1 Vectors and Matrix Norms (2.1)

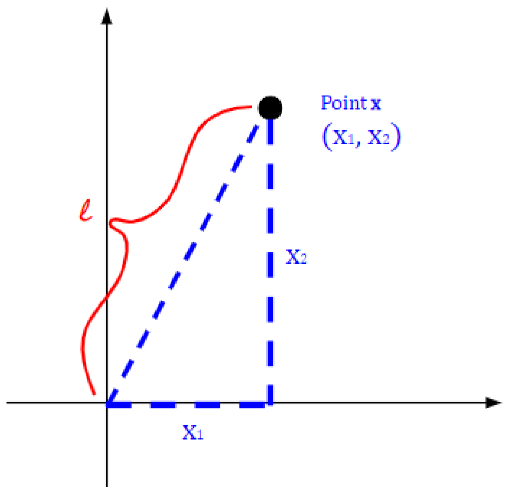
In numerical analysis, we want to find approximate solutions to problems (e.g., ODEs). Some things we want to know are

- How good the approximate solution is?
- How close is the approximate solution to the exact solution?

So, **norms** are a measure of length, or the measure of being close or far apart.

1.1 Vector Norms

The vector norm we're most familiar with is the one in $\mathbb{R}^2$, also known as the *2-norm*. These might look something like



For a point

$$\mathbf{x} = (x_1, x_2),$$

we can define the *2-norm* of a vector to be

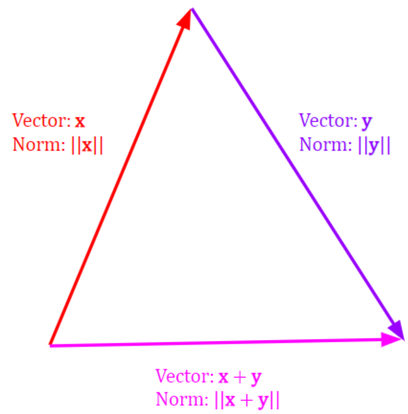
$$||\mathbf{x}||_2 = \sqrt{x_1^2 + x_2^2}.$$

Definition 1.1: Vector Norm

A **norm** of a vector (i.e., a **vector norm**) $\mathbf{x} \in \mathbb{R}^n$ is a real number $||x||$ that is assigned to $\mathbf{x}$. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all $c \in \mathbb{R}$, the following properties are satisfied:

1. Positive Definite Property: $||\mathbf{x}|| > 0$ for $\mathbf{x} \neq 0$ and $||\mathbf{0}|| = 0$.
2. Absolute Homogeneity: $||c\mathbf{x}|| = |c| \cdot ||\mathbf{x}||$.
3. Triangle Inequality: $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.

Remark: With regards to the third property, consider



Note that

$$||\mathbf{x}+\mathbf{y}|| = ||\mathbf{x}|| + ||\mathbf{y}||$$

if $\mathbf{x}$ and $\mathbf{y}$ points at the same direction.

1.1.1 Popular Norms

There are some common norms that we've seen before. As implied, they all satisfy the properties above.

- For $p \geq 1$, we define
$$||\mathbf{x}||_p = \sqrt[p]{|x_1|^p + |x_2|^p + \ldots + |x_n|^p}.$$
Note that some special cases are
$$||\mathbf{x}||_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$
and
$$||\mathbf{x}||_1 = |x_1| + |x_2| + \ldots + |x_n| = \sum_{i=1}^{n} |x_i|.$$
- The infinite norm is defined to be
$$||\mathbf{x}||_\infty = \max_{i=1,2,\ldots,n} |x_i|.$$
It should be noted that
$$\lim_{p \mapsto \infty} ||\mathbf{x}||_p = ||\mathbf{x}||_\infty.$$

1.2 Matrix Norms

We now want to consider norms for a matrix $A \in \mathbb{R}^{n\times n}$. There are two ways we can interpret matrix norms.

1. Interpret matrix as a vector. For example, suppose we have
$$A = \begin{bmatrix} -1 & 0 & 5 \\ 8 & 2 & 7 \\ -3 & 1 & 0 \end{bmatrix}.$$

Then, we can "convert" this matrix to a vector like so:

$$\mathbf{v} = \begin{bmatrix} -1 \\ 0 \\ 5 \\ 8 \\ 2 \\ 7 \\ -3 \\ 1 \\ 0 \end{bmatrix}.$$

Here, $\mathbf{v} \in \mathbb{R}^9$. Notice how the first column of A is the top three elements in $\mathbf{v}$, the second column of A is the middle three elements of $\mathbf{v}$, and the last column of A is the bottom three elements of $\mathbf{v}$.

2. We can also define the matrix as a linear operator. That is, for a function $L : \mathbb{R}^n \mapsto \mathbb{R}^n$, we have

$$L(\mathbf{x}) = A\mathbf{x}.$$

1.2.1 General Definition of Matrix Norms

Definition 1.2: Matrix Norm

A **matrix norm** assigns a real number $||A||$ to a matrix A. This should satisfy the following conditions for all $A, B \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R}$.

1. $||A|| > 0$ if $A \neq 0$, and $||0|| = 0$.
2. $||cA|| = |c| \cdot ||A||$.
3. $||A + B|| \leq ||A|| + ||B||$.
4. Submultiplicity: $||AB|| \leq ||A|| \cdot ||B||$.

Remark: Regarding submultiplicity, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have

$$|\langle x, y\rangle| \leq ||\mathbf{x}||_2 ||y||_2,$$

known as the Cauchy Schwarz inequality.

1.2.2 Vector Viewpoint

Going back to the vector viewpoint, let's suppose we have

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}.$$

Then,

$$\mathbf{v} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \\ a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \\ \vdots \\ a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{bmatrix}.$$

The Frobenius norm of A is defined by

$$||A||_F = ||\mathbf{v}||_2 = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}}.$$

1.2.3 Matrix Norm

Matrix p-norms are defined as follows:

$$||A||_p = \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \mathbf{x} \neq \mathbf{0}}} \frac{||A\mathbf{x}||_p}{||\mathbf{x}||_p}.$$

This measures the *maximum stretch* the linear function $L(\mathbf{x}) = A\mathbf{x}$ can do to a vector (normalized by the length of the vector).

Some of the most important matrix p-norms are

- For $p = 1$: $||A||_1 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{||A\mathbf{x}||_1}{||\mathbf{x}||_1} = \max_{j=1,2,\ldots,n} \sum_{i=1}^{n} |a_{ij}|$. (maximum L_1-norm of each column.)
- For $p = \infty$: $||A||_\infty = \max_{\mathbf{x} \neq \mathbf{0}} \frac{||A\mathbf{x}||_\infty}{||\mathbf{x}||_\infty} = \max_{i=1,2,\ldots,n} \sum_{j=1}^{n} |a_{ij}|$. (maximum L_1-norm of each row.)
- For $p = 2$: $||A||_2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{||A\mathbf{x}||_2}{||\mathbf{x}||_2} = \sigma_1$. (largest[1] singular value of matrix A.)

Remark: Don't confuse $||A||_2$ and $||A||_F$. They are very different! Specifically,

$$||A||_2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{||A\mathbf{x}||_2}{||\mathbf{x}||_2}$$

while

$$||A||_F = ||\mathbf{v}||_2 = \left(\sum_{i,j} (a_{ij})^2 \right)^{\frac{1}{2}}.$$

[1] This is related to SVD, which we will learn later.