8 Ansätze mit Vertauensbereichen (Trust regions)

Bei dem folgenden Text handelt es sich im wesentlichen um den Abschnitt 4.4 aus dem Vorlesungstext von Carl Geiger, *Optimierung*, SS 1998.

Wir haben die Newton-Richtung (vgl.(6.9)) über die quadratischen Modellfunktion

$$\tilde{f}(\boldsymbol{z}) \;:=\; f(\boldsymbol{x}) + \mathbf{g}(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{z}-\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{z}-\boldsymbol{x})^{\mathrm{T}}\boldsymbol{g}'(\boldsymbol{x})(\boldsymbol{z}-\boldsymbol{x})$$

als Abstiegsrichtung gewonnen, und zwar unter der Voraussetzung, daß die Hesse-Matrix $\mathbf{g}'(\boldsymbol{x})$, die zum erstenmal in (6.6) vorkommt, positiv definit ist. Ist $\boldsymbol{g}'(\boldsymbol{x})$ nicht positiv definit, so besitzt $\tilde{f}$ i.a. kein oder kein eindeutig bestimmtes Minimum, und das Prinzip der lokalen Minimierung versagt, und die nach (6.8) gewonnene Richtung $\boldsymbol{h}_j$ ist keine Abstiegsrichtung mehr.

Andererseits ist aber auch in diesem Fall $\tilde{f}$ in einer gewissen Umgebung Ω von $\boldsymbol{x}$ eine gute Approximation von f, man vgl. die Figur 8.1 mit

$$f(\boldsymbol{x}) := 4(x_2 - x_1^2)^2 + (1-x_1)^2. \tag{8.1}$$

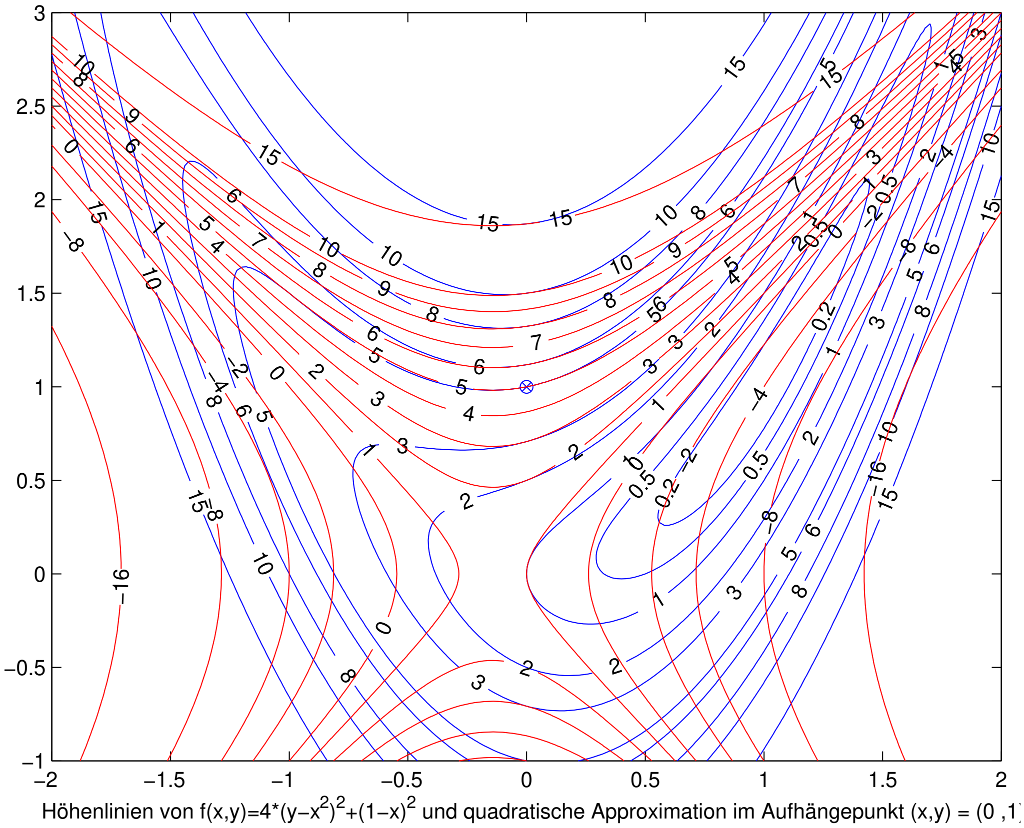

Abb. 8.1. Höhenlinen von f aus (8.1) und q aus (8.2) mit indefiniter Hesse-Matrix

Diese Situation führt zu der Idee, die Minimumsuche für $\tilde{f}$ auf einen *Vertrauensbereich (trust region)* Ω zu beschränken. Setzt man für vorgegebenes $\boldsymbol{x}$ und ein noch zu bestimmendes $h > 0$

$$\Omega := \{\boldsymbol{z} \in \mathbb{R}^n \,:\, ||\boldsymbol{z} - \boldsymbol{x}|| \leq h\},$$

so erhält man die Suchrichtung $\boldsymbol{s}$ nach dieser Idee also aus

$$\begin{array}{ll} \text{minimiere} & q(\boldsymbol{s}) \;:=\; f(\boldsymbol{x}) \;+\; \mathbf{g}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{s} \;+\; \tfrac{1}{2}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{g}\,'(\boldsymbol{x})\boldsymbol{s} \\ \text{unter der Nebenbedingung} & \|\boldsymbol{s}\| \;\leq\; h. \end{array} \tag{8.2}$$

Wie ist nun h in Abhängigkeit von $\boldsymbol{x}$ zu wählen? Da f und $\tilde{f}$ in Ω hinreichend gut übereinstimmen sollen, benötigen wir ein Maß für diese Übereinstimmung; wir nehmen dazu das Verhältnis r der beim Übergang von $\boldsymbol{x}$ zum interessierenden Punkt $\boldsymbol{x} + \boldsymbol{s}$ *eintretenden Reduktion* $f(\boldsymbol{x}) \,-\, f(\boldsymbol{x} + \boldsymbol{s})$ und der durch das quadratische Modell *vorausgesagten Reduktion* $f(\boldsymbol{x}) \,-\, \tilde{f}(\boldsymbol{x} + \boldsymbol{s}) \,=\, q(\mathbf{0}) \,-\, q(\boldsymbol{s})$, also

$$r \;:=\; \frac{f(\boldsymbol{x}) - f(\boldsymbol{x}+\boldsymbol{s})}{q(\mathbf{0}) - q(\boldsymbol{s})} \;=\; \frac{f(\boldsymbol{x}) - f(\boldsymbol{x}+\boldsymbol{s})}{f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x}+\boldsymbol{s})}. \tag{8.3}$$

Je näher r bei 1 liegt, desto besser ist die Übereinstimmung von $f(\boldsymbol{x}+\boldsymbol{s})$ und $\tilde{f}(\boldsymbol{x}+\boldsymbol{s})$.

Diese Überlegungen zur Anpassung von h werden in dem folgenden *Trust-Region-Modellalgorithmus* verwendet :

Algorithmus 8.2. Wähle $\boldsymbol{x} \in \mathbb{R}^n$, $h > 0$, $0 \leq \rho_0 < \rho_1 < \rho_2 < 1$, $0 < \sigma_1 < 1 < \sigma_2$.

`REPEAT`

Bestimme $\boldsymbol{s}$ aus (8.2);
Bestimme r aus (8.3);
`IF` $r < \rho_1$ `THEN` $h := \sigma_1 h$; `%` h `wird verkleinert`
`IF` $r > \rho_2$ `AND` $\|\boldsymbol{s}\| = h$ `THEN` $h := \sigma_2 h$; `%` h `wird vergrößert`
`IF` $r > \rho_0$ `THEN` $\boldsymbol{x} := \boldsymbol{x} + \boldsymbol{s}$;

`UNTIL` Abbruch.

Man mache sich die Plausibilität der einzelnen Anweisungen klar! Ordnet man diesen Algorithmus dem Modellalgorithmus aus Abschnitt 3 unter, so ist die letzte Anweisung als besonders einfache Schrittlängenbestimmung ($\alpha = 1$ oder 0) zu interpretieren; an die Stelle einer aufwendigeren Schrittlängenbestimmung tritt hier die Steuerung der Schranke h. Der Nenner in (8.3) kann nur im Fall $\boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}$ verschwinden.

Man kann zeigen (vgl. Werner [1992], Seite 241), daß im Falle $f \in \mathrm{C}^2$ und beschränkter Niveaumenge $L(\boldsymbol{x}^0)$ die vom Algorithmus erzeugte Folge der Punkte einen Häufungspunkt $\boldsymbol{x}^*$ mit $\mathbf{g}(\boldsymbol{x}^*) = \mathbf{0}$ und $\boldsymbol{g}\,'(\boldsymbol{x}^*)$ positiv semi-definit besitzt.

Um aus dem Trust-Region-Modellalgorithmus (auch *Restricted-Step-Methode* genannt) einen implementierbaren Algorithmus zu machen, muß ein Verfahren zur (näherungsweisen) Lösung des restringierten Optimierungsproblems (8.2) angegeben werden. Glücklicherweise kann, wenn dabei die euklidische Norm verwendet wird, (8.2) für ein $\nu \geq 0$ mit dem linearen Gleichungssystem

$$(\boldsymbol{g}'(\boldsymbol{x}) + \nu\mathbf{I})\boldsymbol{s} = -\mathbf{g}(\boldsymbol{x})\,, \tag{8.4}$$

in Verbindung gebracht werden. Um dies zu verstehen, betrachten wir die für alle $\boldsymbol{s}$, $\boldsymbol{t} \in \mathbb{R}^n$ bestehende Darstellung

$$q(\boldsymbol{t}) - q(\boldsymbol{s}) = (\mathbf{g}(\boldsymbol{x}) + \boldsymbol{g}'(\boldsymbol{x})\boldsymbol{s})^{\mathrm{T}}(\boldsymbol{t} - \boldsymbol{s}) + \frac{1}{2}(\boldsymbol{t} - \boldsymbol{s})^{\mathrm{T}}\boldsymbol{g}'(\boldsymbol{x})(\boldsymbol{t} - \boldsymbol{s}).$$

Ist $\boldsymbol{s}$ Lösung von (8.4), so folgt

$$\begin{aligned} q(\boldsymbol{t}) - q(\boldsymbol{s}) &= -\nu\boldsymbol{s}^{\mathrm{T}}(\boldsymbol{t} - \boldsymbol{s}) + \frac{1}{2}(\boldsymbol{t} - \boldsymbol{s})^{\mathrm{T}}(\boldsymbol{g}'(\boldsymbol{x}) + \nu\mathbf{I})(\boldsymbol{t} - \boldsymbol{s}) - \frac{\nu}{2}\|\boldsymbol{t} - \boldsymbol{s}\|^2 \\ &= \frac{\nu}{2}(\|\boldsymbol{s}\|^2 - \|\boldsymbol{t}\|^2) + \frac{1}{2}(\boldsymbol{t} - \boldsymbol{s})^{\mathrm{T}}(\boldsymbol{g}'(\boldsymbol{x}) + \nu\mathbf{I})(\boldsymbol{t} - \boldsymbol{s}). \end{aligned}$$

Ist überdies $\boldsymbol{g}'(\boldsymbol{x}) + \nu\mathbf{I}$ positiv semi-definit, so hat man

$$q(\boldsymbol{t}) - q(\boldsymbol{s}) \geq \frac{\nu}{2}(\|\boldsymbol{s}\|^2 - \|\boldsymbol{t}\|^2) \quad \text{für alle} \quad \boldsymbol{t}.$$

Gilt nun noch $\|\boldsymbol{s}\| \leq h$ und im Fall $\nu > 0$ sogar $\|\boldsymbol{s}\| = h$, so resultiert

$$q(\boldsymbol{t}) - q(\boldsymbol{s}) \geq 0 \quad \text{für alle } \boldsymbol{t} \text{ mit } \|\boldsymbol{t}\| \leq h,$$

d.h. $\boldsymbol{s}$ ist (globale) Lösung von (8.2). Damit haben wir die eine Richtung des folgenden Lemmas bewiesen (für die andere Richtung vgl. Fletcher [1987], p. 101).

Lemma 8.3. Sei $f \in \mathrm{C}^2$, $\boldsymbol{x} \in \mathbb{R}^n$, $h > 0$. Der Vektor $\boldsymbol{s} \in \mathbb{R}^n$ ist genau dann (globale) Lösung von (8.2), wenn es ein $\nu \geq 0$ gibt, so daß die folgenden Bedingungen erfüllt sind:
(i) $(\boldsymbol{g}'(\boldsymbol{x}) + \nu\mathbf{I})\boldsymbol{s} = -\mathbf{g}(\boldsymbol{x})$,
(ii) $\boldsymbol{g}'(\boldsymbol{x}) + \nu\mathbf{I}$ ist positiv semidefinit,
(iii) $\|\boldsymbol{s}\| \leq h$ falls $\nu = 0$, $\|\boldsymbol{s}\| = h$ falls $\nu > 0$.

Methoden, die mit dem Gleichungssystem (8.4) arbeiten, werden als *Levenberg-Marquardt-Methoden* bezeichnet; ursprünglich waren solche Verfahren für ℓ_2-Approximationsprobleme (mit einer speziellen Approximation für $\boldsymbol{g}'(\boldsymbol{x})$) entwickelt worden.

Man kann zeigen, daß $\|\boldsymbol{s}\|$ mit $\boldsymbol{s}$ nach (8.4) in Abhängigkeit von ν monoton abnimmt. Dies legt es nahe, anstelle von h die Zahl ν als Steuerparameter zu verwenden. Eine mögliche Realisierung ist der folgende Algorithmus.

Algorithmus 8.4. Wähle $\boldsymbol{x} \in \mathbb{R}^n$, $\nu > 0$, $0 \le \rho_0 < \rho_1 < \rho_2 < 1$, $0 < \sigma_1 < 1 < \sigma_2$.

```
REPEAT
    WHILE  g'(x) + νI  nicht positiv definit   DO   ν := σ₂ν ; END WHILE
    Bestimme  s  aus (8.4);
    Bestimme  r  aus (8.3);
    IF  r < ρ₁   THEN  ν := σ₂ν ;
    IF  r > ρ₂   THEN  ν := σ₁ν ;
    IF  r > ρ₀   THEN  x := x + s ;
UNTIL  Abbruch.
```

Für den Definitheitstest und das Lösen von (8.4) bietet sich das *Cholesky-Verfahren* an. Wir erinnern daran durch eine kurze Beschreibung für die $\mathbf{LDL}^{\mathrm{T}}$-Zerlegung einer symmetrischen Matrix $\mathbf{A}$ ($\ell_{ij} = 0$ für $j > i$; $\ell_{ii} = 1$; $D =$ Diagonalmatrix): durch Ausmultiplizieren der rechten Seite von $\mathbf{A} = \mathbf{LDL}^{\mathrm{T}}$ erhält man $(i, j = 1, \ldots, n)$

$$\begin{aligned} a_{jj} &= d_j &&+ \textstyle\sum_{r=1}^{j-1} d_r \ell_{jr}^2 \,, \\ a_{ij} &= d_j \ell_{ij} &&+ \textstyle\sum_{r=1}^{j-1} d_r \ell_{ir} \ell_{jr} \,, \quad i > j \end{aligned}$$

und hieraus die Berechnungsvorschrift

```
FOR  k := 1  TO  n  DO
    FOR  i := 1  TO  k − 1  DO
        r_ik := a_ik − Σ_{j=1}^{i−1} ℓ_ij r_jk
    d_k := a_kk
    FOR  i := 1  TO  k − 1  DO
        ℓ_ki := r_ik / d_i
        d_k := d_k − ℓ_ki r_ik
    IF  d_k ≤ 0  THEN  STOP : A nicht positiv definit
```

Häufig wird $(\mathbf{LD}^{1/2})\,(\mathbf{LD}^{1/2})^{\mathrm{T}}$ als *Cholesky-Zerlegung* bezeichnet. Da bei der obigen Rechnung die Symmetrie von $\mathbf{A}$ ausgenutzt wird, sind Rechen- und Speicherplatzaufwand je etwa halb so hoch wie beim Gaußschen Eliminationsverfahren.

Man kann es als einen Nachteil ansehen, daß in Algorithmus 8.4 $\nu = 0$ (also $\boldsymbol{s} =$ Newton-Richtung) auch im Falle der positiven Definitheit von $\boldsymbol{g}'(\boldsymbol{x})$ nicht vorkommen kann. Zu Weiterentwicklungen der Trust-Region-Idee vgl. Fletcher [1987], p. 103 f. und Moré [1983].

Abschließend sei angemerkt, daß die Trust-Region-Idee auch mit anderen Modellfunktionen als der hier verwendeten Funktion $\tilde{f}$ eingesetzt werden kann.