Einzelschrittverfahren zur Lösung konvexer und

dual-konvexer Minimierungsprobleme

W. Oettli *)

[46]

Zusammenfassung: Es wird ein allgemeines Einzelschrittverfahren zur Lösung konvexer Minimalprobleme untersucht.

Summary: A general single-step method for solving convex minimization problems is discussed.

Mai 1973

*) Universität Mannheim, Fakultät für Mathematik und Informatik

Einzelschrittverfahren zur Lösung konvexer und dual-konvexer Minimierungsprobleme

W. Oettli

Wir untersuchen im folgenden die Verwendung von Einzelschrittverfahren zur Minimierung einer konvexen Funktion $F(x_1,\dots,x_n)$ auf einem cartesischen Produkt konvexer Mengen X_i:

$$\min \{ F(x_1,\dots,x_n) \mid x_i \varepsilon X_i, i=1,\dots,n \}.$$

Als Einzelschrittverfahren bezeichnen wir solche Verfahren, die bei jedem Schritt nur eine Komponente x_i abändern, während alle übrigen Komponenten auf ihrem zuletzt bekannten Wert festgehalten werden. Die Reihenfolge, in der die Komponenten behandelt werden, sei dabei fest vorgegeben. Bekannter Vertreter dieser Verfahrensklasse ist das Verfahren der komponentenweisen Minimierung, bei dem die Funktion in jedem Einzelschritt bezüglich einer Komponente minimiert wird. Wenn die Komponenten x_i selbst wieder mehrdimensional sind, so ist die komponentenweise Minimierung grundsätzlich von gleicher Schwierigkeit wie das Ausgangsproblem. Es liegt daher nahe, die komponentenweise Minimierung zu ersetzen durch leichter zu handhabende Rechenvorschriften. Dies führt uns zum Modell eines allgemeinen Einzelschrittverfahrens, das wir im ersten Abschnitt für die zyklische Reihenfolge und im zweiten Abschnitt für eine beliebige Reihenfolge der zu behandelnden Komponenten diskutieren. Das allgemeine Modell wird durch einige Beispiele illustriert. Die Minimierung einer konvexen Funktion auf einem konvexen cartesischen Produkt ist ein spezieller Fall eines konvexen Programms. Allgemeinere konvexe Programme lassen sich über die Dualitätstheorie auf diesen Fall zurückführen. Im dritten Abschnitt untersuchen wir daher die Bedingungen, unter denen ein konvexes Programm auf dem Umweg über das duale Programm mittels des zyklischen Einzelschrittverfahrens gelöst werden kann. Im vierten Abschnitt behandeln wir den Fall einer quadratischen Zielfunktion mit linearen Nebenbedingungen. Wir interessieren uns hier insbesondere für die Konvergenzgeschwindigkeit des Hildreth-Verfahrens zur Lösung quadratischer Programme.

1. Das zyklische Einzelschrittverfahren

Wir betrachten das folgende Minimierungsproblem:

$$(\mathrm{P}) \quad \begin{cases} \min \{F(x) \mid x \varepsilon X\}, \\ x = (x_1,\dots,x_n),\ X = X_1 \times \dots \times X_n,\ x_i \varepsilon X_i\ (i=1,\dots,n). \end{cases}$$

Hierbei ist $X_i \subseteq \mathbb{R}^{n_i}$. Wir nennen die x_i Komponenten von x. Im weiteren soll stets die folgende *Voraussetzung* erfüllt sein:

$$(\mathrm{V}) \quad \begin{cases} \text{Die Mengen } X_i \text{ seien konvex und abgeschlossen; } F \text{ sei auf } X \\ \text{konvex und differenzierbar}^{*)}. \text{ Die Menge } S^o = \{x\varepsilon X \mid F(x) \le F(x^o)\} \\ \text{sei kompakt.} \end{cases}$$

x^o wird der Startpunkt des Verfahrens sein. Wegen der Stetigkeit von F und der Kompaktheit von S^o hat (P) eine Lösung. Unter den getroffenen Konvexitäts- und Differenzierbarkeitsannahmen gelten die folgenden beiden Hilfssätze, die wohl aus der konvexen Analysis bekannt sind. Wir geben der Vollständigkeit halber den Beweis im Anhang.

Hilfssatz 1: *Es gilt*

$$(1) \qquad F(\xi)-F(\hat{x}) \ge 0 \quad \forall \xi \varepsilon X$$

genau dann, wenn

$$(2) \qquad (\xi-\hat{x})^T \nabla F(\hat{x}) \ge 0 \quad \forall \xi \varepsilon X.$$

---

*) F ist differenzierbar in $\hat{x}\varepsilon X \subseteq \mathbb{R}^n$, wenn mit einem geeignet gewählten n-Vektor t gilt $F(\xi)=F(\hat{x})+(\xi-\hat{x})^T t+|\xi-\hat{x}|\cdot\chi(\xi-\hat{x})$ $\forall \xi\varepsilon X$, $\lim_{\xi\to\hat{x}}\chi(\xi-\hat{x})=0$. Den (von $\hat{x}$ abhängigen und nicht ohne weiteres eindeutigen) Vektor t bezeichnen wir mit $\nabla F(\hat{x})$ ("Gradient"). Wenn $x=(x_1,\dots,x_n)$, so zerlegen wir entsprechend $\nabla F=(\nabla_{x_1} F,\dots,\nabla_{x_n} F)$.

Für $x = (x_1,\dots,x_i,\dots,x_n)$ setzen wir zur Abkürzung

$$x_{i-} = (x_1,\dots,x_{i-1}),\ x_{i+} = (x_{i+1},\dots,x_n),\ x = (x_{i-},x_i,x_{i+}).$$

<u>Hilfssatz 2:</u> *Aus der Gültigkeit von*

(3) $$F(\hat{x}) \le F(\hat{x}_{i-},\xi_i,\hat{x}_{i+})\ \forall \xi_i \in X_i$$

für alle i *folgt*

(4) $$F(\hat{x}) \le F(\xi)\ \forall \xi \in X.$$

Wir sagen, es sei ein Punkt $\hat{x} \in X$

minimal bezüglich der i*-ten Komponenten* (kurz: i-*minimal*),

wenn für diesen Index i Bedingung (3) erfüllt ist. Der Hilfssatz besagt also, daß unter den angegebenen Voraussetzungen ein Punkt, der minimal bezüglich aller Komponenten ist, das globale Minimum von F über X liefert, d.h. eine Lösung von (P) darstellt. (Man überlegt sich leicht, daß der Hilfssatz nicht mehr gelten muß, wenn man auf Konvexität oder Differenzierbarkeit verzichtet.)

Um das allgemeine Modell eines zyklischen Einzelschrittverfahrens für (P) zu formulieren, benötigen wir mengenwertige Abbildungen $\mathcal{N}_i(x)$, durch die jedem $x \in X$ nichtleere Teilmengen $\mathcal{N}_i(x) \subseteq X_i$ $(i=1,\dots,n)$ zugeordnet werden. Die Abbildungen $\mathcal{N}_i(x)$ seien abgeschlossen (α) und monoton (β) im folgenden Sinne:

(α) Aus $x^\nu \in X$, $x^\nu \to \bar{x} \in X$, $\xi_i^\nu \in \mathcal{N}_i(x^\nu)$, $\xi_i^\nu \to \bar{\xi}_i \in X_i$ soll folgen

$$\bar{\xi}_i \in \mathcal{N}_i(\bar{x}).$$

(β) $F(x_{i-},\xi_i,x_{i+}) = F(x)\ \forall \xi_i \in \mathcal{N}_i(x)$ genau dann, wenn x i-*minimal* ist;

$F(x_{i-},\xi_i,x_{i+}) < F(x)\ \forall \xi_i \in \mathcal{N}_i(x)$ genau dann, wenn x *nicht* i-*minimal* ist.

Das zyklische Einzelschrittverfahren verläuft dann mit beliebigem Startpunkt $x^o \varepsilon X$ wie folgt: Wenn $x^k \varepsilon X$ vorliegt, so bestimmt man der Reihe nach (zyklisch) die Komponenten von x^{k+1}, und zwar so, daß

(5) $$x_i^{k+1} \varepsilon \mathcal{N}_i(x_{i-}^{k+1}, x_i^k, x_{i+}^k) \text{ beliebig.}$$

Hierbei werden also für alle Komponenten die zuletzt bekannten Werte zugrunde gelegt. Nach Beendigung eines Zyklus, bestehend aus n Einzelschritten, liegen alle Komponenten des neuen Iterationspunktes x^{k+1} vor.

Offenbar gilt

(6) $$F(x^{k+1}) \leq F(x_{i-}^{k+1}, x_i^{k+1}, x_{i+}^k) \leq F(x_{i-}^{k+1}, x_i^k, x_{i+}^k) \leq F(x^k),$$

da wegen (β) in jedem Einzelschritt der Wert von F nicht zunimmt. Die Folge $\{x^k\}$ hat aus Kompaktheitsgründen mindestens einen Häufungspunkt.

S a t z 1: *Jeder Häufungspunkt* $\bar{x}$ *der Folge* $\{x^k\}$ *ist Lösung von* (P).

Beweis: Da die Folge $\{x^k, x^{k+1}\}$ in der kompakten Menge $S^o \times S^o$ liegt, können wir eine Teilfolge so auswählen, daß (komponentenweise)

$$x^{k_\ell} \to \bar{x} \varepsilon X, \quad x^{k_\ell + 1} \to \bar{y} \varepsilon X.$$

Aus (6) folgt dann im limes für die Teilfolge

(7) $$F(\bar{y}) \leq F(\bar{y}_{i-}, \bar{y}_i, \bar{x}_{i+}) \leq F(\bar{y}_{i-}, \bar{x}_i, \bar{x}_{i+}) \leq F(\bar{x}).$$

Da die gesamte Folge $\{F(x^k)\}$ monoton nicht wachsend ist, haben $F(x^{k_\ell + 1})$ und $F(x^{k_\ell})$ den gleichen Grenzwert, d.h. es ist

(8) $$F(\bar{y}) = F(\bar{x}).$$

Damit erhält man aus (7) insbesondere

(9) $$F(\bar{y}_{i-}, \bar{y}_i, \bar{x}_{i+}) = F(\bar{y}_{i-}, \bar{x}_i, \bar{x}_{i+}).$$

Wegen (α) gilt

(10) $$\bar{y}_i \varepsilon \mathcal{N}_i(\bar{y}_{i-}, \bar{x}_i, \bar{x}_{i+}).$$

Dies, zusammen mit (9), liefert wegen (β), daß $(\bar{y}_{i-}, \bar{x}_i, \bar{x}_{i+})$ i-minimal ist. Wegen (9) ist dann auch $(\bar{y}_{i-}, \bar{y}_i, \bar{x}_{i+})$ i-minimal, d.h., es gilt

(11) $$F(\bar{y}_{i-}, \bar{y}_i, \bar{x}_{i+}) \leq F(\bar{y}_{i-}, \xi_i, \bar{x}_{i+}) \ \forall \xi_i \varepsilon X_i.$$

Aussagen (9) und (11) gelten für $i=1,\dots,n$. Wir zeigen, daß für $j=1,\dots,n$ gilt

(12) $$F(\bar{y}_1,\dots, \bar{y}_j, \bar{x}_{j+1},\dots, \bar{x}_n) \leq F(\xi_1,\dots, \xi_j, \bar{x}_{j+1},\dots, \bar{x}_n) \quad \forall(\xi_1,\dots,\xi_j)\varepsilon X_1\times\dots\times X_j.$$

(12) ist richtig für $j=1$, denn dann ist die Aussage die gleiche, wie (11) für $i=1$. Sei (12) richtig für $j=i-1$, d.h.

$$F(\bar{y}_{i-}, \bar{x}_i, \bar{x}_{i+}) \leq F(\xi_{i-}, \bar{x}_i, \bar{x}_{i+}) \ \forall \xi_{i-} \varepsilon X_{i-}.$$

Aus (11) und (9) folgt

$$F(\bar{y}_{i-}, \bar{x}_i, \bar{x}_{i+}) \leq F(\bar{y}_{i-}, \xi_i, \bar{x}_{i+}) \ \forall \xi_i \varepsilon X_i.$$

Aus diesen beiden Aussagen folgt entsprechend Hilfssatz 2, daß

$$F(\bar{y}_{i-}, \bar{x}_i, \bar{x}_{i+}) \leq F(\xi_{i-}, \xi_i, \bar{x}_{i+}).$$

Hieraus wieder wegen (9)

$$F(\bar{y}_{i-}, \bar{y}_i, \bar{x}_{i+}) \leq F(\xi_{i-}, \xi_i, \bar{x}_{i+}).$$

Dies ist aber (12) für $j=i$. Somit gilt (12) für alle j. Setzen wir insbesondere $j=n$, so liefert (12)

$$F(\bar{y}) \leq F(\xi) \ \forall \xi \varepsilon X.$$

Damit, wegen (8),

(13) $$F(\bar{x}) \leq F(\xi) \ \forall \xi \varepsilon X,$$

und $\bar{x}$ löst (P). q.e.d.

In gleicher Weise, wie der Schluß von (9) und (10) auf (13), ergibt sich das

<u>C o r o l l a r 1:</u> x^k *ist Lösung von* (P) *genau dann, wenn* $F(x^{k+1}) = F(x^k)$.

Ist F zusätzlich streng konvex in jeder Komponente x_i, so ist das Minimum bezüglich einer Komponente eindeutig, und es ergibt sich in diesem Fall, daß x^k Lösung von (P) genau dann ist, wenn $x^{k+1}=x^k$.

Wenn $\mathcal{N}_i(x) = \{N_i(x)\}$ einelementig ist, und wenn $N_i(x)$ stetig von x abhängt, so ist (α) automatisch erfüllt, und man hat nur (β) zu verifizieren.

Man kann die Monotonieforderung (β) dahingehend verschärfen, daß man zusätzlich verlangt

$$\mathcal{N}_i(x) = \{x_i\}, \text{ wenn } x \text{ i-minimal ist}$$

(Eindeutigkeit des komponentenweisen Minimums). Unter dieser verschärften Voraussetzung gilt auch ohne Konvexitäts- und Differenzierbarkeitsannahmen, daß der Häufungspunkt $\bar{x}$ minimal bezüglich jeder Komponente ist. Aus (9) und (10) folgt dann nämlich $\bar{y}_i=\bar{x}_i$ $\forall i$; substituiert man dies in (11), so hat man bereits das gewünschte Resultat.

<u>Beispiele</u>

Wir illustrieren noch kurz das allgemeine Modell durch einige Beispiele. Hierbei sind von Fall zu Fall zusätzliche Voraussetzungen nötig.

<u>1. Zyklische Minimierung</u> [1]. $\mathcal{N}_i(x)$ sei die Menge aller ξ_i, die Lösungen sind von

$$\min\{F(x_{i-},\xi_i,x_{i+})\,|\,\xi_i \in X_i\}. \tag{14}$$

Der Punkt x_i^{k+1} ist also so zu bestimmen, daß er die Funktion $F(x_{i-}^{k+1},\xi_i,x_{i+}^k)$ über X_i minimiert. Man überzeugt sich leicht, daß (α) und (β) erfüllt sind.

2. Zyklische Linearisierung. Zusätzlich zu (V) seien die Mengen X_i kompakt, und F sei stetig differenzierbar. $\mathcal{N}_i(x)$ sei die Menge aller ξ_i, die sich durch die folgende Konstruktion erhalten lassen: Man bestimmt zuerst ein $\eta_i\varepsilon X_i$ als Lösung von

(15) $$\min\{(\eta_i-x_i)^T\nabla_{x_i}F(x)\mid \eta_i\varepsilon X_i\},$$

und hierauf einen Punkt ξ_i auf $[x_i,\eta_i]$ - der Verbindungsstrecke zwischen x_i und η_i - als Lösung von

(16) $$\min\{F(x_{i-},\xi_i,x_{i+})\mid \xi_i\varepsilon [x_i,\eta_i]\}.$$

Die Forderungen (α), (β) sind in diesem Fall erfüllt. Um (α) zu verifizieren, benützt man im wesentlichen, daß die Aussagen

$$(\eta^\nu-x^\nu)^T\nabla F(x^\nu)\le(\eta-x^\nu)^T\nabla F(x^\nu)\ \forall\eta\varepsilon X,\ F(\xi^\nu)\le F(\xi)\ \forall\xi\varepsilon [x^\nu,\eta^\nu]$$

auch im limes gültig bleiben, wenn $\eta^\nu\to\bar\eta$, $\xi^\nu\to\bar\xi$, $x^\nu\to\bar x$. Die Monotonie (β) folgt mittels Hilfssatz 1: Wenn x nicht i-minimal ist, so gilt nach Konstruktion $(\eta_i-x_i)^T\nabla_{x_i}F(x)<0$; hieraus folgt $F(x_{i-},\xi_i,x_{i+})<F(x)$. Die Beweistechnik ist die gleiche wie in [2,p.90-91], [3,p.158-62].

3. Zyklische Projektion. Es bezeichne $|\cdot|$ die euklidische Norm, P_{X_i} die euklidische Projektion auf X_i. Zusätzlich zu (V) sei F stetig differenzierbar mit $|\nabla F(x)-\nabla F(y)|\le M\cdot|x-y|\ \forall x,y\varepsilon X$. $\mathcal{N}_i(x)$ bestehe nur aus einem Element $N_i(x)$ gemäß

(17) $$N_i(x)=P_{X_i}\{x_i-\lambda_i\nabla_{x_i}F(x)\},\ \lambda_i\varepsilon(0,\tfrac{2}{M})\ \text{fest.}$$

Aus der Stetigkeit des Gradienten und des Projektionsoperators folgt die Stetigkeit von N_i; damit ist (α) erfüllt. Man zeigt nun wie in [4,p.123], daß (bei variablem λ_i) x i-minimal ist, wenn $x_i = N_i(x)\ \forall\lambda_i > 0$, während x nicht i-minimal ist, wenn $x_i \neq N_i(x)\ \forall\lambda_i > 0$. Im letzteren Fall gilt ausserdem

$$\begin{aligned} F(x_{i-}, N_i(x), x_{i+}) &\le F(x) - |x_i - N_i(x)|^2 \cdot \left(\frac{1}{\lambda_i} - \frac{M}{2} \right) \\ &< F(x), \text{ wenn } 0 < \lambda_i < \frac{2}{M} \end{aligned}$$

(dies zeigt man wie in [4,p.124]; daselbst allerdings der Faktor 2 für die maximale Schrittweite λ_i verschenkt). Damit ist auch (β) erfüllt.

<u>4. Zyklische Relaxation.</u> Es sei X der ganze $\mathbb{R}^n$, und F sei zweimal stetig differenzierbar. $H(x)$ resp. $H_i(x)$ bezeichne die Matrix der zweiten Ableitungen bezüglich x resp. x_i. Für alle $x \in S^o$ gelte mit $m > 0$: $m|y|^2 \le y^T H(x) y\ \forall y$. Wir setzen

$$(18) \quad \left\{ \begin{aligned} N_i(x) &= x_i - \lambda_i(x) \cdot t_i(x), \\ t_i(x) &= (H_i(x))^{-1} \nabla_{x_i} F(x), \quad \lambda_i(x) = \frac{\omega}{1 + \frac{\mu}{m} \cdot |t_i(x)|}, \end{aligned} \right.$$

wobei $\omega \in (0,2)$ fest, $\mu > 0$ fest.

$N_i(x)$ ist stetig; damit ist (α) erfüllt. Um (β) zu verifizieren, müssen wir weiterhin voraussetzen, daß $|H(z) - H(x)| \le \mu \cdot |t_i(x)|$ gelten soll, wenn $|z - x| \le 2 \cdot |t_i(x)|$. Offenbar ist x genau dann i-minimal, wenn $t_i(x) = 0$, d.h. wenn $N_i(x) = x_i$. Sei nun $t_i(x) \neq 0$. Mit $t_i = t_i(x)$, $\bar{x} = (x_{i-}, N_i(x), x_{i+})$, $\lambda_i(x) = \lambda_i \in (0,2)$ beliebig erhalten wir dann

$$\begin{aligned} F(\bar{x}) - F(x) &= -\lambda_i \cdot t_i^T \nabla_{x_i} F(x) + \tfrac{1}{2}\lambda_i^2 \cdot t_i^T H_i(z) t_i \qquad (z \in [x, \bar{x}]) \\ &= -\lambda_i \cdot t_i^T H_i(x) t_i + \tfrac{1}{2}\lambda_i^2 \cdot t_i^T H_i(z) t_i \\ &= (-\lambda_i + \tfrac{1}{2}\lambda_i^2) \cdot t_i^T H_i(x) t_i + \tfrac{1}{2}\lambda_i^2 \cdot t_i^T (H_i(z) - H_i(x)) t_i \\ &\le (-\lambda_i + \tfrac{1}{2}\lambda_i^2) \cdot m |t_i|^2 + \tfrac{1}{2}\lambda_i^2 \cdot \mu |t_i|^3 \end{aligned}$$

$$= m\lambda_i |t_i|^2 \cdot \left(-1 + \frac{1}{2}\lambda_i \left(1 + \frac{\mu}{m} |t_i|\right) \right)$$

$$< 0, \text{ wenn } \lambda_i < \frac{2}{1 + \frac{\mu}{m} |t_i|} \, .$$

Damit ist für (18) Forderung (β) erfüllt. Bestimmt man $\lambda_i(x)$ durch die Vorschrift

$$\lambda_i(x) = \min \left\{ 1, \frac{\omega}{1 + \frac{\mu}{m} |t_i(x)|} \right\}, \; \omega \varepsilon (1,2) \;\; \text{fest,}$$

so sind (α) und (β) immer noch erfüllt. Außerdem wird dann automatisch $\lambda_i(x) = 1$ in der Umgebung eines Punktes $\hat{x}$ mit $\nabla F(\hat{x}) = 0$, d.h. $N_i(x)$ fällt in der Umgebung eines Minimalpunkts mit dem Newton-Operator zusammen. Würde x nur aus einer Komponente bestehen, so hätte man also eine global konvergente Variante des Newton-Verfahrens. Man vergleiche hierzu auch [5,p.92-94], [6].

2. Einzelschrittverfahren mit beliebiger Ordnung

Wir betrachten nun den Fall, daß die Reihenfolge, in der die Komponenten abgeändert werden (auch *Ordnung* des Verfahrens genannt), nicht mehr zyklisch ist, sondern beliebig vorgegeben. Da es keine Zyklen mehr gibt, ist es sinnvoller, die Iterationspunkte nach den Einzelschritten zu numerieren. Wenn $i(k) \varepsilon \{1,\dots,i,\dots,n\}$ $(k = 0,1,\dots)$ eine beliebig vorgegebene Ordnung ist, so wird im k-ten Einzelschritt, der von x^k zu x^{k+1} führt, die $i(k)$-te Komponente abgeändert; alle übrigen Komponenten bleiben unverändert. Damit werden wir auf die folgende Iterationsvorschrift für das Einzelschrittverfahren mit der Ordnung $i(k)$ geführt:

(19) $\quad x_j^{k+1} = x_j^k$ für $j \neq i(k)$, $x_{i(k)}^{k+1} \varepsilon \mathcal{N}_{i(k)}(x^k)$ beliebig,

wobei $\mathcal{N}_i(x)$ wieder den Bedingungen (α) und (β) genügt. Wenn $i(k)=i$, so werden wir dafür kurz sagen, daß in x^k die i-te Komponente behandelt werde. Für das Einzelschrittverfahren mit Ordnung $i(k)$ gilt

Satz 2: *Zusätzlich zu (V) sei* F *streng konvex. Wenn* $\bar{x}$ *ein Häufungspunkt der Folge* $\{x^k\}$ *ist, so ist* $\bar{x}$ *minimal bezüglich aller Komponenten, die auf der Folge* $\{x^k\}$ *unendlich oft behandelt werden.*

Wir skizzieren die einzelnen Beweisschritte:

— $\bar{x}$ ist minimal bezüglich aller Komponenten, die auf einer gegen $\bar{x}$ konvergenten Teilfolge unendlich oft behandelt werden. Sei nämlich x^{k_ℓ} eine Teilfolge derart, daß $x^{k_\ell} \to \bar{x}$ und $i(k_\ell) = i$ für alle k_ℓ. Aus Kompaktheitsgründen können wir die Teilfolge sogar so wählen, daß $x^{k_\ell+1} \to \bar{y}$. Aus (19) folgt im limes unter Verwendung von $F(\bar{y}) = F(\bar{x})$ - vergl.(8) -, daß $\bar{y}_i \varepsilon \mathcal{N}_i(\bar{x})$, $F(\bar{x}_{i-},\bar{y}_i,\bar{x}_{i+}) = F(\bar{x})$. $\bar{x}$ ist also i-minimal.

— Wir zerlegen $x=(x_A,x_B)$, wobei x_A (x_B) diejenigen Komponenten umfaßt, bezüglich derer $\bar{x}$ minimal (nicht minimal) ist. Analog $X=X_A \times X_B$. Nach den Hilfssätzen 1 und 2 gilt

$$(20) \qquad (x_A - \bar{x}_A)^T \nabla_{x_A} F(\bar{x}) \geq 0 \quad \forall x_A \varepsilon X_A .$$

— Es gibt eine Umgebung U von $\bar{x}$ in X derart, daß in U nur solche Komponenten behandelt werden, die zugleich auch auf einer gegen $\bar{x}$ konvergenten Teilfolge behandelt werden. Bezüglich dieser Komponenten ist aber $\bar{x}$ minimal. Somit werden in U nur Komponenten aus x_A behandelt, d.h.

$$(21) \qquad x^k \varepsilon U \implies x_B^{k+1} = x_B^k .$$

— Es sei für $\theta \geq 0$

$$V_\theta = \{x \varepsilon X \mid F(x) - F(\bar{x}) \leq \theta,\ (x-\bar{x})^T \nabla F(\bar{x}) \geq -\theta\}.$$

V_o besteht wegen der strengen Konvexität von F nur aus $\bar{x}$. Die Mengen V_θ sind für $\theta > 0$ Umgebungen von $\bar{x}$ in X, die mit $\theta > 0$ gegen $V_o = \{\bar{x}\}$ kontrahieren. Wir wählen $\theta > 0$ so klein, daß

$$V_\theta \subseteq U.$$

— Wir wählen K so groß, daß für alle $k \geq K$ gilt

$$F(x^k) - F(\bar{x}) \leq \theta \tag{22}$$

(dies ist möglich, weil $F(x^k)$ monoton fallend gegen $F(\bar{x})$ strebt). Wir wählen ein $k \geq K$ derart, daß

$$x^k \varepsilon V_\theta, \tag{23}$$

$$(x_B^k - \bar{x}_B)^T \nabla_{x_B} F(\bar{x}) \geq -\theta. \tag{24}$$

— Aus $x^k \varepsilon V_\theta \subseteq U$ folgt nach (21) $x_B^{k+1} = x_B^k$. Somit folgt aus (24)

$$(x_B^{k+1} - \bar{x}_B)\nabla_{x_B} F(\bar{x}) \geq -\theta, \tag{25}$$

und aus (20)

$$(x_A^{k+1} - \bar{x}_A)\nabla_{x_A} F(\bar{x}) \geq 0.$$

Addition dieser beiden Ungleichungen liefert

$$(x^{k+1} - \bar{x})\nabla F(\bar{x}) \geq -\theta.$$

Aus (22) folgt

$$F(x^{k+1}) - F(\bar{x}) \leq \theta.$$

Somit ist auch $x^{k+1} \varepsilon V_\theta$, und x^{k+1} erfüllt die gleichen Voraussetzungen (23), (24), wie x^k.

— Wiederholung dieses Arguments liefert insbesondere, daß $x^{k+h} \varepsilon V_\theta \subseteq U$ $\forall\, h \geq 0$. Da aber in U nur Komponenten aus x_A behandelt werden dürfen, müssen alle Komponenten, die unendlich oft behandelt werden, in x_A enthalten sein; bezüglich dieser Komponenten ist also $\bar{x}$ minimal. q.e.d.

3. Anwendung auf duale Probleme

Wir untersuchen nun die Anwendbarkeit des zyklischen Einzelschrittverfahrens auf duale Programme. Gegeben sei eine reellwertige Funktion

$$\phi(x,u): X \times U \to \mathbb{R}.$$

Mittels der beiden erweitert-reellwertigen Funktionen

$$M(x) = \sup_{u\varepsilon U} \phi(x,u), \quad m(u) = \inf_{x\varepsilon X} \phi(x,u)$$

bildet man das duale Programmpaar

$$P{:}\min \{M(x) \mid x\varepsilon X,\ M(x) < \infty\},\ D{:}\ \max \{m(u) \mid u\varepsilon U,\ m(u) > -\infty\}.$$

Offensichtlich gilt stets

$$M(x) \geq \phi(x,u) \geq m(u) \ \forall x\varepsilon X,\ \forall u\varepsilon U.$$

Aus

(26) $\hat{x}\varepsilon X,\ \hat{u}\varepsilon U,\ M(\hat{x}) \leq m(\hat{u})$

folgt daher

(27) $\hat{x}$ löst P, $\hat{u}$ löst D,

sowie

(28) $\hat{x}$ löst $\min \{\phi(x,\hat{u}) \mid x\varepsilon X\}$.

Wir fragen nach Annahmen, die sicherstellen, daß D durch das zyklische Einzelschrittverfahren gelöst werden kann (d.h. daß (V) erfüllt ist, wobei $-m$ die Rolle von F übernimmt), und daß man mittels einer Lösung von D eine Lösung von P konstruieren kann.

Wir nehmen im folgenden an, daß X bzw. U im $\mathbb{R}^n$ bzw. $\mathbb{R}^m$ liegen, und daß $U = \bigtimes_i U_i$ ein cartesisches Produkt ist. Bedingung (V), übertragen auf D, lautet

$$(V_D) \begin{cases} U \text{ sei abgeschlossen und konvex. } m(u) \text{ sei auf } U \text{ endlichwer-} \\ \text{tig, konkav und differenzierbar. } S^o = \{u\varepsilon U \mid m(u) \geq \alpha\} \text{ sei} \\ \text{nichtleer und kompakt für geeignetes } \alpha\in\mathbb{R}. \end{cases}$$

Wir zeigen, daß die folgenden Annahmen (A) - (C) sicherstellen, daß (V_D) erfüllt ist.

(A): U und X seien konvex und abgeschlossen. $\phi(x,u)$ sei stetig in (x,u), konkav in u, streng konvex in x. Für alle $u\varepsilon U$ existiere $x(u)\varepsilon X$ derart, daß $m(u) = \phi(x(u),u)$.

$x(u)$, die Lösung von $\min \{\phi(x,u) \mid x\varepsilon X\}$, ist wegen der strengen Konvexität von ϕ eindeutig. Aus (A) folgt, wie im Anhang bewiesen,

<u>H i l f s s a t z 3:</u> $x(u)$ *hängt stetig von* u *ab.*

Damit ist $m(u)$ stetig auf ganz U. $m(u)$ ist ferner konkav, als Infimum einer Familie konkaver Funktionen (entsprechend ist $M(x)$ konvex, als Supremum einer Familie konvexer Funktionen).

(B): $\nabla_u\phi(x,u)$ existiere und sei stetig auf $X\times U$.

Aus (A) und (B) folgt, wie im Anhang bewiesen,

<u>H i l f s s a t z 4:</u> $m(u)$ *ist auf* U *stetig differenzierbar, und es ist* $\nabla m(u) = \nabla_u\phi(x(u),u)$.

Die folgende Bedingung (C) wollen wir die verallgemeinerte Slater-Bedingung nennen.

(C): Es existiere $(x^o,u^o)\varepsilon X\times U$ sowie $\beta\in\mathbb{R}$ derart, daß $\{u\varepsilon U \mid u^T t^o \geq \beta\}$ für $t^o = \nabla_u\phi(x^o,u^o)$ nichtleer und beschränkt ist.

Wenn (C) für einen speziellen Wert $\beta=\beta^o$ erfüllt ist, so auch für beliebiges $\beta \leq \beta^o$; dies folgt aus einem Satz von Fenchel [7], [8,p.139] über die Niveaumengen konkaver Funktionen. Aus (C) folgt wegen

$$m(u) \leq \phi(x^o,u) \leq \phi(x^o,u^o) + (u-u^o)^T t^o \quad \text{(Stützungleichung)},$$

daß S^o beschränkt ist für alle $\alpha \geq \phi(x^o,u^o) - (u^o)^T t^o + \beta$.

Damit ist gezeigt, daß die Annahmen (A) - (C) hinreichend sind für die Gültigkeit von (V_D). D hat mindestens eine Optimallösung, und das zyklische Einzelschrittverfahren kann zum Auffinden einer Lösung von D benützt werden. Es ergibt sich nun ohne weitere Annahmen, daß P eine eindeutige Optimallösung besitzt, und daß diese sehr einfach zu charakterisieren ist.

Sei $\hat{u}$ eine beliebige Optimallösung von D. Nach Hilfssatz 1, übertragen auf das Maximumproblem D, gilt

$$(u-\hat{u})^T \nabla m(\hat{u}) \leq 0 \;\; \forall u \varepsilon U,$$

somit

$$\begin{aligned}\phi(x(\hat{u}),u) &\leq \phi(x(\hat{u}),\hat{u}) + (u-\hat{u})^T \nabla_u \phi(x(\hat{u}),\hat{u}) \\ &= m(\hat{u}) + (u-\hat{u})^T \nabla m(\hat{u}) \\ &\leq m(\hat{u}) \;\; \forall u \varepsilon U,\end{aligned}$$

somit

$$M(x(\hat{u})) = \sup_{u \varepsilon U} \phi(x(\hat{u}),u) \leq m(\hat{u}).$$

Nach (26), (27) ist also $x(\hat{u})$ Optimallösung von P. Für alle Optimallösungen $\hat{x}$ von P und eine feste Optimallösung $\hat{u}$ von D gilt dann $M(\hat{x}) = m(\hat{u})$, d.h. nach (28) $\hat{x} = x(\hat{u})$. Für festes $\hat{u}$ ist aber $x(\hat{u})$ eindeutig. Somit hat P eine eindeutige Optimallösung $\hat{x}$. Wenn $m(u^n) \to \max_{u \varepsilon U} m(u)$ für eine Folge $\{u^n\} \subseteq U$, so konvergiert, wie man leicht sieht, $x(u^n) \to \hat{x}$.

Wir fassen zusammen.

<u>S a t z 3:</u> *Unter den Annahmen* (A), (B), (C) *hat* D *Optimallösungen. Das zyklische Einzelschrittverfahren ist auf* D *anwendbar.* P *hat eine eindeutige Optimallösung* $\hat{x} = x(\hat{u})$, *wo* $\hat{u}$ *irgendeine Optimallösung von* D *ist.*

Wir spezialisieren nun die vorangehenden Ergebnisse auf den Fall, daß

(29) $$\phi(x,u) = F(x)+u^T f(x),$$

U ein abgeschlossener, konvexer Kegel mit Scheitel im Nullpunkt. Hierbei ist $f: \mathbb{R}^n \to \mathbb{R}^m$. Mit dieser Festsetzung gilt offenbar für $x\varepsilon X$

$$M(x) = \begin{cases} F(x), \text{ wenn } f(x)\varepsilon U^-, \\ +\infty \text{ sonst,} \end{cases}$$

wobei $U^- = \{t \mid t^T u \leq 0 \ \forall u\varepsilon U\}$. P hat dann die Form

(30) $$P: \min \{F(x) \mid x\varepsilon X,\ f(x)\varepsilon U^-\}.$$

Forderung (A) ist offenbar erfüllt, wenn wir annehmen

(A'): X sei konvex und kompakt. f sei auf X komponentenweise konvex und stetig. F sei auf X streng konvex und stetig.

Forderung (B) ist automatisch erfüllt, $\nabla_u\phi(x,u) = f(x)$. Forderung (C) besagt dann, daß mit $t^o = f(x^o)$ die Menge $\{u\varepsilon U \mid u^T t^o \geq 0\}$ kompakt sein soll, d.h. nur aus dem Nullpunkt bestehen darf. Dies ist, wie aus der Theorie der konvexen Kegel bekannt, gleichwertig damit, daß t^o ein innerer Punkt von U^- ist. Forderung (C) erhält also folgende Form:

(C'): Es existiere $x^o\varepsilon X$ derart, daß $f(x^o)\varepsilon$ int U^-.

Dies ist die übliche Slater-Bedingung für (30), die auch bei anderen Fragestellungen in der konvexen Optimierung eine wichtige Rolle spielt. Unter (A'), (C') können wir (30) lösen, indem wir auf das Duale von (30) das zyklische Einzelschrittverfahren anwenden.

Ein Teil der Fragestellungen dieses Abschnitts findet sich auch in [9] behandelt.

4. Der quadratische Fall

Das Verfahren der zyklischen Minimierung zur Bestimmung des freien Minimums der streng konvexen quadratischen Funktion $Q(x) = \frac{1}{2} x^T G\, x + h^T x$ fällt zusammen mit dem Gauß-Seidel-Verfahren zur Lösung des linearen Gleichungssystems $\nabla Q(x) \equiv G\, x + h = 0$. Wir betrachten nun das eingeschränkte Gauß-

Seidel-Verfahren [10], d.h. die Anwendung der zyklischen Minimierung zur Lösung von

(31) $$\min \{Q(x) = \frac{1}{2} x^T G\, x + h^T x \mid \mu \le x\}.$$

Hierbei sei $x = (x_1,\dots,x_n) \in \mathbb{R}^n$, $G = G^T$ positiv-semidefinit. Die Menge $\{x \ge \mu \mid Q(x) \le Q(x^o)\}$ sei beschränkt. Für die Diagonalelemente von G gelte außerdem noch $g_{ii} > 0$. Aus Hilfssatz 1 folgt leicht als notwendige und hinreichende Bedingung dafür, daß $x \ge \mu$ i-minimal ist, die Gültigkeit von

(32) $$\frac{\partial Q(x)}{\partial x_i} \equiv \sum_j g_{ij}\, x_j + h_i \begin{cases} = 0, \text{ wenn } x_i > \mu_i \\ \ge 0, \text{ wenn } x_i = \mu_i \,. \end{cases}$$

Das Verfahren der zyklischen Minimierung bestimmt x_i^{k+1} so, daß

(33) $(x_{i-}^{k+1}, x_i^{k+1}, x_{i+}^k)$ i-minimal ist.

Das ergibt

$$x_i^{k+1} = N_i(x_{i-}^{k+1}, x_i^k, x_{i+}^k)$$

mit

(34) $$N_i(x) = \max\{\mu_i, \rho_i(x)\}, \ \rho_i(x) = -(g_{ii})^{-1} \Big(\sum_{j \ne i} g_{ij} x_j + h_i\Big).$$

Es ist nämlich $\frac{\partial}{\partial x_i} Q(x_{i-}, \rho_i(x), x_{i+}) = 0$. Ist nun $\rho_i(x) < \mu_i$, so gilt wegen der Monotonie von $\frac{\partial Q}{\partial x_i}$, daß $\frac{\partial}{\partial x_i} Q(x_{i-}, N_i(x), x_{i+}) \ge 0$. Damit erfüllt $(x_{i-}, N_i(x), x_{i+})$ die Bedingung (32), x_i^{k+1} also (33). Setzen wir noch voraus, daß (31) eine eindeutige Optimallösung $\hat{x}$ hat, so gilt für dieses Verfahren

<u>S a t z 4:</u> *Es gibt eine Konstante* $\rho < 1$ *derart, daß*

$$Q(x^{k+1}) - Q(\hat{x}) \le \rho \cdot \big(Q(x^k) - Q(\hat{x})\big) \ \forall k.$$

Wir skizzieren den Beweis:

— Durch eine Translation und Addition einer Konstanten zur Zielfunktion können wir, ohne das Konvergenzverhalten zu ändern, erreichen, daß $\hat{x} = 0$, $Q(\hat{x}) = 0$. Die Optimalität von $\hat{x} = 0$ bedingt nach (32), daß

(35) $\mu_i \leq 0,\ h_i \geq 0,\ \mu_i \cdot h_i = 0\ (i = 1,\ldots,n).$

— Für diejenigen i mit $h_i > 0$, $\mu_i = 0$ gilt $\rho_i(0) < 0$, somit aus Stetigkeitsgründen $N_i(x) = \mu_i = 0$ für alle x in einer gewissen Umgebung des Nullpunkts. Da x^k gegen $\hat{x} = 0$ konvergiert, folgt für diese Komponenten $x_i^k = 0\ \forall k \geq K$. Wir können uns diejenigen Komponenten, die für $k \geq K$ stets den Wert Null haben, von vornherein ausgeschieden denken. In den verbleibenden Komponenten ist wegen $h_i = 0$ die Funktion Q quadratisch-homogen um $\hat{x} = 0$:

$$Q(\lambda t) = \lambda^2 Q(t)\ \forall\lambda.$$

Ferner ist $\rho_i(x)$ rein linear.

— Für diejenigen i mit $h_i = 0$, $\mu_i < 0$ gilt $\rho_i(0) = 0 > \mu_i$, somit aus Stetigkeitsgründen $N_i(x) = \rho_i(x)$ für alle x in einer gewissen Umgebung des Nullpunkts. Die Nebenbedingung $x_i \geq \mu_i$ kommt also für $k \geq K$ gar nicht mehr zum Tragen. Wir können uns daher für diese i von vornherein $\mu_i = -\infty$ gesetzt denken. Sei $\tilde{X}$ der derart erweiterte zulässige Bereich.

Für diejenigen i mit $h_i = 0$, $\mu_i = 0$ gilt $N_i(x) = \max\{0, \rho_i(x)\}$. Somit gilt für die verbleibenden Komponenten entweder $N_i(x) = \rho_i(x)$ oder $N_i(x) = \max\{0, \rho_i(x)\}$, wobei ρ_i rein linear. Hieraus folgt $N_i(\lambda t) = \lambda N_i(t)$ $\forall\lambda \geq 0$.

— Es sei $T_i(x) = (x_{i-}, N_i(x), x_{i+})$, $T = T_n T_{n-1} \cdots T_1$. Dann ist $x^{k+1} = T(x^k)$, und es gilt auch

$$T(\lambda t) = \lambda T(t)\ \forall\lambda \geq 0.$$

Entsprechend Corollar 1 gilt $Q(T(x)) < Q(x)\ \forall x \neq \hat{x}$.

— Sei

$$S = \{t \varepsilon \tilde{X} \mid Q(t) = 1\}.$$

Es gilt, weil S nicht die Optimallösung $\hat{x}$ enthält, und weil T stetig und S kompakt ist,

$$Q(T(t)) \le \rho < 1 \ \forall t \in S.$$

Jedes $x \in \tilde{X}$ läßt sich in der Form $x = \lambda t,\ t \in S,\ \lambda \ge 0$ darstellen. Damit

$$Q(x) = Q(\lambda t) = \lambda^2 Q(t) = \lambda^2 \ \forall x \in \tilde{X},$$

$$Q(T(x)) = Q(T(\lambda t)) = \lambda^2 Q(T(t)) \le \lambda^2 \rho \ \forall x \in \tilde{X},$$

$$Q(T(x)) \le \rho \cdot Q(x) \ \forall x \in \tilde{X},$$

$$Q(x^{k+1}) \le \rho \cdot Q(x^k).$$

Wegen $Q(\hat{x}) = 0$ folgt die Behauptung des Satzes. q.e.d.

Wir wollen nun Satz 4 auf das Hildreth-Verfahren der quadratischen Programmierung anwenden. Dazu setzen wir in (29)

$$X = \mathbb{R}^n,\ U = \{u \in \mathbb{R}^m \mid u \ge 0\},\ \phi(x,u) = \frac{1}{2} x^T C x + d^T x + u^T (Ax+b),$$

wobei $C = C^T$ positiv definit. (30) wird dann zu

(36) $\quad$ P: $\min \{F(x) = \frac{1}{2} x^T C x + d^T x \mid Ax+b \le 0\}$.

Forderung (A) des vorangehenden Abschnitts ist erfüllt; $x(u)$ existiert für alle $u \ge 0$ und wird durch Nullsetzen des Gradienten gefunden, $\nabla_x \phi(x,u) \equiv Cx+d+A^T u = 0$. Das ergibt

$$x(u) = -C^{-1}(A^T u + d).$$

Einsetzen liefert

$$m(u) = \phi(x(u),u) = -\frac{1}{2} u^T G u - h^T u - \alpha,$$

wobei

$$G = AC^{-1}A^T,\ h = AC^{-1}d - b,\ \alpha = \frac{1}{2} d^T C^{-1} d.$$

Somit lautet D, bis auf eine additive Konstante in der Zielfunktion,

(37) $\quad$ D: $\max \{Q(u) = -\frac{1}{2} u^T G u - h^T u \mid u \ge 0\}$.

Die Matrix G ist positiv semidefinit, mit $g_{ii} = a_i^T C^{-1} a_i > 0$. Bedingung (C') fordert die Existenz von x^o derart, daß $Ax^o+b < 0$. Wenn sie erfüllt ist, so gilt Satz 3, d.h. man kann auf (37) das Verfahren der zyklischen Maximierung anwenden (das in diesem Fall das eingeschränkte Gauß-Seidel-Verfahren ist), und $x(u^k)$ konvergiert gegen die Lösung von (36). Dies ist das Verfahren von Hildreth und D'Esopo zur Lösung von (36) [11,p.73-77]. Ist $\hat{u}$ eindeutig, so gilt auch Satz 4, und das Verfahren konvergiert linear. Damit haben wir eine Antwort auf die Frage nach der Konvergenzgeschwindigkeit des Hildreth-Verfahrens, die in [12] nur näherungsweise beantwortet wurde. $\hat{u}$ ist übrigens eindeutig, wenn diejenigen Zeilen von A, für die $a_i^T\hat{x}+b_i = 0$ gilt, linear unabhängig sind.

Anhang

Beweis von Hilfssatz 1: a) Sei (2) erfüllt. Wenn $F(\xi)-F(\hat{x}) \leq \leq -\varepsilon < 0$, so folgt hieraus mit $x_\lambda = \hat{x}+\lambda(\xi-\hat{x})$ $(0 < \lambda \leq 1)$ unter Verwendung der Konvexität von F, der Differenzierbarkeit von F, und wegen (2)

$$- \lambda\varepsilon \geq \lambda \cdot \left[F(\xi) - F(\hat{x})\right] \geq F(x_\lambda) - F(\hat{x}) = \lambda(\xi - \hat{x})^T \nabla F(\hat{x}) + \lambda \cdot \chi(\lambda)$$
$$\geq \lambda \cdot \chi(\lambda).$$

Damit ergibt sich $\chi(\lambda) \leq -\varepsilon$, ein Widerspruch für genügend kleines $\lambda > 0$. Somit gilt (1). b) Sei (1) erfüllt. Wenn $(\xi-\hat{x})^T \nabla F(\hat{x}) \leq -\varepsilon < 0$, so folgt unter Verwendung der Differenzierbarkeit von F

$$0 \leq F(x_\lambda) - F(\hat{x}) = \lambda(\xi-\hat{x})^T \nabla F(\hat{x}) + \lambda \cdot \chi(\lambda)$$
$$\leq - \lambda\varepsilon + \lambda \cdot \chi(\lambda).$$

Damit ergibt sich $\chi(\lambda) \geq \varepsilon$, ein Widerspruch für genügend kleines $\lambda > 0$. Somit gilt (2).

Beweis von Hilfssatz 2: Aus (3) folgt nach Hilfssatz 1 als notwendige Bedingung $(\xi_i - \hat{x}_i)^T \nabla_{x_i} F(\hat{x}) \geq 0 \ \forall \xi_i \varepsilon X_i$ für alle i. Addition dieser Ungleichungen liefert $(\xi-\hat{x})^T \nabla F(\hat{x}) \geq 0 \ \forall \xi \varepsilon X$, und dies ist nach Hilfssatz 1 hinreichend für (4).

Beweis von Hilfssatz 3: Sei $\bar{u} \varepsilon U$, $\bar{x} = x(\bar{u})$, $\theta > 0$ beliebig. Wir definieren

$$S = \{x \varepsilon X \mid |x-\bar{x}| \leq \theta\}, \quad \partial S = \{x \varepsilon X \mid |x-\bar{x}| = \theta\}.$$

∂S sei zunächst nicht leer. Wegen der Abgeschlossenheit von X ist ∂S kompakt. Wegen der Eindeutigkeit von $x(\bar{u})$, der Stetigkeit von ϕ, und der Kompaktheit von ∂S folgt

$$\phi(\xi,\bar{u}) - \phi(\bar{x}, \bar{u}) \geq \eta > 0 \ \forall \xi \varepsilon \partial S.$$

Die Funktion $\psi(u) = \min\limits_{\xi\varepsilon\partial S} \big(\phi(\xi,u) - \phi(\bar{x},u)\big)$ ist stetig auf U. Wegen $\psi(\bar{u}) \geq \eta > 0$ gibt es eine Umgebung $V(\bar{u})$ derart, daß $\psi(u) \geq 0$ $\forall u\varepsilon V(\bar{u})$. Für beliebiges $u\varepsilon V(\bar{u})$ gilt also $\phi(\xi,u) \geq \phi(\bar{x},u)$ $\forall\xi\varepsilon\partial S$. Hieraus folgt aber

(38) $$\phi(x,u) > \phi(\bar{x},u) \ \forall x\varepsilon X\backslash S.$$

Gäbe es nämlich ein $x\varepsilon X\backslash S$ mit $\phi(x,u) \leq \phi(\bar{x},u)$, so wäre, wegen der strengen Konvexität von ϕ, $\phi(\xi,u) < \phi(\bar{x},u)$ für alle ξ auf der offenen Strecke $(\bar{x},x)$, insbesondere also für ein ξ aus ∂S. Aus (38) folgt, daß $x(u)\varepsilon S$. Wenn also $u\varepsilon V(\bar{u})$, so gilt $|x(u) - x(\bar{u})| \leq \theta$. Das gleiche Resultat ergibt sich, wenn ∂S leer ist; dann ist nämlich $X \subseteq S$. Somit ist $x(u)$ stetig.

Beweis von Hilfssatz 4: Wegen der Konkavität von ϕ in u gilt die Stützungleichung

$$\phi(x,\eta) - \phi(x,u) \leq (\eta-u)^T\nabla_u\phi(x,u) \ \forall\eta\varepsilon U.$$

Hieraus folgt für beliebiges $u\varepsilon U$

$$\begin{aligned} m(\eta) - m(u) &= \inf_{x\varepsilon X} \phi(x,\eta) - \phi(x(u),u) \leq \phi(x(u),\eta) - \phi(x(u),u) \\ &\leq (\eta-u)^T\nabla_u\phi(x(u),u) \ \forall\eta\varepsilon U. \end{aligned}$$

Wir setzen zur Abkürzung $t(u) = \nabla_u\phi(x(u),u)$. Dann gilt also

(39) $$m(\eta) - m(u) \leq (\eta-u)^T t(u) \ \forall\eta\varepsilon U.$$

Vertauschung von η und u und Anwendung der Schwarz'schen Ungleichung liefert hieraus

$$\begin{aligned} m(u) - m(\eta) &\leq (u-\eta)^T t(\eta) \\ &\leq (u-\eta)^T t(u) + |u-\eta|\cdot|t(\eta)-t(u)|, \end{aligned}$$

d.h.

(40) $$m(\eta) - m(u) \geq (\eta-u)^T t(u) + |\eta-u|\cdot\chi(\eta-u),$$

wobei $\lim\limits_{\eta\to u} \chi(\eta-u) = 0$ wegen der Stetigkeit von t. Aus (39) und (40) folgt, daß m differenzierbar ist, und $\nabla m(u) = t(u)$.

Literatur

[1] WARGA, J., Minimizing certain convex functions. SIAM J. Appl. Math. 11 (1963), 588-593.

[2] BERGE, C., GHOUILA-HOURI, A., Programmes, jeux et réseaux de transport. Dunod, Paris, 1962.

[3] ZANGWILL, W.I., Nonlinear Programming. Prentice-Hall, Englewood Cliffs, N.J., 1969.

[4] GOLDSTEIN, A.A., Constructive Real Analysis. Harper & Row, New York, 1967.

[5] CÉA, J., Optimisation. Dunod, Paris, 1971.

[6] SCHECHTER, S., Minimization of a convex function by relaxation. In: Integer and Nonlinear Programming (J. Abadie ed.), pp. 177-189. North-Holland, Amsterdam, 1970.

[7] FENCHEL, W., Über konvexe Funktionen mit vorgeschriebenen Niveaumannigfaltigkeiten. Math. Z. 63 (1956), 496-506.

[8] STOER, J., WITZGALL, C., Convexity and Optimization in Finite Dimensions I. Springer, Berlin, 1970.

[9] GEBAUER-WALDMEIER, V., Beiträge zur zyklischen Relaxation in der konvexen Programmierung. Dissertation, Universität Zürich, 1971.

[10] CÉA, J., Recherche numérique d'un optimum dans un espace produit. In: Colloquium on Methods of Optimization (Lecture Notes in Mathematics, 112), pp. 33-50. Springer, Berlin, 1970.

[11] KÜNZI, H.P., KRELLE, W., Nichtlineare Programmierung. Springer, Berlin, 1962.

[12] WENDLER, K., Beiträge zum Hildreth-Verfahren und zur ε-Störung quadratischer und linearer Optimierungsaufgaben. Dissertation, Universität Bonn, 1969.

Anschrift: Prof. Dr. W. Oettli, Lehrstuhl für Mathematik VII, Universität Mannheim, D-68 Mannheim, Schloss.