Chapter 11
Data Fitting

Often a set of data points has to be fitted by a continuous function, either to obtain approximate function values in between the data points or to describe a functional relationship between two or more variables by a smooth curve, i.e. to fit a certain model to the data. If uncertainties of the data are negligibly small, an exact fit is possible, for instance with polynomials, spline functions or trigonometric functions (Chap. 2). If the uncertainties are considerable, a curve has to be constructed that fits the data points approximately. Consider a two-dimensional data set

$$(x_i, y_i) \quad i = 1 \ldots m \tag{11.1}$$

and a model function

$$f(x, a_1 \ldots a_n) \quad m \geq n \tag{11.2}$$

which depends on the variable x and $n \leq m$ additional parameters a_j. The errors of the fitting procedure are given by the residuals

$$r_i = y_i - f(x_i, a_1 \ldots a_n). \tag{11.3}$$

The parameters a_j have to be determined such, that the overall error is minimized, which in most practical cases is measured by the mean square difference[1]

$$S_{sd}(a_1 \ldots a_n) = \frac{1}{m} \sum_{i=1}^{m} r_i^2. \tag{11.4}$$

The optimal parameters are determined by solving the system of normal equations. If the model function depends linearly on the parameters, orthogonalization offers a numerically more stable method. The dimensionality of a data matrix can be reduced

[1] Minimization of the sum of absolute errors $\sum |r_i|$ is much more complicated.

P.O.J. Scherer, *Computational Physics*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-61088-7_11

with the help of singular value decomposition, which allows to approximate a matrix by another matrix of lower rank and is also useful for linear regression, especially if the columns of the data matrix are linearly dependent.

11.1 Least Square Fit

A (local) minimum of (11.4) corresponds to a stationary point with zero gradient. For n model parameters there are n, generally nonlinear, equations which have to be solved [123]. From the general condition

$$\frac{\partial S_{sd}}{\partial a_j} = 0 \quad j = 1 \ldots n \tag{11.5}$$

we find

$$\sum_{i=1}^{m} r_i \frac{\partial f(x_i, a_1 \ldots a_n)}{\partial a_j} = 0 \tag{11.6}$$

which can be solved with the methods discussed in Chap. 6. For instance, the Newton–Raphson method starts from a suitable initial guess of parameters

$$(a_1^0 \ldots a_n^0) \tag{11.7}$$

and tries to improve the fit iteratively by making small changes to the parameters

$$a_j^{s+1} = a_j^s + \Delta a_j^s. \tag{11.8}$$

The changes Δa_j^s are determined approximately by expanding the model function

$$f(x_i, a_1^{s+1} \ldots a_n^{s+1}) = f(x_i, a_1^s \ldots a_n^s) + \sum_{j=1}^{n} \frac{\partial f(x_i, a_1^s \ldots a_n^s)}{\partial a_j} \Delta a_j^s + \ldots \tag{11.9}$$

to approximate the new residuals

$$r_i^{s+1} = r_i^s - \sum_{j=1}^{n} \frac{\partial f(x_i, a_1^s \ldots a_m^s)}{\partial a_j} \Delta a_j^s \tag{11.10}$$

and the derivatives

$$\frac{\partial r_i^s}{\partial a_j} = -\frac{\partial f(x_i, a_1^s \ldots a_m^s)}{\partial a_j}. \tag{11.11}$$

Equation (11.6) now becomes

$$\sum_{i=1}^{m}\left(r_i^s - \sum_{j=1}^{n}\frac{\partial f(x_i)}{\partial a_j}\Delta a_j^s\right)\frac{\partial f(x_i)}{\partial a_k} \tag{11.12}$$

which is a system of n (usually overdetermined) linear equations for the Δa_j, the so-called normal equations:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\partial f(x_i)}{\partial a_j}\frac{\partial f(x_i)}{\partial a_k}\Delta a_j^s = \sum_{i=1}^{m} r_i^s\frac{\partial f(x_i)}{\partial a_k}. \tag{11.13}$$

With the definition

$$A_{kj} = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial f(x_i)}{\partial a_k}\frac{\partial f(x_i)}{\partial a_j} \tag{11.14}$$

$$b_k = \frac{1}{m}\sum_{i=1}^{m} y_i\frac{\partial f(x_i)}{\partial a_k} \tag{11.15}$$

the normal equations can be written as

$$\sum_{j=1}^{n} A_{kj}\Delta a_j = b_k. \tag{11.16}$$

11.1.1 Linear Least Square Fit

Especially important are model functions which depend linearly on all parameters (Fig. 11.1 shows an example which is discussed in problem 11.1)

$$f(x, a_1 \ldots a_n) = \sum_{j=1}^{n} a_j f_j(x). \tag{11.17}$$

The derivatives are

$$\frac{\partial f(x_i)}{\partial a_j} = f_j(x_i) \tag{11.18}$$

and the minimum of (11.4) is given by the solution of the normal equations

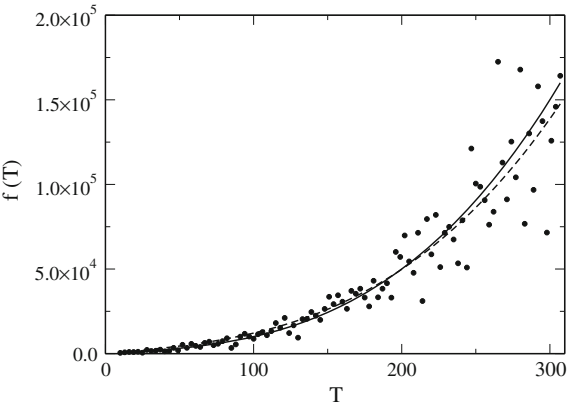

Fig. 11.1 (Least *square* fit) The polynomial $C(T) = aT + bT^3$ (*full curve*) is fitted to a set of data points which are distributed randomly around the "exact" values $C(T) = a_0T + b_0T^3$ (*dashed curve*). For more details see problem 11.1

$$\frac{1}{m}\sum_{j=1}^{n}\sum_{i=1}^{m} f_k(x_i)f_j(x_i)a_j = \frac{1}{m}\sum_{i=1}^{m} y_i f_k(x_i) \tag{11.19}$$

which for a linear fit problem become

$$\sum_{j=1}^{n} A_{kj}a_j = b_k \tag{11.20}$$

with

$$A_{kj} = \frac{1}{m}\sum_{i=1}^{m} f_k(x_i)f_j(x_i) \tag{11.21}$$

$$b_k = \frac{1}{m}\sum_{i=1}^{m} y_i f_k(x_i). \tag{11.22}$$

Example: Linear Regression

For a linear fit function

$$f(x) = a_0 + a_1 x \tag{11.23}$$

the mean square difference is

$$S_{sd} = \frac{1}{m}\sum_{i=1}^{m}(y_i - a_0 - a_1 x_i)^2 \tag{11.24}$$

and we have to solve the equations

$$0 = \frac{\partial S_{sd}}{\partial a_0} = \frac{1}{m}\sum_{i=1}^{m}(y_i - a_0 - a_1 x_i) = \overline{y} - a_0 - a_1\overline{x}$$

$$0 = \frac{\partial S_{sd}}{\partial a_1} = \frac{1}{m}\sum_{i=1}^{m}(y_i - a_0 - a_1 x_i)x_i = \overline{xy} - a_0\overline{x} - a_1\overline{x^2} \tag{11.25}$$

which can be done here with determinants

$$a_0 = \frac{\begin{vmatrix} \overline{y} & \overline{x} \\ \overline{xy} & \overline{x^2} \end{vmatrix}}{\begin{vmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{vmatrix}} = \frac{\overline{y}\,\overline{x^2} - \overline{x}\,\overline{xy}}{\overline{x^2} - \overline{x}^2} \tag{11.26}$$

$$a_1 = \frac{\begin{vmatrix} 1 & \overline{y} \\ \overline{x} & \overline{xy} \end{vmatrix}}{\begin{vmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{vmatrix}} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}. \tag{11.27}$$

11.1.2 Linear Least Square Fit with Orthogonalization

With the definitions

$$\mathbf{x} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \tag{11.28}$$

and the $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_m) & \cdots & f_n(x_m) \end{pmatrix} \tag{11.29}$$

the linear least square fit problem (11.20) can be formulated as a search for the minimum of

$$|A\mathbf{x} - \mathbf{b}| = \sqrt{(A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b})}. \tag{11.30}$$

In the last section we calculated the gradient

$$\frac{\partial |A\mathbf{x} - \mathbf{b}|^2}{\partial \mathbf{x}} = A^T(A\mathbf{x} - \mathbf{b}) + (A\mathbf{x} - \mathbf{b})^T A = 2A^T A\mathbf{x} - 2A^T\mathbf{b} \tag{11.31}$$

and solved the normal equations

$$A^T A\mathbf{x} = A^T\mathbf{b}. \tag{11.32}$$

This method can become numerically unstable. Alternatively we use orthogonalization of the n column vectors $\mathbf{a}_k$ of A to have

$$A = (\mathbf{a}_1 \cdots \mathbf{a}_n) = (\mathbf{q}_1 \cdots \mathbf{q}_n) \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix} \tag{11.33}$$

where $\mathbf{a}_k$ and $\mathbf{q}_k$ are now vectors of dimension m. Since the $\mathbf{q}_k$ are orthonormal $\mathbf{q}_i^T\mathbf{q}_k = \delta_{ik}$ we have

$$\begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{pmatrix} A = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}. \tag{11.34}$$

The $\mathbf{q}_k$ can be augmented by another $(m-n)$ vectors to provide an orthonormal basis of $\mathbb{R}^m$. These will not be needed explicitly. They are orthogonal to the first n vectors and hence to the column vectors of A. All vectors $\mathbf{q}_k$ together form an orthogonal matrix

$$Q = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n & \mathbf{q}_{n+1} & \cdots & \mathbf{q}_m \end{pmatrix} \tag{11.35}$$

and we can define the transformation of the matrix A:

$$\widetilde{A} = \begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \\ \mathbf{q}_{n+1}^T \\ \vdots \\ \mathbf{q}_m^T \end{pmatrix} (\mathbf{a}_1 \cdots \mathbf{a}_n) = Q^T A = \begin{pmatrix} R \\ 0 \end{pmatrix} \quad R = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \vdots \\ & & r_{nn} \end{pmatrix}. \tag{11.36}$$

The vector $\mathbf{b}$ transforms as

$$\tilde{\mathbf{b}} = Q^T\mathbf{b} = \begin{pmatrix} \mathbf{b}_u \\ \mathbf{b}_l \end{pmatrix} \quad \mathbf{b}_u = \begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{pmatrix} \mathbf{b} \quad \mathbf{b}_l = \begin{pmatrix} \mathbf{q}_{n+1}^T \\ \vdots \\ \mathbf{q}_m^T \end{pmatrix} \mathbf{b}. \tag{11.37}$$

Since the norm of a vector is not changed by unitary transformations

$$|\mathbf{b} - A\mathbf{x}| = \sqrt{(\mathbf{b}_u - R\mathbf{x})^2 + \mathbf{b}_l^2} \tag{11.38}$$

which is minimized if

$$R\mathbf{x} = \mathbf{b}_u. \tag{11.39}$$

The error of the fit is given by

$$|\mathbf{b} - A\mathbf{x}| = |\mathbf{b}_l|. \tag{11.40}$$

Example: Linear Regression

Consider again the fit function

$$f(x) = a_0 + a_1 x \tag{11.41}$$

for the measured data (x_i, y_i). The fit problem is to determine

$$\left| \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \right| = \min. \tag{11.42}$$

Orthogonalization of the column vectors

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \mathbf{a}_2 = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \tag{11.43}$$

with the Schmidt method gives:

$$r_{11} = \sqrt{m} \tag{11.44}$$

$$\mathbf{q}_1 = \begin{pmatrix} \frac{1}{\sqrt{m}} \\ \vdots \\ \frac{1}{\sqrt{m}} \end{pmatrix} \tag{11.45}$$

$$r_{12} = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} x_i = \sqrt{m}\,\overline{x} \tag{11.46}$$

$$\mathbf{b}_2 = (x_i - \overline{x}) \tag{11.47}$$

$$r_{22} = \sqrt{\sum (x_i - \overline{x})^2} = \sqrt{m}\sigma_x \tag{11.48}$$

$$\mathbf{q}_2 = (\frac{x_i - \overline{x}}{\sqrt{m}\sigma_x}). \tag{11.49}$$

Transformation of the right hand side gives

$$\begin{pmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \sqrt{m}\,\overline{y} \\ \sqrt{m}\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x} \end{pmatrix} \tag{11.50}$$

and we have to solve the system of linear equations

$$R\mathbf{x} = \begin{pmatrix} \sqrt{m} & \sqrt{m}\,\overline{x} \\ 0 & \sqrt{m}\sigma \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sqrt{m}\,\overline{y} \\ \sqrt{m}\frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x} \end{pmatrix}. \tag{11.51}$$

The solution

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{(x - \overline{x})^2}} \tag{11.52}$$

$$a_0 = \overline{y} - \overline{x}a_1 = \frac{\overline{y}\,\overline{x^2} - \overline{x}\,\overline{xy}}{\overline{(x - \overline{x})^2}} \tag{11.53}$$

coincides with the earlier results since

$$\overline{(x - \overline{x})^2} = \overline{x^2} - \overline{x}^2. \tag{11.54}$$

11.2 Singular Value Decomposition

Computational physics often has to deal with large amounts of data. Singular value decomposition is a very useful tool to reduce redundancies and to extract the most important information from data. It has been used for instance for image compression [124], it is very useful to extract the essential dynamics from molecular dynamics simulations [125, 126] and it is an essential tool of Bio-informatics [127].

11.2.1 Full Singular Value Decomposition

For $m \geq n$,[2] any real[3] $m \times n$ matrix A of rank $r \leq n$ can be decomposed into a product

$$A = U\, \Sigma\, V^T \tag{11.55}$$

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} v_{11} & \dots & v_{n1} \\ \vdots & \ddots & \vdots \\ v_{1n} & \dots & v_{nn} \end{pmatrix} \tag{11.56}$$

where U is a $m \times m$ orthogonal matrix, Σ is a $m \times n$ matrix, in which the upper part is a $n \times n$ diagonal matrix and V is an orthogonal $n \times n$ matrix.

The diagonal elements s_i are called singular values. Conventionally, they are sorted in descending order and the last $n - r$ of them are zero. For a square $n \times n$ matrix singular value decomposition (11.56) is equivalent to diagonalization

$$A = USU^T. \tag{11.57}$$

11.2.2 Reduced Singular Value Decomposition

We write

$$U = (U_n, U_{m-n}) \tag{11.58}$$

with the $m \times n$ matrix U_n and the $m \times (m - n)$ matrix U_{m-n} and

$$\Sigma = \begin{pmatrix} S \\ 0 \end{pmatrix} \tag{11.59}$$

with the diagonal $n \times n$ matrix S. The singular value decomposition then becomes

$$A = (U_n, U_{m-n}) \begin{pmatrix} S \\ 0 \end{pmatrix} V^T = U_n S V^T \tag{11.60}$$

which is known as reduced singular value decomposition. U_n (usually simply denoted by U) is not unitary but its column vectors, called the left singular vectors, are orthonormal

[2] Otherwise consider the transpose matrix.

[3] Generalization to complex matrices is straightforward.

$$\sum_{i=1}^{m} u_{i,r} u_{i,s} = \delta_{r,s} \tag{11.61}$$

as well as the column vectors of V which are called the right singular vectors

$$\sum_{i=1}^{n} v_{i,r} v_{i,s} = \delta_{r,s}. \tag{11.62}$$

Hence the products

$$U_n^T U_n = V^T V = E_n \tag{11.63}$$

give the $n \times n$ unit matrix.

In principle, U and V can be obtained from diagonalization of $A^T A$ and AA^T, since

$$A^T A = (V \Sigma^T U^T)(U \Sigma V^T) = V(S, 0) \begin{pmatrix} S \\ 0 \end{pmatrix} V^T = V S^2 V^T \tag{11.64}$$

$$AA^T = (U \Sigma V^T)(V \Sigma^T U^T) = U \begin{pmatrix} S \\ 0 \end{pmatrix} (S, 0) U^T = U_n S^2 U_n^T. \tag{11.65}$$

However, calculation of U by diagonalization is very inefficient, since usually only the first n rows are needed (i.e. U_n). To perform a reduced singular value decomposition, we first diagonalize

$$A^T A = V D V^T \tag{11.66}$$

which has positive eigenvalues $d_i \geq 0$, sorted in descending order and obtain the singular values

$$S = D^{1/2} = \begin{pmatrix} \sqrt{d_1} & & \\ & \ddots & \\ & & \sqrt{d_n} \end{pmatrix}. \tag{11.67}$$

Now we determine a matrix U such, that

$$A = USV^T \tag{11.68}$$

or, since V is unitary

$$Y = AV = US. \tag{11.69}$$

The last $n-r$ singular values are zero if $r < n$. Therefore we partition the matrices (indices denote the number of rows)

$$\left(Y_r \; 0_{n-r}\right) = \left(U_r \; U_{n-r}\right) \begin{pmatrix} S_r & \\ & 0_{n-r} \end{pmatrix} = \left(U_r S_r \; 0\right). \tag{11.70}$$

We retain only the first r columns and obtain a system of equations

$$\begin{pmatrix} y_{11} & \dots & y_{1r} \\ \vdots & \ddots & \vdots \\ y_{m1} & \dots & y_{mr} \end{pmatrix} = \begin{pmatrix} u_{11} & \dots & u_{1r} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mr} \end{pmatrix} \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_r \end{pmatrix} \tag{11.71}$$

which can be easily solved to give the first r rows of U

$$\begin{pmatrix} u_{11} & \dots & u_{1r} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mr} \end{pmatrix} = \begin{pmatrix} y_{11} & \dots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \dots & y_{mn} \end{pmatrix} \begin{pmatrix} s_1^{-1} & & \\ & \ddots & \\ & & s_r^{-1} \end{pmatrix}. \tag{11.72}$$

The remaining $n-r$ column vectors of U have to be orthogonal to the first r columns but are otherwise arbitrary. They can be obtained for instance by the Gram Schmidt method.

For larger matrices direct decomposition algorithms are available, for instance [128], which is based on a reduction to bidiagonal form and a variant of the *QL* algorithm as first introduced by Golub and Kahan [129].

11.2.3 Low Rank Matrix Approximation

Component-wise (11.60) reads

$$a_{i,j} = \sum_{k=1}^{r} u_{i,k} s_k v_{j,k}. \tag{11.73}$$

Approximations to A of lower rank are obtained by reducing the sum to only the largest singular values (the smaller singular values are replaced by zero). It can be shown [130] that the matrix of rank $l \leq r$

$$a_{i,j}^{(l)} = \sum_{k=1}^{l} u_{i,k} s_k v_{j,k} \tag{11.74}$$

is the rank-l matrix which minimizes

$$\sum_{i,j} |a_{i,j} - a_{i,j}^{(l)}|^2. \tag{11.75}$$

If only the largest singular value is taken into account, A is approximated by the rank-1 matrix

$$a_{i,j}^{(1)} = s_1 u_{i,1} v_{j,1}. \tag{11.76}$$

As an example, consider a $m \times n$ matrix

$$A = \begin{pmatrix} x_1(t_1) & \dots & x_n(t_1) \\ \vdots & & \vdots \\ x_1(t_m) & \dots & x_n(t_m) \end{pmatrix} \tag{11.77}$$

which contains the values of certain quantities $x_1 \dots x_n$ observed at different times $t_1 \dots t_m$. For convenience, we assume that the average values have been subtracted, such that $\sum_{j=1}^{m} x_i = 0$. Approximation (11.76) reduces the dimensionality to 1, i.e. a linear relation between the data. The i-th row of A,

$$\begin{pmatrix} x_1(t_i) & \dots & x_n(t_i) \end{pmatrix} \tag{11.78}$$

is approximated by

$$s_1 \, u_{i,1} \begin{pmatrix} v_{1,1} & \dots & v_{n,1} \end{pmatrix} \tag{11.79}$$

which describes a direct proportionality of different observables

$$\frac{1}{v_{j,1}} x_j(t_i) = \frac{1}{v_{k,1}} x_k(t_i). \tag{11.80}$$

According to (11.75) this linear relation minimizes the mean square distance between the data points (11.78) and their approximation (11.79).

Example: Linear approximation [131]

Consider the data matrix

$$A^T = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2.5 & 3.9 & 3.5 & 4.0 \end{pmatrix}. \tag{11.81}$$

First subtract the row averages

$$\overline{x} = 3 \quad \overline{y} = 2.98 \tag{11.82}$$

to obtain

$$A^T = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ -1.98 & -0.48 & 0.92 & 0.52 & 1.02 \end{pmatrix}. \tag{11.83}$$

Diagonalization of

$$A^T A = \begin{pmatrix} 10.00 & 7.00 \\ 7.00 & 6.308 \end{pmatrix} \tag{11.84}$$

gives the eigenvalues

$$d_1 = 15.393 \quad d_2 = 0.915 \tag{11.85}$$

and the eigenvectors

$$V = \begin{pmatrix} 0.792 & -0.610 \\ 0.610 & -0.792 \end{pmatrix}. \tag{11.86}$$

Since there are no zero singular values we find

$$U = AVS^{-1} = \begin{pmatrix} -0.181 & -0.380 \\ -0.070 & 0.252 \\ 0.036 & 0.797 \\ 0.072 & -0.217 \\ 0.143 & -0.451 \end{pmatrix}. \tag{11.87}$$

This gives the decomposition[4]

$$A = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} \begin{pmatrix} s_1 & \\ & s_2 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{pmatrix} = s_1 \mathbf{u}_1 \mathbf{v}_1^T + s_2 \mathbf{u}_2 \mathbf{v}_2^T$$

$$= \begin{pmatrix} -2.212 & -1.704 \\ -0.860 & -0.662 \\ 0.445 & 0.343 \\ 0.879 & 0.677 \\ 1.748 & 1.347 \end{pmatrix} + \begin{pmatrix} 0.212 & -0.276 \\ -0.140 & 0.182 \\ -0.445 & 0.577 \\ 0.121 & -0.157 \\ 0.252 & -0.327 \end{pmatrix}. \tag{11.88}$$

If we neglect the second contribution corresponding to the small singular value s_2 we have an approximation of the data matrix by a rank $-$ 1 matrix. The column vectors of the data matrix, denoted as $\mathbf{x}$ and $\mathbf{y}$, are approximated by

[4] $\mathbf{u}_i \mathbf{v}_i^T$ is the outer or matrix product of two vectors.

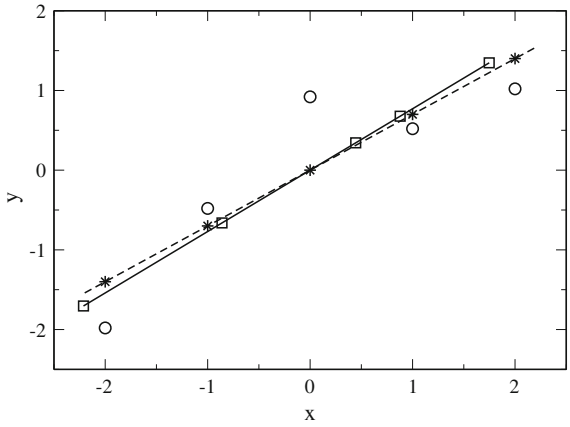

Fig. 11.2 (Linear approximation by singular value decomposition) The data set (11.81) is shown as *circles*. The linear approximation which is obtained by retaining only the dominant singular value is shown by the *squares* and the full line. It minimizes the mean *square* distance to the data points. *Stars* and the *dashed line* show the approximation by linear regression, which minimizes the mean *square* distance in *vertical* direction

$$\mathbf{x} = s_1 v_{11} \mathbf{u}_1 \quad \mathbf{y} = s_1 v_{21} \mathbf{u}_1 \tag{11.89}$$

which describes a proportionality between **x** and **y** (Fig. 11.2).

11.2.4 Linear Least Square Fit with Singular Value Decomposition

The singular value decomposition can be used for linear regression [131]. Consider a set of data, which have to be fitted to a linear function

$$y = c_0 + c_1 x_1 \cdots + c_n x_n \tag{11.90}$$

with the residual

$$r_i = c_0 + c_1 x_{i,1} \cdots + c_n x_{i,n} - y_i. \tag{11.91}$$

Let us subtract the averages

$$r_i - \overline{r} = c_1 (x_{i,1} - \overline{x}_1) \cdots + c_n (x_{i,n} - \overline{x}_n) - (y_i - \overline{y}) \tag{11.92}$$

which we write in matrix notation as

$$\begin{pmatrix} r_1 - \overline{r} \\ \vdots \\ r_m - \overline{r} \end{pmatrix} = \begin{pmatrix} x_{1,1} - \overline{x}_1 \ \ldots \ x_{1,n} - \overline{x}_n \\ \vdots \\ x_{m,1} - \overline{x}_1 \ \ldots \ x_{m,n} - \overline{x}_n \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} - \begin{pmatrix} y_1 - \overline{y} \\ \vdots \\ y_m - \overline{y} \end{pmatrix} \tag{11.93}$$

or shorter

$$\mathbf{r} = X\mathbf{c} - \mathbf{y}. \tag{11.94}$$

Now let us insert the full decomposition of X

$$\mathbf{r} = U\Sigma V^T\mathbf{c} - \mathbf{y}. \tag{11.95}$$

Since U is orthogonal

$$U^T\mathbf{r} = \Sigma V^T\mathbf{c} - U^T\mathbf{y} = \Sigma\mathbf{a} - \mathbf{b} \tag{11.96}$$

where we introduce the abbreviations

$$\mathbf{a} = V^T\mathbf{c} \quad \mathbf{b} = U^T\mathbf{y}. \tag{11.97}$$

The sum of squared residuals has the form

$$|\mathbf{r}|^2 = |U^T\mathbf{r}|^2 = \left| \begin{pmatrix} S_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{a}_r \\ \mathbf{a}_{n-r} \end{pmatrix} - \begin{pmatrix} \mathbf{b}_r \\ \mathbf{b}_{n-r} \end{pmatrix} \right|^2$$

$$= |S_r\mathbf{a}_r - \mathbf{b}_r|^2 + \mathbf{b}_{n-r}^2 \leq |S_r\mathbf{a}_r - \mathbf{b}_r|^2. \tag{11.98}$$

Hence $\mathbf{a}_{n-r}$ is arbitrary and one minimum of S_D is given by

$$\mathbf{a}_r = S_r^{-1}\mathbf{b}_r \quad \mathbf{a}_{n-r} = 0 \tag{11.99}$$

which can be written more compactly as

$$\mathbf{a} = \Sigma^+\mathbf{b} \tag{11.100}$$

with the Moore-Penrose pseudoinverse [132] of Σ

$$\Sigma^+ = \begin{pmatrix} s_1^{-1} & & & \\ & \ddots & & \\ & & s_r^{-1} & \\ & & & 0 \end{pmatrix}. \tag{11.101}$$

Finally we have

$$\mathbf{c} = V\Sigma^{+}U^{T}\mathbf{y} = X^{+}\mathbf{y} \tag{11.102}$$

where

$$X^{+} = V\Sigma^{+}U^{T} \tag{11.103}$$

is the Moore-Penrose pseudoinverse of X.

Example

The following data matrix has rank 2

$$X = \begin{pmatrix} -3 & -4 & -5 \\ -2 & -3 & -4 \\ 0 & 0 & 0 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 1.0 \\ 1.1 \\ 0 \\ -1.0 \\ -1.1 \end{pmatrix}. \tag{11.104}$$

A solution to the linear fit problem is given by

$$\mathbf{c} = X^{+}\mathbf{y} = \begin{pmatrix} -0.917 & 1.167 & 0 & -1.167 & 0.917 \\ -0.167 & 0.167 & 0 & -0.167 & 0.167 \\ 0.583 & -0.833 & 0 & 0.833 & -0.583 \end{pmatrix} \begin{pmatrix} 1.0 \\ 1.1 \\ 0 \\ -1.0 \\ -1.1 \end{pmatrix} = \begin{pmatrix} 0.525 \\ 0.000 \\ -0.525 \end{pmatrix}. \tag{11.105}$$

The fit function is

$$y = 0.525(x_1 - x_3) \tag{11.106}$$

and the residuals are

$$X\mathbf{c} - \mathbf{y} = \begin{pmatrix} 0.05 \\ -0.05 \\ 0 \\ -0.05 \\ 0.05 \end{pmatrix}. \tag{11.107}$$

11.2.5 Singular and Underdetermined Linear Systems of Equations

SVD is also very useful to solve linear systems with a singular or almost singular matrix. Consider a system

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \tag{11.108}$$

with $n > m$, i.e. more unknowns than equations. SVD transforms this system into

$$\begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} \begin{pmatrix} s_1 & & 0 \dots 0 \\ & \ddots & \vdots \\ & s_m & 0 \dots 0 \end{pmatrix} \begin{pmatrix} v_{11} & \dots & v_{n1} \\ \vdots & \ddots & \vdots \\ v_{1n} & \dots & v_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}. \tag{11.109}$$

Substituting

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} v_{11} & \dots & v_{n1} \\ \vdots & \ddots & \vdots \\ v_{1n} & \dots & v_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ and } \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix}^{-1} \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \tag{11.110}$$

it remains to solve

$$\begin{pmatrix} s_1 & & 0 \dots 0 \\ & \ddots & \vdots \\ & s_m & 0 \dots 0 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}. \tag{11.111}$$

For $y_1 \dots y_m$ the solution is

$$y_i = s_i^{-1} c_i \quad i = 1, \dots m \tag{11.112}$$

whereas $y_{m+1} \dots y_n$ are arbitrary and parametrize the solution manifold. Back substitution gives

$$\mathbf{x} = \mathbf{x}_p + \mathbf{z} \tag{11.113}$$

with the particular solution

$$\mathbf{x}_p = \begin{pmatrix} v_{11} & \dots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \dots & v_{nm} \end{pmatrix} \begin{pmatrix} s_1^{-1} & & \\ & \ddots & \\ & & s_m^{-1} \end{pmatrix} \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix}^{-1} \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \tag{11.114}$$

and

$$\mathbf{z} = \begin{pmatrix} v_{1,m+1} & \dots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n,m+11} & \dots & v_{nn} \end{pmatrix} \begin{pmatrix} y_{m+1} \\ \vdots \\ y_n \end{pmatrix} \tag{11.115}$$

which is in the nullspace of A

$$A\mathbf{z} = U\Sigma V^T V \begin{pmatrix} 0 \\ y_{m+1} \\ \vdots \\ y_n \end{pmatrix} = U \begin{pmatrix} S & 0 \end{pmatrix} \begin{pmatrix} 0 \\ y_{m+1} \\ \vdots \\ y_n \end{pmatrix} = 0. \tag{11.116}$$

If $m - r$ singular values are zero (or if the smallest singular values are set to zero) (11.111) becomes

$$\begin{pmatrix} s_1 & & & 0 \dots 0 \\ & \ddots & & \vdots \\ & & s_r & 0 \dots 0 \\ 0 & \dots & 0 & 0 \dots 0 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & 0 \dots 0 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} \tag{11.117}$$

which gives on the one hand

$$y_i = s_i^{-1} c_i \quad i = 1 \dots r \tag{11.118}$$

$$y_i = \text{arbitrary} \quad i = r + 1 \dots m \tag{11.119}$$

but also requires

$$c_i = 0 \quad i = r + 1 \dots m. \tag{11.120}$$

If this condition is not fulfilled, the equations are contradictory and no solution exists (e.g. if two rows of A are the same but the corresponding elements of **b** are different).

Problems

Problem 11.1 Least Square Fit

At temperatures far below Debye and Fermi temperatures the specific heat of a metal contains contributions from electrons and lattice vibrations and can be described by

$$C(T) = aT + bT^3. \tag{11.121}$$

The computer experiment generates data

$$T_j = T_0 + j\Delta t \tag{11.122}$$

$$C_j = (a_0 T_j + b_0 T_j^3)(1 + \varepsilon_j) \tag{11.123}$$

with relative error

$$\varepsilon_j = \varepsilon\, \xi_j. \tag{11.124}$$

Random numbers ξ_j are taken from a Gaussian normal distribution function (Sect. 9.2.6).
The fit parameters a, b are determined from minimization of the sum of squares

$$S = \frac{1}{n} \sum_{j=1}^{n} (C_j - aT_i - bT_i^3)^2. \tag{11.125}$$

Compare the "true values" a_0, b_0 with the fitted values a, b.