Einführung in die Grundlagen der Numerik

Wintersemester 2018/19
Prof. Dr. Ira Neitzel
Fabian Hoppe

UNIVERSITÄT BONN

Übungsblatt 5.

Abgabe am **15. November vor der Vorlesung**.

Aufgabe 1. Sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und zweimal stetig differenzierbar. Es gelte

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2 \text{ für } x, y \in \mathbb{R}^n.$$

Wir betrachten das Gradientenverfahren mit *konstanter Schrittweite*, d.h.

$$x_{k+1} := x_k - \alpha \nabla f(x_k),$$

mit einem a priori *fixierten* $\alpha > 0$. Im Folgenden bezeichnen wir mit x^* ein globales Minimum von f und mit $\nabla^2 f(x)$ die Hesse-Matrix von f an der Stelle x.

Wie wir bereits in den Übungen gesehen haben, kann dieses Verfahren versagen und sollte daher gemieden werden. In Anwendungen, in denen eine bessere Schrittweitensteuerung nicht oder nur mit zu hohem Aufwand möglich ist, kann die konstante Wahl der Schrittweite trotzdem eine Alternative sein. Problematischerweise wird die Lipschitz-Konstante L in der Regel nicht bekannt sein, was eine Verifizierung der Bedingung $\alpha \leq L^{-1}$ (siehe Teilaufgabe c.) in der Praxis unmöglich macht.

a) Zeigen Sie für $z, w \in \mathbb{R}^n$:

$$w^T \nabla^2 f(z) w \leq L\|w\|_2^2.$$

b) Zeigen Sie für alle $k \in \mathbb{N}$:

$$f(x_{k+1}) \leq f(x_k) - \left(1 - \frac{\alpha L}{2}\right) \alpha \|\nabla f(x_k)\|_2^2$$

Hinweis: Taylor-Entwicklung und a).

c) Sei nun $0 < \alpha \leq \frac{1}{L}$. Folgern Sie aus b), dass dann

$$f(x_{k+1}) \leq f(x^*) + \frac{1}{2\alpha}\left(\|x_k - x^*\|_2^2 - \|x_{k+1} - x^*\|_2^2\right)$$

gilt.

Hinweis: Überlegen Sie sich, dass $f(x) \leq f(x^*) + \nabla f(x)^T (x - x^*)$ für alle x gilt.

d) Zeigen Sie schließlich

$$f(x_k) - f(x^*) \leq \frac{1}{k} \frac{\|x_0 - x^*\|_2^2}{2\alpha}.$$

Hinweis: Machen Sie sich klar, dass $(f(x_k))_k$ monoton fällt und nutzen Sie eine Teleskopsumme.

(2+2+3+3 = 10 Punkte)

Aufgabe 2. Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch positiv definit und $b \in \mathbb{R}^n$. Die quadratische Funktion

$$f : \mathbb{R}^n \to \mathbb{R}, \qquad x \mapsto \frac{1}{2}x^T Ax + b^T x,$$

ist dann strikt konvex. Es bezeichne $x^* := -A^{-1}b$ das strikte globale Minimum von f. In dieser Aufgabe werden wir die in der Vorlesung erwähnten Resultate über die Konvergenzgeschwindigkeit des *Gradientenverfahrens mit exakter Schrittweitensteuerung* herleiten.

Mit x_k bezeichnen wir die Iterierten dieses Verfahrens. Die Norm $\|\cdot\|$ ist die euklidische Norm und κ steht für die Konditionszahl bezüglich dieser Norm.

a) Zeigen Sie, dass

$$f(x_{k+1}) - f(x^*) = \left(1 - \frac{\|d_k\|^4}{d_k^T A d_k \cdot d_k^T A^{-1} d_k}\right)(f(x_k) - f(x^*))$$

für alle $k \in \mathbb{N}$ gilt. Hierbei bezeichnet $d_k = -\nabla f(x_k)$ die Abstiegsrichtung der Gradientenschritte.

Hinweis: Zeigen Sie zunächst $f(x_{k+1}) - f(x^*) = f(x_k) - f(x^*) - \frac{1}{2}\frac{\|d_k\|^4}{d_k^T A d_k}$.

b) Folgern Sie, dass

$$f(x_{k+1}) - f(x^*) \le \left(\frac{\kappa(A) - 1}{\kappa(A) + 1}\right)^2 (f(x_k) - f(x^*)),$$

$$\|x_{k+1} - x^*\| \le \sqrt{\kappa(A)} \left(\frac{\kappa(A) - 1}{\kappa(A) + 1}\right)^k \|x_0 - x^*\|$$

für alle $k \in \mathbb{N}$ gilt.

Hinweis: Sie dürfen ohne Beweis die sog. *Kantorovich-Ungleichung* verwenden: Für eine symmetrisch positiv definite Matrix C gilt:

$$\frac{\|x\|^4}{x^T Cx \cdot x^T C^{-1}x} \ge \frac{4\lambda_{\min}(C)\lambda_{\max}(C)}{(\lambda_{\min}(C) + \lambda_{\max}(C))^2}, \qquad \text{für alle } x \in \mathbb{R}^n \setminus \{0\}.$$

Hierbei bezeichnen $\lambda_{\min}(C)$ bzw. $\lambda_{\max}(C)$ den kleinsten bzw. größten Eigenwert von C.

(4+3 = 7 Punkte)

Aufgabe 3. Für $A \in \mathbb{R}^{n\times n}$ symmetrisch positiv definit und $b \in \mathbb{R}^n$ zeige man:

a) Der Vektor $x \in \mathbb{R}^n$ minimiert genau dann das Funktional $\phi(x) = \frac{1}{2}(b - Ax)^T A^{-1}(b - Ax)$, wenn er $F(x) = \frac{1}{2}x^T Ax - b^T x$ minimiert.

b) Die Familie von Vektoren $(d_1, \dots, d_n) \subset \mathbb{R}^n$ sei A-konjugiert, d.h. es gilt

$$\langle d_i, d_j\rangle_A := \langle d_i, Ad_j\rangle_2 = \delta_{ij}.$$

Für beliebige $\gamma_k \in \mathbb{R}$, $k = 1, \dots, n$, gilt dann $F(\sum_{k=1}^{n} \gamma_k d_k) = \sum_{k=1}^{n} F(\gamma_k d_k)$.

Interpretieren Sie dieses Resultat.

(1+2 = 3 Punkte)

Programmieraufgabe 1. (30 Punkte)

Bearbeiten Sie die im jupyter notebook bereitgestellte Programmieraufgabe II.

Abgabe der Programmieraufgabe in der Woche 19. bis 23. November