Matrix Theory

25 Norms

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

Definition

A **norm** on $\mathbb{R}^n$ is a map $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ with

1. $\|x\| \geq 0$, and $\|x\| = 0 \iff x = 0$ (positivity)
2. $\|\alpha x\| = |\alpha| \ \|x\|$ (homogeneity)
3. $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)

Everything else follows from these three.

The p-norms

For $p \geq 1$,

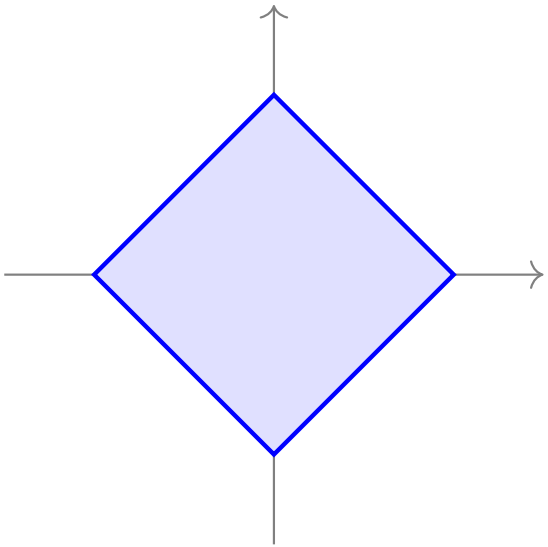
$$\|x\|_p \;=\; \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p}$$

Three cases do almost all the work:

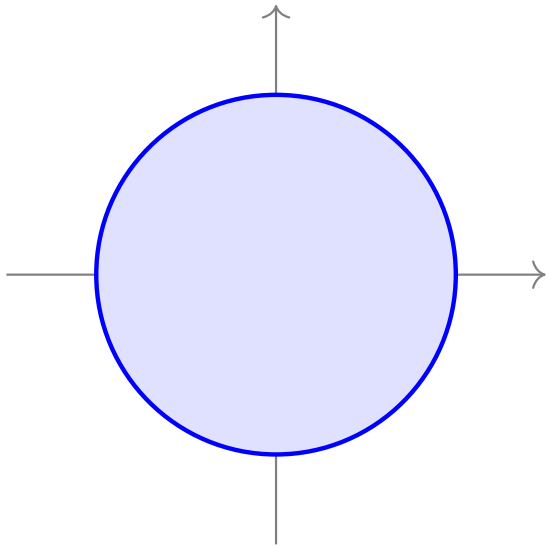
$\|x\|_1 = \sum_i \|x_i\|$	"taxicab"	sparsity, robustness
$\|x\|_2 = \sqrt{\sum_i x_i^2}$	Euclidean	geometry, least squares
$\|x\|_\infty = \max_i \|x_i\|$	max / Chebyshev	worst-case error

Note: $p < 1$ fails the triangle inequality. The "ℓ_0 norm" is not a norm.

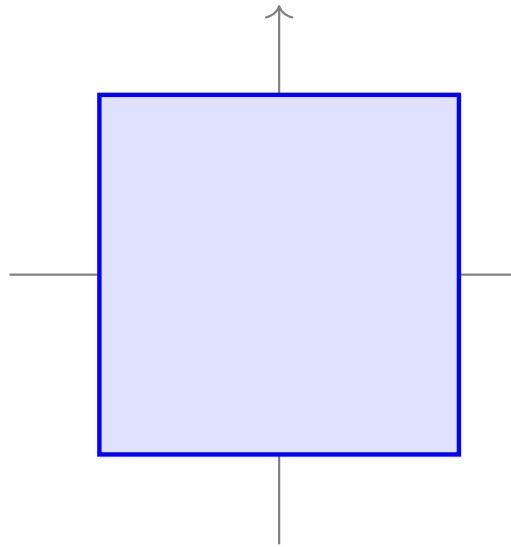
Unit balls $\{x : \|x\| \leq 1\}$ in $\mathbb{R}^2$



$\|x\|_1 \leq 1$



$\|x\|_2 \leq 1$



$\|x\|_\infty \leq 1$

All convex, all symmetric about the origin. The ball *is* the norm.

All norms on $\mathbb{R}^n$ are equivalent

For any two norms $\|\cdot\|_a$, $\|\cdot\|_b$ there exist $0 < c \leq C$ with

$$c\,\|x\|_b \;\leq\; \|x\|_a \;\leq\; C\,\|x\|_b \qquad \text{for all } x.$$

Concretely:

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\,\|x\|_\infty\,, \qquad \|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2\,.$$

Consequence. Convergence does not depend on the choice of norm.

Caveat. The constants grow with n — in high dimensions the choice still matters a great deal in practice.

The 2-norm is special: it comes from an inner product

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

$$\|x\|_2^2 \;=\; x^{\mathsf{T}}x \;=\; \langle x, x\rangle$$

This buys us geometry that the other p-norms do not have:

- angles: $\cos\theta = \dfrac{x^{\mathsf{T}}y}{\|x\|_2\,\|y\|_2}$
- Cauchy–Schwarz: $\left|x^{\mathsf{T}}y\right| \le \|x\|_2\,\|y\|_2$
- Pythagoras: $x \perp y \implies \|x+y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$
- invariance: $\|Qx\|_2 = \|x\|_2$ for orthogonal Q

The last one is why the SVD lives naturally in the 2-norm.

Two ways to measure a matrix

As a list of numbers

Treat $A \in \mathbb{R}^{m \times n}$ as a long vector and apply a vector norm.

$\rightarrow$ the Frobenius norm.

As an operator

Ask how much A can stretch a vector.

$\rightarrow$ the induced norms.

Both are useful. They answer different questions.

Induced (operator) norms

Given a vector norm $\|\cdot\|$,

$$\|A\| \;=\; \max_{x\neq 0} \frac{\|Ax\|}{\|x\|} \;=\; \max_{\|x\|=1} \|Ax\|$$

Immediately, for every x:

$$\boxed{\|Ax\| \;\leq\; \|A\|\;\|x\|}$$

This single inequality is the workhorse — it is how norms pass through products.

The bound is *tight*: some x attains it.

The three induced norms you will meet

$$\|A\|_1 = \max_j \sum_i |a_{ij}|$$

(largest column sum)

$$\|A\|_\infty = \max_i \sum_j |a_{ij}|$$

(largest row sum)

$$\|A\|_2 \;=\; \sigma_{\max}(A)$$

The first two are cheap to read off. The third requires an SVD — but it is the one with the geometry.

Why $\|A\|_2 = \sigma_{\max}$

Write $A = U\Sigma V^{\mathsf{T}}$ and let $\|x\|_2 = 1$. Put $y = V^{\mathsf{T}}x$, so $\|y\|_2 = 1$ too.

$$\|Ax\|_2^2 \;=\; \|U\Sigma y\|_2^2 \;=\; \|\Sigma y\|_2^2 \;=\; \sum_i \sigma_i^2 y_i^2 \;\le\; \sigma_1^2 \sum_i y_i^2 \;=\; \sigma_1^2$$

with equality at $y = e_1$, i.e. $x = v_1$.

Picture. A maps the unit sphere to an ellipsoid with semi-axes $\sigma_1 \ge \cdots \ge \sigma_n$. The norm is the longest axis.

Submultiplicativity

Apply the workhorse inequality twice:

$$\|ABx\| \leq \|A\| \, \|Bx\| \leq \|A\| \, \|B\| \, \|x\| \quad \Longrightarrow \quad \|AB\| \leq \|A\| \, \|B\|$$

Two consequences worth remembering:

- $\|I\| = 1$ for any induced norm
- hence $\kappa(A) = \|A\| \, \|A^{-1}\| \geq \|AA^{-1}\| = 1$

No matrix is better conditioned than perfectly conditioned.

The Frobenius norm

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\operatorname{trace}(A^{\mathsf{T}}A)} = \sqrt{\sum_i \sigma_i^2}$$

- submultiplicative, and $\|Ax\|_2 \le \|A\|_F \, \|x\|_2$
- but **not** an induced norm: $\|I_n\|_F = \sqrt{n} \neq 1$
- cheap to compute — no SVD needed

Sandwich: $\|A\|_2 \le \|A\|_F \le \sqrt{r}\, \|A\|_2, \quad r = \operatorname{rank} A.$

Unitary invariance

A norm is **unitarily invariant** if

$$\|QAZ\| = \|A\| \qquad \text{for all orthogonal } Q, Z.$$

Both $\|\cdot\|_2$ and $\|\cdot\|_F$ have this property. $\|\cdot\|_1$ and $\|\cdot\|_\infty$ do not.

Why it matters. Such a norm cannot see U or V — only Σ. So it must be a function of the singular values alone.

Matrix norms are vector norms of σ

name	formula	vector norm of σ
spectral	$\|A\|_2 = \sigma_1$	$\|\sigma\|_\infty$
Frobenius	$\|A\|_F = \sqrt{\sum \sigma_i^2}$	$\|\sigma\|_2$
nuclear	$\|A\|_* = \sum \sigma_i$	$\|\sigma\|_1$

The nuclear norm is the convex relaxation of rank — the matrix analogue of ℓ_1 promoting sparsity.

Eckart–Young

Let $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^{\mathsf{T}}$ be the truncated SVD. Then for every B with $\operatorname{rank} B \leq k$:

$$\|A - A_k\|_2 = \sigma_{k+1} \;\leq\; \|A - B\|_2$$

$$\|A - A_k\|_F = \sqrt{\textstyle\sum_{i>k} \sigma_i^2} \;\leq\; \|A - B\|_F$$

The truncated SVD is the *best* rank-k approximation — in both norms, and in fact in every unitarily invariant norm.

This is the theorem that justifies PCA.

Back to conditioning

From $Ax = b$ and $A(x + \delta x) = b + \delta b$:

$$\frac{\|\delta x\|}{\|x\|} \leq \underbrace{\|A\| \left\|A^{-1}\right\|}_{\kappa(A)} \cdot \frac{\|\delta b\|}{\|b\|}$$

In the 2-norm this is a statement about the ellipsoid:

$$\kappa_2(A) = \frac{\sigma_{\max}}{\sigma_{\min}} = \frac{\text{largest stretch}}{\text{smallest stretch}}$$

$\sigma_{\min}$ is the distance from A to the nearest singular matrix. So κ_2 measures nearness to singularity — continuously.

Perturbation of singular values

For any perturbation E,

$$|\sigma_i(A+E) - \sigma_i(A)| \;\leq\; \|E\|_2 \qquad \text{for every } i.$$

No amplification factor. Compare with eigenvalues of a non-symmetric matrix, where Bauer–Fike gives a factor of $\kappa(S)$.

This is why numerical rank is determined from the SVD: count $\sigma_i > \text{tol}$, and the count is stable.

Summary

- A norm is positivity + homogeneity + triangle inequality.
- Induced norms measure stretching; $\|Ax\| \le \|A\|\,\|x\|$ does the work.
- $\|A\|_2 = \sigma_1, \quad \|A\|_F = \|\sigma\|_2, \quad \|A\|_* = \|\sigma\|_1.$
- Unitarily invariant norms see only the singular values.
- Eckart–Young: truncating the SVD is optimal.
- $\kappa_2 = \sigma_{\max}/\sigma_{\min}$ — the amplification factor for errors.

Exercises

1. Show $\|A\|_\infty = \left\|A^{\mathsf{T}}\right\|_1$.
2. Find a 2×2 matrix with $\|A\|_2 = 1$ but $\|A\|_F = \sqrt{2}$.
3. Prove $\|A\|_2 \leq \sqrt{\|A\|_1 \, \|A\|_\infty}$.
4. For orthogonal Q, verify $\kappa_2(Q) = 1$ directly from the definition.
5. Let A have rank 1. Show $\|A\|_2 = \|A\|_F = \|A\|_*$.