

Inhaltsübersicht für heute:

Innere-Punkte-Verfahren

Lineare Optimierung über Kegeln

Second-Order-Cone Programme

SOC Anwendung: Regularisierung

SOC Anwendung: Klassifizierung, Support-Vektor

SOC Anwendung: Das Markowitz Modell, Chance Constraints

Inhaltsübersicht für heute:

Innere-Punkte-Verfahren

Lineare Optimierung über Kegeln

Second-Order-Cone Programme

SOC Anwendung: Regularisierung

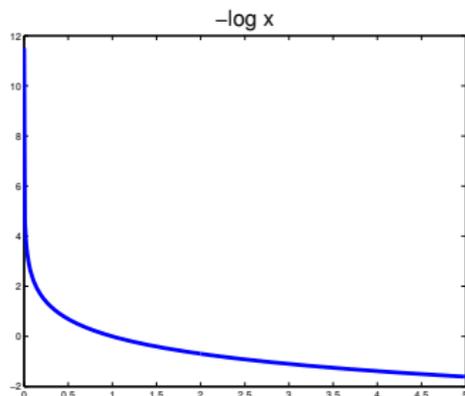
SOC Anwendung: Klassifizierung, Support-Vektor

SOC Anwendung: Das Markowitz Modell, Chance Constraints

Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren.
Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).



Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren.
Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).

Bsp: $\min x$ s.t. $x \in [a = 0.5, b = 3.5]$

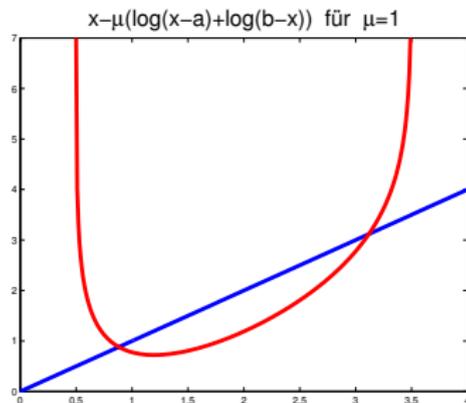
Barriere für $x \geq a$: $-\log(x - a)$

Barriere für $x \leq b$: $-\log(b - x)$

Barriereproblem:

$\min f(x) := x - \mu(\log(x - a) + \log(b - x))$

Starte mit $x \in (a, b)$, suche Minimum, verkleinere μ , suche Minimum, etc.



Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren.
Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).

Bsp: $\min x$ s.t. $x \in [a = 0.5, b = 3.5]$

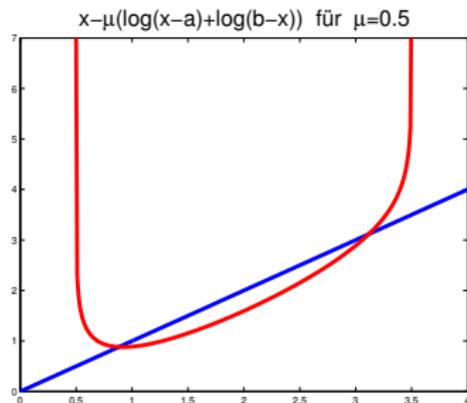
Barriere für $x \geq a$: $-\log(x - a)$

Barriere für $x \leq b$: $-\log(b - x)$

Barriereproblem:

$\min f(x) := x - \mu(\log(x - a) + \log(b - x))$

Starte mit $x \in (a, b)$, suche Minimum, verkleinere μ , suche Minimum, etc.



Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren.
Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).

Bsp: $\min x$ s.t. $x \in [a = 0.5, b = 3.5]$

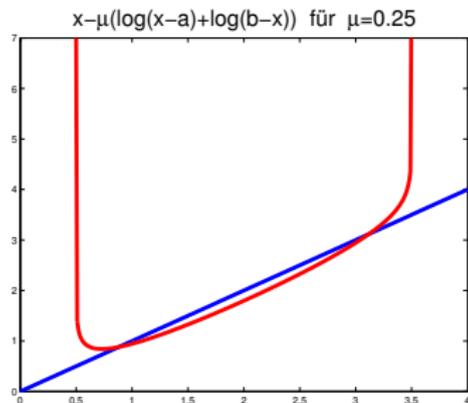
Barriere für $x \geq a$: $-\log(x - a)$

Barriere für $x \leq b$: $-\log(b - x)$

Barriereproblem:

$\min f(x) := x - \mu(\log(x - a) + \log(b - x))$

Starte mit $x \in (a, b)$, suche Minimum, verkleinere μ , suche Minimum, etc.



Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren.
Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).

Bsp: $\min x$ s.t. $x \in [a = 0.5, b = 3.5]$

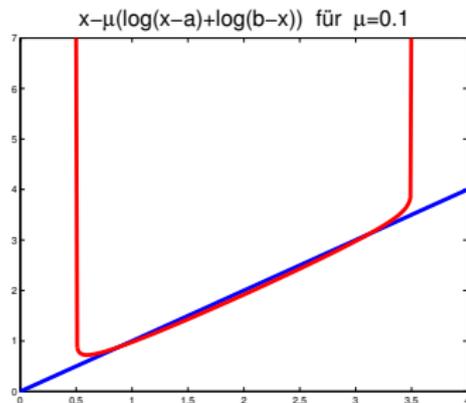
Barriere für $x \geq a$: $-\log(x - a)$

Barriere für $x \leq b$: $-\log(b - x)$

Barriereproblem:

$\min f(x) := x - \mu(\log(x - a) + \log(b - x))$

Starte mit $x \in (a, b)$, suche Minimum, verkleinere μ , suche Minimum, etc.



Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren.
Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).

Bsp: $\min x$ s.t. $x \in [a = 0.5, b = 3.5]$

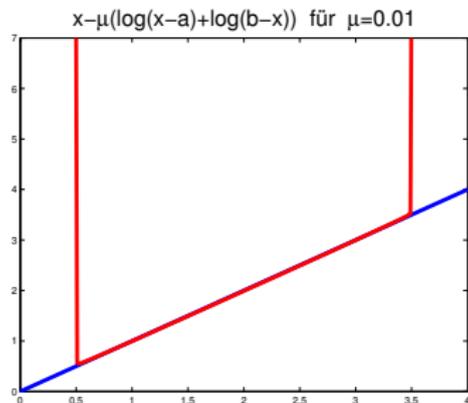
Barriere für $x \geq a$: $-\log(x - a)$

Barriere für $x \leq b$: $-\log(b - x)$

Barriereproblem:

$\min f(x) := x - \mu(\log(x - a) + \log(b - x))$

Starte mit $x \in (a, b)$, suche Minimum, verkleinere μ , suche Minimum, etc.



Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren.
Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).

Bsp: $\min x$ s.t. $x \in [a = 0.5, b = 3.5]$

Barriere für $x \geq a$: $-\log(x - a)$

Barriere für $x \leq b$: $-\log(b - x)$

Barriereproblem:

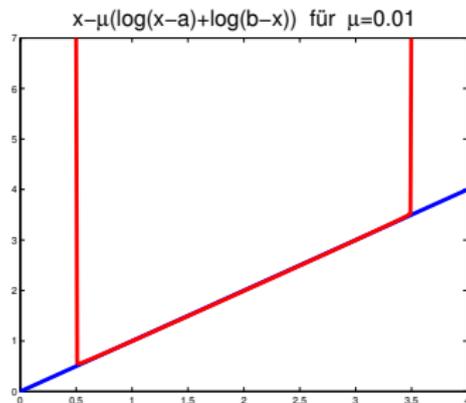
$\min f(x) := x - \mu(\log(x - a) + \log(b - x))$

Starte mit $x \in (a, b)$, suche Minimum, verkleinere μ , suche Minimum, etc.

Primales/duales Problem:

$$(P) \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array}$$

$$(D) \quad \begin{array}{ll} \max & b^T y \\ \text{s.t.} & A^T y + z = c \\ & z \geq 0 \end{array}$$



[für $x \in \mathbb{R}^n$: $\log x = \sum_{i=1}^n \log x_i$]

Innere-Punkte-Verfahren

Nachweislich polynomiales Verfahren! [Simplex *worst case*: exponentiell]

Nutzt Idee der **Barriere-Verfahren**, um durch das Innere des Polyeders zu gehen: Ein mit einem **Barriere-Parameter** $\mu > 0$ kontrollierter Barriere-Term der Zielfunktion verhindert das Verlassen des Inneren. Eine günstige Barriere für $x \geq 0$ ist $-\log x$ (streng konvex).

Bsp: $\min x$ s.t. $x \in [a = 0.5, b = 3.5]$

Barriere für $x \geq a$: $-\log(x - a)$

Barriere für $x \leq b$: $-\log(b - x)$

Barriereproblem:

$\min f(x) := x - \mu(\log(x - a) + \log(b - x))$

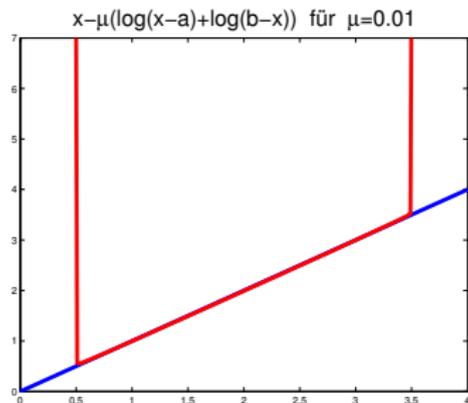
Starte mit $x \in (a, b)$, suche Minimum, verkleinere μ , suche Minimum, etc.

Primales/duales Barriere-Problem:

$$(P_\mu) \quad \begin{array}{ll} \min & c^T x - \mu \log x \\ \text{s.t.} & Ax = b \\ & [x > 0] \end{array}$$

[für $x \in \mathbb{R}^n$: $\log x = \sum_{i=1}^n \log x_i$]

$$(D_\mu) \quad \begin{array}{ll} \max & b^T y + \mu \log z \\ \text{s.t.} & A^T y + z = c \\ & [z > 0] \end{array}$$



Das Primal-Duale KKT-System

Der Name KKT bezieht sich auf die Optimalitätsbedingungen für nichtlineare restringierte Optimierung (näheres dazu später).

$$\begin{array}{ll}
 \min & c^T x - \mu \log x \\
 (P_\mu) \quad \text{s.t.} & Ax = b \\
 & [x > 0]
 \end{array}
 \qquad
 \begin{array}{ll}
 \max & b^T y + \mu \log z \\
 (D_\mu) \quad \text{s.t.} & A^T y + z = c \\
 & [z > 0]
 \end{array}$$

Für die Lagrange-Funktion zum primalen Barriere-Problem (P_μ)

$$L_\mu(x, y) := c^T x + y^T (b - Ax) - \mu \log x$$

sucht man die Nullstellen der Ableitung nach x und y

$$\begin{array}{ll}
 \nabla_x L_\mu = 0 & \rightarrow \quad c - A^T y - \mu x^{-1} = 0 \\
 \nabla_y L_\mu = 0 & \rightarrow \quad b - Ax = 0
 \end{array}
 \qquad
 [x^{-1} := [x_1^{-1}, \dots, x_n^{-1}]^T]$$

Das Primal-Duale KKT-System

Der Name KKT bezieht sich auf die Optimalitätsbedingungen für nichtlineare restringierte Optimierung (näheres dazu später).

$$\begin{array}{ll}
 \min & c^T x - \mu \log x \\
 (P_\mu) \quad \text{s.t.} & Ax = b \\
 & [x > 0]
 \end{array}
 \qquad
 \begin{array}{ll}
 \max & b^T y + \mu \log z \\
 (D_\mu) \quad \text{s.t.} & A^T y + z = c \\
 & [z > 0]
 \end{array}$$

Für die Lagrange-Funktion zum primalen Barriere-Problem (P_μ)

$$L_\mu(x, y) := c^T x + y^T (b - Ax) - \mu \log x$$

sucht man die Nullstellen der Ableitung nach x und y

$$\nabla_x L_\mu = 0 \rightarrow c - A^T y - \mu x^{-1} = 0 \quad [x^{-1} := [x_1^{-1}, \dots, x_n^{-1}]^T]$$

$$\nabla_y L_\mu = 0 \rightarrow b - Ax = 0$$

Setzt man $z = \mu x^{-1}$ erhält man das **Primal-Duale KKT-System**

$$(PD_\mu) \quad \begin{array}{ll}
 A^T y + z = c & \text{duale Zulässigkeit} \quad [z > 0] \\
 Ax = b & \text{primale Zulässigkeit} \quad [x > 0] \\
 x \circ z = \mu \mathbf{1} & \text{perturbierte Komplementarität} \quad [x \circ z := [x_i z_i]]
 \end{array}$$

Für $\mu = 0$: $x \circ z = 0 \Leftrightarrow x^T z = 0$, mit $x, z \in \mathbb{R}_+^n$ Opt.-Bed. für (P) und (D)!

Das Primal-Duale KKT-System

Der Name KKT bezieht sich auf die Optimalitätsbedingungen für nichtlineare restringierte Optimierung (näheres dazu später).

$$\begin{array}{ll}
 \min & c^T x - \mu \log x \\
 (P_\mu) \quad \text{s.t.} & Ax = b \\
 & [x > 0]
 \end{array}
 \qquad
 \begin{array}{ll}
 \max & b^T y + \mu \log z \\
 (D_\mu) \quad \text{s.t.} & A^T y + z = c \\
 & [z > 0]
 \end{array}$$

Für die Lagrange-Funktion zum primalen Barriere-Problem (P_μ)

$$L_\mu(x, y) := c^T x + y^T (b - Ax) - \mu \log x$$

sucht man die Nullstellen der Ableitung nach x und y

$$\nabla_x L_\mu = 0 \rightarrow c - A^T y - \mu x^{-1} = 0 \qquad [x^{-1} := [x_1^{-1}, \dots, x_n^{-1}]^T]$$

$$\nabla_y L_\mu = 0 \rightarrow b - Ax = 0$$

Setzt man $z = \mu x^{-1}$ erhält man das **Primal-Duale KKT-System**

$$(PD_\mu) \quad \begin{array}{ll}
 A^T y + z = c & \text{duale Zulässigkeit} \qquad [z > 0] \\
 Ax = b & \text{primale Zulässigkeit} \qquad [x > 0] \\
 x \circ z = \mu \mathbf{1} & \text{perturbierte Komplementarität} \qquad [x \circ z := [x_i z_i]]
 \end{array}$$

Für $\mu = 0$: $x \circ z = 0 \Leftrightarrow x^T z = 0$, mit $x, z \in \mathbb{R}_+^n$ Opt.-Bed. für (P) und (D)!

Sind (P) und (D) streng zulässig (und hat A vollen Zeilenrang), gibt es für jedes $\mu > 0$ eine eindeutige Lösung $(x_\mu, y_\mu, z_\mu) \in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n$, es gilt x_μ ist OL von (P_μ) und (y_μ, z_μ) ist OL von (D_μ) .

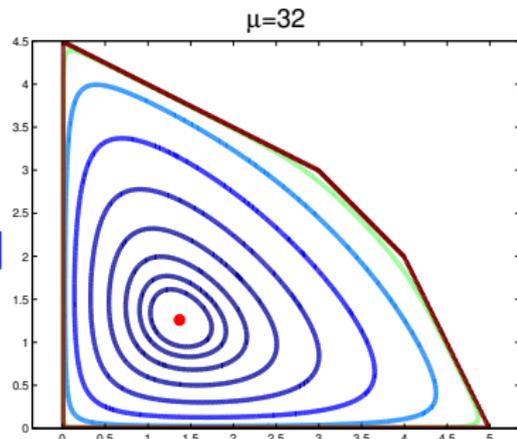
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{array}{l} A^T y + z = c \\ Ax = b \\ x \circ z = \mu \mathbf{1} \end{array} \quad \begin{array}{l} [z > 0] \\ [x > 0] \end{array}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



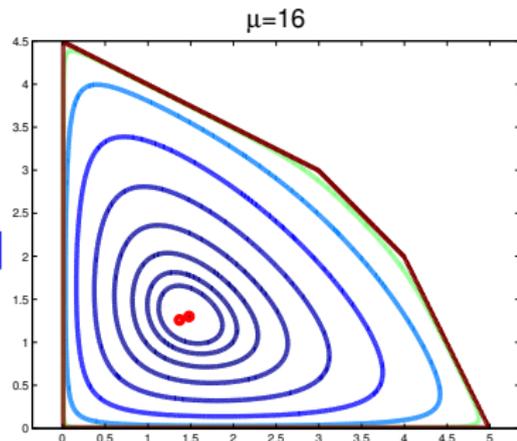
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{cases} A^T y + z = c & [z > 0] \\ Ax = b & [x > 0] \\ x \circ z = \mu \mathbf{1} \end{cases}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



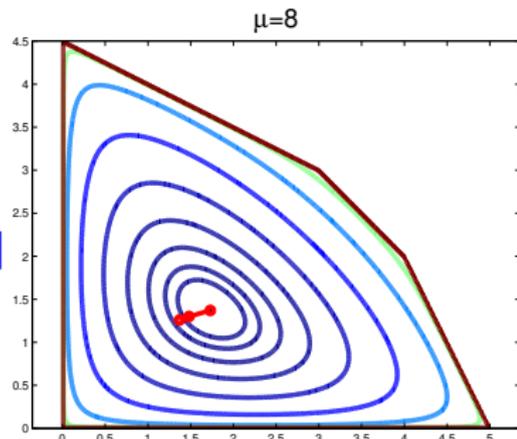
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{array}{l} A^T y + z = c \\ Ax = b \\ x \circ z = \mu \mathbf{1} \end{array} \quad \begin{array}{l} [z > 0] \\ [x > 0] \end{array}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



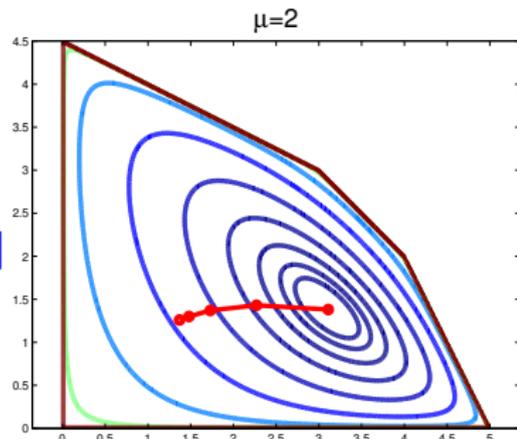
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{array}{l} A^T y + z = c \\ Ax = b \\ x \circ z = \mu \mathbf{1} \end{array} \quad \begin{array}{l} [z > 0] \\ [x > 0] \end{array}$$

$$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} Z(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



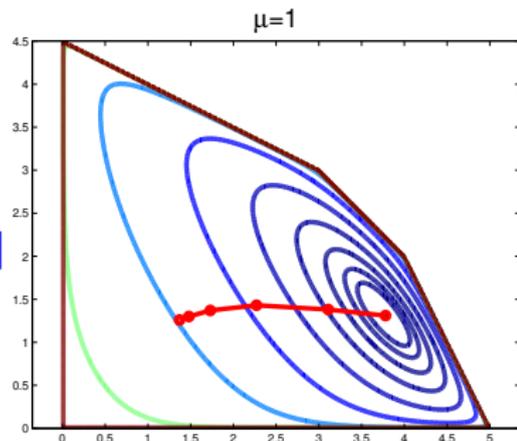
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{cases} A^T y + z = c & [z > 0] \\ Ax = b & [x > 0] \\ x \circ z = \mu \mathbf{1} \end{cases}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} Z(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



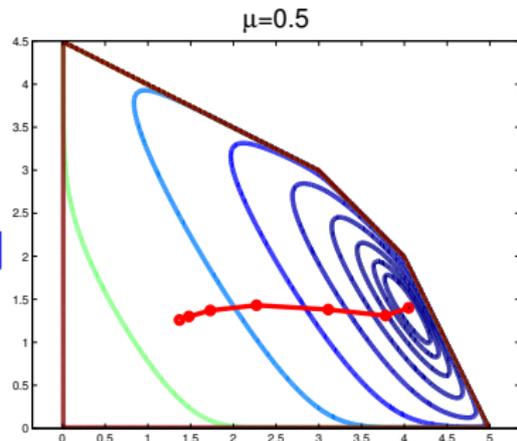
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{array}{l} A^T y + z = c \\ Ax = b \\ x \circ z = \mu \mathbf{1} \end{array} \quad \begin{array}{l} [z > 0] \\ [x > 0] \end{array}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



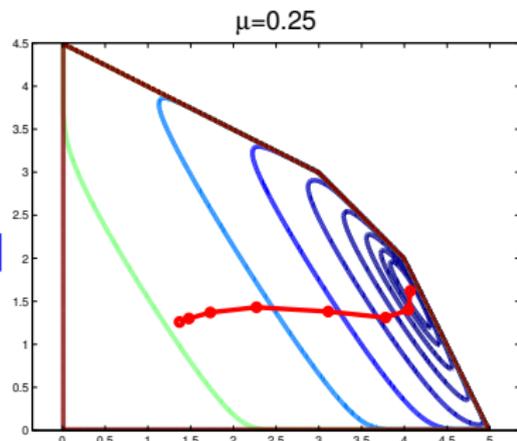
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{array}{l} A^T y + z = c \\ Ax = b \\ x \circ z = \mu \mathbf{1} \end{array} \quad \begin{array}{l} [z > 0] \\ [x > 0] \end{array}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



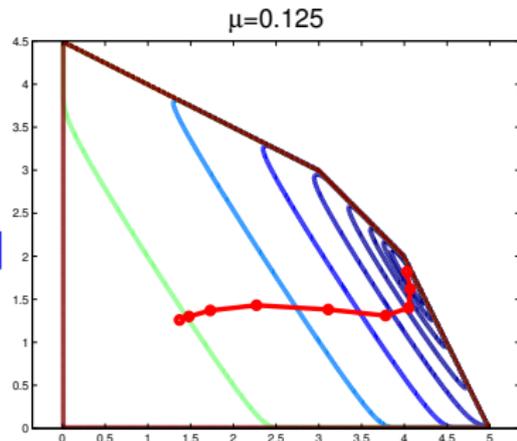
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{cases} A^T y + z = c & [z > 0] \\ Ax = b & [x > 0] \\ x \circ z = \mu \mathbf{1} \end{cases}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



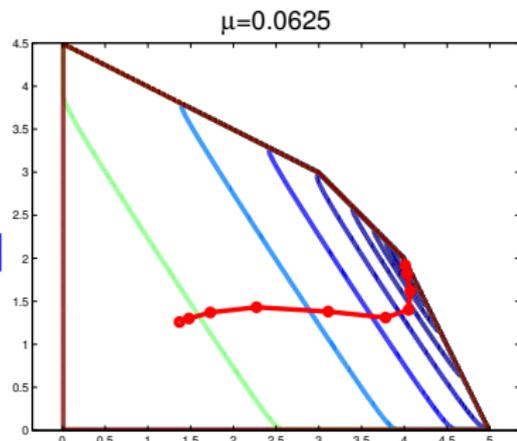
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{cases} A^T y + z = c & [z > 0] \\ Ax = b & [x > 0] \\ x \circ z = \mu \mathbf{1} \end{cases}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



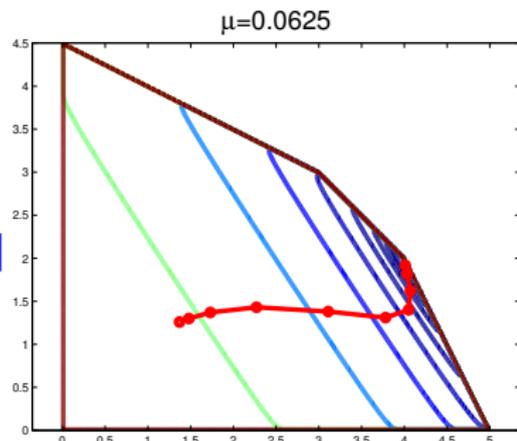
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{cases} A^T y + z = c & [z > 0] \\ Ax = b & [x > 0] \\ x \circ z = \mu \mathbf{1} \end{cases}$$

$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} Z(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



Er liegt innerhalb des primalen/dualen Polyeders und konvergiert für $\mu \rightarrow 0$ gegen das **analytische Zentrum** der Opt.-Lösungen

$$(x^*, y^*, z^*) := \lim_{\mu \rightarrow 0} (x_\mu, y_\mu, z_\mu) \in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n.$$

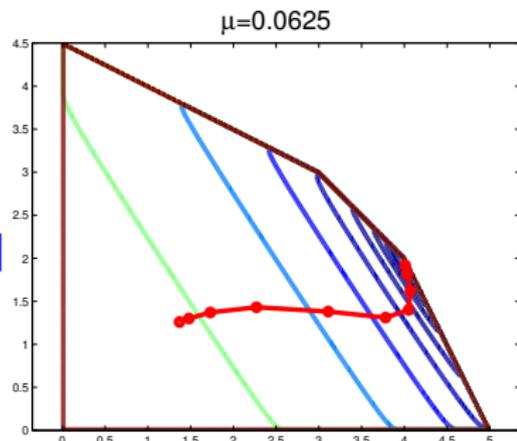
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{cases} A^T y + z = c & [z > 0] \\ Ax = b & [x > 0] \\ x \circ z = \mu \mathbf{1} \end{cases}$$

$$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



Er liegt innerhalb des primalen/dualen Polyeders und konvergiert für $\mu \rightarrow 0$ gegen das **analytische Zentrum** der Opt.-Lösungen

$$(x^*, y^*, z^*) := \lim_{\mu \rightarrow 0} (x_\mu, y_\mu, z_\mu) \in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n.$$

x^* und (y^*, z^*) sind OL von (P) und (D) und **strenge komplementär**, d.h., sie erfüllen $x_i z_i = 0$ mit entweder $x_i > 0$ oder $z_i > 0$.

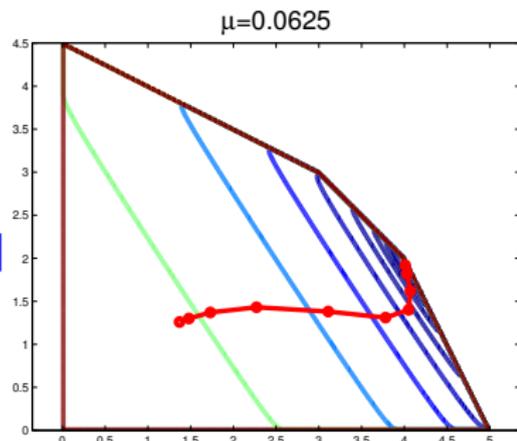
Der zentrale Pfad, strenge Komplementarität

$$(PD_\mu) \quad \begin{cases} A^T y + z = c & [z > 0] \\ Ax = b & [x > 0] \\ x \circ z = \mu \mathbf{1} \end{cases}$$

$$[(P), (D) \text{ str. zul.}, \text{Rang}(A) = m]$$

Der **zentrale Pfad** ist die glatte Kurve

$$\begin{aligned} \mathcal{Z}(\mu) &= (x_\mu, y_\mu, z_\mu), \quad \mu > 0, \\ (x_\mu, y_\mu, z_\mu) &\in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n \text{ löst } (PD_\mu). \end{aligned}$$



Er liegt innerhalb des primalen/dualen Polyeders und konvergiert für $\mu \rightarrow 0$ gegen das **analytische Zentrum** der Opt.-Lösungen

$$(x^*, y^*, z^*) := \lim_{\mu \rightarrow 0} (x_\mu, y_\mu, z_\mu) \in \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^n.$$

x^* und (y^*, z^*) sind OL von (P) und (D) und **streng komplementär**, d.h., sie erfüllen $x_i z_i = 0$ mit entweder $x_i > 0$ oder $z_i > 0$.

Nur falls die Optimallösung von (P) bzw. (D) eindeutig ist, ist x^* bzw. (y^*, z^*) auch eine Ecke des zulässigen Polyeders von (P) bzw. (D)!

Lösung des Primal-Dualen Systems

$\mathcal{Z}(\mu)$ nähert man iterativ mit dem Newton-Verfahren für nichtlineare Gleichungen an (s. später): Bestimme für gegebenes $\bar{x} > 0$, \bar{y} , $\bar{z} > 0$ den nächsten Punkt $(\bar{x} + \Delta x, \bar{y} + \Delta y, \bar{z} + \Delta z)$ aus der Linearisierung:

$$\begin{array}{rcl}
 A^T y + z & = & c \\
 Ax & = & b \\
 x \circ z & = & \mu \mathbf{1}
 \end{array}
 \longrightarrow
 \begin{array}{rcl}
 \text{I.} & A^T(\bar{y} + \Delta y) + \bar{z} + \Delta z & = c \\
 \text{II.} & A(\bar{x} + \Delta x) & = b \\
 \text{III.} & \bar{x} \circ \bar{z} + \Delta x \circ \bar{z} + \bar{x} \circ \Delta z & = \mu \mathbf{1}
 \end{array}$$

Lösung des Primal-Dualen Systems

$\mathcal{Z}(\mu)$ nähert man iterativ mit dem Newton-Verfahren für nichtlineare Gleichungen an (s. später): Bestimme für gegebenes $\bar{x} > 0$, \bar{y} , $\bar{z} > 0$ den nächsten Punkt $(\bar{x} + \Delta x, \bar{y} + \Delta y, \bar{z} + \Delta z)$ aus der Linearisierung:

$$\begin{array}{rcl}
 A^T y + z & = & c \\
 Ax & = & b \\
 x \circ z & = & \mu \mathbf{1}
 \end{array}
 \longrightarrow
 \begin{array}{rcl}
 \text{I.} & A^T(\bar{y} + \Delta y) + \bar{z} + \Delta z & = c \\
 \text{II.} & A(\bar{x} + \Delta x) & = b \\
 \text{III.} & \bar{x} \circ \bar{z} + \Delta x \circ \bar{z} + \bar{x} \circ \Delta z & = \mu \mathbf{1}
 \end{array}$$

- aus I.: $\Delta z = c - A^T \bar{y} - \bar{z} - A^T \Delta y$
- aus III.: $\Delta x = \mu \bar{z}^{-1} - \bar{x} - \bar{x} \circ \bar{z}^{-1} \circ \Delta z \stackrel{(1.)}{=} \bar{x} \circ \bar{z}^{-1} \circ A^T \Delta y + \dots$
- in II.: $A \text{Diag}(\bar{x} \circ \bar{z}^{-1}) A^T \Delta y = b - A \bar{x} + \dots$

Lösung des Primal-Dualen Systems

$\mathcal{Z}(\mu)$ nähert man iterativ mit dem Newton-Verfahren für nichtlineare Gleichungen an (s. später): Bestimme für gegebenes $\bar{x} > 0$, \bar{y} , $\bar{z} > 0$ den nächsten Punkt $(\bar{x} + \Delta x, \bar{y} + \Delta y, \bar{z} + \Delta z)$ aus der Linearisierung:

$$\begin{array}{rcl} A^T y + z & = & c \\ Ax & = & b \\ x \circ z & = & \mu \mathbf{1} \end{array} \quad \longrightarrow \quad \begin{array}{rcl} \text{I.} & A^T(\bar{y} + \Delta y) + \bar{z} + \Delta z & = c \\ \text{II.} & A(\bar{x} + \Delta x) & = b \\ \text{III.} & \bar{x} \circ \bar{z} + \Delta x \circ \bar{z} + \bar{x} \circ \Delta z & = \mu \mathbf{1} \end{array}$$

- aus I.: $\Delta z = c - A^T \bar{y} - \bar{z} - A^T \Delta y$
 - aus III.: $\Delta x = \mu \bar{z}^{-1} - \bar{x} - \bar{x} \circ \bar{z}^{-1} \circ \Delta z \stackrel{(1.)}{=} \bar{x} \circ \bar{z}^{-1} \circ A^T \Delta y + \dots$
 - in II.: $A \text{Diag}(\bar{x} \circ \bar{z}^{-1}) A^T \Delta y = b - A \bar{x} + \dots$
-

Wegen $x > 0, z > 0$ ist die Matrix $M := A \text{Diag}(\bar{x} \circ \bar{z}^{-1}) A^T \in \mathbb{R}^{m \times m}$ symmetrisch positiv semidefinit und positiv definit, falls $\text{Rang}(A) = m$.

\Rightarrow Das Gleichungssystem $M \Delta y = \dots$ kann in etwa $m^3/3$ arithmetischen Operationen mit dem **Cholesky-Verfahren** gelöst werden.

Dieses „faktoriert“ $M = LL^T$ mit L untere Dreiecksmatrix.

Lösung des Primal-Dualen Systems

$Z(\mu)$ nähert man iterativ mit dem Newton-Verfahren für nichtlineare Gleichungen an (s. später): Bestimme für gegebenes $\bar{x} > 0, \bar{y}, \bar{z} > 0$ den nächsten Punkt $(\bar{x} + \Delta x, \bar{y} + \Delta y, \bar{z} + \Delta z)$ aus der Linearisierung:

$$\begin{array}{rcl} A^T y + z & = & c \\ Ax & = & b \\ x \circ z & = & \mu \mathbf{1} \end{array} \quad \longrightarrow \quad \begin{array}{rcl} \text{I.} & A^T(\bar{y} + \Delta y) + \bar{z} + \Delta z & = c \\ \text{II.} & A(\bar{x} + \Delta x) & = b \\ \text{III.} & \bar{x} \circ \bar{z} + \Delta x \circ \bar{z} + \bar{x} \circ \Delta z & = \mu \mathbf{1} \end{array}$$

- aus I.: $\Delta z = c - A^T \bar{y} - \bar{z} - A^T \Delta y$
 - aus III.: $\Delta x = \mu \bar{z}^{-1} - \bar{x} - \bar{x} \circ \bar{z}^{-1} \circ \Delta z \stackrel{(1.)}{=} \bar{x} \circ \bar{z}^{-1} \circ A^T \Delta y + \dots$
 - in II.: $A \text{Diag}(\bar{x} \circ \bar{z}^{-1}) A^T \Delta y = b - A \bar{x} + \dots$
-

Wegen $x > 0, z > 0$ ist die Matrix $M := A \text{Diag}(\bar{x} \circ \bar{z}^{-1}) A^T \in \mathbb{R}^{m \times m}$ symmetrisch positiv semidefinit und positiv definit, falls $\text{Rang}(A) = m$.
 \Rightarrow Das Gleichungssystem $M \Delta y = \dots$ kann in etwa $m^3/3$ arithmetischen Operationen mit dem **Cholesky-Verfahren** gelöst werden.

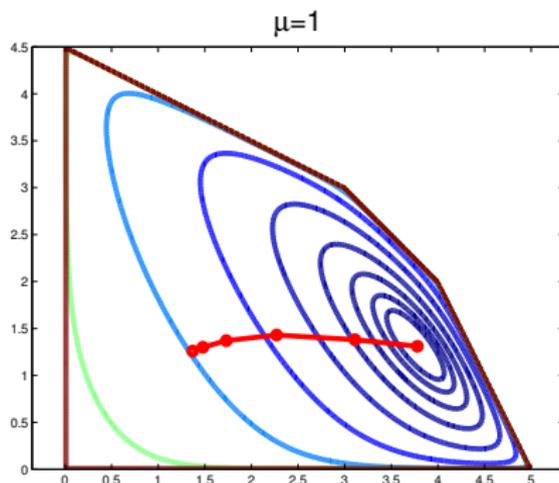
Dieses „faktoriert“ $M = LL^T$ mit L untere Dreiecksmatrix.

Wichtiger Unterschied zum Simplex für die Praxis:

A ist meist nur sehr dünn besetzt, für Simplex in $A_B x_B = b$ gut nutzbar!
 Trotz vieler Tricks ist L meist sehr viel dichter besetzt als A

\rightarrow braucht viel mehr Speicher, jede Iteration sehr langsam, dafür wenige!

Grobe Näherung für $\mathcal{Z}(\mu)$ genügt



Für $\bar{\mu} > \mu > 0$ mit $\bar{\mu} - \mu$ klein, ist $\|\mathcal{Z}(\bar{\mu}) - \mathcal{Z}(\mu)\|$ klein. Ist $(\bar{x}, \bar{y}, \bar{z})$ gute Näherung für $\mathcal{Z}(\bar{\mu})$ mit $\bar{\mu} = \bar{x}^T \bar{z} / n$ und $\mu = \sigma \bar{\mu}$ mit $\sigma < 1$ nahe bei 1, dann genügt ein Newton-Schritt, damit $(\bar{x} + \Delta x, \bar{y} + \Delta y, \bar{z} + \Delta z)$ eine gute Näherung für $\mathcal{Z}(\sigma \bar{\mu})$ ist mit $\sigma \bar{\mu} = (\bar{x} + \Delta x)^T (\bar{z} + \Delta z) / n$.

Algorithmisches Schema für Innere-Punkte-Verfahren

Idee: Folge dem zentralen Pfad \mathcal{Z} mit Newton-Schritten für $\mu \rightarrow 0$.

Input: $A, b, c, \bar{x} > 0, \bar{y}, \bar{z} > 0, \varepsilon > 0, \sigma \in (0, 1)$

1. Setze $\mu \leftarrow \sigma \frac{\bar{x}^T \bar{z}}{n}$.
2. Bestimme $(\Delta x, \Delta y, \Delta z)$ als Newton-Schritt für (PD_μ) .
3. Bestimme großes $\alpha \in (0, 1]$ mit $\bar{x} + \alpha \Delta x > 0, \bar{z} + \alpha \Delta z > 0$
4. setze $(\bar{x}, \bar{y}, \bar{z}) \leftarrow (\bar{x}, \bar{y}, \bar{z}) + \alpha(\Delta x, \Delta y, \Delta z)$
5. Ist $\|b - A\bar{x}\| < \varepsilon$ und $\|c - A^T \bar{y} - \bar{z}\| < \varepsilon$ und $\bar{x}^T \bar{z} < \varepsilon$ STOP, sonst gehe zu 1.

Algorithmisches Schema für Innere-Punkte-Verfahren

Idee: Folge dem zentralen Pfad \mathcal{Z} mit Newton-Schritten für $\mu \rightarrow 0$.

Input: $A, b, c, \bar{x} > 0, \bar{y}, \bar{z} > 0, \varepsilon > 0, \sigma \in (0, 1)$

1. Setze $\mu \leftarrow \sigma \frac{\bar{x}^T \bar{z}}{n}$.
2. Bestimme $(\Delta x, \Delta y, \Delta z)$ als Newton-Schritt für (PD_μ) .
3. Bestimme großes $\alpha \in (0, 1]$ mit $\bar{x} + \alpha \Delta x > 0, \bar{z} + \alpha \Delta z > 0$
4. setze $(\bar{x}, \bar{y}, \bar{z}) \leftarrow (\bar{x}, \bar{y}, \bar{z}) + \alpha(\Delta x, \Delta y, \Delta z)$
5. Ist $\|b - A\bar{x}\| < \varepsilon$ und $\|c - A^T \bar{y} - \bar{z}\| < \varepsilon$ und $\bar{x}^T \bar{z} < \varepsilon$ STOP, sonst gehe zu 1.

Sind (P) und (D) zulässig und $\text{Rang}(A) = m$, kann man für ein leicht umformuliertes Problem einen zulässigen Startpunkt $\bar{x}, \bar{y}, \bar{z}$ und ein $\sigma < 1$ angeben, sodass der Algorithmus nach höchstens $\gamma \sqrt{n \log \frac{\bar{x}^T \bar{z}}{\varepsilon}}$ Iterationen (für eine problemunabhängige Konstante $\gamma > 0$) endet.

Algorithmisches Schema für Innere-Punkte-Verfahren

Idee: Folge dem zentralen Pfad \mathcal{Z} mit Newton-Schritten für $\mu \rightarrow 0$.

Input: $A, b, c, \bar{x} > 0, \bar{y}, \bar{z} > 0, \varepsilon > 0, \sigma \in (0, 1)$

1. Setze $\mu \leftarrow \sigma \frac{\bar{x}^T \bar{z}}{n}$.
2. Bestimme $(\Delta x, \Delta y, \Delta z)$ als Newton-Schritt für (PD_μ) .
3. Bestimme großes $\alpha \in (0, 1]$ mit $\bar{x} + \alpha \Delta x > 0, \bar{z} + \alpha \Delta z > 0$
4. setze $(\bar{x}, \bar{y}, \bar{z}) \leftarrow (\bar{x}, \bar{y}, \bar{z}) + \alpha(\Delta x, \Delta y, \Delta z)$
5. Ist $\|b - A\bar{x}\| < \varepsilon$ und $\|c - A^T \bar{y} - \bar{z}\| < \varepsilon$ und $\bar{x}^T \bar{z} < \varepsilon$ STOP, sonst gehe zu 1.

Sind (P) und (D) zulässig und $\text{Rang}(A) = m$, kann man für ein leicht umformuliertes Problem einen zulässigen Startpunkt $\bar{x}, \bar{y}, \bar{z}$ und ein $\sigma < 1$ angeben, sodass der Algorithmus nach höchstens $\gamma \sqrt{n \log \frac{\bar{x}^T \bar{z}}{\varepsilon}}$ Iterationen (für eine problemunabhängige Konstante $\gamma > 0$) endet.

- Das Ergebnis ist eine Näherung und keine exakte Lösung!
- Für ε klein genug kann für diese in einem „**cross-over**“-Schritt eine optimale Basis bestimmt werden (benötigt oft viele Simplex-Iterationen).
- Erkennen von Unzulässigkeit/Unbeschränktheit ist deutlich schwerer.

Zusammenfassung der Eigenschaften

- Innere-Punkte-Verfahren haben polynomiale Laufzeit.
- Für große Probleme sind sie oft schneller als Simplex, Degeneriertheit der Polyeder ist kaum ein Problem.
- Erkennen von Unzulässigkeit/Unbeschränktheit ist schwerer.
- Da Schnittebenen den zentralen Pfad großräumig verändern, ist kein guter Warmstart von der alten OL aus möglich.
→ erst cross-over, dann dualer Simplex
- Der gleiche Ansatz funktioniert sehr gut für lineare Programme über gutartigen nichtlinearen Kegeln!

Inhaltsübersicht für heute:

Innere-Punkte-Verfahren

Lineare Optimierung über Kegeln

Second-Order-Cone Programme

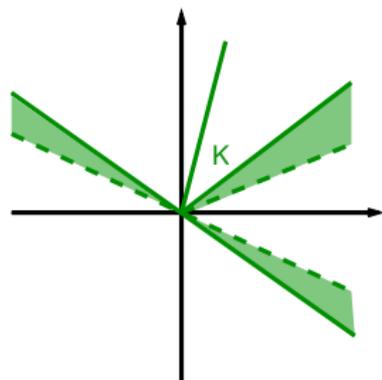
SOC Anwendung: Regularisierung

SOC Anwendung: Klassifizierung, Support-Vektor

SOC Anwendung: Das Markowitz Modell, Chance Constraints

Konvexe Kegel, dualer Kegel

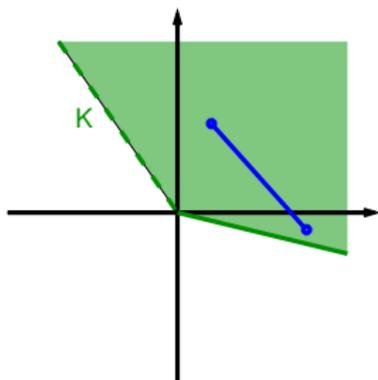
Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.



Konvexe Kegel, dualer Kegel

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **konvexer Kegel**,
wenn mit $x, y \in K$ auch $\alpha(x + y) \in K$ für $\alpha \geq 0$.

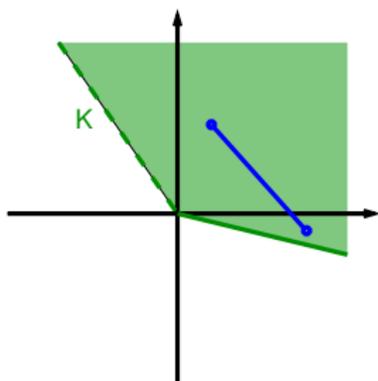


Konvexe Kegel, dualer Kegel

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **konvexer Kegel**,
wenn mit $x, y \in K$ auch $\alpha(x + y) \in K$ für $\alpha \geq 0$.

Bspe: \emptyset , $\{0\}$, \mathbb{R}^n , \mathbb{R}_+^n ,
lineare Unterräume $\{x \in \mathbb{R}^n : Ax = 0\}$.



Konvexe Kegel, dualer Kegel

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.

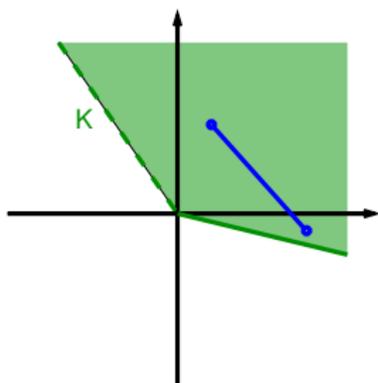
Eine Menge $K \subseteq \mathbb{R}^n$ heißt **konvexer Kegel**,
wenn mit $x, y \in K$ auch $\alpha(x + y) \in K$ für $\alpha \geq 0$.

Bspe: \emptyset , $\{0\}$, \mathbb{R}^n , \mathbb{R}_+^n ,

lineare Unterräume $\{x \in \mathbb{R}^n : Ax = 0\}$.

Sind $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ konvexe Kegel,

so auch $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.



Konvexe Kegel, dualer Kegel

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **konvexer Kegel**,
wenn mit $x, y \in K$ auch $\alpha(x + y) \in K$ für $\alpha \geq 0$.

Bspe: \emptyset , $\{0\}$, \mathbb{R}^n , \mathbb{R}_+^n ,

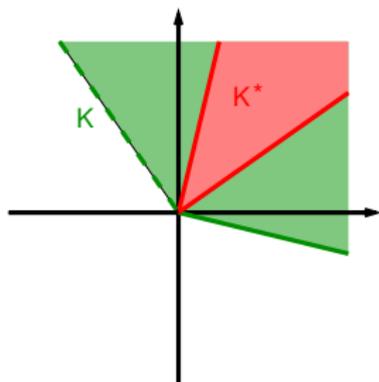
lineare Unterräume $\{x \in \mathbb{R}^n : Ax = 0\}$.

Sind $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ konvexe Kegel,

so auch $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

Ist $K \subseteq \mathbb{R}^n$ konvexer Kegel, nennt man

$K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ den zu K **dualen Kegel**.



Konvexe Kegel, dualer Kegel

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **konvexer Kegel**,
wenn mit $x, y \in K$ auch $\alpha(x + y) \in K$ für $\alpha \geq 0$.

Bspe: \emptyset , $\{0\}$, \mathbb{R}^n , \mathbb{R}_+^n ,

lineare Unterräume $\{x \in \mathbb{R}^n : Ax = 0\}$.

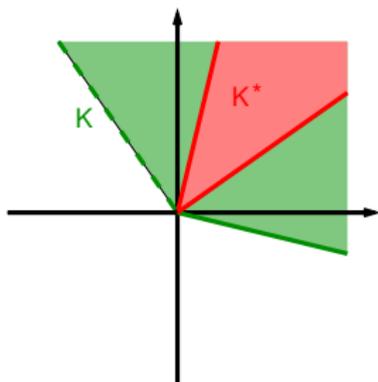
Sind $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ konvexe Kegel,

so auch $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.

Ist $K \subseteq \mathbb{R}^n$ konvexer Kegel, nennt man

$K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \forall x \in K\}$ den zu K **dualen Kegel**.

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



Konvexe Kegel, dualer Kegel

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.

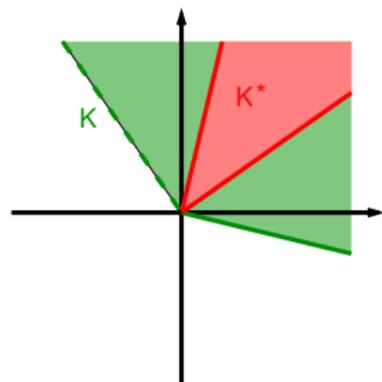
Eine Menge $K \subseteq \mathbb{R}^n$ heißt **konvexer Kegel**,
wenn mit $x, y \in K$ auch $\alpha(x + y) \in K$ für $\alpha \geq 0$.

Bspe: \emptyset , $\{0\}$, \mathbb{R}^n , \mathbb{R}_+^n ,

lineare Unterräume $\{x \in \mathbb{R}^n : Ax = 0\}$.

Sind $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ konvexe Kegel,

so auch $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.



Ist $K \subseteq \mathbb{R}^n$ konvexer Kegel, nennt man

$K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ den zu K **dualen Kegel**.

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

Für die Optimierung wichtig ist die Eigenschaft

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

Konvexe Kegel, dualer Kegel

Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**,
wenn mit $x \in K$ auch $\alpha x \in K$ für $\alpha \geq 0$.

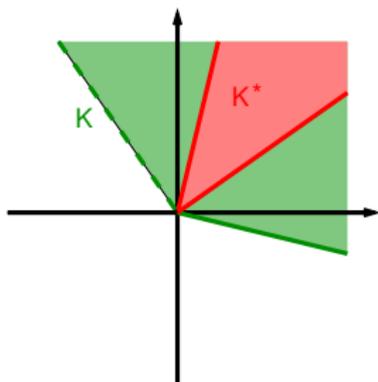
Eine Menge $K \subseteq \mathbb{R}^n$ heißt **konvexer Kegel**,
wenn mit $x, y \in K$ auch $\alpha(x + y) \in K$ für $\alpha \geq 0$.

Bspe: \emptyset , $\{0\}$, \mathbb{R}^n , \mathbb{R}_+^n ,

lineare Unterräume $\{x \in \mathbb{R}^n : Ax = 0\}$.

Sind $K_1 \subseteq \mathbb{R}^n$, $K_2 \subseteq \mathbb{R}^m$ konvexe Kegel,

so auch $K_1 \times K_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in K_1, y \in K_2 \right\}$.



Ist $K \subseteq \mathbb{R}^n$ konvexer Kegel, nennt man

$K^* := \{z \in \mathbb{R}^n : z^T x \geq 0 \ \forall x \in K\}$ den zu K **dualen Kegel**.

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

Für die Optimierung wichtig ist die Eigenschaft

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

Vorsicht! $(K^*)^* = K$ gilt NUR, wenn K ein abgeschlossener konvexer Kegel ist!

Primale/Duale lineare Programme über Kegeln

Ersetze in der linearen Optimierung \mathbb{R}_+^n durch konvexen Kegel $K \subseteq \mathbb{R}^n$:

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

Primale/Duale lineare Programme über Kegeln

Ersetze in der linearen Optimierung \mathbb{R}_+^n durch konvexen Kegel $K \subseteq \mathbb{R}^n$:

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

Definiere die Lagrange-Funktion

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

Für $y \in \mathbb{R}^m$ und $Ax = b$ ist $(b - Ax)^T y = 0$, daher gilt

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

Primale/Duale lineare Programme über Kegeln

Ersetze in der linearen Optimierung \mathbb{R}_+^n durch konvexen Kegel $K \subseteq \mathbb{R}^n$:

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

Definiere die Lagrange-Funktion

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

Für $y \in \mathbb{R}^m$ und $Ax = b$ ist $(b - Ax)^T y = 0$, daher gilt

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

Die beste untere Schranke (Lagrange-Relaxation) liefert

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

Primale/Duale lineare Programme über Kegeln

Ersetze in der linearen Optimierung \mathbb{R}_+^n durch konvexen Kegel $K \subseteq \mathbb{R}^n$:

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

Definiere die Lagrange-Funktion

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

Für $y \in \mathbb{R}^m$ und $Ax = b$ ist $(b - Ax)^T y = 0$, daher gilt

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

Die beste untere Schranke (Lagrange-Relaxation) liefert

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

Das inf ist nur für $z = c - A^T y \in K^*$ endlich, das duale Programm lautet

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

Schwache und starke Dualität

Sei $K \subseteq \mathbb{R}^n$ ein konvexer Kegel.

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

Nach Konstruktion gilt immer **schwache Dualität**, $v(P) \geq v(D)$,
Gleichheit gilt keineswegs immer (Beispiel später)!

Für starke Dualität braucht es Zusatzbedingungen:

Schwache und starke Dualität

Sei $K \subseteq \mathbb{R}^n$ ein konvexer Kegel.

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

Nach Konstruktion gilt immer **schwache Dualität**, $v(P) \geq v(D)$,
Gleichheit gilt keineswegs immer (Beispiel später)!

Für starke Dualität braucht es Zusatzbedingungen:

Ein primal zulässiges \bar{x} heißt **streng zulässig** für (P) , wenn \bar{x} im Inneren von K liegt ($\exists \rho > 0 : B_\rho(\bar{x}) := \{x \in \mathbb{R}^n : \|x - \bar{x}\| \leq \rho\} \subseteq K$). Gibt es so ein \bar{x} , heißt auch (P) **streng zulässig**.

Schwache und starke Dualität

Sei $K \subseteq \mathbb{R}^n$ ein konvexer Kegel.

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

Nach Konstruktion gilt immer **schwache Dualität**, $v(P) \geq v(D)$,
Gleichheit gilt keineswegs immer (Beispiel später)!

Für starke Dualität braucht es Zusatzbedingungen:

Ein primal zulässiges \bar{x} heißt **streng zulässig** für (P), wenn \bar{x} im Inneren von K liegt ($\exists \rho > 0 : B_\rho(\bar{x}) := \{x \in \mathbb{R}^n : \|x - \bar{x}\| \leq \rho\} \subseteq K$). Gibt es so ein \bar{x} , heißt auch (P) **streng zulässig**.

Ein dual zulässiges (\bar{y}, \bar{z}) heißt **streng zulässig** für (D), wenn \bar{z} im Inneren von K^* liegt ($\exists \rho > 0 : B_\rho(\bar{z}) \subseteq K^*$). Gibt es so ein (\bar{y}, \bar{z}) , heißt auch (D) **streng zulässig**.

Satz (Starke Dualität)

Ist (P) streng zulässig, wird das duale Optimum $v(D)$ angenommen.

Ist (D) streng zulässig, wird das primale Optimum $v(P)$ angenommen.

In beiden Fällen gilt $v(P) = v(D)$.

Selbstduale Kegel, Innere-Punkte-Verfahren

Hier werden nur drei spezielle Kegelarten K verwendet:

- $K = \mathbb{R}_+^n$, der nichtnegative Orthant
- $K = \mathcal{Q}^n$, der quadratische Kegel
- $K = \mathcal{S}_+^n$ der Kegel der positiv semidefiniten Matrizen

Die genauen Definitionen von \mathcal{Q}^n und \mathcal{S}_+^n folgen demnächst.

Selbstduale Kegel, Innere-Punkte-Verfahren

Hier werden nur drei spezielle Kegelarten K verwendet:

- $K = \mathbb{R}_+^n$, der nichtnegative Orthant
- $K = \mathcal{Q}^n$, der quadratische Kegel
- $K = \mathcal{S}_+^n$ der Kegel der positiv semidefiniten Matrizen

Die genauen Definitionen von \mathcal{Q}^n und \mathcal{S}_+^n folgen demnächst.

Die wichtigsten Eigenschaften dieser drei sind:

- Sie sind **selbstdual**, d.h., $K = K^*$.
- Es sind geeignete Barriere-Funktionen für sie bekannt.
- Gute Innere-Punkte-Codes, die die Verwendung der drei gleichzeitig erlauben, sind für MATLAB verfügbar (SeDuMi, SDPT3).
- Sie ermöglichen viele wichtige Anwendungen zu modellieren.

Selbstduale Kegel, Innere-Punkte-Verfahren

Hier werden nur drei spezielle Kegelarten K verwendet:

- $K = \mathbb{R}_+^n$, der nichtnegative Orthant
- $K = \mathcal{Q}^n$, der quadratische Kegel
- $K = \mathcal{S}_+^n$ der Kegel der positiv semidefiniten Matrizen

Die genauen Definitionen von \mathcal{Q}^n und \mathcal{S}_+^n folgen demnächst.

Die wichtigsten Eigenschaften dieser drei sind:

- Sie sind **selbstdual**, d.h., $K = K^*$.
- Es sind geeignete Barriere-Funktionen für sie bekannt.
- Gute Innere-Punkte-Codes, die die Verwendung der drei gleichzeitig erlauben, sind für MATLAB verfügbar (SeDuMi, SDPT3).
- Sie ermöglichen viele wichtige Anwendungen zu modellieren.

In den Anwendung ist K oft aus mehreren Einzelkegeln zusammengesetzt,

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

Das wird sich aber natürlich ergeben und $K = K^*$ gilt für diese Kombinationen immer.

Inhaltsübersicht für heute:

Innere-Punkte-Verfahren

Lineare Optimierung über Kegeln

Second-Order-Cone Programme

SOC Anwendung: Regularisierung

SOC Anwendung: Klassifizierung, Support-Vektor

SOC Anwendung: Das Markowitz Modell, Chance Constraints

Der quadratische Kegel (Second Order Cone)

Der **quadratische Kegel**
oder **Second-Order-Cone (SOC)**

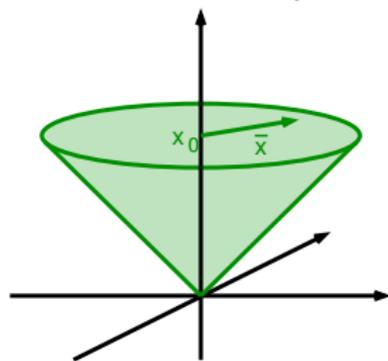
$$Q^n = \left\{ \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \in \mathbb{R}^{n+1} : x_0 \geq \|\bar{x}\| \right\}$$

ist ein konvexer Kegel,

denn für $x, y \in Q^n$, $\alpha \geq 0$ ist

$$\|\alpha(\bar{x} + \bar{y})\| \leq \alpha\|\bar{x}\| + \alpha\|\bar{y}\| \leq \alpha(x_0 + y_0).$$

Q^n ist **selbstdual**, $(Q^n)^* = Q^n$.



Der quadratische Kegel (Second Order Cone)

Der **quadratische Kegel**
oder **Second-Order-Cone (SOC)**

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

ist ein konvexer Kegel,

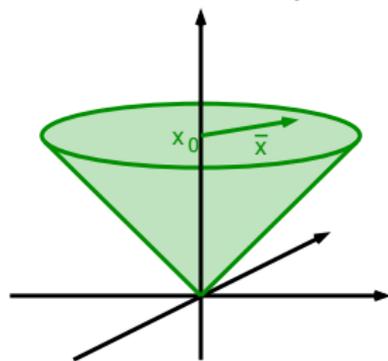
denn für $x, y \in Q^n$, $\alpha \geq 0$ ist

$$\|\alpha(\bar{x} + \bar{y})\| \leq \alpha\|\bar{x}\| + \alpha\|\bar{y}\| \leq \alpha(x_0 + y_0).$$

Q^n ist **selbstdual**, $(Q^n)^* = Q^n$.

Statt $x \in Q^n$ schreiben wir auch $x \geq_Q 0$. Für $a, b \in \mathbb{R}^{n+1}$ ist $a \geq_Q b$ gleichbedeutend mit $a - b \geq_Q 0$, also $a - b \in Q^n$.

Innere-Punkte-Verfahren verwenden $-\log(x_0^2 - \bar{x}^T \bar{x})$ als Barriere.



Der quadratische Kegel (Second Order Cone)

Der **quadratische Kegel**
oder **Second-Order-Cone** (SOC)

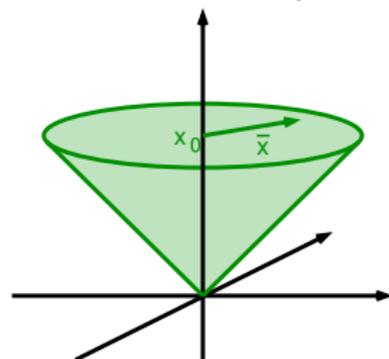
$$Q^n = \left\{ \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \in \mathbb{R}^{n+1} : x_0 \geq \|\bar{x}\| \right\}$$

ist ein konvexer Kegel,

denn für $x, y \in Q^n$, $\alpha \geq 0$ ist

$$\|\alpha(\bar{x} + \bar{y})\| \leq \alpha\|\bar{x}\| + \alpha\|\bar{y}\| \leq \alpha(x_0 + y_0).$$

Q^n ist **selbstdual**, $(Q^n)^* = Q^n$.



Statt $x \in Q^n$ schreiben wir auch $x \geq_Q 0$. Für $a, b \in \mathbb{R}^{n+1}$ ist $a \geq_Q b$ gleichbedeutend mit $a - b \geq_Q 0$, also $a - b \in Q^n$.

Innere-Punkte-Verfahren verwenden $-\log(x_0^2 - \bar{x}^T \bar{x})$ als Barriere.

Lineare Programme, die als Kegel nur \mathbb{R}_+^n und mindestens ein Q^n einsetzen, heißen **Second-Order-Cone Programme** (kurz SOCP).

Ein SOCP mit nur einem Q^n liest sich

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

[Für nur einen einzelnen SOC ist es sogar explizit lösbar.]

Ein SOCP-Beispiel mit Dualitätslücke

$$(P) \quad \begin{array}{ll} \min & -x_2 \\ \text{s.t.} & x_0 - x_1 = 0 \\ & x_0 \geq \sqrt{x_1^2 + x_2^2} \end{array}$$

$$(D) \quad \begin{array}{ll} \max & 0 \\ \text{s.t.} & y_1 + z_0 = 0 \\ & -y_1 + z_1 = 0 \\ & z_2 = -1 \\ & z_0 \geq \sqrt{z_1^2 + z_2^2} \end{array}$$

Ein SOCP-Beispiel mit Dualitätslücke

$$\begin{array}{ll}
 \min & -x_2 \\
 \text{s.t.} & x_0 - x_1 = 0 \\
 & x_0 \geq \sqrt{x_1^2 + x_2^2}
 \end{array}
 \quad (P)
 \qquad
 \begin{array}{ll}
 \max & 0 \\
 \text{s.t.} & y_1 + z_0 = 0 \\
 & -y_1 + z_1 = 0 \\
 & z_2 = -1 \\
 & z_0 \geq \sqrt{z_1^2 + z_2^2}
 \end{array}
 \quad (D)$$

In (P) folgt aus $x_0 = x_1$ und $x \in \mathcal{Q}^2$, dass $x_2 = 0$, also $v(P) = 0$.

Ein SOCP-Beispiel mit Dualitätslücke

$$\begin{array}{ll} \min & -x_2 \\ \text{s.t.} & x_0 - x_1 = 0 \\ & x_0 \geq \sqrt{x_1^2 + x_2^2} \end{array} \quad (P) \qquad \begin{array}{ll} \max & 0 \\ \text{s.t.} & y_1 + z_0 = 0 \\ & -y_1 + z_1 = 0 \\ & z_2 = -1 \\ & z_0 \geq \sqrt{z_1^2 + z_2^2} \end{array} \quad (D)$$

In (P) folgt aus $x_0 = x_1$ und $x \in \mathcal{Q}^2$, dass $x_2 = 0$, also $v(P) = 0$.

In (D) folgt aus $z_0 = -y_1 = -z_1$ und $z \in \mathcal{Q}^2$, dass $z_2 = 0$,
in Widerspruch zu $z_2 = -1$, also $v(D) = -\infty$.

Ein SOCP-Beispiel mit Dualitätslücke

$$\begin{array}{ll}
 \min & -x_2 \\
 \text{s.t.} & x_0 - x_1 = 0 \\
 & x_0 \geq \sqrt{x_1^2 + x_2^2}
 \end{array}
 \quad (P)
 \qquad
 \begin{array}{ll}
 \max & 0 \\
 \text{s.t.} & y_1 + z_0 = 0 \\
 & -y_1 + z_1 = 0 \\
 & z_2 = -1 \\
 & z_0 \geq \sqrt{z_1^2 + z_2^2}
 \end{array}
 \quad (D)$$

In (P) folgt aus $x_0 = x_1$ und $x \in \mathcal{Q}^2$, dass $x_2 = 0$, also $v(P) = 0$.

In (D) folgt aus $z_0 = -y_1 = -z_1$ und $z \in \mathcal{Q}^2$, dass $z_2 = 0$,
in Widerspruch zu $z_2 = -1$, also $v(D) = -\infty$.

Beachte: $v(P) \neq v(D)$ geht nur, wenn beide, Primales und Duales, keinen streng zulässigen Punkt haben.

Ein SOCP-Beispiel mit Dualitätslücke

$$\begin{array}{ll}
 \min & -x_2 \\
 \text{s.t.} & x_0 - x_1 = 0 \\
 & x_0 \geq \sqrt{x_1^2 + x_2^2}
 \end{array}
 \quad (P)
 \qquad
 \begin{array}{ll}
 \max & 0 \\
 \text{s.t.} & y_1 + z_0 = 0 \\
 & -y_1 + z_1 = 0 \\
 & z_2 = -1 \\
 & z_0 \geq \sqrt{z_1^2 + z_2^2}
 \end{array}
 \quad (D)$$

In (P) folgt aus $x_0 = x_1$ und $x \in \mathcal{Q}^2$, dass $x_2 = 0$, also $v(P) = 0$.

In (D) folgt aus $z_0 = -y_1 = -z_1$ und $z \in \mathcal{Q}^2$, dass $z_2 = 0$,
in Widerspruch zu $z_2 = -1$, also $v(D) = -\infty$.

Beachte: $v(P) \neq v(D)$ geht nur, wenn beide, Primales und Duales, keinen streng zulässigen Punkt haben.

In Anwendungen hat eine SOC-Nebenbedingung meist die Struktur

$$\begin{array}{ll}
 \min & c^T x \\
 \text{s.t.} & g^T x + d \geq \|Ax - b\| \\
 & x \in \mathbb{R}^n
 \end{array}$$

Ein SOCP-Beispiel mit Dualitätslücke

$$\begin{array}{ll}
 \min & -x_2 \\
 \text{s.t.} & x_0 - x_1 = 0 \\
 & x_0 \geq \sqrt{x_1^2 + x_2^2}
 \end{array}
 \quad (P)
 \qquad
 \begin{array}{ll}
 \max & 0 \\
 \text{s.t.} & y_1 + z_0 = 0 \\
 & -y_1 + z_1 = 0 \\
 & z_2 = -1 \\
 & z_0 \geq \sqrt{z_1^2 + z_2^2}
 \end{array}
 \quad (D)$$

In (P) folgt aus $x_0 = x_1$ und $x \in \mathcal{Q}^2$, dass $x_2 = 0$, also $v(P) = 0$.

In (D) folgt aus $z_0 = -y_1 = -z_1$ und $z \in \mathcal{Q}^2$, dass $z_2 = 0$,
in Widerspruch zu $z_2 = -1$, also $v(D) = -\infty$.

Beachte: $v(P) \neq v(D)$ geht nur, wenn beide, Primales und Duales, keinen streng zulässigen Punkt haben.

In Anwendungen hat eine SOC-Nebenbedingung meist die Struktur

$$\begin{array}{ll}
 \min & c^T x \\
 \text{s.t.} & g^T x + d \geq \|Ax - b\| \\
 & x \in \mathbb{R}^n
 \end{array}
 \quad \rightarrow \quad
 \begin{array}{ll}
 \min & c^T x \\
 \text{s.t.} & x_0 = g^T x + d \\
 & \bar{x} = Ax - b \\
 & \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \geq_{\mathcal{Q}} 0, x \in \mathbb{R}^n
 \end{array}$$

Inhaltsübersicht für heute:

Innere-Punkte-Verfahren

Lineare Optimierung über Kegeln

Second-Order-Cone Programme

SOC Anwendung: Regularisierung

SOC Anwendung: Klassifizierung, Support-Vektor

SOC Anwendung: Das Markowitz Modell, Chance Constraints

SOC Anwendung: Regularisierte Approximation

Ein Datenvektor $b \in \mathbb{R}^m$ soll einerseits möglichst gut durch eine Linearkombination $\sum_i a^{(i)} x_i$ von Grundfunktionen $a^{(i)} \in \mathbb{R}^m$ dargestellt werden ($\rightarrow \|Ax - b\|$), andererseits möchte man die „Größe“ von x möglichst klein halten oder Eigenschaften von Ax kontrollieren ($\rightarrow f(x)$).

SOC Anwendung: Regularisierte Approximation

Ein Datenvektor $b \in \mathbb{R}^m$ soll einerseits möglichst gut durch eine Linearkombination $\sum_i a^{(i)} x_i$ von Grundfunktionen $a^{(i)} \in \mathbb{R}^m$ dargestellt werden ($\rightarrow \|Ax - b\|$), andererseits möchte man die „Größe“ von x möglichst klein halten oder Eigenschaften von Ax kontrollieren ($\rightarrow f(x)$).

Dieses bikriterielle Problem wird mit Parameter $\gamma > 0$ **skalarisiert** zu

$$\min_{x \in \mathbb{R}_+^n} \|Ax - b\| + \gamma f(x)$$

SOC Anwendung: Regularisierte Approximation

Ein Datenvektor $b \in \mathbb{R}^m$ soll einerseits möglichst gut durch eine Linearkombination $\sum_i a^{(i)} x_i$ von Grundfunktionen $a^{(i)} \in \mathbb{R}^m$ dargestellt werden ($\rightarrow \|Ax - b\|$), andererseits möchte man die „Größe“ von x möglichst klein halten oder Eigenschaften von Ax kontrollieren ($\rightarrow f(x)$).

Dieses bikriterielle Problem wird mit Parameter $\gamma > 0$ **skalarisiert** zu

$$\min_{x \in \mathbb{R}_+^n} \|Ax - b\| + \gamma f(x)$$

Für Anwendungen in der Signalrekonstruktion/-glättung/-entstörung ist

- $b_j = h(t_j)$ das empfangene Signal zum Zeitpunkt t_j
- $a_j^{(i)} = h_i(t_j)$ der Wert der Basisfunktion h_i zum Zeitpunkt t_j
- x_i der zu ermittelnde Anteil der Basisfunktion h_i am Signal

Die Wahl von f unterscheidet sich je nach Zielen, hier:

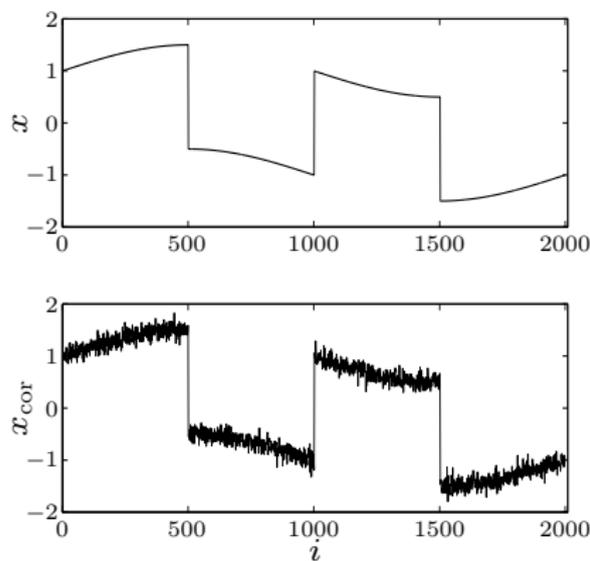
- Glättung bei Sprunganteilen
- Reduktion der Zahl verwendeter Basisfunktionen

Glättung bei Sprunganteilen

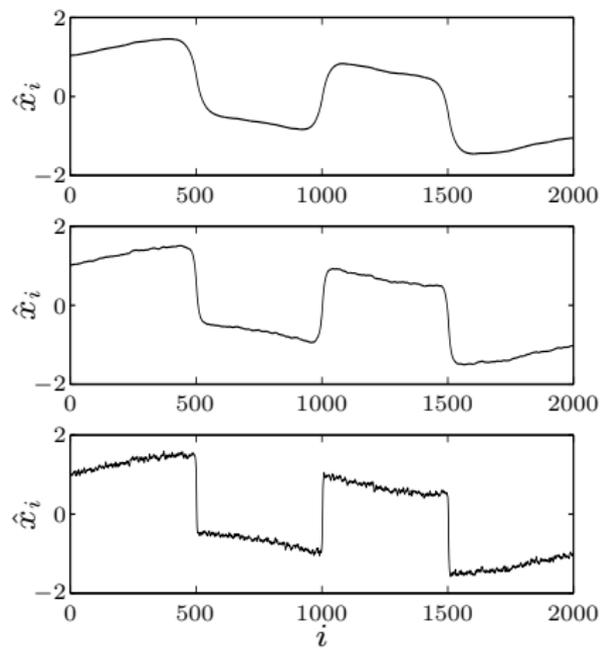
Bei Glättung bestraft f normalerweise starke Änderungen in Ax , bei quadratischer Glättung wäre $f(x) = \sum_{j=1}^{m-1} ([Ax]_{j+1} - [Ax]_j)^2$, aber dadurch werden Sprünge „herausgeglättet“.

[Folie von Boyd, <http://www.stanford.edu/~boyd/cvxbook/>, 25.10.2009]
Andere Notation!

total variation reconstruction example



original signal x and noisy
signal x_{cor}



three solutions on trade-off curve
 $\|\hat{x} - x_{\text{cor}}\|_2$ versus $\phi_{\text{quad}}(\hat{x})$

Glättung bei Sprunganteilen

Bei Glättung bestraft f normalerweise starke Änderungen in Ax , bei quadratischer Glättung wäre $f(x) = \sum_{j=1}^{m-1} ([Ax]_{j+1} - [Ax]_j)^2$, aber dadurch werden Sprünge „herausgeglättet“.

Sollen Sprunganteile erhalten bleiben, verwendet man besser die 1-Norm,

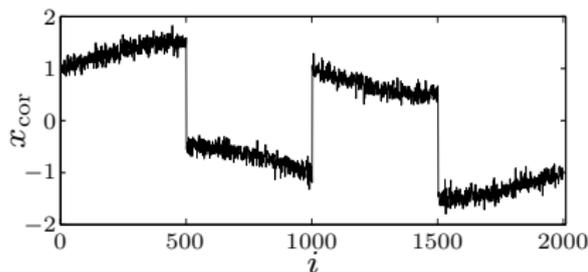
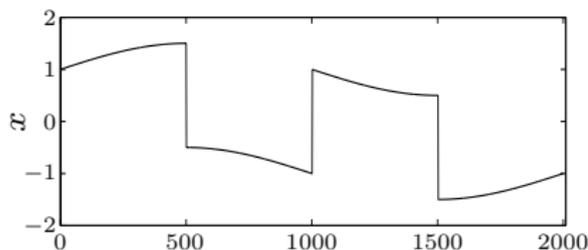
$$f(x) = \sum_{j=1}^{m-1} |[Ax]_{j+1} - [Ax]_j|,$$

für diese dürfen gerne große Terme an wenigen Stellen auftreten, solange dadurch die Summe insgesamt nicht größer wird.

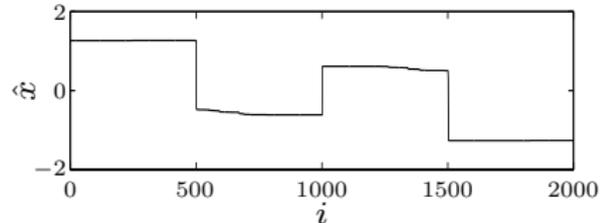
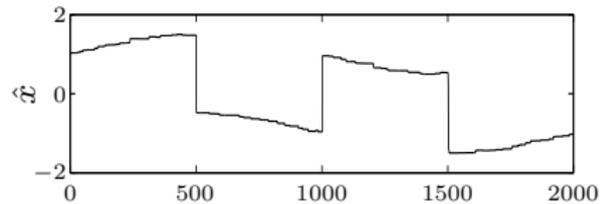
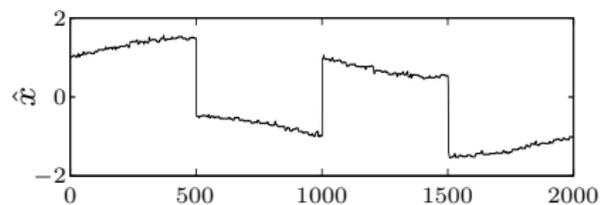
$$\min_{x \in \mathbb{R}_+^n} \|Ax - b\| + \gamma \sum_{j=1}^{m-1} |[Ax]_{j+1} - [Ax]_j|$$

ist ein SOCP:

[Folie von Boyd, <http://www.stanford.edu/~boyd/cvxbook/>, 25.10.2009]
 Andere Notation!



original signal x and noisy
 signal x_{cor}



three solutions on trade-off curve

$$\|\hat{x} - x_{\text{cor}}\|_2 \text{ versus } \phi_{\text{tv}}(\hat{x})$$

Glättung bei Sprunganteilen

Bei Glättung bestraft f normalerweise starke Änderungen in Ax , bei quadratischer Glättung wäre $f(x) = \sum_{j=1}^{m-1} ([Ax]_{j+1} - [Ax]_j)^2$, aber dadurch werden Sprünge „herausgeglättet“.

Sollen Sprunganteile erhalten bleiben, verwendet man besser die 1-Norm,

$$f(x) = \sum_{j=1}^{m-1} |[Ax]_{j+1} - [Ax]_j|,$$

für diese dürfen gerne große Terme an wenigen Stellen auftreten, solange dadurch die Summe insgesamt nicht größer wird.

$$\min_{x \in \mathbb{R}_+^n} \|Ax - b\| + \gamma \sum_{j=1}^{m-1} |[Ax]_{j+1} - [Ax]_j|$$

ist ein SOCP:

Glättung bei Sprunganteilen

Bei Glättung bestraft f normalerweise starke Änderungen in Ax , bei quadratischer Glättung wäre $f(x) = \sum_{j=1}^{m-1} ([Ax]_{j+1} - [Ax]_j)^2$, aber dadurch werden Sprünge „herausgeglättet“.

Sollen Sprunganteile erhalten bleiben, verwendet man besser die 1-Norm,

$$f(x) = \sum_{j=1}^{m-1} |[Ax]_{j+1} - [Ax]_j|,$$

für diese dürfen gerne große Terme an wenigen Stellen auftreten, solange dadurch die Summe insgesamt nicht größer wird.

$$\min_{x \in \mathbb{R}_+^n} \|Ax - b\| + \gamma \sum_{j=1}^{m-1} |[Ax]_{j+1} - [Ax]_j|$$

ist ein SOCP:

[Notation: $A_{j,\bullet}$... Zeile j von A]

$$\begin{array}{ll} \min & \xi_0 + \gamma \sum_{j=1}^{m-1} s_j \\ \text{s.t.} & \xi - Ax = -b \\ & -s_j \leq (A_{j+1,\bullet} - A_{j,\bullet})x \leq s_j \quad j = 1, \dots, m-1 \\ & \begin{bmatrix} \xi_0 \\ \xi \\ \xi \end{bmatrix} \geq_{\mathcal{Q}} 0, x \geq 0, s \geq 0 \end{array}$$

Reduktion der Zahl verwendeter Basisfunktionen

Wenn man weiß, dass sich das Signal nur aus wenigen Basisfunktionen zusammensetzt oder zur Datenkompression nur wenige verwenden will, möchte man ein gemischt-ganzzahliges nichtlineares Problem lösen:

$$\min_{x \in \mathbb{R}^+} \|Ax - b\| + \gamma |\{i = 1, \dots, n : x_i > 0\}|.$$

Reduktion der Zahl verwendeter Basisfunktionen

Wenn man weiß, dass sich das Signal nur aus wenigen Basisfunktionen zusammensetzt oder zur Datenkompression nur wenige verwenden will, möchte man ein gemischt-ganzzahliges nichtlineares Problem lösen:

$$\min_{x \in \mathbb{R}^+} \|Ax - b\| + \gamma |\{i = 1, \dots, n : x_i > 0\}|.$$

Dies ist zu aufwendig, daher gibt man sich mit der 1-Norm zufrieden,

$$\min_{x \in \mathbb{R}^+} \|Ax - b\| + \gamma \sum_{i=1}^n |x_i|,$$

und hofft, dass diese den Wert auf nur wenige Koordinaten konzentriert.

Übung: Erstelle das SOCP zur 1-Norm Variante.

Reduktion der Zahl verwendeter Basisfunktionen

Wenn man weiß, dass sich das Signal nur aus wenigen Basisfunktionen zusammensetzt oder zur Datenkompression nur wenige verwenden will, möchte man ein gemischt-ganzzahliges nichtlineares Problem lösen:

$$\min_{x \in \mathbb{R}^+} \|Ax - b\| + \gamma |\{i = 1, \dots, n : x_i > 0\}|.$$

Dies ist zu aufwendig, daher gibt man sich mit der 1-Norm zufrieden,

$$\min_{x \in \mathbb{R}^+} \|Ax - b\| + \gamma \sum_{i=1}^n |x_i|,$$

und hofft, dass diese den Wert auf nur wenige Koordinaten konzentriert.

Übung: Erstelle das SOCP zur 1-Norm Variante.

Bei bikriteriellen Aufgaben bietet es sich an, das Problem für mehrere Parameterwerte $\gamma > 0$ zu lösen und die jeweils erzeugten Funktionswertpaare in einer Graphik miteinander zu vergleichen.

Inhaltsübersicht für heute:

Innere-Punkte-Verfahren

Lineare Optimierung über Kegeln

Second-Order-Cone Programme

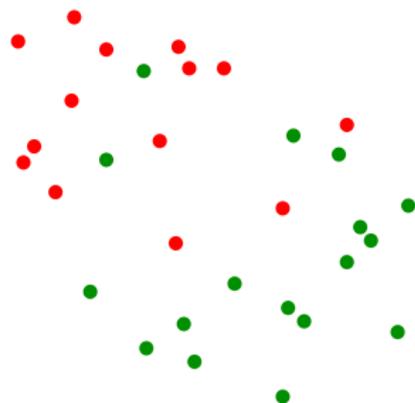
SOC Anwendung: Regularisierung

SOC Anwendung: Klassifizierung, Support-Vektor

SOC Anwendung: Das Markowitz Modell, Chance Constraints

Klassifizierung, Support-Vektor

Für Datenpunkte im \mathbb{R}^n , die eine bestimmte Eigenschaft haben bzw. nicht haben, soll eine Ebene gefunden werden, die die Punkttypen „möglichst gut“ voneinander trennt (Ziel: neue Punkte klassifizieren).



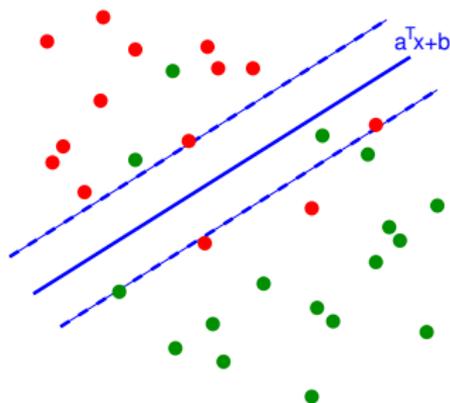
Klassifizierung, Support-Vektor

Für Datenpunkte im \mathbb{R}^n , die eine bestimmte Eigenschaft haben bzw. nicht haben, soll eine Ebene gefunden werden, die die Punkttypen „möglichst gut“ voneinander trennt (Ziel: neue Punkte klassifizieren).

Gegeben zwei disjunkte endliche Mengen $G, R \subset \mathbb{R}^n$, finde $a^T x + b$ (mit Variablen a und b) mit „möglichst“ $a^T x + b \geq 1$ für $x \in G$ und $a^T x + b \leq -1$ für $x \in R$.

Schwierigkeiten:

- Für gute Trennung soll $\|a\|$ klein sein.
- Was tun, wenn Fehlklassifizierungen unvermeidbar sind?



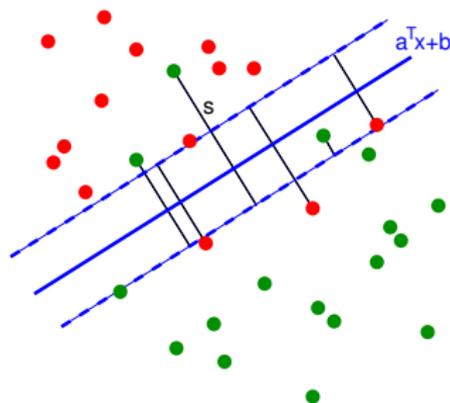
Klassifizierung, Support-Vektor

Für Datenpunkte im \mathbb{R}^n , die eine bestimmte Eigenschaft haben bzw. nicht haben, soll eine Ebene gefunden werden, die die Punkttypen „möglichst gut“ voneinander trennt (Ziel: neue Punkte klassifizieren).

Gegeben zwei disjunkte endliche Mengen $G, R \subset \mathbb{R}^n$, finde $a^T x + b$ (mit Variablen a und b) mit „möglichst“ $a^T x + b \geq 1$ für $x \in G$ und $a^T x + b \leq -1$ für $x \in R$.

Schwierigkeiten:

- Für gute Trennung soll $\|a\|$ klein sein.
- Was tun, wenn Fehlklassifizierungen unvermeidbar sind?



Ein Ansatz: Minimiere gleichzeitig $\|a\|$ und die Summe der Verletzungen der Ungleichungsbedingungen, skalarisiert mit Parameter $\gamma > 0$,

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

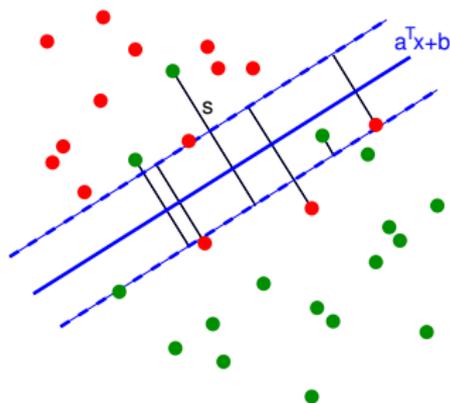
Klassifizierung, Support-Vektor

Für Datenpunkte im \mathbb{R}^n , die eine bestimmte Eigenschaft haben bzw. nicht haben, soll eine Ebene gefunden werden, die die Punktypen „möglichst gut“ voneinander trennt (Ziel: neue Punkte klassifizieren).

Gegeben zwei disjunkte endliche Mengen $G, R \subset \mathbb{R}^n$, finde $a^T x + b$ (mit Variablen a und b) mit „möglichst“ $a^T x + b \geq 1$ für $x \in G$ und $a^T x + b \leq -1$ für $x \in R$.

Schwierigkeiten:

- Für gute Trennung soll $\|a\|$ klein sein.
- Was tun, wenn Fehlklassifizierungen unvermeidbar sind?



Ein Ansatz: Minimiere gleichzeitig $\|a\|$ und die Summe der Verletzungen der Ungleichungsbedingungen, skalarisiert mit Parameter $\gamma > 0$,

$$\begin{array}{ll}
 \min & \|a\| + \gamma \sum_{x \in G \cup R} s_x \\
 \text{s.t.} & x^T a - b \geq 1 - s_x \quad x \in G \\
 & x^T a - b \leq s_x - 1 \quad x \in R \\
 & a \in \mathbb{R}^n, b \in \mathbb{R}, s \in \mathbb{R}_+^{G \cup R}
 \end{array}
 \quad \rightarrow \quad
 \begin{array}{ll}
 \min & a_0 + \gamma \sum_{x \in G \cup R} s_x \\
 \text{s.t.} & x^T a - b \geq 1 - s_x \quad x \in G \\
 & x^T a - b \leq s_x - 1 \quad x \in R \\
 & \begin{bmatrix} a_0 \\ a \end{bmatrix} \geq 0, b \in \mathbb{R}, s \geq 0
 \end{array}$$

Inhaltsübersicht für heute:

Innere-Punkte-Verfahren

Lineare Optimierung über Kegeln

Second-Order-Cone Programme

SOC Anwendung: Regularisierung

SOC Anwendung: Klassifizierung, Support-Vektor

SOC Anwendung: Das Markowitz Modell, Chance Constraints

Das Markowitz Modell

Im Markowitz-Modell der Portfolio-Optimierung soll gegebenes Kapital so mit einem erwarteten Mindestgewinn investiert werden, dass das Risiko minimiert wird.

Das Markowitz Modell

Im Markowitz-Modell der Portfolio-Optimierung soll gegebenes Kapital so mit einem erwarteten Mindestgewinn investiert werden, dass das Risiko minimiert wird.

$x \in \mathbb{R}_+^n$ mit $\mathbf{1}^T x = 1$ gibt den Anteil des Budgets an, der in Aktien $1, \dots, n$ investiert wird. Der Gewinn g pro Anteil ist eine Zufallsvariable mit Erwartungswert $\bar{g} \in \mathbb{R}^n$ und Kovarianzmatrix $G \in S_+^n$ ($n \times n$, positiv semidefinit), $s \in \mathbb{R}$ sei ein gegebener Schwellwert. Als Maß für das Risiko sieht das Markowitz-Modell $x^T G x$ vor. [Andere Maße möglich!]

$$\begin{aligned} \min \quad & x^T G x \\ \text{s.t.} \quad & \bar{g}^T x \geq s \\ & \mathbf{1}^T x = 1 \\ & x \in \mathbb{R}_+^n \end{aligned}$$

Das Markowitz Modell

Im Markowitz-Modell der Portfolio-Optimierung soll gegebenes Kapital so mit einem erwarteten Mindestgewinn investiert werden, dass das Risiko minimiert wird.

$x \in \mathbb{R}_+^n$ mit $\mathbf{1}^T x = 1$ gibt den Anteil des Budgets an, der in Aktien $1, \dots, n$ investiert wird. Der Gewinn g pro Anteil ist eine Zufallsvariable mit Erwartungswert $\bar{g} \in \mathbb{R}^n$ und Kovarianzmatrix $G \in S_+^n$ ($n \times n$, positiv semidefinit), $s \in \mathbb{R}$ sei ein gegebener Schwellwert. Als Maß für das Risiko sieht das Markowitz-Modell $x^T G x$ vor. [Andere Maße möglich!]

$$\begin{aligned} \min \quad & x^T G x \\ \text{s.t.} \quad & \bar{g}^T x \geq s \\ & \mathbf{1}^T x = 1 \\ & x \in \mathbb{R}_+^n \end{aligned}$$

Wegen G positiv semidefinit ist das ein konvex quadratisches Problem.

[Das bikriterielle Ziel Gewinn gegen Risiko wird hier durch eine Nebenbedingung an eines der Kriterien umgesetzt.]

Modellierung als SOCP?

Quadratische Nebenbedingungen mit SOCP

Sei $Q \in S_+^n$ positiv definit, $q \in \mathbb{R}^n$, $d \in \mathbb{R}$. Die konvex quadratische Nebenbedingung

$$\frac{1}{2}x^T Qx + q^T x + d \leq 0$$

lässt sich mit $Q = Q^{\frac{1}{2}} Q^{\frac{1}{2}}$ über

$$\|Q^{\frac{1}{2}}x + Q^{-\frac{1}{2}}q\| \leq \sqrt{q^T Q^{-1}q - 2d}$$

als SOC-Nebenbedingung schreiben (Beweis: beide Seiten quadrieren).

[Eigenwertzerlegung $Q = P\Lambda P^T \rightarrow Q^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P^T$, $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n) \geq 0$]

Quadratische Nebenbedingungen mit SOCP

Sei $Q \in S_+^n$ positiv definit, $q \in \mathbb{R}^n$, $d \in \mathbb{R}$. Die konvex quadratische Nebenbedingung

$$\frac{1}{2}x^T Qx + q^T x + d \leq 0$$

lässt sich mit $Q = Q^{\frac{1}{2}} Q^{\frac{1}{2}}$ über

$$\|Q^{\frac{1}{2}}x + Q^{-\frac{1}{2}}q\| \leq \sqrt{q^T Q^{-1}q - 2d}$$

als SOC-Nebenbedingung schreiben (Beweis: beide Seiten quadrieren).

[Eigenwertzerlegung $Q = P\Lambda P^T \rightarrow Q^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P^T$, $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_n) \geq 0$]

Für das Markowitz-Modell ist es einfacher, $x_0 \geq \|G^{\frac{1}{2}}x\|$ reicht aus

$$\begin{aligned} \min \quad & x_0 \\ \text{s.t.} \quad & \bar{x} = G^{\frac{1}{2}}x \\ & \bar{g}^T x \geq s \\ & \mathbf{1}^T x = 1 \\ & \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} \geq_Q 0, x \geq 0 \end{aligned}$$

Probabilistische Nebenbedingung, Chance Constraint

Unter der Annahme, dass g mit Varianz G um \bar{g} normalverteilt ist, sollen nun zusätzlich nur Investitionspläne erlaubt werden, für die die Wahrscheinlichkeit, dass der Gewinn über einem Schwellwert $\underline{s} < s$ bleibt, mindestens $\eta \in (0, 1)$ ist,

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

Probabilistische Nebenbedingung, Chance Constraint

Unter der Annahme, dass g mit Varianz G um \bar{g} normalverteilt ist, sollen nun zusätzlich nur Investitionspläne erlaubt werden, für die die Wahrscheinlichkeit, dass der Gewinn über einem Schwellwert $\underline{s} < s$ bleibt, mindestens $\eta \in (0, 1)$ ist,

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

Dies modelliert man mit einer Technik der robusten Optimierung: $g^T x \geq \underline{s}$ wird als Ungleichung mit unsicheren Koeffizienten aufgefasst.

Nebenbedingungen mit unsicheren Koeffizienten

Ist in einer Ungleichung $a^T x \leq b$ über die Koeffizienten nur bekannt, dass $a \in \{\bar{a} + Hu : \|u\| = 1\}$ für gegebenes $H \in S_+^n$ (pos. semidef), und soll x diese Ungleichung für all diese a erfüllen, so muss

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

gelten.

Nebenbedingungen mit unsicheren Koeffizienten

Ist in einer Ungleichung $a^T x \leq b$ über die Koeffizienten nur bekannt, dass $a \in \{\bar{a} + Hu : \|u\| = 1\}$ für gegebenes $H \in S_+^n$ (pos. semidef), und soll x diese Ungleichung für all diese a erfüllen, so muss

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

gelten. Letztere Ungleichung ist als SOC darstellbar in der Form

$$\begin{aligned} \xi_0 &= b - \bar{a}^T x \\ \xi &= Hx \\ \begin{pmatrix} \xi_0 \\ \xi \end{pmatrix} &\in SOC_n = \left\{ \begin{pmatrix} \xi_0 \\ \xi \end{pmatrix} : \|\xi\| \leq \xi_0, \xi \in \mathbb{R}^n \right\} \end{aligned}$$

Nebenbedingungen mit unsicheren Koeffizienten

Ist in einer Ungleichung $a^T x \leq b$ über die Koeffizienten nur bekannt, dass $a \in \{\bar{a} + Hu : \|u\| = 1\}$ für gegebenes $H \in S_+^n$ (pos. semidef), und soll x diese Ungleichung für all diese a erfüllen, so muss

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

gelten. Letztere Ungleichung ist als SOC darstellbar in der Form

$$\begin{aligned} \xi_0 &= b - \bar{a}^T x \\ \xi &= Hx \\ \begin{pmatrix} \xi_0 \\ \xi \end{pmatrix} &\in SOC_n = \left\{ \begin{pmatrix} \xi_0 \\ \xi \end{pmatrix} : \|\xi\| \leq \xi_0, \xi \in \mathbb{R}^n \right\} \end{aligned}$$

Für die probabilistische Interpretation sei g normalverteilt um \bar{g} mit Kovarianzmatrix $G = H^2$ und $g^T x \geq \underline{s}$ soll mit Wahrscheinlichkeit $0 < \eta < 1$ erfüllt sein, dann entspricht $\mathbb{P}(g^T x \geq \underline{s}) \geq \eta$ der Nebenbedingung $-\bar{g}^T x + \Phi^{-1}(\eta)\|Hx\| \leq -\underline{s}$. [\[\$\Phi\$... Normalverteilung\]](#)