

Partielle Ableitungen von linearen und quadratischen Funktionen mehrerer Variablen

Dieses Blatt ist eine Ergänzung für die Begeisterten, die Details über Seite 186 im Skriptum und Hinweise für Beispiel 4 auf dem 12. Übungsblatt gerne hätten.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2$$

Wegen der Beziehung zwischen der Euklidischen Norm und dem Skalarprodukt,

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

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

$$f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 = (\mathbf{Ax} - \mathbf{b})^\top (\mathbf{Ax} - \mathbf{b})$$

Wegen der Eigenschaft der Transponierten eines Produkts,

$$(\mathbf{Ax})^\top = \mathbf{x}^\top \mathbf{A}^\top$$

folgt

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

und wenn ausmultipliziert,

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

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

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

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

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

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

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

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 $f(\mathbf{x})$ hat die Form $-2\mathbf{v}^\top \mathbf{x}$ für einen Vektor $\mathbf{v} = A^\top \mathbf{b}$, und die entsprechenden partiellen Ableitungen sind durch das vorletzte Beispiel gegeben

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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