Optimierung I

Prof. Dr. Christoph Helmberg

15. Februar 2004

Inhaltsverzeichnis

Kapitel 0

Einführung

Optimierung beschäftigt sich mit Aufgaben

$$(OA)\begin{cases} \min f(x) & \leftarrow \text{Zielfunktion} \\ \text{s.t.} & \leftarrow \text{subject to} \\ h_i(x) = 0, i \in E & \leftarrow \text{Gleichheitsnebenbedingungen} \\ g_i(x) \leq 0, i \in I & \leftarrow \text{Ungleichungsnebenbedingungen} \\ x \in \Omega & \leftarrow \text{Grundmenge,} \end{cases}$$

wobei meist $\Omega \subseteq \mathbb{R}^n$
$f, h, g : \mathbb{R}^n \to \mathbb{R}$
$|E|, |I| < \infty$

Definition 0.0.1 (Lösungsbegriffe).

- *Für eine Optimierungsaufgabe der Form (OA) heißt*

  $$\mathcal{X} = \{x \in \Omega : h_i(x) = 0 \; \forall i \in E, g_i(x) \leq 0 \; \forall i \in I\}$$

  die Menge der zulässigen Punkte/Lösungen oder die zulässige Menge (feasible set).

- *Falls $\mathcal{X} = \varnothing$ heißt das Problem unzulässig (infeasible).*
- *Ein $x \in \mathcal{X}$ heißt zulässiger Punkt/zulässige Lösung oder einfach zulässig.*
- *Ein $x^\star \in \mathcal{X}$ heißt globales Optimum oder globale Optimallösung, falls*

  $$f(x^\star) \leq f(x) \; \forall x \in \mathcal{X}.$$

- *Ein $x^\star \in \mathcal{X}$ heißt lokales Optimum, wenn es eine Umgebung $U_\varepsilon(x^\star)$ gibt mit*

  $$f(x^\star) \leq f(x) \; \forall x \in \mathcal{X} \cap U_\varepsilon(x^\star).$$

- *Der Wert $f^\star = \inf\{f(x) : x \in \mathcal{X}\}$ heißt Optimalwert.*
  Er ist ∞, falls $\mathcal{X} = \varnothing$. Ist der Wert $-\infty$, so heißt das Problem unbeschränkt.

Fragestellungen: Wann existieren und wie erkennt man Optimallösungen?
Wie bestimmt man sie algorithmisch (möglicherweise annähernd)?
Wie effizient sind die Algorithmen?

Ganz allgemein ist da nicht viel zu machen, selbst ohne Nebenbedingungen. Daher untersucht man Optimierungsprobleme mit Zusatzforderungen an f, g, h und Ω.

0.1 Konvexe Optimierung

f, g_i sind konvexe Funktionen.
h_i sind lineare Funktionen.
Ω ist konvexe Menge.

Definition 0.1.1. *Eine Menge $C \subseteq \mathbb{R}^n$ heißt konvex, falls mit zwei Elementen $x, y \in C$ die gesamte Verbindungsstrecke $[x, y]$ in C ist, d.h. wenn gilt:*

$$x, y \in C \Rightarrow \alpha x + (1-\alpha)y \in C \; \forall \alpha \in (0,1).$$

Beispiel 0.1.1. *$\varnothing$; $\mathbb{R}^n$; für $a \in \mathbb{R}^n, b \in \mathbb{R}$ der Halbraum $H_{a,b} := \{x \in \mathbb{R}^n : a^\tau x \leq b\}$; affine Unterräume*

Definition 0.1.2 (Konvexe Funktionen).

- *Eine Funktion $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ heißt konvex, wenn $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \forall \alpha \in (0,1), x, y \in \mathbb{R}$.*
- *Sie heißt eigentlich konvex, falls f konvex und $\exists x \in \mathbb{R}^n : f(x) < \infty$.*
- *Sie heißt streng konvex, falls die strenge Ungleichung gilt.*
- *Sie heißt konkav, wenn $-f$ konvex ist.*
- *Die Funktion f heißt quasikonvex, falls $f(\alpha x + (1-\alpha)y) \leq \max\{f(x), f(y)\} \forall \alpha \in (0,1)$.*
- *Sie heißt streng quasikonvex, wenn die strenge Ungleichung gilt.*

Definition 0.1.3.

- *Der Epigraph einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist die Menge*

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

- *Die Niveaumenge einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ zum Niveau r ist die Menge*

$$S_r(f) := \{x \in \mathbb{R}^n : f(x) \leq r\}.$$

Beobachtung 0.1.1. $f : \mathbb{R}^n \to \mathbb{R} \iff \text{epi}f$ *konvex.*

Beweis. $''\Rightarrow''$ Seien $(x,r),(y,p) \in \text{epi}f \Rightarrow f(x) \leq r, f(y) \leq p$. Für alle $\alpha \in (0,1)$ gilt nun

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha r + (1-\alpha)p,$$

also ist $(\alpha x + (1-\alpha)y, \alpha r + (1-\alpha)p) = \alpha(x,r) + (1-\alpha)(y,p) \in \text{epi}f$. $''\Leftarrow''$ Mit $(x, f(x))$ und $(y, f(y))$ ist für $\alpha \in (0,1)$

$$(\alpha x + (1-\alpha)y, \alpha f(x) + (1-\alpha)f(y)) \in \text{epi}f,$$

also ist $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)\ \forall \alpha \in (0,1)$. □

Beobachtung 0.1.2. *Der Schnitt einer Familie konvexer Mengen ist konvex.*

Beweis. Seien $x, y \in C := \bigcap_{i\in I} C_i$ mit C_i konvex für $i \in I$. Dann sind $x, y \in C_i\ \forall i \in I$, also auch $[x,y] \subseteq C_i\ \forall i \in I$ und damit $[x,y] \subseteq C$. □

Beobachtung 0.1.3. *Die Niveaumengen einer konvexen Funktion sind konvex.*

Beweis. Bilde $S'_r = \text{epi}f \cap \{(x,r) : x \in \mathbb{R}^n\}$; S'_r ist nach Beobachtung 0.1.2 konvex.
$S_r(f) = \{x : (x,r) \in S'_r\}$ verbleibt dem Leser als Übung. □

Wir betrachten jetzt:

$$\begin{cases} \min f(x) \\ h_i(x) = 0, i \in E \\ g_i(x) \leq 0, i \in I \\ x \in \Omega \end{cases}$$

Die zulässige Menge

$$\mathcal{X} = \underbrace{\Omega}_{\textit{konvex}} \cap \bigcap_{i\in E} \underbrace{\{x : h_i(x) = 0\}}_{\textit{affiner Unterraum}} \cap \bigcap_{i\in I} \underbrace{\{x : g_i(x) \leq 0\}}_{\textit{nach Beobachtung 0.1.3 konvex}}$$

ist nach Beobachtung 0.1.2 konvex.

Satz 0.1.4. *In einem konvexen Optimierungsproblem ist jedes lokale Optimum auch globales Optimum.*

Beweis. Sei $\overline{x}$ ein lokales und $x^\star$ ein globales Optimum und $x^\star \neq \overline{x}$. Da $\mathcal{X}$ konvex ist, ist auch $[\overline{x}, x^\star]$ zulässig. Wähle $x \in (\overline{x}, x^\star) \cap U(\overline{x})$; dann existiert $\alpha \in (0,1) : x = \alpha\overline{x} + (1-\alpha)x^\star$ und

$$f(x) = f(\alpha\overline{x} + (1-\alpha)x^\star) \overset{f\text{konvex}}{\leq} \alpha f(\overline{x}) + (1-\alpha)f(x^\star) \overset{x^\star \text{ global optimal}}{\leq} f(\overline{x}) \overset{\overline{x} \text{ lokal optimal}}{\leq} f(x),$$

also ist $f(\overline{x}) = f(x^\star) = f(x)$. □

Insbesondere gilt: Die Menge der Optimallösungen ist konvex; denn sei $\mathcal{X}$ die zulässige Menge und $f^\star = \inf\{f(x) : x \in \mathcal{X}\}$, dann ist die Menge der Optimallösungen $S_{f^\star}(f) \cap \mathcal{X}$ als Schnitt zweier konvexer Mengen konvex. Sie kann aber leer sein, selbst wenn $f^\star$ endlich ist.

Beispiel: $f : \mathbb{R} \to \mathbb{R}; \mathcal{X} = \mathbb{R}_+; f(x) = \frac{1}{x}$

Die Existenz von Optimallösungen hat immer etwas mit der Abgeschlossenheit der Mengen und Epigraphen zu tun. Wir werden uns bald mit einem besonders wichtigen Spezialfall beschäftigen, wo das kein Problem ist.

0.1.1 Lineare Optimierung

f, g, h sind lineare Funktionen, also von der Form $a^\tau x + b$, Ω ist der nichtnegative Orthant.
"Lineares Programm in Standardform": $\min c^\tau x$
s.t. $Ax = b$
$x \geq 0.$

0.1.2 Nichtlineare Optimierung

Die Funktionen müssen genügend glatt sein, Ω ist meist $\mathbb{R}^n$. Einteilung in:

- freie/unrestringierte nichtlineare Optimierung (unconstrained optimisation): $|E| = |I| = 0, \Omega = \mathbb{R}^n$
- restringierte nichtlineare Optimierung (contrained optimisation) $|E| + |I| > 0$

Der Anspruch beschränkt sich auf das Auffinden eines lokalen Optimums. In der nichtlinearen Optimierung kreist man die richtige Kombination von Abstiegsrichtungen mit dem *Newton-Verfahren* ein.

0.2 Optimalitätsbedingungen für freie nichtlineare Optimierung

Betrachten für $f \in C^2(\mathbb{R}^n)$ die Aufgabe $\min\limits_{x \in \mathbb{R}^n} f(x)$.

In einem Punkt $\overline{x}$ steht zur Verfügung:

- "lineares Modell": $f(\overline{x})$
  $\nabla f(\overline{x})$ ("steilster Anstieg")
  $f(\overline{x}) + \nabla f(\overline{x})^\tau (x - \overline{x})$ beschreibt die Tangentialebene

- "quadratisches Modell": zusätzlich $\nabla^2 f(\overline{x})$ (2. Ableitung; Hessematrix; symmetrisch)
  $f(\overline{x}) + \nabla f(\overline{x})^\tau (x - \overline{x}) + \frac{1}{2}(x - \overline{x})^\tau \nabla^2 f(\overline{x})(x - \overline{x})$

Satz von Taylor: *Seien $x, p \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$, so gilt für ein $t \in (0,1)$:*

$$f(x+p) = f(x) + \nabla f(x+tp)^\tau p.$$

Ist $f \in C^2(\mathbb{R}^n)$, so gilt für ein $t \in (0,1)$:

$$\nabla f(x+p) = \nabla f(x) + \textstyle\int_0^1 \nabla^2 f(x+sp) p \mathrm{d}s \qquad (MWS)$$
$$f(x+p) = f(x) + \nabla f(x)^\tau p + \tfrac{1}{2} p^\tau \nabla^2 f(x+tp) p.$$

Satz 0.2.1 (Notwendige Bedingung 1. Ordnung). *Ist $x^\star$ ein lokales Minimum und f stetig differenzierbar auf einer Umgebung $U(x^\star)$ von $x^\star$, dann ist $\nabla f(x^\star) = 0$.*

Beweis. Sei $\nabla f(x^\star) \neq 0$, setze $p = -\nabla f(x^\star)$. Daher ist $p^\tau \nabla f(x^\star) = -\|\nabla f(x^\star)\|^2_{(L_2)} < 0$. Da $\nabla f(x^\star)$ stetig, existiert ein $\alpha > 0$, so dass $p^\tau \nabla f(x^\star + tp) < 0 \; \forall t \in (0, \alpha)$. Für beliebiges $t \in (0, \alpha]$ ist nach Taylor nun für ein gewisses $\overline{t} \in (0, t)$

$$f(x^\star + tp) = f(x^\star) + \underbrace{t p^\tau \nabla f(x^\star + \overline{t} p)}_{<0} < f(x^\star)$$

$\Rightarrow$ Widerspruch. □

Definition 0.2.1. *Ist $\nabla f(\overline{x}) = 0$, so nennt man $\overline{x}$ einen stationären Punkt von f.*

Satz 0.2.2 (Notwendige Bedingung 2. Ordnung). *Ist $x^\star$ ein lokales Minimum von f und $\nabla^2 f$ ist stetig auf einer Umgebung $U(x^\star)$, dann ist $\nabla f(x^\star) = 0$ und $\nabla^2 f(x^\star)$ positiv semidefinit.*

Beweis. Nach Satz 0.2.1 folgt $\nabla f(x^\star) = 0$.
Sei $\nabla^2 f(x^\star)$ nicht positiv semidefinit $\Rightarrow \exists p : p^\tau \nabla^2 f(x^\star) p < 0$. Wegen der Stetigkeit von $\nabla^2 f$ existiert ein $\alpha > 0$, so dass $p^\tau \nabla^2 f(x^\star + tp) p < 0 \; \forall t \in (0, \alpha)$. Für beliebiges $t \in (0, \alpha]$ ist nach Taylor nun für ein gewisses $\overline{t} \in (0, t)$

$$f(x^\star + tp) = f(x^\star) + tp^\tau \underbrace{\nabla f(x^\star)}_{=0} + t^2 \underbrace{p^\tau \nabla^2 f(x^\star + \overline{t} p) p}_{<0} < f(x^\star).$$

□

Dies ist im allgemeinen nicht hinreichend, z.B.: $f(x,y) = x^2 - y^4$

$$\nabla f(0,0) = \begin{pmatrix} 2x \\ -4y^3 \end{pmatrix}\Big|_{(0,0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \succeq 0$$

Satz 0.2.3 (Hinreichende Bedingung). *$\nabla^2 f$ sei stetig auf einer Umgebung $U(x^\star)$. Ist dann $\nabla f(x^\star) = 0$ und $\nabla^2 f(x^\star)$ positiv definit, dann ist $x^\star$ ein lokales Minimum.*

Beweis. Da $\nabla^2 f$ stetig auf $U(x^\star)$, existiert ein $r > 0$, so dass $\nabla^2 f(x)$ positiv definit $\forall x \in B_r(x^\star) = \{x : \|x - x^\star\| < r\}$. Für beliebiges $p \neq 0$ mit $x + p \in B_r(x^\star)$ gilt nach Taylor für ein $t \in (0, 1)$:

$$f(x^\star + p) = f(x^\star) + p^\tau \underbrace{\nabla f(x^\star)}_{=0} + \underbrace{\frac{1}{2} p^\tau \nabla^2 f(x^\star + tp) p}_{>0} > f(x^\star).$$

□

Für konvexe differenzierbare Funktionen ist die Lage viel leichter!

Satz 0.2.4. *Ist f konvex und differenzierbar und ist $x^\star$ stationärer Punkt, so ist $x^\star$ globales Minimum.*

Beweis. Ann.: $\exists \overline{x}$ mit $f(\overline{x}) < f(x^\star)$. Dann ist

$$f(x^\star + t(\overline{x} - x^\star)) \leq (1 - t) f(x^\star) + t f(\overline{x}) = f(x^\star) + t(f(\overline{x}) - f(x^\star))$$

und

$$0 = \nabla f(x^\star)^\tau (\overline{x} - x^\star) = \lim_{t \to 0} \tfrac{f(x^\star + t(\overline{x} - x^\star)) - f(x^\star)}{t} \leq \lim_{t \to 0} \tfrac{t(f(\overline{x}) - f(x^\star))}{t} = (f(\overline{x}) - f(x^\star)) < 0.$$

□

0.3 Methoden der freien nichtlinearen Optimierung

In der nichtlinearen Optimierung

- versucht man mit Abstiegsverfahren möglichst nah an ein lokales Optimum zu kommen,
- hofft, dass dort die Funktion lokal streng konvex ist und das quadratische Modell eine gute Approximation darstellt,
- bestimmt iterativ das Minimum des quadratischen Modells.

Satz 0.3.1. *Sei $f \in C^2$, $x^\star$ genüge den hinreichenden Optimalitätsbedingungen aus Satz 0.2.3 und $\nabla^2 f$ sei Lipschitzstetig in einer Umgebung $U(x^\star)$, das heißt:*

$$\exists L > 0 : \|\nabla^2 f(x) - \nabla^2 f(\overline{x})\| \leq L \|x - \overline{x}\| \; \forall x, \overline{x} \in U(x^\star).$$

Dann gilt für die durch $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ festgelegte Folge:

1. *Ist Startpunkt x_0 nahe genug bei $x^\star$, dann $x_k \to x^\star$.*
2. *Die Konvergenzrate ist dann quadratisch, also $\|x_{k+1} - x^\star\| \leq c\|x_k - x^\star\|^2$ für ein $c > 0$.*
3. *Die Folge der Gradientennormen $\|\nabla f(x_k)\|$ konvergiert quadratisch gegen 0.*

Interpretation: Für x_k nahe genug an $x^\star$ ist die Hessematrix des quadratischen Modells $\nabla^2 f(x_k)$ positiv definit; folglich ist

$$q(x) = f(x_k) + \nabla f(x_k)^\tau (x - x_k) + \tfrac{1}{2}(x - x_k)^\tau \nabla^2 f(x_k)(x - x_k)$$

streng konvex und hat eine eindeutige Optimallösung, den stationären Punkt x von q:

$$0 = \nabla q(x) = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) \quad \Rightarrow \quad x = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

Alternativ kann man das Newton-Verfahren als Methode zur Bestimmung der Nullstelle einer vektorwertigen Funktion $F : \mathbb{R}^n \to \mathbb{R}^m$ auffassen; gesucht ist x mit $F(x) = 0$. Baue lineares Modell an F in x_k und bestimme dessen Nullstelle: $F(x_k) + J_F(x_k)(x - x_k) = 0$. Ist $J_F(x_k)$ invertierbar, so ist $x = x_k - J_F(x_k)^{-1} F(x_k)$. Für die Optimierung: $F = \nabla f$. Das Newton-Verfahren sucht einen stationären Punkt, der in der Nähe des Startpunktes liegt.

Beweis von Satz 0.3.1. Sei $x^\star$ die Optimallösung; $\nabla f_k := \nabla f(x_k)$, $\nabla^2 f_k := \nabla^2 f(x_k)$.

$$\begin{aligned}
x_{k+1} - x^\star &= x_k - \nabla^2 f_k^{-1} \nabla f_k - x^\star \\
&= \nabla^2 f_k^{-1} (\nabla^2 f_k (x_k - x^\star) - (\nabla f_k - \underbrace{\nabla f_\star}_{=0})) \\
\nabla f_k - \nabla f_\star &= \textstyle\int_0^1 \nabla^2 f(x_k + t(x^\star - x_k))(x_k - x^\star) \mathrm{d}t \\
\|\nabla^2 f_k (x_k - x^\star) - (\nabla f_k - \nabla f_\star)\| &= \|\textstyle\int_0^1 [\nabla^2 f_k - \nabla^2 f(x_k + t(x^\star - x_k))](x_k - x^\star) \mathrm{d}t\| \\
&\leq \textstyle\int_0^1 \|\nabla^2 f_k - \nabla^2 f(x_k + t(x^\star - x_k))\| \|x_k - x^\star\| \mathrm{d}t \\
&\leq L\|x_k - x^\star\|^2 \textstyle\int_0^1 t \mathrm{d}t = \frac{1}{2} L \|x_k - x^\star\|^2
\end{aligned}$$

Schranke für $\|\nabla^2 f_k^{-1}\|$: $\nabla^2 f_k$ ist positiv definit und Lipschitzstetig, also existiert $r > 0$, so dass

$$\|\nabla^2 f^{-1}(x)\| \leq 2\|\nabla^2 f_\star^{-1}\| \quad \forall x \in U(x^\star) : \|x - x^\star\| \leq r.$$

Mit obigen Abschätzungen gilt nun

$$\|x_{k+1} - x^\star\| \leq L\|\nabla f_\star^{-1}\| \|x_k - x^\star\|^2 \quad \forall x \in U(x^\star) : \|x_k - x^\star\| \leq r.$$

Falls $\|x_0 - x^\star\| < \min\left\{r, \frac{1}{L\|\nabla^2 f_\star^{-1}\|}\right\}$, dann geht $x_k \to x^\star$ mit quadratischer Konvergenz $\Rightarrow$ 1. und 2. gelten.

3. $\|\nabla f_{k+1}\| = \|\nabla f_{k+1} - \underbrace{\nabla f_k - \nabla^2 f_k(x_{k+1} - x_k)}_{=0}\|$ □

$$
\begin{aligned}
&= \| \int_0^1 \nabla^2 f(x_k + t(x_{k+1} - x_k))(x_{k+1} - x_k)\mathrm{d}t - \nabla^2 f_k(x_{k+1} - x_k)\| \\
&\leq \int_0^1 \|\nabla^2 f(x_k + t(x_{k+1} - x_k)) - \nabla^2 f(x_k)\|\|x_{k+1} - x_k\|\mathrm{d}t \\
&\leq L\|x_{k+1} - x_k\|^2 \int_0^1 t\mathrm{d}t \\
&= \tfrac{1}{2}L\| \underbrace{x_{k+1} - x_k}_{=-\nabla^2 f_k^{-1}\nabla f_k} \|^2 \\
&\leq \tfrac{1}{2}L\|\nabla f_k^{-1}\|^2\|\nabla f_k\|^2 \\
&\leq 2L\|\nabla f_\star^{-1}\|^2\|\nabla f_k\|^2.
\end{aligned}
$$

Probleme bei Newton: lokales Verfahren,
im Allgemeinen keine Abstiegsrichtung.
Negativer Gradient ist immer Abstiegsrichtung!

Definition 0.3.1. *Eine Schrittrichtung p heißt Abstiegsrichtung in x, wenn $(\nabla f(x))^\tau p < 0$.*

Typische Optimierungsverfahren

- Berechne eine Schrittrichtung $p_k = -B_k^{-1}\nabla f_k$ mit B_k positiv definit, dann ist

  $$\nabla f_k^\tau p_k = -\nabla f_k^\tau B_k^{-1}\nabla f_k < 0, \text{wenn } \nabla f_k \neq 0.$$

- Suche entlang dem Halbstrahl $x_k + \alpha p_k$ annähernd das Minimum von f
  (man muss nur hinreichenden Abstieg garantieren, "sufficient decrease")

- Setze $x_{k+1} = x_k + \alpha_k p_k$. Nutze die Schrittinformation um B_k zu verbessern.

Klasse der "*line search*" Algorithmen

1. $B_k = I$ ist "*steepest descent*"
   Steilster Abstieg hat schlechte Konvergenzeigenschaften für konvexe quadratische Funktionen $f = \frac{1}{2}x^\tau Qx + c^\tau x$.
   Problem bei steepest decsent: stark skalierungsabhängig
   Vorteil bei Newton: $p_k = -(\nabla^2 f(x_k))^{-1}\nabla f_k = -Q^{-1}Qx_k = -x_k$

2. Setzt man $B_k = \nabla^2 f(x_k)$, dann erhält man das Newtonverfahren mit line search. Funktioniert, wenn $\nabla^2 f$ positiv definit, also bei lokal konvexem f. Sonst $B_k = \nabla^2 f + \lambda I$ mit $\lambda > -\min\{0, \lambda_{\min}(\nabla^2 f)\}$

3. Aktuelle Verfahren starten mit $B_k = I$ und versuchen sich eine immer bessere Approximation der Hessematrix aufzubauen ("*Quasi-Newton-Verfahren*").

Kapitel 1

Lineare Optimierung

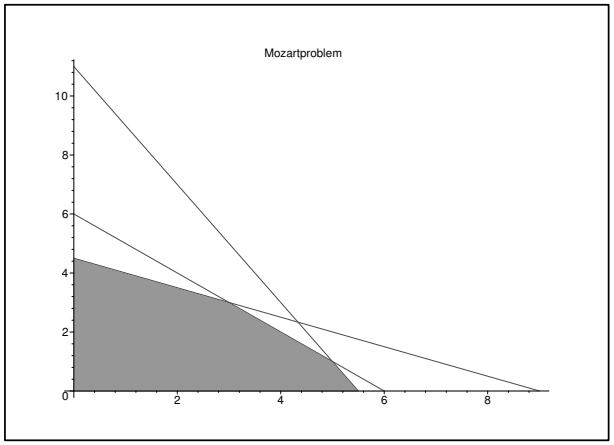
1.1 Lineare Programme

Beispiel 1.1.1. *Möchten Mozartkugeln und Mozarttaler produzieren; brauchen jeweils:*

	Marzipan	*Nougat*	*Bitterschokolade*	*Preis*
Mozartkugeln	*1*	*2*	*1*	*9*
Mozarttaler	*1*	*1*	*2*	*8*
Gesamtmenge	*6*	*11*	*9*	

x_1 *: Menge an Mozartkugeln*
x_2 *: Menge an Mozarttalern*

$$\begin{aligned} \max\ & 9x_1 + 8x_2 \\ \text{s.t.}\ & x_1 + x_2 \leq 6 \\ & 2x_1 + x_2 \leq 11 \\ & x_1 + 2x_2 \leq 9 \\ & x_1 \geq 0; x_2 \geq 0 \end{aligned}$$



"Offensichtlich" ist die Optimallösung eine Ecke. Wie berechnet man die Optimallösung? Bestimme den Schnittpunkt der beiden Geraden!

$$\begin{cases} x_1 + x_2 = 6 \\ 2x_1 + x_2 = 11 \end{cases} \quad \Rightarrow \quad x_1 = 5, x_2 = 1$$

Realistische Probleme haben bis zu einigen Millionen Variablen und Nebenbedingungen
$\Rightarrow$ Brauchen algebraische Verfahren!

- Variablen: Vektor $x \in \mathbb{R}^n$
  oft verlangen wir $x_i \geq 0, i = 1, \ldots, n \Leftrightarrow x \geq 0$
- andere Nebenbedingungen: $\sum a_i x_i \leq \beta \Leftrightarrow a^\tau x \leq \beta$;
  $H_{a,\beta} = \{x \in \mathbb{R}^n : a^\tau x \leq \beta\}$
- Wir fassen mehrere Nebenbedingungen zu einer Matrixungleichung zusammen:

$$\begin{matrix} a_1^\tau x \leq b_1 \\ \vdots \\ a_m^\tau x \leq b_m \end{matrix} \quad \Leftrightarrow \quad Ax \leq b \text{ mit } A = \begin{bmatrix} a_1^\tau \\ \vdots \\ a_m^\tau \end{bmatrix}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Lineares Programm in kanonischer Form:

$$\begin{array}{ll} \max c^\tau x & \\ \text{s.t.} & Ax \leq b \\ & x \geq 0 \end{array}$$

Wie bringt man Probleme mit Gleichungen und Ungleichungen in diese Form?

- $a^\tau x \geq \beta \rightsquigarrow (-a)^\tau x \leq -\beta$
- $a^\tau x = \beta \rightsquigarrow \begin{cases} a^\tau x \geq \beta \rightsquigarrow (-a)^\tau x \leq -\beta \\ a^\tau x \leq \beta \end{cases}$
- negative Variable $x_i \leq 0$ ersetzen durch $\overline{x_i} = -x_i$
- freie Variable (x_i nicht Vorzeichen beschränkt) $x_i = x_i^+ - x_i^-, x_i^+, x_i^- \geq 0$
- Minimierungsproblem: $\min c^\tau x = -\max(-c)^\tau x$

Jede Aufgabe mit linearer Zielfunktion sowie linearen Gleichungs- und Ungleichungsnebenbedingungen kann als lineares Programm in kanonischer Form dargestellt werden.

Die "Schlupfvariablen" oder "slacks" $s_i = b_i - a_i^\tau x \geq 0$ messen die Distanz zur Ungleichung:

$$\begin{array}{|l|}\hline \max c^\tau x \\ \text{s.t. } Ax + s = b \\ \quad\quad x \geq 0, s \geq 0 \\ \hline \end{array}$$

mit $\overline{A} = [A\ I]$ und $\overline{x} = \begin{bmatrix} x \\ s \end{bmatrix}$:

$$\begin{array}{|l|}\hline \max c^\tau x \\ \text{s.t. } \overline{A}\overline{x} = b \\ \quad\quad \overline{x} \geq 0 \\ \hline \end{array}$$

Lineares Programm in Normalform:

$$\begin{array}{|l|}\hline \min c^\tau x \\ \text{s.t. } Ax = b \\ \quad\quad x \geq 0 \\ \hline \end{array}$$

O.B.d.A. hat A vollen Rang

Beispiel 1.1.2 (Mozartproblem in Normalform). *3 Schlupfvariablen (x_3, x_4, x_5) für 3 Ungleichungen*

$$\overline{A} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 6 \\ 11 \\ 9 \end{pmatrix}, c = \begin{pmatrix} -9 \\ -8 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

vorher: Optimallösung war Schnittpunkt der 1. und 2. Ungleichung
$\Rightarrow x_3 = 0, x_4 = 0$
Es bleibt zu lösen:

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_5 \end{pmatrix} = \begin{pmatrix} 6 \\ 11 \\ 9 \end{pmatrix}$$

"Ecke berechnen": Wähle $n - m$ Schlupfvariablen, die Null sein sollen. $\Rightarrow$ Lösung liegt auf den entsprechenden Nebenbedingungen. Lösung des restlichen Systems bestimmt Schlupf der anderen Nebenbedingungen.

Wofür ist x_1 der Schlupf? Für $x_1 \geq 0$.

$Ax = b$: m Gleichungen; $Ix \geq 0$: n Ungleichungen

Ecke wird durch n Gleichungen festgelegt, m stehen fest. $n-m$ aus den Ungleichungen auswählen: $N \subseteq \{1,\ldots,n\} : |N| = n-m$. Müssen $n-m$ so auswählen (alle $= 0$), dass durch das System $Ax = b$ die restlichen eindeutig bestimmt sind (sonst keine Ecke).

$$
\begin{aligned}
&B = \{1,\ldots,n\} \setminus N \\
&A = [A_B\ A_N], x = \begin{bmatrix} x_B \\ x_N \end{bmatrix} \\
&A_B x_B + A_N x_N = b, x_N = 0, A_B \text{ muss regulär sein} \\
&\Rightarrow x_B = A_B^{-1}[b - A_N x_N]
\end{aligned}
$$

Die Spalten von A_B müssen linear unabhängig sein, also eine Basis des $\mathbb{R}^m$ bilden.

Definition 1.1.1. *Für ein Ungleichungssystem $Ax = b, x \geq 0$ mit $A \in \mathbb{R}^{m\times n}$ heißt:*

- *eine reguläre Untermatrix A_B mit Spaltenindizes $B \subseteq \{1,\ldots,n\}, |B| = m$ Basis,*
- *$\overline{x} = A^{-1}[b - A_N \overline{x}_N]$ mit $\overline{x}_N = 0$ Basislösung (zur Basis B),*
- *und zulässige Basislösung, falls $\overline{x}_B \geq 0$.*
- *x_B heißen Basisvariable oder abhängige Variable.*
- *x_N heißen Nichtbasisvariable oder unabhängige Variable.*

Idee des Simplexalgorithmus für Lineare Programme
Springe von einer Ecke zu einer benachbarten, besseren Ecke, bis es keine bessere gibt.

1.2 Der Simplexalgorithmus

Beginnen mit der zulässigen Basislösung $x_B = A_B^{-1}[b - A_N x_N], x_N = 0, x_B \geq 0$. Versuchen zu verbessern, indem wir ein x_i mit $i \in N$ von 0 wegschieben.

$$
\begin{aligned}
c^\tau x &= c_B^\tau x_B + c_N^\tau x_N \\
&= c_B^\tau A_B^{-1}[b - A_N x_N] + c_N^\tau x_N \\
&= c_B^\tau A_B^{-1} b - c_B^\tau A_B^{-1} A_N x_N + c_N^\tau x_N \\
&= c_B^\tau A_B^{-1} + [c_N^\tau - c_B^\tau A_B^{-1} A_N] x_N \\
&= \underbrace{c_B^\tau A_B^{-1} b}_{\substack{\text{Konstante} \\ \text{derzeitiger ZFW}}} + [\underbrace{c_N - A_N^\tau A_B^{-\tau} c_B}_{\text{reduzierte Kosten:}=\tilde{c}_N}]^\tau x_N
\end{aligned}
$$

Reduzierte Kosten sagen aus, wie sich die Zielfunktion verändert, wenn ein x_j mit $j \in N$ vergrößert wird.

Vergrößern von x_j führt zur Verbesserung, falls $\tilde{c}_j < 0$ ist. Was ist falls $\tilde{c}_j \geq 0 \; \forall j \in N$?

Lemma 1.2.1. *Für das LP* $\min\{c^\tau x : Ax = b, x \geq 0\}$ *sei* $\overline{x}$ *eine zulässige Basislösung zur Basis* B*. Erfüllen die reduzierten Kosten*

$$c_N - A_N^\tau A_B^{-\tau} c_B \geq 0,$$

so ist $\overline{x}$ *eine Optimallösung.*

Beweis. Sei $x^\star$ eine Optimallösung, also $Ax^\star = b, x^\star \geq 0$ und $c^\tau x^\star \leq c^\tau x \; \forall x \in \mathcal{X} = \{x \geq 0 : Ax = b\}$. Es gilt $x_B^\star = A_B^{-1}[b - A_N x_N^\star]$ und $x_N^\star \geq \overline{x_N} = 0$ und

$$c^\tau x^\star \;\; = \;\; c_B^\tau A_B^{-1} b + (\underbrace{c_N - A_N^\tau A_B^{-\tau} c_B}_{\geq 0})^\tau x_N^\star \geq c_B^\tau A_B^{-1} b = c^\tau \overline{x}$$

□

Bemerkung 1.2.2. *Aus dem Beweis erkennt man, dass weitere Optimallösungen nur dann existieren können, wenn wenigstens eine Komponente der reduzierten Kosten* $= 0$ *ist.*

Wir nehmen an, der reduzierte-Kosten-Vektor enthält eine negative Komponente: $\exists \hat{\jmath} \in N : \tilde{c}_{\hat{\jmath}} < 0$. Die Berechnung der reduzierten Kosten und die Auswahl von $\hat{\jmath}$ nennt man "Pricing".

Um wieviel können wir $x_{\hat{\jmath}}$ vergrößern, ohne unzulässig zu werden? (Alle anderen x_j mit $j \in N \setminus \{\hat{\jmath}\}$ behalten den Wert 0.) Also: $x_B(x_{\hat{\jmath}}) = A_B^{-1}[b - A_{\cdot,\hat{\jmath}} x_{\hat{\jmath}}]$ soll ≥ 0 bleiben.

$$\Rightarrow \forall i \in \{1, \ldots, m\} : \underbrace{[A_B^{-1} b]_i}_{\geq 0} \geq \underbrace{[A_B^{-1} A_{\cdot,\hat{\jmath}}]_i}_{\text{nur} > 0 \text{ interessant}} \; x_{\hat{\jmath}}$$

$$x_{\hat{\jmath}} \leq \min \left\{ \tfrac{[A_B^{-1} b]_i}{[A_B^{-1} A_{\cdot,\hat{\jmath}}]_i} : i \in \{1, \ldots, m\} : [A_B^{-1} A_{\cdot,\hat{\jmath}}]_i > 0 \right\}.$$

Diesen Schritt nennt man "ratio test".

Was passiert, wenn die Menge derartiger Indizes i leer ist?
$\Rightarrow$ Das Problem ist unbeschränkt, in Richtung des Halbstrahls geht es gegen $-\infty$, denn $x_{\hat{\jmath}}$ kann beliebig vergrößert werden ohne den zulässigen Bereich zu verlassen: Der Halbstrahl

$$x(\alpha) = \begin{pmatrix} x_B \\ x_N \end{pmatrix} + \alpha \left[\begin{pmatrix} -A_B^{-1} A_{\cdot,\hat{\jmath}} \\ 0 \end{pmatrix} + e_{\hat{\jmath}} \right]$$

ist zulässig, also $x(\alpha) \in \mathcal{X} \; \forall \alpha > 0$ und es gilt

$$\inf\{c^\tau x(\alpha) = c_B^\tau A_B^{-1} b + \alpha(\underbrace{c_N - A_N^\tau A_B^{-\tau} c_B}_{< 0})_{\hat{\jmath}}, \alpha > 0\} = -\infty$$

Das Problem ist also unbeschränkt, falls es kein $i \in \{1, \ldots, m\}$ mit $[A_B^{-1} A_{.,\hat{\jmath}}]_i > 0$ gibt.

Nehmen wir nun an, $\hat{\imath}$ sei ein Index, für den obiges Minimum angenommen wird, zu $\hat{\imath}$ gehört eine Basisvariable mit Index $B(\hat{\imath})$. "Setzen"

$$x_{\hat{\jmath}} = \frac{[A_B^{-1} b]_{\hat{\imath}}}{[A_B^{-1} A_{.,\hat{\jmath}}]_{\hat{\imath}}} \geq 0,$$

dann wird $x_{B(\hat{\imath})} = 0$, also ergibt sich eine neue zulässige Basis mit

$$N^+ = N \setminus \{\hat{\imath}\} \cup \{B(\hat{\imath})\}$$
$$B^+ = B \setminus \{B(\hat{\imath})\} \cup \{\hat{\jmath}\}.$$

$x_{\hat{\jmath}}$ heißt eintretende Variable ("entering"); $x_{B(\hat{\imath})}$ heißt austretende Variable ("leaving").

Lemma 1.2.3. *Die neue Indexmenge B^+ beschreibt wieder eine zulässige Basis und*

$$x^+ = x + \frac{[A_B^{-1} b]_{\hat{\imath}}}{[A_B^{-1} A_{.,\hat{\jmath}}]_{\hat{\imath}}} \cdot \left[\begin{pmatrix} -A_B^{-1} A_{.,\hat{\jmath}} \\ 0 \end{pmatrix} + e_{\hat{\jmath}} \right] \geq 0$$

ist eine zulässige Basislösung.

Beweis. $A_B w = A_{.,\hat{\jmath}}$ hat eine eindeutige Lösung, da A_B Basis ist. Wegen $w_{\hat{\imath}} > 0$ ist $A_{.,\hat{\jmath}}$ linear unabhängig von den Spalten $B \setminus \{\hat{\imath}\}$, also ist A_{B^+} wieder eine Basis. Der Rest folgt nach Konstruktion. □

Definition 1.2.1. *Das Paar $(\hat{\imath}, \hat{\jmath})$ wird das Pivot und das entsprechende Element in der Matrix $A_B^{-1} A_N$ Pivot-Element genannt.*

1.2.1 Der (primale) Simplex Algorithmus

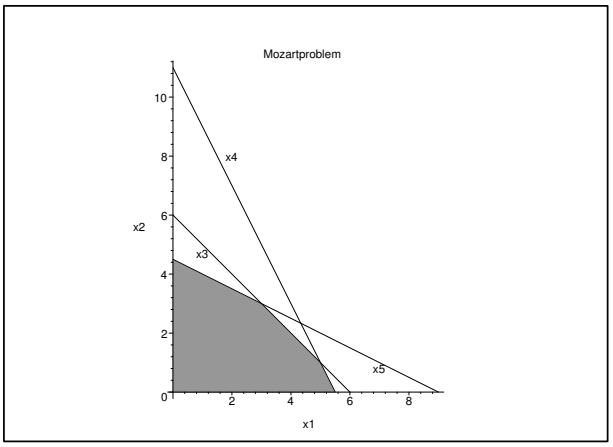
Algorithmus 1.2.4. *Input: A, b, c, eine zulässige Basis B und $\overline{x_B} = A_B^{-1} b$*

1. *BTRAN: Berechne $\overline{y} = A_B^{-\tau} c_B$ durch Lösen von $A_B^\tau \overline{y} = c_B$*
2. *Pricing: Berechne $\overline{z}_N = c_N - A_N^\tau \overline{y}$.*
   Falls $\overline{z}_N \geq 0$, ist $\overline{x}$ Optimallösung. Stop!
   Sonst, wähle $\hat{\jmath} \in N$ mit $\overline{z}_{\hat{\jmath}} < 0$. $x_{\hat{\jmath}}$ ist die eintretende Variable.
3. *FTRAN: Löse $A_B w = A_{.,\hat{\jmath}}$*
4. *Ratio-Test: Falls $w \leq 0$ ist das LP unbeschränkt. Stop!*
   Sonst, berechne $\gamma = \min \left\{ \frac{\overline{x}_{B(i)}}{w_i} : w_i > 0, i \in \{1, \ldots, m\} \right\} = \frac{\overline{x}_{B(i)}}{w_{\hat{\imath}}}$ für ein gewisses $\hat{\imath} \in \{1, \ldots, m\}$.
   $x_{B(\hat{\imath})}$ ist die austretende Variable.

5. *Update: Setze* $\overline{x}_B := \overline{x}_B - \gamma w$,
$x_{\hat{\imath}} := \gamma$,
$B(\hat{\imath}) := \hat{\jmath}$,
$N := N \setminus \{\hat{\jmath}\} \cup \{B(\hat{\imath})\}$.

Beispiel 1.2.1. *Mozartproblem:*

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 6 \\ 11 \\ 9 \end{pmatrix}, c = \begin{pmatrix} -9 \\ -8 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$



Anfang: Starten im Nullpunkt, zugehörige Basis $B = \{3, 4, 5\}$, $N = \{1, 2\}$

$$\overline{x} = (0, 0, 6, 11, 9)^\tau, \overline{x}_B = (6, 11, 9)^\tau$$

1. *Iteration:*

$$A_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, c_B = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \overline{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \overline{z}_N = \begin{bmatrix} -9 \\ -8 \end{bmatrix}.$$

Wählen $\hat{\jmath} = 1$, *dann ist* $w = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \Rightarrow \hat{\imath} = 2, B(\hat{\imath}) = 4, \gamma = 5.5$

$$B = \{3, 1, 5\}, N = \{4, 2\}, \overline{x} = \begin{bmatrix} 5.5 \\ 0 \\ 0.5 \\ 0 \\ 3.5 \end{bmatrix}, \overline{x}_B = \begin{bmatrix} 0.5 \\ 5.5 \\ 3.5 \end{bmatrix}.$$

2. *Iteration:*

$$A_B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}, c_B = \begin{bmatrix} 0 \\ -9 \\ 0 \end{bmatrix}, \overline{y} = \begin{bmatrix} 0 \\ -4.5 \\ 0 \end{bmatrix}, \overline{z}_N = \begin{bmatrix} 4.5 \\ -3.5 \end{bmatrix}.$$

Wählen $\hat{\jmath} = 2$*, dann ist* $w = \begin{bmatrix} 0.5 \\ 0.5 \\ 1.5 \end{bmatrix} \Rightarrow \hat{\imath} = 1, B(\hat{\imath}) = 3, \gamma = 1$

$$B = \{2, 1, 5\}, N = \{4, 3\}, \overline{x} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}, \overline{x}_B = \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}.$$

3. *Iteration:*

$$A_B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 2 & 1 & 1 \end{bmatrix}, c_B = \begin{bmatrix} -8 \\ -9 \\ 0 \end{bmatrix}, \overline{y} = \begin{bmatrix} -7 \\ -1 \\ 0 \end{bmatrix}, \overline{z}_N = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

$\Rightarrow$ *optimal.*

Satz 1.2.5. *Ist im primalen Algorithmus immer* $\gamma > 0$*, dann endet der Simplex-Algorithmus nach endlich vielen Schritten.*

Beweis. In jedem Schritt mit $\gamma > 0$ verbessert sich der Zielfunktionswert:

$$c^\tau x^+ = c_B^\tau A_B^{-1} b + (c_N - A_N^\tau A_B^{-\tau} c_B)^\tau x_N^+ = c_B^\tau A_B^{-1} b + \underbrace{\overline{z}_{\hat{\jmath}}}_{<0} \underbrace{\gamma}_{>0} < c_B^\tau A_B^{-1} b = c^\tau x$$

d.h. keine Basis kommt zweimal vor. Jede Basis entspricht einer Auswahl von m Indizes aus $\{1, \ldots, n\}$; es gibt höchstens $\begin{pmatrix} n \\ m \end{pmatrix}$ unterschiedliche Basen, d.h. der Algorithmus endet nach höchstens $\begin{pmatrix} n \\ m \end{pmatrix}$ Iterationen. □

Wann ist $0 = \gamma = \min\left\{\frac{\overline{x}_{B(i)}}{w_i}, w_i, i \in \{1, \ldots, m\}\right\}$?
$\gamma = 0$ bedeutet, $\exists i \in B : \overline{x}_{B(i)} = 0$, d.h. eine weitere Ungleichung ist bzgl. dieser Basislösung aktiv.

Definition 1.2.2.

- *Eine Basis* B *heißt entartet/degeneriert (degenerate), falls* $x_{B(i)} = 0$ *für ein* $i \in \{1, \ldots, m\}$.
- *Ein lineares Optimierungsproblem heißt entartet/degeneriert, falls es entartete Basen besitzt.*

Entartetes LP: Durch einen von n Gleichungen festgelegten Punkt gehen mehr als n Gleichungen. Für zufällige Daten tritt dies mit Wahrscheinlichkeit 0 ein; in der Praxis aber recht oft.

Im Simplex-Algorithmus legen die Nichtbasisvariablen N den Punkt fest. Im Pricing-Schritt wird eine Ungleichung aus N herausgenommen. Damit wird ein eindimensionaler Halbstrahl festgelegt, entlang dem sich die Lösung weiterbewegen kann. Trifft der Halbstrahl bereits für Schrittweite 0 die nächste Ungleichung, beschreiben die neuen Nichtbasisvariablen denselben Punkt. So etwas kann sich mehrmals wiederholen, oder sogar auf die selbe Basis zurückführen. In diesem Fall sagt man, der Simplexalgorithmus *kreist*; er terminiert dann nicht.

Für die meisten üblichen Auswahlverfahren im Pricing bzw. Ratio Test kann man Beispiele mit Kreisen konstruieren.

Eine einfache Auswahlregel, die das Kreisen verhindert, ist die **Regel von Bland**:

- Im Pricing wähle unter den Variablen mit negativen reduzierten Kosten diejenige mit dem kleinsten Index.
- Im Ratio-Test wähle unter den 0-Variablen diejenige mit kleinsten Index.

Satz 1.2.6. *Werden im Simplexalgorithmus die Auswahlverfahren von Bland verwendet, terminiert der Simplexalgorithmus immer.*

Beweis. Nehmen an, der Algorithmus kreist trotzdem. Seien $B_1, \dots B_k$ die Basen, die der Algorithmus durchkreist: $B_1, \dots, B_k, B_1, \dots$ und

$$\mathcal{I} = \{i \in \{1, \dots, n\} : \exists k_1, k_2 \in \{1, \dots, k\} : i \in B_{k_1}, i \notin B_{k_2}\}$$

die Indexmenge der wechselnden Variablen. Sei $t = \max \mathcal{I}$, o.B.d.A. sei $t \in B_1, t \notin B_2$. Sei $s \in B_2 \setminus B_1$, also $s \in N_1, s \notin N_2$ die neue Basisvariable, die t ersetzt. Sei $\overline{B} = B_h \in \{B_2, \dots, B_k\}$ die erste Basis mit $t \notin N_h, t \in B_{h+1}$. Für $\overline{B}$ müssen die reduzierten Kosten von t negativ sein:

$$\overline{c}_t := c_t - (A_{\cdot,t})^\tau A_{\overline{B}}^{-\tau} c_{\overline{B}} < 0 \quad \text{und} \quad \overline{c}_i \geq 0 \, \forall i \in N \cap I \Rightarrow \overline{c}_s \geq 0.$$

Definieren $\overline{c}_i = 0 \, \forall i \in \overline{B}$. Der Zielfunktionswert ist $(*) \quad \beta + \sum_{j=1}^{n+m} \overline{c}_j x_j$ mit $\beta = c_{\overline{B}}^\tau A_{\overline{B}}^{-\tau} b$.

Für die Basis $\hat{B} = B_1$ ist $s \in \hat{N}$ und $\hat{c}_s := c_s - (A_{\cdot,s})^\tau A_{\hat{B}}^{-\tau} c_{\hat{B}} < 0$. Die Zielfunktion in Abhängigkeit von x_s ist $\beta + \hat{c}_s x_s$.

Aus $(\star)$ mit $x_{\hat{B}} = \underbrace{A_{\hat{B}}^{-1} b}_{\hat{b}} - \underbrace{A_{\hat{B}}^{-1} A_{\cdot,s}}_{\hat{a}} x_s$ erhalten wir

$$\beta + \hat{c}_s x_s = \beta + \textstyle\sum_{j \in \hat{B}} \overline{c}_j (\hat{b}_j - \hat{a}_j x_s) + \overline{c}_{j_s} x_s$$

$$(\hat{c}_s - \overline{c}_{j_s} + \textstyle\sum_{j \in \hat{B}} \overline{c}_j \hat{a}_j) x_s = \textstyle\sum_{j \in \hat{B}} \overline{c}_j \hat{b}_j = \text{const} \quad \forall x_s$$

$$\Rightarrow \underbrace{\hat{c}_s}_{<0} - \underbrace{\overline{c}_{j_s}}_{\geq 0} + \underbrace{\sum_{j \in \hat{B}} \overline{c}_j \hat{a}_j}_{\Rightarrow > 0} = 0$$

$\Rightarrow \exists j \in \hat{B}$ mit $\overline{c}_j \hat{a}_j > 0$, insbesondere $\overline{c}_j \neq 0$, also $j \notin \overline{B}, j \in \hat{B}$.
Also wechselt Variable j in und aus der Basis, $j \neq t$ weil $j \notin \overline{B} \Rightarrow j < t$ damit $\overline{c}_j \geq 0$ (bei $\overline{c}_j < 0$ wäre j statt t im Pricing ausgewählt worden); damit ist auch $\hat{a}_j > 0$.
Da j wechselt, ist $x_j = 0$ und wegen $j \in \hat{B}$ und $j < t$ hätte j statt t im Schritt von $\hat{B} = B_1$ zu B_2 im Ratio-Test gewählt werden müssen. $\Rightarrow$ Widerspruch. □

Korollar 1.2.7. *Ausgehend von einer zulässigen Basis bestimmt der Simplexalgorithmus mit der Auswahlregel von Bland in endlich vielen Schritten eine Optimallösung oder weist die Unbeschränktheit nach.*

Beweis. Folgt aus Lemma 1.2.1, Lemma 1.2.3 und aus Satz 1.2.6. □

Es gibt noch andere Varianten, das Kreisen zu vermeiden:

- zufällige Wahl der austretenden Variable
- zufällige Pertubation der rechten Seite
- symbolisches Pertubieren
  ($0 < \varepsilon_1 \ll \varepsilon_2 \ll \ldots \ll \varepsilon_m$ Pertubationen, die sich nicht auslöschen können)

Diese werden immer erst dann eingesetzt, wenn Anzeichen für das Kreisen beobachtet werden.

Sonst gibt es bessere Regeln:

ursprünglich: negativste reduzierte Kosten.
Nachteil: hängt stark von der Skalierung der einzelnen Variablen ab, z.B.:

$$\tilde{c}_1 x_1 + \tilde{c}_2 x_2 = -0.3 x_1 - 1.5 x_2 \Rightarrow x_2$$

ersetzen $x_1 = 10\overline{x}_1$

$$\Rightarrow -3\overline{x}_1 - 1.5 x_2 \Rightarrow x_1$$

misst nur den Fortschritt pro Einheitsschritt entlang $x_{\hat{j}}$

heute allgemein: steilste Kante ("*steepest edge*")
Idee: Wähle $\hat{j} \in N$, so dass die Richtung $\triangle x$ den Ausdruck $\frac{c^\tau \triangle x}{\|\triangle x\|}$ minimiert $\Rightarrow$ misst den Fortschritt pro Einheitsschritt im Gesamtraum.

Wie geschieht dies effizient?
Bei Wahl von j aus N ist

$$\begin{aligned}
\Delta x &= \begin{bmatrix} \Delta x_B \\ \Delta x_N \end{bmatrix} = \begin{bmatrix} -A_B^{-1}A_{.,j} \\ 0 \end{bmatrix} + e_j \\
x_B &= -A_B^{-1}A_N x_N + A_B^{-1}b \\
c^\tau \Delta x &= c_B^\tau \Delta x_B + c_N^\tau \Delta x_N = c_j - (A_N^\tau A_B^{-\tau} c_B)_j = -y_j \\
\|\Delta x\|^2 &= \|\Delta x_B\|^2 + 1 = \|A_B^{-1}A_N e_j\|^2 + 1
\end{aligned}$$

In jedem Schritt $A_B^{-1}A_N$ zu berechnen ist zu aufwendig. Man kann aber die Normen $\nu_k = \|A_B^{-1}A_N e_k\|^2, k \in N$ mitführen und effizient aktualisieren.
Betrachten Aktualisierung von ν_k bei Basiswechsel $\tilde{B} = B \setminus \{B(\hat{\imath})\} \cup \{\hat{\jmath}\}$:

$$\begin{aligned}
A_{\tilde{B}} &= A_B + (A_{.,\hat{\jmath}} - A_{B(\hat{\imath})})e_{\hat{\imath}}^\tau \\
&= A_B(I + A_B^{-1}(A_{.,\hat{\jmath}} - A_{.,B(\hat{\imath})})e_{\hat{\imath}}^\tau) \\
&= A_B(I + (\underbrace{A_B^{-1}, A_{.,\hat{\jmath}}}_{w} - \underbrace{A_B^{-1}A_{.,B(\hat{\imath})}}_{e_{\hat{\imath}}})e_{\hat{\imath}}^\tau) \\
&= A_B(\underbrace{I + (w - e_{\hat{\imath}})e_{\hat{\imath}}^\tau}_{W}) \\
\tilde{\nu}_k &= \|A_{\tilde{B}}^{-1}A_{.,k}\|^2 = A_{.,k}^\tau A_{\tilde{B}}^{-\tau} A_{\tilde{B}}^{-1} A_{.,k} = A_{.,k}^\tau A_B^{-\tau} W^{-\tau} W^{-1} A_B^{-1} A_{.,k}
\end{aligned}$$

Einschub: Wie bekommt man $(A + uv^\tau)^{-1}$ aus A^{-1} (low rank update)?

$(A + uv^\tau)^{-1} = A^{-1} - \frac{1}{1+v^\tau A^{-1}u} A^{-1}uv^\tau A^{-1}$
für$A = I : (I + uv^\tau)^{-1} = I - \frac{1}{1+v^\tau u} uv^\tau$
$\Rightarrow$ für$W : W^{-1} = I - \frac{1}{w_{\hat{\imath}}}(w - e_{\hat{\imath}})e_{\hat{\imath}}^\tau$.

Also erhält man

$$\begin{aligned}
\tilde{\nu}_k &= \underbrace{A_{.,k}^\tau}_{a_k} A_B^{-\tau} \left(I - \tfrac{e_{\hat{\imath}}(w-e_{\hat{\imath}})^\tau}{w_{\hat{\imath}}}\right)\left(I - \tfrac{(w-e_{\hat{\imath}})e_{\hat{\imath}}^\tau}{w_{\hat{\imath}}}\right) A_B^{-1} \underbrace{A_{.,k}}_{a_k} \\
&= a_k^\tau A_B^{-\tau} A_B^{-1} a_k - 2a_k^\tau \tfrac{A_B^{-\tau} e_{\hat{\imath}}(w-e_{\hat{\imath}})^\tau A_B^{-1}}{w_{\hat{\imath}}} a_k + a_k^\tau \tfrac{A_B^{-\tau} e_{\hat{\imath}}(w-e_{\hat{\imath}})^\tau (w-e_{\hat{\imath}})e_{\hat{\imath}}^\tau A_B^{-1}}{w_{\hat{\imath}}^2} a_k \\
&= \nu_k - 2\tfrac{a_k^\tau u(v-u)^\tau a_k}{w_{\hat{\imath}}} + (a_k^\tau u)\tfrac{\|w-e_i\|^2}{w_i^2}
\end{aligned}$$

benötigen dazu also nur: $u = A_B^{-\tau} e_{\hat{\imath}}$, also lösen von $A_B^\tau u = e_{\hat{\imath}}$
$v = A_B^{-\tau} w$, also lösen von $A_B^\tau v = w$
$\Rightarrow$ also recht günstig zu berechnen.

Laufzeit des Simplexalgorithmus:

- Im schlimmsten Fall exponentiell viele Iterationen
  Klee und Minty 1972: 2^n (n Variablen, n Nebenbedingungen)
  "Klee und Minty Cubes" für negativste reduzierte Kosten

- empirisch für praktische Probleme: $1.5m$ Iterationen; hängt jedoch sehr von der Degeneriertheit des Linearen Programmes ab.

1.3 Zulässigkeit und Fundamentalsatz der Linearen Optimierung

Wie findet man eine zulässige Basis? Betrachten also $\boxed{\begin{array}{ll}\min & c^\tau x \\ \text{s.t.} & Ax = b \\ & x \geq 0,\end{array}}$ wobei keine zulässige Basis bekannt sei.

1.3.1 Zwei-Phasen-Methode

Lösen in Phase I ein künstliches Problem. O.B.d.A. sei $b \geq 0$. Die Aufgabe

$$\boxed{\begin{array}{ll}\min & e^\tau s = \sum s_i \\ \text{s.t.} & Ax + s = b \\ & x \geq 0, s \geq 0\end{array}}$$

hat eine zulässige Basis: $\{s_1, \ldots, s_m\}$

Findet der Simplexalgorithmus eine Optimallösung mit $s_i = 0 \ \forall i$, dann ist die entsprechende Basis zulässig für das Originalproblem. Löse davon ausgehend in Phase II das Originalproblem.

Bemerkung 1.3.1.

- *Sobald ein s_i nach N wechselt, kann die Hilfsvariable entfernt werden.*
- *Sind in der Optimallösung der Phase I noch s_i in der Basis (degeneriert), so kann man diese immer mit einer geeigneten Spalte herauspivotisieren.*

Terminiert Phase I mit Optimalwert > 0, dann gibt es nach Korollar 1.2.7 keine zulässige Lösung.

1.3.2 Groß-M/(Big-M)-Methode

Man versucht, beide Phasen in einer zu behandeln, indem man $M \gg 0$ wählt und

$$\boxed{\begin{array}{ll}\min & c^\tau x + M e^\tau s \\ \text{s.t.} & Ax + s = b \\ & x \geq 0, s \geq 0\end{array}}$$

löst.

Nachteile: Es ist unklar wie groß M gewählt werden muss.
Verursacht oft numerische Schwierigkeiten.

Vorteile: Wenn M klein gewählt werden kann: Simplexalgorithmus geht sofort auf die Suche nach einer einer "guten" Basis.
Künstliche Variablen können wie zuvor entfernt werden, sobald sie in N wechseln.

Satz 1.3.2 (Fundamentalsatz der linearen Optimierung). *Hat ein Lineares Programm eine Optimallösung, wird diese auch in einer zulässigen Basis angenommen. Ein lineares Programm ohne Optimallösung ist entweder unbeschränkt oder unzulässig.*

Beweis. Zwei-Phasen-Methode und Korollar 1.2.7. □

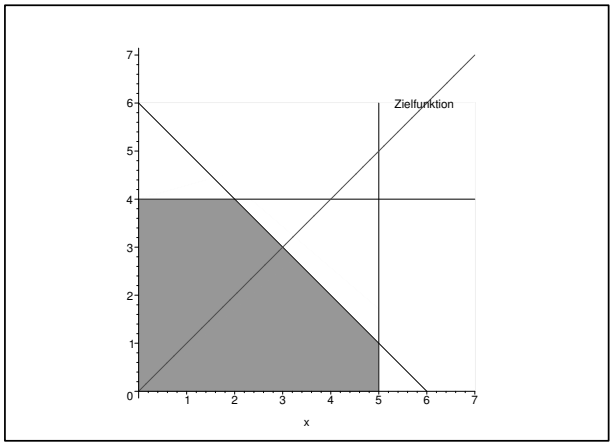
Wichtig: Das Resultat verwendet wesentlich $x \geq 0$. Lässt man auch freie Variable zu, muss es keine Basislösung mehr geben.
Beispiel: $\min ax + by$
s.t.$ax + by = c$.

1.4 Dualität

Zu jedem linearen Programm kann man ein *"duales Programm"* aufschreiben, das immer gleichzeitig mitgelöst wird und wichtige Zusatzinformationen zum ursprünglichen liefert.

Geometrisch lässt es sich als das *"Bestimmen einer gültigen Ungleichung, die den Zielfunktionswert am stärksten einschränkt"*, erklären.

Beispiel 1.4.1.



$$\begin{aligned} &\max c^\tau x \\ &\text{s.t. } Ax \leq b \\ &\quad x \geq 0 \end{aligned}$$

$$\mathcal{X} = \{x \geq 0 : Ax \leq b\}$$

Aus $Ax \leq b$ können neue gültige Ungleichungen gewonnen werden, indem man die Ungleichungen mit nichtnegativen Zahlen multipliziert und addiert ("Nichtnegative Linearkombinationen").

Sei $y \in \mathbb{R}^m : y \geq 0$, $a' = y^\tau A, b' = y^\tau b$. Dann ist $a'^\tau x \leq b'$ gültig $\forall x \in \mathcal{X}$. Gilt zusätzlich $a' \geq c$, dann ist wegen $x \geq 0$ auch $c^\tau x \leq a'^\tau x \leq b' \quad \forall x \in \mathcal{X}$.

Jedes $y \geq 0$ mit $A^\tau y \geq c$ liefert $y^\tau b$ als obere Schranke für den Optimalwert. Das duale Programm sucht die kleinste Schranke über alle y.

$$\boxed{\begin{array}{l}\max c^\tau x \\ \text{s.t. } Ax \leq b \\ \quad\quad x \geq 0\end{array}} \quad \Longrightarrow \quad \boxed{\begin{array}{l}\min b^\tau y \\ \text{s.t. } A^\tau y \geq c \\ \quad\quad y \geq 0\end{array}}$$

Primales Programm in kanonischer Form — Duales Programm in kanonischer Form

Woher kommt der Name *"dual"*? In der linearen Algebra wird für einen Vektorraum V der Vektorraum der linearen Funktionale $f : V \to \mathbb{R}$ als Dualraum $V^\star$ bezeichnet. Für $V = \mathbb{R}^n$ ist $f(x) = \alpha_1 x_1 \cdots \alpha_n x_n$, also $(\mathbb{R}^n)^\star = \mathbb{R}$.. Die Nebenbedingungen enthalten lineare Funktionale. Mit $y \geq 0$ bilden wir eine nichtnegative Linearkombination dieser Vektoren aus dem Dualraum und erhalten ein neues Element aus dem dualen Raum. Im dualen Problem optimieren wir über den dualen Vektorraum. Wegen $(\mathbb{R}^n)^\star = \mathbb{R}^n$ sind beide Probleme "ganz gleich". Insbesondere ist das Duale des dualen Problems wieder das Primale.

Analog leitet man das Duale für ein Primales in Standardform ab:

$$\boxed{\begin{array}{l}\min c^\tau x \\ \text{s.t. } Ax = b \\ \quad\quad x \geq 0\end{array}} \quad \Longleftrightarrow \quad \boxed{\begin{array}{l}\max b^\tau y \\ \text{s.t. } A^\tau y \leq c \\ \quad\quad y \text{ frei}\end{array}} \quad \Longleftrightarrow \quad \boxed{\begin{array}{l}\max b^\tau y \\ \text{s.t. } A^\tau y + z = c \\ \quad\quad y \text{ frei}, z \geq 0\end{array}}$$

Primales in Standardform — Duales in Standardform

Übung: Was macht man, wenn $\leq, \geq, =$-Nebenbedingungen gleichzeitig auftreten?

1.4.1 Schwache Dualität

Betrachten primal-duales Paar in Standardform:

$$\mathcal{X} = \{x \geq 0 : Ax = b\}, \mathcal{Z} = \{(x, y) : A^\tau y + z = c, z \geq 0\}.$$

Offensichtlich gilt nach Konstruktion sogenannte schwache Dualität (weak duality):

$$\boxed{\inf_{x \in \mathcal{X}} c^\tau x \geq \sup_{(x,y) \in \mathcal{Z}} b^\tau y}$$

alternativer Beweis: Sei $x \in \mathcal{X}, (y, z) \in \mathcal{Z}$

$$c^\tau x = (A^\tau y + z)^\tau x = y^\tau A x + z^\tau x = b^\tau y + \underbrace{z^\tau x}_{\geq 0} \geq b^\tau y.$$

Entscheidende Frage: Kann es passieren, dass

$$p^\star := \inf c^\tau x > \sup_{(y,z)\in\mathcal{Z}} b^\tau y =: d^\star,$$

also eine Dualitätslücke auftritt, oder gilt immer $p^\star = d^\star$ (sogenannte starke Dualität)?

Nicht, wenn $p^\star$ endlich ist, denn dann berechnet der primale Simplexalgorithmus gleich die duale Optimallösung mit.

1. BTRAN: Berechne $\overline{y} = A_B^{-\tau} c_B$ durch lösen von $A_B^\tau \overline{y} = c_B$
2. Pricing: Berechne $\overline{z}_N = c_N - A_N^\tau \overline{y}$
   - Falls $\overline{z}_N \geq 0$, ist $\overline{x}$ Optimallösung. Stop!
   - ...

Ist $\overline{x}$ Optimallösung, dann ist $\begin{pmatrix} A_B^\tau \\ A_N^\tau \end{pmatrix} \overline{y} + \underbrace{\begin{pmatrix} 0 \\ \overline{z}_N \end{pmatrix}}_{=:\overline{z}\geq 0} = \begin{pmatrix} c_B \\ c_N \end{pmatrix}$, also $(\overline{y}, \overline{z})$ dual zulässig und der Wert der dualen Lösung ist

$$b^\tau \overline{y} = b^\tau A_B^{-\tau} c_B = \overline{x}_B^\tau c_B = c^\tau \overline{x}$$

gleich dem primalen Optimalwert. Aufgrund der schwachen Dualität ist $(\overline{y}, \overline{z})$ also duale Optimallösung mit dem gleichen Wert. Das zeigt:

Satz 1.4.1 (starke Dualität). *Ein primales Programm hat eine endliche Optimallösung genau dann, wenn auch das dazugehörige duale eine endliche Optimallösung hat. Insbesondere: Ist eines der beiden zulässig, dann gilt*

$$\inf_{x\in\mathcal{X}} c^\tau x = \sup_{(y,z)\in\mathcal{Z}} b^\tau y.$$

Beispiel 1.4.2. *Beide unzulässig ist möglich:*

$$\boxed{\begin{array}{ll} \max & x_1 \\ \text{s.t.} & x_1 - x_2 \leq -1 \\ & -x_1 + x_2 \leq 0 \\ & x_1 \geq 0, x_2 \geq 0 \end{array}} \iff \boxed{\begin{array}{ll} \min & -y_1 \\ \text{s.t.} & y_1 - y_2 \geq 1 \\ & -y_1 + y_2 \geq 0 \\ & y_1 \geq 0, y_2 \geq 0 \end{array}}$$

Korollar 1.4.2 (Satz vom Komplementären Schlupf (complementary slackness theorem)). *Seien $\overline{x}$ und $(\overline{y}, \overline{z})$ primal und dual zulässig (bezüglich der Standardform). $\overline{x}$ und $(\overline{y}, \overline{z})$ sind primale und duale Optimallösung genau dann, wenn $\overline{x}^\tau \overline{z} = 0$.*

Beweis. "⇒": Satz 1.4.1 und schwache Dualität
"⇐": schwache Dualität □

Bemerkung 1.4.3. $\overline{x}^\tau \overline{z} = 0$ *gilt offensichtlich für beliebige primale und duale Optimallösungen. Wegen* $\overline{x} \geq 0$ *und* $\overline{z} \geq 0$ *folgt aus* $\overline{x}^\tau \overline{z} = 0$*, dass* $\overline{x_i}\ \overline{z_i} = 0$*. Wenn also* $\overline{x_i}$ *der primalen Schlupfvariablen der Ungleichung* j *entspricht, heißt dass*

$$\overline{x_i} \cdot \overline{z_i} = \overline{x_i}(\underbrace{c_i}_{=0} - \underbrace{A^\tau_{\cdot,i}}_{=e_j}\overline{y}) = \overline{x_i}(\underbrace{0 - \overline{y_j}}_{-\overline{y_j} \leq 0}) = 0.$$

Falls der primale Schlupf der j-ten Ungleichung in einer Optimallösung $\overline{x}$ größer als 0 ist (die Ungleichung ist nicht bindend), dann muss in jeder dualen Optimallösung $y_j = 0$ sein.
Umgekehrt: Ist y_j Dualvariable einer primalen Ungleichung und ist $y_j \neq 0$ in einer dualen Optimallösung, dann muss diese primale Ungleichung in allen primalen Optimallösungen mit Gleichheit erfüllt sein.

Primale Optimallösung degeneriert $\Longleftrightarrow$ Duale Optimallösung nicht eindeutig.
Duale Optimallösung ist degeneriert $\Longleftrightarrow$ primale Optimallösung ist nicht eindeutig.

Um dualen Simplex zu entwickeln: $\begin{pmatrix} A^\tau_B \\ A^\tau_N \end{pmatrix} y + \begin{pmatrix} z_B \\ z_N \end{pmatrix} = \begin{pmatrix} c_B \\ c_N \end{pmatrix}$.
Setzen also z_B auf 0 und berechnen die anderen (beachte: komplementärer Schlupf):

$$\begin{aligned} y &= A_B^{-\tau} c_B - A_B^{-\tau} z_B \\ z_N &= c_N - A_N^\tau y = c_N - A_N^\tau A_B^{-\tau} c_B + A_N^\tau A_B^{-\tau} z_B \geq 0 \text{ für duale Zulässigkeit} \end{aligned}$$

Zielfunktion in Abhängigkeit von z_B:

$$b^\tau y = bA_B^{-\tau} c_B - bA_B^{-\tau} z_B$$

$z_{B(\hat{\imath})}$ vergrößern hilft dann, wenn $0 > [A_B^{-\tau} b]_{B(\hat{\imath})} = x_{B(\hat{\imath})}$ (also wenn die Nebenbedingung $B(\hat{\imath})$ primal unzulässig ist).
$z_{B(\hat{\imath})}$ vergrößern, solange z_N zulässig bleibt: Finde maximales γ mit

$$z_N = c_N - A_N^\tau A_B^{-\tau} c_B + A_N^\tau A_B^{-\tau} e_i \gamma \geq 0$$

Algorithmus 1.4.4 (dualer Simplexalgorithmus). *Input:* A, b, c*, eine zulässige Basis* B *und* $\overline{z}_N = c_N - A_N^\tau A_B^{-\tau} c_B \geq 0$

1. *BTRAN: Löse* $A_B \overline{x}_B = b$
2. *Pricing: Falls* $\overline{x}_B \geq 0$*, ist B optimal. Stop!*
   Sonst wähle $\hat{\imath} \in \{1, \ldots, m\}$ *mit* $x_{B(\hat{\imath})} < 0$*.* $z_{B(\hat{\imath})}$ *ist die austretende Variable.*
3. *FTRAN: Löse* $A_B^\tau w = e_{\hat{\imath}}$ *und berechne dann* $\alpha_N = -A_N^\tau w$*.*
4. *Ratio-Test: Falls* $\alpha_N \leq 0$*, ist das LP unzulässig. Stop!*
   Sonst setze $\gamma = \min\left\{\frac{\overline{z}_k}{\alpha_k} : \alpha_k > 0, k \in N\right\} = \{\frac{\overline{z}_{\hat{\jmath}}}{\alpha_{\hat{\jmath}}}\}$ *mit* $\hat{\jmath} \in N, \alpha_{\hat{\jmath}}$*.*
   $z_{\hat{\jmath}}$ *ist die eintretende Variable.*

5. *Update: Setze* $\overline{z}_N := \overline{z}_N - \gamma\alpha_N$
   $z_{B(\hat{\imath})} := \gamma$
   $N := N \setminus \{\hat{\jmath}\} \cup \{B(\hat{\imath})\}$
   $B(\hat{\imath}) := \hat{\jmath}$.

1.5 Sensitivität

Typische Frage an die Optimallösung: Wie stabil ist die Optimallösung gegenüber Veränderungen der Kosten oder der rechten Seite?
Etwa weil: man die Preise anpassen will,
man die Daten nicht genau kennt,
man wissen will, welche Nebenbedingungen besonders wichtig sind.

Seien primale und duale Optimallösungen

$$\begin{aligned} x_B^\star &= A_B^{-1}b \\ y^\star &= A_B^{-\tau}c_B \\ z_N^\star &= c_N - A_N^\tau y^\star \end{aligned}$$

sowie die primalen bzw. dualen Optimalwerte

$$\begin{aligned} p:&\ c_B^\tau x_B^\star + c_N^\tau x_N^\star \\ d:&\ b^\tau y^\star. \end{aligned}$$

gegeben.

Änderungen Δc in c: erhalten die primale Zulässigkeit,
aber führen zu Änderungen im Dualen.

Derzeitige Basis ist solange optimal, solange $z_N(t) \geq 0$ sind für $c(t) = c + t\Delta c$ mit $t \in \mathbb{R}$.

$$\begin{aligned} z_N(t) &= c_N + t\Delta c_N - A_N^\tau A_B^{-\tau}(c_B + t\Delta c_B) \\ &= z_N^\star + t(\underbrace{\Delta c_N - A_N^\tau A_B^{-\tau}\Delta c_B}_{=:\Delta z_N}) \end{aligned}$$

Also gilt $z_N(t) \geq 0$ für

$$\max_{[\Delta z_N]_i > 0} \left\{ -\tfrac{[z_N^\star]_i}{[\Delta z_N]_i} \right\} \leq t \leq \min_{[\Delta z_N]_i < 0} \left\{ -\tfrac{[z_N^\star]_i}{[\Delta z_N]_i} \right\}.$$

Für diese t ist der neue Optimalwert einfach $(c + t\Delta c)^\tau x^\star$.
Kein Spielraum, wenn $[z_N^\star]_i = 0$ und $[\Delta z_N]_i \neq 0$ (mit richtigem Vorzeichen) (duale Optimallösung degeneriert).

Änderungen Δb in b: erhalten die duale Zulässigkeit,
führen aber zu Änderungen im Primalen.

Derzeitige Basis ist solange optimal, solange $x_B(t) \geq 0$ bleibt für $b(t) = b + t\Delta b$

$$x_B(t) = x_B^\star + t \underbrace{A_B^{-1}\Delta b}_{\Delta x_B}$$

Also gilt $x_B(t) \geq 0$ für

$$\max_{[\Delta x_B]_i > 0} \left\{ -\frac{[x_B^\star]_i}{[\Delta x_B]_i} \right\} \leq t \leq \min_{[\Delta x_B]_i < 0} \left\{ -\frac{[x_B^\star]_i}{[\Delta x_B]_i} \right\}.$$

Für diese t ist der Optimalwert $= (b + t\Delta b)^\tau y^\star$. Interessiert uns besonders $\Delta b = e_j$ (also wenn wir für eine Ungleichung j die rechte Seite b_j ändern) $\Rightarrow$ $y_j^\star$ gibt uns die marginale Änderung der Zielfunktion bei Einheitsänderung $\Delta b = e_j$ an.

Folglich lassen sich y als *Schattenpreise* bei Ressource-Nebenbedingungen interpretieren. Wenn man b_j durch Zukauf von Ressourcen vergrößern will, dann darf das pro Einheit höchstens y_j kosten, sonst rentiert sich das sicher nicht. $\Rightarrow$ Anbieter der Ressource wird im Gleichgewichtsfall gerade y_j verlangen (sonst könnte man Gewinn steigern). Ist insbesondere ein $y_j = 0$, kann die entsprechende Ungleichung weggelassen werden, ohne die Optimallösung zu verändern, denn: Primales bleibt zulässig, Zielfunktionswert ändert sich nicht, Duales ändert sich nicht.

Definition 1.5.1. *Nebenbedingungen mit $y_j \neq 0$ nennt man aktiv.*

Ebenso kann man natürlich primale Variablen mit $x_i^\star = 0$ weglassen, ohne die Optimallösung zu verändern. Bei sehr großen linearen Programmen versucht man inaktive Ungleichungen und (unnötige) Variablen erst gar nicht aufzunehmen.

1.6 Spaltengenerierung und Schnittebenenverfahren

1.6.1 Spaltengenerierung (column generation)

$Ax = b$ lässt sich nicht speichern, weil n zu groß ist, man weiß aber, wie zu jedem x_j die Spalte $A_{\cdot,j}$ gebildet werden muss.

Solange $x_j = 0$ ist, braucht man die Spalte $A_{\cdot,j}$ nicht. x_j wird $\neq 0$, wenn es im Pricing-Schritt gewählt wird.

Pricing: Berechne $\overline{z}_N = c_N - A_N^\tau \overline{y}$ mit $\overline{y} = A_B^{-\tau} c_B$ und wähle $\hat{\jmath}$ mit $\overline{z}_{\hat{\jmath}} < 0$. Es reicht, wenn $\min_{j \in N}(c_j - A_{\cdot,j}^\tau \overline{y})$ bestimmt werden kann, dazu muss A_N nicht explizit verfügbar sein.

Beispiel 1.6.1. *Schneide Bleche mit Breite b_i mit Gesamtlänge $l_i, i = 1, \ldots, m$, aus Blechrollen der Breite $\overline{b}$, so dass möglichst wenig Verschnitt entsteht.*

Schnittmuster: beliebige Kombination der Breiten b_i, so dass die Gesamtbreite $\leq \overline{b}$. Zulässige Menge also

$$\mathcal{S} = \{s \in \mathbb{N}_0^m : \sum_{i=1}^{m} s_i b_i \leq \overline{b}\}$$

s_i: "wie oft kommt b_i im Schnittmuster s vor"
Variable x_s für $s \in \mathcal{S}$: Länge, mit der Schnittmuster s eingesetzt wird.
Problem: $\min \sum_{s\in\mathcal{S}} x_s$ *möglichst wenige Rollen verwenden*

$$\text{s.t.} \sum_{s\in\mathcal{S}} s x_s \geq l \quad \textit{Bedarf erfüllen}$$
$$x_s \geq 0 \; \forall s \in \mathcal{S}.$$

Für $s \in \mathcal{S}$ ist $A_{.,s} = s$. Das Pricing-Problem ist damit

$$\min_{s\in\mathcal{S}}(1 - s^\tau \overline{y}) \quad \text{oder} \quad \begin{cases} \max \overline{y}^\tau s \\ \text{s.t.} \sum_{i=1}^{m} s_i b_i \leq \overline{b} \\ \quad s_i \in \mathbb{N}_0, i = 1, \ldots, m \end{cases}$$

Dieses sogenannte Rucksackproblem (knapsack problem) löst man für $b_i \in \mathbb{N}$ und $\overline{b} \in \mathbb{N}$ (jeweils nicht zu groß, ≤ 1000) mit dynamischer Programmierung:

- bauen die Lösung sukzessive für Gesamtbreite $b = 1, \ldots, \overline{b}$ und mit Breiten $b_1, \ldots, b_k$ mit $k < m$ auf
- nutzen die Rekursion (für $0 < b \leq \overline{b}, 0 < k \leq m$)

$$\text{opt}(b,k) = \max\{\text{opt}(b, k-1), \text{opt}(b - b_k, k) + \underbrace{\overline{y}_k}_{o.B.d.A.\geq 0}\}$$

- definieren $\text{opt}(z_1, z_2) = 0$ für $z_1 = 0$ oder ($z_1 > 0$ und $z_2 = 0$) und $\text{opt}(z_1, z_2) = -\infty$ für $z_1 < 0$.

Beispiel 1.6.2. $b_1 = 3, b_2 = 5, b_3 = 6, \overline{b} = 10, \overline{y_1} = 2, \overline{y_2} = 4, \overline{y_3} = 7$

	$\overbrace{\quad}^{opt(.,0)}$		$\overbrace{\quad}^{opt(.,1)}$		$\overbrace{\quad}^{opt(.,2)}$		$\overbrace{\quad}^{opt(.,3)}$
10	0	$\Longrightarrow$	6	$\Longrightarrow$	8	$\Longrightarrow$	9
9	0		6		6		9
8	0		4		6		7
7	0		4		4		7
6	0		4		4		7
5	0		2		4		4
4	0		2		2		2
3	0		2		2		2
2	0		0		0		0
1	0		0		0		0
0	0		0		0		0

Zu opt$(10, 3) = 9$ *gehört die Spalte* $s = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, *die als neue Spalte* $A_{\cdot,j}$ *aufgenommen wird.*

In der Praxis:

1. Wähle einige Anfangsmuster, die Zulässigkeit garantieren.
2. Löse für diese optimal.
3. Entferne ungebrauchte Schnittmuster (Spalten).
4. Bestimme neue Schnittmuster durch Spaltengenerierung. Falls keine negativen reduzierten Kosten $\Rightarrow$ optimal über alle Schnittmuster. Stop!
5. Sonst zurück zu Schritt 2.

Nachteil in konkreter Anwendung:

- Länge x_s des Schnittmusters ist leider normalerweise kein ganzzahliges Vielfaches der Längen der Ausgangsrollen.
- Anzahl der ausgewählten Schnittmuster m (Anzahl der Breiten) im allgemeinen noch zu viel für praktische Anwendung (hohe Rüstkosten).

1.6.2 Schnittebenenverfahren

- ist gerade die Spaltengenerierung im Dualen
- besser vorstellbar: im Primalen
    1. Berechne Optimallösung für Auswahl der Ungleichungen.
    2. F üge eine Ungleichung dazu, die von der derzeitigen Optimallösung verletzt wird. Gibt es keine $\rightarrow$ Optimallösung des Gesamtsystems, sonst 1.

Schwierigkeiten hier: Effizientes Finden von Ungleichungen, die die aktuelle Optimallösung verletzen $\rightarrow$ "*Separierungsproblem*".

Beispiel: *Traveling Salesman Problem*(TSP)

Welchen Algorithmus verwendet man wann?

- Spaltengenerierung: Problem bleibt primal zulässig $\Rightarrow$ primaler Simplex
- Schnittebenenverfahren: Problem bleibt dual zulässig $\Rightarrow$ dualer Simplex

Simplex hat sehr gute "Warmstart"-Eigenschaften.
Warmstart: Nach Problemmodifikation von aktueller Lösung aus fortsetzen.

1.7 Ganzzahligkeit von Basislösungen

In vielen Anwendungen sucht man ganzahlige Optimallösungen, meistens ist dies $\mathcal{NP}$-schwer. Es gibt wichtige Spezialfälle, da geht es automatisch; nämlich dann, wenn alle Basislösungen ganzzahlig sind.

Sei $\mathcal{X} = \{x \geq 0 : Ax = b\}$ und o.B.d.A. sei $A \in \mathbb{Z}^{m \times n}$.

Definition 1.7.1. *Eine Matrix $A \in \mathbb{Z}^{m \times n}$ vollen Zeilenranges heißt unimodular, falls die Determinante jeder aus m linear unabhängigen Spalten bestimmten Submatrix, den Wert ± 1 hat.*

Satz 1.7.1. *Sei $A \in \mathbb{Z}^{m \times m}$ regulär. $A^{-1}b$ ist ganzzahlig $\forall b \in \mathbb{Z}^m \iff A$ ist unimodular.*

Beweis. "$\Rightarrow$" Setzen $b = e_i$ für $i \in \{1, \ldots, m\} \Rightarrow A^{-1}e_i$ ist ganzzahlig $\Rightarrow A^{-1}$ ganzzahlig.

$$\Rightarrow \underbrace{\det A}_{\in \mathbb{Z}} \underbrace{\det A^{-1}}_{\in \mathbb{Z}} = \det I = \pm 1 \Rightarrow \det A = \pm 1.$$

"$\Leftarrow$" $Ax = b$ mit Cramerscher Regel lösen:

$$x_i = \frac{\overbrace{\det(A_{.,[1,i-1]}, b, A_{.,[i+1,m]})}^{\in \mathbb{Z}}}{\underbrace{\det A}_{= \pm 1}} \quad \Rightarrow x \in \mathbb{Z}^m.$$

□

Satz 1.7.2. *Es habe $A \in \mathbb{Z}^{m \times n}$ vollen Zeilenrang. Genau dann sind alle zulässigen Basislösungen von $\{x \geq 0 : Ax = b\}$ ganzzahlig für alle $b \in \mathbb{Z}^m$, wenn A unimodular ist.*

Beweis. "$\Leftarrow$" Sei A unimodular, $b \in \mathbb{Z}^m$, und $\overline{x}$ eine zulässige Basislösung von $\{x \geq 0, Ax = b\}$. Dann gibt es eine Basis B mit $\overline{x}_B = A_B^{-1}b$ ganzzahlig (weil A_B unimodular nach Satz 1.7.1) und $\overline{x}_N = 0 \Rightarrow \overline{x} \in \mathbb{Z}^n$.
"$\Rightarrow$" Sei B Basis von A, also A_B regulär. Nach Satz 1.7.1 ist zu zeigen: $A_B^{-1}b \in \mathbb{Z}^m \; \forall b \in \mathbb{Z}^m$.
Sei $b \in \mathbb{Z}^m$; müssen also $A_B^{-1}b$ "zulässig" machen. Wählen $w \in \mathbb{Z}^m$, so dass $w + A_B^{-1}b \geq 0$. Damit ist

$$\overline{b} = A_B(w + A_B^{-1}b) \in \mathbb{Z}^m.$$

Setze $\overline{x}_B = w + A_B^{-1}b$ und $\overline{x}_N = 0$, dann ist $\overline{x}$ zulässige Basislösung von $\{x \geq 0 : Ax = \overline{b}\}$, also ist $\overline{x}$ ganzzahlig und damit auch $A_B^{-1}b = \overline{x}_B - w$. □

Wann geht das auch für $\{x : Ax \leq b\}$ und alle $b \in \mathbb{Z}^m$?
$\Rightarrow [A\ I] \begin{bmatrix} x \\ s \end{bmatrix} = b$, also $[A\ I]$ unimodular.
$\Rightarrow$ Jede quadratische Untermatrix von A muss Determinante ± 1 oder 0 haben (wähle k Spalten aus A und $m - k$ Spalten aus I, mit Laplace-Entwicklungssatz folgt Aussage).

Definition 1.7.2. *Eine Matrix $A \in \mathbb{Z}^{m\times n}$ heißt total unimodular, wenn jede quadratische Untermatrix die Determinante ± 1 oder 0 hat.*

Ist A total unimodular, folgt also $\Rightarrow A \in \{0, +1, -1\}^{m\times n}$.

Beobachtung 1.7.3. *1. A total unimodular $\Longleftrightarrow [A\ I]$ unimodular.*

2. A total unimodular $\Longleftrightarrow [A, I, -A, -I]^\tau$ total unimodular.

3. A total unimodular $\Longleftrightarrow A^\tau$ total unimodular.

Satz 1.7.4 (Hoffmann und Kruskal (1956)). *Sei $A \in \mathbb{Z}^{m\times n}$ total unimodular und $b \in \mathbb{Z}^m$. Dann hat $\min\{c^\tau x \text{ s.t. } Ax \leq b\}$ immer eine ganzzahlige Optimallösung.*

Beweis. $\min\{c^\tau x \text{ s.t. } Ax \leq b\}$ ist äquivalent zu $\min\{c^\tau x \text{ s.t. } [A, -A, I][x_1, x_2, s]^\tau = b, x_1 \geq 0, x_2 \geq 0, s \geq 0\}$. Nach Beobachtung 1.7.3, (2., 3.), ist $[A, -A, I]$ total unimodular, nach Beobachtung 1.7.3, (1.) unimodular, also hat nach Satz 1.7.2 das zweite LP immer eine ganzzahlige Optimallösung $\overline{x_1}, \overline{x_2}$ und $x^\star = \overline{x_1} - \overline{x_2} \in \mathbb{Z}^m$ ist ganzzahlige Optimallösung für $\min\{c^\tau x \text{ s.t. } Ax \leq b\}$. □

Satz 1.7.5. *Ist A total unimodular, und ist bei $c \in \mathbb{Z}^n, b \in \mathbb{Z}^m$*

$$\max\{c^\tau x : Ax \leq b, x \geq 0\} = \min\{b^\tau y : A^\tau y \geq c, y \geq 0\}$$

endlich, dann werden die Optimallösungen auch in ganzzahligen Punkten $x^\star$ und $y^\star$ angenommen.

Beweis. Nach Satz 1.7.4 und nach Beobachtung 1.7.3 (2.) gilt das für $x^\star$ und wegen Beobachtung 1.7.3 (3.) auch für $y^\star$. □

Definition 1.7.3. *Ein (ungerichteter) Graph ist ein Paar $G = (V, E)$ bestehend aus einer (endlichen) Menge V von Knoten (nodes) und einer Menge $E \subseteq \{\{u, v\} : u, v \in V, u \neq v\}$ von Kanten (edges).*

Definition 1.7.4.

- *Eine Kantenmenge $P = \{\{u_1, u_2\}, \{u_2, u_3\}, \ldots, \{u_k, u_{k+1}\}\}$ mit $P \subseteq E$ in Graphen $G = (V, E)$ heißt Weg (der Länge k) (path), falls die u_i paarweise verschieden sind.*
- *Ein Graph heißt zusammenhängend (connected), wenn es zwischen je zwei Knoten $u, v \in V$ mit $u \neq v$ einen Weg gibt.*
- *Ein Graph $G' = (V', E')$ heißt ein Teil- oder Untergraph von $G = (V, E)$, falls $V' \subseteq V, E' \subseteq E$ und $E' \subset V' \times V'$.*
- *Die kantenmaximalen zusammenhängenden Teilgraphen eines Graphen heißen die Zusammenhangskomponenten von G.*

- *Eine Kantenmenge $C = \{\{u_1, u_2\}, \{u_2, u_3\}, \ldots, \{u_k, u_1\}\}$ mit $C \subseteq E$ in einem Graphen $G = (V, E)$ heißt ein Kreis (cycle) (der Länge k), falls die u_i paarweise verschieden sind.*

Definition 1.7.5.

- *Der vollständige Graph auf $|V| = n$ Knoten wird mit K_n bezeichnet.*
- *Ein Graph $G = (V, E)$ heißt bipartit, falls $\exists V_1, V_2$ mit $V = V_1 \cup V_2$ (disjunkt vereinigt) und $E \subseteq \{\{u, v\} : u \in V_1, v \in V_2\}$. Der vollständige bipartite Graph $|V_1| = n$ und $|V_2| = m$ Knoten wird mit $K_{n,m}$ bezeichnet.*

Satz 1.7.6. *Ein Graph $G = (V, E)$ ist bipartit genau dann, wenn er keine Kreise ungerader Länge besitzt.*

Beweis. "⇒" jeweils hin und zurück ⇒ gerade.
"⇐" Graph ist bipartit, wenn seine Zusammenhangskomponenten es sind. O.B.d.A. ist der Graph zusammenhängend. Wähle $u \in V$ und setze:
$V_1 = \{v \in V$: Länge eines kürzesten Weges von u nach v ist ungerade$\}$
$V_2 = \{v \in V$: Länge eines kürzesten Weges von u nach v ist gerade$\} \cup \{u\}$
Annahme: $\exists$ Kante $\{v, w\} \in E$ mit ($v \in V_1$ und $w \in V_1$) oder ($v \in V_2$ und $w \in V_2$). Sei P_v ein kürzester uv-Weg und P_w ein kürzester uw-Weg. Sei $\bar{u}$ der letzte gemeinsame Knoten auf diesen Wegen. Da beide Wege kürzeste sind, ist der Abschnitt u nach $\bar{u}$ jeweils gleich lang, damit sind die Restwege $\bar{P}_v$ von $\bar{u}$ nach v und $\bar{P}_w$ von $\bar{u}$ nach w jeweils beide gerade oder ungerade. $C = \bar{P}_v \cup \bar{P}_w \cup \{v, w\}$ hat ungerade Länge ⇒ Widerspruch. □

Satz 1.7.7 (Heller und Tompkins (1956)). *Sei $A \in \{0, 1, -1\}^{m \times n}$ mit höchstens zwei nichtverschwindenden Einträgen pro Spalte. A ist total unimodular $\Longleftrightarrow$ Die Zeilen von A können in zwei Klassen eingeteilt werden, so dass Zeilen mit einem $+1$ und einem -1 Eintrag in der gleichen Spalte in die gleiche Klasse und Zeilen mit zwei vorzeichengleichen Einträgen in der gleichen Spalte in unterschiedliche Klassen kommen.*

Beweis. "⇒" Spalten mit nur einem Eintrag sind vernachlässigbar. Fassen zuerst Zeilen gemäß 1. zu Zeilenmengen $Z_i, i = 1, \ldots, h$, zusammen.
Annahme: Es kommt dabei zu einem Widerspruch zu 2.; also gibt es eine Untermatrix der Form

$$B = \begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ k \end{array} \begin{pmatrix} 1 & 0 & 0 & \cdots & 1 \\ -1 & 1 & & & 0 \\ 0 & -1 & \ddots & & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

Dann ist $\det B = B_{1,1} \det B_{2:k,2:k} + (-1)^{k-1} B_{1,k} \det B_{2:k,1:k-1}$ ⇒ Widerspruch zu
$= 1 \cdot 1 + (-1)^{k-1} \cdot (-1)^{k-1} = 2.$
"B ist Untermatrix einer total unimodularen Matrix".
Bauen nun Graphen $G = (V, E)$ mit $V = \{Z_i : i = 1, \ldots n\}$, also Zeilenmengen sind Knoten, und $\{Z_i, Z_j\} \in E \Leftrightarrow \exists$ Spalte $\bar{j}$ mit $A_{\bar{z},\bar{j}} = A_{\bar{\bar{z}},\bar{j}} \neq 0, \bar{z} \in Z_i, \bar{\bar{z}} \in Z_j$, also eine Kante wird

eingeführt, wenn gemäß 2. die Zeilenmengen in unterschiedliche Klassen gehören.
Behauptung: G ist bipartit.
Annahme: G ist nicht bipartit. Nach Satz 1.7.6 gibt es einen ungeraden Kreis, also gibt es eine Untermatrix der Form

$$B = \begin{pmatrix} \overbrace{1}^{\text{Kante } e_1} & 0 & 0 & \cdots & \overbrace{1}^{\text{Kante } e_k} \\ 1 & 1 & & & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 1 & 1 \end{pmatrix}$$

mit $\det B = 1 + (-1)^{k-1} \cdot 1 = 2 \Rightarrow$ Widerspruch zu A ist total unimodular (Untermatrix B ist total unimodular).
Beachte: Sind für $Z_i = e_i \cap e_j$ die Zeilen $\bar{z}$ und $\bar{\bar{z}} \in Z_i$ unterschiedlich, so muss eigentlich statt

$$B = \begin{matrix} \bar{z} \to \\ \\ \\ \\ \bar{\bar{z}} \to \end{matrix} \begin{pmatrix} 1 & 0 & 0 & \cdots & 1 \\ 1 & 1 & & & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 1 & 1 \end{pmatrix} \rightsquigarrow \begin{matrix} \bar{z} \to \\ \\ \\ \\ \bar{\bar{z}} \to \end{matrix} \begin{pmatrix} 1 & 0 & 0 & \cdots & 1 \\ 1 & 1 & & & 0 \\ 0 & -1 & \ddots & & \vdots \\ \vdots & & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

geschrieben werden. Aber jede -1 erhöht auch k um eins und verändert das Ergebnis daher nicht.

"$\Leftarrow$" Sei B eine beliebige quadratische Untermatrix. O.B.d.A. hat B nur Spalten mit 2 Einträgen $\neq 0$ (die anderen können leicht mit Laplaceschem Entwicklungssatz behandelt werden). Sei $\{Z_1, Z_2\}$ eine Partition der Zeilen gemäß 1. und 2., multipliziere die Zeilen aus Z_1 mit $+1$, die aus Z_2 mit -1. Dann ist die Summe der Zeilen der Nullvektor, also sind sie linear abhängig. □

1.7.1 Anwendung: Bipartite Paarung (Matching/Zuweisungsproblem)

Definition 1.7.6. *In einem Graphen $G = (V, E)$ heißt eine Kantenmenge $M \subseteq E$ mit $e \cap f = \varnothing \; \forall \, e, f \in M : e \neq f$ Matching oder Paarung.*
Ein Matching heißt perfekt, falls $V(M) := \bigcup_{e \in M} e = V$.

Also: Keine zwei Kanten in M haben einen Knoten gemeinsam. In einem perfekten Matching werden alle Konten überdeckt.

In den Anwendungen sind meist Gewichte w_e für $e \in E$ gegeben und man sucht ein Matching maximalen Gewichts: $\max \left\{ \sum_{e \in M} w_e : M \subseteq E, M \text{ Matching} \right\}$. Ist G bipartit, spricht man von einem bipartiten Matching.

Beispiel 1.7.1. *Arbeiter zu Maschinen zuordnen:*
Eine Kante zwischen einem Arbeiter und einer Maschine gibt an, dass er sie bedienen kann; das entsprechende Gewicht, wie gut er dabei ist. Gesucht ist eine Zuordnung mit maximaler Gesamtgüte.

Modellierung als LP

Definition 1.7.7.

- *In einem Graphen $G = (V, E)$ heißen zwei Knoten u, v adjazent, wenn $\{u, v\} \in E$.*
- *Eine Kante e und ein Knoten u heißen inzident, wenn $u \in e$.*
- *Sei $G = (V, E)$ mit Kantengewichten $w_e \in \mathbb{R}^{|E|}$, dann ist die gewichtete Adjazenzmatrix die (symmetrische) Matrix $A \in \mathbb{R}^{|V|\times|V|}$ mit*

$$a_{i,j} = \begin{matrix} w_{\{i,j\}}, & \textit{falls } \{i,j\} \in E \\ 0, & \textit{sonst.} \end{matrix}$$

- *Die Knoten-Kanteninzidenzmatrix ist die Matrix $A \in \mathbb{R}^{|V|\times|E|}$ mit*

$$a_{u,e} = \begin{cases} 1, & \textit{falls } u \in e \\ 0, & \textit{sonst.} \end{cases}$$

- *Der Inzidenzvektor oder charakteristische Vektor einer Teilmenge B einer Obermenge $\mathcal{O}$ ist der Vektor $\chi(B) \in \{0,1\}^{|\mathcal{O}|}$ mit*

$$\chi(B)_i = \begin{cases} 1, & \textit{falls } i \in B \\ 0, & \textit{sonst.} \end{cases}$$

Sei A die Knoten-Kanteninzidenzmatrix eines bipartiten Graphen $G = (V, E)$. Dann ist

- A total unimodular (nach 1.7.7; Z_1: Zeilen zu V_1, Z_2: Zeilen zu V_2).
- Sei $M \subseteq E$. M ist ein Matching genau dann, wenn $A\chi(M) \leq e = (1, \ldots, 1)^\tau$ ist, d.h. wenn jeder Knoten höchstens einmal überdeckt wird.

Lineares Programm für Matching maximalen Gewichts:

$$\begin{array}{l} \max w^\tau x \\ \text{s.t. } Ax \leq e \\ \quad\quad x \geq 0 \end{array}$$

Simplex liefert optimale Basislösung $x^\star$ und nach dem Satz von Hoffman und Kruskal (1.7.4) ist $x^\star$ auch ganzzahlig, also ist $x^\star \in \{0,1\}^{|E|}$, also charakteristischer Vektor eines Matchings maximalen Gewichtes für bipartite Graphen. Ist $w = e$, so liefert Satz 1.7.5:

$$\max\{e^\tau x : Ax \leq e, x \geq 0\} = \min\{e^\tau y : A^\tau y \geq e, y \geq 0\}$$

und bei beiden sind optimale Basislösungen ganzzahlig.
$A^\tau y \geq e$ bedeutet $y \in \{0,1\}^{|V|}$ und $y_u = 1 \Rightarrow$ alle zu u inzidenten Kanten werden überdeckt.

Satz 1.7.8 (König und Egerváry). *In einem bipartiten Graphen ist die maximale Kardinalität eines Matchings gleich der minimalen Überdeckung aller Kanten durch Knoten.*

Beweis. Siehe obige Überlegungen. □

Spezialfall: perfektes Matching in $K_{n,n}$

$$A = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ n \\ \hline 1 \\ 2 \\ \vdots \\ n \end{array} \left[\begin{array}{cccc|cccc|c|cccc} 1 & 1 & 1 & 1 & & & & & & & & \\ & & & & 1 & 1 & 1 & 1 & & & & \\ & & & & & & & & \cdots & & & \\ & & & & & & & & & 1 & 1 & 1 & 1 \\ \hline 1 & & & & 1 & & & & & 1 & & \\ & 1 & & & & 1 & & & & & 1 & \\ & & \ddots & & & & \ddots & & \cdots & & & \ddots & \\ & & & 1 & & & & 1 & & & & 1 \end{array}\right], e = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$Ax = e$ bedeutet: "Jeder Knoten wird genau einmal überdeckt."

$$\begin{array}{c|ccc} \frac{Kanten}{Knoten} & 1 & & n \\ \hline 1 & x_{\{1,1\}} & \cdots & x_{\{1,n\}} \\ \vdots & \vdots & \ddots & \vdots \\ n & \underbrace{x_{\{n,1\}}}_{\sum=1} & \cdots & x_{\{n,n\}} \end{array} \} \sum = 1$$

$\mathcal{X} := \{x \geq 0 : Ax = e\} \Longleftrightarrow$ Menge der nichtnegativen $n \times n$-Matrizen mit Spalten- und Zeilensumme 1 ("*doppelt stochastische Matrizen*").

Jede Ecke von $\mathcal{X}$ ist ganzzahlig und entspricht einer "*Permutationsmatrix*".

Beispiel: $P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Satz 1.7.9 (Birkhoff). *Doppelt stochastische Matrizen sind Konvexkombinationen der Permutationsmatrizen.*

Das Transportproblem

Gegeben: eine Menge von Produzenten $i \in P$, die alle dasselbe Gut anbieten
eine Menge von Konsumenten $j \in K$, die alle dieses Gut nachfragen
a_i sei die von $i \in P$ angebotene Menge
b_j sei der von $j \in K$ nachgefragte Bedarf

Annahme: $\sum_{i \in P} a_i = \sum_{j \in K} b_j$ (sonst: führen künstlichen Produzenten/Konsumenten ein)

$\Rightarrow$ Modellierung über bipartiten Graphen $G = (V, E)$ mit $V = P \cup K, E = P \times K$
$c_e = c_{\{i,j\}}$ seien die Transportkosten für Transport einer Einheit von i nach j.
$x_e = x_{\{i,j\}}$ sei die von i nach j transportierte Menge.

Gesucht: kostenminimaler Transportplan.

$$\begin{cases} \min \sum_{e \in E} c_e x_e \\ \text{s.t.} \sum_{j \in K} x_{\{i,j\}} = a_i \quad \forall i \in P \\ \quad \sum_{i \in P} x_{\{i,j\}} = b_j \quad \forall j \in K \\ \quad x_e \geq 0 \quad \forall e \in E \end{cases}$$

Betrachten Knoten-Kanteninzidenzmatrix A mit $A_{i,j} = \begin{cases} 1 & \exists u \in V : j = \{i, u\} \\ 0 & \text{sonst.} \end{cases}$

Beispiel:

$$A = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 & 1 \\ 4 & 1 & 0 & 1 & 0 & 1 & 0 \\ 5 & 0 & 1 & 0 & 1 & 0 & 1 \end{array},$$

$x = (x_e)_{e \in E} \in \mathbb{R}_+^{|E|} (x \in \mathbb{R}_+^6)$

$A_{(1,:)}x = x_1 + x_2 = x_{\{1,4\}} + x_{\{1,5\}} = a_1$
$A_{(:,4)}x = x_1 + x_3 + x_5 = x_{\{1,4\}} + x_{\{2,4\}} + x_{\{3,4\}} = b_1$

Damit erhält man die Aufgabe in Matrixschreibweise: $\min \left\{ c^\tau x : Ax = \begin{pmatrix} a \\ b \end{pmatrix}, x \geq 0 \right\}$.

Matrix A ist total unimodular (nach Satz 1.7.7 (Heller und Tompkins): alle Zeilen zu $i \in P$ in eine Klasse, alle Zeilen zu $j \in K$ in andere Klasse).
$\Rightarrow$ Simplexmethode liefert (falls a, b ganzzahlig) ganzzahlige Lösung.

1.7.2 Flüsse in Netzwerken

Rohrleitung, Verkehrsströme, Datenströme,...
Betrachten dazu gerichtete Graphen:

Definition 1.7.8. *Ein gerichteter Graph oder Digraph (directed graph) ist ein Paar $D = (V, E)$ aus einer Knotenmenge V und einer Menge E gerichteter Kanten (Pfeile, geordnete Knotenpaare, arcs), d.h. $E \subseteq \{(u, v) : u, v \in V, u \neq v\}$.*
Die Knoten-Kanteninzidenzmatrix $A \in \{0, 1, -1\}^{|V| \times |E|}$ ist gegeben durch

$$a_{i,j} = \begin{cases} 1 & \text{falls } \exists u \in V : j = (u, i) \in E \text{ (Pfeil } j \text{ endet in } i) \\ -1 & \text{falls } \exists u \in V : j = (i, u) \in E \text{ (Pfeil } j \text{ beginnt in } i) \\ 0 & \text{sonst.} \end{cases}$$

Beobachtung 1.7.10. *Die Knoten-Kanteninzidenzmatrix eines gerichteten Graphen ist total unimodular.*

Beweis. Mit Satz 1.7.7: Weil in jeder Spalte genau eine $+1$ und genau eine -1 ist, liegen alle Zeilen sogar in nur einer Klasse. □

Bemerkung 1.7.11.

- *Mehrfachkanten sind möglich (Beobachtung 1.7.10 bleibt gültig).*
- *$rank(A) < |V|$, weil die Zeilen von A linear abhängig sind (Summe aller Zeilen $= 0$).*

Definition 1.7.9.

- *Ein Netzwerk besteht aus einem gerichteten Graphen $D = (V, E)$ und Kapazitäten auf den Kanten, also $w_e \geq 0, \forall e \in E$.*
- *Ein Fluss ist eine Funktion $f : E \to \mathbb{R}_+$, für die gilt:*

$$\underbrace{\sum_{v:e=(u,v)\in E} f_e}_{\text{vom Knoten u ausfließender Fluss}} = \underbrace{\sum_{v:e=(v,u)\in E} f_e}_{\text{im Knoten u ankommender Fluss}} \qquad \forall u \in V \text{ (“Flusserhaltungsbedingungen”)}$$

- *Gilt zusätzlich $0 \leq f_e \leq w_e \; \forall e \in E$, heißt der Fluss zulässig.*

Mit $x \in \mathbb{R}_+^{|E|}, x_e = f_e = f(e)$ sind also die Flusserhaltungsbedingungen $Ax = 0$ und die Zulässigkeitsbedingungen $0 \leq x \leq w$.

$$\begin{cases} Ax = 0 \\ 0 \leq x \leq w \end{cases} \quad \Rightarrow \quad \underbrace{\begin{bmatrix} A \\ -A \\ I \\ -I \end{bmatrix}}_{\text{total unimodular nach Beobachtung 1.7.10 und 1.7.3}} x \qquad \leq \begin{bmatrix} 0 \\ 0 \\ w \\ 0 \end{bmatrix}$$

$\Rightarrow$ wichtig für ganzzahlige Optimierungsprobleme.

Sowohl untere Kapazitätsschranken $l \leq x$ als auch Zufluss/Abfluss in gewissen Knoten lassen sich mit geeigneten rechten Seiten modellieren.

Definition 1.7.10. *In einem Digraphen $D = (V, E)$ heißt eine Knotenmenge*

$$\{(v_1, v_2), \ldots, (v_k, v_1)\} \subseteq E \text{ mit } v_i \neq v_j \; \forall i \neq j$$

gerichteter Kreis der Länge k. Die Menge aller Kreise in D wird mit $\mathcal{C}(D)$ bezeichnet.

Beobachtung 1.7.12. *Jeder (zulässige) Fluss lässt sich darstellen als nichtnegative Linearkombination der Kreise, d.h. $\exists f_C \geq 0, C \in \mathcal{C}(D)$, so dass*

$$x = \sum_{C \in \mathcal{C}} f_C \chi(C).$$

Dabei bezeichnet $\chi(C) \in \{0,1\}^{|E|}$ den Inzidenzvektor von C, d.h. $(\chi(C))_e = \begin{cases} 1, e \in C \\ 0, e \notin C. \end{cases}$

Beweis. Falls $x = 0$ fertig ($f_C = 0$).
Sei also $x \neq 0$, d.h. $\exists (v_1, v_2) \in E : x_{(v_1,v_2)} > 0 \Rightarrow \exists (v_2, v_3) \in E : x_{(v_2,v_3)} > 0$ (wegen Flusserhaltung) usw. Solange fortsetzen, bis erstmalig ein bereits besuchter Knoten auftritt $\Rightarrow$ gerichteter Kreis $C \in \mathcal{C}(D) \Rightarrow$ wähle $f_C = \min\{x_e : e \in C\} > 0 \Rightarrow$ setze $\bar{x} = x - f_C \chi(C)$ und setze fort mit neuem Fluss $\bar{x}$. Dies endet nach endlich vielen Schritten, weil in jedem Schritt mindestens eine Kante auf Null gesetzt wird. □

Maximaler s-t-Fluss

Für ein gegebenes Netzwerk $D = (V, E)$ mit Kapazitäten w sucht man einen "Fluss", der möglichst viel Material von einem Knoten s (Quelle, source) zu einem Knoten t (Senke, sink) mit $s \neq t$ unter Beachtung der Kantenkapazitäten transportiert.

Modellierung: führe künstliche Kante von t nach s mit "unendlicher" Kapazität ein (endlich reicht: $w_{\{t,s\}} = 1 + \sum_{e \in E} w_e$).

primal:

$$\begin{array}{l} \max x_{(t,s)} \\ \text{s.t. } Ax = 0 \\ \qquad 0 \underbrace{\leq}_{q} x \underbrace{\leq}_{p} w \end{array}$$

dual:

$$\begin{array}{l} \min w^\tau p \\ \text{s.t. } A^\tau y + p - q = e_{(t,s)} \\ \qquad p \geq 0, q \geq 0 \end{array}$$

Äquivalente Formulierung der Dualen:

$$\begin{array}{ll} \min w^\tau p & \\ \text{s.t. } y_j - y_i + p_{(i,j)} - q_{(i,j)} = 0 & \forall (i,j) \neq (t,s) \in E \\ \qquad y_s - y_t + p_{(t,s)} - q_{(t,s)} = 1 & \\ \qquad p \geq 0, q \geq 0. & \end{array}$$

Definition 1.7.11.
- *Sei x ein zulässiger Fluss in $D = (V, E)$ mit Kapazität w. Der Wert des $s-t$-Flusses ist $x_{(t,s)}$.*
- *Sei $D = (V, E)$ ein Digraph und $S \subseteq V$, dann heißt die Kantenmenge*
$$\delta^+(S) = \{(i, j) \in E : i \in S, j \in V \setminus S\}$$
"Schnitt" (cut) in D.
- *Ist $s \in S$ und $t \in V \setminus S$, so heißt $\delta^+(S)$ $s-t$-Schnitt in D.*
- *In einem Netzwerk $D = (V, E)$ mit Kapazität w ist für $S \subseteq V$ der Wert des Schnitts $w(\delta^+(S)) := \sum_{e \in \delta^+(S)} w_e$.*

Beobachtung 1.7.13. *Der Wert jedes $s-t$-Flusses ist kleiner gleich dem Wert jedes $s-t$-Schnitts.*

Beweis. Sei x ein $s-t$-Fluss. Wegen Beobachtung 1.7.12 ist $x = \sum_{C \in \mathcal{C}(D)} f_C \chi(C)$. Der Wert des Flusses ist $x_{(t,s)}$. Sei $S \subseteq V$ mit $s \in S$ und $t \in V \setminus S$. Für alle $C \in \mathcal{C}(D)$ mit $(t, s) \in C$ gibt es mindestens eine Kante $e(C)$ mit $e \in \delta^+(S)$. Sei $F = \bigcup_{C \in \mathcal{C}(D),(t,s) \in C} e(C) \subseteq \delta^+(S)$

$$\Rightarrow\ x_{(t,s)} = \sum_{C \in \mathcal{C}(D),(t,s) \in C} f_C \leq \sum_{e \in F} x_e \leq \sum_{e \in \delta^+(S)} x_e \leq \sum_{e \in \delta^+(S)} w_e = w(\delta^+(S)).$$

□

Satz 1.7.14 (Max-Flow-Min-Cut Theorem, Ford und Fulkerson). *In einem Netzwerk mit Kapazität $w \in \mathbb{Z}_+^{|E|}$ ist der Wert eines maximalen $s-t$-Flusses gleich dem Wert eines minimalen $s-t$-Schnittes.*

Beweis. Nebenbedingungsmatrix ist total unimodular, rechte Seite (der Dualen) ist ganzzahlig $\Rightarrow$ Satz 1.7.4: Duales hat eine ganzzahlige Optimallösung $\Rightarrow$ Primales und Duales haben den gleichen Wert $x_{(t,s)} < w_{(t,s)} = \sum w_i + 1$.
Sei $0 \leq x_{(t,s)} < w_{(t,s)} \Rightarrow$ Satz vom komplementären Schlupf: $p_{(t,s)} = 0 \Rightarrow y_s - y_t = 1 + q_{(t,s)} \geq 1$ und y ganzzahlig. Wähle $\bar{y} \in \{y_s, y_s - 1, \ldots, y_t + 1\}$. Setze $S = \{i : y_i \geq \bar{y}\}$.
Behauptung $w(\delta^+(S)) = x_{(t,s)}$.
Offensichtlich ist $y_s \geq \bar{y}$, also $s \in S$ und $y_t < \bar{y}$, also $t \in V \setminus S$, d.h. S ist ein $s-t$-Schnitt. Sei $i \in S, j \in V \setminus S$ sowie $(i, j) \neq (t, s) \in E$. Dann ist $y_j - y_i \leq -1$, also $p_{(i,j)} - q_{(i,j)} \geq 1$, also muss $p_{(i,j)} \geq 1$ sein. Nach dem Satz vom komplementären Schlupf folgt $x_{(i,j)} = w_{(i,j)}$.
Sei $i \in V \setminus S$ und $j \in S$ mit $(i, j) \in E$, dann ist $y_i - y_j \geq 1$, also $q_{(i,j)} \geq 1$ und nach dem Satz vom komplementäen Schlupf ist $x_{(i,j)} = 0$.
Damit gilt nach Beobachtung 1.7.13

$$x_{(t,s)} = \sum_{(i,j) \in \delta^+(S)} x_{(i,j)} = \sum_{(i,j) \in \delta^+(S)} w_{(i,j)} = w(\delta^+(S)).$$

□

Bemerkung 1.7.15. *Für jedes $\bar{y}$ bekommen wir eine minimalen Schnitt. Nach dem Beweis ist $p_{(i,j)} \geq 1\ \forall (i, j) \in \delta^+(S)$ und $\sum w_{(i,j)} p_{(i,j)} = w(\delta^+(S)) \Rightarrow p_{(i,j)} = 1$; damit ist $y_s = y_t + 1$. Trotzdem kann es noch andere minimale Schnitte geben.*

Minimale Kostenflüsse (Min-Cost-Flow-Problem)

Oft ist bekannt, wieviel in einem Netzwerk von einem Ort zu einem anderen fließen soll. Gesucht wird nur nach einer Routenwahl zu minimalen Kosten.

- Für jeden Knoten wird angegeben, wieviele Einheiten dort entstehen oder verschwinden: pro Knoten $i \in V$ sei $b_i \in \mathbb{R}$ (positiv oder negativ) die "*Balance*".
- Pro Kante kostet der Transport einen gewissen Preis pro Einheit: für $e \in E$ sei dies $c_e \in \mathbb{R}$.

Definition 1.7.12. • *Auf einem Netzwerk $D = (V, E)$ mit Kapazitäten w und Balancen b heißt eine Funktion $f : E \to \mathbb{R}$, die die Flusserhaltungsgleichungen*

$$\sum_{(i,j)\in E} f(i,j) - \sum_{(j,i)\in E} f(j,i) = b_j \; \forall j \in V$$

erfüllt, "balancierter Fluss".

- *Ein balancierter Fluss heißt zulässiger Fluss, falls er die Kapazitätsbedingungen einhält, also*

$$0 \leq f(i,j) \leq w_{(i,j)} \quad \forall (i,j) \in E.$$

Ein Problem mit unteren Schranken an den Fluss $l_{u,v} \leq f \leq w_{u,v}$ kann immer in ein Problem mit unterer Schranke 0 umgewandelt werden: $0 \leq f' \leq w_{u,v} - l_{u,v}$. Das kann nachträglich wieder zurückgesetzt werden: $f = f' + l_{u,v}$. Wie vorher gilt:

- Sind b und w ganzzahlig, sind alle primal optimalen Basislösungen ganzzahlig.
- Ist c ganzzahlig, dann sind die dual optimalen Basislösungen ganzzahlig.

Für alle Netzwerkprobleme dieser Art gibt es eine besonders effiziente Art, den Simplexalgorithmus durchzuführen. Betrachten:

$$\begin{aligned} &\min c^\tau x \\ &\text{s.t. } Ax = b \\ &\quad 0 \leq x \leq w, \end{aligned}$$

wobei A die Knoten-Kanteninzidenzmatrix eines zusammenhängenden Digraphen $D = (V, E)$ mit $|V| = n$ sei.
Zeile i: Flusserhaltungsgleichung (balanciert) für Knoten i.
Zeilen von A sind linear abhängig $\Rightarrow$ Löschen Zeile zu Knoten n: $\rightsquigarrow \tilde{A}$. Damit erhält man

$$\begin{aligned} &\min c^\tau x \\ &\text{s.t. } \tilde{A}x = \tilde{b} \\ &\quad 0 \leq x \leq w. \end{aligned}$$

Eine Basis muss $n - 1$ mal linear unabhängige Spalten enthalten.

Lemma 1.7.16. *Sei $B \subseteq E$ eine Basis von $\tilde{A}$. Definieren $\tilde{G} = (V, \tilde{E})$ mit $\tilde{E} = \{\{i, j\} : (i, j) \in B\}$. Es gilt:*

1. *B enthält mindestens eine Kante e, die mit Knoten n inzidiert und $\tilde{G}$ ist zusammenhängend.*
2. *$\tilde{G}$ enthält keinen Kreis (im ungerichteten Sinn).*

Beweis. 1. Falls nicht, ist in jeder Spalte genau eine $+1$ und eine $-1 \Rightarrow e^\tau \tilde{A}_B = 0$. Also sind die Zeilen nicht linear unabhängig und B keine Basis.
Gleiches Argument gilt, falls $\tilde{G}$ nicht zusammenhängend.
2. $\tilde{G}$ hat $n - 1$ Kanten, $\tilde{G}$ ist zusammenhängend. Nach Übung 5.3 enthält $\tilde{G}$ keinen Kreis ($\tilde{G}$ ist ein Baum). □

Definition 1.7.13. *Ein zusammenhängender ungerichteter Graph ohne Kreise heißt ein Baum. Ein Knoten in einem Baum heißt Blatt, falls nur eine Kante des Baumes mit ihm inzidiert.*

$\Rightarrow$ Jede Basis entspricht im ungerichteten Sinn einem Baum.

Satz 1.7.17. *$B \subseteq E$ ist genau dann eine Basis von $\tilde{A}$, wenn B im ungerichteten Sinn den Kanten eines Baumes auf V entspricht.*

Beweis. "$\Rightarrow$" Lemma 1.7.16
"$\Leftarrow$" Bestehe der Baum nur aus einer Kante $\{v_1, n\}$, dann ist $\tilde{A}_B = (\pm 1)$ und $A_B x = b$ ist eindeutig lösbar. Allgemein hat der Baum mindestens ein Blatt $\neq n$ (Übung), also hat $\tilde{A}_B$ eine Zeile mit nur einem Eintrag $\neq 0$, dies sei in Spalte e. Berechne x_e und eliminiere die Spalte. $\Rightarrow$ Es entsteht ein Baum mit einem Knoten weniger. Wiederhole dies. □

Beispiel 1.7.2.

$$\left(\begin{array}{cccc} 0 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline -1 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{c} x_{(n,2)} \\ x_{(1,4)} \\ x_{(3,1)} \\ x_{(1,n)} \end{array}\right) = \left(\begin{array}{c} 2 \\ 6 \\ -10 \\ 4 \\ \hline 2 \end{array}\right)$$

$$\Rightarrow x_{(n,2)} = 6 \Rightarrow x_{(1,4)} = 4 \Rightarrow x_{(3,1)} = 10 \Rightarrow x_{(1,n)} = 4$$

Also: Das Gleichungssystem $\tilde{A}_B x = b$ ist sehr schnell und einfach lösbar.
Pivot bedeutet: neue Kante aufnehmen $\rightarrow$ es entsteht Kreis (reduzierte Kosten)
Kante aus dem Kreis entfernen (Ratio Test)

reduzierte Kosten: aus der Dualen

$$(D) \quad \begin{cases} \max b^\tau y - w^\tau \bar{z} \\ \text{s.t. } \tilde{A}^\tau y - \bar{z} + z = c \\ \quad \bar{z} \geq 0, z \geq 0 \end{cases}$$

$$\Leftrightarrow \begin{bmatrix} \tilde{A}_B^\tau \\ \tilde{A}_N^\tau \end{bmatrix} y - \begin{bmatrix} \bar{z}_B \\ \bar{z}_N \end{bmatrix} + \begin{bmatrix} z_B \\ z_N \end{bmatrix} = \begin{bmatrix} c_B \\ c_N \end{bmatrix}$$

$$\Leftrightarrow \begin{cases} A_B^\tau y = c_B + \underbrace{\bar{z}_B}_{=0} - \underbrace{z_B}_{=0} \\ -\bar{z}_N + z_N = c_N - \tilde{A}_N^\tau y \quad \text{(höchstens eines der } \bar{z}_i, z_i \neq 0). \end{cases}$$

Löse $\tilde{A}_B^\tau y = c_B$. Die Spalten von $\tilde{A}_B^\tau$ entsprechen den Knoten. y_i entspricht Knoten $i \in V \setminus \{n\}$, $y_n := 0$ (setzen wir so).
Fassen y_i als Knotenpotentiale auf:

$$\left(\begin{array}{cccc|c} 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \hline y_n (= 0) \end{array}\right) = \begin{pmatrix} c_{(n,2)} \\ c_{(1,n)} \\ c_{(3,1)} \\ c_{(1,n)} \end{pmatrix}$$

Von der Wurzel her entlang der Baumkanten auflösen:
Sei $(i,j) \in B$ und i näher an der Wurzel und y_i schon bestimmt $\Rightarrow y_j = c_{(i,j)} + y_i$
j näher an der Wurzel und y_i schon bestimmt $\Rightarrow y_i = y_j - c_{(i,j)}$

$\Rightarrow y$ sehr schnell und einfach bestimmbar.

Reduzierte Kosten für $(i,j) \in N$ je nachdem bestimmen, ob $x_{(i,j)} = 0$ oder $x_{(i,j)} = w_{(i,j)}$:

$x_{(i,j)} = 0$ $\Rightarrow$ (Satz vom kompl. Schlupf) $\bar{z}_{(i,j)} = 0$ (gehört zu $x_{(i,j)} \leq w_{(i,j)}$)
reduzierte Kosten: $z_{(i,j)} = c_{(i,j)} - (1 \cdot y_j - 1 \cdot y_i) = c_{(i,j)} + y_i - y_j$

$x_{(i,j)} = w_{(i,j)}$ $\Rightarrow$ (Satz vom kompl. Schlupf) $z_{(i,j)} = 0$ (gehört zu $x_{(i,j)} \geq 0$)
reduzierte Kosten: $\bar{z}_{(i,j)} = (1 \cdot y_j - 1 \cdot y_i) - c_{(i,j)} = y_j - y_i - c_{(i,j)}$

Kante (i,j) als Pivotspalte wählbar, falls

$$(x_{(i,j)} = 0 \text{ und } z_{(i,j)} < 0) \text{ oder } (x_{(i,j)} = w_{(i,j)} \text{ und } \bar{z}_{(i,j)} < 0).$$

Wieder alle reduzierten Kosten effizient bestimmen. Sei Kante (i,j) die eintretende Variable.

Ratio Test: Wieviel kann $x_{(i,j)} = \begin{matrix} 0 & \text{vergrößert} \\ w_{(i,j)} & \text{verkleinert} \end{matrix}$ werden?
(i,j) bildet einen Kreis $C \subset B \cup \{(i,j)\}$, nur die Variablen aus dem Kreis sind betroffen, weil sie

linear abhängig sind. Wie findet man den Kreis? Zur Wurzel hin nach dem ersten gemeinsamen Punkt suchen. Veränderungen um $\Delta x = \Delta x_{(i,j)}$ in $x_{(i,j)}$ führen für Kante $(k,l) \in C$ zu

$$\Delta x_{(k,l)} = \begin{array}{l} +\Delta x, \text{ falls } (k,l) \text{ im Kreis gleich orientiert ist wie } (i,j) \\ -\Delta x, \text{ falls } (k,l) \text{ im Kreis gegenorientiert ist wie } (i,j). \end{array}$$

Für $\begin{array}{ll} x_{(i,j)} = 0 & \text{finde maximales } \Delta x \\ x_{(i,j)} = w_{(i,j)} & \text{finde minimales } \Delta x, \end{array}$ so dass $0 \leq x_{(k,l)} + \Delta x_{(k,l)} \leq w_{(k,l)} \quad \forall (k,l) \in C.$

Wähle eine begrenzende Kante $(k,l) \in C$ als austretende Variable, setze

$$B^+ = B \setminus \{(k,l)\} \cup \{(i,j)\}$$

und setze fort.

Man erkennt sofort (da nur Addition und Subtraktion): b, w ganzzahlig $\Rightarrow x$ ganzzahlig,
c ganzzahlig $\Rightarrow y, z, \bar{z}$ ganzzahlig.

Wie findet man eine zulässige Basis?

Neuen künstlichen Knoten einfügen mit künstlichen Kanten zu allen anderen Knoten, deren Fluss alle Balancen befriedigt. Phase I beseitigt den Fluss auf den künstlichen Kanten.

Kapitel 2

Konvexe Analysis

Gegenstand: notwendige und hinreichende Optimalitätsbedingungen für konvexe Probleme
Grundlagen für konvexe Optimierungsalgorithmen

2.1 Konvexe Mengen

Definition 2.1.1.

- *Eine Menge $C \subseteq \mathbb{R}^n$ heißt konvex, wenn mit $x, y \in C$ auch*

$$\alpha x + (1-\alpha)y \in C \quad \forall \alpha \in (0,1).$$

- *Eine Menge $C \subseteq \mathbb{R}^n$ heißt konvexer Kegel, wenn mit $x, y \in C$ auch*

$$\lambda(x+y) \in C \quad \forall\, \lambda > 0.$$

Wichtige konvexe Mengen:

Hyperebene: $H_{s,r} = \{x \in \mathbb{R}^n : \langle s, x\rangle = r\}$ für gegebenes $s \in \mathbb{R}^n, r \in \mathbb{R}$
abgeschlossener Halbraum: $\{x \in \mathbb{R}^n : \langle s, x\rangle \leq r\}$
offener Halbraum: $\{x \in \mathbb{R}^n : \langle s, x\rangle < r\}$
Einheitssimplex: $\triangle_k = \{\alpha \in \mathbb{R}^k : \sum_{i=1}^k \alpha_i = 1, \alpha_i \geq 0, i = 1 \ldots k\}$
Kegel: nichtnegativer Orthant: $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq, i = 1 \ldots n\}$

2.1.1 Grundlegende Operationen auf konvexen Mengen

Beobachtung 2.1.1. *1. Sei $\{C_j\}_{j\in J}$ eine beliebige Familie konvexer Mengen. Dann ist $\bigcap_{j\in J} C_j$ konvex.*

2. Seien $C_i \subseteq \mathbb{R}^{n_i}$ konvex für $i = 1 \ldots k$. Dann ist $C_1 \times \ldots \times C_k$ konvex in $\mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_k}$.

3. *Sei $A : \mathbb{R}^n \to \mathbb{R}^m$ eine affine Abbildung. Dann ist für konvexes $C \subseteq \mathbb{R}^n$ das Bild $A(C)$ konvex in $\mathbb{R}^m$ und für konvexes Menge $D \subseteq \mathbb{R}^m$ das Urbild $A^{-1}(D)$ konvex im $\mathbb{R}^n$.*

4. *Sind $C_1, C_2 \subseteq \mathbb{R}^n$ konvex, dann ist die Minkowski Summe $C_1 + C_2$ konvex; allgemeiner ist mit $\alpha_1, \alpha_2 \in \mathbb{R}$ auch $\alpha_1 C_1 + \alpha_2 C_2 = \{\alpha_1 x_1 + \alpha_2 x_2 : x_1 \in C_1, x_2 \in C_2\}$ konvex.*

5. *Ist $C \subseteq \mathbb{R}^n$ konvex, dann ist das Innere $intC$ und der Abschluss clC konvex.*

Beweis. in Übungen □

2.1.2 Konvexkombinationen und konvexe Hüllen

Aus der linearen Algebra ist bekannt:

Definition 2.1.2.

- *Für $x_1 \dots x_k \in \mathbb{R}^n$ und $\alpha_1 \dots \alpha_k \in \mathbb{R}$ heißt $\sum_{i=1}^k \alpha_i x_i$ Linearkombination der Elemente $x_i, i = 1 \dots k$.*
- *Ein (linearer) Unterraum enthält alle Linearkombinationen seiner Elemente.*
- *Der Schnitt linearer Unterräume ist wieder ein linearer Unterraum.*
- *Für beliebiges $S \subseteq \mathbb{R}^n$ ist $spanS$ / $linS$ der kleinste Unterraum, der S enthält, oder äquivalent: der Schnitt aller Unterräume, die S enthalten.*
- *Für $x_1 \dots x_k \in \mathbb{R}^n$ und $\alpha_1 \dots \alpha_k \in \mathbb{R}$ mit $\sum_{i=1}^k \alpha_i = 1$ heißt $\sum_{i=1}^k \alpha_i x_i$ Affinkombination der Elemente $x_i, i = 1 \dots k$.*
- *Eine affine Mannigfaltigkeit (affiner Unterraum) vom $\mathbb{R}^n$ enthält alle Affinkombinationen ihrer Elemente.*
- *Der Schnitt affiner Mannigfaltigkeiten ist wieder eine affine Mannigfaltigkeit.*
- *Für beliebiges $S \subseteq \mathbb{R}^n$ ist $affS$ die kleinste affine Mannigfaltigkeit, die S enthält, oder äquivalent: der Schnitt aller affinen Mannigfaltigkeiten, die S enthalten.*
- *$k+1$ Punkte $x_0, x_1, \dots, x_k$ heißen affin unabhängig, wenn die k Vektoren $\{x_1 - x_0, x_2 - x_0, \dots, x_k - x_0\}$ linear unabhängig sind.*
- *Sind $x_0, \dots, x_k$ affin unabhängig, dann ist jedes $x \in aff\{x_0, \dots, x_k\}$ eindeutig als Affinkombination durch $x = \sum_{i=0}^k \alpha_i x_i$ mit $\sum_{i=0}^k \alpha_i = 1$ darstellbar.*
- *Für $x_1, \dots, x_k \in \mathbb{R}^n, \alpha_1 \dots, \alpha_k \geq 0$ mit $\sum_{i=1}^k \alpha_i = 1$ (oder $\alpha \in \triangle_k$) heißt $\sum_{i=1}^k \alpha_i x_i$ Konvexkombination der x_i.*
- *Für beliebiges $S \subseteq \mathbb{R}^n$ ist die konvexe Hülle $convS = \bigcap \{S \subseteq C, C \text{ konvex}\} C$ der Schnitt aller konvexen Mengen C, die S enthalten.*

- *Für $x_1, \dots, x_k \in \mathbb{R}^n, \lambda_1, \dots, \lambda_k \geq 0$ heißt $\sum_{i=1}^k \lambda_i x_i$ konische Kombination der x_i.*
- *Für beliebiges $S \subseteq \mathbb{R}^n$ ist die konische Hülle*

$$coneS = \left\{ x : \exists k \in \mathbb{N}, (x_1, \dots, x_k) \in S, \lambda_1, \dots, \lambda_k \geq 0 : x = \textstyle\sum_{i=1}^k \lambda_i x_i \right\}.$$

Beobachtung 2.1.2. *1. Eine Menge $C \subseteq \mathbb{R}^n$ ist konvex $\iff$ C enthält jede Konvexkombination ihrer Elemente.*

2. Sei $S \subseteq \mathbb{R}^n$ beliebig. Dann gilt

$$convS = \left\{ x \in \mathbb{R}^n : \exists k \in \mathbb{N}, x_1, \dots, x_k \in S, \alpha \in \triangle_k : \textstyle\sum_{i=1}^k \alpha_i x_i = x \right\}$$

"Die konvexe Hülle ist die Menge der Konvexkombinationen von Elementen aus S."

3. Sei $S \subseteq \mathbb{R}^n$ beliebig. $coneS = \mathbb{R}_+(convS) = conv(\mathbb{R}_+ S)$

Beweis. 1., 2. in Übung

3. Für $\sum_{i=1}^k \lambda_i > 0, \lambda_i \geq 0$ ist $x = \underbrace{\sum_{i=1}^k \lambda_i s_i}_{\in coneS} = \underbrace{(\sum \lambda_j)}_{\in \mathbb{R}_+} \underbrace{\overbrace{\sum_{i=1}^k \frac{\lambda_i}{\sum \lambda_j}}^{=1} s_i}_{\in conv\, S} = \sum_{i=1}^k \underbrace{\frac{\lambda_i}{\sum \lambda_j}}_{\sum = 1} \underbrace{(\sum \lambda_j) s_i}_{\in \mathbb{R}_+ S}$.

(Für $\sum_{i=1}^k \lambda_i = 0$ ist 0-Punkt in allen drei Mengen.) □

Tatsächlich braucht man für $x \in convS$ bzw. $coneS$ nur "recht wenige" Elemente aus S. (Welche, ist von x abhängig!)

Satz 2.1.3 (Satz von Carathéodory, Kegelversion). *Sei $S \subseteq \mathbb{R}^n$. Jedes $x \in coneS$ kann als konische Kombination von linear unabhängigen Vektoren $s_i \in S$ dargestellt werden.*

Beweis. Sei $x = \sum_{i=1}^k \lambda_i s_i \in coneS$ mit $s_i \in S$ und $\lambda_i \geq 0, i = 1, \dots, k$. Falls alle s_i linear unabhängig, so sind wir fertig.
Sind die s_i linear abhängig, dann genügt es zu zeigen, dass wir ein s_i eliminieren können (notfalls wiederholen).

O.B.d.A. sei $\lambda_i > 0$, $\{s_i\}$ linear abhängig $\Rightarrow \exists \beta \in \mathbb{R}^k, \beta \neq 0 : \sum_{i=1}^k \beta_i s_i = 0$, o.B.d.A. ist ein $\beta_j < 0$ (verwende sonst $-\beta$). Bestimme $t > 0$, so dass $\lambda' = \lambda + t\beta \geq 0$ ist, und $\exists \bar{i}$ mit $\lambda'_{\bar{i}} = 0$.

$$\Rightarrow \sum_{i \in \{1,\dots,k\} \setminus \{\bar{i}\}} \lambda'_i s_i = \sum_{i=1}^k (\lambda_i + t\beta_i) s_i = \sum_{i=1}^k \lambda_i s_i + t \sum_{i=1}^k \beta_i s_i = x.$$

□

Jedes $x \in coneS$ kann als konische Kombination von n Elementen aus S (abhängig von x) dargestellt werden. Wie geht das bei $convS$?
Trick: *"Homogenisierung"*: Mache aus konvexer Menge in $\mathbb{R}^n$ einen konvexen Kegel im $\mathbb{R}^{n+1}$:

$$K = cone(C \times \{1\})$$

"Jede konvexe Menge ist als Schnitt eines konvexen Kegels mit einer Hyperebene darstellbar."

Satz 2.1.4 (Satz von Carathéodory, Originalversion). *Sei $S \subseteq \mathbb{R}^n$. Jedes $x \in convS$ kann als Konvexkombination von höchstens $n+1$ Elementen aus S dargestellt werden.*

Beweis. Sei $C = convS$, $K = cone(S \times \{1\})$.

$$\begin{aligned}\begin{pmatrix} x \\ 1 \end{pmatrix} \in K \quad &\Leftrightarrow \exists \lambda_1, \ldots, \lambda_k \in \mathbb{R}_+ : \begin{pmatrix} x \\ 1 \end{pmatrix} = \sum_{i=1}^{k} \lambda_i \begin{pmatrix} s_i \\ 1 \end{pmatrix} \\ &\Leftrightarrow x = \sum_{i=1}^{k} \lambda_i s_i \text{ mit } \lambda_i \geq 0 \text{ und } \sum_{i=1}^{k} \lambda_i = 1 \\ &\Leftrightarrow x \in convS\end{aligned}$$

Nach Satz 2.1.3 kann $k \leq n+1$ gewählt werden. □

2.1.3 Abschluss und relatives Inneres

Definition 2.1.3.

- *Für $S \subseteq \mathbb{R}^n$ ist die abgeschlossene konvexe Hülle $\overline{conv}S$ der Schnitt aller abgeschlossenen Mengen, die S enthalten.*
- *Für $S \subseteq \mathbb{R}^n$ ist die abgeschlossene konische Hülle $\overline{cone}S = cl(coneS)$ der Abschluss der konischen Hülle.*
- *Die Dimension $dimC$ einer konvexen Menge C ist die Dimension von $affC$.*
- *Das relative Innere riC einer konvexen Menge C ist die Menge der Punkte $x \in C$, für die bezüglich $affC$ auch eine Umgebung von x in C enthalten ist:*

  $$riC = \{x \in C : \exists \varepsilon > 0 : B_\varepsilon(x) \cap affC \subset C\}.$$

- *Der relative Rand $rbdC$ (relative boundary) einer konvexen Menge ist der Rand von C bzgl. $affC$: $rbdC := clC \setminus riC$.*

Satz 2.1.5. *Sei $S \subseteq \mathbb{R}^n$. Ist S beschränkt / kompakt, dann ist $convS$ beschränkt / kompakt.*

Beweis. Sei $x = \sum_{i=1}^{k} \alpha_i x_i \in convS, x_i \in S, \alpha \in \triangle_k$ mit $k \leq n+1$ nach Satz 2.1.4.
Sei S beschränkt durch $S \subseteq B_M(0)$.

$$\Rightarrow \|x\| \leq \sum_{i=1}^{k} \alpha_i \|x_i\| \leq M \sum_{i=1}^{k} \alpha_i \leq M,$$

also ist $convS$ beschränkt.
Sei nun S zusätzlich abgeschlossen.
z.z.: $convS$ ist abgeschlossen.
Sei $\{x^k = \sum_{i=1}^{n+1} \alpha_i^k x_i^k\} \in convS$ mit $x_i^k \in S, \alpha^k \in \triangle_{n+1}$ und $x^k \to x$. Da S und $\triangle_{n+1}$ kompakt, gibt es eine Teilfolge $K^0 \subseteq \mathbb{N}$ mit $\alpha \xrightarrow{K^0} \overline{\alpha} \in \triangle_{n+1}$ und für $i = 1, \ldots, n+1$ Teilfolgen $K^i \subseteq K^{i-1}$, so dass $x_i^k \xrightarrow{K^i} \overline{x_i} \in S$ also $x^k \longrightarrow^{K^{n+1}} x = \sum\limits_{i=1}^{n+1} \overline{\alpha_i} \cdot \overline{x_i} \in convS$ □

Beobachtung 2.1.6. *1. Sei $S \subseteq \mathbb{R}^n$. Es gilt $\overline{conv}S = cl(convS)$*

2. *Ist $S \subseteq \mathbb{R}^n$ beschränkt, so gilt $cl(convS) = conv(clS)$, also $\overline{conv}S = conv(clS)$.*

3. *Ist $S \subseteq \mathbb{R}^n$ kompakt und $0 \notin convS$, dann ist*

$$\overline{cone}S = \underbrace{\mathbb{R}_+ convS}_{=coneS}.$$

Beweis. 1. Übung

2. Wegen $S \subseteq convS$ ist auch $clS \subseteq cl(convS)$. Letztere Menge ist konvex nach Beobachtung 2.1.1 (5), also $conv(clS) \subseteq cl(convS) = \overline{conv}S$ nach (1). Nach Satz 2.1.5 ist $conv(clS)$ abgeschlossen, daraus folgt $\overline{conv}S \subseteq conv(clS)$ aus Definition 2.1.3 von $\overline{conv}S$.

3. Nach Satz 2.1.5 ist $C := convS$ kompakt und nach Voraussetzung ist $0 \notin C$.
Die Inklusion $\mathbb{R}_+ C \subseteq \overline{cone}S$ ist offensichtlich.
z.z.: $\mathbb{R}_+ C$ ist abgeschlossen.
Sei $\{\lambda_k x_k\} \in \mathbb{R}_+ C$ und $\lambda_k x_k \longrightarrow x$. Da C kompakt, existiert eine Teilfolge K, so dass $x_k \xrightarrow{K} \overline{x} \neq 0, \overline{x} \in C$. Wegen $\frac{\lambda_k x_k}{\|x_k\|} \xrightarrow{K} \frac{x}{\|\overline{x}\|}$ muss $\lambda_k \xrightarrow{K} \frac{\|x\|}{\|\overline{x}\|} =: \overline{\lambda} \geq 0$. Damit gilt $\lambda_k x_k \xrightarrow{K} \underbrace{\underbrace{\overline{\lambda}}_{\in \mathbb{R}_+ \cdot} \underbrace{\overline{x}}_{\in C}}_{\in \mathbb{R}_+ C}$.

□

Satz 2.1.7. *Sei $C \subseteq \mathbb{R}^n$ konvex. Falls $C \neq \varnothing$ ist $riC \neq \varnothing$ und* $\dim(riC) = \dim(C)$.

Beweis. Ist $dim(affC) = k$, dann enthält C $k+1$ affin unabhängige Elemente $x_0, \ldots, x_k$. Sei $\overline{x} = \sum_{i=0}^{k} \frac{1}{k+1} x_i$ das "*Zentrum*" des Simplex $\triangle = conv\{x_0, \ldots, x_k\} \subset C$. Die Spalten der Matrix $A = [x_1 - x_0, \ldots, x_k - x_0]$ sind linear unabhängig. $affC = \{\overline{x} + A\alpha : \alpha \in \mathbb{R}^k\}$ und für $y \in affC$ gibt es ein eindeutiges $\alpha(y) \in \mathbb{R}^k$ mit $y = \overline{x} + A\alpha(y)$. Wähle $\varepsilon > 0$, so dass

$\forall y \in B_\varepsilon(\overline{x}) \cap affC$ gilt $\|\alpha(y)\| \leq \frac{1}{(k+1)^2}$. Dann ist

$$y = \underbrace{\left(\frac{1}{k+1} - \sum_{i=1}^{k} \alpha_i(y)\right)}_{>0} x_0 + \sum_{i=1}^{k} \underbrace{\left(\frac{1}{k+1} + \alpha_i(y)\right)}_{>0} x_i \in \triangle \subseteq C$$
$$\Rightarrow \overline{x} \in ri\triangle \subseteq riC \text{und} \dim(riC) \geq \dim \triangle = \dim(affC).$$

□

Lemma 2.1.8. *Sei $C \subseteq \mathbb{R}^n$ konvex, $x \in clC$ und $\in riC$. Dann gilt:*

$$(x, y] = \{\alpha x + (1-\alpha)y : 0 \leq \alpha < 1\} \subseteq riC.$$

Beweis. in Übung □

Beobachtung 2.1.9. *Sei $C \subseteq \mathbb{R}^n$ konvex. Die drei Mengen riC, C und clC haben dieselbe affine Hülle (also gleiche Dimension), dasselbe relative Innere und denselben Abschluss, also auch denselben relativen Rand.*

Beweis. Dieselbe affine Hülle folgt aus Satz 2.1.7 (auch für clC, weil $affC$ abgeschlossen).
Der Rest geht mit Lemma 2.1.8.
Zum Beispiel: riC und C haben den gleichen Abschluss, also noch zu zeigen: $clC \subseteq cl(riC)$
Sei $x \in clC$ und $y \in riC$ (gibt es nach Satz 2.1.7). Dann ist $x_k = (1 - \frac{1}{k})x + \frac{1}{k}y \in riC$ nach Lemma 2.1.8 und mit $\lim_{k\to\infty} x_k = x \in cl(riC)$. □

Wie überträgt sich relatives Inneres und Abschluss bei den konvexitätserhaltenden Operationen?

Beobachtung 2.1.10. *Seien $C_1, C_2 \subseteq \mathbb{R}^n$ konvexe Mengen mit $riC_1 \cap riC_2 \neq \varnothing$. Dann gilt:*

$$ri(C_1 \cap C_2) = riC_1 \cap riC_2 \quad \textit{und} \quad cl(C_1 \cap C_2) = clC_1 \cap clC_2.$$

Beweis. zuerst cl: Wegen $riC_1 \cap riC_2 \subseteq C_1 \cap C_2 \subseteq clC_1 \cap clC_2$ gilt:

$$cl(riC_1 \cap riC_2) \subseteq cl(C_1 \cap C_2) \subseteq clC_1 \cap clC_2.$$

Noch zu zeigen: $clC_1 \cap clC_2 \subseteq cl(C_1 \cap C_2)$
Sei $x \in clC_1 \cap clC_2$. Wähle $y \in riC_1 \cap riC_2$; mit Lemma 2.1.8 folgt:

$$(x, y] \subseteq riC_1 \cap riC_2 \subseteq C_1 \cap C_2 \Rightarrow x \in cl(riC_1 \cap riC_2) \subseteq cl(C_1 \cap C_2).$$

Wegen Beobachtung 2.1.9 folgt aus der Gleichheit von cl, dass

$$ri(C_1 \cap C_2) = ri(riC_1 \cap riC_2) \subseteq riC_1 \cap riC_2.$$

Noch zu zeigen: $riC_1 \cap riC_2 \subseteq ri(C_1 \cap C_2)$.
Sei $x \in riC_1 \cap riC_2$. Wählen $y \in ri(C_1 \cap C_2)$ welches es nach Satz 2.1.7 gibt, da $\varnothing \neq riC_1 \cap riC_2 \subseteq C_1 \cap C_2$.
Falls $x = y$, fertig. Sei also $x \neq y$.
Wegen $y \in C_1$ und $x \in riC_1$ existiert $\varepsilon_1 > 0 : [y, y + (1 + \varepsilon_1)(x - y)] \subseteq C_1$ ("stretching");
wegen $x \in C_2$ und $x \in riC_2$ existiert $\varepsilon_2 > 0 : [y, y + (1 + \varepsilon_2)(x - y)] \subseteq C_2$.

$$\Rightarrow \overline{x} := y + (1 + \min\{\varepsilon_1, \varepsilon_2\})(x - y) \in C_1 \cap C_2 \quad \text{und} \quad x \in (\overline{x}, y).$$

Somit folgt nach Lemma 2.1.8

$$\overline{x} \in C_1 \cap C_2, y \in ri(C_1 \cap C_2) \Rightarrow (\overline{x}, y] \subseteq ri(C_1 \cap C_2),$$

also $x \in (\overline{x}, y) \subseteq ri(C_1 \cap C_2)$. □

Beobachtung 2.1.11. *Seien $C_i \subseteq \mathbb{R}^{n_i}$ konvex für $i = 1, \ldots, k$. Dann ist*

$$ri(C_1 \times \ldots \times C_k) = riC_1 \times \ldots \times riC_k,$$
$$cl(C_1 \times \ldots \times C_k) = clC_1 \times \ldots \times C_k.$$

Beweis. Nutze $aff(C_1 \times \ldots \times C_k) = affC_1 \times \ldots \times affC_k$ und die Definition. □

Beobachtung 2.1.12. *Sei $A : \mathbb{R}^n \to \mathbb{R}^m$ eine affine Abbildung.*

1. *Ist $C \subseteq \mathbb{R}^n$ konvex, dann ist $ri(A(C)) = A(riC)$.*
2. *Ist $D \subseteq \mathbb{R}^m$ konvex mit $A^{-1}(riD) \neq \varnothing$ (Urbild), dann ist $ri(A^{-1}(D)) = A^{-1}(riD)$.*

Beweis. 1. Da A stetig (folgt aus A affin), gilt für beliebige $S \subseteq \mathbb{R}^n$: $A(clS) \subseteq cl(A(S))$ und es gilt nach Beobachtung 2.1.9:

$$A(C) \subseteq A(clC) = A(cl(riC)) \subseteq cl[A(riC)] \subseteq cl[A(C)].$$

Also ist $cl[A(riC)] = cl[A(C)]$. Nach Beobachtung 2.1.1 (3) und Beobachtung 2.1.9 folgt

$$ri(A(C)) = ri[A(riC)] \subseteq A(riC).$$

Wir zeigen nun $A(riC) \subseteq riA(C)$.
Sei $\overline{w} = A(\overline{x}) \in A(riC)$ mit $\overline{x} \in riC$. Sei $v \in A(y) \in riA(C)$ mit $y \in C$. Falls $v = \overline{w}$ fertig. Sei also $v \neq \overline{w}$ und damit $\overline{x} \neq y$. Nach dem Dehnungsargument existiert wegen $\overline{x} \in riC, y \in C$ ein $x \in C$, so dass $\overline{x} \in (y, x) \subset C$ und wegen $v \neq \overline{w}$ ist $A(\overline{x}) \in (A(x), A(y))$. Nach Lemma 2.1.8 ist $A(\overline{x}) \in riA(C)$.
2. mit gleicher Technik □

Bemerkung 2.1.13.
1. *In Beobachtung 2.1.12 (2) braucht man $A^{-1}(riD) \neq \varnothing$:*
   Sei $A = 0^{m \times n}$, $D = \mathbb{R}^m_+ \Rightarrow ri(A^{-1}(D)) = \mathbb{R}^n$, aber $A^{-1}(riD) = \varnothing$, da $0 \notin ri\mathbb{R}^m_+$.
2. *Es gilt i. A. nicht, dass $cl(A(C)) = A(clC)$:*
   Sei $A = [0\ 1], C = \{(x, y) : y \geq \frac{1}{x}, x > 0\}$
3. *Wegen $ri(\alpha_1 C_1 + \alpha_2 C_2) = \alpha_1 riC_1 + \alpha_2 riC_2$ folgt die wichtige Charakterisierung:*

$$0 \in ri(C_1 \cap C_2) \Longleftrightarrow riC_1 \cap riC_2 \neq \varnothing.$$

2.1.4 Projektion auf abgeschlossene konvexe Mengen

Definition 2.1.4. *Sei $\varnothing \neq C \subseteq \mathbb{R}^n$ abgeschlossen, konvex und $x \in \mathbb{R}^n$. Dann heißt*

$$p_C(x) = \arg\min_{y \in C} \frac{1}{2}\|y - x\|^2$$

die Projektion von x auf C. Dabei bezeichnet $\arg\min$ *das eindeutige minimierende Element.*

Bemerkung 2.1.14. *p_C ist wohldefiniert, da $f(y) = \frac{1}{2}\|y - x\|^2$ streng konvex ist und daher die Optimallösung eindeutig ist. Die Existenz folgt aus der Stetigkeit von f und der Wahl einer kompakten Kugel um x, die C schneidet. (Radius ist $r = \|x - y\|$ für ein $y \in C$.)*

Satz 2.1.15. *Sei $\varnothing \neq C \subseteq \mathbb{R}^n$ abgeschlossen, konvex und $\overline{y} \in C, x \in \mathbb{R}^n$.*

$$\overline{y} = p_C(x) \Longleftrightarrow \langle x - \overline{y}, y - \overline{y}\rangle \leq 0 \quad \forall y \in C.$$

Beweis. "⇒": Sei $\overline{y} = p_C(x)$, dann ist für $y \in C$ und $\alpha > 0$

$$\tfrac{1}{2}\|x - \overline{y}\|^2 \leq \tfrac{1}{2}\|x - (\overline{y} + \alpha(y - \overline{y}))\|^2 \leq \tfrac{1}{2}\|x - \overline{y}\|^2 - \alpha\langle x - \overline{y}, y - \overline{y}\rangle + \alpha^2\tfrac{1}{2}\|y - \overline{y}\|^2,$$

also ist $0 \leq -\alpha\langle x - \overline{y}, y - \overline{y}\rangle + \alpha^2\frac{1}{2}\|y - \overline{y}\|^2$. Für $\alpha \to 0$ geht α^2 schneller gegen 0 als α $\Rightarrow \langle x - \overline{y}, y - \overline{y}\rangle \leq 0$.
"⇐": Erfüllt $\overline{y}$ die Relation $\langle x - \overline{y}, y - \overline{y}\rangle \leq 0\ \forall y \in C$, dann folgt

$$0 \geq \langle x - \overline{y}, y - x + x - \overline{y}\rangle = \|x - \overline{y}\|^2 + \langle x - \overline{y}, y - \overline{y}\rangle \geq \|x - \overline{y}\|^2 - \|x - \overline{y}\| \cdot \|y - x\|.$$

Division durch $\|x - \overline{y}\|$ zeigt, dass $\|x - \overline{y}\| \leq \|y - x\|\ \forall y \in C$. □

Beobachtung 2.1.16. *Sei $\varnothing \neq C \subseteq \mathbb{R}^n$ abgeschlossen und konvex.*
Dann gilt $\forall (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n : \|p_C(x_1) - p_C(x_2)\|^2 \leq \langle p_C(x_1) - p_C(x_2), x_1 - x_2\rangle$.

Beweis. Nach Satz 2.1.15 mit $x = x_1, \overline{y} = p_C(x_1), y = p_C(x_2)$

$$\Rightarrow \langle p_C(x_2) - p_C(x_1), x_1 - p_C(x_1)\rangle \leq 0.$$

Nach Satz 2.1.15 mit $x = x_2, \overline{y} = p_C(x_2), y = p_C(x_1)$

$$\Rightarrow \langle p_C(x_1) - p_C(x_2), x_2 - p_C(x_2)\rangle \leq 0.$$

Addition beider Ungleichungen liefert

$$\langle p_C(x_1) - p_C(x_2), x_2 - x_1 + p_C(x_1) - p_C(x_2)\rangle \leq 0.$$

□

Bemerkung 2.1.17. *Mit der Cauchy-Schwarz-Ungleichung folgt aus obiger Beobachtung, dass* $\|p_C(x_1) - p_C(x_2)\| \leq \|x_1 - x_2\|$*, d.h. die Projektion ist eine kontrahierende Abbildung. ("Der Abstand dehnt sich durch Projektion nicht aus.")*

Definition 2.1.5. *Sei* $K \subseteq \mathbb{R}^n$ *ein konvexer Kegel (nicht notwendigerweise abgeschlossen). Der polare Kegel (polar cone) ist*

$$K^\circ = \{s \in \mathbb{R}^n : \langle s, x\rangle \leq 0 \; \forall x \in K\}.$$

"Menge aller Normalvektoren auf Hyperebenen, die K *im negativen Halbraum enthalten."*

K° ist abgeschlossen: Sei $\{s_k\} \to s$ mit $s_k \in K^\circ$; dann ist für $x \in K$ beliebig

$$\langle s, x\rangle = \lim\langle s_k, x\rangle \leq \sup\langle s_k, x\rangle \leq 0 \; \forall x, \text{ also } s \in K^\circ.$$

Satz 2.1.18. *Sei* K *ein konvexer abgeschlossener Kegel.*

$$\overline{y} = p_K(x) \Leftrightarrow \overline{y} \in K, x - \overline{y} \in K^\circ, \langle x - \overline{y}, \overline{y}\rangle = 0.$$

Beweis. "⇒" Nach Satz 2.1.15 $\Rightarrow \langle x - \overline{y}, y - \overline{y}\rangle \leq 0 \; \forall y \in K$. Setze $y = \alpha\overline{y}$. Dann folgt für beliebiges $\alpha \geq 0$

$$(\alpha - 1)\langle x - \overline{y}, \overline{y}\rangle \leq 0 \; \forall y \in K.$$

Da $(\alpha - 1)$ sowohl positiv als auch negativ sein kann, folgt $\langle x - \overline{y}, \overline{y}\rangle = 0$ und damit $\langle x - \overline{y}, y\rangle \leq 0 \; \forall y \in K$, also nach Definition $x - \overline{y} \in K^\circ$.
"⇐" Sei $\overline{y}$ wie beschrieben, $y \in K$ beliebig.

$$\begin{aligned}
\tfrac{1}{2}\|x - y\|^2 &= \tfrac{1}{2}\|x - \overline{y} + \overline{y} - y\|^2 = \tfrac{1}{2}\|x - \overline{y}\|^2 + \langle x - \overline{y}, \overline{y} - y\rangle + \tfrac{1}{2}\|\overline{y} - y\|^2 \\
&= \tfrac{1}{2}\|x - \overline{y}\|^2 - \langle x - \overline{y}, y\rangle + \tfrac{1}{2}\|\overline{y} - y\|^2 \\
&\geq \tfrac{1}{2}\|x - \overline{y}\|^2 - \underbrace{\langle x - \overline{y}, y\rangle}_{\leq 0 \text{ wegen } (x-\overline{y}) \in K^\circ} \\
&\geq \tfrac{1}{2}\|x - \overline{y}\|^2.
\end{aligned}$$

Nach Definition ist damit $\Rightarrow \overline{y} = p_K(x)$. □

2.1.5 Separierung konvexer Mengen

Satz 2.1.19. *Sei* $C \subseteq \mathbb{R}^n$*,* C *konvex, abgeschlossen und sei* $x \notin C$*. Dann gibt es ein* $s \in \mathbb{R}^n$ *mit*

$$\langle s, x\rangle > \sup_{y \in C}\langle s, y\rangle.$$

Beweis. Setze $s := x - p_C(x) \neq 0$, da $x \notin C$. Nach Satz 2.1.15 ist

$$\langle 0 \geq s, y - \underbrace{(x-s)}_{p_C(c)} \rangle = \langle s, y \rangle - \langle s, x \rangle + \|s\|^2, \forall y \in C$$

und daher ist

$$\langle s, x \rangle - \|s\|^2 \geq \langle s, y \rangle \; \forall y \in C.$$

□

Korollar 2.1.20 (starke Trennung konvexer Mengen). *Seien C_1, C_2 zwei nichtleere konvexe abgeschlossene Mengen mit $C_1 \cap C_2 = \varnothing$. Ist C_2 beschränkt (und damit kompakt), dann gibt es ein $s \in \mathbb{R}^n$, so dass*

$$\sup_{y_1 \in C_1} \langle s, y_1 \rangle < \min_{y_2 \in C_2} \langle s, y_2 \rangle.$$

Beweis. $C_1 - C_2$ ist konvex und abgeschlossen (weil C_2 kompakt) und $0 \notin C_1 - C_2$ (weil $C_1 \cap C_2 = \varnothing$), nach Satz 2.1.19 gibt es ein $s \in \mathbb{R}^n$ mit

$$\sup\{\langle s, y \rangle : y \in C_1 - C_2\} < \langle s, 0 \rangle = 0,$$

also

$$0 > \sup_{y_1 \in C_1} \langle s, y_1 \rangle + \sup_{y_2 \in C_2} \langle s, -y_2 \rangle = \sup_{y_1 \in C_1} \langle s, y_1 \rangle - \inf_{y_2 \in C_2} \langle s, y_2 \rangle.$$

Wegen C_2 kompakt, wird das $\inf$ angenommen, also

$$\sup_{y_1 \in C_1} \langle s, y_1 \rangle < \min_{y_2 \in C_2} \langle s, y_2 \rangle.$$

□

Das Korollar hilft nachzuweisen, dass konvexe Mengen Stützhyperebenen erlauben.

Definition 2.1.6. *Eine Hyperebene $H_{s,r}$ stützt eine Menge C, wenn C ganz in einem der beiden abgeschlossenen Halbräume enthalten ist.*
Sie stützt C in $x(\in C)$, falls zusätzlich $\langle s, x \rangle = r$ gilt.

Lemma 2.1.21. *Sei $x \in bdC$ mit $\varnothing \neq C \in \mathbb{R}^n$ konvex. Dann gibt es eine Hyperebene, die C in x stützt.*

Beweis. C, clC sowie deren Komplemente haben denselben Rand (nach Beobachtung 2.1.9) $\Rightarrow$ $\exists\{x_k\}_{k\in\mathbb{N}}$ mit $x_k \notin clC$ und $\lim_{k\to\infty} x_k = x$. Für jedes k gibt es nach Satz 2.1.18 ein s_k mit $\|s_k\| = 1$, so dass

$$\langle s_k, x_k - y \rangle > 0 \; \forall y \in C \subseteq clC.$$

Da die s_k auf einer kompakten Kugel liegen, existiert eine Teilfolge K und ein s, so dass $\{s_k\} \xrightarrow{K} s$ und damit $\langle s, x_k - y \rangle \geq 0 \; \forall y \in C$. Die Stützhyperebene ist also $H_{s,r}$: $\langle s, x \rangle = r \geq \langle s, y \rangle \; \forall y \in C$. □

Falls C "flach" ist, kann es vorkommen, dass $C \subset H_{s,r} = \{x : \langle s, x\rangle = r\}$. Vermeidbar, falls $x_0 \in rbdC$, indem die Konstruktion in $affC$ ausgeführt wird.

Satz 2.1.22. *Sei $\varnothing \neq C \subsetneq \mathbb{R}^n$ konvex. Dann ist der Schnitt aller Halbräume, die C enthalten, der Abschluss von C.*

Beweis. Es gilt:

$$clC \subset C^\star := \{x \in \mathbb{R}^n : \langle s, x\rangle \leq r \text{ mit } C \subseteq \{y : \langle s, y\rangle \leq r\}\},$$

denn falls C im abgeschlossenen Halbraum ist, liegt auch clC im abgeschlossenen Halbraum. Sei nun $x \notin clC$. Separiere x von clC (Satz 2.1.19) durch $s_0 \neq 0$, so dass

$$\langle s_0, x\rangle > \sup_{y \in clC} \langle s_0, y\rangle =: r_0,$$

also $x \notin C^\star$. □

Korollar 2.1.23. *Jede abgeschlossene, konvexe Menge C ist der Schnitt aller Halbräume, die C enthalten.*

Klappt auch für $C = \mathbb{R}^n, C = \varnothing$.
Spezialfall: Polyeder.

Definition 2.1.7. *Ein (abgeschlossenes, konvexes) Polyeder ist der Schnitt **endlich** vieler Halbräume*

$$P = \{x \in \mathbb{R}^n : \langle s_j, x\rangle \leq r_j, j = 1, \ldots, m\} = \{x : Ax \leq b\}$$

Sind alle $r_j = 0$ ($b = 0$), dann ist P ein abgeschlossener, konvexer, polyedrischer Kegel.

Satz 2.1.24. *Sei K ein konvexer Kegel mit Polarkegel K°, dann ist der bipolare Kegel $(K^\circ)^\circ$ der Abschluss von K.*

Beweis. Nach Definition ist $K^\circ = \{s \in \mathbb{R}^n : \langle s, x\rangle \leq 0 \; \forall x \in K\}$. Da K ein Kegel ist, ist für $s \in \mathbb{R}^n$

$$\sup_{x \in K} \langle s, y\rangle = \begin{cases} 0 & \text{falls } s \in K^\circ \text{ (betrachte Folge } x_k \to 0) \\ \infty & \text{sonst (für } \langle s, x\rangle > 0 \text{ betrachte } \lambda x \text{ mit } \lambda \to \infty), \end{cases}$$

also beschreibt K° die Menge aller Halbräume, die K engstmöglich umfassen. Nach Satz 2.1.22 ist der Schnitt aller dieser Halbräume der Abschluss von K, also

$$clK = \bigcap_{s \in K^\circ} \{x \in \mathbb{R}^n : \langle s, x\rangle \leq 0\} = \{x \in \mathbb{R}^n : \langle s, x\rangle \leq 0 \; \forall s \in K^\circ\} = (K^\circ)^\circ.$$

□

Für die Optimierung über polyedrische Mengen (z.B. bei LP's) ist von Bedeutung, ob $\mathcal{X} = \{x \geq 0 : Ax = b\}$ eine Lösung hat oder ob nachweisbar ist, dass $\mathcal{X} = \varnothing$ gilt.
Aufgefasst als Frage "$b \in \{\sum A_{\cdot,i} x_i : x_i \geq 0\}$" liest sich das als: Ist b im Kegel, der von den Spalten von A aufgespannt wird? Antwort darauf gibt das Lemma von Farkas.

Lemma 2.1.25 (Lemma von Farkas, Version 1). *Sei $b \in \mathbb{R}^n$ und $A \in \mathbb{R}^{n \times m}$.*
Dann hat genau eines der Systeme

$$Ax = b, x \geq 0 \qquad \textit{und} \qquad A^\tau y \leq 0, b^\tau y > 0$$

eine Lösung.
"Entweder liegt b im Kegel, der von den Spalten von A aufgespannt wird, oder es gibt eine Hyperebene mit Normalenvektor y, die b vom Kegel trennt."

Falls b nicht im Kegel liegt, muss der Kegel nur "abgeschlossen" sein, damit das Lemma aus dem Trennungssatz 2.1.19 folgt.
Das Farkas Lemma wird also durch den Beweis folgender Umformulierung bewiesen.

Lemma 2.1.26 (Lemma von Farkas, Version 2). *Seien $a_1, \ldots, a_m$ aus dem $\mathbb{R}^n$ gegeben. Der konvexe Kegel*

$$K = cone\{a_1, \ldots, a_m\} = \left\{\sum \lambda_i a_i : \lambda_i \geq 0, i = 1, \ldots, m\right\}$$

ist abgeschlossen.

Beweis. Sei $\{b^k\} \to b$ mit $b^k \in K$, also $b^k = \sum \lambda_i^k a_i : \lambda_i^k \geq 0 \; \forall i, k$. Sind die a_i linear unabhängig $\Rightarrow \lambda_i^k \to \lambda_i \geq 0$ mit $b = \sum a_i \lambda_i \in K$, also abgeschlossen.
Falls die a_i linear abhängig sind, dann können λ_i^k "ganz beliebig" aussehen; müssen die lineare Abhängigkeit loswerden. Nach Satz 2.1.3 kann jedes $x \in K$ als konische Kombination von linear unabhängigen Vektoren a_i dargestellt werden. Es gibt nur endlich viele linear unabhängige Teilmengen von $\{a_1, \ldots, a_m\}$, also endlich viele abgeschlossene Teilkegel. In einem davon müssen unendlich viele b^k liegen und nach dem ersten Teil daher auch b. □

2.1.6 Seitenflächen und Extrempunkte

Definition 2.1.8.

- *Seien $F, C \subseteq \mathbb{R}^n$ konvex mit $F \subseteq C$. Falls aus $x \in F$ und $\exists x_1, x_2 \in C$ mit $x \in (x_1, x_2)$ folgt, dass $x_1, x_2 \in F$, dann heißt F Seitenfläche von C.*
  In Worten: "Enthält eine Strecke aus C einen inneren Punkt aus F, dann ist die ganze Strecke in F enthalten."
- *$\varnothing$ und C nennt man triviale Seitenflächen von C.*
- *Eine Seitenfläche F von C mit $\dim(F) = \dim C - 1$ heißt Facette.*

- *Eine Seitenfläche F von C mit $\dim(F) = 1$ heißt Kante (edge).*
- *Eine Seitenfläche F von C mit $\dim(F) = 0$ heißt Extrempunkt.*
- *Eine Teilmenge $F \subseteq C$ (C konvex) heißt exponierte Seitenfläche, falls es eine Stützhyperebene $H_{s,r}$ gibt, mit $F = H_{s,r} \cap C$.*
- *Ein exponierter Extrempunkt heißt Ecke (vertex).*

Beobachtung 2.1.27. *Eine exponierte Seitenfläche ist eine Seitenfläche.*

Beweis. F sei eine exponierte Seitenfläche zur Stützhyperebene $H_{s,r}$ an C, also $F = H_{s,r} \cap C$. Seien $x_1, x_2 \in C$ mit $\langle s, x_1\rangle \leq r, \langle s, x_2\rangle \leq r$ und $\overline{x} = \alpha x_1 + (1-\alpha)x_2 \in F$ für ein $\alpha \in (0,1)$. Also ist

$$\begin{aligned} r &\overset{x\in F = H_{s,r}\cap C}{=} \langle s, \overline{x}\rangle = \langle s, \alpha x_1\rangle + \langle s, (1-\alpha)x_2\rangle \leq \alpha r + (1-\alpha)r = r \\ &\Rightarrow \langle s, x_1\rangle = r, \langle s, x_2\rangle = r \\ &\Rightarrow x_1, x_2 \in F. \end{aligned}$$

□

Beobachtung 2.1.28. *Ist $\varnothing \neq C \subseteq \mathbb{R}^n$ konvex und kompakt, dann hat C Extrempunkte.*

Beweis. Da C kompakt, wird für die stetige Funktion $\|x\|^2$ das Maximum über C angenommen; sei $\overline{x} \in C$ ein Punkt, der dieses Maximum realisiert. Falls $\{\overline{x}\}$ kein Extrempunkt, dann gibt es

$$x_1, x_2 \in C : \overline{x} = \tfrac{1}{2}(x_1 + x_2)$$

(o.B.d.A. ist $\alpha = \frac{1}{2}$). Dann ist

$$\|\overline{x}\|^2 = \|\tfrac{1}{2}(x_1 + x_2)\|^2 < \tfrac{1}{2}(\|x_1\|^2 + \|x_2\|^2) \leq \tfrac{1}{2}(\|\overline{x}\|^2 + \|\overline{x}\|^2) = \|\overline{x}\|^2$$

$\Rightarrow$ Widerspruch. □

Beobachtung 2.1.29. *Sei F eine Seitenfläche von C, C konvex. Dann ist jeder Extrempunkt von F auch ein Extrempunkt von C.*

Beweis. Falls $x \in F$ und kein Extrempunkt von C, dann gibt es offensichtlich $x_1, x_2 \in C$ mit $x \in (x_1, x_2) \Rightarrow x_1, x_2 \in F \Rightarrow x$ ist kein Extrempunkt von F. □

Satz 2.1.30 (Satz von Minkowski). *Sei $C \subseteq \mathbb{R}^n$ konvex und kompakt. Dann ist C die konvexe Hülle ihrer Extrempunkte.*

Beweis. Induktion über $\dim C$.
Wahr für $C = \varnothing$ bzw. $\dim C = 0$.
Sei $x \in C$, zu zeigen: x ist Konvexkombination der Extrempunkte von C.

1. $x \in rbdC$: Nach Lemma 2.1.21 und Bemerkung gibt es eine Hyperebene, die C in x stützt, so dass $\dim(C \cap H) \leq \dim C - 1$, also ist x nach Induktionsvoraussetzung darstellbar als eine Konvexkombination der Extrempunkte von $F = C \cap H$. F ist Seitenfläche nach Beobachtung 2.1.27, deren Extrempunkte sind auch Extrempunkte von C nach Beobachtung 2.1.29.

2. $x \in riC$: Wähle Gerade in $affC$ durch x; diese schneidet $rbdC$ in zwei Punkten $\overline{x}$ und $\overline{\overline{x}}$ (da C kompakt). Diese sind als Konvexkombination darstellbar nach Teil 1. $\Rightarrow x$ ist als Konvexkombination von $\overline{x}$ und $\overline{\overline{x}}$ ebenfalls als Konvexkombination von Extrempunkten von C darstellbar.

□

Beobachtung 2.1.31. *Sei $C \subseteq \mathbb{R}^n$ konvex und kompakt, $s \in \mathbb{R}^n$. Dann gilt:*

$$\max_{x \in C} \langle s, x \rangle = \max\{\langle s, x \rangle : x \textit{ Extrempunkt von } C\}$$
$$\arg\max_{x \in C} \langle s, x \rangle = conv \arg\max\{\langle s, x \rangle : x \textit{ Extrempunkt von } C\},$$

wobei $\arg\max$ *die Menge der Optimalpunkte bezeichnet.*

Beweis. Da C kompakt, nimmt $\langle s, . \rangle$ ihr Maximum $r^\star$ auf C an $\Rightarrow F = C \cap H_{s,r^\star} \neq \varnothing$ ist exponierte Seitenfläche, da $H_{s,r^\star}$ Stützhyperebene. F ist die Menge der Optimallösungen und als konvexe, kompakte Menge nach Satz 2.1.30 die konvexe Hülle ihrer Extrempunkte, die nach Beobachtung 2.1.29 auch die Extrempunkte von C sind. □

"Spezialfall" Hauptsatz der linearen Optimierung: Basislösungen sind Extrempunkte der zulässigen Menge.

2.1.7 Tangenten- und Normalenkegel

In der glatten nichtlinearen Optimierung werden die Funktionen durch Tangentialebenen und Gradienten approximiert. In der konvexen Welt braucht man für die "Knickstellen" Kegel.

Definition 2.1.9. *Sei $\varnothing \neq S \subseteq \mathbb{R}^n$. Wir nennen eine Richtung $d \in \mathbb{R}^n$ tangential zu S in einem Punkt $x \in S$, wenn es eine Folge von Punkten $\{x_k\} \in S$ mit $x_k \to x$ und $\{t_k\} \downarrow 0$ gibt mit*

$$\lim_{k \to \infty} \frac{x_k - x}{t_k} \to d.$$

Die Menge aller solcher Richtungen nennen wir den Tangentialkegel zu S in x und schreiben dafür $T_S(x)$. Stets ist $0 \in T_S(x)$.

Setzt man $d_k = \frac{x_k - x}{t_k} \to d$ erhält man:

Beobachtung 2.1.32.

$$d \in T_S(x) \Leftrightarrow \exists \{d_k\} \to d, \{t_k\} \downarrow 0 : x + t_k d_k \in S \; \forall k.$$

Beobachtung 2.1.33. *Der Tangentenkegel ist abgeschlossen.*

Beweis. Diagonalisierungsargument $\{d_l\} \to d$ mit $d_l \in T_S(x)$. Seien $\{x_{k,l}\}, \{t_{k,l}\}$ die d_l definierenden Folgen. Für jedes l bestimme k_l, so dass

$$\left\| \frac{x_{k_l,l}-x}{t_{k_l,l}} - d_l \right\| \leq \tfrac{1}{2}.$$

$\Rightarrow$ Folge $\{x_{k_l,l}\}, \{t_{k_l,l}\}$ erfüllt $\lim_{l\to\infty} \frac{x_{k_l,l}-x}{t_{k_l,l}} \to d \in T_S(x)$. □

Beobachtung 2.1.34. *Sei C abgeschlossen und konvex, $x \in C$. Der Tangentenkegel ist der Abschluss des durch $C - \{x\}$ erzeugten Kegels:*

$$T_C(x) = \overline{cone}(C - \{x\}) = cl\mathbb{R}_+(C - \{x\}) = cl\{d \in \mathbb{R}^n : d = \alpha(y - x), y \in C, \alpha \geq 0\}.$$

Beweis. $C - \{x\} \subseteq T_C(x)$, weil

$$x + td \in C \; \forall d \in C - \{x\} \quad \forall t \in [0, 1].$$

Da $T_C(x)$ abgeschlossen (Beobachtung 2.1.33), ist also auch $cl\{\mathbb{R}_+(C - \{x\})\} \subseteq T_C(x)$.
Sei nun $d \in T_C(x)$ mit $\{x_k\}, \{t_k\}$ nach Definition, dann ist offensichtlich

$$\tfrac{x_k - x}{t_k} \in \mathbb{R}_+(C - \{x\}),$$

also ist der Limes d auch im Abschluss $cl(\mathbb{R}_+(C - \{x\}))$. □

Definition 2.1.10. *Eine Richtung $s \in \mathbb{R}^n$ heißt normal zu C in x, falls $\langle s, y - x\rangle \leq 0 \; \forall y \in C$. Die Menge aller solcher Richtungen wird Normalenkegel zu C in x genannt und mit $N_C(x)$ bezeichnet.*

Beobachtung 2.1.35. *Der Normalenkegel ist polar zum Tangentenkegel.*

Beweis. Sei $s \in N_C(x), d := (y - x) \in C - \{x\}$. Es gilt

$$\langle s, d\rangle \leq 0 \; \forall d \in C - \{x\} \Rightarrow \langle s, d\rangle \leq 0 \; \forall \mathbb{R}_+(C - \{x\})$$

und wegen der Stetigkeit gilt das auch $\forall d \in cl(\mathbb{R}_+(C - \{x\})) = T_C(x)$. Also ist $N_C(x) \subseteq [T_C(x)]^\circ$.

Sei $s \in [T_C(x)]^\circ$. $\Rightarrow \langle s, d\rangle \leq 0 \; \forall d \in T_C(x) = cl(\mathbb{R}_+(C - \{x\})) \Rightarrow \langle s, d\rangle \leq 0 \; \forall d \in C - \{x\}$. □

Da für konvexe Kegel C die Gleichheit $(C^\circ)^\circ = cl(C)$ gilt (Satz 2.1.24) und der Tangentenkegel bereits abgeschlossen ist (Beobachtung 2.1.33), folgt jetzt sofort:

Korollar 2.1.36. *Der Tangentenkegel ist der Polarkegel des Normalenkegels.*

2.2 Konvexe Funktionen

Definition 2.2.1.
- *Die Menge $Conv\mathbb{R}^n$ bezeichnet die Menge der eigentlich (proper) konvexen Funktionen $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ (mit $f \not\equiv +\infty$).*
- *$\overline{Conv}\mathbb{R}^n$ bezeichnet die abgeschlossenen Funktionen in $Conv\mathbb{R}^n$.*
- *$domf = \{x \in \mathbb{R}^n : f(x) < \infty\}$ (domain) bezeichnet den eigentlichen Definitionsbereich von f. (Er ist für $f \in Conv\mathbb{R}^n$ nichtleer.)*

Satz 2.2.1 (Jensensche Ungleichung). *Sei $f \in Conv\mathbb{R}^n$.*
Für $k \in \mathbb{N}, x_i \in domf, i = 1, \ldots, k, \alpha_i \geq 0, \sum \alpha_i = 1$ gilt:

$$f\left(\sum_{i=1}^{k} \alpha_i x_i\right) \leq \sum_{i=1}^{k} \alpha_i f(x_i).$$

Beweis. Spalte erst $\alpha_n x_n$ ab mit $\overline{x} = \sum \frac{\alpha_i}{1-\alpha_n} x_i$ usw. □

Besondere Bedeutung haben lineare (affine) Funktionen $f(x) = \langle s, x\rangle + r$ mit $s \in \mathbb{R}^n, r \in \mathbb{R}$. Deren Epigraph ist ein Halbraum, der durch einen nichthorizontalen Normalenvektor $(s, -1)$ ausgezeichnet ist.

Beobachtung 2.2.2. *Sei $f \in Conv\mathbb{R}^n$ und $x_0 \in ri(domf)$. Dann gibt es eine lineare Minorante, die f in x_0 stützt, d.h. $\exists s \in \mathbb{R}^n$:*

$$f(x) \geq f(x_0) + \langle s, x - x_0\rangle \quad \forall x \in \mathbb{R}^n.$$

Beweis. O.B.d.A. ist $aff(domf) = \mathbb{R}^n$. Wähle $x_0 \in int(domf)$. $(x_0, f(x_0))$ ist auf dem Rand des Epigraphen. Nach Lemma 2.1.21 gibt es eine Stützhyperebene in $(x_0, f(x_0))$ mit $s \in \mathbb{R}^n, \alpha \in \mathbb{R}$, so dass

$$\langle s, x\rangle + \alpha r \leq \langle s, x_0\rangle + \alpha f(x_0) \quad \forall (x, r) \in epif.$$

α muss ≤ 0 sein, sonst ergibt sich für $r \to \infty$ ein Widerspruch.
Wegen $x_0 \in int(domf)$ ist $\alpha = 0$ unmöglich (wähle $x = x_0 + \varepsilon s$), also $\alpha < 0$. Skaliere α auf den Wert -1

$$\Rightarrow f(x_0) + \langle s, x - x_0\rangle \leq f(x) \; \forall x \in \mathbb{R}^n$$

□

Lemma 2.2.3. *Sei $f \in Conv\mathbb{R}^n$ und es gebe $x_0 \in \mathbb{R}^n, \delta > 0, m, M \in \mathbb{R}$, so dass*

$$m \leq f(x) \leq M \; \forall x \in B_{2\delta}(x_0).$$

Dann ist f Lipschitz-stetig auf $B_\delta(x_0)$:

$$|f(y) - f(y')| \leq \tfrac{M-m}{\delta} \|y - y'\| \; \forall y, y' \in B_\delta(x_0).$$

Beweis. Wähle $y, y' \in B_\delta(x_0)$ und setze $y'' := y' + \delta\frac{y'-y}{\|y'-y\|} \in B_{2\delta}(x_0)$.
Nach Konstruktion gilt $y' \in [y'', y]$, genauer: $y' = \frac{\|y'-y\|}{\delta+\|y'-y\|}y'' + \frac{\delta}{\delta+\|y'-y\|}y =: \alpha y'' + (1-\alpha)y$.
Da f konvex und $m \le f(x) \le M$ ist, folgt

$$\begin{aligned} f(y') - f(y) &\le \alpha f(y'') + (1-\alpha)f(y) - f(y) \\ &= \tfrac{\|y'-y\|}{\delta+\|y'-y\|}(f(y'') - f(y)) \\ &\le \tfrac{1}{\delta}\|y'-y\|(M-m). \end{aligned}$$

Ebenso für y vertauscht mit y'. □

Satz 2.2.4. *Sei $f \in Conv\mathbb{R}^n$ und S eine kompakte Teilmenge von $ri(domf)$. Dann gibt es $L = L(S) \ge 0$ mit*

$$|f(x) - f(x')| \le L\|x - x'\| \; \forall x, x' \in S.$$

Beweis. O.B.d.A. sei $aff(domf) = \mathbb{R}^n$, also $ri(domf) = int(domf)$.
Wir werden zeigen:

$$(*) \quad \forall x_0 \in S \quad \exists \delta = \delta(x_0), L = L(x_0, \delta): \; \begin{array}{l} B_\delta(x_0) \subseteq int(domf) \\ |f(y) - f(y')| \le L\|y - y'\| \; \forall y, y' \in B_\delta(x_0). \end{array}$$

$\Rightarrow$ Bekommen damit offene Überdeckung von S; wegen Kompaktheit lässt sich eine endliche Überdeckung mit den Parametern

$$(x_1, \delta_1, L_1), \ldots, (x_k, \delta_k, L_k)$$

extrahieren. Setzen dann $L := \max\{L_1, \ldots, L_k\}$. Für $[x, x'] \subseteq S$ zerlege

$$[x, x'] = [\underbrace{y_0}_{=x}, y_1] \cup [y_1, y_2] \cup \ldots \cup [y_{l-1}, \underbrace{y_l}_{=x'}]$$

mit $[y_{i-1}, y_i] \subset B_{\delta_{k_i}}(x_{k_i})$, $i = 1, \ldots, l$ und $\|x - x'\| = \sum_{i=1}^{l} \|y_{i-1} - y_i\|$. Dann ist

$$|f(x) - f(x')| \le \sum |f(y_i) - f(y_{i-1})| \le \sum L\|y_i - y_{i-1}\| \le L\|x - x'\|.$$

Zeigen nun $(*)$ mit Lemma 2.2.3. Für $x_0 \in int(domf)$ wähle $\delta > 0$ so, dass

$$B_{2\delta}(x_0) \subseteq conv\{v_0, \ldots, v_n\} \subseteq domf$$

mit geeigneten $v_0, \ldots, v_n \in domf$. Jedes $y \in B_{2\delta}(x_0)$ ist darstellbar als $y = \sum_{i=0}^{n} \alpha_i v_i$ mit $\alpha \in \triangle_{n+1}$. Da f konvex ist, folgt

$$f(y) \le \sum \alpha_i f(v_i) \le \max f(v_i) =: M$$

und nach Beobachtung 2.2.2 (lineare Minorante in x_0)

$$\exists m : f(y) \ge m \; \forall y \in B_{2\delta}(x_0).$$

Also ist nach Lemma 2.2.3 f Lipschitz-stetig auf $B_\delta(x_0)$ mit $L(x_0, \delta) = \frac{M-m}{\delta}$. □

$\Rightarrow$ Jede konvexe Funktion ist stetig auf $ri(domf)$ (bezüglich $aff(domf)$).

2.2.1 Das Subdifferential einer konvexen Funktion

Der Gradient beschreibt mit $f(x)$ eine Tangential- oder Stützhyperebene. Für konvexe Funktionen ist diese nicht eindeutig. Das Subdifferential in einem Punkt beschreibt die Menge der Stützhyperebenen in einem Punkt.

Definition 2.2.2. *Sei $f \in Conv\mathbb{R}^n$. Dann heißt für $x \in \mathbb{R}^n$ die Menge*

$$\partial f(x) := \{s \in \mathbb{R}^n : f(y) \geq f(x) + \langle s, y - x\rangle \quad \forall y \in \mathbb{R}^n\}$$

das Subdifferential von f in x.

Bemerkung 2.2.5.

- *$f(y) \geq f(x) + \langle s, y - x\rangle$ wird Subgradientenungleichung genannt.*
- *Subgradient wirkt, im Gegensatz zum Gradienten, global $\forall y \in \mathbb{R}^n$.*
- *Existenz des Subdifferentials ist für $x \in ri(domf)$ nach Beobachtung 2.2.2 gesichert, also insbesondere für $f : \mathbb{R}^n \to \mathbb{R}$.*
- *Rand von ∂f wird durch Richtungsableitungen beschrieben.*

Definition 2.2.3. *Die Richtungsableitung einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ in Richtung $d \in \mathbb{R}^n$ ist*

$$f(x, d)' := \lim_{t \downarrow 0} \tfrac{f(x+td)-f(x)}{t}.$$

Beobachtung 2.2.6. *Für $f : \mathbb{R}^n \to \mathbb{R}$ konvex, ist*

$$f'(x, d) = \inf\left\{\tfrac{f(x+td)-f(x)}{t} : t > 0\right\}.$$

Beweis. Zeigen: Quotient ist monoton wachsend in t.
Sei $t_1 > t_2 > 0$. Dann ist aufgrund der Konvexität

$$\begin{aligned} & f(x + t_2 d) \leq \tfrac{t_2}{t_1} f(x + t_1 d) + (1 - \tfrac{t_2}{t_1}) f(x) \\ \Rightarrow\ & \tfrac{f(x+t_2 d)-f(x)}{t_2} \leq \tfrac{f(x+t_1 d)+(\frac{t_1}{t_2}-1)f(x)-\frac{t_1}{t_2}f(x)}{t_1} \\ & = \tfrac{f(x+t_1 d)-f(x)}{t_1}. \end{aligned}$$

□

Beobachtung 2.2.7. *Für $f : \mathbb{R}^n \to \mathbb{R}$ konvex und festes $x \in \mathbb{R}^n$ ist $f'(x, .)$ konvex, endlich und positiv homogen (d.h. $f'(x, \lambda d) = \lambda f'(x, d) \quad \forall \lambda > 0$).*

Beweis. Seien $d_1, d_2 \in \mathbb{R}^n, \alpha_1, \alpha_2 \geq 0, \alpha_1 + \alpha_2 = 1$. Da f konvex ist, gilt für alle $t > 0$:

$$\begin{aligned} f(x + t(\alpha_1 d_1 + \alpha_2 d_2)) - f(x) &= f(\alpha_1(x + td_1) + \alpha_2(x + td_2)) - \alpha_1 f(x) - \alpha_2 f(x) \\ &\leq \alpha_1[f(x + td_1) - f(x)] + \alpha_2[f(x + td_2) - f(x)]. \end{aligned}$$

Division durch t und $t \downarrow 0$ liefert

$$\Rightarrow f'(x, \alpha_1 d_1 + \alpha_2 d_2) \leq \alpha_1 f'(x, d_1) + \alpha_2 f'(x, d_2),$$

also ist $f'(x, \cdot)$ konvex.
Die positive Homogenität folgt aus der Definition:

$$f'(x, \lambda d) = \lim_{t \downarrow 0} \lambda \tfrac{f(x+\lambda t d) - f(x)}{\lambda t} = \lambda \lim_{\tau \downarrow 0} \tfrac{f(x+\tau d) - f(x)}{\tau} = \lambda f'(x, d).$$

Die Endlichkeit folgt, da f für $x \in int(dom f)$ nach Satz 2.2.4 lokal Lipschitzstetig ist:
Sei $\|d\| = 1$. $\Rightarrow \exists \varepsilon > 0, L > 0$ mit

$$|f(x + td) - f(x)| \leq Lt \quad \forall 0 \leq t \leq \varepsilon \qquad \Rightarrow f'(x, d) \leq L.$$

Für $\|d\| \neq 1$ folgt aus der positiven Homogenität und dem eben Gezeigten

$$f'(x, d) = \|d\| f'(x, \tfrac{d}{\|d\|}) \leq L\|d\|.$$

□

Definition 2.2.4. *Eine positiv homogene Funktion $f \in Conv\mathbb{R}^n$ nennt man sublinear.*

Bemerkung 2.2.8. *Der Epigraph einer sublinearen Funktion ist ein Kegel.*

Satz 2.2.9. *Für konvexes $f : \mathbb{R}^n \to \mathbb{R}$ ist*

$$\partial f(x) = \{s \in \mathbb{R}^n : \langle s, d\rangle \leq f'(x, d)\ \forall d \in \mathbb{R}^n\}.$$

Insbesondere gilt:

$$f'(x, d) = \max\{\langle s, d\rangle : s \in \partial f(x)\}.$$

Beweis. "$\subseteq$": Sei $s \in \partial f(x)$. Für alle $t > 0$ ist nun $\tfrac{f(x+td)-f(x)}{t} \geq \tfrac{f(x)+t\langle s,d\rangle - f(x)}{t} = \langle s, d\rangle$.

"$\supseteq$": Nach Beobachtung 2.2.6 ist $\forall d \in \mathbb{R}^n\ \forall t > 0$: $\langle s, d\rangle \leq f'(x, d) \leq \tfrac{f(x+td)-f(x)}{t}$
$\Rightarrow f(x) + t\langle s, d\rangle \leq f(x + td)$.
Für $y = x + td \in \mathbb{R}^n$ beliebig ist also $f(x) + \langle s, y - x\rangle \leq f(y)$, also $s \in \partial f(x)$.
$f'(x, .)$ ist konvex, wähle also nach Beobachtung 2.2.2 eine Stützhyperebene s in $\bar{d}$:

$$f'(x, \bar{d}) + \langle s, d - \bar{d}\rangle \leq f'(x, d).\ \forall d \in \mathbb{R}^n$$

Für $d = \lambda\bar{d}$: $(\lambda - 1)\langle s, \bar{d}\rangle \leq (\lambda - 1) f'(x, \bar{d}) \Rightarrow \langle s, \bar{d}\rangle = f'(x, \bar{d})\ \forall \lambda$.
Setzen ein: $\langle s, d\rangle \leq f'(x, d)\ \forall d$. Nach dem ersten Teil heißt das $s \in \partial f(x)$. □

Satz 2.2.10. *Für konvexes $f : \mathbb{R}^n \to \mathbb{R}$ sind äquivalent:*

1. $x \in \arg\min f$ *(d.h.* $f(y) \geq f(x)\ \forall y \in \mathbb{R}^n$)
2. $0 \in \partial f(x)$
3. $f'(x,d) \geq 0\ \forall y \in \mathbb{R}^n$.

Beweis. Aus Definition des Subgradienten folgt:

$$f(y) \geq f(x) = f(x) + \langle 0, y-x\rangle \quad \forall y \quad \Leftrightarrow \quad 0 \in \partial f(x).$$

Aus Satz 2.2.9 folgt: $0 \in \partial f(x) \quad \Leftrightarrow \quad \langle 0,d\rangle = 0 \leq f'(x,d) \quad \forall d \in \mathbb{R}^n$. □

Über den Epigraphen lässt sich das Subdifferential gut geometrisch beschreiben.

Beobachtung 2.2.11. *Sei* $f : \mathbb{R}^n \to \mathbb{R}$ *konvex.*

1. *Ein Vektor* $s \in \mathbb{R}^n$ *ist ein Subgradient von* f *in* x *genau dann, wenn* $(s,-1) \in \mathbb{R}^n \times \mathbb{R}$ *normal auf* epi$f$ *in* $(x, f(x))$ *steht, d.h.*

$$N_{\text{epi}f}(x, f(x)) = \{(\lambda s, -\lambda) : s \in \partial f(x), \lambda \geq 0\}.$$

2. *Der Tangentialkegel von* epi$f$ *in* $(x, f(x))$ *ist gerade der Epigraph von* $d \mapsto f'(x,d)$*, d.h.*

$$T_{\text{epi}f}(x, f(x)) = \{(d,r) : r \geq f'(x,d)\}.$$

Beweis. 1. Aus der Definition des Normalenkegels folgt:

$$\begin{aligned}(s,-1) \in N_{epif}(x,f(x)) &\Leftrightarrow \langle s, y-x\rangle + \langle -1, r - f(x)\rangle \leq 0\ \forall y \in \mathbb{R}^n\ \forall r \geq f(y)\\ &\Leftrightarrow f(x) + \langle s, y-x\rangle \leq f(y)\ \forall y \in \mathbb{R}^n\\ &\Leftrightarrow s \in \partial f(x).\end{aligned}$$

$\lambda \geq 0$ streckt entsprechend.

2. Der Tangentialkegel ist polar zum Normalenkegel (Korollar 2.1.36), also

$$T_{epif}(x,f(x)) = \{(d,r) \in \mathbb{R}^n \times \mathbb{R} : \langle \lambda s, d\rangle + (-\lambda r) \leq 0\ \forall s \in \partial f(x)\ \lambda \geq 0\}.$$

Für $\lambda = 0$ ist dies erfüllt. Sei $\lambda > 0$.

$$\begin{aligned}\langle s,d\rangle \leq r\ \forall s \in \partial f(x) &\Leftrightarrow r \geq \max_{s \in \partial f(x)} \langle s,d\rangle = f'(x,d)\\ &\Leftrightarrow (d,r) \in epif'(x,.).\end{aligned}$$

□

In der Optimierung interessieren uns besonders die Mengen $\{x \in \mathbb{R}^n : f(x) \leq 0\}$, also die Niveaumengen konvexer Nebenbedingungen f:

$$Sf(x) := S_{f(x)}(f) = \{y \in \mathbb{R}^n : f(y) \leq f(x)\}.$$

Lemma 2.2.12. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex. Dann ist $T_{Sf(x)}(x) \subseteq \{d : f'(x,d) \leq 0\}$*
"Der Tangentialkegel der Niveaumenge in x ist Teilmenge der Abstiegsrichtungen."

Beweis. Sei $y \in Sf(x), t > 0$, setze $d := t(y - x)$.

$$0 \geq t[f(y) - f(x)] = \frac{f(x+\frac{1}{t}d)-f(x)}{\frac{1}{t}} \geq f'(x,d)$$
$$\Rightarrow d \in \underbrace{\mathbb{R}^+[Sf(x) - x]}_{\text{Abschluss davon ist Tangentialkegel}} \subset \underbrace{\{d : f'(x,d) \leq 0\}}_{\text{abgeschlossen, weil } f' \text{ stetig von } d \text{ abhängt.}}$$

□

Leider gilt die umgekehrte Inklusion im allgemeinen nicht: Sei $f(x) = \frac{1}{2}\|x\|^2$; für $x = 0$ ist $Sf(x) = \{0\}$ und $f'(0,d) = 0\ \forall d$, also $T_{Sf(x)}(x) = \{0\} \subsetneq \mathbb{R}^n = \{d : f'(0,d) \leq 0\}$. Das, was Probleme macht, ist $0 \in \partial f(x)$. Um das zu sehen brauchen wir ein technisches Lemma.

Beobachtung 2.2.13. *Sei $g : \mathbb{R}^n \to \mathbb{R}$ konvex und es gebe $x_0 \in \mathbb{R}^n$ mit $g(x_0) < 0$. Dann ist*

1. $cl\{z : g(z) < 0\} = \{z : g(z) \leq 0\}$
2. $\{z : g(z) < 0\} = int\{z : g(z) \leq 0\}$.

Beweis. 1. "$\subseteq$": gilt weil g stetig.
"$\supseteq$": Sei $\bar{z} \in \mathbb{R}^n : g(\bar{z}) \leq 0$. Wegen $g(x_0) < 0$ ist für $z_k := \frac{1}{k}x_0 + (1 - \frac{1}{k})\bar{z}$ wegen der Konvexität $g(z_k) < 0$ und für $k \to \infty$ folgt $\bar{z} \in cl\{z : g(z) < 0\}$.

2. Wegen 1. ist $int\, cl\{z : g(z) < 0\} = int\{z : g(z) \leq 0\}$; nach Beobachtung 2.1.9 ist $int\, cl\{z : g(z) < 0\} = int\{z : g(z) < 0\}$.
Wegen der Stetigkeit von g gilt $int\{z : g(z) < 0\} = \{z : g(z) < 0\}$.

□

Ein solches x_0 wird *Slater-Punkt* genannt. Die Annahme, dass x_0 existiert, wird *Slater-Annahme* genannt. Sie entspricht der Bedingung "ri Epipraph $\cap\, ri$ Hyperebene zum Niveau $0 \neq \varnothing$".

Satz 2.2.14. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und $0 \notin \partial f(x)$. Dann gilt für die Niveaumenge $Sf(x)$*

1. $T_{Sf(x)}(x) = \{d \in \mathbb{R}^n : f'(x,d) \leq 0\}$
2. $int[T_{Sf(x)}(x)] = \{d \in \mathbb{R}^n : f'(x,d) < 0\} \neq \varnothing$.

Beweis. Nach Satz 2.2.9 gilt: $\partial f(x) = \{s \in \mathbb{R}^n : \langle s, d\rangle \leq f'(x,d)\ \forall d \in \mathbb{R}^n\}$.
Da $0 \notin \partial f(x) \Rightarrow \exists d \in \mathbb{R}^n : f'(x,d) < 0 \Rightarrow \exists \bar{t} > 0 : f(x + \bar{t}d) < f(x)$.
Es ist $d = \frac{x+td-x}{t}$ und $x + td \in Sf(x) \Rightarrow d \in T_{Sf(x)}(x)$. Wir haben gezeigt, dass

$$\{d : f'(x,d) < 0\} \subseteq \mathbb{R}^+[Sf(x) - \{x\}] \subseteq T_{Sf(x)}(x) \text{ (\textit{nach Beobachtung} 2.1.34)}.$$

Wenden Beobachtung 2.2.13, (1.) auf $g = f'(x,.)$ an (nach Beobachtung 2.2.7 ist g konvex), denn $\exists d : f'(x,d) < 0$, also ist $cl\{d : f'(x,d) < 0\} = \{d : f'(x,d) \leq 0\}$. Wegen Lemma 2.2.12 ist

$$T_{Sf(x)}(x) \subseteq \{d : f'(x,d) \leq 0\} \subseteq cl\mathbb{R}^+[Sf(x) - \{x\}] = T_{Sf(x)}(x)$$

2. folgt aus Beobachtung 2.2.13, (2.). □

Satz 2.2.15. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und $0 \notin \partial f(x)$. Dann gilt:*

$$N_{Sf(x)}(x) = \mathbb{R}^+ \partial f(x).$$

Beweis. Aus Satz 2.2.14, (1.) und Satz 2.2.9 folgt:

$$\begin{aligned} T_{Sf(x)} &= \{d \in \mathbb{R}^n : \langle s, d\rangle \leq 0\ \forall s \in \partial f(x)\} \\ &= \{d \in \mathbb{R}^n : \langle \lambda s, d\rangle \leq 0\ \forall \lambda \geq 0, s \in \partial f(x)\} = [\mathbb{R}^+ \partial f(x)]^\circ. \end{aligned}$$

Also ist nach Beobachtung 2.1.35 $N_{Sf(x)}(x) = [T_{Sf(x)}(x)]^\circ = \mathbb{R}^+ \partial f(x)$. □

In der Optimierung ist die zulässige Menge meist durch konvexe $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$ beschrieben durch

$$\{x : g_i(x) \leq 0, i = 1, \ldots, m\};$$

oder mit $g(x) := \max\{g_i(x) : i = 1, \ldots, m\}$ durch

$$x \in S_0(g) = \bigcap_{i=1,\ldots,m} S_0(g_i).$$

Wie kann man das Subdifferential und den Normalenkegel zum "*Level set*" dafür bestimmen?

Satz 2.2.16. *Seien $g_i : \mathbb{R}^n \to \mathbb{R}$ konvex und $g(x) = \max\{g_i(x) : i = 1, \ldots, m\}$. Dann ist*

$$\partial g(x) = conv \left\{ \bigcup_{i \in I(x)} \partial g_i(x) \right\} \text{ mit } I(x) = \{i : g(x) = g_i(x)\}.$$

Beweis. "$\supseteq$": Sei $s \in C := conv\{\cup_{i\in I(x)}\partial g_i(x)\}$, also

$$s = \sum_{i\in I(x)} \alpha_i s_i \text{ mit } s_i \in \partial g_i(x), i \in I(x), \alpha_i \geq 0, \sum \alpha_i = 1.$$

Dann gilt für beliebiges $y \in \mathbb{R}^n$:

$$\begin{aligned} g(x) + \langle s, y - x\rangle &= \sum_{i\in I(x)} \alpha_i g_i(x) + \sum_{i\in I(x)} \alpha_i \langle s_i, y - x\rangle \\ &= \sum_{i\in I(x)} \alpha_i (g_i(x) + \langle s_i, y - x\rangle) \\ &\leq \sum_{i\in I(x)} \alpha_i g(y) \\ &\leq \max_{i\in I(x)} g_i(y) \\ &\leq g(y). \end{aligned}$$

"$\subseteq$": Sei $\bar{s} \notin C$. Nach Beobachtung 2.2.7 und Satz 2.2.9 sind $\partial g_i(x)$ kompakt; nach Satz 2.1.5 ist dann C konvex und kompakt. Nach dem Trennungssatz 2.1.19 existiert z mit

$$\langle \bar{s}, z\rangle > \sup_{s\in C}\langle s, z\rangle = \max_{i\in I(x)}\{\langle s, z\rangle : s \in \partial g_i(x)\}.$$

Nach Satz 2.2.9 gilt $\forall i \in I(x)$

$$g_i'(x, z) < \langle \bar{s}, z\rangle \quad \Rightarrow g_i(x + tz) < g_i(x) + t\langle \bar{s}, z\rangle \quad \text{für } t > 0 \text{ klein genug.}$$

Damit ist $g(x) + \langle \bar{s}, tz\rangle > \max_{i\in I(x)}\{g_i(x + tz)\}$ für $t > 0$ klein genug. Wegen der Stetigkeit der g_i ist für hinreichend kleines $t > 0$ auch

$$g(x) + \langle \bar{s}, tz\rangle > \max_{i\in\{1,\ldots,m\}\setminus I(x)} g_i(x + tz),$$

also ist $g(x) + \langle \bar{s}, tz\rangle > g(x + tz)$ und damit $\bar{s} \in \partial g(x)$. □

Beobachtung 2.2.17. *Seien $g_i : \mathbb{R}^n \to \mathbb{R}$ konvex, sei $g(x) := \max\{g_1(x), \ldots, g_m(x)\}$ und $I(x) = \{i : g_i(x) = g(x)\}$, $S_0(g) = \{x : g(x) \leq 0\}$.*

1. *Falls $g(x) = 0$ und $0 \notin \partial g(x)$, dann ist*

$$N_{S_0(g)}(x) = \mathbb{R}_+\partial g(x) = \left\{ \sum_{i\in I(x)} \lambda_i s_i : \lambda_i \geq 0, s_i \in \partial g_i(x)\ \forall i \in I(x) \right\}.$$

2. *Falls $\exists x_0 : g_i(x_0) < 0\ \forall i = 1, \ldots, m$ ($\Leftrightarrow g(x_0) < 0$, Slater(regularitäts)bedingung), dann ist $0 \notin \partial g(x)\ \forall x : g(x) = 0$.*

Beweis. 1. Satz 2.2.15 auf Satz 2.2.16 anwenden.

2. Annahme: $g(x) = 0$ und $0 \in \partial g(x)$. Mit der Subgradientenungleichung folgt:

$$0 > g(x_0) \geq g(x) + \langle 0, x_0 - x\rangle = 0 \quad \Rightarrow \quad \text{Widerspruch.}$$

□

2.2.2 Optimalitätsbedingungen für Aufgaben mit Nebenbedingungen

Betrachten für konvexes $f : \mathbb{R}^n \to \mathbb{R}$ und konvexes $C \subseteq \mathbb{R}^n$ die Aufgabe

$$\boxed{\begin{array}{l}\min f(x)\\ \text{s.t. } x \in C\end{array}}$$

Wollen Analogon zu Satz 2.2.10 formulieren.

Satz 2.2.18. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex, $C \subseteq \mathbb{R}^n$ konvex und abgeschlossen.*
Dann ist für $\bar{x} \in C$ äquivalent:

1. *$\bar{x}$ minimiert $f(\cdot)$ über C.*
2. *$f'(\bar{x}, d) \geq 0 \quad \forall d \in T_C(\bar{x})$ (die zulässigen Richtungen sind keine Abstiegsrichtungen).*
3. *$0 \in \partial f(\bar{x}) + N_C(\bar{x})$ (d.h. Abstiegsrichtungen $s \in -\partial f(\bar{x})$ liegen im Normalenkegel).*

Beweis. "1. $\Rightarrow$ 2." Für beliebige $y \in C$ gilt $f(y) \geq f(\bar{x})$.
Für $\bar{x} \neq y \in C$ ist dann $d := y - \bar{x} \in T_C(\bar{x})$ und $f(\bar{x} + td) \geq f(\bar{x})\ \forall t \in [0, 1]$.

$$\Rightarrow f'(\bar{x}, d) = \lim_{t \downarrow 0} \tfrac{1}{t}(f(\bar{x} + td) - f(\bar{x})) \geq 0.$$

Sei $d \in T_C(\bar{x})$ beliebig, d.h. es existieren $y_k \in C, \alpha_k > 0$, so dass $d_k := \alpha_k(y_k - \bar{x})$ und $d_k \to d$. Für diese d_k gilt nach Obigem aber $f'(\bar{x}, d_k) \geq 0$. Wegen der Stetigkeit von $f'(\bar{x}, .)$ gilt damit $f'(\bar{x}, d) \geq 0$.

"2. $\Rightarrow$ 1." Sei $\bar{x} \neq y \in C. \Rightarrow \bar{d} = y - \bar{x} \in T_C(\bar{x})$ und es gilt

$$0 \leq f'(\bar{x}, \bar{d}) = f'(\bar{x}, y - \bar{x}) \leq \tfrac{1}{t}(f(\bar{x} + t(y - \bar{x})) - f(\bar{x}))\ \forall t$$

nach Beobachtung 2.2.6; insbesondere folgt für $t = 1$: $f(y) - f(\bar{x})$, d.h. $f(\bar{x}) \leq f(y), \forall y \in C$.

"1. $\Rightarrow$ 3." Nutzen die Indikatorfunktion $i_C(x) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C. \end{cases}$
Dann sind die Aufgaben $\min\{f(x) : x \in C\}$ und $\min\{f(x) + i_C(x) : x \in \mathbb{R}^n\}$ äquivalent, insbesondere ist also $\bar{x} \in \arg\min\{f(x) + i_C(x)\}$ ($\bar{x}$ ist Optimallösung). Damit ist nach Satz 2.2.10

$$0 \in \partial(f + i_C)(\bar{x}).$$

Es gilt $\partial(f + i_C)(x) = \partial f(x) + \partial i_C(x)$ und für $x \in C$ gilt:

$$\begin{aligned} \partial i_C(x) &= \{s : \underbrace{i_C(x)}_{=0} + \langle s, y - x\rangle \leq i_C(y)\ \forall y \in \mathbb{R}^n\} \\ &= \{s : \langle s, y - x\rangle \leq 0, \forall y \in C\} \\ &= N_C(x), \end{aligned}$$

damit ist $0 \in \partial(f + i_C)(\bar{x}) = \partial f(\bar{x} + \partial i_C(\bar{x})) = \partial f(\bar{x}) + N_C(\bar{x})$.

"3. $\Rightarrow$ 1." $f(\bar{x}) + i_C(\bar{x}) + \langle 0, y - \bar{x}\rangle \leq f(y) + i_C(y)\ \forall y \in \mathbb{R}^n$ (Subgradientenungleichung)
$\Rightarrow f(\bar{x}) \leq f(y)\ \forall y \in C$.

□

Die zulässige Menge C ist meist nicht explizit bekannt, sondern durch Ungleichungen (mit konvexen Funktionen) beschrieben:

$$C = \bigcap_{i=1}^{m} \{x : g_i(x) \leq 0\} = \{x : g(x) \leq 0\} \quad \text{mit} \quad g(x) = \max\{g_1(x), \ldots, g_m(x)\}.$$

Entsprechend Beobachtung 2.2.17 ist (mit $I(x) = \{i : g_i(x) = 0\}$) eine zugängliche Beschreibung für den Normalenkegel $N_C(x)$ in $x \in C$ gegeben durch

$$\begin{aligned} N'(x) &:= \left\{ \sum_{i \in I(x)} \lambda_i s_i : \lambda_i \geq 0, s_i \in \partial g_i(x), i \in I(x) \right\} \cup \{0\} \\ &(= \{0\} \text{ falls } I(x) = \varnothing). \end{aligned}$$

Dies kann zu klein sein.
Beispiel: $g_1(x_1, x_2) = (x_1 - 1)^2 + x_2^2 - 1$
$g_2(x_1, x_2) = (x_1 + 1)^2 + x_2^2 - 1$

$$\Rightarrow C = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} \Rightarrow N_C\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) = \mathbb{R}^2,$$

$$\text{aber } N'\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) = \left\{ \lambda_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} : \lambda_1, \lambda_2 \geq 0 \right\} = \mathbb{R} \times \{0\}.$$

Die Menge $N'(x)$ soll $N_C(x)$ ersetzen in den Optimalitätsbedingungen; dann ist $0 \in \partial f(\bar{x}) + N_C(\bar{x})$, aber $0 \notin \partial f(\bar{x}) + N'(x)$ möglich. Um die Gleichheit $N_C(x) = N'(x)$ zu garantieren, werden im allgemeinen zusätzliche Voraussetzungen benötigt (vgl. Beobachtung 2.2.17).

$N'(x)$ kann offen sein.
Beispiel: $C = \{(x_1, x_2) : \underbrace{x_1 + \|x\|}_{g(x_1, x_2)} \leq 0\} = \{(x_1, 0) : x_1 \leq 0\}$

$$\partial g(0,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + B_1(0)$$

$$\Rightarrow N'(x) = \left\{ \lambda \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix} + B_1(0) \right] : \lambda \geq 0 \right\} = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1 > 0 \right\} \cup \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

ist nicht abeschlossen. Dies kann (nur) passieren, falls $0 \in \partial g(x)$ (vgl. Beobachtung 2.1.6).

Lemma 2.2.19. *Sei $C = \{x : g_i(x) \leq 0\ \forall 1 = 1, \ldots, m\}$ konvex. Dann gilt für $x \in C$:*

1. $T_C(x) \subseteq T'(x) := (N'(x))^\circ$
2. $N_C(x) \supseteq clN'(x) \supseteq N'(x)$.

Beweis. 1. Für $\bar{g}(x) := \max\{0, g_1(x), \ldots, g_m(x)\}$ gilt $\bar{g}(x) = 0 \Leftrightarrow g(x) \leq 0$ und nach Satz 2.2.16 gilt

$$N'(x) = \left\{ \sum_{i \in I(x)} \lambda_i s_i : s_i \in \partial g_i(x), \lambda_i \geq 0 \right\} = \mathbb{R}^+ \partial \bar{g}(x).$$

$$\begin{aligned} \Rightarrow T'(x) = (N'(x))^\circ &= \{d : \langle d, s \rangle \leq 0 \, \forall s \in N'(x)\} \\ &= \{d : \langle d, s \rangle \leq 0 \, \forall s \in \partial \bar{g}(x)\} \\ &= \{d : \bar{g}'(x, d) \leq 0\} \\ &\supseteq T_{S\bar{g}(x)}(x) \\ &= T_{S_0(g)}(x) = T_C(x). \end{aligned}$$

2. $N_C(x) = (T_C(x))^\circ \supseteq (T'(x))^\circ = ((N'(x))^\circ)^\circ = clN'(x) \supseteq N'(x)$.

□

Satz 2.2.20. *Sei $\bar{x} \in C = \{x : g_i(x) \leq 0, \forall i = 1, \ldots, m\}$. Falls $N_C(\bar{x}) = N'(\bar{x})$, dann sind äquivalent:*

1. $\bar{x} \in \arg\min\{f(x) : x \in C\}$
2. $f'(\bar{x}, d) \geq 0 \, \forall d \in T_C(\bar{x})$
3. $0 \in \partial f(\bar{x}) + clN'(\bar{x})$
4. *Es existiert $\lambda = (\lambda_1, \ldots, \lambda_m)$ (Lagrangemultiplikatoren), so dass:*

$$\text{(KKT)} \begin{cases} 0 \in \partial f(\bar{x}) + \sum_{i=1}^m \lambda_i \partial g_i(\bar{x}) \\ \lambda \geq 0 \\ \lambda_i g_i(\bar{x}) = 0 \, \forall i = 1, \ldots, m \text{ (komplementärer Schlupf)} \end{cases}$$

[KKT: Karush-Kuhn-Tucker]
Falls $N_C(\bar{x}) \neq N'(\bar{x})$, gilt 1. $\Leftarrow$ 2. $\Leftrightarrow$ 3. $\Leftarrow$ 4., d.h. (KKT) sind hinreichend, aber im allgemeinen nicht notwendig.

Beweis. Sei $N_C(\bar{x}) = N'(\bar{x})$; da $N_C(\bar{x})$ abgeschlossen ist, gilt dann auch

$$N_C(\bar{x}) = N'(\bar{x}) = clN'(\bar{x}) \Rightarrow T_C(\bar{x}) = T'(\bar{x}).$$

Also: Satz 2.2.18 liefert $1. \Leftrightarrow 2. \Leftrightarrow 3.$
Setze $\lambda_i = 0 \; \forall i \notin I(\bar{x}) \Rightarrow 4.$ ist dann eine explizite Formulierung von $3.$

Sei $N_C(\bar{x}) \neq N'(\bar{x})$: Dann ist $T'(\bar{x}) := (N'(\bar{x}))^\circ = (clN'(\bar{x}))^\circ$
$(T'(\bar{x}))^\circ = ((N'(\bar{x}))^\circ)^\circ = clN'(\bar{x}).$

"$3. \Rightarrow 2.$" $0 \in \partial f(\bar{x}) + clN'(\bar{x})$

$$\Leftrightarrow \exists \quad \underbrace{s \in \partial f(\bar{x})}_{\langle s,d\rangle \leq f(\bar{x},d) \; \forall d \in \mathbb{R}^n} : \underbrace{-s \in clN'(\bar{x}) = (T'(\bar{x}))^\circ}_{\langle -s,d\rangle \leq 0 \; \forall d \in T'(\bar{x})}$$
$$\Rightarrow 0 \leq \langle s, d\rangle \leq f'(\bar{x}, d) \; \forall d \in T'(\bar{x}).$$

"$2. \Rightarrow 3.$" Annahme: $0 \notin \partial f(\bar{x}) + clN'(\bar{x})$, d.h. $\underbrace{\partial f(\bar{x})}_{\text{kompakt, konvex}} \cap \underbrace{(-clN'(\bar{x}))}_{\text{konvex}} = \varnothing$. Nach Korollar 2.1.20 existiert eine trennende Hyperebene, d.h.

$$\exists z \in \mathbb{R}^n : \underbrace{\max_{s \in \partial f(\bar{x})} \langle s, z\rangle}_{= f'(\bar{x},z), \text{ endlich}} < \underbrace{\inf_{s \in -clN'(\bar{x})} \langle s, z\rangle = 0}_{\text{da } -clN'(\bar{x}) \text{Kegel}, \ni 0}$$
$$\Rightarrow \langle s, z\rangle \leq 0 \; \forall s \in clN'(\bar{x}) \Rightarrow z \in (clN'(\bar{x}))^\circ = T'(\bar{x}) \text{ und } f'(\bar{x}, z) < 0$$

$\Rightarrow$ Widerspruch zu 2.

"$4. \Rightarrow 3.$" Nach Definition von $N'(\bar{x})$ folgt aus (KKT)

$$0 \in \partial f(\bar{x}) + \sum_{i \in I(x)} \lambda_i \partial g_i(\bar{x}) = \partial f(\bar{x}) + N'(\bar{x}) \subseteq \partial f(\bar{x}) + clN'(\bar{x}) \; .$$

"$2. \Rightarrow 1.$" Nach Lemma 2.2.19, (2.) ist $clN'(\bar{x}) \subseteq N_C(\bar{x})$, damit ist $0 \in \partial f(\bar{x}) + N_C(\bar{x})$ und nach Satz 2.2.18 folgt $\bar{x} \in \arg\min\{f(x) : x \in C\}$. □

Falls f, g_i zusätzlich differenzierbar sind, lauten die (KKT)-Bedingungen:

$$\exists \lambda \in \mathbb{R}^m_+ : \begin{cases} 0 = \nabla f(\bar{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x}) \\ \lambda_i g_i(\bar{x}) = 0, \forall i = 1, \ldots, m, \end{cases}$$

d.h. die optimalen Lösungen ergeben sich als Lösung eines (im allgemeinen nichtlinearen) Gleichungssystems.

Wesentlich ist die Eigenschaft $N_C(\bar{x}) = N'(\bar{x})$. Diese Bedingung wird als "*basic constraint qualification*" bezeichnet. Die Erfüllbarkeit dieser Forderung hängt nur von der Beschreibung von C durch die g_i ab.

Beobachtung 2.2.21. *Sei* $C = \{x : \langle a_i, x\rangle \leq b_i \; \forall i = 1, \ldots, m\}$. *Dann gilt für* $\bar{x} \in C$:

$$N_C(\bar{x}) = N'(\bar{x}) = cone\{a_i : i \in I(\bar{x})\}.$$

Beweis. Aus den Übungen ist bekannt $T_C(\bar{x}) = \{d : \langle a_i, d\rangle \leq 0 \quad \forall i \in I(\bar{x})\}$
$N_C(\bar{x}) = (T_C(\bar{x}))^\circ = cone\{a_i : i \in I(\bar{x})\}.$

□

Für polyedral beschriebene Mengen C ist also keine weitere Regularitätsforderung nötig, d.h. mit Beobachtung 2.2.21 und Satz 2.2.20 existieren stets Lagrangemultiplikatoren $\lambda_i, i = 1, \ldots, m$, die die Optimalität von $\bar{x}$ nachweisen.

Liegen affine Gleichungsrestriktionen $\langle a_i, x\rangle = b_i$ vor, dann gilt:

$$\begin{aligned} \langle a_i, x\rangle &\leq b_i \\ \langle a_{\hat{\imath}}, x\rangle = \langle -a_i, x\rangle &\leq -b_i. \end{aligned}$$

Für $x \in C$ sind damit $i, \hat{\imath} \in I(\bar{x})$ und damit enthält $N'(\bar{x})$ sowohl $\lambda_i a_i$ als auch $\lambda_{\hat{\imath}} a_{\hat{\imath}} = -\lambda_{\hat{\imath}} a_i$, also $(\lambda_i - \lambda_{\hat{\imath}})a_i \; \forall \lambda_i, \lambda_{\hat{\imath}} \geq 0$, d.h. $\mu_i a_i, \forall \mu_i \in \mathbb{R}$. Also: Für Aufgaben der Form

$$\min\{f(x) : Ax = b, g_i(x) \leq 0 \; i = 1, \ldots, m\}$$

schreibt man die (KKT)-Bedingungen wie folgt:

$$\exists \lambda \in \mathbb{R}^m_+, \exists \mu \in \mathbb{R}^n : \begin{cases} 0 \in \partial f(\bar{x}) + A^\tau \mu + \sum\limits_{i=1}^m \lambda_i g_i(\bar{x}) \\ \lambda_i g_i(\bar{x}) = 0 \; \forall i = 1, \ldots, m. \end{cases}$$

Sei $M(\bar{x})$ die Menge aller derartigen Multiplikatoren (μ, λ).

$M(\bar{x})$ ist konvex.
Beweis: Seien $(\mu^1, \lambda^1), (\mu^2, \lambda^2) \in M(\bar{x})$, sei $\alpha \in (0, 1)$.
zu zeigen: $\alpha(\mu^1, \lambda^1) + (1 - \alpha)(\mu^2, \lambda^2) \in M(\bar{x})$.

$$\begin{aligned} &\alpha\lambda^1 + (1-\alpha)\lambda^2 \in \mathbb{R}^m_+ \quad \surd \\ &(\alpha\lambda^1_i + (1-\alpha)\lambda^2_i) g_i(\bar{x}) = 0 \; \forall i = 1, \ldots, m \quad \surd \\ &\partial f(\bar{x}) + A^\tau(\alpha\mu^1 + (1-\alpha)\mu^2) + \textstyle\sum_{i=1}^m (\alpha\lambda^1_i + (1-\alpha)\lambda^2_i)\partial g_i(\bar{x}) \\ &= \underbrace{\alpha \underbrace{(\partial f(\bar{x}) + A^\tau\mu^1 + \sum_{i=1}^m \lambda^1_i \partial g_i(\bar{x}))}_{\ni 0} + (1-\alpha) \underbrace{(\partial f(\bar{x}) + A^\tau\mu^2 + \sum_{i=1}^m \lambda^2_i \partial g_i(\bar{x}))}_{\ni 0}}_{\ni 0}. \end{aligned}$$

$M(\bar{x})$ ist abgeschlossen.
Beweis: Sei $(\mu^k, \lambda^k) \in M(\bar{x}) \; \forall k$ und $(\mu^k, \lambda^k) \to (\mu, \lambda)$.

zu zeigen: $(\mu, \lambda) \in M(\bar{x})$.

$$\lambda^k \geq 0 \; \forall k \Rightarrow \lambda \geq > 0 \quad \surd$$
$$\lambda_i^k g_i(\bar{x}) = 0 \; \forall i \; \forall k \Rightarrow \lambda_i g_i(\bar{x}) = 0 \; \forall i \quad \surd$$
$$0 \in \underbrace{\partial f(\bar{x})}_{\text{kompakt}} + A^\tau \mu^k + \sum_{i=1}^{m} \lambda_i^k \underbrace{\partial g_i(\bar{x})}_{\text{kompakt}} \; \forall k.$$

Für fixiertes (μ^k, λ^k) sind dies konvexe, kompakte Mengen, die stetig von μ, λ abhängen

$$\underbrace{\Longrightarrow}_{\lambda^k \to \lambda, \mu^k \to \mu} \quad 0 \in \partial f(\bar{x}) + A^\tau \mu + \sum_{i=1}^{m} \lambda_i g_i(\bar{x}).$$

Falls A nicht vollen Zeilenrang hat, dann $\exists \bar{\mu} \neq 0 : A^\tau \bar{\mu} = 0$. Für $(\mu, \lambda) \in M(\bar{x})$ ist also $(\mu, \lambda) + t(\bar{\mu}, 0) \in M(\bar{x}) \; \forall t \in \mathbb{R}$, also ist $M(\bar{x})$ unbeschränkt.
Mit der Slaterbedingung hingegen ist $M(\bar{x})$ sogar kompakt.

Definition 2.2.5. *Die Menge $\{x : Ax = b, g_i(x) \leq 0 \; \forall i = 1, \ldots, m\}$ erfüllt die (starke) Slaterbedingung, falls A vollen Zeilenrang hat und ein x_0 existiert mit $Ax_0 = b$ und $g_i(x_0) < 0 \quad \forall i = 1, \ldots, m$.*

Satz 2.2.22. *Sei $C := \{x : Ax = b, g_i(x) \leq 0 \; \forall i = 1, \ldots, m\}$ und $\bar{x} \in \operatorname{Argmin}\{f(x) : x \in C\}$. Die Menge $M(\bar{x})$ der Lagrangemultiplikatoren ist nichtleer und kompakt genau dann, wenn die starke Slaterbedingung erfüllt ist.*

Beweis. "$\Leftarrow$": Für x_0 sei die starke Slaterbedingung erfüllt. Beobachtung 2.2.17 gilt (mit entsprechend angepasstem Kegel) auch mit den zusätzlichen Nebenbedingungen $Ax = b$ (Beweis: aufwendig), also ist $N_C(\bar{x}) = N'(\bar{x})$ und nach Satz 2.2.20 folgt $M(\bar{x}) \neq \emptyset$.

Für $d := x_0 - \bar{x}$ gilt $Ad = Ax_0 - A\bar{x} = b - b = 0$; für $i \in I(\bar{x})$ gilt unter Verwendung von Beobachtung 2.2.6 $g_i'(\bar{x}, d) \leq g_i(x_0) - g_i(\bar{x}) < 0$, folglich existiert ein $\varepsilon > 0$, so dass $g_i'(\bar{x}, d) < -\varepsilon \; \forall i \in I(\bar{x})$.

Sei nun $(\mu, \lambda) \in M(\bar{x})$; sei $h(x) := f(x) + \mu^\tau Ax + \sum \lambda_i g_i(x)$. Dann gilt

$$\partial h(\bar{x}) = \partial f(\bar{x}) + A^\tau \mu + \sum \lambda_i \partial g_i(\bar{x}) \overset{(\mu,\lambda) \in M(\bar{x})}{\ni} 0.$$

Nach Satz 2.2.9 ist dann $h'(\bar{x}, d) = \max\limits_{s \in \partial h(\bar{x})} \langle d, s \rangle \geq 0$; andererseits ist

$$\begin{aligned}
\max_{s \in \partial h(\bar{x})} \langle d, s \rangle &= \max_{s \in \partial f(\bar{x}) + A^\tau \mu + \sum \lambda_i \partial g_i(\bar{x})} \langle d, s \rangle \\
&\leq \max_{s_0 \in \partial f(\bar{x})} \langle d, s_0 \rangle + \mu^\tau \underbrace{Ad}_{=0} + \sum \lambda_i \max_{s_i \in \partial g_i(\bar{x})} \langle d, s_i \rangle \\
&= \underbrace{f'(\bar{x}, d)}_{\text{endlich}} + \sum \lambda_i \underbrace{g_i'(\bar{x}, d)}_{< -\varepsilon}.
\end{aligned}$$

Schließlich folgt $\sum \lambda_i \leq \frac{1}{\varepsilon} f'(\bar{x}, d)$ und aufgrund von $\lambda \geq 0$ die Beschränktheit der Multiplikatoren λ_i.

Die Bedingung $0 \in \partial f(\bar{x}) + A^\tau \mu + \sum \lambda_i \partial g_i(\bar{x})$ ist äquivalent zu

$$-A^\tau \mu \in \underbrace{\underbrace{\partial f(\bar{x})}_{\text{kompakt}} + \sum \underbrace{\lambda_i}_{\text{beschränkt}} \cdot \underbrace{\partial g_i(\bar{x})}_{\text{kompakt}}}_{\text{kompakt}},$$

d.h. $-A^\tau \mu$ liegt in einer kompakten Menge und aufgrund von $\ker A^\tau = \{0\}$ stammen auch die Multiplikatoren μ aus einer kompakten Menge, insbesondere sind sie beschränkt.

Zusammen mit der bereits oben gezeigten Abgeschlossenheit von $M(\bar{x})$ folgt die Kompaktheit.

"$\Rightarrow$": Die Slaterbedingung gelte nicht; zu zeigen ist dann: $M(\bar{x}) \neq \emptyset \;\Rightarrow\; M(\bar{x})$ unbeschränkt.

Nach obigen Betrachtungen gilt dies offensichtlich, falls A keinen vollen Zeilenrang hat.
Also habe A vollen Zeilenrang und $\nexists x : Ax = b, \quad g_i(x) < 0 \; \forall i = 1, \ldots, m.$

Für die Funktion

$$g(x) := \max\{g_1(x), \ldots, g_m(x), \langle a_1, x\rangle - b_1, b_1 - \langle a_1, x\rangle, \ldots, \langle a_h, x\rangle - b_h, b_h - \langle a_h, x\rangle\}$$

gilt dann $g(x) \geq 0 \; \forall x$ (weil $\nexists x : Ax = b, g_i(x) < 0 \; \forall i = 1, \ldots, m$) und $g(\bar{x}) = 0$ (weil $\bar{x}$ Optimallösung, insbesondere $\bar{x} \in C$); also ist $\bar{x} \in \operatorname{Argmin}\{g(x) : x \in \mathbb{R}^n\}$. Nach Satz 2.2.10 und Satz 2.2.16 existieren also Multiplikatoren

$$(\bar{\mu}, \bar{\lambda}) \in \mathbb{R}^h \times \mathbb{R}^m_+ : 0 \in \partial g(\bar{x}) = A^\tau \bar{\mu} + \sum_{i \in I(\bar{x})} \bar{\lambda}_i \partial g_i(\bar{x}) \text{ und } \bar{\lambda}_i = 0 \text{ für } i \notin I(\bar{x}).$$

Dabei kann $(\bar{\mu}, \bar{\lambda}) \neq 0$ gewählt werden (vgl. Satz 2.1.30).

Also: $(\mu, \lambda) \in M(\bar{x}) \;\Rightarrow\; (\mu, \lambda) + t(\bar{\mu}, \bar{\lambda}) \in M(\bar{x}) \; \forall \, t \geq 0$, also ist $M(\bar{x})$ unbeschränkt. □

2.3 Sattelpunkte

Betrachte für $X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}^m$ eine Funktion

$$l : X \times Y \to \mathbb{R}, \; (x, y) \mapsto l(x, y).$$

Derartige Funktionen entstehen in der Optimierung durch "*Lagrange- Relaxierung*", z.B.

$$\boxed{\begin{array}{l}\min c^\tau x \\ \text{s.t. } Ax = b \\ \quad\; x \geq 0\end{array}} \quad \Longleftrightarrow \quad \inf_{x \geq 0}(c^\tau x + \sup_{y \in \mathbb{R}^m} \langle b - Ax, y\rangle).$$

Die angegebene Äquivalenz folgt, da

$$\sup_{y\in\mathbb{R}^m}\langle b-Ax,y\rangle = \begin{cases}+\infty, \text{ falls } \exists i : b_i - A_{.,i}x \neq 0\\ 0, \text{ falls } b = Ax\end{cases}$$
$$\iff \inf_{x\geq 0}\sup_{y\in\mathbb{R}^m}\{l(x,y) := c^\tau x + \langle b-Ax,y\rangle\}$$

Was geschieht, wenn man nun $\inf$ und $\sup$ vertauscht?

Beobachtung 2.3.1. $\inf_{x\in X}\sup_{y\in Y} l(x,y) \geq \sup_{y\in Y}\inf_{x\in X} l(x,y)$.

Beweis. Für jedes $\bar{x}\in X, y\in Y$ ist $l(\bar{x},\bar{y}) \geq \inf_{x\in X} l(x,\bar{y})$

$$\Rightarrow \sup_{y\in Y} l(\bar{x},y) \geq \sup_{y\in Y}\inf_{x\in X} l(x,y) =: \text{ const}$$
$$\Rightarrow \inf_{x\in X}\sup_{y\in Y} l(x,y) \geq \sup_{y\in Y}\inf_{x\in X} l(x,y).$$

□

Beispiel 2.3.1.
$$\inf_{x\geq 0}\sup_{y\in\mathbb{R}}(c^\tau x + b^\tau y - x^\tau A^\tau y) \geq \sup_{y\in\mathbb{R}^m}\inf_{x\geq 0}(b^\tau y + \langle c - A^\tau y, x\rangle)$$
$$= \sup_{y\in\mathbb{R}}(b^\tau y + \inf_{x\geq 0}\langle c - A^\tau y, x\rangle).$$

Weil offensichtlich $\inf_{x\geq 0}\langle c - A^\tau y, x\rangle = \begin{cases} 0, & c - A^\tau y \geq 0\\ -\infty, & \text{sonst}\end{cases}$ *, ist die letzte Aufgabe äquivalent zu* $\max\{b^\tau y : A^\tau y \leq c, y \text{ frei }\}$.

Sattelpunkte haben mit Dualität zu tun. "Schwache Dualität" gilt immer (vgl. 2.3.1) , auch ohne Konvexitätseigenschaften.

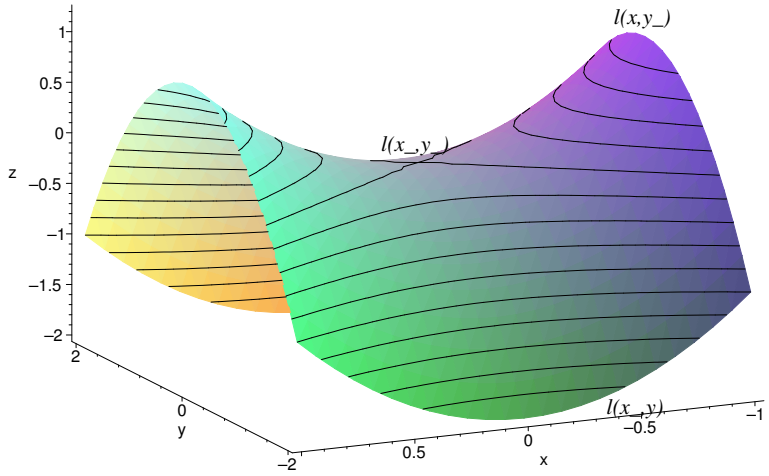
Offene Fragen: Wann ist $\inf_{x\in X}\sup_{y\in Y} l(x,y) = \sup_{y\in Y}\inf_{x\in X} l(x,y)$?
Gibt es dann auch $\bar{x},\bar{y}$, für die der Wert angenommen wird?

Definition 2.3.1. *Ein Paar* $(\bar{x},\bar{y})$ *heißt ein Sattelpunkt einer Funktion* $l : X\times Y\to\mathbb{R}$*, falls*

$$\sup_{y\in Y} l(\bar{x},y) = l(\bar{x},\bar{y}) = \inf_{x\in X} l(x,\bar{y}).$$

Äquivalent dazu sind: $l(\bar{x},y) \leq l(\bar{x},\bar{y}) \leq l(x,\bar{y})\ \forall (x,y)\in X\times Y$
$l(\bar{x},y) \leq l(x,\bar{y})\ \forall (x,y)\in X\times Y.$

Vorstellung:



x kann sich nicht verbessern, wenn sich y nicht bewegt;
y kann sich nicht verbessern, wenn sich x nicht bewegt.

Beobachtung 2.3.2. *Der Wert $l(\bar{x}, \bar{y})$ ist derselbe für alle Sattelpunkte $(\bar{x}, \bar{y})$. Sind $(\bar{x}_1, \bar{y}_1)$ und $(\bar{x}_1, \bar{y}_1)$ Sattelpunkte, dann auch $(\bar{x}_1, \bar{y}_2)$ und $(\bar{x}_2, \bar{y}_1)$.*

Beweis. $\forall (x, y) \in X \times Y$ gilt:

$$\begin{cases} l(\bar{x}_1, y) \leq l(\bar{x}_1, \bar{y}_1) \leq l(x, \bar{y}_1) \leftarrow x = \bar{x}_2, y = \bar{y}_2 \\ l(\bar{x}_2, y) \leq l(\bar{x}_2, \bar{y}_2) \leq l(x, \bar{y}_2) \leftarrow x = \bar{x}_1, y = \bar{y}_1 \end{cases}$$
$$\Longrightarrow \begin{cases} l(\bar{x}_1, \bar{y}_2) \leq l(\bar{x}_1, \bar{y}_1) \leq l(\bar{x}_2, \bar{y}_1) \\ l(\bar{x}_2, \bar{y}_1) \leq l(\bar{x}_2, \bar{y}_2) \leq l(\bar{x}_1, \bar{y}_2) \end{cases}$$

und somit gilt die Gleichheit, damit ist $\forall (x, y) \in X \times Y$

$$l(\bar{x}_1, y) \leq l(\bar{x}_1, \bar{y}_1) = l(\bar{x}_1, \bar{y}_2) = l(\bar{x}_2, \bar{y}_2) \leq l(x, \bar{y}_2)$$

$\Rightarrow (\bar{x}_1, \bar{y}_2)$ ist Sattelpunkt. Analog für $(\bar{x}_2, \bar{y}_1)$. □

Definieren zwei Funktionen:

$$\varphi(x) := \sup_{y \in Y} l(x, y), \quad x \in X \quad \text{“primale Funktion”, kann } +\infty \text{ annehmen}$$
$$\psi(y) := \inf_{x \in X} l(x, y), \quad y \in Y \quad \text{“duale Funktion”, kann } -\infty \text{ annehmen.}$$

Beobachtung 2.3.3. $\psi(y) \leq l(x, y) \leq \varphi(x) \quad \forall (x, y) \in X \times Y.$

Beweis. $\inf_{x \in X} l(x, y) \leq l(x, y) \leq \sup_{y \in Y} l(x, y).$ □

Satz 2.3.4. *Sei* $\Phi := \{\bar{x} \in X : \varphi(\bar{x}) = \inf_{x \in X} \varphi(x)\}, \Psi := \{\bar{y} \in X : \psi(\bar{y}) = \sup_{y \in Y} \psi(y)\}$.
l hat genau dann Sattelpunkte auf $X \times Y$, wenn $\min_{x \in X} \varphi(x) = \max_{y \in Y} \psi(y)$. In diesem Fall ist $\Phi \times \Psi$ die Menge der Sattelpunkte.

Beweis. Sei $(\bar{x}, \bar{y})$ ein Sattelpunkt. Dann ist $(\bar{x}, \bar{y}) \in \Phi \times \Psi$ und erfüllt $\min \varphi = \max \psi$.
Gelte nun $\min \varphi = \max \psi$, dann gibt es $\bar{x} \in \Phi$ und $\bar{y} \in \Psi$ mit $\varphi(\bar{x}) = \psi(\bar{y})$ oder $l(\bar{x}, y) \leq \varphi(\bar{x}) = l(x, \bar{y}) = \psi(\bar{y}) \leq l(x, \bar{y}) \quad \forall (x, y) \in X \times Y$. □

Um die Existenz von Sattelpunkten (d.h. starke Dualität) garantieren zu können, brauchen wir Konvexität und starke zusätzliche Annahmen.

(A1) $X \subseteq \mathbb{R}^n$ und $Y \subseteq \mathbb{R}^m$ sind nichtleer, konvex und abgeschlossen.

(A2) l ist stetig und konvex-konkav auf $X \times Y$, d.h.
für $y \in Y$ ist die Funktion $l(\cdot, y) : X \to \mathbb{R}$ konvex;
für $x \in X$ ist die Funktion $l(x, \cdot) : Y \to \mathbb{R}$ konkav.

(A3) X ist beschränkt oder $\exists y_0 \in Y : l(x, y_0) \to +\infty$ für $\|x\| \to \infty, x \in X$.

(A4) Y ist beschränkt oder $\exists x_0 \in X : l(x_0, y) \to -\infty$ für $\|y\| \to \infty, y \in Y$.

Satz 2.3.5. *Sind (A1) - (A4) erfüllt, dann hat l auf $X \times Y$ eine nichtleere konvexe, kompakte Menge von Sattelpunkten.*

Beweis. Falls es Sattelpunkte gibt, so gibt es nach Beobachtung 2.3.2 einen Sattelwert $\bar{l} = l(\bar{x}, \bar{y})$; dann ist wegen $l(\bar{x}, y) \leq \bar{l} \leq l(x, \bar{y}) \quad \forall (x, y) \in X \times Y$ die Menge $\Phi = \bigcap_{y \in Y} S_{\bar{l}}(l(\cdot, y))$ Schnitt der Niveaumengen der konvexen Funktionen $l(\cdot, y)$. Also ist Φ konvex und abgeschlossen (alle Niveaumengen abgeschlossen, da l stetig) und wegen (A3) ist mindestens eine Niveaumenge kompakt, also ist Φ kompakt.
Gleiches gilt für Ψ und für $\Phi \times \Psi$ (Menge aller Sattelpunkte nach Satz 2.3.4).

Beweisen nun die Existenz eines Sattelpunktes in drei Schritten, die jeweils schwächere Annahmen verwenden.
1. Schritt: Zusätzlich zu (A1) - (A4) seien X und Y beschränkt und $l(x, \cdot)$ streng konkav für $x \in X$. Damit sind die Funktionen

$$h_y(x) = l(x, y) + i_X(x) \in \overline{Conv}\mathbb{R}^n$$

für $y \in Y$ konvex und abgeschlossen. Also ist $\varphi(x) = \sup_{y \in Y} h_y(x)$ konvex und abgeschlossen (Schnitt abgeschlossener Epigraphen). Zu jedem $x \in X$ gibt es ein eindeutiges $y(x)$ mit $\varphi(x) = l(x, y(x))$; da das effektive Definitionsgebiet $dom(\varphi) = X$ kompakt ist, wird also $\min_{x \in X} \varphi(x)$ in einem $\bar{x} \in X$ angenommen. Setze $\bar{y} = y(\bar{x})$, dann gilt:

$$\varphi(\bar{x}) = l(\bar{x}, \bar{y}) \geq l(\bar{x}, y) \quad \forall y \in Y.$$

Für die zweite Ungleichung wähle $x \in X$ beliebig und definiere für $k = 1, 2, \ldots$

$$x_k := \tfrac{1}{k}x + (1 - \tfrac{1}{k})\bar{x}, y_k = y(x_k)$$
$$\Rightarrow \varphi(\bar{x}) \leq \varphi(x_k) = l(x_k, y_k) \underbrace{\leq}_{l(\cdot, y_k) \text{ konvex}} \underbrace{\frac{1}{k}}_{\to 0} l(x, y_k) + (1 - \tfrac{1}{k}) \underbrace{l(\bar{x}, y)}_{l(\bar{x},\bar{y}) \geq \varphi(\bar{x})}.$$

Da Y kompakt, existiert eine konvergente Teilfolge $y_k \overset{K}{\to} \tilde{y}$, so dass der Ausdruck auf der rechten Seite gegen $0 + l(\bar{x}, \tilde{y}) \geq \varphi(\bar{x})$ konvergiert. Da $l(\bar{x}, \cdot)$ streng konkav ist, folgt $\bar{y} = \tilde{y}$.
Andererseits folgt auch

$$\varphi(\bar{x}) \leq \tfrac{1}{k} l(x, y_k) + (1 - \tfrac{1}{k}) \underbrace{l(\bar{x}, \bar{y})}_{=\varphi(\bar{x})};$$

nach Multiplikation mit k folgt $\varphi(\bar{x}) \leq l(x, y_k) \overset{K}{\longrightarrow} l(x, \bar{y})\ \forall x \in X$, d.h. ist $(\bar{x}, \bar{y})$ Sattelpunkt.

2. Schritt: Zusätzlich zu (A1) - (A4) sei nun nur die Beschränktheit von X und Y vorausgesetzt. Definiere für $k = 1, 2, \ldots$ die Funktionen

$$l_k(x, y) = l(x, y) - \tfrac{1}{k}\|y\|^2.$$

Diese Funktionen sind streng konvex in y, folglich hat jedes l_k einen Sattelpunkt $(\bar{x}_k, \bar{y}_k)$. Also gilt für alle $(x, y) \in X \times Y$:

$$l(\bar{x}_k, y) - \tfrac{1}{k}\|y\|^2 \leq l(x, \bar{y}_k) - \tfrac{1}{k}\|\bar{y}_k\|^2.$$

Aufgrund der Beschränktheit existiert eine konvergente Teilfolge K, also $(\bar{x}_k, \bar{y}_k) \overset{K}{\longrightarrow} (\bar{x}, \bar{y})$. Aus obiger Ungleichung folgt

$$l(\bar{x}, y) \leq l(x, \bar{y}) \quad \forall (x, y) \in X \times Y,$$

was der äquivalenten Definition eines Sattelpunktes entspricht.

3.Schritt: Es gelte nur (A1) - (A4). Für $k = 1, 2, \ldots$ definiere

$$X_k := X \cap B_k(0), Y_k := Y \cap B_k(0).$$

Nach dem 2. Schritt hat l für jedes k auf $X_k \times Y_k$ einen Sattelpunkt $(\bar{x}_k, \bar{y}_k) \in X_k \times Y_k$, also

$$l(\bar{x}_k, y) \leq l(x, \bar{y}_k) \quad \forall (x, y) \in X_k \times Y_k.$$

Annahme: $\{\bar{y}_k\}$ ist unbeschränkt, also auch Y unbeschränkt.
Für hinreichend großes k ist nach (A4) $x_0 \in X_k$ und damit $l(\bar{x}_k, y) \leq l(x_0, \bar{y}_k) \to -\infty$, folglich muss für beliebiges $y \in Y$ gelten $l(\bar{x}_k, y) \to -\infty$. Dies geht nur, wenn $\{\bar{x}_k\}$ unbeschränkt ist, also X unbeschränkt.
Nach (A3) ist $y_0 \in Y_k$ für hinreichend großes k und deshalb

$$+\infty \longleftarrow l(\bar{x}_k, y_0) \leq l(x_0, \bar{y}_k) \longrightarrow -\infty.$$

Dieser Widerspruch widerlegt die Annahme, also ist $\{\bar{y}_k\}$ beschränkt; analog auch $\{\bar{x}_k\}$. Folglich gibt es eine konvergente Teilfolge K mit $(\bar{x}_k, \bar{y}_k) \to (\bar{x}, \bar{y})$
$l(\bar{x}, y) \leq l(x, \bar{y}) \quad \forall (x, y) \in X_k \times Y_k$.

□

Unter den Annahmen (A1) - (A4) hat man also anstelle der allgemein gültigen Ungleichung

$$\inf_{x \in X} \sup_{y \in Y} l(x, y) \geq \sup_{y \in Y} \inf_{x \in X} l(x, y)$$

die viel nützlichere Beziehung

$$\min_{x \in X} \sup_{y \in Y} l(x, y) \geq \max_{y \in Y} \inf_{x \in X} l(x, y).$$

2.4 Lagrangefunktion und Dualität

Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex, $g_i : \mathbb{R}^n \to \mathbb{R}$ konvex und $X \subseteq \mathbb{R}^n$ konvex und von einfacher Struktur. Für das Optimierungsproblem $\min f(x)$ s.t. $g_i(x) \leq 0, i = 1, \ldots, m$, $x \in X$ betrachten wir die Lagrangefunktion

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$

Für festes $\lambda \geq 0$ kann man

$$\min_{x \in X} L(x, \lambda)$$

als abgeändertes Optimierungsproblem betrachten, in dem λ die Verletzung der Zulässigkeit bestraft, aber leider auch die Übererfüllung belohnt.

$$\min_{x \in X} \sup_{\lambda \geq 0} L(x, \lambda)$$

ist wieder das Originalproblem: Ist ein $g_i(x) > 0$, wähle $\lambda_i \to \infty$, also $\sup = \infty$,
ist $g_i(x) \leq 0$ für alle i, wähle $\lambda = 0$, also $\sup = 0$.
Nur der Fall $\sup = 0$, also Zulässigkeit von x, ist relevant für das $\min$.

Nach Beobachtung 2.3.1 bekommt man eine Schranke für den Optimalwert, wenn man das "duale Optimierungsproblem"

$$\sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda)$$

lösen kann. Oft ist $\inf_{x \in X} L(x, \lambda)$ für festes λ leicht zu bestimmen; und nachträglich über λ zu optimieren beherrscht man (Subgradienten- oder Bündelverfahren).
Wenn $L(x, \lambda)$ Sattelpunkte besitzt, bekommt man sogar den richtigen Wert: "starke Dualität".

Satz 2.4.1. *Sei $X = \mathbb{R}^n$. Die Sattelpunkte von $L(x, \lambda)$ auf $\mathbb{R}^n \times \mathbb{R}^m_+$ sind Punkte $(\bar{x}, \bar{\lambda})$ mit*

1. *$\bar{x}$ minimiert $L(\cdot, \bar{\lambda})$ auf $\mathbb{R}^n$*
2. *$g_i(\bar{x}) \leq 0 \quad \forall i = 1, \ldots, m$ (primale Zulässigkeit)*
3. *$\lambda_i g_i(\bar{x}) = 0 \quad \forall i = 1, \ldots, m$ (Komplementarität).*

Beweis. 1. ist Teil der Sattelpunkteigenschaft.
Der zweite Teil der Sattelpunkteigenschaft besagt: $\bar{\lambda}$ löst $\max\{L(\bar{x}, \lambda) : \lambda \geq 0\}$.
Da $L(\bar{x}, \cdot)$ linear in λ ist, lässt sich die Lösung direkt charakterisieren:
$\bar{\lambda}_i > 0 \Rightarrow g_i(\bar{x}) = 0$ (sonst $\bar{\lambda}$ verbesserbar)
$g_i(\bar{x}) < 0 \Rightarrow \bar{\lambda}_i = 0$ (sonst $\bar{\lambda}$ verbesserbar)
$g_i(\bar{x}) > 0$ ist unmöglich (dann $\notin \bar{\lambda}$)
$\Rightarrow$ 2.,3. □

Korollar 2.4.2. *Ist $(\bar{x}, \bar{\lambda})$ Sattelpunkt von L über $\mathbb{R}^n \times \mathbb{R}^m_+$, dann löst $\bar{x}$ das Optimierungsproblem für $X = \mathbb{R}^n$.*

Beweis. Nach Satz 2.4.1 ist $\bar{x}$ zulässig und aufgrund von 3. gilt

$$f(\bar{x}) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \; \forall x \in \mathbb{R}^n,$$

und für alle zulässigen x gilt wegen $\bar{\lambda} \geq 0$

$$L(x, \bar{\lambda}) = f(x) + \sum \underbrace{\lambda_i}_{\geq 0} \underbrace{g_i(x)}_{\leq 0} \leq f(x).$$

□

Damit es Sattelpunkte gibt, muss aber die algebraische Bechreibung $\bigcap_i \{x : g_i(x) \leq 0\}$ zur geometrischen passen:

Satz 2.4.3. *Für das konvexe Optimierungsproblem auf $X = \mathbb{R}^n$ sind äquivalent:*

1. *$(\bar{x}, \bar{\lambda})$ ist Sattelpunkt von L auf $\mathbb{R}^n \times \mathbb{R}^m_+$.*
2. *$\bar{x}$ löst das Optimierungsproblem und $\bar{\lambda}$ ist ein Lagrangemultiplikator im Sinne von (4.) aus Satz 2.2.20.*

Beweis. Die 1. Bedingung aus Satz 2.4.1 ist nach Satz 2.2.10 äquivalent zu $0 \in \partial(L(\cdot, \bar{\lambda}))(\bar{x})$ und dies nach Satz 2.2.20 äquivalent zu "$\bar{x}$ ist Optimallösung". □

Nach Satz 2.3.4 sind die Optimallösungen des dualen also gerade die Lagrangemultiplikatoren der primalen Optimallösungen.

Kapitel 3

Innere-Punkte-Verfahren

3.1 Motivation

Simplex-Verfahren läuft entlang dem Rand, besucht unter Umständen exponentiell viele Ecken.
Innere-Punkte-Verfahren: durch die Mitte (hoffentlich schneller).
Ansatz aus der nichtlinearen Optimierung: Starte im Inneren des zulässigen Bereiches und verhindere das Verlassen desselben durch eine Barrierefunktion.

Beispiel 3.1.1. *Für $x \geq 0$ nimmt man die Barrierefunktion $-\log x$.*
$\min_{a \leq x \leq b} cx$ *wird also das Barriereproblem* $\min(cx - \log(x-a) - \log(b-x))$ *zugeordnet.*

Die Barrierefunktion hält das Minimum vom Rand fern. $\rightarrow$ Verwende Barriereparameter $\mu > 0$, um Einfluss sukzessive abzuschwächen.

Barriere-Problem

(P)	$\min c^\tau x$ s.t. $Ax = b$ $x \geq 0$	(BP_μ)	$\min c^\tau x - \mu \sum \log x_i$ s.t. $Ax = b$ $x > 0$
(D)	$\max b^\tau y$ s.t. $A^\tau y + z = c$ $z \geq 0$	(BD_μ)	$\max b^\tau y + \mu \sum \log z_i$ s.t. $A^\tau y + z = c$ $z > 0$

Definition 3.1.1. *x heißt streng zulässig für (P), falls x zulässig für (P) und $x > 0$.*
(y, z) heißt streng zulässig für (D), falls (y, z) zulässig für (D) und $z > 0$.

Annahme 1. *Es existiert für (P) ein streng zulässiges x_0 und für (D) ein streng zulässiges (y_0, z_0).*

Weg: Bestimme Optimallösung von (BP_μ) für $\mu \to 0$. Für μ eng benachbart "gut" möglich.

Zuerst: festes μ. Die Optimallösung von (BP_μ) ist Sattelpunkt der Lagrangefunktion

$$\mathcal{L}_\mu(x, y) = c^\tau x - \mu \sum \log x_i + (b - Ax)^\tau y.$$

Sattelpunkt erfüllt KKT-Bedingungen:

$$\begin{aligned} \nabla_y \mathcal{L}_\mu(x, y) = 0 \text{ (optimal bzgl. } y) \qquad & b - Ax = 0 \text{ (primal zulässig)} \\ \nabla_x \mathcal{L}_\mu(x, y) = 0 \text{ (optimal bzgl. } x) \qquad & c - \underbrace{\mu x^{-1}}_{=z} - A^\tau y = 0 \text{ (dual zulässig)} \end{aligned}$$

Primal-duales KKT-System:

$$\begin{aligned} Ax = b, x > 0 \quad & \text{(primal zulässig)} \\ A^\tau y + z = c, z > 0 \quad & \text{(dual zulässig)} \\ x \circ z = \mu e \quad & \text{("perturbierte" Komplementarität)} \end{aligned}$$

mit $e = (1, \ldots, 1)^\tau$, $\quad x \circ z = (x_1 z_1, \ldots, x_n z_n)^\tau$.

Warum perturbierte Komplementarität? Für $\mu = 0$ ist $x \circ z = 0$, also $\langle x, z \rangle = 0 \Rightarrow$ optimal.

Man kommt auf das gleiche System von (BD_μ) aus. Zielfunktion von (BP_μ) ist streng konvex in x; die von (BD_μ) ist streng konkav in z. Lösung des KKT-Systems liefert eindeutige Optimallösung (x_μ, z_μ) von (BP_μ) und (BD_μ).
y spannt "nur" zulässigen z-Raum auf. y_μ ist eindeutig, falls A vollen Zeilenrang hat, und dann durch z_μ bestimmt. $\to$ Wichtig ist nur (x_μ, z_μ). Mittels dem Satz über implizite Funktionen kann man zeigen: $\{(x_\mu, z_\mu) : \mu > 0\}$ beschreibt eine glatte Kurve, die aus dem Inneren gegen eine ganz spezifische Optimallösung konvergiert.

Definition 3.1.2. *Ein Paar primaler und dualer Optimallösungen $x^\star$ und $(y^\star, z^\star)$ heißt streng komplementär, falls $\forall i$ gilt: $x_i^\star \neq 0$ oder $z_i^\star \neq 0$.*

Lemma 3.1.1. *Für $x', x'' \in \{x : Ax = b\}$; $z', z'' \in \{z : A^\tau y + z = c\}$ ist $(x' - x'')^\tau (z' - z'') = 0$ (d.h. die affinen Unterräume sind orthogonal).*

Beweis. $Ax' - Ax'' = 0$
$A^\tau(y' - y'') = -(z' - z'')$
$\Rightarrow -(z' - z'')^\tau (x' - x'') = ((y' - y'')^\tau A)(x' - x'') = (y' - y'')^\tau (A(x' - x'')) = 0.$

□

Lemma 3.1.2. *Für $\mu_k \downarrow 0$ konvergiert (x_{μ_k}, z_{μ_k}) gegen eine streng komplementäre Optimallösung $x^\star, z^\star)$ von (P) bzw. (D).*

Beweis. Folge bleibt beschränkt für $\mu < \bar{\mu}$:

$$0 = (x_\mu - x_0)^\tau (z_\mu - z_0) = \underbrace{x_\mu^\tau z_\mu}_{=n\mu,\ \text{da } x_\mu \circ z_\mu = \mu e} - x_0^\tau z_\mu - x_\mu^\tau z_0 + x_0^\tau z_0$$
$$\Rightarrow x_0^\tau z_\mu + x_\mu^\tau z_0 = n\mu + x_0^\tau z_0 \leq n\bar{\mu} + x_0^\tau z_0.$$

Wegen $x_0 > 0, z_0 > 0$ müssen x_μ und z_μ beschränkt bleiben.
Also konvergieren $x_\mu \to \bar{x} \geq 0$ und $z_\mu \to \bar{z} \geq 0$ und sind jeweils zulässig. Wegen $x_{\mu_k}^\tau z_{\mu_k} = n\mu_k \to 0$ (für $\mu_k \to 0$) folgt $\bar{x}^\tau \bar{z} = 0$, also ist $\bar{x} = x^\star$ und $\bar{z} = z^\star$ Optimallösung.
Strenge Komplementarität: Ersetze x_0 durch $x^\star$, z_0 durch $z^\star$:

$$\Rightarrow (x^\star)^\tau z_\mu + x_\mu^\tau (z^\star) = n\mu.$$

Aus $x_\mu \circ z_\mu = \mu e$ folgt

$$\tfrac{z_\mu}{\mu} = x_\mu^{-1}, \qquad \tfrac{x_\mu}{\mu} = z_\mu^{-1}, \sum_{i=1}^n \tfrac{x_i^\star}{(x_\mu)_i} + \sum_{i=1}^n \tfrac{z_i^\star}{(z_\mu)_i} = n.$$

Wegen $\frac{(z^\star)_i}{(z_\mu)_i} \to 0$ oder 1 und der Komplementarität (d.h. $x_i^\star$ und $z_i^\star$ können nicht beide $\neq 0$ sein) folgt die Behauptung. □

Verfahren:

- Suchen Nullstelle für $F_\mu(x, y, z) = \begin{pmatrix} Ax - b \\ A^\tau y + z - c \\ x \circ z - \mu e \end{pmatrix} = 0.$
- Newton: $F_\mu + \nabla F_\mu^\tau \cdot \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = 0$

$$\begin{array}{lllll} \text{I} & A\Delta x & = & b - Ax & =: f_p \\ \text{II} & A^\tau \Delta y + \Delta z & = & c - A^\tau y - z & =: f_d \\ \text{III} & \Delta x \circ z + x \circ \Delta z & = & \mu e - x \circ z & =: f_c \end{array}$$

Aus II: $\Delta z = f_d - A^\tau \Delta y$;
aus III: $\Delta x = \mu z^{-1} - x - z^{-1} \circ x \circ \Delta z = \mu z^{-1} - x - z^{-1} \circ x \circ f_d + z^{-1} \circ x \circ A^\tau \Delta y$.

Wegen $v \circ w = \text{diag}(v)w$ ergibt sich nach Einsetzen in I:

$$\underbrace{A\,\text{diag}(z^{-1} \circ x)A^\tau}_{=:M} \Delta y = f_p - A(\mu z^{-1} - x - z^{-1} \circ x \circ f_d).$$

$M \in \mathcal{S}_+^m$ ist positiv definit, falls A vollen Zeilenrang hat. Können Δy durch Cholesky-Zerlegung berechnen ⇒ höchstens $\frac{m^3}{3}$ flops ($\cong \mathcal{O}(\frac{m^3}{3})$).

Algorithmus 3.1.3 (Konzeptioneller Algorithmus). *Input:* A, b, c, x_0, y_0, z_0 *mit* $x_0 > 0, z_0 > 0$.

1. *Wähle* μ.
2. *Berechne* $\Delta x, \Delta y, \Delta z$ *wie oben.*
3. *Wähle Schrittlänge* $\alpha(\leq 1)$*, so dass* $x + \alpha\Delta x > 0, z + \alpha\Delta z > 0$.
4. *Update:* $x_+ = x + \alpha\Delta x, z_+ = z + \alpha\Delta z$.
5. *Falls* $\|f_p\|, \|f_d\|$ *und* $x^\tau z$ *klein genug, STOP. Sonst gehe zu 1.*

Bemerkung 3.1.4. *Wahl von* μ *beeinflusst die Güte der Schrittrichtung und Schrittlänge.*
μ wird langsam verkleinert: guter Schritt, aber langsamer Fortschritt.
μ wird schnell verkleinert: Schritt schlecht, $\rightarrow$ geringe Schrittlänge.

3.2 Algorithmus von Monteiro und Adler

- feasible short-step primal-dual path following algorithm
- beginnt und bleibt bei voller Schrittlänge 1 in einer Nachbarschaft des zentralen Pfades.

Nachbarschaft: (N) $\|x \circ z - \mu(x,z) \cdot e\| \leq \theta\mu(x,z)$ mit $0 < \theta < 1$ und $\mu(x,z) = \frac{x^\tau z}{n}$, so dass $\mu(x,z)e = \langle x \circ z, \frac{e}{\sqrt{n}}\rangle \cdot \frac{e}{\sqrt{n}}$ ("Projektion von $x \circ z$ auf $\frac{e}{\sqrt{n}}$")

Für einen Punkt (x, z) in der Nachbarschaft gilt:

$$(1-\theta)\mu(x,z) \leq x_i z_i \leq (1+\theta)\mu(x,z), \qquad \forall i = 1, \ldots, n.$$

Algorithmus 3.2.1. *Input:* A, b, c *streng zulässiger Startpunkt* (x^0, y^0, z^0) *der* (N) *erfüllt.*

1. *Wähle* $\sigma < 1$ *fest (später wie), setze k=0.*
2. $\mu_k = \frac{x^{k^\tau} z^k}{n}$
3. *Löse:*
$$\begin{aligned} A\Delta x &= 0 \\ A^\tau\Delta y + \Delta z &= 0 \\ \Delta x \circ z^k + x^k \circ \Delta z &= \sigma\mu_k e - x^k \circ z^k. \end{aligned}$$
4. $(x^{k+1}, y^{k+1}, z^{k+1}) = (x^k + \Delta x, y^k + \Delta y, z^k + \Delta z)$ *(kein line search!)*
5. *Falls* $(x^{k+1})^\tau z^{k+1} < 2^{-2L}$ *dann STOP.*
6. $k = k + 1$*; gehe zu 2.*

(L bezeichnet die Anzahl der bits, mit der das Lineare Programm kodiert wird.)

Schreiben ab jetzt (x, y, z) für (x^k, y^k, z^k) und (x_+, y_+, z_+) für $(x^{k+1}, y^{k+1}, z^{k+1})$.
Wegen $A\Delta x = 0$ erfüllt mit x auch x_+ die Bedingung $Ax_+ = b$, ebenso $A^\tau y_+ + z_+ = c$. Im KKT-System ist im Allgemeinen nicht erfüllt: pertubierte Komplementarität, da nicht linear (i.A. $\Delta x \circ \Delta z \neq 0$). Wir werden zeigen, dass (x_+, z_+) (N) erfüllt und damit $x_+ > 0, z_+ > 0$.

Lemma 3.2.2. *1.* $\Delta x^\tau \Delta z = 0$,

2. $\mu_+ = \frac{x_+ \circ z_+}{n} = \sigma\mu$,

3. $x_+ \circ z_+ = \mu_+ e + \Delta x \circ \Delta z$.

Beweis. 1. folgt direkt aus Lemma 3.1.1 (zulässige Mengen sind orthogonal)

$$\begin{aligned} x_+ \circ z_+ = (x + \Delta x) \circ (z + \Delta z) &= x \circ z + x \circ \Delta z + \Delta x \circ z + \Delta x \circ \Delta z \\ &= \sigma\mu e + \Delta x \circ \Delta z \\ n\mu_+ = x_+^\tau z_+ = e^\tau (x_+ \circ z_+) = e^\tau(\sigma\mu e + \Delta x \circ \Delta z) &= n\sigma\mu + \underbrace{\Delta x^\tau \Delta z}_{=0} = n\sigma\mu \end{aligned}$$

$\Rightarrow \sigma\mu = \mu_+$.

□

Wir müssen nur noch zeigen $\|\Delta x \circ \Delta z\| \leq \theta\mu(x, z)$, falls das gilt, ist auch $x_+ > 0, z_+ > 0$ erfüllt, folgt aus folgendem Lemma.

Lemma 3.2.3. $x(\alpha) = x + \alpha\Delta x, z(\alpha) = z + \alpha\Delta z$. *Ist* (N) *für* $\alpha = 0$ *und* $\alpha = 1$ *erfüllt, dann ist* (N) *auch für* $\alpha \in [0, 1]$ *erfüllt.*

Beweis. Übung. □

Lemma 3.2.4. *Sei* $x > 0, z > 0$ *und* $h = \Delta x \circ z + x \circ \Delta z$. *Für* $d = x^{\frac{1}{2}} \circ z^{-\frac{1}{2}}$ *ist*

$$\|d^{-1} \circ \Delta x\|^2 + \|d \circ \Delta z\|^2 + 2\Delta x^\tau \Delta z = \|x^{-\frac{1}{2}} \circ z^{-\frac{1}{2}} \circ h\|^2$$

Beweis. $\Delta x \circ z + \Delta z \circ x = h \quad | \cdot x^{-\frac{1}{2}} \circ z^{-\frac{1}{2}}$

$$\Delta x \circ \underbrace{x^{-\frac{1}{2}} \circ z^{\frac{1}{2}}}_{d^{-1}} + \Delta z \circ \underbrace{x^{\frac{1}{2}} \circ z^{-\frac{1}{2}}}_{d} = h \circ x^{-\frac{1}{2}} \circ z^{-\frac{1}{2}}$$

.

Berechne davon die quadratische Norm. □

Lemma 3.2.5. *Sei* $\gamma = \min\{x_i z_i : i = 1, \ldots, n\}$. *Dann ist* $\|\Delta x \circ \Delta z\| \leq \frac{\|x \circ z - \mu_+ e\|^2}{2\gamma}$.

Beweis. Sei d wie in Lemma 3.2.4.

$$\begin{aligned} \|\Delta x \circ \Delta z\| &= \|\Delta x \circ d^{-1} \circ d \circ \Delta z\| \\ &\leq \|d^{-1} \circ \Delta x\| \cdot \|d \circ \Delta z\| \\ &\leq \tfrac{1}{2}(\|d^{-1} \circ \Delta x\|^2 + \|d \circ \Delta z\|^2) \\ &\overset{L.3.2.4}{=} \tfrac{1}{2}\|x^{-\frac{1}{2}} \circ z^{-\frac{1}{2}} \circ (\mu_+ e - x \circ z)\|^2 \\ &\leq \tfrac{1}{2}\|\gamma^{-\frac{1}{2}}(\mu_+ e - x \circ z)\|^2. \end{aligned}$$

□

Satz 3.2.6. *Werden* $0 < \theta < 1$ *und* $0 < \sigma < 1$ *so gewählt, dass*

$$\frac{\theta^2+n(1-\sigma)^2}{2(1-\theta)} \leq \theta\sigma, \qquad (\star)$$

dann ist $\|x_+ \circ z_+ - \mu_+ e\| \leq \theta\mu_+$ *in jeder Iteration erfüllt.*

Beweis. Wegen $(x \circ z - \mu(x,z)e)^\tau e = 0$ ist

$$\begin{aligned}\|x \circ z - \mu_+ e\|^2 &= \|x \circ z - \mu e\|^2 + \|\mu e - \mu_+ e\|^2 \text{ (Pythagoras)}\\ &\leq (\theta\mu)^2 + \mu^2(1-\sigma)^2\|e\|^2\\ &= (\theta^2 + n(1-\sigma)^2)\mu^2.\end{aligned}$$

Wegen (x,z) in (N) gilt für γ aus Lemma 3.2.5: $(\gamma \geq (1-\theta)\mu)$

$$\|x_+ \circ z_+ - \mu_+ e\| \overset{L.3.2.5}{\leq} \tfrac{\|x\circ z-\mu_+ e\|^2}{2\gamma} \leq \underbrace{\frac{\theta^2+n(1-\sigma)^2}{2(1-\theta)}}_{\leq\theta\sigma} \leq \theta \underbrace{\sigma\mu}_{\mu_+} = \theta\mu_+.$$

□

Für θ konstant und unabhängig von n muss σ die Form $\sigma \sim 1 - \frac{\delta}{n}$ haben, damit $(\star)$ für beliebiges n gilt. Die Bedingung $(\star)$ ist für $\theta = \delta = 0.35$ erfüllt.

Satz 3.2.7. *Mit* θ *und* $\sigma = 1 - \frac{\delta}{\sqrt{n}}$ *so, dass* $(\star)$ *erfüllt ist, und einem streng zulässigen Startpunkt* (x^0, y^0, z^0)*, der* (N) *und* $x^{0\tau} z^0 \leq 2^L$ *erfüllt, terminiert der Algorithmus in* $\mathcal{O}(\sqrt{n}L)$ *Iterationen.*

Beweis. $(x^k)^\tau(z^k) = n\mu_k \overset{L.3.2.2}{=} n\sigma^k\mu_0$ muss $\leq 2^{-2L}$ werden . Kleinstes k mit $\sigma^k \leq \frac{2^{-2L}}{n\mu_0}$:

$$\begin{aligned}&k\log\sigma = k\log(1-\tfrac{\delta}{\sqrt{n}}) \leq -k\tfrac{\delta}{\sqrt{n}} \qquad (\ \log(1+x) \leq x\ )\\ &\Longrightarrow \bar{k} \geq \tfrac{\sqrt{n}}{\delta}[2L\log 2 + \log n\mu_0] = \mathcal{O}(\sqrt{n}L).\end{aligned}$$

□

Pro Iteration ist der Aufwand $\mathcal{O}(m^3)$, meist ist $m = \mathcal{O}(n)$. ⇒ Gesamtaufwand $\mathcal{O}(n^{3.5}L)$ arithmetische Operationen. Theoretisch gibt es sogar $\mathcal{O}(n^3 L)$.

3.3 Zentrierter Startpunkt (keine Annahmen über Zulässigkeit)

Betrachten zuerst das folgende homogene schiefsymmetrische System:

$$\begin{array}{cccccc} & Ax & -\tau b & & = 0 & \\ -A^\tau y & & +\tau c & -z & = 0 & \\ b^\tau y & -c^\tau x & & -\rho & = 0 & \\ x \geq 0, & \tau \geq 0, & z \geq 0, & \rho \geq 0 & & \end{array} \qquad \text{(HS)}$$

- System ist zulässig: setze alle Variablen auf 0.
- Falls es eine Lösung mit $\tau > 0$ gibt, $\Rightarrow \frac{x}{\tau}, \frac{y}{\tau}, \frac{z}{\tau}$ ist primal/dual zulässig und 3. Zeile garantiert $b^\tau \frac{y}{\tau} \geq c^\tau \frac{x}{\tau}$. Mit schwacher Dualität $b^\tau \frac{y}{\tau} \leq c^\tau \frac{x}{\tau}$ folgt Optimalität.

Problem: Falls $\tau > 0$ muss $\rho = 0$ gelten $\Rightarrow \nexists$ streng zulässiger Punkt.

Wir brauchen dafür noch zwei Variablen (ϑ, σ) und zwei Konstanten (α, β) und betrachten die "schiefsymmetrische Einbettung":

$$
\begin{array}{lrrrrrrl}
\min\ \beta\vartheta & & & & & & & \\
\text{s.t.} & Ax & -\tau b & +\vartheta\bar{b} & & & = & 0 \\
-A^\tau y & & +\tau c & -\vartheta\bar{c} & -z & & = & 0 \\
b^\tau y & -c^\tau x & & +\vartheta\alpha & & -\rho & = & 0 \\
-\bar{b}^\tau y & +\bar{c}^\tau x & -\tau\alpha & & & & = & -\beta \\
z, x, \tau, \vartheta, \rho, \sigma \geq 0. & & & & & & &
\end{array}
$$

Konstanten $\alpha, \beta, \bar{b}, \bar{c}$ erhalten wir für $x^0 = e, y^0 = 0, z^0 = e, \tau^0 = \vartheta^0 = \rho^0 = \sigma^0 = 1$:

$$
\begin{aligned}
\bar{b} &= b - Ae \\
\bar{c} &= c - e \\
\alpha &= c^\tau e + 1 \\
\beta &= -\bar{c}^\tau e + c^\tau e + 1 + 1 = e^\tau e + 2 = n + 2
\end{aligned}
$$

Das besondere an dem Programm: *Es ist selbstdual!* (Das Duale ist wieder das Programm selbst.)

Multiplikatoren: 1. Zeile $\tilde{y}$, 2. Zeile $\tilde{x}$, 3. Zeile $\tilde{\tau}$, 4. Zeile $\tilde{\vartheta}$
$\longrightarrow$ Duales hat den gleichen Startpunkt $\tilde{x}^0 = e, \tilde{y}^0 = 0, \tilde{z}^0 = e, \tilde{\tau}^0 = \tilde{\vartheta}^0 = \tilde{\rho}^0 = \tilde{\sigma}^0 = 1$.

Komplementaritätsbedingungen sind: $x\tilde{z} = \mu e, \quad \tilde{x}z = \mu e$
$\tau\tilde{\rho} = \mu, \quad \tilde{\tau}\rho = \mu$
$\vartheta\tilde{\sigma} = \mu, \quad \tilde{\vartheta}\sigma = \mu.$

Für $\mu = 1$ ist der Startpunkt direkt auf dem zentralen Pfad. Der Innere-Punkte-Algorithmus wird beide Variablengruppen genau gleich aktualisieren: $x^k \equiv \tilde{x}^k, y^k \equiv \tilde{y}^k, \ldots \Rightarrow$ brauchen nur eine Variablen- und Gleichungsgruppe (nur primal) $\Rightarrow$ zu lösendes Problem hat nur 2 Zeilen mehr! Lemma 3.1.2 $\Rightarrow$ zentraler Pfad konvergiert gegen eine streng komplementäre Lösung. Diese hilft uns die ursprüngliche Lösung zu rekonstruieren.

Satz 3.3.1. *Das selbstduale Programm hat eine Optimallösung* $(x^\star, y^\star, z^\star, \tau^\star, \rho^\star, \vartheta^\star, \sigma^\star)$ *und* $\tau^\star > 0$ *oder* $\rho^\star > 0$*. Es gilt:*

i) $\tau^\star > 0 \Longleftrightarrow$ *(P) und (D) sind zulässig mit Optimallösung* $\frac{x^\star}{\tau^\star}$ *und* $\left(\frac{y^\star}{\tau^\star}, \frac{z^\star}{\tau^\star}\right)$.

ii) $\tau^\star = 0 \Longleftrightarrow$ *(P) oder (D) hat einen verbessernden Halbstrahl.*

Beweis. Existenz einer Lösung mit $\tau^\star > 0$ oder $p^\star > 0$ folgt aus Lemma 3.2.3.
Aus der Optimallösung folgt $\vartheta = 0$ (setze $\sigma = \beta$, alle anderen auf 0).
$\Rightarrow$ Zeilen 1-3 ergeben das homogene System (HS)

Aus $\tau^\star > 0 \Rightarrow \frac{x^\star}{\tau^\star}$ und $\left(\frac{y^\star}{\tau^\star}, \frac{z^\star}{\tau^\star}\right)$ sind primal/dual zulässig und optimal.

Aus $\tau^\star = 0 \Rightarrow \rho^\star > 0$ (strenge Komplementarität); also ist $Ax^\star = 0, A^\tau y^\star + z^\star = 0$ und $b^\tau y^\star > c^\tau x^\star$
$\Rightarrow b^\tau y^\star > 0$ oder $c^\tau x^\star < 0$
$\Rightarrow$ eines der beiden Probleme hat einen verbessernden Halbstrahl
$\Rightarrow$ mindestens ein Problem ist unzulässig.

Sind $\bar{x}$ und $(\bar{y}, \bar{z})$ primale und duale Optimallösungen, setze $\vartheta^\star = 0, \rho^\star = 0, x^\star = \tau^\star \bar{x}, y^\star = \tau^\star \bar{y}, z^\star = \tau^\star \bar{z}$.
Zeile 1-3 gelten für $\tau^\star > 0$ und

$$\begin{aligned}
&\bar{b}^\tau(\tau^\star \bar{y}) - \bar{c}^\tau(\tau^\star \bar{x}) + \tau^\star \alpha + \sigma^\star = \beta \\
&\tau^\star(-e^\tau \underbrace{A^\tau \bar{y}}_{c-\bar{z}} + \underbrace{b^\tau \bar{y} - c^\tau \bar{x}}_{=0} + e^\tau \bar{x} + c^\tau e + 1) + \sigma^\star = \tau^\star(e^\tau(\bar{x} + \bar{z}) + 1) + \sigma^\star = n + 2 \\
&\Longrightarrow \tau^\star \underbrace{(e^\tau(\bar{x} + \bar{z}) + 1)}_{>0} = n + 2 - \sigma^\star,
\end{aligned}$$

also zulässig für $\tau^\star \leq \frac{n+2}{e^\tau(\bar{x}+\bar{z})+1}$.

zeigen primalen Halbstrahl $A\bar{x} = 0, c^\tau \bar{x} < 0$ mit $\bar{x}$ so, dass $\bar{c}^\tau \bar{x} = c^\tau \bar{x} - e^\tau \bar{x} \geq -\beta$. Setze

$$\bar{x} = x^\star, y^\star = 0, z^\star = 0, \tau^\star = 0, \vartheta^\star = 0, \rho^\star = \underbrace{-c^\tau \bar{x}}_{>0}, \sigma^\star = \bar{c}^\tau \bar{x} + \beta$$

analog für dualen Halbstrahl. □

Lösung des schiefsymetrischen Systems mit Inneren-Punkte-Verfahren liefert
- entweder Optimallösungen
- oder Nachweis, dass wenigstens eines der beiden Probleme unzulässig ist.

In der Praxis verzichtet man auf ϑ und σ und arbeitet "unzulässig".

3.4 Quadratische Optimierung

Sei $Q \in \mathcal{S}_+^n, c \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

$$\begin{array}{lcl}
\min \frac{1}{2} x^\tau Q x + c^\tau x & \text{Lagrange-} & \min \frac{1}{2} x^\tau Q x + c^\tau x + (b - Ax)^\tau y \\
\text{s.t. } Ax \geq b & \longleftrightarrow & \text{s.t. } Qx + c - A^\tau y = 0 \\
x \text{ frei} & \text{dualität} & y \geq 0
\end{array}$$

Starke Dualität folgt aus Beobachtung Beobachtung 2.2.21 , Satz 2.2.20 und Satz 2.4.3.

Barriere-Problem: $\min(\frac{1}{2}x^\tau Qx + c^\tau x - \mu \sum \log(A_{i,\cdot}x - b_i)) =: f(x)$

$\nabla f(x) = 0 : \qquad Qx + c - \mu A^\tau \left(\frac{1}{A_{i,\cdot}x - b_i}\right) = 0$

$s := Ax - b$ "Schlupf-Variable", $y := \mu s^{-1}$ "Dualvariable".

Primal-Duales KKT-System:

$Qx + c - A^\tau y = 0$	duale Zulässigkeit
$Ax - s = b$	primale Zulässigkeit
$s \circ y = \mu e$	pertubierte Komplementarität.

Newton-Schritt:

$$\begin{aligned} Q\Delta x - A^\tau \Delta y &= -(Qx + c - A^\tau y) \ , \\ A\Delta x - \Delta s &= b - Ax + s \\ \Delta s \circ y + s \circ \Delta y &= \mu e - s \circ y \end{aligned}$$

kann mit ähnlicher Strategie in $\mathcal{O}(n^{3.5}L)$ bzw. $\mathcal{O}(n^3 L)$ gelöst werden.

Beispiel: Portfolio-Optimierung (Markowitz-Modell)
- $i = 1, \ldots, n$ mögliche Investitionen in Aktien
- x_i - Anteil, der in Aktie i investiert wird. -erwarteter Gewinn (Verlust): w_i
-Kovarianz-Matrix Q
(aus Vergangenheitsdaten oder "Expertenwissen")

Finde Investition mit minimalem Risiko, deren erwarteter Gewinn einen Mindestwert $\bar{w} \in \mathbb{R}_+$ erreicht:

$\min \frac{1}{2}x^\tau Qx$	"Risikomaß"
s.t. $w^\tau x \geq \bar{w}$	"erwarteter Gewinn"
$e^\tau x = 1$	"Anteil am Gesamtkapital"
$x \geq 0.$	

Beispiel: SQP-Verfahren (sequential quadratic programming)
Löse für $f, h \in \mathbb{C}^2$ nichtlineares Optimierungsproblem

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } h_i(x) \leq 0, \qquad i = 1, \ldots, m \end{aligned}$$

indem Schrittrichtung Δx jeweils durch quadratisches Unterproblem bestimmt wird: Zur Herleitung betrachte Barriere-Problem $\min\limits_x (f(x) - \mu \sum \log(-h_i(x)))$.

$\nabla = 0 \Rightarrow \nabla f(x) + \mu \sum \nabla h_i(x) \frac{1}{-h_i(x)} = 0.$

Setze $s_i = -h_i(x) \geq 0, y = \mu s^{-1}$:

I	$\nabla f(x) + \sum \nabla h_i(x) y_i = 0$
II	$h_i(x) + s_i = 0, \quad i = 1, ..., m$
III	$s \circ y = \mu e$

Newton:

I	$[\nabla^2 f(x) + \sum y_i \nabla^2 h_i(x)]\Delta x + \sum \nabla h_i(x) \Delta y_i = -[\nabla f(x) + \sum y_i \nabla h_i(x)]$
II	$\nabla h_i(x)^\tau \Delta x + \Delta s_i = (-s_i + h_i(x)), \quad i = 1, ...m$

Setze $A := \begin{bmatrix} -\nabla h_1(x)^\tau \\ \vdots \\ -\nabla h_m(x)^\tau \end{bmatrix}, Q := [\nabla^2 f(x) + \sum y_i \nabla^2 h_i(x)], c := \nabla f(x), b := -h(x)$

I $\quad Q\Delta x + c - A^\tau(y + \Delta y) = 0$
II $\quad A\Delta x - (s + \Delta s) = b$
III $\quad (s + \Delta s) \circ (y + \Delta y) = \mu e$

Mit $\bar{x} := \Delta x, \bar{s} := s + \Delta s, \bar{y} := y + \Delta y$ folgt

I $\quad Q\bar{x} + c - A^\tau \bar{y} = 0$
II $\quad A\bar{x} - \bar{s} = b$
III $\quad \bar{s} \circ \bar{y} = \mu e,$

was gerade die Optimalitätsbedingungen für Barriereproblem zu

$$\begin{array}{ll} \min \frac{1}{2}\bar{x}^\tau Q\bar{x} + c^\tau \bar{x} & \Leftrightarrow \quad \min \frac{1}{2}\Delta x^\tau(\nabla^2 f(x) + \sum y_i \nabla^2 h_i(x))\Delta x + \nabla f(x)^\tau \Delta x \\ \text{s.t. } A\bar{x} \geq b & \quad\ \text{s.t. } h(x) + J(h(x))\Delta x \leq 0 \end{array}$$

sind. Also: Nebenbedingungen werden linearisiert, aber der quadratische Term der Kostenfunktion berücksichtigt die Krümmung von f und relevanter $h_i \Rightarrow$ kleiner Schrittanteil in Richtungen, in denen sich diese Funktionen stark ändern. Fortsetzen mit line search in Richtung Δx (Schrittlänge ≤ 1).

3.5 Lineare Optimierung über dem quadratischen Kegel

(second order cone programming)

$$\begin{array}{lll} \min & c^\tau x & \\ \text{s.t.} & Ax = b & \\ & x = \begin{pmatrix} x_0 \\ \bar{x} \end{pmatrix}, \quad x_0 \geq \|\bar{x}\| \end{array} \quad \begin{array}{c} \text{Lagrange-} \\ \longleftrightarrow \\ \text{dualität} \end{array} \quad \begin{array}{lll} \max & b^\tau y & \\ & A^\tau y + z = c & \\ \text{s.t.} & z = \begin{pmatrix} z_0 \\ \bar{z} \end{pmatrix}, z_0 \geq \|\bar{z}\| \end{array}$$

Starke Dualität ist garantiert, wenn es streng zulässige primale und duale Punkte gibt, d.h. falls $\exists \tilde{x} >_Q 0, A\tilde{x} = b$ und $\tilde{z} >_Q 0, \exists \tilde{y} : A^\tau \tilde{y} + \tilde{z} = c$ (ohne Beweis).

Barriere-Problem: $\begin{array}{ll} \min & c^\tau x - \frac{1}{2}\mu \log(x_0^2 - \|\bar{x}\|^2) \\ \text{s.t.} & Ax = b \\ & x >_Q 0 \end{array}$

$$\mathcal{L}(x, y) = c^\tau x - \tfrac{1}{2}\mu \log(x_0^2 - \|\bar{x}\|^2) + y^\tau(b - Ax)$$

$$\nabla_x \mathcal{L} = 0 : c - \tfrac{1}{2}\mu \tfrac{1}{x_0^2 - \|\bar{x}\|^2} \begin{bmatrix} 2x_0 \\ -2\bar{x} \end{bmatrix} - A^\tau y = 0$$
$$\nabla_y \mathcal{L} = 0 : Ax = b$$

Setze $z = \frac{\mu}{x_0^2 - \|\bar{x}\|^2} \begin{bmatrix} x_0 \\ -\bar{x} \end{bmatrix}$. z löst $\underbrace{\begin{bmatrix} x_0 & \bar{x}^\tau \\ \bar{x} & x_0 I \end{bmatrix}}_{=:\mathrm{Arw}(x)} z = \mu \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

$\mathrm{Arw}(x)$ ist p.d. $\Leftrightarrow x_0 > 0$ und $x_0 > \frac{1}{x_0}\bar{x}^\tau \bar{x}$ (Schurkomplement; siehe Übung) $\Leftrightarrow x \in \mathrm{int}\, Q_+^n$; in diesem Fall ist z eindeutig.

Primal-duales KKT-System:
$$\begin{array}{ll} c - A^\tau y - z = 0, & z \in \mathrm{int} Q_+^n \\ Ax = b, & x \in \mathrm{int} Q_+^n \\ \mathrm{Arw}(x) z = \mu e_0. & \end{array}$$

Linearisieren mit Newton, ... für einen Kegel in konstanter Zeit (auch direkt geometrisch).
Nur in Kombination mit anderen Kegeln sinnvoll.

3.6 Semidefinite Optimierung

Betrachten
$$\begin{array}{l} \min \langle C, X \rangle \\ \text{s.t. } \mathcal{A}X = b \\ \quad X \succcurlyeq 0. \end{array}$$

Dabei ist $\mathcal{A}X = \begin{bmatrix} \langle \mathcal{A}_1, X \rangle \\ \vdots \\ \langle \mathcal{A}_m, X \rangle \end{bmatrix}$.

Für Dualaufgabe wird adjungierter Operator zu $\mathcal{A}$ benötigt:

$$\forall X \in \mathcal{S}^n, \forall y \in \mathbb{R}^m : \langle \mathcal{A}X, y \rangle = \sum y_i \langle \mathcal{A}_i, X \rangle = \langle \sum y_i \mathcal{A}_i, X \rangle = \langle \mathcal{A}^\tau y, X \rangle$$

Anfangs leichter vorstellbar:

$$x = \mathrm{vec}(X) = (x_{1,1}, x_{2,1}, \cdots, x_{n,1}, x_{1,2}, x_{2,2}, \cdots, x_{n,n})^\tau$$
$$\mathcal{A}X = \underbrace{\begin{bmatrix} \mathrm{vec}(\mathcal{A}_1)^\tau \\ \mathrm{vec}(\mathcal{A}_2)^\tau \\ \vdots \end{bmatrix}}_{=:A} \mathrm{vec}(X)$$
$$\mathcal{A}^\tau y = \mathrm{vec}^{-1}(A^\tau y) = \mathrm{vec}^{-1}(\sum y_i \mathrm{vec}(\mathcal{A}_i)).$$

Lagrange-Duales:
$$\begin{array}{l} \max b^\tau y \\ \text{s.t. } \mathcal{A}^\tau y + Z = C \\ \quad Z \succcurlyeq 0 \end{array}$$

Z ist Slackmatrix.
Starke Dualität gilt, falls $\exists X \succ 0 : \mathcal{A}X = b$ und $\exists Z \succ 0 : \exists y : \mathcal{A}^\tau y + Z = C$; dann werden primale und duale Optimalwerte angenommen und die Optimalwerte sind gleich (ohne Beweis).

Semidefinite Programmierung enthält:

- lineare Optimierung
$$X = \begin{bmatrix} x_1 & & & 0 \\ & x_2 & & \\ & & \ddots & \\ 0 & & & x_n \end{bmatrix}$$
- konvexe quadratische Optimierung
$$x^\tau Q x + b^\tau x + d \leq 0 \Longleftrightarrow \begin{bmatrix} I & Q^{\frac{1}{2}}x \\ x^\tau Q^{\frac{1}{2}} & -(b^\tau x + d) \end{bmatrix} \succcurlyeq 0$$
- lineare Optimierung über dem quadratischen Kegel
$$x_0 \geq \|\bar{x}\| \Leftrightarrow \mathrm{Arw}(x) = \begin{bmatrix} x_0 & \bar{x}^\tau \\ \bar{x} & x_0^\tau I \end{bmatrix}$$
- Optimierung über mehrere semidefinite Variablen
$$X_1 \succcurlyeq 0, X_2 \succcurlyeq 0, \ldots, X_n \succcurlyeq 0 \Longleftrightarrow \mathcal{X} = \begin{bmatrix} X_1 & & & 0 \\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_n \end{bmatrix}$$

Was ist eine gute Barrierefunktion für SDP? $-\mu \log\det(X)$. Man kann zeigen:

- $-\log\det(X)$ ist streng konvex auf int$\mathcal{S}^n_{++}$
- $\nabla(-\log\det(X)) = \mathrm{vec}(X^{-1})$

Lagrangefunktion:

$$\begin{aligned} &\mathcal{L}(X,y) = \langle C, X\rangle - \mu\log\det(X) + y^\tau(b - \mathcal{A}X) \\ &\nabla_X\mathcal{L} = 0 : C - \mu X^{-1} - \mathcal{A}^\tau y = 0, \quad \Rightarrow Z = \mu X^{-1} \\ &\nabla_y\mathcal{L} = 0 : b - \mathcal{A}X = 0 \end{aligned}$$

primal-duales KKT-System:	$C - Z - \mathcal{A}^\tau y = 0$	primal zulässig
	$b - \mathcal{A}X = 0$	dual zulässig
	$XZ = \mu I$	pertubierte Komplementarität.

Differenz von primalem und dualem Optimalwert soll 0 sein, d.h. $\langle X, Z\rangle = 0$.

$$\begin{aligned} &X = \textstyle\sum \lambda_i(X) v_i^X (v_i^X)^\tau, Z = \sum \lambda_i(Z) v_i^Z (v_i^Z)^\tau \\ &\langle X, Z\rangle = \textstyle\sum \lambda_i(X)\lambda_i(Z)\langle v_i^X(v_i^X)^\tau, v_i^Z(v_i^Z)^\tau\rangle = 0 \\ &\Leftrightarrow XZ = 0. \end{aligned}$$

Linearisieren: $\mathcal{A}^\tau \Delta y + \Delta Z = C - \mathcal{A}^\tau y - Z$
$\mathcal{A}\Delta X = b - \mathcal{A}X$
$\Delta X\, Z + X\, \Delta Z = \mu I.$

Problem: im allgemeinen ist $\Delta X \notin \mathcal{S}^n$!
Man kann zeigen, dass Symmetrisieren durch $\overline{\Delta X} = \frac{\Delta X + \Delta X^\tau}{2}$ ausreicht.

$$(N) \quad \|\tfrac{1}{2}(Z^{\frac{1}{2}} X Z^{-\frac{1}{2}} + Z^{-\frac{1}{2}} X Z^{-\frac{1}{2}}) - \mu(X,Z) I\| \leq \theta \mu(X,Z), \quad 0 < \theta < 1$$
$$\mu(X,Z) = \tfrac{\langle X, Z\rangle}{n}$$

$\mathcal{O}(\sqrt{n} \log\left(\frac{\langle X^0, Z^0\rangle}{\varepsilon}\right))$ Iterationen nötig, um $\langle X^k, Z^k\rangle \leq \varepsilon$ zu erhalten.

- Es ist noch nicht klar, ob SDP polynomial gelöst werden können.
- Schiefsymmetrische Einbettung funktioniert auch.

Beispiel: Robuste Steuerung
Wollen feststellen, ob ein dynamisches System $\frac{\mathrm{d}x}{\mathrm{d}t} = A(t)x$ in einen stabilen Zustand kommen kann. Über $A(t)$ weiß man nur, dass $A(t) \in \operatorname{conv}\{A_1, \ldots, A_k\}$.
Das System heißt stabil, wenn alle Trajektorien gegen 0 gehen (Trajektorien sind Kurven, die der Differentialgleichung genügen). Stimmt sicher, falls es eine Norm $\|x\|_H = \sqrt{x^\tau H x}$ gibt mit $H \succ 0$ und $\frac{\mathrm{d}\|x(t)\|_H^2}{\mathrm{d}t} \leq \delta < 0$. Gibt es so ein H, dann heißt das System quadratisch stabil und $x^\tau H x$ wird Lyapunov-Funktion genannt.

$$\tfrac{\mathrm{d}}{\mathrm{d}t} x^\tau H x = \left(\tfrac{\mathrm{d}x}{\mathrm{d}t}\right)^\tau H x + x^\tau H \left(\tfrac{\mathrm{d}x}{\mathrm{d}t}\right) = x^\tau \underbrace{\left(A(t)^\tau H + H A(t)\right)}_{\text{negativ definit}} x \leq \delta < 0 \quad \forall x.$$

Wegen $A(t) \in \operatorname{conv}\{A_1, \ldots, A_k\}$ müssen wir ein $H \succ 0$ finden mit $A_i^\tau H + H A_i \prec 0$ für $i = 1, \ldots, k$, also:

$$\begin{aligned} &\max \lambda \\ &\text{s.t. } H \succcurlyeq \lambda I \\ &\qquad A_i^\tau H + H A_i \preccurlyeq -\lambda I. \end{aligned}$$

Gut vorstellbar in dualer Form: $y = (\lambda, h_{1,1}, \ldots, h_{1,n}, h_{2,1}, \ldots, h_{n,n})^\tau$.

$$\begin{aligned} &\max \lambda = e_0^\tau y \\ &\text{s.t. } \underbrace{\begin{bmatrix} \lambda I - H & & & \\ & A_1^\tau H + H A_1 & & \\ & & \ddots & \\ & & & A_k^\tau H + H A_k \end{bmatrix}}_{\text{linear in } y} + \begin{bmatrix} Z_0 & & & \\ & Z_1 & & \\ & & \ddots & \\ & & & Z_k \end{bmatrix} = 0 \end{aligned}$$

Falls $\lambda_k > 0$, dann erzeugt $H^\star$ die gesuchte Lyapunov-Funktion.

Kapitel 4

Nichtglatte Optimierungsverfahren

4.1 Das Subgradientenverfahren

Betrachten $\min f(x)$
$\text{s.t.} x \in C.$
Dabei sei $C \subseteq \mathbb{R}^n$ konvex und abgeschlossen; f konvex und Lipschitzstetig auf einer Umgebung von C mit Konstante M.

C sei von einfacher Struktur, so dass die Projektion $P_C(x)$ leicht zu berechnen ist (z.B. Würfel, Kugel, einfacher Kegel). f ist über ein sogenanntes Orakel 1. Ordnung gegeben. Das Orakel liefert für $x \in C$: $f(x)$ und einen Subgradienten $g(x) \in \partial f(x)$, egal welchen.

Annahme: Es existiert eine Optimallösung $x^\star \in C$. Also gilt $\forall\, x \in C$:

$$\begin{aligned} f(x^\star) &\geq f(x) + \langle g(x), x^\star - x\rangle \\ \Rightarrow 0 \leq f(x) - f(x^\star) &\leq \langle g(x), x - x^\star\rangle. \end{aligned}$$

Also zeigt $-g(x)$ "in Richtung" von $x^\star$. Leider gibt $\|g(x)\|$ keinen guten Hinweis auf die Schrittlänge. Notbehelf: Wir wählen Schrittlängen h_k schon im Voraus.

Algorithmus 4.1.1 (Subgradienten-Verfahren). *0. Wähle $x_0 \in C$ und eine Folge $\{h_k\}$, $h_k \geq 0$ mit $\sum_{k=0}^{\infty} h_k \to \infty$, $\sum_{k=0}^{\infty} h_k^2 < \infty$. Setze $k = 0$.*

1. *Rufe Orakel für $x_k \mapsto f(x_k), g(x_k)$.*
2. $x_{k+1} = P_C\left(x_k - h_k \frac{g(x_k)}{\|g(x_k)\|}\right)$.
3. $k = k + 1$, *gehe zu 1.*

Satz 4.1.2. *Sei f konvex, Lipschitzstetig auf einer Umgebung von C mit Konstante M und $x_0 \in C$ mit $r_0 = \|x_0 - x^\star\|$. Dann gilt:*

$$\min_{i=0,\dots,k} f(x_i) - f(x^\star) \leq M \frac{r_0^2 + \sum_{i=0}^{k} h_i^2}{2\sum_{i=0}^{k} h_i} \quad \begin{matrix}(<\infty)\\(\to\infty)\end{matrix}$$

Beweis. Sei $r_i = \|x_i - x^\star\|$. Wegen Beobachtung 2.1.16 und Cauchy-Schwarz-Ungleichung gilt $\|P_C(x) - P_C(y)\| \leq \|x - y\|$, also

$$\begin{aligned} r_{i+1}^2 = \left\| P_C\left(x_i - h_i \tfrac{g(x_i)}{\|g(x_i)\|}\right) - x^\star \right\|^2 &\leq \left\| x_i - h_i \tfrac{g(x_i)}{\|g(x_i)\|} - x^\star \right\|^2 \\ &= \underbrace{\|x_i - x^\star\|}_{=r_i^2} - 2h_i \left\langle \tfrac{g(x_i)}{\|g(x_i)\|}, x_i - x^\star \right\rangle + h_i^2, \end{aligned}$$

also $r_i^2 + h_i^2 \geq r_{i+1}^2 + 2h_i \left\langle \frac{g(x_i)}{\|g(x_i)\|}, x_i - x^\star \right\rangle$ für $i = 0, \dots, k$

$$r_0^2 + \sum_{i=0}^{k} h_i^2 \geq r_{k+1}^2 + \sum_{i=0}^{k} 2h_i \underbrace{\left\langle \frac{g(x_i)}{\|g(x_i)\|}, x_i - x^\star \right\rangle}_{\geq 0} \geq \min_{i=1,\dots,k} \left\langle \tfrac{g(x_i)}{\|g(x_i)\|}, x_i - x^\star \right\rangle \cdot \sum_{i=0}^{k} 2h_i$$

Minimum werde für $\bar{\imath}$ angenommen.

$$\Rightarrow f(x_{\bar{\imath}}) - f(x^\star) \leq \langle g(x_{\bar{\imath}}), x_{\bar{\imath}} - x^\star \rangle \leq \underbrace{\|g(x_{\bar{\imath}})\|}_{\leq M \text{ wegen Lipschitzstetigkeit}} \frac{r_0^2 + \sum_{i=0}^{k} h_i^2}{\sum_{i=0}^{k} 2h_i}.$$

□

Ist obere Schranke $R > r_0$ bekannt, wählt man oft $h_k = \frac{R}{\sqrt{k+1}}$.
Konvergenz recht langsam: $\frac{r_0^2 + R^2 \log(k+1)}{2R\sqrt{k+1}}$.
Typisch: anfangs guter Fortschritt, dann extrem langsam. → Bündelverfahren (sind etwas besser).