

6.3 Singulärwertzerlegung von rechteckigen Matrizen

(dieses Material ist nicht mehr klausurrelevant)

Häufig hat man es mit rechteckigen anstatt mit quadratischen Matrizen zu tun. Ein Bild zum Beispiel ist typischerweise rechteckig, nicht quadratisch, und, wenn es ein schwarz-weiß Bild ist wie im nächsten Kapitel, ist es gegeben durch eine $m \times n$ Matrix

$$A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}^{m \times n} \quad (6.57)$$

mit Einträgen zwischen 0 und 1, wobei jeder Eintrag dann dem Grauwert eines Pixels an der Stelle (i, j) entspricht. Wir wollen uns in diesem Kapitel überlegen, ob wir auch in einem solchen Fall noch eine Formel analog zu (6.42) hinschreiben können. Die Antwort ist ja. Dabei wollen wir nur den reellen Fall $A \in \mathbb{R}^{m \times n}$ betrachten, also keine komplexen Matrizen.

Wir betrachten zunächst die Matrix

$$M := A^T A : \mathbb{R}^n \rightarrow \mathbb{R}^n \quad (6.58)$$

Das M ist eine quadratische $n \times n$ Matrix und es gilt

$$M^T = (A^T A)^T \stackrel{\substack{\text{Ü-Blatt 6} \\ \text{Aufg. 6}}}{=} A^T [A^T]^T = A^T A = M \quad (6.59)$$

Das M ist also eine reell-symmetrische Matrix und sie ist diagonalisierbar: es gibt eine ONB $\vec{v}_1, \dots, \vec{v}_n$ von Vektoren aus dem $\mathbb{R}^n$ mit

$$M \vec{v}_k = \mu_k \vec{v}_k \quad \text{und} \quad (\vec{v}_j, \vec{v}_k) = \delta_{j,k} \quad (6.60)$$

Wegen

$$\begin{aligned} \mu_k (\vec{v}_k, \vec{v}_k) &= (\mu_k \vec{v}_k, \vec{v}_k) = (M \vec{v}_k, \vec{v}_k) = (A^T A \vec{v}_k, \vec{v}_k) \\ &= (A \vec{v}_k, A \vec{v}_k) = \|A \vec{v}_k\|^2 \geq 0 \end{aligned} \quad (6.61)$$

können die Eigenwerte μ_k nicht negativ sein, wir können also eine reelle Wurzel ziehen:

$$\lambda_k := \sqrt{\mu_k} \in \mathbb{R} \quad (6.62)$$

In dem Fall, dass das A selber schon quadratisch und symmetrisch ist, sind die λ_k gerade die Eigenwerte von A oder genauer, die Beträge davon, die Eigenwerte von einem quadratischen A könnten ja auch negativ sein. Wir ordnen dann die λ_k von gross nach klein, der kleinste mögliche Wert wäre dann 0, nehmen wir an, dass etwa p der n λ_k echt grösser als 0 sind,

$$\lambda_1 > \lambda_2 \geq \dots \geq \lambda_p > 0 \quad (6.63)$$

und die restlichen λ_k sind 0,

$$\lambda_{p+1} = \dots = \lambda_n = 0 \quad (6.64)$$

Da das $M = A^T A$ ja mit dem A zunächst mal von $\mathbb{R}^n$ nach $\mathbb{R}^m$ geht und dann mit dem A^T wieder von $\mathbb{R}^m$ nach $\mathbb{R}^n$ zurück geht, kann man sich überlegen, dass das p nicht nur kleiner oder gleich n , sondern auch kleiner oder gleich m sein muss, also

$$p \leq \min\{m, n\} \quad (6.65)$$

Das Analogon zu der Matrix D aus (6.42) ist dann folgendermaßen zu definieren:

$$D := \text{Diag}(\lambda_1, \dots, \lambda_{\min\{m,n\}}) \in \mathbb{R}^{m \times n} \quad (6.66)$$

Dieses D hat also wieder nur auf der Diagonalen potentiell Einträge ungleich 0, aber das ist jetzt eine rechteckige Matrix mit m Zeilen und n Spalten, also keine quadratische Matrix mehr wie in (6.42). Jetzt müssen wir noch eine geeignete ONB im $\mathbb{R}^m$ festlegen, die uns dann ein Analogon für das linke V in $A = VD V^T$ liefert. Dazu setzen wir

$$\vec{w}_j := \frac{1}{\lambda_j} A \vec{v}_j \in \mathbb{R}^m \quad \text{für } 1 \leq j \leq p \quad (6.67)$$

Man beachte, dass wieder für den Fall, dass das A selber schon quadratisch und symmetrisch ist, die λ_k dann (evtl. bis aufs Vorzeichen) die Eigenwerte von A sind und die obige Formel sich dann auf $\vec{w}_j = \pm \vec{v}_j$ reduziert. Weiterhin sind diese $\vec{w}_j$ alle orthogonal zueinander,

$$\begin{aligned} (\vec{w}_j, \vec{w}_k) &= \frac{1}{\lambda_j} \frac{1}{\lambda_k} (A \vec{v}_j, A \vec{v}_k) = \frac{1}{\lambda_j \lambda_k} (A^T A \vec{v}_j, \vec{v}_k) = \frac{1}{\lambda_j \lambda_k} (M \vec{v}_j, \vec{v}_k) \\ &= \frac{\mu_j}{\lambda_j \lambda_k} (\vec{v}_j, \vec{v}_k) = \frac{\lambda_j^2}{\lambda_j \lambda_k} \delta_{j,k} = \delta_{j,k} \end{aligned} \quad (6.68)$$

Im allgemeinen Fall kann es sein, dass $p < m$ ist und die oben angegebenen $\vec{w}_j$ tun dann noch nicht ausreichen, um den ganzen $\mathbb{R}^m$ auszuspannen. In dem Fall ist es tatsächlich so, checken wir gleich, dass man zu den p angegebenen $\vec{w}_j$ noch $m - p$ beliebige weitere $\vec{w}_j$ hinzunehmen kann, die müssen nur auf den bereits gegebenen p $\vec{w}_j$ senkrecht stehen, das kann man immer erreichen. Also man erweitert die bereits gegebenen $\vec{w}_1, \dots, \vec{w}_p$ zur einer ONB

$$\vec{w}_1, \dots, \vec{w}_p, \vec{w}_{p+1}, \dots, \vec{w}_m \quad (6.69)$$

des $\mathbb{R}^m$ und diese Basisvektoren schreibt man dann als Spaltenvektoren in eine $m \times m$ Matrix W ,

$$W := \begin{pmatrix} | & & | \\ \vec{w}_1 & \cdots & \vec{w}_m \\ | & & | \end{pmatrix} \in \mathbb{R}^{m \times m} \quad (6.70)$$

Dann gilt:

$$A = W D V^T \quad (6.71)$$

Um (6.71) zu verifizieren, reicht es zu zeigen, dass die linke und rechte Seite von (6.71) angewendet auf die Basisvektoren $\vec{v}_k$ des $\mathbb{R}^n$ dasselbe Resultat liefert, checken wir das: Zunächst mal ist

$$V^T \vec{v}_k = \vec{e}_k =: \vec{e}_k^{(n)} \in \mathbb{R}^n \quad (6.72)$$

Dabei haben wir an das $\vec{e}_k$ ein superscript $^{(n)}$ angehängt, um deutlicher zu machen, dass das ein Vektor im $\mathbb{R}^n$ ist. Wenn wir jetzt das D darauf anwenden, kommt da ein Vektor aus dem $\mathbb{R}^m$ raus, und zwar:

$$D \vec{e}_k^{(n)} = \begin{cases} \lambda_k \vec{e}_k^{(m)} & \text{falls } k \leq \min\{m, n\} \\ \vec{0}^{(m)} & \text{falls } k > \min\{m, n\} \end{cases} \quad (6.73)$$

Damit bekommen wir: Für $k \leq \min\{m, n\}$:

$$\begin{aligned}
W D V^T \vec{v}_k &= W D \vec{e}_k^{(n)} = W \lambda_k \vec{e}_k^{(m)} = \lambda_k W \vec{e}_k^{(m)} = \lambda_k \vec{w}_k \\
&= \begin{cases} \lambda_k \frac{1}{\lambda_k} A \vec{v}_k & \text{falls } k \leq p \\ 0 \cdot \vec{w}_k = \vec{0}^{(m)} & \text{falls } k > p \end{cases} \\
&= A \vec{v}_k
\end{aligned} \tag{6.74}$$

da $\|A \vec{v}_k\|^2 = \mu_k \|\vec{v}_k\|^2 = 0$ für $k > p$. Und für $k > \min\{m, n\} \geq p$:

$$W D V^T \vec{v}_k = W D \vec{e}_k^{(n)} = W \vec{0}^{(m)} = \vec{0}^{(m)} = A \vec{v}_k \tag{6.75}$$

also insgesamt $A = W D V^T$. Fassen wir die Ergebnisse in einem Theorem zusammen:

Theorem (Singulärwertzerlegung): Es sei $A \in \mathbb{R}^{m \times n}$ eine beliebige reelle $m \times n$ Matrix. Dann gibt es orthogonale Matrizen

$$W \in O(m), \quad V \in O(n) \tag{6.76}$$

und eine rechteckige Diagonalmatrix

$$D := \text{Diag}(\lambda_1, \dots, \lambda_{\min\{m, n\}}) \in \mathbb{R}^{m \times n} \tag{6.77}$$

mit nicht negativen λ_k , so dass

$$A = W D V^T \tag{6.78}$$

Dabei sind die Spaltenvektoren $\vec{v}_k$ der Matrix V die Eigenvektoren von $A^T A$ mit Eigenwerten λ_k^2 und die Spaltenvektoren der Matrix W sind gegeben durch $\frac{1}{\lambda_k} A \vec{v}_k$ solange die $\lambda_k > 0$ sind. Eine äquivalente Schreibweise von (6.78) ist

$$A = \sum_{k=1}^{\min\{m, n\}} \lambda_k \begin{pmatrix} | \\ \vec{w}_k \\ | \end{pmatrix} \begin{pmatrix} - & \vec{v}_k & - \end{pmatrix} \tag{6.79}$$

Beispiel: Schauen wir uns noch ein Beispiel an. Für $a \neq 0$ sind die Matrizen

$$A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \tag{6.80}$$

nicht diagonalisierbar. Eine Singulärwertzerlegung kann man aber für beliebige Matrizen machen, wie würde das hier aussehen? Für $a = 1$ werden die Zahlen etwas krumm, deshalb wählen wir $a = 3/2$, wir betrachten also die Matrix

$$A = \begin{pmatrix} 1 & 3/2 \\ 0 & 1 \end{pmatrix} \tag{6.81}$$

und möchten die Singulärwertzerlegung von diesem A bestimmen.

Lösung: Die $\vec{v}_k$ und die $\mu_k = \lambda_k^2$ sind die Eigenvektoren und Eigenwerte der Matrix

$$M = A^T A = \begin{pmatrix} 1 & 0 \\ 3/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3/2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3/2 \\ 3/2 & 13/4 \end{pmatrix} \quad (6.82)$$

Das charakteristische Polynom ist

$$\begin{aligned} \chi_M(\mu) &= \det[M - \mu Id] = \det \begin{pmatrix} 1 - \mu & 3/2 \\ 3/2 & 13/4 - \mu \end{pmatrix} = (1 - \mu)(13/4 - \mu) - 9/4 \\ &= \frac{13}{4} - \frac{17}{4}\mu + \mu^2 - \frac{9}{4} = \mu^2 - \frac{17}{4}\mu + 1 \end{aligned} \quad (6.83)$$

Die Nullstellen sind

$$\mu_{\pm} = \frac{17}{8} \pm \sqrt{\frac{289}{64} - 1} = \frac{17}{8} \pm \sqrt{\frac{225}{64}} = \frac{17}{8} \pm \frac{15}{8}$$

also

$$\mu_+ = 4, \quad \mu_- = 1/4 \quad (6.84)$$

Die Eigenvektoren $\vec{v}_{\pm}$ sind gegeben durch

$$\begin{pmatrix} 1 - \mu_{\pm} & 3/2 \\ 3/2 & 13/4 - \mu_{\pm} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (6.85)$$

wobei die zweite Zeile notwendigerweise redundant ist, die Determinante ist ja Null. Wir bekommen also

$$(1 - \mu_{\pm}) v_1 + \frac{3}{2} v_2 = 0 \quad (6.86)$$

oder

$$\text{Für } \mu_+: \quad -3 v_1 + \frac{3}{2} v_2 = 0 \quad (6.87)$$

$$\text{Für } \mu_-: \quad \frac{3}{4} v_1 + \frac{3}{2} v_2 = 0 \quad (6.88)$$

Damit bekommen wir die Eigenvektoren

$$\vec{v}_+ = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \vec{v}_- = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad (6.89)$$

Die Eigenvektoren müssen senkrecht aufeinander stehen, checken wir das eben:

$$\vec{v}_+ \cdot \vec{v}_- = 1 \cdot (-2) + 2 \cdot 1 = 0$$

Damit wir sie für die Matrix V nehmen können, müssen sie aber auch noch normiert sein, sie müssen die Länge 1 haben. Also:

$$\vec{v}_1 := \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \vec{v}_2 := \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad (6.90)$$

und in den Spalten von V stehen dann $\vec{v}_1$ und $\vec{v}_2$,

$$V = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \quad (6.91)$$

Die Matrix D ist gegeben durch

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} \sqrt{4} & 0 \\ 0 & \sqrt{1/4} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad (6.92)$$

und die $\vec{w}_j$ waren gegeben durch

$$\vec{w}_j = \frac{1}{\lambda_j} A \vec{v}_j \quad (6.93)$$

solange die λ_j ungleich 0 sind. Also:

$$\begin{aligned} \vec{w}_1 &= \frac{1}{\lambda_1} \begin{pmatrix} 1 & 3/2 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{2\sqrt{5}} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ \vec{w}_2 &= \frac{1}{\lambda_2} \begin{pmatrix} 1 & 3/2 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{2}{\sqrt{5}} \begin{pmatrix} -1/2 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \end{aligned} \quad (6.94)$$

Diese $\vec{w}_1$ und $\vec{w}_2$ müssen dann automatisch eine Orthonormalbasis bilden und das ist offensichtlich der Fall. Die Matrix W ist dann gegeben durch

$$W = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \quad (6.95)$$

und die Singulärwertzerlegung (6.78)

$$A = W D V^T$$

lautet dann konkret:

$$\begin{pmatrix} 1 & 3/2 \\ 0 & 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T \quad (6.96)$$

Rechnen wir das eben nach, die rechte Seite ist

$$\begin{aligned} \frac{1}{5} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} &= \frac{1}{5} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ -1 & 1/2 \end{pmatrix} \\ &= \frac{1}{5} \begin{pmatrix} 5 & 15/2 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 3/2 \\ 0 & 1 \end{pmatrix} \end{aligned} \quad (6.97)$$

das passt also alles.