Norms and Errors
Norms

What’s a norm?

• A generalization of ‘absolute value’ to vectors.
• \( f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}_0^+ \), returns a ‘magnitude’ of the input vector
• In symbols: Often written \( \|\mathbf{x}\| \).

Define norm.

A function \( \|\mathbf{x}\| : \mathbb{R}^n \rightarrow \mathbb{R}_0^+ \) is called a norm if and only if

1. \( \|\mathbf{x}\| > 0 \Leftrightarrow \mathbf{x} \neq \mathbf{0} \).
2. \( \|\gamma \mathbf{x}\| = |\gamma| \|\mathbf{x}\| \) for all scalars \( \gamma \).
3. Obeys triangle inequality \( \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \)
What are some examples of norms?

The so-called $p$-norms:

$$\left\| \begin{pmatrix} x_1 \\ x_n \end{pmatrix} \right\|_p = \sqrt[p]{|x_1|^p + \cdots + |x_n|^p} \quad (p \geq 1)$$

$p = 1, 2, \infty$ particularly important
Demo: Vector Norms (click to visit)
If we’re computing a vector result, the error is a vector. That’s not a very useful answer to ‘how big is the error’. What can we do?

Apply a norm!

How? Attempt 1:

\[
\text{Magnitude of error} \neq \|\text{true value}\| - \|\text{approximate value}\| \quad \text{WRONG!}
\]

Attempt 2:

\[
\text{Magnitude of error} = \|\text{true value} - \text{approximate value}\|\]
What are the absolute and relative errors in approximating the location of Siebel center \((40.114, -88.224)\) as \((40, -88)\) using the 2-norm?

\[
\begin{pmatrix}
40.114 \\
-88.224
\end{pmatrix}
- 
\begin{pmatrix}
40 \\
-88
\end{pmatrix}
= 
\begin{pmatrix}
0.114 \\
-.224
\end{pmatrix}
\]

Absolute magnitude;

\[
\left\| \begin{pmatrix}
40.114 \\
-88.224
\end{pmatrix} \right\|_2 \approx 96.91
\]
Absolute error:

$$\left\| \begin{pmatrix} 0.114 \\ -0.224 \end{pmatrix} \right\|_2 \approx 0.2513$$

Relative error:

$$\frac{0.2513}{96.91} \approx 0.00259.$$ 

**But:** Is the 2-norm really the right norm here?
Demo: Calculate geographic distances using http://tripstance.com

• Siebel Center is at 40.113813,-88.224671. (latitude, longitude)
• Locations in that format are accepted in the location boxes.
• What’s the distance to the nearest integer lat/lon intersection, 40,-88?
• How does distance relate to lat/lon? Only lat? Only lon?
What norms would we apply to matrices?

- Easy answer: ‘Flatten’ matrix as vector, use vector norm. This corresponds to an entrywise matrix norm called the Frobenius norm,

\[ \| A \|_F := \sqrt{\sum_{i,j} a_{ij}^2}. \]
• However, interpreting matrices as linear functions, what we are really interested in is the maximum amplification of the norm of any vector multiplied by the matrix,

\[ \|A\| := \max_{\|x\|=1} \|Ax\| . \]

These are called induced matrix norms, as each is associated with a specific vector norm \(\|\cdot\|\).

• The following are equivalent:

\[
\max_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\| \neq 0} \frac{x}{\|x\|} A \frac{x}{\|x\|} \|y\| = 1 \max_{\|y\| = 1} \|Ay\| = \|A\|. 
\]
• Logically, for each vector norm, we get a different matrix norm, so that, e.g. for the vector 2-norm \( \|x\|_2 \) we get a matrix 2-norm \( \|A\|_2 \), and for the vector \( \infty \)-norm \( \|x\|_\infty \) we get a matrix \( \infty \)-norm \( \|A\|_\infty \).
Demo: Matrix norms (click to visit)

In-class activity: Matrix norms
Matrix norms inherit the vector norm properties:

1. \( \|A\| > 0 \iff A \neq 0 \).
2. \( \|\gamma A\| = |\gamma| \|A\| \) for all scalars \( \gamma \).
3. Obeys triangle inequality \( \|A + B\| \leq \|A\| + \|B\| \)

But also some more properties that stem from our definition:

1. \( \|Ax\| \leq \|A\| \|x\| \)
2. \( \|AB\| \leq \|A\| \|B\| \) (easy consequence)

Both of these are called submultiplicativity of the matrix norm.
What is the 2-norm of an orthogonal matrix?

Linear Algebra recap: For an orthogonal matrix $A$, $A^{-1} = A^T$. In other words: $AA^T = A^TA = I$.

Next:

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$$

where

$$\|Ax\|_2 = \sqrt{(Ax)^T(Ax)} = \sqrt{x^T(A^TA)x} = \sqrt{x^Tx} = \|x\|_2,$$

so $\|A\|_2 = 1$. 
Now, let’s study condition number of solving a linear system $Ax = b$. 

**Input:** $b$ with error $\Delta b$, 

**Output:** $x$ with error $\Delta x$. 
Observe $A(x + \Delta x) = (b + \Delta b)$, so $A\Delta x = \Delta b$.

\[
\frac{\text{rel err. in output}}{\text{rel err. in input}} = \frac{\|\Delta x\|/\|x\|}{\|\Delta b\|/\|b\|} = \frac{\|\Delta x\|\|b\|}{\|\Delta b\|\|x\|} = \frac{\|A^{-1}\Delta b\|\|Ax\|}{\|\Delta b\|\|x\|} \leq \frac{\|A^{-1}\|\|A\|\|\Delta b\|\|x\|}{\|\Delta b\|\|x\|} = \|A^{-1}\|\|A\|.
\]

So we’ve found an upper bound on the condition number. With a little bit of fiddling, it’s not too hard to find examples that achieve this bound, i.e. that it is tight.
So we’ve found the condition number of linear system solving, also called the condition number of the matrix $A$:

$$\text{cond}(A) = \kappa(A) = \|A\| \|A^{-1}\|.$$ 

• **cond** is relative to a given norm. So, to be precise, use $\text{cond}_2$ or $\text{cond}_\infty$.

• If $A^{-1}$ does not exist: $\text{cond}(A) = \infty$ by convention.
Demo: Condition number visualized (click to visit)
Demo: Conditioning of 2x2 Matrices (click to visit)
What is $\text{cond}(A^{-1})$?

$$\text{cond}(A^{-1}) = \|A\| \cdot \|A^{-1}\| = \text{cond}(A).$$

What is the condition number of applying the matrix-vector multiplication $Ax = b$? (I.e. now $x$ is the input and $b$ is the output)

Let $B = A^{-1}$.

Then computing $b = Ax$ is equivalent to solving $Bb = x$.

Solving $Bb = x$ has condition number

$$\text{cond}(B) = \text{cond}(A^{-1}) = \text{cond}(A).$$
So the operation ‘multiply a vector by matrix $A$’ has the same condition number as ‘solve a linear system with matrix $A$’.
Give an example of a matrix that is very well-conditioned. (I.e. has a condition-number that’s good for computation.)

What is the best possible condition number of a matrix?

*Small* condition numbers mean *not a lot of error amplification*. *Small* condition numbers are good.

The identity matrix $I$ should be well-conditioned:

$$
\| I \| = \max_{\|x\|=1} \|Ix\| = \max_{\|x\|=1} \|x\| = 1.
$$

It turns out that this is the smallest possible condition number:

$$
1 = \| I \| = \| A \cdot A^{-1} \| \leq \| A \| \cdot \| A^{-1} \| = \kappa(A).
$$
Both of these are true for any norm $\| \cdot \|$. 
What is the 2-norm condition number of an orthogonal matrix $A$?

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \|A\|_2 \|A^T\|_2 = 1.$$  

That means orthogonal matrices have optimal conditioning. They’re very well-behaved in computation.
In-class activity: Matrix Conditioning