

Lineare Algebra 2

Mitschrieb von Martin Salzer *

*Vorlesung im SS 09, Dozent: Prof. Dr. W. Kratz, Übungsleiterin: K. Setzer

7 Spektralsatz und Hauptachsentransformation

Bemerkung: Es sei $A \in \mathbb{C}$, $A = (a_1, \dots, a_n) = (a_{\mu\nu})$, $a_\nu \in \mathbb{C}^n$ (Spalte).

- (i) A heißt unitär (bzw. orthogonal), falls gilt: $A^*A = I$ (bzw. $A \in \mathbb{R}$) ($A^* = \overline{A}^T = A^{-1} \Leftrightarrow \{a_1, \dots, a_n\}$ bilden eine Orthonormalbasis des \mathbb{C}^n , d.h. $\langle a_\mu, a_\nu \rangle = \delta_{\mu\nu}$, $\|a_\nu\| = 1$, $a_\mu \perp a_\nu \forall \mu \neq \nu$, wobei wir im Folgenden stets das sogenannte *kanonische Skalarprodukt* verwenden, d.h. $\langle a, b \rangle = \overline{a}^T b$ und die euklidische Norm $\|a\| = \sqrt{\langle a, a \rangle}$)
- (ii) Eine Zahl $\lambda \in \mathbb{C}$ heißt ein *Eigenwert von A* , falls ein $x \in \mathbb{C}$ (der zugehörige Eigenwert) existiert mit $Ax = \lambda x$ und $x \neq 0 \Leftrightarrow 0 = P_A(\lambda) = \det(A - \lambda I)$ und $\sigma(A) := \{\lambda \in \mathbb{C} \mid \lambda \text{ ist ein Eigenwert von } A\}$ heißt das Spektrum von A sowie $\rho(A) := \max |\lambda| (\lambda \in \sigma(A))$ der Spektralradius von A
- (iii) A heißt diagonalisierbar, falls eine Matrix $S = (s_1, \dots, s_n) \in GL(n, \mathbb{C})$ existiert sowie $\lambda_1, \dots, \lambda_n$ mit $S^{-1}AS = \text{diag}(\lambda_1, \dots, \lambda_n) \Leftrightarrow s_1, \dots, s_n$ sind linear unabhängige Eigenvektoren zu den Eigenwerten $\lambda_1, \dots, \lambda_n$

Definition 7.0.1. Eine Matrix $A \in \mathbb{C}^{n \times n}$ heißt unitär diagonalisierbar, falls Zahlen $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ und eine unitäre Matrix $U \in \mathbb{C}^{n \times n}$ existieren so daß gilt: $U^{-1}AU = \text{diag}(\lambda_1, \dots, \lambda_n)$

7.1 Unitäre Trigonalisierung

Satz 7.1.1 (Lemma von Schur). Zu jeder Matrix $A \in \mathbb{C}^{n \times n} \exists \lambda_1, \dots, \lambda_n$ und eine unitäre

Matrix U aus $\mathbb{C}^{n \times n}$ so dass $U^{-1}AU = \begin{pmatrix} \lambda_1 & \dots & * \\ & \ddots & \vdots \\ 0 & & \lambda_n \end{pmatrix}$ eine obere Dreiecksmatrix ist.

Beweis. Nach Satz 1.1 \exists Eigenwert $\lambda_1 \in \mathbb{C}$ (d.h. $P_A(\lambda_1) = \det(A - \lambda_1 I) = 0$, $\exists u_1 \in \mathbb{C}^n \setminus \{0\}$ mit $Au_1 = \lambda_1 u_1$ und $\|u_1\| = 1$. Nach Cram-Schmidt (LA2 Kapitel 2 Satz 2.2 Kor. 2.1) kann u_1 zu einer ONB ergänzt werden, d.h. $\exists u_2, \dots, u_n \in \mathbb{C}^n$ so dass $U_1 = (u_1, \dots, u_n) \in \mathbb{C}^{n \times n}$ unitär ist

$$\Rightarrow U_1^{-1}AU_1 = \begin{pmatrix} \overline{u_1}^T \\ \vdots \\ \overline{u_n}^T \end{pmatrix} = (Au_1, \dots, Au_n) = \begin{pmatrix} \lambda_1 & \dots \\ 0 & \\ \vdots & \tilde{A} \\ 0 & \end{pmatrix} \text{ mit } \tilde{A} \in \mathbb{C}^{(n-1) \times (n-1)} \xrightarrow{\text{induktiv}}$$

$$\exists \text{ unitäre Matrix } \tilde{U} \in \mathbb{C}^{(n-1) \times (n-1)} \text{ mit } \tilde{U}^{-1}\tilde{A}\tilde{U} = \begin{pmatrix} \lambda_2 & \dots & * \\ & \ddots & \vdots \\ 0 & & \lambda_n \end{pmatrix}$$

Definiere $U := U_1 \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U} \end{pmatrix}$ Dann folgt

- (i) U ist unitär, weil $\tilde{U}^T U = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U}^T \end{pmatrix} \tilde{U}^T U_1 \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U}^T \tilde{U} \end{pmatrix} = I$
- (ii) $U^{-1}AU = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U}^T \end{pmatrix} \tilde{U}_1^T AU_1 \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U}^T \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots \\ 0 & \tilde{A} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \dots \\ 0 & \tilde{U}^T \tilde{A} \tilde{U} \end{pmatrix}$ ist eine obere Dreiecksmatrix

□

Zusatz: Aus Satz 1.1 und Beweis folgt

- (i) Ist $A \in \mathbb{R}^{n \times n}$ reell und sind alle Eigenwerte reell so kann man $U \in \mathbb{R}^{n \times n}$ reell (also orthogonal) gewählt werden.
- (ii) Es gilt $P_A(\lambda) = \det(A - \lambda I) = \det(\Delta - \lambda I) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda)$, d.h. $\lambda_1, \dots, \lambda_n \in \mathbb{C}$

7.2 Der Spektralsatz

Definition 7.2.1. Eine Matrix $A \in \mathbb{C}^{n \times n}$ heißt normal, falls gilt $A^*A = AA^*$

Satz 7.2.1 (Spektralsatz). Eine Matrix $A \in \mathbb{C}^{n \times n}$ ist genau dann normal, wenn sie unitär diagonalisierbar ist, d.h. es gibt Zahlen $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ und eine Matrix $U \in \mathbb{C}$ mit:

$$U^{-1}AU = \text{diag}(\lambda_1, \dots, \lambda_n) \text{ und } U^{-1} = U^*$$

Beweis. (i) A sei unitär diagonalisierbar $\Rightarrow AA^* = (UDU^{-1})(U\bar{D}^T\bar{U}^T) = U\bar{D}\bar{D}^T U = U\bar{D}^T D U = (U\bar{D}^T\bar{U}^T)(U\bar{U}^T) = A^*A$

(ii) Nun sei A normal. Nach Lemma von Schur (Satz 1.1) existieren $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ und

$$U \in \mathbb{C}^{n \times n} \text{ mit } U^*U = I \text{ und } U^{-1}AU = \Delta = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \text{ nun ist auch } \Delta\Delta^* =$$

$$(\bar{U}^T AU)(\bar{U}^T \bar{A}^T U) = \bar{U}^T A\bar{A}^T U = \bar{U}^T \bar{A}^T AU = (\bar{U}^T \bar{A}^T U)(\bar{U}^T AU) = \Delta^* \Delta \Rightarrow \Delta \text{ ist eine normale obere Dreiecksmatrix} \xrightarrow{\text{Lemma (s.u.)}} \Delta \text{ ist eine Diagonalmatrix} \Rightarrow \text{Beh.}$$

□

Lemma. Eine Dreiecksmatrix ist genau dann normal wenn sie eine Diagonalmatrix ist.

Beweis. (i) Eine Diagonalmatrix ist normal

(ii) Nun sei $\Delta = (\Delta_{\mu\nu})$ obere Dreiecksmatrix und normal

$$\Rightarrow (\Delta * \Delta)_{11} = \begin{pmatrix} \lambda_1 & 0 \\ & \vdots \\ * & \lambda_n \end{pmatrix} \begin{pmatrix} \lambda_1 & * \\ & \vdots \\ 0 & \lambda_n \end{pmatrix} = |\lambda_1|^2 = (\Delta\Delta^*)_{11} = |\lambda_1|^2 + (|\Delta_{12}|^2 + \dots + |\Delta_{1n}|^2) \Rightarrow \text{induktiv folgt nun die Behauptung.}$$

□

Bemerkung:

(i) A ist normal, U unitär $\Rightarrow \bar{U}^T A U$ ist unitär

(ii) Jede diagonale Matrix ist normal

(iii) Jede reell-symmetrische Matrix ($A \in \mathbb{R}^{n \times n}$ und $A = A^T = A^*$) ist normal

(iv) jede unitäre Matrix ist normal, weil $U^{-1}U = UU^{-1} = I$

(v) $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ ist nicht diagonalisierbar \Rightarrow nicht normal

7.3 Die Hauptachsentransformation

Definition 7.3.1. Eine quadratische Matrix A heißt reell-symmetrisch (oder hermitesch), falls gilt $A \in \mathbb{R}^{n \times n}$ und $A = A^T$ (oder $A \in \mathbb{C}^{n \times n}$ und $A = \bar{A}^T$)

Bemerkung: A reell-symmetrisch oder hermitesch $\Rightarrow A$ ist normal

Proposition 7.3.1. Eine reell-symmetrische (oder hermitesche) Matrix besitzt nur reelle Eigenwert und ihre Eigenvektoren zu verschiedenen Eigenwerten sind immer zueinander orthogonal (bzgl. des kanonischen Skalarprodukts)

Beweis. Es sei $A \in \mathbb{C}$ hermitesch $\lambda \in \sigma(A) \Rightarrow \exists x \in \mathbb{C}^n$ mit $Ax = \lambda x \Rightarrow \lambda \|x\|^2 = \lambda \langle x, x \rangle = \langle x, \lambda x \rangle = \langle x, Ax \rangle = \bar{x}^T Ax = \bar{x}^T \bar{A}^T x = (\bar{A}\bar{x})^T x = \langle Ax, x \rangle = \langle \lambda x, x \rangle = \bar{\lambda} \langle x, x \rangle = \bar{\lambda} \|x\|^2 \Rightarrow \lambda = \bar{\lambda} \in \mathbb{R}$

Nun sei $\mu \in \sigma(A)$ mit $\mu \neq \lambda$ und $y \in \mathbb{C}^n \setminus \{0\}$ mit $Ay = \mu y \Rightarrow \mu \in \mathbb{C}$ und $\langle x, y \rangle (\lambda - \mu) = \lambda \langle x, y \rangle - \mu \langle x, y \rangle = \langle \lambda x, y \rangle - \langle x, \mu y \rangle = \langle Ax, y \rangle - \langle x, Ay \rangle = (\bar{A}\bar{x})^T y - \bar{x}^T Ay = \bar{x}^T Ay - \bar{x}^T Ay = 0 \Rightarrow x \perp y$ □

Satz 7.3.1 (Hauptachsentransformation, 1. Teil). Es sei A eine reell-symmetrische (oder hermitesche) $n \times n$ -Matrix. Dann \exists reelle Zahlen $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ und eine orthogonale (oder unitäre) Matrix U , so dass gilt $U^{-1}AU = \text{diag}(\lambda_1, \dots, \lambda_n)$, ($\bar{U}^T = U^{-1}$)

Beweis. A reell-symmetrisch (oder hermitesch) $\Rightarrow A$ ist normal $\xrightarrow{\text{Spektralsatz, Prop 1.1}} \Rightarrow$ Beh. □

Beachte: $\lambda, \mu \in \mathbb{R}$ und $A \in \mathbb{R}^{n \times n} \Rightarrow U$ aus Satz 2.1 ist auch reell/orthogonal

Zusatz: Es gelte $U^*AU = \text{diag}(\lambda_\nu)$ mit $U^* = U^{-1}$ und $\lambda_\nu \in \mathbb{R} \Rightarrow A$ ist hermitesch (also reell-symmetrisch, wenn A reell ist)

Beweis. $A = UDU^{-1} \Rightarrow \bar{A}^T = \overline{(UDU^{-1})}^T = U\bar{D}^T\bar{U}^T = UDU^{-1} = A \Rightarrow A$ ist hermitesch \square

Fazit: Eine reelle Matrix ist genau dann orthogonal diagonalisierbar mit reellen Eigenwerten, wenn sie reell-symmetrisch ist.

Beispiel: $A = \begin{pmatrix} 7 & -5 & -2 \\ -5 & 7 & -2 \\ -2 & -2 & 4 \end{pmatrix}$ ist reell-symmetrisch mit Spektrum $\sigma(A) = \{0, 6, 12\}$

und Eigenvektoren $s_1 = (1, 1, 1)^T, s_2 = (1, 1, -2), s_3 = (-1, 1, 0), S^{-1}AS = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 12 \end{pmatrix}$

für $S = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -2 & 0 \end{pmatrix}$

Beachte, gemäß Prop. 4.1 $\sigma(A) \subset \mathbb{R}$ und $s_1 \perp s_2, s_2 \perp s_3, s_3 \perp s_1$

Nun ist $U = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} & 0 \end{pmatrix}$ orthogonal mit $U^{-1}AU = \text{diag}(0, 6, 12)$

\Rightarrow Hauptachsentransformation durchgeführt!

Beispiel: $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$ ist reell-symmetrisch

(i) Durchführung der Hauptachsentransformation: berechne λ_1, λ_2 und $U \in \mathbb{R}^{n \times n}$ mit

$$U^T U = I, U^T A U = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$P_A(\lambda) = \det(A - \lambda I) = \lambda^2 - 6\lambda + 8 = (\lambda - 2)(\lambda - 4) \Rightarrow \sigma(A) = \{2, 4\}$$

$$(A - 2I)u_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} u_1 = 0 \text{ für } u_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \Rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$(A - 4I)u_1 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} u_2 = 0 \text{ für } u_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\Rightarrow U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \text{ ist orthogonal mit } U^T A U = \text{diag}(2, 4)$$

(ii) Betrachtung der zugehörigen quadratischen Form: $q_A(x) = x^T A x = 3x_1^2 + 2x_1x_2 + 3x_2^2$

Ziel: Schreibe q_A als Summe von Quadraten:

a) quadratisch ergänzen: $q_A(x) = 3(x_1 + \frac{1}{3}x_2)^2 + \frac{8}{3}x_2^2 = y^T \begin{pmatrix} 3 & 0 \\ 0 & \frac{8}{3} \end{pmatrix} y = 3y_1^2 + \frac{8}{3}y_2^2$

für $x = Ty$ mit $T = \begin{pmatrix} 1 & \frac{1}{3} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -\frac{1}{3} \\ 0 & 1 \end{pmatrix}$

b) Hauptachsentransformation: $q_A(x) = x^T A x = x^T U (U^T A U) U^T x \Rightarrow y^T \text{diag}(2, 4) y = 2y_1^2 + 4y_2^2$ mit $x = U y$ bzw. $y = U^T x = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} x \Rightarrow q_A(x) = 2 \left(\frac{x_1 - x_2}{\sqrt{2}} \right)^2 + 4 \left(\frac{x_1 + x_2}{\sqrt{2}} \right)^2$

$E = \{x \in \mathbb{R}^2 \mid q_A(x) = 1\}$, d.h. $2y_1^2 + 4y_2^2 = 1 \Rightarrow$ Normalform einer Ellipse mit den Längen $\frac{1}{\sqrt{2}}$ und $\frac{1}{2}$

7.4 Quadratische Formen

Definition 7.4.1. Es sei $A \in \mathbb{R}^{n \times n}$ reell-symmetrisch (d.h. $A^T = A$)

- (i) $q_A(x) = x^T A x \in \mathbb{R}$ für $x \in \mathbb{R}^n$ heißt die der Matrix A zugeordnete quadratische Formel $= \sum_{\mu, \nu}^n a_{\mu\nu} x_\mu x_\nu$ für $A = (a_{\mu\nu})$, $x = (x_\mu)$
- (ii) $q_A(\cdot)$ oder A heißt positiv definit (resp. positiv semidefinit) bzw. negativ definit (resp. negativ semidefinit), falls gilt $q_A(x) > 0$ (resp. $q_A(x) \geq 0$) bzw. $q_A(x) < 0$ (resp. $q_A(x) \leq 0$) $\forall x \in \mathbb{R}^n \setminus \{0\}$ oder undefinit, falls $\exists x, y \in \mathbb{R}$ mit $q_A(x) < 0 < q_A(y)$

Bemerkung:

- (i) $A = (a_{\mu\nu})$ sei reell-symmetrisch, $x = x_{\mu\nu} \in \mathbb{R}^n \Rightarrow q_A(x) = \sum_{\nu=1}^n a_{\nu\nu} x_\nu^2 + 2 \sum_{1 \leq \mu < \nu \leq n} a_{\mu\nu} x_\mu x_\nu$
- (ii) Nun sei $A = (a_{\mu\nu}) \in \mathbb{R}^{n \times n}$ bel. $x = (x_\mu) \in \mathbb{R}^n$
 $x^T A x = \sum_{\mu, \nu}^n a_{\mu\nu} x_\mu x_\nu = q_B(x)$ für $B = \frac{1}{2}(A + A^T)$, wobei B reell-symmetrisch

Satz 7.4.1 (Hauptachsentransformation, 2. Teil). Es sei $A \in \mathbb{R}^{n \times n}$ reell-symmetrisch. Dann gibt es reelle Zahlen $\lambda_1, \dots, \lambda_n$ und eine orthogonale Matrix $U = (u_1, \dots, u_n) \in \mathbb{R}^{n \times n}$ so daß gilt $q_A(x) = \sum_{\nu=1}^n \lambda_\nu y_\nu^2$

Beweis. Es seien $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ und $U = (u_1, \dots, u_n) \in \mathbb{R}^{n \times n}$ orthogonal gemäß Satz 3.1 d.h. $U^T U = I$ und $U^T A U = \text{diag}(\lambda_\nu)$, dann gilt für $x = U y \in \mathbb{R}^n$: $q_A(x) = x^T A x = (U y)^T A (U y) = y^T (U^T A U) y = y^T \text{diag}(\lambda_\nu) y$ \square

Satz 7.4.2 (Definitheit). Es sei $A \in \mathbb{R}^{n \times n}$ reell-symmetrisch. Folgende Aussagen sind äquivalent

- (i) A ist positiv definit
- (ii) $-A$ ist negativ definit
- (iii) Alle Eigenwerte von A sind positiv
- (iv) alle HM von A sind positiv, d.h. $D_\nu > 0$ für $\nu = 1, \dots, n$, $D_\nu = \det(A_\nu)$, wobei

$$A_\nu := \begin{pmatrix} a_{11} & \dots & a_{1\nu} \\ \vdots & & \vdots \\ a_{\nu 1} & \dots & a_{\nu\nu} \end{pmatrix}$$

Eine Matrix A ist genau dann positiv semi-definit wenn die Eigenwerte ≥ 0 sind und sie ist indefinit wenn sie (mindestens) einen positiven und (mindestens) einen negativen Eigenwert besitzt.

Beweis. Es sei gemäß Satz 3.1 $U^T A U = D, U U^T = I$

(ii) (i) \Leftrightarrow (ii) : $q_A > 0 \Leftrightarrow 0 > -q_A = q_{-A}$

(iii) (i) \Rightarrow (iii) : $q_A(u_\nu) = u_\nu^T A u_\nu = \lambda_\nu u_\nu^T u_\nu = \lambda_\nu > 0$ für $\nu = 1, \dots, n$ falls A positiv definit ist.

(iii) \Rightarrow (i) : $q_A(x) = \sum_{\nu=1}^n \lambda_\nu y_\nu^2 > 0 \forall x = U y \in \mathbb{R}^n \setminus \{0\}$, falls $\lambda_\nu > 0 \forall \nu = 1, \dots, n$

(iv) (i) \Rightarrow (iv) : A sei positiv definit, $\nu \in \{1, \dots, n\}, x = \begin{pmatrix} z \\ 0 \end{pmatrix} \in \mathbb{R}^n$ mit $z \in \mathbb{R}^\nu \Rightarrow$

$q_A(x) = z^T A_\nu z > 0$, falls $z \neq 0 \Rightarrow A_\nu$ ist positiv definit $\forall \nu = 1, \dots, n$

$\Rightarrow D_\nu = \det A_\nu = \prod$ EW von $A_\nu > 0$ gemäß (iii)

(iv) \Rightarrow (i): Es sei also $D_\nu > 0 \forall \nu = 1, \dots, n$ Wir zeigen *induktiv*: A_ν ist positiv definit $\forall \nu = 1, \dots, n$

$\nu = 1$: $D_1 = a_{11} = A_1 > 0 \Rightarrow q_{A_1}(x) = a_{11}x^2 > 0 \forall x \in \mathbb{R}^1 \setminus \{0\}$

$\nu \rightarrow \nu + 1$: $A_{\nu+1} = \begin{pmatrix} A_\nu & a \\ a^T & \alpha \end{pmatrix}$ mit $a \in \mathbb{R}^\nu, \alpha = a_{\nu+1\nu+1} \in \mathbb{R}, A_\nu \in \mathbb{R}^{\nu \times \nu}$ ist positiv definit nach Ind-Hyp.

Für $x = \begin{pmatrix} z \\ x_{\nu+1} \end{pmatrix} \in \mathbb{R} \setminus \{0\}$ folgt (durch quadratisch ergänzen): $q_{A_{\nu+1}} = (z^T, x_{\nu+1}) \begin{pmatrix} A_\nu z + a x_{\nu+1} \\ a^T z + \alpha x_{\nu+1} \end{pmatrix}$
 $z^T A_\nu z + 2z^T a x_{\nu+1} + \alpha x_{\nu+1}^2 = (z + A_\nu^{-1} a x_{\nu+1})^T A_\nu (z + A_\nu^{-1} a x_{\nu+1}) + (\alpha - a^T A_\nu^{-1} a) x_{\nu+1}^2$

beachte: A_ν^{-1} existiert, wegen $\det A_\nu > 0$ und > 0 , weil A_ν positiv definit ist

und $\beta := \alpha - a^T A_\nu^{-1} a > 0$, denn $D_{\nu+1} := \det \left(\begin{pmatrix} A_\nu & a \\ a^T & \alpha \end{pmatrix} \begin{pmatrix} I & -A_\nu^{-1} a \\ 0 & 1 \end{pmatrix} \right) =$

$\det \left(\begin{pmatrix} A_\nu & 0 \\ a^T & \beta \end{pmatrix} \right) = D_\nu \cdot \beta$ und $D_\nu, D_{\nu+1} > 0$ wegen (iv) $\Rightarrow \beta > 0$

Der Zusatz folgt wie die Äquivalenzen (i) bis (iii)

□

Beispiele: [Platzhalter]

8 Singulärwertzerlegung und Moore-Penrose-Inverse

8.1 Polaren- und Singulärwertzerlegung

Das zentrale Hilfsmittel in diesem Kapitel liefert das

Lemma (Hauptlemma). *Es sei $A \in \mathbb{C}^{m \times n}$ mit $m \geq n$ und $r := \text{Rang}(A) \leq n$. Dann gibt es eine unitäre Matrix $W \in \mathbb{C}^{n \times n}$, eine Matrix $V \in \mathbb{C}^{m \times n}$ und eindeutige reelle Zahlen $\sigma_1, \dots, \sigma_n$ so dass gilt: $\sigma_1 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_n$, $V^*V = I_{n \times n}$ und $A = V\Lambda W^*$ mit $\Lambda = \text{diag}(\sigma_1, \dots, \sigma_n)$ (*)*

Beweis. Die Matrix $A^*A \in \mathbb{C}^{n \times n}$ ist hermitesch, $\text{Rang}(A^*A) = \text{Rang } A = n$ (weil $\text{Kern } A = \text{Kern } A^*A$ gemäß Kap. 2 Satz 5.5), und A^*A ist positiv semidefinit, weil $q_{A^*A}(x) = \bar{x}^T \bar{A}^T A x = \|Ax\|^2 \geq 0 \forall x \in \mathbb{C}^n \Rightarrow$ Satz über die Hauptachsentransformation 1. Teil (Kap. 8, Satz 3.1) \exists unitäre Matrix $W \in \mathbb{C}^{n \times n}$ und $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ mit $W^*A^*AW = \text{diag}(\lambda_1, \dots, \lambda_n)$ und $\lambda_\nu \geq 0 \forall \nu$, weil A^*A positiv semidefinit ist, und somit sei (o.B.d.A.) $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0 = \lambda_{r+1} = \dots = \lambda_n$, weil $\text{Rang } \text{diag}(\lambda_\nu) = \text{Rang } A = r$

Setze: $\sigma_\nu := \sqrt{\lambda_\nu}$ für $\nu = 1, \dots, n \Rightarrow \sigma_1 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_n$, $\Lambda := \text{diag}(\sigma_1, \dots, \sigma_n)$

Weiter setze: $V = (v_1, \dots, v_n) \in \mathbb{C}^{m \times n}$ mit $v_\nu := \frac{1}{\sigma_\nu} A w_\nu \in \mathbb{C}^m$ für $\nu = 1, \dots, r$ wobei $W = (w_1, \dots, w_n) \in \mathbb{C}^{n \times n}$. Mit dem Satz von Gram-Schmidt existieren $v_{r+1}, \dots, v_n \in \mathbb{C}^m$ so dass $\{v_1, \dots, v_n\}$ ein ONS bildet.

Für $V = (v_1, \dots, v_n)$ gilt nun $V \in \mathbb{C}^{m \times n}$ mit $V^*V = (\bar{v}_\mu^T v_\nu) = I_{n \times n}$ und $A = V\Lambda W^* \Leftrightarrow AW = V\Lambda \Leftrightarrow Aw_\nu = \sigma_\nu v_\nu$ für $\nu = 1, \dots, n$, weil

(i) $Aw_\nu = \sigma_\nu v_\nu$ für $\nu = 1, \dots, r$ nach Def. der v_ν und

(ii) $A^*Aw_\nu = \lambda_\nu w_\nu = 0$ für $\nu = r+1, \dots, n \Rightarrow Aw_\nu = 0 v_\nu$ für $\nu = r+1, \dots, n$

\Rightarrow Damit haben wir die Existenz von W, V und $\Lambda := \text{diag}(\sigma_\nu)$ mit (*) gezeigt, und zwar konstruktiv

Eindeutigkeit der σ_ν : Es gelte (*) $\Rightarrow A^*A = W\Lambda V^*V\Lambda W^* = W\Lambda^2 W^* = W \text{diag}(\sigma_1^2, \dots, \sigma_n^2) W^* \Rightarrow \sigma_1^2, \dots, \sigma_n^2 \geq 0$ sind die Eigenwerte von A^*A und damit eindeutig bestimmt durch $\sigma_1 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_n$ \square

Korollar 8.1.1. *Es gelte (*) aus dem Hauptlemma*

(i) $\sigma_1^2, \dots, \sigma_n^2$ sind die Eigenwerte von A^*A mit Eigenvektoren w_1, \dots, w_n , wobei $W = (w_1, \dots, w_n)$ unitär ist

(ii) $\sigma_1^2, \dots, \sigma_n^2$ sind die Eigenwerte von AA^* mit Eigenvektoren v_1, \dots, v_n , wobei $W = (w_1, \dots, w_n) \in \mathbb{C}^{m \times n}$ mit $V^*V = I$

(iii) Ist A reell, so können auch W, V reell gewählt werden

Beweis. (i) und (iii) folgen aus obiger Konstruktion

(ii) (*) $\Rightarrow AA^* = V\Lambda W^*W\Lambda V^* = V\Lambda^2 V^* \Leftrightarrow AA^*V = V\Lambda^2 \Leftrightarrow AA^*v_\nu = \sigma_\nu^2 v_\nu \Rightarrow$ (ii) gilt \square

Beispiel: $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 2 \end{pmatrix}$, $m = 3 \geq 2 = n$, $A^T A = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix}$ $P_{A^T A}(\lambda) = (6 - \lambda)(6 - \lambda) -$

$16 = \lambda^2 - 12\lambda + 20 = (\lambda - 10)(\lambda - 2) \Rightarrow \sigma_1 = \sqrt{10} \geq \sqrt{2} = \sigma > 0$

$$\begin{aligned}
(A^T A - 10I)w_1 &= \begin{pmatrix} -4 & 4 \\ 4 & -4 \end{pmatrix} w_1 = 0 \text{ für } w_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
(A^T A - 10I)w_1 &= \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix} w_1 = 0 \text{ für } w_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\
\Rightarrow W &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \\
v_1 &= \frac{1}{\sigma_1} A w_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 2 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \\
&\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
\Rightarrow V &= \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{2}{\sqrt{5}} & 0 \end{pmatrix} \text{ und } V \Lambda W^* \stackrel{!}{=} A
\end{aligned}$$

Im obigen Beweis wurde folgendes gezeigt/benutzt

Korollar 8.1.2. Für jede Matrix $A \in \mathbb{C}^{m \times n}$ ist $A^* A \in \mathbb{C}^{n \times n}$ positiv semidefinit d.h. $A^* A$ ist hermitesch und $\bar{x}^T \bar{A}^T A x = \|Ax\|^2 \geq 0 \forall x \in \mathbb{C}^n$ und damit sind die Eigenwerte von $A^* A > 0$ und dann gilt $\text{Rang } A = \text{Rang } A^* A = \text{Rang } A A^*$

Satz 8.1.1 (Singularwertzerlegung, SVD „singular value decomposition“). Es sei $A \in \mathbb{C}^{m \times n}$ und $r = \text{Rang } A \leq \min\{m, n\} =: q$. Dann gibt es unitäre Matrizen $W \in \mathbb{C}^{n \times n}$ und $V \in \mathbb{C}^{m \times m}$ sowie eindeutige reelle Zahlen $\sigma_1, \dots, \sigma_q$ so daß gilt $A = V \Sigma W^*$ wobei $\Sigma = (\sigma_{\mu\nu}) \in \mathbb{C}^{m \times n}$ mit $\sigma_{\nu\nu} = \sigma_\nu$ für $\nu = 1, \dots, q$ und $\sigma_{\mu\nu} = 0$ sonst und $\sigma_1 \geq \dots \geq \sigma_q > 0 = \sigma_{q+1} = \dots = \sigma_n$

Beweis. Es sei zunächst $n \leq m$ wie im Hauptlemma, dann $q = n \Rightarrow$ Hauptlemma $\exists \sigma_1 \geq \dots \geq \sigma_q > 0 = \sigma_{q+1} = \dots = \sigma_n, W \in \mathbb{C}^{n \times n}$ unitär und $\tilde{V} = (v_1, \dots, v_n) \in \mathbb{C}^{m \times n}$ mit $\tilde{V}^* \tilde{V} = I, A = \tilde{V} \Lambda W^*$ für $\Lambda = \text{diag}(\lambda_\nu)$

\Rightarrow Gram-Schmidt $\exists v_{n+1}, \dots, v_m$ so dass $\{v_1, \dots, v_m\}$ eine ONB ist $\Rightarrow V^* V = I_{m \times m}$

Setze $\Sigma = \begin{pmatrix} \Lambda \\ 0 \end{pmatrix} \in \mathbb{C}^{m \times n} \Rightarrow V \Sigma W^* = (\tilde{V}, v_{n+1}, \dots, v_m) \begin{pmatrix} 1 \\ 0 \end{pmatrix} W^* = \tilde{V} \Lambda W^* = A$

Nun sei $m \leq n \Rightarrow 1$. Teil $q = m$ und \exists unitäre Matrizen W und V und Zahlen σ mit $A^* = V \Sigma W^* \Rightarrow A = W \Sigma^T V^*$ \square

Definition 8.1.1. Die Zahlen $\sigma_1, \dots, \sigma_q$ heißen die Singularwerte der Matrizen A

Beispiel: $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 2 \end{pmatrix}, \tilde{V} \begin{pmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{2} \end{pmatrix} W$ mit $\tilde{V} = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{2}{\sqrt{5}} & 0 \end{pmatrix}, W = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \Rightarrow$

$V = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 & -\frac{2}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{pmatrix}$ ist orthogonal und $A = V \Sigma W^T$ mit $\Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{pmatrix}$ ist die SVD

Korollar 8.1.3. Es sei $A \in \mathbb{C}^{m \times n}$ mit Singulärwerten $\sigma_1, \dots, \sigma_q \geq 0$. Dann gilt $\sigma_1 = \max_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|} =: \|A\|_s$, wobei $\|x\| = \sqrt{x^T x}$. Man nennt $\|A\|_s$ die Spektralnorm von A

Beweis. Es seien V, W unitär $A = V\Sigma W^*$ gemäß Satz 1.1 sowie $x \in \mathbb{C}^n, y = (y_\mu) = W^*x \Rightarrow \|x\| = \|y\|$, weil W^* unitär ist und $\|Ax\| = \|V\Sigma W^*x\| = \|\Sigma y\| = \sqrt{\sum_{\mu=1}^n \sigma_\mu^2 |y_\mu|^2} \leq \sigma_1 \sqrt{\sum_{\mu=1}^n |y_\mu|^2} = \sigma_1 \|y\| = \sigma_1 \|x\|$ für $y = e_1$, d.h. $x = We_1$ \square

Bemerkung: (i) Zwischen *Spektralsatz* und der SVD besteht der folgende *Zusammenhang* für normale Matrizen:

Es sei $A \in \mathbb{C}^{n \times n}$ normal \Rightarrow (Spektralsatz) $\exists \lambda_1, \dots, \lambda_n \in \mathbb{C}$ und $U \in \mathbb{C}^{n \times n}$ mit $U^*AU = \text{diag}(\lambda_\nu)$ d.h. $A = U \text{diag}(\lambda_\nu)U^*$ und o.B.d.A. $|\lambda_1| \geq \dots \geq |\lambda_n|$ Nun gilt (nach Satz vom Argument) $\exists \phi_\nu \in (-\pi; \pi]$ mit $\lambda_\nu = |\lambda_\nu|e^{i\phi_\nu}$. Setze $\sigma_\nu = |\lambda_\nu|$ und $W = U$, sowie $V = U \text{diag}(e^{i\phi_\nu}) = (e^{i\phi_1}u_1, \dots, e^{i\phi_n}u_n) \Rightarrow \sigma_1 \geq \dots \geq \sigma_n \geq 0$, W ist unitär, V ist unitär $(e^{i\phi_\mu}u_\mu)^*(e^{i\phi_\nu}u_\nu) = u_\mu^*u_\nu e^{i(\phi_\nu - \phi_\mu)} = \delta_{\mu\nu}$ und es gilt $A = V\Sigma W^*$ ist die SVD von A

(ii) $\sigma_1^2, \dots, \sigma_n^2$ sind die Eigenwerte von A^*A und von AA^* bis auf die Vielfachheit $m - q$, bzw. $n - q$ des Eigenwerts 0

(iii) A und A^* besitzt dieselben Singulärwerte

(iv) Die Singulärwerte sind die Eigenwerte der sogenannten Polarmatrix

Satz 8.1.2 (Polarzerlegung). Es sei $A \in \mathbb{C}^{m \times n}$ mit $m \geq n$ Dann gibt es eine unitäre Matrix $U \in \mathbb{C}^{m \times m}$ und eine positive semidefinite Matrix $P \in \mathbb{C}^{m \times n}$ so dass gilt $A = UP$ und $U^*U = I$. Die Matrix P ist eindeutig nämlich $P = \sqrt{A^*A}$ Man nennt sie die Polarmatrix.

Beweis. Nach dem *Hauptlemma* existieren $\sigma_1 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_n$, $W \in \mathbb{C}^{n \times n}$ unitär, $V \in \mathbb{C}^{m \times n}$ mit $A = V\Lambda W^*$ mit $V^*V = I_{n \times n}, \Lambda = \text{diag}(\sigma_1, \dots, \sigma_n)$

Setze $P := W\Sigma W^*$ und $U = VW^*$. Damit ist P positiv semidefinit, weil $\sigma_\nu \geq 0 \forall \nu \in \{1, \dots, n\}$ ($\{\sigma_\nu\} = \sigma(P)$) und $U^*U = WV^*VW^* = WW^* = I$ sowie $UP = VW^*W\Lambda W^* = V\Lambda W^* = A \Rightarrow$ Existenz

Eindeutigkeit: $A = UP$ mit $U^*U = I, P$ positiv semidefinit, d.h. $P = P^*, \bar{x}^T P x \geq 0 \forall x \in \mathbb{C}^{n \times n}$. Dann folgt $P^2 = P^*P = P^*(U^*U)P = A^*A \Rightarrow P\sqrt{A^*A}$ mit folgendem Satz folgt damit die Eindeutigkeit \square

Satz 8.1.3 (Quadratwurzel). Jede positiv semi-definite Matrix besitzt eine eindeutig positiv-semidefinit Quadratwurzel, d.h. $\forall A \in \mathbb{C}^{n \times n}$ mit $A = A^*$ und $\bar{x}^T A x \geq 0 \forall x \in \mathbb{C}^n$ existiert genau eine Matrix $B := \sqrt{A} \in \mathbb{C}^{n \times n}$ mit $B^2 = A$ und $B = B^*$

Beweis. A ist positiv semidefinit $\Rightarrow \exists \lambda_1 \geq \dots \geq \lambda_n \geq 0$ und $U \in \mathbb{C}^{n \times n}$ unitär mit $U^*AU = \text{diag}(\lambda_\nu)$, d.h. $A = U\Lambda U^*$.

Definiere $B := U \text{diag}(\sqrt{\lambda_\nu})U^*$, dann folgt B ist hermitesch, positiv semidefinit (weil $\sqrt{\lambda_\nu} \geq 0$) mit $B^2 = A \Rightarrow$ Existenz

Eindeutigkeit: Es sei $\tilde{B} \in \mathbb{C}^{n \times n}$ positiv semidefinit mit $\tilde{B}^2 = A$. Wir zeigen $\tilde{B} = B \Rightarrow \exists \mu_1 \geq \dots \geq \mu_n \geq 0$ und $V = (v_1, \dots, v_n)$ unitär mit $V^*BV = \text{diag}(\mu_1, \dots, \mu_n)$, d.h. $\tilde{B}v_\nu = \mu_\nu v_\nu \forall \nu \in 1, \dots, n \Rightarrow Av_\nu = \tilde{B}\tilde{B}v_\nu = \mu_\nu^2 v_\nu$, d.h. $\mu_1^2 \geq \dots \geq \mu_n^2$ sind die Eigenwerte von A mit $\text{Eig}(\tilde{B}, \mu_\nu) = \text{Eig}(A, \mu_\nu^2)$ und $\lambda_\nu = \mu_\nu^2 \forall \nu = 1, \dots, n$ und wegen $Au_\nu = \lambda_\nu u_\nu$ mit $U = (u_1, \dots, u_n) \Rightarrow u_\nu \in \text{Eig}(\tilde{B}, \mu_\nu)$, d.h. $\tilde{B}u_\nu = \mu_\nu u_\nu \forall \nu = 1, \dots, n$ d.h. $U^*\tilde{B}U = \text{diag}(\mu_\nu) \Leftrightarrow \tilde{B} = U \text{diag}(\mu_\nu)U^* = B$ \square

Bemerkung: Falls die gegebenen Matrizen reell sind so sind auch die Matrizen $V, W, A, \Sigma, U, P, \sqrt{A}$ aus Satz 1.1-1.3 reell (bzw. können reell gewählt werden).

Beispiel: Berechne SVD, Polarenzerlegung, $\sqrt{\cdot}$ (soweit definiert) ff. Matrizen (siehe vorne)

(i) $A = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix}$. Dann gilt $U^T A U = \text{diag}(10, 2)$ mit $U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$

SVD: $A = V \Sigma W^T$ für $V = W = U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ und $\Sigma = \text{diag}(10, 2)$

Polarenzerlegung: $A = UP$, $U = I, P = A$

Quadratwurzel: $A = U \text{diag}(\sqrt{10}, \sqrt{2})U^T = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{10} & \sqrt{10} \\ -\sqrt{2} & \sqrt{2} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \sqrt{10} + \sqrt{2} & \sqrt{10} - \sqrt{2} \\ \sqrt{10} - \sqrt{2} & \sqrt{10} + \sqrt{2} \end{pmatrix}$

(ii) $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 2 \end{pmatrix} = \tilde{V} \Lambda W^T$ für $\Lambda = \text{diag}(\sqrt{10}, \sqrt{2}), W = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ mit $\tilde{V} = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{2}{\sqrt{5}} & 0 \end{pmatrix}$

SVD: $A \Sigma W^*$ für $V = \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 & -\frac{2}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{pmatrix}, W = \text{siehe oben}, \Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{pmatrix}$

Polarenzerlegung ($n = 2 < 3 = m$) $A = UP$ für $U = \tilde{V}W^T, P = W^T A W =$

(Rechnung siehe oben) $= \begin{pmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{2}{\sqrt{5}} & 0 \end{pmatrix}$

(iii) $A = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$ ist reell-symmetrisch ist bereits „auf Hauptachse“, A ist indefinit

SVD: $A = V \Sigma W$ für $\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, V = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$

Polarenzerlegung: $A = UP$ mit $P = \sqrt{A^T A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, U = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

Quadratwurzel: nicht definiert, weil A nicht positiv semidefinit ist.

Bemerkung: (i) Polarenzerlegung. Diese entspricht für komplexe Zahlen $z \in \mathbb{C}^{1 \times 1}$ der Polarkoordinatenzerlegung, nämlich $z = e^{i\varphi} r = up$ für $p = r = |z| \geq 0$ und $u = e^{i\varphi}, \bar{u}u = e^{-i\varphi} e^{i\varphi} = 1, r = \sqrt{\bar{z}z}$

(ii) Quadratwurzel. Für eine gegebene Matrix $A \in \mathbb{R}^{n \times n}$ besitzt die Gleichung $X^2 = A$ i.a. viele Lösungen für $n \neq 2$ (im Falle $n = 1$ gibt es 2 Lösung), z.B. $X^2 = I_{2 \times 2}$ besitzt Lösungen $X = I, -I, \pm \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ -\sin(\alpha) & -\cos(\alpha) \end{pmatrix}$

8.2 Die Moore-Penrose verallgemeinerte Inverse

Satz 8.2.1 (Moore-Penrose Inverse). Für alle Matrizen $A \in \mathbb{C}^{m \times n}$ gibt es genau eine Matrix $A^+ \in \mathbb{C}^{n \times m}$ mit den folgenden Eigenschaften: (*) $A = AA^+A, A^+ = A^+AA^+,$ und AA^+ und A^+A sind hermitesch, Ist A reell, so ist auch A^+ reell.

Definition 8.2.1. Die Matrix $A^+ \in \mathbb{C}^{n \times m}$ heißt die Moore-Penrose Inverse von $A \in \mathbb{C}^{m \times n}$

Beweis. Es sei $A = V\Sigma W^*$ SVD von A gemäß Satz 1.1 (V, W unitär, $\Sigma = \begin{pmatrix} \sigma_1 & \dots & 0 \\ 0 & \ddots & \vdots \\ \vdots & & \sigma_r \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix} \in \mathbb{C}^{m \times n}$).

Setze $A^+ := W\Sigma^+V^* \in \mathbb{C}^{n \times m}$, wobei $\Sigma^+ := \begin{pmatrix} \frac{1}{\sigma_1} & \dots & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \dots & \frac{1}{\sigma_r} & 0 \end{pmatrix}$

$\Rightarrow AA^+ = V\Sigma W^*W\Sigma^+V^* = V \begin{pmatrix} I_{r \times r} & 0 \\ 0 & 0 \end{pmatrix}_{m \times m} V^*$ und

$\Rightarrow A^+A = W\Sigma^+V^*V\Sigma W^* = W \begin{pmatrix} I_{r \times r} & 0 \\ 0 & 0 \end{pmatrix}_{n \times n} W^*$ sind hermitesch mit

$AA^+A = V \begin{pmatrix} I_{r \times r} & 0 \\ 0 & 0 \end{pmatrix}_{m \times m} V^*V\Sigma W^* = V\Sigma W^* = A,$

$A^+AA^+ = W \begin{pmatrix} I_{r \times r} & 0 \\ 0 & 0 \end{pmatrix}_{n \times n} W^*W\Sigma^+V^* = W\Sigma^+V^* = A^+ \Rightarrow$ Existenz

Eindeutigkeit: Es sei $\tilde{A} \in \mathbb{C}^{n \times m}$ mit (*), Zu zeigen: $\tilde{A} = A^+$. Wir verwenden im Beweis, dass $A = B \Leftrightarrow (A - B)^*(A - B) = 0$ gemäß der Bemerkung unten.

(a) $(A^+A - \tilde{A}\tilde{A})^*(A^+A - \tilde{A}\tilde{A}) = (A^+A)^*(A^+A) + (\tilde{A}\tilde{A})^*(\tilde{A}\tilde{A}) - (A^+A)^*(\tilde{A}\tilde{A}) - (\tilde{A}\tilde{A})^*(A^+A) \stackrel{(*)}{=} A^+AA^+A + \tilde{A}\tilde{A}\tilde{A}\tilde{A} - A^+A\tilde{A}\tilde{A} - \tilde{A}\tilde{A}A^+A = A^+A + \tilde{A}\tilde{A} - A^+A - \tilde{A}\tilde{A} = 0 \Rightarrow A^+A = \tilde{A}\tilde{A}$

$$(b) (AA^+ - A\tilde{A})^*(AA^+ - A\tilde{A}) = AA^+AA^+ + A\tilde{A}A\tilde{A} - AA^+A\tilde{A} - A\tilde{A}AA^+ = AA^+ + A\tilde{A} - A\tilde{A} - AA^+ = 0$$

$$\Rightarrow \tilde{A} = \tilde{A}A\tilde{A} \stackrel{(a)}{=} A^+A\tilde{A} \stackrel{(b)}{=} A^+AA^+ = A^+ \quad \square$$

Bemerkung: Für jede Matrix $A \in \mathbb{C}^{m \times n}$ gilt $\text{Rang } A = \text{Rang } A^*A \Rightarrow A = 0$ genau dann wenn $A^*A = 0$, also $A = B \Leftrightarrow (A - B)^*(A - B) = 0$

Beispiele: Berechne A^+ für folgende Matrizen A .

$$(i) A = \begin{pmatrix} 6 & 4 \\ 4 & 6 \end{pmatrix} \text{ ist regulär} \Rightarrow A^+0A^{-1} = \frac{1}{20} \begin{pmatrix} 6 & -4 \\ -4 & 6 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$$

$$(ii) A = \begin{pmatrix} 1 & -1 & 2 \\ 1 & 1 & 2 \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 2 \end{pmatrix} = V\Sigma W^T \text{ mit [Platzhalter]}$$

Proposition 8.2.1 (Eigenschaften der Moore-Penrose-Inversen). Für $A \in \mathbb{C}^{m \times n}$ gilt:

$$(i) (A^+)^+ = A$$

$$(ii) (A^*)^+ = (A^+)^*$$

$$(iii) \text{Rang } A = \text{Rang } A^+$$

$$(iv) A^+ = A^{-1}, \text{ falls } A \text{ regulär ist } (m = n = r)$$

$$(v) A^+ = (A^*A)^{-1}A^*, \text{ falls } \text{Rang } A = n \text{ gilt}$$

$$(vi) (UAV)^+ = V^*A^+U^*, \text{ falls } U, V \text{ unitär sind.}$$

Übungsaufgabe: $(A^*A)^+ = A^+(A^*)^+$, $A^+ = A^*(AA^*)^{-1}$, falls $\text{Rang } A = m$ gilt, $\text{Im } A^+ = \text{Im } A^*$, $\text{Kern } A = \text{Kern}(A^*)^+$

Beweis. (i) gilt, weil $(*)$ symmetrisch in A und A^+ ist

$$(ii) (*) \Rightarrow A^* = A^*(A^+)^*A^*, (A^+)^*A^*(A^+)^*, \text{ und } A^*(A^+)^*, (A^+)^*A^* \text{ sind hermitesch} \\ \Rightarrow (A^+)^* = (A^*)^+$$

$$(iii) \text{Rang } A = r = \text{Rang } \Sigma = \text{Rang } \Sigma^+ = \text{Rang } A^+$$

$$(iv) A \in \text{GL}(n, \mathbb{C}), A = A^{-1}AA^{-1} = A^{-1}$$

$$(v) \text{Rang } A = n = \text{Rang } A^*A \Rightarrow A^*A \in \mathbb{C}^{n \times n} \text{ ist regulär} \Rightarrow \tilde{A} := (A^*A^{-1})A^* \text{ ist} \\ \text{wohldefiniert und } A\tilde{A} = A(A^*A)^{-1}A^* \text{ ist hermitesch, weil es } A^*A \text{ ist, } \tilde{A}A = I \text{ ist} \\ \text{hermitesch} \Rightarrow A\tilde{A}A = AI = A, \tilde{A}A\tilde{A} = I\tilde{A} = \tilde{A} \Rightarrow (*) \text{ gilt} \Rightarrow \tilde{A} = A^+$$

$$(vi) \text{Es seien } U, V \text{ unitär, } B := UAV, \tilde{B} := V^*A^+U^* \Rightarrow B\tilde{B} = UAVV^*A^+U^* = \\ UAA^+U^* \text{ und } \tilde{B}B = V^*A^+AV \text{ sind hermitesch, weil } A^+A \text{ und } AA^+ \text{ es sind} \\ \text{und } B\tilde{B}B = UAA^+(U^*U)AV = UAV = B, \tilde{B}B\tilde{B} = V^*A^+A(VV^*)A^+U^* = \\ V^*A^+U^* = \tilde{B} \Rightarrow \text{Satz 2.1 } B = B^+ \quad \square$$

8.3 Die Methode der kleinsten Fehlerquadrate (Gauß)

Motivation: Ein lineares Gleichungssystem $Ax = b$ sei unlösbar. Berechne x mit $\|Ax - b\| = \min$, d.h. $\sum_{\mu=1}^m |\sum_{\nu=1}^n a_{\mu\nu}x_{\nu} - b_{\mu}|^2 \rightarrow \min$ (und $\|x\| = \min$)

Satz 8.3.1 (least square property). *Es sei $A \in K^{m \times n}$ und $b \in \mathbb{C}^m$. Dann ist $\hat{x} := A^+b \in \mathbb{C}^n$ die eindeutige Lösung des folgenden Optimierungsproblems $\|A\hat{x} - b\| = \min\{\|Ax - b\| : x \in \mathbb{C}^n\}$ und $\|\hat{x}\| = \min\{\|x\| : x \in \mathbb{C}^n \text{ mit } \|Ax - b\| = \|A\hat{x} - b\|\}$*

Beweis. Es sei also $\hat{x} = A^+b$ und $x \in \mathbb{C}^n$. Dann gilt $\langle A\hat{x} - b, A(x - \hat{x}) \rangle = b^*(AA^+ - I)^{*}A(x - \hat{x}) = b^*(AA^+A - A)(x - \hat{x}) = 0$, d.h. $A\hat{x} - b \perp A(x - \hat{x})$ (1).

Nun sei $x_0 \in \text{Kern } A \Rightarrow \langle \hat{x}, x_0 \rangle = b^*(A^+)^{*}x_0 = b^*(A^+AA^+)^{*}x_0 = b^*(A^+)^{*}(A^+A)^{*}x_0 = b^*(A^+)^{*}A^+Ax_0 = 0 \Rightarrow \hat{x} \perp \text{Kern } A$ (2)

Für $x \in \mathbb{C}^n$ folgt $\|Ax - b\|^2 = \|A(x - \hat{x}) + (A\hat{x} - b)\|^2 \stackrel{(1)}{=} \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \geq \|A\hat{x} - b\|^2$ mit = genau dann wenn $A(x - \hat{x}) = 0$, d.h. $x_0 = x - \hat{x} \in \text{Kern } A$ und dann $\|x\|^2 = \|x + \hat{x}_0\|^2 \stackrel{(2)Pyth.}{=} \|x_0\|^2 + \|\hat{x}\|^2 \geq \|\hat{x}\|^2$ und = genau dann wenn $x_0 = 0 \Leftrightarrow x - \hat{x} = 0 \Leftrightarrow x = \hat{x}$ \square

Bemerkung (Anwendungen): Moore-Penrose Inverse haben vielfältige Anwendungen, insbesondere in der Stochastik, u.a. Schätzungen von Parametern, Anpassen linearer Hypothesen via least square estimates

Kurvenanpassung mit der Gaußschen Methode der kleinsten Fehlerquadrate:

Gegeben seien m Daten $(x_{\mu}, y_{\mu})^T \in \mathbb{R}^2$ für $\mu = 1, \dots, m$, sowie n Funktionen $\varphi_{\nu}(x) \forall \nu = 1, \dots, n$ (i.a. $n \leq m$) (etwa $\varphi_{\nu}(x) = x^{\nu-1}$) für $\beta = (\beta_{\nu})$ sei $f(x) = \sum_{\nu=1}^n \varphi_{\nu}(x)$ (etwa $\sum_{\nu=1}^n \beta_{\nu}x^{\nu-1}$)

Bestimme $\hat{\beta} \in \mathbb{R}^n$, so daß für $\beta = \hat{\beta}$ gilt $\sum_{\mu=1}^m |f(x_{\mu}) - y_{\mu}|^2 \rightarrow \min$ und dann $\|beta\| = \min$. Nun gilt $f(x_{\mu}) = \sum_{\nu=1}^n \varphi_{\nu}(x_{\mu})\beta_{\nu} = \sum_{\nu=1}^n a_{\mu\nu}b_{\nu} \Rightarrow \beta^+ = A^+y$ für $A = (a_{\mu\nu}) = (\varphi_{\nu}(x_{\mu})) \in \mathbb{R}^{m \times n}$, $y = (y_{\nu})$ mit angepasster Kurve $\hat{f}(x) = \sum_{\nu=1}^m \hat{\beta}_{\nu}\varphi_{\nu}(x)$

Beispiel (Folie): Anpassung an eine gerade („Ausgleichsgerade“) $\varphi_{\nu}(x) = x^{\nu-1}$ für

$\nu = 1, \dots, 2 = n$ für gegebene Daten $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3, y = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 2 \\ 4 \end{pmatrix}$, $A =$

$$(\varphi_{\nu}(x_{\nu})) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, A^+ = \frac{1}{10} \begin{pmatrix} 7 & 4 & 1 & -2 \\ -3 & -1 & 1 & 3 \end{pmatrix}$$

$$\hat{\beta} = A^+y = \dots = \frac{1}{20} \begin{pmatrix} -1 \\ 24 \end{pmatrix} \Rightarrow \hat{f}(x) = 0,05 + 1,2x$$

8.4 Normen

Definition 8.4.1. *Es sei V ein Vektorraum über $K = \mathbb{R}$ oder $K = \mathbb{C}$. Eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ heißt eine Norm in V (auch Vektornorm), falls die folgenden Eigenschaften (Axiome) gelten $\forall a, b \in V$ und $\alpha \in K$*

(N1) $\|a + b\| \geq \|a\| + \|b\|$ (Dreiecksungleichung)

(N2) $\|\alpha a\| = |\alpha| \|a\|$ (Homogenität)

(N3) $\|a\| \geq 0$ und $= 0$ genau dann wenn $a = 0$ (Positivität)

Man nennt dann V einen normierten Raum

Proposition 8.4.1. In jedem normierten Raum V gilt $\|a - b\| \geq \left| \|a\| - \|b\| \right| \forall a, b \in V$ (N1)

Beweis.

$$\begin{aligned} \|a\| = \|a - b + b\| &\leq \|a - b\| + \|b\| &\Rightarrow \|a\| - \|b\| &\leq \|a - b\| \\ \text{und } \|b\| = \|b - a + a\| &\leq \|b - a\| + \|a\| &\Rightarrow \|b\| - \|a\| &\leq \|b - a\| = \|a - b\| \end{aligned}$$

□

Satz 8.4.1 (Äquivalenz von Normen). Es sei V ein VR über $K = \mathbb{R}$ oder $K = \mathbb{C}$ mit endlicher Dimension $n \leq \dim V < \infty$, $\|\cdot\|_*$, $\|\cdot\|_\sim$ seien Normen in V . Dann gibt es Konstanten $c_1, c_2 > 0$ so dass gilt $c_1 \|a\|_* \leq \|a\|_\sim \leq c_2 \|a\|_*$ für alle $a \in V$

Beweis. Es sei $\{a_1, \dots, a_n\}$ eine Basis von V (LA1). Für $\alpha = (\alpha_\nu) \in K^n$ definieren wir Funktionen $f_*(\alpha) := \|\sum_{\nu=1}^n \alpha_\nu a_\nu\|_*$ und $f_\sim(\alpha) := \|\sum_{\nu=1}^n \alpha_\nu a_\nu\|_\sim$, diese Funktionen sind stetig auf ihrem Definitionsbereich, weil für $\alpha, \beta \in K^n$ gilt:

$$|f_*(\alpha) - f_*(\beta)| = \left| \left\| \sum_{\nu=1}^n \alpha_\nu a_\nu \right\|_* - \left\| \sum_{\nu=1}^n \beta_\nu a_\nu \right\|_* \right| \leq \left\| \sum_{\nu=1}^n (\alpha_\nu - \beta_\nu) a_\nu \right\|_* \leq \sum_{\nu=1}^n |\alpha_\nu - \beta_\nu| \|a_\nu\| \leq c_* \|\alpha - \beta\|_2 \text{ für } c_* := \left(\sum_{\nu=1}^n \|a_\nu\|^2 \right)^{\frac{1}{2}} \text{ also } < \epsilon \text{ für } \|\alpha - \beta\| < \delta = \frac{\epsilon}{c_*},$$

entsprechendes gilt für $\|\cdot\|_\sim$
Nun ist $S := \{\alpha \in K^n \mid \|\alpha\| = 1\} \subset K^n$ ist kompakt $\Rightarrow d_1 := \max_{\alpha \in S} f_\sim(\alpha)$ und $d_2 := \min_{\alpha \in S} f_*(\alpha)$ existieren mit $d_1, d_2 > 0$, da a_1, \dots, a_n linear unabhängig sind und wegen (N3)

Für $a \in V$ (o.B.d.A $a \neq 0$) $\exists (\alpha_\nu) = \alpha \in K^n$ so daß $\sum_{\nu=1}^n \alpha_\nu a_\nu = a$ (weil $\{a_\nu\}$ Basis von V ist) $\Rightarrow \|a\|_\sim = \left\| \sum_{\nu=1}^n \alpha_\nu a_\nu \right\|_\sim = \|\alpha\|_2 \left\| \sum_{\nu=1}^n \beta_\nu a_\nu \right\|_\sim \stackrel{\beta = \frac{\alpha}{\|\alpha\|_2} \in S}{=} \|\alpha\|_2 f_\sim(\beta) \leq \|\alpha\|_2 d_1 \leq \frac{d_1}{d_2} \|\alpha\|_2 f_*(\beta) = \frac{d_1}{d_2} \|\alpha\|_2 \left\| \sum_{\nu=1}^n \beta_\nu a_\nu \right\|_* = \frac{d_1}{d_2} \left\| \sum_{\nu=1}^n \alpha_\nu a_\nu \right\|_* = c_2 \|a\|_*$ für $c_2 := \frac{d_1}{d_2} > 0$, wegen Symmetrie $(\|\cdot\|_* \sim \|\cdot\|_\sim)$ folgt $\exists \tilde{c}_1$ mit $\|a\|_* \leq \tilde{c}_1 \|a\|_\sim$, d.h. $c_1 \|a\|_* \leq \|a\|_\sim \forall a \in V$ mit $c_1 := (\tilde{c}_1)^{-1}$ □

Korollar 8.4.1. Es sei $\|\cdot\|$ eine beliebige Norm in K^n mit $K = \mathbb{C}$ oder $K = \mathbb{R}$ und $(a_k) = \left(a_\nu^{(k)} \right)_{k=1}^\infty$ sei eine Folge aus K^n sowie $a = (a_\nu) \in K^n$. Dann gilt

$\lim_{k \rightarrow \infty} \|a_k - a\| = 0 \Leftrightarrow \lim_{k \rightarrow \infty} a_\nu^{(k)} = a_\nu \quad \forall \nu = 1, \dots, n$ (Koordinatenweise Konvergenz)

Beweis. $\|a_n - a\| \rightarrow 0 \stackrel{\text{Satz 4.1}}{\Leftrightarrow} \|a_k - a\|_\infty \rightarrow 0$, d.h. $\lim_{k \rightarrow \infty} a_\nu^{(k)} = a_\nu \forall \nu = 1, \dots, n$ □

Zusatz: $\sum_{k=1}^\infty \|a_k\| < \infty \Leftrightarrow \sum_{k=1}^\infty \left| a_\nu^{(k)} \right| < \infty \quad \forall \nu = 1, \dots, n$

Definition 8.4.2. Für $A = (a_{\mu\nu}) \in \mathbb{C}^{m \times n}$ heißt

- (i) $\|A\|_F := \left(\sum_{\mu=1}^m \sum_{\nu=1}^n |a_{\mu\nu}|^2 \right)^{\frac{1}{2}}$ die Frobeniusnorm von A
- (ii) $\|A\|_S := \max_{x \in \mathbb{C} \setminus \{0\}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2$ die Spektralnorm von A
- (iii) $\rho(A) = \max\{|\lambda| \mid \lambda \in \sigma(A)\}$, falls $m = n$, der Spektralradius von A

Proposition 8.4.2. (i) Die Frobeniusnorm und die Spektralnorm sind Normen auf $\mathbb{C}^{m \times n}$

- (ii) $\|A\|_F^2 = \sum_{\nu=1}^n \|a_\nu\|_2^2$ für $A = (a_1, \dots, a_n)$
- (iii) $\|AB\|_F \leq \|A\|_F \|B\|_F \forall A \in \mathbb{C}^{m \times n}, B \in \mathbb{C}^{n \times k}$ und insbesondere $\|Ax\|_2 \leq \|A\|_F \|x\|_2$ für $x \in \mathbb{C}^n$
- (iv) $\|A\|_F = \|A^*\|_F$ und $\|UAV\|_F = \|A\|_F \forall A \in \mathbb{C}^{m \times n}$ und $U \in \mathbb{C}^{m \times m}, V \in \mathbb{C}^{n \times n}$ unitär
- (v) $\|AB\|_S \leq \|A\|_S \|B\|_S \quad \forall A \in \mathbb{C}^{m \times n}, B \in \mathbb{C}^{n \times k}$
und $\|A\|_S \leq \|A\|_F, \|Ax\|_2 \leq \|A\|_S \|x\| \quad \forall x \in \mathbb{C}^n$
- (vi) $\|A\|_S = \|A^*\|_S = \sqrt{\rho(A^*A)} = \sigma_1 = \text{maximaler SW von } A, \forall A \in \mathbb{C}^{m \times n}$
- (vii) Die Spektralnorm ist ebenfalls unitär invariant

Beweis. (i) $\|A\|_F = \|A\|_2$ für $A \in \mathbb{C}^{m \times n}$ „=“ $\mathbb{C}^{m \cdot n}$ ist also eine Norm in $\mathbb{C}^{m \times n}$

Die Spektralnorm erfüllt (N2) und (N3) nach Definition 4.1 und für $A, B \in \mathbb{C}^{m \times n}$ und $x \in \mathbb{C}^n$ folgt:

$$\|(A+B)x\|_2 \|x\|_2 \leq (\|Ax\|_2 + \|Bx\|_2) \|x\|_2 \Rightarrow \|A+B\|_S \leq \|A\|_S + \|B\|_S$$

(ii)

- (iii) Es sei $A \in \mathbb{C}^{m \times n}, x \in \mathbb{C}^n \Rightarrow \|Ax\|_2^2 = \sum_{\mu=1}^m \left| \sum_{\nu=1}^n a_{\mu\nu} x_\nu \right|^2 \stackrel{\text{CSU}}{=} \sum_{\mu=1}^m \left[\sum_{\nu=1}^n |a_{\mu\nu}|^2 \cdot \sum_{s=1}^n |x_s|^2 \right] = \|A\|_F^2 \|x\|_2^2$ und es sei $B = (b_1, \dots, b_n) \Rightarrow \|AB\|_F^2 = \|(Ab_1, \dots, Ab_n)\|_F^2 = \sum_{\nu=1}^n \|Ab_\nu\|_2^2 \leq \|A\|_F^2 \sum_{\nu=1}^n \|b_\nu\|_2^2 = \|A\|_F^2 \|B\|_F^2$

- (iv) $\|A\|_F = \|A^*\|_F$ nach Definition 4.1 (ii) und es sei $V \in \mathbb{C}^{m \times m}$ unitär

$$A = (a_1, \dots, a_n) \Rightarrow \|VA\|_F^2 = \|(Va_1, \dots, Va_n)\|_F^2 = \sum_{\nu=1}^n \|Va_\nu\|_2^2 = \sum_{\nu=1}^n \|a_\nu\|_2^2 = \|A\|_F^2$$

$$\|VAU\|_F = \|AU\|_F = \|U^*A^*\|_F = \|A^*\|_F = \|A\|_F$$

- (v) Für $A \in \mathbb{C}^{m \times n}, B \in \mathbb{C}^{n \times k}, x \in \mathbb{C}^k$ folgt $\|ABx\|_2 \leq \|A\|_S \|Bx\|_2 \leq \|A\|_S \|B\|_S \|x\| \Rightarrow \|AB\|_S \leq \|A\|_S \|B\|_S$

$$\text{Aus } \|Ax\|_2 \leq \|A\|_F \|x\|_2 \Rightarrow \|A\|_S \leq \|A\|_F$$

(vi) $\|A\|_S = \|A^*\|_S = (\rho(A^*A))^{\frac{1}{2}}$ gilt gemäß Kor. 1.3, 1.1 und Definition 4.1 (iii)

(vii) (Übungsaufgabe)

□

Definition 8.4.3. Eine Abbildung $\|\cdot\| : \mathbb{C}^{n \times n} \rightarrow \mathbb{R}$ heißt eine Matrixnorm in $\mathbb{C}^{n \times n}$ falls folgende Eigenschaften gelten für alle $A, B \in \mathbb{C}^{n \times n}, \alpha \in \mathbb{C}$

$$(M1) \quad \|A + B\| \leq \|A\| + \|B\|$$

$$(M2) \quad \|\alpha A\| = |\alpha| \|A\|$$

$$(M3) \quad \|A\| \geq 0 \text{ mit } \Leftrightarrow A = 0$$

$$(M4) \quad \|AB\| \leq \|A\| \|B\| \text{ (submultiplikativ)}$$

Aus (M4) ergibt sich die Folgerung $\|A^k\| \leq \|A\|^k$

Beispiele: (i) Die Frobeniusnorm und die Spektralnorm sind Matrixnormen gemäß Prop. 4.2.

(ii) $\|A\| := \max_{\mu, \nu} |a_{\mu\nu}|$ ist eine Norm in $\mathbb{C}^{n \times n}$, aber dies ist keine Matrixnorm für $n \geq 2$, weil für $A = B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \Rightarrow \|A\| \|B\| = 1 < 2 = \left\| \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \right\| = \|AB\|$

Proposition 8.4.3. Für jede Matrixnorm $\|\cdot\|$ in $\mathbb{C}^{n \times n}$ gilt $\|A\| \geq \rho(A) \forall A \in \mathbb{C}^{n \times n}$

Bemerkung:

(i) $\rho(A) = \|A\|_S$, falls A normal (Übungsaufgabe)

(ii) Für $0 \neq A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ gilt $\rho(A) = 0 < \|A\|$ für jede Norm $\|\cdot\|$

Beweis. Es seien $\|\cdot\|$ eine Matrixnorm in $\mathbb{C}^{n \times n}$, $A \in \mathbb{C}^{n \times n}$, $\lambda \in \sigma(A)$ mit $|\lambda| = \rho(A) \Rightarrow \exists x \in \mathbb{C}^n$ mit $x \neq 0$ und $Ax = \lambda x$. Definiere $X = (x, \dots, x) \in \mathbb{C}^{n \times n} \Rightarrow X \neq 0, \|X\| > 0$
(M3). $AX = (\lambda x, \dots, \lambda x) = \lambda X \Rightarrow |\lambda| \|X\| = \|AX\| \leq \|A\| \|X\| \Rightarrow \rho(A) = |\lambda| \leq \|A\|$ □

Proposition 8.4.4. Es sei also $\|\cdot\|$ eine Matrixnorm in $\mathbb{C}^{n \times n}$ und $S \in \mathbb{C}^{n \times n}$ regulär. Dann wird durch $\|A\|_* = \|S^{-1}AS\|$ für $A \in \mathbb{C}^{n \times n}$ wiederum eine Matrixnorm definiert.

Beweis. $\|\cdot\|_* := \mathbb{C}^{n \times n} \rightarrow \mathbb{R}$ erfüllt die Eigenschaften $\forall A, B \in \mathbb{C}^{n \times n}, \alpha \in \mathbb{C}$

$$(M1) \quad \|A + B\|_* = \|S^{-1}(A + B)S\| = \|S^{-1}AS + S^{-1}BS\| \leq \|S^{-1}AS\| + \|S^{-1}BS\| = \|A\|_* + \|B\|_*$$

$$(M2) \quad \|\alpha A\|_* = \|S^{-1}AS\| = |\alpha| \|S^{-1}AS\| = |\alpha| \|A\|_*$$

$$(M3) \quad \|A\|_* \geq 0 \text{ und } \Leftrightarrow A = 0$$

$$(M4) \quad \|AB\|_* = \|S^{-1}ABS\| = \|(S^{-1}AS)(S^{-1}BS)\| \leq \|S^{-1}AS\| \cdot \|S^{-1}BS\| = \|A\|_* \|B\|_*$$

□

Beispiele: (i) $e^A := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$, die sog. Exponentialreihe konvergiert elementweise absolut für alle $A \in \mathbb{C}^{n \times n}$

Beweis. $\sum_{k=0}^{\infty} \left\| \frac{1}{k!} A^k \right\| = \sum_{k=0}^{\infty} \frac{\|A\|^k}{k!} < \infty \quad \forall A \in \mathbb{C}^{n \times n}$ mit einer beliebigen Matrixnorm z.B. $\|\cdot\|_F \Rightarrow$ Beh. mit Kor. 4.1 (Zusatz) □

(ii) $\sum_{k=0}^{\infty} A^k$, die *Neumannsche Reihe* konvergiert elementweise absolut gegen $(I-A)^{-1}$ falls $A \in \mathbb{C}^{n \times n}$ mit $\|A\| < 1$ für eine Matrixnorm $\|\cdot\|$ gilt

Beweis. $\sum_{k=0}^{\infty} \|A^k\| = \sum_{k=0}^{\infty} \|A\|^k < \infty$, falls $\|A\| < 1 \Rightarrow B := \sum_{k=0}^{\infty} A^k$ konvergiert elementweise

$$B(I - A) = \sum_{k=0}^{\infty} A^k - \sum_{k=0}^{\infty} A^{k+1} = A^0 = I \Rightarrow B = (I - A)^{-1} \quad \square$$

Definition 8.4.4. Es sei $\|\cdot\|$ eine Vektornorm in \mathbb{C}^n

(i) $\|A\| := \max \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$ für $A \in \mathbb{C}^{n \times n}$ heißt die durch $\|\cdot\|$ induzierte Matrixnorm (oder Operatornorm)

(ii) Eine Matrixnorm $\|\cdot\|_M$ in $\mathbb{C}^{n \times n}$ heißt verträglich mit der Vektornorm $\|\cdot\|$ falls gilt $\|Ax\| \leq \|A\|_M \|x\| \quad \forall x \in \mathbb{C}^n, A \in \mathbb{C}^{n \times n}$

Bemerkung: (i) Für eine induzierte Matrixnorm gilt $\|Ax\| \leq c\|x\| \quad \forall x \in \mathbb{C}^n$ für $c = \|A\|$ und dies ist die kleinste Zahl mit dieser Eigenschaft

(ii) Die induzierte Matrixnorm ist die kleinste mit der Vektornorm verträglichen Norm

(iii) Die Frobeniusnorm und die Spektralnorm sind verträglich mit der euklidischen Norm und diese induziert die Spektralnorm [Häufige Bezeichnung $\|A\|_2 := \|A\|_S \leq \|A\|_F$]

Proposition 8.4.5. (M1) $\|(A + B)x\| = \|Ax + Bx\| \leq \|Ax\| + \|Bx\| \leq (\|A\| + \|B\|)\|x\| \Rightarrow \|A + B\| \leq \|A\| + \|B\|$

$$(M2) \quad \|\alpha Ax\| = |\alpha| \|Ax\| \Rightarrow \|\alpha A\| = |\alpha| \|A\|$$

$$(M3) \quad \|A\| \geq 0 \text{ und } \|A\| = 0 \Leftrightarrow A = 0$$

$$(M4) \quad \|ABx\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\| \Rightarrow \|AB\| \leq \|A\| \|B\|$$

Übungsaufgaben:

(i) $\|A\|_1 := \max_{1 \leq \nu \leq n} \sum_{\mu=1}^n |a_{\mu\nu}|$ für $A \in \mathbb{C}^{n \times n}$, die sogenannte Spaltennorm wird induziert von $\|x\|_1 = \sum_{\nu=1}^n |x_\nu|$

(ii) $\|A\|_\infty := \max_{1 \leq \mu \leq n} \sum_{\nu=1}^n |a_{\mu\nu}|$ für $A \in \mathbb{C}^{n \times n}$, die sogenannte Zeilennorm wird induziert von $\|x\|_\infty = \max_{1 \leq \nu \leq n} |x_\nu|$

In diesem Unterkapitel (8.4) können $\mathbb{C}^{n \times n}, \mathbb{C}^n$ überall durch $\mathbb{R}^{n \times n}, \mathbb{R}^n$ ersetzt werden.

9 Matrixfunktionen

9.1 Matrixpolynome

[Platzhalter]Wiederholung aus der Vorlesung „Linearen Algebra 1“ [Platzhalter]

Bemerkung: Insbesondere wegen des Fundamentalsatzes der Algebra betrachten wir in diesem Kapitel durchweg komplexe (insbesondere auch reelle) Matrizen $A \in \mathbb{C}^{n \times n}$

Definition 9.1.1. Für eine Matrix $A \in \mathbb{C}^{n \times n}$ und ein Polynom $Q(x) = \sum_{M=0}^l a_M x^M \in \mathbb{C}[x]$ ist das Matrixpolynom $Q(A) \in \mathbb{C}^{n \times n}$ definiert durch $Q(A) = \sum_{M=0}^l a_M A^M$ mit $A^0 := I_{n \times n}$

Proposition 9.1.1. Es seien $Q(x) = \sum_{M=0}^l a_M x^M, R(x) = \sum_{M=0}^l b_M x^M \in \mathbb{C}[x], A, B \in \mathbb{C}^{n \times n}, \alpha, \beta \in \mathbb{C}$. Dann gilt

- (i) $(\alpha Q + \beta R)(A) = \alpha Q(A) + \beta R(A)$
- (ii) $Q(A^T) = (Q(A))^T$
- (iii) $Q(A) = a_l(A - \lambda_1 I) \cdot \dots \cdot (A - \lambda_l I)$, falls $Q(x) = a_l(x - \lambda_1) \cdot \dots \cdot (x - \lambda_l)$
- (iv) $(Q \cdot R)(A) = Q(A) \cdot R(A)$
- (v) $Q \circ R(A) = Q(R(A))$
- (vi) $S^{-1}Q(A)S = Q(S^{-1}AS) \quad \forall S \in \text{GL}(n, \mathbb{C})$
- (vii) $Q(A) = S \text{diag}(Q(\lambda_r))S^{-1}$ insbesondere ist $A^r = S \text{diag}(\lambda_r^r)S^{-1} \quad \forall r \in \mathbb{N}_0$, falls A durch S diagonalisierbar ist, also $S^{-1}AS = \text{diag}(\lambda_r)$
- (viii) $Ax = \lambda x$ mit $x \in \mathbb{C}^n, \lambda \in \mathbb{C} \Rightarrow Q(A)x = Q(\lambda)x$, insbesondere ist $\text{Eig}(A, \lambda) \subset \text{Eig}(Q(A), Q(\lambda))$
- (ix) $P_{Q(A)}(x) = (-1)^n(x - Q(\lambda_1)) \cdot \dots \cdot (x - Q(\lambda_n))$
- (x) $Q(A)R(B) = R(B)Q(A)$, falls $AB = BA$
- (xi) $\frac{d}{dt}Q(tA) = AQ'(tA)$ für $t \in \mathbb{C}$ mit $Q'(x) = \sum_{M=1}^l a_M M x^{M-1}$

Beweis. (i) klar

(ii) $(A^T)^M = (A^M)^T \Rightarrow Q(A^T) = \sum_{M=0}^l a_M (A^T)^M = (Q(A))^T$

(iii) klar

(iv) $(QR)(x) = \sum_{M=0}^l \sum_{r=0}^m a_M b_r x^{M+r} \Rightarrow (QR)(A) = \sum_{M=0}^l \sum_{r=0}^m a_M b_r A^{M+r} = \sum_{M=0}^l a_M A^M \sum_{r=0}^m b_r A^r = Q(A)R(A)$

(v) klar

$$(vi) \quad S^{-1}Q(A)S = S^{-1}\left(\sum_{M=0}^l a_M A^M\right)S = \sum_{M=0}^l a_M S^{-1}A^M S = \sum_{M=0}^l a_M (S^{-1}AS)(S^{-1}AS)\cdots(S^{-1}AS) \\ \sum_{M=0}^l a_M (S^{-1}AS)^M = \sum_{M=0}^l a_M (S^{-1}AS)^M = Q(S^{-1}AS)$$

(vii) gemäß Bem. Wiederholung

$$(viii) \quad Ax = \lambda x \Rightarrow A^M x = \lambda^M x \forall M \in \mathbb{N}_0 \Rightarrow Q(A)x = \sum_{M=0}^l a_M A^M x = Q(\lambda)x$$

(ix) Gemäß dem Lemma von Schur [Kapitel 7, Satz 1.1] existiert eine unitäre Matrix

$$U \in \mathbb{C}^{n \times n} \text{ mit } U^{-1}AU = \Delta = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \Rightarrow P_A(x) = (-1)^n (x - \lambda_1) \cdots (x - \lambda_n) \\ \text{und } P_{Q(A)} = P_{U^{-1}Q(A)U}(x) = P_{Q(\Delta)}(x) = (-1)^n (x - Q(\lambda_1)) \cdots (x - Q(\lambda_n))$$

$$(x) \quad AB = BA \Rightarrow A^M B^R = B^R A^M \forall M, R \in \mathbb{N}_0 \Rightarrow Q(A)R(B) = \sum_{M=0}^l \sum_{R=0}^m a_M A^M b_R B^R = \\ \sum_{R=0}^m b_R B^R \sum_{M=0}^l a_M A^M = R(B)Q(A)$$

$$(xi) \quad \frac{d}{dt}Q(tA) = \frac{d}{dt} \sum_{M=0}^l a_M t^M A^M = \sum_{M=1}^l a_M M t^{M-1} A^{M-1+1} = A \sum_{M=1}^l a_M M t^{M-1} A^{M-1} = \\ A Q'(tA)$$

□

Proposition 9.1.2. Für $A, B \in \mathbb{C}^{n \times n}$ gilt $(A + B)^\nu = \sum_{M=0}^\nu \binom{\nu}{M} A^M B^{\nu-M} \forall \nu \in \mathbb{N}_0$, falls $AB = BA$ gilt.

Beweis. wie Binomialformel

□

Bemerkung: $(A + B)^2 = (A + B)(A + B) = A^2 + AB + BA + B^2 = A^2 + 2AB + B^2 \Leftrightarrow AB = BA$

Proposition 9.1.3 (Leibniz-Formel). Bei Existenz der Ableitungen gilt: $(f \cdot g)^{(\nu)}(t) = \sum_{M=0}^\nu \binom{\nu}{M} f^{(M)} g^{(\nu-M)}(t)$

Beweis. (dito mit Produktregel)

□

Satz 9.1.1 (Hermite-Interpolation). Gegeben seien $k, n_1, \dots, n_k \in \mathbb{N}, n := \sum_{\nu=1}^k n_\nu$, beliebige Zahlen $c_{\mu\nu} (= f^{(\mu)}(\lambda_\nu)) \in \mathbb{C}$ für $0 \leq \mu \leq n_\nu - 1, \nu = 1, \dots, k$, λ_ν paarweise verschiedene Zahlen.

Dann gibt es genau ein Polynom $Q \in \mathbb{C}[x]$ vom Grade $\leq n - 1$ mit $Q^{(\mu)}(\lambda_\nu) = c_{\mu\nu} \quad \forall 0 \leq \mu \leq n_\nu - 1, \nu = 1, \dots, k$

und dieses Hermitesche Interpolationspolynom ist gegeben durch

$$(\cdot) \quad Q(\lambda) = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{1}{\mu!} \left(\frac{f}{q_\nu}\right)^{(\mu)}(\lambda_\nu) (x - \lambda_\nu)^\mu q_\nu(x) \text{ mit } q_\nu(x) = \prod_{\mu=0, \mu \neq \nu}^k (x - \lambda_\mu)^{n_\mu}$$

Beweis. Existenz: Es sei Q gegeben durch $(\cdot) \Rightarrow Q \in \mathbb{C}[x]$ mit $\text{Grad } Q \leq \max_{\nu=1, \dots, k} (n_\nu - 1 + n - n_\nu) = n - 1$ und für $0 \leq r \leq n_s - 1, s = 1, \dots, k$ folgt

$$\begin{aligned}
Q^{(r)}(\lambda_s) &= \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu} \frac{1}{\mu!} \left(\frac{f}{q_\nu}\right)^{(\mu)}(\lambda_\nu) \left(\frac{d}{dx}\right)^r [(x - \lambda_\nu)^\mu q_\nu(x)]_{x=\lambda_s} \\
&= \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu} \frac{1}{\mu!} \left(\frac{f}{q_\nu}\right)^{(\mu)}(\lambda_\nu) \delta_{s\nu} \mu! q_\nu^{(r-\mu)}(\lambda_s) \binom{\nu}{\mu} \\
&= \sum_{\mu=0}^r \left(\frac{f}{q_s}\right)^{(\mu)}(\lambda_s) q_s^{(r-\mu)}(\lambda_s) \binom{r}{\mu} \underset{=0 \text{ für } \mu > r}{=} \\
&= \left(\frac{f}{q_s} \cdot q_s\right)^{(r)}(\lambda_s) = f^{(r)}(\lambda_s) = c_{rs}
\end{aligned}$$

Eindeutigkeit: Es erfüllen Q_1, Q_2 die Behauptung $\Rightarrow Q := Q_1 - Q_2 \in \mathbb{C}[x] \Rightarrow \text{Grad } Q \leq n-1, Q^{(\mu)}(\lambda_\nu) = c_{\mu\nu} - c_{\mu\nu} = 0 \quad \forall 0 \leq \mu \leq n_\nu$ für $\nu = 1, \dots, k \Rightarrow$ Das Polynom besitzt $\sum_{\nu=1}^k n_\nu = n$ Nullstellen $\Rightarrow Q = 0 \quad \square$

Für $n_\nu = 1, \nu = 1, \dots, k$ also $k = n$ folgt das

Korollar 9.1.1 (Lagrange-Interpolation). Für $n \in \mathbb{N}, c_1, \dots, c_n \in \mathbb{C}$ paarweise verschiedene $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ gibt es genau ein Polynom $Q(x) \in \mathbb{C}[x]$ vom Grade $\leq n-1$ mit $Q(\lambda_\nu) = c_\nu \forall \nu = 1, \dots, n$ und dieses sogenannte Lagrange-Interpolations-Polynom: $Q(x) = \sum_{\nu=1}^k f(\lambda_\nu) \frac{q_\nu(x)}{q_\nu(\lambda_\nu)}$ mit $q_\nu(x) = \prod_{\mu=1, \mu \neq \nu}^n (x - \lambda_\mu)$

9.2 Frobeniussche Kovarianten

Wir verwenden durchweg folgende Bezeichnungen: $A \in \mathbb{C}^{n \times n}$ mit $P_A(x) = \det(A - xI) = \sum_{\mu=0}^n \alpha_\mu x^\mu = (-1)^n (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k}, \lambda_1, \dots, \lambda_k \in \mathbb{C}$ paarweise verschieden mit $n_\nu = n(\lambda_\nu)$ für $\nu = 1, \dots, k, q_\nu(x) = \prod_{\mu=1, \mu \neq \nu}^k (x - \lambda_\mu)^{n_\mu}, q(x) := \sum_{\nu=1}^k q_\nu(x)$ [für $\nu = 1, q(x) = q_1(x) = 1]$
 $B_\nu = q_\nu(A)$, für $\nu = 1, \dots, k, B = q(A) = \sum_{\nu=1}^k B_\nu$ ($n = 1 \Rightarrow B_1 = B = I$)

Lemma. (i) B ist regulär

$$(ii) (A - \lambda_\nu I)^{n_\nu} B_\nu = 0$$

$$(iii) \text{Rang}(A - \lambda_\nu I)^m \geq n - n_\nu \quad \forall m \in \mathbb{N}$$

Beweis. (i) $\det B = \prod$ Eigenwerte von $B (= q(A)) \stackrel{\text{Prop. 1.1(ix)}}{=} q(\lambda_1) \cdots q(\lambda_n) \neq 0$, weil $q(\lambda_\nu) = q_\nu(\lambda_\nu) \neq 0$ weil $\lambda_\mu \neq \lambda_\nu$

$$(ii) (A - \lambda_\nu I)^{n_\nu} B_\nu = (A - \lambda_\nu I)^{n_\nu} \prod_{\mu=1, \mu \neq \nu}^k (A - \lambda_\mu I)^{n_\mu} = (-1)^n P_A(A) = 0 \text{ gemäß dem Satz Cayley-Hamilton}$$

$$(iii) \text{ nach Lemma von Schur } \exists U \text{ unitär mit } U^{-1}AU = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_1 & & * \\ & & & \lambda_2 & \\ & & & & \ddots \\ & & & & & \lambda_n \end{pmatrix} \Rightarrow$$

$$\text{Rang}(A - \lambda_1 I)^m = \text{Rang} \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & * \\ & & & \lambda_2 - \lambda_1 & \\ & & & & \ddots \\ & & & & & \lambda_n - \lambda_1 \end{pmatrix} \geq n - n_1$$

o.B.d.A. $\nu = 1$ □

Definition 9.2.1. Die Matrizen $C_\nu = B_\nu B^{-1}$ für $\nu = 1, \dots, k$ heißen die Frobeniusschen Kovarianten der Matrix A

Bemerkung: A, B_ν, B, C_ν kommutieren, da sie Polynome von A sind. Im Falle der Existenz ist $P_A(\lambda) =: \sum_{\mu=1}^n \alpha_\mu x^\mu \Rightarrow A^{-1} = \sum_{\mu=1}^n -\frac{\alpha_\mu}{\alpha_0} A^{\mu-1}$

Proposition 9.2.1 (Eigenschaften der Frobeniusschen Kovarianten). (i) $\sum_{\nu=1}^k C_\nu = I$

$$(ii) C_\mu C_\nu = 0 \quad \forall \mu \neq \nu \text{ und } C_\nu^2 = C_\nu \quad \nu = 1, \dots, k$$

$$(iii) \mathbb{C}^n = \text{Im } C_1 \oplus \dots \oplus \text{Im } C_n$$

$$(iv) (A - \lambda_\nu I)^{n_\nu} C_\nu = 0 \quad \forall \nu = 1, \dots, k$$

$$(v) \text{Rang } C_\nu = n_\nu, \text{Rang}(A - \lambda_\nu I)^{n_\nu} = n - n_\nu$$

$$\text{Beweis.} \quad (i) \sum_{\nu=1}^k C_\nu = (\sum_{\nu=1}^k B_\nu B^{-1} = BB^{-1} = I$$

$$(ii) (A - \lambda_\nu I)^{n_\nu} B_\nu = 0 \Rightarrow C_\mu C_\nu = 0 \quad \forall \mu \neq \nu \text{ (da } P(A) | C_\mu C_\nu \text{) und } C_\nu^2 = C_\nu (\sum_{\mu=1}^k C_\mu) = C_\nu$$

$$(iii) \mathbb{C}^n = \text{Im } C_1 + \dots + \text{Im } C_n \text{ gemäß (i), und die Summe ist direkt, denn es sei } x \in C_\nu \cap \sum_{\mu=1, \mu \neq \nu}^k C_\mu d_\mu \Rightarrow x = C_\nu d = C_\nu^2 d = C_\nu^2 d = C_\nu x = C_\nu \sum_{\mu=1, \mu \neq \nu}^k C_\mu d_\mu = 0$$

$$(iv) (A - \lambda_\nu)^{n_\nu} C_\nu = (A - \lambda_\nu)^{n_\nu} B_\nu B^{-1} = 0 B^{-1} = 0$$

$$(v) \text{Im } C_\nu \subset \text{Kern}(A - \lambda_\nu I) \Rightarrow \text{Rang } C_\nu \leq \dim \text{Kern}(A - \lambda_\nu I)^{n_\nu} = n - \text{Rang}(A - \lambda_\nu I)^{n_\nu} \leq n - (n - n_\nu) = n_\nu$$

$$n = \sum_{\nu=1}^k \text{Rang } C_\nu \leq \sum_{\nu=1}^k n_\nu = n \Rightarrow \text{Rang } C_\nu = n_\nu \Rightarrow \text{Rang}(A - \lambda_\nu I)^{n_\nu} = n - n_\nu$$
□

Beispiel: $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$, $\lambda_1 = 1, n_1 = 2, \lambda_2 = 2, n_2 = 1 \Rightarrow q_1(x) = x - 2, q_2(x) = (x - 1)^2$

$$B_1 = A - 2I = \begin{pmatrix} -1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} 0 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, B = B_1 + B_2 = \begin{pmatrix} -1 & 2 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, B^{-1} = \begin{pmatrix} -1 & -2 & 8 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, C_1 = B_1 B^{-1} = \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}, C_2 = B_2 B^{-1} = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Satz 9.2.1 (Matrixpotenzen und polynome). Für $A \in \mathbb{C}^{n \times n}$, $Q(x) = \sum_{\mu=0}^k a_\mu x^\mu \in \mathbb{C}[x]$

$$(1) Q(A) = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{Q^{(\mu)}(\lambda_\nu)}{\mu!} (A - \lambda_\nu I)^\mu C_\nu$$

$$(2) A^r = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \binom{r}{\mu} \lambda_\nu^{r-\mu} (A - \lambda_\nu I)^\mu C_\nu$$

Beweis. $r \in \mathbb{N}_0, \nu = \{1, \dots, k\}$. Dann gilt $(A - \lambda_\nu)^\mu C_\mu = 0$ für $\mu \geq n_\nu$ (Prop 2.1. (iv)) und $\binom{r}{\mu} = 0$ für $\mu > r$

$$\Rightarrow \sum_{\mu=0}^{n_\nu-1} \binom{r}{\mu} \lambda_\nu^{r-\mu} (A - \lambda_\nu I) C_\nu = \sum_{\mu=0}^r \binom{r}{\mu} \lambda_\nu^{r-\mu} (A - \lambda_\nu I) C_\nu$$

$$= (\lambda_\nu I + (A - \lambda_\nu I))^r C_\nu = A^r C_\nu \text{ Also gilt } A^r = I A^r = \sum_{\nu=1}^k C_\nu A^r = (2) \quad \square$$

Nun gilt $\binom{r}{\mu} \lambda_\nu^{r-\mu} = \frac{1}{\mu!} (x^r)^\mu (\lambda_\nu) \Rightarrow Q(A) = \sum_{r=0}^k a_r A^r \stackrel{(1)}{=} \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \left[\sum_{r=0}^l a_r (x^r)^{(\mu-r)} \right] \frac{1}{\mu!} (A - \lambda_\nu I)^\mu C_\nu = (2)$

Satz 9.2.2 (Diagonalisierbarkeit, 2. Teil). $A \in \mathbb{C}^{n \times n}$ ist genau dann diagonalisierbar wenn gilt:

$$(3) A = \sum_{\nu=1}^k \lambda_\nu C_\nu \text{ und dann gilt } Q(A) = \sum_{\nu=1}^n Q(\lambda_\nu) C_\nu$$

Beweis. A sei diagonalisierbar $\Rightarrow m_A(x) = (x - \lambda_1) \cdots (x - \lambda_k) \Rightarrow (A - \lambda_1 I) \cdots (A - \lambda_k I) = 0 \Rightarrow (A - \lambda_\nu I)^1 C_\nu = 0 \forall 1 \leq \nu \leq k \stackrel{(2)}{\Rightarrow} A = \sum_{\nu=1}^k \lambda_\nu^1 C_\nu \Rightarrow (3)$

Es gelte (3) $\Rightarrow A C_\nu = (\sum_{\mu=1}^k \lambda_\mu) C_\mu C_\nu = \lambda_\nu C_\nu$, d.h. $\text{Im } C_\nu \subset \text{Eig}(A, \lambda_\nu) \Rightarrow \rho(\lambda_\nu) = \dim \text{Kern}(A - \lambda_\nu) \geq \text{Rang } C_\nu = n_\nu \geq \rho(\lambda_\nu) \Rightarrow \rho(\lambda_\nu) = n_\nu \quad \forall \nu = 1, \dots, k \Rightarrow A$ ist diagonalisierbar \square

9.3 Die Jordansche Normalform

Der vorherige Beweis aus Satz 2.2 lässt sich ausbauen zum Beweis des folgenden Satzes

Satz 9.3.1 (Jordansche Normform). Für jede Matrix $A \in \mathbb{C}^{n \times n}$ gibt es $r \in \mathbb{N}, m_1, \dots, m_r \in \mathbb{N}$ mit $\sum_{\nu=1}^r m_\nu = n, \lambda_1, \dots, \lambda_r \in \mathbb{C}$ und $S \in \text{GL}(n, \mathbb{C})$ so daß gilt

$$S^{-1}AS = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_r \end{pmatrix} \text{ mit } J_\nu = \begin{pmatrix} \lambda_\nu & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_\nu \end{pmatrix} \in \mathbb{C}^{m_\nu \times m_\nu}$$

Definition 9.3.1. Ein Vektor $x \in \mathbb{C}^n$ heißt ein Hauptvektor p.ter Stufe zum Eigenwert λ der Matrix $A \in \mathbb{C}^{n \times n}$ falls gilt:

$$(A - \lambda I)^p x = 0 \text{ und } (A - \lambda I)^{p-1} x \neq 0$$

Beachte: $(A - \lambda_\nu I)^{n_\nu} C_\nu = 0$ gemäß Proposition 2.1.

Bemerkung: (i) Hauptwerte der 1. Stufe sind Eigenvektoren

(ii) Es gilt $k \leq r \leq n$ und natürlich gilt $\sigma(A) = \{\lambda_1, \dots, \lambda_k\} = \{\lambda_\nu | \nu = 1, \dots, r\} (m_\nu \leq n_\nu)$

(iii) A ist diagonalisierbar $\Leftrightarrow m_\nu = 1 \forall \nu = 1, \dots, r = n$

(iv) Die Jordansche Normalform $J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_r \end{pmatrix}$ ist eindeutig bis auf die Reihenfolge

9.4 Matrixfunktionen

Der Satz 2.1 führt zur folgenden Definition (mit den Bezeichnungen aus Unterkapitel 2)

Definition 9.4.1. Es seien $A \in \mathbb{C}^{n \times n}$, $f : D \rightarrow \mathbb{C}$ mit $\sigma(A) \subset D$ und $f^{(n_\nu-1)}(\lambda_\nu)$ existiert $\forall \nu = 1, \dots, k$

Dann ist die Matrixfunktion $f(A) \in \mathbb{C}^{n \times n}$ definiert durch

$$(*) f(A) = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{f^{(\mu)}(\lambda_\nu)}{\mu!} (A - \lambda_\nu I)^\mu C_\nu$$

Beispiele: $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ wie in Unterkapitel 2 mit $\lambda_1 = 1, n_1 = 2, \lambda_2 = 2, n_2 = 1$

und $C_1 = \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}, C_2 = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$

$$\stackrel{\text{Def. 4.1}}{\Rightarrow} e^A = e^1 C_1 + e^2 \begin{pmatrix} 0 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = e^1 \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} + e^2 \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} e & 2e & 3e^2 - 5e \\ 0 & e & e^2 - e \\ 0 & 0 & e^2 \end{pmatrix}$$

Proposition 9.4.1. *Es sei $A \in \mathbb{C}^{n \times n}$, f und f_r für $r \in \mathbb{N}$ wie in Definition 4.1. mit $\lim_{r \rightarrow \infty} f_r^{(\mu)}(\lambda_\nu) = f^{(\mu)}(\lambda_\nu)$.*

Dann gilt $\lim_{r \rightarrow \infty} f_r^{(\mu)}(\lambda_\nu) = f(A)$ [Elementweise Konvergenz]

Beweis. Voraussetzung & Definition 4.1 □

Korollar 9.4.1. *Es sei $A \in \mathbb{C}^{n \times n}$ und $f(z) = \sum_{r=0}^{\infty} a_r z^r$ sei eine Potenzreihe mit $R > \rho(A)$. Dann gilt $f(A) = \sum_{r=0}^{\infty} a_r A^r$*

Proposition 9.4.2. *Seien A, f und g wie in Definition 4.1.*

(i) $f(A) = \sum_{r=0}^{n_\nu-1} c_r A^r$ ist ein Polynom in A vom Grade $\leq n - 1$ mit Koeffizienten $c_r = c_r(f, A) \in \mathbb{C}$

(ii) $f(A) = g(A)$, falls $f^{(\mu)}(\lambda_\nu) = g^{(\mu)}(\lambda_\nu) \forall \nu = 1, \dots, k, 0 \leq \mu \leq n_\nu - 1$

(iii) $f(A) = Q_f(A)$ für $Q_f(x) = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{1}{\mu!} \left(\frac{f}{q_\nu}\right)^{(\mu)}(\lambda_\nu)(x - \lambda_\nu)^\mu q_\nu(x)$ für $q_\nu(x) = \prod_{\nu \neq \mu} (x - \lambda_\nu)^{n_\nu} \Rightarrow q(A) = B_\nu$

Beweis. (i) folgt aus (ii), (ii) folgt aus Def. 4.1, (iii) folgt aus (ii) und Satz 1.1 □

Satz 9.4.1 (Rechenregeln für Matrixfunktionen). $A, B \in \mathbb{C}^{n \times n}, \alpha, \beta \in \mathbb{C}$ und f, g seien gegebene Funktionen. Dann gilt, falls die auftretenden Matrixfunktionen jeweils gemäß Definition 4.1 wohldefiniert sind folgendes:

(i) $(\alpha f + \beta g)(A) = \alpha f(A) + \beta g(A)$

(ii) $f(A^T) = (f(A))^T$

(iii) $(f \cdot g)(A) = f(A)g(A) = g(A)f(A)$

(iv) $(f \circ g)(A) = f(g(A))$

(v) $S^{-1}f(A)S = f(S^{-1}AS) \quad \forall S \in \text{GL}(n, \mathbb{C})$

(vi) $f(A) = S \text{diag}(f(\lambda_\nu))S^{-1} = \sum_{\nu=1}^k f(\lambda_\nu)C_\nu$, falls A diagonalisierbar ist.

(vii) $Ax = \lambda x$ für $x \in \mathbb{C}^n, \lambda \in \mathbb{C} \Rightarrow f(A)x = f(\lambda)x$ insbesondere $\text{Eig}(A, \lambda) \subset \text{Eig}(f(A), f(\lambda))$

(viii) $P_{f(A)}(x) = (-1)^n(x - f(\lambda_1)) \cdots (x - f(\lambda_n))$, falls $P_A(\lambda) = (x - \lambda_1) \cdots (x - \lambda_n)$

(ix) $f(A)g(B) = g(B)f(A)$ falls $AB = BA$

Beweis. Die Behauptungen folgen aus Proposition 1.1 und 4.2 (ii) via Hermite-Interpolation mit Satz 1.2 sowie Satz 2.1 □

Bemerkung (Matrixexponentialfunktion e^A): Für $A \in \mathbb{C}^{n \times n}$ gilt

$$e^A = \sum_{r=0}^{\infty} \frac{1}{r!} A^r = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{e^{\lambda_\nu}}{\mu!} (A - \lambda_\nu I)^\mu C_\nu \text{ und } e^{A+B} = e^A e^B \text{ für } B \in \mathbb{C}^{n \times n} \text{ falls } AB = BA$$

Beispiel: e^{tA} für $t \in \mathbb{R}$ und $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ (Betrachte die Funktion $f(z) = e^{tz}$ bei festen t)

$$P_A(x) = x^2 + 1, \lambda_1 = i, \lambda_2 = -i, A \text{ somit diagonalisierbar. } B_1 = \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix}, B_2 = \begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix}, B = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}, B^{-1} = \frac{1}{2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, C_1 = B_1 B^{-1} = \frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \Rightarrow e^{-it} \dots = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix}$$

Korollar 9.4.2. Es sei $A \in \mathbb{C}, I \subset \mathbb{R}$ ein Intervall und $f(tA)$ existiere für alle $t \in I$. Dann gilt:

$$(i) f(tA) = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{t^\mu}{\mu!} f^{(\mu)}(t\lambda_\nu) (A - \lambda_\nu I)^\mu C_\nu \text{ und}$$

$$(ii) \frac{d}{dt} f(tA) = A f'(tA) \text{ für } t \in I, \text{ falls } f'(tA) \text{ ebenfalls auf } I \text{ existiert}$$

Beweis. (i) Es sei $t \in I$ fest, $f(tz) = (f \circ g)(z) =: h(z)$ für $g(z) := tz$

$$\Rightarrow h^{(\mu)}(\lambda_\nu) = t^\mu f^{(\mu)}(t\lambda_\nu) \text{ für } 0 \leq \mu \leq n_\nu - 1, \nu = 1, \dots, k \text{ und } g(A) = tA$$

$$f(tA) = f(g(A)) = (f \circ g)(A) = h(A) = \sum_{\mu=1}^k \sum_{\nu=0}^{n_\nu-1} \frac{h^{(\mu)}(\lambda_\nu)}{\mu!} (A - \lambda_\nu I)^\mu C_\nu \Rightarrow \text{Behauptung}$$

(ii)

$$\begin{aligned} A f'(tA) &= (A - \lambda_\nu I + \lambda_\nu I) \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{t^\mu}{\mu!} f^{(\mu+1)}(tA) (A - \lambda_\nu I)^\mu C_\nu \\ &= \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{t^\mu}{\mu!} f^{(\mu+1)}(tA) (A - \lambda_\nu I)^{\mu+1} C_\nu + \lambda_\nu \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{t^\mu}{\mu!} f^{(\mu+1)}(tA) (A - \lambda_\nu I)^\mu C_\nu \\ &= \sum_{\nu=1}^k \sum_{\rho=1}^{n_\nu-1} \frac{t^{\rho-1}}{(\rho-1)!} f^{(\rho)}(tA) (A - \lambda_\nu I)^\rho C_\nu + \lambda_\nu \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{t^\mu}{\mu!} f^{(\mu+1)}(tA) (A - \lambda_\nu I)^\mu C_\nu \\ &= \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \left[\frac{t^{\mu-1}}{\mu!} \mu f^{(\mu)}(t\lambda_\nu) + \frac{t^\mu}{\mu!} \lambda_\nu f^{(\mu+1)}(t\lambda_\nu) \right] (A - \lambda_\nu I)^\mu C_\nu \\ &= \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{d}{dt} \frac{t^\mu}{\mu!} f^{(\mu)}(t\lambda_\nu) \\ &= \frac{d}{dt} f(tA) \end{aligned}$$

□

Für $f(z) = e^z$ ergibt sich folgendes

Satz 9.4.2 (Matrixexponentialfunktion). Für $A, B \in \mathbb{C}^{n \times n}$ gilt:

$$(i) \quad e^{tA} = \sum_{\nu=0}^{\infty} \frac{A^\nu}{\nu!} t^\nu = \sum_{\nu=1}^k \sum_{\mu=0}^{n_\nu-1} \frac{t^\mu}{\mu!} e^{t\lambda_\nu} (A - \lambda_\nu I)^\mu C_\nu \quad \forall t \in \mathbb{R}$$

$$(ii) \quad \frac{d}{dt} e^{tA} = A e^{tA} \quad \forall t \in \mathbb{R}$$

$$(iii) \quad e^{tA} e^{tB} = e^{t(A+B)} \Leftrightarrow AB = BA$$

$$(iv) \quad e^{tA} e^{sA} = e^{(t+s)A}, (e^{tA})^{-1} = e^{-tA} \text{ für alle } t \in \mathbb{R} \text{ und } e^0 = I$$

Beweis: (i) & (ii) folgen aus Korollar 4.2 und 4.1 für $f(z) := z$

$$(iii) \quad AB = BA \Rightarrow (tA)(tB) = (tB)(tA) \quad \forall t \in \mathbb{R} \xrightarrow{\ddot{U}24} e^{tA} e^{tB} = e^{t(A+B)}$$

$$\text{Nun gelte } e^{tA} e^{tB} = e^{t(A+B)} \quad \forall t \in \mathbb{R}$$

$$\begin{aligned} \left. \left(\frac{d}{dt} \right)^2 e^{t(A+B)} \right|_{t=0} &= (A+B)^2 = A^2 + AB + BA + B^2 \\ \left. \left(\frac{d}{dt} \right)^2 e^{tA} e^{tB} \right|_{t=0} &= \frac{d}{dt} (Ae^{tA} e^{tB} + e^{tA} B e^{tB}) \Big|_{t=0} \\ &= A^2 e^{tA} e^{tB} + Ae^{tA} B e^{tB} + Ae^{tA} B e^{tB} + e^{tA} B^2 e^{tB} \Big|_{t=0} \\ &= A^2 + AB + BA + B^2 \Rightarrow BA = AB \end{aligned}$$

(iv) folgt aus 3. [s = -t]

□

9.5 Matrixwertige Funktion einer Variablen

Definition 9.5.1. Für eine $m \times n$ -matrixwertige Funktion $X(t) = (x_{\mu\nu}(t)) : I \rightarrow \mathbb{C}^{m \times n}$ definiert man Differenziation und Integrale Elementweise, falls existent, also:

$$\frac{d}{dt} X(t) = X'(t) = \dot{X}(t) := (x'_{\mu\nu}(t)) \in \mathbb{C}^{m \times n} \text{ und } \int_a^b X(t) dt = \left(\int_a^b x_{\mu\nu}(t) dt \right) \in \mathbb{C}^{m \times n}$$

Beispiele: (i) $X(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} : \mathbb{R} \rightarrow \mathbb{R}^{2 \times 1}$ mit $\dot{X}(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}, \int_0^\pi X(t) dt =$

$$\begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

(ii) $X(t) = e^{tA} : \mathbb{R} \rightarrow \mathbb{C}^{n \times n}$, falls $A \in \mathbb{C}^{n \times n}$ mit $\dot{X}(t) = A e^{tA} = A X(t)$

(iii) $X(t) = \{\cos(tA)\}^{-1}$, falls existent $\Rightarrow \dot{X}(t) = A - \sin(At)(\cos(tA))^{-2}$

(iv) $X(t) = \begin{pmatrix} 1 & t \\ 0 & t \end{pmatrix}$, $\dot{X}(t) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$, $X^{-1}(t) = \begin{pmatrix} 1 & -1 \\ 0 & \frac{1}{t} \end{pmatrix}$, $\frac{d}{dt}X^{-1}(t) = \begin{pmatrix} 0 & 0 \\ 0 & -\frac{1}{t^2} \end{pmatrix}$ mit
 $\dot{X}(t)X^{-2}(t) = \begin{pmatrix} 0 & -\frac{1}{t^2} \\ 0 & -\frac{1}{t^2} \end{pmatrix} \neq \frac{d}{dt}X^{-1}(t)$

Proposition 9.5.1 (Rechenregeln).

(i) $\frac{d}{dt}(\alpha X(t) + \beta Y(t)) = \alpha \dot{X}(t) + \beta \dot{Y}(t)$, $\int_a^b (\alpha X(t) + \beta Y(t)) dt = \alpha \int_a^b X(t) dt + \beta \int_a^b Y(t) dt$
 im Fall der Existenz, wobei $\alpha, \beta \in \mathbb{C}$, $X(t), Y(t) \in \mathbb{C}^{m \times n}$

(ii) $\frac{d}{dt}(X(t)Y(t)) = \dot{X}(t)Y(t) + X(t)\dot{Y}(t)$ (Produktregel) und $\int_a^b \dot{X}(t)Y(t) dt = X(t)Y(t)|_a^b - \int_a^b X(t)\dot{Y}(t) dt$ (Partielle Integration)

im Falle der Existenz, wobei $X(t) \in \mathbb{C}^{m \times n}$, $Y(t) \in \mathbb{C}^{n \times k}$, wobei die Reihenfolge der Faktoren zu beachten ist.

Beweis. (klar, ergibt sich elementweise aus den Rechenregeln aus Analysis 1) □

Beispiel: $\frac{d}{dt}X^2(t) = \dot{X}(t)X(t) + X(t)\dot{X}(t) \neq 2X(t)\dot{X}(t)$

Satz 9.5.1 (Differenziation der Inversen). *Es sei $X(t) \in \mathbb{C}^{n \times n}$ regulär und elementweise differenzierbar für $t \in I$ ($I \subset \mathbb{R}$ ein Intervall)*

Dann ist $X^{-1}(t)$ differenzierbar mit $\frac{d}{dt}X^{-1}(t) = -X^{-1}(t)\dot{X}(t)X^{-1}(t)$

Beweis. $X(t)$ regulär $\forall t \in I \Rightarrow \det X(t) \neq 0$ mit der Formel für Inverse (LA1, Kapitel 4, Satz 3.2) $X^{-1}(t) = \frac{1}{\det X(t)} ((-1)^{\nu+\mu} \det(\dots))^T \Rightarrow X^{-1}(t)$ ist differenzierbar $\forall t \in I$

$$0 = \frac{d}{dt}I = \frac{d}{dt}(X(t)X^{-1}(t)) = \dot{X}(t)X^{-1}(t) + X\dot{X}^{-1}(t) \Rightarrow \text{Behauptung}$$

□

9.6 Lineare Systeme mit konstanten Koeffizienten

Beispiel: Gegeben $A \in \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Löse die Differenzialgleichung $\dot{x} = Ax$, d.h. man

bestimme $x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \in \mathbb{R}^2$ mit $\frac{d}{dt}x(t) = Ax(t) \forall t \in \mathbb{R}$ also $\dot{x}_1(t) = \dot{x}_2(t)$ und $\dot{x}_2(t) = -x_1(t) \Leftrightarrow y(t) = x_1(t), \dot{y}(t) = x_2(t)$ mit $y(t) + \ddot{y}(t) = 0$

Es sei $A \in \text{bb}C^{n \times n}$. Dann heißt $\dot{x} = Ax$ ein homogenes lineares Differenzialgleichungssystem (kurz: ein lineares System) mit konstanten Koeffizienten $A = (a_{\mu\nu})$, d.h. $\dot{x}_\mu(t) = \sum_{\nu=1}^n a_{\mu\nu}x_\nu(t)$ für $\mu = 1, \dots, n, t \in \mathbb{R}$

Satz 9.6.1 (Fundamentalmatrix). *A Es sei $A \in \mathbb{C}^{n \times n}$. Dann gilt*

(i) $\dot{X}(t) = AX(t)$ mit $\det X(t) \neq 0 \forall t \in \mathbb{R}$ für $X(t) = e^{tA} \in \mathbb{C}^{n \times n}$, d.h. $X(t) = e^{tA}$ ist eine Fundamentalmatrix und die Spalten bilden ein sogenanntes Fundamentalsystem des Gleichungssystems $\dot{x} = Ax$

(ii) Für $x \in \mathbb{C}^n$ gilt $\dot{x} = Ax$ auf $\mathbb{R} \Leftrightarrow x(t) = e^{tA}c \forall t \in \mathbb{R}$ und ein $c \in \mathbb{C}^n$, d.h. $x(t) = \sum_{\nu=1}^n c_\nu s_\nu(t)$, wobei $e^{tA} = (s_1(t), \dots, s_n(t))$

kurz: Lösungsgesamtheit $\{x | \dot{x} = Ax\} = \{\sum_{\nu=1}^n c_\nu [\text{Platzhalter}]\}$

Beweis. (i) folgt aus Satz 4.2 (ii) und (iv)

(ii) Für $c \in \mathbb{C}^n$, $x = e^{tA}c$ gilt $\frac{d}{dt}x(t) = Ax(t)$.

Nun gelte $\dot{x}(t) = Ax(t)$. Betrachte eine neue Funktion $z(t) = e^{-tA}x(t) \Rightarrow \dot{z}(t) \stackrel{e^{tA}, A \text{ komm.}}{=} e^{-tA}(\dot{x}(t) - Ax(t)) = 0 \Rightarrow z(t) = c \in \mathbb{C}^n$ konstant auf $\mathbb{R} \Rightarrow x(t) = e^{tA}z(t) = e^{tA}c$ □

Korollar 9.6.1 (Anfangswertproblem). *Es seien $A \in \mathbb{C}^{n \times n}$, $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{C}^n$. Dann ist $x(t) = e^{(t-t_0)A}x_0 \forall t \in \mathbb{R}$ die eindeutige Lösung des Anfangswertproblems $\dot{x} = Ax$ mit $x(t_0) = x_0$*

Beweis. $\dot{x} = Ax \Leftrightarrow \exists c \in \mathbb{C}^n x = e^{tA}c$ und dann ist $x(t_0) = c \Leftrightarrow e^{t_0A}c = x_0$ □

Übungsaufgaben. (i) *Variationen der Konstanten.* *Es sei $A \in \mathbb{C}^{n \times n}$, $b, x_0 \in \mathbb{C}^n$, $t \in \mathbb{R}$. Dann ist $x(t) = e^{(t-t_0)A}x_0 + \int_{t_0}^t e^{(t-s)A}b ds$ die eindeutige Lösung des Anfangswertproblems $\dot{x} = Ax + b$ mit $x(t_0) = x_0$*

(ii) *Lineare Differenzialgleichungen höherer Ordnung.* *Es seien $a_0, \dots, a_{n-1} \in \mathbb{C}$ und $y(t)$ seien normal differenzierbar auf \mathbb{R} . Dann gilt $L(y) = y^{(n)} + \sum_{\nu=0}^{n-1} a_\nu y^{(\nu)} = 0 \Leftrightarrow$*

$$\dot{x} = Ax \text{ für } x = (y, \dot{y}, \dots, y^{(n-1)})^T \text{ mit der Frobenius-Matrix } A = \begin{pmatrix} 0 & 1 & & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ -a_0 & \dots & & & a_{n-1} \end{pmatrix}$$

Zusatz: Vereinfachtes Lösungsverfahren für $L(y) = 0$

- charakteristisches Polynom $P(x) = x^n + \sum_{\nu=0}^{n-1} a_\nu x^\nu = (x - \lambda_1)^{n_1} \dots (x - \lambda_k)^{n_k}$ [Faktorisierung]
- Die Lösung von $L(y) = 0$ sind genau alle Linearkombinationen der n -Faktoren $t^\mu e^{\lambda_\nu t} = y_{\mu\nu}(t)$ für $0 \leq \mu \leq n_\nu - 1, \nu = 1, \dots, k$

Beispiel: (i) $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ mit $e^{tA} = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix}$, $\dot{x} = Ax \Leftrightarrow x(t) = e^{tA}c \Rightarrow c_1 \begin{pmatrix} \cos(t) \\ -\sin(t) \end{pmatrix} + c_2 \begin{pmatrix} \sin(t) \\ \cos(t) \end{pmatrix}$

(ii) $y''' - 2y'' - y' + 2y = 0, P(x) = x^3 - 2x^2 - x + 2 = (x - 1)(x^2 - x - 2) = (x - 1)(x + 1)(x - 2)$

Alle Lösungen $c_1e^{-t} + c_2e^t + c_3e^{2t}$

(iii) $y^{(4)} - 2y'' + y = 0 \Rightarrow P(x) = (x^2 - 1)^2 = (x - 1)^2(x + 1)^2$

Alle Lösungen $y(t) = c_1e^{-t} + c_2te^{-t} + c_3e^t + c_4te^t$

10 Lineares Programmieren

10.1 Beispiele linearer Optimierungsprobleme und graphischer Lösungen

Beispiel (Nr.1, Textilfabrik): Eine Textilfabrik stellt 2 Produkte her und benötigt 2 Stunden für eine Gro-Einheit (GE) und die Maschinen laufen 24 h/Tag. Verarbeitet werden die Rohstoffe Wolle und Synthetik mit:

- Produkt 1: 80% Wolle, 20% Synthetik und 2000 € / Groß-Einheit
- Produkt 2: 20% Wolle, 80% Synthetik und 1000 € / Groß-Einheit

Die Maschinen können $24/5$ Einheiten Wolle und $36/5$ Einheiten Synthetik am Tag verarbeiten \Rightarrow Maximiere den Gewinn

Das führt zu folgendem linearen Optimierungsproblem [x_1, x_2 Anzahl produzierter GE/Tag]

Maximiere $G := 2000x_1 + 1000x_2$ unter den Nebenbedingungen

$$2x_1 + 2x_2 \leq 24$$

$$0,8x_1 + 0,2x_2 \leq \frac{23}{5}$$

$$0,2x_1 + 0,8x_2 \leq \frac{36}{5}$$

$$\text{und } x_1 \geq 0, x_2 \geq 0$$

Beispiel (2., Produktionsproblem): Produkte 1 und 2, Produktfaktoren A, B, C

	1	2	Kapazität n Stück
A	2	10	60
B	6	6	60
C	15	5	85
Gewinn	45	30	

x_1 produzierte Stückzahlen 1 x_2 produzierte Stückzahlen 2

Führt zu folgendem Optimierungsproblem:

Maximiere $G = 45x_1 + 30x_2$ unter

$$2x_1 + 10x_2 \leq 60$$

$$6x_1 + 6x_2 \leq 60$$

$$15x_1 + 5x_2 \leq 85$$

$$\text{und } x_1, x_2 \geq 0$$

Graphische Lösung [Platzhalter]

10.2 Der Begriff des linearen Optimierungsproblems

Definition 10.2.1. Es seien $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. Dann heißt

$$(LP) \begin{cases} \text{Maximiere } G = G(x) = c^T x = \sum_{\nu=1}^n x_\nu c_\nu \text{ unter den Nebenbedingungen} \\ Ax \leq b, x \geq 0, x \in \mathbb{R}^n \end{cases}$$

ein lineares Optimierungsproblem in Standardform (kurz: ein lineares Programm) (oder ein Standard-Maximum-Problem). Man nennt $G(x) = c^T x$ die sogenannte Zielfunktion und die Menge der zulässigen Punkte $Z := \{x \in \mathbb{R}^n | Ax \leq b, x \geq 0\}$, den zugelassenen Bereich.

Ein zulässiger Punkt $x^* \in Z$ mit $c^T x^* \geq c^T x \quad \forall x \in Z$ heißt optimale Lösung von (LP)

Bemerkung: (i) Im allgemeinen ist $n \leq m$

(ii) Ein Standard-Minimum-Problem ist von der Form [Kostenminimierung]:

$$\text{Minimiere } K(x) = c^T x \text{ unter } Ax \geq b, x \geq 0$$

(iii) Transformation auf Standardform und Schlupfvariablen [slack variables]

- Transformation von Standard-Maximum auf Standard-Minimum-Problem

$$c^T x \rightarrow \max \text{ unter } Ax \leq b, x \geq 0 \Leftrightarrow \tilde{c}^T \rightarrow \min \text{ unter } \tilde{A}x \geq \tilde{b}, x \geq 0 \text{ mit } \tilde{c} = -c, \tilde{A} = -A, \tilde{b} = -b$$

- Keine Vorzeichenrestriktion [oder nur einige]: $x \in \mathbb{R}^n, x = x^+ - x^-$ mit $x^+, x^- \geq 0$

$$c^T x \rightarrow \max \text{ unter } Ax \leq b \Leftrightarrow \tilde{c}^T \tilde{x} \rightarrow \max \text{ unter } \tilde{A}\tilde{x} \leq \tilde{b}, \tilde{x} \geq 0 \text{ mit } \tilde{x} = (x^+ x^-)^T \in \mathbb{R}^{2n}, \tilde{A} = (A | -A), \tilde{c} = (c, -c)^T$$

- Gleichungen als Ungleichungen: $Ax = b \Leftrightarrow \tilde{A}x \leq \tilde{b}$ für $\tilde{A} = \begin{pmatrix} A \\ -A \end{pmatrix}, \tilde{b} = (b, -b)^T \in \mathbb{R}^{2n}$

- Standardform nach Einführung von Schlupfvariablen:

$$c^T x \rightarrow \max \text{ unter } Ax \leq b, x \geq 0 \Leftrightarrow c^T x \rightarrow \max \text{ unter } Ax + y = b, x \geq 0 \text{ und } y \geq 0$$

Lemma. Das Standard-Maximum-Problem $c^T x \rightarrow \max$ unter $Ax \leq b$ und $x \geq 0$ ist äquivalent mit dem linearen Programm $Ax + y = b$ und $x \geq 0, y \geq 0, z = \begin{pmatrix} x \\ y \end{pmatrix}$

Definition 10.2.2. Es seien $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. Das lineare Programm

Maximiere $G(x) = c^T x \rightarrow \max$ unter $Ax + y = b$ und $z = \begin{pmatrix} x \\ y \end{pmatrix} \geq 0$ heißt Standardform mit Schlupfvariable y

Bemerkung: Trichotomie: Bei einem linearen Programm gemäß Definition 2.2. mit zulässigem Bereich $Z = \left\{ z = \begin{pmatrix} x \\ y \end{pmatrix} | Ax + b, z \geq 0 \right\}$ können die folgenden 3 Fälle auftreten:

- $Z = \emptyset$, d.h. es gibt keine zulässigen Punkte

- $Z \neq \emptyset$ und $\sup\{c^T x \mid x \in Z\} = \infty$, d.h. es gibt keine optimale Lösung (insbesondere ist dann Z unbeschränkt)
- Es gibt eine optimale Lösung

Bemerkung: Wir formulieren zunächst weitgehend Fakten ohne die zugrunde liegende Theorie mit Beweisen. Dies erfolgt nach Beschreibung des Simplex-Verfahrens [George Danzig, 1947]

10.3 Das Simplex-Verfahren

Wir betrachten durchweg das Standardproblem. Gegeben $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$

(LP) Maximiere $G(x) := c^T x$ unter $Ax \leq b$ und $x \geq 0$ mit zulässigem Bereich $Z := \{z = (x, y)^T \in \mathbb{R}^{m+n} \mid Ax + y = b, z \geq 0\}$

Definition 10.3.1. Ein Punkt $z \in Z$ heißt eine Ecke (oder Extrempunkt) von Z , falls z keine echte Konvexkombination 2er verschiedener Punkte von Z ist, d.h. z besitzt keine Darstellung der folgenden Form.

$$z = \alpha z_1 + (1 - \alpha) z_2 \text{ mit } 0 < \alpha < 1, z_1, z_2 \in Z, z_1 \neq z_2$$

Man spricht statt von Ecken auch von zulässigen Basislösungen von (LP)

Bemerkung: Wir zeigen später: $z = \begin{pmatrix} x \\ y \end{pmatrix} \in Z$ ist genau dann eine Ecke von Z , falls mindestens n Koordinaten von Z gleich 0 sind und die Spalten von $(A \mid I) \in \mathbb{R}^{m+n}$ zu positiven Koordinaten von z linear unabhängig. Man nennt z entartet, wenn mehr als n Koordinaten gleich Null sind.

Mit dieser Bemerkung folgt unmittelbar die

Proposition 10.3.1. Es sei $b \geq 0$. Dann ist $z = (0 \ b)^T \in Z$ eine Ecke von Z , die sogenannte Ausgangsecke des Simplexverfahrens.

Satz 10.3.1. Falls Z eine Ecke besitzt und falls (LP) eine Lösung besitzt, dann ist stets eine Ecke von Z eine optimale Lösung

10.3.1 Herleitung

Grundidee des Simplexverfahrens: Ausgehend von der Ausgangsecke bestimmt man rekursiv „benachbarte Ecken“ durch Austausch einer Basisvariablen $z_\nu > 0$ durch eine Nichtbasisvariable $z_\mu = 0$ so daß die Zielfunktion G jeweils größer wird.

Herleitung des Austauschschritts [Pivot-Schritt] mit Ausgangsecke $z = \begin{pmatrix} 0 \\ b \end{pmatrix}$, d.h. $x = 0, y = b \geq 0$

		x_1	\dots	x_q	\dots	x_n	
Simplextableau	y_1	a_{11}	\dots		\dots	a_{1n}	b_1
	\vdots	\vdots		a_{iq}		\vdots	\vdots
	y_m	a_{m1}	\dots		\dots	a_{mn}	b_m
		c_1	\dots		\dots	c_n	z

	\tilde{x}_1	\dots	\tilde{x}_q	\dots	\tilde{x}_n	
\tilde{y}_1	\tilde{a}_{11}	\dots		\dots	\tilde{a}_{1n}	\tilde{b}_1
\vdots	\vdots		\tilde{a}_{iq}		\vdots	\vdots
\tilde{y}_m	\tilde{a}_{m1}	\dots		\dots	\tilde{a}_{mn}	\tilde{b}_m
	\tilde{c}_1	\dots		\dots	\tilde{c}_n	\tilde{z}

Fallunterscheidung:

(i) $c_\nu \leq 0 \forall \nu = 1, \dots, n \Rightarrow G(x) \leq \sum_{\nu=1}^n c_\nu x_\nu \leq 0 = G(0) \forall x \geq 0 \Rightarrow x = 0$ ist optimale Lösung

(ii) $c_q > 0$ für ein $q \in \{1, \dots, n\}$

Für $\tilde{x}_q = x_q + \varepsilon, \varepsilon \geq 0, \tilde{x}_\nu = x_\nu \forall \nu \neq q$ folgt $\tilde{G} = c^T \tilde{x} = G + \varepsilon c_q \geq G$, also $\varepsilon \geq 0$ (möglichst gro)

Der neue Punkt \tilde{x} muß zulässig sein d.h. $\tilde{x} \geq 0$ ist erfüllt, falls $\varepsilon \geq 0$ und $b_j \geq (A\tilde{x})_j = \sum_{\mu \neq q} a_{j\mu} x_\mu + a_{jq} \varepsilon = a_{jq} \varepsilon$, d.h. $b_j \geq a_{jq} \varepsilon \forall j = 1, \dots, n$

a) $a_{jq} \leq 0 \forall j = 1, \dots, n \Rightarrow \tilde{x} = \tilde{x}(\varepsilon)$ ist erlaubt $\forall \varepsilon \geq 0$ mit $\tilde{G} = G + \varepsilon c_q \rightarrow \infty$

b) $a_{jq} > 0$ für ein $j \in \{1, \dots, n\}$ man erhält das größte ε durch $\varepsilon := \min\{\frac{b_j}{a_{jq}} \mid a_{jq} > 0\}$

Es folgt $\tilde{G} = G + \frac{b_j}{a_{jq}} c_q \geq G, b_i = \varepsilon a_{iq} = 0$ [Schlupf = 0]

Man nennt a_{jq} das Pivotelement mit Pivotzeile j und Spalte q . Es ergibt sich folgende ...:

$\tilde{x}_\nu u = x_\nu \forall \nu \neq q, \tilde{y}_i = y_i \forall j \neq i, \tilde{x}_q = y_i, \tilde{y}_i = x_q$. Forderung $\tilde{A}\tilde{x} + \tilde{y} = \tilde{b}$

$b_i = \sum_{\mu=1}^n a_{i\mu} x_\mu + y_i = \sum_{\mu \neq q} a_{i\mu} \tilde{x}_\mu + a_{iq} \tilde{y}_i + \tilde{x}_q \Leftrightarrow \tilde{b}_i = \sum \tilde{a}_{i\mu} x_\mu + y_i$

mit Pivotzeile $\tilde{a}_{iq} = \frac{1}{a_{iq}}, \tilde{a}_{i\nu} = \frac{a_{i\nu}}{a_{iq}}, \tilde{b}_i = \frac{b_i}{a_{iq}}$

$(j \neq i) : b_j = \sum_{\nu=1}^n a_{j\nu} x_\nu + y_j = \sum_{\nu \neq q} a_{j\nu} \tilde{x}_\nu + \tilde{y}_j + a_{jq} x_q \quad x_q = \frac{1}{a_{iq}} [b_i - y_i - \sum_{\nu \neq q} a_{i\nu}]$

$= \sum_{\nu \neq q} a_{j\nu} \tilde{x}_\nu + \tilde{y}_j + [b_i - y_i - \sum_{\nu \neq q} a_{i\nu}] \frac{a_{jq}}{a_{iq}}$

$\Leftrightarrow \tilde{b}_j = \sum_{\nu=1}^n \tilde{a}_{j\nu} \tilde{x}_\nu + y_\nu$

$\Rightarrow b_j - \frac{b_i a_{jq}}{a_{iq}}, \tilde{a}_{j\nu} = a_{j\nu} - \frac{a_{i\nu} a_{jq}}{a_{iq}} \forall j \neq i, \nu \neq q$

und Pivotspalte $= \tilde{a}_{jq} = \frac{a_{jq}}{a_{iq}}$

(iii) entsprechend 2. setzen wir $\tilde{c}_\nu = c_\nu - \frac{c_q a_{i\nu}}{a_{iq}} \forall \nu = 1, \dots, n \neq q$

$\tilde{c}_q = -\frac{c_q}{a_{iq}}$ sowie $\tilde{z} = z - \frac{c_q b_i}{a_{iq}}$ (gemäß obiger Rechnung)

10.3.2 Die Zweiphasenmethode für beliebige, lineare Programme

1. Phase: Berechnung einer Ausgangsecke
2. Phase: Simplexverfahren gemäß 10.3.2 ab 2. (Demo an einem Beispiel (Folie))

10.4 Konvexität

Definition 10.4.1. Eine Menge $M \subset \mathbb{R}^n$ heißt konvex, falls gilt
 $\overline{z_1 z_2} = \{\alpha z_1 + (1 - \alpha)z_2 \mid \alpha \in [0, 1]\}$

Beispiele: Es sei $B = (b_1, \dots, b_n) \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$

- (i) $M := \{x \in \mathbb{R}^n \mid Bx = b \text{ und } x \geq 0\} \subset \mathbb{R}^n$ ist konvex, weil für $z_1, z_2 \in M, \alpha \in [0, 1]$ gilt $\alpha z_1 + (1 - \alpha)z_2 \geq 0$ und $B(\alpha z_1 + (1 - \alpha)z_2) = \alpha Bz_1 + (1 - \alpha)Bz_2 = \alpha b + (1 - \alpha)b = b$

- (ii) [Platzhalter]

Proposition 10.4.1. Eine Menge $M \subset \mathbb{R}^n$ ist genau dann konvex, falls gilt $M = \text{conv}(M)$, wobei $\text{conv}(M)$, die sogenannten konvexe Hülle von M , definiert ist durch $\text{conv}(M) = \{\text{Konvexkombinationen von Punkten aus } M\} := \{\sum_{\nu=1}^k \alpha_\nu z_\nu \mid k \in \mathbb{N}, \alpha_\nu \geq 0, \sum_{\nu=1}^k \alpha_\nu = 1\}$

Beweis. Es gilt $M \subset \text{conv}(M)$ [für $k = 0$]

- (i) $\text{conv}(M) = M \Rightarrow M$ ist konvex [für $k = 0$]

- (ii) Sei M konvex \Rightarrow z.Z. $\text{conv}(M) \subset M$

$k = 1$: klar

$k \rightarrow k + 1$ Es seien $z_1, \dots, z_{k+1} \in M, \alpha_\nu > 0, \sum_{\nu=1}^n \alpha_\nu = 1$, o.B.d.A. $\alpha_{n+1} < 1$

$\sum_{\nu=1}^k \frac{\alpha_\nu}{1 - \alpha_{k+1}} z_\nu \in M$ nach Induktions-Hypothese $\Rightarrow \sum_{\nu=1}^{k+1} z_\nu = (1 - \alpha_{k+1}) \sum_{\nu=1}^k \frac{\alpha_\nu}{1 - \alpha_{k+1}} z_\nu + \alpha_{k+1} z_{k+1} \in M$ (Definition 4.1, M konvex).

□

Bemerkung: \emptyset ist konvex und $\text{conv } \emptyset = \emptyset$

Definition 10.4.2. Eine Funktion $f : M \rightarrow \mathbb{R}$ heißt konvex auf M , falls $M \subset \mathbb{R}^n$ konvex ist und falls $f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2) \quad \forall x_1, x_2 \in M$

Satz 10.4.1 (Globale Minima). Es sei $f : M \rightarrow \mathbb{R}^n$ konvex auf M und $z \in M$ sei ein lokales Minimum auf $f(M)$. Dann ist z globales Minimum auf M

Beweis. zZ $f(z) \leq f(x) \quad \forall x \in M$. Dazu sei $x \in M$

$f(\alpha x + (1 - \alpha)z) \leq \alpha f(x) + (1 - \alpha)f(z)$ und $\exists 0 < \varepsilon < 1$ mit

$f(\alpha x + (1 - \alpha)z) = f(z + \alpha(x - z)) \geq f(z) \quad \forall 0 < \alpha < \varepsilon$

$\Rightarrow f(z) = \alpha f(x) + (1 - \alpha)f(z) \Leftrightarrow \alpha f(z) \leq \alpha f(x) \Leftrightarrow f(z) \leq f(x)$

□

Definition 10.4.3. Es sei $M \subset \mathbb{R}^n$. Ein Punkt $z \in M$ heißt eine Ecke (oder Extrempunkt) von M , falls z keine echte Konvexkombination 2er verschiedener Punkte von M ist, d.h. z besitzt keine Darstellung der folgenden Form.

Satz 10.4.2 (Extrema in Ecken). Es sei $M \subset \mathbb{R}^n$ nicht leer, kompakt und konvex, sowie $c \in \mathbb{R}^n, f(x) = c^T x$ für $x \in M$. Dann gibt es eine Ecke $z \in M$ von M mit $f(z) = \min_{x \in M} f(x)$

Beweis. $M \neq \emptyset$ und kompakt und f ist stetig auf $M \Rightarrow$ [Analysis: Weierstraß] $\exists z \in M$ mit $f(z) = \min_{x \in M} f(x)$ und $\|z\|_2 = \max\{\|x\|_2 \mid x \in M \text{ und } f(x) = \min_{x \in M} f(x)\}$.

Wir zeigen: z ist eine Ecke von M . Dazu seien $z_1, z_2 \in M, 0 < \alpha < 1$ mit $z = \alpha z_1 + (1 - \alpha)z_2$.

Zu zeigen: $z_1 = z_2 (= z)$. Nun gilt

$$f(z) = f(\alpha z_1 + (1 - \alpha)z_2) = \alpha f(z_1) + (1 - \alpha)f(z_2) \geq \min\{f(z_1), f(z_2)\} \geq f(z) \Rightarrow f(z_1) = f(z_2) = f(z)$$

$$f(\beta z_1 + (1 - \beta)z_2) = \beta f(z_1) + (1 - \beta)f(z_2) = f(z) \forall \beta \in [0, 1], z(\beta) = \beta z_1 + (1 - \beta)z_2 = z_2 + \beta(z_1 - z_2)$$

Nach der Konstruktion von z folgt $\|z\|_2^2 = \|z(\alpha)\|_2^2 \geq \|z(\beta)\|_2^2 = \|z_2\|_2^2 + \beta^2\|z_1 - z_2\|_2^2 + 2\beta z_2^T(z_1 - z_2) \forall \beta \in [0, 1]$

$$\alpha \in (0, 1) \Rightarrow (\text{Notw. Bed. f. Max}) 0 \geq \left(\frac{d}{d\beta}\right)^2 \|z(\beta)\|_2^2 \Big|_{\beta=\alpha} = 2\|z_1 - z_2\|^2 \Rightarrow z_1 = z_2 \quad \square$$

Korollar 10.4.1. Jede nicht-leere kompakte und konvexe Teilmenge des \mathbb{R}^n besitzt mindestens eine Ecke

Satz 10.4.3 (Trennungssatz). Es sei $M \subset \mathbb{R}^n$ nicht leer, konvex und abgeschlossen sowie $b \in \mathbb{R}^n$ mit $b \notin M$. Dann gibt es $y \in \mathbb{R}^n \setminus \{0\}$ und $\gamma > 0$ mit $y^T x \geq \gamma + y^T b > y^T b$ für alle $x \in M$ [„Es gibt eine trennende Hyperebene“]

Beweis. Es sei zunächst $b = 0 \notin M$. Weiter sei $r > 0$ mit $K := \{x \in \mathbb{R}^n \mid \|x\| \leq r\} \Rightarrow K \neq \emptyset$, kompakt und konvex \Rightarrow (Weierstr.) $\exists y \in K \cap M \neq \emptyset$ mit $\|y\|_2 = \min$ \square