Nichtlineare Optimierung - 2. Hausaufgabe

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

Tutor: Mathieu Rosière

17. Mai 2017

2.1	2.2	2.3	2.4	Σ

Anmerkungen:

Aufgabe 2.1

Es sei $f : \mathbb{R}^n \to \mathbb{R}$ konvex und $x \in \mathbb{R}^n$. Für $h \in \mathbb{R}^n$ bezeichne $f'(x,h)$ die Richtungsableitung von f im Punkt x in Richtung h und weiter definieren wir

$$g_h : \mathbb{R} \to \mathbb{R}, \quad g_h(t) = f(x + th).$$

(i) Sei $h \in \mathbb{R}^n$ beliebig, aber fest gewählt. Mit der Konvexität von f folgt auch die Konvexität von g_h; betrachte dazu für $t_1, t_2 \in \mathbb{R}$, $\lambda \in [0,1]$

$$\begin{aligned} g_h(\lambda t_1 + (1-\lambda)t_2) &= f((\lambda + (1-\lambda))x + (\lambda t_1 + (1-\lambda)t_2)h) \\ &= f(\lambda(x + t_1 h) + (1-\lambda)(x + t_2 h)) \\ &\leq \lambda f(x + t_1 h) + (1-\lambda) f(x + t_2 h) = \lambda g_h(t_1) + (1-\lambda) g_h(t_2). \end{aligned}$$

Dann gilt für $t, t_1, t_2 \in \mathbb{R}$ mit $t_1 < t < t_2$ wie aus Analysis I bekannt

$$\frac{g(t) - g(t_1)}{t - t_1} \leq \frac{g(t_2) - g(t_1)}{t_2 - t_1} \leq \frac{g(t_2) - g(t)}{t_2 - t}.$$

Diese Ungleichung zeigt, dass der Differenzenquotient monoton wachsend ist (fixiere dazu jeweils t_1, t_2 beziehungsweise t). Damit ist insbesondere $\frac{g(t)-g(0)}{t}$ für $t \searrow 0$ monoton fallend und durch $\frac{g(0)-g(-1)}{1}$ nach unten beschränkt. Somit existiert

$$\lim_{t \searrow 0} \frac{g(t) - g(0)}{t} = \lim_{t \searrow 0} \frac{f(x + th) - f(x)}{t} = f'(x,h).$$

(ii) Sei $h \in \mathbb{R}^n$ und $\lambda > 0$, dann gilt mit $s = \lambda t$

$$\begin{aligned} f'(x, \lambda h) &= \lim_{t \to 0} \frac{f(x + t\lambda h) - f(x)}{t} = \lambda \lim_{t \to 0} \frac{f(x + t\lambda h) - f(x)}{\lambda t} \\ &= \lambda \lim_{s \to 0} \frac{f(x + sh) - f(x)}{s} = \lambda f'(x,h). \end{aligned}$$

(iii) Seien $h_1, h_2 \in \mathbb{R}^n$. Mit (ii) gilt dann

$$\begin{aligned} f'(x, h_1 + h_2) &= \lim_{t \to 0} \frac{f(x + t(h_1 + h_2)) - f(x)}{t} \\ &= \lim_{t \to 0} \frac{f((\frac{1}{2} + \frac{1}{2})x + \frac{1}{2}2th_1 + \frac{1}{2}2th_2)) - (\frac{1}{2} + \frac{1}{2})f(x)}{t} \\ &\leq \lim_{t \to 0} \frac{\frac{1}{2}f(x + 2th_1) - \frac{1}{2}f(x) + \frac{1}{2}f(x + 2th_2) - \frac{1}{2}f(x)}{t} \\ &= \frac{1}{2}f'(x, 2h_1) + \frac{1}{2}f'(x, 2h_2) \\ &= f'(x, h_1) + f'(x, h_2). \end{aligned}$$

(iv) Als Gegenbeispiel untersuchen wir die euklidische Norm $|| \cdot ||: \mathbb{R}^2 \to \mathbb{R}$ und ihre Richtungsableitung in Null. Dann ist

$$f'(0, e_1) = 1 \quad \text{und} \quad f'(0, e_2) = 1 \quad \text{sowie} \quad f'(0, e_1 + e_2) = \sqrt{2}.$$

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 es 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 koerzitiv 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.

Außerdem ist f_λ Summe der streng konvexen Funktion $n : x \mapsto \|x-z\|_2^2$ und der konvexen Funktion $x \mapsto \|x\|_1$. Bemerke dazu, dass $n''(x) = \mathrm{id}_{\mathbb{R}^n}$ positiv definit und n daher streng konvex ist und dass für alle $x, y \in \mathbb{R}^n$ und $t \in [0,1]$ wegen der Dreiecksungleichung

$$\|tx + (1-t)y\|_1 \leq t\|x\|_1 + (1-t)\|y\|_1$$

gilt. Also ist f_λ selbst streng konvex und somit insbesondere stetig. Aus der Stetigkeit und Koerzitivität folgt, dass f_λ einen eindeutigen Minimierer x_λ besitzt.

(ii) Nach (i) ist der Minimierer $x_\lambda = \underset{y \in \mathbb{R}^n}{\operatorname{argmin}} f_\lambda(x)$ eindeutig bestimmt. Löst ein $x \in \mathbb{R}^n$ nun die Variationsungleichung, gilt also $f'(x,h) \geq 0$ für alle $h \in \mathbb{R}^n$, so gilt $x = x_\lambda$.

Dann gilt insbesondere $f'(x_\lambda, h) \geq 0$ für jeden k-ten Einheitsvektoren $e_k = (\delta_{ik})_{i=1,\ldots,n}$ und ihre additiven Inversen $-e_k = (-\delta_{ik})_{i=1,\ldots,n}$.

Mit $x_\lambda = (x_k)_{k=1,\ldots,n}$ gilt für $k = 1, \ldots, n$ also:

$$f'_\lambda(x_\lambda, e_k) = \lim_{t \searrow 0} \frac{f(x_\lambda + te_k) - f(x_\lambda)}{t} = (x_k - z_k) + \lambda \operatorname{sgn}(x_k) \geq 0$$

und

$$f'_\lambda(x_\lambda, -e_k) = \lim_{t \searrow 0} \frac{f(x_\lambda - te_k) - f(x_\lambda)}{t} = -((x_k - z_k) + \lambda \operatorname{sgn}(x_k)) \leq 0.$$

Aus beiden Ungleichungen ergibt sich die Gleichung

$$x_k - z_k + \lambda \operatorname{sgn}(x_k) = 0$$

und durch Umstellen die implizite Identität

$$x_k = z_k - \lambda \operatorname{sgn}(x_k).$$

Ist $z_k > \lambda$, so folgt $\operatorname{sgn}(x_k) = 1$, da für $x_k < 0$ der Widerspruch $0 > x_k = z_k + \lambda > 0$ entsteht. Ist $z_k < \lambda$, so folgt analog $\operatorname{sgn}(x_k) = -1$. Für $z_k = \lambda$ gilt aus ähnlichen Gründen $\operatorname{sgn}(x_k) = 0$, also $x_k = 0$. Daher erhalten wir sogar die explizite Formel

$$x_k = z_k - \lambda \operatorname{sgn}(z_k - \lambda).$$

Bemerkung:

Sei nun $|z_k| \leq \lambda$, d.h. $z_k < \lambda$ und $-z_k < \lambda$. Aus $z_k < \lambda$ folgt wie oben $\operatorname{sgn}(x_k) = -1$ da sonst $0 > x_k = z_k + \lambda > 0$ zum Widerspruch führt. Es folgt also $\operatorname{sgn}(x_k) = 0$ und damit $x_k = 0$. Also kommen für große λ dünnbesetzte Vektoren als Lösungen heraus.

Sei nun $\lambda \geq \|z\|_\infty = \max\limits_{k=1,\dots,n} |z_k|$. Für $k = 1, ..., n$ folgt $|z_k| \leq \lambda$ und daher $x_k = 0$. Also gilt $x_\lambda = 0$.

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 semidefinite 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 $A^T b \in \operatorname{Bild} A^T A$ ist. Wir zeigen sogar die Gleichheit $\operatorname{Bild} A^T = \operatorname{Bild} A^T A$. Dabei ist die Inklusion $\operatorname{Bild} A^T \supset \operatorname{Bild} A^T A$ klar. Für $x \in \mathbb{R}^n$ gilt weiterhin

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

Damit folgt aus $x \in \operatorname{Kern} A^T A$, d.h. $A^T Ax = 0$, dass

$$\|Ax\|^2 = \langle Ax, Ax\rangle = \langle A^T Ax, x\rangle = \langle 0, x\rangle = 0.$$

Also ist $\|Ax\| = 0$ und wegen der positiven Definitheit $Ax = 0$, d.h. $x \in \operatorname{Kern}(A)$. Umgekehrt folgt aus $x \in \operatorname{Kern} A$, d.h. $Ax = 0$, dass $A^T Ax = 0$, d.h. $x \in \operatorname{Kern} A^T A$. Also ist $\operatorname{Kern} A^T A = \operatorname{Kern} A$ und $\operatorname{Rang} A^T A = \operatorname{Rang} A = \operatorname{Rang} A^T$. Daher muss bereits $\operatorname{Bild} A^T = \operatorname{Bild} A^T A$ gelten, d.h. es gibt eine Lösung $\bar{x} \in \mathbb{R}^n$ von $A^T Ax = A^T b$.

(iv) Die quadratische Matrix $A^T A \in \mathbb{R}^{n,n}$ hat wegen (iii) $\operatorname{Rang} A^T A = \operatorname{Rang} A = n$ und ist somit invertierbar, d.h. die zugehörige Abbildung $\mathbb{R}^n \to \mathbb{R}^n$ ist bijektiv. Folglich erfüllt nur $x = (A^T A)^{-1} A^T b$ die notwendige Optimalitätsbedingung und ist nach (ii) einziger Minimierer.

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 $f'(\bar{x}) = [f'(\bar{x}, e_1), ..., f'(\bar{x}, e_n)] = 0$. 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 $\bar{x} = 0$ und $d = \begin{pmatrix} x & y \end{pmatrix}^T \in \mathbb{R}^2 \setminus \{0\}$ ist

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

Im Fall $x = 0$ und $y \neq 0$ gilt für $t \neq 0$

$$g_d(t) = t^2y^2 > 0 = g(0)$$

Im Fall $y = 0$ und $x \neq 0$ gilt für $t \neq 0$ analog

$$g_d(t) = 2t^4x^4 > 0 = g(0).$$

Im verbleibenden Fall $x \neq 0$ und $y \neq 0$ gilt für $t < 0$

$$g_d(t) = (ty - t^2x^2)(ty - 2t^2x^2) = t^2(y - tx^2)(y - 2tx^2) > t^2y^2 > 0$$

und für $t \in]0, \frac{|y|}{2x^2}[$ wegen $\operatorname{sgn}(y - tx^2) = \operatorname{sgn}(y - 2tx^2) \neq 0$ schließlich ebenfalls

$$g_d(t) = (ty - t^2x^2)(ty - 2t^2x^2) = t^2(y - tx^2)(y - 2tx^2) > 0$$

Es gibt also stets eine punktierte Umgebung U um 0, so dass für alle $t \in U$ gilt:

$$g_d(t) > 0 = g_d(0).$$

Daher ist $t = 0$ für $\bar{x} = 0$ und für jedes $d \in \mathbb{R}^2\setminus\{0\}$ ein lokaler Minimierer von g_d.

Allerdings ist f auf der Menge $M = \{(x, y) \in \mathbb{R}^2\setminus\{0\} \mid y = \frac{3}{2}x^2\}$ negativ:

$$f(x, y) = (y - x^2)(y - 2x^2) = \left(\frac{1}{2}x^2\right)\left(-\frac{1}{2}x^2\right) = -\frac{1}{4}x^4 < 0$$

Für alle $r > 0$ gilt $B_r(0) \cap M \neq \emptyset$. Daher ist $\bar{x} = 0$ kein lokaler Minimierer von f.