1 Ordinary least squares problem

1.1 Introduction

A system of linear equations is considered *overdetermined* if there are more equations than unknown variables. If all equations of an overdetermined system are linearly independent, the system has no exact solution.

An *ordinary least-squares problem* (also called *linear least-squares problem*) is the problem of finding an approximate solution to an overdetermined linear system. It often arises in applications where a theoretical model is fitted to experimental data.

1.2 Ordinary least-squares problem

Consider a linear system

$$A\mathbf{c} = \mathbf{b} \, , \tag{1}$$

where A is a $n \times m$ matrix, $\mathbf{c}$ is an m-component vector of unknown variables and $\mathbf{b}$ is an n-component vector of the right-hand side terms. If the number of equations n is larger than the number of unknowns m, the system is overdetermined and generally has no solution.

However, it is still possible to find an approximate solution — the one where $A\mathbf{c}$ is only approximately equal $\mathbf{b}$ — in the sence that the Euclidean norm of the difference between $A\mathbf{c}$ and $\mathbf{b}$ is minimized,

$$\mathbf{c} : \ \min_{\mathbf{c}} \|A\mathbf{c} - \mathbf{b}\|^2 \, . \tag{2}$$

The problem (2) is called the ordinary least-squares problem and the vector $\mathbf{c}$ that minimizes $\|A\mathbf{c} - \mathbf{b}\|^2$ is called the *least-squares solution.*

1.3 Least-squares solution via QR-decomposition

The linear least-squares problem can be solved by QR-decomposition. The matrix A is factorized as $A = QR$, where Q is $n \times m$ matrix with orthogonal columns, $Q^{\mathsf{T}}Q = 1$, and R is an $m \times m$ upper triangular matrix. The Euclidean norm $\|A\mathbf{c} - \mathbf{b}\|^2$ can then be rewritten as

$$
\begin{aligned}
\|A\mathbf{c} - \mathbf{b}\|^2 &= \|QR\mathbf{c} - \mathbf{b}\|^2 \\
&= \|R\mathbf{c} - Q^{\mathsf{T}}\mathbf{b}\|^2 + \|(1 - QQ^{\mathsf{T}})\mathbf{b}\|^2 \\
&\geq \|(1 - QQ^{\mathsf{T}})\mathbf{b}\|^2 \, .
\end{aligned}
\tag{3}
$$

The term $\|(1 - QQ^{\mathsf{T}})\mathbf{b}\|^2$ is independent of the variables $\mathbf{c}$ and can not be reduced by their variations. However, the term $\|R\mathbf{c} - Q^{\mathsf{T}}\mathbf{b}\|^2$ can be reduced down to zero by solving the $m \times m$ system of linear equations

$$R\mathbf{c} = Q^{\mathsf{T}}\mathbf{b} \, . \tag{4}$$

The system is right-triangular and can be readily solved by back-substitution.

Thus the solution to the ordinary least-squares problem (2) is given by the solution of the triangular system (4).

1.4 Least-squares solution via Singular Value Decomposition

Under the *thin singular value decomposition* we shall understand a representation of a tall $n \times m$ ($n > m$) matrix A in the form

$$A = USV^\mathsf{T} , \tag{5}$$

where U is an orthogonal $n \times m$ matrix ($U^\mathsf{T}U = 1$), S is a square $m \times m$ diagonal matrix with non-negative real numbers on the diagonal (called singular values of matrix A), and V is a square $m \times m$ orthoginal matrix ($V^\mathsf{T}V = 1$).

Singular value decomposition can be used to solve our linear least squares problem $A\mathbf{c} = \mathbf{b}$. Indeed inserting the decomposition into the equation gives

$$USV^\mathsf{T}\mathbf{c} = \mathbf{b} . \tag{6}$$

Multiplying from the left with U^T and using the orthogonality of U one gets the projected equation

$$SV^\mathsf{T}\mathbf{c} = U^\mathsf{T}\mathbf{b} . \tag{7}$$

This is a square system which can be easily solved first by solving the diagonal system

$$S\mathbf{y} = U^\mathsf{T}\mathbf{b} \tag{8}$$

for $\mathbf{y}$ and then obtaining $\mathbf{c}$ as

$$\mathbf{c} = V\mathbf{y} . \tag{9}$$

The covariance matrix (22) can be calculated as

$$\Sigma = (A^\mathsf{T}A)^{-1} = (VS^2V^\mathsf{T})^{-1} = VS^{-2}V^T . \tag{10}$$

Singular value decomposition can be found by diagonalising the $m \times m$ symmetric positive semi-definite matrix $A^\mathsf{T}A$ (although this method is not the best for practical calculations, it would do as an educational tool),

$$A^\mathsf{T}A = VDV^\mathsf{T} , \tag{11}$$

where D is a diagonal matrix with eigenvalues of the matrix $A^\mathsf{T}A$ on the diagonal and V is the matrix of the corresponding eigenvectors. Indeed it is easy to check that the sought decomposition can the be constructed as $A = USV^\mathsf{T}$ where $S = D^{1/2}$, $U = AVD^{-1/2}$.

1.5   Ordinary least-squares curve fitting

Ordinary least-squares curve fitting is a problem of fitting n (experimental) data points $\{x_i, y_i \pm \Delta y_i\}_{i=1,\ldots,n}$, where Δy_i are experimental errors, by a linear combination, $F_{\mathbf{c}}$, of m functions $\{f_k(x)\}_{k=1,\ldots,m}$,

$$F_{\mathbf{c}}(x) = \sum_{k=1}^{m} c_k f_k(x) , \tag{12}$$

where the coefficients c_k are the fitting parameters.

The objective of the least-squares fit is to minimize the square deviation, called χ^2, between the fitting function $F_{\mathbf{c}}(x)$ and the experimental data [1],

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{F(x_i) - y_i}{\Delta y_i} \right)^2 . \tag{13}$$

where the individual deviations from experimental points are weighted with their inverse errors in order to promote contributions from the more precise measurements.

Minimization of χ^2 with respect to the coefficiendt c_k in (12) is apparently equivalent to the least-squares problem (2) where

$$A_{ik} = \frac{f_k(x_i)}{\Delta y_i} , \; b_i = \frac{y_i}{\Delta y_i} . \tag{14}$$

If QR = A is the QR-decomposition of the matrix A, the formal least-squares solution to the fitting problem is

$$\mathbf{c} = \mathrm{R}^{-1}\mathrm{Q}^{\mathsf{T}}\mathbf{b} . \tag{15}$$

In practice of course one rather back-substitutes the right-triangular system

$$\mathrm{R}\mathbf{c} = \mathrm{Q}^{\mathsf{T}}\mathbf{b} . \tag{16}$$

1.5.1   Variances and correlations of fitting parameters

Suppose δy_i is a small deviation of the measured value of the physical observable at hand from its exact value. The corresponding deviation δc_k of the fitting coefficient is then given as

$$\delta c_k = \sum_i \frac{\partial c_k}{\partial y_i} \delta y_i . \tag{17}$$

In a good experiment the deviations δy_i are statistically independent and distributed normally with the standard deviations Δy_i. The deviations (17) are then also distributed normally with *variances*

$$\langle \delta c_k \delta c_k \rangle = \sum_i \left(\frac{\partial c_k}{\partial y_i} \Delta y_i \right)^2 = \sum_i \left(\frac{\partial c_k}{\partial b_i} \right)^2 . \tag{18}$$

The standard errors in the fitting coefficients are then given as the square roots of variances,

$$\Delta c_k = \sqrt{\langle \delta c_k \delta c_k \rangle} = \sqrt{\sum_i \left(\frac{\partial c_k}{\partial b_i} \right)^2} \; . \tag{19}$$

The variances are diagonal elements of the *covariance matrix*, Σ, made of *covariances*,

$$\Sigma_{kq} \equiv \langle \delta c_k \delta c_q \rangle = \sum_i \frac{\partial c_k}{\partial b_i} \frac{\partial c_q}{\partial b_i} \; . \tag{20}$$

Covariances $\langle \delta c_k \delta c_q \rangle$ are measures of to what extent the coefficients c_k and c_q change together if the measured values y_i are varied. The normalized covariances,

$$\frac{\langle \delta c_k \delta c_q \rangle}{\sqrt{\langle \delta c_k \delta c_k \rangle \langle \delta c_q \delta c_q \rangle}} \tag{21}$$

are called *correlations*.

Using (20) and (15) the covariance matrix can be calculated as

$$\Sigma = \left(\frac{\partial \mathbf{c}}{\partial \mathbf{b}} \right) \left(\frac{\partial \mathbf{c}}{\partial \mathbf{b}} \right)^{\mathsf{T}} = R^{-1}(R^{-1})^{\mathsf{T}} = (R^{\mathsf{T}} R)^{-1} = (A^{\mathsf{T}} A)^{-1} \; . \tag{22}$$

The square roots of the diagonal elements of this matrix provide the estimates of the errors $\Delta \mathbf{c}$ of the fitting coefficients,

$$\Delta c_k = \sqrt{\Sigma_{kk}} \, \Big|_{k=1\ldots m} \; , \tag{23}$$

and the (normalized) off-diagonal elements provide the estimates of their correlations.

Table 1.5.1 shows how a Csharp implementation of the ordinary least squares fit via QR decomposition could look like.

An illustration of a fit is shown on Figure 1 where a polynomial is fitted to a set of data.

References

[1] R.M. Barnett. Review of particle properties. *Physical Review D*, 54:1, 1996.

```
static (vector,matrix) lsfit
(Func<double,double>[] fs, vector x, vector y, vector dy){
    int n = x.size, m=fs.Length;
    var A = new matrix(n,m);
    var b = new vector(n);
    for(int i=0;i<n;i++){
        b[i]=y[i]/dy[i];
        for(int k=0;k<m;k++)A[i,k]=fs[k](x[i])/dy[i];
        }
    var qra = new GSQR(A);
    vector c = qra.solve(b);
    var pinvA = qra.pinverse();
    var S = pinvA*pinvA.T;
    return (c,S);
    }
```

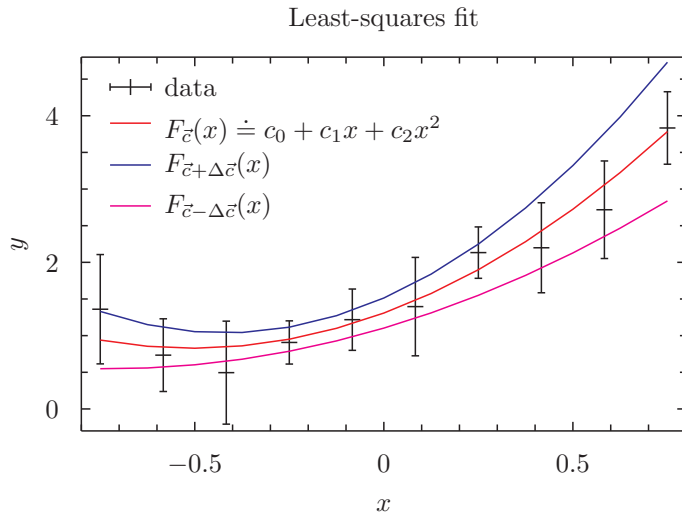Table 1: A Csharp implemetation of the ordinary least-squares fit.

Figure 1: Ordinary least squares fit of $F_{\mathbf{c}}(x) = c_1 + c_2x + c_3x^2$ to a set of data. Shown are fits with optimal coefficiens $\mathbf{c}$ as well as with $\mathbf{c} + \Delta\mathbf{c}$ and $\mathbf{c} - \Delta\mathbf{c}$.