MATH3511/6111: Scientific Computing

08. Norms

Lindon Roberts

Semester 1, 2022

Office: Hanna Neumann Building #145, Room 4.87
Email: lindon.roberts@anu.edu.au
Based on lecture notes written by S. Roberts, L. Stals, Q. Jin, M. Hegland, K. Duru.

Australian National University

Measuring Errors

So far, we have looked at algorithms for solving problems where the answer is a single real number:

- Rootfinding: find $x \in \mathbb{R}$ such that $f(x) \approx 0$.
- Approximation: find $y \in \mathbb{R}$ such that $y \approx f(x)$ (interpolation or splines)
- Differentiation: find $d \in \mathbb{R}$ such that $d \approx f'(x)$
- Integration/quadrature: find $Q \in \mathbb{R}$ such that $Q \approx \int_a^b f(x)dx$.

In all cases, we need to measure how far away our estimates are.

So far, we have measured errors using the absolute value function. For example, for the trapezoidal rule with $|f''(x)| \le M$ we had

$$\left| \underbrace{\int_a^b f(x)dx}_{\text{truth}} - \underbrace{\sum_{i=0}^{n-1} \frac{f(x_i) + f(x_{i+1})}{2} h}_{\text{approx.}} \right| \le \frac{M}{12}(b-a)h^2.$$

Measuring Errors in $\mathbb{R}$

The absolute value is a useful way to measure the size of numbers on $\mathbb{R}$:

$$|x| = \text{“Size of } x\text{”}.$$

This function has some useful properties:

- Positivity: $|x| \geq 0$ for all $x \in \mathbb{R}$.
- Scaling: $|cx| = |c| \cdot |x|$ for all $x \in \mathbb{R}$ and any scaling factor $c \in \mathbb{R}$.
- Triangle inequality: $|x + y| \leq |x| + |y|$ for all $x, y \in \mathbb{R}$.

In our case, we usually have an approximation $\widetilde{x}$ to a true value x, and we measure the size of the error as $|e|$ for $e = x - \widetilde{x}$.

Measuring Errors: General Case

Question

What happens if we are doing approximation of vectors or functions? How do we measure the "size of the error" for these?

We need a general way to define the "size" of a mathematical object. In our case, we only need to think about objects in vector spaces (e.g. vectors, matrices, functions).

Definition

If V is a vector space, a function $\|\cdot\| : V \to \mathbb{R}$ is called a norm if

- $\|\boldsymbol{v}\| \geq 0$ for all $\boldsymbol{v} \in V$.
- $\|\boldsymbol{v}\| = 0$ if and only if $\boldsymbol{v} = \boldsymbol{0}$.
- $\|c\boldsymbol{v}\| = |c| \cdot \|\boldsymbol{v}\|$ for all $\boldsymbol{v} \in V$ and all $c \in \mathbb{R}$.
- $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$.

Note we are using bold lower-case letters to represent vectors in V.

Measuring Errors: General Case

Any norm can be used as a measure of distance between vectors in V.

Suppose $\widetilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an approximation of $\boldsymbol{x} \in \mathbb{R}^n$. Just like $|\widetilde{x} - x|$ is the distance between $\widetilde{x}$ and x in $\mathbb{R}$, for a given norm $\|\cdot\|$, we define the absolute error to be

$$\text{error}_{\text{abs}} = \|\widetilde{\boldsymbol{x}} - \boldsymbol{x}\|$$

and the relative error

$$\text{error}_{\text{rel}} = \frac{\|\widetilde{\boldsymbol{x}} - \boldsymbol{x}\|}{\|\boldsymbol{x}\|}.$$

Note that the value of the error will depend on the choice of norm (there are many different norms on $\mathbb{R}^n$, which we will see shortly).

Common Norms on $\mathbb{R}^n$

There are several common norms for $\mathbb{R}^n$.

Definition

The ℓ_1 norm on $\mathbb{R}^n$ is given by

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|,$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$.

For example:

$$\left\| \begin{bmatrix} -1 \\ 2 \\ -4 \end{bmatrix} \right\|_1 = 1 + 2 + 4 = 7.$$

Proof for ℓ_1 norm

Let's prove that $\|\cdot\|_1$ is a norm on $\mathbb{R}^2$.

$\boxed{\|\boldsymbol{x}\|_1 \geq 0}$

Since $\|\boldsymbol{x}\|_1 = |x_1| + |x_2|$, this is follows from $|x_1| \geq 0$ and $|x_2| \geq 0$.

$\boxed{\|\boldsymbol{x}\|_1 = 0 \quad \Leftrightarrow \quad \boldsymbol{x} = \mathbf{0}}$

First, if $\boldsymbol{x} = \mathbf{0}$, clearly $\|\boldsymbol{x}\|_1 = 0$.

Now suppose $\|\boldsymbol{x}\|_1 = 0$ for some $\boldsymbol{x} \in \mathbb{R}^2$. This means that $|x_1| + |x_2| = 0$, and so $|x_1| = 0$ and $|x_2| = 0$. Hence $x_1 = x_2 = 0$ and so $\boldsymbol{x} = \mathbf{0}$.

Proof for ℓ_1 norm

$\boxed{\|c\boldsymbol{x}\|_1 = |c| \cdot \|\boldsymbol{x}\|_1}$

Choose $c \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^2$. Then

$$\|c\boldsymbol{x}\|_1 = |cx_1| + |cx_2| = |c| \cdot |x_1| + |c| \cdot |x_2| = |c| \left(|x_1| + |x_2|\right) = |c| \cdot \|\boldsymbol{x}\|_1.$$

$\boxed{\|\boldsymbol{x} + \boldsymbol{y}\|_1 \leq \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1}$

Choose $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$. Then using the triangle inequality for the absolute value function, we get

$$\|\boldsymbol{x} + \boldsymbol{y}\|_1 = |x_1 + y_1| + |x_2 + y_2| \leq \left(|x_1| + |y_1|\right) + \left(|x_2| + |y_2|\right) = \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1.$$

And so $\|\cdot\|_1$ is a norm on $\mathbb{R}^2$.

Common Norms on $\mathbb{R}^n$

There are several common norms for $\mathbb{R}^n$.

Definition

The ℓ_∞ norm on $\mathbb{R}^n$ is given by

$$\|\boldsymbol{x}\|_\infty = \max(|x_1|, |x_2|, \dots, |x_n|),$$

where $\boldsymbol{x} = [x_1, x_2, \dots, x_n]^T$.

For example:

$$\left\| \begin{bmatrix} -1 \\ 2 \\ -4 \end{bmatrix} \right\|_\infty = \max(1, 2, 4) = 4.$$

Common Norms on $\mathbb{R}^n$

Perhaps the norm you are most familiar with is the usual Euclidean norm.

Definition

The ℓ_2 norm on $\mathbb{R}^n$ is given by

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2},$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$.

For example:

$$\left\| \begin{bmatrix} -1 \\ 2 \\ -4 \end{bmatrix} \right\|_2 = \sqrt{1+4+16} \approx 4.5826.$$

Common Norms on $\mathbb{R}^n$

All of these are examples of Hölder ℓ_p norms (Otto Hölder, 1889).

Definition

The ℓ_p norm on $\mathbb{R}^n$ is given by

$$\|\boldsymbol{x}\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p},$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$. It is a norm for all $p \geq 1$ (not just integers).

If $p = 1, 2$, we get the ℓ_1 and ℓ_2 norms defined above. Note that the ℓ_∞ is called this because

$$\|\boldsymbol{x}\|_p \to \|\boldsymbol{x}\|_\infty \qquad \text{as } p \to \infty.$$

Note: sometimes we abbreviate the ℓ_p norm to the "p-norm".

More on the ℓ_2 norm

Note that the ℓ_2 norm can be defined in terms of the usual inner product (dot product) on $\mathbb{R}^n$:

$$\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = \sqrt{\boldsymbol{x}\cdot\boldsymbol{x}}.$$

So, if $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix ($Q^TQ = I$, e.g. rotation or reflection), we have

$$\|Q\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^TQ^TQ\boldsymbol{x}} = \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = \|\boldsymbol{x}\|_2.$$

So, orthogonal matrices do not change the 2-norm (we say the 2-norm is invariant under orthogonal transformations).

Hölder's Inequality

An important relationship between the p-norms is Hölder's inequality:

Theorem (Hölder's inequality)

If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $p \geq 1$, then

$$|\boldsymbol{x}^T \boldsymbol{y}| \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q, \qquad \text{where } \frac{1}{p} + \frac{1}{q} = 1.$$

Note: if $p = 1$ then $q = \infty$ (and vice versa)

In the case where $p = q = 2$, this is the Cauchy-Schwarz inequality (Augustin-Louis Cauchy 1821, Hermann Schwarz, 1888):

$$|\boldsymbol{x}^T \boldsymbol{y}| \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2.$$

Relationships between ℓ_p norms

There are some important inequalities between the most important ℓ_p norms ($p = 1, 2, \infty$). For all $\boldsymbol{x} \in \mathbb{R}^n$,

$$\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_1,$$

plus

$$\|\boldsymbol{x}\|_1 \leq \sqrt{n}\|\boldsymbol{x}\|_2 \leq n\|\boldsymbol{x}\|_\infty.$$

When we are using norms to measure errors, this means that if $\|\widetilde{\boldsymbol{x}} - \boldsymbol{x}\| \to 0$ in any of these norms, then it approaches zero in all other norms too.

Matrix Norms

Just like with $\mathbb{R}^n$, there are many different norms for matrices. The most important for us will be the operator norms.

Here, we think about $m \times n$ matrices as linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, and compare the norm of the input vs. norm of the output.

Definition (Matrix operator norm)

Suppose we have norms $\|\cdot\|_n$ on $\mathbb{R}^n$ and $\|\cdot\|_m$ on $\mathbb{R}^m$. If $A \in \mathbb{R}^{m\times n}$, the matrix operator norm induced by $\|\cdot\|_n$ and $\|\cdot\|_m$ is

$$\|A\| = \max\left\{\frac{\|A\boldsymbol{x}\|_m}{\|\boldsymbol{x}\|_n} : \boldsymbol{x} \in \mathbb{R}^n, \ \boldsymbol{x} \neq \boldsymbol{0}\right\}.$$

Matrix Norms

Most commonly, we will use the vector p-norms on both $\mathbb{R}^n$ and $\mathbb{R}^m$ (same p for both).

Definition (Matrix p-norm)

If $A \in \mathbb{R}^{m \times n}$, the matrix p-norm is

$$\|A\|_p = \max\left\{\frac{\|A\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p} : \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{x} \neq \boldsymbol{0}\right\}.$$

Note we are a bit vague with notation: $\|\cdot\|_p$ can either be a vector p-norm or matrix p-norm (which one should be clear from context).

Matrix Norms

The matrix p-norms satisfy all the usual norm properties (from the definition). It also has some extra useful properties.

Theorem

Suppose $A \in \mathbb{R}^{m \times n}$ and $p \geq 1$ (including $p = \infty$). Then

- *For all $\boldsymbol{x} \in \mathbb{R}^n$, $\|A\boldsymbol{x}\|_p \leq \|A\|_p \|\boldsymbol{x}\|_p$.*
- *There exists $\boldsymbol{x}^* \in \mathbb{R}^n$ such that $\|\boldsymbol{x}^*\|_p = 1$ and $\|A\|_p = \|A\boldsymbol{x}^*\|_p$.*
- *If $B \in \mathbb{R}^{n \times q}$, then $\|AB\|_p \leq \|A\|_p \|B\|_p$.*

Proof

The first result follows straight from the definition of $\|A\|_p$.

The second result comes from rewriting

$$\max\left\{\frac{\|A\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p} : \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{x} \neq \mathbf{0}\right\} = \max\left\{\|A\boldsymbol{x}\|_p : \|\boldsymbol{x}\|_p = 1\right\}.$$

This means we are trying to maximise $\|A\boldsymbol{x}\|_p$ (a continuous function of $\boldsymbol{x}$) over a compact (closed and bounded) set $\|\boldsymbol{x}\|_p = 1$. Therefore the maximum is attained for some $\boldsymbol{x}^*$.

The third result is an exercise.

Computing Matrix Norms

For some values of p, computing the matrix p-norm is quite easy.

Theorem

Let the (i,j) entry of $A \in \mathbb{R}^{m\times n}$ be $a_{i,j}$. Then

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

and

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

The 1-norm of a matrix is the maximum absolute column sum.

The ∞-norm of a matrix is the maximum absolute row sum.

Computing Matrix Norms

For example, take the matrix:

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 1 & 3 & 5 \\ -4 & 6 & 6 \end{bmatrix}.$$

The absolute column sums are 6, 12 and 13, so $\|A\|_1 = 13$.

The absolute row sums are 6, 9 and 16, so $\|A\|_\infty = 16$.

Frobenius Norm

Question

The only matrix norms we have seen are operator norms. Are there any norms based just on the entries of the matrix?

Yes! The most common is the Frobenius norm (Ferdinand Frobenius, late 1800s(?))

Definition (Frobenius norm)

If $A \in \mathbb{R}^{m \times n}$, its Frobenius norm is

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j}^2}.$$

It turns out that this norm can be calculated from an inner product on matrices:

$$\|A\|_F = \sqrt{(A, A)_F}, \qquad \text{where} \qquad (A, B)_F = \operatorname{trace}(A^T B).$$

2-norm and Frobenius noorm

The matrix 2-norm and Frobenius norm have several similarities. Both are invariant to transposition and multiplication by orthogonal matrices:

$$\|A\|_2 = \|A^T\|_2, \qquad \text{and} \qquad \|A\|_F = \|A^T\|_F,$$

plus

$$\|Q_1 A Q_2\|_2 = \|A\|_2, \qquad \text{and} \qquad \|Q_1 A Q_2\|_F = \|A\|_F,$$

for any orthogonal matrices $Q_1 \in \mathbb{R}^{m\times m}$ and $Q_2 \in \mathbb{R}^{n\times n}$.

They also have a submultiplicative relationship:

$$\|AB\|_F \le \|A\|_2 \|B\|_F \qquad \text{and} \qquad \|AB\|_F \le \|A\|_F \|B\|_2.$$

So, often the Frobenius norm is useful because it is easy to compute (directly from the entries of A), and it can be used instead of the 2-norm.

Computing the matrix 2-norm

We have given simple ways to compute the matrix 1-, ∞- and Frobenius norms in terms of the entries of A. What about $\|A\|_2$?

There is not a simple formula for computing $\|A\|_2$ from its entries. It can be computed in two ways:

- Compute the largest eigenvalue of $A^T A$ (which will be real and ≥ 0 since it is symmetric positive semidefinite). Then $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$.
- Compute the singular value decomposition (SVD) of A. Then $\|A\|_2 = \sigma_1$, the largest singular value.

Singular Value Decomposition

The singular value decomposition is one of the most important linear algebra results. It says that every matrix $A \in \mathbb{R}^{m\times n}$ can be written as

$$A = U\Sigma V^T,$$

where $U \in \mathbb{R}^{m\times m}$ is orthogonal, $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with sorted entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ and $V \in \mathbb{R}^{n\times n}$ is orthogonal.

If you know this factorisation, you can easily know many important things about A:

- Rank of A (number of nonzero σ_i).
- An orthonormal basis for $\text{col}(A)$ and $\text{nul}(A)$ (columns of U and V).
- A unit vector $\boldsymbol{x}^*$ for which $\|A\boldsymbol{x}^*\|_2 = \|A\|_2$ (first column of V).
- An easy way to compute the best possible low-rank approximations to A.

Note that $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sigma_1^2 + \cdots \sigma_{\min(m,n)}^2}$.

Matrix Norm Comparison

Let's consider again the matrix

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 1 & 3 & 5 \\ -4 & 6 & 6 \end{bmatrix}.$$

We can compute (we did $p = 1$ and $p = \infty$ above):

$$\begin{aligned} \|A\|_1 &= 13, \\ \|A\|_2 &\approx 10.6252, \\ \|A\|_\infty &= 16, \\ \|A\|_F &\approx 11.7047. \end{aligned}$$

Matrix Norm Comparison

Just like for vectors, there are inequalities relating different matrix norms: if $A \in \mathbb{R}^{m \times n}$,

$$\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty,$$
$$\frac{1}{\sqrt{m}}\|A\|_1 \le \|A\|_2 \le \sqrt{n}\|A\|_1,$$
$$\frac{1}{\sqrt{n}}\|A\|_F \le \|A\|_2 \le \|A\|_F.$$

Definition (Norm equivalence)

If V is a vector space with two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$, we say that the norms are equivalent if $c_1\|\boldsymbol{v}\|_\alpha \le \|\boldsymbol{v}\|_\beta \le c_2\|\boldsymbol{v}\|_\alpha$ for all $\boldsymbol{v} \in V$ and some $c_1, c_2 > 0$ (independent of $\boldsymbol{v}$).