3 Abstiegsverfahren

Wir beschäftigen uns in diesem Abschnitt mit einer Verfahrensklasse für glatte, unrestringierte Optimierungsaufgaben der Form:

$$\text{Minimiere} \qquad f(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n \,. \tag{3.1}$$

Hierbei sei $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ also hinreichend oft stetig differenzierbar.

Sei $\boldsymbol{x} \in \mathbb{R}^n$ ein Näherungswert für ein gesuchtes (lokales) Minimum der Funktion f (aktuelle Iterierte). Wir nehmen an, daß $\boldsymbol{x}$ nicht schon ein stationärer Punkt von f ist, also $\boldsymbol{\nabla} f(\boldsymbol{x}) \neq \mathbf{0}$. Dann gibt es eine Richtung $\boldsymbol{s} \in \mathbb{R}^n$, $\boldsymbol{s} \neq \mathbf{0}$, mit der Eigenschaft $\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{s} \neq \mathbf{0}$. Wir untersuchen nun das Verhalten von f auf dem Halbstrahl, der bei $\boldsymbol{x}$ beginnt und in Richtung von $\boldsymbol{s}$ verläuft. Dazu betrachten wir folgende Funktion einer Variablen

$$\phi(\alpha) \; := \; f(\boldsymbol{x} + \alpha\, \boldsymbol{s}), \;\; \alpha \geq 0 \,. \tag{3.2}$$

Ist $\boldsymbol{s}$ ein Einheitsvektor (Länge 1), so gibt α die **Schrittlänge** an, um die wir uns von $\boldsymbol{x}$ entfernen. Die Taylor-Entwicklung von $\phi(\alpha)$ um $\alpha = 0$ liefert

$$\phi(\alpha) \; := \; f(\boldsymbol{x}) \; + \; \alpha \, (\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{s}) \; + \; o(\alpha) \,. \tag{3.3}$$

Man nennt $\boldsymbol{s}$ nun eine **Abstiegsrichtung** für f im Punkte $\boldsymbol{x}$, falls $\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{s} < 0$ gilt.

An der Gleichung (3.3) liest man unmittelbar ab: Ist $\boldsymbol{s}$ eine Abstiegsrichtung, so gilt $\phi(\alpha) < f(\boldsymbol{x})$ für alle hinreichend kleinen $\alpha > 0$, genauer:

$$\exists\, \alpha_0 > 0 \;:\; \forall \alpha \;:\; 0 < \alpha \leq \alpha_0 \;\; \Longrightarrow \;\; f(\boldsymbol{x} + \alpha\, \boldsymbol{s}) \; < \; f(\boldsymbol{x}) \,. \tag{3.4}$$

Abstiegsverfahren sind numerische Minimierungsalgorithmen, die diese Eigenschaft ausnützen. Sie laufen generell nach dem folgenden Modellalgorithmus ab:

Modellalgorithmus (3.5)

(i) $\boldsymbol{x}_0 \in \mathbb{R}^n$: Startvektor

(ii) Für $k = 0, 1, \ldots$:
  1.) Test auf Abbruch (etwa, falls $\|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\|$ hinreichend klein ist)
  2.) Bestimmung einer Abstiegsrichtung $\boldsymbol{s}_k$ mit $\boldsymbol{\nabla} f(\boldsymbol{x}_k)^{\mathrm{T}} \boldsymbol{s}_k < 0$
  3.) Bestimmung einer Schrittlänge $\alpha_k > 0$ mit $f(\boldsymbol{x}_k + \alpha_k\, \boldsymbol{s}_k) \; < \; f(\boldsymbol{x}_k)$
  4.) Neue Näherung $\boldsymbol{x}_{k+1} \; := \; \boldsymbol{x}_k + \alpha_k\, \boldsymbol{s}_k$.

Für ein konkretes Abstiegsverfahren ist also (neben einer Abbruchstrategie) ein Verfahren zur Festlegung einer Abstiegsrichtung und eine sogenannte Schrittlängenstrategie anzugeben. Hier ergibt sich natürlich die Frage, wie man denn solche Strategien wählen sollte, und damit im Zusammenhang die Frage, für welche Strategien die Konvergenz eines Abstiegsverfahrens gesichert ist und schließlich die Frage nach der Konvergenzgeschwindigkeit des Verfahrens.

Bevor wir hierzu allgemeinere Aussagen treffen können, wollen wir uns zunächst einige konkrete Realisierungen ansehen.

Eine (zumindest theoretische) sinnvolle Wahl für die Schrittlänge ergibt sich, wenn man $\phi(\alpha)$ auf der Halbgeraden $\alpha \geq 0$ minimiert:

$$\alpha_{\text{exakt}} \;:=\; \arg\min\{ f(\boldsymbol{x} + \alpha \boldsymbol{s}) \;:\; \alpha > 0 \} \,. \tag{3.6}$$

Ist die Niveaumenge $L(\boldsymbol{x})$ kompakt, so besitzt (3.6) eine (nicht notwendig eindeutige) Lösung. Man spricht hierbei auch von der **exakten** (oder optimalen) **Schrittlänge**. Zur numerischen Realisierung von (3.6) hätte man im allgemeinen eine eindimensionale Minimierung durchzuführen. Der Aufwand für diese Schrittlängenstrategie ist daher, wenn nicht spezielle Situationen vorliegen (wie etwa im folgenden Beispiel), nicht unerheblich.

Beispiel (3.7)

Für eine **quadratische Zielfunktion** $f(\boldsymbol{x}) = \frac{1}{2}\, \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x} + c$ mit einer symmetrischen und positiv definiten Matrix $\boldsymbol{A} \in \mathbb{R}^{(n,n)}$ läßt sich die exakte Schrittlänge explizit angeben.

Wegen $\boldsymbol{\nabla} f(\boldsymbol{x}) = \boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{b}$ und

$$\phi'(\alpha) \;=\; \boldsymbol{\nabla} f(\boldsymbol{x} + \alpha \boldsymbol{s})^{\mathrm{T}}\, \boldsymbol{s} \;=\; \boldsymbol{s}^{\mathrm{T}} \boldsymbol{A}(\boldsymbol{x} + \alpha \boldsymbol{s}) - \boldsymbol{s}^{\mathrm{T}} \boldsymbol{b}$$

erhält man die (eindeutig bestimmte) exakte Schrittlänge

$$\alpha_{\text{exakt}} \;:=\; -\, \frac{\boldsymbol{s}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})}{\boldsymbol{s}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{s}} \,. \tag{3.8}$$

In dieser Gleichung ist der Nenner positiv (da $\boldsymbol{A}$ positiv definit ist) und der Zähler negativ (da $\boldsymbol{s}$ Abstiegsrichtung ist). Man erhält also wie gewünscht einen positiven Wert für α_{exakt}.

Ebenfalls für die Strategie zur Bestimmung einer Abstiegsrichtung ergibt sich ein naheliegendes kanonisches Verfahren, nämlich die Wahl

$$\boldsymbol{s}_{\mathrm{G}} \;:=\; -\boldsymbol{\nabla} f(\boldsymbol{x}) \,. \tag{3.9}$$

Es ist offensichtlich, daß (3.9) tatsächlich eine Abstiegsrichtung liefert, denn

$$\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{s}_{\mathrm{G}} \;=\; -\, \|\boldsymbol{\nabla} f(\boldsymbol{x})\|^2 < 0 \,.$$

Man nennt $\boldsymbol{s}_{\mathrm{G}}$ auch die **Richtung des steilsten Abstiegs** (steepest descent method) und die Verfahren, die hiermit arbeiten, auch **Gradientenverfahren**. Der Grund für diese Bezeichnung wird im folgenden Satz genannt.

Satz (3.10)

Die (normierte) Richtung des steilsten Abstiegs $\boldsymbol{s}^* = -\nabla f(\boldsymbol{x})/\|\nabla f(\boldsymbol{x})\|$ *minimiert den (linearisierten) Anstieg* $\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{s}$ *über allen Einheitsvektoren* $\boldsymbol{s} \in \mathbb{R}^n$, $\|\boldsymbol{s}\| = 1$.

Beweis: Die Cauchy–Schwarzsche Ungleichung (vgl. zum Beisp. Ansorge, Oberle [...]) liefert für beliebige Einheitsvektoren $\boldsymbol{s}$

$$|\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{s}| \;\leq\; \|\nabla f(\boldsymbol{x})\|\,\|\boldsymbol{s}\| \;=\; \|\nabla f(\boldsymbol{x})\|.$$

Diese Schranke wird speziell für $\boldsymbol{s} = \boldsymbol{s}^*$ angenommen: $\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{s}^* = -\|\nabla f(\boldsymbol{x})\|$.
D.h. $\boldsymbol{s}^*$ gibt tatsächlich die Richtung des (lokal) steilsten Abstiegs an. □

Beispiel (3.11) (nach Luenberger [...])

Es sei $n = 4$ und f die quadratische Zielfunktion $f(\boldsymbol{x}) \;=\; \frac{1}{2}\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \;-\; \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$ mit

$$\boldsymbol{A} \;=\; \begin{pmatrix} 0.78 & -0.02 & -0.12 & -0.14 \\ -0.02 & 0.86 & -0.04 & 0.06 \\ -0.12 & -0.04 & 0.72 & -0.08 \\ -0.14 & 0.06 & -0.08 & 0.74 \end{pmatrix}$$

und $\boldsymbol{b} = (0.76,\, 0.08,\, 1.12,\, 0.68)^{\mathrm{T}}$. Die Voraussetzung zur Anwendung des Verfahrens des steilsten Abstiegs ist hier erfüllt: $\boldsymbol{A}$ ist symmetrisch und diagonaldominant mit positiven Diagonalelementen und daher auch positiv definit. Bei Wahl der exakten Schrittlänge (3.8) erhält man die Rekursion ($k = 0, 1, 2, \ldots$)

$$\boldsymbol{x}_{k+1} \;=\; \boldsymbol{x}_k \;-\; \frac{\boldsymbol{g}_k^{\mathrm{T}}\,\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\,\boldsymbol{A}\,\boldsymbol{g}_k}\;\boldsymbol{g}_k$$

wobei $\boldsymbol{g}_k := \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}$ den Gradienten von f bezeichnet. Dies ist zugleich das Residuum des zugehörigen linearen Gleichungssystems.

In der folgenden Tabelle sind die Ergebnisse des Verfahrens für den Startvektor $\boldsymbol{x}_0 = \boldsymbol{0}$ und dem Abbruchkriterium $\|\boldsymbol{g}_k\| < 10^{-5}$ angegeben. $\mathrm{Res}(\boldsymbol{x}_k)$ bezeichnet dabei die Norm des Residuenvektors $\boldsymbol{g}_k$.

Man erhält hiernach numerische Konvergenz nach bereits acht Iterationen. Für das gesuchte Minimum ergibt sich bei etwa fünfstelliger Genauigkeit

$$\boldsymbol{x}_8 = (1.53496,\, 0.12201,\, 1.97515,\, 1.41295)^{\mathrm{T}} .$$

Tabelle 3.1. : Ergebnisse zu Beispiel (3.11)

k	f(x_k)	Res(x_k)
0	0	1.516838818068683e+000
1	-2.156362688357058e+000	1.693217346054398e-001
2	-2.174406506409742e+000	1.803089812780436e-002
3	-2.174644122895887e+000	4.522365290280037e-003
4	-2.174658547707589e+000	1.137438467649929e-003
5	-2.174659484894549e+000	2.959845031455454e-004
6	-2.174659546618460e+000	7.495760100757472e-005
7	-2.174659550688035e+000	1.951658824016860e-005
8	-2.174659550956394e+000	4.942963507688019e-006

Daß das Verfahren des steilsten Abstiegs selbst bei exakter Schrittlängenbestimmung jedoch nicht immer so zufriedenstellend verläuft, zeigt das folgende Beispiel.

Beispiel (3.12)

Es sei $n = 2$ und $f(x_1, x_2) = x_1^2 + 10\,x_2^2$. f ist also eine quadratische Funktion mit

$$\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 0 & 20 \end{pmatrix}$$

und $\boldsymbol{b} = \boldsymbol{0}$, $c = 0$ und dem eindeutig bestimmten Minimum $\boldsymbol{x}^* = \boldsymbol{0}$. Das Verfahren des steilsten Abstiegs mit exakter Schrittlänge lautet dann (vgl. den Modellalgorithmus (3.5) und das Beispiel (3.7)):

(i) Startvektor: $x_{1,0} = 1\,,\ x_{2,0} = 0.1$

(ii) Für $k = 0, 1, \dots$:

1.) Abbruch: Falls $\sqrt{4\,x_{1,k}^2 + 400\,x_{2,k}^2} \leq 10^{-5}$: STOP

2.) $s_1 := -\,2\,x_{1,k}\,,\ \ s_2 := -\,20\,x_{2,k}$

3.) $\alpha := -\,\dfrac{s_1\,x_{1,k} + 10\,s_2\,x_{2,k}}{s_1^2 + 10\,s_2^2}$

4.) $x_{j,k+1} := x_{j,k} + \alpha\,s_{j,k}\,,\ \ j = 1, 2$.

Für den Startvektor $\boldsymbol{x}_0 = (1, 0.1)^{\mathrm{T}}$ benötigt das Verfahren etwa 60 Iterationen um $\|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\| < 10^{-5}$ zu erreichen. Die Wahl des negativen Gradienten als Abstiegsrichtung ist in diesem Beispiel offenbar wenig geeignet. Man erkennt dies an den Zickzackweg der Iterierten.

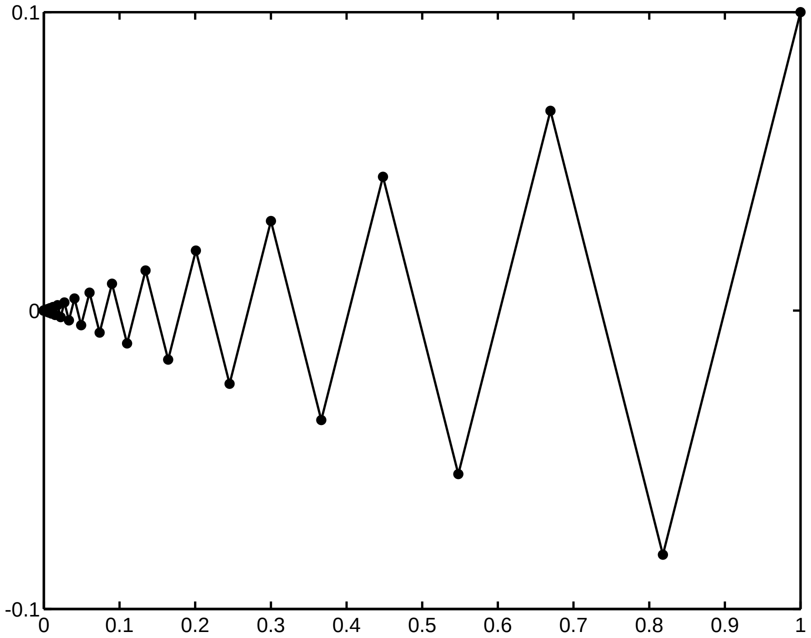

Abb. 3.1. Zickzackweg für Beispiel (3.12)

Den Grund für das unterschiedliche Verhalten des Verfahrens in den obigen beiden Beispielen liefert der folgende Satz über die Konvergenz und Konvergenzgeschwindigkeit des Verfahrens des steilsten Abstiegs mit exakter Schrittlänge angewendet auf quadratische Zielfunktionen.

Satz (3.13)

Sei $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *eine quadratische Zielfunktion*

$$f(\boldsymbol{x}) \;=\; \frac{1}{2}\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \;-\; \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} \;+\; c$$

mit einer symmetrischen und positiv definiten Matrix $\boldsymbol{A}$ *und dem eindeutig bestimmten Minimum* $\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$. *Für das Verfahren des steilsten Abstiegs mit exakter Schrittlänge gilt dann*

$$f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*) \;\leq\; \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \;(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*))\;. \tag{3.14}$$

Hierbei bezeichnen $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ *die (sämtlich positiven) Eigenwerte von* $\boldsymbol{A}$.

Bemerkungen (3.15)

- Da der Faktor in Gleichung (3.14) kleiner eins ist, ist die Konvergenz des Verfahrens für alle quadratischen Zielfunktionen mit positiv definiter Matrix $\boldsymbol{A}$ gesichert.

– Mit der sogenannten **Konditionszahl** $\kappa(\boldsymbol{A}) := \lambda_n/\lambda_1 \geq 1$ läßt sich die Abschätzung (3.14) auch folgendermaßen schreiben

$$f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*) \;\leq\; \left(1 \;-\; \frac{2}{\kappa(\boldsymbol{A})+1}\right)^2 \;(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*))\;. \tag{3.16}$$

Je größer also die Kondition der Matrix $\boldsymbol{A}$ ist, desto langsamer wird das Verfahren des steilsten Abstieges konvergieren.

– Für das Beispiel (3.11) berechnet man (beispielsweise mit der MATLAB Routine **eig** oder **cond**) $\lambda_{\text{min}} \approx 0.52$, $\lambda_{\text{max}} \approx 0.94$. Damit ist $\kappa \approx 1.8$ und für den konvergenzerzeugenden Faktor in (3.14) ergibt sich

$$\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \;\approx\; 0.081\,.$$

Die erklärt die schnelle Konvergenz für dieses Beispiel.

Für das Beispiel (3.12) ergibt sich $\kappa(\boldsymbol{A}) = 10$ und man erhält den konvergenzerzeugenden Faktor ≈ 0.67.

Beweis zu (3.13):

Zunächst einige Vorbemerkungen über Normen:

Wir betrachten neben der Euklidischen Vektornorm $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$ noch die Vektornorm $\|\boldsymbol{x}\|_{\boldsymbol{A}} := \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}$. Da $\boldsymbol{A}$ positiv definit ist, ist durch $\|\,.\,\|_{\boldsymbol{A}}$ tatsächlich eine Norm für $\mathbb{R}^n$ gegeben. Man zeigt hierzu einfach, daß $<\boldsymbol{x},\boldsymbol{y}>_{\boldsymbol{A}} := \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$ ein Skalarprodukt ist. Die zugehörige Matrixnorm ist dann gegeben durch

$$\|\boldsymbol{Q}\|_{\boldsymbol{A}} \;:=\; \max\{\,\|\boldsymbol{Q}\boldsymbol{x}\|_{\boldsymbol{A}} \;:\; \|\boldsymbol{x}\|_{\boldsymbol{A}} \;=\; 1\,\}$$

und hierfür gilt (wie für alle zu einer Vektornorm zugehörigen Matrixnormen) die Abschätzung der Submultiplikativität $\|\boldsymbol{Q}\boldsymbol{x}\|_{\boldsymbol{A}} \leq \|\boldsymbol{Q}\|_{\boldsymbol{A}}\,\|\boldsymbol{x}\|_{\boldsymbol{A}}$.

Zur Berechnung von $\|\boldsymbol{A}\|_{\boldsymbol{A}}$ verwenden wir, daß es zu der symmetrischen Matrix $\boldsymbol{A}$ eine orthonormale Basis $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$ des $\mathbb{R}^n$ aus Eigenvektoren gibt. Es gelte also $\boldsymbol{A}\,\boldsymbol{x}_j \;=\; \lambda_j\,\boldsymbol{x}_j\;,\; j=1,\ldots,n$ und $\boldsymbol{x}_j^{\mathrm{T}}\,\boldsymbol{x}_i \;=\; \delta_{i\,j}$ (Kronecker-Symbol). Setzt man nun $\boldsymbol{x} = \sum_{j=1}^n \alpha_j \boldsymbol{x}_j$ in die Definition von $\|\boldsymbol{A}\|_{\boldsymbol{A}}$ ein, so folgt

$$\begin{aligned}
\|\boldsymbol{A}\|_{\boldsymbol{A}} \;&=\; \max\{\,\|\boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{A}} \;:\; \|\boldsymbol{x}\|_{\boldsymbol{A}} \;=\; 1\,\}\\
&=\; \max\{\,\sqrt{\boldsymbol{x}^{\mathrm{T}}\,\boldsymbol{A}^3\,\boldsymbol{x}} \;:\; \sqrt{\boldsymbol{x}^{\mathrm{T}}\,\boldsymbol{A}\,\boldsymbol{x}} \;=\; 1\,\}\\
&=\; \max\{\,\sqrt{\sum_j \lambda_j^3\,\alpha_j^2} \;:\; \sqrt{\sum_j \lambda_j\,\alpha_j^2} \;=\; 1\,\}\\
&=\; \lambda_{\max} \;=\; \lambda_n\,.
\end{aligned}$$

Mit der gleichen Beweistechnik zeigt man, daß für ein beliebiges (skalares) Polynom $p(\lambda)$ gilt

$$\|p(\boldsymbol{A})\|_{\boldsymbol{A}} \;=\; \max\{\, |p(\lambda)| \;:\; \lambda = \lambda_1, \ldots, \lambda_n \,\} \;. \tag{3.17}$$

Nun zum eigentlichen Beweis:

Unter Verwendung der Lösung $\boldsymbol{x}^*$ des Minimierungsproblems und des Fehlers $\boldsymbol{e}(\boldsymbol{x}) := \boldsymbol{x} - \boldsymbol{x}^*$ läßt sich f folgendermaßen darstellen

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{2}\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \;-\; \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} \;+\; c \\ &= \frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^*) \;-\; \frac{1}{2}\,\boldsymbol{x}^{*\,\mathrm{T}}\boldsymbol{A}\boldsymbol{x}^* \;+\; c \\ &= \frac{1}{2}\,\|\boldsymbol{e}(\boldsymbol{x})\|_{\boldsymbol{A}}^2 \;+\; f(\boldsymbol{x}^*)\;. \end{aligned} \tag{3.18}$$

Nun minimiert die exakte Schrittlänge α_k im k-ten Schritt die Funktion $f(\boldsymbol{x}_k + \alpha\,\boldsymbol{s}_k)$ und damit wegen (3.18) auch $\|\boldsymbol{e}(\boldsymbol{x}_k + \alpha\,\boldsymbol{s}_k)\|_{\boldsymbol{A}}$. Es gilt also

$$\|\boldsymbol{e}(\boldsymbol{x}_{k+1})\|_{\boldsymbol{A}} \;=\; \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_{\boldsymbol{A}} \;\le\; \|\boldsymbol{x}_k + \alpha\,\boldsymbol{s}_k - \boldsymbol{x}^*\|_{\boldsymbol{A}}$$

für alle $\alpha \ge 0$. Nun ist

$$\boldsymbol{x}_k + \alpha\,\boldsymbol{s}_k - \boldsymbol{x}^* \;=\; \boldsymbol{e}(\boldsymbol{x}_k) - \alpha\,\boldsymbol{A}\,\boldsymbol{e}(\boldsymbol{x}_k) \;=\; (\boldsymbol{I}_n - \alpha\,\boldsymbol{A})\,\boldsymbol{e}(\boldsymbol{x}_k)$$

Wegen der Submultiplikativität der Matrixnorm folgt hieraus für alle $\alpha \ge 0$:

$$\|\boldsymbol{e}(\boldsymbol{x}_{k+1})\|_{\boldsymbol{A}} \;\le\; \|\boldsymbol{I}_n - \alpha\,\boldsymbol{A}\|_{\boldsymbol{A}}\,\|\boldsymbol{e}(\boldsymbol{x}_k)\|_{\boldsymbol{A}}\;. \tag{3.19}$$

Mit der Beziehung (3.17) läßt sich der Faktor explizit berechnen:

$$\begin{aligned} \|\boldsymbol{I}_n - \alpha\,\boldsymbol{A}\|_{\boldsymbol{A}} &= \max\{\, |1-\alpha\,\lambda| \;:\; \lambda = \lambda_1, \ldots, \lambda_n \,\} \\ &= \begin{cases} |1-\alpha\,\lambda_n|, & \text{falls } \dfrac{1}{\alpha} < \dfrac{\lambda_1+\lambda_n}{2} \\ |1-\alpha\,\lambda_1|, & \text{falls } \dfrac{1}{\alpha} \ge \dfrac{\lambda_1+\lambda_n}{2} \\ 1, & \text{falls } \alpha = 0\;. \end{cases} \end{aligned}$$

Die kleinste obere Schranke erhält man für $\alpha = \alpha^*$ mit $1/\alpha^* = (\lambda_1+\lambda_n)/2$. Dies eingesetzt ergibt

$$\|\boldsymbol{I}_n - \alpha^*\,\boldsymbol{A}\|_{\boldsymbol{A}} \;=\; \left|\, 1 \;-\; \frac{2}{\lambda_1+\lambda_n}\,\lambda_n \,\right| \;=\; \frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\;.$$

Hiermit liefert (3.19) schließlich die Abschätzung

$$\|\boldsymbol{e}(\boldsymbol{x}_{k+1})\|_{\boldsymbol{A}} \;\le\; \left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)\;\|\boldsymbol{e}(\boldsymbol{x}_k)\|_{\boldsymbol{A}}\;, \tag{3.20}$$

woraus vermöge (3.18) die Behauptung folgt. □

Bemerkungen (3.21)

- Mit (3.20) haben wir zugleich eine Abschätzung für den absoluten Fehler gefunden haben, allerding wird dieser in der $\|\,.\,\|_{\boldsymbol{A}}$ – Norm gemessen.
- Eine ähnliche Aussage wie in Satz (3.13) kann auch für nichtquadratische Zielfunktionen gezeigt werden (vgl. Luenberger [...]). Hierbei sind λ_1 bzw. λ_n kleinster bzw. größter Eigenwert der (als positiv definit vorausgesetzten) Hesse-Matrix $\boldsymbol{\nabla}^2 f(\boldsymbol{x}^*)$ im Lösungspunkt.

Zur Beschleunigung der Methode des steilsten Abstiegs könnte man daran denken, anstelle von $f(\boldsymbol{x})$ eine geeignet umskalierte Funktion zu minimieren. Ziel einer Umskalierung wäre es, die Eigenwerte der transformierten Hesse-Matrix günstig zu beeinflussen (im Sinn von $\lambda_n - \lambda_1$ „klein“).

Ein ähnliches Verfahren besteht darin, die Richtung des steilsten Abstiegs bezüglich einer skalierten Norm – etwa $\|\,.\,\|_{\boldsymbol{C}}$ – zu wählen. Hierbei ist $\boldsymbol{C}$ eine (geeignet vorzugebende) symmetrische und positiv definite Matrix. Es gelten dann zu den Sätzen (3.10) und (3.13) analoge Eigenschaften:

Satz (3.22)

- *Die (normierte) Richtung* $\boldsymbol{s}^* = -\boldsymbol{C}^{-1}\,\boldsymbol{\nabla} f(\boldsymbol{x})/\|\boldsymbol{C}^{-1}\,\boldsymbol{\nabla} f(\boldsymbol{x})\|_{\boldsymbol{C}}$ *minimiert den (linearisierten) Anstieg* $\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{s}$ *über allen Einheitsvektoren* $\boldsymbol{s} \in \mathbb{R}^n$, $\|\boldsymbol{s}\|_{\boldsymbol{C}} = 1$.
- *Wendet man das Abstiegsverfahren mit exakter Schrittlänge und obiger Absteigsrichtung auf eine quadratische Zielfunktion mit symmetrischer und positiv definiter Koeffizientenmatrix* $\boldsymbol{A}$ *an, so gilt die Fehlerabschätzung (3.14), wobei* λ_1 *bzw.* λ_n *kleinster bzw. größter Eigenwert von* $\boldsymbol{C}^{-1}\,\boldsymbol{A}$ *ist.*

Beispiel (3.23)

Wählt man in Beispiel (3.12) als Abstiegsrichtung $\boldsymbol{s} = -\boldsymbol{C}^{-1}\,\boldsymbol{\nabla} f(\boldsymbol{x})$ mit

$$\boldsymbol{C} \;=\; \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix}$$

so erhält man $\boldsymbol{s} = -\boldsymbol{C}^{-1}\,(2x_1, 20x_2)^{\mathrm{T}} = -2\,\boldsymbol{x}$. Bereits der erste Iterationsschritt des Verfahrens führt daher (unabhängig vom Startvektor) ins Minimum $\boldsymbol{x}^* = \boldsymbol{0}$.

Wir wenden uns nun wieder den allgemeinen Abstiegsverfahren zu und fragen nach Bedingungen an die Wahl der Abstiegsrichtung und an die Schrittlängenstrategie, unter denen die Konvergenz des Verfahrens gesichert werden kann.

Für die Wahl der Abstiegsrichtung ist eine Bedingung offensichtlich: die Abstiegsrichtung sollte (zumindest für kleine Schrittlängen) ins Innere der Niveaumenge führen.

Oder, geometrisch beschrieben: Der Winkel γ zwischen Abstiegsrichtung und negativem Gradienten soll von $\pi/2$ weg beschränkt sein. Wegen

$$\cos\gamma \;=\; -\,\frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}}{\|\boldsymbol{\nabla} f(\boldsymbol{x})\|\;\|\boldsymbol{s}\|}$$

führt dies auf die folgende Definition.

Definition (3.24)

Eine Strategie, die jedem Punkt $\boldsymbol{x}$ mit $\boldsymbol{\nabla} f(\boldsymbol{x})$ eine Abstiegsrichtung $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{x})$ zuordnet, erfüllt die **Winkelbedingung**, falls es eine (von $\boldsymbol{x}$ unabhängige) Konstante $\nu > 0$ gibt mit

$$-\,\frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}}{\|\boldsymbol{\nabla} f(\boldsymbol{x})\|\;\|\boldsymbol{s}\|} \;\geq\; \nu \qquad \text{für alle } \boldsymbol{x} \text{ mit } \boldsymbol{\nabla} f(\boldsymbol{x}) \neq \boldsymbol{0}\,.$$

Man prüft leicht nach, daß die Richtung des steilsten Abstiegs - auch in der skalierten Variante (3.22) - die Winkelbedingung erfüllt.

Von einer sinnvollen Schrittlängenstrategie verlangen wir, daß diese in jedem Iterationsschritt den Funktionswert „ausreichend“ verkleinert. Was hierbei „ausreichend“ heißt, kann man sich am „quadratischen Modell“ klarmachen. Wir nehmen an, daß das Abstiegsverfahren in einem Punkt $\boldsymbol{x}_0$ mit kompakter Niveaumenge $L(\boldsymbol{x}_0)$ gestartet wird. Dann ist die Hesse-Matrix von f auf $L(\boldsymbol{x}_0)$ beschränkt, $\|\boldsymbol{\nabla}^2 f(\boldsymbol{x})\| \leq \gamma$. Der Taylorsche Satz liefert die Abschätzung

$$f(\boldsymbol{x} + \alpha\,\boldsymbol{s}) \;\leq\; f(\boldsymbol{x}) \;+\; \alpha\,\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s} \;+\; \frac{\alpha^2}{2}\,\gamma\,\|\boldsymbol{s}\|^2 \;\;=:\;\; p(\alpha)\,. \tag{3.25}$$

Diese Ungleichung gilt für alle $\boldsymbol{x} \in L(\boldsymbol{x}_0)$, $\boldsymbol{s}$ Abstiegsrichtung von f in $\boldsymbol{x}$ und für alle Schrittlängen $\alpha > 0$ mit $\forall \beta \in [0, \alpha] \,:\, \boldsymbol{x} + \beta\boldsymbol{s} \in L(\boldsymbol{x}_0)$.

Die rechte Seite von (3.25) ist, als Funktion von α, eine nach oben geöffnete Parabel $p(\alpha)$ mit $p'(0) < 0$ und Minimum (Scheitel) in

$$\alpha^* \;=\; \frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}}{\gamma\,\|\boldsymbol{s}\|^2}\,.$$

Wegen der vorausgesetzten Kompaktheit von $L(\boldsymbol{x}_0)$ ist α^* eine zulässige Schrittweite, es gilt sogar $\forall \beta \in [0, \alpha^*] \,:\, \boldsymbol{x} + \beta\boldsymbol{s} \in L(\boldsymbol{x})$. (3.25) gilt also auch für α^* und man erhält

$$f(\boldsymbol{x} + \alpha^*\,\boldsymbol{s}) \;\leq\; p(\alpha^*) \;=\; f(\boldsymbol{x}) \;-\; \frac{1}{2\,\gamma}\left(\frac{\boldsymbol{\nabla} f(\boldsymbol{x})^{\mathrm{T}}\,\boldsymbol{s}}{\|\boldsymbol{s}\|}\right)^2 .$$

Diese Abschätzung legt nun die folgende Definition nahe.

Definition (3.26)

Eine Strategie, die jedem Punkt $\boldsymbol{x}$ mit $\boldsymbol{\nabla} f(\boldsymbol{x}) \neq \mathbf{0}$ und jeder Abstiegsrichtung $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{x})$ von f in $\boldsymbol{x}$ eine Schrittlänge $\alpha = \alpha(\boldsymbol{x}, \boldsymbol{s})$ zuordnet, heißt **effizient**, wenn es eine (von $\boldsymbol{x}$ und $\boldsymbol{s}$ unabhängige) Konstante $\delta > 0$ gibt mit

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

Der obigen Konstruktion von α^* entnimmt man, daß die exakte Schrittlängenstrategie auch effizient ist.

Mit diesen beiden Forderungen an Abstiegsrichtung und Schrittlänge läßt sich nun der folgende Konvergenzsatz zeigen:

Satz (3.27) (Konvergenzsatz)

Ein Abstiegsverfahren (Modellalgorithmus (3.5) ohne Abbruch) arbeite mit einer Abstiegsrichtung $\boldsymbol{s}(\boldsymbol{x})$, die die Winkelbedingung erfüllt, und mit einer effizienten Schrittlängenstrategie $\alpha(\boldsymbol{x}, \boldsymbol{s})$. Dann gilt eine der folgenden drei Möglichkeiten

- *Die Iteration bricht nach endlich vielen Schritten mit* $\boldsymbol{\nabla} f(\boldsymbol{x}_k) \;=\; \mathbf{0}$ *ab.*
- *Die Funktionswerte* $f(\boldsymbol{x}_k)$ *sind nach unten unbeschränkt.*
- *Die Folge der Gradienten* $\boldsymbol{\nabla} f(\boldsymbol{x}_k)$ *konvergiert gegen Null, d.h. jeder Häufungspunkt der Folge* $(\boldsymbol{x}_k)$ *ist ein stationärer Punkt von* f.

Beweis:

Mit $\nu_k \;:=\; -\dfrac{\boldsymbol{\nabla} f(\boldsymbol{x}_k)^{\mathrm{T}}\, \boldsymbol{s}_k}{\|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\|\; \|\boldsymbol{s}_k\|}$ liefert die Effizienzbedingung (3.26)

$$f(\boldsymbol{x}_{k+1}) \;-\; f(\boldsymbol{x}_k) \;\leq\; -\,\delta\, \nu_k^2\, \|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\|^2 \;.$$

Bricht die Iteration also nicht ab, so folgt hieraus für $m \in \mathbb{N}$:

$$f(\boldsymbol{x}_m) \;-\; f(\boldsymbol{x}_0) \;=\; \sum_{k=0}^{m-1} \left(f(\boldsymbol{x}_{k+1}) \;-\; f(\boldsymbol{x}_k)\right) \;\leq\; -\,\delta \sum_{k=0}^{m-1} \nu_k^2\, \|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\|^2 \,.$$

Nehmen wir weiterhin an, daß die Funktionswerte $f(\boldsymbol{x}_k)$ nach unten beschränkt sind, so ist die Reihe $\sum_{k=0}^{\infty} \nu_k^2\, \|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\|^2$ demnach (absolut) konvergent. Aufgrund der Winkelbedingung ist ν_k nun von Null weg beschränkt ist; daher folgt $\boldsymbol{\nabla} f(\boldsymbol{x}_k) \longrightarrow \mathbf{0}$ für $k \longrightarrow \infty$. □

Abschließend gehen wir noch kurz auf die Frage des Abbruchkriteriums ein (Modellalgorithmus (3.5), erster Schritt). Es ist keineswegs einfach, ein Abbruchkriterium zu formulieren, welches ohne äußeren Eingriff in allen denkbaren Fällen mit einiger Sicherheit anspricht, die gewünschte Genauigkeit liefert und zum anderen auch nicht zu viele unnötige Iterationsschritte zuläßt. Letztlich beruhen brauchbare Abbruchkriterien zumeist auf heuristischen Überlegungen. Wir geben hier die Vorschläge nach Gill, Murray und Wright [...] wieder.

Es bezeichne $\mathrm{TOL} = 10^{-r}$ eine vom Benutzer vorgegebene Genauigkeitsschranke; r entspricht etwa der Zahl der gewünschten gültigen Dezimalziffern. Gill et al. schlagen die folgenden Abbruchkriterien vor:

$$\begin{array}{lrcl}
\text{(K1)} & |f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k-1})| & \leq & \mathrm{TOL}\ (1 + |f(\boldsymbol{x}_k)|) \\
\text{(K2)} & \|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\| & \leq & \sqrt{\mathrm{TOL}}\ (1 + \|\boldsymbol{x}_k\|) \\
\text{(K3)} & \|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\| & \leq & \sqrt[3]{\mathrm{TOL}}\ (1 + |f(\boldsymbol{x}_k)|) \\
\text{(K4)} & \|\boldsymbol{\nabla} f(\boldsymbol{x}_k)\| & \leq & \mathrm{EPS}\ \text{(Maschinengenauigkeit)} \\
\text{(K5)} & k & \geq & k_{\mathrm{max}}
\end{array}$$

(K1) bis (K3) sind die Abbruchkriterien im Fall einer erfolgreichen numerischen „Konvergenz“. (K4) und (K5) sind die „Notbremsen“. (K1) bis (K3) fragen den relativen Fehler ab, falls $|f(\boldsymbol{x}_k)|$ bzw. $\|\boldsymbol{x}_k\|$ hinreichend groß sind, sie fragen den absoluten Fehler ab, falls diese Größen klein sind. Das Auftreten der Wurzel in (K2) ist durch den Taylorschen Satz begründbar:

$$f(\boldsymbol{x}_{k-1}) \;=\; f(\boldsymbol{x}_k) + \boldsymbol{\nabla} f(\boldsymbol{x}_k)^{\mathrm{T}} (\boldsymbol{x}_{k-1} - \boldsymbol{x}_k) + O(\|\boldsymbol{x}_{k-1} - \boldsymbol{x}_k\|^2)\,.$$

Ist $\boldsymbol{\nabla} f(\boldsymbol{x}_k)$ hinreichend klein und hat $f(\boldsymbol{x}_{k-1}) - f(\boldsymbol{x}_k)$ die Größenordnung TOL, so hat $\|\boldsymbol{x}_{k-1} - \boldsymbol{x}_k\|$ die Größenordnung $\sqrt{\mathrm{TOL}}$.

Das Auftreten der dritten Wurzel in (K3) wird von Gill et al. heuristisch begründet. Sie dient dazu, die Bedingung (K3) abzuschwächen, die mit der (theoretisch zu begründenden) Quadratwurzel zu restriktiv sei. Gill et al. schlagen auf der anderen Seite vor, daß ein (erfolgreicher) Abbruch nur erfolgen sollte, wenn (K1) bis (K3) *gleichzeitig* erfüllt sind.