Linear Algebra II

Math 232

instructor:

John Terilla

Orthogonal Projections and the Minimization problem

May 8, 2020

In a normed vector space $(V, \|\cdot\|)$ one often faces the following problem:

Problem: Given a subspace $W \subseteq V$ and a vector $v \in V \setminus W$, find the vector in W that is closest to v. That is, find $\underset{w \in W}{\operatorname{argmin}} \|v - w\|$.

This problem shows up in many places. For example, in approximation of functions or in

fitting data to a model.

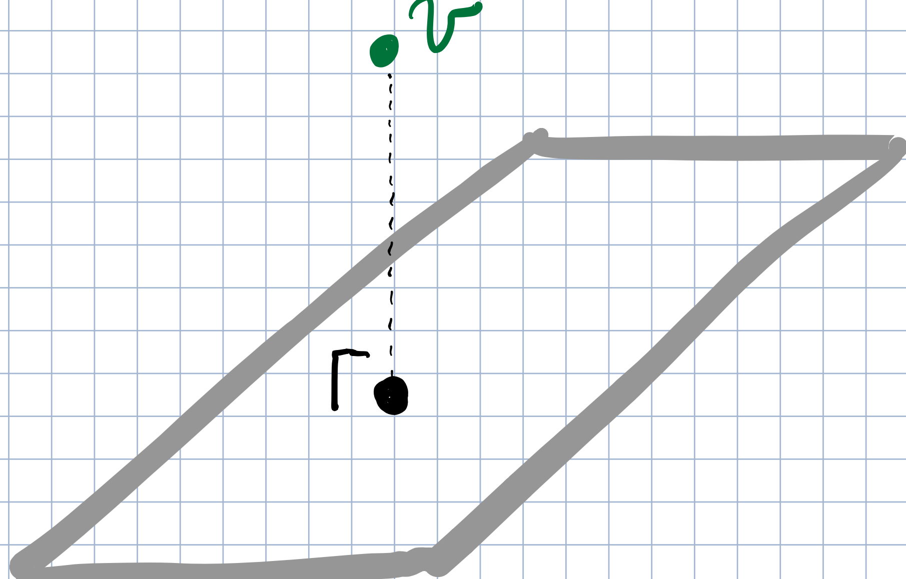
Imagine you'd like to find a polynomial $p \in \mathcal{P}_5(\mathbb{R})$ that is closest to $\cos(t)$. First, you need to make "closest" a precise concept, say by equipping the space $V = \{\text{continuous functions } [-1,1] \to \mathbb{R}\}$ with a norm.

Then, you're trying to find the vector $p \in \mathcal{P}_5(\mathbb{R}) \subseteq V$ that minimizes $\|\cos t - p\|$.

Now, in general this minimization problem is <u>hard</u>.

However, if $\|\cdot\|$ comes from an inner product, the problem is <u>easy</u>. The reason is that you have <u>orthogonal</u> <u>projection</u>!

Here's a picture:



I'm going to present this a little differently than Axler since we'll be interested in the case when we have an "almost inner product" defined on V that restricts to an inner product on a finite dimensional subspace W.

For example, if $V = \{\text{continuous functions } [0,10] \to \mathbb{R}\}$ and

$$\langle f, g \rangle := f(0)g(0) + f(2)g(2) + f(4)g(4) + f(6)g(6) + f(8)g(8) + f(10)g(10)$$

then $\langle , \rangle$ isn't quite an inner product since it's possible that $\langle f, f \rangle = 0$ even though $f \neq 0$.

For example

$$f(x) = \frac{x(x-2)(x-4)(x-6)(x-8)(x-10)\sin(x)}{e^{5x^2+1} + \sqrt{x^2+5x^4+1}}.$$

Nonetheless, if we define $\|f\|$ by $\|f\| = \sqrt{\langle f, f \rangle}$, then we have

$$\|f\| \geq 0$$
$$\|\alpha f\| = |\alpha| \, \|f\|$$
$$\|f+g\| \leq \|f\| + \|g\|$$

it just may happen that

$\|f\| = 0$ for nonzero $f \in V$.

Moreover, if we define distance by $d(f,g)$ by

$d(f,g) := \|f-g\|$ then we have

$$d(f,g) \geq 0$$
$$d(f,g) = d(g,f)$$
$$d(f,g) \leq d(f,h) + d(h,g)$$

it just may happen that $d(f,g) = 0$ when $f \neq g$.

Notice that $\|f\| = \sqrt{\langle f,f \rangle}$

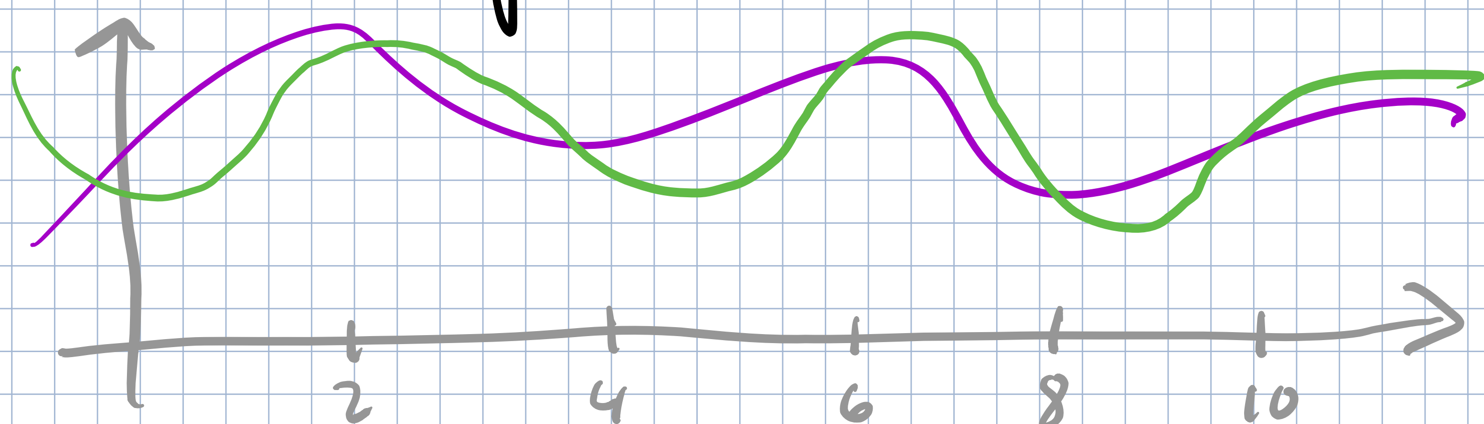
where

$$\langle f,g\rangle := f(0)g(0)+f(2)g(2)+f(4)g(4)$$
$$+f(6)g(6)+f(8)g(8)+f(10)g(10)$$

captures a useful idea of distance:

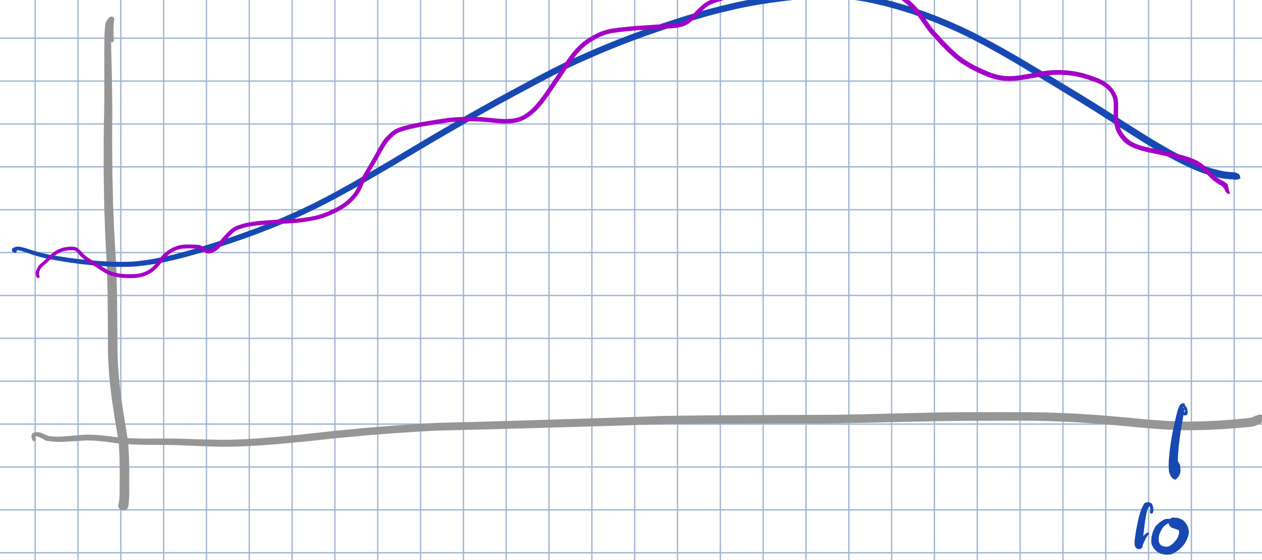
$d(f,g)$ is small $\Leftrightarrow \langle f-g, f-g\rangle$ is small $\Leftrightarrow (f(0)-g(0))^2 + (f(2)-g(2))^2 + \dots + (f(10)-g(10))^2$ is small $\Leftrightarrow f(x) \approx g(x)$ when $x = 0, 2, 4, 6, 8, 10$.

Here's a picture:

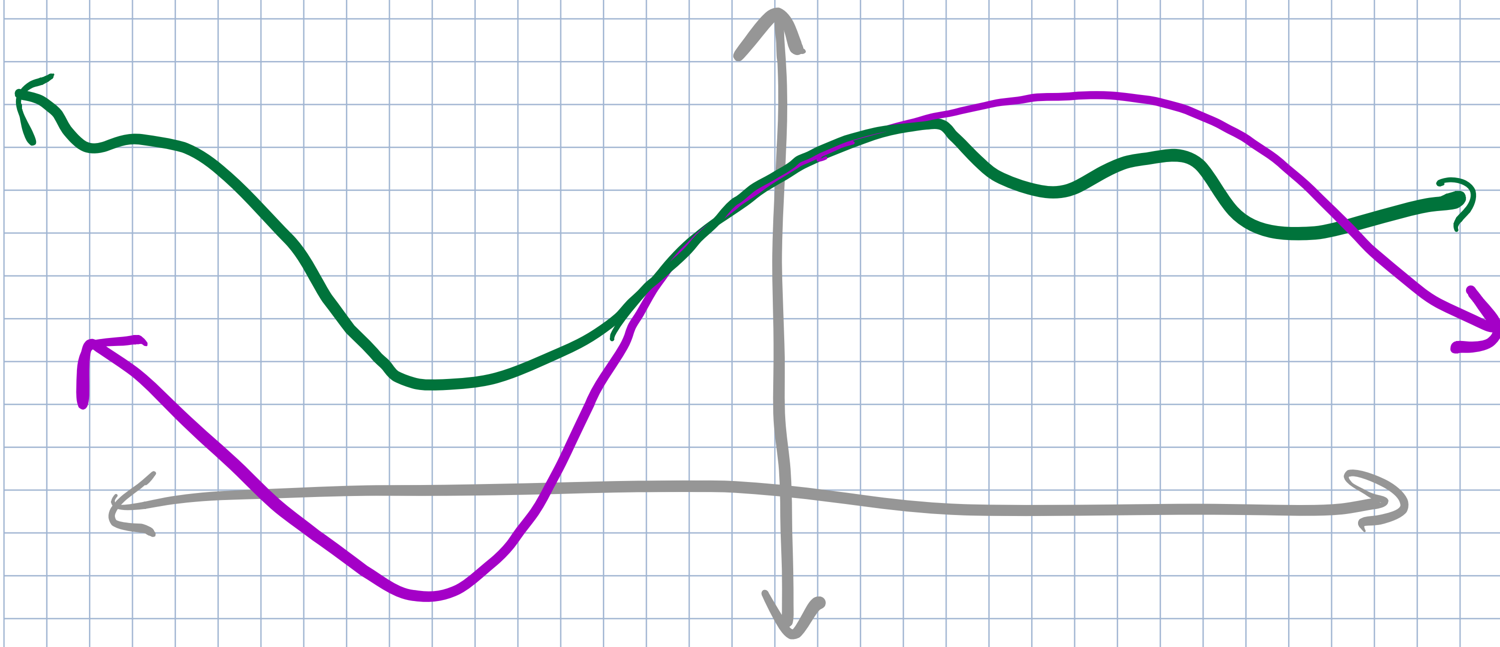


Other choices for "almost" inner products on V that capture reasonable notions of distance are

$$\langle f,g\rangle = \int_0^{10} f(x)g(x)\,dx$$



Also, $\langle f,g\rangle = f(0)g(0) + f'(0)g'(0) + f''(0)g''(0) + f'''(0)g'''(0)$



Etc...

In all these cases, what we have is a symmetric, nonnegative bilinear pairing $\langle , \rangle$ on a real vector space V, meaning a bilinear function $\langle , \rangle : V \times V \to \mathbb{R}$ satisfying

$$\langle v, w\rangle = \langle w, v\rangle$$
$$\langle v+v', w\rangle = \langle v, w\rangle + \langle v', w\rangle$$
$$\langle \alpha v, w\rangle = \alpha\langle v, w\rangle$$
$$\langle v, v\rangle \geq 0$$

Not necessarily assuming $\langle v, v\rangle = 0 \Rightarrow v = 0$

If $W \leq V$ is a finite dimensional subspace on which $\langle\,,\rangle$ is nondegenerate meaning $w \in W$ and $\langle w, w\rangle = 0 \Rightarrow w = 0$, then we have an orthogonal projection operator $P: V \to W$

that solves the minimization problem. That is, for any $v \in V$, we have

$$\|v - Pv\| \leq \|v - w\|$$

for any $w \in W$.

P has a simple formula:

$$Pv := \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \cdots + \langle v, e_n \rangle e_n$$

where $e_1, e_2, \ldots, e_n$ is any orthonormal basis for W.

What follows is a proof that

- the definition of $P: V \to W$ does not depend on the choice of basis $e_1, e_2, \ldots, e_n$ of W

- P_V solves the minimization problem.

To get started let $B = e_1, \ldots, e_n$ be an orthonormal basis for W and let

$$P_B : V \to W$$
$$v \longmapsto \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

Note that $P_B|_W = \mathrm{id}_W$ and so

$$P_B^2 = P_B.$$

Therefore $V = W \oplus \mathrm{null}(P_{\mathcal{B}})$

Since every vector $v \in V$ can be written $v = P_{\mathcal{B}} v + v - P_{\mathcal{B}} v$ and $W \cap \mathrm{null}(P_{\mathcal{B}}) = \{0\}$.

Check: $P_{\mathcal{B}}(v - P_{\mathcal{B}} v) = P_{\mathcal{B}} v - P_{\mathcal{B}}^2 v$

$= P_{\mathcal{B}} v - P_{\mathcal{B}} v = 0$ So $v - P_{\mathcal{B}} v \in \mathrm{null}(P_{\mathcal{B}})$. Also, $w \in W \cap \mathrm{null}\, P_{\mathcal{B}}$

$\Rightarrow P_{\mathcal{B}} w = 0 \Rightarrow \mathrm{id}_{\mathcal{B}} w = 0 \Rightarrow w = 0$

So $W \cap \mathrm{null}\, P_{\mathcal{B}} = \{0\}$.

Define $W^\perp = \{v \in V : \langle v, w \rangle = 0$ for all $w \in W\}$.

Claim: $W^\perp = \operatorname{null} P_B$.

Proof: Suppose $v \in W^\perp$ and notice that $P_B v = \sum_{i=1}^{n} \langle v, e_i \rangle e_i = 0$ since each $\langle v, e_i \rangle = 0$.

On the other hand, suppose $v \in \operatorname{null}(P_B)$. Then $Pv = 0 \Rightarrow$ $0 = \langle Pv, Pv \rangle = \sum_{i=1}^{n} \langle v, e_i \rangle^2 \Rightarrow$ $\langle v, e_i \rangle = 0 \ \forall i \Rightarrow v \in W^\perp$. ∎

Conclusion: $P_{\mathcal{B}}$ doesn't depend on the basis $\mathcal{B}$. In fact, $P_{\mathcal{B}}$ is simply the map

$$W \oplus W^{\perp} \xrightarrow{P} W.$$

Notice, we have the Pythagorean theorem: If $\langle v, w \rangle = 0$ then $\|v+w\|^2 = \|v\|^2 + \|w\|^2$ since

$$\|v+w\|^2 = \langle v+w, v+w \rangle$$
$$= \langle v,v \rangle + \langle v,w \rangle + \langle w,v \rangle + \langle w,w \rangle$$
$$= \|v\|^2 + \|w\|^2.$$

Let's return to $P: V \to W$.

Theorem: For all $v \in V$, $w \in W$
$$\|v - Pv\| \leq \|v - w\|.$$

Proof: Let $v \in V$ and $w \in W$.

$$\|v - Pv\|^2 \leq \|v - Pv\|^2 + \|Pv - w\|^2$$
$$= \|v - Pv + Pv - w\|^2 = \|v - w\|^2$$

This equality follows from $v - Pv \in W^\perp$ and $Pv - w \in W$.

$\blacksquare$