Nichtlineare Optimierung - 4. Hausaufgabe

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

Tutor: Mathieu Rosière

22. Juni 2017

4.1	4.2	4.3	4.4	4.5	Σ

Anmerkungen:

Aufgabe 4.1

Sei $x^0, b \in \mathbb{R}^n$ und $A \in \mathbb{R}^{n\times n}$ positiv definit und symmetrisch. Außerdem seien g_k die im CG-Verfahren zur Lösung des Gleichungssystems $Ax = b$ bei Start in x^0 berechneten Gradienten. Dann ist die Folge $(\|g_k\|_{A^{-1}})_{k=1}^N = \left(\sqrt{g_k^T A^{-1} g_k}\right)_{k=1}^N$ monoton fallend.

Beweis. Für alle $k \in \mathbb{N}$ gilt

$$g_{k+1} = g_k + \sigma_k A d_k \tag{1}$$

mit

$$d_k = -g_k + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1} = -g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} d_{k-1} \tag{2}$$

und

$$\sigma_k = \frac{\|g_k\|^2}{\langle d_k, A d_k\rangle} = \frac{g_k^T g_k}{d_k^T A d_k} \tag{3}$$

Wegen der Symmetrie von A gilt $A = A^T$. Wegen der positiven Definitheit von A gilt $x^T A x \geq 0$ für alle $x \in \mathbb{R}^n$. Außerdem wird in der folgenden Rechnung die Aussage $\langle g_k, d_{k-1}\rangle = g_k^T d_{k-1} = 0$ aus Lemma 4.8.2 genutzt. Mit diesen Hilfsmitteln folgt

$$\begin{aligned}
\|g_{k+1}\|_{A^{-1}} &= \sqrt{g_{k+1}^T A^{-1} g_{k+1}} \\
&\overset{(1)}{=} \sqrt{(g_k + \sigma_k A d_k)^T A^{-1} (g_k + \sigma_k A d_k)} \\
&= \sqrt{g_k^T A^{-1} g_k + g_k^T A^{-1} \sigma_k A d_k + (\sigma_k A d_k)^T A^{-1} g_k + (\sigma_k A d_k)^T A^{-1} \sigma_k A d_k} \\
&= \sqrt{g_k^T A^{-1} g_k + g_k^T A^{-1} \sigma_k A d_k + d_k^T A^T \sigma_k A^{-1} g_k + d_k^T A^T \sigma_k A^{-1} \sigma_k A d_k} \\
&\overset{A=A^T}{=} \sqrt{g_k^T A^{-1} g_k + g_k^T A^{-1} \sigma_k A d_k + d_k^T A \sigma_k A^{-1} g_k + d_k^T A \sigma_k A^{-1} \sigma_k A d_k} \\
&= \sqrt{g_k^T A^{-1} g_k + 2\sigma_k g_k^T d_k + \sigma_k^2 d_k^T A d_k} \\
&\overset{(2)}{=} \sqrt{g_k^T A^{-1} g_k - 2\sigma_k g_k^T g_k + 2\sigma_k g_k^T \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} d_{k-1} + \sigma_k^2 d_k^T A d_k} \\
&\overset{g_k^T d_{k-1}=0}{=} \sqrt{g_k^T A^{-1} g_k - 2\sigma_k g_k^T g_k + \sigma_k^2 d_k^T A d_k} \\
&\overset{(3)}{=} \sqrt{g_k^T A^{-1} g_k - 2 \frac{g_k^T g_k}{d_k^T A d_k} g_k^T g_k + \frac{(g_k^T g_k)^2}{(d_k^T A d_k)^2} d_k^T A d_k} \\
&= \sqrt{g_k^T A^{-1} g_k - \frac{(g_k^T g_k)^2}{d_k^T A d_k}} \\
&\overset{d_k^T A d_k \geq 0}{\leq} \sqrt{g_k^T A^{-1} g_k} \\
&= \|g_k\|_{A^{-1}}
\end{aligned}$$

Also ist $(\|g_k\|_{A^{-1}})_{k=1}^N = \left(\sqrt{g_k^T A^{-1} g_k}\right)_{k=1}^N$ monoton fallend. $\square$

Aufgabe 4.2

Es sei $K \subset \mathbb{R}^d$ ein konvexer Kegel mit $0 \in K$.

(i) Wir zeigen, dass K^* abgschlossen ist. Sei dazu $(x_n)_n \subset K^*$ eine in $\mathbb{R}^d$ konvergente Folge mit Grenzwert $x \in \mathbb{R}^d$. Sei $y \in K$ beliebig. Nach Definition des Dualkegels gilt

$$\langle x_n, y \rangle \leq 0$$

und mit der Stetigkeit der dualen Paarung folgt

$$\langle x, y \rangle = \lim_{n \to \infty} \langle x_n, y \rangle \leq 0.$$

Weil y beliebig war, ist $x \in K^*$.

(ii) Wir zeigen, dass K genau dann abgeschlossen ist, wenn $K = K^{**}$ ist.

$\Leftarrow$ Wegen $K^{**} = (K^*)^*$ folgt mit (i) die Abgeschlossenheit von $K^{**} = K$.

$\Rightarrow$ Für jeden Kegel gilt $K \subset K^{**}$, denn ist $x \in K$ beliebig, so gilt für alle $s \in K^*$

$$\langle x, s \rangle \leq 0,$$

d.h. $x \in K^{**}$. Bleibt noch die zweite Inklusion $K \supset K^{**}$ zu zeigen. Angenommen, es gäbe $y \in K^{**} \backslash \{K\}$. Nach dem Trennungssatz gibt es ein $s \in \mathbb{R}^d$ mit

$$\langle s, y - x \rangle > 0$$

für alle $x \in K$. Wegen $0 \in K$ ist insbesondere

$$\langle s, y \rangle > 0.$$

Für $\lambda > 0$ und $x \in K$ gilt

$$0 < \langle s, y - \lambda x \rangle = \langle s, y \rangle - \lambda \langle s, x \rangle.$$

Weil λ beliebig war, folgt $\langle s, x \rangle \leq 0$ für alle $x \in K$. Damit ist $s \in K^*$. Dann müsste aber auch

$$\langle s, y \rangle \leq 0$$

sein, was uns den ersehnten Widerspruch liefert. Folglich ist $K \supset K^{**}$.

(iii) Es sei $f \in K^*$ und es gelte $\langle f, x_0 \rangle \leq 0$ für einen einen inneren Punkt $x_0 \in K$. Wir zeigen mittels Widerspruchsbeweis, dass $f = 0$ ist. Sei also $f \neq 0$. Da x_0 ein innerer Punkt von K ist, gibt es ein $\varepsilon > 0$ mit $U_\varepsilon(x_0) \subset K$. Damit ist insbesondere $x_0 + \frac{\varepsilon}{2\|f\|} f \in K$ und es gilt

$$0 \geq \langle f, x_0 + \frac{\varepsilon}{2\|f\|} f \rangle = \langle f, x_0 \rangle + \langle f, \frac{2\varepsilon}{\|f\|} f \rangle = 0 + \frac{\varepsilon}{2} \|f\|.$$

Dies kann aber nur gelten, wenn $\|f\| = 0$ ist, was im Widerspruch zur Annahme $f \neq 0$ steht.

(iv) Diese Aussage ist falsch. Dazu betrachten wir für $d = 2$ den Kegel

$$K \coloneqq \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : 0 < x_1 \land 0 < x_2 \right\} \cup \{0\}$$

und

$$K^* = \left\{ \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} : 0 \geq s_1 \land 0 \geq s_2 \right\}.$$

Dass K und K^* Kegel sind, ist offensichtlich. Außerdem gilt für jedes $s \in K^*$

$$\langle s, x \rangle = s_1 x_1 + s_2 x_2 \leq 0$$

für alle $x \in K$. Andererseits ist für $y \in \mathbb{R}^2 \setminus K^*$ entweder $y_1 > 0$ oder $y_2 > 0$; sei o.B.d.A. $y_1 > 0$. Dann gilt für $x \coloneqq (y_1 + |y_2|, \frac{y_1}{2})^T \in K$

$$\langle x, y \rangle \geq y_1^2 + \frac{|y_2| y_1}{2} > 0.$$

Damit ist gezeigt, dass K^* tatsächlich der Dualkegel von K ist. Offensichtlich ist der erste Einheitsvektor nicht in K enthalten, dennoch gilt

$$\langle e_1, s \rangle \leq 0$$

für alle $s \in K^*$.

Aufgabe 4.3

Es sei $C \in \mathbb{R}^d$ eine nichtleere, konvexe und abgeschlossene Menge. Für einen Punkt $x_0 \in \mathbb{R}^d$ definieren wir
$$f : \mathbb{R}^d \to \mathbb{R},\ f(x) = \frac{1}{2}\|x - x_0\|^2$$
und
$$P_C(x_0) := \arg\min_{x \in C} f(x) = \arg\min_{x \in C} \|x - x_0\|.$$

(i) Seien $x, x_0 \in \mathbb{R}^d$. Da f strikt konvex ist, ist die Lösung des Problems eindeutig. Und es gilt genau dann $p(x_0) = x$, wenn x die Variationsungleichung
$$0 \leq \langle \nabla f(x), y - x\rangle = \langle x - x_0, y - x\rangle \quad \text{für alle } y \in C$$
erfüllt. Dies ist äquivalent zu
$$0 \geq \langle x_0 - x, y - x\rangle \quad \text{für alle } y \in C,$$
mit anderen Worten $x_0 - x \in N(C, x)$.

(ii) Sei zusätzlich $x_1 \in \mathbb{R}^d$. Wegen $P_C(x_0), P_C(x_1) \in C$ haben wir mit Aufgabenteil (ii)
$$0 \geq \langle x_0 - P_C(x_0), P_C(x_1) - P_C(x_0)\rangle$$
und
$$0 \geq \langle x_1 - P_C(x_1), P_C(x_0) - P_C(x_1)\rangle.$$
Durch Addition ergibt sich
$$\begin{aligned} 0 \quad &\geq \langle x_0 - P_C(x_0), P_C(x_1) - P_C(x_0)\rangle + \langle x_1 - P_C(x_1), P_C(x_0) - P_C(x_1)\rangle \\ &= \langle x_0 - P_C(x_0), P_C(x_1) - P_C(x_0)\rangle + \langle P_C(x_1) - x_1, P_C(x_1) - P_C(x_0)\rangle \\ &= \langle P_C(x_1) - P_C(x_0), P_C(x_1) - P_C(x_0)\rangle + \langle x_0 - x_1, P_C(x_1) - P_C(x_0)\rangle. \end{aligned}$$
Somit ist mit Cauchy-Schwarz
$$\|P_C(x_1) - P_C(x_0)\|^2 \leq \langle x_1 - x_0, P_C(x_1) - P_C(x_0)\rangle \leq \|x_1 - x_0\|\|P_C(x_1) - P_C(x_0)\|$$
und damit insbesondere
$$\|P_C(x_1) - P_C(x_0)\| \leq \|x_1 - x_0\|,$$
was gerade die Stetigkeit von P_C bedeutet.

Aufgabe 4.4

Bevor wir die gestellten Aufgaben lösen, beweisen wir den folgenden Hilfssatz.

Lemma. Ist $S := \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 1\}$, so gilt $K(S,0) = \mathbb{R}^2$.

Beweis. Sei $x \in \mathbb{R}^2$ vorgegeben. Dann können wir dieses x durch die Wahl eines geeigneten $\varphi_x \in [0, 2\pi[$ in Polarkoordinaten wie folgt darstellen:

$$x = \|x\|_2 \cdot (\cos(\varphi_x), \sin(\varphi_x)).$$

Setzen wir $\widetilde{x} := (\cos(\varphi_x), \sin(\varphi_x))$, so ist $\|\widetilde{x}\|_2 = 1$ und damit $\widetilde{x} \in S$. Nach Konstruktion folgt dann aus

$$x = \|x\|_2 \cdot (\widetilde{x} - 0) \in K(S,0)$$

die Behauptung. $\square$

Nun widmen wir uns den gestellten Aufgaben. Dabei übernehmen wir jeweils die gegebenen Voraussetzungen und Notationen.

(i) Da die Eigenschaft der Konvexität translationsinvariant ist, können wir ohne Einschränkung annehmen, dass $x_0 = 0$ gilt. Da x_0 nach Voraussetzung ein innerer Punkt von C ist, muss ein $\varepsilon > 0$ existieren, mit

$$\overline{U_\varepsilon(0)} \subseteq C.$$

Da Konvexität auch unter Skalierung invariant bleibt, können wir ohne Beschränkung der Allgemeinheit $\varepsilon = 1$ annehmen.

Ist nun ein $x \in \mathbb{R}^n$ vorgegeben, dann gibt es einen Unterraum E_x mit $\dim(E) = 2$ sodass $x \in E_x$ und trivialerweise auch $x_0 \in E$ gilt. Da E_x isomorph zu $\mathbb{R}^2$ ist und wir $E_x \cap \overline{U_\varepsilon(0)}$ mit S identifizieren können, folgt aus dem eingangs bewiesenen Lemma $K(C,0) \cap E_x = E_x$. Da x beliebig gewählt war, gilt folglich

$$K(C,0) = \bigcup_{x \in \mathbb{R}^n} E_x = \mathbb{R}^n.$$

Des Weiteren gilt $N(C,0) = \{0\}$. Wäre dies nicht der Fall, so müsste eine Normalenrichtung $s \in \mathbb{R}^n \setminus \{0\}$ existieren, die für alle $y \in C$ die Abschätzung

$$\langle s, y \rangle \leq 0$$

erfüllt. Wegen $C = \mathbb{R}^n$ muss diese Bedinungen insbesondere für $y = s$ erfüllt sein. Das ist aber der positiven Definitheit von Skalarprodukten wegen nicht möglich. Es gilt nämlich $\langle s, s \rangle > 0$.

(ii) Es sei $H^+ := \{(x_1, x_2)^T \in \mathbb{R}^2 \mid x_2 > 0\}$ die offene obere Halbebene des $\mathbb{R}^2$. Wir werden die Mengengleichheit $K(M,0) = H^+$ beweisen.

Beweis. Zunächst bemerken wir, dass $K(H^+,0) = H^+$ gilt. Wäre dem nicht so, dann ließe sich ein $y = (y_1, y_2)^T \in \mathbb{R}^2$ mit $y_2 \leq 0$ und $y \in K(H^+,0)$ finden. Wegen $K(H^+,0) = \{(\alpha x_1, \alpha x_2)^T \in \mathbb{R}^2 \mid \alpha > 0, x_2 > 0\}$ kann aber so ein y nicht existieren. Wegen $M \subseteq H^+$ folgt dann unmittelbar $K(M,0) \subseteq K(H^+,0) = H^+$.

Nun zur anderen Mengeninklusion: Betrachten wir zunächst den Rand ∂M von $\overline{M}$. Dann ist die eindeutig bestimmte Tangente von ∂M an x_0 durch den Unterraum $\operatorname{Span}(1,0)^T$ gegeben. Ist nun ein $x = (x_1, x_2) \in H^+$ gegeben, dann folgt aus $x_2 > 0$ für den Winkel $\measuredangle((1,0), x)$ zwischen x und dem ersten Einheitsvektor

$$\measuredangle((1,0), x) \in]0, \pi[.$$

Demnach muss die Verbindungsstrecke $\overline{x_0 x}$ oberhalb der Tangente liegen. Da nun aber $x_0 \in \partial M$ gilt, hat $\overline{x_0 x}$ zwei Schnittpunkte mit ∂M. Es gibt daher ein von x_0 und x verschiedenes Element $y \in \overline{x_0 x}$ mit $y \in M$. Wegen

$$x = \frac{\|x\|}{\|y\|} \cdot (y - 0)$$

folgt somit aus $x \in K(H^+, 0)$ die gewünschte Gleichheit. $\square$

Zum Abschluss zeigen wir noch die Mengengleichheit

$$N' := \{(0, x_2) \in \mathbb{R}^2 \mid x_2 \leq 0\} = N(M, 0).$$

Sei dafür zunächst ein $s = (0, s_2) \in N'$ gegeben. Dann gilt für alle $y = (y_1, y_2) \in H^+$ die Identität $\langle s, y - 0 \rangle = s_2 y_2$. Wegen $s_2 \leq 0$ und $y_2 > 0$, folgt $s_2 y_2 \leq 0$ und damit $s \in N(M, 0)$.

Sei nun $s = (s_1, s_2) \in N(M, 0)$. Dann gilt nach Definition für alle $y = (y_1, y_2) \in M$

$$\langle s, y - 0 \rangle = s_1 y_1 + s_2 y_2 \leq 0. \tag{4}$$

Ist $y = (0, y_2)$ für ein $y_2 \in \mathbb{R}$, so folgt mit (4) gerade $s_1 y_1 + s_2 y_2 = s_2 y_2$. Wegen $y_2 > 0$ kann die geforderte Ungleichung nur im Falle $s_2 \leq 0$ erfüllt werden. Ist $s_1 \neq 0$, dann folgt mit $y = (-s_2/s_1 + 1, 1)$

$$s_1 y_1 + s_2 y_2 = 1 > 0.$$

Demnach muss $s_1 = 0$ gelten. Damit folgt dann endlich $N(M, 0) = N'$. $\square$

(iii) Es sei $f \colon \mathbb{R}^3 \to \mathbb{R}, f(x) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2)$ und $g \colon \mathbb{R}^3 \to \mathbb{R}, g(x) = x_1 + x_2 + x_3 - 6$. Dann ist die Lagrange-Funktion für das in der Aufgabe gestellte Minimierungsproblem durch

$$L(x, \lambda) = f(x) + \lambda g(x)$$

gegeben. Wir untersuchen die notwendige Optimalitätsbedingung

$$\begin{pmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{pmatrix} = \begin{pmatrix} x_1 + \lambda \\ x_2 + \lambda \\ x_3 + \lambda \\ x_1 + x_2 + x_3 - 6 \end{pmatrix} \overset{!}{=} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Die Erfüllung der Optimalitätsbedingung ist damit äquivalent zur Existenz einer Lösung des linearen Gleichungssystems

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 6 \end{pmatrix}.$$

Dieses System besitzt die eindeutige Lösung

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \\ -2 \end{pmatrix},$$

d.h. bei $x = (2, 2, 2)^T$ wird das Minimum von $f(x) = 6$ angenommen.

Aufgabe 4.5

Es sei $A\colon \mathbb{R}^n \to \mathbb{R}^n$ symmetrisch und positiv definit sowie $b \in \mathbb{R}^n$. Ferner bezeichne $\|\cdot\|_A$ die von A induzierte Norm und $\|\cdot\|_{A^{-1}}$ die von A^{-1} induzierte Norm. Es sei $x \in \mathbb{R}^n$ die Lösung von $Ax = b$ und sei $\tilde{x} \in \mathbb{R}^n$ ein völlig beliebiger Vektor.

Dann gilt

$$\begin{aligned}
\|x - \tilde{x}\|_A^2 &= (x - \tilde{x})^T A (x - \tilde{x}) \\
&= (A^{-1}b - \tilde{x})^T (b - A\tilde{x}) \\
&= (A^{-1}(b - A\tilde{x}))^T (b - A\tilde{x}) \\
&= \|b - A\tilde{x}\|_{A^{-1}}^2.
\end{aligned}$$

Mit der Dreiecksungleichung folgt dann für beliebiges $w \in \mathbb{R}^n$

$$\|x - \tilde{x}\|_A = \|b - (w - w) - A\tilde{x}\|_{A^{-1}} \leq \|b - w\|_{A^{-1}} + \|A\tilde{x} - w\|_{A^{-1}}.$$

Wir betrachten nun $n = 1$ und $A = \mathrm{id}$. Dann gilt z.B. für $x = 0$, $\tilde{x} = 1$ und $w = 0$ die Gleichheit

$$|x - \tilde{x}| = |0 - 1| = 1 = |0 - 0| + |1 - 0| = |b - w| + |A\tilde{x} - w|.$$