STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2023
LECTURE 2

1. CONTINUITY OF NORMS

- all norms are continuous functions — a simple but important observation
- we know that $\|v-w\| \leq \|v\|+\|w\|$ but we can obtain another useful relationship as follows:
$$\|v\| = \|(v-w)+w\| \leq \|v-w\| + \|w\|$$
we obtain
$$\|v-w\| \geq \|v\| - \|w\|$$
- thirdly, from
$$\|w\| = \|w-v+v\| \leq \|v-w\| + \|v\|$$
it follows that
$$\|v-w\| \geq \|w\| - \|v\|$$
and therefore
$$\big|\|v\| - \|w\|\big| \leq \|v-w\| \tag{1.1}$$
- the inequality (1.1) 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

2. 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
$$c_1\|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq c_2\|\mathbf{x}\|_\alpha \tag{2.1}$$
  for all $\mathbf{x} \in V$
  - example: from the definition of the ∞-norm, we have
$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$$
  - example: also not hard to show that
$$\frac{1}{n}\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1$$
  - in fact, no matter what crazy choices of norms that we make, say
$$\|\mathbf{x}\|_\alpha = \left(\sum_{i=1}^n i|x_i|^n\right)^{1/n}, \qquad \|\mathbf{x}\|_\beta = \mathbf{x}^{\mathsf{T}} \begin{bmatrix} 3 & -1 & & \\ -1 & 3 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 3 \end{bmatrix} \mathbf{x},$$
  we know that there are c_1 and c_2 so that (2.1) holds

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

- it is the same for $\mathbb{C}^{m\times n}$; for example we always have

$$\|A\|_{\mathsf{H},\infty} \le \|A\|_{\mathsf{F}} \le \sqrt{mn}\|A\|_{\mathsf{H},\infty}$$
$$\frac{1}{\sqrt{m}}\|A\|_2 \le \|A\|_\infty \le \sqrt{n}\|A\|_2$$
$$\|A\|_2 \le \|A\|_{\mathsf{F}} \le \sqrt{n}\|A\|_2$$

  for all $A \in \mathbb{C}^{m\times n}$

- 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

3. INNER PRODUCTS

- an *inner product* is a complex-valued function on a product of a vector space (over $\mathbb{C}$) with itself, denoted $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$, satisfying
  (1) $\langle v, v\rangle \ge 0$ for all $v \in V$
  (2) $\langle v, v\rangle = 0$ if and only if $v = 0_V$
  (3) $\langle v, \alpha_1 w_1 + \alpha_2 w_2\rangle = \alpha_1\langle v, w_1\rangle + \alpha_2\langle v, w_2\rangle$ for all $\alpha_1, \alpha_2 \in \mathbb{C}$ and $v, w_1, w_2 \in V$
  (4) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for any $v, w \in V$
- by virtue of the last two conditions $\langle \alpha_1 v_1 + \alpha_2 v_2, w\rangle = \overline{\alpha}_1\langle v_1, w\rangle + \overline{\alpha}_2\langle v_2, w\rangle$ for all $\alpha_1, \alpha_2 \in \mathbb{C}$ and $v_1, v_2, w \in V$
- for real vector spaces, an inner product is a real-valued function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ and as a result, the last two conditions become:
  (3) $\langle \alpha_1 v_1 + \alpha_2 v_2, w\rangle = \alpha_1\langle v_1, w\rangle + \alpha_2\langle v_2, w\rangle$ for all $\alpha_1, \alpha_2 \in \mathbb{R}$ and $v_1, v_2, w \in V$
  (4) $\langle v, w\rangle = \langle w, v\rangle$ for any $v, w \in V$
- just as norms are an abstraction of length, inner products are an abstraction of angles (or rather, inverse cosines of angles)
- the defining properties of an inner product tell us that

$$\|v\| \coloneqq \sqrt{\langle v, v\rangle}$$

  defines a norm called the *norm induced by the inner product*

- Cauchy–Schwartz inequality in fact holds for any inner product and the norm induced by that inner product

$$|\langle v, w\rangle| \le \|v\|\|w\|$$

- given a norm $\|\cdot\|$, how can we tell if it is a norm induced by some inner product?
- the answer is: if and only if the norm satisfies the *parallelogram law*

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2$$

- the only inner products we care about in this course are the *Hermitian inner product* or l^2-inner product for vectors

$$\langle \mathbf{x}, \mathbf{y} \rangle \coloneqq \mathbf{x}^* \mathbf{y} = \sum_{i=1}^{n} \overline{x}_i y_i, \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n$$

  and *trace inner product* for matrices

$$\langle X, Y \rangle \coloneqq \operatorname{tr}(X^* Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} \overline{x}_{ij} y_{ij}, \qquad \text{for all } X, Y \in \mathbb{C}^{m \times n}$$

- over reals, we have

$$\langle \mathbf{x}, \mathbf{y} \rangle \coloneqq \mathbf{x}^{\mathsf{T}} \mathbf{y} = \sum_{i=1}^{n} x_i y_i, \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

  called the *Euclidean* (instead of Hermitian) inner product, and

$$\langle X, Y \rangle \coloneqq \operatorname{tr}(X^{\mathsf{T}} Y) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} y_{ij}, \qquad \text{for all } X, Y \in \mathbb{R}^{m \times n}$$

- the norms induced by these inner products are precisely the Euclidean norm and Frobenius norm respectively since

$$(\mathbf{x}^* \mathbf{x})^2 = \|\mathbf{x}\|_2^2 \quad \text{and} \quad \operatorname{tr}(X^* X) = \|X\|_{\mathsf{F}}^2$$

  for any $\mathbf{x} \in \mathbb{C}^n$ and $X \in \mathbb{C}^{m \times n}$
- Cauchy–Schwartz inequality yields

$$|\mathbf{x}^* \mathbf{y}| \le \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \quad \text{and} \quad |\operatorname{tr}(X^* Y)| \le \|X\|_{\mathsf{F}} \|Y\|_{\mathsf{F}}$$

- using the parallelogram law, we can show that no other vector p-norms or matrix Hölder p-norm are induced by inner products when $p \neq 2$
- the parallelogram law also tells us that matrix (p, q)-norm are not induced by inner products, whatever the value of p and q (including $p = q = 2$, so the spectral norm is not induced by an inner product either)
- there are two well-known generalizations of the Cauchy–Schwarz inequality:
  (i) the *Hölder inequality*

$$|\mathbf{x}^* \mathbf{y}| \le \|\mathbf{x}\|_p \|\mathbf{y}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

  when $p = q = 2$, we get back the Cauchy–Schwarz inequality
  (ii) the *Bessel inequality* in Homework 0: for any $\mathbf{x}_1, \dots, \mathbf{x}_r$ pairwise orthogonal unit vectors, i.e., $\|\mathbf{x}_i\|_2 = 1$ for all $i = 1, \dots, r$, and $\mathbf{x}_i^{\mathsf{T}} \mathbf{x}_j = 0$ for all $i \neq j$,

$$\sum_{i=1}^{r} (\mathbf{x}^{\mathsf{T}} \mathbf{x}_i)^2 \le \|\mathbf{x}\|_2^2$$

  when $r = 1$, we get back the Cauchy–Schwarz inequality

4. Outer product

- for $\mathbf{x} = [x_1, \dots, x_m]^{\mathsf{T}} \in \mathbb{C}^m$ and $\mathbf{y} = [y_1, \dots, y_n]^{\mathsf{T}} \in \mathbb{C}^n$, the product

$$\mathbf{x}\mathbf{y}^{\mathsf{T}} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{bmatrix} \in \mathbb{C}^{m \times n}$$

or

$$\mathbf{x}\mathbf{y}^* = \begin{bmatrix} x_1\bar{y}_1 & x_1\bar{y}_2 & \cdots & x_1\bar{y}_n \\ x_2\bar{y}_1 & x_2\bar{y}_2 & \cdots & x_2\bar{y}_n \\ \vdots & \vdots & & \vdots \\ x_m\bar{y}_1 & x_m\bar{y}_2 & \cdots & x_m\bar{y}_n \end{bmatrix} \in \mathbb{C}^{m\times n}$$

is often called the *outer product* of $\mathbf{x}$ and $\mathbf{y}$

- if neither $\mathbf{x}$ nor $\mathbf{y}$ is the zero vector, then

$$\operatorname{rank}(\mathbf{x}\mathbf{y}^{\mathsf{T}}) = \operatorname{rank}(\mathbf{x}\mathbf{y}^*) = 1$$

- furthermore if $\operatorname{rank}(A) = 1$, then there exists $\mathbf{x} \in \mathbb{C}^m$ and $\mathbf{y} \in \mathbb{C}^n$ so that $A = \mathbf{x}\mathbf{y}^*$
- as such a matrix of this form is often called a *rank*-1 *matrix*

5. Matrix product

- the following are some useful observations regarding matrix-matrix and matrix-vector products, assuming over $\mathbb{R}$ for simplicity
- by definition, multiplying two matrices

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^{\mathsf{T}} \\ \vdots \\ \boldsymbol{\alpha}_m^{\mathsf{T}} \end{bmatrix} \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B = [\mathbf{b}_1, \dots, \mathbf{b}_p] \in \mathbb{R}^{n\times p}$$

is the same as forming the matrix of inner products of the row vectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m \in \mathbb{R}^n$ and the column vectors $\mathbf{b}_1, \dots, \mathbf{b}_p \in \mathbb{R}^n$,

$$AB = \begin{bmatrix} \boldsymbol{\alpha}_1^{\mathsf{T}}\mathbf{b}_1 & \boldsymbol{\alpha}_1^{\mathsf{T}}\mathbf{b}_2 & \cdots & \boldsymbol{\alpha}_1^{\mathsf{T}}\mathbf{b}_p \\ \boldsymbol{\alpha}_2^{\mathsf{T}}\mathbf{b}_1 & \boldsymbol{\alpha}_2^{\mathsf{T}}\mathbf{b}_2 & \cdots & \boldsymbol{\alpha}_2^{\mathsf{T}}\mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_m^{\mathsf{T}}\mathbf{b}_1 & \boldsymbol{\alpha}_m^{\mathsf{T}}\mathbf{b}_2 & \cdots & \boldsymbol{\alpha}_m^{\mathsf{T}}\mathbf{b}_p \end{bmatrix} \in \mathbb{R}^{m\times p} \tag{5.1}$$

- multiplying two matrices

$$A = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B = \begin{bmatrix} \boldsymbol{\beta}_1^{\mathsf{T}} \\ \vdots \\ \boldsymbol{\beta}_n^{\mathsf{T}} \end{bmatrix} \in \mathbb{R}^{n\times p}$$

is the same as taking the sum of outer products of the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ and the row vectors $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_n \in \mathbb{R}^n$,

$$AB = \mathbf{a}_1\boldsymbol{\beta}_1^{\mathsf{T}} + \cdots + \mathbf{a}_n\boldsymbol{\beta}_n^{\mathsf{T}} \in \mathbb{R}^{m\times p} \tag{5.2}$$

- multiplying two matrices

$$A \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B = [\mathbf{b}_1, \dots, \mathbf{b}_p] \in \mathbb{R}^{n\times p}$$

is the same as multiplying A to each of the column vectors $\mathbf{b}_1, \dots, \mathbf{b}_p \in \mathbb{R}^n$ of B,

$$AB = [A\mathbf{b}_1, \dots, A\mathbf{b}_p] \in \mathbb{R}^{m\times p}$$

- multiplying two matrices

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^{\mathsf{T}} \\ \vdots \\ \boldsymbol{\alpha}_n^{\mathsf{T}} \end{bmatrix} \in \mathbb{R}^{m\times n} \qquad \text{and} \qquad B \in \mathbb{R}^{n\times p}$$

is the same as multiplying each of the row vectors $\boldsymbol{\alpha}_1^\mathsf{T}, \dots, \boldsymbol{\alpha}_n^\mathsf{T} \in (\mathbb{R}^m)^* = \mathbb{R}^{m\times 1}$ by B on the right:

$$AB = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} B \\ \vdots \\ \boldsymbol{\alpha}_n^\mathsf{T} B \end{bmatrix} \in \mathbb{R}^{m\times p}$$

- the following are special cases when one of the matrices is a vector or is a diagonal matrix
- multiplying $A = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ on the right by $\mathbf{x} = [x_1, \dots, x_n]^\mathsf{T} \in \mathbb{R}^n$ is the same as taking linear combinations of the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$,

$$A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \in \mathbb{R}^m$$

- multiplying

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\alpha}_n^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

  on the left by $\mathbf{y}^\mathsf{T} = [y_1, \dots, y_m] \in \mathbb{R}^m$ is the same as taking linear combinations of the row vectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m \in \mathbb{R}^n$,

$$\mathbf{y}^\mathsf{T} A = y_1\boldsymbol{\alpha}_1^\mathsf{T} + \cdots + y_m\boldsymbol{\alpha}_m^\mathsf{T} \in \mathbb{R}^{n*}$$

  where $\mathbb{R}^{n*} = \mathbb{R}^{1\times n}$ is the dual space of $\mathbb{R}^n = \mathbb{R}^{n\times 1}$
- multiplying $A = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$ on the right by a diagonal matrix

$$D = \operatorname{diag}(d_1, \dots, d_n) = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \in \mathbb{R}^{n\times n}$$

  is the same as scaling the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ by $d_1, \dots, d_n \in \mathbb{R}$,

$$AD = [d_1\mathbf{a}_1, \dots, d_n\mathbf{a}_n] \in \mathbb{R}^{m\times n}$$

- multiplying

$$A = \begin{bmatrix} \boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{\alpha}_m^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

  on the left by a diagonal matrix

$$\Delta = \operatorname{diag}(\delta_1, \dots, \delta_m) = \begin{bmatrix} \delta_1 & & \\ & \ddots & \\ & & \delta_m \end{bmatrix} \in \mathbb{R}^{m\times m}$$

  is the same as scaling the row vectors $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n \in \mathbb{R}^m$, by $\delta_1, \dots, \delta_m \in \mathbb{R}$,

$$\Delta A = \begin{bmatrix} \delta_1\boldsymbol{\alpha}_1^\mathsf{T} \\ \vdots \\ \delta_m\boldsymbol{\alpha}_m^\mathsf{T} \end{bmatrix} \in \mathbb{R}^{m\times n}$$