Lecture 9

- Least Squares Solutions

- Inner Product Spaces

Least Square Solutions

A linear system $\mathbf{Ax} = \mathbf{b}$ has either:

- A unique solution
- Infinite number of solutions
- No solutions
  - Matrix $\mathbf{A}$ is constructed as a result of measurement
  - Data in $\mathbf{A}$ is based on Survey

Least Square Solutions

- One of the main tasks in statistical analysis is to understand relations between several quantities that vary.
- Observational data are used to build and verify a formula that predicts the value of one variable as a function of other variables.
- The simplest case is the linear equation $y = \beta_0 + \beta_1 x$.
- This linear model may be useful when observed data points

$$(x_1, y_1), \ldots, (x_n, y_n)$$

  lie close to a line when graphed.

Least Square Solutions

The problem is to find the parameters β_0, β_1 that makes the line $y = \beta_0 + \beta_1 x$ as 'close' to the points as possible.

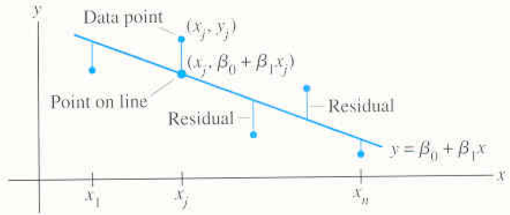


FIGURE 1 Fitting a line to experimental data.

Least Square Solutions

Consider a line $y = \beta_0 + \beta_1 x$ for particular fixed values of β_0, β_1.

- For each data point (x_j, y_j) there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinates.
- We call y_j the *observed* value of y and $\hat{y}_j = \beta_0 + \beta_1 x_j$ the *predicted* y-value.
- The difference between an observed y-value and a predicted y-value is called a *residual*.
- There are many ways to measure how 'close' the line is to the data. One is to add the squares of the residuals. The **least-squares line** is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals.
- This is called a **line of regression of y on x** because any errors in the data are assumed to be in y-coordinates.
- The coefficients β_0, β_1 of the line are called **regression coefficients**.

Least Square Solutions

If the data points lie exactly on the line $y = \beta_0 + \beta_1 x$ then

$$\begin{aligned} \beta_0 + \beta_1 x_1 &= y_1, \\ \beta_0 + \beta_1 x_2 &= y_2, \\ &\vdots \\ \beta_0 + \beta_1 x_n &= y_n. \end{aligned}$$

We can write this system as $\mathbf{X}\beta = \mathbf{y}$, where

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

- The matrix $\mathbf{X}$ is called the **design matrix**,
- $\boldsymbol{\beta}$ is the **parameter vector** and
- $\mathbf{y}$ is the **observation vector**.

If the data points don't lie on a line then there are no parameters β_0, β_1 for which the predicted y-values in $\mathbf{X}\beta$ equal the observed y-values in $\mathbf{y}$ and $X\beta = \mathbf{y}$ does not have a solution.

Least-Squares Problems

- When a solution does not exist the best that we can do is to find an $\mathbf{x}$ that makes $\mathbf{Ax}$ as close as possible to $\mathbf{b}$. This value of $\mathbf{Ax}$ is an approximation to $\mathbf{b}$.
- The smaller the distance between $\mathbf{b}$ and $\mathbf{Ax}$ (given by $\|\mathbf{b} - \mathbf{Ax}\|$) the better the approximation.
- The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible.
- The term *least squares* originated from the definition of the distance $\|\mathbf{b} - \mathbf{Ax}\|$, which is given by the square root of a sum of squares.

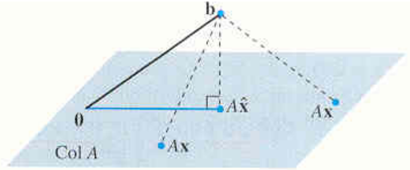


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to

Least-Squares Problems

Definition: If $\mathbf{A}$ is a matrix with sizes $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $\mathbf{Ax} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

- The vector $\mathbf{Ax}$ is in the column space of $\mathbf{A}$.
- Therefore we seek an $\mathbf{x}$ that makes $\mathbf{Ax}$ the closest point in *Col* $\mathbf{A}$ to $\mathbf{b}$ . . . use orthogonal projection.

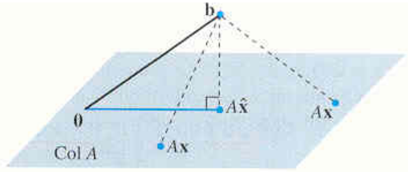


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Least-Squares Solution

We apply the Best Approximation Theorem to the subspace *Col* $\mathbf{A}$. Let

$$\hat{\mathbf{b}} = \mathrm{proj}_{Col\,\mathbf{A}}\mathbf{b}.$$

Then $\hat{\mathbf{b}} \in Col\,\mathbf{A}$ so the equation $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent, i.e. $\exists \hat{\mathbf{x}} \in \mathbb{R}^n$ s.t.

$$\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}.$$

Given the vector $\hat{\mathbf{b}}$ is the closest point in *Col* $\mathbf{A}$ to $\mathbf{b}$ a vector $\hat{\mathbf{x}}$ is a least square solution of $\mathbf{Ax}$ if and only if $\hat{\mathbf{x}}$ satisfies

$$\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}.$$

By Orthogonal Decomposition Theorem, $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to *Col* $\mathbf{A}$, so $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ is orthogonal to each column $\mathbf{a}_j$ of $\mathbf{A}$:

$$\mathbf{a}_j \cdot (\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}, \text{ or } \mathbf{a}_j^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}, \quad \mathbf{j} = \mathbf{1}, \ldots, \mathbf{n}$$
$$\text{or } \mathbf{A}^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}.$$

Hence $\mathbf{A}^\mathbf{T}\mathbf{b} - \mathbf{A}^\mathbf{T}\mathbf{A}\hat{\mathbf{x}} = \mathbf{0}$ and

$$\mathbf{A}^\mathbf{T}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^\mathbf{T}\mathbf{b}.$$

This matrix equation is called the **normal equation** for $\mathbf{Ax} = \mathbf{b}$. The solution of this equation is denoted by $\hat{\mathbf{x}}$.

Least-Squares Solution

- **Theorem:** The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation
$$\mathbf{A}^\mathbf{T}\mathbf{A}\mathbf{x} = \mathbf{A}^\mathbf{T}\mathbf{b}.$$

- **Note:** If $\hat{\mathbf{x}}$ satisfies the normal equation then the vector $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ is orthogonal to all of *Col* $\mathbf{A}$. Therefore
$$\mathbf{b} = \mathbf{A}\hat{\mathbf{x}} + (\mathbf{b} - \mathbf{A}\hat{\mathbf{x}})$$
is a decomposition of $\mathbf{b}$ into the sum of a vector $\mathbf{A}\hat{\mathbf{x}}$ in *Col* $\mathbf{A}$ and a vector $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ orthogonal to *Col* $\mathbf{A}$.
- Therefore $\mathit{Col}\,\mathbf{A} = \mathit{Row}\,\mathbf{A}^\mathbf{T} \perp \mathit{Nul}\,\mathbf{A}^\mathbf{T}$. So $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}} \in \mathit{Nul}\,\mathbf{A}^\mathbf{T}$
- Null space of $\mathbf{A}^\mathbf{T}$ is a set of all $\mathbf{y}$ vectors in $\mathbb{R}^m$ such that $\mathbf{A}^\mathbf{T}\mathbf{y} = \mathbf{0}$

Least-Squares Solution

- When a least squares solution $\hat{\mathbf{x}}$ is used to produce $\mathbf{A}\hat{\mathbf{x}}$ as an approximation to $\mathbf{b}$ the distance from $\mathbf{b}$ to $\mathbf{A}\hat{\mathbf{x}}$ is the **least squares error** of this approximation.

- The normal equation can be derived from the condition that the error $\varepsilon^{\mathbf{2}} = (\mathbf{b} - \mathbf{A}\mathbf{x}) \cdot (\mathbf{b} - \mathbf{A}\mathbf{x}) = (\mathbf{b} - \mathbf{A}\mathbf{x})^{\mathbf{T}}(\mathbf{b} - \mathbf{A}\mathbf{x})$ be minimal.

- So by requiring that the partial derivatives of this length in terms of x_j be minimal we arrive to $\mathbf{2A^T A x} - \mathbf{2A^T b} = \mathbf{0}$.

Least-Squares Solution

Example 1: Find a least-squares solution of the inconsistent system $\mathbf{Ax} = \mathbf{b}$ for

$$\mathbf{A} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}.$$

Solution: Construct the normal equation $\mathbf{A^T A x} = \mathbf{A^T b}$.
We compute $\mathbf{A^T A}$

$$\mathbf{A^T A} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix},$$

$$\mathbf{A^T b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

Least-Squares Solution

The equation becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

$$\mathbf{A^T A x} = \mathbf{A^T b}.$$

$$(\mathbf{A^T A})^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

$$\hat{\mathbf{x}} = (\mathbf{A^T A})^{-1} A^T \mathbf{b} = \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Least-Squares Solution

Given **A** and **b** as in **Example 1**, determine the least squares error of the least square solution of $\mathbf{Ax} = \mathbf{b}$ for

$$\mathbf{A} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}, \ \hat{\mathbf{x}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Solution: We calculate $\mathbf{A}\hat{\mathbf{x}}$

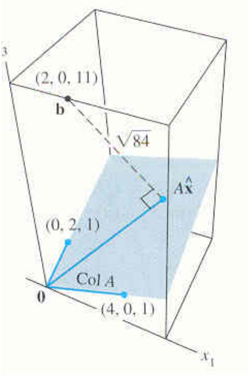
$$\mathbf{A}\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

Therefore

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

and $\|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$. For any $\mathbf{x}$ in $\mathbb{R}^2$ the distance between $\mathbf{b}$ and the vector $A\mathbf{x}$ is at least $\sqrt{84}$.

Least-Squares Solution



IGURE 3

$$A\hat{\mathbf{x}} = \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}.$$

The solution $\hat{\mathbf{x}}$ is not on the figure.

$\mathbf{A^T A}$ is often invertible but sometimes it is not.

Example 2: Find the least square solution of $A\mathbf{x} = \mathbf{b}$ for

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix}.$$

Solution: We compute

$$\mathbf{A^T A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}.$$

Least-Squares Solution

$$\mathbf{A}^\mathsf{T}\mathbf{b} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -4 \\ 2 \\ 6 \end{pmatrix}.$$

The augmented matrix for $\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{x} = \mathbf{A}^\mathsf{T}\mathbf{b}$ is

$$\left(\begin{array}{cccc|c} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Hence

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ -5 \\ -2 \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{cases} x_1 + x_4 & = 3, \\ x_2 - x_4 & = -5, \\ x_3 - x_4 & = -2 \end{cases}$$

Least-Squares Solution

The general solution is

$$\hat{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 - x_4 \\ -5 + x_4 \\ -2 + x_4 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ -5 \\ -2 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

This is the general least squares solution of $\mathbf{Ax} = \mathbf{b}$.

Projection Matrix

- **Theorem:** Matrix $\mathbf{A^T A}$ has the same null space as $\mathbf{A}$.
  Proof: If $\mathbf{Ax} = \mathbf{0}$ then of course $\mathbf{A^T Ax} = \mathbf{0}$.
  Let assume now that $\mathbf{A^T Ax} = \mathbf{0}$. Lets dot product of this equation from the right by $\mathbf{x}$. We obtain $\mathbf{x} \cdot \mathbf{A^T Ax} = \mathbf{0}$ or $\mathbf{x} \cdot \mathbf{A^T Ax} = \mathbf{x^T A^T Ax} = (\mathbf{Ax})^{\mathbf{T}}(\mathbf{Ax}) = \|\mathbf{Ax}\|^{\mathbf{2}} = \mathbf{0}$. So $\mathbf{Ax} = \mathbf{0}$.

- As a consequence if columns of $\mathbf{A}$ are linearly independent than the only solution for $\mathbf{Ax} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. Therefore the only solution of $\mathbf{A^T Ax} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$ and $\mathbf{A^T A}$ is invertible.

- In this case the equation $\mathbf{Ax} = \mathbf{b}$ has only one least-square solution $\hat{\mathbf{x}}$ and it is given by

$$\hat{\mathbf{x}} = (\mathbf{A^T A})^{-\mathbf{1}}\mathbf{A^T b}.$$

Projection Matrix

- If the least square solution for $\mathbf{Ax} = \mathbf{b}$ is
$$\hat{\mathbf{x}} = (\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{b}$$
Then the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$ is
$$\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\mathbf{b} = \mathbf{Pb}$$
- Matrix $\mathbf{P}$ is the projection matrix
$$\mathbf{P} = \mathbf{A}(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}$$
- If matrix $\mathbf{A}$ has sizes $[m \times n]$ then $\mathbf{P}$ matrix has sizes $[m \times m]$.
- The projection matrix $\mathbf{P}$ has the property
$$\mathbf{P}^\mathbf{2} = \mathbf{P}$$
- Projection matrix is symmetric
$$\mathbf{P}^\mathbf{T} = \mathbf{P}$$

.

Projection Matrix

- Lets prove that $\mathbf{P^2} = \mathbf{P}$. Indeed
$$\mathbf{P^2} = \mathbf{A}(\mathbf{A^T}\mathbf{A})^{-1}\mathbf{A^T}\mathbf{A}(\mathbf{A^T}\mathbf{A})^{-1}\mathbf{A^T} = \mathbf{A}(\mathbf{A^T}\mathbf{A})^{-1}\mathbf{A^T} = \mathbf{P}$$
- **Case 1**: Lets assume $\mathbf{b}$ is in the column space of $\mathbf{A}$, so $\mathbf{b} = \mathbf{Ax}$ then
$$\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A^T}\mathbf{A})^{-1}\mathbf{A^T}\mathbf{Ax} = \mathbf{Ax} = \mathbf{b}$$
So $\hat{\mathbf{b}} = \mathbf{b}$ as this should be.
- **Case 2**: Lets assume $\mathbf{b}$ is orthogonal to the column space of $\mathbf{A}$, $\mathbf{b} \perp Col(\mathbf{A})$ then $\mathbf{b}$ is perpendicular to every column of $\mathbf{A}$, therefore $\mathbf{A^T}\mathbf{b} = \mathbf{0}$. So
$$\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A^T}\mathbf{A})^{-1}\mathbf{A^T}\mathbf{b} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A^T}\mathbf{A})^{-1}\mathbf{0} = \mathbf{0}$$
So $\hat{\mathbf{b}} = \mathbf{0}$ as this should be.
- **Case 3**: Lets assume that $\mathbf{A}$ is square and invertible then
$$\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}(\mathbf{A^T}\mathbf{A})^{-1}\mathbf{A^T}\mathbf{b} = \mathbf{A}\mathbf{A}^{-1}(\mathbf{A^T})^{-1}\mathbf{A^T}\mathbf{b} = \mathbf{b}$$
So in this case the projection matrix $\mathbf{P}$ is an Identity matrix $\mathbf{I}$.

Projection Matrix

Example 3: Find the projection matrix $\mathbf{P}$ onto space spanned by the columns of matrix $\mathbf{A}$

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

Solution: We need to calculate: $\mathbf{P} = \mathbf{A}(\mathbf{A^T A})^{-\mathbf{1}}\mathbf{A^T}$
So

$$\mathbf{A^T} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \;\; \mathbf{A^T A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \;\; (\mathbf{A^T A})^{-\mathbf{1}} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Then

$$\mathbf{P} = \frac{1}{3}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}.$$

Projection Matrix

Example 4: Find the projection matrix $\mathbf{P}$ onto space spanned by the vector $\mathbf{a}$. Find the projection of vector $\mathbf{b}$ onto vector $\mathbf{a}$.

$$\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Solution: We need to calculate:
$\mathbf{P} = \mathbf{A}(\mathbf{A^T A})^{-1}\mathbf{A^T} = \mathbf{a}(\mathbf{a^T a})^{-1}\mathbf{a^T}$
So

$$\mathbf{a^T} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}, \ \mathbf{a^T a} = 2, \ (\mathbf{a^T a})^{-1} = \frac{1}{2}.$$

Then

$$\mathbf{P} = \mathbf{a}(\mathbf{a^T a})^{-1}\mathbf{a^T} = \frac{\mathbf{aa^T}}{\mathbf{a^T a}}$$

And

$$\mathbf{P} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \hat{\mathbf{b}} = \mathbf{Pb} = \frac{\mathbf{aa^T b}}{\mathbf{a^T a}} = \frac{\mathbf{a}(\mathbf{a}\cdot\mathbf{b})}{\mathbf{a}\cdot\mathbf{a}} = \frac{1}{2}\mathbf{a}$$

Projection Matrix

Example 5: Find the projection matrix $\mathbf{P}$ for matrix $\mathbf{A}$ with orthonormal columns.
Solution: We need to calculate:

$$\mathbf{P} = \mathbf{A}(\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}$$

Given that columns of $\mathbf{A}$ are orthonormal we have $\mathbf{A}^\mathbf{T}\mathbf{A} = \mathbf{I}$. So $\mathbf{P} = \mathbf{A}\mathbf{A}^\mathbf{T}$ and

$$\hat{\mathbf{b}} = \mathbf{P}\mathbf{b} = \mathbf{A}\mathbf{A}^\mathbf{T}\mathbf{b}$$

. If the columns of $\mathbf{A}$ are $\mathbf{a}_1$, $\mathbf{a}_2, \ldots,$ $\mathbf{a}_n$ then the orthogonal

projection of $\mathbf{b}$ onto $Col\,\mathbf{A}$ is

$$\hat{\mathbf{b}} = \mathbf{A}\mathbf{A}^\mathbf{T}\mathbf{b} = \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 + \cdots + \frac{\mathbf{b}\cdot\mathbf{a}_n}{\mathbf{a}_n\cdot\mathbf{a}_n}\mathbf{a}_n$$

.

$$\hat{\mathbf{b}} = (\mathbf{b}\cdot\mathbf{a}_1)\mathbf{a}_1 + (\mathbf{b}\cdot\mathbf{a}_2)\mathbf{a}_2 + \ldots + (\mathbf{b}\cdot\mathbf{a}_n)\mathbf{a}_n$$

Given $\mathbf{a}_j\cdot\mathbf{a}_j = 1$.

Least-Squares Solution

Example 6: Find the least squares solution of $\mathbf{Ax} = \mathbf{b}$ for

$$\mathbf{A} = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}.$$

Solution: The columns $\mathbf{a}_1$, $\mathbf{a}_2$ of $\mathbf{A}$ are orthogonal so the orthogonal projection of $\mathbf{b}$ onto *Col* $\mathbf{A}$ is given by

$$\hat{\mathbf{b}} = \frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b} \cdot \mathbf{a}_2}{\mathbf{a}_2 \cdot \mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2.$$

$$\hat{\mathbf{b}} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}.$$

Least-Squares Solution

Solutions (cont'd): So $\hat{\mathbf{b}}$ is known and we can solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. But we know the solution of this equation given the relation

$$\frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 = \hat{\mathbf{b}}.$$

Therefore

$$\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}.$$

This is the least squares solution.

Least-Squares Solution

In some cases the equation

$$\mathbf{A^T A x} = \mathbf{A^T b}$$

is ill conditioned. Small errors in $\mathbf{A^T A}$ can lead to large errors in $\hat{\mathbf{x}}$. If the columns of $\mathbf{A}$ are linearly independent the solution can be found by a $\mathbf{QR}$ factorization of $\mathbf{A}$.

Theorem: Given an $m \times n$ matrix $\mathbf{A}$ with linearly independent columns, and $\mathbf{A} = \mathbf{QR}$ is a QR factorization of $\mathbf{A}$. Then for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $\mathbf{Ax} = \mathbf{b}$ has a unique least-squares solution given by

$$\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q^T b}.$$

Proof: Let $\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q^T b}$. Then

$$\mathbf{A}\hat{\mathbf{x}} = \mathbf{QR}\hat{\mathbf{x}} = \mathbf{QRR}^{-1}\mathbf{Q^T b} = \mathbf{QQ^T b}.$$

The columns of $\mathbf{Q}$ form an orthonormal basis for *Col* $\mathbf{A}$. Therefore $\mathbf{QQ^T b}$ is the orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto *Col* $\mathbf{A}$. Then $\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}$ and $\hat{\mathbf{x}}$ is the least squares solution for $\mathbf{Ax} = \mathbf{b}$.

Least-Squares Solution

Note: The least-squares solution can be found from

$$\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{b}.$$

We can write this as

$$\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^{\mathsf{T}}\mathbf{b}.$$

However, $\mathbf{R}$ is upper triangular so the solution can be found by back-substitution or row operations rather than calculating $\mathbf{R}^{-1}$.

Applications to Linear Models

Example: Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2,1),(5,2),(7,3),(8,3)$.

Solution: The linear system $X\beta = \mathbf{y}$ is

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

The normal equation for this system is

$$\mathbf{X}^{\mathbf{T}}\mathbf{X}\beta = \mathbf{X}^{\mathbf{T}}\mathbf{y}.$$

We need to calculate:

$$\mathbf{X}^{\mathbf{T}}\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}.$$

Applications to Linear Models

$$\mathbf{X}^{\mathsf{T}}\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

The normal equation is

$$\mathbf{X}^{\mathsf{T}}\mathbf{X}\beta = \mathbf{X}^{\mathsf{T}}\mathbf{y}.$$

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 59 \\ 57 \end{bmatrix}.$$

Solution (cont'd):

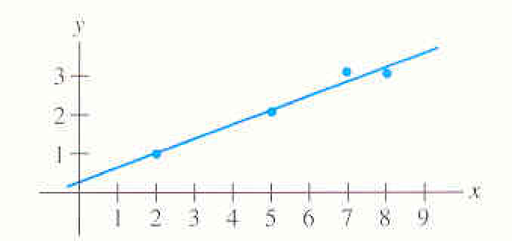
The solution is

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} =$$

$$= \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

The least squares solution is

$$y = \frac{2}{7} + \frac{5}{14}x$$



The General Linear Model

- In statistics usually a **residual vector** ε is introduced, defined by $\varepsilon = \mathbf{y} - \mathbf{X}\beta$.
- The equation for $\mathbf{y}$ is $\mathbf{y} = \mathbf{X}\beta + \varepsilon$. Any equation of this form is referred to as a **linear model**. The goal is to minimize the length of ε. This is equivalent to finding a least squares solution of $\mathbf{X}\beta = \mathbf{y}$.

- Sometimes we need to fit the data with more complex curves.
- The matrix equation is still the same $\mathbf{X}\beta = \mathbf{y}$, but the form of $\mathbf{X}$ depends on the problem.
- In each case the least-squares solution $\hat{\beta}$ is a solution of the normal equation $\mathbf{X}^{\mathbf{T}}\mathbf{X}\beta = \mathbf{X}^{\mathbf{T}}\mathbf{y}$.

Least-Square Fittings of Curves

When the data points $(x_1, y_1), \ldots, (x_n, y_n)$ do not lie close to any line it may be appropriate to postulate other functional relationships between x and y.

The general linear form is

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \ldots + \beta_k f_k(x),$$

where

$f_0, \ldots, f_k$ are known functions and

$\beta_0, \ldots, \beta_k$ are parameters that must be determined.

This is a linear equation in terms of the β_j.

For a particular value of x the relation gives a predicted value of y. The difference between the predicted and observed value is the residual. The parameters $\beta_0, \ldots, \beta_k$ must be determined to minimize the sum of the squares of the residuals.

Least-Squares Fittings of Curves

Example: Suppose data points $(x_1, y_1), \ldots, (x_n, y_n)$ appear to lie along some kind of parabola. We approximate the data using the model function

$$y = \beta_0 + \beta_1 x + \beta_2 x^2.$$

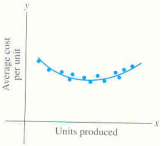


FIGURE 3
Average cost curve.

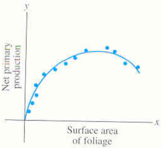


FIGURE 4
Production of nutrients.

Least-Squares Fittings of Curves

Each data point determines an equation.

$$\begin{aligned}
\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon_1 &= y_1, \\
\beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \varepsilon_2 &= y_2, \\
&\vdots \\
\beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \varepsilon_n &= y_n,
\end{aligned}$$

where the ε_i is the residual error. This is a linear system in terms of the β_j in the form

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

or

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

To find the least-squares solution solve $\mathbf{X^T X}\beta = \mathbf{X^T y}$.

Inner Product Spaces

Inner Product Spaces

- Notions of length, distance, and orthogonality are important in applications in general vector spaces V.
- So far we considered $\mathbb{R}^n$ spaces.
- The concept of the inner product can be introduced in general vector spaces V.
- Definition : An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and all scalars c

  1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
  2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
  3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
  4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Inner Product Spaces

- The vector space $\mathbb{R}^n$ with the standard inner product $\mathbf{u} \cdot \mathbf{v}$ is an Inner Product Space.
- The general vector space V with the supplied inner product is called an Inner Product Space.
- Such spaces are foundations for applications in engineering, physics mathematics and statistics.
- Example: Let consider two vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in $\mathbb{R}^2$ and lets set

$$\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1 v_1 + 5u_2 v_2$$

- This is an inner product.
- This inner product can be defined for $\mathbb{R}^n$.
  They arise in the weighted-least-square problems when more importance is given to more reliable measurement.

Inner Product Spaces

- Let consider the vector space of polynomials $\mathbb{P}_n$.
- Let $t_0, \ldots, t_n$ be distinct real numbers. For two polynomials p and q in $\mathbb{P}_n$ we define

  $$\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \ldots + p(t_n)q(t_n).$$

  Inner product Axioms 1-3 are satisfied. This is easy to check. For Axiom 4

  $$\langle p, p \rangle = [p(t_0)]^2 + [p(t_1)]^2 + \ldots + [p(t_n)]^2 \geq 0$$

  If $\langle p, p \rangle = 0$ then p must vanish at $n+1$ points $t_0, \ldots, t_n$. This is possible only for zero polynomial $p = 0$ because the degree of the polynomial is less then $n+1$.
- Therefore

  $$\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \ldots + p(t_n)q(t_n).$$

  is an inner product.

Example 2: Let V be $\mathbb{P}^2$ with the inner product defined as in the previous slide, where $t_0 = 0, t_1 = 1/2, t_2 = 1$. Let $p(t) = 12t^2$ and $q(t) = 2t - 1$.

Compute $\langle p, q \rangle$ and $\langle q, q \rangle$.
Solution:

$$\langle p, q \rangle = p(t_0)q(t_0) + p(1/2)q(1/2) + p(1)q(1) =$$

$$= 0(-1) + 3 \times 0 + 12 \times 1 = 12$$

$$\langle q, q \rangle = q(0)q(0) + q(1/2)q(1/2) + 1(1)q(1) =$$
$$= (-1)^2 + 0^2 + 1^2 = 2$$

Length, Distance, Orthogonality

- Let V be an inner product space, with the inner product denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$. Just as in $\mathbb{R}^n$ we define the length or norm of a vector $\mathbf{v}$ to be the scalar

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

- Equivalently $||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle$.
  This definition makes sense because $\langle \mathbf{v}, \mathbf{v} \rangle \geq \mathbf{0}$.
- A **unit vector** is one whose length is 1.
- The **distance between two vectors u and v** is $||\mathbf{u} - \mathbf{v}||$
  Vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$

Example 4: Let $\mathbb{P}_2$ have the inner product as

$$\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \ldots + p(t_n)q(t_n).$$

Compute the length of the vectors $p(t) = 12t^2$ and $q(t) = 2t - 1$ and $t_0 = 0, t_1 = 1/2, t_2 = 1$.
Solution: $p(t) = 12t^2$ and $q(t) = 2t - 1$

$$||p||^2 = \langle p, p \rangle = [p(0)]^2 + [p(1/2)]^2 + [p(1)]^2 =$$
$$= 0 + 3^2 + 12^2 = 153$$

$$||q||^2 = [q(0)]^2 + [q(1/2)]^2 + [q(1)]^2 = 2.$$

So $||q|| = \sqrt{2}$.

The Gram-Schmidt Process

- The Gram-Schmidt process can be applied to construct an orthogonal basis for a subspace of a vector space V in the similar way as in $\mathbb{R}^n$.
- An orthogonal projection of a vector onto a subspace W with the orthogonal basis can be constructed as in $\mathbb{R}^n$.
- The projection has the properties described in the Orthogonal Decomposition Theorem and the Best Approximation Theorem.
- **Example 5**: Let V be $\mathbb{P}_4$ with the inner product defined as

$$\langle p, q\rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + p(t_2)q(t_2) +$$

$$+ p(t_3)q(t_3) + p(t_4)q(t_4).$$

  involving evaluation of polynomials at $-2, -1, 0, 1, 2$. Produce an orthogonal basis for $\mathbb{P}_2$ by applying the Gram-Schmidt process to the polynomials : $p_0(t) = 1$, $p_1(t) = t$, $p_2(t) = t^2$.

- **Solution**: The inner product depends on the values of the polynomials calculated at points $-2, -1, 0, 1, 2$. The values of the polynomials $p_0(t) = 1$, $p_1(t) = t$, $p_2(t) = t^2$ at points $-2, -1, 0, 1, 2$ are

$$\mathbf{v_0} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v_1} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} 4 \\ 1 \\ 0 \\ 1 \\ 4 \end{bmatrix}.$$

  These are vectors in $\mathbb{R}^5$.
- The inner product of two polynomials in V equals to the inner product of their corresponding vectors in $\mathbb{R}^5$.
- We can see that $\langle p_0(t) p_1(t) \rangle = \mathbf{v_0} \cdot \mathbf{v_1} = 0$.
  These polynomials are orthogonal and we can choose them in the new basis vectors.

The Gram-Schmidt Process

- To find the new p_2 we project p_2 into subspace spanned by p_0, p_1. The inner product we use is:

$$\langle p, q \rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + p(t_2)q(t_2) +$$
$$+ p(t_3)q(t_3) + p(t_4)q(t_4).$$

- We need to use the Gram-Schmidt process to construct the third basis vector.

$$proj_W \; p_2 = \frac{\langle p_0 p_2 \rangle}{\langle p_0 p_0 \rangle} p_0 + \frac{\langle p_1 p_2 \rangle}{\langle p_1 p_1 \rangle} p_1$$
$$\langle p_0 p_2 \rangle = \langle p_0 \; t^2 \rangle = 4 + 1 + 0 + 1 + 4 = 10$$
$$\langle p_0 p_0 \rangle = 1 + 1 + 1 + 1 + 1 = 5$$
$$\langle p_1 p_2 \rangle = \langle t \; t^2 \rangle = -8 + (-1) + 0 + 1 + 8 = 0$$

- Then the orthogonal projection of t^2 onto $1, t$ is

$$proj_W \; p_2 = 10/5 p_0 + 0 p_1 = 2$$

The Gram-Schmidt Process

The new basis vector is

$$p_2'(t) = p_2(t) - proj_W\, p_2 = t^2 - 2$$

The orthogonal basis for the subspace $\mathbb{P}_2$ of V is

$$\mathbf{v_0} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v_1} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \mathbf{v_2} = \begin{bmatrix} 2 \\ -1 \\ -2 \\ -1 \\ 2 \end{bmatrix}.$$

The polynomials are

$$p_0(t) = 1,\ p_1(t) = t,\ p_2'(t) = t^2 - 2.$$

Best Approximation

- A common vector space in Applied Mathematics is a vector space V whose elements are functions.
- The problem is to approximate a function f in V by a function g from a specified subspace W of V.
- The 'closeness' of the approximation of f depends on the way $||f - g||$ is defined.
- When the distance between f and g is given by an inner product the best approximation to f by functions in W is the orthogonal projection of f onto the subspace of W.

Best Approximation

Example: Let V be $\mathbb{P}^4$ with the inner product

$$\langle p, q\rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + p(t_2)q(t_2) +$$
$$+ p(t_3)q(t_3) + p(t_4)q(t_4)$$

calculated at the points $-2, -1, 0, 1, 2$.

- Let $p_0(t) = 1, p_1(t) = t, p_2(t) = t^2 - 2$ be the orthogonal basis for the subspace $\mathbb{P}^2$. Find the best approximation to $p(t) = 5 - \frac{1}{2}t^4$.
- The best approximation is

$$proj_W\, p = \frac{\langle p_0 p\rangle}{\langle p_0 p_0\rangle} p_0 + \frac{\langle p_1 p\rangle}{\langle p_1 p_1\rangle} p_1 + \frac{\langle p_2 p\rangle}{\langle p_2 p_2\rangle} p_2.$$

- We calculate
  $\langle p\, p_0\rangle = 8,\ \langle p\, p_1\rangle = 0, \langle p\, p_2\rangle = -31$
  $\langle p_0\, p_0\rangle = 5, \langle p_2\, p_2\rangle = 14$

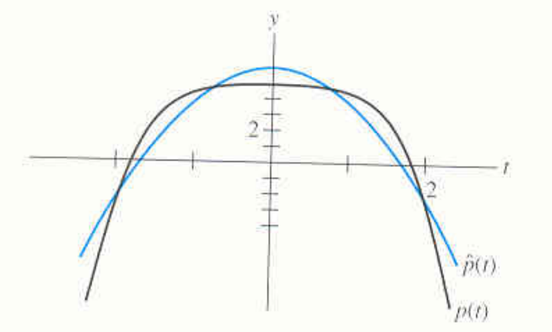
$$\hat{p} = proj_{\mathbb{P}_2}\, p = \frac{8}{5} p_0 + \frac{-31}{14} p_2 = \frac{8}{5} - \frac{31}{14}(t^2 - 2).$$

Best Approximation

The polynomial $\hat{p}(t)$ is the closest to $p(t)$ (which is part of $\mathbb{P}^4$) of all polynomials in $\mathbb{P}^2$, when the distance measured at points $-2, -1, 0, 1, 2$.

$$\hat{p}(t) = proj_{\mathbb{P}_2}\, p(t) = \frac{8}{5} - \frac{31}{14}(t^2 - 2).$$

$$p(t) = 5 - \frac{1}{2}t^4.$$



An inner product for $C[a, b]$

When the sum (in used above inner product) is weighted by the widths $\frac{b-a}{n}$ (with a minor adjustment at the endpoints) and $n \to \infty$, we get an integral definition of $\langle f, g \rangle$ in the interval $[a, b]$ as

$$\langle f, g \rangle = \int_a^b f(t)g(t)\,dt.$$

To show this defines an inner product on $C[a, b]$:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

$$\langle f, g \rangle = \int_a^b f(t)g(t)\,dt = \int_a^b g(t)f(t)\,dt = \langle g, f \rangle.$$

2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

$$\begin{aligned} \langle f + h, g \rangle &= \int_a^b (f(t) + h(t))g(t)\,dt \\ &= \int_a^b f(t)g(t)\,dt + \int_a^b h(t)g(t)\,dt \\ &= \langle f, g \rangle + \langle h, g \rangle. \end{aligned}$$

An inner product for $C[a, b]$

3. $\langle c\mathbf{u}, \mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$

$$\langle cf, g\rangle = \int_a^b (cf(t))g(t)\,dt = c\langle f, g\rangle.$$

4. $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u}\rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

$$\langle f, f\rangle = \int_a^b f(t)f(t)\,dt = \int_a^b f^2(t)\,dt \geq 0.$$

The function $f^2(t)$ is continuous and nonnegative on $[a, b]$.
If the definite integral of $f^2(t)$ is zero it means $f(t) \equiv 0$ on $[a, b]$.
Therefore $\langle f, f\rangle = 0$ implies that $f \equiv 0$.

Hence

$$\langle f, g\rangle = \int_a^b f(t)g(t)\,dt$$

is an inner product on $C[a, b]$.

An inner product for $C[a, b]$

Example: Let V be the space $C[a, b]$ with the inner product

$$\langle f, g \rangle = \int_0^1 f(t)g(t)\, dt$$

and let W be the subspace spanned by the polynomials

$$p_1(t) = 1,\ p_2(t) = 2t - 1,\ p_3(t) = 12t^2.$$

Use the Gram-Schmidt process to find an orthogonal basis for W.
Solution: We need to use Gram–Schmidt to build a new set

$$\{q_1(t),\ q_2(t),\ q_3(t)\}$$

which will be an orthogonal basis for W.
As usual, let $q_1 = p_1$. We compute

$$\langle p_2, q_1 \rangle = \int_0^1 (2t - 1)(1)dt = (t^2 - t)|_0^1 = 0.$$

So p_2 is already orthogonal to q_1 and we choose $q_2 = p_2$

An inner product for $\mathbb{C}[a, b]$

As usual

$$p_1(t) = 1,\ p_2(t) = 2t - 1,\ \ p_3(t) = 12t^2.$$

The q_3 is given as usual by

$$q_3(t) = p_3 - proj_{W_2}\, p_3$$

We need to find the projection of p_3 into subspace $W_2 = Span\{q_1, q_2\}$. This is given by

$$proj_W\, p_3 = \frac{\langle q_1 p_3 \rangle}{\langle q_1 q_1 \rangle} q_1 + \frac{\langle q_2 p_3 \rangle}{\langle q_2 q_2 \rangle} q_2.$$

$$\langle p_3, q_1 \rangle = \int_0^1 12t^2(1)dt = 4t^3|_0^1 = 4.$$

$$\langle q_1, q_1 \rangle = \int_0^1 (1)(1)dt = t|_0^1 = 1.$$

$$\langle p_3, q_2 \rangle = \int_0^1 12t^2(2t - 1)dt = 2.$$

An inner product for $\mathbb{C}[a, b]$

,

$$\langle q_2, q_2 \rangle = \int_0^1 (2t-1)^2(1)dt = t|_0^1 = \frac{1}{3}.$$

Then the projection is

$$proj_W\ p_3 = \frac{\langle q_1 p_3 \rangle}{\langle q_1 q_1 \rangle} q_1 + \frac{\langle q_2 p_3 \rangle}{\langle q_2 q_2 \rangle} q_2 = \frac{4}{1} q_1 + \frac{2}{1/3} q_2.$$

Therefore

$$proj_W\ p_3 = \frac{\langle q_1 p_3 \rangle}{\langle q_1 q_1 \rangle} q_1 + \frac{\langle q_2 p_3 \rangle}{\langle q_2 q_2 \rangle} q_2 = \frac{4}{1} q_1 + \frac{2}{1/3} q_2.$$

and

$$q_3(t) = p_3 - proj_{W_2}\ p_3 =$$

$$= p_3 - 4q_1 - 6q_2 = 12t^2 - 4 - 6(2t-1) = 12t^2 - 12t + 2.$$

given

$$q_1(t) = p_1(t) = 1,\ q_2(t) = p_2(t) = 2t - 1, p_3(t) = 12t^2$$

The orthogonal basis for W is $\{q_1, q_2, q_3\}$

Trend Analysis of Data

- Let f be unknown function whose values are approximately known at $t_0, \ldots, t_n$.
- If there is a linear trend of the data $f(t_0), \ldots, f(t_n)$ we may approximate the data by $y = \beta_0 + \beta_1 x$.
- If there is a quadratic trend then we may try $y = \beta_0 + \beta_1 x + \beta_1 x^2$.
- However the coefficient β_2 may not give the desired information about the quadratic trend in the data because it may not be independent from β_0, β_1.
- To make what is known as a **trend analysis** of the data we introduce an inner product on the space $\mathbb{P}^n$ in the form

$$\langle p, q \rangle = p(t_0)q(t_0) + \ldots + p(t_n)q(t_n)$$

Trend Analysis of Data

- Let p_0, p_1, p_2, p_3 be an orthogonal basis of the subspace $\mathbb{P}_3$ of $\mathbb{P}_n$ obtained from basis $1, t, t^2, t^3$ using the Gram-Schmidt process.
- There is a polynomial g in $\mathbb{P}_n$ whose values at $t_0, \ldots, t_n$ coincide with the unknown function f.
- Let $\hat{g}$ be the orthogonal projection of g onto $\mathbb{P}_3$.

$$\hat{g} = c_0 p_0 + c_1 p_1 + c_2 p_2 + c_3 p_3.$$

- Then $\hat{g}$ is called the trend function and the coefficients $c_0, \ldots, c_n$ the trend coefficients of the data.
- If the data have certain properties the c_j are independent. Since $p_0, \ldots, p_3$ are orthogonal the trend coefficients can be calculated one at a time

$$c_i = \frac{\langle p_j, \hat{g} \rangle}{\langle p_j, p_j \rangle}$$

Trend Analysis of Data

Example : Fit a quadratic trend function to the data $(-2,3),(-1,5),(0,5),(1,4),(2,3)$.
$g(-2)=3, g(-1)=5, g(0)=5, g(1)=4,$
$g(2)=3$

- The orthogonal polynomials are $p_0(t)=1, p_1(t)=t, p_2(t)=t^2-2$

$$\mathbf{p_0}=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix},\ \mathbf{p_1}=\begin{bmatrix}-2\\-1\\0\\1\\2\end{bmatrix},\mathbf{p_2}=\begin{bmatrix}2\\-1\\-2\\-1\\2\end{bmatrix},\mathbf{g}=\begin{bmatrix}3\\5\\5\\4\\3\end{bmatrix}.$$

Then the best approximation to the data by $\mathbb{P}^2$ is given by the orthogonal projection

$$\hat{p}=\frac{\langle p_0 g\rangle}{\langle p_0 p_0\rangle}p_0+\frac{\langle p_1 g\rangle}{\langle p_1 p_1\rangle}p_1+\frac{\langle p_2 g\rangle}{\langle p_2 p_2\rangle}p_2=$$
$$=20/4p_0-1/10p_1-7/10p_2.$$

Trend Analysis of Data

$$\hat{p}(t) = 4 - 0.1t - 0.5(t^2 - 2)$$

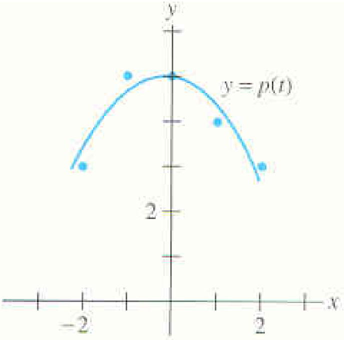


FIGURE 2
Approximation by a quadratic trend function.