7 Fitting functions to data

If you want to fit parameters in a functional relation to experimental data, the best method is a least-squares analysis: Find the parameters that minimize the sum of squared deviations of the measured values from the values predicted by your function. In this chapter both linear and nonlinear least-squares fits are considered. It is explained how you can test the validity or effectiveness of the fit and how you can determine the expected inaccuracies in the optimal values of the parameters.

7.1 Introduction

Consider the following task: you wish to devise a function $y = f(x)$ such that this function fits as accurately as possible to a number of data points (x_i, y_i), $i = 1, \ldots, n$. Usually you have – based on theoretical considerations – a set of functions to choose from, and those functions may still contain one or more yet undetermined *parameters*. In order to select the "best" function and parameters you must use some kind of *measure* for the deviation of the data points from the function. If this deviation measure is a single value, you can then select the function that minimizes this deviation.

This task is not at all straightforward and you may be lured into pitfalls during the process. For example, your choice of functions and parameters may be so large and your set of data may be so small that you can choose a function that *exactly* fits your data. If you have n data points, you can fit an $(n-1)$th degree polynomial exactly through all points. But you can also fit a smooth cubic spline through all points. In fact, there are an infinite number of functions that fit all points and by choosing one you have achieved nothing else than a fancy description of your data set (see Fig. 7.1). At best you have found a non-exclusive way to interpolate your data.

Two things are needed to improve the quality of your task. First, there must be a valid theory behind your choice of functions. The better your theory is, the more restricted is the range of functions and parameters from which you can choose. Second, your deviation measure must have statistical relevance

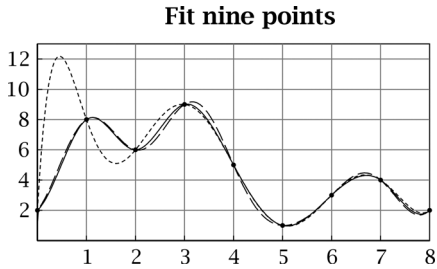

Figure 7.1 Several functions fitting exactly through nine equidistant points. The first and last point are taken equal to allow periodic solutions. Drawn curve: periodic cubic spline (piecewise third-degree polynomials with continuous first and second derivatives). Broken curve: each point is expanded by the function $\sin \pi x/(\pi x)$ (Nyquist–Shannon formula; the Fourier transform of the resulting function has no components with wavelength smaller than two units). Dotted curve: fit to a (non-periodic) eighth-degree polynomial using Lagrange's formula (Press *et al.*, 1992). Global polynomial fits are almost never satisfactory.

in the sense that it must have an associated *probability* that a given deviation occurs. For example, you have n independent data points ($n \gg 3$) and a theory that allows either a linear relation (two parameters) or a quadratic relation (three parameters). It is clear that a quadratic relation (which includes linear relations as a subset) will always fit better than a linear relation and a deviation measure will then be lower for the quadratic equation. Without a probability attached to your deviation measure you may always decide in favor of the quadratic relation, but with a proper probability measure you may decide that the quadratic relation fits too well and the linear relation is more probable. As you will see, with certain assumptions deviation measures can be devised to which meaningful probabilities can be attached.

Assuming that the independent variables x_i are accurate, and the function $f(x)$ is the correct relation, you expect that the deviations of the dependent variables y_i from the function values $f_i = f(x_i)$ behave as independent random samples from a probability distribution with zero mean and finite variance:

$$y_i = f(x_i) + \varepsilon_i; \tag{7.1}$$

$$E[\varepsilon_i] = 0, \tag{7.2}$$

$$E[\varepsilon_i \varepsilon_j] = \sigma_i^2 \delta_{ij}. \tag{7.3}$$

The ε_i's are called the *residuals* of the fitting procedure. The assumption that the x_i's are accurate is for convenience only; in Section 7.2 it is shown how you should treat your data when the x_i's are themselves samples of a

probability distribution (see (7.11) on page 88). Also the assumption that the residues are independent (at least: uncorrelated) is for convenience only: in Appendix A9 it is shown how you should treat your data if the residues are correlated.

If you are lucky you may have some advance information on the probability distribution of the residues because you know something about the random process that generates the deviations. For example, if you know that the residue ε_i is an independent sample from a normal distribution with known variance σ_i^2, you can estimate the probability that the set of (independent) residues $\varepsilon_1, \ldots, \varepsilon_n$ occurs:

$$P(\varepsilon_1, \ldots, \varepsilon_n) = \Pi_{i=1}^n p(\varepsilon_i) \propto \exp\left[-\sum_{i=1}^n \frac{\varepsilon_i^2}{2\sigma_i^2}\right] = \exp\left[-\frac{1}{2}\chi^2\right], \quad (7.4)$$

where χ^2 is defined as the weighted sum of square deviations:

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - f(x_i))^2}{\sigma_i^2}. \quad (7.5)$$

This probability product can be considered as the *likelihood* of the fit: a function with a higher likelihood is more likely to occur and the best fit can be considered to be the one that *minimizes* χ^2. Having found the best fit, you can use the minimum value of χ^2 to assess the quality of your fit by performing a *chi-squared analysis*. This will be considered in Section 7.4. Also the inaccuracy of the parameters in the function (their variances and covariances) can be derived (Section 7.5). In Chapter 8 the principles behind choosing the "best" function are considered more carefully.

In practice you may not be so lucky as to know beforehand what random process generates the deviations. Often you don't know the individual variances, but you do know the *relative weights* w_i of the deviations. For example, if data point i was the average of 100 measurements and data point j was the average of 25 similar measurements, then point i should be given a relative weight 4 times larger than point j. Or, if you have a series of measurements t_i with similar uncertainty, but you use $y_i = \log t_i$ for your fitting procedure, you should give the value y_i a weight proportional to t_i^2. Refer to Exercise 7.6 on page 109 for an explanation. Instead of minimizing χ^2 you can now minimize the *weighted sum of square residues* S:

$$S = \sum_{i=1}^n w_i (y_i - f_i)^2, \quad (7.6)$$

but – of course – you can no longer use the minimal value of S to assess the quality of the fit. If you can trust the functional form, and you have reasons to *assume* that the residues are just random samples from a distribution

with unknown variance, you can *derive* an estimate of the variance of the distribution. From that you can in turn derive the inaccuracy (variances and covariances) of the parameters in the function.

So, for the determination of accuracies in a fitting procedure there are two possibilities: either use the known uncertainties in the data (if available) or use the observed magnitude of the sum of square deviations. If both are compatible, use the more reliable of the two, or – in doubt – choose the higher uncertainty. If they are both reliable but not compatible (according to a chi-squared analysis), then lean back, check your data and your error estimates – possibly measure again – and revise your theory.

7.2 Linear regression

Linear regression is a least-squares fit of the parameters in a *linear* function to a data set:

$$f(x) = ax + b, \tag{7.7}$$

where a and b are the adjustable parameters of the function. Given a set of independent data (x_i, y_i), $i = 1, \ldots, n$, with optional individual weights w_i, it is now your task to minimize the sum of (weighted) square deviations:

$$S = \sum_{i=1}^{n} w_i (y_i - f_i)^2 \text{ minimal} \tag{7.8}$$

by adjusting the parameters a and b. Here

$$f_i = f(x_i) = ax_i + b. \tag{7.9}$$

Here we use x to denote the independent variable, which is also called the *explanatory* variable, as the values of y – but for a random deviation – follow from x. There may well be more than one explanatory variable so that $f_i = f(\boldsymbol{x})$, where $\boldsymbol{x}$ is a vector and also the parameter a becomes a vector. This complicates the least-squares solution a bit; the multidimensional linear regression can be found in Appendix A9.2 on page 161.

While (7.7) is linear in x, it is the *linearity in the parameters* a and b that allows an analytical solution of the minimization problem (7.8). Thus also for functions like $ax^2 + bx + c$ or $a + b \log x + c/x$ least-squares minimizations can be solved by linear regression. How this is done is explained in Appendix A9. Here we consider only linear functions of x.

The factors w_i are the statistical weights of the data points. It is quite common that all points have the same weight because they come from the same statistical distribution; in those cases w_i can all be taken $= 1$. If the weights are not equal because the data points have different standard deviations σ_i,

then the weights must be taken equal to (or proportional to) $1/\sigma_i^2$ (note: *not* proportional to $1/\sigma$!).

Uncertainties in *x*

When the uncertainties in x are negligible (which is the common case), the standard deviation σ_i is simply the standard deviation of y_i. When the uncertainty in x is not negligible (but independent of the deviation of y_i), then σ_i^2 must be replaced by

$$\sigma_i^2 = \sigma_{yi}^2 + \left(\frac{\partial f}{\partial x}\right)_{x=x_i}^2 \sigma_{xi}^2, \tag{7.10}$$

because we deal with the uncertainty in $y_i - f(x_i)$. For the linear relation (7.7) this reduces to

$$\sigma_i^2 = \sigma_{yi}^2 + a^2\sigma_{xi}^2. \tag{7.11}$$

To evaluate this, you need to know the value of a which is yet to be determined. However, a rough estimate (e.g. from a graphic sketch) suffices at this point.

The best parameter estimates

In general you will need a computer program to find the solution to the least-squares minimization problem of (7.8). However, for the linear relation (7.7) the solution can be expressed in simple terms. The solution follows from setting the two derivatives $\partial S/\partial a$ and $\partial S/\partial b$ to zero and is worked out in Appendix A9. Here the resulting equations are given.

The parameters a and b follow from a number of averages over the measured data points. The weights must be taken into account to determine the averages, just as was done in Section 5.6 (see, for example, (5.20) on page 62). For example:

$$\langle xy\rangle = \frac{1}{w}\sum_{i=1}^{n} w_i x_i y_i; \quad w = \sum_{i=1}^{n} w_i. \tag{7.12}$$

The parameters are:

$$a = \frac{\langle(\Delta x)(\Delta y)\rangle}{\langle(\Delta x)^2\rangle}; \quad b = \langle y\rangle - a\langle x\rangle, \tag{7.13}$$

where

$$\Delta x = x - \langle x\rangle; \quad \Delta y = y - \langle y\rangle. \tag{7.14}$$

These averages can also be computed without first subtracting the averages of x and y because

$$\langle(\Delta x)(\Delta y)\rangle = \langle xy\rangle - \langle x\rangle\langle y\rangle; \tag{7.15}$$

$$\langle(\Delta x)^2\rangle = \langle x^2\rangle - \langle x\rangle^2. \tag{7.16}$$

Beware of numerical precision when you subtract two large numbers (see note on page 58).

From the equation for b (7.13) you see that the optimal function passes through the point $(\langle x\rangle, \langle y\rangle)$. This is the "center of mass" of the set of points. This fact was used in the discussion of graphical estimates in Section 6.3.

Uncertainties in the parameters

Estimates for the standard uncertainties σ_a and σ_b in a and b are the square root of the estimated variances $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$. The latter follow from the behavior of the function $\chi^2(a,b)$ and are derived from the likelihood function (7.4):

$$p(a,b) \propto \exp\left(-\frac{1}{2}\chi^2(a,b)\right). \tag{7.17}$$

The function $\chi^2(a,b)$ is a quadratic function in a and b and hence the probability distribution $p(a,b)$ is a bivariate normal distribution in the deviations Δa and Δb from the parameter values at the minimum. The coefficients of the terms $(\Delta a)^2$, $(\Delta a)(\Delta b)$ and $(\Delta b)^2$ determine the variances and covariance of a and b as explained in Section 7.5 and in Appendix A9. When χ^2 is estimated from the data themselves, i.e. from the minimum value S_0 of S, the results for the estimated (co)variances are

$$\text{var}\,(a) = \hat{\sigma}_a^2 = \frac{S_0}{w(n-2)\langle(\Delta x)^2\rangle}, \tag{7.18}$$

$$\text{var}\,(b) = \hat{\sigma}_b^2 = \hat{\sigma}_a^2\langle x^2\rangle, \tag{7.19}$$

$$\text{cov}\,(a,b) = -\hat{\sigma}_a^2\langle x\rangle \tag{7.20}$$

where w is the total weight of all points together. In the common case that all weights have been taken $= 1$, w is simply equal to the number of observations n.

The $n-2$ in (7.18) has the meaning of the number of *degrees of freedom*: the number of (independent) data points minus the number of parameters in the function. Appendix A9 explains the details, but one could loosely say that two points are needed to determine two parameters and only $n-2$ points are

left to determine deviations from the fit. This makes sense: you can always draw a straight line through two points; for $n = 2, S = 0$ and the inaccuracies remain undetermined.

Covariances between parameters

The covariance $\text{cov}(a, b)$ indicates whether deviations in a and b are correlated with each other: to what extent can a deviation in a be compensated by a deviation in b? Covariances must be used to determine the uncertainty in quantities that are a function of the parameters, e.g. for interpolation and extrapolation of data. See the remarks on page 22 and in Appendix A1.

The covariance is often expressed relative to the product $\hat{\sigma}_a\hat{\sigma}_b$ and is then called the *correlation coefficient* between a and b (a dimensionless number between -1 and $+1$):

$$\rho_{ab} = \frac{\text{cov}(a, b)}{\hat{\sigma}_a\hat{\sigma}_b} = -\frac{\langle x\rangle}{\sqrt{\langle x^2\rangle}}. \tag{7.21}$$

Note that a and b are uncorrelated ($\rho_{ab} = 0$) when $\langle x\rangle = 0$, i.e., when the zero of x is chosen in the "center of mass." So when you choose as the linear function

$$f(x) = a(x - \langle x\rangle) + b, \tag{7.22}$$

you can be sure that a and b are uncorrelated. In addition, you know right away that

$$b = \langle y\rangle. \tag{7.23}$$

Extrapolations are now much simplified: if you wish to determine the inaccuracy in $f(x)$ at an arbitrary x you can simply add the contributions quadratically:

$$\sigma_f^2 = \sigma_a^2(x - \langle x\rangle)^2 + \sigma_b^2, \tag{7.24}$$

while using the formula $f(x) = ax + b$ a correction is needed: (see Appendix A1):

$$\sigma_f^2 = \sigma_a^2x^2 + \sigma_b^2 + 2\rho_{ab}\sigma_a\sigma_bx. \tag{7.25}$$

Should you use S_0 or χ_0^2?

As you have noticed, the (co)variances are proportional to the minimal sum of squares: we have used the measured deviations to determine the uncertainties in the parameters. This is the only choice we have if the individual standard deviations σ_i are not known beforehand. If they *are* known and reliable, they could also be used to determine the uncertainties in the parameters. In that case the term $S_0/[w(n - 2)]$ must be replaced by the known

$1/\sum \sigma_i^{-2}$. Before deciding which choice to make you should always perform a chi-squared analysis (Section 7.4). These matters are discussed more fully in Sections 7.4 and 7.5.

Correlation coefficient between *x* and *y* values of a data series

There is a quantity that indicates how well a series of points lie on a straight line. This is the *correlation coefficient* r between the x and y values of the data series. Points approach a straight line only when there is a strong correlation between their x and y values. Don't confuse this correlation coefficient with the correlation coefficient ρ_{ab} (7.21) between a and b as discussed above. While the latter is derived from expectations over an estimated probability distribution, r is a property of the data set itself:

$$r = \frac{\langle(\Delta x)(\Delta y)\rangle}{\sqrt{\langle(\Delta x)^2\rangle\langle(\Delta y)^2\rangle}} \tag{7.26}$$

$$= a\sqrt{\frac{\langle(\Delta x)^2\rangle}{\langle(\Delta y)^2\rangle}}. \tag{7.27}$$

For $r = \pm 1$ there is complete correlation: the points lie exactly on a straight line; for $r = 0$ there is no correlation and it makes no sense to fit a linear function to the data. For a reasonable correlation r should be above 0.9.

A correlation coefficient less than 1.0 indicates *that*, but not *how* the data deviate from a linear relation. The two sets of data plotted in Figs 7.2a and 7.2b have the same correlation coefficient of 0.900, but both deviate in very

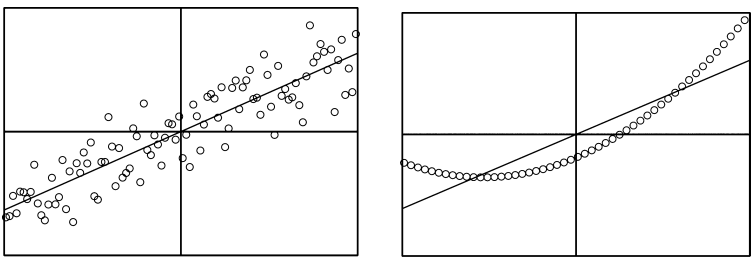

Figure 7.2 Two data sets x_i, y_i with the same correlation coefficient r between x and y of 0.900. There are no numbers along the axes: correlation coefficients are independent of linear scaling or translation of the axes. The drawn lines are the best linear fits through the points.

different ways from the best-fitted straight line. You also see in this figure that $r = 0.9$ does not guarantee the fit to a straight line to be very satisfactory.

7.3 General least-squares fit

When the function to be fitted $f(x, \boldsymbol{\theta})$ with parameters $\boldsymbol{\theta} = \theta_1, \ldots, \theta_n$ is more general than $ax + b$, the following cases should be distinguished:

(i) f **linear in $\boldsymbol{\theta}$.** For functions linear in all parameters, such as

$$f(x) = ax^2 + bx + c \tag{7.28}$$

$$f(x) = a + b\exp(-k_1 x) + c\exp(-k_2 x); \tag{7.29}$$

$$(k_1, k_2 \text{ known constants})$$

$$f(x) = ax + b/x + c, \tag{7.30}$$

an analytical (weighted) least-squares solution minimizing S (7.8) is still possible, but requires some matrix algebra. Appendix A9 gives details.

(ii) f **linear in several variables.** Functions linear in more than one independent ("explanatory") variable, such as

$$f(\xi, \eta, \zeta) = a\xi + b\eta + c\zeta + d, \tag{7.31}$$

where ξ, η, ζ are the independent variables and a, b, c, d are the parameters, similarly yield analytical least-squares solutions using some matrix algebra. Such functions are also treated in Appendix A9.

(iii) f **nonlinear but linearizable.** Functions that are nonlinear in the parameters can often be transformed to linear functions, as was done in Section 6.2 on page 73 in order to obtain linear graphs. For example, the function $f(t) = a\exp(-kt)$ is not linear in the parameter k. But if you take the logarithm:

$$\ln f(t) = -kt + \ln a, \tag{7.32}$$

you obtain a function that *is* linear in k and has the form $ax + b$. So you can apply linear regression to the points $(t_i, \ln y_i)$ and determine k and $\ln a$. But take proper care of the weights: if all values of y have equal standard deviations σ, the values $\ln y_i$ have different weights:

$$\sigma_{\ln y} = \left|\frac{d\ln y}{dy}\right| \sigma_y = \sigma_y / y_i \tag{7.33}$$

In this case you should take the weights w_i equal to (or proportional to) y_i^2. Negative values of y, which may occur by random deviations for

large values of t, cannot be handled. It is not allowed to selectively omit negative values, as that will bias the result. The best way to proceed is to omit all points with larger values for t than the value for which a negative y first occurred. Even better is the use of a general nonlinear fitting procedure.

(iv) f **nonlinear: general case.** Functions that are nonlinear in the parameters cannot always be linearized. For example, the function

$$f(t) = a\exp(-kt) + b \tag{7.34}$$

cannot be transformed to a linear function of the parameters a, b, and k. A nonlinear least-squares fitting procedure must then be used. In this case there are no analytical solutions and solutions are obtained by iterative function minimizers. There are several minimizers available, some that require analytical derivatives and some that do not. The latter are more easy to use. In all cases an initial guess of the parameters is required; for some functions a bad guess may lead to failure of the minimization procedure. A graphical analysis is a good source for a reasonable initial guess.

Below an example is given of a nonlinear least square minimization using Python.

Example: nonlinear fit

Consider the data on enzyme kinetics given in Table 6.2 on page 76. Given are six data points S_i, v_i with equal weights and our task is to fit two parameters $p_0 = v_{max}$ and $p_1 = K_m$ in the function

$$f(S,p) = \frac{p_0 S}{p_1 + S}; \tag{7.35}$$

$$p = [v_{max}, K_m], \tag{7.36}$$

such that the sum of squares

$$SSQ = \sum_i [v_i - f(S_i,p)]^2 \tag{7.37}$$

is minimal. One possibility is to use the Python least-squares minimization procedure `leastsq` that comes with the module `optimize` of `SciPy`. This function requires a specification of the residues $y_i - f_i$ (or $(y_i - f_i)/\sigma_i$ if s.d.'s are known), which are a function of the parameters p, but does not require any derivatives. It must be called with an initial guess for p, for which we choose the values found by graphical inspection in Exercise 7.6:

$$p_{\text{init}} = [15, 105]. \tag{7.38}$$

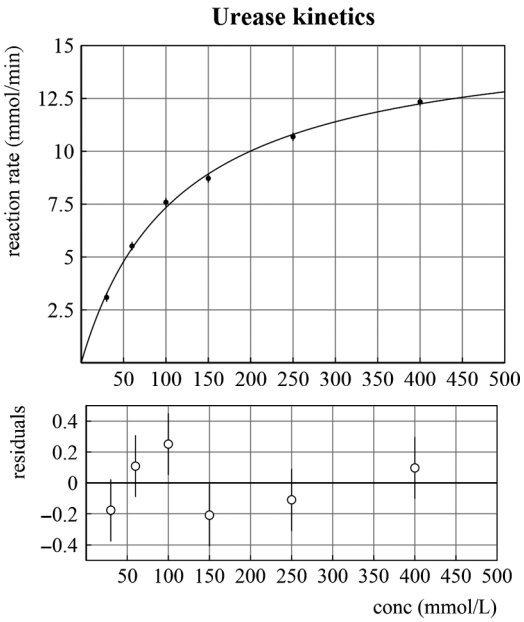

Figure 7.3 *Upper panel*: The urease reaction rate data plotted together with the least-squares fitted function. *Lower panel*: Plotting the residuals $y_i - f_i$ with error bars more clearly show whether the deviations have a random character or not.

The SSQ using this initial guess equals 0.375. After applying the minimization procedure the parameters appear to be

$$p_{\min} = [15.75, 114.65] \tag{7.39}$$

and the minimal SSQ is 0.171. Figure 7.3 shows the fit, together with a plot of the residues. The latter plot is able to show the size of the error bars, and gives a visual impression of any systematic deviations. Later we shall see how large the uncertainties in the parameters appear to be.

Another possibility is to use the Python procedure `fmin_powell`, also in the module `optimize` of `SciPy`. With this minimizer the function to be minimized must be specified. This routine is less accurate than `leastsq` and should preferably be applied more than once.

A Python code which performs these minimizations is **Python code** 7.1 on page 183

Having determined the best values of the parameters, your problems are not yet solved! You also wish to assess the validity of the fit and you wish to have an estimate of the inaccuracies in the parameters. The key to the answers to these problems lies in the value of χ^2 as a function of the parameters. The next two sections explain all this.

7.4 The chi-squared test

Suppose you have performed a least-squares fit of the parameters in a function to a set of data. Is the fit reasonable, i.e., are the data – within experimental errors – compatible with the functional relation? What criteria can you apply to answer this question? What is "reasonable?"

Deviations from an acceptable fit can be expected both ways: the fit can be *not good enough* but it can also be *too good*. When the function has too few parameters or the functional shape is incorrect, the data will have systematic deviations that exceed the expected random errors. When there are too many parameters, the functional fit (if it succeeds at all) will follow the data too closely and the deviations will be less than expected from random errors.

You should always first check how the deviations $y_i - f_i$ (i.e., the residues) depend on x. A successful fit will yield residues that are samples from a random, generally normal, distribution. Systematic deviations are generally immediately apparent from a plot of the residues versus x. If you see such deviations, your fit is not acceptable and you should reconsider the functional choice you have made.

If there are no obvious deviations, the next step is to perform a *chi-squared test*. This test checks whether the sum of square deviations is compatible with the expectation based on the assumption that the deviations are samples from a (known) probability distribution. *The chi-squared test can only be performed if you have reliable prior estimates for the standard inaccuracies σ_i of each data point.* How to proceed when you don't have prior knowledge about the expected deviations is explained at the end of this section. Determine the minimal value of χ^2, defined as

$$\chi_0^2 \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{(y_i - f_i)^2}{\sigma_i^2}, \tag{7.40}$$

where the f_i are evaluated at the optimal parameter values. This is just the value of S_0 (7.6) when all weights are taken equal to the inverse variances:

$$w_i = \sigma_i^{-2}. \tag{7.41}$$

Note: if you have determined S using weights that are proportional but not equal to inverse variances, you can derive χ_0^2 from

$$\chi_0^2 = \frac{S_0}{w} \sum_{i=1}^{n} \sigma_i^{-2}, \tag{7.42}$$

where w is the total weight $\sum w_i$. See Exercise 7.6.

Since each of the terms in the chi-squared sum should have an expectation of 1, you expect the sum to be close to n. This is not quite correct: for a linear regression two degrees of freedom have been "used" for the determination of the two parameters a and b. The number of degrees of freedom ν equals $n - 2$ and χ^2 will be approximately equal to $\nu = n - 2$. In general, when there are m adjustable parameters, the remaining number of degrees of freedom is $n - m$. But χ^2 has a probability distribution around this value. This probability function depends on ν and is – if based on a normal distribution of the contributing deviations – denoted by $f(\chi^2|\nu)$. The larger the number of degrees of freedom, the narrower the chi-squared distribution becomes. The function $f(\chi^2|\nu)$ is fairly complex (see the data sheet CHI-SQUARED DISTRIBUTION on page 199 for equations), but has the convenient property that it approaches a normal distribution for large ν. The mean is equal to ν and the standard deviation equals $\sqrt{2\nu}$; this is not only true in the limit of large ν but is valid for all ν.

Tables of the chi-squared distribution do not give the probability density, but give the *cumulative distribution function* (cdf) $F(\chi^2|\nu)$: the probability that a sum of squares does *not* exceed χ^2. The survival function (sf) $1 - F(\chi^2|\nu)$ then indicates the probability that the value χ^2 *is* exceeded. The data sheet CHI-SQUARED DISTRIBUTION gives a table of values of χ^2 that are still acceptable for acceptance limits of 1%, 10%, 50%, 90% and 99%. Most books on statistics (and the *Handbook of Chemistry and Physics*) contain larger tables, but you may find it easier to use a Python routine from SciPy.

See **Python code** 7.2 on page 184 to generate chi-squared probabilities.

Use the table or Python code as follows. First set an acceptance criterion, e.g. between 1% and 99%, or between 10% and 90%. The choice is subjective and you should report your choice when you publish your results. In the following example we choose the 10–90% limits. If your least-squares χ^2-value is less than the 10% value, you do not accept the outcome as a random occurrence and conclude that the fit is *too good*. Your function has too many parameters and you should try a simpler function with the same data: *your data do not justify the complexity of your function*. On the other hand, if your least-squares χ^2-value exceeds the 90% value, you do not accept the outcome as a random occurrence either and conclude that the data deviate significantly from the function. In that case: look for a function that better describes the data, possibly with more parameters. In both cases it is

also good practice to review your data and your original estimates of the variances.

Example 1

In the urease kinetics example (see page 7.3) the least-squares fit gave a minimal value for the sum of squared deviations SSQ of 0.171. Since the s.d. of the data y_i were given as 0.2 mmol/min, the minimum value of χ^2 appears to be $0.171/0.2^2 = 4.275$. This is quite close to the number of degrees of freedom $\nu = n - m = 6 - 2 = 4$, so you may conclude that the deviations are compatible with random fluctuations. Indeed, the cdf of the chi-squared distribution `cdf(4.275,4)` equals 0.63, a completely insignificant deviation.

Example 2

You have 10 independent measurements (x, y), with x being accurate and y having a known standard uncertainty σ. You have a simple theory that predicts y to be a linear function $ax + b$ of x, but a more refined theory predicts a second-degree function $px^2 + qx + r$. Do your data justify the refined theory above the simple one at a confidence level of 90%? You perform a linear least-squares fit to both functions, using $1/\sigma^2$ as weight factor for all points. For the linear function you find $S = \chi^2 = 14.2$ and for the quadratic function you find $S = \chi^2 = 7.3$. Inspection of the table in data sheet CHI-SQUARED DISTRIBUTION for 8 degrees of freedom shows that 14.2 lies above the 90% limit and hence is not acceptable according to the chosen acceptance criterion. The quadratic function (with 7 degrees of freedom) is indeed acceptable and the data justify its use. Had these values been 12.3 (linear function) and 6.5 (quadratic function), then the conclusion that a quadratic function should be used would not have been warranted, despite the fact that the quadratic function gives a better fit than the linear function.

What should you do if the standard uncertainties of the data are unknown or inaccurately known? In that case a chi-squared test cannot be used. It could well happen that you overestimate the experimental uncertainties by a factor of 2; this makes χ^2 a factor of 4 smaller. With e.g. 10 degrees of freedom you may then find a value of 2, while the accurate value is 8. The value 2 is below the 1% probability limit, making the fit unacceptable, while the accurate value of 8 is perfectly acceptable. Thus you may draw the wrong conclusion based on a wrong prior estimate of the uncertainties. This example shows that your prior knowledge of the inaccuracies in the data points should be rather precise for a valid chi-squared analysis.

With a sufficient number of data you may *use* the data themselves to determine the uncertainties of the measurements y. The minimal sum of squared

deviations S_0 provides the information to estimate the individual variances $\hat{\sigma}_i^2$ if you take for χ^2 the best estimate $\hat{\chi}_0^2 = n - m$. The relation is

$$\hat{\sigma}_i^2 = \frac{S_0}{(n-m)w_i}, \tag{7.43}$$

which is easily obtained by setting $w_i = c/\sigma_i^2$ and solving c from

$$\hat{\chi}_0^2 = \frac{S_0}{c} = n - m. \tag{7.44}$$

Note that in the common case of equal weights and $w_i = 1$, $\hat{\sigma}^2 = S_0/(n-m)$.

Of course you cannot use $\hat{\chi}_0^2$ now to assess the quality of the fit. Therefore you should apply other criteria to analyze the random character of the residues $\varepsilon_i = y_i - f_i$. A graph versus x should not show systematic deviations. The cumulative distribution function should resemble a symmetric normal distribution. Statistical attributes such as mean and variance, if taken over sections of the data, should not differ significantly for different sections.

7.5 Accuracy of the parameters

Suppose you have performed a least-squares fit and your residues stood the tests for randomness so that you can trust the values of the standard uncertainties in the data. Either you had accurate prior knowledge of the standard uncertainties of the data points or you have scaled your uncertainties such that χ^2 in the minimum exactly equals $n - m$; in any case you know χ^2 as a function of the parameters. Now you can compute the variances and covariances of the fitted parameters.

In this section some general equations are given; their derivations are given in Appendix A9. We start again from n data points x_i, y_i and perform a least-squares fit of a (linear or nonlinear) function of m parameters $\theta_k, k = 1, \ldots, m$. The procedure yields χ^2 as a function of the parameters, with a minimum χ_0^2 for the parameter values $\hat{\theta}_i$, which are considered as best estimates. Because χ_0^2 is a minimum, the function $\chi^2(\theta_1, \theta_2, \ldots, \theta_m)$ is *quadratic* in the neighborhood of the minimum (the quadratic term is the first term in a Taylor expansion of χ^2 about the minimum, also for a fit function which is not linear in the parameters).

An important role in the derivation of (co)variances for the parameters is played by the matrix $\boldsymbol{B}$ with elements

$$B_{kl} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \frac{\partial f_i}{\partial \theta_k} \frac{\partial f_i}{\partial \theta_l}. \tag{7.45}$$

The partial derivatives of f with respect to θ are constants for functions that are linear in the parameters; for nonlinear functions it is required to take the derivatives at the best-fit values $\hat{\theta}$ of the parameters. This matrix is also half the matrix of *second* derivatives of the function $\chi^2(\boldsymbol{\theta})$ (see Appendix A9):

$$B_{kl} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial\theta_k \partial\theta_l}, \tag{7.46}$$

meaning that $\boldsymbol{B}$ quantifies the curvature of the function $\chi^2(\theta_1, \theta_2, \ldots, \theta_m)$ at the minimum:

$$\Delta\chi^2 = \chi^2(\theta) - \chi^2(\hat{\theta}) \tag{7.47}$$

$$\approx \sum_{k,l=1}^{m} B_{kl}\Delta\theta_k \Delta\theta_l, \tag{7.48}$$

where $\Delta\theta_k = \theta_k - \hat{\theta}_k$. For linear functions the $\approx$ sign can be replaced by an $=$ sign.

Covariances of the parameters

The (co)variances of the parameters follow from the likelihood (7.4) on page 86:

$$p(\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2}\Delta\chi^2(\boldsymbol{\theta})\right]. \tag{7.49}$$

By inserting (7.48) into (7.49) a bivariate normal distribution is obtained. As is more fully explained in Appendix A9, the (co)variances of the parameters are given by the *inverse* of the matrix $\boldsymbol{B}$. Denoting this inverse by $\boldsymbol{C}$:

$$\boldsymbol{C} = \boldsymbol{B}^{-1}, \tag{7.50}$$

then

$$\text{cov}\,(\theta_k, \theta_l) = C_{kl}. \tag{7.51}$$

From the covariance matrix the standard deviation σ_{θ_k} of θ_k (which is its standard inaccuracy) can be found:

$$\sigma_{\theta_k} = \sqrt{C_{kk}}. \tag{7.52}$$

The correlation coefficient ρ_{kl} between θ_k and θ_l is

$$\rho_{kl} = \frac{C_{kl}}{\sqrt{C_{kk}C_{ll}}}. \tag{7.53}$$

Table 7.1 *Relation between* $\Delta\chi^2$ *and the probability distribution of a single parameter* θ.

$\Delta\chi^2$	$p(\Delta\theta)/p(0)$	$\Delta\theta$	$P(-\Delta\theta, \Delta\theta)$
0.00	1.00000	0.0000	0.00%
0.50	0.77880	0.7071	52.05%
1.00	0.60653	1.0000	68.27%
1.50	0.47237	1.2247	77.93%
2.00	0.36788	1.4142	84.27%
2.50	0.28650	1.5811	88.62%
3.00	0.22313	1.7321	91.67%
3.50	0.17377	1.8708	93.86%
4.00	0.13534	2.0000	95.45%
4.50	0.10540	2.1213	96.61%
5.00	0.08208	2.2361	97.47%

Thus knowledge of the matrix $\boldsymbol{C}$ suffices to estimate inaccuracies of the parameters and their mutual correlations. In the following two paragraphs the one- and two-dimensional cases are graphically illustrated.

Relation between χ^2 and 1-D parameter distribution

In Table 7.1 the relations are given between $\Delta\chi^2$ and the distribution function of $\Delta\theta$ in the case of only one parameter, for which

$$p(\Delta\theta) = p(0)\exp\left[-\frac{1}{2}\Delta\chi^2(\Delta\theta)\right], \tag{7.54}$$

$$\Delta\chi^2(\Delta\theta) = b(\Delta\theta)^2. \tag{7.55}$$

Here b is the matrix element B_{11} in the expansion (7.48); $1/b$ is the variance σ_θ^2. The standard deviation is reached when $\Delta\chi^2 = 1$ (note that this is a fraction $1/\nu$ of $\Delta\chi_0^2$, the latter having an expectation of the number of degrees of freedom ν). Twice the standard deviation is reached for $\Delta\chi^2 = 4$. Figure 7.4 depicts the relevant relations.

This relation between $\Delta\chi^2$ and $p(\Delta\theta)$ is not only valid in the single-parameter case. When there are several parameters, and you wish to know the *marginal* probability distribution of a particular parameter θ_1, then it suffices to compute the function $\Delta\chi^2(\Delta\theta_1)$, while minimizing $\Delta\chi^2$ with respect to all other parameters. So the standard deviation of $\Delta\theta_1$ is reached

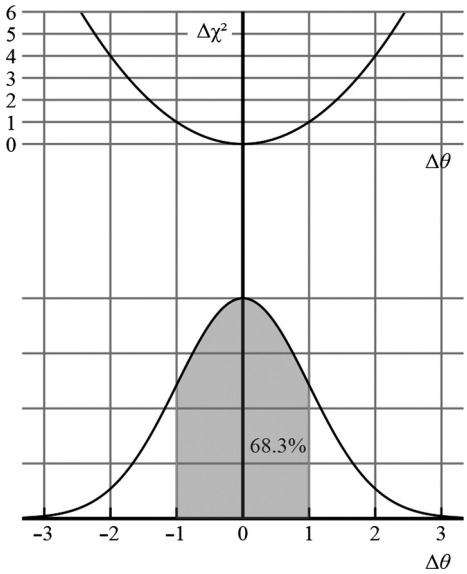

Figure 7.4 Relation between $\Delta\chi^2$ and the probability distribution of a single parameter θ.

when $\Delta\chi^2(\Delta\theta_1) = 1$, while all other parameters have their least-squares values. Why this is so is explained in Appendix A9.

Relation between χ^2 and 2-D parameter distribution

In the case of two parameters, $\Delta\chi^2$ is a quadratic function of two parameters $\Delta\theta_1, \Delta\theta_2$. The important feature to remember is that the vertical and horizontal *tangents to the contour* $\Delta\chi^2 = 1$ are positioned at $\theta_1 = \pm\sigma_1$ and $\theta_2 = \pm\sigma_2$, respectively. Appendix A9 explains why. Table 7.2 gives the relations between $\Delta\chi^2$, their tangent projections and the integrated percentage probability that the combined $(\Delta\theta_1, \Delta\theta_2)$ lies within the contour corresponding to $\Delta\chi^2$. The integrated probability $P(\Delta\chi^2)$ within the contour is a simple function of $\Delta\chi^2$:

$$P = 1 - \exp\left(-\frac{1}{2}\Delta\chi^2\right), \tag{7.56}$$

with inverse:

$$\Delta\chi^2 = -2\ln(1 - P). \tag{7.57}$$

For example, the θ values that represent 99 percent of the integrated joint probability are contained in the contour $\Delta\chi^2 = -2\ln 0.01 = 9.21$.

Table 7.2 *Relation between* $\Delta\chi^2$ *and the 2-D probability distribution of two parameters* $\Delta\theta_1, \Delta\theta_2$.

$\Delta\chi^2$ contour	*tangent projection* in units σ	P(*contour*) integrated
0.0	0.000	0.00%
0.5	0.707	22.12%
1.0	1.000	39.35%
1.5	1.225	52.76%
2.0	1.414	63.21%
2.5	1.581	71.35%
3.0	1.732	77.69%
3.5	1.871	82.62%
4.0	2.000	86.47%
4.5	2.121	89.46%
5.0	2.236	91.79%
5.5	2.345	93.61%
6.0	2.449	95.02%

An example is given in Fig. 7.5, with the following choice for $\boldsymbol{B}$ and $\boldsymbol{C} = \boldsymbol{B}^{-1}$:

$$\boldsymbol{B} = \frac{1}{3}\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}; \quad \boldsymbol{C} = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix},$$

the latter meaning that

$$\sigma_1 = 2; \quad \sigma_2 = 1; \quad \rho_{12} = 0.5.$$

Figure 7.5 represents a *contour plot*: Each contour encloses all values of the two parameters for which the *joint* probability exceeds a given level. The integrated joint probability for the parameter values within the contour $\Delta\chi^2 = 1$ (dark grey area) equals 39 percent (see Table 7.2); the projection on the $\Delta\theta_1$-axis indicates the standard deviation $\sigma_1 = 2$. Thus the integrated *marginal* probability of $\Delta\theta_1$ (light grey area) equals the usual 68 percent between $\pm\sigma$ you already know from the normal distribution. It is also possible to read the value of the correlation coefficient ρ from the contour $\Delta\chi^2 = 1$: The contour intersects the $\Delta\theta_1$-axis at the value $\sigma_1\sqrt{1-\rho^2}$ and likewise for $\Delta\theta_2$. The larger $|\rho|$, the more elongated is the ellipse in diagonal direction. For positive ρ the long axis is in the SW–NE direction; for negative ρ it is in the NW–SE direction.

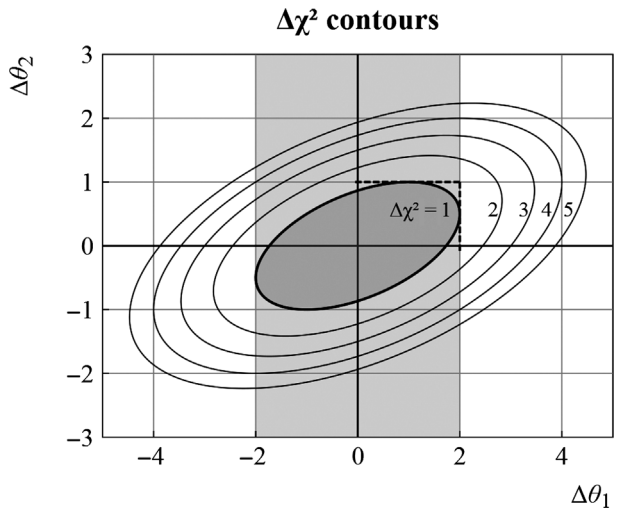

Figure 7.5 Contour plot of $\Delta\chi^2$ for the two-parameter case. The contour $\Delta\chi^2 = 1$ projects onto the axes at the values of the standard deviations ($\sigma_1 = 2$; $\sigma_2 = 1$).

See **Python code** 7.3 on page 184 to generate a contour of a two-dimensional function at a prescribed level.

When there are more than two parameters, a 2-D contour plot with the same properties can still be made for any selected pair of parameters $\Delta\theta_1, \Delta\theta_2$, but it is then required that $\Delta\chi^2(\Delta\theta_1, \Delta\theta_2)$ is minimized with respect to all other parameters. The 2-D contours are projections of an m-dimensional ellipsoid representing $\Delta\chi^2$ in the space of all parameters.

Example

Consider the urease kinetics example (page 76; data from Table 6.2). A least-squares fit has already been performed (page 93). You have a function $S(v_{max}, K_m)$, which in this case is the unweighted sum of square deviations. The minimum of this function is $S_0 = 0.171$, occurring at the parameter values [15.75, 114.64]. You have also seen before (Example 1 on page 97) that S_0 is compatible with the known inaccuracies in the measurements, according to a chi-squared analysis. What you need is χ^2 as a function of the parameters and this you obtain by scaling S such that the minimum value scales to the expected value $n - m = 4$:

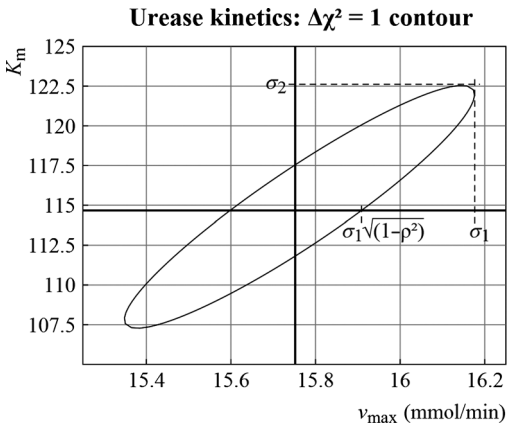

Figure 7.6 Contour plot of $\Delta\chi^2 = 1$ for the urease kinetics example. The contour projects onto the axes at the values of the standard deviations ($\sigma_1 = 0.41$; $\sigma_2 = 7.6$). The contour intersects each axis at a fraction $\sqrt{1-\rho^2}$ of the corresponding standard deviation.

$$\chi^2(v_{max}, K_m) = \frac{n-m}{S_0} S = \frac{4}{0.171} S(v_{max}, K_m). \tag{7.58}$$

Figure 7.6 shows the contour of $\chi^2 = 1$ in 2-D parameter space. You can read the standard uncertainties in the parameters from the projections of the ellipse on the axes and derive the correlation coefficient ρ from the intersection of the contour with the axes through the minimum. These values can also be taken from the array of contour points from which the contour plot is generated. The results are:

$$\sigma_1 = 0.41; \quad \sigma_2 = 7.63; \quad \rho = 0.93.$$

As you see, the two parameters are strongly correlated. A simultaneous deviation of the two parameters in the same direction (both positive or both negative) is much more probable than a simultaneous deviation in opposite directions. This is also clear from Fig. 7.6. The correlation coefficient is important if you need to predict the inaccuracy of the reaction rate at a given concentration.

See **Python code** 7.4 on page 186 to generate the χ^2-contour and to derive the uncertainties and correlation from the contour data.

If you had used your data errors (given as $\sigma_y = 0.2$ in this case) you would have found $\chi_0^2 = 0.171/0.2^2 = 4.3$ instead of the expectation 4.0.

Table 7.3 *Standard deviations and correlation coefficient for v_{max} and K_m in the urease kinetics example, computed with different methods.*

method	σ_1	σ_2	ρ
from leastsq routine	0.41	7.6	0.92
from $\Delta\chi^2 = 1$	0.41	7.6	0.92
from $\boldsymbol{C} = \boldsymbol{B}^{-1}$; $\delta = [0.2, 3.5]$	0.42	7.8	0.92
as above with $\delta = [0.0004, 0.007]$	0.36	7.0	0.93

This would have made your standard deviations an insignificant 4 percent larger.

In order to obtain the standard deviations and correlations without producing a $\Delta\chi^2 = 1$ contour, the covariance matrix $\boldsymbol{C}$ is required. This is the general procedure for more than two parameters when a 2-D plot is less suitable. If an appropriate least-squares program is used, the program can provide this matrix, as its elements are built up during the minimization procedure. Another route to the covariance matrix is the construction and subsequent inversion of the matrix $\boldsymbol{B} = \boldsymbol{C}^{-1}$, see (7.48). The elements of $\boldsymbol{B}$ can be found by evaluating $\Delta\chi^2$ at grid points near the minimum, e.g. at displacements δ_i in each of the parameters i and at displacements δ_i, δ_j for all pairs. When you perform the latter evaluation and you take the test displacements about equal to the standard deviations, you find results that are similar (but not equal) to covariances produced by the least-squares procedure (see Table 7.3). They are also close to the results obtained from the $\Delta\chi^2 = 1$ contour. However, if you take very small test displacements, the covariance matrix differs (in this example the values are 10 to 20 percent lower). The reason is that $\Delta\chi^2$ is not a pure quadratic function of the parameters as a result of the nonlinear character of the fit function. The likelihood $\exp\left(-\frac{1}{2}\Delta\chi^2\right)$ is not a pure bivariate normal distribution. It is best to use the covariances derived from $\Delta\chi^2$ in the neighborhood of 1, but be aware that the non-normal character will influence the tails of the likelihood distribution. Don't trust confidence intervals based on normal distributions and take the standard deviations with a pinch of salt.

See **Python code** 7.5 on page 187 to generate the covariance matrix for the urease kinetics example from the least-squares routine.

See **Python code** 7.6 on page 187 to generate the covariance matrix for the urease kinetics example from construction of the B matrix.

See **Python code** 7.7 on page 189 for a program that reports the results of a general least-squares fit of a predefined function to given data.

7.6 F-test on significance of the fit

If you have fitted a theoretical relation to your data points, the first question to ask is whether the fit has any significance at all. Does the fit significantly reduce the sum of square deviations compared to the sum of square deviations with respect to the average of y_i? If not, the function does not add anything to explain the data. But what is "significant" in this context?

The *total* sum of squared deviations of the data y_i with respect to their average $\langle y \rangle$ is given by (for simplicity we take all weight factors equal to one):

$$\mathrm{SST} = \sum_{i=1}^{n} (y_i - \langle y \rangle)^2. \tag{7.59}$$

This is the relevant sum of square deviations if you had no model at all. The number of degrees of freedom is $n - 1$ since you have used the data to determine one parameter (the average). *With* a model you predict values f_i that are matched as closely as possible to y_i. The residual sum of squared deviations, now called the *error sum of squares* SSE, is

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - f_i)^2. \tag{7.60}$$

If your functional relation contains m parameters to determine the f_i's, the number of degrees of freedom is $n - m$. This sum of squares is the least you can obtain with your model; it is entirely due to random errors. The difference SST − SSE is the part of the total sum of squared deviations that is explained by the model. It is called SSR, the *regression sum of squares*. Its magnitude is

$$\mathrm{SSR} = \mathrm{SST} - \mathrm{SSE} = \sum_{i=1}^{n} (f_i - \langle f \rangle)^2. \tag{7.61}$$

The number of degrees of freedom associated with SSR is $m - 1$ since the f_i's are determined by m variables, but one is used for the average. You see that all degrees of freedom are accounted for.

The validity of this equality is not immediately clear. It is valid when the f_i's have been determined such that the residues $\varepsilon_i = y_i - f_i$ form a set of independent samples from a probability distribution with zero mean. The latter implies that $\langle y \rangle = \langle f \rangle$, while independence means that ε_i is not related to f_i: $\sum_i \varepsilon_i (f_i - \langle f \rangle) = 0$. From this it follows that

$$\begin{aligned}\text{SST} &= \sum (y_i - \langle y \rangle)^2 = \sum_i (y_i - f_i + f_i - \langle f \rangle)^2 \\ &= \text{SSE} + \text{SSR} + 2\sum_i (y_i - f_i)(f_i - \langle f \rangle) \qquad (7.62) \\ &= \text{SSE} + \text{SSR}. \qquad (7.63)\end{aligned}$$

In practice this relation may not be exactly fulfilled because the last term in (7.62) may not be exactly zero.

Having separated the total sum of squared deviations into a part SSR explained by the model and a random-error part SSE, it is clear that the larger SSR is (relative to SSE), the more significant your model is. The proper statistical test is the F test (see Chapter 4, page 106, and the datasheet F-DISTRIBUTION on page 201). The F-ratio (which is the ratio of estimated variances) is given by

$$F_{m-1,n-m} = \frac{\text{SSR}/(m-1)}{\text{SSE}/(n-m)}. \qquad (7.64)$$

The cumulative F-distribution gives the probability that both sets of deviations are sampled from distributions with the same variance. It tests the null hypothesis "the model does not explain the data to a significant extent" or – equivalently – "the model is insignificant." You may reject the null hypothesis and accept the model as significant when the F-ratio exceeds the critical value F_c for which $F(F_c) > 1 - \alpha$, where α is the significance level. For example, if you have 10 data points and 3 adjustable parameters, so that $\nu_{\text{SSR}} = 2$ and $\nu_{\text{SSE}} = 7$, and you set the significance level at 1 percent, then the critical F-ratio equals 9.55 (see the second table of data sheet F-DISTRIBUTION on page 201). For larger values than 9.55 you can be confident that the model is significant.

Example

Let us return to the urease kinetics example (page 76; data from Table 6.2). The fact that the two parameters $v_{max} = 15.8 \pm 0.4$ and $K_m = 115 \pm 8$ are highly significant (see example on page 103) already suggests that the fit is very relevant. Indeed, almost all variation of the measured values is explained by the model, as you can see by evaluating the sum of squared deviations: SST = 57.02; SSR = 56.24; SSE = 0.17. Note that the sum of SSR and SSE is not quite equal to SST, which is to be expected for a nonlinear least-squares fit. The F-ratio equals [SSR/1]/[SSE/4] = 1315 and the cumulative probability of the F-distribution is 0.9999966. You can be (very) confident that the model is relevant! See Exercise 7.7 for an example with a more dubious outcome.

See **Python code** 7.8 on page 193 to compute these results.

Summary *You are now able to perform least-squares fits of parameters in a functional relation to a given data set. For functions that are linear in the parameters, the least-squares method is robust as long as the parameters are not mutually dependent. For nonlinear functions a minimum can usually be found by appropriate iterative computer programs. You can test your fit in two ways: the first question is whether the fit to the specified function is significantly better than a fit to just an average: use an F-test for this assessment. The second question is whether the residual deviations with respect to the fit behave as random samples from a distribution with variance compatible with your prior knowledge of the uncertainties: use a chi-squared test for this assessment. If OK, then you can compute the covariance matrix of the parameters from the observed sum S of square deviations. Use the dependence of S on the parameters to find this matrix.*

Exercises

7.1 Perform a linear regression on the enzyme kinetics data of Table 6.2 (page 76). Do this according to the Lineweaver–Burk plot, i.e., for $x = 1/[\mathrm{S}]$ en $y = 1/v$. Use correct standard inaccuracies for y. Give the values for v_{max} and K_m with their standard inaccuracies (be aware that for an s.d. in a combination of a and b also the correlation coefficient between a and b is needed). Compare your values with the graphical estimates from Fig. 6.2 and with the nonlinear least-squares solution of the example on page 93. Plot the data points with the best-fitted straight line.

7.2 The ΔG of a reaction was determined by measuring equilibrium constants at various temperatures. The following values were obtained:

T/K	ΔG/kJ mol^{-1}
270	40.3
280	38.2
290	36.1
300	32.2
310	29.1
320	28.0
330	25.3

The uncertainty in T is negligible and the weight factors are equal for all cases. Determine the reaction entropy $\Delta S = -d\Delta G/dT$ by fitting the values of ΔG to a linear function of T. What is the standard inaccuracy in ΔS? Extrapolate ΔG to $T = 350\,\mathrm{K}$ and give the standard inaccuracy that follows from the variance and covariance of the parameters, as found from the least-squares fit. Now do the same thing, but take for x the values of $T - 300$ instead of T itself. Discuss the differences (if any) between the two calculations.

7.3 Explain why the weight given to a data point $y_i = \log t_i$ should be proportional to t_i^2 if the t_i's are all random samples from probability distributions with the same variance. Start by deriving the variance of y assuming a constant variance σ_t^2 of t. Then relate the weight to the variance.

7.4 Perform a least-squares fit of the four-parameter double exponential function $a\exp(-px) + b\exp(-qx)$ to the data x, y from Table 6.1 on page 74. Use the Python program `fit` (code **7.7**). As initial parameter guess use the values that were determined graphically in Section 6.2. If the minimization does not reach a result after the maximum number of trials, then take the last values of the parameters as initial ones and minimize again.

7.5 You wish to measure the focal length of a positive lens on an optical bench with a ruler graduated in mm from 0 to 1000. Your lens is placed somewhere near 190 mm but it is an encased thick lens and you are not sure about its exact position. Your object (a lamp) is placed at position x and you observe the image to be sharp at position y. You estimate the s.d. σ_y of y. All data are in mm.

x	y	σ_y
60	285	1
80	301	2
100	334	3
110	383	4
120	490	5
125	680	10

Set up the parameterized equation $y \approx f(x, p)$, assuming the thin-lens formula to be valid: $1/f = 1/s_1 + 1/s_2$, where s_1 and s_2 are the distances from the lens to the object and the image, respectively. Find the least-squares solution and evaluate the best value for f and its standard inaccuracy. Discuss the validity of the functional fit. Use Python.

7.6 Prove (7.42) by setting $w_i = c/\sigma_i^2$ and eliminating c.

7.7 The F-test for relevance of a least-squares fit can be used to test if there is a *drift* in a time series, e.g. a time-dependent variable generated by a simulation that is supposed to produce stationary fluctuating quantities. Generate 100 random numbers from a normal distribution $N(0, 1)$. Perform a least-squares fit to $f_i = ai + b$, yielding the best estimates $\hat{a}$ and $\hat{b}$, and compute the standard inaccuracy in a. There are two ways to check if the drift you find is significant. First you can assess the probability that a differs from zero, i.e., the two-sided probability that $|a| \leq |\hat{a}|$. Although Student's t-distribution is the appropriate distribution, you can also use a normal distribution because with a large number of degrees of freedom the t-distribution is nearly equivalent to a normal distribution. The second method is using an F-test. Compute SSR and SSE and assess the significance of the linear regression model with an F-test. Do all this for a few fresh series of random samples and compare the two results. In order to get significant results you may add a drift term to the data.

Remark: This is done most easily by calling the routine `report`, see **Python code 6.2** on page 176.