Chapter 3

Hilbert spaces

Contents

Key missing geometrical concepts: **angle** and **orthogonality** (“right angles”).

3.1

Introduction

We now turn to the subset of normed spaces called **Hilbert spaces**, which must have an **inner product**. These are particularly useful spaces in applications/analysis.

Why not introduce Hilbert first then? For generality: it is helpful to see which properties are general to vector spaces, or to normed spaces, vs which require additional assumptions like an inner product.

Overview

- inner product
- orthogonality
- orthogonal projections
- applications
  - least-squares minimization
  - orthonormalization of a basis
  - Fourier series

General forms of things you have seen before: Cauchy-Schwarz, Gram-Schmidt, Parseval’s theorem

3.2

Inner products

Definition. A **pre-Hilbert space**, aka an **inner product space** , is a vector space $\mathcal{X}$ defined on the field $\mathcal{F} = \mathbb{R}$ or $\mathcal{F} = \mathbb{C}$, along with an **inner product** operation $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X} \to \mathcal{F}$, which must satisfy the following axioms $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X},\ \alpha \in \mathcal{F}$.

1. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle^*$ (**Hermitian symmetry**), where * denotes complex conjugate.
2. $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ (**additivity**)
3. $\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ (**scaling**)
4. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ and $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ iff $\boldsymbol{x} = \boldsymbol{0}$. (**positive definite**)

Properties of inner products

Bilinearity property:

$$\left\langle \sum_i \alpha_i \boldsymbol{x}_i,\ \sum_j \beta_j \boldsymbol{y}_j \right\rangle = \sum_i \sum_j \alpha_i \beta_j^* \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle.$$

Lemma. In an inner product space, if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ for all $\boldsymbol{y}$, then $\boldsymbol{x} = \boldsymbol{0}$.
Proof. Let $\boldsymbol{y} = \boldsymbol{x}$. □

Cauchy-Schwarz inequality

Lemma. For all $\boldsymbol{x}$,$\boldsymbol{y}$ in an inner product space,

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} \sqrt{\langle \boldsymbol{y}, \boldsymbol{y} \rangle} = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \ \textit{\textbf{(see induced norm below)}},$$

with equality iff $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent.
Proof. For any $\lambda \in \mathcal{F}$ the positive definiteness of $\langle \cdot, \cdot \rangle$ ensures that

$$0 \leq \langle \boldsymbol{x} - \lambda \boldsymbol{y}, \boldsymbol{x} - \lambda \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle - \lambda \langle \boldsymbol{y}, \boldsymbol{x} \rangle - \lambda^* \langle \boldsymbol{x}, \boldsymbol{y} \rangle + |\lambda|^2 \langle \boldsymbol{y}, \boldsymbol{y} \rangle.$$

If $\boldsymbol{y} = \boldsymbol{0}$, the inequality holds trivially. Otherwise, consider $\lambda = \langle \boldsymbol{x}, \boldsymbol{y} \rangle / \langle \boldsymbol{y}, \boldsymbol{y} \rangle$ and we have

$$0 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle - |\langle \boldsymbol{y}, \boldsymbol{x} \rangle|^2 / \langle \boldsymbol{y}, \boldsymbol{y} \rangle.$$

Rearranging yields $|\langle \boldsymbol{y}, \boldsymbol{x} \rangle| \leq \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle} = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|$.
The proof about equality conditions is *Problem 3.1*. □
This result generalizes all the “Cauchy-Schwarz inequalities” you have seen in previous classes, *e.g.*, vectors in $\mathbb{R}^n$, random variables, discrete-time and continuous-time signals, each of which corresponds to a particular inner product space.

Angle

Thanks to this inequality, we can generalize the notion of the **angle** between vectors to any general inner product space as follows:

$$\theta = \cos^{-1}\left(\frac{|\langle \boldsymbol{x},\, \boldsymbol{y} \rangle|}{\|\boldsymbol{x}\|\, \|\boldsymbol{y}\|}\right), \qquad \forall \boldsymbol{x}, \boldsymbol{y} \neq \boldsymbol{0}.$$

This definition is legitimate since the argument of $\cos^{-1}$ will always be between 0 and 1 due to the Cauchy-Schwarz inequality.

Induced norm

Proposition. In an inner product space $(\mathcal{X}, \langle \cdot,\, \cdot \rangle)$, the **induced norm** $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x},\, \boldsymbol{x} \rangle}$ is indeed a norm.

Proof. What must we show?

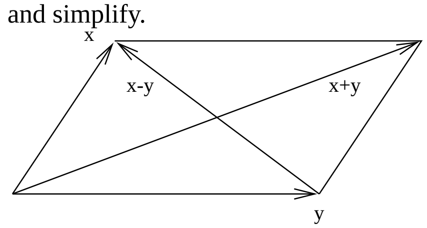
- The first axiom ensures that $\langle \boldsymbol{x},\, \boldsymbol{x} \rangle$ is real.
- $\|\boldsymbol{x}\| \geq 0$ with equality iff $\boldsymbol{x} = \boldsymbol{0}$ follows from Axiom 4.
- $\|\alpha \boldsymbol{x}\| = \sqrt{\langle \alpha \boldsymbol{x},\, \alpha \boldsymbol{x} \rangle} = \sqrt{\alpha \langle \boldsymbol{x},\, \alpha \boldsymbol{x} \rangle} = \sqrt{\alpha \langle \alpha \boldsymbol{x},\, \boldsymbol{x} \rangle^*} = \sqrt{\alpha \alpha^* \langle \boldsymbol{x},\, \boldsymbol{x} \rangle^*} = |\alpha| \sqrt{\langle \boldsymbol{x},\, \boldsymbol{x} \rangle} = |\alpha|\, \|\boldsymbol{x}\|$, using Axioms 1 and 3.
- The only condition remaining to be verified is the triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \langle \boldsymbol{x},\, \boldsymbol{x} \rangle + \langle \boldsymbol{x},\, \boldsymbol{y} \rangle + \langle \boldsymbol{y},\, \boldsymbol{x} \rangle + \langle \boldsymbol{y},\, \boldsymbol{y} \rangle$ $= \|\boldsymbol{x}\|^2 + 2\,\mathrm{real}(\langle \boldsymbol{x},\, \boldsymbol{y} \rangle) + \|\boldsymbol{y}\|^2 \leq \|\boldsymbol{x}\|^2 + 2\,|\langle \boldsymbol{x},\, \boldsymbol{y} \rangle| + \|\boldsymbol{y}\|^2 \leq \|\boldsymbol{x}\|^2 + 2\,\|\boldsymbol{x}\|\, \|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 = (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2$. □
  (Recall if $z = a + \imath b$, then $a = \mathrm{real}(z) \leq \sqrt{a^2 + b^2} = |z|$.)

Any inner product space is necessarily a normed space. Is the reverse true? Not in general.
The following property distinguishes inner product spaces from mere normed spaces.

Lemma. (The **parallelogram law**.) In an inner product space:

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 + \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2\,\|\boldsymbol{x}\|^2 + 2\,\|\boldsymbol{y}\|^2, \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}. \tag{3-1}$$

Proof. Expand the norms into inner products and simplify. □



Remarkably, the converse of this Lemma also holds (see, *e.g.*, problem [2, p. 175]).

Proposition. If $(\mathcal{X}, \|\cdot\|)$ is a normed space over $\mathbb{C}$ or $\mathbb{R}$, *and* its norm satisfies the parallelogram law (3-1), then $\mathcal{X}$ is also an inner product space, with inner product:

$$\langle \boldsymbol{x},\, \boldsymbol{y} \rangle = \frac{1}{4}\left(\|\boldsymbol{x} + \boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2 + i\,\|\boldsymbol{x} + i\boldsymbol{y}\|^2 - i\,\|\boldsymbol{x} - i\boldsymbol{y}\|^2\right).$$

Proof. homework challenge problem.

Continuity of inner products

Lemma. In an inner product space $(\mathcal{X}, \langle \cdot,\, \cdot \rangle)$, if $\boldsymbol{x}_n \to \boldsymbol{x}$ and $\boldsymbol{y}_n \to \boldsymbol{y}$, then $\langle \boldsymbol{x}_n,\, \boldsymbol{y}_n \rangle \to \langle \boldsymbol{x},\, \boldsymbol{y} \rangle$.
Proof.

$$\begin{aligned}
|\langle \boldsymbol{x}_n,\, \boldsymbol{y}_n \rangle - \langle \boldsymbol{x},\, \boldsymbol{y} \rangle| &= |\langle \boldsymbol{x}_n,\, \boldsymbol{y}_n \rangle - \langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle + \langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle - \langle \boldsymbol{x},\, \boldsymbol{y} \rangle| \leq |\langle \boldsymbol{x}_n,\, \boldsymbol{y}_n \rangle - \langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle| + |\langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle - \langle \boldsymbol{x},\, \boldsymbol{y} \rangle| \\
&= |\langle \boldsymbol{x}_n - \boldsymbol{x},\, \boldsymbol{y}_n \rangle| + |\langle \boldsymbol{x},\, \boldsymbol{y}_n - \boldsymbol{y} \rangle| \\
&\leq \|\boldsymbol{x}_n - \boldsymbol{x}\|\, \|\boldsymbol{y}_n\| + \|\boldsymbol{x}\|\, \|\boldsymbol{y}_n - \boldsymbol{y}\| \text{ by Cauchy-Schwarz} \\
&\leq \|\boldsymbol{x}_n - \boldsymbol{x}\|\, M + \|\boldsymbol{x}\|\, \|\boldsymbol{y}_n - \boldsymbol{y}\| \text{ since } \boldsymbol{y}_n \text{ is convergent and hence bounded} \\
&\quad \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Thus $\langle \boldsymbol{x}_n,\, \boldsymbol{y}_n \rangle \to \langle \boldsymbol{x},\, \boldsymbol{y} \rangle$. □

Examples

Many of the normed spaces we considered previously are actually induced by suitable inner product space.

Example. In Euclidean space, the usual inner product (aka "**dot product**") is

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} a_i b_i, \text{ where } \boldsymbol{x} = (a_1, \ldots, a_n) \text{ and } \boldsymbol{y} = (b_1, \ldots, b_n).$$

Verifying the axioms is trivial. The induced norm is the usual ℓ_2 norm.

Example. For the space ℓ_2 over the complex field, the usual inner product is[1] $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_i a_i b_i^*$.
The Hölder inequality, which is equivalent to the Cauchy-Schwarz inequality for this space, ensures that $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2$.
So the inner product is indeed finite for $\boldsymbol{x}, \boldsymbol{y} \in \ell_2$. Thus ℓ_2 is not only a Banach space, it is also an inner product space.

Example. What about ℓ_p for $p \neq 2$? Do suitable inner products exist?
Consider $\mathcal{X} = \left(\mathbb{R}^2, \|\cdot\|_p \cdot\right)$ with $\boldsymbol{x} = (1, 0)$ and $\boldsymbol{y} = (0, 1)$.
The parallelogram law holds (for this $\boldsymbol{x}$ and $\boldsymbol{y}$) iff $2(1+1)^{2/p} = 2 \cdot 1^2 + 2 \cdot 1^2$, *i.e.*, iff $2^{2/p} = 2$.
Thus ℓ_2 is only inner product space in the ℓ_p family of normed spaces.

Example. The space of measurable functions on $[a, b]$ with inner product

$$\langle f, g \rangle = \int_a^b w(t) f(t) g^*(t) \, \mathrm{d}t,$$

where $w(t) > 0, \ \forall t$ is some (real) weighting function. Choosing $w = 1$ yields $\mathcal{L}_2[a, b]$.

Hilbert space

Definition. A *complete* inner product space is called a **Hilbert space**.

In other words, a Hilbert space is a Banach space along with an inner product that induces its norm. The addition of the inner product opens many analytical doors, as we shall see.

The concept "complete" is appropriate here since any inner product space is a normed space.

All of the preceding examples of inner product spaces were complete vector spaces (under the induced norm).

Example. The following is an inner product space, but *not* a Hilbert space, since it is incomplete:

$$R_2[a, b] = \left\{ f : [a, b] \to \mathbb{R} : \text{Riemann integral } \int_a^b f^2(t) \, \mathrm{d}t < \infty \right\},$$

with inner product (easily verified to satisfy the axioms): $\langle f, g \rangle = \int_a^b f(t) g(t) \, \mathrm{d}t$.

[1] Note that the conjugate goes with the second argument become of Axiom 3. I have heard that some treatments scale the second argument in Axiom 3, which affects where the conjugates go in the inner products.

Minimum norm optimization problems

Section *3.3* is called "the projection theorem" and it is about a certain type of minimum norm problem. Before focusing on that specific minimum norm problem, we consider the broad family of such problems.

Consider the following problem, which arises in many applications such as in approximation problems:
Given $\boldsymbol{x}$ in a normed space $(\mathcal{X}, \|\cdot\|)$, and a subset S in $\mathcal{X}$, find "the" vector $\boldsymbol{s} \in S$ that minimizes $\|\boldsymbol{x} - \boldsymbol{s}\|$.

Example. Control subject to energy constraint. See Section 3.11.

Example. Least-squares estimation: $\min_{\boldsymbol{\theta}} \|\boldsymbol{y} - \sum_{i=1}^n \theta_i \boldsymbol{x}_i\|$ is equivalent to $\min_{\boldsymbol{m} \in [\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}]} \|\boldsymbol{y} - \boldsymbol{m}\|$.

What questions should we ask about such problems?

- Is there any best $\boldsymbol{s}$? I.e., does there exist $\boldsymbol{s}^\star \in S$ s.t. $\|\boldsymbol{x} - \boldsymbol{s}^\star\| = d(\boldsymbol{x}, S)$?
  What answers do we have so far for this question? **??**
- If so, is $\boldsymbol{s}^\star$ unique? Answers thus far? **??**
- How is $\boldsymbol{s}^\star$ characterized? (Better yet would be an explicit formula for $\boldsymbol{s}^\star$.)

Chebyshev sets

One way to address such questions is "answer by definition."

Definition. A set S in a normed space $(\mathcal{X}, \|\cdot\|)$ is called **proximinal** [16] iff $\forall \boldsymbol{x} \in \mathcal{X}$, there *exists at least one point* $\boldsymbol{s} \in S$ s.t. $\|\boldsymbol{x} - \boldsymbol{s}\| = d(\boldsymbol{x}, S)$.

Definition. In a normed space, a set S is called a **Chebyshev set** iff $\forall \boldsymbol{x} \in \mathcal{X}$, there exists a *unique* $\boldsymbol{s} \in S$ s.t. $\|\boldsymbol{x} - \boldsymbol{s}\| = d(\boldsymbol{x}, S)$.

Fact. Any proximinal set is closed. (The points in $\overline{S} - S$ do not have a closest point in S.)
Fact. Any Chebyshev set is a proximinal set.
Fact. Any compact set is a proximinal set (due to Weierstrass theorem).

Note that we have not said anything about inner products here, so why not study minimum norm problems in detail in Banach spaces? The answer is due to one of the most famous unsolved problems in functional analysis: characterization of Chebyshev sets in general Banach spaces and in infinite-dimensional Hilbert spaces. What is known includes the following.
(See Deutsch paper [17], a scanned version of which is available on the course web page.)

- inner product space
  - In finite-dimensional Hilbert spaces, any Chebyshev set is closed, convex, and nonempty.
  - "Conversely," in any inner product space, any complete and convex set is Chebyshev. (We will prove this later in *3.12*).
- normed space
  - Are *all* Chebyshev sets convex? In general: no. A nonconvex Chebyshev set (in an incomplete infinite-dimensional normed space within the space of finite-length sequences) is given in [18].
  - In [3, p. 285], an example is given in a Banach space of a closed (and thus complete) subspace (hence convex) that is not a Chebyshev set.

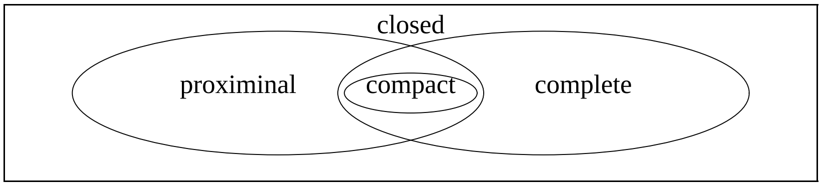
There is continued effort to characterize Chebyshev sets, *e.g.*, [19–21].

Since the characterization of Chebyshev sets is unsolved in normed spaces, we focus primarily on closed convex sets in inner product spaces hereafter.

However, this fact is encouraging. If S is a nonempty closed subset of $\mathbb{R}^n$, then $\left\{\boldsymbol{x} \in \mathbb{R}^n : \arg\min_{\boldsymbol{y} \in S} \|\boldsymbol{x} - \boldsymbol{y}\| \text{ is nonunique}\right\}$ has Lebesgue measure zero [16, p. 493].

Are all complete and bounded sets also proximinal?
The answer is yes in finite-dimensional normed spaces since there all closed and bounded sets are compact.
But the answer is "not necessarily" in infinite dimensional normed spaces, even in a Hilbert space in fact.

Example. Let $S = \{(1 + 1/n)\boldsymbol{e}_n : n \in \mathbb{N}\} \subset \ell_p$. Then S is bounded, and is complete since there are no "non-trivial" Cauchy sequences in S. Since $d(\mathbf{0}, (1 + 1/n)\boldsymbol{e}_n) = 1 + 1/n$, we have $d(\mathbf{0}, S) = 1$, yet there is no $\boldsymbol{s} \in S$ for which $\|\boldsymbol{s} - \mathbf{0}\| = 1$.



Projectors

If S is a Chebyshev set in a normed space $(\mathcal{X}, \|\cdot\|)$, then we can define a **projector** $\mathrm{P} : \mathcal{X} \to S$ that, for each point $\boldsymbol{x} \in \mathcal{X}$, gives the closest point $\mathrm{P}(\boldsymbol{x}) \in S$. In other words, for a Chebyshev set S, we can define legitimately

$$\mathrm{P}(\boldsymbol{x}) = \arg\min_{\boldsymbol{s} \in S} \|\boldsymbol{x} - \boldsymbol{s}\|,$$

and "arg min" is well defined since there exists a unique minimizer when S is Chebyshev.

When needed for clarity, we will write P_S rather than just P.

Such a projector satisfies the following properties.
- $\mathrm{P}(\boldsymbol{x}) \in S, \ \forall \boldsymbol{x} \in S$
- $\|\boldsymbol{x} - \mathrm{P}(\boldsymbol{x})\| \leq \|\boldsymbol{x} - \boldsymbol{y}\|, \ \forall \boldsymbol{y} \in S$, *i.e.*, $\|\boldsymbol{x} - \mathrm{P}(\boldsymbol{x})\| = d(\boldsymbol{x}, S)$
- $\mathrm{P}(\mathrm{P}(\boldsymbol{x})) = \mathrm{P}(\boldsymbol{x})$ or more concisely: $\mathrm{P}^2 = \mathrm{P}$.

As noted above, closedness of S is a necessary condition for existence of a projector defined on all of $\mathcal{X}$.

. . . .

Example. Consider $\mathcal{X} = \mathbb{R}$ and $S = [0, \infty)$. This is a Chebyshev set with $\mathrm{P}(x) = \max\{x, 0\}$.

. . . .

Example. Consider $\mathcal{X} = \mathbb{R}^2$ and the (compact) set $K = \left\{\boldsymbol{y} \in \mathbb{R}^2 : \|\boldsymbol{y}\| = 1\right\}$ (the unit circle). There is not a unique minimizer of the distance to $\boldsymbol{x} = \boldsymbol{0}$, the center of the unit circle.

This why there is not a plethora of signal processing papers on "projections onto compact sets" (POKS?) methods.

3.3

Orthogonality

Definition. In an inner product space, two vectors $\boldsymbol{x}$,$\boldsymbol{y}$ are called **orthogonal** iff $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, which is denoted $\boldsymbol{x} \perp \boldsymbol{y}$.
(This is consistent with the earlier $\cos^{-1}$ definition of angle.)

Definition. A vector $\boldsymbol{x}$ is called orthogonal to a set S iff $\forall \boldsymbol{s} \in S, \ \boldsymbol{x} \perp \boldsymbol{s}$, in which case we write $\boldsymbol{x} \perp S$.

Definition. Two sets S and T in an inner product space $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ are called **orthogonal** and written $S \perp T$ iff

$$\langle \boldsymbol{s}, \boldsymbol{t} \rangle = 0, \ \forall \boldsymbol{s} \in S, \ \forall \boldsymbol{t} \in T.$$

Exercise. Show $\boldsymbol{x} \perp S = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ iff $\boldsymbol{x} \perp \boldsymbol{y}_i, \ i = 1, \ldots, n$.

Lemma. (**Pythagorean theorem**) $\boxed{\boldsymbol{x} \perp \boldsymbol{y} \Longrightarrow \|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2.}$

Proof. $\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\,\mathrm{real}(\langle \boldsymbol{x}, \boldsymbol{y} \rangle)$. □

The converse does not hold. Consider $\mathbb{C}$ and the vectors $\boldsymbol{x} = 1$ and $\boldsymbol{y} = \imath$. Then $\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 = 2$ but $\boldsymbol{x}$ and $\boldsymbol{y}$ are not perpendicular since here $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = xy^* = -\imath \neq 0$.

We first consider the easiest version of the general minimum norm problem, where the set of interest is actually a **subspace** of $\mathcal{X}$.

Theorem. ***(Pre-projection theorem)***
Let $\mathcal{X}$ be an inner product space, M a subspace of $\mathcal{X}$, and $\boldsymbol{x}$ a vector in $\mathcal{X}$.
- *If there exists a vector $\boldsymbol{m}_0 \in M$ such that $\|\boldsymbol{x} - \boldsymbol{m}_0\| = d(\boldsymbol{x}, M)$, then that $\boldsymbol{m}_0$ is unique.*
- *A necessary and sufficient condition that $\boldsymbol{m}_0 \in M$ be the unique minimizing vector in M is that the* ***error vector*** *$\boldsymbol{x} - \boldsymbol{m}_0$ be orthogonal to M.*

Proof.

Claim 1. If $\exists \boldsymbol{m} \in M$ (not necessarily unique) s.t. $\|\boldsymbol{x} - \boldsymbol{m}_0\| = d(\boldsymbol{x}, M)$, then $\boldsymbol{x} - \boldsymbol{m}_0 \perp M$.

Pf. We show by contradiction. Suppose $\boldsymbol{m}_0 \in M$ is a minimizer, but $\boldsymbol{x} - \boldsymbol{m}_0 \not\perp M$.
Then $\exists \boldsymbol{m} \in M$ s.t. $\langle \boldsymbol{x} - \boldsymbol{m}_0,\, \boldsymbol{m} \rangle = \delta \neq 0$, for some δ, where w.l.o.g. we assume $\|\boldsymbol{m}\| = 1$.
Consider $\boldsymbol{m}_1 \triangleq \boldsymbol{m}_0 + \delta \boldsymbol{m} \in M$.

$$\|\boldsymbol{x} - \boldsymbol{m}_1\|^2 = \|\boldsymbol{x} - \boldsymbol{m}_0 - \delta \boldsymbol{m}\|^2 = \|\boldsymbol{x} - \boldsymbol{m}_0\|^2 - \delta^* \langle \boldsymbol{x} - \boldsymbol{m}_0,\, \boldsymbol{m} \rangle - \delta \langle \boldsymbol{m},\, \boldsymbol{x} - \boldsymbol{m}_0 \rangle + |\delta|^2 = \|\boldsymbol{x} - \boldsymbol{m}_0\|^2 - |\delta|^2 < \|\boldsymbol{x} - \boldsymbol{m}_0\|^2,$$

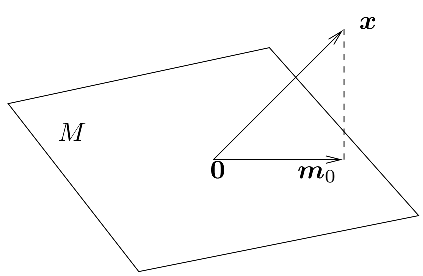
contradicting the assumption that $\boldsymbol{m}_0$ is a minimizer.

Claim 2. If $\exists \boldsymbol{m}_0 \in M$ s.t. $\boldsymbol{x} - \boldsymbol{m}_0 \perp M$, then $\boldsymbol{m}_0$ is a unique minimizer.

Pf. For any $\boldsymbol{m} \in M$, since $\boldsymbol{m}_0 - \boldsymbol{m} \in M$, by the Pythagorean theorem:

$$\|\boldsymbol{x} - \boldsymbol{m}\|^2 = \|(\boldsymbol{x} - \boldsymbol{m}_0) + (\boldsymbol{m}_0 - \boldsymbol{m})\|^2 = \|\boldsymbol{x} - \boldsymbol{m}_0\|^2 + \|\boldsymbol{m}_0 - \boldsymbol{m}\|^2 > \|\boldsymbol{x} - \boldsymbol{m}_0\|^2$$

if $\boldsymbol{m} \neq \boldsymbol{m}_0$. So $\boldsymbol{m}_0$, and only $\boldsymbol{m}_0$, is the minimizer. □



In words: in an inner product space, if a subspace is proximinal, then it is Chebyshev.

The "good thing" about this theorem is that it does not require completeness of $\mathcal{X}$, only the presence of an inner product. However, it does not prove the existence of a minimizer!

Is this lack of an existence proof simply because "we" have not been clever enough to find it?
Or, are there (incomplete) inner product spaces in which no such minimizer exists?
(We cannot find an example drawing 2d or 3d pictures, since $\mathbb{E}^n$ is complete!)

<u>Example</u>. Consider the Hilbert space ℓ_2 with the incomplete (and hence non-closed) subspace M that consists of sequences with only finitely many nonzero terms, and consider $\boldsymbol{x} = (1, 1/2, 1/3, 1/4, \ldots)$.

For $n \in \mathbb{N}$, let $\boldsymbol{m}_n$ be identical to $\boldsymbol{x}$ for the first n terms, and then zero thereafter. Then $\|\boldsymbol{x} - \boldsymbol{m}_n\|_2^2 = \sum_{k=n}^{\infty} (1/k)^2$, which approaches 0 as $n \to \infty$, but $\{\boldsymbol{m}_n\}$ converges to $\boldsymbol{x} \notin M$. So $d(\boldsymbol{x}, M) = \inf_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\| = 0$.

But clearly no $\boldsymbol{m} \in M$ achieves this minimum since $\boldsymbol{x}$ has an infinite number of nonzero terms.

But how about an example where $\mathcal{X}$ is incomplete and M is a closed subspace?

Example. (See [3, p. 289].)
Consider the (incomplete) inner product space $\mathcal{X} = $ ("finite-length" sequences, $\|\cdot\|_2$). Let $\boldsymbol{u} \in \ell_2$ denote the sequence of reals $u_i = 1/2^i$, and define the following (uncountably infinite) subspace: $M = \{\boldsymbol{x} \in \mathcal{X} : \sum_{i=1}^{\infty} x_i u_i = 0\}$.

Claim 1. M is closed.
Suppose $\{\boldsymbol{y}_n\} \in M$ and $\boldsymbol{y}_n \to \boldsymbol{y} \in \mathcal{X}$. Then (borrowing the ℓ_2 inner product): $|\langle \boldsymbol{y}, \boldsymbol{u} \rangle| = |\langle \boldsymbol{y} - \boldsymbol{y}_n, \boldsymbol{u} \rangle| \leq \|\boldsymbol{y} - \boldsymbol{y}_n\|_2 \|\boldsymbol{u}\|_2 \to 0$ as $n \to \infty$ (using Cauchy-Schwarz inequality) so $\langle \boldsymbol{y}, \boldsymbol{u} \rangle = 0$ and hence $\boldsymbol{y} \in M$.

Claim 2. $\inf_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\|$ is not achieved, where $\boldsymbol{x} = (1, 0, 0, \ldots) \notin M$ but $\boldsymbol{x} \in \mathcal{X}$.
Suppose $\boldsymbol{m} \in M$ achieves that infimum. Then by the pre-projection theorem, $\boldsymbol{z} \triangleq \boldsymbol{x} - \boldsymbol{m} \perp M$.
Since $\boldsymbol{z} \in \mathcal{X}$, $\boldsymbol{z} = (z_1, \ldots, z_N, 0, 0, \ldots)$ for some $N \in \mathbb{N}$.
Define $\boldsymbol{w}_i = \frac{1}{u_i}\boldsymbol{e}_i - \frac{1}{u_{N+1}}\boldsymbol{e}_{N+1}$ for $i = 1, \ldots, N$ where $\{\boldsymbol{e}_i\}$ denotes the standard basis vectors for ℓ_2. Since $\boldsymbol{w}_i \in M$, $\langle \boldsymbol{z}, \boldsymbol{w}_i \rangle = 0$. But therefore $z_i = 0$, $i = 1, \ldots, N$, so $\boldsymbol{z} = \boldsymbol{0}$, *i.e.*, $\boldsymbol{x} = \boldsymbol{m} \in M$. This contradicts the fact $\boldsymbol{x} \notin M$.

To establish *existence* of a minimizer, we make a stronger assumption: **completeness** of the subspace. Or, more frequently, we assume the inner product space is complete, so that all closed subspaces within it are complete. Why? **??**

> **Theorem.** *(The **classical projection theorem**)*
> *Let M be a complete **subspace** of an inner product space $\mathcal{H}$ (e.g., M may be a closed **subspace** of a Hilbert space).*
> • *For any $\boldsymbol{x} \in \mathcal{H}$, there exists a unique $\boldsymbol{m}_0 \in M$ such that $\|\boldsymbol{x} - \boldsymbol{m}_0\| = d(\boldsymbol{x}, M)$, i.e., M is Chebyshev.*
> • *Furthermore, a necessary and sufficient condition that $\boldsymbol{m}_0 \in M$ be the unique minimizer is that $\boldsymbol{x} - \boldsymbol{m}_0 \perp M$.*

Proof. Uniqueness and the characterization of $\boldsymbol{m}_0$ in terms of orthogonality was established in the pre-projection theorem, so we focus on existence.

Clearly, if $\boldsymbol{x} \in M$, then $\boldsymbol{m}_0 = \boldsymbol{x}$ and we are done. For $\boldsymbol{x} \notin M$, $\delta = d(\boldsymbol{x}, M) > 0$. Why? If $d(\boldsymbol{x}, M)$ were zero, then we could find a sequence $\boldsymbol{m}_i \in M$ such that $d(\boldsymbol{x}, \boldsymbol{m}_i) \to 0$, meaning $\boldsymbol{m}_i \to \boldsymbol{x}$, but that would imply $\boldsymbol{x} \in M$ because M is closed, contradicting $\boldsymbol{x} \notin M$. So $\delta > 0$.

Let $\{\boldsymbol{m}_i\}$ denote a sequence of vectors in M such that $\|\boldsymbol{x} - \boldsymbol{m}_i\| < \delta + 1/i$, which is possible by definition of $d(\boldsymbol{x}, M)$.

Claim 1. $\{\boldsymbol{m}_i\}$ is Cauchy.

By the parallelogram law:

$$\|(\boldsymbol{m}_j - \boldsymbol{x}) + (\boldsymbol{x} - \boldsymbol{m}_i)\|^2 + \|(\boldsymbol{m}_j - \boldsymbol{x}) - (\boldsymbol{x} - \boldsymbol{m}_i)\|^2 = 2\|\boldsymbol{m}_i - \boldsymbol{x}\|^2 + 2\|\boldsymbol{x} - \boldsymbol{m}_j\|^2$$

so

$$\|\boldsymbol{m}_i - \boldsymbol{m}_j\|^2 = 2\|\boldsymbol{m}_i - \boldsymbol{x}\|^2 + 2\|\boldsymbol{x} - \boldsymbol{m}_j\|^2 - 4\|\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{m}_j + \boldsymbol{m}_i)\|^2.$$

However, $\frac{1}{2}(\boldsymbol{m}_j + \boldsymbol{m}_i) \in M$, so $\|\boldsymbol{x} - \frac{1}{2}(\boldsymbol{m}_j + \boldsymbol{m}_i)\| \geq \delta$. Thus

$$\|\boldsymbol{m}_i - \boldsymbol{m}_j\|^2 \leq 2\|\boldsymbol{m}_i - \boldsymbol{x}\|^2 + 2\|\boldsymbol{x} - \boldsymbol{m}_j\|^2 - 4\delta^2 \to 2\delta^2 + 2\delta^2 - 4\delta^2 = 0 \text{ as } i, j \to \infty.$$

Since $\{\boldsymbol{m}_i\}$ is Cauchy, and M is complete, $\exists \boldsymbol{m}_0 \in M$ s.t. $\boldsymbol{m}_i \to \boldsymbol{m}_0$.

Since norms are continuous,

$$\|\boldsymbol{x} - \boldsymbol{m}_0\| = \left\|\boldsymbol{x} - \lim_{i\to\infty} \boldsymbol{m}_i\right\| = \lim_{i\to\infty} \|\boldsymbol{x} - \boldsymbol{m}_i\| \leq \lim_{i\to\infty} \delta + 1/i = \delta = d(\boldsymbol{x}, M).$$

□

Remark. The key step in the proof was (the clever use of) the parallelogram law, a defining property of inner product spaces.

Remark. The proof uses only completeness of M, not of $\mathcal{H}$. We will use this generality in a subsequent example.

Polynomial approximation example

Consider the problem of approximating the function $x(t) = \sin^{-1} t$ over the interval [-1,1] by a third-order polynomial.

If we want the approximation to fit better at the ends than in the middle, then the following inner product space is reasonable:

$$\mathcal{X} = \{f : [-1,1] \to \mathbb{R} : f \text{ is continuous}\}, \ \langle f, g \rangle = \int_{-1}^{1} w(t) f(t) g(t) \, dt, \text{ where } w(t) = 1 + t^2. \qquad \textit{\textbf{(Picture)}}$$

Since $x(t)$ is an odd function, the following subspace of $\mathcal{X}$ suffices:

$$M = \left\{ at + bt^3 : a, b \in \mathbb{R} \right\}.$$

Is $\mathcal{X}$ complete? **??** Is M? **??**

To find the best 3rd-order polynomial approximation, *i.e.*, $\boldsymbol{m}_\star = \arg\min_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\|$, we apply the projection theorem, which characterizes the minimizer through $\boldsymbol{x} - \boldsymbol{m}_\star \perp M$. Denoting $\boldsymbol{m}_\star(t) = ct + dt^3$, then

$$\int_{-1}^{1} w(t)(\sin^{-1} t - ct - dt^3)(at + bt^3) \, dt = 0, \ \forall a, b \in \mathbb{R}.$$

Since $aq + br = 0, \forall a, b \in \mathbb{R} \iff q = r = 0$, we can reduce the problem to the following finite-dimensional system of equations:

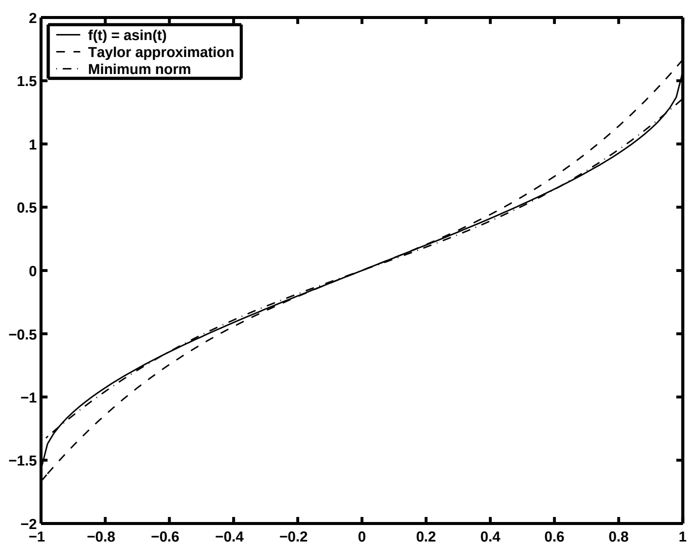
$$\begin{bmatrix} \int_{-1}^{1} (1+t^2) t \sin^{-1} t \, dt \\ \int_{-1}^{1} (1+t^2) t^3 \sin^{-1} t \, dt \end{bmatrix} = \begin{bmatrix} \int_{-1}^{1} (1+t^2) t^2 \, dt & \int_{-1}^{1} (1+t^2) t^4 \, dt \\ \int_{-1}^{1} (1+t^2) t^4 \, dt & \int_{-1}^{1} (1+t^2) t^6 \, dt \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}.$$

Using MATLAB's symbolic toolbox for the integration yields:

$$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 16/15 & 24/35 \\ 24/35 & 32/63 \end{bmatrix}^{-1} \begin{bmatrix} 13/32 \\ 13/48 \end{bmatrix} \pi = \begin{bmatrix} 296/1027 \\ 148/1027 \end{bmatrix} \pi.$$

Thus $\boldsymbol{m}_\star(t) = \frac{296\pi}{1027} t + \frac{148\pi}{1027} t^3$, so $\dot{\boldsymbol{m}}_\star(0) = 296\pi/1027 \approx 0.91$.

The following figure shows $x(t)$, the Taylor approximation $t + t^3/3!$, and the minimum norm approximation $\boldsymbol{m}_\star(t)$. Although the Taylor approximation fits the best near $t = 0$, the minimum norm approximation has a better overall fit.



It would be fair to argue that we did not really need the general version of the projection theorem for this problem. We could have solved $\min_{a,b \in \mathbb{R}} \int_{-1}^{1} w(t)(x(t) - at - bt^3)^2 \, dt$ by conventional methods. The forthcoming Fourier series examples, where M is infinite dimensional, are (perhaps?) more compelling.

What about $\|\cdot\|_1$? **??**

Complete subspaces versus Chebyshev subspaces

The preceding theorems and examples might lead one to conjecture that completeness of a subspace M is a *necessary* condition for M to be Chebyshev. The next example shows otherwise.

Example. Consider the (incomplete) inner product space $\mathcal{X}$ consisting of sequences with finitely many nonzero terms, with the usual ℓ_2 inner product, and define the subspace $M = \{(a_1, a_2, \ldots) \in \mathcal{X} : a_1 = 0\}$. This subspace is closed, but is not complete for the same reasons that $\mathcal{X}$ is incomplete. Nevertheless, M is Chebyshev, and $\boldsymbol{x} = (a_1, a_2, a_3, \ldots) \Longrightarrow \mathrm{P}_M(\boldsymbol{x}) = (0, a_2, a_3, \ldots)$.

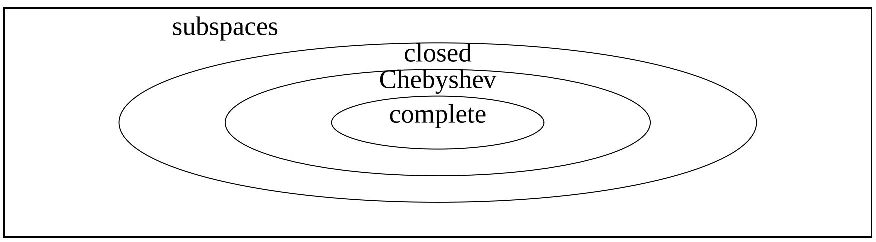
These complications are eliminated if we focus on Hilbert spaces. In a Hilbert space, completeness of a subspace becomes a *necessary* condition for the subspace to be Chebyshev.

Theorem. *[22, p. 192] In a Hilbert space $\mathcal{H}$, a subspace M is Chebyshev if and only if M is complete.*

Proof.
M complete $\Longrightarrow$ M Chebyshev follows from the projection theorem.
M Chebyshev $\Longrightarrow$ M closed $\Longrightarrow$ M is complete (since $\mathcal{H}$ is also a complete normed space).

Summary

The following Venn diagram summarizes the situation with subspaces in any inner product space.



In Hilbert spaces (including all finite-dimensional inner product spaces), the three ellipses coincide!

Example. In signal processing problems, the subspace of **band-limited** (continuous-time) signals with a certain band-limit is important. Is this subspace complete?

Consider $\mathcal{X} = \mathcal{L}_2[\mathbb{R}]$ and define $M \subset \mathcal{X}$ to be the subspace of all square-integrable functions having Fourier transform supported on the frequency interval $[-\nu_{\max}, \nu_{\max}]$. Since $\mathcal{X}$ is a Hilbert space, the preceding theorem says that the question of whether M is complete is equivalent to determining whether M is Chebyshev. In this case, a simple way to answer that is to construct a projector for M. Given $\boldsymbol{x} \in \mathcal{X}$, can we find a (unique, in the $\mathcal{L}_2$ sense) band-limited function $\boldsymbol{y}_\star$ such that

$$\|\boldsymbol{y}_\star - \boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{x}\|, \ \forall \boldsymbol{y} \in M.$$

Applying Parseval's theorem from Fourier analysis:

$$\begin{aligned} \boldsymbol{y} \in M \Longrightarrow \|\boldsymbol{y} - \boldsymbol{x}\|^2 &= \int_{-\infty}^{\infty} |y(t) - x(t)|^2 \,\mathrm{d}t = \int_{-\infty}^{\infty} |Y(\nu) - X(\nu)|^2 \,\mathrm{d}\nu \\ &= \int_{-\nu_{\max}}^{\nu_{\max}} |Y(\nu) - X(\nu)|^2 \,\mathrm{d}\nu + \int_{|\nu| > \nu_{\max}} |X(\nu)|^2 \,\mathrm{d}\nu\,. \end{aligned}$$

Clearly this is minimized by taking $\boldsymbol{y}_\star$ to be the (unique in $\mathcal{L}_2$ sense) signal having Fourier transform

$$Y_\star(\nu) = \begin{cases} X(\nu), & |\nu| \leq \nu_{\max} \\ 0, & \text{otherwise.} \end{cases}$$

Since a projector exists, M is Chebyshev, and hence M is complete.

Of course, in this case M is also convex.

3.4

Orthogonal complements (The key to **duality**)

We saw in the projection theorem that an **orthogonality** condition characterizes the closest point in a complete subspace of an inner product space. We now examine orthogonality in more detail.

In $\mathbb{R}^2$, we decompose any vector into a sum of an “x-component” vector and a “y-component vector,” the two of which are orthogonal. We can extend this concept considerably in general Hilbert spaces.

Definition. If S is a subset of an inner product space $(\mathcal{X}, \langle\cdot,\,\cdot\rangle)$, then the **orthogonal complement** of S is the *subspace*:

$$S^\perp \triangleq \{\boldsymbol{x} \in \mathcal{X} : \boldsymbol{x} \perp S\}.$$

Clearly
- $\boldsymbol{x} \in S^\perp \iff \boldsymbol{x} \perp S$
- $\boldsymbol{y} \in S \Longrightarrow \boldsymbol{y} \perp S^\perp$.

Example. What is $\{\mathbf{0}\}^\perp$? **??**

Example. In 3-space, what is the orthogonal complement of the “cone” $\{\alpha\boldsymbol{x} : \alpha \in [0,\infty)\}$ for some $\boldsymbol{x} \neq \mathbf{0}$? **??**

Proposition. If S and T are **subsets** of an inner product space $(\mathcal{X}, \langle\cdot,\,\cdot\rangle)$, then the following hold.
(a) $S^\perp$ is a closed subspace of $\mathcal{X}$.
(b) $S \subseteq S^{\perp\perp}$ When equal? See next theorem!
(c) If $S \subseteq T$, then $T^\perp \subseteq S^\perp$.
(d) $S^{\perp\perp\perp} = S^\perp$
(f) $S^\perp = \left(\overline{S}\right)^\perp = \overline{S^\perp}$

Proof.

(a) $S^\perp$ is a subspace since linear combinations of vectors orthogonal to S are also orthogonal to S.
Suppose $\{\boldsymbol{x}_n\} \in S^\perp$ with $\boldsymbol{x}_n \to \boldsymbol{x} \in \mathcal{X}$. Since $\langle \boldsymbol{x}_n,\, \boldsymbol{s}\rangle = 0,\ \forall \boldsymbol{s} \in S$, by continuity of the inner product we have:
$\langle \boldsymbol{x},\, \boldsymbol{s}\rangle = \langle \lim_{n\to\infty} \boldsymbol{x}_n,\, \boldsymbol{s}\rangle = \lim_{n\to\infty}\langle \boldsymbol{x}_n,\, \boldsymbol{s}\rangle = 0$, so $\boldsymbol{x} \in S^\perp$.

(b) $\boldsymbol{y} \in S \Longrightarrow \boldsymbol{y} \perp S^\perp \Longrightarrow \boldsymbol{y} \in S^{\perp\perp}$, so $S \subseteq S^{\perp\perp}$.

(c) If $S \subseteq T$, then $\boldsymbol{y} \in T^\perp \Longrightarrow \boldsymbol{y} \perp \boldsymbol{x},\ \forall \boldsymbol{x} \in T \Longrightarrow \boldsymbol{y} \perp \boldsymbol{x},\ \forall \boldsymbol{x} \in S \Longrightarrow \boldsymbol{y} \in S^\perp$. So $T^\perp \subseteq S^\perp$.

(d) From above $S^\perp \subseteq S^{\perp\perp\perp}$. Also $S \subseteq S^{\perp\perp}$ so by 3rd property: $S^{\perp\perp\perp} \subseteq S^\perp$. Thus $S^{\perp\perp\perp} = S^\perp$.

(f) From (a), $S^\perp = \overline{S^\perp}$, which is the second equality. Now pick any $\boldsymbol{x} \perp S$. If $\boldsymbol{y} \in \overline{S}$, then $\exists \boldsymbol{s}_n \in S$ s.t. $\boldsymbol{s}_n \to \boldsymbol{y}$. Since $\boldsymbol{x} \perp S$, we have $\boldsymbol{x} \perp \boldsymbol{s}_n,\ \forall n$, so $\langle \boldsymbol{x},\, \boldsymbol{y}\rangle = \langle \boldsymbol{x},\, \lim_{n\to\infty} \boldsymbol{s}_n\rangle = \lim_{n\to\infty}\langle \boldsymbol{x},\, \boldsymbol{s}_n\rangle = 0$. Thus $\boldsymbol{x} \perp S \Longrightarrow \boldsymbol{x} \perp \overline{S}$.
Since $\boldsymbol{x}$ was arbitrary, $S^\perp \subseteq \left(\overline{S}\right)^\perp$. Furthermore, $S \subseteq \overline{S} \Longrightarrow \left(\overline{S}\right)^\perp \subseteq S^\perp$. Thus $S^\perp = \left(\overline{S}\right)^\perp$. □

Proposition. If S is a **subset** of a Hilbert space, then the following hold.
(e) $S^{\perp\perp} = \overline{[S]}$, *i.e.*, $S^{\perp\perp}$ is the *smallest* closed subspace containing S.
(g) $S^\perp$ is complete

Proof.

(e) See problem *3.9*. **??**

(g) follows since $S^\perp$ is closed and a Hilbert space is complete. □

Example. $\mathbb{E}^2$ with $S = \{(1,0)\}$. Then $S^\perp = [(0,1)]$, $S^{\perp\perp} = [(1,0)] \neq S$, $S^{\perp\perp\perp} = [(0,1)] = S^\perp$.
Example. $\mathbb{E}^2$ with $S = \{(1,0),\ (2,0)\}$. and $T = \{(1,0)\}$. Then $S^\perp = [(0,1)] = T^\perp$, yet it is *not* the case that $S \subset T$. So the converse of (c) above fails.

Orthogonal projection

The projection theorem allows us to extend the geometric projection property of Euclidean space to general inner product spaces.

Definition. Let M be a **Chebyshev subspace** of an inner product space $\mathcal{X}$. For each $\boldsymbol{x} \in \mathcal{X}$, let $\mathrm{P}(\boldsymbol{x})$ be the unique point in M closest to $\boldsymbol{x}$:

$$\mathrm{P}(\boldsymbol{x}) = \arg\min_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\| .$$

Then $\mathrm{P} : \mathcal{X} \to M$ is called the **orthogonal projection** of $\mathcal{X}$ onto M. When needed for clarity, we write P_M rather than just P.

Lemma. If M is a **Chebyshev subspace** of an inner product space, then
- $M^\perp$ is a Chebyshev set,
- $\mathrm{P}_{M^\perp}(\boldsymbol{x}) = \mathrm{P}_M^\perp(\boldsymbol{x}) \triangleq \boldsymbol{x} - \mathrm{P}_M(\boldsymbol{x})$, which is called the **projection onto the orthogonal complement**[2],
- $\boldsymbol{x} = \mathrm{P}_M(\boldsymbol{x}) + \mathrm{P}_M^\perp(\boldsymbol{x})$ where $\mathrm{P}_M(\boldsymbol{x}) \in M, \quad \mathrm{P}_M^\perp(\boldsymbol{x}) \in M^\perp$.

Proof. (Exercise)

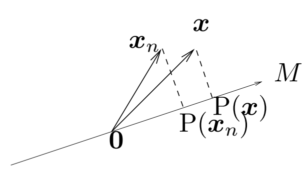
Recall the following properties of the projector P onto a Chebyshev subset S.
- $\mathrm{P}(\boldsymbol{x}) \in S$,
- $\|\boldsymbol{x} - \mathrm{P}(\boldsymbol{x})\| = d(\boldsymbol{x}, S)$
- $\mathrm{P}(\mathrm{P}(\boldsymbol{x})) = \mathrm{P}(\boldsymbol{x})$ (**idempotent**)

Here are some trivial properties of *orthogonal* projections (onto subspaces) that follow directly from the projection theorem.
- $\boldsymbol{x} - \mathrm{P}(\boldsymbol{x}) \perp M$
- $\mathrm{P}^\perp(\boldsymbol{x}) \in M^\perp$
- $\mathrm{P}(\boldsymbol{x}) \perp \mathrm{P}^\perp(\boldsymbol{x})$
- $\boldsymbol{x} = \mathrm{P}(\boldsymbol{x}) + \mathrm{P}^\perp(\boldsymbol{x})$

Here are some less trivial properties.

Proposition. If M is a Chebyshev subspace of an inner product space $\mathcal{X}$, then the orthogonal projector $\mathrm{P} = \mathrm{P}_M$ has the following properties.
(a) $\mathrm{P}(\boldsymbol{x}) = \boldsymbol{0}$ iff $\boldsymbol{x} \perp M$ (*cf.* earlier figure)
(b) $\mathrm{P} : \mathcal{X} \to M$ is a **linear operator** (*cf.* earlier figure)
(c) $|\!|\!| \mathrm{P} |\!|\!| = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \|\mathrm{P}(\boldsymbol{x})\| \,/\, \|\boldsymbol{x}\| = 1$ (see p. 105) (provided M is at least 1 dimensional, *i.e.*, not simply $\{\boldsymbol{0}\}$)
(d) P is continuous, *i.e.*, $\boldsymbol{x}_n \to \boldsymbol{x} \Longrightarrow \mathrm{P}(\boldsymbol{x}_n) \to \mathrm{P}(\boldsymbol{x})$.

Proof.
(a) $\mathrm{P}(\boldsymbol{x}) = \boldsymbol{0} \iff \boldsymbol{x} \perp M$ follows directly from the "characterization" part of the pre-projection theorem.
(b) $\mathrm{P}(\boldsymbol{x}_1) = \boldsymbol{m}_1$ and $\mathrm{P}(\boldsymbol{x}_2) = \boldsymbol{m}_2 \Longrightarrow \boldsymbol{x}_1 - \boldsymbol{m}_1 \perp M$ and $\boldsymbol{x}_2 - \boldsymbol{m}_2 \perp M$.
Thus $\alpha(\boldsymbol{x}_1 - \boldsymbol{m}_1) + \beta(\boldsymbol{x}_2 - \boldsymbol{m}_2) \perp M$ so $(\alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2) - (\alpha\boldsymbol{m}_1 + \beta\boldsymbol{m}_2) \perp M$.
Thus $\mathrm{P}(\alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2) = \alpha\boldsymbol{m}_1 + \beta\boldsymbol{m}_2 = \alpha\,\mathrm{P}(\boldsymbol{x}) + \beta\,\mathrm{P}(\boldsymbol{x}_2)$ by the pre-projection theorem.
(c) $\boldsymbol{x} \in M \Longrightarrow \mathrm{P}(\boldsymbol{x}) = \boldsymbol{x}$ so $|\!|\!| \mathrm{P} |\!|\!| \geq 1$. $\boldsymbol{x} \in \mathcal{H} \Longrightarrow \boldsymbol{x} - \mathrm{P}(\boldsymbol{x}) \perp M \Longrightarrow \boldsymbol{x} - \mathrm{P}(\boldsymbol{x}) \perp \mathrm{P}(\boldsymbol{x})$ since $\mathrm{P}(\boldsymbol{x}) \in M$.
So by Pythagorean: $\|\boldsymbol{x}\|^2 = \|\boldsymbol{x} - \mathrm{P}(\boldsymbol{x}) + \mathrm{P}(\boldsymbol{x})\|^2 = \|\boldsymbol{x} - \mathrm{P}(\boldsymbol{x})\|^2 + \|\mathrm{P}(\boldsymbol{x})\|^2 \geq \|\mathrm{P}(\boldsymbol{x})\|^2$ so $\|\mathrm{P}(\boldsymbol{x})\| \,/\, \|\boldsymbol{x}\| \leq 1$.
(d) **Exercise.** **??**



□

[2]Notice the reuse of the symbol $^\perp$ here. This is reasonable since $\mathrm{P}_{M^\perp} = \mathrm{P}_M^\perp$ when M is a Chebyshev subspace in an inner product space.

Direct sum

Recall that if S,T are subsets of a common vector space $\mathcal{X}$, then $S + T = \{\boldsymbol{s} + \boldsymbol{t} : \boldsymbol{s} \in S,\ \boldsymbol{t} \in T\}$. However, in general, if $\boldsymbol{x} \in S + T$, decompositions of the form $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{t}$ need not be unique.

Definition. A vector space $\mathcal{X}$ is called the **direct sum** of subspaces M and N iff each $\boldsymbol{x} \in \mathcal{X}$ has a *unique* representation as $\boldsymbol{x} = \boldsymbol{m} + \boldsymbol{n}$ where $\boldsymbol{m} \in M$ and $\boldsymbol{n} \in N$. We notate this situation as follows:

$$\mathcal{X} = M \oplus N.$$

Fact. If $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ are a linearly independent set of vectors, then

$$[\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}] = [\{\boldsymbol{u}_1\}] \oplus \cdots \oplus [\{\boldsymbol{u}_n\}]$$

This is an algebraic concept, so we could have introduced it much earlier. But its main use is in the context of Hilbert spaces.

Theorem. *If M is a **Chebyshev subspace** of an inner product space $\mathcal{X}$, then*

$$\mathcal{X} = M \oplus M^\perp, \text{ and } M^{\perp\perp} = M.$$

Proof. As shown previously, $\boldsymbol{x} = \boldsymbol{m}_\star + \boldsymbol{n}_\star$ where $\boldsymbol{m}_\star = \mathrm{P}(\boldsymbol{x}) \in M$ and $\boldsymbol{n}_\star = \mathrm{P}^\perp(\boldsymbol{x}) \in M^\perp$.

(However, uniqueness of $\boldsymbol{m}_\star$ as the minimizer in M of $\|\boldsymbol{x} - \boldsymbol{m}\|$ does not alone ensure uniqueness of the decomposition $\boldsymbol{m}_\star + \boldsymbol{n}_\star$, so we must prove uniqueness next.)

Suppose $\boldsymbol{x} = \boldsymbol{m} + \boldsymbol{n}$ with $\boldsymbol{m} \in M$ and $\boldsymbol{n} \in M^\perp$. Then $\mathbf{0} = \boldsymbol{x} - \boldsymbol{x} = (\boldsymbol{m}_\star + \boldsymbol{n}_\star) - (\boldsymbol{m} + \boldsymbol{n}) = (\boldsymbol{m}_\star - \boldsymbol{m}) + (\boldsymbol{n}_\star - \boldsymbol{n})$, but $\boldsymbol{m}_\star - \boldsymbol{m} \perp \boldsymbol{n}_\star - \boldsymbol{n}$, so by the Pythagorean theorem: $0 = \|\mathbf{0}\|^2 = \|\boldsymbol{m}_\star - \boldsymbol{m}\|^2 + \|\boldsymbol{n}_\star - \boldsymbol{n}\|^2$. Thus $\boldsymbol{m}_\star = \boldsymbol{m}$ and $\boldsymbol{n}_\star = \boldsymbol{n}$.

Since $M \subseteq M^{\perp\perp}$ was shown previously, we need to show $M^{\perp\perp} \subseteq M$.
Suppose $\boldsymbol{x} \in M^{\perp\perp}$. By the first part of this theorem, $\boldsymbol{x} = \boldsymbol{m} + \boldsymbol{n}$ where $\boldsymbol{m} \in M \subseteq M^{\perp\perp}$ and $\boldsymbol{n} \in M^\perp$.
Since both $\boldsymbol{x}$ and $\boldsymbol{m}$ are in $M^{\perp\perp}$, $\boldsymbol{n} = \boldsymbol{x} - \boldsymbol{m} \in M^{\perp\perp}$. But also $\boldsymbol{n} \in M^\perp$, so $\boldsymbol{n} \perp \boldsymbol{n} = 0 \Longrightarrow \boldsymbol{n} = \mathbf{0}$.
Thus $\boldsymbol{x} = \boldsymbol{m} \in M$ and hence $M^{\perp\perp} \subseteq M$ since $\boldsymbol{x}$ was arbitrary. □

Corollary. For any subset S of a Hilbert space $\mathcal{X}$: $\mathcal{X} = \overline{[S]} \oplus \overline{[S]}^\perp$.

Summarizing previous results:

- In any inner product space, for any *subset* S we have $S \subseteq S^{\perp\perp}$.
- In any inner product space, for any **Chebyshev subspace** M we have $M = M^{\perp\perp}$.

Example. A subspace M in a Hilbert space where $M \neq M^{\perp\perp}$.
Take $\mathcal{X} = \ell_2$ and $M = \{\text{sequences with finitely many nonzero terms}\}$ (not closed). Then $M^\perp = \{\mathbf{0}\}$ and $M^{\perp\perp} = \ell_2 = \overline{M} \neq M$.

Example of a closed subspace in an inner product space where $M \neq M^{\perp\perp}$? (Exercise). **??**

Having established the fundamental theory of inner product spaces, we now move towards “applications:” Fourier series and other minimum norm problems like approximation.

3.5

Orthogonal sets

Orthogonal sets of vectors in a Hilbert space (such as complex exponentials for ordinary Fourier series) are very useful in applications such as signal analysis.

Definition. A set S of vectors in an inner product space is called an **orthogonal set** iff

$$\boldsymbol{u}, \boldsymbol{v} \in S,\ \boldsymbol{u} \neq \boldsymbol{v} \Longrightarrow \boldsymbol{u} \perp \boldsymbol{v}.$$

If in addition each vector in S has unity norm, then S is called an **orthonormal set**.

Remark. An orthogonal set can include the zero vector. An orthonormal set cannot.

Example. $\mathcal{L}_2[0,1]$ with $S = \{\cos(2\pi k t) : k \in \mathbb{N}\}$ (countable).

Do uncountable orthogonal sets exist? The example $S = \left\{1_{\{t=a\}} : a \in (0,1)\right\}$ fails since $S \equiv \mathbf{0}$ in $\mathcal{L}_2$.

Proposition. In any inner product space, an orthogonal set of nonzero vectors is a linearly independent set.

Proof. Suppose $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\} \subset S$ and $\sum_{i=1}^n \alpha_i \boldsymbol{u}_i = \mathbf{0}$ yet $\boldsymbol{u}_i \neq \mathbf{0},\ \forall i$. Then

$$0 = \langle \mathbf{0},\ \boldsymbol{u}_k \rangle = \left\langle \sum_{i=1}^n \alpha_i \boldsymbol{u}_i,\ \boldsymbol{u}_k \right\rangle = \alpha_k \|\boldsymbol{u}_k\|^2,$$

so $\alpha_k = 0,\ k = 1, \ldots, n$. Thus the vectors are linearly independent.
Since n and the α_i's and $\boldsymbol{u}_i$'s were arbitrary, the set is linearly independent. □

Fact. (Projection onto the span of a single vector.) Using a convenient shorthand:

$$\mathrm{P}_{\boldsymbol{u}}(\boldsymbol{x}) \triangleq \mathrm{P}_{[\{\boldsymbol{u}\}]}(\boldsymbol{x}) = \begin{cases} \frac{\langle \boldsymbol{x},\, \boldsymbol{u} \rangle}{\|\boldsymbol{u}\|^2} \boldsymbol{u}, & \boldsymbol{u} \neq \mathbf{0} \\ \mathbf{0}, & \boldsymbol{u} = \mathbf{0}. \end{cases}$$

Proof. $\langle \alpha \boldsymbol{u},\ \boldsymbol{x} - \mathrm{P}_{\boldsymbol{u}}(\boldsymbol{x}) \rangle = \alpha \left\langle \boldsymbol{u},\ \boldsymbol{x} - \frac{\langle \boldsymbol{x},\, \boldsymbol{u} \rangle}{\|\boldsymbol{u}\|^2} \boldsymbol{u} \right\rangle = \alpha \left(\langle \boldsymbol{u},\ \boldsymbol{x} \rangle - \langle \boldsymbol{x},\ \boldsymbol{u} \rangle^* \right) = 0$, so $\boldsymbol{x} - \mathrm{P}_{\boldsymbol{u}}(\boldsymbol{x}) \perp [\{\boldsymbol{u}\}]$.

Fact. If $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ form an orthogonal set, then

$$\mathrm{P}_{[\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}]}(\boldsymbol{x}) = \mathrm{P}_{\boldsymbol{u}_1}(\boldsymbol{x}) + \cdots + \mathrm{P}_{\boldsymbol{u}_n}(\boldsymbol{x}).$$

Proof. Exercise.

Fact. If M and N are orthogonal **Chebyshev subspace**s of an inner product space, then *(Prob. 3.7)*

- $M + N = M \oplus N$,
- $M \oplus N$ is a **Chebyshev subspace**, and
- $\mathrm{P}_{M \oplus N}(\boldsymbol{x}) = \mathrm{P}_M(\boldsymbol{x}) + \mathrm{P}_N(\boldsymbol{x})$.

Gram-Schmidt procedure

Since orthonormal sets are so convenient, it is fortunate the we can always create such sets by using the **Gram-Schmidt orthogonalization procedure** described in the proof of the following theorem.

This is another generalization of a familiar method in finite dimensions to general inner product spaces.

Theorem. ***(Gram-Schmidt)***
If $\{\boldsymbol{x}_i\}$ is a finite or countable sequence of linearly independent vectors in an inner product space $(\mathcal{X}, \langle \cdot, \cdot \rangle)$, then there exists an orthonormal sequence $\{\boldsymbol{e}_i\}$ such that

$$[\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}] = [\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}], \qquad \forall n \in \mathbb{N}.$$

Proof. Linearly independent vectors are necessarily nonzero, so $\|\boldsymbol{x}_i\| \neq 0$.

Take $\boldsymbol{e}_1 = \boldsymbol{x}_1 / \|\boldsymbol{x}_1\|$, which clearly has unity norm and spans the same space as $\boldsymbol{x}_1$.

Form the remaining vectors recursively:

$$\boldsymbol{z}_n = \boldsymbol{x}_n - \sum_{i=1}^{n-1} \langle \boldsymbol{x}_n, \boldsymbol{e}_i \rangle \boldsymbol{e}_i, \qquad \boldsymbol{e}_n = \boldsymbol{z}_n / \|\boldsymbol{z}_n\|, \qquad n = 2, 3, \ldots.$$

Being a linear combination of linearly independent vectors $\boldsymbol{z}_n$ is nonzero. And $\boldsymbol{z}_n \perp \boldsymbol{e}_i$ for $i = 1, \ldots, n-1$ is easily verified. Since we can write $\boldsymbol{x}_n$ as a linear combination of the $\boldsymbol{e}_i$ vectors, by an induction argument the span of $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ equals the span of $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. □

Note:

$$\boldsymbol{z}_n = \boldsymbol{x}_n - \sum_{i=1}^{n-1} \langle \boldsymbol{x}_n, \boldsymbol{e}_i \rangle \boldsymbol{e}_i = \boldsymbol{x}_n - \sum_{i=1}^{n-1} P_{\boldsymbol{e}_i}(\boldsymbol{x}_n) = \boldsymbol{x}_n - P_{[\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}\}]}(\boldsymbol{x}_n) = \mathrm{P}^{\perp}_{[\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1}\}]}(\boldsymbol{x}_n) = \mathrm{P}^{\perp}_{[\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}\}]}(\boldsymbol{x}_n).$$

. . . .

Corollary. Any finite-dimensional inner product space has an orthonormal basis.

Example. The polynomials $x_i(t) = t^i,\ i = 0, 1, \ldots,$ are linearly independent in $\mathcal{L}_2[-1, 1]$.
Why? **??**

Applying Gram-Schmidt yields the orthogonal polynomials:

$$\begin{aligned} e_0(t) &= \frac{x_0(t)}{\|\boldsymbol{x}_0\|} = \frac{1}{\sqrt{2}}, \quad z_1(t) = x_1(t) - \langle \boldsymbol{x}_1, \boldsymbol{e}_0 \rangle e_0(t) = t - \left(\int_{-1}^{1} t \cdot \frac{1}{\sqrt{2}} \, \mathrm{d}t \right) \cdot \frac{1}{\sqrt{2}} = t, \quad \|\boldsymbol{z}_1\|^2 = \int_{-1}^{1} t^2 \, \mathrm{d}t = 2/3, \\ e_1(t) &= \frac{z_1(t)}{\|\boldsymbol{z}_1\|} = \sqrt{\frac{3}{2}} t, \quad z_2(t) = x_2(t) - \langle \boldsymbol{x}_2, \boldsymbol{e}_0 \rangle e_0(t) - \langle \boldsymbol{x}_2, \boldsymbol{e}_1 \rangle e_1(t) = t^2 - \frac{2}{3}, \ \ldots. \end{aligned}$$

One can show by induction that

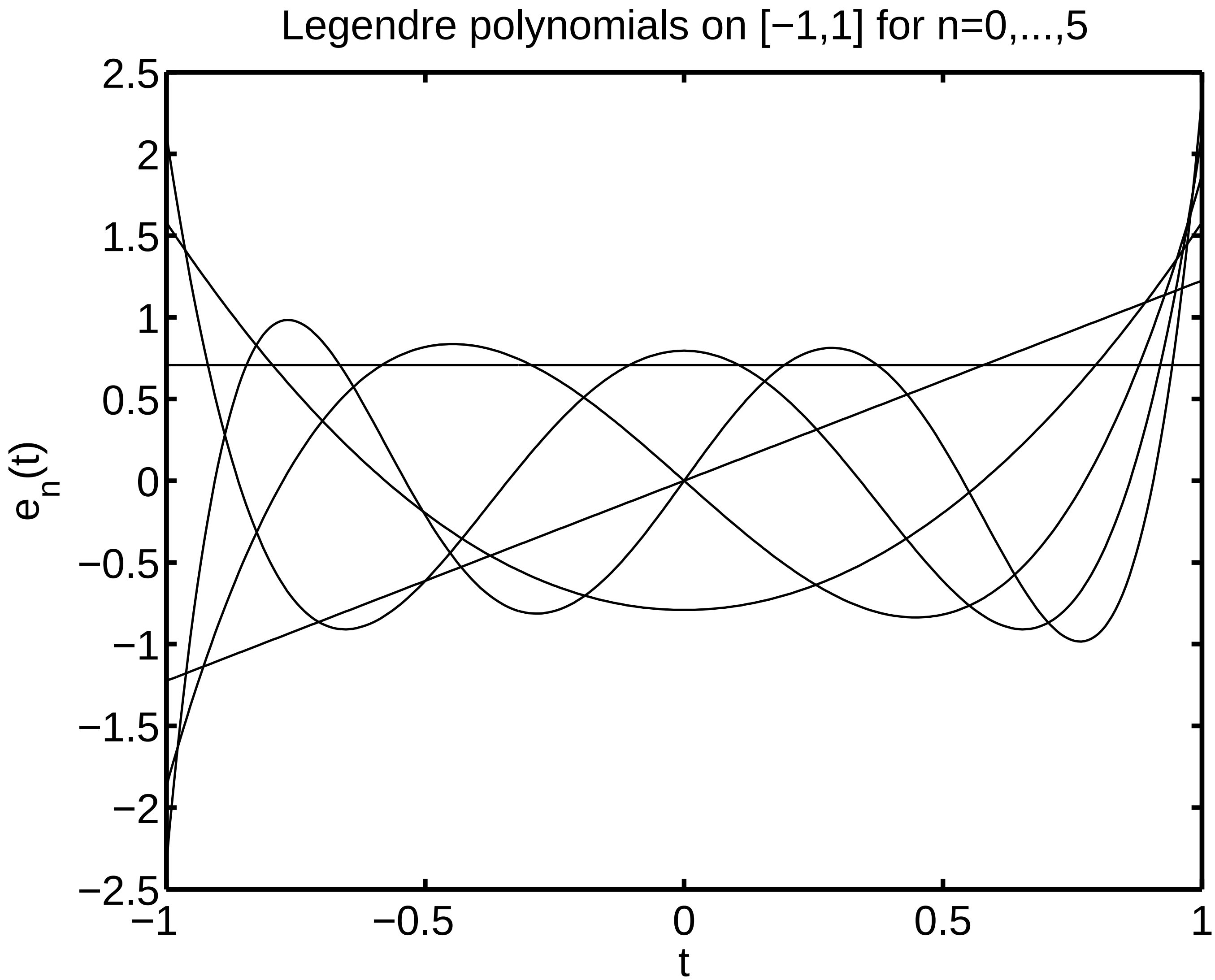
$$e_n(t) = \sqrt{\frac{2n+1}{2}} P_n(t), \qquad n = 0, 1, 2, \ldots,$$

where the $P_n(t)$ are the **Legendre polynomials**

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} (t^2 - 1)^n.$$

Clearly some subset of these $e_i(t)$'s will be an orthonormal basis for any finite-dimensional space of polynomials.

Is the entire collection some type of "basis" for $\mathcal{L}_2[-1, 1]$? (It is not a Hamel basis for $\mathcal{L}_2$.)
We will return to this question soon!


Legendre polynomials on [−1,1] for n=0,...,5
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2.5
$e_n(t)$
−1
−0.5
0
0.5
1
t

Approximation

A previous example considered the problem of approximating $\arcsin(t)$ by a 3rd-order polynomial, which reduced to a 2 by 2 system of equations since only 2 of the 4 coefficients were relevant.

We now consider such **approximation** problems more generally, and see that such reductions to a finite system of equations is the general behavior when the subspace M is finite dimensional.

Suppose M is a *finite-dimensional subspace* of an inner product space $(\mathcal{X}, \langle\cdot,\, \cdot\rangle)$. Then by definition, $M = [\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}]$ for some vectors $\boldsymbol{y}_i \in \mathcal{X}$. Furthermore, being finite dimensional, M is *complete*, so by the projection theorem, M is a Chebyshev set. Thus, given an arbitrary vector $\boldsymbol{x} \in \mathcal{X}$, there exists a unique *approximation* $\hat{\boldsymbol{x}} \in M$ that is *closest* to $\boldsymbol{x}$, as measured, of course, by the norm induced by the inner product:

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| = d(\boldsymbol{x}, M) = \inf_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\|\,.$$

Now we would like to find an explicit *formula* for $\hat{\boldsymbol{x}}$, since "existence and uniqueness" alone is inadequate for most practical applications.

3.6

Normal equations

Since $\hat{\boldsymbol{x}} \in M \Longrightarrow \hat{\boldsymbol{x}} = \sum_{i=1}^n \alpha_i \boldsymbol{y}_i$, we must find the n scalars $\{\alpha_i\}$ that minimize $\|\boldsymbol{x} - \sum_{i=1}^n \alpha_i \boldsymbol{y}_i\|$.

The **projection theorem** ensures that $\hat{\boldsymbol{x}}$ exists, and is characterized by $\boldsymbol{x} - \hat{\boldsymbol{x}} \perp M$, or equivalently $\boldsymbol{x} - \hat{\boldsymbol{x}} \perp \boldsymbol{y}_j$ for $j = 1, \ldots, n$. Thus

$$0 = \langle \boldsymbol{x} - \hat{\boldsymbol{x}},\, \boldsymbol{y}_j \rangle = \left\langle \boldsymbol{x} - \sum_{i=1}^n \alpha_i \boldsymbol{y}_i,\ \boldsymbol{y}_j \right\rangle = \langle \boldsymbol{x},\, \boldsymbol{y}_j \rangle - \sum_{i=1}^n \alpha_i \langle \boldsymbol{y}_i,\, \boldsymbol{y}_j \rangle, \qquad j = 1, \ldots, n.$$

Rearranging yields the $n \times n$ system of linear equations for the scalar coefficients, called the **normal equations**:

$$\begin{bmatrix} \langle \boldsymbol{y}_1,\, \boldsymbol{y}_1 \rangle & \cdots & \langle \boldsymbol{y}_n,\, \boldsymbol{y}_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{y}_1,\, \boldsymbol{y}_n \rangle & \cdots & \langle \boldsymbol{y}_n,\, \boldsymbol{y}_n \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{x},\, \boldsymbol{y}_1 \rangle \\ \vdots \\ \langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle \end{bmatrix}.$$

If the $\boldsymbol{y}_i$'s are vectors in $\mathbb{C}^m$, with the usual inner product $\langle \boldsymbol{x},\, \boldsymbol{y} \rangle = \boldsymbol{y}'\boldsymbol{x}$, then defining the $m \times n$ matrix $\boldsymbol{Y} = [\boldsymbol{y}_1 \ \ldots \ \boldsymbol{y}_n]$ we have

$$\boldsymbol{\alpha} = (\boldsymbol{Y}'\boldsymbol{Y})^{-1} \boldsymbol{Y}'\boldsymbol{x}.$$

In particular, if $n = 1$, then $\hat{\boldsymbol{x}} = \langle \boldsymbol{x},\, \boldsymbol{y}/\,\|\boldsymbol{y}\| \rangle \boldsymbol{y}/\,\|\boldsymbol{y}\|$ (*cf.* the picture we draw).

Example. See previous polynomial approximation to $\arcsin(t)$.

Gram matrices

The $n \times n$ matrix above is called the **Gram matrix** of $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$.
Its determinant $g(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ is called the **Gram determinant**.

To find the best α_i's, we must solve the above system of equations, which has a unique solution iff the Gram determinant is nonzero.

Proposition. The Gram determinant is nonzero iff the vectors $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ are linearly independent.

Proof. If the $\boldsymbol{y}_i$'s are linearly dependent, then $\exists\ \alpha_i$'s not all zero such that $\sum_i \alpha_i \boldsymbol{y}_i = \mathbf{0}$.
Thus the rows of the Gram matrix are also linearly dependent, so the determinant is zero.

Conversely, if the determinant is zero, then the rows of the Gram matrix are linearly dependent, so $\exists\ \alpha_i$'s not all zero such that $\sum_i \alpha_i \langle \boldsymbol{y}_i,\, \boldsymbol{y}_j \rangle = 0$ for all j. Thus $\langle \sum_i \alpha_i \boldsymbol{y}_i,\, \boldsymbol{y}_j \rangle = 0$ for all j, so $\sum_j \alpha_j^* \langle \sum_i \alpha_i \boldsymbol{y}_i,\, \boldsymbol{y}_j \rangle = 0$. Thus $\|\sum_i \alpha_i \boldsymbol{y}_i\| = 0$ so $\sum_i \alpha_i \boldsymbol{y}_i = \mathbf{0}$ so the $\boldsymbol{y}_i$'s are linearly dependent. □

Remark. If the $\boldsymbol{y}_i$'s are linearly dependent, then there are multiple solutions to the normal equations, all of which are equally good approximations.
Often, at least in signal processing, of these many solutions we prefer the one that minimizes the Euclidean norm of $(\alpha_1, \ldots, \alpha_n)$.

However, no matter which solution for $\boldsymbol{\alpha}$ we choose, when added up via $\hat{\boldsymbol{x}} = \sum_{i=1}^n \alpha_i \boldsymbol{y}_i$ we will get *the same* $\hat{\boldsymbol{x}}$ since $\hat{\boldsymbol{x}}$ is unique by the projection theorem! Uniqueness of $\hat{\boldsymbol{x}}$ is different than uniqueness of $\boldsymbol{\alpha}$.

The text also describes the explicit error norm formula (which does not seem particularly useful):

$$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 = \frac{g(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n, \boldsymbol{x})}{g(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)}.$$

We see now that the approximation problem has an easily computed solution when M is a finite dimensional subspace.
We will see later in *3.10* that the same is true if $M^\perp$ is finite dimensional!

Orthogonal bases

What happens if the $\boldsymbol{y}_i$'s are orthonormal?
Then the Gram matrix is just the identity matrix, and we can immediately write down the optimal approximation:

$$\hat{\boldsymbol{x}} = \sum_{i=1}^n \alpha_i \boldsymbol{y}_i, \qquad \alpha_i = \langle \hat{\boldsymbol{x}},\, \boldsymbol{y}_i \rangle, \qquad \text{or equivalently } \hat{\boldsymbol{x}} = \sum_{i=1}^n P_{\boldsymbol{y}_i}(\boldsymbol{x}).$$

To generalize this result, we want to consider *infinite* dimensional approximating subspaces, since that is often of most interest in practice (*e.g.*, ordinary Fourier series).

3.9

Approximation and Fourier series

Returning to the case of finite-dimensional subspace M, we can find the minimum norm solution by first applying Gram-Schmidt to orthonormalize a (linearly independent) basis for M, and then use the Fourier series:

$$\hat{\boldsymbol{x}} = \sum_{i=1}^n \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle \boldsymbol{e}_i.$$

Thus applying Gram-Schmidt is equivalent to inverting the Gram matrix. Pick your poison...

Weighted least-squares FIR filter design

Example. Suppose we have a given desired frequency response $D(\omega)$ that we would like to approximate by a FIR filter with impulse response

$$h[n] = \sum_{k=0}^{M} h[k]\,\delta[n-k]$$

and corresponding frequency response

$$H(\omega) = \sum_{k=0}^{M} h[k]\,\mathrm{e}^{-\imath\omega k} \in \left[1, \mathrm{e}^{-\imath\omega}, \mathrm{e}^{-\imath 2\omega}, \ldots, \mathrm{e}^{-\imath M\omega}\right].$$

The natural inner product space here is $\mathcal{L}_2[-\pi,\pi]$ but with a *weighted* inner product

$$\langle H_1,\, H_2\rangle = \int_{-\pi}^{\pi} H_1(\omega) H_2^*(\omega) W(\omega)\,\mathrm{d}\omega,$$

where the positive weighting function $W(\omega)$ can influence which frequency bands require the closest match between $D(\omega)$ and $H(\omega)$ since the induced norm is

$$\|D - H\|^2 = \int_{-\pi}^{\pi} |D(\omega) - H(\omega)|^2\, W(\omega)\,\mathrm{d}\omega\,.$$

The Gram matrix $\boldsymbol{G}$ has elements

$$G_{kl} = \langle \mathrm{e}^{-\imath\omega k},\, \mathrm{e}^{-\imath\omega l}\rangle = \int_{-\pi}^{\pi} \mathrm{e}^{-\imath\omega(k-l)}\, W(\omega)\,\mathrm{d}\omega, \qquad k,l = 0,\ldots,M,$$

so the WLS optimal filter design has coefficients $\boldsymbol{h} = \boldsymbol{G}^{-1}\boldsymbol{d}$, where $\boldsymbol{d} = (d_0,\ldots,d_M)$ with

$$d_k = \langle D,\, \mathrm{e}^{-\imath\omega k}\rangle = \int_{-\pi}^{\pi} D(\omega)\,\mathrm{e}^{\imath\omega k}\, W(\omega)\,\mathrm{d}\omega, \qquad k = 0,\ldots,M.$$

Because the complex exponentials $\mathrm{e}^{-\imath\omega k}$ are linearly independent, the Gram matrix is invertible.

What if we want to minimize $\|D - H\|_\infty$ instead? Use Remez algorithm...

We now explore other minimum norm problems, in particular infinite dimensional ones where neither the normal equations nor the Fourier series solutions are applicable directly. In particular, we consider the broad family of problems involving linear varieties.

Linear varieties (from *2.3* and *3.10*)

Definition. A subset V of a vector space $\mathcal{X}$ is called a **linear variety** if $V = \boldsymbol{x}_0 + M$ for some $\boldsymbol{x}_0 \in \mathcal{X}$ and some subspace M of $\mathcal{X}$. Another term used is **affine subspace**.

Exercise. If $(\mathcal{X}, \mathcal{F})$ is a vector space and $V \subset \mathcal{X}$, then the following are equivalent.

- V is a linear variety.
- For any $\boldsymbol{x}_\star \in V$, the set $M = \{\boldsymbol{x} - \boldsymbol{x}_\star : \boldsymbol{x} \in V\}$ is a subspace.
- $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V,\ \alpha \in \mathcal{F}, \quad \alpha(\boldsymbol{x} - \boldsymbol{y}) + \boldsymbol{z} \in V$.

Example. In $\mathbb{R}^2$, consider $V = \{(a, b) : a + b = 1,\ a, b \in \mathbb{R}\}$. (A line *not* through the origin.)

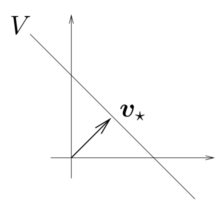
Exercise. If $V = \boldsymbol{x}_0 + M$, then V is complete / Chebyshev / closed iff M is complete / Chebyshev / closed.

- Any subspace is a linear variety (take $\boldsymbol{x}_0 = \mathbf{0}$), so linear varieties are a small generalization of subspaces.
- A single point $\boldsymbol{x}_0$ is a linear variety (take $M = \{\mathbf{0}\}$).

Fact. The point $\boldsymbol{x}_0$ need not be unique. If $V = \boldsymbol{x}_0 + M$ is a linear variety, then for any $\boldsymbol{v}_0 \in V$ we can write $V = \boldsymbol{v}_0 + M$, because $\boldsymbol{v}_0 = \boldsymbol{x}_0 + \boldsymbol{m}_0$ for some $\boldsymbol{m}_0 \in M$, so if $\boldsymbol{v} \in V$ then $\boldsymbol{v} = \boldsymbol{x}_0 + \boldsymbol{m} = \boldsymbol{v}_0 - \boldsymbol{m}_0 + \boldsymbol{m} = \boldsymbol{v}_0 + (\boldsymbol{m} - \boldsymbol{m}_0)$ where $\boldsymbol{m} - \boldsymbol{m}_0 \in M$.

In a "variety" of problems one would like to find the $\boldsymbol{x}$ having minimum norm (*e.g.*, minimum energy) subject to certain constraints.

Example. Continuing the previous example, the following figure shows the $\boldsymbol{v}_\star \in V$ with minimum $\|\cdot\|_2$ norm.



3.10

Dual approximation problem

The following theorem shows that such problems always have a unique solution and characterizes that solution.

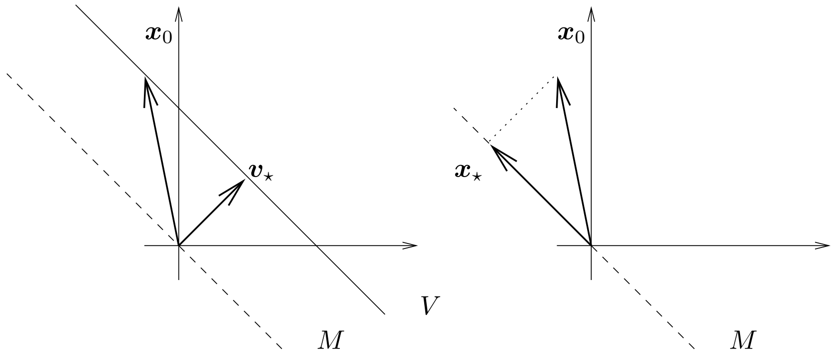
Theorem. *(Modified projection theorem.)* *Let M be a Chebyshev subspace of an inner product space $\mathcal{X}$. If $V = \boldsymbol{x}_0 + M$ is a linear variety where $\boldsymbol{x}_0 \in \mathcal{X}$, then there exists a unique $\boldsymbol{v}_\star \in V$ having minimum norm, and $\boldsymbol{v}_\star$ is characterized completely by the two conditions $\boldsymbol{v}_\star \in V$ and $\boldsymbol{v}_\star \perp M$.*

Proof. Simply translate V by $-\boldsymbol{x}_0$ and apply the projection theorem:

$$\inf_{\boldsymbol{v} \in V} \|\boldsymbol{v}\| = \inf_{\boldsymbol{m} \in M} \|\boldsymbol{x}_0 + \boldsymbol{m}\| = \inf_{\boldsymbol{m} \in M} \|\boldsymbol{x}_0 - \boldsymbol{m}\| = \|\boldsymbol{x}_0 - \boldsymbol{x}_\star\| \text{ where } \boldsymbol{x}_\star = \mathrm{P}_M(\boldsymbol{x}_0) \text{ and } \boldsymbol{x}_0 - \boldsymbol{x}_\star \perp M.$$

So use $\boldsymbol{v}_\star \triangleq \boldsymbol{x}_0 - \boldsymbol{x}_\star = \mathrm{P}_M^\perp(\boldsymbol{x}_0) \in V$ and $\boldsymbol{v}_\star \perp M$ and $\boldsymbol{v}_\star$ has minimum norm in V. □

Remark. Note that $\boldsymbol{v}_\star \perp M$, not $\boldsymbol{v}_\star \perp V$, *cf.* preceding figure.



Why called **dual**? Perhaps:

$$\boldsymbol{x}_\star = \arg\min_{\boldsymbol{m}\in M} \|\boldsymbol{x}_0 - \boldsymbol{m}\| = \boldsymbol{x}_0 - \boldsymbol{v}_\star \text{ where } \boldsymbol{v}_\star = \arg\min_{\boldsymbol{v}\in V=\boldsymbol{x}_0+M} \|\boldsymbol{v}\|.$$

Exercise. Generalize to the problem $\arg\min_{\boldsymbol{v}\in V} \|\boldsymbol{x}-\boldsymbol{v}\|$ if possible. **??**

One important application of linear varieties is in the study of problems with **constraints**.

In particular, the projection theorem led to the very convenient normal equations for the case of finite-dimensional subspaces. But there are also problems where something akin to the normal equations still apply even though the problem appears to be infinite dimensional.

Example. The set V in the previous example could be written $V = \left\{\boldsymbol{x} \in \mathbb{R}^2 : \langle \boldsymbol{x}, \boldsymbol{y}\rangle = 1\right\}$, where $\boldsymbol{y} = (1,1)$.

The following proposition shows that such sets are always linear varieties.

Proposition. Let $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ be a finite set of *linearly independent* vectors in an inner product space $\mathcal{X}$.
Then for given scalars $c_1, \ldots, c_n$, the following set is a closed linear variety:

$$V = \{\boldsymbol{x} \in \mathcal{X} : \langle \boldsymbol{x}, \boldsymbol{y}_i\rangle = c_i\}.$$

Moreover, there exists a unique $\boldsymbol{x}_0 \in [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ such that $V = \boldsymbol{x}_0 + [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\perp$.

Proof.
The $\boldsymbol{y}_i$'s are linearly independent, so the Gram matrix is nonsingular, and there is a unique $\boldsymbol{x}_0 \in [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ such that $\langle \boldsymbol{x}_0, \boldsymbol{y}_i\rangle = c_i$ for $i = 1, \ldots, n$. Thus V is nonempty and consists at least of this single point $\boldsymbol{x}_0$.

Claim. $V = \boldsymbol{x}_0 + [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\perp$.

$$\begin{aligned}
\boldsymbol{x} \in V \quad &\iff \quad \langle \boldsymbol{x}, \boldsymbol{y}_i\rangle = c_i,\ i = 1, \ldots, n \\
&\iff \quad \langle \boldsymbol{x} - \boldsymbol{x}_0, \boldsymbol{y}_i\rangle = \langle \boldsymbol{x}, \boldsymbol{y}_i\rangle - \langle \boldsymbol{x}_0, \boldsymbol{y}_i\rangle = c_i - c_i = 0,\ i = 1, \ldots, n \\
&\iff \quad \boldsymbol{x} - \boldsymbol{x}_0 \perp [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n] \iff \boldsymbol{x} - \boldsymbol{x}_0 \in [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\perp.
\end{aligned}$$

Thus $V = \boldsymbol{x}_0 + [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\perp$. Recall that orthogonal complements are closed, so $[\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\perp$ is closed.
(We did not use completeness to show this, rather just the continuity of the inner product!) $\therefore V$ is closed by a preceding Exercise.

Suppose $V = \boldsymbol{x}_1 + [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\perp$. Then $\boldsymbol{x}_0 \in V \Longrightarrow \boldsymbol{x}_0 = \boldsymbol{x}_1 + \boldsymbol{n}$ where $\boldsymbol{n} \in [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\perp$. Thus $\langle \boldsymbol{x}_0, \boldsymbol{y}_i\rangle = \langle \boldsymbol{x}_1 + \boldsymbol{n}, \boldsymbol{y}_i\rangle = \langle \boldsymbol{x}_1, \boldsymbol{y}_i\rangle = c_i$ for $i = 1 \ldots, n$, so $\boldsymbol{x}_1 = \boldsymbol{x}_0$ by the uniqueness discussed above. □

Remark. In general a linear variety V can be infinite dimensional. However, for the specific type of V in the above proposition, we say V has **codimension** n since the orthogonal complement of the underlying subspace has dimension n.

Applications

Two types of linear varieties are of particular interest in optimization problems.

- $V = \{\boldsymbol{x} + \sum_{i=1}^n c_i \boldsymbol{x}_i : c_i \in \mathbb{R}\}$, where the $\boldsymbol{x}_i$'s are linearly independent
- $V = \{\boldsymbol{x} \in \mathcal{X} : \langle \boldsymbol{x}, \boldsymbol{y}_i \rangle = c_i, \; i = 1, \ldots, n\}$

Both reduce to finite dimensional problems thanks to the projection theorem.

Theorem. *Let $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ be a set of linearly independent vectors in an inner product space $\mathcal{X}$. Let*

$$V = \{\boldsymbol{x} \in \mathcal{X} : \langle \boldsymbol{x}, \boldsymbol{y}_i \rangle = c_i, \; i = 1, \ldots, n\}.$$

Then there exists a unique $\boldsymbol{v}_\star \in V$ with minimum norm. Moreover, $\boldsymbol{v}_\star = \sum_{i=1}^n \beta_i \boldsymbol{y}_i$, where the β_i's satisfy the normal equations

$$\begin{bmatrix} \langle \boldsymbol{y}_1, \boldsymbol{y}_1 \rangle & \cdots & \langle \boldsymbol{y}_n, \boldsymbol{y}_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{y}_1, \boldsymbol{y}_n \rangle & \cdots & \langle \boldsymbol{y}_n, \boldsymbol{y}_n \rangle \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Remark. Note that V is necessarily nonempty by the previous Proposition, due to the linear independence.

Proof. From the previous proposition, $V = \boldsymbol{x}_0 + M^\perp$, where $M = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ and $\boldsymbol{x}_0 \in M$.
Being finite dimensional, M is Chebyshev, so $M^\perp$ is also Chebyshev.
So existence of a unique minimizing $\boldsymbol{v}_\star$ follows from the modified projection theorem.
Likewise, $\boldsymbol{v}_\star \perp M^\perp$ follows from that theorem. Thus $\boldsymbol{v}_\star \in M^{\perp\perp} = M$ since $M^\perp$ is Chebyshev.
Since $\boldsymbol{v}_\star \in M$, we have $\boldsymbol{v}_\star = \sum_{j=1}^n \beta_j \boldsymbol{y}_i$ for some β's.
We also need $\boldsymbol{v}_\star \in V$, *i.e.*, $\boldsymbol{v}_\star$ must satisfy the constraints $\langle \boldsymbol{v}_\star, \boldsymbol{y}_i \rangle = c_i$ or equivalently

$$\left\langle \sum_{j=1}^n \beta_j \boldsymbol{y}_j, \, \boldsymbol{y}_i \right\rangle = \sum_{j=1}^n \beta_j \langle \boldsymbol{y}_j, \boldsymbol{y}_i \rangle = c_i, \; i = 1, \ldots, n,$$

leading to the normal equations. □

Remark. Combining the projection theorem and the derivation of the normal equations yields the following theorem, which should be contrasted with the previous theorem.

Theorem. *((3.10-2) Really just a corollary to projection theorem.)*
If $M = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ is a finite-dimensional subspace of an inner product space $\mathcal{X}$, then given $\boldsymbol{x} \in \mathcal{X}$, there exists a unique $\boldsymbol{x}_\star \in M$ s.t. $\|\boldsymbol{x} - \boldsymbol{x}_\star\| = \inf_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\|$. Furthermore, $\boldsymbol{x} - \boldsymbol{x}_\star \perp M$, and $\boldsymbol{x}_\star = \sum_{i=1}^n \alpha_i \boldsymbol{y}_i$ where

$$\begin{bmatrix} \langle \boldsymbol{y}_1, \boldsymbol{y}_1 \rangle & \cdots & \langle \boldsymbol{y}_n, \boldsymbol{y}_1 \rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{y}_1, \boldsymbol{y}_n \rangle & \cdots & \langle \boldsymbol{y}_n, \boldsymbol{y}_n \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{x}, \boldsymbol{y}_1 \rangle \\ \vdots \\ \langle \boldsymbol{x}, \boldsymbol{y}_n \rangle \end{bmatrix}.$$

As the nice figure at the bottom of p. 67 shows, if either M or $M^\perp$ is finite dimensional, then minimum norm problems reduce to a finite set of linear equations.

Skim 3.11: control problem example

Example. (See [4, p. 66].)

Consider the linear system

$$y(t) = y(0) + \int_0^t h(t,\tau) u(\tau) \, d\tau$$

where $h \in C([0,T] \times [0,T])$,
i.e., $h(t,\tau)$ is a real-valued function that is continuous on $[0,T] \times [0,T]$,

Problem: find $u(\cdot)$ that minimizes $\int_0^T |u(t)|^2 \, dt$ subject to $y(T) = y_f$.
This is a minimum energy control problem.

Solution. Let $\mathcal{H} = \mathcal{L}_2[0,T]$ with $\langle x, y \rangle = \int_0^T x(t) y(t) \, dt$.

Now $h(T,\cdot) \in \mathcal{L}_2[0,T]$ since $\phi(t) = h(T,t)$ is a continuous function on the compact set $[0,T]$.

Thus

$$y_f = y(0) + \int_0^T h(T,\tau) u(\tau) \, d\tau = y(0) + \langle h(T,\cdot), u(\cdot) \rangle$$

so we have the following constraint set (with codimension = 1):

$$V = \{u \in \mathcal{L}_2[0,T] : y(t) = y_f\} = \Big\{ u \in \mathcal{L}_2[0,T] : \underbrace{y_f - y(0)}_{\text{“}c_1\text{”}} = \langle \underbrace{h(T,\cdot)}_{\text{“}\boldsymbol{y}_1\text{”}}, u(\cdot) \rangle \Big\}.$$

So in function space notation our problem is

$$\min_{u \in V} \|\boldsymbol{u}\| .$$

By a previous theorem, there is a unique solution $u^\star(t)$ that satisfies

$$u^\star(t) = \beta h(T,t), \text{ where } \langle h(T,\cdot), h(T,\cdot) \rangle \beta = y_f - y(0) \qquad \text{(normal equation)},$$

so the general solution is

$$u^\star(t) = \frac{y_f - y(0)}{\int_0^T h^2(T,\tau) \, d\tau} h(T,t).$$

Interestingly, the system itself, through $h(T,t)$, determines the *form* of the solution; this particular constraint affects only a scale factor.

3.7

Fourier series

Recall that an **infinite series** of the form $\sum_{i=1}^{\infty} \boldsymbol{x}_i$ is said to **converge** to $\boldsymbol{x}$ in a normed space iff the sequence of partial sums $\boldsymbol{s}_n = \sum_{i=1}^{n} \boldsymbol{x}_i$ converges to $\boldsymbol{x}$, in which case we write $\boldsymbol{x} = \sum_{i=1}^{\infty} \boldsymbol{x}_i$ as short hand for $\boldsymbol{x} = \lim_{n\to\infty} \sum_{i=1}^{n} \boldsymbol{x}_i$.

Lemma. If $\{\boldsymbol{e}_i\}$ is an **orthonormal sequence** in an inner product space, then

$$\left\| \sum_{i=1}^{n} c_i \boldsymbol{e}_i \right\|^2 = \sum_{i=1}^{n} |c_i|^2, \qquad \forall c_i \in \mathcal{F}, \quad \forall n \in \mathbb{N}.$$

This lemma is a form of **Parseval's relationship**.

The following theorem gives necessary and sufficient conditions for convergence of an infinite series of orthogonal vectors.

Theorem. *If $\{\boldsymbol{e}_i\}$ is an orthonormal sequence in a Hilbert space $\mathcal{H}$, then a series of the form $\sum_{i=1}^{\infty} c_i \boldsymbol{e}_i$ converges to some $\boldsymbol{x} \in \mathcal{H}$ iff $\sum_{i=1}^{\infty} |c_i|^2 < \infty$. In that case, we have $c_i = \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle$.*

Proof. Let $\boldsymbol{s}_n = \sum_{i=1}^{n} c_i \boldsymbol{e}_i$ and $\beta_n = \sum_{i=1}^{n} |c_i|^2$. Note $\{\beta_n\}$ is an increasing sequence.

Claim. $\{\boldsymbol{s}_n\}$ is Cauchy iff $\sum_{i=1}^{\infty} |c_i|^2 < \infty$.

$$\|\boldsymbol{s}_n - \boldsymbol{s}_m\|^2 = \left\| \sum_{i=n+1}^{m} c_i \boldsymbol{e}_i \right\|^2 = \sum_{i=n+1}^{m} |c_i|^2 = |\beta_n - \beta_m|.$$

So $\{\boldsymbol{s}_n\}$ is Cauchy $\iff$ $\{\beta_n\}$ is Cauchy $\iff$ $\{\beta_n\}$ converges in $\mathbb{R}$ (since $\mathbb{R}$ is complete) $\iff$ $\sum_{i=1}^{\infty} |c_i|^2 < \infty$.

Since $\mathcal{H}$ is complete, when $\{\boldsymbol{s}_n\}$ is Cauchy it converges to some limit $\boldsymbol{x} \in \mathcal{H}$ s.t. $\boldsymbol{x} = \sum_{i=1}^{\infty} c_i \boldsymbol{e}_i$.

Now $\langle \boldsymbol{s}_n,\, \boldsymbol{e}_i \rangle = c_i$, so by continuity of the inner product:

$$\langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle = \left\langle \lim_{n\to\infty} \boldsymbol{s}_n,\, \boldsymbol{e}_i \right\rangle = \lim_{n\to\infty} \langle \boldsymbol{s}_n,\, \boldsymbol{e}_i \rangle = \lim_{n\to\infty} c_i = c_i.$$

□

This "continuity of the inner product" technique is very useful in such proofs.

The $c_i = \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle$ values are called the **Fourier coefficients** of $\boldsymbol{x}$ w.r.t. $\{\boldsymbol{e}_i\}$.

Remark. The above theorem does *not* yet ensure that *any* $\boldsymbol{x}$ in $\mathcal{H}$ can be written as a Fourier series. That will come soon though.

Lemma. (**Bessel's inequality**)
If $\boldsymbol{x}$ is an element of an inner product space and $\{\boldsymbol{e}_i\}$ is an orthonormal sequence in that space, then

$$\sum_{i=1}^{\infty} |\langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle|^2 \leq \|\boldsymbol{x}\|^2 .$$

Proof. Let $c_i = \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle$.

$$0 \leq \left\| \boldsymbol{x} - \sum_{i=1}^{n} c_i \boldsymbol{e}_i \right\|^2 = \left\langle \boldsymbol{x} - \sum_{i=1}^{n} c_i \boldsymbol{e}_i,\, \boldsymbol{x} - \sum_{i=1}^{n} c_i \boldsymbol{e}_i \right\rangle = \|\boldsymbol{x}\|^2 - \sum_{i=1}^{n} |c_i|^2,$$

so $\forall n \in \mathbb{N},\ \sum_{i=1}^{n} |c_i|^2 \leq \|\boldsymbol{x}\|^2$, so $\sum_{i=1}^{\infty} |c_i|^2 \leq \|\boldsymbol{x}\|^2$. □

Remark. Bessel's inequality guarantees that in a Hilbert space $\mathcal{H}$, there exists $\hat{\boldsymbol{x}} \in \mathcal{H}$ such that $\hat{\boldsymbol{x}} = \sum_{i=1}^{\infty} \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle \boldsymbol{e}_i$, thanks to the preceding theorem.

Now we need to characterize $\hat{\boldsymbol{x}}$.

Theorem. *If $\boldsymbol{x}$ is an element of a Hilbert space $\mathcal{H}$, and $\{\boldsymbol{e}_i\}$ is an orthonormal sequence in $\mathcal{H}$, then*

$$\hat{\boldsymbol{x}} \triangleq \sum_{i=1}^{\infty} \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle \boldsymbol{e}_i \in M \triangleq \overline{[\{\boldsymbol{e}_i\}_{i=1}^{\infty}]},$$

which is called the ***closed subspace generated by*** *the $\boldsymbol{e}_i$'s. Furthermore, $\boldsymbol{x} - \hat{\boldsymbol{x}} \perp M$.*

Why do we need a closure above? **??**

Proof. Convergence of the series follows from Bessel's inequality and the preceding theorem.

Clearly $\hat{\boldsymbol{x}} \in M$ since $\hat{\boldsymbol{x}}$ is the limit of partial sums $\boldsymbol{s}_n = \sum_{i=1}^{n} c_i \boldsymbol{e}_i \in M$, where $c_i = \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle$.

By continuity of the inner product:

$$\langle \boldsymbol{x} - \hat{\boldsymbol{x}},\, \boldsymbol{e}_i \rangle = \langle \boldsymbol{x} - \lim_{n\to\infty} \boldsymbol{s}_n,\, \boldsymbol{e}_i \rangle = \lim_{n\to\infty} \langle \boldsymbol{x} - \boldsymbol{s}_n,\, \boldsymbol{e}_i \rangle = \lim_{n\to\infty} c_i - c_i = 0,$$

so $\boldsymbol{x} - \hat{\boldsymbol{x}} \perp [\{\boldsymbol{e}_i\}_{i=1}^{\infty}]$. Using (f) of proposition on orthogonal complements, we conclude $\boldsymbol{x} - \hat{\boldsymbol{x}} \perp \overline{[\{\boldsymbol{e}_i\}_{i=1}^{\infty}]}$. □

Corollary. If M is a closed subspace of a Hilbert space $\mathcal{H}$ and $\{\boldsymbol{e}_i\}$ is an orthonormal sequence such that $M = \overline{[\{\boldsymbol{e}_i\}_{i=1}^{\infty}]}$, then $P : \mathcal{H} \to M$, the orthogonal projection, is given by

$$P(\boldsymbol{x}) = \sum_{i=1}^{\infty} \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle \boldsymbol{e}_i .$$

Now the key question is when is $\overline{[\{\boldsymbol{e}_i\}_{i=1}^{\infty}]} = \mathcal{H}$? When the closed subspace generated by the $\boldsymbol{e}_i$'s is all of $\mathcal{H}$, then we can expand *any* vector in $\mathcal{H}$ as a series of the $\boldsymbol{e}_i$'s with the Fourier coefficients.

3.8

Complete orthonormal sequences / countable orthonormal bases

See "review of bases".

Definition. An orthonormal sequence $\{\boldsymbol{e}_i\}$ in a Hilbert space $\mathcal{H}$ is called **complete** (Luenberger) or a **countable orthonormal basis** (Naylor and others) iff the closed subspace generated by the $\boldsymbol{e}_i$'s is $\mathcal{H}$, *i.e.*, iff $\mathcal{H} = \overline{[\{\boldsymbol{e}_i\}_{i=1}^\infty]}$.

Lemma. If $\{\boldsymbol{e}_i\}$ is an orthonormal sequence in a Hilbert space $\mathcal{H}$, then the following are equivalent.
- $\{\boldsymbol{e}_i\}$ is complete.
- The only vector that is orthogonal to all of the $\boldsymbol{e}_i$'s is the zero vector, *i.e.*, $[\bigcup_{i=1}^\infty \{\boldsymbol{e}_i\}]^\perp = \{\boldsymbol{0}\}$
- Every vector in $\mathcal{H}$ has a Fourier series representation, *i.e.*, $\forall \boldsymbol{x} \in \mathcal{H},\ \boldsymbol{x} = \sum_{i=1}^\infty \langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle \boldsymbol{e}_i$.
- $\forall \boldsymbol{x} \in \mathcal{H}$, $\|\boldsymbol{x}\|^2 = \sum_{i=1}^\infty |\langle \boldsymbol{x},\, \boldsymbol{e}_i \rangle|^2$, which is called **Parseval's equality** [23, p. 24].

Proof. (Left to reader)

When does a Hilbert space $\mathcal{H}$ have a countable orthonormal basis? When (and only when) $\mathcal{H}$ is **separable** [23, p 21].
For a (complicated) example of a nonseparable Hilbert space, see [3, p. 230].

Example. Completeness of the orthogonal polynomials in $\mathcal{L}_2[-1, 1]$.

Sketch of proof. ***see text***
It suffices to show that the only function that is orthogonal to all the polynomials is the zero function.
Suppose $f \perp t^n$ for all $n = 0, 1, \ldots$ for some $f \in \mathcal{L}_2[-1, 1]$.
Then its integral F (which is continuous) is also orthogonal to the polynomials, by integration by parts.
Use Weierstrass and Cauchy-Schwarz to show that $\|F\|$ can be made arbitrarily small by choosing a sufficiently accurate polynomial approximation to F. So F must be zero, and hence f must be zero a.e.

Example. Completeness of the complex exponentials $e_k(t) = e^{ikt}/\sqrt{2\pi}$ for $k = 0, \pm 1, \pm 2, \ldots$ in $\mathcal{L}_2[0, 2\pi]$.

see text

Caution. Equality in $\mathcal{L}_2$ is meant in the $\mathcal{L}_2$ sense, *i.e.*, $\boldsymbol{x} = \boldsymbol{y}$ means $x(t) = y(t)$ a.e.

Remark. Completeness of the complex exponentials does not contradict Gibbs phenomena for truncated Fourier series. Completeness in $\mathcal{L}_2$ implies that $\|\boldsymbol{x} - \boldsymbol{s}_n\|_2 \to 0$, but it can still be the case that $\|\boldsymbol{x} - \boldsymbol{s}_n\|_\infty$ does not go to zero, and indeed it does not for functions with discontinuities. There is a big difference between convergence in integrated squared error and pointwise convergence.

Remark. The Fourier series coefficients of $x(t) = 1_{\{t \in \mathbb{Q}\}}$ are all zero.

Wavelets

What major post-1969 topic is missing here? Wavelets, of course.

There are both orthonormal wavelets and non-orthogonal wavelets.

A set $\{\boldsymbol{y}_n\} \in \mathcal{H}$ (a Hilbert space) is called a **frame** iff $\exists A > 0,\ B < \infty$ such that

$$A\,\|\boldsymbol{x}\|^2 \le \sum_n |\langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle|^2 \le B\,\|\boldsymbol{x}\|^2\,, \qquad \forall \boldsymbol{x} \in \mathcal{H}.$$

If the **frame bounds** A and B are equal, then we call the frame **tight**.

In a tight frame [23, p. 27]

$$\begin{aligned} \|\boldsymbol{x}\|^2 &= \frac{1}{A} \sum_n |\langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle|^2 \\ \boldsymbol{x} &= \frac{1}{A} \sum_n \langle \boldsymbol{x},\, \boldsymbol{y}_n \rangle \boldsymbol{y}_n. \end{aligned}$$

Despite how similar this last expression looks to the Fourier series representation for an orthonormal basis, the $\boldsymbol{y}_n$'s here need not be orthogonal, and in fact may be linearly dependent [23, p. 27,320], in which case we call it an **overcomplete expansion**.

Bases

Vector spaces

Definition. A *finite* sum $\sum_{i=1}^n \alpha_i \boldsymbol{x}_i$ for $\boldsymbol{x}_i \in \mathcal{X}$ and $\alpha_i \in \mathcal{F}$ is called a **linear combination**.

Definition. If S is a *subset* of a vector space $\mathcal{X}$, then the **span** of S is the subspace of linear combinations drawn from S:

$$[S] = \left\{ \boldsymbol{x} \in \mathcal{X} : \boldsymbol{x} = \sum_{i=1}^n \alpha_i \boldsymbol{x}_i, \text{ for } \boldsymbol{x}_i \in S, \ \alpha_i \in \mathcal{F}, \text{ and } n \in \mathbb{N} \right\}.$$

Definition. A set S of vectors is called **linearly independent** iff each vector in the set is linearly independent of the others, *i.e.*,

$$\forall \boldsymbol{x} \in S, \quad \boldsymbol{x} \notin [S - \{\boldsymbol{x}\}].$$

S can even be uncountable.

Definition. A set S is called a **Hamel basis** [3, p 183] for $\mathcal{X}$ iff S is linearly independent and $[S] = \mathcal{X}$.
Luenberger says "finite set" but Naylor [3, p. 183] and Maddox [2, p. 74] do not. Let us agree to use the above definition, rather than Luenberger's.

If S is linearly independent set in a vector space $\mathcal{X}$, then there exists a basis B for $\mathcal{X}$ such that $S \subseteq B$ [3, p 184].
Thus, every vector space has a Hamel basis, (since the empty set is linearly independent).

However, "Hamel basis is not the only concept of basis that arises in analysis. There are concepts of basis that involve topological as well as linear structure. ... In applications involving infinite-dimensional spaces, a useful basis, if one even exists, is usually something other than a Hamel basis. For example, a complete orthonormal set is far more useful in an infinite-dimensional Hilbert space than a Hamel basis" [3, p. 183].

Normed spaces

A (Hamel) basis for a Banach space is either finite or uncountably infinite [3, p. 218].

What is the closure of a span?

- If S is finite, then $[S]$ is finite-dimensional so $\overline{[S]} = [S]$.
- If S is countably infinite, *i.e.*, $S = \bigcup_{i=1}^\infty \boldsymbol{x}_i$, then $\overline{[S]}$ contains (at least) all the convergent series formed from S:

$$\overline{[S]} \supset \left\{ \sum_{i=1}^\infty \alpha_i \boldsymbol{x}_i : \alpha_i \in \mathcal{F}, \ \sum_{i=1}^\infty \alpha_i \boldsymbol{x}_i \text{ is convergent} \right\}.$$

  An example where $\overline{[S]}$ contains more than its convergent series is given in Naylor [3, p. 316], where S is a countable collection of linearly independent vectors!

Definition. In a normed space, $\{\boldsymbol{x}_n\} \in \mathcal{X}$ is a **Schauder basis** for $\mathcal{X}$ iff for each $\boldsymbol{x} \in \mathcal{X}$, there exists a unique sequence $\{\lambda_n\}$ such that $\boldsymbol{x} = \sum_{n=1}^\infty \lambda_n \boldsymbol{x}_n$ [2, p. 98].

The famous Banach conjecture that every separable Banach space has a Schauder basis was shown to be incorrect by Elfon in 1973 [2, p. 100].

So for more satisfactory answers we need to turn to Hilbert spaces, which have better structure.

Inner product spaces

Any finite-dimensional inner product space has an orthonormal basis (via Gram-Schmidt).

Definition. An **orthonormal set** $S = \{\boldsymbol{x}_\alpha\}$ is **maximal** iff there is no unit vector $\boldsymbol{x}_0 \in \mathcal{X}$ such that $S \cup \boldsymbol{x}_0$ is an orthonormal set.

Definition. A maximal orthonormal set B in a Hilbert space $\mathcal{H}$ is called an **orthonormal basis**.

Theorem. *If $\{\boldsymbol{x}_n\}$ be an orthonormal set in a Hilbert space, then [3, p. 307]:*

$$\{\boldsymbol{x}_n\} \text{ is an orthonormal basis} \iff \boldsymbol{x} = \sum_n \langle \boldsymbol{x}, \boldsymbol{x}_n \rangle \boldsymbol{x}_n, \qquad \forall \boldsymbol{x} \in \mathcal{X}.$$

When $S = \{\boldsymbol{x}_n\}$ is an orthonormal basis in Hilbert space $\mathcal{H}$, we have

$$\mathcal{H} = \overline{[S]} = \left\{ \sum_{i=1}^{\infty} \alpha_i \boldsymbol{x}_i : \sum_{i=1}^{\infty} |\alpha_i|^2 < \infty \right\}.$$

Orthonormal sets can be countable or uncountable, and the previous theorem applies to both cases [3, p. 314].
In engineering we usually work in separable Hilbert spaces (notably ℓ_2 and $\mathcal{L}_2$).

Theorem. *A Hilbert space has a countable orthonormal basis iff it is separable [3, p. 314].*

In other words:

Luenberger's **complete orthonormal sequence** $\equiv$ Naylor's **countable orthonormal basis**

and Naylor's terminology is more common in signal processing literature, *e.g.*, Vetterli's wavelet book.

Naylor notes that he deliberately avoids the term "complete" to describe certain orthonormal sets.

Summary

In Hilbert spaces, we have two kinds of bases:
- **Hamel bases**, which are not very useful, (uncountable for infinite-dimensional spaces)
- **orthonormal bases**, which are extremely useful.

3.12

Minimum distance to a convex set

Thus far we have considered only minimum distances to subspaces and linear varieties. Many applications need broader sets.

Theorem.

- *Let K be a nonempty complete convex subset in an inner product space $\mathcal{X}$ (e.g., K may be a nonempty closed convex subset of a Hilbert space.) Then K is Chebyshev: for any $\boldsymbol{x} \in \mathcal{X}$, there exists a unique vector $\boldsymbol{k}_\star \in K$ such that*

$$\|\boldsymbol{x} - \boldsymbol{k}_\star\| = d(\boldsymbol{x}, K) = \inf_{\boldsymbol{k} \in K} \|\boldsymbol{x} - \boldsymbol{k}\| .$$

- *Let K be a convex Chebyshev subset in an inner product space $\mathcal{X}$. Then the projector P_K for K is characterized in the following necessary and sufficient sense: $\boldsymbol{k}_\star = \mathrm{P}_K(\boldsymbol{x}) \iff \mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle) \le 0$ for all $\boldsymbol{k} \in K$.*
- *Let K be a subset of an inner product space $\mathcal{X}$ with the property that for every $\boldsymbol{x} \in \mathcal{X}$, there exists a unique point $\boldsymbol{k}_\star$ such that $\mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle) \le 0$ for all $\boldsymbol{k} \in K$. Then K is Chebyshev and $\boldsymbol{k}_\star = \mathrm{P}_K(\boldsymbol{x})$.*

Proof.

First we prove existence. Let $\{\boldsymbol{k}_i\}$ be a sequence in K such that $\|\boldsymbol{x} - \boldsymbol{k}_i\| \to \delta = d(\boldsymbol{x}, K)$.

. . . .

Claim 1. $\{\boldsymbol{k}_i\}$ is Cauchy.

By the parallelogram law, $\|\boldsymbol{k}_i - \boldsymbol{k}_j\|^2 = \|(\boldsymbol{k}_i - \boldsymbol{x}) - (\boldsymbol{k}_j - \boldsymbol{x})\|^2 = 2\|\boldsymbol{k}_i - \boldsymbol{x}\|^2 + 2\|\boldsymbol{k}_j - \boldsymbol{x}\|^2 - 4\left\|\boldsymbol{x} - \frac{\boldsymbol{k}_i + \boldsymbol{k}_j}{2}\right\|^2$.

Since K is **convex**, $\frac{\boldsymbol{k}_i + \boldsymbol{k}_j}{2} \in K$, so $\left\|\boldsymbol{x} - \frac{\boldsymbol{k}_i + \boldsymbol{k}_j}{2}\right\|^2 \ge \delta^2$.

Thus $\|\boldsymbol{k}_i - \boldsymbol{k}_j\|^2 \le 2\|\boldsymbol{k}_i - \boldsymbol{x}\|^2 + 2\|\boldsymbol{k}_j - \boldsymbol{x}\|^2 - 4\delta^2 \to 0$ as $i, j \to \infty$.

Since $\{\boldsymbol{k}_i\}$ is Cauchy and K is **complete**, $\{\boldsymbol{k}_i\}$ converges to some $\boldsymbol{k}_\star \in K$.
By continuity of the norm, $\|\boldsymbol{x} - \boldsymbol{k}_\star\| = \lim_{i\to\infty} \|\boldsymbol{x} - \boldsymbol{k}_i\| = \delta$.

. . . .

Claim 2. $\boldsymbol{k}_\star$ is unique. (Proof by contradiction)

Suppose $\boldsymbol{k}_1 \in K$ also satisfies $\|\boldsymbol{x} - \boldsymbol{k}_1\| = \delta$. Then the sequence $\boldsymbol{k}_n = \begin{cases} \boldsymbol{k}_\star, & n \text{ even} \\ \boldsymbol{k}_1, & n \text{ odd} \end{cases}$ is Cauchy by the same argument used for Claim 1, so $\{\boldsymbol{k}_n\}$ is convergent, which can only happen if $\boldsymbol{k}_\star = \boldsymbol{k}_1$.

. . . .

Claim 3. $\mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle) \le 0$ for all $\boldsymbol{k} \in K$. (Proof by contradiction.)

Suppose $\exists \boldsymbol{k} \in K$ s.t. $\mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle) = \varepsilon > 0$. Define $\boldsymbol{k}_\alpha = \alpha \boldsymbol{k} + (1-\alpha)\boldsymbol{k}_\star \in K$ for $\alpha \in [0,1]$ since K is **convex**.
Define $f(\alpha) = \|\boldsymbol{x} - \boldsymbol{k}_\alpha\|^2$, so $f(0) = \delta^2$. Now

$$f(\alpha) = \|\boldsymbol{x} - \alpha\boldsymbol{k} - (1-\alpha)\boldsymbol{k}_\star\|^2 = \|(\boldsymbol{x} - \boldsymbol{k}_\star) - \alpha(\boldsymbol{k} - \boldsymbol{k}_\star)\|^2 = \delta^2 + \alpha^2\|\boldsymbol{k} - \boldsymbol{k}_\star\|^2 - 2\alpha\,\mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle)$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} f(\alpha)\Big|_{\alpha=0} = \alpha\delta^2 - 2\,\mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle)\Big|_{\alpha=0} = -2\varepsilon < 0.$$

Thus $\exists \alpha > 0$ s.t. $f(\alpha) < f(0) = \delta^2$, contradicting the minimizing norm property of $\boldsymbol{k}_\star$.

. . . .

Claim 4. If $\mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle) \le 0$ for some $\boldsymbol{k} \in K$, then $\|\boldsymbol{x} - \boldsymbol{k}_\star\| \le \|\boldsymbol{x} - \boldsymbol{k}\|$. (So "is characterized by" means "iff.")

$$\|\boldsymbol{x} - \boldsymbol{k}\|^2 = \|\boldsymbol{x} - \boldsymbol{k}_\star - (\boldsymbol{k} - \boldsymbol{k}_\star)\|^2 = \|\boldsymbol{x} - \boldsymbol{k}_\star\|^2 + \|\boldsymbol{k} - \boldsymbol{k}_\star\|^2 - 2\,\mathrm{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star,\, \boldsymbol{k} - \boldsymbol{k}_\star \rangle) \ge \|\boldsymbol{x} - \boldsymbol{k}_\star\|^2 .$$

□

Remark.

- The convexity of K was used for *both* the existence and characterization parts.
- However, Claim 4 did not use convexity, which explains the last item in the Theorem.
- In this case the characterization is an inequality, which is usually harder to work with.
- Although P_K exists since K is Chebyshev, we do not have a general formula for it other than the "characterization" inequality.
- Can we generalize further? No. In any finite-dimensional inner product space, K Chebyshev $\Longrightarrow$ K closed, convex, and nonempty.

Revisiting subspaces

What happens when K is in fact a **subspace**? Then for any $\boldsymbol{k}_0 \in K$ we can pick $\boldsymbol{k} = \boldsymbol{k}_\star - \boldsymbol{k}_0 \in K$ and $\boldsymbol{k} = \boldsymbol{k}_\star + \boldsymbol{k}_0 \in K$ to show that $\operatorname{real}(\langle \boldsymbol{x} - \boldsymbol{k}_\star, \boldsymbol{k}_0 \rangle) = 0$. For complex subspaces we can also pick $\boldsymbol{k} = \boldsymbol{k}_\star - \imath \boldsymbol{k}_0 \in K$ to show that the imaginary part is zero, to conclude that $\boldsymbol{x} - \boldsymbol{k}_\star \perp K$. Conversely, if $\boldsymbol{k}_\star \in K$ and $\boldsymbol{x} - \boldsymbol{k}_\star \perp K$, then $\langle \boldsymbol{x} - \boldsymbol{k}_\star, \boldsymbol{k} - \boldsymbol{k}_\star \rangle = 0$ for all $\boldsymbol{k} \in K$.
So we could have presented convex sets first, and then specialized to subspaces.

Example. See text p. 71 for a somewhat unsatisfying example involving nonnegativity constraints.

Example. In $\mathbb{R}^2$, consider the half plane $K = \left\{(a, b) \in \mathbb{R}^2 : a + b \leq 0\right\}$, which is a closed convex set.
By sketching this set and using geometric intuition, one can conjecture that the projector is given by

$$\mathrm{P}_K((a, b)) = (a - p, b - p), \text{ where } p = \left[\frac{a+b}{2}\right]_+, \text{ and } [x]_+ = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Note that $[x]_+ (x - [x]_+) = 0$.

To verify that the above projector is correct, we can check it against the characterization condition in the preceding Theorem.

If $\boldsymbol{x} = (a, b)$ and $\boldsymbol{k}_\star = \mathrm{P}_K(\boldsymbol{x})$ then

$$\begin{aligned}
\langle \boldsymbol{x} - \boldsymbol{k}_\star, \boldsymbol{k} - \boldsymbol{k}_\star \rangle &= \langle (a, b) - (a-p, b-p), (k_1, k_2) - (a-p, b-p) \rangle = \langle (p, p), (k_1 - a + p, k_2 - b + p) \rangle \\
&= 2p^2 - p(a+b) + p(k_1 + k_2) \\
&\leq 2p^2 - p(a+b) \text{ since } \boldsymbol{k} \in K \Longrightarrow k_1 + k_2 \leq 0 \\
&= 2p\left(p - \frac{a+b}{2}\right) = 0.
\end{aligned}$$

Thus the above projector is correct.

Projection onto convex sets (POCS)

In light of the previous theorem, if K is a nonempty closed convex subset of a Hilbert space $\mathcal{H}$, then K is Chebyshev and we can legitimately define a projection operator $\mathrm{P}_K : \mathcal{H} \to K$ by

$$\mathrm{P}_K(\boldsymbol{x}) = \arg\min_{\boldsymbol{k} \in K} \|\boldsymbol{x} - \boldsymbol{k}\|.$$

It is *not* an orthogonal projection in general, *i.e.*, $\boldsymbol{x} - \mathrm{P}_K(\boldsymbol{x}) \not\perp K$, (unless of course K happens to be a subspace).

Indeed, in general this projection inherits only the trivial properties of projectors given previously:
- $\mathrm{P}_K(\boldsymbol{x}) \in K$
- $\mathrm{P}_K(\mathrm{P}_K(\boldsymbol{x})) = \mathrm{P}_K(\boldsymbol{x})$
- $\|\boldsymbol{x} - \mathrm{P}_K(\boldsymbol{x})\| = d(\boldsymbol{x}, K)$

In addition, $\mathrm{P}_K(\cdot)$ is **continuous** [24] (homework problem).

There are a variety of convex sets of interest in signal and image processing problems, such as
- The subspace of band-limited signals with a given band-limit.
- The set of nonnegative signals.
- The subspace of signals with a given time or spatial support.

Because of such examples, POCS methods are (somewhat) useful in signal processing. A typical problem would be "find the signal with a given spatial support whose spectrum is given over certain frequency ranges only."

Example. In $\mathbb{R}^n$ consider $K = \{\boldsymbol{x} : x_j \geq 0, \ j = 1, \ldots, n\}$. Then if $\hat{\boldsymbol{k}} = P_K(\boldsymbol{x})$ we have $\hat{k}_j = \begin{cases} x_j, & x_j \geq 0 \\ 0, & \text{otherwise.} \end{cases}$
If $\boldsymbol{k} \in K$ then $\langle \boldsymbol{x} - \hat{\boldsymbol{k}}, \boldsymbol{k} - \hat{\boldsymbol{k}} \rangle = \sum_{j=1}^n (x_j - \hat{k}_j)(k_j - \hat{k}_j) = \sum_{j : x_j < 0} (x_j - 0)(k_j - 0) \leq 0$, as required, since $k_j \geq 0$.

??

Summary

- Inner products and properties
  - Cauchy-Schwarz
  - induced norm
  - parallelogram law
  - continuity
- Orthogonality
  - Pythagorean theorem
  - orthogonal complements
  - direct sum
  - orthogonal sets
  - Gram-Schmidt procedure
  - orthonormal bases
- Minimum norm problems
  - (orthogonal) projections onto subspaces
  - normal equations
  - Fourier series
  - complete orthonormal sequences (countable orthonormal bases)
  - minimum norm within linear variety (constraints)
  - minimum distance to convex sets

---

1. P. Enflo. A counterexample to the approximation problem in Banach spaces. *Acta Math*, 130:309–17, 1973.
2. I. J. Maddox. *Elements of functional analysis*. Cambridge, 2 edition, 1988.
3. A. W. Naylor and G. R. Sell. *Linear operator theory in engineering and science*. Springer-Verlag, New York, 2 edition, 1982.
4. D. G. Luenberger. *Optimization by vector space methods*. Wiley, New York, 1969.
5. J. Schauder. Zur theorie stetiger abbildungen in funktionenrumen. *Math. Zeitsch.*, 26:47–65, 1927.
6. A. M. Ostrowski. *Solution of equations in Euclidian and Banach spaces*. Academic, 3 edition, 1973.
7. R. R. Meyer. Sufficient conditions for the convergence of monotonic mathematical programming algorithms. *J. Comput. System. Sci.*, 12(1):108–21, 1976.
8. M. Rosenlicht. *Introduction to analysis*. Dover, New York, 1985.
9. A. R. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Tr. Med. Im.*, 12(2):328–33, June 1993.
10. A. R. De Pierro. On the convergence of the iterative image space reconstruction algorithm for volume ECT. *IEEE Tr. Med. Im.*, 6(2):174–175, June 1987.
11. A. R. De Pierro. Unified approach to regularized maximum likelihood estimation in computed tomography. In *Proc. SPIE 3171, Comp. Exper. and Num. Meth. for Solving Ill-Posed Inv. Imaging Problems: Med. and Nonmed. Appl.*, pages 218–23, 1997.
12. J. A. Fessler. Grouped coordinate descent algorithms for robust edge-preserving image restoration. In *Proc. SPIE 3170, Im. Recon. and Restor. II*, pages 184–94, 1997.
13. A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Tr. Med. Im.*, 14(1):132–137, March 1995.
14. J. A. Fessler and A. O. Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Tr. Im. Proc.*, 4(10):1417–29, October 1995.
15. M. W. Jacobson and J. A. Fessler. Properties of optimization transfer algorithms on convex feasible sets. *SIAM J. Optim.*, 2003. Submitted.
16. P. L. Combettes and H. J. Trussell. Method of successive projections for finding a common point of sets in metric spaces. *J. Optim. Theory Appl.*, 67(3):487–507, December 1990.
17. F. Deutsch. The convexity of Chebyshev sets in Hilbert space. In A. Yanushauskas Th. M. Rassias, H. M. Srivastava, editor, *Topics in polynomials of one and several variables and their applications*, pages 143–50. World Sci. Publishing, River Edge, NJ, 1993.
18. M. Jiang. On Johnson's example of a nonconvex Chebyshev set. *J. Approx. Theory*, 74(2):152–8, August 1993.
19. V. S. Balaganskii and L. P. Vlasov. The problem of convexity of Chebyshev sets. *Russian Mathematical Surveys*, 51(6):1127–90, November 1996.
20. V. Kanellopoulos. On the convexity of the weakly compact Chebyshev sets in Banach spaces. *Israel Journal of Mathematics*, 117:61–9, 2000.
21. A. R. Alimov. On the structure of the complements of Chebyshev sets. *Functional Analysis and Its Applications*, 35(3):176–82, July 2001.
22. Y. Bresler, S. Basu, and C. Couvreur. Hilbert spaces and least squares methods for signal processing, 2000. Draft.
23. M. Vetterli and J. Kovacevic. *Wavelets and subband coding*. Prentice-Hall, New York, 1995.
24. D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part I—Theory. *IEEE Tr. Med. Im.*, 1(2):81–94, October 1982.