2.1 Norms. If up to this point we have talked about exactly solving systems of equations, in the next module we will start thinking about approximation, or finding best fits, or sometimes projections.

The first step toward this is defining distances, so we can measure how good the approximation is (or, in certain cases, to ensure we can prove that iterative methods converge, or to understand how the matrix properties may affect the size of the floating point error).

The simplest distance between two vectors is, naturally, the Euclidean one:

$$d(x, y) = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + \ldots + |x_n - y_n|^2}.$$ 

What makes this distance very natural is that it is given by a norm (which we define below); it is a handy way to measure distance, but it isn’t always the best, and we can define many, many more. The most important distances are given by norms.

**Definition** A norm is a function $||·|| : \mathbb{R}^n \to \mathbb{R}$ which satisfies the following three properties:

- $||x|| > 0$ for all $x \neq 0$; (positivity)
- $||αx|| = |α| · ||x||$, for all $α \in \mathbb{R}$ and $x \in \mathbb{R}^n$, (multiplication by a constant)
- $||x + y|| \leq ||x|| + ||y||$, for all $x, y \in \mathbb{R}^n$. (triangle inequality)

**Remark 1.** *Any norm defines a distance via $d(x, y) = ||x - y||$.***

**Examples of norms:**

- $||x||_1 = |x_1| + |x_2| + \ldots + |x_n|,$
- $||x||_2 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2},$
- $||x||_p = \left( |x_1|^p + |x_2|^p + \ldots + |x_n|^p \right)^{1/p},$

for any $p \geq 1$ a real number. The above are called “$p$”-norms; the 1-norm is sometimes known as the “Manhattan norm”, the 2-norm is the Euclidean norm.

Also if $||·||$ is a norm and $A$ is a matrix, under certain circumstances, $||x||_A := ||Ax||$ will be a norm.

One interesting feature of these norms is the shape of the unit sphere (the set of points of norm 1). For example, for $n = 2$, we can look at the picture below.

The unit sphere in 1-norm is the square whose vertices are at $(\pm 1, 0)$ and $(0, \pm 1)$. In the 2-norm, the unit sphere is simple the unit circle. As we take a larger and larger $p$-norm, the unit sphere flattens, and approaches the square with vertices at $(\pm 1, \pm 1)$. This is not accidental, but rather due to the existence of the $\infty$-norm:

$$||x||_\infty = \max_i |x_i|,$$
Matrix norms. All matrices can be seen as vectors with extra structure. We will focus mostly on square, $n \times n$ matrices, although one can define matrix (semi)norms for rectangular matrices as well.

Definition A matrix norm is a function defined $\| \cdot \| : \mathbb{R}^{n \times n} \to \mathbb{R}$ which satisfies the following four properties:

- $\|A\| > 0$ for all $A \neq 0$; (positivity)
- $\|\alpha A\| = |\alpha| \cdot \|A\|$, for all $\alpha \in \mathbb{R}$ and $A \in \mathbb{R}^{n \times n}$, (multiplication by a constant)
- $\|A + B\| \leq \|A\| + \|B\|$, for all $A, B \in \mathbb{R}^{n \times n}$, (triangle inequality)
- $\|AB\| \leq \|A\| \cdot \|B\|$, for all $A, B \in \mathbb{R}^{n \times n}$. (submultiplicativity)

Remark 2. Note that the first three conditions are the same as for vectors, but the fourth (submultiplicativity) is new.

For example, examine the following matrix norm, also known as the Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}.$$ 

This, at first glance, looks like the 2-norm for vectors. Indeed, the first three properties of this matrix norm follow straightforwardly from the fact that they are true for a vector norm. However, the fourth one is not immediate. Let us examine it:

$$\|AB\|_F^2 = \sum_{i,j} |AB_{ij}|^2 = \sum_{i,j} \left( \sum_{k=1}^n a_{ik} b_{kj} \right)^2 .$$ (1)
The Cauchy-Schwarz theorem says that
\[
\left( \sum_{i=1}^{n} a_i b_i \right)^2 \leq \left( \sum_{i=1}^{n} a_i^2 \right) \left( \sum_{i=1}^{n} b_i^2 \right) ;
\]
if we apply this to (1), we obtain that
\[
||AB||_F^2 \leq \sum_{i,j} \left( \sum_{k=1}^{n} |a_{ik}|^2 \right) \left( \sum_{k=1}^{n} |b_{kj}|^2 \right),
\]
and so we may split the right hand side as
\[
\sum_{i,j} \left( \sum_{k=1}^{n} |a_{ik}|^2 \right) \left( \sum_{k=1}^{n} |b_{kj}|^2 \right) = \left( \sum_{i,k} |a_{ik}|^2 \right) \left( \sum_{k,j} |b_{kj}|^2 \right)
= ||A||_F^2 \cdot ||B||_F^2 .
\]
This proves submultiplicativity.

**Induced matrix norms.** Any vector norm induces a matrix one, as follows.

**Definition** Given the \( n \)-dimensional vector norm \( ||\cdot|| \), the induced matrix norm on \( n \times n \) matrices is defined as
\[
||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||} .
\]

The first three properties of the induced norm follow fairly easily from the three properties of the corresponding vector norm. The fourth is more subtle:
\[
||AB|| = \max_{x \neq 0} \frac{||ABx||}{||x||} = \max_{x \neq 0, Bx \neq 0} \frac{||ABx||}{||Bx||} \cdot \max_{x \neq 0} \frac{||Bx||}{||x||} .
\]

Note that, by denoting \( y = Bx \), we can rewrite
\[
\max_{x \neq 0, Bx \neq 0} \frac{||ABx||}{||Bx||} = \max_{y \neq 0} \frac{||Ay||}{||y||} = ||A|| ,
\]
and thus
\[
||AB|| \leq ||A|| \max_{x \neq 0} \frac{||Bx||}{||x||} = ||A|| \cdot ||B|| .
\]

**Important matrix norms.** \( ||A||_1 \), the induced matrix 1-norm, can be seen to be
\[
||A||_1 = \max_j \sum_{i=1}^{n} |a_{ij}| ;
\]
this is the maximum column 1-norm; $\|A\|_\infty$, the induced $\infty$-norm, is actually

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|,$$

this is the maximum row 1-norm (NOT the maximum row $\infty$-norm!) Also, $\|A\|_2$ and $\|A\|_F$ are different, as you can see from the fact that if $I$ is the $n$-dimensional identity, $\|I\|_2 = 1$, whereas $\|I\|_F = \sqrt{n}$. 