Chapter 10
Data Fitting

Often a set of data points have 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

$$(x_i, y_i) \quad i = 1 \dots N \tag{10.1}$$

and a model function

$$f(x, a_1 \dots a_m) \quad m \le N \tag{10.2}$$

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

$$r_i = y_i - f(x_i, a_1 \dots a_m). \tag{10.3}$$

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]

$$S(a_1 \dots a_m) = \frac{1}{N} \sum_{i=1}^{N} r_i^2. \tag{10.4}$$

10.1 Least Square Fit

A (local) minimum of (10.4) corresponds to a stationary point with zero gradient. For m model parameters there are m, generally nonlinear, equations which have to be solved. From the general condition

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

P.O.J. Scherer, *Computational Physics*, DOI 10.1007/978-3-642-13990-1_10,

$$\frac{\partial S}{\partial a_j} = 0 \quad j = 1 \ldots m \tag{10.5}$$

we find

$$\sum_{i=1}^{N} r_i \frac{\partial f(x_i, a_1 \ldots a_m)}{\partial a_j} = 0. \tag{10.6}$$

In principle, the methods discussed in (6.2) are applicable. For instance, the Newton–Raphson method starts from a suitable initial guess of parameters

$$(a_1^0 \ldots a_m^0) \tag{10.7}$$

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

$$a_j^{n+1} = a_j^n + \Delta a_j^n. \tag{10.8}$$

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

$$f(x_i, a_1^{n+1} \ldots a_m^{n+1}) = f(x_i, a_1^n \ldots a_m^n) + \sum_{j=1}^{m} \frac{\partial f(x_i, a_1^n \ldots a_m^n)}{\partial a_j} \Delta a_j^n + \ldots \tag{10.9}$$

to approximate the new residuals by

$$r_i^{n+1} = r_i^n - \sum_{j=1}^{m} \frac{\partial f(x_i, a_1^n \ldots a_m^n)}{\partial a_j} \Delta a_j^n \tag{10.10}$$

and the derivatives by

$$\frac{\partial r_i^n}{\partial a_j} = -\frac{\partial f(x_i, a_1^n \ldots a_m^n)}{\partial a_j}. \tag{10.11}$$

Equation (10.6) now becomes

$$\sum_{i=1}^{N} \left(r_i^n - \sum_{j=1}^{m} \frac{\partial f(x_i)}{\partial a_j} \Delta a_j^n \right) \frac{\partial f(x_i)}{\partial a_k} \tag{10.12}$$

which is a system of m linear equations for the Δa_j, the so-called normal equations:

$$\sum_{ij} \frac{\partial f(x_i)}{\partial a_j} \frac{\partial f(x_i)}{\partial a_k} \Delta a_j^n = \sum_{i=1}^{N} r_i^n \frac{\partial f(x_i)}{\partial a_k}. \tag{10.13}$$

With

$$A_{kj} = \frac{1}{n}\sum_{i=1}^{N}\frac{\partial f(x_i)}{\partial a_k}\frac{\partial f(x_i)}{\partial a_j} \tag{10.14}$$

$$b_k = \frac{1}{n}\sum_{i=1}^{N} y_i \frac{\partial f(x_i)}{\partial a_k} \tag{10.15}$$

the normal equations can be written as

$$\sum_{j=1}^{p} A_{kj}\Delta a_j = b_k. \tag{10.16}$$

10.1.1 Linear Least Square Fit

Especially important are model functions which depend linearly on the parameters

$$f(x, a_1 \dots a_m) = \sum_{j=1}^{m} a_j f_j(x). \tag{10.17}$$

The derivatives are simply

$$\frac{\partial f(x_i)}{\partial a_j} = f_j(x_i). \tag{10.18}$$

The minimum of (10.4) is now determined by the normal equations

$$\frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n} f_k(x_i) f_j(x_i) a_j = \frac{1}{n}\sum y_i f_k(x_i) \tag{10.19}$$

which become

$$\sum_{j=1}^{p} A_{kj} a_j = b_k \tag{10.20}$$

with

$$A_{kj} = \frac{1}{n}\sum_{i=1}^{n} f_k(x_i) f_j(x_i) \tag{10.21}$$

$$b_k = \frac{1}{n}\sum_{i=1}^{n} y_i f_k(x_i). \tag{10.22}$$

Example: Linear Regression

For a linear fit function

$$f(x) = a_0 + a_1 x \tag{10.23}$$

we have

$$S = \frac{1}{n}\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2 \tag{10.24}$$

and we have to solve the equations

$$\begin{aligned} 0 &= \frac{\partial S}{\partial a_0} = \frac{1}{n}\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i) = \bar{y} - a_0 - a_1\bar{x} \\ 0 &= \frac{\partial S}{\partial a_1} = \frac{1}{n}\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)x_i = \overline{yx} - a_0\bar{x} - a_1\overline{x^2} \end{aligned} \tag{10.25}$$

which can be done in this simple case using determinants:

$$a_0 = \frac{\begin{vmatrix} \bar{y} & \bar{x} \\ \overline{xy} & \overline{x^2} \end{vmatrix}}{\begin{vmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{vmatrix}} = \frac{\bar{y}\overline{x^2} - \bar{x}\overline{xy}}{\overline{x^2} - \bar{x}^2} \tag{10.26}$$

$$a_1 = \frac{\begin{vmatrix} 1 & \bar{y} \\ \bar{x} & \overline{xy} \end{vmatrix}}{\begin{vmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{vmatrix}} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}. \tag{10.27}$$

10.1.2 Least Square Fit Using Orthogonalization

The problem to solve the linearized problem (10.12) can be formulated with the definitions

$$\mathbf{x} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \tag{10.28}$$

and the $N \times m$ matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{Nm} \end{pmatrix} = \begin{pmatrix} \frac{\partial f(x_1)}{\partial a_1} & \cdots & \frac{\partial f(x_1)}{\partial a_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x_N)}{\partial a_1} & \cdots & \frac{\partial f(x_N)}{\partial a_m} \end{pmatrix} \tag{10.29}$$

as the search for the minimum of

$$S = |A\mathbf{x} - \mathbf{b}| = \sqrt{(A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b})}. \tag{10.30}$$

In the last section we calculated

$$\frac{\partial S^2}{\partial \mathbf{x}} = A^T(A\mathbf{x} - \mathbf{b}) + (A\mathbf{x} - \mathbf{b})^T A = 2A^T A\mathbf{x} - 2A^T\mathbf{b} \tag{10.31}$$

and solved the system of linear equations[2]

$$A^T A\mathbf{x} = A^T\mathbf{b}. \tag{10.32}$$

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

$$A = (\mathbf{a}_1 \cdots \mathbf{a}_m) = (\mathbf{q}_1 \cdots \mathbf{q}_m) \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ & r_{22} & \cdots & r_{2m} \\ & & \ddots & \vdots \\ & & & r_{mm} \end{pmatrix}, \tag{10.33}$$

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

$$\begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_m^T \end{pmatrix} A = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ & r_{22} & \cdots & r_{2m} \\ & & \ddots & \vdots \\ & & & r_{mm} \end{pmatrix}. \tag{10.34}$$

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

$$Q = \left(\mathbf{q}_1 \ \cdots \ \mathbf{q}_m \ \mathbf{q}_{m+1} \ \cdots \ \mathbf{q}_N\right) \tag{10.35}$$

and we can define the transformation of the matrix A:

$$\widetilde{A} = \begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_m^T \\ \mathbf{q}_{m+1}^T \\ \vdots \\ \mathbf{q}_N^T \end{pmatrix} (\mathbf{a}_1 \cdots \mathbf{a}_m) = Q^H A = \begin{pmatrix} R \\ 0 \end{pmatrix} \quad R = \begin{pmatrix} r_{11} & \cdots & r_{1N} \\ & \ddots & \vdots \\ & & r_{NN} \end{pmatrix}. \tag{10.36}$$

[2] Also known as normal equations.

The vector **b** transforms as

$$\widetilde{\mathbf{b}} = Q^H \mathbf{b} = \begin{pmatrix} \mathbf{b}_u \\ \mathbf{b}_l \end{pmatrix} \quad \mathbf{b}_u = \begin{pmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_m^T \end{pmatrix} \mathbf{b} \quad \mathbf{b}_l = \begin{pmatrix} \mathbf{q}_{m+1}^T \\ \vdots \\ \mathbf{q}_n^T \end{pmatrix} \mathbf{b}. \tag{10.37}$$

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

$$|\mathbf{b} - A\mathbf{x}| = \sqrt{(\mathbf{b}_u - R\mathbf{x})^2 + \mathbf{b}_l^2} \tag{10.38}$$

which is minimized if

$$R\mathbf{x} = \mathbf{b}_u. \tag{10.39}$$

The error of the fit is given by

$$|\mathbf{b} - A\mathbf{x}| = |\mathbf{b}_l|. \tag{10.40}$$

Example: Linear Regression

Consider again the fit function

$$f(x) = a_0 + a_1 x \tag{10.41}$$

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

$$\left| \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \right| = \min. \tag{10.42}$$

Orthogonalization of the column vectors

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \mathbf{a}_2 = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \tag{10.43}$$

with the Schmidt method gives

$$r_{11} = \sqrt{N} \tag{10.44}$$

$$\mathbf{q}_1 = \begin{pmatrix} \frac{1}{\sqrt{N}} \\ \vdots \\ \frac{1}{\sqrt{N}} \end{pmatrix} \tag{10.45}$$

$$r_{12} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i = \sqrt{N}\bar{x} \tag{10.46}$$

$$\mathbf{b}_2 = (x_i - \bar{x}) \tag{10.47}$$

$$r_{22} = \sqrt{\sum (x_i - \bar{x})^2} = \sqrt{N}\sigma_x \tag{10.48}$$

$$\mathbf{q}_2 = \left(\frac{x_i - \bar{x}}{\sqrt{N}\sigma_x} \right). \tag{10.49}$$

Transformation of the right-hand side gives

$$\begin{pmatrix} \mathbf{q}_1^t \\ \mathbf{q}_2^t \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \sqrt{N}\bar{y} \\ \sqrt{N}\frac{\overline{yx} - \bar{x}\bar{y}}{\sigma_x} \end{pmatrix} \tag{10.50}$$

and we have to solve the system of linear equations

$$R\mathbf{x} = \begin{pmatrix} \sqrt{N} & \sqrt{N}\bar{x} \\ 0 & \sqrt{N}\sigma \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sqrt{N}\bar{y} \\ \sqrt{N}\frac{\overline{yx} - \bar{x}\bar{y}}{\sigma_x} \end{pmatrix}. \tag{10.51}$$

The solution

$$a_1 = \frac{\overline{yx}/\bar{x}\bar{y}}{\overline{(x - \bar{x})^2}} \tag{10.52}$$

$$a_0 = \bar{y} - \bar{x}a_1 = \frac{\bar{y}\overline{x^2} - \bar{x}\overline{xy}}{\overline{(x - \bar{x})^2}} \tag{10.53}$$

coincides with the earlier results since

$$\overline{(x - \bar{x})^2} = \overline{x^2} - \bar{x}^2. \tag{10.54}$$

10.2 Singular Value Decomposition

Computational physics often has to deal with large amounts of data. The method of singular value decomposition is very useful to reduce redundancies and to extract the most important information from data. It has been used, for instance, for image compression [39], it is very useful to extract the essential dynamics from molecular dynamics simulations [40, 41], and it is an essential tool of bio-informatics [42]. The general idea is to approximate an $m \times n$ matrix of data of rank r $(m \geq n \geq r)$ by a matrix with smaller rank $l < r$. This can be formally achieved by the decomposition

$$X = U\,S\,V^{\mathrm{T}}$$
$$\begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix}$$
$$= \begin{pmatrix} u_{11} & \dots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mn} \end{pmatrix} \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \end{pmatrix} \begin{pmatrix} v_{11} & \dots & v_{n1} \\ \vdots & \ddots & \vdots \\ v_{1n} & \dots & v_{nn} \end{pmatrix}, \tag{10.55}$$

where U is an $m \times n$ matrix, S is an $n \times n$ diagonal matrix, and V is another $n \times n$ matrix. The column vectors of U are called the left singular vectors and are orthonormal

$$\sum_{i=1}^{m} u_{i,r} u_{i,s} = \delta_{r,s} \tag{10.56}$$

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

$$\sum_{i=1}^{n} v_{i,r} v_{i,s} = \delta_{r,s}. \tag{10.57}$$

The diagonal elements of S are called the singular values. For a square $n \times n$ matrix Eq. (10.55) is equivalent to diagonalization:

$$X = USU^{\mathrm{T}}. \tag{10.58}$$

Component wise (10.55) reads[3]

$$x_{r,s} = \sum_{i=1}^{r} u_{r,i} s_i v_{s,i}. \tag{10.59}$$

Approximations to X of lower rank are obtained by reducing the sum to only the largest singular values. It can be shown that the matrix of rank $l \le r$

$$x_{r,s}^{(l)} = \sum_{i=1}^{l} u_{r,i} s_i v_{s,i} \tag{10.60}$$

is the rank-l matrix which minimizes

[3] The singular values are ordered in descending order and the last $(n - r)$ singular values are zero.

$$\sum_{r,s} |x_{r,s} - x^{(l)}_{r,s}|^2. \tag{10.61}$$

One way to perform the singular value decomposition is to consider

$$X^{\mathrm{T}}X = (VSU^{\mathrm{T}})(USV^{\mathrm{T}}) = VS^2V^{\mathrm{T}}. \tag{10.62}$$

Hence V and the singular values can be obtained from diagonalization of the square $n \times n$ matrix:

$$X^{\mathrm{T}}X = V\,D\,V^{\mathrm{T}}, \tag{10.63}$$

where the (non-negative) eigenvalues d_i are ordered in descending order. The singular values are

$$S = D^{1/2} = \begin{pmatrix} \sqrt{d_1} & & \\ & \ddots & \\ & & \sqrt{d_n} \end{pmatrix}. \tag{10.64}$$

Now we have to determine a matrix U such that

$$X = USV^{\mathrm{T}} \tag{10.65}$$

$$XV = US. \tag{10.66}$$

We have to be careful since some of the s_i might be zero. Therefore we consider only the nonzero singular values and retain from the equation

$$\begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix} \begin{pmatrix} v_{11} & \dots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \dots & v_{nn} \end{pmatrix}$$
$$= \begin{pmatrix} u_{11} & \dots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mn} \end{pmatrix} \begin{pmatrix} s_1 & & & & & \\ & \ddots & & & & \\ & & s_r & & & \\ & & & 0_{r+1} & & \\ & & & & \ddots & \\ & & & & & 0_n \end{pmatrix} \tag{10.67}$$

only the first r columns of the matrix U

$$\begin{pmatrix} xv_{11} & \dots & xv_{1r} \\ \vdots & \ddots & \vdots \\ xv_{m1} & \dots & xv_{mr} \end{pmatrix} = \begin{pmatrix} u_{11} & \dots & u_{1r} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mr} \end{pmatrix} \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_r \end{pmatrix} \tag{10.68}$$

which can be solved by

$$\begin{pmatrix} u_{11} & \dots & u_{1r} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mr} \end{pmatrix} = \begin{pmatrix} xv_{11} & \dots & xv_{1n} \\ \vdots & \ddots & \vdots \\ xv_{m1} & \dots & xv_{mn} \end{pmatrix} \begin{pmatrix} s_1^{-1} & & \\ & \ddots & \\ & & s_r^{-1} \end{pmatrix}. \tag{10.69}$$

The remaining 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.

Example:

Consider the data matrix

$$X^{\mathrm{T}} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2.1 & 3.05 & 3.9 & 4.8 \end{pmatrix}. \tag{10.70}$$

Diagonalization of

$$X^{\mathrm{T}}X = \begin{pmatrix} 55 & 53.95 \\ 53.95 & 52.9625 \end{pmatrix}$$

gives the eigenvalues

$$d_1 = 107.94 \quad d_2 = 0.0216 \tag{10.71}$$

and the eigenvectors

$$V = \begin{pmatrix} -0.714 & 0.7004 \\ -0.7004 & -0.714 \end{pmatrix}. \tag{10.72}$$

Since there are no zero singular values we find

$$\begin{aligned} U &= XVS^{-1} \\ &= \begin{pmatrix} -1.414 & -0.013 \\ -2.898 & -0.098 \\ -4.277 & -0.076 \\ -5.587 & 0.018 \\ -6.931 & 0.076 \end{pmatrix} \begin{pmatrix} 10.39 & \\ & 0.147 \end{pmatrix}^{-1} = \begin{pmatrix} -0.136 & -0.091 \\ -0.279 & -0.667 \\ -0.412 & -0.515 \\ -0.538 & 0.122 \\ -0.667 & 0.517 \end{pmatrix}. \end{aligned} \tag{10.73}$$

This gives the decomposition[4]

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

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

$$= \begin{pmatrix} 1.009 & 0.990 \\ 2.069 & 2.030 \\ 3.053 & 2.996 \\ 3.987 & 3.913 \\ 4.947 & 4.854 \end{pmatrix} + \begin{pmatrix} -0.009 & 0.0095 \\ -0.069 & 0.0700 \\ -0.053 & 0.0541 \\ 0.013 & -0.0128 \\ 0.053 & -0.0542 \end{pmatrix}. \quad (10.74)$$

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. If the column vectors of the data matrix are denoted as $\mathbf{x}$ and $\mathbf{y}$ they are approximated by

$$\mathbf{x} = s_1 v_{11} \mathbf{u}_1 \quad \mathbf{y} = s_1 v_{21} \mathbf{u}_1 \quad (10.75)$$

which is a linear relationship between $\mathbf{x}$ and $\mathbf{y}$ (Fig. 10.1)

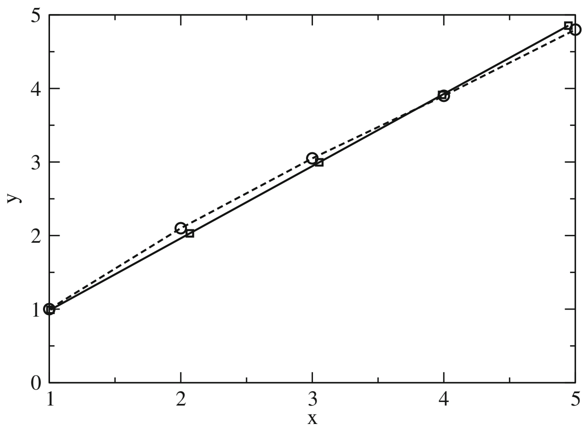

Fig. 10.1 Linear approximation by singular value decomposition. The data set (10.70) is shown as *circles*. The linear approximation which is obtained by retaining only the dominant singular value is shown by the *squares* and the *solid line*

Problems

Problem 10.1 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

$$C(T) = aT + bT^3$$

The computer experiment generates data with a random relative error

$$T_j = T_0 + j\Delta t$$
$$C_j = (a_0 T_j + b_0 T_j^3)(1 + \varepsilon_j)$$

and minimizes the sum of squares

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

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