Aplikace matematiky

Gerhard Donath; Karl-Heinz Elster

Zur Konvergenz des Verfahrens der Koordinatenweisensuche

Aplikace matematiky, Vol. 23 (1978), No. 3, 161–173

Persistent URL: http://dml.cz/dmlcz/103742

ZUR KONVERGENZ DES VERFAHRENS DER KOORDINATENWEISENSUCHE

GERHARD DONATH und KARL-HEINZ ELSTER

(Eingegangen 16. Januar 1976)

1. EINLEITUNG

Zur Lösung restriktionsfreier nichtlinearer Optimierungsprobleme

$$\min\{f(x) \mid x \in R^n\}, \tag{1.1}$$

d. h. von Problemen zur Bestimmung eines Punktes $x^* \in R^n$ mit

$$f(x^*) \leqq f(x) \; \forall x \in R^n,$$

wurden in den letzten Jahren in zunehmendem Maße direkte oder ableitungsfreie Verfahren entwickelt. Sie besitzen gegenüber Gradientenverfahren oder Verfahren mit höheren Ableitungen den Vorteil des Wegfalls der mitunter komplizierten Bestimmung dieser Ableitungen bzw. der meist mit empfindlichen numerischen Fehlern behafteten finiten Approximation der Werte dieser Ableitungen.

Wir führen die Lösungsmenge

$$G^* =_{\mathrm{Df}} \{x^* \in R^n \mid f(x^*) \leqq f(x) \; \forall x \in R^n\}$$

ein. Von Verfahren zur Lösung von (1.1) wird Konvergenz in folgendem Sinne verlangt (vgl. Zangwill [12]):

Definition 1. *Gegeben sei das Problem* (1.1). *Ein Verfahren zur Ermittlung eines Punktes $x^* \in G^*$ mittels einer Folge $(x^k) \subset R^n$ heißt konvergent, wenn die folgenden Bedingungen erfüllt sind:*

a) $x^k \in G^* \Rightarrow x^{k+1} \in G^* \; \forall k \geqq 1$.

b) $x^k \notin G^* \; \forall k \geqq 1 \Rightarrow \begin{cases} \text{Für jede konvergente Teilfolge} \\ (x^{k_r}) \subset (x^k) \text{ gilt } \lim_{r \to \infty} x^{k_r} \in G^* . \end{cases}$

Existieren keine konvergenten Teilfolgen von (x^k), so ist $G^* = \emptyset$.

Leider fehlen in der Regel sowohl in den Originalarbeiten über direkte Verfahren als auch in den meisten in den letzten Jahren erschienenen Monographien Bedingungen für die Konvergenz im Sinne von Definition 1, oder aber es werden nicht hinreichende bzw. zu starke Bedingungen dafür angegeben. Obwohl dies in vielen Fällen für praktische Bedürfnisse ausreichen mag, besteht die allgemeine Aufgabe der Ermittlung von Konvergenzbedingungen für direkte Verfahren.

Im Falle $n = 1$ läßt sich für das Problem (1.1) sogar ein effektivstes Verfahren in dem Sinne angeben, daß die Anzahl K der Funktionswertberechnungen minimal ist, wobei das Verhältnis zwischen der Länge eines x^* enthaltenden Ausgangsintervalls und einer Abbruchschranke $\delta \geqq |x^k - x^*|$ als gegeben angenommen wird (vgl. Kiefer [9]). In [9] werden dabei jedoch stärkere Konvergenzbedingungen als z. B. bei Donath/Elster [2] benötigt.

Im Falle $n > 1$ ist u.a. das Pattern-Search-Verfahren von Hooke/Jeeves [7] anwendbar; Bedingungen für seine Konvergenz wurden von Donath/Elster [3] angegeben.

In der vorliegenden Arbeit werden Voraussetzungen für die Konvergenz des Verfahrens der koordinatenweisen Suche von Friedman/Savage [5] mitgeteilt.

Dieses Verfahren, im folgenden mit (KS) bezeichnet, erzeugt eine Punktfolge $(x^k) \subset R^n$ nach folgender Vorschrift:

$$(1.2) \qquad x^1 \in R^n\,, \quad \text{beliebig}\,;$$

$$x^{k+1} \in L_k \quad \text{mit} \quad f(x^{k+1}) = \min\{f(x) \mid x \in L_k\}\,,$$

$$(1.3) \qquad L_k = L(x^k) =_{\mathrm{Df}} \{x \in R^n \mid x_i = x_i^k\; \forall i \not\equiv m \bmod n\}\,;$$

$$f(x^{k+1}) = f(x^k) \Rightarrow x^{k+1} = x^k\; \forall k \geqq 1\,.$$

Weil die L_k Geraden durch x^k darstellen, die parallel zur x_j-Achse ($j \equiv k \bmod n$) verlaufen, bietet sich zur Bestimmung von x^{k+1} gemäß (1.2) ein eindimensionales direktes Verfahren an.

2. EIGENSCHAFTEN DER ZIELFUNKTION

Von Kiefer [9] wurden im Zusammenhang mit direkten Verfahren im eindimensionalen Fall sogen. unimodale Funktionen verwendet. Der von uns im Abschnitt 3. angegebene Konvergenzbeweis für das Verfahren (KS) benötigt Eigenschaften der Zielfunktion, die als Verallgemeinerungen des Begriffes der Unimodalität betrachtet werden können. Dazu nennen wir zunächst:

Definition 2. *Eine Funktion $f: R \to R$ heißt quasi-unimodal (nach unten), wenn es ein nicht leeres (offenes, halboffenes oder abgeschlossenes) Intervall $I^* \subseteq R$ mit*

$c =_{\mathrm{Df}} \inf \{x \mid x \in I^*\} \geqq -\infty$ *und* $d =_{\mathrm{Df}} \sup \{x \mid x \in I^*\} \leqq +\infty$ *derart gibt, daß gilt:*

$$(2.1) \qquad f(x) \begin{cases} \textit{streng fallend auf } \{x \in R \mid x \leqq c\}, \\ = \text{const.} \leqq \min \{f(c), f(d)\} \textit{ auf } I^*, \\ \textit{streng wachsend auf } \{x \in R \mid x \geqq d\}. \end{cases}$$

Man erkennt leicht, daß f auf I^* das globale Minimum annimmt. Außerhalb I^* kann es wegen der strengen Monotonie von f kein lokales, aber nicht globales Minimum geben. Als unmittelbare Folgerung aus (2.1) erhält man

Hilfssatz 1. *Seien $f: R \to R$ quasi-unimodal und $\bar{x} \in R$ ein beliebiger Punkt. Dann gilt:*

1. *f nicht streng wachsend für alle $x \geqq \bar{x} \Rightarrow f$ fallend für alle $x \leqq \bar{x}$.*
2. *f nicht streng fallend für alle $x \leqq \bar{x} \Rightarrow f$ wachsend für alle $x \geqq \bar{x}$.*

Zum Beweis verifiziert man beide Implikationen für die drei Fälle $\bar{x} \leqq c$, $\bar{x} \in I^*$ und $\bar{x} \geqq d$.

Auf Funktionen $f: R^n \to R$ $(n \geqq 2)$ läßt sich der Begriff der Quasi-Unimodalität in folgender Weise übertragen:

Definition 3. *Eine Funktion $f: R^n \to R$ heißt linear quasi-unimodal (nach unten), wenn f auf jeder Geraden $L(x^1, x^2) =_{\mathrm{Df}} \mathrm{aff}\{x^1, x^2\} = \{x \in R^n \mid x = x(\lambda) = \lambda x^1 + (1-\lambda) x^2, \lambda \in R\}$ mit $x^1, x^2 \in R^n$ quasi-unimodal in dem Sinne ist, daß*

$$F(\lambda) =_{\mathrm{Df}} f(x(\lambda)) \textit{ der Definition 2 genügt.}$$

Diese Definition verallgemeinert den Begriff der linearen Unimodalität von Funktionen (vgl. Donath/Elster [1]).

Damit wird f wie folgt in die Menge der verallgemeinert konvexen Funktionen eingeordnet:

Satz 1. *Sei $f: R^n$. Dann gilt:*

$$\left.\begin{matrix} f \textit{ linear} \\ \textit{quasi-unimodal} \end{matrix}\right\} \Rightarrow \left\{\begin{matrix} f \textit{ explizit} \\ \textit{quasikonvex.} \end{matrix}\right.$$

Beweis. Für linear quasi-unimodale Funktionen f gilt:

$$(2.2) \qquad \left.\begin{matrix} x^1, x^2 \in R^n, \\ f(x^1) > f(x^2), \\ \lambda \in (0,1) \end{matrix}\right\} \Rightarrow f(\lambda x^1 + (1-\lambda) x^2) < f(x^1) \,^{*)},$$

außerdem

$$\left.\begin{matrix} x^1, x^2 \in R^n, \\ f(x^1) = f(x^2), \\ \lambda \in [0,1] \end{matrix}\right\} \Rightarrow f(\lambda x^1 + (1-\lambda) x^2) \leqq f(x^1),$$

---

*) Diese Eigenschaft von f nennen wir streng quasikonvex. Man beachte, daß daraus nicht die Quasikonvexität folgt (vgl. das Beispiel bei Karamardian [8]).

mithin erst recht

$$\left.\begin{array}{l} x^1, x^2 \in R^n, \\ f(x^1) \geqq f(x^2), \\ \lambda \in [0, 1] \end{array}\right\} \Rightarrow f(\lambda x^1 + (1 - \lambda) x^2) = f(x^1),$$

d. h. f ist quasikonvex. Zusammen mit (2.2) folgt hieraus die Behauptung des Satzes.

Die Umkehrung von Satz 1 gilt nicht, da explizit quasikonvexe Funktionen ihr Minimum nicht einmal auf kompakten Mengen annehmen müssen. Sie ist jedoch unter der Generalprämisse der Existenz einer lokalen Minimumstelle möglich, da dort auch das globale Minimum angenommen wird (vgl. Greenberg/Pierskalla [6]). Ferner gilt:

Satz 2. *Sei $f: R^n \to R$ halbstetig nach unten. Dann gilt*

$$f \text{ streng quasikonvex} \Rightarrow f \text{ quasikonvex}.$$

Beweis. Angenommen, f sei nicht quasikonvex. Es genügt dann offenbar, den Fall x^1, $x^2(\neq x^1) \in R^n$ mit $f(x^1) = f(x^2)$ zu betrachten, wobei gleichzeitig eine Zahl $\lambda^* \in (0, 1)$ existiert, so daß $f(\lambda^* x^1 + (1 - \lambda^*) x^2) > f(x^1)$ ist. Wegen der strengen Quasikonvexität von f muß dann gelten

$$f(x) = f(x^1) \quad \forall x \in [x^1, x^2] \setminus \{\lambda^* x^1 + (1 - \lambda^*) x^2\}.$$

Dann ist aber f in $\lambda^* x^1 + (1 - \lambda^*) x^2$ nicht halbstetig nach unten im Widerspruch zur Voraussetzung.

Nach einem bekannten Satz der Analysis nehmen nach unten halbstetige Funktionen auf einer kompakten Menge das globale Minimum an. Deshalb sind unter der Voraussetzung der Halbstetigkeit nach unten für eine konvexe und kompakte Menge $G \subset R^n$ die Aussageformen

„f ist auf G linear quasi-unimodal",

„f ist auf G explizit quasikonvex" und

„f ist auf G streng quasikonvex"

äquivalent.

Ferner ist auch jede pseudokonvexe Funktion streng quasikonvex auf R^n (vgl. etwa [4]), und die bekannte notwendige Bedingung $\nabla f(x^*) = 0$ für Minimumstellen $x^* \in R^n$ mit $f(x^*) \leqq f(x)\ \forall x \in R^n$ ist dann hinreichend.

3. KONVERGENZBEDINGUNGEN

Wir wenden nun die im Abschnitt 2. beschriebenen Eigenschaften von Funktionen als Konvergenzbedingungen an. Soll ein eindimensionales direktes Verfahren wie etwa die Fibonacci-Suche (vgl. [9]) zur Bestimmung der x^{k+1} gemäß (1.2) eingesetzt werden, so ist nach [2] Quasi-Unimodalität von f für die Konvergenz dieses Verfahrens hinreichend. Für das Verfahren (KS) ist die lineare Quasi-Unimodalität von f in folgendem Sinne eine notwendige Konvergenzbedingung:

Satz 3. *Wenn das Verfahren (KS) unabhängig von der Wahl der Koordinatenachsenrichtungen und des Startpunktes x^1 konvergiert, so ist f linear quasi-unimodal.*

Beweis. Sei f nicht linear quasi-unimodal. Dann gibt es eine Gerade $L(x', x'')$, auf der f nicht quasi-unimodal ist. Wir setzen $x^1 =_{\mathrm{Df}} x'$ und wählen außerdem die Koordinatenachsen so, daß mit einer gewissen reellen Zahl $\mu \neq 0$ gilt:

$$x'' - x' = \mu e^1 ,$$

wobei e^i den i-ten Einheitsvektor des R^n bezeichnet. Offenbar ist dann f auf L_1 nicht quasi-unimodal. Dann ist bereits das angewendete eindimensionale direkte Verfahren nicht konvergent, denn es kann damit ein Punkt x^2 bestimmt werden, zu dem ein Punkt $\bar{x} \in L_1$ existiert, so daß gilt:

$$f(\bar{x}) < f(x^2) .$$

Dann wird die Vorschrift (1.2) zur Bildung der Folge (x^k) verletzt. Folglich ist die Annahme einer nicht linear quasi-unimodalen Funktion f falsch.

Wilde/Beightler [11] erwähnen bei der Beschreibung des Verfahrens (KS), daß es Funktionen f gibt, bei denen einerseits die Mengen

$$G_c =_{\mathrm{Df}} \{x \in R^n \mid f(x) < c\} , \quad c \in R ,$$

konvex sind, aber andererseits das Verfahren (KS) nicht konvergiert. Die Konvexität von G_c charakterisiert bekanntlich die quasikonvexen Funktionen. In der Tat gibt es quasikonvexe Funktionen, die nicht linear quasi-unimodal sind. Die Ursache, weshalb aber selbst lineare Quasi-Unimodalität von f nicht hinreichend für die Konvergenz des Verfahrens (KS) ist, liegt jedoch darin, daß stärkere generelle Voraussetzungen über die Funktionenklasse gemacht werden müssen, der f angehört. Enthält nämlich für einen gewissen Punkt $k^n \in R^n$ die Menge $G_{f(x^k)}$ keinen Punkt der Form $x^k + \lambda e^i$, $i \in \{1, \ldots, n\}$, $\lambda \neq 0$, reell, so hat man gemäß (1.2)

$$x^l = x^k \, \forall l < k .$$

Damit ist $\lim_{l \to \infty} x^l = x^k$. Aber auch bei linear quasi-unimodalen Funktionen f folgt daraus nicht unbedingt

$$x^k \in G^* ,$$

wie es bei Konvergenz von (KS) zwingend wäre. Ist dagegen f stetig differenzierbar und sind die Niveauhyperflächen

$$A(x^k) =_{\mathrm{Df}} \{x \in R^n \mid f(x) = f(x^k)\}$$

für alle $k \geqq 1$ glatt, so gilt

$$\left.\begin{array}{l} x^k + \lambda e^i \notin G_{f(x^k)}, \\ \forall \lambda \neq 0, \\ \forall i \in \{1, \ldots, n\} \end{array}\right\} \Rightarrow x^k \in G^* .$$

Für diese Eigenschaft ist hinreichend:

$$\nabla f(x) \neq 0 \quad \forall x \in R^n \setminus G^* .$$

Diese Bedingung trifft, wie am Ende von Abschnitt 2. festgestellt wurde, auf pseudokonvexe Funktionen zu, falls $G^* \neq \emptyset$ ist. Sie sind dann auch streng quasikonvex und damit linear quasi-unimodal. Es gilt daher:

Satz 4. *Wenn $f: R^n \to R$ stetig differenzierbar und pseudokonvex ist und es eine kompakte Menge $K \subset R^n$ mit $(x^k) \subset K$ gibt, dann konvergiert das Verfahren (KS).*

Beweis. 1. Sei $x^k \in G^*$. Dann gibt es kein $x \in L_k$ mit $f(x) < f(x^k)$. Also gilt unter Berücksichtigung von (1.2) auch $x^{k+1} = x^k \in G^*$ gemäß (1.3).
2. Sei

$$x^k \notin G^* \quad \forall_k \geqq 1 . \tag{3.1}$$

Aus der Kompaktheit von K folgt die Existenz einer konvergenten Teilfolge $(x^{k_r}) \subset (x_\gamma)$ mit

$$\lim_{r \to \infty} x^{k_r} = x' \in K .$$

Nach Zangwill [12], Lemma 4.1, folgt aus der Stetigkeit von f und

$$f(x^{k+1}) \leqq f(x^k) \quad \forall_k \geqq 1 , \tag{3.2}$$

daß gilt:

$$\lim_{k \to \infty} f(x^k) = f(x') . \tag{3.3}$$

Angenommen, es sei $x' \notin G^*$, d. h.

$$\|\nabla f(x')\| > 0 .$$

Dann existiert ein $i \in \{1, \ldots, n\}$ mit

$$\left|\frac{\partial f(x')}{\partial x_i}\right| = \min \left\{\left|\frac{\partial f(x')}{\partial x_j}\right| \neq 0 \,\middle|\, j \in \{1, \ldots, n\}\right\} =_{\mathrm{Df}} \sigma > 0 . \tag{3.4}$$

Wegen der stetigen Differenzierbarkeit von f gibt es für alle $\varepsilon > 0$ ein $r_0 = r_C(\varepsilon) \geqq 1$, so daß für alle $j \in \{1, \ldots, n\}$ gilt:

$$\left|\frac{\partial f(x')}{\partial x_j} - \frac{\partial f(x^{k_r})}{\partial x_j}\right| < \varepsilon \quad \forall r \geqq r_0 . \tag{3.5}$$

Es gilt aber auch

$$\left|\frac{\partial f(x')}{\partial x_j} - \frac{\partial f(x^{k_r})}{\partial x_j}\right| \geqq \left|\frac{\partial f(x')}{\partial x_j}\right| - \left|\frac{\partial f(x^{k_r})}{\partial x_j}\right|,$$

also

$$(3.5') \qquad \left|\frac{\partial f(x^{k_r})}{\partial x_j}\right| \geqq \left|\frac{\partial f(x')}{\partial x_j}\right| - \left|\frac{\partial f(x')}{\partial x_j} - \frac{\partial f(x^{k_r})}{\partial x_j}\right| \forall r \geqq r_0, \quad \forall j \in \{1, \ldots, n\}.$$

Ohne Einschränkung der Allgemeinheit läßt sich die konvergente Teilfolge (x^{k_r}) so wählen, daß für alle $k_r \geqq 1$ gilt: $x^{k_r} \neq x^{k_r+1}$. Dann hat man für $j \in \{1, \ldots, n\}$ mit $j \equiv k_r \bmod n$:

$$\left|\frac{\partial f(x^{k_r})}{\partial x_j}\right| \neq 0 \quad \forall k_r \geqq 1.$$

Es können nämlich im Verfahren (KS) höchstens $n-1$ Punkte $x^l, x^{l+1}, \ldots, x^{l+n-2}$ in der Folge (x^k) aufeinanderfolgen, für die gilt:

$$\frac{\partial f(x^l)}{\partial x_j} = 0, \quad \frac{\partial f(x^{l+1})}{\partial x_{j+1}} = 0, \ldots, \frac{\partial f(x^{l+p})}{x_n} = 0,$$

$$\frac{\partial f(x^{l+p+1})}{\partial x_1} = 0, \ldots, \frac{\partial f(x^{l+n-2})}{\partial x_{j-2}} = 0$$

mit $j \in \{1, \ldots, n\}$ und $j \equiv l \bmod n$. Wegen (1.2), (1.3) ist dann

$$x^l = x^{l+1} = \ldots = x^{l+n-2} = x^{l+n-1}.$$

Wäre nun auch

$$\frac{\partial f(x^{l+n-1})}{\partial x_{j-1}} = 0,$$

so gilt

$$\nabla f(x^l) = 0.$$

Wegen der vorausgesetzten Pseudokonvexität von f folgt daraus

$$x^l \in G^*$$

im Widerspruch zu (3.1).

Daher bedeutet die beschriebene Wahl der Teilfolge (x^{k_r}) ein Weglassen von höchstens endlich vielen gleichen Gliedern, was ohne Einfluß auf das Konvergenzverhalten ist. Es folgt mit (3.4), (3.5) und $\varepsilon = \sigma/2$ nach (3.5'): Es gibt ein $r_0 \geqq 1$, so daß für alle $j \in \{1, \ldots, n\}$ mit

$$(3.6) \qquad j \equiv k_r \bmod n \quad \text{und} \quad \partial f(x')/\partial x_j \neq 0$$

gilt:

$$(3.7) \qquad \left|\frac{\partial f(x^{k_r})}{\partial x_j}\right| \geqq \sigma - \frac{\sigma}{2} = \frac{\sigma}{2} \quad \forall r \geqq r_0.$$

Wegen der stetigen Differenzierbarkeit von f gilt nach dem Satz von Taylor:

$$(3.8) \qquad f(x^{k_r} + \alpha e^j) = f(x^{k_r}) + \alpha \frac{\partial f(x^{k_r})}{\partial x_j} + 0(\alpha) \quad \forall \alpha \neq 0\,, \quad \text{reell.}$$

Sei $\operatorname{sgn} \alpha = -\operatorname{sgn} \partial f(x^{k_r})/\partial x_j$. Dann lautet (3.8)

$$f(x^{k_r} + \alpha e^j) = f(x^{k_r}) - \left| \alpha \frac{\partial f(x^{k_r})}{\partial x_j} \right| + 0(\alpha)$$

und es gilt mit (3.7)

$$(3.9) \qquad f(x^{k_r} + \alpha e^j) = f(x^{k_r}) - \tfrac{1}{2}|\alpha|\, \sigma + 0(\alpha) \quad \forall r \geqq r_0\,, \quad \forall f \quad \text{gemäß (3.6)}\,.$$

Nach (3.9) gibt es ein α_0 mit $\operatorname{sgn} \alpha_0 = -\operatorname{sgn} \partial f(x^{k_r})/\partial x_j$, so daß gilt:

$$(3.10) \qquad f(x^{k_r} + \alpha e^j) \leqq f(x^{k_r}) - \tfrac{1}{4}|\alpha|\, \sigma \quad \forall \alpha \quad \text{mit} \quad \operatorname{sgn} \alpha = \operatorname{sgn} \alpha_0$$

und $|\alpha| \in (0, |\alpha_0|]$, $\forall r \geqq r_0$, $\forall j$ gemäß (3.6).
Daraus ergibt sich mit (3.2) und (3.3)

$$(3.11) \qquad f(x^k) > f(x') \quad \forall k \geqq 1\,.$$

Insbesonders gibt es für alle $|\alpha| > 0$ ein $r_1 \geqq 1$, so daß gilt:

$$(3.12) \qquad f(x^{k_r}) \leqq f(x') + \tfrac{1}{8}|\alpha|\, \sigma \quad \forall r \geqq r_1\,.$$

Andererseits ist für $r \geqq \max\{r_0, r_1\}$ wegen (3.10)

$$f(x^{k_r+1}) \leqq f(x^{k_r} + \alpha e^j) \leqq f(x^{k_r}) - \tfrac{1}{4}|\alpha|\, \sigma$$

und mit (3.12)

$$f(x^{k_r+1}) \leqq f(x') - \tfrac{1}{8}|\alpha|\, \sigma\,,$$

was (3.11) widerspricht. Mithin gilt $x' \in G^*$.

4. ANWENDBARKEIT DES VERFAHRENS FÜR PROBLEME MIT RESTRIKTIONEN

Das Verfahren (KS) kann auch zur Lösung von Problemen der Form

$$(4.1) \qquad \min\{f(x) \mid x \in G\}\,,$$

$$G =_{\mathrm{Df}} \{x \in R^n \mid a_i \leqq x_i,\ i \in I;\ x_j \leqq b_j,\ j \in J;\ I, J \subseteq \{1, \ldots, n\};\ a_i, b_j \in R\} \neq \emptyset\,,$$

d. h. zur Bestimmung eines Punktes $x^* \in G$ mit

$$f(x^*) \leqq f(x) \quad \forall x \in G$$

verwendet werden. O. B. d. A. sei dabei $a_k < b_k\ \forall k \in I \cap J$. Dazu muß in (1.2) die $\in$-Relation überall auf G anstelle R^n bezogen werden. Selbstverständlich gilt für die erzeugte Punktfolge $(x^k) \subset G$.

Definition 2 wird auf abgeschlossene (evtl. unbeschränkte) zusammenhängende Punktmengen $G \subseteq R$ anstelle von R erweitert; die Existenz eines nicht leeren Intervalls $I^* \subseteq G$ mit den angegebenen Eigenschaften definiert die Quasi-Unimodalität auf G. Definition 3 läßt sich auf konvexe, abgeschlossene Bereiche $G \subseteq R^n$ anstelle R^n erweitern. Dazu fordert man, daß f quasi-unimodal auf jeder Strecke $[x^1, x^2] =_{\mathrm{Df}} \operatorname{conv}\{x^1, x^2\} = \{x \in G \mid x = \lambda x^1 + (1-\lambda)x^2;\ x^1, x^2 \in G,\ \lambda \in [0, 1]\}$ ist. Die Sätze 1 und 2 gelten auch für die entsprechenden Eigenschaften von f auf abgeschlossenen, konvexen Mengen G.

Für pseudokonvexe Funktionen f ist aber $\nabla f(x^*) = 0$ nicht mehr notwendig dafür, daß $f(x^*) \leqq f(x^*)$ $\forall x \in G$ ist. Vielmehr gilt für Probleme der Form (4.1) unter Benutzung der Kuhn-Tucker-Bedingungen:

Satz 5. *Seien $X \subseteq R^n$ eine offene Menge und $f: X \to R$ pseudokonvex. Dann ist $x^* \in G \subset X$ Lösung des Problems* (4.1), *falls gilt:*

$$\min\left\{\frac{\partial f(x^*)}{\partial x_j}(b_j - x_j^*), \frac{\partial f(x^*)}{\partial x_j}(a_j - x_j^*)\right\} \geqq 0 \quad \forall j \in \{1, \ldots, n\}\,. \tag{4.2}$$

Zum Beweis vgl. Mangasarian [10].

Satz 3 läßt sich dagegen nicht aufrechterhalten: Es wird sich als entscheidend erweisen, daß G koordinatenachsenparallele Kanten besitzt, und somit entfällt die Unabhängigkeit von der Wahl der Koordinatenachsenrichtungen. Anstelle Satz 4 gilt nachstehende Aussage:

Satz 4'. *Wenn $f: X \to R$ stetig differenzierbar und pseudokonvex auf der offenen Menge $X \subseteq R^n$ ist, $G \subset X$ gemäß* (4.1) *definiert ist und es eine kompakte Menge $K \subseteq G$ mit $(x^k) \subset K$ gibt, dann konvergiert das Verfahren* (*KS*).

Beweis. 1. Für $x^k \in G^* =_{\mathrm{Df}} \{x^* \in G \mid f(x^*) \leqq f(x)\ \forall x \in G\}$ gilt dieselbe Überlegung wie unter 1. im Beweis von Satz 7. 2. Im entgegengesetzten Fall ist wegen der Kompaktheit von K wieder die Existenz einer konvergenten Teilfolge $(x^{k_r}) \subseteq (x^k)$ mit

$$\lim_{r \to \infty} x^{k_r} = x' \in G$$

gesichert. Die Beziehungen (3.2) und (3.3) gelten weiter. Wir nehmen an, es sei

$$x' \notin G^*\,. \tag{4.3}$$

Aus der Bemerkung über pseudokonvexe Funktionen am Ende des Abschnitts 2 folgt, daß dann ein Index $j \in \{1, \ldots, n\}$ existiert mit

$$\frac{\partial f(x')}{\partial x_j} \neq 0\,. \tag{4.4}$$

Wir nehmen nun eine Fallunterscheidung vor: Fall 1 : $x' \in \operatorname{int} G$. Dann gelten dieselben Aussagen wie unter (3.4) ... (3.12) mit der Maßgabe, daß in (3.10) α_0 so zu wählen ist, daß gilt

$$x^{k_r} + \alpha_0 e^j \in G \quad \forall r \geqq r_0\,, \quad \forall j \text{ gemäß (3.6).}$$

Fall 2 : $x' \in \operatorname{bd} G$. Die Annahme (4.4) verlangt dann, daß gemäß Satz 5 (4.2) nicht gilt, d. h. es existiert ein Index $j \in \{1, \ldots, n\}$ mit

$$\frac{\partial f(x')}{\partial x_j}(b_j - x'_j) < 0 \quad \text{oder} \quad \frac{\partial f(x')}{\partial x_j}(a_j - x'_j) < 0\,.$$

Hieraus folgt

$$\frac{\partial f(x')}{\partial x_j} < 0 \quad \text{und} \quad x'_j < b_j$$

oder

$$\frac{\partial f(x')}{\partial x_j} > 0 \quad \text{und} \quad x'_j > a_j\,.$$

Dann gibt es sicher eine reele Zahl α_0 mit $\operatorname{sgn} \alpha_0 = -\operatorname{sgn} \partial f(x')/\partial x_j$, so daß

$$f(x' + \alpha e^j) < f(x') \quad \text{und} \quad x' + \alpha e^j \in G \quad \forall \alpha \quad \text{mit} \quad \operatorname{sgn} \alpha = \operatorname{sgn} \alpha_0$$
$$\text{und} \quad |\alpha| \in (0, |\alpha_0|]\,.$$

Nach der Konstruktionsvorschrift für die Folge (x^k) kann dann jedoch x' nicht Limeselement einer konvergenten Teilfolge (x^{k_r}) sein. Folglich gilt (4.3) nicht und Satz 4' ist vollständig bewiesen.

5. NUMERISCHE ERGEBNISSE

Das Verfahren (KS) wurde für die Anlage Odra 1204 in ALGOL 1204 programmiert. Als eindimensionales direktes Verfahren zur Bestimmung von x^{k+1} gemäß (1.2) wurde das Verfahren des Goldenen Schnitts (vgl. [11]) verwendet. Der Einsatz dieses Verfahrens innerhalb einer Strecke L_k wird jeweils dann abgebrochen, wenn der Abstand von zwei durch dieses Verfahren erzeugten Hilfspunkten kleiner als eine vorgegebene positive Konstante δ wird. Der Hilfspunkt mit dem kleineren Wert der Zielfunktion f wird dann x^{k+1}. Die Implementation von (KS) verlangt stets die Angabe von a_i, b_i, $i \in \{1, \ldots, n\}$, derart, daß sich $G \subset R^n$ gemäß (4.1) als n-dimensionaler Quader ergibt. Sind im gegebenen Problem gewisse Variable x_i nach oben bzw. nach unten unbeschränkt, so ist als b_i die größte bzw. als a_i die kleinste Maschinenzahl verwendbar. (KS) bricht ab, wenn für einen gewissen Index $k > n$ gilt:

$$\|x^k - x^{k-n}\|_T < \delta\,,$$

wobei $\|.\|_T$ die Tschebyschew-Norm bedeutet. Der Startpunkt $x^1 \in G$, der gleichzeitig der erste Hilfspunkt innerhalb des Verfahrens des Goldenen Schnitts ist, wird wie folgt vom Programm gewählt:

$$x_1^1 =_{\mathrm{Df}} 0{\cdot}381966 a_1 + 0{\cdot}618034 b_1 ,$$
$$x_i^1 =_{\mathrm{Df}} 0{\cdot}5 a_i + 0{\cdot}5 b_i \quad \text{für} \quad i \in \{2, \ldots, n\} .$$

Als Testbeispiele wurden u. a. die folgenden Probleme verwendet.

Beispiel A. $n = 4$, $\delta = 10^{-4}$; $f(x) = x_1^2 + 0{\cdot}5x_2^2 + x_3^2 + 0{\cdot}5x_4^2 - x_1x_3 + x_3x_4 - x_1 - 3x_2 + x_3 - x_4$.

Diese Funktion ist im R^4 differenzierbar und auf jeder konvexen Teilmenge dieses Raumes streng konvex, also auch pseudokonvex.

Zum Vergleich des Einflusses der Anzahl der aktiven Restriktionen wurde das Problem bei verschiedenen Lagen des zulässigen Bereichs G gelöst.

1. $a_i = 0{\cdot}5$, $b_i = 1{\cdot}5$ für $i \in \{1, 2, 3, 4\}$;

$x^* = (0{\cdot}75, 1{\cdot}5, 0{\cdot}5, 0{\cdot}5)^{\mathrm{T}}$, $f(x^*) = -3{\cdot}3125$.

Ermittelte Werte:

$$\tilde{x}^* = \begin{pmatrix} 0{\cdot}750064 \\ 1{\cdot}49989 \\ 0{\cdot}500107 \\ 0{\cdot}500107 \end{pmatrix}, \quad f(\tilde{x}^*) = -3{\cdot}31215 ;$$

Rechenzeit: 8·6 sec.

2. $a_1 = -2{\cdot}5$, $b_1 = 0$, $a_2 = 2{\cdot}5$, $b_2 = 4{\cdot}5$,

$a_3 = -3{\cdot}5$, $b_3 = -1{\cdot}2$, $a_4 = 3{\cdot}0$, $b_4 = 6{\cdot}7$;

$x^* = (-1, 3, -3, 4)^{\mathrm{T}}$, $f(x^*) = -7{\cdot}5$.

Ermittelte Werte:

$$\tilde{x}^* = \begin{pmatrix} -1{\cdot}00026 \\ 3{\cdot}00013 \\ -3{\cdot}00039 \\ 4{\cdot}00040 \end{pmatrix}, \quad f(x^*) = -7{\cdot}50000 ;$$

Rechenzeit: 22·2 sec.

Beispiel B: $n = 2$, $\delta = 10^{-4}$; $f(x) = -e^{-(x_1{}^2 + x_2{}^2)}$. Diese Funktion ist im R^2 differenzierbar und auf jeder konvexen Teilmenge dieses Raumes streng pseudokonvex (aber nicht konvex). Mit $a_i = -1{\cdot}0$, $b_i = 1{\cdot}0$ für $i \in \{1, 2\}$ sind die Voraussetzungen des Satzes 4′ erfüllt. Es ist dann $x^* = (0, 0)^{\mathrm{T}}$, $f(x^*) = -1{\cdot}0$.

Ermittelte Werte:

$$\tilde{x}^* = \begin{pmatrix} 0{\cdot}000011 \\ -0{\cdot}000009 \end{pmatrix}, \quad f(\tilde{x}^*) = -1{\cdot}00000\,;$$

Rechenzeit: 3·8 sec.

Beispiel C. $n = 2$, $\delta = 10^{-4}$; $f(x) = \max\{|x_1|, |x_2|\}$. Diese Funktion ist zwar auf jeder konvexen Teilmenge des R^2 konvex und damit linear quasi-unimodal, aber sie ist nicht überall differenzierbar. Die Voraussetzungen des Satzes 4′ sind mit $a_i = -2{\cdot}0$, $b_i = 1{\cdot}0$ für $i \in \{1, 2\}$ nicht erfüllt. In der Tat ist $x^* = (0, 0)^{\mathsf{T}}$, $f(x^*) = 0$, aber es wurde ermittelt:

$$\tilde{x} = \begin{pmatrix} -0{\cdot}500010 \\ -0{\cdot}500000 \end{pmatrix}, \quad f(\tilde{x}) = 0{\cdot}500010\,.$$

Sieht man von dem unterhalb der δ-Schranke liegenden Abbruchfehler bei $\tilde{x}_1$ ab, so ist

$$\tilde{x} \approx \begin{pmatrix} -0{\cdot}5 \\ -0{\cdot}5 \end{pmatrix} = x'\,.$$

In x' ist aber die Niveaulinie $A(x') = \{x \in R^2 \mid f(x) = 0{\cdot}5\}$ nicht glatt, so daß das Verfahren (KS) von dort aus keinen Punkte mit kleinerem Zielfunktionswert findet.

Literatur

[1] *Donath, G./K.-H. Elster:* Über Eigenschaften unimodaler Funktionen. Wiss. Z. TH Ilmenau 18 (1972), Heft 3, 103—120.

[2] *Donath, G./K.-H. Elster:* Über eine Verallgemeinerung unimodaler Funktionen.Godišnik na visšite techničeski učebni zavedenija. Matematika IX (1973), Heft 3, 7—21.

[3] *Donath, G./K.-H. Elster:* Zur Konvergenz eines Verfahrens der nichtlinearen Optimierung Теоретична и приложна механика V (1974), No. 4, 23—28.

[4] *Elster, K.-H./G. Folgmann:* Über Verallgemeinerungen konvexer Funktionen und deren Anwendung in der Theorie der nichtlinearen Optimierung. Wiss. Z. TH Ilmenau 16 (1970), Heft 4, 23—34.

[5] *Friedman, M./L. S. Savage:* Selected Techniques of Statistical Analysis. McGraw-Hill Book Co. New York 1947.

[6] *Greenberg, H. J./W. P. Pierskalla:* A review of quasiconvex functions. Op. Res. 19 (1971), 1553—1570.

[7] *Hooke, R./T. A. Jeeves:* "Direct Search" Solution of Numerical and Statistical Problems. Journ. ACM 8 (1962), 212—229.

[8] *Karamardian, S.:* Strictly quasiconvex (concave) functions and duality in mathematical programming. J. Math. Anal. Appl. 20 (1967), 344—358.

[9] *Kiefer, J.:* Sequential minimax search for a maximum. Proc. Am. Math. Soc. 4 (1953), 502—506.

[10] *Mangasarian, O. L.:* Pseudo-convex functions. J. SIAM, Ser. A, 3 (1965), 281—290.

[11] *Wilde, D. J./C. S. Beightler:* Foundations of Optimization. Prentice-Hall Inc. Englewood Cliffs 1967.

[12] *Zangwill, W. I.:* Nonlinear Programming: A Unified Approach. Prentice-Hall Inc., Englewood Cliffs 1969.

Souhrn

KE KONVERGENCI METODY HLEDÁNÍ PO SOUŘADNICÍCH

GERHARD DONATH, KARL-HEINZ ELSTER

V předložené práci jsou uvedeny předpoklady pro konvergenci jisté metody pro řešení nelineárních optimalizačních úloh bez omezení. Vyšetřovaná metoda patří do třídy přímých metod nebo metod neobsahujících derivace, pro které zpravidla nejsou uvedeny podmínky pro konvergenci. Při těchto podmínkách hrají roli vlastnosti cílové funkce, které representují zobecnění unimodality, ale podmínky souvisí také se zobecněnými pojmy konvexnosti. Autoři rozšiřují výroky o konvergenci s použitím Kuhnových-Tuckerových podmínek pro úlohy s proměnnými, které jsou zdola nebo shora omezené.

Anschriften der Verfasser: Prof. Dr. rer. nat. habil. *Karl-Heinz Elster*, Technische Hochschule Ilmenau, Sektion Mathematik, Rechentechnik und ökonomische Kybernetik, DDR-63 Ilmenau, und Dr. rer. nat. *Gerhard Donath*, Pädagogische Hochschule ,,N. K. Krupskaja" Halle, Sektion Mathematik/Physik, DDR-402 Halle.