Prof. Dr. M. Hinze
Dr. J. Sternberg-Kaletta SoSe 2010

Übungen zur Vorlesung Optimierung
Blatt 3
Abgabetermin: Theoretische Aufgaben 30.04.2010 vor der Übung. Vorführung Numerik 06.05.2010, 16.00 in GEOM 142

Aufgabe 8: (4 Punkte ($2+2$))

a) Approximieren Sie den Gradienten und die Hesse-Matrix der Funktion f mit zentralen Differenzen:

$$\nabla_h f(x) = \frac{\hat{f}(x+h) - \hat{f}(x-h)}{2h}.$$

Zeigen Sie:

$$\varepsilon_{\nabla^2 f} = \mathcal{O}(\varepsilon_{\nabla f}^{\frac{2}{3}}) = \mathcal{O}(\varepsilon_f^{\frac{4}{9}}).$$

Dabei bezeichnet $\hat{f}(x)$ die gestörte Funktion $f(x)$ definiert als:

$$\hat{f}(x) = f(x) + \varepsilon_f(x).$$

ε_f, $\varepsilon_{\nabla f}$ und $\varepsilon_{\nabla^2 f}$ sind jeweils die Störungen der Funktion f, des Gradienten ∇f und der Hesse-Matrix $\nabla^2 f$.

b) Approximieren Sie den Gradienten der Funktion f mit zentralen Differenzen und die Hesse-Matrix der Funktion f mit Vorwärtsdifferenzen. Zeigen Sie, dass in diesem Fall gilt:

$$\varepsilon_{\nabla^2 f} = \mathcal{O}(\sqrt{\varepsilon_{\nabla f}}) = \mathcal{O}(\varepsilon_f^{\frac{1}{3}}).$$

Warum wird diese Vorgehensweise in der Praxis i.d.R. angewendet? Vergleichen Sie dazu auch den Satz 2.3.4 im Buch von Kelley.

Aufgabe 9: (4 Punkte)

Betrachten Sie das vereinfachte Newtonverfahren zur Minimierung der zweimal stetig differenzierbaren Funktion $f : \mathbb{R}^n \to \mathbb{R}$:

$x^0 \in \mathbb{R}^n$ ist gegeben.

Die Iterationsvorschrift ist:

$$x^{[i+1]} = x^{[i]} - \left(\nabla^2 f(x^{[0]})\right)^{-1} \nabla f(x^{[i]}), \quad i = 0, 1, 2, \ldots$$

Es sei $x^\star$ ein stationärer Punkt von f und $\nabla^2 f(x^\star)$ sei invertierbar. Zeigen Sie:

Es existiert ein $\delta > 0$, so daß das vereinfachte Newtonverfahren für alle $x^{[0]} \in B_\delta(x^\star)$ wohldefiniert ist und eine gegen $x^\star$ linear konvergente Folge $\{x^{[i]}\}$ liefert. Setzen Sie dabei den Satz 2.3.4 aus dem Buch von Kelley als bekannt voraus.

Aufgabe 10: (4 Punkte)

Zeigen Sie, daß das lokale Newtonverfahren invariant bezüglich affin-linearer Transformationen $x = Ay + b$ ist.

Ein Verfahren heißt invariant bezüglich affin-linearer Transformationen, wenn das Verfahren angewendet auf $f(x)$ eine Folge $\{x^{[i]}\}$ und angewendet auf $g(y) := f(Ay + b)$ eine Folge $\{y^{[i]}\}$ erzeugt, welche die Implikation

$$x^{[0]} = Ay^{[0]} + b \quad \Rightarrow \quad x^{[i]} = Ay^{[i]} + b \quad \forall i \in \mathbb{N}$$

erfüllen. Dabei sei $A \in \mathbb{R}^{n \times n}$ invertierbar und $b \in \mathbb{R}^n$.

Numerische Aufgabe:

i) Implementieren Sie das globalisierte inexakte Newton-CG-Verfahren (siehe die Vorlesung am 23.04.2010).

Globalisiertes inexaktes Newton-CG-Verfahren

1). $x^0 \in \mathbb{R}^n$, $K > 0$, $p > 2$, $\rho \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$, $\epsilon \geq 0$, $k := 0$, $d^{-1} := 0$.

2). IF $\|\nabla f(x^k)\| \leq \epsilon$ STOP.

3). Berechne d^k mit dem CG-Verfahren angewendet auf $\nabla^2 f(x^k) d^k = -\nabla f(x^k)$ gestartet von $d^k = 0$: Wähle $\eta_k \geq 0$, $k_{max} > 0$.

CG $(x^k, f, \eta_k, k_{max}, d^k)$.

Terminiert CG-Verfahren nicht oder gilt

$$\nabla f(x^k)^\top d^k > -K \mid d^k \mid^p,$$

so $d^k = -\nabla f(x^k)$.

4). Wähle die Schrittweite t_k durch:

$t_k := \max(\beta^l : l = 0, 1, 2, ...; f(x^k + \beta^l d^k) \leq f(x^k) + \sigma \beta^l \nabla f(x^k)^\top d^k)$.

5). $x^{k+1} = x^k + t_k d^k$, $k := k + 1$, gehe nach 2).

Verwenden sie dabei:

$$\eta_k = \min\left\{\frac{c_1}{k+1}, c_2 \|\nabla f(x^k)\|\right\}, \qquad c_1 = c_2 = 1.$$

CG-Verfahren: (CG (x, f, η, k_{max}, d)) (vgl. auch Alg. 2.5.1 im Buch von Kelley)

1). $r = -\nabla f(x)$, $\rho_0 = \|r\|_2^2$, $k = 1$, $d = 0$.

2). IF ($\sqrt{\rho_{k-1}} \leq \eta \|\nabla f(x)\|$ oder $k \geq k_{max}$) STOP.

(a) IF $k = 1$, THEN $p = r$,

ELSE $\beta = \rho_{k-1}/\rho_{k-2}$, $p = r + \beta p$.

(b) $w = \nabla^2 f(x) p$

IF $p^\top w = 0$ (Indefinitheit), STOP.

IF $p^\top w < 0$ (negative Krümmung), STOP.

(c) $\alpha = \rho_{k-1}/p^\top w$.

(d) $d = d + \alpha p$.

(e) $r = r - \alpha w$.

(f) $\rho_k = \|r\|^2$.

(g) $k = k + 1$.

ii) Wenden Sie das Programm zur Minimierung von Browns schlecht skalierter Funktion $f : I\!R^n \to I\!R$,

$$f(x) = (x_1 - 10^6)^2 - (x_2 - 2 \cdot 10^{-6})^2 + (x_1 x_2 - 2)^2$$

mit Startpunkt $x^0 = (1,1)^T$ an.
Parameterwerte: $\sigma = 10^{-4}, \beta = 0.5, K > 0, p > 2, k_{max}$ geeignet, Abbruchbedingung: $\left\|\nabla f(x^k)\right\| \leq 10^{-8}$. Wieviele Iterationen, Funktions-, Gradienten- und Hesse-Matrix-Auswertungen werden benötigt?

iii) Lassen Sie die Höhenlinien der Funktion und den Pfad der globalisierten inexakten Newton-CG-Iteration zeichen.