§ 2 Notation und Wiederholung von Differenzierbarkeitsbegriffen

- Die Mengeninklusion $A \subset B$ bedeutet immer: $A \subseteq B$.
- Matrizen werden üblicherweise mit lateinischen Großbuchstaben bezeichnet, Vektoren mit lateinischen Kleinbuchstaben und Skalare mit griechischen oder lateinischen Kleinbuchstaben.
- Unendliche Folgen bezeichnen wir mit $\{x^{(k)}\}$ und nicht mit $\{x_k\}$ etc., um einen Konflikt mit der Bezeichnung der Komponenten eines Vektors $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ zu vermeiden. Endlich viele Vektoren werden dennoch auch mit x_1, x_2 etc. bezeichnet.
- Für Vektoren $x, y \in \mathbb{R}^n$ bezeichnet $x^\top y$ das Euklidische Skalarprodukt (Innenprodukt) (statt $\langle x,\, y\rangle$ oder $x \cdot y$) und $\|x\|$ die euklidische Norm:
$$\|x\| = \sqrt{x^\top x}$$
- Ist $M \in \mathbb{R}^{n\times n}$ eine symmetrisch positiv definite Matrix, so erzeugt sie ein Skalarprodukt $x^\top M y$ und eine Norm $\|x\|_M = \sqrt{x^\top M x}$ auf $\mathbb{R}^n$. Es gilt $\|x\| = \|x\|_I$.
- Für $\delta > 0$ und $x^* \in \mathbb{R}^n$ ist
$$U_\delta(x^*) = \{x \in \mathbb{R}^n : \|x - x^*\| < \delta\}$$
die offene δ-**Umgebung** von x^* oder auch δ-**Kugel** um x^*.
- Für eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ und gegebenes $x \in \mathbb{R}^n$ heißt die Ableitung der partiellen Funktion $t \mapsto f(x + t\, e_i)$ an der Stelle $t = 0$ die i-te **partielle Ableitung** von f an der Stelle x, kurz: $\frac{\partial}{\partial x_i} f(x)$. Dabei ist $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ einer der Vektoren der Standardbasis von $\mathbb{R}^n$. Mit anderen Worten:
$$\frac{\partial}{\partial x_i} f(x) = \lim_{t\to 0} \frac{f(x + t\, e_i) - f(x)}{t}.$$
- Allgemeiner heißt die Ableitung der Funktion $t \mapsto f(x + t\, d)$ an der Stelle $t = 0$ die **(beidseitige) Richtungsableitung** von f an der Stelle x in Richtung $d \neq 0$, kurz:
$$\frac{\partial}{\partial d} f(x) = \lim_{t\to 0} \frac{f(x + t\, d) - f(x)}{t}.$$
- Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heißt **differenzierbar** (**diffbar**) an der Stelle $x \in \mathbb{R}^n$, falls ein Vektor $v \in \mathbb{R}_n$ (Zeilenvektor) existiert, sodass gilt:
$$\frac{f(x + h) - f(x) - v\, h}{\|h\|} \to 0 \quad \text{für } h \to 0.$$
Der Vektor v heißt in dem Fall die **(totale) Ableitung** von f an der Stelle x und wird mit $f'(x)$ bezeichnet.
- Für diffbare Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ gilt
$$f'(x) = \left(\tfrac{\partial f(x)}{\partial x_1}, \quad \cdots\ , \quad \tfrac{\partial f(x)}{\partial x_n} \right).$$

Den transponierten Vektor (Spaltenvektor)

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix} = f'(x)^\top$$

bezeichnen wir als den **Gradienten** (bzgl. des Euklidischen Skalarprodukts) von f an der Stelle x.

- Für diffbare Funktionen gilt:
$$\delta f(x; d) = f'(x)\, d = \nabla f(x)^\top d.$$
- Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heißt **stetig partiell diffbar** oder kurz: C^1, wenn alle partiellen Ableitungen $\frac{\partial f(x)}{\partial x_i}$ als Funktionen von der Stelle x stetig sind. C^1-Funktionen sind überall diffbar.
- Die Matrix
$$\nabla^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial x_i\, \partial x_j}\right)_{i,j=1}^n = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}.$$
bestehend aus den zweiten partiellen Ableitungen der Funktion $f : \mathbb{R}^n \to \mathbb{R}$ an der Stelle x heißt die **Hessematrix**.
- Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heißt **zweimal stetig partiell differenzierbar** oder kurz: C^2, wenn alle Einträge in $\nabla^2 f(x)$ als Funktionen von der Stelle x stetig sind. In diesem Fall ist $\nabla^2 f(x)$ symmetrisch (Satz von Schwarz).

Schließlich benötigen wir häufig den Satz von Taylor:

Satz 2.1 (Taylor[6]).

Es sei $G \subset \mathbb{R}^n$ offen und $f : G \to \mathbb{R}$ $(k+1)$-mal stetig partiell diffbar, kurz: C^{k+1}. Falls x_0 und $x_0 + d$ und die gesamte Verbindungsstrecke in G liegen, dann existiert $\xi \in (0,1)$, sodass gilt:

$$k = 0: \quad f(x_0 + d) = f(x_0) + \nabla f(x_0 + \xi\, d)^\top d \quad \textbf{(Mittelwertsatz)}$$

$$k = 1: \quad f(x_0 + d) = f(x_0) + \nabla f(x_0)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_0 + \xi\, d)\, d$$

[6]siehe z. B. [Heuser, 2002, Satz 168.1]