Fakultät für Luft- und Raumfahrttechnik
Institut für Mathematik und Rechneranwendung
Prof. Dr. M. Gerdts, Dr. A. Dreves

Universität der Bundeswehr München

Frühjahrstrimester 2019

5. Übung in Optimierung

16) Gesucht ist ein Minimum der Funktion

$$f(x_1, x_2) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2.$$

(a) Zeigen Sie, dass das Gradientenverfahren mit exakter Liniensuche und dem Startvektor $x^{[0]} = (0,0)^\top$ die Folge $\{x^{[k]}\}$ mit

$$x^{[k]} = \left(\frac{2}{3^k} - 2, \left(-\frac{1}{3}\right)^k - 1 \right)^\top, \quad k = 0, 1, 2, \ldots$$

liefert und berechnen Sie den Grenzwert dieser Folge.

(b) Was liefert das lokale Newton-Verfahren?

17) Berechnen Sie mit Hilfe des Quasi-Newton-Verfahrens aus Algorithmus 3.7.1 das Minimum der Funktion

$$f(x_1, x_2) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2$$

beginnend vom Startpunkt $x^{[0]} = (0,0)^\top$. Verwenden Sie die exakte Minimierungsregel als Schrittweitenstrategie, die Einheitsmatrix als Startmatrix H_0, und die BFGS-Update Formel

$$H_+ = H + \frac{yy^\top}{y^\top d} - \frac{(Hd)(Hd)^\top}{d^\top H d}, \qquad y = \nabla f(x + \alpha d) - \nabla f(x)$$

18) Zeigen Sie mithilfe einer Rechnung per Hand, dass das Trust-Region-Newton-Verfahren (Algorithmus 3.8.2) mit dem Startwert $x^{[0]} = 0$ und den Parametern $\sigma_2 = 2$ und $\Delta_0 = 1$ in zwei Schritten das Minimum der quadratischen Funktion $f(x) = x^2 + x + \frac{1}{2}$ liefert.

Lösungen

16) (a) Wir beweisen die Formel mittels vollständiger Induktion. Für $k = 0$ gilt

$$x^{[0]} = (0,0)^\top = \left(\frac{2}{3^0} - 2, \left(-\frac{1}{3}\right)^0 - 1\right)^\top .$$

Nun gelte für ein $k \in \mathbb{N}$.

$$x^{[k]} = \left(\frac{2}{3^k} - 2, \left(-\frac{1}{3}\right)^k - 1\right)^\top$$

Dann folgt:

$$d^{[k]} = -\nabla f(x^{[k]}) = -\begin{pmatrix} 2x_1^{[k]} + 4 \\ 4x_2^{[k]} + 4 \end{pmatrix} = -\begin{pmatrix} 2\left(\frac{2}{3^k} - 2\right) + 4 \\ 4\left(\left(-\frac{1}{3}\right)^k - 1\right) + 4 \end{pmatrix} = \begin{pmatrix} -\frac{4}{3^k} \\ -4\left(-\frac{1}{3}\right)^k \end{pmatrix}$$

Bei der exakten Liniensuche bestimmen wir

$$\min_\alpha \varphi(\alpha) = f(x + \alpha d).$$

Aus der notwendigen Bedingung für ein Minimum erhalten wir

$$\begin{aligned} 0 = \varphi'(\alpha) &= \nabla f(x_1 + \alpha d_1, x_2 + \alpha d_2)^\top d \\ &= (2(x_1 + \alpha d_1) + 4)d_1 + (4(x_2 + \alpha d_2) + 4)d_2 \\ &= (2x_1 d_1 + 4d_1 + 4x_2 d_2 + 4d_2) + (2d_1^2 + 4d_2^2)\alpha \end{aligned}$$

und somit

$$\alpha = -\frac{x_1 d_1 + 2x_2 d_2 + 2d_1 + 2d_2}{d_1^2 + 2d_2^2}. \tag{1}$$

Wegen $\varphi''(\alpha) = (2d_1^2 + 4d_2^2) > 0$ liegt ein Minimum vor, solange $d \neq 0$ gilt. Damit ist

$$\begin{aligned} \alpha_k &= -\frac{4d_1^{[k]} + 4d_2^{[k]} + 2x_1^{[k]} d_1^{[k]} + 4x_2^{[k]} d_2^{[k]}}{2(d_1^{[k]})^2 + 4(d_2^{[k]})^2} \\ &= -\frac{-\frac{16}{3^k} - \frac{16(-1)^k}{3^k} - \frac{16}{3^{2k}} + \frac{16}{3^k} - \frac{16}{3^{2k}} + \frac{16(-1)^k}{3^k}}{\frac{6 \cdot 16}{3^{2k}}} = \frac{1}{3}. \end{aligned}$$

Wir erhalten im Gradientenverfahren also

$$x^{[k+1]} = x^{[k]} + \alpha_k d^{[k]} = \begin{pmatrix} \frac{2}{3^k} - 2 \\ \left(-\frac{1}{3}\right)^k - 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} -\frac{4}{3^k} \\ -4\left(-\frac{1}{3}\right)^k \end{pmatrix} = \begin{pmatrix} \frac{2}{3^{k+1}} - 2 \\ \left(-\frac{1}{3}\right)^{k+1} - 1 \end{pmatrix}.$$

Nach dem Prinzip der vollständigen Induktion ist die Formel damit bewiesen.

Für den Grenzwert erhalten wir

$$\lim_{k \to \infty} x^{[k]} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}.$$

(b) Beim lokalen Newton-Verfahren berechnen wir die Newton-Richtung mittles

$$\begin{aligned} d^{[k]} &= -H_f(x)^{-1}\nabla f(x^{[k]}) = -\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 2x_1^{[k]}+4 \\ 4x_2^{[k]}+4 \end{pmatrix} \\ &= -\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 2x_1^{[k]}+4 \\ 4x_2^{[k]}+4 \end{pmatrix} = -\begin{pmatrix} x_1^{[k]}+2 \\ x_2^{[k]}+1 \end{pmatrix} \end{aligned}$$

Es folgt $x^{[k+1]} = x^{[k]} + d^{[k]} = (-2,-1)^\top$. Somit konvergiert das lokale Newton-Verfahren in einem Schritt (unabhängig vom Startpunkt) und wir haben $x^{[1]} = (-2,-1)^\top$.

17) Für $f(x_1,x_2) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2$ erhalten wir

$$\nabla f(x_1,x_2) = \begin{pmatrix} 2x_1+4 \\ 4x_2+4 \end{pmatrix}.$$

Für die Schrittweite α gilt (1). Weiter erhalten wir

$$y = \nabla f(x+\alpha d) - \nabla f(x) = \begin{pmatrix} 2(x_1+\alpha d_1)+4 \\ 4(x_2+\alpha d_2)+4 \end{pmatrix} - \begin{pmatrix} 2x_1+4 \\ 4x_2+4 \end{pmatrix} = \begin{pmatrix} 2\alpha d_1 \\ 4\alpha d_2 \end{pmatrix}$$

Nach dieser Vorüberlegung erhalten wir nun beginnend von $x^{[0]} = (0,0)^\top$:

$$\begin{aligned} d^{[0]} &= -H_0^{-1}\nabla f(x^{[0]}) = -I\begin{pmatrix} 2x_1^{[0]}+4 \\ 4x_2^{[0]}+4 \end{pmatrix} = \begin{pmatrix} -4 \\ -4 \end{pmatrix}, \\ \alpha_0 &= -\frac{x_1^{[0]}d_1^{[0]} + 2x_2^{[0]}d_2^{[0]} + 2d_1^{[0]} + 2d_2^{[0]}}{(d_1^{[0]})^2 + 2(d_2^{[0]})^2} = -\frac{-8-8}{16+32} = \frac{1}{3}, \\ x^{[1]} &= x^{[0]} + \alpha_0 d^{[0]} = \begin{pmatrix} -4/3 \\ -4/3 \end{pmatrix} \\ y^{[0]} &= \begin{pmatrix} 2\alpha_0 d_1^{[0]} \\ 4\alpha_0 d_2^{[0]} \end{pmatrix} = \begin{pmatrix} -8/3 \\ -16/3 \end{pmatrix}, \\ H_1 &= H_0 + \frac{y^{[0]}(y^{[0]})^\top}{(y^{[0]})^\top d} - \frac{(H_0 d^{[0]})(H_0 d^{[0]})^\top}{(d^{[0]})^\top H_0 d^{[0]}} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{32}\begin{pmatrix} 64/9 & 128/9 \\ 128/9 & 256/9 \end{pmatrix} - \frac{1}{32}\begin{pmatrix} 16 & 16 \\ 16 & 16 \end{pmatrix} = \frac{1}{18}\begin{pmatrix} 13 & -1 \\ -1 & 25 \end{pmatrix}, \\ d^{[1]} &= -H_1^{-1}\nabla f(x^{[1]}) = -\frac{1}{18}\begin{pmatrix} 25 & 1 \\ 1 & 13 \end{pmatrix}\begin{pmatrix} 4/3 \\ -4/3 \end{pmatrix} = \begin{pmatrix} -16/9 \\ 8/9 \end{pmatrix}, \\ \alpha_1 &= -\frac{x_1^{[1]}d_1^{[1]} + 2x_2^{[1]}d_2^{[1]} + 2d_1^{[1]} + 2d_2^{[1]}}{(d_1^{[1]})^2 + 2(d_2^{[1]})^2} = \frac{3}{8}, \\ x^{[2]} &= x^{[1]} + \alpha_1 d^{[1]} = \begin{pmatrix} -4/3 \\ -4/3 \end{pmatrix} + \frac{3}{8}\begin{pmatrix} -16/9 \\ 8/9 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}. \end{aligned}$$

Für $x^{[2]}$ gilt $\nabla f(x^{[2]}) = 0$, das Quasi-Newton-Verfahren konvergiert also in 2 Schritten.

18) Für $f(x) = x^2 + x + \frac{1}{2}$ gilt $\nabla f(x) = 2x + 1$ und $\nabla^2 f(x) = 2$.

Die Zielfunktion des Trust-Region-Teilproblems lautet

$$\begin{aligned}\hat{f}(d) &= f(x) + \nabla f(x)^\top d + \frac{1}{2} d^\top \nabla^2 f(x) d \\ &= x^2 + x + \frac{1}{2} + (2x+1)d + d^2 \\ &= (x+d)^2 + (x+d) + \frac{1}{2} \\ &= f(x+d).\end{aligned}$$

Berechnung des Quotienten q liefert

$$q = \frac{f(x) - f(x+d)}{\hat{f}(0) - \hat{f}(d)} = \frac{f(x) - f(x+d)}{f(x) - f(x+d)} = 1$$

für alle d mit $\hat{f}(d) \neq \hat{f}(0)$. Es treten also stets erfolgreiche Schritte auf.

Damit vereinfacht sich das Trust-Region-Verfahren zu:

(i) Wähle einen Startvektor $x^{[0]} \in \mathbb{R}^n$ und $\sigma_2 > 1$, $\Delta_0 > 0$. Setze $i = 0$.

(ii) Berechne

$$d^{[i]} = \arg \min_{\|d\| \leq \Delta_i} f(x^{[i]} + d).$$

(iii) Falls $f(x^{[i]}) = \hat{f}_i(d^{[i]}) = f(x^{[i]} + d^{[i]})$, STOP.

(iv) Setze $x^{[i+1]} = x^{[i]} + d^{[i]}$. Wähle $\Delta_{i+1} \in [\Delta_i, \sigma_2 \Delta_i]$. Setze $i := i + 1$ und gehe zu (ii).

Wir führen dieses Verfahren nun für die gegebenen Parameter durch:

In (i) haben wir $x^{[0]} = 0, \sigma_2 = 2$ und $\Delta_0 = 1$.

$i = 0$:

(ii) $d^{[0]} = \arg\min_{\|d\| \leq \Delta_0 = 1} f(x^{[0]} + d) = \arg\min_{\|d\| \leq 1} (d^2 + d + 1/2) = -1/2$

(iii) $f(x^{[0]}) = 1/2$, $\hat{f}_0(d^{[0]}) = f(0 - 1/2) = 1/4$, also bricht das Verfahren nicht ab.

(iv) $x^{[1]} := 0 - 1/2 = -1/2$, $\Delta_1 \in [1, 2]$, z.B. $\Delta_1 = 2$.

$i = 1$:

(ii) $d^{[1]} = \arg\min_{\|d\| \leq \Delta_1 = 2} f(x^{[1]} + d) = \arg\min_{\|d\| \leq 2} (d^2 + 1/4) = 0$

(iii) $f(x^{[1]}) = 1/4$, $\hat{f}_1(d^{[1]}) = f(-1/2 + 0) = 1/4$, also STOP.

Das gefundene Minimum ist $x = -1/2$, und dies ist auch das globale Minimum von f.