Primer on matrix norms

Arash A. Amini

April 13, 2021

These notes provide a short introduction to common matrix norms. (This is a rough draft. There are most likely mistakes.)

Terminology: PSD = Positive Semi-Definite matrices.

1 Vector ℓ_p norms

The ℓ_p vector norms are defined as

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

for any $x \in \mathbb{R}^n$ and $p \in [1, \infty)$. The ℓ_∞ norm is defined as $\|x\|_\infty = \max_j |x_j|$.

Exercise 1.1. *Show that* $\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$.

Exercise 1.2. *Show that* $\|\cdot\|_p$ *is a (proper) norm for* $p \in [1, \infty]$.

For $p \in (0, 1)$, $\|\cdot\|_p$ defines a quasi-norm, i.e., it fails the triangle inequality, but $x \mapsto \|x\|_p^p$ is subadditive:

$$\|x + y\|_p^p \le \|x\|_p^p + \|y\|_p^p.$$

It follows that $x \mapsto \|x\|_p^p$ defines a so-called F-norm. We denote the unit ball of ℓ_p as

$$\mathbb{B}_p := \{x : \|x\|_p \le 1\ \}.$$

If we want to emphasize the dimension we write $\mathbb{B}_p^n := \{x \in \mathbb{R}^n : \|x\|_p \le 1\}$. The space $\mathbb{R}^n$ equipped with ℓ_p norm is usually written as $\ell_p^n = (\mathbb{R}^n, \|\cdot\|_p)$. This is a finite-dimensional Banach space.

Exercise 1.3. *Show that* $\mathbb{B}_p$ *is a convex set for* $p \in [1, \infty]$.

1.1 Duality

Let $p, p' \in [1, \infty]$ be dual exponents, i.e. $1/p + 1/p' = 1$. Then, Hölder inequality gives

$$|\langle x, y\rangle| \le \|x\|_p \, \|y\|_{p'}, \quad \forall x, y \in \mathbb{R}^n.$$

Since for every x, the inequality is achieved by some y, it follows that $\|\cdot\|_p$ and $\|\cdot\|_{p'}$ are dual norms, e.g.:

$$\|x\|_{p'} = \max_{\|y\|_p \le 1} \langle x, y\rangle = \max_{y \in \mathbb{B}_p} \langle x, y\rangle = \max_{y \in \mathbb{B}_p} |\langle x, y\rangle|. \tag{1}$$

1.2 Interpolation

Let $p_0, p_1 \in (0, \infty]$, and for $\theta \in (0,1)$ let $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$. Then,

$$\|x\|_{p_\theta} \le \|x\|_{p_0}^{1-\theta} \|x\|_{p_1}^{\theta} \quad \text{(log-convexity)}. \tag{2}$$

This is saying that $1/p \mapsto \log \|x\|_p$ is convex over $[0, \infty]$ for every fixed x.

2 Matrix norms

2.1 Operator norms

Consider a matrix $\mathbb{R}^{m\times n}$. We can view the matrix as an operator $A : (\mathbb{R}^n, \|\cdot\|_p) \to (\mathbb{R}^m, \|\cdot\|_q)$. The corresponding operator norm is

$$\|A\|_{p,q} := \|A\|_{p\to q} := \max_{\|x\|_p \,\le\, 1} \|Ax\|_q. \tag{3}$$

It measures the radius of the smallest ℓ_q ball (centered at origin) that contains the image of ℓ_p ball under A, that is:

Exercise 2.1. *Show that* $\|A\|_{p\to q} = \inf\{t \ge 0 : A\,\mathbb{B}_p^n \subset t\,\mathbb{B}_q^m\}$.

Exercise 2.2. *Show the following alternative representations of the operator norm:*

$$\|A\|_{p\to q} = \max_{\|x\|_p \,=\, 1} \|Ax\|_q = \max_{x\neq 0} \frac{\|Ax\|_q}{\|x\|_p}.$$

The definition of the operator norm has the following basic consequences:

Exercise 2.3. *Show that* $\|Ax\|_q \le \|A\|_{p,q}\|x\|_p$ *for all* $x \in \mathbb{R}^n$ *and all* $p, q \in (0, \infty]$.

Exercise 2.4. *Prove the* multiplicative *property for two conformal matrices A and B:*

$$\|AB\|_{p,q} \le \|A\|_{p,r}\|B\|_{r,q}. \tag{4}$$

One usually writes $\|A\|_p = \|A\|_{p,p}$. The special case $\|A\|_2 = \|A\|_{2,2}$ is often called "the" operator norm. Other notations for "the" operator norm are

$$\|A\|_{\text{op}} = \|A\| = \|A\|_2 := \max_{\|x\|_2 \,=\, 1} \|Ax\|_2 = \max_{x,y\colon\, \|x\|_2 \,=\, \|y\|_2 \,=\, 1} \langle Ax, y\rangle$$

For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, the operator norm has the following alternative characterization:

$$\|A\|_2 = \max_{\|x\|_2=1} |\langle Ax, x\rangle|. \tag{5}$$

Exercise 2.5. *Show (5), using the eigenvalue decomposition of A (Section 3).*

Exercise 2.6. *Argue that the absolute value in (5) cannot be dropped in general.*

Exercise 2.7. *Give a counterexample to show that (5) does not necessarily hold for nonsymmetric matrices.*

Exercise 2.8. *Show that* ℓ_1 *and* ℓ_∞ *operator norms are given by*

$$\|A\|_1 = \|A\|_{1\to 1} = \max_j \sum_{i=1}^m |a_{ij}|, \qquad \|A\|_\infty = \|A\|_{\infty\to\infty} = \max_i \sum_{j=1}^n |a_{ij}|$$

that is, $\|A\|_1$ *is the maximum absolute column sum, and* $\|A\|_\infty$ *is the maximum absolute row sum.*

Exercise 2.9. *Show that for any* $p \in [1, \infty]$,

$$\|A\|_{p\to\infty} = \max_i \|A_{i*}\|_{p'}$$

where p' *is the dual exponent to* p *and* A_{i*} *is the* i*th row of* A.

2.1.1 Interpolation

The following theorem interpolates between operator norms:

Theorem 1 (Riesz–Thorin)**.** *For* $p_i, q_i \in [1, \infty]$, $i = 0, 1$, *and* $\theta \in (0, 1)$,

$$\|A\|_{p_\theta,\, q_\theta} \le \|A\|_{p_0,\, q_0}^{1-\theta} \|A\|_{p_1,\, q_1}^{\theta}.$$

where $p_\theta = (1-\theta)p_0 + \theta p_1$ *and* $q_\theta = (1-\theta)q_0 + \theta q_1$.

A special case of Riesz–Thorin is the following:

$$\|A\|_2 \le \sqrt{\|A\|_1 \|A\|_\infty}, \quad \text{and if } A \text{ is symmetric } \|A\|_2 \le \|A\|_1 = \|A\|_\infty.$$

This special case is very useful in bounding the operator norm by bounding row sums and column sums of A. An alternative equivalent statement is this: Assume that all the row sums of A are bounded by a and all the column sums by b. Then, $\|A\|_{\mathrm{op}} \le \sqrt{ab}$.

2.1.2 Duality

Here A^* is the adjoint of A, which for us is the same as A^T. Recall that p' is the Hölder exponent dual to p. We have

$$\begin{aligned}\|A\|_{p,q} = \max_{x \,\in\, \mathbb{B}_p} \|Ax\|_q &= \max_{x \,\in\, \mathbb{B}_p,\, y \,\in\, \mathbb{B}_{q'}} \langle Ax, y\rangle \\ &\underset{(a)}{=} \max_{x \,\in\, \mathbb{B}_p,\, y \,\in\, \mathbb{B}_{q'}} \langle x, A^* y\rangle = \max_{y \,\in\, \mathbb{B}_{q'}} \|A^* y\|_{p'} = \|A^*\|_{q',p'}.\end{aligned}$$

Equality (a) uses the defining property of the adjoint A^*. In can be verified directly in our case using $A^* = A^T$ and $\langle x, y\rangle = x^T y$.

2.2 Frobenius norm

The Frobenius norm a matrix $A \in \mathbb{R}^{n\times m}$ is defined as

$$\|A\|_F := \Big(\sum_{i,j} A_{ij}^2\Big)^{1/2}. \tag{6}$$

Let us write $\mathrm{vec}(A)$ for the vector obtained by concatenating the columns of A. For $A \in \mathbb{R}^{n\times m}$ we have $\mathrm{vec}(A) \in \mathbb{R}^{mn}$. We note that $\|A\|_F = \|\mathrm{vec}(A)\|_2$.

By viewing matrices as vectors, we can go further and extend the usual Euclidean inner product to matrices, by defining

$$\langle A, B\rangle := \langle \mathrm{vec}(A), \mathrm{vec}(B)\rangle = \mathrm{vec}(A)^T \mathrm{vec}(B). \tag{7}$$

Exercise 2.10. *Show that $\langle A, B\rangle = \mathrm{tr}(A^T B)$.*

Note that $\|A\|_F = \sqrt{\langle A, A\rangle}$. The space of $n \times m$ matrices equipped the above inner product is a Hilbert space, with norm being the Frobenius norm, which is also referred to as the Hilbert–Schmidt norm.

Exercise 2.11. *Let $A \in \mathbb{R}^{n\times m}$ and $\{e_i\}$ be any basis for $\mathbb{R}^m$. Show that $\|A\|_F^2 = \sum_i \|Ae_i\|_2^2$.*

2.3 Unitarily-invariant norms

A matrix norm is unitarily invariant if $\|A\| = \|UAW\|$ for unitary (or orthogonal) matrices U and W.

Let $A = U\Sigma V^T$ be a SVD of A, where $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n)$ contains the singular values of A, nonnegative by definition. We order them as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. We also write $\sigma_i = \sigma_i(A)$ to emphasize that we are talking about the singular values of A. Let $\sigma = (\sigma_i)$ and $\sigma(A) = (\sigma_i(A))$ denote the vector of singular values.

1. The operator norm $\|\cdot\|_{\mathrm{op}} = \|\cdot\|_2$ is unitarily invariant, a direct consequence of unitary invariance of the ℓ_2 norm. Hence, $\|A\|_{\mathrm{op}} = \|\Sigma\|_{\mathrm{op}} = \max\{\sum_i \sigma_i x_i^2 : \sum_i x_i^2 = 1\} = \|\sigma\|_\infty$. That is, $\|A\|_{\mathrm{op}} = \sigma_1(A)$.

2. The Frobenius norm defined as $\|A\|_F = (\sum_{ij} A_{ij}^2)^{1/2}$ is unitarily invariant. This can be seen by writing $\|A\|_F^2 = \langle A, A\rangle = \mathrm{tr}(A^T A) = \mathrm{tr}(\Sigma^2) = \|\sigma\|_2^2$ using invariance of trace under circular permutations. That is, $\|A\|_F = (\sum_i \sigma_i^2(A))^{1/2}$.

3. The nuclear norm, also known as the trace norm, defined by $\|A\|_* := \|\sigma\|_1 = \sum_{i=1}^n \sigma_i(A)$ is clearly unitarily invariant.

Using the relation between ℓ_∞, ℓ_2 and ℓ_1 norm we have

$$\|A\|_2 \leq \|A\|_F \leq \|A\|_*$$

In general, one defines the Schatten-p matrix norms as $\|A\|_{S_p} := \|\sigma(A)\|_p$. The cases $p = 1, 2, \infty$ correspond to the nuclear, Frobenius and operator norms, respectively.

2.4 Inequalities

Exercise 2.12. *Show that for any two matrices $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{r \times n}$:*

$$\|AB\|_F \leq \|A\|_{op} \|B\|_F \tag{8}$$

From (8) deduce that Frobenious norm is multiplicative: $\|AB\|_F \leq \|A\|_F \|B\|_F$.

Exercise 2.13. *Show that for two PSD matrices A and B,*

$$\operatorname{tr}(AB) \leq \operatorname{tr}(A) \|B\|_{op}.$$

Give a counterexample to show that the PSD assumption for both A and B cannot be dropped.

3 Matrix decompositions

Every symmetric matrix has the following eigenvalue decomposition (EVD); it is also referred to as the spectral decomposition.

Theorem 2 (EVD)**.** *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then, there is an orthogonal matrix $U = [u_1 \mid u_2 \mid \cdots \mid u_n] \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ such that*

$$A = U \Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T.$$

4 Hints for exercises

- Exercise 1.2: For the triangle inequality use $|x_j + y_j|^p \leq |x_j||x_j + y_j|^{p-1} + |y_j||x_j + y_j|^{p-1}$ and the Hölder inequality. The triangle inequality for ℓ_p norms is also called the Minkowski inequality.
- Exercise 1.3: It is a consequence of $\|\cdot\|_p$ being a norm.
- Exercise 2.6: Consider a negative definite matrix.
- Exercise 2.12: Let $B = [b_1 \mid b_2 \mid \cdots \mid b_n]$ be the column decomposition of B.
- Exercise 2.13: Reduce to the case where B is diagonal.