

In diesem Kapitel wird eine Grundlage für einheitliche und verständliche Notation geschaffen. Zudem werden einige Sätze und Definitionen eingeführt, die im weiteren Verlauf dieser Arbeit verwendet werden.

Die Elemente einer Matrix A werden mit a_{ij} bezeichnet. Dabei gibt i die Zeilen- und j die Spaltenposition des Elements an. Sei $A \in \mathbb{K}^{n \times m}$, dann hat A die Form

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}. \quad (0.1)$$

Weiter bezeichnen wir A^\top als Transponierte und A^H als Adjungierte von A . Matrizen sind orthogonal bezüglich des standard Skalarproduktes. Das komplex konjugierte von a bezeichnen wir als \bar{a} .

Definition 0.1 (Frobeniusnorm). Sei $X \in \mathbb{C}^{n \times m}$ eine Matrix und x_{ij} mit $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$. Die *Frobeniusnorm* ist definiert durch

$$\|X\|_F := \sqrt{\sum_{i=1}^n \sum_{j=1}^m |x_{ij}|^2}.$$

Lemma 0.2. Sei $X \in \mathbb{C}^{n \times m}$ eine Matrix. Dann gilt

$$\|X\|_F = \sqrt{\text{tr}(X^H X)}.$$

Beweis. Sei x_i ein Spaltenvektor von X . Dann ist

$$\begin{aligned} X^H X &= \begin{pmatrix} - & x_1^H & - \\ \vdots & & \vdots \\ - & x_m^H & - \end{pmatrix} \begin{pmatrix} | & & | \\ x_1 & \dots & x_m \\ | & & | \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^n x_{1j}^H x_{1j} & * & \\ \ddots & \ddots & \\ * & \sum_{j=1}^n x_{mj}^H x_{mj} & \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n \bar{x}_{1j} x_{1j} & * & \\ \ddots & \ddots & \\ * & \sum_{j=1}^n \bar{x}_{mj} x_{mj} & \end{pmatrix}. \end{aligned}$$

Mit $\bar{x}x = |x|^2$ erhalten wir

$$\mathrm{tr}(X^H X) = \sum_{i=1}^m \sum_{j=1}^n \bar{x}_{ij} x_{ij} = \sum_{i=1}^m \sum_{j=1}^n |x_{ij}|^2 = \|X\|_F^2. \quad (0.2)$$

□

Die Frobeniusnorm wird im weiteren Verlauf aufgrund ihrer Invarianz bezüglich unitärer Transformation verwendet, welche wir durch folgendes Lemma zeigen.

Lemma 0.3. Sei $X \in \mathbb{C}^{n \times m}$ eine Matrix und $U \in \mathbb{C}^{n \times n}$ und $V \in \mathbb{C}^{m \times m}$ unitäre Matrizen. Dann gilt

1. $\|X\|_F = \|X^H\|_F,$
2. $\|UX\|_F = \|X\|_F,$
3. $\|XV\|_F = \|X\|_F.$

Beweis.

1. Aus Gleichung (0.2) folgt

$$\|X\|_F^2 = \mathrm{tr}(X^H X) = \sum_{i=1}^m \sum_{j=1}^n \bar{x}_{ij} x_{ij} = \sum_{j=1}^n \sum_{i=1}^m \bar{x}_{ij} x_{ij} = \mathrm{tr}(XX^H) = \|X^H\|_F^2.$$

2. Es ist

$$\|UX\|_F^2 = \mathrm{tr}(X^H U^H UX) = \mathrm{tr}(X^H \mathbf{1} X) = \mathrm{tr}(X^H X) = \|X\|_F^2.$$

3. Mit 1. und 2. folgt

$$\|XV\|_F = \|V^H X^H\|_F = \|X^H\|_F = \|X\|_F. \quad \square$$

Satz 0.4 (Singulärwertzerlegung [1]). Sei $X \in \mathbb{R}^{n \times m}$ eine Matrix und $r := \mathrm{rang} X$ mit $r \leq \min(n, m)$. Dann existieren orthogonale Matrizen $U \in \mathbb{R}^{n \times n}$ und $V \in \mathbb{R}^{m \times m}$ und eine Diagonalmatrix $\hat{\Sigma} = \mathrm{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$ mit $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ so, dass

$$X = U \Sigma V^\top \text{ mit } \Sigma = \begin{pmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n \times m}$$

gilt.

Der Beweis des Satzes 0.4 folgt etwas später, da dafür noch etwas Vorbereitung benötigt wird.

Bemerkung 0.5. Unter den in Satz 0.4 genannten Voraussetzungen, ist $\hat{\Sigma}$ durch X eindeutig bestimmt. Die beiden orthogonalen Matrizen U und V hingegen sind nicht eindeutig.

Lemma 0.6. Sei $X \in \mathbb{R}^{n \times m}$ und λ_i die Eigenwerte von $X^\top X$. Dann gilt für die Singulärwerte unter den Voraussetzungen aus Satz 0.4

$$\sigma_i = \sqrt{\lambda_i}.$$

Beweis. Durch

$$\begin{aligned} X^\top X &= (U\Sigma V^\top)^\top (U\Sigma V^\top) \\ &= V\Sigma^\top U^\top U\Sigma V^\top \\ &= V\Sigma^\top \Sigma V^\top \end{aligned}$$

ist $X^\top X$ ähnlich zu $\Sigma^\top \Sigma$. Also haben $X^\top X$ und $\Sigma^\top \Sigma$ auch die gleichen Eigenwerte. \square

Bemerkung 0.7. Damit gilt für die Diagonalelemente von $\hat{\Sigma}$ aus Satz 0.4,

$$\sigma_i = \sqrt{\lambda_i},$$

dass deren Wurzeln gleich der Eigenwerte von $X^\top X$ sind.

Definition 0.8. Wir nennen die Diagonalelemente σ_i von $\hat{\Sigma}$ *Singulärwerte*.

Bemerkung 0.9. Die aus der Singulärwertzerlegung in 0.4 entstehenden Singulärwerte sind invariant bezüglich Transposition vom X .

Beweis. Seien die Voraussetzungen aus Satz 0.4 gegeben, dann gilt

$$X^\top = (U\Sigma V^\top)^\top = V\Sigma^\top U^\top.$$

Da die Singulärwerte in Σ auf der Diagonalen liegen, erhalten wir

$$X^\top = V\Sigma U^\top.$$

\square

Lemma 0.10 ([1, S. 191]). Sei $X \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix. Dann existiert eine orthogonale Matrix $U \in \mathbb{R}^{n \times n}$ und eine Diagonalmatrix $D \in \mathbb{R}^{n \times n}$ so, dass

$$U^\top X U = D$$

gilt.

Lemma 0.11 ([1, S. 227]). Sei $X \in \mathbb{R}^{n \times m}$ eine Matrix mit $\text{rang } X = r < \min(n, m)$. Dann existieren orthogonale Matrizen $U \in \mathbb{R}^{n \times n}$ und $V \in \mathbb{R}^{m \times m}$ sowie eine reguläre obere Dreiecksmatrix $\hat{R} \in \mathbb{R}^{r \times r}$ so, dass

$$U^\top X V = R \text{ mit } R = \begin{pmatrix} \hat{R} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n \times m}$$

gilt.

Beweis Satz 0.4. Nach Bemerkung 0.9 nehmen wir o.B.d.A an, dass $m \leq n$ gilt. Zunächst betrachten wir den Fall des vollen Ranges von X , also $\text{rang } X = r = m$. Bei vollem Rang ist $X^\top X \in \mathbb{R}^{m \times m}$ sowohl symmetrisch als auch positiv definit. Die Eigenwerte λ_i von $X^\top X$ sortieren wir nach ihrer Größe, also

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0.$$

Mit Lemma 0.10 finden wir ein orthogonales $V \in \mathbb{R}^{m \times m}$, sodass

$$V^\top X^\top X V = D = \text{diag}\{\lambda_1, \dots, \lambda_m\} \in \mathbb{R}^{m \times m}.$$

Weiter definieren wir

$$\hat{\Sigma} := \text{diag}\{\sigma_1, \dots, \sigma_m\}$$

mit $\sigma_i := \sqrt{\lambda_i} \in (0, \infty]$, für $i = 1, \dots, m$, und setzen

$$\hat{U} := X V \hat{\Sigma}^{-1} \in \mathbb{R}^{n \times m}.$$

Jetzt kann man durch Umformung von

$$\begin{aligned}
\hat{U}^\top \hat{U} &= \hat{\Sigma}^{-\top} V^\top X^\top X V \hat{\Sigma}^{-1} \\
&= \hat{\Sigma}^{-1} D \hat{\Sigma}^{-1} \\
&= \hat{\Sigma}^{-1} \hat{\Sigma}^{-1} D \\
&= D^{-1} D \\
&= \mathbb{1}
\end{aligned}$$

die Orthonormalität der Spalten von \hat{U} sehen.

Da \hat{U} in $\mathbb{R}^{n \times m}$ mit $m < n$ liegt und aus orthonormalen Spaltenvektoren besteht, finden wir ein $Z \in \mathbb{R}^{n \times (n-m)}$, welches \hat{U} so um $n - m$ orthonormale Spaltenvektoren ergänzt, dass

$$U = (\hat{U} \ Z) \in \mathbb{R}^{n \times n}$$

eine orthogonale Matrix ist. Es gilt also

$$\begin{aligned}
Z^\top X V \hat{\Sigma}^{-1} &= Z^\top \hat{U} \\
&= 0 \in \mathbb{R}^{(n-m) \times n}.
\end{aligned}$$

Da $\hat{\Sigma} \hat{\Sigma}^{-1} = \mathbb{1}$ ist, folgt auch direkt

$$Z^\top X V = 0.$$

Zusammengetragen haben wir dann

$$\begin{aligned}
U^\top X V &= \begin{pmatrix} \hat{U}^\top \\ Z^\top \end{pmatrix} X V = \begin{pmatrix} \hat{\Sigma}^{-1} V^\top X^\top X V \\ Z^\top X V \end{pmatrix} \\
&= \begin{pmatrix} \hat{\Sigma}^{-1} D \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \hat{\Sigma} \\ 0 \end{pmatrix},
\end{aligned}$$

womit wir den Fall des vollen Ranges bewiesen hätten.

Seien λ_i wie oben definiert wieder die Eigenwerte von $X^\top X$ und sei $X_r \in \mathbb{R}^{n \times m}$ eine Matrix mit $\text{rang } X_r = r < m$. Nach Lemma 0.11 gibt es zwei orthogonale Matrizen

$\tilde{U} \in \mathbb{R}^{n \times n}$ und $\tilde{V} \in \mathbb{R}^{m \times m}$, sowie eine obere Dreiecksmatrix $\hat{R} \in \mathbb{R}^{r \times r}$, sodass

$$\tilde{U}^\top X_r \tilde{V} = R \text{ mit } R = \begin{pmatrix} \hat{R} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n \times m}$$

gilt. Nach obigem Fall mit vollem Rang, wissen wir, dass es orthogonale Matrizen $U_r \in \mathbb{R}^{n \times n}$ und $V_r \in \mathbb{R}^{r \times r}$, sodass

$$U_r^\top \begin{pmatrix} \hat{R} \\ 0 \end{pmatrix} V_r = \begin{pmatrix} \hat{\Sigma}_r \\ 0 \end{pmatrix} \in \mathbb{R}^{n \times r}.$$

Dabei ist $\hat{\Sigma}_r := \text{diag}\{\sigma_1, \dots, \sigma_r\}$ mit $\sigma_i = \sqrt{\lambda_i}$. Im Folgenden erweitern wir

$$\begin{pmatrix} \hat{\Sigma}_r \\ 0 \end{pmatrix} \in \mathbb{R}^{n \times r} \quad \text{zu} \quad \begin{pmatrix} \hat{\Sigma}_r & 0 \\ 0 & 0 \end{pmatrix} = \Sigma \in \mathbb{R}^{n \times m}$$

und finden eine orthogonale Fortsetzung für V_r zu

$$\bar{V}_r = \begin{pmatrix} V_r & 0 \\ 0 & \mathbb{1} \end{pmatrix} \in \mathbb{R}^{m \times m}.$$

Wenn wir uns jetzt die orthogonalen Matrizen

$$U := \tilde{U} U_r \in \mathbb{R}^{n \times n} \quad \text{und} \quad V := \tilde{V} \bar{V}_r \in \mathbb{R}^{m \times m}$$

definieren, erhalten wir durch

$$\begin{aligned} U^\top X_r V &= U_r^\top \tilde{U} X_r \tilde{V} \bar{V}_r \\ &= U_r^\top \begin{pmatrix} \hat{R} & 0 \\ 0 & 0 \end{pmatrix} \bar{V}_r \\ &= \begin{pmatrix} \hat{\Sigma}_r & 0 \\ 0 & 0 \end{pmatrix} =: \Sigma \end{aligned}$$

eine erfolgreiche Zerlegung von X_r durch orthogonale Matrizen. \square

Satz 0.12 (Moore-Penrose-Inverse [1]). Sei $X \in \mathbb{R}^{n \times m}$ eine Matrix und $U, \hat{\Sigma}$ und V die Matrizen aus der Singulärwertzerlegung in Satz 0.4. Dann existiert genau eine Moore-

Penrose-Inverse der Form

$$X^\dagger = V \begin{pmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^\top \in \mathbb{R}^{m \times n}.$$

Definition 0.13 ($\arg \min$). Sei $f: X \rightarrow \mathbb{R}$ eine Funktion. Dann ist die Menge $\arg \min$ definiert durch

$$\arg \min_{x \in X} f(x) := \{x \in X \mid f(x) \text{ ist minimal}\}.$$