Let $W$ be a subspace of a Hilbert space $V$ that is spanned by the linearly independent (but not orthogonal) set of vectors $w_1, \ldots, w_n \in V$.

The projection theorem shows that $\hat{v} \in W$ is the best approximation of $v \in V$ if and only if $(v - \hat{v}) \perp w_j$ for $j = 1, \ldots, n$. This implies that

$$\langle v - \hat{v}, w_j \rangle = \left\langle v - \sum_{i=1}^{n} s_i w_i, w_j \right\rangle = 0$$

or, equivalently, the normal equations

$$\sum_{i=1}^{n} s_i \langle w_i, w_j \rangle = \langle v, w_j \rangle.$$

The gives a system of $n$ linear equations in $n$ unknowns defined by

$$\begin{bmatrix}
\langle w_1, w_1 \rangle & \langle w_2, w_1 \rangle & \cdots & \langle w_n, w_1 \rangle \\
\langle w_1, w_2 \rangle & \langle w_2, w_2 \rangle & \cdots & \langle w_n, w_2 \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle w_1, w_n \rangle & \langle w_2, w_n \rangle & \cdots & \langle w_n, w_n \rangle
\end{bmatrix}
\begin{bmatrix}
s_1 \\
s_2 \\
\vdots \\
s_n
\end{bmatrix}
= 
\begin{bmatrix}
\langle v, w_1 \rangle \\
\langle v, w_2 \rangle \\
\vdots \\
\langle v, w_n \rangle
\end{bmatrix}.$$
4.2: The Gramian

**Definition**

For \( w_1, \ldots, w_n \), the \( n \times n \) Gramian matrix is defined to be

\[
G = \begin{bmatrix}
\langle w_1, w_1 \rangle & \langle w_2, w_1 \rangle & \cdots & \langle w_n, w_1 \rangle \\
\langle w_1, w_2 \rangle & \langle w_2, w_2 \rangle & \cdots & \langle w_n, w_2 \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle w_1, w_n \rangle & \langle w_2, w_n \rangle & \cdots & \langle w_n, w_n \rangle 
\end{bmatrix}
\]

Since \( g_{ij} = \langle w_j, w_i \rangle \), we see \( G \) is Hermitian symmetric (i.e. \( G^H = G \)).

**Definition**

A matrix \( M \in F^{n \times n} \) is positive-semidefinite if \( M = M^H \) and \( \mathbf{v}^H M \mathbf{v} \geq 0 \) for all \( \mathbf{v} \in F^n - \{0\} \). If the inequality is strict, then it is positive-definite.

**Theorem**

A Gramian matrix \( G \) is always positive-semidefinite. It is positive-definite if and only if the vectors \( w_1, \ldots, w_n \) are linearly independent.

Proof in live session.
4.3: Least-Squares Solution of a Linear System

For $V = F^m$, let $A \in F^{m \times n}$ be a matrix whose $i$-th column is $w_i \in V$. Then, a vector $\hat{v} \in W = \text{colspace}(A)$ can be written as

$$\hat{v} = As = \sum_{i=1}^{n} s_iw_i.$$ 

Also, the best approximation of $v$ by vectors in $W$ is found by solving

$$\min_{\hat{v} \in W} \|v - \hat{v}\| = \min_s \|v - As\|.$$ 

For the induced norm, any solution must satisfy the normal equations

$$\langle v - \hat{v}, w_j \rangle = \langle v - As, w_j \rangle = 0, \quad j \in [n].$$

For the standard inner product, these equations can be expressed as

$$0 = \begin{bmatrix} w_1^H \\ \vdots \\ w_n^H \end{bmatrix} (v - As) = A^Hv - A^HAS = t - GS,$$

where $G = A^HA$ is the Gramian and $t$ is the cross-correlation vector.
4.3.2: Pseudo-Inverse and Projection

When the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are linearly independent, the Gramian matrix is positive definite and hence invertible. Thus, the optimal solution for the least-squares problem is given by

$$\mathbf{s} = \mathbf{G}^{-1} \mathbf{t} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{v},$$

where the matrix $(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ is the pseudoinverse of $\mathbf{A}$ in this case.

Using this, the best approximation of $\mathbf{v} \in \mathcal{V}$ by vectors in $\mathcal{W}$ is equal to

$$\hat{\mathbf{v}} = \mathbf{A} \mathbf{s} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{v}.$$ 

The matrix $\mathbf{P} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ is the projection matrix for the range of $\mathbf{A}$. It defines an orthogonal projection onto the range of $\mathbf{A}$ (i.e., the subspace spanned by the columns of $\mathbf{A}$).
4.3.3: Weighted Least-Squares Solution of a Linear System

For the standard Euclidean norm \( \|v\|_E = \sqrt{v^H v} \) and any invertible \( B \), consider the weighted least-squares problem

\[
\min_{\hat{v} \in W} \|B(v - \hat{v})\|_E = \min_{\hat{s}} \|B(v - As)\|_E
\]

But, \( \|Bv\|_E \) equals the induced norm \( \|v\| \) for the weighted inner product

\[
\langle u, v \rangle \triangleq v^H B^H Bu.
\]

For the weighted inner product, the normal equations look the same

\[
\langle v - \hat{v}, w_j \rangle = \langle v - As, w_j \rangle = 0, \quad j \in [n].
\]

but they solve a different problem and they reduce to

\[
0 = \begin{bmatrix}
    w_1^H \\
    \vdots \\
    w_n^H
\end{bmatrix}
\]

\[
B^H B (v - As) = A^H B^H Bv - A^H B^H BA_s
\]
4.4.2: Linear Minimum Mean-Squared Error Estimation

Let $Y, X_1, \ldots, X_n$ be zero-mean random variables. Linear minimum mean-squared error (LMMSE) estimation finds $s_1, \ldots, s_n$ such that

$$\hat{Y} = s_1 X_1 + \cdots + s_n X_n$$

minimizes the mean squared-error $E[|Y - \hat{Y}|^2]$. Using the inner product

$$\langle X, Y \rangle = E[XY],$$

the normal equations for the LMMSE estimate $\hat{Y}$ are $G\mathbf{s} = \mathbf{t}$, where

$$G = \begin{bmatrix}
E[X_1X_1] & E[X_1X_2] & \cdots & E[X_1X_n] \\
E[X_2X_1] & E[X_2X_2] & \cdots & E[X_2X_n] \\
\vdots & \vdots & \ddots & \vdots \\
E[X_nX_1] & E[X_nX_2] & \cdots & E[X_nX_n]
\end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix}
E[YX_1] \\
E[YX_2] \\
\vdots \\
E[YX_n]
\end{bmatrix}.$$  

If $G$ is invertible, then $\mathbf{s} = G^{-1}\mathbf{t}$ implies $E[|\hat{Y}|^2] = \mathbf{s}^H G \mathbf{s} = \mathbf{s}^H \mathbf{t}$ and

$$\|Y\|^2 - \|\hat{Y}\|^2 = E[|Y|^2] - E[|\hat{Y}|^2] = E[|Y|^2] - \mathbf{s}^H \mathbf{t}.$$
An underdetermined system of linear equations has an infinite number of solutions. It often makes sense to prefer the minimum-norm solution.

Let $V$ be a Hilbert space and $w_1, w_2, \ldots, w_n$ be a basis for subspace $W$. For any $v \in V$, the best approximation of $v$ in $W$ can be found by solving

$$
\begin{bmatrix}
\langle w_1, w_1 \rangle & \langle w_2, w_1 \rangle & \cdots & \langle w_n, w_1 \rangle \\
\langle w_1, w_2 \rangle & \langle w_2, w_2 \rangle & \cdots & \langle w_n, w_2 \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle w_1, w_n \rangle & \langle w_2, w_n \rangle & \cdots & \langle w_n, w_n \rangle \\
\end{bmatrix} \begin{bmatrix}
s_1 \\
s_2 \\
\vdots \\
s_n \\
\end{bmatrix} = \begin{bmatrix}
\langle v, w_1 \rangle \\
\langle v, w_2 \rangle \\
\vdots \\
\langle v, w_n \rangle \\
\end{bmatrix}.
$$

(1)

Theorem

Let $V$ be a Hilbert space and $w_1, w_2, \ldots, w_n$ be a basis for $W \subseteq V$. The dual approximation problem is to find the minimum-norm vector $w \in V$ satisfying $\langle w, w_i \rangle = c_i$ for $i = 1, \ldots, n$. Then, the solution $w$ satisfies

$$w = \sum_{i=1}^{n} s_i w_i \in W,$$

where $s_1, s_2, \ldots, s_n$ can be found by solving (1) with $\langle v, w_i \rangle = c_i$. 
4.5.2: Minimum-Norm Solutions

For $A \in \mathbb{C}^{m \times n}$ with $m < n$ and $v \in \mathbb{C}^m$, consider the underdetermined linear system $As = v$. Then, the dual approximation theorem can be applied to solve the minimum-norm problem

$$\min_{s : As = v} \|s\|.$$

To see this as a dual approximation, we can rewrite the constraint $As = v$ as $B^H s = v$ where $B = A^H$. Then, the theorem concludes that the minimum-norm solution lies in the column space of $B = A^H$.

Using $s \in \mathcal{R}(A^H)$, there is a $t$ such that $\hat{s} = A^H t$ and the constraint gives $A(A^H t) = v$. If the rows of $A$ are linearly independent, then the columns of $B = A^H$ are linearly independent and $(B^H B)^{-1} = (AA^H)^{-1}$ exists.

Thus, the solution $\hat{s}$ can be obtained in closed form and is given by

$$\hat{s} = A^H (AA^H)^{-1} v.$$
Next Steps

To continue studying after this video –

- Try the required reading: Course Notes EF 4.2 - 4.3.4
- Or the recommended reading: LADR 6C
- Also, look at the problems in Assignment 7