Intro to Computational Math, Spring 2026
Section 5, Feb 24 (not due or graded).

Section will give you time to work on/discuss questions in small groups and then present your solutions. The information below is more of a summary of useful topics. There are no problems this week, but the following may be useful to garner intuition for matrix norms and condition numbers.

Reviewing Norm Inequalities: Recall that a norm on a vector space (i.e. $\mathbb{R}^n$) is a mapping that takes in a vector and spits out a nonnegative scalar. $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$. Furthermore, it must satisfy three properties

1. $\|x\| = 0 \iff x = 0$
2. $\|ax\| = |a|\|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|$

In a sense, these rules but restrictions on the operations that make a vector space a vector space. The first tells us how to handle the zero vector. The second explains how scalar multiplication works under the norm. The third explains how the norm behaves under addition of vectors. Recall a *vector space is just a set where you can add/subtract, multiply by scalars, and notions of a "zero" exist.* Therefore, the set of matrices also exhibit a vector space structure, and we may put a norm on them. In particular, we will *induce* a norm from the underlying vector spaces that the matrix is sending vectors from and sending vectors to.

Definition 1. *Let A be a $m \times n$ matrix. Given a p-norm $\|\cdot\|_p$ we define the **induced matrix p-norm** as following*

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|=1} \|Ax\|_p$$

This definition can be interpreted as how much does the matrix scale a vector, relative to the vectors size and relative to the geometry induced by the norm. Note that different norms induce different notions of distance, which induce different geometries.

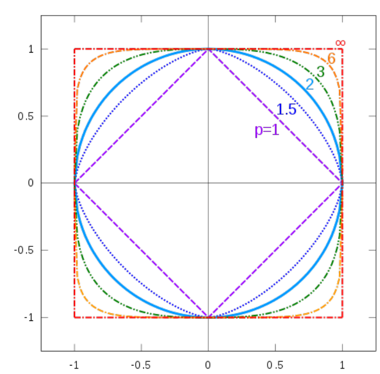

Figure 1: p-norm "circles"

Remark 2. *For any norm $\|\cdot\|$, the identity matrix has norm* 1*, as $Ix = x$ for all x, so the ratio of the norms is always* 1*. That is, there is no scaling/stretching.*

Remark 3. *Note, we have to be really careful of our intuition when it comes to "stretching" as the underlying geometry induced by the norm is important here. A matrix that rotated a vector in $\mathbb{R}^2$ 45 degrees clockwise may not seem to stretch the vector at all, so we would image its induced matrix norm is 1. This is the case only when $p = 2$ and circles look like circles. Otherwise, when rotating, the length (with respect to this norm) changes.*

We can now prove a general theorem that bounds various norms. Similar to the definitional bounds we say for a vector space norm, these provide some induced structure on the operations a matrix applies (i.e. matrix-vector multiplication and matrix-matrix multiplication).

Theorem 3.1. *Let $\|\cdot\|$ designate a matrix norm and the vector norm that induced it. Then for all matrices and vectors of compatible sizes,*

$$\|Ax\| \leq \|A\| \cdot \|x\|.$$

For all matrices of compatible sizes

$$\|AB\| \leq \|A\| \cdot \|B\|$$

Both of these are proven simply by considering the definition; however, both also have geometric intuition that will be discussed.

Proof. If $x = 0$ the result is trivial. Otherwise for any given x

$$\frac{\|Ax\|}{\|x\|} \leq \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \|A\|$$

where the inequality holds because we are considering the largest value of that ratio, and the equality holds by definition. Then for compatible matrices A and B and considering any nonzero x, we can utilize the previous result twice to say

$$\|ABx\| \leq \|A\| \cdot \|Bx\| \leq \|A\| \cdot \|B\| \cdot \|x\|.$$

Dividing both sides by $\|x\|$ and taking a max yields

$$\|AB\| := \max_{x \neq 0} \frac{\|ABx\|}{\|x\|} \leq \|A\| \cdot \|B\|$$

□

Intuition of the above result Both results just follow from the definition, but having a geometric understanding is always nice. The first result

$$\|Ax\| \leq \|A\| \cdot \|x\|$$

says that the size of the output vector is bounded. This really highlights how the matrix norm is some notion of "stretchy-ness" of the matrix. If the vector x is mapped to y, then the norm of is scaled by no more than the matrix norm.

The second result is a little less simple, but it is not terribly complicated. If a matrix norm is a quantity on how much a matrix stretches vectors, then $\|AB\|$ is looking at how much the matrix AB stretches a single vector. However, $\|A\|\|B\|$ allows for A and B is acts on separate vectors, taking their respective maximum stretching. The left hand side has more restriction, so it must be smaller.

Another look at conditioning of linear systems Given data of a matrix and an output vector A and b, we may ask *is there a solution?* Does there exist a vector x such that $Ax = b$?. Later on we may learn there is no solution, so we aim to find the best approximation, finding an x such that Ax is "as close to" b as possible. For now, we assume a solution exists and pay attention to the computations that may need to take place to solve this system.

When considering condition numbers before we looked at the relative change to the output given a change to the input:

$$\frac{\frac{|f(x)-f(\tilde{x})|}{|f(x)|}}{\frac{|x-\tilde{x}|}{|x|}}$$

where $\tilde{x}$ is some perturbation of x. If we write $\tilde{x} = x(1+\epsilon)$ for some small $\epsilon > 0$ then this implies to

$$\frac{|f(x) - f(x+\epsilon x)|}{|\epsilon f(x)|}.$$

Taking a limit as $\epsilon \to 0$ we have seen before that we can define

$$\kappa_f(x) = \lim_{\epsilon \to 0} \frac{|f(x) - f(x+\epsilon x)|}{|\epsilon f(x)|} = \left| \frac{x f'(x)}{f(x)} \right|$$

if $f : \mathbb{R} \to \mathbb{R}$ is differentiable. We can extend this to the vector mappings that a matrix does by considering norms. If $f : \mathbb{R}^n \to \mathbb{R}^m$ we can imaging a generalizing with the Jacobian (the matrix of all the possible partials)

$$\kappa_f(x) = \lim_{\epsilon \to 0} \frac{\|f(x) - f(x+\epsilon x)\|}{\|\epsilon f(x)\|} = \frac{\|x^T J f(x)|}{\|f(x)\|} \leq \frac{\|x\| \|J f(x)\|}{\|f(x)\|}$$

where the last inequality just applies our bound from Theorem 3.1.

In the case where $f(x) = Ax$ for some invertible matrix A then we have $Jf(x) = A$ and so $\|Jf(x)\| = \|A\|$ for all x. As far as conditioning a linear system goes (**with respect to a specific norm**!), we may want to find the worst case over all nonzero x (for division by $f(x)$ to be defined),

$$\kappa_{\|\cdot\|}(A) = \sup_{x \neq 0} \kappa_f(x) \leq \sup_{x \neq 0} \frac{\|x\| \|J f(x)\|}{\|f(x)\|} = \|A\| \sup_{x \neq 0} \frac{\|x\|}{\|Ax\|}$$

Then since A is invertible (i.e. injective and surjective) we may rewrite the last sup with respect to a given $y = Ax$ or equivalently $A^{-1}y = x$. Since A is invertible, this covers all nonzero x and all nonzero y, so this change of variables does not affect the value . That is

$$\kappa_{\|\cdot\|}(A) \leq \|A\| \sup_{x \neq 0} \frac{\|x\|}{\|Ax\|} = \|A\| \sup_{y \neq 0} \frac{\|A^{-1}y\|}{\|y\|} = \|A\| \cdot \|A^{-1}\|.$$

This is exactly the formula seen in class using direct perturbations, and the deriving method is pretty much the same as well. Here we show that the definitions from chapter one still have some use for us.