Norms

By equipping a vector space $\mathcal{S}$ with a norm, we are imbuing it with a sense of **length** and **distance**. Another way to say this is that a norm adds a layer **topological structure** on top of the algebraic structure defining a linear space.

Definition. A **norm** $\|\cdot\|$ on a vector space $\mathcal{S}$ is a mapping

$$\|\cdot\| \;:\; \mathcal{S} \to \mathbb{R}$$

with the following properties for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$:

1. $\|\boldsymbol{x}\| \geq 0$, and $\|\boldsymbol{x}\| = 0 \;\Leftrightarrow\; \boldsymbol{x} = \boldsymbol{0}$.
2. $\|\boldsymbol{x} + \boldsymbol{y}\| \;\leq\; \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ (triangle inequality)
3. $\|a\boldsymbol{x}\| = |a| \cdot \|\boldsymbol{x}\|$ for any scalar a (homogeneity)

Other related definitions:

- The **length** of $\boldsymbol{x} \in \mathcal{S}$ is simply $\|\boldsymbol{x}\|$
- The **distance** between $\boldsymbol{x}$ and $\boldsymbol{y}$ is $\|\boldsymbol{x} - \boldsymbol{y}\|$
- A linear vector space equipped with a norm is called a **normed linear space**.

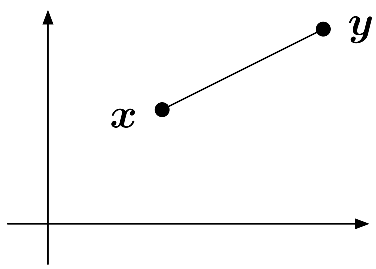
Examples:

1. $\mathcal{S} = \mathbb{R}^N$,

$$\|\boldsymbol{x}\|_2 = \left(\sum_{n=1}^{N} |x_n|^2\right)^{1/2}$$

This is called the "ℓ_2 norm" or "standard Euclidean norm"

In $\mathbb{R}^2$:



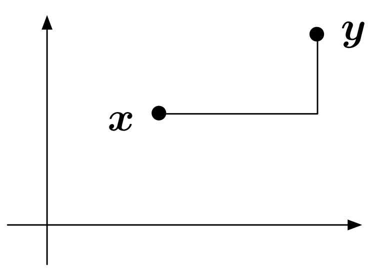
$$\|\boldsymbol{x}-\boldsymbol{y}\|_2 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

2. $\mathcal{S} = \mathbb{R}^N$

$$\|\boldsymbol{x}\|_1 = \sum_{n=1}^{N} |x_n|$$

This is the "ℓ_1 norm" or "taxicab norm" or "Manhattan norm"

In $\mathbb{R}^2$:



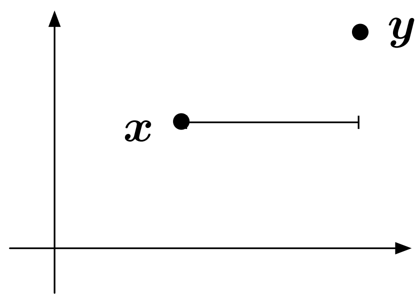
$$\|\boldsymbol{x} - \boldsymbol{y}\|_1 = |x_1 - y_1| + |x_2 - y_2|$$

3. $\mathcal{S} = \mathbb{R}^N$

$$\|\boldsymbol{x}\|_\infty = \max_{n=1,\ldots,N} |x_n|$$

This is the "ℓ_∞ norm" or "Chebyshev norm" or "max norm"

In $\mathbb{R}^2$:



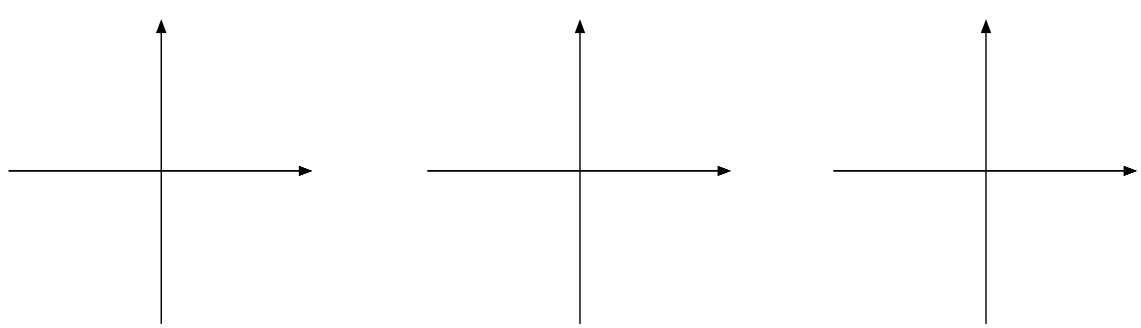
$$\|\boldsymbol{x}-\boldsymbol{y}\|_\infty = \max\left(|x_1 - y_1|,\ |x_2 - y_2|\right)$$

4. $\mathcal{S} = \mathbb{R}^N$

$$\|\boldsymbol{x}\|_p = \left(\sum_{n=1}^{N} |x_n|^p\right)^{1/p} \qquad \text{for some } 1 \le p < \infty$$

This is the "ℓ_p norm".

Draw the "ℓ_p unit balls" $\mathcal{B}_p = \{\boldsymbol{x} \in \mathbb{R}^2\ :\ \|\boldsymbol{x}\|_p \le 1\}$



$p = 1$ $\qquad$ $p = 2$ $\qquad$ $p = \infty$

Georgia Tech ECE 7750 Notes by J. Romberg. Minor updates for Fall 22 offering by Ashwin Pananjady: 12:24, August 31, 2022

5. The same definitions extend easily to infinite sequences:
$\mathcal{S}$ = sequences $\{x_n\}_{n\in\mathbb{Z}}$ indexed by the integers $n \in \mathbb{Z}$

$$\|\boldsymbol{x}\|_p = \left(\sum_{n=-\infty}^{\infty} |x_n|^p\right)^{1/p}$$

You can verify at home that the set of all sequences that have $\|\boldsymbol{x}\|_p < \infty$ is a (normed) linear space; we call this space ℓ_p.

6. $\mathcal{S}$ = real-valued functions on an interval $[a, b]$:

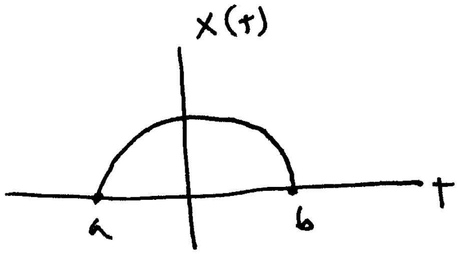
$$\|\boldsymbol{x}\|_p = \left(\int_a^b |x(t)|^p\right)^{1/p}$$
$$\|\boldsymbol{x}\|_\infty = \sup_{t\in[a,b]} |x(t)|$$

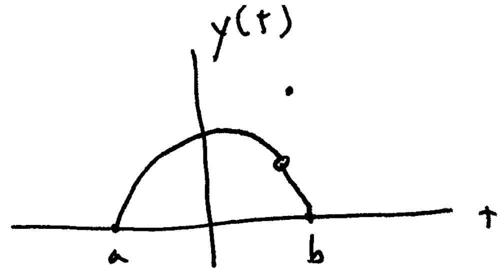
The normed linear space of all signals on the interval $[a, b]$ with finite L_p norm is called $L_p([a, b])$. This definition extends in the obvious way to complex-valued functions as well.

In a normed linear space, we say that

$$\boldsymbol{x} = \boldsymbol{y} \quad \text{if} \quad \|\boldsymbol{x} - \boldsymbol{y}\| = 0.$$

For example, in $L_2([a, b])$, say $y(t) = x(t)$ except at one point





Then

$$\|\boldsymbol{x}-\boldsymbol{y}\|_2 \;=\; \left(\int_a^b |x(t)-y(t)|^2\right)^{1/2} \;=\; 0$$

and so we still say that $\boldsymbol{x}=\boldsymbol{y}$. In general, if $\boldsymbol{x},\boldsymbol{y}\in L_p$ differ only on a "set of measure zero", then $\boldsymbol{x}=\boldsymbol{y}$.

(A set $\Gamma\subset\mathbb{R}$ has measure zero if

$$\int I_\Gamma(t)\;dt = 0,$$

where

$$I_\Gamma(t) = \begin{cases} 1 & t\in\Gamma \\ 0 & t\notin\Gamma \end{cases}$$

is an indicator function.)

Exercises

1. "When traveling from point $\boldsymbol{x}$ to $\boldsymbol{y}$, passing through an intermediate point $\boldsymbol{z}$ can only make your trip longer." Show that this common-sense statement holds in every normed linear space. That is, show

$$\|\boldsymbol{x}-\boldsymbol{y}\| \;\leq\; \|\boldsymbol{x}-\boldsymbol{z}\|+\|\boldsymbol{y}-\boldsymbol{z}\| \tag{1}$$

for every $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ in the space. Draw a picture that illustrates this inequality[1].

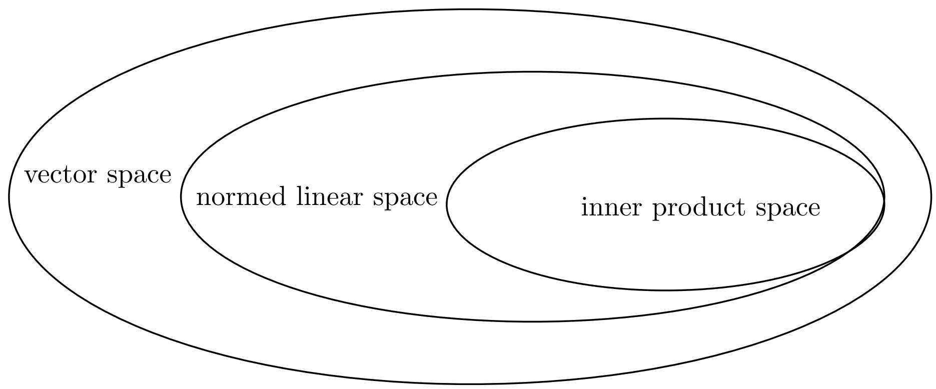
2. Show that equality in (1) holds if $\boldsymbol{z}$ is on a line between $\boldsymbol{x}$ and $\boldsymbol{y}$, that is

$$\boldsymbol{z}=\tau \boldsymbol{x}+(1-\tau) \boldsymbol{y}, \quad \text{for some } \; 0 \leq \tau \leq 1.$$

[1] Your argument should be a very short application of the triangle inequality, and your picture should illustrate why this is a good name.

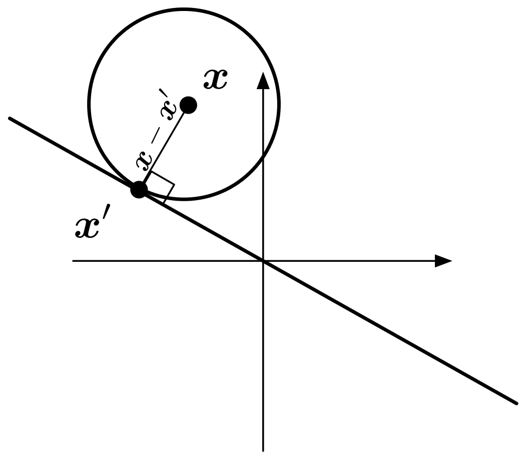
Inner product spaces

Layers of structure:



The abstract definition of an inner product, which we will see very shortly, is simple (and by itself is pretty boring). But it gives us just enough mathematical structure to make sense of many important and fundamental problems.

Consider the following motivating example in the plane $\mathbb{R}^2$. Let $\mathcal{T}$ be a one dimensional subspace (i.e. a line through the origin). Now suppose we are given a vector $\boldsymbol{x}$. What is the closest point $\boldsymbol{x}'$ in $\mathcal{T}$ to $\boldsymbol{x}$?



The salient feature of this point $\boldsymbol{x}'$ is that

$$\boldsymbol{x}-\boldsymbol{x}' \perp \boldsymbol{v} \quad \text{for all } \boldsymbol{v} \in \mathcal{T}.$$

So all we need to define this optimality property is the notion of **orthogonality** which follows immediately from defining an inner product. More on this later ...

Definition: An **inner product** on a vector space $\mathcal{S}$ is a mapping

$$\langle \cdot, \cdot \rangle \ : \ \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$$

that obeys[2]

1. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$

2. For any $a, b \in \mathbb{R}$

$$\langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{z} \rangle \ = \ a\langle \boldsymbol{x}, \boldsymbol{z} \rangle + b\langle \boldsymbol{y}, \boldsymbol{z} \rangle$$

[2] One can also define inner products as mappings to complex numbers, but we will avoid this for the moment.

3. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0 \quad \text{and} \quad \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$

Standard Examples:

1. $\mathcal{S} = \mathbb{R}^N$,
$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \;=\; \sum_{n=1}^{N} x_n y_n \;=\; \boldsymbol{y}^{\mathrm{T}} \boldsymbol{x}$$

2. $\mathcal{S} = L_2([a, b])$,
$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \;=\; \int_a^b x(t) y(t)\, dt$$

Slightly less standard examples:

1. $\mathcal{S} = \mathbb{R}^{M \times N}$ (the set of $M \times N$ matrices with real entries)
$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \operatorname{trace}(\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{X}) = \sum_{m=1}^{M} \sum_{n=1}^{N} X_{m,n} Y_{m,n}$$
(Recall that $\operatorname{trace}(\boldsymbol{X})$ is the sum of the entries on the diagonal of $\boldsymbol{X}$.) This is called the *trace inner product* or *Frobenius inner product* or *Hilbert-Schmidt inner product.*

2. $\mathcal{S} =$ zero-mean Gaussian random variables with finite variance,
$$\langle X, Y \rangle = \mathrm{E}[XY]$$

3. $\mathcal{S}$ = differentiable real-valued functions on $\mathbb{R}$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \int_{-\infty}^{\infty} x(t)y(t)\; dt + \int_{-\infty}^{\infty} x'(t)y'(t)dt,$$

where x' is the derivative of x. This is called a *Sobolev inner product*.

A linear vector space equipped with an inner product is called an **inner product space**.

Induced norms

A valid inner product induces a valid norm by

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

(Check this on your own as an exercise.)

In $\mathbb{R}^N$, the standard inner product induces the ℓ_2 norm.

Properties of induced norms

In addition to the triangle inequality,

$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|,$$

induced norms obey some very handy inequalities. See the end of this note for proofs of these inequalities.

Note: these are not necessarily true for norms in general, only for norms induced by an inner product.

1. **Cauchy-Schwarz Inequality**

   $$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \ \leq \ \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|$$

   Equality is achieved above when (and only when) $\boldsymbol{x}$ and $\boldsymbol{y}$ are **colinear**:

   $$\exists\, a \in \mathbb{R} \quad \text{such that} \quad \boldsymbol{y} = a\boldsymbol{x}.$$

2. **Pythagorean Theorem**

   $$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \ \Rightarrow \ \|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$$

The left-hand side above also implies that
$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

3. **Parallelogram Law**

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 + \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2.$$

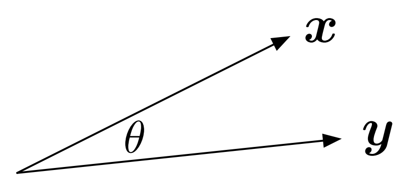
You can prove this by expanding $\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle$ and similarly for $\|\boldsymbol{x} - \boldsymbol{y}\|^2$.

4. **Polarization Identity**

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \ = \ \frac{\|\boldsymbol{x} + \boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2}{4}$$

Angles between vectors

In $\mathbb{R}^2$ (and $\mathbb{R}^3$), we are very familiar with the geometrical notion of an angle between two vectors.



We have

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \, \cos\theta$$

Notice that this relationship depends only on norms and inner products. Therefore, we can extend the definition to any inner product space.

Definition: The **angle** between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in an inner product space is

$$\cos\theta \;=\; \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|},$$

where the norm is the one induced by the inner product.

Definition: Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in an inner product space are **orthogonal** to one another if

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle \;=\; 0.$$

Example: (Weighted inner product)

$$\mathcal{S} = \mathbb{R}^2, \quad \langle \boldsymbol{x}, \boldsymbol{y} \rangle_Q = \boldsymbol{y}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{x}, \quad \text{where} \quad \boldsymbol{Q} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

so $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 4x_1y_1 + x_2y_2$. What is the norm induced by this inner product? Draw the unit ball $\mathcal{B}_Q = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ \|\boldsymbol{x}\|_Q \leq 1\}$.

Find a vector which is orthogonal to $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ under $\langle \cdot, \cdot \rangle_Q$.

Technical Details: Completeness, convergence, and Hilbert Spaces

As we can see from the properties above, inner product spaces have almost all of the geometrical properties of the familiar Euclidean space $\mathbb{R}^3$ (or more generally $\mathbb{R}^N$). In fact, in the coming sections, we will see that any space with an inner product defined (which comes with its induced norm) is directly analogous to Euclidean space.

To make all of this work nicely in infinite dimensions, we need a technical condition on $\mathcal{S}$ called **completeness**. Roughly, this means that there are no points "missing" from the space. Let us make this precise.

Let $\|\cdot\|$ denote the norm induced by the inner product, and define a *Cauchy sequence*—that is, a sequence $(\boldsymbol{x}_n)_{n=1}^{\infty}$ with elements in $\mathcal{S}$ is Cauchy if, for all $\epsilon > 0$, there is some integer $N(\epsilon)$ such that

$$\|\boldsymbol{x}_n - \boldsymbol{x}_m\| < \epsilon, \qquad \text{for all } n, m \geq N(\epsilon).$$

An inner product space in which every Cauchy sequence $(\boldsymbol{x}_n)_{n=1}^{\infty}$ converges to some element $\boldsymbol{x}^*$ of the space is said to be *complete*. A complete inner product space is called a *Hilbert space*.

An example of a metric space which is **not** complete is the space of continuous functions on $[0,1]$ equipped with the L_2 norm. It is easy to come up with a sequence of functions which are all continuous but converge to a discontinuous function[3].

[3]To do this, note that we can construct a sequence of continuous functions (f_n), where f_n takes constant value 0 left of $1/2 - 1/n$ and constant value 1 from $1/2$ to 1, with a linear interpolator between $1/2 - n$ and $1/2$. As an exercise, show that this sequence is Cauchy. However, note that this sequence converges to a discontinuous function in the limit!

However, when the space of continuous functions continuous functions is equipped with the L_∞ norm, it is complete. For more on this, and for an overview of the mathematics that make all of this discussion more concrete, see the "Analysis Fundamentals" auxiliary notes.

All finite dimensional inner product spaces are complete, which is a basic result in mathematical analysis. Determining whether or not a space is complete is far outside the scope of this course; it is enough for us to know what this concept means.

The point of asking that the space be complete is that it gives us confidence in writing expressions like

$$x(t) = \sum_{n=1}^{\infty} \alpha_n \psi_n(t).$$

What is on the left is a sum of an infinite number of terms; the equality above means that as we include more and more terms in this sum, it converges to something which we call $x(t)$. There are different ways we might define convergence, depending on how much of a role we want the order of terms to play in the result. But we say that $\sum_{n=1}^{\infty} \alpha_n \psi_n(t)$, where the $\psi_n(t)$ are in a Hilbert space $\mathcal{S}$, is convergent if there is an $x(t)$ such that

$$\left\| x(t) - \sum_{n=1}^{N} \alpha_n \psi_n(t) \right\| \to 0 \quad \text{as} \quad N \to \infty.$$

If $\mathcal{S}$ is complete, we know that $x(t)$ will also be in $\mathcal{S}$.

Finally, we note that this notion of completeness only depends on having a norm defined. A complete normed linear space (where the norm is not necessarily induced by an inner product) is called a **Banach space**.

Technical Details: Proofs of properties of induced norms

1. **Cauchy-Schwarz**
   Set
$$\boldsymbol{z} = \boldsymbol{x} - \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}\boldsymbol{y},$$
   and notice that $\langle \boldsymbol{z}, \boldsymbol{y} \rangle = 0$, since
$$\langle \boldsymbol{z}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}\|\boldsymbol{y}\|^2 = 0.$$
   We can write $\boldsymbol{x}$ in terms of $\boldsymbol{y}$ and $\boldsymbol{z}$ as
$$\boldsymbol{x} = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}\boldsymbol{y} + \boldsymbol{z},$$
   and since $\boldsymbol{y} \perp \boldsymbol{z}$,
$$\begin{aligned}\|\boldsymbol{x}\|^2 &= \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2}{\|\boldsymbol{y}\|^4}\|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 \\ &= \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2}{\|\boldsymbol{y}\|^2} + \|\boldsymbol{z}\|^2.\end{aligned}$$
   Thus
$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \ = \ \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 - \|\boldsymbol{z}\|^2\|\boldsymbol{y}\|^2 \ \leq \ \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2.$$
   We have equality above if and only if $\boldsymbol{z} = \boldsymbol{0}$. If $\boldsymbol{z} = \boldsymbol{0}$, then $\boldsymbol{x}$ is co-linear with $\boldsymbol{y}$, as
$$\boldsymbol{x} = \alpha\boldsymbol{y}, \quad \text{with} \ \ \alpha = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}.$$
   Conversely, if $\boldsymbol{x} = \alpha\boldsymbol{y}$ for some $\alpha \in \mathbb{R}$, then
$$\boldsymbol{z} = \alpha\boldsymbol{y} - \frac{\alpha\langle \boldsymbol{y}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}\boldsymbol{y} = \boldsymbol{0}.$$

2. **Pythagorean Theorem**

$$
\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}\rangle &= \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle \\
&= \|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 \qquad (\text{since } \langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0)
\end{aligned}
$$

3. **Parallelogram Law**. As above, we have

$$
\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2, \\
\|\boldsymbol{x}-\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 - \langle \boldsymbol{x}, \boldsymbol{y}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2,
\end{aligned}
$$

and so

$$
\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2.
$$

4. **Polarization Identity**. Using the expansions above, we quickly see that

$$
\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 4\langle \boldsymbol{x}, \boldsymbol{y}\rangle.
$$