Computing 19, 59—72 (1977)

Computing

Effiziente Schrittweitenfunktionen bei unrestringierten Optimierungsaufgaben

W. Warth und **J. Werner**, Göttingen

Eingegangen am 17. September 1976

Zusammenfassung — Abstract

Effiziente Schrittweitenfunktionen bei unrestringierten Optimierungsaufgaben. Von einem allgemeinen Standpunkt aus werden verschiedene Schrittweitenfunktionen diskutiert und ihr Einfluß auf die Konvergenz von Verfahren der unrestringierten Optimierung betrachtet. Es werden effiziente Schrittweitenfunktionen definiert und gezeigt, daß die bekannten Schrittweitenalgorithmen effizient sind. Schließlich werden notwendige und hinreichende Konvergenzkriterien für Abstiegsverfahren angegeben und auf Verfahren der konjugierten Gradienten angewendet.

Efficient Step-Size Functions for Unconstrained Optimization Problems. In the present paper we discuss several steplength procedures from a general point of view. We consider their influence on the convergence of algorithms for the numerical treatment of optimization problems without constraints. We define efficient step-size functions and show that well known steplength procedures are efficient. Necessary and sufficient conditions for convergence of descent methods with efficient step-size functions and applications to conjugate gradient methods are given.

0. Einleitung

Abstiegsverfahren zur näherungsweisen Lösung der unrestringierten Optimierungsaufgabe, eine stetig differenzierbare Funktion $f: \mathbb{R}^n \to \mathbb{R}$ zu minimieren, haben bekanntlich die folgende Form:

Ist x ein Näherungswert für die Lösung und kein kritischer Punkt, so bestimme man zunächst eine Abstiegsrichtung p, d. h. ein $p \in \mathbb{R}^n$ mit $f(x-t\,p) < f(x)$ für alle hinreichend kleinen Schrittweiten $t>0$ und anschließend nach einer geeigneten Vorschrift eine Schrittweite $t(x,p)>0$ mit $f\big(x-t(x,p)\,p\big)<f(x)$.

Im allgemeinen (d. h. wenn kein Abbruch nach endlich vielen Schritten durch Erreichen eines kritischen Punktes erfolgt) werden also durch ein Abstiegsverfahren eine Näherungsfolge $\{x^k\}$, eine Richtungsfolge $\{p^k\}$ und eine Schrittweitenfolge $\{t_k\}$ mit $t_k = t(x^k, p^k)$ definiert. Wichtig ist nun die Bemerkung (siehe unter anderem Daniel [1], Ortega-Rheinboldt [5]), daß für die gebräuchlichen Schrittweitenvorschriften die Bedingung

$$\lim_{k\to\infty} \frac{(\nabla f(x^k), p^k)}{\| p^k \|} = 0 \qquad (0.1)$$

nachgewiesen wird, falls nur $\{p^k\}$ eine Folge von Abstiegsrichtungen ist. Häufig kann etwas mehr als (0.1) gezeigt werden, nämlich daß eine Konstante $\theta>0$ existiert mit

$$f(x^k)-f(x^k-t_k p^k)\geq\theta\,\frac{(\nabla f(x^k),p^k)^2}{\|p^k\|^2}\quad\text{für alle } k. \tag{0.2}$$

Dies führt auf den Begriff der effizienten Schrittweite bzw. der effizienten Schrittweitenfunktion.

Über f werden wir generell in diesem Aufsatz voraussetzen

V) i) $f:\mathbb{R}^n\to\mathbb{R}$ ist stetig differenzierbar.

ii) Es existiert ein $\delta\in\mathbb{R}$ mit:

$$M_\delta:=\{x\in\mathbb{R}^n: f(x)\leq\delta\}\text{ ist kompakt.}$$

iii) Es existiert ein $\gamma\in\mathbb{R}^+$ mit

$$\|\nabla f(x)-\nabla f(y)\|\leq\gamma\,\|x-y\|\quad\text{für alle } x,y\in M_\delta.$$

(Hier und im folgenden bedeutet $\|\ \|$ stets die euklidische Vektornorm im $\mathbb{R}^n$.)

Nun präzisieren wir durch die folgende Definition, was wir unter einer effizienten Schrittweitenfunktion verstehen.

Definition 0.1: Sei $M_\delta^*:=\{x\in M_\delta:\nabla f(x)\neq\theta\}$ die Menge der nichtkritischen Punkte in M_δ, $F(x):=\{p\in\mathbb{R}^n:(\nabla f(x),p)>0\}$ die Menge der Abstiegsrichtungen für f in x und $A:=\{(x,p)\in M_\delta^*\times\mathbb{R}^n: p\in F(x)\}$. Dann heißt eine Abbildung $T:A\to p(\mathbb{R}^+)$ (= Menge aller Teilmengen von $\mathbb{R}^+$) eine *Schrittweitenfunktion*. Eine Schrittweitenfunktion T heißt *effizient*, falls eine Konstante $\theta>0$, der sogenannte *Effizienzfaktor*, existiert mit

$$f(x)-f(x-t(x,p)\,p)\geq\theta\,\frac{(\nabla f(x),p)^2}{\|p\|^2}\quad\text{für alle } t(x,p)\in T(x,p)\text{ und alle }(x,p)\in A.$$

Wir werden im folgenden von einigen bekannten Schrittweitenstrategien nachweisen, daß sie auf eine effiziente Schrittweitenfunktion zurückgeführt werden können. Wir werden uns dabei bemühen, den optimalen Effizienzfaktor zu erhalten, da man dessen Größe als ein Maß für den Abstieg der Zielfunktion und damit für die Konvergenzgüte ansehen kann.

Im darauffolgenden Abschnitt werden wir notwendige und hinreichende Konvergenzbedingungen für Abstiegsverfahren mit einer effizienten Schrittweitenfunktion angeben. Genauer wird gezeigt, daß die Zoutendijk-Bedingung

$$\sum_{k=1}^{\infty}\frac{(\nabla f(x^k),p^k)^2}{\|\nabla f(x^k)\|^2\,\|p^k\|^2}=\infty,$$

bei gleichmäßig konvexer Zielfunktion eine hinreichende Konvergenzbedingung ist und im allgemeinen Fall notwendig für Konvergenz ist, wodurch Ergebnisse von Lenard [4] verallgemeinert werden. Im letzten Abschnitt werden wir unsere Ergebnisse auf Verfahren mit konjugierten Gradienten anwenden.

1. Beispiele effizienter Schrittweitenfunktionen

Im folgenden sei $\psi(t) := f(x - t\,p)$ bei festem $(x, p) \in A$.

a) *Curry-Altmann*: Sei $\alpha \in [0, 1)$ vorgegeben. $\hat{t}(x, p)$ sei die erste positive Nullstelle von ψ' (Curry-Schrittweite). Dann heißt

$$T_{CA}(x, p) := \{t_{CA} \in (0, \hat{t}(x, p)] : \alpha\, \psi'(0) \leq \psi'(t_{CA})\}$$

die Curry-Altmann-Schrittweitenfunktion.

Bemerkung: Für $\alpha = 0$ erhält man die Curry-Schrittweite. Unter der Curry-Altmann-Schrittweite versteht man im allgemeinen (siehe Ortega-Rheinboldt [5]) die kleinste positive Nullstelle von $\phi(t) = \psi'(t) - \alpha\,\psi'(0)$. Bei uns (siehe auch Lenard [4], Wolfe [9]) wird die Schrittweite nicht durch eine Gleichung, sondern durch eine Ungleichung bestimmt, was vom praktischen Standpunkt aus wichtig ist.

Satz 1.1: *Die Curry-Altmann-Schrittweitenfunktion existiert* $\big($*d. h.* $T_{CA}(x, p) \neq \emptyset$ *für alle* $(x, p) \in A\big)$ *und ist effizient.*

Beweis: α) Existenz: Betrachte $\phi(t) := \psi'(t) - \alpha\,\psi'(0)$ auf $[0, \hat{t}(x, p)]$. Dann ist $\phi(0) = (1 - \alpha)\,\psi'(0) < 0$ und $\phi\big(\hat{t}(x, p)\big) = -\alpha\,\psi'(0) \geq 0$ $\big(>0$ falls $\alpha \in (0, 1)\big)$, womit z. B. $\hat{t}(x, p) \in T_{CA}(x, p)$ gezeigt ist und im Falle $\alpha > 0$ die Existenz eines $\varepsilon > 0$ mit $[\hat{t}(x, p) - \varepsilon, \hat{t}(x, p)] \subset T_{CA}(x, p)$ folgt.

β) Effizienz: Im folgenden sei $(x, p) \in A$ fest, $\hat{t} = \hat{t}(x, p)$ und $t_{CA} \in T_{CA}(x, p)$ beliebig. Dann gilt

$$\alpha\,\psi'(0) \leq \psi'(t_{CA}) = \psi'(0) + \psi'(t_{CA}) - \psi'(0) \text{ bzw.}$$

$$-\alpha\,(\nabla f(x), p) \leq -(\nabla f(x), p) + (\nabla f(x) - \nabla f(x - t_{CA}\,p), p)$$

$$\leq -(\nabla f(x), p) + t_{CA}\,\gamma\,\|p\|^2 \quad \text{wegen} \quad x - t_{CA}\,p \in M_\delta$$

und daher

$$\tilde{t}_{CA} := (1 - \alpha)\,\frac{(\nabla f(x), p)}{\gamma\,\|p\|^2} \leq t_{CA}.$$

Auf $[0, \hat{t}]$, speziell auf $[0, t_{CA}]$, ist ψ monoton nicht wachsend. Damit wird

$$f(x) - f(x - t_{CA}\,p) = \psi(0) - \psi(t_{CA}) \geq \psi(0) - \psi(\tilde{t}_{CA}) =$$

$$= -\psi'(0)\,\tilde{t}_{CA} - \int_0^{\tilde{t}_{CA}} (\psi'(\tau) - \psi'(0))\,d\tau$$

$$\geq (\nabla f(x), p)\,\tilde{t}_{CA} - \frac{\tilde{t}_{CA}^2\,\gamma\,\|p\|^2}{2} = \frac{1 - \alpha^2}{2\gamma}\,\frac{(\nabla f(x), p)^2}{\|p\|^2},$$

womit die Effizienz der Curry-Altmann-Schrittweitenfunktion mit Effizienzfaktor $\theta_{CA} = \dfrac{1 - \alpha^2}{2\gamma}$ gezeigt ist.

b) *Goldstein*: Sei $\beta \in (0, \frac{1}{2})$ vorgegeben. Dann heißt

$$T_G(x, p) := \{t_G \in \mathbb{R}^+ : t_G \beta (\nabla f(x), p) \leq \psi(0) - \psi(t_G) \leq t_G (1 - \beta)(\nabla f(x), p)\}$$

die Goldstein-Schrittweitenfunktion.

Der Vorteil der Goldstein-Schrittweitenfunktion besteht, genau wie bei der oben angegebenen Version der Curry-Altmann-Schrittweitenfunktion, darin, daß man zur Berechnung der Schrittweite keine Gleichung lösen muß.

Satz 1.2: *Die Goldstein-Schrittweitenfunktion T_G existiert und ist effizient.*

Beweis: α) Existenz: Siehe Ortega-Rheinboldt [5].

β) Effizienz: Sei $(x, p) \in A$ und $t_G \in T_G(x, p)$ beliebig.

i) $t_G \leq t_C :=$ Curry-Schrittweite.
Dann ist

$$t_G (1 - \beta)(\nabla f(x), p) \geq \psi(0) - \psi(t_G) \geq (\nabla f(x), p)\, t_G - \frac{t_G^2 \gamma \| p \|^2}{2},$$

so daß $\tilde{t}_G := \dfrac{2\beta}{\gamma} \dfrac{(\nabla f(x), p)}{\| p \|^2} \leq t_G$. Wegen $\tilde{t}_G \leq t_G \leq t_C$ wird damit

$$\begin{aligned} f(x) - f(x - t_G p) &\geq \psi(0) - \psi(\tilde{t}_G) \geq (\nabla f(x), p)\, \tilde{t}_G - \frac{\tilde{t}_G^2 \gamma \| p \|^2}{2} \\ &= \frac{2\beta(1-\beta)}{\gamma} \frac{(\nabla f(x), p)^2}{\| p \|^2}. \end{aligned}$$

ii) $t_C < t_G$.
Dann ist $\dfrac{(\nabla f(x), p)}{\gamma \| p \|^2} \leq t_C < t_G$, so daß

$$f(x) - f(x - t_G p) \geq t_G \beta (\nabla f(x), p) \geq \frac{\beta}{\gamma} \frac{(\nabla f(x), p)^2}{\| p \|^2},$$

womit die Effizienz der Goldstein-Schrittweitenfunktion mit $\theta_G := \dfrac{\beta}{\gamma}$ bewiesen ist.

c) *Danilin-Psenichnyi*: Seien $\varepsilon \in (0, 1)$, $\beta \in (0, 1)$ und $\sigma \in \mathbb{R}^+$ vorgegeben. Sei $t^*(x, p) := \sigma \cdot \dfrac{(\nabla f(x), p)}{\| p \|^2}$ und $q(x, p)$ die kleinste nichtnegative ganze Zahl mit

$$f(x) - f\left(x - \beta^{q(x,p)} t^*(x, p)\, p\right) \geq \varepsilon\, \beta^{q(x,p)} t^*(x, p) (\nabla f(x), p).$$

Dann heißt $T_{DP}(x, p) = \{\beta^{q(x,p)} t^*(x, p)\}$ die Danilin-Psenichnyi-Schrittweitenfunktion.

Die Danilin-Psenichnyi-Schrittweitenfunktion (siehe Danilin-Psenichnyi [2]) ist eine leichte Spezialisierung der Goldstein-Armijo-Schrittweitenfunktion (siehe z. B. Ortega-Rheinboldt [5], Polak [7]).

Satz 1.3: *Die Danilin-Psenichnyi-Schrittweitenfunktion existiert und ist effizient.*

Beweis: α) Existenz: Betrachte $\phi(t):=\psi(0)-\psi(t)+\varepsilon\, t\, \psi'(0)$. Dann ist $\phi(0)=0$, $\phi'(0)=-(1-\varepsilon)\psi'(0)>0$, so daß $\phi(t)>0$ für alle hinreichend kleinen positiven t.

β) Effizienz: Sei $t_{DP}(x,p)=\beta^{q(x,p)}\, t^*(x,p)$.

i) $q(x,p)=0$. Dann ist $t_{DP}(x,p)=t^*(x,p)$ und

$$f(x)-f(x-t_{DP}\,p)\geq \varepsilon\,\sigma\,\frac{(\nabla f(x),p)^2}{\|p\|^2}.$$

ii) $q(x,p)\geq 1$. Dann sind die folgenden beiden Ungleichungen erfüllt:

$$\psi(0)-\psi(t_{DP})\geq \varepsilon\, t_{DP}\left(\nabla f(x),p\right)$$
$$\psi(0)-\psi(\beta^{-1}\,t_{DP})<\varepsilon\,\beta^{-1}\,t_{DP}\left(\nabla f(x),p\right).$$

1. Fall: $\beta^{-1}\,t_{DP}\leq t_C=$Curry-Schrittweite.

Dann ist

$$\varepsilon\,\beta^{-1}\,t_{DP}\left(\nabla f(x),p\right)>\psi(0)-\psi(\beta^{-1}\,t_{DP})\geq\left(\nabla f(x),p\right)\beta^{-1}\,t_{DP}-\frac{(\beta^{-1}\,t_{DP})^2\,\gamma\,\|p\|^2}{2},$$

so daß

$$\tilde{t}_{DP}:=\frac{2(1-\varepsilon)\beta}{\gamma}\,\frac{(\nabla f(x),p)}{\|p\|^2}\leq t_{DP}\leq\beta^{-1}\,t_{DP}\leq t_C$$

und daher

$$f(x)-f(x-t_{DP}\,p)\geq\psi(0)-\psi(\tilde{t}_{DP})\geq\frac{2(1-\varepsilon)\beta}{\gamma}\,(1-\beta+\varepsilon\,\beta)\,\frac{(\nabla f(x),p)^2}{\|p\|^2}.$$

2. Fall: $t_C<\beta^{-1}\,t_{DP}$.

Dann ist

$$\frac{\beta}{\gamma}\,\frac{(\nabla f(x),p)}{\|p\|^2}\leq t_{DP}$$

und daher

$$f(x)-f(x-t_{DP}\,p)\geq\frac{\varepsilon\,\beta}{\gamma}\,\frac{(\nabla f(x),p)^2}{\|p\|^2}.$$

Insgesamt ist damit die Effizienz der Danilin-Psenichnyi-Schrittweitenfunktion mit Effizienzfaktor

$$\theta_{DP}=\min\left(\varepsilon\,\sigma,\frac{2(1-\varepsilon)\beta}{\gamma}\,(1-\beta+\varepsilon\,\beta),\frac{\varepsilon\,\beta}{\gamma}\right)$$

bewiesen.

d) *Polak*: Sei $t_M=t_M(x,p)>0$ das erste globale Minimum von ψ auf $[0,\infty)$. Dann heißt

$$T_P(x,p):=\{t_P\in(0,\infty):f(x)-f(x-t_P\,p)\geq|t_P-t_M|\;\|p\|\}$$

die Polak-Schrittweitenfunktion (siehe Polak [7]).

Satz 1.4: *Die Polak-Schrittweitenfunktion existiert und ist effizient.*

Beweis: α) Existenz: Sei $\phi(t)=f(x)-f(x-t\,p)-|t-t_M|\,\|p\|$. Dann ist $\phi(t_M)>0$, woraus die Existenz folgt.

β) Effizienz: Sei $t_P\in T_P(x,p)$ beliebig. Dann ist

$$\begin{aligned} f(x)-f(x-t_P\,p)&=\psi(0)-\psi(t_P)\\ &=\psi(0)-\psi(t_M)+\psi(t_M)-\psi(t_P)\\ &\geq\frac{1}{2\gamma}\frac{(\nabla f(x),p)^2}{\|p\|^2}-|\psi'(\tilde t_M)|\,|t_M-t_P| \end{aligned}$$

mit $\tilde t_M=\lambda\,t_M+(1-\lambda)\,t_P$, $\lambda\in(0,1)$. Mit M_δ ist auch die konvexe Hülle $\mathrm{co}\,M_\delta$ kompakt, daher existiert eine Konstante $L>0$ mit $\|\nabla f(y)\|\leq L$ für alle $y\in\mathrm{co}\,M_\delta$. Speziell ist $x-\tilde t_M\,p\in\mathrm{co}\,M_\delta$ (da $x-t_M\,p$, $x-t_P\,p\in M_\delta$). Folglich ist

$$\begin{aligned} f(x)-f(x-t_P\,p)&\geq\frac{1}{2\gamma}\frac{(\nabla f(x),p)^2}{\|p\|^2}-L\,\|p\|\cdot|t_M-t_P|\\ &\geq\frac{1}{2\gamma}\frac{(\nabla f(x),p)^2}{\|p\|^2}-L\,(f(x)-f(x-t_P\,p)), \end{aligned}$$

so daß $f(x)-f(x-t_P\,p)\geq\dfrac{1}{2\gamma(L+1)}\dfrac{(\nabla f(x),p)^2}{\|p\|^2}$, womit die Effizienz der Polak-Schrittweitenfunktion bewiesen ist.

Bemerkung: Statt der Schrittweitenfunktion $T_M(x,p)=\{t_M(x,p)\}$ kann man offenbar eine beliebige andere effiziente Schrittweitenfunktion nehmen und hierdurch eine effiziente Schrittweitenfunktion erzeugen. Etwa folgendermaßen: Sei T eine effiziente Schrittweitenfunktion und $\alpha>0$ eine Konstante. Dann sei für $(x,p)\in A$

$$T_V(x,p):=\{t_V\in(0,\infty):\exists\,t\in T(x,p)\text{ mit }f(x)-f(x-t_V\,p)\geq\alpha\,|t-t_V|\,\|p\|\}.$$

Dann ist auch T_V eine effiziente Schrittweitenfunktion.

Zum Schluß dieses Abschnitts wollen wir noch eine Bemerkung über die auftretenden Effizienzfaktoren machen. Der Effizienzfaktor der Curry-Schrittweitenfunktion ist $\theta_C=\frac{1}{2\gamma}$ ($\alpha=0$ bei Curry-Altmann). Bei der Goldstein-Wahl erhält man $\theta_G(\beta)=\frac{\beta}{\gamma}\to\frac{1}{2\gamma}$ für $\beta\to\frac{1}{2}$, während man bei der Danilin-Psenichnyi-Wahl für $\sigma\geq\frac{1}{\gamma}$, $\varepsilon=\frac{1}{2}$ erhält $\theta_{DP}(\beta)\to\frac{1}{2\gamma}$ für $\beta\to 1$, so daß im Grenzfall die ersten drei auftretenden Schrittweitenfunktionen den gleichen Effizienzfaktor besitzen.

2. Konvergenz von Abstiegsverfahren mit effizienter Schrittweitenfunktion

Zur Lösung des in Abschnitt 0 angegebenen unrestringierten Minimierungsproblems wollen wir den folgenden Algorithmus verwenden, wobei über f die gleichen Voraussetzungen wie in den vorhergehenden Abschnitten gemacht werden.

Algorithmus A:

0) Es sei T eine effiziente Schrittweitenfunktion.

1) Bestimme $x^0 \in M_\delta^*$, setze $k := 0$.
2) Bestimme $p^k \in F(x^k)$.
3) Bestimme $t_k \in T(x^k, p^k)$.
4) Setze $x^{k+1} := x^k - t_k p^k$.
5) Ist $\| \nabla f(x^{k+1}) \| = 0$ dann STOP; andernfalls setze $k := k+1$ und gehe nach 2).

Dieser Algorithmus ist durchführbar, und wenn er nicht nach endlich vielen Schritten mit einem kritischen Punkt abbricht, so liefert er Folgen $\{x^k\}$, $\{p^k\}$ mit

$$\lim_{k \to \infty} \frac{(\nabla f(x^k), p^k)}{\| p^k \|} = 0.$$

Kann hieraus geschlossen werden, daß die Folge $\{x_k\}$ kritisch ist, d. h. daß $\lim_{k \to \infty} \| \nabla f(x^k) \| = 0$ (eine Richtungsfolge $\{p_k\}$ mit dieser Eigenschaft heißt bei Daniel [1] „admissible"), so folgt die gewünschte Konvergenzaussage, daß jeder Häufungspunkt von $\{x^k\}$ ein kritischer Punkt ist. Diese Implikation kann man erzwingen, indem man notfalls den Gradienten der Zielfunktion als Abstiegsrichtung wählt. Wir geben zwei Möglichkeiten an.

Algorithmus B:
Ersetze in Algorithmus A den Schritt 0) durch

0 b) Es sei T eine effiziente Schrittweitenfunktion, $\alpha \in \mathbb{R}^+$, $\sigma \in \mathbb{R}^+$ und $\Phi \in \mathbb{R}^+ \cup \{0\}$;

und den Schritt 2) durch

2 b) Bestimme $\bar{p}^k \in \mathbb{R}^n$, setze

$$\delta_k := \begin{cases} \min(\alpha, \Phi \| \nabla f(x^k) \|) & \text{falls } \Phi \neq 0 \\ \alpha & \text{falls } \Phi = 0 \end{cases}$$

und

$$p^k := \begin{cases} \bar{p}^k & \text{falls } (\nabla f(x^k), \bar{p}^k) > \delta_k \| \nabla f(x^k) \|^\sigma \| \bar{p}^k \| \\ \nabla f(x^k) & \text{sonst.} \end{cases}$$

Algorithmus C:
Ersetze in Algorithmus A den Schritt 0) durch

0 c) Es sei T eine effiziente Schrittweitenfunktion, $\alpha \in \mathbb{R}^+$, $K \in \mathbb{R}^+$, $\sigma \in \mathbb{R}^+$ und $\Phi \in \mathbb{R}^+ \cup \{0\}$;

und den Schritt 2) durch

2 c) Bestimme $\bar{p}^k \in \mathbb{R}^n$, setze

$$\delta_k := \begin{cases} \min(\alpha, \Phi \| \nabla f(x^k) \|) & \text{falls } \Phi \neq 0 \\ \alpha & \text{falls } \Phi = 0 \end{cases}$$

und

$$p^k := \begin{cases} \bar{p}^k & \text{falls } \| \bar{p}^k \| \leq K \text{ und } (\nabla f(x^k), \bar{p}^k) > \delta_k \| \nabla f(x^k) \|^\sigma \\ \nabla f(x^k) & \text{sonst.} \end{cases}$$

In beiden Fällen folgt aus $\lim_{k\to\infty} \frac{(\nabla f(x^k), p^k)}{\| p^k \|} = 0$ daß $\lim_{k\to\infty} \| \nabla f(x^k) \| = 0$ und damit Konvergenz der $\{x^k\}$ im oben angegebenen Sinne.

Die Algorithmen B und C liefern Verfahren, die mindestens so gut sind wie Verfahren des steilsten Abstiegs. Es stellt sich das Problem, wie man im k-ten Iterationsschritt $\bar{p}^k$ zu bestimmen hat, so daß die Wahl $p^k = \bar{p}^k$ möglich ist und das Verfahren besser ist als das Verfahren des steilsten Abstiegs.

Oft wird $\bar{p}^k$ in der Form $\bar{p}^k = H_k \nabla f(x^k)$ mit einer $n \times n$-Matrix H_k angegeben. Dies ist vernünftig, wenn H_k eine gute Approximation an $H_f^{-1}(x^k)$ darstellt, da unter dieser und weiteren Differenzierbarkeits- und Konvexitätsvoraussetzungen superlineare Konvergenz nachgewiesen werden kann. Unter diesen Typ von Verfahren fallen die Quasi-Newton-Verfahren sowie die Verfahren von Goldstein-Price [3] (für dieses Verfahren wird von Polak [7] eine modifizierte Form angegeben), Ritter [8], und Danilin-Psenichnyi [2]. Die Ergebnisse von Ritter und Danilin-Psenichnyi kann man so zusammenfassen:

Ist f zweimal stetig differenzierbar und uniform konvex auf M_δ, so existiert ein $K \in \mathbb{R}^+$ derart, daß der Algorithmus C mit $\Phi = 1$ und $\sigma = 2$ (Ritter) bzw. $\sigma = 1$ (Danilin-Psenichnyi) und beliebigen $\alpha \in \mathbb{R}^+$ entweder nach endlich vielen Schritten mit der Lösung $\hat{x}$ abbricht oder Folgen $\{x^k\}$, $\{p^k\}$ liefert mit $p^k = \bar{p}^k$ für alle hinreichend großen k und $\lim_{k\to\infty} x^k = \hat{x}$ superlinear. Die Matrizen H_k werden hierbei nach Ritter bzw. Danilin-Psenichnyi berechnet, das Verfahren ist mit der Schrittweitenfunktion T_{GP} oder T_C (Ritter) bzw. T_{DP} (Danilin-Psenichnyi) durchzuführen. Ob superlineare Konvergenz auch für andere Schrittweitenfunktionen nachgewiesen werden kann, ist uns nicht bekannt. Es ist uns nicht gelungen, das Verfahren von Goldstein-Price-Polak in Beziehung zu Algorithmus C zu setzen.

Wir wenden uns nun der Untersuchung der Konvergenz von Algorithmus A zu.

Lemma 2.1: *Der Algorithmus A breche nicht nach endlich vielen Schritten ab und liefere Folgen $\{x^k\}$, $\{p^k\}$. f sei zweimal stetig differenzierbar und uniform konvex auf M_δ, M_δ konvex. Setzt man für*

$$k = 0, 1, \ldots \quad \alpha_k := \frac{(\nabla f(x^k), p^k)}{\| \nabla f(x^k) \| \, \| p^k \|},$$

so existiert eine Konstante $\xi \in \mathbb{R}^+$ derart, daß

$$f(x^{k+1}) - f(\hat{x}) \le \exp\left(-\theta \xi \cdot \sum_{l=0}^{k} \alpha_l^2 \right) (f(x^0) - f(\hat{x}))$$

fur $k = 0, 1, \ldots$ gilt; hierbei ist $\hat{x}$ das eindeutig bestimmte Minimum von f auf M_δ und θ der Effizienzfaktor der Schrittweitenfunktion von Algorithmus A.

Beweis: Es existieren $m, M \in \mathbb{R}^+$ mit

$$m \| h \|^2 \le (h, H_f(x) h) \le M \| h \|^2 \text{ für alle } h \in \mathbb{R}^n,\ x \in M_\delta,$$

wobei $H_f(x)$ die Hessesche von f an der Stelle x ist.

Für alle $x \in M_\delta$ ist dann

$$f(x)-f(\hat{x}) \leq \frac{M}{2m^2} \| \nabla f(x) \|^2$$

und

$$\frac{m}{2M^2} \| \nabla f(x) \|^2 \leq f(x)-f(\hat{x}).$$

Damit wird

$$f(x^k)-f(x^{k+1}) \geq \theta \frac{(\nabla f(x^k), p^k)^2}{\| p^k \|^2} = \theta \alpha_k^2 \| \nabla f(x^k) \|^2$$

$$\geq \theta \frac{2m^2}{M} \alpha_k^2 \left(f(x^k)-f(\hat{x})\right).$$

Setzt man $\xi := \frac{2m^2}{M}$ so erhalten wir für $k=0, 1, \ldots$

$$f(x^{k+1})-f(\hat{x}) \leq (1-\theta\,\xi\,\alpha_k^2)\left(f(x^k)-f(\hat{x})\right),$$

hieraus

$$f(x^{k+1})-f(\hat{x}) \leq \prod_{l=0}^{k} (1-\theta\,\xi\,\alpha_l^2)\left(f(x^0)-f(\hat{x})\right)$$

und hieraus die Behauptung.

Hieraus erhält man

Satz 2.2: *Der Algorithmus A breche nicht nach endlich vielen Schritten ab und liefere Folgen $\{x^k\}$, $\{p^k\}$. Die Folge $\{\alpha_k\}$ sei wie in Lemma 2.1 definiert. Dann gilt:*

i) Ist $\sum_{k=0}^{\infty} \alpha_k^2 = \infty$, so besitzt $\{x^k\}$ einen Häufungspunkt, der kritisch für f ist.

ii) Ist f zweimal stetig differenzierbar und uniform konvex auf M_δ, M_δ konvex, so ist die Zoutendijk-Bedingung $\sum_{k=0}^{\infty} \alpha_k^2 = \infty$ notwendig und hinreichend dafür, daß $\lim_{k\to\infty} x^k = \hat{x}$, wobei $\hat{x}$ das eindeutig bestimmte Minimum von f auf M_δ ist.

Beweis: i) θ sei der Effizienzfaktor der Schrittweitenfunktion in Algorithmus A. Dann gilt $f(x^k)-f(x^{k+1}) \geq \theta \alpha_k^2 \| \nabla f(x^k) \|^2$ für $k=0, 1, \ldots$ und daher $\sum_{k=0}^{\infty} \alpha_k^2 \| \nabla f(x^k) \|^2 < \infty$. Ist jeder Häufungspunkt von $\{x^k\}$ nichtkritisch, so existiert $\varepsilon \in \mathbb{R}^+$ mit $\| \nabla f(x^k) \| \geq \varepsilon$ für $k=0, 1, \ldots$, so daß

$$\varepsilon^2 \sum_{k=0}^{\infty} \alpha_k^2 \leq \sum_{k=0}^{\infty} \alpha_k^2 \| \nabla f(x^k) \|^2 < \infty,$$

ein Widerspruch.

ii) a) Ist $\sum_{k=0}^{\infty} \alpha_k^2 = \infty$, so folgt aus Lemma 2.1, daß $\lim_{k=\infty} f(x^k) = f(\hat{x})$ und hieraus folgt $\lim_{k\to\infty} x^k = \hat{x}$.

b) Sei $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$. $m, M \in \mathbb{R}^+$ seien wie im Beweis von Lemma 2.1 gewählt. Dann gilt

$$f(x^{k+1}) - f(x^k) \geq -t_k (\nabla f(x^k), p^k) + t_k^2 \frac{m}{2} \| p^k \|^2$$

$$\geq -\frac{1}{2m} \frac{(\nabla f(x^k), p^k)^2}{\| p^k \|^2} = -\frac{1}{2m} \alpha_k^2 \| \nabla f(x^k) \|^2$$

$$\geq -\frac{M^2}{m^2} \alpha_k^2 (f(x^k) - f(\hat{x})) \quad \text{und damit}$$

$$f(x^{k+1}) - f(\hat{x}) \geq \left(1 - \frac{M^2}{m^2} \alpha_k^2\right) (f(x^k) - f(\hat{x})).$$

Wegen $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ gilt $\lim_{k \to \infty} \alpha_k = 0$, daher existiert ein $k_0 \in \mathbb{N}$ mit $\frac{M^2}{m^2} \alpha_k^2 \leq \frac{1}{2}$ bzw. $1 - \frac{M^2}{m^2} \alpha_k^2 \geq \frac{1}{2}$ für alle $k \geq k_0$.

Dann gilt für alle $k \geq k_0$

$$f(x^k) - f(\hat{x}) \geq \prod_{l=0}^{k-k_0-1} \left(1 - \frac{M^2}{m^2} \alpha_{k_0+l}^2\right) (f(x^{k_0}) - f(\hat{x}))$$

$$\geq \exp\left(-\frac{2M^2}{m^2} \sum_{l=0}^{k-k_0-1} \alpha_{k_0+l}^2\right) (f(x^{k_0}) - f(\hat{x}))$$

$$\geq \exp\left(-\frac{2M^2}{m^2} \sum_{k=0}^{\infty} \alpha_k^2\right) (f(x^{k_0}) - f(\hat{x})),$$

d. h. $x^k \not\to \hat{x}$.

Bemerkung 1: Die Voraussetzungen von Satz 2.2 seien gegeben, es sei $N \subset \mathbb{N}$ eine nicht-endliche Teilmenge mit $p^k = \nabla f(x^k)$ für alle $k \in N$. Dann ist $\alpha_k = 1$ für alle $k \in N$ und daher $\sum_{k=0}^{\infty} \alpha_k^2 = \infty$. Ist der Algorithmus A also so beschaffen, daß unendlich oft eine Reinitialisierung auf die Gradientenrichtung erfolgt, so existiert eine Teilfolge von $\{x^k\}$, die gegen einen kritischen Punkt von f konvergiert (siehe hierzu Ortega-Rheinboldt [6]).

Bemerkung 2: Die Schrittweitenfunktion in Algorithmus A werde folgendermaßen modifiziert: T sei eine gegebene effiziente Schrittweitenfunktion mit Effizienzfaktor θ.

Schritt 3) werde ersetzt durch

3′) Bestimme $t_k > 0$ mit $f(x^k) - f(x^k - t_k p^k) > 0$.

Dann gilt: Existiert zu jedem $t_k > 0$ ein $t_k' \in T(x^k, p^k)$ und ist $\sum_{k=0}^{\infty} \sigma_k < \infty$ mit $\sigma_k := | t_k - t_k' |$, existiert ferner eine Konstante $K > 0$ mit $\| p^k \| \leq K$ für alle k, so bleiben die Aussagen von Satz 2.2 bestehen.

Denn: Sei $\gamma>0$ eine Konstante mit $|f(y)-f(z)|\leq\gamma \| y-z \|$ für alle $y, z\in M_\delta$. Dann ist

$$\begin{aligned} f(x^k)-f(x^{k+1}) &= f(x^k)-f(x^k-t_k' p^k)+f(x^k-t_k' p^k)-f(x^k-t_k p^k) \\ &\geq \theta \frac{(\nabla f(x^k), p^k)^2}{\| p^k \|^2} - \sigma_k \gamma \| p^k \| \\ &\geq \theta \frac{(\nabla f(x^k), p^k)^2}{\| p^k \|^2} - \sigma_k \gamma \cdot K, \end{aligned}$$

woraus man die Behauptung leicht ableiten kann (siehe Ritter [8]).

Bemerkung 3: Existieren $\alpha\in\mathbb{R}^+$, $\Phi\in\mathbb{R}^+\cup\{0\}$ so, daß mit

$$\delta_k := \begin{cases} \alpha & \text{falls } \Phi=0 \\ \min(\alpha, \Phi \| \nabla f(x^k) \|) & \text{falls } \Phi\neq 0 \end{cases}$$

gilt $\alpha_k\geq\delta_k$ $(k=0, 1, \ldots)$, so ist jeder Häufungspunkt der durch den Algorithmus A erzeugten Folge $\{x^k\}$ kritisch.

Denn aus $\frac{(\nabla f(x^k), p^k)}{\| p^k \|}=\alpha_k^2 \| \nabla f(x^k) \|^2\to 0$ folgt $\delta_k \| \nabla f(x^k) \| \to 0$ und hieraus $\| \nabla f(x^k) \| \to 0$.

So folgt z. B. unter den Voraussetzungen von Bemerkung 2 mit $\lim_{k\to\infty} \sigma_k=0$ statt $\sum_{k=0}^{\infty} \sigma_k<\infty$ daß $\frac{(\nabla f(x^k), p^k)}{\| p^k \|}\to 0$; eine Kombination mit obiger Bemerkung ergibt eine hinreichende Bedingung für die Aussage, daß jeder Häufungspunkt der Folge $\{x^k\}$ kritisch ist, wobei $\{x^k\}$ gemäß Bemerkung 2 konstruiert ist.

Die Zoutendijk-Bedingung $\sum_{k=0}^{\infty} \alpha_k^2=\infty$ findet sich bei Zoutendijk [11] und auch bei Wolfe [10], Lenard [4].

Der folgende Algorithmus gibt eine Überleitung zu den *CG*-Verfahren (Verfahren der konjugierten Gradienten).

Algorithmus D:

Ersetze in Algorithmus A den Schritt 0) durch

0 d) Es sei T eine effiziente Schrittweitenfunktion, $\eta\in\mathbb{R}^+$, $\lambda\in\mathbb{R}^+$ und den Schritt 2) durch

2 d) Bestimme $\beta_{k-1}\in\mathbb{R}^+$ (falls $k\geq 1$) und setze

$$\bar{p}^k := \begin{cases} \nabla f(x^0) & \text{falls } k=0 \\ \nabla f(x^k)-\beta_{k-1} p^{k-1} & \text{falls } k\geq 1 \end{cases}$$

Dann sei

$$p^k := \begin{cases} \bar{p}^k \text{ falls } \| \beta_{k-1} p^{k-1} \| \leq\eta \| \nabla f(x^k) \| \text{ und } (\nabla f(x^k), \bar{p}^k)>\lambda \| \nabla f(x^k) \|^2 \\ \nabla f(x^k) \text{ sonst.} \end{cases}$$

Dieser Algorithmus ist ein Spezialfall von Algorithmus C, man setze dort $\Phi=0$, $\sigma=2$, $\alpha=\lambda$ und $K=(1+\eta)\max_{x\in M_\delta}\|\nabla f(x)\|$. Ein Algorithmus dieser Art ist von Mc Cormick und Ritter [15] angegeben worden. Zusätzlich zu unseren Forderungen wird unendlich oft $p^k=\nabla f(x^k)$ gewählt, jedoch $\lambda=0$ ist dann zugelassen. Die Konvergenzaussage folgt dann mit Berücksichtigung von Bemerkung 1.

3. CG-Verfahren mit effizienter Schrittweitenfunktion

Der Algorithmus D stellt die allgemeinste Form eines CG-Verfahrens dar. Wir wollen nun die Konvergenz spezieller, praktisch durchführbarer CG-Verfahren untersuchen.

a) Das Verfahren von Fletcher-Reeves [12].

1) Bestimme $x_0\in M_\delta^*$; $\alpha,\lambda\in[0,1)$.

2) Setze $p^0:=\nabla f(x^0)$, $k:=0$.

3) Bestimme $t_k\in(0,\hat{t}(x^k,p^k)]$ ($\hat{t}(x^k,p^k)$ ist hierbei wieder die Curry-Schrittweite, d. h. die erste positive Nullstelle von $(\nabla f(x^k-t\,p^k),p^k)$) so, daß

$$(\nabla f(x^k-t_k\,p^k),p^k)\le\min\left(\alpha(\nabla f(x^k),p^k),\ \lambda\,\|\nabla f(x^k)\|^2\right).$$

4) Wähle $\beta_k\in\mathbb{R}$ mit $|\beta_k|\le\dfrac{\|\nabla f(x^k-t_k\,p^k)\|^2}{\|\nabla f(x^k)\|^2}$.

5) Setze $x^{k+1}:=x^k-t_k\,p^k$, $p^{k+1}:=\nabla f(x^{k+1})-\beta_k\,p^k$.

6) Ist $\nabla f(x^{k+1})=\theta$, dann STOP; andernfalls setze $k:=k+1$ und gehe nach 3).

Es handelt sich hier um ein Verfahren mit effizienter Schrittweitenfunktion; das Verfahren ist durchführbar, d. h. t_k und β_k sind wählbar und wenn $\nabla f(x^{k+1})\neq\theta$ ist, so ist $(\nabla f(x^{k+1}),p^{k+1})\ge(1-\lambda)\,\|\nabla f(x^{k+1})\|^2>0$ und somit $p^{k+1}\in F(x^{k+1})$.

Satz 3.1: *Es seien die generellen Voraussetzungen aus Abschnitt* 0 *gegeben, man betrachte das oben angegebene Verfahren von Fletcher-Reeves und nehme an, daß der Algorithmus nicht schon nach endlich vielen Schritten mit einem kritischen Punkt abbricht. Dann besitzt die Folge* $\{x^k\}$ *einen kritischen Häufungspunkt. Ist* f *gleichmäßig konvex auf* M_δ, M_δ *konvex, so folgt die Konvergenz der Folge* $\{x^k\}$ *gegen das Minimum.*

Beweis: Sei $K:=\max_{x\in M_\delta}\|\nabla f(x)\|$. Dann gilt

$$\frac{(\nabla f(x^{k+1}),p^{k+1})}{\|\nabla f(x^{k+1})\|\ \|p^{k+1}\|}\ge(1-\lambda)\,\frac{\|\nabla f(x^{k+1})\|}{\|p^{k+1}\|}\ge\frac{(1-\lambda)}{K}\,\frac{\|\nabla f(x^{k+1})\|^2}{\|p^{k+1}\|}.$$

Eine einfache Abschätzung liefert dann

$$\|p^{k+1}\|^2\le(1+2\lambda)\,\|\nabla f(x^{k+1})\|^2+\frac{\|\nabla f(x^{k+1})\|^4}{\|\nabla f(x^k)\|^4}\,\|p^k\|^2$$

und hieraus folgt

$$\frac{\| p^{k+1} \|^2}{\| \nabla f(x^{k+1}) \|^4} \le \frac{1+2\lambda}{\| \nabla f(x^{k+1}) \|^2} + \frac{\| p^k \|^2}{\| \nabla f(x^k) \|^4}$$

$$\le (1+2\lambda) \sum_{l=0}^{k+1} \frac{1}{\| \nabla f(x^l) \|^2}.$$

Wir nehmen nun an, daß ein $\varepsilon > 0$ mit $\| \nabla f(x^k) \| \ge \varepsilon$ für alle $k \in \mathbb{N}$ existiert. Dann erhalten wir

$$\alpha_{k+1}^2 = \frac{(\nabla f(x^{k+1}), p^{k+1})^2}{\| \nabla f(x^{k+1}) \|^2 \cdot \| p^{k+1} \|^2} \ge \frac{(1-\lambda)^2}{K^2} \frac{\| \nabla f(x^{k+1}) \|^4}{\| p^{k+1} \|^2}$$

$$\ge \frac{(1-\lambda)^2}{K^2} \frac{2}{(1+2\lambda)(k+2)},$$

und damit $\sum_{k=0}^{\infty} \alpha_k^2 = \infty$. Aus Satz 2.2 i) folgt die Existenz eines kritischen Häufungspunktes, ein Widerspruch zur Annahme, womit die Existenz eines kritischen Häufungspunktes schließlich doch bewiesen ist.

Bemerkung: Wählt man $\beta_k = \frac{\| \nabla f(x^{k+1}) \|^2}{\| \nabla f(x^k) \|^2}$ und $t_k = \hat{t}(x^k, p^k)$, so erhält man das klassische Verfahren von Fletcher-Reeves. Die hier vorgeschlagene Wahl von β_k stammt von Mangasarian (nach Zitat bei McCormick-Ritter [15]).

b) Die Verfahren von Daniel [1], Polak-Ribière [13] und Wolfe (nach Zitat bei Lenard [4]).

1) Bestimme $x_0 \in M_\delta^*$, $\alpha \in [0, 1)$ und $\lambda \in (0, 1]$.

2) Setze $p^0 := \nabla f(x^0)$, $k := 0$.

3) Bestimme die Curry-Schrittweite $\hat{t}_k = \hat{t}(x^k, p^k)$. Wenn $\nabla f(x^k - \hat{t}_k p^k) = \theta$, dann STOP, andernfalls:

4) Bestimme $t_k \in (0, \hat{t}_k]$ so, daß

$$(\nabla f(x^k - t_k p^k), p^k) \le \alpha (\nabla f(x^k), p^k)$$

und

$$(1-\lambda) \| \nabla f(x^k - t_k p^k) \|^2 - \beta_k(t_k) (\nabla f(x^k - t_k p^k), p^k) \ge 0$$

gilt. Hierbei β_k eine der folgenden Funktionen:

$$\beta_k(t) = \begin{cases} \dfrac{(\nabla f(x^k - t p^k), H_f(x^k - t p^k) p^k)}{(p^k, H_f(x^k - t p^k) p^k)} & \text{(Daniel)} \\[2ex] -\dfrac{(\nabla f(x^k - t p^k), \nabla f(x^k - t p^k) - \nabla f(x^k))}{\| \nabla f(x^k) \|^2} & \text{(Polak-Ribière)} \\[2ex] \dfrac{(\nabla f(x^k - t p^k), \nabla f(x^k - t p^k) - \nabla f(x^k))}{(p^k, \nabla f(x^k - t p^k) - \nabla f(x^k))} & \text{(Wolfe)} \end{cases}$$

(Bei dem Verfahren von Daniel muß vorausgesetzt werden, daß f zweimal stetig differenzierbar ist.)

5) Setze $x^{k+1} := x^k - t_k p^k$, $p^{k+1} := \nabla f(x^{k+1}) - \beta_k(t_k) p^k$. Gehe nach 3).

Es handelt sich hier um Verfahren mit effizienter Schrittweitenfunktion, die Verfahren sind durchführbar, d. h. t_k und $\beta_k = \beta_k(t_k)$ sind wählbar, wenn $\nabla f(x^k - \hat{t}_k p^k) \neq \theta$ ist. Es ist dann $(\nabla f(x^{k+1}), p^{k+1}) \geq \lambda \| \nabla f(x^{k+1}) \|^2 > 0$, also $p^{k+1} \in F(x^{k+1})$.

Ist f zweimal stetig differenzierbar und gleichmäßig konvex auf M_δ, so können wir die Existenz einer Konstanten $\eta \in \mathbb{R}^+$ mit $\| \beta_k p^k \| \leq \eta \| \nabla f(x^{k+1}) \|$ nachweisen, wobei β_k nach Daniel, Polak-Ribière oder Wolfe zu bestimmen ist. Dann können wir das Verfahren als Spezialfall von Algorithmus D auffassen. Die klassischen Verfahren erhält man, wenn man stets $t_k = \hat{t}_k$ wählt. Das Verfahren von Polak-Ribière wurde auch von Polyak [14] angegeben, dort mit Reinitialisierung. Nach unseren Überlegungen im vorigen Abschnitt sind jedoch Konvergenzaussagen für solche Verfahren einfach.

Literatur

[1] Daniel, J. W.: The approximate minimization of functionals. Englewood Cliffs, N.J.: Prentice-Hall 1971.

[2] Danilin, Y. M., Psenichnyi, B. N.: On methods of minimization with accelerated convergence. USSR, Comp. Math. a. Math. Phys. *10*, 4—19 (1972).

[3] Goldstein, A. A., Price, I. F.: An effective algorithm for minimization. Num. Math. *10*, 184—189 (1967).

[4] Lenard, M. L.: Practical convergence conditions for unconstrained optimization. Math. Prog. *4*, 309—323 (1973).

[5] Ortega, J. M., Rheinboldt, W. C.: Iterative solution of nonlinear equations in several variables. New York: Academic Press 1970.

[6] Ortega, J. M., Rheinboldt, W. C.: A general convergence result for unconstrained minimization methods. SIAM J. Num. Anal. *9*, 40—43 (1972).

[7] Polak, E.: Computational methods in optimization. New York: Academic Press 1971.

[8] Ritter, K.: A superlinearly convergent method for unconstrained minimization problems, in: Nonlinear programming (Rosen, I. B., Mangasarian, O. L., Ritter, K., eds.). New York: Academic Press 1970.

[9] Wolfe, Ph.: Convergence conditions for ascent methods. SIAM Rev. *11*, 226—235 (1969).

[10] Wolfe, Ph.: Convergence theory in nonlinear programming, in: Integer and nonlinear programming (Abadie, J., ed.). Amsterdam: North-Holland 1970.

[11] Zoutendijk, G.: Some algorithms based on the principle of feasible directions, in: Integer and nonlinear programming (Abadie, J., ed.). Amsterdam: North-Holland 1970.

[12] Fletcher, R., Reeves, C. M.: Function minimization by conjugate gradients. Comp. J. *7*, 149—154 (1964).

[13] Polak, E., Ribière, G.: Note sur la convergence de méthodes de direction conjugées. RIRO *3*, 35—44 (1969).

[14] Polyak, B. T.: Conjugate gradient methods in extremum problems. USSR Comp. Math. a. Math. Phys. *9*, 94—112 (1969).

[15] Mc Cormick, G. P., Ritter, K.: Alternative proofs of the convergence of the conjugate method. JOTA *13*, 497—518 (1974).

Dr. W. Warth und Prof. Dr. J. Werner
Lehrstuhl für Numerische
und Angewandte Mathematik
Universität Göttingen
Lotzestraße 16-18
D-3400 Göttingen
Bundesrepublik Deutschland