Nichtlineare Optimierung
7. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Prof. Dr. Irwin Yousept
Hannes Meinlschmidt

WS 2012/13
6.12.2012

Nächste Woche wieder Rechnerübung in S2|15 K313!

Gruppenübung

Aufgabe G1 (Verfahren der konjugierten Gradienten [Conjugated Gradient Method])
Gegeben sei die quadratische Funktion

$$q : \mathbb{R}^n \to \mathbb{R} : y \mapsto c^T y + \frac{1}{2} y^T C y$$

mit $c \in \mathbb{R}^n$ und $C \in \mathbb{R}^{n\times n}$ symmetrisch positiv definit. Zur Bestimmung des (eindeutigen) globalen Minimums $-C^{-1}c$ von q betrachten wir folgenden Algorithmus:

Algorithmus 1 : Verfahren der konjugierten Gradienten

1 Wähle y^0 und berechne $g^0 := c + Cy^0$;
2 **if** $g^0 = 0$ **then**
3   STOP mit Ergebnis y^0.
4 **else**
5   Setze $k \leftarrow 0$ und $d^0 = g^0$;
6 **end**
7 Berechne $\alpha_k := \frac{{g^k}^T g^k}{{d^k}^T C d^k}$;
8 Setze $y^{k+1} := y^k - \alpha_k d^k$ sowie $g^{k+1} := g^k - \alpha_k C d^k$;
9 **if** $g^{k+1} = 0$ **then**
10   STOP mit Ergebnis y^{k+1}.
11 **end**
12 Berechne $\beta_k := \frac{{g^{k+1}}^T g^{k+1}}{{g^k}^T g^k}$;
13 Setze $d^{k+1} := g^{k+1} + \beta_k d^k$;
14 Setze $k \leftarrow k+1$ und gehe zu 7.

Sei V_{k+1} definiert als $\mathrm{Span}\{g^0, Cg^0, \ldots, C^k g^0\}$. Folgende Aussagen dürfen Sie als gegeben annehmen: Solange $g^k \neq 0$ ist, gilt:

(a) $d^k \neq 0$,

(b) $V_{k+1} = \mathrm{Span}\{g^0, \ldots, g^k\} = \mathrm{Span}\{d^0, \ldots, d^k\}$,

(c) die Vektoren $d^0, \ldots d^k$ sind *paarweise C-konjugiert*, d.h.

$$ {d^i}^T C d^j = 0 \quad \text{für alle } i, j \in \{0, \ldots, k\} \text{ mit } i \neq j,$$

(d) g^{k+1} ist orthogonal zum Unterraum V_{k+1}, also $g^{k+1} \perp V_{k+1}$.

Zeigen Sie damit: Es gilt $q(y^{k+1}) = \min_{y \in V_{k+1}} q(y^0 + y)$ und das Verfahren berechnet in höchstens n Schritten das globale Minimum von q.
Hinweis: Interpretieren Sie $(d^0, \ldots, d^k)$ als Basis von V_{k+1} und formulieren Sie das Minimierungsproblem auf dem Unterraum um in ein Minimierungsproblem auf $\mathbb{R}^m$ für ein geeignetes m. Nutzen Sie dann die Konstruktion der y^k aus dem Algorithmus und die in (a) bewiesenen Eigenschaften.

Aufgabe G2 (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 aus Aufgabe G1 verwenden. Die Idee ist dabei, das CG-Verfahren auf die Funktion $q_k(s) := \frac{1}{2}s^T\nabla^2 f(x^k)s + \nabla f(x^k)^T s$ loszulassen und den Algorithmus 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 verwenden wir also zur Berechnung der Suchrichtung s^k das folgende, leicht modifizierte CG-Verfahren:

Algorithmus 2 : Inexaktes CG-Newton-Verfahren zur Bestimmung der Suchrichtung

Input : $\alpha, \nu \in (0,1)$ beliebig, aber fest, aktuelle Iterierte x^k des allgemeinen Abstiegsverfahrens
1 Wähle $y^0 = 0$, setze $g^0 = \nabla f(x^k)$ und $d^0 := \nabla f(x^k)$, sowie $j := 0$;
2 **if** $\|g^j\| \leq \min\{\nu, \|\nabla f(x^k)\|\}\|\nabla f(x^k)\|$ **then** (relatives Residuum klein genug)
3 STOP mit $s^k = y^j$.
4 **end**
5 **if** ${d^j}^T\nabla^2 f(x^k)d^j \leq 0$ **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 $\gamma \in (0, 1/2)$ 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 $\|s^k\| \geq \delta\|\nabla f(x^k)\|$ und $\|y^j\| \geq \delta\|\nabla f(x^k)\|$ mit einem $\delta > 0$.

(b) Die erzeugten Suchrichtungen s^k sind zulässig.

(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 bzw. sogar q-quadratisch, falls $\nabla^2 f$ in $\bar{x}$ 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 nur deshalb abbricht, weil das Residuum klein genug ist. Zeigen Sie nun, dass die Voraussetzungen des Konvergenzsatzes über Inexakte Newton-Verfahren 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.

Hausübung

Aufgabe H1 (Rang-1-Updates für Quasi-Newton-Verfahren) (4 Punkte)

(a) 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*

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

(b) In der Vorlesung haben wir gesehen, dass der Ansatz eines symmetrischen Rang-1 Quasi-Newton Updates zum Update

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

führt. 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.

Aufgabe H2 (BFGS-Aufdatierung) (5 Punkte)

(a) Zeigen Sie, dass die inverse BFGS-Aufdatierung 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 $B_0 \in \mathbb{R}^{n\times n}$ gegeben. Zur Berechnung der Suchrichtung im Schritt k eines BFGS-Verfahrens, 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 $v = \texttt{bfgsrek}(k, w)$ 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 $w_2 = \texttt{bfgsrek}(k-1, w_1)$;
return $w_2 + (\alpha - \rho {y^{k-1}}^T w_2) d^{k-1}$

Aufgabe H3 (DFP und BFGS) (6 Punkte)

Sei H_k symmetrisch und invertierbar. Zeigen Sie, dass die DFP- und BFGS-Updates jeweils durch gegenseitige inverse Updates erzeugt werden können:

(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.