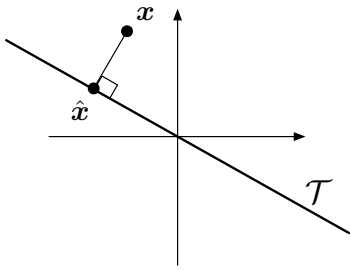
So far, we have introduced vector spaces as well as notions of a norm and inner product, which allowed us to measure distances and angles, respectively. As alluded to before, the point of working in inner product spaces (and with norms induced by inner products) is that natural notions in Euclidean spaces (e.g., that of a "closest point") extend to more general inner product spaces.

Linear approximation in a Hilbert space

Consider the following problem:

Let $\mathcal{S}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$, and let $\mathcal{T}$ be a subspace of $\mathcal{S}$.

Given a $\boldsymbol{x} \in \mathcal{S}$, what is the **closest point** $\hat{\boldsymbol{x}} \in \mathcal{T}$?



In other words, find $\hat{\boldsymbol{x}} \in \mathcal{T}$ that minimizes $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|$; 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 the **orthogonality principle**.

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 exists exactly one point $\hat{\boldsymbol{x}} \in \mathcal{T}$ that obeys

$$\boldsymbol{x} - \hat{\boldsymbol{x}} \;\perp\; \mathcal{T}, \tag{2}$$

   meaning $\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{y} \rangle = 0$ for all $\boldsymbol{y} \in \mathcal{T}$, and
2. this point $\hat{\boldsymbol{x}}$ is the (unique) minimizer of (1).

We will start by showing that any vector $\hat{\boldsymbol{x}}$ that satisfies (2) must also be the unique minimizer of (1). We will then give a constructive technique for finding such a point, and show that it must be unique.

Let $\hat{\boldsymbol{x}}$ be a vector in $\mathcal{T}$ that obeys

$$\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{y} \rangle = 0, \quad \text{for all } \boldsymbol{y} \in \mathcal{T}, \tag{3}$$

and let $\boldsymbol{y}$ any vector in $\mathcal{T}$. Then (squared) distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ can be written as

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= \|(\boldsymbol{x} - \hat{\boldsymbol{x}}) - (\boldsymbol{y} - \hat{\boldsymbol{x}})\|^2 \\ &= \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{x}}\|^2, \end{aligned}$$

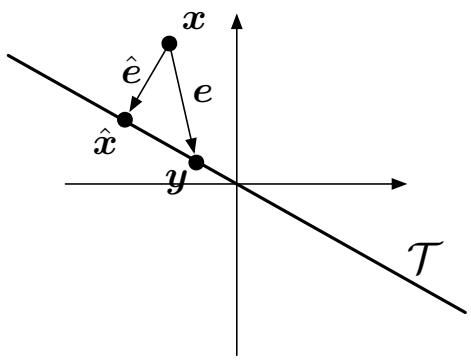
where the last equality comes from combing the fact that

$$\boldsymbol{y}, \hat{\boldsymbol{x}} \in \mathcal{T} \quad \Rightarrow \quad \boldsymbol{y} - \hat{\boldsymbol{x}} \in \mathcal{T},$$

with (3) and the Pythagorean Theorem. Thus

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 \;>\; \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 \quad \text{for all} \quad \boldsymbol{y} \neq \hat{\boldsymbol{x}}.$$

[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, distracts from the overall geometrical picture presented here.



Proving uniqueness

Instead of presenting a uniqueness proof by contradiction (which would similar to what we did in the previous lecture), we will proceed to a more useful result: a computational procedure for evaluating $\hat{\boldsymbol{x}}$ that will automatically show uniqueness (think about why as you read).

Computing the best approximation

The orthogonality principle 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 $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ such that

$$\hat{\boldsymbol{x}} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_N \boldsymbol{v}_N.$$

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 α_n must obey

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

or moving things around,

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

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:

$$\begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle & \cdots & \langle \boldsymbol{v}_N, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle & & \langle \boldsymbol{v}_N, \boldsymbol{v}_2 \rangle \\ \vdots & & \ddots & \vdots \\ \langle \boldsymbol{v}_1, \boldsymbol{v}_N \rangle & \cdots & & \langle \boldsymbol{v}_N, \boldsymbol{v}_N \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \\ \langle \boldsymbol{x}, \boldsymbol{v}_2 \rangle \\ \vdots \\ \langle \boldsymbol{x}, \boldsymbol{v}_N \rangle \end{bmatrix}$$

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 $\boldsymbol{\alpha} \in \mathbb{R}^N$ such that

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

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

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

- $\boldsymbol{G}$ is **symmetric** ("Hermitian"[2]):

$$\boldsymbol{G} = \boldsymbol{G}^{\mathrm{T}}.$$

  This fact has algorithmic implications when it comes time to actually solve the system of equations.

Existence and Uniqueness

The method above shows us how to construct an $\hat{\boldsymbol{x}}$ such that (2) holds. This construction centered on solving a system of equations $\boldsymbol{G}\boldsymbol{\alpha} = \boldsymbol{b}$; since this system is invertible, this system always has a unique solution. This likewise maps to a unique $\hat{\boldsymbol{x}}$, since the $\boldsymbol{v}_n$ are a basis for $\mathcal{T}$. Finally, our first argument showed that (2) holding means that $\hat{\boldsymbol{x}}$ is the solution to (1).

[2]When dealing with complex numbers, Hermitian matrices are conjugate symmetric matrices. But as we said before, this will not be a detour that we take in this class.

Example: Let

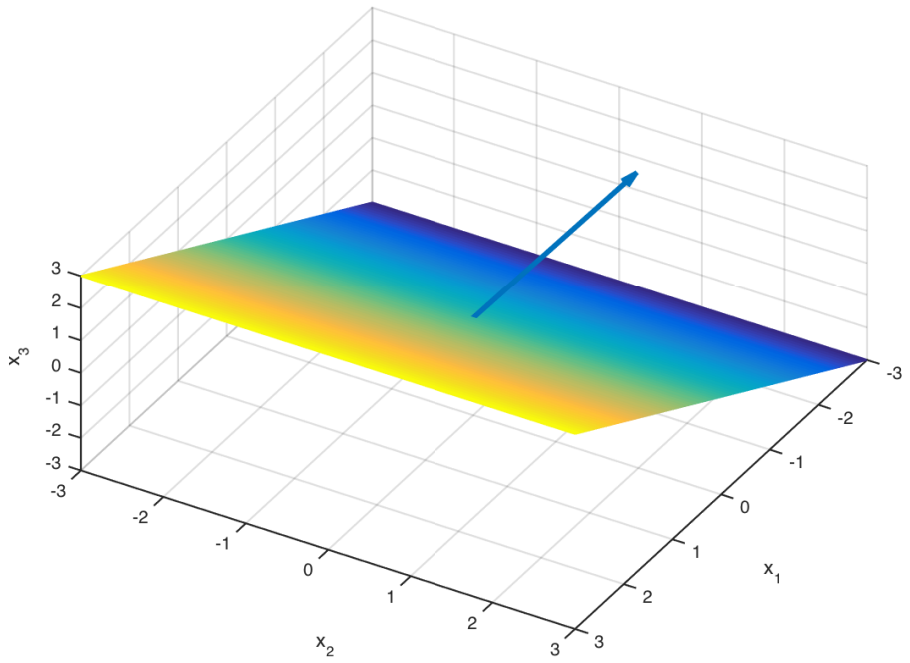
$$\mathcal{T} = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right), \qquad \boldsymbol{x} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}$$

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

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

and so

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

This means that

$$\boldsymbol{\alpha} =$$

from which we synthesize the answer

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

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

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

Example: Polynomial approximation of e^t

1. We want to calculate a quadratic approximation of $x(t) = e^t$ over the interval $[0, 1]$.

   (a) Sketch $x(t)$ on $[0, 1]$.

   (b) One way to do the approximation is to truncate the Taylor expansion

   $$e^t = 1 + t + \frac{t^2}{2} + \frac{t^3}{6} + \frac{t^4}{24} + \cdots$$

   In this case the approximation is

   $$\tilde{x}_{\text{taylor}}(t) = 1 + t + t^2/2.$$

   Sketch $\tilde{x}_{\text{taylor}}(t)$.

   (c) Is the truncated Taylor approximation the best possible quadratic approximation? The answer is no if we are interested in minimizing the energy of the error. We want to find the second order polynomial

   $$\tilde{x}(t) = a_1 + a_2 t + a_3 t^2$$

   that minimizes

   $$\|e^t - \tilde{\boldsymbol{x}}\|_{L_2([0,1])} = \sqrt{\int_0^1 |e^t - \tilde{x}(t)|^2 \ dt}.$$

   We set this up as a subspace approximation problem. Set

   $$v_1(t) = 1, \quad v_2(t) = t, \quad v_3(t) = t^2,$$

   and set $\mathcal{T} = \text{Span}\left(\{v_1, v_2, v_3\}\right)$. We of course now know a systematic way to find the best $\tilde{x} \in \mathcal{T}$. Start by calculating the Gram matrix $\boldsymbol{G}$ (recall that $G_{ij} = \langle v_i, v_j \rangle$).

   (d) Calculate the right-hand-side $\boldsymbol{b}$ (recall that $b_i = \langle x, v_i \rangle$). Write down the system of equations that need to be solved. You can use MATLAB to solve this system.

   (e) Calculate the error for both the Taylor approximation and the optimal approximation computed above. Plot $x(t), \tilde{x}_{\text{taylor}}(t)$, and $\tilde{x}(t)$ on the same set of axes.

2. We consider the same problem, but calculate the error in a different way. The norm we use to measure the error is

$$\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_S^2 = \int_0^1 w(t)|x(t) - \tilde{x}(t)|^2 \, dt$$

where

$$w(t) = 16(t - 1/2)^2.$$

(a) Plot $w(t)$ and argue that measuring the error in this way will penalize mismatch at the ends of the interval than in the middle.

(b) Write down the inner product that induces $\|\cdot\|_S$.

(c) Find the second order polynomial that is the best approximation to e^t in the $\|\cdot\|_S$ norm.