Nichtlineare Optimierung

HP Butzmann

Vorlesung im HWS 10

Inhaltsverzeichnis

Kapitel 1

Einführung

Definition 1.1 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$, $f : \mathcal{D} \to \mathbb{R}$ eine Abbildung und $K \subseteq \mathcal{D}$. Der formale Ausdruck*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \end{array}$$

*heißt **Minimierungsproblem** (MP). Ein Punkt $x^* \in K$ heißt **Lösung** von (MP), wenn gilt:*

$$f(x^*) \leq f(x) \qquad \textit{für alle } x \in K \ .$$

Also ist $x^* \in K$ genau dann Lösung von (MP), wenn gilt $f(x^*) = \min f(K)$.

*Man nennt K den **zulässigen Bereich**, die Punkte aus K **zulässige** Punkte und f die **Zielfunktion**.*

Bemerkung 1.2 Es seien $K \subseteq \mathcal{D} \subseteq \mathbb{R}^n$ und $f : \mathcal{D} \to \mathbb{R}$ eine Abbildung. Weiter sei

$$g : D \longrightarrow \mathbb{R}$$

definiert durch $g(x) = -f(x)$. Dann existiert $\max f(K)$ genau dann, wenn $\min g(K)$ existiert und es gilt für alle $x^* \in K$:

$$f(x^*) = \max f(K) \quad \Longleftrightarrow \quad g(x^*) = \min g(K) \ .$$

Also kann man jedes Maximierungsproblem ohne Mühe auf ein Minimierungsproblem zurückführen.

Beispiele 1.3

(i) Es seien A eine reelle Matrix mit p Zeilen und n Spalten, $b \in \mathbb{R}^p$ und $c \in \mathbb{R}^n$. Man setze

$$K = \{x \in \mathbb{R}^n : Ax \leq b\}$$

und definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch

$$f(x) = c^t x$$

Dann heißt das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

linear. Man schreibt dafür üblicherweise

$$\begin{array}{ll} \min & c^t x \\ \text{bez.} & Ax \leq b \end{array}$$

und nennt das MP ein lineares Programm. Die Theorie linearer Programme ist sehr gut entwickelt, zur praktischen Lösung setzt man meistens das Simplex-Verfahren ein. Allerdings betrachtet man mittlerweile sogenannte "Innere-Punkte-Methoden" als echte Alternative dazu.

(ii) Ein neuer Flugplatz mit den Ortskoordinaten $x^* \in \mathbb{R}^2$ soll so gelegt werden, dass einerseits eine gewichtete Summe der Entfernungen zu r Nachbarflugplätzen mit den Ortskoordinaten $x_0, \ldots, x_r$ möglichst klein wird, andererseits s Gebiete, hier durch Kreissscheiben mit den Mittelpunkten $y_1, \ldots, y_s$ und den Radien $r_1, \ldots, r_s$ beschrieben, nicht berührt werden. Zur Lösung dieses Problems seien

$$K = \{x \in \mathbb{R}^n : \|x - y_i\| \geq r_i \text{ für } i = 1, \ldots, s\}$$

sowie $f : K \to \mathbb{R}$ definiert durch

$$f(x) = \sum_{i=1}^{r} \gamma_i \|x - x_i\| \ ,$$

dann ist das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

zu lösen.

(iii) Es seien $[a, b]$ ein reelles Intervall, $a = x_0 < x_0 \cdots < x_k = b$ eine Zerlegung von $[a, b]$ sowie $t_0, \ldots, t_k$ reelle Zahlen. Schließlich sei G ein n-dimensionaler Vektorraum von Abbildungen von $[a, b]$ nach $\mathbb{R}$. Dann ist eine Abbildung $g^* \in G$ so gesucht, dass g^* die Stützpunkte $(x_0, t_0), \ldots (x_k, t_k)$ im Sinne der Methode der kleinsten Quadrate am besten approximiert. Also soll gelten:

$$\sum_{i=0}^{k} (g^*(x_i) - t_i)^2 \leq \sum_{i=0}^{k} (g(x_i) - t_i)^2 \qquad \text{für alle } g \in G$$

Zur Lösung dieses Problems sei $\{g_1, \ldots, g_n\}$ eine Basis von G. Man definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch

$$f(\alpha_1, \ldots, \alpha_n) = \sum_{i=0}^{k} (\sum_{j=1}^{n} \alpha_j g_j(x_i) - t_i)^2$$

Dann ist das MP

$$\begin{array}{ll} \min & f(\alpha) \\ \text{bez.} & \alpha \in \mathbb{R}^n \end{array}$$

zu lösen.

Bezeichnungsweisen 1.4

(i) Es ist in der Optimierung üblich, die Elemente des $\mathbb{R}^n$ als Spaltenvektoren aufzufassen. Abweichend von dieser Konvention werden die Argumente von Abbildungen aus dem $\mathbb{R}^n$ als Zeilenvektoren geschrieben. Wenn also $\mathcal{D} \subseteq \mathbb{R}^n$ gilt und $f : \mathcal{D} \to \mathbb{R}^p$ eine Abbildung ist, schreibt man $f(u_1, \ldots, u_n)$ für alle $(u_1, \ldots, u_n)^t \in \mathcal{D}$.

Ich werde mich an diese Konvention (jedenfalls meistens) halten.

Falls nicht anders bemerkt, trägt der $\mathbb{R}^n$ die **euklidische Norm**, die mit $\|\cdot\|$ bezeichnet wird.

Für alle $(u_1, \ldots, u_n)^t, (v_1, \ldots, v_n)^t \in \mathbb{R}^n$ definiere man

$$(u_1, \ldots, u_n)^t \leq (v_1, \ldots, v_n)^t \quad \Longleftrightarrow \quad u_i \leq v_i \quad \text{für alle } i \ .$$

(ii) Es sei $f : \mathcal{D} \to \mathbb{R}^p$ eine Abbildung, dann bezeichnen $f_1, \ldots, f_p$ die **Komponentenabbildungen** von f, d.h. es gilt

$$f(x) = (f_1(x), \ldots, f_p(x))^t \qquad \text{für alle } x \in \mathcal{D} \ .$$

(iii) Es sei $f : \mathcal{D} \to \mathbb{R}$ eine partiell differenzierbare Abbildung. Dann bezeichnet

$$D_i f = \frac{\partial f_j}{\partial x_i}$$

die **i.-te partielle Ableitung** von f.

(iv) Es sei $f : \mathcal{D} \to \mathbb{R}^p$ eine partiell differenzierbare Abbildung. Dann heißt die Matrix mit p Zeilen und n Spalten

$$Df(x) = (D_i f_j(x))_{j,i}$$

die **Jakobi-Matrix** von f an der Stelle $x \in \mathcal{D}$. Man setzt

$$\nabla f(x) = Df(x)^t \ .$$

Im Fall $p = 1$ heißt $\nabla f(x) = (D_1 f(x), \ldots, D_n f(x))^t$ der **Gradient** von f in x und es gilt offenbar für alle p und alle x:

$$\nabla f(x) = (\nabla f_1(x), \ldots, \nabla f_p(x)) \ .$$

(iv) Es sei $f : \mathcal{D} \to \mathbb{R}$ eine zweimal partiell differenzierbare Abbildung. Dann heißt

$$Hf(x) = (D_{i,j} f(x))$$

die **Hesse-Matrix** von f in x. Man setzt auch $\nabla^2 f(x) = Hf(x)^t$.

Bekanntlich ist die Hesse-Matrix einer zweimal stetig differenzierbaren Abbildung in jedem Punkt symmetrisch.

Erinnerung 1.5 Es sei $\mathcal{D} \subseteq \mathbb{R}^n$ offen.

(i) Es sei $f : \mathcal{D} \to \mathbb{R}^p$ eine differenzierbare Abbildung und $x_0 \in \mathcal{D}$. Dann gibt es eine Abbildung $R : \mathcal{D} \to \mathbb{R}^p$ so dass gelten:

$$\begin{aligned} (a) \quad f(x) &= f(x_0) + Df(x_0)(x - x_0) + R(x) \\ &= f(x_0) + \nabla f(x_0)^t (x - x_0) + R(x) \end{aligned}$$

$$(b) \quad \lim_{x \to x_0} \frac{1}{\|x - x_0\|} R(x) = 0.$$

(ii) Es sei $f : \mathcal{D} \to \mathbb{R}$ eine zweimal stetig differenzierbare Abbildung. Dann gibt es eine Abbildung $R : \mathcal{D} \to \mathbb{R}$ so dass gelten:

$$\begin{aligned} (a) \quad f(x) &= f(x_0) + Df(x_0)(x - x_0) + \tfrac{1}{2}(x - x_0)^t Hf(x_0)(x - x_0) + R(x) \\ &= f(x_0) + \nabla f(x_0)^t (x - x_0) + \tfrac{1}{2}(x - x_0)^t Hf(x_0)(x - x_0) + R(x) \end{aligned}$$

$$(b) \quad \lim_{x \to x_0} \frac{1}{\|x - x_0\|^2} R(x) = 0.$$

Proposition 1.6 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}$ differenzierbar, $x_0 \in \mathcal{D}$ und $a \in \mathbb{R}^n$. Man wähle ein $\varepsilon > 0$ so dass gilt $x_0 + \alpha a \in \mathcal{D}$ für alle $\alpha \in (-\varepsilon, \varepsilon)$. Dann ist die Abbildung $\varphi : (-\varepsilon, \varepsilon) \to \mathbb{R}$ definiert durch*

$$\varphi(\alpha) = f(x_0 + \alpha a)$$

differenzierbar und es gilt

$$\varphi'(\alpha) = Df(x_0 + \alpha a)a = \nabla f(x_0 + \alpha a)^t a \qquad \text{für alle } \alpha \in (-\varepsilon, \varepsilon)$$

Beweis Man definiere $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ durch $\gamma(\alpha) = x_0 + \alpha a$. Dann ist γ differenzierbar und es gilt $D\gamma(\alpha) = a$ für alle α. Weiterhin gilt $\varphi = f \circ \gamma$. Also ist φ differenzierbar und es gilt nach der Kettenregel für alle $\alpha \in (-\varepsilon, \varepsilon)$

$$\varphi'(\alpha) = Df(\gamma(\alpha))D\gamma(\alpha) = Df(x_0 + \alpha a)a$$

■

Kapitel 2

Verfahren (1)

VEREINBARUNG

In diesem Kapitel seien, falls nicht anders bemerkt, $f : \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Abbildung. Betrachtet wird das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in \mathbb{R}^n \end{array}$$

Bemerkung 2.1 Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $f : \mathcal{D} \to \mathbb{R}$ eine differenzierbare Abbildung. Wenn $x^* \in \mathcal{D}$ das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in \mathcal{D} \end{array}$$

löst, gilt $\nabla f(x^*) = 0$. So ein Punkt heißt auch **stationärer** Punkt des MPs und die Verfahren dieses Kapitels suchen einen stationären Punkt.

Definition 2.2 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $f : \mathcal{D} \to \mathbb{R}$ eine Abbildung.*

*(i) Es sei $x_0 \in \mathcal{D}$. Ein Vektor $d \in \mathbb{R}^n$ heißt **Abstiegsrichtung** von f in x_0, wenn es ein $\varepsilon > 0$ so gibt, dass gilt $x_0 + \alpha d \in \mathcal{D}$ für alle $0 \leq \alpha \leq \varepsilon$ und*

$$f(x_0 + \alpha d) < f(x_0) \qquad \text{für alle } 0 < \alpha \leq \varepsilon \, .$$

*(ii) Es seien $K \subseteq \mathcal{D}$ und $x_0 \in K$. Ein Vektor $d \in \mathbb{R}^n$, $d \neq 0$ heißt **zulässige Richtung** in x_0 bzgl. K, wenn es ein $\varepsilon > 0$ so gibt, dass gilt*

$$x_0 + \alpha d \in K \qquad \text{für alle } 0 < \alpha \leq \varepsilon \, .$$

Proposition 2.3 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}$ stetig differenzierbar und $x_0 \in \mathbb{R}^n$ sowie $d \in \mathbb{R}^n$. Dann gelten:*

(i) Es gelte $\nabla f(x_0)^t d < 0$. Dann ist d eine Abstiegsrichtung von f in x_0.

(ii) Es sei d eine Abstiegsrichtung. Dann gilt $\nabla f(x_0)^t d \leq 0$.

Beweis Man wähle ein $r > 0$ mit $x_0 + \alpha d \in \mathcal{D}$ für alle $|\alpha| < r$ und definiere $\varphi : (-r, r) \longrightarrow \mathbb{R}$ durch

$$\varphi(\alpha) = f(x_0 + \alpha d) \ .$$

Dann gilt nach 1.6 für alle α:

$$\varphi'(\alpha) = \nabla f(x_0 + \alpha d)^t d \ ,$$

also folgt

$$\varphi'(0) = (\nabla f(x_0))^t d \ .$$

(i) Wenn $\varphi'(0) = (\nabla f(x_0))^t d < 0$ gilt, ist φ in einer Umgebung von 0 streng monoton fallend, also gibt es ein $0 < \varepsilon \leq r$ so dass gilt

$$f(x_0) = \varphi(0) > \varphi(\alpha) = f(x_0 + \alpha d) \qquad \text{für alle } 0 < \alpha \leq \varepsilon \ .$$

(ii) Wenn d eine Abstiegsrichtung ist, gibt es ein $\varepsilon > 0$ so dass gilt $\varphi(\alpha) < \varphi(0)$ für alle $0 < \alpha < \varepsilon$ und für diese α folgt

$$\frac{\varphi(\alpha) - \varphi(0)}{\alpha - 0} \leq 0$$

und daher $(\nabla f(x_0))^t d = \varphi'(0) \leq 0$. ■

Bemerkung 2.4 *(Abstiegsverfahren) Vorgegeben sei das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \end{array}$$

Die Grundstruktur eines ***Abstiegsverfahrens*** *zur Bestimmung eines stationären Punktes des MPs ist wie folgt:*

(S1) Man wähle $x_0 \in K$ beliebig.

(S2) Es seien $x_0, \ldots, x_k$ bestimmt. Wenn x_k ein stationärer Punkt ist, bricht man das Verfahren ab. Wenn x_k kein stationärer Punkt des MPs ist, sucht man eine zulässige Abstiegsrichtung d_k.

(S3) Man wähle ein $\alpha_k > 0$ so dass gilt $f(x_k + \alpha_k d_k) < f(x_k)$ und setze

$$x_{k+1} = x_k + \alpha_k x_k \ .$$

Unter geeigneten Voraussetzungen kann man beweisen, dass die Folge (x_k) Häufungspunkte besitzt und dass die Häufungspunkte stationäre Punkte des MPs sind.

In (S3) benutzt man oft die folgende Variante:

Es sei

$$I := \{t : x_k + t d_k \in K\} \ ,$$

dann sucht man ein Minimum der Abbildung $\varphi : I \longrightarrow \mathbb{R}$ definiert durch

$$\varphi(\alpha) = f(x_k + \alpha d_k)$$

auf I oder einer geeigneten Teilmenge von I.
Bei einem Abstiegsverfahren bietet es sich an, als Abstiegsrichung (also d_k) die Richtung "mit dem steilsten Abstieg" zu wählen. Dazu muss man die "Steilheit" s_d einer Abstiegsrichtung d definieren. Eine nahliegende Möglichkeit ist die folgende:

$$s_d = \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\|\alpha d\|}$$

Aber diesen Limes kann man ohne Mühe ausrechnen: Definiert man wieder φ durch $\varphi(\alpha) = f(x + \alpha d)$, dann gilt

$$\frac{f(x + \alpha d) - f(x)}{\alpha} = \frac{\varphi(\alpha) - \varphi(0)}{\alpha} \to \varphi'(0) = \nabla f(x)^t d$$

und daher

$$s_d = \frac{1}{\|d\|} \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha} = \frac{1}{\|d\|} \nabla f(x)^t d$$

Lemma 2.5 *Es seien $x_0 \in \mathcal{D}$ und $\nabla f(x_0) \neq 0$. Dann ist $d_0 = -\nabla f(x_0)$ eine zulässige Abstiegsrichtung, in der Tat ist es die Richtung des steilsten Abstiegs in x_0, d.h. es gilt*

$$\frac{1}{\|d_0\|} \nabla f(x_0)^t d_0 \leq \frac{1}{\|d\|} \nabla f(x_0)^t d \qquad \textit{für alle } d \neq 0$$

Beweis Es gilt

$$\begin{aligned} \frac{1}{\|d_0\|} \nabla f(x_0)^t d_0 &= -\frac{1}{\|\nabla f(x_0)\|} \nabla f(x_0)^t \nabla f(x_0) \\ &= -\frac{1}{\|\nabla f(x_0)\|} \|\nabla f(x_0)\|^2 \\ &= -\|\nabla f(x_0)\| < 0 \end{aligned}$$

Also ist d_0 nach 2.3 eine Abstiegsrichtung. Weiter gilt nach der Cauchy-Schwarz'schen Ungleichung gilt für alle $d \in \mathbb{R}^n$:

$$|\frac{1}{\|d\|} \nabla f(x_0)^t d| \leq \frac{1}{\|d\|} \|\nabla f(x_0)\| \, \|d\| = \|\nabla f(x_0)\|$$

und daher

$$\frac{1}{\|d_0\|} \nabla f(x_0)^t d_0 = -\|\nabla f(x_0)\| \leq \frac{1}{\|d\|} \nabla f(x_0)^t d \qquad \blacksquare$$

Verfahren 2.6 *(Gradientenverfahren, Methode des steilsten Abstiegs)*

Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Abbildung. Man betrachte den Algorithmus:

(S1) Man wähle ein $x_0 \in \mathbb{R}^n$.

(S2) $x_0, \ldots, x_k$ seien bestimmt. Wenn x_k stationär ist, bricht man ab. Wenn x_k nicht stationär ist, seien $d_k = -\nabla f(x_k)$ und α_k eine Lösung des MPs

$$\begin{array}{ll} \min & f(x_k + \alpha d_k) \\ \textit{bez.} & \alpha \in [0, \infty) \end{array}$$

(S3) Man setze

$$x_{k+1} = x_k + \alpha_k d_k$$

Proposition 2.7 *Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Abbildung. Wenn das Gradientenverfahren wohldefiniert ist und nicht abbricht, ist jeder Häufungspunkt der Folge ein stationärer Punkt des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in \mathbb{R}^n \end{array}$$

Beweis Ich nehme an, dass das Verfahren nicht abbricht. Nach 2.5 ist d_k für alle k eine Abstiegsrichtung, also ist $(f(x_k))$ eine streng monoton fallende Folge. Es sei (x_{k_j}) eine Teilfolge, die gegen ein x^* konvergiert. Dann konvergiert $(f(x_{k_j}))$ gegen $f(x^*)$. Da $(f(x_k))$ monoton ist, konvergiert $(f(x_k))$ gegen $f(x^*)$. Für alle k und $\alpha \geq 0$ gilt nun:

$$f(x_{k+1}) \leq f(x_k + \alpha d_k)$$

und es folgt für alle j:

$$f(x_{k_j+1}) \leq f(x_{k_j} + \alpha d_{k_j}) \; .$$

Da f stetig differenzierbar ist, konvergiert $(d_{k_j}) = (-\nabla f(x_{k_j}))$ gegen $d_0 := -\nabla f(x^*)$ und es folgt

$$f(x^*) \leq f(x^* + \alpha d_0) \qquad \text{für alle } \alpha > 0 \; .$$

Also ist d_0 keine Abstiegsrichtung und aus 2.3 folgt

$$-\|\nabla f(x^*)\|^2 = \nabla f(x^*)^t d_0 \geq 0 \; ,$$

und daher ist x^* ein stationärer Punkt. ■

Korollar 2.8 *Es sei $x_0 \in \mathbb{R}^n$ und die Menge*

$$K := \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$$

sei beschränkt. Wenn das Gradientenverfahren nicht abbricht, bleibt es in K und ist daher beschränkt. Jeder Häufungspunkt ist ein stationärer Punkt.

Beweis Ich nehme an, dass das Verfahren nicht abbricht.

Ich zeige induktiv, dass x_k für alle $k \in \mathbb{N}_0$ wohldefiniert ist und in K liegt.

Dies ist offenbar für x_0 der Fall, es gelte für $x_0, \ldots, x_k$. Dann gilt

$$f(x_0) \geq f(x_0) \geq \cdots \geq f(x_k)$$

Man setze $d_k = -\nabla f(x_k) \neq 0$. Sei

$$I := \{\alpha \geq 0 : x_k + \alpha d_k \in K\} ,$$

dann ist I nicht-leer, abgeschlossen und beschränkt, also kompakt. Man wähle ein $\alpha_k \in I$ so dass gilt

$$f(x_k + \alpha_k d_k) \leq f(x_k + \alpha d_k) \qquad \text{für alle } \alpha \in I .$$

Weiter gilt $x_k + \alpha d_k \notin K$ für alle $\alpha \notin I$ und daher für diese α:

$$f(x_k + \alpha d_k) \geq f(x_0) \geq f(x_k) \geq f(x_k + \alpha_k d_k) .$$

Es folgt

$$f(x_k + \alpha_k d_k) \leq f(x_k + \alpha d_k) \qquad \text{für alle } \alpha \geq 0 .$$

also löst α_k das MP

$$\begin{array}{ll} \min & f(x_k + \alpha d_k) \\ \text{bez.} & \alpha \in [0, \infty) \end{array}$$

und es gilt $x_{k+1} = x_k + \alpha_k d_k \in K$. ■

Bemerkung 2.9 Das Gradientenverfahren konvergiert oft in der Nähe eines stationären Punktes nicht sehr schnell, es tritt der Fall ein, dass man einen "Zick-Zack-Weg" erhält. Um "Zick-Zack-Wege zu vermeiden, kann man das Verfahren entweder modifizieren oder nur benutzen, um in die Nähe eines stationären Punktes zu kommen und dann ein anderes Verfahren zu benutzen.

In (S2) des Gradientenverfahrens muss man eine Abbildung, die auf $[0, \infty)$ definiert ist, minimieren. Das kann sehr aufwendig sein, insbesondere, wenn man bedenkt, dass eine exakte Lösung dieses Teilproblems ja in der Regel garnicht notwendig ist. Daher gibt es eine Reihe von Verfahren, die an dieser Stelle anders vorgehen. Ich werde die sogenannte **Armijo-Regel** vorstellen. Die Grundgedanken sind dabei die folgenden:

(1) Beim Gradientenverfahren ist die Wahl deswegen auf die neue Suchrichtung $d_k = -\nabla f(x_k)$ gefallen, weil sie in x_k den steilsten Abstieg hat. Nun ist diese Information lokal, so dass es zweifelhaft ist, ob ein Minimum in dieser Suchrichtung "sehr weit draußen" nützlich ist. Also ist denkbar, dass man z.B. nur in $\{x_k + \alpha d_k : 0 \leq \alpha \leq 1\}$ sucht und dann einen neuen Test macht.

(2) Mit der Wahl der Richtung verbindet man die Hoffnung auf eine gewisse Abstiegsgeschwindigkeit. Nun gilt nach 1.5:

$$f(x_k + \alpha d_k) - f(x_k) = \nabla f(x_k)^t(\alpha d_k) + R(\alpha d_k) = \alpha \nabla f(x_k)^t d_k + R(\alpha d_k)$$

Nun konvergiert $R(\alpha d_k)$ sehr schnell gegen 0, also hat man

$$f(x_k + \alpha d_k) - f(x_k) \sim \alpha \nabla f(x_k)^t d_k = -\alpha \|\nabla f(x_k)\|^2$$

Also sollte der Abstieg in dieser Größenordnung liegen. Wenn dies nicht der Fall ist, gibt man diese Suchrichtung auf. Bei der Armijo-Regel wählt man ein $\sigma \in (0,1)$ und betrachtet die Menge

$$\{\alpha \geq 0 : f(x_k + \alpha d_k) - f(x_k) \leq \sigma \alpha \nabla f(x_k)^t d_k\}$$

(3) Nun ist die Minimierungsaufgabe nach (2) ziemlich kompliziert geworden. Daher begnügt man sich damit, nur wenige Punkte zu testen: Man wählt ein $\beta \in (0,1)$ und betrachtet die Punkte

$$x_k + \beta^j d_k : j \in \mathbb{N}_0$$

Auch in dieser Menge sucht man nicht den minimalen Funktionswert, sondern den größten Punkt, der der Bedingung aus (2) genügt:

$$\alpha_k = \max\{\beta^j : f(x_k + \beta^j d_k) - f(x_k) \leq \sigma \beta^j \nabla f(x_k)^t d_k\} \ .$$

Diese Regelung hat den Vorteil, dass man nur das kleinste j finden muss, für das gilt

$$f(x_k + \beta^j d_k) \leq f(x_k) + \sigma \beta^j \nabla f(x_k)^t d_k$$

und das tut man natürlich, indem man nacheinander $j = 0, 1, \ldots$ setzt.

Verfahren 2.10 *(Modifiziertes Gradientenverfahren) Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine differenzierbare Abbildung. Man betrachte den Algorithmus:*

(S1) Man wähle ein $x_0 \in \mathbb{R}^n$ und $\sigma, \beta \in (0,1)$.

(S2) $x_0, \ldots, x_k$ seien bestimmt. Wenn x_k stationär ist, bricht man ab. Wenn x_k nicht stationär ist, seien $d_k = -\nabla f(x_k)$ und

$$\alpha_k := \max\{\beta^j : f(x_k + \beta^j d_k) \leq f(x_k) + \sigma \beta^j \nabla f(x_k)^t d_k\}$$

(S3) Man setze

$$x_{k+1} = x_k + \alpha_k d_k \ .$$

Zum Beweis der wesentlichen Eigenschaften des modifizierten Gradientenverfahrens brauche ich noch ein techisches Lemma:

Lemma 2.11 *Es seien $f : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, $x_0, d_0 \in \mathbb{R}^n$ und (x_k) bzw. (d_k) Folgen in $\mathbb{R}^n$, die gegen x_0 bzw. d_0 konvergieren. Schließlich sei (α_k) eine Nullfolge in $\mathbb{R} \setminus \{0\}$. Dann gilt*

$$\frac{f(x_k + \alpha_k d_k) - f(x_k)}{\alpha_k} \longrightarrow \nabla f(x_0)^t d_0$$

Beweis Definiert man für alle k die Abbildung $\varphi_k : \mathbb{R} \to \mathbb{R}$ durch

$$\varphi_k(\alpha) = f(x_k + \alpha d_k)$$

dann ist φ stetig differenzierbar und es gilt

$$\varphi_k'(\alpha) = \nabla f(x_k + \alpha d_k)^t d_k$$

Nach dem Mittelwertsatz gibt es ein $|\lambda_k| \leq |\alpha_k|$ so dass gilt

$$\frac{f(x_k + \alpha_k d_k) - f(x_k)}{\alpha_k} = \frac{\varphi_k(\alpha_k) - \varphi(0)}{\alpha_k} = \varphi_k'(\lambda_k) = \nabla f(x_k + \lambda_k d_k)^t d_k$$

Da f stetig diffbar ist, folgt die Behauptung. ■

Proposition 2.12 *Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Abbildung. Dann ist das modifizierte Gradientenverfahren wohldefiniert. Wenn es nicht abbricht, ist jeder Häufungspunkt der Folge ein stationärer Punkt von f.*

Beweis Es seien $x_0, \ldots, x_k$ definiert. Wenn x_k nicht stationär ist, gilt

$$\frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha} \longrightarrow_{\alpha \to 0} \nabla f(x_k)^t d_k = -\|\nabla f(x_k)\|^2 < -\sigma \|\nabla f(x_k)\|^2$$

$$= \sigma \nabla f(x_k)^t d_k$$

Also gibt es ein $\varepsilon > 0$ so dass gilt

$$\frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha} < \sigma \nabla f(x_k)^t d_k \qquad \text{für alle } |\alpha| \leq \varepsilon$$

Da (β^j) gegen 0 konvergiert, gibt es ein j so dass gilt

$$\frac{f(x_k + \beta^j d_k) - f(x_k)}{\beta^j} < \sigma \nabla f(x_k)^t d_k$$

Also ist das Verfahren wohldefiniert.

Es sei (x_k) die Folge und x^* ein Häufungspunkt sowie $(x_k)_{k \in I}$ eine Teilfolge, die gegen x^* konvergiert. Dann konvergiert $(f(x_k))_{k \in I}$ gegen $f(x^*)$. Da $(f(x_k))_{k \in \mathbb{N}}$ eine

monotone Folge ist, konvergiert dann aber $(f(x_k))_{k\in\mathbb{N}}$ gegen $f(x^*)$. Insbesondere konvergiert dann $(f(x_k)-f(x_{k+1}))$ gegen 0. Nach Konstruktion gilt für alle $k \in I$:

$$f(x_{k+1}) = f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma\alpha_k \nabla f(x_k)^t d_k$$

und daher

$$\sigma\alpha_k \|\nabla f(x_k)\|^2 \leq f(x_k) - f(x_{k+1})$$

Angenommen, es gilt $\nabla f(x^*) \neq 0$, dann konvergiert $(\alpha_k)_{k\in I}$ gegen 0. Es gelte $\alpha_k = \beta^{j_k}$, dann folgt $j_k \geq 1$ für alle $k \in I, k \geq k_0$ und daraus

$$f(x_k + \beta^{j_k-1} d_k) > f(x_k) + \sigma\beta^{j_k-1} \nabla f(x_k)^t d_k$$

also

$$\frac{f(x_k + \beta^{j_k-1} d_k) - f(x_k)}{\beta^{j_k-1}} > \sigma \nabla f(x_k)^t d_k$$

Da $(\alpha_k)_{k\in I}$ gegen 0 konvergiert, konvergiert $(j_k)_{k\in I}$ gegen ∞ und daher (β^{j_k-1}) gegen 0. Es folgt aus 2.11:

$$-\nabla f(x^*)^t \nabla f(x^*) \geq -\sigma \|\nabla f(x^*)\|^2$$

und daraus

$$(1-\sigma)\|\nabla f(x^*)\|^2 \leq 0$$

also $\nabla f(x^*) = 0$ im Widerspruch zur Annahme. ■

Die Suche nach einem stationären Punkt ist ja gerade die Suche nach einer Nullstelle der Gradientenabbildung ∇f. Nullstellen einer reellwertigen, auf einem Intervall definierten Abbildung f kann man oft sehr effektiv mit dem Newton-Verfahren finden. Dieses Verfahren hat ein mehrdimensionales Analogon, das hier zum Einsatz kommen wird, allerdings erfordert der Konvergenzbeweis einige Vorbereitungen.

Erinnerung 2.13 *Es sei $M(p,n)$ die Menge aller reellwertigen Matrizen mit p Zeilen und n Spalten. Dann ist die Abbildung $\|\cdot\| : M(p,n) \to \mathbb{R}$ definiert durch*

$$\|A\| = \max\{\|Ax\| : \|x\| \leq 1\}$$

eine Norm auf $M(p,n)$. Sie heißt auch die ***euklidische*** *oder* ***Spektralnorm****. Weiterhin gilt für alle $A \in M(p,n)$ und $B \in (n,q)$:*

(i) $\|Ax\| \leq \|A\| \, \|x\|$ *für alle* $x \in \mathbb{R}^n$

(ii) $\|AB\| \leq \|A\| \, \|B\|$

(iii) $\max\{|a_{i,j}| : i,j\} \leq \|A\| \leq n\sqrt{p}\max\{|a_{i,j}| : i,j\}$

Insbesondere 2.13(i) wird im Folgenden sehr oft, und in der Regel ohne explizite Referenz, benutzt.

Bemerkung 2.14 Die Elemente aus $M(p,n)$ kann man in natürlicher Weise als Elemente aus $\mathbb{R}^{pn}$ auffassen Nach 2.13 konvergiert eine Folge (A_k) in $M(p,n)$ bezüglich der (Matrizen-)Norm genau dann gegen eine Matrix A, wenn sie komponentenweise konvergiert. Betrachtet man also die Matrizen aus $M(p,n)$ als Elemente des $\mathbb{R}^{pn}$, so induziert die Matrixnorm gerade die konvergenten Folgen, die im $\mathbb{R}^{np}$ konvergieren.

Weiterhin sind sind die Begriffe Stetigkeit und Differenzierbarkeit von Abbildungen aus $M(p,n)$ nach $M(p',n')$ oder nach $\mathbb{R}^q$ in diesem Sinn definiert.

Proposition 2.15

(i) Die Normabbildung

$$\|\cdot\| : M(k,n) \longrightarrow \mathbb{R}$$

ist stetig.

(ii) Die Determinantenabbildung

$$\det : M(n,n) \longrightarrow \mathbb{R}$$

ist stetig.

(iii) Es sei $M(n,n)^*$ die Menge aller regulären $n \times n$-Matrizen. Dann ist die Abbildung

$$M(n,n)^* \longrightarrow M(n,n)^*$$
$$A \mapsto A^{-1}$$

stetig.

Beweis

(i) Allgemein gilt: Wenn $(E, \|\cdot\|$ ein normierter Raum ist, ist $\|\cdot\| : E \to \mathbb{R}$ stetig: Eine leichte Überlegung zeigt, dass für alle $x, y \in E$ gilt:

$$|\,\|x\| - \|y\|\,| \leq \|x - y\|$$

also folgt für alle $x, x_0 \in E$:

$$|\,\|x\| - \|x_0\|\,| \leq \|x - x_0\|$$

und daher ist $\|\cdot\|$ stetig.

(ii) Nach einem Ergebnis der Linearen Algebra gilt:

$$\det(a_{i,j}) = \sum_{\pi \in \mathcal{S}_n} sgn(\pi) a_{1,\pi(1)} \cdots a_{n,\pi(n)}$$

(iii) Für eine Matrix $A = (a_{i,j})$ sei $A_{i,j}$ die Matrix, die man erhält, wenn man in A die i.-te Zeile und j.-te Spalte streicht. Dann gilt, ebenfall nach einem Ergebnis der Linearen Algebra, für jede reguläre Matrix A:

$$A^{-1} = \frac{1}{\det(A)}((-1)^{i+j}\det(A_{j,i}))_{i,j}$$

Lemma 2.16

(i) Für alle $x, y \in \mathbb{R}^n$ gilt $|x^t y| \leq \|x\| \, \|y\|$.

(ii) Für alle $A \in M(n,n)$ und alle $x, y \in \mathbb{R}^n$ gilt

$$|x^t A y| \leq \|A\| \, \|x\| \, \|y\| \, .$$

Beweis (i) Es sei $< \cdot, \cdot >$ das euklidische Skalarprodukt auf $\mathbb{R}^n$. Dann folgt aus der Ungleichung von Cauchy-Schwarz:

$$|x^t y| = | < x, y > | \leq \|x\| \, \|y\| \, .$$

(ii) Es gilt

$$|x^t A y| = |x^t (Ay)| \leq \|x\| \, \|Ay\| \leq \|x\| \, \|A\| \, \|y\| \, . \qquad \blacksquare$$

Proposition 2.17 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}$ stetig differenzierbar und $x_0, x_2 \in \mathcal{D}$ so gewählt, dass für die Verbindungsgerade*

$$[x_0, x_2] := \{x_0 + t(x_2 - x_0) : 0 \leq t \leq 1\}$$

gilt $[x_0, x_2] \subseteq \mathcal{D}$. Dann gibt es ein $\xi \in (0,1)$ so dass gilt

$$f(x_2) - f(x_0) = Df(x_0 + \xi(x_2 - x_0))(x_2 - x_0) = \nabla f(\xi)(x_0 + \xi(x_2 - x_0))^t (x_2 - x_0)$$

Speziell gilt

$$\begin{aligned} |f(x_2) - f(x_0)| &\leq \|x_2 - x_0\| \max\{\|Df(x)\| : x \in [x_0, x_2]\} \\ &= \|x_2 - x_0\| \max\{\|\nabla f(x)\| : x \in [x_0, x_2]\} \end{aligned}$$

Beweis Man definiere $\varphi : [0,1] \to \mathbb{R}$ durch

$$\varphi(t) = f(x_0 + t(x_2 - x_0)) \, ,$$

dann ist φ stetig differenzierbar und es gilt

$$\varphi'(t) = Df(x_0 + t(x_2 - x_0))(x_2 - x_0)$$

Nach dem reellen Mittelwertsatz gibt es ein $\xi \in (0,1)$ so dass gilt

$$f(x_2) - f(x_0) = \varphi(1) - \varphi(0) = \varphi'(\xi)(1 - 0) = Df(x_0 + \xi(x_2 - x_0))(x_2 - x_0)$$

Es folgt

$$|f(x_2) - f(x_0)| = |\varphi(1) - \varphi(0)| = |\varphi'(\xi)(1-0)| = |Df(x_0 + \xi(x_2 - x_0))(x_2 - x_0)|$$
$$\leq \|Df(x_0+\xi(x_2-x_0))\| \, \|(x_2-x_0)\| \leq \sup\{\|Df(x)\| : x \in [x_0, x_2]\}\|(x_2-x_0)\|\}$$

Da Df und $\|\cdot\|$ stetige Abbildungen sind, ist die Abbildung

$$[x_0, x_2] \to \mathbb{R} \quad x \mapsto \|Df(x)\|$$

stetig, und da $[x_0, x_2]$ kompakt ist, nimmt diese Abbildung ein Maximum an. ■

Satz 2.18 *(Mittelwertsatz) Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}^p$ stetig differenzierbar und $x_0, x_2 \in \mathcal{D}$ so gewählt, dass $[x_0, x_2] \subseteq \mathcal{D}$ gilt. Dann folgt*

$$\|f(x_2) - f(x_0)\| \leq \|x_2 - x_0\| \max\{\|Df(x)\| : x \in [x_0, x_2]\}$$

Beweis Man wähle ein $a \in \mathbb{R}^p$ und definiere $g : \mathcal{D} \to \mathbb{R}$ durch

$$g(x) = f(x)^t a = \sum f_i(x) a_i$$

Dann ist g differenzierbar und es gilt

$$Dg(x) = \sum Df_i(x) a_i = Df(x)a$$

Aus 2.17 folgt dann:

$$\begin{aligned}
|(f(x_2) - f(x_0))^t a| &= |g(x_2) - g(x_0)| \\
&\leq \|x_2 - x_0\| \max\{\|Dg(x)\| : x \in [x_0, x_2]\} \\
&= \|x_2 - x_0\| \max\{\|Df(x)a\| : x \in [x_0, x_2]\} \\
&\leq \|x_2 - x_0\| \max\{\|Df(x)\| \, \|a\| : x \in [x_0, x_2]\} \\
&= \|x_2 - x_0\| \, \|a\| \max\{\|Df(x)\| : x \in [x_0, x_2]\}
\end{aligned}$$

Setzt man nun $a = f(x_0) - f(x_2)$, erhält man

$$\|f(x_2) - f(x_0)\|^2 = \|(f(x_2) - f(x_0))^t (f(x_2) - f(x_0))\|$$
$$\leq \|x_2 - x_0\| \, \|f(x_2) - f(x_0)\| \max\{\|Df(x)\| : x \in [x_0, x_2]\}$$

und daraus die Behauptung. ■

Korollar 2.19 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $x_0, x_2 \in \mathcal{D}$ und es gelte $[x_0, x_2] \subseteq \mathcal{D}$.*
(i) Es sei und $f : \mathcal{D} \to \mathbb{R}^p$ stetig differenzierbar. Dann gilt

$$\|f(x_2) - f(x_0) - Df(x_0)(x_2 - x_0)\|$$
$$\leq \|x_2 - x_0\| \max\{\|Df(x) - Df(x_0)\| : x \in [x_0, x_2]\}|$$

(ii) Es sei $f : \mathcal{D} :\longrightarrow \mathbb{R}$ zweimal stetig differenzierbar. Dann gilt

$$|f(x_2) - f(x_0) - \nabla f(x_0)^t (x_2 - x_0) - \tfrac{1}{2}(x_2 - x_0)^t Hf(x_0)(x_2 - x_0)|$$
$$\leq \tfrac{1}{2}\|x_2 - x_0\|^2 \max\{\|Hf(x) - Hf(x_0)\| : x \in [x_0, x_2]\}$$

Beweis

(i) Man definiere $g : \mathcal{D} \to \mathbb{R}^p$ durch $g(x) = f(x) - Df(x_0)x$, dann ist g stetig differenzierbar und es gilt $Dg(x) = Df(x) - Df(x_0)$. Also folgt aus dem Mittelwertsatz:

$$\begin{aligned}\|f(x_2) - f(x_0) - Df(x_0)(x_2 - x_0)\| &= \|g(x_2) - g(x_0)\| \\ &\leq \|x_2 - x_0\| \max\{\|Dg(y)\| : y \in [x_0, x_2]\} \\ &= \|x_2 - x_0\| \max\{\|Df(y) - Df(x_0)\| : y \in [x_0, x_2]\}\end{aligned}$$

(ii) Man definiere $\varphi : [0, 1] \to \mathbb{R}$ durch

$$\varphi(t) = f(x_0 + t(x_2 - x_0))$$

dann ist φ zweimal stetig differenzierbar und es gilt

$$\varphi'(t) = \nabla f(x_0 + t(x_2 - x_0))^t (x_2 - x_0)$$

und

$$\varphi''(t) = (x_2 - x_0)^t Hf(x_0 + t(x_2 - x_0))(x_2 - x_0)$$

Nach dem Satz von Taylor gibt es ein $\xi \in (0, 1)$ so dass gilt

$$\varphi(1) = \varphi(0) + \varphi'(0)(1 - 0) + \frac{1}{2}\varphi''(\xi)(1 - 0)^2$$

und daher

$$f(x_2) = f(x_0) + \nabla f(x_0)^t (x_2 - x_0) + \frac{1}{2}(x_2 - x_0)^t Hf(x_0 + \xi(x_2 - x_0))(x_2 - x_0)$$

Es folgt:

$$\begin{aligned}&|f(x_2) - f(x_0) - \nabla f(x_0)^t (x_2 - x_0) - \tfrac{1}{2}(x_2 - x_0)^t Hf(x_0)(x_2 - x_0)| \\ &\quad = \tfrac{1}{2}|(x_2 - x_0)^t (Hf(x_0 + \xi(x_2 - x_0)) - Hf(x_0))(x_2 - x_0)| \\ &\quad \leq \tfrac{1}{2}\|Hf(x_0 + \xi(x_2 - x_0)) - Hf(x_0)\| \, \|x_2 - x_0\|^2 \\ &\quad \leq \tfrac{1}{2}\|x_2 - x_0\|^2 \max\{\|Hf(x) - Hf(x_0)\| : x \in [x_0, x_2]\}\end{aligned}$$

■

2.19 ist offenbar eine Verschärfung von 1.5 und beschreibt den Fehler, den man macht, wenn man eine Funktion durch ihre lineare oder quadratische Approximation ersetzt.

Zur Motivation des Newton-Verfahrens betrachte man eine stetig differenzierbare Abbildung $f : \mathbb{R}^n \to \mathbb{R}^n$. Gesucht ist eine Nullstelle von f. Dazu sei $x_k \in \mathbb{R}^n$ eine Näherung. dann gilt für alle $x \in \mathbb{R}^n$.

$$f(x) = f(x_k) + Df(x_k)(x - x_k) + R(x)$$

Um eine neue Näherung für die Nullstelle zu bekommen, ersetzt man f durch die lineare Approximation $f(x_k) + Df(x_k)(x - x_k)$ und sucht davon eine Nullstelle. Dies ergibt als Bedingung für eine neue Näherung:

$$f(x_k) + Df(x_k)(x - x_k) = 0$$

und dies impliziert

$$x_{k+1} = x_k - Df(x_k)^{-1} f(x_k)$$

und dies ist gerade die Newton-Iteration.

Definition 2.20 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}^n$ eine stetig differenzierbare Abbildung und $x_0 \in \mathcal{D}$. Wenn die induktiv durch*

$$x_{k+1} = x_k - Df(x_k)^{-1} f(x_k)$$

*definierte Folge wohldefiniert ist, heißt sie **Newton-Folge** oder **Newton-Iteration** für f mit dem Startwert x_0.*

Der angekündigte Konvergenzsatz für das (n-dimensionale) Newton-Verfahren lautet nun:

Satz 2.21 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}^n$ eine stetig differenzierbare Abbildung und $x^* \in \mathcal{D}$ eine Nullstelle von f, für die $Df(x^*)$ regulär ist. Es gebe ein $r_0 > 0$ so dass $B(x^*, r_0] \subseteq \mathcal{D}$ gilt und dass Df in $B(x^*, r_0]$ einer Lipschitzbedingung genügt. Dann gibt es ein $r > 0$, so dass die Newton-Folge mit dem Startwert x_0 für alle $x_0 \in B(x^*, r]$ wohldefiniert ist und gegen x^* konvergiert. In der Tat konvergiert die Folge quadratisch, d.h. es gibt ein $c > 0$ so dass gilt:*

$$\|x_{k+1} - x^*\| \leq c\|x_k - x^*\|^2 \qquad \text{für alle } k \in \mathbb{N}$$

Beweis

(a) Da Df in $B(x^*, r_0]$ einer Lipschitzbedingung genügt, gibt es ein $L > 0$ so dass gilt:

$$\|Df(x) - Df(y)\| \leq L\|x - y\| \qquad \text{für alle } x, y \in B(x^*, r_0] \,.$$

b) Da die Abbildung

$$x \mapsto \det(Df(x))$$

stetig ist, gibt es ein $s > 0$ so dass gilt

$$\det(Df(x)) \neq 0 \qquad \text{für alle } x \in B(x^*, s] \,,$$

also ist $Df(x)$ für alle $x \in B(x^*, s]$ regulär. OBdA gelte $s \leq r_0$.

c) Da die Abbildung von $B(x^*, s]$ nach $\mathbb{R}$

$$x \to \|Df(x)^{-1}\|$$

stetig ist, ist sie beschränkt, also gibt es ein $M > 0$ so dass gilt:

$$\|Df(x)^{-1}\| \leq M \qquad \text{für alle } x \in B(x^*, s] \ .$$

Es gilt für alle k mit $x_k \in B(x^*, s]$:

$$x_{k+1} = x_k - Df(x_k)^{-1} f(x_k) \ ,$$

es folgt:

$$\begin{aligned} x_{k+1} - x^* &= x_k - x^* - Df(x_k)^{-1} f(x_k) \\ &= Df(x_k)^{-1}(-f(x_k) - Df(x_k)(x^* - x_k)) \\ &= Df(x_k)^{-1}(f(x^*) - f(x_k) - Df(x_k)(x^* - x_k)) \end{aligned}$$

und daraus:

$$\begin{aligned} \|x_{k+1} - x^*\| &= \|Df(x_k)^{-1}(f(x^*) - f(x_k) - Df(x_k)(x^* - x_k))\| \\ &\leq \|Df(x_k)^{-1}\| \, \|f(x^*) - f(x_k) - Df(x_k)(x^* - x_k)\| \\ &\leq M\|x^* - x_k\| \max_{y \in [x^*, x_k]} \|Df(y) - Df(x_k)\| \\ &\leq M\|x^* - x_k\| \max_{y \in [x^*, x_k]} L\|y - x_k\| \\ &\leq LM\|x^* - x_k\|^2 \end{aligned}$$

Man wähle nun $r \leq s$ so, dass gilt $LMr \leq 1/2$, dann folgt für alle k mit $x_k \in B(x^*, r]$:

$$\|x_{k+1} - x^*\| \leq \frac{1}{2}\|x_k - x^*\| \leq r \ ,$$

also folgt mit vollständiger Induktion, dass $x_k \in B(x^*, r]$ für alle k gilt, wenn $x_0 \in B(x^*, r]$ gilt. Weiterhin konvergiert die Folge gegen x^* und es gilt

$$\|x_{k+1} - x^*\| \leq LM\|x_k - x^*\|^2 \ .$$

■

Lemma 2.22 *Es seien $f : \mathcal{D} \to \mathbb{R}^p$ stetig differenzierbar, $x_0 \in \mathcal{D}$ und $r > 0$ so gewählt, dass $B(x_0, r] \subseteq \mathcal{D}$ gilt. Dann genügt f in $B(x_0, r]$ einer Lipschitzbedingung.*

Beweis Da Df und $\|\cdot\|$ stetig sind, ist die Abbildung $\|\Delta \cdot\|$ stetig und nimmt daher auf der kompakten Menge $B(x_0, r]$ ihrm Maximum an. Also gilt mit $L = \max\{\|Df(z)\| : z \in B\}$: nach dem Mittelwertsatz (2.17) für alle $x, y \in B := B(x_0, r]$:

$$\|f(x) - f(y)\| \leq \|x - y\| \max\{\|Df(z)\| : z \in B\} = L\|x - y\|$$

Korollar 2.23 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}^n$ zweimal stetig differenzierbar und $x^* \in \mathcal{D}$ eine Nullstelle von f so das $Df(x^*)$ regulär ist. Dann gibt es ein $r > 0$ so dass das Newton-Verfahren für jeden Startwert $x_0 \in B(x^*, r]$ quadratisch gegen x^* konvergiert.*

Beweis Da f zweimal stetig differenzierbar ist, $Df : \mathcal{D} \to \mathbb{R}^n$ stetig differenzierbar. Die Behauptung folgt dann aus 2.21 und 2.22. ■

Bemerkung 2.24 Um beim Newton-Verfahren x_{k+1} aus x_k zu bestimmen, muss man $Df(x_k)$ nicht invertieren, denn $d_k = x_{k+1} - x_k$ ist die Lösung des linearen Gleichungssystems

$$Df(x_k)X = -f(x_k) \ .$$

Dennoch ist die Lösung dieses Gleichungssystems der aufwendigste Schritt des Newton-Verfahrens und es gibt eine Reihe von Varianten, die diesen Schritt vereinfachen, dann aber nicht mehr so schnell konvergieren.

Satz 2.25 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $f : \mathcal{D} \to \mathbb{R}$ eine dreimal stetig differenzierbare Abbildung. Weiterhin seien $x^* \in \mathcal{D}$ so gewählt, dass $\nabla f(x^*) = 0$ gilt und dass $Hf(x^*)$ regulär ist. Dann gibt es ein $r > 0$ so dass die Iteration*

$$x_{k+1} = x_k - Hf(x_k)^{-1} \nabla f(x_k)$$

für alle $x_0 \in B(x^*, r]$ quadratisch gegen x^* konvergiert.

Beweis Man wende 2.23 auf ∇f an. ■

Das Newton-Verfahren konvergiert nur lokal und dann auch nur gegen einen stationären Punkt. Nun ist aber auch ein Punkt, der das Maximumproblem löst, ein stationärer Punkt. Man sucht nun Verfahren, die diese beiden Nachteile vermeiden, aber die schnelle Konvergenz des Newton-Verfahrens in einer Umgebung einer Lösung ausnutzen. Im folgenden wird so ein Verfahren beschrieben und studiert.

Verfahren 2.26 *(Globalisiertes Newton-Verfahren) Es sei* $f : \mathbb{R}^n \to \mathbb{R}$ *eine zweimal stetig differenzierbare Abbildung. Man wähle* $\rho > 0,\ p > 2,\ \beta \in (0,1)$ *sowie* $\sigma \in (0, 1/2)$ *und betrachte den folgenden Algorithmus:*

(S1) Man wähle $x_0 \in \mathbb{R}^n$ *beliebig.*

(S2) $x_0, \ldots, x_k$ *seien konstruiert. Wenn* x_k *stationär ist, bricht das Verfahren ab. Andernfalls wähle man ein* $d_k \in \mathbb{R}^n$ *so dass gilt*

$$Hf(x_k)d_k = -\nabla f(x_k)$$

Falls dies nicht möglich ist oder wenn nicht gilt

$$\nabla f(x_k)^t d_k \leq -\rho \|d_k\|^p$$

setze man $d_k = -\nabla f(x_k)$

(S3) Man setze

$$\alpha_k := \max\{\beta^j : f(x_k + \beta^j d_k) \leq f(x_k) + \sigma\beta^j \nabla f(x_k)^t d_k\}$$

(S4) Man setze

$$x_{k+1} = x_k + \alpha_k d_k$$

Wenn also das Newton-Verfahren nicht durchführbar ist oder keine befriedigende Abstiegseigenschaft hat, wechselt man auf einen Schritt des modifizierten Gradientenverfahrens. In der Tat garantiert der Algorithmus, dass d_k in jedem Schritt eine Abstiegsrichtung ist.

Bemerkung 2.27

(i) Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}$ zweimal stetig differerenzierbar und $x^* \in \mathcal{D}$ eine Nullstelle von f so dass $Df(x^*)$ regulär ist. Dann gibt es ein $r > 0$ so dass für alle $x \in B(x^*, r]$ die Newton-Iteration x' definiert ist und dass gilt $\|x^* - x'\| \leq \|x^* - x\|$.

(ii) Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Abbildung. Wenn das modifizierte Gradientenverfahren nicht abbricht, ist jeder Häufungspunkt x^* der Folge nach 2.12 ein stationärer Punkt von f. Eine Analyse des Beweises zeigt, dass x^* auch dann ein stationärer Punkt ist, wenn man d_k und α_k für alle $k \notin I$ nur so wählt, dass $f(x_k + \alpha_k d_k) \leq f(x_k)$ gilt.

Proposition 2.28 *Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine zweimal stetig differenzierbare Abbildung. Dann ist das globalisierte Newton-Verfahren wohldefiniert. Wenn es nicht abbricht, ist jeder Häufungspunkt ein stationärer Punkt von f.*

Beweis Ich zeige zunächst, dass das Verfahren wohldefiniert ist. Dazu ist zu zeigen, dass (S3) in jedem Fall durchführbar ist. Also seien $x_0, \ldots, x_k$ definiert und $\nabla f(x_k) \neq 0$. Falls x_{k+1} wie im modifizierten Gradientenverfahren definiert ist, ist ist (S3) offenbar durchführbar. Andernfalls gilt $d_k \neq 0$ und daher

$$\nabla f(x_k)^t d_k < -\rho \|d_k\|^p < 0$$

Also ist d_k eine Abstiegsrichtung. Es folgt:

$$\frac{f(x_k + td_k) - f(x_k)}{t} \longrightarrow_{t \to 0} \nabla f(x_k)^t d_k < 0$$

Dies impliziert $\nabla f(x_k)^t d_k < \sigma \nabla f(x_k)^t d_k$, also gibt es ein $\varepsilon > 0$ so dass gilt

$$\frac{f(x_k + td_k) - f(x_k)}{t} < \sigma \nabla f(x_k)^t d_k \qquad \text{für alle } 0 < t \leq \varepsilon$$

Sei j so gewählt, dass gilt $\beta^j \leq \varepsilon$, dann gilt offenbar

$$f(x_k + \beta^j d_k) < f(x_k) + \sigma \beta^j \nabla f(x_k)^t d_k$$

Ich nehme an, dass das Verfahren nicht abbricht. Es seien x^* ein Häufungspunkt des Verfahrens und $(x_k)_{k \in I}$ eine Folge, die gegen x^* konvergiert. Wenn x_{k+1} unendlich oft durch einen Gradientenschritt erzeugt wird, folgt die Behauptung aus 2.27. Also kann man oBdA annehmen, dass x_{k+1} für alle $k \in I$ durch einen "Newton-Schritt" erzeugt wird.

Annahme $\nabla f(x^*) \neq 0$

Ich zeige als Erstes:

Behauptung Es gibt $c_1, c_2 > 0$ so, dass gilt

(1) $$c_1 \leq \|d_k\| \leq c_2 \qquad \text{für alle } k \in I$$

Beweis Für alle $k \in I$ gilt

$$\|\nabla f(x_k)\| = \|Hf(x_k) d_k\| \leq \|Hf(x_k)\| \, \|d_k\|$$

Angenommen, es gibt eine Teilfolge $(d_k)_{k \in J}$ von $(d_k)_{k \in I}$, die gegen 0 konvergiert. Da $(Hf(x_k))_{k \in I}$ gegen $Hf(x^*)$ konvergiert, konvergiert dann $(\nabla f(x_k))_{k \in J}$ gegen 0, im Widerspruch zu $\nabla f(x^*) \neq 0$. Also gibt es ein k_0 und ein $c > 0$ so dass gilt $c \leq \|d_k\|$ für alle $k \in I$, $k \geq k_0$ und man kann

$$c_1 = \min(\{\|d_k\| : k \in I,\, k < k_0\} \cup \{c\})$$

setzen. Andererseits folgt aus $\nabla f(x_k)^t d_k \leq -\rho \|d_k\|^p$ für alle $d \in I$, dass gilt

$$\rho \|d_k\|^p \leq |\nabla f(x_k)^t d_k| \leq \|\nabla f(x_k)\| \, \|d_k\|$$

und daraus

$$\rho \|d_k\|^{p-1} \leq \|\nabla f(x_k)\|$$

Da $(\nabla f(x_k))_{k \in I}$ gegen $\nabla f(x^*)$ konvergiert, ist die Folge beschränkt Wegen $p-1 > 1$ ist auch die Folge $(\|d_k\|)_{k \in I}$ beschränkt, und es gibt ein c_2 so dass gilt

$$\|d_k\| \leq c_2 \qquad \text{für alle } k \in I \; .$$

Da $(x_k)_{k \in I}$ gegen x^* konvergiert und $(f(x_k))$ monoton fällt, konvergiert $(f(x_k))_{k \in \mathbb{N}}$ gegen $f(x^*)$. Aus

$$f(x_{k+1}) - f(x_k) \leq \sigma \alpha_k \nabla f(x_k)^t d_k \qquad \text{für alle } k \in I$$

folgt :

$$(\alpha_k \nabla f(x_k)^t d_k)_{k \in I} \longrightarrow 0 \tag{2}$$

Ich zeige als Nächstes, dass gilt:

Behauptung Es gibt ein $\varepsilon > 0$ so dass gilt $|\alpha_k| \geq \varepsilon$ für alle $k \in I$.

Beweis Angenommen, es gibt eine Teilfolge $(\alpha_k)_{k \in J}$ von $(\alpha_k)_{k \in I}$, die gegen 0 konvergiert. Es gelte $\alpha_k = \beta^{j_k}$, dann konvergiert j_k gegen ∞ und es folgt für alle $k \in J$ nach Wahl von α_k:

$$f(x_k + \beta^{j_k - 1} d_k) \geq f(x_k) + \sigma \beta^{j_k - 1} \nabla f(x_k)^t d_k$$

Man erhält:

$$\frac{f(x_k + \beta^{j_k - 1} d_k) - f(x_k)}{\beta^{j_k - 1}} \geq \sigma \nabla f(x_k)^t d_k$$

Da (d_k) nach (1) beschränkt ist, kann man oBdA annehmen, dass (d_k) gegen ein d^* konvergiert. Man erhält:

$$\nabla f(x^*)^t d^* \geq \sigma \nabla f(x^*)^t d^*$$

und daraus $(1 - \sigma) \nabla f(x^*)^t d^* \geq 0$, also $\nabla f(x^*)^t d^* \geq 0$. Andererseits folgt aus

$$\nabla f(x_k)^t d_k \leq -\rho \|d_k\|^p \qquad \text{für alle } k \in I \; ,$$

dass $\nabla f(x^*)^t d^* \leq -\rho \|d^*\|^p < -\rho c_1^p < 0$ gilt. W.!

Also gilt $|\alpha_k| \geq \varepsilon$ für alle $k \in I$ und aus (2) folgt, dass $(\nabla f(x_k)^t d_k)_{k \in I}$ gegen 0 konvergiert. Dies impliziert $\nabla f(x^*)^t d^* = 0$, erneut im Widerspruch zu $\nabla f(x^*)^t d^* < 0$. Also ist die Annahme $\nabla f(x^*) \neq 0$ falsch und es folgt die Behauptung des Satzes. ■

Im Folgenden will ich zeigen: Wenn x^* ein Häufungspunkt des globalisierten Newton-Verfahrens ist und $Hf(x^*)$ positiv definit ist, geht das Verfahren in das Newton-Verfahren über und konvergiert quadratisch gegen x^*. Dazu reicht es zu zeigen, dass d_k in (S2) nach der ersten Bedingung gewählt wird, wenn $k \geq k_0$ gilt und dass $\alpha_k = 1$ für diese k gilt.

Proposition 2.29 *Es seien $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar, $x^* \in \mathbb{R}^n$ und $Hf(x^*)$ sei positiv definit. Dann gibt es ein $r > 0$ und ein $c > 0$ so dass gilt*

$$h^t Hf(x)h \geq c\|h\|^2 \qquad \text{für alle } x \in B(x^*, r] \text{ und alle } h \in \mathbb{R}^n$$

Beweis Angenommen, die Behauptung ist falsch, dann gibt es zu jedem $k \in \mathbb{N}$ ein $x_k \in B(x^*, 1/k)$ und ein $h_k \in \mathbb{R}^n$ so dass gilt

$$h_k^t Hf(x_k)h_k < \frac{1}{k}\|h_k\|^2$$

Indem man $\frac{1}{\|h_k\|}h_k$ betrachtet, kann man oBdA annehmen, dass $\|h_k\| = 1$ für alle $k \in \mathbb{N}$ gilt. Dann gibt es eine Teilfolge (h_{k_j}), die gegen ein $h_0 \in \mathbb{R}^n$ konvergiert. Offenbar gilt $\|h_0\| = 1$. Aus

$$h_k^t Hf(x_k)h_k < 1/k\|h_k\|^2 \qquad \text{für alle } k \in \mathbb{N}$$

folgt $h_0^t Hf(x^*)h_0 = 0$ im Widerspruch zur positiven Definitheit von $Hf(x^*)$. ■

Lemma 2.30 *Es seien $f : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar, $x^* \in \mathbb{R}^n$ ein stationärer Punkt von f und $Hf(x^*)$ sei positiv definit. Weiterhin seien $\rho > 0$, $p > 2$ und $0 < \sigma < 1/2$. Dann gibt es ein $r > 0$ so dass $Hf(x)$ für alle $x \in B(x^*, r]$ regulär ist und dass für $d_x = -Hf(x)^{-1}\nabla f(x)$ gilt*

$$\nabla f(x)^t d_x \leq -\rho\|d_x\|^p$$

sowie

$$f(x + d_x) \leq f(x) + \sigma\nabla f(x)^t d_x$$

Beweis Nach 2.29 gibt es ein $r_1 > 0$ und ein $c > 0$ so dass gilt

$$d^t Hf(x)d \geq c\|d\|^2 \qquad \text{für alle } x \in B(x^*, r_1] \text{ und alle } d \in \mathbb{R}^n$$

Also ist insbesondere $Hf(x)$ für alle $x \in B(x^*, r_1]$ positiv definit und daher regulär. Nun gilt für diese x:

$$\nabla f(x)^t d_x = -d_x^t Hf(x)d_x \leq -c\|d_x\|^2 = -\rho\|d_x\|^p \frac{c}{\|d_x\|^{p-2}\rho}$$

Weiterhin gilt

$$d_x = Hf(x)^{-1}\nabla f(x) \longrightarrow_{x \to x^*} Hf(x^*)^{-1}\nabla f(x^*) = 0$$

Wegen $p > 2$ folgt daraus

$$\|d_x\|^{p-2} \longrightarrow_{x \to x^*} 0$$

und daher gibt es ein $r_2 \leq r_1$ so dass gilt

$$\|d_x\|^{p-2} \leq c/\rho \qquad \text{für alle } x \in B(x^*, r_2]$$

Es folgt dann für diese x:

$$\nabla f(x)^t d_x \leq -\rho \|d_x\|^p \frac{c}{\|d_x\|^{p-2}\rho} \leq -\rho \|d_x\|^p$$

Weiterhin gilt

$$\begin{aligned} & f(x+d_x) - f(x) - \sigma \nabla f(x)^t d_x \\ &= f(x+d_x) - f(x) - \nabla f(x)^t d_x + (1-\sigma)\nabla f(x)^t d_x \\ &= f(x+d_x) - f(x) - \nabla f(x)^t d_x - (1-\sigma) d_x^t Hf(x) d_x \\ &= f(x+d_x) - f(x) - \nabla f(x)^t d_x - \tfrac{1}{2} d_x^t Hf(x) d_x \\ &\quad -(1/2-\sigma) d_x^t Hf(x) d_x \end{aligned}$$

Nach 2.29 gibt es ein $r_3 \leq r_2$ und ein $c > 0$ so dass gilt

$$d^t Hf(x) d \geq c\|d\|^2 \qquad \text{für alle } x \in B(x^*, r_3] \text{ und alle } d$$

Andererseits gilt nach 2.19 für alle x und d:

$$|f(x+d) - f(x) - \nabla f(x)^t - \frac{1}{2} d^t Hf(x) d| \leq \|d\|^2 \max\{\|Hf(x) - Hf(y)\| : y \in [x, x+d]\}$$

Also gibt es ein $r \leq r_3$ so dass für alle $x \in B(x^*, r]$ gilt:

$$|f(x+d_x) - f(x) - \nabla f(x)^t d_x - \frac{1}{2} d_x^t Hf(x) d_x| \leq \|d_x\|^2 (\frac{1}{2} - \sigma) c \leq (\frac{1}{2} - \sigma) d_x^t Hf(x) d_x$$

und es folgt die Behauptung. ■

Satz 2.31 *Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine zweimal stetig differenzierbare Abbildung und (x_k) eine Folge, die nach dem Algorithmus des globalisierten Newton-Verfahrens erzeugt worden ist. Es sei x^* ein Häufungspunkt der Folge und $Hf(x^*)$ sei positiv definit. Dann besitzt f in x^* ein (striktes) lokales Minimum, die Folge (x_k) konvergiert gegen x^* und es gibt ein j_0 so dass die Iteration für alle $j \geq j_0$ die Newton-Iteration ist.*

Beweis Nach 2.28 ist x^* ein stationärer Punkt. Da $Hf(x^*)$ positiv definit ist, besitzt f in x^* ein (isoliertes) lokales Minimum. Nach 2.30 gibt es ein $r > 0$ so dass x_{k+1} für alle $x_k \in B(x^*, r]$ nach dem Newton-Verfahren bestimmt wird. Da $(x_k)_{k \in I}$ gegen x^* konvergiert, gibt es ein $k_0 \in I$ so dass $x_{k_0} \in B(x^*, r]$ liegt. Nach 2.27 kann man oBdA annehmen, dass das Newton-Verfahren für alle $k \geq k_0$ in $B(x^*, r]$ liegt und daher wird x_{k+1} für alle $k \geq k_0$ nach dem Newton-Verfahren bestimmt. ■

Kapitel 3

Konvexe Mengen

Bei konvexen Minimierungsproblemen sind der zulässige Bereich und die Zielfunktion konvex. In diesem Fall gibt es eine Reihe wesentlicher Vereinfachungen: Jede lokale Lösung so eines MPs ist eine Lösung (4.6). Wenn die Zielfunktion differenzierbar ist, sind die stationären Punkte (in geeigneter Weise definiert) genau die Lösungen des MPs (4.12). Jeder Kuhn-Tucker-Punkt ist eine Lösung (6.3), und die Umkehrung gilt unter der leicht zu verifizierenden Slater-Bedingung (6.7). Man beachte auch das letzte Zitat am Ende des Manuskripts. Allerdings sind einige dieser Resultate sehr tief und brauchen eine aufwendige Vorbereitung, so zum Beispiel den Trennungssatz, der in (3.25) bewiesen wird.

Bemerkung 3.1 *Es seien E ein reeller Vektorraum und $a, b \in E$. Dann gilt:*

$$\{\alpha a + (1-\alpha)b : 0 \leq \alpha \leq 1\} = \{b + \alpha(a-b) : 0 \leq \alpha \leq 1\}$$

$$= \{\beta b + (1-\beta)a : 0 \leq \beta \leq 1\} = \{a + \beta(b-a) : 0 \leq \beta \leq 1\}$$

Diese Menge heißt die Verbindungsstrecke zwischen a und b und wird mit $[a, b]$ bezeichnet. Offenbar gilt $[a, b] = [b, a]$. (Beachten Sie, dass dies natürlich für $n = 1$ nicht gilt. Aber das kommt so selten vor, dass ich diese Inkonsistenz in Kauf nehme.)

Definition 3.2 *Es sei E ein reller Vektorraum. Eine Menge $A \subseteq E$ heißt **konvex**, wenn gilt:*

$$\alpha a + (1-\alpha)b \in A \qquad \text{für alle } a, b \in A \text{ und alle } \alpha \in \mathbb{R},\ 0 \leq \alpha \leq 1$$

wenn also die Verbindungsstrecke zweier Punkte aus A wieder in A liegt.

Beispiele 3.3

(i) Es sei E ein normierter Raum. Dann sind für alle $x_0 \in E$ und alle $r > 0$ die Mengen

$$B(x_0, r) := \{x \in E : \|x - x_0\| < r\}$$

und

$$B(x_0, r] := \{x \in E : \|x - x_0\| \leq r\}$$

konvexe Mengen.

(ii) Es seien E ein reeller Vektorraum und $\varphi : E \to \mathbb{R}$ eine lineare Abbildung sowie $b \in \mathbb{R}$. Dann sind die Mengen

$$\{x \in E : \varphi(x) \geq b\} \text{ und } \{x \in E : \varphi(x) = b\}$$

konvex. Speziell sind für alle $A \in M(p, n)$ und $b \in \mathbb{R}^p$ die Mengen

$$\{x \in \mathbb{R}^n : Ax \geq b\} \text{ und } \{x \in \mathbb{R}^n : Ax = b\}$$

konvex.

(iii) Eine Menge $A \subseteq \mathbb{R}$ ist genau dann konvex, wenn sie ein Intervall ist.

Beweis Alle Beweise sind Routine, ich zeige daher nur einen Teil von (i):

Es seien $a, b \in B(x_0, r)$ und $0 \leq \alpha \leq 1$. Dann gilt:

$$\begin{aligned} \|\alpha a + (1-\alpha)b - x_0\| &= \|\alpha(a - x_0) + (1-\alpha)(b - x_0)\| \\ &\leq \|\alpha(a - x_0)\| + \|(1-\alpha)(b - x_0)\| \\ &= \alpha\|a - x_0\| + (1-\alpha)\|b - x_0\| \\ &< (\alpha + (1-\alpha))r = r \end{aligned}$$

Also ist $B(x_0, r)$ konvex. ■

Proposition 3.4 *Es seien E ein reeller Vektorraum und $A, B \subseteq E$ konvexe Mengen und $\alpha \in \mathbb{R}$. Dann gelten:*

(i) $A + B := \{a + b : a \in A\ ,\ b \in B\}$ ist eine konvexe Menge.

(ii) $\alpha A := \{\alpha a : a \in A\}$ ist eine konvexe Menge.

(iii) Es sei $(A_i)_{i \in I}$ eine Familie konvexer Mengen. Dann ist $\bigcap_{i \in I} A_i$ eine konvexe Menge.

Wenn E ein normierter Raum ist, gilt weiterhin:

(iv) $\overline{A}$ ist eine konvexe Menge.

Beweis (i) - (iii) sind Routine.

(iv): Es seien $a, b \in \overline{A}$ und $0 \leq \alpha \leq 1$. Dann gibt es Folgen (a_i) und (b_i) in A, die gegen a bzw. b konvergieren. Es folgt $\alpha a_i + (1-\alpha)b_i \in A$ für alle i und da $(\alpha a_i + (1-\alpha)b_i)$ gegen $\alpha a + (1-\alpha)b$ konvergiert, folgt $\alpha a + (1-\alpha)b \in \overline{A}$. ■

Proposition 3.5 *Es sei E ein reller Vektorraum. Eine Menge $A \subseteq E$ ist genau dann konvex, wenn für alle $a_1, \ldots, a_k \in A$ und alle $\alpha_1, \ldots, \alpha_k \geq 0$ mit $\sum_{i=1}^k \alpha_i = 1$ gilt $\sum_{i=1}^k \alpha_i a_i \in A$.*

Beweis Offenbar ist A konvex, wenn alle Elemente dieser Form zu A gehören (Im Fall $k = 2$ gilt $\alpha_2 = 1 - \alpha_1$.) Also sei A konvex, dann zeige ich durch Induktion nach k, dass alle Elemente dieser Form zu A gehören. Der Fall $k = 1$ ist klar, die Behauptung gelte für k. Es sei

$$a = \sum_{i=1}^{k+1} \alpha_i a_i \ ,$$

dann kann man oBdA annehmen, dass $\alpha_i > 0$ für alle i gilt. Man setze

$$\alpha = \sum_{i=1}^{k} \alpha_i > 0$$

dann gilt $\alpha_{k+1} = 1 - \alpha$ und aus der Induktionsannahme folgt:

$$c := \frac{1}{\alpha} \sum_{i=1}^{k} \alpha_i a_i = \sum_{i=1}^{k} \frac{\alpha_i}{\alpha} a_i \in A \ .$$

Die Konvexität von A impliziert dann:

$$a = \alpha c + (1 - \alpha) a_{k+1} \in A \ .$$

■

Definition 3.6 *Es seien E reeller Vektorraum Raum und $A \subseteq E$. Dann heißt*

$$co(A) = \bigcap \{C \subseteq \mathbb{R}^n : A \subseteq C \ , \ C \text{ konvex}\}$$

*die **konvexe Hülle** von A.*

Nach 3.4 ist $co(A)$ konvex.

Proposition 3.7 *Es seien E ein reeller Vektorraum und $A \subseteq E$. Dann gilt:*

$$co(A) = \{\sum_{i=1}^{k} \alpha_i a_i : k \in \mathbb{N},\ a_i \in A,\ \alpha_i \geq 0,\ \sum_{i=1}^{k} \alpha_i = 1\} \ .$$

Beweis Es sei B die Menge auf der rechten Seite. Man sieht leicht, dass B konvex ist. Wenn nun $C \supseteq A$ eine konvexe Menge ist, gilt $\sum_{i=1}^{k} \alpha_i a_i \in C$ für alle $k \in \mathbb{N},\ a_i \in C,\ \alpha_i \geq 0,\ \sum_{i=1}^{k} \alpha_i = 1$ nach 3.5 und daher erst recht für alle $a_i \in A$. Also folgt $B \subseteq C$ für jede konvexe Menge $C \supseteq A$ und daraus die Behauptung.

Bemerkung 3.8 *Elemente der Form $\sum_{i=1}^{k} \alpha_i a_i$ für die gilt $\alpha_1, \ldots, \alpha_k \geq 0$ und $\sum_{i=1}^{k} \alpha_i = 1$ nennt man **Konvexkombination** der Elemente $a_1, \ldots, a_k$. Nach 3.7 besteht die konvexe Hülle einer Menge A also aus allen Konvexkombinationen von Elementen aus A.*

Beispiele 3.9

(i) Natürlich gilt $co(A) = A$ genau dann, wenn A konvex ist.

(ii) Es seien E ein reelleer Vektorraum und $a, b \in E$. Dann gilt

$$co(\{a, b\}) = \{\alpha a + (1 - \alpha)b : 0 \leq \alpha \leq 1\} = [a, b] = [b, a]$$

d.h. $co(\{a, b\})$ ist die Strecke zwischen b und a.

(iii) Es seien $r, s > 0$ und $A = \{(0, 0), (r, 0), (0, s)\} \subseteq \mathbb{R}^2$. Dann gilt

$$co(A) = \{(u, v)^t \in \mathbb{R}^2 : u, v \geq 0\ ,\ v \leq u - \frac{s}{r}u\}$$

Also ist die konvexe Hülle von A das Dreieck (mit Inhalt), mit den Ecken $(0, 0), (r, 0)$ und $(0, s)$.

Lemma 3.10 *Es seien $a_1, \ldots, a_k$ linear abhängige Vektoren in $\mathbb{R}^n$ und $\alpha_1, \ldots, \alpha_k \geq 0$. Dann gibt es $\beta_1, \ldots, \beta_k \geq 0$ so dass gilt $\beta_j = 0$ für ein j und*

$$\sum_{i=1}^{k} \alpha_i a_i = \sum_{i=1}^{k} \beta_i a_i$$

Beweis Sei $x = \sum_{i=1}^{k} \alpha_i a_i$. Es gibt $(\gamma_1, \ldots, \gamma_k) \neq (0, \ldots, 0)$ so dass gilt $\sum_{i=1}^{k} \gamma_i a_i = 0$. OBdA gebe es ein r mit $\gamma_r > 0$. Es gilt für alle $\delta \geq 0$:

$$x = \sum (\alpha_i - \delta\gamma_i)a_i$$

Setzt man $\beta_i = \alpha_i - \delta\gamma_i$ für alle i, dann gilt $\beta_i \geq 0$ für alle i genau dann, wenn gilt

$$\delta\gamma_i \leq \alpha_i \qquad \text{für alle } \gamma_i > 0\ ,$$

also

$$\delta \leq \frac{\alpha_i}{\gamma_i} \qquad \text{für alle } \gamma_i > 0\ .$$

Sei

$$\delta = \min\{\frac{\alpha_i}{\gamma_i} : \gamma_i > 0\}\ ,$$

dann genügen $\beta_1, \ldots, \beta_k$ der Bedingung des Lemmas. ■

Der Satz von Caratheodory hat zwei schöne Anwendungen, die ich zeigen will, obwohl sie in dieser Vorlesung nicht gebraucht werden:

Satz 3.11 *(Caratheodory) Es seien $A \subseteq \mathbb{R}^n$ und $x \in co(A)$. Dann kann man x als Konvexkombination von höchstens $n + 1$ Elementen aus A darstellen.*

Beweis Es sei $x \in co(A)$, dann gibt es $a_1, \ldots, a_k \in A$ und $\alpha_1, \ldots, \alpha_k \geq 0$ mit $\sum_{i=1}^{k} \alpha_i = 1$ so dass gilt

$$x = \sum_{i=1}^{k} \alpha_i a_i$$

Man wähle ein minimales k mit dieser Eigenschaft. Angenommen, $k \geq n+2$. Dann sind die Vektoren

$$\begin{pmatrix} a_1 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} a_k \\ 1 \end{pmatrix}$$

linear abhängig und es gilt:

$$\begin{pmatrix} x \\ 1 \end{pmatrix} = \sum_{i=1}^{k} \alpha_i \begin{pmatrix} a_i \\ 1 \end{pmatrix} .$$

Nach 3.10 gibt es $\beta_1, \ldots, \beta_k \geq$ mit $\beta_j = 0$ für ein j so dass gilt

$$\begin{pmatrix} x \\ 1 \end{pmatrix} = \sum_{i=1}^{k} \beta_i \begin{pmatrix} a_i \\ 1 \end{pmatrix} .$$

Es folgt $x = \sum_{i \neq j} \beta_i a_i$ und $\sum_{i \neq j} \beta_i = 1$. Also ist k nicht minimal. Widerspruch! ■

Korollar 3.12 *Es sei $K \subseteq \mathbb{R}^n$ kompakt. Dann ist $co(K)$ kompakt.*

Beweis Es sei

$$L = \{\alpha \in \mathbb{R}^{n+1} : 0 \leq \alpha \leq 1 \, , \sum_{i=1}^{n+1} a_i = 1\} \times K^{n+1} \, ,$$

dann ist L kompakt. Man definiere die stetige Abbildung $\varphi : L \to \mathbb{R}^n$ durch $\varphi(\alpha_1, \ldots, \alpha_{n+1}, a_1, \ldots, a_{n+1}) = \sum_{i=1}^{n+1} \alpha_i a_i$. Dann gilt $\varphi(L) = co(K)$ nach dem Satz von Caratheodory und als stetiges Bild einer kompakten Menge ist $co(K)$ kompakt. ■

Proposition 3.13 *Es sei $a_1, \ldots, a_k \in \mathbb{R}^n$. Dann ist*

$$A = \{\sum_{i=1}^{k} \alpha_i a_i : \alpha_i \geq 0\}$$

eine konvexe, abgeschlossene Menge.

Beweis Offenbar ist A konvex. Ich zeige die Abgeschlossenheit von A zunächst für den Fall, dass $a_1, \ldots, a_k$ linear unabhängig sind:

Man ergänze $a_1, \ldots, a_k$ zu einer Basis $a_1, \ldots, a_n$ des $\mathbb{R}^n$. Es gibt eine lineare, also stetige Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ so dass gilt $\varphi(a_i) = e_i$ für alle i. Weiterhin sei

$$L = \{\alpha \in \mathbb{R}^n : \alpha \geq 0 \ , \ \alpha_{k+1} = \cdots = \alpha_n = 0\} \ ,$$

dann ist L abgeschlossen und daher auch $\varphi^{-1}(L)$ als Urbild unter einer stetigen Abbildung. Nun sei $x \in \varphi^{-1}(L)$. Es gibt α_i, so dass gilt $x = \sum \alpha_i a_i$. Es folgt:

$$\begin{aligned} x \in \varphi^{-1}(L) &\Leftrightarrow \varphi(x) \in L \\ &\Leftrightarrow \textstyle\sum \alpha_i \varphi(a_i) \in L \\ &\Leftrightarrow \textstyle\sum \alpha_i e_i \in L \\ &\Leftrightarrow (\alpha_1, \ldots, \alpha_n)^t \in L \\ &\Leftrightarrow \alpha_1, \cdots, \alpha_k \geq 0 \ , \ \alpha_{k+1} = \cdots = \alpha_n = 0 \\ &\Leftrightarrow x \in A \ . \end{aligned}$$

Also folgt $A = \varphi^{-1}(L)$ und A ist in diesem Fall abgeschlossen.

Ich zeige nun die Behauptung der Proposition durch vollständige Induktion nach k. Im Fall $k = 1$ gilt $a_1 = 0$ und daher $A = \{0\}$ oder $a_1 \neq 0$ und in diesem Fall ist a_1 linear unabhängig. Die Behauptung gelte also für $k - 1$. Es ist zu zeigen, dass

$$A = \{\sum_{i=1}^{k} \alpha_i a_i : \alpha_i \geq 0\}$$

abgeschlossen ist.

Falls $a_1, \ldots, a_k$ linear unabhängig sind, ist die Behauptung schon bewiesen. Andernfalls setze man für $1 \leq j \leq k$

$$A_j = \{\sum_{i=1}^{k} \alpha_i a_i : \alpha_i \geq 0 \ , \ \alpha_j = 0\} \ ,$$

dann sind alle A_j nach Induktionsannahme abgeschlossen und es reicht zu zeigen, dass $A = A_1 \cup \ldots \cup A_k$ gilt. :

Es sei $x \in A$, $x = \sum \alpha_i a_i$ mit $\alpha_i \geq 0$. Dann gibt es nach 3.10 $\beta_1, \ldots, \beta_k \geq 0$ mit $\beta_j = 0$ für ein j so dass gilt $x = \sum \beta_i a_i$, d.h. es gilt $x \in A_j$. ■

Nach diesen vorbereitenden Bemerkungen über konvexe Mengen komme ich nun wieder zur Optimierung. Es seien $K \subseteq \mathcal{D} \subseteq \mathbb{R}^n$. Wenn $f : \mathcal{D} \to \mathbb{R}$ eine differenzierbare Abbildung ist, und $x^* \in K$ das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

löst, dann gilt $\nabla f(x^*) = 0$, wenn x^* ein **innerer Punkt** von K ist. So ein Punkt heißt dann stationär. Wenn nun K konvex ist, kann man eine Abschwächung dieser Aussage beweisen. Unglücklicherweise heißt so ein Punkt auch stationär:

Definition 3.14 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $K \subseteq \mathcal{D}$ und $f : \mathcal{D} \to \mathbb{R}$ eine differenzierbare Abbildung. Vorgegeben sei das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \end{array}$$

*Ein Punkt $x^* \in K$ heißt **stationärer Punkt** des MPs, wenn gilt*

$$\nabla f(x^*)^t (x - x^*) \geq 0 \qquad \textit{für alle } x \in K$$

Bemerkung 3.15 *Es sei $x^* \in K$ ein innerer Punkt, der im Sinn der vorherigen Definition stationär ist. Dann gilt*

$$\nabla f(x^*)^t (x - x^*) \geq 0 \qquad \textit{für alle } x \in K$$

Sei nun $d \in \mathbb{R}^n$ beliebig, dann gibt es ein $\alpha \neq 0$ so dass gilt $x = x^* \pm \alpha d \in K$. Es folgt

$$\pm\alpha \nabla f(x^*)^t d = \nabla f(x^*)^t (x - x^*) \geq 0$$

und daraus $\nabla f(x^*)^t d = 0$. Da dies für alle $d \in \mathbb{R}^n$ gilt, folgt $\nabla f(x^*) = 0$ und x^* ist ein stationärer Punkt des MPs im ursprünglichen Sinn.

Proposition 3.16 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}$ differenzierbar und $K \subseteq \mathcal{D}$ konvex. Es sei $x^* \in K$ eine lokale Lösung des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \end{array}$$

d.h. es gebe ein $r > 0$ so dass x^* Lösung des MPs

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \cap B(x^*, r] \end{array}$$

ist. Dann ist x^* ein stationärer Punkt des MPs.

Beweis Es sei $x \in K$ beliebig. Da K konvex ist, gilt

$$x^* + \alpha(x - x^*) = \alpha x + (1 - \alpha) x^* \in K \qquad \text{für alle } 0 \leq \alpha \leq 1$$

Da x^* eine lokale Lösung des MPs ist, gibt es ein $\varepsilon > 0$ so dass gilt

$$f(x^*) \leq f(x^* + \alpha(x - x^*)) \qquad \text{für alle } 0 \leq \alpha \leq \varepsilon$$

Es folgt für alle $\alpha \in (0, \varepsilon]$:

$$0 \leq \frac{f(x^* + \alpha(x - x^*)) - f(x^*)}{\alpha} \longrightarrow_{\alpha \to 0} \nabla f(x^*)^t (x - x^*) \qquad \blacksquare$$

Ich komme nun zu einem der wichtigsten Hilfsmittel der Theorie konvexer Minimalprobleme, den Trennungssätzen. Trennungssätze spielen in der Funktionalanalysis eine wichtige Rolle, allerdings beruhen ihre Beweise schon für normierte Räume auf dem Satz von Hahn-Banach. Um diesen zu vermeiden, werde ich mich im folgenden auf den endlich-dimensionalen Fall beschränken. Ich weise aber darauf hin, dass die Sattze 3.21 und 3.26 in jedem normierten Raum oder allgemeiner in jedem lokalkonvexen topologischen Vektorraum gelten, während 3.25 im wesentlichen nur in $\mathbb{R}^n$ richtig ist.

Definition 3.17 *Es seien $A, B \subseteq \mathbb{R}^n$. Man sagt, dass man die Mengen A und B* ***trennen*** *kann, wenn es ein $c \in \mathbb{R}^n \setminus \{0\}$ und ein $\gamma \in \mathbb{R}$ so gibt, dass gilt:*

$$c^t x \leq \gamma \leq c^t y \qquad \text{für alle } x \in A \text{ und alle } y \in B$$

Man sagt, dass man die Mengen A und B ***strikt trennen*** *kann, wenn es ein $c \in \mathbb{R}^n \setminus \{0\}$ und ein $\gamma \in \mathbb{R}$ so gibt, dass gilt:*

$$c^t x < \gamma < c^t y \qquad \text{für alle } x \in A \text{ und alle } y \in B$$

Man sagt dann auch, dass c die beiden Mengen (strikt) trennt.

Die Menge $\{x \in \mathbb{R}^n : c^t x = \gamma\}$ nennt man auch **Hyperebene**, im Fall $n = 2$ ist sie eine Gerade, im Fall $n = 3$ eine Ebene. Die Tatsache, dass man zwei Mengen (strikt) trennen kann bedeutet dann geometrisch, dass es eine Hyperebene gibt, so dass die beiden Mengen "auf verschiedenen Seiten" der Hyperebene liegen (wobei sie im strikten Fall die Hyperebene nicht schneiden).

Lemma 3.18 *Es seien $K \subseteq \mathbb{R}^n$ eine abgeschlossene, konvexe Menge und $0 \notin K$. Dann gibt es ein $c \in \mathbb{R}^n \setminus \{0\}$, und ein $\gamma > 0$ so dass gilt*

$$c^t x \geq \gamma > 0 \qquad \text{für alle } x \in K \ .$$

Beweis Falls $K = \emptyset$ gilt, ist nichts zu zeigen, sei also $K \neq \emptyset$. Man definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch $f(x) = \|x\|^2 = x^t x$, dann ist f stetig differenzierbar und es gilt $\nabla f(x) = 2x$ für alle $x \in \mathbb{R}^n$.

Man wähle ein $r > 0$ so dass gilt $K \cap B(0, r] \neq \emptyset$. Dann ist $K \cap B(x_0, r]$ kompakt und daher nimmt f auf $K \cap B(x_0, r]$ sein Minimum an. Also ist das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \cap B(0, r] \end{array}$$

lösbar. Sei $x^* \in K$ eine Lösung. Dann ist x^* offenbar auch eine Lösung des MPs

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

Da K konvex ist, ist x^* nach 3.16 ein stationärer Punkt, also folgt:

$$\nabla f(x^*)^t(x - x^*) \geq 0 \qquad \text{für alle } x \in K \ .$$

Da $0 \notin K$ gilt, ist $\nabla f(x^*) = 2x^* \neq 0$ und es gilt mit $c = \nabla f(x^*) = 2x^*$ sowie $\gamma = \|x^*\|^2$:

$$c^t x \geq c^t x^* = 2(x^*)^t x^* = 2\|x^*\|^2 = 2\gamma > 0 \qquad \text{für alle } x \in K \ . \qquad \blacksquare$$

Lemma 3.19 *Es seien $A, B \subseteq \mathbb{R}^n$, A abgeschlossen und B kompakt. Dann ist $B - A$ abgeschlossen.*

Beweis Es sei $(y_j - x_j)$ eine Folge in $B - A$, die gegen ein z_0 konvergiert. Da B kompakt ist, gibt es eine Teilfolge (y_{j_k}), die gegen ein $y_0 \in B$ konvergiert. Dann konvergiert $(x_{j_k}) = (y_{j_k} - (y_{j_k} - x_{j_k}))$ gegen $y_0 - z_0$ Da A abgeschlossen ist, folgt $y_0 - z_0 \in A$ und daraus $z_0 = y_0 - (y_0 - z_0) \in B - A$. $\blacksquare$

Beispiel 3.20 *Es seien*

$$B = \{(u, v)^t \in \mathbb{R}^2 : u > 0 \ , \ 1/u \leq v\}$$

und

$$A = \{(w, 0)^t \in \mathbb{R}^2 : w \in \mathbb{R}\} \ .$$

Dann sind A und B konvexe, abgeschlossenen Mengen, aber $B - A$ ist nicht abgeschlossen.

Weiterhin sind A und B konvexe, abgeschlossene, disjunkte Mengen, die man nicht strikt trennen kann.

Satz 3.21 *Es seien $A, B \subseteq \mathbb{R}^n$ disjunkte, nicht-leere, konvexe Mengen, A abgeschlossen, B kompakt. Dann gibt es ein $c \in \mathbb{R}^n \setminus \{0\}$, ein $\gamma \in \mathbb{R}$ und ein $\varepsilon > 0$ so dass gilt:*

$$c^t x \leq \gamma < \gamma + \varepsilon \leq c^t y \qquad \textit{für alle } x \in A \textit{ und alle } y \in B \ .$$

Insbesondere kann man A und B strikt trennen.

Beweis $B-A$ ist nach 3.4 konvex, nach 3.19 abgeschlossen und es gilt $0 \notin B-A$, da die Mengen disjunkt sind. Nach 3.18 gibt es ein $c \in \mathbb{R}^n$ und ein ε so dass gilt:

$$c^t(y-x) \geq \varepsilon > 0 \qquad \text{für alle } x \in A\ ,\ y \in B\ .$$

Es folgt

$$c^t x + \varepsilon \leq c^t y \qquad \text{für alle } x \in A\ ,\ y \in B\ .$$

Also ist $c^t A$ nach oben und $c^t B$ nach unten beschränkt und es gilt:

$$\sup\{c^t x : x \in A\} + \varepsilon \leq \inf\{c^t y : y \in B\}\ .$$

Die Behauptung folgt jetzt z.B. mit $\gamma = \sup\{c^t x : x \in A\}$. Weiterhin gilt

$$c^t x < \gamma + \varepsilon/2 < c^t y \qquad \text{für alle } x \in A,\ y \in B$$

also kann man A und B strikt trennen. ■

Zum Beweis des allgemeinen Trennungssatzes muss man nun noch ein wenig arbeiten. Es gibt viele Wege dahin. Der hier vorgestellte Weg benutzt die sogenannte "Heine-Borel-Eigenschaft" kompakter Mengen:

Satz 3.22 *Es seien $K \subseteq \mathbb{R}^n$ eine kompakte Menge und $\mathcal{C}$ eine Menge offener Teilmengen von $\mathbb{R}^n$, die K überdeckt, d.h. es gelte*

$$K \subseteq \bigcup \mathcal{C} = \bigcup \{U : U \in \mathcal{C}\}$$

Dann gibt es eine endliche Teilüberdeckung von K, d.h. es gibt $U_1, \ldots, U_k \in \mathcal{C}$ so dass gilt

$$K \subseteq U_1 \cup \ldots \cup U_k$$

Beweis Ich beweise zunächst den Fall, dass $\mathcal{C}$ abzählbar ist, d.h. es gilt $\mathcal{C} = \{U_n : n \in \mathbb{N}\}$. Angenommen, das ist falsch, dann wähle man zu jedem $k \in \mathbb{N}$ ein $x_k \in K \setminus (U_1 \cup \ldots \cup U_k)$. Da K kompakt ist, besitzt die Folge (x_k) einen Häufungspunkt $x_0 \in K$. Dann gibt es ein k_0 so dass gilt $x_0 \in U_{k_0}$. Da U_{k_0} offen ist, gibt es unendlich viele k so dass gilt $x_k \in U_{k_0}$, im Widerspruch zu $x_k \notin U_{k_0}$ für alle $k \geq k_0$.

Im allgemeinen Fall gibt es zu jedem $x \in K$ ein $U_x \in \mathcal{C}$ so dass gilt $x \in U_x$. Da U_x offen ist, gibt es ein $r_x \in \mathbb{Q},\ r_x > 0$ so dass gilt

$$x \in B(x, 2r_x) \subseteq U_x$$

Man wähle ein $q_x \in B(x, r_x) \cap \mathbb{Q}^n$, dann gilt

$$x \in B(q_x, r_x) \subseteq B(x, 2r_x) \subseteq U_x$$

Also gilt

$$K \subseteq \bigcup_{x \in K} B(q_x, r_x)$$

Da die Menge $\{B(q_x, r_x) : x \in K\}$ abzählbar ist, gibt es nach dem Bewiesenen $x_1, \ldots, x_k \in K$ so dass gilt $K \subseteq B(q_{x_1}, r_{x_1}) \cup \ldots \cup B(q_{x_k}, r_{x_k})$. Es folgt $K \subseteq U_{x_1} \cup \ldots \cup U_{x_k}$. ■

Korollar 3.23 *Es sei $\mathcal{A}$ eine nicht-leere Familie abgeschlossener Teilmengen von $\mathbb{R}^n$ und $K \subseteq \mathbb{R}^n$ kompakt. Es gelte*

$$A_1 \cap \ldots \cap A_n \cap K \neq \emptyset \qquad \textit{für alle } A_1, \ldots, A_n \in \mathcal{A}$$

(man sagt, $\mathcal{A}$ hat die endliche Durchschnittseigenschaft EDE in K), dann gilt

$$\bigcap \{A \in \mathcal{A}\} \cap K \neq \emptyset \,,$$

d.h. es gibt ein $x \in K$ so dass gilt $x \in A$ für alle $A \in \mathcal{A}$.

Beweis Angenommen, die Behauptung ist falsch, dann gilt $K \subseteq \mathbb{R}^n \setminus \bigcap \{A : A \in \mathcal{A}\}$. Man setze

$$\mathcal{C} = \{\mathbb{R}^n \setminus A : A \in \mathcal{A}\}$$

dann ist $\mathcal{C}$ eine Menge offener Teilmengen von $\mathbb{R}^n$ und es gilt nach de Morgan:

$$\bigcup \{U : U \in \mathcal{C}\} = \bigcup \{\mathbb{R}^n \setminus A : A \in \mathcal{A}\} = \mathbb{R}^n \setminus \bigcap \{A : A \in \mathcal{A}\} \supseteq K$$

Nach 3.22 gibt es $A_1, \ldots, A_k \in \mathcal{C}$ so dass gilt

$$K \subseteq (\mathbb{R}^n \setminus A_1) \cup \ldots \cup (\mathbb{R}^n \setminus A_k)$$

und es folgt

$$K \cap A_1 \cap \ldots \cap A_k = \emptyset$$

im Widerspruch zur Voraussetzung. ■

Lemma 3.24 *Es seien $A \subseteq \mathbb{R}^n$ eine konvexe Menge und $0 \notin A$. Dann gibt es eine Hyperebene, die $\{0\}$ und A trennt.*

Beweis Man setze $K = \{d \in \mathbb{R}^n : \|d\| = 1\}$, dann ist K kompakt. Für alle $x \in A$ sei

$$A_x = \{d \in \mathbb{R}^n : \|d\| = 1 \,,\; d^t x \geq 0\} \,.$$

Es reicht zu zeigen, dass $\{A_x : x \in K\}$ EDE in K besitzt, die Behauptung folgt dann aus 3.23. Also seien $x_1, \ldots x_k \in A$, dann gilt $B := co(\{x_1, \ldots, x_k\}) \subseteq A$ und daher $0 \notin B$. Nach 3.12 ist B kompakt, also abgeschlossen und nach 3.21 gibt es ein $d \neq 0$ und γ so dass gilt

$$0 = d^t 0 \leq \gamma \leq d^t y \qquad \text{für alle } y \in B \,.$$

Es folgt $\frac{1}{\|d\|} d \in A_{x_1} \cap \ldots \cap A_{x_k} \cap K$. ■

Beachten Sie, dass der Beweis dieses harmlos aussehenden Lemmas, das das Herz des nachfolgenden Trennungssatzes ist, die beiden nicht-trivialen Ergebnisse 3.21 und 3.23 benutzt, also in keiner Weise trivial ist. Andererseits ist es nur ein Spezialfall dieses Trennungssatzes, so dass es nach meiner Philosophie nicht selbst "Satz" genannt werden sollte.

Satz 3.25 *Es seien $A, B \subseteq \mathbb{R}^n$ konvexe, nicht-leere, disjunkte Mengen. Dann kann man A und B trennen.*

Beweis Die Menge $B - A$ ist konvex und es gilt $0 \notin B - A$. Nach 3.24 gibt es ein $c \neq 0$ so dass gilt $0 \leq c^t(y - x)$ für alle $x \in A, y \in B$. Es folgt

$$c^t x \leq c^t y \qquad \text{für alle } x \in A,\ y \in B$$

Dies impliziert

$$\sup\{c^t x : x \in A\} \leq \inf\{c^t y : y \in B\}$$

und man kann jedes $\gamma \in [\sup\{c^t x : x \in A\}, \inf\{c^t y : y \in B\}]$ wählen. ■

Korollar 3.26 *Es seien $A, B \subseteq \mathbb{R}^n$ nicht-leere konvexe, disjunkte Mengen und A offen. Dann gibt es ein $c \in \mathbb{R}^n$ und ein $\gamma \in \mathbb{R}$ so dass gilt:*

$$c^t x < \gamma \leq c^t y \qquad \textit{für alle } x \in A\ ,\ y \in B$$

Beweis Nach 3.25 gibt es ein $c \neq 0$ so dass gilt

$$c^t x \leq c^t y \qquad \text{für alle } x \in A,\ y \in B$$

Sei $x \in A$. Da A offen ist, folgt $x + \varepsilon c \in A$ für ein $\varepsilon > 0$ und es folgt

$$\gamma \geq c^t(x + \varepsilon c) = c^t x + \varepsilon \|c\|^2 > c^t x$$

■

Das folgende Lemma von Farkas geht an zentraler Stelle beim Beweis der Existenz von Lagrange-Multiplikatoren ein:

Satz 3.27 *(Lemma von Farkas) Es seien $a_1, \ldots, a_k, b \in \mathbb{R}^n$. Dann sind äquivalent:*

(i) Es gibt $\lambda_1, \ldots, \lambda_k \geq 0$ so dass gilt

$$b = \lambda_1 a_1 + \cdots + \lambda_k a_k\ .$$

(ii) Für alle $x \in \mathbb{R}^n$ folgt aus $a_i^t x \leq 0$ für $i = 1, \ldots, k$ stets $b^t x \leq 0$.

Beweis "(i) $\Rightarrow$ (ii)" Es gelte $a_i^t x \leq 0$ für $i = 1, \ldots, k$ und es gebe $\lambda_1, \ldots, \lambda_k \geq 0$ so dass gilt:

$$b = \lambda_1 a_1 + \cdots + \lambda_k a_k \ ,$$

dann folgt für alle $x \in \mathbb{R}^n$:

$$b^t x = \lambda_1 a_1^t x + \cdots + \lambda_k a_k^t x \leq 0 \ .$$

"(ii) $\Rightarrow$ (i)" Es sei

$$A = \{\lambda_1 a_1 + \cdots + \lambda_k a_k : \lambda_i \geq 0\} \, ,$$

dann ist zu zeigen, dass $b \in A$ gilt. Angenommen, $b \notin A$. Da A nach 3.13 abgeschlossen und konvex ist, gibt es nach 3.21 ein $c \in \mathbb{R}^n$ und ein $\gamma \in \mathbb{R}$ so dass gilt:

$$c^t a \leq \gamma < c^t b \qquad \text{für alle } a \in A \ .$$

Wegen $0 \in A$ gilt $\gamma \geq 0$. Sei $a \in A$, dann folgt $ka \in A$ für alle $k \in \mathbb{N}$ und daraus

$$kc^t a = c^t(ka) \leq \gamma \qquad \text{für alle } k \in \mathbb{N} \ .$$

Es folgt $c^t a \leq 0$ für alle $a \in A$, speziell also $c^t a_i \leq 0$ für alle i, und $c^t b > 0$ im Widerspruch zu (i). ■

Bemerkung 3.28 Für die nicht-triviale Richtung (ii) $\Rightarrow$ (i) von 3.27 gibt es einen schönen Beweis, der die Dualitätstheorie der linearen Optimierung benutzt:

Man setze $A = (a_1, \ldots, a_r)^t$, dann gilt:

Aus $A^t x \geq 0$ folgt stets $b^t x \geq 0$.

Also hat das lineare MP

$$\begin{array}{ll} \min & b^t x \\ \text{bez.} & A^t x \geq 0 \end{array}$$

die Lösung $x^* = 0$. Daher ist auch das duale Programm lösbar. Dieses hat die Form:

$$\begin{array}{ll} \max & 0^t y \\ \text{bez.} & Au = b \\ & u \geq 0 \end{array}$$

Also gibt es ein $u \geq 0$ mit $Au = b$ und es folgt

$$b = u_1 a_1 + \cdots + u_r a_r \ .$$

Kapitel 4

Konvexe Abbildungen

Definition 4.1 *Es sei $K \subseteq \mathbb{R}^n$.*

*(i) Eine Abbildung $f : K \to \mathbb{R}^p$ heißt **konvex**, wenn K konvex ist und weiterhin gilt:*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$$
$$\text{für alle } x, y \in K \text{ und alle } \alpha \in [0,1]$$

*(ii) Eine Abbildung $f : K \to \mathbb{R}$ heißt **streng konvex**, wenn K konvex ist und außerdem gilt:*

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$$
$$\text{für alle } x, y \in K,\ x \neq y \text{ und alle } \alpha \in (0,1)$$

Bemerkung 4.2 *Es sei $K \subseteq \mathbb{R}^n$ konvex.*

(i) Es sei $I \subseteq \mathbb{R}$ ein Intervall. Eine Abbildung $f : I \to \mathbb{R}$ ist genau dann konvex, wenn für alle $x, y \in I$ die Verbindungsgerade zwischen den Punkten $(x, f(x))$ und $(y, f(y))$ nicht unterhalb des Graphen von f liegt.

(ii) Eine Abbildung $f : K \to \mathbb{R}^p$ ist genau dann konvex, wenn alle f_i konvex sind.

(iii) Es sei $K \subseteq \mathbb{R}^n$. Eine Abbildung $f : K \to \mathbb{R}^p$ ist genau dann konvex, wenn

$$Epi(f) = \{(x, r) \in K \times \mathbb{R}^p : f(x) \leq r\}$$

*konvex ist. Man nennt $Epi(f)$ den **Epigraphen** von f.*

Beweis Das ist eine Übungsaufgabe. ■

Ich werde in 4.13 ein sehr einfaches Kriterium für die (strenge) Konvexität einer differenzierbaren Abbildung beweisen, das man in der Regel anwenden kann, und beschränke mich daher auf zwei einfache Beispiele:

Beispiele 4.3

(i) $|\cdot| : \mathbb{R} \to \mathbb{R}$ *ist konvex, aber nicht streng konvex. Allgemeiner ist jede Norm* $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ *konvex, aber nicht streng konvex.*

(ii) Eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}^p$ *heißt* ***affin****, wenn es eine Matrix* $A \in M(p,n)$ *und ein* $d \in \mathbb{R}^p$ *so gibt, dass gilt*

$$f(x) = Ax + d \qquad \text{für alle } x \in \mathbb{R}^n$$

Jede affine Abbildung $f : \mathbb{R}^n \to \mathbb{R}^p$ *ist konvex, für* $p = 1$ *ist sie nicht streng konvex.*

Beweis

(i) Für alle $x, y \in \mathbb{R}^n$ und alle $\alpha \in (0,1)$ gilt:

$$\|\alpha x + (1-\alpha)y\| \leq \|\alpha x\| + \|(1-\alpha)y\| = \alpha\|x\| + (1-\alpha)\|y\| \,,$$

also ist $\|\cdot\|$ konvex. Für $\lambda \geq 0$ und $y = \lambda x$ gilt:

$$\begin{aligned} \|\alpha x + (1-\alpha)y\| &= \|\alpha x + \lambda(1-\alpha)x\| \\ &= |(\alpha + \lambda(1-\alpha))|\,\|x\| \\ &= (\alpha + \lambda(1-\alpha))\|x\| \\ &= \alpha\|x\| + (1-\alpha)\|y\| \,, \end{aligned}$$

also ist f nicht streng konvex.

(ii) Für alle $x, y \in \mathbb{R}^n$ und alle $\alpha \in \mathbb{R}$ gilt:

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= A(\alpha x + (1-\alpha)y) + d \\ &= \alpha Ax + (1-\alpha)Ay + \alpha d + (1-\alpha)d \\ &= \alpha f(x) + (1-\alpha)f(y) \end{aligned}$$

■

Lemma 4.4 *Es seien* $K \subseteq \mathbb{R}^n$ *konvex und* $f : K \to \mathbb{R}$ *eine Abbildung. Dann gelten:*

(i) f *ist genau dann konvex, wenn gilt:*

$$f(x + \alpha(y-x)) \leq f(x) + \alpha(f(y) - f(x)) \qquad \text{für alle } x, y \in K \text{ und alle } \alpha \in [0,1] \,.$$

f *ist genau dann streng konvex, wenn gilt:*

$$f(x + \alpha(y-x)) < f(x) + \alpha(f(y) - f(x)) \qquad \text{für alle } x, y \in K,\ x \neq y \text{ und alle } \alpha \in (0,1) \,.$$

(ii) f *ist genau dann konvex, wenn für alle* $x_1, \ldots, x_k \in K$ *und alle* $\alpha_1, \ldots, \alpha_k \geq 0$ *mit* $\alpha_1 + \cdots + \alpha_k = 1$ *gilt:*

$$f(\alpha_1 x_1 + \cdots + \alpha_k x_k) \leq \alpha_1 f(x_1) + \cdots + \alpha_k f(x_k) \,.$$

Beweis Der Beweis von (i) und (ii) folgt unmittelbar aus der Tatsache, dass

$$x + \alpha(y - x) = \alpha y + (1 - \alpha)x$$

gilt.

(iii) Offenbar impliziert die Bedingung die Konvexität. Es seien f konvex, $x_0, \ldots, x_k$ Elemente aus K und $\alpha_1, \ldots, \alpha_k \in [0, 1]$, $\sum \alpha_i = 1$. Dann gilt $(x_i, f(x_i)) \in Epi(f)$ für alle i. Da f konvex ist, ist $Epi(f)$ nach 4.2 konvex und es folgt

$$(\sum \alpha_i x_i, \sum \alpha_i f(x_i)) = \sum \alpha_i (x_i, f(x_i)) \in Epi(f)$$

nach 3.5 und daraus die Behauptung. ■

Lemma 4.5 *Es seien $K \subseteq \mathbb{R}^n$ konvex und $f, g : K \to \mathbb{R}^p$ konvexe Abbildungen sowie $\alpha \geq 0$. Dann sind $f + g$ und αf konvexe Abbildungen.*

Beweis Der Beweis ist trivial. ■

Satz 4.6 *Es sei $f : K \to \mathbb{R}$ eine konvexe Abbildung und $x^* \in K$ eine lokale Lösung des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \end{array}$$

Dann ist x^* eine Lösung des MPs. Wenn f streng konvex ist, ist x^* die einzige Lösung des MPs.

Beweis Es sei $x \in K$ beliebig. Dann gibt es ein $\varepsilon \in (0, 1)$ so dass für alle $0 < \alpha \leq \varepsilon$ gilt $f(x^*) \leq f(x^* + \alpha(x - x^*))$. Es folgt:

$$f(x^*) \leq f(x^* + \alpha(x - x^*)) \leq f(x^*) + \alpha(f(x) - f(x^*))$$

und daraus $f(x) \geq f(x^*)$.

Wenn f streng konvex ist, ist die erste Ungleichung strikt und es folgt die Behauptung. ■

4.6 ist einer der Gründe für die Wichtigkeit konvexer Abbildungen in der Optimierung. Wie schon mehrmals betont, liefern Verfahren zur Bestimmung von Lösungen von Minimalproblemen in der Regel nur lokale Lösungen. Wenn nun das Problem konvex ist, ist jede lokale Lösung automatisch eine Lösung.

Erinnerung 4.7 Eine symmetrische Matrix $A \in M(n, n)$ heißt **positiv semidefinit**, wenn gilt

$$x^t A x \geq 0 \qquad \text{für alle } x \in \mathbb{R}^n$$

A heißt **positiv definit**, wenn gilt

$$x^t A x > 0 \qquad \text{für alle } x \in \mathbb{R}^n,\ x \neq 0$$

Da jede symmetrische Matrix A diagonalisierbar ist, sieht man leicht, dass A genau dann positiv definit bzw. semidefinit ist, wenn alle Eigenwerte positiv bzw. nicht-negativ sind.

Bekanntlich ist die Hesse-Matrix einer zweimal stetig differenzierbaren Abbildung in jedem Punkt symmetrisch.

Bemerkung 4.8 *Es sei $A \in M(n,n)$ symmetrisch, dann gilt für alle $x, y \in \mathbb{R}^n$:*

$$(x+y)^t A(x+y) = x^t Ax + 2x^t Ay + y^t Ay \; .$$

Beweis Es gilt

$$(x+y)^t A(x+y) = x^t Ax + x^t Ay + y^t Ax + y^t Ay$$

und die Behauptung folgt aus

$$y^t Ax = (y^t Ax)^t = x^t A^t y = x^t Ay \; .$$

■

Mit Hilfe von positiv (semi-)definiten Matrizen erhält man eine wichtige Klasse (streng) konvexer Abbildungen:

Proposition 4.9 *Es sei $A \in M(n,n)$ eine symmetrische Matrix. Man definiere*

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

definiert durch

$$f(x) = x^t Ax$$

Dann gelten:

(i) f ist genau dann konvex, wenn A positiv semi-definit ist.

(ii) f ist genau dann streng konvex, wenn A positiv definit ist.

Beweis

(ii) Es sei A positiv definit, dann gilt für alle $x, y \in \mathbb{R}^n,\ x \neq y$ und alle $0 < \alpha < 1$:

$$\begin{aligned}
f(x+\alpha(y-x)) &= x^t Ax + 2\alpha x^t A(y-x) + \alpha^2 (y-x)^t A(y-x) \\
&< f(x) + 2\alpha x^t A(y-x) + \alpha (y-x)^t A(y-x) \\
&= f(x) + 2\alpha x^t A(y-x) + \alpha y^t A(y-x) - \alpha x^t A(y-x) \\
&= f(x) + \alpha x^t A(y-x) + \alpha y^t A(y-x) \\
&= f(x) - \alpha x^t Ax + \alpha y^t Ay \\
&= f(x) + \alpha (f(y) - f(x))
\end{aligned}$$

Wenn A nicht positiv definit ist, gibt es ein $x \neq 0$ mit $x^t Ax \leq 0$ und es folgt für alle $\alpha \in (0,1)$:

$$f(\alpha x + (1-\alpha)0) = f(\alpha x) = \alpha^2 x^t Ax \geq \alpha x^t Ax = \alpha f(x) + (1-\alpha) f(0)$$

und f ist streng nicht konvex.

(ii) beweist man wie (i) mit den entprechenden Modifikationen. ■

Beispiel 4.10 *Die Abbildung $f : \mathbb{R}^n \to \mathbb{R}$ definiert durch $f(x) = x^t x = \|x\|^2$ ist streng konvex.*

Proposition 4.11 *Es seien $K \subseteq \mathbb{R}^n$ konvex, $\mathcal{D} \supseteq K$ offen und $f : \mathcal{D} \to \mathbb{R}$ differenzierbar. Dann gelten:*

(i) f ist genau dann konvex auf K, d.h. $f|K$ ist konvex, wenn gilt:

$$f(x) + \nabla f(x)^t (y-x) \leq f(y) \qquad \text{für alle } x, y \in K \ .$$

(ii) f ist genau dann streng konvex auf K, gilt:

$$f(x) + \nabla f(x)^t (y-x) < f(y) \quad \text{für alle } x, y \in K,\ x \neq y \ .$$

Beweis

(i) Die Behauptung gelte und es seien $x, y \in K,\ 0 \leq \alpha \leq 1$. Man setze $z = \alpha x + (1-\alpha) y$, dann gilt:

$$f(z) + \nabla f(z)^t (x-z) \leq f(x)$$

$$f(z) + \nabla f(z)^t (y-z) \leq f(y)$$

und es folgt:

$$\alpha f(z) + \alpha \nabla f(z)^t (x-z) + (1-\alpha) f(z) + (1-\alpha) \nabla f(z)^t (y-z) \leq \alpha f(x) + (1-\alpha) f(y)$$

Nun gilt

$$\alpha(x-z) + (1-\alpha)(y-z) = \alpha x + (1-\alpha) y - z = 0 \ ,$$

und daher

$$\begin{aligned} \alpha \nabla f(z)^t (x-z) + (1-\alpha) \nabla f(z)^t (y-z) &= \nabla f(z)^t (\alpha(x-z) + (1-\alpha)(y-z)) \\ &= 0 \ , \end{aligned}$$

also ist f konvex.

Umgekehrt sei f konvex, dann gilt für alle $x, y \in K,\ x \neq y,\ 0 < \alpha \leq 1$:

$$f(x + \alpha(y-x)) \leq f(x) + \alpha(f(y) - f(x))$$

und es folgt

$$\frac{f(x+\alpha(y-x))-f(x)}{\alpha} \leq f(y)-f(x)$$

Nun gilt

$$\lim_{\alpha\to 0}\frac{f(x+\alpha(y-x)-f(x)}{\alpha} = \nabla f(x)^t(y-x)$$

und es folgt

$$\nabla f(x)^t(y-x) \leq f(y)-f(x)$$

(ii) Eine leichte Modifikation des Beweises von (i) zeigt, dass f streng konvex ist, wenn die Bedingung erfüllt ist. Umgekehrt sei f streng konvex. Angenommen, es gibt $x, y \in K,\ x \neq y$ so dass gilt

$$f(y) = f(x) + \nabla f(x)^t(y-x)\ ,$$

dann folgt für alle $0 < \alpha < 1$

$$\begin{aligned} f(x)+\alpha(f(y)-f(x)) &> f(x+\alpha(y-x)) \\ &\geq f(x)+\nabla f(x)^t(\alpha(y-x)) \\ &= f(x)+\alpha(f(y)-f(x)) \end{aligned}$$

und daraus ein Widerspruch. ■

Die Abbildung

$$y \mapsto f(x) + \nabla f(x)^t(y-x)$$

ist die Tangente von f in x, also besagt 4.11 gerade, dass eine Abbildung genau dann konvex ist, wenn Sie stets oberhalb jeder Tangente verläuft. Analoges gilt für strenge Konvexität.

Satz 4.12 *Es seien $K \subseteq \mathbb{R}^n$ konvex, $\mathcal{D} \supseteq K$ offen sowie $f : \mathcal{D} \to \mathbb{R}$ differenzierbar und konvex auf K. Ein Punkt $x^* \in K$ ist genau dann Lösung des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \end{array}$$

wenn x^* ein stationärer Punkt des MPs ist.

Beweis Wenn x^* eine Lösung des MPs ist, folgt die Behauptung aus 3.16.

Also sei x^* ein stationärer Punkt, dann folgt für alle $y \in K$:

$$f(y) \geq f(x^*) + \nabla f(x^*)(y-x^*) \geq f(x^*)\ .$$ ■

Auch 4.12 scheint wegen des schnellen Beweises eine harmlose Angelegenheit zu sein, aber es ist natürlich ein weitere ganz wichtige Eigenschaft konvexer MPe: Die Lösungen sind genau die stationären Punkte.
xxx

Proposition 4.13 *Es seien $K \subseteq \mathbb{R}^n$ konvex, $\mathcal{D} \supseteq K$ offen und $f : \mathcal{D} \to \mathbb{R}$ zweimal stetig differenzierbar. Dann gelten:*

(i) Wenn $Hf(x)$ für alle $x \in K$ positiv semidefinit ist, ist f in K konvex.

(ii) Wenn $Hf(x)$ für alle $x \in K$ positiv definit ist, ist f in K streng konvex.

Beweis Man wähle $x, y \in K$ und definiere $\varphi : [0,1] \longrightarrow \mathbb{R}$ durch

$$\varphi(\alpha) = f(x + \alpha(y - x))$$

Dann ist φ zweimal stetig differenzierbar und es gilt

$$\varphi'(\alpha) = \nabla f(x + \alpha(y - x))^t (y - x)$$

$$\varphi''(\alpha) = (y - x)^t H f(x + \alpha(y - x))(y - x)$$

Nach dem Satz von Taylor gibt es ein $\xi \in (0,1)$ so dass gilt

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\xi)$$

und es folgt

$$f(y) = f(x) + \nabla f(x)^t (y - x) + \frac{1}{2}(y - x)^t H f(x + \xi(y - x))(y - x)$$

Die Behauptung folgt jetzt aus 4.11. ■

Korollar 4.14 *Es sei $I \subseteq \mathbb{R}$ ein Intervall. Eine zweimal stetig differenzierbare Abbildung $f : I \to \mathbb{R}$ ist konvex, wenn $f''(x) \geq 0$ für alle $x \in I$ gilt und streng konvex wenn $f''(x) > 0$ für alle $x \in I$ gilt.*

Beispiel 4.15 *Die Abbildungen $x \mapsto x^2$ und* exp *sind streng konvex. Weiterhin ist die Abbildung $f : (0, \infty) \to \mathbb{R}$ definiert durch $f(x) = 1/x$ streng konvex und daher ist*

$$Epi(f) = \{(u, v) \in \mathbb{R}^2 : f(u) \leq v\} = \{(u, v) \in \mathbb{R}^2 : u > 0,\ 1/u \leq v\}$$

konvex.

Bemerkung 4.16 *In 4.13(i) gilt die Umkehrung, wenn K einen inneren Punkt besitzt, die Umkehrung in (ii) gilt nicht, wie die Abbildung $f : \mathbb{R} \to \mathbb{R}$ definiert durch $f(x) = x^4$ zeigt.*

Eine der möglichen Iterationen zur Lösung des MPs

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

besteht darin, dass man f durch das Taylor-Polynom zweiten Grades T_2 ersetzt: Wenn also x_k bestimmt ist, betrachtet man

$$T_2(x) = f(x_k) + \nabla f(x_k)^t(x - x_k) + \frac{1}{2}(x - x_k)^t Hf(x_k)(x - x_k)$$

Wenn nun $Hf(x_k)$ positiv semidefinit ist, ist T_2 konvex und man kann dann Verfahren der konvexen Optimierung einsetzen.

Ich beschließe dieses Kapitel mit zwei interessanten Eigenschaften konvexer Funktionen auf offenen Mengen: Sie sind automatisch stetig, und wenn sie differenzierbar sind, sind sie stetig differenzierbar:

Proposition 4.17 *Es seien $K \subseteq \mathbb{R}^n$ konvex und offen sowie $f : K \to \mathbb{R}^p$ konvex. Dann ist f stetig.*

Beweis Es reicht, den Fall $p = 1$ zu betrachten. Es sei $x_0 \in K$. Man setze $K_0 = K - x_0$ und definiere $g : K_0 \to \mathbb{R}$ durch

$$g(x) = f(x + x_0) - f(x_0)$$

dann ist g konvex und es gilt $0 \in K_0$ sowie $g(0) = 0$. Wenn ich nun zeigen kann, dass g in 0 stetig ist, ist f in x_0 stetig und daher nehme ich im folgenden an, dass gilt:

$$0 \in K \quad \text{und} \quad f(0) = 0$$

Ich zeige nun zunächst:

Behauptung Es gibt ein $r > 0$ so dass $f|B(0, r]$ nach oben beschränkt ist.

Beweis Da K offen ist, gibt es ein $s > 0$ so dass gilt

$$B := \{x \in \mathbb{R}^n : \sum |x_i| \leq s\} \subseteq K$$

Sei

$$A = \{\alpha e_i : |\alpha| = s\}$$

dann ist A endlich und es gilt $co(A) = B$. Zu jedem $x \in B$ gibt es $x_0, \ldots x_k \in A$ und $\alpha_1, \ldots, \alpha_k \geq 0$ mit $\sum \alpha_i = 1$ so dass gilt $x = \sum \alpha_i x_i$. Es folgt

$$f(x) = f(\sum \alpha_i x_i) \leq \sum \alpha_i f(x_i) \leq \max\{f(x) : x \in A\} \sum \alpha_i = \max\{f(x) : x \in A\}$$

Wählt man nun ein $r > 0$ mit $B(0, r] \subseteq B$, dann gibt es ein $C > 0$ so dass gilt

$$f(x) \leq C \qquad \text{für alle } x \in B(0, r]$$

Für $x \in B(0, r],\ x \neq 0$ setze man $x_0 = \dfrac{r}{\|x\|}\, x \in B(0, r]$, dann gilt $x = \dfrac{\|x\|}{r}\, x_0$ und es folgt:

$$f(x) = f\left(\frac{\|x\|}{r}x_0 + (1 - \frac{\|x\|}{r})\,0\right) \leq \frac{\|x\|}{r}f(x_0) + (1 - \frac{\|x\|}{r})f(0) \leq \frac{C}{r}\|x\|$$

Weiterhin gilt für alle $x \in B(0,r]$:

$$0 = f(0) = f((1/2)x + (-1/2)x) \leq (1/2)f(x) + (1/2)f(-x)$$

und es folgt:

$$-f(x) \leq f(-x) \leq \frac{C}{r}\| - x\| = \frac{C}{r}\|x\|$$

Insgesamt erhält man

$$|f(x)| \leq \frac{C}{r}\|x\| \qquad \text{für alle } x \in B(0,r]$$

und es folgt die Behauptung. ■

Beispiel 4.18 *Die Abbildung $f : [0,1] \longrightarrow \mathbb{R}$ definiert durch*

$$f(x) = \begin{cases} 1 & x = 0 \\ 0 & x > 0 \end{cases}$$

ist konvex aber unstetig. Also kann man in 4.17 die Voraussetzung, dass K offen ist, nicht weglassen.

Proposition 4.19 *Es seien $K \subseteq \mathbb{R}^n$ konvex, offen und $f : K \to \mathbb{R}^p$ eine differenzierbare, konvexe Abbildung. Dann ist f stetig differenzierbar.*

Beweis OBdA gelte $p = 1$. Sei $x_0 \in K$, dann setze man $K_0 = K - x_0$ und betrachte die Abbildung

$$g : K - x_0 \longrightarrow \mathbb{R}$$

definiert durch

$$g(x) = f(x_0 + x) - \nabla f(x_0)^t x - f(x_0)$$

Dann ist g konvex und gilt $0 \in K_0$, $g(0) = 0$ sowie $\nabla g(0) = 0$. Also kann man oBdA annehmen, dass $0 \in K$ und $f(0) = 0$ sowie $\nabla f(0) = 0$ gelten.

Da f konvex ist, gilt für alle $x \in K$:

$$f(x) \geq f(0) + \nabla f(0)^t(x - 0) = 0$$

also gilt $f(x) \geq 0$ für alle $x \in K$.

Da f differenzierbar ist, gilt

$$f(0 + h) = f(0) + \nabla f(0)^t(h - 0) + R(h)$$

und $\lim_{h\to 0} \frac{R(h)}{\|h\|} = 0$, also folgt $\lim_{h\to 0} \frac{f(h)}{\|h\|} = 0$. Daher gibt es zu vorgegebenem $\varepsilon > 0$ ein $\delta > 0$ so dass gilt

$$\frac{|f(h)|}{\|h\|} \leq \varepsilon \quad \text{für alle } 0 < \|h\| \leq \delta$$

Seien $\|x\| < \delta/2$ und $y \neq 0$. Dann gibt es ein $\alpha > 0$ mit $\|x + \alpha y\| = \delta$. Da f konvex ist, gilt

$$f(x + \alpha y) \geq f(x) + \nabla f(x)^t(\alpha y) \geq \alpha \nabla f(x)^t y$$

und daher

$$\nabla f(x)^t y \leq (1/\alpha) f(x + \alpha y)$$

Weiterhin gilt

$$\delta = \|x + \alpha y\| \leq \|x\| + \alpha\|y\| \leq \delta/2 + \alpha\|y\|$$

und es folgt

$$2\alpha\|y\| \geq \delta = \|x + \alpha y\|$$

Man erhält:

$$\nabla f(x)^t y \leq 2\|y\| \frac{f(x + \alpha y)}{\|x + \alpha y\|} \leq 2\|y\|\varepsilon$$

Setzt man nun $y = \nabla f(x)$, dann folgt

$$\|\nabla f(x)\|^2 \leq 2\|\nabla f(x)\|\varepsilon \qquad \text{für alle } \|x\| \leq \delta/2$$

also

$$\|\nabla f(x)\| \leq 2\varepsilon \qquad \text{für alle } \|x\| \leq \delta/2$$

und daraus die Behauptung. ■

Kapitel 5

Differenzierbare Minimierungsprobleme

Erinnerung 5.1

(i) Es seien $K \subseteq \mathbb{R}^n$ kompakt und $f : K \to \mathbb{R}$ eine stetige Abbildung. Dann hat das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

eine Lösung.

Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $K \subseteq \mathcal{D}$ und $f : \mathcal{D} \to \mathbb{R}$ eine zweimal stetig differenzierbare Abbildung. Dann gelten weiterhin:

(ii) Es sei x^* eine lokale Lösung des MPs

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

Wenn x^* ein innerer Punkt von K ist (d.h. es gibt eine $r > 0$ mit $B(x^*, r) \subseteq K$), gilt $\nabla f(x^*) = 0$ und $Hf(x^*)$ ist positiv semidefinit.

(iii) Es sei $x^* \in K$. Es gelte:

(a) $\nabla f(x^*) = 0$

(b) $Hf(x^*)$ ist positiv definit

Dann ist x^* eine lokale Lösung des MPs

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in \mathcal{D} \end{array}$$

und damit auch des MPs

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

Bezeichnungsweisen 5.2 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ und $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^p$ sowie $h : \mathcal{D} \to \mathbb{R}^q$ Abbildungen. Man setze*

$$K(g,h) = \{x \in \mathcal{D} : g(x) \leq 0 \ , \ h(x) = 0\}$$

Dann schreibt man für das MP

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K(g,h) \end{array}$$

suggestiver

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Weiterhin sei $x_0 \in \mathbb{R}^n$ ein zulässiger Punkt. Dann setzt man

$$I(x_0) = \{i \in \{1, \ldots, p\} : g_i(x_0) = 0\}$$

*Man sagt, dass die Ungleichung $g_i(x) \leq 0$ in x_0 **aktiv** ist, wenn $i \in I(x_0)$, d.h. $g_i(x_0) = 0$ gilt.*

VEREINBARUNG

Falls nicht anders bemerkt, seien im Rest dieses Kapitels $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $K \subseteq \mathcal{D}$ und

$$\begin{array}{l} f : \mathcal{D} \to \mathbb{R} \\ g : \mathcal{D} \to \mathbb{R}^p \\ h : \mathcal{D} \to \mathbb{R}^q \end{array}$$

stetig differenzierbare Abbildungen. Es wird das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

betrachtet.

Lemma 5.3 *Es sei $x^* \in \mathcal{D}$ ein zulässiger Punkt des MPs*

$$\textit{(MP1)} \qquad \begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

x^* ist genau dann eine lokale Lösung von (MP1), wenn x^* eine lokale Lösung des MPs

$$\textit{(MP2)} \qquad \begin{array}{lll} \min & f(x) & \\ \textit{bez.} & g_i(x) \leq 0 & \textit{für alle } i \in I(x^*) \\ & h(x) = 0 & \end{array}$$

ist.

Beweis Wenn x^* eine lokale Lösung von (MP2) ist, ist x^* offenbar auch eine lokale Lösung von (MP1). Sei also x^* eine lokale Lösung von (MP1), dann gibt es ein $\varepsilon > 0$ so dass gilt $f(x^*) \leq f(x)$ für alle $x \in K(f,g) \cap B(x^*, \varepsilon)$. Sei nun $x \in \mathcal{D}$ so gewählt dass gilt $g_i(x) \leq 0$ für alle $i \in I(x_0)$ und $h(x) = 0$. Weiter sei $i \notin I(x_0)$, dann gilt $g_i(x^*) < 0$, da x^* zulässig ist. Da g_i stetig ist, gibt es ein $\varepsilon_i > 0$ so dass gilt $g_i(x) \leq 0$ für alle $x \in B(x^*, \varepsilon_i)$. Dann gilt aber $f(x^*) \leq f(x)$ für alle $x \in B(x^*, \varepsilon) \cap \bigcap\{B(x^*, \varepsilon_i) : i \notin I(x_0)\}$.

Bemerkung 5.4 Wenn x^* eine lokale Lösung von (MP1) ist, ist x^* auch eine lokale Lösung von (MP2) und daher eine lokale Lösung des MPs

$$\text{(MP3)} \qquad \begin{array}{lll} \min & f(x) & \\ \text{bez.} & g_i(x) = 0 & \text{für alle } i \in I(x^*) \\ & h(x) = 0 & \end{array}$$

Wenn nun $(\nabla g_i(x^*)_{i \in I(x_0)}, \nabla h(x^*))$ linear unabhängig sind, gibt es Langrange Multiplikatoren $(\lambda_i)_{i \in I(x_0)}$ und (μ_j) so dass gilt

$$\nabla f(x^*) + \sum_{i \in I(x_0)} \lambda_i \nabla g_i(x^*) + \sum \mu_j \nabla h_j(x^*) = 0$$

Setzt man noch $\lambda_i = 0$ für alle $i \notin I(x_0)$, dann erhält man

$$\begin{aligned} &\nabla f(x^*) + \textstyle\sum \lambda_i \nabla g_i(x^*) + \sum \mu_j \nabla h_j(x^*) = 0 \\ &\lambda_i g_i(x^*) = 0 \qquad \text{für alle } i. \end{aligned}$$

Es bleibt die Frage, welche Verbesserungen man aus der Tatsache bekommen kann, dass x^* sogar (MP1) löst, und ob man die Voraussetzung der linearen Unabhängigkeit der Gradienten nicht abschwächen kann. Dies ist möglich, zu diesem Zweck muss man die "Geometrie" des zulässigen Bereichs näher studieren.

Definition 5.5 *Es seien $K \subseteq \mathbb{R}^n$ und $x_0 \in K$. Ein Vektor $d \in K$ heißt **tangential** an K in x_0, wenn es eine Folge (α_k) in $[0, \infty)$ und eine Folge $(x_k) \in K$ gibt, so dass gelten:*

$$\lim_{k \to \infty} x_k = x_0 \quad \textit{und} \quad d = \lim_{k \to \infty} \alpha_k (x_k - x_0)$$

*Die Menge aller tangentialen Vektoren in x_0 wird mit $\mathcal{T}_K(x_0)$ oder in der Regel mit $\mathcal{T}(x_0)$ bezeichnet, sie heißt auch der **Tangentialkegel** von x_0.*

Beispiele 5.6

(i) Es seien $K \subseteq \mathbb{R}^n$ und x_0 ein innerer Punkt von K, dann gilt $\mathcal{T}(x_0) = \mathbb{R}^n$.

(ii) Es sei

$$K = \{(u,v)^t \in \mathbb{R}^2 : u \geq 1 \ , \ v \geq 1\}$$

dann gilt für $x_0 = (u_0, v_0)^t \in K$*:*

$$\mathcal{T}(x_0) = \begin{cases} \mathbb{R}^2 & : \quad u_0 > 1, v_0 > 1 \\ \{(u,v)^t : u \geq 0\} & : \quad u_0 = 1, v_0 > 1 \\ \{(u,v)^t : v \geq 0\} & : \quad u_0 > 1, v_0 = 1 \\ \{(u,v)^t : u, v \geq 0\} & : \quad u_0 = 1, v_0 = 1 \end{cases}$$

(iii) Man definiere wieder $g : \mathbb{R}^2 \to \mathbb{R}^3$ *durch*

$$g(u,v) = (v - (1-u)^3, -u, -v)^t$$

und setze

$$K = \{x \in \mathbb{R}^2 : g(x) \leq 0\}$$

dann gilt für $e_1 = (1,0)^t$*:*

$$\mathcal{T}_K(e_1) = \{(u,0)^t : u \leq 0\}$$

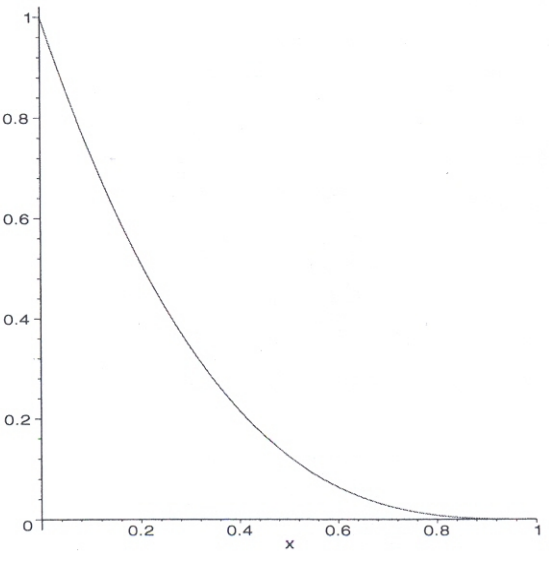
Beweis

(i) Es sei $d \in \mathbb{R}^n$, dann gilt $x_k = x_0 + \frac{1}{k}d \in K$ für alle $i \geq i_0$ und es folgt

$$k(x_k - x_0) = d \to d$$

(ii) Übungsaufgabe

(iii) K hat die Form:



Es sei $d = (\delta_1, \delta_2) \in \mathcal{T}(e_1)$, dann gibt es Folgen (α_k) in $(0, \infty)$ und $((u_k, v_k)^t)$ in K so dass gilt:

$$(u_k, v_k) \to (1, 0) \quad \text{und} \quad \alpha_k((u_k, v_k) - (1, 0)) \longrightarrow (\delta_1, \delta_2)$$

Es folgt:

$$\alpha_k(u_k - 1) \to \delta_1 \qquad \text{und} \qquad \alpha_k v_k \to \delta_2$$

Wegen $\alpha_k(u_k - 1) \leq 0$ folgt $\delta_1 \leq 0$ und aus $0 \leq v_k \leq (1 - u_k)^3$ folgt:

$$0 \leq \alpha_k v_k \leq \alpha_k(1 - u_k)^3 = \alpha_k(1 - u_k)(1 - u_k)^2 \to -\delta_1 \cdot 0 = 0$$

und daraus $\delta_2 = 0$. Also folgt $d = (\delta, 0)^t$ mit $\delta \leq 0$.

Sei umgekehrt $\delta \leq 0$, dann setze man $x_k = (1 - 1/k, 0)^t$, und $\alpha_k = -k\delta$, dann gilt:

$$\alpha_k(x_k - (1, 0)) = -k\alpha(-1/k, 0)^t \to (\delta, 0)^t$$

Bemerkung 5.7 Es seien $K \subseteq \mathbb{R}^n$ und $x_0 \in K$. Weiterhin sei $d \neq 0$ ein tangentialer Vektor an K in x_0. Dann gilt für eine Folge (x_k) in K, die gegen x_0 konvergiert und eine Folge (α_i) in $[0, \infty)$:

$$\alpha_k(x_k - x_0) \to d$$

Es folgt

$$\alpha_k \|x_k - x_0\| \to \|d\|$$

Weiterhin gilt $x_k \neq x_0$ für alle $k \geq k_0$ und es folgt für diese k:

$$\frac{1}{\|x_k - x_0\|}(x_k - x_0) = \frac{1}{\alpha_k \|x_k - x_0\|}\alpha_k(x_k - x_0) \to \frac{1}{\|d\|}d$$

Also gibt es zu jedem tangentialen Vektor $d \neq 0$ eine Folge (x_k) in $K \setminus \{x_0\}$ so dass $(\frac{\|d\|}{\|x_k - x_0\|}(x_k - x_0))$ gegen d konvergiert.

Wenn umgekehrt (x_k) eine Folge in K ist, ist $(\frac{1}{\|x_k - x_0\|}(x_k - x_0))$ eine beschränkte Folge in $\mathbb{R}^n$, und jeder Häufungspunkt ist d ein tangentialer Vektor.

Proposition 5.8 *Es seien $K \subseteq \mathbb{R}^n$ und $x_0 \in K$. Dann ist $\mathcal{T}(x_0)$ ein abgeschlossener Kegel.*

(Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**, wenn $\alpha x \in K$ für alle $\alpha \in [0, \infty)$ und $x \in K$ gilt.)

Beweis Offenbar ist $\mathcal{T}(x_0)$ ein Kegel. Sei (d_k) eine Folge in $\mathcal{T}(x_0)$, die gegen ein $d \in \mathbb{R}^n$ konvergiert. OBdA gelte $\|d_k - d\| < 1/k$ für alle k. Wegen $d \in \mathcal{T}(x_0)$ gibt es ein $x_k \in K$ und ein $\alpha_k \geq 0$ so dass gilt $\|x_k - x_0\| < 1/k$ und $\|\alpha_k(x_k - x_0) - d_k\| < 1/k$. Es folgt $\|\alpha_k(x_k - x_0) - d\| < \frac{2}{k}$ und daraus $d \in \mathcal{T}(x_0)$. ■

Das folgende Ergebnis zeigt die Bedeutung des Tangentialkegels:

Proposition 5.9 *Vorgelegt sei das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in K \end{array}$$

Wenn $x^* \in K$ eine lokale Lösung des MPs ist, dann gilt

$$\nabla f(x^*)^t d \geq 0 \qquad \textit{für alle } d \in \mathcal{T}(x^*)$$

Beweis Es sei $d \in \mathcal{T}(x^*)$, dann gibt es eine Folge (x_k) in K, die gegen x_0 konvergiert, so dass gilt

$$\frac{\|d\|}{\|x_k - x^*\|}(x_k - x^*) \to d$$

Nun gilt für alle i :

$$0 \leq f(x_k) - f(x^*) = \nabla f(x^*)^t (x_k - x^*) + R(x_k)$$

mit $\lim_{x \to x^*} \frac{R(x)}{\|x - x^*\|} = 0$. Es folgt:

$$0 \leq \nabla f(x^*)^t \frac{\|d\|}{\|x_k - x^*\|}(x_k - x^*) + \frac{\|d\|}{\|x_k - x^*\|} R(x_i)$$

und daraus die Behauptung. ■

5.9 besagt also, dass es in $\mathcal{T}(x^*)$ keine Abstiegsrichtungen von f gibt, wenn x^* das MP löst.

Proposition 5.10 *Es seien $g : \mathbb{R}^n \to \mathbb{R}^p$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ affine Abbildungen, dann gibt es $a_1, \ldots, a_p, b_1, \ldots, b_q \in \mathbb{R}^n$ und $c_i, d_j \in \mathbb{R}$ so dass für $1 \leq i \leq p$ und $1 \leq j \leq q$ gilt:*

$$g_i(x) = a_i^t x + c_i \quad , \quad h_j(x) = b_j^t x + d_j$$

Dann gilt für alle $x_0 \in K(g,h)$:

$$\mathcal{T}(x_0) = \{d \in \mathbb{R}^n : a_i^t d \leq 0 \textit{ für alle } i \in I(x_0),\ b_j^t d = 0 \textit{ für alle } j\}$$

Beweis Es sei $d \in \mathcal{T}(x_0)$, dann gibt es Folgen (x_k) in K und (α_k) in $[0, \infty)$ so dass gilt

$$x_k \to x_0 \quad \text{und} \quad \alpha_k(x_k - x_0) \to d$$

Für alle i gilt dann

$$\alpha_k a_i^t (x_k - x_0) \to a_i^t d$$

Weiterhin gilt für alle $i \in I(x_0)$:

$$a_i^t(x_k - x_0) = (a_i^t x_k + c_i) - (a_i^t x_0 + c_i) = g_i(x_k) - g_i(x_0) = g_i(x_k) \leq 0$$

Es folgt für alle $i \in I(x_0)$:

$$a_i^t(x_k - x_0) \leq 0$$

und daraus $a_i^t d \leq 0$. Analog zeigt man, dass $b_j^t d = 0$ für alle j gilt.

Nun gelte $a_i^t d \leq 0$ für alle $i \in I(x_0)$ und $b_j^t d = 0$ für alle j für ein $d \in \mathbb{R}^n$. Für alle $\varepsilon > 0$ und alle $i \in I(x_0)$ erhält man:

$$g_i(x_0 + \varepsilon d) = a_i^t(x_0 + \varepsilon d) + c_i = a_i^t x_0 + c_i + \varepsilon a_i^t d = g(x_0) + \varepsilon a_i^t d \leq 0$$

und analog $h_j(x_0 + \varepsilon d) = 0$ für alle j. Nun sei $i \notin I(x_0)$, dann gilt $g_i(x_0) < 0$. Da g_i stetig ist, gibt es ein $\varepsilon_i > 0$ so dass gilt $g_i(x_0) \leq 0$ für alle $x \in B(x_0, \varepsilon_i)$. Also gibt es ein $\varepsilon_0 > 0$ so dass gilt

$$g_i(x) \leq 0 \qquad \text{für alle } i \notin I(x_0) \text{ und alle } x \in B(x_0, \varepsilon_0)$$

Also gilt $x_0 + \frac{1}{k} d \in K$ für alle $k \geq k_0$ und es folgt die Behauptung. ■

Die Berechnung von $\mathcal{T}_{K(g,h)}(x_0)$ ist oft sehr schwierig. Wenn nun g und h differenzierbar sind, betrachtet man den Tangentialkegel, den man erhält, wenn man g und h durch die "affinen Approximationen" ersetzt. Diese sind definiert durch:

$$\varphi_g(x) = g(x_0) + \nabla g(x_0)^t (x - x_0)$$

und

$$\varphi_h(x) = h(x_0) + \nabla h(x_0)^t (x - x_0)$$

Die Tangentialkegel dieser Abbildungen habe ich gerade berechnet und man kommt zu der folgenden

Definition 5.11 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $g : \mathcal{D} \to \mathbb{R}^p$ und $h : \mathcal{D} \to \mathbb{R}^q$ differenzierbare Abbildungen, dann definiert man für alle $x_0 \in K(g,h)$:*

$$\mathcal{Z}_{(g,h)}(x_0) = \{d \in \mathbb{R}^n : \nabla g_i(x_0)^t d \leq 0 \text{ für alle } i \in I(x_0),\ \nabla h_j(x_0)^t d = 0 \text{ für alle } j\}$$

In der Regel schreibt man $\mathcal{Z}(x_0)$ für $\mathcal{Z}_{(g,h)}(x_0)$.

Man nennt $\mathcal{Z}_{(g,h)}(x_0)$ auch den **linearisierten** oder **linearisierenden** Tangentialkegel, eine Bezeichnung, die nicht sehr glücklich ist, weil ja nicht der Kegel, sondern die Abbildungen linearisiert worden sind. Ich werde in 5.13 zeigen, dass $\mathcal{Z}_{(g,h)}(x_0)$ in der Tat nicht nur von der Menge $K(g,h)$, sondern von g und h abhängt.

Proposition 5.12 *Es seien $g : \mathcal{D} \to \mathbb{R}^p$ und $h : \mathcal{D} \to \mathbb{R}^q$ differenzierbare Abbildungen, dann gilt für alle $x_0 \in K$:*

$$\mathcal{T}(x_0) \subseteq \mathcal{Z}(x_0)$$

oder, genauer $\mathcal{T}_{K_{(g,h)}}(x_0) \subseteq \mathcal{Z}_{(g,h)}(x_0)$.

Beweis Es sei $d \in \mathcal{T}(x_0),\ d \neq 0$, dann gibt es eine Folge (x_k) in K, die gegen x_0 konvergiert, so dass $(\frac{\|d\|}{\|x_k - x_0\|}(x_k - x_0))$ gegen d konvergiert. Sei $i \in I(x_0)$. Da g_i in x_0 differenzierbar ist, gilt für alle i und alle $x \in \mathcal{D}$:

$$g_i(x) = g_i(x_0) + \nabla g(x_0)^t(x - x_0) + R(x) = \nabla g(x_0)^t(x - x_0) + R(x)$$

mit

$$\lim_{x \to x_0} \frac{1}{\|x - x_0\|} R(x) = 0$$

Es folgt für alle i:

$$0 \geq g_i(x) = \nabla g(x_0)^t(x_k - x_0) + R(x_k)$$

und daraus

$$0 \geq \nabla g(x_0)^t \frac{\|d\|}{\|x_k - x_0\|}(x_k - x_0) + \frac{\|d\|}{\|x_k - x_0\|} R(x_k)$$

und schließlich

$$0 \geq \nabla g(x_0)^t d$$

Den 2. Teil beweist man analog. ■

Beispiel 5.13

(i) Das folgende Beispiel zeigt, dass i.a. $\mathcal{T}(x) \neq \mathcal{Z}(x)$ gilt:

Man definiere wieder $g : \mathbb{R}^2 \to \mathbb{R}^3$ durch

$$g(u, v) = (v - (1 - u)^3, -u, -v)^t$$

Dann gilt $\mathcal{T}(e_1) = \{(u, 0)^t : u \leq 0\}$ nach 5.6.

Es gilt

$$g_1(e_1) = 0\ ,\ g_2(e_1) = -1\ ,\ g_3(e_1) = 0$$

also $I(e_1) = \{1, 3\}$. Weiterhin gilt für alle (u, v):

$$\nabla g_1(u, v) = (3(1 - u)^2, 1)^t \quad , \quad \nabla g_3(u, v) = (0, -1)^t$$

und daher

$$\nabla g_1(e_1) = e_2 \quad , \quad \nabla g_3(e_1) = -e_2$$

also

$$\begin{aligned}
\mathcal{Z}(e_1) &= \{(u, v)^t : \nabla g_1(e_1)^t(u, v)^t \leq 0\ ,\ \nabla g_3(e_1)^t(u, v)^t \leq 0\} \\
&= \{(u, v)^t : e_2^t(u, v)^t \leq 0\ ,\ -e_2^t(u, v)^t \leq 0\} \\
&= \{(u, v) : v = 0\} \\
&= \{(u, 0) : u \in \mathbb{R}\} \\
&\neq \mathcal{T}(e_1)
\end{aligned}$$

(ii) Man definiere nun $\tilde{g} : \mathbb{R}^2 \to \mathbb{R}^4$ durch

$$\tilde{g}(u,v) = (g_1(u,v), g_2(u,v), g_3(u,v), u-1)^t$$

dann gilt

$$\{(u,v) : \tilde{g}(u,v) \leq 0\} = \{(u,v) : g(u,v) \leq 0\}$$

Weiterhin gilt $\tilde{g}_4(e_1) = 0$ und daher $4 \in I(e_1)$ sowie $\nabla g_4(u,v) = e_1$. Also folgt mit (i):

$$\mathcal{Z}(e_1) = \{(u,v)^t : v \leq 0\ ,\ -v \leq 0\ ,\ (u,v)^t e_1^t \leq 0\} = \mathcal{T}(e_1)$$

also hängt $\mathcal{Z}_{(g,h)}(x_0)$ nicht nur von $K(g,h)$ ab, sondern auch von g und h selbst. Für die linearisierte Form von g_1 gilt:

$$\varphi_{g_1}(u,v) = g_1(e_1) + \nabla g_1(e_1)^t (u-1,v)^t = e_2^t (u-1,v)^t = v$$

also gilt für die linearisierte Form von g:

$$\varphi_g(u,v) = (v, -u, -v)$$

und es folgt:

$$\varphi_g(u,v) \leq 0 \iff u \geq 0,\ v = 0$$

Damit ist bei der Linearisierung die Bedingung $u \leq 1$ weggefallen. Dies hat zur Folge, dass $\mathcal{Z}(e_1)$ im ersten Fall größer ist als $\mathcal{T}(e_1)$. Da diese Bedingung im zweiten Fall explizit aufgenommen wird, erhält man hier die Gleichheit.

Der folgende Satz wird zeigen, dass die Bedingung $\mathcal{T}(x_0) = \mathcal{Z}(x_0)$ weitreichende Konsequenzen hat. Daher hat diese Bedingung einen eigenen Namen:

Definition 5.14 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $g : \mathcal{D} \to \mathbb{R}^p$ sowie $h : K \to \mathbb{R}^q$ differenzierbare Abbildungen. Schließlich sei $f : K(g,h) \to \mathbb{R}$ eine Abbildung. Ein zulässiger Punkt x_0 des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

*genügt der **Regularitätsbedingung von Abadie** (engl.: Abadie constraint qualification, Abadie CQ), wenn gilt*

$$\mathcal{T}_{K(g,h)}(x_0) = \mathcal{Z}_{(g,h)}(x_0)$$

Damit bin ich in der Lage, einen der fundamentalen Sätze der Theorie zu formulieren und beweisen:

Satz 5.15 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^p, h : \mathcal{D} \to \mathbb{R}^q$ differenzierbare Abbildungen. Es sei x^* eine Lösung des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Wenn x^* der Regularitätsbedingung von Abadie genügt, gibt es $\lambda_1, \ldots, \lambda_p \geq 0$ und $\mu_1, \ldots \mu_q \in \mathbb{R}$ so dass gelten:

$$(1) \quad \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0$$
$$(2) \quad \lambda_i g_i(x^*) = 0 \qquad \textit{für alle } i \ .$$

Beweis OBdA sei $I(x^*) = \{1, \ldots, r\}$. Ich will das Lemma von Farkas (3.27) auf

$$\nabla g_1(x^*), \ldots, \nabla g_r(x^*), \nabla h_1(x^*), \ldots, \nabla h_q(x^*), -\nabla h_1(x^*), \ldots, -\nabla h_q(x^*)$$

und $-\nabla f(x^*)$ anwenden: Es gelte für ein $x \in \mathbb{R}^n$

$$\begin{array}{rl} \nabla g_i(x^*)^t x \leq 0 & i = 1, \ldots, r \\ \nabla h_j(x^*)^t x \leq 0 & j = 1, \ldots, q \\ -\nabla h_j(x^*)^t x \leq 0 & j = 1, \ldots, q \end{array}$$

dann gilt $x \in \mathcal{Z}(x^*)$. Aus der Voraussetzung folgt dann $x \in \mathcal{T}(x^*)$ und daraus $\nabla f(x^*)^t x \geq 0$ nach 5.9, also $-\nabla f(x^*)^t x \leq 0$. Nach dem Lemma von Farkas (3.27) gibt es Zahlen $\lambda_1, \ldots, \lambda_r, \sigma_1, \ldots \sigma_q, \tau_1, \ldots, \tau_q \geq 0$ so dass gilt:

$$-\nabla f(x^*) = \sum_{i=1}^{r} \lambda_i \nabla g_i(x^*) + \sum \sigma_j \nabla h_j(x^*) + \sum \tau_j \nabla h_j(x^*)$$

Setzt man nun $\mu_j = \sigma_j - \tau_j$ für alle j und $\lambda_i = 0$ für alle $i \notin I(x^*)$, dann folgt die Behauptung. ■

Die Bedingungen (1) und (2) heißen **Kuhn-Tucker-Bedingungen** (oder kurz KT-Bedingungen) und ein zulässiger Punkt x^*, der ihnen genügt, heißt auch **Kuhn-Tucker-Punkt** (KTP).

Man nennt die (λ_i) und (μ_j) aus 5.15 wieder **Langrange-Multiplikatoren**. Für jede Ungleichung $g_i(x) \leq 0$ ist also der zugehörige Lagrange-Multiplikator ≥ 0 und wenn er $\neq 0$ ist, ist $g_i(x^*) = 0$.

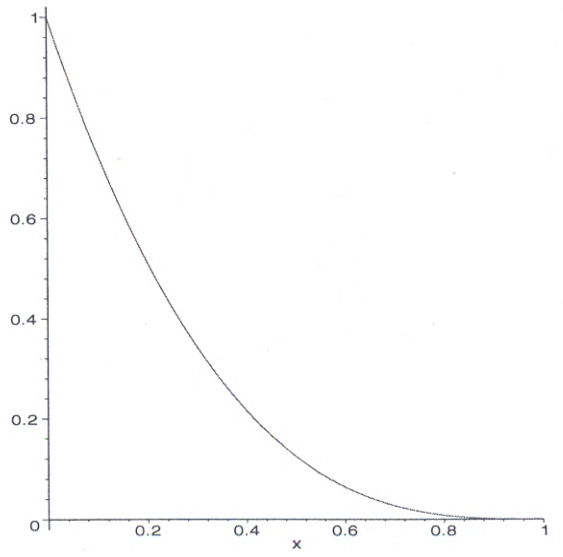
Beispiel 5.16 Man definiere $f : \mathbb{R}^2 \to \mathbb{R}$ und $g : \mathbb{R}^2 \to \mathbb{R}^3$ für alle $(u, v) \in \mathbb{R}^2$ durch:

$$f(u, v) = -u \ , \ g_1(u, v) = v - (1 - u)^3 \ , \ g_2(u, v) = -u \ , \ g_3(u, v) = -v$$

und betrachte das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & g(x) \leq 0 \end{array}$$

Der zulässige Bereich hat ja die Form:



und offenbar ist $x^* = (1,0)^t$ die Lösung des MPs.

Es gilt für alle $(u,v)^t \in \mathbb{R}^2$:

$$\nabla g_1(u,v) = (3(1-u^2),1)^t \ , \ \nabla g_2(u,v) = (-1,0)^t \ , \ \nabla g_3(u,v) = (0,-1)^t$$

Für $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ gelte

$$\nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) + \lambda_2 \nabla g_3(x^*) = 0$$

dann folgt

$$\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

und daraus

$$\begin{aligned} -1 - \lambda_2 &= 0 \\ \lambda_1 - \lambda_3 &= 0 \end{aligned}$$

also folgt $\lambda_2 = -1 < 0$. Man kann daher im allgemeinen nicht erwarten, dass alle Lagrange-Multiplikatoren nicht-negativ sind.

Definiert man zusätzlich $g_4 : \mathbb{R} \to \mathbb{R}$ durch $g_4(u,v) = u - 1$, ist das MP dasselbe, wird also von x^* gelöst. Weiterhin gilt $\nabla g_4(x^*) = (1,0)^t$ und es folgt

$$\nabla f(x^*) + 1 \cdot \nabla g_4(x^*) = 0$$

so dass es in der Tat $\lambda_i \geq 0$ gibt mit $\nabla f(x^*) + \sum \lambda_i \nabla g_i(x^*) = 0$.

Da der Satz 5.15 von fundamentaler Bedeutung ist, ist die Frage, wann die Regularitätsbedingung von Abadie erfüllt ist, von großer Bedeutung. Vor allen Dingen ist es sehr wichtig, einfache hinreichende Bedingungen dafür zu finden. Davon gibt es ziemlich viele, ich gebe einige davon an, wobei die letzte wohl die wichtigste ist. Da diese oft leicht verifizierbar ist, warte ich mit Beispielen, bis ich diese bewiesen habe.

Bemerkung 5.17 Es sei $I \subseteq \mathbb{R}$ ein (nicht notwendigerweise offenes) Intervall. Eine Abbildung $f : I \to \mathbb{R}^n$ heißt **differenzierbar** in $x_0 \in I$, wenn alle Komponentenabbildungen in x_0 differenzierbar sind. Man setzt in diesem Fall

$$f'(x_0) = (f_1'(x_0), \ldots, f_n'(x_0))^t$$

Offenbar fällt diese Definition mit der üblichen Definition zusammen, wenn I offen ist und es gilt in diesem Fall

$$f'(x_0) = Df(x_0)$$

Weiterhin gilt natürlich

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Lemma 5.18 *Es seien $\varepsilon > 0$ und $\chi : [0, \varepsilon) \to K$ eine differenzierbare Abbildung. Dann gilt $\chi'(0) \in \mathcal{T}(\chi(0))$.*

Beweis Es sei (α_k) eine Nullfolge in $(0, \varepsilon)$ Für alle $k \in \mathbb{N}$ setze man $x_k = \chi(\alpha_k)$, dann gilt

$$\frac{1}{\alpha_k}(x_k - \chi(0)) = \frac{\chi(\alpha_k) - \chi(0)}{\alpha_k} \longrightarrow \chi'(0)$$ ■

Proposition 5.19 *Es sei $x_0 \in K(g, h)$ und es gebe zu jedem $d \in \mathcal{Z}(x_0)$ eine differenzierbare Abbildung $\chi : [0, \varepsilon) \to K(g, h)$ so dass gelten $\chi(0) = x_0$ und $\chi'(0) = d$. Dann gilt $\mathcal{T}(x_0) = \mathcal{Z}(x_0)$.*

Beweis 5.12 und 5.18. ■

Man nennt die Bedingung aus 5.19 auch die **Kuhn-Tucker-Restriktionsqualifikation** (Kuhn-Tucker-constraint-qualification (KTCQ)). Diese Bedingung ist offenbar auch nicht besonders handlich, daher sucht man Bedingungen, die die Existenz der Abbildung χ garantieren. Hier geht nun der Satz über implizite Funktionen in seiner allgemeinen Form ein, den ich daher zunächst noch einmal formuliere:

Satz über implizite Funktionen Es seien $W \subseteq \mathbb{R}^p \times \mathbb{R}^q$ offen, $F : W \to \mathbb{R}^q$ eine stetig differenzierbare Abbildung und $(x_0, y_0) \in W$. Weiterhin gelte $F(x_0, y_0) = 0$ und $D_y F(x_0, y_0)$ sei regulär. Dann gibt es offene Mengen $U \subseteq \mathbb{R}^p$ und $V \subseteq \mathbb{R}^q$ mit $(x_0, y_0) \in U \times V \subseteq W$ und eine stetig differenzierbare Abbildung $g : U \to \mathbb{R}^q$ mit $g(U) \subseteq V$ so dass gelten

(i) $F(x, g(x)) = 0$ für alle $x \in U$

(ii) $g(x_0) = y_0$

(iii) $Dg(x) = -D_y F(x, g(x))^{-1} D_x F(x, g(x))$ für alle $x \in U$

Proposition 5.20 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $h : \mathcal{D} \to \mathbb{R}^q$ eine stetig differenzierbare Abbildung sowie $x_0 \in \mathcal{D}$ und $d \in \mathbb{R}^n$. Es gelte:*

(i) $h(x_0) = 0$.

(ii) Die Vektoren $\nabla h_1(x_0), \ldots, \nabla h_q(x_0)$ sind linear unabhängig.

(iii) $\nabla h_j(x_0)^t d = 0$ *für alle j.*

Dann gibt es ein $\varepsilon > 0$ und eine stetig differenzierbare Abbidung $\chi : [0, \varepsilon) \to \mathbb{R}^n$ so dass gelten:

(i) $h(\chi(t)) = 0$ *für alle t*

(ii) $\chi(0) = x_0,\ \chi'(0) = d$.

Beweis Die Matrix

$$A = (\nabla h_1(x_0), \ldots, \nabla h_q(x_0))$$

hat maximalen Rang, man wähle eine Matrix $B = (b_1, \ldots, b_m)$ so dass (A, B) regulär ist. Weiterhin definiere man $F : \mathbb{R} \times \mathcal{D} \to \mathbb{R}^n$ durch

$$F(\alpha, x) = \begin{pmatrix} h(x) \\ B^t(x - x_0) \end{pmatrix} - \alpha \begin{pmatrix} A^t \\ B^t \end{pmatrix} d$$

Dann ist F stetig differenzierbar. Weiterhin gilt $F(0, x_0) = 0$ und

$$D_x F(0, x_0) = \begin{pmatrix} A^t \\ B^t \end{pmatrix}$$

sowie

$$D_\alpha F(0, x_0) = -\begin{pmatrix} A^t \\ B^t \end{pmatrix} d$$

also ist $D_x F(0, x_0)$ regulär. Nach dem Satz über implizite Funktionen gibt es ein $\varepsilon > 0$ und eine stetig differenzierbare Abbildung $\chi : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ so dass für alle $\alpha \in (-\varepsilon, \varepsilon)$ gelten $(\alpha, \chi(\alpha)) \in \mathcal{D},\ F(\alpha, \chi(\alpha)) = 0,\ \chi(0) = x_0$ und

$$D\chi(\alpha) = -(D_x F(\alpha, \chi(\alpha)))^{-1} D_\alpha F(\alpha, \chi(\alpha))$$

also

$$\chi'(0) = D\chi(0) = -(D_x F(0, \chi(0)))^{-1} D_\alpha F(0, \chi(0)) = d$$

Aus $\nabla h_j(x_0)^t d = 0$ für alle j folgt, dass $A^t d = 0$ gilt. Man erhält für alle t:

$$0 = F(\alpha, \chi(\alpha)) = \begin{pmatrix} h(\chi(\alpha)) \\ B^t(\chi(\alpha) - x_0) \end{pmatrix} - \alpha \begin{pmatrix} A^t \\ B^t \end{pmatrix} d = \begin{pmatrix} h(\chi(\alpha)) \\ B^t(\chi(\alpha) - x_0) \end{pmatrix} - \begin{pmatrix} 0 \\ \alpha B^t d \end{pmatrix}$$

und daraus

$$h(\chi(\alpha)) = 0 \qquad \text{für alle } \alpha \qquad \blacksquare$$

Proposition 5.21 *Es sei x_0 ein zulässiger Punkt des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \le 0 \\ & h(x) = 0 \end{array}$$

Es gelten:

(i) Die Vektoren $\nabla h_1(x_0), \ldots, \nabla h_q(x_0)$ sind linear unabhängig.

(ii) Es gibt einen Vektor $d_0 \in \mathbb{R}^n$ so dass gelten

$$\nabla g_i(x_0)^t d_0 < 0,\ i \in I(x_0) \qquad \textit{und} \qquad \nabla h_j(x_0)^t d_0 = 0,\ j = 1, \ldots, q\ .$$

Dann gilt $\mathcal{T}(x_0) = \mathcal{Z}(x_0)$.

Beweis Es sei $d \in \mathcal{Z}(x_0)$, dann gilt $\nabla h_i(x_0)^t d = 0$ für alle i. Man wähle $k \in \mathbb{N}$ fest und setze $d_k = d + \frac{1}{k} d_0$. Dann gilt $\nabla h_j(x_0)^t d_k = 0$ für alle j. Nach 5.20 gibt es ein $\varepsilon > 0$ und eine stetig differenzierbare Abbildung $\chi : [0, \varepsilon) \to \mathbb{R}^n$ so dass gelten $h(\chi(\alpha)) = 0,\ \chi(0) = x_0$ und $\chi'(0) = d_k$. Weiterhin gilt $g_i(\chi(0)) = g_i(x_0) \le 0$ und

$$(g_i \circ \chi)'(\alpha) = \nabla g_i(\chi(\alpha))^t \chi'(\alpha) \qquad \text{für alle } \alpha$$

Es folgt für alle $i \in I(x_0)$:

$$(g_i \circ \chi)'(0) = \nabla g_i(x_0)^t d_k = \nabla g_i(x_0)^t d + \frac{1}{k} \nabla g_i(x_0)^t d_0 < 0$$

Also gibt es ein $\varepsilon_i > 0,\ \varepsilon_i \leq \varepsilon$ so dass für alle $i \in I(x_0)$ gilt

$$g_i(\chi(\alpha)) \leq 0 \qquad \text{für alle } 0 \leq \alpha \leq \varepsilon_i$$

Für alle $i \notin I(x_0)$ gilt $g_i(x_0) < 0$ und daher gibt es ein $\varepsilon_i > 0$ so dass gilt

$$g(\chi(\alpha)) \leq 0 \qquad \text{für alle } 0 \leq \alpha \leq \varepsilon_i$$

Sei $\varepsilon_0 = \min \varepsilon_i$, dann gilt $\chi(\alpha) \in K(g,h)$ für alle $0 \leq \alpha \leq \varepsilon_0$, Nach 5.18 gilt $d_k \in \mathcal{T}(x_0)$ für alle k. Da (d_k) gegen d konvergiert, folgt $d \in \mathcal{T}(x_0)$ nach 5.8. ■

Man nennt die Bedingung aus 5.21 die **Mangasarian-Fromovitz-Bedingung** (**Mangasarian-Fromovitz-constraint-qualification**, MFCQ). Die am häufigsten benutzte Bedingung hat ebenfalls einen Namen, den ich explizit definieren will:

Definition 5.22 *Vorgegeben sei das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

*Man sagt, ein zulässiger Punkt $x_0 \in \mathcal{D}$ genügt der **Regularitätsbedingung der linearen Unabhängigkeit** (linear independence constraint qualification, LICQ), wenn die Vektoren*

$$(\nabla g_i(x_0))_{i \in I(x_0)}, (\nabla h_i(x_0))_{i=1,\ldots,q}$$

*linear unabhängig sind. Ich werde so einen Punkt kurz **regulär** bezüglich des MPs nennen.*

Proposition 5.23 *Es sei x_0 ein zulässiger Punkt des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Die Vektoren

$$\nabla g_i(x_0) : i \in I(x_0), \qquad \nabla h_j(x_0) : j = 1, \ldots, q$$

seien linear unabhäng. Dann gibt es eine Vektor $d_0 \in \mathbb{R}^n$ so dass gilt:

$$\nabla g_i(x_0)^t d_0 < 0,\ i \in I(x_0) \qquad \textit{und} \qquad \nabla h_i(x_0)^t d_0 = 0,\ i = 1, \ldots, q$$

Beweis Da die Vektoren linear unabhängig sind, gibt es eine lineare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}$ so dass gelten

$$\varphi(\nabla g_i(x_0)) = -1,\ i \in I(x_0) \qquad \text{und} \qquad \varphi(\nabla h_i(x_0)) = 0,\ i = 1, \ldots, q$$

Man wähle d_0 so dass gilt $d_0^t x = \varphi(x)$ für alle $x \in \mathbb{R}^n$. ■

Also gilt

$$\text{LICQ} \implies \text{MFCQ} \implies \text{KTCQ} \implies \text{Abadie CQ} \iff \mathcal{T}(x_0) = \mathcal{Z}(x_0)$$

Aus 5.15 folgt dann unmittelbar:

Satz 5.24 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $f : \mathcal{D} \to \mathbb{R}, g : \mathcal{D} \to \mathbb{R}^p, h : \mathcal{D} \to \mathbb{R}^q$ stetig differenzierbare Abbildungen. Dann ist jede reguläre Lösung des MPs*

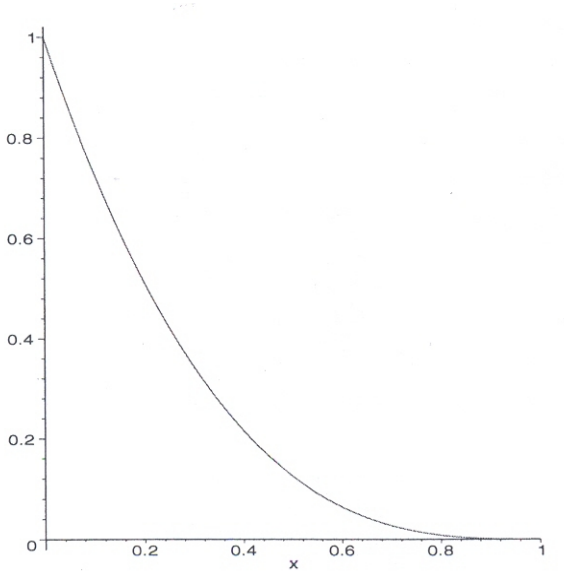
$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

ein KTP.

Beispiel 5.25 Vorgegeben sei das Maximierungsproblem

$$\begin{array}{ll} \max & u + v \\ \text{bez.} & v \leq (1-u)^3 \\ & u, v \geq 0 \end{array}$$

Das folgende Bild zeigt den zulässigen Bereich:



Also wird das Maximimum offenbar in $e_1 = (1, 0)^t$ und $e_2 = (0, 1)^t$ angenommen.

Definiert man $f : \mathbb{R}^2 \to \mathbb{R}$ und $g : \mathbb{R}^2 \to \mathbb{R}^3$ durch

$$f(u, v) = -u - v$$

und

$$\begin{aligned} g_1(u, v) &= v - (1 - u)^3 \\ g_2(u, v) &= -u \\ g_3(u, v) &= -v \end{aligned}$$

dann reicht es offenbar, das folgende MP zu lösen:

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & g(x) \leq 0 \end{array}$$

(1) **Lösbarkeit des MPs** Wie man am Bild sieht und ohne Probleme nachrechnet, ist der zulässige Bereich kompakt. Da f stetig ist, ist das Problem lösbar.

(2) **Suche der nicht regulären Punkte** Es sei $x_0 = (u, v) \in \mathbb{R}^2$, dann gilt:

$$\begin{aligned} \nabla f(u, v) &= (-1, -1)^t \\ \nabla g_1(u, v) &= (3(1 - u)^2, 1)^t \\ \nabla g_2(u, v) &= (-1, 0)^t \\ \nabla g_3(u, v) &= (0, -1)^t \end{aligned}$$

Die Antwort auf die Frage, ob x_0 regulär ist, wird wesentlich dadurch erschwert, dass $I := I(x_0)$ nicht bekannt ist. Daher muss man alle Möglichkeiten in Betracht ziehen.
(a) $I = \{1, 2, 3\}$, dann folgt $g_1(x_0) = g_2(x_0) = g_3(x_0) = 0$ und daraus $u = v = 0$ sowie $0 - (1 - 0)^3 = 0$, was offenbar nicht möglich ist.

(b) $I = \{2, 3\}$. Da $\nabla g_2(u, v)$ und $\nabla g_3(u, v)$ linear unabhängig sind, ist x_0 in diesem Fall regulär.

(c) $I = \{1, 3\}$, dann gilt $v = 0$ und daher $u = 1$. Wegen

$$\nabla g_1(1, 0) = (0, 1)^t = -\nabla g_3(1, 0)$$

ist e_1 nicht regulär. (Das folgt übrigens auch schon aus 5.13.)

(d) $I = \{1, 2\}$. Die Vektoren

$$\nabla g_1(x_0), \nabla g_2(x_0) = \begin{pmatrix} 3(1 - u)^2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

sind offenbar linear unabhängig.

(e) Da alle drei Gradienten $\neq 0$ sind, ist x_0 regulär, wenn I nur aus einem Punkt besteht. Wenn $I = \emptyset$ gilt, ist x_0 offenbar regulär.

Also ist e_1 der einzige nicht reguläre Punkt des MPs.

Bestimmung der Lösungen Es sei $x^* = (u, v)^t$ eine Lösung des MPs. Wenn $x^* \neq e_1$ gilt, ist x^* ein KTP des MPs, also gibt es $\lambda_1, \lambda_2, \lambda_3 \geq 0$ so dass gelten

(i) $\nabla f(u, v) + \sum_{i=1}^{3} \lambda_i \nabla g_i(u, v) = 0$

(ii) $\lambda_i g_i(u, v) = 0 \qquad$ für $i = 1, 2, 3$

(b) $I = \{2, 3\}$, dann gilt $\lambda_1 = 0$ und es folgt:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Man erhält $\lambda_2 = -1$, also einen Widerspruch.

(c) $I = \{1, 3\}$, dann gilt $x^* = e_1$ und dieser Fall war ausgeschlossen.

(d) $I = \{1, 2\}$. Dann gilt $x^* = e_2$ und $\lambda_3 = 0$. Es folgt

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Diese Gleichung ist erfüllt für $(\lambda_1, \lambda_2) = (1, 2)$.

(e1) $I = \{1\}$, dann gilt $\lambda_2 = \lambda_3 = 0$ und $v = (1 - u)^3$ sowie

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 3(1-u)^2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Es folgt $\lambda_1 = 1$ und daraus $3(1-u)^2 = 1$, also $u = 1 \pm \frac{1}{\sqrt{3}}$. Da $(1 + \frac{1}{\sqrt{3}}, v)$ für kein v zulässig ist, folgt $u = 1 - \frac{1}{\sqrt{3}}$ und daraus $v = \frac{1}{\sqrt{3}^3}$, also

$$x^* = (1 - \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}^3})^t$$

(e2) $I = \{2\}$, dann folgt $\lambda_1 = \lambda_3 = 0$ und $u = 0$. Man erhält:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Es folgt $\lambda_2 = -1$, also ein Widerspruch.

(e3) $I = \{3\}$, dann folgt $\lambda_1 = \lambda_2 = 0$ und $v = 0$. Man erhält:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

also einen Widerspruch.

(e4) $I = \emptyset$, dann folgt $\lambda_1 = \lambda_2 = \lambda_3 = 0$ und daraus $\nabla f(x^*) = 0$ also einen Widerspruch Also ist x^* einer folgenden Vektoren:

$$e_1, e_2, \ (1 - \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}^3})^t$$

Nun gilt

$$f(e_1) = f(e_2) = -1$$

und

$$f(1 - \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}^3})^t = -1 + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}^3} > -1$$

und damit sind e_1 und e_2 die Lösungen des MPs.

Bei der Formulierung der KT-Bedingungen, insbesondere bei den Bedingungen zweiter Ordnung, leistet die sogenannte Lagrange-Funktion gute Dienste:

Definition 5.26 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen, $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^p$ und $h : \mathcal{D} \to \mathbb{R}^q$ Abbildungen. Dann heißt die Abbildung*

$$\begin{aligned} L \ : \ & \mathcal{D} \times \mathbb{R}^p \times \mathbb{R}^q \longrightarrow \mathbb{R} \qquad \textit{definiert durch} \\ & L(x, \lambda, \mu) = f(x) + \textstyle\sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \mu_j h_j(x) \\ & = f(x) + \lambda^t g(x) + \mu^t h(x) \end{aligned}$$

die zu dem MP

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

assoziierte ***Lagrange-Abbildung****.*

Bemerkung 5.27 *Es seien f, g und h partiell differenzierbar. Dann gelten für alle $(x, \lambda, \mu) \in \mathcal{D} \times \mathbb{R}^p \times \mathbb{R}^q$:*

(i) $\nabla_x L(x, \lambda, \mu) = \nabla f(x) + \sum \lambda_i \nabla g_i(x) + \sum \mu_j \nabla h_j(x)$

(ii) $\nabla_\lambda L(x, \lambda, \mu) = g(x)$

(iii) $\nabla_\mu L(x, \lambda, \mu) = h(x)$

Wenn f, g und h zweimal partiell differenzierbar sind, gilt darüber hinaus:

(iv) $H_x L(x, \lambda, \mu) = Hf(x) + \sum \lambda_i Hg_i(x) + \sum \mu_j Hh_j(x)$

Korollar 5.28 *Vorgegeben sei das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

(i) Ein Punkt $x^* \in \mathbb{R}^n$ ist genau dann ein KTP des MPs, wenn es $\lambda^* \in \mathbb{R}^p$, $\mu^* \in \mathbb{R}^q$ so gibt, dass gelten:

$$\begin{array}{rcl} \nabla_x L(x^*, \lambda^*, \mu^*) & = & 0 \\ \nabla_\lambda L(x^*, \lambda^*, \mu^*) & \leq & 0 \\ \nabla_\mu L(x^*, \lambda^*, \mu^*) & = & 0 \\ \lambda^* & \geq & 0 \\ (\lambda^*)^t \nabla_\lambda L(x^*, \lambda^*, \mu^*) & = & 0 \end{array}$$

(ii) Wenn x^* der Regularitätsbedingung von Abadie genügt, gibt es $\lambda^* \in \mathbb{R}^p$, $\lambda^* \geq 0$ und $\mu^* \in \mathbb{R}^q$ so dass gelten:

$$\begin{array}{rcl} \nabla_x L(x^*, \lambda^*, \mu^*) & = & 0 \\ (\lambda^*)^t \nabla_\lambda L(x^*, \lambda^*, \mu^*) & = & 0 \end{array}$$

Beweis

(i) Die zweite und dritte Bedingung garantieren, dass x^* zulässig ist. Wenn aber x^* zulässig ist und $\lambda^* \geq 0$ gilt folgt $\lambda_i^* g_i(x^*) \leq 0$ für alle i und daher

$$0 = (\lambda^*)^t \Delta_\lambda L(x^*, \lambda^*, \mu^*) = (\lambda^*)^t g(x^*) = \sum \lambda_i^* g_i(x^*) \Leftrightarrow \lambda_i^* g_i(x^*) = 0 \text{ für alle } i$$

(ii) Das folgt jetzt direkt aus 5.15 und (i). ■

Kapitel 6

Konvexe Optimierung

Konvexe Optimierung beschäftigt sich naturgemäß mit der Optimierung konvexer Abbildungen auf konvexen Mengen. Die Wichtigkeit der konvexen Optimierung wird u.a. durch die Ergebnisse 4.6 und 4.12 unterstrichen: Jede lokale Lösung eines konvexen Optimierungsproblems ist automatisch eine Lösung und ein Punkt löst ein konvexes Optimierungsproblem genau dann, wenn er stationär ist. Nun ist es so, dass der bisher am häufigsten behandelte Typ eines MPs die Form hat:

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

und dieses Problem ist in der Regel selbst dann nicht konvex, wenn f, g, h konvex sind: Während dann $\{x : g(x) \leq 0\}$ in der Tat eine konvexe Menge ist, ist $\{x : h(x) = 0\}$ im allgemeinen keine konvexe Menge, wie das Beispiel $h(x) = x^t x - 1$ zeigt. Also ist $K(g, h)$ in der Regel nicht konvex, wenn g und h konvex sind. Dieses Problem ist nicht so einfach zu umgehen und die übliche Methode ist die, dass man verlangt, dass h affin ist.

Nun kann man konvexe MPe mit und ohne Differenzierbarkeitsvoraussetzungen studieren, wobei man im zweiten Fall oft längere Beweis in Kauf nehmen muss. Ich werde das in dieser Vorlesung nicht tun und daher im Allgemeinen die Differenzierbarkeit der auftretenden Abbildungen verlangen. Zur Vereinfachung der Sprechweise definiere ich:

Definition 6.1 *Es seien $K \subseteq \mathbb{R}^n$ und $f : K \to \mathbb{R}$, $g : K \to \mathbb{R}^p$ und $h : K \to \mathbb{R}^q$ Abbildungen. Das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

*heißt **konvex**, wenn f und g konvex sind und h affin ist. Das (MP) heißt **differenzierbar**, wenn K offen und f, g und h differenzierbar sind.*

Proposition 6.2 *Es seien $K \subseteq \mathbb{R}^n$ konvex und $g : K \to \mathbb{R}^p$ konvex sowie $h : K \to \mathbb{R}^q$ affin. Dann gelten für alle $x, y \in K$ und $\alpha \in \mathbb{R}$:*

(i) $h(\alpha x + (1-\alpha)y) = \alpha h(x) + (1-\alpha)h(y)$

(ii) Für alle j gilt

$$h_j(y) = h_j(x) + \nabla h_j(x)^t(y - x)$$

(iii) $K(g, h)$ ist konvex.

Beweis Alle drei Beweise sind ziemlich einfach: Es gelte $h(x) = Ax + b$ und $h_j(x) = a_j^t x + b_j$, dann folgt:

$$\begin{aligned} h(\alpha x + (1-\alpha)y) &= A(\alpha x + (1-\alpha)y) + b \\ &= \alpha Ax + (1-\alpha)Ay + \alpha b + (1-\alpha)b \\ &= \alpha(Ax + b) + (1-\alpha)(Ay + b) \\ &= \alpha h(x) + (1-\alpha)h(y) \end{aligned}$$

also (i). Weiterhin gilt $\nabla h_j(x) = a_j$ für alle j und alle x und daher

$$h_j(x) + \nabla h_j(x)^t(y - x) = a_j^t x + b_j + a_j^t(y - x) = a_j^t y + b_j = h_j(y)$$

also (ii).

(iii) Für alle $x, y \in K(g, h)$ und $\alpha \in [0, 1]$ gilt

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y) \leq 0$$

sowie nach (i)

$$h(\alpha x + (1-\alpha)y) = \alpha h(x) + (1-\alpha)h(y) = 0$$

insgesamt also $\alpha x + (1-\alpha)y \in K(g, h)$. ■

Im folgenden will ich den Zusammenhang zwischen der Lösung eines differenzierbaren, konvexen MPs und einem KTP dieses MPs studieren. Wenn eine Lösung eines differenzierbaren MPs der Regularitätsbedingung von Abadie genügt, ist sie ein KTP, die Umkehrung ist falsch. Im konvexen, differenzierbaren Fall ist nun jeder KTP eine Lösung und für die recht unhandliche Bedingung von Abadie gibt es eine einfacher zu verifizierende hinreichende Bedingung (die sog. Slater-Bedingung). Der Beweis der ersten Behauptung geht ganz einfach:

Proposition 6.3 *Vorgelegt sei das konvexe, differenzierbare MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Dann ist jeder KTP des MPs eine Lösung des MPs.

Beweis Es sei x^* ein KTP. Nach 4.12 muss gezeigt werden, dass x^* ein stationärer Punkt des MPs ist. Da x^* ein KTP ist, gibt es $\lambda^* \geq 0$ und μ^* so dass gelten:

$$\nabla f(x^*) + \sum \lambda_i^* \nabla g_i(x^*) + \sum \mu_j^* \nabla h_j(x^*) = 0$$

und

$$\lambda_i^* g_i(x^*) = 0$$

und daher

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) + \sum \mu_j^* \nabla h_j(x^*) = 0$$

Nun gilt für alle $i \in I(x^*)$ und $x \in K(g,h)$ nach 4.11:

$$\nabla g_i(x^*)^t (x - x^*) \leq g_i(x) - g_i(x^*) = g_i(x) \leq 0$$

und daher

$$-\lambda_i^* \nabla g_i(x^*)^t (x - x^*) \geq 0$$

Weiter gilt für alle j und $x \in K(g,h)$:

$$\nabla h_j(x^*)^t (x - x^*) = h_j(x) - h_j(x^*) = 0$$

Es folgt für alle $x \in K(g,h)$ nach 6.2:

$$\nabla f(x^*)^t (x - x^*) = - \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*)^t (x - x^*) - \sum \mu_j^* \nabla h_j(x^*)^t (x - x^*) \geq 0$$

■

Satz 6.4 *Es seien $K \subseteq \mathbb{R}^n$ konvex und offen, $f : K \to \mathbb{R}$ und $g : K \to \mathbb{R}^p$ konvex und differenzierbar. Weiterhin sei $h : K \to \mathbb{R}^q$ affin. Vorgelegt sei das konvexe MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Wenn ein Punkt $x^* \in K(g,h)$ der Regularitätsbedingung von Abadie genügt, löst er das MP genau dann, wenn er ein KT-Punkt ist.

Beweis Nach 6.3 ist jeder KT-Punkt eine Lösung, die Umkehrung folgt aus 5.15. ■

Der Nachweis der Tatsache, dass ein Punkt der Regularitätsbedingung von Abadie genügt ist in der Regel aufwendig. Er wird wesentlich vereinfacht, wenn das MP konvex ist. Der zugehörige Begriff wird definiert in:

Definition 6.5 *Man sagt, das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

genügt der ***Slater-Bedingung****, wenn es ein* $\overline{x} \in K(g,h)$ *so gibt, dass gilt*

$$g_i(\overline{x}) < 0 \qquad \textit{für alle nicht-affinen } g_i$$

Proposition 6.6 *Das konvexe, differenzierbare MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

genüge der Slater-Bedingung. Dann genügt jeder Punkt $x \in K(g,h)$ *der Regularitätsbedingung von Abadie, d.h. es gilt* $\mathcal{T}(x) = \mathcal{Z}(x)$.

Beweis Es seien $x_0 \in K(g,h)$ und I_0 die Menge aller i so dass g_i nicht affin ist. Man setze

$$\mathcal{Z}_0 = \{d \in \mathcal{Z}(x_0) : \nabla g_i(x_0)^t d < 0 \qquad \text{für alle } i \in I_0 \cap I(x_0)\}$$

Ich zeige zunächst, dass $\mathcal{Z}_0 \subseteq \mathcal{T}(x_0)$ gilt, sei also $d \in \mathcal{Z}_0(x_0)$.

Für alle $i \in I_0 \cap I(x_0)$ gilt $\nabla g_i(x_0)^t d < 0$ also ist d nach nach 2.3 eine Abstiegsrichtung, und daher gibt es ein $\varepsilon > 0$ so dass für alle $i \in I(x_0) \cap I_0$ gilt:

$$g_i(x_0 + \alpha d) < g_i(x_0) = 0 \qquad \text{für alle } 0 < \alpha \leq \varepsilon$$

Für alle $i \in I_0 \setminus I(x_0)$ gilt $g_i(x_0) < 0$, also gibt es ein $\varepsilon' \leq \varepsilon$ so dass gilt

$$g_i(x_0 + \alpha d) \leq 0 \qquad \text{für alle } 0 < \alpha \leq \varepsilon'$$

Für alle $i \notin I_0$ ist g_i affin, also gilt für alle $\alpha \geq 0$:

$$g_i(x_0 + \alpha d) = g_i(x_0) + \alpha \nabla g_i(x_0)^t d \leq 0$$

und analog gilt für alle j und alle $\alpha \geq 0$:

$$h_j(x_0 + \alpha d) = h_j(x_0) + \alpha \nabla h_j(x_0)^t d = 0$$

Es folgt $x_0 + \frac{1}{k} d \in K(g,h)$ für alle $k \geq k_0$ und daraus

$$d = \lim k((x_0 + \frac{1}{k} d) - x_0) \in \mathcal{T}(x_0)$$

Ich zeige nun dass gilt:

$$\mathcal{Z}(x_0) \subseteq \overline{\mathcal{T}(x_0)}$$

Da $\mathcal{T}(x_0)$ nach 5.8 abgeschlossen ist, folgt daraus die Behauptung.

Sei also $d \in \mathcal{Z}(x_0)$. Nach Voraussetzung gibt es ein $\overline{x} \in K(g,h)$ so dass gilt $g_i(\overline{x}) < 0$ für alle $i \in I_0$. Da g_i konvex ist, folgt aus 4.11 für alle $i \in I(x_0) \cap I_0$ und alle $\alpha > 0$:

$$\nabla g_i(x_0)^t(d + \alpha(\overline{x} - x_0)) = \nabla g_i(x_0)^t d + \alpha \nabla g_i(x_0)^t(\overline{x} - x_0) \leq \alpha(g_i(\overline{x}) - g_i(x_0)) < 0$$

Weiterhin gilt für alle $i \in I(x_0) \setminus I_0$:

$$\nabla g_i(x_0)^t(d + \alpha(\overline{x} - x_0)) = \nabla g_i(x_0)^t d + \alpha \nabla g_i(x_0)^t(\overline{x} - x_0) \leq \alpha(g_i(\overline{x}) - g_i(x_0)) \leq 0$$

und schließlich für alle j:

$$\nabla h_j(x_0)^t(d + \alpha(\overline{x} - x_0)) = \nabla h_j(x_0)^t d + \alpha \nabla h_j(x_0)^t(\overline{x} - x_0) = \alpha(h_j(\overline{x}) - h_j(x_0)) = 0$$

Also folgt $d + \frac{1}{k}(\overline{x} - x_0) \in \mathcal{Z}_0 \subseteq \mathcal{T}(x_0)$ und daraus $d \in \mathcal{T}(x_0)$. ■

Satz 6.7 *Das konvexe, differenzierbare MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

genüge der Slater-Bedingung. Ein Punkt $x^* \in K(g,h)$ löst das MP genau dann, wenn er ein KTP ist.

Beweis Das folgt direkt aus 6.4 und 6.6. ■

Korollar 6.8 *Vorgegeben sei das konvexe, differenzierbare MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Wenn g affin ist, löst ein Punkt $x^* \in K(g,h)$ dieses MP genau dann, wenn er ein KTP des MPs ist.

Beweis Das MP genügt der Slater-Bedingung. ■

Beispiel 6.9 Man definiere $f, g : \mathbb{R} \to \mathbb{R}$ durch $f(x) = x$ und $g(x) = x^2$ für alle $x \in \mathbb{R}$ und betrachte das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & g(x) \leq 0 \end{array}$$

Dann sind f und g konvex und $x^* = 0$ ist die einzige Lösung. Angenommen, x^* ist ein KT-Punkt, dann gibt es ein $\lambda^* \geq 0$ so dass gilt

$$\nabla f(x^*) + \lambda^* \nabla g(x^*) = 0$$

Es folgt

$$0 = 1 + 2\lambda^* x^* = 1$$

also ein Widerspruch. Daher ist x^* kein KT-Punkt. Dies zeigt, dass man die Voraussetzung der Slater-Bedingung in 6.7 nicht ersatzlos streichen kann.

Eine anderes Optimalitätskriterium erhält man durch den Begriff des Kuhn-Tucker-Sattelpunktes:

Definition 6.10 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$, $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^p$ und $h : \mathcal{D} \to \mathbb{R}^q$ Abbildungen. Ein Punkt $(x^*, \lambda^*, \mu^*) \in \mathcal{D} \times \mathbb{R}^p_+ \times \mathbb{R}^q$ heißt* ***Kuhn-Tucker-Sattelpunkt*** *(KTSP) des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

wenn gilt

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*)$$
$$\textit{für alle } x \in \mathcal{D},\ \lambda \in \mathbb{R}^p_+,\ \mu \in \mathbb{R}^q$$

(x^*, λ^*, μ^*) ist also genau dann ein Sattelpunkt, wenn die Abbildung $L(\cdot, \lambda^*, \mu^*)$ in x^* ein Minimum besitzt und die Abbildung $L(x^*, \cdot, \cdot)$ in (λ^*, μ^*) ein Maximum besitzt, wenn man sie auf $\mathbb{R}^p_+ \times \mathbb{R}^q$ einschränkt. Der Vorteil des Begriffs des KTSPs ist die Tatsache, dass er keine Differenzierbarkeitsvoraussetzungen braucht. Daher ist er beim Studium nicht notwendig differenzierbarer (konvexer) Abbildungen von großem Nutzen.

Proposition 6.11 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$, $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^p$ und $h : \mathcal{D} \to \mathbb{R}^q$ Abbildungen. Vorgegeben sei das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

und (x^*, λ^*, μ^*) ein KTSP des MPs. Dann gelten:

(i) x^* ist eine Lösung des MPs und es gilt $(\lambda^*)^t g(x^*) = 0$.

(ii) Es seien $\mathcal{D}$ offen und f, g und h differenzierbar. Dann ist x^* ein KT-Punkt mit Lagrange-Multiplikatoren λ^*, μ^*.

Beweis

(i) Für alle $\lambda \geq 0$ und alle μ gilt:

$$f(x^*) + \sum \lambda_i^* g_i(x^*) + \sum \mu_j^* h_j(x^*) = L(x^*, \lambda^*, \mu^*)$$
$$\leq L(x^*, \lambda, \mu) = f(x^*) + \sum \lambda_i g_i(x^*) + \sum \mu_j h_j(x^*)$$

und daher

$$\sum (\lambda_i - \lambda_i^*) g_i(x^*) + \sum (\mu_j - \mu_j^*) h_j(x^*) \geq 0$$

Da man λ_i beliebig groß und μ_j beliebig wählen kann, folgt daraus $g_i(x^*) \leq 0$ für alle i und $h_j(x^*) = 0$ für alle j, d.h. $x^* \in K(g,h)$. Setzt man $\lambda_i = 0$ für alle i und $\mu_j = \mu_j^*$ für alle j, erhält man:

$$\sum \lambda_i^* g_i(x^*) \geq 0$$

Aus $\sum \lambda_i^* g_i(x^*) \leq 0$ folgt dann

$$(\lambda^*)^t g(x^*) = \sum \lambda_i^* g_i(x^*) = 0$$

Andererseits gilt für alle $x \in K(g,h)$:

$$f(x^*) + \sum \lambda_i^* g_i(x^*) + \sum \mu_j^* h_j(x^*) = L(x^*, \lambda^*, \mu^*)$$
$$\leq L(x, \lambda^*, \mu^*) = f(x) + \sum \lambda_i^* g_i(x) + \sum \mu_j^* h_j(x)$$

Daraus folgt:

$$f(x^*) \leq f(x) + \sum \lambda_i^* g_i(x) \leq f(x)$$

Also löst x^* das MP.

(ii) Nach (i) gilt $x^* \in K(g,h)$ und $(\lambda^*)^t g(x^*) = 0$. Da die Abbildung $L(\cdot, \lambda^*, \mu^*)$ in x^* in Minimum besitzt, folgt $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$. ■

Proposition 6.12 *Es seien $\mathcal{D} \subseteq \mathbb{R}^n$ offen und $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^p$, $h : \mathcal{D} \to \mathbb{R}^q$ Abbildungen. Vorgegeben sei das konvexe, differenzierbare MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Ein Punkt x^* ist genau dann ein KTP des MPs, wenn es $\lambda^* \geq 0$ und μ^* so gibt, dass (x^*, λ^*, μ^*) ein KTSP des MPs ist.

Beweis Die eine Richtung ist gerade 6.11(ii). Sei also x^* ein KTP mit zugehörigen Lagrange-Multiplikatoren λ^*, μ^*. Da die Abbildung $L(\cdot, \lambda^*, \mu^*)$ konvex ist und x^* ein stationärer Punkt ist, besitzt sie in x^* ein Minimum nach 4.12. Also gilt

$$L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*) \qquad \text{für alle } x \in K$$

Andererseits gilt:

$$\begin{aligned} & L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \\ \Leftrightarrow \quad & f(x^*) + \lambda^t g(x^*) + \mu^t h(x^*) \leq f(x^*) + (\lambda^*)^t g(x^*) + (\mu^*)^t h(x^*) \\ \Leftrightarrow \quad & \lambda^t g(x^*) \leq (\lambda^*)^t g(x^*) \\ \Leftrightarrow \quad & \lambda^t g(x^*) \leq 0 \end{aligned}$$

die letzte Aussage ist offenbar richtig. ■

Korollar 6.13 *Vorgegeben sei das konvexe, differenzierbare MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Wenn das MP der Slater-Bedingung genügt, sind für einen Punkt x^* äquivalent:

(i) x^* *löst das MP.*

(ii) x^* *ist ein KTP des MPs.*

(iii) Es gibt λ^*, μ^* *so dass* (x^*, λ^*, μ^*) *ein KTSP des MPs ist.*

Die Äquivalenz von (i) und (iii) in 6.13 gilt in der Tat für jedes konvexe MP, allerdings ist der Beweis erheblich aufwendiger. (Vgl. z.B. das Buch von Blum/Oettli.)

Kapitel 7

Quadratische Minimierungsprobleme

Definition 7.1 *Es seien $A \in M(n,n)$ symmetrisch, $c \in \mathbb{R}^n, B \in M(p,n)$ und $d \in \mathbb{R}^p$. Dann heißt das MP*

$$\begin{array}{ll} \min & x^t A x + c^t x \\ \textit{bez.} & Bx \leq d \end{array}$$

quadratisches MP.

Bemerkung 7.2 Es sei $A_0 \in M(n,n)$ beliebig. Dann ist

$$A = (1/2)(A_0 + A_0^t)$$

symmetrisch und es gilt für alle $x \in \mathbb{R}^n$:

$$x^t A x = x^t A_0 x \ ,$$

also ist die Voraussetzung der Symmetrie von A bei der Definition eines quadratischen MPs keine Einschränkung.

Proposition 7.3 *Es seien $A \in M(n,n)$ symmetrisch, $c \in \mathbb{R}^n$ und $f : \mathbb{R}^n \to \mathbb{R}$ definiert durch*

$$f(x) = x^t A x + c^t x \ ,$$

dann gelten für alle $x, y \in \mathbb{R}^n$:

(i) $\nabla f(x) = 2Ax + c$

(ii) $Hf(x) = 2A$

(iii) $f(y) = f(x) + \nabla f(x)^t (y-x) + \frac{1}{2}(y-x)^t Hf(x)(y-x)$
$= f(x) + \nabla f(x)^t (y-x) + (y-x)^t A (y-x)$

(iv) *f ist genau dann konvex bze. streng konvex, wenn A positiv semidefinit bzw. positiv definit ist.*

Beweis (iii) ist eine etwas längere Rechnung, folgt aber auch aus 2.19(ii), die anderen Teile sind Übungsaufgaben. ■

Das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in K \end{array}$$

hat bekanntlich eine Lösung, wenn f stetig und K nicht-leer und kompakt (also beschränkt und abgeschlossen) ist. Man kann nun keine der beiden letzten Voraussetzungen weglassen, selbst wenn f konvex und nach unten beschränkt ist:

Beispiel 7.4 *Man betrachte die MPe*

$$\begin{array}{ll} \min & x \\ \textit{bez.} & x \in (0,1) \end{array}$$

und

$$\begin{array}{ll} \min & exp(x) \\ \textit{bez.} & x \in \mathbb{R} \end{array}$$

dann sind beide MPe nach unten beschränkt, aber nicht lösbar.

Daher ist es nun etwas überraschend, dass man für quadratische MPe das folgende Resultat hat:

Satz 7.5 *Es seien $A \in M(n,n)$ symmetrisch und $c \in \mathbb{R}^n$. Weiter seien $B \in M(p,n)$ und $d \in \mathbb{R}^p$. Wenn das quadratische MP*

$$\begin{array}{ll} \min & x^t Ax + c^t x \\ \textit{bez.} & Bx \leq d \end{array}$$

einen zulässigen Punkt besitzt und nach unten beschränkt ist (wenn also die Menge $\{x^t Ax + c^t x : Bx \leq d\}$ nach unten beschränkt ist), besitzt das MP eine Lösung.

Beweis Man setze $K = \{x \in \mathbb{R}^n : Bx \leq d\}$ und definiere $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ durch

$$f(x) = x^t Ax + c^t x \; .$$

Dann ist $f(K)$ nicht-leer und nach unten beschränkt, besitzt also ein Infimum.

Beweisidee Für alle $k \in \mathbb{N}$ setze man $K_k = K \cap B(0,k]$ und oBdA gelte $K_1 \neq \emptyset$. Man wähle $x_k \in K_k$ so dass gilt $f(x_k) = \min f(K_k)$. Dann konvergiert $(f(x_k))$ gegen $\inf f(K)$. Wenn (x_k) einen Häufungspunkt x^* besitzt, ist x^* eine Lösung

des MPs. Aber ist es möglich, dass die Folge keinen Häufungspunkt besitzt: Definiert man z.B. $f(u,v) = u^2$ und $B = 0\ ,\ d = 0$, dann kann man $x_k = (1/k, k)$ für alle k wählen und diese Folge hat keinen Häufungspunkt. Also muss man bei der Auswahl von x_k Vorsicht walten lassen: Man wählt x_k so, dass $\|x_k\|$ minimal ist. Dann gilt $f(x) > f(x_k)$ für alle $\|x\| < \|x_k\|$. Die Annahme, dass (x_k) keinen Häufungspunkt besitzt, führt man auf die folgende Weise zum Widerspruch: Die Folge $(\frac{1}{\|x_k\|}x_k)$ besitzt einen Häufungspunkt w. OBdA konvergiere sie gegen w, dann zeigt man, dass $-w$ für $k \geq k_0$ eine zulässige Abstiegsrichtung ist und dass $\|x_k - \alpha w\| < \|x_k\|$ für $0 \leq \alpha \leq \varepsilon_k$ und für diese k gilt.

Fortsetzung des Beweises Wegen $K \neq \emptyset$ gilt $K_k \neq \emptyset$ für alle $k \geq k_0$, oBdA gelte $K_k \neq \emptyset$ für alle k. Dann ist die Menge

$$M_k := \{\|x\| : x \in K_k,\ f(x) = \min f(K_k)\}$$

kompakt und nicht leer, besitzt also ein Minimum. Man wähle $x_k \in K_k$ so dass gilt $\|x_k\| = \min M_k$, dann folgt

(i) $f(x_k) < f(x)$ für alle $x \in K$ mit $\|x\| < \|x_k\|$

(ii) $f(x_k) \to \inf f(K)$

Ich zeige nun, dass (x_k) einen Häufungspunkt x^* besitzt. Wenn dies gilt, ist x^* eine Lösung des MPs.

Angenommen, (x_k) besitzt keinen Häufungspunkt, dann konvergiert $(\|x_k\|)$ gegen ∞. OBdA gelte $x_k \neq 0$ für alle k. Dann hat die Folge $(w_k) := (\frac{1}{\|x_k\|}x_k)$ einen Häufungspunkt w, oBdA konvergiere $(\frac{1}{\|x_k\|}x_k)$ gegen w. Aus $Bx_k \leq d$ für alle k folgt $Bw_k \leq \frac{1}{\|x_k\|}d$ und daraus

$$Bw \leq 0 \tag{1}$$

Da $f(K)$ nach unten beschränkt ist, gibt es ein $\beta \in \mathbb{R}$ so dass gilt $\beta \leq f(x)$ für alle $x \in K$. Da $(f(x_k))$ monoton fällt, gibt es ein $\gamma \in \mathbb{R}$ so dass gilt $f(x_k) \leq \gamma$ für alle k. Es folgt $\beta \leq f(x_k) \leq \gamma$ für alle k und daraus

$$\beta \leq x_k^t A x_k + c^t x_k \leq \gamma \qquad \text{für alle } k$$

Dies impliziert

$$\frac{\beta}{\|x_k\|^2} \leq w_k^t A w_k + \frac{1}{\|x_k\|} c^t w_k \leq \frac{\gamma}{\|x_k\|^2}$$

und es folgt

$$w^t A w = 0 \tag{2}$$

Aus (1) folgt für alle k und alle $\alpha \geq 0$:

$$B(x_k + \alpha w) = Bx_k + \alpha Bw \leq Bx_k \leq d$$

und daher $x_k + \alpha w \in K$ für alle $\alpha \geq 0$ und alle k. Für alle $\alpha \in \mathbb{R}$ gilt nach 7.3:

$$f(x_k + \alpha w) = f(x_k) + \alpha \nabla f(x_k)^t w + \alpha^2 w^t A w = f(x_k) + \alpha \nabla f(x_k)^t w$$

und es folgt für alle $\alpha \geq 0$:

$$\beta \leq f(x_k + \alpha w) = f(x_k) + \alpha \nabla f(x_k)^t w$$

Dies impliziert $\nabla f(x_k)^t w \geq 0$ und

$$f(x_k - \alpha w) \leq f(x_k) \qquad \text{für alle } \alpha \geq 0 \text{ und alle } k \tag{3}$$

Also ist w eine Abstiegsrichtung. Es seien $b_1, \ldots, b_p$ die Zeilenvektoren von B. Dann gilt für alle $x \in \mathbb{R}^n$

$$Bx \leq 0 \Longleftrightarrow b_j^t x \leq 0 \qquad \text{für alle } j$$

und

$$Bx \leq d \Longleftrightarrow b_j^t x \leq d_j \qquad \text{für alle } j$$

Es gilt $Bw \leq 0$, also folgt $b_j^t w \leq 0$ für alle j. Wenn $b_j^t w = 0$ gilt, folgt für alle α:

$$b_j^t(x_k - \alpha w) = b_j^t x_k - \alpha b_j^t w \leq d_j$$

Wenn $b_j^t w < 0$ gilt, gibt es ein $\varepsilon > 0$ mit $b_j^t \dfrac{x_k}{\|x_k\|} \leq -\varepsilon$ für alle $k \geq k_0$. Es folgt

$$b_j^t x_k \leq -\varepsilon \|x_k\|$$

und daraus

$$b_j^t(x_k - \alpha w) = b_j^t x_k - \alpha b_j^t w \leq -\varepsilon \|x_k\| - \alpha b_j^t w \; .$$

Also gibt es ein k_j so dass gilt:

$$b_j^t(x_k - \alpha w) \leq d_j \qquad \text{für alle } k \geq k_j \text{ und alle } \alpha \leq 1$$

Daher gibt es ein k^* so dass gilt

$$x_k - \alpha w \in K \qquad \text{für alle } k \geq k^*,\ \alpha \leq 1 \tag{4}$$

Schließlich konvergiert $(\frac{1}{\|x_k\|} x_k^t w) = (w_k^t w)$ gegen $w^t w = 1$, also gibt es ein k' so dass gilt $x_k^t w > 0$ für alle $k \geq k'$. Es folgt für alle $\alpha > 0$:

$$\|x_k - \alpha w\|^2 = \|x_k\|^2 - 2\alpha x_k^t w + \alpha^2 \|w\|^2 = \|x_k\|^2 - \alpha(2x_k^t w - \alpha)$$

Daher gibt es ein $1 \geq \varepsilon_k > 0$ so dass gilt

$$\|x_k - \alpha w\| < \|x_k\| \quad \text{für alle } k \geq k' \text{ und alle } 0 < \alpha \leq \varepsilon_k. \tag{5}$$

Wählt man nun ein $k \geq \max\{k^*, k'\}$, dann gilt für alle $0 < \alpha \leq \varepsilon_k$:

(a) $x_k - \alpha w \in K$ nach (4)
(b) $f(x_k - \alpha w) \leq f(x_k)$ nach (3)
(c) $\|x_k - \alpha w\| < \|x_k\|$ nach (5)

im Widerspruch zu (ii). ■

Korollar 7.6 *Es sei $A \in M(n,n)$ positiv definit. Wenn das quadratische MP*

$$\begin{array}{ll} \min & x^t A x + c^t x \\ \textit{bez.} & Bx \leq d \end{array}$$

einen zulässigen Punkt besitzt, ist es (eindeutig) lösbar.

Beweis Man definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch $f(x) = x^t A x + c^t x$. Dann ist f streng konvex nach 7.3. Weiterhin gilt $\nabla f(x) = 2Ax + c$ für alle x. Setzt man nun $x^* = -\frac{1}{2} A^{-1} c$, dann gilt $\nabla f(x^*) = 0$, also ist x^* ein stationärer Punkt. Nach 4.12 löst x^* dann das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & x \in \mathbb{R}^n \end{array}$$

Also ist f nach unten beschränkt, insbesondere ist $\{f(x) : Bx \leq d\}$ nach unten beschränkt und die Behauptung folgt aus 7.5. ■

Die bisher erzielten Resultate erlauben einen schnellen Beweis des Dualitätssatzes der linearen Optimierung.

Satz 7.7 *Es seien $A \in M(q,n)$, $b \in \mathbb{R}^q$ und $c \in \mathbb{R}^n$. Man betrachte das MP*

$$\textit{(P)} \qquad \begin{array}{ll} \min & c^t x \\ \textit{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

sowie das Maximierungsproblem

$$\textit{(D)} \qquad \begin{array}{ll} \max & b^t \mu \\ \textit{bez.} & A^t \mu \leq c \end{array}$$

(Man nennt (P) das primale und (D) das duale MP.)

Dann gilt für alle (P)-zulässigen x und (D)-zulässigen μ:

$$b^t \mu \leq c^t x$$

Weiterhin sind äquivalent:

(i) (P) ist lösbar.

(ii) (D) ist lösbar.

(iii) (P) und (D) besitzen einen zulässigen Punkt.

Wenn dies der Fall ist, haben beide MPe denselben optimalen Wert der Zielfunktion, d.h. es gilt

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

Beweis Es seien x ein (P)-zulässiger und μ ein (D)-zulässiger Punkt, dann folgt:

$$b^t\mu = x^t A^t\mu \leq x^t c = c^t x$$

wobei die letzte Ungleichung aus $x \geq 0$ folgt.

"(iii) $\Rightarrow$ (i), (ii): Es seien x_0 ein (P)-zulässiges Element und μ_0 ein (D)-zulässiges Element. Dann folgt

$$b^t\mu_0 \leq c^t x \qquad \text{für alle (P)-zulässigen } x$$

und

$$b^t\mu \leq c^t x_0 \qquad \text{für alle (D)-zulässigen } \mu$$

Also ist (P) nach unten beschränkt. Da (P) nach Voraussetzung ein zulässiges Element besitzt, ist (P) nach 7.5 lösbar. Analog ist (D) lösbar.

Offenbar folgt (iii) aus (i) und (ii), so dass es reicht zu zeigen, dass (i) äquivalent zu (ii) ist.

"(i) $\Rightarrow$ (ii)": Man definiere $f : \mathbb{R}^n \to \mathbb{R},\ g : \mathbb{R}^n \to \mathbb{R}^n$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ durch

$$f(x) = c^t x, \qquad g(x) = -x, \qquad h(x) = b - Ax$$

dann ist (P) gerade das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Sei x^* eine Lösung von (P). Da das MP konvex ist und alle Nebenbedingungen affin sind, ist x^* nach 6.8 ein KT-Punkt des MPs. Also gibt es $\lambda^* \geq 0$ und μ^* so dass gelten

(i) $\nabla f(x^*) + \sum \lambda_i^* \nabla g_i(x^*) + \sum \mu_j^* \nabla h_j(x^*) = 0$

(ii) $(\lambda^*)^t g(x^*) = 0$

Nun gilt $\nabla f(x) = c$ und $\nabla g_i(x) = -e_i$ für alle i. Es seien $a_1^t, \dots a_q^t$ die Zeilenvektoren von B, dann gilt $h_j(x) = b_j - a_j^t x$ und daher $\nabla h_j(x) = -a_j$ für alle j. Es folgt

(i) $c - \sum \lambda_i^* e_i - \sum \mu_j^* a_j = 0$

(ii) $(\lambda^*)^t x^* = 0$

Dies impliziert

$$c = \sum \lambda_i^* e_i + \sum \mu_j^* a_j = \lambda^* + A^t \mu^*$$

und daraus $A^t \mu^* \leq c$, also ist μ^* ein (D)-zulässiger Punkt, sowie

$$b^t \mu^* = (Ax^*)^t \mu^* = (x^*)^t A^t \mu^* = (x^*)^t (c - \lambda^*) = (x^*)^t c - (x^*)^t \lambda^* = c^t x^*$$

Nun folgt für alle (D)-zulässigen μ:

$$b^t \mu \leq c^t x^* = b^t \mu^*$$

also ist μ^* eine Lösung von (D).

"(ii) $\Rightarrow$ (i)" Dieser Beweis verläuft im Wesentlichen analog: Definiert man dieses Mal $f : \mathbb{R}^p \to \mathbb{R}$ und $g : \mathbb{R}^q \to \mathbb{R}^n$ durch

$$f(\mu) = -b^t \mu \qquad \text{und} \qquad g(\mu) = A^t \mu - c$$

dann ist μ^* genau dann eine Lösung von (D), wenn μ^* eine Lösung von

$$\begin{array}{ll} \min & f(\mu) \\ \text{bez.} & g(\mu) \leq 0 \end{array}$$

ist. Sei μ^* eine Lösung von (D). Da f linear und g affin ist, ist μ^* ein KTP des MPs, also gibt es $x^* \geq 0$ so dass gelten:

(i) $\nabla f(\mu^*) + \sum x_i^* \nabla g_i(\mu^*) = 0$

(ii) $(x^*)^t g(\mu^*) = 0$

Es folgt $-b + Ax^* = 0$, also $Ax^* = b$, daher ist x^* ein (P)-zulässiger Punkt, und $(x^*)^t (A^t \mu^* - c) = 0$. Man erhält

$$b^t \mu^* = (x^*)^t A^t \mu^* = (x^*)^t c = c^t x^*$$

und daher ist x^* eine Lösung von (P).

Alternativ kann man "(ii) $\Rightarrow$ (i)" auch folgendermaßen beweisen: (D) ist genau dann lösbar, wenn das MP

$$\text{(D')} \qquad \begin{array}{ll} \max & b^t(\mu_1 - \mu_2) \\ \text{bez.} & A^t(\mu_1 - \mu_2) + \nu = c \\ & \mu_1, \mu_2, \nu \geq 0 \end{array}$$

lösbar ist. Und (D') ist genau dann lösbar, wenn das MP

$$\text{(D")} \qquad \begin{array}{ll} \min & (-b)^t(\mu_1 - \mu_2) \\ \text{bez.} & -A^t(\mu_1 - \mu_2) - \nu = -c \\ & \mu_1, \mu_2, \nu \geq 0 \end{array}$$

lösbar ist. Setzt man nun

$$\tilde{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \nu \end{pmatrix}, \quad \tilde{c} = \begin{pmatrix} -b \\ b \\ 0 \end{pmatrix}, \quad \tilde{A} = (-A^t, A^t, -I_n), \quad \tilde{b} = -c$$

Dann ist (D') gerade das MP

$$\begin{array}{ll} \min & \tilde{c}^t\tilde{\mu} \\ \text{bez.} & \tilde{A}\tilde{\mu} = \tilde{b} \\ & \tilde{\mu} \geq 0 \end{array}$$

Nach dem Bewiesenen ist das zu (D') duale Problem

$$\begin{array}{ll} \max & \tilde{b}^t x \\ \text{bez.} & \tilde{A}^t x \leq \tilde{c} \end{array}$$

lösbar. Also ist auch das MP

$$\text{(P')} \qquad \begin{array}{ll} \min & -\tilde{b}^t x \\ \text{bez.} & \tilde{A}^t x \leq \tilde{c} \end{array}$$

lösbar. Nun gilt

$$-\tilde{b}^t x = c^t x$$

und

$$\tilde{A}^t x = \begin{pmatrix} -A \\ A \\ -I_n \end{pmatrix} x = \begin{pmatrix} -Ax \\ Ax \\ -x \end{pmatrix}$$

also

$$\tilde{A}^t x \leq \tilde{c} = \begin{pmatrix} -b \\ b \\ 0 \end{pmatrix} \Leftrightarrow -Ax \leq -b,\ Ax \leq b,\ -x \leq 0 \Leftrightarrow Ax = b,\ x \geq 0$$

d.h. (P') ist gerade (P). ■

Es seien $A \in M(n,n)$ symmetrisch, positiv definit und $c \in \mathbb{R}^n$, dann betrachtet man das MP

$$\begin{array}{ll} \min & x^t Ax + c^t x \\ \text{bez.} & x \in \mathbb{R}^n \end{array}$$

Definiert man $f : \mathbb{R}^n \to \mathbb{R}$ durch

$$f(x) = x^t Ax + c^t x$$

dann ist f streng konvex und nach 4.12 löst ein Vektor $x^* \in \mathbb{R}^n$ das MP genau dann, wenn er ein stationärer Punkt des MPs ist, wenn also gilt $\nabla f(x^*) = 0$. Nun gilt $\nabla f(x) = 2Ax + c$ und daher löst x^* das MP genau dann, wenn gilt $2Ax^* + c = 0$. Daher ist x^* die Lösung eines linearen Gleichungssystems. Hier geht es nun darum, effektive Verfahren zu finden, die die besondere Struktur von A ausnutzen.

Um den Faktor $1/2$ bei der Lösung des MPs zu vermeiden, betrachte ich, wie allgemein üblich, das MP

$$\begin{array}{ll} \min & \frac{1}{2}\, x^t Ax + c^t x \\ \text{bez.} & x \in \mathbb{R}^n \end{array}$$

Definition 7.8 *Es sei $A \in M(n,n)$ symmetrisch und positiv definit. Zwei Vektoren $u, v \in \mathbb{R}^n$ heißen A-**konjugiert** oder kurz **konjugiert**, wenn gilt $u^t Av = 0$.*

Es sei $A \in M(n,n)$ symmetrisch und positiv definit. Dann ist die Abbildung

$$\beta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

definiert durch

$$\beta(u, v) = u^t Av$$

ein Skalarprodukt auf $\mathbb{R}^n$. Zwei Vektoren $u, v \in \mathbb{R}^n$ sind genau dann A-konjugiert, wenn sie orthogonal in $(\mathbb{R}^n, \beta)$ sind. Also stehen hier die Methoden der Theorie der euklidischen Vektorräume zur Verfügung. Insbesondere gilt:

(i) Es seien $v_1, \ldots, v_k$ paarweise A-konjugierte Vektoren, die alle von Null verschieden sind. Dann sind $v_1, \ldots, v_k$ linear unabhängig.

(ii) $\mathbb{R}^n$ besitzt eine Basis aus A-konjugierten Vektoren. In der Tat gibt es zu jeder Basis $u_1, \ldots, u_n$ des $\mathbb{R}^n$ eine Basis A-konjugierter Vektoren $v_1, \ldots, v_n$ so dass $u_1, \ldots, u_k$ und $v_1, \ldots, v_k$ für alle k denselben Untervektorraum erzeugen. Praktisch kann man diese Vektoren mit dem Gram-Schmidt-Verfahren finden.

Proposition 7.9 *Es seien $A \in M(n,n)$ symmetrisch, positiv semidefinit und $c \in \mathbb{R}^n$. Man definiere $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ durch*

$$f(x) = \frac{1}{2} x^t Ax + c^t x$$

Weiterhin seien $x_0, v_1, \ldots, v_k \in \mathbb{R}^n$ sowie

$$L = x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_k = \{x_0 + \alpha_1 v_1 + \cdots + \alpha_k v_k : \alpha_1, \ldots, \alpha_k \in \mathbb{R}\}$$

Ein Punkt $x^* \in L$ löst das MP

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & x \in L \end{array}$$

genau dann, wenn gilt

$$\nabla f(x^*)^t v_i = 0 \qquad \textit{für } i = 1, \ldots, k \; .$$

Beweis Nach 4.12 löst x^* das MP genau dann, wenn x^* ein stationärer Punkt ist, wenn also gilt

$$\nabla f(x^*)^t (x - x^*) \geq 0 \qquad \text{für alle } x \in L \; .$$

Sei $x^* = x_0 + \alpha_1^* v_1 + \cdots + \alpha_k^* v_k$, dann ist dies äquivalent zu:

$$\nabla f(x^*)^t (x_0 + \alpha_1 v_1 + \cdots + \alpha_k v_k - (x_0 + \alpha_1^* v_1 + \cdots + \alpha_k^* v_k)) \geq 0 \qquad \text{für alle } \alpha \in \mathbb{R}^k$$

also

$$\nabla f(x^*)^t ((\alpha_1 - \alpha_1^*) v_1 + \cdots + (\alpha_k - \alpha_k^*) v_k) \geq 0 \qquad \text{für alle } \alpha \in \mathbb{R}^k$$

Dies ist aber offenbar äquivalent zu

$$\nabla f(x^*)^t v_i = 0 \qquad \text{für } i = 1, \ldots, k$$

■

Korollar 7.10 *Es seien $A \in M(n,n)$ symmetrisch, positiv definit und $c \in \mathbb{R}^n$. Man definiere $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ durch*

$$f(x) = \frac{1}{2} x^t A x + c^t x \; .$$

Weiterhin seien $x_0 \in \mathbb{R}^n$ und $v \in \mathbb{R}^n$, $v \neq 0$. Dann nimmt f sein Minimum auf $x_0 + \mathbb{R}v$ in dem Punkt $x^* = x_0 + \alpha^* v$ mit

$$\alpha^* = -\frac{(Ax_0 + c)^t v}{v^t A v} = -\frac{\nabla f(x_0)^t v}{v^t A v}$$

an.

Beweis Nach 7.9 nimmt f sein Minimum auf $x_0 + \mathbb{R}v$ genau dann in $x^* = x_0 + \alpha^* v$ an, wenn gilt

$$\nabla f(x^*)^t v = 0$$

Nun gilt $\nabla f(x) = Ax + c$ für alle x und daher

$$\nabla f(x_0 + \alpha v) = Ax_0 + \alpha Av + c$$

Also folgt

$$\nabla f(x_0 + \alpha v)^t v = x_0^t A v + \alpha v^t A v + c^t v$$

und daher gilt $\nabla f(x_0 + \alpha^* v)^t v = 0$ genau dann, wenn gilt

$$\alpha^* = -\frac{(x_0^t A + c^t)v}{v^t A v} = -\frac{(Ax_0 + c)^t v}{v^t A v}$$

■

Die folgende Proposition ist der zentrale Punkt beim Verfahren der konjugierten Gradienten:

Proposition 7.11 *Es seien $A \in M(n,n)$ symmetrisch, positiv semidefinit, $c \in \mathbb{R}^n$ und $f : \mathbb{R}^n \to \mathbb{R}$ definiert durch $f(x) = \frac{1}{2}x^t Ax + c^t x$. Weiterhin seien $v_1, \ldots, v_k$ paarweise konjugierte Vektoren, und $x_0 \in \mathbb{R}^n$. Man setze*

$$L = x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_{k-1}$$

und für $x^* \in L$ gelte $f(x^*) = \min f(L)$. Schließlich gelte für $\alpha^* \in \mathbb{R}$:

$$f(x^* + \alpha^* v_k) = \min f(x^* + \mathbb{R}v_k)$$

Dann gilt

$$f(x^* + \alpha^* v_k) = \min f(L + \mathbb{R}v_k) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_k)\ .$$

Beweis Nach 7.9 gilt

$$\nabla f(x^*)^t v_j = 0 \quad \text{für } j = 1, \ldots, k-1$$

und

$$\nabla f(x^* + \alpha^* v_k)^t v_k = 0$$

Es gilt $\nabla f(x) = Ax + c$ und es folgt für alle $1 \le j \le k-1$:

$$\begin{aligned} \nabla f(x^* + \alpha^* v_k)^t v_j &= (A(x^* + \alpha^* v_k) + c)^t v_j \\ &= (Ax^* + c)^t v_j + \alpha^* v_k^t A v_j \\ &= \nabla f(x^*)^t v_j + \alpha^* v_k^t A v_j \\ &= 0 \end{aligned}$$

da v_k und v_j konjugiert sind. Die Behauptung folgt dann wieder aus 7.9. ■

Proposition 7.12 *Vorgegeben sei das quadratische MP*

$$\begin{array}{ll} \min & \frac{1}{2}\, x^t Ax + c^t x \\ bez. & x \in \mathbb{R}^n \end{array}$$

Es seien A symmetrisch, positiv definit und $v_1, \ldots, v_n$ eine Basis paarweise A-konjugierter Vektoren des $\mathbb{R}^n$. Man definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch

$$f(x) = \frac{1}{2}x^t Ax + c^t x$$

und betrachte das folgende Verfahren:

(S1) Man wähle $x_0 \in \mathbb{R}^n$ beliebig.

(S2) Es seien $x_0, \ldots, x_k$ definiert. Dann setze man

$$\alpha_k = -\frac{\nabla f(x_k)^t v_k}{v_k^t A v_k}$$

und

$$x_{k+1} = x_k + \alpha_k v_k$$

Dann löst x_{n+1} das MP.

Beweis Ich zeige induktiv, dass für alle $k \geq 2$ gilt:

$$f(x_k) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_{k-1}) \ ,$$

dann folgt $f(x_{n+1}) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_n) = f(\mathbb{R}^n)$.

Nach 7.10 gilt:

$$f(x_2) = \min f(x_0 + \mathbb{R}v_1)$$

die Behauptung gelte für k. Nach 7.10 gilt dann $f(x_{k+1}) = \min f(x_k + \mathbb{R}v_k)$ und aus 7.11 folgt dann $f(x_{k+1}) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_k)$. ■

Das Verfahren aus 7.12 setzt voraus, dass eine Basis aus konjugierten Vektoren bekannt ist. Wenn dies nicht der Fall ist, kann man eine solche Basis natürlich konstruieren, indem man die Standardbasis des $\mathbb{R}^n$ "A-orthogonalisiert". Es ergibt sich die Frage, ob es eine bessere Basis für die A-Orthogonalisierung gibt. Dies ist in der Tat der Fall:
Nehmen wir an, $x_0, \ldots, x_k$ und $v_1, \ldots, v_k$ seien so konstruiert, dass die v_i paarweise A-konjugiert sind und dass gilt $f(x_k) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_{k-1})$, dann bestimmt man x_{k+1} so dass gilt $f(x_{k+1}) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_k)$. Um einen Vektor v_{k+1} zu finden, der A-konjugiert zu $v_1, \ldots, v_k$ ist, braucht man nun zunächst einen Vektor, der von $v_1, \ldots, v_k$ linear unabhängig ist. Und dafür gibt es einen guten Kandidaten: Nach 7.9 gilt ja $\nabla f(x_{k+1})^t v_i = 0$ für $i = 1, \ldots, k$. Falls nun $\nabla f(x_{k+1}) = 0$ gilt, ist x_{k+1} ein stationärer Punkt und damit eine Lösung des MPs. Anderfalls ist $\nabla f(x_{k+1})$ linear unabhängig von $v_1, \ldots, v_k$ und man kann diesen Vektor dann so modifizieren, dass er in der Tat A-orthogonal zu $v_1, \ldots, v_k$ ist. Nun ist ja $-\nabla f(x_k)$ eine Abstiegsrichtung und deswegen ersetzt man in den obigen Überlegungen $\nabla f(x_k)$ durch $-\nabla f(x_k)$. Die technische Ausführung dieser Überlegungen führt dann zu dem

Verfahren 7.13 *(Verfahren der konjugierten Gradienten, conjugate gradients method, CG-Verfahren) Es seien $A \in M(n,n)$ symmetrisch und positiv definit sowie $c \in \mathbb{R}^n$. Vorgegeben sei das quadratische MP*

$$\begin{array}{ll} \min & \frac{1}{2}x^t A x + c^t x \\ \textit{bez.} & x \in \mathbb{R}^n \end{array}$$

Man definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch

$$f(x) = \frac{1}{2}x^t Ax + c^t x$$

und betrachte den folgenden Algorithmus:

(S1) Man wähle ein $x_0 \in \mathbb{R}^n$ beliebig und setze $v_1 = -\nabla f(x_0)$.

(S2) Es seien $x_0, \ldots, x_k$ und $v_1, \ldots, v_k$ definiert. Wenn $\nabla f(x_k) = 0$ gilt, bricht man das Verfahren ab. Wenn dies nicht der Fall ist, setze man

$$\begin{aligned} \alpha_k &= -\frac{\nabla f(x_k)^t v_k}{v_k^t A v_k} \\ x_{k+1} &= x_k + \alpha_k v_k \\ \beta_k &= \frac{\nabla f(x_{k+1})^t A v_k}{v_k^t A v_k} \\ v_{k+1} &= -\nabla f(x_{k+1}) + \beta_k v_k \end{aligned}$$

Proposition 7.14 *Es seien $A \in M(n,n)$ symmetrisch und positiv definit sowie $c \in \mathbb{R}^n$. Vorgegeben sei das quadratische MP*

$$\begin{array}{ll} \min & \frac{1}{2}x^t Ax + c^t x \\ \textit{bez.} & x \in \mathbb{R}^n \end{array}$$

Dann bricht das Verfahren der konjugierten Gradienten mit einem beliebigen Startwert nach höchstens n Schritten mit einer Lösung des MPs ab.

Zusatz
Wenn das Verfahren bis zum $(k-1)$-ten Schritt nicht abgebrochen ist, gilt $v_1, \ldots, v_k$, $\nabla f(x_0), \ldots, \nabla f(x_k) \neq 0$ und

(1) $f(x_k) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_{k-1})$ *für alle* $k \geq 2$

(2) $v_i^t A v_j = 0$ *für alle* $i \neq j$ *mit* $i, j \leq k$

(3) $\nabla f(x_i)^t \nabla f(x_j) = 0$ *für alle* $i \neq j$ *mit* $i, j \leq k$

(4) $\nabla f(x_k)^t v_k = -\|\nabla f(x_k)\|^2$

Beweis Ich zeige zunächst die Behauptungen des Zusatzes durch vollständige Induktion nach k. Der Fall $k = 1$ ist einfach. Also gelte die Behauptung für k, dann sind $v_1, \ldots, v_k$ paarweise konjugiert und $\nabla f(x_0), \ldots, \nabla f(x_k)$ paarweise orthogonal, und die $2k$ Vektoren sind alle von 0 verschieden. Wenn das Verfahren im (k+1)-ten Schritt nicht abbricht, gilt $\nabla f(x_{k+1}) \neq 0$.

Beweis von (1): Nach 7.10 gilt

$$f(x_{k+1}) = \min f(x_k + \mathbb{R}v_k)$$

Aus 7.11 folgt dann

$$f(x_{k+1}) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_k)$$

und damit (1).

Beweis von (4): Nach 7.9 gilt

$$\nabla f(x_{k+1})^t v_i = 0 \quad \text{für alle } i \leq k$$

Es folgt

$$\nabla f(x_{k+1})^t v_{k+1} = \nabla f(x_{k+1})^t(-\nabla f(x_{k+1}) + \beta_k v_k) = -\|\nabla f(x_{k+1})\|^2$$

und damit (4).

Beweis von (3):Für alle $i \leq k$ gilt:

$$\nabla f(x_{k+1})^t \nabla f(x_i) = \nabla f(x_{k+1})^t(\beta_{i-1}v_{i-1} - v_i) = 0$$

und damit (3).

Beweis von (2):
Nach (3) sind $\nabla f(x_0), \ldots, \nabla f(x_{k+1})$ linear unabhängig, insbesondere folgt dann $x_{i+1} \neq x_i$ für alle $i \leq k$. Nun sei $i \leq k$, dann gilt:

$$x_{i+1} - x_i = \alpha_i v_i$$

und daher $\alpha_i \neq 0$. Es folgt:

$$\alpha_i A v_i = A(x_{i+1} - x_i) = \nabla f(x_{i+1}) - \nabla f(x_i)$$

Man erhält für alle $i < k$:

$$\begin{aligned} v_{k+1}^t A v_i &= (-\nabla f(x_{k+1})^t + \beta_k v_k) A v_i \\ &= -\nabla f(x_{k+1})^t A v_i \\ &= \frac{1}{\alpha_i} \nabla f(x_{k+1})^t (\nabla f(x_{i+1}) - \nabla f(x_i)) \\ &= 0 \end{aligned}$$

Schließlich gilt:

$$v_{k+1}^t A v_k = -\nabla f(x_{k+1})^t A v_k + \beta_k v_k^t A v_k = -\nabla f(x_{k+1})^t A v_k + \nabla f(x_{k+1})^t A v_k = 0$$

und daher (2).

Da das Verfahren nicht abgebrochen ist, gilt $\nabla f(x_{k+1}) \neq 0$. Wegen $\nabla f(x_{k+1})^t v_k = 0$ sind $\nabla f(x_{k+1})$ und v_k linear unabhängig und daher gilt

$$v_{k+1} = \nabla f(x_{k+1}) + \beta_k v_k \neq 0$$

Damit ist der Zusatz bewiesen.

Da $v_1, \ldots, v_n$ von 0 verschieden sind, sind sie linear unabhängig und daher eine Basis des $\mathbb{R}^n$. Falls das Verfahren bis zum (n-1)-ten Schritt nicht abgebrochen ist, folgt nach (1):

$$f(x_{n+1}) = \min f(x_0 + \mathbb{R}v_1 + \cdots + \mathbb{R}v_n) = \min f(\mathbb{R}^n) \qquad \blacksquare$$

Es seien nun $A \in M(n,n)$ positiv definit und $c \in \mathbb{R}^n$. Wenn man das lineare Gleichungssystem $Ax + c = 0$ mit dem CG-Verfahren löst, ist der aufwendigste Teil jedes Schrittes die Berechnung von Av_k oder $\nabla f(x_{k+1})^t A$, also die Multiplikation einer Matrix mit einem Vektor. Dazu braucht man im allgemeinen n^2 Multiplikationen. Da das Verfahren spätestens nach n Schritten abbricht, liegt die Anzahl der notwendigen Multiplikationen in der Größenordnung von n^3, im Gegensatz zum Gauß-Verfahren, bei dem diese Zahl in der Größenordnung von $n^3/3$ liegt, also erheblich geringer ist. Das CG-Verfahren ist aber dennoch dem Gauß-Verfahren in wenigstens zwei Fällen überlegen: Wenn die Matrix sehr groß, aber dünn besetzt ist (also viele Nullen enthält). Während das Gauß-Verfahren diese Eigenschaft eher zerstört, profitiert das Verfahren konjugierter Gradienten davon erheblich. Den anderen Fall erhält man, wenn man die Konvergenz des CG-Verfahrens näher untersucht: Es sei κ die Kondition der Matrix, also der Quotient aus dem größten und kleinsten Eigenwert von A, dann gilt für die Lösung x^*:

$$\|x_{k+1} - x^*\| \leq 2\sqrt{\kappa} \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|x_0 - x_0\|$$

Wenn A also gut konditioniert ist, konvergiert das CG-Verfahren sehr schnell. Gelegentlich nimmt man in der Tat eine konditionsverbessernde Koordinatentransformation vor, um die Konvergenzgeschwindigkeit zu verbessern.

Bei der Verallgemeinerung des CG-Verfahrens auf beliebige ein- oder zweimal stetig differenzierbare Abbildungen kann man nun bedenken, dass für eine quadratische Abbildung ja $A = Hf(x)$ für alle x gilt und A dann entsprechend ersetzen. Dieses erfordet natürlich die Berechnung zweiter Ableitungen, was oft aufwendig ist. Die folgende Bemerkung zeigt, dass man im quadratischen Fall ohne zweite Ableitungen auskommt:

Bemerkung 7.15 Beim Verfahren der konjugierten Gradienten gilt für alle k:

$$\nabla f(x_{k+1}) - \nabla f(x_k) = Ax_{k+1} - Ax_k = \alpha_k A v_k$$

also folgt:

$$\alpha_k \nabla f(x_{k+1})^t A v_k = \nabla f(x_{k+1})^t (\nabla f(x_{k+1}) - \nabla f(x_k)) = \|\nabla f(x_{k+1})\|^2$$

und

$$\begin{aligned} \alpha_k v_k^t A v_k &= (-\nabla f(x_k) + \beta_{k-1} v_{k-1})^t (\nabla f(x_{k+1}) - \nabla f(x_k)) \\ &= \|\nabla f(x_k)\|^2 \end{aligned}$$

und daher gilt

$$\beta_k = \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2}$$

Daher kann man beim Verfahren der konjugierten Gradienten v_{k+1} auch aus der Gleichung $v_{k+1} = -\nabla f(x_{k+1}) + \beta_k v_k$ mit

$$\beta_k = \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2}$$

bestimmen.

Verfahren 7.16 *(Verfahren von Fletcher und Reeves) Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Abbildung. Man betrachte den folgenden Algorithmus:*

(S1) Man wähle ein $x_0 \in \mathbb{R}^n$ beliebig und setze $v_1 = -\nabla f(x_0)$.

(S2) Es seien $x_0, \ldots, x_k$ und $v_1, \ldots, v_k$ definiert. Wenn $\nabla f(x_k) = 0$ gilt, bricht man das Verfahren ab.

Wenn dies nicht der Fall ist, wähle man α_k so dass gilt

$$f(x_k + \alpha_k v_k) = \min f(x_k + \mathbb{R} v_k)$$

und setzt

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k v_k \\ \beta_k &= \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2} \\ v_{k+1} &= -\nabla f(x_{k+1}) + \beta_k v_k \end{aligned}$$

Man beachte, dass beim CG-Verfahren α_k gerade so bestimmt worden ist, dass $f(x_k + \alpha_k v_k) = \min f(x_k + \mathbb{R} v_k)$ gilt, so dass das Verfahren von Fletscher und Reeves für quadratische MPe mit positiv definiter Matrix A gerade das CG-Verfahren ist.

Das quadratische MP

$$\begin{array}{ll} \min & x^t A x + c^t x \\ \text{bez.} & Bx \leq d \end{array}$$

genügt (z.B. nach Übungsaufgabe 27) der Regularitätsbedingung von Abadie. Also ist nach 5.15 jede Lösung des MPs ein KT-Punkt. Die Beschreibung eines KT-Punktes ist nun sehr einfach, wenn die Nebenbedingungen nur aus Gleichheitsbedingungen bestehen:

Proposition 7.17 *Es seien $A \in M(n,n)$ symmetrisch, $c \in \mathbb{R}^n, B \in M(q,n)$ und $d \in \mathbb{R}^q$. Ein Punkt $x^* \in \mathbb{R}^n$ ist genau dann ein KTP des quadratischen MPs*

$$\begin{array}{ll} \min & \frac{1}{2}x^tAx + c^tx \\ \textit{bez.} & Bx = d \end{array}$$

wenn es ein $\mu^* \in \mathbb{R}^n$ so gibt, dass (x^*, μ^*) eine Lösung des folgenden linearen Gleichungssystems ist:

$$\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} -c \\ d \end{pmatrix}$$

Beweis Man definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch $f(x) = \frac{1}{2}x^tAx + c^tx$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ durch $h(x) = Bx - d$. Ein Punkt $x^* \in \mathbb{R}^n$ ist genau dann ein KTP, wenn es ein $\mu^* \in \mathbb{R}^q$ so gibt, dass gilt

(1) $$\nabla f(x^*) + \sum \mu_j^* \nabla h_j(x^*) = 0$$

Es seien $b_1, \ldots, b_q$ die Zeilenvektoren von B, dann gilt für alle j:

$$h_j(x) = b_j^t x - d_j \qquad \text{und} \qquad \nabla h_j(x) = b_j$$

Also ist (1) äquivalent zu:

(2) $$Ax^* + c + \sum \mu_j^* b_j = 0$$

und dies ist äquivalent zu:

(3) $$Ax^* + c + \sum B^t \mu^* = 0$$

Dies ist gerade die erste Zeile, die zweite beschreibt die Zulässigkeit von x^*. ■

Natürlich ist es nützlich, hinreichende Kriterien zu haben, die die Existenz eines KTPes garantieren. Ein häufig nützliches Kriterium ist das folgende:

Proposition 7.18 *Es seien $A \in M(n,n)$ symmetrisch und positiv definit und für $B \in M(q,n)$ gelte $rg(B) = q$. Dann ist die Matrix*

$$\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix}$$

regulär. Also besitzt für alle $s \in \mathbb{R}^n$ und $d \in \mathbb{R}^q$ das MP

$$\begin{array}{ll} \min & \frac{1}{2}x^tAx + c^tx \\ \textit{bez.} & Bx = d \end{array}$$

genau einen KTP x^*. Weiterhin ist x^* die einzige Lösung des MPs.

Beweis Ich zeige zunächst, dass die Matrix regulär ist:

Es gelte

$$\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Dann folgt $Ax + B^t\mu = Bx = 0$ und daraus

$$0 = x^tAx + x^tB^t\mu = x^tAx + (Bx)^t\mu = x^tAx\ .$$

Da A positiv definit ist, folgt $x = 0$ und daraus $B^t\mu = 0$. Es seien $b_1^t, \ldots, b_q^t$ die Zeilenvektoren von B, dann sind $b_1^t, \ldots, b_q^t$ wegen $rg(B) = q$ linear unabhängig. Es folgt

$$0 = B^t\mu = \sum_{j=1}^{q} \mu_j b_j$$

und daraus $\mu = 0$.

Da A positiv definit ist, ist die Abbildung f definiert durch

$$f(x) = \frac{1}{2}x^tAx + c^tx$$

konvex und nach 6.8 löst ein Punkt das MP genau dann, wenn er ein KT-Punkt ist. Da die Matrix

$$\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix}$$

regulär ist, besitzt das MP nach 7.17 genau einen KTP. ■

Beispiel 7.19

$$\begin{array}{ll} \min & u^2 + v^2 + w^2 \\ \text{bez.} & u + 2v - w = 4 \\ & u - v + w = -2 \end{array}$$

Das folgende LGS ist zu lösen:

$$\begin{pmatrix} 2 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 2 & -1 \\ 0 & 0 & 2 & -1 & 1 \\ 1 & 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 4 \\ -2 \end{pmatrix}$$

und die Lösung ist

$$(x^*, \mu^*) = \frac{1}{7}(2, 10, -6, -8, 4)$$

Also ist $x^* = \frac{1}{7}(2, 10, -6)$ der einzige KTP mit zugehörigem Lagrange-Multiplikator $\mu^* = \frac{1}{7}(-8, 4)$ und x^* die einzige Lösung des MPs. ■

Es seien $f : \mathbb{R}^n \to \mathbb{R}$ eine quadratische Abbildung und $g : \mathbb{R}^n \to \mathbb{R}^p$ sowie $h : \mathbb{R}^n \to \mathbb{R}^q$ affine Abbildungen. Vorgegeben sei das MP

$$\begin{array}{lll} & \min & f(x) \\ \text{(MP)} & \text{bez.} & g(x) \leq 0 \\ & & h(x) = 0 \end{array}$$

Wenn x^* eine Lösung des MPs ist, ist x^* ein KTP, also gibt es $\lambda^* \geq 0$ und μ^* so dass gelten:

$$\begin{array}{ll} \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{q} \mu_j^* \nabla h_j(x^*) = 0 & \\ \lambda_i^* g_i(x^*) = 0 & \text{für alle } i \end{array}$$

Dies kann man auch schreiben in der Form:

$$\begin{array}{ll} \nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{q} \mu_j^* \nabla h_j(x^*) = 0 & \\ \lambda_i^* = 0 & \text{für alle } i \notin I(x^*) \end{array}$$

Bei der Lösung dieses Gleichungssystems erweist es sich als ein Problem, dass $I(x^*)$ nicht bekannt ist. Bei den Beispielen in Kapitel 5 und den zugehörigen Übungen bin ich davon ausgegangen, dass jede Menge $I \subseteq \{1, \ldots, p\}$ ein Kandidat für $I(x^*)$ ist und habe dann das entsprechende Gleichungssystem untersucht. Das ist natürlich kein praktisches Verfahren, da die Anzahl der Teilmengen sehr schnell sehr groß wird. Bei der Methode der aktiven Mengen (active set method) konstruiert man nicht nur eine Folge zulässiger Punkte (x_k), die einen KTP x^* approximieren soll, sondern auch eine Folge (I_k) von Teilmengen von $\{1, \ldots, p\}$, die $I(x^*)$ "annähern" soll (und in in guten Fällen mit $I(x^*)$ endet). Das Verfahren beginnt wie üblich mit einem beliebigen zulässigen Punkt x_0 und dieses Mal mit einer Menge $I_0 \subseteq I(x_0)$. Wenn nun x_k und $I_k \subseteq I(x_k)$ konstruiert sind, bedenkt man das Folgende:

Wenn x^* eine Lösung des MPs ist, ist x^* eine (MP)-zulässige Lösung des MPs

$$\begin{array}{lll} \min & f(x) & \\ \text{bez.} & g_i(x) = 0 & \text{für alle } i \in I(x^*) \\ & h(x) = 0 & \end{array}$$

Wenn nun $I_k = I(x^*)$ gälte, dann wäre x^* eine (MP)-zulässige lokale Lösung des MPs

$$\begin{array}{lll} \min & f(x) & \\ \text{bez.} & g_i(x) = 0 & \text{für alle } i \in I_k \\ & h(x) = 0 & \end{array}$$

Also löst man dieses MP, man betrachtet zwei Fälle:

Fall 1 x_k löst dieses MP. Dann ist x_k ein KTP dieses MPs und daher gibt es $(\lambda_i)_{i \in I_k}$ und (μ_j) so dass gilt

$$\nabla f(x_k) + \sum_{i \in I_k} \lambda_i \nabla g_i(x_k) + \sum \mu_j \nabla h_j(x_k) = 0$$

Falls nun $\lambda_i \geq 0$ für alle $i \in I_k$ gilt, setze man $\lambda_i = 0$ für alle $i \notin I_k$. Dann gilt $\lambda_i \geq 0$ für alle i und

$$\lambda_i g_i(x_k) = 0 \qquad \text{für alle } i \,.$$

Also ist x_k ein KTP von (MP) und das Verfahren bricht ab. Nehmen wir an, es gelte $\lambda_j < 0$ für ein j. Wenn x_k regulär bzgl. (MP) ist, ist die obige Darstellung eindeutig und x_k kein KTP von (MP) und daher keine Lösung von (MP). Also gilt $f(x^*) < f(x_k)$ und daher $I(x^*) \neq I_k$. Nun kann man im nächsten Schritt I_k vergrößern oder verkleinern. Wenn man I_k vergrößert, verkleinert man den zulässigen Bereich und vergrößert das Minimum. Das ist offenbar nicht nützlich und daher verkleinert man I_k. Die Frage ist, welches Element man aus I_k herausnehmen soll und hier bietet sich natürlich jedes $\lambda_i < 0$ an. Üblicherweise nimmt man das kleinste und setzt $x_{k+1} = x_k$ und $I_{k+1} = I_k \setminus \{\lambda_i\}$.

Fall 2 x_k löst dieses MP nicht. Dann sei y_k eine Lösung dieses MPs. Wenn y_k zulässig für (MP) ist, setzt man $x_{k+1} = y_k$ und $I_{k+1} = I_k$. Offenbar gilt in diesem Fall $f(x_{k+1}) < f(x_k)$. Wenn y_k nicht (MP)-zulässig ist, setze man

$$\alpha_k = \max\{\alpha \geq 0 : g_i(x_k + \alpha(y_k - x_k)) \leq 0 \text{ für alle } i\}$$

und $x_{k+1} = x_k + \alpha_k(y_k - x_k)$. Also ist x_{k+1} "letzte" Punkt auf der Verbindungsgerade, der noch zulässig ist. Weiterhin wählt man ein r, das "verhindert", dass α_k größer wird. Dann gilt $g_r(x_{k+1}) = 0$ und man setzt $I_{k+1} = I_k \cup \{r\}$.

Der Punkt des Verfahrens ist nun die Tatsache, dass für unendlich viele k der 1. Fall eintritt und ebenso für unendlich viele k der 2. Fall eintritt, wenn das Verfahren nicht abbricht: Wenn für k der 1. Fall eintritt, ist I_{k+1} echt in I_k enthalten. Wenn auch für $k+1$ der 1. Fall eintritt, ist I_{k+2} echt in I_{k+1} enthalten. Dies ist offenbar nur endlich oft möglich, so dass es ein r gibt so dass für $k+r$ der 2. Fall gilt. Wenn andererseits für k der 2. Fall gilt und für $k+1$ nicht der 1. Fall gilt, ist I_{k+1} echt größer als I_k. Wenn auch für $k+1$ nicht der 1. Fall eintritt ist I_{k+2} echt größer als I_{k+1}. Also tritt nach endlich vielen Schritten wieder der 1. Fall ein.
Wenn für k der 1. Fall gilt, löst x_k das MP

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & g_i(x) = 0 \qquad \text{für alle } i \in I_k \\ & h(x) = 0 \end{array}$$

Da es nur endlich viele Teilmengen von $\{1, \ldots, p\}$ gibt, gibt es eine Menge $I_0 \subseteq \{1, \ldots, p\}$ und eine unendliche Menge $J \subseteq \mathbb{N}$ so dass für alle $k \in J$ der 1. Fall zutrifft und dass $I_k = I_0$ gilt. Dann folgt aber $f(x_k) = f(x_r)$ für alle $k, r \in J$ und die Folge $f(x_k)$ wird konstant.

Wenn andererseits für k der 2. Fall gilt und y_k zulässig ist, gilt $f(x_{k+1}) = f(y_k) < f(x_k)$. Schwieriger wird die Sache, wenn $x_{k+1} = x_k + \alpha_k(y_k - x_k)$ für $\alpha_k < 1$ gilt, denn dann ist nicht gesichert, dass $f(x_{k+1}) \leq f(x_k)$ oder gar $f(x_{k+1}) < f(x_k)$ gilt.

Im allgemeinen ist durchaus möglich, dass $f(x_{k+1}) > f(x_k)$ gilt und man muss diesen Fall durch geeignete Voraussetzungen an f vermeiden. Wenn nun aber für alle k, die dem 2. Fall genügen, $f(x_{k+1}) < f(x_k)$ gilt, bricht das Verfahren ab. Damit erhält man prinzipiell das folgende Verfahren:

Verfahren 7.20 *(der aktiven Mengen, Prototyp) Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine stetig differenzierbare Abbildung und $g : \mathbb{R}^n \to \mathbb{R}^p$ sowie $h : \mathbb{R}^n \to \mathbb{R}^q$ affine Abbildungen. Vorgegeben sei das MP*

$$\begin{array}{lll} & \min & f(x) \\ (MP) & \textit{bez.} & g(x) \leq 0 \\ & & h(x) = 0 \end{array}$$

Man betrachte das folgende Verfahren:

(S1) Man wähle einen zulässigen Vektor $x_0 \in \mathbb{R}^n$ und $I_0 \subseteq I(x_0)$.

(S2) Es seien (MP)-zulässige Punkte $x_0, \ldots, x_k$ und Mengen $I_0 \subseteq I(x_0), \ldots, I_k \subseteq I(x_k)$ konstruiert. Man betrachte das MP

$$\begin{array}{lll} \min & f(x) & \\ \textit{bez.} & g_i(x) = 0 & \textit{für alle } i \in I_k \\ & h(x) = 0 & \end{array}$$

Fall 1 *x_k löst dieses MP, dann suche man Lagrange-Multiplikatoren $(\lambda_i)_{i \in I_k}$ und $\mu \in \mathbb{R}^q$ so dass gilt*

$$\nabla f(x_k) + \sum_{i \in I_k} \lambda_i \nabla g_i(x_k) + \sum \mu_j \nabla h_j(x_k) = 0$$

a) *Wenn $\lambda_i \geq 0$ für alle $i \in I_k$ gilt, ist x_k ein stationärer Punkt von (MP) und das Verfahren bricht ab.*
b) *Andernfalls wähle man r so dass gilt $\lambda_r = \min\{\lambda_i : i \in I_k\}$ und setze*

$$x_{k+1} = x_k \qquad \textit{und} \qquad I_{k+1} = I_k \setminus \{\lambda_r\}$$

Fall 2 *x_k ist keine Lösung dieses MPs. Dann sei y_k eine Lösung.*

a) *Wenn y_k (MP)-zulässig ist, setze man $x_{k+1} = y_k$ und $I_{k+1} = I_k$.*

b) *y_k ist nicht (MP)-zulässig. Dann setze man*

$$\alpha_k = \max\{\alpha \leq 1 : g_i(x_k + \alpha(y_k - x_k)) \leq 0\}$$

und

$$x_{k+1} = x_k + \alpha_k(y_k - x_k)$$

Schließlich wählt man ein $r \notin I_k$ so dass gilt $g_r(x_{k+1}) = 0$ und setzt

$$I_{k+1} = I_k \cup \{r\}$$

Man wählt also im Fall 2b) den zulässigen Punkt auf der Verbindungsgerade zwischen x_k und y_k, der am Nächsten zu y_k liegt. Bevor ich den wesentlichen Satz über das Verfahren der aktiven Mengen beweise, brauche ich eine einfache Tatsche über konvexe Abbildungen:

Lemma 7.21 *Es seien $I \subseteq \mathbb{R}$ ein Intervall und $f : I \to \mathbb{R}$ eine streng konvexe, zweimal stetig differenzierbare Abbildung. Schließlich sei $x_0 \in I$ ein stationärer Punkt. Dann ist f in $I \cap (-\infty, x_0]$ streng monoton fallend und in $I \cap [x_0, \infty)$ streng monoton wachsend.*

Beweis ÜA

Proposition 7.22 *Es seien $A \in M(n,n)$ positiv definit und $c \in \mathbb{R}^n$. Weiterhin seien $g : \mathbb{R}^n \to \mathbb{R}^p$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ affine Abbildungen. Man definiere $f : \mathbb{R}^n \to \mathbb{R}$ durch*

$$f(x) = \frac{1}{2} x^t A x + c^t x$$

und betrachte das MP

$$\begin{array}{lll} & \min & f(x) \\ \textit{(MP)} & \textit{bez.} & g(x) \leq 0 \\ & & h(x) = 0 \end{array}$$

Schließlich seien $x_0 \in \mathbb{R}^n$ ein zulässiger Punkt und $I_0 \subseteq I(x_0)$. Dann ist das Verfahren der aktiven Mengen wohldefiniert. Wenn im 2. Fall (b) stets $\alpha_k > 0$ gilt, bricht das Verfahren mit einer Lösung ab.

Beweis x_k und I_k seien bestimmt. In Fall 1a) bricht das MP ab, in Fall 1b) sind x_{k+1} und I_{k+1} offenbar wohldefiniert. Also ist in Fall 2 zunächst einmal zu zeigen, dass das MP

$$\begin{array}{lll} & \min & f(x) \\ (\mathrm{MP}_k) & \text{bez.} & g_i(x) = 0 \quad \text{für alle } i \in I_k \\ & & h(x) = 0 \end{array}$$

lösbar ist. Dies gilt nach 7.6. Also bleibt zu zeigen, dass es im Fall 2b) stets ein $r \in \{1, \ldots, n\} \setminus I_k$ gibt mit $g(x_{k+1}) = 0$. Es gilt $g_i(y_k) = 0$ für alle $i \in I_k$ und daher für alle α:

$$g_i(x_k + \alpha(y_k - x_k)) = g_i(x_k) + \alpha(g_i(y_k) - g_i(x_k)) = 0$$

Weiterhin gilt für alle α:

$$h(x_k + \alpha(y_k - x_k)) = h(x_k) + \alpha(h(y_k) - h(x_k)) = 0$$

Da alle g_i stetig sind, gilt $g_i(x_k + \alpha_k(y_k - x_k)) \leq 0$ für alle i. Angenommen, es gilt $g_i(x_k + \alpha_k(y_k - x_k)) < 0$ für alle $i \notin I_k$. Da $y_k = x_k + 1 \cdot (y_k - x_k)$ nicht (MP)-zulässig ist, gilt $\alpha_k < 1$. Da alle g_i stetig sind, gibt es dann ein $\alpha_k < \alpha' \leq 1$

so dass gilt $g_i(x_k + \alpha'(y_k - x_k)) \leq 1$ für alle $i \notin I_k$ also ist $x_k + \alpha'(y_k - x_k)$ ein (MP)-zulässiger Vektor im Widerspruch zur Wahl von α_k.

Ich zeige nun, dass im 2. Fall stets $f(x_{k+1}) < f(x_k)$ gilt:

Im Fall 2a) ist das klar. Im Fall 2b) betrachte man die Abbildung $\varphi : \mathbb{R} \to \mathbb{R}$ definiert durch $\varphi(\alpha) = f(x_k + \alpha(y_k - x_k))$. Diese Abbildung ist streng konvex und nimmt daher ihr Minimum in einem $\alpha_0 \in \mathbb{R}$ an. Nach 7.21 ist φ streng monoton fallend in $(\infty, \alpha_0]$ und streng monoton wachsend in $[\alpha_0, \infty)$. Es folgt $1 \leq \alpha_0$ und daraus wegen $\alpha_k > 0$:

$$f(x_{k+1}) = f(x_k + \alpha_k(y_k - x_k)) = \varphi(\alpha_k) > \varphi(0) = f(x_k)$$

Angenommen, das Verfahren bricht nicht ab, dann gibt es nach der Vorbemerkung eine unendliche Menge $J_0 \subseteq \mathbb{N}$ so dass für alle $k \in J_0$ der Fall 1 gilt. Da es nur endlich viele Teilmengen von $\{1, \ldots, n\}$ gibt, gibt es dann eine unendliche Teilmenge $J \subseteq \{1, \ldots, n\}$ so dass $I_k = I_r$ für alle $k, r \in J$ gilt. Da f streng konvex ist, hat (MP_k) genau eine Lösung. Es folgt $x_k = x_r$ für alle $k, r \in J$. Da J unendlich ist, besagt dies, dass die Folge $(f(x_k))$ konstant wird. Da es aber nach endlich vielen Schritten ein r gibt, so dass der 2. Fall eintritt, folgt $f(x_{r+1}) < f(x_r)$ und daraus ein Widerspruch. ■

Für die Durchführung des Verfahrens der aktiven Mengen sind nun einige Dinge zu bestimmen: Im Fall 1 braucht man Lagrange-Multiplikatoren, im Fall 2 muss man ein MP lösen und im Fall 2b) muss man darüber hinaus α_k bestimmen und r finden. Am Einfachsten kann man α_k und r bestimmen:

Lemma 7.23 *Im Verfahren der aktiven Mengen gilt im Fall 2b)*

$$\alpha_k = \min\{\frac{-g_i(x_k)}{\nabla g_i(x_k)^t(y_k - x_k)} : \nabla g_i(x_k)^t(y_k - x_k) > 0\}$$

Wählt man ein r so dass gilt $\nabla g_r(x_k)^t z_k > 0$ und

$$\frac{-g_r(x_k)}{\nabla g_r(x_k)^t(y_k - x_k)} = \alpha_k$$

dann gilt $r \notin I_k$ und $g_r(x_k + \alpha_k(y_k - x_k)) = 0$.

Beweis Man setze $z_k = y_k - x_k$. Da alle g_i affin sind, folgt aus 6.2 für alle i und α:

$$g_i(x_k + \alpha z_k) = g_i(x_k) + \alpha \nabla g_i(x_k)^t z_k$$

Also gilt für alle i und α

$$g_i(x_k + \alpha z_k) \leq 0 \quad \iff \quad \alpha \nabla g_i(x_k)^t z_k \leq -g_i(x_k)$$

Wegen $-g_i(x_k) \geq 0$ folgt dann für alle $\alpha \geq 0$:

$$g_i(x_k + \alpha z_k) \leq 0 \quad \Longleftrightarrow \quad \alpha \leq \frac{-g_i(x_k)}{\nabla g_i(x_k)^t z_k} \quad \text{für alle } i \text{ mit } \nabla g_i(x_k)^t z_k > 0$$

Daher gilt

$$\alpha_k = \min\{\frac{-g_i(x_k)}{\nabla g_i(x_k)^t z_k} : \nabla g_i(x_k)^t z_k > 0\}$$

Man wähle ein r so dass gilt $\nabla g_r(x_k)^t z_k > 0$ und $\frac{-g_r(x_k)}{\nabla g_r(x_k)^t z_k} = \alpha_k$, dann gilt offenbar $g_r(x_k + \alpha_k z_k) = 0$. Angenommen, $r \in I_k$, dann folgt

$$0 = g_r(x_k + \alpha_k z_k) = g_r(x_k) + \alpha_k \nabla g_r(x_k)^t z_k = \nabla g_r(x_k)^t z_k > 0$$

und daraus ein Widerspruch. Es folgt $r \notin I_k$. ■

Bei der Lösung des k-ten MPs kennt man ja einen zulässigen Punkt, nämlich x_k. Es ist nun üblich, zur Bestimmung der Lösung y_k den Differenzvektor $z_k = y_k - x_k$ zu berechnen:

Lemma 7.24 *Es sei x_0 ein zulässiger Punkt des MPs*

$$\begin{array}{ll} \min & \frac{1}{2} x^t A x + c^t x \\ \textit{bez.} & Bx = d \end{array}$$

Der Punkt $x^* = x_0 + z^*$ ist genau dann ein KTP mit zugehörigem Lagrange-Multiplikator μ^*, wenn (z^*, μ^*) eine Lösung des folgenden LGSs ist:

$$\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix} \begin{pmatrix} z \\ \mu \end{pmatrix} = \begin{pmatrix} -(Ax_0 + c) \\ 0 \end{pmatrix}$$

Beweis Nach 7.17 ist $x_0 + z^*$ genau dann ein KTP mit Lagrange-Multiplikator μ^*, wenn gilt

$$\begin{pmatrix} A & B^t \\ B & 0 \end{pmatrix} \begin{pmatrix} x_0 + z^* \\ \mu^* \end{pmatrix} = \begin{pmatrix} -c \\ d \end{pmatrix}$$

Wegen $Bz^* = d$ ist dies offenbar äquivalent zur Behauptung. ■

Damit erhält man die folgende algorithmische Beschreibung des Verfahrens der aktiven Mengen für quadratische MPe:

Verfahren 7.25 *(der aktiven Mengen) Vorgegeben sei das quadratische MP*

$$\begin{array}{ll} \min & \frac{1}{2}x^tAx + c^tx \\ bez. & Bx \leq d \\ & Dx = f \end{array}$$

(S1) Man wähle einen zulässigen Punkt x_0 und setze $I_0 = I(x_0)$.

(S2) Wenn x_k und I_k bestimmt sind, sei B_k die Matrix, deren Zeilenvektoren gerade $(b_i)_{i\in I_k}$ sind. Es sei (z_k, λ_k, μ_k) eine Lösung des LGS

$$\begin{pmatrix} A & B_k^t & D^t \\ B_k & 0 & 0 \\ D & 0 & 0 \end{pmatrix} \begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} -(Ax_k + c) \\ 0 \\ 0 \end{pmatrix}$$

Fall 1a) Es gilt $z_k = 0$ und $\lambda_k \geq 0$. Dann ist x_k ein KTP und man bricht ab.

Fall 1b) Es gilt $z_k = 0$ aber $\lambda_{k,i} < 0$ für ein $i \in I_k$. Dann wähle man $r \in I_k$ so dass gilt $\lambda_{k,r} = \min\{\lambda_{k,i} : i \in I_k\}$ und setze

$$x_{k+1} = x_k \qquad und \qquad I_{k+1} = I_k \setminus \{r\}$$

Fall 2a) Es gilt $z_k \neq 0$ und $x_k + z_k$ ist zulässig. Dann setze man

$$x_{k+1} = x_k + z_k \qquad und \qquad I_{k+1} = I_k$$

Fall 2b) $x_k + z_k$ ist nicht zulässig. Dann setze man

$$\alpha_k = \min\{\frac{d_i - b_i^t x_k}{b_i^t z_k} : b_i^t z_k > 0\}$$

Es gelte

$$\alpha_k = \frac{d_r - b_r^t x_k}{b_r^t z_k}$$

dann setze man

$$x_{k+1} = x_k + \alpha_k z_k \qquad und\ I_{k+1} = I_k \cup \{r\}$$

Beispiel 7.26 Man betrachte das MP

$$\begin{array}{ll} \min & u^2 - uv + v^2 - 3u \\ \text{bez.} & u \geq 0 \\ & v \geq 0 \\ & u + v \leq 2 \end{array}$$

Man setze

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad \text{und} \quad c = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$$

$$B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{und} \quad d = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

Dann ist das MP

$$\begin{array}{ll} \min & \frac{1}{2}x^t Ax + c^t x \\ \text{bez.} & Bx \leq b \end{array}$$

zu lösen.

$\mathbf{k} = 0$: Es seien

$$x_0 = (0,0)^t \quad \text{und} \quad I_0 = I(x_0) = \{1,2\}$$

$\mathbf{k} = 1$: Es gilt

$$-(Ax_0 + c) = -c = (3,0)$$

also ist zu lösen:

$$\begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ \lambda' \\ \lambda'' \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Die Lösung ist

$$(z_0, \lambda_0) = ((0,0)^t, (-3,0)^t)$$

Also liegt Fall 2b) vor. Es gilt

$$\lambda_{0,1} = \min\{\lambda_{0,i} : i \in I_0\} = -3$$

Also setzt man

$$x_1 = x_0 = (0,0)^t \quad \text{und} \quad I_1 = I_0 \setminus \{1\} = \{2\}$$

$\mathbf{k} = 2$: Es gilt $-(Ax_1 + c) = -(Ax_0 + c) = (3, 0)$ und damit ist zu lösen:

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ \lambda \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$$

Die Lösung ist

$$(z_1, \lambda_1) = ((3/2, 0)^t, -3/2)$$

Weiterhin ist $x_1 + z_1 = (3/2, 0)$ zulässig und daher gilt

$$x_2 = x_1 + z_1 = (3/2, 0) \qquad \text{und } I_2 = I_1 = \{2\}$$

$\mathbf{k} = 3$: Da $x_2 = x_1 + z_1$ eine Lösung von $(\mathrm{MP}_1) = (\mathrm{MP}_2)$ mit Lagrange-Multiplikator $-3/2$ ist, liegt Fall 1b) vor und man setzt $x_3 = x_2$ und $I_3 = I_2 \setminus \{2\} = \emptyset$.

$\mathbf{k} = 4$: Es gilt

$$(Ax_3 + c) = (Ax_2 + c) = (0, -3/2)$$

also ist zu lösen:

$$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 3/2 \end{pmatrix}$$

Die Lösung ist $z_3 = (1/2, 1)$. Weiterhin ist $x_3 + z_3 = (2, 1)$ nicht zulässig. Es gilt

$$\begin{array}{rcl} b_1^t z_3 & = & -1/2 \\ b_2^t z_3 & = & -1 \\ b_3^t z_3 & = & 3/2 \end{array}$$

Also gilt

$$\alpha_3 = \frac{d_3 - b_3^t x_3}{b_3^t z_3} = \frac{2 - 3/2}{3/2} = 1/3$$

und daher

$$x_4 = x_3 + \alpha_3 z_3 = (3/2, 0) + \frac{1}{3}(1/2, 1) = (5/3, 1/3) \qquad \text{und} \qquad I_4 = \{3\}$$

$\mathbf{k} = 5$: Es gilt:

$$(Ax_4 + c) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 5/3 \\ 1/3 \end{pmatrix} + \begin{pmatrix} -3 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

und daher ist zu lösen:

$$\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Dieses hat die Lösung:

$$(z_4, \lambda_4) = ((-1/6, 1/6), 1/2)$$

Weiterhin ist

$$x_4 + z_4 = (5/3, 1/3) + (-1/6, 1/6) = (3/2, 1, 2)$$

zulässig. Daher gilt

$$x_5 = x_4 + z_4 = (3/2, 1/2) \qquad \text{und} \qquad I_5 = I_4 = \{3\}$$

k $= 6$: Da $x_5 = x_4 + z_4$ eine Lösung von $(\mathrm{MP}_4) = (\mathrm{MP}_5)$ mit Lagrange-Multiplikator $1/2$ ist, liegt Fall 1a) vor und das Verfahren bricht mit dem KTP x_5 ab. Da das MP streng konvex ist, ist x_5 die einzige Lösung des MPs. ■

Das Verfahren minimiert also die Funktion zunächst unter der Nebenbedingung $u = v = 0$, was offenbar $x_0 = (0, 0)$ als Lösung liefert. Da der Lagrange-Multiplikator nicht nicht-negativ ist, ist $(0, 0)$ nicht die Lösung des MPs. Danach minimiert es die Funktion unter der Nebenbedingung $v = 0$, die Lösung ist $(3/2, 0)$, auch das ist keine Lösung. Also bestimmt es das Minimum der Funktion ohne Nebenbedingungen, dies wird in $(2, 1)$ angenommen, und dieser Punkt ist nicht zulässig. Daher sucht es auf der Verbindungsstrecke von $(3/2, 0)$ mit $(2, 1)$ den (MP)-zulässigen Punkt, der am nächsten zu $(2, 1)$ liegt. Das ist $(5/3, 1/3)$ und er erfüllt die Nebenbedingung $u + v = 2$. Als sucht das Verfahren das Minimum von f unter der Nebenbedingung $u + v = 2$. Dies wird in $(3/2, 1/2)$ angenommen. Da der Lagrange-Multiplikator nicht-negativ ist, ist das die Lösung des MPs.

Man kann zeigen, dass die Matrix $\begin{pmatrix} B_k \\ D \end{pmatrix}$ für alle k maximalen Rang hat, wenn dies für B_0 gilt. Wenn also A positiv definit ist, ist die Matrix

$$\begin{pmatrix} A & B_k^t & D^t \\ B_k & 0 & 0 \\ D & 0 & 0 \end{pmatrix}$$

nach 7.18 regulär, so dass das lineare Gleichungssystem

$$\begin{pmatrix} A & B_k^t & D^t \\ B_k & 0 & 0 \\ D & 0 & 0 \end{pmatrix} \begin{pmatrix} z \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} -(Ax_k + c) \\ 0 \\ 0 \end{pmatrix}$$

eindeutig lösbar ist.

Exkurs: Lineare Programmierung

Definition 7.27 *Es seien $A \in M(q,n), C \in M(p,n), b \in \mathbb{R}^q, d \in \mathbb{R}^q$ und $c \in \mathbb{R}^n$. Dann heißt das MP*

$$\begin{array}{ll} \min & c^t x \\ \textit{bez.} & Ax = b \\ & Cx \leq d \\ & x \geq 0 \end{array}$$

auch ***Lineares Programm***

Proposition 7.28 *Vorgegeben seien die Linearen Programme*

$$\textit{(MP)} \quad \begin{array}{ll} \min & c^t x \\ \textit{bez.} & Ax = b \\ & Cx \leq d \\ & x \geq 0 \end{array} \quad \textit{und} \quad \textit{(MPe)} \quad \begin{array}{ll} \min & c^t x \\ \textit{bez.} & Ax = b \\ & (C, I_p) \begin{pmatrix} x \\ u \end{pmatrix} = d \\ & x, u \geq 0 \end{array}$$

Dann gelten:

a) Es sei x^* eine Lösung von (MP). Man setze $u^* = d - Cx^*$, dann ist $\begin{pmatrix} x^* \\ u^* \end{pmatrix}$ eine Lösung von (MPe).

b) Es sei $\begin{pmatrix} x^* \\ u^* \end{pmatrix}$ eine Lösung von (MPe), dann ist x^* eine Lösung von (MP).

Man nennt die Komponenten von u auch ***Schlupfvariablen****.*

Beweis

a) Es sei $\begin{pmatrix} x \\ u \end{pmatrix}$ (MPe)-zulässig. Dann gilt $Cx + u = d$ und daher $Cx \leq d$. Also ist x (MP)-zulässig und es gilt $c^t x^* \leq c^t x$.

b) Es s ei x (MP)-zulässig. Man setze $u = d - Cx$, dann ist $\begin{pmatrix} x \\ u \end{pmatrix}$ (MPe)-zulässig und es folgt $c^t x^* \leq c^t x$. ■

Wegen 7.28 reicht es, sich bei der Lösung Linearer Programme auf Programme des Typs

$$\begin{array}{ll} \min & c^t x \\ \text{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

zu beschränken.

Definition 7.29 *Es seien E ein reeller Vektorraum und $K \subseteq E$ konvex. Ein Punkt $x \in K$ heißt* ***Extremalpunkt*** *von K, wenn für alle $y, z \in K$ und $0 < \alpha < 1$*

aus $x = \alpha y + (1-\alpha)z$ stets folgt $x = y = z$, d.h. wenn x nicht im Inneren der Verbindungsgerade von zwei verschiedenen Punkten aus K liegt. Die Menge aller Extremalpunkte von K wird mit $ex\,K$ bezeichnet.

Beispiele 7.30

(i) Für alle $a < b$ gilt

$$ex[a,b] = \{a,b\}$$

und

$$ex(a,b) = \emptyset$$

sowie

$$ex[a,b) = ex[a,\infty) = \{a\}$$

(ii) Es sei

$$S = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

Dann ist S konvex und es gilt

$$ex\,S = \{x \in \mathbb{R}^n : \|x\| = 1\}$$

Beweis Einfach.

Bezeichnungsweisen 7.31 *Es seien $A = (a_1, \ldots, a_n) \in M(q,n)$ und $b \in \mathbb{R}^q$. Dann setze man*

$$K(A,b) = \{x \in \mathbb{R}^n : x \geq 0,\ Ax = b\} = \{x \in \mathbb{R}^n : x \geq 0,\ x_1a_1 + \cdots + x_na_n = b\}$$

Offenbar ist $K(A,b)$ konvex und der zulässige Bereich des LPs

$$\begin{array}{ll} \min & c^tx \\ \text{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

Wie 7.30 auch zeigt, kann die Menge der Extremalpunkte einer Menge leer oder aber auch unendlich sein. Die im Zusammenhang mit der Linearen Programmierung wesentliche Eigenschaft von Extremalpunkten ist die Tatsache, dass $ex\,K(A,b)$ endlich sind und dass ein lösbares LP auch eine Lösung aus $ex\,K(A,b)$ besitzt. Daher ist es prinzipiell möglich, so ein LP in endlich vielen Schritten zu lösen (indem man z.B. alle Extremalpunkte abklappert). Aber um das zu beweisen, muss man ein wenig arbeiten. Ich beginne mit einer Charakterisierung der Extremalpunkte von $K(A,B)$.

Proposition 7.32 *Es seien $A = (a_1, \ldots, a_n) \in M(k,n,\mathbb{R})$ und $b \in \mathbb{R}^k$. Ein Punkt $x \in K(A,b)$ ist genau dann ein Extremalpunkt von $K(A,b)$, wenn die Vektoren $(a_i)_{x_i \neq 0}$ linear unabhängig sind.*

Beweis Es sei $x = (x_1, \ldots, x_n)^t \in K(A,b)$. Ich nehme oBdA an, dass gilt $x_1, \ldots, x_p \neq 0$ und $x_{p+1} = \cdots = x_n = 0$, dann ist zu zeigen, dass genau dann $x \in ex\, K(A,b)$ gilt, wenn $a_1, \ldots, a_p$ linear unabhängig sind.

Es gilt $x \in K(A,b)$ genau dann, wenn gilt

$$x_1 a_1 + \cdots + x_n a_n = b$$

Es gelte zunächst $x \in ex\, K(A,b)$ und $\alpha_1 a_1 + \cdots + \alpha_p a_p = 0$. Man wähle $\varepsilon > 0$ und setze $\beta_i = \varepsilon \alpha_i$. Dann gilt $\beta_1 a_1 + \cdots + \beta_p a_p = 0$. Weiterhin wähle man ε so dass gilt

$$|\beta_i| \leq \min\{x_1, \ldots, x_p\}$$

Setzt man noch $\beta_i = 0$ für $i = p+1, \ldots, n$ und $\beta = (\beta_1, \ldots, \beta_n)^t$, dann gilt $x + \beta \in K(A,b)$ und $x - \beta \in K(A,b)$ und weiterhin $x = \frac{1}{2}(x+\beta) + \frac{1}{2}(x-\beta)$ und daher $x \in [x+\beta, x-\beta]$. Da x ein Extremalpunkt ist, folgt $\beta = 0$ und daraus $\alpha_1 = \cdots = \alpha_p = 0$. Also sind die Vektoren $a_1, \ldots, a_p$ linear unabhängig.

Umgekehrt seien $a_1, \ldots, a_p$ linear unabhängig und es gelte $x = \alpha y + (1-\alpha) z$ mit $0 < \alpha < 1$ und $y, z \in K(A,b)$. Dann folgt

$$x_i = \alpha y_i + (1-\alpha) z_i \qquad \text{für alle } i$$

und daher für alle $i \geq p+1$:

$$0 = \alpha y_i + (1-\alpha) z_i$$

Wegen $y_i, z_i \geq 0$ erhält man dann $y_i = z_i = 0$ für alle $i \geq p+1$. Es folgt

$$y_1 a_1 + \cdots + y_p a_p = y_1 a_1 + \cdots + y_n a_n = b = z_1 a_1 + \cdots + z_n a_n = z_1 a_1 + \cdots + z_p a_p$$

und daraus

$$(y_1 - z_1) a_1 + \cdots + (y_p - z_p) a_p = 0$$

Aus der linearen Unabhängigkeit von $a_1, \ldots, a_p$ folgt dann $y_i = z_i$ für alle $i \leq p$ und daraus $y = z$. Aus $x = \alpha y + (1-\alpha) z$ folgt dann unmittelbar, dass $y = z = x$ gilt. ■

Die Extremalpunkte von $K(A,b)$ nennt man auch **Ecken**.

Korollar 7.33 *$K(A,b)$ besitzt nur endlich viele Extremalpunkte.*

Beweis Für alle $x \in \mathbb{R}^n$ sei

$$I(x) = \{i : x_i \neq 0\}$$

Ich behaupte, dass für alle $x, y \in ex\, K(A,b)$ gilt:

$$I(x) = I(y) \Rightarrow x = y$$

Beweis OBdA gelte $I(x) = \{1, \ldots, p\}$, dann folgt

$$x_1a_1 + \cdots + x_pa_p = x_1a_1 + \cdots + x_na_n = b = y_1a_1 + \cdots + y_na_n = y_1a_1 + \cdots + y_pa_p$$

Nach 7.32 sind die Vektoren $a_1, \ldots, a_p$ linear unabhängig, also folgt $x_i = y_i$ für alle $i \leq p$ und daher für alle i. Da es nur endlich viele Teilmengen von $\{1, \ldots, n\}$ gibt, ist $ex\, K(A, b)$ also ebenfalls endlich. ■

Proposition 7.34 *Es seien $A = (a_1, \ldots, a_n) \in M(k, n, \mathbb{R})$ und $b \in \mathbb{R}^k$. Wenn $K(A, b)$ nicht leer ist, besitzt die Menge einen Extremalpunkt.*

Beweis Für $x \in \mathbb{R}^n$ setze man wieder

$$I(x) = \{i : x_i \neq 0\}$$

Man wähle ein $x \in K(A, b)$, so dass $I(x)$ minimal ist, d.h. es gibt kein $y \in K(A, b)$, so dass gilt $I(y) \subset I(x)$. Ich behaupte, dass $x \in ex\, K(A, b)$ gilt:

Falls $I(x) = \emptyset$ gilt, sind die Vektoren $(x_i)_{i \in I(x)}$ offenbar linear unabhängig.

Also gelte $I(x) \neq \emptyset$ und oBdA gelte $x_1, \ldots, x_p \neq 0$ und $x_{p+1} = \cdots = x_n = 0$. Nach 7.32 muss ich zeigen, dass $a_1, \ldots, a_p$ linear unabhängig sind: Also gelte $\alpha_1a_1 + \cdots + \alpha_pa_p = 0$ und $\alpha_k > 0$ für ein k. Dann folgt $\varepsilon\alpha_1a_1 + \cdots + \varepsilon\alpha_pa_p = 0$ für alle ε und daraus

$$b = x_1a_1 + \cdots + x_na_n = x_1a_1 + \cdots + x_pa_p = (x_1 - \varepsilon\alpha_1)a_1 + \cdots + (x_p - \varepsilon\alpha_p)a_p$$

Setzt man nun

$$\varepsilon = \min\{\frac{x_i}{\alpha_i} : \alpha_i > 0\}$$

und

$$y = (x_1 - \varepsilon\alpha_1, \ldots, x_p - \varepsilon\alpha_p, 0, \ldots, 0)$$

dann gilt $y \in K(A, b)$ und $I(y) \subset I(x)$, also ein Widerspruch. ■

Interessanterweise kann man 7.34 benutzen, um zu beweisen, dass ein lösbares LP auch eine Lösung besitzt, die ein Extremalpunkt des zulässigen Bereichs ist:

Satz 7.35 *Wenn das LP*

$$\begin{array}{ll} \min & c^tx \\ \textit{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

lösbar ist, gibt es einen Punkt $x \in ex\, K(A, b)$, der das LP löst.

Beweis Es seien $x^* \in K(A, b)$ eine Lösung des LPs, dann setze man $\mu^* = c^tx^*$. Es folgt

$$\mu^* \leq c^tx \qquad \text{für alle } x \in K(A, b)$$

Man setze

$$\widetilde{A} = \begin{pmatrix} A \\ c^t \end{pmatrix} \quad \text{und} \quad \widetilde{b} = \begin{pmatrix} b \\ \mu^* \end{pmatrix}$$

Dann gilt

$$\widetilde{A}x^* = \begin{pmatrix} Ax^* \\ c^t x^* \end{pmatrix} = \begin{pmatrix} b \\ \mu^* \end{pmatrix} = \widetilde{b}$$

und daher $x^* \in K(\widetilde{A}, \widetilde{b})$. Also gilt $K(\widetilde{A}, \widetilde{b}) \neq \emptyset$. Nach 7.34 gibt es ein $x_0 \in ex\, K(\widetilde{A}, \widetilde{b})$. Ich zeige, dass $x_0 \in ex\, K(A, b)$ gilt: Es gelte $x_0 = \alpha y + (1-\alpha)z$ mit $0 < \alpha < 1$ und $y, z \in K(A, b)$. Dann folgt

$$\mu^* = c^t x_0 = \alpha c^t y + (1-\alpha) c^t z$$

Wegen $c^t y, c^t z \in [\mu^*, \infty)$ und $\mu^* \in ex[\mu^*, \infty)$ folgt $c^t y = c^t z = \mu^*$ und daraus $y, z \in K(\widetilde{A}, \widetilde{b})$. Da x_0 ein Extremalpunkt von $K(\widetilde{A}, \widetilde{b})$ ist, folgt $y = z = x_0$. ■

Damit haben wir das erste Ziel erreicht: Wenn ein LP lösbar ist, gibt es einen Extremalpunkt, der das LP löst, und es gibt nur endlich viele Extremalpunkte. Diese bekommt man, indem man ein linear unabhängige Spaltenvektoren betrachtet und ein lineares Gleichungssystem löst. Allerdings ist diese Prozedur in dieser Form nur für kleine LPe praktikabel: Nehmen wir an, es gilt $rg(A) = k$, dann gibt es maximal $\binom{n}{k}$ Extremalpunkte, und diese Zahl wird für großes n und nicht so großes k sehr groß. Daher ist es notwendig, ein Verfahren zu finden, dass diese Prozedur möglichst abkürzt.
Das im Folgenden vorgestellte sogannte **Simplexverfahren** hat die merkwürdige Eigenschaft, dass es in der Regel einigermaßen schnell funktioniert, obgleich es im schlechtesten Fall alle Ecken abklappert (und damit natürlich extrem langsam ist). Es ist einer dieser in der Numerik gelegentlich eintretende Fall, dass ein Verfahren "in der Praxis" sehr gut funktioniert, obgleich man nicht genau weiß, warum. Ein typisches weiteres Verfahren dieses Typs ist das Newton-Verfahren. Zur Erklärung des Simplex-Verfahrens beginne ich mit einer

Vorbemerkung 7.36 Es seien wieder $A = (a_1, \ldots, a_n) \in M(k, n, \mathbb{R})$, $b \in \mathbb{R}^k$ und $c \in \mathbb{R}^n$. Man betrachte das LP

$$\begin{array}{ll} \min & c^t x \\ \text{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

und es sei $x \in ex\, K(A, b)$, d.h. x sei eine Ecke. Nach 7.32 sind dann die Vektoren $(a_i)_{x_i \neq 0}$ linear unabhängig und man kann sie zu einer Basis des Spaltenraums $L(\{a_1, \ldots, a_n\})$ ergänzen. Ich nehme an, $\{a_1, \ldots, a_p\}$ sei so eine Basis. Dann gilt $x \geq 0$, $x_i = 0$ für alle $i \geq p+1$ und

$$x_1 a_1 + \cdots + x_p a_p = b$$

Weiter gibt es für alle $j > p$ reelle Zahlen $\alpha_{j,i}$ so dass gilt

$$a_j = \sum_{i=1}^{p} \alpha_{j,i} a_i$$

In einem Simplex-Schritt wird nun x durch eine andere Ecke x' ersetzt, wobei nur ein Basisvektor ausgetauscht wird, d.h. es gibt $1 \leq s \leq p < r \leq n$ so dass gilt $x'_s = 0$ und $x'_i = 0$ für alle $i \geq p+1,\ i \neq r$. Um r und s zu finden betrachte man ein $r \geq p+1$. Weiter sei $y = (y_1, \ldots, y_n)^t \in \mathbb{R}^n$ und es gelte $y_i = 0$ für alle $i \geq p+1, i \neq r$. Dann gilt

$$\sum_{i=1}^{n} y_i a_i = \sum_{i=1}^{p} y_i a_i + y_r a_r = \sum_{i=1}^{p} y_i a_i + y_r (\sum_{i=1}^{p} \alpha_{r,i} a_i) = \sum_{i=1}^{p} (y_i + \alpha_{r,i} y_r) a_i$$

Nun gilt $y \in K(A,b)$ genau dann, wenn gilt $y \geq 0$ und $\sum_{i=1}^{n} y_i a_i = b = \sum_{i=1}^{p} x_i a_i$. Da $\{a_1, \ldots, a_p\}$ eine Basis ist, erhält man

$$y \in K(A,b) \quad \Leftrightarrow \quad y \geq 0 \quad \text{und} \quad x_i = y_i + \alpha_{r,i} y_r \quad \text{für alle } i \leq p$$

also

$$y \in K(A,b) \quad \Leftrightarrow \quad y \geq 0 \quad \text{und} \quad y_i = x_i - \alpha_{r,i} y_r \quad \text{für alle } i \leq p$$

und daher

$$y \in K(A,b) \quad \Leftrightarrow \quad y_r \geq 0 \quad \text{und} \quad y_i = x_i - \alpha_{r,i} y_r \geq 0 \quad \text{für alle } i \leq p$$

Weiterhin gilt:

$$\begin{aligned} c^t y &= \textstyle\sum_{i=1}^{n} c_i y_i \\ &= \textstyle\sum_{i=1}^{p} c_i (x_i - \alpha_{r,i} y_r) + c_r y_r \\ &= \textstyle\sum_{i=1}^{p} c_i x_i - \sum_{i=1}^{p} c_i \alpha_{r,i} y_r + c_r y_r \\ &= c^t x - (\textstyle\sum_{i=1}^{p} c_i \alpha_{r,i} - c_r) y_r \end{aligned}$$

Man setze $\delta_r = \sum_{i=1}^{p} c_i \alpha_{r,i} - c_r$, dann gibt es drei Fälle:

1. Fall $\delta_r \leq 0$
Dann kann man den Wert der Zielfunktion nicht verkleinern, indem man a_r gegen einen der Vektoren $a_1, \ldots, a_p$ austauscht.

2. Fall $\delta_r > 0, \alpha_{r,i} \leq 0$ für alle $1 \leq i \leq p$
Dann gilt $y_i = x_i - \alpha_{r,i} y_r \geq 0$ für alle $y_r \geq 0$ und daher $y \geq 0$ für alle $y_r \geq 0$. Dann gilt aber

$$c^t y = c^t x - \delta_r y_r \longrightarrow_{y_r \to \infty} -\infty$$

und die Zielfunktion ist auf $K(A,b)$ nicht nach unten beschränkt. In diesem Fall ist das LP nicht lösbar.

3. Fall $\delta_r > 0, \alpha_{r,\nu} > 0$ für ein $1 \leq \nu \leq p$
Es sei $y_r \geq 0$, dann gilt

$$\begin{aligned} y \geq 0 &\iff y_i \geq 0 && \text{für alle } i \leq p \\ &\iff x_i - \alpha_{r,i} y_r \geq 0 && \text{für alle } i \leq p \\ &\iff \alpha_{r,i} y_r \leq x_i && \text{für alle } i \leq p \\ &\iff y_r \leq \frac{x_i}{\alpha_{r,i}} && \text{für alle } \alpha_{r,i} > 0 \end{aligned}$$

Setzt man also

$$y_r = \min\{\frac{x_i}{\alpha_{r,i}} : \alpha_{r,i} > 0\}$$

dann gilt $y \geq 0$. Schließlich wähle man ein $1 \leq s \leq p$ so dass gilt $\alpha_{r,s} > 0$ und $y_r = \frac{x_s}{\alpha_{r,s}}$. Man definiere nun $x' \in \mathbb{R}^n$ durch:

$$x'_i = \begin{cases} x_i - \dfrac{\alpha_{r,i}}{\alpha_{r,s}} x_s & 1 \leq i \leq p \\ \dfrac{x_s}{\alpha_{r,s}} & i = r \\ 0 & p+1 \leq i \leq n,\ i \neq r \end{cases}$$

dann gilt $x' \in K(A,b)$. Weiterhin gilt $x'_s = 0$. Wegen $a_r = \sum_{i=1}^{p} \alpha_{r,i} a_i$ und $\alpha_{r,s} \neq 0$, ist

$$\{a_1, \ldots, a_{s-1}, a_r, a_{s+1}, \ldots, a_n\}$$

nach dem Austauschlemma (LA , 4.15) eine Basis des Spaltenraums und es folgt $x' \in ex\, K(A,b)$.

Schließlich gilt

$$c^t x' = c^t x - \delta_r \frac{x_s}{\alpha_{r,s}}$$

Also gilt $c^t x' < c^t x$ genau dann, wenn gilt $x_s > 0$.

Also ist das LP im 2. Fall nicht lösbar, im 3. Fall erhält man in der Regel eine Ecke mit kleinerem Funktionswert der Zielfunktion. Bleibt die Frage, was im 1. Fall los ist, wenn es also nicht möglich ist, dass durch einen Austauschschritt der Wert der Zielfunktion zumindest nicht vergrößert wird. Und das ist nun der springende Punkt des Verfahrens: In diesem Fall ist x eine Lösung des LPs:

Lemma 7.37 *Mit den Bezeichnungen von 7.36 gelte $\delta_j \leq 0$ für alle $p+1 \leq j \leq n$, dann ist x eine Lösung des LPs.*

Beweis Es sei $y \in K(A,b)$, dann gilt

$$\begin{aligned}
\sum_{i=1}^{p} x_i a_i = b = \sum_{i=1}^{n} y_i a_i &= \sum_{i=1}^{p} y_i a_i + \sum_{i=p+1}^{n} y_i (\sum_{j=1}^{p} \alpha_{i,j} a_j) \\
&= \sum_{i=1}^{p} y_i a_i + \sum_{i=p+1}^{n} \sum_{j=1}^{p} y_i \alpha_{i,j} a_j \\
&= \sum_{i=1}^{p} y_i a_i + \sum_{j=p+1}^{n} \sum_{i=1}^{p} y_j \alpha_{j,i} a_i \\
&= \sum_{i=1}^{p} y_i a_i + \sum_{i=1}^{p} \sum_{j=p+1}^{n} y_j \alpha_{j,i} a_i \\
&= \sum_{i=1}^{p} (y_i + \sum_{j=p+1}^{n} y_j \alpha_{j,i}) a_i
\end{aligned}$$

und daher für alle $1 \leq i \leq p$:

$$x_i = y_i + \sum_{j=p+1}^{n} y_j \alpha_{j,i}$$

also

$$y_i = x_i - \sum_{j=p+1}^{n} y_j \alpha_{j,i}$$

Dies ergibt:

$$\begin{aligned}
c^t y = \sum_{i=1}^{n} c_i y_i &= \sum_{i=1}^{p} c_i (x_i - \sum_{j=p+1}^{n} y_j \alpha_{j,i}) + \sum_{i=p+1}^{n} c_i y_i \\
&= \sum_{i=1}^{p} c_i x_i - \sum_{j=p+1}^{n} \sum_{i=1}^{p} c_i y_j \alpha_{j,i} + \sum_{j=p+1}^{n} c_j y_j \\
&= c^t x - \sum_{j=p+1}^{n} (\sum_{i=1}^{p} c_i \alpha_{j,i} - c_j) y_j \\
&= c^t x - \sum_{j=p+1}^{n} \delta_j y_j \\
&\geq c^t x
\end{aligned}$$

■

Nun ist so so, dass in der Regel $\{a_1, \ldots, a_p\}$ nicht die zugehörige Basis zu einer Ecke ist, sondern dass die Basis die Form$\{a_{i_1}, \ldots, a_{i_p}\}$ hat. Aber die entsprechenden Modifikationen sind ganz einfach. Damit kann man einen Simplex-Schritt beschreiben:

Proposition 7.38 *(Simplex-Schritt) Es seien $A = (a_1, \ldots, a_n) \in M(k,n,\mathbb{R})$, $b \in \mathbb{R}^k$ und $c \in \mathbb{R}^n$. Weiter seien $x \in ex\, K(A,b)$ eine Ecke, $\{a_i : i \in I\}$ eine zugehörige Basis, d.h. es gelte $x_i = 0$ für alle $i \notin I$ und $J = \{1, \ldots, n\} \setminus I$. Weiter gelte:*

$$a_j = \sum_{i \in I} \alpha_{j,i} a_i \qquad \text{für alle } j \in J$$

Für alle $j \in J$ setze man

$$\delta_j = \sum_{i \in I} c_i \alpha_{j,i} - c_j$$

Dann gelten:

(i) Wenn $\delta_j \leq 0$ für alle $j \in J$ gilt, ist x eine Lösung des LPs

$$\begin{array}{ll} \min & c^t x \\ \textit{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

(ii) Es gebe ein $j \in J$ mit $\delta_j > 0$ und $\alpha_{j,i} \leq 0$ für alle $i \in I$, dann hat das LP keine Lösung.

(iii) Es gebe ein $r \in J$ und $j \in I$ mit $\delta_r > 0$ und $\alpha_{r,j} > 0$, dann wähle man ein $s \in I$ mit $\alpha_{r,s} > 0$ so dass gilt

$$\frac{x_s}{\alpha_{r,s}} = \min\{\frac{x_i}{\alpha_{r,i}} : i \in I, \alpha_{r,i} > 0\}$$

Man definiere weiterhin $x' \in \mathbb{R}^n$ durch

$$x'_i = \begin{cases} x_i - \frac{\alpha_{r,i}}{\alpha_{r,s}} x_s & i \in I \\ \frac{x_s}{\alpha_{r,s}} & i = r \\ 0 & i \in J, i \neq r \end{cases}$$

Dann gilt $x' \in ex\, K(A,b)$, $I' = \{a_i : i \in I, i \neq s\} \cup \{a_r\}$ ist eine zugehörige Basis und es gilt

$$c^t x' = c^t x - \delta_r \frac{x_s}{\alpha_{r,s}}$$

Zusatz *Setzt man $I' = (I \setminus \{s\}) \cup \{r\},\ J' = \{1, \ldots, n\} \setminus I'$ und*

$$a'_j = \sum_{i \in I'} \alpha'_{j,i} a_i \qquad \textit{für alle } j \in J'$$

sowie

$$\delta'_j = \sum_{i \in I'} c_i \alpha'_{j,i} - c_j$$

Dann gelten für alle $i \in I', j \in J'$:

$$\begin{array}{lcll} \alpha'_{s,r} & = & \frac{1}{\alpha_{r,s}} & \\ \alpha'_{s,i} & = & -\frac{\alpha_{r,i}}{\alpha_{r,s}} & i \neq r \\ \alpha'_{j,r} & = & \frac{\alpha_{j,s}}{\alpha_{r,s}} & j \neq s \\ \alpha'_{j,i} & = & \alpha_{j,i} - \frac{\alpha_{j,s}}{\alpha_{r,s}} \alpha_{r,i} & j \neq s, i \neq r \end{array}$$

sowie

$$\begin{array}{lcll} \delta'_s & = & -\frac{\delta_r}{\alpha_{r,s}} & \\ \delta'_j & = & \delta_j - \frac{\alpha_{j,s}}{\alpha_{r,s}} \delta_r & j \neq s \end{array}$$

Beweis Für alle $j \in J$ gilt

$$a_j = \sum_{i \in I} \alpha_{j,i} a_i$$

also insbesondere

$$a_r = \sum_{i \in I} \alpha_{r,i} a_i = \sum_{i \in I, i \neq s} \alpha_{r,i} a_i + \alpha_{r,s} a_s$$

und daher

$$a_s = \frac{1}{\alpha_{r,s}} (a_r - \sum_{i \in I, i \neq s} \alpha_{r,i} a_i) = - \sum_{i \in I, i \neq s} \frac{\alpha_{r,i}}{\alpha_{r,s}} a_i + \frac{1}{\alpha_{r,s}} a_r$$

Es folgt

$$\alpha'_{s,i} = \begin{cases} \frac{1}{\alpha_{r,s}} & i = r \\ -\frac{\alpha_{r,i}}{\alpha_{r,s}} & i \in I', i \neq r \end{cases}$$

Weiterhin gilt für alle $j \in J$:

$$\begin{aligned} a_j = \textstyle\sum_{i \in I} \alpha_{j,i} a_i &= \textstyle\sum_{i \in I, i \neq s} \alpha_{j,i} a_i + \alpha_{j,s} a_s \\ &= \textstyle\sum_{i \in I, i \neq s} \alpha_{j,i} a_i - \sum_{i \in I, i \neq s} \frac{\alpha_{j,s}}{\alpha_{r,s}} \alpha_{r,i} a_i + \frac{\alpha_{j,s}}{\alpha_{r,s}} a_r \\ &= \textstyle\sum_{i \in I, i \neq s} (\alpha_{j,i} - \frac{\alpha_{j,s}}{\alpha_{r,s}} \alpha_{r,i}) a_i + \frac{\alpha_{j,s}}{\alpha_{r,s}} a_r \end{aligned}$$

Es folgt für alle $j \in J', j \neq s$:

$$\alpha'_{j,i} = \begin{cases} \alpha_{j,i} - \frac{\alpha_{j,s}}{\alpha_{r,s}} \alpha_{r,i} & i \neq r \\ \frac{\alpha_{j,s}}{\alpha_{r,s}} & i = r \end{cases}$$

Schließlich gilt für alle $j \in J', j \neq s$:

$$\begin{aligned} \delta'_j = \sum_{i \in I'} c_i \alpha'_{j,i} - c_j &= \sum_{i \in I', i \neq r} c_i \alpha'_{j,i} + c_r \alpha'_{j,r} - c_j \\ &= \sum_{i \in I', i \neq r} c_i (\alpha_{j,i} - \frac{\alpha_{j,s}}{\alpha_{r,s}} \alpha_{r,i}) + c_r \frac{\alpha_{j,s}}{\alpha_{r,s}} - c_j \\ &= \sum_{i \in I, i \neq s} c_i \alpha_{j,i} - \sum_{i \in I, i \neq s} c_i \frac{\alpha_{j,s}}{\alpha_{r,s}} \alpha_{r,i} + c_r \frac{\alpha_{j,s}}{\alpha_{r,s}} - c_j \\ &= \sum_{i \in I} c_i \alpha_{j,i} - c_s \alpha_{j,s} - \sum_{i \in I} c_i \frac{\alpha_{j,s}}{\alpha_{r,s}} \alpha_{r,i} + c_s \alpha_{j,s} + c_r \frac{\alpha_{j,s}}{\alpha_{r,s}} - c_j \\ &= \sum_{i \in I} c_i \alpha_{j,i} - c_j - \frac{\alpha_{j,s}}{\alpha_{r,s}} (\sum_{i \in I} c_i \alpha_{r,i} - c_r) \\ &= \delta_j - \frac{\alpha_{j,s}}{\alpha_{r,s}} \delta_r \end{aligned}$$

sowie

$$\begin{aligned}
\delta'_s = \sum_{i\in I'} c_i\alpha'_{s,i} - c_s &= -\sum_{i\in I', i\neq r} c_i\frac{\alpha_{r,i}}{\alpha_{r,s}} + \frac{c_r}{\alpha_{r,s}} - c_s \\
&= -\frac{1}{\alpha_{r,s}}\Big(\sum_{i\in I', i\neq r} c_i\alpha_{r,i} + c_s\alpha_{r,s} - c_r\Big) \\
&= -\frac{1}{\alpha_{r,s}}\Big(\sum_{i\in I} c_i\alpha_{r,i} - c_r\Big) \\
&= -\frac{1}{\alpha_{r,s}}\delta_r
\end{aligned}$$

■

Definition 7.39 *Die Hintereinanderausführung mehrerer Simplex-Schritte nennt man das* ***Simplex-Verfahren****.*

Die Ecken des zulässigen Bereichs eines LPs nennt man auch ***Basislösungen****.*

Eine schnelle Konsequenz aus 7.38 ist:

Korollar 7.40 *Wenn beim Simplex-Verfahren keine ausgearteten Ecken entstehen, d.h. wenn für jede Iteration x die Menge $\{a_i : x_i \neq 0\}$ eine Basis des Spaltenraums ist, bricht das Verfahren ab.*

Beweis Wenn Fall (ii) in 7.38 nicht auftritt, wird nach 7.38 der Wert der Zielfunktion in jedem Schritt verkleinert, wenn keine ausgearteten Ecken entstanden sind. Da es nur endlich viele Ecken gibt, bricht das Verfahren ab. ■

Die Tabellarisierung des Simplex-Verfahrens geschieht folgendermaßen: Es seien

$$I = \{i_1, \ldots, i_p\} \qquad \text{und} \qquad J = \{j_1, \ldots, j_q\}$$

Dann betrachtet man das folgende Tableau:

	j_1	j_2	$\ldots$	r	$\ldots$	j_q	
i_1	α_{j_1,i_1}	α_{j_2,i_1}	$\ldots$	α_{r,i_1}	$\cdots$	α_{j_q,i_1}	x_{i_1}
i_2	α_{j_1,i_2}	α_{j_2,i_2}	$\ldots$	α_{r,i_2}	$\cdots$	α_{j_q,i_2}	x_{i_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s	$\alpha_{j_1,2}$	$\alpha_{j_2,s}$	$\ldots$	$\alpha_{r,s}$	$\ldots$	$\alpha_{j_q,s}$	x_s
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i_p	α_{j_1,i_p}	α_{j_2,i_p}	$\ldots$	α_{r,i_p}	$\cdots$	α_{j_q,i_p}	x_{i_p}
	δ_{j_1}	δ_{j_2}	$\ldots$	δ_r	$\ldots$	δ_{j_q}	$c^t x$

Fall 1 Es gilt $\delta_{j_\nu} \leq 0$ für $1 \leq \nu \leq q$, dann ist x eine Lösung.

Fall 2 Andernfalls wähle man ein r mit $\delta_r > 0$.

Fall 2a Es gilt $\alpha_{r,i_\mu} \leq 0$ für alle $1 \leq \mu \leq p$. Dann hat das LP keine Lösung.

Fall 2b Andernfalls ergänze man das Tableau durch die $\frac{x_{i_\mu}}{\alpha_{r,\mu}}$, für die $\alpha_{r,\mu} > 0$ gilt und bestimme s so dass gilt

$$\frac{x_s}{\alpha_{r,s}} = \min\{\frac{x_i}{\alpha_{r,i}} : \alpha_{r,i} > 0\}$$

Man nennt dann $\alpha_{r,s}$ das **Pivot-Element**, die Zeile, in der es steht, die **Pivot-Zeile** und die Spalte, in der es steht, die **Pivot-Spalte**. Das neue Tableau erhält man dann folgendermaßen:

a) Man vertauscht r und s.

b) Man ersetzt $\alpha_{r,s}$ durch $\frac{1}{\alpha_{r,s}}$.

c) Alle weiteren Elemente der Pivot-Zeile ersetzt man durch ihr $\frac{1}{\alpha_{r,s}}$- faches.

d) Alle weiteren Elemente der Pivot-Spalte ersetzt man durch ihr $-\frac{1}{\alpha_{r,s}}$- faches.

e) Alle übrigen Elemente ersetzt man nach der sogenannten Rechteck-Regel:

$$\begin{array}{ll} \alpha_{r,s} & b \\ c & d \end{array}$$

Man ersetzt d durch $d - \frac{bc}{\alpha_{r,s}}$.

Diese Regeln gelten auch für die letzte Zeile und letzte Spalte des Tableaus.

Das Simplex-Verfahren eignet sich in der Regel nicht dafür, mit der Hand gerechnet zu werden. Sollte man es dennoch tun, gibt es einige kleinere Vereinfachungen:

Wenn man das neue Pivot-Element berechnet hat, erhält man die weiteren neuen Elemente der Pivot-Zeile, indem man die alten Elemente mit dem neuen Pivot-Element multipliziert und die weiteren neuen Elemente der Pivot-Spalte, indem man die alten Elemente mit dem inversen des neuen Pivot-Element multipliziert. (Das erspart die erneute Berechnung des Inversen des alten Pivot-Elements.) Man erhält die weiteren Elemente einer beliebigen Zeile, indem man das neue Element dieser Zeile in der Pivot-Spalte mit der alten Pivot-Zeile multipliziert und zu der alten Zeile hinzufügt. (Hier hilft nur Ausprobieren.)

Also ist das Verfahren selbst im Gegensatz zu seiner Herleitung ziemlich einfach!!

Beispiel 7.41 Man betrachte das LP

$$\begin{array}{ll} \min & -30x_1 - 12x_2 \\ \text{bez.} & 3x_1 + x_2 \leq 90 \\ & 2x_1 + x_2 \leq 75 \\ & 4x_1 + 3x_2 \leq 210 \\ & x_1, x_2 \geq 0 \end{array}$$

Durch die Einführung von Schlupfvariablen erhält man das LP:

$$\begin{array}{ll} \min & -30x_1 - 12x_2 \\ \text{bez.} & 3x_1 + x_2 + x_3 = 90 \\ & 2x_1 + x_2 + x_4 = 75 \\ & 4x_1 + 3x_2 + x_5 = 210 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{array}$$

Setzt man also

$$A = \begin{pmatrix} 3 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 4 & 3 & 0 & 0 & 1 \end{pmatrix}$$

und

$$b = (90, 75, 210)^t \quad \text{und} \quad c = (-30, -12, 0, 0, 0)^t$$

dann ist das folgende LP zu lösen:

$$\begin{array}{ll} \min & c^t x \\ \text{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

Offenbar ist $x = (0, 0, 90, 75, 210)^t$ eine Basislösung des LPs und es gelten: $I = \{3, 4, 5\},\ J = \{1, 2\}$ und weiterhin

$$a_1 = (3, 2, 4)^t = 3a_3 + 2a_4 + 4a_5$$

sowie

$$a_2 = (1, 1, 3)^t = 1a_3 + 1a_4 + 3a_5$$

und

$$\begin{array}{rcrcl} \delta_1 & = & c_3\alpha_{1,3} + c_4\alpha_{1,4} + c_5\alpha_{1,5} - c_1 & = & 30 \\ \delta_2 & = & c_3\alpha_{2,3} + c_4\alpha_{2,4} + c_5\alpha_{2,5} - c_2 & = & 12 \end{array}$$

und schließlich

$$c^t x = 0$$

Damit erhält man das folgende Ausgangstableau:

	1	2	
3	3	1	90
4	2	1	75
5	4	3	210
	30	12	0

Man wählt die erste Spalte als Pivot-Spalte und fügt in der letzten Spalte $\frac{x_j}{\alpha_{1,j}}$ für $\alpha_{1,j} > 0$ hinzu:

	1	2		
3	3	1	90	30
4	2	1	75	37.5
5	4	3	210	52.5
	30	12	0	

Also wird die 1. Zeile die Pivot-Zeile. Im folgenden Diagramm ist das Pivot-Element fett geschrieben:

	1	2		
3	**3**	1	90	30
4	2	1	75	37.5
5	4	3	210	52.5
	30	12	0	

Der Simplex-Schritt ergibt dann:

	3	2	
1	1/3	1/3	30
4	−2/3	1/3	15
5	−4/3	5/3	90
	−10	2	−900

Die einzige mögliche Pivot-Spalte ist die zweite, ich ergänze sie:

	3	2		
1	1/3	1/3	30	90
4	−2/3	1/3	15	45
5	−4/3	5/3	90	54
	−10	2	−900	

also wird die 2. Zeile die Pivot-Zeile:

	3	2	
1	1/3	1/3	30
4	−2/3	**1**/3	15
5	−4/3	5/3	90
	−10	2	−900

Der Simplex-Schritt ergibt:

	3	4	
1	1	−1	15
2	−2	3	45
5	2	−5	15
	−6	−6	−990

Also bricht das Simplex-Verfahren mit der Lösung $\hat{x} = (15, 45, 0, 0, 15)$ des modifizierten LP ab und daher löst $x = (15, 45)$ das LP und der minimale Wert der Zielfunktion ist -990.

Um das Simplex-Verfahren starten zu können, braucht man eine Ecke des zulässigen Bereichs. Auch die kann man mit dem Simplex-Verfahren finden. Dabei gibt es einen einfachen Spezialfall, der vor allen Dingen in ökomischen Anwendungen vorliegt.

Bemerkung 7.42

(i) Vorgegeben sei das LP

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

mit $b \geq 0$.

Durch Einführung von Schlupfvariablen erhält man das LP

$$\begin{array}{ll} \min & c^t x \\ \text{bez.} & Ax + u = b \\ & x, u \geq 0 \end{array}$$

Setzt man nun

$$\widetilde{A} = (A, I) = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} & 1 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{k,1} & \dots & a_{k,n} & 0 & \dots & 1 \end{pmatrix} = (a_1, \dots, a_n, e_1, \dots, e_k)$$

und

$$\tilde{c} = (c_1, \dots, c_n, 0 \dots, 0)^t$$

dann ist das LP

$$\begin{array}{ll} \min & \tilde{c}^t \begin{pmatrix} x \\ u \end{pmatrix} \\ \text{bez.} & \widetilde{A} \begin{pmatrix} x \\ u \end{pmatrix} = b \\ & \begin{pmatrix} x \\ u \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{array}$$

zu lösen. Wegen $b \geq 0$ ist $\tilde{b} = \begin{pmatrix} 0 \\ b \end{pmatrix}$ ein für dieses LP zulässiger Vektor und $(e_1, \dots, e_k)$ eine zugehörige Basis des Spaltenraums von $\widetilde{A}$. Also gilt in der Tat $b \in ex\, K(\widetilde{A}, b)$ und daher ist b ein zulässiger Startvektor für das Simplex-Verfahren (für das zweite LP).

Die Bestimmung der weiteren Werte für das Anfangstableau ist nun ganz einfach: Setzt man $a_{n+i} = e_i$, dann erhält man für alle $1 \leq j \leq k$:

$$a_j = a_{1,j} e_1 + \cdots a_{k,j} e_k = a_{1,j} a_{n+1} + \cdots a_{k,j} a_{n+k}$$

und daher

$$\alpha_{j,n+i} = a_{i,j} \qquad \text{für alle } 1 \leq j \leq n,\ 1 \leq i \leq k$$

Weiterhin gilt

$$\delta_j = \sum_{i=1}^{k} \alpha_{j,k+i}\tilde{c}_{k+i} - c_j = -c_j$$

sowie

$$\tilde{c}^t\tilde{b} = 0$$

Also erhält man als Anfangstableau:

$$\begin{array}{c|ccc|c} & 1 & \dots & n & \\ \hline n+1 & a_{1,1} & \dots & a_{1,n} & b_1 \\ \vdots & \vdots & & \vdots & \vdots \\ n+k & a_{k,1} & \dots & a_{k,n} & b_k \\ \hline & -c_1 & \dots & -c_n & 0 \end{array} \quad = \quad \begin{array}{c|ccc|c} & 1 & \dots & n & \\ \hline n+1 & & & & \\ \vdots & & A & & b \\ n+k & & & & \\ \hline & & -c & & 0 \end{array}$$

(ii) Im allgemeinen Fall muss man einen Eckpunkt von $K(A,b)$ finden. Indem man die Zeilen gegebenenfalls mit -1 multipliziert, kann man hier oBdA annehmen, dass $b \geq 0$ gilt. Vorgegeben sei also das LP

$$\begin{array}{lll} & \min & c^t x \\ \text{(LP)} & \text{bez.} & Ax = b \\ & & x \geq 0 \end{array}$$

mit $b \geq 0$. Man betrachte das LP

$$\begin{array}{lll} & \min & y_1 + \cdots + y_k \\ \text{(LP')} & \text{bez.} & Ax + y = b \\ & & x \geq 0 \end{array}$$

Setzt man

$$\hat{A} = (A, I_k)$$

dann hat dieses LP die Form

$$\begin{array}{ll} \min & y_1 + \cdots + y_k \\ \text{bez.} & \hat{A}\begin{pmatrix} x \\ y \end{pmatrix} = b \\ & x, y \geq 0 \end{array}$$

Wenn (LP) einen zulässigen Vektor x_0 besitzt, ist $\begin{pmatrix} x_0 \\ 0 \end{pmatrix}$ eine Lösung von (LP') mit dem optimalen Wert der Zielfunktion 0. Wenn umgekehrt $z^* = \begin{pmatrix} x^* \\ y^* \end{pmatrix}$ eine Lösung von (MP') ist und $y^* \neq 0$ gilt, hat (LP) keinen zulässigen Punkt. Wenn andererseits $y^* = 0$ gilt, ist x^* eine Ecke von (LP), denn

$$(a_i)_{x_i^* \neq 0} = (\hat{a}_i)_{z_i^* \neq 0}$$

Weiterhin gilt $\begin{pmatrix} 0 \\ b \end{pmatrix} \in ex\, K(\hat{A}, b)$, d.h. $\begin{pmatrix} 0 \\ b \end{pmatrix}$ ist eine Ecke von $K(\hat{A}, b)$, so dass man das Simplex-Verfahren auf dieses LPs anwenden kann. Schließlich sei $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ eine Ecke, die dieses LP löst, dann gilt $y_0 = 0$ und $x_0 \in ex\, K(A, b)$.

Die Frage der Lösbarkeit eines Linearen Programms kann man mit 7.5 ohne Probleme beantworten:

Satz 7.43 *Es seien $A \in M(k, n, \mathbb{R}), b \in \mathbb{R}^k$ und $c \in \mathbb{R}^n$. Vorgegeben sei das LP*

$$\begin{array}{ll} \min & c^t x \\ \textit{bez.} & Ax \leq b \\ & x \geq 0 \end{array}$$

Wenn der zulässige Bereich nicht leer ist und die Zielfunktion auf dem zulässigen Bereich nach unten beschränkt ist, ist das LP lösbar.

Korollar 7.44 *Vorgegeben sei das LP*

$$\begin{array}{ll} \min & c^t x \\ \textit{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

Wenn der zulässige Bereich nicht leer ist und die Zielfunktion auf dem zulässigen Bereich nach unten beschränkt ist, ist das LP lösbar.

Beweis Man betrachte das LP:

$$\begin{array}{ll} \min & c^t x \\ \text{bez.} & Ax \leq b \\ & -Ax \leq -b \\ & x \geq 0 \end{array}$$

Dann sind der zulässiger Bereich und die Ziefunktion beider LPs gleich, so dass die Behauptung aus 7.43 folgt. ■

Nachtrag Natürlich ist es mathematisch nicht befriedigend, dass das Simplex-Verfahren nicht in jedem Fall nach endlich vielen Schritten abbricht. Um dieses Problem zu vermeiden, gibt es eine Reihe von Zusatzregeln, die verhindern, dass das Verfahren im Kreis läuft. Das folgende "lexikographische" Simplex-Verfahren ist so eins:

Zunächst definiert man auf $\mathbb{R}^n$ die sogenannte **lexikographische Ordnung**:

Für alle $x = (x_1, \ldots, x_n)^t,\ y = (y_1, \ldots, y_n)^t \in \mathbb{R}^n$ definiere man $x \prec y$ genau dann, wenn es ein $i_0 \in \{1, \ldots n\}$ so gibt, dass gilt

$$x_i = y_i \text{ für alle } i < i_0 \qquad \text{und} \qquad x_{i_0} < y_{i_0}$$

sowie

$$x \preceq y \quad \Leftrightarrow \quad x = y \text{ oder } x \prec y$$

Man sieht leicht, dass $\preceq$ eine Ordnung auf $\mathbb{R}^n$ ist, die linear ist, d.h. für je zwei Elemente $x, y \in \mathbb{R}^n$ gilt $x \preceq y$ oder $y \preceq x$.

Nun sei das LP

$$\begin{array}{ll} \min & c^t x \\ \text{bez.} & Ax = b \\ & x \geq 0 \end{array}$$

vorgegeben und oBdA sei $\{a_1, \ldots, a_p\}$ eine Basis des Spaltenraums von A. Wenn man nun mit dem Simplex-Verfahren eine Basis $\{a_i : i \in I\}$ konstruiert hat, gibt es reelle Zahlen $\alpha_{j,i}, i \in I, j \in \{1, \ldots, n\}$ so dass für **alle** $j \in \{1, \ldots, n\}$ gilt

$$b_j = \sum_{i \in I} \alpha_{j,i} b_i$$

Weiterhin setzt man für **alle** $j \in \{1, \ldots, n\}$:

$$\delta_j = \sum_{i \in I} \alpha_{j,i} c_i - \ c_j$$

(Beachten Sie, dass das keine wesentlich Erweiterung der bisherigen Definition ist.) Da $a_1, \ldots, a_p$ linear unabhängig sind, sind die Vektoren der Form

$$\frac{1}{\alpha_{r,i}}(x_i, \alpha_{1,i}, \ldots, \alpha_{p,i})$$

paarweise verschieden. Nun gelte wieder $\delta_r > 0$ für ein $r \notin I$, dann gibt es also ein eindeutig bestimmtes s so dass gilt $\alpha_{r,s} > 0$ und

$$\frac{1}{\alpha_{r,s}}(x_s, \alpha_{1,s}, \ldots, \alpha_{p,s}) \prec \frac{1}{\alpha_{r,i}}(x_i, \alpha_{1,i}, \ldots, \alpha_{p,i}) \qquad \text{für alle } i \in I \setminus \{s\} \text{ mit } \alpha_{r,i} > 0$$

Nach der **Zusatzregel** wählt man dieses s. Man beachte, dass die Zusatzregel nur in Kraft tritt, wenn es zwei mögliche Wahlen für die Pivot-Spalte gibt.

Man kann zeigen, dass gilt

$$(c^t x', \delta'_1, \ldots \delta'_p) \prec (c^t x, \delta_1, \ldots \delta_p)$$

d.h. dieser Vektor wird bei einem Simplex-Schritt immer (lexikographisch) echt verkleinert. Also kann keine Ecke zweimal beim Simplex-Verfahren auftauchen. Da es nur endlich viele Ecken gibt, bricht das Simplex-Verfahren mit Zusatzregel also ab.

Kapitel 8

SQP-Verfahren

Es seien $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ stetig differenzierbare Abbildungen. Wenn x^* eine reguläre Lösung des MPs

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & h(x) = 0 \end{array}$$

ist, ist x^* ein KTP, also gibt es ein μ^* so dass gelten

(i) $\nabla f(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0$

(ii) $h(x^*) = 0$

Nun gilt ja $\nabla h(x) = (\nabla h_1(x), \ldots, \nabla h_q(x))$ und daher $\nabla h(x)\mu = \sum \mu_j \nabla h_j(x)$ und daher ist x^* genau dann ein KTP, wenn es ein μ^* so gibt, dass gelten

(i) $\nabla f(x^*) + \nabla h(x^*)\mu^* = 0$

(ii) $h(x^*) = 0$

Betrachtet man wieder die Lagrangefunktion L definiert durch

$$L(x, \mu) = f(x) + \sum_{j=1}^{q} \mu_j h_j(x)$$

dann gilt für alle x, μ:

$$\nabla_x L(x, \mu) = \nabla f(x) + \sum \mu_j \nabla h_j(x) = \nabla f(x) + \nabla h(x)\mu$$

und

$$\nabla_\mu L(x, \mu) = h(x)$$

Also ist x^* genau dann ein KTP, wenn es ein μ^* so gibt, dass gilt

$$\nabla L(x^*, \mu^*) = \begin{pmatrix} \nabla_x L(x^*, \mu^*) \\ \nabla_\mu L(x^*, \mu^*) \end{pmatrix} = 0 \, .$$

Wenn nun L zweimal stetig differenzierbar ist, und $H(x^*, \mu^*)$ regulär ist, kann man eine Nullstelle von ∇L mit dem Newton-Verfahren suchen.

Wenn also (x_k, μ_k) bestimmt ist, berechnet man (x_{k+1}, μ_{k+1}) nach 2.25 aus der Formel

$$HL(x_k, \mu_k)((x_{k+1}, \mu_{k+1}) - (x_k, \mu_k)) = -\nabla L(x_k, \mu_k)$$

Nun gilt für alle (x, μ):

$$HL(x, \mu) = \begin{pmatrix} H_x L(x, \mu) & \nabla h(x) \\ \nabla h(x)^t & 0 \end{pmatrix}$$

Also gilt

$$\begin{pmatrix} H_x L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^t & 0 \end{pmatrix} \begin{pmatrix} x_{k+1} - x_k \\ \mu_{k+1} - \mu_k \end{pmatrix} = -\begin{pmatrix} \nabla_x L(x_k, \mu_k) \\ h(x_k) \end{pmatrix}$$

$$= -\begin{pmatrix} \nabla f(x_k) + \nabla h(x_k)\mu_k \\ h(x_k) \end{pmatrix}$$

und daher

$$H_x L(x_k, \mu_k)(x_{k+1} - x_k) + \nabla h(x_k)(\mu_{k+1} - \mu_k) = -\nabla f(x_k) - \nabla h(x_k)\mu_k$$
$$\nabla h(x_k)^t(x_{k+1} - x_k) = -h(x_k)$$

Dies ist nun äquivalent zu:

$$\nabla f(x_k) + H_x L(x_k, \mu_k)(x_{k+1} - x_k) + \nabla h(x_k)\mu_{k+1} = 0$$
$$h(x_k) + \nabla h(x_k)^t(x_{k+1} - x_k) = 0$$

Also ist (x_{k+1}, μ_{k+1}) genau dann die Newton-Iteration von (x_k, μ_k), wenn x_{k+1} ein KT-Punkt des quadratischen MPs

$$\begin{array}{ll} \min & \frac{1}{2}(x - x_k)^t H_x L(x_k, \mu_k)(x - x_k) + \nabla f(x_k)^t(x - x_k) \\ \text{bez.} & h(x_k) + \nabla h(x_k)^t(x - x_k) = 0 \end{array}$$

mit zugehörigem Lagrange-Multiplikator μ_{k+1} ist. Also bietet sich das folgende Verfahren an, dabei steht "SQP" für "sequential quadratic programming":

Verfahren 8.1 *(lokales SQP-Verfahren) Es seien $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ zweimal stetig differenzierbare Abbildungen. Man betrachte das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & h(x) = 0 \end{array}$$

(S1) Man wähle $x_0 \in \mathbb{R}^n$ und $\mu_0 \in \mathbb{R}^q$ beliebig.

(S2) Es seien $x_0, \ldots, x_k, \mu_0, \ldots, \mu_k$ definiert. Dann sei x_{k+1} ein KT-Punkt des quadratischen MPs

$$(MP_k) \qquad \begin{array}{ll} \min & \frac{1}{2}(x-x_k)^t H_x L(x_k,\mu_k)(x-x_k) + \nabla f(x_k)^t(x-x_k) \\ \textit{bez.} & \nabla h(x_k)^t(x-x_k) + h(x_k) = 0 \end{array}$$

und μ_{k+1} ein zugehöriger Lagrange-Multiplikator.

Bemerkung 8.2 In vielen Varianten des lokalen SQP-Verfahrens ersetzt man $H_x(x_k, \mu_k)$ durch eine Matrix H_k, die "in der Nähe" von $H_x(x_k, \mu_k)$ liegt. Dies ist besonders dann nützlich, wenn $H_x(x_k, \mu_k)$ schwer zu berechnen ist.

Lemma 8.3 *Es seien $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ zweimal stetig differenzierbar und x^* ein regulärer KTP des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & h(x) = 0 \end{array}$$

mit Lagrange-Multiplikator μ^*. Weiterhin sei $H_x L(x^*, \mu^*)$ positiv definit auf $\mathcal{Z}(x^*)$. Dann ist $HL(x^*, \mu^*)$ regulär.

Beweis Es gilt

$$HL(x^*, \mu^*) = \begin{pmatrix} H_x L(x^*, \mu^*) & \nabla h(x^*) \\ \nabla h(x^*)^t & 0 \end{pmatrix}$$

Aus $HL(x^*, \mu^*) \begin{pmatrix} u \\ v \end{pmatrix} = 0$ folgt

$$\begin{aligned} H_x L(x^*, \mu^*)u + \nabla h(x^*)v &= 0 \\ \nabla h(x^*)^t u &= 0 \end{aligned}$$

Dies impliziert $u \in \mathcal{Z}(x^*)$ und aus

$$u^t H_x L(x^*, \mu^*)u = u^t H_x L(x^*, \mu^*)u + u^t \nabla h(x^*) = 0$$

folgt $u = 0$ und daraus $v = 0$, da x^* regulär ist. ■

Proposition 8.4 *Es seien $f : \mathbb{R}^n \to \mathbb{R}$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ zweimal stetig differenzierbare Abbildungen und x^* ein regulärer KTP des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & h(x) = 0 \end{array}$$

mit zugehörigem Lagrange-Multiplikator μ^*. Weiterhin seien $H_x L(x^*, \mu^*)$ positiv definit auf $\mathcal{Z}(x^*)$. Dann gibt es ein $r > 0$ so dass für alle $(x_0, \mu_0) \in B((x^*, \mu^*), r])$ gilt:

Das SQP-Verfahren mit dem Startpunkt (x_0, μ_0) ist wohldefiniert und konvergiert quadratisch gegen (x^*, μ^*).

Beweis x_{k+1} ist genau dann ein KTP des MPs

$$\begin{array}{ll} \min & \frac{1}{2}(x-x_k)^t H_x L(x_k,\mu_k)(x-x_k) + \nabla f(x_k)^t(x-x_k) \\ \text{bez.} & \nabla h(x_k)^t(x-x_k) + h(x_k) = 0 \end{array}$$

mit zugehörigem Lagrange-Multiplikator μ_{k+1}, wenn gilt

$$\begin{aligned} HL(x_k,\mu_k)\begin{pmatrix} x_{k+1}-x_k \\ \mu_{k+1} \end{pmatrix} &= \begin{pmatrix} H_x L(x_k,\mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^t & 0 \end{pmatrix}\begin{pmatrix} x_{k+1}-x_k \\ \mu_{k+1} \end{pmatrix} \\ &= \begin{pmatrix} -\nabla f(x_k) \\ -h(x_k) \end{pmatrix} \end{aligned}$$

Da $HL(x^*,\mu^*)$ nach 8.3 regulär ist, gibt es ein $r > 0$ so dass $HL(x,\mu)$ für alle $(x,\mu) \in B((x^*,\mu^*),r]$ regulär ist. Also hat das MP genau einen KTP und dieser besitzt genau einen Lagrange-Multiplikator. Daher ist die SQP-Iteration die Newton-Iteration. ■

Das SQP-Verfahren erhält man, indem man anstelle des MPs

$$\begin{array}{ll} \min & f(x) \\ \text{bez.} & h(x) = 0 \end{array}$$

das MP

$$\begin{array}{ll} \min & \frac{1}{2}w^t H_x L(x_k,\mu_k)w + \nabla f(x_k)^t w \\ \text{bez.} & h(x_k) + \nabla h(x_k)^t w = 0 \end{array}$$

löst. Nun ist die Abbildung

$$w \mapsto \frac{1}{2}w^t H_x L(x_k,\mu_k)w + \nabla f(x_k)^t w$$

die quadratische Approximation von f in x_k und die Abbildung

$$w \mapsto h(x_k) + \nabla h(x_k)^t w$$

ist die lineare (affine) Approximation von h in x_k. Diese Beobachtung erlaubt es nun, ein analoges Verfahren für MPe mit Ungleichungsrestriktionen zu definieren, das wieder SQP-Verfahren heißt.

Verfahren 8.5 *(lokales SQP-Verfahren) Es seien $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ und $h : \mathbb{R}^n \to \mathbb{R}^q$ zweimal stetig differenzierbare Abbildungen. Man betrachte das MP*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

Das folgende Verfahren heißt auch lokales SQP-Verfahren:

(S1) Man wähle $x_0 \in \mathbb{R}^n,\ \lambda_0 \in \mathbb{R}^p,\ \lambda_0 \geq 0$ *und* $\mu_0 \in \mathbb{R}^q$ *beliebig.*

(S2) Es seien $x_0, \ldots, x_k$ *und* $\lambda_0, \ldots, \lambda_k, \mu_0, \ldots, \mu_k$ *definiert. Wenn* x_k *ein KT-Punkt des MPs ist, bricht man ab. Sonst sei* x_{k+1} *ein KTP des quadratischen MPs*

$$(MP_k) \qquad \begin{array}{ll} \min & \nabla f(x_k)^t(x - x_k) + \frac{1}{2}(x - x_k)^t H_x L(x_k, \lambda_k, \mu_k)(x - x_k) \\ \textit{bez.} & g(x_k) + \nabla g(x_k)^t(x - x_k) \leq 0 \\ & h(x_k) + \nabla h(x_k)^t(x - x_k) = 0 \end{array}$$

mit zugehörigen Lagrange-Multiplikator $(\lambda_{k+1}, \mu_{k+1})$.

Der Beweis der Konvergenz des lokalen SQP-Verfahrens 8.5 soll wieder auf das Newton-Verfahren zurückgeführt werden. Dazu muss man KT-Punkte mit Hilfe von Gleichungen beschreiben. Das ist aber ziemlich einfach:

Bemerkung 8.6 *Ein Punkt* (x^*, λ^*, μ^*) *ist genau dann ein KT-Punkt des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

mit den Lagrange-Multiplikatoren λ^* *und* μ^**, wenn gelten:*

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$
$$\min\{-g(x^*), \lambda^*\} = 0$$
$$h(x^*) = 0$$

Dabei setzt man

$$\min\{-g(x), \lambda\} = (\min\{-g_i(x), \lambda_i\})_{i=1,\ldots,p}$$

Offenbar ist die Abbildung

$$(x, \lambda) \mapsto \min\{-g(x), \lambda\}$$

in der Regel nicht differenzierbar, so dass die Beschreibung in der in 8.6 gegebenen Form nicht immer anwendbar ist. Hier hilft:

Lemma 8.7 *Es seien* $g : \mathbb{R}^n \to \mathbb{R}^p$ *(zweimal) stetig differenzierbar,* $x^* \in \mathbb{R}^n$ *und* $\lambda^* \in \mathbb{R}^p$*. Man definiere* $\varphi : \mathbb{R}^n \to \mathbb{R}^p$ *durch*

$$\varphi(x, \lambda) = \min\{-g(x), \lambda\}$$

Es gelte $g_i(x^*) + \lambda_i^* \neq 0$ *für alle* i*. Dann gibt es ein* $r > 0$ *so dass* φ *in* $B((x^*, \lambda^*), r)$ *(zweimal) stetig differenzierbar ist.*

Beweis Es sei $i \in \{1, \ldots, p\}$, dann gilt

$$g_i(x^*) + \lambda_i^* > 0 \qquad \text{oder} \qquad g_i(x^*) + \lambda_i^* < 0$$

Also gibt es ein $r_i > 0$ so dass gilt

$$g_i(x) + \lambda_i > 0 \qquad \text{für alle } (x, \lambda) \in B_i = B(x^*, \lambda^*), r_i)$$

oder

$$g_i(x) + \lambda_i < 0 \qquad \text{für alle } (x, \lambda) \in B_i = B(x^*, \lambda^*), r_i)$$

Im ersten Fall folgt $\varphi_i(x, \lambda) = \lambda_i$ für alle $(x, \lambda) \in B_i$ und im zweiten Fall folgt $\varphi_i(x, \lambda) = -g_i(x)$ für alle $(x, \lambda) \in B_i$. Setzt man noch $r = \min\{r_1, \ldots, r_p\}$, dann ist φ in $B((x^*, \lambda^*), r)$ (zweimal) stetig differenzierbar. ■

Proposition 8.8 *Es seien $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ und und $h : \mathbb{R}^n \to \mathbb{R}^q$ zweimal stetig differenzierbare Abbildungen und x^* ein regulärer KTP des MPs*

$$\begin{array}{ll} \min & f(x) \\ \textit{bez.} & g(x) \leq 0 \\ & h(x) = 0 \end{array}$$

mit zugehörigen Lagrange-Multiplikatoren λ^*, μ^*. Weiterhin gelten:

(i) $g_i(x^*) + \lambda_i^* \neq 0$ *für alle i*

(ii) $H_x L(x^*, \lambda^*, \mu^*)$ *ist positiv definit auf*

$$\mathcal{G}(x^*, \lambda^*) = \{d \in \mathcal{Z}(x^*) : d^t \nabla g_i(x^*) = 0 \text{ für alle } i \text{ mit } \lambda_i^* > 0\}$$

(iii) Falls (MP_k) mehrerer KTPe besitzt, sei $(x_{k+1}, \lambda_{k+1}, \mu_{k+1})$ so gewählt, dass $\|(x_{k+1}, \lambda_{k+1}, \mu_{k+1}) - (x_k, \lambda_k, \mu_k)\|$ minimal ist

Dann gibt es ein $r > 0$ so dass gilt: Für alle $(x_0, \lambda_0, \mu_0) \in B((x^*, \lambda^*, \mu^*), r]$ ist das Verfahren aus 8.5 wohldefiniert und die Folge (x_k, λ_k, μ_k) konvergiert quadratisch gegen (x^*, λ^*, μ^*).

Beweisidee Man definiere $\Phi : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ durch

$$\Phi(x, \lambda, \mu) = \begin{pmatrix} \nabla_x L(x, \lambda, \mu) \\ \min\{-g(x), \lambda\} \\ h(x) \end{pmatrix}$$

Dann ist (x^*, λ^*, μ^*) eine Nullstelle von Φ und Φ ist nach 8.7 in einer Kugel um (x^*, λ^*, μ^*) stetig differenzierbar. Also reicht es zu zeigen, dass das lokale SQP-Verfahren in der Tat das Newton-Verfahren für Φ ist.

Den Beweis findet man z.B. in Geiger/Kantzow "Theorie und Numerik restringierter Optimierungsaufgaben", 5.31.

Ich möchte die Vorlesung mit zwei Zitaten beenden, die mir besonders bemerkenswert erscheinen:

"Nonetheless, it must be appreciated that the existence of convergence and order of convergence results for any algorithm is not a guarantee of good performance in practice. Not only do the results themselves fall short of a guarantee of acceptable behaviour, but also they neglect computer round-off errors which can be crucial. Often the results impose certain restrictions on the function which may not be easy to verify, and in some cases (for example when it is assumed to be a *convex* function) these conditions may not be satisfied in practice. Thus the development of an optimization method also relies on *experimentation*. That is to say, the algorithm is shown to have acceptable behaviour on a variety of *test functions* which should be chosen to represent the different features which might arise in general (insofar as this is possible). Clearly experimentation can never give a guarantee of good performanence in the sense of a mathematical proof. My experience however is that well-chosen experimental testing is often the most reliable indication of good performance. The ideal of course is a good selection of experimental testing backed up by convergence and order of convergence proofs."

(R. Fletcher: Practical methods of optimization)

"From the viewpoint of numerical processing of a minimization problem, there exista a "solvable case" - the one of convex optimization problems, those where the domain is a closed convex subset of $\mathbb{R}^n$ and the objective function and the constraints are convex functions. (...)
In contrast to this, general-type nonconvex problems are too difficult for numerical solutions; the computational effort required to solve such a problem, by the best numerical methods known, grows prohibitively fast with the dimensions of the problem and the number of accuracy digits. Moreover, there are serious theoretical reasons to conjecture that this is an intrinsic feature of nonconvex problems rather than a drawback of the existing optimization techniques."

(Ben-Tal, Nemirowski: Lectures on modern convex optimization)

Zu dem 2. Zitat ist natürlich anzumerken, dass es von Autoren eines Buches über *konvexe* Optimierung stammt.

Literatur

- Geiger, Carl und Kanzow, Christian: Theorie und Numerik restringierter Optimierungsaufgaben
- Geiger, Carl und Kanzow, Christian: Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben
- Mangasarian, Olvi L.: Nonlinear programming
- Blum, Eugen; Oettli, Werner: Mathematische Optimierung : Grundlagen u. Verfahren
- Collatz, Lothar; Wetterling, Wolfgang: Optimierungsaufgaben
- Fletcher, Roger: Practical methods of optimization
- Spellucci, Peter: Numerische Verfahren der nichtlinearen Optimierung
- Ben-Tal, Aharon; Nemirovskij, Arkadij S.: Lectures on modern convex optimization : analysis, algorithms, and engineering applications
- Nesterov, Yurii und Nemirowskii, Arkadii: Interior-Point Polynomial Algorithms in Convex Programming

Korrekturen

Seite, Zeile	Fehler	Korrektur	Datum
34, -2	$\gamma \leq \gamma + \varepsilon$	$\gamma < \gamma + \varepsilon$	02.11.2010
35, 1	3.18	3.19	02.11.2010
35, 2	3.17	3.18	02.11.2010
36, 11	Für alle $a \in \mathcal{A}$	Für alle $A \in \mathcal{A}$	02.11.2010
36, 17	3.21	3.22	02.11.2010
36, -5	3.22	3.23	02.11.2010
36, -4	3.20	3.21	02.11.2010
37, 8	3.23	3.24	02.11.2010
38, 6	"(i) $\Rightarrow$ (ii)"	"(ii) $\Rightarrow$ (i)"	02.11.2010
38, 9	3.20	3.21	02.11.2010
38, 14	(ii)	(i)	02.11.2010
38, 15	"(i) $\Rightarrow$ (ii)"	"(ii) $\Rightarrow$ (i)"	02.11.2010
54, 7	$(x_i - x^*)$	$(x_k - x^*)$	02.11.2010
55, 4	$d^t b_j$	$b_j^t d$	02.11.2010
61, 13	(ii)	(iii)	02.11.2010
67, 14	$\sum \lambda_i f_i(x) + \sum \mu_j g_j(x)$	$\sum \lambda_i g_i(x) + \sum \mu_j h_j(x)$	02.11.2010
71-73	Ich habe 6.6 hinter 6.3 gestellt. Damit ergibt sich die Umbenennung: 6.6 –> 6.4, 6.4 –> 6.5, 6.5 –> 6.6 Außerdem ist 6.7 jetzt ein Satz.		04.11.2010
12, -9	$\frac{\cdots}{t}$	$\frac{\cdots}{\alpha}$	05.11.2010
13, 7	$\alpha_k = \delta^{j_k}$	$\alpha_k = \beta^{j_k}$	05.11.2010
23,8	$f(x^*)$	x^*	05.11.2010
40, -8	$\leq f(x) + (1-\alpha)f(y)$	$\leq f(x) + \alpha(f(y) - f(x))$	18.11.2010
40, -5	$< f(x) + (1-\alpha)f(y)$	$< f(x) + \alpha(f(y) - f(x))$	18.11.2010
75, 4-6	hier wurden λ_i und λ_i^* sowie μ_j und μ_j^* vertauscht		18.11.2010
85, -1	α_r	α_k	22.11.2010
40, -3	$x_1, \ldots, x_n$	$x_1, \ldots, x_k$	07.01.2011
45, 7	φ stetig diffbar	φ zweimal stetig diffbar	07.01.2011
51, 8	auch x^*	x^* auch	07.01.2011
58, 11	$\nabla h_q(x^*)$	$-\nabla h_q(x^*)$	07.01.2011
58, 17	$\nabla f(x^*)^t d \geq 0$	$\nabla f(x^*)^t x \geq 0$	07.01.2011
58, -12	$\sigma_j \nabla h_i(x^*)$	$\sigma_j \nabla h_j(x^*)$	07.01.2011
62, -11	$h_q(x_0)$	$\nabla h_q(x_0)$	07.01.2011

Seite, Zeile	**Fehler**	**Korrektur**	**Datum**
72, -3	$h_j(x_0) + \alpha \nabla h_j(x_0)^t d \leq 0$	$h_j(x_0) + \alpha \nabla h_j(x_0)^t d = 0$	07.01.2011
88, -15	$x_k = ...$	$f(x_k) = ...$	07.01.2011
95, -1	$\nabla g(x_k)$	$\nabla g_i(x_k)$	07.01.2011
101, 1	$Cx = f$	$Dx = f$	07.01.2011
110, 7	$r \geq p + 1$	$i \geq p + 1$	07.01.2011
119, 13	$\widetilde{A} \begin{pmatrix} x \\ u \end{pmatrix} \leq b$	$\widetilde{A} \begin{pmatrix} x \\ u \end{pmatrix} = b$	07.01.2011
123, 7	$\nabla f(x^*) + \sum_{j=1}^{q} \nabla h_j(x^*)$	$\nabla f(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0$	07.01.2011
123, -7	$\sum_{j=1}^{q} h_j(x)$	$\sum_{j=1}^{q} \mu_j h_j(x)$	07.01.2011
123, -4	$\nabla h(x)\mu$	$\nabla f(x) + \nabla h(x)\mu$	07.01.2011

Bei der Abarbeitung der Korrekturen ist mir aufgefallen, dass der Seitenumbruch in Kapitel 3 gegenber der ersten Version verändert worden ist. Für den Seitenumbruch ist natürlich in der Regel LaTeX verantwortlich, ich weiß nicht, was der Anlass für diese Änderung war. Ich habe mich bemüht, in der aktuellen Version die ursprüngliche Version wiederherzustellen.