Optimierung III
Nichtlineare Optimerung

Dr. Ralf Gollmer
Universität Duisburg-Essen
Institut für Mathematik

27. Juli 2006

Inhaltsverzeichnis

1 Einleitung

Gegenstand der Optimierung ist die Untersuchung von Extremalaufgaben, insbesondere unter Nebenbedingungen, wichtige Aspekte dabei sind:

- Existenz von Lösungen
- Charakterisierung von Lösungen (notwendige und hinreichende Bedingungen)
- Lösungsverfahren
- Empfindlichkeit der Lösungen gegenüber Fehlern in den Problemfunktionen bzw. Änderungen derselben (Sensitivitätsanalyse, parametrische Optimierung)

In dieser Vorlesung werden uns Probleme im Endlich-dimensionalen der allgemeinen Form

$$\min\{f(x) \,:\, x \in M\} \quad \text{mit} \quad M = \left\{ x \in \mathbb{R}^n \,:\, \begin{array}{rcll} g_i(x) & \leq & 0, & i = 1, \ldots, m, \\ h_j(x) & = & 0, & j = 1, \ldots, l \end{array} \right\}$$

beschäftigen. Die Funktion f in diesem Problem heißt Zielfunktion, ein Punkt $x \in M$ heißt zulässiger Punkt des Problems, die Menge M Restriktionsbereich oder zulässige Menge, die Ungleichungen und Gleichungen in der Beschreibung von M heißen Restriktionen.

Extremalaufgaben sind aus den bisherigen Vorlesungen nicht unbekannt, wir wiederholen die aus der Analysis-Vorlesung bekannten Aussagen.

Definition 1.1 *Gegeben sei eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$, $M \subseteq \mathbb{R}^n$, $x^0 \in M$.*
$B(x^0, \varepsilon)$ bezeichne die offene Kugel um x^0 mit Radius ε: $B(x^0, \varepsilon) = \{x \in \mathbb{R}^n \,:\, \|x - x^0\|_2 < \varepsilon\}$.

i) x^0 heißt globales Minimum von f auf M, falls
$$f(x^0) \leq f(x) \quad \forall x \in M;$$

ii) x^0 heißt lokales Minimum von f auf M, falls
$$\exists \varepsilon > 0 : \; f(x^0) \leq f(x) \quad \forall x \in M \cap B(x^0, \varepsilon);$$

iii) x^0 heißt striktes lokales Minimum von f auf M, falls
$$\exists \varepsilon > 0 : \; f(x^0) < f(x) \quad \forall x \in M \cap B(x^0, \varepsilon) \setminus \{x^0\};$$

iv) x^0 heißt isoliertes lokales Minimum von f auf M, falls es ein $\varepsilon > 0$ gibt, so daß x^0 einziges lokales Minimum in $M \cap B(x^0, \varepsilon)$ ist;

v) Besitzt f alle partiellen Ableitungen erster Ordnung in x^0 und ist $\frac{\partial f}{\partial x_i}(x^0) = 0 \quad \forall i = 1, \ldots, n$, so heißt x^0 stationärer (oder kritischer) Punkt von f.

Existenzsatz:

Satz 1.2 (Satz von Weierstraß - Satz 4.14 Ana I)
Seien (X, d) ein metrischer Raum, $\emptyset \neq K \subseteq X$ kompakt und $f : X \to \mathbb{R}$ stetig.
Dann existieren $x^0, x^1 \in K$, so daß
$f(x^0) = \inf\{f(x) : x \in K\}$ und $f(x^1) = \sup\{f(x) : x \in K\}$.

Abschwächung für Minimum-Problem:

Satz 1.3 (Satz 4.16 Ana I)
Seien (X, d) metrischer Raum, $\emptyset \neq K \subseteq X$ kompakt $f : X \to \mathbb{R}$ unterhalbstetig
(d.h. $\forall x \in X \;\; \forall (x_n) : \quad x_n \underset{n \to \infty}{\to} x \;\Rightarrow\; f(x) \leq \liminf\limits_{n \to \infty} f(x_n)$)

Dann existiert ein $x^0 \in K$, so daß $\quad f(x^0) = \inf\{f(x) : x \in K\}$.

notwendige Optimalitätsbedingung:

Satz 1.4 (Satz 5.50 Ana II)
$f : \mathbb{R}^n \to \mathbb{R}$ besitze in $x^0 \in C$ alle Richtungsableitungen, $C \subseteq \mathbb{R}^n$ sei konvex und x^0 lokales Minimum von f auf C. Dann gilt

$$f'(x^0, x - x^0) \geq 0 \quad \forall x \in C.$$

Ist außerdem $x^0 \in int\, C$, so ist x^0 stationärer Punkt von f.

Die notwendige Bedingung für innere Punkte führt im Fall differenzierbarer Funktionen und $C = \mathbb{R}^n$ zu einem Verfahren, bei welchem stationäre Punkte bestimmt werden:
Es sind Punkte zu finden, welche Lösung eines nichtlinearen Gleichungssystems sind. Dafür bietet sich das Newton-Verfahren an, welches eine quadratische Konvergenzrate (Numerik-Vorlesung) besitzt.

hinreichende Optimalitätsbedingung:

Satz 1.5 (Satz 5.52 Ana II)
Seien $G \subseteq \mathbb{R}^n$ offen und konvex, $f \in C^2(G)$, $C \subseteq G$ konvex und es gelte für $x^0 \in C$:
i) $< \nabla f(x^0), x - x^0 > \geq 0 \quad \forall x \in C$ und
ii) $< \nabla^2 f(x^0)h, h > \; > 0 \quad \forall h \in \mathbb{R}^n \setminus \{0\}$ (Hessematrix positiv definit)
dann ist x^0 striktes lokales Minimum von f auf C.

Folgerung 1.6 (Folgerung 5.54 Ana II)
Es seien die Voraussetzungen von Satz 1.5 erfüllt, wobei die Voraussetzung ii) ersetzt wird durch $< \nabla^2 f(y)(x - x^0), x - x^0 > \; \geq 0 \quad \forall x, y \in C$.
Dann ist x^0 sogar globales Minimum von f auf C.

notwendige Optimalitätsbedingung unter Gleichungsrestriktionen (Lagrange-Multiplikatoren):

Satz 1.7 (Satz 5.62 Ana II)
Es sei $X \subseteq \mathbb{R}^n$ offen, $f : X \to \mathbb{R}$ sowie $h : X \to \mathbb{R}^m \quad (m < n)$ seien stetig differenzierbar. f besitze in x^0 ein lokales Minimum auf $C = \{x \in X \; : \; h(x) = 0\}$. Außerdem besitze die Jacobimatrix $J_h(x^0)$ Vollrang. Dann existieren $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, so daß

$$\nabla f(x^0) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^0) = 0.$$

(Beweis mittels Satz über implizite Funktionen)

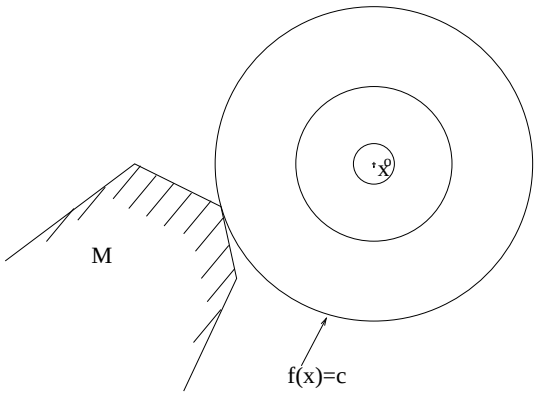
Dieses Prinzip der notwendigen Bedingung wird auf den Fall von Gleichungs- und Ungleichungsrestriktionen verallgemeinert. Die vorangegangenen Sätze zeigen die besondere Bedeutung der Konvexität der Restriktionsmenge und der Zielfunktion in diesem Zusammenhang.

Ein spezieller Fall solcher Probleme, in denen sowohl die Zielfunktion als auch die Restriktionsfunktionen g_i, h_j linear (und somit konvex) sind, wird unter Ausnutzung der speziellen Eigenschaften behandelt und ist Gegenstand der Vorlesung Lineare Optimierung (Optimierung I).

Es folgen nun einige Beispiele nichtlinearer Probleme.

Beispiel 1.8 (minimaler Abstand) *Gegeben sei eine Menge $M \subseteq \mathbb{R}^n$ welche durch lineare Ungleichungen beschrieben ist, d.h. $M = \{x \in \mathbb{R}^n \; : \; Ax \leq b\}$, und ein Punkt $x^0 \in \mathbb{R}^n$. Gesucht ist ein Punkt in M, welcher den euklidischen Anstand zu x^0 minimiert. Da die Wurzel eine monoton wachsende Funktion ist, ist dies äquivalent zur Minimierung des Quadrates des Abstandes, was ein einfacheres Problem liefert:*

$$\min\{(x - x^0)^T (x - x^0) \; : \; Ax \leq b\}$$



Dies ist der einfachste Spezialfall eines nichtlinearen Problems: die Zielfunktion ist quadratisch, die Restriktionsfunktionen linear. Ein solches Problem wird als quadratisches Optimierungsproblem bezeichnet.

Beispiel 1.9 (optimaler Entwurf eines Trägers) *Wir betrachten einen homogenen Träger mit rechteckigem Querschnitt und gegebener Länge l. Seine Höhe x_1 und die Breite x_2 sind so zu bestimmen, daß der Träger minimales Gewicht hat und die folgenden Restriktionen erfüllt sind:*

- *Die auftretenden Spannungen unter der maximal zulässigen Last dürfen einen vorgegebenen Wert nicht überschreiten. Dies führt zu einer Ungleichung der Form*
$$u \leq x_1^2 x_2$$
mit gegebenem u.
- *Damit der Träger nicht zu dünn wird, werden folgende Beziehungen für seine Maße gefordert:*
$$\begin{array}{rcl} x_1 & \leq & 4x_2 \\ x_2 & \leq & x_1 \end{array}$$
- *Außerdem sind natürlich nur nichtnegative Werte für x_1 und x_2 sinnvoll.*

Insgesamt lautet das Problem in der obigen Standardform

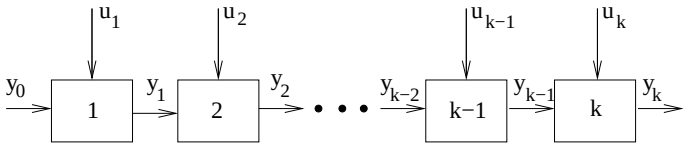
$$\min\{\, lx_1x_2 \;:\; \left.\begin{array}{rcl} u - x_1^2x_2 & \leq & 0 \\ x_1 - 4x_2 & \leq & 0 \\ -x_1 + x_2 & \leq & 0 \\ -x_1 & \leq & 0 \\ -x_2 & \leq & 0 \end{array}\right\}$$

Optimalsteuerungsprobleme:

Beispiel 1.10 (Diskrete Optimalsteuerung) *Gegeben seien k Perioden fixierter Länge. Es wird ein System betrachtet, dessen Zustand zu Beginn von Periode i durch den Zustandsvektor y_{i-1} repräsentiert wird. Der Anfangszustand y_0 sei gegeben. In der Periode i wirkt ein Steuerungsvektor u_i, so daß am Ende der Periode sich der neue Zustand gemäß*

$$y_i = \Phi(y_{i-1}, u_i), \quad i = 1, \ldots, k$$

verändert.



Sowohl für Zustand als auch für Steuerungen gelten Beschränkungen

$$\begin{aligned} y_i &\in Y_i, \quad i = 1, \ldots, k, \\ u_i &\in U_i, \quad i = 1, \ldots, k, \end{aligned}$$

weiterhin ist eine sogenannte Trajektorien-Restriktion der Form

$$\Psi(y_0, y_1, \ldots, y_k, u_1, \ldots, u_k) \in D$$

möglich. Dabei sind Y_i, U_i, D vorgegebenen Mengen, Ψ eine gegebene Funktion. Unter diesen Restriktionen ist eine Zielfunktion $f(y_0, y_1, \ldots, y_k, u_1, \ldots, u_k)$ zu minimieren.
$\longrightarrow$ Verbindung zu Optimierung I, dynamische Optimierung

Beispiel 1.11 (Stetige Optimalsteuerung) *In Fortsetzung des vorigen Beispiels betrachten wir nun nicht Intervalle positiver Länge und die Steuerung wirkt nicht an diskreten Punkten, sondern kontinuierlich. Wir erhalten ein unendlichdimensionales Problem:*
Sei ein Problem mit fixiertem Planungshorizont $[0, T]$ betrachtet, dann ist der Gegenstand der Optimierung eine Funktion $u : [0, T] \to \mathbb{R}^m$. Die Transformationsgleichung wird als gewöhnliche Differentialgleichung angesetzt:

$$y' = \Phi(y(t), u(t)), \quad t \in [0, T]$$

Die Zustands-, Steuer- und Trajektorien-Restriktionen erhalten die Form

$$\begin{aligned} y &\in Y \\ u &\in U \\ \Psi(y, u) &\in D \end{aligned}$$

Als typisches Beispiel hat man U als Teilmenge der stückweise stetigen Funktionen auf $[0, T]$ mit $a \leq u(t) \leq b \quad \forall t \in [0, T]$.
Die Zielfunktion hat jetzt Integralform, so daß das Problem insgesamt lautet:

$$\left.\begin{aligned} \min \{ \int_0^T f[y(t), u(t)]dt \ : \quad & y' = \Phi(y(t), u(t)) \quad t \in [0, T], \\ & y(t) \in Y \\ & u(t) \in U \\ & \Psi(y, u) \in D \end{aligned}\right\}$$

Bekanntestes Beispiel in fast jedem Buch über Steuerungstheorie ist die weiche Landung einer Rakete mit minimalem Treibstoffverbrauch.
Solche Probleme sind Gegenstand einer eigenständigen (unendlichdimensionalen) Theorie.
Approximativ kann man jedoch das Problem durch Diskretisierung der Zeit und Ersetzen der Differentialgleichung durch Differenzengleichungen (z.B. Eulerschen Polygonzug) als nichtlineares Optimierungsproblem behandeln und so mittels Verfahren der endlich-dimensionalen Optimierung näherungsweise Lösungen erhalten.

2 Konvexe Analysis

2.1 Konvexe Mengen

Definition 2.1 *Die Menge $S \subseteq \mathbb{R}^n$ heißt konvex, wenn für alle $x, y \in S$ und alle $\lambda \in [0,1]$ gilt $\lambda x + (1-\lambda)y \in S$.*

Bemerkung 2.2

Übung: *S ist konvex $\Leftrightarrow$ für jedes $k \in \mathbb{N}$ und Punkte $x^1, \ldots, x^k \in S$ liegt jede Konvexkombination*

$$\sum_{j=1}^{k} \lambda_j x^j \quad mit \quad \lambda_j \in \mathbb{R},\ \lambda_j \geq 0\ j = 1, \ldots, k, \quad \sum_{j=1}^{k} \lambda_j = 1$$

ebenfalls in S.

Bezeichnung: $\sum_{j=1}^{k} \lambda_j x^j$ *mit* $\lambda_j \in \mathbb{R},\ j = 1, \ldots, k, \quad \sum_{j=1}^{k} \lambda_j = 1$ *heißt affine Kombination,*

$\sum_{j=1}^{k} \lambda_j x^j$ *mit* $\lambda_j \in \mathbb{R}$ *beliebig heißt Linearkombination der x^j.*

Beispiele konvexer Mengen:

a) $a \in \mathbb{R}^n,\ a \neq 0,\ \alpha \in \mathbb{R}$ beliebig fix.
$\{\ x \in \mathbb{R}^n\ :\ a^T x = \alpha\}$ (Hyperebene),
$\{\ x \in \mathbb{R}^n\ :\ a^T x \leq \alpha\}$ (Halbraum).

b) A eine $m \times n$-Matrix, $b \in \mathbb{R}^m$
$\{\ x \in \mathbb{R}^n\ :\ Ax \leq b\ \}$ (konvexes Polyeder)

c) $B(x^0, \varepsilon) = \{\ x \in \mathbb{R}^n\ :\ \|x - x^0\|_2 < \varepsilon\ \}$ (offene Kugel mit Radius ε um x^0)

Lemma 2.3 *i) Sei S_i, $i \in I$ eine Familie konvexer Teilmengen der $\mathbb{R}^n$ mit beliebiger Indexmenge I. Dann ist $\bigcap_{i \in I} S_i$ konvex,*

ii) Seien S_1, S_2 konvexe Teilmengen des $\mathbb{R}^n$. Dann sind
$S_1 \oplus S_2 = \{x \in \mathbb{R}^n : x = x^1 + x^2,\ x^1 \in S_1,\ x^2 \in S_2\ \}$ konvex
(Minkowski-Summe zweier Mengen),
und $S_1 \ominus S_2 = \{x \in \mathbb{R}^n : x = x^1 - x^2,\ x^1 \in S_1,\ x^2 \in S_2\ \}$ konvex.

iii) Es sei A eine $m \times n$-Matrix, $b \in \mathbb{R}^m$. Das Bild $f(C)$ einer konvexen Menge C bei der affinen Abbildung $f :\ x \in \mathbb{R}^n \longrightarrow Ax + b \in \mathbb{R}^m$ ist konvex.

Beweis: ÜA

Definition 2.4 *Sei $S \subseteq \mathbb{R}^n$ beliebig. Die Menge*

a) $$conv\ S = \bigcap_{\substack{C \supseteq S \\ C\ konvex}} C \quad heißt\ konvexe\ Hülle\ von\ S.$$

b) $$aff\ S = \bigcap_{\substack{A \supseteq S \\ A\ affiner\ UR}} A \quad heißt\ affine\ Hülle\ von\ S.$$

c) $$lin\ S = \bigcap_{\substack{L \supseteq S \\ L\ linearer\ UR}} L \quad heißt\ lineare\ Hülle\ von\ S.$$

Bemerkung 2.5

a) conv S ist die bezüglich der Mengeninklusion kleinste konvexe Menge, die S enthält, es gilt:

$$conv\ S = \{\ x \in \mathbb{R}^n\ :\ \left.\begin{array}{l} \exists\ k \in \mathbb{N},\ x^1, \ldots, x^k \in S,\ \lambda_1, \ldots, \lambda_k \in \mathbb{R}\quad mit \\ x = \sum_{j=1}^{k} \lambda_j x^j,\ \ \sum_{j=1}^{k} \lambda_j = 1,\ \ \lambda_j \geq 0\ \ j = 1, \ldots, k \end{array}\right\}$$

b) aff S ist der bezüglich der Mengeninklusion kleinste affine Unterraum, der S enthält, es gilt:

$$aff\ S = \{\ x \in \mathbb{R}^n\ :\ \left.\begin{array}{l} \exists\ k \in \mathbb{N},\ x^1, \ldots, x^k \in S,\ \lambda_1, \ldots, \lambda_k \in \mathbb{R}\quad mit \\ x = \sum_{j=1}^{k} \lambda_j x^j,\ \ \sum_{j=1}^{k} \lambda_j = 1 \end{array}\right\}$$

c) lin S ist der bezüglich der Mengeninklusion kleinste lineare Unterraum, der S enthält, es gilt:

$$lin\ S = \{\ x \in \mathbb{R}^n\ :\ \left.\begin{array}{l} \exists\ k \in \mathbb{N},\ x^1, \ldots, x^k \in S,\ \lambda_1, \ldots, \lambda_k \in \mathbb{R}\quad mit \\ x = \sum_{j=1}^{k} \lambda_j x^j \end{array}\right\}$$

Bezeichnung 2.6

int S	*Inneres von S:*	$x \in int\ S$	$\iff$	$\exists \varepsilon > 0\ \ B(x,\varepsilon) \subseteq S$
cl S	*Abschließung von S:*	$x \in cl\ S$	$\iff$	$\forall \varepsilon > 0\ \ S \cap B(x,\varepsilon) \neq \emptyset$
bd S	*Rand von S:*	$x \in bd\ S$	$\iff$	$\forall \varepsilon > 0\ \ S \cap B(x,\varepsilon) \neq \emptyset\ \wedge$ $(\mathbb{R}^n \setminus S) \cap B(x,\varepsilon) \neq \emptyset$

Lemma 2.7 *S konvex* $\implies$ *cl S konvex.*

Beweis: B bezeichne die Einheitskugel um den Ursprung: $B = B(0,1)$.
$y \in cl\ S$ gdw. $\forall\ \varepsilon > 0\ \exists x \in S \cap B(y,\varepsilon) = S \cap y \oplus \varepsilon B$. Damit natürlich $y \in x \oplus \varepsilon B$. Wir betrachten die Vereinigung aller solcher Kugeln um die Punkte aus S. $S \oplus \varepsilon B$ ist konvex für alle $\varepsilon > 0$ als Linearkombination konvexer Mengen. Es ist nun $cl\ S = \bigcap_{\varepsilon > 0} (S \oplus \varepsilon B)$. Folglich ist $cl\ S$ als Durchschnitt konvexer Mengen wieder konvex. □

Lemma 2.8 *Es sei* $S \subseteq \mathbb{R}^n$ *konvex.*
Wenn $x^1 \in cl\ S$ *und* $x^2 \in int\ S$, *so* $y = \lambda x^1 + (1-\lambda)x^2 \in int\ S$ *für jedes* $\lambda \in]0,1[$.

Beweis: Es existiert $\varepsilon > 0$ mit $B(x^2,\varepsilon) \subseteq S$. Wir zeigen, daß $B(y,(1-\lambda)\varepsilon) \subseteq S$. Sei $z \in B(y,(1-\lambda)\varepsilon)$ beliebig. $x^1 \in cl\ S \implies \exists z^1 \in S$ mit

$$\|z^1 - x^1\| < \frac{(1-\lambda)\varepsilon - \|z-y\|}{\lambda}$$

Betrachten dann

$$z^2 = \frac{z - \lambda z^1}{1-\lambda}$$

Es gilt

$$\begin{array}{rcll} \|z^2 - x^2\| & = & \|\frac{z-\lambda z^1}{1-\lambda} - \frac{y-\lambda x^1}{1-\lambda}\| & \\ & = & \frac{1}{1-\lambda}\|(z-y) + \lambda(x^1 - z^1)\| & \\ & \leq & \frac{1}{1-\lambda}(\|z-y\| + \lambda\|x^1 - z^1\|) & < \varepsilon \end{array}$$

$\implies z^2 \in B(x^2,\varepsilon) \implies z = \lambda z^1 + (1-\lambda)z^2 \in S$ □

Folgerung 2.9 *S konvex* $\implies$ *int S konvex.*

Folgerung 2.10 *Es sei S konvex, int* $S \neq \emptyset$. *Dann gilt*

i) $cl(int\ S) = cl\ S$

ii) $int(cl\ S) = int\ S$

Beweis:

i) Trivialerweise $cl(int\ S) \subseteq cl\ S$. Sei nun $x \in cl\ S$. Wählen $y \in int\ S$, dann gilt
$$\lambda x + (1-\lambda)y \in int\ S \quad \forall \lambda \in]0,1[.$$
Grenzwert $\lambda \uparrow 1$ liefert $x \in cl(int\ S)$.

ii) Trivial: $int\ S \subseteq int(cl\ S)$. Sei nun $x^1 \in int(cl\ S)$, es existiert also $\varepsilon > 0$, so daß aus $\|y - x^1\| < \varepsilon$ folgt $y \in cl\ S$. Sei nun $x^2 \neq x^1$ aus $int\ S$ gewählt und $y = (1+\delta)x^1 - \delta x^2$ mit
$$\delta = \frac{\varepsilon}{2\|x^1 - x^2\|}$$
Dann ist $\|y - x^1\| = \|\delta(x^1 - x^2)\| = \frac{\varepsilon}{2}$, somit $y \in cl\ S$. Aber x^1 ist als Konvexkombination von y und x^2 darstellbar:
$$x^1 = \lambda y + (1-\lambda)x^2 \quad \text{mit} \quad \lambda = \frac{1}{1+\delta} \in]0,1[$$
und Lemma 2.8 liefert $x^1 \in int\ S$. □

2.2 Stützhyperebenen und Trennungssätze

In diesem Abschnitt werden wir uns mit der Trennbarkeit zweier konvexer Mengen durch eine Hyperebene und mit Stützhyperebenen an konvexe Mengen beschäftigen.

Der folgende Satz behandelt ein spezielles Optimierungsproblem und liefert gleichzeitig ein wesentliches Hilfsmittel für die nachfolgenden Beweise.

Satz 2.11 *Sei $S \neq \emptyset$ eine konvexe abgeschlossene Teilmenge der $\mathbb{R}^n$ und $y \notin S$. Dann gibt es genau einen Punkt $\bar{x} \in S$, welcher globaler Minimalpunkt für das Problem*
$$\min \{\ \|y - x\|_2 :\ x \in S\ \}$$
ist (d.h. den minimalen euklidischen Abstand zum Punkt y realisiert).
Weiterhin ist die Bedingung
$$(y - \bar{x})^T (x - \bar{x}) \leq 0 \quad \forall x \in S$$
notwendig und hinreichend für die Optimalität von $\bar{x}$.

Beweis: Existenz: Wegen $S \neq \emptyset$ existiert ein $\hat{x} \in S$ und das Problem kann auf die Menge $\hat{S} = S \cap \{x \in \mathbb{R}^n : \|y - x\| \leq \|y - \hat{x}\|\}$ eingeschränkt werden. Nun ist aber die Norm eine stetige Funktion, $\hat{S}$ eine kompakte Menge, der Satz von Weierstraß (Satz 1.2) liefert die Existenz eines Minimalpunktes.
Eindeutigkeit: Sei $x' \in S$ mit $\|y - \bar{x}\| = \|y - x'\| = \gamma$. Aufgrund der Konvexität von S ist $\frac{1}{2}(\bar{x} + x') \in S$ und die Dreiecksungleichung liefert
$$\|y - \frac{1}{2}(\bar{x} + x')\| \leq \frac{1}{2}\|y - \bar{x}\| + \frac{1}{2}\|y - x'\| = \gamma$$
Die strikte Ungleichung würde einen Widerspruch zur Optimalität von $\bar{x}$ liefern, also gilt Gleichheit. Damit muß das Dreieck entartet sein und es ex. ein $\lambda \in \mathbb{R}$, so daß $(y - \bar{x}) = \lambda(y - x')$. Wegen der Gleichheit der Norm gilt $|\lambda| = 1$. $\lambda = -1$ ist unmöglich, da sonst $y = \frac{1}{2}(\bar{x} + x') \in S$, also $\lambda = 1$, d.h. $\bar{x} = x'$.
Bed. ist hinreichend: Sei $x \in S$ beliebig. Dann ist
$$\|y - x\|^2 = \|y - \bar{x} + \bar{x} - x\|^2 = \|y - \bar{x}\|^2 + \underbrace{\|\bar{x} - x\|^2}_{\geq 0} + 2\underbrace{(\bar{x} - x)^T(y - \bar{x})}_{\geq 0 \text{ (Vor.)}} \geq \|y - \bar{x}\|$$
Bed. ist notwendig: Sei $\|y - x\|^2 \geq \|y - \bar{x}\|^2 \quad \forall x \in S$. Wir wählen $x \in S$ beliebig $\Longrightarrow$
$\bar{x} + \lambda(x - \bar{x}) \in S \quad \forall \lambda \in [0,1] \quad \Longrightarrow \quad \|y - \bar{x} - \lambda(x - \bar{x})\|^2 \geq \|y - \bar{x}\|^2$ und es gilt wieder
$$\|y - \bar{x} - \lambda(x - \bar{x})\|^2 = \|y - \bar{x}\|^2 + \lambda^2\|x - \bar{x}\|^2 - 2\lambda(y - \bar{x})^T(x - \bar{x})$$
$$\Longrightarrow \quad 2\lambda(y - \bar{x})^T(x - \bar{x}) \leq \lambda^2\|x - \bar{x}\|^2 \quad \forall \lambda \in [0,1]$$

Wir betrachten $\lambda > 0$, dividieren die letzte Ungleichung durch 2λ und lassen λ gegen Null streben
$\Longrightarrow \quad (y - \bar{x})^T(x - \bar{x}) \leq 0.$ □

Die Bedingung besagt, daß S in dem Halbraum $a^T(x - \bar{x}) \leq 0$ liegt, dessen zugehörige Hyperebene den Punkt $\bar{x} \in S$ enthält und für deren Normalenvektor gilt $a = y - \bar{x}$. Diesen Begriff wollen wir nun in eine Definition fassen.

Definition 2.12 *Sei S eine nichtleere Menge in $\mathbb{R}^n$ und sei $\bar{x} \in bd\, S$. Eine Hyperebene $H = \{x \in \mathbb{R}^n : a^T(x - \bar{x}) = 0\}$ heißt Stützhyperebene (supporting hyperplane) an S in $\bar{x}$, wenn $S \subseteq H^+ = \{x \in \mathbb{R}^n : a^T(x - \bar{x}) \geq 0\}$ oder $S \subseteq H^- = \{x \in \mathbb{R}^n : a^T(x - \bar{x}) \leq 0\}$.*
Ist weiterhin $S \not\subseteq H$, so heißt H echte Stützhyperebene.

Zunächst soll uns jedoch die Trennung konvexer Mengen beschäftigen.

Definition 2.13 *Seien $S_1,\ S_2 \subseteq \mathbb{R}^n$ nichtleer. Eine Hyperebene $H = \{x \in \mathbb{R}^n : a^T x = \alpha\}$ heißt trennende Hyperebene (separating hyperplane) für S_1 und S_2, wenn $S_1 \subseteq H^+$ und $S_2 \subseteq H^-$.*
Ist außerdem $S_1 \cup S_2 \not\subset H$, so sagt man, H sei eine echte trennende Hyperebene (properly separating hyperplane).
H heißt strikt trennende Hyperebene (strictly separating hyperplane) für S_1 und S_2, wenn $a^T x > \alpha \quad \forall x \in S_1 \quad$ und $\quad a^T x < \alpha \quad \forall x \in S_2$.
H heißt stark trennende Hyperebene (strongly separating hyperplane) für S_1 und S_2, wenn $\exists \varepsilon > 0 : \quad a^T x \geq \alpha + \varepsilon \quad \forall x \in S_1 \quad$ und $\quad a^T x \leq \alpha \quad \forall x \in S_2$

Bemerkung: Die Definition der strikten Trennung mit zwei strikten Ungleichungen ist aus Bazaraa, Sherali, Shetty.

Nun folgt zunächst der einfachste Trennungssatz über die Trennung einer konvexen Menge von einem Punkt. Aus diesem werden sich weitere Trennungsaussagen als Folgerung ergeben.

Satz 2.14 *Seien $\emptyset \neq S \subseteq \mathbb{R}^n$ konvex und abgeschlossen sowie $y \notin S$. Dann existiert eine stark trennende Hyperebene für S und $\{y\}$.*

Beweis: Nach Satz 2.11 existiert in S genau ein Punkt $\bar{x}$ minimalen Abstands zu y und es gilt $(x - \bar{x})^T(y - \bar{x}) \leq 0 \quad \forall x \in S$. Wir wählen $a = y - \bar{x} \neq 0$ und $\alpha = a^T\bar{x}$. Dann gilt $a^T x \leq \alpha \quad \forall x \in S$ und $\quad a^T y - \alpha = \|y - \bar{x}\|^2 > 0$ (wegen $y \notin S$). □

Bemerkung: Im vorliegenden Falle, wo eine der beiden Mengen einpunktig ist und die andere abgeschlossen, fallen die Begriffe starke und strikte Trennung zusammen.

Folgerung 2.15 *Sei $\emptyset \neq S \subseteq \mathbb{R}^n$ konvex und abgeschlossen. Dann ist S gleich dem Durchschnitt aller abgeschlossenen Halbräume, welche S enthalten.*

Beweis: ÜA

Folgerung 2.16 *Seien $\emptyset \neq S \subseteq \mathbb{R}^n$ und $y \notin cl\ conv\ S$. Dann existiert eine stark trennende Hyperebene für S und $\{y\}$.*

Beweis: ÜA: Setze *cl conv S* an die Stelle von S im Satz 2.14.

Satz 2.17 *Seien $\emptyset \neq S \subset \mathbb{R}^n$ konvex und $\bar{x} \in bd\, S$. Dann existiert eine Stützhyperebene an S in $\bar{x}$.*

Beweis: Da $\bar{x} \in bd\ S$, existiert eine Folge $y^k \notin cl\ S$ mit $\lim\limits_{k\to\infty} y^k = \bar{x}$. Nach Satz 2.14 existiert zu jedem y^k ein p_k (welches auf Norm 1 normiert sei), so daß ${p_k}^T y^k > {p_k}^T x \quad \forall x \in cl\ S$. Da (p_k) eine beschränkte Folge ist, besitzt sie eine konvergente Teilfolge, für deren Grenzwert p ebenfalls $\|p\| = 1$ gilt. Betrachten wir diese Teilfolge und ein fixiertes $x \in cl\ S$. Es gilt ${p_k}^T y^k > {p_k}^T x$, im Grenzwert erhalten wir $p^T\bar{x} \geq p^T x$. Da also für jedes $x \in cl\ S$ gilt $p^T(x - \bar{x}) \leq 0$, ist die Behauptung bewiesen. □

Folgerung 2.18 *Seien $\emptyset \neq S \subseteq \mathbb{R}^n$ konvex und $\bar{x} \notin int\ S$. Dann existiert ein Vektor $a \neq 0$, so daß $a^T(x - \bar{x}) \leq 0 \quad \forall x \in cl\ S$.*

Beweis: Für $\bar{x} \notin cl\ S$ folgt das Resultat aus Satz 2.14, für $\bar{x} \in bd\ S$ ist es gerade die Aussage von Satz 2.17

Folgerung 2.19 *Seien $\emptyset \neq S \subseteq \mathbb{R}^n$ und $y \notin int\ conv\ S$. Dann existiert eine Hyperebene, die S und $\{y\}$ trennt.*

Folgerung 2.20 *Seien $\emptyset \neq S \subseteq \mathbb{R}^n$ und $\bar{x} \in bd\ S \cap bd\ conv\ S$. Dann existiert eine Stützhyperebene an S in $\bar{x}$.*

Satz 2.21 (Trennungssatz) *Es seien S_1 und S_2 konvexe, nichtleere Teilmengen des $\mathbb{R}^n$ und es sei $S_1 \cap S_2 = \emptyset$. Dann existiert eine Hyperebene, welche S_1 und S_2 trennt.*

Beweis: Wir betrachten die Menge $S = S_1 \ominus S_2$, welche nach Lemma 2.3 konvex ist. Nach Voraussetzung ist weiterhin $0 \notin S$. Folgerung 2.18 liefert die Existenz eines Vektors $a \neq 0$ mit $a^T x \leq 0 \quad \forall x \in S$, d.h. $a^T x^1 \leq a^T x^2 \quad \forall x^1 \in S_1,\ x^2 \in S_2$. □

Die Voraussetzung der Disjunktheit der zu trennenden Mengen läßt sich abschwächen.

Folgerung 2.22 *Es seien S_1 und S_2 Teilmengen des $\mathbb{R}^n$ mit $int\ conv\ S_i \neq \emptyset \quad i = 1, 2$, aber $int\ conv\ S_1 \cap int\ conv\ S_2 = \emptyset$. Dann existiert eine Hyperebene, welche S_1 und S_2 trennt.*

Beweis: Man ersetzt im Satz 2.21 die Menge S_i durch *int conv* S_i und beachtet, daß $\sup\{a^T x\ :\ x \in S\} = \sup\{a^T x\ :\ x \in int\ S\}$ □

Satz 2.23 (Starke Trennung) *Es seien S_1 und S_2 konvexe, abgeschlossene, nichtleere Teilmengen des $\mathbb{R}^n$ und es sei S_1 beschränkt. Wenn $S_1 \cap S_2 = \emptyset$, so existiert eine Hyperebene, welche S_1 und S_2 stark trennt.*
(d.h. $\inf\{a^T x\ :\ x \in S_1\} \geq \varepsilon + \sup\{a^T x\ :\ x \in S_2\}$ für ein $\varepsilon > 0$)

Beweis: Es sei wieder $S = S_1 \ominus S_2$ betrachtet, welches konvex ist und $0 \notin S$ erfüllt. Nun soll gezeigt werden, daß unter den Voraussetzungen des Satzes S abgeschlossen ist. Dazu sei eine konvergente Folge (x^k) von Punkten aus S betrachtet. Ihr Grenzwert sei x. Nach der Bildungsvorschrift für S existieren $y^k \in S_1,\ z^k \in S_2$ mit $x^k = y^k - z^k$. Da S_1 kompakt ist, hat die Folge (y^k) eine konvergente Teilfolge (y^{k_i}) mit Grenzwert y in S_1. Aus $y^{k_i} - z^{k_i} \longrightarrow x$ und $y^{k_i} \longrightarrow y$ für $i \longrightarrow \infty$ folgt $z^{k_i} \longrightarrow z$. Wegen der Abgeschlossenheit von S_2 folgt $z \in S_2$. Also ist $x = y - z \in S$ und dies zeigt die Abgeschlossenheit von S.
Satz 2.14 liefert nun die Existenz einer stark trennenden Hyperebene für S und $\{0\}$, woraus sich die Behauptung ergibt. □

ÜA: Geben Sie ein Beispiel im $\mathbb{R}^2$ an, welches zeigt, daß für starke Trennung auf die Voraussetzung der Beschränktheit wenigstens einer der Mengen nicht verzichtet werden kann!

Die Voraussetzungen der angegebenen Trennungssätze lassen sich weiter abschwächen. Für die echte Trennung geschieht dies, indem wir ein weiteres topologisches Konzept einführen, welches für konvexe Mengen geeignet ist. Solche Mengen haben ein "natürliches" Inneres, auch wenn ihr Inneres im Sinne der Definition in Bezeichnung 2.6 leer ist. Betrachten Sie als Beispiel ein ebenes Dreieck im $\mathbb{R}^3$.
Das Problem besteht, wenn die Menge eine Dimension kleiner als die des Raumes hat.
Als Dimension einer konvexen Menge bezeichnen wir die Dimension ihrer affine Hülle.

Bevor eine Aussage zur Dimension einer konvexen Menge bewiesen wird, wollen wir noch den Satz von Carathéodory beweisen, welcher Bemerkung 2.5 a) verschärft. Er besagt, daß es für die Bildung der konvexen Hülle einer Menge im $\mathbb{R}^n$ ausreicht, alle Konvexkombinationen von höchstens $n+1$ Punkten zu betrachten (bzw. genau $n+1$ Punkten, wenn man nicht fordert, daß alle diese Punkte paarweise verschieden sind):

Satz 2.24 (Satz von Carathéodory) *Sei $S \subseteq \mathbb{R}^n$. Ist $x \in conv\ S$, so existieren $n+1$ Punkte $x^1, \ldots, x^{n+1} \in S$ mit $x \in conv\ \{x^1, \ldots, x^{n+1}\}$.*

Beweis: Nach Bemerkung 2.5 existieren ein $k \in \mathbb{N}$, reelle Zahlen $\lambda_1, \ldots, \lambda_k$ und Punkte $x^1, \ldots, x^k$ mit $x = \sum_{j=1}^{k} \lambda_j x^j$, $\sum_{j=1}^{k} \lambda_j = 1$ und $\lambda_j \geq 0$, $x^j \in S \quad \forall j = 1, \ldots, k$. Für $k = n+1$ ist nichts zu zeigen, für $k < n+1$ genügt es, die Summe durch $\lambda_j = 0$, $x^j = x^k \quad j = k+1, \ldots, n+1$ zu ergänzen.
Sei nun ein $k > n+1$ und $\lambda_j > 0$, $j = 1, \ldots, k$. Dann sind die Punkte x^j, $j = 1, \ldots, k$ im n-dimensionalen Raum affin abhängig, d.h. $x^2 - x^1, \ldots, x^k - x^1$ sind linear abhängig.
Es existieren also Koeffizienten $\mu_j \in \mathbb{R}$, $j = 2, \ldots, k$, die nicht alle gleich Null sind und für welche $\sum_{j=2}^{k} \mu_j (x^j - x^1) = 0$. Setzen wir dann $\mu_1 = -\sum_{j=2}^{k} \mu_j$, so ist $\sum_{j=1}^{k} \mu_j x^j = 0$, $\sum_{j=1}^{k} \mu_j = 0$ und mindestens einer der Koeffizienten μ_j, $j = 1 \ldots, k$ ist positiv. Wir setzen nun

$$\alpha = \min \left\{ \frac{\lambda_j}{\mu_j}, : \mu_j > 0,\ j \in \{1, \ldots, k\} \right\} \ (> 0)$$

und i_0 sei ein Index, für welchen dieses Minimum realisiert wird. Es gilt für alle $\alpha \in \mathbb{R}$, also speziell auch für das so definierte

$$x = \sum_{j=1}^{k} \lambda_j x^j = \sum_{j=1}^{k} \lambda_j x^j - \alpha \sum_{j=1}^{k} \mu_j x^j = \sum_{j=1}^{k} (\lambda_j - \alpha \mu_j) x^j$$

Aufgrund der Wahl von α und i_0 gilt $\lambda_j - \alpha \mu_j \geq 0 \quad \forall j \in \{1, \ldots, k\}$ und $\lambda_{i_0} - \alpha \mu_{i_0} = 0$ sowie $\sum_{j=1}^{k} (\lambda_j - \alpha \mu_j) = \sum_{j=1}^{k} \lambda_j - \alpha \sum_{j=1}^{k} \mu_j = 1$. Somit haben wir eine Darstellung von x als Konvexkombination von $k-1$ Punkten gefunden. Dieses Argument läßt sich nun so lange wiederholen, bis $k = n+1$ erreicht ist. □

Lemma 2.25 *Die Dimension einer konvexen Menge C ist gleich dem Maximum der Dimensionen aller in C enthaltenen Simplizes.*

Beweis: Indem wir uns auf den Raum $aff\ C$ beschränken, ließe sich aus dem Satz 2.24 schließen, daß die Dimension der Menge größer oder gleich der Dimension jedes Simplex in C ist. Der Beweis läßt sich jedoch auch ohne Benutzung des Satzes führen:
Da C konvex ist, enthält es die konvexe Hülle jeder seiner Teilmengen. Die maximale Dimension der in C enthaltenen Simplizes ist also gleich dem größtem m, so daß C eine Menge von $m+1$ affin unabhängigen Punkten enthält. Sei $V = \{v^0, v^1, \ldots, v^m\}$ solch eine Menge mit maximalem m und $M = aff\ V$. Dann ist $dim\ M = m$ und $M \subseteq aff\ C$. Weiterhin gilt $C \subseteq M$, denn andernfalls enthielte $C \setminus M$ ein Element v^{m+1} und die Menge $\{v^0, v^1, \ldots, v^{m+1}\} \subset C$ enthielte $m+2$ affin unabhängige Punkte im Widerspruch zur Maximalität von m. Da per Definition $aff\ C$ der kleinste affine Unterraum ist, welcher C enthält, folgt $aff\ C = M$ und somit $dim\ C = m$. □

Definition 2.26 *Es sei $C \subseteq \mathbb{R}^n$ eine konvexe Menge. Als relatives Inneres (Bez.: ri C) bezeichnen wir das Innere in Bezug auf die affine Hülle von C:*

$$ri\ C = \{x \in aff\ C : \exists \varepsilon > 0 \quad B(x, \varepsilon) \cap aff\ C \subseteq C\}$$

C heißt relativ offen, wenn $C = ri\ C$.
Die Menge $rbd\ C = (cl\ C) \setminus (ri\ C)$ heißt der relative Rand von C.

Für Mengen voller Dimension (d.h. $aff\ C = R^n$) fallen die Begriffe relatives Inneres und Inneres zusammen.
Während gilt: $C_1 \subseteq C_2 \implies int\ C_1 \subseteq int\ C_2$, überträgt sich diese Monotonie-Eigenschaft nicht auf das relativ Innere. (Bsp.: C_2 Würfel im R^3, C_1 eine seiner Seitenflächen).

Abschließungen und relativ Innere werden unter bijektiven affinen Abbildungen des $\mathbb{R}^n$ auf sich selbst erhalten. Damit vereinfachen sich viele Betrachtungen in Beweisen: Wenn $aff\ C$ ein m-dimensionaler affiner Unterraum des $\mathbb{R}^n$ ist, so existiert eine bijektive affine Transformation, welche man sogar winkeltreu wählen kann (Translation und Drehung), die $aff\ C$ auf den linearen Unterraum

$$L = \{x \in \mathbb{R}^n \ : \ x_i = 0,\ i = m+1, \dots n\}$$

abbildet. L kann als eine Kopie des $\mathbb{R}^m$ angesehen werden und so lassen sich Fragen für allgemeine konvexe Mengen auf die Betrachtung volldimensionaler Mengen zurückführen. Ein Beispiel hierfür ist die Verallgemeinerung von Lemma 2.8:

Lemma 2.27 *Sei $S \subseteq \mathbb{R}^n$ konvex und es seien $x^1 \in cl\ S$, $x^2 \in ri\ S$.*
Dann ist $y = \lambda x^1 + (1-\lambda)x^2 \in ri\ S$ für jedes $\lambda \in]0,1[$.

Beweis: Nach den vorangegangenen Bemerkungen beschränken wir uns auf die Betrachtung des $\mathbb{R}^m$, wenn die Menge S m-dimensional ist. In diesem Raum fallen aber $int\ S$ und $ri\ S$ zusammen und Lemma 2.8 liefert das Ergebnis. □

Der nächste Satz zeigt eine wichtige Eigenschaft der Operationen cl und ri auf dem System aller konvexen Mengen, welche für ri eine interessante Konsequenz hat.

Satz 2.28 *Sei $C \subseteq \mathbb{R}^n$ konvex. Dann sind $cl\ C$ und $ri\ C$ konvex und haben die selbe affine Hülle und folglich auch die selbe Dimension wie C. Insbesondere ist $ri\ C \neq \emptyset$, falls $C \neq \emptyset$.*

Beweis: $cl\ C$ ist nach Lemma 2.7 konvex. $aff\ C$ ist abgeschlossen, somit gilt $cl\ C \subseteq aff\ C$. Offensichtlich ist $aff\ C \subseteq aff\ cl\ C$, was mit der obigen Inklusion die Gleichheit der affinen Hüllen liefert.
Die Konvexität von $ri\ C$ folgt aus Lemma 2.27 . Um den Beweis zu vervollständigen, reicht es zu zeigen, daß eine volldimensionale konvexe Menge im $\mathbb{R}^n$ mit $n > 0$ ein nichtleeres Inneres hat (vgl. Bemerkungen vor Lemma 2.27). Eine n-dimensionale konvexe Menge enthält nach Lemma 2.25 ein n-dimensionales Simplex. Es reicht also zu zeigen, daß ein solches Simplex S ein nichtleeres Inneres hat.
Im Falle $n = 0$ ist die Behauptung trivial, da dann $C = \{x^0\} = ri\ C = cl\ C = aff\ C$. Sei also $n > 0$. Mittels einer bijektiven affinen Transformation können die Ecken von S in die Punkte $(0,0,\dots,0),\ (1,0,\dots,0),\dots,\ (0,\dots,0,1)$ überführt werden. Das so entstehende Simplex hat die Beschreibung

$$\tilde{S} = \{x \in \mathbb{R}^n \ : \ x_i \geq 0,\ \sum_{j=1}^{n} x_j \leq 1\}$$

und man überzeugt sich mit Hilfe von Satz 2.11 leicht, daß eine Kugel um den Punkt $\bar{x} = \frac{1}{n+1}\mathbb{1}$ mit Radius $\varepsilon = \frac{1}{2\sqrt{n}(n+1)}$ in $\tilde{S}$ enthalten ist und somit ist $int\ C$ nichtleer.

□

Für den Begriff der relativen Inneren lassen sich nun Eigenschaften analog zu Folgerung 2.10 *i)* und *ii)* ohne weitere Voraussetzungen beweisen.

Folgerung 2.29 *Es sei S konvex. Dann gilt*

i) $cl(ri\ S) = cl\ S$

ii) $ri(cl\ S) = ri\ S$

Beweis: Für $S = \emptyset$ ist die Aussage klar, da dann alle auftretenden Mengen leer sind. Sei also $S \neq \emptyset$, woraus nach Satz 2.28 folgt $ri\ S \neq \emptyset$.

i) Analog zum Beweis von Folgerung 2.10, wobei Lemma 2.27 an die Stelle von Lemma 2.8 tritt.

ii) Wegen $S \subseteq cl\ S$ und da nach Satz 2.28 $aff\ S = aff(cl\ S)$, gilt $ri\ S \subseteq ri(cl\ S)$. Der weitere Beweis erfolgt ebenfalls genau wie in Folgerung 2.10. □

Hieraus ergibt sich sofort

Folgerung 2.30 *Seien C_1 und C_2 konvexe Teilmengen des $\mathbb{R}^n$, so gilt*
$$cl\ C_1 = cl\ C_2 \iff ri\ C_1 = ri\ C_2 \iff ri\ C_1 \subseteq C_2 \subseteq cl\ C_1$$

Nun beweisen wir noch eine Charakterisierung relativ innerer Punkte: ein Punkt z gehört zum relativen Inneren von C genau dann, wenn sich jede Strecke in C über z hinaus verlängern läßt:

Satz 2.31 *Sei $\emptyset \neq C \subseteq \mathbb{R}^n$ konvex. Dann gilt $z \in ri\ C$ genau dann, wenn für jedes $x \in C$ ein $\mu > 1$ existiert, so daß $(1-\mu)x + \mu z \in C$.*

Beweis: "$\Longrightarrow$" trivial
"$\Longleftarrow$" z genüge der Bedingung. Nach Satz 2.28 existiert ein $x \in ri\ C$. Sein nun $y = (1-\mu)x+\mu z \in C$ für ein $\mu > 1$ der nach Voraussetzung existierende Punkt. Dann ist aber $z = \lambda y + (1-\lambda)x$ mit $0 < \lambda = \frac{1}{\mu} < 1$ und somit nach Lemma 2.27 $z \in ri\ C$. □

Bemerkung: Aus dem Beweis ersieht man: es genügt, die Existenz dieser Verlängerung für ein $x \in ri\ C$ zu fordern, d.h. es gilt:
Sei $\emptyset \neq C \subseteq \mathbb{R}^n$ konvex, $z \in C$. Existieren dann ein $x \in ri\ C$ und ein $\mu > 1$, so daß $(1-\mu)x+\mu z \in C$, so gilt $z \in ri\ C$.

Satz 2.32 *Sei $C \subseteq \mathbb{R}^n$ konvex und mit A $m \times n$-Matrix, $b \in \mathbb{R}^m$ $f : x \in \mathbb{R}^n \longrightarrow Ax + b \in \mathbb{R}^m$ eine affine Transformation. Dann gilt $f(cl\ C) \subseteq cl(f(C))$ und $ri(f(C)) = f(ri\ C)$.*

Beweis: Die Behauptung für die Abschließung beruht nur darauf, daß affine Abbildungen in endlich-dimensionalen Räumen stetig sind, hierfür wird die Konvexität von C nicht benötigt.
Um die Beziehung der relativen Inneren zu zeigen, überlegen wir uns zunächst, daß
$$cl\ f(ri\ C) \supseteq f(cl(ri\ C)) = f(cl\ C) \supseteq f(C) \supseteq f(ri\ C)$$
Daraus folgt, daß die Menge $f(C)$ die selbe Abschließung hat wie $f(ri\ C)$, woraus sich mit Folgerung 2.30 ergibt, daß ebenfalls das relative Innere beider Mengen übereinstimmt. Also gilt $ri(f(C)) \subseteq f(ri\ C)$. Sei nun $z \in f(ri\ C)$. Mit Hilfe der Charakterisierung von Satz 2.31 wollen wir nun zeigen, daß $z \in ri(f(C))$.
Sei $x \in f(C)$ beliebig. Wählen wir nun $z' \in ri\ C$ und $x' \in C$, so daß $z = f(z')$, $x = f(x')$, so existiert nach Satz 2.31 ein $\mu > 1$, so daß $y = (1-\mu)x' + \mu z' \in C$. Es ist aber damit wegen der Form der Transformation $f(y) = Ay + b = (1-\mu)(Ax' + b) + \mu(Az' + b) = (1-\mu)x + \mu z \in f(C)$ und somit wiederum nach Satz 2.31 $z \in ri(f(C))$. □

Folgerung 2.33

i) Seien C eine konvexe Teilmenge des $\mathbb{R}^n$ und $\lambda \in \mathbb{R}$ beliebig, so gilt
$ri(\lambda C) = \lambda(ri\ C)$.

ii) Seien C_1 und C_2 konvexe Teilmengen des $\mathbb{R}^n$, so gilt
$ri(C_1 \oplus C_2) = ri\ C_1 \oplus ri\ C_2$.

Beweis:
i) Man wählt $A :\ x \longrightarrow \lambda x$.
ii) Es gilt (**Beweis ÜA**)
$$ri(C_1 \times C_2) = (ri\ C_1) \times (ri\ C_1) \subseteq \mathbb{R}^{2n}$$
Nun betrachte man die lineare Abbildung $A :\ (x_1, x_2) \in \mathbb{R}^{2n} \longrightarrow x_1 + x_2 \in \mathbb{R}^n$ und Satz 2.32 liefert die Behauptung. □

Um im allgemeinen Trennungssatz die Notwendigkeit der Bedingung für die echte Trennbarkeit zeigen zu können, brauchen wir noch zwei weitere Hilfssätze über Eigenschaften des relativen Inneren.

Lemma 2.34 *Seien C_i konvexe Mengen im $\mathbb{R}^n$ für $i \in I$ und es sei $\bigcap_{i\in I} ri\ C_i \neq \emptyset$. Dann gilt*

$$cl\left(\bigcap_{i\in I} C_i\right) = \bigcap_{i\in I} cl\ C_i$$

Ist weiterhin I endliche Indexmenge, so gilt

$$ri\left(\bigcap_{i\in I} C_i\right) = \bigcap_{i\in I} ri\ C_i$$

Beweis: ÜA Hinweis: beweisen Sie zunächst, daß gilt

$$\bigcap_{i\in I} cl\ C_i \subseteq cl \bigcap_{i\in I} ri\ C_i \subseteq cl \bigcap_{i\in I} C_i \subseteq \bigcap_{i\in I} cl\ C_i$$

Sei ein $x \in \bigcap_{i\in I} ri\ C_i$ fest gewählt und $y \in \bigcap_{i\in I} cl\ C_i$ beliebig. Dann liefert Lemma 2.27, daß $(1-\lambda)x + \lambda y \in ri\ C_i \quad \forall \lambda \in]0,1[,\ i \in I$ und y ist der Grenzwert dieses Vektors für $\lambda \uparrow 1$. Es folgt

$$\bigcap_{i\in I} cl\ C_i \subseteq cl \bigcap_{i\in I} ri\ C_i \subseteq cl \bigcap_{i\in I} C_i \subseteq \bigcap_{i\in I} cl\ C_i$$

Dies beweist die Formel für die Abschließung und zeigt zugleich, daß $\bigcap_{i\in I} ri\ C_i$ und $\bigcap_{i\in I} C_i$ die selbe Abschließung haben. Nach Folgerung 2.30 haben sie dann auch das selbe relative Innere. Somit ergibt sich

$$ri \bigcap_{i\in I} C_i \subseteq \bigcap_{i\in I} ri\ C_i$$

Für den Beweis der umgekehrten Inklusion benötigen wir die Endlichkeit von I. Sei $z \in \bigcap_{i\in I} ri\ C_i$. Nach Satz 2.31 existiert für jedes $i \in I$ ein $\mu_i > 1$, so daß $(1-\mu_i)x + \mu_i z \in C_i$ und wegen der Konvexität gilt dies für alle $\mu \in [0, \mu_i]$. Aufgrund der Endlichkeit gilt $1 < \bar{\mu} := \min_{i\in I} \mu_i$ und wegen Konvexität $(1-\bar{\mu})x + \bar{\mu} z \in \bigcap_{i\in I} C_i$, was wiederum nach Satz 2.31 liefert $z \in ri \bigcap_{i\in I} C_i$. □

Unter Zusatzvoraussetzungen gilt auch für das relative Innere Monotonie:

Lemma 2.35 *Es seien C_1, C_2 konvex, $C_2 \subseteq cl\ C_1$ und $C_2 \setminus rbd\ C_1 \neq \emptyset$. Dann gilt $ri\ C_2 \subseteq ri\ C_1$.*

Beweis: ÜA

Jetzt können wir einen Trennungssatz beweisen, welcher eine schwächere hinreichende Voraussetzung als Satz 2.21 für die echte Trennbarkeit zweier konvexer Mengen angibt, die gleichzeitig auch notwendig ist.
Um diesen allgemeinen Satz zu beweisen, verwenden wir eine Verallgemeinerung von Satz 2.14.

Satz 2.36 *Es seien C eine nichtleere konvexe Teilmenge des $\mathbb{R}^n$ und $y \notin ri\ C$. Dann existiert eine echte trennende Hyperebene für C und $\{y\}$.*

Beweis:
Fall a) $y \notin cl\ C$, Satz 2.14 liefert die Behauptung sogar für starke Trennung von $cl\ C$ und $\{y\}$.
Fall b) $y \in rbd\ C$. Die Dimension von C sei $m \leq n$. Wir betrachten gemäß den Bemerkungen vor Lemma 2.27 den durch bijektive affine Transformation in

$$L = \{x \in \mathbb{R}^n \ :\ x_i = 0,\ i = m+1, \ldots n\}$$

überführten affinen Unterraum $aff\ C$ als zugrundeliegenden Raum. In diesem Exemplar des $\mathbb{R}^m$ liefert Satz 2.17 die Existenz einer Stützhyperebene an C in y, d.h. die Existenz eines $0 \neq a \in \mathbb{R}^m$

mit $y \in H = \{x \in \mathbb{R}^m \ : \ a^T(x-y) = 0\}$ und $C \subseteq H^-$. Da die Menge C bezüglich dieses Raumes volle Dimension hat, kann sie nicht in der Hyperebene enthalten sein, diese ist also eine echte Stützhyperebene. Ergänzen wir den Vektor a zu einem n-dimensionalen Vektor mit

$$\tilde{a}_j = a_j, \quad j = 1, \ldots, m, \quad \tilde{a}_j = 0, \quad j = m+1, \ldots, n,$$

so definiert dieser eine echte Stützhyperebene bezüglich des gesamten (transformierten) $\mathbb{R}^n$, das Urbild ist also die gesuchte Stützhyperebene (denn die jetzt $n-1$-dimensionale Ebene wurde nur um Richtungen orthogonal zu $aff\ C$ ergänzt). □

Satz 2.37 (allgemeiner Trennungssatz)
Es seien C_1 und C_2 nichtleere konvexe Teilmengen des $\mathbb{R}^n$. Es existiert eine echte trennende Hyperebene für C_1 und C_2 genau dann, wenn $ri\ C_1 \cap ri\ C_2 = \emptyset$.

Beweis: Genau wie für den Satz 2.21 betrachten wir die konvexe Menge $C = C_1 \ominus C_2$. Aufgrund der schwächeren Voraussetzung kann man nicht schließen, daß diese Differenz den Ursprung nicht enthält, aber aufgrund von Folgerung 2.33 gilt $ri\ C = ri\ C_1 \ominus ri\ C_2$, woraus folgt:

$$0 \notin ri\ C \quad \Longleftrightarrow \quad ri\ C_1 \cap ri\ C_2 = \emptyset$$

Satz 2.36 liefert die Existenz einer echten trennenden Hyperebene für $\{0\}$ und C, d.h. es existiert ein Vektor a, so daß

$$0 \leq \inf_{x \in C} a^T x = \inf_{x^1 \in C_1} a^T x^1 - \sup_{x^2 \in C_2} a^T x^2$$

und C ist nicht in dieser Hyperebene enthalten, d.h.

$$0 < \sup_{x \in C} a^T x = \sup_{x^1 \in C_1} a^T x^1 - \inf_{x^2 \in C_2} a^T x^2$$

Dies bedeutet gerade, daß C_1 und C_2 durch eine Hyperebene mit Normalenvektor a echt trennbar sind.
Umgekehrt folgen aus der Existenz einer echten trennenden Hyperebene für C_1 und C_2 die obigen Bedingungen und damit die Existenz einer Hyperebene, so daß $C \subseteq H^+$ und das (relative) Innere des Halbraums $ri\ H^+ = \left\{x \in \mathbb{R}^n \ : \ a^T x > 0\right\}$ mit $ri\ C$ einen nichtleeren Durchschnitt hat. (Begründung: es existiert ein Punkt aus C im rel. Inneren des Halbraums, betrachten Verbindungsstrecke mit Punkt aus dem relativen Inneren von C. Nach Lemma 2.27 ist das relative Innere dieser Strecke Teilmenge von $ri\ C$ und dieses hat mit dem offenen Halbraum nichtleeren Durchschnitt.) Lemma 2.35 liefert dann $ri\ C \subseteq ri\ H^+$ und somit $0 \notin ri\ C$. □

2.3 Konvexe Funktionen

Die übliche Definition einer konvexen Funktion, welche Sie wahrscheinlich in der Analysis-Vorlesung kennengelernt haben, lautet (für Funktionen vom $\mathbb{R}^n$ in $\mathbb{R}$):

Es sei $D \subseteq \mathbb{R}^n$ konvex (der Definitionsbereich von f). Die Funktion $f: D \longrightarrow \mathbb{R}$ heißt konvex gdw. für jedes Paar von Punkten aus D der Graph der Funktion entlang der Verbindungsstrecke unterhalb der Sekante verläuft. f heißt konkav, wenn $-f$ konvex ist.

Diese Definition führt durch die Angabe des Definitionsbereiches zu zusätzlichem technischem Aufwand bei der Betrachtung von Operationen zwischen konvexen Funktionen mit verschiedenem Definitionsbereich, außerdem lassen sich Funktionen, welche als Infimum oder Supremum definiert sind (und damit auch die Werte $+\infty$ oder $-\infty$ annehmen könnten) nicht ohne Zusatzbetrachtung hinsichtlich des Bereiches endlicher Funktionswerte mit dieser Definition behandeln.
Diese technischen Schwierigkeiten umgeht man, indem man dem Buch *"Convex Analysis"* von *R.T. Rockafellar* folgend, den Begriff zunächst erweitert und mit einem erweiterten Wertebereich Funktionen betrachtet, deren Definitionsbereich stets der gesamte $\mathbb{R}^n$ ist. Die obige Definition wird sich dann als Folgerung ergeben.

2.3.1 Konvexitätsbegriffe

Für die oben erwähnte erweiterte Konvexitätsdefinition erweitern wir die rellen Zahlen:

Bezeichnung 2.38 $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$

In diesem erweiterten Zahlenbereich müssen wir noch die Rechenregeln für Operationen mit den hinzugefügten Elementen festlegen:

$$\begin{array}{ll}
\alpha + \infty = \infty + \alpha = \infty & \text{für } -\infty < \alpha \leq \infty \\
\alpha - \infty = -\infty + \alpha = -\infty & \text{für } -\infty \leq \alpha < \infty \\
\alpha\infty = \infty\alpha = \infty, \;\; \alpha(-\infty) = (-\infty)\alpha = -\infty & \text{für } 0 < \alpha \leq \infty \\
\alpha\infty = \infty\alpha = -\infty, \;\; \alpha(-\infty) = (-\infty)\alpha = \infty & \text{für } -\infty \leq \alpha < 0 \\
0\infty = \infty 0 = 0(-\infty) = (-\infty)0 = 0, \;\; -(-\infty) = \infty &
\end{array}$$

Außerdem wird wie üblich das Infimum über die leeren Menge als ∞ und das Supremum über die leere Menge als $-\infty$ festgelegt.
In den Operationen bleiben die Kombinationen $\infty - \infty$ und $-\infty + \infty$ undefiniert. Die Kommutativ-, Assoziativ- und Distributivgesetze für Multiplikation und Addition bleiben gültig, sofern die beiden verbotenen Kombinationen nicht vorkommen.
Indem wir Funktionen außerhalb ihres Definitionsbereiches auf $+\infty$ festlegen, können wir in weiteren immer erweitert-reellwertige Funktionen betrachten, welche auf dem gesamten $\mathbb{R}^n$ definiert sind. Wir werden sehen, daß diese Erweiterung des Definitionsbereiches an Konvexitätseigenschaften nichts ändert.

Definition 2.39 *Sei* $f: \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$.

i) Die Menge $epi\, f := \left\{ \begin{pmatrix} x \\ \mu \end{pmatrix} \in R^{n+1} : \mu \geq f(x) \right\}$ *heißt Epigraph von f.*

ii) Die Menge $hyp\, f := \left\{ \begin{pmatrix} x \\ \mu \end{pmatrix} \in R^{n+1} : \mu \leq f(x) \right\}$ *heißt Hypograph von f.*

iii) Die Menge $dom\, f := \left\{ x \in R^n : \exists \mu \in \mathbb{R}, \begin{pmatrix} x \\ \mu \end{pmatrix} \in epi\, f \right\}$ *heißt effektiver Definitionsbereich von f.*

iv) Die Funktion f heißt konvex, wenn $epi\, f$ konvex ist.

v) Die Funktion f heißt konkav, wenn $hyp\, f$ konvex ist.

Man sieht leicht die Gültigkeit der folgenden Beziehungen:

- $epi\ f = \emptyset \iff dom\ f = \emptyset$
- $dom\ f$ ist die orthogonale Projektion von $epi\ f$ auf den $\mathbb{R}^n$.
- $dom\ f$ ist konvex für eine konvexe Funktion f.
- $\forall x \in dom\ f: \ -\infty \leq f(x) < +\infty$
- f ist konkav genau dann, wenn $-f$ konvex ist.

Um dem Problem der undefinierten Operation $\infty - \infty$ aus dem Wege zu gehen, wird oft vorausgesetzt werden, daß die Funktion der nachfolgenden Definition genügt.

Definition 2.40 *Sei $f: \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$. Die Funktion f heißt eigentliche (proper) Funktion, wenn ihr Epigraph nichtleer ist und keine vertikale Gerade enthält, d.h. wenn*

a) $dom\ f = \{x \in \mathbb{R}^n : \ f(x) < \infty\} \neq \emptyset$

b) $\{x \in \mathbb{R}^n : \ f(x) = -\infty\} = \emptyset$

Folgerung 2.41 *Jede eigentliche konvexe Funktion $f: \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ läßt sich in der Form schreiben:*

$$f(x) = \begin{cases} g(x) & x \in C \\ +\infty & x \in \mathbb{R}^n \setminus C \end{cases}$$

wobei C eine konvexe Menge und $g: C \longrightarrow \mathbb{R}$ eine konvexe Funktion ist.

Satz 2.42 *Sei $f: \mathbb{R}^n \longrightarrow]-\infty, \infty]$. f ist konvex genau dann, wenn*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \quad \forall \lambda \in [0,1],\ \forall x, y \in \mathbb{R}^n$$

Satz 2.43 *Sei $f: \mathbb{R}^n \longrightarrow [-\infty, \infty]$. f ist konvex genau dann, wenn*

$$f((1-\lambda)x + \lambda y) < (1-\lambda)\alpha + \lambda\beta \quad \forall \lambda \in [0,1],\ \forall x, y \in \mathbb{R}^n,\ \forall \alpha, \beta \in \mathbb{R}: \ \alpha > f(x),\ \beta > f(y)$$

Satz 2.44 (Jensen's Ungleichung)
Sei $f: \mathbb{R}^n \longrightarrow]-\infty, \infty]$. f ist konvex genau dann, wenn

$$f(\sum_{i=1}^{m} \lambda_i x^i) \leq \sum_{i=1}^{m} \lambda_i f(x^i) \quad \forall m \in \mathbb{N},\ \forall x^i \in \mathbb{R}^n,\ \forall \lambda_i \in [0,1],\ \ i = 1, \ldots, m \ \ mit \ \ \sum_{i=1}^{m} \lambda_i = 1$$

Die Beweise dieser Beziehungen werden als Übung überlassen.

Für differenzierbare Funktionen können weitere Kriterien für Konvexität angegeben werden. Hierfür müssen wir uns natürlich wieder auf reellwertige Funktionen und somit die Angabe ihrer Definitionsbereiche zurückziehen.
(Zur Erinnerung: der Gradient $\nabla f(x)$ war ein Zeilenvektor der Länge n, so daß das Produkt im folgenden Satz im Sinne des Produkts von Matrizen gebildet werden kann.)

Satz 2.45
Sei $f: C \longrightarrow \mathbb{R}$ differenzierbar auf der offenen und konvexen Menge C.
f ist konvex auf C genau dann, wenn für alle $x, y \in C$ gilt $f(y) - f(x) \geq \nabla f(x)(y - x)$.

Beweis:
"$\Longleftarrow$" Seien $\lambda \in [0,1]$, $x, y \in C$ beliebig, also $z = \lambda x + (1-\lambda)y \in C$.
Die Voraussetzung liefert

$$\begin{array}{rcrcl} \lambda f(x) & - & \lambda f(z) & \geq & \lambda \nabla f(z)(x - z) \\ (1-\lambda)f(y) & - & (1-\lambda)f(z) & \geq & (1-\lambda)\nabla f(z)(y - z) \quad \text{, Addition führt zu:} \\ \lambda f(x) + (1-\lambda)f(y) & - & f(z) & \geq & \nabla f(z)\underbrace{(\lambda x + (1-\lambda)y - z)}_{=0} = 0 \end{array}$$

und Satz 2.42 liefert somit Konvexität von f.
"$\Longrightarrow$" Über $[0,1]$ definieren wir die reelle Funktion F:
$F(\lambda) = \lambda f(y) + (1-\lambda)f(x) - f(\lambda y + (1-\lambda)x)$ für beliebige feste $x, y \in C$.
Da C offen ist, existiert ein $\varepsilon > 0$, so daß diese Funktion auch auf dem offenen Intervall $]-\varepsilon, 1+\varepsilon[$ erklärt und differenzierbar ist. Es ist dann

$$F'(\lambda) = f(y) - f(x) - \nabla f\,(\lambda y + (1-\lambda)x)\,(y-x)$$

$x = y \implies F'(\lambda) \equiv 0$
$x \neq y \implies F(0) = 0$ und $F(\lambda) \geq 0\ \forall \lambda \in [0,1]$ wegen Konvexität von f. $\implies F'(0) \geq 0$.
Somit $0 \leq F'(0) = f(y) - f(x) - \nabla f\,(x)\,(y-x)$. □

Satz 2.46
Sei $f : C \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar auf der offenen und konvexen Menge C.
f ist konvex auf C genau dann, wenn die Hessematrix $\nabla^2 f(x)$ positiv semidefinit ist für alle $x \in C$.

Beweis: Mittels der Taylorformel (Satz 5.47 Analysis I):
Es existiert ein $\theta \in]0,1[$ mit $f(y) = f(x) + \nabla f(x)(y-x) + \frac{1}{2}(y-x)^T \left[\nabla^2 f(x+\theta(y-x))\right](y-x)$.
Ist also die Hessematrix an dem Zwischenpunkt $(1-\theta)x + \theta y$ positiv semidefinit, so folgt $f(y) \geq f(x) + \nabla f(x)(y-x)$ für alle $x, y \in C$, also mit Satz 2.45 f konvex.
Umgekehrt betrachten wir einen beliebigen Punkt $z \in C$. Da C offen ist, existiert zu jedem $x \in \mathbb{R}^n$ ein $\bar{\lambda} > 0$, so daß $z + \lambda x \in C\ \forall \lambda \in [0, \bar{\lambda}[$. Nun verwenden wir die Formulierung der Taylorformel mit Restglied zweiter Ordnung:

$$f(z+\lambda x) = f(z) + \lambda \nabla f(z) x + \frac{1}{2}\lambda^2 x^T \left[\nabla^2 f(z)\right] x + \lambda^2 \|x\|^2 \rho(\lambda x)$$

mit $\lim\limits_{\|\lambda x\| \to 0} \rho(\lambda x) = 0$. Wegen Satz 2.45 folgt

$$\frac{1}{2}\lambda^2 x^T \left[\nabla^2 f(z)\right] x + \lambda^2 \|x\|^2 \rho(\lambda x) \geq 0$$

Dividieren wir diese Ungleichung durch λ^2 und lassen λ gegen Null streben, erhalten wir $x^T \nabla^2 f(z) x \geq 0$. □

Ein verschärfter Konvexitätsbegriff ist wichtig für Anwendungen in der Optimierung, dieser läßt sich nur für reellwertige Funktionen definieren (und somit mit Angabe eines Definitionsbereiches):

Definition 2.47 *Sei C eine konvexe Menge, $f : C \longrightarrow \mathbb{R}$. Die Funktion f heißt streng konvex (strictly convex) auf C, wenn für alle $x, y \in C$ mit $x \neq y$ gilt*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \quad \forall \lambda \in]0,1[$$

Kriterien für strenge Konvexität differenzierbarer Funktionen analog zu den Sätzen 2.45 und 2.46 lassen sich auf die gleiche Weise beweisen:

Satz 2.48
Sei $f : C \longrightarrow \mathbb{R}$ differenzierbar auf der offenen und konvexen Menge C.
f ist streng konvex auf C genau dann, wenn für alle $x, y \in C$ mit $x \neq y$ gilt
$f(y) - f(x) > \nabla f(x)(y-x)$.

Beweis: "$\Longleftarrow$" Analog zum Beweis von Satz 2.45:
Seien $\lambda \in]0,1[$, $x, y \in C$ beliebig, also $z = \lambda x + (1-\lambda)y \in C$, dann liefert die Voraussetzung:

$$\begin{array}{rcrcl}
\lambda f(x) & - & \lambda f(z) & > & \lambda \nabla f(z)(x-z) \\
(1-\lambda) f(y) & - & (1-\lambda) f(z) & > & (1-\lambda)\nabla f(z)(y-z) \quad \text{, Addition führt zu:} \\
\lambda f(x) + (1-\lambda) f(y) & - & f(z) & > & \nabla f(z)\underbrace{(\lambda x + (1-\lambda)y - z)}_{=0} = 0
\end{array}$$

dies ist die Definition der strengen Konvexität von f.

"$\Longrightarrow$" Angenommen, es exitieren Punkte $x \neq y$ mit $f(y)-f(x) \leq \nabla f(x)(y-x)$. Aus der strengen Konvexität von f folgt die Konvexität, also gilt nach Satz 2.45 in der Ungleichung "$\geq$", somit "$=$", d.h.

$$f(y) = f(x) + \nabla f(x)(y-x)$$

Wir betrachten $z = \frac{1}{2}(x+y)$ den Mittelpunkt der Strecke und setzen die gewonnene Gleichung in die Ungleichung aus der Definition der strengen Konvexität ein:

$$\begin{array}{rcl} f(z) & < & \frac{1}{2}(f(x)+f(y)) \\ & = & \frac{1}{2}(f(x)+f(x)+\nabla f(x)(y-x)) \\ & = & f(x)+\nabla f(x)(\frac{1}{2}(y-x)) \\ & = & f(x)+\nabla f(x)(\frac{1}{2}(y+x)-x) \\ & = & f(x)+\nabla f(x)(z-x) \end{array}$$

Dies steht nach Satz 2.45 im Widerspruch zur Konvexität von f. □

Satz 2.49
Sei $f : C \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar auf der offenen und konvexen Menge C. Ist die Hessematrix $\nabla^2 f(x)$ positiv definit für alle $x \in C$, so ist f streng konvex auf C.

Beweis: Analog zum Beweis von Satz 2.46 liefert hier die positive Definitheit mit dem Satz von Taylor für alle $x \neq y$:

$$\begin{array}{rcl} f(y) & = & f(x)+\nabla f(x)(y-x)+\underbrace{\frac{1}{2}(y-x)^T\left[\nabla^2 f(x+\theta(y-x))\right](y-x)}_{>0} \\ & > & f(x)+\nabla f(x)(y-x) \quad \forall x \neq y \end{array}$$

und Satz 2.48 liefert die Behauptung □

Der umgekehrte Schluss wie im Beweis von Satz 2.48 mittels Grenzwert liefert keine strenge Ungleichung, also ließe sich so nur positive Semidefinitheit der Hessematrix schließen. Dies liegt nicht nur an der Beweistechnik. Die Umkehrung gilt hier nicht, wie das eindimensionale Beispiel $f(x) = x^4$ zeigt. Diese Funktion ist streng konvex, ihre zweite Ableitung im Nullpunkt jedoch gleich Null. Für quadratische Funktionen ($f(x) = x^T Cx + p^T x + q$) gilt auch die Umkehrung.

Eine weitere Eigenschaft konvexer Funktionen ist wichtig für unsere Zwecke in der Optimierung:

Lemma 2.50
Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ konvex. Dann ist für jedes $\alpha \in \mathbb{R}$ die Niveaumenge (level set oder lower level set)

$$L_f(\alpha) := \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$$

konvex.
Beweis: ÜA

Die Umkehrung des Lemmas gilt nicht. Für manche Zwecke reicht jedoch diese schwächere Eigenschaft aus und dies führt zur Formulierung eines weiteren (schwächeren) Konvexitätsbegriffes.

Definition 2.51 *Sei C eine konvexe Menge, $f : C \longrightarrow \mathbb{R}$. Die Funktion f heißt quasikonvex auf C, wenn für alle $x, y \in C$ gilt*

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\} \quad \forall \lambda \in]0,1[$$

f heißt quasikonkav, wenn $-f$ quasikonvex ist.

Die analytisch formulierte Definition der Quasikonvexität ist äquivalent geometrischen charakterisierbar:

Satz 2.52
Sei C eine konvexe Menge, $f: C \longrightarrow \mathbb{R}$.
f ist quasikonvex auf C genau dann, wenn für alle $\alpha \in \mathbb{R}$ die Niveaumenge $L_f(\alpha) = \{x \in C : f(x) \leq \alpha\}$ konvex ist.

Beweis: ÜA
"$\Longrightarrow$" Seien $x, y \in Ł_f(\alpha)$ und $\lambda \in [0,1]$ beliebig. Dann gilt wegen der Konvexität von C $z = \lambda x + (1-\lambda)y \in C$ und wegen Quasikonvexität von f ist $f(z) \leq \max\{f(x), f(y)\} \leq \alpha$, also $z \in L_f(\alpha)$.
"$\Longleftarrow$" Seien $x, y \in C$ und $\lambda \in [0,1]$ beliebig sowie z wie oben. Es gilt wieder $z \in C$. Wir wählen $\alpha = \max\{f(x), f(y)\}$. Es gilt dann $x, y \in L_f(\alpha)$ und wegen der Konvexität von $L_f(\alpha)$ auch $z \in L_f(\alpha)$, also $f(z) \leq \alpha = \max\{f(x), f(y)\}$, was gerade die Quasikonvexität von f liefert. □

Auch für diesen Begriff existiert eine notwendige und hinreichende Charakterisierung im Falle der Differenzierbarkeit:

Satz 2.53
Sei $f: C \longrightarrow \mathbb{R}$ differenzierbar auf der offenen und konvexen Menge C.
f ist quasikonvex auf C genau dann, wenn für alle $x, y \in C$ gilt:
Wenn $f(y) \leq f(x)$, so $\nabla f(x)(y-x) \leq 0$.

Beweis: siehe *Bazaraa, Sherali, Shetty "Nonlinear Programming"*, Lemma 3.5.4

Zwei weitere Begriffe werden aus dem der Quasikonvexität abgeleitet:

Definition 2.54 *Sei C eine konvexe Menge, $f: C \longrightarrow \mathbb{R}$.*
Die Funktion f heißt streng quasikonvex (strictly quasiconvex) auf C, wenn für alle $x, y \in C$ mit $f(x) \neq f(y)$ gilt

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\} \quad \forall \lambda \in]0,1[$$

Die Funktion f heißt stark quasikonvex (strongly quasiconvex) auf C, wenn für alle $x, y \in C$ mit $x \neq y$ gilt

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\} \quad \forall \lambda \in]0,1[$$

Während für streng konvexe Funktionen stets Konvexität gilt, überträgt sich dies nicht auf das Verhältnis von streng quasikonvex und quasikonvex, wie das folgende Beispiel zeigt:

$$f: \mathbb{R} \longrightarrow \mathbb{R}, \quad f(x) = \begin{cases} 1 & \text{für } x = 0 \\ 0 & \text{sonst} \end{cases}$$

Diese Funktion ist streng quasikonvex, jedoch nicht quasikonvex (z.B. $x = -1,\ y = 1,\ \lambda = \frac{1}{2}$).
Fordert man zusätzlich Unterhalbstetigkeit von f, so folgt auch aus strenger Quasikonvexität die Quasikonvexität von f (*Bazaraa, Sherali, Shetty*, Lemma 3.5.7).

Für Quasikonvexität gelten die Beziehungen:

f konvex	$\Longrightarrow$	f streng quasikonvex
f streng konvex	$\Longrightarrow$	f stark quasikonvex
f stark quasikonvex	$\Longrightarrow$	f streng quasikonvex
f stark quasikonvex	$\Longrightarrow$	f quasikonvex

Diese Eigenschaften haben nützliche Folgerungen für Optimierungsprobleme:

Satz 2.55
Sei $f: C \longrightarrow \mathbb{R}$ streng quasikonvex auf der konvexen nichtleeren Menge C.
Ist x ein lokales Minimum von f auf C, so ist es auch globales Minimum.

Beweis: Nehmen wir im Widerspruch an, es gäbe ein $y \in C$ mit $f(y) < f(x)$. Die strenge Quasikonvexität liefert
$f(\lambda y + (1-\lambda)x) < \max\{f(x), f(y)\} = f(x) \quad \forall \lambda \in]0,1[$.
Wegen der Konvexität von C gilt $\lambda y + (1-\lambda)x \in C \quad \forall \lambda \in [0,1]$ und es existiert ein $\delta \in]0,1]$, so daß diese Punkte für $\lambda \in [0, \delta[$ in der Umgebung liegen, in welcher x Minimalpunkt ist. Dies steht im Widerspruch zur obigen Ungleichung. □

Satz 2.56
Sei $f: C \longrightarrow \mathbb{R}$ stark quasikonvex auf der konvexen nichtleeren Menge C.
Ist x ein lokales Minimum von f auf C, so ist es das einzige globales Minimum.

Beweis: Nehmen wir im Widerspruch zur Behauptung an, es gäbe ein $y \in C$ mit $y \neq x$ und $f(y) \leq f(x)$. Wegen starker Quasikonvexität gilt dann

$$f(\lambda y + (1-\lambda)x) < \max\{f(x), f(y)\} = f(x) \quad \forall \lambda \in]0,1[$$

jedoch die Punkte auf dieser Strecke liegen für hinreichend kleine λ in der Umgebung um x, in welcher dieses Minimalpunkt ist. Dies liefert einen Widerspruch. □

Trotz dieser positiven Eigenschaften haben die Begriffe der strengen bzw. starken Quasikonvexität den Mangel, daß $\nabla f(x) = 0$ keine hinreichende Bedingung dafür ist, daß x Minimalpunkt ist. Als Beispiel dafür sei $f(x) = x^3$ betrachtet. Diese Funktion ist streng monoton und damit ist starke Quasikonvexität erfüllt, im Punkt $x = 0$ verschwindet die Ableitung, es liegt jedoch kein Minimum vor, sondern ein Wendepunkt.
Dies gibt Anlaß zu einem weiteren Konvexitätsbegriff für differenzierbare Funktionen:

Definition 2.57 *Sei C eine offene konvexe Menge, $f: C \longrightarrow \mathbb{R}$ sei differenzierbar auf C.*
Die Funktion f heißt pseudokonvex auf C, wenn für alle $x, y \in C$ gilt:
$f(x) > f(y) \implies \nabla f(x)(y-x) < 0$.
Die Funktion f heißt streng pseudokonvex auf C, wenn für alle $x, y \in C$ mit $x \neq y$ gilt:
$f(x) \geq f(y) \implies \nabla f(x)(y-x) < 0$.

Man sieht sofort, daß mit Pseudokonvexität das Verschwinden des Gradienten hinreichende Bedingung für das Vorliegen sogar eines globalen Minimums ist, bei strenger Pseudokonvexität sogar für das Vorliegen eines eindeutigen globalen Minimums.
Nun soll noch eine Verschärfung des Begriffes der strengen Konvexität definiert werden, auf die man in manchen Sätzen über Optimierungsprobleme stößt. Auch diese ist nur für reellwertige Funktionen definiert und fordert eine "mindestens quadratische" Krümmung der Funktion:

Definition 2.58 *Sei C eine konvexe Menge, $f: C \longrightarrow \mathbb{R}$.*
Die Funktion f heißt stark konvex auf C, wenn ein $\sigma > 0$ existiert, so daß für alle $x, y \in C$ und alle $\lambda \in [0,1]$ gilt:
$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \sigma\lambda(1-\lambda)\|x-y\|^2$.

2.3.2 Stetigkeit konvexer Funktionen

Konvexe Funktionen sind stetig auf dem Inneren ihres effektiven Definitionsbereiches, einen Beweis dafür finden Sie in *Bazaraa, Sherali, Shetty*, Theorem 3.1.3. Da jedoch die Behandlung dieses Themas nach *R.T. Rockafellar* zusätzliche Einsichten liefert, wollen wir diese hier skizzieren.

Wir wiederholen die Definitionen der Halbstetigkeit von unten bzw. oben, hier für erweitert-reellwertige Funktionen:

Definition 2.59 *Sei* $f: \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$.
Die Funktion f heißt unterhalbstetig (lower semicontinuous, kurz: uhs) in x, wenn für alle Folgen (x^n) mit $x^n \underset{n\to\infty}{\longrightarrow} x$ gilt $f(x) \le \liminf\limits_{n\to\infty} f(x^n)$

Die Funktion f heißt oberhalbstetig (upper semicontinuous, kurz: ohs) in x, wenn für alle Folgen (x^n) mit $x^n \underset{n\to\infty}{\longrightarrow} x$ gilt $f(x) \ge \limsup\limits_{n\to\infty} f(x^n)$

Zur Erinnerung: $\liminf\limits_{n\to\infty} a_n$ ist der kleinste Häufungswert der Folge (a_n) reeller Zahlen oder anders ausgedrückt das Infimum über die Grenzwerte aller konvergenten Teilfolgen von (a_n), bzw. ist als $-\infty$ definiert für nach unten unbeschränkte Folgen.
Uns soll zunächst die Unterhalbstetigkeit beschäftigen. Ihre Bedeutung zeigt der nächste Satz:

Satz 2.60
Sei $f: \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$. Die folgenden Bedingungen sind äquivalent:

i) f ist unterhalbstetig auf $\mathbb{R}^n$

ii) $L_f(\alpha)$ ist abgeschlossen für jedes $\alpha \in \mathbb{R}$

iii) epi f ist abgeschlossen

Beweis:
i) $\Longrightarrow$ iii) Sei $\begin{pmatrix} x^n \\ \mu_n \end{pmatrix}$ eine konvergente Folge von Punkten aus *epi* f. Es gilt also $\mu_n \ge f(x^n) \quad \forall n \in \mathbb{N}$ und es ist $x^n \underset{n\to\infty}{\longrightarrow} x$, $\mu_n \underset{n\to\infty}{\longrightarrow} \mu$. Da f unterhalbstetig ist, gilt

$$f(x) \le \liminf_{n\longrightarrow\infty} f(x^n) \le \lim_{n\longrightarrow\infty} \mu_n = \mu \quad \Longrightarrow \quad \begin{pmatrix} x \\ \mu \end{pmatrix} \in epi\ f$$

iii) $\Longrightarrow$ ii) Sei (x^n) eine in $\mathbb{R}^n$ konvergente Folge von Punkten aus $L_f(\alpha)$, ihr Grenzwert sei x. Es ist $(x^n, \alpha)^T \in epi\ f \ \forall n$. Dies ist eine in $\mathbb{R}^{n+1}$ konvergente Folge. Wegen der Abgeschlossenheit von *epi* f liegt ihr Grenzwert $(x, \alpha)^T$ in *epi* f, d.h. $f(x) \le \alpha$, also $x \in L_f(\alpha)$.
ii) $\Longrightarrow$ i) Sei (x^n) eine beliebige in $\mathbb{R}^n$ konvergente Folge, ihr Grenzwert sei x.
Fall a) $f(x^n)$ hat einen Häufungspunkt $\mu \in \mathbb{R} \cup \{-\infty\}$.
Wir betrachten eine beliebige Teilfolge, für welche $f(x^{n_i})$ gegen diesen Häufungspunkt konvergiert (bzw. bestimmt gegen $-\infty$ divergiert). Für jedes $\alpha \in \mathbb{R}$ mit $\alpha > \mu$ existiert ein i_0, so daß gilt $f(x^{n_i}) < \alpha$ für $i \ge i_0$. Es folgt $x \in cl\,(L_f(\alpha)) = L_f(\alpha)$, d.h. $f(x) \le \alpha \quad \forall \alpha > \mu$, woraus wiederum folgt $f(x) \le \mu$.
Fall b) Einziger Häufungspunkt μ von $f(x^n)$ in $\bar{\mathbb{R}}$ ist $+\infty$.
Betrachten wir eine beliebige bestimmt divergente Teilfolge ($\lim\limits_{i\longrightarrow\infty} f(x^{n_i}) = +\infty$), so gilt die Ungleichung $f(x) \le \mu$ für diese Teilfolge trivialerweise.
Da die Ungleichung somit für jeden Häufungswert gilt, folgt $f(x) \le \liminf\limits_{n\longrightarrow\infty} f(x^n)$. □

Definition 2.61 *Sei* $f: \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$.
Dann existiert eine größte unterhalbstetige Funktion g, für welche $g(x) \le f(x) \ \forall x \in \mathbb{R}^n$, nämlich diejenige Funktion g mit epi $g = cl(epi\ f)$. Die Funktion g heißt die unterhalbstetige Hülle von f.
Ist f konvex, so wird als Abschließung von f (Schreibweise: cl f) die uhs Hülle von f bezeichnet, falls f eigentlich ist. Für eine uneigentliche konvexe Funktion f wird als cl f die Funktion definiert, die identisch $-\infty$ ist.
Eine konvexe Funktion f heißt abgeschlossen, wenn $f = cl\ f$.

Bemerkung 2.62 *$f: \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ sei eine eigentliche konvexe Funktion.*

i) Abgeschlossenheit von f ist gleichbedeutend mit Unterhalbstetigkeit.

ii) $epi(cl\ f) = cl(epi\ f)$ (nach Definition)

iii) $(cl\ f)(x) = \liminf\limits_{y\longrightarrow x} f(y)$

iv) $cl(dom\ f) \supseteq dom(cl\ f) \supseteq dom\ f$

Beispiele für echte Teilmengenbeziehungen in *iv)* (**ÜA**):

$$f : \mathbb{R} \longrightarrow \bar{\mathbb{R}} \quad f(x) = \begin{cases} \frac{1}{x} & \text{für } x > 0 \\ +\infty & \text{sonst} \end{cases}$$

$$f : \mathbb{R} \longrightarrow \bar{\mathbb{R}} \quad f(x) = \begin{cases} 0 & \text{für } x > 0 \\ +\infty & \text{sonst} \end{cases}$$

Satz 2.63
Ist $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine uneigentliche konvexe Funktion, so ist* $f(x) = -\infty \quad \forall x \in ri(dom\ f)$.

Beweis: Ist $dom\ f \neq \emptyset$, so existiert ein Punkt $y \in dom\ f$ mit $f(y) = -\infty$, da f uneigentlich. Sei $x \in ri(dom\ f)$ beliebig. Dann existiert nach Satz 2.31 ein $\mu > 1$ mit $z = (1-\mu)y + \mu x \in dom\ f$. Es ist dann $x = (1-\lambda)y + \lambda z$ mit $0 < \lambda = \frac{1}{\mu} < 1$. Nach Satz 2.43 gilt dann

$$f(x) < (1-\lambda)\alpha + \lambda\beta \quad \forall \lambda \in [0,1],\ \forall \alpha, \beta \in \mathbb{R} :\ \alpha > f(y),\ \beta > f(z)$$

Wegen $f(z) < +\infty$ und $f(y) = -\infty$ folgt daraus $f(x) = -\infty$ (man wählt ein festes β und läßt α gegen $-\infty$ streben). □

Folgerung 2.64 *Sei* $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine unterhalbstetige uneigentliche konvexe Funktion. Dann nimmt f keine endlichen Werte an.*

Beweis: Wegen f uhs muß die Menge der Punkte, in denen der Wert $-\infty$ angenommen wird, $cl(ri(dom\ f))$ enthalten und es ist $cl(ri(dom\ f)) = cl(dom\ f) \supseteq dom\ f$. □

Hieraus ergibt sich, daß für eine uneigentliche konvexe Funktion die Abschließung nach der obigen Definition mit der unterhalbstetigen Hülle auf $cl(dom\ f)$ übereinstimmt. Außerhalb ist die uhs Hülle $+\infty$, wogegen $cl\ f$ als $-\infty$ erklärt wurde.

Folgerung 2.65 *Sei* $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine konvexe Funktion, deren effektiver Definitionsbereich relativ offen ist. Dann ist entweder* $f(x) > -\infty \quad \forall x \quad$ *oder* $\quad f(x) \in \{-\infty, +\infty\} \quad \forall x$.

Für die weitere Analyse wichtig ist die folgende topologische Aussage:

Lemma 2.66
Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *konvex. Dann gilt*

$$ri(epi\ f) = \left\{(x,\mu)^T \in \mathbb{R}^{n+1} :\ x \in ri(dom\ f),\ \ f(x) < \mu < \infty\right\}$$

Beweis: Es genügt, sich auf den volldimensionalen Fall zu beschränken (Bem. vor Lemma 2.27). Die Inklusion "$\subseteq$" ist offensichtlich. Zu zeigen bleibt die umgekehrte Richtung.
Seien $\bar{x} \in int(dom\ f)$ und $\bar{\mu}$ eine reelle Zahl mit $\bar{\mu} > f(\bar{x})$. Mit dem Satz von Carathéodory folgt dann die Existenz von $n+1$ Punkten $x^1, \ldots, x^{n+1} \in dom\ f$, so daß $\bar{x}$ im Inneren des von diesen Punkten aufgespannten Simplex S liegt. (*brauchen endlich viele Punkte, um im folgenden ein* $\alpha < +\infty$ *zu haben*) Wir definieren $\alpha = \max\{f(x^i) :\ i = 1, \ldots, n+1\}$ $(\in \mathbb{R} \cup \{-\infty\})$. Für jedes $x \in S$ gibt es eine Darstellung $x = \sum_{i=1}^{n+1} \lambda_i x^i$ mit $\lambda_i \geq 0,\ \sum_{i=1}^{n+1} \lambda_i = 1$.
Wegen Konvexität von f gilt dann für alle $x \in int\ S$: $f(x) \leq \sum_{i=1}^{n+1} \lambda_i f(x^i) \leq \alpha$. Diese Ungleichung gilt nach Satz 2.44, falls f nirgends den Wert $-\infty$ annimmt. Wenn f den Wert $-\infty$ annimmt, so nach Satz 2.63 in jedem Punkt von $ri(dom\ f)$, also auch in x. Somit gilt die Ungleichung auch in diesem Fall. Folglich ist die offene Menge

$$\{(x,\mu)^T \in \mathbb{R}^{n+1} :\ x \in int\ S,\ \ \alpha < \mu < \infty\}$$

Teilmenge von $epi\ f$. Insbesondere gilt für jedes $\mu > \alpha$, daß $(\bar{x},\mu)^T \in int(epi\ f)$. Somit ist $(\bar{x},\bar{\mu})^T$ ein relativ innerer Punkt einer "vertikalen" Strecke in $epi\ f$, welche mit $int(epi\ f)$ einen nichtleeren Durchschnitt hat. Daraus folgt nach Lemma 2.35 $(\bar{x},\bar{\mu})^T \in int(epi\ f)$. □

Folgerung 2.67 *Seien α eine reelle Zahl und $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ eine konvexe Funktion, so daß ein x existiert mit $f(x) < \alpha$. Dann existiert ein $y \in ri(dom\ f)$ mit $f(y) < \alpha$.*

Beweis: Nach Voraussetzung hat der offene Halbraum $\{(x,\mu)^T \in \mathbb{R}^{n+1} : \mu < \alpha\}$ einen nichtleeren Durchschnitt mit $epi\ f$. Die Behauptung folgt nun sofort aus einer weiteren Folgerung aus Folgerung 2.29:

Sind O eine offene Menge und C eine konvexe Menge, so gilt: $O \cap cl\ C \neq \emptyset \implies O \cap ri\ C \neq \emptyset$

(Beweis in der Übung) Dies ist wie folgt zu sehen: Nach Voraussetzung ist $O \cap cl\ C \neq \emptyset$, Folgerung 2.29 liefert $cl\ C = cl(ri\ C)$. Es existiert also ein x in $O \cap cl(ri\ C)$, somit existiert eine Folge von Punkten (x^n) aus $ri\ C$, welche gegen x konvergiert. Wegen O offen liegen fast alle Folgenglieder in O. Es reicht bereits die Existenz eines solchen Folgengliedes für den Beweis der Behauptung aus. □

Folgerung 2.68 *Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ eine konvexe Funktion, C eine konvexe Menge mit $ri\ C \subseteq dom\ f$ und α eine reelle Zahl, so daß ein $x \in cl\ C$ existiert mit $f(x) < \alpha$. Dann existiert ein $y \in ri\ C$ mit $f(y) < \alpha$.*

Beweis: Wir definieren eine Hilfs-Funktion $g : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ durch

$$g(x) = \begin{cases} f(x) & \text{für } x \in cl\ C \\ +\infty & \text{sonst} \end{cases}$$

g ist dann ebenfalls konvex und die Anwendung von Folgerung 2.67 liefert die Behauptung. □

Folgerung 2.69 *Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ eine konvexe Funktion, C eine konvexe Menge auf welcher f endliche Werte hat. Ist $f(x) \geq \alpha\ \forall x \in C$, so gilt auch $f(x) \geq \alpha\ \forall x \in cl\ C$.*

Beweis: Direkt aus Folgerung 2.68. □

Eine weitere direkte Konsequenz ist die Tatsache, daß für eine konvexe Funktion f ihre Abschließung $cl\ f$ durch die Werte von f auf $ri(dom\ f)$ vollständig bestimmt ist:

Folgerung 2.70 *Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ und $g : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ konvexe Funktionen, für welche $ri(dom\ f) = ri(dom\ g)$ und $f(x) = g(x)\ \ \forall x \in ri(dom\ f)$. Dann gilt $cl\ f = cl\ g$.*

Beweis: Aus den Voraussetzungen folgt $ri(epi\ f) = ri(epi\ g)$ und somit nach Folgerung 2.30 $cl(epi\ f) = cl(epi\ g)$. Dies ist für eigentliche f und g gerade die Behauptung.
Für uneigentliche Funktionen ist die Behauptung trivial nach Definition von $cl\ f$. □

Satz 2.71
Ist $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ eine eigentliche konvexe Funktion, so ist $cl\ f(x)$ eine abgeschlossene eigentliche konvexe Funktion. f und $cl\ f$ unterscheiden sich höchstens in relativen Randpunkten von $dom\ f$.

Beweis: Da $epi\ f$ konvex ist und $epi(cl\ f) = cl(epi\ f)$ ist $epi(cl\ f)$ eine abgeschlossene konvexe Menge, $cl\ f$ also eine unterhalbstetige konvexe Funktion (Def. und Satz 2.60).
f ist endlich auf $dom\ f$. Sei $x \in ri(dom\ f)$ beliebig. Wir betrachten die vertikale Gerade $M = \{(x,\mu)^T : \mu \in \mathbb{R}\}$. Nach Lemma 2.66 schneidet diese $ri(epi\ f)$. Aus Lemma 2.34 folgt wegen $ri\ M = M = cl\ M$ die Beziehung $M \cap epi(cl\ f) = M \cap cl(epi\ f) = cl(M \cap epi\ f) = M \cap epi\ f$, denn $M \cap epi\ f = \{(x,\mu)^T : \mu \geq f(x)\}$ nach Definition von $epi\ f$. Somit ist $(cl\ f)(x) = f(x)$.
Ist hingegen $x \notin cl(dom\ f)$, so liefert Bemerkung 2.62 *iii)*, daß $(cl\ f)(x) = +\infty = f(x)$ (das ist die Aussage $cl(dom\ f) \supseteq dom(cl\ f)$ von Bemerkung 2.62 *iv)*. Somit ist gezeigt, daß sich f und $cl\ f$ höchstens in relativen Randpunkten von $dom\ f$ unterscheiden.
Damit ist auch klar, daß $cl\ f$ eigentliche Funktion ist, denn f ist endlich auf $dom\ f$ und nach

Folgerung 2.64 kann somit $cl\ f$ nirgends den Wert $-\infty$ annehmen. $cl\ f$ ist somit gleich seiner uhs Hülle und damit abgeschlossen. □

Dieser Satz liefert uns die Unterhalbstetigkeit jeder konvexen Funktion relativ zum relativen Inneren ihres effektiven Definitionsbereiches. Es gilt sogar noch mehr, dazu sei zunächst die Stetigkeit relativ zu einer Teilmenge erklärt.

Definition 2.72 *Eine auf $\mathbb{R}^n$ definierte Funktion heißt stetig relativ zu $S \subseteq \mathbb{R}^n$, wenn ihre Einschränkung auf S stetig ist.*

Dies bedeutet, daß für $x \in S$ $f(x^n)$ gegen $f(x)$ konvergieren muß für jede Folge (x^n) in S, welche gegen x konvergiert, dies muß jedoch nicht bei Annäherung an x über Punkte außerhalb von S gelten.

Satz 2.73
Eine konvexe Funktion $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ ist stetig relativ zu jeder relativ offenen konvexen Teilmenge C von $dom\ f$, insbesondere relativ zu $ri(dom\ f)$.

Beweis: Wir betrachten die Funktion g, die mit f auf C übereinstimmt und sonst als $+\infty$ erklärt wird.:

$$g(x) = \begin{cases} f(x) & x \in C \\ +\infty & \text{sonst} \end{cases}$$

Es ist $dom\ g = C$. f ist stetig relativ zu C gdw. g stetig relativ zu C. Außerdem können wir uns o.B.d.A. auf den Fall $dim(dom\ g) = n$ beschränken (Einschränkung auf $aff\ C$, vgl. Bem. vor Lemma 2.27), d.h. $dom\ g$ offen.
Ist f uneigentlich, so $g(x) = f(x) = -\infty \quad \forall x \in C$ nach Satz 2.63, Stetigkeit ist in diesem Falle trivial.
Sei nun g eigentlich, d.h. g endlich auf $dom\ g$. Satz 2.71 liefert uns $(cl\ g)(x) = g(x) \quad \forall x \in dom\ g$. Somit ist g unterhalbstetig auf $dom\ g$. Zu zeigen bleibt g oberhalbstetig, d.h. $-g$ unterhalbstetig auf $dom\ g$. Nach Satz 2.60 (in welchem die Funktion nicht konvex sein mußte) ist das äquivalent zu $L_{-g}(\alpha) = \{x \in \mathbb{R}^n : -g(x) \leq \alpha\} = \{x \in \mathbb{R}^n : g(x) \geq -\alpha\}$ abgeschlossen für alle $\alpha \in \mathbb{R}$. Da $dom\ g$ offen ist, haben wir nach Lemma 2.66

$$int(epi\ g) = \left\{ \begin{pmatrix} x \\ \mu \end{pmatrix} \in \mathbb{R}^{n+1} :\ x \in dom\ g,\ g(x) < \mu \right\} = \left\{ \begin{pmatrix} x \\ \mu \end{pmatrix} \in \mathbb{R}^{n+1} :\ g(x) < \mu \right\}$$

Somit ist die Menge $\{x \in \mathbb{R}^n : g(x) < -\alpha\}$ (das Komplement der Menge, von welcher wir Abgeschlossenheit zeigen wollen) die orthogonale Projektion des (offenen und konvexen) Durchschnitts von $int(epi\ g)$ mit dem offenen Halbraum $\{(x, \mu)^T \in \mathbb{R}^{n+1} : \mu < -\alpha\}$ auf den $\mathbb{R}^n$. Da das Bild einer offenen Menge bei einer orthogonalen Projektion offen ist, ist diese Menge offen, ihr Komplement $\{x \in \mathbb{R}^n : g(x) \geq -\alpha\}$ also abgeschlossen. □

Folgerung 2.74 *Sei $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ eine konvexe Funktion, so ist f stetig auf $\mathbb{R}^n$.*

Diese Tatsache ist nützlich, da Konvexität auch unter einigen Operationen erhalten bleibt, welche nicht notwendigerweise Stetigkeit erhalten.

Bemerkung 2.75 *Für konvexe Funktionen f und $g : \mathbb{R}^n \longrightarrow]-\infty, \infty]$ und relle Zahlen $\lambda,\ \mu \geq 0$ ist $\lambda f + \mu g$ konvex.*
*Sind f_i konvex für $i \in I$ (beliebige Indexmenge), so ist $f(x) := \sup\{f_i(x) :\ i \in I\}$ konvex (**ÜA**).*

Sei z.B. $f : \mathbb{R}^n \times T :\longrightarrow \mathbb{R}$ mit einer beliebigen Menge T, so daß $f(., t)$ konvex für alle $t \in T$ und $f(x, .)$ nach oben beschränkt für alle x (z.B. erfüllt für $f(x, .)$ stetig für alle $x \in \mathbb{R}^n$ und $T \subseteq \mathbb{R}$ abgeschlossenenes Intervall). Dann ist $h(x) := \sup\{f(x, t) :\ t \in T\}$ stetig auf $\mathbb{R}^n$.
Dies folgt, da nach Voraussetzung h nur endliche Werte annimmt und als punktweises Supremum einer Menge konvexer Funktionen wieder konvex ist.

Aus dem folgenden Satz ergibt sich, daß eigentliche konvexe Funktionen sogar *lokal Lipschitzstetig* über dem relativen Inneren ihres effektiven Definitionsbereiches sind.

Definition 2.76 *Eine Funktion* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *heißt lokal Lipschitzstetig (locally Lipschitzian) in* x*, falls ein* $L > 0$ *und ein* $\varepsilon > 0$ *existieren, so daß*

$$|f(x^1) - f(x^2)| \leq L\|x^1 - x^2\|$$

für alle x^1, x^2 *mit* $\|x^i - x\| \leq \varepsilon, \quad i = 1, 2.$

Satz 2.77
Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine eigentliche konvexe Funktion und* $S \subset ri(dom\ f)$ *eine beliebige kompakte Menge. Dann ist* f *auf* S *lipschitzstetig.*

Beweis: O.B.d.A. (siehe Bem. vor Lemma 2.27) sei $dom\ f$ volldimensional, d.h. $S \subset int(dom\ f)$. Sei $K = \{x \in \mathbb{R}^n : \|x\| \leq 1\}, \quad M_\varepsilon(S) = S \oplus \varepsilon K.$

Es ist $M_\varepsilon(S)$ kompakt für alle $\varepsilon > 0$. Dies folgt daraus, daß $S \times K$ kompakt und $M_\varepsilon(S) = \phi_\varepsilon(S \times K)$ mit der stetigen Abbildung $\phi_\varepsilon : \mathbb{R}^{2n} \longrightarrow \mathbb{R}^n, \quad \phi_\varepsilon(x, y) = x + \varepsilon y.$

Aufgrund der Voraussetzungen $S \subset int(dom\ f)$ und S kompakt existiert ein $\varepsilon^* > 0$, so daß $M_{\varepsilon^*}(S) \subset int(dom\ f)$, denn:
$S \subset int(dom\ f) \implies$ zu jedem $x \in S$ ex. $\varepsilon_x > 0$ mit $B(x, \varepsilon_x) \subseteq int(dom\ f)$. Die monotone Mengenfamilie $N_\varepsilon = (S \oplus \varepsilon K) \cap (\mathbb{R}^n \setminus int(dom\ f)) \quad \varepsilon > 0$ hat also einen leeren Durchschnitt. Angenommen, $N_\varepsilon \neq \emptyset \ \ \forall \varepsilon > 0$. Wir betrachten die N_{ε_n} für die Folge $\varepsilon_n = \frac{1}{n}$, dann ist $N_{\varepsilon_n} \subseteq N_1$. Enthält nun jede dieser Mengen einen Punkt x^n, so liegt diese Folge in der kompakten Menge N_1, sie besitzt also eine konvergente Teilfolge mit Grenzwert $\bar{x}$. Nach Konstruktion der N_ε gilt $\bar{x} \in \mathbb{R}^n \setminus int(dom\ f)$ (da diese Menge abgeschlossen ist) und $\bar{x} \in \bigcap_{n=1} S \oplus \frac{1}{n}K = S$ im Widerspruch zur Voraussetzung. Es existiert also ein ε^* mit $N_{\varepsilon^*} = \emptyset$.

Da nach Satz 2.73 f stetig auf der kompakten Menge $M_{\varepsilon^*}(S)$ ist, nimmt sie dort ihr Minimum α_1 und ihr Maximum α_2 an.
Seien nun $x, y \in S, \ x \neq y$ beliebig. Dann ist $\tilde{x} = y + \varepsilon^* \frac{y-x}{\|y-x\|} \in M_{\varepsilon^*}(S) \subset int(dom\ f)$. Damit ist *(umgestellt)*

$$y\left(1 + \frac{\varepsilon^*}{\|y - x\|}\right) = \tilde{x} + \frac{\varepsilon^*}{\|y - x\|}x$$

$$\text{und mit } \ \lambda := \frac{\|y - x\|}{\|y - x\| + \varepsilon^*} \text{ gilt dann } \ y = \lambda\tilde{x} + (1 - \lambda)x.$$

Die Konvexität von f liefert $f(y) \leq \lambda f(\tilde{x}) + (1 - \lambda)f(x)$, daraus folgt

$$f(y) - f(x) \leq \lambda(f(\tilde{x}) - f(x)) \leq \lambda(\alpha_2 - \alpha_1)$$

Durch Einsetzen der Definition von λ erhalten wir

$$f(y) - f(x) \leq \frac{\|y - x\|}{\|y - x\| + \varepsilon^*}(\alpha_2 - \alpha_1) \leq \underbrace{\frac{\alpha_2 - \alpha_1}{\varepsilon^*}}_{=:L}\|y - x\|$$

Vertauscht man nun die Rolle von x und y in den Betrachtungen, so erhält man insgesamt die Gültigkeit der Ungleichung für den Betrag, d.h. die Behauptung

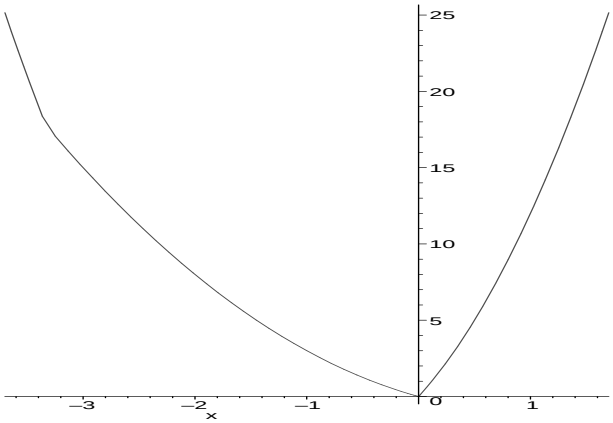
$$|f(y) - f(x)| \leq L\|y - x\|$$

□

2.3.3 Differenzierbarkeit konvexer Funktionen, Subdifferential

Nachdem wir die Stetigkeit als Folgerung aus der Konvexität einer reellwertigen Funktion erhalten haben, werden jetzt die Differenzierbarkeitseigenschaften betrachtet. Bereits die reelle Betragsfunktion zeigt, daß eine konvexe Funktion nicht differenzierbar zu sein braucht. Das folgende Beispiel einer stark konvexen Funktion zeigt, daß auch die eingeführten stärkeren Konvexitätsbegriffe die Differenzierbarkeit nicht garantieren:

$$f(x) = max(4x^2 + 8x, x^2 - 2x) = \begin{cases} 4x^2 + 8x & x \leq -\frac{10}{3} \\ x^2 - 2x & x \in [-\frac{10}{3}, 0] \\ 4x^2 + 8x & x \geq 0 \end{cases}$$



Die Verwendung eines schwächeren Differenzierbarkeitsbegriffes führt zu einer ersten positiven Aussage. Dabei wird der Begriff der Richtungsableitung aus der Analysis-Vorlesung auf einseitige Grenzwerte eingeschränkt.

Definition 2.78 *Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$, $x \in \mathbb{R}^n$ mit $f(x) \in \mathbb{R}$ und $d \in \mathbb{R}^n$. Existiert der Grenzwert ($+\infty$ und $-\infty$ als Grenzwerte zugelassen)*

$$f'(x,d) := \lim_{\lambda \longrightarrow +0} \frac{f(x+\lambda d) - f(x)}{\lambda}$$

so heißt er einseitige Richtungsableitung von f in x bzgl. Richtung d.

Wegen $f'(x,-d) = -\lim_{\lambda \longrightarrow -0} \frac{f(x+\lambda d)-f(x)}{\lambda}$ ist die Richtungsableitung $f'(x,d)$ zweiseitig (d.h. im Sinne der Defintion aus Ana der Grenzwert $\lambda \longrightarrow 0$ gebildet), wenn auch $f'(x,-d)$ existiert und $f'(x,-d) = -f'(x,d)$. Ist f differenzierbar in x, so existieren alle Richtungsableitungen, sind endlich und zweiseitig und es gilt $f'(x,d) = \nabla f(x)d \ \ \forall d$.

Satz 2.79
Sei $f : \mathbb{R}^n \longrightarrow]-\infty, +\infty]$ konvex, $x \in \mathbb{R}^n$ ein Punkt mit $f(x) \in \mathbb{R}$ und $0 \neq d \in \mathbb{R}^n$ eine Richtung. Dann ist der Differenzenquotient in der Definition der Richtungsableitung eine monoton wachsende Funktion in λ für $\lambda > 0$, es existiert die einseitige Richtungsableitung und

$$f'(x,d) := \inf_{\lambda > 0} \frac{f(x+\lambda d) - f(x)}{\lambda}$$

Beweis: Seien $\lambda_2 > \lambda_1 > 0$ gewählt. Aufgrund der Konvexität von f gilt

$$f(x+\lambda_1 d) = f\left[\frac{\lambda_1}{\lambda_2}(x+\lambda_2 d) + \left(1 - \frac{\lambda_1}{\lambda_2}\right)x\right] \leq \frac{\lambda_1}{\lambda_2} f(x+\lambda_2 d) + \left(1 - \frac{\lambda_1}{\lambda_2}\right) f(x)$$

Aus dieser Ungleichung ergibt sich

$$\frac{f(x+\lambda_1 d) - f(x)}{\lambda_1} \leq \frac{f(x+\lambda_2 d) - f(x)}{\lambda_2}$$

Dies zeigt, daß der Differenzenquotient monoton wachsend in λ ist für $\lambda > 0$.
Damit ist jede Folge von Differenzenquotienten mit monoton fallenden $\lambda_k > 0$, $\lambda_k \longrightarrow 0$ monoton fallend. Die Differenzenquotienten sind damit identisch $+\infty$ für $\lambda > 0$, divergieren bestimmt gegen $-\infty$ oder konvergieren gegen einen endlichen Wert. Da dies für jede Folge $\lambda_k \longrightarrow +0$ gilt, sind alle diese Grenzwerte gleich und die Richtungsableitung existiert. (Beweis indirekt: Angenommen, es gäbe zwei Folgen (λ_k^1), (λ_k^2) mit verschiedenen Grenzwerten der Differenzenquotienten. Dann betrachten wir die Mischfolge, bei welcher die Folgenglieder von (λ_k^2) so in die Folge (λ_k^1) einsortiert werden, daß wieder eine monoton fallende Folge entsteht. Die Folge der zugehörigen Differenzenquotienten hätte zwei verschiedene Häufungspunkte, was im Widerspruch dazu steht, daß sie monoton ist.) Wegen der Monotonie kann der Grenzwert durch das Infimum ersetzt werden. □

Definition 2.80
Eine Funktion $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *heißt positiv homogen, wenn* $f(\lambda x) = \lambda f(x) \quad \forall x \in \mathbb{R}^n, \forall \lambda > 0$.
Eine Funktion $f : \mathbb{R}^n \longrightarrow]-\infty, +\infty]$ *heißt subadditiv, wenn* $f(x+y) \leq f(x) + f(y) \quad \forall x, y \in \mathbb{R}^n$.

Ein Beispiel für eine positiv homogene und subadditive Funktion, welche nicht linear ist, ist jede Norm, da nach Normaxiomen gilt $\|\lambda x\| = |\lambda| \|x\|$ und $\|x + y\| \leq \|x\| + \|y\|$.

Diese Eigenschaft steht zu einer geometrischen Eigenschaft des Epigraphen in Beziehung.

Definition 2.81 *Eine Menge* $K \subseteq \mathbb{R}^n$ *heißt Kegel mit Scheitel in 0, wenn*

$$x \in K \implies \lambda x \in K \quad \forall \lambda > 0$$

Da wir in dieser Vorlesung nur Kegel mit Scheitel im Ursprung betrachten, werden wir diese auch kurz als Kegel bezeichnen.
Sei $M \subseteq \mathbb{R}^n$ *eine beliebigen Menge. Als konische Hülle von M (Bezeichnung: con(M)) wird der kleinste Kegel mit Scheitel in 0 bezeichnet, welcher M enthält, d.h.*

$$con(M) = \bigcap_{\substack{K \supseteq M, \\ K Kegel}} K$$

Ein Kegel nach dieser Definition muß nicht konvex sein (im Unterschied zur Def. in Opt. I). Bsp.: Vereinigung zweier Strahlen, welche vom Ursprung ausgehen.
In der Definition eines Kegels wird hier nicht verlangt, daß der Scheitel im Kegel enthalten ist, der Begriff wird in dieser Beziehung nicht einheitlich in der Literatur verwendet.

Lemma 2.82
$f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *ist positiv homogen genau dann, wenn epi f ein Kegel mit Scheitel in 0 ist*
Beweis: ÜA

Lemma 2.83
Sei $f : \mathbb{R}^n \longrightarrow]-\infty, +\infty]$ *positiv homogen. f ist konvex genau dann, wenn f subadditiv ist.*

Beweis: ÜA
"$\Longrightarrow$" $f(x+y) = 2f(\frac{x}{2} + \frac{y}{2}) \leq 2\left(\frac{1}{2}f(x) + \frac{1}{2}f(x)\right) = f(x) + f(y)$
"$\Longleftarrow$" $f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda) f(y)$ □

Lemma 2.84
Sei $f : \mathbb{R}^n \longrightarrow]-\infty, +\infty]$ *konvex. Dann gilt:*

i) $f'(x, 0) = 0 \quad \forall x \in \mathbb{R}^n$

ii) Für alle x ist $f'(x, .)$ *positiv homogen (als Funktion der Richtung).*

iii) Für alle x ist $f'(x, .)$ *subadditiv.*

Beweis:

i) Nach Definition der Richtungsableitung.

ii) Sei $\lambda > 0$.
$\lambda f'(x,y) = \lambda \lim\limits_{\mu \longrightarrow +0} \frac{f(x+\mu y)-f(x)}{\mu} = \lim\limits_{\mu \longrightarrow +0} \frac{f(x+\mu y)-f(x)}{\lambda^{-1}\mu} = \lim\limits_{\mu^* \longrightarrow +0} \frac{f(x+\lambda\mu^* y)-f(x)}{\mu^*} = f'(x,\lambda y)$
mit $\mu^* = \lambda^{-1}\mu$. Es ist wegen $\lambda > 0$ fix und $\mu > 0$ auch $\mu^* > 0$ und $\mu^* \longrightarrow +0 \Longleftrightarrow \mu \longrightarrow +0$.

iii) Seien $y, z \in \mathbb{R}^n$ beliebig, $\lambda > 0$ beliebig. Wegen der Konvexität von f gilt dann

$$f\left(x + \lambda(y+z)\right) = f\left(\frac{1}{2}(x+2\lambda y) + \frac{1}{2}(x+2\lambda z)\right) \leq \frac{1}{2}f(x+2\lambda y) + \frac{1}{2}f(x+2\lambda z)$$

Damit folgt

$$\begin{aligned} f'(x, y+z) &= \lim_{\lambda \longrightarrow +0} \frac{f(x+\lambda(y+z))-f(x)}{\lambda} \\ &\leq \lim_{\lambda \longrightarrow +0} \left(\frac{f(x+2\lambda y)-f(x)}{2\lambda} + \frac{f(x+2\lambda z)-f(x)}{2\lambda} \right) = f'(x,y) + f'(x,z) \end{aligned}$$

□

Folgerung 2.85 *Sei* $f : \mathbb{R}^n \longrightarrow]-\infty, +\infty]$ *eine konvexe Funktion, so ist* $f'(x,.)$ *konvex für alle* $x \in \mathbb{R}^n$.

Satz 2.86
Sei $f : \mathbb{R}^n \longrightarrow]-\infty, +\infty]$ *konvex,* $x \in int(dom\ f)$. *Dann ist* $f'(x,d)$ *endlich für alle* $d \in \mathbb{R}^n$.

Beweis: Wegen $x \in int(dom\ f)$ existiert zu jedem $d \in \mathbb{R}^n$ ein $\varepsilon > 0$, so daß $x - \varepsilon d \in dom\ f$. Für den Beweis der Endlichkeit der Richtungsableitung können wir uns wegen $f'(x, \varepsilon d) = \varepsilon f'(x,d)$ also o.B.d.A. auf den Fall $x - d \in dom\ f$ beschränken.
Ist $\lambda > 0$ so ergibt sich wegen der Konvexität von f

$$f(x) = f\left[\frac{\lambda}{1+\lambda}(x-d) + \frac{1}{1+\lambda}(x+\lambda d)\right] \leq \frac{\lambda}{1+\lambda}f(x-d) + \frac{1}{1+\lambda}f(x+\lambda d)$$

und damit

$$\begin{array}{lll} (1+\lambda)f(x) & \leq & \lambda f(x-d) + f(x+\lambda d) \quad | -\lambda f(x) - f(x+\lambda d) \\ \Longrightarrow \quad f(x) - f(x+\lambda d) & \leq & \lambda\left(f(x-d) - f(x)\right) \quad | : (-\lambda) \\ \Longrightarrow \quad \frac{f(x+\lambda d)-f(x)}{\lambda} & \geq & f(x) - f(x-d) \end{array}$$

Für jede Folge von Differenzenquotienten mit monoton fallenden $\lambda_k > 0$, $\lambda_k \longrightarrow 0$ gilt somit: wegen $x \in int(dom\ f)$ sind fast alle Folgenglieder kleiner als $+\infty$, nach Satz 2.79 ist die Folge monoton fallend und durch $f(x) - f(x-d)$ nach unten beschränkt, besitzt also einen endlichen Grenzwert. □

In Randpunkten von *dom f* kann die Richtungsableitung $-\infty$ sein, wie das Beispiel der rechtsseitigen Ableitung von

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

im Punkt $x = 0$ zeigt.
Beispiel mit einer im Randpunkt unstetigen Funktion:

$$f(x) = \begin{cases} 1 & \text{für } x = 0 \\ 0 & \text{für } x > 0 \\ +\infty & \text{sonst} \end{cases}$$

Nun befassen wir uns mit dem wichtigen Konzept des Subgradienten konvexer Funktionen.

Satz 2.87
Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine eigentliche konvexe Funktion und* $x^0 \in int(dom\ f)$. *Dann existiert eine echte Stützhyperebene an epi* f *in* $(x^0, f(x^0))^T$ *und jede echte Stützhyperebene ist nicht vertikal, d.h. es existiert ein Vektor* $a \in \mathbb{R}^n$, *so daß*

$$epi\ f \subseteq \left\{ \begin{pmatrix} x \\ \mu \end{pmatrix} \in \mathbb{R}^{n+1} : \begin{pmatrix} a \\ -1 \end{pmatrix}^T \begin{pmatrix} x \\ \mu \end{pmatrix} \leq a^T x^0 - f(x^0) \right\}$$

Insbesondere gilt für alle $x \in \mathbb{R}^n$

$$f(x) \geq f(x^0) + a^T(x - x^0)$$

Beweis: Lemma 2.66 liefert $\begin{pmatrix} x^0 \\ f(x^0) \end{pmatrix} \in rbd(epi\ f)$. Die Existenz einer echten Stützhyperebene folgt deshalb aus Satz 2.37 (allg. Trennungssatz).
Angenommen, diese sei vertikal, d.h. die letzte Komponente des Normalenvektors sei Null. Es existiert also ein $0 \neq a \in \mathbb{R}^n$ mit

$$epi\ f \subseteq \left\{ \begin{pmatrix} x \\ \mu \end{pmatrix} \in \mathbb{R}^{n+1} : a^T x \leq a^T x^0 \right\} = H^-$$

Da $dom\ f$ die orthogonale Projektion von $epi\ f$ auf den $\mathbb{R}^n$ ist, folgt daraus sofort

$$dom\ f \subseteq \left\{ x \in \mathbb{R}^n : a^T x \leq a^T x^0 \right\} = \Pi(H^-)$$

Die orthogonale Projektion der zugehörigen Hyperebene $H_n := \{x \in \mathbb{R}^n : a^T x = a^T x^0\}$ ist also eine Stützhyperebene an $dom\ f$ in x^0. Diese Hyperebene H_n ist eine trennende Hyperebene für $dom\ f$ und $\{x^0\}$, es ist aber nach Voraussetzung $ri(dom\ f) \cap ri\ \{x^0\} \neq \emptyset$, nach Satz 2.37 kann also keine echte trennende Hyperebene existieren, demzufolge muß gelten $dom\ f \subseteq H_n$. Dies ist ein Widerspruch zur Voraussetzung $x^0 \in int(dom\ f)$ (denn es muss ja $dom\ f$ volldimensional sein).

$epi\ f$ enthält den vertikale Strahl $\{(x^0, \mu)^T \in \mathbb{R}^{n+1} : \mu \geq f(x^0)\}$ und somit kann die letzte Komponente des Normalenvektors einer Hyperebene H^- nicht positiv sein, sie kann also durch Skalierung immer gleich -1 gemacht werden.
Ist $f(x) = +\infty$, so ist die Ungleichung $f(x) \geq f(x^0) + a^T(x - x^0)$ trivialerweise erfüllt. Für Punkte $x \in dom\ f$ ist $(x, f(x))^T \in epi\ f$, also gilt $a^T x - f(x) \leq a^T x^0 - f(x^0)$, was die zu beweisende Ungleichung liefert. □

Folgerung 2.88
Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine eigentliche konvexe Funktion und* $x^0 \in ri(dom\ f)$. *Dann existiert eine nicht vertikale echte Stützhyperebene an epi* f *in* $(x^0, f(x^0))^T$.

Beweis: Sei $dim(aff(dom\ f) = m$. Schränken wir dann die Betrachtung auf das Exemplar des $\mathbb{R}^m$ ein, welcher $aff(dom\ f)$ repräsentiert (vgl. Bem. vor Lemma 2.27) und betrachten $epi\ f$ im entprechenden $\mathbb{R}^{m+1}$ (entspr. $aff(dom\ f) \times \mathbb{R}$). Die Transformation betrifft nur die ersten m Komponenten. Wir haben in disesm Raum die Situation von Satz 2.87. Es existiert also ein m-dimensionaler Vektor a, so daß $(a, -1)^T$ in diesem Raum eine echte Stützhyperebene an $epi\ f$ in $(x^0, f(x^0))^T$ liefert. Die Ergänzung des Vektors a mit Nullen in den restlichen $n - m$ Dimensionen und Rücktransformation liefern den Normalenvektor der Stützhyperebene aus der Behauptung. □

Definition 2.89 *Ein Vektor* $a \in \mathbb{R}^n$ *heißt Subgradient einer konvexen Funktion* $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *im Punkt* x^0, *wenn*

$$f(x) \geq f(x^0) + a^T(x - x^0) \quad \forall x \in \mathbb{R}^n$$

Die Menge aller Subgradienten von f *in* x^0 *heißt das Subdifferential von* f *in* x^0.
Schreibweise: $\partial f(x^0)$. *Die Punkt-Menge-Abbildung* $\partial f : x \longrightarrow \partial f(x)$ *heißt Subdifferential von* f.
f *heißt subdifferenzierbar in* x^0, *wenn* $\partial f(x^0) \neq \emptyset$.

Beispiel 2.90

i) $f(x) = \|x\|$ *ist für* $x \neq 0$ *differenzierbar, nicht jedoch im Ursprung.* f *besitzt in diesem Punkt einen Subgradienten, denn es gilt* $\|x\| \geq 0$*, also erfüllt der Nullvektor die Definition eines Subgradienten im Ursprung.*
Nun wollen wir das Subdifferential in diesem Punkt bestimmen.
Es ist $\partial f(0) = \{a \in \mathbb{R}^n : \|x\| \geq a^T x \quad \forall x \in \mathbb{R}^n\}$*. Sei* $a \in \partial f(0)$ *, so muß die Ungleichung insbesondere für* $x = a$ *gelten, d.h.* $\|a\| \geq a^T a \implies \|a\| \leq 1$*. Sei umgekehrt* $a \in \mathbb{R}^n$ *mit* $\|a\| \leq 1$ *beliebig, so gilt nach der Cauchy-Schwarz'schen Ungleichung (Satz 1.54 Ana I)*

$$a^T x \leq \|a\|\|x\| \leq \|x\| \quad \forall x \in \mathbb{R}^n,$$

also ist $\partial f(0) = \{a \in \mathbb{R}^n : \|a\| \leq 1\}$.

ii) Sei $\emptyset \neq C \subseteq \mathbb{R}^n$ *konvex und* $\delta_C(x) = \begin{cases} 0 & x \in C \\ +\infty & \text{sonst} \end{cases}$ *die Indikatorfunktion von* C *(***Achtung:** *diese Definition der Indikatorfunktion unterscheidet sich von der in Ana III benutzten!). Dann ist nach Definition* $a \in \partial\delta_C(x^0)$ *gdw.*

$$\delta_C(x) \geq \delta_C(x^0) + a^T(x - x^0) \quad \forall x \in \mathbb{R}^n$$

Für $x^0 \in C$ *bedeutet diese Bedingung* $0 \geq a^T(x - x^0) \quad \forall x \in C$*, d.h.* a *ist Element des Normalenkegels an* C *in* x^0 *(insbesondere* $\partial\delta_C(x^0) = \{0\}$ *für* $x^0 \in int\, C$*).*
Für $x^0 \notin C$ *ist* $\partial\delta_C(x^0) = \emptyset$

Folgerung 2.91

i) Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *konvex und es existiere ein* $\bar{x} \in \mathbb{R}^n$ *mit* $f(\bar{x})$ *endlich und* $\partial f(\bar{x}) \neq \emptyset$*. Dann ist* f *eigentlich.*

ii) Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine eigentliche konvexe Funktion, so gilt*
$\partial f(x) \neq \emptyset \quad \forall x \in ri(dom\, f)$ *und*
$\partial f(x) = \emptyset \quad \forall x \notin dom\, f$.

Beweis: i) nach Vor. ist $dom\, f \neq \emptyset$ und f majorisiert eine affine Funktion, kann also nirgends den Wert $-\infty$ annehmen.
ii) Folgerung 2.88 liefert direkt die Existenz eines $a \in \partial f(x)$ für $x \in ri(dom\, f)$.
Ist hingegen $f(x^0) = +\infty$ und wählt man in der Definitions-Ungleichung des Subgradienten $x \in dom\, f$, so ist klar, daß die Ungleichung für kein $a \in \mathbb{R}^n$ erfüllbar ist.

ÜA: Geben Sie eine konvexe Funktion an, welche in mindestens einem Punkt endlich, aber nicht subdifferenzierbar ist.

Satz 2.92
Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *konvex und* $x^0 \in \mathbb{R}^n$ *ein beliebiger Punkt. Das Subdifferential* $\partial f(x^0)$ *ist eine abgeschlossene konvexe Menge.*

Beweis:

$$\begin{aligned} a \in \partial f(x^0) &\iff f(x) \geq f(x^0) + a^T(x - x^0) \quad \forall x \in \mathbb{R}^n \\ &\iff a \in \bigcap_{x \in \mathbb{R}^n} \{b \in \mathbb{R}^n : f(x^0) + b^T(x - x^0) \leq f(x)\} \end{aligned}$$

Die Menge $\partial f(x^0)$ ist also gegeben als Durchschnitt einer überabzählbar unendlichen Menge von abgeschlossenen Halbräumen (welche abgeschlossen und konvex sind), falls $f(x^0)$ endlich ist.
Ist $f(x^0) = +\infty$, so ist $\partial f(x^0) = \emptyset$, wenn $dom\, f \neq \emptyset$, bzw. gleich $\mathbb{R}^n$ sonst.
Im Falle $f(x^0) = -\infty$ ist $\partial f(x^0) = \mathbb{R}^n$. □

Nach Satz 2.79 existieren alle Richtungsableitungen einer konvexen Funktion, über $ri(dom\ f)$ ist sie nach Folgerung 2.91 subdifferenzierbar. Wie hängen nun diese beiden Ableitungsbegriffe zusammen?

Satz 2.93
Seien $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *konvex und* $x^0 \in \mathbb{R}^n$ *ein beliebiger Punkt mit* $f(x^0)$ *endlich. Dann gilt:*

$$a \in \partial f(x^0) \iff f'(x^0, d) \geq a^T d \quad \forall d \in \mathbb{R}^n$$

Beweis:

$$a \in \partial f(x^0) \iff f(x) \geq f(x^0) + a^T(x - x^0) \quad \forall x \in \mathbb{R}^n$$

Dies können wir äquivalent umschreiben mit dem Ansatz $x = x^0 + \lambda d$ mit $\lambda > 0$, wobei d den $\mathbb{R}^n$ durchläuft. Somit erhalten wir die äquivalente Bedingung ($f(x^0)$ kann auf die andere Seite der Ungleichung, da endlich)

$$f(x^0 + \lambda d) - f(x^0) \geq \lambda a^T d \quad \forall d \in \mathbb{R}^n \; \forall \lambda > 0$$

Division durch λ liefert wiederum äquivalent

$$\frac{f(x^0 + \lambda d) - f(x^0)}{\lambda} \geq a^T d \quad \forall d \in \mathbb{R}^n \; \forall \lambda > 0$$

Die rechte Seite der Ungleichung ist unabhängig von λ, somit können wir zum Limes inferior für $\lambda \longrightarrow +0$ übergehen und erhalten die Ungleichung $f'(x^0, d) \geq a^T d$.
Umgekehrt hatten wir gezeigt, daß der Differenzenquotient monoton in λ ist, also gilt $\frac{f(x^0+\lambda d)-f(x^0)}{\lambda} \geq f'(x^0, d) \quad \forall \lambda > 0$ und damit folgt aus $f'(x^0, d) \geq a^T d$ die obige Bedingung. □

Folgerung 2.94 *Sei* $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine eigentliche konvexe Funktion und* $x \in int(dom\, f)$. *Dann ist* $\partial f(x)$ *nichtleer und beschränkt.*

Beweis: Nach Folgerung 2.91 ist $\partial f(x)$ nichtleer. Nach Satz 2.86 ist $f'(x, d)$ endlich für alle $d \in \mathbb{R}^n$. Somit ist $\partial f(x)$ nach Satz 2.93 enthalten in einem n-dimensionalen Quader, also beschränkt. □

Bemerkung: R.T. Rockafellar beweist in "Convex Analysis", Theorem 23.4 noch mehr:
Ist $x \in ri(dom\, f)$, so gilt $f'(x, d) = sup\{a^T d : \; a \in \partial f(x)\}$.
In der Folgerung 2.94 gilt sogar Äquivalenz: $x \in int(dom\, f)$ gdw. $\partial f(x)$ nichtleer und beschränkt.

Nun soll uns noch der Zusammenhang Gradient-Subgradient für differenzierbare konvexe Funktionen interessieren. Der Differenzierbarkeitsbegriff läßt sich für Endlichkeitspunkte erweitert-reellwertiger Funktionen erweitern:
Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$, $x^0 \in \mathbb{R}^n$ mit $f(x^0) \in \mathbb{R}$. f heißt differenzierbar in x^0, wenn es einen (eindeutig bestimmten) Vektor $a \in \mathbb{R}^n$ gibt, so daß

$$f(x) = f(x^0) + a^T(x - x^0) + o(\|x - x^0\|)$$

oder mit anderen Worten

$$\lim_{x \longrightarrow x^0} \frac{f(x) - f(x^0) - a^T(x - x^0)}{\|x - x^0\|} = 0$$

Wenn ein solches a existiert, heißt es Gradient von f in x^0 und wird mit $\nabla f(x^0)$ bezeichnet.

Satz 2.95
Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *konvex,* $x^0 \in \mathbb{R}^n$ *mit* $f(x^0) \in \mathbb{R}$. *Ist* f *differenzierbar in* x^0, *so gilt* $\partial f(x^0) = \{\nabla f(x^0)\}$. *Ist umgekehrt* $\partial f(x^0)$ *einelementig, so ist* f *in* x^0 *differenzierbar.*

Beweis:
"$\Longrightarrow$" Ist f in x^0 differenzierbar, so gilt (Satz 5.28 der Analysis-Vorlesung) $f'(x^0, d) = \nabla f(x^0)d$.
Nach Satz 2.93 ist $a \in \partial f(x^0) \iff f'(x^0, d) \geq a^T d \quad \forall d \iff \nabla f(x^0)d \geq a^T d \quad \forall d$
Dies ist gleichbedeutend mit

$$(\nabla f(x^0) - a)^T d \geq 0 \quad \forall d \in \mathbb{R}^n,$$

woraus folgt $\nabla f(x^0) - a = 0$, also $\partial f(x^0) = \{\nabla f(x^0)\}$.
"$\Longleftarrow$" Sei nun $\partial f(x^0) = \{a\}$, d.h. es existiert genau ein $a \in \mathbb{R}^n$, so daß

$f(x) \geq f(x^0) + a^T(x - x^0) \quad \forall x \in \mathbb{R}^n$.
Wir betrachten die konvexe Funktion $g(y) = f(x^0 + y) - f(x^0) - a^T y$. Es ist $g(0) = 0$ und $g(y) \geq 0 \quad \forall y \in \mathbb{R}^n$ und somit

$$\begin{aligned} \partial g(0) &= \{b \in \mathbb{R}^n : \ g(y) \geq g(0) + b^T(y - 0) \quad \forall y \in \mathbb{R}^n\} \\ &= \{b \in \mathbb{R}^n : \ g(y) \geq b^T y \quad \forall y \in \mathbb{R}^n\} \\ \Longrightarrow \quad 0 &\in \partial g(0) \end{aligned}$$

Nehmen wir nun an, es gäbe einen Subgradienten $b \neq 0$ von g in 0, so würde folgen
$f(x^0 + y) - f(x^0) - a^T y = g(y) \geq g(0) + b^T(y - 0) \quad \forall y \in \mathbb{R}^n \iff$
$f(x^0 + y) \geq f(x^0) + (a + b)^T y \quad \forall y \in \mathbb{R}^n$,
mit dem Ansatz $x = x^0 + y$ bedeutet dies aber $f(x) \geq f(x^0) + (a + b)^T(x - x^0) \quad \forall x \in \mathbb{R}^n$,
d.h. $a + b \neq a$ wäre ebenfalls Subgradient von f in x^0 im Widerspruch zur Voraussetzung. Es ist also $\partial g(0) = \{0\}$. Zu zeigen ist nun

$$\lim_{y \longrightarrow 0} \frac{g(y)}{\|y\|} = 0$$

Mit Satz 2.93 folgt aus $0 \in \partial g(0)$, daß $g'(0, d) \geq 0 \quad \forall d \in \mathbb{R}^n$. Es gilt $0 \in int(dom\ g)$, andernfalls würde eine Stützhyperebene an $dom\ g$ im Ursprung existieren, es würde also ein $a \neq 0$ existieren mit $a^T x \leq 0 \quad \forall x \in dom\ g$. Dann wäre aber $g(x)(\geq 0) \geq a^T x \quad \forall x \in dom\ g$ und diese Ungleichung gilt für $x \notin dom\ g$ trivialerweise, a wäre also ebenfalls Subgradient von g im Ursprung $\Longrightarrow$ Wid.

Die nach Folgerung 2.94 zitierte Aussage von Theorem 23.4 aus *R. T. Rockafellar "Convex Analysis"* liefert dann aus $\partial g(0) = \{0\}$ sogar

$$g'(0, d) = \lim_{\lambda \longrightarrow +0} \frac{g(\lambda d) - g(0)}{\lambda} = \lim_{\lambda \longrightarrow +0} \frac{g(\lambda d)}{\lambda} = 0 \quad \forall d \in \mathbb{R}^n$$

Der Differenzenquotient ist nach Satz 2.79 eine monoton wachsende Funktion von λ für $\lambda > 0$. Die konvexen Funktionen $h_\lambda(d) = \frac{g(\lambda d)}{\lambda}$ für $\lambda > 0$ konvergieren also für $\lambda \longrightarrow +0$ punktweise monoton fallend gegen die Funktion, welche konstant gleich Null ist. Seien $\{a^1, \ldots, a^{n+1}\}$ Punkte mit $K = \{x \in \mathbb{R}^n : \|x\| \leq 1\} \subset conv\{a^1, \ldots, a^{n+1}\}$. Jedes $u \in K$ besitzt dann eine Darstellung $u = \sum_{i=1}^{n+1} \lambda_i a^i$ und es gilt $0 \leq h_\lambda(u) \leq \sum_{i=1}^{n+1} \lambda_i h_\lambda(a^i) \leq \max_{i=1,\ldots,n+1} h_\lambda(a^i)$.
Wegen $h_\lambda(a^i) \downarrow 0$ für $\lambda \downarrow 0$ folgt daraus, daß $h_\lambda(u) \downarrow 0$ für $\lambda \downarrow 0$ gleichmäßig in $u \in K$.
Ist also ein $\varepsilon > 0$ gegeben, so existiert ein $\delta > 0$, so daß $\frac{g(\lambda u)}{\lambda} \leq \varepsilon \quad \forall \lambda \in]0, \delta] \quad \forall u \in K$. Zu jedem $y \in \mathbb{R}^n$ mit $0 < \|y\| \leq \delta$ existiert genau ein $u \in K$ mit $y = \lambda u$ und $\lambda = \|y\|$. Also gilt

$$0 \leq \frac{g(y)}{\|y\|} \leq \varepsilon \quad \text{wenn } 0 < \|y\| \leq \delta \quad \Longrightarrow \quad \lim_{y \longrightarrow 0} \frac{g(y)}{\|y\|} = 0 \qquad \square$$

Wir haben den Begriff des Subgradienten nur für konvexe Funktionen definiert. Umgekehrt kann man aus der Existenz eines Vektors mit der definierenden Eigenschaft für jeden Punkt auf Konvexität der Funktion schließen:

Satz 2.96
Sei $C \subseteq \mathbb{R}^n$ eine konvexe Menge und für $f : C \longrightarrow \mathbb{R}$ existiere in jedem Punkt $x^0 \in ri\ C$ ein Vektor $a \in \mathbb{R}^n$ mit

$$f(x) \geq f(x^0) + a^T(x - x^0) \quad \forall x \in C$$

Dann ist f konvex über $ri\ C$.

Beweis: Seien $x^1, x^2 \in ri\ C$ und $\lambda \in [0, 1]$ beliebig. Nach Lemma 2.27 ist $ri\ C$ konvex und somit $\bar{x} := \lambda x^1 + (1 - \lambda)x^2 \in ri\ C$. Nach Voraussetzung existiert dann ein Vektor mit der angegebenen Eigenschaft, und insbesondere gilt

$$\begin{aligned} f(x^1) &\geq f(\bar{x}) + (1 - \lambda)a^T(x^1 - x^2) \\ f(x^2) &\geq f(\bar{x}) + \lambda a^T(x^2 - x^1) \end{aligned}$$

Multipliziert man nun die erste Ungleichung mit λ, die zweite mit $(1-\lambda)$ und addiert die Ergebnisse, so erhält man

$$\lambda f(x^1) + (1-\lambda) f(x^2) \geq f(\bar{x}) = f(\lambda x^1 + (1-\lambda) x^2)$$

Dies liefert nach Satz 2.42 die behauptete Konvexität von $f + \delta_{ri\, C}$. □

Abschließend sei in diesem Abschnitt noch ein Satz ohne Beweis zitiert, siehe *R.T. Rockafellar "Convex Analysis"*, Theorem 25.5.

Satz 2.97
Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ eine eigentliche konvexe Funktion und D die Menge der Punkte, in welchen f differenzierbar ist. Dann ist D eine dichte Teilmenge von $int(dom\ f)$, ihre Komplementmenge $int(dom\ f) \setminus D$ hat Maß Null. Weiterhin ist die Gradienten-Funktion $\nabla f : x \longrightarrow \nabla f(x)$ stetig relativ zu D.

3 Konvexe Optimierung - Optimalitätsbedingungen

3.1 Allgemeine konvexe Probleme

Wir betrachten zunächst ein allgemeines konvexes Problem in der Form

$$\min\{h(x) : \ x \in C\} \tag{1}$$

wobei $h : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ eine eigentliche konvexe Funktion und $C \subseteq \mathbb{R}^n$ eine konvexe Menge mit $C \cap dom\, h \neq \emptyset$ ist (d.h. es existiert ein $x^0 \in C$ mit $f(x^0) \in \mathbb{R}$). Die Voraussetzungen beschränken die Betrachtung auf die "interessanten" Probleme, denn für uneigentliche konvexe Funktionen bzw. $h(x) = +\infty \quad \forall x \in C$ ist das Problem trivial.
Dieses Problem kann für Untersuchungen, welche nicht die Struktur der Menge C benutzen, äquivalent in ein freies (d.h. nichtrestriktioniertes) Minimierungsproblem überführt werden. Dazu betrachten wir die Funktion $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$, welche durch Addition der Indikatorfunktion (siehe Bsp. 2.90) zu h definiert wird:

$$f(x) := h(x) + \delta_C(x) = \begin{cases} h(x) & x \in C \\ +\infty & x \notin C \end{cases}$$

Offensichtlich ist das Problem (1) äquivalent zu

$$\min\{f(x) : \ x \in \mathbb{R}^n\} \tag{2}$$

in dem Sinne, daß die Mengen der Minimalpunke gleich sind, insbesondere besitzt Problem (1) eine Lösung genau dann, wenn Problem (2) lösbar ist.
Die für differenzierbare Funktionen bekannte Bedingung des verschwindenden Gradienten kann nun mit Hilfe des Begriffes des Subdifferentials auf beliebige konvexe Funktionen ausgedehnt werden.
Wir bezeichnen $\alpha^* = \inf\limits_{x \in \mathbb{R}^n} f(x) \in \bar{\mathbb{R}}$. Offenbar gilt für die zugehörige Niveaumenge

$$\begin{aligned} L_f(\alpha^*) &= \{x \in \mathbb{R}^n : \ f(x) \leq \alpha^*\} \\ &= \{x \in \mathbb{R}^n : \ f(x) = \alpha^*\} \end{aligned}$$

Wir nennen diese Menge die Minimalmenge von f. Wir haben bereits gezeigt (siehe Satz 2.56, Lemma 2.50, Satz 2.60)

i) f streng konvex $\implies$ f stark quasikonvex $\implies$ $L_f(\alpha^*)$ einelementig oder leer

ii) $L_f(\alpha^*)$ konvex

iii) $L_f(\alpha^*)$ abgeschlossen, wenn f abgeschlossen (d.h. f uhs, da eigentlich)

Satz 3.1 *Sei* $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eigentliche konvexe Funktion. Dann gilt*

$$x^0 \in L_f(\alpha^*) \iff 0 \in \partial f(x^0)$$

Beweis: Anwendung der Definitionen:

$$\begin{aligned} x^0 \in L_f(\alpha^*) &\iff f(x^0) \leq f(x) \quad \forall x \in \mathbb{R}^n \\ &\iff f(x^0) + \underbrace{0^T(x - x^0)}_{=0} \leq f(x) \quad \forall x \in \mathbb{R}^n \\ &\iff 0 \in \partial f(x^0) \end{aligned}$$

□

Bevor wir für den Fall $f = h + \delta_C$ speziellere Aussagen machen, benötigen wir weitere Begriffe aus der konvexen Analysis.

Definition 3.2 *Sei* $C \subseteq \mathbb{R}^n$ *eine konvexe Menge. Ein Vektor* $y \in \mathbb{R}^n$ *heißt Rezessionsrichtung von* C *(direction of recession of* C*), wenn* $x + \lambda y \in C \quad \forall \lambda \geq 0 \; \forall x \in C$ *(Bezeichnung:* O^+C*).*

Satz 3.3 *Sei* $C \subseteq \mathbb{R}^n$ *eine konvexe Menge. Die Menge aller Rezessionsrichtungen von* C *ist ein konvexer Kegel mit Scheitel in 0 und es gilt*

$$O^+C = \{y \in \mathbb{R}^n : \{y\} \oplus C \subseteq C\}.$$

Beweis: Wir beweisen zunächst die zweite Aussage. Jedes $y \in O^+C$ hat die Eigenschaft, daß $x + y \in C \quad \forall x \in C$, d.h. $\{y\} \oplus C \subseteq C$. Ist andererseits $\{y\} \oplus C \subseteq C$, so folgt mittels $(m+1)\{y\} \oplus C = \{y\} \oplus (m\{y\} \oplus C) \subseteq C$ durch vollständige Induktion $x + my \subseteq C \quad \forall x \in C \; \forall m \in \mathbb{N}$. Betrachten wir ein beliebig fixiertes $x \in C$, so folgt aus der Konvexität von C, daß auch die Verbindungsstrecken aller dieser Punkte mit x in C enthalten sind, d.h. $x + \lambda y \in C \quad \forall \lambda \geq 0$. Folglich gilt $y \in O^+C$.
Konvexer Kegel ist ÜA: Die Eigenschaft O^+C Kegel mit Scheitel in 0 sieht man sofort. Zu zeigen bleibt Konvexität. Seien $y^1, y^2 \in O^+C$ und $\lambda \in [0, 1]$ beliebig. Dann gilt

$$(\{\lambda y^1 + (1-\lambda)y^2\} \oplus C = \lambda(\{y^1\} \oplus C) \oplus (1-\lambda)(\{y^2\} \oplus C) \subseteq \lambda C \oplus (1-\lambda)C = C$$

□

Definition 3.4 *Sei* $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine eigentliche konvexe Funktion. Ein Vektor* $y \in \mathbb{R}^n$ *heißt Rezessionsrichtung von* f*, wenn*

$$\begin{pmatrix} y \\ 0 \end{pmatrix} \in O^+(epi\ f)$$

ÜA: Die Menge aller dieser Richtungen ist ein konvexer Kegel mit Scheitel in 0 (Durchschnitt des Rezessionskegels von *epi* f mit der Hyperebene $\{(y, 0)^T : y \in \mathbb{R}^n\}$) und ist abgeschlossen, wenn f eine abgeschlossene Funktion ist. Es gilt

$$\begin{aligned} \begin{pmatrix} y \\ 0 \end{pmatrix} \in O^+(epi\ f) &\iff \begin{pmatrix} x + \lambda y \\ \mu \end{pmatrix} \in epi\ f \quad \forall \lambda \geq 0 \; \forall \begin{pmatrix} x \\ \mu \end{pmatrix} \in epi\ f \\ &\iff f(x + \lambda y) \leq f(x) \qquad \forall \lambda \geq 0 \; \forall x \in dom\ f \end{aligned}$$

Mit Satz 3.3 folgt: y ist Rezessionsrichtung von f gdw. $\sup\{f(x + y) - f(x) : x \in dom\ f\} \leq 0$.

Der folgende Satz besagt, daß für unterhalbstetige eigentliche konvexe Funktionen, welche keine Richtung außer dem Nullvektor mit dieser Eigenschaft besitzen, das Minimierungsproblem lösbar ist. Außerdem läßt sich eine Eigenschaft beweisen, die für Optimierungsverfahren wichtig ist, da für eine Reihe von Verfahren nur bewiesen werden kann, daß sie eine sogenannte Minimalfolge generieren, d.h. eine Folge (x^n) mit $\lim\limits_{n \longrightarrow \infty} f(x^n) = \inf\limits_{x \in \mathbb{R}^n} f(x)$. Diese Aussagen sollen hier ohne Beweis angegeben werden. Die Beweise dieser Aussagen werden in *R.T. Rockafellar "Convex Analysis"* (dort Satz 27.2, Corollar 27.2.1 und 27.2.2) mit Hilfe von Fenchel-konjugierten Funktionen geführt, welche wir in dieser Vorlesung nicht eingeführt haben.

Satz 3.5

i) Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *abgeschlossene eigentliche konvexe Funktion, welche keine Rezessionsrichtung* $y \neq 0$ *besitzt. Dann ist das Infimum von* f *endlich und wird angenommen. Sei weiterhin* (x^n) *eine Folge von Punkten im* $\mathbb{R}^n$*, so daß* $\lim\limits_{n\longrightarrow\infty} f(x^n) = \inf\limits_{x\in\mathbb{R}^n} f(x)$*. Dann ist die Folge* (x^n) *beschränkt und alle ihre Häufungspunkte liegen in der Minimalmenge von* f*.*

ii) Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *abgeschlossene eigentliche konvexe Funktion, welche ihr Minimum in genau einem Punkt* $\bar{x} \in \mathbb{R}^n$ *annimmt. Ist* (x^n) *eine Folge von Punkten im* $\mathbb{R}^n$ *mit* $\lim\limits_{n\longrightarrow\infty} f(x^n) = \inf\limits_{x\in\mathbb{R}^n} f(x)$*, so konvergiert diese Folge gegen* $\bar{x}$*.*

Zur Erinnerung: h abgeschlossen war im Falle h eigentlich gleichbedeutend mit h unterhalbstetig. Zunächst noch eine Folgerung hinsichtlich der Spezialisierung von Satz 3.5 auf den Fall $f = h+\delta_C$.

Folgerung 3.6 *Es sei* $h : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine abgeschlossene eigentliche konvexe Funktion und* $C \subseteq \mathbb{R}^n$ *eine abgeschlossene konvexe Menge mit* $dom\, h \cap C \neq \emptyset$*. Wenn* h *und* C *keine gemeinsame Rezessionsrichtung besitzen, so wird das Infimum von* h *über* C *angenommen. Die Aussagen von Satz 3.5 über Minimalfolgen gelten analog.*

Beweis: Die Funktion $f = h + \delta_C$ ist unter den gemachten Voraussetzungen eine abgeschlossene eigentliche konvexe Funktion, deren Rezessionsrichtungen gerade die gemeinsamen Rezessionsrichtungen von h und C sind. □

Nun kommen wir zur Spezialisierung von Satz 3.1.

Satz 3.7 *Sei* h *eine eigentliche konvexe Funktion und* C *eine konvexe Menge. Weiterhin sei ein Punkt* $x^0 \in C$ *gegeben. Wenn es einen Subgradienten* $a \in \partial h(x^0)$ *gibt, welcher die Ungleichung* $a^T(x - x^0) \geq 0$ *für alle* $x \in C$ *erfüllt, so ist* x^0 *Minimum von* h *über* C*.*
Die Bedingung ist notwendig und hinreichend, wenn $ri(dom\, h) \cap ri\, C \neq \emptyset$*.*

Beweis: Zunächst ist die Bedingung offensichtlich hinreichend, denn

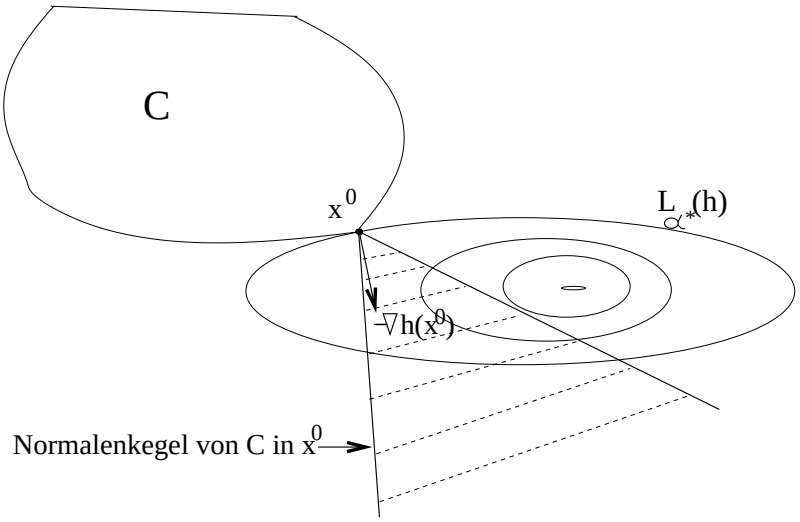
$$a \in \partial h(x^0) \implies h(x) \geq h(x^0) + \underbrace{a^T(x - x^0)}_{\geq 0} \geq h(x^0) \quad \forall x \in C$$

Sei umgekehrt analog zu oben $\alpha^0 = \inf\limits_{x\in C} h(x)$. Wir betrachten im $\mathbb{R}^{n+1}$ die konvexen Mengen $C_1 = epi\, h$ und $C_2 = \{(x,\mu)^T \in \mathbb{R}^{n+1} : x \in C,\ \mu \leq \alpha^0\}$. Mit der Beschreibung von $ri(epi\, h)$ aus Lemma 2.66 ist klar, daß $ri\, C_1 \cap ri\, C_2 = \emptyset$. Diese Mengen können also durch eine Hyperebene echt getrennt werden, unter der Voraussetzung $ri(dom\, h) \cap ri\, C \neq \emptyset$ ist sogar jede echt trennende Hyperebene nichtvertikal, denn mit einem Punkt $\bar{x} \in ri(dom\, h) \cap ri\, C$ ist der Strahl $\{(\bar{x},\mu)^T : \mu \geq h(\bar{x}) + 1\}$ in $ri\ C_1$ enthalten, der Strahl $\{(\bar{x},\mu)^T : \mu \leq \alpha^0 - 1\}$ in $ri\ C_2$. Wäre die Hyperebene vertikal, so enthielte sie diese Punkte aus dem relativen Inneren beider Mengen. Sie ist aber trennende Hyperebene, d.h. die gesamte Menge und somit die Umgebung jedes relativ inneren Punktes bezüglich der affinen Hülle liegt ganz in einem der von dieser Hyperebene definierten abgeschlossenen Halbräume. Dies ist nur möglich, wenn die affinen Hüllen und somit beide Mengen in der Hyperebene enthalten sind im Widerspruch zur echten Trennung.

Die Hyperebene ist der Graph einer affinen Funktion, beschrieben durch eine Gleichung der Form $\mu = a^T x + \beta$. Ist nun x^0 ein Punkt, in welchem das Minimum angenommen wird, so ist $(x^0, \alpha^0)^T \in C_1 \cap C_2$. Es gilt also $\alpha^0 = a^T x^0 + \beta$ $(*)$ und die Hyperebene ist Stützhyperebene an $epi\, f$ in $(x^0, \alpha^0)^T$. Der Vektor a aus der Beschreibung der Hyperebene ist also Subgradient von h in diesem Punkt. Außerdem ist $C_2 \subseteq H^+ = \{(x,\mu)^T : \mu \leq a^T x + \beta\}$, betrachten wir insbesondere die Punkte $(x, \alpha^0)^T \in C_2$ für $x \in C$, so ergibt sich unter Benutzung der Gleichung $(*)$:
$\alpha^0 \leq a^T x + \beta = \alpha^0 + a^T(x - x^0)$ d.h. $a^T(x - x^0) \geq 0$ für alle $x \in C$. □

Die Aussage des Satzes bedeutet unter der Verschärfung unserer generellen Voraussetzung, daß die Mengen C und $dom\, h$ relativ innere Punkte gemeinsam haben ($ri(dom\, h) \cap ri\, C \neq \emptyset$ gilt), daß ein

Punkt Minimalpunkt des Problems ist genau dann, wenn es einen Subgradienten a der Funktion in diesem Punkt gibt, so daß $-a$ im Normalenkegel von C liegt. Im Falle differenzierbarer Zielfunktion h ist die Bedingung gleichbedeutend damit, daß $-\nabla h(x^0)$ im Normalenkegel an C ist (der für innere Punkte von C gerade nur aus dem Nullvektor besteht).



Für differenzierbare Zielfunktionen h können wir sogar die Menge aller Lösungen des Problems (1) charakterisieren:

Satz 3.8 *Sei* $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ *eine differenzierbare konvexe Funktion und* C *eine konvexe Menge. Weiterhin sei der Punkt* $x^0 \in C$ *eine Lösung des Problems (1) (d.h.* $h(x^0) \leq h(x) \quad \forall x \in C$*). Dann gilt für die Menge* C_{opt} *aller Lösungen*

$$C_{opt} = N := \{x \in C : \ \nabla h(x) = \nabla h(x^0), \ \ \nabla h(x^0)(x - x^0) = 0\}$$

Beweis:
"$N \subseteq C_{opt}$" h ist konvex und differenzierbar, also nach Satz 2.95

$$x \in N \Longrightarrow \ h(x^0) \geq h(x) + \underbrace{\nabla h(x)}_{=\nabla h(x^0)}(x^0 - x) = h(x) + \underbrace{\nabla h(x^0)(x^0 - x)}_{=0} = h(x) \ \Longrightarrow \ x \in C_{opt}$$

"$C_{opt} \subseteq N$" Sei $x \in C_{opt}$, d.h. $x \in C, \ h(x) = h(x^0)$. Die Menge C_{opt} ist gerade die Minimalmenge der zugeordneten Funktion $f = h + \delta_C$ und somit konvex. Wir betrachten die Richtung $d := x - x^0$. Dann ist $h'(x^0, d) = \lim\limits_{\lambda \longrightarrow +0} \frac{h(x^0 + \lambda d) - h(x^0)}{\lambda} = 0$.
Andererseits gilt $h'(x^0, d) = \nabla h(x^0)d$, d.h. $\nabla h(x^0)(x - x^0) = 0$.
Es bleibt zu zeigen $\nabla h(x) = \nabla h(x^0)$. Wir betrachten die Hilfsfunktion

$$g(z) = h(z) - \nabla h(x^0)(z - x^0)$$

Diese ist als Summe zweier konvexer Funktionen konvex. Für $z = x$ ist

$$g(x) = h(x) = h(x^0) = g(x^0), \text{ außerdem ist } \nabla g(z) = \nabla h(z) - \nabla h(x^0)$$

Angenommen, $\nabla g(x) \neq 0$. Für $y = -\nabla g(x)$ ist $\nabla g(x)y < 0$.
Damit ist $0 > g'(x, y) = \lim\limits_{\lambda \longrightarrow +0} \frac{g(x + \lambda y) - g(x)}{\lambda}$ und somit gilt für hinreichend kleine positive λ: $g(x + \lambda y) < g(x) = g(x^0) = h(x^0)$. Daraus folgt mit der Definition von g:
$h(x + \lambda y) < h(x^0) + \nabla h(x^0)((x + \lambda y) - x^0)$ im Widerspruch zur Konvexität von h. Somit ist $0 = \nabla g(x) = \nabla h(x) - \nabla h(x^0)$. □

Bemerkung: Im Theorem 3.4.4 des Buches von *Bazaraa, Sherali, Shetty "Nonlinear Programming"* ist (dort unter der Voraussetzung zweimaliger Differenzierbarkeit von h) bewiesen:

$$C_{opt} = \{x \in C : \ \nabla h(x) = \nabla h(x^0), \ \ \nabla h(x^0)(x - x^0) \leq 0\}$$

Diese Beschreibung ist nur formal verschieden von der oben angegebenen, da x^0 ja als optimal vorausgesetzt ist und somit nach Satz 3.7 (hier ist *dom* $h = \mathbb{R}^n$!) unter Differenzierbarkeit gilt

$$\nabla h(x^0)(x - x^0) \geq 0 \quad \forall x \in C$$

Die Optimalitätsbedingungen für das allgemeine konvexe Problem (1) lassen sich auch sehr anschaulich geometrisch formulieren. Die dabei verwendeten Ideen lassen sich auch in viel allgemeineren Räume verwenden (lineare topologische Räume, Banachräume) sowie für nichtkonvexe Probleme. Wir bleiben jedoch im euklidischen Raum.
Dazu führen wir zunächst weitere Begriffe ein.

Definition 3.9 *Sei* $\emptyset \neq M \subseteq \mathbb{R}^n$ *eine Menge und* $x \in cl\ M$. *Ein Vektor* $d \in \mathbb{R}^n$ *heißt zulässige Richtung von* M *in* x, *falls ein* $\delta > 0$ *existiert, so daß* $x + \alpha d \in M \quad \forall \alpha \in]0, \delta[$.
Die Menge aller zulässigen Richtungen von M *in* x^0 *bezeichnen wir mit* $F_M(x)$.

Beispiele:

i) $M = \{x \in \mathbb{R}^2 : \max(|x_1|, |x_2|) \leq 1\}$, $x = (1,1)^T$. M ist abgeschlossenes Quadrat der Seitenlänge 2, $F_M(x) = \{d \in \mathbb{R}^2 : d_i \leq 0 \ \ i = 1,2\}$ ist der negative Orthant (abgeschlossen).

ii) $M = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$, $x = (0,1)^T$. M ist Kreis um Ursprung mit Radius 1, $F_M(x) = \{d \in \mathbb{R}^2 : d_2 < 0\}$ ist offener Halbraum.

iii) $M = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1,\ x_1 \geq 0\}$, $x = (0,1)^T$. M ist Halbkreis,
$F_M(x) = \{d \in \mathbb{R}^2 : d_1 \geq 0,\ d_2 < 0\}$ ist weder offen noch abgeschlossen.

Definition 3.10 *Sei* $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$. *Ein Vektor* $d \in \mathbb{R}^n$ *heißt Abstiegsrichtung für* f *in* x, *falls ein* $\delta > 0$ *existiert, so daß* $f(x + \alpha d) < f(x) \quad \forall \alpha \in]0, \delta[$.
Die Menge aller Abstiegsrichtungen von f *in* x *bezeichnen wir mit* $A_f(x)$.

Zunächst Aussagen zur Struktur dieser Mengen.

Lemma 3.11 *Sei* $M \subseteq \mathbb{R}^n$ *eine beliebige Menge und* $x \in cl\ M$. *Dann gilt:*

i) $F_M(x)$ *ist ein Kegel mit Scheitel in 0 und ist konvex, wenn* M *konvex.*

ii) $0 \in F_x(M) \iff x \in M$

iii) $F_M(x) \subseteq con(M \ominus \{x\})$

iv) Ist M *konvex und ist zusätzlich eine der Voraussetzungen* $x \in M$ *oder* M *abgeschlossen oder* M *offen erfüllt, so* $F_M(x) = con(M \ominus \{x\})$

Beweis:

i) Kegel klar nach der Definition, man wähle $\bar{\delta} = \frac{\delta}{\lambda} > 0$. Konvexität: Seien $d^1,\ d^2 \in F_M(x)$ mit zugehörigen $\delta_1,\ \delta_2 > 0$ sowie $\lambda \in]0,1[$. Dann ist mit $\bar{\delta} = \min\{\delta_1, \delta_2\} > 0$
$x + \alpha(\lambda d^1 + (1-\lambda)d^2) = \lambda \underbrace{(x + \alpha d^1)}_{\in M} + (1-\lambda) \underbrace{(x + \alpha d^2)}_{\in M} \in M \quad \forall \alpha \in]0, \bar{\delta}[$ falls M konvex.

ii) trivial

iii) Sei $d \in F_M(x)$, dann existiert ein $\delta > 0$ mit $x + \alpha d \in M \quad \forall \alpha \in (0, \delta)$. Dies bedeutet aber $\alpha d \in M \ominus \{x\}$, somit $d \in con(M \ominus \{x\})$.

iv) Sei $z \in M$ beliebig. Ist M abgeschlossen, so gilt wie in der ersten zusätzlichen Voraussetzung $x \in M$ und wegen Konvexität von M gilt $\lambda z + (1-\lambda)x = x + \lambda(z - x) \in M \quad \forall \lambda \in]0,1[$. Im Falle M offen gilt diese Beziehung nach Lemma 2.8. Somit ist $z - x \in F_M(x)$. Da dies für jedes $z \in M$ gilt, bedeutet das $M \ominus \{x\} \subseteq F_M(x)$. Nach i) ist $F_M(x)$ Kegel, somit ist $F_M(x)$ einer der Kegel, über welche der Durchschnitt gemäß der Definition von $con(M \ominus \{x\})$ gebildet wird, also

$$con(M \ominus \{x\}) \subseteq F_M(x)$$

Mit iii) zusammen folgt die Behauptung. □

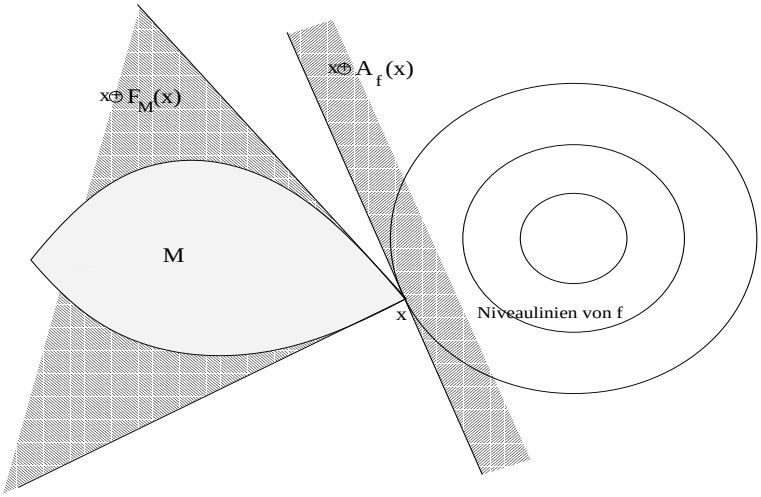
ÜA: Beispiel dafür, daß die Behauptung *iv)* nicht gelten muß, wenn M weder offen noch abgeschlossen ist:
Wähle $M = \{x \in \mathbb{R}^2 : \ 0 < x_i < 1 \ \ i = 1,2\} \cup \{(0,1)^T\}, \quad \bar{x} = (0,0)^T \in cl\ M \setminus M$. Hier ist $F_M(\bar{x}) = \{d \in \mathbb{R}^2 : \ d_i > 0 \ \ i = 1,2\}$, aber $con(M \ominus \{\bar{x}\}) = \{d \in \mathbb{R}^2 : \ d_1 \geq 0, \ d_2 > 0\}$.

Bemerkung 3.12 *Mit der Menge* $M(x,f) = \{y \in \mathbb{R}^n : \ f(y) < f(x)\}$ *gilt* $A_f(x) = F_{M(x,f)}(x)$. *Damit übertragen sich die Struktureigenschaften aus Lemma 3.11 auch auf diese Menge. Die Menge* $M(x,f)$ *ist für konvexe reellwertige Funktionen* f *offen und konvex.*

Mit diesen Begriffen ergibt sich eine ganz allgemeine notwendige Bedingung, die an keinerlei Konvexitäts- oder Differenzierbarkeitsvoraussetzung geknüpft ist.

Satz 3.13 (notwendige Optimalitätsbedingung in geometrischer Form)
Sei $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ *eine beliebige Funktion,* $M \subseteq \mathbb{R}^n$ *eine beliebige Menge. Ist* x^0 *lokales Minimum von* f *auf* M*, so gilt* $F_M(x^0) \cap A_f(x^0) = \emptyset$.

Beweis: indirekt: Angenommen, es existiere $d \in F_M(x^0) \cap A_f(x^0)$. Dann existiert $\delta_1 > 0$, so daß $x^0 + \alpha d \in M \quad \forall \alpha \in]0, \delta_1[$ und ein $\delta_2 > 0$, so daß $f(x^0 + \alpha d) < f(x^0) \quad \forall \alpha \in]0, \delta_2[$. Dann gelten aber mit $\bar{\delta} = \min(\delta_1, \delta_2) > 0$ beide Eigenschaften für alle $\alpha \in]0, \bar{\delta}[$ im Widerspruch zur Voraussetzung, daß x^0 lokales Minimum (die Strecke schneidet jede Umgebung von x^0). □

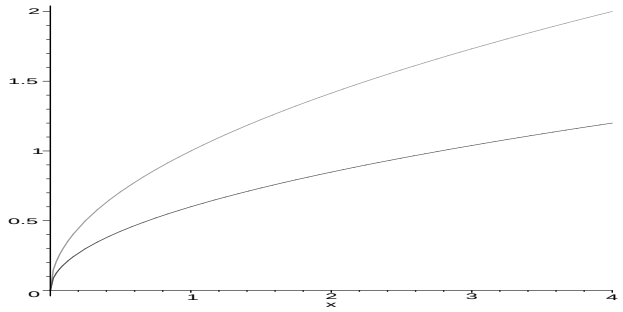


An dem kurzen Beweis ist bereits zu sehen, daß diese Bedingung wohl zunächst nicht viel mehr an Überprüfbarkeit bringen wird als die Definition eines lokalen Optimums. Sie zeigt aber eine andere Anschauung, da wir es jetzt mit der Disjunktheit von Kegeln zu tun haben. So können wir unter weiteren Voraussetzungen auch Trennungssätze anwenden.

Wie so oft ist auch diese allgemein gültige Bedingung nur schwer nachprüfbar und nur unter zusätzlichen Annahmen ergeben sich handhabbare Kriterien.
Allerdings kann ohne Konvexität und Regularitätsvoraussetzungen bereits die Menge $F_M(x^0)$ leer sein, wie das folgende Beispiel veranschaulicht.

$$\min\{f(x) : \ x \in M\} \quad M = \{x = (x_1, x_2)^T \in \mathbb{R}^2 : \ x_1 \geq 0, \ x_2 \leq \sqrt{x_1}, \ x_2 \geq 0.6\sqrt{x_1}\}$$

Im Koordinatenursprung existiert keine zulässige Richtung, da beide Restriktions-Funktionen bei Annäherung an $x_1 = 0$ sich asymptotisch der x_2-Achse nähern.



Somit ist für jede beliebige Funktion f im Koordinatenursprung diese notwendige Bedingung erfüllt.
$\longrightarrow$ Regularitätsvoraussetzungen nötig.
Aus diesem Grunde werden auch Bedingungen auf Basis weiterer Mengen betrachtet, wie z.B. der Menge der tangential zulässigen Richtungen (was hier nicht ausgeführt wird).

Es gilt der folgende Satz über die Beschreibung des Kegels der zulässigen Richtungen bei spezieller Menge M.

Satz 3.14
Es sei $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ *richtungsdifferenzierbar in* x *und* $M = \{z \in \mathbb{R}^n : \ g(z) \leq g(x)\}$. *Dann gilt*

$$\{d \in \mathbb{R}^n : \ g'(x,d) < 0\} \subseteq F_M(x) \subseteq \{d \in \mathbb{R}^n : \ g'(x,d) \leq 0\}$$

Beweis: Sei zunächst d so gewählt, daß $g'(x,d) < 0$. Dann gilt $g(x+\alpha d) - g(x) = \alpha g'(x,d) + o(\alpha)$ für $\alpha > 0$. Wegen $\lim\limits_{\alpha \longrightarrow +0} \frac{o(\alpha)}{\alpha} = 0$ existiert ein $\delta > 0$, so daß $\left|\frac{o(\alpha)}{\alpha}\right| < |g'(x,d)| \quad \forall \alpha \in]0,\delta[$.
Damit ist $g(x+\alpha d) < g(x) \quad \forall \alpha \in]0,\delta[$, also $x + \alpha d \in M \ \forall \alpha \in]0,\delta[$, d.h. $d \in F_M(x)$.
Nun sei $d \in F_M(x)$. Dann existiert $\delta > 0$, so daß $x + \alpha d \in M \ \forall \alpha \in]0,\delta[$, mit der Beschreibung vom M also $g(x+\alpha d) \leq g(x) \quad \forall \alpha \in]0,\delta[$. Daraus folgt direkt $g'(x,d) \leq 0$. □

g ist nur als richtungsdifferenzierbar vorausgesetzt, könnte also z.B. als Maximum konvexer Funktionen definiert sein.

Der vorangegangene Satz liefert nur eine Inklusionskette, welche für allgemeine Funktionen strenge Inklusionen beinhalten kann. Unter einer zusätzlichen Voraussetzung an die Funktion g läßt sich Gleichheit beweisen:

Lemma 3.15
Es sei $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ *differenzierbar und streng pseudokonvex,* $M = \{z \in \mathbb{R}^n : \ g(z) \leq g(x)\}$. *Dann gilt*

$$F_M(x) = \{d \in \mathbb{R}^n : \ g'(x,d) < 0\} = \{d \in \mathbb{R}^n : \ \nabla g(x) d < 0\}$$

Beweis: Strenge Pseudokonvexität von g bedeutet nach Definition, daß

$$g(x) \geq g(y) \implies \nabla g(x)(y-x) < 0$$

Ist also $d \in F_M(x)$, d.h. $g(x) \geq g(x+\alpha d) \quad \forall \alpha \in]0,\delta[$ mit einem $\delta > 0$, so folgt
$\alpha \nabla g(x) d = \nabla g(x)(\alpha d) < 0 \quad \forall \alpha \in]0,\delta[$ und damit die Behauptung. □

Völlig analog läßt sich die für die Abstiegsrichtungen wichtige Beschreibung beweisen.

Lemma 3.16
Es sei $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ *differenzierbar und pseudokonvex,* $M = \{z \in \mathbb{R}^n : \ g(z) < g(x)\}$.
Dann gilt

$$F_M(x) = \{d \in \mathbb{R}^n : \ g'(x,d) < 0\} = \{d \in \mathbb{R}^n : \ \nabla g(x) d < 0\}$$

Dies besagt für den Kegel der Abstiegsrichtungen einer in x differenzierbaren pseudokonvexen Funktion f:
$A_f(x) = \emptyset$, falls $\nabla f(x) = 0$, andernfalls ist $A_f(x)$ ein offener Halbraum.

Die notwendige Bedingung von Satz 3.13 ist im allgemeinen Fall nicht hinreichend, wie das folgende Beispiel zeigt.
Wir betrachten $C = \mathbb{R}^2$ und die Funktion $f(x) = (x_2 - x_1^2)(x_2 - 2x_1^2)$ sowie den Punkt $x^* = (0,0)^T$. Es gilt für jede Richtung $d \neq 0$:

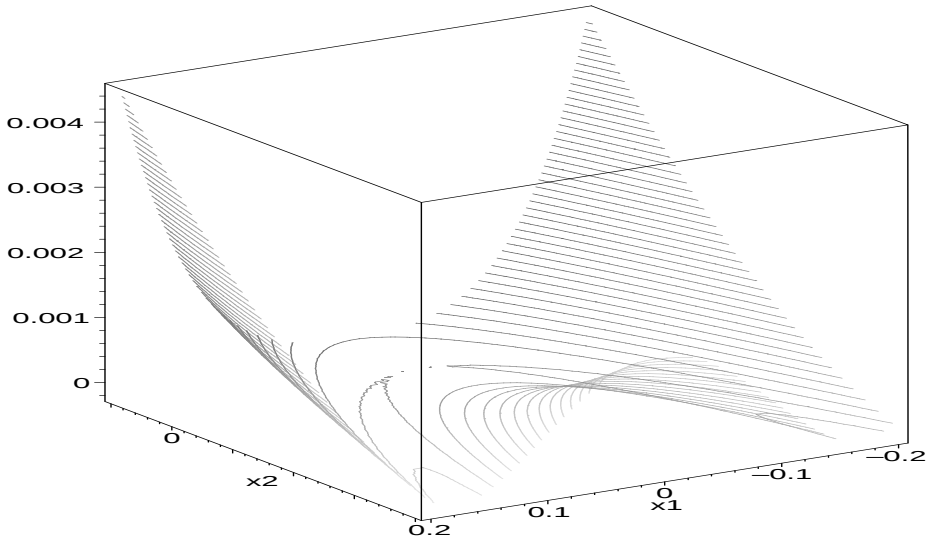
$$f(x^* + \alpha d) = \alpha^2(d_2 - \alpha d_1^2)(d_2 - 2\alpha d_1^2) > 0 = f(x^*) \quad \forall \alpha \in]0, \bar{\alpha}[$$

mit

$$\bar{\alpha} = \begin{cases} \frac{d_2}{2d_1^2} & \text{falls } d_1 \neq 0 \wedge d_2 > 0 \\ +\infty & \text{sonst} \end{cases}$$

Somit besitzt diese Funktion keine Abstiegsrichtung. Trotzdem ist x^* kein lokales Minimum von f, denn

$$f(\varepsilon, \frac{3}{2}\varepsilon^2) = -\frac{1}{4}\varepsilon^4 < 0 \quad \forall \varepsilon \neq 0$$



Im Falle konvexer Optimierungsprobleme der Form (1) ist die Bedingung aus Satz 3.13 notwendig und hinreichend:

Satz 3.17
Sei $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *eine konvexe Funktion,* $C \subseteq \mathbb{R}^n$ *eine konvexe Menge und es sei* $x^0 \in C$.
x^0 *ist lokales (und damit globales) Minimum von* f *auf* C *genau dann, wenn* $F_C(x^0) \cap A_f(x^0) = \emptyset$.

Beweis: ÜA
"$\Longrightarrow$" Satz 3.13
"$\Longleftarrow$" Nehmen wir an, x^0 sei nicht optimal. Dann gibt es einen Vektor $d \neq 0$, so daß

$$x^0 + d \in C \quad \text{und} \quad f(x^0 + d) < f(x^0)$$

Aus der Konvexität von C und f folgt dann

$$x^0 + \alpha d \in C \quad \text{und} \quad f(x^0 + \alpha d) < f(x^0) \quad \forall \alpha \in]0, 1]$$

und somit $d \in F_C(x^0) \cap A_f(x^0)$ im Widerspruch zur Voraussetzung.
Oder etwas abstrakter:
Lemma 3.11 *iv)* liefert unter den Voraussetzungen des Satzes
$C \subseteq \{x^0\} \oplus F_C(x^0)$ und (mit Bemerkung 3.12) auch $M(x^0, f) \subseteq \{x^0\} \oplus A_f(x^0)$, somit folgt aus $F_M(x^0) \cap A_f(x^0) = \emptyset$, daß $C \cap M(x^0, f) = \emptyset$, also x^0 (globales) Minimum von f auf C. □

3.2 Sattelpunkte, Fritz John und Karush-Kuhn-Tucker-Bedingungen

In diesem Abschnitt spezialisieren wir nun die betrachteten Probleme, indem wir voraussetzen, daß die Restriktionsmenge als Durchschnitt von Niveaumengen von Funktionen gegeben ist. Mit der sich damit ergebenden Struktur des Problems lassen sich natürlich weitergehende Aussagen gewinnen.

Wir betrachten nun ein Optimierungsproblem in der Form

$$\begin{gathered} \min\{f(x):\ x\in M\} \\ \text{mit}\quad M=\{x\in\mathbb{R}^n:\ g_i(x)\le 0\ \ i=1,\dots,m,\ \ h_j(x)=0,\ \ j=1,\dots l\} \end{gathered} \tag{3}$$

Dabei seien $m,l\in\mathbb{N}$ und $f,\ g_i\ \ i=1,\dots,m,\ h_j\ \ j=1,\dots,l$ reellwertige Funktionen über dem $\mathbb{R}^n$.

Der Schwerpunkt unseres Interesses liegt auf *konvexen Optimierungsproblemen*, das sind solche Probleme der Form (3), bei denen $f,\ g_i\ \ i=1,\dots,m$ konvexe Funktionen und alle h_j affin sind. Für die Restriktionsmenge gilt $M=\bigcap\limits_{i=1}^{m} L_{g_i}(0)\cap\bigcap\limits_{i=1}^{l}\{x\in\mathbb{R}^n:\ h_j(x)=0\}$ und somit ist sie im Falle konvexer g_i und affiner h_j konvex als Durchschnitt konvexer Mengen. Beim Vorhandensein nichtlinearer Gleichungen würde im allgemeinen die Konvexität der Restriktionsmenge nicht mehr garantiert sein.

Die hierfür entwickelte Theorie stellt eine Verallgemeinerung der für Minimierungsprobleme unter Gleichungsrestriktionen entwickelten Methode der Lagrange-Multiplikatoren (vgl. Satz 1.7) auf Probleme mit Ungleichungsrestriktionen dar.

In völliger Analogie zum klassischen Fall von Gleichungsnebenbedingungen ordnen wir den Restriktionen einen Vektor von *Lagrange-Multiplikatoren* $(u,v)\in\mathbb{R}^{m+l}$ zu und definieren die *Lagrange-Funktion* $\Phi:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^l\longrightarrow\mathbb{R}$ zum Problem (3) als

$$\Phi(x,u,v)=f(x)+\sum_{i=1}^{m}u_ig_i(x)+\sum_{i=1}^{l}v_jh_j(x)=f(x)+u^Tg(x)+v^Th(x)$$

mit den vektorwertigen Funktionen $g:\mathbb{R}^n\longrightarrow\mathbb{R}^m,\ h:\mathbb{R}^n\longrightarrow\mathbb{R}^l$. Für das klassische Problem wurden nun alle Lagrange-Multiplikatoren im $\mathbb{R}^{m+l}$ zugelassen. Dies entspricht anschaulich einer Bestrafung der Verletzung der Gleichung durch einen Lagrange-Multiplikator mit passendem Vorzeichen. So wird bei der Supremumsbildung bezüglich der Lagrange-Multiplikatoren die Zulässigkeit erzwungen.

Wir haben es hier mit Ungleichungen $g_i(x)\le 0$ als Restriktionen zu tun, eine "Abweichung nach unten" des Wertes $g_i(x)$ ist also nicht schädlich für die Zulässigkeit des Punktes. Demzufolge sind nur positive Werte der Ungleichungs-Restriktionsfunktionen "zu bestrafen", deshalb sind nur nichtnegative Lagrange-Multiplikatoren u von Interesse. Um die Formulierung im weiteren zu vereinfachen, verändern wir *Rockafellar* folgend die Definition der Funktion Φ zu einer erweitert-reellwertigen Funktion $L:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^l\longrightarrow\bar{\mathbb{R}}$ mit der Festlegung

$$L(x,u,v)=\begin{cases} f(x)+u^Tg(x)+v^Th(x) & \text{falls}\ \ u_i\ge 0\quad i=1,\dots,m \\ -\infty & \text{sonst}\end{cases}$$

Diese Funktion ist für konvexe Probleme konvex in x für festes u und affin über dem Endlichkeitsbereich und damit insgesamt konkav als Funktion von u für festes x, sie ist affin in v für festes x.

Wir formulieren nun unser Problem (3) mit Hilfe der Lagrange-Funktion um.

3.2.1 Sattelpunktbedingungen

Definition 3.18 *Der Punkt $(x^0,u^0,v^0)^T\in\mathbb{R}^{n+m+l}$ heißt Sattelpunkt von L, wenn*

$$L(x^0,u,v)\le L(x^0,u^0,v^0)\le L(x,u^0,v^0)\quad\forall x\in\mathbb{R}^n\ \forall u\in\mathbb{R}^m\ \forall v\in\mathbb{R}^l$$

Ohne Konvexität erhalten wir eine hinreichende Optimalitätsbedingung.

Satz 3.19 *Ist $(x^0, u^0, v^0)^T \in \mathbb{R}^{n+m}$ Sattelpunkt von L, dann ist x^0 optimal für das Problem (3).*

Beweis:

i) $x^0 \in M$: Zunächst ist $u^0 \geq 0$ (komponentenweise), andernfalls wäre $L(x^0, u^0, v^0) = -\infty$ und der Punkt könnte nicht Sattelpunkt sein, da $L(x^0, 0, 0) = f(x^0) > L(x^0, u^0, v^0)$. Somit gilt

$$f(x^0) + u^T g(x^0) + v^T h(x^0) = L(x^0, u, v) \leq L(x^0, u^0, v^0) = f(x^0) + u^{0T} g(x^0) + v^{0T} h(x^0) \quad \forall u \geq 0,\ \forall v$$

Dies bedeutet

$$\sum_{i=1}^{m} (u_i - u_i^0) g_i(x^0) + \sum_{j=1}^{l} (v_j - v_j^0) h_j(x^0) \leq 0 \quad \forall u \geq 0\ \forall v$$

und daraus folgt $g_i(x^0) \leq 0 \quad i = 1, \ldots, m$ und $h_j(x^0) = 0 \quad j = 1, \ldots . l.$

ii) $f(x^0) \leq f(x) \quad \forall x \in M$: Betrachten wir speziell $u = 0$, so bedeutet die eben abgeleitete Ungleichung $\sum\limits_{i=1}^{m} u_i^0 g_i(x^0) \geq 0$ und wegen $g_i(x^0) \leq 0,\ u_i^0 \geq 0 \quad i = 1, \ldots, m$ folgt daraus $\sum\limits_{i=1}^{m} u_i^0 g_i(x^0) = 0$. Nun benutzen wir die Ungleichung bezüglich x in der Sattelpunktbedingung.

$$\begin{aligned} f(x^0) &= f(x^0) + \underbrace{u^{0T} g(x^0)}_{=0} + \underbrace{v^{0T} h(x^0)}_{=0} \\ &= L(x^0, u^0, v^0) \\ &\leq L(x, u^0, v^0) && \forall x \in \mathbb{R}^n \\ &= f(x) + \underbrace{u^{0T} g(x)}_{\leq 0} + \underbrace{v^{0T} h(x)}_{=0} \leq f(x) && \forall x \in M \end{aligned}$$

□

Dieser Satz ist selbst für konvexe Probleme ohne zusätzliche Voraussetzungen nicht umkehrbar, wie folgendes Beispiel zeigt:

$$\min\{-x :\ x \in M\} \quad \text{mit} \quad M = \{x \in \mathbb{R} :\ x^2 \leq 0\}$$

Es ist $M = \{0\}$, somit ist dieser einzige zulässige Punkt optimal. Existiert nun ein $u^0 \geq 0$, so daß $(0, u^0)^T$ Sattelpunkt für die Lagrange-Funktion ist? Es ist $L(0, u^0) = 0$. Nun müßte aber gelten $0 = L(0, u^0) \leq L(x, u^0) = -x + u^0 x^2 \quad \forall x \in \mathbb{R}$. Diese Ungleichung ist im Falle $u^0 = 0$ für jedes positive x verletzt, für $u^0 > 0$ ist $L(\frac{1}{2u^0}, u^0) = -\frac{1}{4u^0} < 0$.

Satz 3.20 *Das Problem (3) sei ein konvexes Optimierungsproblem (f, g_i konvex, h_j affin $\forall i, j$) und es sei die Slater-Bedingung erfüllt: $\exists \tilde{x}$ mit $g_i(\tilde{x}) < 0,\ \ i = 1, \ldots, m,\ \ h_j(\tilde{x}) = 0\ \ j = 1. \ldots, l.$*
Dann gilt:
Ist x^0 optimal für das Problem (3), so existiert ein Vektor von Lagrange-Multiplikatoren $(u^0, v^0)^T$, so daß $(x^0, u^0, v^0)^T \in \mathbb{R}^{n+m+l}$ Sattelpunkt von L ist.

Beweis: Zunächst sei o.B.d.A. das System der affinen Funktionen h_j nicht redundant (Zeilen-Rang der Koeffizientenmatrix voll), sonst können die redundanten Gleichungen ohne Veränderung des Problems aus der Beschreibung gelöscht werden.
Wir konstruieren die Lagrange-Multiplikatoren mit Hilfe des Trennungssatzes. Dazu betrachten wir die Mengen

$$M_1 := \left\{ \begin{pmatrix} y \\ z \\ \mu \end{pmatrix} \in \mathbb{R}^{m+l+1} :\ \exists x \in \mathbb{R}^n \begin{array}{ll} y_i \geq g_i(x) & i = 1, \ldots, m \\ z_j = h(x) & j = 1, \ldots, l \\ \mu \geq f(x) & \end{array} \right\}$$

und

$$M_2 := \left\{ \begin{pmatrix} y \\ z \\ \mu \end{pmatrix} \in \mathbb{R}^{m+l+1} : \begin{array}{ll} y_i \leq 0 & i = 1, \ldots, m \\ z_j = 0 & j = 1, \ldots, l \\ \mu \leq f(x^0) & \end{array} \right\}$$

Beide Mengen sind nichtleer. Konvexität des Problems liefert die Konvexität von M_1, M_2 ist als unendlicher Quader ebenfalls konvex. Es ist $M_1 \cap ri\, M_2 = \emptyset$, denn andernfalls würden $x^*,\ y^*,\ z^*,\ \mu^*$ existieren mit

$$0 > y_i^* \geq g_i(x^*)\ \ i = 1, \ldots, m, \quad 0 = z_j^* = h_j(x^*)\ \ j = 1 \ldots, l, \quad f(x^0) > \mu^* \geq f(x^*)$$

und damit wäre x^* zulässig für (3) mit einem kleineren Zielfunktionswert im Widerspruch zur vorausgesetzten Optimalität von x^0.
Es existiert also nach dem Trennungssatz 2.37 eine echte trennende Hyperebene, d.h. ein Vektor $a = (u, v, \alpha)^T \in \mathbb{R}^{m+l+1} \neq 0$ und ein $\beta \in \mathbb{R}$ mit $M_1 \subseteq \{b \in \mathbb{R}^{m+l+1} : a^T b \geq \beta\}$ und $M_2 \subseteq \{b \in \mathbb{R}^{m+l+1} : a^T b \leq \beta\}$ und die Vereinigung beider Mengen ist nicht in der Hyperebene enthalten.
Es gilt $u \geq 0,\ \alpha \geq 0$, denn die entsprechenden Komponenten y, μ von Punkten in der Menge M_2 können beliebig klein gewählt werden.
Wir wählen nun für ein beliebiges $x \in \mathbb{R}^n$ insbesondere $b^1 = (g(x), h(x), f(x))^T \in M_1$ und $b^2 = (0, 0, f(x^0))^T \in M_2$. Es gilt also

$$u^T g(x) + v^T h(x) + \alpha f(x) \geq \beta \geq \alpha f(x^0) \quad \forall x \in \mathbb{R}^n \qquad (*)$$

Nun zeigen wir, daß $\alpha > 0$ gilt. Angenommen, es sei $\alpha = 0$. Dann würde aus der obigen Beziehung folgen $u^T g(x) + v^T h(x) \geq 0 \quad \forall x \in \mathbb{R}^n$. Nach der Slater-Bedingung existiert aber ein $\tilde{x}$ mit $g_i(\tilde{x}) < 0, \quad i = 1, \ldots, m, \quad h_j(\tilde{x}) = 0 \quad j = 1. \ldots, l$, woraus wegen $u \geq 0$ folgt $u = 0$. Damit würde gelten $v^T h(x) \geq 0 \quad \forall x \in \mathbb{R}^n$. Die linke Seite ist wegen der Affinität von h eine affine Funktion und somit müßte sie konstant in x sein. Da sie im Slaterpunkt $\tilde{x}$ den Wert 0 annimmt, wäre sie also identisch Null. Aufgrund der Voraussetzung, dass der Rang der Koeffizientenmatrix von h voll sei, würde dies nur für verschwindende Lagrange-Multiplikatoren $v = 0$ möglich sein. Insgesamt wäre also der Vektor a der Nullvektor im Widerspruch dazu, daß er eine Hyperebene definiert.
Nun definieren wir $u^0 = \frac{1}{\alpha} u, \quad v^0 = \frac{1}{\alpha} v$ und erhalten aus der Beziehung (*)

$$L(x, u^0, v^0) = {u^0}^T g(x) + {v^0}^T h(x) + f(x) \geq f(x^0) \quad \forall x \in \mathbb{R}^n \qquad (**)$$

Setzen wir in dieser Beziehung $x = x^0$, so erhalten wir ${u^0}^T g(x^0) \geq 0$. Andererseits ist $g(x^0) \leq 0, \quad u^0 \geq 0$, also insgesamt ${u^0}^T g(x^0) = 0$. Damit ist $f(x^0) = L(x^0, u^0, v^0)$ und die Beziehung (**) liefert

$$L(x, u^0, v^0) \geq L(x^0, u^0, v^0) \quad \forall x \in \mathbb{R}^n$$

Wegen der Zulässigkeit von x^0 ist $u^T g(x^0) \leq 0 \ \forall u \geq 0$ und $h(x^0) = 0$, also

$$L(x^0, u^0, v^0) = f(x^0) \geq f(x^0) + u^T g(x^0) + v^T h(x^0) = L(x^0, u, v) \quad \forall u \geq 0,\ \forall v$$

und da L als $-\infty$ festgesetzt wurde für Lagrange-Multiplikatoren, welche nicht die Bedingung $u \geq 0$ erfüllen, ist die Sattelpunkteigenschaft gezeigt. □

Im Beweis haben wir auch gezeigt, daß für den Vektor der Lagrange-Multiplikatoren zu Ungleichungen die Komplementaritätsbeziehung ${u^0}^T g(x^0) = 0$ gilt, daß also nur Lagrange-Multiplikatoren zu sogenannten aktiven Ungleichungen, d.h. solchen, die vom Punkt x^0 als Gleichungen erfüllt sind, positiv sein können.
Die Indexmenge der aktiven Ungleichungen $I(x^0) := \{i \in \{1, \ldots, m\} : g_i(x^0) = 0\}$ spielt in theoretischen Aussagen und auch in einer ganzen Reihe von Algorithmen eine herausragende Rolle (Stichwort: Aktive-Indexmengen-Strategie. Betrachte die im aktuellen Iterationspunkt aktiven Restriktionen (d.h. die aktiven Ungeichungen und die Gleichungen) als Gleichungsrestriktionen, optimiere über der Mannigfaltigkeit. Das erfolgt meist mittels Prädiktor-Korrektor-Verfahren. Dazu

wird von einem Iterationspunkt aus entlang einer Abstiegsrichtung eine Schrittweite mit hinreichendem Abstieg gesucht. Wird dabei eine der inaktiven Ungleichungen verletzt, so bestimme Schnittpunkt und erweitere die Menge der aktiven Restriktionen. Wechselt ein Lagrange-Multiplikator zu Ungleichungen das Vorzeichen, so bestimme Nullstelle und lasse entspr. Ungleichung aus Menge der aktiven Restriktionen weg.).

Es ist auch anschaulich klar, daß Restriktionen, welche im Optimalpunkt als strenge Ungleichung gelten, die Optimalität des Punktes nicht beeinflussen. Dies wird im folgenden Satz ausgedrückt.

Satz 3.21 *(3) sei ein konvexes Optimierungsproblem, d.h.* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *konvex,* $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *konvex für alle* i, $h_j : \mathbb{R}^n \longrightarrow \mathbb{R}$ *affin für alle* j. *Ist* x^0 *optimal für das Problem (3), so ist dieser Punkt auch optimal für das Problem*

$$(3^*) \qquad \min\{f(x) : \ x \in M^*\}$$

mit

$$M^* = \{x \in \mathbb{R}^n : \ \begin{array}{ll} g_i(x) \leq 0 & i \in I(x^0) \\ h_j(x) = 0 & j = 1, \ldots, l \end{array} \}$$

und $I(x^0) := \{i \in \{1, \ldots, m\} : \ g_i(x^0) = 0\}$.

Beweis: ÜA
Angenommen, es existiere ein $x^1 \in M^*$ mit $f(x^0) > f(x^1)$. Es ist $M \subseteq M^*$ und M^* konvex. Aus $x^0, x^1 \in M^*$ folgt somit $x^\lambda = x^0 + \lambda(x^1 - x^0) \in M^* \quad \forall \lambda \in [0,1]$. Aus Konvexität der g_i und Affinität der h_j folgt sofort, daß $g_i(x^\lambda) \leq 0 \quad i \in I(x^0)$ und $h_j(x^\lambda) = 0 \quad j = 1, \ldots, l \ \forall \lambda \in [0,1]$. Die reellwertigen konvexen Funktionen g_i sind stetig, es existiert somit eine Umgebung um x^0, für welche die Werte dieser Funktionen für $i \in \{1, \ldots, m\} \setminus I(x^0)$ negativ bleiben. Da es endlich viele Funktionen sind, existiert ein $\lambda_0 \in]0,1]$, so daß $x^{\lambda_0} \in M$.
Es ist aber $f(x^{\lambda_0}) \leq (1 - \lambda_0) f(x^0) + \lambda_0 f(x^1) < f(x^0)$ im Widerspruch zur Optimalität von x^0 für das Problem (3). □

3.2.2 Lagrange-Dualität

Bereits in der Motivation für die Formulierung der Lagrange-Funktion war die Rede von der Supremumsbildung bezüglich der Lagrange-Multiplikatoren. Hier soll kurz der darauf beruhende Dualitätsansatz der konvexen Optimierung und seine Beziehung zur Sattelpunktbedingung erläutert werden (dies ist nicht der einzige Vorschlag für die Konstruktion von Dualproblemen, die Fenchel-konjugierten Funktionen liefern einen anderen Ansatz, siehe z.B. *R.T. Rockafellar*).
Das Problem (3) wollen wir als *Primalproblem* bezeichnen. Das zugehörige *Lagrange-Dualproblem* ist

$$\max\{\Theta(u,v) : \ u \geq 0\} \tag{4}$$

mit

$$\Theta(u,v) = \inf\{L(x,u,v) : \ x \in \mathbb{R}^n\}$$

Die Zielfunktion des Lagrange-Dualproblems $\Theta(u,v)$ ist definiert durch ein Infimum bezüglich x und kann für eine Menge von Vektoren von Lagrange-Multiplikatoren den Wert $-\infty$ annehmen. Das Dualproblem besteht in der Maximierung dieses Infimums bezüglich der Menge $\mathbb{R}^m_+ \times \mathbb{R}^l$ und wird deshalb auch als max-min-Problem bezeichnet. Man kann diesen Ansatz verfeinern, indem nicht alle Restriktionsfunktionen in die Lagrange-Funktion aufgenommen werden und die übrigen (einfacheren) eine Menge X definieren, welche im Problem zur Bestimmung von Θ als Restriktionsmenge auftritt. Der Sattelpunktsatz läßt sich auch mit solch zusätzlicher Menge beweisen, der Einfachheit der Schreibweise halber wurden hier stets alle Restriktionen in die Lagrange-Funktion aufgenommen.

Das erste Ergebnis zur Dualität ist die sogenannte schwache Dualität, welche besagt, daß ein zulässiger Punkt des Dualproblems stets eine untere Schranke für den Zielfunktionswert zulässiger Punkte des Primalproblems liefert.

Satz 3.22 *Sei x ein zulässiger Punkt für Problem (3) (d.h. $x \in M$) und sei $(u,v)^T \in \mathbb{R}^{m+l}$ zulässig für (4) (d.h. $u \geq 0$). Dann gilt $f(x) \geq \Theta(u,v)$.*

Beweis: $\Theta(u,v) = \inf\{L(y,u,v) : \ y \in \mathbb{R}^n\} \leq f(x) + \underbrace{u^T g(x)}_{\leq 0} + \underbrace{v^T h(x)}_{=0} \leq f(x)$,
da $u \geq 0,\ g(x) \leq 0,\ h(x) = 0$. □

Folgerung 3.23 $\inf\{f(x) : \ x \in M\} \geq \sup\{\Theta(u,v) : \ u \geq 0\}$

Folgerung 3.24 *Existieren $x^0 \in M,\ u^0 \geq 0,\ v^0 \in \mathbb{R}^l$ mit $f(x^0) = \Theta(u^0,v^0)$, so ist x^0 optimal für (3) und (u^0,v^0) optimal für (4).*

Folgerung 3.25 *Ist $\sup\{\Theta(u,v) : \ u \geq 0\} = +\infty$, so ist $M = \emptyset$.*

Bei nichtkonvexen Problemen kann eine sog. *Dualitätslücke* auftreten, d.h. der optimale Zielfunktionswert des Primalproblems ist größer als der des Dualproblems.

Selbst bei konvexen lösbaren Problemen ist nicht garantiert, daß die Situation von Folgerung 3.24 vorliegt, d.h. das Dualproblem ebenfalls lösbar ist. Diese Situation ist, wie wir gleich sehen werden, gegeben, wenn die Existenz eines Sattelpunktes garantiert ist. Hierfür bedarf es einer zusätzlichen *Regularitätsbedingung* wie der *Slater-Bedingung*, wie wir bereits im vorigen Abschnitt gesehen haben.
Wir wollen das Beispiel vor Satz 3.20 noch einmal im Hinblick auf die Eigenschaften des Dualproblems interpretieren: $\min\{-x : \ x \in C\}$ mit $C = \{x \in \mathbb{R} : \ x^2 \leq 0\}$
Bestimmen wir nun die Zielfunktion $\Theta(u)$ des Dualproblems: Es ist $L(x,u) = -x + ux^2$, diese Funktion ist linear für $u = 0$ und streng konvex mit einem (globalen) Minimum in $x = \frac{1}{2u}$ für $u > 0$. Daraus ergibt sich

$$\Theta(u) = \inf\{-x + ux^2 : \ x \in \mathbb{R}\} = \begin{cases} -\infty & u = 0 \\ -\frac{1}{4u} & u > 0 \end{cases}$$

und es ist klar, daß das Supremum der Funktion $\Theta(u)$ über alle $u \geq 0$ gleich dem optimalen primalen Zielfunktionswert (Null) ist, jedoch kein u existiert, in welchem dieser Wert angenommen wird.

Satz 3.26 *Ein Punkt* $(x^0, u^0, v^0)^T \in \mathbb{R}^{n+m+l}$ *ist ein Sattelpunkt von* $L(x,u,v)$ *genau dann, wenn* x^0 *Optimalpunkt für (3) ist,* $(u^0, v^0)^T$ *optimal für (4) ist und* $f(x^0) = \Theta(u^0, v^0)$ *gilt.*

Beweis:
"$\Longrightarrow$" $(x^0, u^0, v^0)^T \in \mathbb{R}^{n+m+l}$ sei Sattelpunkt von L. Im Beweis von Satz 3.19 hatten wir dann gezeigt, daß x^0 optimal für (3) ist und daß $u^0 \geq 0$, also $(u^0, v^0)^T$ zulässig für (4). Außerdem ergab sich $f(x^0) = L(x^0, u^0, v^0)$ und die Sattelpunkteigenschaft liefert $L(x^0, u^0, v^0) \leq L(x, u^0, v^0) \quad \forall x \in \mathbb{R}^n$, also $L(x^0, u^0, v^0) = \Theta(u^0, v^0)$. Folgerung 3.24 liefert dann Optimalität von (u^0, v^0) für das Dualproblem.
"$\Longleftarrow$" Seien nun x^0 optimal für (3), (u^0, v^0) optimal für (4) und es gelte $f(x^0) = \Theta(u^0, v^0)$. Wir haben wegen der primalen und dualen Zulässigkeit $u^0 \geq 0$ und (wie im Beweis von Satz 3.22)
$\Theta(u^0, v^0) = \inf\{L(y, u^0, v^0) : \ y \in \mathbb{R}^n\} \leq f(x^0) + \underbrace{{u^0}^T g(x^0)}_{\leq 0} + \underbrace{{v^0}^T h(x^0)}_{=0} \leq f(x^0).$
Da nun nach Voraussetzung Gleichheit gilt, folgt ${u^0}^T g(x^0) = 0$. Somit gilt
$L(x^0, u^0, v^0) = f(x^0) = \Theta(u^0, v^0) = \inf\{L(x, u^0, v^0) : \ x \in \mathbb{R}^n\} \leq L(x, u^0, v^0) \quad \forall x \in \mathbb{R}^n.$
Weiterhin ist
$L(x^0, u^0, v^0) = f(x^0) \geq f(x^0) + \underbrace{u^T g(x^0)}_{\leq 0} + \underbrace{v^T h(x^0)}_{=0} = L(x^0, u, v) \quad \forall u \geq 0 \ \forall v \in \mathbb{R}^l,$
insgesamt gilt also die Sattelpunkteigenschaft. □

Das Vorliegen eines Sattelpunktes liefert also starke Dualität, d.h.
(3) lösbar $\Longrightarrow$ (4) lösbar und optimale Zielfunktionswerte sind gleich.

3.2.3 Fritz John Bedingungen

Für den Fall differenzierbarer Funktionen in der Beschreibung von (3) lassen sich die geometrischen Bedingungen durch Bedingungen an die Gradienten ausdrücken. Diese Optimalitätsbedingungen wurden von Fritz John 1948 veröffentlicht.
Zum Beweis brauchen wir einen Alternativsatz, welcher auch im Teil Optimierung I eine Rolle spielt und dort im Zusammenhang mit dem Lemma von Farkas bewiesen wird. Dieses Lemma soll hier direkt mit dem Trennungssatz bewiesen werden.

Lemma 3.27 (Lemma von Gordan)
Es sei A *eine* $k \times n$*-Matrix.*
Dann hat genau eines der beiden Systeme

$$(I) \qquad Ax < 0, \ x \in \mathbb{R}^n$$

und

$$(II) \qquad A^T y = 0, \quad y \in \mathbb{R}^k, \ y \geq 0, \ y \neq 0$$

eine Lösung.

Beweis: Zunächst kann System (II) nicht lösbar sein, wenn System (I) lösbar ist. Angenommen, $\hat{x}$ sei Lösung von (I). Dann folgt aus $A\hat{x} < 0$ daß $y^T A\hat{x} < 0$ für alle y, welche die Bedingungen $y \geq 0, \ y \neq 0$ erfüllen, d.h. es kann nicht $A^T y = 0$ gelten.
Das System (I) sein nun unlösbar. Dann sind die beiden nichtleeren konvexen Mengen

$$C_1 = \{z \in \mathbb{R}^k : \ z = Ax, \ x \in \mathbb{R}^n\} \ \text{ und } \ C_2 = \{z \in \mathbb{R}^k : \ z < 0\}$$

disjunkt. Es existiert also eine trennende Hyperebene für C_1 und $cl\ C_2$, d.h. ein Vektor $y \neq 0$ mit

$$y^T Ax \geq y^T z \quad \forall x \in \mathbb{R}^n \ \forall z \in cl\ C_2$$

Da jede Komponente eines Vektors in C_2 betragsmäßig beliebig große negative Werte annehmen kann, muß $y \geq 0$ gelten. Betrachten wir $z = 0 \in cl\ C_2$, so folgt $y^T Ax \geq 0 \quad \forall x \in \mathbb{R}^n$. Wählt man nun speziell $x = -A^T y$, so ergibt sich $-\|A^T y\| \geq 0$, also $A^T y = 0$. Somit ist y Lösung von (II). □

Satz 3.28 (notwendige Fritz John Bedingung)
Es sei $\bar{x}$ zulässig für (3) und die Indexmenge der in $\bar{x}$ aktiven Ungleichungen sei bezeichnet mit $I(\bar{x}) = \{i \in \{1,\dots,m\} : g_i(\bar{x}) = 0\}$. f, g_i seien differenzierbar in $\bar{x}$ für alle $i \in I(\bar{x})$, g_i stetig in $\bar{x}$ für alle $i \in \{1,\dots,m\} \setminus I(\bar{x})$ und h_j seien stetig differenzierbar für alle $j \in \{1,\dots,l\}$. Wenn $\bar{x}$ ein lokales Optimum für (3) ist, so existieren $\alpha \in \mathbb{R}$, $u \in \mathbb{R}^{|I(\bar{x})|}$ und $v \in \mathbb{R}^l$ mit

$$\begin{aligned} \alpha\nabla f(\bar{x}) + \sum_{i\in I(\bar{x})} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) &= 0 \\ \alpha &\geq 0 \\ u_i &\geq 0 \quad \forall i \in I(\bar{x}) \\ (\alpha, u, v)^T &\neq 0 \end{aligned}$$

Beweis:
Im Falle $l = 0$, d.h. beim Fehlen von Gleichungsrestriktionen ist die Aussage des Satzes leicht aus den bisher bewiesenen Sätzen und dem *Lemma von Gordan* ableitbar:
Die notwendige Optimalitätsbedingung Satz 3.13 liefert, daß es keinen Vektor gibt, welcher gleichzeitig zulässige Richtung der Restriktionsmenge und Abstiegsrichtung der Zielfunktion ist. Diese Aussage wenden wir auf das reduzierte Problem von Satz 3.21 an, dessen Restriktionsmenge in diesem Falle die Form

$$\bigcap_{i\in I(\bar{x})} \{x \in \mathbb{R}^n : \; g_i(x) \leq g_i(\bar{x})\}$$

hat. Benutzung von Satz 3.14 liefert, daß das System

$$\begin{aligned} \nabla f(\bar{x})d &< 0 \\ \nabla g_i(\bar{x})d &< 0 \quad i \in I(\bar{x}) \end{aligned}$$

unlösbar ist, Lemma 3.27 liefert dann die Behauptung in diesem Spezialfall.

Nun betrachten wir das Problem mit Gleichungsrestriktionen. Sind die Gradientenvektoren $\nabla h_j(\bar{x}),\ j = 1,\dots,l$ linear abhängig, so ist die Behauptung mit $\alpha = 0,\ u_i = 0, i \in I(\bar{x})$ erfüllbar. Seien die Gradienten der Gleichungsrestriktionen in $\bar{x}$ linear unabhängig. Dann ist das System

$$\begin{aligned} \nabla f(\bar{x})d &< 0 \\ \nabla g_i(\bar{x})d &< 0, \quad i \in I(\bar{x}) \\ \nabla h_j(\bar{x})d &= 0, \quad j = 1,\dots,l \end{aligned} \tag{5}$$

unlösbar. Angenommen, es existiere $\bar{d} \in \mathbb{R}^n$, welches das System (5) löst. Dann läßt sich mittels des Differentialgleichungssystems

$$x'(t) = P(x(t))\bar{d}, \;\; x(0) = \bar{x}$$

eine C^1-Kurve von zulässigen Punkten $x(t),\ t \in [-\varepsilon, \varepsilon]$ konstruieren. Dabei bezeichnet $P(x(t))$ die Matrix, welche jeden Vektor orthogonal auf den Nullraum der Jacobimatrix $\nabla_x h(x(t))$ projiziert (für eine $k \times n$-Matrix A mit Rang k tut dies $(I - A^T(AA^T)^{-1}A)$). $P(x(t))$ ist wegen der vorausgesetzten stetigen Differenzierbarkeit der h_j stetig in t. Der Existenzsatz von Cauchy/Peano liefert die Existenz der Kurve. Es gilt $x'(0) = P(\bar{x})\bar{d} = \bar{d}$, da nach Annahme $\bar{d}$ im Nullraum von $\nabla_x h(x(t))$ liegt. Damit ist nach Kettenregel

$$\frac{d}{dt} g_i(x(0)) = \nabla g_i(\bar{x})P(\bar{x})\bar{d} = \nabla g_i(\bar{x})\bar{d} < 0 \quad \forall i \in I(\bar{x})$$

Daraus folgt $g_i\left(x(t)\right) < 0 \quad i \in I(\bar{x})$ für $t > 0$ hinreichend klein. Wegen der Stetigkeit der g_i und $g_i(\bar{x}) < 0$ für $i \notin I(\bar{x})$ gilt die Ungleichung auch für $x(t)$ mit hinreichend kleinen t. Es bleibt zu zeigen, daß die Gleichungsrestriktionen entlang der Kurve erfüllt bleiben. Dies liefert der Mittelwertsatz zusammen mit der Kettenregel: Es gilt

$$h_j\left(x(t)\right) = h_j(\bar{x}) + t\frac{d}{dt} h_j(x(\mu)) = t\frac{d}{dt} h_j(x(\mu))$$

für ein $\mu \in]0, t[$ und

$$\frac{d}{dt} h_j(x(\mu)) = \nabla h_j(x(\mu))P(x(\mu))\bar{d}$$

Da nach Konstruktion $P(x(\mu))\bar{d}$ im Nullraum von $\nabla h_j(x(\mu))$ liegt, folgt $h_j\left(x(t)\right) = 0$. Dies gilt für jedes j, wir haben also Zulässigkeit von $x(t)$ für $t \in [0, \varepsilon]$. Analog zum Beweis der Zulässigkeit hinsichtlich der Ungleichungen für $i \in I(\bar{x})$ folgt $\frac{d}{dt} f(x(0)) < 0$ und somit

$$f(x(t)) < f(\bar{x}) \quad \text{für } t \text{ hinreichend klein.}$$

Das obige System besitzt also keine Lösung, da wir sonst einen Widerspruch zur lokalen Optimalität von $\bar{x}$ erhielten. Analog zum Beweis des Lemmas von Gordan liefert der Trennungssatz angewandt auf die beiden Mengen

$$C_1 = \{(z_1, z_2)^T \in \mathbb{R}^{k+l} : \; z_1 = A_1 x, \; z_2 = A_2 x, \; x \in \mathbb{R}^n\} \;\text{ und }\; C_2 = \{(z_1, z_2) : \; z_1 < 0, \; z_2 = 0\}$$

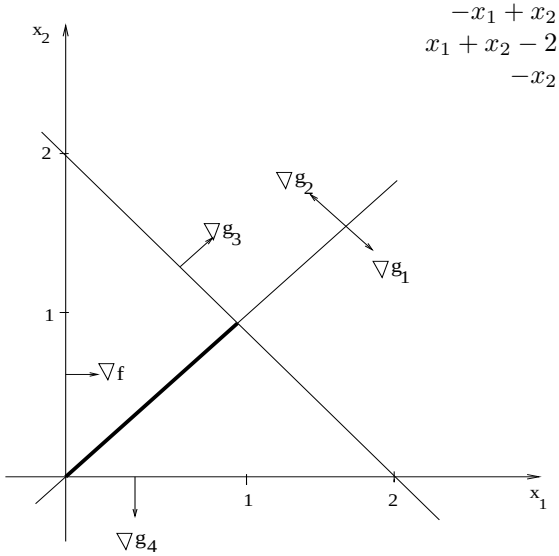
die Existenz eines Vektors $(p_1, p_2) \neq 0$ mit $p_1 \geq 0$ und $A_1^T p_1 + A_2^T p_2 = 0$. Dies angewandt auf das obige System liefert die Behauptung. □

Unter der zusätzlichen Voraussetzung der Affinität der h_j, linearer Unabhängigkeit der $\nabla h(\bar{x})$, Pseudokonvexität der Zielfunktion f und strenger Pseudokonvexität der g_i für $i \in I(\bar{x})$ auf einer Umgebung von $\bar{x}$ ist die Fritz-John-Bedingung hinreichend (siehe *Bazaraa, Sherali, Shetty* Theorem 4.3.6).

3.2.4 Karush-Kuhn-Tucker-Bedingungen

In den Friz John Bedingungen muß der Multiplikator α der Zielfunktion nicht notwendig positiv sein. Weiterhin garantiert selbst für so einfache Probleme wie lineare Optimierungsprobleme das Erfülltsein der Fritz John Bedingungen nicht die Optimalität eines zulässigen Punktes.
Beispiel:

$$\min\{x_1 : \quad \left.\begin{array}{rcl} x_1 - x_2 & \leq & 0, \\ -x_1 + x_2 & \leq & 0, \\ x_1 + x_2 - 2 & \leq & 0, \\ -x_2 & \leq & 0 \end{array}\right\}$$



In diesem Beispiel ist durch die ersten beiden Ungleichungen eine Gleichung ausgedrückt. In jedem zulässigen Punkt sind die beiden Ungleichungen $g_1(x) \leq 0, \; g_2(x) \leq 0$ aktiv, ihre Gradienten sind linear abhängig und so sind z.B. mit $\alpha = 0, \; u = (1, 1, 0, 0)^T$ für jeden zulässigen Punkt die Fritz John Bedingungen erfüllt.

Analog zu den Sattelpunktbedingungen liefert eine zusätzliche Regularitätsbedingung (*constraint qualification*) die Existenz von Lagrange-Multiplikatoren, für welche $\alpha > 0$ gilt, welches somit auf 1 normalisiert werden kann.

Obwohl auch hier analog zu Satz 3.28 die Differenzierbarkeit nur für die aktiven Restriktionen erforderlich ist, werden wir den Satz unter der schärferen Voraussetzung der Differenzierbarkeit aller Problemfunktionen formulieren.

Satz 3.29 (notwendige Karush-Kuhn-Tucker-Bedingungen)
Es sei $\bar{x}$ zulässig für (3). f, g_i seien differenzierbar in $\bar{x}$ für $i = 1, \ldots, m$ und die h_j seien stetig differenzierbar in $\bar{x}$ für alle $j \in \{1, \ldots, l\}$.
Weiterhin gelte die linear independence constraint qualification oder kurz LICQ, d.h. die Gradienten der in $\bar{x}$ aktiven Restriktionen $\nabla g_i(\bar{x})$, $i \in I(\bar{x})$, $\nabla h_j(\bar{x})$, $j = 1, \ldots, l$ seien linear unabhängig.

Wenn $\bar{x}$ lokaler Optimalpunkt für (3) ist, so existieren eindeutig bestimmte Vektoren von Lagrange-Multiplikatoren $u \in \mathbb{R}^m$, $v \in \mathbb{R}^l$ mit

$$\begin{aligned} \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) &= 0 \\ u_i g_i(\bar{x}) &= 0 \qquad i = 1, \ldots, m \\ u_i &\geq 0 \qquad i = 1, \ldots, m \end{aligned}$$

Beweis: Nach Satz 3.28 existieren Multiplikatoren α, u_i $i \in I(\bar{x})$, v_j $j = 1, \ldots, l$ mit

$$\begin{aligned} \alpha \nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) &= 0 \\ \alpha \geq 0,\ u_i &\geq 0 \quad \forall i \in I(\bar{x}) \\ (\alpha, u, v)^T &\neq 0 \end{aligned}$$

Wäre $\alpha = 0$, so hätten wir eine Darstellung des Nullvektors als nichttriviale Linearkombination aus den Gradientenvektoren der aktiven Restriktionen in $\bar{x}$, was der LICQ widerspricht. Die Gleichung kann also durch α dividiert werden. Die lineare Unabhängigkeit liefert die Eindeutigkeit dieser Lagrange-Multiplikatoren. Den inaktiven Restriktionen werden Lagrange-Multiplikatoren gleich Null zugeordnet, was durch die Komplementaritätsbedingung ausgedrückt wird. □

Die Bedingung besagt also, daß unter den Voraussetzungen des Satzes der negative Gradient der Zielfunktion als Linearkombination der Gradienten der aktiven Restriktionen ausgedrückt werden kann, wobei die Faktoren zu Ungleichungsrestriktionen nichtnegativ sind. Im Falle $l = 0$ heißt das gerade, daß der negative Gradient der Zielfunktion in dem Kegel liegt, welcher durch die Gradienten der aktiven Restriktionen aufgespannt wird. Damit sind wir im Falle konvexer Optimierungsprobleme bei einer Bedingung angelangt, die in enger Beziehung zum Satz 3.7 steht.

Im nächsten Satz wird unter verallgemeinerten Konvexitätsbedingungen gezeigt, daß die KKT-Bedingungen hinreichend sind. Hierfür reicht es aus, wenn die verallgemeinerten Bedingungen entsprechend der nachfolgenden Definition im betrachteten Punkt erfüllt sind.

Definition 3.30 *Es sei $S \subseteq \mathbb{R}^n$, $\bar{x} \in S$ und $f : S \longrightarrow \mathbb{R}$.*
f heißt quasikonvex in $\bar{x}$, wenn für alle $x \in S$ und alle $\lambda \in]0, 1[$ gilt

$$f(\lambda \bar{x} + (1 - \lambda)x) \leq \max\{f(x), f(\bar{x})\}.$$

Ist f weiterhin differenzierbar in $\bar{x}$, so heißt f pseudokonvex in $\bar{x}$, wenn für alle $x \in S$ gilt

$$\nabla f(\bar{x})(x - \bar{x}) \geq 0 \Longrightarrow f(x) \geq f(\bar{x})$$

Satz 3.31 (hinreichende Karush-Kuhn-Tucker-Bedingungen)
Es sei $\bar{x}$ zulässig für (3). f, g_i seien differenzierbar in $\bar{x}$ für $i = 1, \ldots, m$ und die h_j seien stetig differenzierbar in $\bar{x}$ für alle $j \in \{1, \ldots, l\}$.
In $\bar{x}$ seien die KKT-Bedingungen erfüllt, d.h. es existieren $u \in \mathbb{R}^m$, $v \in \mathbb{R}^l$ mit

$$\begin{aligned} \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) &= 0 \\ u_i g_i(\bar{x}) &= 0 \qquad i = 1, \ldots, m \\ u_i &\geq 0 \qquad i = 1, \ldots, m \end{aligned} \tag{6}$$

Es sei $J = \{j : v_j > 0\}$, $K = \{j : v_j < 0\}$. Sind dann f pseudokonvex in $\bar{x}$, g_i quasikonvex in $\bar{x}$ für alle $i \in I(\bar{x})$, h_j quasikonvex in $\bar{x}$ für alle $j \in J$ und h_j quasikonkav in $\bar{x}$ für alle $j \in K$, so ist $\bar{x}$ globales Minimum für (3).

Beweis: Sei x ein beliebiger zulässiger Punkt für (3). Dann gilt für alle $i \in I(\bar{x})$ $g_i(x) \leq g_i(\bar{x})$. Aus der Quasikonvexität folgt für $x_\lambda = \lambda x + (1-\lambda)\bar{x}$

$$g_i(x_\lambda) \leq \max\{g_i(x), g_i(\bar{x})\} = g_i(\bar{x}),$$

d.h. von $\bar{x}$ aus in Richtung $x - \bar{x}$ ist der Funktionswert von g_i nicht wachsend. Daraus folgt

$$\nabla g_i(\bar{x})(x - \bar{x}) \leq 0$$

Genauso folgt aus Quasikonvexität bzw. Quasikonkavität der h_j für j aus J bzw. K

$$\begin{array}{rcll} \nabla h_j(\bar{x})(x-\bar{x}) & \leq & 0 & j \in J \\ \nabla h_j(\bar{x})(x-\bar{x}) & \geq & 0 & j \in K \end{array}$$

Werden diese Ungleichungen mit den entsprechenden Lagrange-Multiplikatoren multipliziert und addiert, so erhält man

$$-\nabla f(\bar{x})(x-\bar{x}) = \left[\sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} v_j \nabla h_j(\bar{x})\right](x-\bar{x}) \leq 0$$

Die Pseudokonvexität von f in $\bar{x}$ liefert nun $f(x) \geq f(\bar{x})$. □

Aus Konvexität folgt Quasikonvexität und unter Differenzierbarkeit auch Pseudokonvexität, somit sind die verallgemeinerten Konvexitäts-Bedingungen für konvexe Optimierungsprobleme (f und die g_i konvex, h_j affin, damit konvex und konkav) erfüllt. Zu beachten ist, daß auch für konvexe Optimierungsprobleme eine Regularitätsbedingung erforderlich für die Notwendigkeit der KKT-Bedingungen ist. Als Beispiel betrachte man

$$\min\{x_1 : \; x_1^2 + (x_2-1)^2 - 1 \leq 0, \; x_1^2 + (x_2+1)^2 - 1 \leq 0\}$$

Der Nullpunkt ist der einzige zulässige Punkt, damit optimal, die KKT-Bedingungen sind jedoch nicht erfüllt.
Bei nichtkonvexen Problemen können die KKT-Bedingungen in Punkten erfüllt sein, welche nicht einmal lokale Minima sind:

$$\min\{-x_1^2 - x_2^2 : \; x_1 \leq 0\}$$

Im Nullpunkt sind mit $u = 0$ die KKT-Bedingungen erfüllt ($\nabla f(0) = 0$), der Punkt ist jedoch kein lokales Minimum, sondern globales Maximum des Problems.

Wir haben gezeigt, daß für konvexe Optimierungsprobleme unter der LICQ die KKT-Bedingungen notwendig und hinreichend für Optimalität sind. Diese Bedingungen stehen in engem Zusammenhang zur Sattelpunktbedingung, für welche wir unter einer anderen Regularitätsbedingung, der Slater-Bedingung, die Notwendigkeit gezeigt haben. Auch für die KKT-Bedingungen ist die Slater-Bedingung zusammen mit einer Konvexitätsbedingung eine mögliche Regularitätsbedingung, um Notwendigkeit zu erhalten. Dies wird im folgenden Satz (jetzt nur für den Fall konvexer Optimierungsprobleme) formuliert.

Satz 3.32
(3) sei ein konvexes Optimierungsproblem und es sei $\bar{x}$ zulässig für (3). f, g_i seien differenzierbar in $\bar{x}$ für $i = 1, \ldots, m$ und die h_j seien stetig differenzierbar in $\bar{x}$ für alle $j \in \{1, \ldots, l\}$.
Weiterhin sei die Slater-Bedingung erfüllt.
In $\bar{x}$ sind die KKT-Bedingungen erfüllt (d.h. es existieren Lagrange-Multiplikatoren u, v, so daß (6) erfüllt ist) genau dann, wenn $\bar{x}$ ist globales Optimum von (3).

Beweis: Satz 3.31 liefert die hinreichende Richtung unter der Konvexitätsvoraussetzung.
Satz 3.20 liefert die Sattelpunktbedingung unter den Voraussetzungen des Satzes. Den Beweis der notwendigen Richtung führen wir, indem wir zeigen, daß die KKT-Bedingungen aus der Sattelpunktbedingung folgen. Es sei $(\bar{x}, \bar{u}, \bar{v})^T$ ein Sattelpunkt von $L(x, u, v)$. Wir hatten im Beweis von Satz 3.19 gezeigt, daß dann $\bar{u} \geq 0$ und $\bar{u}^T g(\bar{x}) = 0$. Nehmen wir also an, es existiere eine Komponente i^* des Vektors

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} \bar{v}_j \nabla h_j(\bar{x})$$

ungleich Null. Diese Komponente ist gerade $\frac{\partial}{\partial x_{i^*}} L(\bar{x}, \bar{u}, \bar{v})$. Wir betrachten den Punkt x^1 mit $x^1_i = \bar{x}_i$ für $i \neq i^*$ und $x^1_{i^*} = \bar{x}_{i^*} + 1$, falls diese partielle Ableitung negativ ist, $x^1_{i^*} = \bar{x}_{i^*} - 1$ sonst. Dann ist

$$\nabla_x L(\bar{x}, \bar{u}, \bar{v})^T (x^1 - \bar{x}) = - \left| \frac{d}{dx_{i^*}} L(\bar{x}, \bar{u}, \bar{v}) \right| < 0$$

und somit ist $x^1 - \bar{x}$ eine Abstiegsrichtung für $L(., \bar{u}, \bar{v})$ im Punkt $\bar{x}$. Das steht im Widerspruch dazu, daß wegen der Sattelpunkteigenschaft diese Funktion ein Minimum in $\bar{x}$ haben muß. □

Der Satz läßt sich mit anderen Mitteln unter einer schwächeren Voraussetzung beweisen. Die Slater-Bedingung als Regularitätsvoraussetzung wird nur für die nicht affinen Ungleichungsfunktionen gebraucht. Der Satz und eine Folgerung daraus (*R.T. Rockafellar, "Convex Analysis"*, Theorem 28.2 und Corollar 28.2.2) werden hier nur zitiert.

Satz 3.33 *Sei (3) ein konvexes Optimierungsproblem und sei I die Menge der Indizes von Ungleichungen, für welche die Funktion g_i nicht affin ist. Wenn der Optimalwert von (3) endlich ist und (3) eine zulässige Lösung besitzt, welche die Ungleichungen für $i \in I$ strikt erfüllt, dann existiert ein (nicht notwendig eindeutig bestimmter) Vektor $(x, u, v)^T$, so daß die KKT-Bedingungen erfüllt sind.*

Folgerung 3.34 *Sei (3) ein konvexes Optimierungsproblem mit nur affinen Restriktionsfunktionen. Wenn der Optimalwert von (3) endlich ist und (3) eine zulässige Lösung besitzt, dann existiert ein Vektor $(x, u, v)^T$, so daß die KKT-Bedingungen erfüllt sind.*

Eine wichtige andere (schwächere) Regularitätsbedingung, welche bei Konvexität und Differenzierbarkeit sowohl aus der Slater-Bedingung als auch aus der LICQ folgt und welche die Notwendigkeit der KKT-Bedingungen garantiert, ist die Mangasarian-Fromovitz-Bedingung (MFCQ):

$$\begin{aligned} &\nabla h_j(\bar{x}),\ j = 1, \ldots, l \ \text{ linear unabhängig und} \\ &\exists d \in \mathbb{R}^n : \quad \nabla g_i(\bar{x}) d < 0, \quad i \in I(\bar{x}) \\ &\qquad\qquad\ \ \nabla h_j(\bar{x}) d = 0, \quad j = 1, \ldots, l \end{aligned}$$

Dies kann hier aus Zeitgründen nicht bewiesen werden, es sei auf Kapitel 5 des Buches von *Bazaraa, Sherali, Shetty* verwiesen, wo auch die Beziehung zum Kegel der tangential zulässigen Richtungen erläutert ist.

Abschließend sei in diesem Abschnitt noch eine hinreichende Bedingung zweiter Ordnung angegeben.

Satz 3.35 (hinreichende KKT-Bedingungen zweiter Ordnung)
$f,\ g_i,\ h_j$ seien zweimal differenzierbar für $i = 1, \ldots, m,\ j = 1, \ldots, l$. Es sei $\bar{x}$ zulässig für (3) und erfülle die KKT-Bedingungen mit Lagrange-Multiplikatoren $\bar{u},\ \bar{v}$. Es sei bezeichnet
$I(\bar{x}) = \{i \in \{1, \ldots, m\} : g_i(\bar{x} = 0\}$, $I^+(\bar{x}) = \{i \in I(\bar{x}) : u_i > 0\}$, $I^0(\bar{x}) = \{i \in I(\bar{x}) : u_i = 0\}$.
Mit $\nabla^2_{xx} L(\bar{x}, \bar{u}, \bar{v})$ sei die Hesse-Matrix bezüglich x der Lagrange-Funktion im Punkt $(\bar{x}, \bar{u}, \bar{v})^T$ bezeichnet:

$$\nabla^2_{xx} L(\bar{x}, \bar{u}, \bar{v}) = \nabla^2 f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \nabla^2 g_i(\bar{x}) + \sum_{j=1}^{l} \bar{v}_j \nabla^2 h_j(\bar{x})$$

wobei $\nabla^2 f(\bar{x})$, $\nabla^2 g_i(\bar{x})$, $\nabla^2 h_j(\bar{x})$ die Hessematrizen der Zielfunktion bzw. der Restriktionsfunktionen in $\bar{x}$ bezeichnen. Es sei der Kegel

$$C = \{d \in \mathbb{R}^n : \left.\begin{array}{lcll} \nabla g_i(\bar{x}) d & = & 0 & i \in I^+(\bar{x}) \\ \nabla g_i(\bar{x}) d & \leq & 0 & i \in I^0(\bar{x}) \\ \nabla h_j(\bar{x}) d & = & 0 & j = 1, \ldots, l \\ d \neq 0 & & & \end{array}\right\}$$

definiert. Ist dann die Hessematrix der Lagrangefunktion positiv definit über C, d.h. gelte

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{u}, \bar{v}) d > 0 \quad \forall d \in C,$$

so ist $\bar{x}$ ein striktes lokales Minimum von (3).

Beweis: Angenommen, $\bar{x}$ sei kein striktes lokales Minimum von (3). Dann existiert eine Folge (x^k) zulässiger Punkte mit $x^k \neq \bar{x}$, welche gegen $\bar{x}$ konvergiert und für welche gilt $f(x^k) \leq f(\bar{x})\ \forall k$. Definieren wir $\lambda_k = \|x^k - \bar{x}\|,\ d^k = \frac{1}{\lambda_k}(x^k - \bar{x})$, so haben wir $x^k = \bar{x} + \lambda_k d^k$ mit $\|d^k\| = 1\ \forall k$ und $\lambda_k \longrightarrow +0$ für $k \longrightarrow \infty$. Da die d^k in einer kompakten Menge liegen, existiert eine konvergente Teilfolge. Um Doppelindizierung zu vermeiden, nehmen wir o.B.d.A. an, die gesamte Folge konvergiere, es gilt also $d^k \longrightarrow d$ mit $\|d\| = 1$.
Weiterhin gilt

$$\begin{array}{lll}
0 \geq f(\bar{x} + \lambda_k d^k) - f(\bar{x}) & = \lambda_k \nabla f(\bar{x}) d^k + \frac{1}{2}\lambda_k^2 {d^k}^T \nabla^2 f(\bar{x}) d^k + \lambda_k^2 \alpha_f(\bar{x}, \lambda_k d^k) & \\
0 \geq g_i(\bar{x} + \lambda_k d^k) - g_i(\bar{x}) & = \lambda_k \nabla g_i(\bar{x}) d^k + \frac{1}{2}\lambda_k^2 {d^k}^T \nabla^2 g_i(\bar{x}) d^k + \lambda_k^2 \alpha_{g_i}(\bar{x}, \lambda_k d^k) & i \in I(\bar{x}) \\
0 = h_j(\bar{x} + \lambda_k d^k) - h_j(\bar{x}) & = \lambda_k \nabla h_j(\bar{x}) d^k + \frac{1}{2}\lambda_k^2 {d^k}^T \nabla^2 h_j(\bar{x}) d^k + \lambda_k^2 \alpha_{h_j}(\bar{x}, \lambda_k d^k) & j = 1, \ldots, l
\end{array}$$

wobei die $\alpha_f,\ \alpha_{g_i},\ \alpha_{h_j}$ gegen Null streben, wenn $k \longrightarrow \infty$ für $i \in I(\bar{x}),\ j = 1, \ldots, l$.
Teilt man nun jeden der obigen Ausdrücke durch $\lambda_k > 0$ und betrachtet den Grenzwert für $k \longrightarrow \infty$, so erhält man

$$\nabla f(\bar{x}) d \leq 0,\ \ \nabla g_i(\bar{x}) d \leq 0 \text{ für } i \in I(\bar{x}),\ \ \nabla h_j(\bar{x}) d = 0 \text{ für } j = 1, \ldots, l. \qquad (*)$$

Da $(\bar{x}, \bar{u}, \bar{v})$ die KKT-Bedingungen erfüllt, gilt

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{l} \bar{v}_j \nabla h_j(\bar{x}) = 0 \qquad (**)$$

Wenn wir nun (**) skalar mit d multiplizieren und die Beziehungen (*) benutzen, erhalten wir

$$\nabla f(\bar{x}) d = 0,\ \ \nabla g_i(\bar{x}) d = 0\ \ i \in I^+(\bar{x}),\ \ \nabla g_i(\bar{x}) d \leq 0\ \ i \in I^0(\bar{x}),\ \ \nabla h_j(\bar{x}) d = 0\ \ j = 1, \ldots, l.$$

Damit gilt $d \in C$.
Multiplizieren wir nun jede der Ungleichungen zu Restriktionen in obigem System mit dem zugeordneten Lagrange-Multiplikator und addieren auf, so erhalten wir unter Benutzung von (**)

$$0 \geq \frac{1}{2}\lambda_k^2 \underbrace{\left[{d^k}^T \nabla^2 f(\bar{x}) d^k + \sum_{i \in I(\bar{x})} \bar{u}_i {d^k}^T \nabla^2 g_i(\bar{x}) d^k + \sum_{j=1}^{l} \bar{v}_j {d^k}^T \nabla^2 h_j(\bar{x}) d^k \right]}_{= {d^k}^T \nabla^2 L(\bar{x}, \bar{u}, \bar{v}) d^k} +$$

$$+ \lambda_k^2 \left[\alpha_f(\bar{x}, \lambda_k d^k) + \sum_{i \in I(\bar{x})} \bar{u}_i \alpha_{g_i}(\bar{x}, \lambda_k d^k) + \sum_{j=1}^{l} \bar{v}_j \alpha_{h_j}(\bar{x}, \lambda_k d^k) \right]$$

Diese Ungleichung dividieren wir nun durch $\lambda_k^2 > 0$ und bilden den Grenzwert für $k \longrightarrow \infty$. Wir erhalten $d^T \nabla^2 L(\bar{x}, \bar{u}, \bar{v}) d \leq 0$, wobei $d \in C$. Dies widerspricht der Voraussetzung, daß die Hessematrix von L positiv definit über C ist, also ist $\bar{x}$ ein striktes lokales Minimum. □

Als notwendige Bedingungen zweiter Ordnung für das Vorliegen eines Minimums in $\bar{x}$ erweist sich unter LICQ, daß die Hessematrix der Lagrangefunktion in $\bar{x}$ positv semidefinit ist über dem Kegel

$$C' = \{ d \in \mathbb{R}^n : \begin{array}{lll} \nabla g_i(\bar{x}) d & \leq & 0 \quad i \in I(\bar{x}) \\ \nabla h_j(\bar{x}) d & = & 0 \quad j = 1, \ldots, l \\ d \neq 0 & & \end{array} \}$$

Für den Beweis siehe *Bazaraa, Sherali, Shetty "Nonlinear Programming"*, Theorem 4.4.3. Offensichtlich gilt $C \subseteq C'$.

Wir waren in diesem Abschnitt weitgehend dem Buch von *Bazaraa, Sherali, Shetty "Nonlinear Programming"* gefolgt und haben die KKT-Bedingungen nur für glatte Probleme (d.h. Probleme mit differenzierbaren Problemfunktionen) formuliert und bewiesen. Dabei brauchen wir Konvexitätsbedingungen nur für die "hinreichende" Richtung, während für die "notwendige" Richtung eine Regularitätsbedingung erforderlich ist.
Analoge Bedingungen lassen sich auch für allgemeine konvexe (nichtdifferenzierbare) Probleme formulieren und beweisen, wobei das Subdifferential an die Stelle der Gradienten tritt und die Gleichung ersetzt wird durch die Bedingung, daß der Nullvektor in der entsprechenden Menge enthalten ist - vgl. *R. T. Rockafellar "Convex Analysis"*, aus welchem hier das Theorem 28.3, (übersetzt in unsere Notation) ohne Beweis angegeben werden soll:

Satz 3.36 (KKT-Bedingungen für allg. konvexe Probleme)
Betrachtet sei ein konvexes Optimierungsproblem, d.h. $f,\ g_i$ *seien konvex für* $i = 1, \ldots, m,\ h_j$ *seien affin für* $j = 1, \ldots, l$.
$(\bar{x}, \bar{u}, \bar{v})$ *ist ein Sattelpunkt für* $L(x, u, v)$ *genau dann, wenn die folgenden Bedingungen erfüllt sind:*

a) $0 \in \left[\partial f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \partial g_i(\bar{x}) + \sum_{j=1}^{l} \bar{v}_j \partial h_j(\bar{x})\right]$

b) $\bar{u}_i \geq 0,\ g_i(\bar{x}) \leq 0,\ \bar{u}_i g_i(\bar{x}) = 0 \quad i = 1, \ldots, m$

c) $h_j(\bar{x}) = 0 \quad j = 1, \ldots, l$

4 Optimierungsverfahren

Wir haben notwendige und hinreichende Optimalitätsbedingungen hergeleitet. Im Falle konvexer differenzierbarer Optimierungsprobleme ist z.B. unter Slater-Bedingung die Bestimmung eines globalen Optimums äquivalent zur Bestimmung von $(x, u, v)^T$, so daß x zulässig und die KKT-Bedingungen erfüllt sind. Damit ist die Suche eines (globalen) Optimums auf die Lösung eines nichtlinearen Ungleichungs-/Gleichungssystems zurückgeführt. Für bestimmte Klassen von Problemen (wie quadratische Optimierungsprobleme) liefert dies anwendbare Verfahren. Im allgemeinen Fall wissen Sie, daß für die Lösung nichtlinearer Gleichungssysteme lokal konvergente Verfahren existieren (wie das Newton-Verfahren). Sind keine Ungleichungsrestriktionen im Problem vorhanden (also auch keine Vorzeichenbeschränkungen an Multiplikatoren), so ist bei Konvergenz des Verfahrens (d.h. Kenntnis einer Anfangsnäherung im Konvergenzbereich) die Lösung des so zugeordneten Gleichungssystems eine Möglichkeit zur Lösung des Optimierungsproblems.
Da man in praktischen Problemen eventuell noch eine gute Anfangsnäherung für die Problemvariablen kennt, gute Näherungen für die Lagrange-Multiplikatoren jedoch schwieriger zu "raten" sind, ist das Problem auf diese Weise noch nicht vom Tisch, ebenso stellt das Finden einer Lösung mit nichtnegativen Multiplikatoren zu Ungleichungen eine zusätzliche Herausforderung dar.
Es sollen nun einige Ideen von Optimierungsverfahren vorgestellt werden.

4.1 Suchverfahren

Eine Gruppe von *ad hoc* Optimierungsverfahren existiert für unrestriktionierte Probleme bzw. Probleme mit einfachen Restriktionen (wie z.B. sog. *box constraints* der Form $x^{lb} \leq x \leq x^{ub}$), welche oft ohne Ableitungsinformationen der Zielfunktion auskommen. Dazu zählen

deterministische Suchverfahren wie z.B.:

- Koordinatenweise Suche, wo ausgehend von einem Startpunkt jeweils parallel zu einer Koordinatenachse das eindimensionale Minimum bestimmt wird. Die Koordinaten werden dabei zyklisch durchlaufen.

- Die sog. Simplexmethode (*Spendley et al., Nelder/Mead*), bei welcher ein Simplex (durch bloßen Vergleich von Funktionswerten) durch die Operationen *Spiegelung* eines Eckpunktes an der gegenüberliegenden Seite (wenn diese in Anwendung auf den Eckpunkt mit größtem Zielfunktionswert einen neuen Punkt mit kleinerem Zielfunktionswert liefert), oder *Kontraktion* (wenn Spiegelung nicht erfolgreich ist) verändert wird. Abbruch bei Unterschreiten einer vorgegebenen Seitenlänge des Simplex, dessen Schwerpunkt wird dann als Näherung des Optimums angesehen.

stochastische Suchverfahren wie z.B.:

- Monte-Carlo-Methode, wobei zufällige Punkte in einer Umgebung des aktuellen Punktes (mit einem Pseudo-Zufallszahlengenerator) erzeugt werden und deren Zielfunktionswerte verglichen werden. Ist der neue Suchpunkt besser als der bisherige, so dient er als neue Näherung. Liefert eine Anzahl von Suchoperationen keine Verbesserung, so wird der Suchradius verkleinert. Ist die Zielfunktion nichtkonvex, so werden "ab und zu" wieder weit entfernte Suchpunkte generiert, um evtl. von lokalen Minima wegzukommen.

- genetische Algorithmen oder Evolutionsverfahren, wobei ausgehend von einer eine Menge von Suchpunkten (der sog. *Population*) mittels Zufallsgeneratoren neue Punkte erzeugt werden. Der Name stammt daher, daß in Anlehnung an reale Vererbungsprozesse aus zwei "Eltern" ein neuer Punkt durch "Rekombination" (Berechnung der Koordinaten aus denen der "Eltern" mit evtl. Vertauschung) und "Mutationen" (zufällige Änderung von zufällig ausgewählten Koordinaten) ein neuer Punkt erzeugt wird und dieser in der Population nur überlebt, wenn seine "fitness" (sein Zielfunktionswert) klein genug im Verhältnis zu dem der restlichen Population ist.

In Fällen, wo die Zielfunktion z.B. nur implizit gegeben ist bzw. Ergebnis einer (aufwendigen) Berechnung oder Simulation und keinerlei Ableitungsinformationen direkt verfügbar sind, kann

der Einsatz solcher Verfahren die einzige Möglichkeit sein. Ebenso kann der Einsatz sinnvoll sein, wenn eine stark nichtkonvexe Zielfunktion vorliegt, keine Information über die Lage des globalen Optimums vorliegt und das Auffinden eines lokalen Optimums (welches die anderen Verfahren im besten Falle liefern) ohne Wert ist. Allgemein sind solche Verfahren wenig effektiv und nur für Probleme in wenigen Variablen sinnvoll (*curse of dimensionality*).

4.2 Abstiegsverfahren

In einem Optimierungsverfahren wird eine Folge von Iterationspunkten erzeugt. Im Falle unrestriktionierter Probleme ist das einzige Kriterium für die Entscheidung, ob ein neuer Iterationspunkt akzeptabel ist, die Verbesserung (Verkleinerung) des Zielfunktionswertes. Daher rührt der Name Abstiegsverfahren. Diese Verfahren liefern in der Regel bei Konvergenz einen Punkt, welcher die notwendigen Optimalitätsbedingungen erfüllt. Dies könnte ein Sattelpunkt sein, im Falle lokaler Konvexität ist das Vorliegen eines lokalen Minimums gesichert und im günstigen Falle (z.B. bei Konvexität des Problems) ein globales Minimum. Durch die Konstruktion des Verfahrens ist zumindest das Stoppen in einem lokalen Maximum (außer im Startpunkt) recht unwahrscheinlich.

Ein generelles Algorithmenschema lautet für das Problem $\min\{f(x) : x \in M\}$:

S0 **Initialisierung** Wähle einen Punkt $x^0 \in M$, setze $k = 0$.

S1 **Abbruchtest**: Untersuche, ob das vorgegebene Abbruchkriterium erfüllt ist. Wenn ja, terminiere den Algorithmus, sonst setze gehe zu S2.

S2 **Bestimmung einer Suchrichtung**: Bestimme einen Vektor $p^k \neq 0$, die Suchrichtung, so daß diese Abstiegsrichtung für f und zulässige Richtung für M ist.

S3 **Schrittweitenbestimmung**: Bestimme eine reelle Zahl $\alpha_k > 0$, so daß $x^{k+1} = x^k + \alpha_k p^k \in M$ und $f(x^{k+1}) < f(x^k)$. Setze $k := k + 1$ und gehe zu **S1**

In diesem Schema sind eine Reihe von Punkten zu konkretisieren.

- Wie wählt man einen Punkt x^0?
- Wie findet man in **S2** eine zulässige Abstiegsrichtung p^k?
- Wie wählt man die Schrittweite α_k?
- Welche Abbruchkriterien sind sinnvoll?

Die Antwort auf diese Fragen hängt vom Typ des vorliegenden Optimierungsproblems ab. Insbesondere sind natürlich nur solche Punkte akzeptabel, welche zulässig sind. Dies betrifft bereits den Anfangspunkt x^0, welcher in der Regel vom Benutzer vorgegeben werden muß.
Das ist am einfachsten im unrestriktionierten Fall, wo jeder Punkt zulässig ist.
Etwas komplizierter, aber noch behandelbar ist der Fall affiner Restriktionsfunktionen. Hier existiert mit der ersten Phase der Simplexmethode ein Algorithmus zum Bestimmen eines zulässigen Punktes bzw. zur Entscheidung, daß kein solcher existiert. Im Schritt **S3** ist bei affinen Restriktionen auch leicht zu entscheiden, wie groß maximal die Schrittweite sein kann, damit Zulässigkeit erhalten bleibt (lineare Ungleichungen für α_k). Das größte Problem stellt die Bestimmung einer Abstiegsrichtung dar, wenn die Richtung des negativen Gradienten keine zulässige Richtung ist. Dann muß eine Projektion auf den Restriktionsbereich erfolgen, was im Falle linearer Gleichungsrestriktionen noch relativ einfach ist. Selbst im Falle linearer Ungleichungsrestriktionen wäre hierfür bereits ein quadratisches Optimierungsproblem zu lösen (was im Falle einer quadratischen Zielfunktion evtl. nicht viel einfacher als das Ausgangsproblem ist). Aus diesem Grunde wird im restriktionierten Falle meist eine Aktive-Indexmengen-Strategie verfolgt, wodurch das Problem auf die Behandlung der aktiven Restriktionen als Gleichungsrestriktionen zurückgeführt wird (z.B. Verfahren der reduzierten Gradienten).
Im allgemeinen nichtlinearen Fall wird oft vom Nutzer verlangt, daß er einen zulässigen Punkt

bereitstellt. Bereits die Entscheidung darüber, ob ein solcher Punkt existiert, kann jedoch schwer sein. Ebenso ist die Bestimmung der maximalen Schrittweite zur Erhaltung der Zulässigkeit komplizierter. Wie eben gesagt, stellt die Bestimmung einer zulässigen Abstiegsrichtung das schwierigste Problem dar.
Weiterhin ist klar, daß für einen Erfolg des Verfahrens (d.h. Konvergenz gegen ein Minimum) nicht allein das Vorliegen eines (evtl. beliebig kleinen) Abstiegs ausreichend sein kann, sondern daß ein hinreichend großer Abstieg erforderlich ist.

4.2.1 Verfahren des steilsten Abstiegs - Gradientenverfahren

Wir wollen voraussetzen, daß die Zielfunktion differenzierbar ist und den unrestriktionierten Fall betrachten. Wir wissen aus dem Satz 3.14 zusammen mit Bemerkung 3.12, daß die Bedingung

$$\nabla f(x)d < 0$$

hinreichend dafür ist, daß d eine Abstiegsrichtung im Punkt x ist. Bei pseudokonvexem f ist diese Bedingung nach Lemma 3.15 auch notwendig.

Nun wollen wir den maximalen Abstieg pro Einheit erreichen, d.h. wir wollen die Richtung mit der größten Änderungsgeschwindigkeit (=Richtungsableitung) der Zielfunktion wählen. Dabei ist natürlich die Länge zu normieren, da das Problem sonst unbeschränkt wäre.

$$\min\{f'(x,d) : \ \|d\| = 1\}$$

Im Falle der euklidischen Norm ist die Lösung des Problems gerade der normierte negative Gradient $\bar{d} = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$, wie man mittels der KKT-Bedingungen leicht findet, die hier (im Falle einer Gleichung) mit der klassischen Methode der Lagrange-Multiplikatoren zusammenfallen (bzw. mit der Schwarz'schen Ungleichung beweist).
Als Abbruchkriterium ergibt sich hier ganz natürlich das Unterschreiten einer vorgegebenen Schranke für die Norm des Gradienten.
Bleibt die Schrittweitenbestimmung zu spezifizieren. Nehmen wir an, das eindimensionale Problem entlang der Suchrichtung könne exakt gelöst werden, so spricht man von exakter Strahlminimierung *exact line search*.

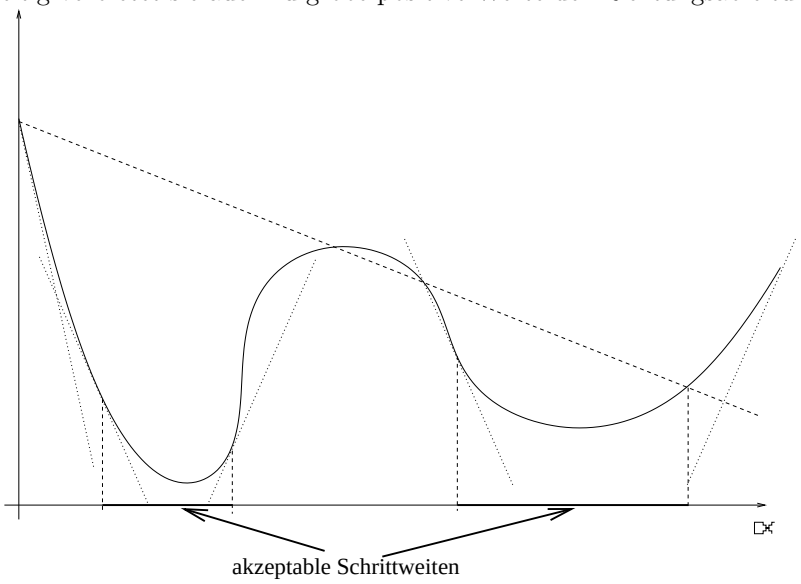
Für das Gradientenverfahren mit exakter Strahlminimierung gilt:

- Der Algorithmus stoppt nach endlich vielen Iterationen in $\bar{x}$ mit $\nabla f(\bar{x}) = 0$ oder
- Jede konvergente Teilfolge von Iterationspunkten hat einen Grenzwert mit verschwindendem Gradienten. Ist die Folge der Iterationspunkte in einer kompakten Menge enthalten, so konvergiert sie.
- Das Verfahren konvergiert linear und die Konvergenzrate ist im schlechtesten Fall beliebig nahe an 1. Es kann gezeigt werden, daß im schlechtesten Fall eine Konvergenzrate von $\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2$ erreicht wird, wobei λ_1 den größten und λ_n den kleinsten Eigenwert von $\nabla^2 f(\bar{x})$ im Minimum $\bar{x}$ bezeichnen (selbst im einfachsten Falle quadratischer Zielfunktion).
- Das Gradientenverfahren arbeitet normalerweise recht gut, solange man weit vom Minimum entfernt ist und zeigt bei Annäherung an das Minimum das sog. *zigzagging* (Aufeinanderfolgende Suchrichtungen sind bei exakter eindimensionaler Minimierung zueinander orthogonal, die Schrittweiten gehen gegen Null und in der Nähe eines stationären Punktes macht der Algorithmus sehr geringe Fortschritte).

Da die Annahme der Möglichkeit einer exakten Bestimmung des eindimensionalen Minimums nur in Spezialfällen real umsetzbar ist, stellt sich die Frage nach Bedingungen für die Schrittweite, welche Konvergenz des Verfahrens garantieren.
Ein Paar solcher Bedingungen sind die *Goldstein-Armijo-Bedingungen*, in *Nocedal/Wright "Numerical Optimization"* mit schwächerer Forderung an ρ als strenge Wolfe-Bedingungen bezeichnet:

$$\begin{array}{lll} f(x^k + \alpha s^k) & \leq f(x^k) + \alpha\rho\nabla f(x^k)^T s^k & \text{mit } \rho \in]0, \frac{1}{2}[\\ |\nabla f(x^k + \alpha s^k)^T s^k| & \leq \sigma |\nabla f(x^k)^T s^k| & \text{mit } \sigma \in]\rho, 1[\end{array} \tag{7}$$

Unter diesen Bedingungen hat das Verfahren auch bei *inexact line search* die angegebenen Konvergenzeigenschaften.
Die erste Bedingung fordert einen hinreichenden Abstieg proportional zur Schrittweite und der Richtungsableitung - der Funktionswert im neuen Punkt muss unterhalb einer affinen Funktion mit negativem Anstieg $\rho\nabla f(x^k)^T s^k$ liegen. Wegen $\rho \in]0, \frac{1}{2}[$ ist dies für hinreichend kleine α erfüllt. Die zweite Bedingung sichert, daß der Schritt nicht zu klein wird. Sie bedeutet wegen $f(x^k)^T s^k < 0$ insbesondere, daß die Richtungsableitung im neuen Punkt größer oder gleich dem σ-fachen der Ableitung im alten Punkt ist, damit sind wegen $0 < \rho < \sigma$ sehr kleine Schrittweiten ausgeschlossen. Gleichzeitig verbietet sie auch zu große positive Werte der Richtungsableitung im neuen Punkt.



Das folgende Lemma zeigt unter nicht einschränkenden Voraussetzungen die Existenz von Schrittweiten, welche diesen Bedingungen genügen.

Lemma 4.1 *Sei $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ stetig differenzierbar, s^k eine Abstiegsrichtung für f in x^k und es sei f nach unten beschränkt über dem Strahl $\{x = x^k + \alpha s^k : \alpha \geq 0\}$. Dann exitieren Intervalle von Schrittweiten, welche den Goldstein-Armijo-Bedingungen genügen.*

Beweis: Da $\Phi(\alpha) := f(x^k + \alpha s^k)$ nach unten beschränkt ist für alle $\alpha > 0$ und wegen $0 < \rho < \frac{1}{2}$ muß die affine Funktion $l(\alpha) := f(x^k) + \alpha\rho\nabla f(x^k)^T s^k$ den Graphen von Φ in mindestens einem $\alpha > 0$ schneiden. Sei $\alpha' > 0$ der kleinste dieser Werte, d.h.

$$f(x^k + \alpha' s^k) = f(x^k) + \alpha'\rho\nabla f(x^k)^T s^k$$

Die Bedingung hinreichenden Abstiegs gilt dann als strenge Ungleichung für alle $\alpha \in]0, \alpha'[$.
Nach dem Mittelwertsatz existiert ein $\bar{\alpha} \in]0, \alpha'[$, so daß

$$f(x^k + \alpha' s^k) - f(x^k) = \alpha'\nabla f(x^k + \bar{\alpha} s^k)^T s^k$$

Unter Verwendung der beiden Gleichungen erhalten wir

$$\nabla f(x^k + \bar{\alpha} s^k)^T s^k = \rho\nabla f(x^k)^T s^k > \sigma\nabla f(x^k)^T s^k$$

wegen $0 < \rho < \sigma$ und $\nabla f(x^k)^T s^k < 0$. Der Term auf der linken Seite ist negativ, somit ist auch die zweite Ungleichung der Goldstein-Armijo-Bedingungen in $\bar{\alpha}$ als strenge Ungleichung erfüllt. Wegen der stetigen Differenzierbarkeit existiert dann eine Umgebung um $\bar{\alpha}$, in welcher beide Ungleichungen gelten. □

Für linear restriktionierte Probleme der Form

$$\min\{f(x) : \quad \left.\begin{array}{lll} {a^i}^T x - b_i & \leq 0, & i \in I \\ {c^j}^T x - d_j & = 0, & j \in J \end{array}\right\}$$

ist das Verfahren durch die Bestimmung einer zulässigen Abstiegsrichtung mittels der Lösung eines quadratischen Optimierungsproblems modifizierbar. Dieses quadratische Optimierungsproblem

könnte z.B. mit den Verfahren von Lemke (durch eine modifizierte Simplexmethode der linearen Optimierung) gelöst werden. Eine zulässige Abstiegsrichtung existiert, wenn das *Richtungssuchproblem*

$$\left.\begin{array}{lll} \min\{\nabla f(x^k)^T p: & {a^i}^T p & \leq 0, \quad i \in I(x^k) \\ & {c^j}^T p & = 0, \quad j \in J \\ & p^T p & = 1 \end{array}\right\}$$

einen negativen optimalen Zielfunktionswert besitzt. Die Lösung ist eine Suchrichtung. Das angegebene Problem hat eine nichtlineare Restriktion, die Normierungsbedingung. Man kann jedoch äquivalent das folgende quadratische Optimierungsproblem (d.h. quadratische Zielfunktion und lineare Restriktionen) betrachten

$$\left.\begin{array}{lll} \min\{p^T p: & {a^i}^T p & \leq 0, \quad i \in I(x^k) \\ & {c^j}^T p & = 0, \quad j \in J \\ & \nabla f(x^k)^T p & = -1 \end{array}\right\}$$

Reduzierte Gradienten

Für den Fall unrestriktionierter Problem haben wir den normierten negativen Gradienten als Suchrichtung im Verfahren des steilsten Abstiegs gewählt. Ist nun ein Problem mit linearen Gleichungsrestriktionen der Form

$$\min\{f(x): \ Ax = b\}$$

gegeben, so wird die Gradienten-Richtung allgemein keine zulässige Richtung sein. Bei der Idee der reduzierten Gradienten geht es nun darum, zulässige Such-Richtungen in einem reduzierten Raum (der Lösungsmenge des Gleichungssystems) abzuleiten.
Dazu seien o.B.d.A. die m Zeilen von A linear unabhängig. Dann ist der Lösungsraum des Gleichungssystems von der Dimension $n - m$. Ein direktes Vorgehen wäre nun, das Gleichungssystem nach einer Teilmenge von m Variablen umzustellen, dieses Ergebnis in die Zielfunktion einzusetzen und damit ein freies Optimierungsproblem in $n - m$ Variablen zu erhalten - analog zur Darstellung des Problems im Raum der Nichtbasisvariablen in der linearen Optimierung ($x_B = A_B^{-1} b - A_B^{-1} A_N x_N$).
Direkte Variablen-Elimination ist nicht die einzige Möglichkeit, eine verallgemeinerte Eliminationsmethode ist im Wesentlichen eine Lineartransformation der Variablen.
Jeder Vektor des des $\mathbb{R}^n$ läßt sich eindeutig als Summe zweier orthogonaler Vektoren darstellen - eines Vektors im Bildraum der Matrix A^T und eines Vektors im dazu komplementären Nullraum der Matrix A. Es existieren eine $n \times m$-Matrix Y und eine $n \times (n-m)$-Matrix, Z mit $AY = I$, AZ die Nullmatrix und $(Y|Z)$ ist regulär.
Für den Fall der einfachen Elimination ist mit der Schreibweise analog zur linearen Optimierung $A = (A_B A_N)$ mit der regulären Teilmatrix A_B

$$Y = \begin{pmatrix} A_B^{-1} \\ 0 \end{pmatrix}, \quad Z == \begin{pmatrix} -A_B^{-1} A_N \\ I \end{pmatrix}$$

Zu jedem Vektor $x \in \mathbb{R}^n$ existieren dann eindeutig bestimmte Vektoren $x_Y \in \mathbb{R}^m$ und $x_Z \in \mathbb{R}^{n-m}$ mit $x = Y x_Y + Z x_Z$.
Ist nun $x^* = Y x_Y^* + Z x_Z^*$ ein zulässiger Punkt, so gilt

$$Ax^* = AY x_Y^* + AZ x_Z^* = x_Y^* = b$$

jeder zulässige Punkt ist also von der Form $Yb + Z x_Z$ mit $x_Z \in \mathbb{R}^n - m$. Ist der aktuelle Iterationspunkt x^k zulässig, so kann jeder zulässiger Punkt ausgedrückt werden als

$$x = x^k + Zy \quad \text{ mit } y \in \mathbb{R}^{n-m}$$

Betrachtet man das unrestriktionierte Problem der Minimierung von

$$\Psi(y) := f(x^k + Zy)$$

so ist nach der Kettenregel

$$\nabla_y \Psi(y) = Z^T \nabla_y f(x)$$

was als reduzierter Gradientenvektor bezeichnet wird und

$$\nabla_y^2 \Psi(y) = Z^T \nabla_x^2 f(x) Z$$

die reduzierte Hessematrix.
Die Aussagen und Techniken für unrestriktionierte Probleme sind auf das reduzierte Problem anwendbar. Sucht man analog zu oben die Richtung des steilsten Abstiegs als Lösung von

$$\min\{\nabla f(x^k)^T Z p_Z : \ \|Z p_Z\|_2^2 = 1\},$$

so erhält man aus den KKT-Bedingungen als Suchrichtung im vollen Raum

$$s^k = -\frac{1}{\|p^k\|_2} p^k \quad \text{mit} \ \ p^k = -Z(Z^T Z)^{-1} Z^T \nabla f(x^k)$$

Liegt ein Problem mit linearen Gleichungs- und Ungleichungsrestriktionen vor, so kommt eine Aktive-Indexmengen-Strategie zum Einsatz, bei welcher stets ein Gleichungs-restriktioniertes Problem mit den aktiven Restriktionen gelöst wird. Für die Entscheidung über Optimalität (Lagrange-Multiplikatoren zu Ungleichungen sind nichtnegativ) und die eventuelle Verkleinerung der aktiven Indexmenge wird eine Näherung für die Lagrange-Multiplikatoren benötigt. Eine Möglichkeit der Berechnung der Lagrange-Multiplikatoren liefert die Multiplikation der Bestimmungsgleichung $\nabla f(x^k) + A^T v^k = 0$ von links mit der Matrix Y^T, wodurch sich ergibt $v^k = -Y^T \nabla f(x^k)$.
Verbesserte Konvergenzeigenschaften ergeben sich bei Benutzung zweiter Ableitungen, wobei die Suchrichtung z.B. mittels der Newton-Iteration

$$Z^T \nabla_x^2 f(x^k) Z p_z = -Z^T \nabla f(x^k)$$

bestimmt werden könnte. Ein Problem dabei ist, daß die reduzierte Hessematrix nicht positiv definit zu sein braucht, weshalb positiv definite Approximationen dieser Matrix zum Einsatz kommen.

Hat das Problem nichtlineare Restriktionen, so wird zunächst die Matrix A und damit auch Z aus der Linearisierung der Restriktionen im aktuellen Iterationspunkt gewonnen, ist also in jedem Punkt neu zu bestimmen. Die reduzierte Suchrichtung liegt dann im Tangentialraum zu den aktiven Restriktionen. Eine nachfolgende Korrektur (z.B. Newton-Verfahren) ist also für jede Schrittweite nötig, um Zulässigkeit bezüglich nichtlinearer Restriktionen herzustellen, die Strahlminimierung erfordert also in jedem Suchpunkt einen weiteren Iterationsprozeß.

4.2.2 Verfahren der konjugierten Gradienten

Das Verfahren der konjugierten Gradienten wurde zunächst als iteratives Verfahren zur Lösung des linearen Gleichungssystems $Ax = b$ mit symmetrischer positiv definiter $n \times n$-Koeffizientenmatrix bzw. (was äquivalent ist) zur Minimierung der Funktion

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

entwickelt. Dabei werden auf effiziente Weise Richtungen $p^1, p^2, \ldots, p^n$ erzeugt, welche bezüglich der Matrix A konjugiert sind, d.h. es gilt

$${p^i}^T A p^j = 0 \quad \forall i \neq j$$

Die so erzeugten Richtungen sind linear unabhängig und es gilt:
wird die Folge (x^k) durch die Vorschrift

$$x^{k+1} = x^k + \alpha_k p^k \tag{8}$$

erzeugt, wobei α_k durch exakte Strahlminimierung bestimmt wird, so liefert der Algorithmus die Lösung des Gleichungssystems und damit ein globales Minimum von f in höchstens n Schritten.

Die Schrittweite für exakte Strahlminimierung läßt sich in dem Falle der quadratischen Funktion explizit angeben: Die streng konvexe Funktion entlang des Strahls wird minimiert, wenn

$$0 = \frac{d}{d\alpha} f(x^k + \alpha_k p^k) = \big(A(x^k + \alpha_k p^k) - b\big)^T p^k = \big(\alpha_k A p^k + (Ax^k - b)\big)^T p^k, \tag{9}$$

woraus folgt

$$\alpha_k = -\frac{{r^k}^T p^k}{{p^k}^T A p^k} \quad \text{mit } r^k = r(x^k) = Ax^k - b \tag{10}$$

$r^k = r(x^k)$ bezeichnet das Residuum des Gleichungssystems im Punkt x^k, was gleich dem Gradienten von f in x^k ist. Dieses Ergebnis wollen wir im folgenden Satz formulieren (wegen der linearen Unabhängigkeit spannen n konjugierte Richtungen den gesamten Raum auf). Zunächst nehmen wir an, wir hätten bereits eine Menge zueinander bzgl. A konjugierter Richtungen.

Satz 4.2 *Sei $x^1 \in \mathbb{R}^n$ ein beliebiger Startpunkt und mit der Menge konjugierter Richtungen $\{p^1, \ldots, p^k\}$ werde die Folge (x^k) gemäß (8) und (10) gebildet. Dann ist x^{k+1} das Minimum von f über der Menge*

$$\big\{x \in \mathbb{R}^n : \; x^1 + lin\{p^1, \ldots, p^k\}\big\}$$

Beweis: Wir betrachten die Hilfsfunktion

$$h(\xi) = f(x^1 + \sum_{i=1}^{k} \xi_i p^i)$$

Dies ist eine streng konvexe quadratische Funktion und somit existiert genau ein Minimum ξ^* dieser Funktion und die notwendige Bedingung ist hinreichend. Anwendung der Kettenregel liefert

$$0 = \frac{\partial}{\partial \xi_i} h(\xi^*) = \nabla f(x^1 + \sum_{j=1}^{k} \xi_j^* p^j)^T p^i = r(x^*)^T p^i, \quad i = 1, \ldots, k$$

mit $x^* = x^1 + \sum_{j=1}^{k} \xi_j^* p^j$ und $r(x^*) = Ax^* - b$ bezeichnet analog zu oben das Residuum des Gleichungssystems (bzw. den Gradienten von f) in x^*.
x^* ist also das Minimum über dem betrachteten affinen Unterraum genau dann, wenn

$$r(x^*)^T p^i = 0, \quad i = 1, \ldots, k$$

Diese Eigenschaft beweisen wir nun für die Iterationspunkte mittels vollständiger Induktion.
Wegen (9) gilt zunächst ${r^2}^T p^1 = 0$. Es gelte nun (Induktionsvoraussetzung):

$${r^k}^T p^i = 0, \quad i = 1, \ldots, k-1$$

Nun gilt

$$r^{k+1} = Ax^{k+1} - b = A(x^k + \alpha_k p^k) - b = \big(Ax^k - b\big) + \alpha_k A p^k = r^k + \alpha_k A p^k \tag{11}$$

und daraus folgt mit der Definition von α_k durch (10)

$${p^k}^T r^{k+1} = {p^k}^T r^k + \alpha_k {p^k}^T A p^k = 0 \tag{12}$$

Für die Vektoren p^i, $i = 1, \ldots, k-1$ folgt mit Hilfe der Induktionsvoraussetzung und der Konjugiertheit der Richtungen

$${p^i}^T r^{k+1} = \underbrace{{p^i}^T r^k}_{=0\ (IV)} + \alpha_k \underbrace{{p^i}^T A p^k}_{=0\ (konjug.)} = 0$$

□

Soweit ist die vorgestellte Methode eine *Methode der konjugierten Richtungen* und ihre Eigenschaften sind für quadratische streng konvexe Probleme attraktiv. Es bleibt für diese Methode die Frage: woher bekommt man eine Menge von n zueinander bezüglich A konjugierten Richtungen? Eine Wahl wäre ein vollständiger Satz von Eigenvektoren, jedoch deren Bestimmung ist recht aufwendig und würde vor allem für Probleme großer Dimension einen unverhältnismäßig großen Rechenaufwand erfordern.
Es gibt weitere Ansätze, wie eine modifizierte Gram-Schmidt-Orthogonalisierung, jedoch auch diese erfordert recht großen Rechenaufwand und die Speicherung der gesamten Menge von Richtungen. Ein anderer Ansatz führt zur eigentlichen Methode der konjugierten Gradienten: die Richtungen werden nicht vorher bestimmt, sondern im Laufe des Lösungsalgorithmus werden die Richtungen

zusammen mit den Iterationspunkten erzeugt. Es erweist sich dabei, daß nicht die gesamte Menge von Richtungen gespeichert werden muß, sondern daß die letzte benutzte Richtung p^k zur Bestimmung von p^{k+1} ausreicht. Die Berechnungsvorschrift garantiert dann, daß die neue Richtung automatisch auch konjugiert zu allen vorhergehenden Richtungen bestimmt wird. Auf diese Weise wird Speicherplatz und Rechenaufwand gespart.

Als erste Richtung wählen wir die des steilsten Abstiegs, also

$$p^1 = -\nabla f(x^1) = -r^1 \tag{13}$$

Für alle weiteren machen wir den Ansatz als Linearkombination aus der negativen Gradientenrichtung und der zuletzt benutzten Richtung (wenn wir $p^0 = 0$ setzen, gilt die folgende Gleichung für alle k):

$$p^k = -r^k + \beta_k p^{k-1}, \tag{14}$$

wobei die Zahl β_k so zu bestimmen ist, daß p^{k-1} und p^k zueinander bzgl A konjugiert sind. Der obige Ansatz wird also von links mit ${p^{k-1}}^T A$ multipliziert und wir erhalten aus der Gleichung ${p^{k-1}}^T A p^k = 0$

$$\beta_k = \frac{{p^{k-1}}^T A r^k}{{p^{k-1}}^T A p^{k-1}} \tag{15}$$

Satz 4.3 *Sei $x^1 \in \mathbb{R}^n$ ein beliebiger Startpunkt und die Folge (x^k) gemäß (8), (10), (13) und (14), (15) gebildet. Dann terminiert der Algorithmus nach $m \leq n$ Iterationen mit einem stationären Punkt x^{m+1} (d.h. $r^{m+1} = \nabla f(x^{m+1}) = 0$) und es gilt für alle $k \leq m$*

i) ${p^k}^T A p^i = 0 \qquad i = 1, \ldots, k-1$

ii) ${r^k}^T r^i = 0 \qquad i = 1, \ldots, k-1$

iii) ${p^k}^T r^k = -{r^k}^T r^k$

Beweis: Aus Satz 4.2 folgt mit i), daß der Algorithmus spätestens nach n Schritten terminiert.
Für $m = 0$ sind die Aussagen trivial. Ist $m \geq 1$, so beweisen wir die Aussagen mittels vollständiger Induktion.
Für $k = 1$ ist für die Aussagen i) und ii) nichts zu zeigen. iii) gilt, da $p^1 = -r^1$.
Gelten nun die Aussagen für ein $k < m$. Wegen iii) und $k < m$, d.h. $r^k \neq 0$ ist

$$\alpha_k = -\frac{{r^k}^T p^k}{{p^k}^T A p^k} = \frac{{r^k}^T r^k}{{p^k}^T A p^k} > 0 \tag{16}$$

Wir benutzen wieder (11) und (14) und erhalten für $i \in \{1, \ldots, k\}$

$$\begin{aligned} {r^{k+1}}^T r^i &= {r^k}^T r^i + \alpha_k {p^k}^T A r^i \\ &= {r^k}^T r^i + \alpha_k {p^k}^T A(\beta_i p^{i-1} - p^i) \\ &= {r^k}^T r^i + \alpha_k \beta_i {p^k}^T A p^{i-1} - \alpha_k {p^k}^T A p^i \end{aligned}$$

Für $i = k$ ist der Ausdruck gleich Null wegen (16) und der Gültigkeit von i) für k. Für $i < k$ folgt aus der Gültigkeit von i) und ii) für k, daß der Ausdruck gleich Null ist. Somit ist die Gültigkeit von ii) auch für $k+1$ gezeigt.
Wiederum mit (14) und (11) ergibt sich

$$\begin{aligned} {p^{k+1}}^T A p^i &= -{r^{k+1}}^T A p^i + \beta_{k+1} {p^k}^T A p^i \\ &= \tfrac{1}{\alpha_i} {r^{k+1}}^T (r^i - r^{i+1}) + \beta_{k+1} {p^k}^T A p^i \end{aligned}$$

Für $i = k$ ist dies (vgl. erste Zeile) gleich Null wegen (15). Für $i < k$ ist der zweite Summand der zweiten Zeile gleich Null wegen i) für k, der erste wegen der eben gezeigten Gültigkeit von ii) für $k+1$.

Wenden wir schließlich (14) und die Orthogonalität von Gradient (=Residuum) und Suchrichtung (12) an, so erhalten wir die Gültigkeit von iii) für $k+1$

$$\begin{aligned} -{r^{k+1}}^T p^{k+1} &= {r^{k+1}}^T r^{k+1} - \beta_{k+1} {r^{k+1}}^T p^k \\ &= {r^{k+1}}^T r^{k+1} \end{aligned}$$

Somit sind alle drei Beziehungen induktiv bewiesen, insbesondere zeigt i) die Konjugiertheit der Richtungen und damit das Erfülltsein der Voraussetzung von Satz 4.2. □

Bemerkung: Für den Beweis der Eigenschaften ist es wichtig, daß die erste Suchrichtung gerade die negative Gradientenrichtung ist.

Wir haben im Beweis bereits eine effizientere Formel für die Bestimmung der Schrittweite angegeben.

$$\alpha_k = \frac{{r^k}^T r^k}{{p^k}^T A p^k}$$

Ebenso läßt sich die Bestimmungsgleichung (15) für β_k vereinfachen, indem wir (11) benutzen, was liefert

$$A p^{k-1} = \frac{1}{\alpha_{k-1}} (r^k - r^{k-1})$$

Dies eingesetzt in (15) liefert mit der obigen Formel für α_{k-1} unter Benutzung von Aussage ii) von Satz 4.3:

$$\begin{aligned} \beta_k &= \frac{{r^k}^T A p^{k-1}}{{p^{k-1}}^T A p^{k-1}} \\ &= \frac{{p^{k-1}}^T A p^{k-1}}{{r^{k-1}}^T r^{k-1}} \cdot \frac{{r^k}^T (r^k - r^{k-1})}{{p^{k-1}}^T A p^{k-1}} \end{aligned}$$

und damit (mit iii) von Satz 4.3)

$$\beta_k = \frac{{r^k}^T r^k}{{r^{k-1}}^T r^{k-1}}$$

Mit dieser Formel kann das in der vorigen Iteration berechnete Skalarprodukt ${r^{k-1}}^T r^{k-1}$ wiederverwendet werden für die Berechnung von β_k.

Der Algorithmus liefert in höchstens n Schritten eine stationären Punkt, die Konvergenz ist noch schneller, wenn die Eigenwerte von A in wenigen Clustern liegen. Damit kann das Verfahren beschleunigt werden, wenn die Eigenwertverteilung der Matrix A verändert werden kann. Dies geschieht durch sog. *preconditioning*, wo durch eine nichtsinguläre Matrix C eine Koordinatentransformation

$$\hat{x} = Cx$$

vorgenommen wird, wodurch die Funktion in den neuen Variablen die Matrix

$${C^{-1}}^T A C^{-1}$$

erhält. Man kann nun C so wählen, daß in den neuen Variablen die Matrix bessere Eigenschaften hinsichtlich der Verteilung der Eigenwerte hat (*preconditioned cg-Verfahren*), siehe z.B. *J. Nocedal, S.J. Wright "Numerical Optimization"*.

Es stellt sich nun die Frage, ob dieser Algorithmus auch für die Minimierung allgemeiner nichtlinearer Funktionen f adaptiert werden kann. *Fletcher* und *Reeves* haben gezeigt, daß dies mit kleinen Änderungen möglich ist. Die explizite Angabe der Schrittweite zur exakten Strahlminimierung ist dann natürlich nicht mehr möglich, sondern diese muß mit eindimensionaler Suche bestimmt werden. Das Residuum r^k wird in völliger Analogie durch den Gradienten von f im Iterationspunkt ersetzt und wir erhalten den Fletcher-Reeves-Algorithmus

Initialisierung: Eingabe: x^0. Berechne $\nabla f_0 = \nabla f(x^0)$, setze $p^0 = -\nabla f_0$, $k = 0$
WHILE $\|p^k\| > \varepsilon$
Bestimme die Schrittweite α_k, so daß die Goldstein-Armijo-Bedingungen mit $0 < \rho < \sigma < \frac{1}{2}$ erfüllt sind
Setze $x^{k+1} = x^k + \alpha_k p^k$
Berechne $\nabla f_{k+1} = \nabla f(x^{k+1})$
$\beta_{k+1} := \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k \nabla f_k}$
$p^{k+1} := -\nabla f_{k+1} + \beta_{k+1} p^k$
$k := k + 1$
END WHILE

Mit Schrittweiten, welche den Goldstein-Armijo-Bedingungen genügen, liefert das Verfahren stets Richtungen p^k, welche Abstiegsrichtungen sind:

Lemma 4.4 *Die Funktion f sei zweimal stetig differenzierbar, die Niveaumenge $L_f(f(x^0) = \{x \in \mathbb{R}^n : f(x) \le f(x^0)\}$ sei beschränkt. Der Fletcher-Reeves-Algorithmus mit Schrittweiten gemäß (7) erzeugt in jeder Iteration eine Abstiegsrichtung p^k, welche den folgenden Ungleichungen genügt:*

$$-\frac{1}{1-\sigma} \le \frac{\nabla f_k p^k}{\|\nabla f_k\|^2} \le \frac{2\sigma - 1}{1-\sigma} \tag{17}$$

Beweis: Zunächst liefert Lemma 4.1, daß in jedem Schritte eine Schrittweite α_k existiert, welche den Goldstein-Armijo-Bedingungen genügt.
Betrachten wir die Funktion $t :]-\infty, 1[\longrightarrow \mathbb{R}$ mit $t(\xi) := \frac{2\xi - 1}{1-\xi}$. Es ist $t'(\xi) = \frac{1}{(1-\xi)^2} > 0$, also ist t monoton wachsend. Weiterhin ist $t(0) = -1$, $t(\frac{1}{2}) = 0$ und somit wegen $\sigma \in]0, \frac{1}{2}[$

$$-1 < \frac{2\sigma - 1}{1-\sigma} < 0 \tag{18}$$

Aus $\sigma > 0$ ergibt sich sofort

$$-\frac{1}{1-\sigma} < -1 \tag{19}$$

Wenn wir die Ungleichung (17) bewiesen haben, so ergibt sich somit sofort, daß die Richtung p^k eine Abstiegsrichtung ist.
Der Beweis erfolgt nun mittels Induktion.
Für $k = 0$ ist der mittlere Term in (17) gleich -1, mit (18) und (19) sehen wir also, daß beide Ungleichungen erfüllt sind.
Es sei also (17) für ein k erfüllt. Aus den Formeln für p^{k+1} und β_k im Algorithmus ergibt sich

$$\frac{\nabla f_{k+1}^T p^{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \beta_{k+1} \frac{\nabla f_{k+1}^T p^k}{\|\nabla f_{k+1}\|^2} = -1 + \frac{\nabla f_{k+1}^T p^k}{\|\nabla f_k\|^2} \tag{20}$$

Die zweite Ungleichung der Goldstein-Armijo-Bedingungen (7) liefert wegen $f_k^T p^k < 0$

$$\sigma \nabla f_k p^k \le \nabla f_{k+1}^T p^k \le -\sigma \nabla f_k p^k$$

und dies liefert mit (20)

$$-1 + \sigma \frac{\nabla f_k p^k}{\|\nabla f_k\|^2} \le \frac{\nabla f_{k+1}^T p^{k+1}}{\|\nabla f_{k+1}\|^2} \le -1 - \sigma \frac{\nabla f_k p^k}{\|\nabla f_k\|^2}$$

Setzen wir nun die linke Ungleichung aus der Induktionsvoraussetzung für den Term $\frac{\nabla f_k p^k}{\|\nabla f_k\|^2}$ ein, so erhalten wir

$$-1 - \frac{\sigma}{1-\sigma} \le \frac{\nabla f_{k+1}^T p^{k+1}}{\|\nabla f_{k+1}\|^2} \le -1 + \frac{\sigma}{1-\sigma}$$

was die Gültigkeit von (17) für $k + 1$ zeigt. □

Es gilt die folgende Aussage über globale Konvergenz des Verfahrens (*Nocedal/Wright "Numerical Optimization"*, Theorem 5.8):
Ist die Niveaumenge $L_f(f(x^0)$ beschränkt und existiert eine Umgebung $N \supset L_f(f(x^0)$, auf welcher f Lipschitzstetig differenzierbar ist, d.h. $\exists L > 0 \ \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in N$, dann gilt für die vom Fletcher-Reeves-Verfahren erzeugte Folge

$$\liminf_{k \longrightarrow \infty} \nabla ||f(x^k)|| = 0.$$

Es gibt eine Reihe von Varianten des Verfahrens, welche sich vor allem in der Bestimmung des Parameters β_k unterscheiden. Eine wichtige andere Variante ist das *Polak-Ribière-Verfahren*, bei welchen die Formel geändert wird zu

$$\beta_{k+1}^{PR} := \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\nabla f_k \nabla f_k}$$

Für streng konvexes quadratische Funktion f liefern beide Formeln identische Ergebnisse, da wir in Satz 4.3 ii) erhalten hatten, daß die Gradienten paarweise orthogonal sind. Für allgemeine Funktionen hat sich jedoch die Variante Polak-Ribière als robuster und effizienter erwiesen. Überraschenderweise sichern die Bedingungen für die Schrittweitenbestimmung bei dieser Variante nicht, daß jedes p^k eine Abstiegsrichtung ist, was zu der zusätzlichen Vorschrift $\beta_{k+1} := \max\{\beta_{k+1}^{PR}, 0\}$ führt.

Für allgemeine Funktionen gilt nun nicht mehr die Aussage, daß nach maximal n Schritten ein stationärer Punkt erreicht wird. Es wird nun ein regelmäßiger *Restart* (z.B. nach n Schritten) durchgeführt, bei welchem die bisher aufgesammelten Informationen in der Suchrichtung verworfen werden und wieder mit der negativen Gradientenrichtung gestartet wird. Motivation: nach dem Satz von Taylor kann die Funktion in der Nähe der Lösung gut durch eine streng konvexe quadratische Funktion approximiert werden. Die positiven Eigenschaften der endlichen Termination beruhten wesentlich mit darauf, daß als Anfangsrichtung die des negativen Gradienten gewählt wurde.

Da bei großdimensionalen Problemen ein Restart evtl. nie erreicht würde, werden in Implementationen der Methode auch andere Kriterien für einen Restart herangezogen. So sind bei quadratischer Funktion die Gradienten in verschiedenen Iterationspunkten orthogonal zueinander, ein mögliches Kriterium ist, daß zwei aufeinanderfolgende Gradienten "deutlich nicht-orthogonal" sind, d.h. mit z.B. $\nu = 0.1$

$$\frac{\nabla f_k \nabla f_{k-1}}{\nabla f_k \nabla f_k} \geq \nu$$

Praktisch funktioniert die Methode recht gut, im Gegensatz zum Fall streng konvexer quadratischer Funktionen f gibt es jedoch keine vollständige Konvergenztheorie. Es gibt sogar Beispiele für sehr langsame Konvergenz des Fletcher-Reeves-Verfahrens (schlechter als steilster Abstieg), wo beliebig kleine Iterationsschritte erzeugt werden. Die Polak-Ribière-Variante verhält sich in diesen Fällen besser.

4.3 Straf- und Barriereverfahren

In diesem Abschnitt wollen wir Methodiken diskutieren, welche die Lösung restriktionierter Optimierungsprobleme durch die Lösung einer Folge von unrestriktionierten Problemen approximieren.

4.3.1 Strafmethoden

Die Idee ist bereits bei der Diskussion der Sattelpunktbedingungen aufgetreten: eine Verletzung der Restriktionen wird durch Terme in der Zielfunktion (analog zur Lagrangefunktion) "bestraft". Die Zielfunktion soll für zulässige Punkte nicht verändert werden.
Für Gleichungen $h_j(x) = 0$ leistet ein Summand der Form $\mu h_j^2(x)$ das Gewünschte. Ist die Gleichung verletzt, so wird bei hinreichend großem Multiplikator μ der Punkt "unattraktiv" in der Minimierung.
Für Ungleichungen könnte man die Funktion

$$g_i^+(x) = \max\{g_i(x), 0\}$$

verwenden, diese braucht jedoch auch bei differenzierbaren Problemfunktionen nicht differenzierbar zu sein. Deshalb verwendet man meist Terme der Form $\mu(g_i^+(x))^2$, welche bei stetig differenzierbarem g_i ebenfalls stetig differenzierbar sind.

Insgesamt können wir für ein Problem der Form (3) die quadratische Straffunktion definieren als

$$Q(x, \mu) = f(x) + \mu \left(\sum_{i=1}^{m} (g_i^+(x))^2 + \sum_{j=1}^{l} h_j^2(x) \right)$$

Für $\mu \longrightarrow \infty$ können wir erwarten, die Lösung des Originalproblems approximativ zu finden.

Ein generelles Schema für den Algorithmus lautet

Gegeben: Startpunkt x_S^0, Strafparameter μ_0, Toleranz τ_0
FOR $k = 0, 1, \ldots$
Bestimme mit Startpunkt x_S^k näherungsweise ein Minimum x^k für $Q(x, \mu_k)$
mit Abbruchkriterium $\|\nabla_x Q(x^k, \mu_k)\| \leq \tau_k$
Teste auf Konvergenz.
Wenn Konvergenztest erfüllt, STOP mit Näherungslösung x^k
Wähle neuen Startpunkt x_S^{k+1},
neue Toleranz $\tau_{k+1} < \tau_k$
und Strafparameter $\mu_{k+1} > \mu_k$
END FOR

Die Parameter des Verfahrens müssen folgenden Bedingungen genügen:

$$\tau_k \longrightarrow 0$$
$$\mu_k \longrightarrow \infty$$

Diese Methode wird manchmal auch als *äußere Strafmethode* bezeichnet, weil Iterationspunkte nicht notwendig zulässig für das Ausgangsproblem sind.
Die Lösung des Ausgangsproblems erfolgt so durch die wiederholte Lösung freier Minimum-Probleme, welche z.B. mit den im vorigen Abschnitt dargestellten Verfahren erfolgen kann. Bei der Anwendung des cg-Verfahrens wird allerdings mit wachsendem Strafparameter μ die Konditionszahl der Hessematrix größer, so daß das Verfahren immer schlechter konvergiert. Es müssen speziell angepaßte Verfahren verwendet werden.

Trotz der numerisch nicht idealen Eigenschaften liefert die Strafmethode einen möglichen Ansatz zur Behandlung von Problemen mit Restriktionen (auch bei nichtkonvexen Problemen). Wir wollen dies für ein abstrahiertes Modell mit exakter Lösung der freien Minimum-Probleme beweisen.

Dazu formulieren wir zunächst die Eigenschaften, welche eine Straffunktion erfüllen muß (und

welche von der angegebenen quadratischen Straffunktion $Q(x,\mu)$ bei glatten Problemfunktionen erfüllt werden).

Definition 4.5 *M bezeichne die Restriktionsmenge von (3).*
Eine über $\mathbb{R}^n \times \mathbb{R}_+$ definierte Funktion $p(x,\mu)$ heißt Straffunktion für (3), wenn

i) $p(x,\mu)$ stetig in $x \quad \forall \mu > 0$

ii) für beliebige Folgen (x^k), (μ_k) mit $x^k \in \mathbb{R}^n$, $\mu_k > 0$ und $x^k \longrightarrow y$, $\mu_k \longrightarrow \infty$ gilt

$$\liminf p(x^k,\mu_k) \quad \begin{cases} = +\infty & \text{für } y \notin M \\ \geq f(y) & \text{für } y \in M \end{cases}$$

iii) für $y \in M$ und beliebige Folgen (μ_k) mit $\mu_k > 0$ und $\mu_k \longrightarrow \infty$ gilt $p(y,\mu_k) \longrightarrow f(y)$.

Satz 4.6 *Es seien $p(x,\mu)$ eine Straffunktion für (3), (μ_k) mit $\mu_k > 0$ und $\mu_k \longrightarrow \infty$ und die Folge (x^k) erfülle $x^k \in \arg\min\limits_{x\in\mathbb{R}^n} p(x,\mu_k)$.*

Ist die Menge M_{opt} der globalen Minima von (3) nicht leer, so gehört jeder Häufungspunkt von (x^k) zu M_{opt}.

Beweis: Es sei ein Punkt $\bar{x} \in M_{opt}$ beliebig fest gewählt. x^* sei ein beliebiger Häufungspunkt der Folge (x^k) (sofern ein solcher existiert), dann existiert eine Teilfolge (x^{k_i}) mit $x^{k_i} \longrightarrow x^*$. Aus der Voraussetzung, daß für jedes k gilt $p(x^k,\mu_k) \leq p(x,\mu_k) \quad \forall x \in \mathbb{R}^n$, folgt insbesondere

$$p(x^{k_i},\mu_{k_i}) \leq p(\bar{x},\mu_{k_i}) \quad \forall i \qquad (*)$$

Aus der Definition 4.5 iii) folgt

$$\lim_{i\longrightarrow\infty} p(\bar{x},\mu_{k_i}) = f(\bar{x})$$

und mit (*) zusammen ergibt sich

$$\limsup p(x^{k_i},\mu_{k_i}) \leq \lim_{i\longrightarrow\infty} p(\bar{x},\mu_{k_i}) = f(\bar{x})$$

Aus Definition 4.5 ii) folgt wegen

$$x^{k_i} \longrightarrow x^*, \quad \mu_{k_i} \longrightarrow \infty \quad \text{und} \quad \liminf p(x^{k_i},\mu_{k_i}) \leq \limsup p(x^{k_i},\mu_{k_i}) \leq f(\bar{x}) < \infty,$$

daß $x^* \in M$. Weiterhin ergibt sich aus dieser Forderung an die Straffunktion p und der obigen Beobachtung

$$f(x^*) \leq \liminf p(x^{k_i},\mu_{k_i}) \leq f(\bar{x})$$

woraus wegen der Wahl vom $\bar{x}$ und der Zulässigkeit von x^* folgt $x^* \in M_{opt}$ □

Es kann gezeigt werden, daß exakte Minimierung in den freien Problemen nicht erforderlich ist. Die Forderung $\tau_k \longrightarrow 0$ reicht aus, um zu garantieren, daß jeder Häufungspunkt der Folge (x^k), in dem die LICQ erfüllt ist, die KKT-Bedingungen erfüllt.

Ein Problem des Strafmethoden-Ansatzes ist, daß nur im Grenzwert $\mu \longrightarrow \infty$ das Minimum der Straffunktion eine Lösung des Ausgangsproblems liefert. Damit ist eine Folge von freien Optimierungsproblemen zu lösen (d.h. in jedem Schritt ein nichtabbrechender Iterationsprozeß). Dieses Problem wird vermieden, wenn man **exakte Straffunktionen** verwendet. Dies sind solche, für welche bereits bei einem endlichen Wert des Strafparameters μ das freie Minimum eine Lösung des Ausgangsproblems liefert. Dies bedeutet, daß die Lösung eines einzigen freien Optimierungsproblems ausreicht. Dies unter der Annahme, man würde die dafür notwendige Größe des Strafparameters bestimmen können. Eine exakte Straffunktion ist z.B. die l_1-Straffunktion

$$p(x,\mu) = f(x) + \mu(\sum_{i=1}^{m} g_i^+(x) + \sum_{j=1}^{l} |h_j(x)|)$$

Es kann gezeigt werden, daß für diese Funktion der Strafparameter größer als die ∞-Norm des optimalen Lagrangemultiplikator-Vektors zu wählen ist. Wäre dieser bekannt oder wenigstens eine gute Schätzung dafür, so könnte man also den Strafparameter so wählen, daß die Probleme äquivalent sind. Diese Straffunktion bringt aber ein anderes Problem für die freie Minimierung - sie ist nicht differenzierbar. Es gibt auch differenzierbare Ansätze für exakte Straffunktionen, welche in enger Beziehung zu den *augmented* Lagrangefunktionen stehen, auf welche wir noch kurz eingehen werden.

4.3.2 Barrieremethoden

Im Gegensatz zu den im vorigen Punkt betrachteten *äußeren Strafmethoden* wird bei den Barriere-Methoden oder *inneren Strafmethoden* eine Folge von freien Optimierungsproblemen zugeordnet, welche stets zulässige Punkte liefern. Dazu betrachten wir Probleme, welche nur Ungleichungsrestriktionen enthalten.

Definition 4.7 *M bezeichne die Restriktionsmenge von (3) mit $l = 0$, es sei*

$$\tilde{M} := \{x \in \mathbb{R}^n : \ g_i(x) < 0, \ \ i = 1, \ldots, m\} \neq \emptyset$$

Eine über $\tilde{M} \times \mathbb{R}_+$ definierte Funktion $b(x, \mu)$ heißt Barrierefunktion für (3), wenn

i) $b(x, \mu)$ stetig in $x \in \tilde{M} \quad \forall \mu > 0$

ii) für jede Folge (x^k) mit $x^k \in \tilde{M}$, $x^k \longrightarrow x^B$ und $\exists i_0 : \ g_{i_0}(x^B) = 0$ gilt für alle $\mu > 0$

$$\lim_{k \longrightarrow \infty} b(x^k, \mu) = +\infty$$

iii) für $x \in \tilde{M}$ beliebig gilt: $\mu_1 > \mu_2 > 0 \Longrightarrow b(x, \mu_1) > b(x, \mu_2) \geq f(x)$ und für jede Folge (μ_k) mit $\mu_k \longrightarrow +0$ gilt $\lim\limits_{k \longrightarrow \infty} b(x, \mu_k) = f(x)$.

Beispiele: a) $b_1(x, \mu) = f(x) - \mu \sum\limits_{i=1}^{m} ln(-g_i(x))$ *(logarithmische Barrierefunktion)*

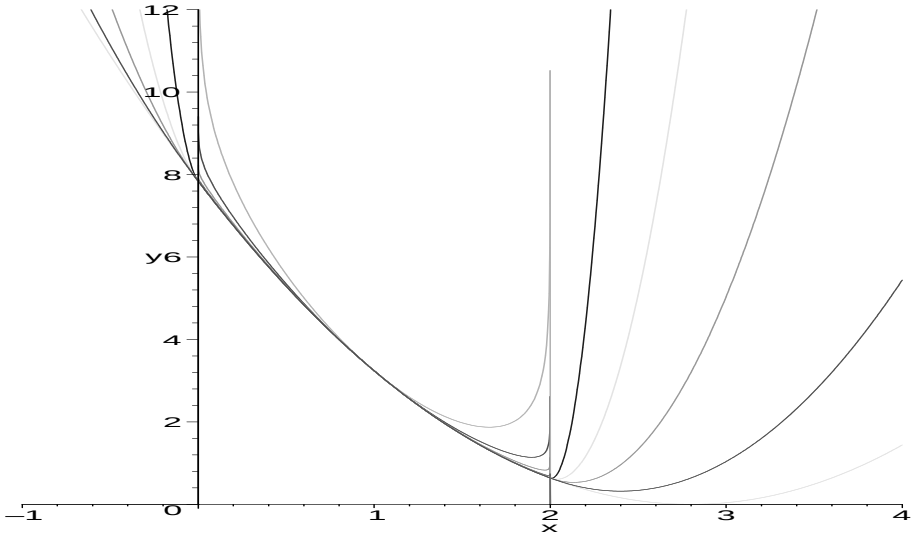
b) $b_2(x, \mu) = f(x) - \mu \sum\limits_{i=1}^{m} \frac{1}{g_i(x)}$

Durch ihre Konstruktion erzwingt eine Barrierefunktion die Generierung von Punkten in $\tilde{M}$ bei freier Minimierung, da der Funktionswert bei Annäherung an den Rand von M gegen $+\infty$ geht.

Das Verhalten der quadratischen Straffunktion und der logarithmischen Barrierefunktion sollen graphisch am Problem

$$\min\{(x - 2.8)^2 : \quad \left.\begin{array}{rl} -x & \leq 0 \\ x - 2 & \leq 2 \end{array}\right\}$$

veranschaulicht werden. Die Zielfunktion und die beiden Funktionen werden für verschiedene Straf- bzw. Barriereparameter dargestellt (Parameter $\mu = 1, \ 5, \ 20, \ 100$ für die Straffunktion bzw. deren Reziproke für die Barrierefunktion).



Der prinzipielle Algorithmus gleicht dem im vorigen Abschnitt, außer daß der Barriereparameter gegen Null streben muß, d.h. es ist zu wählen $\mu_{k+1} \in]0, \mu_k[$.
Es sind im Prinzip Algorithmen der freien Minimierung anwendbar, da das Minimum stets in $\tilde{M}$ liegt. Das Problem der zunehmend schlechteren Konditionierung der Probleme gilt genauso wie

bei den Strafmethoden.
Prinzipiell ist das Verfahren jedoch für konvexe Probleme mit beschränkter Optimalmenge erfolgreich, wir wollen den Satz über die Eigenschaften der logarithmischen Barrierefunktion ohne Beweis zitieren. (siehe *M. Wright, "Interior methods for constrained optimization", Acta Numerica 1992, Cambridge Univ. Press, S. 341-401*)

Satz 4.8 *Es seien (3) mit $l = 0$, $\tilde{M} \neq \emptyset$ und die Menge der Lösungen M_{opt} sei nichtleer und beschränkt. Der Optimalwert sei mit f^* bezeichnet.*
Die Funktionen f, g_i $i = 1, \ldots, m$ seien konvex und zweimal stetig differenzierbar. Die Folge (μ_k) genüge $\mu_k > 0 \; \forall k$ und $\mu_k \longrightarrow +0$. Dann gilt

i) Für jedes $\mu > 0$ ist die logarithmische Barrierefunktion $b_1(x, \mu)$ konvex über $\tilde{M}$ und hat ein (globales, nicht notwendig eindeutiges) Minimum $x(\mu)$ über $\tilde{M}$.

ii) Jede Folge von Minima $x(\mu_k)$ hat eine konvergente Teilfolge, alle ihre Häufungspunkte liegen in M_{opt}.

iii) $f\left(x(\mu_k)\right) \longrightarrow f^*$ und $b_1\left(x(\mu_k), \mu_k\right) \longrightarrow f^*$ für jede Folge von Minima $(x(\mu_k))$.

Die Funktion b_2 ist z.B. in *A.V. Fiacco, G.P. McCormick "Nonlinear Programming: Sequential Unconstrained Minimization Techniques", Wiley 1968* diskutiert.

4.3.3 Multiplikatormethoden - augmented Lagrangians

Der nun zu diskutierende Ansatz stellt eine Kombination aus der Lagrangefunktion und quadratischer Straffunktion dar. Die Lagrangefunktion wird mit einem quadratischen Strafterm ergänzt (engl.: augmented). Auf diese Weise werden die Eigenschaften der Lagrangefunktion verbessert und die Nachteile der schlechten Konditionierung der reinen Strafmethode durch die Verwendung expliziter Näherungen für die Lagrangemultiplikatoren vermieden. Andererseits wird ein glattes Problem gelöst, was leichter ist als die Verwendung der exakten l_1-Straffunktion. Algorithmen können so mit Standard-Bausteinen der differenzierbaren Optimierung implementiert werden. Eine erfolgreiche Implementation auf Basis dieses Ansatzes ist **LANCELOT**.

Bei der näherungsweisen Minimierung der quadratischen Straffunktion werden nicht zulässige Punkte erzeugt und es gilt für jede konvergente Teilfolge, falls im Grenzpunkt die LICQ erfüllt ist

$$\begin{array}{rcll} \lim\limits_{i \longrightarrow \infty} 2\mu_{k_i} g_j^+(x^{k_i}) & = & u_j^* & j = 1, \ldots, m \\ \lim\limits_{i \longrightarrow \infty} 2\mu_{k_i} h_j(x^{k_i}) & = & v_j^* & j = 1, \ldots, l \end{array}$$

so daß die Produkte Näherungen für die Lagrangemultiplikatoren liefern. Dies sieht man wie folgt: O.B.d.A. konvergiere die gesamte Folge, x^* sei der Grenzwert und alle Funktionen seien zweimal differenzierbar. Im freien Minimum x^k der quadratischen Straffunktion mit Strafparameter μ_k ist die notwendige Bedingung erfüllt

$$0 = \nabla_x Q(x^k, \mu_k) = \nabla f(x^k) + \sum_{i=1}^{m} (2\mu_k g_i^+(x^k)) \nabla g_i(x^k) + \sum_{j=1}^{l} (2\mu_k h_j(x^k)) \nabla h_j(x^k)$$

Für $i \notin I(x^*)$ gilt dann für hinreichend große k, daß $g_i(x^k) < 0$, also $g_i^+(x^k) = 0$, so daß in der Summe nur die Ungleichungen mit $i \in I(x^*)$ vorkommen. Vergleicht man diese Gleichung mit der Gleichung aus den KKT-Bedingungen (unter der Voraussetzung der LICQ im Grenzpunkt ist diese in einem glatten Problem für hinreichend große k ebenso erfüllt, unter LICQ sind die Lagrangemultiplikatoren eindeutig bestimmt), so erhält man

$$\begin{array}{rcll} u_i^k & = & 2\mu_k g_i^+(x^k) & i \in I(x^*) \\ u_i^k & = & 2\mu_k g_i^+(x^k) = 0 & i \notin I(x^*) \\ v_j^k & = & 2\mu_k h_j(x^k) & j = 1, \ldots, l \end{array}$$

und es gilt insbesondere $u_i^k > 0$ für $i \in I(x^*)$ und k hinreichend groß. Wenn alle Funktionen stetig differenzierbar sind, gilt $u_i^k \longrightarrow u_i^* \; i \in I(x^*)$, $v_j^k \longrightarrow v_j^* \; j = 1, \ldots, l$.

Die augmented Lagrangefunktion lautet für das Problem mit Gleichungsrestriktionen

$$\mathcal{L}(x,\lambda,\mu) = f(x) + \sum_{j=1}^{l} \lambda_j h_j(x) + \mu \sum_{j=1}^{l} h_j^2(x)$$

Differentiation nach x liefert

$$\nabla_x \mathcal{L}(x,\lambda,\mu) = \nabla f(x) + \sum_{j=1}^{l} [\lambda_j + 2\mu h_j(x)] \nabla h_j(x)$$

Im Algorithmus wird nun in jeder Iteration ein Strafparameter μ_k und eine Näherung λ^k fixiert. Ist dann x^k ein näherungsweises Minimum der Funktion $\mathcal{L}(.,\lambda^k,\mu_k)$, so gilt für den Vektor der optimalen Lagrangemultiplikatoren λ^*

$$\lambda_j^* \approx \lambda_j^k + 2\mu_k h_j(x^k) \quad j = 1,\ldots,l$$

und damit

$$h_j(x^k) \approx \frac{1}{2\mu_k}(\lambda_j^* - \lambda_j^k) \quad j = 1,\ldots,l$$

Daraus folgt, daß bei guter Näherung λ^k die Unzulässigkeit deutlich kleiner ist als bei der quadratischen Straffunktion, wo sie proportional $\frac{\lambda_j^*}{\mu_k}$ ist.
Nun stellt sich die Frage, wie mittels der vorliegenden Informationen in der nächsten Iteration die Näherung für die Lagrangemultiplikatoren verbessert werden kann. Die obige Beziehung legt folgende Vorschrift nahe

$$\lambda_j^{k+1} = \lambda_j^k + 2\mu_k h_j(x^k) \quad j = 1,\ldots,l$$

Damit haben wir das folgende Algorithmenschema

Gegeben: Startpunkte x_S^0, λ^0, Strafparameter μ_0, Toleranz τ_0
FOR $k = 0, 1, \ldots$
Bestimme mit Startpunkt x_S^k näherungsweise ein Minimum x^k für $\mathcal{L}(x,\lambda^k,\mu_k)$
mit Abbruchkriterium $\|\nabla_x \mathcal{L}(x^k,\lambda^k,\mu_k)\| < \tau_k$
Teste auf Konvergenz.
Wenn Konvergenztest erfüllt, STOP mit Näherungslösung x^k
Berechne neue Schätzung $\lambda^{k+1} = \lambda^k + 2\mu_k h(x^k)$,
Wähle neuen Startpunkt $x_S^{k+1} = x^k$,
neue Toleranz $\tau_{k+1} < \tau_k$
und Strafparameter $\mu_{k+1} > \mu_k$
END FOR

Es kann nun gezeigt werden, daß das Verfahren bereits mit endlichem Strafparameter konvergiert, wenn die hinreichende KKT-Bedingungen zweiter Ordnung im Grenzpunkt erfüllt sind. Somit kann diese Funktion auch als eine exakte Straffunktion bezeichnet werden und ist ein Beispiel für eine glatte exakte Straffunktion. Im Gegensatz zu den vorher betrachteten Straffunktionsmethoden wird hier auch eine Näherung für die optimalen Lagrangemultiplikatoren ermittelt.

Ungleichungsrestriktionen können im Rahmen der augmented Lagrangians einbezogen werden, indem sie mit Schlupfvariablen als Gleichungen geschrieben werden. Das Verfahren kann nun wie im Programm LANCELOT so modifiziert werden, daß Box-Constraints mit berücksichtigt werden. Um die Vorzeichenbeschränkungen an die Schlupfvariablen zu vermeiden, kann man die Ungleichungen $g_i(x) \leq 0 \;\; i = 1,\ldots,m$ auch in der Form

$$g_i(x) + s_i^2 = 0 \quad i = 1,\ldots,m$$

schreiben. Existiert dann ein KKT-Punkt $(\bar{x},\bar{u},\bar{v})^T$für das Problem (3), in welchem die hinreichende Bedingung zweiter Ordnung erfüllt ist und **strikte Komplementarität** gilt, d.h.

$$\bar{u}_i g_i(\bar{x}) = 0 \;\; i = 1,\ldots,m \;\text{ und }\; \bar{u}_i > 0 \;\; i \in I(\bar{x}),$$

dann gilt auch für diesen Ansatz, daß für hinreichend großes μ die Lösung $(\bar{x},\bar{s})^T$ ein striktes

lokales Minimum der zugeordneten augmented Lagrangian ist. Die Update-Formel für die Lagrange-Multiplikatoren zu Ungleichungen ergibt sich zu

$$u_i^{k+1} = u_i^k + \max\{2\mu_k g_i(x^k), -u_i^k\} \quad i = 1, \ldots, m$$

4.3.4 Sequential Quadratic Programming (SQP)-Verfahren

Eine sehr erfolgreiche Methode beruht auf der iterativen Lösung quadratischer Approximationen des Optimierungsproblems (d.h. quadratische Zielfunktion und affine Restriktionen). Dies entspricht einer quadratischen Approximation der Zielfunktion und Linearisierung der Restriktionsfunktionen in jedem Iterationspunkt. Anders interpretiert ist dies die Anwendung des Newtonverfahrens oder von Quasi-Newtonverfahren auf das KKT-System.

Zur Darstellung des Prinzips verwenden wir ein Problem mit Gleichungsrestriktionen ($m = 0$ in (3)). Die Problemfunktionen f und h_j, $j = 1, \ldots, l$ seien zweimal stetig differenzierbar.

Zunächst zeigen wir die Interpretation über das Newtonverfahren. Die KKT-Bedingungen für dieses Problem fordern die Existenz von Vektoren $x \in \mathbb{R}^n$ und $v \in \mathbb{R}^l$, so daß

$$\begin{aligned} \nabla f(x) + \sum_{j=1}^{l} v_j \nabla h_j(x) &= 0 \\ h_j(x) = 0 \quad j &= 1, \ldots, l \end{aligned} \tag{21}$$

Dieses System schreiben wir kurz als $W(x, v) = 0$. Wird nun das Newton-Verfahren auf das Gleichungssystem angewandt, so wird in jedem Schritt eine Approximation erster Ordnung des Gleichungssystems gelöst, d.h.

$$W(x^k, v^k) + \nabla W(x^k, v^k) \begin{pmatrix} x - x^k \\ v - v^k \end{pmatrix} = 0$$

Mit der Hessematrix der Lagrangefunktion $\nabla^2_{xx} L(x, v)$ und der Jacobi-Matrix der Restriktionsfunktionen $\nabla h(x)$ ist

$$\nabla W(x, v) = \begin{pmatrix} \nabla^2_{xx} L(x, v) & \nabla h(x) \\ \nabla h(x) & 0 \end{pmatrix}$$

und das Gleichungssystem aus dem Newton-Verfahren wird mit $d = x - x^k$ zu

$$\begin{aligned} \nabla^2_{xx} L(x^k, v^k) d \quad + \nabla h(x^k) v \quad &= -\nabla f(x^k) \\ \nabla h(x^k) d \qquad\qquad\quad &= -h(x^k) \end{aligned} \tag{22}$$

Unter der Annahme, daß auch in den Iterationspunkten LICQ erfüllt ist und die Hessematrix der Lagrangefunktion positiv definit ist über dem Tangentialraum der Restriktionen, ist dieses System lösbar, die Lösung bezeichnen wir mit (d^k, v^{k+1}). Die Annahmen sind erfüllt in der Nähe einer Lösung, in der LICQ und hinreichende Bedingung zweiter Ordnung erfüllt sind. Mit $x^{k+1} = x^k + d$ setzen wir dann $k := k + 1$ und setzen die Iteration fort, bis $\|d\| < \varepsilon$.

Diese Iteration kann alternativ auch aufgefaßt werden als Lösung eines quadratischen Optimierungsproblems, dessen Optimalitätsbedingungen (22) sind:

$$\min \left\{ \frac{1}{2} d^T \nabla^2_{xx} L(x^k, v^k) d + \nabla f(x^k) d : \; h_j(x^k) + \nabla h_j(x^k) d = 0 \quad j = 1, \ldots, l \right\}$$

Die Zielfunktion dieses Problems stellt nicht nur eine quadratische Approximation der Zielfunktion des Originalproblems im Punkt x^k dar, sondern enthält noch den Term $\frac{1}{2} \sum_{j=1}^{l} d^T \nabla^2 h(x^k) d$, welcher die Krümmung der Restriktionsfunktionen im Iterationspunkt widerspiegelt. Andererseits kann in der Zielfunktion der Term $\nabla f(x^k) d$ äquivalent durch $\nabla_x L(x^k, v^k)^T d = (\nabla f(x^k) + {v^k}^T \nabla h(x^k)) d$ ersetzt werden, denn die Restriktion liefert $\nabla h_j(x^k) d = -h_j(x^k)$ und somit wird durch die Substitution nur der konstante Term ${v^k}^T h(x^k)$ zur Zielfunktion hinzugefügt. Die Zielfunktion ist also eine quadratische Approximation der Lagrangefunktion und das Problem stellt somit eine Approximation des Problems der Minimierung der Lagrangefunktion bzgl. der Gleichungsrestriktionen dar.

Unter den angegebenen Voraussetzugen hat dieses Problem eine eindeutig bestimmte Lösung (x^{k+1}, v^{k+1}), welche dem Gleichungssystem (22) genügt. Aus dieser Interpretation stammt der Name SQP - sequentielle quadratische Programmierung (=Optimierung).
Es liegt so eine Äquivalenz zwischen der Anwendung des Newtonverfahrens und der sequentiellen Lösung quadratischer Optimierungsprobleme vor. Der Blickwinkel des Newtonverfahrens ist geeignet für die Konvergenzanalyse und liefert lokale quadratische Konvergenz des Verfahrens unter den Annahmen und bei stetiger Differenzierbarkeit von f und allen h_j mit Lipschitzstetigen zweiten Ableitungen.
Der Blickwinkel der sequentiellen Lösung quadratischer Probleme ist geeignet für die Ableitung praktischer Algorithmen und zur Ausdehnung auf den allgemeinen Fall mit Ungleichungsrestriktionen.

Ungleichungsrestriktionen lassen sich in dieses quadratische Problem analog linearisiert aufnehmen:

$$\min \quad \{\tfrac{1}{2} d^T \nabla^2_{xx} L(x^k, , u^k, v^k) d + \nabla f(x^k) d : \quad \left.\begin{array}{ll} g_i(x^k) + \nabla g_i(x^k) d \leq 0 & i = 1, \ldots, m \\ h_j(x^k) + \nabla h_j(x^k) d = 0 & j = 1, \ldots, l \end{array}\right\} \tag{23}$$

Lokaler SQP-Algorithmus:

Gegeben: Startpunkt und Start-Näherung für Lagrangevektor x^0, u^0, v^0
FOR $k = 0, 1, \ldots$
Löse Problem (23), seine Lösung sei d^k
Setze $x^{k+1} = x^k + d^k$ und u^{k+1}, v^{k+1} gleich Näherungen der optimalen Lagrangeparameter von (23)
Teste auf Konvergenz.
Wenn Konvergenztest erfüllt, STOP mit Näherungslösung x^{k+1} und Näherungen u^{k+1}, v^{k+1}.
END FOR

Unter Annahme strikter Komplemetarität läßt sich zeigen, daß im Verfahren in der Nähe einer Lösung die aktive Indexmange konstant bleibt. Das Problem verhält sich dann genauso als wären die aktiven Restriktionen als Gleichungen vorhanden, es ergibt sich also auch quadratische Konvergenz.

Die Lösung der jeweiligen Probleme erfordert die Berechnung der Hessematrix der Lagrangefunktion und die Existenz einer Lösung ist nur dann sicher, wenn diese positiv definit auf dem Tangetialraum der aktiven Restriktionen ist. Selbst wenn wir das im Lösungspunkt fordern, braucht dies in den Iterationspunkten nicht unbedingt zu gelten. Eine mögliche Abhilfe dafür ist die Verwendung von Quasi-Newton-Approximationen für $\nabla^2_{xx} L(x^k, v^k)$. Eine andere Möglichkeit ist, die Lagrangefunktion durch eine augmented Lagrangefunktion zu ersetzen und so gewisse Konvexitätseigenschaften zu sichern. Noch ein andrer Ansatz ist die Verwendung von sog. *trust region* Ansätzen, bei welchen die Schrittweite zusätzlich beschräkt wird. Damit wird der neue Iterationspunkt in einer Umgebung von x^k gewählt, in welcher das quadratische Modell als zuverlässig (*to trust* = vetrauen) betrachtet wird. Damit wird das Verfahren auch bei nicht positiv definiten Hessematrizen anwendbar. Probleme können dadurch entstehen, dass durch die zusätzliche Restriktion unlösbare Teilaufgaben entstehen können.

Ein anderer Punkt, welcher weitere Änderungen an Algorithmus erfordert, ist die zunächst nur lokale Konvergenz des Verfahrens. Durch die Verwendung einer geeigneten Merit-Funktion (z.B. der l_1-Straffunktion) und die Modifikation des Updates $x^{k+1} = x^k + d^k$ zu einer eindimensionalen Minimierung dieser Merit-Funktion über den Punkten der Form $x^k + \lambda d^k, \quad \lambda \geq 0$ läßt sich globale Konvergenz des Verfahrens erreichen.

4.4 Quasi-Newton-Approximation - die BFGS-Formel

An verschiedenen Stellen wurden im vorangehenden Quasi-Newton Approximationen der Hessematrix erwähnt. Nun soll dargestellt werden, welche Idee diesen Approximationen zugrunde liegt und wie sie im Verlaufe eines Verfahrens gewonnen werden (entnommen *Nodedal/Wright*).
Die Grund-Idee wurde Mitte der 50er Jahre vom Physiker W.C. Davidon am Argonne National Laboratory entwickelt. Er versuchte, ein großes Optimierungsproblem mit einem Abstiegsvergfahren zu lösen. Dioe Computer jener Zeit liefen jedoch noch nicht sehr stabil und es kam immer zu einem Crash, ehe das Programm zum Ende gekommen war. So suchte Davidon nach einer Möglichkeit, das Verfahren zu beschleunigen. Der von ihm entwickelte erste Quasi-Newton Algorithmus erwies sich als eine der revolutionärsten Entwicklungen in der nichtlinearen Optimierung dieser Zeit. Von Fletcher und Powell wurde gezeigt, daß der neue Algorithmus schneller und zuverlässiger war als die anderen existierenden Methoden. Eine Ironie der Geschichte ist, daß Davidon's Artikel nicht zur Publikation angenommen wurde und für mehr als 30 Jahre nur als Technical Report verfügbar war, bis er 1991 im *SIAM Journal on Optimization* erschien. Da die Methode durch die Untersuchungen Fletcher und Powell bekannt gemacht wurde, ist die ursprüngliche Form als **DFP**-Formel (**D**avidon/**F**letcher/**P**owell) bekannt geworden. Eine später durch die Autoren **B**royden, **F**letcher, **G**oldfarb und **S**hanno etwickelte Variante ist populärer geworden und trägt die Bezeichnung **BFGS**-Methode.

Die Idee geht aus von einem streng konvexen quadratischen Modell der Zielfunktion im aktuellen Iterationspunkt x^k:

$$m_k(p) = f(x^k) + \nabla f(x^k)p + \frac{1}{2}p^T B_k p$$

Hierbei ist B_k eine positiv definite symmetrische Matrix, welche in jeder Iteration mittels eine Update-Formel bestimmt wird. Für $p = 0$ stimmen Funktionswert und Gradient des Modells mit denen der Originalfunktion f überein.
Das Minimum dieser Funktion $m_k(.)$ kann explizit bestimmt werden

$$p^k = -B_k^{-1}\nabla f(x^k) \tag{24}$$

und wird als Suchrichtung verwendet und der neue Iterationspunkt durch Richtungssuche in der Form

$$x^{k+1} = x^k + \alpha_k p^k$$

bestimmt, wobei die Schrittweite den Goldstein-Armijo-Bedingungen genügt. Der Unterschied zur Newton-Methode besteht in der Verwendung einer positiv definiten Approximation B_k anstelle der Hessematrix von f in x^k.
Die Matrix B_k wird in jeder Iteration modifiziert. Davidon schlug vor, sie auf einfache Art so zu modifizieren, daß sie die Krümmung der Zielfunktion im aktuellen Schritt widerspiegelt. Eine naheliegende Forderung an das neue Modell

$$m_{k+1}(p) = f(x^{k+1}) + \nabla f(x^{k+1})^T p + \frac{1}{2}p^T B_{k+1} p$$

ist, daß der Gradient von m_{k+1} mit dem von f in den beiden Iterierten x^k und x^{k+1} übereinstimmen soll. Die zweite Bedingung ist durch die Form von $m_G k + 1$ gesichert, die erste Forderung liefert die Gleichung

$$\nabla f(x^{k+1}) - \alpha_k B_{k+1} p^k = \nabla f(x^k)$$

Umstellen liefert mit $s^k = x^{k+1} - x^k$, $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ die Sekantengleichung

$$B_{k+1} s^k = y^k \tag{25}$$

Diese Gleichung ist nur lösbar, wenn die *Krümmungsbedingung* ${s^k}^T y^k > 0$ gilt (Gleichung von links mit ${s^k}^T$ multiplizieren). Diese Bedingung ist für streng konvexes f stets erfüllt, im allgemeinen Fall muß diese durch die Schrittweitenwahl gesichert werden. Die zweite Goldstein-Armijo-Bedingung liefert $\nabla f(x^k + \alpha p^k)^T s^k \geq \sigma \nabla f(x^k) p^k$ und somit

$${y^k}^T s^k \geq (\sigma - 1)\nabla f(x^k) p^k$$

Wegen $\sigma < 1$ und da p^k eine Abstiegsrichtung ist, ist der Term auf der rechten Seite positiv, die Krümmungsbedingung ist also erfüllt.
Ist diese Bedingung erfüllt, so hat (25) immer eine Lösung. Das Gleichungssystem ist jedoch nicht eindeutig lösbar, da es $n(n+1)/2$ Freiheitsgrade hat, das Gleichungssystem liefert n Gleichungen. Die positive Definitheit liefert weitere n Bedingungen. Es gibt also unendlich viele Lösungen.
Um die Matrix B_{k+1} eindeutig zu bestimmen, wird eine weitere Forderung gestellt - daß sie in gewissem Sinne unter allen in Frage kommenden Matrizen den Abstand zu B_k minimiert. Es wird das Problem

$$\min_B \{\|B - B_k\| : \ B = B^T, \ Bs^k = y^k, \ B \text{ positiv definit}\} \tag{26}$$

In Abhängigkeit von der gewählten Matrix-Norm ergeben sich verschiedene Quasi-Newton-Methoden. Wird eine gewichtete Frobenius-Norm

$$\|A\|_W = \|W^{1/2} A W^{1/2}\|_F$$

mit $\|C\|_F = \sum_{i=1}^n \sum_{j=1}^n c_{ij}^2$. Die Gewichtsmatrix erfülle die Bedingung $Ws^k = y^k$. Dies kann z.B. $W = \bar{G}_k^{-1}$ mit der mitteleren Hessematrix

$$\bar{G}_k = \int_0^1 \nabla^2 f(x^k + \tau \alpha_k p^k) d\tau$$

sein. Mit dieser Norm und dieser Gewichtung ist die Norm skalierungsinvariant und die eindeutige Lösung des Problems (26) ist

$$(DFP) \quad B_{k+1} = (I - \gamma_k y^k s^{k^T}) B_k (I - \gamma_k y^k s^{k^T}) + \gamma_k y^k y^{k^T}$$

mit

$$\gamma_k = \frac{1}{y^{k^T} s^k}$$

Berechnet man nicht B_k, sondern direkt deren Inverse $H_k = B_k^{-1}$, so kann die Methode leicht implemetiert werden, da die Bestimmung der Suchrichtung gemäß (24) eine einfache Matrix-Vektor-Multiplikation wird. Unter Benutzung der Sherman-Morrison-Formel läßt sich ein expliziter Ausdruck zur Berechnung von H_{k+1} aus H_k gewinnen. Für eine Rang-1-Änderung lautet diese:

Satz 4.9 *Die $n \times n$ Matrix A sei regulär und es wird die Rang-1-Änderung zu $A' = A + uv^T$ mit $u, v \in \mathbb{R}^n$ betrachtet. Ist dann $1 + v^T A^{-1} u \neq 0$ so ist A' regulär und es gilt*

$$(A')^{-1} = A^{-1} - \frac{1}{1 + v^T A^{-1} u} A^{-1} u v^T A^{-1}$$

Beweis: Der Beweis erfolgt einfach durch Ausmultiplizieren,

$$(A + uv^T)(A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + v^T A^{-1} u}) = I + uv^T A^{-1} - \frac{uv^T A^{-1}}{1 + v^T A^{-1} u} - \frac{u(v^T A^{-1} u) v^T A^{-1}}{1 + v^T A^{-1} u} = I$$

Dabei wird der Skalar $(v^T A^{-1} u)$ vor das Matrizenprodukt gezogen. □

Die Formel läßt sich verallgemeinern für Änderungen vom Rang p mit $1 \leq p \leq n$

Satz 4.10 *Die $n \times n$ Matrix A sei regulär. Es seien U und V $n \times p$-Matrizen. Ist dann $(I + V^T A^{-1} U)$ regulär, so auch $A' = A + UV^T$ und es gilt*

$$(A')^{-1} = A^{-1} - A^{-1} U (I + V^T A^{-1} U)^{-1} V^T A^{-1}$$

Mit Hilfe dieser Formel läßt sich die DFP-Update-Formel für H_k ableiten:

$$(DFP) \quad H_{k+1} = H_k - \frac{1}{y^{k^T} H_k y^k} H_k y^k y^{k^T} H_k + \frac{s^k s^{k^T}}{y^{k^T} s^k} \tag{27}$$

Die BFGS-Methode unterscheidet sich nun in der Festlegung des Problems zur Bestimmung der Matrix. Statt die Matrix B nahe B_k zu wählen, wird dies hier für die Inversen getan:

$$\min_H \{\|H - H_k\| : \ H = H^T, \ Hy^k = s^k, \ H \text{ positiv definit}\} \tag{28}$$

Die eindeutige Lösung dieses Problem (wieder mit der Frobenius-Norm) ist nun

$$(BFGS) \quad H_{k+1} = (I - \gamma_k y^k {s^k}^T) H_k (I - \gamma_k y^k {s^k}^T) + \gamma_k s^k {s^k}^T$$

mit $\gamma_k = \frac{1}{{y^k}^T s^k}$ wie oben.
Die einzige noch zu klärende Frage ist nun: wie fimdet man eine erste Approximation H_0? Es gibt keine Wahl, die in allen Fällen gut funktioniert. Oft wird einfach die Einheitsmatrix eingesetzt. Man kann auch spezifische Probleminformationen einsetzen und z.B. mittels finiter Differenzen eine Näherung für die Hessematrix bestimmen (positive Definitheit ist evtl. durch Modifikation der Hauptdiagonalen zu erzwingen).