

Kapitel 10: Optimierung

Definition 10.1 Sei $X \subset \mathbb{R}^{n \times 1}$ offen und $f: X \rightarrow \mathbb{R}$ differenzierbar. Dann
heißt $x \in X$ kritischer Punkt von f wenn $\nabla f(x) = 0$.

Satz 10.2: (notwendige Extremalitätsbedingungen)

$X \subset \mathbb{R}^{n \times 1}$ offen, $f: X \rightarrow \mathbb{R}$ habe lokale Extremalstelle in $x \in X$.

- Ist f differenzierbar in X dann ist x kritischer Punkt von f
- Ist f zweimal differenzierbar in X dann ist die Hessematrix von f in x semi-definit (per falls lok. Minimum, neg. falls lok. Maximum)

Beweis: Kürzungen auf 1D

Sei $B_\epsilon(x) \subset X$, $\xi \in \partial B_\epsilon(0)$, $g(t) = f(x + t\xi)$ $t \in (-\epsilon, \epsilon)$

Dann hat g lok. Extremstelle in 0 also (AI) $g'(0) = 0$, $g''(0) \leq 0$

Es ist $g'(t) = \partial_\xi f(x + t\xi)$, $g''(t) = \partial_\xi \partial_\xi f(x + t\xi)$

also $g'(0) = \partial_\xi f(x) = f'(x)(\xi) = \xi \cdot \nabla f(x)$

und $g''(0) = f''(x)(\xi, \xi) = \sum_{i,j=1}^n \xi_i \xi_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \xi^T H \xi$

wegen $0 = \xi \cdot \nabla f(x)$ für alle $\xi \in \partial B_1(0)$ folgt $\nabla f(x) = 0$ (sonst Widerspruch mit $\xi = \frac{\nabla f(x)}{\|\nabla f(x)\|}$)

und $\xi^T H \xi \geq 0$ für alle $\xi \in \partial B_1(0)$ bedeutet H pos/neg semi definit

Zusammenfassung: 1) Suche Optima nur unter kritischen Punkten statt unter allen Punkten

kontinuierlich durch Lösen der Gleichung $\nabla f(x) = 0$

2) Kontrolliere Turnere an kritischen Punkten

kleinster Wert: Kandidat für globales Minimum

größter Wert: Kandidat für globales Maximum

3) Kontrolliere "Randverhalten" der Funktionswerte

wenn potentielle globale Extremwerte an(2) nicht über/untere Schranken werden, dann sind Optima gefunden.

Beispiel: $f(x) = x_1^4 + x_2^4 - 4x_1^3 + 5x_1^2 - x_2^2 - 2x_1 + 1$ $x \in \mathbb{R}^{2 \times 1}$

$$\partial_1 f(x) = 4x_1^3 - 12x_1^2 + 10x_1 - 2 \stackrel{!}{=} 0$$

$$\partial_2 f(x) = 4x_2^3 - 2x_2 \stackrel{!}{=} 0$$

(norm. Bed.)

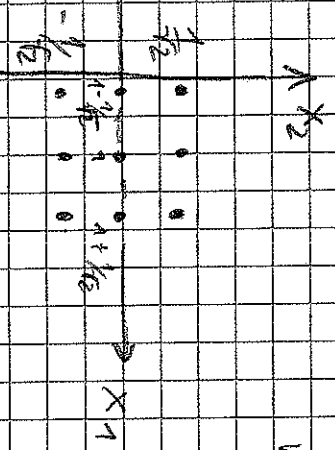
$$x_2(2x_2^2 - 1) = 0 \Leftrightarrow x_2 = 0 \text{ oder } x_2 = \frac{1}{\sqrt{2}} \text{ oder } x_2 = -\frac{1}{\sqrt{2}}$$

da $x_1 = 1$ Nullstelle von $\partial_1 f(x)$ mit Polynomdivision

$$\partial_1 f(x) = (x_1 - 1)(4x_1^2 - 8x_1 + 2) = 0$$

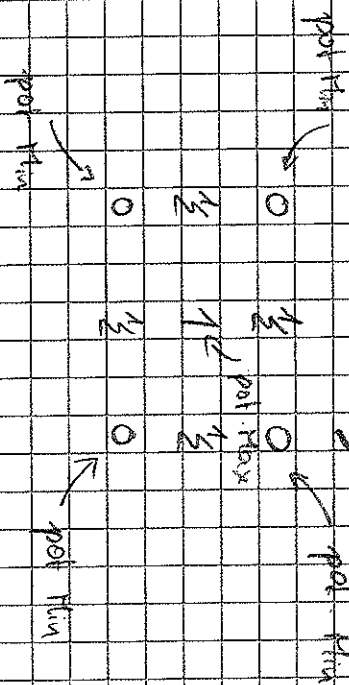
$$\Leftrightarrow x_1 = 1 \text{ oder } x_1 = 1 + \frac{1}{\sqrt{2}} \text{ oder } x_1 = 1 - \frac{1}{\sqrt{2}}$$

9 krit. Punkte:



mit Funktionswerten

2)



wichtig: Randbelohnung 3)

aufserhalb

einen großen Würfel $\|x\|_\infty \geq R$ gilt

$x_1^4 + x_2^4 \geq R^4$ und für R hinreichend groß damit $f(x) \geq \frac{1}{2} R^4$

dann

$$\begin{aligned} & \left| -4x_1^3 + 5x_1^2 - x_2^2 - 2x_1 + 1 \right| \\ & \leq \frac{4\|x\|_\infty^3 + 6\|x\|_\infty^2 + 2\|x\|_\infty + 1}{\|x\|_\infty^4} \end{aligned}$$

$$\leq \frac{4}{R} + \frac{6}{R^2} + \frac{2}{R^3} + \frac{1}{R^4} \xrightarrow{R \rightarrow \infty} 0$$

Zusammenfassend:

f ist nach oben unbeschränkt
hat also kein Maximum (bei (0) ebenfalls lok. Max) also $< \frac{1}{2}$ für $R \geq R$

Betrachte großen Würfel mit Kantenlänge ≥ 2 und Randwerten ≥ 1

f nimmt im abg. Würfel Min an (kompakt!)

Da $f(\frac{1}{2}, \frac{1}{2}) = 0 < 1$ ist Min nicht am Rand

Daher liegt Min im offenen Würfel und nach Satz 10.2 am krit. Punkt

$\leadsto \min f = 0$

Zur Charakterisierung kritischer Punkte dient

Satz 10.3 Sei $X \subset \mathbb{R}^m$ offen, $f \in C^2(X, \mathbb{R})$, $x \in X$ kritischer Punkt von f und H Hessematrix von f in x . Ist H ① pos. definit ② neg. definit ③ indefinit, dann hat f ① globales Minimum, ② lok. Maximum ③ kein lok. Extremum in x

Beweis: Satz von Taylor für $\|h\| < r$ mit $\mathcal{B}_r(x) \subset X$, $h \neq 0$

$$f(x+h) = f(x) + \underbrace{f'(x)}_{=h \cdot \nabla f(x)=0}(h) + \frac{1}{2} f''(x+\partial h)(h, h)$$

$$= f(x) + \frac{1}{2} h^T H h + \frac{1}{6} \|h\|^3 \left(f''(x+\partial h) \left(\frac{h}{\|h\|}, \frac{h}{\|h\|} \right) - f''(x) \left(\frac{h}{\|h\|}, \frac{h}{\|h\|} \right) \right)$$

beachte f' stetig in x d.h. zu jedem $\varepsilon > 0$ gibt es $\delta(\varepsilon) > 0$ mit

$$\|f''(y) - f''(x)\|_{\mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}))} < \varepsilon \quad \text{falls} \quad \|y - x\| < \delta(\varepsilon)$$

$$\|f''(y)(h) - f''(x)(h)\|_{\mathcal{L}(\mathbb{R}^m, \mathbb{R})} < \varepsilon \quad \text{für alle} \quad \|h\| = 1$$



$$|f''(y)(k)(h) - f''(x)(k,h)| < \varepsilon \quad \text{falls } \|k\|, \|h\| < 1 \text{ und } \|y - x\| < \delta(\varepsilon)$$

damit

$$\frac{1}{2} h^T H h - \frac{1}{2} \|h\|^2 \leq f(x+h) - f(x) \leq \frac{1}{2} h^T H h + \frac{1}{2} \|h\|^2 \varepsilon$$

$$\text{falls } \|h\| < \min\{r, \delta(\varepsilon)\}$$

H pos. def: $\xi^T H \xi > 0$ für alle $\xi \neq 0$

also $\xi \mapsto \underbrace{\xi^T H \xi}_{\text{steig}} > 0$ auf $\underbrace{\partial B_1(0)}_{\text{kompakt}}$

$$\Rightarrow \xi^T H \xi \geq \lambda_{\min} > 0 \quad \xi \in \partial B_1(0)$$

$$\leadsto h^T H h \geq \lambda_{\min} \|h\|^2 \quad \text{für alle } h$$

mit $\varepsilon = \lambda_{\min}/2$ $f(x+h) - f(x) \geq \frac{1}{4} \lambda_{\min} \|h\|^2$ lok. Minimum

H neg. def: $\xi^T H \xi < 0$ für alle $\xi \neq 0 \Rightarrow h^T H h \leq -\underbrace{\lambda_{\max}}_{>0} \|h\|^2$

mit $\varepsilon = \lambda_{\max}/2$ $f(x+h) - f(x) \leq -\frac{1}{4} \lambda_{\max} \|h\|^2$ lok. Maximum

H indef: $\exists \eta, \mu \in \mathbb{R}_{>0}$:

$$\eta^T H \eta = \min_{\xi \in \partial B_\eta(0)} \xi^T H \xi \leq -c < 0$$

$$\mu^T H \mu = \max_{\xi \in \partial B_\mu(0)} \xi^T H \xi \geq c > 0$$

wähle $\varepsilon = \frac{c}{2}$
 $\Rightarrow$ auch in Richtung η, μ
 $\Rightarrow$ x kein lok. Extremum

Beispiel (fortsetzung)

$$\partial_1^2 f(x) = 12x_1^2 - 24x_1 + 10$$

$$\partial_1 \partial_2 f(x) = 0$$

$$\partial_2^2 f(x) = 12x_2^2 - 2$$

an der Stelle $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$:

$$H = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

hat neg. Eigenwerte $\Rightarrow$ neg. def. $\Rightarrow$ lok. Max

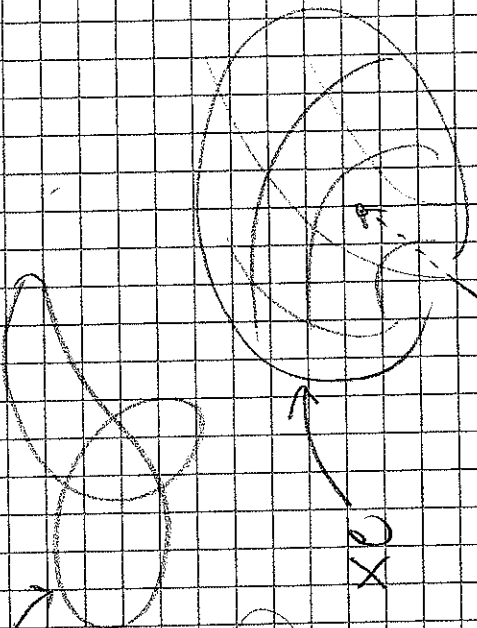
an der Stelle $\begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$:

$$H = \begin{pmatrix} -2 & 0 \\ 0 & 4 \end{pmatrix}$$

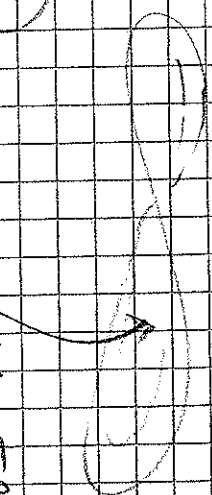
pos. und neg. Eigenwerte $\Rightarrow$ indef. $\Rightarrow$ kein Ext.

bisher: Optimalitätsbedingung an inneren Punkten

X: Volumen hat innere Punkte



X hat keine inneren Punkte



X : Fläche hat keine inneren Punkte

X : Kurve hat keine inneren Punkte

Besonders häufig möglichkeiten für "duinne" Mengen $\rightarrow$ in frei wählbare Parameter

1) explizit:

$$X = \varphi(A) \quad \text{mit} \quad \varphi: \mathbb{R}^{(m \times 1)} \supset A \rightarrow \mathbb{R}^{(n \times 1)} \quad m < n$$

2) implizit:

$$X = \{x \in \mathbb{R}^{(n \times 1)} \mid \phi(x) = 0\} \quad \text{mit} \quad \phi: \mathbb{R}^{(n \times 1)} \rightarrow \mathbb{R}^{(k \times 1)}$$

k -Bedingungen

Bsp: k Variablen frei d.h. $m = n - k$ frei wählbare Komponenten

1)

Wir wissen schon:

$$\varphi: A \xrightarrow{\text{n. offen}} \mathbb{R}^{n \times 1}$$

Spalten von $\varphi'(u)$

$\partial_1 \varphi(u), \partial_2 \varphi(u), \dots, \partial_m \varphi(u)$ Tangentialvektoren

an $X = \varphi(A)$ im Punkt $\varphi(u)$

$X = \varphi(A)$ heißt m-dimensionale Fläche

Wenn $\text{rang } \varphi' = m$ auf A

dann ist Tangentialraum in jedem Punkt $\varphi(u)$ m-dimensional

und φ heißt Parameterisierung von X

angenommen: $f: \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}$ differenzierbar

und $f|_X$ hat lokales Extremum in $x = \varphi(u)$

d.h. $f \circ \varphi$ hat lokales Extremum in u

d.h. (Satz 10.2) $(f \circ \varphi)'(u) = 0 \iff$ notwendige Bedingung

d.h. $(f \circ \varphi)'(u)(h) = 0$ für alle $h \in \mathbb{R}^{n \times 1}$ ↓ geometrische Interpretation

d.h. $\underbrace{f'(\varphi(u))}_{\substack{\text{Gradient} \\ \text{von } f \text{ in } \varphi(u)}}(\underbrace{\varphi'(u)(h)}_{\substack{\text{Tangentialvektor} \\ \text{in } \varphi(u)}}) = 0$ für $h \in \{e_1, \dots, e_m\}$

$$\varphi'(u)(e_i) = \partial_i \varphi(u)$$

d.h. $\nabla f(x) \cdot \underbrace{\partial_i \varphi(u)}_{\text{Tangententialvektor}} = 0$ für $i=1, \dots, m$

d.h. $\nabla f(x) \perp$ Tangentialraum an x im Punkt x

ist gewisser. Version der norm. Optimalitätsbedingung $\nabla(f \circ \varphi)(u) = 0$

Beispiel:

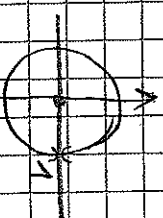
$$f(x) = 2x_1 + x_2$$

$$\varphi(x) = \begin{pmatrix} \cos x \\ \sin x \end{pmatrix}$$

$$x \in (0, 2\pi) = A$$

$$f(\varphi(x)) = 2\cos x + \sin x$$

$$x = \varphi(A) =$$



$$\nabla f(\varphi(x)) = \frac{d}{dx} f(\varphi(x)) = -2\sin x + \cos x = 0$$

$$= -2x_2 + x_1$$

$$\leadsto x_1 = 2x_2$$

$$\text{außer dem } x_1^2 + x_2^2 = 1$$

also

$$4x_2^2 + x_2^2 = 1 \leadsto x_2 = \pm \frac{1}{\sqrt{5}}$$

$$\text{krit. Punkte } \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, -\frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

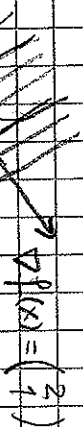
mit Einsetzen

$$\sqrt{5} \quad -\sqrt{5}$$

Randkandidat:

$$f\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = 2 \quad \text{nicht optimal/unkritisch} \quad \sqrt{5} / -\sqrt{5}$$

$\leadsto$ Max, Min gefunden



$$f(x) = c \Leftrightarrow x_2 = c - 2x_1$$

$\nabla f \perp$ Tangentialraum

2)

Nur wissen schon: $\phi: \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{k \times 1}$

$\nabla \phi_i(x)$ steht senkrecht auf dem Tangentialraum der Nullmenge $\{x \in \mathbb{R}^{n \times 1} \mid \phi_i(x) = 0\}$ im Punkt x

und $X = \{x \in \mathbb{R}^{n \times 1} \mid \phi(x) = 0\} = \bigcap_{i=1, \dots, k} \{\phi_i = 0\}$ ist Schnitt der Nullmengen $\{\phi_i = 0\}$ $i=1, \dots, k$

$\Rightarrow$ der Tangentialraum in $x \in X$ ist senkrecht zu $\text{span}\{\nabla \phi_1(x), \dots, \nabla \phi_k(x)\}$ "Normalraum" $\text{an } x \text{ in } X$

notw. Optimalitätsbedingung für $f: X \rightarrow \mathbb{R}$: $\nabla f(x) \perp$ Tangentialraum in $x \in X$

$$\Leftrightarrow \nabla f(x) \in \text{Normalraum in } x \in X$$

$$\Leftrightarrow \nabla f(x) = \lambda_1 \nabla \phi_1(x) + \dots + \lambda_k \nabla \phi_k(x)$$

für Parameter $\lambda_1, \dots, \lambda_k$

(sog. Lagrange-Multiplikatoren)

andere ausgedrückt mit

Lagrange Funktion: $L(x, \lambda) = f(x) - \sum_{i=1}^k \lambda_i \phi_i(x)$

notwendige Optimalitätsbedingung:

$$L'(x, \lambda) = 0 \quad n+k \text{ Gleichungen}$$

für $n+k$ Unbekannte

$$\text{denn } 0 = \frac{\partial}{\partial x_i} (x, \lambda) = \frac{\partial}{\partial x_i} f(x) - \sum_{j=1}^k \lambda_j \frac{\partial}{\partial x_i} \phi_j(x) \Rightarrow \nabla f(x) = \sum_{j=1}^k \lambda_j \nabla \phi_j(x) \quad \text{d.h. } \nabla f \in \text{Tangraum}$$

$$0 = \frac{\partial}{\partial x_i} L(x, \lambda) = -\phi_j(x) \Rightarrow \phi(x) = 0 \quad \text{d.h. Nebenbedingung erfüllt}$$

Satz 10. Sei $\mathcal{O} \subset \mathbb{R}^{n+k}$ offen und $f \in C^1(\mathcal{O}, \mathbb{R})$, $\phi \in C^1(\mathcal{O}, \mathbb{R}^{k \times 1})$ für ein $k \in \mathbb{N}$.

Wahr sei $\text{rang } \phi' = k$ in $\mathcal{O}$ und $X = \{x \in \mathcal{O} \mid \phi(x) = 0\}$. Hat $\frac{\partial f}{\partial x}$ ein lokales

Extremum in $x \in X$, so gibt es "Lagrange-Multiplikatoren" $\hat{\lambda} \in \mathbb{R}^{k \times 1}$, so dass

die Lagrange-Funktion $L: \mathcal{O} \times \mathbb{R}^{k \times 1} \rightarrow \mathbb{R}$ mit $L(x, \lambda) = f(x) - \lambda \cdot \phi(x)$ in $(\hat{x}, \hat{\lambda})$

einen kritischen Punkt hat: $L'(\hat{x}, \hat{\lambda}) = 0$.

Der Beweis benutzt die Aussage von oben und den Satz über implizite Funktionen

dann man die k Gleichungen $\phi_1(x) = 0, \dots, \phi_k(x) = 0$

in der Nähe der Lösung $\hat{x}$ nach k der Variablen auflösen kann

Beispiel: $n=5$ $k=2$

$$O(X_1, \dots, X_n)$$

$$D_2(x_1, \dots, x_5) = 0$$

3 aufgestellt nach

$$x_2 = a(x_1, x_3, x_4)$$

$$X = (x_1, x_2, x_3, x_4)$$

in der Nähe von ~~2~~

June 1st

$$f(x_1, x_2, x_3)$$

$$\begin{pmatrix} x_1 \\ a(x_2, x_3, x_4) \\ x_3 \\ x_4 \\ b(x_1, x_3, x_4) \end{pmatrix}$$

Parameterisierung von $X = \mu + \sigma Z$

in der Nähe von X

also

$$0(x_1, x_2, x_3) = 0$$

Details from site report and FamilySearch ... copied 11

Beispiel für Optimierung mit Nebenbedingungen:

$A \in \mathbb{R}^{m \times n}$ sei symmetrisch

$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$

$$x \mapsto x^T A x$$

$$X = \{x \in \mathbb{R}^n \mid x^T x \leq 1\} \quad \text{abg.} \quad \text{euklid. Einheitskugel}$$

Suche Min/Max von f

1) krit. Punkte in X° : $\nabla f(x) = 0$

$$\frac{\partial}{\partial x_k} \sum_{i=1}^n x_i x_j A_{ij} = \sum_{i=1}^n \left(A_{ij} x_i \frac{\partial}{\partial x_k} x_j + A_{ij} x_j \frac{\partial}{\partial x_k} x_i \right) = \sum_{i=1}^n A_{ik} x_i + \sum_{j=1}^n A_{kj} x_j$$

$$= \delta_{kj} = \begin{cases} 1 & k=j \\ 0 & \text{sonst} \end{cases} = \delta_{ik} = A_{ki} x_i \quad \text{da } A = A^T$$

* Kronecker-Delta

$$= \sum_{i=1}^n A_{ki} x_i = (A x)_k$$

also $\nabla f(x) = 2Ax$

erwünschte Punkte: $\{x \in X^\circ \mid Ax = 0\} = \text{Kern } A \cap X^\circ$ Frühantwort abend: 0

2) Randbetrachtung : $\partial X = \{x \in \mathbb{R}^{n \times 1} \mid \underbrace{x^T x - 1}_{\neq 1} = 0\}$

$\leadsto$ Suche Optimum unter Nebenbedingung $\{ \phi = 0 \}$ d.h. suche krit. Punkte von

$$L(x, \lambda) = f(x) - \lambda \phi(x) \quad \phi(x) = x^T x - 1 = x^T I x - 1 \quad \leadsto \nabla \phi(x) = 2I x = 2x$$

$\xrightarrow{\text{Einsubstitut.}} \text{symm.}$

$$0 = \nabla_x L(x, \lambda) = 2Ax - 2\lambda x, \quad 0 = \nabla_\lambda L(x, \lambda) = \phi(x) = \|x\|^2 - 1$$

$\uparrow$ x-Gradient

also $Ax = \lambda x, \quad \|x\| = 1 \quad \text{d.h. kritische Punkte von } f \text{ sind Eigenvektoren}$

da A symm. gibt es n Eigenvektoren
Lagrange multipl. sind zugeh. Eigenwerte

Hauptideiwerte dort : $x^T Ax = x^T (\lambda x) = \lambda x^T x = \lambda \quad \text{d.h. die Eigenwerte}$

Optima auf ∂X : $\lambda_{\min}$ (Minimum) $\lambda_{\max}$ (Maximum)

3) A pos. def. $\rightarrow$ $\max f = \lambda_{\max}, \min f = 0$ $\rightarrow$ scarst $\rightarrow$ $\max f = \lambda_{\max}$
 A neg. def. $\rightarrow$ $\min f = \lambda_{\min}, \max f = 0$ $\rightarrow$ $\text{auf } \partial X$
 A indef. $\rightarrow$ $\min f = \lambda_{\min}, \max f = 0$ $\rightarrow$ $\text{auf } \partial X$