Fakultät für Luft- und Raumfahrttechnik
Institut für Mathematik und Rechneranwendung
Lineare und nichtlineare Optimierung
Dr. habil. S.-J. Kimmerle

Universität der Bundeswehr München

2. Übung

Herbsttrimester 2019

4) Die mehrdimensionale Taylor-Formel mit Restglied liefert für eine dreimal stetig differenzierbare Funktion $f : \mathbb{R}^n \to \mathbb{R}$ und einen Entwicklungspunkt x

$$f(x+d) = f(x) + \nabla f(x)^\top d + \frac{1}{2} d^\top \nabla^2 f(x) d + R(d),$$

wobei für das Restglied $R(d)$

$$\lim_{\|d\|\to 0} \frac{R(d)}{\|d\|^3} = 0$$

gilt.

Bestimmen Sie hiermit eine Approximation 1. und 2. Ordnung der Funktion

$$f(x_1, x_2) = x_1 \sin(x_2)$$

um den Entwicklungspunkt $\left(1, \frac{\pi}{2}\right)^\top$.

5) Gegeben sei die Funktion $f(x, y) = x^2 y^2 + (x^2 - 1)^2$.

   (a) Welche Informationen liefern die notwendigen und hinreichenden Bedingungen über die lokalen Minima und Maxima der Funktion f?

   (b) Bestimmen Sie alle lokalen und globalen Minima und Maxima von f.

6) Gegeben sei die Funktion $f(x, y) := 3x^4 - 4x^2 y + y^2$. Zeigen Sie:

   (a) $(0, 0)$ ist ein stationärer Punkt von f.

   (b) f besitzt längs jeder Ursprungsgeraden ein lokales Minimum in $(0, 0)$.

   (c) $(0, 0)$ ist kein lokales Minimum der Funktion f .

7) Implementieren Sie das Gradientenverfahren mit Armijo-Regel aus Algorithmus 3.3.1 mit den Parametern $\beta = 0.5$ und $\sigma = 10^{-4}$. Testen Sie es mit der Rosenbrock-Funktion

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

und dem Startvektor $x^0 = (-1.2, 1)$.

Lösungen

4) Hier arbeiten wir mit $d = y - x$, x der gegebene Entwicklungspunkt. Wir bestimmen den Gradienten und die Hessematrix von $f(x_1, x_2) = x_1 \sin(x_2)$ und erhalten:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \sin(x_2) \\ x_1 \cos(x_2) \end{pmatrix} \quad \text{und} \quad H_f(x_1, x_2) = \begin{pmatrix} 0 & \cos(x_2) \\ \cos(x_2) & -x_1 \sin(x_2) \end{pmatrix}.$$

Damit ergibt sich für die 1. Näherung um den Entwicklungspunkt $(x_1, x_2)^\top = \left(1, \frac{\pi}{2}\right)^\top$:

$$\begin{aligned} f(y_1, y_2) &= f(x_1, x_2) + \nabla f(x_1, x_2)^\top \begin{pmatrix} y_1 - x_1 \\ y_2 - x_2 \end{pmatrix} \\ &= 1 \cdot \sin(\pi/2) + (\sin(\pi/2), 1 \cdot \cos(\pi/2)) \begin{pmatrix} y_1 - 1 \\ y_2 - \pi/2 \end{pmatrix} \\ &= y_1. \end{aligned}$$

Für die 2. Näherung erhalten wir

$$\begin{aligned} f(y_1, y_2) &= f(x_1, x_2) + \nabla f(x_1, x_2)^\top \begin{pmatrix} y_1 - x_1 \\ y_2 - x_2 \end{pmatrix} + \frac{1}{2}(y_1 - x_1, y_2 - x_2) H_f(x_1, x_2) \begin{pmatrix} y_1 - x_1 \\ y_2 - x_2 \end{pmatrix} \\ &= y_1 + \frac{1}{2}(y_1 - 1, y_2 - \pi/2) \begin{pmatrix} 0 & \cos(\pi/2) \\ \cos(\pi/2) & -\sin(\pi/2) \end{pmatrix} \begin{pmatrix} y_1 - 1 \\ y_2 - \pi/2 \end{pmatrix} \\ &= y_1 - \frac{1}{2}\left(y_2 - \frac{\pi}{2}\right)^2. \end{aligned}$$

5) (a) Für die Funktion $f(x, y) = x^2 y^2 + (x^2 - 1)^2$ gilt

$$\nabla f(x, y) = \begin{pmatrix} 2xy^2 + 4x(x^2 - 1) \\ 2x^2 y \end{pmatrix}.$$

Die notwendige Bedingung 1. Ordnung $\nabla f(x, y) = (0, 0)^T$ liefert also die Kandidaten $x = 0$ und y beliebig, oder $y = 0$ und $x = \pm 1$. Das hinreichende Kriterium 2. Ordnung liefert, für Punkte mit $\nabla f(x, y) = (0, 0)^T$, dass wenn $\nabla^2 f(x, y)$ positiv bzw. negativ definit ist, (x, y) ein Minimum bzw. Maximum ist. Es gilt

$$\nabla^2 f(x, y) = \begin{pmatrix} 2y^2 + 12x^2 - 4 & 4xy \\ 4xy & 2x^2 \end{pmatrix}.$$

Nun ist $\nabla^2 f(0, y) = \begin{pmatrix} 2y^2 - 4 & 0 \\ 0 & 0 \end{pmatrix}$ positiv semidefinit für alle $|y| \geq \sqrt{2}$ und negativ semidefinit für alle $|y| < \sqrt{2}$, liefert also keine weitere Aussage über Minima oder Maxima für die Punkte $(0, y)$. Weiterhin ist $\nabla^2 f(\pm 1, 0) = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}$ positiv definit, und somit hat f in $(\pm 1, 0)$ lokale Minima.

(b) Wegen $\lim_{x\to\infty} f(x,y) = +\infty$ gibt es keine globalen Maxima. Wegen $f(x,y) \geq 0$ sind die Punkte $(\pm 1, 0)$ wegen $f(\pm 1, 0) = 0$ die einzigen globalen Minima.

Sei $\varepsilon > 0$ gegeben und seien δ_x, δ_y so, dass $(0+\delta_x, y+\delta_y) \in \mathcal{B}_\varepsilon(0,y)$ ist. Dann gilt

$$f(0+\delta_x, y+\delta_y) = \delta_x^2(y+\delta_y)^2 + (\delta_x^2 - 1)^2 = 1 + \delta_x^2(y^2 - 2 + 2y\delta_y + \delta_x^2 + \delta_y^2).$$

Für $|y| > \sqrt{2}$ existiert ein hinreichend kleines $\varepsilon > 0$, so dass $(y^2 - 2 + 2y\delta_y + \delta_x^2 + \delta_y^2) > 0$ für alle zulässigen δ_x, δ_y, und damit folgt in der gesamten Umgebung $f(0,y) = 1 \leq f(0+\delta_x, y+\delta_y)$, also ist $(0,y)$ ein lokales Minimum.

Für $|y| < \sqrt{2}$ existiert ein hinreichend kleines $\varepsilon > 0$, so dass $(y^2 - 2 + 2y\delta_y + \delta_x^2 + \delta_y^2) < 0$ für alle zulässigen δ_x, δ_y, und damit folgt in der gesamten Umgebung $f(0,y) = 1 \geq f(0+\delta_x, y+\delta_y)$, also ist $(0,y)$ ein lokales Maximum.

Für $|y| = \sqrt{2}$ existieren (δ_x, δ_y) mit $\delta_x \neq 0$, so dass $(y^2 - 2 + 2y\delta_y + \delta_x^2 + \delta_y^2)$ sowohl negative als auch positive Werte annimmt. Also können keine lokalen Extrema vorliegen.

Wegen $f(0,y) = 1$ für alle $y \in \mathbb{R}$ sind die lokalen Extrema nicht strikt.

6) (a) Für $f(x,y) := 3x^4 - 4x^2y + y^2$ gilt $\nabla f(x,y) = \begin{pmatrix} 12x^3 - 8xy \\ -4x^2 + 2y \end{pmatrix}$.

Damit ist $\nabla f(0,0) = (0,0)^T$, also ist $(0,0)$ ein stationärer Punkt.

(b) Entlang der x-Achse gilt $y = 0$, also $f(x,y) = 3x^4$, und entlang der y-Achse ist $x = 0$ und damit $f(x,y) = y^2$. In beiden Fällen hat f in $(0,0)$ offensichtlich ein striktes lokales Minimum. Alle anderen Ursprungsgeraden haben die Form $y = kx$ mit $k \neq 0$. Dann folgt für $g(x) := f(x, kx)$:

$$g(x) = 3x^4 - 4kx^3 + k^2x^2 = 3x^4 + x^2(k^2 - 4kx)$$

Damit ist $g(x) > 0$ für alle $x \neq 0$ mit $|x| \leq \frac{|k|}{4}$. Wegen $g(0) = 0$, ist 0 ein striktes lokales Minimum von g, und somit ist $(0,0)$ ein striktes lokales Minimum von f entlang jeder Ursprungsgeraden.

(c) Für $y = 2x^2$ gilt

$$f(x,y) = 3x^4 - 8x^4 + 4x^4 = -x^4 < 0 = f(0,0)$$

für alle $x \neq 0$. Es gibt also in jeder Umgebung von $(0,0)$ einen Punkt $(x, 2x^2)$ mit $f(x, 2x^2) < f(0,0)$, und somit ist $(0,0)$ kein lokales Minimum.

7)
```
function x=Gradientenverfahren(x,f,Df)

sigma=1e-4; %Parameter Armijo-Liniensuche
beta=0.5;
eps=1e-4; %Fehlertoleranz Abbruchkriterium
kmax=1e5; %5; %max. Schritte in Gradientenrichtung
alphamin=1e-4; %minimale Schrittweite (ergibt Maximalzahl an Armijo-
    Iterationen)
k=0;
fcount=0; %zaehlt Funktionsauswertungen
Dfcount=0; %zaehlt Gradientenauswertungen

%Ausgabe #Schritte, Schrittweite, Abbruchkriterium, Zwischenwert, #
    Funktions-/Gradientenauswertungen
%fprintf('%5.0f %6.4f %12.7f %12.7f %12.7f %6.0f %6.0f\n',k,alpha,
    term,x(1),x(2),fcount,Dfcount);
fprintf('    k  alpha         term          x(1)          x(2) fcount
    Dfcount \n');

fx=feval(f,x);
fcount=fcount+1;
nablafx=feval(Df,x);
Dfcount=Dfcount+1;
term=norm(nablafx); %fuer Abbruchkriterium

while term > eps && k<=kmax %Gradientenverfahren

    alpha=1;
    fxalphad=feval(f,x-nablafx);
    fcount=fcount+1;
    while fxalphad > fx -alpha*sigma*term^2 && alpha > alphamin %
        Armijo-Liniensuche
        alpha=alpha*beta;
        fxalphad=feval(f,x-alpha*nablafx);
        fcount=fcount+1;
    end
    x=x-alpha*nablafx; %Update
    fx=fxalphad; %letzter Funktionswert wird angenommen
    nablafx=feval(Df,x);
    Dfcount=Dfcount+1;
    term=norm(nablafx); %fuer Abbruchkriterium
```

```
    k=k+1;
    %Ausgabe #Schritte, Schrittweite, Abbruchkriterium, Zwischenwert,
        #Funktions-/Gradientenauswertungen
    %if mod(k, 25) == 0 %falls k durch 100 teilbar Pause zur Ansicht
        der Ausgabe beim Testen
    %     pause
    %     fprintf('    k   alpha            term           x(1)           x(2)
          fcount Dfcount \n');
    %end
    fprintf('%5.0f %6.4f %12.7f %12.7f %12.7f %6.0f %6.0f\n',k,alpha,
        term,x(1),x(2),fcount,Dfcount);
end
fprintf('\n')
end

function fx=Rosenbrock(x)
fx=100*(x(2)-x(1)^2)^2+(1-x(1))^2;
end

function fx=DRosenbrock(x)
fx=[ 400*(x(1)^2-x(2))*x(1)+2*(x(1)-1);
     200*(x(2)-x(1)^2)];
end
```

Mit diesen Funktionen erhalten wir:

```
>> x0 = [-1.2; 1];
>> Gradientenverfahren(x0, @Rosenbrock, @DRosenbrock)
    k   alpha         term          x(1)          x(2)  fcount  Dfcount
    1  0.0010   43.8985209    -0.9894531     1.0859375      12        2


.
.
.

 8058  0.0020    0.0000998     0.9999206     0.9998407   80045     8059

ans =
    0.9999
    0.9998
```