4.4 Linear Least Squares: Matrix formulation

Consider n given pairs of data (x_i, y_i) and a model function $f(x, \vec{a})$, where for simplicity the set of p parameters (that the model function depends on) is written as a vector $\vec{a}$ (of length p). To obtain meaningful results for the parameters, obviously the condition $p \le n$ is necessary. The general least squares problem would be to minimize the objective function

$$\chi^2(\vec{a}) = \sum_{i=1}^{n} [y_i - f(x_i, \vec{a})]^2 \tag{4.15}$$

with respect to $\vec{a}$. As stated above, here we consider the case that f depends linearly on $\vec{a}$. This means that f can be expressed as a linear combination of p other functions F_k, with the parameters being the weighting coefficients:

$$f(x) = \sum_{k=1}^{p} a_k F_k(x)\,. \tag{4.16}$$

For the calculation of the objective function (here: χ^2, Eq. 4.15) we need these values at the x_i:

$$f(x_i) = \sum_{k=1}^{p} a_k F_k(x_i)\,. \tag{4.17}$$

This can be interpreted as a scalar product between the vector $\vec{a}$ and another vector $\vec{F}_i$ consisting of the functional values of the model functions F_k evaluated at x_i. Therefore, Eq. (4.17) can simply be written as

$$f(x_i) = \vec{a} \cdot \vec{F}_i\,. \tag{4.18}$$

(Note that the index i at $\vec{F}$ tells that there are n such vectors; it has nothing to do with the components of $\vec{F}$.)

Using the latter representation of f, the objective function, Eq. (4.15), for the case of a linear least squares problem can be written as

$$\chi^2(\vec{a}) = \sum_{i=1}^{n} (y_i - \vec{F}_i \cdot \vec{a})^2 =: \sum_{i=1}^{n} (y_i - m_i)^2. \tag{4.19}$$

This can be interpreted as the calculation of the length of a vector difference squared, where the "data vector" $\vec{y}$ consists of the y values of the given data and the "model vector" $\vec{m} = \mathsf{M}\vec{a}$ results from the multiplication of a matrix M with the vector $\vec{a}$, where the *rows* of M are the n transposed vectors $\vec{F}_i$ (i.e., M is a matrix with n rows and p columns); obviously, one has that $m_i = f(x_i)$. Then, one can finally write the objective function as

$$\chi^2(\vec{a}) = |\vec{y} - \mathsf{M}\vec{a}|^2. \tag{4.20}$$

This has to be minimized with respect to $\vec{a}$, a task which can be done analytically. It is known from the math lecture that the result is given by the following expression:

$$\vec{a}_{\text{min}} = \left(\mathsf{M}^{\mathrm{T}}\mathsf{M}\right)^{-1} \mathsf{M}^{\mathrm{T}}\vec{y}\,, \tag{4.21}$$

where M^{T} indicates the transpose of matrix M.

This expression can easily be evaluated in MATLAB, since it is routinely used for matrix division in case of an overdetermined system, i.e. where $n > p$. MATLAB has two variants of matrix division, which differ with respect to the order of the matrices (corresponding to the left and right inverse). Roughly speaking, it is like this: For given matrices A and B,
`X = B/A` denotes the solution to the matrix equation $\mathsf{XA} = \mathsf{B}$, whereas
`X = A\B` denotes the solution to the matrix equation $\mathsf{AX} = \mathsf{B}$.
For a non-degenerate square matrix A, these operations (`/` and `\`) correspond to a matrix multiplication with the inverse of A (but are not computed that way), whereas for non-square matrices with fewer columns than rows Eq. (4.21) is used.[4]

[4] Search the MATLAB help for `mldivide` to obtain more information.