Mathematik für InformatikerInnen 2

Frank-Olaf Schreyer

Universität des Saarlandes, SS 2020

Singulärwertzerlegung und Pseudoinverse

Die Themen heute sind

- Singulärwertzerlegung
- Pseudoinverse
- Penrose-Relationen

Die Abbildung

$$\mathbb{R}^{m \times n} \to \mathbb{Z}, \; A \mapsto \operatorname{rang} A$$

ist nicht stetig, da die Werte diskret sind. Die Singulärwertzerlegung ist eine stetige Version des Normalformsatzes für lineare Abbildungen.

Normalform für lineare Abbildungen

Satz. *Sei* $A \in \mathbb{R}^{m \times n}$ *eine Matrix vom Rang* $\operatorname{rang} A = r$. *Dann existieren* $S \in \mathrm{GL}(m, \mathbb{R})$, $T \in \mathrm{GL}(n, \mathbb{R})$, *so dass* $SAT^{-1} = D$ *in folgender Normalform ist:*

$$D = \underbrace{\left(\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ \hline & 0 & & 0 \end{array}\right)}_{r \quad\quad n-r} \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\ \ddots\\ 1\end{array}}\right\} r \\ \left.\vphantom{\begin{array}{c}0\\0\end{array}}\right\} m-r \end{array}$$

Die Abbildung

$$\mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}, \quad A \mapsto D$$

ist nicht stetig, da der Wertebereich diskret ist. Die Singulärwertzerlegung ist eine stetige Version dieses Satzes.

Singulärwertzerlegung

Satz. *Sei* $A \in \mathbb{R}^{m \times n}$ *eine Matrix und* $p = \min(m, n)$. *Dann existieren eindeutig bestimmte reelle Zahlen*

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_p \geqslant 0$$

und zwei orthogonale Matrizen $U \in \mathrm{O}(m)$, $V \in \mathrm{O}(n)$, *so dass* $U^t A V^t = \Sigma$ *die Gestalt*

$$\left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & 0 \\ & & & \sigma_p & \end{array}\right) \quad \textit{oder} \quad \left(\begin{array}{cccc} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \\ \hline & & & \\ & 0 & & \\ & & & \end{array}\right)$$

hat.

Man nennt die σ_j die **Singulärwerte** von A und

$$A = U \Sigma V$$

die **Singulärwertzerlegung** von A.

Numerische Rangberechnung

Sei $A \in \mathbb{R}^{m \times n}$ eine Matrix, von der wir nur eine numerische Approximation $\tilde{A}$ mit Rundungsfehlern kennen. Erfüllen die Singulärwerte $\tilde{\sigma}_j$ von $\tilde{A}$ die Bedingung

$$\tilde{\sigma}_1 \geqslant \ldots \geqslant \tilde{\sigma}_r \gg \tilde{\sigma}_{r+1} \geqslant \ldots \geqslant \tilde{\sigma}_p \geqslant 0,$$

so kann man vermuten, dass A in Wirklichkeit nur Rang r hat. Um mehr Zuversicht in diese Vermutung zu bekommen, verschafft man sich oft eine genauere Approximation $\overline{A}$ von A. Gilt für deren Singulärwerte $\overline{\sigma}_j$

$$\tilde{\sigma}_k \approx \overline{\sigma}_k \text{ für } k \leqslant r$$

und

$$\tilde{\sigma}_k \gg \overline{\sigma}_k \text{ für } k \geqslant r + 1,$$

so können wir recht zuversichtlich sein, dass $\operatorname{rang} A = r$ gilt, und $\overline{\sigma}_k$ für $k \leqslant r$ gute Approximationen der von Null verschiedenen Singulärwerte σ_k von A sind.

Beste Rang-r-Approximation: das Eckhart-Young-Theorem

Sei $A \in \mathbb{R}^{m\times n}$, $m \geqslant n$, mit Singulärwertzerlegung $A = U\Sigma V$ mit

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{pmatrix}. \text{ Für } \quad \Sigma_r = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix}$$

gilt dann:

Satz. *$A_r = U\Sigma_r V$ ist die beste Rang-r-Approximation von A. Von allen Matrizen vom Rang $\leqslant r$ hat A_r zu A den kleinsten Abstand bzgl. der Frobenius-Norm:*

$$\|A - A_r\|_F = \min_{\operatorname{rang} B \leqslant r} \|A - B\|_F = \sqrt{\sigma_{r+1}^2 + \ldots + \sigma_n^2}.$$

Ende Teil 1

Beweis des Satzes über die Singulärwertzerlegung

Beweis. Wir betrachten den Fall $p = \min(m, n) = n$, also gilt $m \geqslant n$. (Ist $m < n$, dann argumentieren wir mit A^t.) Die Matrix

$$B = A^t \cdot A$$

ist symmetrisch und hat nur nichtnegative Eigenwerte: Ist v ein normierter Eigenvektor von B zum Eigenwert λ, dann gilt

$$\begin{aligned} \lambda &= \lambda\langle v, v\rangle = \langle v, Bv\rangle \\ &= \langle v, A^t Av\rangle = \langle Av, Av\rangle \geqslant 0. \end{aligned}$$

Sei $V \in \mathrm{O}(n)$ eine orthogonale Matrix ,die B in Diagonalgestalt

$$VBV^t = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

überführt mit

$$\lambda_1 \geqslant \ldots \geqslant \lambda_n \geqslant 0.$$

Dann sind $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}$ die gesuchten Singulärwerte von A.

Ist $U^t A V^t = \Sigma$ wie im Satz, so folgt

$$VBV^t = VA^t UU^t AV^t = \Sigma^t \Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}.$$

$\sigma_1 \ldots, \sigma_n$ sind also als Quadratwurzeln der Eigenwerte von $A^t A$ durch A eindeutig bestimmt. Sei nun $r = \operatorname{rang} A$, d.h.,

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0$$

Um U zu bestimmen, betrachten wir

$$u_i = \frac{1}{\sqrt{\lambda_i}} A v_i \in \mathbb{R}^m \text{ für } i = 1, \ldots, r,$$

wobei $\mathcal{B} = \{v_1, \ldots, v_n\}$ die Orthonormalbasis von $\mathbb{R}^n$ ist, die aus den Spalten von V^t besteht. Dann gilt

$$\sigma_i u_i = A v_i.$$

Die Vektoren $\{u_1, \dots, u_r\}$ bilden ein Orthonormalsystem im $\mathbb{R}^m$:

$$\begin{aligned}\langle u_i, u_j\rangle &= \frac{1}{\sigma_i\sigma_j}\langle Av_i, Av_j\rangle = \frac{1}{\sigma_i\sigma_j}\langle v_i, A^t Av_j\rangle \\ &= \frac{1}{\sigma_i\sigma_j}\langle v_i, Bv_j\rangle = \frac{\lambda_j}{\sigma_i\sigma_j}\langle v_i, v_j\rangle = \delta_{ij}.\end{aligned}$$

Wir ergänzen diese nun zu einer Orthonormalbasis

$$\mathcal{A} = \{u_1, \dots, u_r, u_{r+1}, \dots, u_m\}$$

von $\mathbb{R}^m$. Bzgl. dieser Basen hat A die Darstellung

$$U^t AV^t = M_{\mathcal{B}}^{\mathcal{A}}(A) = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix},$$

wobei U^t die Spalten $u_1, \dots u_m$ hat. $\square$

Beispiele.

1. Die Singulärwerte einer orthogonalen Matrix $S \in \mathrm{O}(n)$ sind alle 1. In der Tat, für $U = S$ und $V = E$, die Einheitsmatrix, gilt:
$$U^t S E^t = S^t S E = E.$$
2. Die Singulärwerte einer symmetrischen Matrix sind die Beträge der Eigenwerte
$$|\lambda_1|, \ldots, |\lambda_n|.$$
In der Tat,
$$VAV^t = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \Rightarrow VA^t V^t VAV^t = \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{pmatrix}.$$
Also
$$\sigma_1 = \sqrt{\lambda_1^2} = |\lambda_1|, \ldots, \sigma_n = \sqrt{\lambda_n^2} = |\lambda_n|$$
bei geeigneter Anordnung der λ_j.

3. Die Matrix

$$\begin{aligned} A &= \begin{pmatrix} 0.36 & 1.6 & 0.48 \\ 0.48 & -1.2 & 0.64 \end{pmatrix} \\ &= \begin{pmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0.6 & 0 & 0.8 \\ -0.8 & 0 & 0.6 \end{pmatrix} \end{aligned}$$

hat die Singulärwerte $2 > 1$. Hier ist $U \in \mathrm{SO}(2)$ und $V \in \mathrm{O}(3) \setminus \mathrm{SO}(3)$.

4. Die Matrix $A_\epsilon = \begin{pmatrix} 0 & 1 \\ \epsilon & 0 \end{pmatrix}$ hat die Singulärwerte 1 und $|\epsilon|$:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \epsilon & 0 \end{pmatrix} \begin{pmatrix} 0 & \pm 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & |\epsilon| \end{pmatrix} .$$

Für $\epsilon \to 0$ strebt $\sigma_2 \to 0$, während σ_1 konstant 1 bleibt.

Ende Teil 2

Pseudoinverse

Sei $A \in \mathbb{R}^{m \times n}$ eine Matrix und

$$f_A \colon \mathbb{R}^n \to \mathbb{R}^m, f_A(x) = Ax$$

die zugehörige lineare Abbildung. Im allgemeinen ist f_A weder injektiv noch surjektiv. Stattdessen haben wir die ausgezeichneten Untervektorräume $\ker A \subset \mathbb{R}^n$ und $\operatorname{Bild} A \subset \mathbb{R}^m$.

Wir wollen trotzdem eine Art Umkehrabbildung

$$A^+ : \mathbb{R}^m \to \mathbb{R}^n$$

definieren, welche in gewisser Weise die beste Approximation an eine 'Umkehrabbildung' liefert.

Gegeben sei $b \in \mathbb{R}^m$. Wir wollen $x = A^+ b \in \mathbb{R}^n$ finden, so dass

a) $\|Ax - b\|$ minimal wird

b) unter allen solchen, soll $\|x\|$ minimal sein:

$$\|x\| = \min_{z : \|Az - b\| = \min} \|z\|.$$

Mit anderen Worten: Wir betrachten die orthogonale Projektion

$$p : \mathbb{R}^m \to \text{Bild}\, A.$$

Anschließend wählen wir in

$$f_A^{-1}(p(b)) = x_0 + \ker A$$

den Punkt aus, der am dichtesten zum Nullpunkt $0 \in \mathbb{R}^n$ liegt, also

$$\{A^+(b)\} = f_A^{-1}(p(b)) \cap (\ker A)^\perp.$$

Definition. Die so definierte Abbildung (Matrix)

$$A^+ : \mathbb{R}^m \to \mathbb{R}^n$$

heißt **Pseudoinverse** von A.

Eigenschaften der Pseudoinversen

Satz. *Sei $A \in \mathbb{R}^{m\times n}$ und $A^+ \in \mathbb{R}^{n\times m}$ deren Pseudoinverse. Dann gilt:*

1)
$$P = AA^+ : \mathbb{R}^m \to \mathbb{R}^m$$
definiert die orthogonale Projektion von $\mathbb{R}^m$ auf $\operatorname{Bild} A \subset \mathbb{R}^m$. *Insbesondere gilt: $P^2 = P$.*

2)
$$Q = A^+A : \mathbb{R}^n \to \mathbb{R}^n$$
definiert die orthogonale Projektion von $\mathbb{R}^n$ auf $(\ker A)^\perp \subset \mathbb{R}^n$. *Insbesondere gilt: $Q^2 = Q$.*

3) *A und A^+ induzieren zueinander inverse Isomorphismen*
$$(\ker A)^\perp \underset{A^+}{\overset{A}{\rightleftarrows}} \operatorname{Bild} A.$$

4) *Es gilt* $\quad A^+AA^+ = A^+, \quad AA^+A = A,$
und $\quad (A^+A)^t = A^+A, \quad (AA^+)^t = AA^+.$

5) $(A^+)^+ = A$. *Ist A invertierbar, also insbesondere $n = m$, dann gilt $A^+ = A^{-1}$.*

Man nennt die Relationen in 4) die **Penrose-Relationen**.

Beweis. Klar aus der Konstruktion: 1)

$$\begin{aligned} b \mapsto p(b) \in \operatorname{Bild} A \quad &\mapsto \quad A^+ b \in f_A^{-1}(p(b)) \cap (\ker A)^\perp \\ &\mapsto \quad AA^+ b = p(b). \end{aligned}$$

Liegt $b \in \operatorname{Bild} A$ so ist $p(b) = b$, also $p|_{\operatorname{Bild} A} = \mathrm{id}_{\operatorname{Bild} A}$ und deshalb $P^2 = P$.

2) Für $x \in \mathbb{R}^n$ ist $f_A^{-1}(Ax) = x + \ker A$. Wegen

$$\ker A \cap (\ker A)^\perp = 0$$

hat $(x + \ker A) \cap (\ker A)^\perp$ nur einen Punkt, und dieser ist $Qx = A^+ Ax$. Also

$$x \mapsto Qx$$

ist die orthogonale Projektion auf $(\ker A)^\perp$ und $Q^2 = Q$.

3)

$$f_A|_{(\ker A)^\perp} \colon (\ker A)^\perp \to \operatorname{Bild}$$

ist injektiv und wegen

$$\dim(\ker A)^\perp = n - \dim\ker A = \dim \operatorname{Bild} A$$

auch surjektiv, also ein Isomorphismus.

Da $A^+A = Q$ auf $(\ker A)^\perp$ die identische Abbildung induziert, wird die Umkehrabbildung von A^+ induziert.
4)
$$A^+AA^+ = A^+P = A^+, AA^+A = PA = A$$
folgt mit der Konstruktion. Dass A^+A und AA^+ symmetrische Matrizen sind, folgt am klarsten aus dem nächsten Satz, der einen Zusammenhang mit der Singulärwertzerlegung herstellt.
5) $\ker A^+ = (\operatorname{Bild} A)^\perp$ und $\operatorname{Bild} A^+ = (\ker A)^\perp$ sind klar. Also $(\ker A^+)^\perp = (\operatorname{Bild} A)^{\perp\perp} = \operatorname{Bild} A$. Die Konstruktion von $(A^+)^+$ gibt:

also $(A^+)^+ = A$. Schließlich für invertierbare A ist $A^+ = A^{-1}$ klar. $\square$

Ende Teil 3

Pseudoinverse aus der Singulärwertzerlegung

Satz. *Sei* $A \in \mathbb{R}^{m \times n}$ *eine Matrix vom Rang* $\operatorname{rang} A = r$, $A = U\Sigma V$ *deren Singulärwertzerlegung und* A^+ *ihre Pseudoinverse. Dann gilt*

$$A^+ = V^t \Sigma^+ U^t,$$

wobei

$$\Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

Es gilt $\sigma_{r+1} = \ldots = \sigma_{\min(m,n)} = 0$, wegen $\operatorname{rang} A = r$.

Beweis. Die geometrische Beschreibung von A^+ macht klar, dass A^+ unabängig von der Wahl der Orthonormalbasen ist. Also gilt

$$(U^t A V^t)^+ = U^t A^+ V^t$$

für beliebige Matrizen $U \in \mathrm{O}(m)$ und $V \in \mathrm{O}(n)$.
Es reicht also

$$\Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

einzusehen. Dies ist klar:

$$\ker \Sigma = \{0\}^r \times \mathbb{R}^{n-r} \subset \mathbb{R}^n, \quad (\ker \Sigma)^\perp = \mathbb{R}^r \times \{0\}^{n-r}$$
$$\operatorname{Bild} \Sigma = \mathbb{R}^r \times \{0\}^{m-r} \subset \mathbb{R}^m$$

Der Isomorphismus

$$\mathbb{R}^r = (\ker \Sigma)^\perp \to \operatorname{Bild} \Sigma = \mathbb{R}^r$$

wird durch die Diagonalmatrix

$$\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

beschrieben. Deren Inverse ist

$$\begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r} \end{pmatrix},$$

und die Behauptung folgt. □

Beispiel. Sei $A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix}$. Eine mit numerischen Methoden bestimmte Singulärwertzerlegung ist

$$A = U \begin{pmatrix} 6.75317 & 0 & 0 \\ 0 & .628244 & 0 \\ 0 & 0 & 8.94254 \cdot 10^{-17} \end{pmatrix} V$$

mit

$$U = \begin{pmatrix} -.354171 & .841366 & .408248 \\ -.553852 & .163039 & -.816497 \\ -.753532 & -.515288 & .408248 \end{pmatrix}$$

und

$$V = \begin{pmatrix} -.246041 & -.551218 & -.797259 \\ .778544 & -.60235 & .176194 \\ .57735 & .57735 & -.57735 \end{pmatrix}$$

Ganz offensichtlich ist $\sigma_3 = 8.94254 \cdot 10^{-17}$ in Wirklichkeit null.

Die Pseudoinverse ergibt sich approximativ mit

$$A^+ \approx U^t \Sigma_2^+ V^t = \begin{pmatrix} 1.05556 & .222222 & -.611111 \\ -.777778 & -.111111 & .555556 \\ .277778 & .111111 & -.0555556 \end{pmatrix}.$$

Der genaue Wert ist

$$A^+ = \frac{1}{18} \begin{pmatrix} 19 & 4 & -11 \\ -14 & -2 & 10 \\ 5 & 2 & -1 \end{pmatrix} \in \mathbb{Q}^{3\times 3} \subset \mathbb{R}^{3\times 3},$$

denn $\frac{1}{18} = 0.0\overline{5} = 0.0555...$.

Ist $A \in \mathbb{Q}^{m\times n}$ eine Matrix mit rationalen Einträgen, dann hat die Pseudoinverse ebenfalls rationale Einträge.

Satz.

$$A \in \mathbb{Q}^{m\times n} \Rightarrow A^+ \in \mathbb{Q}^{n\times m}.$$

Rationale Matrizen haben rationale Pseudoinverse

Beweis. Hat A rationale Einträge, dann besitzen $\ker A$ und $\operatorname{Bild} A$ jeweils eine Basis aus Vektoren mit rationalen Einträgen. Die orthogonalen Projektionen

$$P\colon \mathbb{R}^m \to \operatorname{Bild} A \subset \mathbb{R}^m \text{ und } Q\colon \mathbb{R}^n \to (\ker A)^\perp \subset \mathbb{R}^n$$

werden durch Matrizen mit rationalen Einträgen definiert, da sich die Quadratwurzeln, die sich im Gram-Schmidt-Verfahren ergeben, beim Projizieren wieder herausheben. Die Bilder Pe_i im $\operatorname{Bild} A$ der Standarbasisvektoren e_i von $\mathbb{R}^n$ haben daher rationale Koordinaten. Der Gauß-Algorithmus gibt uns eine rationale Lösung $x = x^{(i)}$ des Gleichungssystems

$$Ax = Pe_i.$$

Diese projiziert mit Q auf einen rationale Vektor

$$Qx^{(i)} = A^+ e_i \in (\ker A)^\perp \subset \mathbb{R}^n.$$

Die Vektoren $Qx^{(1)}, \ldots Qx^{(n)} \in \mathbb{Q}^m$ bilden also die Spalten von A^+. $\square$

Ende Teil 4