STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2019
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
  (1) $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in V$
  (2) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
  (3) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all $\alpha \in \mathbb{C}$ and $\mathbf{x} \in V$
  (4) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for any $\mathbf{x}, \mathbf{y} \in V$
- the triangle inequality generalizes directly to sums of more than two vectors:
$$\|\mathbf{x}+\mathbf{y}+\mathbf{z}\| \leq \|\mathbf{x}+\mathbf{y}\| + \|\mathbf{z}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| + \|\mathbf{z}\|$$
- more generally,
$$\left\|\sum_{i=1}^{m} \mathbf{x}_i\right\| \leq \sum_{i=1}^{m} \|\mathbf{x}_i\|$$
- 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 $\mathbb{C}^{m\times n}$

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 $|\alpha x| = |\alpha||x|$
    (4) using the triangle inequality for complex numbers, we obtain
$$\|\mathbf{x}+\mathbf{y}\|_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} |x_i| + |y_i| \leq \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1$$
  - therefore the function defines a norm, called the *1-norm* or *Manhattan norm*

Date: October 2, 2019, version 1.0. Comments, bug reports: `lekheng@galton.uchicago.edu`.

- example: more generally, for $p \geq 1$ (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

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

  - from which it follows that

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

  - 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. Continuity of norms

- all norms are continuous functions — an simple but important observation
- what can we say about the norm of the difference of two vectors? we know that $\|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ but we can obtain a more useful relationship as follows:

$$\|\mathbf{x}\| = \|(\mathbf{x} - \mathbf{y}) + \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\|$$

  we obtain

$$\|\mathbf{x} - \mathbf{y}\| \geq \|\mathbf{x}\| - \|\mathbf{y}\|$$

- thirdly, from
$$\|\mathbf{y}\| = \|\mathbf{y} - \mathbf{x} + \mathbf{x}\| \le \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{x}\|$$
it follows that
$$\|\mathbf{x} - \mathbf{y}\| \ge \|\mathbf{y}\| - \|\mathbf{x}\|$$
and therefore
$$\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \le \|\mathbf{x} - \mathbf{y}\| \tag{3.1}$$
- the inequality (3.1) yields a very important property of norms, namely, they are all (uniformly) continuous functions of the entries of their arguments — in fact, they are *Lipschitz functions* if you know what those are

4. Equivalence of norms

- there are also interesting relationships for two different norms
- first and foremost, on finite dimensional spaces (which include $\mathbb{C}^n$ and $\mathbb{C}^{m\times n}$) all norms are *equivalent*
  - that is, given two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$, there exist constants c_1 and c_2 with $0 < c_1 < c_2$ such that
$$c_1\|\mathbf{x}\|_\alpha \le \|\mathbf{x}\|_\beta \le c_2\|\mathbf{x}\|_\alpha \tag{4.1}$$
for all $\mathbf{x} \in V$
  - example: from the definition of the ∞-norm, we have
$$\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2 \le \sqrt{n}\|\mathbf{x}\|_\infty$$
  - example: also not hard to show that
$$\frac{1}{n}\|\mathbf{x}\|_1 \le \|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_1$$
  - in fact, no matter what crazy choices of norms that we make, say
$$\|x\|_\alpha = \left(\sum_{i=1}^{n} i|x_i|^n\right)^{1/n}, \qquad \|x\|_\beta = \mathbf{x}^{\mathsf{T}} \begin{bmatrix} 3 & -1 & & \\ -1 & 3 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 3 \end{bmatrix} \mathbf{x},$$
we know that there are c_1 and c_2 so that (4.1) holds
- by definition, a sequence of vectors $\mathbf{x}_0, \mathbf{x}_1, \dots$ converges to a vector $\mathbf{x}$ if and only if
$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}\| = 0$$
for any norm (you may also write down a formal version in terms of ε and N)
- the equivalence of norms on finite dimensional vector spaces tells us that
$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}\|_\alpha = 0 \quad \text{if and only if} \quad \lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{x}\|_\beta = 0$$
for any choice of norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ (why?)
- if we can establish convergence of an algorithm in a specific norm convergence in every other norm follows automatically
- for this reason, norms are very useful to measure the error in an approximation
- secondly we have a relationship that applies to products of norms, the *Hölder inequality*
$$|\mathbf{x}^*\mathbf{y}| \le \|\mathbf{x}\|_p\|\mathbf{y}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$
  - a well-known corollary arises when $p = q = 2$, the *Cauchy–Schwarz inequality*
$$|\mathbf{x}^*\mathbf{y}| \le \|\mathbf{x}\|_2\|\mathbf{y}\|_2$$

- you will see a generalization of Cauchy–Schwarz inequality called the *Bessel inequality* in Homework 0
- by setting $\mathbf{x} = [1, 1, \dots, 1]^\mathsf{T}$, the Hölder inequality yields the relationships

$$\left|\sum_{i=1}^n y_i\right| \le \sum_{i=1}^n |y_i|$$

and

$$\left|\sum_{i=1}^n y_i\right| \le n \max_{i=1,\dots,n} |y_i|$$

and

$$\left|\sum_{i=1}^n y_i\right| \le \sqrt{n} \left(\sum_{i=1}^n |y_i|^2\right)^{1/2}$$

5. 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*

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

6. Hölder norms

- example: *Frobenius norm*

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

which is submultiplicative since

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

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

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

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

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

and

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

– 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 $\boldsymbol{\alpha} = [a_{11}, a_{12}, \ldots, a_{mn}]^{\mathsf{T}} \in \mathbb{C}^{mn}$ with mn entries, this is often written as

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

– we have $\|A\|_{H,p} = \|\operatorname{vec}(A)\|_p$
– clearly $\|A\|_{H,2} = \|A\|_F = \|\operatorname{vec}(A)\|_2$

- in general Hölder p-norms are not submultiplicative for $p \neq 2$
  – 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

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