

Singulärwertzerlegung I - Niedrigrangapproximation mit Beispielen

Singulärwertzerlegung

Motivation:

- Zerlegung *jeder* Matrix in unitäre und diagonale Matrizen
- Verhalten/Eigenschaften von A und A^{-1} lassen sich leichter untersuchen (Rang, Bild, Kern, Norm ... direkt ablesbar)
- numerische Vorteile (Invarianzen unitärer Matrizen, Bestandteil vieler stabiler Algorithmen zur Bestimmung von Eigenwerten, Kleinste Quadrate, ...)

Definition: Seien $m, n \in \mathbb{N}$ beliebig. Sei $A \in \mathbb{C}^{m \times n}$ eine Matrix (nicht notwendigerweise voller Rang). Die Zerlegung von A in $A = U\Sigma V'$ mit $U \in \mathbb{C}^{m \times m}$ unitär, $V \in \mathbb{C}^{n \times n}$ unitär und $\Sigma \in \mathbb{R}^{m \times n}$ diagonal heißt die Singulärwertzerlegung (SVD) von A . Zusätzlich nehmen wir an, dass die Diagonaleinträge σ_j von Σ folgendermaßen geordnet sind: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ mit $p = \min\{n, m\}$.

Begriffe:

- Singulärwert (SW) von $A \hat{=} \sigma_j$ (Diagonaleinträge von Σ)
- linksseitiger Singulärvektor (l.SV) $\hat{=} u_j$ (Spaltenvektoren von U)
- rechtsseitiger Singulärvektor (r.SV) $\hat{=} v_j$ (Spaltenvektoren von V)

Satz (SVD): Jede Matrix $A \in \mathbb{C}^{m \times n}$ hat eine Singulärwertzerlegung. Außerdem sind die Singulärwerte σ_j eindeutig bestimmt. Falls A quadratisch und die σ_j paarweise verschieden, sind die links- und rechtsseitigen Singulärvektoren u_j und v_j bis auf das Vorzeichen eindeutig bestimmt.

Beweis:

- Existenz

Betrachte die größte Wirkung von A . Setze dazu $\sigma_1 = \|A\|_2$. Damit existiert nach Definition der Matrixnorm ein Vektor $v_1 \in \mathbb{C}^n$ mit $\|v_1\|_2 = 1$ und $\|u_1\|_2 = \sigma_1$ mit $u_1 = Av_1$. Erweitere v_1 zu einer Orthonormalbasis $\{v_1, \dots, v_n\}$ von \mathbb{C}^n und u_1 zu einer Orthonormalbasis $\{u_1, \dots, u_m\}$ von \mathbb{C}^m . Bezeichne V_1 bzw. U_1 als die unitären Matrizen mit den Basisvektoren der jeweiligen ONB in den Spalten.

$$U_1'AV_1 = S = \begin{pmatrix} \sigma_1 & w' \\ 0 & B \end{pmatrix} \text{ mit } 0 \in \mathbb{C}^{m-1}, w \in \mathbb{C}^{n-1}, B \in \mathbb{C}^{(m-1) \times (n-1)}$$

$$\begin{aligned} \|S\|_2 \cdot \left\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2 &\geq \left\| \begin{pmatrix} \sigma_1 & w' \\ 0 & B \end{pmatrix} \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} \sigma_1^2 + w'w \\ Bw \end{pmatrix} \right\|_2 = \left((\sigma_1^2 + w'w)^2 + \underbrace{(Bw)'Bw}_{\geq 0} \right)^{\frac{1}{2}} \\ &\geq \sigma_1^2 + w'w = (\sigma_1^2 + w'w)^{\frac{1}{2}} \cdot (\sigma_1^2 + w'w)^{\frac{1}{2}} = (\sigma_1^2 + w'w)^{\frac{1}{2}} \cdot \left\| \begin{pmatrix} \sigma_1 \\ w \end{pmatrix} \right\|_2 \end{aligned}$$

$$\Rightarrow \|S\|_2 \geq (\sigma_1^2 + w'w)^{\frac{1}{2}}$$

$$V_1 \text{ und } U_1 \text{ sind unitär} \Rightarrow \|S\|_2 = \|U_1'AV_1\| = \|A\|_2 = \sigma_1$$

$$\Rightarrow \sigma_1 \geq (\sigma_1^2 + w'w)^{\frac{1}{2}} \Rightarrow w = 0 \text{ (fertig für } n=1 \text{ oder } m=1)$$

B beschreibt Wirkung von A auf den Unterraum orthogonal zu v_1 . Nach Induktionsannahme gilt $B = U_2\Sigma_2V_2'$

$$\Rightarrow A = \underbrace{U_1}_{U} \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix} \underbrace{\begin{pmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & V_2 \end{pmatrix}'}_{V'} V_1'$$

- Eindeutigkeit

σ_1 ist durch $\sigma_1 = \|A\|_2$ eindeutig bestimmt

Angenommen, es existiert ein zweiter, linear unabhängiger Vektor w mit $\|w\|_2 = 1$ und $\|Aw\|_2 = \sigma_1$. Sei $v_2 := \frac{w - (v_1'w)v_1}{\|w - (v_1'w)v_1\|_2}$, also $v_2 \perp v_1$. Dann gilt $\|Av_2\|_2 \leq \sigma_1$, es muss aber Gleichheit gelten, sonst ist $\|Aw\|_2 < \sigma_1$ mit

$w = c \cdot v_1 + s \cdot v_2$ und $|c|^2 + |s|^2 = 1$. ($\|Aw\|_2^2 = \|A(cv_1 + sv_2)\|_2^2 \leq (|c|\|A\|_2 + |s|\|A\|_2)^2 = \|A\|_2^2 (|c| + |s|)^2 < \sigma_2^2$)

$\Rightarrow v_2$ ist also zweiter r.SV zu $\sigma_1 \Rightarrow \exists y$ mit $\|y\|_2 = 1$ bestehend aus den letzten $n - 1$ Komponenten von $V_1'v_2$ und $\|By\|_2 = \sigma_1 \Rightarrow \sigma_1$ nicht einfach ζ

Rest folgt induktiv

□

Eigenschaften von A via SVD

- $\text{rg}(A) = r$ (Anzahl der SW ungleich Null)
- $\text{im}(A) = \langle u_1, \dots, u_r \rangle$, $\text{ker}(A) = \langle v_{r+1}, \dots, v_n \rangle$
- $\|A\|_2 = \sigma_1$ und $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2}$
- $\sigma_j = \sqrt{|\lambda_j|}$ mit σ_j SW von A, λ_j EW von $A'A$ bzw. AA' (haben die selben EW und sind positiv semidefinit). Daraus folgt auch die Eindeutigkeit der σ_j .
- $A = A' \Rightarrow \sigma_j = |\lambda_j|$, σ_j SW von A, λ_j EW von A
- $A \in \mathbb{C}^{m \times m} \Rightarrow |\det(A)| = \prod_{j=1}^m \sigma_j$

Bemerkung: Die Beweise dazu sind sehr kurz und folgen stets dem selben Prinzip: A singularär zerlegen und die Eigenschaften an der Diagonalmatrix Σ ablesen.

Niedrigrangapproximation

Satz 3.8: A ist die Summe aus r Rang-1-Matrizen, $A = \sum_{j=1}^r \sigma_j u_j v_j'$

Beweis: Entspricht lediglich $A = U\Sigma V'$ durch äußere Produkte gebildet.

Satz 3.9: Definiere für ein μ mit $0 \leq \mu \leq r$ $A_\mu = \sum_{j=1}^\mu \sigma_j u_j v_j'$. Falls $\mu = \min\{m, n\}$, setze $\sigma_{\mu+1} = 0$. Dann gilt:

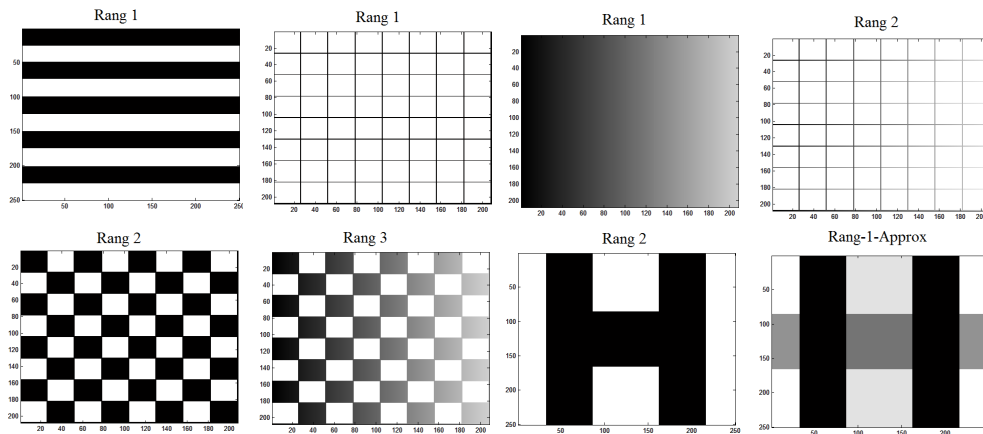
$$\|A - A_\mu\|_2 = \inf_{\substack{B \in \mathbb{C}^{m \times n} \\ \text{rg}(B) \leq \mu}} \|A - B\|_2 = \sigma_{\mu+1}$$

Beweis: Angenommen, es gäbe ein B mit $\text{rg}(B) \leq \mu$, sodass $\|A - B\|_2 < \|A - A_\mu\| = \sigma_{\mu+1}$. $\Rightarrow \exists (n - \mu)$ -dim. Unterraum $W \subset \mathbb{C}^n$, sodass $w \in W \Rightarrow Bw = 0$.

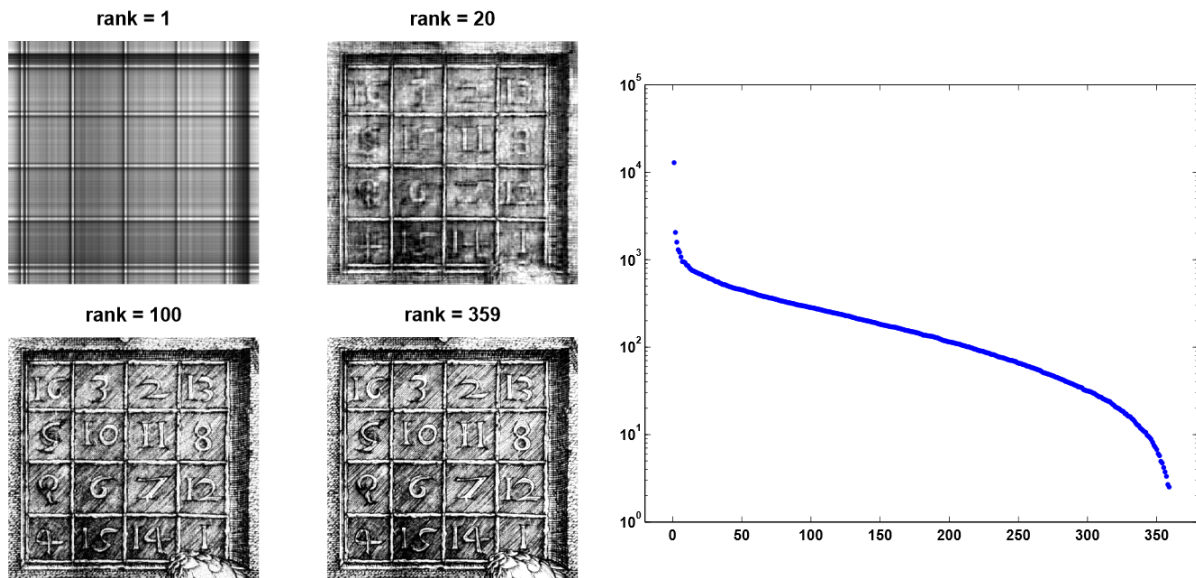
$$\|Aw\|_2 = \|(A - B)w\|_2 \leq \|A - B\|_2 \|w\|_2 < \sigma_{\mu+1} \|w\|_2$$

Aber: es gibt auch einen $(\mu+1)$ -dim. Unterraum mit $\|Aw\|_2 \geq \sigma_{\mu+1} \|w\|_2$ (aufgespannt von $v_1, \dots, v_{\mu+1}$). Damit gibt es ein w , welches in beiden Unterräumen liegt. Dies ist ein Widerspruch. \square

Rangbeispiele und Illustration der Niedrigrangapproximation



Den Zahlenwerten der Matrizen lassen sich Grauwerte zuweisen und man kann so die Matrix bildlich darstellen. Alle Rang-1-Matrizen lassen sich als äußeres Produkt von zwei Vektoren darstellen.

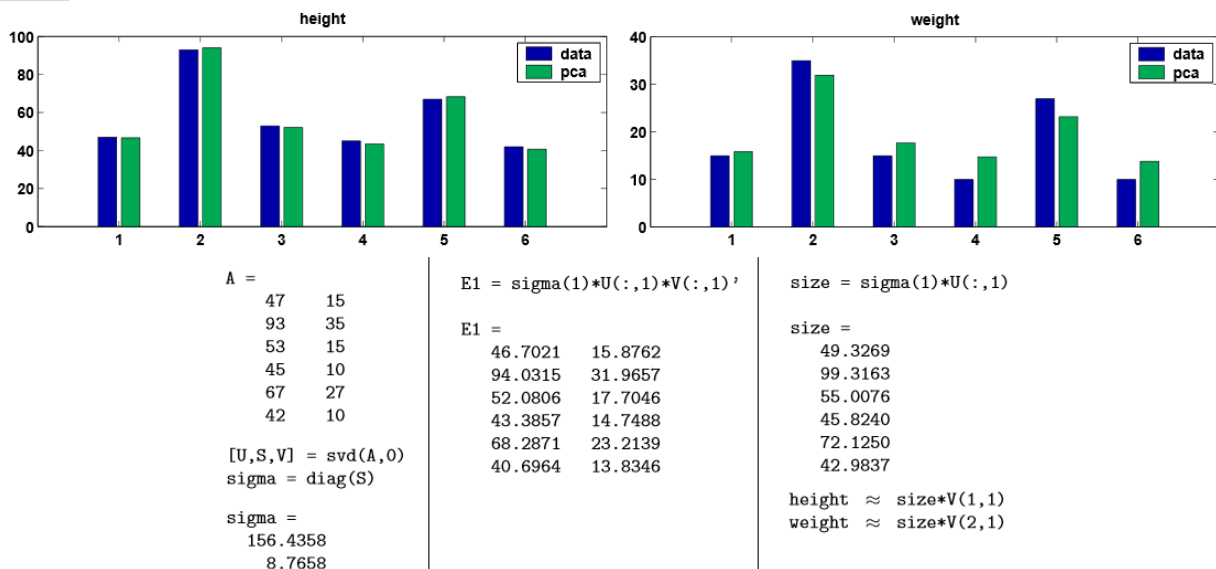


Wie man an diesem Beispiel sieht, braucht es nur wenig Approximations-Schritte, um das fertige Bild zu sehen. Diese schnelle Konvergenz lässt sich anschaulich durch die schnell fallenden Größen der Singulärwerte (logarithmischer Plot rechts) erklären, sowie durch den immer kleiner werdenden Fehler aus Satz 3.9. Je größer der Singulärwert, umso mehr „Information“ steckt also in der zugehörigen Approximation.

Hauptkomponentenanalyse (PCA)

- benutzt Niedrigrangapproximation
- Einsatzgebiete der PCA: Statistik, Geowissenschaften, Archäologie, Bildverarbeitung (insbesondere Mustererkennung), ...

Beispiel:



Die Datenmatrix A möge die Höhe (1. Spalte) und das Gewicht (2. Spalte) von Paketen erfassen. Zerlegen wir diese singular, erkennen wir, dass der erste Singulärwert deutlich größer ist als der zweite. Die Rang-1-Approximation $E1$ liefert daher schon ein recht gutes Ergebnis (vergleiche mit A und der Grafik). Wir gehen von einem starken Zusammenhang von Höhe und Gewicht aus und definieren uns eine Hauptkomponente „size“ wie angegeben. Aus dieser kann man dann die Höhe und das Gewicht berechnen.