Least Squares Identification

Given $y(x) = \sin(x)$, $-\pi \le x \le \pi$ find a, b, c, d such that

$$y_P(x) = a + bx + cx^2 + dx^3$$

is a "good" approximation to $y = \sin(x)$. Let

$$\mathbf{x} = \begin{bmatrix} x_1 = -\pi & x_2 & x_3 & \cdots & x_{n-1} & x_n = \pi \end{bmatrix}$$

and

$$\mathbf{y} = \sin(\mathbf{x}) = \begin{bmatrix} \sin(x_1) = 0 & \sin(x_2) & \sin(x_3) & \cdots & \sin(x_{n-1}) & \sin(x_n) = 0 \end{bmatrix}.$$

Then we want to find a, b, c, d such that

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix}}_{Y \in \mathbb{R}^n} = \begin{bmatrix} \sin(x_1) \\ \sin(x_2) \\ \sin(x_3) \\ \vdots \\ \sin(x_{n-1}) \\ \sin(x_n) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & x_{n-1}^3 \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}}_{W \in \mathbb{R}^{n \times 4}} \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}}_{P \in \mathbb{R}^4}$$

This can be written compactly as

$$Y = WP.$$

Multiply both sides by W^T to get

$$W^T Y = W^T W P$$

where

$$W^T W = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ x_1 & x_2 & x_3 & \cdots & x_{n-1} & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_{n-1}^2 & x_n^2 \\ x_1^3 & x_2^3 & x_3^3 & \cdots & x_{n-1}^3 & x_n^3 \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & x_{n-1}^3 \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix} \in \mathbb{R}^{4 \times 4}.$$

If the 4×4 matrix $W^T W$ is invertible then set

$$P^* = \begin{bmatrix} a^* \\ b^* \\ c^* \\ d^* \end{bmatrix} \triangleq (W^T W)^{-1} W^T Y$$

It turns out in this example that $W^T W$ will be invertible. However, it does *not* mean that

$$\sin(x) = a^* + b^* x + c^* x^2 + d^* x^3 \quad \text{for} \quad -\pi \le x \le \pi.$$

The interpretation of P^* is that it is the set of coefficients that minimizes the mean squared error which is defined by

$$e \triangleq \sum_{i=1}^{n} \Big(y_i - \ (a + bx_i + cx_i^2 + dx_i^3) \Big)^2 .$$

To see this let

$$E \triangleq \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix} - \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-1} & x_{n-1}^2 & x_{n-1}^3 \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = Y - WP$$

so that

$$\begin{aligned} e = E^T E &= (Y - WP)^T (Y - WP) \\ &= Y^T Y - P^T W^T Y - Y^T W P + P^T W^T W P \\ &= Y^T Y - \begin{bmatrix} a & b & c & d \end{bmatrix} W^T Y - Y^T W \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} + \begin{bmatrix} a & b & c & d \end{bmatrix} W^T W \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \\ &= Y^T Y - \begin{bmatrix} a & b & c & d \end{bmatrix} W^T Y - \begin{bmatrix} a & b & c & d \end{bmatrix} W^T Y + \begin{bmatrix} a & b & c & d \end{bmatrix} W^T W \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}. \end{aligned}$$

Then

$$\frac{\partial e}{\partial P} = \begin{bmatrix} \frac{\partial e}{\partial a} \\ \frac{\partial e}{\partial b} \\ \frac{\partial e}{\partial c} \\ \frac{\partial e}{\partial d} \end{bmatrix} = -2W^T Y + 2W^T W P$$

Setting this gradient to zero we solve

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = -W^T Y - (YW)^T + 2W^T W P = -2W^T Y + 2W^T W P$$

to obtain

$$P^* = (W^T W)^{-1} W^T Y.$$

To show this is a global minimum rewrite the error

$$\begin{aligned} e(P) = \underbrace{Y^T Y}_{R_Y} - P^T \underbrace{W^T Y}_{R_{WY}} - \underbrace{YW}_{R_{YW}} P + P^T \underbrace{W^T W}_{R_W} P &= R_Y - P^T R_{WY} - R_{YW} P + P^T R_W P \\ &= R_Y - R_{YW} R_W^{-1} R_{WY} + \left(P - R_W^{-1} R_{WY}\right)^T R_W \left(P - R_W^{-1} R_{WY}\right) \end{aligned}$$

where $R_{YW}^T = R_{YW}, (R_W^{-1})^T = R_W^{-1}$ and $R_W^T = R_W$ were used. R_W is symmetric and positive semidefinite, i.e., $z^T R_W z \geq 0$ for all $z \in \mathbb{R}^4$. As R_W turns out to be invertible it follows that R_W is positive definite, that is, $z^T R_W z \geq 0$ for all $z \in \mathbb{R}^4$ and $z^T R_W z = 0$ only if $z = 0 \in \mathbb{R}^4$. Thus

$$e(P) = R_Y - R_{YW} R_W^{-1} R_{WY} + \left(P - R_W^{-1} R_{WY}\right)^T R_W \left(P - R_W^{-1} R_{WY}\right)$$

is minimized for $P = P^* = R_W^{-1} R_{WY} = (W^T W)^{-1} W^T Y$.

Gradient Descent If this minimum is found by gradient descent then we first compute the gradient of the error which is

$$\begin{aligned}
\frac{\partial e}{\partial a} &= 2\sum_{i=1}^{n}\Big(y_i - \ (a + bx_i + cx_i^2 + dx_i^3)\Big) \\
\frac{\partial e}{\partial b} &= 2\sum_{i=1}^{n}\Big(y_i - \ (a + bx_i + cx_i^2 + dx_i^3)\Big)\, x_i \\
\frac{\partial e}{\partial c} &= 2\sum_{i=1}^{n}\Big(y_i - \ (a + bx_i + cx_i^2 + dx_i^3)\Big)\, x_i^2 \\
\frac{\partial e}{\partial d} &= 2\sum_{i=1}^{n}\Big(y_i - \ (a + bx_i + cx_i^2 + dx_i^3)\Big)\, x_i^3.
\end{aligned}$$

Then update according to

$$\begin{bmatrix} a^{(n)} \\ b^{(n)} \\ c^{(n)} \\ d^{(n)} \end{bmatrix} = \begin{bmatrix} a^{(n-1)} \\ b^{(n-1)} \\ c^{(n-1)} \\ d^{(n-1)} \end{bmatrix} - \eta \begin{bmatrix} \frac{\partial e}{\partial a} \\ \frac{\partial e}{\partial b} \\ \frac{\partial e}{\partial c} \\ \frac{\partial e}{\partial d} \end{bmatrix}\Bigg|_{P^{(n-1)}} \qquad \text{where } P^{(n-1)} \triangleq \begin{bmatrix} a^{(n-1)} \\ b^{(n-1)} \\ c^{(n-1)} \\ d^{(n-1)} \end{bmatrix}$$

Power Series Expansion A power series expansion of $\sin(x)$ about $x = 0$ is

$$\sin(x) = \sin(0) + \frac{d\sin(x)}{dx}\Big|_{x=0} x + \frac{d^2\sin(x)}{dx^2}\Big|_{x=0} \frac{x^2}{2!} + \frac{d^3\sin(x)}{dx^3}\Big|_{x=0} \frac{x^3}{3!} + \cdots$$

Taking only the first four terms we have

$$\sin(x) = 0 + x - \frac{x^2}{2!} + 0\frac{x^3}{3!}.$$

This power series expansion is motivated as way to approximate $\sin(x)$ about $x = 0$ where the above least squares approach is motivated as a way approximate $\sin(x)$ over the interval $[-\pi, \pi]$.