Lecture 3. Norms

The essential notions of size and distance in a vector space are captured by norms. These are the yardsticks with which we measure approximations and convergence throughout numerical linear algebra.

Vector Norms

A *norm* is a function $\|\cdot\| : \mathbb{C}^m \to \mathbb{R}$ that assigns a real-valued length to each vector. In order to conform to a reasonable notion of length, a norm must satisfy the following three conditions. For all vectors x and y and for all scalars $\alpha \in \mathbb{C}$,

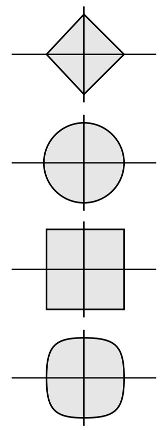
$$\begin{aligned} &(1) \quad \|x\| \geq 0, \text{ and } \|x\| = 0 \text{ only if } x = 0, \\ &(2) \quad \|x + y\| \leq \|x\| + \|y\|, \\ &(3) \quad \|\alpha x\| = |\alpha|\,\|x\|. \end{aligned} \tag{3.1}$$

In words, these conditions require that (1) the norm of a nonzero vector is positive, (2) the norm of a vector sum does not exceed the sum of the norms of its parts—the *triangle inequality*, and (3) scaling a vector scales its norm by the same amount.

In the last lecture, we used $\|\cdot\|$ to denote the Euclidean length function (the square root of the sum of the squares of the entries of a vector). However, the three conditions (3.1) allow for different notions of length, and at times it is useful to have this flexibility.

The most important class of vector norms, the p-norms, are defined below. The closed unit ball $\{x \in \mathbb{C}^m : \|x\| \leq 1\}$ corresponding to each norm is illustrated to the right for the case $m = 2$.

$$
\begin{aligned}
\|x\|_1 &= \sum_{i=1}^{m} |x_i|, \\
\|x\|_2 &= \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2} = \sqrt{x^* x}, \\
\|x\|_\infty &= \max_{1 \leq i \leq m} |x_i|, \\
\|x\|_p &= \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p} \quad (1 \leq p < \infty).
\end{aligned}
\tag{3.2}
$$



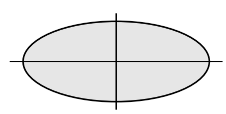
The 2-norm is the Euclidean length function; its unit ball is spherical. The 1-norm is used by airlines to define the maximal allowable size of a suitcase. The Sergel plaza in Stockholm, Sweden has the shape of the unit ball in the 4-norm; the Danish poet Piet Hein popularized this "superellipse" as a pleasing shape for objects such as conference tables.

Aside from the p-norms, the most useful norms are the *weighted p-norms*, where each of the coordinates of a vector space is given its own weight. In general, given any norm $\|\cdot\|$, a weighted norm can be written as

$$\|x\|_W = \|Wx\|. \tag{3.3}$$

Here W is the diagonal matrix in which the ith diagonal entry is the weight $w_i \neq 0$. For example, a weighted 2-norm $\|\cdot\|_W$ on $\mathbb{C}^m$ is specified as follows:

$$\|x\|_W = \left(\sum_{i=1}^{m} |w_i x_i|^2 \right)^{1/2}. \tag{3.4}$$



One can also generalize the idea of weighted norms by allowing W to be an arbitrary nonsingular matrix, not necessarily diagonal (Exercise 3.1).

The most important norms in this book are the unweighted 2-norm and its induced matrix norm.

Matrix Norms Induced by Vector Norms

An $m \times n$ matrix can be viewed as a vector in an mn-dimensional space: each of the mn entries of the matrix is an independent coordinate. Any mn-dimensional norm can therefore be used for measuring the "size" of such a matrix.

However, in dealing with a space of matrices, certain special norms are more useful than the vector norms (3.2)–(3.3) already discussed. These are the *induced matrix norms*, defined in terms of the behavior of a matrix as an operator between its normed domain and range spaces.

Given vector norms $\|\cdot\|_{(n)}$ and $\|\cdot\|_{(m)}$ on the domain and the range of $A \in \mathbb{C}^{m\times n}$, respectively, the induced matrix norm $\|A\|_{(m,n)}$ is the smallest number C for which the following inequality holds for all $x \in \mathbb{C}^n$:

$$\|Ax\|_{(m)} \le C\|x\|_{(n)}. \tag{3.5}$$

In other words, $\|A\|_{(m,n)}$ is the supremum of the ratios $\|Ax\|_{(m)}/\|x\|_{(n)}$ over all vectors $x \in \mathbb{C}^n$—the maximum factor by which A can "stretch" a vector x. We say that $\|\cdot\|_{(m,n)}$ is the matrix norm induced by $\|\cdot\|_{(m)}$ and $\|\cdot\|_{(n)}$.

Because of condition (3) of (3.1), the action of A is determined by its action on unit vectors. Therefore, the matrix norm can be defined equivalently in terms of the images of the unit vectors under A:

$$\|A\|_{(m,n)} = \sup_{\substack{x\in\mathbb{C}^n \\ x\neq 0}} \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \sup_{\substack{x\in\mathbb{C}^n \\ \|x\|_{(n)}=1}} \|Ax\|_{(m)}. \tag{3.6}$$

This form of the definition can be convenient for visualizing induced matrix norms, as in the sketches in (3.2) above.

Examples

Example 3.1. The matrix

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \tag{3.7}$$

maps $\mathbb{C}^2$ to $\mathbb{C}^2$. It also maps $\mathbb{R}^2$ to $\mathbb{R}^2$, which is more convenient if we want to draw pictures and also (it can be shown) sufficient for determining matrix p-norms, since the coefficients of A are real.

Figure 3.1 depicts the action of A on the unit balls of $\mathbb{R}^2$ defined by the 1-, 2-, and ∞-norms. From this figure, one can see a graphical interpretation of these three norms of A. Regardless of the norm, A maps $e_1 = (1,0)^*$ to the first column of A, namely e_1 itself, and $e_2 = (0,1)^*$ to the second column of A, namely $(2,2)^*$. In the 1-norm, the unit vector x that is amplified most by A is $(0,1)^*$ (or its negative), and the amplification factor is 4. In the ∞-norm, the unit vector x that is amplified most by A is $(1,1)^*$ (or its negative), and the amplification factor is 3. In the 2-norm, the unit vector that is amplified most by A is the vector indicated by the dashed line in the figure (or its negative), and the amplification factor is approximately 2.9208. (Note that it must be at least $\sqrt{8} \approx 2.8284$, since $(0,1)^*$ maps to $(2,2)^*$.) We shall consider how to calculate such 2-norm results in Lecture 5. □

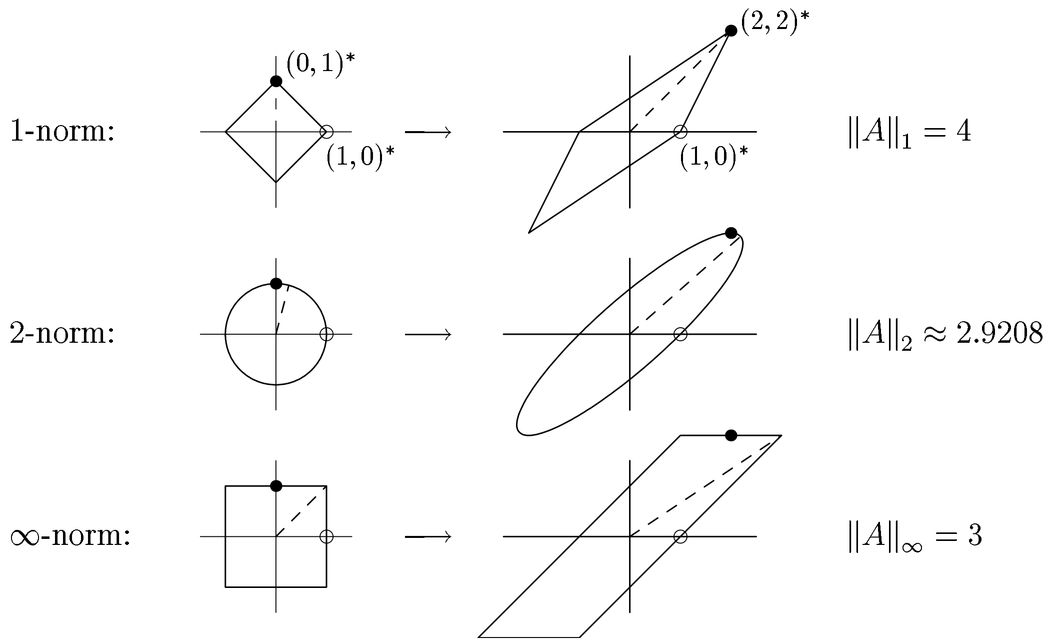


Figure 3.1. *On the left, the unit balls of* $\mathbb{R}^2$ *with respect to* $\|\cdot\|_1$, $\|\cdot\|_2$, *and* $\|\cdot\|_\infty$. *On the right, their images under the matrix* A *of* (3.7). *Dashed lines mark the vectors that are amplified most by* A *in each norm.*

Example 3.2. The p-Norm of a Diagonal Matrix. Let D be the diagonal matrix

$$D = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \end{bmatrix}.$$

Then, as in the second row of Figure 3.1, the image of the 2-norm unit sphere under D is an m-dimensional ellipse whose semiaxis lengths are given by the numbers $|d_i|$. The unit vectors amplified most by D are those that are mapped to the longest semiaxis of the ellipse, of length $\max_i\{|d_i|\}$. Therefore, we have $\|D\|_2 = \max_{1\le i\le m}\{|d_i|\}$. In the next lecture we shall see that *every* matrix maps the 2-norm unit sphere to an ellipse—properly called a *hyperellipse* if $m > 2$—though the axes may be oriented arbitrarily.

This result for the 2-norm generalizes to any p: if D is diagonal, then $\|D\|_p = \max_{1\le i\le m}|d_i|$. □

Example 3.3. The 1-Norm of a Matrix. If A is any $m \times n$ matrix, then $\|A\|_1$ is equal to the "maximum column sum" of A. We explain and derive this result as follows. Write A in terms of its columns

$$A = \left[\begin{array}{c|c|c} a_1 & \cdots & a_n \end{array} \right], \tag{3.8}$$

where each a_j is an m-vector. Consider the diamond-shaped 1-norm unit ball in $\mathbb{C}^n$, illustrated in (3.2). This is the set $\{x \in \mathbb{C}^n : \sum_{j=1}^n |x_j| \le 1\}$. Any vector Ax in the image of this set satisfies

$$\|Ax\|_1 \;=\; \Big\|\sum_{j=1}^n x_j a_j\Big\|_1 \;\le\; \sum_{j=1}^n |x_j|\,\|a_j\|_1 \;\le\; \max_{1\le j\le n}\|a_j\|_1.$$

Therefore the induced matrix 1-norm satisfies $\|A\|_1 \le \max_{1\le j\le n}\|a_j\|_1$. By choosing $x = e_j$, where j maximizes $\|a_j\|_1$, we attain this bound, and thus the matrix norm is

$$\|A\|_1 \;=\; \max_{1\le j\le n}\|a_j\|_1. \tag{3.9}$$

□

Example 3.4. The ∞-Norm of a Matrix. By much the same argument, it can be shown that the ∞-norm of an $m\times n$ matrix is equal to the "maximum row sum,"

$$\|A\|_\infty \;=\; \max_{1\le i\le m}\|a_i^*\|_1, \tag{3.10}$$

where a_i^* denotes the ith row of A. □

Cauchy–Schwarz and Hölder Inequalities

Computing matrix p-norms with $p \neq 1, \infty$ is more difficult, and to approach this problem, we note that inner products can be bounded using p-norms. Let p and q satisfy $1/p + 1/q = 1$, with $1 \le p, q \le \infty$. Then the *Hölder inequality* states that, for any vectors x and y,

$$|x^*y| \le \|x\|_p\|y\|_q. \tag{3.11}$$

The *Cauchy–Schwarz inequality* is the special case $p = q = 2$:

$$|x^*y| \le \|x\|_2\|y\|_2. \tag{3.12}$$

Derivations of these results can be found in linear algebra texts. Both bounds are tight in the sense that for certain choices of x and y, the inequalities become equalities.

Example 3.5. The 2-Norm of a Row Vector. Consider a matrix A containing a single row. This matrix can be written as $A = a^*$, where a is a column vector. The Cauchy–Schwarz inequality allows us to obtain the induced matrix 2-norm. For any x, we have $\|Ax\|_2 = |a^*x| \le \|a\|_2\|x\|_2$. This bound is tight: observe that $\|Aa\|_2 = \|a\|_2^2$. Therefore, we have

$$\|A\|_2 \;=\; \sup_{x\neq 0}\{\|Ax\|_2/\|x\|_2\} \;=\; \|a\|_2.$$

□

Example 3.6. The 2-Norm of an Outer Product. More generally, consider the rank-one outer product $A = uv^*$, where u is an m-vector and v is an n-vector. For any n-vector x, we can bound $\|Ax\|_2$ as follows:

$$\|Ax\|_2 = \|uv^*x\|_2 = \|u\|_2|v^*x| \leq \|u\|_2\|v\|_2\|x\|_2. \tag{3.13}$$

Therefore $\|A\|_2 \leq \|u\|_2\|v\|_2$. Again, this inequality is an equality: consider the case $x = v$. □

Bounding $\|AB\|$ in an Induced Matrix Norm

The induced matrix norm of a matrix product can also be bounded. Let $\|\cdot\|_{(\ell)}$, $\|\cdot\|_{(m)}$, and $\|\cdot\|_{(n)}$ be norms on $\mathbb{C}^l$, $\mathbb{C}^m$, and $\mathbb{C}^n$, respectively, and let A be an $l \times m$ matrix and B an $m \times n$ matrix. For any $x \in \mathbb{C}^n$ we have

$$\|ABx\|_{(\ell)} \leq \|A\|_{(\ell,m)}\|Bx\|_{(m)} \leq \|A\|_{(\ell,m)}\|B\|_{(m,n)}\|x\|_{(n)}.$$

Therefore the induced norm of AB must satisfy

$$\|AB\|_{(\ell,n)} \leq \|A\|_{(\ell,m)}\|B\|_{(m,n)}. \tag{3.14}$$

In general this inequality is not an equality. For example, the inequality $\|A^n\| \leq \|A\|^n$ holds for any square matrix in any matrix norm induced by a vector norm, but $\|A^n\| = \|A\|^n$ does not hold in general for $n \geq 2$.

General Matrix Norms

As noted above, matrix norms do not have to be induced by vector norms. In general, a matrix norm must merely satisfy the three vector norm conditions (3.1) applied in the mn-dimensional vector space of matrices:

$$\begin{aligned} &(1)\quad \|A\| \geq 0, \text{ and } \|A\| = 0 \text{ only if } A = 0,\\ &(2)\quad \|A + B\| \leq \|A\| + \|B\|,\\ &(3)\quad \|\alpha A\| = |\alpha|\,\|A\|. \end{aligned} \tag{3.15}$$

The most important matrix norm which is not induced by a vector norm is the *Hilbert–Schmidt* or *Frobenius norm*, defined by

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}. \tag{3.16}$$

Observe that this is the same as the 2-norm of the matrix when viewed as an mn-dimensional vector. The formula for the Frobenius norm can also be written in terms of individual rows or columns. For example, if a_j is the jth column of A, we have

$$\|A\|_F = \left(\sum_{j=1}^{n} \|a_j\|_2^{\,2} \right)^{1/2}. \tag{3.17}$$

This identity, as well as the analogous result based on rows instead of columns, can be expressed compactly by the equation

$$\|A\|_F = \sqrt{\operatorname{tr}(A^*A)} = \sqrt{\operatorname{tr}(AA^*)}, \tag{3.18}$$

where $\operatorname{tr}(B)$ denotes the *trace* of B, the sum of its diagonal entries.

Like an induced matrix norm, the Frobenius norm can be used to bound products of matrices. Let $C = AB$ with entries c_{ik}, and let a_i^* denote the ith row of A and b_j the jth column of B. Then $c_{ij} = a_i^* b_j$, so by the Cauchy–Schwarz inequality we have $|c_{ij}| \le \|a_i\|_2 \|b_j\|_2$. Squaring both sides and summing over all i, j, we obtain

$$\begin{aligned} \|AB\|_F^2 &= \sum_{i=1}^{n} \sum_{j=1}^{m} |c_{ij}|^2 \\ &\le \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\|a_i\|_2 \|b_j\|_2 \right)^2 \\ &= \sum_{i=1}^{n} \left(\|a_i\|_2 \right)^2 \sum_{j=1}^{m} \left(\|b_j\|_2 \right)^2 \;=\; \|A\|_F^2 \|B\|_F^2. \end{aligned}$$

Invariance under Unitary Multiplication

One of the many special properties of the matrix 2-norm is that, like the vector 2-norm, it is invariant under multiplication by unitary matrices. The same property holds for the Frobenius norm.

Theorem 3.1. *For any $A \in \mathbb{C}^{m\times n}$ and unitary $Q \in \mathbb{C}^{m\times m}$, we have*

$$\|QA\|_2 = \|A\|_2, \qquad \|QA\|_F = \|A\|_F.$$

Proof. Since $\|Qx\|_2 = \|x\|_2$ for every x, by (2.10), the invariance in the 2-norm follows from (3.6). For the Frobenius norm we note that by (3.17), it is enough to show that the jth column of QA has the same 2-norm as the jth column of A, and this follows from (1.6) and (2.10). $\square$

Exercises

1. Prove that if W is an arbitrary nonsingular matrix, the function $\|\cdot\|_W$ defined by (3.3) is a vector norm.

2. Let $\|\cdot\|$ denote any norm on $\mathbb{C}^m$ and also the induced matrix norm on $\mathbb{C}^{m\times m}$. Show that $\rho(A) \le \|A\|$, where $\rho(A)$ is the *spectral radius* of A, i.e., the largest absolute value $|\lambda|$ of an eigenvalue λ of A.

3. Vector and matrix p-norms are related by various inequalities, often involving the dimensions m or n. For each of the following, verify the inequality and give an example of a nonzero vector or matrix (for general m, n) for which equality is achieved. In this problem x is an m-vector and A is an $m \times n$ matrix.
   (a) $\|x\|_\infty \le \|x\|_2$
   (b) $\|x\|_2 \le \sqrt{m}\,\|x\|_\infty$
   (c) $\|A\|_\infty \le \sqrt{n}\,\|A\|_2$
   (d) $\|A\|_2 \le \sqrt{m}\,\|A\|_\infty$

4. Let A be an $m \times n$ matrix and let B be a submatrix of A, that is, an $\mu \times \nu$ matrix ($\mu \le m$, $\nu \le n$) obtained by selecting certain rows and columns of A.
   (a) Explain how B can be obtained by multiplying A by certain row and column "deletion matrices" as in step (7) of Exercise 1.1.
   (b) Using this product, show that $\|B\|_p \le \|A\|_p$ for any p with $1 \le p \le \infty$.

5. Example 3.6 shows that if E is an outer product $E = uv^*$, then $\|E\|_2 = \|u\|_2\|v\|_2$. Is the same true for the Frobenius norm, i.e., $\|E\|_F = \|u\|_F\|v\|_F$? Prove it or give a counterexample.

6. Let $\|\cdot\|$ denote any norm on $\mathbb{C}^m$. The corresponding *dual norm* $\|\cdot\|'$ is defined by the formula $\|x\|' = \sup_{\|y\|=1} |y^*x|$.
   (a) Prove that $\|\cdot\|'$ is a norm.
   (b) Let $x, y \in \mathbb{C}^m$ with $\|x\| = \|y\| = 1$ be given. Show that there exists a rank-one matrix $B = yz^*$ such that $Bx = y$ and $\|B\| = 1$, where $\|B\|$ is the matrix norm of B induced by the vector norm $\|\cdot\|$. You may use the following lemma, without proof: given $x \in \mathbb{C}^m$, there exists a nonzero $z \in \mathbb{C}^m$ such that $|z^*x| = \|z\|'\|x\|$.