WIKIPEDIA

Matrix norm

In mathematics, a **matrix norm** is a vector norm in a vector space whose elements (vectors) are matrices (of given dimensions).

Contents

Definition

In what follows, K will denote a field of either real or complex numbers.

Let $K^{m\times n}$ denote the vector space of all matrices of size $m \times n$ (with m rows and n columns) with entries in the field K.

A matrix norm is a norm on the vector space $K^{m\times n}$. Thus, the matrix norm is a function $\|\cdot\| : K^{m\times n} \to \mathbb{R}$ that must satisfy the following properties:

For all scalars $\alpha \in K$ and for all matrices $A, B \in K^{m\times n}$,

- $\|\alpha A\| = |\alpha| \|A\|$ (being *absolutely homogeneous*)
- $\|A + B\| \leq \|A\| + \|B\|$ (being *sub-additive* or satisfying the *triangle inequality*)
- $\|A\| \geq 0$ (being *positive-valued*)
- $\|A\| = 0$ iff $A = 0_{m,n}$ (being *definite*)

Additionally, in the case of square matrices (thus, $m = n$), some (but not all) matrix norms satisfy the following condition, which is related to the fact that matrices are more than just vectors:

- $\|AB\| \leq \|A\| \|B\|$ for all matrices A and B in $K^{n\times n}$.

A matrix norm that satisfies this additional property is called a **sub-multiplicative norm** (in some books, the terminology *matrix norm* is used only for those norms which are sub-multiplicative). The set of all $n \times n$ matrices, together with such a sub-multiplicative norm, is an example of a Banach algebra.

The definition of sub-multiplicativity is sometimes extended to non-square matrices, for instance in the case of the induced p-norm, where for $A \in K^{m\times n}$ and $B \in K^{n\times k}$ holds that $\|AB\|_q \leq \|A\|_p \|B\|_q$. Here $\|\cdot\|_p$ and $\|\cdot\|_q$ are the norms induced from K^n and K^k, respectively, and $p,q \geq 1$.

There are three types of matrix norms which will be discussed below:

- Matrix norms induced by vector norms,
- Entrywise matrix norms, and
- Schatten norms.

Matrix norms induced by vector norms

Suppose a vector norm $\|\cdot\|$ on K^m is given. Any $m \times n$ matrix A induces a linear operator from K^n to K^m with respect to the standard basis, and one defines the corresponding *induced norm* or *operator norm* on the space $K^{m\times n}$ of all $m \times n$ matrices as follows:

$$\begin{aligned}\|A\| &= \sup\{\|Ax\| : x \in K^n \text{ with } \|x\| = 1\}\\ &= \sup\left\{\frac{\|Ax\|}{\|x\|} : x \in K^n \text{ with } x \neq 0\right\}.\end{aligned}$$

In particular, if the p-norm for vectors ($1 \leq p \leq \infty$) is used for both spaces K^n and K^m, then the corresponding induced operator norm is:

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

These induced norms are different from the "entrywise" p-norms and the Schatten p-norms for matrices treated below, which are also usually denoted by $\|A\|_p$.

> **Note:** We have described above the *induced operator norm* when the same vector norm was used in the "departure space" K^n and the "arrival space" K^m of the operator $A \in K^{m\times n}$. This is not a necessary restriction. More generally, given a norm $\|\cdot\|_\alpha$ on K^n, and a norm $\|\cdot\|_\beta$ on K^m, one can define a matrix norm on $K^{m\times n}$ induced by these norms:
>
> $$\|A\|_{\alpha,\beta} = \max_{x\neq 0} \frac{\|Ax\|_\beta}{\|x\|_\alpha}.$$
>
> The matrix norm $\|A\|_{\alpha,\beta}$ is sometimes called a subordinate norm. Subordinate norms are consistent with the norms that induce them, giving
>
> $$\|Ax\|_\beta \leq \|A\|_{\alpha,\beta}\|x\|_\alpha.$$

Any induced operator norm is a sub-multiplicative matrix norm: $\|AB\| \leq \|A\|\|B\|$; this follows from

$$\|ABx\| \leq \|A\|\|Bx\| \leq \|A\|\|B\|\|x\|$$

and

$$\max_{\|x\|=1} \|ABx\| = \|AB\|.$$

Moreover, any induced norm satisfies the inequality

$$\|A^r\|^{1/r} \geq \rho(A), \qquad (1)$$

where ρ(A) is the spectral radius of A. For symmetric or hermitian A, we have equality in (1) for the 2-norm, since in this case the 2-norm *is* precisely the spectral radius of A. For an arbitrary matrix, we may not have equality for any norm; a counterexample being given by

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

which has vanishing spectral radius. In any case, for square matrices we have the spectral radius formula:

$$\lim_{r\to\infty} \|A^r\|^{1/r} = \rho(A).$$

Special cases

In the special cases of $p = 1, 2, \infty,$ the induced matrix norms can be computed or estimated by

$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|,$$

which is simply the maximum absolute column sum of the matrix;

$$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|,$$

which is simply the maximum absolute row sum of the matrix;

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

where $\sigma_{\max}(A)$ represents the largest singular value of matrix A. There is an important inequality for the case $p = 2$:

$$\|A\|_2 = \sigma_{\max}(A) \le \|A\|_{\mathrm{F}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}},$$

where $\|A\|_{\mathrm{F}}$ is the Frobenius norm. Equality holds if and only if the matrix A is a rank-one matrix or a zero matrix. This inequality can be derived from the fact that the trace of a matrix is equal to the sum of its eigenvalues.

When $p = 2$ we have an equivalent definition for $\|A\|_2$ as $\sup\{x^T Ay : x, y \in K^n \text{ with } \|x\|_2 = \|y\|_2 = 1\}$. It can be shown to be equivalent to the above definitions using the Cauchy–Schwarz inequality.

For example, for

$$A = \begin{bmatrix} -3 & 5 & 7 \\ 2 & 6 & 4 \\ 0 & 2 & 8 \end{bmatrix},$$

we have

$$\|A\|_1 = \max(|-3| + 2 + 0; 5 + 6 + 2; 7 + 4 + 8) = \max(5, 13, 19) = 19,$$
$$\|A\|_\infty = \max(|-3| + 5 + 7; 2 + 6 + 4; 0 + 2 + 8) = \max(15, 12, 10) = 15.$$

In the special case of $p = 2$ (the Euclidean norm or ℓ_2-norm for vectors), the induced matrix norm is the *spectral norm*. The spectral norm of a matrix A is the largest singular value of A i.e. the square root of the largest eigenvalue of the matrix A^*A where A^* denotes the conjugate transpose of A:[1]

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^*A)} = \sigma_{\max}(A).$$

"Entrywise" matrix norms

These norms treat an $m \times n$ matrix as a vector of size $m \cdot n$, and use one of the familiar vector norms. For example, using the p-norm for vectors, $p \ge 1$, we get:

$$\|A\|_p = \|\mathrm{vec}(A)\|_p = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^p\right)^{1/p}$$

This is a different norm from the induced p-norm (see above) and the Schatten p-norm (see below), but the notation is the same.

The special case $p = 2$ is the Frobenius norm, and $p = \infty$ yields the maximum norm.

$L_{2,1}$ and $L_{p,q}$ norms

Let $(a_1, \ldots, a_n)$ be the columns of matrix A. The $L_{2,1}$ norm[2] is the sum of the Euclidean norms of the columns of the matrix:

$$\|A\|_{2,1} = \sum_{j=1}^{n} \|a_j\|_2 = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}|^2 \right)^{\frac{1}{2}}$$

The $L_{2,1}$ norm as an error function is more robust since the error for each data point (a column) is not squared. It is used in robust data analysis and sparse coding.

The $L_{2,1}$ norm can be generalized to the $L_{p,q}$ norm, $p, q \geq 1$, defined by

$$\|A\|_{p,q} = \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}.$$

Frobenius norm

When $p = q = 2$ for the $L_{p,q}$ norm, it is called the **Frobenius norm** or the **Hilbert–Schmidt norm**, though the latter term is used more frequently in the context of operators on (possibly infinite-dimensional) Hilbert space. This norm can be defined in various ways:

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\operatorname{trace}(A^* A)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(A)},$$

where $\sigma_i(A)$ are the singular values of A. Recall that the trace function returns the sum of diagonal entries of a square matrix.

The Frobenius norm is an extension of the Euclidean norm to $K^{n \times n}$ and comes from the Frobenius inner product on the space of all matrices.

The Frobenius norm is sub-multiplicative and is very useful for numerical linear algebra. The sub-multiplicativity of Frobenius norm can be proved using Cauchy–Schwarz inequality.

Frobenius norm is often easier to compute than induced norms and has the useful property of being invariant under rotations and, more generally, under unitary operations, that is, $\|A\|_{\mathrm{F}} = \|AU\|_{\mathrm{F}} = \|UA\|_{\mathrm{F}}$ for any unitary matrix U. This property follows from the cyclic nature of the trace ($\operatorname{trace}(XYZ) = \operatorname{trace}(ZXY)$):

$$\|AU\|_{\mathrm{F}}^2 = \operatorname{trace}((AU)^* AU) = \operatorname{trace}(U^* A^* AU) = \operatorname{trace}(UU^* A^* A) = \operatorname{trace}(A^* A) = \|A\|_{\mathrm{F}}^2,$$

and analogously

$$\|UA\|_{\mathrm{F}}^2 = \operatorname{trace}((UA)^* UA) = \operatorname{trace}(A^* U^* UA) = \operatorname{trace}(A^* A) = \|A\|_{\mathrm{F}}^2,$$

where we have used the unitary nature of U (that is, $U^* U = UU^* = \mathbf{I}$).

It also satisfies

$$\|A^* A\|_{\mathrm{F}} = \|AA^*\|_{\mathrm{F}} \leq \|A\|_{\mathrm{F}}^2$$

and

$$\|A+B\|_{\mathrm{F}}^2 = \|A\|_{\mathrm{F}}^2 + \|B\|_{\mathrm{F}}^2 + 2\langle A, B\rangle_{\mathrm{F}},$$

where $\langle A, B\rangle_{\mathrm{F}}$ is the Frobenius inner product.

Max norm

The **max norm** is the elementwise norm with $p = q = \infty$:

$$\|A\|_{\max} = \max_{ij} |a_{ij}|.$$

This norm is not sub-multiplicative.

Note that in some literature (such as Communication complexity) an alternative definition of max-norm, also called the γ_2-norm, refers to the factorization norm:

$$\gamma_2(A) = \min_{U,V:A=UV^T} \|U\|_{2,\infty}\|V\|_{2,\infty} = \min_{U,V:A=UV^T} \max_{i,j} \|U_{i,:}\|_2 \|V_{j,:}\|_2$$

Schatten norms

The Schatten p-norms arise when applying the p-norm to the vector of singular values of a matrix. If the singular values of the $m \times n$ matrix A are denoted by σ_i, then the Schatten p-norm is defined by

$$\|A\|_p = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^p(A)\right)^{\frac{1}{p}}.$$

These norms again share the notation with the induced and entrywise p-norms, but they are different.

All Schatten norms are sub-multiplicative. They are also unitarily invariant, which means that $\|A\| = \|UAV\|$ for all matrices A and all unitary matrices U and V.

The most familiar cases are $p = 1, 2, \infty$. The case $p = 2$ yields the Frobenius norm, introduced before. The case $p = \infty$ yields the spectral norm, which is the operator norm induced by the vector 2-norm (see above). Finally, $p = 1$ yields the **nuclear norm** (also known as the *trace norm*, or the Ky Fan 'n'-norm[3]), defined as

$$\|A\|_* = \operatorname{trace}\left(\sqrt{A^*A}\right) = \sum_{i=1}^{\min\{m,n\}} \sigma_i(A),$$

where $\sqrt{A^*A}$ denotes a positive semidefinite matrix B such that $BB = A^*A$. More precisely, since A^*A is a positive semidefinite matrix, its square root is well-defined. The nuclear norm $\|A\|_*$ is a convex envelope of the rank function $\operatorname{rank}(A)$, so it is often used in mathematical optimization to search for low rank matrices.

Consistent norms

A matrix norm $\|\cdot\|$ on $K^{m\times n}$ is called *consistent* with a vector norm $\|\cdot\|_a$ on K^n and a vector norm $\|\cdot\|_b$ on K^m if:

$$\|Ax\|_b \leq \|A\|\|x\|_a$$

for all $A \in K^{m\times n}, x \in K^n$. All induced norms are consistent by definition.

Compatible norms

A matrix norm $\|\cdot\|$ on $K^{n\times n}$ is called *compatible* with a vector norm $\|\cdot\|_a$ on K^n if:

$$\|Ax\|_a \leq \|A\|\|x\|_a$$

for all $A \in K^{n\times n}, x \in K^n$. Induced norms are compatible with the inducing vector norm by definition.

Equivalence of norms

For any two matrix norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$, we have

$$r\|A\|_\alpha \leq \|A\|_\beta \leq s\|A\|_\alpha$$

for some positive numbers r and s, for all matrices $A \in K^{m\times n}$. In other words, all norms on $K^{m\times n}$ are *equivalent*; they induce the same topology on $K^{m\times n}$. This is true because the vector space $K^{m\times n}$ has the finite dimension $m \times n$.

Moreover, for every vector norm $\|\cdot\|$ on $\mathbb{R}^{n\times n}$, there exists a unique positive real number k such that $l\|\cdot\|$ is a sub-multiplicative matrix norm for every $l \geq k$.

A sub-multiplicative matrix norm $\|\cdot\|_\alpha$ is said to be *minimal* if there exists no other sub-multiplicative matrix norm $\|\cdot\|_\beta$ satisfying $\|\cdot\|_\beta < \|\cdot\|_\alpha$.

Examples of norm equivalence

Let $\|A\|_p$ once again refer to the norm induced by the vector p-norm (as above in the Induced Norm section).

For matrix $A \in \mathbb{R}^{m\times n}$ of rank r, the following inequalities hold:[4][5]

- $\|A\|_2 \leq \|A\|_F \leq \sqrt{r}\|A\|_2$
- $\|A\|_F \leq \|A\|_* \leq \sqrt{r}\|A\|_F$
- $\|A\|_{\max} \leq \|A\|_2 \leq \sqrt{mn}\|A\|_{\max}$
- $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\|A\|_\infty$
- $\frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1.$

Another useful inequality between matrix norms is

$$\|A\|_2 \leq \sqrt{\|A\|_1\|A\|_\infty},$$

which is a special case of Hölder's inequality.

Notes

1. Carl D. Meyer, Matrix Analysis and Applied Linear Algebra, §5.2, p.281, Society for Industrial & Applied Mathematics, June 2000.
2. Ding, Chris; Zhou, Ding; He, Xiaofeng; Zha, Hongyuan (June 2006). "R1-PCA: Rotational Invariant L1-norm Principal Component Analysis for Robust Subspace Factorization". *Proceedings of the 23rd International Conference on Machine Learning*. ICML '06. Pittsburgh, Pennsylvania, USA: ACM. pp. 281–288. doi:10.1145/1143844.1143880 (https://doi.org/10.1145%2F1143844.1143880). ISBN 1-59593-383-2.
3. Fan, Ky. (1951). "Maximum properties and inequalities for the eigenvalues of completely continuous operators" (http://www.pnas.org/content/37/11/760.citation). *Proceedings of the National Academy of Sciences of the United States of America*. **37** (11): 760–766. doi:10.1073/pnas.37.11.760 (https://doi.org/10.1073%2Fpnas.37.11.760). PMC 1063464 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1063464). PMID 16578416 (https://pubmed.ncbi.nlm.nih.gov/16578416).
4. Golub, Gene; Charles F. Van Loan (1996). Matrix Computations – Third Edition. Baltimore: The Johns Hopkins University Press, 56–57. ISBN 0-8018-5413-X.
5. Roger Horn and Charles Johnson. *Matrix Analysis,* Chapter 5, Cambridge University Press, 1985. ISBN 0-521-38632-2.

References

- James W. Demmel, Applied Numerical Linear Algebra, section 1.7, published by SIAM, 1997.

- Carl D. Meyer, Matrix Analysis and Applied Linear Algebra, published by SIAM, 2000. [1] (http://www.matrixanalysis.com)
- John Watrous, Theory of Quantum Information, 2.3 Norms of operators (https://web.archive.org/web/20160304053759/https://cs.uwaterloo.ca/~watrous/CS766/LectureNotes/02.pdf), lecture notes, University of Waterloo, 2011.
- Kendall Atkinson, An Introduction to Numerical Analysis, published by John Wiley & Sons, Inc 1989