Nichtlineare Optimierung
7. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Prof. Dr. Stefan Ulbrich
Hannes Meinlschmidt

WS 2011-2012
2.12.2011

Gruppenübung

Aufgabe G1 (Inexaktes CG-Newton-Verfahren)
In der Vorlesung wurden inexakte Newton-Verfahren zur Minimierung einer zweimal stetig differenzierbaren Funktion $f : \mathbb{R}^n \to \mathbb{R}$ vorgestellt. Um nun eine inexakte Lösung der Newton-Gleichung zu berechnen, kann man das CG-Verfahren (siehe auch Aufgabe H2 vom 2. Rechnerübungsblatt bzw. Übung 6) verwenden. Wie dort gesehen, liefert dieses für symmetrische, positiv definite Matrizen $C \in \mathbb{R}^{n,n}$ und $c \in \mathbb{R}^n$ die Lösung des Gleichungssystems $Cy = -c$. Dazu wird (in höchstens n Schritten) die streng konvexe Funktion $q(y) = \frac{1}{2}y^T C y + c^T y$ minimiert.
Die Idee ist nun, das CG-Verfahren auf die Funktion $q_k(s) := \frac{1}{2}s^T \nabla^2 f(x_k)s + \nabla f(x_k)^T s$ anzuwenden und abzubrechen, wenn das Residuum $\|\nabla f(x_k) + \nabla^2 f(x_k)s\|$ klein genug ist. Damit erhält man eine näherungsweise Lösung der Newton-Gleichung $\nabla^2 f(x_k)s = -\nabla f(x_k)$.
Im allgemeinen Abstiegsverfahren (Algorithmus 3 im Skript) verwenden wir also zur Berechnung der Suchrichtung s_k das folgende modifizierte CG-Verfahren:

Algorithmus 1 : Inexaktes CG-Newton-Verfahren zur Bestimmung der Suchrichtung

Input : $\alpha, \nu \in (0,1)$ beliebig, aber fest.
1 Wähle $y_0 = 0$, setze $g_0 = \nabla f(x_k)$ und $d_0 := \nabla f(x_k)$, sowie $j := 0$;
2 **if** $\|g_j\| \le \min\{\nu, \|\nabla f(x_k)\|\}\|\nabla f(x_k)\|$ **then** (relatives Residuum klein genug)
3 STOP mit $s_k = y_j$.
4 **end**
5 **if** $d_j^T \nabla^2 f(x_k) d_j \le 0$ **then** (Richtung nichtpositiver Krümmung)
6 STOP mit Ergebnis $s_k = y_j - \operatorname{sign}(\nabla f(x_k)^T d_j)\|\nabla f(x_k)\| \frac{d_j}{\|d_j\|}$.
7 **end**
8 Berechne $\alpha_j = \frac{g_j^T g_j}{d_j^T \nabla^2 f(x_k) d_j}$;
9 Setze $y_{j+1} = y_j - \alpha_j d_j$ sowie $g_{j+1} := g_j - \alpha_j \nabla^2 f(x_k) d_j$;
10 **if** $-\nabla f(x_k)^T y_{j+1} < \min\{\alpha, \|\nabla f(x_k)\|\}\|\nabla f(x_k)\|\|y_{j+1}\|$ **then** (Abstiegsrichtung wird unzureichend)
11 STOP mit Ergebnis $s_k = y_j$.
12 **end**
13 Berechne $\beta_j := \frac{g_{j+1}^T g_{j+1}}{g_j^T g_j}$ und setze $d_{j+1} := g_{j+1} + \beta_j d_j$;
14 Setze $j \leftarrow j+1$ und gehe nach 2;

Zur Bestimmung der Schrittweite im inexakten Newton-Verfahren werde die Armijo-Regel mit Parametern $\gamma \in (0, \frac{1}{2})$ und $\beta \in (0,1)$ verwendet. Sei $x_0 \in \mathbb{R}^n$ und die Niveaumenge $N_f(x_0)$ kompakt. Zeigen Sie:

(a) Es gilt $\|s_k\| \ge \delta \|\nabla f(x_k)\|$ und $\|y_j\| \ge \delta \|\nabla f(x_k)\|$ mit einem $\delta > 0$.

(b) Die erzeugten Suchrichtungen s_k sind zulässig.
Hinweis: Zeigen Sie hierzu, dass die verallgemeinerte Winkelbedingung erfüllt ist.

(c) Die mit der Armijo-Regel erzeugten Schrittweiten σ_k sind zulässig.

(d) Ist $\nabla^2 f(\bar{x})$ positiv definit und gilt $x_k \to \bar{x}$, so konvergiert $x_k \to \bar{x}$ Q-superlinear oder sogar Q-quadratisch, falls $\nabla^2 f(x)$ in einer Umgebung von $\bar{x}$ lokal Lipschitz-stetig ist.

Hinweis: Zeigen Sie, dass es ein $K \in \mathbb{N}$ gibt, so dass das inexakte CG-Verfahren in allen Iterationen $k \geq K$ des Abstiegsverfahrens die Suchrichtungsbestimmung nur abbricht, weil das Residuum klein genug ist. Zeigen Sie nun, dass die Voraussetzungen des Satzes 2.9.1 ii) bzw. iii) für $F(x) = \nabla f(x)$ erfüllt sind.

Hinweis: Die beiden Ungleichungen

$$-\nabla f(x_k)^T s_k \geq \frac{\|\nabla f(x_k)\|^2}{1+2\|\nabla^2 f(x_k)\|} \quad \text{und} \quad -\nabla f(x_k)^T y_j \geq \frac{\|\nabla f(x_k)\|^2}{1+2\|\nabla^2 f(x_k)\|} \tag{1}$$

dürfen ohne weiteren Beweis verwendet werden.

Aufgabe G2 (Rang-1-Updates für Quasi-Newton-Verfahren)

(a) Zeigen Sie: Der Ansatz für einen symmetrischen Rang-1-Update der Form

$$H_{k+1} = H_k + \alpha_k v_k v_k^T$$

mit $\alpha_k \in \{\pm 1\}$, $0 \neq v_k \in \mathbb{R}^n$ und einer symmetrischen Matrix $H_k \in \mathbb{R}^{n,n}$, bei dem H_{k+1} die Quasi-Newton-Gleichung erfüllen soll, führt unter der Voraussetzung $(y_k - H_k d_k)^T d_k \neq 0$ auf die Formel

$$H_{k+1} = H_k + \frac{(y_k - H_k d_k)(y_k - H_k d_k)^T}{(y_k - H_k d_k)^T d_k}. \tag{SR1}$$

(b) Sei $H \in \mathbb{R}^{n \times n}$ regulär und $u, v \in \mathbb{R}^n$ gegeben. Zeigen Sie: Die Matrix $H + uv^T$ ist regulär, wenn $1 + v^T H^{-1} u \neq 0$ ist, und es gilt die sogenannte *Sherman-Morrison-Formel*

$$(H + uv^T)^{-1} = \left(I - \frac{H^{-1} u v^T}{1 + v^T H^{-1} u} \right) H^{-1}.$$

(c) Leiten Sie mit der Sherman-Morrison-Formel die zur Formel (SR1) gehörige inverse Aufdatierungsformel

$$B_{k+1} = B_k + \frac{(d_k - B_k y_k)(d_k - B_k y_k)^T}{(d_k - B_k y_k)^T y}$$

her.

Hausübung

Aufgabe H1 (BFGS-Aufdatierung) (4 Punkte)

(a) Zeigen Sie, dass die inverse BFGS-Aufdatierung, d.h. Gleichung (2.65) im Skript, auch in der Form

$$B_{k+1}^{BFGS} = V_k^T B_k V_k + \rho_k d_k d_k^T$$

geschrieben werden kann, wobei $V_k = I - \rho_k y_k d_k^T$ und $\rho_k = \frac{1}{d_k^T y_k}$.

(b) Sei $B_0 \in \mathbb{R}^{n,n}$ gegeben. Zur Berechnung der Suchrichtung im Schritt k, d.h. $s_k = -B_k \nabla f(x_k)$, ist ein rekursives Verfahren in Funktion `bfgsrek(k,w)` angegeben. Zeigen Sie, dass der Aufruf $v =$ `bfgsrek`(k, w) das Ergebnis $v = B_k w$ liefert, wobei B_k die k-te inverse BFGS-Matrix ist.

Funktion `bfgsrek`(k,w)

if $k = 0$ **then**
  return $B_0 w$;
end
Berechne $\rho = \frac{1}{d_{k-1}^T y_{k-1}}$ und $\alpha = \rho d_{k-1}^T w$;
Setze $w_1 = w - \alpha y_{k-1}$;
Berechne $w_2 =$ `bfgsrek`$(k-1, w_1)$;
return $w_2 + (\alpha - \rho y_{k-1}^T w_2) d_{k-1}$

Aufgabe H2 (DFP und BFGS) (6 Punkte)

Sei H_k symmetrisch und invertierbar. Zeigen Sie Lemma 2.10.2:

(a) Gilt $y_k^T d_k \neq 0$, $d_k^T H_k d_k \neq 0$ und $y_k^T H_k^{-1} y_k \neq 0$, so sind H_{k+1}^{DFP} sowie H_{k+1}^{BFGS} invertierbar und es gilt

$$(H_{k+1}^{DFP})^{-1} = \Phi^{BFGS}(H_k^{-1}, y_k, d_k)$$

und

$$(H_{k+1}^{BFGS})^{-1} = \Phi^{DFP}(H_k^{-1}, y_k, d_k).$$

Hinweis: Wegen $y_k^T d_k \neq 0$ lässt sich jeder Vektor $v \in \mathbb{R}^n$ schreiben als $v = u + \lambda d_k$, wobei $u \perp y_k$ (orthogonale Zerlegung). Berechnen Sie zunächst $H_{k+1}^{DFP} v$, um die erste Gleichung zu zeigen. Benutzen Sie weiter, dass H_{k+1}^{DFP} die Quasi-Newton-Gleichung erfüllt.

(b) Ist H_k symmetrisch positiv definit und $y_k^T d_k > 0$, dann sind H_{k+1}^{DFP}, H_{k+1}^{BFGS} und $H_{k+1}^{B,\lambda}$ für alle $\lambda \in [0, 1]$ wieder positiv definit.