VORLÄUFIGE, TEILWEISE ÜBERARBEITETE VERSION (DIE NUMMERIERUNG IST ERST TEILWEISE VON TEX AUF LATEX UMGESTELLT UND ENTSPRECHEND NICHT GANZ KONSISTENT).

5 Konvexe Mengen, Trennungssätze

In diesem Kapitel werden einige Grundlagen über konvexe Mengen und konvexe Funktionen aufgeführt, die in den beiden folgenden Kapiteln zur Herleitung von Optimalitätsbedingungen benutzt werden.

Wir erinnern zunächst an die Definition von "konvex" aus dem Kapitel zur linearen Optimierung.

(5.1a) Definition: Eine Menge $\mathcal{K} \subset \mathbb{R}^n$ heißt *konvex* genau dann, wenn

$$\forall x_1, x_2 \in \mathcal{K}, \quad \forall \lambda \in [0,1]: \qquad \lambda x_1 + (1-\lambda) x_2 \in \mathcal{K},$$

d.h. mit x_1 und x_2 liegt auch die ganze Strecke zwischen x_1 und x_2 in $\mathcal{K}$.
$\mathcal{K} \in \mathbb{R}^n$ heißt ein *Kegel* genau dann, wenn

$$\forall x \in \mathcal{K}, \quad \forall \lambda \geq 0: \qquad \lambda x \in \mathcal{K}$$

Darauf aufbauend benutzen wir noch folgende

(5.1b) Definition: Sei $\mathcal{S} \subset \mathbb{R}^n$ eine beliebige Menge. Mit $\mathrm{conv}(\mathcal{S})$ bezeichnen wir die kleinste konvexe Menge, die $\mathcal{S}$ enthält,

$$\mathrm{conv}(\mathcal{S}) = \bigcap_{\mathcal{S} \subset C \subset \mathbb{R}^n,\ C \text{ ist konvex}} C.$$

Es gilt folgende Darstellung,

$$\mathrm{conv}(\mathcal{S}) = \left\{ x \,\middle|\, \exists k \in \mathbb{N}, \exists \lambda_1, \ldots, \lambda_k \geq 0 \text{ mit } \sum_{i=1}^{k} \lambda_i = 1, \quad \exists x_1, \ldots, x_k \in \mathcal{S} : x = \sum_{i=1}^{k} \lambda_i x_i \right\}.$$

(Zum Beweis dieser Aussage überzeuge man sich, daß die Menge auf der rechten Seite konvex ist und daß jede konvexe Menge, die $\mathcal{S}$ enthält auch die Konvexkombinationen $\sum \lambda_i x_i$ enthalten muß.)

Ist $\mathcal{S}$ endlich, d. h. $\mathcal{S} = \{x_1, \ldots x_m\}$ für ein $m \in \mathbb{N}$, so ist $\mathrm{conv}(\mathcal{S})$ ein Polyeder.

Interessante Aussagen über konvexe Mengen und die konvexe Hülle findet man in Stoer und Witzgall [5]. So ist die konvexe Hülle einer kompakten Menge stets kompakt, die konvexe Hülle der abgeschlossenen Menge $\{(x,y) \in \mathbb{R}^2 \mid x = 0\} \cup \{(0,1)\}$ ist hingegen nicht abgeschlossen. Weiter sagt der Satz von Carathéodory, daß in obiger Definition $k = d+1$ ausreicht, wobei $d \leq n$ die Dimension von $\mathrm{conv}(\mathcal{S})$ ist.

Wir beschränken uns in diesem Kapitel auf die Beweise, die wir im Folgenden noch benötigen. Trotz der Kürze ist dieses Kapitel etwas trocken; viele Aussagen sind anschaulich klar, und der Leser kann die Beweise dazu ohne weiteres überspringen.

(5.2) Definition: Sei $a \in \mathbb{R}^n$, $a \neq 0$, $H := \{x \mid a^T x = \alpha\}$ eine affine Hyperebene, und seien $H_+ := \{x \mid a^T x \geq \alpha\}$ und $H_- := \{x \mid a^T x \leq \alpha\}$ Halbräume. Dann gilt folgende Definition:

- *H trennt $\mathcal{K}_1$ und $\mathcal{K}_2$*: $\Longleftrightarrow \mathcal{K}_1 \subset H_+$ und $\mathcal{K}_2 \subset H_-$ oder umgekehrt.
- *H trennt $\mathcal{K}_1$ und $\mathcal{K}_2$ strikt*: $\Longleftrightarrow \mathcal{K}_1 \subset H_+^\circ$ und $\mathcal{K}_2 \subset H_-^\circ$ oder umgekehrt, wobei $H_\pm^\circ$ das Innere von $H_\pm$ bezeichne.
- *H trennt $\mathcal{K}_1$ und $\mathcal{K}_2$ eigentlich*: $\Longleftrightarrow$ H trennt $\mathcal{K}_1$ und $\mathcal{K}_2$ und es gibt $x_1 \in \mathcal{K}_1$ und $x_2 \in \mathcal{K}_2$ mit $a^T x_1 < a^T x_2$ oder umgekehrt.

Beispiel: Anschaulich besagt die strikte Trennung, daß man zwischen $\mathcal{K}_1$ und $\mathcal{K}_2$ eine Hyperebene (im $\mathbb{R}^2$ also eine Gerade) dazwischenschieben kann. Im Fall der nicht strikten Trennung kann die Hyperebene $\mathcal{K}_1$ oder $\mathcal{K}_2$ (oder beide) berühren, darf die Mengen $\mathcal{K}_1$ und $\mathcal{K}_2$ aber nicht schneiden. Im $\mathbb{R}^2$ wählen wir z.B. für $\mathcal{K}_1$ die x-Achse und für $\mathcal{K}_2$ die abgeschlossene obere Halbebene. Dann ist z.B. $H = \mathcal{K}_2$ selbst ein trennender Halbraum, der $\mathcal{K}_1$ und $\mathcal{K}_2$ eigentlich trennt, obwohl in diesem Fall $\mathcal{K}_1$ ganz in $\mathcal{K}_2$ enthalten ist. Eine strikte Trennung ist in diesem Fall offenbar nicht möglich.

(5.3) Satz: Sei $\mathcal{K} \subset \mathbb{R}^n$ abgeschlossen und konvex. Falls $0 \notin \mathcal{K}$ dann kann 0 strikt von $\mathcal{K}$ getrennt werden, d. h. $\exists a \neq 0;\ \alpha > 0$, so daß

$$a^T x > \alpha > 0 \qquad \forall x \in \mathcal{K}.$$

Beweis: Im Fall $\mathcal{K} = \emptyset$ ist nichts zu zeigen.
Wir setzen also voraus daß $\exists \bar{x} \in \mathcal{K}$. Dann existiert

$$\min\{\|x\| \mid x \in \mathcal{K}\} = \min\{\|x\| \mid \underbrace{x \in \mathcal{K} \cap \{z \mid \|z\| \leq \|\bar{x}\|\}}_{kompakt}\} > 0.$$

Sei y der Minimalpunkt, also $y \in \mathcal{K}$ und $y \neq 0$ (weil $0 \notin \mathcal{K}$). (Als Übung überlege man sich, dass y eindeutig ist.) Für beliebige $x \in \mathcal{K}$ setze:

$$\begin{aligned} \varphi(\lambda) := \| \underbrace{\lambda x + (1-\lambda) y}_{\in \mathcal{K} \text{ für } \lambda \in [0,1]} \|^2 &= (\lambda(x-y)+y)^T(\lambda(x-y)+y) \\ &= \lambda^2 (x-y)^T(x-y) + 2\lambda y^T(x-y) + y^T y. \end{aligned}$$

Dann ist $\varphi'(0) = 2y^T(x-y) \geq 0$, denn $\varphi(\lambda) \geq \varphi(0) = \|y\|^2$ für $\lambda \in [0,1]$. Also gilt $y^T x \geq y^T y$, und somit trennt $a := y$, $\alpha := \frac{1}{2} y^T y > 0$ die Null strikt von $\mathcal{K}$. □

Wir arbeiten nun auf eine "leichte" Verallgemeinerung dieses Satzes hin, nämlich die eigentliche Trennung zweier konvexer Mengen.

(5.4) Satz: Sei $\mathcal{K} \subset \mathbb{R}^n$ konvex, $0 \notin \mathcal{K}$. Dann kann 0 von $\mathcal{K}$ getrennt werden.

Beweis: Wir nehmen wieder an, daß $\mathcal{K} \neq \emptyset$. Für $x \in \mathcal{K}$ sei $A_x := \{y \mid \|y\|_2 = 1, y^T x \geq 0\}$.
Es ist A_x kompakt und $A_x \neq \emptyset$.

Teilbehauptung: Für $x_1, \ldots, x_k \in \mathcal{K}$ ist $A_{x_1} \cap A_{x_2} \cap \ldots \cap A_{x_k} \neq \emptyset$.
Beweis: $P := \text{conv}(x_1, \ldots, x_k)$ ist ein (abgeschlossenes) Polyeder und es ist $P \subset \mathcal{K}$. Wegen $0 \notin P$ kann laut Satz (5.3) die 0 strikt von P getrennt werden, d. h. $\exists \hat{y} \neq 0$, $\|\hat{y}\|_2 = 1$: $\hat{y}^T x > 0$, $\forall x \in P$. Insbesondere ist $\hat{y}^T x_i \geq 0$ für $1 \leq i \leq k$. Aus der Definition von A_x folgt $\hat{y} \in A_{x_i}$ für alle i, d. h.

$$\hat{y} \in A_{x_1} \cap \ldots \cap A_{x_k} \neq \emptyset.$$

Damit ist die Teilbehauptung gezeigt. □
Sei $\mathcal{S} := \{y \mid \|y\|_2 = 1\}$. Wir definieren die offenen Mengen $C_x := \mathbb{R}^n \setminus A_x$.
Behauptung: Es gibt ein $y \in \mathcal{S}$ mit $y \notin \bigcup_{x \in \mathcal{K}} C_x$. Wir beweisen die Behauptung durch Widerspruch und nehmen dazu an daß $\mathcal{S} \subseteq \bigcup_{x \in \mathcal{K}} C_x$.
Da $\mathcal{S}$ kompakt ist, folgt aus dem Satz von Heine Borel, daß es eine endliche Überdeckung von $\mathcal{S}$ gibt, $\mathcal{S} \subseteq C_{x_1} \cup \ldots \cup C_{x_k}$ für gewisse $x_i \in \mathcal{K}$. Also gibt es für jedes $y \in \mathcal{S}$ ein i mit $y \in C_{x_i}$, d.h. $y \notin A_{x_i}$, oder kurz ausgedrückt $\mathcal{S} \cap A_{x_1} \cap \ldots \cap A_{x_k} = \emptyset$. Nach Definition von $\mathcal{S}$ und A_x ist aber $A_{x_i} \subset \mathcal{S}$ für jedes i und somit folgt aus $\hat{y} \in A_{x_1} \cap \ldots \cap A_{x_k} \cap \mathcal{S}$ der gesuchte Widerspruch. □
Also gibt es ein $y \in \mathcal{S}$ mit $y \notin \bigcup_{x \in \mathcal{K}} C_x$; d. h. $y \in A_x$, $\forall x \in \mathcal{K}$, d. h. $y^T x \geq 0$ für alle $x \in \mathcal{K}$.
Damit ist Satz 5.4 gezeigt. □

Es ist eine leichte Aufgabe jetzt zu zeigen, daß in Satz 5.4 die Null sogar eigentlich von $\mathcal{K}$ getrennt werden kann sofern $\mathcal{K} \neq \emptyset$. Dies ist die wesentliche Aussage, die in Kapitel 6 benötigt wird. Für die allgemeinen Optimalitätsbedingungen in Kapitel 7 werden wir auch das Konzept relativ innerer Punkte benutzen. Wir holen dazu an dieser Stelle bereits aus und zeigen anschließend die eigentliche Trennung von 0 und $\mathcal{K}$ in Korollar (5.11) in einem etwas allgemeineren Zusammenhang.

(5.5) Satz: Seien $\mathcal{K}_1$, $\mathcal{K}_2 \subset \mathbb{R}^n$ konvex mit

$$\mathcal{K}_1 \neq \emptyset \neq \mathcal{K}_2 \quad \text{und} \quad \mathcal{K}_1 \cap \mathcal{K}_2 = \emptyset.$$

Dann gibt es eine Hyperebene H, die $\mathcal{K}_1$ und $\mathcal{K}_2$ trennt.

Beweis: Setze $\mathcal{K} := \mathcal{K}_1 - \mathcal{K}_2 = \{x_1 - x_2 \mid x_1 \in \mathcal{K}_1,\ x_2 \in \mathcal{K}_2\}$.
Wie man leicht zeigt, ist $\mathcal{K}$ konvex. Außerdem ist $0 \notin \mathcal{K}$, sonst gäbe es $x_1 = x_2 \in \mathcal{K}_1 \cap \mathcal{K}_2$. Somit existiert aufgrund von Satz (5.4) ein y mit $x^T y \geq 0$, für alle $x \in \mathcal{K}$, d. h. $(x_1 - x_2)^T y \geq 0$, für alle $x_1 \in \mathcal{K}_1$ und für alle $x_2 \in \mathcal{K}_2$. Mit $\alpha := \inf_{x \in \mathcal{K}_1} x^T y$ folgt

$$x_1^T y \geq \alpha \geq x_2^T y \qquad \forall x_1 \in \mathcal{K}_2, \quad x_2 \in \mathcal{K}_2.$$

Man beachte, aus $\mathcal{K}_1 \neq \emptyset \neq \mathcal{K}_2$ folgt, daß α endlich ist. □

Beispiel: Die konvexen und abgeschlossenen Mengen $\{(x, y) \mid y \geq e^x\}$ und $\{(x, y) \mid y \leq 0\}$ in $\mathbb{R}^2$ können nicht strikt getrennt werden obwohl sie abgeschlossen sind und keine gemeinsamen Punkte haben. Wenn aber $\mathcal{K}_1$ und $\mathcal{K}_2 \subset I\!R^n$ abgeschlossen sind und $\mathcal{K}_1$ kompakt ist, dann ist $\mathcal{K}_1 - \mathcal{K}_2$ abgeschlossen (Übung). In diesem Falle ist die strikte Trennung von $\mathcal{K}_1$ und $\mathcal{K}_2$ möglich falls $\mathcal{K}_1 \cap \mathcal{K}_2 = \emptyset$ (denn dann kann 0 von $\mathcal{K}_1 - \mathcal{K}_2$ strikt getrennt werden).

(5.6) Definition: Für $\mathcal{K} \subset \mathbb{R}^n$ ist

$$\text{aff}\,(\mathcal{K}) := \{x \in \mathbb{R}^n \mid \exists x_0, \ldots, x_n \in \mathcal{K}, \exists \lambda_0, \ldots, \lambda_n \in \mathbb{R} : \sum_{i=0}^{n} \lambda_i = 1 \quad \text{und} \quad x = \sum_{i=0}^{n} \lambda_i x_i\}$$

die affine Hülle von $\mathcal{K}$ = *kleinste affine Mannigfaltigkeit*, die $\mathcal{K}$ enthält. Die Festlegung auf $n+1$ Punkte $x_0, \ldots, x_n$ ist eine obere Schranke für die Anzahl der benötigten "Stützpunkte" x_i. Es können auch weniger Stützpunkte ausreichen (dann sind einige $\lambda_i = 0$). Die Normierung $\sum \lambda_i = 1$ ist folgendermaßen zu verstehen: Ein Punkt $x \in \text{aff}(\mathcal{K})$ hat die Darstellung

$x = x_0 + \alpha_1(x_1 - x_0) + \ldots + \alpha_n(x_n - x_0)$. "Man marschiert zu einem Punkt $x_0 \in \mathcal{K}$ und bewegt sich anschließend frei in der linearen Mannigfaltigkeit, die von $\mathcal{K} - x_0$ aufgespannt wird, d.h. die $\alpha_i \in \mathbb{R}$ sind völlig beliebig." Die Zahlen $\lambda_i = \alpha_i$ für $i \geq 1$ und $\lambda_0 = 1 - \sum \alpha_i$ erfüllen offenbar die Bedingung $\sum \lambda_i = 1$.

Ein Punkt $x \in \mathcal{K}$ heißt *relativ innerer* Punkt von $\mathcal{K}$, in Zeichen $x \in \mathcal{K}^i$, falls es eine ε-Umgebung

$$U_\varepsilon(x) = \{z \mid \|z - x\| \leq \varepsilon\}$$

gibt ($\varepsilon > 0$), so daß $U_\varepsilon(x) \cap \operatorname{aff}(\mathcal{K}) \subset \mathcal{K}$ gilt. x heißt *relativer Randpunkt* von $\mathcal{K} \iff x \in \bar{\mathcal{K}} \setminus \mathcal{K}^i$. (Man beachte, daß der obere Index i in $\mathcal{K}^i$ für das Wort "Innere" steht und nicht mit der i-ten Komponente x_i eines Vektors $x \in \mathbb{R}^n$ zu verwechseln ist.)

(5.7) Satz: Sei $\mathcal{K} \neq \emptyset$, $\mathcal{K} \subset \mathbb{R}^n$, $\mathcal{K}$ konvex. Dann ist $\mathcal{K}^i \neq \emptyset$.

Beweis: (Der Satz ist an sich anschaulich klar, soll aber hier trotzdem genau bewiesen werden.)
Sei $\dim \mathcal{K} := \dim \operatorname{aff}(\mathcal{K}) = m \geq 0$.
Dann gibt es affin unabhängige Punkte $x_0, \ldots, x_m \in \mathcal{K}$ und jedes $x \in \operatorname{aff}\mathcal{K}$ läßt sich eindeutig schreiben als $x = \lambda_0 x_0 + \cdots + \lambda_m x_m$, mit $\sum \lambda_i = 1$, d. h.

$$\underbrace{\begin{pmatrix} 1 & \ldots & 1 \\ x_0 & \ldots & x_m \end{pmatrix}}_{=:M} \begin{pmatrix} \lambda_0 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} 1 \\ x \end{pmatrix}$$

ist für alle $x \in \operatorname{aff}\mathcal{K}$ eindeutig nach λ lösbar. Die Matrix $M \in \mathbb{R}^{(n+1)\times(m+1)}$ hat Rang $m+1$ und es gilt

$$\begin{pmatrix} \lambda_0 \\ \vdots \\ \lambda_m \end{pmatrix} = (M^T M)^{-1} M^T \begin{pmatrix} 1 \\ x \end{pmatrix}.$$

(Obige Gleichung wird ersichtlich, wenn man sie von links mit der regulären Matrix $M^T M$ multipliziert.)

Behauptung: $\bar{x} := \frac{1}{m+1} \sum_{i=0}^m x_i \in \mathcal{K}^i$. ($\bar{x}$ ist der Schwerpunkt der x_i.)

Beweis: Sei $\tilde{x} \in \operatorname{aff}\mathcal{K}$ mit $\|\tilde{x} - \bar{x}\|_\infty \leq \varepsilon$ für ein positives $\varepsilon < 1/((m+1)\operatorname{lub}_\infty((M^T M)^{-1} M^T))$.
Dann ist

$$\tilde{x} = \sum_{i=0}^m \tilde{\lambda}_i x_i \quad \text{mit} \quad \sum_{i=0}^m \tilde{\lambda}_i = 1, \qquad \text{bzw.} \qquad \begin{pmatrix} \tilde{\lambda}_0 \\ \vdots \\ \tilde{\lambda}_n \end{pmatrix} = (M^T M)^{-1} M^T \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix}.$$

Aus

$$\frac{1}{m+1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \bar{\lambda}_0 \\ \vdots \\ \bar{\lambda}_n \end{pmatrix} = (M^T M)^{-1} M^T \begin{pmatrix} 1 \\ \bar{x} \end{pmatrix}$$

folgt somit

$$\left\| \begin{pmatrix} \tilde{\lambda}_0 \\ \vdots \\ \tilde{\lambda}_m \end{pmatrix} - \frac{1}{m+1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\|_\infty \leq \operatorname{lub}_\infty((M^T M)^{-1} M^T) \left\| \begin{pmatrix} 1 \\ \tilde{x} \end{pmatrix} - \begin{pmatrix} 1 \\ \bar{x} \end{pmatrix} \right\|_\infty \leq \frac{1}{m+1}.$$

Also ist $\tilde{\lambda}_0, \ldots, \tilde{\lambda}_m \geq 0$, d. h. $\tilde{x}$ ist Konvexkombination der $x_i \in \mathcal{K}$ und somit ist $\tilde{x} \in \mathcal{K}$. $\square$

(5.8) Lemma ("Accessibility Lemma")

a) Sei $\mathcal{K} \subset \mathbb{R}^n$ konvex, $\bar{y} \in \bar{\mathcal{K}}$, $x \in \mathcal{K}^i$, dann gilt

$$[x, \bar{y}) = \left\{ z = (1-\lambda)x + \lambda\bar{y} \;\; | \, 0 \leq \lambda < 1 \right\} \subset \mathcal{K}^i$$

b) $\mathcal{K}^i$ und $\bar{\mathcal{K}}$ sind konvex und es gilt $\overline{\mathcal{K}^i} = \bar{\mathcal{K}} \subset$ aff $\mathcal{K}$ sowie $(\bar{\mathcal{K}})^i = \mathcal{K}^i$. (Wir bezeichnen mit $\bar{\mathcal{K}} := \{y \mid \exists x_k \in \mathcal{K}, \; \lim_{k\to\infty} x_k = y\}$ den Abschluß von $\mathcal{K}$.)

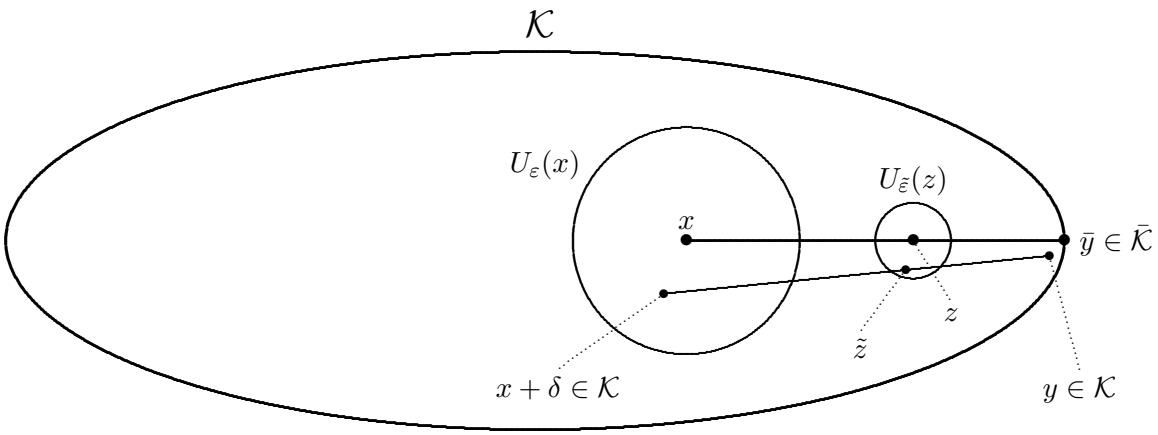
Beweis:

a) Für $\lambda = 0$ ist nichts zu zeigen. Sei $z := (1-\lambda)x + \lambda\bar{y}$ mit $0 < \lambda < 1$ und sei $\varepsilon > 0$ so, daß $U_\varepsilon(x) \cap$ aff$(\mathcal{K}) \subset \mathcal{K}$.

Sei weiter $\tilde{\varepsilon} := \frac{1}{3}(1-\lambda)\varepsilon$ und $\tilde{z} \in U_{\tilde{\varepsilon}}(z) \cap$ aff$(\mathcal{K})$. Es genügt dann zu zeigen daß $\tilde{z} \in \mathcal{K}$. Wegen $\bar{y} \in \bar{\mathcal{K}}$ ist $U_\nu(\bar{y}) \cap \mathcal{K} \neq \emptyset$ für jedes $\nu > 0$ und daher gibt es ein $y \in U_{\tilde{\varepsilon}}(\bar{y}) \cap \mathcal{K}$. Es ist nach Definition von z

$$z = (1-\lambda)x + \lambda\bar{y} \Longrightarrow (1-\lambda)\bar{y} + z - \bar{y} = (1-\lambda)x \Longrightarrow \bar{y} + \frac{1}{1-\lambda}(z - \bar{y}) = x.$$

In obiger Darstellung von x ersetzen wir nun $\bar{y}$ und z durch y und $\tilde{z}$ und erhalten einen Punkt $x + \delta$ mit kleinem $\|\delta\|$ (weil die Punkte y und $\bar{y}$ bzw. z und $\tilde{z}$ nahe beieinander liegen):



Wegen $\tilde{z} \in$ aff $\mathcal{K}$ und $y \in$ aff $\mathcal{K}$ ist

$$\underbrace{y + \frac{1}{1-\lambda}(\tilde{z} - y)}_{\in \text{aff}\,\mathcal{K}} = \underbrace{\bar{y} + \frac{1}{1-\lambda}(z - \bar{y})}_{x} + \underbrace{(y - \bar{y}) + \frac{1}{1-\lambda}(\tilde{z} - z - (y - \bar{y}))}_{\delta} = x + \delta$$

$$\text{mit } \|\delta\| \leq \tilde{\varepsilon} + \frac{1}{1-\lambda} 2\tilde{\varepsilon} < \varepsilon.$$

Nach Definition von ε ist $y + \frac{1}{1-\lambda}(\tilde{z} - y) \in \mathcal{K}$.

Aus der Konvexität von $\mathcal{K}$ und aus $\tilde{z} = \lambda y + (1-\lambda)\Big(y + \frac{1}{1-\lambda}(\tilde{z} - y)\Big)$ folgt $\tilde{z} \in \mathcal{K}$. □

b) folgt mit ähnlichen Überlegungen unter Benutzung von Teil a) □

(5.9) Satz: Sei $\mathcal{K} \subset \mathbb{R}^n$ konvex. Dann gilt $x \in \mathcal{K}^i \iff$ Für jedes $y \in \text{aff } \mathcal{K}$ existiert ein $\varepsilon > 0$ so daß $x \pm \varepsilon(y - x) \in \mathcal{K}$.

Beweis:

"$\Longrightarrow$" Sei $x \in \mathcal{K}^i$, dann gibt es ein $\tilde{\varepsilon} > 0$ so daß $U_{\tilde{\varepsilon}}(x) \cap \text{aff}\,\mathcal{K} \subset \mathcal{K}$. Sei $y \in \text{aff}\,\mathcal{K}$. Ohne Einschränkung sei $y \neq x$. Für $\varepsilon := \tilde{\varepsilon}/\|y - x\|$ ist dann

$$\underbrace{x \pm \varepsilon(y - x)}_{\in \text{aff}\,\mathcal{K}} \in U_{\tilde{\varepsilon}}(x) \cap \text{aff}\,\mathcal{K} \subset \mathcal{K}.$$

"$\Longleftarrow$" Wähle $y \in \mathcal{K}^i$. (So ein y existiert nach Satz (5.7)). Nach Voraussetzung existiert ein $\varepsilon > 0$, so daß $\hat{x} := x - \varepsilon(y - x) \in \mathcal{K}$. Die Definition von $\hat{x}$ läßt sich auch als

$$x = \frac{1}{1+\varepsilon}\hat{x} + \frac{\varepsilon}{1+\varepsilon} y,$$

schreiben, woraus folgt, daß x im Inneren der Verbindungsstrecke zwischen $\hat{x} \in \mathcal{K}$ und $y \in \mathcal{K}^i$ liegt. Nach dem Accessibility Lemma 5.8 liegt x somit in $\mathcal{K}^i$.

□

(5.10) Satz: Seien $\mathcal{K}_1, \mathcal{K}_2 \subset \mathbb{R}^n$ konvex und nicht leer. Dann gilt: Es existiert eine Hyperebene $H = \{x \mid a^T x = \alpha\}$, $a \neq 0$, die $\mathcal{K}_1$ und $\mathcal{K}_2$ *eigentlich* trennt (d. h. H trennt $\mathcal{K}_1$ und $\mathcal{K}_2$ und $\mathcal{K}_1 \cup \mathcal{K}_2 \not\subset H$), genau dann, wenn $\mathcal{K}_1^i \cap \mathcal{K}_2^i = \emptyset$.

Beweisskizze:

"$\Longrightarrow$" H trenne $\mathcal{K}_1$ und $\mathcal{K}_2$ eigentlich, d. h. $\exists a \neq 0$ und $\alpha \in \mathbb{R}$: $a^T x_1 \geq \alpha \geq a^T x_2$ für alle $x_1 \in \mathcal{K}_1$, $x_2 \in \mathcal{K}_2$ und es gibt $\bar{x}_1 \in \mathcal{K}_1$ und $\bar{x}_2 \in \mathcal{K}_2$ mit $a^T \bar{x}_1 > a^T \bar{x}_2$.

Behauptung: Für alle $x_1 \in \mathcal{K}_1^i$ und $x_2 \in \mathcal{K}_2^i$ gilt $a^T x_1 > a^T x_2$ (daraus folgt dann $\mathcal{K}_1^i \cap \mathcal{K}_2^i = \emptyset$).

Annahme: Das gilt nicht, d. h. $\exists x_1 \in \mathcal{K}_1^i$ und $x_2 \in \mathcal{K}_2^i$ mit $a^T x_1 = a^T x_2$. Aus Satz (5.9) folgt: $\exists \varepsilon > 0$: $\tilde{x}_1 = x_1 - \varepsilon(\bar{x}_1 - x_1) \in \mathcal{K}_1$ und $\tilde{x}_2 = x_2 - \varepsilon(\bar{x}_2 - x_2) \in \mathcal{K}_2$. Dann ist $a^T(\tilde{x}_1 - \tilde{x}_2) = -\varepsilon a^T(\bar{x}_1 - \bar{x}_2) < 0$. Also $a^T \tilde{x}_1 < a^T \tilde{x}_2$ im Widerspruch zur Trennung von $\mathcal{K}_1$ und $\mathcal{K}_2$ durch H. □

"$\Longleftarrow$" Wir zeigen zunächst: Für konvexe Mengen $\emptyset \neq \mathcal{K}_1, \mathcal{K}_2 \subseteq \mathbb{R}^n$ gilt allgemein

$$\mathcal{K}_1^i + \mathcal{K}_2^i = (\mathcal{K}_1 + \mathcal{K}_2)^i.$$

Beweis:

"$\subseteq$" Sei $x \in \mathcal{K}_1^i + \mathcal{K}_2^i$. Setze $x = x_1 + x_2$ mit $x_k \in \mathcal{K}_k^i$ für $k = 1, 2$. Sei $y \in \text{aff}(\mathcal{K}_1 + \mathcal{K}_2) \subset \text{aff}(\mathcal{K}_1) + \text{aff}(\mathcal{K}_2)$ beliebig, $y = y_1 + y_2$ mit $y_k \in \text{aff}(\mathcal{K}_k)$. Dann existiert aufgrund von Satz (5.9) ein $\varepsilon > 0$, so daß $x_k \pm \varepsilon(y_k - x_k) \in \mathcal{K}_k$ für $k = 1, 2$. Also $x \pm \varepsilon(y - x) \in \mathcal{K}_1 + \mathcal{K}_2$ und somit folgt wiederum aus Satz (5.9) daß $x \in (\mathcal{K}_1 + \mathcal{K}_2)^i$.

"$\supseteq$"

$$(\mathcal{K}_1 + \mathcal{K}_2)^i \subseteq (\bar{\mathcal{K}}_1 + \bar{\mathcal{K}}_2)^i = \left(\overline{\mathcal{K}_1^i} + \overline{\mathcal{K}_2^i}\right)^i \subseteq \left(\overline{\mathcal{K}_1^i + \mathcal{K}_2^i}\right)^i = (\mathcal{K}_1^i + \mathcal{K}_2^i)^i \subseteq \mathcal{K}_1^i + \mathcal{K}_2^i$$

wegen $\bar{\mathcal{K}} = \overline{\mathcal{K}^i}$, $\bar{A} + \bar{B} \subset \overline{A + B}$ für $A, B \subset \mathbb{R}^n$ und $(\bar{\mathcal{K}})^i = \mathcal{K}^i$ (Accessibility Lemma). □

Sei nun $\mathcal{K}_1^i \cap \mathcal{K}_2^i = \emptyset$. Wir zeigen, daß es eine eigentlich trennende Hyperebne H gibt. Nach Voraussetzung ist $0 \notin \mathcal{K}_1^i - \mathcal{K}_2^i = (\mathcal{K}_1 - \mathcal{K}_2)^i$. Wir betrachten zunächst den Fall, daß $\text{aff}\,(\mathcal{K}_1 - \mathcal{K}_2) = \mathbb{R}^n$ ist. Dann ist $\dim(\mathcal{K}_1 - \mathcal{K}_2) = n$ und $0 \notin (\mathcal{K}_1 - \mathcal{K}_2)^i = (\mathcal{K}_1 - \mathcal{K}_2)^\circ$, wobei A° das Innere einer Menge A bezeichne. $(\mathcal{K}_1 - \mathcal{K}_2)^\circ$ ist eine konvexe Menge. Wegen Satz (5.4) gibt es ein $a \neq 0$, so daß $a^T(x_1 - x_2) \geq 0$ für alle $x_1 \in \mathcal{K}_1$ und $x_2 \in \mathcal{K}_2$. Für die inneren Punkte von $\mathcal{K}_1 - \mathcal{K}_2$ ist diese Ungleichung sogar strikt. Also gibt es $\bar{x}_k \in \mathcal{K}_k$: $a^T(\bar{x}_1 - \bar{x}_2) > 0$. Mit $\alpha = \inf_{x_1 \in \mathcal{K}_1} a^T x_1$ definiert $\{x \mid a^T x = \alpha\}$ eine Hyperebene H, die $\mathcal{K}_1$, $\mathcal{K}_2$ eigentlich trennt. Falls $\dim(\mathcal{K}_1 - \mathcal{K}_2) = m < n$, so läßt sich wie eben gezeigt $\mathcal{K}_1 - \mathcal{K}_2$ innerhalb von $\text{aff}(\mathcal{K}_1 - \mathcal{K}_2)$ durch eine Hyperebene H_m von 0 eigentlich trennen. Setzt man H_m orthogonal zu $\text{aff}(\mathcal{K}_1 - \mathcal{K}_2)$ zu einer Hyperebene im $\mathbb{R}^n$ fort, so folgt die Behauptung. □

(5.11) Korollar:

1) Sei $\mathcal{K}$ konvex und $0 \notin \mathcal{K}$. Dann kann 0 eigentlich von $\mathcal{K}$ getrennt werden.

2) Für nichtleere konvexe Mengen $\mathcal{K}_1, \mathcal{K}_2 \subset \mathbb{R}^n$ gilt: $\mathcal{K}_1^i + \mathcal{K}_2^i = (\mathcal{K}_1 + \mathcal{K}_2)^i$.

3) Falls A eine $m \times n$-Matrix ist und $\mathcal{K} \subset \mathbb{R}^n$ konvex ist, gilt $A(\mathcal{K}^i) = (A\mathcal{K})^i$.

Beweis zu 1). Folgt direkt aus Satz (5.10), wegen $\{0\}^i = \{0\}$.
Beweis zu 2). Siehe oben.
Beweis zu 3). $A\mathcal{K} = \{z = Ax \mid x \in \mathcal{K}\}$ ist konvex. Sei $\tilde{x} \in A(\mathcal{K}^i)$, d. h. $\tilde{x} = A\tilde{k}$ mit $\tilde{k} \in \mathcal{K}^i$. Sei $y \in \text{aff}(A\mathcal{K}) = A\,\text{aff}(\mathcal{K})$, d. h. $y = Ak$ mit $k \in \text{aff}(\mathcal{K})$. Aufgrund von Satz (5.9) existiert ein $\varepsilon > 0$ mit $\tilde{k} \pm \varepsilon(k - \tilde{k}) \in \mathcal{K}$. Also ist auch $\tilde{x} \pm \varepsilon(y - \tilde{x}) = A(\tilde{k} \pm \varepsilon(k - \tilde{k}))$ in $A\mathcal{K}$. Wegen Satz (5.9) reicht dies aus, um zu folgern, $\tilde{x} \in (A\mathcal{K})^i$. Sei nun umgekehrt $\tilde{x} \in (A\mathcal{K})^i$. Wäre $\tilde{x} \notin A(\mathcal{K}^i)$, so könnte $\tilde{x}$ von $A(\mathcal{K}^i)$ nach (5.10) durch eine Hyperebene H eigentlich getrennt werden, $a^T\tilde{x} \leq a^T(Ak)$ für alle $k \in \mathcal{K}^i$ und $a^T\tilde{x} < a^T(A\bar{k})$ für ein $\bar{k} \in \mathcal{K}$. Wegen Lemma (5.8) trennt H auch $\tilde{x}$ von $A\mathcal{K}$ eigentlich. Dies ist aber ein Widerspruch zu Satz (5.10), denn $\{\tilde{x}\} \cap (A\mathcal{K})^i = \{\tilde{x}\} \neq \emptyset$. □

(5.12) Definition: Sei $A \subseteq \mathbb{R}^n$ beliebig, dann heißt

$$A^P := \{y \in \mathbb{R}^n \mid y^T x \leq 0 \text{ für alle } \ x \in A\}$$

polare Kegelmenge zu A.

(5.13) Satz: Seien A, A_1, $A_2 \subseteq \mathbb{R}^n$ beliebig. Dann gilt

1) A^P ist ein abgeschlossener konvexer Kegel.

2) $A_1 \subset A_2 \Longrightarrow A_1^P \supset A_2^P$.

3) $A^{PP} = \overline{\mathcal{C}(A)}$, wobei $\mathcal{C}(A)$ der kleinste konvexe Kegel ist, der A enthält.

   $A^{PP} = A \Longleftrightarrow A$ ist ein abgeschlossener konvexer Kegel.

4) Es gilt $(\overline{\mathcal{C}(A)})^P = A^P$.

5) Ist A ein linearer Teilraum von $\mathbb{R}^n$, so ist $A^P = A^\perp = \{y \in \mathbb{R}^n \mid y^T x = 0 \quad \forall x \in A\}$.

Beweis:

1) $A^P = \bigcap_{x \in A} \{y \mid y^T x \leq 0\}$ ist als Schnitt abgeschlossener Halbräume selbst abgeschlossen. Die Konvexität und die Kegeleigenschaft lassen sich an der obigen Definition leicht verifizieren.

2) Trivial.

3) "$\overline{\mathcal{C}(A)} \subset A^{PP}$": Sei $x \in A \Longrightarrow y^T x \leq 0 \; \forall y \in A^P \Longrightarrow x \in (A^P)^P$ d. h. $A \subset A^{PP}$. Außerdem ist A^{PP} ein abgeschlossener konvexer Kegel. Aus $A \subset A^{PP}$ folgt somit $\overline{\mathcal{C}(A)} \subset A^{PP}$.

   "$A^{PP} \subset \overline{\mathcal{C}(A)}$": Wäre dies falsch, dann gäbe es ein $x_0 \in A^{PP}$ mit $x_0 \notin \overline{\mathcal{C}(A)}$. Nach Trennungssatz (5.3) existiert dann ein $y \neq 0$ mit $y^T x_0 > 0 \geq y^T x \; \forall x \in \overline{\mathcal{C}(A)} \supset A$. $\Longrightarrow y \in A^P$. Wegen $x_0 \in (A^P)^P$, folgt $y^T x_0 \leq 0$ im Widerspruch zu $y^T x_0 > 0$.

   Die zweite Behauptung aus 3) folgt aus der ersten.

4) $(\overline{\mathcal{C}(A)})^P = (A^{PP})^P = (A^P)^{PP} = A^P$, da A^P abgeschlossen und konvex ist.

5) Sei A ein linearer Teilraum. Dann ist mit $x \in A$ auch $-x \in A$ und

$$A^P = \{y \mid y^T x = 0 \quad \forall x \in A\} = A^\perp.$$

(5.14) Definition:

a) $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ heißt *konvexe Funktion* falls

   1) $\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\} \neq \emptyset$. Der *eigentliche* Definitionsbereich von f ist nicht leer.

   2) $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ für alle $0 \leq \lambda \leq 1$ und für alle $x, y \in \mathbb{R}^n$.

b) $g : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ heißt *konkav* falls $-g$ ist konvex ist. Mit $\operatorname{dom} g$ bezeichnen wir dann $\{x \in \mathbb{R}^n \mid g(x) > -\infty\} \neq \emptyset$.

c) f heißt *streng konvex* falls f konvex ist und $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$ $\forall x \neq y$ und $0 < \lambda < 1$.

Bemerkung: Es gibt auch Definitionen (siehe z.B. Rockafellar [4]), die konvexe Funktionen als Funktionen nach $\mathbb{R} \cup \{-\infty, \infty\}$ zulassen. Eine konvexe Funktion, die an einem Punkt den Wert $-\infty$ annimmt nimmt aber notwendigerweise fast überall nur Werte aus $\{-\infty, \infty\}$ an und ist für uns daher uninteressant. Um Fallunterscheidungen zu vermeiden schließen wir solche Funktionen im Folgenden aus.

(5.15) Satz: Sei $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ konvex und $C := \operatorname{dom} f$. Sei C° das Innere von C, dann ist $f(x)$ für alle $x \in C^\circ$ stetig.

Beweis: Falls $C^\circ = \emptyset$ ist nicht zu zeigen. Sei $x_0 \in C^\circ$ und $\dim C = n$.

Für $n+1$ Punkte $y_0, \ldots, y_n \in C$, die affin unabhängig sind, d. h. $\{y_i - y_0\}_{i \in \{1,\ldots,n\}}$ sind linear unabhängig, setzen wir $Q := \{x = \sum_{i=0}^n \lambda_i y_i \mid \sum \lambda_i = 1, \lambda_i \geq 0\}$.

Q ist ein Simplex, der in C enthalten ist (da C konvex ist). Außerdem ist $Q^\circ \neq \emptyset$. Seien die y_i nun so gewählt, daß $x_0 = \frac{1}{n+1} \sum_{i=0}^n y_i \in Q^\circ$. Für alle $x = \sum \lambda_i y_i \in Q$ ist $f(x) = f(\sum_{i=0}^n \lambda_i y_i) \leq \sum_{i=0}^n \lambda_i f(y_i) \leq \max\limits_{0 \leq i \leq n} f(y_i) =: M$. Also ist f auf Q beschränkt. Sei $\varepsilon > 0$ so daß $U_\varepsilon(x_0) \subset Q$ gilt. Für $\Delta x \in \mathbb{R}^n$ mit $\|\Delta x\| < \varepsilon$ gilt dann wegen $x_0 + \sigma \Delta x = \sigma(x_0 + \Delta x) + (1 - \sigma) x_0$ für $0 \leq \sigma \leq 1$ die Ungleichung

$$f(x_0 + \sigma \Delta x) \leq \sigma f(x_0 + \Delta x) + (1 - \sigma) f(x_0).$$

Daraus folgt

$$f(x_0 + \sigma \Delta x) - f(x_0) \leq \sigma(f(x_0 + \Delta x) - f(x_0)) \leq \sigma(M - f(x_0)).$$

Wegen $x_0 = \frac{\sigma}{1+\sigma}(x_0 - \Delta x) + \frac{1}{1+\sigma}(x_0 + \sigma \Delta x)$ ist auch $f(x_0) \leq \frac{\sigma}{1+\sigma} f(x_0 - \Delta x) + \frac{1}{1+\sigma} f(x_0 + \sigma \Delta x)$.

Multipliziert man dies mit $1 + \sigma$, so folgt

$$f(x_0 + \sigma \Delta x) - f(x_0) \geq \sigma(f(x_0) - f(x_0 - \Delta x)) \geq \sigma(f(x_0) - M).$$

Also: $|f(x_0 + \sigma \Delta x) - f(x_0)| \leq \sigma(M - f(x_0))$ und somit ist f in x_0 stetig. □

Bemerkung: Satz (5.15) besagt auch daß die Einschränkung $f\big|_C$ von f auf C in C^i stetig ist. Der Satz gilt nicht in unendlichdimensionalen Räumen, dort können sogar lineare Abbildungen (die selbstverständlich konvex sind) unstetig sein.

6 Konvexe Ungleichungssysteme und der Satz von Kuhn & Tucker für konvexe Optimierungsprobleme

In diesem Kapitel werden Bedingungen hergeleitet, die es erlauben, anhand der Daten eines konvexen Optimierungsproblems abzulesen, ob ein gegebener Punkt optimal ist oder nicht. Diese Frage ist bei Funktionen von mehreren Unbekannten—und bei gegebenen Nebenbedingungen an die Unbekannten—in der Tat nicht leicht zu beantworten. Die Antwort aus diesem Kapitel ist die Grundlage für viele numerische Verfahren zur Bestimmung einer Optimallösung und ist für das Verständnis dieser Verfahren sehr wichtig. Es "lohnt sich" daher, die Optimalitätsbedingungen genau zu beleuchten. Salopp ausgedrückt, die Bedeutung der Optimalitätsbedingungen ändert sich nicht, wohingegen die praktische Bedeutung der einzelnen Optimierungsverfahren doch gewissen Modeschwankungen unterliegt; die Vorzüge

der einzelnen Verfahren hängen nicht zuletzt auch sehr von der Struktur des jeweiligen Problems und der benutzten Computerarchitektur ab.

Eine differenzierbare konvexe Funktion f hat auf $\mathbb{R}^n$ genau dann ein Minimum bei $\bar{x}$ wenn $Df(\bar{x}) = 0$ gilt. (Wir überlassen den Nachweis dieser Aussage als einfache Übung.) Ziel der folgenden Betrachtungen ist es, eine Verallgemeinerung dieser Bedingung für den Fall konvexer Nebenbedingungen zu finden. Wir beginnen mit einem einfachen Hilfsresultat.

(6.1) Satz: Seien f_i; $i = 1, \ldots, m$ konvexe Funktionen auf $\mathbb{R}^n$, $C \subset \mathbb{R}^n$ konvex mit $\emptyset \neq C \cap \bigcap_{i=1}^m \operatorname{dom} f_i$. Dann gilt:

$$\text{Die Ungleichung } F(x) := \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} < 0 \text{ mit } x \in C \text{ ist unlösbar}$$

$$\iff \quad \exists z \in \mathbb{R}^m\colon\ z \geq 0,\ z \neq 0,\ z^T F(x) = \sum_{i=1}^m z_i f_i(x) \geq 0 \text{ für alle } x \in C.$$

Beweis:

"$\Longleftarrow$" Klar, denn die zweite Bedingung impliziert, daß für jedes $x \in C$ mindestens ein i existiert mit $f_i(x) \geq 0$.

"$\Longrightarrow$" Die Menge $A := \{v \in \mathbb{R}^m \mid \exists x \in C\colon\ \ v > F(x)\}$ ist konvex. Weiter ist $A \subset \mathbb{R}^m$, $A \neq \emptyset$ und $0 \notin A$. Nach Satz (5.4) kann $\{0\}$ von A getrennt werden, d. h. $\exists z \in \mathbb{R}^m$, $z \neq 0$, so daß $z^T v \geq z^T 0 = 0\ \forall v \in A$. Da mit $v \in A$ auch $v + \lambda e_i \in A$ für alle $\lambda \geq 0$ (e_i bezeichnet den i-ten Einheitsvektor) folgt $z \geq 0$. (Wäre ein $z_i < 0$ so erhielte man für große λ einen Widerspruch.)

Insgesamt folgt $z \geq 0$, $z^T F(x) \geq 0$, $\forall x \in C$. □

Die Bedingungen an den Vektor z aus Satz 6.1 sollen im Folgenden etwas verschärft und auf das Problem (6.2) angewandt werden.

Problem (6.2):

$$\min_{x \in \mathcal{S}} f(x) \quad \text{mit} \quad \mathcal{S} := \{x \in C \mid f_i(x) \leq 0 \quad 1 \leq i \leq p, \quad f_j(x) = 0 \quad j = p+1, \ldots, m\}$$

Für $f(x)$ schreiben wir im Folgenden auch $f_0(x) := f(x)$ und treffen stets die folgende Voraussetzung

(**V**): Es sei $C \subset \mathbb{R}^n$ konvex, $C \subset \operatorname{dom} f_i$ für $0 \leq i \leq m$, f_i konvex für $0 \leq i \leq p$ und f_j affin für $p+1 \leq j \leq m$.

Desweiteren setzen wir

$$F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix}; \quad F_1(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix}, \quad F_2(x) = \begin{pmatrix} f_{p+1}(x) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

(6.3) Satz: Voraussetzung:

1) Es gelte Voraussetzung (V).

2) Setze $\mathcal{S} := \{x \in C \mid f_i(x) \leq 0,\ i = 1, \ldots, p,\ \ f_j(x) = 0,\ j = p+1, \ldots, m\}$ ($\mathcal{S}$ ist konvex). Für $\mathcal{S}$ gelte

a) $\exists \bar{x} \in \mathcal{S} \cap C^i$.

b) Für alle nichtaffinen Funktionen $f_i(x)$, $(i \leq p)$ gilt: $\exists x^i \in \mathcal{S}$: $f_i(x_i) < 0$.

3) Es gibt kein $x \in \mathcal{S}$ mit $f(x) < 0$, (d. h. $\inf_{x \in \mathcal{S}} f(x) \geq 0$).

Dann gilt

(6.3.a) $\exists y \in \mathbb{R}^m$: $y_i \geq 0$ für $i = 1, \ldots, p$, und $f(x) + y^T F(x) \geq 0$ für alle $x \in C$.

Bemerkung: 2.b) ist äquivalent zu $\exists \hat{x} \in \mathcal{S}$: f: $f_i(\hat{x}) < 0$ für alle $i = 1, \ldots, p$, für die f_i nichtaffin ist.

Beweis: Seien $f_1, \ldots, f_l$ nichtaffine konvexe Funktionen und $f_{l+1} \ldots, f_p$ affin. Wegen 2.b) gibt es für $1 \leq i \leq l$ je ein $x^i \in \mathcal{S}$ mit $f_i(x^i) < 0$ und $f_j(x^i) \leq 0$ für $j \neq i$, $j = 1, \ldots, p$, sowie $f_j(x^i) = 0$ für $j = p+1, \ldots, m$. Setze $\hat{x} = \frac{1}{l}(x^1 + x^2 + \cdots + x^l)$. Dann ist

$$f_i(\hat{x}) = f_i(\frac{1}{l}(x^1 + \cdots + x^i + \cdots + x^l)) \leq \frac{1}{l}(\underbrace{f_i(x^1)}_{\leq 0} + \cdots + \underbrace{f_i(x^i)}_{<0} + \cdots + \underbrace{f_i(x^l)}_{\leq 0}) < 0.$$

Beweis von Satz (6.3): Um die Fallunterscheidungen zu vereinfachen, setzen wir voraus:

$$\exists \hat{x} \in \mathcal{S}: \quad f_i(\hat{x}) < 0 \quad \text{für } \{1 \leq i \leq p\},$$

d. h. $f_i(\hat{x}) < 0$ auch für die affinen f_i mit $i \in \{1, \ldots, p\}$.

(Die affinen Ungleichungen, die für alle $x \in \mathcal{S}$ mit Gleichheit erfüllt sind, kann man als Gleichungen behandeln. Aus Satz 6.2 läßt sich dann jedoch nicht ablesen, daß eine Wahl von y mit $y_i \geq 0$ auch für alle affinen $i \in \{1, \ldots, p\}$ möglich ist. Weiter ist dann die Menge der $y \in \mathbb{R}^m$, die (6.3 a) erfüllen konvex, und die Annahme, daß keines der y die Bedingung $y_i \geq 0$ für $i \leq p$ erfüllt läßt sich mit Trennungssatz (5.10) auf einen Widerspruch führen. Wir verzichten hier auf die Details.)

Wie im Beweis der vorangegangenen Bemerkung können wir durch eine Konvexkombination von $\bar{x}$ aus Bedingung 2.a) und $\hat{x}$ (von oben) sogar annehmen, daß

$$\exists \hat{x} \in \mathcal{S} \cap C^i: \quad f_i(\hat{x}) < 0 \quad \text{für } \{1 \leq i \leq p\}.$$

Der Beweis teilt sich in zwei Schritte auf:

1) $\exists z = (z_0, \ldots, z_m)^T \in \mathbb{R}^{m+1} \setminus \{0\}$ mit $z_i \geq 0$ für $0 \leq i \leq p$ und $\sum_{i=0}^m z_i f_i(x) \geq 0$ für alle $x \in C$.

2) Es ist $z_0 > 0$ und mit $y := \left(\frac{z_1}{z_0}, \ldots, \frac{z_m}{z_0}\right)^T$ ist die Behauptung des Satzes gezeigt.

Zu 1) Sei (in leichter Abwandlung zum Beweis von Satz (6.1))

$$A := \left\{ v = \begin{pmatrix} v_0 \\ \vdots \\ v_m \end{pmatrix} \in \mathbb{R}^{m+1} \quad \middle| \quad \exists x \in C : v_0 > f_0(x),\ v_i \geq f_i(x) \text{ für } 1 \leq i \leq p \right.$$
$$\left. \text{und } v_j = f_j(x) \text{ für } p+1 \leq j \leq m \right\}.$$

A ist konvex (trivial), $A \neq \emptyset$, und wegen Voraussetzung 3) ist $0 \notin A$. Also kann 0 laut Korollar (5.11), 1) von A eigentlich getrennt werden, d. h. $\exists z \neq 0$ mit $z^T v \geq 0$ ($= z^T 0$) $\forall v \in A$ und $z^T \bar{v} > 0$ für ein $\bar{v} \in A$.

Aus der Definition von A folgt wieder $z_i \geq 0$ für $1 \leq i \leq p$. Außerdem ist $z_0 f(x) + \sum z_i f_i(x) \geq 0\ \forall x \in C$. (Wähle $v = v_\varepsilon = \binom{f_0(x)+\varepsilon}{F(x)} \in A$ für $\varepsilon > 0$ und bilde den Grenzwert $\epsilon \to 0$.) Damit ist Teil 1) gezeigt.

Zu 2) Falls $z_0 = 0$ so wähle $x = \hat{x}$ und $v = (f_0(\hat{x}) + 1, f_1(\hat{x}), \ldots, f_p(\hat{x}), 0, \ldots, 0)^T$. Wegen $v \in A$ ist $z^T v \geq 0$. Weiter ist wegen $z_0 = 0$, $z_1 \geq 0, \ldots, z_p \geq 0$ und $f_1(\hat{x}) < 0, \ldots, f_p(\hat{x}) < 0$ (Definition von $\hat{x}$!) notwendigerweise $z_1 = \ldots = z_p = 0$. Also ist aufgrund der Definition von A

$$z_{p+1} f_{p+1}(x) + \cdots + z_m f_m(x) \geq 0 \quad \forall x \in C. \tag{67}$$

Aus der eigentlichen Trennung folgt sogar $\exists \tilde{x} \in C$: $z_{p+1} f_{p+1}(\tilde{x}) + \cdots + z_m f_m(\tilde{x}) > 0$. Für kleine $\varepsilon > 0$ ist wegen $\hat{x} \in C^i$ auch $\hat{x} \pm \varepsilon(\tilde{x} - \hat{x}) \in C$, und weil f_j affin ist, ist

$$f_j(\hat{x} - \varepsilon(\tilde{x} - \hat{x})) = \underbrace{f_j(\hat{x})}_{=0} - \varepsilon(f_j(\tilde{x}) - f_j(\hat{x})) = -\varepsilon f_j(\tilde{x}) \qquad \text{für } j \geq p+1$$

also ist

$$z_{p+1} f_{p+1}(\hat{x} - \varepsilon(\tilde{x} - \hat{x})) + \cdots + z_m f_m(\hat{x} - \varepsilon(\tilde{x} - \hat{x})) = -\varepsilon(z_{p+1} f_{p+1}(\tilde{x}) + \cdots + z_m f(\tilde{x})) < 0$$

im Widerspruch zu (67). □

Die Bedingung 2.a) und 2.b) heißt

(6.4) **Constraint qualification von Slater**

und besagt $\exists x^1 \in C^i \cap \mathcal{S}$: $\quad f_i(x) < 0$ für alle nichtaffinen f_i mit $i \in 1, \ldots, p$ (Verbot einer gewissen Form von Entartung bei den nichtaffinen Nebenbedingungen). Wir erläutern kurz die Voraussetzung 2 des Satzes (6.3).

Beispiel zur Bedeutung der Voraussetzung 3b) in Satz (6.3): Sei

1) $C := \mathbb{R}$ konvex, $f(x) := x$ und $f_1(x) := x^2$ konvex, d. h. ($n = m = p = 1$). Dann ist $C \subset \operatorname{dom} f_i$, und

2) $\mathcal{S} := \{x \in C \mid f_1(x) \leq 0\} = \{0\}$ erfüllt

   a) $\exists \bar{x} \in \mathcal{S} \cap C^i$, aber

   b) $\nexists x^1 \in \mathcal{S}$ mit $f_1(x^1) < 0$. Die Aussage

3) $\nexists x \in \mathcal{S}$ mit $f(x) < 0$ ist wieder richtig, und auch die Folgerung $\exists z \in \mathbb{R}^{m+1}$: $z_0 x + z_1 x^2 \geq 0\ \forall x \in \mathbb{R}$ gilt, wie das Beispiel $(z_0, z_1) = (0, 1)$ zeigt. Aber $z_0 \neq 0$ ist nicht möglich, egal wie groß z_1 gewählt ist.

Beispiel zur Bedeutung der Voraussetzung 2 a) in Satz (6.3): Sei

$$f(x) := \begin{cases} -\sqrt{x} & \text{für } x \geq 0 \\ \infty & \text{sonst} \end{cases}$$

und $f_1(x) := x$ sowie $C := \{x \in \mathbb{R} \mid x \geq 0\}$, $\mathcal{S} = \{x \in C \mid f_1(x) \leq 0\} = \{0\}$. Jetzt ist die Bedingung 3a) verletzt: $C^i \cap \mathcal{S} = \emptyset$. Wir prüfen die Existenz von $y_1 \geq 0$ mit

$$f(x) + y_1 f_1(x) \geq 0 \quad \forall x \in C,$$

d. h. $-\sqrt{x} + y_1 x \geq 0\ \forall x \geq 0$. Es existiert kein solches y_1, denn: Wähle zu $y_1 > 0$ die Zahl $x = \frac{1}{4y_1^2}$, dann ist $-\sqrt{x} + y_1 x = -\frac{1}{2y_1} + \frac{1}{4y_1} < 0$.

Diskussion: Der 1. Teil des Beweises von Satz (6.3) besagt $\exists z \in \mathbb{R}^{m+1}$: $z_0, \ldots, z_p \geq 0$, $z \neq 0$ mit

$$\phi(x) := \sum_{i=0}^{m} z_i f_i(x) \geq 0 \qquad \forall x \in C. \tag{68}$$

Für den Spezialfall daß die Voraussetzungen 1) bis 3) in Satz (6.3) durch die Bedingungen $C = \mathbb{R}^n$ und $\exists x^* \in \mathcal{S}$ mit $\min_{x \in \mathcal{S}} f(x) = f(x^*)$ sowie $f(x^*) = 0$ ergänzt werden, und alle f_i in x^* differenzierbar sind so folgt aus (68): Es gibt ein $z \in \mathbb{R}^{m+1}$ mit $z_0, z_1, \ldots, z_p \geq 0$ und und für $x = x^*$ gilt

$$\begin{gathered} \textstyle\sum_{i=0}^{m} z_i \nabla f_i(x) = 0 \\ f_i(x) z_i = 0,\ \ f_i(x) \leq 0 \quad \text{für} \quad 0 \leq i \leq p \\ f_j(x) = 0 \quad \text{für} \quad p+1 \leq j \leq m \end{gathered} \tag{69}$$

auch *ohne constraint qualification* (Bed. 2.a), 2.b)). Dabei folgt die erste Zeile von (69) weil $\phi(x)$ eine konvexe Funktion ist mit $\phi(x^*) \leq 0$ (wegen $z_i \geq 0$ für $0 \leq i \leq p$ und $f_i(x^*) \leq 0$ für $0 \leq i \leq p$ sowie $f_j(x^*) = 0$ für $j \geq p+1$). Da außerdem $\phi(x) \geq 0$ für alle $x \in I\!R^n$ gilt nimmt ϕ bei x^* sein Minimum an, d.h. der Gradient von ϕ bei $x = x^*$ ist Null. Dies ist genau die erste Zeile von (69). Die zweite Zeile folgt aus $z_i \geq 0,\ f_i(x) \leq 0$ für $x \in \mathcal{S}$ und $1 \leq i \leq p$. Wäre nämlich eines der Produkte von Null verschieden, so müßte es strikt negativ sein, und dann wäre $\phi(x^*) < 0$ ein Widerspruch. Die dritte Zeile schließlich folgt wieder wegen $x \in \mathcal{S}$.

Der Nutzen von obigem Resultat ist durch die Forderung $f(x^*) = 0$ stark eingeschränkt. Ist $f(x^*) = 0$, so könnte man versuchen das System (69) zu lösen um eine Optimallösung von $\min_{x \in \mathcal{S}} f(x)$ zu bestimmen. Da für $\lambda > 0$ mit z auch λz eine Lösung dieses Systems ist, kann man zusätzlich z. B. $\sum_{i=0}^{p} z_i = 1$ verlangen.

Dieser Zugang behandelt f_0 und f_i für $1 \leq i \leq p$ völlig „gleichberechtigt“. In der Regel ist aber $f_0(x^*)$ nicht bekannt. Obige Idee ist dann nicht anwendbar und f_0 und f_i für $i \geq 1$ sind daher in ihrer Rolle verschieden.

Die Bedingungen 2.a) und 2.b) erlauben nun ohne Kenntnis von $f(x^*)$ zunächst zu dem äquivalenten Problem überzugehen, bei welchem die Funktion $f^*(x)$ zu minimieren ist, wobei $f^*(x) := f(x) - f(x^*)$ gar nicht bekannt zu sein braucht! Die beiden Gleichungen

$z_0 f^*(x) = 0$ und $\sum_{i=0}^{p} z_i = 1$ lassen sich nämlich durch eine Gleichung $z_0 = 1$ ersetzen. (Die Summe $\sum_{i=0}^{p} z_i$ ist bei Festlegung von $z_0 = 1$ in der Regel größer als 1; sie hängt z. B. von den f_i ab). Das so entstandene System hat die Form

$$\begin{aligned} \nabla f(x) + \textstyle\sum_{i=1}^{m} z_i \nabla f_i(x) &= 0 \\ f_i(x) z_i = 0, \;\; f_i(x) \le 0 \quad &\text{für} \quad 1 \le i \le p\,, \\ f_j(x) = 0 \quad &\text{für} \quad p+1 \le j \le m \end{aligned} \tag{70}$$

d.h. es liegt ein nichtlineares Vorzeichen-beschränktes Gleichungssystem mit $n + m$ Unbekannten und ebensovielen Gleichungen vor. Beachte, daß $\nabla f(x) = \nabla f^*(x)$ und daß der unbekannte Wert $f(x^*)$ in (70) garnicht eingeht! Falls Problem (6.2) die Slaterbedingung erfüllt, ist die Lösung von (70) äquivalent[1] zur Lösung des Problems (6.2). Eine Variante des Newtonverfahrens zur Lösung des Systems (70) ist die Grundlage der sogenannten "primal-dualen Innere-Punkte-Verfahren" aus Kapitel 8.

Übung: Aus Satz (6.3) läßt sich Farka's Lemma

$$\text{“}(A^T x \le 0 \Longrightarrow c^T x \le 0) \Longleftrightarrow (\exists u \ge 0 \colon c = Au)\text{”}$$

herleiten.

(6.5) Definition:

$$L \colon C \times D \longrightarrow \mathbb{R}, \quad D := \{y \in \mathbb{R}^m \mid y_i \ge 0 \text{ für } 1 \le i \le p\}$$

$$L(x, y) := f(x) + \sum_{i=1}^{m} y_i f_i(x) = f(x) + y^T F(x)$$

heißt *Lagrangefunktion* von (6.2). Der Punkt $(\bar{x}, \bar{y}) \in C \times D$ heißt *Sattelpunkt* von L auf $C \times D$ falls

$$L(\bar{x}, y) \le L(\bar{x}, \bar{y}) \le L(x, \bar{y}) \quad \forall (x, y) \in C \times D.$$

Satz (6.3) läßt sich mit dieser Definition schreiben als

(6.6) Satz von (Karusch), Kuhn & Tucker
Sei (**V**) für Problem (6.2) erfüllt, dann gilt:

1) Falls $(\bar{x}, \bar{y})$ Sattelpunkt der Langrangefunktion auf $C \times D$ ist, dann ist $\bar{x}$ optimal für (6.2) und $\bar{y}_i f_i(\bar{x}) = 0$ für $1 \le i \le m$, d. h.

$$L(\bar{x}, \bar{y}) = f(\bar{x}).$$

2) Falls $\bar{x}$ Optimallösung von (6.2) ist und die Constraint Qualification (6.4) gilt, gibt es ein $\bar{y} \in D$, so daß $(\bar{x}, \bar{y})$ Sattelpunkt von L ist.

Beweis:

[1] Streng genommen haben wir nur gezeigt, daß es zu einer Lösung x von Problem (6.2) einen Vektor z geben muß, der zusammen mit x das System (70) löst sofern die Slaterbedingung erfüllt ist. Die Umkehrung folgt mit Argumenten ähnlich wie im nachfolgenden Satz (6.6).

1) Sei $(\bar{x}, \bar{y})$ ein Sattelpunkt von L auf $C \times D$. Dann ist

$$L(\bar{x}, \bar{y}) \geq L(\bar{x}, y) = f(\bar{x}) + \underbrace{\sum_{i=1}^{p} y_i f_i(\bar{x})}_{\Rightarrow f_i(\bar{x}) \leq 0} + \underbrace{\sum_{j=p+1}^{m} y_j f_j(\bar{x})}_{\Rightarrow f_j(\bar{x}) = 0} \qquad \forall y \in D$$

denn die linke Seite ist beschränkt und die $y_i \geq 0$ bzw. $y_j \in \mathbb{R}$ können beliebig gewählt werden. Also ist $\bar{x} \in \mathcal{S}$.
Falls $f_i(\bar{x})\bar{y}_i \neq 0$ für ein $i \in \{1, \ldots, p\}$, so muß $f_i(\bar{x}) < 0$ und $\bar{y}_i > 0$ sein. Wir setzen dann

$$\left.\begin{array}{l} y_i = 0 \text{ für dieses } i \\ \text{und } y_i = \bar{y}_i \text{ sonst} \end{array}\right\} \Longrightarrow L(\bar{x}, y) > L(\bar{x}, \bar{y})$$

im Widerspruch zur Definition des Sattelpunktes, also $\bar{y}_i f_i(\bar{x}) = 0 \; \forall i = 1, \ldots, m$.
Für beliebige $x \in \mathcal{S}$ ist

$$f(\bar{x}) = L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) = f(x) + \sum_{i=1}^{p} \underbrace{f_i(x)}_{\leq 0} \underbrace{\bar{y}_i}_{\geq 0} + \sum_{j=p+1}^{m} \underbrace{f_j(x)}_{=0} \bar{y}_j \leq f(x)$$

d. h. $f(\bar{x})$ ist das Minimum von (6.2).

2) Falls $\bar{x}$ für (6.2) optimal ist und die Slater-Bedingung erfüllt ist, ist Satz (6.3) mit $\bar{f}(x) := f(x) - f(\bar{x})$ anwendbar, d. h. $\exists \bar{y} \in D$ mit

$$\bar{f}(x) + \bar{y}^T F(x) \geq 0 \qquad \forall x \in C$$

beziehungsweise

$$L(x, \bar{y}) = f(x) + \bar{y}^T F(x) \geq f(\bar{x}) \qquad \forall x \in C.$$

Für $x = \bar{x}$ folgt $\bar{y}^T F(\bar{x}) \geq 0$. Wegen $f_j(\bar{x}) = 0$ für $j \geq p+1$ ist daher $\sum_{i=1}^{p} \bar{y}_i f_i(\bar{x}) \geq 0$, und wegen $\bar{y}_i \geq 0$, $f_i(\bar{x}) \leq 0$ gilt $\bar{y}^T F(x) = 0$. Zusammenfassend erhält man

$$L(x, \bar{y}) \geq L(\bar{x}, \bar{y}) = f(\bar{x}) \geq f(\bar{x}) + \sum_{i=1}^{p} \underbrace{y_i}_{\geq 0} \underbrace{f_i(\bar{x})}_{\leq 0} + \sum_{p+1}^{m} y_j \underbrace{f_j(\bar{x})}_{=0} = L(\bar{x}, y)$$

für alle $(x, y) \in C \times D$, also ist $(\bar{x}, \bar{y})$ ein Kuhn-Tucker-Punkt. □

Satz 6.6 zeigt den Satz von Kuhn und Tucker in einer sehr allgemeinen Form. Von praktischer Bedeutung ist in erster Linie die Formulierung (70), deren erste Zeile sich auch mithilfe der Lagrangefunktion mittels $\nabla_x L(x, z) = 0$ schön kompakt ausdrücken läßt.

6.1 Dualität bei konischen konvexen Programmen

In Anlehnung an das Buch [3] schildern wir hier noch eine weitere sehr elegante Möglichkeit, für konvexe Probleme, welche die Slaterbedingung erfüllen, ein duales Problem zu formulieren.

Die Grundidee beruht auf der Beobachtung, daß sich ein konvexes Problem stets in einer konischen Standardform schreiben läßt. Dazu erinnern wir zunächst an die Definition

des polaren Kegels aus Kapitel 5. Eine eng verwandte, gebräuchliche Definition ist die des dualen Kegels. Sei $\mathcal{K} \subset I\!R^n$ ein Kegel, dann ist

$$\mathcal{K}^D := -\mathcal{K}^P = \{y \in \mathbb{R}^n \mid y^T x \geq 0 \text{ für alle } \; x \in \mathcal{K}\} \tag{71}$$

der **duale Kegel** zu $\mathcal{K}$. Man überzeugt sich leicht, daß $(I\!R^n_+)^D = I\!R^n_+$, d.h. der $I\!R^n_+$ ist selbstdual.

Das Konzept des dualen (oder polaren) Kegels ist dabei nicht an die spezielle Struktur des zugrunde liegenden Raumes gebunden, es genügt ein Raum mit einem Skalarprodukt. Bezeichnet man z.B. mit $\mathcal{S}^n$ den Raum der symmetrischen $n \times n$-Matrizen, so kann man für $X, Z \in \mathcal{S}^n$ das Skalarprodukt

$$\langle X, Z \rangle := \text{Spur}(X^T Z) = \sum_{i=1}^{n} \sum_{j=1}^{n} X_{i,j} Z_{i,j}$$

definieren. (Die obige Definition gilt auch für nichtsymmetrische Matrizen X und Z; für symmetrisches X kann man oben natürlich kürzer $\text{Spur}(XZ)$ an Stelle von $\text{Spur}(X^T Z)$ schreiben.) Dieses Skalarprodukt definiert die Frobeniusnorm,

$$\|X\|_F := \sqrt{\langle X, X \rangle} = \sqrt{\sum_{i,j} X_{i,j}^2}.$$

In $\mathcal{S}^n$ ist die Menge der symmetrischen positiv semidefiniten Matrizen $\mathcal{S}^n_+$ durch

$$\mathcal{S}^n_+ = \{X \in \mathcal{S}^n \mid h^T X h \geq 0 \text{ für alle } \; h \in I\!R^n\}$$

gegeben. Man überzeugt sich leicht, daß $\mathcal{S}^n_+$ ein konvexer Kegel ist. Der Satz von Fejer (siehe z.B. [2]) besagt: Eine symmetrische Matrix Z is positiv semidefinit genau dann, wenn $\langle X, Z \rangle \geq 0$ für alle positiv semidefiniten Matrizen X gilt. Anders ausgedrückt heißt das, daß auch $\mathcal{S}^n_+$ selbstdual ist,

$$(\mathcal{S}^n_+)^D = \{Z \in \mathcal{S}^n \mid \langle X, Z \rangle \geq 0 \text{ für alle } \; X \in \mathcal{S}^n_+\} = \mathcal{S}^n_+. \tag{72}$$

In (72) haben wir das Skalarprodukt $y^T x$ aus (71) durch die Schreibweise $\langle y, x \rangle$ bzw. $\langle X, Z \rangle$ ersetzt, das Produkt $X^T Z$ ist hier ja eine Matrix. Wir möchten diese Schreibweise $\langle \ . \ , \ . \ \rangle$ des Skalarproduktes für den Rest dieses Kapitels beibehalten.

Wir zeigen nun, wie ein konvexes Programm (6.2)

$$\min_{x \in \mathcal{S}} f(x) \quad \text{mit} \quad \mathcal{S} := \{x \in C \mid f_i(x) \leq 0 \quad 1 \leq i \leq p, \quad f_j(x) = 0 \quad j = p+1, \ldots, m\}$$

in eine konische Standardform umgewandelt werden kann. Zunächst können wir ohne Einschränkung der Allgemeinheit annehmen, daß $f(x) = c^T x$ linear ist. (Dies kann man stets erreichen, indem man z.B. eine neue Variable x_{n+1} einführt und die zusätzliche Nebenbedingung $f(x) \leq x_{n+1}$ fordert und dann x_{n+1} minimiert. Letzteres ist natürlich eine lineare Funktion des erweiterten Vektors (x, x_{n+1}) von Unbekannten.) Im folgenden sei wieder $x \in I\!R^n$ d.h. $\mathcal{S} \subset I\!R^n$.

Wir "heben" die Menge $\mathcal{S}$ in einen konvexen Kegel im $I\!R^{n+1}$ und definieren

$$\widehat{\mathcal{K}} := \{(x, x_{n+1}) \in I\!R^{n+1} \mid x_{n+1} > 0, \quad \frac{1}{x_{n+1}} x \in \mathcal{S}\}.$$

Offensichtlich ist $\widehat{\mathcal{K}}$ ein Kegel. Weiter kann man leicht nachrechnen, wenn $\mathcal{S}$ konvex ist, dann auch $\widehat{\mathcal{K}}$. Man beachte aber, daß $\widehat{\mathcal{K}} \cup \{0\}$ im allgemeinen nicht abgeschlossen ist, selbst wenn $\mathcal{S}$ abgeschlossen sein sollte. (Als Beispiel halte man sich

$$n = 1, \quad \mathcal{S} = \{x \in I\!R \mid x \geq 0\} \quad \text{und} \quad \mathcal{K} \cup \{0\} = \{(x, x_2) \in I\!R^2 \mid x \geq 0,\ x_2 > 0\} \cup \{(0,0)\}$$

vor Augen.) Wir definieren $\mathcal{K}$ als den Abschluss $\mathcal{K} = \overline{\widehat{\mathcal{K}}}$ von $\widehat{\mathcal{K}}$.

Da $\mathcal{S}$ abgeschlossen ist gilt nun

$$x \in \mathcal{S} \Longleftrightarrow \tilde{x} := (x, x_{n+1}) \in \mathcal{K} \text{ und } x_{n+1} = 1$$

und die Bedingung $x_{n+1} = 1$ ist natürlich linear. Somit kann für abgeschlossene konvexe Mengen $\mathcal{S}$ das Problem

$$\min\{\langle c, x\rangle \mid x \in \mathcal{S}\}$$

in ein **konisches konvexes Problem** der Form

$$(P) \qquad \min\{\langle \tilde{c}, \tilde{x}\rangle \mid \tilde{x} \in \mathcal{K},\ \ \tilde{x} \in \mathcal{L} + \tilde{b}\}$$

umgewandelt werden, wobei $\mathcal{K}$ ein abgeschlossener konvexer Kegel und $\mathcal{L}$ ein linearer Teilraum ist und $\tilde{b}$ irgendein Vektor, der die linearen Gleichungen (z.B. $x_{n+1} = 1$) erfüllt.

Das duale Problem ist hier völlig symmetrisch durch

$$(D) \qquad \min\{\langle \tilde{b}, \tilde{s}\rangle \mid \tilde{s} \in \mathcal{K}^D,\ \ \tilde{s} \in \mathcal{L}^\perp + \tilde{c}\}$$

gegeben. Dabei haben wir mit $\mathcal{L}^\perp$ den Senkrechtraum zu $\mathcal{L}$ bezeichnet, $\mathcal{L}^\perp = \{\tilde{s} \mid \langle \tilde{x}, \tilde{s}\rangle = 0 \ \ \forall\, \tilde{x} \in \mathcal{L}\}$. Es gilt nun folgender Dualitätssatz

Satz 6.1 *Sofern die Slaterbedingung "$\exists \tilde{x} \in \mathcal{K}^i \cap \mathcal{L} + \tilde{b}$" erüllt ist, und (P) eine Optimallösung $\tilde{x}^*$ besitzt, so besitzt auch (D) eine Optimallösung $\tilde{s}^*$ und die Optimalwerte erfüllen*

$$\langle \tilde{c}, \tilde{x}^*\rangle + \langle \tilde{b}, \tilde{s}^*\rangle = \langle \tilde{b}, \tilde{c}\rangle. \tag{73}$$

Der Beweis folgt am Ende dieses Abschnitts. Die Dualitätsbeziehung (73) ist von daher sehr ansprechend, dass hier das primale und das duale Problem völlig symmetrisch auftreten. Die Formulierung ist allerdings in gewisser Hinsicht "abstrakt"; typischerweise ist ein Optimierungsproblem in der Form 6.2 gegeben, und bei der Umformung in die Form (P) sind die Nebenbedingungen $f_i(x) \leq 0$ beispielsweise durch $f_i(x/x_{n+1}) \leq 0$ zu ersetzen, was etwas komplizierter aussieht. Trotzdem wurden mit dieser Formulierung bereits sehr effiziente Programme entwickelt, die konvexe Probleme der Form (6.2) lösen, siehe z.B. [1].

6.2 Semidefinite Programme

Eine weiteres Beispiel wo die Dualitätsbeziehung (73) in natürlicher Weise auftritt sind die sogenannten semidefinite Programme: In einer Reihe von Anwendungen treten—wie wir später noch sehen werden—Optimierungsprobleme auf, bei denen eine unbekannte symmetrische Matrix X so zu bestimmen ist, dass X positiv semidefinit ist und gegebene lineare Gleichungen erfüllt. Konkret sieht ein lineares semidefinites Programm wie folgt aus:

$$(SDP) \qquad \text{minimiere } \langle C, X\rangle \mid A(X) = b,\ X \succeq 0.$$

Dabei ist A eine lineare Abbildung von $\mathcal{S}^n$ nach $\mathbb{R}^m$ und $X \succeq 0$ bedeutet dass X im Kegel $X \in \mathcal{S}^n_+$ liegt. Mann kann sich die Abbildung A wie folgt vorstellen: Gegeben seien m symmetrische Matrizen $A^{(i)}$ $(1 \leq i \leq m)$. Dann ist

$$A(X) := \begin{pmatrix} \langle A^{(1)}, X \rangle \\ \vdots \\ \langle A^{(m)}, X \rangle \end{pmatrix} \tag{74}$$

eine lineare Abbildung von $\mathcal{S}^n$ nach $\mathbb{R}^m$. Wir wollen (73) nun auf das Problem (SDP) anwenden. Dazu definieren wir die lineare Abbildung A^* durch die Beziehung

$$\langle y, A(X) \rangle = \langle A^*(y), X \rangle \qquad \text{für alle } X, y. \tag{75}$$

(Falls X ein n-Vektor ist, d.h. $X \in \mathbb{R}^n$ und y ein m-Vektor, $y \in \mathbb{R}^m$, dann gilt für die Matrizen $\mathbf{A}$ und $\mathbf{A}^*$, die die Abbildungen A und A^* beschreiben stets $\mathbf{A}^* = \mathbf{A}^T$. Allgemein heißt A^* die adjungierte Abbildung. Im Fall (74) ist z.B.

$$A^*(y) = A^{(1)} y_1 + \ldots + A^{(m)} y_m.$$

Falls X eine Matrix ist kann man A zwar auch 'irgendwie' in Tableauform d.h. als Matrix $\mathbf{A}$ darstellen und A^* als die transponierte Matrix; diese Darstellung ist aber zeimlich technisch. Es ist an dieser Stelle wirklich von Vorteil hier etwas abstrakter mit den linearen Abbildungen A und A^* anstelle der Matrizen $\mathbf{A}$ und $\mathbf{A}^T$ zu arbeiten.)

Sei weiter B eine symmetrische Matrix mit $A(B) = b$ und $\mathcal{L} = \{X \mid A(X) = 0\}$. (Da hier nicht $B \succeq 0$ gefordert ist, ist es—mittels Gaußelimination—einfach ein solches B zu bestimmen sofern es existiert. Wenn es kein solches B gibt, so hat (SDP) auch keine Lösung.) Es ist dann

$$\mathcal{L} + B = \{X \mid A(X) = b\}$$

und

$$\mathcal{L}^\perp = \{S \mid S = A^*(y), \quad y \in \mathbb{R}^m\}.$$

(Für $S \in \mathcal{L}^\perp$ und $X \in \mathcal{L}$ gilt dann offenbar

$$\langle S, X \rangle = \langle A^*(y), X \rangle = \langle y, A(X) \rangle = 0$$

wie gefordert.) Setzen wir die Definition von $\mathcal{L} + B$ und

$$\mathcal{L}^\perp + C = \{S \mid S = A^*(y) + C, \quad y \in \mathbb{R}^m\}$$

in das Paar (P) und (D) ein, so erhalten wir unter Ausnutzung von $(\mathcal{S}^n_+)^D = \mathcal{S}^n_+$ als duales Programm von (SDP):

$$(DSDP) \qquad \text{minimiere } \langle B, S \rangle \mid S = A^*(y) + C, \ S \succeq 0.$$

Sofern die Slaterbedingung für (SDP) erfüllt ist, d.h. hier also, dass eine Matrix $X \succ 0$ mit $A(X) = b$ existiert, so folgt für die Optimalwerte X^*, S^* von (SDP) und $(DSDP)$ die Beziehung

$$\langle C, X^* \rangle + \langle B, S^* \rangle = \langle B, C \rangle$$

(sofern X^* existiert). Wegen

$$\langle B, S\rangle = \langle B, A^*(y) + C\rangle = \langle B, A^*(y)\rangle + \langle B, C\rangle = \langle A(B), y\rangle + \langle B, C\rangle = b^T y + \langle B, C\rangle$$

kann man als duales Programm von (SDP) auch das Problem

$$\text{minimiere } b^T y \mid S = A^*(y) + C,\ S \succeq 0 \tag{76}$$

definieren. Man beachte, dass sich für (76) der additive Term $\langle B, C\rangle$ in der Dualitätsbeziehung (73) wegkürzt, d.h. der Optimalwert von (76) stimmt genau mit dem von (SDP) überein. Wir formen dieses Resultat noch etwas um und fassen es als Satz zusammen:

Satz 6.2 *Sofern es ein $X \succ 0$ mit $A(X) = b$ gibt gilt*

$$\min\{\langle C, X\rangle \mid A(X) = b,\ X \succeq 0\} \quad = \quad \max\{b^T y \mid A^*(y) + S = C,\ S \succeq 0\}$$

mit der üblichen Konvention dass das Maximum einer Funktion über der leeren Menge $-\infty$ ist.

Wenn die Matrix X eine Diagonalmatrix ist, d.h. wenn die linearen Gleichungen $A(X) = b$ nur für Diagonalmatrizen X erfüllbar sind, dann kann man (SDP) als eine komplizierte Art auffassen um ein lineares Programm zu formulieruen. Der Dualitätssatz stimmt dann mit dem der linearen Programmierung überein (man überlege kurz dass das wirklich so ist!); allerdings gilt die hier hergeleitete Dualität nur unter der Voraussetzung der Slaterbedingung. Wir werden später noch auf dieses Paar dualer Programme zurückkommen.

6.3 Beweis von Satz 6.1

Wir zeigen nun die Dualitätsbeziehung (73). Der Beweis lässt sich mittels Satz 6.6 führen:

Sei $\tilde{x}^*$ optimal für (P) und sei $A \in \mathbb{R}^{m\times n}$ eine lineare Abbildung mit

$$\mathcal{L} = \{\tilde{x} \mid A\tilde{x} = 0\} \qquad \text{bzw.} \qquad \mathcal{L} + \tilde{b} = \left\{\tilde{x} \mid A(\tilde{x} - \tilde{b}) = 0\right\}.$$

Damit hat (P) die Form (6.2) mit $f(x) = \tilde{c}^T\tilde{x}$, $C = \mathcal{K}$, $p = 0$ (keine Ungleichungen) und

$$F(\tilde{x}) = (f_1(\tilde{x}), \ldots, f_m(\tilde{x}))^T = A(\tilde{x} - \tilde{b}).$$

Die Lagrangefunktion (6.5) für (P) ist

$$L:\ \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}:\quad L(\tilde{x}, \tilde{y}) = \langle \tilde{c}, \tilde{x}\rangle + \langle \tilde{y}, A(\tilde{x} - \tilde{b})\rangle.$$

Teil 2) des Satzes 6.6 besagt: $\exists \tilde{y}^* \in \mathbb{R}^m$ so dass $(\tilde{x}^*, \tilde{y}^*)$ Sattelpunkt von L auf $\mathcal{K} \times \mathbb{R}^m$ ist, d.h.

$$L(\tilde{x}^*, \tilde{y}) \le L(\tilde{x}^*, \tilde{y}^*) \le L(\tilde{x}, \tilde{y}^*) \qquad \forall (\tilde{x}, \tilde{y}) \in \mathcal{K} \times \mathbb{R}^m.$$

Setzt man die Lagrangefunktion ein, so besagt die zweite Ungleichung dass

$$\langle \tilde{c}, \tilde{x}^*\rangle + \langle \tilde{y}^*, A(\tilde{x}^* - \tilde{b})\rangle \le \langle \tilde{c}, \tilde{x}\rangle + \langle \tilde{y}^*, A(\tilde{x} - \tilde{b})\rangle \qquad \forall \tilde{x} \in \mathcal{K}$$

d.h.

$$0 \ge \langle \tilde{c}, \tilde{x}^* - \tilde{x}\rangle + \langle \tilde{y}^*, A(\tilde{x}^* - \tilde{x})\rangle = \langle \tilde{c} + A^*\tilde{y}^*, \tilde{x}^* - \tilde{x}\rangle \quad \forall \tilde{x} \in \mathcal{K}.$$

Setzen wir $\tilde{s}^* := \tilde{c} + A^*\tilde{y}^*$ so besagt die letzte Ungleichung

$$\langle \tilde{s}^*, \tilde{x} - \tilde{x}^* \rangle \geq 0 \quad \forall \tilde{x} \in \mathcal{K}. \tag{77}$$

Nun ist mit $\tilde{x} \in \mathcal{K}$ auch $\lambda\tilde{x} \in \mathcal{K}$ für alle $\lambda \geq 0$ (weil $\mathcal{K}$ ein Kegel ist). Aus (77) folgt dann

$$\langle \tilde{s}^*, \tilde{x} \rangle \geq 0 \quad \forall \tilde{x} \in \mathcal{K},$$

d.h. $\tilde{s}^* \in \mathcal{K}^D$. Wählt man $\tilde{x} = 0$ in (77), so folgt weiter $\langle \tilde{s}^*, \tilde{x}^* \rangle \leq 0$. Wegen $\tilde{s}^* \in \mathcal{K}^D$ muss daher $\langle \tilde{s}^*, \tilde{x}^* \rangle = 0$ gelten. Wegen

$$\langle A^*\tilde{y}^*, \tilde{x} \rangle = \langle \tilde{y}^*, A\tilde{x} \rangle = \langle \tilde{y}^*, 0 \rangle$$

für alle $\tilde{x} \in \mathcal{L}$ ist außerdem $A^*\tilde{y}^* \in \mathcal{L}^\perp$ und $\tilde{s}^* \in \mathcal{L}^\perp + \tilde{c}$. Also ist $\tilde{s}^*$ zulässig für das Problem

$$\min\{\langle \tilde{s}, \tilde{x}^* \rangle \mid \tilde{s} \in \mathcal{K}^D \cap \mathcal{L}^\perp + \tilde{c}\}. \tag{78}$$

Wegen $\tilde{x}^* \in \mathcal{K}$ und $\langle \tilde{s}, \tilde{x}^* \rangle \geq 0$ für alle $\tilde{s} \in \mathcal{K}^D$ ist $\tilde{s}^*$ sogar optimal für (78).

Man beachte nun, dass wegen $\tilde{x}^* - \tilde{b} \in \mathcal{L}$ folgt $A\tilde{x}^* = A\tilde{b}$ und daher ist für jedes $\tilde{s} \in \mathcal{L}^\perp + \tilde{c}$, d.h. für jedes $\tilde{s} = A^*\tilde{y} + \tilde{c}$

$$\langle \tilde{s}, \tilde{x}^* \rangle = \langle A^*\tilde{y} + \tilde{c}, \tilde{x}^* \rangle = \langle A^*\tilde{y}, \tilde{x}^* \rangle + \langle \tilde{c}, \tilde{x}^* \rangle$$

$$= \langle \tilde{y}, A\tilde{x}^* \rangle + \langle \tilde{c}, \tilde{x}^* \rangle = \langle \tilde{y}, A\tilde{b} \rangle + \langle \tilde{c}, \tilde{b} \rangle + \langle \tilde{c}, \tilde{x}^* - \tilde{b} \rangle$$

$$= \langle A^*\tilde{y}, \tilde{b} \rangle + \langle \tilde{c}, \tilde{b} \rangle + \langle \tilde{c}, \tilde{x}^* - \tilde{b} \rangle = \langle A^*\tilde{y} + \tilde{c}, \tilde{b} \rangle + \langle \tilde{c}, \tilde{x}^* - \tilde{b} \rangle = \langle \tilde{s}, \tilde{b} \rangle + \langle \tilde{c}, \tilde{x}^* - \tilde{b} \rangle.$$

Vergleichen wir den ersten und den letzten Term in obiger Kette von Gleichungen, dann sehen wir dass $\tilde{s}^*$ nicht nur für (78) sondern auch für (D) optimal ist (der additive Term $\langle \tilde{c}, \tilde{x}^* - \tilde{b} \rangle$ hängt gar nicht von der Wahl von $\tilde{s}$ ab). Setzen wir $\tilde{s} = \tilde{s}^*$ in obiger Kette ein, so folgt

$$0 = \langle \tilde{s}^*, \tilde{x}^* \rangle = \langle \tilde{s}^*, \tilde{b} \rangle + \langle \tilde{c}, \tilde{x}^* - \tilde{b} \rangle.$$

Formulieren wir diese letzte Gleichung um, so erhalten wir

$$\langle \tilde{s}^*, \tilde{b} \rangle + \langle \tilde{c}, \tilde{x}^* \rangle = \langle \tilde{c}, \tilde{b} \rangle,$$

was zu zeigen war. #

7 Optimalitätsbedingungen für allgemeine Optimierungsprobleme

Sei $\mathcal{S} \subseteq \mathbb{R}^n$, $\mathcal{S} \neq \emptyset$, $f: \mathcal{S} \to \mathbb{R}$. Wir betrachten das Problem, eine Lösung von

$$\min\{f(x) \mid x \in \mathcal{S}\}$$

zu finden.

Definition 7.1: $\bar{x} \in \mathcal{S}$ heißt *globales* (bzw. *lokales*) *Minimum* von f auf $\mathcal{S}$ (oder *globale* (bzw. *lokale*) *Optimallösung*), falls $f(\bar{x}) \leq f(x)$ für alle $x \in \mathcal{S}$ (bzw. für alle $x \in \mathcal{S} \cap U_\delta(\bar{x})$ mit einem $\delta > 0$.)

Satz 7.2: Sei $\mathcal{S} \subseteq \mathbb{R}^n$ konvex, $f: \mathcal{S} \to \mathbb{R}$ konvex und $\bar{x} \in \mathcal{S}$ ein lokales Minimum von f auf $\mathcal{S}$. Dann ist $\bar{x}$ auch globales Minimum von f auf $\mathcal{S}$. Falls f streng konvex ist, dann ist $\bar{x}$ eindeutig.

Beweis: Übung.

Definition 7.3: Für $\mathcal{S} \subseteq \mathbb{R}^n$ und $\bar{x} \in \mathcal{S}$ heißt

$$T(\mathcal{S}, \bar{x}) := \{s \in \mathbb{R}^n \quad | \quad \exists \{\lambda_m\}_m,\ \exists \{x_m\}_m \colon\ \lambda_m \geq 0,\ x_m \in \mathcal{S},\ \lim_{m\to\infty} x_m = \bar{x},\ \lim_{m\to\infty} \lambda_m (x_m - \bar{x}) = s\}$$

Tangentialkegel von $\mathcal{S}$ in $\bar{x}$.

Anschaulich heißt das folgendes: Wenn es eine Punktfolge x^m in $\mathcal{S}$ gibt, die sich $\bar{x}$ beliebig nahe annähert, dann ist die Richtung $s := \lim_{m\to\infty} \frac{x^m - \bar{x}}{\|x^m - \bar{x}\|}$ aus der x^m sich auf $\bar{x}$ zubewegt (und alle positiven Vielfachen davon) in $T(\mathcal{S}, \bar{x})$. (Falls $\lim_{m\to\infty} \frac{x^m - \bar{x}}{\|x^m - \bar{x}\|}$ nicht existiert, so liegen alle konvergenten Teilfolgen von $\frac{x^m - \bar{x}}{\|x^m - \bar{x}\|}$ in $T(\mathcal{S}, \bar{x})$.)

Satz 7.4: $T(\mathcal{S}, \bar{x})$ ist ein abgeschlossener Kegel.

Beweis:
Kegeleigenschaft: Falls $s \in T(\mathcal{S}, \bar{x})$, so ist $\lambda s \in T(\mathcal{S}, \bar{x})$ für alle $\lambda > 0$ (Siehe Definition!).
Abgeschlossenheit: Sei $s_n \in T(\mathcal{S}, \bar{x})$ mit $s_n \to \bar{s}$ $(n \to \infty)$. Dann ist o.B.d.A. $\|s_n - \bar{s}\| \leq \frac{1}{n}$ für alle n (bilde nötigenfalls eine Teilfolge!). Da $s_n \in T(\mathcal{S}, \bar{x})$, gibt es für jedes n eine Folge $(\lambda_{n,j})_j$ und eine Folge $(x_{n,j})_j \subset \mathcal{S}$ mit $\lim_{j\to\infty} x_{n,j} = \bar{x}$, $\lim_{j\to\infty} \lambda_{n,j}(x_{n,j} - \bar{x}) = s_n$ und $\lambda_{n,j} \geq 0$. Wir wählen nun $j(n)$ so groß, daß für $j \geq j(n)$ der Abstand $\|x_{n,j} - \bar{x}\| \leq \frac{1}{n}$ ist und auch $\|\lambda_{n,j}(x_{n,j} - \bar{x}) - s_n\| \leq \frac{1}{n}$, also $\|\bar{s} - \lambda_{n,j(n)}(x_{n,j(n)} - \bar{x})\| \leq \frac{2}{n}$ (Dreiecksungleichung) und $\|x_{n,j(n)} - \bar{x}\| \leq \frac{1}{n}$. Somit beweisen die Folgen $(\lambda_{n,j(n)})_n$ und $(x_{n,j(n)})_n$, daß $\bar{s} \in T(\mathcal{S}, \bar{x})$.

Satz 7.5: Sei $\bar{x}$ ein lokales Minimum von f auf $\mathcal{S}$. Sei $f \in C^1(\bar{x})$, d. h. $f(x) = f(\bar{x}) + Df(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|)$ für alle $x \in \mathrm{dom}(f)$. Dann gilt: $Df(\bar{x})s \geq 0$ für alle $s \in T(\mathcal{S}, \bar{x})$.

Beweis: Sei $s \in T(\mathcal{S}, \bar{x})$, $s = \lim_{n\to\infty} \lambda_n(x_n - \bar{x})$ mit $x_n \in \mathcal{S}$, $\lim_{n\to\infty} x_n = \bar{x}$. Weil $\bar{x}$ lokales Minimum ist folgt $f(\bar{x}) \leq f(x)$ für $x \in U_\delta(\bar{x}) \cap \mathcal{S}$ mit einem $\delta > 0$. Für genügend große n ist auch $\|x_n - \bar{x}\| \leq \delta$, also $f(\bar{x}) \leq f(\bar{x}) + Df(\bar{x})(x_n - \bar{x}) + \|x_n - \bar{x}\| \cdot o(1)$, d. h. $0 \leq Df(\bar{x}) \cdot \lambda_n(x_n - \bar{x}) + \lambda_n \|x_n - \bar{x}\| \cdot o(1)$. Mit $n \to \infty$ folgt $0 \leq Df(\bar{x})s$.

7.6 Korollar: Falls $\bar{x} \in \mathcal{S}^o$, so ist $T(\mathcal{S}, \bar{x}) = \mathbb{R}^n$ und falls $\bar{x} \in S^\circ$ lokales Minimum ist, so ist $Df(\bar{x}) = 0$.

Betrachte nun

$$(P'): \qquad \min\{f(x) \mid f_i(x) \leq 0, \quad 1 \leq i \leq p, \quad f_j(x) = 0, \quad p+1 \leq j \leq m\}$$

Dies ist äquivalent zu

$$\min\{f(x) \mid F(x) \in -\mathcal{K}\}$$

mit $F(x) = (f_1(x), \ldots, f_m(x))^T$ und

$$\mathcal{K} = \{u \in \mathbb{R}^m \mid u_i \geq 0, \quad 1 \leq i \leq p, \quad u_j = 0, \quad p+1 \leq j \leq m\}$$

Offenbar ist $\mathcal{K}$ ein abgeschlossener konvexer Kegel.

Sei $\mathcal{K}$ ein beliebiger abgeschlossener konvexer Kegel. Definiere

$$u \leq_{\mathcal{K}} 0 \Longleftrightarrow u \in -\mathcal{K},$$

$$u \geq_{\mathcal{K}} 0 \Longleftrightarrow u \in \mathcal{K}.$$

(Für $\mathcal{K} = \{x \in \mathbb{R}^n \mid x \geq 0\}$ ist „ $\leq_{\mathcal{K}}$ “ das übliche „$\leq$“.)

7.7 Bemerkung:

$$\left.\begin{array}{l} u \leq_{\mathcal{K}} 0 \\ v \leq_{\mathcal{K}} 0 \end{array}\right\} \Longrightarrow \lambda u + \mu v \leq_{\mathcal{K}} 0 \quad \text{für alle} \quad \lambda, \mu > 0.$$

In Verallgemeinerung zu (P') (und zur Vereinfachung der Notation) gelte nun die Voraussetzung

(V) $C \subseteq \mathbb{R}^n$ abgeschlossen, konvex, $\neq \emptyset$
$\mathcal{K} \subseteq \mathbb{R}^n$ abgeschlossener, konvexer Kegel, $\neq \emptyset$
$F\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, $f, F \in C^1(\mathbb{R}^n)$.

Wir betrachten

$$(P) \qquad \min\{f(x) \mid x \in C, \quad F(x) \leq_{\mathcal{K}} 0\} = \min\{f(x) \mid x \in \mathcal{S}\}$$

mit $\mathcal{S} := \{x \in C \mid F(x) \leq_{\mathcal{K}} 0\}$ (d. h. $F(x) + k = 0$ für ein $k \in \mathcal{K}$).
Sei $\bar{x}$ eine lokale Optimallösung von (P). Betrachte das zu (P) assoziierte linearisierte Problem (P_L):

$$(P_L): \qquad \min\{f(\bar{x}) + Df(\bar{x})(x - \bar{x}) \mid F(\bar{x}) + DF(\bar{x})(x - \bar{x}) \leq_{\mathcal{K}} 0, \quad x \in C\}$$

bzw. für den Spezialfall (P'):

$$(P'_L): \qquad \min\{f(\bar{x}) + Df(\bar{x})(x - \bar{x}) \quad | \quad \begin{array}{l} f_i(\bar{x}) + Df_i(\bar{x})(x - \bar{x}) \leq 0, \quad 1 \leq i \leq p, \\ f_j(\bar{x}) + Df_j(\bar{x})(x - \bar{x}) = 0, \quad p+1 \leq j \leq m\} \end{array}$$

Sei $\mathcal{S}_L := \{x \in C \mid F(\bar{x}) + DF(\bar{x})(x - \bar{x}) \leq_{\mathcal{K}} 0\}$ die zulässige Menge des linearisierten Problems (P_L) und

$$\begin{array}{rcll} L(\mathcal{S}, \bar{x}) : & = & \{s \in \mathbb{R}^n \mid \exists \lambda \geq 0, \ \exists x \in C\colon \ s = \lambda(x - \bar{x}), \ F(\bar{x}) + DF(\bar{x})(x - \bar{x}) \leq_{\mathcal{K}} 0\} & \\ & = & \{s \in \mathbb{R}^n \mid \exists \lambda \geq 0, \ \exists x \in \mathcal{S}_L\colon s = \lambda(x - \bar{x})\} & \textbf{(7.8)} \end{array}$$

der linearisierte Kegel von $\mathcal{S}$ in $\bar{x}$. Dann ist

$$L(\mathcal{S}, \bar{x}) = \bigcup_{\lambda \geq 0} \lambda(\mathcal{S}_L - \bar{x}) \subseteq T(\mathcal{S}_L, \bar{x}).$$

“Der Tangentialkegel der linearisierten Menge enthält den linearisierten Kegel.”

Beweis: Sei $s \in L(\mathcal{S}, \bar{x})$. Dann ist $s = \lambda(x - \bar{x})$ mit $x \in C$ und $F(\bar{x}) + DF(\bar{x})(x - \bar{x}) \leq_{\mathcal{K}} 0$, und einem $\lambda \geq 0$. Es ist $F(\bar{x}) \leq_{\mathcal{K}} 0$, $\bar{x} \in C$, da $\bar{x}$ zulässig für (P) ist. Setze $x_n := \frac{1}{n}x + \frac{n-1}{n}\bar{x}$.

Da C konvex ist, ist $x_n \in C$. Dann ist $s = n\lambda(x_n - \bar{x})$ und $\lim_{n\to\infty} x_n = \bar{x}$. Außerdem ist $x_n \in \mathcal{S}_L$, denn

$$F(\bar{x}) + DF(\bar{x})(x_n - \bar{x}) = \frac{1}{n}\underbrace{(F(\bar{x}) + DF(\bar{x})(x - \bar{x}))}_{\leq_{\mathcal{K}} 0} + \frac{n-1}{n}\underbrace{F(\bar{x})}_{\leq_{\mathcal{K}} 0} \leq_{\mathcal{K}} 0.$$

Also ist $s \in T(\mathcal{S}_L, \bar{x})$. □

Falls $\mathcal{K}$ kein polyedrischer Kegel ist, kann $L(\mathcal{S}, \bar{x}) \neq T(\mathcal{S}_L, \bar{x})$ vorkommen. (Man prüfe dies am Beispiel $\mathcal{K} = \{z \in \mathbb{R}^3 : \sqrt{z_1^2 + z_2^2} \leq z_3\}$ und $F(x) = (x_1, x_2, -1)^T$ nach.)

Spezialfall: (P'_L):

$$L(\mathcal{S}, \bar{x}) = \{s \in \mathbb{R}^n \mid \begin{array}{ll} Df_i(\bar{x})s \leq 0 & \text{für alle } i \in I(\bar{x}), \\ Df_j(\bar{x})s = 0 & \text{für alle } j \geq p+1\} \end{array}$$

mit $I(\bar{x}) = \{i \in \{1, \ldots, p\} \mid f_i(\bar{x}) = 0\}$ (alle in $\bar{x}$ „aktiven" Ungleichungen).

Relation zwischen $T(\mathcal{S}, \bar{x})$ und $L(\mathcal{S}, \bar{x})$:
Falls $L(\mathcal{S}, \bar{x}) \subseteq T(\mathcal{S}, \bar{x})$, dann gilt für alle $s \in L(\mathcal{S}, \bar{x})$ die Ungleichung $Df(\bar{x})s \geq 0$, sofern $\bar{x}$ ein lokales Minimum von (P) ist, d. h. $\bar{x}$ ist dann auch lokales Minimum von (P_L). Da (P_L) konvex ist, ist $\bar{x}$ globales Minimum von (P_L). Diese Beobachtung ist für $L(\mathcal{S}, \bar{x}) \not\subseteq T(\mathcal{S}, \bar{x})$ jedoch leider nicht richtig:

Beispiel:

$$\mathcal{S} = \left\{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_2 \geq 0,\ x_2 \leq x_1^3\right\}, \quad \bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hier ist $L(\mathcal{S}, \bar{x}) = \{\binom{x_1}{x_2} \mid x_2 = 0\}$ und $T(\mathcal{S}, \bar{x}) = \{\binom{x_1}{x_2} \mid x_2 = 0 \text{ und } x_1 \geq 0\}$.

Wir versuchen daher, eine Bedingung zu identifizieren, die $L(\mathcal{S}, \bar{x}) \subset T(\mathcal{S}, \bar{x})$ impliziert.

Definition (7.10): Regularitätsbedingung von Robinson. (P) bzw. $\mathcal{S}$ heißt *regulär* in $\bar{x} \in \mathcal{S} :\Longleftrightarrow 0 \in (F(\bar{x}) + DF(\bar{x})(C - \bar{x}) + \mathcal{K})^\circ$, d. h. 0 ist im Inneren von $F(\bar{x}) + DF(\bar{x})(C - \bar{x}) + \mathcal{K}$.

Die Regularitätsbedingung ist hier sehr abstrakt formuliert und wird im Anschluß an Satz 7.13 noch erläutert werden. Wir stellen zunächst zwei zentrale Sätze vor.

Satz (7.11): Seien (V) und (7.10) erfüllt. Dann ist $L(\mathcal{S}, \bar{x}) \subseteq T(\mathcal{S}, \bar{x})$, und insbesondere ist $\bar{x}$ globales Minimum von (P_L), falls $\bar{x}$ lokales Minimum von (P) ist.

Der Beweis von Satz 7.11 folgt am Ende des Kapitels.

(7.12) Satz von Kuhn und Tucker für (P):

Voraussetzungen: Es gelte (V), d. h. C sei abgeschlossen und konvex, $\mathcal{K}$ sei ein abgeschlossener und konvexer Kegel, $f, F \in C^1(\mathbb{R}^n)$. Weiter sei $\bar{x}$ lokales Optimum von (P), und (P) sei in $\bar{x}$ regulär. Dann gibt es ein $\bar{y} \in \mathbb{R}^m$ mit

a) $\bar{y} \in -\mathcal{K}^P = \{y \in \mathbb{R}^m \mid y^T x \geq 0\ \forall x \in \mathcal{K}\}$, dem polaren Kegel von $\mathcal{K}$

b) $\bar{y}^T F(\bar{x}) = 0$ (complementary slackness)

c) $D_x L(\bar{x}, \bar{y})(x - \bar{x}) = [Df(\bar{x}) + \bar{y}^T DF(\bar{x})](x - \bar{x}) \geq 0 \quad \forall x \in C.$

Jedes y mit a)–c) heißt *Kuhn-Tucker-Vektor* oder *Lagrangemultiplikator.*

Beweis: [vgl. Beweis von Satz (6.3)]
Nach Satz 7.11 folgt aus der Regularitätsbedingung daß $\bar{x}$ globales Minimum von (P_L) ist. Setze

$$A := \Big\{ \begin{pmatrix} v_0 \\ v \end{pmatrix} \in \mathbb{R}^{m+1} \mid \exists x \in C,\ \exists k \in \mathcal{K}: \quad v_0 > Df(\bar{x})(x-\bar{x}) \\ v = F(\bar{x}) + DF(\bar{x})(x-\bar{x}) + k \Big\}$$

(d. h. $v \geq_{\mathcal{K}} F(\bar{x}) + DF(\bar{x})(x-\bar{x})$.)
Es folgt: A ist konvex und $0 \notin A$
Also kann 0 eigentlich von A getrennt werden, d.h.

$$\exists (z_0, z^T)^T \in \mathbb{R}^{m+1}\ \forall (v_0, v^T)^T \in A\colon z_0 v_0 + z^T v \geq 0 \quad \text{(i)}$$
$$\exists (\bar{v}_0, \bar{v}^T)^T \in A\colon z_0 \bar{v}_0 + z^T \bar{v} > 0 \quad \text{(ii)}.$$

Wegen (i) und da v_0 in A nach oben unbeschränkt ist, folgt $z_0 \geq 0$. Wegen (i) und nach Definition von A ist $z \in -\mathcal{K}^P$. (Ersetze $\bar{v}$ durch $\bar{v} + \lambda k$, betrachte $\lambda \to +\infty \Longrightarrow z^T k \geq 0$).
Falls $z_0 \neq 0$, setze $\bar{y} = \frac{z}{z_0} \in -\mathcal{K}^P$. Dann ist $Df(\bar{x})(x-\bar{x}) + \bar{y}^T(F(\bar{x}) + DF(\bar{x})(x-\bar{x})) \geq 0$ für alle $x \in C$. Daraus folgt (c).
Für $x = \bar{x}$ folgt $\bar{y}^T F(\bar{x}) \geq 0$. Andererseits: $\bar{y}^T F(\bar{x}) \leq 0$ (da $F(\bar{x}) \in -\mathcal{K}$), also $\bar{y}^T F(\bar{x}) = 0$. Daraus folgt (b).
Falls $z_0 = 0$ so ist $z \neq 0$ und dann folgt $z^T v \geq 0$ für alle $v \in M := F(\bar{x}) + DF(\bar{x})(C - \bar{x}) + \mathcal{K}$. Mit (ii): $z^T \bar{v} > 0$ (mit $\bar{v} \in M$) folgt ein Widerspruch denn (7.10) besagt $0 \in M^\circ$, d.h. $\exists \tilde{v} \in M\colon z^T \tilde{v} < 0$. Daraus folgt die Behauptung.

Spezialfall: Satz von Kuhn-Tucker für (P'):
Sei $C = \mathbb{R}^n$, $\mathcal{K} = \{y \in \mathbb{R}^m \mid y_1, \ldots, y_p \geq 0, y_{p+1} = \ldots = y_m = 0\}$, und es gelte:

1) $f, F \in C^1(\mathbb{R}^n)$

2) $\bar{x}$ sei eine lokale Optimallösung von (P').

3) (P') sei regulär in $\bar{x}$

Dann gibt es ein $\bar{y} \in \mathbb{R}^m$ mit

a) $\bar{y}_i \geq 0$ für $1 \leq i \leq p$

1. b) $\bar{y} f_i(\bar{x}) = 0$ für $1 \leq i \leq m$ (complementary slackness)

2. c) $D_x L(\bar{x}, \bar{y}) = Df(\bar{x}) + \bar{y}^T\, DF(\bar{x}) = 0$

Die Regularitätsbedingung (7.10) ist auch mit $\operatorname{aff}(C)$ und $\operatorname{aff}(K)$ durch Gleichungssysteme beschreibbar. Sei hierzu

$$\begin{aligned} \operatorname{aff}(C) &= \{x \in \mathbb{R}^n \mid Gx = g\}, \quad G \in \mathbb{R}^{k \times n}, \quad \operatorname{rang}(G) = k = n - \dim \operatorname{aff}(C) \\ \operatorname{aff}(\mathcal{K}) &= \{x \in \mathbb{R}^m \mid Hx = 0\}, \quad H \in \mathbb{R}^{l \times m}, \quad \operatorname{rang}(H) = l = m - \dim \operatorname{aff}(\mathcal{K}). \end{aligned}$$

(Beachte, $\operatorname{aff}(\mathcal{K})$ ist ein linearer Raum, d.h. $0 \in \operatorname{aff}(\mathcal{K})$.)

Satz (7.13): (P) ist in $\bar{x}$ regulär genau dann, wenn

1) $\operatorname{rang}\begin{pmatrix} H\,DF(\bar{x}) \\ G \end{pmatrix} = k + l$

2) $\exists x_1 \in C^i$: $F(\bar{x}) + DF(\bar{x})(x_1 - \bar{x}) \in -\mathcal{K}^i$

Beweis:

"$\Longrightarrow$": Sei (P) in $\bar{x}$ regulär.

1) Annahme: $\begin{pmatrix} H\,DF(\bar{x}) \\ G \end{pmatrix}$ habe nicht maximalen Rang. Dann gibt es ein Paar $(s,t) \neq (0,0)$: $s^T H\,DF(\bar{x}) + t^T G = 0$. Nun ist $s^T H \neq 0$, da sonst $s = 0$ wäre (denn H hat vollen Zeilenrang) und dann aus $t^T G = 0$, also $t = 0$ (weil G ebenfalls vollen Zeilenrang hat), ein Widerspruch zu $(s,t) \neq (0,0)$ folgen würde. Also gilt:

$$s^T H(F(\bar{x}) + DF(\bar{x})(C - \bar{x}) + \mathcal{K}) \quad \underbrace{\overset{-F(\bar{x})\in\mathcal{K},\quad H\mathcal{K}=0}{=}} \; s^T H\,DF(\bar{x})(C - \bar{x})$$
$$= \quad -t^T G(C - \bar{x})$$
$$\underbrace{\overset{G(C-\bar{x})=0}{=}} \; 0$$

Wegen (7.10) gibt es eine Kugel B mit $0 \in B \subset M := F(\bar{x}) + DF(\bar{x})(C - \bar{x}) + \mathcal{K}$. Die vorangegangene Überlegung besagt daher $s^T H u = 0$ für alle $u \in B$ und damit auch für alle $u \in \mathbb{R}^n$. Also ist $s^T H = 0$, was den gewünschten Widerspruch liefert.

2) Wegen (7.10) ist $0 \in M^\circ \subset M^i = F(\bar{x}) + DF(\bar{x})(C^i - \bar{x}) + \mathcal{K}^i$. (Beweis von (5.10))
Anders ausgedrückt: $\exists\, x_1 \in C^i$ mit $0 \in F(\bar{x}) + DF(\bar{x})(x_1 - \bar{x}) + \mathcal{K}^i$, was zu zeigen war.

"$\Longleftarrow$": Es mögen (1) und (2) gelten. Aus (2) folgt $0 \in M^i$. Wäre $M^\circ \neq M^i$, so wäre $\dim \operatorname{aff}(M) < m$, und es gäbe ein $v \in \mathbb{R}^m$, $v \neq 0$ mit $v^T M = 0$. Also insbesondere $v^T F(\bar{x}) = 0$ und

a) $v^T\,DF(\bar{x})(C - \bar{x}) = 0 \Longrightarrow v^T\,DF(\bar{x})(x - \bar{x}) = 0$ für alle $x \in \operatorname{aff}(C)$.

b) $v^T \mathcal{K} = 0 \Longrightarrow v^T k = 0$ für alle $k \in \operatorname{aff}(\mathcal{K})$.

Aus (a) folgt: Falls $G\tilde{x} = 0$ (d. h. $x = \tilde{x} + \bar{x} \in \operatorname{aff}(C)$) so ist $v^T\,DF(\bar{x})\tilde{x} = 0$. Also wird $v^T DF(\bar{x})$ von den Zeilen von G aufgespannt, d. h. $\exists t \in \mathbb{R}^k$: $v^T\,DF(\bar{x}) = t^T G$.

Aus (b) folgt: Falls $Hk = 0$, so ist $v^T k = 0$. Also $\exists s \in \mathbb{R}^l$: $v^T = s^T H$. Wegen $v \neq 0$ ist $s \neq 0$ und somit $s^T H\,DF(\bar{x}) = v^T\,DF(\bar{x}) = t^T G$. Also ist

$$(s, -t)^T \begin{pmatrix} H\,DF(\bar{x}) \\ G \end{pmatrix} = 0$$

und daraus folgt wegen (1), daß $(s,-t) = (0,0)$ und somit der gesuchte Widerspruch. □

Spezialfall (P'): Sei

$$\bar{x} \in \mathcal{S} := \{x \in \mathbb{R}^n \mid f_i(x) \leq 0, \quad i = 1, \ldots, p, \quad f_j(x) = 0, \quad j = p+1, \ldots, m\}.$$

$\mathcal{S}$ ist in $\bar{x}$ regulär genau dann, wenn die Gradienten $Df_{p+1}(\bar{x}), \ldots, Df_m(\bar{x})$ linear unabhängig sind und es ein $s \in \mathbb{R}^n$ gibt mit $Df_j(\bar{x})s = 0$ für $p+1 \leq j \leq m$, und $Df_i(\bar{x})s < 0$, für die aktiven Indices $i \in I \subset \{1, \ldots, p\}$ (d.h. die Indices $i \leq p$ mit $f_i(\bar{x}) = 0$). Hinreichend hierfür ist, daß die $Df_i(\bar{x})$ für $i \in I \cup \{p+1, \ldots, m\}$ linear unabhängig sind.

Zur Begründung untersuchen wir die Frage: Wann gilt $0 \in (F(\bar{x}) + DF(\bar{x})(\mathbb{R}^n - \bar{x}) + \mathcal{K})^\circ$ mit $F = (f_1, \ldots, f_m)^T$ und

$$\mathcal{K} = \{u \in \mathbb{R}^m \mid u_i \geq 0, \quad i = 1, \ldots, p, \quad u_j = 0, \quad j = p+1, \ldots, m\}?$$

Hier ist $C = \mathbb{R}^n$ (G= leere Abbildung), $\operatorname{aff}(\mathcal{K}) = \{u \in \mathbb{R}^m \mid Hu = 0\}$ mit

$$H = \begin{pmatrix} 0 & \cdots & 0 & 1 & & \\ \vdots & & \vdots & & \ddots & \\ 0 & \cdots & 0 & & & 1 \end{pmatrix}$$
$$\begin{matrix} \uparrow & & \uparrow & \uparrow & & \uparrow \\ 1 & & p & p{+}1 & & m \end{matrix}$$

1) Es ist $\operatorname{rang}(H\, DF(\bar{x})) = l = m - p$ genau dann, wenn die Gradienten der Gleichungsbedingungen $Df_{p+1}(\bar{x}), \ldots, Df_m(\bar{x})$ linear unabhängig sind.

2) Die zweite Bedingung aus Satz 7.13 besagt: Es existiert ein $x_1 \in \mathbb{R}^n (= C^i)$ mit

$$F(\bar{x}) + DF(\bar{x})(x_1 - \bar{x}) \in -\mathcal{K}^i$$

Dies ist genau dann der Fall wenn es ein $s \in \mathbb{R}^n$ gibt mit $Df_j(\bar{x})s = 0$ für $p+1 \leq j \leq m$, und $Df_i(\bar{x})s < 0$ für $i \in I$. Um die hinreichende Bedingung zu sehen, ignorieren wir die inaktiven Ungleichungen und nehmen o.B.d.A. an, daß $I = \{1, \ldots, p\}$. Wir unterteilen $DF(\bar{x}) = (M_1, M_2)$ wobei M_1 nach Voraussetzung so gewählt werden kann, daß M_1 quadratisch und nichtsingulär ist. Wähle u mit $u_I < 0$, $u_{\{p+1,\ldots,m\}} = 0$, dann liefert $s_1 := M_1^{-1}u$, $s_2 := 0$, $s := (s_1, s_2)^T$ die gewünschten Beziehungen.

Beispiel:
Sei $f(x) = x_1 + x_2$ und

$$F(x) = \begin{pmatrix} -x_2 \\ x_2 - x_1^3 \end{pmatrix} \overset{!}{\leq} 0, \quad \bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Die Menge $\mathcal{S}$ ist dann das Gebiet oberhalb der x_1-Achse und unterhalb der Kurve $x_2 = x_1^3$. Dann ist $\bar{x}$ Minimum von f auf $\mathcal{S} = \{x \mid F(x) \leq 0\}$ und

$$T(\mathcal{S}, \bar{x}) = (\mathbb{R}^+ \cup \{0\}) \times \{0\} = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;\middle|\; x_1 \geq 0,\ x_2 = 0 \right\}.$$

Es ist $\nabla f(x) = (1, 1)^T$ und

$$DF(x) = \begin{pmatrix} 0 & -1 \\ -3x_1^2 & 1 \end{pmatrix}.$$

Daher ist

$$F(\bar{x}) + DF(\bar{x})(x - \bar{x}) \leq_{\mathcal{K}} 0 \Longleftrightarrow \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq 0.$$

Für $L(\mathcal{S}, \bar{x})$ ergibt sich somit $L(\mathcal{S}, \bar{x}) = \mathbb{R} \times \{0\}$. Offenbar gilt nicht „$Df(\bar{x})s \geq 0$ für alle $s \in L(\mathcal{S}, \bar{x})$“, d.h. $\bar{x}$ ist nicht die Optimallösung des linearisierten Problems. (Dies ist weil die zulässige Menge im Optimalpunkt nicht regulär ist.)

Mit der zusätzlichen Restriktion $x_1 \geq 0$ erreicht man $L(\mathcal{S}, \bar{x}) = T(\mathcal{S}, \bar{x})$. In diesem Falle ist $Df(\bar{x})s \geq 0$ für alle $s \in L(\mathcal{S}, \bar{x})$. Nun ist $s \in L(\mathcal{S}, \bar{x})$ genau dann wenn $s = \lambda(x - \bar{x}) = \lambda x$ mit $\lambda \geq 0$ und $DF(\bar{x})x \leq 0$. Mit dem Lemma von Farkas: $(A^T y \leq 0 \Longrightarrow b^T y \leq 0) \Longleftrightarrow (\exists w \geq 0$: $b = Aw)$ folgt die Existenz eines Multiplikators $y \geq 0$ mit $Df(\bar{x}) + y^T DF(\bar{x}) = 0$, d.h. die Aussage aus Satz 7.12 gilt “ausnahmsweise”, obwohl die zulässige Menge im Optimalpunkt immer noch nicht regulär ist im Sinne von Robinson.

Die Bedingung 7.11 bzw. 7.13 verlangt mehr als die Slaterbedingung. Naiv ausgedrückt ist 7.11 für eine größere Problemklasse definiert als Slaterbedingung. Um auch in der größeren Problemklasse alle “pathologischen Fälle” auszuschließen ist eine schärfere Bedingung nötig.

Hilfssatz (7.14): Sei (P) in $\bar{x}$ regulär im Sinne von (7.10), d. h. insbesondere $\exists x^1 \in C^i$: $F(\bar{x}) + DF(\bar{x})(x^1 - \bar{x}) \in -\mathcal{K}^i$. Dann gilt für jedes solche x^1 und $s := x^1 - \bar{x}$:

1) Es gibt ein $\varepsilon > 0$ und eine auf $[0; \varepsilon]$ differenzierbare Funktion $x\colon [0; \varepsilon] \to \mathbb{R}^n$ mit $x(0) = \bar{x}$ und $\dot{x}(0) = s$, so daß für alle $t \in [0; \varepsilon]$ gilt:

$$x(t) \in \operatorname{aff}(C), \qquad F(x(t)) \in \operatorname{aff}(\mathcal{K})$$

2) $\exists \varepsilon_1 \in (0; \varepsilon]$ $\forall t \in [0; \varepsilon_1]$: $x(t) \in C$, $F(x(t)) \leq_{\mathcal{K}} 0$, (d. h. $x(t)$ ist zulässige Lösung von (P)). In Worten gefaßt ist $x(t)$ eine Kurve mit $x(0) = \bar{x}$, die für kleine $t \geq 0$ im zulässigen Bereich verläuft und (lokal) in Richtung s von $\bar{x}$ wegläuft.

Beweis: Sei

$$A(x) := \begin{pmatrix} H\, DF(x) \\ G \end{pmatrix}$$

wie in (7.13), d. h. $\operatorname{aff}(C) = \{x \in \mathbb{R}^n \mid Gx = g\}$, $\operatorname{aff}(\mathcal{K}) = N(H)$. Regularität $\Longrightarrow A(\bar{x})$ hat vollen Zeilenrang (Satz (7.13)). Stetigkeit $\Longrightarrow A(x)$ hat vollen Zeilenrang für $\|x - \bar{x}\| \leq \tilde{\varepsilon}$ mit einem festen $\tilde{\varepsilon} > 0$. Setze $P(x) := A(x)^T (A(x)A(x)^T)^{-1} A(x)$ für $\|x - \bar{x}\| \leq \tilde{\varepsilon}$. (Falls $z \in N(A(x))$ im Nullraum von $A(x)$ liegt, so ist offenbar $P(x)z = 0$, und falls $z \in R(A^T(x))$ im Bild von $A^T(x)$ liegt, so folgt $P(x)z = z$ (Projektionseigenschaft).)
Es ist

$$A(\bar{x})s = \begin{pmatrix} H\, DF(\bar{x})s \\ Gs \end{pmatrix} = 0,$$

denn wegen $x^1 \in C^i \subseteq \operatorname{aff}(C)$ ist

$$Gs = G(x^1 - \bar{x}) = g - g = 0$$

und

$$DF(\bar{x})s \in -F(\bar{x}) - \mathcal{K}^i \subset -\mathcal{K} - \mathcal{K} \subseteq \operatorname{aff}(\mathcal{K}),$$

d. h. $H\, DF(\bar{x})s = 0$.

Betrachte folgendes AWP: $\dot{x} = (I - P(x))s$; $x(0) = \bar{x}$. Die Abbildung $x \mapsto (I - P(x))s$ ist eine stetige Vektorfunktion für $\|x - \bar{x}\| < \tilde{\varepsilon} \Longrightarrow$ (Satz von Peano): Es gibt ein $\varepsilon > 0$ und

eine Funktion $x(t)$ mit $x(0) = \bar{x}$ und $\dot{x}(t) = (I - P(x(t)))s$ für $t \in [0;\varepsilon]$. Da $A(\bar{x})s = 0$ ist, folgt $\dot{x}(0) = (I - P(\bar{x}))s = s$. Es ist

$$\begin{pmatrix} HF(x(t)) \\ Gx(t) \end{pmatrix} = \begin{pmatrix} 0 \\ g \end{pmatrix} \Longleftrightarrow \begin{cases} F(x(t)) & \in & \text{aff}(\mathcal{K}) \\ x(t) & \in & \text{aff}(C). \end{cases}$$

Die linke Seite gilt für $t = 0$. Wegen

$$\frac{d}{dt}\begin{pmatrix} HF(x(t)) \\ Gx(t) \end{pmatrix} = \begin{pmatrix} H\,DF(x(t)) \\ G \end{pmatrix} \dot{x}(t) = A(x(t))\Big(I - P(x(t))\Big) \overbrace{=}^{\text{Projektionseigenschaft}} 0$$

ist die linke Seite für $t \in [0;\varepsilon]$ ebenfalls erfüllt, d. h. es gilt (1).

Da C konvex ist, $x(0) = \bar{x} \in C$, $x(0) + \dot{x}(0) = \bar{x} + s = x^1 \in C^i$, ist $x(t) \in C$ für kleine $t > 0$ (Übungsaufgabe — wegen Teil (1) genügt es, den Fall $x^1 \in C^\circ$ zu untersuchen). Ebenso ist $F(\bar{x}) = F(x(0)) \in -\mathcal{K}$,

$$F(x(0)) + \frac{d}{dt}F(x(t))\Big|_{t=0} = F(\bar{x}) + DF(\bar{x}) \cdot s \in -\mathcal{K}^i$$

d. h. die Tangente an die Kurve $F(x(t)) \in \mathbb{R}^m$ zeigt in $t = 0$ in $-\mathcal{K}^i$, und wie oben folgt aus der Konvexität von $\mathcal{K}$, daß $F(x(t)) \in -\mathcal{K}$ für kleine $t \geq 0$. Dies zeigt (2).

Beweis von Satz (7.11): Sei

$$s_0 \in L(\mathcal{S}, \bar{x}) = \{s \in \mathbb{R}^n \mid \exists \lambda \geq 0 \quad \exists x \in C\colon\ s = \lambda(x - \bar{x}), \quad F(\bar{x}) + DF(\bar{x})(x - \bar{x}) \leq_{\mathcal{K}} 0\}.$$

Wir zeigen $s_0 \in T(S, \bar{x})$ und nehmen o.B.d.A. an, daß $s_0 = x_0 - \bar{x}$ mit $x_0 \in C$ und $F(\bar{x}) + DF(\bar{x})(x_0 - \bar{x}) \leq_{\mathcal{K}} 0$. (Sonst zeige, daß $\frac{1}{\lambda} \cdot s_0 \in T(\mathcal{S}, \bar{x})$.)
Regularität $\Longrightarrow$ Es gibt $x_1 \in C^i$ mit $F(\bar{x}) + DF(\bar{x})(x_1 - \bar{x}) \in -\mathcal{K}^i$. Setze $s_1 := x_1 - \bar{x}$ und $w_k := \frac{k-1}{k}s_0 + \frac{1}{k}s_1 = \frac{k-1}{k}x_0 + \frac{1}{k}x_1 - \bar{x}$. Wegen $x_0 \in C$, $x_1 \in C^i$ gilt $\frac{k-1}{k}x_0 + \frac{1}{k}x_1 \in C^i$ für $k \in \mathbb{N}$ nach Lemma (5.8). Außerdem ist $F(\bar{x}) + DF(\bar{x})w_k \in -\mathcal{K}^i$. Nach Hilfssatz (7.14) gibt es für jedes k ein $\varepsilon_k > 0$ und eine Kurve $x_k := [0;\varepsilon_k] \longrightarrow \mathbb{R}^n$, so daß $x_k(t)$ zulässig für (P) ist für $t \in [0;\varepsilon_k]$ und $x_k(0) = \bar{x}$ und $\dot{x}_k(0) = w_k$. Also ist für jedes feste $k \in \mathbb{N}$ der Punkt $w_k = \lim_{t\to 0,\ t>0} \frac{x_k(t)-\bar{x}}{t} \in T(\mathcal{S}, \bar{x})$ nach Definition des Tangentialkegels (bilde Folge $(t_j)_{j\in\mathbb{N}}$ mit $\lim_{j\to\infty} t_j = 0$). Da $T(\mathcal{S}, \bar{x})$ abgeschlossen ist und $\lim_{k\to\infty} w_k = s_0$, folgt $s_0 \in T(\mathcal{S}, \bar{x})$. Dies zeigt $L(\mathcal{S}, \bar{x}) \subseteq T(\mathcal{S}, \bar{x})$.

Bemerkung: Im Fall (P') ist

$$C = \mathbb{R}^n, \quad \mathcal{K} = \{u \in \mathbb{R}^m \mid u_i \geq 0, \quad 1 \leq i \leq p, \quad u_j = 0, \quad p+1 \leq j \leq m\}$$

Lemma (7.15): Falls für (P') die Voraussetzung (V) gilt, so ist $T(\mathcal{S}, \bar{x}) \subseteq L(\mathcal{S}, \bar{x})$.

Beweis: Übung.

Korollar (7.16): Falls für (P') die Voraussetzung (V) und (7.10) gilt, so ist $T(\mathcal{S}, \bar{x}) = L(\mathcal{S}, \bar{x})$.

8 Optimalitätsbedingungen 2. Ordnung

Betrachte (P'):

$$\min\{f(x) \mid f_i(x) \leq 0, \;\; i = 1, \ldots, p, \quad f_j(x) = 0, \;\; j = p+1, \ldots, m\}.$$

Zusätzlich zu den Optimalitätsbedingungen aus Kapitel 7 suchen wir notwendige und hinreichende Bedingungen für ein Minimum, falls alle Funktionen zweimal stetig differenzierbar sind.

Beispiel:

Falls $p = m = 0$ ist, so ist für ein lokales Minimum von f in $\bar{x}$ notwendig: $Df(x) = 0$ und $D^2 f(x) \geq 0$ (positiv semidefinit).

Hinreichend ist: $Df(x) = 0$ und $D^2 f(x) > 0$ (positiv definit).

Für $f\colon \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) = x^n$, $n \geq 2$ ist z.B.:

$$Df(0) = 0, \quad \text{und} \quad D^2 f(0) = \begin{cases} 2 & \text{für } n = 2 \\ 0 & \text{für } n \geq 3 \end{cases}$$

Dabei hat f für gerades n ein lokales Minimum bei $x = 0$ und für ungerades n nicht.Die Lücke zwischen notwendiger und hinreichender Bedingung zweiter Ordnung (d.h. einer Bedingung, die nur auf den ersten beiden Ableitungen beruht) läßt sich also selbst in diesem einfachen Fall nicht ganz schließen.

Ziel dieses Kapitels ist die Verallgemeinerung dieser Bedingungen auf Probleme der Form (P'):

Hilfssatz (8.1):

Sei $\varphi \in C^2(\mathbb{R}^n)$, $\varphi\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, $S \subseteq \mathbb{R}^n$ beliebig, $\bar{x} \in S$ und $\nabla\varphi(\bar{x}) = 0$. Dann gilt:

1) Falls $\bar{x}$ ein lokales Minimum von φ auf S ist, so ist $s^T D^2\varphi(\bar{x}) s \geq 0$ für alle $s \in (T(S, \bar{x}))$.

2) Falls $s^T D^2\varphi(\bar{x}) s > 0$ für alle $s \in T(S, \bar{x})$, dann ist $\bar{x}$ ein striktes lokales Minimum von φ auf S, und es gibt ein $\varepsilon > 0$ und ein $\delta > 0$, so daß

$$\varphi(x) \geq \varphi(\bar{x}) + \epsilon \cdot \|x - \bar{x}\|_2^2$$

für alle $x \in S$ mit $\|x - \bar{x}\|_2 \leq \delta$.

Beweis:

1) Sei $s \in T(S, \bar{x})$, d. h. $s = \lim_{k\to\infty}(x_k - \bar{x})$ mit $\lambda_k \geq 0$, $x_k \in S$ und $\lim_{k\to\infty} x_k = \bar{x}$. Es folgt mit der Taylorentwicklung aus der Voraussetzung $D\varphi(\bar{x}) = 0$ für große k:

$$0 \leq \varphi(x_k) - \varphi(\bar{x}) = \tfrac{1}{2}(x_k - \bar{x})^T D^2\varphi(\bar{x})(x_k - \bar{x}) + o(\|x_k - \bar{x}\|^2),$$

also auch

$$0 \leq \tfrac{1}{2}\lambda_k(x_k - \bar{x})^T D^2\varphi(\bar{x})\lambda_k(x_k - \bar{x}) + \lambda_k^2 \cdot o(\|x_k - \bar{x}\|^2).$$

Wegen

$$\lambda_k^2 \cdot o(\|x_k - \bar{x}\|^2) = \underbrace{\|\lambda_k(x_k - \bar{x})\|^2}_{\to s} \cdot \frac{o(\|x_k - \bar{x}\|^2)}{\|x_k - \bar{x}\|^2} \overset{k\to\infty}{\longrightarrow} 0$$

folgt für $k \to \infty$:

$$0 \leq \tfrac{1}{2} s^T D^2\varphi(\bar{x}) s.$$

2) Annahme: Die Behauptung sei falsch. Zu gegebenem $\varepsilon > 0$ gibt es also kein $\delta > 0$ mit

$$\varphi(x_k) \geq \varphi(\bar{x}) + \varepsilon \|x_k - \bar{x}\|^2,$$

sofern nur $\|x_k - \bar{x}\| \leq \delta$. Anders ausgedrückt gibt es also eine Folge, z. B. $\varepsilon_k = \frac{1}{k}$, und zu jedem ε_k stets ein $x_k \in S$ mit $\|x_k - \bar{x}\|_2 \leq \frac{1}{k} =: \delta_k$ und $\varphi(x_k) < \varphi(\bar{x}) + \frac{1}{k}\|x_k - \bar{x}\|_2^2$. Insbesondere ist $x_k \neq \bar{x}$ für alle k (sonst wäre $\varphi(x_k) = \varphi(\bar{x}) + \frac{1}{k}\|x_k - \bar{x}\|_2^2$). Die Folge

$$\left(\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|_2}\right)_k$$

ist beschränkt und hat somit einen Häufungspunkt s, d. h. es gibt eine Teilfolge $(k_i)_i$ mit

$$\lim_{i\to\infty} \frac{x_{k_i} - \bar{x}}{\|x_{k_i} - \bar{x}\|_2} = s.$$

Wähle

$$\lambda_i := \frac{1}{\|x_{k_i} - \bar{x}\|_2} > 0.$$

Dann ist $\lim_{i\to\infty} \lambda_i(x_{k_i} - \bar{x}) = s$ und $x_{k_i} \in S$. Also ist $s \in T(S, \bar{x})$. Es folgt

$$\varphi(\bar{x}) + \frac{1}{k_i}\|x_{k_i} - \bar{x}\|_2^2 > \varphi(x_{k_i})$$

$$= \varphi(\bar{x}) + D\varphi(\bar{x})(x_{k_i} - \bar{x}) + \tfrac{1}{2}(x_{k_i} - \bar{x})^T D^2\varphi(\bar{x})(x_{k_i} - \bar{x}) + o(\|x_{k_i} - \bar{x}\|_2^2).$$

Subtrahiert man $\varphi(\bar{x})$ von beiden Seiten, so erhält man nach Multiplikaktion mit $\lambda_{k_i}^2$

$$\frac{1}{k_i}(\lambda_{k_i}\|x_{k_i} - \bar{x}\|_2)^2 > \tfrac{1}{2}\lambda_{k_i}(x_{k_i} - \bar{x})^T D^2\varphi(\bar{x})\lambda_{k_i}(x_{k_i} - \bar{x}) + \lambda_{k_i}^2 o(\|x_{k_i} - \bar{x}\|_2^2),$$

und für $i \to \infty$ wie oben

$$0 \geq \tfrac{1}{2} s^T D^2\varphi(\bar{x}) s,$$

im Widerspruch zur Voraussetzung.

Bemerkung:
Falls $\bar{x} \in S^\circ$, so ist $T(S, \bar{x}) = \mathbb{R}^n$ und der Hilfssatz liefert die Aussage aus dem voranstehenden Beispiel.

Satz 8.2) Notwendige Bedingungen 2. Ordnung
Für das Problem (P') gelte:

1) $f \in C^2(\mathbb{R}^n)$ und $F = (f_1, \ldots, f_m)^T \in C^2(\mathbb{R}^n)$. Weiter sei

$$L(x, y) := f(x) + y^T F(x) = f(x) + \sum_{i=1}^{m} y_i f_i(x)$$

2) $\bar{x} \in S$ sei ein lokales Minimum für (P'), wobei

$$S := \{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \text{ für } 1 \leq p, \qquad f_j(x) = 0 \text{ für } p+1 \leq j \leq m\}.$$

3) S sei in $\bar{x}$ regulär, d. h. nach (7.13) ist

a) $Df_j(\bar{x})$, $j = p+1, \ldots, m$, sind linear unabhängig

b) $\exists\, s \in \mathbb{R}^n$:

$$\begin{array}{rcl} Df_i(\bar{x})s & < & 0 \text{ für } i \in I(\bar{x}) := \{i \leq p \mid f_i(\bar{x}) = 0\} \\ Df_j(\bar{x})s & = & 0 \text{ für } p+1 \leq j \leq m \end{array}$$

Dann gibt es ein $\bar{y} \in \mathbb{R}^m$ mit:

$\alpha)$ $\bar{y}_i \geq 0$ für $1 \leq i \leq p$,

$\beta)$ $\bar{y}_i f_i(\bar{x}) = 0$ für $1 \leq i \leq p$,

$\gamma)$ $\nabla_x L(x, \bar{y}) \mid_{x=\bar{x}} = \nabla f(\bar{x}) + (DF(\bar{x}))^T \bar{y} = 0$,

$\delta)$ $s^T (D_x^2 L(x, \bar{y}) \mid_{x=\bar{x}}) s \geq 0$ für alle $s \in T(S_1, \bar{x})$, wobei

$$S_1 := \{x \in S \mid \forall_{i \in \tilde{I}(\bar{x})} : f_i(x) = 0\} \text{ mit } \tilde{I}(\bar{x}) := \{i \in I(\bar{x}) \mid \bar{y}_i > 0\}.$$

(Hierbei können wir $\tilde{I}$ als die Menge der „wichtigen“ aktiven Indizes interpretieren).

Beweis: $\alpha)$–$\gamma)$ gelten wegen (7.12). $\delta)$: Sei $x \in S_1$. Dann ist

$$L(x, \bar{y}) = f(x) + \sum_{i=1}^{p} \bar{y}_i f_i(x) + \sum_{j=p+1}^{m} \bar{y}_j f_j(x) = f(x).$$

Denn für $i \leq p$ und $x \in S_1$ ist entweder $\bar{y}_i = 0$ oder $f_i(x) = 0$. Für $j \geq p+1$ ist $f_j(x) = 0$. Wegen $S_1 \subseteq S$ ist $\bar{x}$ somit lokales Minimum von $\varphi(x) := L(x, \bar{y})$ auf S_1 und

$$D\varphi(\bar{x}) = D_x L(x, \bar{y}) \mid_{x=\bar{x}} = 0.$$

Wegen Hilfssatz (8.1) ist (mit S_1 statt S)

$$0 \leq s^T D^2 \varphi(\bar{x}) s = s^T (D_x^2 L(x, \bar{y}) \mid_{x=\bar{x}}) s \text{ für alle } s \in T(S_1, \bar{x}). \qquad \#$$

Falls auch S_1 in $\bar{x}$ regulär ist, ist wegen Korollar (7.16) $T(S_1, \bar{x}) = L(S_1, \bar{x})$, wobei

$$\begin{array}{rll} L(S_1, \bar{x}) := \{s \in \mathbb{R}^n \mid Df_i(\bar{x})s = 0 & \text{für} & i \in \tilde{I}(\bar{x}) \cup \{p+1, \ldots, m\}, \\ Df_i(\bar{x})s \leq 0 & \text{für} & i \in I(\bar{x}) \setminus \tilde{I}(\bar{x})\}. \end{array}$$

Regularität von S_1 in $\bar{x}$ ist wegen (7.13) äquivalent zu

(8.3) 1) $Df_i(\bar{x})$, $i \in \tilde{I}(\bar{x}) \cup \{p+1, \ldots, m\}$, sind linear unabhängig,

2) $\exists s \in \mathbb{R}^n$: $Df_i(\bar{x})s = 0$ für $i \in \tilde{I}(\bar{x}) \cup \{p+1, \ldots, m\}$, $Df_i(\bar{x})s < 0$ für $i \in I(\bar{x}) \setminus \tilde{I}(\bar{x})$.

Die Bedingung (8.3) heißt *„constraint qualification“ 2.Ordnung.*

Korollar (8.4):

Es mögen die Voraussetzungen von Satz (8.2) gelten und zusätzlich noch (8.3). Dann kann in Satz (8.2) die Aussage $\delta)$ durch

δ'): $s^T D_x^2 L(\bar{x}, \bar{y}) s \geq 0$ für alle $s \in L(S_1, \bar{x})$,
d. h. für alle s mit $Df_i(\bar{x})s = 0$ für $i \in \tilde{I}(\bar{x}) \cup \{p+1, \ldots, m\}$ und $Df_i(\bar{x})s \leq 0$ für $i \in I(\bar{x}) \setminus \tilde{I}(\bar{x})$

ersetzt werden.

Satz (8.5): Hinreichende Bedingung 2.Ordnung

Es gelte die Voraussetzung (1) von Satz (8.2), und zusätzlich gebe es ein $\bar{y}$ mit α), β), γ) aus (8.2) sowie

δ''): $s^T D_x^2 L(\bar{x}, \bar{y}) s > 0$ für alle $s \in L(S_1, \bar{x})$ mit $s \neq 0$.

Dann ist $\bar{x}$ ein striktes lokales Minimum von (P'), und es gibt $\varepsilon > 0$, $\delta > 0$, so daß

$$f(x) \geq f(\bar{x}) + \varepsilon \cdot \|x - \bar{x}\|_2^2 \text{ für alle } x \in N_\delta(\bar{x}) \text{ mit } N_\delta(\bar{x}) := \{x \in S \mid \|x - \bar{x}\|_2 \leq \delta\}.$$

Beweis:

Aus Lemma (7.15) folgt $T(S_1, \bar{x}) \subseteq L(S_1, \bar{x})$.
Annahme: Die Behauptung ist falsch, d. h. wie im Beweis von Hilfssatz (8.1) gibt es zu jedem $k \in \mathbb{N}$ ein $x_k \in S$ mit $\|x_k - \bar{x}\| \leq \frac{1}{k}$ und

$$f(x_k) < f(\bar{x}) + \tfrac{1}{k}\|x_k - \bar{x}\|_2^2. \qquad \text{(i)}$$

Insbesondere folgt $x_k \neq \bar{x}$ für alle k. Die Folge

$$\left(\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|_2}\right)_k$$

besitzt einen Häufungspunkt s. Wir nehmen O.B.d.A. an, daß

$$\lim_{k \to \infty} \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|_2} = s \in T(S, \bar{x}).$$

Setze

$$g(x) := \sum_{i \in \tilde{I}(\bar{x})} \bar{y}_i f_i(x),$$

dann gilt für $x \in S$

$$L(x, \bar{y}) = f(x) + \bar{y}^T F(x) = f(x) + g(x).$$

Somit gilt wegen (i)

$$\begin{aligned} \tfrac{1}{k}\|x_k - \bar{x}\|_2^2 &\geq f(x_k) - f(\bar{x}) = \\ &= (L(x_k, \bar{y}) - L(\bar{x}, \bar{y})) - (g(x_k) - g(\bar{x})) \\ &\geq L(x_k, \bar{y}) - L(\bar{x}, \bar{y}), \end{aligned}$$

da $g(\bar{x}) = 0$ und $g(x_k) \leq 0$. Also folgt

$$\frac{1}{k} > \frac{L(x_k, \bar{x}) - L(\bar{x}, \bar{y})}{\|x_k - \bar{x}\|_2^2}.$$

Wegen

$$L(x_k,\bar{y}) = L(\bar{x},\bar{y}) + D_xL(\bar{x},\bar{y})(x_k-\bar{x}) + \tfrac{1}{2}(x_k-\bar{x})^T D_x^2 L(\bar{x},\bar{y})(x_k-\bar{x}) + o(\|x_k-\bar{x}\|_2^2)$$

und $D_xL(\bar{x},\bar{y}) = 0$ ist somit

$$\frac{1}{k} > \frac{1}{\|x_k-\bar{x}\|_2^2}\Big(\tfrac{1}{2}(x_k-\bar{x})^T D_x^2 L(\bar{x},\bar{x})(x_k-\bar{x}) + o(\|x_k-\bar{x}\|_2^2)\Big)$$

und für $k \to \infty$ folgt

$$0 \geq s^T D_x^2 L(\bar{x},\bar{y}) s \tag{79}$$

für ein $s \in T(S,\bar{x}) \subseteq L(S,\bar{x})$. Aus (79):

$$\frac{1}{k} \geq \underbrace{\frac{L(x_k,\bar{x}) - L(\bar{x},\bar{y})}{\|x_k-\bar{x}\|_2^2}}_{\longrightarrow s^T D_x^2 L(\bar{x},\bar{y}) s} - \frac{g(x_k)-g(\bar{x})}{\|x_k-\bar{x}\|_2^2}$$

folgt wegen $g(x_k) \leq g(\bar{x}) \leq 0$

$$\frac{|g(x_k)-g(\bar{x})|}{\|x_k-\bar{x}\|_2^2} \leq M$$

für alle k, insbesondere

$$\lim_{k\to\infty} \frac{|g(x_k)-g(\bar{x})|}{\|x_k-\bar{x}\|_2} = 0.$$

Es ist

$$\begin{aligned} g(x_k)-g(\bar{x}) &= \sum_{i\in\tilde{I}(\bar{x})} \bar{y}_i(f_i(x_k)-f_i(\bar{x})) = \\ &= \sum_{i\in\tilde{I}(\bar{x})} \bar{y}_i\Big(Df_i(\bar{x})(x_k-\bar{x}) + o(\|x_k-\bar{x}\|)\Big) \end{aligned}$$

Für $k \to \infty$ folgt:

$$0 = \sum_{i\in\tilde{I}(\bar{x})} \bar{y}_i \cdot Df_i(\bar{x}) s.$$

Mit $\bar{y}_i > 0$ und $Df_i(\bar{x})s \leq 0$ für $i \in \tilde{I}(\bar{x})$ folgt also $Df_i(\bar{x}) = 0$ für alle $i \in \tilde{I}(\bar{x})$, d. h. $s \in L(S_1,\bar{x})$. Damit steht (79) im Widerspruch zu δ''). #

Setze nun

$$\Phi(x,y) := \begin{pmatrix} \nabla L(x,y) \\ y_1 f_1(x) \\ \vdots \\ y_p f_p(x) \\ \vdots \\ f_m(x) \end{pmatrix} \quad \text{für } x \in \mathbb{R}^n,\ y \in \mathbb{R}^m,$$

d. h. Φ: $\mathbb{R}^{n+m} \longrightarrow \mathbb{R}^{n+m}$. Das System $\Phi(x,y) = 0$ besteht dann aus $n+m$ Gleichungen in $n+m$ Unbekannten.

Eine Lösung von $\Phi(x,y) = 0$ mit $y_1 \geq 0, \ldots, y_p \geq 0$ und $f_1(x) \leq 0, \ldots, f_p(x) \leq 0$ heißt *Kuhn-Tucker-Punkt zu* (P') (weil sie die notwendige Bedingung 1.Ordnung für ein Minimum erfüllt). Mit

$$L(x,y) = f(x) + \sum_{i=1}^{m} y_i f_i(x)$$

und

$$\nabla_x L(x,y) = \nabla f(x) + \sum_{i=1}^{m} y_i \nabla f_i(x) = \nabla f(x) + (DF(x))^T y$$

ergibt sich

$$D\Phi(x,y) = \begin{pmatrix} D_x^2 L(x,y) & & (DF_1(x))^T & & (DF_2(x))^T \\ y_1 Df_1(x) & f_1(x) & & & \\ \vdots & & \ddots & & 0 \\ y_p Df_p(x) & & & f_p(x) & \\ Df_{p+1}(x) & & & & \\ \vdots & & 0 & & 0 \\ Df_m(x) & & & & \end{pmatrix}$$

$$= \begin{pmatrix} H(x,y) & (DF_1(x))^T & (DF_2(x))^T \\ Y_1\, DF_1(x) & \mathrm{Diag}(F_1(x)) & 0 \\ DF_2(x) & 0 & 0 \end{pmatrix} =: J(x,y),$$

mit

$$H(x,y) = D^2 f(x) + \sum_{l=1}^{m} y_l D^2 f_l(x) \quad \text{und} \quad Y_1 = \begin{pmatrix} y_1 & & \\ & \ddots & \\ & & y_p \end{pmatrix}.$$

Satz (8.6):

Es mögen folgende Voraussetzungen gelten:

i) $f,\ f_i \in C^2(\mathbb{R})^n, \quad 1 \leq i \leq m.$

ii) $(\bar{x}, \bar{y})$ ist Lösung von $\Phi(x,y) = 0$ mit $\bar{y}_i \geq 0$ und $f_i(\bar{x}) \leq 0$ für $1 \leq i \leq p$ (d. h. $(\bar{x}, \bar{y})$ ist ein Kuhn-Tucker-Punkt zu (P').)

Dann gilt:

1) $J(\bar{x}, \bar{y})$ ist nicht singulär, falls

   a) $\bar{y}_i > 0$ für $i \in I(\bar{x})$ (d. h. $I(\bar{x}) = \tilde{I}(\bar{x})$) (strikte Komplementarität)

   b) die constraint qualification 2.Ordnung erfüllt ist, d. h. $Df_i(\bar{x})$ sind linear unabhängig für $i \in I(x) \cup \{p+1, \ldots, m\}$.

   c) die hinreichenden Bedingungen 2.Ordnung für ein lokales Minimum erfüllt sind, d. h. $s^T D_x^2 L(\bar{x}, \bar{y}) s > 0$ für alle $s \neq 0$ mit $Df_i(\bar{x}) s = 0$ für alle $i \in I(\bar{x}) \cup \{p+1, \ldots, m\}$.

2) Falls $J(\bar{x},\bar{y})$ nichtsingulär ist, gelten 1a), 1b), und falls $s^T D_x^2 L(\bar{x},\bar{y})s \geq 0$ für alle $s \in L(S_1,\bar{x})$ (notwendige Bedingung 2.Ordnung), dann gilt auch 1c).

Beweis:
Es sei $I(\bar{x}) = \{1,\ldots,p_1\}$, d. h. die nichtaktiven Ungleichungen entsprechen den Indizes $p_1+1,\ldots,p$, sowie

$$F_{11}(x) := \begin{pmatrix} f_1(x) \\ \vdots \\ f_{p_1}(x) \end{pmatrix}, \quad F_{12}(x) := \begin{pmatrix} f_{p_1+1}(x) \\ \vdots \\ f_p(x) \end{pmatrix}, \quad Y_{11} = \begin{pmatrix} y_1 & & \\ & \ddots & \\ & & y_{p_1} \end{pmatrix}, \quad \text{usw.}$$

Dann sind $F_{11}(\bar{x}) = 0$ und $\bar{Y}_{12} = 0$. Wir erhalten

$$D\Phi(\bar{x},\bar{y}) = \begin{pmatrix} H(\bar{x},\bar{y}) & (DF_{11}(\bar{x}))^T & (DF_{12}(\bar{x}))^T & (DF_2(\bar{x}))^T \\ \bar{Y}_{11}\,DF_{11}(\bar{x}) & 0 & 0 & 0 \\ 0 & 0 & \mathrm{Diag}(F_{12}(\bar{x})) & 0 \\ DF_2(\bar{x}) & 0 & 0 & 0 \end{pmatrix}.$$

$D\Phi$ ist regulär, wenn das Gleichungssystem

$$\begin{pmatrix} H(\bar{x},\bar{y}) & (DF_{11}(\bar{x}))^T & (DF_{12}(\bar{x}))^T & (DF_2(\bar{x}))^T \\ \bar{Y}_{11}\,DF_{11}(\bar{x}) & 0 & 0 & 0 \\ 0 & 0 & \mathrm{Diag}(F_{12}(\bar{x})) & 0 \\ DF_2(\bar{x}) & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

nur die Lösung 0 hat. Dies ist wegen $F_{12}(\bar{x}) < 0$ äquivalent zu $w = 0$ und

$$\begin{pmatrix} H(\bar{x},\bar{y}) & (DF_{11}(\bar{x}))^T & (DF_2(\bar{x}))^T \\ \bar{Y}_{11}\,DF_{11}(\bar{x}) & 0 & 0 \\ DF_2(\bar{x}) & 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \qquad (**)$$

Wir können daher o.B.d.A. $p_1 = p$ annehmen („w fällt weg“). Offenbar ist $J(\bar{x},\bar{y})$ singulär, falls ein $\bar{y}_i = 0$ für $i \in I(\bar{x}) = \{1,\ldots,p\}$ (Nullzeile in $(**)$!) $\Longrightarrow$ 1a) ist notwendig.
Falls $\bar{Y}_{11}$ nur positive Diagonalelemente hat, kann man die 2.Blockzeile von (**) mit Y_{11}^{-1} durchmultiplizieren, ohne die Regularität zu ändern. Wir erhalten:

$$\bar{J}(\bar{x},\bar{y}) := \begin{pmatrix} H(\bar{x},\bar{y}) & (DF(\bar{x}))^T \\ DF(\bar{x}) & 0 \end{pmatrix}.$$

Gäbe es ein $u \in \mathrm{Kern}((DF(x))^T)$, d. h. $(DF(\bar{x}))^T u = 0$, $u \neq 0$, so wäre

$$\bar{J}(\bar{x},\bar{y}) \begin{pmatrix} 0 \\ u \end{pmatrix} = 0,$$

d. h. $\bar{J}$ wäre singulär in $(\bar{x},\bar{y})$.
Also ist auch 1b) notwendig für die Regularität von $J(\bar{x},\bar{y})$ bzw. von $\bar{J}(\bar{x},\bar{y})$.
Löse $\bar{J}(\bar{x},\bar{y})\binom{v}{u} = 0$, d. h.

$$\begin{pmatrix} H(\bar{x},\bar{y}) & (DF(\bar{x}))^T \\ DF(\bar{x}) & 0 \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad (***)$$

nach (u, v) auf.

Zu zeigen: $u = 0$, $v = 0$ ist die einzige Lösung, falls 1a)–c) erfüllt sind.
Aus der 1. Zeile folgt: $H(\bar{x}, \bar{y})v \in R((DF(\bar{x}))^T)$.
Aus der 2. Zeile folgt: $v \in N(DF(\bar{x}))$.
Damit ist $v^T H(\bar{x}, \bar{y})v = 0$.

Wegen 1c) ist $\tilde{v}H(\bar{x}, \bar{y})\tilde{v} > 0$ für alle $v \in N(DF(\bar{x})) \setminus \{0\}$. Also ist $v = 0$ und somit auch $u = 0$, d. h. $\bar{J}(\bar{x}, \bar{y})$ ist regulär.

Falls umgekehrt $v^T H(\bar{x}, \bar{y})v \geq 0$ für alle $v \in N(DF(\bar{x}))$, so ist mit

$$P_N := I - (DF(\bar{x})^T (DF(\bar{x}) \cdot (DF(\bar{x}))^T)^{-1} (DF(\bar{x}))$$

die Matrix $M := P_N^T H(\bar{x}, \bar{y}) P_N$ positiv semidefinit. (Beachte, daß P_N wegen 1b) existiert und die Orthogonalprojektion auf $N(DF(\bar{x}))$ beschreibt, denn für $z \in N(DF(\bar{x}))$ ist $P_N z = z$ und für $z \in R((DF(\bar{x}))^T)$ ist $z = (DF(\bar{x}))^T w$ und somit $P_N z = z - z = 0$.)
Zu zeigen: $J(\bar{x}, \bar{y})$ regulär $\Longrightarrow v^T H(\bar{x}, \bar{y})v > 0$ für alle $v \in N(DF(\bar{x})) \setminus \{0\}$.
Annahme: $v^T H(\bar{x}, \bar{y})v = 0$ für ein $v \in N((DF(\bar{x}))$, $v \neq 0$.
Dann ist $v^T M v = 0$ und somit $Mv = 0$. Daraus folgt

$$P_N^T H(\bar{x}, \bar{y})v = P_N^T H(\bar{x}, \bar{y}) P_N v = 0$$

und

$$H(\bar{x}, \bar{y})v \in N(P_N^T) = R((DF(\bar{x}))^T),$$

d. h. es gibt ein u mit

$$H(\bar{x}, \bar{y})v = DF(\bar{x})^T u.$$

Dieser Vektor v liefert somit eine von 0 verschiedene Lösung von $(***)$

$$\bar{J}(\bar{x}, \bar{y}) \begin{pmatrix} v \\ -u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Somit gilt auch 1c). #

Sensitivität der Lösung: Für $t \in \mathbb{R}^q$ seien f, f_i: $\mathbb{R}^{n+q} \to \mathbb{R}$ in $C^2(\mathbb{R}^{n+q})$ $(1 \leq i \leq m)$ und (P_t'):

$$\min_{x \in \mathbb{R}^n} \{f(x,t) \mid f_i(x,t) \leq 0, \quad 1 \leq i \leq p, \quad f_j(x,t) = 0, \quad p+1 \leq j \leq m\}$$

und $x(t)$ eine zugehörige Lösung (falls existent). (z. B. $f_i(x,t) := f_i(x) - t_i$, $f(x,t) := f(x) + t_0 \cdot \tilde{c}^T x$).

Lemma 8.7: Für $t = 0$ möge $x(0) = \bar{x}$ die hinreichenden Bedingungen 2.Ordnung erfüllen. (D.h. es seien die Bedingungen erster Ordnung erfüllt: es gibt ein $\bar{y} \in \mathbb{R}^m$ mit $\bar{y}_i \geq 0$, $\bar{y}_i f_i(\bar{x}) = 0$ für $1 \leq i \leq p$ und $D_x L(\bar{x}, \bar{y}, 0) = 0$, wobei $L(x, y, t) := f(x,t) + \sum_{i=1}^m y_i \cdot f_i(x,t)$ die Lagrangefunktion zu dem Problem (P_t') ist, sowie die Bedingungen zweiter Ordnung: $s^T D_x L(\bar{x}, \bar{y}, 0)s > 0$ für alle $s \neq 0$ mit $Df_i(\bar{x}, 0)s = 0$ für $i \in I(\bar{x}) \cup \{p+1, \ldots, m\}$).
Weiter sei $\bar{y}_i - f_i(\bar{x}, 0) > 0$ (strikte Komplementarität) und

$$\{D_x f_i(\bar{x}, 0) \mid i \in I(\bar{x}) \cup \{p+1, \ldots, m\}\}$$

linear unabhängig. Dann gibt es $\delta > 0$ und ein $\varepsilon > 0$, so daß es für jedes $t \in \mathbb{R}^q$ mit $\|t\| \leq \delta$ genau ein $x(t)$ gibt mit folgenden Eigenschaften. $x(t)$ erfüllt die hinreichenden Bedingungen 2. Ordnung für (P'_t) sowie $\|x(t) - \bar{x}\| \leq \varepsilon$. $x(t)$ ist strikt komplementär und die Gradienten aus $I(\bar{x}) \cup \{p+1, \ldots, m\}$ sind linear unabhängig.

Beweis:
$x(0)$ erfüllt $\Phi(x, y, t) = 0$ für $y = \bar{y}$ und $t = 0$. Hierbei ist

$$\Phi(x, y, t) = \begin{pmatrix} D_x L(x, y, t) \\ y_1 f_1(x, t) \\ \vdots \\ y_p f_p(x, t) \\ f_{p+1}(x, t) \\ \vdots \\ f_m(x, t) \end{pmatrix}$$

O.B.d.A. sei $I(\bar{x}) = \{1, \ldots, p\}$ (vgl. Beweis von Satz 8.6)). Es ist

$$D_{x,y}\Phi(x, y, t) \mid_{t=0} = \begin{pmatrix} D_x^2 L(x, y, 0) & (D_x F(x, 0))^T \\ Y_1 D_x F_1(x, 0) & 0 \\ D_x F_2(x, 0) & 0 \end{pmatrix}$$

nach Satz (8.6) regulär in $(x, y) = (\bar{x}, \bar{y})$. Nach dem Satz über implizite Funktionen gibt es ein $\delta > 0$ und ein $\tilde{\varepsilon} > 0$, so daß $\Phi(x, y, t) = 0$ eine eindeutige Lösung $x(t)$, $y(t)$ hat mit

$$\left\| \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} - \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \right\| \leq \tilde{\varepsilon} \quad \text{für} \quad \|t\| \leq \delta.$$

Dabei hängen $x(t)$, $y(t)$ stetig von t ab.
Die strikte Komplementarität $y_i(t) - f_i(x(t), t) > 0$ bleibt aufgrund der Stetigkeit für kleine $\|t\|$ erhalten, ebenso die lineare Unabhängigkeit der $Df_i(x(t))$ und die Definitheit von M (im Beweis von Satz (8.6)) auf Kern $DF(x(t), t)$. (Die positiven Eigenwerte von M hängen stetig von t ab, die Null-Eigenwerte bleiben aufgrund der Projektionseigenschaft erhalten). Sämtliche Bedingungen 2. Ordnung sind somit für kleine $\|t\|$ in $x(t)$, $y(t)$ erfüllt.

Um die Empfindlichkeit von t bezüglich Änderungen in den Eingabedaten zu messen, betrachtet man

$$\varphi(t) := f(x(t), t).$$

Es ist

$$D_t \varphi(t) \mid_{t=0} = D_t L(x(t), y(t), t) \mid_{t=0} .$$

Beweis: Übung.

9 Primal-duale Innere-Punkte-Methoden für nichtlineare Probleme

Vorüberlegung: Sei $\bar{x}$ ein lokales Minimum von

$$(P') : \qquad \min\{f(x) \mid f_i(x) \leq 0 \text{ für } 1 \leq i \leq p, \quad f_j(x) = 0 \text{ für } p+1 \leq j \leq m\}$$

Wir suchen einfache notwendige bzw. hinreichende Bedingungen für ein lokales Minimum, die wir in einem Algorithmus zum Auffinden des Minimums ausnutzen könnten.

1. Die naheliegendste Charakterisierung ist folgende: Es gibt keine zulässige Abstiegsrichtung, d.h. für alle $s \in \mathbb{R}^n$ und für $t \geq 0$ genügend klein ist entweder

$$\bar{x} + ts \notin S := \{x \in \mathbb{R}^n \mid f_i(x) \leq 0,\ 1 \leq i \leq p, \quad f_j(x) = 0,\ p+1 \leq j \leq m\}$$

   oder $f(\bar{x} + ts) \geq f(\bar{x})$.
   Problem: Diese Bedingung ist in dieser Form recht schwer nachzuprüfen und Vektoren s, die diese Bedingung verletzen sind im allgemeinen keine effizienten Suchrichtungen. Außerdem ist diese Bedingung auch nicht immer hinreichend, wie das Beispiel $p(x, y) := x^4 + y^2 + 4x^2y$ zeigt. Diese Funktion p hat in $(x, y) = (0, 0)$ *kein* lokales Minimum aber jede (geradlinige) Richtung ist lokal eine Anstiegsrichtung für p. (Beweis durch nachrechnen!)

2. Falls $\bar{x}$ regulär ist, dann gibt es nach dem Satz von Kuhn und Tucker ein $\bar{y} \in \mathbb{R}^m$, $\bar{y} = (\bar{y}_{(1)}, \bar{y}_{(2)})^T$ (wobei $\bar{y}_{(1)} \in \mathbb{R}^p$, $\bar{y}_{(2)} \in \mathbb{R}^{m-p}$) mit $\bar{y}_{(1)} \geq 0$,

$$F_1(\bar{x}) \leq 0,\ F_2(\bar{x}) = 0 \quad \text{für} \quad F(x) := \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix} = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix},\ \bar{y}_1^T F_1(\bar{x}) = 0$$

   und $\nabla f(\bar{x}) + (\bar{y}^T\, DF(\bar{x}))^T = 0$.

(Um bei Vektoren die Partition $y_{(1)}$ von der Komponente y_1 unterscheiden zu können, schreiben wir die '1' in Klammern (), bei F werden die einzelnen Komponenten durch Kleinbuchstaben f_i bezeichnet und F_1 bezeichnet die $y_{(1)}$ entsprechende Partition.)

Innere-Punkte-Ansatz: Für einen kleinen festen Parameter $\mu > 0$ löse:

$$\left.\begin{array}{l} F_1(x) + s_{(1)} = 0, \quad s_{(1)} > 0 \\ F_2(x) = 0 \\ \nabla f(x) + (y^T\, DF(x))^T = 0 \\ Y_{(1)} s_{(1)} = \mu e, \quad y_{(1)} > 0 \end{array}\right\} (*)$$

wobei

$$Y_{(1)} = \begin{pmatrix} y_1 & & \\ & \ddots & \\ & & y_p \end{pmatrix}, \quad s_{(1)} = \begin{pmatrix} s_1 \\ \vdots \\ s_p \end{pmatrix}, \quad e = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^p.$$

Als Startpunkt können dabei beliebige $x \in \mathbb{R}^n$, $y_{(1)} > 0$, $s_{(1)} > 0$, $y_{(2)} \in \mathbb{R}^{m-p}$ gewählt werden. (Dann sind alle Ungleichungen erfüllt, möglicherweise auf Kosten eines Residuums in den Gleichungen).
Idee: Löse $(*)$ mit dem gedämpften Newtonverfahren unter Bewahrung der Ungleichungen. Genauer: Man linearisiere die Gleichungen in $(*)$ und wähle die Schrittweiten beim Newtonverfahren stets so, daß die Ungleichungen strikt gewahrt bleiben. Falls die Lösung von $(*)$ hinreichend gut approximiert ist, so reduziere μ (z.B. auf $\mu^+ = 0.1\mu$) und wiederhole das Verfahren.

Daß dieser Ansatz tatsächlich gegen ein lokales Minimum von (P') konvergiert und nicht z.B. irgendwo mit immer kleiner werdenden Schrittweiten "hängen bleibt" oder wegen einer singulären Jacobimatrix abbricht, ist bislang erst für spezielle Klassen von (konvexen) Programmen gezeigt worden und hängt wesentlich davon ab, daß stets $\mu > 0$ gewählt wird. Insbesondere ist der nachfolgende Algorithmus zum einen wegen fehlender Prediktor-Korrektor-Strategie für praktische Zwecke viel zu langsam, und zum anderen für nichtkonvexe Probleme in dieser Form außerdem nicht immer konvergent. Er dient nur als Motivation um Analogien zur linearen Programmierung aufzuzeigen und eine konzeptionelle "Schwäche" der klassischen Barrieremethode aufzuzeigen.

Ein primal-duales Verfahren Wir benutzen die Notation

$$H(x,y) := \mathcal{D}^2 f(x) + \sum_{l=1}^{m} y_l \mathcal{D}^2 f_l(x).$$

Konzeptioneller Algorithmus (p-d.V.):
x^0, y^0, s^0 seien gegeben mit $y^0_{(1)} > 0$ und $s^0_{(1)} > 0$, $(y_{(1)}, s_{(1)} \in I\!R^p)$.
Für $k = 0, 1, 2, \ldots$:

1. Wähle $\mu^k \geq 0$ (Ziel: $\lim_{k\to\infty} \mu^k = 0$).

2. Löse die Linearisierung von $(*)$:

$$\begin{aligned} F_1(x) + DF_1(x)\Delta x + s_{(1)} + \Delta s_{(1)} &\stackrel{!}{=} 0 \\ F_2(x) + DF_2(x)\Delta x &\stackrel{!}{=} 0 \\ \nabla f(x) + (DF(x))^T y + H(x,y)\Delta x + (DF(x))^T \Delta y &\stackrel{!}{=} 0 \\ Y_{(1)} s_{(1)} + Y_{(1)} \Delta s_{(1)} + S_{(1)} \Delta y_{(1)} &\stackrel{!}{=} \mu^k e \end{aligned}$$

   mit $(x,y,s) = (x^k, y^k, s^k)$ nach $(\Delta x, \Delta y, \Delta s)$ auf.

3. Bestimme eine Schrittweite $\alpha_k \in (0,1]$ mit $y^k_{(1)} + \alpha_k \Delta y_{(1)} > 0$ und $s^k_{(1)} + \alpha_k \Delta s_{(1)} > 0$.

4. Setze
$$\left(x^{k+1}, y^{k+1}, s^{k+1}\right) := (x^k, y^k, s^k) + \alpha_k (\Delta x, \Delta y, \Delta s)$$

Zusammenhang mit Barrieremethoden Sei $b : I\!R_+ \to I\!R$ eine streng monoton fallende, glatte, konvexe Barrierefunktion für $I\!R_+$, d.h. $\lim_{t\to 0} b'(t) = -\infty$.
Beispiele:

$$b(t) = -\log t, \quad b(t) = -\sqrt{t}, \quad b(t) = \frac{1}{t}, \quad b(t) = \frac{1}{t^\alpha} \quad (\alpha > 0)$$

Zur Lösung von (P') betrachte folgende Hilfsprobleme:

$$(B) \qquad \min_x \Big\{ f(x) + \sum_{i=1}^{p} w_i b(d_i - f_i(x)) \mid f_j(x) = 0, \qquad j \geq p+1, \Big\}$$

wobei $w_i > 0$ kleine Gewichte und $d_i \geq 0$ kleine " Shifts" seien, so daß $d_i - f_i(x) > 0$. In (B) nehmen wir implizit die Ungleichungen $f_i(x) < d_i$ mit an, da andernfalls der Barriereterm $b(d_i - f_i(x))$ nicht definiert wäre. Falls $d_i = 0$, dann garantiert der Summand $w_i b(0 - f_i(x))$ in der Zielfunktion von (B) für jedes noch so kleine $w_i > 0$, daß x bezüglich "$f_i(x) \leq 0$" strikt zulässig bleibt.

Definition:

$$\Phi(x;\, w, d) := f(x) + \sum_{i=1}^{p} w_i b(d_i - f_i(x))$$

Lemma 1: Falls f, f_i $(i = 1. \ldots, p)$ konvex sind, so ist auch Φ konvex.

Beweis: Falls g, h konvex sind, dann auch $\lambda g + \mu h$ für $\lambda, \mu \geq 0$. (Einsetzen in die Definition von Konvexität). Es genügt daher zu zeigen, daß $\varphi_i(x) := b(d_i - f_i(x))$ für jedes $i = 1, \ldots, p$ konvex ist. Für $\varrho \in [0; 1]$ gilt:

$$\begin{aligned}
d_i - f_i(\varrho x + (1-\varrho)y) &\geq d_i - (\varrho f_i(x) + (1-\varrho) f_i(y)) \\
&= \varrho(d_i - f_i(x)) + (1-\varrho)(d_i - f_i(y)). \\
\varrho\varphi_i(x) + (1-\varrho)\varphi_i(y) &= \varrho b(d_i - f_i(x)) + (1-\varrho) b(d_i - f_i(y)) \\
&\geq b\Big(\varrho(d_i - f_i(x)) + (1-\varrho)(d_i - f_i(y))\Big) \\
&\geq b\Big(d_i - f_i(\varrho x + (1-\varrho)y)\Big) \\
&= \varphi_i(\varrho x + (1-\varrho)y)
\end{aligned}$$

aufgrund der Konvexität und Monotonie von b. #

Lemma 2: Falls f, f_i $(i = 1, \ldots, p)$ konvex und f_j $(j = p+1, \ldots, m)$ affin sind, sowie $w_i = \mu$ und $d_i = 0$ für $i = 1, \ldots, p$ gewählt sind, so stimmen für die logarithmische Barrierefunktion die Minima von (B) und die Lösungen von $(*)$ überein.

Beweis: Φ ist nach Lemma 1 konvex. Weiter ist für (B) die Slaterbedingung erfüllt (da nur affine Gleichungsrestriktionen vorliegen). Also sind folglende Aussagen notwendig und hinreichend für ein Minimum von (B):

$$(**) \begin{cases} \nabla f(x) - \sum_{i=1}^{p} w_i b'(d_i - f_i(x)) \nabla f_i(x) + \sum_{j=p+1}^{m} y_j \nabla f_j(x) = 0 \\ f_j(x) = 0, \quad j = p+1, \ldots, m \end{cases}$$

Definiert man $y_i := -w_i b'(d_i - f_i(x)) > 0$ und $s_{(1)} := d - F_1(x)$, dann sind die ersten 3 Bedingungen von $(*)$ erfüllt. Weiter ist

$$y_i s_i = -\mu b'(d_i - f_i(x))(d_i - f_i(x)) = \mu,$$

falls $b'(t) = -\frac{1}{t}$, d.h. falls $b(t) = -\log t$. Umgekehrt: Falls eine Lösung von $(*)$ gegeben ist, so folgt $-f_i(x) = \frac{\mu}{y_i}$ für $i = 1, \ldots, p$ (denn $Y_1 F_1(x) = -Y_1 s_1 = -\mu e$) und

$$-\mu b'(-f_i(x)) = -\frac{\mu}{-\mu/y_i} = y_i, \qquad i = 1, \ldots, p,$$

so daß aus der Gleichung $\nabla f(x) + (DF(x))^T y = 0$ von $(*)$ die erste Gleichung von $(**)$ folgt. (Die zweite Gleichung von $(*)$ und die zweite Gleichung von $(**)$ stimmen überein.) #

Wählt man die Gewichte w_i anders, so lässt sich obiger Zusammenhang auch für andere Barrierefunktioenn herleiten.

Lemma 3: Es gelte zusätzlich $\lim_{t\to\infty} b'(t) = 0$. Falls f, f_i $(i = 1, \ldots, p)$ konvex sind und f_j $(j = p+1, \ldots, m)$ affin, und die Menge der Optimallösungen von (P') nicht leer und beschränkt ist, dann hat (B) für jedes $w > 0$ und jedes $d > 0$ $(d \geq 0$, falls die Slaterbedingung für (P') erfüllt ist) ein Minimum $x(w, d)$. Außerdem ist

$$\lim_{\lambda\to 0+} \Big(\inf\{\|\bar{x} - x(\lambda w,\, \lambda d)\| \;\Big|\; \bar{x} \text{ ist Optimallösung von } (P')\}\Big) = 0$$

Beweis: ohne Beweis.

Lemma 3 legt folgendes Verfahren zum Lösen von konvexen Programmen (P') mit f, $f_i \in C^2(\mathbb{R}^n)$ nahe:

Klassische Barrieremethode (k.Bm.): Gegeben $x^0 \in \mathbb{R}^n$. Wähle $w^0 > 0$ und $d^0 \geq 0$, so daß $d_i^0 > f_i(x^0)$ $(1 \leq i \leq p)$. Für $k = 1, 2, 3, \ldots$:

1. Wähle $\lambda_k \in (0, 1)$ so, daß mit $(w^k, d^k) := \lambda_k(w^{k-1}, d^{k-1})$ gilt: $f_i(x^{k-1}) < d_i^k$ $(1 \leq i \leq p)$.

2. Ausgehend von x^{k-1} führe einige Schritte des Newtonverfahrens (mit *line search*) zum Lösen von (B) mit $w = w^k$ und $d = d^k$ aus.

Motivation: Da $x(w, d)$ unter schwachen Voraussetzungen stetig (sogar glatt) von (w, d) abhängt, wird $x(w^{k-1}, d^{k-1})$ eine gute Näherung an $x(w^k, d^k)$ sein, wenn

$$\|(w^k, d^k) - (w^{k-1}, d^{k-1})\|$$

klein ist. Wegen $(w^k, d^k) \to 0$ $(k \to \infty)$ ist letzteres sicher für genügend große k der Fall.

Schwierigkeit: Der Einzugsbereich des Newtonverfahrens zur Minimierung von Φ, d.h. zur Lösung des Barriereproblems wird mit $w^k \to 0$ immer kleiner. Diese Schwierigkeit, in Verbindung mit der Tatsache, daß die Hessematrizen von $\Phi(x; w, d)$ für kleine $w > 0$, $d > 0$ und x in der Nähe von $\bar{x}$ im allgemeinen beliebig schlecht konditioniert sind, haben in der Vergangenheit dazu geführt, daß diese Methode als numerisch unbrauchbar eingestuft wurde. Ein Teil dieser Schwierigkeiten kann durch Verfeinerungen an der **(k.Bm.)** allerdings behoben werden. Trotzdem wird die **(k.Bm.)** auch heute nur selten eingesetzt.

Wir wollen die Suchrichtungen (Newtonrichtungen) der klassischen Barrieremethode und des primal-dualen Innere-Punkte-Verfahren vergleichen.

Vergleich der Newtonschritte (1) In (B) führen wir den Newtonschritt für

$$\begin{aligned} \nabla_x \Phi(x_i;\, w, d) + (DF_2(x))^T y_{(2)} &= 0 \\ F_2(x) &= 0 \end{aligned}$$

aus, wobei

$$D_x \Phi(x_i;\, w, d) = Df(x) - \sum_{i=1}^{p} w_i b'(d_i - f_i(x)) Df_i(x).$$

Wir definieren

$$y_i := y_i(x) := -w_i b'(d_i - f_i(x)) > 0,$$
$$H(x,y) := D^2 f(x) + \sum_{l=1}^{m} y_l D^2 f_l(x).$$

Die Linearisierung von

$$\begin{aligned} \nabla f(x+\Delta x) + (DF_1(x+\Delta x))^T y_{(1)}(x+\Delta x) + (DF_2(x+\Delta x))^T (y_{(2)} + \Delta y_{(2)}) &\stackrel{!}{=} 0 \\ F_2(x+\Delta x) &\stackrel{!}{=} 0 \end{aligned}$$

liefert das Gleichungssystem (GLS):

$$\begin{aligned} H(x,y)\Delta x + (DF_1(x))^T D_x y_{(1)}(x)\Delta x + (DF_2(x))^T \Delta y_{(2)} &= -\nabla f(x) - (DF(x))^T y \\ DF_2(x)\Delta x &= -F_2(x) \end{aligned}$$

wobei

$$D_x y_i(x) = D_x(-w_i \cdot b'(d_i - f_i(x))) = +w_i \cdot b''(d_i - f_i(x)) Df_i(x)$$

Wir erhalten somit

$$D_x y_{(1)}(x) = W_1 \cdot \mathrm{Diag}\left((b''(d_i - f_i(x)))_{i=1,\ldots,p} \right) \cdot DF_1(x)$$

mit $W_1 = \mathrm{Diag}(w_{(1)})$.
(Wir benutzen die allgemein übliche Notation wonach für $x \in I\!R^k$ mit $\mathrm{Diag}(x)$ die Diagonalmatrix $\begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_k \end{pmatrix}$, und für $M \in I\!R^{k\times k}$ mit $\mathrm{diag}(M)$ der Vektor $\begin{pmatrix} M_{11} \\ \vdots \\ M_{kk} \end{pmatrix}$ bezeichnet wird.)
Es sei $\tilde{r}_1 := -\nabla f(x) - (DF(x))^T y$ und $\tilde{r}_2 := -F_2(x)$. Wir betrachten nun die logarithmische Barrierefunktion $b(t) = -\log t$ mit $b'(t) = -\frac{1}{t}$, $b''(t) = \frac{1}{t^2}$ und erhalten das System

$$\begin{pmatrix} H(x,y) + A_1^T Y_1 S_1^{-1} A_1 & A_2^T \\ A_2 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y_{(2)} \end{pmatrix} = \begin{pmatrix} \tilde{r}_1 \\ \tilde{r}_2 \end{pmatrix}$$

mit

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} DF_1(x) \\ DF_2(x) \end{pmatrix},$$

$s_{(1)} := d - F_1(x)$, $S_1 := \mathrm{Diag}(s_{(1)})$ und

$$Y_1 = \mathrm{Diag}\left(\left(\frac{w_i}{d_i - f_i(x)} \right)_{i=1,\ldots,p} \right).$$

(2) Newtonschritt für $(*)$: Setzt man in Schritt 2) des **(p-d.V.)** $\Delta s_{(1)} = -F_1(x) - s_1 - A_1 \Delta x$, $\Delta y_{(1)} = S_1^{-1}(\mu e + Y_1(F_1(x) + A_1 \Delta x))$ so folgt (nach kurzer Rechnung)

$$\begin{pmatrix} H(x,y) + A_1^T Y_1 S_1^{-1} A_1 & A_2^T \\ A_2 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y_2 \end{pmatrix} = \begin{pmatrix} \tilde{r_1} - A_1^T S_1^{-1}(\mu e + Y_1 F_1(x)) \\ \tilde{r_2} \end{pmatrix}.$$

Der Einfachheit halber nehen wir nun an, daß ein strikt zulässiger Startpunkt x^0 mit $f_i(x^0) < 0$ für $1 \leq i \leq p$ bekannt sei und daher $d = 0$ gewählt werden kann. Sei nun $(x, y_{(2)})$ gegeben.

Dann sind für (B) die Größen $s_{(1)} = -F_1(x)$ und $y_{(1)} = S_1^{-1}\mu e$ definiert.
Falls μe in Schritt 2) vom **(p-d.V.)** durch $w_{(1)} = -Y_1F_1(x)$ ersetzt wird, so ist $w_{(1)} + Y_1F_1(x) = 0$, d.h. der Newtonschritt für $(*)$ stimmt mit dem für (B) überein. (Das Verfahren $(*)$ erzeugt die gleichen Suchrichtungen wie (B).) Die Änderung von μe zu $-Y_1F_1(x)$ ist dabei unwesentlich, sie wurde nur deshalb nicht in die Beschreibung des **(p-d.V.)** mit aufgenommen weil sie praktisch keine Vorteile hat.

Wo liegt nun der Unterschied zwischen den beiden Verfahren?

Wir betrachten wieder den Fall $d = 0$.
Es sei x das Minimum von (B) und $y_{(2)}$ der zugehörige Lagrange-Multiplikator. Setze

$$Y_1 = \mathrm{Diag}\left(\left(-\frac{w_i}{f_i(x)}\right)_{i=1,\ldots,p}\right)$$

und $s_{(1)} = -F_1(x)$, dann erfüllen (x, y, s) das System $(*)$ mit rechter Seite $(0, 0, 0, w_{(1)})$ anstelle von $(0, 0, 0, \mu e)$.
Annahme: Es sei $(*)$ in $(0, 0, 0, 0)$ *regulär*, d.h. die Linearisierung von $(*)$ in dem zugehörigen Lösungsvektor $\bar{x}, \bar{y}, \bar{s}$ habe nur $\bar{x}, \bar{y}, \bar{s}$ als Lösung (gleichbedeutend damit daß die Jacobimatrix von $(*)$ in $(x, y, s)(0)$ regulär ist).
Seien weiter f, f_i $(i = 1, \ldots, p)$, f_j $(j = p + 1, \ldots, m)$ 2-mal stetig differenzierbar. $\Longrightarrow$ Die Jacobimatrix von $(*)$ ist eine stetige Funktion von (x, y, s) und somit in einer kleinen Umgebung von $(\bar{x}, \bar{y}, \bar{s})$ regulär, und (x, y, s) sind stetige Funktionen von w_1 für kleine $\|w_1\|$. Falls $\|w_1\|$ klein ist, so ist $(x, y, s)(w_{(1)}) \approx (x, y, s)(\lambda w_{(1)})$ für $\lambda \in [0, 1]$. Insbesondere bleibt auch $y = y(x, w_{(1)})$ beim Übergang von $w_{(1)}$ zu $\lambda w_{(1)}$ nahezu unverändert. Aber in (B) ist $y_{(1)}(x, \lambda w_{(1)}) = \lambda y_{(1)}(x, w_{(1)})$, d.h. beim Übergang von $w_{(1)}$ zu $\lambda w_{(1)}$ in (B) werden die (impliziten) Multiplikatoren $y_{(1)}$ weitgehend " zerstört". Der Newtonschritt für (B) mit dem neuen λw ist aber wie oben erkannt gerade der Newtonschritt für $(*)$, wobei in $(*)$ die " guten" Schätzwerte $y_{(1)}$ durch " schlechte" $\lambda y_{(1)}$ ersetzt werden (und man sich für z.B. $\lambda = 0.1$ so " sinnlos" von der gefundenen Näherung (x, y, s) entfernt).

Eine "Reparatur" der klassischen Barrieremethode, die diese Zerstörung der Schätzwerte $y_{(1)}$ unterbindet wurde 1994 von Conn, Gould und Toint vorgeschlagen. Sie läuft allerdings im wesentlichen auf ein primal-duales Verfahren hinaus.

Literatur

[1] E.D. Andersen and Y. Ye, "A computational study of the homogeneous algorithm for large-scale convex optimization", Publications from Department of Management no. 3/1996, Odense University, Denmark (1996).

[2] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press (1985).

[3] J.E. Nesterov and A.S. Nemirovsky, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications* (SIAM, Philadelphia, 1994).

[4] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N.J. (1970).

[5] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions*, Grundlehren der Mathematischen Wissenschaften 163, Springer (1970).