1 Linear least-squares problem

If in a linear system $A\mathbf{c} = \mathbf{b}$, where A is an $n \times m$ matrix, $\mathbf{c}$ is an m-component vector and $\mathbf{b}$ is an n-component vector, the number of unknowns n is larger than the number of equations m, the system generally has no solution. However it is still possible to find the "best possible" solution which minimizes the Euclidean norm of the difference between $A\mathbf{c}$ and $\mathbf{b}$, $|A\mathbf{c} - \mathbf{b}|^2 \to \min$. The vector c which minimizes $|A\mathbf{c}-\mathbf{b}|^2$ is called the *least-squares solution* to the *least-squares problem* $Ac = b$.

The problem can be solved by QR-decomposition. The matrix A factorizes as $A = QR$, where Q is $n \times m$ matrix with orthogonal columns and R is an $m \times m$ upper triangular matrix. The Euclidean norm

$$\begin{aligned} |A\mathbf{c} - \mathbf{b}|^2 &= |QR\mathbf{c} - \mathbf{b}|^2 \\ &= |R\mathbf{c} - Q^T\mathbf{b}|^2 + |(1 - QQ^T)\mathbf{b}|^2 \\ &\geq |(1 - QQ^T)\mathbf{b}|^2 \end{aligned} \tag{1}$$

can then be minimized by solving an $m \times m$ system of linear equations

$$R\mathbf{c} - Q^T\mathbf{b} = 0 \tag{2}$$

by back-substitution.

1.1 Linear least-squares fit

Linear least-squares fit is a problem of fitting n data points $\{x_i, y_i \pm \sigma_i\}$ by a linear combination of m functions

$$F(x) = \sum_{k=1}^{m} c_k f_k(x). \tag{3}$$

The least-squares fit minimizes the square deviation (called χ^2)

$$\chi^2 = \sum_i \left(\frac{y_i - F(x_i)}{\sigma_i} \right)^2 \tag{4}$$

One can recognize the above problem of minimizing $|A\mathbf{c} - \mathbf{b}|^2$ where

$$A_{ik} = \frac{f_k(x_i)}{\sigma_i} \;,\; b_i = \frac{y_i}{\sigma_i} \;. \tag{5}$$

The formal solution is $\mathbf{c} = R^{-1}Q^T\mathbf{b}$, however in practice it is better to back-substitute the system $R\mathbf{c} = Q^T\mathbf{b}$.

1.1.1 The error of the least-squares fit: covariance matrix

The errors Δc_k of the obtained coefficients c_k can be estimated as

$$(\Delta c_k)^2 = \sum_i \left(\frac{\partial c_k}{\partial y_i} \sigma_i \right)^2 = \sum_i \left(\frac{\partial c_k}{\partial b_i} \right)^2 \tag{6}$$

The *error matrix*

$$E_{kq} \equiv \langle \Delta c_k \Delta c_q \rangle = \sum_i \frac{\partial c_k}{\partial b_i} \frac{\partial c_q}{\partial b_i} \tag{7}$$

(also called the *covariance matrix*) is then given as

$$E = \frac{\partial \mathbf{c}}{\partial \mathbf{b}} \frac{\partial \mathbf{c}}{\partial \mathbf{b}}^T = (R^T R)^{-1} = (A^T A)^{-1} \tag{8}$$

The diagonal elements of the covariance matrix are the squares of the errors in the corresponding coefficients $\Delta c_k = \sqrt{E_{kk}}$. The off-diagonal elements characterize correlations in the data: if the (normalized) off-diagonal element is close to one, $E_{kq}(E_{kk}E_{qq})^{-1/2} \approx 1$, then the coefficients c_k and c_q are correlated and can not be reliably estimated from the data.