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 of **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**.

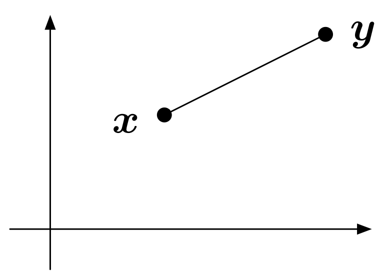
Examples:

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 $\mathbb{R}^2$:



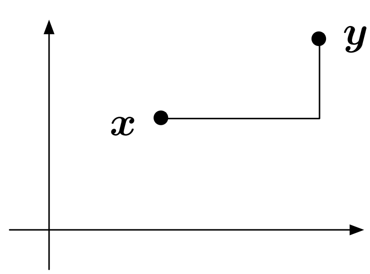
$$\|\boldsymbol{x}-\boldsymbol{y}\|_2 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

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"

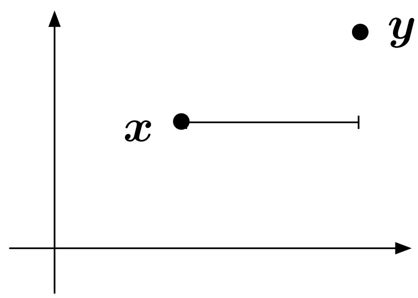
In $\mathbb{R}^2$:



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

3. $\mathcal{S} = \mathbb{R}^N$
$$\|\boldsymbol{x}\|_\infty = \max_{n=1,\ldots,N} |x_n|$$
This is the "ℓ_∞ norm" or "Chebyshev norm"

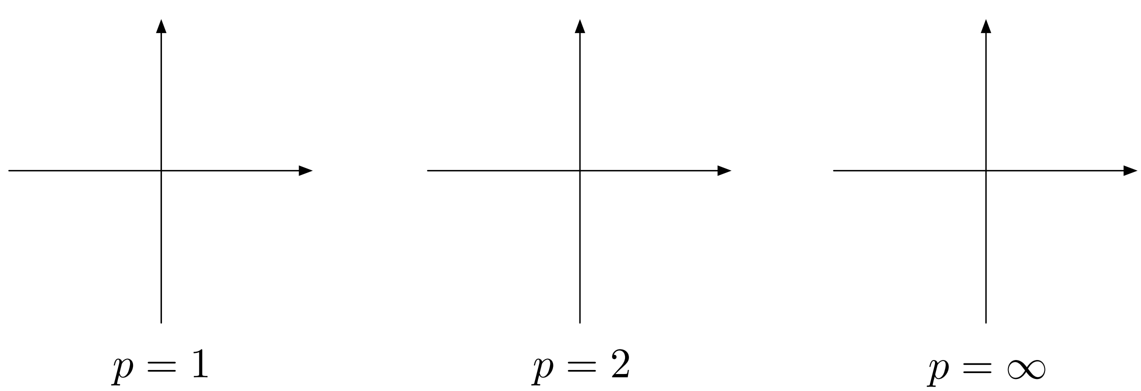
In $\mathbb{R}^2$:



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

4. $\mathcal{S} = \mathbb{R}^N$
$$\|\boldsymbol{x}\|_p = \left(\sum_{n=1}^{N} |x_n|^p\right)^{1/p} \qquad \text{for some } 1 \leq p < \infty$$
This is the "ℓ_p norm".

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



5. The same definitions extend straightforwardly to infinite sequences:
   $\mathcal{S}$ = sequences (discrete-time signals) $x[n]$ indexed by the integers $n \in \mathbb{Z}$

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

   It is easy to verify that the set of all discrete-time signals that have $\|\boldsymbol{x}\|_p < \infty$ is a (normed) linear space; we call this space ℓ_p.

6. $\mathcal{S}$ = continuous-time signals on the real line

$$\|x(t)\|_2 = \left(\int_{-\infty}^{\infty} |x(t)|^2 \ \mathrm{d}t \right)^{1/2}$$

   This is called the L_2 norm[1]. In engineering, we often refer to $\|x(t)\|_2^2$ as the **energy** in the signal.

   Similarly,

$$\|x(t)\|_p = \left(\int_{-\infty}^{\infty} |x(t)|^p \ \mathrm{d}t \right)^{1/p}$$

   and

$$\|x(t)\|_\infty = \sup_{t \in \mathbb{R}} |x(t)|, \quad \text{where sup = "least upper bound"}$$

[1]The L is for Lebesgue, the mathematician who formalized the modern theory of integration in the early 1900s.

Note that we are also using $\|\cdot\|_p$ for the discrete version of these norms, but I do not expect this will cause any confusion.

The set of continuous-time signals that have finite L_p norm are a normed linear space; we call this space $L_p(\mathbb{R})$.

7. $\mathcal{S}$ = continuous-time functions on an interval $[a, b]$:

$$\|x(t)\|_p = \left(\int_a^b |x(t)|^p\right)^{1/p}$$

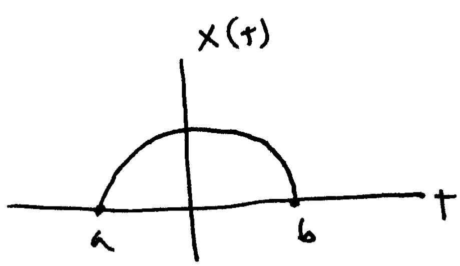
$$\|x(t)\|_\infty = \sup_{t\in[a,b]} |x(t)|$$

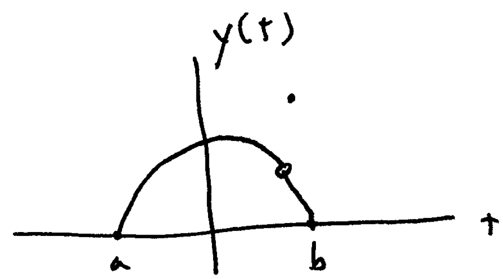
The normed linear space of all signals on the interval $[a, b]$ with finite L_p norm is called $L_p([a, b])$.

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

$$\int I_\Gamma(t) \ \mathrm{d}t = 0,$$

where

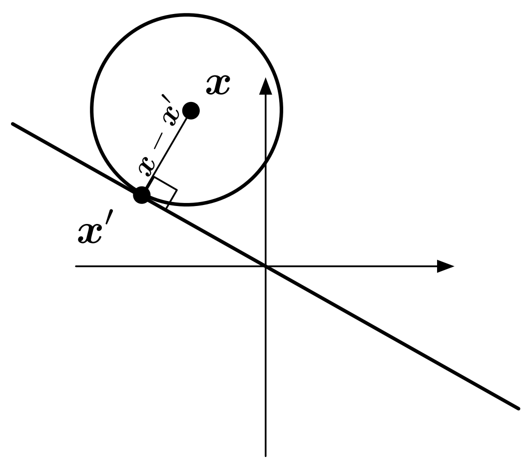
$$I_\Gamma(t) = \begin{cases} 1 & t \in \Gamma \\ 0 & t \notin \Gamma \end{cases}$$

is an indicator function.)

Inner products

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 another vector $\boldsymbol{x}$. What is the closest point 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 (real- or complex-valued) vector space $\mathcal{S}$ is a mapping

$$\langle \cdot, \cdot \rangle \ : \ \mathcal{S} \times \mathcal{S} \to \mathbb{C}$$

that obeys[2]

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

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

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

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

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \ = \ \sum_{n=1}^{N} x_n \, \overline{y_n} \ = \ \boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}$$

3. $\mathcal{S} = L_2([a, b])$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \ = \ \int_a^b x(t)\overline{y(t)} \ \mathrm{d}t$$

[2] We are using $\bar{a}$ to denote the complex conjugate of a scalar a, and $\boldsymbol{x}^{\mathrm{H}}$ to denote the conjugate transpose of a vector $\boldsymbol{x}$.

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 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,

$$\langle X, Y \rangle = \mathrm{E}[XY]$$

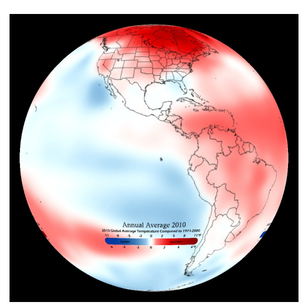
3. $\mathcal{S}$ = differentiable real-valued continuous-time signals on $\mathbb{R}$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \int x(t) y(t)\ \mathrm{d}t + \int x'(t) y'(t)\, \mathrm{d}t,$$

where $x'(t)$ is the derivative of $x(t)$. This is called a Sobolev inner product.

4. $\mathcal{S}$ = signals $x(\theta, \phi)$ on the sphere in $3D$

Difference in average temperature, 2010 vs. 1971-2000



A natural (and valid) inner product is

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} x(\theta, \phi)\, y(\theta, \phi) \sin\theta \;\, \mathrm{d}\phi \;\, \mathrm{d}\theta$$

If we think of θ as being latitude and ϕ as longitude, the $\sin\theta$ can be interpreted as a weight for the size of the "circle" of equal latitude (these get smaller as you go towards the poles).

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.)

It is not hard to see that in $\mathbb{R}^N/\mathbb{C}^N$, the standard inner product induces the ℓ_2 norm.

Properties of induced norms

In addition to the triangle inequality,

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

induced norms obey some very handy inequalities (note that these are not necessarily true for norms in general, only for norms induced by an inner product):

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

Proof.

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

2. **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**:

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

Proof. 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 $\boldsymbol{z} = \mathbf{0}$. If $\boldsymbol{z} = 0$, 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 $\alpha \in \mathbb{C}$, then

$$\boldsymbol{z} = \alpha \boldsymbol{y} - \frac{\alpha \langle \boldsymbol{y}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2} \boldsymbol{y} = \mathbf{0}.$$

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

Proof. As above, we have

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

and so

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

4. **Polarization Identity**

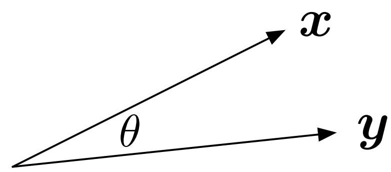
$$\mathrm{Re}\left\{\langle \boldsymbol{x}, \boldsymbol{y}\rangle\right\} = \frac{\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2}{4}$$

Proof. Using the expansions above, we quickly see that

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

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 $\mathcal{B}_Q = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ \|\boldsymbol{x}\|_Q = 1\}$.

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