Part I

Framework of Least-Squares Method

Chapter 1

Introduction to Data-Fitting Problems

1.1 What is data fitting?

The classical setup for data fitting is based on an experiment or measurement. The measurements yield different results, dependent on the experimental conditions. For instance, we can think of measuring the electrical current I (*observation*) in a resistor R dependent on the given voltage U (*condition*) across this resistor.

Typically, the underlying functional relation between the outcome of the measurement and the corresponding conditions is known or assumed. This relation is described by a mathematical model, which is called *model function* throughout the book. If we expect, for instance, the resistor R to be constant, the functional connection would be $I = U/R$ according to Ohm's law. If the electrical component is of any other type, the relation between I and U could be different and another suitable model has to be used.

The goal of data fitting is to find those parameters of the model function that best describe the connection between observations and conditions.

The mathematical description is as follows.
The outcome of the measurement or experiment is called observation y. It is dependent on several conditions, described by a vector $\mathbf{x}$. The parameters of the model are combined in a vector $\mathbf{a}$ with M elements a_j, $j = 1 \ldots M$. They are unknown and have to be determined by data fitting methods. Thus, the model function is

$$y = f(\mathbf{x}|\mathbf{a}) \, . \tag{1.1}$$

In case of the aforementioned Ohm's law, the current I is equivalent to the observation y, the vector of conditions reduces to a scalar $x = U$, and $\mathbf{a} = a = R$ is the only parameter to be determined.

Sometimes more than one value is observed for a single experiment. We could, for instance, think of measuring not only the current I flowing through the resistor, but also its temperature T. However, since the corresponding model function $T = f(I)$ will be different, y is only considered as a scalar for the moment and not as a vector of observations.

It is impossible to change the experimental conditions continuously in practice. Discrete points indicated by variable i have to be selected instead. Besides of this, the observations are affected by random errors (noise). That is why the problem has to be re-formulated as

$$y_i = f(\mathbf{x}_i|\mathbf{a}) + \varepsilon_i \ . \tag{1.2}$$

The vector $\mathbf{x}_i$ contains the condition for experiment i and can be regarded as a set of independent variables, whereas the y_i are considered as dependent variables. It is assumed that $\mathbf{x}_i$ is known exactly or that the error in $\mathbf{x}_i$ is at least negligibly small compared to the error in the measured observation y_i. Only $\mathbf{a}$ is unknown and has to be calculated using suitable data-fitting methods. The repetition of a certain experiment under the same condition $\mathbf{x}_i$ leads to different values of y_i, because observations are subject to the random error ε_i. Doing the experiment an infinite number of times would reveal the true *parent distribution* of y_i. In practice, however, the parent distribution is unknown and can only be estimated to a certain accuracy.

Since we are looking for an optimal solution for $\mathbf{a}$, data fitting is a kind of optimisation problem. Strictly speaking, it concerns minimisation, as will be shown later on. It must be emphasized that the optimum of $\mathbf{a}$ is never absolute, but rather always related to the measured data and the assumptions taken with regard to the reliability of the observations y_i or the distribution of their errors [Mey84]. When the model parameters $\mathbf{a}$ are found, the model function can be used for the interpolation and, provided that the model is not only an approximation for a limited subset of conditions, also for the extrapolation of data.

Data fitting is often referred as *data reduction*, since the number of N observations is reduced to a mere M parameters plus the selection of a suitable model function. Keeping in mind that the model does not represent all observations exactly, but instead, acts as approximation, it also can be regarded as special kind of lossy data compression [Str09a].

In some applications, the true functional connection between the experimental conditions $\mathbf{x}_i$ and the observations y_i is not known. In this case, it is a common method to approximate y for a limited range of $\mathbf{x}$ by using a superposition of special functions of higher order like polynomials, trigonometric functions, or something else. However, the decision for or against a certain model function must be made very carefully, since it will heavily influence the result of the data-fitting procedure. In particular, model functions that are not based on a true or realistic physical, chemical, or other model must not be used extrapolate data!

1.2 Notation

Three basic types of variables are used in this book. Scalar variables are always denoted in *italics*, for instance, y, $f(x)$, a_1 etc. Vectors are expressed in small, bold-faced letters, such as $\mathbf{x} = (x_1, x_2, \dots x_N)^{\mathrm{T}}$. The exponent T means "transposed", making $\mathbf{x}$ a column vector. Please note that T is not in *italics* because it is not a variable. The same holds true for indices, which are not consecutive numbers such as i in y_i but rather simply indicate that a certain variable is related to something else, like σ_{y}^2. y is not a serial number here; it signals that σ^2 is the variance in y. Finally, matrices are denoted with capital, bold-faced letters such as $\mathbf{J}$ and $\mathbf{W}$, for example.

The elements of matrices are scalars, thus the elements of matrix $\mathbf{A}$ are denoted by a_{ij} with $i = 1, \dots, N$ and $j = 1, \dots, M$ as the row and column indices, respectively. The set of integers is denoted by $\mathbb{Z}$. $||\mathbf{x}||$ is the Euclidean norm of column vector $\mathbf{x}$, i.e., $||\mathbf{x}|| = \left(\mathbf{x}^{\mathrm{T}} \cdot \mathbf{x}\right)^{1/2} = \left(\sum_i x_i^2\right)^{1/2}$. Pages 241f list most of the variables, which are used throughout the book.

Readers who are not familiar with matrix algebra are requested to read Section 5.1 first.

1.3 Linear vs. nonlinear problems

The discussion of the linearity of a data-fitting problem is indispensable, since the kind of treatment and its complexity depend heavily on this property.

In engineering, the term "linear" is associated with things like linear time invariant (LTI) systems or linear signal transformations and indicates that an additive superposition of input signals leads to an additive superposition of the according output signals. In addition, a characteristic curve $y = f(x)$ is called linear if it is a straight line, which is basically the same because of

$$y_1 = f(x_1)\ , \qquad y_2 = f(x_2) \qquad \longrightarrow \qquad y_1 + y_2 = f(x_1 + x_2)\ . \tag{1.3}$$

In the field of data fitting, the terms "linear" and "nonlinear" are used with different meanings. They are not assigned to the relation between the observation

y and the vector of conditions $\mathbf{x}$ of the model function $y = f(\mathbf{x}|\mathbf{a})$, but instead to the relation between y and the model parameters a_j $(j = 1 \dots M)$! For instance, the exponential function

$$y = f(\mathbf{x}|\mathbf{a}) = a_1 \cdot \mathrm{e}^{-x} \tag{1.4}$$

is regarded as linear in $\mathbf{a}$, despite the nonlinear term e^{-x}, since the first derivative is $\partial y/\partial a_1 = \mathrm{e}^{-x}$ and no longer dependent on $\mathbf{a}$. In contrast to that

$$y = \mathrm{e}^{a_2 \cdot x} \tag{1.5}$$

is obviously nonlinear due to a_2, because $\partial y/\partial a_2 = x \cdot \mathrm{e}^{a_2 \cdot x}$ is still dependent on a_2.

Sometimes a chosen model function seems to be nonlinear at first glance, but turns to linear if treated properly. Since data fitting of nonlinear functions is in general more difficult than data fitting of linear ones, the desired model should always be inspected carefully to determine whether it is in fact nonlinear or simply *pseudo-nonlinear.*

Example (1):
The model function

$$y = b_1 + (x - b_2)^2 \ , \tag{1.6}$$

for example, seems to be nonlinear, since b_2 is part of a squared term. However, the function can be transformed into

$$\begin{aligned} y &= b_1 + x^2 - 2 \cdot b_2 \cdot x + b_2^2 \\ &= b_1 + b_2^2 - 2 \cdot b_2 \cdot x + x^2 \end{aligned}$$

and at closer look it becomes apparent that the equation can be rewritten as

$$y = a_1 + a_2 \cdot x + x^2 \ . \tag{1.7}$$

with

$$a_1 = b_1 + b_2^2 \quad \text{and} \quad a_2 = -2 \cdot b_2 \ .$$

It turns out that the model is linear in $\mathbf{a}$!
□□□

Example (2):
Another example is the cosine function

$$y = b_1 + b_2 \cdot \cos(x - b_3) \ . \tag{1.8}$$

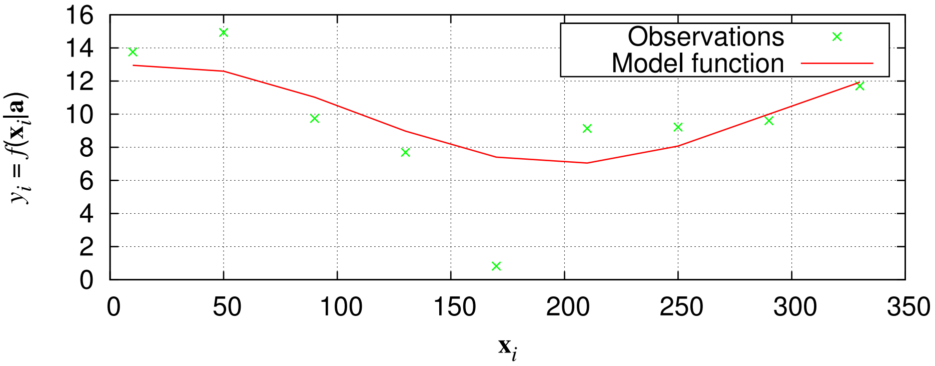

Figure 1.1: Example of a pseudo-nonlinear model function: $y = b_1 + b_2 \cdot \cos(x - b_3)$

This model could be used to describe a circular movement, if the observation is restricted to the projection of the movement onto a screen perpendicular to the plane of movement as for instance described in [Pot06]. **Figure 1.1** shows the recording of nine observations in dependence of known rotation angles x_i.

There are three unknowns to be determined via data fitting: b_2 is the radius of the circular movement, b_3 is a phase offset dependent on the direction of the projection, and b_1 indicates the centre of the performed movement.

It is somewhat more difficult to discover a proper transformation here. Looking for a suitable trigonometric operation, we find

$$\cos(a - b) = \cos(a)\cos(b) + \sin(a)\sin(b)$$

and after its application to the initial model function (1.8)

$$y = b_1 + b_2 \cdot [\cos(x)\cos(b_3) + \sin(x)\sin(b_3)] \ .$$

Using

$$a_1 = b_1, \quad a_2 = b_2 \cdot \cos(b_3), \quad \text{and} \quad a_3 = b_2 \cdot \sin(b_3)$$

we get

$$y = f(\mathbf{x}|\mathbf{a}) = a_1 + a_2 \cdot \cos(x) + a_3 \cdot \sin(x) \ , \tag{1.9}$$

which is a linear model function with respect to **a**. After fitting of (1.9) it is straightforward to determine b_j from a_j.
□□□

This text was taken from the book:

T.Strutz: *Data Fitting and Uncertainty – An introduction to weighted least squares and beyond*. ISBN-10: 3834810223, ISBN-13: 978-3834810229

„A text guiding you step-by-step from the basics to the details of the method of data fitting“