Wiederholung von Linearer Algebra und Differentialrechnung im $\mathbb{R}^n$

1 Lineare Algebra

1.1 Matrizen

Notation:

- **Vektor** $x \in \mathbb{R}^n$**:** $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_i)_{i=1}^n$, mit den **Komponenten** $x_i,\ i \in \{1, \dots, n\}$

  zugehörige **Indexmenge:** $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$

  $$\Longrightarrow\ x_I = (x_i)_{i \in I} = \begin{pmatrix} x_{i_1} \\ \vdots \\ x_{i_k} \end{pmatrix} \begin{matrix} \leftarrow \\ \leftarrow \\ \leftarrow \end{matrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

  Matlab-Beispiel:

  ```
  >>  x = [9 8 7 6],  N = [2 3]
  x =
       9 8 7 6
  N =
       2 3
  >> x(N)
  ans =
       8 7
  ```

- **Matrix** $A \in \mathbb{R}^{n \times m}$**:** $A = (a_{ij})_{i=1,\ j=1}^{n,\quad m} = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix}$

  zugehörige **Indexmengen:** $\begin{matrix} I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\} \\ J = \{j_1, \dots, j_l\} \subseteq \{1, \dots, m\} \end{matrix}$

  $$\Longrightarrow\ A_{I,J} = (a_{ij})_{i \in I,\ j \in J} = \begin{array}{c} \begin{matrix} j_1 & j_2 & \dots & j_l \end{matrix} \\ \begin{bmatrix} * & * & \dots & * \\ * & * & \dots & * \\ \vdots & \vdots & & \vdots \\ * & * & \dots & * \end{bmatrix} \end{array} \begin{matrix} \\ i_1 \\ i_2 \\ \vdots \\ i_k \end{matrix}$$

  j**-te Spalte:** $A_{\cdot,j} = (a_{ij})_{i=1}^n$ (Matlabnotation: $A(:,j)$)

i-te Zeile: $A_{i,\cdot} = (a_{ij})_{j=1}^m$ (Matlabnotation: $A(i,:)$)

speziell $A \in \mathbb{R}^{n\times n},\ I \subseteq \{1,\dots,n\} \ \rightarrow\ A_I := A_{I,I}$

Matlab-Beispiel:

```
>>  A = [1 2 3 4; 5 6 7 8; 1 2 3 4], I = [2 3], J = [2 1]
A =
     1 2 3 4
     5 6 7 8
     1 2 3 4
I =
     2 3
J =
     2 1
>> A(I, J)
ans =
     6 5
     2 1
```

Definition 1. *Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch.*
*A heißt **positiv semidefinit** (schreibweise: $A \succeq 0$) $:\Leftrightarrow\ x^\top Ax \geq 0\ \forall x \in \mathbb{R}^n$.*
*A heißt **positiv definit** ($A \succ 0$) $:\Leftrightarrow\ x^\top Ax > 0\ \forall x \in \mathbb{R}^n \setminus \{0\}$.*
*A heißt **negativ semidefinit** ($A \preceq 0$) $:\Leftrightarrow\ x^\top Ax \leq 0\ \forall x \in \mathbb{R}^n$ bzw. falls $-A$ positiv semidefinit ist.*
*A heißt **negativ definit** ($A \prec 0$) $:\Leftrightarrow\ x^\top Ax < 0\ \forall x \in \mathbb{R}^n \setminus \{0\}$ bzw. falls $-A$ positiv definit ist.*

$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n |\ x_i \geq 0\ i = 1,\dots,n\}$.
S^n Menge der symmetrischen Matrizen.
S^n_+ Menge der positiv semidefiniten Matrizen.
S^n_{++} Menge der positiv definiten Matrizen.

Satz 2. *Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch. Dann sind äquivalent:*

(i) $A \succ 0$;

(ii) $\lambda_i > 0\ \forall i = 1,\dots,n$, *mit λ_i ist i-ter Eigenwert von A ($\lambda_i \in \mathbb{R}$, da A symmetrisch);*

(iii) $A = C^\top C$, *mit $C \in \mathbb{R}^{n\times n}$ regulär;*

(iv) $I_1 \subset I_2 \subset \dots \subset I_{n-1} \subset I_n = \{1,\dots,n\}$ *sei eine beliebige Folge geschachtelter Indexmengen, mit $|I_i| = i$.*

$$\det A_{I_i} > 0\ \forall i = 1,\dots,n.$$

(Die Determinanten einer beliebigen Folge geschachtelter Hauptuntermatrizen sind <u>alle</u> positiv.)

Bemerkung 3. *In (iii) kann C als obere Dreiecksmatrix gewählt werden (**Cholesky-Zerlegung**), d.h*

$$C = \begin{bmatrix} * & \dots & * \\ & \ddots & \vdots \\ 0 & & * \end{bmatrix}$$

(vgl. Numerik).

Satz 4. *Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch. Dann sind äquivalent:*

(i) $A \succeq 0$;

(ii) $\lambda_i \geq 0 \; \forall i = 1, \dots, n$;

(iii) $A = C^\top C$, *mit* $C \in \mathbb{R}^{n \times n}$ *und* $\mathrm{rank} C = \mathrm{rank} A$.

1.2 Normen

Definition 5. *Eine Funktion $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ heißt **(Vektor-) Norm**, wenn $\forall x, y \in \mathbb{R}^n$ gilt:*

1. $\|x\| \geq 0$

1a. $\|x\| = 0 \;\Leftrightarrow\; x = 0$

2. $\|cx\| = |c| \|x\| \; \forall c \in \mathbb{R}$

3. $\|x + y\| \leq \|x\| + \|y\|$

Beispiel 6.

- ***Euklidische Norm*** *oder l_2-Norm*

$$\|x\|_2 := (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$$

  (Wir verwenden fast ausschließlich diese Norm. Diese Norm ist bei uns meist gemeint, falls kein Index angegeben ist.)

- ***Summennorm*** *oder l_1-Norm*

$$\|x\|_1 := |x_1| + |x_2| + \ldots + |x_n|$$

- ***Maximumnorm*** *oder l_∞-Norm*

$$\|x\|_\infty := \max\{|x_1|, |x_2|, \ldots, |x_n|\}$$

- l_p-***Norm***

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Satz 7. *Seien $\|\cdot\|_\alpha$ und $\|\cdot\|_\beta$ zwei Normen im $\mathbb{R}^n$. Dann gibt es Konstanten $\overline{c} \geq \underline{c} > 0$, so dass*

$$\underline{c}\|x\|_\alpha \leq \|x\|_\beta \leq \overline{c}\|x\|_\alpha \ \forall x \in \mathbb{R}^n.$$

(D.h. im $\mathbb{R}^n$ sind alle Normen äquivalent).

Definition 8. *Eine Folge $\{x^{(k)}\} \subseteq \mathbb{R}^n$ von Vektoren konvergiert gegen $x \in \mathbb{R}^n$ bzgl. der Norm $\|x\|$ $:\Leftrightarrow$ $\|x^{(k)} - x\| \to 0$ für $k \to \infty$.*

Schreibweise: $x^{(k)} \stackrel{\|\cdot\|}{\to} x$ oder $\lim\limits_{k\to\infty} x^{(k)} = x$ bzgl. $\|\cdot\|$.

Korollar 9. *Im $\mathbb{R}^n$ sind äquivalent:*

(i) $x^{(k)} \stackrel{\|\cdot\|_\alpha}{\to} x,\ k \to \infty;$

(ii) $x^{(k)} \stackrel{\|\cdot\|_\beta}{\to} x,\ k \to \infty;$

(iii) *(**punktweise Konvergenz**)*

$$\begin{aligned} x_1^{(k)} &\to x_1,\ k \to \infty \\ &\vdots \\ x_n^{(k)} &\to x_n,\ k \to \infty. \end{aligned}$$

Definition 10. *Sei $M \subseteq \mathbb{R}^n$. Der **Abschluss von** M ($\mathrm{cl}M$) ist die Menge der Grenzwerte aller konvergenten Folgen aus M.*

$x, y \in \mathbb{R}^n$, **Skalarprodukt** $\langle x, y\rangle := x_1y_1 + x_2y_2 + \ldots + x_ny_n$.
Ein wichtiges Hilfsmittel bei Abschätzungen ist

Satz 11 (Cauchy-Schwarzsche Ungleichung)**.**

$$|\langle x, y\rangle| \leq \|x\|_2 \cdot \|y\|_2 \ \forall x, y \in \mathbb{R}^n$$

Normen lassen sich auf Matrizen im $\mathbb{R}^{n\times m}$ verallgemeinern:

Definition 12. *Eine Funktion $\|\cdot\| : \mathbb{R}^{n\times m} \to \mathbb{R}$ heißt **(Matrix-) Norm**, wenn $\forall A, B \in \mathbb{R}^{n\times m}$ gilt:*

1. $\|A\| \geq 0$

1a. $\|A\| = 0 \ \Leftrightarrow \ A = 0$

2. $\|cA\| = |c|\|A\| \ \forall c \in \mathbb{R}$

3. $\|A+B\| \leq \|A\| + \|B\|$

Beispiel 13.

- ***Frobeniusnorm***

$$\|A\|_2 = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)^{1/2}.$$

- ***induzierte Matrixnorm***

  Sei $\|\cdot\|_V$ eine Vektornorm auf $\mathbb{R}^n$. Wir definieren eine Funktion $\|\cdot\|_M$ auf $\mathbb{R}^{n\times n}$ durch

$$\|A\|_M := \max_{\|x\|_V=1} \|Ax\|_V.$$

  Satz 14. *Die so definierte Funktion $\|\cdot\|_M : \mathbb{R}^{n\times n} \to \mathbb{R}$ ist eine Norm, und es gilt*

$$\|Ax\|_V \leq \|A\|_M \cdot \|x\|_V \ \forall A \in \mathbb{R}^{n\times n},\ \forall x \in \mathbb{R}^n.$$

- *Die durch die Vektornorm $\|\cdot\|_2$ induzierte Matrixnorm ist die* ***Spektralnorm***

$$\|A\| := \max\{\sqrt{\lambda} \mid \lambda \text{ ist Eigenwert von } A^\top A\}.$$

Im Weiteren meint $\|x\|$ immer die l_2-Norm $\|x\|_2$.

1.3 Differentialrechnung im $\mathbb{R}^n$

$f:\ \mathbb{R}^n \to \mathbb{R}$

- **partielle Ableitung nach x_1 im Punkt** $(\xi_1,\ldots,\xi_n)$

$$\begin{aligned}\frac{\partial f}{\partial x_1}(\xi_1,\ldots,\xi_n) &:= \lim_{t\to 0}\frac{f(\xi_1+t,\xi_2,\ldots,\xi_n)-f(\xi_1,\xi_2,\ldots,\xi_n)}{t}\\ &= \frac{\mathrm{d}}{\mathrm{d}t} f(t,\xi_2,\ldots,\xi_n)|_{t=\xi_1}\end{aligned}$$

  Analog die partielle Ableitung nach x_i: $\frac{\partial f}{\partial x_i}(\xi_1,\ldots,\xi_n)$

- **Gradient** $\nabla f = \begin{pmatrix}\frac{\partial f}{\partial x_1}\\ \vdots \\ \frac{\partial f}{\partial x_n}\end{pmatrix}$

- **Richtungableitung nach** $p \in \mathbb{R}^n$

$$\frac{\partial f}{\partial p}(x) := \lim_{t\to 0}\frac{f(x+tp)-f(x)}{t}$$

  (Differenzenquotient in Richtung p)

$$f:\ \mathbb{R}^n \to \mathbb{R}^m,\ f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

- **Jacobimatrix** $\mathrm{J}_f(\xi) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\xi) & \dots & \frac{\partial f_1}{\partial x_n}(\xi) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\xi) & \dots & \frac{\partial f_m}{\partial x_n}(\xi) \end{pmatrix} = \begin{pmatrix} \nabla f_1(\xi)^\top \\ \vdots \\ \nabla f_m(\xi)^\top \end{pmatrix}$

Definition 15. *Sei $G \subseteq \mathbb{R}^m$ offen. Eine Funktion $f:\ G \to \mathbb{R}^m$ heißt* ***differenzierbar in*** *$x \in G$, wenn es eine $(m \times n)$-Matrix A gibt, so dass für hinreichend kleines h ($\forall x + h \in U_\epsilon(x)$)*

$$f(x+h) - f(x) = Ah + r(h),$$

mit

$$\lim_{h \to 0} \frac{r(h)}{\|h\|} = 0$$

(d.h. $r(h) = o(\|h\|)$) gilt.
Die Matrix A heißt ***Ableitung von f an der Stelle*** *x.*
f heißt ***differenzierbar*** *(auf G), wenn f in jedem $x \in G$ differenzierbar ist.*
f heißt ***stetig differenzierbar****, wenn f differenzierbar und*

$$f' : G \to \mathbb{R}^{m \times n},\ x \mapsto A(x)$$

stetig ist.

Satz 16. *$f : G \to \mathbb{R}^m$ sei differenzierbar in $\xi \in G$. Dann gilt*

- *jedes f_i ist in ξ partiell nach allen Variablen $x_1, \dots, x_n$ differenzierbar;*
- *A ist eindeutig bestimmt, $A = \mathrm{J}_f(\xi)$.*

Satz 17. *$f : G \to \mathbb{R}^m$ ist stetig differenzierbar auf $G \iff \forall i, j:\ f_j$ ist auf G partiell nach x_i differenzierbar und $\frac{\partial f_j}{\partial x_i}$ ist stetig auf G.*

$f : \mathbb{R}^n \to \mathbb{R}$, $\mathrm{J}_f(x) = \nabla f(x)^\top$. Obige Sätze:

- f ist differenzierbar in $x \ \Rightarrow\ \exists \nabla f(x)$ und $f(x+h) - f(x) = \nabla f^\top(x) h + r(h)$ (d.h. Gradient ist die Ableitung);
- f ist stetig differenziebar auf $\mathbb{R}^n \iff \exists \nabla f(x)$ auf $\mathbb{R}^n$ und ist stetig.

Zusammenhang von $\nabla f(x)$ und Richtungsableitung?

Satz 18. *Ist $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbar in x, so existiert für jedes $p \in \mathbb{R}^n$ die Richtungsableitung und ist*

$$\frac{\partial f}{\partial p}(x) = \nabla f(x)^\top p.$$

Sei $\|p\| = 1$. Cauchy-Schwarz

$$|\nabla f(x)^\top p| \leq \|\nabla f(x)\| \cdot \|p\| = \|\nabla f(x)\|.$$

Damit können spezielle Richtungen identifiziert werden:
Sei $p_0 := \frac{\nabla f(x)}{\|\nabla f(x)\|}$

$$\nabla f(x)^\top p_0 = \|\nabla f(x)\|,$$

d.h. der Gradient zeigt in Richtung des steilsten Anstiegs.
Sei $p_1 := \frac{-\nabla f(x)}{\|-\nabla f(x)\|}$

$$\nabla f(x) \top p_1 = -\|\nabla f(x)\|,$$

d.h. der Antigradient zeigt in Richtung des steilsten Abstiegs.

$$\nabla f(x) = 0 \;\Rightarrow\; \frac{\partial f}{\partial p}(x) = 0 \;\forall p \in \mathbb{R}^n.$$

Höhere Ableitungen: $f : \mathbb{R}^n \to \mathbb{R}$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} := \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j}$$

Hesse-Matrix

$$\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} = J_{\nabla f}$$

Definition 19. *Sei $G \subseteq \mathbb{R}^n$ offen. $C^m(G)$ ist die Menge der Funktionen $f : G \to \mathbb{R}$, deren partielle Ableitung der Ordnung $\leq m$ alle auf G existieren und stetig sind.*

Satz 20 (Satz von Schwarz)**.** *Für jedes $f \in C^m(G)$ sind die partiellen Ableitungen der Ordnung $\leq m$ unabhängig von der Reihenfolge der Differentiation.*

Speziell gilt also

$$f \in C^2(G) \;\Rightarrow\; \nabla^2 f \text{ ist symmetrisch.}$$

Folgende Sätze sind wichtige Hilfsmittel.

Satz 21 (Satz von Taylor)**.** *Seien $x,\ p \in \mathbb{R}^n$ und $f : \mathbb{R}^n \to \mathbb{R}$.*
Ist $f \in C^1(\mathbb{R}^n)$, dann gilt

$$f(x+p) = f(x) + \nabla f(x+tp)^\top p$$

für ein $t \in (0,1)$.
Ist $f \in C^2(\mathbb{R}^n)$, dann gilt

$$f(x+p) = f(x) + \nabla f(x)^\top p + \frac{1}{2} p^\top \nabla^2 f(x+tp) p$$

für ein $t \in (0,1)$ *und*

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p\mathrm{d}t.$$

Sei $x = (y_1, \dots, y_k, z_{k+1}, \dots, z_n)^\top = (y,z)^\top$.

$$[\mathrm{J}_f(\xi)]_y := \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(\xi) & \dots & \frac{\partial f_1}{\partial y_k}(\xi) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial y_1}(\xi) & \dots & \frac{\partial f_m}{\partial y_k}(\xi) \end{pmatrix},$$

d.h. die Spalten der Jacobimatrix bzgl. y.

Satz 22 (Satz über implizite Funktionen)**.** *Seien* $G \subseteq \mathbb{R}^n$ *und* $H \subseteq \mathbb{R}^m$ *nichtleer und offen und die Funktion* $F : G \times H \to \mathbb{R}^m$, $(x,y) \mapsto F(x,y)$ *sei stetig differenzierbar. Ferner seien* $\xi \in G$ *und* $\nu \in H$ *Punkte, für die*

$$F(\xi, \nu) = 0 \text{ und } [\mathrm{J}_F(\xi,\nu)]_y \text{ invertierbar}$$

ist.
Dann gibt es eine δ*-Umgebung* $U \subseteq G$ *von* ξ*, und eine* ϵ*-Umgebung* $V \subseteq H$ *von* ν *und genau eine stetige Funktion* $f : U \to V$ *mit* $f(\xi) = \nu$ *und* $F(x, f(x)) = 0\ \forall x \in U$.
Für jedes feste $x \in U$ *ist* $f(x)$ *sogar die einzige in* V *liegende Lösung der Gleichung* $F(x,y) = 0$.