STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2023
LECTURE 1

1. NORMS

- a *norm* is a real-valued function on a vector space (over $\mathbb{R}$ or $\mathbb{C}$), denoted $\|\cdot\| : V \to \mathbb{R}$ satisfying
  (i) $\|v\| \geq 0$ for all $v \in V$
  (ii) $\|v\| = 0$ if and only if $v = 0_V$
  (iii) $\|\alpha v\| = |\alpha|\|v\|$ for all $\alpha \in \mathbb{C}$ and $v \in V$
  (iv) $\|v + w\| \leq \|v\| + \|w\|$ for any $v, w \in V$
- the triangle inequality generalizes directly to sums of more than two vectors:

$$\|u + v + w\| \leq \|u + v\| + \|w\| \leq \|u\| + \|v\| + \|w\|$$

- more generally,

$$\left\|\sum_{i=1}^{m} v_i\right\| \leq \sum_{i=1}^{m} \|v_i\|$$

- we will be interested in two specific choices of V
  - $V = \mathbb{R}^n$ or $\mathbb{C}^n$
  - $V = \mathbb{R}^{m\times n}$ or $\mathbb{C}^{m\times n}$
- our convention: $\mathbb{R}^n \equiv \mathbb{R}^{n\times 1}$ and $\mathbb{C}^n \equiv \mathbb{C}^{n\times 1}$
- in other words, vectors $\mathbf{x} \in \mathbb{R}^n$ or $\mathbb{C}^n$ will *always* be a column vector
- if we ever need to denote a row vector, we will write $\mathbf{x}^\mathsf{T} \in \mathbb{R}^{1\times n}$ or $\mathbb{C}^{1\times n}$

2. VECTOR NORMS

- if $V = \mathbb{C}^n$ or $V = \mathbb{R}^n$, we call a norm on V a *vector norm*
- example: consider $\|\cdot\|_1 : \mathbb{C}^n \to \mathbb{R}$ defined by

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

  for $\mathbf{x} = [x_1, \dots, x_n]^\mathsf{T} \in \mathbb{C}^n$ and where $|x|$ denotes the modulus/absolute value of $x \in \mathbb{C}$
  - check that this is a norm:
    (1) clearly $\|\mathbf{x}\|_1 \geq 0$
    (2) the only way a sum nonnegative entries $\|\mathbf{x}\|_1 = 0$ is if all entries $|x_i| = 0$ and so $\mathbf{x} = [0, \dots, 0]^\mathsf{T} = \mathbf{0}$
    (3) we have

$$\|\alpha\mathbf{x}\|_1 = \sum_{i=1}^{n} |\alpha x_i| = |\alpha| \sum_{i=1}^{n} |x_i| = |\alpha|\|\mathbf{x}\|_1$$

    since complex modulus satisfies $|\alpha x| = |\alpha||x|$

Date: October 2, 2023, version 1.0. Comments, bug reports: `lekheng@uchicago.edu`.

(4) using the triangle inequality for complex numbers, we obtain

$$\|\mathbf{x}+\mathbf{y}\|_1 = \sum_{i=1}^{n}|x_i+y_i| \le \sum_{i=1}^{n}|x_i|+|y_i| \le \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1$$

– therefore the function defines a norm, called the 1-*norm* or *Manhattan norm*

- example: more generally, for $p \ge 1$ (can be any real number, not necessarily an integer), we define the p-*norm* $\|\mathbf{x}\|_p$ by

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

  – most commonly used p-norms is the 2-*norm* or *Euclidean norm*:

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2}$$

  – easy to see that for any p, we have

$$\left(\max_{i=1,\dots,n}|x_i|^p\right)^{1/p} \le \|\mathbf{x}\|_p = \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p} \le \left(n\max_{i=1,\dots,n}|x_i|^p\right)^{1/p}$$

  – from which it follows that

$$\max_{i=1,\dots,n}|x_i| \le \|\mathbf{x}\|_p \le n^{1/p}\max_{i=1,\dots,n}|x_i|$$

  – as $p \to \infty$, we obtain the *infinity norm*

$$\|\mathbf{x}\|_\infty = \lim_{p\to\infty}\|\mathbf{x}\|_p = \max_{i=1,\dots,n}|x_i|$$

  which is also known as the *Chebyshev norm*

  – easy to verify that p-norms for any $p \in [1,\infty]$ are indeed norms

- generalization of the p-norm is the *weighted* p-*norm*, defined by

$$\|\mathbf{x}\|_{p,\mathbf{w}} = \left(\sum_{i=1}^{n} w_i|x_i|^p\right)^{1/p}$$

  – again it can be shown that this is a norm as long as the *weights* w_i, $i = 1,\dots,n$, are strictly positive real numbers

- example: a vast generalization of all of the above is the A-*norm* or *Mahalanobis norm*, defined in terms of a matrix A by

$$\|\mathbf{x}\|_A = (\mathbf{x}^* A\mathbf{x})^{1/2} = \left(\sum_{i,j=1}^{n} a_{ij}\overline{x}_i x_j\right)^{1/2}$$

  – this defines a norm provided that the matrix A is positive definite

  – note that if $W = \operatorname{diag}(\mathbf{w})$, then

$$\|\mathbf{x}\|_W = \|\mathbf{x}\|_{2,\mathbf{w}}$$

3. Matrix norms

- note that the space of complex $m\times n$ matrices $\mathbb{C}^{m\times n}$ is a vector space over $\mathbb{C}$ (ditto for real matrices over $\mathbb{R}$) of dimension mn
- we write O for the $m\times n$ zero matrix, i.e., all entries are 0
- a norm on either $\mathbb{C}^{m\times n}$ or $\mathbb{R}^{m\times n}$ is called a *matrix norm*
- recall that these means $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ satisfies

  (1) $\|A\| \ge 0$ for all $A \in \mathbb{C}^{m\times n}$

(2) $\|A\| = 0$ if and only if $A = O$
(3) $\|\alpha A\| = |\alpha|\|A\|$
(4) $\|A + B\| \le \|A\| + \|B\|$

- often we add a fifth condition that $\|\cdot\|$ satisfies the *submultiplicative property*

$$\|AB\| \le \|A\|\|B\|$$

4. Hölder norms

- example: *Frobenius norm*

$$\|A\|_{\mathsf{F}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}$$

which is submultiplicative since

$$\|AB\|_{\mathsf{F}}^2 = \sum_{i=1}^{m}\sum_{k=1}^{p}\left|\sum_{j=1}^{n} a_{ij}b_{jk}\right|^2 \le \sum_{i=1}^{m}\sum_{k=1}^{p}\left[\left(\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{j=1}^{n}|b_{jk}|^2\right)\right]$$

by the Cauchy–Schwarz inequality and the last expression is equal to

$$\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{k=1}^{p}\sum_{j=1}^{n}|b_{jk}|^2\right) = \|A\|_{\mathsf{F}}^2\|B\|_{\mathsf{F}}^2$$

- example: more generally we have Hölder p-norm for any $p \in [1, \infty]$,

$$\|A\|_{\mathsf{H},p} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p\right)^{1/p}$$

and

$$\|A\|_{\mathsf{H},\infty} = \max_{i,j}|a_{ij}|$$

  - Hölder norms are obtained by viewing an $m \times n$ matrix $A = [a_{ij}]_{i,j=1}^{m,n} \in \mathbb{C}^{m\times n}$ as a vector $\boldsymbol{\alpha} = [a_{11}, a_{12}, \dots, a_{mn}]^{\mathsf{T}} \in \mathbb{C}^{mn}$ with mn entries, this is often written as

$$\boldsymbol{\alpha} = \operatorname{vec}(A)$$

  - we have $\|A\|_{\mathsf{H},p} = \|\operatorname{vec}(A)\|_p$
  - clearly $\|A\|_{\mathsf{H},2} = \|A\|_{\mathsf{F}} = \|\operatorname{vec}(A)\|_2$
- in general Hölder p-norms are not submultiplicative for $p \ne 2$
  - example: take

$$A = \begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}, \qquad B = \begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}, \qquad AB = \begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix}$$

    but

$$\|AB\|_{\mathsf{H},\infty} = 2 > 1 = \|A\|_{\mathsf{H},\infty}\|B\|_{\mathsf{H},\infty}$$

5. OPERATOR NORMS

- a very important class of matrix norms are the so called *operator* or *induced* or *natural norms* defined as
$$\|A\|_{a,b} := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_b}{\|\mathbf{x}\|_a} \tag{1}$$
for any $A \in \mathbb{C}^{m\times n}$ and any vector norms $\|\cdot\|_a : \mathbb{C}^n \to \mathbb{R}$ and $\|\cdot\|_b : \mathbb{C}^m \to \mathbb{R}$ defined on the domain and codomain of A respectively
- note that a and b here are not numbers, just used to distinguish the two norms
- the operator norm may also be written as
$$\|A\|_{a,b} = \max\{\|A\mathbf{x}\|_b : \|\mathbf{x}\|_a \leq 1\} \tag{2}$$
or as
$$\|A\|_{a,b} = \max\{\|A\mathbf{x}\|_b : \|\mathbf{x}\|_a = 1\} \tag{3}$$
- in other words, the operator norm measures how far the operator A sends points in the unit disc (or the unit circle)
- proof is simple, for example, here's how you would prove (3):
$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_b}{\|\mathbf{x}\|_a} = \max_{\mathbf{x}\neq\mathbf{0}} \left\| \frac{1}{\|\mathbf{x}\|_a} A\mathbf{x} \right\|_b = \max_{\mathbf{x}\neq\mathbf{0}} \left\| A\left(\frac{\mathbf{x}}{\|\mathbf{x}\|_a}\right) \right\|_b = \max_{\|\mathbf{v}\|_a=1} \|A\mathbf{v}\|_b\,,$$
the first equality uses the property that $|\alpha|\|\mathbf{v}\|_b = \|\alpha\mathbf{v}\|_b$, the second equality uses $\alpha A\mathbf{x} = A(\alpha\mathbf{x})$, and the last equality uses the observation that $\mathbf{v} = \mathbf{x}/\|\mathbf{x}\|_a$ always has unit a-norm
- exercise: prove (3) and (2) are equal
- another exercise: prove that
$$\|A\mathbf{x}\|_b \leq \|A\|_{a,b}\|\mathbf{x}\|_a \tag{4}$$
for any $\mathbf{x} \in \mathbb{C}^n$; this more restrictive form of submultiplicativity is called *consistency*
- a note on the use of *supremum* and *maximum*: for $S \subseteq \mathbb{C}^n$ and a real-valued function f whose domain includes S,
  - we write $\sup_{\mathbf{x}\in S} f(\mathbf{x})$ for the smallest $\mu \in \mathbb{R}$ such that $f(\mathbf{x}) \leq \mu$ for every $\mathbf{x} \in S$ (and we set $\mu = +\infty$ if f is unbounded on S)
  - we write $\max_{\mathbf{x}\in S} f(\mathbf{x})$ if the supremum is attained by some element in S, i.e., there is an $\mathbf{x}_{\max} \in S$ such that $f(\mathbf{x}_{\max}) = \sup_{\mathbf{x}\in S} f(\mathbf{x})$
  - $\mathbf{x}_{\max}$ is called a *maximizer* of f on S
  - likewise for infimum and minimum (and minimizer)
  - by the extreme value theorem, if f is continuous and S is compact, then supremum and infimum are always attained
- in the above $S = \{\mathbf{x} \in \mathbb{C} : \|\mathbf{x}\|_a \leq 1\}$ and $S = \{\mathbf{x} \in \mathbb{C} : \|\mathbf{x}\|_a = 1\}$ are compact and the function $f = \|\cdot\|_b : \mathbb{C}^m \to \mathbb{R}$ is continuous
- in other words, we can always find an $\mathbf{x}_{\max}$ with $\|\mathbf{x}_{\max}\|_a = 1$ such that
$$\|A\mathbf{x}_{\max}\|_b = \|A\|_{a,b}$$
- that's why we may always write max in (3) and (2), and therefore in (1); although strictly speaking we should have written (1)
$$\|A\|_{a,b} := \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_b}{\|\mathbf{x}\|_a}$$
- the operator norm is *not* submultiplicative in general: take
$$A = \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

since every $\mathbf{x} \in \mathbb{R}^2$ with $\|\mathbf{x}\| = 1$ has the form $\mathbf{x} = (\cos\theta, \sin\theta)^\mathsf{T}$, we see that

$$\begin{aligned}
\|A\|_{2,\infty} &= \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_\infty = \max_\theta |2\cos\theta + 2\sin\theta| = 2\sqrt{2} \\
\|B\|_{2,\infty} &= \max_{\|\mathbf{x}\|_2=1} \|B\mathbf{x}\|_\infty = \max_\theta |\cos\theta| = 1 \\
\|AB\|_{2,\infty} &= \max_{\|\mathbf{x}\|_2=1} \|AB\mathbf{x}\|_\infty = \max_\theta |4\cos\theta| = 4
\end{aligned}$$

but

$$\|AB\|_{2,\infty} = 4 > 2\sqrt{2} = \|A\|_{2,\infty}\|B\|_{2,\infty}$$

(thanks to Lijun Ding for this example)

- however given $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times p}$ it is always true that

$$\|AB\|_{a,c} \le \|A\|_{b,c}\|B\|_{a,b}$$

for any norms $\|\cdot\|_c$ on $\mathbb{C}^p$, $\|\cdot\|_b$ on $\mathbb{C}^m$, $\|\cdot\|_a$ on $\mathbb{C}^n$
- the most interesting operator norms are the ones obtained when $\|\cdot\|_a$ and $\|\cdot\|_b$ are vector ℓ^p-norms, we write

$$\|A\|_{p,q} := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_q}{\|\mathbf{x}\|_p} \quad \text{and} \quad \|A\|_p := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

for any $A \in \mathbb{C}^{m\times n}$ and $p, q \in [1, \infty]$
- we call $\|\cdot\|_{p,q}$ the matrix (p,q)-norm and $\|\cdot\|_p$ the matrix p-norm
- the matrix 2-norm

$$\|A\|_2 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$

is very widely used and has its own special name, *spectral norm*, because of its relation to the spectrum of a matrix (i.e., the eigenvalues); we will discuss it in the next two lectures
- the matrix 1-norm and ∞-norm are also very widely used, largely because, they can be easily computed
- let $A = [a_{ij}]_{i,j=1}^{m,n} \in \mathbb{C}^{m\times n}$, then

$$\|A\|_1 = \max_{j=1,\dots,n} \left[\sum_{i=1}^m |a_{ij}|\right] \tag{5}$$

and

$$\|A\|_\infty = \max_{i=1,\dots,m} \left[\sum_{j=1}^n |a_{ij}|\right] \tag{6}$$

- an easy way to remember these is that $\|A\|_1$ is the maximum column sum and $\|A\|_\infty$ is the maximum row sum of A
- let us prove (6) and leave (5) as an exercise:

- we use (3), so

$$
\begin{aligned}
\|A\|_\infty &= \max\{\|A\mathbf{x}\|_\infty : \|\mathbf{x}\|_\infty = 1\} \\
&= \max_{\|\mathbf{x}\|_\infty=1} \left\{ \max_{i=1,\dots,m} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \right\} \\
&\le \max_{\|\mathbf{x}\|_\infty=1} \left\{ \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}||x_j| \right] \right\} \\
&\le \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right] \qquad (7)
\end{aligned}
$$

  where the last inequality follows because $\|\mathbf{x}\|_\infty = 1$ and so we must have $|x_j| \le 1$
- to show equality, we just need to exhibit one single $\mathbf{x}^*$ with $\|\mathbf{x}^*\|_\infty = 1$ so that

$$\|A\mathbf{x}^*\|_\infty \ge \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right]$$

- we know that the maximum in (7) is attained by some row $i = k \in \{1, \dots, m\}$, so

$$\max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right] = \sum_{j=1}^{n} |a_{kj}|$$

- now we define $\mathbf{x}^* = [x_1^*, \dots, x_n^*] \in \mathbb{C}^n$ as the vector whose coordinates are given by

$$x_j^* = \begin{cases} |a_{kj}|/a_{kj} & \text{if } a_{kj} \ne 0, \\ 0 & \text{if } a_{kj} = 0, \end{cases}$$

  for $j = 1, \dots, n$
- observe that $\mathbf{x}^*$ has $\|\mathbf{x}^*\|_\infty = 1$ as well as the effect of attaining the requisite bound

$$\|A\mathbf{x}^*\|_\infty = \max_{i=1,\dots,m} \left| \sum_{j=1}^{n} a_{ij} x_j^* \right| \ge \sum_{j=1}^{n} |a_{kj}| = \max_{i=1,\dots,m} \left[\sum_{j=1}^{n} |a_{ij}| \right]$$