Basics in Applied Mathematics – Part I: Numerics

Lecture 6: The singular value decomposition — the pseudoinverse and best low-rank approximation

JProf. Dr. Diyora Salimova

Department of Applied Mathematics, University of Freiburg

5 November 2026

Theorem 3.13 (least squares via QR)

Let $\operatorname{rank} A = n$ and $A = Q\begin{pmatrix} R_1 \\ 0 \end{pmatrix} = Q_1 R_1$ as in Theorem 3.12. Then the least squares solution of $Ax = b$ is the solution of the regular triangular system $R_1 x^* = Q_1^\top b$, and $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = \|Q_2^\top b\|_2$.

$\operatorname{cond}_2(R_1) = \operatorname{cond}_2(A)$ — not $\operatorname{cond}_2(A^\top A) = \operatorname{cond}_2(A)^2$ (Remark 3.14).

Householder: $H_v = I - 2\frac{vv^\top}{v^\top v}$. For $x \neq 0$: $v := x - \alpha e_1$, $\alpha := -\operatorname{sgn}(x_1)\|x\|_2$ ($\operatorname{sgn} 0 := 1$), then $H_v x = \alpha e_1$ (Lemma 3.21). Column by column: $H_p \cdots H_1 A = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$, $H_p \cdots H_1 b = Q^\top b$ — Q is never formed.

Why this sign of α? $v_1 = x_1 + \operatorname{sgn}(x_1)\|x\|_2$ adds numbers of the same sign: no cancellation.

Today: $\operatorname{rank} A < n$ — R_1 is singular and the least squares solution is not unique; the SVD selects the one of minimal norm (Section 4.3).

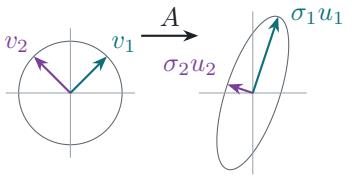
Theorem 4.1 (singular value decomposition)

Let $A \in \mathbb{R}^{m\times n}$ and $p := \min(m,n)$. There exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ and a matrix $\Sigma \in \mathbb{R}^{m\times n}$ with $\Sigma_{ij} = 0$ for $i \neq j$, $\Sigma_{ii} = \sigma_i$, $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$, such that $A = U\Sigma V^\top$. The **singular values** $\sigma_1, \ldots, \sigma_p$ are uniquely determined by A. The columns u_i of U and v_i of V (left and right **singular vectors**) satisfy $Av_i = \sigma_i u_i$ and $A^\top u_i = \sigma_i v_i$, $i = 1, \ldots, p$.

Reduced SVD ($r :=$ number of nonzero σ_i):

$$A = U_r \Sigma_r V_r^\top = \sum_{i=1}^{r} \sigma_i \, u_i v_i^\top \qquad (4.1)$$

A maps the unit sphere to an **ellipsoid** with semi-axes $\sigma_i u_i$.



$A = \left(\begin{smallmatrix} 3 & 0 \\ 4 & 5 \end{smallmatrix}\right)$: unit circle $\mapsto$ ellipse

Today: the proof of Theorem 4.1 (blackboard), then Example 4.2 and Sections 4.2–4.4.

Example 4.2: singular values by hand

$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 3}$. The nonzero eigenvalues of $AA^\top$ and $A^\top A$ coincide (proof of Proposition 1.23(i)) — use the smaller matrix.

Recipe: $AA^\top = ?$ its eigenvalues ? $\sigma_1, \sigma_2 = ?$

$AA^\top = \begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix}$, eigenvalues 12 and 10 $\Rightarrow$ $\sigma_1 = \sqrt{12} \approx 3.46$, $\sigma_2 = \sqrt{10} \approx 3.16$;
$\operatorname{rank} A = 2$, and $A^\top A \in \mathbb{R}^{3\times 3}$ has the eigenvalues 12, 10, 0.

Proposition 4.3

Let $A = U\Sigma V^\top \in \mathbb{R}^{m\times n}$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_p$. Then:

(i) $\operatorname{rank} A = r$, $\operatorname{Im} A = \operatorname{span}(u_1, \ldots, u_r)$ and $\ker A = \operatorname{span}(v_{r+1}, \ldots, v_n)$;

(ii) $\|A\|_2 = \sigma_1$ and $\|A\|_F = \left(\sigma_1^2 + \cdots + \sigma_r^2\right)^{1/2}$;

(iii) $m = n$, A regular: $\|A^{-1}\|_2 = 1/\sigma_n$ and $\operatorname{cond}_2(A) = \sigma_1/\sigma_n$;

(iv) $m = n$, A symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$: the singular values are $|\lambda_1|, \ldots, |\lambda_n|$ (in decreasing order);

(v) $m = n$: $\sigma_n = \min\{\|A - B\|_2 : B \in \mathbb{R}^{n\times n} \text{ singular}\}$, the distance of A to the singular matrices.

Blackboard: proof.

$\operatorname{rank} A < n$: least squares solutions exist (Theorem 3.4), but with x^* every $x^* + z$, $z \in \ker A \neq \{0\}$, is one as well; $A^\top A$ and R_1 are singular. **Choose the solution of minimal norm.**

Definition 4.4 (pseudoinverse)

Let $A = U\Sigma V^\top \in \mathbb{R}^{m\times n}$ with $\operatorname{rank} A = r$. The **(Moore–Penrose) pseudoinverse** of A is

$$A^+ := V\Sigma^+ U^\top \in \mathbb{R}^{n\times m},$$

where $\Sigma^+ \in \mathbb{R}^{n\times m}$ is the matrix with $(\Sigma^+)_{ii} = \sigma_i^{-1}$ for $i \le r$ and all other entries equal to 0. Equivalently, $A^+ = \sum_{i=1}^{r} \sigma_i^{-1} v_i u_i^\top = V_r \Sigma_r^{-1} U_r^\top$.

For instance $m = 3$, $n = 2$, $r = 1$: $\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\Sigma^+ = \begin{pmatrix} \sigma_1^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Invert the nonzero σ_i, keep the zeros, transpose.

The minimal-norm least squares solution

Theorem 4.5 (minimal-norm least squares solution)

Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The vector $x^+ := A^+ b$ is the least squares solution of $Ax = b$ **of minimal Euclidean norm**: it minimises $\|Ax - b\|_2$, and among all minimisers it is the unique one with the smallest norm $\|x\|_2$. Moreover:

(i) if $\operatorname{rank} A = n$, then $A^+ = (A^\top A)^{-1} A^\top$;

(ii) if $m = n$ and A is regular, then $A^+ = A^{-1}$;

(iii) $AA^+ = U_r U_r^\top$ is the orthogonal projection onto $\operatorname{Im} A$, and $A^+ A = V_r V_r^\top$ is the orthogonal projection onto $(\ker A)^\perp = \operatorname{Im} A^\top$.

Blackboard: proof.

Example 4.6: a rank-deficient problem

$A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$, $b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\operatorname{rank} A = 1$.

$Ax = b \iff x_1 + x_2 = 1$: a whole line of exact solutions.

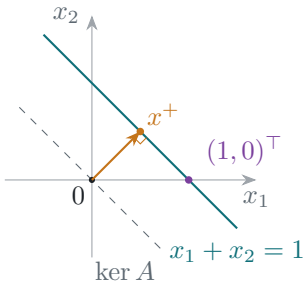
Which one has minimal norm?

$A^\top A = \begin{pmatrix} 5 & 5 \\ 5 & 5 \end{pmatrix}$, eigenvalues $10,\ 0$: $\sigma_1 = \sqrt{10}$,

$v_1 = \frac{1}{\sqrt{2}}(1,1)^\top$, $u_1 = Av_1/\sigma_1 = \frac{1}{\sqrt{5}}(1,2)^\top$.

$A^+ = \sigma_1^{-1} v_1 u_1^\top = ?$ $\frac{1}{10}\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$, $x^+ = A^+ b = (\frac{1}{2}, \frac{1}{2})^\top$.

$x^+ \perp \ker A = \operatorname{span}\left((1,-1)^\top\right)$: the point of the line closest to 0 (e.g. $\|(1,0)^\top\|_2 = 1 > \|x^+\|_2$).



Cut the reduced SVD (4.1) off after k terms:

$$A_k := \sum_{i=1}^{k} \sigma_i \, u_i v_i^\top = U_k \Sigma_k V_k^\top \qquad (k < r = \operatorname{rank} A). \tag{4.2}$$

$\operatorname{rank} A_k = k$; storage: $k(m+n+1)$ numbers instead of mn. **How good is A_k?**

Notation: $\sigma_1(M) \geq \sigma_2(M) \geq \dots$ are the singular values of $M \in \mathbb{R}^{m\times n}$, $\sigma_l(M) := 0$ for $l > \min(m,n)$.

Lemma 4.8

Let $M \in \mathbb{R}^{m\times n}$ with SVD $M = U\Sigma V^\top$ and let $x \in \mathbb{R}^n$ be a unit vector.

(i) If $x \in \operatorname{span}(v_1, \dots, v_k)$ for some $k \leq n$, then $\|Mx\|_2 \geq \sigma_k(M)$.

(ii) If $x \in \operatorname{span}(v_k, \dots, v_n)$ for some $k \leq n$, then $\|Mx\|_2 \leq \sigma_k(M)$.

Blackboard: proof.

Singular values under perturbations

Lemma 4.9 (Weyl's inequality for singular values)

For $X, Y \in \mathbb{R}^{m\times n}$ and all $i, j \geq 1$,

$$\sigma_{i+j-1}(X+Y) \leq \sigma_i(X) + \sigma_j(Y).$$

In particular, $|\sigma_i(A+\Delta A) - \sigma_i(A)| \leq \|\Delta A\|_2$ for all $A, \Delta A \in \mathbb{R}^{m\times n}$ and all i.

Proof*: in the notes.

Best approximation of rank k

Theorem 4.10 (Eckart–Young–Mirsky)

Let $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank} A = r$, let $k < r$ and let A_k be defined by (4.2). Then

$$\min_{\operatorname{rank} B \le k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1},$$

$$\min_{\operatorname{rank} B \le k} \|A - B\|_F = \|A - A_k\|_F = \Big(\sum_{i=k+1}^{r} \sigma_i^2 \Big)^{1/2},$$

where the minima are taken over all $B \in \mathbb{R}^{m\times n}$ of rank at most k.

Blackboard: proof.

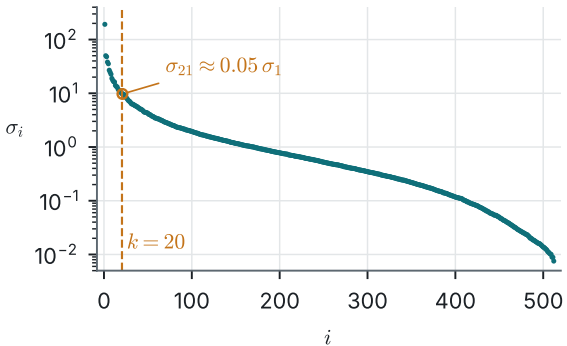
A_k attains both minima and is read off from the SVD. Relative errors:

$$\frac{\|A - A_k\|_2}{\|A\|_2} = \frac{\sigma_{k+1}}{\sigma_1}, \qquad \frac{\|A - A_k\|_F}{\|A\|_F} = \Big(\frac{\sigma_{k+1}^2 + \cdots + \sigma_r^2}{\sigma_1^2 + \cdots + \sigma_r^2} \Big)^{1/2}.$$

Experiment: image compression (1/2)

Section 4.4.1

A grey-scale image is a matrix: 600×512, entries in $[0, 1]$, rank 512.



k	storage	$\frac{\|A - A_k\|_F}{\|A\|_F}$	$\frac{\sigma_{k+1}}{\sigma_1}$
5	1.8 %	0.35	0.14
20	7.2 %	0.19	0.05
50	18.1 %	0.11	0.02

storage: $k(m + n + 1)/(mn)$

A few large singular values dominate ($\sigma_{21} \approx 0.05\,\sigma_1$): a small k should capture most of the image. What do A_5, A_{20}, A_{50} look like?

Experiment: image compression (2/2)

The image A and its truncated SVDs A_k of (4.2), with the storage needed:



A (rank 512) A_5 (1.8 %) A_{20} (7.2 %) A_{50} (18.1 %)

Already A_{20} is clearly recognisable — and by Theorem 4.10 no matrix of rank at most 20 is closer to A, in $\|\cdot\|_2$ or in $\|\cdot\|_F$.

Summary of Chapter 4

- **SVD**: every $A \in \mathbb{R}^{m\times n}$ is $A = U\Sigma V^\top$ with U, V orthogonal and Σ diagonal, $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$; $\sigma_1^2, \ldots, \sigma_p^2$ are the p largest eigenvalues of $A^\top A$
- read off the SVD: $\operatorname{rank} A = r$, $\operatorname{Im} A$, $\ker A$, $\|A\|_2 = \sigma_1$, $\|A\|_F$, $\operatorname{cond}_2(A) = \sigma_1/\sigma_n$, the distance σ_n to the singular matrices
- **pseudoinverse** $A^+ = V\Sigma^+U^\top$: A^+b is the least squares solution of minimal norm, for every rank
- **best rank-k approximation**: $A_k = \sum_{i\leq k} \sigma_i u_i v_i^\top$, error σ_{k+1} in the 2-norm and $\big(\sum_{i>k} \sigma_i^2\big)^{1/2}$ in the Frobenius norm (Eckart–Young–Mirsky) — compression; singular values change by at most $\|\Delta A\|_2$ (Weyl)