3b. Abstiegsverfahren

Konvergenz

Optimierung SoSe 2020

Dr. Alexey Agaltsov

Plan

- Globale Konvergenz
- Konvergenzgeschwindigkeit
- Gradientenverfahren

Globale Konvergenz

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

$f \in C^1(\mathbb{R}^n)$ strikt konvex

- Angenommen, x_* ist eine (eindeutige) optimale Lösung
- Ein Abstiegsverfahren erzeugt eine Folge $x_1, x_2, \ldots$ für ein gegebenes x_0
- Wir bezeichnen das Verfahren als global konvergent falls $\forall x_0 \in \mathbb{R}^n$:

$$x_k \to x_*$$

- $\nabla f(x_k) \to 0$ ist einfacher zu kontrollieren (x_* ist nicht bekannt)

Lemma 3.4. Globale Konvergenz

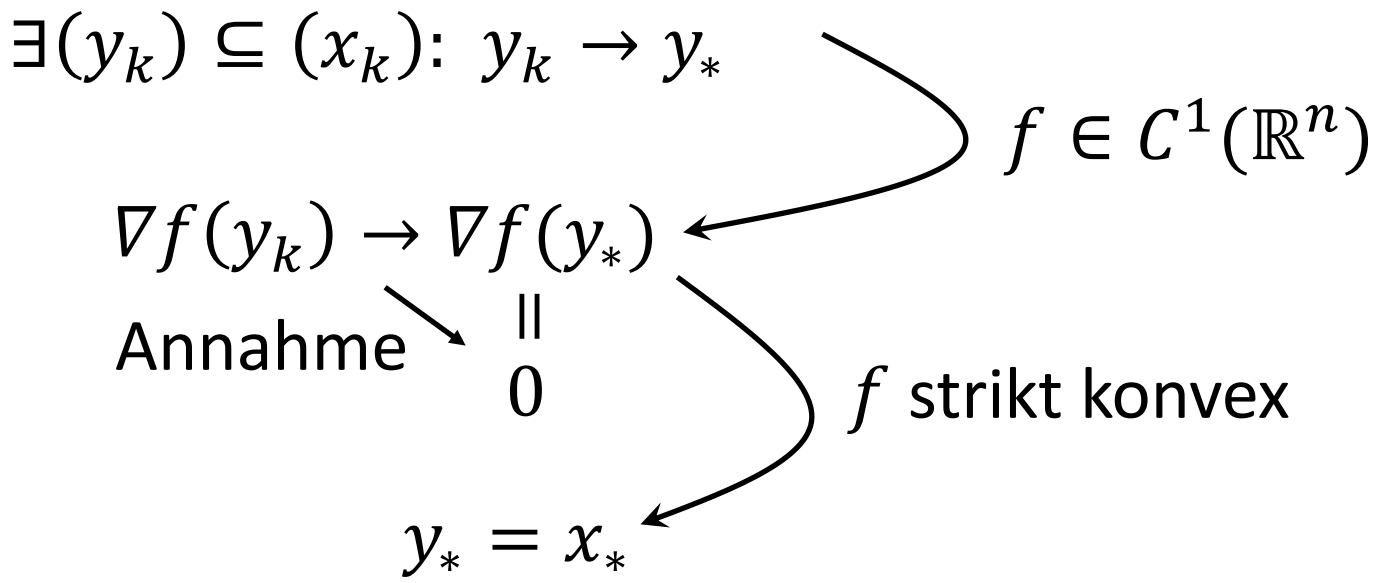
- Sei $f \in C^1(\mathbb{R}^n)$ strikt konvex und x_* ihr Minimum
- Sei $x_0, x_1, \ldots$ ist eine beschränkte Folge mit $\nabla f(x_k) \to 0$

Dann $x_k \to x_*$

Bemerkung: Man kann zeigen, dass die Beschränktheit von (x_k) folgt aus der anderen Voraussetzungen

Beweis

Behauptung: Ist y_* ein Häufungspunkt von (x_k), so ist $y_* = x_*$



□

Konvergenz der Gradienten

- Als nächstes werden wir die Bedingungen für $\nabla f(x_k) \to 0$ herleiten
- Wir wenden ein Abstiegsverfahren (oder kurz AV) an und erhalten:

  Näherungslösungen (x_k)

  Abstiegsrichtungen (d_k)

  Schrittweiten (α_k)
- Hierbei ist der Winkel θ_k zwischen d_k und $-\nabla f(x_k)$ wichtig

$$\cos\theta_k = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|_2 \|d_k\|_2}$$

Satz 3.5. Satz von Zoutendijk

Sei $f \in C^1(\mathbb{R}^n)$. Seien $(x_k), (d_k), (\alpha_k)$ mit einem AV erzeugt. Angenommen:

- $\inf_S f > -\infty$, wobei $S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$
- (α_k) erfüllen die Wolfe-Bedingungen
- $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2$ mit $L > 0\ \forall x, y \in S$

Dann ist die Zoutendijk-Bedingung erfüllt:

$$\sum_{k=0}^{\infty} \cos^2(\theta_k)\, \|\nabla f(x_k)\|_2^2 < \infty$$

$\Rightarrow$ [NW, Theorem 3.2]

Beweis

Behauptung: $\alpha_k \geq \frac{\nu-1}{L} \frac{\nabla f(x_k)^T d_k}{\|d_k\|_2^2}$ mit $\nu \in (0,1)$

$\in (0,1)$

$$\nu \nabla f(x_k)^T d_k \leq \nabla f(x_{k+1})^T d_k$$ *Krümmungsbedingung*

subtrahiere $\nabla f(x_k)^T d_k$

$$0 < \underbrace{(\nu - 1)}_{<0} \underbrace{\nabla f(x_k)^T d_k}_{<0} \leq \left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)^T d_k$$ Cauchy-Schwarz

$$\leq \|\nabla f(x_{k+1}) - \nabla f(x_k)\|_2 \, \|d_k\|_2$$

Annahme

$$\leq L \|x_{k+1} - x_k\|_2 \|d_k\|_2$$

$x_{k+1} = x_k + \alpha_k d_k$

$$= L \alpha_k \|d_k\|_2^2$$

Beweis

$$\alpha_k \geq \frac{\nu-1}{L}\frac{\nabla f(x_k)^T d_k}{\|d_k\|_2^2}$$

Behauptung: $\sum_{k=0}^{\infty} \cos^2(\theta_k)\,\|\nabla f(x_k)\|_2^2 < \infty$

$\in (0,1)$

$$f(x_{k+1}) \leq f(x_k) + \bar{\mu}\alpha_k \nabla f(x_k)^T d_k \quad \textit{Armijo-Bedingung}$$

$$\leq f(x_k) + \mu\frac{\nu-1}{L}\frac{\left(\nabla f(x_k)^T d_k\right)^2}{\|d_k\|_2^2}$$

$$\leq f(x_k) + \mu\frac{\nu-1}{L}\cos^2(\theta_k)\,\|\nabla f(x_k)\|_2^2$$

summiere über k

$$\underbrace{\mu\frac{1-\nu}{L}\sum_{k=0}^{N}\cos^2(\theta_k)\,\|\nabla f(x_k)\|_2^2}_{N\to\infty} \leq f(x_0) - f(x_{N+1})$$

$$\leq f(x_0) - \inf_{x\in S} f(x) < \infty$$

Annahme

Richtungswahl

- Angenommen, f und (α_k) erfüllen die Bedingungen von Satz 3.5

$$\Longrightarrow \quad \sum_{k=0}^{\infty} \cos^2(\theta_k) \, \|\nabla f(x_k)\|_2^2 < \infty$$

- Für welche (d_k) ist das Abstiegsverfahren global konvergent?

$$\cos(\theta_k) \geq \delta > 0 \quad \forall k \quad \Longrightarrow \quad \nabla f(x_k) \to 0$$

Lemma 3.4

globale Konvergenz
für strikt konvexe f

Beispiel: Gradientenverfahren

$$d_k = -\nabla f(x_k)$$

$$\cos(\theta_k) = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|_2 \|d_k\|} = 1$$

global konvergent

Beispiel: Newton-artige Verfahren

$$d_k = -B_k \nabla f(x_k)$$

$$mI \preccurlyeq B_k \preccurlyeq MI$$

$$M > m > 0$$

$$\cos(\theta_k) = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\|_2 \|d_k\|_2}$$

$$= \frac{\nabla f(x_k)^T B_k \nabla f(x_k)}{\|\nabla f(x_k)\|_2 \|B_k \nabla f(x_k)\|_2} \geq \frac{m}{M} > 0$$

global konvergent

Plan

- Globale Konvergenz
- Konvergenzgeschwindigkeit
- Gradientenverfahren

Konvergenzgeschwindigkeit

- Sei $x_0, x_1, ..$ eine durch ein Optimierungsverfahren erzeugte Folge
- Der Fortschritt des Verfahrens wird durch eine Fehlerfunktion gemessen:

$$e(x_k) = \|x_k - x_*\|_2$$

$$e(x_k) = f(x_k) - f_*$$

$$e(x_k) = \|\nabla f(x_k)\|_2$$

Nächstes Ziel: Wie kann die Konvergenzgeschwindigkeit von $z_k = e(x_k)$ gegen 0 qualitativ gemessen werden?

Konvergenzordnung

- Sei $(z_k) \subseteq \mathbb{R}$ mit $z_k \to z_*$
- z_k konvergiert mit der Q-Ordnung (wenigstens) p, falls $\exists\, c > 0$:

$$|z_{k+1} - z_*| \leq c|z_k - z_*|^p \qquad \forall k \geq 0$$

Quadratische Konvergenz

- Für $p = 2$ spricht man von der quadratischen Konvergenz

$$|z_{k+1} - z_*| \leq c|z_k - z_*|^2$$

- $z_k = 0.1^{2^k}$ konvergiert quadratisch gegen $z_* = 0$: $|z_{k+1}| = |z_k|^2$

$$0.1,\ 0.01,\ 0.0001,\ 0.00000001,\ \ldots \rightarrow 0.000000$$

die Anzahl von signifikanten Dezimalstellen wird $\approx$ verdoppelt in jedem Schritt

Lineare Konvergenz

- Für $p = 1$ unterscheidet man drei Fälle je nach β:

$$\limsup_{k\to\infty} \frac{|z_{k+1} - z_*|}{|z_k - z_*|} = \beta$$

$\beta \in (0,1)$ lineare Konvergenz

$\beta = 1$ unterlineare Konvergenz

$\beta = 0$ superlineare Konvergenz

Lineare Konvergenz

- $z = 0.1^k$ konvergiert linear gegen $z_* = 0$:

$$\frac{|z_{k+1}|}{|z_k|} = \frac{0.1^{k+1}}{0.1^k} = 0.1 \qquad \Rightarrow \beta = 0.1$$

$$0.1,\ 0.01,\ 0.001,\ 0.0001,\ \ldots \rightarrow 0.000000$$

Eine weitere signifikante Dezimalstelle in jedem Schritt

Plan

- Globale Konvergenz
- Konvergenzgeschwindigkeit
- Gradientenverfahren

Gradientenverfahren

1. *Initialisierung:* Startwert x_0, Toleranzwert ϵ
2. **for** $k = 0,1,2, \ldots$ **do**:
3. **if** $\|\nabla f(x_k)\|_2 < \epsilon$ **then** break
4. bestimme eine Schrittweite α_k
5. $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$
6. **end for**

- Minimierungsregel:

$$\alpha_k \in \operatorname{Argmin}_{\alpha \geq 0} f\left(x_k - \alpha \nabla f(x_k)\right)$$

Satz 3.6. Konvergenzrate des GVs

Seien $f \in C^2(\mathbb{R}^n)$, sei x_* ein globales Minimum und $f_* = f(x_*)$

Seien $x_0 \in \mathbb{R}^n$ und $S = \{x: f(x) \leq f(x_0)\}$. Angenommen:

1. $mI \preccurlyeq \nabla^2 f(x) \preccurlyeq MI$ mit $M \geq m > 0 \;\; \forall x \in S$ (starke Konvexität) $\rightarrow \kappa\left(\nabla^2 f(x)\right) \leq \frac{M}{m}$
2. Seien $(x_k), (\alpha_k)$ mit dem GV mit der Minimierungsregel erzeugt

Dann gilt:

$$f(x_{k+1}) - f_* \leq c(f(x_k) - f_*)$$

$$c = 1 - \frac{m}{M}$$

Beweis

Behauptung: $f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{M}{2} \|y - x\|_2^2 \quad \forall x, y \in S$

Taylor-Formel für $x, y \in S$ $\Rightarrow \exists z \in [x, y]$

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} (y - x)^T \underbrace{\nabla^2 f(z)}_{\preccurlyeq MI} (y - x)$$

Beweis

$$f(x_{k+1}) \leq f(x_k + \alpha \overbrace{d_k}^{-\nabla f(x_k)}) \quad \forall \alpha \geq 0 \qquad \text{(Minimierungsregel)}$$

letzte Abschätzung

$$\leq f(x_k) + \alpha \nabla f(x_k)^T d_k + \frac{M}{2}\alpha^2 \|d_k\|_2^2$$

$$= f(x_k) + \left(-\alpha + \frac{M}{2}\alpha^2\right)\|\nabla f(x_k)\|_2^2 \qquad \left(d_k = -\nabla f(x_k)\right)$$

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2M}\|\nabla f(x_k)\|_2^2 \qquad \alpha = \frac{1}{M}$$

subtrahiere f_*

$$f(x_{k+1}) - f_* \leq f(x_k) - f_* - \frac{1}{2M}\underbrace{\|\nabla f(x_k)\|_2^2}_{\geq 2m(f(x_k) - f_*) \text{ (Lemma 2.22)}}$$

$$f(x_{k+1}) - f_* \leq c(f(x_k) - f_*) \text{ mit } c = 1 - \frac{m}{M}$$

□

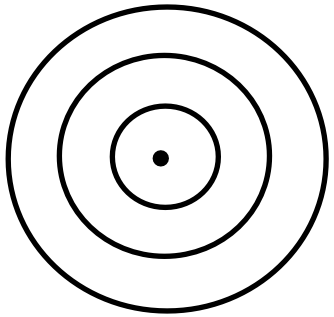
Interpretation

$$mI \preccurlyeq \nabla^2 f \preccurlyeq MI \text{ auf } S \quad \Longrightarrow \quad f(x_k) - f_* \leq \left(1 - \frac{m}{M}\right)^k (f(x_0) - f_*)$$

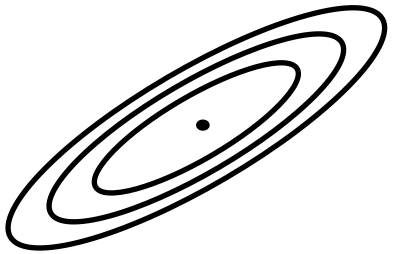
$$S = \{x : f(x) \leq f(x_0)\}$$

Satz 2.23

$$\kappa(S) \leq \frac{M}{m}$$



schnelle Konvergenz



langsame Konvergenz

Quadratische Zielfunktion

Minimiere $f(x) = \frac{1}{2}x^T Q x - c^T x + r$ über $x \in \mathbb{R}^n$

$$Q \in \mathbb{S}^n_{\succ}$$

$$\nabla^2 f(x) = Q$$

Min. Eigenwert Max. Eigenwert

$$mI \preccurlyeq Q \preccurlyeq MI$$

Satz 3.6

$$f(x_k) - f_* \leq c^k (f(x_0) - f_*) \qquad k \geq 0$$

$$c = 1 - \frac{1}{\kappa(Q)} \qquad \kappa(Q) = \frac{M}{m}$$ Kondition von Q

Kondition und Konvergenzrate

$$f(x_k) - f_* \leq c^k (f(x_0) - f_*), \quad c = 1 - \frac{1}{\kappa(Q)}$$

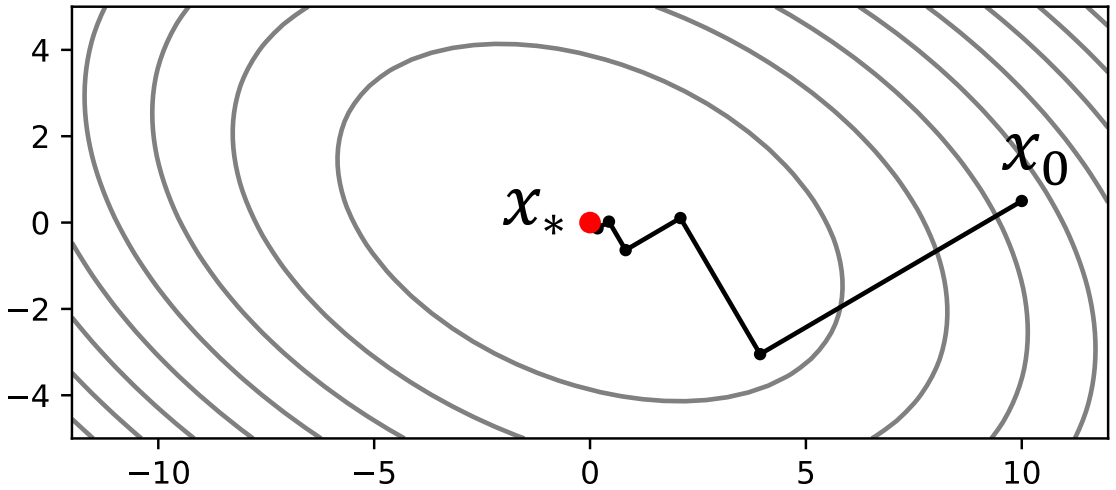
- Aufgabe. $N = \log(0.1)/\log(c)$ Schritte pro signifikante Dezimalstelle
- $N \approx -\kappa(Q)/\log(0.1) \approx 0.43\kappa(Q)$ für großes $\kappa(Q)$

N	c	$\kappa(Q)$
1	0.1	1.11
10	0.79	4.86
100	0.98	43.93
1000	0.998	434.79

Beispiel

$$f(x) = x^T Q x, \qquad Q = \begin{bmatrix} 10 & 5 \\ 5 & 20 \end{bmatrix}$$

$$\kappa(Q) = 2.78, \qquad c = 0.64$$



$$x_0 = (10, 0.5)$$

k	$\|x_k - x_*\|_2$	$f(x_k) - f_*$	c_k
0	10	530	
1	5	110	0,21
2	2,1	23	0,21
3	1	4,8	0,21
4	0,44	1	0,21
5	0,22	0,21	0,21
6	0,092	0,044	0,21
7	0,046	0,0093	0,21
8	0,019	0,0019	0,21
9	0,0097	0,00041	0,21

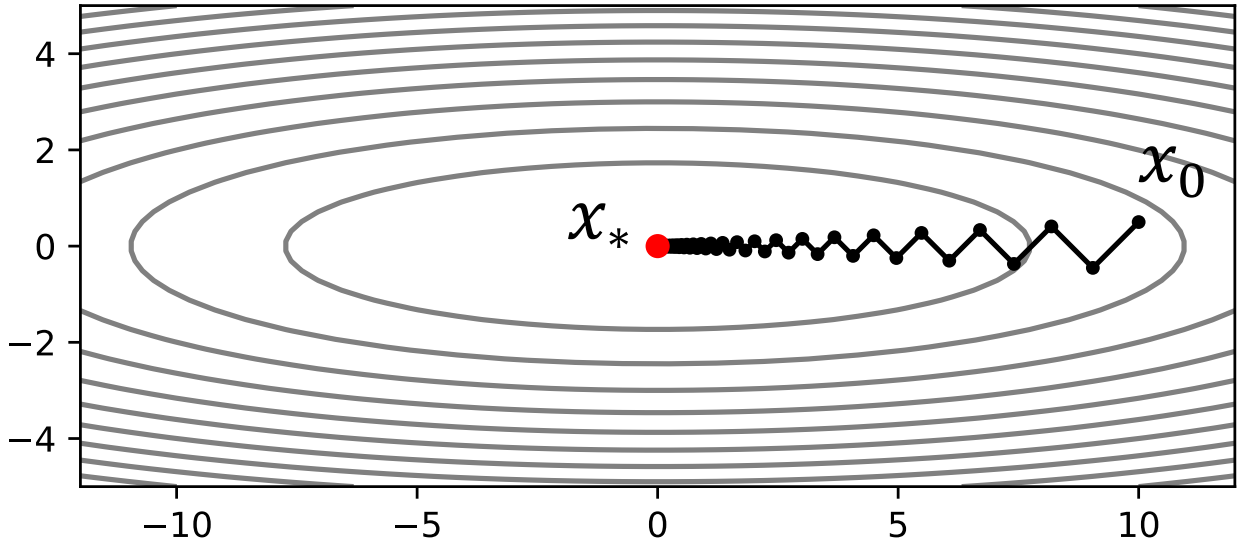
$$f(x_k) - f_* = c_k (f(x_{k-1}) - f_*)$$

$$c_k \approx 0{,}21 < 0.64$$

Beispiel

$$f(x) = x^T Q x, \qquad Q = \begin{bmatrix} 1 & 0 \\ 0 & 20 \end{bmatrix}$$

$$\kappa(Q) = 20, \qquad c = 0.95$$



$$x_0 = (10, 0.5)$$

k	$\|x_k - x_*\|_2$	$f(x_k) - f_*$	c_k
0	10	52	
1	9,1	43	0,82
2	8,2	35	0,82
3	7,4	29	0,82
4	6,7	24	0,82
5	6,1	19	0,82
6	5,5	16	0,82
7	5	13	0,82
8	4,5	11	0,82
9	4,1	8,7	0,82

$$f(x_k) - f_* = c_k(f(x_{k-1}) - f_*)$$

$$c_k \approx 0{,}82 < 0.95$$

Koordinatentransformation

$$mI \preccurlyeq \nabla^2 f(x) \preccurlyeq MI \implies f(x_{k+1}) - f_* \leq \left(1 - \frac{m}{M}\right)^k (f(x_0) - f_*)$$

$$\left.\begin{aligned} \bar{x} &:= P^{1/2} x \text{ mit } P \in \mathbb{S}^n_{\succ} \\ \bar{f}(\bar{x}) &:= f\left(P^{-1/2}\bar{x}\right) = f(x) \end{aligned}\right] \quad \nabla^2 \bar{f}(\bar{x}) = P^{-1/2} \nabla^2 f(x) P^{-1/2}$$

- Ist $P \approx \nabla^2 f(x_*)$, so $\nabla^2 \bar{f}(\bar{x}) \approx I$ für $\bar{x} \approx \bar{x_*} = P^{1/2} x_*$

  Gradientenverfahren in $\bar{x}$ konvergiert schneller als in x

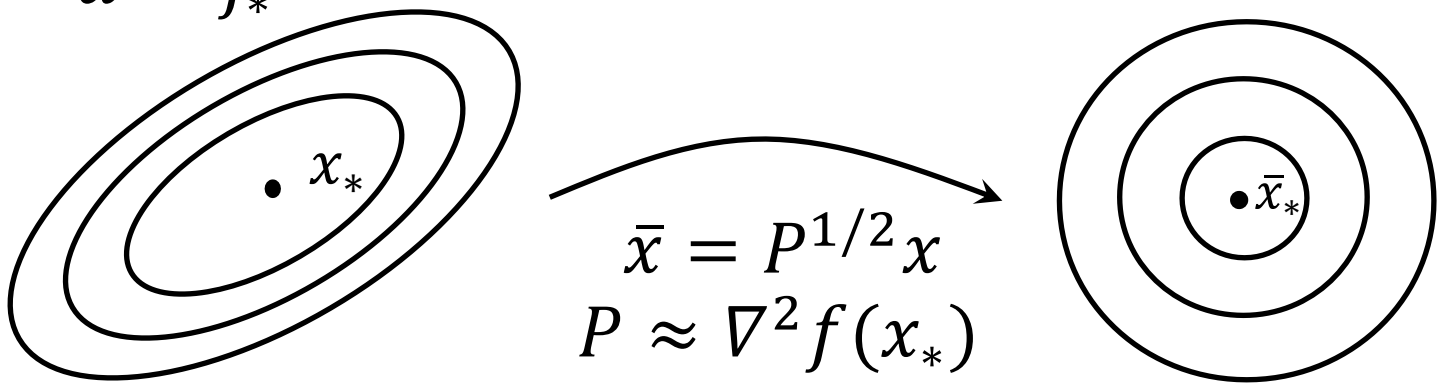
- Gradientenverfahren in $\bar{x}$ entspricht der Methode des steilsten Abstiegs in x bezüglich der Norme $\|x\|_P = \sqrt{x^T P x}$

Geometrische Interpretation

$$f(x) \approx f_* + \tfrac{1}{2}(x - x_*)^T \nabla^2 f(x_*)(x - x_*), \; x \approx x_*$$

$$\{x: f(x) \leq \alpha\} \approx \{x: (x - x_*)^T \nabla^2 f(x_*)(x - x_*) \leq 2(\alpha - f_*)\}$$

$$\alpha \approx f_*$$



$$\{\bar{x}: \bar{f}(\bar{x}) \leq \alpha\} \approx \{\bar{x}: \|\bar{x} - \bar{x}_*\|_2^2 \leq 2(\alpha - f_*)\}$$

Zusammenfassung

- Globale Konvergenz
- Konvergenzgeschwindigkeit
- Gradientenverfahren

Nächstes Video

- 4a. Newtonartige Verfahren: Das Newton-Verfahren