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

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

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 $\overline{a}$.

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

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

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

$$||X||_F = \sqrt{\operatorname{tr}\left(X^H X\right)}\,.$$

Beweis. Sei x_i ein Spaltenvektor von X. Dann ist

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

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

$$\operatorname{tr}\left(X^H X\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \overline{x_{ij}} x_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} |x_{ij}|^2 = ||X||_F^2 \,. \tag{0.2}$$

$\square$

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

$$||X||_F^2 = \operatorname{tr}\left(X^H X\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \overline{x_{ij}} x_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} \overline{x_{ij}} x_{ij} = \operatorname{tr}\left(X X^H\right) = ||X^H||_F^2 \,.$$

2. Es ist

$$||UX||_F^2 = \operatorname{tr}\left(X^H U^H U X\right) = \operatorname{tr}\left(X^H \mathbb{1} X\right) = \operatorname{tr}\left(X^H X\right) = ||X||_F^2 \,.$$

3. Mit 1. und 2. folgt

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

Satz 0.4 (Singulärwertzerlegung [1])**.** Sei $X \in \mathbb{R}^{n\times m}$ eine Matrix und $r := \operatorname{rang} X$ 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 $\hat{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r\times r}$ 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

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

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

Definition 0.7. Die Diagonalelemente von $\hat{\Sigma}$ aus Satz 0.4, wobei $\sigma_i = \sqrt{\lambda_i}$ sind die Wurzeln der Eigenwerte von $X^\top X$ und werden als Singulärwerte bezeichnet.

Definition 0.8. Wir bezeichnen bei der Singulärwertzerlegung die Spaltenvektoren u_i mit $i = 1, \dots, n$ von U als Links-Singulärvektoren und die Spaltenvektoren v_i mit $i = 1, \dots, m$ von V als Rechts-Singulärvektoren. Lasse ich erstmal drinnen, für den Fall, dass ich diese Begriffe später noch verwenden werde

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

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

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

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

□

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 $\operatorname{rang} X = r < \min(n, m)$. 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 $\operatorname{rang} X = r = m$. 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

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0\,.$$

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

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

Weiter definieren wir

$$\hat{\Sigma} := \operatorname{diag}\{\sigma_1, \ldots, \sigma_m\}$$

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

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

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

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

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 $\operatorname{rang} X_r = r < m$. Nach Lemma 0.11 gibt es zwei orthognonale 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 $\hat{\Sigma}_r \coloneqq \operatorname{diag}\{\sigma_1, \dots, \sigma_r\}$ 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

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

definieren, erhalten wir durch

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

eine erfolgreiche Zerlegung von X_r durch orthogonale Matrizen. □

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 $f\colon X \to \mathbb{R}$ eine Funktion. Dann ist die Menge arg min definiert durch

$$\operatorname*{arg\,min}_{x \in X} f(x) \coloneqq \{x \in X \mid f(x) \text{ ist minimal}\} .$$