7.8 Best Approximation and Least Squares

The minimum distance problem in $\mathbb{R}^n$

Given a subspace W and a vector $\mathbf{b}$ in $\mathbb{R}^n$, find a vector $\hat{\mathbf{w}}$ in W that is closest to $\mathbf{b}$ in the sense that $\|\mathbf{b} - \hat{\mathbf{w}}\| < \|\mathbf{b} - \mathbf{w}\|$ for every vector $\mathbf{w}$ in W that is distinct from $\hat{\mathbf{w}}$. Such a vector $\hat{\mathbf{w}}$, if exists, is called a best approximation to $\mathbf{b}$ from W.

Minimum distance problem

Theorem (Best approximation theorem)

If W is a subspace of $\mathbb{R}^n$, and $\mathbf{b}$ is a point in $\mathbb{R}^n$, then there is a unique best approximation to $\mathbf{b}$ from W, namely $\hat{w} = proj_W \mathbf{b}$.

Least squares solutions of linear system

Definition

If A is an $m \times n$ matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^m$, then a vector $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is called a best approximate solution or a least squares solution of $A\mathbf{x} = \mathbf{b}$ if

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

for all $\mathbf{x}$ in $\mathbb{R}^n$. The vector $\mathbf{b} - A\hat{\mathbf{x}}$ is called the least squares error vector, and the scalar $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the least squares error.

Least squares solutions of linear systems

The solutions of $A\mathbf{x} = \mathrm{proj}_{\mathrm{col}(A)}\mathbf{b}$ are the least squares solutions of $A\mathbf{x} = \mathbf{b}$.

Least squares solutions of linear systems

The least square solution of $A\mathbf{x} = \mathbf{b}$ are obtained by solving the equation $A^T A\mathbf{x} = A^T\mathbf{b}$, called the normal equation or normal system associated with $A\mathbf{x} = \mathbf{b}$.

Least squares solutions of linear systems

Theorem

(a) *The least squares solutions of a linear system* $A\mathbf{x} = \mathbf{b}$ *are the exact solution of the normal equation*

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

(b) *If* A *has full column rank, the normal equation has a unique solution, namely*

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

(c) *If* A *does not have full column rank, then the normal equation has infinitely many solutions, but there is a unique solution in the row space of* A*. Moreover, among all solutions of the normal equation, the solution in the row space of* A *has the smallest norm.*

Least squares solutions of linear systems

Theorem

A vector $\hat{\mathbf{x}}$ is a least sqare solution of $A\mathbf{x} = \mathbf{b}$ if and only if the error vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the column space of A.

Least squares solutions of linear systems

Example

Find the least squares solutions and least squares error for

$$\begin{array}{rcrcrcr} 3x_1 & + & 2x_2 & - & x_3 & = & 2 \\ x_1 & - & 4x_2 & + & 3x_3 & = & -2 \\ x_1 & + & 10x_2 & - & 7x_3 & = & 1 \end{array}$$

Fitting a curve to experimental data

Mathematical model

A common problem in experimental work is to obtain a mathematical relationship between two variables x and y by fitting a curve $y = f(x)$ of a specified form to a set of points in the plane that correspond to experimentally determined values of x and y. The curve $y = f(x)$ is called a mathematical model.

Least squares fit by linear functions

Linear models:

Least squares fit by linear functions

Example

Find the least squares line of best fit to the four points $(0, 1), (1, 3), (2, 4)$, and $(3, 4)$.

Least squares fit by higher-degree polynomials

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m$$

Least squares fit by higher-degree polynomials

Example

According to Newton's second law of motion, a body near the surface of the Earth falls vertically downward according to the equation

$$y = y_0 + v_0 t + \frac{1}{2} g t^2.$$

Find the least squares estimates of y_0, v_0, and g from the data:

Time t	0.1	0.2	0.3	0.4	0.5
Displacement y	-0.18	0.31	1.03	2.48	3.73