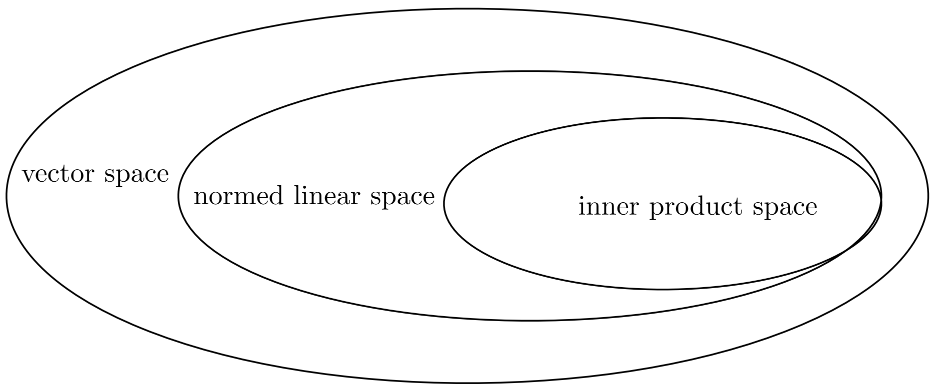
Moving to infinite dimensions

So far we have described the properties of general linear vector spaces, normed linear spaces, and inner product spaces. At each step we require that the space be equipped with functions (i.e. a norm and inner product) that impose ever more geometrical structure on the vector space.



As we can see from the properties provided for an inner product space (and the accompanying induced norm), 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. More precisely, if we have an infinite sequence of vectors in $\mathcal{S}$ that get closer and closer to one another, then they converge to something in $\mathcal{S}$. Even more precisely, we call a normed linear space complete if every Cauchy sequence is a convergent sequence; that is, for every sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in \mathcal{S}$ for which

$$\lim_{\min(m,n)\to\infty} \|\boldsymbol{x}_m - \boldsymbol{x}_n\| = 0, \quad \text{will also have} \quad \lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^* \in \mathcal{S},$$

where $\|\cdot\|$ is the norm induced by the inner product in the case of an inner product space.

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

All finite dimensional normed linear spaces and inner product spaces are complete, as is $L_2([a,b])$ and $L_2(\mathbb{R})$. The former is a basic result in mathematical analysis, the latter is a result from the modern theory of Lebesgue integration. In fact, every example of a normed linear space or inner product space we have encountered so far except for continuous, bounded functions is complete. 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.

A normed linear space which is also complete is called a **Banach space**. An inner product space which is also complete is called a **Hilbert space**. Note that every Hilbert space is automatically a Banach space when equipped with the induced norm.

The Wikipedia pages on these topics are actually pretty good:
http://en.wikipedia.org/wiki/Complete_metric_space
http://en.wikipedia.org/wiki/Banach_space
http://en.wikipedia.org/wiki/Hilbert_space

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 Banach (or 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 $\mathcal{S}$.

This allows us to offer a more general definition of a basis that is sensible when discussing infinite-dimensional vector spaces.

Definition: A **basis** of a Banach space $\mathcal{S}$ is a countable sequence of vectors $\mathcal{B} = \{\boldsymbol{v}_n\}$ such that for any $\boldsymbol{x} \in \mathcal{S}$, there exist unique scalars a_n such that

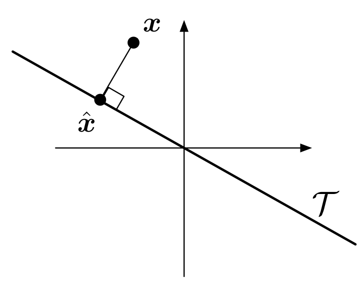
$$\boldsymbol{x} = \sum_n a_n \boldsymbol{v}_n.$$

Recall that the above equality is equivalent to $\|\boldsymbol{x} - \sum_n a_n \boldsymbol{v}_n\| = 0$, and hence for infinite-dimensional spaces we need to be in a Banach space to ensure that the sum converges and that these statements are even coherent.

Linear approximation in a Hilbert space

Hilbert spaces will play a fundamental role in this course. Perhaps the biggest reason is the following approximation problem:

Let $\mathcal{S}$ be a Hilbert space, and let $\mathcal{T}$ be a subspace of $\mathcal{S}$. Given an $\boldsymbol{x} \in \mathcal{S}$, what is the **closest point** $\hat{\boldsymbol{x}} \in \mathcal{T}$?



In other words, find $\hat{\boldsymbol{x}} \in \mathcal{T}$ that minimizes $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|$; i.e. given $\boldsymbol{x}$, we want to solve the following optimization program

$$\underset{\boldsymbol{y} \in \mathcal{T}}{\text{minimize}} \ \|\boldsymbol{x} - \boldsymbol{y}\|, \tag{1}$$

where the norm above is the one induced by the inner product. This problem has a unique solution which is characterized by what we will refer to as the **orthogonality principle**.

The orthogonality principle simply states that for the optimal approximation to $\boldsymbol{x}$ in $\mathcal{T}$ (i.e. the $\hat{\boldsymbol{x}}$ defined by (1)) the error $\hat{\boldsymbol{x}} - \boldsymbol{x}$ is orthogonal to $\mathcal{T}$. We state this more formally in the following theorem.

Theorem: Let $\mathcal{S}$ be a Hilbert space, and let $\mathcal{T}$ be a finite dimensional subspace[1]. Given an arbitrary $\boldsymbol{x} \in \mathcal{S}$,

1. there is exactly one $\hat{\boldsymbol{x}} \in \mathcal{T}$ such that
$$\boldsymbol{x} - \hat{\boldsymbol{x}} \;\perp\; \mathcal{T}, \tag{2}$$
meaning $\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{y} \rangle = 0$ for all $\boldsymbol{y} \in \mathcal{T}$, and
2. this $\hat{\boldsymbol{x}}$ is the closest point in $\mathcal{T}$ to $\boldsymbol{x}$; that is, $\hat{\boldsymbol{x}}$ is the unique minimizer to (1).

Proof: We will show that the first part is true in the next section of the notes, where we show how to explicitly calculate such an $\hat{\boldsymbol{x}}$.

For the second part, let $\hat{\boldsymbol{x}}$ be the vector which obeys
$$\hat{\boldsymbol{e}} = \boldsymbol{x} - \hat{\boldsymbol{x}} \;\perp\; \mathcal{T}.$$
Let $\boldsymbol{y}$ be any other vector in $\mathcal{T}$, and set
$$\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{y}.$$
We will show that
$$\|\boldsymbol{e}\| \;>\; \|\hat{\boldsymbol{e}}\| \quad (\text{i.e. that } \|\boldsymbol{x} - \boldsymbol{y}\| \;>\; \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|)\,.$$

Note that
$$\begin{aligned}\|\boldsymbol{e}\|^2 = \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= \|\hat{\boldsymbol{e}} - (\boldsymbol{y} - \hat{\boldsymbol{x}})\|^2 \\ &= \langle \hat{\boldsymbol{e}} - (\boldsymbol{y} - \hat{\boldsymbol{x}})\,,\; \hat{\boldsymbol{e}} - (\boldsymbol{y} - \hat{\boldsymbol{x}}) \rangle \\ &= \|\hat{\boldsymbol{e}}\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{x}}\|^2 - \langle \hat{\boldsymbol{e}}, \boldsymbol{y} - \hat{\boldsymbol{x}} \rangle - \langle \boldsymbol{y} - \hat{\boldsymbol{x}}, \hat{\boldsymbol{e}} \rangle.\end{aligned}$$

[1] The same results hold when $\mathcal{T}$ is infinite dimensional and is *closed*. We do not prove the infinite dimensional case just because it requires some analysis of infinite sequences which, while not really that difficult, kind of distract from the overall geometrical picture we are trying to paint here.

Since $\boldsymbol{y} - \hat{\boldsymbol{x}} \in \mathcal{T}$ and $\hat{\boldsymbol{e}} \perp \mathcal{T}$,

$$\langle \hat{\boldsymbol{e}}, \boldsymbol{y} - \hat{\boldsymbol{x}} \rangle = 0, \quad \text{and} \quad \langle \boldsymbol{y} - \hat{\boldsymbol{x}}, \hat{\boldsymbol{e}} \rangle = 0,$$

and so

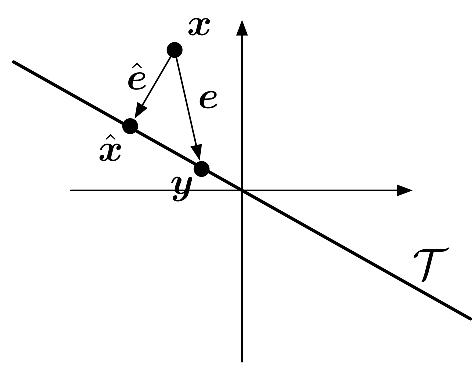
$$\|\boldsymbol{e}\|^2 = \|\hat{\boldsymbol{e}}\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{x}}\|^2.$$

Since all three quantities in the expression above are positive and

$$\|\boldsymbol{y} - \hat{\boldsymbol{x}}\| > 0 \quad \Leftrightarrow \quad \boldsymbol{y} \neq \hat{\boldsymbol{x}},$$

we see that

$$\boldsymbol{y} \neq \hat{\boldsymbol{x}} \quad \Leftrightarrow \quad \|\boldsymbol{e}\| > \|\hat{\boldsymbol{e}}\|.$$



Computing the best approximation

The orthogonality principle also gives us a concrete procedure for actually **computing** the optimal point $\hat{\boldsymbol{x}}$.

Let N be the dimension of $\mathcal{T}$, and let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ be a basis for $\mathcal{T}$. We want to find coefficients $a_1, \ldots, a_N \in \mathbb{C}$ such that

$$\hat{\boldsymbol{x}} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_N\boldsymbol{v}_N.$$

The orthogonality principle tells us that

$$\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{v}_n \rangle = 0 \quad \text{for} \quad n = 1, \ldots, N.$$

This means the a_n must obey

$$\langle \boldsymbol{x} - \sum_{k=1}^{N} a_k \boldsymbol{v}_k \ , \ \boldsymbol{v}_n \rangle = 0 \quad \text{for} \quad n = 1, \ldots, N,$$

or moving things around,

$$\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle = \sum_{k=1}^{N} a_k \langle \boldsymbol{v}_k, \boldsymbol{v}_n \rangle \quad \text{for} \quad n = 1, \ldots, N.$$

Since $\boldsymbol{x}$ and the $\{\boldsymbol{v}_n\}$ are given, we know both the $\langle \boldsymbol{x}, \boldsymbol{v}_n \rangle$ and the $\langle \boldsymbol{v}_k, \boldsymbol{v}_n \rangle$. We are left with a set of N **linear equations** with N unknowns:

$$\begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle & & \langle \boldsymbol{v}_N, \boldsymbol{v}_2 \rangle \\ \vdots & & \ddots & \vdots \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_N \rangle & \cdots & & \langle \boldsymbol{v}_N, \boldsymbol{v}_N \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{x}, \boldsymbol{v}_2 \rangle \\ \vdots \\ \langle \boldsymbol{x}, \boldsymbol{v}_N \rangle \end{bmatrix}$$

The matrix on the left hand side above is called the **Gram matrix** or **Grammian** of the basis $\{\boldsymbol{v}_n\}$.

In more compact notation, we want to find $\boldsymbol{a} \in \mathbb{C}^N$ such that

$$\boldsymbol{G}\boldsymbol{a} = \boldsymbol{b},$$

where $b_n = \langle \boldsymbol{x}, \boldsymbol{v}_n \rangle$ and $G_{k,n} = \langle \boldsymbol{v}_n, \boldsymbol{v}_k \rangle$.

Two notes on the structure of $\boldsymbol{G}$:

- $\boldsymbol{G}$ is guaranteed to be invertible because the $\{\boldsymbol{v}_n\}$ are linearly independent. We can comfortably write

$$\boldsymbol{a} = \boldsymbol{G}^{-1}\boldsymbol{b}.$$

- $\boldsymbol{G}$ is **conjugate symmetric** ("Hermitian"):

$$\boldsymbol{G} = \boldsymbol{G}^{\mathrm{H}},$$

where $\boldsymbol{G}^{\mathrm{H}}$ is the conjugate transpose of $\boldsymbol{G}$ (take the transpose, then take the complex conjugate of all the entries). This fact has algorithmic implications when it comes time to actually solve the system of equations.

Uniqueness

It should be clear that if $\langle \boldsymbol{e}, \boldsymbol{v}_k \rangle = 0$ for all of the basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$, then $\langle \boldsymbol{e}, \boldsymbol{y} \rangle = 0$ for all $\boldsymbol{y} \in \mathcal{T}$. The converse is also true: if $\langle \boldsymbol{e}, \boldsymbol{y} \rangle \neq 0$ for some $\boldsymbol{y} \in \mathcal{T}$ not equal to $\mathbf{0}$, then $\langle \boldsymbol{e}, \boldsymbol{v}_k \rangle \neq 0$ for at least one of the $\boldsymbol{v}_k$.

With the work above, this means that a necessary and sufficient condition for $\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{y} \rangle = 0$ for all $\boldsymbol{y} \in \mathcal{T}$ is to have

$$\hat{\boldsymbol{x}} = \sum_{n=1}^{N} a_n \boldsymbol{v}_n, \quad \text{where } \boldsymbol{a} \text{ satisfies} \quad \boldsymbol{G}\boldsymbol{a} = \boldsymbol{b}.$$

Since $\boldsymbol{G}$ is square and invertible, there is exactly one such $\boldsymbol{a}$, and hence exactly one $\hat{\boldsymbol{x}}$ that obeys the condition

$$\boldsymbol{x} - \hat{\boldsymbol{x}} \perp \mathcal{T}.$$

Example: Let

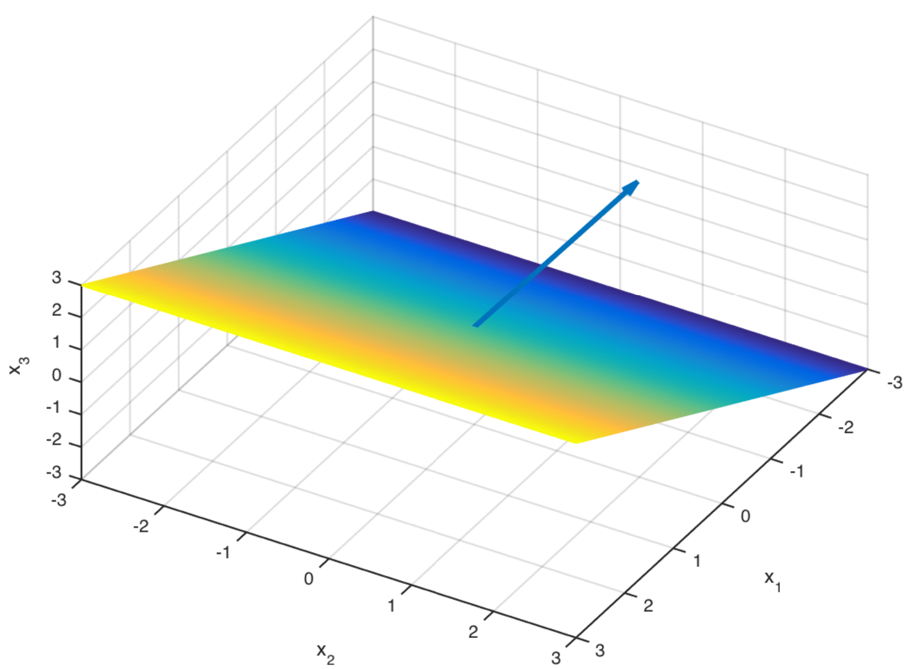
$$\mathcal{T} = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right), \qquad \boldsymbol{x} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}$$

Find the solution to

$$\underset{\boldsymbol{y} \in \mathcal{T}}{\text{minimize}} \ \|\boldsymbol{x} - \boldsymbol{y}\|_2.$$

(Recall that $\|\cdot\|_2$ in $\mathbb{R}^3$ is induced by the standard inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{n=1}^{3} x_n y_n$.)

Here is a plot of $\mathcal{T}$ and $\boldsymbol{x}$:



Solution: We have

$$\boldsymbol{G} = \qquad\qquad\qquad\qquad \boldsymbol{b} =$$

and so

$$\boldsymbol{G}^{-1} =$$

This means that

$$\boldsymbol{a} =$$

from which we synthesize the answer

$$\hat{\boldsymbol{x}} =$$

We can also check that "the error is orthogonal to the approximation"

$$\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \hat{\boldsymbol{x}} \rangle =$$