ECE 586: Vector Space Methods
Lecture 17: Best Approximation

Henry D. Pfister
Duke University
Let $W$ be a subspace of a Banach space $V$ and, for any $v \in V$, consider finding a vector $w \in W$ such that $\|v - w\|$ is as small as possible.

**Definition**

The vector $w \in W$ is a **best approximation** of $v \in V$ by vectors in $W$ if

$$\|v - w\| \leq \|v - w'\|,$$

for all $w' \in W$.

**Example**

If $W$ is spanned by the vectors $w_1, \ldots, w_n \in V$, then we can write

$$v = w + e = s_1 w_1 + \cdots + s_n w_n + e,$$

where $e = v - w$ is the approximation error.
Vector Projection Revisited

Let \( \mathbf{u}, \mathbf{v} \) be vectors in an inner-product space \( V \) with inner product \( \langle \cdot, \cdot \rangle \).

**Lemma**

If \( \langle \mathbf{w}, \mathbf{v} \rangle = 0 \), then \( \| \mathbf{w} + \mathbf{v} \|^2 = \| \mathbf{w} \|^2 + 2 \text{Re}\{\langle \mathbf{w}, \mathbf{v} \rangle \} + \| \mathbf{v} \|^2 = \| \mathbf{w} \|^2 + \| \mathbf{v} \|^2 \).

**Definition (Vector Projection)**

The projection of \( \mathbf{w} \) onto \( \mathbf{v} \) is defined to be

\[
\mathbf{u} = \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\| \mathbf{v} \|^2} \mathbf{v}
\]

**Lemma**

Let \( \mathbf{u} \) be the projection of \( \mathbf{w} \) onto \( \mathbf{v} \). If \( \langle \mathbf{w}, \mathbf{v} \rangle \neq 0 \), then \( \| \mathbf{w} - \mathbf{u} \| < \| \mathbf{w} \| \).

**Proof.**

\( \langle \mathbf{w} - \mathbf{u}, \mathbf{u} \rangle = 0 \) implies \( \| \mathbf{w} \|^2 = \| (\mathbf{w} - \mathbf{u}) + \mathbf{u} \|^2 = \| \mathbf{w} - \mathbf{u} \|^2 + \| \mathbf{u} \|^2 \). □
4.1: Orthogonal Projection

In an arbitrary Banach space, finding a best approximation can be hard. For the induced norm of a Hilbert space, orthogonal projection simplifies this!

**Theorem (Projection)**

Suppose $W$ is a subspace of a Hilbert space $V$ and $v \in V$. Then,

1. The vector $w \in W$ is a best approximation of $v \in V$ by vectors in $W$ if and only if $v - w$ is orthogonal to every vector in $W$.

2. If a best approximation of $v \in V$ by vectors in $W$ exists, it is unique.

3. If $W$ is a closed subspace with a countable orthogonal basis $w_1, w_2, \ldots$, then the best approximation of $v$ by vectors in $W$ is

   $$w = \sum_{i=1}^{\dim(W)} \frac{\langle v, w_i \rangle}{\|w_i\|^2} w_i.$$ 

   Note: the implied linear mapping $T: V \to W$ defined by $T(v) = w$ is called the **orthogonal projection** of $V$ onto $W$.

Proof in live session.
Orthogonal Projection Example

Example

For the standard inner product space $V = \mathbb{R}^3$, let $W$ be the subspace spanned by

$$v_1 = (2, 2, 1),$$
$$v_2 = (3, 6, 0).$$

Then, the Gram-Schmidt process generates the orthogonal basis

$$w_1 = (2, 2, 1),$$
$$w_2 = (-1, 2, -2)$$

and the orthogonal projection of $v \in V$ onto $W$ is defined by

$$T_v = \sum_{i=1}^{2} \frac{\langle v, w_i \rangle}{\|w_i\|^2} w_i$$

$$= \frac{1}{9} \langle v, (2, 2, 1) \rangle (2, 2, 1) + \frac{1}{9} \langle v, (-1, 2, -2) \rangle (-1, 2, -2).$$
4.1.1: Orthogonal Projection onto an Orthonormal Set

Let $V = \mathbb{C}^n$ be the standard $n$-dimensional complex Hilbert space and $U \in \mathbb{C}^{n \times m}$ be a matrix with orthonormal columns $u_1, \ldots, u_m$:

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$$

Then, the best approximation of $v \in V$ by vectors in $\mathcal{R}(U)$, given by

$$w = \sum_{i=1}^{m} \frac{\langle v, u_i \rangle}{\|u_i\|^2} u_i,$$

can also be written as

$$w = U U^H v = \sum_{i=1}^{m} u_i (u_i^H v).$$
4.1.1: What is a Projection? (1)

**Definition**
A function $F : X \to Y$ with $Y \subseteq X$ is **idempotent** if $F(F(x)) = F(x)$. When $F$ is a linear transformation, this reduces to $F^2 = F \cdot F = F$.

**Definition**
Let $V$ be a vector space and $T : V \to V$ be a linear transformation. If $T$ is idempotent, then $T$ is called a **projection** because $T\mathbf{v} = \mathbf{v}$ if $\mathbf{v} \in \mathcal{R}(T)$.

**Example**
The idempotent matrix $A$ is a projection onto the first two coordinates.

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$
Theorem

Let $V$ be a vector space and $T : V \to V$ be a (linear) projection operator. Then, the range $\mathcal{R}(T)$ and the $\mathcal{N}(T)$ are disjoint subspaces of $V$.

Proof.

For a non-zero $v \in \mathcal{R}(T)$, there is a non-zero $w \in V$ such that $v = Tw$. Thus, $Tv = T^2w = Tw = v \neq 0$. But, if $v \in \mathcal{N}(T)$ was also true, then one would get the contradiction $Tv = 0$.

Example

Consider the linear transform $T : V \to V$ defined by $T = I - P$, where $P$ is a projection. It is easy to verify that $T$ is a projection operator because

$$T^2 = (I - P)(I - P) = I - P - P + P^2 = I - P = T.$$ 

In fact, $T$ is a projection onto $\mathcal{R}(T) = \mathcal{N}(P)$ because $Pv = 0$ (i.e., $v \in \mathcal{N}(P)$) if and only if $(I - P)v = v$ (i.e., $v \in \mathcal{R}(T)$).
4.1.1: Orthogonal Projection Operators

**Definition**

Let \( V \) be an inner-product space and \( P: V \to V \) be a (linear) projection operator. If \( \mathcal{R}(P) \perp \mathcal{N}(P) \), then \( P \) is called an orthogonal projection.

**Example**

Let \( V \) be an inner-product space \( P: V \to V \) be an orthogonal projection. Then, \( \mathbf{v} = P\mathbf{v} + (I - P)\mathbf{v} \) gives an orthogonal decomposition of \( \mathbf{v} \) because \( P\mathbf{v} \in \mathcal{R}(P), (I - P)\mathbf{v} \in \mathcal{N}(P) \), and \( \mathcal{R}(P) \perp \mathcal{N}(P) \).

**Theorem**

For \( V = F^n \) with standard inner product, \( P \) is an orthogonal projection matrix if it is idempotent and Hermitian (i.e. \( P^2 = P \) and \( P^H = P \)).

**Proof.**

Since \( \mathcal{R}(P) = \{ Pu | u \in V \} \) and \( \mathcal{N}(P) = \{ \mathbf{v} \in V | P\mathbf{v} = 0 \} \), the general condition is \( \langle Pu, (I - P)\mathbf{v} \rangle = 0 \) for all \( u, \mathbf{v} \in V \). Simplifying this gives

\[
\mathbf{v}^H(I - P)^H Pu = \mathbf{v}^H(P - P^H P)u = \mathbf{v}^H(P - P^2)u = 0.
\]

\( \square \)
To continue studying after this video –

- Try the required reading: Course Notes EF 4.1 - 4.1.1
- Or the recommended reading: LADR 6C
- Also, look at the problems in Assignment 7