

[A2]

a) $x^1, \dots, x^m \in \mathbb{R}^n$

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \sum_{k=1}^m \|x^k - x\|_2^2$$

$$f(x) = \sum_{k=1}^m \|x^k - x\|_2^2 = \sum_{k=1}^m (x^k - x)^T (x^k - x)$$

$$= \sum_{k=1}^m (x^T x - 2x^T x^k + (x^k)^T x^k)$$

$$= mx^T x - 2x^T \sum_{k=1}^m x^k + \underbrace{\sum_{k=1}^m (x^k)^T x^k}_{\text{unabh. von } x}$$

$$\Rightarrow \nabla f(x) = \nabla (mx^T x - 2x^T \sum_{k=1}^m x^k)$$

$$mx^T x = m \cdot \sum_{i=1}^n (x_i)^2 \Rightarrow \nabla (mx^T x) = \begin{pmatrix} 2mx_1 \\ \vdots \\ 2mx_n \end{pmatrix} = 2mx$$

i-te Komponente

$$2x^T \sum_{k=1}^m x^k = 2 \cdot \sum_{i=1}^n \left(x_i \sum_{k=1}^m x_i^k \right)$$

i-te Komponente

$$\Rightarrow \nabla \left(2x^T \sum_{k=1}^m x^k \right) = 2 \begin{pmatrix} \sum_{k=1}^m x_1^k \\ \vdots \\ \sum_{k=1}^m x_n^k \end{pmatrix} = 2 \sum_{k=1}^m x^k$$

$$\Rightarrow \nabla f(x) = 2mx - 2 \sum_{k=1}^m x^k$$

$$\nabla f(x) = 0 \iff 2mx = 2 \sum_{k=1}^m x^k$$

$$\iff x = m^{-1} \sum_{k=1}^m x^k$$

D.h. $x^* := \frac{1}{m} \sum_{k=1}^m x^k$ ist Minimum von f .

Alternative Argumentation über den Erwartungswert als „konstante Approximation“ einer Zufallsvariable,

wobei das Problem $\min_{c \in \mathbb{R}} \mathbb{E}[(X - c)^2]$ betrachtet wurde

$c^* = \mathbb{E}[X]$ ist minimales dieses Problems.

(Stochastik 2)

$\Rightarrow \min_{x \in \mathbb{R}^n} f(x)$ ist n -dimensionale Variante des Problems

②
b)

$$f(x) = \frac{1}{2} x^T A x + x^T b + c = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \cdot x_i \cdot x_j + \sum_{i=1}^n x_i \cdot b_i + c$$

Allgemeine Ableitung nach $1 \leq k \leq n$ ist somit:

$$\frac{\partial}{\partial x_k} Ax = \sum_{i=1}^n a_{ki} \cdot x_k \cdot x_i + \sum_{j=1}^n a_{jk} \cdot x_j \cdot x_k = \frac{1}{2} (Ax)_k + \frac{1}{2} (A^T x)_k$$

$$\frac{\partial}{\partial x_k} x^T b = b_k$$

$$\Rightarrow \nabla f(x) = \frac{1}{2} (Ax) + \frac{1}{2} (A^T x) + b$$

stationäre Punkte sind die $x \in \mathbb{R}^n$,

die $0 = \nabla f(x)$ erfüllen, also für die gilt:

$$\frac{1}{2} (A + A^T)x + b = 0. \quad (1)$$

Da $A + A^T$ nicht invertierbar sein muss, kann es mehrere oder auch keine $x \in \mathbb{R}^n$ geben, die (1) erfüllen.

Die Menge der stationären Punkte ist: $M = \{x \in \mathbb{R}^n : \frac{1}{2} (A^T + A)x = -b\}$

③

a)

$$\text{Nutzen } \log_\alpha(x) = \frac{\ln(x)}{\ln(\alpha)} \quad \text{für } \alpha > 0 \quad (*)$$

$$h(x) = f(g(x)), \quad f: \mathbb{R} \rightarrow \mathbb{R}, \quad g: \mathbb{R}^n \rightarrow \mathbb{R}$$

$$f(x) = \log_\alpha(x) \stackrel{(*)}{=} \ln(x) \cdot \frac{1}{\ln(\alpha)} \quad g(x) = \sum_{i=1}^n e^{x_i}$$

Damit sind

$$\nabla f(x) = f'(x) = \frac{1}{x} \cdot \frac{1}{\ln(\alpha)}$$

und

$$\nabla g(x) = \begin{pmatrix} e^{x_1} \\ e^{x_2} \\ \vdots \\ e^{x_n} \end{pmatrix}$$

$$\text{Und somit: } \nabla h(x) = \nabla g(x) \cdot \nabla f(g(x)) = \begin{pmatrix} e^{x_1} \\ \vdots \\ e^{x_n} \end{pmatrix} \cdot \frac{1}{\sum_{i=1}^n e^{x_i} \cdot \ln(\alpha)}$$

$\nabla h(x)$ kann nie 0 sein, da $e^x \neq 0 \quad \forall x \in \mathbb{R}$ gilt. □

(3)

b) Zur Lösung betrachten wir dies als Ausgleichsproblem und nutzen Aussagen zur Singulärwertzerlegung:

Da $\text{Rang}(A) < n$ und damit $\text{Rang}(A^T) < n$ folgt:

$\text{Rang}(A^T A) < n$. Damit ist das LGS $A^T A x = A^T b$ (1)

Es ist $A^T A \in \mathbb{R}^{n \times n}$. Da $\text{Rang}(A) < n$, ist

mindestens ein Eigenwert von $A^T A$ gleich 0, d.h.

ein Singulärwert von A ist 0. Damit ist $\text{Rang}(\Sigma) \leq n-1$

und damit hat die Moore-Penrose Pseudoinverse $A^+ = V \Sigma^+ U^H$ $\text{Rang} \leq n-1$.

Sei nun $x^* = A^+ b$ eine Lösung von (1) mit mindestens einem Freiheitsgrad. Sei dieser an Stelle j.

Nehmen wir nun an, $x^* = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{pmatrix}$ wäre ein starker lokaler Minimierer, d.h.

$f(x^*) < f(y)$, $\|x^* - y\| < \varepsilon$ Wählen $\hat{x} = x^* + e_j \cdot \frac{\varepsilon}{2}$.

Dann gilt \hat{x} ist Minimierer von $\|Ax - b\|_2$, d.h. $f(x^*) = f(\hat{x})$, aber $\|\hat{x} - x^*\| = \frac{\varepsilon}{2} < \varepsilon$ - Widerspruch.

□

c) z.B. $\langle \nabla f(g(s)), g'(s) \rangle = 0$.

$$\langle \nabla f(g(s)), g'(s) \rangle = (g'(s))^T \cdot \nabla f(g(s)) = \nabla f(g(s)) = 0, \text{ denn:}$$

Auf jeder Kurve $g(s)$ ist f konstant (per Definition),
Daher ist $\nabla f(g(s)) = 0$ in jedem Punkt. □