Partielle Ableitungen von linearen und quadratischen Funktionen mehrerer Variablen

Dieses Blatt ist eine Ergänzung für die Begeisterten, die Details über Seite 201 im Skriptum und Hinweise für Beispiel 4 auf dem 12. Übungsblatt gerne hätten.

Beispiel: Gegeben seien $n \in \mathbb{N}$ und ein Vektor $\boldsymbol{v} \in \mathbb{R}^n$ mit Komponenten

$$\boldsymbol{v} = \{v_i\}_{i=1}^n = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Für $\boldsymbol{x} \in \mathbb{R}^n$ mit Komponenten

$$\boldsymbol{x} = \{x_i\}_{i=1}^n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

sei $f(\boldsymbol{x})$ definiert durch das Skalarprodukt (eine lineare Form)

$$f(\boldsymbol{x}) = \boldsymbol{v} \cdot \boldsymbol{x} = \boldsymbol{v}^\top \boldsymbol{x} = v_1 x_1 + v_2 x_2 + \cdots v_n x_n, \quad D = \mathbb{R}^n, \quad B \subset \mathbb{R}$$

Die partielle Ableitung von $f(\boldsymbol{x})$ nach x_k mit $1 \leq k \leq n$ ist

$$\partial_{x_k} f(\boldsymbol{x}) = \partial_{x_k}(v_1 x_1 + \cdots + \ v_k x_k \ + \cdots + v_n x_n) = v_k$$

und daher ist der Vektor dieser partiellen Ableitungen (der *Gradient* $\nabla f(x)$ von $f(\boldsymbol{x})$) gegeben durch

$$\nabla f(x) = \ \{\partial_{x_k} f(\boldsymbol{x})\}_{k=1}^n = \{v_k\}_{k=1}^n = \boldsymbol{v}$$

Beispiel: Gegeben seien $n \in \mathbb{N}$ und eine Matrix $M \in \mathbb{R}^{n \times n}$ mit Komponenten

$$M = \{m_{i,j}\}_{i,j=1}^n = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,n} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & m_{n,n} \end{bmatrix}$$

Für $\boldsymbol{x} \in \mathbb{R}^n$ mit Komponenten

$$\boldsymbol{x} = \{x_i\}_{i=1}^n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

sei $f(\boldsymbol{x})$ definiert durch das Vektor-Matrix-Vektor Produkt (eine quadratische Form)

$$\begin{aligned} f(\boldsymbol{x}) = \boldsymbol{x}^\top M \boldsymbol{x} = \boldsymbol{x} \cdot (M\boldsymbol{x}) \ &= \ x_1(M\boldsymbol{x})_1 + x_2(M\boldsymbol{x})_2 + \cdots + x_n(M\boldsymbol{x})_n \\ &= \ x_1(m_{1,1}x_1 + m_{1,2}x_2 + \cdots + m_{1,n}x_n) \\ &+ \ x_2(m_{2,1}x_1 + m_{2,2}x_2 + \cdots + m_{2,n}x_n) \\ &\ \ \vdots \\ &+ \ x_n(m_{n,1}x_1 + m_{n,2}x_2 + \cdots + m_{n,n}x_n), \quad D = \mathbb{R}^n, \quad B \subset \mathbb{R} \end{aligned}$$

Die partielle Ableitung von $f(\boldsymbol{x})$ nach x_k mit $1 \le k \le n$ ist

$$\begin{aligned}
\partial_{x_k} f(\boldsymbol{x}) &= \partial_{x_k}\, x_1(m_{1,1}x_1 + \cdots + \; m_{1,k}x_k \; + \cdots + m_{1,n}x_n) \\
&\;\;\vdots \\
&+ \;\partial_{x_k}\, x_k\,(m_{k,1}x_1 + \cdots + \; m_{k,k}x_k \; + \cdots + m_{k,n}x_n) \\
&\;\;\vdots \\
&+ \;\partial_{x_k}\, x_n(m_{n,1}x_1 + \cdots + \; m_{n,k}x_k \; + \cdots + m_{n,n}x_n) \\
&= x_1 m_{1,k} \\
&\;\;\vdots \\
&+ \;(m_{k,1}x_1 + \cdots + m_{k,n}x_n) + x_k m_{k,k} \\
&\;\;\vdots \\
&+ \;x_n m_{n,k} \\
&= (m_{k,1}x_1 + \cdots + m_{k,n}x_n) \\
&+ \;(m_{1,k}x_1 + \cdots + m_{n,k}x_n)
\end{aligned}$$

und daher ist der Vektor dieser partiellen Ableitungen (der *Gradient* $\nabla f(x)$ von $f(\boldsymbol{x})$) gegeben durch

$$\nabla f(x) = \; \{\partial_{x_k} f(\boldsymbol{x})\}_{k=1}^n = \left\{ \begin{array}{c} (m_{k,1}x_1 + \cdots + m_{k,n}x_n) \\ +(m_{1,k}x_1 + \cdots + m_{n,k}x_n) \end{array} \right\}_{k=1}^n = (M + M^\top)\boldsymbol{x}$$

Beispiel: Gegeben seien $m, n \in \mathbb{N}$, $\boldsymbol{b} \in \mathbb{R}^m$ und $A \in \mathbb{R}^{m \times n}$ mit Komponenten

$$\boldsymbol{b} = \{b_i\}_{i=1}^m = \begin{bmatrix} b_1 \\ b_2 \\ \\ \vdots \\ \\ b_m \end{bmatrix} \quad \text{und} \quad A = \{a_{i,j}\}_{i=1,m}^{j=1,n} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,1} & \cdots & a_{2,n} \\ \\ \vdots & \vdots & \cdots & \vdots \\ \\ a_{m,1} & a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

Für $\boldsymbol{x} \in \mathbb{R}^n$ mit Komponenten

$$\boldsymbol{x} = \{x_i\}_{i=1}^n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

sei $\boldsymbol{f}(\boldsymbol{x})$ definiert durch die Euklidischen Norm

$$f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$$

Wegen der Beziehung zwischen der Euklidischen Norm und dem Skalarprodukt,

$$x_1^2 + x_2^2 + \cdots + x_n^2 = \|\boldsymbol{x}\|^2 = \boldsymbol{x}^\top \boldsymbol{x}$$

kann $f(\boldsymbol{x})$ so umgeschrieben werden,

$$f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = (A\boldsymbol{x} - \boldsymbol{b})^\top (A\boldsymbol{x} - \boldsymbol{b})$$

Wegen der Eigenschaft der Transponierten eines Produkts,

$$(A\boldsymbol{x})^\top = \boldsymbol{x}^\top A^\top$$

folgt

$$f(\boldsymbol{x}) = (\boldsymbol{x}^\top A^\top - \boldsymbol{b}^\top)(A\boldsymbol{x} - \boldsymbol{b})$$

und wenn ausmultipliziert,

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top A^\top A\boldsymbol{x} - \boldsymbol{x}^\top A\boldsymbol{b} - \boldsymbol{b}^\top A\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{b}$$

Da $\boldsymbol{x}^\top A^\top \boldsymbol{b}$ ein Skalar ist, ist sie gleich ihre Transponierte. Wegen der obigen Eigenschaft der Transponierten eines Produkts gilt

$$\boldsymbol{x}^\top A^\top \boldsymbol{b} = (\boldsymbol{x}^\top A^\top \boldsymbol{b})^\top = \boldsymbol{b}^\top A\boldsymbol{x} = (A^\top \boldsymbol{b})^\top \boldsymbol{x}$$

Daher kann $f(\boldsymbol{x})$ schließlich folgendermaßen umgeschrieben werden,

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top (A^\top A)\boldsymbol{x} - 2(A^\top \boldsymbol{b})^\top \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{b}$$

Der erste Term in $f(\boldsymbol{x})$ hat die Form $\boldsymbol{x}^\top M\boldsymbol{x}$ für eine Matrix $M = A^\top A$, und die entsprechenden partiellen Ableitungen sind durch das letzte Beispiel gegeben,

$$\begin{gathered}\{\partial_{x_k} \boldsymbol{x}^\top (A^\top A)\boldsymbol{x}\}_{k=1}^n \\ = \{\partial_{x_k} \boldsymbol{x}^\top M\boldsymbol{x}\}_{k=1}^n = (M + M^\top)\boldsymbol{x} \\ = ((A^\top A) + (A^\top A)^\top)\boldsymbol{x} = ((A^\top A) + (A^\top A))\boldsymbol{x} = 2A^\top A\boldsymbol{x}\end{gathered}$$

wobei die Eigenschaft $(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A$ der Transponierten eines Produkts verwendet worden ist. Der zweite Term in $f(\boldsymbol{x})$ hat die Form $-2\boldsymbol{v}^\top \boldsymbol{x}$ für einen Vektor $\boldsymbol{v} = A^\top \boldsymbol{b}$, und die entsprechenden partiellen Ableitungen sind durch das vorletzte Beispiel gegeben

$$\{\partial_{x_k} (A^\top \boldsymbol{b})^\top \boldsymbol{x}\}_{k=1}^n = \{\partial_{x_k} \boldsymbol{v}^\top \boldsymbol{x}\}_{k=1}^n = \boldsymbol{v} = A^\top \boldsymbol{b}$$

Der dritte Term in $f(\boldsymbol{x})$ ist ein Skalar $\boldsymbol{b}^\top \boldsymbol{b}$, der von $\boldsymbol{x}$ nicht abhängt, und daher sind die partiellen Ableitungen dieses Terms Null. Zusammengefasst ist der Vektor der partiellen Ableitungen (der *Gradient* $\nabla f(x)$ von $f(\boldsymbol{x})$) gegeben durch

$$\nabla f(x) = \{\partial_{x_k} f(\boldsymbol{x})\}_{k=1}^n = 2A^\top A\boldsymbol{x} - 2A^\top \boldsymbol{b}$$

Beispiel: Gegeben seien $n \in \mathbb{N}$ und $M \in \mathbb{R}^{n\times n}$ mit Komponenten

$$M = \{m_{i,j}\}_{i,j=1}^n = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,n} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & m_{n,n} \end{bmatrix}$$

Für $\boldsymbol{x} \in \mathbb{R}^n$ mit Komponenten

$$\boldsymbol{x} = \{x_i\}_{i=1}^n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

sei $\boldsymbol{F}(\boldsymbol{x})$ definiert durch das Matrix-Vektor Produkt

$$\boldsymbol{F}(\boldsymbol{x}) = M\boldsymbol{x} = \begin{bmatrix} m_{1,1}x_1 + m_{1,2}x_2 + \cdots + m_{1,n}x_n \\ m_{2,1}x_1 + m_{2,2}x_2 + \cdots + m_{2,n}x_n \\ \vdots \\ m_{n,1}x_1 + m_{n,2}x_2 + \cdots + m_{n,n}x_n \end{bmatrix} \qquad D = \mathbb{R}^n, \quad B \subset \mathbb{R}^n$$

Die partielle Ableitung von $\boldsymbol{F}(\boldsymbol{x})$ nach x_k mit $1 \le k \le n$ ist

$$\partial_{x_k} \boldsymbol{F}(\boldsymbol{x}) = \partial_{x_k} \begin{bmatrix} m_{1,1}x_1 + \cdots + & m_{1,k}x_k & + \cdots + m_{1,n}x_n \\ m_{2,1}x_1 + \cdots + & m_{2,k}x_k & + \cdots + m_{2,n}x_n \\ & \vdots & \\ m_{n,1}x_1 + \cdots + & m_{n,k}x_k & + \cdots + m_{n,n}x_n \end{bmatrix} = \begin{bmatrix} m_{1,k} \\ m_{2,k} \\ \vdots \\ m_{n,k} \end{bmatrix}$$

und daher ist die Matrix dieser partiellen Ableitungen (die Jakobi-Matrix $\nabla F(x)$ von $\boldsymbol{F}(\boldsymbol{x})$) gegeben durch

$$\nabla F(x) = \{\partial_{x_k} \boldsymbol{F}(\boldsymbol{x})\}_{k=1}^n = \left\{ \begin{bmatrix} m_{1,k} \\ m_{2,k} \\ \vdots \\ m_{n,k} \end{bmatrix} \right\}_{k=1}^n = M$$

Beispiel: Gegeben seien $m, n \in \mathbb{N}$, $\boldsymbol{b} \in \mathbb{R}^m$ und $A \in \mathbb{R}^{m \times n}$ mit Komponenten

$$\boldsymbol{b} = \{b_i\}_{i=1}^m = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \quad \text{und} \quad A = \{a_{i,j}\}_{i=1,m}^{j=1,n} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,1} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

Für $\boldsymbol{x} \in \mathbb{R}^n$ mit Komponenten

$$\boldsymbol{x} = \{x_i\}_{i=1}^n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

sei $\boldsymbol{f}(\boldsymbol{x})$ definiert durch die Euklidischen Norm

$$f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$$

Wie im vorletzten Beispiel gezeigt, ist der Vektor der partiellen Ableitungen von $f(\boldsymbol{x})$ gegeben durch

$$\{\partial_{x_k} f(\boldsymbol{x})\}_{k=1}^n = 2A^\top A\boldsymbol{x} - 2A^\top \boldsymbol{b}$$

Der zweite Term ist ein Vektor, der von $\boldsymbol{x}$ nicht abhängt, und daher sind die partiellen Ableitungen dieses Terms Null. Der erste Term hat die Form $2M\boldsymbol{x}$ für eine Matrix $M = A^\top A$, und die entsprechenden partiellen Ableitungen sind durch das letzte Beispiel gegeben

$$\{\partial_{x_k} A^\top A\boldsymbol{x}\}_{k=1}^n = \{\partial_{x_k} M\boldsymbol{x}\}_{k=1}^n = M = A^\top A$$

Daher ist die Matrix der partiellen Ableitungen zweiter Ordnung (die Hesse-Matrix $\nabla^2 f(x)$ von $f(\boldsymbol{x})$) gegeben durch

$$\nabla^2 f(x) = \{\partial_{x_l} \partial_{x_k} f(\boldsymbol{x})\}_{k,l=1}^n = 2M = 2A^\top A$$