ECEn 671: Mathematics of Signals and Systems

Randal W. Beard

Brigham Young University

September 1, 2023

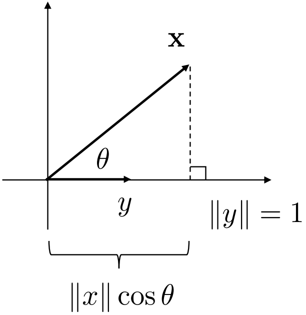
Section 1

Approximation Theory

Projection and Inner Product

- How does inner product represent a projection?
- Recall that

$$\langle x, y \rangle = \|x\| \, \|y\| \, cos\theta$$



- In 2-D $\langle x, y \rangle$ represents the length of the projection of x in the direction of y.
- In general, inner products represent (non-orthogonal) projection of one vector onto another.

Approximation Problem

- Let $\mathbb{S}$ be a Hilbert space, and let $x \in \mathbb{S}$.
- Let $\{p_1, \ldots, p_n\}$ be a set of vectors, all in $\mathbb{S}$.
- Find $\hat{x} \in \text{span}\{p_1, \ldots p_n\}$ that minimizes $\|x - \hat{x}\|$.

Approximation Problem, cont

- Let $\hat{x} = c_1 p_1 + \ldots + c_n p_n \in \text{span}\{p_1, \ldots, p_n\}$.
- By the projection theorem, the error

  $$\begin{aligned} e &= x - \hat{x} \\ &= x - c_1 p_1 - \ldots - c_n p_n \end{aligned}$$

  is minimized if

  $$e \perp \text{span}\{p_1, \ldots, p_n\}.$$

Approximation Problem, cont

$$e \perp \operatorname{span}\{p_1, \ldots, p_n\}.$$

iff

$$\begin{aligned} \langle e, p_1 \rangle &= 0 \\ \langle e, p_2 \rangle &= 0 \\ &\vdots \\ \langle e, p_n \rangle &= 0 \end{aligned}$$

iff

$$\begin{aligned} \langle x - c_1 p_1 - \ldots c_n p_n, p_1 \rangle &= 0 \\ &\vdots \\ \langle x - c_1 p_1 - \ldots c_n p_n, p_n \rangle &= 0 \end{aligned}$$

Approximation Problem, cont

By properties of the inner product we can write this as

$$\begin{aligned}
\langle x, p_1 \rangle - c_1 \langle p_1, p_1 \rangle - \cdots - c_n \langle p_n, p_1 \rangle &= 0 \\
&\vdots \\
\langle x, p_n \rangle - c_1 \langle p_1, p_n \rangle - \cdots - c_n \langle p_n, p_n \rangle &= 0
\end{aligned}$$

or in matrix notation,

$$\underbrace{\begin{pmatrix} \langle p_1, p_1 \rangle & \cdots & \langle p_n, p_1 \rangle \\ \vdots & & \vdots \\ \langle p_1, p_n \rangle & \cdots & \langle p_n, p_n \rangle \end{pmatrix}}_{R} \underbrace{\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}}_{\mathbf{c}} = \underbrace{\begin{pmatrix} \langle x, p_1 \rangle \\ \vdots \\ \langle x, p_n \rangle \end{pmatrix}}_{\mathbf{p}}$$

R is called the Grammian of the set $\{p_1, \ldots, p_n\}$.

The Grammian of a set

Definition (Grammian)

Given a set $\{p_1, \ldots, p_n\}$ of vectors in $\mathbb{S}$, the <u>Grammian</u> of the set is the matrix

$$R = \begin{pmatrix} \langle p_1, p_1 \rangle & \cdots & \langle p_n, p_1 \rangle \\ \vdots & & \vdots \\ \langle p_1, p_n \rangle & \cdots & \langle p_n, p_n \rangle \end{pmatrix}$$

Note that $R^H = R$

We also have the following theorem:

Theorem (Moon, Theorem 3.1)

The Grammian R is positive definite iff the set of vectors $\{p_1, \ldots p_n\}$ are linearly independent.

Proof

Let $y \in \mathbb{S}$ then

$$y^H R y = (\bar{y}_1 \cdots \bar{y}_n) \begin{pmatrix} \langle p_1, p_1 \rangle & \dots & \langle p_n, p_1 \rangle \\ \vdots & & \vdots \\ \langle p_1, p_n \rangle & \dots & \langle p_n, p_n \rangle \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \left(\sum_{i=1}^{n} \bar{y}_i \langle p_1, p_i \rangle \dots \bar{y}_i \sum_{i=1}^{n} \langle p_n, p_i \rangle \right) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} \bar{y}_i y_j \langle p_j, p_i \rangle$$

$$= \left\langle \sum y_j p_j, \sum y_i, p_i \right\rangle = \| \sum y_i p_i \|^2 \geq 0$$

Therefore R is always positive semi-definite.

Proof, cont.

$(\Rightarrow)$: Suppose that R is pd then

$$y^H Ry = \left\|\sum y_i p_i\right\|^2 > 0$$
$$\Rightarrow \sum y_i p_i \neq 0 \text{ for all nonzero } y \in \mathbb{S}$$
$$\Rightarrow \{p_1, \cdots p_n\} \text{ is linearly independent}$$

$(\Leftarrow)$: Conversely suppose $\{p_1, \cdots p_n\}$ is linearly independent, but R is only psd. R is psd implies that $\exists y \neq 0$ such that

$$y^H Ry = \left\|\sum y_i p_i\right\|^2 = 0$$
$$\Rightarrow \sum y_i p_i = 0$$
$$\Rightarrow \{p_1, \cdots p_n\} \text{is linearly dependent.}$$

Which contradicts the assumption that R is psd.

Orthogonality Theorem

Theorem (Moon, Theorem 3.2)

Let $p_1, p_2, \cdots p_n$ be data vectors (or basis vectors) in a Hilbert space $\mathbb{S}$. Let $x \in \mathbb{S}$. Let e be defined as

$$e \triangleq x - \hat{x} = x - \sum_{j=1}^{n} c_j p_j,$$

then e is minimized when it is orthogonal to each of the data vectors, i.e.

$$\langle e, p_j \rangle = 0 \qquad j = 1, \ldots, n$$

Equivalently

$$R\mathbf{c} = \mathbf{p}.$$

Proof.

Follows directly from projection theorem. $\square$

Calculus-Based Approach (Alternative proof)

Rather than using the projection theorem, we can derive the same result using calculus.

Problem Statement: Let $\mathbf{e} = x - \sum_{i=1}^{n} c_i p_i$. Find $\mathbf{c} = (c_1, \ldots, c_n)^\top$ that minimizes $\|\mathbf{e}\|$.

Solution: First note that minimizing $\|\mathbf{e}\|^2$ is equivalent to minimizing $\|e\|$. Also note that

$$\begin{aligned}\|e\|^2 &= \left\langle x - \sum c_j p_j, x - \sum c_j p_j \right\rangle \\ &= \|x\|^2 - 2Re\{\sum_{i=1}^{n} \bar{c}_i \langle x, p_i \rangle\} + \sum \sum c_j \bar{c}_i \langle p_j, p_i \rangle \\ &= \|x\|^2 - 2Re\{\mathbf{c}^H \mathbf{p}\} + \mathbf{c}^H R \mathbf{c}.\end{aligned}$$

Calculus-Based Approach, cont.

To minimize

$$\|e\|^2 = \|x\|^2 - 2Re\{\mathbf{c}^H\mathbf{p}\} + \mathbf{c}^H R\mathbf{c}$$

differentiate with respect to **c** and set to zero. This will be a local minima if the second derivative is psd.

Calculus-Based Approach, cont.

From Moon Appendix we have

$$\frac{\partial}{\partial \bar{\mathbf{c}}} Re\{\mathbf{c}^H \mathbf{p}\} = \frac{1}{2}\mathbf{p}$$
$$\frac{\partial}{\partial \bar{\mathbf{c}}} \mathbf{c}^H R \mathbf{c} = R\mathbf{c}$$

Therefore

$$\frac{\partial \|e\|^2}{\partial \bar{\mathbf{c}}} = -\mathbf{p} + R\mathbf{c} = 0 \qquad \Rightarrow \qquad R\mathbf{c} = \mathbf{p}$$

In addition, we have that

$$\frac{\partial^2 \|e\|^2}{\partial \bar{\mathbf{c}}} = R \geq 0.$$

Therefore the solution of $R\mathbf{c} = \mathbf{p}$ minimize $\|e\|$.
$R\mathbf{c} = \mathbf{p}$ is the same equation we obtained using the projection theorem.

Matrix Representation

- Stack the vectors $\{p_1, \dots p_n\}$ in a matrix

$$A = \begin{pmatrix} p_1 & p_2 & \dots & p_n \end{pmatrix}$$
$$\mathbf{c} = \begin{pmatrix} c_1 & c_2 & \dots & c_n \end{pmatrix}^\top$$

- Then $\hat{x} = \sum c_j p_j = A\mathbf{c}$.
- Therefore $\mathbf{e} = x - \hat{x} = x - A\mathbf{c}$.

Matrix Representation, cont.

- Project $\mathbf{e}$ onto $\{p_1 \ldots p_n\}$:

$$\langle x - A\mathbf{c}, p_1 \rangle = p_1^H (x - A\mathbf{c}) = 0$$
$$\vdots$$
$$\langle x - A\mathbf{c}, p_n \rangle = p_n^H (x - A\mathbf{c}) = 0$$

- Note that $A^H = \begin{bmatrix} p_1^H \\ \vdots \\ p_n^H \end{bmatrix}$.
- Rewrite as

$$A^H (x - A\mathbf{c}) = 0$$
$$\Rightarrow \underbrace{A^H A}_{R} \mathbf{c} = \underbrace{A^H x}_{\mathbf{p}}$$

Matrix Representation, cont.

- If $\{p_1, \ldots p_n\}$ are linearly independent then $R > 0$ which implies that R^{-1} exists, so

$$\mathbf{c} = (A^H A)^{-1} A^H x$$

- Since $\hat{x} = A\mathbf{c}$ we have that

$$\hat{x} = A(A^H A)^{-1} A^H x$$

  is the best approximation to x in $\text{span}\{p_1, \ldots, p_n\}$.
- **Fact:** $P_A = A(A^H A)^{-1} A^H$ is a projection operator from S to $span\{p_1, \ldots, p_n\}$

Application:Polynomial Approximation

- Suppose you are given a real continuous function $f(t)$ and you would like to approximate it by an m^{th} order polynomial on the interval $[a, b]$.
- Let the basis vectors be $\{1, t, \ldots, t^m\}$.
- Then $\hat{f}(t) = c_1 + c_2 t + \cdots + c_{m+1} t^m$
- Define the inner product as $\langle f, g \rangle = \int_a^b f(t)g(t)dt$

Application: Polynomial Approximation, cont.

Then the orthogonality theorem implies that the "best" approximation is given by

$$\left\langle f - \hat{f}, 1 \right\rangle = 0$$
$$\vdots$$
$$\left\langle f - \hat{f}, t^m \right\rangle = 0$$

or

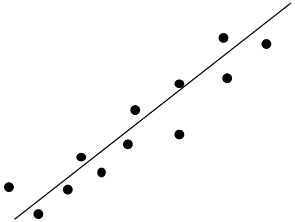
$$\underbrace{\begin{pmatrix} \langle 1, 1\rangle & \cdots & \langle t^m, 1\rangle \\ \vdots & & \vdots \\ \langle 1, t^m\rangle & \cdots & \langle t^m, t^m\rangle \end{pmatrix}}_{\text{Grammian Matrix}} \begin{pmatrix} c_1 \\ \vdots \\ c_{m+1} \end{pmatrix} = \begin{pmatrix} \langle f, 1\rangle \\ \vdots \\ \langle f, t^m\rangle \end{pmatrix}$$

or

$$R\mathbf{c} = \mathbf{p}.$$

Application: Linear Regression

- Suppose you have a number of data points that you are trying to fit to a line.



- Given $(x_i, y_i) \qquad i = 1, \ldots N$
- The equation for a line is $y = ax + b$
- **Problem:** Find a and b that minimizes the mean squared error $\sum_{i=1}^{N} |y_i - ax_i - b|^2$

Application: Linear Regression, cont.

- For each data point we have

$$e_i = y_i - ax_i - b$$

  where e_i is the error for the i^{th} data point.

- Let

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix}, \quad A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} a \\ b \end{pmatrix}$$

- Then $\mathbf{e} = \mathbf{y} - A\mathbf{c}$.

Application: Linear Regression, cont.

- Project the error $\mathbf{e}$ on the data vector (columns of A) and set to zero:

$$A^H\mathbf{e} = A^H(\mathbf{y} - A\mathbf{c}) = 0$$

- Therefore

$$A^H A\mathbf{c} = A^H\mathbf{y}$$

- Giving the minimum least squares solution

$$\mathbf{c} = (A^H A)^{-1} A^H\mathbf{y}.$$