Least Squares and the Generalized Inverse

An important problem is to find a polynomial curve $y = f(x)$ which 'best fits' a given set of m data points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$. If $f(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$ then the given points lie on the curve $y = f(x)$ if and only if $y_i = f(x_i)$ for $1 \le i \le m$ which is is equivalent to the system of equations

$$\begin{gathered} a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{n-1} x_1^{n-1} = y_1 \\ a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{n-1} x_2^{n-1} = y_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_0 + a_1 x_m + a_2 x_m^2 + \cdots + a_{n-1} x_m^{n-1} = y_m \end{gathered}$$

having a solution $(a_0, a_1, \ldots, a_{n-1})$. In general this system does not have a solution and we therefore try to find $(a_0, a_1, \ldots, a_{n-1})$ which minimizes

$$\sum_{i=1}^{m} (y_i - f(x_i))^2.$$

Such a vector is called a **least squares solution** of our system of equations. If $m \ge n$ and at least n of the numbers $x_1, \ldots, x_m$ are distinct then we will show that there is only one least squares solution. This gives a unique polynomial curve $y = f(x)$ of degree n which best fits our data in the sense of least squares.

More generally, if A is a real $m \times n$ matrix and $Y \in \mathbb{R}^{n \times 1}$ then by a **least squares solution** of the system of equations $AX = Y$ we mean any $X \in \mathbb{R}^{n \times 1}$ which minimizes $||AX - Y||$, the norm being the usual norm in $\mathbb{R}^{m \times 1}$. It follows that AX is the orthogonal projection B of Y on the column space of A. Thus, the least squares solutions of $AX = Y$ are those column vectors X with $AX = B$. If rank(A) $\neq$ n then the a squares solution is not unique. In the above example of finding a polynomial curve of best fit the matrix

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix}$$

which is of rank n if $m \ge n$ and at least n of $x_1, \ldots, x_m$ are distinct.

If $AX = B$ and Z is the orthogonal projection of X on the column space of A^t then Z is the unique solution of $AX = B$ which is in the column space of A^t. Indeed, $X - Z \in$ Null(A) since the orthogonal complement of the column space of A^t is the null space of A so that $AZ = AX = B$. If Z_1 is any vector in the column space of A^t with $AZ_1 = B$ then $Z - Z_1$ is in the intersection of the null space of A and the column space of A^t and so must be the zero vector. Since $||Z|| \le ||X||$ with equality if and only if $X = Z$, we see that Z is also the unique vector X of smallest norm with $AX = B$.

The function $T : \mathbb{R}^{m \times 1} \to \mathbb{R}^{n \times 1}$ defined by $T(Y) = Z$ is a linear mapping and so $T = T_{A^+}$ for a unique $n \times m$ matrix A^+. This matrix is called the **generalized inverse** of A. It is characterized by the property that A^+Y is the least squares solution of $AX = Y$ of smallest norm.

Example 1. *Let's find the least squares solution of $AX = Y$ where*

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \quad \textit{and} \quad Y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

The first step is to find the orthogonal projection of Y on the column space of A. The matrix A has rank 2 and the first two columns A_1, A_2 are a basis for the column space of A. Since these columns are not orthogonal, we apply the Gram-Schmidt process to to them to get the orthogonal basis

$$A_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad A_2 - \frac{< A_2, A_1 >}{< A_1, A_1 >} A_1 = \frac{2}{3} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

Multiplying the second vector by $3/2$, we get the orthogonal basis

$$C_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

for the column space of A. The orthogonal projection of Y on the column space of A is then

$$B = \frac{< Y, C_1 >}{< C_1, C_1 >} C_1 + \frac{< Y, C_2 >}{< C_2, C_2 >} C_2 = \frac{1}{3} C_1 + \frac{1}{6} C_2 = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

The second step is to find a solution of $AX = B$. In this case

$$X = \begin{bmatrix} 0 \\ 0 \\ 1/4 \end{bmatrix}$$

is a least squares solution.

The third step is to find the orthogonal projection of this vector X on the column space of

$$A^t = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 0 & 2 \end{bmatrix}.$$

Here, the first two columns D_1, D_2 are already an orthogonal basis for the column space of A^t and so the orthogonal projection of X on the column space of A^t is

$$Z = \frac{< X, D_1 >}{< D_1, D_1 >} D_1 + \frac{< X, D_2 >}{< D_2, D_2 >} D_2 = \frac{1}{12} D_1 = \frac{1}{12} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

Note that $||Z|| = \frac{\sqrt{6}}{12} < \frac{1}{4} = ||X||$.

Problem 1. *Prove that* $A^+ = \frac{1}{12}\begin{bmatrix} 1 & 6 & 1 \\ 1 & -6 & 1 \\ 2 & 0 & 2 \end{bmatrix}$.

Example 2. *Find the line of best fit for the data points* $(1,1),(2,0),(3,4)$.

The required line $y = a + bx$ has the property that $X = [a,b]^t$ is the least squares solution of the system

$$\begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$$

If A_1, A_2 are the columns of the coefficient matrix A then

$$A_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad A_2 - \frac{< A_2, A_1 >}{< A_1, A_1 >} A_1 = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix} - \frac{14}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -11 \\ -2 \\ 13 \end{bmatrix}$$

are an orthogonal basis for the column space of A. Multiplying the second of these vectors by 3, we get the orthogonal basis

$$C_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} -11 \\ -2 \\ 13 \end{bmatrix}.$$

The orthogonal projection of $Y = [1,0,4]^t$ on the column space of A is

$$B = \frac{< Y, C_1 >}{< C_1, C_1 >} C_1 + \frac{< Y, C_2 >}{< C_2, C_2 >} C_2 = \frac{5}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{41}{294}\begin{bmatrix} -11 \\ -2 \\ 13 \end{bmatrix} = \frac{1}{98}\begin{bmatrix} 13 \\ 136 \\ 341 \end{bmatrix}.$$

Solving $\begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{98}\begin{bmatrix} 13 \\ 136 \\ 341 \end{bmatrix}$, we get $a = \frac{-14}{49}, b = \frac{41}{98}$.