COMPUTING THE NORM OF A MATRIX

KEITH CONRAD

1. Introduction

In $\mathbf{R}^n$ there is a standard notion of length: the size of a vector $v = (a_1, \ldots, a_n)$ is

$$\|v\| = \sqrt{a_1^2 + \cdots + a_n^2}.$$

We will discuss in Section 2 the general concept of length in a vector space, called a norm, and then look at norms on matrices in Section 3. In Section 4 we'll see how the matrix norm that is closely connected to the standard norm on $\mathbf{R}^n$ can be computed from eigenvalues of an associated symmetric matrix.

2. Norms on Vector Spaces

Let V be a vector space over $\mathbf{R}$. A *norm* on V is a function $\|\cdot\| : V \to \mathbf{R}$ satisfying three properties:

(1) $\|v\| \geq 0$ for all $v \in V$, with equality if and only if $v = 0$,
(2) $\|v + w\| \leq \|v\| + \|w\|$ for all v and w in V,
(3) $\|cv\| = |c| \, \|v\|$ for all $c \in \mathbf{R}$ and $v \in V$.

The same definition applies to complex vector spaces. From a norm on V we get a metric on V by $d(v, w) = \|v - w\|$. The triangle inequality for this metric is a consequence of the second property of norms.

Example 2.1. The standard norm on $\mathbf{R}^n$, using the standard basis $e_1, \ldots, e_n$, is

$$\left\| \sum_{i=1}^n a_i e_i \right\| = \sqrt{\sum_{i=1}^n a_i^2}.$$

This gives rise to the Euclidean metric on $\mathbf{R}^n$: $d(\sum a_i e_i, \sum b_i e_i) = \sqrt{\sum (a_i - b_i)^2}$.

Example 2.2. The sup-norm on $\mathbf{R}^n$ is

$$\left\| \sum_{i=1}^n a_i e_i \right\|_{\sup} = \max_i |a_i|.$$

This gives rise to the sup-metric on $\mathbf{R}^n$: $d(\sum a_i e_i, \sum b_i e_i) = \max |a_i - b_i|$.

Example 2.3. On $\mathbf{C}^n$ the standard norm and sup-norm are defined similarly to the case of $\mathbf{R}^n$, but we need $|z|^2$ instead of z^2 when z is complex:

$$\left\| \sum_{i=1}^n a_i e_i \right\| = \sqrt{\sum_{i=1}^n |a_i|^2}, \quad \left\| \sum_{i=1}^n a_i e_i \right\|_{\sup} = \max_i |a_i|.$$

A common way of placing a norm on a real vector space V is by an *inner product*, which is a pairing $(\cdot, \cdot) \colon V \times V \to \mathbf{R}$ that is

(1) bilinear: linear in each component when the other is fixed. Linearity in the first component means $(v + v', w) = (v, w) + (v', w)$ and $(cv, w) = c(v, w)$ for $v, v', w \in V$ and $c \in \mathbf{R}$, and similarly in the second component.

(2) symmetric: $(v, w) = (w, v)$.

(3) positive-definite: $(v, v) \geq 0$, with equality if and only if $v = 0$. The standard inner product on $\mathbf{R}^n$ is the dot product:

$$\left(\sum_{i=1}^{n} a_i e_i, \sum_{i=1}^{n} b_i e_i \right) = \sum_{i=1}^{n} a_i b_i.$$

For an inner product $(\cdot, \cdot)$ on V, a norm can be defined by the formula

$$\|v\| = \sqrt{(v, v)}.$$

That this is actually a norm on V follows from the Cauchy–Schwarz inequality

$$(2.1) \qquad\qquad |(v, w)| \leq (v, v)(w, w) = \|v\| \, \|w\|$$

as follows. For all v and w in V,

$$
\begin{aligned}
\|v + w\|^2 &= (v + w, v + w) \\
&= (v, v) + (v, w) + (w, v) + (w, w) \\
&= \|v\|^2 + 2(v, w) + \|w\|^2 \\
&\leq \|v\|^2 + 2|(v, w)| + \|w\|^2 \quad \text{since } a \leq |a| \text{ for all } a \in \mathbf{R} \\
&\leq \|v\|^2 + 2\|v\| \, \|w\| + \|w\|^2 \quad \text{by } (2.1) \\
&= (\|v\| + \|w\|)^2.
\end{aligned}
$$

Taking (positive) square roots of both sides yields $\|v + w\| \leq \|v\| + \|w\|$. A proof of the Cauchy–Schwarz inequality is in the appendix.

Using the standard inner product on $\mathbf{R}^n$, the Cauchy–Schwarz inequality assumes its classical form, proved by Cauchy in 1821:

$$\left| \sum_{i=1}^{n} a_i b_i \right| = \sqrt{\sum_{i=1}^{n} a_i^2 \cdot \sum_{I=1}^{n} b_i^2}.$$

The Cauchy–Schwarz inequality (2.1) is true for every inner product on a real vector space, not just the standard inner product on $\mathbf{R}^n$.

While the norm on $\mathbf{R}^n$ that comes from the standard inner product is the standard norm, the sup-norm on $\mathbf{R}^n$ does *not* arise from an inner product, *i.e.*, there is no inner product whose associated norm is the sup-norm.

Even though the sup-norm and the standard norm on $\mathbf{R}^n$ are not equal, they are each bounded by a constant multiple of the other one:

$$(2.2) \qquad\qquad \max_i |a_i| \leq \sqrt{\sum_{i=1}^{n} a_i^2} \leq \sqrt{n} \max_i |a_i|.$$

i.e., $\|v\|_{\sup} \leq \|v\| \leq \sqrt{n} \, \|v\|_{\sup}$ for all $v \in \mathbf{R}^n$. Therefore the metrics these two norms give rise to determine the same notions of convergence: a sequence in $\mathbf{R}^n$ that is convergent with

respect to one of the metrics is also convergent with respect to the other metric. Also $\mathbf{R}^n$ is complete with respect to both of these metrics.

The standard inner product on $\mathbf{R}^n$ is closely tied to transposition of $n \times n$ matrices. For $A = (a_{ij}) \in \mathrm{M}_n(\mathbf{R})$, let $A^\top = (a_{ji})$ be its transpose. Then for all $v, w \in \mathbf{R}^n$,

$$(2.3) \qquad (Av, w) = (v, A^\top w), \quad (v, Aw) = (A^\top v, w),$$

where $(\cdot, \cdot)$ is the standard inner product on $\mathbf{R}^n$.

We now briefly indicate what inner products are on complex vector spaces. An *inner product* on a complex vector space V is a pairing $(\cdot, \cdot) \colon V \times V \to \mathbf{C}$ that is

   (1) linear on the left and conjugate-linear on the right: it is additive in each component with the other one fixed and $(cv, w) = c(v, w)$ and $(v, cw) = \overline{c}(v, w)$ for $c \in \mathbf{C}$.
   (2) conjugate-symmetric: $(v, w) = \overline{(w, v)}$,
   (3) positive-definite: $(v, v) \geq 0$, with equality if and only if $v = 0$.

The standard inner product on $\mathbf{C}^n$ is not the dot product, but has a conjugation in the second component:

$$(2.4) \qquad \left(\sum_{i=1}^n a_i e_i, \sum_{i=1}^n b_i e_i \right) = \sum_{i=1}^n a_i \overline{b_i}.$$

This standard inner product on $\mathbf{C}^n$ is closely tied to conjugate-transposition of $n \times n$ complex matrices. For $A = (a_{ij}) \in \mathrm{M}_n(\mathbf{C})$, let $A^* = \overline{A}^\top = (\overline{a_{ji}})$ be its conjugate-transpose. Then for all $v, w \in \mathbf{C}^n$,

$$(2.5) \qquad (Av, w) = (v, A^* w), \quad (v, Aw) = (A^* v, w).$$

An inner product on a complex vector space satisfies the Cauchy–Schwarz inequality, so it can be used to define a norm just as in the case of inner products on real vector spaces.

Although we will be focusing on norms on finite-dimensional spaces, the extension of these ideas to infinite-dimensional spaces is quite important in both analysis and physics (quantum mechanics). Speaking of physics, physicists define inner products on complex vector spaces as being linear on the right and conjugate-linear on the left. It's just a difference in notation, but leads to more differences, *e.g.*, the physicist's standard inner product on $\mathbf{C}^n$ is $(\sum_{i=1}^n a_i e_i, \sum_{i=1}^n b_i e_i) = \sum_{i=1}^n \overline{a_i} b_i$, which is the complex conjugate of the mathematician's standard inner product on $\mathbf{C}^n$ as defined earlier.

Exercises.

   1. Letting $(\cdot, \cdot)_m$ be the dot product on $\mathbf{R}^m$ and $(\cdot, \cdot)_n$ be the dot product on $\mathbf{R}^n$, show for each $m \times n$ matrix A, $n \times m$ matrix B, $v \in \mathbf{R}^n$, and $w \in \mathbf{R}^m$ that $(Av, w)_m = (v, A^\top w)_n$ and $(v, Bw)_n = (B^\top v, w)_m$. When $m = n$ this becomes (2.3).
   2. Verify (2.5).

3. Defining Norms on Matrices

From now on, the norm and inner product on $\mathbf{R}^n$ and $\mathbf{C}^n$ are the standard ones.

The set of $n \times n$ real matrices $\mathrm{M}_n(\mathbf{R})$ forms a real vector space. How should we define a norm on $\mathrm{M}_n(\mathbf{R})$? One idea is to view $\mathrm{M}_n(\mathbf{R})$ as $\mathbf{R}^{n^2}$ and use the sup-norm on $\mathbf{R}^{n^2}$

$$\|(a_{ij})\|_{\sup} = \max_{i,j} |a_{ij}|.$$

or the standard norm on $\mathbf{R}^{n^2}$ (also called the *Frobenius norm* on $\mathrm{M}_n(\mathbf{R})$): $\|(a_{ij})\| = \sqrt{\sum a_{ij}^2}$. While these are used in numerical work, in pure math there is a better choice of a norm on $\mathrm{M}_n(\mathbf{R})$ than either of these. Before describing it, we use the sup-norm on $\mathrm{M}_n(\mathbf{R})$ to show that each $n \times n$ matrix changes the (standard) length of vectors in $\mathbf{R}^n$ by a uniformly bounded amount that depends only on n and the matrix. For $v = (c_1, \ldots, c_n) \in \mathbf{R}^n$,

$$
\begin{aligned}
\|Av\| &\leq \sqrt{n}\,\|Av\|_{\mathrm{sup}} \quad \text{by (2.2)} \\
&= \sqrt{n}\,\max_i \left| \sum_{j=1}^n a_{ij}c_j \right| \\
&\leq \sqrt{n}\,\max_i \sum_{j=1}^n |a_{ij}||c_j| \\
&\leq \sqrt{n}\,\max_i \sum_{j=1}^n |a_{ij}|\,\|v\|_{\mathrm{sup}} \\
&\leq n\sqrt{n}\,\max_{i,j} |a_{ij}|\,\|v\|_{\mathrm{sup}} \\
&\leq n\sqrt{n}\,\max_{i,j} |a_{ij}|\,\|v\| \quad \text{by (2.2).}
\end{aligned}
$$

Let $C = n\sqrt{n}\,\max |a_{ij}|$. This is a constant depending on the dimension n of the space and the matrix A, but not on v, and $\|Av\| \leq C\,\|v\|$ for all v. By linearity, $\|Av - Aw\| \leq C\,\|v - w\|$ for all $v, w \in \mathbf{R}^n$.

Let's write down the above calculation as a lemma.

Lemma 3.1. *For each $A \in \mathrm{M}_n(\mathbf{R})$, there is a $C \geq 0$ such that $\|Av\| \leq C\,\|v\|$ for all $v \in \mathbf{R}^n$.*

The constant C we wrote down might not be optimal. Perhaps there is a smaller constant $C' < C$ such that $\|Av\| \leq C'\,\|v\|$ for all $v \in \mathbf{R}^n$. We will get a norm on $\mathrm{M}_n(\mathbf{R})$ by assigning to each $A \in \mathrm{M}_n(\mathbf{R})$ the least $C \geq 0$ such that $\|Av\| \leq C\,\|v\|$ for all $v \in \mathbf{R}^n$, where the vector norms in $\|Av\|$ and $\|v\|$ are the standard ones on $\mathbf{R}^n$.

Theorem 3.2. *For each $A \in \mathrm{M}_n(\mathbf{R})$, there is a unique real number $b \geq 0$ such that*

(i) *$\|Av\| \leq b\,\|v\|$ for all $v \in \mathbf{R}^n$,*
(ii) *b is minimal: if $\|Av\| \leq C\,\|v\|$ for all $v \in \mathbf{R}^n$, then $b \leq C$.*

Proof. We first show by a scaling argument that $\|Av\| \leq b\,\|v\|$ for all $v \in \mathbf{R}^n$ if and only if $\|Av\| \leq b$ for all $v \in \mathbf{R}^n$ with $\|v\| = 1$. The direction ($\Rightarrow$) is clear by using $\|v\| = 1$. For the direction ($\Leftarrow$), when $v = 0$ we trivially have $\|Av\| = 0 = b\,\|v\|$. When $v \neq 0$ let $c = \|v\|$, so $c > 0$ and the vector v/c has norm 1 ($\|v/c\| = |1/c|\,\|v\| = (1/|c|)\,\|v\| = (1/\|v\|)\,\|v\| = 1$), so by hypothesis $\|A(v/c)\| \leq b$, which implies $(1/|c|)\,\|Av\| \leq b$ by linearity. Now multiply both sides by $|c|$ to get $\|Av\| \leq bc = b\,\|v\|$.

Therefore the theorem is saying the set $\{\|Av\| : \|v\| = 1\}$ has a (finite) maximum value, which is b. This is what we will prove.

The matrix A as a function $\mathbf{R}^n \to \mathbf{R}^n$ is continuous since the components of Av are linear functions of the components of v, and hence they are each continuous in v. The standard norm $\|\cdot\| : \mathbf{R}^n \to \mathbf{R}$ is also continuous since it is the square root of a polynomial function of the coordinates. Finally, since the unit sphere in $\mathbf{R}^n$, $\{v \in \mathbf{R}^n : \|v\| = 1\}$, is compact

its continuous image $\{\|Av\| : \|v\| = 1\}$ in $\mathbf{R}$ is also compact. Every compact subset of $\mathbf{R}$ contains a maximum point, so we are done. $\qquad\square$

Definition 3.3. For $A \in \mathrm{M}_n(\mathbf{R})$, $\|A\|$ is the smallest nonnegative real number satisfying the inequality $\|Av\| \leq \|A\| \|v\|$ for all $v \in \mathbf{R}^n$. This is called the *operator norm* of A.

Theorem 3.2 shows $\|A\|$ exists and is the maximum of $\|Av\|$ when v runs over the unit sphere in $\mathbf{R}^n$, so in plain English, $\|A\|$ is the largest amount by which A stretches a vector on the unit sphere of $\mathbf{R}^n$. The next theorem shows the operator norm on $\mathrm{M}_n(\mathbf{R})$ is a vector space norm and has a host of other nice properties.

Theorem 3.4. *For $A, B \in \mathrm{M}_n(\mathbf{R})$ and $v, w \in \mathbf{R}^n$,*

(i) $\|A\| \geq 0$, *with equality if and only if $A = O$.*
(ii) $\|A + B\| \leq \|A\| + \|B\|$.
(iii) $\|cA\| = |c| \|A\|$ *for $c \in \mathbf{R}$.*
(iv) $\|AB\| \leq \|A\| \|B\|$. *It is typically false that $\|AB\| = \|A\| \|B\|$.*
(v) $\|A\| = \|A^\top\|$.
(vi) $\|AA^\top\| = \|A^\top A\| = \|A\|^2$. *Thus $\|A\| = \sqrt{\|AA^\top\|} = \sqrt{\|A^\top A\|}$.*
(vii) $|(Av, w)| \leq \|A\| \|v\| \|w\|$.
(viii) $\|A\|_{\mathrm{sup}} \leq \|A\| \leq n\sqrt{n} \|A\|_{\mathrm{sup}}$ *and $\mathrm{M}_n(\mathbf{R})$ is complete with respect to the metric coming from the operator norm.*

Proof. (i) It is obvious that $\|A\| \geq 0$. If $\|A\| = 0$ then for all $v \in \mathbf{R}^n$ we have $\|Av\| \leq 0 \|v\| = 0$, so $\|Av\| = 0$. Thus $Av = 0$ for all $v \in \mathbf{R}^n$, so $A = O$. The converse is trivial.

(ii) For all $v \in \mathbf{R}^n$,

$$
\begin{aligned}
\|(A + B)v\| &= \|Av + Bv\| \\
&\leq \|Av\| + \|Bv\| \\
&\leq \|A\| \|v\| + \|B\| \|v\| \\
&= (\|A\| + \|B\|) \|v\|.
\end{aligned}
$$

Since $\|A + B\|$ is the *least* $C \geq 0$ such that $\|(A + B)v\| \leq C \|v\|$ for all $v \in \mathbf{R}^n$, $\|A + B\| \leq \|A\| + \|B\|$.

(iii) Left to the reader.

(iv) For all $v \in \mathbf{R}^n$, $\|(AB)v\| = \|A(Bv)\| \leq \|A\| \|Bv\| \leq \|A\| \|B\| \|v\|$, so the minimality property of the operator norm implies $\|AB\| \leq \|A\| \|B\|$.

To show that generally $\|AB\| \neq \|A\| \|B\|$, note that if $\|AB\| = \|A\| \|B\|$ for all A and B in $\mathrm{M}_n(\mathbf{R})$ then for nonzero A and B we'd have $\|AB\| \neq 0$, so $AB \neq O$. That is, the product of two nonzero matrices is always nonzero. This is false when $n > 1$, since there are many nonzero matrices whose square is zero. (We sometimes do have $\|AB\| = \|A\| \|B\|$: see Exercise 3.2.)

(v) For all $v \in \mathbf{R}^n$, we get by (2.3) that

$$
\|Av\|^2 = |(Av, Av)| = |(v, A^\top Av)| \leq \|v\| \|A^\top Av\|
$$

by Cauchy–Schwarz. This last expression is $\leq \|A^\top A\| \|v\|^2$, so

$$
\|Av\| \leq \sqrt{\|A^\top A\|} \|v\|.
$$

The *least* $C \geq 0$ such that $\|Av\| \leq C \|v\|$ for all $v \in \mathbf{R}^n$ is $\|A\|$, so $\|A\| \leq \sqrt{\|A^\top A\|}$. Squaring both sides,

(3.1) $$\|A\|^2 \leq \|A^\top A\| \leq \|A^\top\| \|A\|.$$

Dividing by $\|A\|$ when $A \neq O$, we get $\|A\| \leq \|A^\top\|$. This is also obvious if $A = O$, so

(3.2) $$\|A\| \leq \|A^\top\|$$

for all $A \in \mathrm{M}_n(\mathbf{R})$. Now replace A by $A^\top$ in (3.2) to get

$$\|A^\top\| \leq \|(A^\top)^\top\| = \|A\|,$$

so $\|A\| = \|A^\top\|$.

(vi) Feeding the conclusion of (v) back into (3.1),

$$\|A\|^2 \leq \|A^\top A\| \leq \|A^\top\| \|A\| = \|A\|^2,$$

so $\|A\|^2 = \|A^\top A\|$. Using $A^\top$ in place of A here, we get $\|A\|^2 = \|AA^\top\|$ since $\|A^\top\| = \|A\|$.

(vii) Use Cauchy–Schwarz: $|(Av, w)| \leq \|Av\| \|w\| \leq \|A\| \|v\| \|w\|$.

(viii) Set $v = e_j$ and $w = e_i$ in (vii):

$$|a_{ij}| = |(Ae_j, e_i)| \leq \|A\|.$$

Therefore $\|A\|_{\sup} \leq \|A\|$. The other inequality follows from the calculation leading up to Lemma 3.1.

That $\mathrm{M}_n(\mathbf{R})$ is complete with respect to the metric $d(A, B) = \|A - B\|$ coming from the operator norm follows from completeness of $\mathrm{M}_n(\mathbf{R})$ with respect to the metric coming from the sup-norm (view $\mathrm{M}_n(\mathbf{R})$ as $\mathbf{R}^{n^2}$) and the fact that these two norms on $\mathrm{M}_n(\mathbf{R})$ are bounded by constant multiples of each other. $\qquad\square$

The operator norm on $\mathrm{M}_n(\mathbf{R})$ interacts nicely with the multiplicative structure on $\mathrm{M}_n(\mathbf{R})$ and the standard inner product on $\mathbf{R}^n$ (parts (iv) through (vii) of Theorem 3.4). However, unlike the standard norm on $\mathbf{R}^n$, the operator norm on $\mathrm{M}_n(\mathbf{R})$ is impossible to calculate from its definition in all but the simplest cases. For instance, it is clear that $\|I_n\| = 1$, so $\|cI_n\| = |c|$ for all $c \in \mathbf{R}$. But what is

$$\left\| \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \right\|?$$

By the last part of Theorem 3.4, this norm is bounded above by $2\sqrt{2}(4) = 8\sqrt{2} \approx 11.3$. In the next section we will give a formula for the operator norm on $\mathrm{M}_n(\mathbf{R})$ that will allow us to compute $\|(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix})\|$ easily, and it will turn out to be ≈ 5.5.

Exercises.

1. Rework this section for rectangular matrices that need not be square. For A in $\mathrm{M}_{m,n}(\mathbf{R})$, define its operator norm $\|A\|_{m,n}$ to be the least $b \geq 0$ such that $\|Av\|_m \leq b \|v\|_n$ for all $v \in \mathbf{R}^n$, where the subscripts in the inequality indicate the standard norm on the Euclidean space of the relevant dimension (m or n). Show $\|A\|_{m,n}$ exists and

$$\|A\|_{m,n} = \|A^\top\|_{n,m} = \sqrt{\|AA^\top\|_{m,m}} = \sqrt{\|A^\top A\|_{n,n}}.$$

You will want to use the relationship between the transpose on $\mathrm{M}_{m,n}(\mathbf{R})$ and dot products on $\mathbf{R}^m$ and $\mathbf{R}^n$ in Exercise 2.1.

2. An $n \times n$ *rotation matrix* is a matrix $R \in \mathrm{M}_n(\mathbf{R})$ such that $RR^\top = I_n$ (equivalently, $R^\top R = I_n$), which is the same as the condition $\langle Rv, Rw \rangle = \langle v, w \rangle$ for all v and w in $\mathbf{R}^n$. Prove $\|R\| = 1$ and for every $n \times n$ real matrix A, $\|AR\| = \|A\|$ and $\|RA\| = \|A\|$. (Hint: use the submultiplicativity property of the operator norm in Theorem 3.4(iv).)

3. Define an operator norm on $\mathrm{M}_n(\mathbf{C})$ and establish an analogue of Theorem 3.4. Is it true that a matrix in $\mathrm{M}_n(\mathbf{C})$ generally has the same operator norm as its transpose? For a real $n \times n$ matrix, show its operator norm as an element of $\mathrm{M}_n(\mathbf{R})$ equals its operator norm as an element of $\mathrm{M}_n(\mathbf{C})$.

4. A computational formula for the operator norm on $\mathrm{M}_n(\mathbf{R})$

The key idea to compute the operator norm of $A \in \mathrm{M}_n(\mathbf{R})$ is that Theorem 3.4(vi) tells us it is also the square root of the operator norm of $AA^\top$ and $A^\top A$. What makes $AA^\top$ and $A^\top A$ special is that they are symmetric (equal to their own transposes), *e.g.*, $(AA^\top)^\top = (A^\top)^\top A^\top = AA^\top$.[1] The following theorem gives a method to compute operator norms of symmetric matrices and then general square matrices.

Theorem 4.1. (1) *If $A \in \mathrm{M}_n(\mathbf{R})$ satisfies $A^\top = A$ then all the eigenvalues of A are real and*

$$(4.1) \qquad \|A\| = \max_{\text{eigenvalues } \lambda \text{ of } A} |\lambda|.$$

(2) *For all $A \in \mathrm{M}_n(\mathbf{R})$, the eigenvalues of $AA^\top$ and $A^\top A$ are all nonnegative.*

Proof. (1) To prove that when $A^\top = A$ all the eigenvalues of A are real, let A act on $\mathbf{C}^n$ in the obvious way. Using the standard inner product (2.4) on $\mathbf{C}^n$ (not the dot product!), we have by (2.5) that for all $v \in \mathbf{C}^n$,

$$(Av, v) = (v, A^*v) = (v, \overline{A}^\top v) = (v, A^\top v) = (v, Av).$$

For an eigenvalue $\lambda \in \mathbf{C}$ of A, let $v \in \mathbf{C}^n$ be a corresponding eigenvector. Then

$$(Av, v) = (\lambda v, v) = \lambda(v, v), \qquad (v, Av) = (v, \lambda v) = \overline{\lambda}(v, v),$$

so $\lambda = \overline{\lambda}$ since $(v, v) = \|v\|^2 \neq 0$ (eigenvectors are nonzero). Thus λ is real.

To relate $\|A\|$ to the eigenvalues of A as in (4.1), we will use a fundamental property of real symmetric matrices that is called the *Spectral Theorem*. It asserts that every symmetric matrix $A \in \mathrm{M}_n(\mathbf{R})$ has a basis of mutually orthogonal eigenvectors in $\mathbf{R}^n$.[2] Let $v_1, \ldots, v_n$ be a basis of mutually orthogonal eigenvectors for A, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. What is special about orthogonal vectors is that their squared lengths add (the Pythagorean theorem): if $(v, w) = 0$ then

$$\|v + w\|^2 = (v + w, v + w) = (v, v) + 2(v, w) + (w, w) = (v, v) + (w, w) = \|v\|^2 + \|w\|^2$$

and likewise for a sum of more than two mutually orthogonal vectors.

Order the eigenvalues of A so that $|\lambda_1| \leq \ldots \leq |\lambda_n|$. For each $v \in \mathbf{R}^n$, write it in terms of the basis of eigenvectors as

$$v = c_1 v_1 + \cdots + c_n v_n.$$

---

[1]This is also true if $A \in \mathrm{M}_{m,n}(\mathbf{R})$ is a rectangular matrix, which makes Section 4 applicable to operator norms of rectangular matrices by Exercise 3.1.

[2]The Spectral Theorem includes the assertion that all eigenvalues of A are real, which we showed above.

Then
$$Av = c_1 A(v_1) + \cdots + c_n A(v_n) = c_1 \lambda_1 v_1 + \cdots + c_n \lambda_n v_n.$$

Since the v_i's are mutually perpendicular, their scalar multiples $c_i \lambda_i v_i$ are mutually perpendicular. Therefore

$$
\begin{aligned}
\|Av\|^2 &= \|c_1 \lambda_1 v_1\|^2 + \cdots + \|c_n \lambda_n v_n\|^2 \\
&= c_1^2 \lambda_1^2 \|v_1\|^2 +, \cdots + c_n^2 \lambda_n^2 \|v_n\|^2 \\
&\leq c_1^2 \lambda_n^2 \|v_1\|^2 +, \cdots + c_n^2 \lambda_n^2 \|v_n\|^2 \quad \text{since } |\lambda_i| \leq |\lambda_n| \\
&= \lambda_n^2 (c_1^2 \|v_1\|^2 +, \cdots + c_n^2 \|v_n\|^2) \\
&= \lambda_n^2 (\|c_1 v_1\|^2 + \cdots + \|c_n v_n\|^2) \\
&= \lambda_n^2 \|c_1 v_1 + \cdots + c_n v_n\|^2 \\
&= \lambda_n^2 \|v\|^2,
\end{aligned}
$$

so $\|Av\| \leq |\lambda_n| \|v\|$. Since this inequality holds for all v in $\mathbf{R}^n$, we have $\|A\| \leq |\lambda_n|$. To prove $\|A\| = |\lambda_n|$ it now suffices to find a single nonzero vector v such that $\|Av\| = |\lambda_n| \|v\|$. For that we can use $v = v_n$ since $Av_n = \lambda_n v_n$.

(2) Since $AA^\top$ and $A^\top A$ are both symmetric, all their eigenvalues are real. Let $\lambda \in \mathbf{R}$ be an eigenvalue of $AA^\top$ with corresponding eigenvector $v \in \mathbf{R}^n$, and let $\mu \in \mathbf{R}$ be an eigenvalue of $A^\top A$ with corresponding eigenvector $w \in \mathbf{R}^n$. Using the standard inner product on $\mathbf{R}^n$,

$$0 \leq (Aw, Aw) = (w, A^\top A w) = (w, \mu w) = \mu(w, w).$$

Then $\mu \geq 0$ since $(w, w) = \|w\|^2 > 0$. Similarly,

$$0 \leq (A^\top v, A^\top v) = (v, AA^\top v) = (v, \lambda v) = \lambda(v, v).$$

Since $(v, v) = \|v\|^2 > 0$, it follows that $\lambda \geq 0$. $\qquad \square$

Corollary 4.2. *For $A \in \mathrm{M}_n(\mathbf{R})$, $\|A\|$ is the square root of the largest eigenvalue of $AA^\top$ and is the square root of the largest eigenvalue of $A^\top A$.*

Proof. By Theorem 3.4(vi), $\|A\| = \sqrt{\|AA^\top\|} = \sqrt{\|A^\top A\|}$. Now use (4.1), with $AA^\top$ and $A^\top A$ in place of A. $\qquad \square$

Remark 4.3. This corollary, without proof, goes back to Peano [2, p. 454] using $A^\top A$. On the same page Peano introduced the operator norm on $\mathrm{M}_n(\mathbf{R})$ from Definition 3.3 and proved Theorem 3.4(ii) and (iv). In the same year (1888) Peano [3] introduced the first axiomatic treatment of real vector spaces (which he called "linear systems") of arbitrary dimension and linear operators on them; it was ahead of its time and largely forgotten, including by Peano himself. The main inspiration for the development of abstract linear algebra came from work on normed vector spaces by Banach in the 1920s [1].

Example 4.4. Let's compute the operator norm of the 2×2 matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Since

$$AA^\top = \begin{pmatrix} 5 & 11 \\ 11 & 25 \end{pmatrix},$$

the characteristic polynomial of $AA^\top$ is $X^2 - 30X + 4$, which has eigenvalues $15 \pm \sqrt{221} \approx$ .13, 29.86. Therefore the operator norm of A is $\sqrt{15 + \sqrt{221}}$, so for all $\binom{x}{y} \in \mathbf{R}^2$,

$$\left\| \begin{pmatrix} 5x + 11y \\ 11x + 25y \end{pmatrix} \right\| \leq \sqrt{15 + \sqrt{221}} \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|,$$

and $\sqrt{15 + \sqrt{221}} \approx 5.46$ is the smallest number with that property.

Computing the operator norm of A amounts to finding the largest eigenvalue of a related symmetric matrix ($AA^\top$ or $A^\top A$). In practice, for large symmetric matrices the largest eigenvalue is *not* computed by calculating the roots of its characteristic polynomial. More efficient algorithms for calculating eigenvalues are available (*e.g.*, QR algorithm or Lanczos algorithm).

Exercises.

1. For $A \in \mathrm{M}_{m,n}(\mathbf{R})$, show the eigenvalues of the $m \times m$ matrix $AA^\top$ and the $n \times n$ matrix $A^\top A$ are nonnegative. (The nonzero eigenvalues of these matrices are also equal. For all $A \in \mathrm{M}_{m,n}(\mathbf{R})$ and $B \in \mathrm{M}_{n,m}(\mathbf{R})$, the matrices $AB \in \mathrm{M}_m(\mathbf{R})$ and $BA \in \mathrm{M}_n(\mathbf{R})$ have the same nonzero eigenvalues.)
2. If you worked out properties of operator norms of rectangular matrices in Exercise 3.1, determine the operator norm of the 2×3 matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

3. If we use a norm on $\mathbf{R}^n$ other than the standard norm, the corresponding operator norm on $\mathrm{M}_n(\mathbf{R})$ will be different from the one we have worked with here. When $\mathbf{R}^n$ has a norm $\|\cdot\|'$, the related operator norm of a matrix $A \in \mathrm{M}_n(\mathbf{R})$ is the least $b \geq 0$ such that $\|Av\|' \leq b \|v\|'$ for all $v \in \mathbf{R}^n$. Let's look at an example. Give $\mathbf{R}^n$ the sup-norm, so the associated operator norm of an $n \times n$ matrix A is the least b satisfying $\|Av\|_{\sup} \leq b \|v\|_{\sup}$ for all $v \in \mathbf{R}^n$. What is b?
    (a) For all $v \in \mathbf{R}^n$, show $\|Av\|_{\sup} \leq b \|v\|_{\sup}$ where $b = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$.
    (b) For each row $(a_{i1}, \ldots, a_{in})$ of A, show there is a vector v_i in $\mathbf{R}^n$ with coordinates from $\{\pm 1\}$ such that the ith entry of Av equals $|a_{i1}| + \cdots + |a_{in}|$. Conclude that there is a $v \in \mathbf{R}^n$ such that $\|Av\|_{\sup} = b \|v\|_{\sup}$ where b is the number in part (a). Therefore the operator norm of A when using the sup-norm on $\mathbf{R}^n$ is b.
4. Generalize Theorem 4.1 to give a computational formula for the operator norm of matrices in $\mathrm{M}_n(\mathbf{C})$.

Appendix A. Proof of Cauchy-Schwarz inequality

We give a proof of the Cauchy–Schwarz inequality that was found by Schwarz [4, p. 344] in 1888. It is a clever trick with quadratic polynomials and the context in which Schwarz discovered it is described in [5, pp. 10–11]. (The whole book [5] is recommended as a lively account of fundamental inequalities in mathematics.)

Let V be a real vector space with an inner product $(\cdot, \cdot)$. Pick v and w in V. Our goal is to show

$$|(v, w)| \leq \|v\| \|w\|.$$

This is obvious if v or w is 0, so assume both are nonzero. For all $t \in \mathbf{R}$, $(v+tw, v+tw) \geq 0$. The left side can be expanded to be a quadratic polynomial in t:

$$
\begin{aligned}
(v + tw, v + tw) &= (v, v) + (v, tw) + (tw, v) + (tw, tw) \\
&= (v, v) + t(v, w) + t(w, v) + t^2(w, w) \\
&= \|v\| + 2(v, w)t + \|w\| t^2.
\end{aligned}
$$

This is quadratic since $\|w\| > 0$. A quadratic polynomial in t has nonnegative values for all t if and only if its discriminant is ≤ 0, so

$$
(2(v, w))^2 - 4\|v\|\|w\| \leq 0,
$$

which is equivalent to $|(v, w)| \leq \|v\|\|w\|$, and that completes the proof!

Remark A.1. We have equality $|(v, w)| = \|v\|\|w\|$ if and only if the quadratic polynomial above has a double real root t, and at that root we get $(v+tw, v+tw) = 0$, so $v+tw = 0$ in V and thus v and w are linearly independent. The converse direction, that linear dependence implies equality in the Cauchy–Schwarz inequality, is left to the reader (also in the case that v or w is 0).

References

[1] G. H. Moore, The Axiomatization of Linear Algebra: 1875–1940, *Historia Mathematica* **22** (1995), 262–303. Online at https://core.ac.uk/download/pdf/82128888.pdf.
[2] G. Peano, Intégration par séries des équations différentielles linéaires, *Mathematische Annalen* **32** (1888), 450–456.
[3] G. Peano, "Calcolo Geometrico secondo l'Ausdehnungslehre di H. Grassmann, preceduto dalle operazioni della logica deduttiva," Fratelli Bocca, Turin, 1888. Online at http://mathematica.sns.it/opere/138/.
[4] H. A. Schwarz, Über ein die Flächen kleinsten Flächeninhalts betreffendes Problem der Variationsrechnung, *Acta Soc. Scient. Fenn.* **15** (1888), 315–362.
[5] J. M. Steele, "The Cauchy–Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities," Cambridge Univ. Press, Cambridge, 2004.