Notizen zur Vorlesung Analysis 3

Henrik Schumacher

17. November 2016

1 Unrestringierte Optimierung

1.1 Notwendige Bedingungen für lokale Extrema

Satz 1.1 *Sei $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ partiell differenzierbar und sei $x_0 \in D$ ein* innerer *Punkt von D (d. h., es gibt einen Radius $r > 0$ mit $B_r(x_0) \subset D$).*

i) Ist $x_0 \in D$ eine Extremalstelle von f, so gilt $\nabla f(x_0) = 0$.

ii) Sei f zusätzlich zweimal stetig partiell differenzierbar bei x_0. Ist x_0 ein lokales Minimum (Maximum), so ist die Hessematrix $H_f(x_0)$ positiv semidefinit (negativ semidefinit).

Erinnerung: Eine symmetrische Matrix $M \in \mathbb{R}^{n\times n}$ ist positiv semidefinit (negativ semidefinit), falls für alle $x \in \mathbb{R}^n$ gilt: $x^{\mathsf{T}}Mx \geq 0$ (oder $x^{\mathsf{T}}Mx \leq 0$).

Proof. Nachweis für i): Letzte Vorlesung.

Nachweis für ii): Wir gehen nur den Fall durch, dass x_0 ein lokales Minimum von f ist. Betrachte wieder zur Richtung $h \in \mathbb{R}^n$ die Funktion $g(t) = f(x_0 + t\,h)$. Wir haben

$$\begin{aligned} g'(t) &= (h \cdot \nabla) f(x_0 + t\,h) = (\nabla f(x_0 + t\,h))^{\mathsf{T}} h \\ g''(t) &= (h \cdot \nabla)^2 f(x_0 + t\,h) = h^{\mathsf{T}} H_f(x_0 + t\,h)\, h. \end{aligned}$$

Aus i) folgt:

$$g'(0) = (\nabla f(x_0))^{\mathsf{T}} h = 0.$$

Sei $r > 0$ mit $B_r(x_0) \subset D$. Wir verwenden die eindimensionale Taylorformal mit Lagrange-Restglied: Für jedes t mit $|t| < r/|h|$ gibt es ein $\xi(t) \in \,]-|t|, |t|[$ so dass:

$$g(0) = f(x_0) \leq f(x_0 + t\,h) = g(t) = g(0) + \cancel{g'(0)}\,t + \frac{1}{2!} g''(\xi(t))\,t^2 = g(0) + \frac{1}{2!} g''(\xi(t))\,t^2$$

Wir erhalten

$$h^{\mathsf{T}} H_f(x_0 + \xi(t)\, h)\, h = g''(\xi(t)) \geq 0 \quad \text{für alle hinreichend kleinen } |t|.$$

Da $x \mapsto H_f(x)$ nach Voraussetzung stetig ist, erhalten wir durch Grenzübergang $|t| \to 0$ die Ungleichung

$$h^{\mathsf{T}} H_f(x_0)\, h \geq 0. \qquad \square$$

1.2 Hinreichende Bedingungen für lokale Extrema

Satz 1.2 *Sei $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ zweimal stetig partiell differenzierbar und sei $x_0 \in D$ ein* innerer *Punkt von D.*

Gilt $\nabla f(x_0) = 0$ und ist die Hessematrix $H_f(x_0)$ positiv definit, so ist x_0 ein striktes lokales Minimum von f.

Analog: Gilt $\nabla f(x_0) = 0$ und ist die Hessematrix $H_f(x_0)$ negativ definit, so ist x_0 ein striktes lokales Maximum von f.

Erinnerung: Eine symmetrische Matrix $M \in \mathbb{R}^{n\times n}$ ist positiv definit (negativ definit), falls für alle $x \in \mathbb{R}^n$ gilt: $x^{\mathsf{T}} M x > 0$ (oder $x^{\mathsf{T}} M x < 0$). Insbesondere hat dies zur Folge, dass für den betragsmäßig kleinsten Eigenwert $\lambda_1(M)$ von M gilt: $|\lambda_1(M)| > 0$. Daher hat man

$$x^{\mathsf{T}} M x \geq |\lambda_1(M)|\, |x|^2 \quad \Big(\text{bzw.} \quad x^{\mathsf{T}} M x \leq -|\lambda_1(M)|\, |x|^2\Big).$$

Proof. Wir behandeln nur den Fall, dass $H_f(x_0)$ positiv definit ist. Mit den Bezeichnungen aus dem vorangegangenen Beweis erhalten wir wieder mit der Taylorformel:

$$g(t) = g(0) + \frac{1}{2!} g''(\xi(t))\, t^2 = g(0) + \frac{1}{2}(t\, h)^{\mathsf{T}} H_f(x_0 + t\, h)\, (t\, h).$$

Benutzen wir nun, dass der betragsmaeßig kleinste Eigenwerte $\lambda_1(M)$ stetig von M abhängt, so sehen wir, dass $\lambda_1(H_f(x_0 + t\, h)) \geq \frac{1}{2}\lambda_1(H_f(x_0)) > 0$ für hinreichend kleine $|t|$ gelten muss. Insbesondere ist dann auch $H_f(x_0 + t\, h)$ positiv definit und wir erhalten

$$g(t) \geq g(0) + \frac{1}{2}\, \lambda_1(H_f(x_0 + t\, h))\, |t\, h|^2 > g(0). \qquad \square$$

Beispiel 1.3 Sei $f\colon \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x_1, x_2) = \exp(-|x|^2) = \exp(-x_1^2 - x_2^2)$. Man rechnet schnell nach, dass

$$\nabla f(x) = \begin{pmatrix} -2\, x_1 \\ -2\, x_2 \end{pmatrix} \exp(-|x|^2), \quad \text{und} \quad H_f(x) = \begin{pmatrix} (4x_1^2 - 2) & 4\, x_1\, x_2 \\ 4\, x_1\, x_2 & (4x_2^2 - 2) \end{pmatrix} \exp(-|x|^2).$$

Folglich ist $x_0 = (0, 0)^{\mathsf{T}}$ der einzige kritische Punkt von f und die Hessematrix dort ist $H_f(x_0) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$, also negativ definit. Folglich liegt bei $x_0 = (0, 0)^{\mathsf{T}}$ ein striktes lokales Maximum. (Das ist im Übrigen auch das globale Maximum.)

Wir wissen bereits aus der Linearen Algebra: Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ ist positiv definit genau dann wenn alle ihre Eigenwerte positiv sind. In der Regel ist es aber recht mühsam, die Eigenwerte einer Matrix zu bestimmen. Ein etwas praktikableres Verfahren, um die Positivdefinitheit einer symmetrischen Matrix zu testen liefert das *Hurwitz-Kriterium* (auch *Hauptminorenkriterium* genannt).

Satz 1.4 (Hurwitz Kriterium) *Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ ist genau dann positiv definit wenn alle* führenden Hauptminoren $\det(A_1), \det(A_2), \dots, \det(A_n)$ *positiv sind. Dabei seien*

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \quad \textit{und} \quad A_k := \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}, \quad \textit{für } k \in \{1, 2, \dots, n\}.$$

Bemerkung 1.5 Achtung! *Negativ* definite Matrizen besitzen in der Regel auch *positive* Hauptminoren! Als Beispiel sei die negativ definite Matrix

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

mit den führenden Hauptminore $\det(A_2) = 1 > 0$ genannt.

Daher überprüft man besser eine Matrix A auf Negativdefinitheit, indem man das Hurwitz-Kriterium auf $-A$ anwendet.

Bemerkung 1.6 Achtung! Für semidefinite Matrizen kann man das Hurwitz-Kriterium nicht benutzen. Die führenden Hauptminoren der Matrix

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

sind alle identisch zu 0, also nicht negativ. Trotzdem ist A nicht positiv semidefinit.

Beispiel 1.7 (Hurwitz Kriterium im $\mathbb{R}^2$) Ist $f\colon \mathbb{R}^2 \to \mathbb{R}$ eine zweimal stetig partiell differenzierbare Function und sei $x_0 \in \mathbb{R}^2$ ein kritischer Punkt von f (also sei $\nabla f(x_0) = 0$). Setze

$$A := H_f(x_0) = \begin{pmatrix} f_{x_1x_1}(x_0) & f_{x_1x_2}(x_0) \\ f_{x_2x_1}(x_0) & f_{x_2x_2}(x_0). \end{pmatrix}$$

Dann hat man

$$\det(A_1) = f_{x_1x_1}(x_0) \quad \text{und} \quad \det(A_2) = f_{x_1x_1}(x_0) f_{x_2x_2}(x_0) - f_{x_1x_2}^2(x_0).$$

Daher gilt:

1. Ist $f_{x_1x_1}(x_0)f_{x_2x_2}(x_0) - f^2_{x_1x_2}(x_0) > 0$ und $f_{x_1x_1}(x_0) > 0$, so ist $H_f(x_0)$ positiv definit und x_0 ist ein striktes lokales Minimum.
2. Ist $f_{x_1x_1}(x_0)f_{x_2x_2}(x_0) - f^2_{x_1x_2}(x_0) > 0$ und $f_{x_1x_1}(x_0) < 0$, so ist $-H_f(x_0)$ positiv definit und x_0 ist ein striktes lokales Maximum.

Ist aber $f_{x_1x_1}(x_0)f_{x_2x_2}(x_0) - f^2_{x_1x_2}(x_0) < 0$, so muss $H_f(x_0)$ einen positiven und einen negativen Eigenwert haben. Man sagt dann, x_0 habe einen *Sattelpunkt*.

Definition 1.8 (Sattelpunkt) Sei $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ zweimal stetig partiell differenzierbar und sei $x_0 \in D$ ein stationärer Punkt, d. h., $\nabla f(x_0) = 0$. Besitzt $H_f(x_0)$ sowohl positive als auch negative Eigenwerte, so nennt man x_0 einen *Sattelpunkt*.

Beispiel 1.9 Sei $f\colon \mathbb{R}^2 \to \mathbb{R}$ durch $f(x_1, x_2) = 3x_1x_2 - x_1^3 - x_2^3$ gegeben. Man hat

$$\nabla f(x) = (3\,x_2 - 3x_1^2, 3\,x_1 - 3x_2^2)^\mathsf{T}$$

und die einzigen (reellen) Nullstellen sind $P_1 = (1, 1)$ und $P_2 = (0, 0)$. Man hat

$$H_f(P_1) = \begin{pmatrix} -6 & 3 \\ 3 & -6 \end{pmatrix}, \quad \det(H_f(P_1)) = 27 > 0.$$

Wegen $f_{x_1x_1}(P_1) = -6 < 0$ liegt bei P_1 ein striktes lokales Maximum vor.

Andererseits hat man

$$H_f(P_2) = \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}, \quad \det(H_f(P_2)) = -9 < 0.$$

Also liegt bei P_2 ein Sattelpunkt vor.

2 Ausblick: Optimierung unter Nebenbedingungen

Die bisherigen Beispiele behandelten vornehmlich Funktionen $f\colon M \subset \mathbb{R}^n \to \mathbb{R}$, bei denen der Definitionsbereich M ausschließlich aus inneren Punkten bestand.

Wir möchten an einigen Beispielen auch etwas allgemeinere Optimierungsprobleme der folgenden Form betrachten:

Finde $x_0 \in M$ mit $f(x_0) \le f(x)$ für alle $x \in M \subset \mathbb{R}^n$.

Hierbei nennt man M die *zulässige Menge*.

2.1 Konvexe Mengen

Sei $M = [a, b] \subset \mathbb{R}$ ein abgeschlossenes Intervall und $f\colon [a, b] \to \mathbb{R}$.

Minimierungsaufgabe: Finde $x_0 \in [a, b]$ mit $f(x_0) \leq f(x)$ für alle $x \in [a, b]$.

Beispiel 2.1 Betrachten wir zum Beispiel die Funktion $f(t) = \sin(4\,x) + 2\,x$ auf dem Intervall $[\frac{\pi}{24}, \pi]$. Diese hat mehrere kritische Punkte im Inneren des Intervalls, nämlich $\{\frac{\pi}{3}, \frac{2\pi}{3}, \frac{4\pi}{3}, \frac{5\pi}{3}\}$. Das globale Minimum liegt aber beim linken Randpunkt $x_0 = a = \frac{\pi}{24}$ und hier hat man $f'(x_0) = 2 + 2\sqrt{3} > 0$. Insbesondere gilt also $f'(x_0) \neq 0$.

Um eine *notwendige Bedingung erster Ordnung für lokale Minima* zu bekommen kann man hier wie folgt vorgehen:

Sei x_0 lokales Minimum von $[a, b] \subset \mathbb{R}$. Für jedes $y \in [a, b]$ und jedes $s \in \,]0, 1[$ ist $x = x_0 + s(y - x_0)$ wieder ein Element des Intervalls $[a, b]$. Deshalb hat man

$$f(x_0) \leq f(x) = f(x_0 + s(y - x_0)).$$

Wegen $s > 0$ folgt

$$\frac{f(x_0 + s(y - x_0)) - f(x_0)}{s} \geq 0 \quad \text{für alle } s \in \,]0, 1[$$

und im einseitigen Grenzwert erhält man

$$f'(x_0)\,(y - x_0) = \lim_{s \to 0+} \frac{f(x_0 + s(y - x_0)) - f(x_0)}{s} \geq 0 \quad \text{für alle } y \in [a, b].$$

Diese Argumentation kann man analog auf partiell differenzierbare Funktionen $f\colon M \to \mathbb{R}$ für konvexe Mengen $M \subset \mathbb{R}^n$ übertragen und man erhält als notwendige Bedingung erster Ordnung für ein Minimum bei $x_0 \in M$:

$$(\nabla f(x_0))^\mathsf{T} (y - x_0) \geq 0 \quad \text{für alle } y \in M.$$

Ersetzt man $h = (y - x_0)$, so erhält man sinngemäß:

Lemma 2.2 *Ist x_0 ein lokales Minimum der differenzierbaren Funktion $f\colon M \to \mathbb{R}$, so gilt*

$$(\nabla f(x_0))^\mathsf{T} h \geq 0 \quad \textit{für alle} \text{ nach innen zeigenden Richtungen } h \in \mathbb{R}^n.$$

Dies kann man auch für noch größere Klassen von Mengen verallgemeinern, insofern man einen geeigneten Begriff von “nach innen zeigenden Richtungen” entwickelt. Man spricht dann auch von zulässigen Richtungen.

2.2 Niveaumengen

Minimierungsaufgabe: Finde $x_0 \in M$ mit $f(x_0) \leq f(x)$ für alle $x \in M$, wobei M von der Form $M = \{ x \in \mathbb{R}^n \mid F(x) = 0 \}$ mit einer Funktion $F \colon \mathbb{R}^n \to \mathbb{R}^m$ ist.

Beispiel 2.3 Sei $F \colon \mathbb{R}^2 \to \mathbb{R}$, $F(x) = |s|^2 - 1$. Dann ist $M = \{ x \in \mathbb{R}^2 \mid F(x) = 0 \}$ der Rand des Einheitskreises. Für Punkte $x = (x_1, x_2) \in M$ mit $x_2 \neq 0$ kann man eine Umgebung von x in M lokal als Graph schreiben:

$$x_2 = g_1(x_1) = \sqrt{1 - x_1^2} \quad \text{oder} \quad x_2 = g_2(x_1) = -\sqrt{1 - x_1^2}.$$

Ebenso kann man für $x_1 \neq 0$ eine Umgebung von x in M durch Graphen eineindeutig parametrisieren:

$$x_1 = g_3(x_2) = \sqrt{1 - x_2^2} \quad \text{oder} \quad x_1 = g_4(x_2) = -\sqrt{1 - x_2^2}.$$

Lokale Minima von f kann man dann finden, indem man die Funktionen $t \mapsto f(g_1(t), t)$ und $t \mapsto f(g_2(t), t)$, $t \mapsto f(t, g_3(t))$ und $t \mapsto f(t, g_4(t))$ untersucht.

Dies wirft folgende Frage auf:

> *Unter welchen Voraussetzungen ist es möglich, die Lösungsmenge einer Gleichung $F(x_1, x_2) = 0$ lokal als Graph einer Funktion g zu schreiben?*

Satz 2.4 (Satz über implizite Funktionen) *Seien $U_1 \subset \mathbb{R}^k$ und $U_2 \subset \mathbb{R}^m$ offene Mengen und die Funktion $F \colon U_1 \times U_2 \to \mathbb{R}^m$ sei stetig differenzierbar. Der Punkt $(x_0, y_0) \in U_1 \times U_2$ erfülle die Gleichung $F(x_0, y_0) = 0$ und die $m \times m$-Matrix $D_y F(x_0, y_0)$ sei invertierbar. Dann gibt es eine offene Umgebungen $V_1 \subset U_1$ von x_0, eine offene Umgebungen $V_2 \subset U_2$ von y_0 und eine stetige Funktion $g \colon V_1 \to V_2$ so dass gilt:*

1. *Für alle $x \in V_1$ gilt $F(x, g(x)) = 0$.*
2. *Ist $(x, y) \in V_1 \times V_2$ mit $F(x, y) = 0$, so gilt $y = g(x)$.*

Ferner ist g stetig differenzierbar mit Ableitung

$$Dg(x) = -D_y F(x, g(x))^{-1} \, D_x F(x, g(x)).$$

Der Satz besagt also tatsächlich, dass sich die Lösungsmenge von $F(x, y) = 0$ in einer geeigneten Umbegung vom Punkt (x_0, y_0) als Graph der stetig differenzierbaren Funktion g darstellen lässt:

$$\{ (x, y) \in V_1 \times V_2 \mid F(x, y) = 0 \} = \{ (x, g(x)) \mid x \in V_1 \}.$$

In diesem Sinne lässt sich die Niveaumenge zumindest lokal (fast) genauso gut behandeln, wie eine offene Menge des $\mathbb{R}^m$.

Beispiel 2.5 Sei $F\colon \mathbb{R}^2 \to \mathbb{R}$, $F(x, y) = x^2 + y^2 - 1$, d. h. $k = m = 1$. Wir haben $D_y F(x, y) = 2\,y$. Für Lösungen (x_0, y_0) der Gleichung $F(x, y) = 0$ mit $y \neq 0$ sind also die Voraussetzungen des Satzes über implizite Funktionen erfüllt. Am Punkt $(x_0, y_0) = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ haben wir zum Beispiel $D_y F(x, y) = \sqrt{2} \neq 0$. Tatsächlich finden wir auch (durch analytische Rechnung) die Funktion $g(x) = \sqrt{1 - x^2}$, so dass $F(x, g(x)) = 0$ gilt. Der Satz über implizite Funktionen besagt für $Dg(x)$:

$$g'(x) = Dg(x) = -D_y F(x, g(x))^{-1}\, D_x F(x, g(x)) = -\frac{2\,x}{2\,\sqrt{1 - x^2}} = -\frac{x}{\sqrt{1 - x^2}}.$$

Diese Ergebnis bekommt man natürlich auch, wenn man $g'(x)$ direkt aus dem analytischen Ausdruck mit der Kettenregel berechnet.