ECE 586: Vector Space Methods
Lecture 18: Best Approximation Formulas

Henry D. Pfister
Duke University

Let W be a subspace of a Hilbert space V that is spanned by the linearly independent (but not orthogonal) set of vectors $\underline{w}_1, \ldots, \underline{w}_n \in V$.

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

$$\left\langle \underline{v} - \underline{\hat{v}}, \underline{w}_j \right\rangle = \left\langle \underline{v} - \sum_{i=1}^{n} s_i \underline{w}_i \, , \underline{w}_j \right\rangle = 0$$

or, equivalently, the normal equations

$$\sum_{i=1}^{n} s_i \left\langle \underline{w}_i \, , \underline{w}_j \right\rangle = \left\langle \underline{v} \, , \underline{w}_j \right\rangle.$$

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

$$\underbrace{\begin{bmatrix} \langle \underline{w}_1 , \underline{w}_1 \rangle & \langle \underline{w}_2 , \underline{w}_1 \rangle & \cdots & \langle \underline{w}_n , \underline{w}_1 \rangle \\ \langle \underline{w}_1 , \underline{w}_2 \rangle & \langle \underline{w}_2 , \underline{w}_2 \rangle & \cdots & \langle \underline{w}_n , \underline{w}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \underline{w}_1 , \underline{w}_n \rangle & \langle \underline{w}_2 , \underline{w}_n \rangle & \cdots & \langle \underline{w}_n , \underline{w}_n \rangle \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}}_{\underline{s}} = \underbrace{\begin{bmatrix} \langle \underline{v} , \underline{w}_1 \rangle \\ \langle \underline{v} , \underline{w}_2 \rangle \\ \vdots \\ \langle \underline{v} , \underline{w}_n \rangle \end{bmatrix}}_{\underline{t}}.$$

Definition

For $\underline{w}_1, \ldots, \underline{w}_n$, the $n \times n$ Gramian matrix is defined to be

$$G = \begin{bmatrix} \langle \underline{w}_1 , \underline{w}_1 \rangle & \langle \underline{w}_2 , \underline{w}_1 \rangle & \cdots & \langle \underline{w}_n , \underline{w}_1 \rangle \\ \langle \underline{w}_1 , \underline{w}_2 \rangle & \langle \underline{w}_2 , \underline{w}_2 \rangle & \cdots & \langle \underline{w}_n , \underline{w}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \underline{w}_1 , \underline{w}_n \rangle & \langle \underline{w}_2 , \underline{w}_n \rangle & \cdots & \langle \underline{w}_n , \underline{w}_n \rangle \end{bmatrix}$$

Since $g_{ij} = \langle \underline{w}_j , \underline{w}_i \rangle$, 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 $\underline{v}^H M \underline{v} \geq 0$ for all $\underline{v} \in F^n - \{\underline{0}\}$. 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 $\underline{w}_1, \ldots, \underline{w}_n$ 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 $\underline{w}_i \in V$. Then, a vector $\underline{\hat{v}} \in W = \text{colspace}(A)$ can be written as

$$\underline{\hat{v}} = A\underline{s} = \sum_{i=1}^{n} s_i \underline{w}_i.$$

Also, the best approximation of $\underline{v}$ by vectors in W is found by solving

$$\min_{\underline{\hat{v}} \in W} \|\underline{v} - \underline{\hat{v}}\| = \min_{\underline{s}} \|\underline{v} - A\underline{s}\|.$$

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

$$\left\langle \underline{v} - \underline{\hat{v}}, \underline{w}_j \right\rangle = \left\langle \underline{v} - A\underline{s}, \underline{w}_j \right\rangle = 0, \quad j \in [n].$$

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

$$\underline{0} = \begin{bmatrix} \underline{w}_1^H \\ \vdots \\ \underline{w}_n^H \end{bmatrix} (\underline{v} - A\underline{s}) = A^H \underline{v} - A^H A \underline{s} = \underline{t} - G\underline{s},$$

where $G = A^H A$ is the Gramian and $\underline{t}$ is the cross-correlation vector.

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

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

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

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

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

The matrix $P = A\left(A^H A\right)^{-1} A^H$ is the projection matrix for the range of A. It defines an orthogonal projection onto the range of A (i.e., the subspace spanned by the columns of A).

4.3.3: Weighted Least-Squares Solution of a Linear System

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

$$\min_{\underline{\hat{v}} \in W} \|B(\underline{v} - \underline{\hat{v}})\|_E = \min_{\underline{s}} \|B(\underline{v} - A\underline{s})\|_E$$

But, $\|B\underline{v}\|_E$ equals the induced norm $\|\underline{v}\|$ for the weighted inner product

$$\langle \underline{u}, \underline{v} \rangle \triangleq \underline{v}^H B^H B \underline{u}.$$

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

$$\left\langle \underline{v} - \underline{\hat{v}}, \underline{w}_j \right\rangle = \left\langle \underline{v} - A\underline{s}, \underline{w}_j \right\rangle = 0, \quad j \in [n].$$

but they solve a different problem and they reduce to

$$\underline{0} = \begin{bmatrix} \underline{w}_1^H \\ \vdots \\ \underline{w}_n^H \end{bmatrix} B^H B \left(\underline{v} - A\underline{s}\right) = A^H B^H B \underline{v} - A^H B^H B A \underline{s}$$

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 $\mathrm{E}[|Y - \hat{Y}|^2]$. Using the inner product

$$\langle X, Y \rangle = \mathrm{E}\left[X\overline{Y}\right],$$

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

$$G = \begin{bmatrix} \mathrm{E}\left[X_1\overline{X}_1\right] & \mathrm{E}\left[X_2\overline{X}_1\right] & \cdots & \mathrm{E}\left[X_n\overline{X}_1\right] \\ \mathrm{E}\left[X_1\overline{X}_2\right] & \mathrm{E}\left[X_2\overline{X}_2\right] & \cdots & \mathrm{E}\left[X_n\overline{X}_2\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}\left[X_1\overline{X}_n\right] & \mathrm{E}\left[X_2\overline{X}_n\right] & \cdots & \mathrm{E}\left[X_n\overline{X}_n\right] \end{bmatrix}, \quad \underline{t} = \begin{bmatrix} \mathrm{E}\left[Y\overline{X}_1\right] \\ \mathrm{E}\left[Y\overline{X}_2\right] \\ \vdots \\ \mathrm{E}\left[Y\overline{X}_n\right] \end{bmatrix}.$$

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

$$\|Y\|^2 - \left\|\hat{Y}\right\|^2 = \mathrm{E}\left[|Y|^2\right] - \mathrm{E}\left[|\hat{Y}|^2\right] = \mathrm{E}\left[|Y|^2\right] - \underline{s}^H \underline{t}.$$

4.5.1: Dual Approximation and Minimum-Norm Solutions

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 $\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n$ be a basis for subspace W. For any $\underline{v} \in V$, the best approximation of $\underline{v}$ in W can be found by solving

$$\begin{bmatrix} \langle \underline{w}_1 , \underline{w}_1 \rangle & \langle \underline{w}_2 , \underline{w}_1 \rangle & \cdots & \langle \underline{w}_n , \underline{w}_1 \rangle \\ \langle \underline{w}_1 , \underline{w}_2 \rangle & \langle \underline{w}_2 , \underline{w}_2 \rangle & \cdots & \langle \underline{w}_n , \underline{w}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \underline{w}_1 , \underline{w}_n \rangle & \langle \underline{w}_2 , \underline{w}_n \rangle & \cdots & \langle \underline{w}_n , \underline{w}_n \rangle \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} \langle \underline{v} , \underline{w}_1 \rangle \\ \langle \underline{v} , \underline{w}_2 \rangle \\ \vdots \\ \langle \underline{v} , \underline{w}_n \rangle \end{bmatrix}. \qquad (1)$$

Theorem

Let V be a Hilbert space and $\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n$ be a basis for $W \subseteq V$. The dual approximation problem is to find the minimum-norm vector $\underline{w} \in V$ satisfying $\langle \underline{w} , \underline{w}_i \rangle = c_i$ for $i = 1, \ldots, n$. Then, the solution $\underline{w}$ satisfies

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

where $s_1, s_2, \ldots, s_n$ can be found by solving (1) with $\langle \underline{v} , \underline{w}_i \rangle = c_i$.

4.5.2: Minimum-Norm Solutions

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

$$\min_{\underline{s}:A\underline{s}=\underline{v}} \|\underline{s}\|.$$

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

Using $\underline{s} \in \mathcal{R}(A^H)$, there is a $\underline{t}$ such that $\hat{\underline{s}} = A^H\underline{t}$ and the constraint gives $A(A^H\underline{t}) = \underline{v}$. 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 $\hat{\underline{s}}$ can be obtained in closed form and is given by

$$\hat{\underline{s}} = A^H \left(AA^H\right)^{-1} \underline{v}.$$

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