1 part

1.1 part-sec

A typical linear least squares problem involves fitting a curve to some data given in the form of a set of $m \in \mathbb{N}$ ordered pairs

$$D := \{(x_i, y_i) \in \mathbb{R}^2 \,|\, i = 1, 2, \ldots, m\}. \tag{1}$$

If the independent and dependent variables x and y, respectively, are known (or suspected) to be related like

$$y = \sum_{j=1}^{n} \alpha_j \varphi_j(x), \tag{2}$$

where the φ_j, $j = 1, 2, \ldots, n$, are $n \in \mathbb{N}$ known functions, the problem lies in the determination of the n constants $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$, where the number m of data points is greater than the number n of unknown parameters $\alpha_1, \ldots, \alpha_n$.

Defining column vectors

$$\mathbf{y} := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^{m\times 1}, \quad \boldsymbol{\alpha} := \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^{n\times 1}, \tag{3}$$

and the matrix

$$\mathbf{\Phi} := \begin{bmatrix} \varphi_1(x_i) & \varphi_2(x_1) & \cdots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(x_m) & \varphi_2(x_m) & \cdots & \varphi_n(x_m) \end{bmatrix} \in \mathbb{R}^{m\times n}, \tag{4}$$

the system

$$\forall i \in \{1,\ldots,m\} : y_i = \sum_{j=1}^{n} \alpha_j \varphi_j(x_i) \tag{5}$$

can be expressed in matrix form as

$$\mathbf{\Phi}\boldsymbol{\alpha} = \mathbf{y}. \tag{6}$$

The matrix $\mathbf{\Phi}$ can be uniquely associated with a linear transformation $T_\Phi : \mathbb{R}^n \to \mathbb{R}^m : \mathbf{x} \mapsto \mathbf{\Phi}\mathbf{x}$, where the range $T_\Phi(\mathbb{R}^n) \subset \mathbb{R}^m$ of T_Φ is just the column space

$$\mathrm{col}(\mathbf{\Phi}) := \{\mathbf{v} \in \mathbb{R}^{m\times 1} \mid \exists \mathbf{x} \in \mathbb{R}^{n\times 1} : \mathbf{v} = \mathbf{\Phi}\mathbf{x}\}$$

of the matrix $\mathbf{\Phi}$. Interpreting $\mathrm{col}(\mathbf{\Phi})$ as the space spanned by the columns of $\mathbf{\Phi}$, which are to be thought of as column vectors in $\mathbb{R}^{m\times 1}$, if $\mathbf{\Phi}$ has rank n then $\dim(\mathrm{col}(\mathbf{\Phi})) = n$, hence $T_\Phi(\mathbb{R}^n)$ is an n-dimensional subspace of $\mathbb{R}^m$. This means that, fixing $\mathbf{y} \in \mathbb{R}^{m\times 1}$, the equation (**??**) is unlikely to have a solution $\boldsymbol{\alpha} \in \mathbb{R}^{n\times 1}$. In most cases, the best that can be done is to find that

$\boldsymbol{\alpha}$ which minimizes the magnitude of the residual vector $\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\alpha}$, so that $\boldsymbol{\Phi}\boldsymbol{\alpha}$ is as close as possible to $\mathbf{y}$.

Geometrically, this corresponds to the problem of solving the matrix equation

$$\boldsymbol{\Phi}\boldsymbol{\alpha} = \mathrm{proj}_{T_\Phi(\mathbb{R}^n)}(\mathbf{y}) \,, \tag{7}$$

where $\mathrm{proj}_{T_\Phi(\mathbb{R}^n)}(\,\cdot\,) : \mathbb{R}^m \to T_\Phi(\mathbb{R}^n)$ is the orthogonal projection operator onto the subspace $T_\Phi(\mathbb{R}^n) \subset \mathbb{R}^m$. It is well known that this projection operator is uniquely associated with the projection matrix $\mathbf{P}_\Phi \in \mathbb{R}^{n\times m}$ so that

$$\forall \mathbf{y} \in \mathbb{R}^m : \mathrm{proj}_{T_\Phi(\mathbb{R}^n)}(\mathbf{y}) = \mathbf{P}_\Phi \mathbf{y},$$

where

$$\mathbf{P}_\Phi := \boldsymbol{\Phi}\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T, \tag{8}$$

and so the problem becomes finding a solution to the matrix equation

$$\boldsymbol{\Phi}\boldsymbol{\alpha} = \boldsymbol{\Phi}\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\mathbf{y}. \tag{9}$$

Multiplying both sides of equation (9) by $\boldsymbol{\Phi}^T$ and carrying out the appropriate simplifications yields the so-called *normal equations*

$$\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\alpha} = \boldsymbol{\Phi}^T\mathbf{y}. \tag{10}$$

References