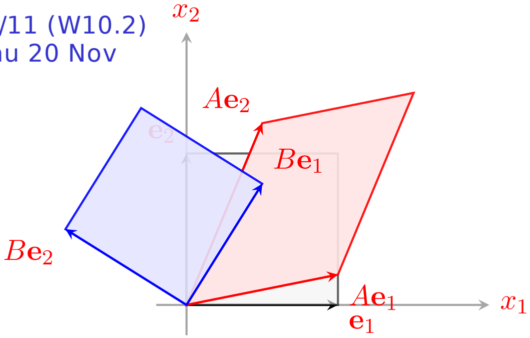
MA385 Part 4: Linear Algebra 2

4.2: Matrix Norms

Dr Niall Madden

November 2025

Started Thu 13/11 (W10.2) and finished Thu 20 Nov (W11.2)



$\mathbf{x} \mapsto A\mathbf{x}$ (red), $\mathbf{x} \mapsto B\mathbf{x}$ (blue)

1. Outline Section 4.2

For more, see Section 2.7 of Suli and Mayers:
https://ebookcentral.proquest.com/lib/nuig/reader.action?docID=221072&ppg=51&c=UERG

Vector norms are related to the magnitude of the entries of the vector.
Now we want to generalise to the concept of a **matrix norm**. In a sense, we can just consider the magnitude of the matrix's entries.

However, if we think of a matrix as a linear transformation, or simply as a function that maps (via matrix multiplication) from $\mathbb{R}^n$ to $\mathbb{R}^n$, we should think about how much it changes a vector.

Definition 4.2.1

Given any (vector) norm $\|\cdot\|$ on $\mathbb{R}^n$, there is a **subordinate matrix norm** on $\mathbb{R}^{n\times n}$ defined by

$$\|A\| = \max_{\mathbf{v}\in\mathbb{R}^n_\star} \frac{\|A\mathbf{v}\|}{\|\mathbf{v}\|}, = \max_{\substack{v\in\mathbb{R}^n \\ \|v\|=1}} \|Av\| \quad (1)$$

where $A \in \mathbb{R}^{n\times n}$ and $\mathbb{R}^n_\star = \mathbb{R}^n/\{\mathbf{0}\}$.

We define a matrix norm like this because we think of A as an *operator* on $\mathbb{R}^n$: if $\mathbf{v} \in \mathbb{R}^n$ then $\mathbf{A}v \in \mathbb{R}^n$. So the norm of A gives us information on how much the matrix can change the size of a vector.

"induced Matrix Norm" "Operator Norm".

3. Computing Matrix Norms

It is not obvious from the above definition how to calculate the norm of a given matrix. We'll see that

- The ∞-norm of a matrix is also the largest absolute-value row sum.
- The 1-norm of a matrix is also the largest absolute-value column sum.
- The 2-norm of the matrix A is the square root of the largest eigenvalue of $A^T A$.

Finished here W10.2. 13 Nov.

Theorem 4.2.1

For any $A \in \mathbb{R}^{n\times n}$ the subordinate matrix norm associated with $\|\cdot\|_\infty$ on $\mathbb{R}^n$ can be computed by

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

Eg

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & -5 \\ 4 & -1 & -2 \end{pmatrix}$$

Then

$$\max_{i=1\ldots3} \sum_{j=1}^{3} |a_{ij}| = \max \left\{ \begin{array}{ccc} i=1 & i=2 & i=3 \\ 1+2+3, & 1+0+5, & 4+1+2 \\ 6 & 6 & 7 \end{array} \right\}$$

So $\|A\|_\infty = 7$

4. The max-norm on $\mathbb{R}^{n\times n}$

Proof: Let v be any vector in $\mathbb{R}^n/\{0\}$ and let $K = \|v\|_\infty = \max_i |v_i|$

The

$$(Av)_i = \sum_{j=1}^{n} a_{ij} v_j$$

So $$|(Av)_i| = \left|\sum_{j=1}^{n} a_{ij} v_j\right| \leq \sum_{j=1}^{n} |a_{ij}| \cdot \overbrace{|v_j|}^{\leq K}$$

by the Triangle Inequality

$$|(Av)_i| \leq K \sum_{j=1}^{n} |a_{ij}|$$

So $$\max_i |(Av)_i| \leq K \max_i \sum_{j=1}^{n} |a_{ij}|.$$

That is

$$\|Av\|_\infty \leq K \max_i \sum_j |a_{ij}|$$

So, since $K = \|v\|_\infty \neq 0$

$$\frac{\|Av\|_\infty}{\|v\|_\infty} \leq \max_i \sum_j |a_{ij}|.$$ for any vector v. Therefore

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

To finish we need a vector v such that $\|Av\|_\infty = \max_i \sum_j |a_{ij}|$ & $\|v\|_\infty = 1$.

Say $\|A\|_\infty = \sum_j |a_{ij}|$ for a given i.

4. The max-norm on $\mathbb{R}^{n\times n}$

Let v be the vector with all entries $+1$ or -1, and

$$v_j = \text{sign}(a_{ij})$$

$$(\text{so} \quad a_{ij} v_j \geq 0).$$

A similar result holds for the 1-norm, the proof of which is left as an exercise.

Theorem 4.2.2

For any $A \in \mathbb{R}^{n\times n}$ the subordinate matrix norm associated with $\|\cdot\|_1$ on $\mathbb{R}^n$ can be computed by

$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{n} |a_{i,j}|. \qquad (2)$$

Computing the 2-norm of a matrix is a little harder that computing the 1- or ∞-norms. However, later we'll need estimates not just for $\|A\|$, but also $\|A^{-1}\|$. And, unlike the 1- and ∞-norms, we can estimate $\|A^{-1}\|_2$ without explicitly forming A^{-1}.

We begin by recalling some important facts about eigenvalues and eigenvectors.

Definition 4.2.2

Let $A \in \mathbb{R}^{n\times n}$. We call $\lambda \in \mathbb{C}$ an *eigenvalue* of A if there is a non-zero vector $\mathbf{x} \in \mathbb{C}^n$ such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

We call any such $\mathbf{x}$ an *eigenvector of A associated with λ*.

(i) If A is a real symmetric matrix (i.e., $A = A^T$), its eigenvalues and eigenvectors are all real-valued.

(ii) If λ is an eigenvalue of A, the $1/\lambda$ is an eigenvalue of A^{-1}.

(iii) If $\mathbf{x}$ is an eigenvector associated with the eigenvalue λ then so too is $\eta\mathbf{x}$ for any non-zero scalar η.

(iv) An eigenvector may be *normalised* as $\|\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{x} = 1$.

If $Ax = \lambda x$ and A^{-1} exists,

then $\underbrace{A^{-1}A}_{I}x = \lambda A^{-1}x$

$\Rightarrow x = \lambda A^{-1}x \Rightarrow A^{-1}x = \frac{1}{\lambda}x$

(v) There are n eigenvectors λ_1, λ_2, $\ldots$, λ_n associated with the real symmetric matrix A. Let $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, ..., $\mathbf{x}^{(n)}$ be the associated normalised eigenvectors. The eigenvectors are linearly independent and so form a basis for $\mathbb{R}^n$. That is, any vector $\mathbf{v} \in \mathbb{R}^n$ can be written as a linear combination:

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{x}^{(i)}.$$

(vi) Furthermore, these eigenvectors are *orthogonal* and *orthonormal*:

$$\left(\mathbf{x}^{(i)}\right)^T \mathbf{x}^{(j)} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Here is a useful consequence of (v) and (vi), which we will use repeatedly.

$$Ax^{(i)} = \lambda_i x^{(i)}$$

(v) $$v = \sum_{i=1}^{n} \alpha_i x^{(i)}$$

(vi) $$\left(x^{(i)}\right)^T x^{(j)} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Then

$$v^T v = \left(\sum_{i=1}^{n} \alpha_i x^{(i)}\right)^T \sum_{j=1}^{n} \alpha_j x^{(j)}$$

$$= \sum_{i=1}^{n} \alpha_i^2$$

The *singular values* of a matrix A are the square roots of the eigenvalues of $A^T A$. They play a very important role in matrix analysis, applied linear algebra, and statistics (principal component analysis).

Our interest here is in their relationship to $\|A\|_2$.

But first we'll prove a theorem about certain matrices (so called, "normal matrices").

Theorem 4.2.3

For any matrix $A \in \mathbb{R}^{n\times n}$, the eigenvalues of A^TA are real and non-negative.

Let $B = A^TA$. Then $B^T = (A^TA)^T = A^T(A^T)^T$

So $B^T = A^TA = B$. So B is symmetric

So any eigenvalue of B is real valued.

Let $Bx = \lambda x$.

So $(A^TA)x = \lambda x$

$\Rightarrow x^T(A^TA)x = \lambda x^Tx \Rightarrow (x^TA^T)(Ax) = \lambda x^Tx$.

$\Rightarrow (Ax)^TAx = \lambda x^Tx \Rightarrow \lambda = \dfrac{\|Ax\|_2^2}{\|x\|_2^2} \geq 0$

Part of the above proof involved showing that, if $(A^T A)\mathbf{x} = \lambda x$, then

$$\sqrt{\lambda} = \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}.$$

This at the very least tells us that

$$\|A\|_2 := \max_{\mathbf{x}\in\mathbb{R}^n_\star} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \geq \max_{i=1,\dots,n} \sqrt{\lambda_i}.$$

With a bit more work, we can show that if $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the the eigenvalues of $B = A^T A$, then

$$\|A\|_2 = \sqrt{\lambda_n}.$$

Theorem 4.2.4

Let $A \in \mathbb{R}^{n\times n}$. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, be the eigenvalues of $B = A^T A$. Then

$$\|A\|_2 = \max_{i=1,\dots,n} \sqrt{\lambda_i} = \sqrt{\lambda_n},$$

Let $Bx^{(i)} = \lambda_i \mathbf{x}^{(i)}$ for $i = 1, \dots, n$. That is, λ_i is an eigenvalue of B, with corresponding eigenvector $\mathbf{x}^{(i)}$. We may assume that the $\mathbf{x}^{(i)}$ are orthogonal and normalised so that

$$(\mathbf{x}^{(i)})^T x^{(j)} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The set $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$ forms a basis for $\mathbb{R}^n$.

Therefore, we can write any $\mathbf{v} \in \mathbb{R}^n$ as

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{x}^{(i)}.$$

Then

$$A^T A\mathbf{v} = B\mathbf{v} = B\Big(\sum_{i=1}^{n} \alpha_i \mathbf{x}^{(i)}\Big) = \sum_{i=1}^{n} \alpha_i B\mathbf{x}^{(i)} = \sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{x}^{(i)}.$$

Next, note that

$$\|A\mathbf{v}\|_2^2 = (A\mathbf{v})^T A\mathbf{v} = \mathbf{v}^T(A^T A\mathbf{v}) = \overbrace{\left(\sum_{i=1}^{n} \alpha_i \cancel{\lambda_i} \mathbf{x}^{(i)}\right)}^{\mathbf{v}}{}^T \left(\sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{x}^{(i)}\right).$$

Because the $\mathbf{x}^{(i)}$ are orthonormal and orthogonal, this simplifies to

$$\|A\mathbf{v}\|_2^2 = \sum \lambda_i \alpha_i^2 \leq \cancel{\lambda_n \|\mathbf{v}\|_2^2} \leq \lambda_n \sum_{i=1}^{n} \alpha_i^2 = \lambda_n \|\mathbf{v}\|_2^2$$

It follows that, for any vector $\mathbf{v}$,

$$\frac{\|A\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \leq \sqrt{\lambda_n}.$$

In addition

$$\frac{\|A\mathbf{x}^{(n)}\|_2}{\|\mathbf{x}^{(n)}\|_2} = \sqrt{\lambda_n}.$$

Therefore,

$$\|A\|_2 := \max_{\mathbf{v}\in\mathbb{R}^n_\star} \frac{\|A\mathbf{v}\|_2}{\|\mathbf{v}\|_2} = \sqrt{\lambda_n}.$$

It follows that, for any vector $\mathbf{v}$,

$$\frac{\|A\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \leq \sqrt{\lambda_n}.$$

In addition

$$\frac{\|A\mathbf{x}^{(n)}\|_2}{\|\mathbf{x}^{(n)}\|_2} = \sqrt{\lambda_n}.$$

Therefore,

$$\|A\|_2 := \max_{\mathbf{v}\in\mathbb{R}^n_\star} \frac{\|A\mathbf{v}\|_2}{\|\mathbf{v}\|_2} = \sqrt{\lambda_n}.$$