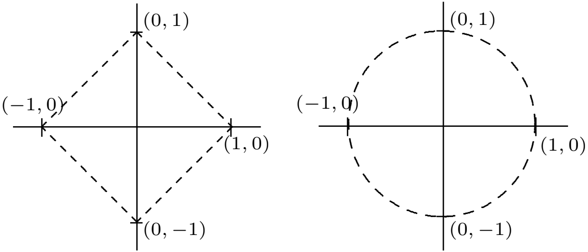
§3 Solving linear systems

§3.5 Vector and Matrix Norms

MA385 – Numerical Analysis 1

November 2019



All computer implementations of algorithms that involve floating-point numbers (roughly, finite decimal approximations of real numbers) contain errors due to round-off error.

It transpires that computer implementations of LU-factorization, and related methods, lead to these round-off errors being greatly magnified: this phenomenon is the main focus of this final section of the course.

You might remember from earlier sections of the course that we had to assume functions where well-behaved in the sense that

$$\frac{|f(x) - f(y)|}{|x - y|} \leq L,$$

for some number L, so that our numerical schemes (e.g., fixed point iteration, Euler's method, etc) would work. If a function *doesn't* satisfy a condition like this, we say it is "ill-conditioned".

One of the consequences is that a small error in the inputs gives a large error in the outputs.

We'd like to be able to express similar ideas about matrices: that $A(u - v) = Au - Av$ is not too "large" compared to $u - v$. To do this we used the notion of a "norm" to describing the relatives sizes of the vectors $\boldsymbol{u}$ and $A\boldsymbol{u}$.

When we want to consider the size of a real number, without regard to sign, we use the *absolute value*. Important properties of this function are:

1. $|x| \geq 0$ for all x.
2. $|x| = 0$ if and only if $x = 0$.
3. $|\lambda x| = |\lambda||x|$.
4. $|x + y| \leq |x| + |y|$ (triangle inequality).

This notion can be extended to vectors and matrices.

Definition 3.18

Let $\mathbb{R}^n$ be all the vectors of length n of real numbers. The function $\|\cdot\|$ is called a **norm** of $\mathbb{R}^n$ if, for all $u, v \in \mathbb{R}^n$

1. $\|v\| \geq 0$,
2. $\|v\| = 0$ if and only if $v = 0$.
3. $\|\lambda v\| = |\lambda|\|v\|$ for any $\lambda \in \mathbb{R}$,
4. $\|u + v\| \leq \|u\| + \|v\|$ (triangle inequality).

Norms on vectors in $\mathbb{R}^n$ quantify the *size* of the vector. But there are different ways of doing this...

Definition 3.19

Let $\boldsymbol{v} \in \mathbb{R}^n$: $\boldsymbol{v} = (v_1, v_2, \ldots, v_{n-1}, v_n)^T$.

(i) The 1-*norm* (a.k.a. the *Taxi cab* norm) is $\|v\|_1 = \sum_{i=1}^{n} |v_i|$.

(ii) The 2-*norm* (a.k.a. the *Euclidean norm*) $\|v\|_2 = \Big(\sum_{i=1}^{n} v_i^2\Big)^{1/2}$.

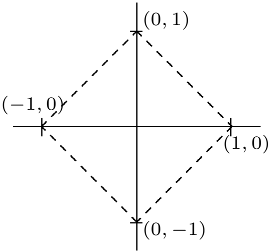
Note, if $\boldsymbol{v}$ is a vector in $\mathbb{R}^n$, then

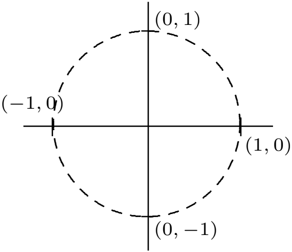
$$\boldsymbol{v}^T\boldsymbol{v} = v_1^2 + v_2^2 + \cdots + v_n^2 = \|\boldsymbol{v}\|_2^2.$$

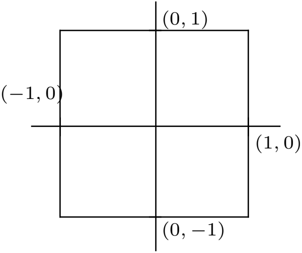
(iii) The ∞-*norm* (a.k.a. the $\max$-*norm*) $\|v\|_\infty = \max_{i=1}^{n} |v_i|$.

Example: $\boldsymbol{v} = (-2, 4, -4)$

The unit balls in $\mathbb{R}^2$ given by $\|\cdot\|_1$ (top left), $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2} = 1$ (top right), and $\|\cdot\|_\infty$.







It is easy to show that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are norms. And it is not hard to show that $\|\cdot\|_2$ satisfies conditions (1), (2) and (3) of Definition 3.18.

It takes a little bit of effort to show that $\|\cdot\|_2$ satisfies the triangle inequality; details are given in Section 3.5.9 of the notes.

Definition 3.20

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

$$\|A\| = \max_{\boldsymbol{v}\in\mathbb{R}^n_\star} \frac{\|A\boldsymbol{v}\|}{\|\boldsymbol{v}\|}, \tag{7}$$

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

You might wonder why we define a matrix norm like this. The reason is that we like to think of A as an *operator* on $\mathbb{R}^n$: if $\boldsymbol{v} \in \mathbb{R}^n$ then $\boldsymbol{A}v \in \mathbb{R}^n$. So rather than the norm giving us information about the "size" of the entries of a matrix, it tells us how much the matrix can change the size of a vector.

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 $A^T A$.

Theorem 3.21

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

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

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

Theorem 3.22

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

$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{n} |a_{i,j}|. \tag{8}$$

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 3.23

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 $\boldsymbol{x} \in \mathbb{C}^n$ such that

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

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

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

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

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

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

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

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

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

The *singular values* of a matrix A are the square roots of the eigenvalues of A^TA. They play a very important role in matrix analysis and in areas of applied linear algebra, such as image and text processing. 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 3.24

For any matrix A, the eigenvalues of $A^T A$ are real and non-negative.

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

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

This at the very least tells us that

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

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 3.25

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

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

Here is the main idea. For full details, see the text-book.

Exercise 3.12 ($\star$)

Show that, for any vector $\boldsymbol{x} \in \mathbb{R}^n$, $\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{x}\|_1 \|\boldsymbol{x}\|_\infty$. For each of these inequalities, give an example for which the equality holds. Deduce that $\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_1$.

Exercise 3.13

Show that if $x \in \mathbb{R}^n$, then $\|\boldsymbol{x}\|_1 \leq n\|\boldsymbol{x}\|_\infty$ and that $\|\boldsymbol{x}\|_2 \leq \sqrt{n}\|\boldsymbol{x}\|_\infty$.

Exercise 3.14

Show that, for *any* subordinate matrix norm on $\mathbb{R}^{n \times n}$, the norm of the identity matrix is 1.

Exercise 3.15 ($\star$)

Prove that

$$\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |a_{i,j}|.$$

Hint: Suppose that

$$\sum_{i=1}^{n} |a_{ij}| \leq C, \qquad j = 1, 2, \ldots n,$$

show that for *any* vector $\boldsymbol{x} \in \mathbb{R}^n$

$$\sum_{i=1}^{n} |(A\boldsymbol{x})_i| \leq C\|\boldsymbol{x}\|_1.$$

Now find a vector $\boldsymbol{x}$ such that $\sum_{i=1}^{n} |(A\boldsymbol{x})_i| = C\|\boldsymbol{x}\|_1$. Now deduce the result.

As mentioned on Slide 59, it takes a little effort to show that $\|\cdot\|_2$ is indeed a norm on $\mathbb{R}^2$; in particular to show that it satisfies the triangle inequality, we need the Cauchy-Schwarz inequality.

Lemma 1 (Cauchy-Schwarz)

$$|\sum_{i=1}^{n} u_i v_i| \leq \|\boldsymbol{u}\|_2 \|\boldsymbol{v}\|_2, \qquad \forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n.$$

The proof can be found in any text-book on analysis.

Now can now apply Cauchy-Schwatz to show that

$$\|\boldsymbol{u} + \boldsymbol{v}\|_2 \leq \|\boldsymbol{u}\|_2 + \|\boldsymbol{v}\|_2.$$

(PTO)

This is because

$$\begin{aligned}
\|u+v\|_2^2 &= (u+v)^T(u+v) \\
&= u^Tu + 2u^Tv + v^Tv \\
&\leq u^Tu + 2|u^Tv| + v^Tv \quad \text{(by the triangle-inequality)} \\
&\leq u^Tu + 2\|u\|\|v\| + v^Tv \quad \text{(by Cauchy-Schwarz)} \\
&= (\|u\| + \|v\|)^2.
\end{aligned}$$

It follows directly that

Corollary 2

$\|\cdot\|_2$ *is a norm.*