Nichtlineare Optimierung - 2. Hausaufgabe

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

Tutor: Mathieu Rosière, Di 8-10 Uhr

13. Mai 2017

2.1	2.2	2.3	2.4	Σ

Anmerkungen:

Aufgabe 2.1

Aufgabe 2.2

Gegeben sei $z \in \mathbb{R}^n$. Wir untersuchen zu $\lambda \geq 0$

$$f_\lambda \colon \mathbb{R}^n \to \mathbb{R}, \quad f_\lambda(x) = \frac{1}{2}\|x - z\|_2^2 + \lambda\|x\|_1 = \frac{1}{2}\sum_{k=1}^{n} |x_k - z_k|^2 + \lambda \sum_{k=1}^{n} |x_k|.$$

(i) Wir zeigen, dass für jedes $\lambda \geq 0$ ein eindeutiges $x_\lambda = \arg\min_{x \in \mathbb{R}^n} f_\lambda(x)$ gibt. Sei also $\lambda \geq 0$ beliebig. Wir zeigen, dass f_λ koerziv und streng konvex ist. Mit

$$f_\lambda(x) = \frac{1}{2}\|x-z\|_2^2+\lambda\|x\|_1 \geq \frac{1}{2}\|x-z\|_2^2 \geq \|x\|_2^2-\|x\|_2\|z\|_2+\|z\|_2^2 \geq \|x\|_2(\|x\|_2-\|z\|_2)$$

folgt die Koerzitivität von f_λ.
Für den Nachweis der Konvexität seien $x, y \in \mathbb{R}^n$, $x \neq y$ und $t \in (0,1)$. Mit der Parallelogrammgleichung gilt

$$\begin{aligned}
f_\lambda(tx + (1-t)y) \quad &= \frac{1}{2}\|tx + (1-t)y - z\|_2^2 + \lambda\|tx + (1-t)y\|_1 \\
&\leq \frac{1}{2}\|tx + (1-t)y - tz - (1-t)z\|_2^2 + \lambda(t\|x\|_1 + (1-t)\|y\|_1) \\
&= \frac{1}{2}\|t(x-z) + (1-t)(y-z)\|_2^2 + \lambda(t\|x\|_1 + (1-t)\|y\|_1) \\
&= \|t(x-z)\|_2^2 + \|(1-t)(y-z)\|_2^2 \\
&\quad - \frac{1}{2}\|t(x-z) - (1-t)(y-z)\|_2^2 + \lambda(t\|x\|_1 + (1-t)\|y\|_1) \\
&\leq \|t(x-z)\|_2^2 + \|(1-t)(y-z)\|_2^2 + \lambda(t\|x\|_1 + (1-t)\|y\|_1) \\
&= t^2\|x-z\|_2^2 + \lambda t\|x\|_1 + (1-t)^2\|(y-z)\|_2^2 + \lambda(1-t)\|y\|_1 \\
&< \frac{t}{2}\|x-z\|_2^2 + \lambda t\|x\|_1 + \frac{(1-t)}{2}\|(y-z)\|_2^2 + \lambda(1-t)\|y\|_1 \\
&= tf_\lambda(x) + (1-t)f_\lambda(y).
\end{aligned}$$

Insgesamt ist f_λ also streng konvex und somit insbesondere stetig sowie koerziv und besitzt damit einen eindeutigen Minimierer x_λ.

Aufgabe 2.3

Sei $A\colon \mathbb{R}^n \to \mathbb{R}^m$ linear und $b \in \mathbb{R}^m$. Wir betrachten die Abbildung $f\colon \mathbb{R}^n \to \mathbb{R}$ mit

$$f(x) = \frac{1}{2}\|Ax - b\|^2 = \frac{1}{2}\langle Ax - b, Ax - b\rangle = \frac{1}{2}(\langle A^T Ax, x\rangle - 2\langle A^T x, b\rangle + \langle b, b\rangle).$$

(i) Dann ist die Ableitung in $x \in \mathbb{R}^n$, wie aus Analysis II bekannt, gegeben durch

$$f'(x) = A^T Ax - A^T b$$

gegeben. Damit ergibt sich der Gradient

$$\nabla f(x) = f'(x)^T = x^T A^T A - b^T A^T.$$

(ii) Ist $\bar{x} \in \mathbb{R}^n$ ein Minimierer von f, dann ist insbesondere die notwendige Optimalitätsbedingung erfüllt, d.h. es gilt $f'(\bar{x}) = 0$ und somit $A^T A\bar{x} = A^T b$. Wie aus Numerik bekannt, ist $A^T A$ eine symmetrische, positiv semi-definite Matrix. Wegen $f''(x) = A^T A$ für alle $x \in \mathbb{R}^n$ ist f also konvex. Damit ist die Erfüllung der notwendigen Optimalitätsbedingung bereits hinreichend für Minimalität in einem Punkt.

(iii) Nach (i) genügt es zu zeigen, dass $A^T b \in \operatorname{im} A^T A$ ist. Wir zeigen sogar $\operatorname{im} A^T = \operatorname{im} A^T A$. Dabei ist die Inklusion $\operatorname{im} A^T \supset \operatorname{im} A^T A$ klar. Für $x \in \mathbb{R}^n$ gilt weiterhin

$$\langle A^T Ax, x\rangle = \langle Ax, Ax\rangle$$

d.h. $x \in \ker A^T A$ genau dann, wenn $x \in \ker A$ ist, kurz $\ker A^T A = \ker A$. Damit ist

$$\dim A^T A = \dim A = \dim A^T.$$

Daher muss bereits $\operatorname{im} A^T = \operatorname{im} A^T A$ gelten, d.h. es gibt eine Lösung $\bar{x} \in \mathbb{R}^n$ von $A^T Ax = A^T b$.

(iv) Wenn A injektiv ist, so ist jeder Eigenwert ungleich Null. Das gilt dann natürlich auch für A^T. Insgesamt sind also A und A^T bijektiv. Somit gibt es genau ein $\bar{x} = A^{-1}b \in \mathbb{R}^n$, das die notwendige Optimalitätsbedingung $A^T Ax = A^T b$ erfüllt.

Aufgabe 2.4

Gegeben sei eine stetig differenzierbare Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ und ein Punkt $\bar{x} \in \mathbb{R}^n$. Für $d \in \mathbb{R}^n$ definieren wir

$$g_d\colon \mathbb{R} \to \mathbb{R}, \quad g_d(t) = f(\bar{x} + td).$$

(i) Sei f konvex und besitze für jedes $d \in \mathbb{R}^n$ die Funktion g_d in $\bar{x} \in \mathbb{R}$ ein lokales Minimum bei $t = 0$. Mit f ist auch g_d für $d \in \mathbb{R}^n$ differerenzierbar und mit der notwendigen Optimalitätsbedingung folgt

$$0 = g_d'(0) = \lim_{t\to 0} \frac{g_d(t) - g(0)}{t} = \lim_{t\to 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} = f'(\bar{x}, d).$$

Da f stetig differenzierbar ist, so ist die Ableitung in einem Punkt durch die partiellen Ableitungen gegeben. Insgesamt ist also $f'(\bar{x}) = 0$. Zusammen mit der Konvexität folgt bereits, dass $\bar{x}$ ein Minimierer von f ist.

(ii) Wir betrachten die stetig differenzierbare Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R}, \quad f(x, y) = (y - x^2)(y - 2x^2).$$

Für $d = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}\backslash\{0\}$ ist

$$g_d(t) = (ty - t^2x^2)(ty - 2t^2x^2).$$

Für den Fall $x = 0$ und $y \neq 0$ vereinfacht sich die Funktion sogar zu

$$g_d(t) = t^2y^2 > 0 = g(0) \text{ für } t \neq 0$$

und falls $y = 0$ und $x \neq 0$ zu

$$g_d(t) = 2t^4x^4 > 0 = g(0) \text{ für } t \neq 0.$$

Andernfalls ist für $|t| < \frac{2|y|}{x^2}$ auch

$$g_d(t) > 0 = g_d(0) \text{ mit } t \neq 0.$$

Insgesamt ist $t = 0$ für jedes $d \in \mathbb{R}^2\backslash\{0\}$ ein lokaler Minimierer von g_d. Trotzdem ist

$$f(0, 0) = 0 > -1 = f(1, 1).$$