Kapitel III

Eine Fehlerabschätzung zum Verfahren von Hildreth und d'Esopo

1. Kuhn-Tucker-Satz und einige Eigenschaften der konvexen Funktionen

Satz 1 (*Kuhn-Tucker*[1]):
Es sei eine Funktion $F(x),\ x \in R^n$, und die Menge

$$S_{III} := \left\{x \mid x \in R^n,\ f_i(x) \le 0,\ (i = 1,2,...,n)\right\}$$

gegeben. Es existiere für alle $y \in R^n$, die den Bedingungen

(3. 1.) $$\sum_{j=1}^{n} \frac{\partial f_i(x^0)}{\partial x_j} y_i \le 0 \text{ für } i \in S^0 := \left\{i \mid f_i(x^0) = 0\right\}$$

bzw.

(3. 1'.) $$y \cdot \nabla f_i(x^0) \le 0 \text{ für } i \in S^0$$

Genügen, eine differenzierbare Kurve $x = h(\alpha)$, die für $0 \le \alpha \le 1$ in der Menge S_{III} liegt und für die

(3. 2.) $$x^0 = h(0)$$

und

(3. 3.) $$\frac{dh(0)}{d\alpha} = \lambda y,\ \lambda > 0$$

gilt[2].

---

[1] Diesen Satz haben erstmals Kuhn und Tucker im Jahre 1950 [43] bewiesen. Er hat in der Theorie der nichtlinearen Optimierung zentrale Bedeutung. In den letzten Jahren erschienen mehrere Varianten des Beweises, wobei jede Variante von einer anderen Regularitätsvoraussetzung ausgeht [7], [16], [21], [24], [63], [70]. Der ursprüngliche Beweis von Kuhn und Tucker basiert auf einer gewissen Eigenschaft des Systems der linearen Ungleichungen. Sie haben dabei die Nichtnegativitätsbedingungen $x_j \ge 0$ von den Restriktionen $f_i(x) \le 0$ getrennt betrachtet. Unser Beweis ist trotz gewisser Modifikationen den Ideen des ursprünglichen Beweises naheliegend.

[2] Diese *Regularitätsvoraussetzung* ist sehr allgemein und daher in konkreten Fällen meistens gegeben. In zwei Fällen ist sie von vornherein erfüllt (Siehe [16]):

1. Wenn die Menge $S^0 = \{i \mid f_i(x) = 0\}$ ist. Mit anderen Worten, wenn x^0 ein innerer Punkt der Menge S_{III} ist, also $f_i(x) < 0$ für alle i gilt.
2. Wenn die Funktionen $f_i(x)$ linear sind.

Es wird weiter vorausgesetzt, dass die Funktionen $F(x)$ und $f_i(x)$ in der Umgebung von x^0 stetige partielle Ableitungen besitzen.
Die *notwendige Bedingungen* dafür, dass x^0 ein lokales Minimum der Funktion $F(x)$ bezüglich der Menge S_{III} darstellt, lauten:

(3. 4.) $$f_i(x^0) \le 0, \quad i = 1,2,...,m,$$

(3. 5.) $$u_i^0 \ge 0, \quad i = 1,2,...,m,$$

(3. 6.) $$u_i^0 f_i(x^0) = 0, \quad i = 1,2,...,m,$$

(3. 7.) $$\frac{\partial F(x^0)}{\partial x_j} + \sum_{i=1}^{m} u_i^0 \frac{\partial f_i(x^0)}{\partial x_j} = 0, \quad j = 1,2,...,n.$$

B e w e i s:
Wir setzen voraus, dass der Punkt x^0 ein lokales Minimum der Funktion $F(x)$ bezüglich der Menge S_{III} darstellt. Es sind zwei Fälle zu unterscheiden:

a) x^0 ist ein innerer Punkt der Menge S_{III}, d.h. für alle i gilt $f_i(x^0) < 0$, und somit ist die Menge $S^0 = \{i \mid f_i(x^0) = 0\}$ leer.
In diesem Falle hat die Funktion $F(x)$ im Punkt x^0 ein freies lokales Minimum, und nach einem bekannten Satz der mathematischen Analysis gilt

(3. 8.) $$\frac{\partial F(x^0)}{\partial x_j} = 0, \quad j = 1,2,...,n.$$

Setzt man zugleich $u_i^0 = 0,\ i = 1,2,...,m,$ so sind die Bedingungen (3. 4.) – (3. 7.) offenbar erfüllt.

b) x^0 ist ein Randpunkt der Menge S_{III}, d.h. $S^0 \neq \varnothing$. Wir werden die Anzahl der Elemente der Menge S^0 mit s_0 $(0 < s_0 \le m)$ bezeichnen und den Satz von Farkas[3] für den Fall

$$A := \left(\frac{\partial f_i(x^0)}{\partial x_j} \right), \quad j = 1,2,...,n;\ i \in S^0,$$

$$b = -\nabla F(x^0), \qquad y \in R^n, \qquad u \in R^{s_0}$$

anwenden.

[3] **Satz** (***Farkas***): Es seien A eine $(m \cdot n)$-Matrix, u ein m-Komponentenvektor und y und b n-dimensionale Vektoren.
Wenn für alle Lösungen des Ungleichungssystems $yA \le 0$ die Beziehung $yb \le 0$ gilt, so existiert eine nichtnegative Lösung des Systems $Au = b$.

Es wird nun das Ungleichungssystem (3. 1‘.), also

$$y \cdot \nabla f_i(x^0) \le 0, \quad i \in S^0 \qquad \text{(bzw. } y \cdot A \le 0\text{)}$$

Näher untersucht:

Nach der Regularitätsvoraussetzung (3. 2.) und (3. 3.) gilt:

(3. 9.)
$$\frac{dF(h(0))}{d\alpha} = \sum_{j=1}^{n} \frac{\partial F(x^0)}{\partial x_j} \cdot \frac{dh_j(0)}{d\alpha}$$

$$= \lambda \sum_{j=1}^{n} \frac{\partial F(x^0)}{\partial x_j} y_j$$

$$= \lambda y \cdot \nabla F(x^0).$$

Da x^0 ein lokales Minimum der Funktion $F(x)$ darstellt, gilt

$$\frac{dF(h(0))}{d\alpha} \ge 0$$

und gemäß (3. 9.)

(3. 10.)
$$-y \cdot \nabla F(x^0) \le 0 \qquad \text{(bzw. } y \cdot b \le 0\text{)}.$$

Damit sind die Voraussetzungen des Satzes von Farkas erfüllt:
Für jeden Vektor y, der das System (3. 1.) erfüllt, gilt (3. 10.). Daraus folgt, dass das System

(3. 11.)
$$\sum_{i \in S^0} \frac{\partial f_i(x^0)}{\partial x_j} \cdot u_i = -\frac{\partial F(x^0)}{\partial x_j}, \quad j = 1,2,...,n \qquad \text{(bzw. } Au = b\text{)}$$

eine nichtnegative Lösung $u_i^0 \ge 0,\ i \in S^0$, aufweist.
Setzt man $u_i^0 = 0$ für $i \notin S^0$, so ist das System (3. 11.) äquivalent der Beziehung

$$\sum_{i=1}^{m} \frac{\partial f_i(x^0)}{\partial x_j} \cdot u_i^0 = -\frac{\partial F(x^0)}{\partial x_j}, \quad j = 1,2,...,n,$$

welche gleich der Beziehung (3. 7.) ist. Damit ist die Notwendigkeit der Bedingungen (3. 5.) und (3. 7.) bewiesen. Da die Bedingung (3. 4.) selbstverständlich ist, ist nur noch die Bedingung (3. 6.) zu beweisen:
Offensichtlich gilt

$$f_i(x^0) = 0 \quad \text{für } i \in S^0$$

und

$u_i^0 = 0$ für $i \notin S^0$.

In beiden Fällen muss also $u_i^0 \cdot f_i(x^0) = 0$ gelten.

Damit ist der Beweis beendet.

Anmerkung:

1. Mit Hilfe der Lagrange-Funktion

(3. 12.) $$L(x,u) = F(x) + \sum_{i=1}^{m} u_i f_i(x)$$

kann man die Bedingungen (3. 4.) bis (3. 7.) folgendermaßen schreiben:

(3. 13.)
$$\begin{cases} \dfrac{\partial L(x^0,u^0)}{\partial x_j} = 0 & j = 1,2,...,n \\ \dfrac{\partial L(x^0,u^0)}{\partial u_i} \le 0 & i = 1,2,...,m \\ u_i^0 \dfrac{\partial L(x^0,u^0)}{\partial u_i} = 0 & i = 1,2,...,m \\ u_i^0 \ge 0 & i = 1,2,...,m \end{cases}$$

2. Ohne Beweis führen wir noch die entsprechenden Kuhn-Tucker-Bedingungen an für die Menge

$$S_I := \{x \mid x \in R^n, f_i(x) \le 0,\ i = 1,2,...,m;\ x_j \ge 0,\ j = 1,2,...,n\}:$$

(3. 14.)
$$\begin{cases} \dfrac{\partial L(x^0,u^0)}{\partial x_j} \ge 0 & j = 1,2,...,n \\ \dfrac{\partial L(x^0,u^0)}{\partial u_i} \le 0 & i = 1,2,...,m \\ x_j^0 \cdot \dfrac{\partial L(x^0,u^0)}{\partial x_j} = 0 & j = 1,2,...,n \\ u_i^0 \cdot \dfrac{\partial L(x^0,u^0)}{\partial u_i} = 0 & i = 1,2,...,m \\ x_j^0 \ge 0 & j = 1,2,...,n \\ u_i^0 \ge 0 & i = 1,2,...,m \end{cases}$$

und die Menge

$$S_{II} := \{x \mid x \in R^n, f_i(x) = 0,\ i = 1,2,...,m;\ x_j \ge 0,\ j = 1,2,...,n\}:$$

(3. 15.)
$$\begin{cases} \dfrac{\partial L(x^0,u^0)}{\partial x_j} \geq 0 & j=1,2,...,n \\ x_j^0 \cdot \dfrac{\partial L(x^0,u^0)}{\partial x_j} = 0 & j=1,2,...,n \\ x_j^0 \geq 0 & j=1,2,...,n \\ \dfrac{\partial L(x^0,u^0)}{\partial u_i} = 0 & i=1,2,...,m \end{cases}$$

Definition 1:
Es sei eine Funktion $F(x),\ x \in R^n$, über einer konvexen Menge $K \subset R^n$ definiert.

a) $F(x)$ heißt *konvex*, wenn für je zwei Punkte $x^1, x^2 \in K$ und einen beliebigen Skalar $\lambda,\ 0 \leq \lambda \leq 1,$ gilt:

(3. 16.) $$F\left\{\lambda x^1 + (1-\lambda)x^2\right\} \leq \lambda F(x^1) + (1-\lambda)F(x^2).$$

b) $F(x)$ heißt *streng konvex*, wenn für je zwei Punkte $x^1, x^2 \in K,\ x^1 \neq x^2,$ und einen beliebigen Skalar $\lambda,\ 0 < \lambda < 1,$ gilt:

(3. 16.) $$F\left\{\lambda x^1 + (1-\lambda)x^2\right\} < \lambda F(x^1) + (1-\lambda)F(x^2).$$

Lemma 1:
Es sei $f(x)$ im offenen Intervall $(a,\ b)$ definiert, wobei $-\infty \leq a < b < +\infty.$ Es existiere in jedem Punkt $x \in (a,\ b)$ die Ableitung $| f'(x) | < \infty$.

a) $f(x)$ ist konvex dann und nur dann, wenn $f'(x)$ eine nichtfallende Funktion darstellt.
b) $f(x)$ ist streng konvex dann und nur dann, wenn $f'(x)$ eine steigende Funktion darstellt.

B e w e i s:
Wir setzen voraus, dass $a < x_1 < x_2 < b$ ist

1a) *Es sei $f(x)$ konvex.*

Für alle $x^1, x^2 \in (a,\ b)$ und $\lambda \in (0,\ 1)$ muss dann gelten:

(3. 18.) $$f\left\{\lambda x^1 + (1-\lambda)x^2\right\} \leq \lambda f(x^1) + (1-\lambda) f(x^2).$$

Setzt man $x := \lambda x^1 + (1-\lambda)x^2$, so wird (3. 18.) äquivalent der folgenden Beziehung:

(3. 19.) $$\frac{f(x)-f(x^1)}{x-x_1} \le \frac{f(x^2)-f(x)}{x_2-x_1}.$$

Wird in (3. 19.) zu den Grenzwerten $\lim_{x\to x_1}$ und $\lim_{x\to x_2}$ übergegangen, so erhält man

(3. 20.) $$f'(x_1) \le \frac{f(x_2)-f(x_1)}{x_2-x_1} \le f'(x_2),$$

woraus unmittelbar folgt, dass $f'(x)$ eine nichtfallende Funktion ist.

1b) Es sei $f(x)$ streng konvex.

Wir nehmen an, dass $f'(x_1) = f'(x_2)$ ist. Mit anderen Worten ist $f'(x)$ im Intervall $(x_1,\ x_2)$ linear. Das widerspricht aber der Voraussetzung über die strenge Konvexität von $f(x)$. Es muss also

$$f'(x_1) < f'(x_2)$$

gelten, wobei

$$x_1 < x_2$$

ist, d.h., $f'(x)$ stellt eine steigende Funktion dar.

IIa) Es sei $f'(x)$ nichtfallend.

Wählt man einen Punkt x mit $x_1 < x < x_2$, so müssen nach dem Lagrangeschen Mittelwertsatz zwei Zahlen ξ_1 und ξ_2 mit $x_1 < \xi_1 < x < \xi_2 < x_2$ existieren derart, dass

(3. 21.) $$f'(\xi_1) = \frac{f(x)-f(x_1)}{x-x_1}; \qquad f'(\xi_2) = \frac{f(x_2)-f(x)}{x_2-x}$$

ist. Auf Grund der Voraussetzung $f'(\xi_1) \le f'(\xi_2)$ folgt die Beziehung (3. 19.), d.h. $f(x)$ ist eine konvexe Funktion.

IIb) Es sei $f'(x)$ steigend.

Die Beweisführung ist analog dem Fall *IIa)*, wobei das Vorzeichen ($\le$) durch ($<$) ersetzt wird. Daraus folgt die strenge Konvexität von $f(x)$.

<u>Folgerung</u>:
Existiert in jedem Punkt $x \in (a,\ b)$ die zweite Ableitung $f''(x)$, so kann man das Lemma 1 folgendermaßen umformulieren:

a) $f(x)$ ist konvex dann und nur dann, wenn für alle $x \in (a,\ b)$ gilt $f''(x) \geq 0$.
b) $f(x)$ ist streng konvex, wenn für alle $x \in (a,\ b)$ gilt $f''(x) > 0$
(Mit Ausnahme der isolierten Punkte, für welche $f''(x) = 0$ ist.)

Lemma 2:
Es sei

(3. 22.) $$\psi(\lambda) = F(x + \lambda t), \quad x, t \in R^n,\ \lambda \in R^1.$$

a) $F(x)$ ist konvex dann und nur dann, wenn $\psi(\lambda)$ für alle $x, t \in R^n$ konvex ist.
b) $F(x)$ ist streng konvex dann und nur dann, wenn $\psi(\lambda)$ für alle $x, t \in R^n$ streng konvex ist.

B e w e i s:
Es sei $0 \leq \mu \leq 1$. offensichtlich gilt:

(3. 23.) $$\psi(\mu\lambda_1 + (1-\mu)\lambda_2) = F(x + (\mu\lambda_1 + (1-\mu)\lambda_2)t)$$

$$= F(\mu(x + \lambda_1 t) + (1-\lambda)(x + \lambda_2 t))$$

$$= F(\mu x^1 + (1-\mu)x^2),$$

mit

(3. 24.) $$x^1 = x + \lambda_1 t; \quad x^2 = x + \lambda_2 t.$$

(Wir bemerken, dass man zu den gegebenen $x^1, x^2 \in R^n$ immer zwei Punkte x und t und die beiden Skalare λ_1 und λ_2 bestimmen kann, derart, dass (3. 24.) gelten wird.)

Aus (3. 22.) und (3. 24.) folgt:

$$\psi(\lambda_1) = F(x + \lambda_1 t) = F(x^1)$$
$$\psi(\lambda_2) = F(x + \lambda_2 t) = F(x^2),$$

und nach (3. 23.) und (3. 24.) gilt

$$F(\mu\lambda_1 + (1-\mu)\lambda_2) = F(\mu x^1 + (1-\mu)x^2).$$

Aus dem Angeführten ergibt sich, dass die Konvexitätsbedingung für $F(x)$, d.h.

$$F(\mu x^1 + (1-\mu)x^2) \leq \mu F(x^1) + (1-\mu)F(x^2)$$

Äquivalent der Konvexitätsbedingung für $\Psi(\lambda)$ ist, also

$$\psi(\mu\lambda_1 + (1-\mu)\lambda_2) \leq \mu\psi(\lambda_1) + (1-\mu)\psi(\lambda_2)$$

gilt.
Damit ist der Beweis beendet.

Anmerkung:
Lemma 2 ermöglicht, die Aussagen über die konvexen Funktionen von n Variablen zu vereinfachen, indem diese Aussagen auf Aussagen über Funktionen einer Variablen reduziert werden können.

Folgerung (*Verallgemeinerung des Lemmas 1*)
Es sei $x, t \in R^n$ eine stetig differenzierbare Funktion.

a) $F(x)$ ist konvex, wenn die Funktion

$$\omega(\lambda) = \sum_{j=1}^{n} \frac{\partial F(x+t)}{\partial x_j} \cdot t_j = t \cdot \nabla F(x + \lambda t)$$

für alle $x, t \in R^n$ nichtfallend ist.

b) $F(x)$ ist streng konvex, wenn $\omega(\lambda)$ für alle $x, t \in R^n$ steigend ist.

B e w e i s:

$$\omega'(\lambda) = \frac{d}{d\lambda} F(x + \lambda t) = \sum_{j=1}^{n} \frac{\partial F(x+t)}{\partial x_j} \cdot t_j = \omega(\lambda)$$

und die Behauptung der Folgerung ergibt sich unmittelbar aus Lemma 1 und Lemma 2.

Satz 2:
Es sei $F(x), x \in R^n$, eine stetig differenzierbare Funktion.

a) $F(x)$ ist konvex dann und nur dann, wenn für alle beliebigen Punkte $x^1, x^2 \in R^n$

(3. 25.) $$F(x^2) - F(x^1) \geq \sum_{j=1}^{n} \frac{\partial F(x^1)}{\partial x_j} \cdot (x_j^2 - x_j^1)$$

bzw. in Vektorenschreibweise (mit Hilfe des Gradienten der Funktion

(3. 25‘.) $$F(x^2) - F(x^1) \geq (x^2 - x^1) \cdot \nabla F(x^1)$$

gilt.

b) $F(x)$ ist streng konvex dann und nur dann, wenn für alle beliebigen Punkte $x^1, x^2 \in R^n,\ x^1 \neq x^2$

(3. 26.) $$F(x^2) - F(x^1) > (x^2 - x^1) \cdot \nabla F(x^1)$$

gilt.

B e w e i s:

Ia) *Es sei $f(x)$ konvex.*

Dann ist nach Lemma 2 die Funktion $\psi(\lambda) = F(x+t)$ konvex für alle $x, t \in R^n$. Für $\lambda_1 < \lambda_2$ muss also gelten (siehe den Beweis von Lemma 1, Beziehung (3. 20.)):

$$\psi'(\lambda_1) \le \frac{\psi(\lambda_2) - \psi(\lambda_1)}{\lambda_2 - \lambda_1}$$

bzw.

(3. 27.) $$\psi(\lambda_2) - \psi(\lambda_1) \ge \psi'(\lambda_1)(\lambda_2 - \lambda_1).$$

Zunächst zeigen wir, dass (3. 27.) äquivalent der Beziehung (3. 25.) (bzw. (3. 25‘.) ist:
Setzt man $x + \lambda_1 t = x^1$ und $x + \lambda_2 t = x^2$, so gilt

(3. 28.) $$\psi(\lambda_2) - \psi(\lambda_1) = F(x^2) - F(x^1)$$

und (siehe Folgerung zum Lemma 2):

(3. 29.) $$\psi'(\lambda_1)(\lambda_2 - \lambda_1) = (\lambda_2 - \lambda_1)t' \cdot \nabla F(x^1)$$

$$= (\lambda_2 t - \lambda_1 t)' \nabla F(x^1)$$

$$= ((x + \lambda_2 t) - (x + \lambda_1 t)) \nabla F(x^1)$$

$$= (x^2 - x^1) \cdot \nabla F(x^1).$$

Damit ist die Beziehung (3. 25.) bzw. (3. 25‘.) beendet.

Ib) *Es sei $f(x)$ streng konvex.*

Dann ist $\psi'(\lambda)$ streng konvex, und gemäß Lemma 1 stellt $\psi'(\lambda)$ eine steigende Funktion für alle $x, t \in R^n$ dar.
Es sei $\lambda_1 < \lambda_0 < \lambda_2$. Nach (3. 27.) muss dann gelten:

$$\psi(\lambda_2) - \psi(\lambda_0) \ge \psi'(\lambda_0)(\lambda_2 - \lambda_0) > \psi'(\lambda_1)(\lambda_2 - \lambda_0)$$

$$\psi(\lambda_0) - \psi(\lambda_1) \qquad\qquad \ge \psi'(\lambda_1)(\lambda_0 - \lambda_1).$$

Durch Addition der obigen Ungleichungen erhält man

(3. 30.) $$\psi(\lambda_2) - \psi(\lambda_1) > \psi'(\lambda_1)(\lambda_2 - \lambda_1).$$

IIa) *Es möge für alle* $x^1, x^2 \in R^n$ *die Beziehung (3. 25.) gelten.*

Dann gilt gemäß (3. 28.) und (3. 29.) die Formel (3. 27.). Tauscht man in (3. 27.) λ_1 und λ_2 gegen einander aus, so erhält man

(3. 31.) $$\psi(\lambda_1) - \psi(\lambda_2) \geq \psi'(\lambda_2)(\lambda_1 - \lambda_2).$$

Durch Kombination der Beziehungen (3. 27.) und (3. 28.) hat man

$$\psi'(\lambda_1) \leq \frac{\psi(\lambda_2) - \psi(\lambda_1)}{\lambda_2 - \lambda_1} \leq \psi'(\lambda_2).$$

Daraus ergibt sich unter Berücksichtigung der Folgerung zum Lemma 2, dass $F(x)$ konvex ist.

IIb) Man verfährt analog zum Fall IIa), nur anstelle der Beziehung (3. 27.) verwendet man (3. 30.).

Folgerung:
Die Kuhn-Tucker-Bedingungen (3. 4.) bis (3. 7.) für die Aufgabe der konvexen Optimierung sind gleich *hinreichende Bedingungen* für ein *globales Minimum*[4].
Genauer gesagt:
Es sei die folgende Aufgabe gegeben:

$$\min\{F(x) \mid x \in R^n,\ f_i(x) \leq 0,\ i = 1,2,...,m\},$$

wobei $F(x)$ und $f_i(x)$ konvexe Funktionen sind.
Erfüllt x^0 die Kuhn-Tucker-Bedingungen (3. 4.) bis (3. 7.), so stellt dieser Punkt das globale Minimum der Funktion $F(x)$ auf der Menge

$$S_{III}^k := \{x \mid x \in R^n,\ f_i(x) \leq 0,\ i = 1,2,...,m\}$$

dar.

B e w e i s:
Aus der Konvexität von $F(x)$ und $f_i(x)$ ergibt sich nach dem obigen Satz

(3. 32.) $$F(x) - F(x^0) \geq \nabla F(x^0)(x - x^0)$$

(3. 33.) $$f_i(x) - f_i(x^0) \geq \nabla f_i(x^0)(x - x^0),\ i = 1,2,...,m.$$

Verwendet man die Beziehungen (3. 32.), (3. 7.), (3. 33.), (3. 6.) und (3. 5.), so erhält man für all $x \in S_{III}^k$:

---

[4] **Satz:** Ein *lokales* Minimum der konvexen Funktion $F(x),\ x \in R^n$ auf der konvexen Menge $S^k \subset R^n$ realisiert zugleich das *globale* Minimum der Funktion $F(x)$ auf $S^k \subset R^n$ -

(3. 34.)
$$\begin{aligned} F(x) - F(x^0) &\geq \nabla F_i(x^0)(x - x^0) \\ &= -\sum_{i=1}^{m} u_i^0 \nabla f_i(x^0)(x - x^0) \\ &\geq= -\sum_{i=1}^{m} u_i^0 (f_i(x) - f_i(x^0)) \\ &= -\sum_{i=1}^{m} u_i^0 f_i(x) \geq 0, \end{aligned}$$

oder

$$F(x) \geq F(x^0).$$

Damit ist die Behauptung der Folgerung bewiesen.

Anmerkung:
Es ist zu bemerken, dass der obige Beweis im Gegenteil zum Satz 1 auf keiner Regularitäts-voraussetzung basiert.

Satz 3 (*Kuhn-Tucker*)
Wir gehen von der Aufgabe der konvexen Optimierung

(3. 35.)
$$\min\{F(x) \mid f_i(x) \leq 0,\ i = 1,2,...,m;\ x_j \geq 0,\ j = 1,2,...,n\}$$

(wobei $F(x)$ und $f_i(x)$ konvexe Funktionen sind) aus[5].

Es sei der zulässige Bereich

$$S_I^k := \{x \mid x \in R^n,\ f_i(x) \leq 0,\ i = 1,2,...,m;\ x_j \geq 0,\ j = 1,2,...,n\}$$

regulär im Sinne des Satzes 1, und die die Funktionen $F(x)$ und $f_i(x)$ haben stetige partielle Ableitungen.
Die *notwendigen* und *hinreichenden Bedingungen* dafür, dass x^* eine optimale Lösung des Problems (3. 35.) darstellt, lauten:

(3. 36.)
$$\begin{cases} \dfrac{\partial L(x^*, u^*)}{\partial x} \geq 0 & (A) \\ x^* \cdot \dfrac{\partial L(x^*, u^*)}{\partial x} = 0 & (B) \\ x^* \geq 0 & (C) \\ \dfrac{\partial L(x^*, u^*)}{\partial u} \leq 0 & (D) \\ u^* \cdot \dfrac{\partial L(x^*, u^*)}{\partial x} = 0 & (E) \\ u^* \geq 0 & (F) \end{cases}$$

[5] Der Satz gilt natürlich für alle Normalformen der Aufgabe der konvexen Optimierung. Wir haben die obige Form explizit angeführt, da die symmetrische Form der zugehörigen Kuhn-Tucker-Bedingung am besten auch die anderen Normalformen umfasst.

Der *Beweis* folgt unmittelbar aus Satz 1 und der entsprechende Verallgemeinerung der Folgerung aus Satz 2 auf (3. 35.).

2. Herleitung einer Fehlerabschätzung

<u>Satz 4</u>:

Sind die Bedingungen (*A*), (*C*), (*D*) und (*F*) im Satz 3 für den Punkt $(\bar{x},\ \bar{u})$ erfüllt, d.h., stellt $\bar{x}$ eine zulässige Lösung des Problems der konvexen Optimierung dar, für welche ein Vektor $\bar{u} \in R^m$ existiert, so dass

$$\nabla F(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla f_i(\bar{x}) \geq 0; \qquad \bar{u} \geq 0 \tag{3. 37.}$$

ist, dann gilt

$$0 \leq F(\bar{x}) - F(x^*) \leq \bar{x} \cdot \frac{\partial L(\bar{x},\bar{u})}{\partial x} - \bar{u} \cdot \frac{\partial L(\bar{x},\bar{u})}{\partial u}. \tag{3. 38.}$$

B e w e i s:
Offensichtlich gilt

$$\begin{aligned} \nabla F(\bar{x}) &= \{\nabla F(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla f_i(\bar{x})\} - \sum_{i=1}^{m} \bar{u}_i \nabla f_i(\bar{x}) \\ &= \frac{\partial L(\bar{x},\bar{u})}{\partial x} - \sum_{i=1}^{m} \bar{u}_i \nabla f_i(\bar{x}). \end{aligned} \tag{3. 39.}$$

Da aber x^* eine optimale (minimale) Lösung des Problems (3. 35.) darstellt, muss

$$0 \geq F(x^*) - F(\bar{x})$$

gelten.

Aus der Konvexität der Funktion $F(x)$ und der Beziehung (3. 39.) folgt nach Satz 2:

$$0 \geq F(x^*) - F(\bar{x}) \geq (x^* - \bar{x}) \cdot \nabla F(\bar{x})$$

$$= (x^* - \bar{x}) \frac{\partial L(\bar{x},\bar{u})}{\partial x} - \sum_{i=1}^{m} \bar{u}_i (x^* - \bar{x}) \nabla f_i(\bar{x}).$$

Da die Funktionen $f_i(x)$ ebenfalls konvex sind (Satz 2), erhält man unter Verwendung der Beziehung (3. 37.):

$$(x^*-\bar{x})\frac{\partial L(\bar{x},\bar{u})}{\partial x}-\sum_{i=1}^{m}\bar{x}_i(x^*-\bar{x})\geq(x^*-\bar{x})\frac{\partial L(\bar{x},\bar{u})}{\partial x}-\sum_{i=1}^{m}\bar{u}_i(f_i(x^*)-f_i(\bar{x}))$$

$$\geq-\bar{x}\frac{\partial L(\bar{x},\bar{u})}{\partial x}+\sum_{i=1}^{m}\bar{u}_i f_i(\bar{x})$$

$$=-\bar{x}\frac{\partial L(\bar{x},\bar{u})}{\partial x}+\bar{u}\frac{\partial L(\bar{x},\bar{u})}{\partial u}.$$

Aus dem Ausgeführten folgt unmittelbar

$$0\leq F(\bar{x})-F(x^*)\leq-\bar{x}\frac{\partial L(\bar{x},\bar{u})}{\partial x}+\bar{u}\frac{\partial L(\bar{x},\bar{u})}{\partial u}.$$

<u>Satz 5</u> (*Spezialisierung des Satzes 4 auf den Fall der streng konvexen quadratischen Optimierung:* $F(x)=Q(x)$ <u>)</u>:

Es sei:

(3. 40.) $$Q(\bar{x})-Q(x^*)=\Delta Q(\bar{x})$$

(1. 12.) $$G=\frac{1}{4}AC^{-1}A^T$$

(1. 13.) $$h=b+\frac{1}{2}AC^{-1}p.$$

Dann gilt folgende Beziehung:

(3. 41.) $$0\leq\Delta Q(\bar{x})\leq 2\bar{u}^T G\bar{u}+h^T\bar{u}.$$

B e w e i s:
Substituiert man die für den Fall der quadratischen Optimierung spezialisierten Ausdrücke

(3. 42.) $$\frac{\partial L(\bar{x},\bar{u})}{\partial x}=2Cx+A^T\bar{u}+p$$

(3. 43.) $$\frac{\partial L(\bar{x},\bar{u})}{\partial u}=A\bar{x}-b$$

in (3. 38.), so erhält man unter Verwendung von (3. 40.), (1. 12.) und (1. 13.):

$$0\leq\Delta(\bar{x})=\bar{x}^T\{2C\bar{x}+p+A^T\bar{u}\}-\bar{u}^T\{A\bar{x}-b\}$$

$$= \bar{x}^T \left\{ -2C \cdot \frac{1}{2} C^{-1}(A^T \bar{u} + p) + p + A^T \bar{u} \right\} - \bar{u}^T \left\{ A \left[-\frac{1}{2} C^{-1}(A^T \bar{u} + p) - b \right] \right\}$$

$$= -\bar{u}^T \left\{ -\frac{1}{2} AC^{-1} A^T \bar{u} - \frac{1}{2} AC^{-1} p - b \right\}$$

$$= 2\bar{u}^T G\bar{u} + h^T \bar{u},$$

also

$$0 \le \Delta Q(\bar{x}) \le 2\bar{u}^T G\bar{u} + h^T \bar{u}.$$

Die obige Beziehung[6] gibt eine Abschätzung der Differenz des Zielfunktionswertes des Primalproblems $Q(x)$ im zulässigen Punkt $\bar{x}$ und der optimalen Lösung an. Diese Differenz kann ein Zweifaches des Zielfunktionswertes der Dualaufgabe minus ihres linearen Teils nicht übersteigen.
Durch die Formel (3. 4.) kann man eine Abschätzung der streng konvexen quadratischen Optimierungsaufgaben erzielen, die auf dem Dualitätsprinzip basieren. Das bekannteste *„Dualverfahren"* ist das Verfahren von Hildreth und d'Esopo. Diese Fehlerabschätzung kann aber theoretisch auch z.B. beim Verfahren der Schnittebenen von J. E. Kelley, jr. [39] für den Fall der streng konvexen quadratischen Optimierung Verwendung finden[7]. In unserer Betrachtung werden wir uns allerdings nur auf das Verfahren von Hildreth und d'Esopo beschränken[8].

Im Weiteren werden zuerst zwei andere, aus der bisherigen Literatur bekannte Fehlerabschätzungen der Dualverfahren der streng konvexen quadratischen Optimierung wiedergegeben[9]. Anschließend soll eine Einschätzung der einzelnen Fehlerabschätzungen erfolgen und dazu ein Illustrationsbeispiel angegeben werden.

3. Andere Fehlerabschätzungen

3. 1. Fehlerabschätzung des Dualproblems (C. Hildreth)

Hildreth hat in [33] eine Fehlerabschätzung des Dualproblems (2. 3.) angedeutet, die hier entwickelt werden soll:

Es sei

(3. 44.) $$u^* := \bar{u} - b \quad (\bar{u} \neq 0).$$

---

[6] Das Wesen einiger Verfahren der quadratischen Optimierung (z.B. das Verfahren von Barankin und Dorfman und das Verfahren von Frank und Wolfe) besteht in der Tat darin, die rechte Seite dieses Ausdrucks bzw. der Beziehung (3. 38.) unter Beachtung der übrigen linearen Kuhn-Tucker-Restriktionen zur Null zu reduzieren.

[7] Dieses Verfahren kann zwar im $x-$ durchgeführt werden, durch den Übergang zur dualen Optimierungsaufgabe wird jedoch der scheinbar erheblichen Rechenaufwand in erträglichen Grenzen gehalten.

[8] Es ist leicht zu verifizieren, dass der Algorithmus von Hildreth und d'Esopo stets die Erfüllung der Voraussetzung (3. 37.) automatisch garantiert.

[9] Diese Fehlerabschätzungen waren ursprünglich nur für das Verfahren von Hildreth und d'Esopo gedacht; sie können jedoch theoretisch auch für andere Dualverfahren verallgemeinert werden.

Wegen (2. 3.) gilt:

$$\Theta(u^*) = (\bar{u} - b)^T G(\bar{u} - b) + h^T(\bar{u} - b)$$

$$= \bar{u}^T G\bar{u} + h^T \bar{u} - 2\bar{u}^T Gb + b^T Gb - h^T b$$

$$= \Theta(\bar{u}) - (2G\bar{u} + h)^T b + b^T Gb.$$

Daher ist

(3. 45.) $$\Theta(u^*) - \Theta(\bar{u}) = -\left(\frac{\partial\Theta(\bar{u})}{\partial u}\right)^T b + b^T Gb.$$

Aus der Nichtnegativität von $b^T Gb$ folgt

$$\Theta(u^*) - \Theta(\bar{u}) \geq -\left(\frac{\partial\Theta(\bar{u})^T}{\partial u}\right)^T b.$$

Setzt man nun

$$\Theta(u^*) - \Theta(\bar{u}) = \Delta\Theta(\bar{u}) \geq 0\,,$$

so erhält man die folgende Abschätzung der $\bar{u}-$ Iteration von der optimalen Lösung des Dualproblems

(3. 46.) $$0 \leq \Delta\Theta(\bar{u}) \leq (2G\bar{u} + h)^T b.$$

<u>Anmerkung</u>:
Stellt das Primalproblem eine Maximierungsaufgabe dar, so setzt man statt (3. 44.)

(3. 44‘.) $$u^* := \bar{u} + b \quad (\bar{u} \neq 0),$$

woraus sich die Fehlerabschätzung folgendermaßen ergibt:

(3. 46‘.) $$0 \leq \Delta\Theta(\bar{u}) \leq -(2G\bar{u} + h)^T b.$$

3. 2. Abweichung zweier aufeinander folgender Primallösungen (H. Werner[10])

u^k und u^{k+1} (bzw. x^k und x^{k+1}) seien zwei aufeinander folgender Iterationslösungen. Es gelte

[10] Siehe [68].

(3. 47.) $$d = u^{k+1} - u^k$$

und

(3. 48.) $$R = \| r_{ij} \| = C^{-1} A^T .$$

Dann folgt aus $2Cx + A^T u + p = 0$ die Beziehung

(3. 49.)
$$\begin{aligned} x^{k+1} - x^k &= -\frac{1}{2} C^{-1} \left(A^T u^{k+1} + p \right) + \frac{1}{2} C^{-1} \left(A^T u^k + p \right) \\ &= -\frac{1}{2} C^{-1} A^T d \\ &= -\frac{1}{2} R d, \end{aligned}$$

oder Komponentenweise

(3. 50.) $$| x_j^{k+1} - x_j^k | = -\frac{1}{2} | \sum_{i=1}^{m} r_{ji} d_i | .$$

Verwendet man nun eine „mittlere Abweichung“ Δx_m , so erhält man

(3. 51.)
$$\begin{aligned} \Delta x_m &= \frac{1}{n} \sum_{j=1}^{n} | x_j^{k+1} - x_j^k | \\ &\le \frac{1}{2n} \sum_{j=1}^{n} | \sum_{i=1}^{m} r_{ji} d_i | \\ &\le \frac{1}{2n} \sum_{j=1}^{n} \sum_{i=1}^{m} | r_{ij} | \cdot | d_i | \\ &= \frac{1}{2n} \sum_{i=1}^{m} \sum_{j=1}^{n} | r_{ij} | \cdot | d_i | \\ &= \sum_{i=1}^{m} f_i | d_i | \end{aligned}$$

mit

(3. 52.) $$f_i := \frac{1}{2n} \sum_{j=1}^{n} | r_{ji} |, \quad i = 1, 2, ..., m.$$

Ist eine „zulässige mittlere Abweichung „ Δx vorhanden, so wird im programmierten Verfahren die Iteration im u – Raum abgerochen, wenn

(3. 53.) $$\sum_{i=1}^{m} f_i | d_i | \le \Delta x$$

gilt.

4. Ein Illustrationsbeispiel und Diskussion der Ergebnisse

Die vorliegenden Fehlerabschätzungen wurden auf das im zweiten Kapitel berechnete Beispiel angewandt.[11]
Das Wesen der im Abschnitt 4 entwickelten Fehlerabschätzungen steht im gewissen Zusammenhang mit der Problematik der Dualität in der mathematischen Optimierung. Das Dualitätsprinzip der linearen Optimierung ist seit längerer Zeit bekannt; es ist Gegenstand intensiver theoretischer und numerischer Untersuchungen gewesen. Die Erforschung der Dualität in der nichtlinearen Optimierung befindet sich dagegen erst im Anfangsstadium. Die bisherigen Untersuchungen auf diesem Gebiet scheinen jedoch die Vermutung zu bestätigen, dass die Dualität in der nichtlinearen Optimierung bei weitem nicht die Bedeutung dieser Kategorie in der linearen Optimierung annimmt. Diese Behauptung ergibt sich u. a. aus folgenden Tatbeständen:

1. Die geringere Symmetrie des Primal- und Dualproblems in der nichtlinearen Optimierung, d.h. hier hat man im Allgemeinen nicht die Erscheinung, dass sich durch zweimaliger Dualisierung wieder die ursprüngliche Aufgabe ergibt. Außerdem sind die symmetrischen Beziehungen der beiden Probleme hinsichtlich der Dualitätseigenschaften nicht immer eindeutig bestimmt und umgekehrt.

2. Das Auftreten eines Teils des Primalproblems in der Zielfunktion des Dualproblems.

Ist das nichtlineare Problem quadratisch mit streng konvexer Zielfunktion, so ist es möglich (Siehe Kapitel I, Abschnitt 3), die beiden genannten Nachteile zu beseitigen. Durch die Elimination der Primalvariablen in der dualen Zielfunktion wird einerseits die Symmetrie zwischen dem Primal- und dem Dualproblem hergestellt; andererseits wird der Dualraum von den Elementen des Primalproblems bereinigt.
Hat man die angeführten Schwierigkeiten beseitigt, so besteht grundsätzlich die *Möglichkeit*, durch die Transformationsformel (1. 11.) für jedes u ein x anzugeben. Die Zulässigkeit der Duallösung muss natürlich nicht unbedingt die Zulässigkeit der Lösung im Primalproblem bedeuten. Das ist aber immerhin im Vergleich zu den am meisten verbreiteten Methoden der linearen Optimierung, z.B. der Simplexmethode ein Vorteil. Während bei der Simplexmethode alle Zwischenschritte grundsätzlich dual unzulässig bleiben[12], hat man bei der Lösung des Problems (2. 3.) die *Möglichkeit*, durch eine zulässige Lösung des Primalproblems eine zulässige Lösung des Dualproblems zu finden.

Der Frage der Zulässigkeit scheint trotzdem auch hier sowohl aus theoretischer Sicht als auch von der praktischen Anwendbarkeit her entscheidende Bedeutung zuzukommen.

Bei unseren empirischen Berechnungen an Hand der im Anhang angeführten Probleme und weiterer großer Aufgaben konnten hinsichtlich der Zulässigkeit drei Typen von Problemen festgestellt werden:

[11] Aus Platzgründen wird auf die Widergabe der detaillierten Ergebnisse verzichtet.

[12] Der Simplexalgorithmus beginnt mit einer zulässigen Lösung des Primalproblems, die Schrittweise verbessert wird. Er beginnt jedoch zugleich mit einer unzulässigen Lösung des Dualproblems (bei der entweder einige Strukturvariablen oder Schlupfvariablen negativ sind), und mit jeder Iteration wird die „Unzulässigkeit (das heiß das Gewicht der negativen Lösungskoordinaten in der Zielfunktion) erniedrigt.

Typ I umfasst solche Probleme, bei denen alle Zwischenschritte primal zulässig bleiben, z.B. das Illustrationsbeispiel, die Aufgaben II $II\ C_1, p_4, II\ C_2, p_4$ usw.

Zum *Typ II* gehören solche Probleme, bei denen alle Zwischenschritte primal zulässig bleiben, z.B. die Aufgaben $II\ C_1, p_6, II\ C_2, p_7$ und $III\ p_3$.

Dem *Typ III* werden solche Probleme zugeordnet, bei denen die Zwischenschritte teilweise zulässig und teilweise unzulässig sind, wobei im Laufe des Verfahrens die Unzulässigkeit erfahrungsgemäß abnimmt. Zu diesem Typ gehören z.B. die Aufgaben $I\ p_2$, $I\ p_5$ und $II\ C_2, p_7$ usw.

Typ I stellt offenbar hinsichtlich der Anwendung der Dualfehlerabschätzungen den idealen Typ dar, da im Falle der schlechten Konvergenz die Berechnung, nachdem die gewünschte Genauigkeit erreicht worden ist, bei beliebigem Schritt abgebrochen werden kann. Problematischer ist es in dieser Hinsicht bei den beiden anderen Typen, wobei der *Typ II* der ungünstigste Typ ist, da die optimale Lösung die einzige zulässige Lösung darstellt. Bei asymptotischen Verfahren besteht sogar die theoretische Möglichkeit, dass nicht einmal das „exakte“ Optimum erreicht wird, so dass streng genommen keine zulässige Lösung vorliegt, so dass die Dualfehlerabschätzungen ihre praktische Anwendbarkeit verlieren, da sie nie im Laufe der Berechnung angewandt werden können.[13]

Außer dem oben behandelten größten Nachteil, der im gewissen Sinne mit der Dualisierung an sich verbunden ist, und daher auf *alle* vorliegenden Abschätzungen übertragen wird, ist noch auf folgende Vor- und Nachteile der einzelnen Dualfehlerabschätzungen hinzuweisen:

Der überwiegende Vorteil der Fehlerabschätzung (3. 41.) gegenüber den anderen Fehlerabschätzungen dürfte darin bestehen, dass die die einzige *Fehlerabschätzung des Primalzielfunktionswertes* darstellt. Durch die Hildrethsche Formel (3. 46.) kann lediglich eine Abschätzung des Dualproblems erzielt werden. Es ist jedoch im Allgemeinen nicht unbedingt immer möglich, aus dem Verhalten der Dualfolge eine Aussage über das Verhalten der Primalzielfunktion zu treffen. Ein anderer Nachteil der Hildrethschen Fehlerabschätzung ist darin zu sehen, dass sie im Allgemeinen nicht gegen Null konvergiert. Das hat u. a. zur Folge, dass die Differenz

$$2G\bar{u}+h)^T b-\Delta\Theta(\bar{u})$$

relativ groß ist, was empirisch nachgewiesen wurde.

Die von H. Werner hergeleitete Beziehung (3. 53.) ist keine mathematisch begründete Fehlerabschätzung, sondern mehr eine empirische Normformel, deren Anwendung jedoch unter Umständen gerechtfertigt ist. Daher unsere Bezeichnung „Abweichung“. Es ist aller4dings möglich, weitere Normformeln dieser Art, z.B.

(3. 54.) $$\sigma^k=\sum_{i=1}^{m}(\sum_{j=1}^{n}|a_{ji}|)\cdot|d_i|$$

[13] Es scheint zweckmäßig zu sein, für den Zulässigkeitstest einen Toleranzbereich anzugeben, in dem die Verletzung der Nebenbedingungen zugelassen wird.

Vorzuschlagen, die noch einfacher zu berechnen sind.[14]

Die *Vorteile* unserer Fehlerabschätzung (3. 41.) ergeben sich aus folgenden Tatsachen:

1. Sie ist die einzige Fehlerabschätzung des Primalzielfunktionswertes.
2. Der Ausdruck $2\bar{u}^T G\bar{u} + h^T \bar{u}$ konvergiert gegen Null.
3. Es ist empirisch nachgewiesen, dass die Differenz $(2\bar{u}^T G\bar{u} + h^T \bar{u}) - \Delta Q(\bar{x})$ gewöhnlich sehr klein ist.
4. Die Fehlerabschätzung (3. 41.) ist nicht verfahrensgebunden; sie kann für alle Dualverfahren der streng konvexen quadratischen Optimierung angewandt werden.

Als *Nachteile* der Fehlerabschätzung (3. 41.) sind zu nennen:

1. Die linke Ungleichung in (3. 41.) gilt nur für zulässige x. Die Tatsache, dass die rechte Ungleichung für alle x (zulässig und unzulässig) gelten muss, kann dazu führen, dass der Ausdruck $2\bar{u}^T G\bar{u} + h^T \bar{u}$ negativ wird. In diesem Falle verliert die Fehlerabschätzung (3. 41.) ihre Aussagekraft, da faktisch keine untere Schranke festgelegt werden kann.
2. Der Rechenaufwand (außer der Zulässigkeitsprobe, die bei jeder Dualabschätzung zusätzlich auftritt), verbunden mit der neuen Fehlerabschätzung, ist etwas umfangreicher als bei den anderen Fehlerabschätzungen, wobei jedoch die Ähnlichkeit zwischen Fehlerabschätzung und Dualziefunktion bezüglich ihrer mathematischen Formulierung bei der rechentechnischen Realisierung zur Verringerung des Rechenaufwandes ausgenutzt werden kann.
3. Aus der Beziehung (3. 41.) ist es nicht möglich, die Abweichung vom optimalen Programm *x direkt* abzuschätzen[15]. Das dürfte jedoch in der Ökonomie weniger interessant sein als die Abweichung vom optimalen Zielfunktionswert $Q(x^*)$.

Auf Grund der angeführten Überlegungen ist das untere Flussdiagramm aufgestellt worden. α, β und ε sind hinreichen kleine positive Zahlen. Die Größe β hängt ausschließlich vom konkreten Problem ab. ε sollte gewöhnlich kleiner als α gewählt werden. λ ist der sogenannte *Toleranzvektor*, dessen Struktur vom praktischen Modell bestimmt wird.

---

[14] Setzt man im Satz 1 des Kapitels 1 $\bar{x} = x^k$, so ergibt sich wegen (3. 47.) die vorgeschlagene Normformel (3. 54.) unmittelbar aus

$$\begin{aligned} C(x^k - x^{k+1}) &= C\left[-\frac{1}{2}C^{-1}(A^T u^k + p) + \frac{1}{2}C^{-1}(A^T u^{k+1} + p)\right] \\ &= \frac{1}{2}A^T(u^{k+1} - u^k) \\ &= \frac{1}{2}A^T d. \end{aligned}$$

[15] Die gilt übrigens auch für die anderen Fehlerabschätzungen.