Übung 2 (Lösung)

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?

Lösung.

(a) Da die Folge monoton fallend ist, kann sie keine kleineren Werte als den des Häufungspunktes annehmen. Denn gäbe es ein $l_0 \in \mathbb{N}_0$ mit $f^{(l_0)} < f^*$, dann wäre auch $f^{(l)} < f^*$ für alle $l \geq l^0$ und somit lägen für $\delta = (f^* - f^{(l_0)})/2 > 0$ nur endliche viele Folgenglieder in der δ-Umgebung um f^*. Widerspruch zur Häufungspunkteigenschaft. (1 Punkt)

Es gibt also eine monoton fallende Teilfolge $\left(f^{(k^{(l)})}\right)_{l\in\mathbb{N}_0}$, die von oben gegen f^* konvergiert. Für gegebenes $\varepsilon > 0$ gibt es also einen Index $l_0 \in \mathbb{N}_0$, so dass (1 Punkt)

$$f^* \leqslant f^{(k)} \leqslant f^{(k^{(l_0)})} \leqslant f^* + \varepsilon \quad \forall k \geq k^{(l_0)}.$$

(b) Zu $\left(x^{(k)}\right)_{k\in\mathbb{N}_0}$ und den zwei konvergenten Teilfolgen, die gegen x^*, x^{**} konvergieren, existieren die entsprechende Folge der Funktionswerte $\left(f(x^{(k)})\right)_{k\in\mathbb{N}_0}$ mit den jeweiligen Teilfolgen, die auf Grund der Stetigkeit von f gegen die Funktionswerte $f(x^*)$ und $f(x^{**})$ konvergieren. (1 Punkt)

Somit sind $f(x^*)$ und $f(x^{**})$ Häufungspunkte der monoton fallenden Funktionswertfolge. Nach Punkt (a) konvergiert die gesamte Folge gegen diese Werte, die daher übereinstimmen müssen. (1 Punkt)

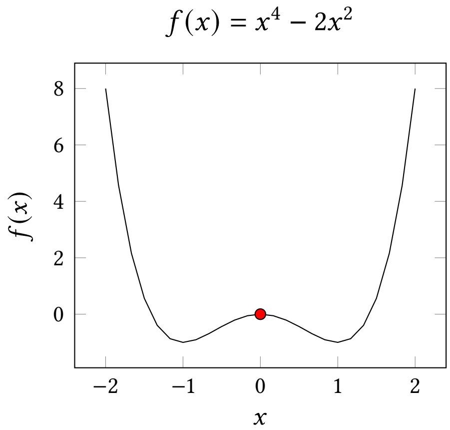
(c) Wenn x^* ein isolierter Punkt der Folge ist, dann kann x^* noch ein strikter Maximierer sein. Beispiel dafür ist $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = -x^2$ und die Folge mit $x^{(k)} = 0$ für alle $k \in \mathbb{N}_0$.

Anderenfalls existiert eine von x^* verschiedene Teilfolge $\left(x^{(k^{(l)})}\right)_{l\in\mathbb{N}_0}$ mit $x^{(k^{(l)})} \xrightarrow{l\to\infty} x^*$. Wie im Beweis von Punkt (b) erhalten wir, dass $f(x^*)$ ein Häufungspunkt der Funktionswertfolge ist. Wie im Beweis von Punkt (a) konvergiert also $f(x^{(k^{(l)})})$ (von oben) gegen $f(x^*)$. Damit kann x^* kein strikter lokaler Maximierer sein. (2 Punkte)

(d) Für Abstiegsverfahren, wo $f(x^{(k+1)}) < f(x^{(k)})$, können keine isolierten Punkte auftreten, die gleichzeitig Häufungspunkte sind, denn die Abstiegseigenschaft garantiert, dass kein Punkt mehr als einmal in der Folge vorkommt. In diesem Fall existiert also eine von x^* verschiedene Teilfolge $\left(x^{(k^{(l)})}\right)_{l\in\mathbb{N}_0}$ mit strikt fallenden Funktionswerten und mit $x^{(k^{(l)})} \xrightarrow{l\to\infty} x^*$. Analog zu Punkt (a) erhält man, dass dann sogar $f(x^*) < f^{(k^{(l)})}$ für alle $l \in \mathbb{N}_0$ gelten muss, so dass x^* gar kein Maximierer sein kann. (1 Punkt)

Hierfür genügt allerdings auch schon, dass in der Funktionswertfolge der Iterierten unendlich oft $f(x^{(k+1)}) < f(x^{(k)})$ gilt, was garantiert, dass jeder Punkt nur endlich oft in der Folge auftreten kann. (1 Punkt)

(e) Bricht das Gradientenverfahren an einem Punkt x^* mit $\nabla_M f(x^*) = 0$ ab, kann x^* natürlich zufällig ein lokales Maximum gewesen sein. Startet man bspw. das Gradientenverfahren für die Funktion $f = x^4 - 2x^2$ in $x = 0$, einem lokalen Maximum, wird es sofort mit $\nabla_M f(0) = 0$ terminieren und das Maximum als letzte Iterierte zurückgeben statt in die lokalen Minima zu konvergieren. (2 Punkte)



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) Zeigen Sie, dass Gleichung (∗) nur durch die Auswertungen $\varphi(0)$ und $\varphi'(0)$ ausgedrückt werden kann, so dass Q nicht explizit vorliegen muss und f' nicht an beliebigen x auswertbar sein muss.

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.

Lösung.

(a) Es gilt

$$\begin{aligned} f(x + t\,d) &= \frac{1}{2}(x + t\,d)^\intercal Q(x + t\,d) + c^\intercal(x + t\,d) + \gamma \\ &= \frac{1}{2}t^2\, d^\intercal Qd + t\,(Qx + c)^\intercal d + \frac{1}{2}x^\intercal Qx + c^\intercal x + \gamma. \end{aligned}$$

Wir bestimmen die exakte Schritte als eindeutigen lokalen Minimierer entlang des Schnitts über die notwendigen Optimalitätsbedingungen erster Ordnung: (1 Punkt)

$$\frac{d}{dt}f(x + t\,d) = t\,d^\intercal Qd + (Qx + c)^\intercal d \overset{!}{=} 0.$$

Das gilt, wenn

$$t_{\min} = -\frac{(Qx + c)^\intercal d}{d^\intercal Qd} = -\frac{f'(x)\,d}{d^\intercal Qd}.$$

(Damit ist $t_{\min}$ positiv, falls d eine Abstiegsrichtung ist.) (2 Punkte)

Wegen $\frac{d^2}{dt^2}f(x + t\,d) = d^\intercal\, Qd > 0$ wird in $t_{\min}$ tatsächlich ein eindeutiges lokales Minimum angenommen. (1 Punkt)

(b) Falls $d = -\nabla_M f(x) = -M^{-1}f'(x)$ gewählt wird, so folgt direkt

$$t_{\min} = \frac{d^\intercal M\,d}{d^\intercal Q\,d}.$$

(1 Punkt)

Ersetzt man M durch $\tilde{M} = \alpha M$, dann ergeben sich in einer Iteration

$$\begin{aligned} d &= -M^{-1} f'(x) & \tilde{d} &= -\tilde{M}^{-1} f'(x) = -\frac{1}{\alpha} M^{-1} f'(x) = \frac{1}{\alpha} d \\ t_{\min} &= \frac{d^\top M\, d}{d^\top Q\, d} & \tilde{t}_{\min} &= \frac{\tilde{d}^\top \tilde{M}\, \tilde{d}}{\tilde{d}^\top Q\, \tilde{d}} = \alpha \frac{d^\top M\, d}{d^\top Q\, d} = \alpha t_{\min} \\ \Delta x &= t_{\min} & \Delta \tilde{x} &= \tilde{t}_{\min} \tilde{d} = t_{\min} d = \Delta x \end{aligned}$$

Bei Umskalierung des Vorkonditionierers wird die Suchrichtung entsprechend antiproportional skaliert und die Schrittweite proportional skaliert, so dass beide Korrekturen gleich sind und das Verfahren also mit beiden Vorkonditionierern die gleichen Iterierten besucht. Der Vorkonditionierer ist also allein zuständig für die Bestimmung der Richtung, während die Liniensuche allein zuständig für die Skalierung ist. (2 Punkte)

Für andere Schrittweitensteuerungen als die exakte bzw. für nichtquadratische Probleme kann das ganz anders aussehen, hier kann sich ein kompliziertes Zusammenspiel aus Schrittweitensteuerung und Vorkonditionierer ergeben.

(c) Zur Verfügung stehen im Gradientenverfahren von vornherein die Werte

$$\begin{aligned} f(x + t\, d)|_{t=0} &= \varphi(0), \\ \frac{d}{dt} f(x + t\, d)|_{t=0} &= \varphi'(0), \end{aligned}$$

wir erhalten also den Zähler des Quotienten direkt als (1 Punkt)

$$-f'(x)\, d = -\varphi'(0),$$

während wir für den Quotienten noch die Auswertung von $\varphi''(0) = d^\top Q d$ gern vermeiden möchten, da die Auswertung der zweiten Ableitung meist kostspielig ist (diese liegt häufig nicht explizit vor, weil sie Inverse enthält). (1 Punkt)

Stattdessen werten wir noch zusätzlich $\varphi(1) = f(x + d)$ aus und erhalten über

$$\frac{1}{2} \varphi''(0) + \varphi'(0) + \varphi(0) = \varphi(1)$$

sofort $\varphi''(0) = 2(\varphi(1) - \varphi(0) - \varphi'(0))$ und somit endgültig

$$t_{\min} = -\frac{\varphi'(0)}{2(\varphi(1) - \varphi(0) - \varphi'(0))}.$$

(2 Punkte)

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.

Lösung.

Sei $(x^{(k^{(l)})})_{l\in\mathbb{N}_0}$ eine gegen x^* konvergente Teilfolge. Hausaufgabe 2.1 liefert, dass $f(x^{(k)}) \searrow f(x^*)$ und aus dem globalen Konvergenzsatz für das Gradientenverfahren (Satz 4.7) folgt, dass $f'(x^*) = 0$. (2 Punkte)

Die hinreichende Bedingung zweiter Ordnung mit quadratischer Wachstumsbedingung (Satz 3.3) liefert für ein $\beta \in (0, \mu)$ die Existenz einer Umgebung $U_\beta(x^*)$ mit dem kleinsten Eigenwert $\mu > 0$ von $f''(x^*)$, so dass

$$f(x) - f(x^*) \geq \frac{\beta}{2}\|x - x^*\|^2 \qquad \forall x \in U_\beta(x^*). \tag{$*$}$$

Damit ist x^* strikter lokaler Minimierer. (2 Punkte)

Es verbleibt die Konvergenz der Folge gegen x^* zu zeigen. Die Strategie für den Nachweis ist, die oben genannte Konvergenz der **Teil**folge $(x^{(k^{(l)})})_{l\in\mathbb{N}_0}$ und der **gesamten** Folge der Funktionswerte $f(x^{(k)})$ mit Hilfe quadratischen Wachstumsbedingung (einer Verknüpfung von Funktionswerten und dem Verhalten von Auswertungspunkten) so zu verknüpfen, dass wir die Konvergenz der **gesamten** Folge $(x^{(k)})_{k\in\mathbb{N}_0}$ gegen x^* erhalten.

Sei dafür ein $\varepsilon > 0$ beliebig gegeben. Wir zeigen, dass es einen Index k_0 gibt, so dass $x^{(k)} \in B_\varepsilon(x^*)$ für alle $k \geqslant k_0$. Dabei können wir o. B. d. A. davon ausgehen, dass $B_\varepsilon(x^*) \subset U_\beta(x^*)$ (ansonsten verkleinern wir ε einfach). Der zentrale Schritt des Nachweises, ist zu zeigen, dass es ein $\delta \in (0, \varepsilon)$ gibt, so dass für jedes $x^{(k)} \in B_\delta(x^*)$ der nächste Schritt $x^{(k+1)} \in U_\beta(x^*)$ ist (also dass wir, wenn wir nah genug an x^* sind, mit dem nächsten Gradientenschritt nicht außerhalb der Umgebung des quadratischen Wachstums landen). (2 Punkte)

Ist so ein δ gegeben, dann gibt es (wegen der Konvergenz der Teilfolge $(x^{(k^{(l)})})_{l\in\mathbb{N}_0}$ gegen x^* und der Funktionswerte $f(x^{(k)})_{k\in\mathbb{N}_0}$ der gesamten Folge von oben gegen $f(x^*)$) einen Index $l_0 \in \mathbb{N}_0$, sodass

$$x^{(k^{(l)})} \in B_\delta(x^*) \quad \text{und} \quad 0 \leqslant f(x^{(k^{(l)})}) - f(x^*) < \frac{\beta}{2} \cdot \delta^2$$

für alle $l \geq l_0$. Wir setzen $k_0 = k^{(l_0)}$, womit dann gilt, dass

$$\begin{aligned}\|x^{(k_0+1)} - x^*\|^2 &\leq \frac{2}{\beta}\Big(f(x^{(k_0+1)}) - f(x^*)\Big) \quad \text{(quadratische Wachstumsbedingung)}\\ &\leq \frac{2}{\beta}\Big(f(x^{(k_0)}) - f(x^*)\Big) \quad \text{(Monotonie der Funktionswerte)}\\ &< \delta^2.\end{aligned}$$

(Die entscheidende Eigenschaft von δ ist hier in der ersten Abschätzung eingeflossen.) Wegen der Monotonie der Funktionswerte gilt dann also wieder, dass

$$x^{(k_0+1)} \in B_\delta(x^*) \quad \text{und} \quad 0 \leqslant f(x^{(k_0+1)}) - f(x^*) < \frac{\beta}{2} \cdot \delta^2.$$

Wiederholte Anwendung dieser Argumentation zeigt, dass $x^{(k)} \in B_\delta(x^*) \subseteq B_\varepsilon(x^*)$ für alle $k \geq k_0$ und damit die Behauptung. (4 Punkte)

Es bleibt also zu zeigen, dass solch ein δ tatsächlich existiert. Wir wissen, dass der Schritt von $x^{(k)}$ zu $x^{(k+1)}$ aus einem (skalierten) Gradientenupdate besteht und wollen also zeigen, dass wir die Norm des Gradientenupdates beschränken können. Das ist kein Problem, denn $f'(x^*) = 0$ und die Abbildung $x \mapsto \nabla_M f(x) = M^{-1} f'(x)$ ist wegen der Stetigkeit von f' ebenfalls stetig. Für unser gegebenes $\varepsilon > 0$ gibt es also ein ein $\delta(\varepsilon)$ so dass

$$\|\nabla_M f(x)\| = \|\nabla_M f(x) - \nabla_M f(x^*)\| < \frac{\varepsilon}{2s} \quad \forall x \in B_\delta(x^*),$$

wobei wir $\delta(\varepsilon)$ o. B. d. A. kleiner als $\frac{\varepsilon}{2}$ wählen können. Damit ist dann

$$\begin{aligned}\|x^{(k+1)} - x^*\| &\leq \|x^{(k+1)} - x^{(k)}\| + \|x^{(k)} - x^*\|\\ &= \|-t^{(k)}\, \nabla_M f(x^{(k)})\| + \|x^{(k)} - x^*\|\\ &= t^{(k)} \|\nabla_M f(x^{(k)})\| + \|x^{(k)} - x^*\|\\ &\leqslant s\|\nabla_M f(x^{(k)})\| + \|x^{(k)} - x^*\|\\ &< \frac{\varepsilon}{2} + \delta < \varepsilon\end{aligned}$$

Das heißt, dass $x^{(k+1)} \in B_\varepsilon(x^*) \subseteq U_\beta(x^*)$. Wenn wir eine Iterierte in $B_\delta(x^*) \subseteq U_\beta(x^*)$ haben, dann gilt sowohl an dieser als auch an der folgenden Iterierten die quadratische Wachstumsbedingung, was genau die geforderte Bedingung für δ ist. (4 Punkte)

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

Gegeben sei die quadratische Funktion

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

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 ε 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 $\|\nabla_M f(x^{(k)})\|_M <$ ATOL terminiert?

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

Lösung.

(a) Wir gehen zweistufig vor und zeigen:

(i) Die Aussage gilt für eine beliebige, endliche Anzahl von Iterierten, d. h. für gegebenes $k_0 \in \mathbb{N}_0$ und $\varepsilon > 0$ existiert ein $\delta(\varepsilon, k_0) > 0$, so dass

$$\left\| x^{(k)} - \widetilde{x}^{(k)} \right\| < \varepsilon \quad \text{für alle } k \in \{0, \ldots, k_0\},$$

wenn

$$\left\| x^{(0)} - \widetilde{x}^{(0)} \right\| < \delta(\varepsilon, k_0).$$

Das bedeutet, dass wir immer endlich viele Schritte kontrolliert kriegen. Den Beweis hierzu führen wir induktiv mittels der Dreiecksungleichung.

(ii) Es existiert ein $\delta(\varepsilon)$ und ein endlicher Index $k_0(x^{(0)}) \in \mathbb{N}_0$, so dass

$$\left\| x^{(k)} - \widetilde{x}^{(k)} \right\| < \varepsilon \quad \text{für alle } k \geq k_0(x^{(0)}),$$

wenn

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

Das bedeutet, dass wir auch nur endlich viele Schritte kontrollieren müssen. Den Beweis hierzu führen wir über die Konvergenzaussage des Gradientenverfahrens bei quadratischen Funktionen.

Kombiniert man die beiden Schritte, so folgt die Behauptung sofort. (1 Punkt)

Wir beginnen mit dem Beweis von Punkt (i) per Induktion. Sei also $\varepsilon > 0$ gegeben. Für den Induktionsbeginn mit $k_0 = 0$ ist die Aussage einfach, denn mit $\delta(\varepsilon, 0) = \varepsilon$ folgt die Aussage sofort. (1 Punkt)
Für den Induktionsschritt sei nun $k_0 \in \mathbb{N}_0$ gegeben, so dass die Teilbehauptung gilt. Dann ist

$$\begin{aligned}\left\|x^{(k_0+1)} - \widetilde{x}^{(k_0+1)}\right\| &= \left\|x^{(k_0)} + t^{(k_0)}d^{(k_0)} - \left(\widetilde{x}^{(k_0)} + \widetilde{t}^{(k_0)}\widetilde{d}^{(k_0)}\right)\right\| \\ &\le \left\|x^{(k_0)} - \widetilde{x}^{(k_0)}\right\| + \left\|t^{(k_0)}d^{(k_0)} - \widetilde{t}^{(k_0)}\widetilde{d}^{(k_0)}\right\|\end{aligned}$$

Nach Induktionsvoraussetzung gibt es ein $\delta(\varepsilon/2, k_0)$, so dass $\left\|x^{(k)} - \widetilde{x}^{(k)}\right\| < \frac{\varepsilon}{2}$ für alle $k \in \{0, \dots, k_0\}$, wenn $\left\|x^{(0)} - \widetilde{x}^{(0)}\right\| < \delta(\varepsilon, k_0)$. Die Abbildung $(\mathbb{R}^{n\times n}, \mathbb{R}^n) \ni (A, d) \mapsto d^\intercal A d$ ist außerdem stetig in d und A, damit ist

$$t^{(k_0)} = \frac{d^{(k_0)\intercal} M d^{(k_0)}}{d^{(k_0)\intercal} Q d^{(k_0)}}$$

stetig in $d^{(k_0)}$. Weiterhin ist $d^{(k_0)} = -M^{-1}\nabla f(x^{(k_0)}) = -M^{-1}(Qx^{(k_0)} + c)$ auch stetig in $x^{(k_0)}$. Damit gibt es also ein $\rho(\frac{\varepsilon}{2}, k_0) > 0$, so dass

$$\left\|t^{(k_0)}d^{(k_0)} - \widetilde{t}^{(k_0)}\widetilde{d}^{(k_0)}\right\| < \frac{\varepsilon}{2}$$

wenn $\left\|x^{(0)} - \widetilde{x}^{(0)}\right\| < \rho(\frac{\varepsilon}{2}, k_0)$. Hierbei wurde noch einmal die Induktionsvoraussetzung ausgenutzt. Setzt man $\delta(\varepsilon, k_0 + 1) = \min(\delta(\varepsilon/2, k_0), \rho(\frac{\varepsilon}{2}, k_0))$ zeigt das also den Induktionsschritt und damit Punkt (i). (4 Punkte)

An dieser Stelle ist nun klar, dass wir zu festem ε für jedes $k_0 \in \mathbb{N}_0$ ein entsprechendes δ für die Behauptung erhalten, nicht aber, dass es ein δ gibt, das gleichmäßig für alle k_0 verwendet werden kann (wir bilden ja in jedem Schritt ein Minimum). Bleibt also Punkt (ii) zu zeigen.

Sei dazu $\varepsilon > 0$ gegeben. Wegen Satz 4.16 des Skripts konvergieren die Folgen der Iterierten des Gradientenverfahrens $(x^{(k)})_{k\in\mathbb{N}_0}$, $(\widetilde{x}^{(k)})_{k\in\mathbb{N}_0}$ gegen den eindeutigen strikten globalen Minimierer $x^* = -Q^{-1}c$. Für ein $C < 1$ liefert der Satz sogar die Abschätzungen

$$\begin{aligned}
\|x^{(k)} - x^*\|_Q &\leqslant C^k \|x^{(0)} - x^*\|_Q < \frac{\varepsilon}{4} \leq \frac{\varepsilon}{2},\\
\|\widetilde{x}^{(k)} - x^*\|_Q &\leqslant C^k \|\widetilde{x}^{(0)} - x^*\|_Q\\
&\leqslant \|\widetilde{x}^{(0)} - x^{(0)}\|_Q + C^k \|x^{(0)} - x^*\|_Q\\
&< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}
\end{aligned}$$

für alle $k \geq k_0(\varepsilon, x^{(0)})$ mit einem $k_0(\varepsilon, x^{(0)})$ und wenn $\|\widetilde{x}^{(0)} - x^{(0)}\|_Q \leq \frac{\varepsilon}{4}$. Es ist dabei ausschlaggebend, dass $k^{(0)}$ nicht von $\widetilde{x}^{(0)}$ abhängt, sondern nur von $x^{(0)}$. (4 Punkte)

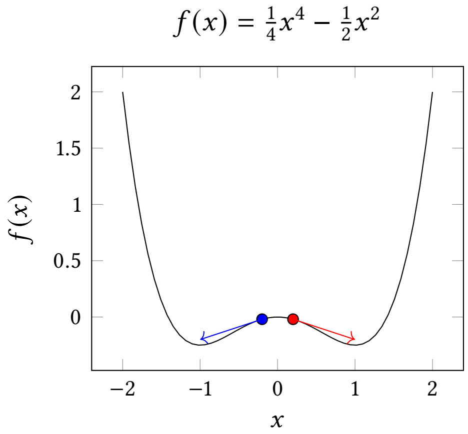
Somit gilt abschließend, dass

$$\|x^{(k)} - \widetilde{x}^{(k)}\|_Q \leq \|x^{(k)} - x^*\|_Q + \|x^* - \widetilde{x}^{(k)}\|_Q < \varepsilon \qquad (*)$$

wenn $k \geq k_0(\varepsilon, x^{(0)})$ und $\|\widetilde{x}^{(0)} - x^{(0)}\|_Q \leq \frac{\varepsilon}{4}$. Wegen der Äquivalenz der Energienorm und der Euklidischen Norm gilt Gleichung $(*)$ nach Anpassung des ursprünglichen ε und der Umgebung um $x^{(0)}$ auch für die Euklidische Norm. (2 Punkte)

(b) Da der Ausdruck in dem Abbruchkriterium stetig von der Iterierten abhängt wissen wir, dass wir das Problem leicht stören können und trotzdem mit der gleichen Anzahl an Iterationen das Abbruchkriterium erfüllen werden. (2 Punkte)

(c) Nein, die Aussage kann nicht verallgemeinert werden, denn im voll nichtlinearen Fall können äußerst "unstetige" Verhalten beobachtet werden. Ein Beispiel dafür ist die Funktion $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = 1/4x^4 - 1/2x^2$ mit $f'(x) = x^3 - x = x(x^2 - 1)$. Startet man das Gradientenverfahren mit einem beliebigen Vorkonditionierer mit den Startwerten $x^{(0)} = -\frac{\delta}{2}$ und $\widetilde{x}^{(0)} = \frac{\delta}{2}$ dann startet das Gradientenverfahren in die jeweils anderen Täler und konvergiert dort gegen die Minimierer -1 bzw. 1, sofern die Liniensuche "vernünftig" eingestellt ist. (2 Punkte)



Zusatzaufgabe 2.5 (Implementierung eines Gradientenverfahrens) 15 Bonuspunkte

Bearbeiten Sie `P1_Gradientenverfahren.ipynb`.

Lösung.

Siehe `P1_Gradientenverfahren_solutions.ipynb`. (15 Punkte)

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