Last time:

1) Solving systems of linear equations via Gaussian elimination

- A linear system has either a unique solution, infinitely many solutions, or no solutions.
- Any linear system with m equations and n unknowns can be written in the form $Ax = b$ where $A$ is an $m \times n$ matrix, $x$ is $n \times 1$ and $b$ is $m \times 1$.
- Any matrix can be reduced to reduced row echelon form (rref) by Gaussian eliminations.
- The possible solution(s) (or no solution) can be seen by identifying pivots and free variables in the rref system.

2) Norms of vectors

o) A norm $\|x\|$ (i.e. the 'length' of the vector $x$) is a real-valued function satisfying the following three properties:

i) $\|x\| \geq 0$ with $\|x\| = 0$ if and only if (iff) $x = 0$

ii) $\|sx\| = |s| \|x\|$ for any scalar $s$ and vector $x$

iii) $\|x+y\| \leq \|x\| + \|y\|$ for any vectors $x, y$. (Triangle inequality)
Example of norms in $\mathbb{C}^n$

1) Euclidean norm ($\ell_2$-norm)

$$\lVert x \rVert_2 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2}$$

for $x \in \mathbb{C}^n$

2) $\ell_1$-norm:

$$\lVert x \rVert_1 = |x_1| + |x_2| + \ldots + |x_n| = \sum_{j=1}^{n} |x_j|$$

Remark: Important in statistical learning theory, machine learning, and sparse signal processing.

3) $\ell_p$-norm for $1 \leq p < \infty$

$$\lVert x \rVert_p := \left( |x_1|^p + |x_2|^p + \ldots + |x_n|^p \right)^{1/p} = \left( \sum_{j=1}^{n} |x_j|^p \right)^{1/p}$$

4) $\ell_\infty$-norm (maximum norm)

$$\lVert x \rVert_\infty := \max \{|x_1|, |x_2|, \ldots, |x_n|\}$$
Why is $\|\cdot\|_1$ a norm?

1. First of all, a sum of non-negative numbers will be non-negative. And only $x = 0$ gives $\|x\|_1 = 0$.

2. Let $s$ be a scalar

$$\|sx\|_1 = |sx_1| + |sx_2| + \ldots + |sx_n| = |s||x_1| + |s||x_2| + \ldots + |s||x_n|$$

$$= |s|\|x\|_1$$

3. Let $x, y$ be vectors

$$\|x + y\|_1 = \sum_{j=1}^{n} |(x + y)_j| = |x_1 + y_1| + |x_2 + y_2| + \ldots + |x_n + y_n|$$

$$\leq |x_1 + y_1| + |x_2 + y_2| + \ldots + |x_n + y_n|$$

$$= |x_1| + |x_2| + \ldots + |x_n| + |y_1| + |y_2| + \ldots + |y_n|$$

$$= \|x\|_1 + \|y\|_1$$
5) What is with $p < 1$?
   For $p \in (0, 1)$, the object $\| \cdot \|_p$ is not a norm.
   It does not satisfy the triangle inequality!
   It is sometimes called a semi-norm.

6) What about $\| \mathbf{x} \| = x_1 + x_2 + \ldots + x_n$
   This is not a norm!
   \[ \| [-1, 1, 0, 0] \| = 0 \quad \text{but} \quad [-1, 1, 0] + [0, 0, 0] \]

Example:
\[
\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}
\]
Let's calculate the $\ell_1$, $\ell_2$ and $\ell_\infty$ norms.
\[ \|a\|_1 = 111 + (101 + 12) + 1-11 = 4 \]
\[ \|a\|_2 = \sqrt{111^2 + 101^2 + 12^2 + 1-11^2} = \sqrt{6} \approx 2.45 \]
\[ \|a\|_\infty = \max \{111, 101, 12, 1-11\} = 2 \]
\[ \|b\|_1 = 1-11 + 11 + 1-11 = 3 \]
\[ \|b\|_2 = \sqrt{1-11^2 + 11^2 + 1-11^2} = \sqrt{3} \approx 1.73 \]
\[ \|b\|_\infty = \max \{1-11, 11, 1-11\} = 1 \]

We see \[ \|a\|_\infty < \|a\|_2 < \|a\|_1 \], and \[ \|b\|_\infty < \|b\|_2 < \|b\|_1 \]

Is this always the case? (The answer is yes, but \( \leq \) not \( < \))

Let \( x \) be a vector in \( \mathbb{C}^n \)
\[ \|x\|_\infty = \max \{|x_1|, |x_2|, \ldots, |x_n|\} \]
\[ \leq |x_1| + |x_2| + \ldots + |x_n| = \|x\|_2 \]
\[ \|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2} \geq \sqrt{|x_1|^2} = |x_1| \]
In the same way, \( \|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \) for all \( i \in \{1, \ldots, n\} \).

Therefore
\[
\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \leq \|x\|_2
\]

Now to the relation between the \( l_1 \) and \( l_2 \) norm:
\[
\|x\|_2^2 = |x_1|^2 + |x_2|^2 + \ldots + |x_n|^2
\]
\[
\|x\|_1^2 = (|x_1| + |x_2| + \ldots + |x_n|)^2 = |x_1|^2 + |x_2|^2 + \ldots + |x_n|^2 + \text{a lot of non-negative terms}
\]

\[
\Rightarrow \quad \|x\|_2^2 \leq \|x\|_1^2 \quad \Rightarrow \quad \|x\|_2 \leq \|x\|_1
\]

Fact:

For any vector \( x \in \mathbb{C}^n \) we have

\[
\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1
\]
However, $\|x\|_\infty$ gives us the largest magnitude of the elements of $x$. Let $\|x\|_2 = x_n$ without loss of generality.

$$\|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2} \leq \sqrt{|x_n|^2 + (x_n)^2 + \ldots + |x_n|^2}$$

$$= \sqrt{n|x_n|^2} = \sqrt{n} |x_n| = \sqrt{n} \|x\|_\infty$$

$$\Rightarrow \|x\|_2 \leq \sqrt{n} \|x\|_\infty$$

As for $\|x\|_2$ and $\|x\|_1$, consider

$$\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n| = \underbrace{|x_1| + |x_2| + \ldots + |x_n|}_{\text{Cayley-Schwarz inequality}}$$

Therefore:

$$\|x\|_1 \leq n \|x\|_\infty$$
Matrix norms

We want to measure the 'size' of a matrix $A$

Example:

$$A = \begin{bmatrix} 1 & 7 \\ 3 & 2 \end{bmatrix}$$

Let's think about $A$ as a vector with entries written in a box and not in a column like

$$a = \begin{bmatrix} 1 \\ 3 \\ 7 \\ 2 \end{bmatrix} \quad \Rightarrow \quad \|a\|_2 = \sqrt{1^2 + 3^2 + 7^2 + 2^2} = \sqrt{63}$$

We not just say $\|A\|_2 = \sqrt{63}$

In fact this can be done and leads to the *Hilbert–Schmidt norm* of a matrix.
Defn (Hilbert–Schmidt norm)

Let $A$ be an $m \times n$ matrix with entries $a_{ij}$. Then the Hilbert–Schmidt norm of $A$ is defined by

$$\|A\|_{HS} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

Example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$\|A\|_{HS} = \sqrt{\sum_{i=1}^{2} \sum_{j=1}^{3} |a_{ij}|^2}$$

$$= \sqrt{(111^2 + 121^2 + 131^2 + 141^2 + 151^2 + 161^2)}$$

This norm is unfortunately not closely tied to the action of $A$ as a linear transformation.
The operator norm

An $m \times n$ matrix $A$ can be thought of as a map

$\mathbb{R}^n \rightarrow \mathbb{R}^m \ ; \ x \mapsto Ax$

We want to say that a matrix is big if it increases the size of vectors, that is to say if $\|Ax\|$ is big compared to $\|x\|$. Thus let's look at the stretching ratio

$$\frac{\|Ax\|}{\|x\|}$$

This ratio depends on $x$ and is not defined if $x = 0$!

Moreover it makes little sense to measure the vector $Ax$ in for example 2-norm and the vector $x$ in another norm. We will always use the same norm for both.
Def (Operator norm)

The operator norm (or matrix norm) of $A$ is given as the maximum stretching factor (the largest stretching ratio) of $A$

$$\|A\|_{(op)} = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

It measures the maximum factor by which $A$ can stretch a vector.

Remark: One can check that this is indeed a norm!

Example: It is hard to calculate $\|A\|$ for general matrices $A$ but it is easier for diagonal matrices.

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$
\[ \| D x \|_2 = \left\| \begin{bmatrix} 2x_1 \\ 3x_2 \\ x_3 \end{bmatrix} \right\|_2 = \sqrt{12x_1^2 + 13x_2^2 + 13x_3^2} \]

This works for all \( x \in \mathbb{C}^n \)

\[ \leq \sqrt{13x_1^2 + 13x_2^2 + 13x_3^2} = 3 \sqrt{1x_1^2 + 1x_2^2 + 1x_3^2} = 3 \| x \|_2 \]

\[ \Rightarrow \frac{\| D x \|_2}{\| x \|_2} \leq 3 \quad \text{for all} \ x \ \text{in} \ \mathbb{C}^n \quad \Rightarrow \quad \| D \| \leq 3 \quad (*) \]

On the other hand let \( \tilde{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \), then \( \| \tilde{x} \|_1 = 1 \) and

\[ D \tilde{x} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} \quad \Rightarrow \quad \| D \tilde{x} \|_2 = 3 \quad \Rightarrow \quad \frac{\| D \tilde{x} \|_2}{\| \tilde{x} \|_2} = 3 \]

Therefore using the definition of the operator norm and \( (*) \) we get

\[ \| D \| = 3 \]