5 Schrittlängensteuerung

Wir knüpfen an den Abschnitt 3 über Abstiegsverfahren an und beschäftigen uns mit dem dritten Punkt des Modellalgorithmus (3.5):

$$\text{3.) Bestimmung eine Schrittlänge } \alpha_k > 0 \text{ mit } f(\boldsymbol{x}_k + \alpha_k \boldsymbol{s}_k) < f(\boldsymbol{x}_k)\,. \tag{5.1}$$

Um die Konvergenz des Abstiegsverfahrens garantieren zu können, wurde über (5.1) hinaus die Effizienzbedingung für die Schrittlänge α_k gefordert, d.h. wir versuchen, die Schrittlänge so zu bestimmen, daß eine Abschätzung der Form

$$f(\boldsymbol{x}_k + \alpha_k\, \boldsymbol{s}_k) \;\leq\; f(\boldsymbol{x}_k) \,-\, \delta \left(\frac{\boldsymbol{\nabla} f(\boldsymbol{x}_k)^{\mathrm{T}}\, \boldsymbol{s}_k}{\|\boldsymbol{s}_k\|}\right)^2 . \tag{5.2}$$

mit einer von $\boldsymbol{x}_k$ und $\boldsymbol{s}_k$ unabhängigen Konstanten $\delta > 0$ gilt.

Wir haben gesehen, daß die sogenannte exakt Schrittlänge

$$\alpha^* \;:=\; \alpha^*(\boldsymbol{x}, \boldsymbol{s}) \;:=\; \arg\min\{f(\boldsymbol{x} + \alpha \boldsymbol{s}) : \alpha \geq 0\} \tag{5.3}$$

(Existenz vorausgesetzt) die Effizienzbedingung (5.2) erfüllt. Allerdings ist die tatsächliche (genaue) Berechnung der exakten Schrittlänge α^* i.a. viel zu aufwendig. Man hat ja hierzu in jedem Iterationsschritt eine eindimensionale Optimierungsaufgabe zu lösen. Es ist daher naheliegend, (5.3) durch eine weniger aufwendige Strategie zu ersetzen. Eine Reihe von Standard-Verfahren zu Schrittlängenbestimmung sind hierfür gebräuchlich, von denen wir im folgenden einige vorstellen wollen.

Generell setzen wir wieder voraus, daß $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ eine hinreichend oft stetig differenzierbare Funktion und $\boldsymbol{x} \in \mathbb{R}^n$ eine aktuelle Näherung mit $\boldsymbol{\nabla} f(\boldsymbol{x}) \neq \boldsymbol{0}$ ist. Ferner sei die Niveaumenge $L(\boldsymbol{x})$ kompakt (vgl. (1.2)) und $\boldsymbol{s} \in \mathbb{R}^n$ sei eine Abstiegsrichtung von f in $\boldsymbol{x}$.

A. Schrittlänge nach Goldstein

Unter den genannten Voraussetzungen gibt es eine kleinste Schrittlänge $\alpha_1 > 0$ mit $f(\boldsymbol{x} + \alpha_1 \boldsymbol{s}) = f(\boldsymbol{x})$. Insbesondere ist daher

$$\forall\, \alpha \;:\; 0 < \alpha < \alpha_1 \;\Longrightarrow\; f(\boldsymbol{x} + \alpha \boldsymbol{s}) < f(\boldsymbol{x})\,.$$

Die (stetig differenzierbare) Funktion $\phi(\alpha) := f(\boldsymbol{x} + \alpha \boldsymbol{s})$ hat demnach die in Abb. 5.1 angegebene Gestalt.

Wir sind an Schrittlängen α interessiert, für die $(\alpha, \phi(\alpha))$ in dem von den Halbgeraden $\phi = \phi(0) + \alpha\phi'(0)$ und $\phi \equiv \phi(0)$, $\alpha \geq 0$ berandeten Kegel liegt ($0 < \alpha < \alpha_1$). Um die Randlagen ($\alpha \approx 0$ und $\alpha \approx \alpha_1$) zu vermeiden, schränken wir den Kegel weiter ein und setzen für einen Parameter ρ mit $0 < \rho < 0.5$:

$$\begin{aligned} \phi_1(\alpha) &:= \phi(0) + (1-\rho)\,\alpha\,\phi'(0) \\ \phi_2(\alpha) &:= \phi(0) + \rho\,\alpha\,\phi'(0)\,. \end{aligned} \tag{5.4}$$

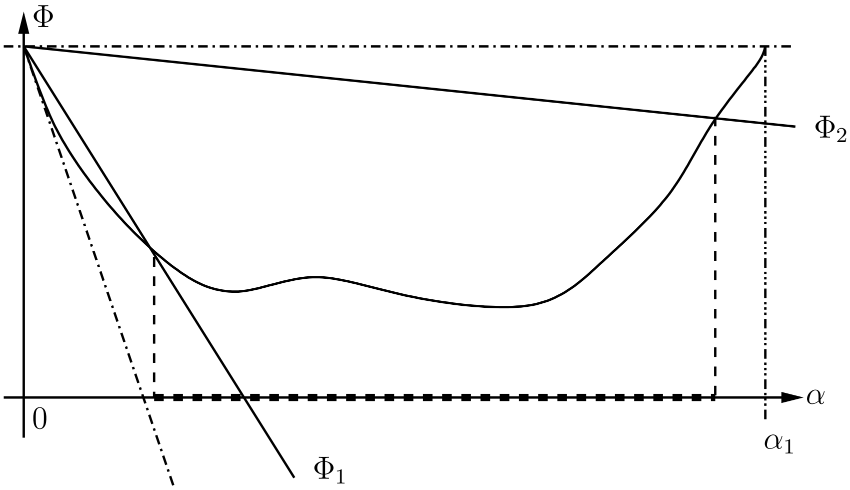

Abb. 5.1. Goldstein Schrittlänge

Wir sagen nun, daß eine Schrittlänge α den **Goldstein–Bedingungen** genügt, falls

$$\phi_1(\alpha) \;\le\; \phi(\alpha) \;\le\; \phi_2(\alpha)\,. \tag{5.5}$$

Aufgrund der genannten Voraussetzungen an die Funktion f und $\rho \in]0, 0.5[$ gibt es stets Schrittlängen, die den Goldstein–Bedingungen genügen. Zur numerischen Berechnung einer solchen Schrittlänge läßt sich folgender Algorithmus verwenden:

Algorithmus (5.6) (Goldstein-Schrittlänge)

(i) Setze $\alpha := 0$, $\beta > 0$ („geeignet“).

(ii) Solange $(\phi(\beta) < \phi_1(\beta))$: $\alpha := \beta$, $\beta := 2\,\alpha$.

(iii) Gilt $\phi(\beta) \le \phi_2(\beta)$, so setze man $\alpha := \beta$.

Andernfalls: **Wiederhole :** $\gamma := (\alpha+\beta)/2$,

Falls $\phi(\gamma) < \phi_1(\gamma)$: $\alpha := \gamma$,

Falls $\phi(\gamma) > \phi_2(\gamma)$: $\beta := \gamma$,

solange bis : $\phi_1(\gamma) \le \phi(\gamma) \le \phi_2(\gamma)$.

$\alpha := \gamma$.

Bemerkungen (5.7)

a) Der Schritt (ii) (Expansionsphase) des Algorithmus liefert ein Intervall $[\alpha, \beta]$, für das $\phi(\alpha) < \phi_1(\alpha)$ und $\phi(\beta) \geq \phi_1(\beta)$ gelten. Gilt nun sogar $\phi(\beta) \leq \phi_2(\beta)$, so ist β bereits eine zulässige Goldstein-Schrittlänge. Andernfalls wird das Intervall in Schritt (iii) (Kontraktionsphase) mit einer Halbierungsstrategie verkleinert, wobei α und β stets die Ungleichungen $\phi(\alpha) < \phi_1(\alpha)$ und $\phi(\beta) \geq \phi_2(\beta)$ erfüllen. Da es Schrittlängen gibt, die die Goldstein-Bedingungen erfüllen, muß die Schleife nach endlich vielen Schritten mit einer zulässigen Goldstein-Schrittweite abbrechen.

b) Natürlich kann man sowohl die Expansions- wie die Kontraktionsphase geeignet modifizieren. So kann man etwa in Schritt (ii) anstelle von $\beta := 2\,\alpha$ setzen $\beta := \lambda\,\alpha$ mit einem $\lambda \geq \lambda_{\min} > 1$.

Analog kann man die Halbierungsstrategie ersetzen durch

$$\gamma \;:=\; \alpha \;+\; \mu\,(\beta - \alpha)\,, \;\; \mu \in [\mu_{\min}, \mu_{\max}],$$

wobei $0 < \mu_{\min} < \mu_{\max} < 1$ fest vorgegeben sind.

c) Eine praxisgerechte Wahl für den Parameter ρ ist $\rho = 1/4$ oder $\rho = 1/10$.

d) Um bei Wahl der Newton-Suchrichtung die schnelle (quadratische) Konvergenz des Abstiegsverfahrens zu sichern (vgl. Abschnitt 6), ist es sinnvoll, vor dem Start des obigen Algorithmus die Schrittlänge $\alpha = 1$ auf die Gültigkeit von (5.5) zu testen.

Satz (5.8)

Eine Schrittlängenstrategie, die die Goldstein-Bedingungen erfüllt, ist effizient. Genauer gilt mit $C := \max\{\|\boldsymbol{\nabla}^2 f(\boldsymbol{x})\| \,:\, \boldsymbol{x} \in L(\boldsymbol{x}^0)\}$ *($\boldsymbol{x}^0$ fester universeller Startvektor) für alle* $\boldsymbol{x} \in L(\boldsymbol{x}^0)$, $\boldsymbol{\nabla} f(\boldsymbol{x}) \neq \boldsymbol{0}$, *$\boldsymbol{s}$: Abstiegsrichtung von f in $\boldsymbol{x}$ und $\alpha > 0$ mit (5.5):*

$$f(\boldsymbol{x} + \alpha\,\boldsymbol{s}) \;\leq\; f(\boldsymbol{x}) \;-\; \frac{\rho}{C}\left(\frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}}{\|\boldsymbol{s}\|}\right)^2 .$$

Beweis: Es sei α_0 das kleinste positive lokale Minimum der Funktion $\phi(\alpha)$. Dann ist $\phi(\alpha)$ im Interval $[0, \alpha_0]$ monoton fallend.

1. Fall: Es gelte $\alpha \leq \alpha_0$. Nach Voraussetzung (5.5) folgt

$$\phi_1(\alpha) \;=\; \phi(0) \;+\; (1-\rho)\,\alpha\,\phi'(0) \;\leq\; \phi(\alpha),$$

woraus sich durch Umformung und Anwendung des Taylorschen Satzes ergibt

$$(1-\rho)\,\alpha\,\phi'(0) \;\leq\; \phi(\alpha) \;-\; \phi(0) \;=\; \alpha\,\phi'(0) \;+\; \frac{\alpha^2}{2}\,\phi''(\theta\,\alpha) \;\leq\; \alpha\,\phi'(0) \;+\; \frac{\alpha^2}{2}\,C\,\|\boldsymbol{s}\|^2 .$$

Hieraus folgt die Abschätzung $\alpha \;\geq\; -\,\dfrac{2\,\rho}{C}\,\dfrac{\phi'(0)}{\|\boldsymbol{s}\|^2} \;=:\; \tilde{\alpha} > 0.$

Einsetzen von $\tilde{\alpha}$ und Ausnutzung der Monotonie von ϕ liefert

$$
\begin{aligned}
\phi(\alpha) \;&\leq\; \phi(\tilde{\alpha}) \\
&\leq\; \phi(0) \;+\; \tilde{\alpha}\,\phi'(0) \;+\; \tilde{\alpha}^2/2\,C\,\|\boldsymbol{s}\|^2 \\
&=\; \phi(0) - \frac{2\,\rho\,(1-\rho)}{C}\,\left(\frac{\phi'(0)}{\|\boldsymbol{s}\|}\right)^2 \\
&\leq\; \phi(0) - \frac{\rho}{C}\,\left(\frac{\phi'(0)}{\|\boldsymbol{s}\|}\right)^2 .
\end{aligned}
$$

2. Fall: $\alpha > \alpha_0$. Aus $0 \;=\; \phi'(\alpha_0) \;=\; \phi'(0) \;+\; \alpha_0\,\phi''(\theta\alpha_0) \;\leq\; \phi'(0) \;+\; \alpha_0\,C\,\|\boldsymbol{s}\|^2$ folgt zunächst die Abschätzung

$$
\alpha_0 \;\geq\; -\,\frac{\phi'(0)}{C\,\|\boldsymbol{s}\|^2}\,.
$$

Mit der Voraussetzung (5.5) folgt dann wie oben

$$
\begin{aligned}
\phi(\alpha) \;&\leq\; \phi_2(\alpha) \\
&=\; \phi(0) \;+\; \rho\,\alpha\,\phi'(0) \\
&\leq\; \phi(0) \;+\; \rho\,\alpha_0\,\phi'(0) \\
&\leq\; \phi(0) - \frac{\rho}{C}\,\left(\frac{\phi'(0)}{\|\boldsymbol{s}\|}\right)^2 .
\end{aligned}
$$

□

B. Schrittlänge nach Armijo

Eine (einfachere) Variante des obigen Algorithmus besteht in einem Verfahren, welches nur die zweite Goldstein-Bedingung

$$
\phi(\alpha) \;\leq\; \phi_2(\alpha) \;=\; \phi(0) \;+\; \rho\,\alpha\,\phi'(0) \tag{5.9}
$$

garantiert und mit einer einfachen Halbierungsstrategie auskommt. Hier läßt sich direkt mit der für das Newton-Verfahren wichtigen Schrittlänge $\alpha = 1$ starten, man vergleiche die Bemerkung (5.7) und Abschnitt 6.

Algorithmus (5.10) (Armijo-Schrittlänge)

(i) Setze $\alpha \;:=\; 1\,.$

(ii) Solange $(\phi(\alpha) \;>\; \phi_2(\alpha))\;:\; \alpha \;:=\; \alpha/2\,.$

Bemerkungen (5.11)

a) Wie bei der Goldstein-Schrittlänge läßt sich die Halbierungsstrategie modifizieren.

So kann man etwa in Schritt (ii) setzen $\alpha := \lambda\,\alpha$ mit einem $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, wobei $0 < \lambda_{\min} \leq \lambda_{\max} < 1$ fest vorgegeben werden.

b) Eine Schrittlängenbestimmung nach Armijo ist i.a. nicht effizient. Allerdings läßt sich eine abgeschwächte Aussage beweisen: Es gibt eine von $\boldsymbol{x}$ und $\boldsymbol{s}$ unabhängige Konstante $\delta > 0$ mit

$$f(\boldsymbol{x} + \alpha\,\boldsymbol{s}) \;\leq\; f(\boldsymbol{x}) \;-\; \delta\,\min\left\{ -\,\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}\,,\; \left(\frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}}{\|\boldsymbol{s}\|} \right)^2 \right\}.$$

Unter der Voraussetzung dieser abgeschwächten Effizienz läßt sich ferner ebenfalls ein Konvergenzsatz - ähnlich zu Satz (3.27) - beweisen; man vergleiche hierzu Werner[...].

C. Die Bedingungen nach Wolfe, Powell und Fletcher

Die erste Goldstein-Bedingung

$$\phi(\alpha) \;\geq\; \phi_1(\alpha) \;=\; \phi(0) \,+\, (1-\rho)\,\alpha\,\phi'(0) \tag{5.12}$$

hat den Nachteil, daß sie möglicherweise das exakte Minimum der Funktion $\phi(\alpha)$ ausschließt. Nun läßt sich aber etwa bei Newton-ähnlichen Verfahren, die mit der Hesse-Matrix bzw. mit einer Approximation dieser Matrix arbeiten, in der Nähe des gesuchten Minimums eine gute Näherung für die exakte Schrittlänge erhalten; man vergleiche hierzu das Beispiel (3.7). In dieser Situation macht es nun keinen Sinn, die Verwendung dieser Schrittlänge durch die Goldstein-Bedingung zu verhindern. (Auch die zweite Goldstein-Bedingung (5.9) kann die exakte Schrittlänge ausschließen, allerdings ist dies bei hinreichend kleinem Parameter ρ eher unwahrscheinlich.)

Um diesen Nachteil zu vermeiden, sind Bedingungen vorgeschlagen worden, die die erste Goldstein-Bedingung ersetzen sollen. Gebräuchlich sind die sogenannte **Wolfe-Powell-Bedingung**

$$\phi'(\alpha) \;\geq\; \sigma\,\phi'(0) \tag{5.13}$$

(σ fest vorgegeben mit $\rho < \sigma < 1$), sowie die sogenannte **Wolfe-Powell-Fletcher-Bedingung**

$$|\phi'(\alpha)| \;\leq\; -\,\sigma\,\phi'(0)\,. \tag{5.14}$$

Beide Bedingungen werden von der exakten Schrittlänge α^* wegen $\phi'(\alpha^*) = 0$ erfüllt und - wegen $\sigma < 1$ - vermeiden auch beide Bedingungen (wie (5.12)), daß die Schrittlänge α zu klein gewählt wird. Die Wolfe-Powell-Fletcher-Bedingung schließt darüberhinaus auch Schrittlängen aus, für die $\phi(\alpha)$ zu stark wächst.

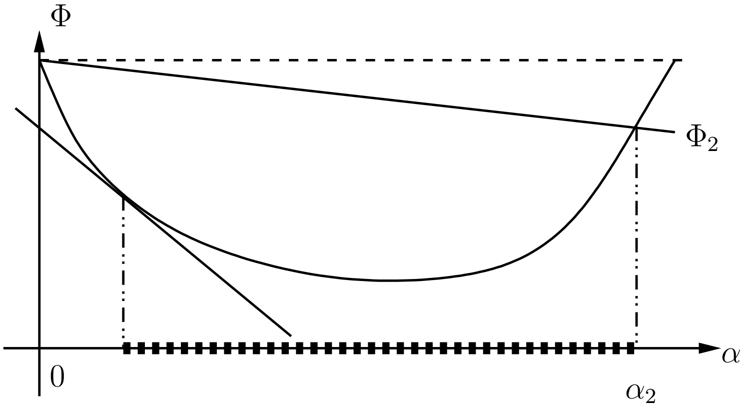

Abb. 5.2. Wolfe-Powell Schrittlänge

Satz (5.15)

Unter den eingangs genannten Voraussetzungen gibt es stets Schrittlängen, die den Wolfe-Powell Bedingungen (5.9) und (5.13) genügen.

Jede dieser Schrittlängen erfüllt die Effizienzbedingung

$$f(\boldsymbol{x} + \alpha\, \boldsymbol{s}) \;\leq\; f(\boldsymbol{x}) \;-\; \frac{\rho\,(1-\sigma)}{C} \left(\frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\, \boldsymbol{s}}{\|\boldsymbol{s}\|} \right)^2 ,$$

wobei wie in Satz (5.8) $C := \max\{\|\boldsymbol{\nabla}^2 f(\boldsymbol{x})\| \,:\, \boldsymbol{x} \in L(\boldsymbol{x}^0)\}$ *ist.*

Beweis: Es sei $0 < \alpha_2 < \alpha_1$ die kleinste Schrittlänge mit $\phi(\alpha_2) = \phi_2(\alpha_2)$ (siehe Abb. 5.2). Damit gilt

$$\forall\, \alpha \;:\; 0 < \alpha < \alpha_2 \;\Longrightarrow\; \phi(\alpha) < \phi_2(\alpha) = \phi(0) + \alpha\,\rho\,\phi'(0)\,.$$

Die Anwendung des Mittelwertsatzes zeigt: Es gibt ein α, $0 < \alpha < \alpha_2$, mit

$$\phi'(\alpha) \;=\; \frac{\phi(\alpha_2) - \phi(0)}{\alpha_2 - 0} \;=\; \rho\,\phi'(0) \;\geq\; \sigma\,\phi'(0)\,.$$

Diese Schrittlänge α erfüllt somit (5.9) und (5.13).

Sei wie im Beweis zu Satz (5.8) α_0 das kleinste positive lokale Minimum der Funktion $\phi(\alpha)$. Für $\alpha \leq \alpha_0$ schließt man mit (5.13)

$$\phi'(\alpha) \;\geq\; \sigma\,\phi'(0) \;=\; \phi'(0) \;-\; (1-\sigma)\,\phi'(0)$$

Hieraus folgt (wiederum mit dem Mittelwertsatz)

$$(1-\sigma)\,\phi'(0) \;\geq\; \phi'(0) - \phi'(\alpha) \;=\; -\,\alpha\,\phi''(\tilde{\alpha}) \;\geq\; -\,\alpha\, C\,\|\boldsymbol{s}\|^2\,, \quad 0 < \tilde{\alpha} < \alpha\;,$$

also $\alpha \;\geq\; -\dfrac{(1-\sigma)\,\phi'(0)}{C\,\|\boldsymbol{s}\|^2}$. Mit (5.9) erhält man schließlich

$$\phi(\alpha) \;\leq\; \phi_2(\alpha) \;=\; \phi(0) \;+\; \rho\,\alpha\,\phi'(0) \;\leq\; \phi(0) \;-\; \frac{\rho\,(1-\sigma)}{C}\left(\frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}}{\|\boldsymbol{s}\|}\right)^2 .$$

Im zweiten Fall, $\alpha > \alpha_0$, folgt wie im Beweis zu Satz (5.8)

$$\phi(\alpha) \;\leq\; \phi(0) - \frac{\rho}{C}\left(\frac{\phi'(0)}{\|\boldsymbol{s}\|}\right)^2 \;\leq\; \phi(0) - \frac{\rho\,(1-\sigma)}{C}\left(\frac{\phi'(0)}{\|\boldsymbol{s}\|}\right)^2 .$$

□

Zur numerischen Berechnung einer Schrittlänge, die die Wolfe-Powell Bedingungen (5.9) und (5.13) erfüllt, kann man einen Algorithmus analog zu (5.6) formulieren. Dieser arbeitet ebenfalls mit einer Expansions- und einer Kontraktionsphase:

Algorithmus (5.16) (Wolfe-Powell-Schrittlänge)

(i) Man wähle $\alpha > 0$ („geeignet“).

(ii) Gilt $\phi(\alpha) \geq \phi_2(\alpha)$, so setze man $a := 0$, $b := \alpha$ und gehe zu (iii).

Gilt $\phi'(\alpha) \geq \sigma\,\phi'(0)$, so gehe man zu (iv).

Man setze $\alpha := 2\,\alpha$ und wiederhole (ii).

(iii) Man setze $\alpha := (a+b)/2$.

Gilt $\phi(\alpha) \geq \phi_2(\alpha)$, so setze man $b := \alpha$ und wiederhole (iii) von Anfang.

Gilt $\phi'(\alpha) \geq \sigma\phi'(0)$, so gehe man zu (iv).

Man setze $a := \alpha$ und wiederhole (iii).

(iv) Die Schrittlänge α erfüllt (5.9) und (5.13).

Bemerkungen (5.17)

a) Wie für den Algorithmus (5.6) läßt sich zeigen, daß die Schleifen in den Phasen (ii) und (iii) jeweils nur endlich oft durchlaufen werden und der Algorithmus daher nach endlich vielen Schritten eine zulässige Schrittweite liefert.

b) Wiederum lassen sich bei der Expansions- wie bei der Kontraktionsphase zusätzliche Freiheiten dadurch gewinnen, daß man in Schritt (ii) anstelle von $\alpha := 2\,\alpha$ setzt $\alpha := \lambda\,\alpha$ mit einem $\lambda \geq \lambda_{\min} > 1$ und anstelle der Halbierungsstrategie

$$\alpha \;:=\; a \;+\; \mu\,(b-a)\,, \quad \mu \in [\mu_{\min}, \mu_{\max}],$$

verwendet, wobei $0 < \mu_{\min} < \mu_{\max} < 1$ fest vorgegeben sind.

c) Die durch die obigen Modifikationen gewonnenen Freiheiten lassen sich nutzen, um mittels Interpolationstechniken geeignete Testwerte für λ bzw. μ zu erhalten.

So läßt sich für die Kontraktionsphase beispielsweise aus den bekannten Daten $\phi(a)$, $\phi'(a)$, $\phi(b)$, $\phi'(b)$ ein kubisches Hermite-Interpolationspolynom konstruieren, dessen Minimum (quadratische Gleichung!) als Testwert für α dienen kann. Durch derartige Interpolationstechniken lassen sich Algorithmen zur Schrittlängenbestimmung mitunter wesentlich beschleunigen.

d) Es gibt eine Variante des obigen Algorithmus mit der sich eine Schrittlänge bestimmen läßt, die auch der strengeren Wolfe-Powell-Fletcher-Bedingung (5.14) genügt.