# Übung 2

Ausgabedatum: 21. Oktober 2024
Abgabedatum: 28. Oktober 2024

**Hausaufgabe 2.1** (Eigenschaften von Abstiegsverfahren) 2 + 2 + 2 + 2 + 2 = 10 Punkte

Abstiegsverfahren sind iterative Optimierungsverfahren mit der Eigenschaft, dass die Folge der Funktionswerte $\big(f^{(k)}\big)_{k\in\mathbb{N}_0} \coloneqq \big(f(x^{(k)})\big)_{k\in\mathbb{N}_0}$ der Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ bzgl. der Iterierten $\big(x^{(k)}\big)_{k\in\mathbb{N}_0}$ strikt fällt, also dass $f(x^{(k+1)}) < f(x^{(k)})$ für alle $k \in \mathbb{N}_0$. Ein Vertreter dieser Algorithmen aus dem Skript ist das Gradientenverfahren (Algorithmen 4.5 und 4.11).

Wir wollen zeigen, dass die Iterierten von Abstiegsverfahren nicht gegen lokale Maximierer konvergieren können, und dass für einige interessante Eigenschaften schon "nicht-Aufstieg" der Funktionswertfolge ausreicht.

(a) Es sei $\big(f^{(k)}\big)_{k\in\mathbb{N}_0}$ eine monoton fallende Folge, (also $f^{(k+1)} \leqslant f^{(k)}$) mit Häufungspunkt $f^*$. Zeigen Sie, dass dann bereits die gesamte Folge $\big(f^{(k)}\big)_{k\in\mathbb{N}_0}$ gegen $f^*$ konvergiert (also $f^{(k)} \xrightarrow{k\to\infty} f^*$).

(b) Es seien $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig und $\big(x^{(k)}\big)_{k\in\mathbb{N}_0}$ eine Folge mit zwei Häufungspunkten $x^*$ und $x^{**}$ sowie $f(x^{(k+1)}) \leqslant f(x^{(k)})$. Zeigen Sie, dass dann $f(x^*) = f(x^{**})$.

(c) Es seien $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig und $\big(x^{(k)}\big)_{k\in\mathbb{N}_0}$ eine Folge mit einem Häufungspunkt $x^*$ sowie $f(x^{(k+1)}) \leqslant f(x^{(k)})$. Zeigen Sie, dass $x^*$ nur dann ein strikter lokaler Maximierer von $f$ sein kann, wenn $x^*$ ein isolierter Punkt der Folge ist.

(d) Zeigen Sie, dass der Punkt $x^*$ in Punkt (c) für eine von einem Abstiegsverfahren generierte Folge von Iterierten gar kein lokaler Maximierer – also auch kein nicht strikter Maximierer – sein kann. Können Sie die Bedingung hierfür abschwächen?

(e) Gilt die Aussage in Punkt (d) auch für das Gradientenverfahren, wenn es nach endlich vielen Schritten in einem Punkt $x^*$ mit $\nabla_M f(x^*) = 0$ abbricht?

**Hausaufgabe 2.2** (Exakte Liniensuche für eine quadratische Funktion) 4 + 3 + 4 = 11 Punkte

Die Vorschrift der exakten Liniensuche für eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ entlang einer Abstiegsrichtung $d \in \mathbb{R}^n$ lautet:

$$\text{Bestimme } t_{\min} \text{ so, dass } f(x + t_{\min} d) = \min_{t \geqslant 0} f(x + t\,d),$$

siehe Gleichung (4.2) im Skript. Gegeben sei die quadratische Funktion

$$f(x) = \frac{1}{2} x^\top Q\, x + c^\top x + \gamma$$

mit einer symmetrischen, positiv definiten Matrix $Q \in \mathbb{R}^{n \times n}$, $c \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$, einer Iterierten $x \in \mathbb{R}^n$ und einer Richtung $d \in \mathbb{R}^n$, $d \neq 0$.

(a) Beweisen Sie die Darstellung

$$t_{\min} = -\frac{f'(x)\, d}{d^\top Q\, d}. \tag{$*$}$$

(b) Zeigen Sie, dass im vorkonditionierten Gradientenverfahren für $d = -M^{-1} f'(x)^\top = -\nabla_M f(x)$ die Darstellung

$$t_{\min} = \frac{d^\top M\, d}{d^\top Q\, d} \tag{4.16}$$

aus dem Skript gilt und erklären Sie, wie sich die Schrittweite $t_{\min}$ und die *Korrektur* $\Delta x \coloneqq t_{\min} d$ ändern, wenn Sie $M$ durch $\tilde{M} \coloneqq \alpha\, M$ für $\alpha \in \mathbb{R}, \alpha \neq 0$ ersetzen. Was bedeutet das für die Wahl des Vorkonditionierers bei Verwendung der exakten Schrittweite?

(c) Drücken Sie Gleichung ($*$) nur durch $\varphi(0)$ und $\varphi'(0)$ aus.

**Beachte:** In der Implementierung von Abstiegsverfahren wird die Liniensuchfunktion für gewöhnlich als separater, austauschbarer Parameter an das Verfahren übergeben. Das Verfahren übergibt dann eine Routine zur Auswertung des Schnitts $\varphi(t) \coloneqq f(x + td)$ durch die Funktion entlang der Suchrichtung und ihrer Ableitungen an die Liniensuche. Der Term ($*$) kann dann (weil $Q$ nicht konkret auswertbar vorliegt) nicht ausgewertet werden. Die Werte $\varphi(0)$ und $\varphi'(0)$ werden allerdings außerhalb berechnet und dann nach unten weitergereicht.

**Hausaufgabe 2.3** (Häufungspunkte des Gradientenverfahrens mit spd Hessematrix sind Attraktoren) 14 Punkte

Es seien $f\colon \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und $(x^{(k)})_{k \in \mathbb{N}_0}$ eine vom vorkonditionierten Gradientenverfahren (Algorithmus 4.11) mit Armijo-Backtracking erzeugte Folge.

Zeigen Sie: Ist $x^*$ ein Häufungspunkt der Folge $(x^{(k)})_{k \in \mathbb{N}_0}$ und $f''(x^*)$ positiv definit, dann konvergiert die gesamte Folge gegen $x^*$, und $x^*$ ist ein strikter lokaler Minimierer.

**Zusatzaufgabe 2.4** (Stetige Abh. der Iterierten im Gradientenverfahren vom Startwert) 12 + 2 + 2 = 16 Bonuspunkte

Gegeben sei die quadratische Funktion

$$f(x) = \frac{1}{2} x^\top Q\, x + c^\top x + \gamma$$

mit symmetrischer, positiv definiter Matrix $Q \in \mathbb{R}^{n \times n}$ und $c \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$.

(a) Es seien $(x^{(k)})_{k \in \mathbb{N}_0}$, $(\widetilde{x}^{(k)})_{k \in \mathbb{N}_0}$ zwei vom vorkonditionierten Gradientenverfahren (Algorithmus 4.11) mit *exakter Liniensuche* erzeugte Folgen von Iterierten zu den Startwerten $x^{(0)}$, respektive $\widetilde{x}^{(0)}$. Zeigen Sie, dass für jedes $\varepsilon$ ein $\delta(\varepsilon, x^{(0)}) > 0$ existiert, so dass

$$\left\| x^{(k)} - \widetilde{x}^{(k)} \right\| < \varepsilon \quad \text{für alle } k \in \mathbb{N}_0,$$

wenn

$$\left\| x^{(0)} - \widetilde{x}^{(0)} \right\| < \delta.$$

**Beachte:** Das bedeutet, dass die Iterierten stetig vom Startwert abhängen.

(b) Was bedeutet Punkt (a) für die Anzahl der Iterationen, wenn das Gradientenverfahren mit der Abbruchbedingung $\|\nabla_M f(x^{(k)})\|_M < \text{ATOL}$ terminiert?

(c) Kann die Aussage in Punkt (a) auf allgemeine (nicht-quadratische) Funktionen erweitert werden?

**Zusatzaufgabe 2.5** (Implementierung eines Gradientenverfahrens) 15 Bonuspunkte

Bearbeiten Sie `P1_Gradientenverfahren.ipynb`.

| Bitte reichen Sie Ihre Lösungen als ein PDF auf Mampf ein. |
|---|