భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad
॥ नवोन्मेषेण देश निर्माणम् ॥

Matrix Theory

23 SVD and PCA

Dr. Konda Reddy Mopuri
Dept. of AI, IIT Hyderabad
July-Nov 2026

What we established last time

For a *symmetric* matrix $A \in \mathbb{R}^{n \times n}$:

- All eigenvalues are real.
- Eigenvectors for distinct eigenvalues are orthogonal; one can always choose an *orthonormal* basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of eigenvectors.
- Hence $A = Q \Lambda Q^{\mathsf{T}}$ with $Q^{\mathsf{T}} Q = I$ and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

The form we will keep reusing

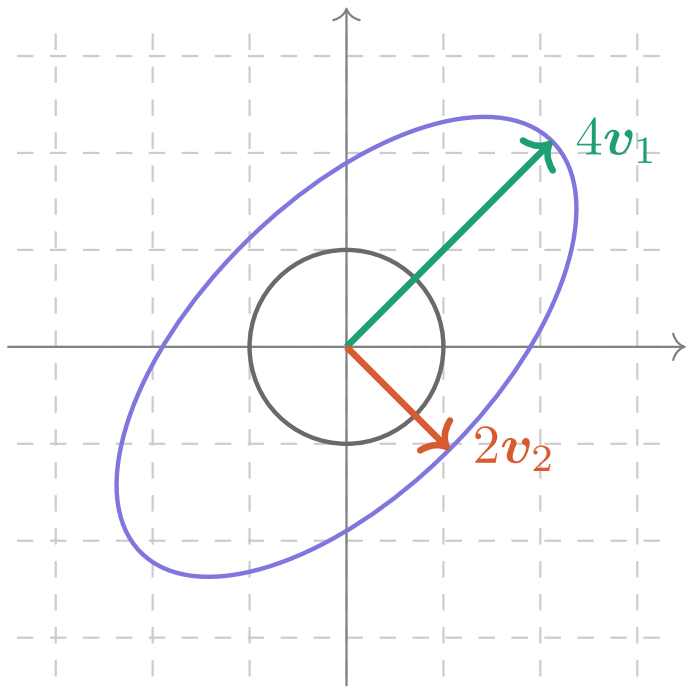
$$A = \sum_{k=1}^{n} \lambda_k \, \boldsymbol{v}_k \boldsymbol{v}_k^{\mathsf{T}}, \qquad \boldsymbol{v}_k \boldsymbol{v}_k^{\mathsf{T}} = \text{orthogonal projector onto } \operatorname{span}\{\boldsymbol{v}_k\}.$$

Each summand has rank one. A acts independently on n orthogonal directions.

The running example

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \qquad \lambda_1 = 4,\ \boldsymbol{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \lambda_2 = 2,\ \boldsymbol{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$A = 4 \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} + 2 \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$$



The unit circle maps to an ellipse with semi-axes λ_1, λ_2 along $\boldsymbol{v}_1, \boldsymbol{v}_2$.

Functions of a matrix come for free

Because Q is orthogonal, $A^k = Q\Lambda^k Q^{\mathsf{T}}$, and more generally $f(A) = Q f(\Lambda) Q^{\mathsf{T}}$:

Object	Rank-one expansion
the matrix	$A = \sum_k \lambda_k \boldsymbol{v}_k \boldsymbol{v}_k^{\mathsf{T}}$
k discrete steps	$A^m = \sum_k \lambda_k^m \, \boldsymbol{v}_k \boldsymbol{v}_k^{\mathsf{T}}$
flow of $\dot{\boldsymbol{x}} = A\boldsymbol{x}$	$e^{At} = \sum_k e^{\lambda_k t} \, \boldsymbol{v}_k \boldsymbol{v}_k^{\mathsf{T}}$
inverse (if $\lambda_k \neq 0$)	$A^{-1} = \sum_k \lambda_k^{-1} \, \boldsymbol{v}_k \boldsymbol{v}_k^{\mathsf{T}}$

Key Idea

Diagonalization replaces a coupled n-dimensional problem by n uncoupled scalar problems.

But the spectral theorem is fragile

It needs A to be square *and* symmetric. Outside that, things break:

Rectangular

$A \in \mathbb{R}^{m \times n}$ with $m \neq n$: "$A\boldsymbol{v} = \lambda \boldsymbol{v}$" is not even type-correct.

Defective

$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ has only a one-dimensional eigenspace: no eigenbasis.

Complex eigenvalues

$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has $\lambda = \pm i$: no real eigenvector at all.

Non-orthogonal eigenvectors

Even when $A = S\Lambda S^{-1}$ exists, S may be badly conditioned, so the "decomposition" is numerically useless.

The question for today

Is there a decomposition with all the good properties (orthogonal bases, a diagonal middle, a rank-one expansion) that exists for *every* matrix?

Minor deviation - Condition Number

What is the Condition Number?

- The **condition number** measures how sensitive a linear system ($Ax = b$) is to errors or inaccuracies in the data.
- It quantifies how much an error in the input (the vector b or matrix A) will be amplified in the output solution (x).

Mathematical Definition

For a square, invertible matrix A, the condition number $\kappa(A)$ is defined using matrix norms:

General Definition

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

Mathematical Definition

For a square, invertible matrix A, the condition number $\kappa(A)$ is defined using matrix norms:

General Definition

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

For solving Ax

Suppose we solve with a slightly wrong right-hand side, $b+\delta b$ instead of b. The solution shifts to $x + \delta x$. The bound is $\frac{\|\delta x\|}{\|x\|} \leq \kappa(A) \cdot \frac{\|\delta b\|}{\|b\|}$

Mathematical Definition

For a square, invertible matrix A, the condition number $\kappa(A)$ is defined using matrix norms:

General Definition

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

Using the L_2 (Spectral) Norm

For a normal matrix, it equals the ratio of its maximum to minimum singular values:

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

Interpreting the Condition Number

- $\kappa(A) = 1$**:** Ideal scenario. Occurs for orthogonal/unitary matrices. Errors are not amplified.
- **Low** $\kappa(A)$ **(Well-conditioned):** The matrix is stable. Small input changes yield small output changes.
- **High** $\kappa(A)$ **(Ill-conditioned):** The matrix is highly sensitive. Small input errors cause massive swings in the output.
- $\kappa(A) = \infty$**:** The matrix is singular (not invertible). No unique solution exists.

Interpreting the Condition Number

- $\kappa(A) = 1$**:** Ideal scenario. Occurs for orthogonal/unitary matrices. Errors are not amplified.
- **Low** $\kappa(A)$ **(Well-conditioned):** The matrix is stable. Small input changes yield small output changes.
- **High** $\kappa(A)$ **(Ill-conditioned):** The matrix is highly sensitive. Small input errors cause massive swings in the output.
- $\kappa(A) = \infty$**:** The matrix is singular (not invertible). No unique solution exists.

The condition number is a continuous measure of nearness to singularity.

The "Rule of Thumb" for Precision

The condition number directly relates to loss of precision in floating-point arithmetic.

The "Rule of Thumb" for Precision

The condition number directly relates to loss of precision in floating-point arithmetic.

Loss of Significance

If your computer uses a precision of d decimal digits, and $\kappa(A) \approx 10^k$, you can expect to **lose up to k digits of accuracy** in your solution.

Example: In standard 16-digit double precision, if $\kappa(A) = 10^7$, your final answer may only be accurate to about 9 digits $(16 - 7 = 9)$.

Why It Matters in Practice

- **Numerical Analysis & Computing:** Rounding errors are inevitable. If a matrix is ill-conditioned, computer outputs can be entirely wrong even with flawless algorithms.
- **Data Science & Regression:** In linear regression, highly correlated independent variables (multicollinearity) lead to ill-conditioned matrices, making calculated coefficients highly unstable.

Back to the motivation

What we must give up, and what we can keep

For a symmetric A we had *one* orthonormal basis doing two jobs at once:

$$A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k \quad \text{(input direction} = \text{output direction)}$$

What we must give up, and what we can keep

For a symmetric A we had *one* orthonormal basis doing two jobs at once:

$$A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k \quad \text{(input direction = output direction)}$$

For a general $A : \mathbb{R}^n \to \mathbb{R}^m$ input and output even live in different spaces. So allow *two* bases:

$$\boxed{A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k}$$

$$\{\boldsymbol{v}_k\} \subset \mathbb{R}^n \text{ orthonormal}, \quad \{\boldsymbol{u}_k\} \subset \mathbb{R}^m \text{ orthonormal}, \quad \sigma_k \geq 0.$$

What we must give up, and what we can keep

For a symmetric A we had *one* orthonormal basis doing two jobs at once:

$$A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k \quad \text{(input direction = output direction)}$$

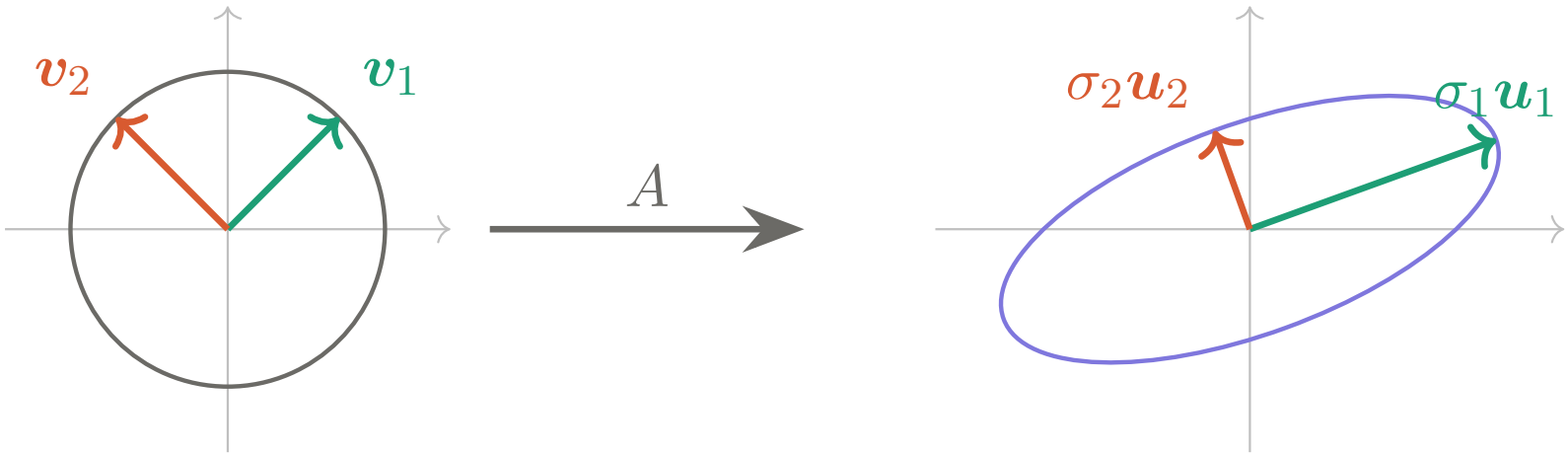
For a general $A : \mathbb{R}^n \to \mathbb{R}^m$ input and output even live in different spaces. So allow *two* bases:

$$\boxed{A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k}$$

$$\{\boldsymbol{v}_k\} \subset \mathbb{R}^n \text{ orthonormal}, \quad \{\boldsymbol{u}_k\} \subset \mathbb{R}^m \text{ orthonormal}, \quad \sigma_k \geq 0.$$

Give up "same basis in and out". Keep orthonormality on both sides and a non-negative diagonal scaling in between. Remarkably, this is always possible.

A geometric promise



unit sphere in $\mathbb{R}^n$ hyper-ellipse in $\mathbb{R}^m$

Claim. Every linear map sends the unit sphere to a (possibly degenerate) hyper-ellipse. The σ_k are the semi-axis lengths, the $\boldsymbol{u}_k$ their directions, and the $\boldsymbol{v}_k$ are the preimages of those axes.

Singular Value Decomposition (SVD)

Every $A \in \mathbb{R}^{m \times n}$ can be written

$$A = U \Sigma V^{\mathsf{T}}$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal and $\Sigma \in \mathbb{R}^{m \times n}$ is "diagonal" with entries

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0, \qquad p = \min(m, n).$$

- σ_k: **singular values**; u_k: **left singular vectors**; v_k: **right singular vectors**.
- A can be a Rectangular, singular, defective, etc.

Three equivalent ways to read $A = U\Sigma V^{\mathsf{T}}$

भारतीय सांकेतिक विज्ञान संस्थ हैदराबाद
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad
॥ नवोन्मेषेण देश निर्माणम् ॥

1. **As a factorisation.**

$$A = \underbrace{U}_{\text{rotate}} \quad \underbrace{\Sigma}_{\text{stretch}} \quad \underbrace{V^{\mathsf{T}}}_{\text{rotate}}$$

Every linear map is a rotation, then an axis-aligned scaling, then another rotation.

Three equivalent ways to read $A = U\Sigma V^{\mathsf{T}}$

भारतीय सांकेतिक विज्ञान संस्थ हैदराबाद
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad
॥ नवोन्मेषेण देश निर्माणम् ॥

1. **As a factorisation.**

$$A = \underbrace{U}_{\text{rotate}} \quad \underbrace{\Sigma}_{\text{stretch}} \quad \underbrace{V^{\mathsf{T}}}_{\text{rotate}}$$

Every linear map is a rotation, then an axis-aligned scaling, then another rotation.

2. **As a basis statement.**

$$A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k, \quad k = 1, \ldots, p, \qquad (\text{and } A\boldsymbol{v}_k = \mathbf{0} \text{ for } k > r)$$

Three equivalent ways to read $A = U\Sigma V^{\mathsf{T}}$

भारतीय सांकेतिक विज्ञान संस्थ हैदराबाद
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad
॥ नवोन्मेषेण देश निर्माणम् ॥

1. **As a factorisation.**

$$A = \underbrace{U}_{\text{rotate}} \quad \underbrace{\Sigma}_{\text{stretch}} \quad \underbrace{V^{\mathsf{T}}}_{\text{rotate}}$$

Every linear map is a rotation, then an axis-aligned scaling, then another rotation.

2. **As a basis statement.**

$$A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k, \quad k = 1, \ldots, p, \qquad (\text{and } A\boldsymbol{v}_k = \mathbf{0} \text{ for } k > r)$$

3. **As a sum of rank-one pieces.** (the one we will actually use)

$$\boxed{A = \sum_{k=1}^{r} \sigma_k \, \boldsymbol{u}_k \boldsymbol{v}_k^{\mathsf{T}}}, \qquad r = \operatorname{rank} A.$$

Three equivalent ways to read $A = U\Sigma V^{\mathsf{T}}$

भारतीय सांकेतिक विज्ञान संस्थ हैदराबाद
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad
॥ नवोन्मेषेण देश निर्माणम् ॥

1. **As a factorisation.**

$$A = \underbrace{U}_{\text{rotate}} \quad \underbrace{\Sigma}_{\text{stretch}} \quad \underbrace{V^{\mathsf{T}}}_{\text{rotate}}$$

Every linear map is a rotation, then an axis-aligned scaling, then another rotation.

2. **As a basis statement.**

$$A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k, \quad k = 1, \ldots, p, \qquad (\text{and } A\boldsymbol{v}_k = \mathbf{0} \text{ for } k > r)$$

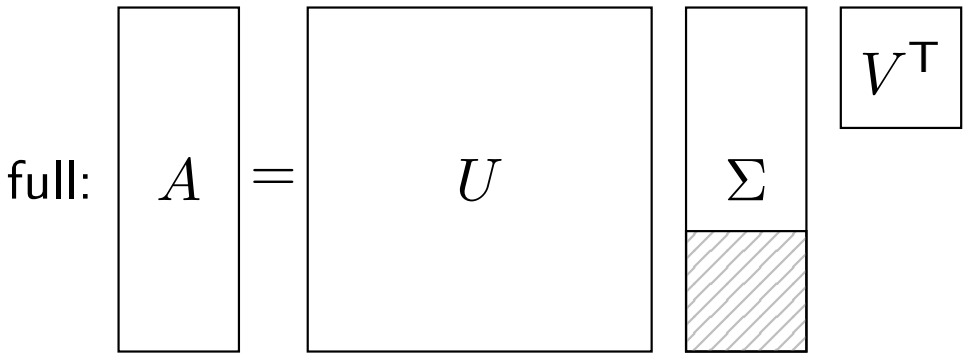
3. **As a sum of rank-one pieces.** (the one we will actually use)

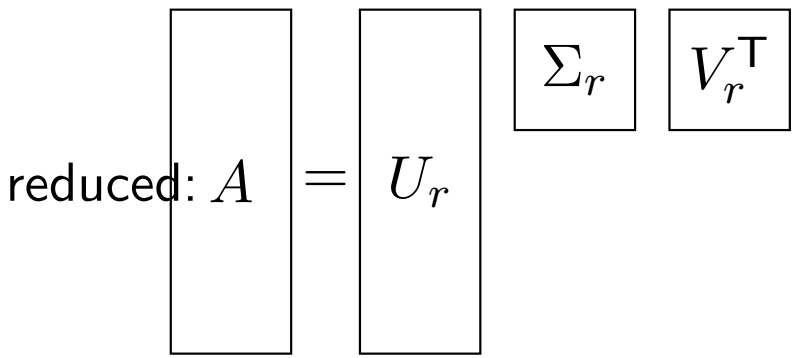
$$\boxed{A = \sum_{k=1}^{r} \sigma_k \, \boldsymbol{u}_k \boldsymbol{v}_k^{\mathsf{T}}}, \qquad r = \operatorname{rank} A.$$

4. Note the shape of each term: $\boldsymbol{u}_k \boldsymbol{v}_k^{\mathsf{T}}$ is $(m \times 1)(1 \times n) = m \times n$. Same as A.

Full versus reduced ("thin") SVD

Suppose $m > n$. Then Σ has $m - n$ rows of zeros, so the last $m - n$ columns of U multiply nothing:





- **Reduced:** keep only the $r = \operatorname{rank} A$ columns with $\sigma_k > 0$. Then $U_r^{\mathsf{T}} U_r = I_r$, $V_r^{\mathsf{T}} V_r = I_r$ (but $U_r U_r^{\mathsf{T}} \neq I$ — it is a projector).

Construction: SVD from the spectral theorem

The trick: A may be troublesome, but $G = A^{\mathsf{T}}A$ is well-behaved.

Two facts about G

1. **Symmetric:** $G^{\mathsf{T}} = (A^{\mathsf{T}}A)^{\mathsf{T}} = A^{\mathsf{T}}A = G$. So the spectral theorem applies: $G = V\Lambda V^{\mathsf{T}}$ with V orthogonal.

Construction: SVD from the spectral theorem

The trick: A may be troublesome, but $G = A^{\mathsf{T}}A$ is well-behaved.

Two facts about G

1. **Symmetric:** $G^{\mathsf{T}} = (A^{\mathsf{T}}A)^{\mathsf{T}} = A^{\mathsf{T}}A = G$. So the spectral theorem applies: $G = V\Lambda V^{\mathsf{T}}$ with V orthogonal.
2. **Positive semidefinite:** for any $\boldsymbol{x}$, $\boldsymbol{x}^{\mathsf{T}}G\boldsymbol{x} = \boldsymbol{x}^{\mathsf{T}}A^{\mathsf{T}}A\boldsymbol{x} = \|A\boldsymbol{x}\|^2 \geq 0$. Hence every eigenvalue $\lambda_k \geq 0$.

Construction: SVD from the spectral theorem

The trick: A may be troublesome, but $G = A^{\mathsf{T}}A$ is well-behaved.

So we may define

$$\sigma_k := \sqrt{\lambda_k(A^{\mathsf{T}}A)} \geq 0, \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n.$$

The right singular vectors $\boldsymbol{v}_k$ are *defined* to be the eigenvectors of $A^{\mathsf{T}}A$.

Construction, continued: producing the $\boldsymbol{u}_k$

Let r be the number of strictly positive σ_k. For $k \leq r$ define

$$\boldsymbol{u}_k := \frac{1}{\sigma_k} A\boldsymbol{v}_k.$$

Construction, continued: producing the $\boldsymbol{u}_k$

Let r be the number of strictly positive σ_k. For $k \leq r$ define

$$\boldsymbol{u}_k := \frac{1}{\sigma_k} A\boldsymbol{v}_k.$$

These are orthonormal:

$$\boldsymbol{u}_j^{\mathsf{T}} \boldsymbol{u}_k = \frac{1}{\sigma_j \sigma_k} \boldsymbol{v}_j^{\mathsf{T}} A^{\mathsf{T}} A \boldsymbol{v}_k = \frac{1}{\sigma_j \sigma_k} \boldsymbol{v}_j^{\mathsf{T}} (\sigma_k^2 \boldsymbol{v}_k) = \frac{\sigma_k}{\sigma_j} \boldsymbol{v}_j^{\mathsf{T}} \boldsymbol{v}_k = \delta_{jk}.$$

Construction, continued: producing the $\boldsymbol{u}_k$

Let r be the number of strictly positive σ_k. For $k \leq r$ define

$$\boldsymbol{u}_k := \frac{1}{\sigma_k} A\boldsymbol{v}_k.$$

These are orthonormal:

$$\boldsymbol{u}_j^{\mathsf{T}} \boldsymbol{u}_k = \frac{1}{\sigma_j \sigma_k} \boldsymbol{v}_j^{\mathsf{T}} A^{\mathsf{T}} A \boldsymbol{v}_k = \frac{1}{\sigma_j \sigma_k} \boldsymbol{v}_j^{\mathsf{T}} (\sigma_k^2 \boldsymbol{v}_k) = \frac{\sigma_k}{\sigma_j} \boldsymbol{v}_j^{\mathsf{T}} \boldsymbol{v}_k = \delta_{jk}.$$

And for $k > r$: $\|A\boldsymbol{v}_k\|^2 = \boldsymbol{v}_k^{\mathsf{T}} A^{\mathsf{T}} A \boldsymbol{v}_k = \sigma_k^2 = 0$, so $A\boldsymbol{v}_k = \boldsymbol{0}$.

Construction, continued: producing the $\boldsymbol{u}_k$

Let r be the number of strictly positive σ_k. For $k \le r$ define

$$\boldsymbol{u}_k := \frac{1}{\sigma_k} A\boldsymbol{v}_k.$$

These are orthonormal:

$$\boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{u}_k = \frac{1}{\sigma_j\sigma_k}\boldsymbol{v}_j^{\mathsf{T}}A^{\mathsf{T}}A\boldsymbol{v}_k = \frac{1}{\sigma_j\sigma_k}\boldsymbol{v}_j^{\mathsf{T}}(\sigma_k^2\boldsymbol{v}_k) = \frac{\sigma_k}{\sigma_j}\,\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{v}_k = \delta_{jk}.$$

And for $k > r$: $\|A\boldsymbol{v}_k\|^2 = \boldsymbol{v}_k^{\mathsf{T}}A^{\mathsf{T}}A\boldsymbol{v}_k = \sigma_k^2 = 0$, so $A\boldsymbol{v}_k = \boldsymbol{0}$.

Extend $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ to an orthonormal basis of $\mathbb{R}^m$ (Gram–Schmidt). Then for every k,

$$A\boldsymbol{v}_k = \sigma_k\boldsymbol{u}_k \quad\Longrightarrow\quad AV = U\Sigma \quad\Longrightarrow\quad A = U\Sigma V^{\mathsf{T}}. \qquad \blacksquare$$

The two Gram matrices

Substituting $A = U\Sigma V^{\mathsf{T}}$:

$$A^{\mathsf{T}}A = V\Sigma^{\mathsf{T}}U^{\mathsf{T}}U\Sigma V^{\mathsf{T}} = V(\Sigma^{\mathsf{T}}\Sigma)V^{\mathsf{T}}, \qquad AA^{\mathsf{T}} = U(\Sigma\Sigma^{\mathsf{T}})U^{\mathsf{T}}.$$

	eigenvectors	eigenvalues
$A^{\mathsf{T}}A$ $(n \times n)$	right singular vectors $\boldsymbol{v}_k$	σ_k^2
AA^{T} $(m \times m)$	left singular vectors $\boldsymbol{u}_k$	σ_k^2

Worked example

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \qquad \Longrightarrow \qquad A^{\mathsf{T}}A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$$

Worked example

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \qquad \Longrightarrow \qquad A^{\mathsf{T}}A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$$

Eigenvalues of $A^{\mathsf{T}}A$: $\lambda = 45,\ 5$ with eigenvectors $\frac{1}{\sqrt{2}}(1,1)$ and $\frac{1}{\sqrt{2}}(1,-1)$. Hence

$$\sigma_1 = \sqrt{45} = 3\sqrt{5} \approx 6.708, \qquad \sigma_2 = \sqrt{5} \approx 2.236.$$

Worked example

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \qquad \Longrightarrow \qquad A^{\mathsf{T}}A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$$

Eigenvalues of $A^{\mathsf{T}}A$: $\lambda = 45,\ 5$ with eigenvectors $\frac{1}{\sqrt{2}}(1,1)$ and $\frac{1}{\sqrt{2}}(1,-1)$. Hence

$$\sigma_1 = \sqrt{45} = 3\sqrt{5} \approx 6.708, \qquad \sigma_2 = \sqrt{5} \approx 2.236.$$

$$\boldsymbol{u}_1 = \frac{A\boldsymbol{v}_1}{\sigma_1} = \frac{1}{3\sqrt{5}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ 9 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad \boldsymbol{u}_2 = \frac{A\boldsymbol{v}_2}{\sigma_2} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Worked example

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \qquad \Longrightarrow \qquad A^{\mathsf{T}}A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$$

Eigenvalues of $A^{\mathsf{T}}A$: $\lambda = 45,\ 5$ with eigenvectors $\frac{1}{\sqrt{2}}(1,1)$ and $\frac{1}{\sqrt{2}}(1,-1)$. Hence

$$\sigma_1 = \sqrt{45} = 3\sqrt{5} \approx 6.708, \qquad \sigma_2 = \sqrt{5} \approx 2.236.$$

$$\boldsymbol{u}_1 = \frac{A\boldsymbol{v}_1}{\sigma_1} = \frac{1}{3\sqrt{5}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 3 \\ 9 \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \qquad \boldsymbol{u}_2 = \frac{A\boldsymbol{v}_2}{\sigma_2} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Note A is not symmetric: its *eigenvalues* are 3 and 5, quite different from its singular values.

Singular values vs. eigenvalues

	Eigenvalues	Singular values
exist for	square matrices	all matrices
may be	complex, negative	always real, ≥ 0
bases	one, possibly non-orthogonal	two, always orthonormal
always exist?	no (defective matrices)	yes
perturbation-stable?	no	yes, by Weyl's inequality

When do they coincide?

If A is symmetric positive semidefinite, $\sigma_k = \lambda_k$ and $\boldsymbol{u}_k = \boldsymbol{v}_k$.
If A is symmetric but indefinite, $\sigma_k = |\lambda_k|$ and $\boldsymbol{u}_k = \pm \boldsymbol{v}_k$.

Our running $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ is SPD, so its SVD *is* its eigendecomposition: $\sigma = 4, 2$.

SVD never fails

- Full SVD always produces a complete orthonormal basis of $\mathbb{R}^n$ for V and a complete orthonormal basis of $\mathbb{R}^m$ for U

SVD never fails

- Full SVD always produces a complete orthonormal basis of $\mathbb{R}^n$ for V and a complete orthonormal basis of $\mathbb{R}^m$ for U
- For every matrix: rectangular, rank-deficient, even the zero matrix.

SVD never fails

- Full SVD always produces a complete orthonormal basis of $\mathbb{R}^n$ for V and a complete orthonormal basis of $\mathbb{R}^m$ for U
- For every matrix: rectangular, rank-deficient, even the zero matrix.
- Rank affects only how many singular values are nonzero, never whether the bases exist.

SVD never fails

Example: $\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$

Rank and the four fundamental subspaces

Let r = number of nonzero singular values. Then $\operatorname{rank} A = r$, and

Subspace	Basis	Dimension
$\operatorname{Col}(A) \subseteq \mathbb{R}^m$	$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$	r
$\operatorname{Null}(A^{\mathsf{T}}) \subseteq \mathbb{R}^m$	$\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$	$m - r$
$\operatorname{Row}(A) \subseteq \mathbb{R}^n$	$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$	r
$\operatorname{Null}(A) \subseteq \mathbb{R}^n$	$\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$	$n - r$

Rank and the four fundamental subspaces

Let r = number of nonzero singular values. Then $\operatorname{rank} A = r$, and

Subspace	Basis	Dimension
$\mathrm{Col}(A) \subseteq \mathbb{R}^m$	$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$	r
$\mathrm{Null}(A^{\mathsf{T}}) \subseteq \mathbb{R}^m$	$\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$	$m - r$
$\mathrm{Row}(A) \subseteq \mathbb{R}^n$	$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$	r
$\mathrm{Null}(A) \subseteq \mathbb{R}^n$	$\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$	$n - r$

The SVD hands you *orthonormal* bases for all four subspaces at once — something Gaussian elimination cannot do.

Norms and conditioning

Spectral (operator 2-) norm

$$\|A\|_2 = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \sigma_1$$

The longest semi-axis of the image ellipse. Attained at $\boldsymbol{x} = \boldsymbol{v}_1$.

Frobenius norm

$$\|A\|_F = \Big(\sum_{i,j} a_{ij}^2\Big)^{1/2} = \Big(\sum_{k=1}^{r} \sigma_k^2\Big)^{1/2}$$

Because $\|A\|_F^2 = \operatorname{tr}(A^{\mathsf{T}}A) = \sum \sigma_k^2$, and the trace is basis-independent.

Condition number

$$\kappa_2(A) = \frac{\sigma_1}{\sigma_r}$$

How much a relative error in $\boldsymbol{b}$ can be amplified when solving $A\boldsymbol{x} = \boldsymbol{b}$. Large κ = the ellipse is a thin sliver.

Distance to singularity

$$\min_{\operatorname{rank} B < r} \|A - B\|_2 = \sigma_r$$

$\sigma_{\min}$ measures how far A is from being rank deficient — a far better test than $\det A$.

The pseudoinverse

If $A = U\Sigma V^{\mathsf{T}}$, define Σ^{+} by inverting the nonzero entries and transposing the shape. Then

$$A^{+} = V\Sigma^{+}U^{\mathsf{T}} = \sum_{k=1}^{r} \frac{1}{\sigma_k} \boldsymbol{v}_k \boldsymbol{u}_k^{\mathsf{T}}.$$

- If A is invertible, $A^{+} = A^{-1}$.
- Otherwise $\boldsymbol{x}^{\star} = A^{+}\boldsymbol{b}$ is the *minimum-norm least-squares solution*: among all $\boldsymbol{x}$ minimising $\|A\boldsymbol{x} - \boldsymbol{b}\|_2$, it is the shortest.

Truncated SVD as regularisation

- Tiny σ_k give enormous $1/\sigma_k$ and amplify noise.

Truncated SVD as regularisation

- Tiny σ_k give enormous $1/\sigma_k$ and amplify noise.
- Dropping terms with $\sigma_k < \varepsilon$ gives a stable, slightly biased solution.

Truncated SVD as regularisation

- Tiny σ_k give enormous $1/\sigma_k$ and amplify noise.
- Dropping terms with $\sigma_k < \varepsilon$ gives a stable, slightly biased solution.
- This is the same move as Tikhonov / ridge regularisation, which replaces $1/\sigma_k$ by $\sigma_k/(\sigma_k^2 + \alpha)$.

Truncating the expansion

Order the terms by size and keep the first k:

$$A = \underbrace{\sum_{j=1}^{k} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^{\mathsf{T}}}_{=:A_k} + \underbrace{\sum_{j=k+1}^{r} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^{\mathsf{T}}}_{\text{discarded}}$$

Because the $\boldsymbol{u}_j$ and the $\boldsymbol{v}_j$ are orthonormal, the error is clean:

$$\|A - A_k\|_2 = \sigma_{k+1}, \qquad \|A - A_k\|_F = \Big(\sum_{j=k+1}^{r} \sigma_j^2 \Big)^{1/2}.$$

Truncating the expansion

Order the terms by size and keep the first k:

$$A = \underbrace{\sum_{j=1}^{k} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^{\mathsf{T}}}_{=:A_k} + \underbrace{\sum_{j=k+1}^{r} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^{\mathsf{T}}}_{\text{discarded}}$$

Because the $\boldsymbol{u}_j$ and the $\boldsymbol{v}_j$ are orthonormal, the error is clean:

$$\|A - A_k\|_2 = \sigma_{k+1}, \qquad \|A - A_k\|_F = \Big(\sum_{j=k+1}^{r} \sigma_j^2 \Big)^{1/2}.$$

The singular values tell you *in advance* how good a rank-k approximation can be — before you compute it.

Next

- PCA

Rough