Nichtlineare Optimierung - 3. Hausaufgabe

Claudia Wohlgemuth	*366323*
Thorsten Lucke	*363089*
Felix Thoma	*358638*

Tutor: Mathieu Rosière

8. Juni 2017

3.1	3.2	3.3	3.4	Σ

Anmerkungen:

Aufgabe 3.1

Obwohl die in der Aufgabenstellung formulierte vermeintliche Charakterisierung der Abstiegsrichtung außerordentlich intuitiv erscheint, ist sie im Allgemeinen nicht mit dem in der Vorlesung eingeführten Begriff der Abstiegsrichtung vereinbar. Das folgende Gegenbeispiel dient als Untermauerung dieser kühnen Behauptung.

Für $n = 1$ bemühen wir die Sinusfunktion $\sin\colon \mathbb{R} \to \mathbb{R}, x \mapsto \sin(x)$ und betrachten den Punkt $x_0 = \pi/2$. Dann ist mit $d = 1$ wegen

$$\nabla \sin(\pi/2) \cdot d = \cos(\pi/2) \cdot 1 = 0 \not< 0$$

keine Abstiegsrichtung im Sinne der Definition (4.1.1) gegeben. Setzen wir allerdings $\widetilde{\sigma} := \pi/2$, so ist für alle $\sigma \in (0, \widetilde{\sigma}]$ die geforderte Abschätzung

$$0 \leq \sin(x + \sigma d) < 1 = \sin(x)$$

dennoch erfüllt. Es kann also keine Äquivalenz gelten.

Aufgabe 3.2

Es sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, streng konvex und koerziv. Damit hat f ein eindeutiges Minimum $\bar{x} \in \mathbb{R}^n$ und $\nabla f(x) = 0$ genau dann, wenn $x = \bar{x}$ ist. (†) Weiter genüge die im Abstiegsverfahren gewählte Richtung $x \mapsto d(x)$ folgender Bedingung: Es gibt Konstanten $c_1, c_2 > 0$ und $p > -1$ mit

$$\frac{|\langle \nabla f(x), d(x)\rangle|}{\|\nabla f(x)\|\|d(x)\|} \geq \min\{c_1, c_2\|\nabla f(x)\|^p\} \tag{$\star$}$$

für alle $x \in \mathbb{R}^n$. Weiter seien die Schrittweiten $(x, d(x)) \mapsto \sigma(x)$ effizient gewählt, d.h. es gibt ein $c_3 > 0$ mit

$$f(x + d(x)\sigma(x)) \leq f(x) - c_3 \left(\frac{\langle \nabla f(x), d(x)\rangle}{\|d(x)\|}\right)^2 \tag{$\star\star$}$$

für alle $x \in \mathbb{R}^n$.

Sei $x^0 \in \mathbb{R}^n$ beliebig. Wir zeigen zunächst, dass die Folge der Gradienten $(\nabla f(x^k))_{k\in\mathbb{N}}$ des Verfahrens gegen Null konvergiert.

Wegen

$$c_3 \left(\frac{\langle \nabla f(x^k), d(x^k)\rangle}{\|d(x^k)\|}\right)^2 \geq 0$$

und mit ($\star\star$) ist die Folge $(f(x^k))_{k\in\mathbb{N}}$ monoton fallend und aufgrund der Beschränktheit von f auch konvergent. Von ($\star\star$) ausgehend, erhalten wir zudem für $k \in \mathbb{N}$

$$c_3 \left(\frac{\langle \nabla f(x^k), d(x^k)\rangle}{\|d(x^k)\|}\right)^2 \leq f(x^k) - f(x^{k+1}) \overset{k\to\infty}{\to} 0.$$

Setzen wir nun ($\star$) ein, so ergibt sich

$$\begin{aligned} c_3\|\nabla f(x^k)\|^2(\min\{c_1, c_2\|\nabla f(x^k)\|^p\})^2 &\leq c_3 \left(\frac{\langle \nabla f(x^k), d(x^k)\rangle}{\|d(x^k)\|}\right)^2 \\ &\leq f(x^k) - f(x^{k+1}) \overset{k\to\infty}{\to} 0. \end{aligned}$$

Damit gilt $c_3c_1^2\|\nabla f(x^k)\|^2 \to 0$ oder $c_3c_2^2\|\nabla f(x^k)\|^{2+2p} \to 0$ für $k \to \infty$, was genau dann erfüllt ist, falls $(\nabla f(x^k))_k$ gegen Null konvergiert.

Aus der Koerzivität und der Stetigkeit von f folgt, dass die Niveaumenge $N(f, f(x_0))$ kompakt ist. Damit hat die jede Teilfolge der Folge $(x^k)_{k\in\mathbb{N}}$ eine konvergente Teilfolge $(x^{k_{l_m}})_{m\in\mathbb{N}}$, deren Grenzwert wir mit $\tilde{x}_l \in \mathbb{R}^n$ bezeichnen wollen. Mit der Stetigkeit der Ableitung von f folgt

$$\nabla f(\tilde{x}_l) = \lim_{m\to\infty} \nabla f(x^{k_{l_m}}) = 0.$$

Mit (†) folgt, dass $\tilde{x}_l$ der (eindeutige) Minimierer $\bar{x}$ von f ist. Insgesamt konvergiert also eine Teilfolge jeder Teilfolge gegen $\bar{x}$ und somit konvergiert die ganze Folge.

Aufgabe 3.3

Gegeben sei eine stetig differenzierbare Funktion $f\colon R^n \to \mathbb{R}$ und ein Startwert $x^0 \in \mathbb{R}^n$.

(i) Es besitze f einen Minimierer; insbesondere ist f dann beschränkt. Weiter sei $(x^k)_{k\in\mathbb{N}}$ eine durch ein Abstiegsverfahren mit effizienter Schrittweite erzeugte Folge, d.h. es gibt ein $c > 0$ mit

$$f(x^{k+1}) \leq f(x^k) - c\left(\frac{\langle \nabla f(x^k), d(x^k)\rangle}{\|d(x^k)\|}\right)^2.$$

Damit ist die Folge $(f(x^k))_{k\in\mathbb{N}}$ monoton fallend und beschränkt, weil f beschränkt ist. Somit ist $(f(x^k))_{k\in\mathbb{N}}$ konvergent. Durch Kontraposition folgt die Behauptung: Ist $(f(x^k))_{k\in\mathbb{N}}$ divergent, so hat f keinen Minimierer.

(ii) Sei $(x^k)_{k\in\mathbb{N}}$ eine durch ein Abstiegsverfahren mit effizienter Schrittweite erzeugte Folge und sei x ein Häufungspunkt von $(x^k)_{k\in\mathbb{N}}$, sowie $(x^{k_l})_{l\in\mathbb{N}}$ die gegen x konvergente Teilfolge. Weil die Schrittweite effizient gewählt ist und für ein Abstiegsverfahren stets $\langle \nabla f(x), d(x)\rangle < 0$ gilt, ist wegen

$$f(x^{k+1}) \leq f(x^k) - c\left(\frac{\langle \nabla f(x^k), d(x^k)\rangle}{\|d(x^k)\|}\right)^2.$$

die Bildfolge $(f(x^k))_{k\in\mathbb{N}}$ *streng* monoton fallend. Insbesondere konvergiert dann wegen der Stetigkeit von f auch $(f(x^{k_l}))_{l\in\mathbb{N}}$ streng monoton von *oben* gegen x, d.h. für jedes $\varepsilon > 0$ gibt es ein $l \in \mathbb{N}$, sodass $0 < f(x^{k_l}) - f(x) < \varepsilon$ ist. Somit kann x kein lokaler Maximierer sein.

(iii) Wir zeigen, dass diese Aussage nicht wahr sein kann. Wir betrachten dazu für $n = 1$ die Funktion $f = \sin(\sqrt{\frac{3\pi}{2}}\cdot)$ und den Startwert $x_0 = 0$. Dann ist

$$f(x^1) = f(x_0 - \nabla f(x^0)) = \sin\left(\sqrt{\frac{3\pi}{2}}\left(0 - \sqrt{\frac{3\pi}{2}}\right)\right) = 1 > 0 = \sin(x^0) = f(x).$$

Damit gilt insbesondere für alle $c > 0$

$$f(x_1) > f(x^0) - c\left(\frac{\langle \nabla f(x_0), \nabla f(x_0)\rangle}{\|\nabla f(x_0)\|}\right)^2,$$

d.h. $\sigma = 1$ ist keine effiziente Schrittweite.

(iv) Auch diese Aussage trifft nicht zu. Wir untersuchen im Fall $n = 1$ die Funktion $f\colon x \mapsto x^3$ und $x_0 = -1$. Dann ist $x_1 = -1 - 1 \cdot 2(-1)^2 = -3$ und es gilt

$$|f(x_0)| = |2(-1)^2| < |2(-3)^2| = |f(x_1)|.$$

Aufgabe 3.4

Wir übernehmen die Notationen und Voraussetzungen aus der Aufgabenstellung. Da mit f ein quadratisches Optimierungsproblem gegeben ist, setzen wir analog zur Bemerkung auf der Seite 36 im Skript

$$\varphi(\sigma) = f(x + \sigma d) = \frac{1}{2}\langle A(x + \sigma d), x + \sigma d\rangle + \langle b, x + \sigma d\rangle + c$$

und lösen die Gleichung $\varphi'(\sigma) = 0$.

Berechnen wir zunächst die Ableitung von φ nach σ. Mit den üblichen Regeln der Differenziation erhalten wir

$$\varphi'(\sigma) = \frac{\langle Ax, d\rangle + \langle Ad, x\rangle}{2} + \sigma\langle Ad, d\rangle + \langle b, d\rangle.$$

Die bereits erwähnte Forderung $\varphi'(\sigma) = 0$ führt zur optimalen Schrittweite

$$\sigma = -\frac{\frac{\langle Ax,d\rangle + \langle Ad,x\rangle}{2} + \langle b, d\rangle}{\langle Ad, d\rangle}.$$