=1}^{n} X_{ij} Y_{ij},$$

wobei $Tr(X) = \sum_{i=1}^{n} X_{ii}$ die Spur („trace") von X ist.

$(S^n, \langle \cdot, \cdot \rangle)$ ist isometrisch zum Raum $\left(\mathbb{R}^{\frac{n(n+1)}{2}}, (x,y) \mapsto x^T y\right)$

mittels der Abbildung

$$T : S^n \to \mathbb{R}^{\frac{n(n+1)}{2}}$$

$$\begin{aligned} X \longmapsto (X_{11}, \sqrt{2}\, X_{12}, \sqrt{2}\, X_{13}, \ldots, \sqrt{2}\, X_{1n}, \\ X_{22}, \sqrt{2}\, X_{23}, \ldots, \sqrt{2}\, X_{2n}, \\ \ldots \\ X_{nn}) \end{aligned}$$

obere Dreiecksmatrix.

$X \in S^n$ heißt <u>positiv semidefinit</u>, falls

$$\forall x \in \mathbb{R}^n: \quad x^T X x \geq 0.$$

<u>Notation:</u> $X \succeq 0$

`$\succeq$`

$S^n_{\succeq 0} = \{X \in S^n : X \succeq 0\}$ ist der Kegel der positiv semidefiniten Matrizen („PSD-Kegel")

<u>Theorem</u> („Spektralzerlegung")
Sei $X \in S^n$. Dann gibt es eine Orthonormalbasis $u_1, \dots, u_n \in \mathbb{R}^n$ und $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, so dass

$$X = \sum_{i=1}^{n} \lambda_i u_i u_i^T. \qquad \text{(eine \underline{Spektralzerlegung} von } X\text{)}$$

Dabei sind $\lambda_1, \dots, \lambda_n$ die Eigenwerte von X und $u_1, \dots, u_n$ zugehörige Eigenvektoren.

<u>In Matrixform:</u> $X = PDP^T$, wobei $P \in O(n)$ eine orthogonale Matrix und $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ eine Diagonalmatrix.

<u>Bew.:</u> $\longrightarrow$ Vorlesung Lineare Algebra 2.

<u>Korollar</u> a) $X \in S^n_{\geq 0} \iff \lambda_1, \dots, \lambda_n \geq 0$.

b) $\left(S^n_{\geq 0}\right)^* = S^n_{\geq 0}$.

<u>Bew.:</u> a) <u>„$\Longrightarrow$"</u> Spektralzerlegung $X = \sum_{i=1}^n \lambda_i u_i u_i^T$.

Dann gilt für $j = 1, \dots, n$:

$$0 \leq u_j^T X u_j = u_j^T \sum_{i=1}^n \lambda_i u_i u_i^T u_i = \sum_{i=1}^n \lambda_i \underbrace{\left(u_i^T u_j\right)^2}_{= \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}} = \lambda_j.$$

<u>„$\Longleftarrow$"</u> Sei $x \in \mathbb{R}^n$. Dann

$$x^T X x = x^T \sum_{i=1}^n \lambda_i u_i u_i x = \sum_{i=1}^n \underbrace{\lambda_i}_{\geq 0} \underbrace{(u_i^T x)^2}_{\geq 0} \geq 0.$$

b) <u>„$\subseteq$"</u> Für $x \in \mathbb{R}^n$ ist $x x^T \in S^n_{\geq 0}$. Sei $Y \in \left(S^n_{\geq 0}\right)^*$.

Dann $\quad 0 \leq \langle x x^T, Y \rangle = x^T Y x. \Rightarrow Y \in S^n_{\geq 0}$.

<u>„$\supseteq$"</u> Sei $X \in S^n_{\geq 0}$. Dann gilt für $Y \in S^n_{\geq 0}$ mit Spektralzerlegung $Y = \sum_{i=1}^n \lambda_i u_i u_i^T$ nach a) $\lambda_1, \dots, \lambda_n \geq 0$.

Desweiteren

$$\langle X, Y \rangle = \sum_{i=1}^n \underbrace{\lambda_i}_{\geq 0} \underbrace{u_i^T X u_i}_{\geq 0} \geq 0 \Rightarrow X \in \left(S^n_{\geq 0}\right)^*.$$

$\boxtimes$

Mehr zu $S^n_{\geq 0}$: $\rightarrow$ Kapitel II.

<u>Bsp. 11</u> Endlich erzeugte Kegel

$K = \operatorname{cone}\{a_1, \dots, a_N\} \subseteq \mathbb{R}^n$.

Nach dem Theorem von Minkowski-Weyl gibt es $B \in \mathbb{R}^{M \times n}$ mit $K = \{x \in \mathbb{R}^n : Bx \leq 0\}$.

<u>Dualer Kegel:</u> $K^* = \operatorname{cone}\{-b_1, \dots, -b_M\}$,

wobei $b_1^T, \dots, b_M^T$ die Zeilen von B sind.

<u>Bew.:</u> <u>„$\supseteq$":</u> Sei $y = \sum_{j=1}^{M} \beta_j (-b_j)$ mit $\beta_1, \dots, \beta_M \geq 0$.

Dann gilt für $x \in K$:

$$x^T y = -\sum_{j=1}^{M} \beta_j \underbrace{b_j^T x}_{\leq 0} \geq 0. \implies y \in K^*.$$

<u>„$\subseteq$":</u> Ang. $y \in K^* \setminus \operatorname{cone}\{-b_1, \dots, -b_M\}$. Dann gibt es nach Satz 6 ein $c \in \mathbb{R}^n$ mit

$$c^T y < 0 \quad \text{und} \quad c^T x \geq 0 \quad \forall x \in \operatorname{cone}\{-b_1, \dots, -b_M\}.$$

Also $Bc \leq 0$, d.h. $c \in K$. Dann kann aber nicht $y \in K^*$ sein.

Bem. 12 Koecher - Vinberg Klassifikation

Sei $K \subseteq \mathbb{R}^n$ ein ordentlicher, selbst-dualer ($K = K^{**}$), konvexer Kegel. Er heißt homogen, wenn die Menge der Kegelautomorphismen, $\Gamma \subseteq GL_n(\mathbb{R})$ mit $\Gamma K = K$, eine abgeschlossene Untergruppe ist, die transitiv operiert (d.h. $\forall x, y \in K \; \exists \gamma \in \Gamma : \gamma x = y$).

Solche Kegel sind direkte Produkte von

(i) $\mathcal{L}^{n+1}$

(ii) $S^n_{\geq 0}$

(iii) Kegel der pos. semidefiniten hermiteschen Matrizen über $\mathbb{C}$

(iv) ——— " ——— über $\mathbb{H}$, den Quaternionen

(v) einem speziellen 27-dim. Kegel von 3×3 hermiteschen Matrizen über $\mathbb{O}$, den Oktaven (oder Oktonionen).

→ Stichwort: Euklidische Jordanalgebren.

§ 3 Konische Programme

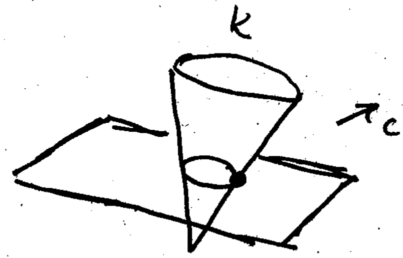
Def. 1 Sei $K \subseteq \mathbb{R}^n$ ein ordentlicher Kegel.

Gegeben seien $c, a_1, \dots, a_m \in \mathbb{R}^n$, $b_1, \dots, b_m \in \mathbb{R}$. Ein primales konisches Programm (in Standardform) ist das folgende Maximierungsproblem:

$$
(\mathrm{P}) \qquad
\begin{aligned}
\sup \quad & c^T x \\
& x \in \mathbb{R}^n \\
& a_j^T x = b_j \quad \text{für } j \in [m] = \{1, \dots, m\} \\
& x \in K.
\end{aligned}
$$

Das zugehörige duale konische Programm ist das folgende Minimierungsproblem:

$$
(\mathrm{D}) \qquad
\begin{aligned}
\inf \quad & \sum_{j=1}^m b_j y_j = b^T y \\
& y \in \mathbb{R}^m \\
& \sum_{j=1}^m y_j a_j - c \in K^*.
\end{aligned}
$$



* primales konisches Programm: maximiere eine lineare Fkt über Durchschnitt eines Kegels mit einem affinen Unterraum.

* Anstatt $a_j^T x = b_j$ für $j \in [m]$ schreibt man auch oft $Ax = b$ mit $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

Genauso: $A^T y - c$ für $\sum_{j=1}^m y_j a_j - c$.

* $x \in \mathbb{R}^n$ heißt <u>zulässige Lösung</u> von (P), falls $Ax = b$ und $x \in K$ gilt. x heißt <u>strikt</u> <u>zulässig</u>, falls x zulässig und $x \in \operatorname{int} K$.

Genauso: $y \in \mathbb{R}^m$ heißt <u>zulässig</u> für (D), falls $A^T y - c \in K^*$, und strikt zulässig, falls $A^T y - c \in \operatorname{int} K^*$.

Geometrische Interpretation von (P) und (D)

Definiere die Untervektorräume

$$L = \{ x \in \mathbb{R}^n : Ax = 0 \}$$

$$L^\perp = \{ y \in \mathbb{R}^n : x^T y = 0 \text{ für alle } x \in L \}$$
$$= \{ y^T A : y \in \mathbb{R}^m \}.$$

Ang. es gibt ein $x_0 \in \mathbb{R}^n$ mit $Ax_0 = b$. [Ansonsten hätte (P) keine zulässige Lösung].

Dann ist

$$b^T y = (Ax_0)^T y = x_0^T A^T y = x_0^T (A^T y - c) + x_0^T c.$$

D.h.

$$(P) = \sup\{ c^T x : x \in K \cap (x_0 + L)\}$$

$$(D) = c^T x_0 + \inf\{ x_0^T z : z \in K^* \cap (-c + L^\perp)\}.$$

Wenn man nun (D) in primaler Form schreibt

$$c^T x_0 - \sup\{ -x_0^T z : z \in K^* \cap (-c + L^\perp)\}$$

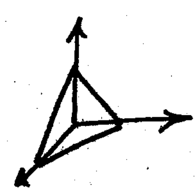
und davon das Duale bildet, erhält man (P) zurück, weil nach Satz 2.5 gilt $(K^*)^* = K$.

Beispiele

Bsp. 2 $K = \mathbb{R}^n_{\geq 0}$. Lineare Programme (LP)

(P) $\sup c^T x$
$x \in \mathbb{R}^n$
$Ax = b$
$x \geq 0$ $(x \in \mathbb{R}^n_{\geq 0})$
↑
komponentenweise

(D) $\inf b^T y$
$y \in \mathbb{R}^m$
$A^T y - c \geq 0$



$(1\ 1\ 1)x = 1$
$x \geq 0$

Bsp. 3 $\quad K = \mathcal{L}^{n_1+1} \times \mathcal{L}^{n_2+1} \times \dots \times \mathcal{L}^{n_r+1}$

Konische quadratische Programmierung (CQP)

$$(P) \quad \sup \; ((c_1,\gamma_1),\dots,(c_r,\gamma_r))^T((x_1,t_1),\dots,(x_r,t_r))$$

$$((x_1,t_1),\dots,(x_r,t_r)) \in \mathbb{R}^{n_1+1} \times \dots \times \mathbb{R}^{n_r+1}$$

$$((a_{j1},\alpha_{j1}),\dots,(a_{jr},\alpha_{jr}))^T((x_1,t_1),\dots,(x_r,t_r)) = b_j, \quad j \in [m]$$

$$((x_1,t_1),\dots,(x_r,t_r)) \in \mathcal{L}^{n_1+1} \times \dots \times \mathcal{L}^{n_r+1}$$

$$(D) \quad \inf \; b^T y$$

$$y \in \mathbb{R}^m$$

$$\sum_{j=1}^m y_j \, ((a_{j1},\alpha_{j1}),\dots,(a_{jr},\alpha_{jr})) - ((c_1,\gamma_1),\dots,(c_r,\gamma_r)) \in \mathcal{L}^{n_1+1} \times \dots \times \mathcal{L}^{n_r+1}.$$

Die letzte Bedingung kann schöner / intuitiver aufgeschrieben werden: Definiere die Matrizen

$$A_i = \left[a_{1i} \; \dots \; a_{mi} \right] \in \mathbb{R}^{n_i \times m}, \quad i = 1,\dots,r$$

und die Vektoren $d_i = (\alpha_{1i}, \dots, \alpha_{mi})^T$, $i = 1, \dots, r$.

Dann ist die letzte Bed. äquivalent zu

$$\| A_i y - c_i \| \leq d_i^T y - \gamma_i \quad , \quad i = 1, \dots, r.$$

Insbesondere sieht man, dass LP ein Spezialfall von CQP ist (setze $A_i = 0$, $c_i = 0$).

<u>Bsp. 4</u> $K = S^n_{\succeq 0}$ Semidefinite Programmierung (SDP).

$$\text{(P)} \quad \sup \langle C, X \rangle$$
$$X \in S^n$$
$$\langle A_j, X \rangle = b_j \, , \quad j = 1, \dots, m$$
$$X \succeq 0$$

$$\text{(D)} \quad \inf b^T y$$
$$y \in \mathbb{R}^m$$
$$\sum_{j=1}^m y_j A_j - C \succeq 0$$

Ungleichungen der Form $\sum_{j=1}^m y_j A_j - C \succeq 0$ heissen <u>lineare Matrixungleichungen</u> (LMI = linear matrix inequality).

Lemma 5 (Schurkomplement)

Sei $X \in \mathcal{S}^n$ eine Matrix mit Blockgestalt

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \quad \text{mit} \quad A \in \mathcal{S}^p, \; C \in \mathcal{S}^{n-p}, \; B \in \mathbb{R}^{p \times (n-p)}.$$

Falls A invertierbar ist gilt

$$X \succeq 0 \iff A \succeq 0 \text{ und } C - B^T A^{-1} B \succeq 0.$$

[Die Matrix $C - B^T A^{-1} B$ heißt das Schurkomplement von A in X.]

Bew.: Durch Nachrechnen:

$$X = P^T \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix} P \quad \text{mit} \quad P = \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \quad \boxtimes$$

Korollar 6

$$\mathcal{L}^{n+1} = \{ (x,t) \in \mathbb{R}^{n+1} : \|x\| \le t \}$$
$$= \left\{ (x,t) \in \mathbb{R}^{n+1} : \begin{bmatrix} t I_n & x \\ x^T & t \end{bmatrix} \succeq 0 \right\}.$$

Bew.: Falls $t = 0$ ist, muß $x = 0$ sein.

Falls $t > 0$ ist, ist $t I_n \succ 0$.

Betrachte Schurkomplement tI_n in $\begin{bmatrix} tI_n & x \\ x^T & t \end{bmatrix}$.

Die letztere Matrix ist nach Lemma 5 psd gdw

$$t - x^T \frac{1}{t} I_n x \geq 0 \iff t^2 \geq x^T x \iff \|x\| \leq t \quad \boxtimes$$

Somit ist CQP ein Spezialfall von SDP.

§4 Alternativsätze

<u>Erinnerung</u>: Farkas Lemma für LPs

<u>Lemma 1</u> Sei $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Dann gilt

entweder i) $\exists\, x \geq 0 : Ax = b$

oder ii) $\exists\, y : A^T y \geq 0, \quad b^T y < 0.$

Verallgemeinerungsversuch auf konische Programme:

Es gilt entweder

i) $\exists\, x \in K : Ax = b$

oder ii) $\exists\, y \in \mathbb{R}^m : A^T y \in K^*, \ b^T y < 0.$

Beweisversuch:

$i) \Rightarrow \neg ii)$: Ang. x ist zulässig für i) und y ist zulässig für ii). Dann

$$0 \leq (A^T y)^T x = y^T A x = y^T b < 0 \quad ↯.$$

$\neg i \Rightarrow ii)$: Ang. i) hat keine zulässige Lösung. Dann ist $b \notin \{Ax : x \in K\}$ und $\{Ax : x \in K\}$ ist ein konvexer Kegel. Dann gibt es nach Satz 2.6 ein $y \in \mathbb{R}^m$ mit $y^T b < 0$ und $y^T A x \geq 0$ für alle $x \in K$. D. h. $(y^T A)^T = A^T y \in K^*$, also folgt ii).

Bsp. 2 Definiere für $i \leq j$ die Matrix

$$E_{ij} = \frac{1}{2}\left(e_i e_j^T + e_j e_i\right).$$

und betrachte die Systeme

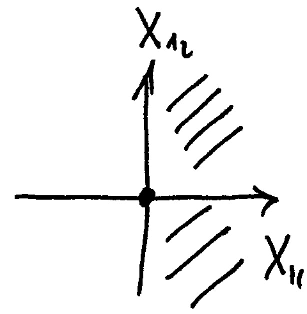
i) $\langle E_{11}, X \rangle = 0, \quad \langle E_{12}, X \rangle = 1, \quad X \succeq 0, \quad X \in S^2$

ii) $y_1 E_{11} + y_2 E_{12} \succeq 0 \qquad y_2 < 0.$

Beide haben keine zulässige Lösung!

? Wo ist das Problem ?

Der konvexe Kegel



$$A S^2_{\succeq 0} = \left\{ \begin{bmatrix} \langle E_{11}, X \rangle \\ \langle E_{12}, X \rangle \end{bmatrix} : X \in S^2_{\succeq 0} \right\}$$

$$= \left\{ \begin{bmatrix} X_{11} \\ X_{12} \end{bmatrix} : X_{11} > 0,\ X_{12} \in \mathbb{R} \right\} \cup \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$$

ist nicht abgeschlossen.

<u>Def. 3</u> Sei $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Das System

$$Ax = b, \quad x \in K$$

heißt <u>schwach zulässig</u>, falls

$$\forall \varepsilon > 0 \ \exists x \in K : \quad \| Ax - b \| \leq \varepsilon .$$

<u>Satz 4</u> Sei $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Dann gilt

entweder i) Das System $Ax = b$, $x \in K$ ist schwach zulässig.

oder ii) $\exists y \in \mathbb{R}^m : \ A^T y \in K^*, \ b^T y < 0.$

<u>Bew.</u>: Wie oben, nur verwende den Abschluss $\overline{\{ Ax : x \in K \}}$.

Zurück zu Bsp. 2: Das System $\langle E_{11}, X\rangle = 0$, $\langle E_{12}, X\rangle = 1$, $X \geq 0$ ist schwach zulässig.

Korollar 5 Sei $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$. Dann gilt

entweder i) $\exists\, x \in K$: $Ax = 0$ und $c^T x > 0$

oder ii) Das System $A^T y - c \in K^*$ ist schwach zulässig, d.h. $\forall\, \varepsilon > 0\ \exists\, y \in \mathbb{R}^m$ und $z \in K^*$ mit $\| A^T y - c - z \| \leq \varepsilon$.

Bew.: $\rightarrow$ Aufgabe 3.1.

Satz 6 Sei $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Ang. $Ax = b$ hat Lösung x_0. Dann gilt entweder

i) $\exists\, x \in \operatorname{int} K$: $Ax = b$ (strikt zulässige Lösung)

oder

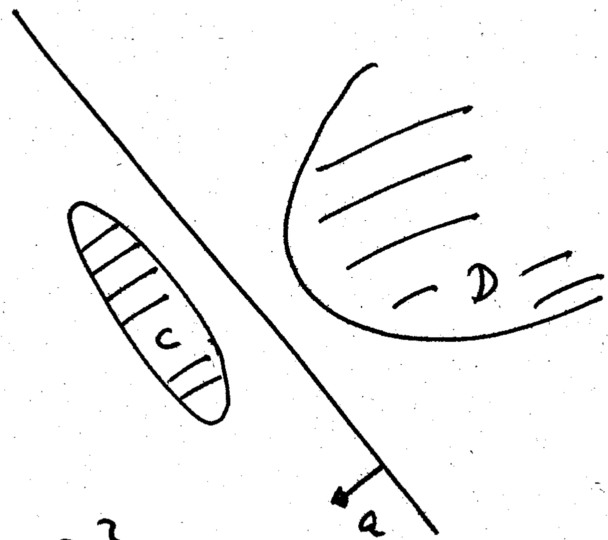
ii) $\exists\, y \in \mathbb{R}^m$: $A^T y \in K^* \setminus \{0\}$ und $b^T y \leq 0$.

Zurück zu Bsp. 2: Hier Bed. ii) ist erfüllt.

Für den Beweis von Satz 6 ist die folgende Variante des Satzes über die Existenz von Trennhyperebenen erforderlich.

<u>Satz 7</u> Seien $C, D \subseteq \mathbb{R}^n$ nichtleere konvexe Mengen mit $C \cap D = \emptyset$. Dann gibt es ein $a \in \mathbb{R}^n \setminus \{0\}$ mit

$$\sup_{x \in C} a^T x \leq \inf_{x \in D} a^T x$$



<u>Bew. skizze:</u> Betrachte die konv. Menge

$$C - D = \{c - d : c \in C, d \in D\}.$$

Nach Vor. ist $0 \notin C - D$. Falls $0 \notin \overline{C-D}$, verwende metrische Projektion zur Def. von a. Falls $0 \in \overline{C-D}$, dann konstruiere Stützhyperebene H durch 0 an $\overline{C-D}$ und verwende einen zu H orthogonal stehenden Vektor als a.

Details: siehe Skript „semidefinite opt." Kapitel 1.1.3.

<u>Bew.</u> (Satz 6)

<u>i) $\Longrightarrow \neg$ ii)</u> Ang. $\exists\, x \in \operatorname{int} K : Ax = b$ und

$\exists\, y : A^T y \in K^* \setminus \{0\}$ mit $b^T y \leq 0$.

Dann

$$0 \leq (A^T y)^T x = y^T A x = y^T b \leq 0.$$

D.h. $(A^T y)^T x = 0$. ↯ zu A 2.1 $\left[x \in \operatorname{int} K \iff y^T x > 0 \ \forall y \in K^* \setminus \{0\} \right]$

<u>¬i) ⟹ ii)</u> Ang. i) hat keine Lösung. Betrachte

$$L = \{x : Ax = 0\}.$$

Dann ist $x_0 + L \cap \operatorname{int} K = \emptyset$. Nach Satz 7 gibt es ein $a \in \mathbb{R}^n \setminus \{0\}$ und $\alpha \in \mathbb{R}$ mit

$$a^T x \geq \alpha \quad \forall x \in K \quad \text{und} \quad a^T x \leq \alpha \quad \forall x \in x_0 + L.$$

Dann ist $0 \geq \alpha$, weil $0 \in K$.

Außerdem $a \in K^*$, weil $a^T(tx) \geq \alpha \quad \forall t > 0, x \in K$.

Gleichzeitig $a \in L^\perp$, weil $a^T(tx + x_0) \leq \alpha \quad \forall t \in \mathbb{R}, x \in L$,

d.h. $a^T x = 0 \quad \forall x \in L$.

Also ist $a = A^T y$ für ein $y \in \mathbb{R}^m$, weil $L^\perp = \{A^T y : y \in \mathbb{R}^m\}$.

Somit $A^T y \in K^* \setminus \{0\}$. Desweiteren

$$y^T b = y^T A x_0 = a^T x_0 \leq \alpha \leq 0. \qquad \boxtimes$$

§ 5 Dualitätstheorie

(P) $p^* = \sup c^T x$
$$Ax = b$$
$$x \in K$$

(D) $d^* = \inf y^T b$
$$A^T y - c \in K^*.$$

<u>Def. 1</u> Die Differenz $d^* - p^*$ heißt <u>Dualitätslücke</u> (duality gap)

<u>Theorem 2</u> a) (schwache Dualität)

Sei x zulässig für (P) und y zulässig für (D). Dann $c^T x \leq y^T b$. Insbesondere $p^* \leq d^*$.

b) (komplementärer Schlupf)

Sei x optimal für (P) und y optimal für (D). Dann
$$(A^T y)^T x = 0.$$

c) (Optimalitätsbedingung)

Sei x zulässig für (P), y zulässig für (D). Dann sind x, y optimal, falls $(A^T y)^T x = 0$.

d) (starke Dualität).

Falls $d^* > -\infty$ und (D) strikt zulässig. Dann gibt es eine zulässige Lösung x von (P) mit $p^* = c^T x$ und es ist $p^* = d^*$.

Falls $p^* < \infty$ und (P) strikt zulässig, dann gibt es eine zulässige Lösung y von (D) mit $d^* = y^T b$ und es ist $p^* = d^*$.

<u>Bem.:</u> c) heißt oft auch <u>KKT-Bedingung</u>
(KKT = Karush-Kuhn-Tucker)

d) heißt oft <u>Slater-Bedingung</u>.

<u>Bew.:</u>

<u>a) - c)</u> $y^T b - c^T x = y^T A x - c^T x = (A^T y - c)^T x \geq 0.$

d) Ang. $d^* > -\infty$ und (D) ist strikt zulässig

<u>z.z.:</u> ~~$\exists x^* \in \mathbb{R}^n$~~: $\exists\, x^* \in \mathbb{R}: A x^* = b, \quad c^T x^* \geq d^*$.

Falls $b = 0$, dann ist $d^* = 0$. Setze $x^* = 0$ und fertig.

Sei also $b \neq 0$ von nun an.

Betrachte die nicht-leere Menge

$$M = \{ A^T y - c : y \in \mathbb{R}^m, \ b^T y \leq d^* \}.$$

Es ist

$$M \cap \operatorname{int} K^* = \emptyset.$$

[Ansonsten gäbe es ein $y \in \operatorname{int} K^*$ mit $y^T b = d^*$. Desweiteren ein $\varepsilon > 0$ mit $B(y, \varepsilon) \subseteq K^*$. Für $y - \varepsilon b \in K^*$ wäre dann $(y - \varepsilon b)^T b = y^T b - \varepsilon b^T b < d^*$. $\lightning$]

Da M und $\operatorname{int} K^*$ konvex sind, gibt es nach Satz 4.7 ein $x \in \mathbb{R}^n \setminus \{0\}$ mit

$$\sup_{z \in M} x^T z \leq \inf_{z \in \operatorname{int} K^*} x^T z = \inf_{z \in K^*} x^T z. \qquad (*)$$

<u>Beh.</u>: 1) $x \in K$

2) $\exists \mu > 0$: $Ax = \mu b$ und $c^T x \geq \mu d^*$

3) $x^* = \frac{x}{\mu}$ ist optimal für (P) [folgt sofort aus 1), 2) und schwacher D.]

1) <u>z.z.:</u> $\inf_{z \in K^*} x^T z \geq 0$. Dann $x \in K^{**} = K$.

Ang. $\exists z \in K^*$ mit $x^T z < 0$. Dann $x^T t z \to -\infty$ für $t \to +\infty$. Widerspruch zu $(*)$ und $d^* > -\infty$.

2) Weil $0 \in K^*$ folgt aus 1): $\inf\limits_{z \in K^*} x^T z = 0$.

Also in (*) $\sup\limits_{z \in M} x^T z \le 0$.

Also

$$b^T y \le d^* \implies x^T (A^T y - c) \le 0$$

$$\iff x^T A^T y \le x^T c.$$

D.h. der Halbraum

$$\{ y : b^T y \le d^* \}$$

ist in dem Halbraum

$$\{ y : (Ax)^T y \le x^T c \}$$

enthalten. D.h. die Vektoren b und Ax sind lin. abh. und zeigen in die gleiche Richtung. Somit $\exists \mu \ge 0 : Ax = \mu b$ und $\mu d^* \le x^T c$. Ang. $\mu = 0$. Da (D) strikt zulässig, $\exists y' : A^T y' - c \in \operatorname{int} K^*$. Nach A.2.1 ist

$$0 < (A^T y' - c)^T x = (y')^T A x - c^T x = (y')^T \mu b - c^T x = -c^T x$$

Also $c^T x < 0$. $\lightning$ $\boxtimes$

Die 2. Beh. in d) folgt aus der 1. Beh durch Dualisieren.

Kapitel II Semidefinite Optimierung

§ 1 Grundlagen

$$(P) \qquad p^* = \sup_{\substack{X \in S^n_{\succeq 0} \\ \langle A_j, X \rangle = b_j, \; j \in [m]}} \langle C, X \rangle$$

$$(D) \qquad d^* = \inf_{\substack{y \in \mathbb{R}^m \\ \sum_{j=1}^m y_j A_j - C \succeq 0}} y^T b$$

wobei $C, A_1, \dots, A_m \in S^n$, $b_1, \dots, b_m \in \mathbb{R}$ gegeben.

Bem. 1 Lösungsverfahren für SDPs

* SDPs sind i.A. effizient (in polynomieller Zeit) lösbar. Durch Anwendung der Ellipsoidmethode und dem Theorem von Grötschel - Lovász - Schrijver ($\rightarrow$ später mehr)
* In der Praxis löst man SDPs mit Innere Punkte Verfahren. Komplexität davon gut, aber nicht vollständig verstanden ($\rightarrow$ Seminar SoSe 2014).

* Online SDP Löser

google: neos sdp csdp

Dateiformat: sparse sdpa

↑ dünn besetzt, nur Matrixeinträge $\neq 0$ werden angegeben

Bsp.:

```
2            ← m = 2
2            ← direktes Produkt von 2 PSD-Kegeln
2 2          ← S^2_{≥0} × S^2_{≥0}
10.0 20.0    ← b = (10.0, 20.0)
0 1 1 1 1.0
0 1 2 2 2.0
0 2 1 1 3.0
0 2 2 2 4.0
1 1 1 1 1.0
1 1 2 2 1.0
2 1 2 2 1.0
2 2 1 1 5.0
2 2 1 2 2.0
2 2 2 2 6.0
```

$$C = \begin{bmatrix} \boxed{\begin{matrix} 1.0 & \\ & 2.0 \end{matrix}} & \\ & \boxed{\begin{matrix} 3.0 & \\ & 4.0 \end{matrix}} \end{bmatrix}$$

$$A_1 = \begin{bmatrix} \boxed{\begin{matrix} 1.0 & \\ & 1.0 \end{matrix}} & \\ & \boxed{\begin{matrix} & \\ & \end{matrix}} \end{bmatrix}$$

$$A_2 = \begin{bmatrix} \boxed{\begin{matrix} & \\ & 1.0 \end{matrix}} & \\ & \boxed{\begin{matrix} 5.0 & 2.0 \\ 2.0 & 6.0 \end{matrix}} \end{bmatrix}$$

(Pathologische) Beispiele

Bsp. 2 d^* nicht angenommen

$$p^* = \sup \langle E_{12}, X \rangle \qquad\qquad d^* = \inf y_1$$

$$X \in S^2_{\geq 0} \qquad\qquad \underbrace{y_1 E_{11} + y_2 E_{22} - E_{12}}_{= \begin{bmatrix} y_1 & -\frac{1}{2} \\ -\frac{1}{2} & y_2 \end{bmatrix}} \succeq 0$$

$$\langle E_{11}, X \rangle = 1, \ \langle E_{22}, X \rangle = 0$$

Hier ist $p^* = d^* = 0$, aber d^* wird nicht angenommen.

Problem: (P) nicht strikt zulässig.

Bsp. 3 positive Dualitätslücke

$$p^* = \sup \langle -E_{11} - E_{22}, X \rangle \qquad d^* = \inf \; y_2$$
$$X \in S^3_{\succeq 0} \qquad y_1 E_{11} + y_2(E_{22} + 2E_{13}) -$$
$$\langle E_{11}, X \rangle = 0, \qquad (-E_{11} - E_{22}) \succeq 0$$
$$\langle E_{22} + 2E_{13}, X \rangle = 1 \qquad \|$$
$$\begin{bmatrix} y_1+1 & 0 & y_2 \\ 0 & y_2+1 & 0 \\ y_2 & 0 & 0 \end{bmatrix} \succeq 0$$

Jede zulässige Lösung von (P) erfüllt $X_{11} = 0 = X_{13}$, $X_{22} = 1$. Also $p^* = -1$. Jede zulässe Lösung von (D) erfüllt $y_2 = 0$, also $d^* = 0$.

§ 2 Eigenwertoptimierung

Rayley-Ritz-Prinzip Sei $C \in S^n$ und seien $\lambda_{max}(C)$, $\lambda_{min}(C)$ der größte, bzw. kleinste Eigenwert von C. Dann ist

$$\lambda_{max}(C) = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^T C x}{x^T x} = \max_{x \in S^{n-1}} x^T C x,$$

$$\lambda_{min}(C) = \min_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^T C x}{x^T x} = \min_{x \in S^{n-1}} x^T C x, \quad \text{mit}$$
$$S^{n-1} = \{x : x^T x = 1\}$$

Das sieht man leicht mit Hilfe von Lagrange-Multiplikatoren.

Berechnung der Gradienten: $\nabla x^T C x = 2Cx$, $\nabla x^T x = 2x$.

Falls x^* den Wert $x^T C x$ unter der Nebenbedingung $x^T x = 1$ maximiert, dann gibt es ein λ mit

$$2Cx^* + \lambda 2x^* = 0,$$

d.h. x^* ist Eigenvektor von C.

<u>Satz 1</u> Sei $C \in S^n$. Dann gilt

$$\lambda_{max}(C) = \max_{\substack{X \in S^n_{\succeq 0} \\ \langle I, X\rangle = 1}} \langle C, X\rangle = \min_{y I_n - C \succeq 0} y.$$

Entsprechendes für $\lambda_{min}(C)$ durch Vertauschung von max & min.

<u>Bew.</u>: Es ist $d^* = \lambda_{max}(C)$, weil für $x \in \mathbb{R}^n \setminus \{0\}$ gilt

$$0 \leq x^T(y I_n - C)x = y\, x^T x - x^T C x \Rightarrow y \geq \frac{x^T C x}{x^T x}.$$

Das primale SDP ist strikt zulässig: $\frac{1}{n} I_n$. Es ist $p^* < \infty$, weil die Einträge von zulässigen Matrix X im Intervall $[-1, +1]$ liegen. Nach Theorem I.5.2 d) folgt $p^* = d^*$ und d^* wird angenommen.

Nachtrag zu II.1 Grundlagen

Satz 3 Sei $X \in S^n$. Die folgenden Aussagen sind äquivalent:

a) X ist positiv semidefinit.

b) Der kleinste Eigenwert von X ist ≥ 0.

c) $\exists\, L \in \mathbb{R}^{n \times k}$ mit $X = LL^T$, eine Cholesky-Zerlegung von X.

d) $\exists\, v_1, \dots, v_n \in \mathbb{R}^k$ mit $X_{ij} = v_i^T v_j$, eine Gram-Darstellung von X.

e) Alle Hauptminoren von X sind ≥ 0.

Bew.: $\rightarrow$ Lineare Algebra.

Zusatz: Falls X positiv definit, dann ist in: b) kl. EW > 0

~~b~~c) $L \in \mathbb{R}^{n \times n}$ mit Rank $L = n$, d) $v_1, \dots, v_n$ sind lin. unabh.,

e) die führenden Hauptminoren sind > 0.

Zurück zur Eigenwertoptimierung

Def. 2 Sei $C \subseteq \mathbb{R}^n$ eine konvexe Menge. Ein Punkt $x \in C$ heißt Extrempunkt, falls für alle $y, z \in C$, $\alpha \in (0,1)$ mit $x = \alpha y + (1-\alpha) z$ folgt $x = y = z$.

Notation: $\operatorname{ext}(C) = \{x : x \in C \text{ Extrempunkt von } C\}$.

Satz 3 (Minkowski; Krein-Milman)

Sei $C \subseteq \mathbb{R}^n$ eine kompakte und konvexe Menge. Dann

$$C = \operatorname{conv}(\operatorname{ext}(C)).$$

Bew.: „$\supseteq$": ✓

„$\subseteq$": per Induktion nach $\dim C$.

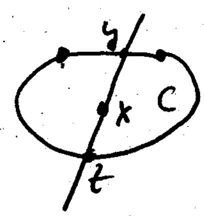
$\dim C = 0$: Dann $C = \{x\}$. ✓

$\dim C > 0$: **1. Fall** $x \in \partial C$

Dann $\exists$ nach Satz I.4.7 eine Stützhyperebene H an C durch x. Betrachte $F = H \cap C$. Dies ist konvex u. kompakt, $\dim F < \dim C$. Nach I.V. ist $x \in \operatorname{conv}(\operatorname{ext}(F))$. Da $\operatorname{ext}(F) \subseteq \operatorname{ext}(C)$, folgt die Beh.

2. Fall $x \notin \partial C$.

Dann ist der Schnitt einer Geraden durch x mit C ein Liniensegment $[y, z]$ mit $y, z \in \partial C$. Nach 1. sind $y, z \in \operatorname{conv}(\operatorname{ext}(C))$, also auch $x \in \operatorname{conv}(\operatorname{ext}(C))$.



Satz 4 (Fan)

Sei $C \in S^n$ mit Eigenwerten $\lambda_1 \geq \ldots \geq \lambda_n$. Dann ist

$$\lambda_1 + \cdots + \lambda_k = \max \{ \langle C, X \rangle : X \in S^n, \ Tr(X) = k, \ I_n \succeq X \succeq 0 \}$$

$$= \max \{ \langle C, YY^T \rangle : Y \in \mathbb{R}^{n \times k}, \ Y^T Y = I_k \}.$$

Lemma 5

Definiere

$$K_1 = \{ X \in S^n : Tr(X) = k, \ I_n \succeq X \succeq 0 \},$$

$$K_2 = \{ YY^T : Y \in \mathbb{R}^{n \times k}, \ Y^T Y = I_k \}.$$

Es ist $ext(K_1) = K_2$. Also $conv(K_2) = K_1$.

Bew.: Sei $P \in O(n)$. Dann ist $X \in K_1 \Longleftrightarrow PXP^T \in K_1$ und $X \in ext(K_1) \Longleftrightarrow PXP^T \in ext(K_1)$.
Definiere das Polytop $Q = \{ x \in [0,1]^n : e^T x = k \}$, wobei $e^T = [1, \ldots, 1]^T$. Schreibe $X \in S^n$ als $X = PDP^T$, mit $P \in O(n)$ und D Diagonalmatrix. Dann ist $X \in K_1 \Longleftrightarrow D \in Q$ und $X \in ext(K_1) \Longleftrightarrow D \in ext(Q)$. Nun ist $ext(Q) = \{ x \in \{0,1\}^k : e^T x = k \}$, also ist $X_1 \in ext(K_1) \Longleftrightarrow X$ hat genau k Eigenwerte, die gleich 1 sind und $n-k$ Eigenwerte, die gleich 0 sind, d.h. X ist eine Projektion auf einen k-dim. UVR. $\boxtimes$

Bew.: (von Satz 4)

Die beiden Maxima stimmen überein: Folgt unmittelbar aus

A 4.3: Sei $C \subseteq \mathbb{R}^n$, $C \neq \emptyset$, konvex u. kompakt, sei $c \in \mathbb{R}^n$,

dann $\max\{c^T x : x \in C\} = \max\{c^T x : x \in \operatorname{ext} C\}$.

Sei p^* das Maximum.

$\lambda_1 + \dots + \lambda_k \leq p^*$: $C = \sum_{i=1}^n \lambda_i u_i u_i^T$ Spektralzerlegung von C.

Definiere $Y = [u_1, \dots, u_k] \in \mathbb{R}^{n \times k}$. Dann ist $YY^T \in K_2$ und

$\langle C, YY^T \rangle = \lambda_1 + \dots + \lambda_k$.

$\lambda_1 + \dots + \lambda_k \geq p^*$: $C = PDP^T$, $P \in O(n)$, $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ Spektralzerlegung von C. Dann ist für $YY^T \in K_2$

$$\langle C, YY^T \rangle = \operatorname{Tr}(CYY^T) = \operatorname{Tr}(PDP^T YY^T) = \operatorname{Tr}(DP^T Y Y^T P)$$
$$= \langle D, P^T Y (P^T Y)^T \rangle,$$

wobei $P^T Y (P^T Y)^T \in K_2$. Also

$$p^* = \max\left\{ \langle D, M \rangle = \sum_{i=1}^n \lambda_i M_{ii} : M \in K_2 \right\}.$$

Nun ist für $M \in K_2$ die Diagonale $(M_{11}, \dots, M_{nn})$ im Polytop Q,

$Q = \{x \in [0,1]^n : e^T x \leq k\}$, d.h.

$$p^* \leq \max\left\{ \sum_{i=1}^n \lambda_i x_i : x \in Q \right\} = \lambda_1 + \dots + \lambda_k.$$

$\boxtimes$

§ 3 Relaxierung quadratischer Programme

Def. 1 Ein *quadratisches Programm* (QP) ist von der Form

$$\begin{aligned} \max \quad & x^T Q_0 x + b_0^T x + \alpha_0 \\ & x \in \mathbb{R}^n \\ & x^T Q_j x + b_j^T x + \alpha_j = 0 \quad \text{für } j \in [m], \end{aligned}$$

wobei $Q_0, Q_1, \dots, Q_m \in S^n$, $b_0, \dots, b_m \in \mathbb{R}^n$, $\alpha_0, \dots, \alpha_m \in \mathbb{R}$ gegeben sind.

Mit quadratischen Programmen kann man sehr viele Optimierungsprobleme modellieren, z.B. 0/1-Bedingungen:

$$x_j^2 - x_j = 0$$

oder $-1/+1$-Bedingungen

$$x_j^2 - 1 = 0.$$

Im allgemeinen sind QP keine konvexen Programme und es ist NP-schwer eine optimale Lösung zu finden.

Bsp. 2 (MAX CUT)

Sei $G = (V, E)$ ein ungerichteter Graph und $w = (w_{ij}) \in \mathbb{R}^E_{\geq 0}$ eine Gewichtsfkt.

Eine Teilmenge $S \subseteq V$ definiert einen <u>Schnitt</u> $\delta_G(S) \subseteq E$:

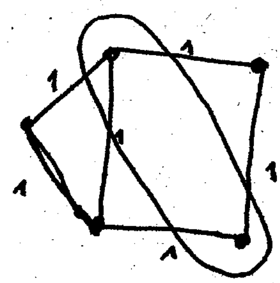
$$\delta_G(S) = \{ \{i,j\} \in E : \ |\{i,j\} \cap S| = 1 \}.$$

Das <u>Gewicht</u> des Schnittes $\delta_G(S)$ ist

$$w(\delta_G(S)) = \sum_{\{i,j\} \in \delta_G(S)} w_{ij},$$

und das <u>MAX CUT-Problem</u> fragt nach einem Schnitt mit maximalem Gewicht

$$mc(G,w) = \max_{S \subseteq V} w(\delta_G(S)).$$



$mc(G,w) = 4$

<u>Formulierung von MAX CUT als QP:</u>

$$\max \frac{1}{2} \sum_{\{i,j\} \in E} w_{ij}(1 - x_i x_j) = mc(G,w)$$

$$x \in \mathbb{R}^V$$

$$x_i^2 - 1 = 0 \quad \text{für } i \in V.$$

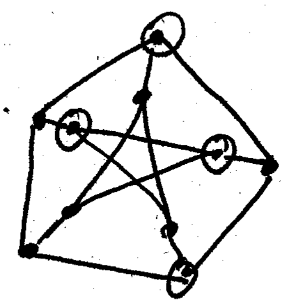
Wir werden später sehen, dass MAX CUT NP-schwer ist, also auch QP.

<u>Bsp. 3</u> (Unabhängigkeitszahl)

$G = (V,E)$. $S \subseteq V$ heißt <u>unabhängig</u>, falls für alle $i,j \in S$

stets $\{i,j\} \notin E$ gilt. Die <u>Unabhängigkeitszahl</u> von G ist definiert als

$$\alpha(G) = \max\{|S| : S \subseteq V \text{ unabhängig}\}.$$



$\alpha(G) = 4$

Berechnung von α ist auch NP-schwer.

<u>Formulierungen von α als QP.</u>

$$\begin{aligned}
\alpha(G) = \max \ & \sum_{i \in V} x_i \\
& x \in \mathbb{R}^V, \ s \in \mathbb{R}^E \\
& x_i^2 - x_i = 0 \qquad \text{für } i \in V \\
& s_{ij}^2 + x_i + x_j - 1 = 0 \quad \text{für } \{i,j\} \in E.
\end{aligned}$$

$$\begin{aligned}
= \max \ & \sum_{i \in V} x_i^2 \\
& x \in \mathbb{R}^V \\
& x_i^2 - x_i = 0 \qquad \text{für } i \in V \\
& x_i x_j = 0 \qquad \text{für } \{i,j\} \in E.
\end{aligned}$$

<u>Satz 4</u> (SDP-Relaxierung von QP).
Es ist $qp \leq sdp$, wobei

$$\begin{aligned}
qp = \max \ & x^T Q_0 x + b_0^T x + \alpha_0 \\
& x \in \mathbb{R}^n \\
& x^T Q_j x + b_j^T x + \alpha_j = 0 \qquad \text{für } j \in [m]
\end{aligned}$$

und

$$\mathrm{sdp} = \inf \; y_0$$

$$y_0, \ldots, y_m \in \mathbb{R}$$

$$y_0 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \sum_{j=1}^{m} y_j \begin{bmatrix} \alpha_j & \frac{1}{2} b_j^T \\ \frac{1}{2} b_j & Q_j \end{bmatrix} - \begin{bmatrix} \alpha_0 & \frac{1}{2} b_0^T \\ \frac{1}{2} b_0 & Q_0 \end{bmatrix} \in S_{\geq 0}^{n+1}.$$

Bew.: Es gilt

$$x^T Q_j x + b_j^T x + \alpha_j = \left\langle \begin{bmatrix} \alpha_j & \frac{1}{2} b_j^T \\ \frac{1}{2} b_j & Q_j \end{bmatrix}, \begin{bmatrix} 1 & x^T \\ x & x x^T \end{bmatrix} \right\rangle$$

und $\begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix} \in S_{\geq 0}^{n+1}$, da $\begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} = \begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix}$.

Also

$$\mathrm{qp} \leq \sup \left\langle \begin{bmatrix} \alpha_0 & \frac{1}{2} b_0^T \\ \frac{1}{2} b_0 & Q_0 \end{bmatrix}, X \right\rangle$$

$$X \in S_{\geq 0}^{n+1}$$

$$\left\langle \begin{bmatrix} \alpha_j & \frac{1}{2} b_j^T \\ \frac{1}{2} b_j & Q_j \end{bmatrix}, X \right\rangle = 0 \quad \text{für } j \in [m].$$

Dualisieren liefert nun die Beh. (zusammen mit schwacher Dualität). $\boxtimes$

Bem. • Falls man im Maximierungsproblem zusätzlich fordert, dass $\operatorname{rank} X = 1$ ist, hätte man $\mathrm{qp} = \sup$. Man "relaxiert" diese Bedingung in Satz 4.

• Vorteile von sdp: – effizient berechenbar
– einfache Methode, um obere Schranken für qp zu finden.

Zur Komplexität von MAX CUT

Def. 5 MAX CUT als Entscheidungsproblem

Gegeben: $G=(V,E)$ Graph, $w \in \mathbb{Z}^E_{\geq 0}$, $k \in \mathbb{N}$.

Entscheide: Gibt es einen Schnitt $\delta_G(S)$ mit $w(\delta_G(S)) \geq k$.

Def. 6 PARTITION

Gegeben: $a_1, \dots, a_n \in \mathbb{N}$

Entscheide: Gibt es ein $S \subseteq [n]$ mit

$$\sum_{i \in S} a_i = \sum_{i \notin S} a_i.$$

Es ist bekannt (siehe Klassiker Garey-Johnson), dass PARTITION NP-vollständig ist.

Satz 7 MAX CUT ist NP-vollständig.

Bew.: Klar: MAX CUT $\in$ NP. [Angabe von $S \subseteq V$ genügt, um in polynomieller Zeit zu verifizieren, dass $\delta_G(S) \geq k$].

Zu zeigen: ~~MAX CUT $\leq_p$ PARTITION, d.h.~~

PARTITION $\leq_p$ MAXCUT, d.h. falls MAXCUT $\in$ P, dann auch PARTITION $\in$ P. Gegeben seien $a_1, \dots, a_n \in \mathbb{N}$.

Betrachte $G = K_n$ den vollständigen Graph auf n Knoten mit $w_{ij} = a_i a_j$, $\{i,j\} \in E$.

Definiere $\sigma = \sum_{i=1}^{n} a_i$ und $k = \sigma^2/4$. Für $S \subseteq [n]$ definiere $a(S) = \sum_{i \in S} a_i$. Dann gilt

$$w(\delta_G(S)) = \sum_{i \in S, j \notin S} w_{ij} = \sum_{i \in S, j \notin S} a_i a_j = \left(\sum_{i \in S} a_i\right)\left(\sum_{j \notin S} a_j\right)$$

$$= a(S)\,(\sigma - a(S)) \leq \sigma^2/4 = k$$

mit Gleichheit gdw. $a(S) = \frac{\sigma}{2}$, d.h. $a(S) = a([n] \setminus S)$.
Es gibt also einen Schnitt mit Gewicht $\geq k$ gdw. $a_1, \dots, a_n$ partitioniert werden kann. $\boxtimes$

Approximationen von MAXCUT:

Wdh:

- MAXCUT: Sei $G=(V,E)$ ein Graph mit Gewichten $w \in \mathbb{Z}_+^E$.
  Für festes $k \in \mathbb{N}$, entscheide ob es einen Schnitt mit Gewicht $\geq k$ gibt.

- Formulierung als quadratisches Programm:

$$\boxed{mc(G,w)} = \max \frac{1}{2} \sum_{\{i,j\} \in E} w_{ij}(1 - x_i x_j)$$

$$x \in \mathbb{R}^V \qquad (n := |V|)$$

$$x_i^2 = 1 \qquad \forall i \in [n]$$

- $X := xx^T \Rightarrow X_{ij} = x_i x_j$ und $X \in S^n_{\succeq 0}$.

$$\Rightarrow \quad \boxed{sdp(G,w)} = \max \frac{1}{2} \sum_{\{i,j\} \in E} w_{ij}(1 - X_{ij})$$

$$X_{ii} = 1$$

$$X \in S^n_{\succeq 0} \qquad \text{(relaxiere rank(X}$$

$$\Rightarrow \quad mc(G,w) \leq sdp(G,w)$$

Wieviel verlieren wir durch die Relaxierung?

Satz 8 (Goemans, Williamson - 1995)

Sei $G=(V,E)$ Graph mit nicht-negativen Gewichten $w \in \mathbb{R}_+^{|E|}$, dann gilt:

$$\operatorname{sdp}(G,w) \geq mc(G,w) \geq 0.878\ldots \cdot \operatorname{sdp}(G,w).$$

Der Beweis liefert einen randomisierten Algorithmus mit Approximationsfaktor $\geq 0.878\ldots$:

1. Löse $\operatorname{sdp}(G,w)$. Sei X eine optimale Lösung.
2. Finde eine Cholesky-Zerlegung von X mit Einheitsvektoren $v_i \in \mathbb{R}^n$, $i \in [n]$, so dass
$$X_{ij} = v_i^T v_j \quad , \quad i,j \in [n].$$
3. Wähle $r \in \mathbb{R}^n$, $\|r\|=1$, gemäß einer Zufallsverteilung auf der Einheitssphäre, die invariant unter Rotationen ist.
4. Definiere einen Schnitt $\delta_G(S)$ durch
$$i \in S \iff \operatorname{sign}(v_i^T r) \geq 0 \qquad \forall\, i \in [n]$$
(also $x_i := \operatorname{sign}(v_i^T r)$)
5. Teste ob $w(\delta_G(S)) \geq 0.878\ldots \cdot \operatorname{sdp}(G,w)$.
Falls nicht, gehe zurück zu Schritt 3.

Schritte 3. und 4. werden auch als Methode des randomisierten Rundens (randomized rounding procedure) bezeichnet.

Werden zeigen: $\mathbb{E}[w(\delta_G(S))] \geq 0.878\ldots \cdot sdp(G,w)$

$\Rightarrow$ Keine Endlos-Schleife.

Allgemeiner: Mit hoher W'keit finden wir schnell einen ausreichend großen Schnitt.

<u>Lemma 9</u> (Grothendieck-Identität)

Seien $u,v \in \mathbb{R}^d$ Einheitsvektoren, und sei $r \in \mathbb{R}^d$, $\|r\|=1$, gemäß einer Zufallsverteilung gewählt, die invariant unter Rotationen ist.

Dann gilt:

1. $\mathbb{P}[\operatorname{sign}(u^Tr) \neq \operatorname{sign}(v^Tr)] = \dfrac{\arccos(u^Tv)}{\pi}$.
2. $\mathbb{E}[\operatorname{sign}(u^Tr)\operatorname{sign}(v^Tr)] = \frac{2}{\pi}\arcsin(u^Tv)$.

<u>Beweis:</u>

1. Wähle r so, dass u,v und r auf einer Ebene liegen (Eig. der ZV).

   $\Rightarrow$ r kann mit der Gleichverteilung auf dem Kreis definiert werden.

   $\Rightarrow$ W'keit für $\operatorname{sign}(u^Tr) \neq \operatorname{sign}(v^Tr)$ hängt nur vom Winkel zwischen u und v ab:

gleiches Vorzeichen


$$\Rightarrow \mathbb{P}[\mathrm{sign}(u^T r) \neq \mathrm{sign}(v^T r)] = 2 \cdot \frac{\arccos(u^T v)}{2\pi} = \frac{\arccos(u^T v)}{\pi}$$

2. $$\mathbb{E}[\mathrm{sign}(u^T r)\,\mathrm{sign}(v^T r)] = (+1)\,\mathbb{P}[\mathrm{sign}(u^T r) = \mathrm{sign}(v^T r)] + (-1)\,\mathbb{P}[\mathrm{sign}(u^T r) \neq \mathrm{sign}(v^T r)]$$

$$= 1 - 2\,\mathbb{P}[\mathrm{sign}(u^T r) \neq \mathrm{sign}(v^T r)] = 1 - 2\,\frac{\arccos(u^T v)}{\pi}$$

$\arcsin t + \arccos t = \frac{\pi}{2}$

$$= \frac{\pi - 2\left(\frac{\pi}{2} - \arcsin(u^T v)\right)}{\pi} = \frac{2}{\pi} \arcsin(u^T v).$$

$\square$

Woher kommt die Konstante in Satz 8?

<u>Lemma 10</u>: Für alle $t \in [-1, 1)$ gilt

$$\frac{2}{\pi} \, \frac{\arccos t}{1 - t} \geq 0{,}878\ldots$$

"Sieht man" am Graph der Funktion: Konvex auf dem Intervall, und minimal ungefähr bei $-0{,}69$

<u>Beweis von Satz 8</u>: Sei X eine optimale Lösung von $\mathrm{sdp}(G, w)$ und $v_1, \ldots, v_n$ Einheitsvektoren mit $X_{ij} = v_i^T v_j$.

Sei $\delta_G(S)$ der Schnitt, der in den Schritten 3. und 4. konstruiert wurde.

Dann gilt:

$$\mathbb{E}[w(\delta_G(S))] = \sum_{\{i,j\}\in E} w_{ij}\, \mathbb{P}[\{i,j\} \text{ ist im Schnitt}]$$

$$= \sum_{\{i,j\}\in E} w_{ij}\, \mathbb{P}[\operatorname{sign}(v_i^T r) \neq \operatorname{sign}(v_j^T r)]$$

$$= \sum_{\{i,j\}\in E} w_{ij} \frac{\arccos(v_i^T v_j)}{\pi}$$

$$= \sum_{\{i,j\}\in E} w_{ij} \left(\frac{1 - v_i^T v_j}{2}\right) \cdot \left(\frac{2}{\pi} \frac{\arccos(v_i^T v_j)}{1 - v_i^T v_j}\right)$$

$v_i^T v_j = 1 \Rightarrow \arccos(v_i^T v_j) = 0$ $\quad v_i v_j \neq 1$

$$\geq \sum_{\{i,j\}\in E} w_{ij} \left(\frac{1 - v_i^T v_j}{2}\right) \cdot 0{,}878\ldots$$

$$= 0{,}878\ldots \cdot \operatorname{sdp}(G,w).$$

Da $mc(G,w) \geq \mathbb{E}[w(\delta_G(S))]$, sind wir fertig. $\square$

<u>Bemerkungen:</u>

- Eine unter Rotationen invariante Verteilung kann wie folgt definiert werden:

  $N(0,1)$ mit Dichte $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

  wähle $x_1, \ldots, x_n$ unabh. nach $N(0,1)$

  $\Rightarrow r = \frac{1}{\sqrt{x_1^2 + \ldots + x_n^2}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in S^{n-1}$ ist invariant unter Rotationen gewählt.

- Der Algorithmus kann modifiziert werden, so dass er deterministisch ist und den gleichen Approximationsfaktor liefert.

Nesterovs Approximationsalgorithmus

Sei $G=(V,E)$ Graph mit Gewichten w.

Wir definieren die Laplace-Matrix zu G und w durch

$$(L_w)_{ij} = \begin{cases} \sum\limits_{k:\{i,k\}\in E} w_{ik} & \text{falls } i=j \\ -w_{ij} & \text{falls } \{i,j\}\in E \\ 0 & \text{sonst}. \end{cases}$$

Dann ist $L_w \in S^n$.

Lemma 11:

(a) Für $x \in \{\pm 1\}^n$ gilt

$$\frac{1}{4} x^T L_w x = \frac{1}{2} \sum_{\{i,j\}\in E} w_{ij}(1-x_i x_j)$$

(b) Falls $w \geq 0$, dann ist $L_w \succeq 0$

Beweis: (a): Ausrechnen;
(b) folgt aus (a). $\square$

Dann gilt:

$$mc(G,w) = \max_{\substack{x\in\mathbb{R}^n \\ x_i^2=1}} \frac{1}{4} x^T L_w x$$

und

$$sdp(G,w) = \max_{\substack{X_{ii}=1 \\ X\in S^n_{\succeq 0}}} \frac{1}{4} \langle L_w, X\rangle$$

Für $w \geq 0$ taucht in beiden Formulierungen die pos. semidefinite Matrix $\frac{1}{4} L_w$ auf.

$\rightsquigarrow$ Verallgemeinerung:

Für $A \in S^n$ definieren wir

$$\boxed{qp(A)} = \max \sum_{i,j=1}^{n} A_{ij} x_i x_j$$
$$x \in \mathbb{R}^n$$
$$x_i^2 = 1 \quad \forall i \in [n]$$

und

$$\boxed{sdp(A)} = \max \langle A, X \rangle$$
$$X_{ii} = 1$$
$$X \in S^n_{\succeq 0}$$

$\Rightarrow \quad qp(\frac{1}{4} L_w) = mc(G, w)$.

Wie zuvor: $\quad qp(A) \leq sdp(A) \quad$ (Relaxiere Rang von X)

Bekommen wir immer noch eine untere Schranke mit konstantem Faktor?

Wieder: Randomisiertes Runden.

Lemma 12: Sei $X \in S^n$ und definiere $\tilde{X}$ durch

$$\tilde{X}_{ij} = \arcsin X_{ij} - X_{ij}$$

Dann gilt:

$$X \succeq 0 \Rightarrow \tilde{X} \succeq 0$$

Beweis: Falls $X \succeq 0$, dann für $k \in \mathbb{Z}_+$ auch die Matrix mit den Einträgen X_{ij}^k (Aufgabe)

$$\arcsin t = t + \left(\frac{1}{2}\right)\frac{t^3}{3} + \left(\frac{1\cdot 3}{2\cdot 4}\right)\frac{t^5}{5} + \ldots$$

Alle Koeffizienten sind nicht-negativ

$\Rightarrow$ Limes-bildung liefert die Aussage $\square$

Satz 13: Für $A \in S^n_{\geq 0}$ gilt

$$\mathrm{sdp}(A) \geq \mathrm{qp}(A) \geq \frac{2}{\pi}\,\mathrm{sdp}(A)$$

$\left(\frac{2}{\pi} = 0{,}636\ldots\right)$

Beweis: Sei X eine optimale Lösung von $\mathrm{sdp}(A)$ und $v_1, \ldots, v_n$ Einheitsvektoren mit $X_{ij} = v_i^T v_j$. Wähle wie zuvor $r \in \mathbb{R}^n$, $\|r\| = 1$ und setze $x_i = \mathrm{sign}(v_i^T r)$

$$\Rightarrow \mathbb{E}\left[\sum_{i,j=1}^{n} A_{ij} x_i x_j\right] = \left\langle A, \left(\mathbb{E}[x_i x_j]\right)_{i,j=1}^{n} \right\rangle$$

Lemma 9(2)

$$\overset{\text{Lemma 9(2)}}{=} \frac{2}{\pi}\langle A, (\arcsin(v_i^T v_j))_{i,j=1}^n \rangle$$

$$= \frac{2}{\pi}\langle A, (\arcsin X_{ij})_{i,j=1}^n \rangle$$

$$= \frac{2}{\pi}\langle A, X\rangle + \underbrace{\frac{2}{\pi}\langle A, (\arcsin X_{ij} - X_{ij})_{i,j=1}^n \rangle}_{\geq 0}$$

$$\geq \frac{2}{\pi}\langle A, X\rangle = \frac{2}{\pi}\,\mathrm{sdp}(A). \qquad \square$$

II.4 Grothendieck-Ungleichung

(Original-Kontext von Lemma 9)

Für $A \in \mathbb{R}^{m\times n}$ definiere das quadratische Programm

$$\|A\|_{\infty\to 1} := \max \sum_{i=1}^m \sum_{j=1}^n A_{ij} x_i y_j$$
$$x \in \mathbb{R}^m,\ y \in \mathbb{R}^n$$
$$x_i^2 = 1 \quad \forall i \in [m]$$
$$y_j^2 = 1 \quad \forall j \in [n]$$

(Notation wegen Operatornorm für Operatoren $A: \ell_\infty^n \to \ell_1^m$)

Dann ist die semidefinite Relaxierung gegeben durch

$$\mathrm{sdp}_{\infty\to 1}(A) := \max \sum_{i=1}^m \sum_{j=1}^n A_{ij} u_i^T v_j$$
$$u_i \in \mathbb{R}^{m+n},\ v_j \in \mathbb{R}^{m+n}$$
$$\|u_i\| = 1 \quad \forall i \in [m]$$
$$\|v_j\| = 1 \quad \forall j \in [n]$$

<u>Satz 1</u> (Grothendieck-Ungleichung)

Es gibt eine Konstante K, so dass für alle $A \in \mathbb{R}^{m \times n}$ gilt:

$$\|A\|_{\infty \to 1} \leq \mathrm{Sdp}_{\infty \to 1}(A) \leq K \cdot \|A\|_{\infty \to 1} .$$

Eine fundamentale Ungleichung in der Theorie der Banach-Räume, und seit einigen Jahren auch immer ~~wichtiger~~ ausserhalb der Funktionalanalysis.

Bekannt: Das kleinstmögliche K (Grothendieck-Konstante K_G) ist zwischen $1,676\ldots$ und $1,782\ldots$

Werden hier die obere Schranke

$$K_G \leq \frac{\pi}{2\ln(1+\sqrt{2})} = 1,782\ldots$$

zeigen.

§ 4 Grothendieck-Ungleichung

$A \in \mathbb{R}^{m \times n}$

$$\|A\|_{\infty \to 1} = \max \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} x_i y_j$$
$$x \in \mathbb{R}^m, \; y \in \mathbb{R}^n$$
$$x_i^2 = 1 \qquad i \in [m]$$
$$y_j^2 = 1 \qquad j \in [n]$$

ist ein quadratisches Programm.

[Aufgabe 6.3 : $\|A\|_{\infty \to 1} = mc(G, w)$ für passendes G, w
$\rightarrow$ Berechnung von $\|A\|_{\infty \to 1}$ ist NP-schwer].

Semidefinite Relaxierung :

$$sdp_{\infty \to 1}(A) = \max \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} u_i^T v_j$$
$$u_i \in \mathbb{R}^{m+n}, \; v_j \in \mathbb{R}^{m+n}$$
$$\|u_i\| = 1 \qquad i \in [m]$$
$$\|v_j\| = 1 \qquad j \in [n]$$

Satz 1 (Grothendieck-Ungleichung)

$\exists$ Konstante K : $\forall A \in \mathbb{R}^{m \times n}$:

$$\|A\|_{\infty \to 1} \le sdp_{\infty \to 1}(A) \le K \, \|A\|_{\infty \to 1}$$

Kleinstmögliches K heißt <u>Grothendieck Konstante</u> K_G. Man weiß z. Zt. nur, dass $K_G \in (1{,}676\ldots, 1{,}782\ldots)$.

Wir zeigen $K_G \leq \frac{\pi}{2\ln(1+\sqrt{2})} = 1{,}782\ldots$

<u>Bew.</u> (nach Krivine 1979)

Wieder: Approximationsalgo. mit randomisiertem Runden.

1. Löse $sdp_{\infty\to 1}(A)$. Seien $u_1, \ldots, u_m, v_1, \ldots, v_n \in S^{m+n-1}$ optimale Einheitsvektoren.
2. Krivines Trick (Lemma 2): Verwende diese Vektoren, um neue Einheitsvektoren $u_1', \ldots, u_m', v_1', \ldots, v_n' \in S^{m+n-1}$ zu konstruieren.
3. Wähle $r \in S^{m+n-1}$ zufällig
4. Runden: $x_i = \text{sign}(r^T u_i') \quad i \in [m]$

   $y_j = \text{sign}(r^T v_j') \quad j \in [n]$.

Erwartete Qualität der Lösung:

$$\|A\|_{\infty\to 1} \geq \mathbb{E}\Big[\sum_{i=1}^m \sum_{j=1}^n A_{ij} x_i y_j\Big]$$

$$= \sum_{i=1}^m \sum_{j=1}^n A_{ij}\, \mathbb{E}\big[\text{sign}(r^T u_i')\, \text{sign}(r^T v_j')\big]$$

$$\overset{\text{Lemma 2}}{=} \sum_{i=1}^m \sum_{j=1}^n A_{ij}\, \beta\, u_i^T v_j = \beta\, sdp_{\infty\to 1}(A)$$

wobei $\beta = \dfrac{2 \ln(1+\sqrt{2})}{\pi}$, also $K_G \leq \beta^{-1}$.

<u>Lemma 2</u> (Krivines Trick)

Seien $u_1, \dots, u_m, v_1, \dots, v_n \in S^{m+n-1}$. Dann gibt es $u_1', \dots, u_m'$, $v_1', \dots, v_n' \in S^{m+n-1}$, so dass

$$\mathbb{E}\left[\operatorname{sign}(r^T u_i') \operatorname{sign}(r^T v_j') \right] = \beta\, u_i^T v_j,$$

wobei $\beta = \frac{2}{\pi} \ln(1+\sqrt{2}) = 0{,}561\dots$

<u>Kurzer Exkurs</u>: Tensorprodukte

$\mathbb{R}^n$ ist n-dim. euklidischer Vektorraum mit Skalarprodukt $x^T y$ und Orthonormalbasis $e_1, \dots, e_n$.

Das <u>k-fache Tensorprodukt</u> $(\mathbb{R}^n)^{\otimes k}$ ist n^k-dim. eukl. VR mit ONB $e_{i_1} \otimes \dots \otimes e_{i_k}$, $i_j \in \{1, \dots, n\}$ mit Skalarprodukt

$$(e_{i_1} \otimes \dots \otimes e_{i_k})^T (e_{j_1} \otimes \dots \otimes e_{j_k}) = \prod_{r=1}^{k} e_{i_r}^T e_{j_r}.$$

Sei $v \in \mathbb{R}^n$ durch $v = v_1 e_1 + v_2 e_2 + \dots + v_n e_n$ gegeben.

Dann $v^{\otimes k} \in (\mathbb{R}^n)^{\otimes k}$ mit

$$v^{\otimes k} = (v_1 e_1 + \dots + v_n e_n) \otimes \dots \otimes (v_1 e_1 + \dots + v_n e_n)$$

$$\overset{\substack{\text{distributiv} \\ \text{ausrechnen}}}{=} \sum_{i_1 \dots i_k} v_{i_1} \dots v_{i_k}\, e_{i_1} \otimes \dots \otimes e_{i_k}.$$

Für $v, w \in \mathbb{R}^n$ gilt dann

$$\left(v^{\otimes k}\right)^T w^{\otimes k} = \left(v^T w\right)^k .$$

<u>Bew.</u> (Lemma 2)

Definiere die Fkt. $E : [-1,+1] \to [-1,+1]$ durch $E(t) = \frac{2}{\pi} \arcsin t$. Nach der Identität von Grothendieck (Lemma 3.9) ist

$$E\left((u_i')^T v_j' \right) = \mathbb{E}\left[\operatorname{sign}(r^T u_i') \operatorname{sign}(r^T v_j') \right] \overset{!}{=} \beta\, u_i^T v_j .$$

<u>Idee</u>: Um u_i', v_j' und β zu finden, invertiere die Fkt E:

$$(u_i')^T v_j = E^{-1}\left(\beta\, u_i^T v_j \right)$$

und benutze dann die Reihenentwicklung

$$E^{-1}(t) = \sum_{r=0}^{\infty} g_{2r+1}\, t^{2r+1} ,$$

die für alle $t \in [-1,+1]$ gilt. [Da E eine ungerade Fkt. ist, $E(-t) = -E(t)$, ist E^{-1} ebenfalls ungerade].

Dazu definiere den unendlich-dimensionalen Hilbertraum

$$H = \bigoplus_{r=0}^{\infty} \left(\mathbb{R}^{m+n}\right)^{\otimes 2r+1} .$$

Definiere $\mu_i', v_j' \in H$ komponentenweise

$$(\mu_i')_r = \operatorname{sign}(g_{2r+1}) \sqrt{|g_{2r+1}| \beta^{2r+1}} \; \mu_i^{\otimes 2r+1}$$

$$(v_j')_r = \sqrt{|g_{2r+1}| \beta^{2r+1}} \; v_j^{\otimes 2r+1}$$

Dann

$$(\mu_i')^T v_j' = \sum_{r=0}^{\infty} g_{2r+1} \beta^{2r+1} (\mu_i^T v_j)^{2r+1} = E^{-1}(\beta \mu_i^T v_j)$$

und

$$1 = (\mu_i')^T \mu_i' = (v_j')^T v_j' = \sum_{r=0}^{\infty} |g_{2r+1}| \beta^{2r+1},$$

was β definiert.

Bestimmung von β:

$$E(t) = \frac{2}{\pi} \arcsin t$$

$$E^{-1}(t) = \sin\left(\frac{\pi}{2} t\right) = \sum_{r=0}^{\infty} \underbrace{\frac{(-1)^{2r+1}}{(2r+1)!} \left(\frac{\pi}{2}\right)^{2r+1}}_{g_{2r+1}} t^{2r+1}$$

Also

$$1 = \sum_{r=0}^{\infty} \left| \frac{(-1)^{2r+1}}{(2r+1)!} \left(\frac{\pi}{2}\right)^{2r+1} \right| \beta^{2r+1} = \sinh\left(\frac{\pi}{2} \beta\right)$$

D.h.

$$\beta = \frac{2}{\pi} \operatorname{arsinh} 1 = \frac{2}{\pi} \ln(1 + \sqrt{2}), \quad \text{weil} \quad \operatorname{arsinh} t = \ln\left(t + \sqrt{t^2+1}\right).$$

☒

Verwendung der Grothendieck-Ungleichung in der statistischen Physik

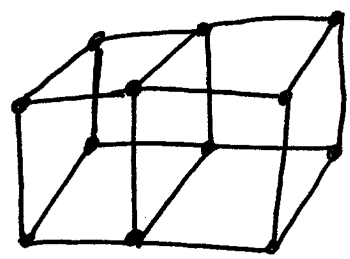
Ising-Modell / Spingläser zur Beschreibung von Ferromagnetismus

$$V = [0,N]^d \cap \mathbb{Z}^d \quad \text{mit } N \in \mathbb{N} \quad \underline{d\text{-dim. Raster}}$$

bestehend aus $(N+1)^d$ Knoten.

Knoten $i \in V$ entsprechen Atomen.

Zwei Knoten $i, j \in V$ interagieren, falls sie im Raster benachbart sind, d.h. falls $\|i-j\| = 1$ gilt.



$J_{ij} \in \mathbb{R}$ Koppelungskonstante, $J_{ij} > 0$: ferromagnetische Kop.

$J_{ij} < 0$: antiferromagnetische K.

Jeder Knoten hat entweder Zustand $+1$ ("spin up") oder Zustand -1 ("spin down").

$x_i^2 = 1$, x_i Variable für gerade Knoten $i_1 + \dots + i_d = 0 \pmod 2$

$y_j^2 = 1$, y_j Variable für ungerade Knoten $j_1 + \dots + j_d = 1 \pmod 2$

Energie des Systems wird durch Hamiltonfkt. beschrieben

$$H = -\frac{1}{2} \sum_{\substack{i,j \in V \\ \|i-j\|=1 \\ i \text{ gerade},\, j \text{ ungerade}}} J_{ij}\, x_i y_j$$

<u>Ziel</u>: Finde Grundzustand, d.h. Spins, die H minimieren.

Dies ist äquivalent zur Berechnung von $\|J\|_{\infty\to 1}$ und das ist NP-schwer für $d \geq 3$ (Barahona 1982).
$\mathrm{sdp}_{\infty\to 1}(J)$ gibt eine $0{,}561\ldots$ Approximation.

§ 5 Approximation von MAX-2-SAT

Seien $a = (a_{ij})$, $b = (b_{ij}) \geq 0$ gegeben. Definiere quadratische Programm

$$\mathrm{qp}(a,b) = \max \sum_{\{i,j\}\in E_1} a_{ij}(1 - x_i x_j) + \sum_{\{i,j\}\in E_2} b_{ij}(1 + x_i x_j)$$
$$x_i^2 = 1 \quad \text{für } i \in [n]$$

und die semidefinite Relaxierung

$$\mathrm{sdp}(a,b) = \max \sum_{\{i,j\}\in E_1} a_{ij}(1 - X_{ij}) + \sum_{\{i,j\}\in E_2} (1 + X_{ij})$$
$$X \in S^m_{\geq 0}$$
$$X_{ii} = 1 \quad \text{für } i \in [n].$$

<u>Satz 1</u> $\quad \mathrm{sdp}(a,b) \geq \mathrm{qp}(a,b) \geq 0{,}878\ldots\ \mathrm{sdp}(a,b)$

Bew.: Aufgabe 6.4, Hinweis: $\arccos(-t) = \pi - \arccos(t)$.

Def. 2 (SAT = satisfiability problem) Erfüllbarkeitsproblem

gegeben: Aussagenlogische Formel in konjunktiver Normalform, d.h. eine Konjunktion von Klauseln $C_1, \dots, C_m$

$$F = C_1 \wedge C_2 \wedge \dots \wedge C_m,$$

wobei jede Klausel aus Disjunktion von Literalen besteht

$$C_i = (l_{j_1} \vee \dots \vee l_{j_i})$$

und ein Literal ist eine Variable $z_1, \dots, z_n$ oder deren Negation $\neg z_1, \dots, \neg z_n$.

gesucht: Belegung der Variablen mit "wahr" oder "falsch", so dass F wahr ist.

Bsp. $F = (z_1 \vee z_2 \vee z_3) \wedge (\neg z_1 \vee z_4)$

Erfüllende Belegung z.B. $z_1 = w$, $z_2 = f$, $z_3 = f$, $z_4 = w$

Theorem (Cook-Levine ~ 1970)

Das Entscheidungsproblem SAT ist NP-vollständig.

Bew.: $\to$ Informatik

Optimierungsvariante MAX-2-SAT

Def. 3 (MAX-2-SAT)

geg.: $F = C_1 \wedge \ldots \wedge C_m$

jedes C_i besteht aus Disjunktion von ≤ 2 Literalen

$w_1, \ldots, w_m \geq 0$ Gewichte

ges.: Belegung der Variablen, so dass Gewicht der erfüllten Klauseln maximiert wird.

Theorem (Håstad 2001) Falls $P \neq NP$, dann $\nexists$ polynomialzeit-Algorithmus, der MAX-2-SAT bis auf einen Faktor von $21/22 = 0{,}954\ldots$ berechnet.

Satz 4 Man kann MAX-2-SAT in polynomieller Zeit mit einem Faktor von $0{,}878\ldots$ berechnen.

[sogar $0{,}943\ldots$ möglich].

Bew.: Modellierung von MAX-2-SAT als $qp(a,b)$ und Anwendung von Satz 1.

Für jede Variable z_i der Formel F benutze Variable x_i mit $x_i^2 = 1$. Extravariable x_0 mit $x_0^2 = 1$ mit folgender Interpretation z_i ist wahr $\Leftrightarrow x_i = x_0$.

Sei C eine Klausel, dann ist der Wert $v(C)$ definiert als

$$v(C) = \begin{cases} 1, & \text{falls } C \text{ erfüllt} \\ 0, & \text{sonst.} \end{cases}$$

Also

$$v(z_i) = \frac{1 + x_0 x_i}{2}, \qquad v(\neg z_i) = 1 - v(z_i) = \frac{1 - x_0 x_i}{2}$$

$$v(z_i \vee z_j) = 1 - v(\neg z_i \wedge \neg z_j) = \cancel{1 - v(\neg z_i) v(\neg z_j)}\ 1 - v(\neg z_i) v(\neg z_j)$$

$$= 1 - \frac{1 - x_0 x_i}{2} \cdot \frac{1 - x_0 x_j}{2}$$

$$= 1 - \frac{1}{4}\left(1 - x_0 x_j - x_0 x_i + x_0^2 x_i x_j\right)$$

$$= \frac{1 + x_0 x_i}{4} + \frac{1 + x_0 x_j}{4} + \frac{1 - x_i x_j}{4}$$

genauso

$v(z_i \vee \neg z_j)$ und $v(\neg z_i \vee \neg z_j)$

D.h. MAX-2-SAT ist äquivalent zu

$$\max \left\{ \sum_{j=1}^{m} w_j v(C_j) \;:\; x_1^2 = \ldots = x_n^2 = x_0^2 = 1 \right\}$$

und das kann in der Form $qp(a, b)$ geschrieben werden. $\boxtimes$

Kapitel III Packungen und Färbungen in Graphen

§ 1 Definitionen

Sei $G = (V, E)$ ein ungerichteter Graph.

$\overline{G} = (V, \overline{E})$ ist der zu G komplementäre Graph,

wobei $\overline{E} = \{ \{i,j\} \in \binom{V}{2} : \{i,j\} \notin E \}$.

$S \subseteq V$ heißt unabhängig, falls für alle $i, j \in S$ stets $\{i,j\} \notin E$ gilt.

$C \subseteq V$ heißt eine Clique, falls für alle $i, j \in C$ stets $\{i,j\} \in E$ gilt.

$\alpha(G) = \max \{ |S| : S \subseteq V \text{ unabhängig} \}$ ist die Unabhängigkeitszahl von G (oder auch Stabilitätszahl)

$\omega(G) = \alpha(\overline{G})$ ist die Cliquenzahl von G

$= \max \{ |C| : C \subseteq V \text{ Clique} \}$.

Sei $k \in \mathbb{N}$. Eine <u>k-Färbung</u> von G ist eine Abbildung $\varphi : \{1, \dots, k\} \to V$ mit $\varphi(i) \neq \varphi(j)$ für alle $\{i,j\} \in E$. I.a.W. eine k-Färbung ist eine Partition von $V = S_1 \cup \dots \cup S_k$ in k unabh. Mengen $S_1, \dots, S_k$. Die <u>chromatische Zahl</u> von G ist

$$\chi(G) = \min\{k : \exists\ k\text{-Färbung von } G\}$$

Die Entscheidungsprobleme

- $\alpha(G) \geq k$?
- $\omega(G) \geq k$?
- $\chi(G) \leq k$?

sind alle NP-vollständig.

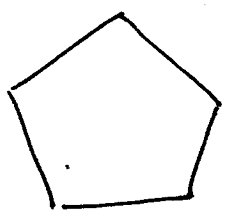
Es ist sogar „$\chi(G) \leq 3$?" für planare Graphen NP-vollständig.

[Dagegen ist jeder planare Graph 4-färbbar ($\to$ 4-Farben Satz von Appel und Haken (1976)]

Außerdem „$\chi(G) \leq 2$?" effizient entscheidbar. Dies ist die Frage nach Bipartitheit.

Klarerweise gilt $\omega(G) \leq \chi(G)$.

Im allgemeinen $\omega(G) < \chi(G)$, zum Bsp.: Kreisgraph C_{2n+1} ungerader Länge:



C_5

$$\omega(C_{2n+1}) = 2$$
$$\chi(C_{2n+1}) = 3.$$

Komplementärgraph $\overline{C_{2n+1}}$: $\omega(\overline{C_{2n+1}}) = n$

$$\chi(\overline{C_{2n+1}}) = n+1.$$

§ 2 Semidefinite Programmierungsschranken

<u>Def. 1</u> Sei $G=(V,E)$. Definiere das semidefinite Programm

$$\vartheta(G) = \max \langle J, X \rangle$$
$$X \succeq 0$$
$$\mathrm{Tr}(X) = 1$$
$$X_{ij} = 0 \quad \text{für } \{i,j\} \in E.$$

ϑ heißt die <u>Lovász ϑ-Zahl</u>.

<u>Bem.:</u> $\vartheta(G)$ = semidefinite Relaxierung des quadratischen Programms, das $\alpha(G)$ beschreibt und in Bsp. II.3.3 gegeben wurde.

Beweis davon $\rightarrow$ Aufgabenblatt 8.

<u>Satz 2</u> (Lovász Sandwich Theorem)

$$\alpha(G) \le \vartheta(G) \le \chi(\overline{G}).$$

<u>Bew.:</u> <u>$\alpha(G) \le \vartheta(G)$</u>

Sei $S \subseteq V$ eine unabhängige Menge mit $|S| = \alpha(G)$.
Definiere den charakteristischen Vektor $\chi^S \in \mathbb{R}^V$ durch

$$(\chi^S)_v = \begin{cases} 1 & \text{falls } v \in S \\ 0 & \text{sonst.} \end{cases}$$

Definiere die Matrix

$$X = \frac{1}{|S|} \chi^S (\chi^S)^T \in \mathcal{S}^V.$$

Dann ist X eine zulässige Lösung von $\vartheta(G)$ mit $\langle J, X \rangle = |S|$.
Also $\alpha(G) \le \vartheta(G)$.

$\vartheta(G) \le \chi(\overline{G})$. Es sei X eine zulässige Lösung von $\vartheta(G)$. Sei $V = C_1 \cup \ldots \cup C_k$ eine Partition von V in k Cliquen.

Beh.: $\langle J, X \rangle \le k \quad (\Rightarrow \vartheta(G) \le \chi(\overline{G}))$.

Bew.: Es ist

$$\sum_{i=1}^{k} \chi^{C_i} = e = (1, \ldots, 1)^T.$$

und

$$\begin{aligned} Y &:= \sum_{i=1}^{k} (k\chi^{C_i} - e)(k\chi^{C_i} - e)^T \\ &= k^2 \sum_{i=1}^{k} \chi^{C_i}(\chi^{C_i})^T - e\underbrace{\left(\sum_{i=1}^{k} k\chi^{C_i}\right)^T}_{ke^T} - \underbrace{\sum_{i=1}^{k} k\chi^{C_i}}_{ke} e^T \\ &\quad + \sum_{i=1}^{k} ee^T \\ &= k^2 \sum_{i=1}^{k} \chi^{C_i}(\chi^{C_i})^T - kJ \quad , \quad \text{wobei } J = ee^T. \\ &\succeq 0. \end{aligned}$$

Desweiteren

$$\begin{aligned} 0 \le \langle X, Y \rangle &= k^2 \langle X, \sum_{i=1}^{k} \chi^{C_i}(\chi^{C_i})^T \rangle - k \langle J, X \rangle \\ &= k^2 \underbrace{\operatorname{Tr}(X)}_{=1} - k\langle J, X \rangle \end{aligned}$$

$\Longrightarrow \langle J, X \rangle \le k.$ $\boxtimes$

§ 3 Perfekte Graphen

Def. 1 Sei $G = (V, E)$ ein Graph und $S \subseteq V$.

$$G[S] = (S, \{\{i,j\} \in E : i \in S, j \in S\})$$

ist der durch S induzierte Teilgraph von G.

Ein Graph H heißt induzierter Teilgraph von G, falls es ein $S \subseteq V$ mit $H = G[S]$ gibt. Notation: $H \trianglelefteq G$

Def. 2 Ein Graph heißt perfekt, falls für alle induzierten Teilgraphen H stets $\omega(H) = \chi(H)$ gilt.

Bsp. 3 a) bipartite Graphen sind perfekt ($\omega(H) = \underbrace{\chi(H)}_{=2}$ klar)

b) C_{2n+1}, $\overline{C_{2n+1}}$ nicht perfekt.

c) Kantengraphen $L(G)$ von bipartiten Graphen G sind perfekt ($\rightarrow$ Blatt 8), dabei

$$L(G) = (E, \{\{e_1, e_2\} : e_1 \cap e_2 \neq \emptyset\}).$$

„Strong perfect graph theorem" (Chudnovsky, Robertson, Seymour, Thomas 2006)

Ein Graph ist perfekt genau dann, wenn weder C_{2n+1} noch $\overline{C}_{2n+1}$ induzierte Teilgraphen sind.

<u>Bew.</u>: ist ca. 180 Seiten lang.

<u>Theorem 4</u> („perfect graph theorem", Lovász 1972)

G ist perfekt $\Longleftrightarrow$ $\overline{G}$ ist perfekt.

<u>Bew.</u>: Zeigen: G ist perfekt $\Longleftrightarrow$ für jeden induzierten Teilgraphen $H \subseteq G$ gilt $|V(H)| \leq \omega(H)\,\alpha(H)$.

[Da $\omega(H) = \alpha(\overline{H})$ und $\alpha(H) = \omega(\overline{H})$ folgt Theorem 4.].

<u>„$\Rightarrow$"</u>: Sei G perfekt und $H \subseteq G$. Dann $\chi(H) = \omega(H)$, d.h. $V(H)$ kann von $\omega(H)$ vielen unabhängigen Mengen überdeckt werden. Also

$$|V(H)| \leq \omega(H)\,\alpha(H).$$

„$\Leftarrow$" Ang. G erfüllt die Ungleichungen

$|V(H)| \le \omega(H)\alpha(H)$, aber G ist nicht perfekt.

Wähle G mit $|V(G)|$ minimal. D.h.

$\omega(G) < \chi(G)$ und für jeden induzierten Teilgraphen

$H \subseteq G$ mit $H \neq G$ gilt $\omega(H) = \chi(H)$.

1. Beh.: $\exists\ \alpha(G)\omega(G)+1$ unabh. Mengen $S_0, \dots, S_{\alpha(G)\omega(G)}$, so dass jeder Knoten in G von exakt $\alpha(G)$ dieser Mengen überdeckt wird.

Bew.: Sei S_0 eine unabh. Menge mit $|S| = \alpha(G)$.

$\forall v \in S_0$: $G[V \setminus \{v_0\}]$ ist perfekt (wg. der Minimalität von G).

Dann

$$\chi(G[V\setminus\{v_0\}]) = \omega(G[V\setminus\{v_0\}]) \le \omega(G).$$

Also kann $V \setminus \{v_0\}$ in $\omega(G)$ unabh. Mengen partitioniert werden. Dies machen wir für jeden der $\alpha(G)$ Knoten in S_0 und bekommen $\alpha(G)\,\omega(G)$ unabh. Mengen $S_1, \dots, S_{\alpha(G)\omega(G)}$, die zusammen mit S_0 die Beh. erfüllen.

2. Beh.: $\forall i = 0, \dots, \alpha(G)\omega(G)$ $\exists K_i$ Clique, $|K_i| = \omega(G)$ mit $K_i \cap S_i = \emptyset$ und $K_i \cap S_j \neq \emptyset$ für $j \neq i$.

Bew.: Da $G[V \setminus S_i]$ perfekt ist gilt

$$\chi(G[V \setminus S_i]) = \omega(G[V \setminus S_i]) \le \omega(G).$$

Dann muss $\chi(G[V \setminus S_i]) = \omega(G)$ sein, denn falls $\chi(G[V \setminus S_i]) \le \omega(G) - 1$ wäre, könnte man G mit $\omega(G)$ Farben färben. Also $\omega(G[V \setminus S_i]) = \omega(G)$ und somit gibt es eine Clique K_i mit $K_i \cap S_i = \emptyset$ und $|K_i| = \omega(G)$.

Außerdem gilt $K_i \cap S_j \neq \emptyset$ für $j \neq i$, weil jedes der $\omega(G)$ Elemente von K_i gehört zu $\alpha(G)$ unabh. Mengen S_j (Beh. 1) und $|K_i \cap S_j| \le 1$, also $|K_i \cap S_j| = 1$.

Definiere Matrizen $M, N \in \mathbb{R}^{|V| \times (\alpha(G)\omega(G)+1)}$ durch

$$M = \left[\chi^{S_0} \; \dots \; \chi^{S_{\alpha(G)\omega(G)+1}} \right]$$

$$N = \left[\chi^{K_0} \; \dots \; \chi^{K_{\alpha(G)\omega(G)+1}} \right].$$

Dann gilt $M^T N = J - I$ wegen Beh. 2.

Nun ist $J-I$ eine reguläre Matrix, d.h.

$$\operatorname{rank}(M^T N) = \operatorname{rank}(J-I) = \alpha(G)\,\omega(G)+1.$$

Andererseits

$$\operatorname{rank}(M^T N) \le \operatorname{rank}(N) \le |V(G)|.$$

Also $\quad \alpha(G)\,\omega(G) < |V(G)|. \quad$ $\lightning$.

<u>Korollar 5</u> Sei G perfekt. Dann gilt

$$\alpha(G) = \vartheta(G) = \chi(\overline{G}).$$

<u>Bew.</u>: Nach Satz III.2.2 und Theorem 4 folgt

$$\alpha(G) \le \vartheta(G) \le \chi(\overline{G}) = \omega(\overline{G}) = \alpha(G),$$

also Gleichheit überall.

D.h. man kann α, ω, χ in perfekten Graphen effizient bestimmen.

Jetzt: Finde $S \subseteq V$ mit $|S| = \alpha(G)$ bzw.

Partition $V = S_1 \cup \ldots \cup S_k$ mit $k = \chi(G)$.

Satz 6 Sei $G = (V, E)$ perfekt. Man kann ein $S \subseteq V$ mit $|S| = \alpha(G)$ effizient bestimmen.

Bew.: **geg.:** $G = (V, E)$ mit $V = \{ i_1, \dots, i_n \}$.

ges.: $S \subseteq V$ mit $|S| = \alpha(G)$

$G_0 = G$

for $j = 1$ to n

Berechne $\alpha(G_{j-1} \setminus i_j)$

if $\alpha(G_{j-1} \setminus i_j) = \alpha(G)$

$G_j = G_{j-1} \setminus i_j$

else

$G_j = G_{j-1}$

$S = G_n$

Falls G perfekt ist, kann man $\alpha(G)$ mit $\vartheta(G)$ berechnen. Insgesamt werden $n+1$ Berechnungen von ϑ benötigt.

Beh.: $|S| = \alpha(G)$ und S ist unabh.

Bew.: klar: $\alpha(G) = |S|$.

Ang. S ist nicht unabh. Dann $\exists\, i_j$ mit $\alpha(G_n \setminus i_j) =$
$= \alpha(G)$.

Aber G_n ist induzierter Teilgraph von $\mathcal{G}_{i_{j-1}}$, also

$$\alpha(G_n \setminus i_j) \leq \alpha(G_{i_{j-1}} \setminus i_j) = \alpha(G),$$

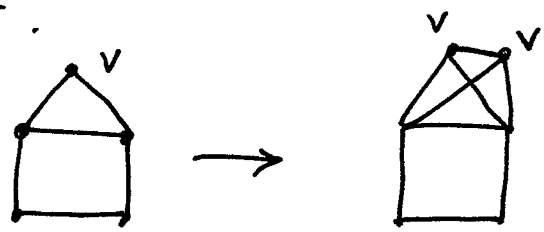
so dass der Knoten i_j schon im j-ten Schritt gelöscht werden würde.

Zur effizienten Bestimmung einer $\chi(G)$-Färbung:

Def. 7 $\mathcal{G} = (V, E)$ Graph, $v \in V$.

$$H = (V \cup \{v'\},\ E \cup \{\{v', w\} : \{v,w\} \in E\} \cup \{v,v'\})$$

ist der Graph, der durch Duplikation des Knoten v aus $\mathcal{G}$ entsteht.



Satz 8 Sei G perfekt und H entstehe aus $\mathcal{G}$ durch Duplikation eines Knoten. Dann ist H perfekt.

Bew.: Zeige: $\overline{H}$ ist perfekt (und wende das perfekte Graphentheorem III.3.4 an).

1. $\alpha(H) = \chi(\overline{H})$.

Nach Konstruktion ist

$$\alpha(H) = \alpha(\mathcal{G}) \underset{\substack{\uparrow \\ G \text{ perfekt}}}{=} \chi(\overline{G})$$

Sei $V = C_1 \dot\cup \ldots \dot\cup C_t$ eine Partition von V in Cliquen.
Ang. $v \in C_1$. Dann ist $V \cup \{v'\} = (C_1 \cup \{v'\}) \dot\cup \ldots \dot\cup C_t$ eine Partition von $V \cup \{v'\}$ in Cliquen. Also $\chi(\overline{H}) = \chi(\overline{G})$ und somit $\alpha(H) = \chi(\overline{H})$.

2. $\alpha(H') = \chi(\overline{H'})$ für $H' \subseteq H$.

Geht genauso, falls $v' \in V(H')$. Falls $v' \notin V(H')$, dann ist $H' \subseteq \mathcal{G}$ und $\alpha(H') = \chi(\overline{H'})$ wegen der Perfektheit von $\mathcal{G}$. ☒

<u>Def. 9</u> Sei $G = (V, E)$ und $w: V \to \mathbb{Z}_{\geq 0}$ eine Gewichtsfkt. Die <u>gewichtete Unabhängigkeitszahl</u> ist

$$\alpha_w(\mathcal{G}) = \max\left\{\sum_{v \in S} w(v) \;:\; S \subseteq V \text{ unabh. in } \mathcal{G}\right\}.$$

Es gilt $\alpha_w(G) = \alpha((\mathcal{G})_w)$, wobei $(\mathcal{G})_w$ aus $\mathcal{G}$ entsteht, indem jeder Knoten $v \in V$ $w(v)$-fach dupliziert wird, und alle Kanten, die einen Knoten zu seinem Duplikat verbinden wieder gelöscht werden.

Lemma 10 Es gibt eine unabh. Menge S in einem perfekten Graph G, die mit allen Cliquen C in $\mathcal{G}$ mit $|C| = \omega(\mathcal{G})$ einen nicht-leeren Schnitt hat.

Bew.: Sei $V = S_1 \dot\cup \ldots \dot\cup S_{\omega(\mathcal{G})}$ eine optimale Färbung von $\mathcal{G}$. Dann gilt für jedes S_i:

$$\omega(\mathcal{G}[V \setminus S_i]) = \chi(\mathcal{G}[V \setminus S_i]) = \omega(\mathcal{G}) - 1.$$

Idee für Algorithmen: Finde eine solche unabh. Menge S und färbe $\mathcal{G}[V \setminus S]$ rekursiv mit $\omega(\mathcal{G}[V \setminus S]) = \omega(\mathcal{G}) - 1$ Farben.

Satz 11 Sei $\mathcal{G} = (V, E)$ perfekt. Man kann eine Partition $V = S_1 \dot\cup \ldots \dot\cup S_{\chi(\mathcal{G})}$ bestehend aus unabh. Mengen S_i effizient bestimmen.

Bew.: Algorithmus 1   input: $\mathcal{G} = (V, E)$ perfekter Graph

output: $\chi(\mathcal{G})$-Färbung $V = S_1 \dot\cup \ldots \dot\cup S_{\chi(\mathcal{G})}$

```
i = 1
while V(G) ≠ ∅ do
    S_i = Ergebnis von Algo. 2 (Lemma 10)
    i = i+1,  G = G[V(G) \ S_i]
```

Algorithmus 2   input: $\mathcal{G} = (V, E)$ perfekter Graph

output: unabh. Menge S mit

$\forall C \subseteq V$ Clique in $\mathcal{G}$, $|C| = \omega(\mathcal{G})$:

$|C \cap S| = 1$.

$t = 0$

repeat

$w = \sum_{i=1}^{t} \chi^{C_i}$

Finde mit Algo von Satz 6 unabh. Menge $S \subseteq V$ $(*)$

mit $\sum_{v \in S} w(v) = \alpha_w(\mathcal{G})$. [Wende Satz 6 auf perfekten Graph G_w an].

if $\omega(\mathcal{G}[V \setminus S]) = \omega(\mathcal{G})$

$t = t+1$

Berechne eine Clique C_t in $\mathcal{G}[V \setminus S]$ mit $|C_t| = \omega(\mathcal{G})$. [Satz 6]

until $\omega(\mathcal{G}[V \setminus S]) < \omega(\mathcal{G})$

output S.

Der Schritt $(*)$ stellt sicher, dass S alle Cliquen $C_1, \dots, C_t$ schneidet.

Falls $\omega(\mathcal{G}[V \setminus S]) < \omega(\mathcal{G})$, dann schneidet S jede Clique C in $\mathcal{G}$ mit $|C| = \omega(\mathcal{G})$.

Die Anzahl der Iterationen ist $\leq |V|$:

Betrachte den affinen Unterraum

$$L_t = \{ x \in \mathbb{R}^V : \sum_{i \in C_j} x_i = 1, \quad j = 1, \dots, t \}.$$

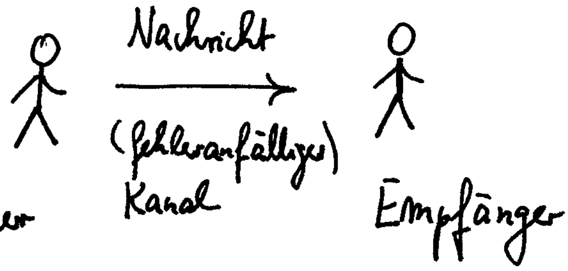
Es gilt $L_t \supsetneq L_{t+1}$, da $X^S \in L_t \setminus L_{t+1}$ für die im t-ten Schritt konstruierte unabh. Menge S.

D.h. $\dim L_{t+1} < \dim L_t$, also $t \leq \dim \mathbb{R}^V = |V|$.

$\boxtimes$

§4 Shannon-Kapazität

Informationstheorie

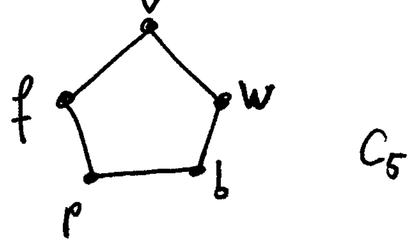


Nachricht: Folge von Symbolen aus V

Problem: Man kann Symbole verwechseln.

Verwechselungsgraph: $G = (V, E)$

z.B.



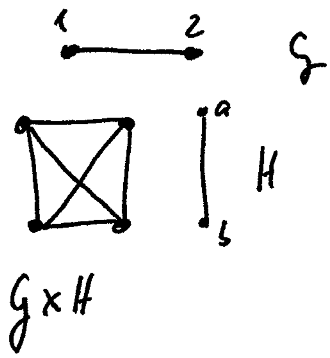
Ziel: Fehlerfreie Kommunikation; verwende nur Symbole, die nicht verwechselt werden können.

Informationsrate: $\alpha(G)$ (hier $\alpha(C_5) = 2$)

Trick zur Verbesserung der Informationsrate: Sende mehrere Symbole gleichzeitig.

<u>Def. 1</u> Seien $G = (V, E)$, $H = (W, F)$. Das <u>starke Graphenprodukt</u> $G \times H$ ist

$$G \times H = \Big(V \times W, \big\{ \{(v_1, w_1), (v_2, w_2)\} : \begin{array}{l} v_1 = v_2 \text{ u. } \{w_1, w_2\} \in F, \\ w_1 = w_2 \text{ u. } \{v_1, v_2\} \in E, \\ \{v_1, v_2\} \in E, \{w_1, w_2\} \in F \end{array} \big\} \Big)$$



Notation: $G^k = \underbrace{G \times \cdots \times G}_{k\text{-mal}}$.

<u>Lemma 2</u> $\alpha(G \times H) \geq \alpha(G)\,\alpha(H)$.

Insbesondere $\sqrt[k]{\alpha(G^k)} \geq \alpha(G)$.

<u>Bew.:</u> Sei $S \subseteq V$ unabh. in G mit $\alpha(G) = |S|$,

$T \subseteq W$ unabh. in ~~G~~ H mit $\alpha(H) = |T|$.

Dann ist $S \times T$ unabh. in $G \times H$.

<u>Def. 3</u>: $\Theta(G) = \sup\limits_{k \in \mathbb{N}} \sqrt[k]{\alpha(G^k)}$ heißt Shannon-Kapazität von G.

Bsp.: $\Theta(C_5) \geq \sqrt{5}$ (≥ 2)

Da $\alpha(C_5^2) \geq 5$, denn $\{(1,1), (2,3), (3,5), (4,2), (5,4)\}$ ist unabh. in C_5^2.

Shannon (1956): 1) Ist $\Theta(C_5) = \sqrt{5}$?

2) Wie berechnet man $\Theta(G)$?

1) Lovász (1979): ja

2) Keine Ahnung, selbst $\Theta(C_7)$ ist unbekannt.

Exkurs: Kroneckerprodukt von Matrizen.

Def. 4 Seien $A = (A_{ij}) \in \mathbb{R}^{m \times n}$, $B = (B_{hk}) \in \mathbb{R}^{p \times q}$.
Das Kroneckerprodukt $A \otimes B \in \mathbb{R}^{mp \times nq}$ ist definiert als

$$A_{ih,jk} = A_{ij} B_{hk}.$$

$A \otimes B$ ist eine $m \times n$ Blockmatrix mit Blöcken der Größe $p \times q$.

$$A \otimes B = \begin{bmatrix} A_{11}B & A_{12}B & \dots & A_{1n}B \\ & \dots & & \\ A_{m1}B & A_{m2}B & \dots & A_{mn}B \end{bmatrix}$$

Lemma 5 a) $(A\otimes B)(C\otimes D) = (AC)\otimes(BD)$.

b) $A\in S^m$, $B\in S^n$ mit Eigenwerten $\alpha_1,\dots,\alpha_m$, bzw. $\beta_1,\dots,\beta_n$. Dann sind die Eigenwerte von $A\otimes B$ gleich $\alpha_i\beta_j$ $\quad i\in[m],\ j\in[n]$.

Bew.: direktes Nachrechnen.

Satz 6 $\vartheta(G\times H) = \vartheta(G)\,\vartheta(H)$.

Bew.: 1) $\vartheta(G\times H) \geq \vartheta(G)\,\vartheta(H)$

Betrachte das primale SDP, das ϑ definiert

$$\begin{aligned} \max\ & \langle J, X\rangle \\ & \mathrm{Tr}(X) = 1,\ X\succeq 0 \\ & X_{ij} = 0 \text{ für } \{i,j\}\in E. \end{aligned}$$

Sei X eine zulässige Lösung für $\vartheta(G)$ und Y eine für $\vartheta(H)$. Dann ist $X\otimes Y$ eine zulässige Lösung für $\vartheta(G\times H)$ mit

$$\langle J, X\otimes Y\rangle = \langle J, X\rangle\,\langle J, Y\rangle,$$

was man durch direktes Nachrechnen verifiziert.

2) $\vartheta(G\times H) \leq \vartheta(G)\,\vartheta(H)$: $\rightarrow$ Blatt 9.

Korollar 7 $\Theta(G) \leq \vartheta(G)$

Bew.: $\sqrt[k]{\alpha(G^k)} \leq \sqrt[k]{\vartheta(G^k)} = \sqrt[k]{\vartheta(G)^k} = \vartheta(G)$.

Satz 8 $\vartheta(C_5) = \sqrt{5}$

Bew.: Aufgabe 3.4 und Aufgabe 7.3.

In §5 werden wir zwei weitere Beweise dafür sehen.

§5 Berechnung von ϑ für symmetrische Graphen

Def. 1 $G = (V,E)$ Graph.

a) Die Automorphismengruppe von G ist

$$\mathrm{Aut}(G) = \{\sigma : V \to V : \sigma \text{ Permutation und } \{i,j\} \in E \Leftrightarrow \{\sigma(i), \sigma(j)\} \in E\}.$$

b) G heißt homogen (knotentransitiv), falls

$$\forall i,j \in V \; \exists \sigma \in \mathrm{Aut}(G) : \sigma(i) = j.$$

Entscheidungsproblem: „$\mathrm{Aut}(G) = \emptyset$?" ist komplexitätstheoretisch interessant. Wissen nicht, ob es in P liegt.

Andererseits liegt es wahrscheinlich nicht in NPC.

Sei $\sigma : V \to V$ eine Permutation. Betrachte die Permutationsmatrix $P^\sigma \in \mathbb{R}^{V \times V}$, die durch

$$P^\sigma_{i,j} = \begin{cases} 1 & , \text{ falls } i = \sigma(j) \\ 0 & , \text{ sonst} \end{cases}$$

bestimmt ist.

Sei $X \in S^V$ eine symmetrische Matrix. Definiere die Linksoperation

$$\sigma(X) = P^\sigma X (P^\sigma)^T = \left(X_{\sigma^{-1}(i), \sigma^{-1}(j)} \right)_{i,j \in V}.$$

<u>Lemma 2</u> Sei $\mathcal{G} = (V, E)$ ein Graph. Betrachte

$$\vartheta(\mathcal{G}) = \max \{ \langle J, X \rangle : X \in S^V_{\geq 0}, \ Tr(X) = 1, \ X_{ij} = 0, \ \{i,j\} \in E \}.$$

Dann gibt es eine optimale Lösung X^* von $\vartheta(\mathcal{G})$, die invariant unter der Operation von $Aut(\mathcal{G})$ ist, d.h.

$$\sigma(X^*) = X^* \quad \forall \sigma \in Aut(\mathcal{G}).$$

<u>Bew.</u>: Sei X eine optimale Lösung von $\vartheta(\mathcal{G})$. Dann ist $\sigma(X)$ ebenfalls optimal für jedes $\sigma \in Aut(\mathcal{G})$.

Definiere $X^* = \frac{1}{|\mathrm{Aut}(\mathcal{G})|} \sum_{\sigma \in \mathrm{Aut}(\mathcal{G})} \sigma(X)$.

Da die Menge der optimalen Lösungen von $\vartheta(\mathcal{G})$ konvex ist, ist X^* optimal. Nach Konstruktion ist $\sigma(X^*) = X^*$ $\forall \sigma \in \mathrm{Aut}(\mathcal{G})$

Satz 3 Sei G knotentransitiv. Dann gibt es eine optimale Lösung von $\vartheta(\mathcal{G})$ mit $X^*_{ii} = \frac{1}{n}$ für alle $i \in V$ und $X^* e = \frac{\vartheta(\mathcal{G})}{n} e$

Bew.: Nach Lemma 2 $\exists X^*$ optimal mit $\sigma(X^*) = X^*$ $\forall \sigma \in \mathrm{Aut}(\mathcal{G})$. Da $\mathcal{G}$ knotentransitiv ist, sind alle Diagonalelemente von X^* gleich: Sei $i, j \in V$ und $\sigma \in \mathrm{Aut}(\mathcal{G})$ mit $\sigma(i) = j$. Dann ist

$$X^*_{jj} = X^*_{\sigma^{-1}(i), \sigma^{-1}(i)} = \sigma(X^*)_{ii} = X^*_{ii}.$$

Weil $\mathrm{Tr}(X^*) = 1$, folgt $X^*_{ii} = \frac{1}{n}$.
Mit dem gleichen Argument zeigt ~~man~~ man, dass die Zeilensummen von X^* alle übereinstimmen, somit $X^* e = \lambda e$. Da $\vartheta(\mathcal{G}) = \langle J, X^* \rangle = e^T X^* e = e^T \lambda e = \lambda n$ folgt die Beh.

Satz 4 Sei G knotentransitiv, dann ist $\vartheta(G)\,\vartheta(\overline{G}) = |V|$.

Bew.: $\underline{\vartheta(G)\,\vartheta(\overline{G}) \geq |V|}$: Aufgabe 7.4

$\underline{\vartheta(G)\,\vartheta(\overline{G}) \leq |V|}$:

$$
\begin{aligned}
\vartheta(\overline{G}) = \min\ & t \\
& Y - J \succeq 0 \\
& Y_{ii} = t \qquad \forall i \in V \\
& Y_{ij} = 0 \qquad \forall \{i,j\} \in E
\end{aligned}
$$

[Aufgabe 7.3].

Verwende optimale Lösung X^* aus Satz 3, um

$Y = \frac{n^2}{\vartheta(G)} X^*$, $t = \frac{n}{\vartheta(G)}$ zu definieren.

Überprüfe, dass (Y, t) gültige Lösung von $\vartheta(\overline{G})$ ist: Dann

$\frac{n^2}{\vartheta(G)} X^* - J \succeq 0$, weil

$$
\frac{n^2}{\vartheta(G)} X^* e - Je = \frac{n^2}{\vartheta(G)} \frac{\vartheta(G)}{n} e - ne = 0.
$$

Also ist $\vartheta(\overline{G}) \leq \frac{n}{\vartheta(G)}$, und die Beh. folgt. $\boxtimes$

Korollar 5 $\vartheta(C_n)\,\vartheta(\overline{C_n}) = n$, insbesondere $\vartheta(C_5) = \sqrt{5}$.

§6 ϑ für Cayley-Graphen endlicher, abelscher Gruppen

Sei $(G,+)$ eine endliche abelsche Gruppe. Nach dem Hauptsatz über endliche abelsche Gruppen ist

$$G \cong \mathbb{Z}/_{n_1}\mathbb{Z} \times \dots \times \mathbb{Z}/_{n_r}\mathbb{Z}.$$

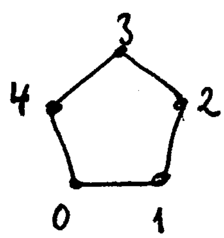
[Stichwort: Moduln über Hauptidealringen; Elementarteiler]

Def. 1 Sei $\Sigma \subseteq G$, mit $\Sigma = -\Sigma$.

$$\mathrm{Cayley}(G,\Sigma) = (G, \{\{x,y\} : x-y \in \Sigma\})$$

heißt der Cayley-Graph von G und Σ.

Bsp.: $C_n = \mathrm{Cayley}(\mathbb{Z}/_n\mathbb{Z}, \{1,-1\})$



Beobachtung: $G \subseteq \mathrm{Aut}(\mathrm{Cayley}(G,\Sigma))$

ist eine Untergruppe, da für $a \in G$ gilt

$$x-y \in \Sigma \iff (a+x)-(a+y) \in \Sigma$$

Ziel: Berechnung von $\vartheta(\mathrm{Cayley}(G,\Sigma))$.

Def. 2 a) $(\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}, \cdot)$ ist der Torus von $\mathbb{C}$

b) Ein Gruppenhomomorphismus $\chi : (G, +) \to (\mathbb{T}, \cdot)$ heißt Charakter von G, d.h. $\chi(x+y) = \chi(x)\chi(y)$.

Bsp.: Sei $a \in \mathbb{Z}/n\mathbb{Z}$. Dann ist $\chi_a(x) = e^{\frac{2\pi i a x}{n}}$ ein Charakter von $\mathbb{Z}/n\mathbb{Z}$, weil

$$\chi_a(x+y) = e^{\frac{2\pi i a(x+y)}{n}} = e^{2\pi i a x/n} e^{2\pi i a y/n}$$
$$= \chi_a(x)\chi_a(y).$$

Wir betrachten Charaktere $\chi : G \to \mathbb{T}$ als Elemente $\chi \in \mathbb{C}^G$ des unitären Vektorraum $\mathbb{C}^G$ mit Skalarprodukt

$$\langle f, g \rangle = \frac{1}{|G|} \sum_{x \in G} \overline{f(x)}\, g(x).$$

Lemma 3 (Orthogonalitätsrelation)

Seien χ, ψ Charaktere von G, dann gilt

$$\langle \chi, \psi \rangle = \begin{cases} 1, & \chi = \psi \\ 0, & \text{sonst}. \end{cases}$$

<u>Bew.:</u>

Falls ψ ein Charakter von G ist, dann ist ψ^{-1} auch einer.

Es gilt

$$\langle \chi, \psi \rangle = \frac{1}{|G|} \sum_{x \in G} \chi(x) \overline{\psi(x)} = \frac{1}{|G|} \sum_{x \in G} \chi(x) \psi^{-1}(x) \cdot \overline{1(x)}$$

$$= \langle \chi \psi^{-1}, 1 \rangle,$$

wobei $1 : G \to \mathbb{T}$ der triviale Charakter (mit $1(x) = 1$) ist.

Außerdem ist das Produkt zweier Charakter wieder ein Charakter, d.h. wir können uns auf die Situation „$\langle \chi, 1 \rangle = ?$" zurückziehen.

Falls $\chi = 1$, dann ist $\langle \chi, 1 \rangle = 1$.

Falls $\chi \neq 1$, dann gibt es ein $y \in G$ mit $\chi(y) \neq 1$.

Es gilt

$$\chi(y) \langle \chi, 1 \rangle = \chi(y) \frac{1}{|G|} \sum_{x \in G} \chi(x) = \frac{1}{|G|} \sum_{x \in G} \chi(y+x)$$

$$= \frac{1}{|G|} \sum_{x \in G} \chi(x) = \langle \chi, 1 \rangle,$$

also $\langle \chi, 1 \rangle = 0$, da $\chi(y) \neq 0$.

$\boxtimes$

<u>Korollar 4</u> Sei $G = \mathbb{Z}/_{n_1\mathbb{Z}} \times \dots \times \mathbb{Z}/_{n_r\mathbb{Z}}$. Dann bilden die Charaktere

$$\chi_{a_1 \dots a_r}(x_1, \dots, x_r) = \prod_{j=1}^{r} e^{\frac{2\pi i\, a_j x_j}{n_j}}$$

mit $(a_1, \dots, a_r), (x_1, \dots, x_r) \in G$ [beachte, dass χ wohldefiniert ist] eine Orthonormalbasis von $\mathbb{C}^G$.

<u>Def. 5</u> $(\hat{G} = \{\chi : \chi \text{ ist Charakter von } G\}, \cdot)$ heisst die zu G <u>duale Gruppe</u>.

Nach Korollar 4 ist $(G, +) \cong (\hat{G}, \cdot)$

<u>Lemma 6</u> Sei $\mathrm{Cayley}(G, \Sigma)$ gegeben.
Sei $X \in S^G$ mit $g(X) = X$ für alle $g \in G$. Dann gilt $X_{x,y} = X_{x',y'}$ für alle x, x', y, y' mit $x - y = x' - y'$, oder $-(x-y) = x' - y'$.

<u>Bsp</u>: $\mathbb{Z}/_{5\mathbb{Z}} = \{0, 1, -1, 2, -2\}$, $\Sigma = \{\pm 1\}$

$$X = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & x_0 & x_1 & x_2 & x_2 & x_1 \\ 1 & x_1 & x_0 & x_1 & x_2 & x_2 \\ 2 & x_2 & x_1 & x_0 & x_1 & x_2 \\ 3 & x_2 & x_2 & x_1 & x_0 & x_1 \\ 4 & x_1 & x_2 & x_2 & x_1 & x_0 \end{array}$$

Bew.: Es ist

$$g(X)_{x,y} = X_{x-g,y-g} = X_{x,y}.$$

Wähle $g = x - x'$. Dann ist

$$x - g = x - (x - x') = x',$$

$$y - g = y - (x - x') = -(x - y) + x' = -(x' - y') + x' = y'.$$

Außerdem ist wegen der Symmetrie von X

$$X_{x,y} = X_{y-g,x-g}.$$

Wähle $g = y - x'$. Dann ist

$$y - g = y - (y - x') = x',$$

$$x - g = x - (y - x') = -(x' - y') + x' = y'. \qquad \boxtimes$$

Satz 7 Sei $f \in \mathbb{C}^G$ mit $f(x) = \sum_{\chi \in \hat{G}} f_\chi \chi(x)$.

Definiere die Matrix $F \in \mathbb{C}^{G \times G}$ durch

$$F_{xy} = f(x - y), \quad x, y \in G.$$

a) F ist hermitesch ($F = \overline{F^T} = F^*$) $\iff$ $f_\chi \in \mathbb{R}$ $\forall \chi \in \hat{G}$.

b) F ist herm. positiv semidefinit ($z^* F z \geq 0$ $\forall z \in \mathbb{C}^G$) $\iff$ $f_\chi \geq 0$ $\forall \chi \in \hat{G}$.

c) $F \in \mathbb{R}^{G \times G}$ $\iff$ $f_\chi = f_{\overline{\chi}}$ $\forall \chi \in \hat{G}$.

<u>Bew.:</u>

a) $$F = F^* \iff f(x-y) = \overline{f(y-x)}$$
$$\iff \sum_\chi f_\chi\, \chi(x-y) = \sum_\chi \overline{f_\chi}\, \overline{\chi(y-x)}$$
$$\iff \sum_\chi f_\chi\, \chi(x)\, \overline{\chi(y)} = \sum_\chi \overline{f_\chi}\, \chi(x)\, \overline{\chi(y)}$$
$$\iff f_\chi = \overline{f_\chi}\,.$$

b) <u>„$\Longleftarrow$"</u>: Sei $z \in \mathbb{C}^G$. Dann

$$z^* F z = \sum_x \sum_y \bar{z}_x \sum_\chi f_\chi\, \chi(x-y)\, z_y$$
$$= \sum_\chi f_\chi\, (z^* \chi)^2 \geq 0.$$

<u>„$\Longrightarrow$"</u>: Sei $\psi \in \hat{G}$. Dann

$$0 \leq \psi^* F \psi = \sum_\chi f_\chi\, (\psi^* \chi)^2 = f_\psi\, (\psi^* \psi)^2$$

Orthogonalitäts-
relation (Lemma 3).

Also $f_\psi \geq 0$.

c) selbst.

Satz 8

$$\vartheta(\mathrm{Cayley}(G,\Sigma)) = \max f_0 \cdot |G|$$
$$f_\chi \geq 0, \quad f_\chi = f_{\bar\chi}, \quad \chi \in \hat G$$
$$\sum_\chi f_\chi = 1$$
$$\sum_\chi f_\chi \, \chi(x) = 0, \quad x \in \Sigma$$

Das SDP vereinfacht sich zu einem LP.

Bew.: Nach Lemma 5.2, Lemma 6.6, Satz 6.7 können wir annehmen, dass eine optimale Lösung X^* für ϑ die Form

$$X^*_{xy} = \sum_\chi f_\chi \, \chi(x-y) \quad \text{mit} \quad f_\chi \geq 0, \quad f_\chi = f_{\bar\chi}$$

hat.

Es gilt $\mathrm{Tr}(X^*) = \sum_{x\in G} \sum_\chi f_\chi \, \chi(x-x) = |G| \sum_\chi f_\chi$

Außerdem $\langle J, X^* \rangle = \sum_{x,y\in G} \sum_\chi f_\chi \, \chi(x-y)$

$$= \sum_\chi f_\chi \sum_x \chi(x) \sum_y \overline{\chi(y)}$$
$$= f_0 \, |G|^2$$

Wegen der Orthogonalitätsrelation (angewendet auf den trivialen Charakter).

Dividieren durch $|G|$ liefert die Beh. $\boxtimes$

<u>Korollar 9</u> $\vartheta(C_n) = \dfrac{n\cos(\frac{\pi}{n})}{1+\cos(\frac{\pi}{n})}$ für n ungerade

<u>Bew.:</u> $C_n = \mathrm{Cayley}(\mathbb{Z}/n\mathbb{Z}, \{\pm 1\})$.

Nach Satz 8

$$\vartheta(C_n) = \max f_0$$

$$f_a \geq 0, \ f_a = f_{-a}, \quad a \in \mathbb{Z}/n\mathbb{Z}$$

$$\sum_a f_a = 1,$$

$$\sum_a f_a e^{\frac{2\pi i a \cdot 1}{n}} = 0,$$

$$\sum_a f_a e^{\frac{2\pi i a \cdot (-1)}{n}} = 0.$$

Weil $f_a = f_{-a}$ ist

$$\sum f_a e^{\frac{2\pi i a \cdot 1}{n}} = f_0 + \sum_{a=1}^{(n-1)/2} f_a \left(e^{\frac{2\pi i a \cdot 1}{n}} + e^{\frac{2\pi i (-a) 1}{n}}\right)$$

$$= f_0 + \sum_{a=1}^{(n-1)/2} 2 f_a \cos\frac{2\pi a}{n}.$$

und $1 = f_0 + 2\sum_{a=1}^{(n-1)/2} f_a$.

Um f_0 maximal zu wählen, muss

$$\sum_{a=1}^{(n-1)/2} 2 f_a \cos \frac{2\pi a}{n}$$

minimal sein unter der Nebenbedingung $f_a \geq 0$, $f_0 + 2 \sum_{a=1}^{\frac{(n-1)}{2}} f_a = 1$.

Es ist

$$\min_{a=1,\dots,\frac{(n-1)}{2}} \cos \frac{2\pi a}{n} = \cos \frac{\pi (n-1)}{n} = -\cos \frac{\pi}{n}.$$

Also

$$1 = f_0 + 2 f_{(n-1)/2}$$

$$0 = f_0 + 2 f_{(n-1)/2} \left(-\cos \frac{\pi}{n}\right)$$

$$= f_0 + (-1 + f_0) \cos \frac{\pi}{n}$$

$$\Rightarrow f_0 = \frac{\cos \frac{\pi}{n}}{1 + \cos \frac{\pi}{n}}. \qquad \boxtimes$$

Kapitel IV: Das Kusszahlproblem

$$\tau_2 = 6$$



$$\tau_3 \geq 12$$

Anno MDCXCIV

Isaac Newton

David Gregory

Richtige Antwort: 12

Erste mathematische Beweis davon:

Schütte, van der Waerden (1953)

mit Hilfe Elementargeometrie

Math. Annalen, Bd. 125, S. 325—334 (1953).

Das Problem der dreizehn Kugeln.

Von

K. Schütte in Marburg/Lahn und B. L. van der Waerden in Zürich.

Die Frage, wieviel Kugeln gleicher Größe an eine ebenso große Kugel angelegt werden können, ist offenbar gleichbedeutend mit der Frage, wieviel Punkte mit Mindestabstand 1 auf einer Kugel vom Radius 1 Platz haben (Fig. 1).

Zwölf Kugeln kann man sicher anlegen, nämlich an den Ecken eines Ikosaeders. (Die Kantenlänge desselben ist größer als der Radius der umbeschriebenen Kugel.) Es ist eine alte Streitfrage zwischen Newton und Gregory, ob sich auch 13 Kugeln anlegen lassen. Im folgenden werden zwei Beweise dafür gegeben, daß dies nicht möglich ist. Es wird nämlich gezeigt, daß die Minimalkugel, auf der 13 Punkte mit Mindestabstand 1 Platz haben, einen Radius $r > 1$ besitzt.

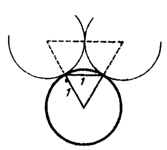


Fig. 1.

Die Beweise erfolgen unter Benutzung von Methoden, die zur Auffindung der Minimalkugel für N Punkte mit Mindestabstand 1 entwickelt wurden[1], [2]. Hinweise auf Paragraphen und Sätze der unter [2] zitierten Arbeit, deren Begriffsbildungen hier verwendet werden, sind in eckige Klammern gesetzt.

Der allgemeine Ansatz für diese Untersuchung (§§ 1, 2) stammt von beiden Verfassern. Den im § 3 dargestellten Beweis hat van der Waerden gefunden, den anderen Beweis (§ 6) mit den dazu benötigten Hilfssätzen (§§ 4, 5) Schütte.

Die Priorität hat Schütte. Sein Beweis (§ 6) ist zwar komplizierter, aber es erscheint uns trotzdem erwünscht, ihn mitzuteilen, weil die Hilfssätze, auf denen er beruht, sich für weitere Untersuchungen als nützlich erweisen könnten, so z. B. für die genaue Bestimmung der Minimalkugel für 10, 11 oder 13 Punkte.

§ 1. Der irreduzible Graph.

Satz 1. Ein irreduzibler Graph auf einer Kugel vom Radius $r \leqq 1$ enthält keine anderen Polygone als Dreiecke, Vierecke und Fünfecke.

Beweis: Bei $r \leqq 1$ ist der zum Abstand 1 gehörende Bogen $a \geqq 60°$ (Fig. 2). Folglich gilt für ein Polygon mit n Ecken

$$n \cdot 60° \leqq n \cdot a < 360° \qquad [§ 6, Satz 6]$$

und somit $n < 6$.

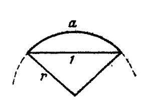


Fig. 2.

Satz 2. Auf einer Minimalkugel für $N > 12$ hat im Innern eines Fünfecks mit den Seiten a kein Punkt ohne Störung Platz.

[1] W. Habicht u. B. L. van der Waerden: Lagerung von Punkten auf der Kugel. Math. Ann. 123, 223—234 (1951).

[2] K. Schütte u. B. L. van der Waerden: Auf welcher Kugel haben 5, 6, 7, 8 oder 9 Punkte mit Mindestabstand Eins Platz? Math. Ann. 123, 96—124 (1951).

Beweis: Hätte ein Punkt im Innern Platz, so könnte man diesen heranschieben, so daß ein Sechseck und ein Dreieck oder ein Fünfeck und ein Viereck entstehen würde. Der gesamte Flächeninhalt wäre (nach Habicht und van der Waerden) $\geqq 5\,\triangle$ (Fig. 3), wenn $\triangle$ der Inhalt des sphärischen Dreiecks mit den Seiten a ist. Das regelmäßige Fünfeck, das ja maximalen Inhalt besitzt, hat aber bei $N > 12$ einen Inhalt $< 5\,\triangle$ (da die Kugel dann notwendig größer ist als für $N = 12$, wo dieser Inhalt gemäß der Ikosaederfigur gleich $5\,\triangle$ ist).

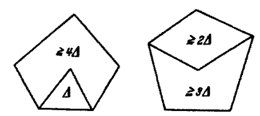


Fig. 3.

Satz 3. Als Minimalfigur für $N = 13$ auf einer Kugel vom Radius $r \leqq 1$ kommt nur ein irreduzibler Graph aus Dreiecken, Vierecken und Fünfecken in Betracht.

Beweis: Auf der Minimalkugel tritt mindestens ein irreduzibler Graph auf [§ 5]. Dieser enthält nach Satz 1 nur Dreiecke, Vierecke und Fünfecke. Nach Satz 2 (und [§ 6, Satz 7]) können ihm keine Punkte ohne Störung eingefügt werden. Er enthält also bereits alle 13 Punkte.

§ 2. Der Winkelüberschuß.

Nach Habicht und van der Waerden ist der Inhalt eines Polygons (mit den Seiten a) mindestens gleich dem Inhalt, der sich ergibt, wenn das Polygon vollständig in Dreiecke zerfällt. Der Inhalt eines n-eckigen Polygons ist also $\geqq (n-2)\,\triangle$. Dann ist die Summe σ der Innenwinkel $\geqq (n-2) \cdot 3\,\alpha$, wenn α den Winkel des sphärischen Dreiecks mit den Seiten a bezeichnet. Die (demnach nicht negative) Differenz

$$w = \sigma - (n-2) \cdot 3\,\alpha$$

bezeichnen wir als den „*Winkelüberschuß*" des betreffenden Polygons.

Der „*Gesamt-Winkelüberschuß*" W der Kugel ergibt sich, wenn über alle Polygone summiert wird.

$$W = \Sigma\,\sigma - \Sigma\,(n-2) \cdot 3\,\alpha\,.$$

Es ist $\Sigma\,\sigma = N \cdot 2\,\pi$ und $\Sigma\,(n-2) = \Sigma\,n - 2G = 2K - 2G$, wo K die Anzahl der Strecken und G die Anzahl der Polygone bedeutet. Nach dem Eulerschen Polyedersatz ist $K - G = N - 2$, also

$$W = N \cdot 2\,\pi - 2\,(N-2) \cdot 3\,\alpha\,.$$

Insbesondere ist für $N = 13$

$$W = 26\,\pi - 66\,\alpha\,. \tag{1}$$

Das „Quadrat" (regelmäßige sphärische Viereck) mit dem Winkel β hat den Winkelüberschuß $Q = 4\,\beta - 6\,\alpha$. Im Falle $r \leqq 1$ ist $\cos a \leqq {}^1/_2$, $\cos \alpha \leqq {}^1/_3$ [§ 7, (2)] und daher (nach [§ 11, (11)])

$$\cos \beta = \frac{\cos a - 1}{\cos a + 1} \leqq -\frac{1}{3} \leqq -\cos \alpha\,,$$

$$\alpha + \beta \geqq \pi\,, \tag{2}$$

folglich

$$Q \geqq 4\,\pi - 10\,\alpha\,. \tag{3}$$

Für den Winkelüberschuß eines Fünfecks aus einem Graphen gemäß Satz 3 läßt sich eine untere Schranke w_5 angeben. Da der Graph irreduzibel ist, sind alle inneren Winkel $< \pi$, also alle Polygone konvex. Das konvexe Fünfeck kleinsten Inhalts besitzt zwei (nicht benachbarte) gestreckte Winkel, d. h. es wird von einem gleichschenkligen Dreieck mit den Schenkeln $2a$ und der Grundlinie a gebildet [§ 13, Beweis von Satz 12]. Bei $r = 1$ zerfällt das gleichschenklige Dreieck in ein gleichseitiges Dreieck und in ein regelmäßiges Viereck (Fig. 4), da in diesem Falle $\alpha + \beta = \pi$ ist. Die Summe der Innenwinkel ist dann $\sigma = 2\,\pi + \alpha + 2\,\beta$. In dem allgemeineren Falle $r \leqq 1$ hat das gleichschenklige Dreieck einen Scheitelwinkel $\geqq \alpha$ und einen Basiswinkel $\geqq \beta$. Dann ist $\sigma \geqq 2\,\pi + \alpha + 2\,\beta \geqq 4\,\pi - \alpha$, und für die untere Schranke $w_5 = \sigma - 9\,\alpha$ des Fünfeck-Winkelüberschusses gilt

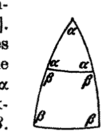


Fig. 4.

$$w_5 \geqq 4\,\pi - 10\,\alpha\,. \tag{4}$$

Aus diesen Abschätzungen folgt:

Satz 4. Falls die Minimalkugel für $N = 13$ einen Radius $r \leqq 1$ besitzt, enthält der Minimalgraph höchstens ein Quadrat und höchstens ein Fünfeck, auch nicht ein Quadrat und ein Fünfeck zugleich.

Beweis: Bei $r \leqq 1$ ist, wie aus $\cos \alpha \leqq 1/3$ hervorgeht, $\alpha > 70{,}5°$. Dann gilt nach (1) und (3)

$$Q - \frac{1}{2}\,W \geqq (4\,\pi - 10\,\alpha) - (13\,\pi - 33\,\alpha) = 23\,\alpha - 9\,\pi$$

$$> 1621{,}5° - 1620° > 0\,,$$

also $2\,Q > W$ und nach (4) ebenso $2\,w_5 > W$. Durch zwei Quadrate, zwei Fünfecke oder ein Quadrat und ein Fünfeck wäre also bereits der Gesamt-Winkelüberschuß W der Kugel überschritten.

§ 3. Erster Beweis für $r > 1$ bei $N = 13$.

Angenommen, die Minimalkugel für $N = 13$ habe einen Radius $r \leqq 1$. Nach Satz 3 wird dann die Minimalfigur von einem irreduziblen Graphen aus Dreiecken, Vierecken und Fünfecken gebildet. Es soll gezeigt werden, daß ein solcher Graph nicht existiert.

I. Fall. Der Graph enthält nur Dreiecke und Vierecke.

Wir bezeichnen alle Dreieckwinkel α und diejenigen Viereckwinkel, die kleiner als der Quadratwinkel β sind, als „kleine Winkel". Alle Viereckwinkel, die größer als β sind, nennen wir „groß". Von einem Quadrat sollen ganz willkürlich zwei Winkel zu den kleinen, die anderen beiden zu den großen Winkeln gerechnet werden. Dann enthält jedes Viereck zwei kleine und zwei große Winkel. Für jeden kleinen Winkel $\varkappa$ gilt

$$\alpha \leqq \varkappa \leqq \beta \tag{5}$$

und für jeden großen Winkel γ

$$\beta \leqq \gamma < 2\,\alpha\,. \tag{6}$$

Der gesamte Winkelüberschuß aller Vierecke um einen Punkt P herum kann kleiner gemacht werden, indem immer ein Winkel verkleinert und ein anderer vergrößert wird, so lange bis alle Winkel bis auf einen den nach (5)

und (6) kleinst- oder
wir so vor: Solange n
sind, vergrößern wir
großen Winkel gleic
gleich α (Fall B).
Im Fall A ändern
ihrer Vierecke verkle
einen gleich $2\,\pi - 4$
gleich α und der letz
lich, da zwei Quadr
Im Fall B ände
geworden sind bis a
übrigen bis auf eine
Fall B_1 ist aber nicht
nur möglich für $r =$
zwei Winkel $= \beta$ u
Nun sei w_P der
Punkt P einen klein
ecks mit Winkel $2\,\pi$

(7)

Summiert man (
Viereck zwei „kleine
erhält man $13\,w_0$. S

Das ist aber ein
ergibt. Für $r = 1$ ste
Für $r < 1$ ist es no
Seite $2 \cdot (26\,\pi - 66$
II. Fall. Der Gr
Nach Satz 4 ka
Winkelüberschuß vo

Wir summieren n
und addieren noch 2
die mit einer kleinen
So ergibt sich

ein eklatanter Wide
mehr als $8 \cdot 4° + 2 \cdot$

Um noch einen z
$N = 13$ und $r \leqq 1$ ge
Eigenschaften konve

Seitdem viele Alternativbeweise: Leech (1956)

Anstreicher (2004), Musin (2006), Maehara (2007),

Bachoc, V. (2008)

Alle sind ziemlich rechenintensiv.

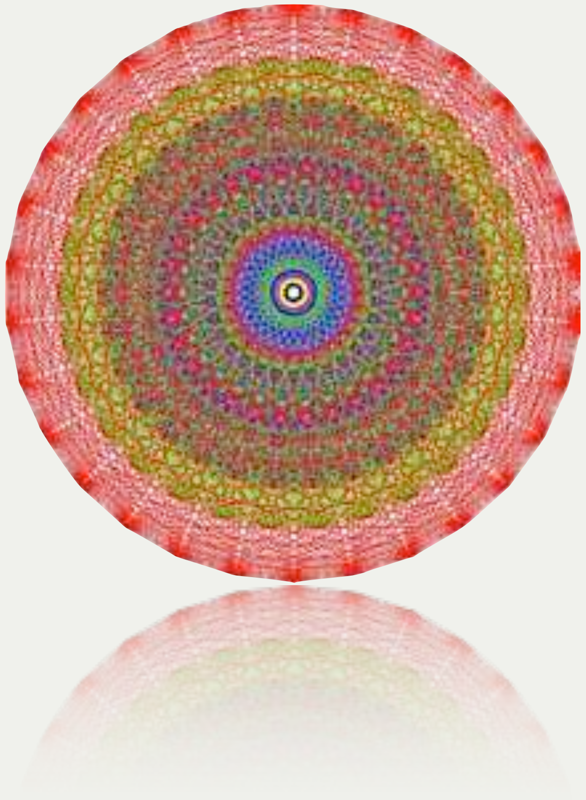
BV (2008): Das folgende Polynom ist ein Zertifikat für $\tau_3 = 12$

$$
\begin{aligned}
&145.5146532u^8v^8 + 988.2590039u^8v^7 + 363.0801883u^8v^6 - 672.8504875u^8v^5 - 225.3291719u^8v^4+\\
&191.5409281u^8v^3 + 59.48900200u^8v^2 - 19.01121967u^8v - 6.236063758u^8 + 988.2590039u^7v^8+\\
&2884.321650u^7v^7t + 2098.660249u^7v^7 + 3568.090777u^7v^6t + 18.16433328u^7v^6 + 344.9966795u^7v^5t-\\
&1553.436106u^7v^5 - 1392.615104u^7v^4t - 147.6093630u^7v^4 - 437.8480785u^7v^3t + 446.8628584u^7v^3+\\
&123.8174512u^7v^2t + 71.67183278u^7v^2 + 62.08524794u^7vt - 49.22376389u^7v - 6.539547051u^7t-\\
&6.216856357u^7 + 363.0801883u^6v^8 + 3568.090777u^6v^7t + 18.16433328u^6v^7 - 1165.053624u^6v^6t^2+\\
&5312.362070u^6v^6t - 1024.256466u^6v^6 - 71.15855319u^6v^5t^2 + 196.7871602u^6v^5t - 471.6846248u^6v^5+\\
&110.7356966u^6v^4t^2 - 2163.927390u^6v^4t + 464.1969368u^6v^4 - 78.18545046u^6v^3t^2-\\
&505.3867129u^6v^3t + 203.1645374u^6v^3 + 41.35985267u^6v^2t^2 + 199.8129603u^6v^2t - 54.08669500u^6v^2\\
&- 17.33067199u^6vt^2 + 77.24220247u^6vt - 21.52580028u^6v + 16.53555569u^6t^2 - 10.78477271u^6t+\\
&0.8926140522u^6 - 672.8504875u^5v^8 + 344.9966795u^5v^7t - 1553.436106u^5v^7 - 71.15855319u^5v^6t^2+\\
&196.7871602u^5v^6t - 471.6846248u^5v^6 + 1168.212790u^5v^5t^3 + 134.9381447u^5v^5t^2-\\
&599.4393672u^5v^5t + 752.0835370u^5v^5 + 386.8314957u^5v^4t^3 + 37.58791494u^5v^4t^2-\\
&219.0874465u^5v^4t + 274.0310180u^5v^4 - 377.6693697u^5v^3t^3 - 116.2455138u^5v^3t^2+\\
&135.0094523u^5v^3t - 132.6882963u^5v^3 - 29.72556641u^5v^2t^3 - 35.53845637u^5v^2t^2+\\
&37.13287924u^5v^2t - 50.62106262u^5v^2 - 55.38456003u^5vt^3 - 6.687980644u^5vt^2 + 13.24127784u^5vt+\\
&16.77621554u^5v + 47.57727922u^5t^3 + 37.98428531u^5t^2 - 14.54446445u^5t - 3.453914175u^5-\\
&+ 464.1969368u^4v^6 + 386.8314957u^4v^5t^3 + 37.58791494u^4v^5t^2 - 219.0874465u^4v^5t+\\
&274.0310180u^4v^5 - 73.36820000u^4v^4t^4 + 392.6271541u^4v^4t^3 + 222.2882726u^4v^4t^2+\\
&735.4659250u^4v^4t - 207.3205267u^4v^4 + 83.02988588u^4v^3t^4 - 118.0842510u^4v^3t^3-\\
&77.09075609u^4v^3t^2 + 244.0885057u^4v^3t - 88.19709904u^4v^3 - 37.62483034u^4v^2t^4-\\
&113.7068820u^4v^2t^3 - 8.326854722u^4v^2t^2 - 45.23470569u^4v^2t + 12.09797900u^4v^2-\\
&77.65166209u^4vt^4 - 55.08896234u^4vt^3 + 31.85022557u^4vt^2 - 17.02325455u^4vt + 10.36532594u^4v+\\
&28.47777799u^4t^4 + 46.28598839u^4t^3 - 2.917811290u^4t^2 - 7.358099446u^4t - 1.944166441u^4+\\
&- 29.72556641u^2v^5t^3 - 35.53845637u^2v^5t^2 + 37.13287924u^2v^5t - 50.62106262u^2v^5-\\
&37.62483034u^2v^4t^4 - 113.7068820u^2v^4t^3 - 8.326854722u^2v^4t^2 - 45.23470569u^2v^4t+\\
&12.09797900u^2v^4 + 2.757110052u^2v^3t^5 - 23.04651883u^2v^3t^4 + 34.97617188u^2v^3t^3+\\
&56.64805815u^2v^3t^2 - 36.40483748u^2v^3t + 4.218095336u^2v^3 - 31.31247252u^2v^2t^6+\\
&14.32494116u^2v^2t^5 + 13.18700960u^2v^2t^4 - 37.71556777u^2v^2t^3 - 39.18630288u^2v^2t^2+
\end{aligned}
$$

$$
\begin{aligned}
&+ 0.7941054727uv^3t + 5.569355446uv^3 + 0.3706844433uv^2t^6 - 1.226131739uv^2t^5+\\
&+ 4.448344730uv^2t^3 - 12.51651111uv^2t^2 + 2.559716007uv^2t + 0.2868981952uv^2+\\
&- 3.865608217uvt^6 - 18.70371050uvt^5 + 14.88704132uvt^4 + 23.94378706uvt^3+\\
&- 4.333409705uvt - 1.954761337uv - 0.1528426696ut^7 + 0.07806634840ut^6-\\
&2.244106275ut^4 + 1.817221500ut^3 + 2.535123200ut^2 - 0.5798954908ut - 0.3852239752u-\\
&6.539547051v^7t - 6.216856357v^7 + 16.53555569v^6t^2 - 10.78477271v^6t + 0.8926140522v^6\\
&t^3 + 37.98428531v^5t^2 - 14.54446445v^5t - 3.453914175v^5 + 28.47777799v^4t^4+\\
&- 2.917811290v^4t^2 - 7.358099446v^4t - 1.944166441v^4 + 7.486743724v^3t^5+\\
&- 17.88091824v^3t^3 - 20.64369889v^3t^2 + 5.781887281v^3t + 2.801486009v^3-\\
&- 2.282422116v^2t^5 - 2.376206451v^2t^4 - 5.756516944v^2t^3 + 2.285069305v^2t^2+\\
&- 0.1807740107v^2 - 0.1528426696vt^7 + 0.07806634840vt^6 - 1.066251935vt^5-\\
&1.817221500vt^3 + 2.535123200vt^2 - 0.5798954908vt - 0.3852239752v\\
&+ 0.6067608795t^6 + 2.504492501t^5 + 6.632189667t^4 + 3.925884209t^3 - 1.191716995t^2-\\
&168270210
\end{aligned}
$$

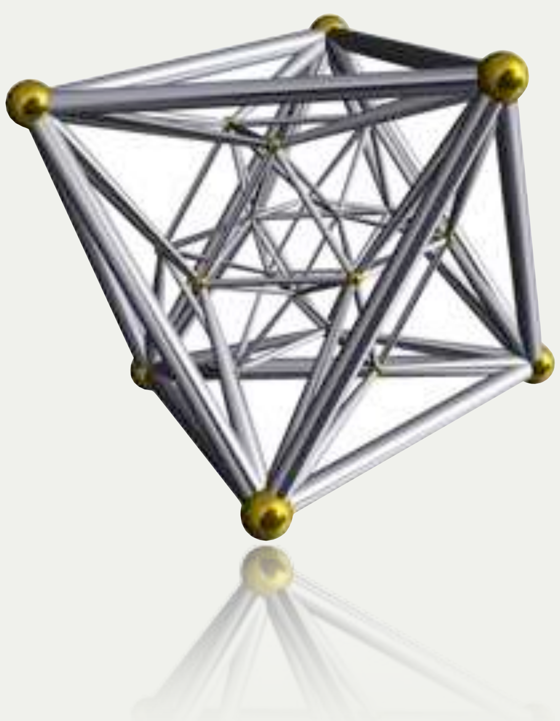
- ⋆ Gefunden mit semidefiniter Programmierung
- ⋆ Übrige Rechnungen sind relativ einfach

Bekannte Kusszahlen



$\tau_8 = 240$
$\tau_{24} = 196560$

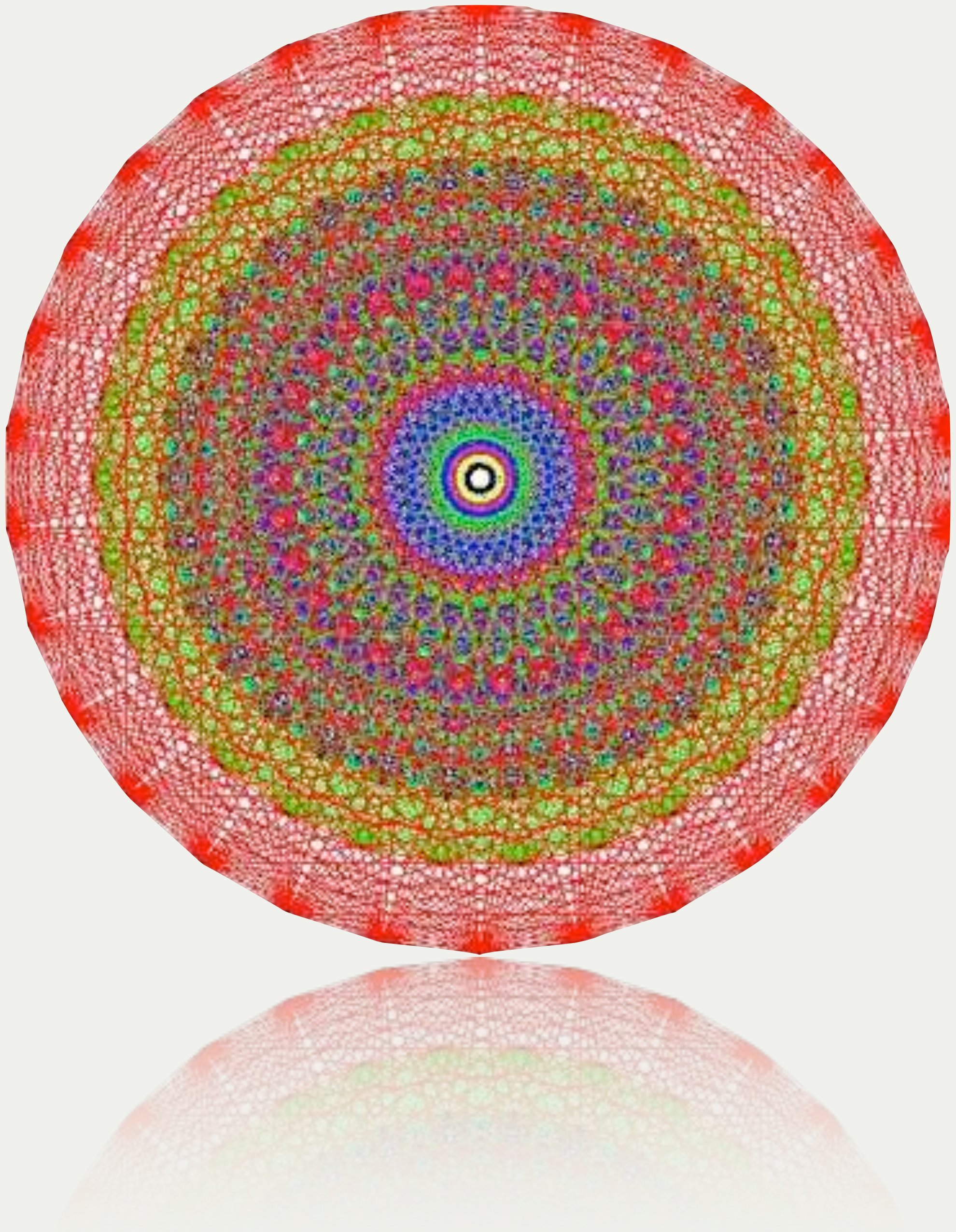
Odlyzko, Sloane und Levenshtein (1979)
mit linearer Programmierung
(relativ leicht, sehr elegant)



$\tau_4 = 24$

Musin (2008)
mit linearer Programmierung und Tricks
(nicht so leicht)

IV.1 Das Kusszahlproblem in Dimension 8

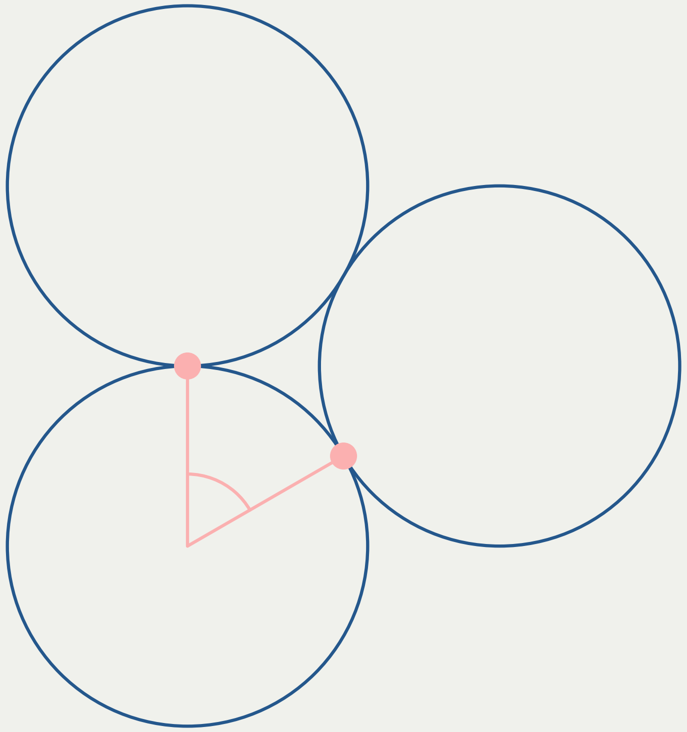


Packungsgraph

$$G(n,(t,1)) = (V,E)$$

$$V = S^{n-1} = \{x \in \mathbb{R}^n : x^\mathsf{T} x = 1\}$$

$$\{x,y\} \in E \iff x^\mathsf{T} y \in (t,1)$$



unabhängige Mengen in $G(n,(t,1))$
= mögliche Berührpunkte

Also: $\tau_n = \alpha(G(n,(t,1)))$

Berechnen von α ist i.A. schwierig

Die ϑ-Funktion …

$\star$ wurde 1979 von Lovász in der berühmten Arbeit
"On the Shannon capacity of a graph" eingeführt

Strategie: Verallgemeinere ϑ auf unendliche Graphen

$\star$ kann durch eine Hierarchie von größer werdenden SDPs,
die gegen α konvergieren, verbessert werden.
(Lovász-Schrijver 1991, Lasserre 2001)

Das Original: für $|V| < \infty$

$$\vartheta'(G) = \min \quad t \qquad Y \in \mathcal{S}^V_{\succeq 0} \qquad Y_{x,x} = t - 1 \text{ für } x \in V \qquad Y_{x,y} \le -1 \text{ für } \{x,y\} \notin E$$

Verallgemeinerung: für $|V| = \infty$

$$\vartheta'(G) = \inf \quad t \qquad Y \in \mathcal{C}(V \times V)_{\succeq 0} \qquad Y(x,x) = t - 1 \text{ für } x \in V \qquad Y(x,y) \le -1 \text{ für } \{x,y\} \notin E$$

positiv semidefinite, stetige Hilbert-Schmidt Kerne

$$\alpha(G) \leq \vartheta'(G) = \inf \quad \begin{array}{l} t \\ Y \in \mathcal{C}(V \times V)_{\succeq 0} \\ Y(x,x) = t-1 \text{ für } x \in V \\ Y(x,y) \leq -1 \text{ für } \{x,y\} \notin E \end{array}$$

Sei $C \subseteq V$ unabhängig

und Y, t eine zulässige Lösung von $\vartheta'(G)$

$$\begin{aligned} 0 &\leq \sum_{x \in C} \sum_{y \in C} Y(x,y) \\ &= \underbrace{\sum_{x \in C} Y(x,x)}_{=|C|(t-1)} + \underbrace{\sum_{x \neq y} Y(x,y)}_{\leq (-1)(|C|^2 - |C|)} \\ &\leq |C|(t-1) - (|C|^2 - |C|) \qquad \Longrightarrow |C| \leq t \end{aligned}$$

Frage: Wie berechnet man $\vartheta'(G)$?

Antwort: Verwende Symmetrie von G!

$\mathrm{Aut}(G(n,(t,1)))$: orthogonale Gruppe $\mathsf{O}(n)$

$$\mathsf{O}(n) = \{A \in \mathbb{R}^{n \times n} : A^{\mathsf{T}} A = I_n\}$$

Falls $Y \in \mathcal{C}(S^{n-1} \times S^{n-1})_{\succeq 0}$ zulässig, dann auch $\overline{Y}$:

$$\overline{Y}(x,y) = \int_{A \in \mathsf{O}(n)} Y(Ax, Ay)\, d\mu(A).$$

Also: Könnnen uns auf $\mathsf{O}(n)$-invariante Kerne beschränken.

$$\forall A \in \mathsf{O}(n)\, \forall x,y \in S^{n-1} : Y(Ax, Ay) = Y(x,y)$$

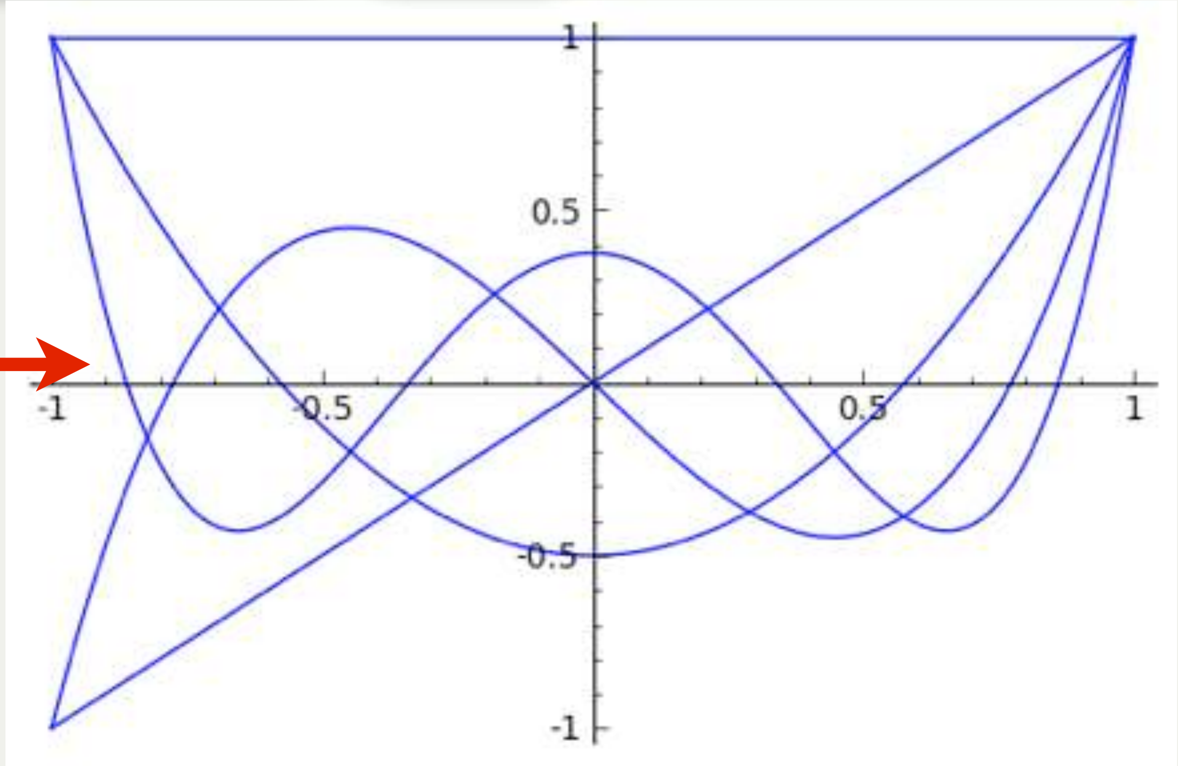
Theorem von Schoenberg (1942)

$Y \in \mathcal{C}(S^{n-1} \times S^{n-1})_{\succeq 0}$ ist $\mathsf{O}(n)$-invariant g.d.w.

$$Y(x, y) = \sum_{k=0}^{\infty} f_k P_k^{((n-3)/2,(n-3)/2)}(x^{\mathsf{T}} y)$$

$f_k \geq 0$

Jacobi Polynome



ϑ' als semi-infinites lineares Programm

$$
\begin{aligned}
\vartheta'(G(n,(t,1))) = \inf \quad & t \\
& f_0 \geq 0, f_1 \geq 0, \ldots \\
& \sum_{k=0}^{\infty} f_k P^{((n-3)/2,(n-3)/2)}(1) = t - 1 \\
& \sum_{k=0}^{\infty} f_k P^{((n-3)/2,(n-3)/2)}(u) \leq -1 \text{ für } u \in [-1, t]
\end{aligned}
$$

$$\vartheta'(G(8,(1/2,1))) = 240 = \tau_8$$
$$\vartheta'(G(24,(1/2,1))) = 196560 = \tau_{24}$$

ABER

$$\vartheta'(G(3,(1/2,1))) > 13$$
$$\vartheta'(G(4,(1/2,1))) > 24$$

Normalisierte Jacobi Polynome mit (α, α), $\alpha = (n-3)/2$

$$P_0^n(t) = P_0^{(\alpha,\alpha)}(t) = 1,$$

$$P_1^n(t) = P_1^{(\alpha,\alpha)}(t) = t,$$

$$P_2^n(t) = P_2^{(\alpha,\alpha)}(t) = \frac{n}{n-1}t^2 - \frac{1}{n-1}.$$

```
sage: x = PolynomialRing(QQ, 'x').gen()
sage: n = 4
sage: a = (n-3)/2
sage: for k in range(0,5):
sage:    print(jacobi_P(k,a,a,x)/jacobi_P(k,a,a,1))

1
x
4/3*x^2 - 1/3
2*x^3 - x
16/5*x^4 - 12/5*x^2 + 1/5
```

$\tau_8 \geq 240$

alle $\binom{8}{2}2^2 = 112$ Permutationen und Vorzeichenwechsel von

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, 0, 0\right)^\top$$

alle $2^7 = 128$ geraden Vorzeichenwechsel von

$$\left(\frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}\right)^\top$$

Mögliche Skalarprodukte:

$$\left\{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\right\}.$$

$\tau_8 \leq 240$

“Glücklicher” Ansatz

$$F(t) = -1 + \beta(t+1)\left(t+\frac{1}{2}\right)^2 t^2 \left(t-\frac{1}{2}\right)$$

Eigenschaften von F:

$$F(-1) = F(-1/2) = F(0) = F(1/2) = -1$$

$F(u) \leq -1$ für $u \in [-1, 1/2]$

Mit $\beta = \frac{320}{3}$: $F(1) = t - 1 = 240 - 1 = 239$

Überprüfe: $F(u) = \sum_{k=0}^{6} f_k P_k^8(u)$ mit $f_k \geq 0$

$\Longrightarrow Y(x,y) = F(x^{\mathsf{T}}y)$ zulässig für ϑ'

$\Longrightarrow \tau_8 \leq \vartheta'(G(8,(1/2,1))) \leq 240$

Kapitel V Determinantenmaximierung

§ 1 Minkowskis Determinantenungleichung

<u>Satz 1</u> Die Fkt.

$$F : S^n_{>0} \to \mathbb{R} \quad , \qquad X \mapsto -(\det X)^{1/n}$$

ist konvex. Zusätzlich gilt

$$F(\alpha X + (1-\alpha) Y) < \alpha F(X) + (1-\alpha) F(Y) \quad , \qquad \alpha \in (0,1),$$

falls X, Y lin. unabh. sind.

<u>Korollar 2</u> (Minkowski, 1905)

Für $X, Y \in S^n_{>0}$ gilt

$$(\det(X+Y))^{1/n} \geq (\det X)^{1/n} + (\det Y)^{1/n}.$$

<u>Bew.:</u> Verwende Satz 1 mit $\alpha = \frac{1}{2}$.

Zum Beweis von Satz 1 benötigen wir zwei Lemmata:

Lemma 3 Seien $X \in S^n_{\succ 0}$, $Y \in S^n$ gegeben. Dann gibt es eine Matrix $S \in GL_n(\mathbb{R})$, so dass

$$S^T X S \quad \text{und} \quad S^T Y S$$

beides Diagonalmatrizen sind.

Bew.: Spektralzerlegung von X:

$$\exists A \in O(n): \quad A^T X A = \operatorname{diag}(\lambda_1, \dots, \lambda_n).$$

Dann ist

$$\begin{bmatrix} 1/\sqrt{\lambda_1} & & \\ & \ddots & \\ & & 1/\sqrt{\lambda_n} \end{bmatrix}^T A^T X \underbrace{A \begin{bmatrix} 1/\sqrt{\lambda_1} & & \\ & \ddots & \\ & & 1/\sqrt{\lambda_n} \end{bmatrix}}_{T \in GL_n(\mathbb{R})} = I_n .$$

Spektralzerlegung von $T^T Y T$:

$$\exists B \in O(n): \quad B^T T^T Y T B = \operatorname{diag}(\mu_1, \dots, \mu_n).$$

Definiere $S = TB$.

Lemma 4 (AM-GM-Ungleichung)

a) Für $x_1, \dots, x_n \geq 0$ gilt

$$\left(\prod_{i=1}^n x_i \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i ,$$

wobei Gleichheit gilt gdw. $x_1 = \dots = x_n$.

[Das Volumen eines Parallelepiped mit Kantenlängensumme $\sum_{i=1}^{n} x_i = S$ ist maximal gdw. $x_1 = \dots = x_n = \frac{S}{n}$ gilt.].

b) Für $x_1, \dots, x_n > 0$, $y_1, \dots, y_n > 0$ gilt

$$\left(\prod_{i=1}^{n} x_i\right)^{1/n} + \left(\prod_{i=1}^{n} y_i\right)^{1/n} \leq \left(\prod_{i=1}^{n} (x_i + y_i)\right)^{1/n},$$

wobei Gleichheit gilt gdw. $(x_1, \dots, x_n)$, $(y_1, \dots, y_n)$ linear abhängig sind.

<u>Bew.:</u> b) folgt aus a): Aufgabe 11.2.

a) George Pólya (1887-1985) bester mathematischer Traum:

Es ist $1 + y \leq e^y$ für alle $y \in \mathbb{R}$. Also $y \leq e^{y-1}$ und $y^{\frac{1}{n}} \leq \left(e^{y-1}\right)^{1/n} = e^{\frac{y}{n} - \frac{1}{n}}$. Setze $y_i = \dfrac{x_i}{\frac{1}{n}\sum_j x_j}$. Dann

$$\left(\prod_{i=1}^{n} y_i\right)^{1/n} \leq e^{\frac{1}{n}\sum_{i=1}^{n} y_i - 1} = e^0 = 1.$$

Somit folgt die Beh, wobei $1 + y = e^y$ gdw. $y = 0$.

Bew.: (Satz 1)

Seien $X, Y \in S^n_{>0}$, $\alpha \in (0,1)$ gegeben. Falls X, Y lin. abh. sind, gilt

$$F(\alpha X + (1-\alpha) Y) = \alpha F(X) + (1-\alpha) F(Y),$$

wie man durch direktes Nachrechnen überprüft. Seien also X, Y lin. unabh. Nach Lemma 3 gibt es $S \in GL_n(\mathbb{R})$ mit

$$S^T X S = \operatorname{diag}(x_1, \dots, x_n), \quad S^T Y S = \operatorname{diag}(y_1, \dots, y_n).$$

Definiere $T = S^{-1}$. Dann ist

$$F(\alpha X + (1-\alpha) Y) = -\det\big(\alpha T^T \operatorname{diag}(x_1, \dots, x_n) T + (1-\alpha) T^T \operatorname{diag}(y_1, \dots, y_n) T\big)^{1/n}$$

$$= -\left(\det T^T T\right)^{1/n} \left(\prod_{i=1}^n (\alpha x_i + (1-\alpha) y_i)\right)^{1/n}$$

$$\overset{\text{AMGM}}{<} -\left(\det T^T T\right)^{1/n} \left(\left(\prod_{i=1}^n \alpha x_i\right)^{1/n} + \left(\prod_{i=1}^n (1-\alpha) y_i\right)^{1/n}\right)$$

$$= \alpha\left(-\det(T^T \operatorname{diag}(x_1, \dots, x_n) T)^{1/n}\right) + (1-\alpha)\left(-\det(T^T \operatorname{diag}(y_1, \dots, y_n) T)^{1/n}\right)$$

$$= \alpha F(X) + (1-\alpha) F(Y). \qquad \boxtimes$$

§ 2 Konvexe, spektrale Funktionen

Die Fkt. $F(X) = -(\det X)^{1/n}$ ist konvex und hängt nur von den Eigenwerten von X ab (d.h. $F(A^T X A) = F(X)\ \forall A \in O(n)$).

Ziel: Charakterisierung all solcher Fkt.

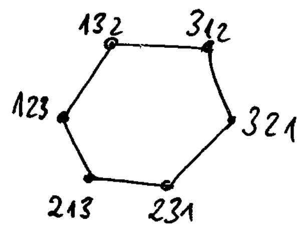
Def. 1: a) Das Schur-Horn Orbitop von $X \in S^n$ ist

$$SH(X) = \operatorname{conv}\{AXA^T : A \in O(n)\} \subseteq S^n.$$

b) Das Permutaeder von $(x_1, \dots, x_n) \in \mathbb{R}^n$ ist

$$\Pi(x_1, \dots, x_n) = \operatorname{conv}\{(x_{\sigma(1)}, \dots, x_{\sigma(n)}) : \sigma \in S_n\}.$$

Bsp.: $\Pi(1,2,3)$ ist ein Sechseck.



Satz 2 Sei $X \in S^n$ mit Eigenwerten $\lambda_1, \dots, \lambda_n$. Betrachte Projektion $\operatorname{Diag} : S^n \to \mathbb{R}^n$, $\operatorname{Diag}(Y) = (Y_{11}, \dots, Y_{nn})$.

Dann gilt

$$\operatorname{Diag}(SH(X)) = \Pi(\lambda_1, \dots, \lambda_n).$$

Bew.: „$\supseteq$": Das Bild $\mathrm{Diag}(SH(X))$ ist konvex, weil Diag linear ist. Außerdem enthält das Bild die Vektoren $(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)})$, $\sigma \in S_n$ (Warum?). Da das Permutaeder $\Pi(\lambda_1, \ldots, \lambda_n)$ die kleinste konvexe Menge ist, die diese Vektoren enthält, folgt die Inklusion.

„$\subseteq$": Spektralzerlegung $X = \sum_{j=1}^{n} \lambda_j u_j u_j^T$.

Sei $A \in O(n)$. Setze $Y = AXA^T$. Dann ist

$$Y_{ii} = e_i^T Y e_i = e_i^T \sum_{j=1}^{n} \lambda_j A u_j u_j^T A^T e_i$$

$$= \sum_{j=1}^{n} \lambda_j \underbrace{\left(e_i^T A u_j\right)^2}_{=: S_{ij}}.$$

Beh.: Die Matrix $(S_{ij})_{1 \le i,j \le n}$ ist doppelt stochastisch.

Bew.:

- $S_{ij} \geq 0$: ✓
- $\sum_{i=1}^{n} S_{ij} = 1$: $\sum_{i=1}^{n} \left(e_i^T A u_j\right)^2 = \| A u_j \|^2 = 1$.
- $\sum_{j=1}^{n} S_{ij} = 1$: genauso.

Nach dem Theorem von Birkhoff und von Neumann (Matchingpolytop des vollständigen bipartiten Graphen $K_{n,n}$) gilt:

$$\{ S \in \mathbb{R}^{n\times n} : S \text{ doppelt-stochastisch} \} = \operatorname{conv} \{ P^\sigma : \sigma \in S_n \},$$

P^σ Permutationsmatrix, $\quad P^\sigma_{ij} = \begin{cases} 1, & \text{falls } i = \sigma(j) \\ 0, & \text{sonst.} \end{cases}$

Also gibt es Koeffizienten $\alpha_\sigma \geq 0$, $\sum\limits_{\sigma \in S_n} \alpha_\sigma = 1$ mit

$S = \sum\limits_{\sigma} \alpha_\sigma P^\sigma$. Einsetzen liefert:

$$\begin{aligned} Y_{ii} &= \sum_{j=1}^{n} \lambda_j \sum_{\sigma \in S_n} \alpha_\sigma (P^\sigma)_{ij} \\ &= \sum_{\sigma \in S_n} \cancel{\lambda_j}\, \alpha_\sigma \sum_{j=1}^{n} \lambda_j (P^\sigma)_{ij} \\ &= \sum_{\sigma \in S_n} \alpha_\sigma \lambda_{\sigma^{-1}(i)}, \end{aligned}$$

also $(Y_{11}, \dots, Y_{nn}) \in \Pi(\lambda_1, \dots, \lambda_n)$. $\boxtimes$

Def. 3 Eine Fkt. $F: S^n \to \mathbb{R} \cup \{\infty\}$ heißt *spektral*, falls $F(X)$ nur von den Eigenwerten $\lambda_1, \dots, \lambda_n$ von X abhängt. D. h. $F(X) = F(AXA^T)$ für alle $A \in \mathcal{O}(n)$ gilt.

Eine spektrale Fkt. F definiert eine symmetrische Fkt. $f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ durch $F(X) = f(\lambda_1, \dots, \lambda_n)$, wobei f invariant unter Permutation der Koordinaten ist.

Satz 4 (Davis 1957)

$F: S^n \to \mathbb{R} \cup \{\infty\}$ ist konvex und spektral $\Longleftrightarrow$ $f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ ist konvex und symmetrisch.

Bew.: „$\Longrightarrow$" f symmetrisch: ✓

f konvex: Sei $X = \operatorname{diag}(x_1, \dots, x_n)$, $Y = \operatorname{diag}(y_1, \dots, y_n)$, $\alpha \in [0,1]$. Dann

$$\begin{aligned} & f(\alpha(x_1, \dots, x_n) + (1-\alpha)(y_1, \dots, y_n)) \\ = {} & F(\alpha X + (1-\alpha) Y) \\ \le {} & \alpha F(X) + (1-\alpha) F(Y) \\ = {} & \alpha f(x_1, \dots, x_n) + (1-\alpha) f(y_1, \dots, y_n). \end{aligned}$$

„$\Leftarrow$": Zeige, dass

$$F(X) = \max_{A \in O(n)} f(\mathrm{Diag}(AXA^T)) \qquad (*)$$

gilt. Dann ist F konvex, weil es das Maximum der Familie von konvexen Fkt.

$$g_A : S^n \to \mathbb{R}, \quad g_A(X) = f(\mathrm{Diag}(AXA^T))$$

ist.

„$\leq$": in $(*)$. Spektralzerlegung $X = \sum_i \lambda_i u_i u_i^T$.

Definiere $A = \begin{bmatrix} - u_1 - \\ \vdots \\ - u_n - \end{bmatrix} \in O(n)$. Dann ist

$$F(X) = f(\lambda_1, \dots, \lambda_n) = f(\mathrm{Diag}(AXA^T)).$$

„$\geq$": in $(*)$

$$\begin{aligned}
\max_{A \in O(n)} f(\mathrm{Diag}(AXA^T)) &\leq \max_{Y \in SH(X)} f(\mathrm{Diag}(Y)) \\
&\overset{\text{Satz 2}}{=} \max_{(x_1, \dots, x_n) \in \Pi(\lambda_1, \dots, \lambda_n)} f(x_1, \dots, x_n) \\
&\overset{f \text{ sym. + konv.}}{=} f(\lambda_1, \dots, \lambda_n) \\
&= F(X).
\end{aligned}$$

$\boxtimes$

Korollar 5 $F(X) = \begin{cases} -(\det X)^{1/n}, & \text{falls } X \succeq 0 \\ \infty, & \text{sonst} \end{cases}$

ist konvex und spektral.

Bew.: Es genügt z.z., dass $f(x_1, \dots, x_n) = -\left(\prod_{i=1}^{n} x_i\right)^{1/n}$ symmetrisch und konvex ist.

Symmetrie: ✓

Konvexität: AM-GM-Ungleichung.

§3 MAXDET Probleme

Satz 1 Die Menge

$$\mathcal{D}^n = \{ (X, s) \in S^n \times \mathbb{R} : X \succeq 0, (\det X)^{1/n} \geq s \}$$

ist ein ordentlicher konvexer Kegel.

Bew.: * $\mathcal{D}^n$ ist konvexer Kegel:

Seien $\alpha \geq 0$, $(X, s), (Y, t) \in \mathcal{D}^n$. Dann

$(\alpha X, \alpha s) \in \mathcal{D}^n$, weil $\alpha X \succeq 0$ und $(\det \alpha X)^{1/n} =$

$= \alpha (\det X)^{1/n} \geq \alpha s.$

$(X+Y, s+t) \in \mathcal{D}^n$, weil

$$\left(\det(X+Y)\right)^{1/n} \overset{\text{Minkowski}}{\geq} (\det X)^{1/n} + (\det Y)^{1/n} \geq s+t.$$

* $\mathcal{D}^n$ ist spitz: Ang. $(X,s), (-X,-s) \in \mathcal{D}^n$.

  Weil $S^n_{\geq 0}$ spitz ist, ist $X=0$. Desweiteren

  $0 = (\det X)^{1/n} \geq s$ und $0 = (\det(-X))^{1/n} \geq -s$.

  Also $s=0$.

* $\mathcal{D}^n$ hat nicht-leeren Inneren: Betrachte eine offene Umgebung von $(I_n, 1)$.

* $\mathcal{D}^n$ ist abgeschlossen: ✓ weil det stetig. ⊠

<u>Def. 2</u> <u>primales MAXDET Problem</u>

$$p^* = \sup \langle (C,c), (X,s) \rangle$$
$$(X,s) \in \mathcal{D}^n$$
$$\langle (A_j, a_j), (X,s) \rangle = b_j, \quad j \in [m].$$

Falls $(C,c) = (0,1)$ und $a_j = 0$ vereinfacht sich das zu:

$$\begin{array}{ll} \sup & s \\ & X \succeq 0,\ (\det X)^{1/n} \geq s \\ & \langle A_j, X \rangle = b_j,\ j \in [m] \end{array} \iff \begin{array}{ll} \sup & (\det X)^{1/n} \\ & X \succeq 0 \\ & \langle A_j, X\rangle = b_j. \end{array}$$

Satz 3

$$(\mathcal{D}^n)^* = \{ (Y,t) \in \mathcal{S}^n \times \mathbb{R} : \ Y \succeq 0,\ (\det Y)^{1/n} \geq -\frac{t}{n} \}.$$

Lemma 4 Sei $X \in \mathcal{S}^n_{\succeq 0}$. Dann

$(*)$ $\operatorname{Tr}(X) - n(\det X)^{1/n} \geq 0$,

mit Gleichheit gdw. X ein Vielfaches von I_n.

Bew.: OBdA $X = \operatorname{diag}(x_1, \dots, x_n)$. Dann

$(*) \iff \sum_{i=1}^n x_i - n\left(\prod_{i=1}^n x_i\right)^{1/n} \geq 0$ (AM-GM Ungleichung).

Bew.: (Satz 3)

„$\supseteq$": Sei (Y,t) mit $Y \in \mathcal{S}^n_{\succeq 0}$ und $(\det Y)^{1/n} \geq -\frac{t}{n}$.

Sei $(X,s) \in \mathcal{D}^n$. Dann ist

$$\langle (X,s), (Y,t) \rangle = \operatorname{Tr}(XY) + st.$$

Möchte Lemma 4 anwenden. <u>Problem</u>: XY evtl. nicht symmetrisch. <u>Ausweg</u>: Cholesky-Zerlegung $X = LL^T$.

Dann $\mathrm{Tr}(XY) = \mathrm{Tr}(LL^TY) = \mathrm{Tr}(L^TYL)$ und

$L^TYL \succeq 0$.

Also

$$\begin{aligned}
\mathrm{Tr}(XY) + st &= \mathrm{Tr}(L^TYL) + st \\
&\geq n \det(L^TYL)^{1/n} + st \\
&= n \det(L^TL)^{1/n} \det(Y)^{1/n} + st \\
&\geq n\left(-\frac{t}{n}\right) s + st \\
&= 0.
\end{aligned}$$

<u>„$\subseteq$"</u>: Sei $(Y,t) \in (\mathcal{D}^n)^*$. Dann ist für $(X,0) \in \mathcal{D}^n$:

$$0 \leq \langle X,Y\rangle + t \cdot 0 = \langle X,Y\rangle.$$

Weil $(S^n_{\succeq 0})^* = S^n_{\succeq 0}$ folgt $Y \succeq 0$.

Nun sei $(X,s) \in \mathcal{D}^n$ beliebig. Dann, mit $s = (\det X)^{1/n}$ und $X \succ 0$

$$0 \leq \mathrm{Tr}(XY) + (\det X)^{1/n} \cdot t \implies -t \leq \frac{\mathrm{Tr}(XY)}{(\det X)^{1/n}}.$$

Minimiere die rechte Seite $\frac{\mathrm{Tr}(XY)}{(\det X)^{1/n}}$. (in Abhängigkeit von X)

Falls Y positiv semidefinit, aber nicht definit, dann ist das Infimum 0. (Warum?)

Falls Y pos. definit, dann ist das Minimum $(\det Y)^{1/n} \cdot n$, und wird bei $X = Y^{-1}$ angenommen (AM-GM-Ungleichung).

Also
$$-t \leq n\,(\det Y)^{1/n},$$
wie gewünscht. $\boxtimes$

<u>Def. 5</u> <u>duales MAXDET Problem</u>

$$d^* = \inf \sum_{j=1}^{m} y_j b_j$$

$$\sum_{j=1}^{m} y_j (A_j, a_j) - (C, c) \in (\mathcal{D}^n)^* .$$

Das Dualitätstheorem I.5.2 gilt auch hier.

Besonders hilfreich: Optimalitätskriterium für MAXDET.

Satz 6 Ang. $p^* = d^*$ und ang. (X, s) ist zulässige Lösung für primale MAXDET und $(y_1, \ldots, y_m)$ ist zulässig für duales MAXDET.

(X, s), $(y_1, \ldots, y_m)$ sind beide optimal $\iff$

i) $\left(\sum y_j A_j - C\right)^{-1} = X$

ii) $s = (\det X)^{1/n}$

iii) $-\dfrac{\sum y_j a_j - c}{n} = (\det Y)^{1/n}$.

Bew.: folgt aus Gleichheitsbed. von Lemma 4. & Thm. I.5.2

$$0 = \langle (X, s), \left(\sum y_j A_j - C, \sum y_j a_j - c\right)\rangle$$

$$= \operatorname{Tr}\left(X \left(\sum y_j A_j - C\right)\right) + s\left(\sum y_j a_j - c\right)$$

$$\geq n (\det X)^{1/n} \left(\det\left(\sum y_j A_j - C\right)\right)^{1/n} + s\left(\sum y_j a_j - c\right)$$

$$\geq n\, s \left(-\left(\frac{\sum y_j a_j - c}{n}\right)\right) + s\left(\sum y_j a_j - c\right)$$

$$= 0. \qquad \boxtimes$$

§ 4 Approximation von Polytopen durch Ellipsoide

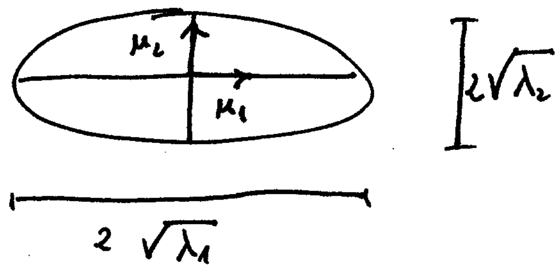
Def. 1 $A \in S^n_{>0}$, $x \in \mathbb{R}^n$ definieren das <u>Ellipsoid</u>

$$\mathcal{E}(A,x) = \{ y \in \mathbb{R}^n : (y-x)^T A^{-1} (y-x) \leq 1 \}.$$

Bsp.: $\mathcal{E}(r^2 I_n, 0) = r B_n$, $\quad B_n = \{ y \in \mathbb{R}^n : \|y\| \leq 1 \}$

n-dim. Einheitskugel

Eigenschaften: Spektralzerlegung $A = \sum_{i=1}^n \lambda_i u_i u_i^T$



u_i = Hauptachsenrichtungen

$$\mathrm{vol}\, \mathcal{E}(A,x) = \sqrt{\det A}\; \underbrace{\mathrm{vol}\, B_n}_{= \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}}.$$

Ellipsoid als affin linearen Bild der Einheitskugel:

$$\mathcal{E}(A^2, x) = \{ Ay + x : y \in B_n \}.$$

Satz 2 (innere Approximation)

Das Ellipsoid $\mathcal{E}(A^2, x)$ ist in dem Polytop

$$P = \{ y \in \mathbb{R}^n : a_1^T y \leq b_1, \dots, a_m^T y \leq b_m \}$$

enthalten (d.h. $\mathcal{E}(A^2, x) \subseteq P$) gdw.

$$\| A a_j \| \leq b_j - a_j^T x \qquad \forall j \in [m].$$

Bew.:

$$\forall y : \|y\| \leq 1 \qquad a_j^T (Ay + x) \leq b_j \qquad j = 1, \dots, m$$

$$\iff \max_{\substack{y \in \mathbb{R}^n \\ \|y\| \leq 1}} a_j^T A y \leq b_j - a_j^T x \quad , \quad j = 1, \dots, m.$$

Nach Cauchy-Schwarz ist das Maximum gleich

$$\max_{\|y\| \leq 1} (Aa_j)^T y = (Aa_j)^T \frac{Aa_j}{\|Aa_j\|} = \| Aa_j \| ,$$

und die Beh. folgt.

Satz 3 (äußere Approximation)

Das Ellipsoid $\mathcal{E}(A,x)$ enthält das Polytop

$P = \operatorname{conv}\{x_1,\dots,x_N\}$ (d.h. $P \subseteq \mathcal{E}(A,x)$).

gdw. die Matrix $\begin{pmatrix} s & d^T \\ d & A^{-1} \end{pmatrix} \in S^{n+1}_{\succeq 0}$ mit $d = A^{-1}x$

und $$x_i^T A^{-1} x_i - 2x_i^T d + s \le 1 \qquad \forall\, i \in [N].$$

Bew.: $x_i \in \mathcal{E}(A,x)$

$$\begin{aligned}
&\iff (x_i - x)^T A^{-1} (x_i - x) \le 1 \\
&\iff x_i^T A^{-1} x_i - 2x_i^T A^{-1} x + x^T A^{-1} x \le 1 \qquad [d = A^{-1}x] \\
&\iff x_i^T A^{-1} x_i - 2x_i^T d + d^T A d \le 1 \\
&\iff x_i^T A^{-1} x_i - 2x_i^T d + s \le 1 \quad \text{mit } s \ge d^T A d.
\end{aligned}$$

Weil A pos. def. ist die Bed. $s \ge d^T A d$ nach dem Schurkomplement äquivalent zu $\begin{bmatrix} s & d^T \\ d & A^{-1} \end{bmatrix} \succeq 0$, und die Beh. folgt.

Beste innere Approximation:

Problem (I): geg.: Polytop $P = \{y \in \mathbb{R}^n : a_1^T y \le b_1, \dots, a_m^T y \le b_m\}$.

ges.: Ellipsoid $\mathcal{E}(A^2, x)$ mit $P \subseteq \mathcal{E}(A^2, x)$ und $\operatorname{vol} \mathcal{E}(A^2, x)$ maximal.

Formulierung als konisches Programm:

$$\max s$$
$$(A, s) \in \mathcal{D}^n, \quad x \in \mathbb{R}^n$$
$$\|A a_j\| \le b_j - a_j^T x, \quad j \in [m]$$

$$= \max s$$
$$(A, s) \in \mathcal{D}^n, \quad x \in \mathbb{R}^n, \quad (y_j, t_j) \in \mathcal{L}^{n+1}, \quad j \in [m]$$
$$y_j = A a_j, \quad t_j = b_j - a_j^T x, \quad j \in [m]$$

Problem (I) hat eindeutige Lösung, weil die Fkt. $F(X) = -(\det X)^{1/n}$ strikt konvex ist entlang von Strecken $[X, Y]$ mit $X \ne \alpha Y$. (siehe Satz 1.1).

Somit ist auch das Ellipsoid eindeutig bestimmt.

Def. 4 a) Sei P ein Polytop. Das Ellipsoid E mit $E \subseteq P$, das das größte Volumen mit dieser Eigenschaft hat, heißt Loewner-John-Ellipsoid $\mathcal{E}_{in}(P)$.

b) Ellipsoid E mit $P \subseteq E$ und minimalem Volumen heißt Loewner-John Ellipsoid $\mathcal{E}_{out}(P)$.

Beste äußere Approximation

Problem (Ä): geg.: Polytop $P = \operatorname{conv}\{x_1, \dots, x_N\}$.

ges.: Ellipsoid $\mathcal{E}(A,x)$ mit $P \subseteq \mathcal{E}(A,x)$ und $\operatorname{vol}(\mathcal{E}(A,x))$ minimal.

Formulierung als konisches Programm:

$$\max \left(\det A^{-1}\right)^{1/n}$$

$$\begin{pmatrix} s & d^T \\ d & A^{-1} \end{pmatrix} \in \mathcal{S}^{n+1}_{\succeq 0}$$

$$\cancel{x_i^T A^{-1}} \quad x_i^T A^{-1} x_i - 2 x_i^T d + s \leq 1 \,, \quad i \in [N].$$

$$= \max t$$

$$(B,t) \in \mathcal{D}^n, \quad \begin{pmatrix} s & d^T \\ d & B \end{pmatrix} \in S^{n+1}_{\geq 0}$$

$$s_1, \dots, s_N \geq 0$$

$$x_i^T B x - 2 x_i^T d + s + s_i = 1, \qquad i \in [N]$$

$$= \max t$$

$$(B,t) \in \mathcal{D}^n, \; Y \in S^{n+1}_{\geq 0}, \quad s \in \mathbb{R}^N_{\geq 0}$$

$$\langle E_{ij}, B \rangle + \langle -E_{i+1,j+1}, Y \rangle = 0 \qquad 1 \leq i \leq j \leq n$$

$$\left\langle \begin{bmatrix} 1 & -x_i^T \\ -x_i & x_i x_i^T \end{bmatrix}, Y \right\rangle + s_i = 1 \quad , \qquad i \in [N].$$

duale MAXDET Programm:

$$\min \sum_{i=1}^{N} y_i$$

$$\left(\sum z_{ij} E_{ij}, -1\right) \in (\mathcal{D}^n)^*$$

$$-\sum z_{ij} E_{i+1,j+1} + \sum y_i \begin{bmatrix} 1 & -x_i^T \\ x_i & x_i x_i^T \end{bmatrix} \succeq 0,$$

$$y_i \geq 0, \quad i \in [N].$$

Beide Programme strikt zulässig $\Rightarrow$ starke Dualität.

Satz 5 (Johns Optimalitätskriterium (1948))

[Historisch die erste Optimalitätsbed. eines nichtlinearen Optimierungsproblems].

Sei $P \subseteq \mathbb{R}^n$ ein Polytop. Es gilt

a) $\mathcal{E}_{out}(P) = B_n \iff P \subseteq B_n$,

$\exists\, \lambda_1, \dots, \lambda_M > 0,\ x_1, \dots, x_M$ Ecken von P:

(i) $\|x_i\| = 1, \quad i \in [M]$

(ii) $\sum_{i=1}^{M} \lambda_i x_i = 0$

(iii) $\sum_{i=1}^{M} \lambda_i x_i x_i^T = I_n$

b) $\mathcal{E}_{in}(P) = B_n \iff B_n \subseteq P$,

$\exists\, \lambda_1, \dots, \lambda_M > 0,\ x_1, \dots, x_M \in \partial P$:

(i)–(iii) wie oben.

Bem.: * Dass das Loewner-John Ellipsoid gleich B_n ist, kann man immer durch eine affine Transformation von P erreichen.

* Bed. (ii): Nicht alle Vektoren $x_1, \dots, x_M$ liegen auf einer Seite von S^{n-1}, z.B.  ist nicht möglich.

* Bed. (iii): $x_1, \dots, x_M$ verhalten sich wie eine Orthonormalbasis

$$x^T y = x^T I_n y = \sum_{i=1}^{M} \lambda_i (x^T x_i)(x_i^T y).$$

* Bed. (ii) & (iii): $M \geq n+1$.

<u>Bew.</u>: b) folgt aus a) durch Übergang zum polaren Polytop.

$$P^* = \{ y \in \mathbb{R}^n : x^T y \leq 1 \quad \forall x \in P \}$$

a) <u>„$\Leftarrow$"</u> Anwendung schwacher Dualität

Es ist $(y_1, \dots, y_M, y_{M+1}, \dots, y_N) = \left(\frac{\lambda_1}{n}, \dots, \frac{\lambda_M}{n}, 0, \dots, 0\right)$,

$$\sum z_{ij} E_{ij} = \frac{1}{n} I_n$$

eine zulässige Lösungen des dualen Programms.

Weil $B_n \supseteq P$ ist $(B, t) = (I_n, 1)$, $Y = \begin{bmatrix} 0 & 0 \\ 0 & I_n \end{bmatrix}$

eine zulässige Lösung des primalen Programms. Nach Satz 3.6 folgt nun die Beh.

"$\Longrightarrow$" Anwendung starker Dualität.

Klar: $P \subseteq B_n$

Sei $y_1^*, \dots, y_N^*$ optimale Lösung vom Dualen.

O.B.d.A. $y_1^* > 0, \dots, y_M^* > 0$, $y_{M+1}^* = 0, \dots, y_N^* = 0$.

Dann ist wegen der komplementären Schlupfes

$$\left\langle \begin{bmatrix} 1 & -x_i^T \\ -x_i & x_i x_i^T \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & I_n \end{bmatrix} \right\rangle = 1 \quad \text{für } i \in [M],$$

d.h. $\|x_i\| = 1$ und (i) folgt.

Sei $Z^* = \sum z_{ij}^* E_{ij}$ optimale Lösung des Dualen. Dann ist $I_n Z^* = \alpha I_n$ für ein $\alpha > 0$, also $Z^* = \frac{1}{n} I_n$.

Sei Y^* optimal für das Primale. Dann

$$Y^* \left(- \sum z_{ij}^* E_{i+1,j+1} + \sum y_i^* \begin{bmatrix} 1 & -x_i^T \\ -x_i & x_i x_i^T \end{bmatrix} \right) = 0$$

$$\Longleftrightarrow \sum y_i^* \begin{bmatrix} 1 & -x_i^T \\ -x_i & x_i x_i^T \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{n} I_n \end{bmatrix}$$

Definiere $\lambda_i = n y_i^*$ und (ii) & (iii) folgen aus der obigen Gleichung.

$\boxtimes$

Korollar 6 Sei $P \subseteq \mathbb{R}^n$ ein Polytop.

(a) Falls $\mathcal{E}_{out}(P) = B_n$, dann $\frac{1}{n} B_n \subseteq P \subseteq B_n$.

(b) Falls $\mathcal{E}_{in}(P) = B_n$, dann $B_n \subseteq P \subseteq n B_n$.

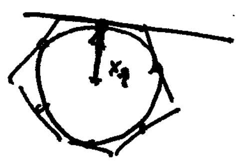
[Beide Inklusionen sind bestmöglich → Blatt 12].

Bew.: (a) folgt aus (b) durch Polarität.

(b) Anwendung von Johns Opt.-bed.

$$\sum_{i=1}^{M} \lambda_i x_i = 0, \quad \sum_{i=1}^{M} \lambda_i x_i x_i^T = I_n, \quad \lambda_i > 0.$$

Die stützenden Hyperebenen durch $x_i \in B_n \cap \partial P$ sind orthogonal zu x_i:



Also $B_n \subseteq P \subseteq Q = \{ x \in \mathbb{R}^n : \underset{i \in [M]}{x_i^T x \leq 1,} \}$.

Falls $x \in Q$, dann $x^T x_i \in [-\|x\|, 1]$. Also

$$0 \leq \sum_{i=1}^{M} \lambda_i (1 - x^T x_i)(\|x\| + x^T x_i)$$
$$= \|x\| \sum_{i=1}^{M} \lambda_i + (1 - \|x\|) \sum_{i=1}^{M} \lambda_i x^T x_i - \sum_{i=1}^{M} \lambda_i (x^T x_i)^2$$
$$= \|x\| n + 0 - \|x\|^2 \qquad \left(\text{weil } \sum \lambda_i = \sum \lambda_i x_i^T x_i = \operatorname{Tr}\left(\sum \lambda_i x_i x_i^T\right) = \operatorname{Tr}(I_n) = n\right)$$

$\Rightarrow \|x\| \leq n.$ ⊠

Kapitel VI Polynomielle Optimierung

§1 Nichtnegative Polynome und SOS

SOS = sum of squares („Quadratsummen")

Notation Polynome in n Unbestimmten $x_1, \dots, x_n$.

- $x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdots x_n^{\alpha_n}$, $\alpha_1, \dots, \alpha_n \in \mathbb{N}$ heißt Monom.
- Kurzschreibweise: x^α, $\alpha = (\alpha_1, \dots, \alpha_n)$
- Grad von x^α: $|\alpha| = \sum_{i=1}^n \alpha_i$
- $p = \sum_\alpha p_\alpha x^\alpha$ endl. Linearkomb. von Monomen x^α mit Koeffizienten $p_\alpha \in \mathbb{R}$ heißt Polynom
- $\mathbb{R}[x_1, \dots, x_n] = \mathbb{R}[x] = \{p : p \text{ Polynom}\}$ Polynomring
- $\deg p = \max\{|\alpha| : p_\alpha \neq 0\}$ Grad von p
- $\mathbb{R}[x]_{=d} = \{p \in \mathbb{R}[x] : \deg p = d,\ p \text{ homogen}\} \cup \{0\}$

  p homogen: falls $p_\alpha \neq 0$, dann $|\alpha| = d$.

  [oder $p(\beta x) = \beta^d p(x)$ $\forall \beta \in \mathbb{R}$ $\forall x \in \mathbb{R}^n$]

Monome vom Grad d sind Basis von $\mathbb{R}[x]_{=d}$.

$$\dim \mathbb{R}[x]_{=d} = \binom{n+d-1}{n-1} = \binom{n+d-1}{d}$$

z.B. $n=3,\ d=4$ $\qquad \binom{n+d-1}{n-1} = \binom{6}{2}$

$$\begin{matrix} \bullet & \bullet & * & \bullet & * & \bullet \\ x_1 & x_1 & & x_2 & & x_3 \end{matrix} \qquad \text{Monom } x_1^2 x_2 x_3$$

- $\mathbb{R}[x]_d = \{ p : p \text{ Polynom, } \deg p \leq d \}$

$$= \mathbb{R}[x]_{=0} \oplus \mathbb{R}[x]_{=1} \oplus \cdots \oplus \mathbb{R}[x]_{=d}.$$

$$\dim \mathbb{R}[x]_d = \sum_{k=0}^{d} \binom{n+k-1}{k} = \binom{n+d}{d}$$

Def. 1 a) $p \in \mathbb{R}[x]$ heißt SOS, falls es $q_1, \ldots, q_m \in \mathbb{R}[x]$ gibt mit $p = q_1^2 + \cdots + q_m^2$.

b) $\Sigma_{n,d} = \{ p \in \mathbb{R}[x]_d : p \text{ ist SOS} \}$ SOS-Kegel
($\longrightarrow$ Aufgabe 4.2)

c) $P_{n,d} = \{ p \in \mathbb{R}[x] : p(x) \geq 0 \ \forall x \in \mathbb{R}^n \}$ Kegel der nicht-negativen Polynome ($\rightarrow$ Aufgabe 1.2)

Klar: $\Sigma_{n,2d+1} = \Sigma_{n,2d}$, $P_{n,2d+1} = P_{n,2d}$

$$\Sigma_{n,d} \subseteq P_{n,d}.$$

Satz 2 $\Sigma_{n,2d} = \{ p \in \mathbb{R}[x]_{2d} : \exists\, Q \in S_{\geq 0}^{\binom{n+d}{d}} : p = [x]_d^T Q [x]_d \}$,

wobei $[x]_d$ der Vektor aller Monome x^α in $\mathbb{R}[x]_d$ ist.

Bsp.: $n=2$, $d=2$ $\quad [x]_2 = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}$

Bew.: „$\subseteq$": Sei $p = q_1^2 + \dots + q_m^2$.

Schreibe $q_i = \sum_\alpha q_{i\alpha} x^\alpha = \underline{q_i}^T [x]_d$,

wobei $\underline{q_i} = (q_{i\alpha})_\alpha \in \mathbb{R}^{\binom{n+d}{d}}$

Dann $q_i^2 = (\underline{q_i}^T [x]_d)^2$ und

$$p = \sum_{i=1}^m [x]_d^T \underbrace{\underline{q_i}\,\underline{q_i}^T}_{\in S_{\geq 0}^{\binom{n+d}{d}}} [x]_d = [x]_d^T Q [x]_d,$$

wobei $Q = \sum_{i=1}^m \underline{q_i}\,\underline{q_i}^T \in S_{\geq 0}^{\binom{n+d}{d}}$.

„$\supseteq$": Sei $p = [x]_d^T Q [x]_d$ mit $Q \succeq 0$.

Spektralzerlegung: $Q = \sum_{i=1}^{\binom{n+d}{d}} \lambda_i u_i u_i^T$.

Dann
$$p = \sum_{i=1}^{\binom{n+d}{d}} \Big(\underbrace{(\sqrt{\lambda_i}\, u_i)^T [x]_d}_{= q_i} \Big)^2$$
$\boxtimes$

Bem.: a) Entscheidungsproblem „$p \in \Sigma_{n,2d}$?" ist algorithmisch einfach ($\rightarrow$ SDP Zulässigkeit).

b) Rang von Q gibt an, wie viele Summanden ein SOS-Darstellung mind. gebraucht werden.

Bsp.: $p(x) = x^4 + 4x^3 + 6x^2 + 4x + 5 \in \Sigma_{1,4}$

$$= \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}^T \begin{bmatrix} q_{00} & q_{01} & q_{02} \\ q_{01} & q_{11} & q_{12} \\ q_{02} & q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}.$$

$$= q_{22} x^4 + 2 q_{12} x^3 + (q_{11} + 2 q_{02}) x^2 + 2 q_{01} + q_{00}$$

$\Longrightarrow$ $4 = q_{22}$, $q_{12} = 3$, $q_{11} + 2 q_{02} = 6$, $q_{01} = 2$, $q_{00} = 5$

Matrix $Q = \begin{bmatrix} 5 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 1 \end{bmatrix} \succeq 0 \Longrightarrow p \in \Sigma_{1,4}$.

Frage: Verhältnis zwischen $\Sigma_{n,2d}$ und $P_{n,2d}$?

Satz 3 $\quad \Sigma_{1,2d} = P_{1,2d}$.

Bew.: Sei $p \in P_{1,2d}$ mit $\deg p = 2d$.

Fundamentalsatz der Algebra:

$$p(x) = c \prod_{i=1}^{r} (x-\alpha_i)^2 \prod_{i=1}^{s} (x-\beta_i)(x-\overline{\beta_i})$$

α_i: reelle Nullstellen,

$\beta_i, \overline{\beta_i}$ Paare komplexer Nullstellen

$c > 0$ Koeffizient von Monom x^{2d}.

Falls $(*)$ $\quad \prod_{i=1}^{s} (x-\beta_i)(x-\overline{\beta_i}) = \sum_{j=1}^{m} (q_j(x))^2$, dann

$$p(x) = \sum_{j=1}^{m} \left(\sqrt{c} \prod_{i=1}^{r} (x-\alpha_i)\, q_j\right)^2.$$

Nun ist $\quad (x-\beta_i)(x-\overline{\beta_i}) = x^2 - (\beta_i + \overline{\beta_i})\, x + \beta_i \overline{\beta_i}$

$$= x^2 - (2 \operatorname{Re} \beta_i)\, x + (\operatorname{Re} \beta_i)^2 + (\operatorname{Im} \beta_i)^2$$

$$= (x - \operatorname{Re} \beta_i)^2 + (\operatorname{Im} \beta_i)^2$$

und die Existenz der Darstellung $(*)$ folgt. $\boxtimes$

Zusatz: Kann jedes $p \in \Sigma_{1,d}$ als Summe von 2 Quadraten schreiben.

Bsp. 4 Das Motzkinpolynom

$$M(x,y) = x^2y^2(x^2+y^2-3)+1 \in P_{2,6} \setminus \Sigma_{2,6}.$$

Bew.:

$M \in P_{2,6}$: Falls $x^2+y^2-3 \geq 0$, dann $M(x,y) \geq 0$.

Falls $x^2+y^2-3 < 0$, dann def. $z^2 = 3-x^2-y^2$. (>0).

AM-GM-Ungleichung:

$$\underbrace{\left(x^2y^2z^2\right)^{1/3}}_{=\left(-M(x,y)+1\right)^{1/3}} \leq \frac{x^2+y^2+z^2}{3} = 1 \implies M(x,y) \geq 0.$$

$M \notin \Sigma_{2,6}$: Ang. $M = \sum_{k=1}^{m} q_k^2$.

$\deg q_k \leq 3$. Koeffizienten von q_k:

x^3	$: 0$
y^3	$: 0$
x^2	$: 0$
y^2	$: 0$
x	$: 0$
y	$: 0.$

Also $q_k = a_k xy^2 + b_k x^2y + c_k xy + d_k$.

Koeffizient von x^2y^2 in M ist -3. D.h. $-3 = \sum_{k=1}^{m} c_k^2$. $\lightning$

Satz 5 (Hilbert 1888)

$$\Sigma_{n,2d} = P_{n,2d} \iff \begin{array}{l} n=1 \quad \text{oder} \\ d=1 \quad \text{oder} \\ n=2,\ d=2. \end{array}$$

Bew.: $n=1$: Satz 3

$d=1$: Choleskyzerlegung

$n=2,\ d=2$: recht schwierig (nicht hier)

alle anderen Fälle: Varianten des Motzkin-Polynoms.

Bem.:
- Entscheidungsproblem „$p \in P_{n,4}$?" ist NP-schwer.
- Blekherman (2006): Im allg. ist $P_{n,2d}$ viel größer als $\Sigma_{n,2d}$.

§ 2 Globale Optimierung mit Polynomen

Polynomielles Optimierungsproblem (POP)

<u>geg.:</u> Polynome $f, g_1, \dots, g_m \in \mathbb{R}[x]$

$$K = \{ x \in \mathbb{R}^n : g_1(x) \geq 0, \dots, g_m(x) \geq 0 \}$$ Menge der zulässigen Lösungen

f Zielfkt.

<u>ges.:</u> $$p_{min} = \inf_{x \in K} f(x)$$

<u>Bem.:</u> a) POP ist sehr allgemein.

b) POP ist sehr schwer.

Dualer Ansatz

$$\mathcal{P}(K) = \{ p \in \mathbb{R}[x] : p(x) \geq 0 \ \ \forall x \in K \}$$ konvexer Kegel.

Dann $$p_{min} = \sup \{ \lambda : p - \lambda \in \mathcal{P}(K) \}.$$

<u>Klar:</u> Optimieren über $\mathcal{P}(K)$ ist schwer.

<u>Relaxierung</u> (Parrilo; Lasserre 2001)

Finde Sequenz von Kegeln $C_1 \subseteq C_2 \subseteq \dots \subseteq \mathcal{P}(K)$ für die $\sup \{ \lambda : p - \lambda \in C_i \}$ einfacher zu lösen ist.

→ gibt Sequenz

$$p_1 \le p_2 \le \dots \le p_{min}$$

von unteren Schranken, die p_{min} approximieren.

→ Optimieren über C_i ist leichter als Optimieren über $C_{i+1}, \dots$

SOS-Relaxierung

$$p_{sos,i} = \sup \{ \lambda : p - \lambda \in \Sigma_{n,2i} + g_1 \Sigma_{n,2i} + \dots + g_m \Sigma_{n,2i} \}$$

Dann

$$p_{sos,1} \le p_{sos,2} \le \dots \le p_{min}.$$

<u>Theorem von Putinar</u>: Falls K kompakt (und K „schön" durch $g_1, \dots, g_m$ dargestellt), dann $p_{min} = \lim_{i \to \infty} p_{sos,i}$.

* wichtiges theoretisches Resultat mit praktischem Nutzen
* oft ist $p_{sos,i} = p_{min}$ für kleines i.
* Ausgangspkt. aktueller Forschung.

Def. 1 a) $M \subseteq \mathbb{R}[x]$ heißt <u>quadratischer Modul</u>, falls $1 \in M$, $M + M \subseteq M$, $\sum M \subseteq M$.

b) Seien $g_1, \dots, g_m \in \mathbb{R}[x]$.

$$M(g_1, \dots, g_m) = \Big\{ \mu_0 + \sum_{j=1}^{m} \mu_j g_j \; : \; \mu_j \in \Sigma, \; j = 0, \dots, m \Big\}$$

ist der von $g_1, \dots, g_m$ <u>erzeugte</u> q. M.

c) $M(g_1, \dots, g_m)$ heißt <u>Archimedisch</u>, falls

$$\forall f \in \mathbb{R}[x] \; \exists N \in \mathbb{N} : \quad N \pm f \in M(g)$$

[Kann zeigen: $M(g_1, \dots, g_m)$ ist archimedisch $\iff$

$$\exists N \in \mathbb{N} : \quad N - \sum_{i=1}^{n} x_i^2 \in M(g) \;]$$

Insbesondere: Falls $M(g_1, \dots, g_m)$ archimedisch, dann K kompakt.

Theorem 2 (Putinar 1993) „Putinars Positivstellensatz"

Angenommen $M(g_1, \dots, g_m)$ ist archimedisch. Dann gilt

$$p(x) > 0 \quad \forall x \in K \implies p \in M(g_1, \dots, g_m).$$

[„Umkehrung": $p \in M(g_1, \dots, g_m) \implies p(x) \geq 0 \quad \forall x \in K$].
ist klar.

§ 3 Beweis von Putinars Positivstellensatz

Def. 1 a) $I \subseteq \mathbb{R}[x]$ heißt <u>Ideal</u>, falls $I + I \subseteq I$ und $\mathbb{R}[x]\, I \subseteq I$.

b) $(g_1, \dots, g_m) = \left\{ \sum_{j=1}^{m} u_j g_j : u_j \in \mathbb{R}[x] \right\}$ ist das von $g_1, \dots, g_m$ <u>erzeugte</u> Ideal

Lemma 2 Sei $M \subseteq \mathbb{R}[x]$ ein quadratischer Modul. Dann ist $I = M \cap (-M)$ ein Ideal.

Lemma 3 Sei $M \subseteq \mathbb{R}[x]$ ein maximaler echter q.M. (d.h. $-1 \notin M$). Dann ist $M \cup (-M) = \mathbb{R}[x]$.

Lemma 4 Sei $M \subseteq \mathbb{R}[x]$ ein maximaler echter, archimedischer q.M. Dann: $\forall f \in \mathbb{R}[x]\ \exists!\, a \in \mathbb{R} : f - a \in I$, mit $I = M \cap (-M)$.

Lemma 5 Ang. $p(x) > 0\ \forall x \in K$. Dann: $\exists\, s \in \Sigma : sp - 1 \in M(g_1, \dots, g_m)$

Lemma 6 Ang. $p(x) > 0\ \forall x \in K$. Dann:

$$\exists\, N \in \mathbb{N},\ h \in M(g_1, \dots, g_m) : \quad N - h \in \Sigma, \quad hp - 1 \in M(g_1, \dots, g_m).$$

Bew. (Putinars Positivstellensatz)

Seien N, h wie in Lemma 6. Da $M(g_1, \dots, g_m)$ archimedisch $\exists\, k \in \mathbb{N}: \quad k + p \in M(g_1, \dots, g_m)$.

Können annehmen: $N > 0$, $k > 0$

Es ist

$$\left(k - \frac{1}{N}\right) + p = \frac{1}{N} \underbrace{\left(\underbrace{(N-h)}_{\in \Sigma} \underbrace{(k+p)}_{\in M(g)} + \underbrace{(hp-1)}_{\in M(g)} + \underbrace{kh}_{\in M(g)}\right)}_{\in M(g)}$$

D. h. wir haben gezeigt: $k + p \in M(g_1, \dots, g_m) \Rightarrow \left(k - \frac{1}{N}\right) + p \in M(g_1, \dots, g_m)$

kN-fache Iteration liefert $\underbrace{\left(k - \frac{kN}{N}\right) + p}_{= p} \in M(g_1, \dots, g_m)$. $\boxtimes$

Bew.: (Lemma 2)

$I + I \subseteq I$: ✓

$\mathbb{R}[x]\, I \subseteq I$: Seien $f \in \mathbb{R}[x]$, $g \in I$. Dann

$$\underbrace{\left(\frac{f+1}{2}\right)^2 g + \left(\frac{f-1}{2}\right)^2 (-g)}_{\in I = M \cap (-M).} = fg$$

$\boxtimes$

Bew.: (Lemma 3)

Ang. $\exists f \notin M \cup (-M)$. Definiere die q. M.

$$M' = \sum f + M, \quad M'' = \sum (-f) + M.$$

M', M'' enthalten M strikt. Weil M maximal echt ist:

$\exists\ g_1, g_2 \in M,\ s_1, s_2 \in \Sigma$:

$$-1 = s_1 f + g_1 \quad , \quad -1 = s_2(-f) + g_2.$$

$$\Longrightarrow\ -s_2 - s_1 = s_2(s_1 f + g_1) + s_1(s_2(-f) + g_2)$$
$$= \underbrace{s_2 g_1 + s_1 g_2}_{\in M}$$

$$\Longrightarrow\ -s_1$$
$$= s_2 g_1 + s_1 g_2 + s_2 \in M$$

$$\Longrightarrow\ s_1, s_2 \in -M.$$

Andererseits : $s_1, s_2 \in \Sigma \subseteq M$.

Also $s_1, s_2 \in I = M \cap (-M)$ (nach Lemma 2 ein Ideal).

$\Longrightarrow\ s_1 f \in I \subseteq M \Longrightarrow\ -1 = s_1 f + g_1 \in M$ ↯. ⊠.

Bew.: (Lemma 4)

Definiere

$$A = \{a \in \mathbb{R} : f - a \in M\}, \quad B = \{b \in \mathbb{R} : b - f \in M\}.$$

Weil M archimedisch: $A, B \neq \emptyset$.

Zu zeigen: $|A \cap B| = 1$. [Dann $f - a \in M$, $a - f \in M$,
d.h. $f - a \in M \cap (-M) = I$].

1) Es ist $a \leq b \quad \forall a \in A, b \in B$.

Ang. $a > b$. Dann $b - a = \underbrace{(f-a) + (b-f)}_{\in M} < 0$. $\lightning$ zu $-1 \notin M$.

2) $a_0 = \sup A$, $b_0 = \inf B$. Dann $a_0 = b_0$.

Ang. $\exists\, c : \quad a_0 < c < b_0$. Dann $f - c \notin M \cup (-M) \overset{\text{Lemma 3}}{=} \mathbb{R}[x]$ $\lightning$

3) $a_0 = b_0 \in A \cap B$.

Ang. $f - a_0 \notin M$. Definiere q. M. $M' = M + (f - a_0)\Sigma$.
Weil M max. echt:

$$\exists\, g \in M, s \in \Sigma : \quad -1 = g + (f - a_0)s.$$

Da M archimedisch: $\exists\, N \in \mathbb{N} : \quad N - s \in M$.

Sei ε, so dass $0 < \varepsilon < 1/N$.

Dann $a_0 - \varepsilon \in A$, $\quad f - (a_0 - \varepsilon) = (f - a_0) + \varepsilon \in M$.

Also
$$-1 + \varepsilon s = g + (f - a_0 + \varepsilon)s \in M.$$

Weiter
$$-1 + \varepsilon s + \underbrace{\varepsilon(N - s)}_{\in M} = \underbrace{-1 + \varepsilon N}_{< 0} \in M.$$

$\lightning$.

$\boxtimes$

Bew.: (Lemma 5)

z.z.: $M_0 = M(g_1, \dots, g_m) - p\Sigma$ ist nicht echt.

Ang. M_0 ist echt. Ziel: Finde $a \in K$ mit $p(a) < 0$.

Sei M ein maximaler echter q.M., der M_0 enthält (gibt es nach dem Lemma von Zorn). M ist archimedisch, weil $M \supseteq M(g_1, \dots, g_m)$. Nach Lemma 4 $\exists\, a_i \in \mathbb{R}$:

$$X_i - a_i \in I = M \cap (-M). \qquad i = 1, \dots, n.$$

Definiere $a = (a_1, \dots, a_n)$.

Beh.: $\forall f \in \mathbb{R}[x]$: $\quad f - f(a) \in I$.

Wende Beh. auf g_j an. Dann

$$g_j(a) = g_j - (g_j - g_j(a)) \in M.$$

Weil M echt, muss $g_j(a) \geq 0$ sein. Also $a \in K$.

Desweiteren

$$-p(a) = (p - p(a)) - p \in M.$$

Weil M echt: $-p(a) \geq 0$. $\lightning$.

Bew. (der Beh.)

Sei $f = \sum_\alpha f_\alpha x^\alpha$. Dann $\sum_\alpha f_\alpha (x^\alpha - a^\alpha) = f - f(a)$

z.z.: $x^\alpha - a^\alpha \in I$.

Induktion nach $|\alpha|$

$|\alpha| = 0$: ✓

$|\alpha| > 0$: Ang. $\alpha_1 \geq 1$. $\beta = \alpha - e_1$.

$$x^\alpha - a^\alpha = X_1 \underbrace{(x^\beta - a^\beta)}_{\underset{\text{I.V.}}{\in I}} + a^\beta \underbrace{(X_1 - a_1)}_{\in I} \in I.$$

$\boxtimes$

Bew.: (Lemma 6)

Wähle $s \in \Sigma$ wie in Lemma 5.

Da $M(g_1, \dots, g_m)$ archimedisch: $\exists k \in \mathbb{N}$: $2k - s$, $2k - s^2 p - 1 \in M(g_1, \dots, g_m)$

Def. $h = s(2k - s)$, $N = k^2$.

Dann $h \in M(g_1, \dots, g_m)$ und

$$N - h = (k - s)^2 \in \Sigma \quad \text{und}$$

$$hp - 1 = s(2k - s)p - 1 = 2k \underbrace{(sp - 1)}_{\in M(g)} + \underbrace{(2k - s^2 p - 1)}_{\in M(g)} \in M(g).$$

$\boxtimes$