PALSfit: A computer program for analysing positron lifetime spectra

Peter Kirkegaard, Jens V Olsen, Morten Eldrup, and Niels Jørgen Pedersen.

Final update 30-10-2014

For citation refer to official February 2009 report see http://palsfit.dk

DTU Risø Campus, Technical University of Denmark
October 2014
Abstract. A Windows based computer program PALSfit has been developed for analysing positron lifetime spectra. The program is built upon the well tested PATFIT package. The present document describes the mathematical foundation of the PALSfit model as well as a number of features of the program. The cornerstones in PALSfit are two least-squares fitting modules: PositronFit extracts lifetimes and intensities from the measured lifetime spectra. ResolutionFit determines the time resolution function from lifetime spectra which have been recorded for this purpose. PALSfit also contains a Monte-Carlo spectrum simulator. The PALSfit software can be obtained from DTU Risø Campus (www.palsfit.dk).

e-mail address: palsfit@risoe.dtu.dk

ISBN XXX-XX-XXX-XXXX-X (Internet)
ISSN XXXX-XXXX

2014
Contents

1 Introduction and overview  5

2 About PALSfit  6
   2.1 General fitting criterion  6
   2.2 PALSfit input  7
   2.3 PALSfit output in general  7
   2.4 The PositronFit model  8
   2.5 The Main Output from PositronFit  10
   2.6 The ResolutionFit model  12
   2.7 The Main Output from ResolutionFit  12

3 Experience with PALSfit  14
   3.1 PositronFit experience  14
   3.2 ResolutionFit experience  15

4 Least-squares fit and statistics  17
   4.1 Unconstrained NLLS data fitting  17
   4.2 Constraints  19
   4.3 Statistical analysis  20
   4.4 Statistical interpretations  22
   4.5 Marquardt minimization  24
   4.6 Separable least-squares technique  26
   4.7 Various mathematics, statistics, and numerics  27

5 Mathematical model details  35
   5.1 PositronFit  35
   5.2 ResolutionFit  37

6 Control files  37
   6.1 PositronFit control file  38
   6.2 ResolutionFit control file  41
   6.3 Channel ranges in PALSfit  42

7 Simulation of lifetime spectra  43

References  44
1 Introduction and overview

An important aspect of doing experiments by positron annihilation lifetime spectroscopy (PALS) is carrying out an analysis of the measured spectra to extract physically meaningful parameters. A number of computer programs have been developed over the last many years by various authors for this purpose. At our laboratory we have concentrated on developing programs for least squares fitting of such spectra.

PALSfit is our newest program of this kind. It is based on the well tested PATFIT software [1, 2], which has been used extensively by the positron annihilation community. A brief, preliminary description of the PALSfit program was given in [3]. Taking advantage of a new Graphical User Interface (GUI) for Windows, a number of user friendly facilities have been incorporated, in particular graphics displays.

The two cornerstones in PALSfit are the following least-squares fitting modules:

- **PositronFit** extracts lifetimes and intensities from lifetime spectra.
- **ResolutionFit** determines the lifetime spectrometer time resolution function to be used in PositronFit analyses.

Correspondingly PALSfit may run in either of two modes, producing a PositronFit analysis or a ResolutionFit analysis, respectively.

Common for both modules is that a model function will be fitted to a measured spectrum. The model function consists of a sum of decaying exponentials convoluted with a time resolution function, plus a constant background. The time resolution function is described by a sum of Gaussians which may be displaced with respect to each other. Various types of constraints may be imposed on the fitting parameters.

An example of a PALSfit spectrum window is shown in Fig. 1 on p. 6.

In ResolutionFit, parameters determining the shape of the resolution function can be fitted, normally by analysing lifetime spectra which contain mainly one component. The extracted resolution curve may then be used in PositronFit to analyse more complicated spectra. PALSfit provides facilities for accomplishing this process.

Note that in PositronFit the shape of the resolution function is fixed. A correction for positrons annihilating outside the sample can be made during the PositronFit analysis [1].

In Chapter 2 we give a presentation of PositronFit and ResolutionFit. This includes a short overview of the mathematical models as well as output examples. In Chapter 3 we convey some experiences we and others have gained with PALSfit and its predecessors. Chapters 4 and 5 contain mathematical and statistical details, while Chapter 6 displays examples of so-called control files produced by PALSfit (either the PositronFit or ResolutionFit module).

PALSfit is available from the website www.palsfit.dk.

A contemporary edition of the PATFIT package, roughly equivalent to PALSfit without its GUI, is available too. It contains command-driven versions of PositronFit and ResolutionFit and might be useful for batch processing or in a Linux environment.
Fig. 1 An example of a window in the PALSfit program, which shows a spectrum to be analysed, some of the input parameters for the analysis as well as icons, buttons, menus and tabs that are used to define the analysis and to display the results in numerical or graphical form.

2 About PALSfit

2.1 General fitting criterion

Common for PositronFit and ResolutionFit is that they fit a parameterized model function to a distribution (a “spectrum”) of experimental data values $y_i$. In the actual case these are count numbers which are recorded in “channels”. We use the least-squares criterion, i.e. we seek values of the $k$ model parameters $b_1, \ldots, b_k$ that minimizes

$$\phi \equiv \sum_{i=1}^{n} w_i (y_i - f_i(b_1, \ldots, b_k))^2$$

where $n$ is the number of data values, $f_i(b_1, \ldots, b_k)$ the model prediction corresponding to data value no. $i$, and $w_i$ a fixed weight attached to $i$; in this work we use “statistical weighting”,

$$w_i = \frac{1}{\sigma_i^2}$$

(2)
where $s_i^2$ is the variance of $y_i$. As some of the parameters enter our models nonlinearly, we must use an iterative fitting technique. In PALSfit we use separable least-square methods to obtain the parameter estimates. Details of the solution methods and the statistical inferences are given in Chapter 4. As a result of the calculations, a number of fitting parameters are estimated that characterize the fitted model function and hence the measured spectrum (e.g. lifetimes and intensities). A number of different constraints may be imposed on the fitting parameters. The two most important types of constraints are that 1) a parameter can be fixed to a certain value, and 2) a linear combination of lifetime intensities is put equal to zero (this latter constraint can be used to fix the ratio of intensities).

2.2 PALSfit input

PALSfit requires — together with the spectrum to be analysed — a set of input data, e.g. some characteristic parameters of the lifetime spectrometer, guesses of the parameters to be fitted, and possible constraints on these parameters. These data are organised in block structured control data sets which in turn are collected in control files. A control file may thus contain several control data sets.

PALSfit can interactively generate and/or edit the control file for either PositronFit or ResolutionFit. Previously generated control data can be used as default input values. A number of checks on the consistency of the generated control data are built into PALSfit. PALSfit is largely self-explanatory regarding input editing.

We recommend that the user reads the model descriptions in the following sections and consults Chapter 6, where the structure of the PositronFit and ResolutionFit control files is discussed. This will give a clear impression of the input possibilities in PALSfit.

Note that control files may also be produced and edited by other means than PALSfit.

2.3 PALSfit output in general

After a successful PositronFit or ResolutionFit analysis PALSfit produces three files: an analysis report (result file), a graphics file and a table file. The analysis report (result file) has the following contents:

a) An edited result section, which is the Main Output for the analysis. It contains the final estimates of the fitting parameters and their standard deviations. In addition, all the guessed input parameters as well as information on constraints are quoted. Furthermore, three statistical numbers, “chi-square”, “reduced chi-square”, and “significance of imperfect model” are shown. They inform about the agreement between the measured spectrum and the model function (Section 4.4). A few key numbers are displayed for quick reference, giving the number of components and the various types of constraints; they are identified by letters or abbreviations.

b) An input echo (optional). This is a raw copy of all the input data contained in the Control data set.

c) Fitting parameters after each iteration (optional). The parameters shown are internal; after convergence they may need a transformation prior to presentation in the Main Output.

d) An estimated correlation matrix for the parameters (optional). This matrix and its interpretation is discussed in Section 4.4.

As indicated above, the outputs b)–d) are optional, while the Main Output is always produced.
The graphics file contains data necessary for the generation of plots of measured and fitted spectra and the table file contains a table with the measured and the fitted spectrum values as a function of the channel number.

Apart from showing a result file, PALSfit displays data and results in graphical form on the screen. The PALSfit plots include measured or fitted spectra or superpositions of both, as well as deviations between them.

2.4 The PositronFit model

The PositronFit (and ResolutionFit) model function consists of a sum of decaying exponentials convoluted with the resolution function of the lifetime spectrometer, plus a constant background. Let \( t \) be the time, \( k \) the number of lifetime components, \( a_j \) the decay function for component \( j \), \( R \) the time-resolution function, and \( B \) the background. The resulting expression is given in full detail in Section 5.1; here we state the model in an annotated form using the symbol * for convolution:

\[
f(t) = \sum_{j=1}^{k} (a_j * R)(t) + B
\]

where

\[
a_j(\tau) = \begin{cases} 
A_j \exp\left(-\left(\frac{\tau - T_{0j}}{\tau_j}\right)\right), & \tau > T_0 \\
0, & \tau < T_0
\end{cases}
\]

In (4) \( \tau_j \) is the mean lifetime of the \( j \)th component, and \( A_j \) is a pre-exponential factor. The integral

\[
\int_{T_0}^{\infty} A_j \exp\left(-\left(\frac{\tau - T_{0j}}{\tau_j}\right)\right) d\tau = A_j \tau_j
\]

is called the area or the absolute intensity of the component. If not for the resolution function \( R \), \( t = T_0 \) would be the onset time for the decaying exponentials, hence \( T_0 \) is called "time-zero". We assume, furthermore, that \( R \) is given by a sum of \( k \) Gaussians which may be displaced with respect to each other:

\[
R(\tau) = \sum_{p=1}^{k} \omega_p G_p(\tau)
\]

where

\[
G_p(\tau) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{(\tau - \Delta_p)^2}{2\sigma_p^2}\right)
\]

and

\[
\sum_{p=1}^{k} \omega_p = 1
\]

The Gaussian (7) is centered around the shift \( \Delta_p \). Its standard deviation \( \sigma_p \) is related to its Full Width at Half Maximum by

\[
\text{FWHM}_p = 2\sqrt{2\ln 2} \sigma_p
\]

We also see that

\[
\int_{-\infty}^{\infty} R(\tau) d\tau = 1
\]

Regarding the time scale, the choices of \( t = 0 \) and the time unit are arbitrary. Considering the actual physical experiment, the positron annihilation lifetimes \( \tau_j \) are measured in ns (or ps). In this physical time representation all the other temporal model parameters in (3-10), i.e. \( t, \tau, T_0, \Delta_p, \) and \( \sigma_p \), will be in ns too, and it is natural to set \( T_0 = 0 \).
However, in our internal analysis it is more convenient to use a time scale directly related to the spectrum recording system. Let the multichannel analyser have \( n_{ch} \) channels, numbered \( i_{ch} = 1, 2, \ldots, n_{ch} \). Each channel represents a time slot whose common width will be used as the time unit. We let channel No. \( i_{ch} \) begin at \( t = i_{ch} - 1 \) and end at \( t = i_{ch} \). This implies that \( t = 0 \) corresponds to the beginning of the first channel. In this channel scale the “time-zero” \( T_0 \) will usually take some positive fractional value, say \( T_0 = 120.36 \). The time \( t \) defined in this way is called the channel time.

Our way of defining the channel time is by no means standard. Others may prefer to let \( t = 0 \) fall in the middle of a channel, or may choose to number the channels 0, 1, 2, \ldots In earlier versions of our software \( t = 0 \) corresponded to the left end of a fictive channel 0. Of course such differences have no influence on the analysis itself, except for the nominal value of \( T_0 \).

The curve given by (3) is continuous, but since the spectra are recorded in channels of a multichannel analyser, this curve shall for proper comparison be transformed into a histogram by integration over intervals each being one channel wide.

If all the \( n_{ch} \) channels in the spectrum are used in the least-squares analysis via (1), we have \( n = n_{ch} \) and we simply identify the channel number \( i_{ch} \) with the data value number \( i \) from (1). Thus we substitute for \( f_i \) in (1) the channel average of the model count,

\[
f_i = \int_{i-1}^{i} f(t) \, dt, \quad i = 1, \ldots, n
\]  
(11)

with \( f(t) \) given by (3), so that (11) is fitted to the measured spectrum. However, often only a subset of the channels are used in the analysis. If this subset starts in channel \( i_{ch}^{\text{min}} \) and ends in \( i_{ch}^{\text{max}} \) (inclusive), where

\[
1 \leq i_{ch}^{\text{min}} \leq i_{ch}^{\text{max}} \leq n_{ch}
\]  
(12)

we should generalize (11) to

\[
f_i = \int_{i_{ch}^{\text{min}}+i-2}^{i_{ch}^{\text{min}}+i-1} f(t) \, dt, \quad i = 1, \ldots, n
\]  
(13)

where now

\[
n = i_{ch}^{\text{max}} - i_{ch}^{\text{min}} + 1
\]  
(14)

Anyway, we obtain as the result a model for the least-squares analysis of the form

\[
f_i = \sum_{j=1}^{n_I} F_{ij} + B
\]  
(15)

where \( F_{ij} \) is the contribution from lifetime component \( j \) in spectrum channel \( i_{ch}^{\text{min}} + i - 1 \). (We relegate the full write-up of \( F_{ij} \) to Section 5.1.) We recall that \( f_i \) in (11), (13), and (15) corresponds to \( f_i(b_1, \ldots, b_k) \) in Section 2.1, formula (1).

The fitting parameters in PositronFit are the lifetimes \( \tau_j \), the relative intensities defined as

\[
I_j = \frac{A_j \tau_j}{\sum_{k=1}^{n_I} A_k \tau_k}
\]  
(16)

the time-zero \( T_0 \), and the background \( B \). Each of these parameters may be fixed to a chosen value. In another type of constraint you may put one or more linear combinations of intensities equal to zero in the fitting, i.e.

\[
\sum_{j=1}^{n_I} h_{ij} I_j = 0
\]  
(17)

Riso–R–1652-update 9
These constraints can be used to fix ratios of intensities. Finally, it is possible to fix the total area of the spectrum in the fitting,

\[ \sum_{j=1}^{k_0} A_j \tau_j + \text{background area} = \text{constant} \]  

This may be a useful option if, for example, the peak region of the measured spectrum is not included in the analysis.

Normally in an experiment a fraction \( \alpha \) of the positrons will not annihilate in the sample, but for example in the source or at surfaces. In PositronFit it is possible to make a correction for this ("source correction"). First, the raw data are fitted in a first iteration cycle. Then, the spectrum of the source correction is subtracted and the corrected spectrum fitted again in a second iteration cycle. In this second cycle it is optional to choose another number of lifetime components as well as type and number of constraints than were used in the first iteration cycle. The source correction spectrum \( f_i^s \) itself is composed of \( k_0 \) lifetime components and expressed in analogy with (15) (with \( B = 0 \)) as follows:

\[ f_i^s = \sum_{j=1}^{k_0} F_i^{s,j} \]  

If \( \tau_j^s \) and \( A_j^s \) are the lifetime and pre-exponential factor, respectively, of source-correction component \( j \), then

\[ \sum_{j=1}^{k_0} A_j^s \tau_j^s = \alpha \sum_{j=1}^{k_0} A_j \tau_j \]  

The necessary mathematical processing of the PositronFit model for the least-squares analysis is outlined in Section 5.1.

### 2.5 The Main Output from PositronFit

In the following we give an example of the Main Output part of a PositronFit analysis report produced by PALSfit, with a brief explanation of its contents (for more details about the input possibilities consult Section 6.1):

---

**PALSfit - Version 2.75 10-oct-2014 - Licensed to Jens V. Olsen**

**Input file: C:\PALSfit_report\My1_crystal.pfc**

**P S I T A N F I T . Version 2.75 11:16:32.67 30-05-14**

**Intrinsic spectrum**

---

**Data set 1**

<table>
<thead>
<tr>
<th>L</th>
<th>T</th>
<th>I</th>
<th>B</th>
<th>Z</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

**Time scale**

- **ns/channel**: 0.077300
- **Area range**
  - starts in ch. 35 and ends in ch. 512
- **Resolution**
  - FWHM (ns): 0.5296, 0.3639, 0.4096

**Function Intensities (%)**

- 20.0000, 20.0000, 60.0000

**Shifts (ns)**

- 0.0412, -0.0749, 0.0000

**Initial Parameters**

- **Time-zero (ch.ms)**: 136.30000
- **Lifetimes (ns)**: 0.23000F, 0.40000G, 2.65000
- **Intensities (%)**: 20.0000F

**Background**

- Fixed input value: 680.0000
- **Area**
  - Fixed to measured spectrum ch. 35 \( \rightarrow \) 512: 9.08171E+06

---

**Results before Source Correction**

---

**Riso-R-1652-update**
Convergence obtained after 6 iterations
Chi-square = 557.48 with 362 degrees of freedom
Lifetimes (ns) : 0.2500F 0.4401 2.6699
Intensities (%) : 30.6011 41.3989 28.0000F
Time-zero Channel number : 135.2061
Total-area From fit : 9.06171E+06 From table : 9.06171E+06

------------------ Source Correction ------------------
Lifetimes (ns) : 0.2940 0.3856
Intensities (%) : 8.0000 0.4527
Total (%) : 100.0000

------------------ 2nd cycle Parameters ------------------
Lifetimes (ns) : 0.1200G 0.3000G 1.2000G 2.9000G
Lin. comb. coeff. : -3.0000 0.0000 1.0000 1.0000
Area Fixed to measured spectrum ch. 35 -> 512 : 8.31407E+06

############################### Final Results ###############################
Data set 1

L T I B Z A G
4 0-1 2 0 1 3

Convergence obtained after 6 additional iterations
Chi-square = 437.05 with 369 degrees of freedom
Reduced chi-square = Chi-square/dof = 1.217 with std deviation 0.075
Significance of imperfect model = 99.70 %
Lifetimes (ns) : 0.1374 0.3654 1.2085 2.7592
Std deviations : 0.0048 0.0012 0.0075 0.0154
Intensities(%) LC: 11.1032 55.6674 5.0632 28.2463
Std deviations : 0.0963 0.3850 0.2996 0.4631

Background Counts/channel : 680.0000
Std deviations : fixed

Time-zero Channel number : 135.3537
Std deviations : 0.0113
Total-area From fit : 8.31407E+06 From table : 8.31407E+06

############################### Positron Fit ###############################

This output was obtained by running PALSfit with the control data set listed in Section 6.1. It does not represent a typical analysis of a spectrum, but is rather meant to illustrate a number of features of the program.

After a heading which contains the spectrum headline the key numbers are displayed in the upper right hand corner. L indicates the number of lifetime components (k0), T the number of fixed lifetimes, I the number and type of intensity constraints (a positive number for fixed intensities, a negative number for linear combinations of intensities, i.e. the number m, Section 6.1, Block 5), B the type of background constraint (KB, Section 6.1, Block 6), Z whether time-zero is free or fixed (0 = free, 1 = fixed), A the type of area constraint (KAR, Section 6.1, Block 7), and G the number of Gaussians used to describe the time resolution function (kG). The rest of the upper part of the output reproduces various input parameters, in particular those for the resolution function, the shape of which is fixed, and the initial values (G for guessed and F for fixed) of the fitting parameters.

The next part (results before source correction) contains the outcome of the first iteration cycle. If convergence could not be obtained, a message will be given and the iteration procedure discontinued, but still the obtained results are presented. Then follows information about the goodness of the fit (Section 4.4).

The next part (source correction) shows the parameters of the chosen source correction, which accounts for those positrons that annihilate outside the sample, as well as optional initial values of the fitting parameters for the second iteration cycle.
The “Final Results” part contains the number of iterations in the final cycle, followed by three lines with information about the goodness of the fit (Section 4.4). Then follows a survey of the final estimates of the fitted (and fixed) parameters and their standard deviations. The “LC” in the intensity line indicates that we have intensity constraints of the linear-combination type (cf. the negative I in the upper right hand corner). The “total area from fit” is calculated as $\sum A_j \tau_j$ plus the background inside the “area range” specified in the beginning of the Main Output. The “total area from table” is the total number of counts in the (source corrected) measured spectrum inside the “area range”.

2.6 The ResolutionFit model

The ResolutionFit model function is the same as for PositronFit, Eqs. (3–15). A few additional formulas relevant to ResolutionFit are given in Section 5.2. The purpose of ResolutionFit is to extract the shape of the resolution function. The widths and displacements (Eqs. (9) and (7)) of the Gaussians in the resolution function are therefore included as fitting parameters. However, in order not to have too many fitting parameters (which may lead to ill-defined estimates of the parameters) the intensities of the Gaussians are fixed parameters. For the same reason it is normally advisable to determine resolution functions by fitting only simple lifetime spectra, i.e. spectra containing only one major lifetime component. The extracted resolution function may then be used in PositronFit to analyse more complicated spectra. Along the same line, ResolutionFit does not include as many features as PositronFit, e.g. there is no source correction and there are no constraints possible on time-zero or on the area. However, the background can be free or fixed, just like in PositronFit.

Hence, the fitting parameters in ResolutionFit are the lifetimes ($\tau_j$), their relative intensities ($I_j$), the background ($B$), the time-zero ($T_0$), and the widths and displacements of the Gaussians in the resolution function. Each of these parameters, except $T_0$, may be constrained to a fixed value and, as in PositronFit, linear combinations of lifetime intensities may be constrained to zero in the fitting.

The various input options will be illustrated in the sample output in the next section.

2.7 The Main Output from ResolutionFit

In the following we give an example of the Main Output part of a ResolutionFit analysis report produced by PALSfit, with a brief explanation of its contents (for more details about the input possibilities consult Section 6.2):

PALSfit - Version 2.75 10-oct-2014 - Licensed to Jens V. Olsen
Input file: C:\PALSfit_report\Ch-ref.rtc

RESOLUTIONFIT Version 2.75 Job time 11:22:45.46 29-oct-14
**************************************************
39699 CU-ANNEALED
in file: \Metal defects spectra.DAT
**************************************************

Data set 1
Time scale ns/channel : 0.013400
Area range starts in ch. 5 and ends in ch. 1000
Fit range starts in ch. 140 and ends in ch. 800

Initial FWHM (ns) : 0.2600G 0.3000G 0.4000G
Resolution Intensities (%) : 77.0000 19.0000 4.0000
Function Shifts (ns) : 0.0000F 0.0223G -0.0462G

Other init. Time-zero (ch.no): 173.0000
Parameters Lifetimes (ns) : 0.1100F 0.1800F 0.4000G

Riso-R-1652-update
Background fixed to mean from ch. 650 to ch. 1000 = 12.0342

Final Results

Convergence obtained after 17 iterations
Chi-square = 315.32 with 351 degrees of freedom
Reduced chi-square = Chi-square/dof = 0.898 with std deviation 0.075
Significance of imperfect model = 8.54 %

Resolution function:

\[
\begin{array}{c|c|c|c}
\text{FWHM (ns)} & 0.2986 & 0.3100 & 0.2844 \\
\text{Std deviations} & 0.0007 & 0.0143 & 0.0104 \\
\text{Intensities (\%)} & 77.0000 & 19.0000 & 4.0000 \\
\text{Shifts (ns)} & 0.0000 & 0.0575 & -0.1900 \\
\text{Std deviations} & \text{fixed} & 0.0097 & 0.0056 \\
\end{array}
\]

Lifetime components:

\[
\begin{array}{c|c|c|c}
\text{L} & \text{T} & \text{I} \\
3 & 2 & 0 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c}
\text{Lifetimes (ns)} & 0.1100 & 0.1800 & 0.4023 \\
\text{Std deviations} & \text{fixed} & \text{fixed} & 0.0143 \\
\text{Intensities (\%)} & 86.5085 & 10.6963 & 2.7951 \\
\text{Std deviations} & 1.0833 & 1.3525 & 0.2093 \\
\end{array}
\]

Background:

\[
\begin{array}{c|c|c|c|c}
\text{Counts/channel} & 12.0342 \\
\text{Std deviation} & \text{mean} \\
\text{Time-zero} & \text{Channel number} & 170.3136 \\
\text{Std deviations} & \text{0.1380} \\
\end{array}
\]

Total-area From fit : 1.93633E+06 From table : 1.93633E+06

Shape parameters for resolution curve (nsec):

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\text{FW at} & 1/N & 0.2552 & 0.3977 & 0.4966 & 0.6217 & 0.7661 & 0.8826 & 0.9928 \\
\text{MIDP at} & 1/N & 0.0006 & 0.0010 & 0.0004 & -0.0042 & -0.0134 & -0.0193 & -0.0226 \\
\end{array}
\]

Peak position of resolution curve: Channel # 170.6092

This output was obtained by running PALSfit with the control data set listed in Section 6.2.

After a heading which includes the spectrum headline, the upper part of the output reproduces various input parameters in a way that is very similar to the PositronFit output. The important difference is that in ResolutionFit all the FWHMs and all of the displacements (called “shifts” for brevity) except one, may be fitting parameters. In addition, the background is displayed as well as the channels between which it is calculated, if the background is fixed to the mean value between these channel limits.

In the “Final Results” part the number of iterations used to obtain convergence is given first. The next three lines contain information about how good the fit is (for definition of the terms see Section 4.4).

The main part of the output, i.e. the estimated values of the fitted (and fixed) parameters and their standard deviations, follows next (for fixed parameters FIXED is written instead of the standard deviation). This part is divided into three, one giving the parameters for the resolution curve, one with the lifetimes and their intensities, and one showing the background. Each part has one or three key numbers displayed in the upper right hand corner. For the resolution function the G indicates the number of Gaussians (kG), W the number of fixed widths, and S the number of fixed displacements (shifts). For the lifetime components the L indicates the number of these (kL), T the number of fixed lifetimes, and I
the number and type of intensity constraints. As in PositronFit, a positive value means fixed intensities, while a negative value indicates constraints on linear combinations of intensities, the absolute value giving the number of constraints. Next follows the background output, where B indicates the type of background constraint (KB, Section 6.2, Block 6), and after the estimated time-zero the “total area from fit” and “total area from table” are given, both calculated as in PositronFit. Finally, for easy comparison of the extracted resolution curve with other such curves, a table of the full width of this curve at different fractions of its peak value is displayed, as well as of the midpoints of the curve compared to the peak position. The latter number clearly shows possible asymmetricities in the resolution curve. Also the channel number of the position of the peak (maximum value) of the resolution curve is given.

3 Experience with PALSfit

In this section we shall give a short account of some of the experiences we (and others) have had with PALSfit and its predecessor versions, specifically the program components PositronFit and ResolutionFit. In general, these fitting programs have proved to be very reliable and easy to use. Further discussion can be found in [1].

The aim of fitting measured spectra will normally be to extract as much information as possible from the spectra. This often entails that one tries to resolve as many lifetime components as possible. However, this has to be done with great care. Because of the correlations between the fitting parameters, and between the fitting parameters and other input parameters, the final estimates of the parameters may be very sensitive to small uncertainties in the input parameters. Therefore, in general, extreme caution should be exercised in the interpretation of the fitted parameters. This is further discussed in e.g. [4, 5, 6, 7, 8, 9, 10, 11]. In this connection, an advantage of the software is the possibility of various types of constraints which makes it possible to select meaningful numbers and types of fitting parameters.

3.1 PositronFit experience

The experience gained with PositronFit over a number of years shows that in metallic systems with lifetimes in the range 0.1 - 0.5 ns it is possible to obtain information about at most three lifetime components in unconstrained analyses [12, 13, 14] while in some insulators where positronium is formed, up to four components (unconstrained analysis) may be extracted e.g. [15, 16, 17, 18]. (This does not mean of course that the spectra cannot be composed of more components than these numbers. This problem is briefly discussed in, e.g. [4, 10]. Various other aspects of the analysis of positron lifetime spectra are discussed in for example [19, 20, 21, 22]). In this connection it is very useful to be able to change the number of components from the first to the second iteration cycle. In this way, the spectrum can be fitted with two different numbers of components within the same analysis (it is also advantageous to use this feature when a source correction removes, e.g., a long-lived lifetime component from the raw spectrum).

In our experience PositronFit always produces the same estimates of the fitted parameters after convergence, irrespective of the initial guesses (except in some extreme cases). However, others have informed us that for spectra containing very many counts (of the order of $10^7$) one may obtain different results, depending upon the initial guesses on the fitting parameters, i.e. local minima exist in the $\chi^2$ as function of the fitting parameters; these minima are often quite shallow. When this happens, PositronFit as well as most other
least-squares fitting codes are in trouble, because they just find some local minimum. From a single fitting you cannot know whether the absolute minimum in the parameter space has been found. The problem of “global minimization” is much harder to solve, but even if we could locate the deepest minimum we would have no guarantee that this would give the “best” parameter values from a physical point of view. In such cases it may be necessary to make several analyses of each spectrum with different initial parameter guesses or measure more than one spectrum under the same conditions, until enough experience has been gained about the analysis behaviour for a certain type of spectra.

3.2 ResolutionFit experience

When using a software component as ResolutionFit an important question of course is whether it is possible in practice to separate the resolution function reliably from the lifetime components. Our experience and that of others [10, 18, 23, 24, 25] suggest that this separation is possible, although in general great care is necessary to obtain well-defined results [1, 10]. The reason for this is the same as mentioned above, viz. that more than one minimum for $\chi^2$ may exist.

From a practical point of view the question arises as to whether there is too strong correlation between some of the parameters defining the resolution function and the lifetime parameters, in particular when three Gaussians (or more) are used to describe the resolution function. As in the example used in this report (Section 2.7), we have often measured annealed copper in order to deduce the resolution function. Even with different settings of the lifetime spectrometer, the copper lifetime normally comes out from a ResolutionFit analysis within a few ps (statistical scatter) of 110 ps (in agreement with results of others, e.g. [26]). Thus, the lifetime is well-defined and separable from the resolution function, even though many parameters are free to vary in the fitting procedure. However, because of the many parameters used to describe the resolution function, one frequently experiences that two (or more) different sets of resolution function parameters may be obtained from the same spectrum in different analyses, if different initial guesses are applied. The lifetimes and intensities come out essentially the same in the different analyses, the fits are almost equally good, and a comparison of the widths at the various heights of the resolution curves obtained in the analyses show that they are essentially identical. Thus, in spite of the many fitting parameters (i.e. so many that the same resolution curve may be described by more than one set of parameters), it still seems possible to separate the lifetimes and resolution function reliably, at least when the lifetime spectrum contains a short-lived component of about 90% intensity or more.

On the other hand, one cannot be sure that the lifetimes can always be separated from the resolution function easily. If, for example, the initial guesses for the fitting parameters are far from the correct parameters, the result of the fitting may be that, for instance, the fitted resolution function is strongly asymmetrical thereby describing in part the slope of the spectrum which arises from the shorter lifetime component. This latter component will then have a shorter lifetime than the correct one. Such cases — where the resolution function parameters will be strongly correlated to the main lifetime — will be more likely the shorter the lifetime is and the broader the resolution function is.

In principle, it is impossible from the analysis alone to decide whether lifetimes and resolution function are properly separated. However, in practice it will normally be feasible. If the main lifetime and the resolution curve parameters are strongly correlated, it is an indication that they are not properly separated. This correlation may be seen by looking for the changes in the lifetimes or resolution function when a small change is made in one of the resolution function parameters (intensity or one of the fitting parameters using a constraint). Other indications that the lifetimes and resolution function are not properly
separated will be that the resulting lifetime deviates appreciably from established values for the particular material or that the half width of the resolution function deviates clearly from the width measured directly with, e.g., a Co-60 source. If the lifetime and the resolution function cannot be separated without large uncertainties on both, one may have to constrain the lifetime to an average or otherwise determined value. Thus, it will always be possible to extract a resolution function from a suitably chosen lifetime spectrum.

A separate question is whether a sum of Gaussians can give a proper representation of the “true” lifetime spectrometer resolution curve, or if some other functional form, e.g., a Gaussian convoluted with two exponentials [23, 25], is better. Of course, it will depend on the detailed shape of the spectrometer resolution curve, but practical experience seems to show that the two descriptions give only small differences in the extracted shape of the curve [10, 25], and the better the resolution is, the less does a small difference influence the extracted lifetime parameters [10]. The sum-of-Gaussians used in PALSfit was chosen because such a sum in principle can represent any shape.

Once a resolution function has been determined from one lifetime measurement, another problem arises: Can this function be used directly for another set of measurements? This problem is not directly related to the software, but we shall discuss it briefly here. The accuracy of the determined resolution function will of course depend on the validity of the basic assumption about the measured lifetime spectrum from which it is extracted. This assumption is that the spectrum consists of a known number of lifetime components (e.g. essentially only one as discussed above) in the form of decaying exponentials convoluted with the resolution function. However, this “ideal” spectrum may be distorted in various ways in a real measurement. For example, instead of one lifetime, the sample may give rise to two almost equal lifetimes which cannot be separated. This will, of course, influence the resulting resolution function. So will source or surface components which cannot be clearly separated from the main component. Another disturbance of the spectrum may be caused by gamma-quanta which are scattered from one detector to the other in the lifetime spectrometer. Such scattered photons may give rise to quite large distortions of a lifetime spectrum. How large they are will depend on energy window settings and source-sample-detector arrangement of the lifetime spectrometer [10, 27, 28]. (Apart from the distortions, these spectrometer characteristics will, of course, also influence the width and shape of the correct resolution function.) In digital lifetime spectrometers that have been developed in recent years it seems possible to discriminate more effectively against some of these undesired distortions of measured spectra [29, 30, 31].

Finally, by means of an example let us briefly outline the way we try to obtain the most accurate resolution function for a set of measurements. Let us say that we do a series of measurements under similar conditions (e.g. an annealing sequence for a defect-containing metal sample). In between we measure an annealed reference sample of the same metal, with — as far as possible — the same source and in the same physical arrangement, and thereby determine the resolution curve. This is done for example on January 2, 7, 12, etc. to keep track of possible small changes due to electronic drift. We then make reasonable interpolations between these resolution curves and use the interpolated values in the analysis of the lifetime spectra for the defect containing samples. Sometimes it is not feasible to always measure the annealed sample in exactly the same physical arrangement as the defect containing sample (for example if the annealing sequence takes place in a cryostat). Then we determine resolution curves from measurements on the annealed sample inside and outside the cryostat (the results may be slightly different) before and after the annealing sequence. The possible time variation (due to electronic drift) of the resolution function is then determined from measurements with the annealed sample outside the cryostat. The same variation is finally applied to the resolution curve valid for measurements inside the cryostat.

As we often use many parameters to describe a resolution function these parameters may
appear with rather large scatter. To obtain well-defined variations with time it is often useful in a second analysis of the annealed metal spectra to constrain one or two of the parameters to some average values. With this procedure we believe that we come as close as possible to a reliable resolution function. We are reluctant to determine the resolution function directly from the spectra for the defected metal sample, as we feel that the lack of knowledge of the exact number of lifetime components makes the determination too uncertain.

Let us finally point to one more useful result of an ordinary ResolutionFit analysis apart from the extraction of the resolution curve, viz. the determination of the “source correction” (Section 2.4). If the sample gives rise to only one lifetime component, any remaining components must be due to positrons annihilating outside the sample and is therefore normally considered as a source correction. In the ResolutionFit Main Output (Section 2.7) the 0.110 ns is the annealed-Cu lifetime, while the 0.18 ns, 10.6062% component is the estimated lifetime and intensity component for the positrons annihilating in the 0.5 mg/cm² nickel foil surrounding the source material. The 0.4023 ns, 2.7951% component, that is determined by the analysis, is believed to arise from positrons annihilating in the NaCl source material and on surfaces. This component may be different for different sources and different samples (due to different backscattering). We consider the latter two components as corrections to the measured spectrum in any subsequent PositronFit analysis (when the same source and similar sample material have been used).

4 Least-squares fit and statistics

The first four sections of the present chapter contain general information about nonlinear least-squares (NLLS) methods and their statistical interpretations with relevance for PALSFit, but without going into details with the specific models involved; these are discussed in Chapters 2 and 5.

The last three sections are of a more technical nature. Section 4.5 presents essential principles of modern NLLS solution methods. Section 4.6 documents the separable least-squares technique which is of utmost importance for the efficiency and robustness of PALSFit, and Section 4.7 contains various mathematical and numerical details.

4.1 Unconstrained NLLS data fitting

We shall first present an overview of the classical unconstrained nonlinear least-squares (NLLS) method for data fitting.

In the classical setup it is assumed that some general model is given,

$$y = f(x; b_1, b_2, \ldots, b_k) = f(x; b)$$

where $x$ and $y$ are the independent and dependent variable, respectively, and $b = (b_j)$ is a parameter vector with $k$ components. (All vectors in this work are considered as column vectors.) The components $b_j$ may enter linearly or nonlinearly in (21), and so we may talk about linear and nonlinear parameters $b_j$. Further, a set of $n$ data points $(x_i, y_i)$ ($i = 1, \ldots, n$) is given, $x_i$ being the independent and $y_i$ the dependent variable; we shall here introduce the data vector $\mathbf{y} = (y_i)$, also called the spectrum. Such a spectrum is usually the result of an experiment. We assume $n \geq k$. According to the least squares principle we
should determine \( b \in \mathbb{R}^k \) such that

\[
\phi(b) = \sum_{i=1}^{n} w_i |y_i - f(x_i; b)|^2
\]

is minimized. The \( w_i \) are the weights of the data; until further notice they are just arbitrary fixed positive numbers accompanying our sample. (In many applications weights are omitted which corresponds to equal weighting, \( w_i = 1 \).)

When setting up equation (22) it was assumed that the \( x_i \) were sample points corresponding to the independent variable \( x \) in (21). In practice, however, we do not always have this situation. For example, if \( x \) represents time, and the equipment records certain events in fixed time intervals \((t_{i-1}, t_i)\) called channels, it would be natural to compare \( y_i \) with an average of the model function in (21) over \((t_{i-1}, t_i)\). Hence it is appropriate to replace (22) by

\[
\phi(b) = \sum_{i=1}^{n} w_i (y_i - f_i(b))^2
\]

In general we shall need a “recipe” \( f_i(b) \) to compute the model values to be compared with the data values \( y_i \). No sample points \( x_i \) of the independent variable enter (23) directly but may possibly be needed for calculating \( f_i(b) \). The reformulation (23) is just a generalisation of the pointwise formulation (22) who has \( f_i(b) = f(x_i, b) \). This has no influence on the least squares analysis to be described presently. In the following we shall assume that the functions \( f_i \) are sufficiently smooth in the argument \( b \).

By introducing the matrix \( W = \text{diag}(w_i) \) and the \( n \)-vector \( f(b) = (f_i(b)) \) we can express (23) in vector notation as follows:

\[
\phi(b) = \| W^{1/2} (y - f(b)) \|^2
\]

Here \( \| \cdot \| \) denotes the usual Euclidean norm. The corresponding minimization problem reads

\[
\phi_{\text{min}} = \min_{b \in \mathbb{R}^k} \{ \| W^{1/2} (y - f(b)) \|^2 \}
\]

A solution \( b \) to (25) satisfies the gradient equation

\[
\nabla \phi(b) = 0
\]

which is equivalent to the \( k \) equations

\[
\frac{\partial \phi(b)}{\partial b_j} = 0, \quad j = 1, \ldots, k
\]

By (23) and (27) we obtain

\[
\sum_{i=1}^{n} w_i (y_i - f_i(b)) p_{ij} = 0, \quad j = 1, \ldots, k
\]

where

\[
p_{ij} = \frac{\partial f_i(b)}{\partial b_j}
\]

It is practical to collect the derivatives (29) in the \( n \times k \) matrix

\[
P = (p_{ij})
\]

The equations (28) are called the normal equations for the problem. They are in general nonlinear and must be solved iteratively. Solution methods will be discussed in Sections 4.5 and 4.6. Only for linear or linearized models the normal equations are linear.

It is instructive to consider the linear case in some detail. Here (21) takes the form

\[
y = \sum_{j=1}^{k} g_j(x) b_j
\]
The $x$-dependence in $g_j(x)$ is arbitrary and may very well be nonlinear; what matters is that the fitting parameters $b_j$ should enter linearly in the model. The derivatives $p_{ij} = g_j(x_i)$ are independent of $b_j$, and (23) can be written

$$
\phi(b) = \sum_{i=1}^n w_i \left( y_i - \sum_{j=1}^k p_{ij} b_j \right)^2
$$

(32)

The normal equations take the classical form

$$
\sum_{j=1}^k \sum_{i=1}^n w_i p_{ij} p_{i'j'} = \sum_{i=1}^n w_i y_i p_{ij}, \quad j = 1, \ldots, k
$$

(33)

The equations (24–25) can be written

$$
\phi(b) = \|W^{1/2}(y - Pb)\|^2
$$

(34)

$$
\phi_{\min} = \min_{b \in \mathbb{R}^k} \{ \|W^{1/2}(y - Pb)\|^2 \}
$$

(35)

The problem (35) is solved by (33) which can be written

$$
P^T W P b = P^T W y
$$

(36)

where $T$ stands for transpose. For unweighted data we have $W = I_n$ ($I_n$ is the unit matrix of order $n$), and so

$$
P^T P b = P^T y
$$

(37)

Assuming that the coefficient matrix $P^T W P$ in (36) is nonsingular, it must be positive definite too. The same applies to $P^T P$ in (37). The case described in (31–37) represents a general linear regression model. It is a fundamental building block in NLLS procedures and their statistical analysis.

Returning to the nonlinear case we shall ignore the complications from possible non-uniqueness when solving the normal equations (28). Here we just assume that a usable solution $b$ can be found.

### 4.2 Constraints

It is important to be able to impose constraints on the free variation of the model parameters. In principle a constraint could be an equality, $h(b) = 0$, as well as an inequality $h(b) \geq 0$, where $h(b)$ is an arbitrary function of the parameter vector.

Although inequality constraints could sometimes be useful, we abandon them in this work because they would lead to quadratic programming problems, and thereby complicate our models considerably. In our algorithm there is, however, a built-in sign check on some of the nonlinear parameters (e.g. annihilation rates). Should an iteration step make such a parameter negative, a new iterate is determined by halving the correction vector from the old one. As a rule, many such “sign excursions” means an inadequate model parameterizing. On the other hand, no sign checks are made on the linear parameters.

Incorporation of general equality constraints would be possible in the framework of our least-squares method. However, apart from trivial single-parameter constraints, $b_j = c$, linear constraints on the linear parameters are sufficient for our purpose, and as we shall see, involve straightforward generalizations of the unconstrained setup discussed previously.

In Section 4.6 we shall describe the separable least-squares technique used in PALSfit. The effect of this method is to define subproblems in which the minimization takes place in the space of the linear parameters only. Hence the incorporation of constraints can just as well be discussed in terms of the linear model (31) where $\phi(b)$ is given by (32). In other words, in the constraints analysis we replace $k$ by the number $p$ of linear parameters in
the model and consider an all-linear model where \( b \) is replaced by the “linear" parameter vector \( \alpha \in \mathbb{R}^p \).

Thus we assume that \( m \) independent and consistent linear constraints on the \( p \) components of \( \alpha \) are given \( (m \leq p) \):

\[
h_1\alpha_1 + \cdots + h_p\alpha_p = \gamma_l, \quad l = 1, \ldots, m
\]

(38)

In vector form (38) reads

\[
H\alpha = \gamma
\]

(39)

where \( H = (h_{ij}) \) is an \( m \times p \) matrix and \( \gamma = (\gamma_l) \) is an \( m \)-vector. Both \( H \) and the augmented matrix \( (H, \gamma) \) are of rank \( m \).

A number of technical questions about how the constraints (38) or (39) influence the NLLS procedure will be discussed in Section 4.7.

4.3 Statistical analysis

In this and the following section we address the question of the statistical scatter in the parameters and \( \phi_{\text{min}} \) that can be expected in NLLS parameter estimation.

Suppose the spectrum \( \{y_i\} \) contains experimental values subject to statistical fluctuations, while the weights \( \{w_i\} \) are fixed. Ideally we should imagine an infinite ensemble of similar spectra \( y = (y_i) \) be given. Let us first consider the unconstrained case. Through solution of the normal equations (28) each spectrum \( y \) gives rise to a parameter estimate \( b = b(y) \). Hence also \( b \) becomes a random (vector) variable with a certain joint distribution.

We shall use the symbol \( E[\cdot] \) for expected value (ensemble mean) and \( \text{Var}[\cdot] \) for variance. We introduce the "ensemble-mean spectrum"

\[
\eta = (\eta_i) = E[y]
\]

(40)

and the corresponding hypothetic estimate

\[
b_0 = (b_0) = b(\eta)
\]

(41)

Thus \( b_0 \) is the solution of (28) corresponding to the particular spectrum \( (\eta_i) \). Now, given an arbitrary spectrum \( (y_i) \), let the corresponding parameter vector be \( b = (b_j) \). If we assume that \( b - b_0 = \Delta b = (\Delta b_j) \) is so small that our model is locally linear in \( b \) around \( b_0 \), we have to a first-order approximation

\[
f_i(b) = f_i(b_0) + \sum_{j=1}^{k} p_{ij}\Delta b_j
\]

(42)

where \( p_{ij} \) are the derivatives (29) evaluated at \( b_0 \). We insert (42) into the normal equations (28) and obtain a linear equation system of order \( k \) with \( \Delta b_j \) as unknowns. In vector notation this system reads

\[
P^TWP \Delta b = P^TW \Delta y
\]

(43)

where \( \Delta y \) is a vector with components

\[
\Delta y_i = y_i - f_i(b_0), \quad i = 1, \ldots, n
\]

(44)

We note the similarity with the linear case (36). The system (43) has the solution

\[
\Delta b = K \Delta y
\]

(45)

where

\[
K = (P^TWP)^{-1}P^TW
\]

(46)

The covariance matrix of a vector variable \( \psi \) will here be denoted \( \Sigma(\psi) \). (Other names for this matrix are dispersion matrix and variance-covariance matrix, since the diagonal row
contains the component variances. It is well-known that if two vectors \( \mathbf{v} \) and \( \mathbf{w} \) are related by a linear transformation
\[
\mathbf{w} = \mathbf{A}\mathbf{v}
\]
then
\[
\Sigma(\mathbf{w}) = \mathbf{A}\Sigma(\mathbf{v})\mathbf{A}^T
\]
Our primary goal is to estimate the covariance matrix
\[
\Sigma(\mathbf{b}) = (\sigma_{jj'})
\]
Equation (45) shows that \( \Delta \mathbf{b} \) is related to \( \Delta \mathbf{y} \) by a (locally) linear transformation, and so we obtain from (48) the approximate result
\[
\Sigma(\mathbf{b}) = \mathbf{K}\Sigma(\mathbf{y})\mathbf{K}^T
\]
We now assume that the measurements \( y_i \) are independent. Let
\[
\text{Var}[y_i] = s_i^2, \quad i = 1, \ldots, n
\]
such that \( s_i \) is the standard deviation of \( y_i \). Then \( \Sigma(\mathbf{y}) = \text{diag}(s_i^2) \). We also assume that the variances \( s_i^2 (i = 1, \ldots, n) \) are known, or at least that estimates are available. With this knowledge it is appropriate to use the statistical weighting introduced in (2) in Section 2.1. We can show that this leads to a simple form of \( \Sigma(\mathbf{b}) \). By using (50) and observing that (2) implies
\[
\mathbf{W}\Sigma(\mathbf{y}) = \mathbf{I}_n
\]
we obtain after reduction the formula
\[
\Sigma(\mathbf{b}) = (\mathbf{P}^T\mathbf{W}\mathbf{P})^{-1}
\]
which holds at least approximately. It is exact in the linear regression case (36).

Still under the assumption of a locally linear model and of statistical weighting as described, we shall next study the distribution of \( \phi_{\min} \) in (25). Here we make the additional assumptions that we have an ideal model, i.e.,
\[
f_i(\mathbf{b}_0) = \eta_i
\]
and that each measurement \( y_i \) has a Gaussian distribution,
\[
y_i \in N(\eta_i, s_i^2)
\]
Then by (44) and (54)
\[
\Delta y_i \in N(0, s_i^2)
\]
In Section 4.7 it is shown that \( \phi_{\min} \) under these assumptions has a \( \chi^2 \)-distribution with \( f \) degrees of freedom,
\[
\phi_{\min} \in \chi^2(f)
\]
where \( f \) is the number of data values minus the number of fitted parameters,
\[
f = n - k
\]
For this reason \( \phi_{\min} \) is often called \( \chi^2 \). Thus
\[
\chi^2 \equiv \phi_{\min} = \min_{b \in \mathbb{R}^k} \sum_{i=1}^n \left( \frac{y_i - f_i(\mathbf{b})}{s_i} \right)^2
\]
The results derived for \( \Sigma(\mathbf{b}) \) and \( \phi_{\min} \) are independent of the applied fitting technique. But we have assumed an unconstrained variation of all components of the \( k \)-vector \( \mathbf{b} \). When linear constraints on the linear parameters are included, the analysis still holds for a "basic subset" of \( k_{\text{ree}} \) independent parameter components, as will be shown in Section 4.7. Thus in the distribution (57–58) for \( \phi_{\min} \) we should replace \( k \) with \( k_{\text{ree}} \). To obtain \( \Sigma(\mathbf{b}) \) we incorporate the linear constraints (38) or (39) and express the remaining components
(deterministically) in terms of the free ones. These operations as well as the resulting formula for $\Sigma(b)$ are given in Section 4.7.

If the parameter vector $b$ is transformed to another vector $b_1$ before the output is presented, the covariance matrix of $b_1$ is computed as

$$\Sigma(b_1) = J \Sigma(b) J^T$$  \hspace{1cm} (60)

where

$$J = \frac{db_1}{db}$$  \hspace{1cm} (61)

is the Jacobian of the transformation, cf. (48). In PALSfit we use only simple transformations when passing from $b$ to $b_1$, or no transformation at all. Examples are lifetimes $\tau_j$ in ns instead of annihilation rate $\lambda_j$ in channels$^{-1}$, and widths in FWHM instead of in standard deviations. These give rise to trivial diagonal elements in $J$. On the other hand, the presentation of relative intensities $I_j$ instead of absolute intensities $J_j$ induces a diagonal block in the upper-left corner of $J$ with the $(j,j')$-entry $I_j(\delta_{jj'} - I_j)/J_j$.

### 4.4 Statistical interpretations

In Section 4.3 we discussed the estimation of the covariance matrix

$$\Sigma(b) = \begin{pmatrix} \sigma_{11} & \ldots & \sigma_{1k} \\ \vdots & \ddots & \vdots \\ \sigma_{k1} & \ldots & \sigma_{kk} \end{pmatrix}$$  \hspace{1cm} (62)

The standard deviations of the estimated parameters are extracted from its diagonal as $\sigma_j = \sqrt{\sigma_{jj}}$, while the off-diagonal entries contain the covariances. In the usual way we construct the correlation matrix

$$R = \begin{pmatrix} 1 & \ldots & \rho_{1k} \\ \vdots & \ddots & \vdots \\ \rho_{k1} & \ldots & 1 \end{pmatrix}$$  \hspace{1cm} (63)

by the formula

$$\rho_{jj'} = \sigma_{jj'}/(\sigma_j \sigma_{j'})$$  \hspace{1cm} (64)

A consequence of the assumed normal distribution of $y_i$ is that the parameter estimates too will be (approximately) normally distributed and their joint distribution is completely determined by the covariance matrix $(\sigma_{jj'})$. The natural statistical interpretation of $(\sigma_{jj'})$ or $(\sigma_j, \rho_{jj'})$ is an estimate of the covariance structure of the computed parameters in a series of repetitions of the spectrum recording under identical physical conditions. We would expect that $(\sigma_{jj'})$ might show a good deal of scatter in such a series.

The standard deviations and correlation coefficients may be used to compute estimated standard deviations of new parameters that are functions of the primary parameters presented in the output from PALSfit, e.g. a mean lifetime or a trapping rate. The standard deviation of such a parameter, $z$, is given (to a first-order approximation) by:

$$\sigma_z = \left\{ \sum_{j=1}^{k} \frac{\partial z}{\partial b_j} \frac{\partial z}{\partial b_j} \rho_{jj'} \sigma_{jj'} \right\}^{1/2} = \left\{ \sum_{j=1}^{k} \left( \frac{\partial z}{\partial b_j} \right)^2 \sigma_j^2 + 2 \sum_{j=1}^{k} \sum_{j'=j+1}^{k} \frac{\partial z}{\partial b_j} \frac{\partial z}{\partial b_{j'}} \rho_{jj'} \sigma_{jj'} \right\}^{1/2}$$  \hspace{1cm} (65)

This result follows by the transformation rule (60).

There is another property of the correlation matrix which might be useful in practice. Suppose an analysis of a given spectrum results in an estimated parameter vector $b = (b_j)$, $j = 1, \ldots, k$. One may ask: What happens to the remaining components if one of them,
say \( b_1 \), is forced to be shifted a small amount \( \Delta b_1 \), and the analysis is repeated with the same spectrum? It can be shown that the other components will be shifted according to the formula

\[
\Delta b_j = (\sigma_j/\sigma_1) \rho_{ij} \Delta b_1, \quad j = 2, \ldots, k
\]

(66)

A proof can be found in Section 4.7.4. The formula (66) refers to a single spectrum and is therefore deterministic. In principle its validity is restricted to small shifts due to the nonlinearity of our models. In our experience the formula is applicable up to at least \( \Delta b_1 \approx 2 \) for well-defined fitting problems with small \( \sigma_j \). For fits with large \( \sigma_j \) it seems to be valid only up to \( \Delta b_1 \approx 0.1 \) for fits with small \( \sigma_j \). For fits with large \( \sigma_j \) it seems to be valid only up to \( \Delta b_1 \approx 0.1 \times \sigma_1 \), and in certain pathological cases it fails completely; such failures may be ascribed to imperfect models or strong nonlinearities.

We saw in Section 4.3 that \( \chi^2 = \phi_{\text{min}} \) in (59) under certain assumptions has a \( \chi^2 \) distribution with \( f \) degrees of freedom, i.e. (57) holds good with

\[
f = n - k + m = n - k_{\text{free}}
\]

(67)

Here \( m \) is the number of constraints, so that \( k_{\text{free}} = k - m \) is the effective number of free parameters in the estimation. The mean and variance of \( \phi_{\text{min}} = \chi^2 \) are

\[
E[\chi^2] = f
\]

(68)

and

\[
\text{Var}[\chi^2] = 2f
\]

(69)

From the \( \chi^2 \) statistics one can derive a “goodness-of-fit” significance test for the validity of the asserted ideal model, cf. (54). In such a \( \chi^2 \)-test we compute the probability \( P\{\chi^2 < \chi^2_{\text{obs}}\} \) that a \( \chi^2 \)-distributed variable with \( f \) degrees of freedom will not exceed the observed value \( \chi^2_{\text{obs}} \). A value close to 100% indicates systematic deviation from the assumed model, and we use the phrase “significance of imperfect model” for this probability. We also compute the quantity

\[
V = \chi^2 / f,
\]

(70)

with mean

\[
E[V] = 1
\]

(71)

and variance

\[
\text{Var}[V] = 2/f
\]

(72)

\( V \) is sometimes called the “reduced chi-square” or the “variance of the fit”; with a good fit this quantity should be close to unity.

We conclude this section with some comments on the underlying assumptions in the statistical NLLS analysis which were:

1. Small fluctuations of each data value \( y_i \), i.e. \( \text{Var}[y_i] \) small.
2. Our model is only weakly nonlinear in the parameter vector \( b \).
3. An ideal model which means that (54) holds.
4. The data values \( y_i \) are independent.
5. Each \( y_i \) has a Gaussian distribution.
6. “Statistical weighting” (2).
7. The population variances \( \text{Var}[y_i] \) are known in advance.

Assumptions 1, 2, 3 should be considered together; for example, violation of 1 and 3 may both invalidate the linear approximation (42).
For NLLS problems with strong non-linearities it is well-known that our covariance matrix formula, which is based on linear expansion of the model, may in general produce over-optimistic standard deviation estimates. Nevertheless, such estimates could be useful for qualitative purposes.

Assumption 1 is a fair approximation in PALSfit applications, where it should be understood in the relative meaning; it holds provided the counts \( y_i \) are not too small.

Assumption 3 expresses that our model “explains” the observed data perfectly, apart from the inevitable statistical noise. This hypothesis was subject to a chi-square goodness-of-fit test as explained.

Assumption 4 is natural in many applications; however in practice some measurements might show a certain correlation between neighboring data values.

Assumption 5 is needed only for the analysis of the goodness-of-fit. Many distributions encountered in practice do not deviate much from the normal distribution and thus admit an approximately correct analysis. In particular, this is true for Poisson counting statistics, again provided the counts are large enough.

Regarding Assumption 6, statistical weighting is a convenient means to equalize the impact from the individual observations \( y_i \) on the fit. To accomplish it we shall need (estimates of) the variances \( \text{Var}[y_i] \) [see also Assumption 7].

Regarding Assumption 7, the theoretical values of the population variances \( \text{Var}[y_i] \) are sometimes unavailable and need to be estimated. In some applications the variances are only known up to a constant of proportionality. By using statistical weighting nevertheless, this would not affect the outcome of the NLLS parameter estimation itself. However, the chi-square analysis would not be possible in the usual way due to the lack of normalization.

### 4.5 Marquardt minimization

As mentioned in Section 4.1 the normal equations (28) are in general nonlinear and must be solved iteratively. We now describe such an iterative method called Marquardt’s principle, which is an efficient combination of two classical unconstrained minimization procedures; constraints will be taken care of as described in Sections 4.2 and 4.7.

Basically, we use Newton’s iterative method (other names are the Gauss-Newton or the Taylor series method), which we shall presently explain. However, first we shall prove the following expansion formula which is approximately correct provided \( d \) is small, the fit is good, the model locally linear, and \( b \) is close to the solution of the NLLS problem:

\[
\phi(b + d) = \phi(b) + \nabla \phi(b) \cdot d + d^T P^T WP d
\]

with the usual meaning of \( W \) and \( P \). Using a quadratic Taylor expansion we obtain

\[
\phi(b + d) = \phi(b) + \nabla \phi(b) \cdot d + \frac{1}{2} d^T S d + O(||d||^3)
\]

Here \( S = \{ s_{ij} \} \) is the Hessian of \( \phi(b) \). From the expression (23) we find

\[
s_{jj} = \partial^2 \phi(b) = 2 \sum_{i=1}^{n} w_i \left( \frac{\partial f_i}{\partial b_j \partial b_j} - (y_i - f_i) \frac{\partial^2 f_i}{\partial b_j \partial b_j} \right)
\]

with \( f_i = f_i(b) \). We shall neglect the term \( \sum_{i=1}^{n} w_i (y_i - f_i) \frac{\partial^2 f_i}{\partial b_j \partial b_j} \) in (75). The reason for doing so is that we expect some cancellation to take place in the summation process, because the residuals \( y_i - f_i \) are supposed to fluctuate around zero when the fit is good. We have also assumed that the second derivatives \( \frac{\partial^2 f_i}{\partial b_j \partial b_j} \) which express the nonlinearity of the model, are not too large. Hence, approximately

\[
S = 2P^T WP
\]
Inserting this in (74) we establish (73). Returning to Newton’s method, let \( \mathbf{b} \) be a guessed or previously iterated parameter vector. Newton’s correction step \( \mathbf{d} \) now solves the local minimization problem
\[
\min_{\mathbf{d} \in \mathbb{R}^k} \{ \phi(\mathbf{b} + \mathbf{d}) \}
\]
where \( \phi(\mathbf{b} + \mathbf{d}) \) is approximated by (73). For brevity we shall write
\[
\mathbf{A} = \mathbf{P}^T \mathbf{W} \mathbf{P}
\]
Assuming that \( \mathbf{P} \) has full rank, \( \mathbf{A} \) will be positive definite. By taking gradients we obtain
\[
\nabla \phi(\mathbf{b} + \mathbf{d}) = \nabla \phi(\mathbf{b}) + 2\mathbf{A}\mathbf{d} + \mathcal{O}(||\mathbf{d}||^2)
\]
We equate this to zero and then compute the Newton step from the normal equation system (cf. (36) and (43))
\[
\mathbf{A}\mathbf{d} = \mathbf{g}
\]
Here the vector \( \mathbf{g} \) is given by
\[
\mathbf{g} = -\frac{1}{2} \nabla \phi(\mathbf{b})
\]
According to (23) its components are
\[
g_j = \sum_{i=1}^{n} w_i(y_i - f_i(\mathbf{b}))(y_i - f_i(\mathbf{b})), \quad j = 1, \ldots, k
\]
Subsequently \( \mathbf{b} + \mathbf{d} \) replaces \( \mathbf{b} \) as the new iterate, and the iterations continue. With the pure Newton method we cannot guarantee that the new \( \phi = \phi(\mathbf{b} + \mathbf{d}) \) is smaller than the old one. Indeed the procedure often tends to diverge due to strong nonlinearities, in particular when the initial guess is bad. To ensure a decrease in \( \phi \) we introduce the Marquardt [32] modification of (80).
\[
(\mathbf{A} + \lambda \mathbf{D})\mathbf{d} = \mathbf{g}
\]
where \( \mathbf{D} \) is a diagonal matrix with the same diagonal row as the positive definite matrix \( \mathbf{A} \). \( \lambda \) is a parameter that is at our disposal. It provides interpolation between Newton’s method and a gradient-like method. The former is obtained by setting \( \lambda = 0 \), cf. (80). On the other hand, when \( \lambda \to \infty \) we obtain a solution vector proportional to \( \mathbf{D}^{-1}\mathbf{g} \). According to (81) \( \mathbf{g} \) is proportional to the negative gradient vector \( -\nabla \phi \), so \( \mathbf{D}^{-1}\mathbf{g} \) becomes a scaled version of \( -\nabla \phi \) and shares with this the property that \( \phi \) (assumed > \( \phi_{\text{min}} \)) certainly decreases initially along the correction vector, although it need not have the steepest descent direction. We can now roughly sketch Marquardt’s procedure. The equation to be solved at iteration number \( r \) reads
\[
(\mathbf{A}^{(r)} + \lambda^{(r)} \mathbf{D}^{(r)})\mathbf{d}^{(r)} = \mathbf{g}^{(r)}
\]
From its solution \( \mathbf{d}^{(r)} \) we calculate
\[
\mathbf{b}^{(r+1)} = \mathbf{b}^{(r)} + \mathbf{d}^{(r)}
\]
and a new \( \phi \)-value, \( \phi^{(r+1)} \). Now it is essential that \( \lambda^{(r)} \) is so chosen that
\[
\phi^{(r+1)} \leq \phi^{(r)}
\]
If we are not already at the minimum, it is always possible to satisfy (86) by selecting a sufficiently large \( \lambda^{(r)} \), and so we avoid the divergence problems encountered in Newton’s method. However, \( \lambda^{(r)} \) should not be chosen unnecessarily large, because we then get a small correction vector of gradient-like type which would give slow convergence. In the later iterations, when convergence is approached, \( \lambda \) should be small. Then we approach Newton’s method which has a fast (quadratic) rate of convergence near the minimum. The procedure has converged when \( \phi^{(r)} \) and \( \mathbf{b}^{(r)} \) are stationary with increasing \( r \). For the detailed strategy we refer to Marquardt [32]. The algorithm is sometimes called the Levenberg-Marquardt method (LM) since Levenberg [33] already in 1944 put forward essential parts of the ideas taken up by Marquardt in 1963 [32].

Riso–R–1652-update 25
Over the years LM has undergone a number of refinements, adding more robustness to it. In earlier versions of PALSfit we used LM as in [32]. But pure LM puts no bounds on the step vector $d(\lambda)$. Modern LM implementations use a "trust-region" enhancement and replace the unrestricted minimization of $\phi$ by the quadratic programming problem

$$\min_{d(\lambda)} \left\{ \phi : \|D^{1/2}d(\lambda)\| \leq \Delta \right\} \quad (87)$$

The effect is to restrict the size of $d = d(\lambda)$. The bound $\Delta$ is adjusted each time a major iteration step begins and is decreased if the solution of (87) does not provide a suitable correction. We have adopted this idea for use in PALSfit from the work of Moré [34], as implemented in the software package MINPACK-1 [35] for unconstrained least-squares minimization. A subroutine from this package, LMPAR, performs minor iterations by finding a value of $\lambda$ that solves (87) approximately. The optimal $\lambda$ is saved for use as an initial estimate in the next major step. Details of this technique are found in Moré [34].

### 4.6 Separable least-squares technique

A substantial gain in computing efficiency can be obtained when some of the $k$ components of the parameter vector $b$ enter our model linearly. Indeed this is the case in PALSfit. The least-squares problem is then called separable or semilinear. Separable procedures have been studied by several authors [36, 37, 38, 39, 40, 41] and have proved successful in this work and in many other applications as well. In the separable case (21) can be written

$$y = f(x; b) = f(x; \alpha, \beta) = \sum_{j=1}^{p} \alpha_j f_j(x; \beta) \quad (88)$$

This means a partitioning of the $k$-vector

$$b = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad (89)$$

into a "linear" $p$-vector $\alpha = (\alpha_j)$ and a "nonlinear" $q$-vector $\beta = (\beta_j)$, where $p + q = k$. Corresponding to (88) we have

$$f_i(b) = \sum_{j=1}^{p} \alpha_j f_{ij}(\beta) = (F(\beta)\alpha)_i, \quad i = 1, \ldots, n \quad (90)$$

where $F = F(\beta)$ is an $n \times p$ matrix with elements $f_{ij} = f_{ij}(\beta)$. With these definitions $f(b)$ entering (25) can be written

$$f(b) = F(\beta)\alpha \quad (91)$$

In separable NLLS we consider the linear subproblems of (25) where $\beta$ is fixed and $\alpha$ varies:

$$\min_{\alpha \in \mathbb{R}^p} \{ ||W^{1/2}(y - F(\beta)\alpha)||^2 : \beta \text{ fixed} \} \quad (92)$$

Considering first the unconstrained case, the standard linear least-squares analysis tells that $\alpha = \alpha(\beta)$ is the solution of the $p$th-order normal equation system

$$F^T WF\alpha = F^T Wy \quad (93)$$

cf. the linear regression case (31–37). Turning to the determination of the nonlinear part $\beta$ of the parameter vector $b$, we realize that an iterative method is needed. In fact, there will be an outer loop, where each step provides a correction vector $d$ to $\beta$, and an inner procedure which invokes a linear minimization (92–93) each time a new trial value of $\beta$ is chosen. We can formulate the nonlinear outer minimization as follows:

$$\min_{\beta \in \mathbb{R}^q} \{ \phi(\beta) = ||W^{1/2}(y - F(\beta)\alpha(\beta))||^2 \} \quad (94)$$
We solve (94) by a modified Marquardt procedure as explained in Section 4.5, where \( b \) should be replaced by \( \beta \). Indeed, equation (83) takes the form
\[
(P^T WP + \lambda D) d = P^T W (y - F(\beta)\alpha)
\] (95)
where \( P \) is now the \( n \times q \) matrix with elements
\[
p_{ij} = \frac{\partial f_i}{\partial \beta_j}
\] (96)
and \( D \) is a diagonal matrix with the same diagonal row as \( P^T WP \).

A crucial point in the separable procedure is the evaluation of (96), which can be accomplished by considering the vector \( f = f(b) \) in (91). Note that \( f \) depends on \( \beta \) directly through \( F \) and indirectly through \( \alpha \); hence
\[
\frac{\partial f}{\partial \beta_j} = \frac{\partial F}{\partial \beta_j} \alpha + F \frac{\partial \alpha}{\partial \beta_j}
\] (97)
To find \( \partial \alpha / \partial \beta_j \) we take the derivative of both members of (93). This leads to
\[
F^T WF \frac{\partial \alpha}{\partial \beta_j} = \frac{\partial F^T}{\partial \beta_j} W (y - F \alpha) - F^T W \frac{\partial F}{\partial \beta_j} \alpha, \quad j = 1, \ldots, q
\] (98)
For an ideal model the term in (98) containing the residual vector \( y - F \alpha \) is negligible when the minimum is approached, but is important when the current iterate is far from convergence.

Now we can give a summary of the complete strategy for the unconstrained separable minimization of \( \phi \): Start the outer iterations from a guessed value of \( \beta \), and select suitable initial values for \( \lambda \) and the bound \( \Delta \). For each outer iteration, solve the linear subproblem (92–93) for \( \alpha \) and calculate \( \phi \). Compute the Jacobian elements \( \partial f / \partial \beta_j \) from (98) and (97), and form \( P \) and \( D \). Then enter an inner procedure and find near-optimal values of \( \lambda \) and the correction vector to \( \beta \), \( d = d(\lambda) \), using Marquardt’s method with Moré’s modification. Update the bound \( \Delta \), replace \( \beta \) by \( \beta + d \), and resume the outer iteration loop. The procedure is finished when \( \phi \) has proved to be stationary.

When implementing our separable algorithm, there is a practical difficulty in handling \( \partial f / \partial \beta_j \) in (97). For each data value we must evaluate a \( p \times q \) matrix of scalar derivatives which means altogether \( n \times p \times q \) values. To reduce the memory demand we use a packed ("sparse-matrix") scheme for storing only the nonzero derivatives.

Linear constraints on linear model parameters, as they occur in PALSfit, are readily integrated in the separable NLLS procedure, cf. Sections 4.2 and 4.7.

The numerical solution of many of the linear-algebraic and optimization subproblems in our algorithm is accomplished by software from the standard packages LINPACK [42] and MINPACK-1 [35]. To accommodate application of this software we found it convenient to rescale the NLLS problem, cf. Section 4.7. In that section we also give some comments on the practical numerical solution of NLLS subproblems.

4.7 Various mathematics, statistics, and numerics

In this section a number of technical details are collected. They all have relevance to the previous sections in this chapter.

4.7.1 Estimation of background and weight smoothing

Considering Assumptions 6 and 7 in Section 4.4 we face the problem that we do not know \( s_i^2 = \text{Var}[y_i] \) when setting up the statistical weighting (2). Each count \( y_i \) is distributed in a
Poisson distribution with a certain mean value \( \eta_i \),

\[
y_i \in P(\eta_i)
\]

implying that

\[
E[y_i] = \eta_i \quad (100)
\]

\[
\text{Var}[y_i] = \eta_i \quad (101)
\]

The ideal weighting

\[
w_i = \frac{1}{\eta_i} \quad (102)
\]

would provide approximately central (i.e. unbiased) least-squares estimates of the model parameters. Since the \( \eta_i \) are unknown we must use some kind of approximation. We may simply take

\[
w_i = \frac{1}{y_i} \quad (103)
\]

or rather use the modified formula

\[
w_i = \frac{1}{\max(y_i, 1)} \quad (104)
\]

because it is possible to record \( y_i = 0 \). However, in the following we shall assume that the probability of this is negligible,

\[
P\{y_i = 0\} \approx 0 \quad (105)
\]

The “raw” weights (103) will normally fluctuate, and we shall now show that this may induce a bias on the parameter estimates. In a fitting model (3) with a free background \( B \) it is often the case that \( B \) is virtually independent of the other parameters. This means that obtaining a least-squares estimate \( B^* \) of \( B \) is essentially a one-parameter problem,

\[
B^* = \arg\min_{B \in \mathbb{R}} \sum_{i=1}^{n} w_i (y_i - B)^2
\]

which has the unique solution

\[
B^* = \frac{\sum_{i} w_i y_i}{\sum_{i} w_i} \quad (106)
\]

In the “theoretical” case (102) we evaluate \( B^* = B^*_{0} \) by inserting (102) in (107). Because \( B \) is the only parameter we have \( \eta_i = B \) and then

\[
B^*_{0} = \frac{\sum_{i} y_i}{n} = \langle y_i \rangle \quad (108)
\]

where \( \langle \cdot \rangle \) stands for averaging over channels. \( B^*_{0} \) is indeed central as

\[
E[B^*_{0}] = E[y_i] = \eta_i = B
\]

On the other hand, in the “real” case (103), we evaluate \( B^* = B^*_{1} \) by inserting (103) in (107), and then it turns out that \( B^*_{1} \) is not central. In fact we can show that \( B^*_{1} \) underestimates the background roughly by 1,

\[
E[B^*_{1}] \approx B - 1 \quad (110)
\]

under the additional assumptions that the number of channels \( n \) in the analysis as well as the counts \( y_i \) and the background \( B \) are reasonably large. We find

\[
B^*_{1} = \frac{n}{\sum_{i} y_i} = \frac{1}{\langle y_i \rangle} \quad (111)
\]

From the Poisson distribution (99) we obtain the probability

\[
P\{y_i = k\} = \frac{B^k}{k!} e^{-B} \quad (112)
\]
We then have
\[ E[B_1^4] \approx \frac{1}{E[B_1]} \]  
(113)

Now (112) and (105) give
\[ E\left[ \frac{1}{y_i} \right] \approx \sum_{k=1}^{\infty} \frac{1}{k} \frac{B^k}{k!} e^{-B} \]  
(114)

This sum can be evaluated with the result
\[ E\left[ \frac{1}{y_i} \right] \approx e^{-B} (\text{Ei}(B) - \gamma - \log B) \]  
(115)

where Ei is the exponential integral and \( \gamma \) is Euler’s constant. Since \( B \) was assumed fairly large, we can use the asymptotic expansion
\[ \text{Ei}(B) \sim e^B \left\{ 1 + \sum_{r=1}^{\infty} \frac{r!}{B^r} \right\}, \quad B \to +\infty \]  
(116)

We shall use the first-order approximand, and discarding \( \gamma \) and \( \log B \) in (115) we then obtain
\[ E\left[ \frac{1}{y_i} \right] \approx \frac{1}{B} \left( 1 + \frac{1}{B} \right) \]  
(117)

Finally (113) gives the approximate result (110) that the bias is \( -1 \).

An approximate removal of the background bias can be accomplished by *weight smoothing* which can be done in several ways. Earlier PALSfit versions carried out an extra iteration cycle after the first cycle had converged. (This should be done anyway in case of source correction in *PositionFit*.) Between the two iteration cycles the weight fluctuations were removed by replacing (103) by the weights
\[ w_i = \frac{1}{y_i} \]  
(118)

where \( y_i^0 \) is the model-predicted \( y_i \)-estimate at the end of the first cycle. This procedure did in fact remove the bias, but also has some drawbacks. When the fit is not perfect, the smoothing not only removes the fluctuations but also distorts the overall shape of the weight function, which results in an unreliable statistical analysis.

The present PALSfit version uses instead a heuristic *non-parametric* (i.e. model-independent) weight smoothing, once and for all at the beginning of the analysis. Guided by log-plots of various typical spectra \( y(t) \), we make a continuous piecewise linear fit of \( \log_{10} y(t) \), in which the knot abscissas and knot values are the fitting parameters. Our algorithm uses separable least squares optimization where the inner minimisation determines the ordinates for given knot abscissas, while the outer minimisation furnishes the optimal position of the knot abscissas. The latter is controlled by an “amoeba type” procedure (Nelder and Mead, in Numerical Recipies [43]). We obtain good results by dividing the spectrum in at most 4 parts and using the piecewise linear algorithm on each of these. There will be fewer than 4 parts when, for example, the peak position is outside the interval from \( t_{ch}^{\min} \) to \( t_{ch}^{\max} \), cf. (12). The number of segments used in each part is normally 4 but less when there are few channels in a part. Common fit values at contiguous parts are obtained by a simple averaging. This ensures an overall continuity of the fit. When the minimum value of the counts within some part exceeds a critical value, say 1000, we abandon the smoothing there, since the spectrum within that part may already be considered smooth. From the computed fit values we finally obtain the smoothed weights \( w_i \) by taking antilogarithm and reciprocal.

The new way of estimating the weights has very little influence on the values of the fitted parameters. Typically the parameters change by an amount which is an order of magnitude smaller than the parameter standard deviations, or less. The main effect is that the program has become more robust against some extreme choices of input parameters to the analysis. Moreover the reliability of the statistical analysis for imperfect fits is improved.
4.7.2 Implementation of linear constraints

We consider the constrained linear least-squares problem (cf. (38–39) and (35)),
\[
\phi_{\text{min}} = \min_{\alpha \in \mathbb{R}^p} \{ ||W^{1/2}(y - P\alpha)||^2 : H\alpha = \gamma \} \tag{119}
\]
This subproblem is part of the separable NLLS method discussed in Section 4.6, where an optimal linear parameter vector \( \alpha \in \mathbb{R}^p \) was computed for a given nonlinear parameter vector \( \beta \in \mathbb{R}^q \). Thus the derivative matrix \( P = (p_{ij}) = (\partial f_i/\partial x_j) \) is here of size \( n \times p \). One way of handling this constraints problem would be to use Lagrange multipliers. This method was used in early predecessors of PALSfit. As a result, the normal equation system (36) was extended to a block matrix system:
\[
\begin{pmatrix}
P^TWP & H^T \\
H & 0
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\mu
\end{pmatrix}
=
\begin{pmatrix}
P^TWy \\
\gamma
\end{pmatrix} \tag{120}
\]
where the vector \( 2\mu \) contains the Lagrange multipliers. Although (120) is simple enough, there are some drawbacks in this procedure. The coefficient matrix in (120) is not positive definite as was \( P^TWP \). This could reduce the numerical stability of the calculations. We also note that the constraints increase the size of the “effective normal equation system” from \( p \times p \) to \( (p + m) \times (p + m) \). Below we describe an elimination method which is now in use in PALSfit. It offers better stability, reduced computer time, and reduced storage demand. Since the rank of \( H \) is \( m \), we can construct a nonsingular matrix by picking \( m \) independent columns from \( H \). A suitable permutation will bring these columns to the \( m \) first positions. This can be expressed in terms of a \( p \)-th order permutation matrix \( \Pi \) by
\[
\Pi H \Pi = H' = \begin{bmatrix}
m & p-m \\
m & p-m
\end{bmatrix}
\begin{bmatrix}
B & N
\end{bmatrix} \tag{121}
\]
In the language of linear programming we call \( B \) a “basis matrix” for \( H \), whereas the columns in \( N \) are called “nonbasic”. Because \( \Pi \) is orthogonal, \( \Pi H^T = I_p \), (39) can be written
\[
H'\alpha' = \gamma \tag{122}
\]
with
\[
\alpha' = \Pi^T \alpha \tag{123}
\]
Equation (122) has the complete solution
\[
\alpha' = \alpha'_0 + Y't \tag{124}
\]
where
\[
\alpha'_0 = \begin{bmatrix}
m \\
p-m
\end{bmatrix}
\begin{bmatrix}
B^{-1} \gamma \\
0
\end{bmatrix} \tag{125}
\]
\[
Y' = \begin{bmatrix}
m \\
p-m
\end{bmatrix}
\begin{bmatrix}
p-m \\
p-m
\end{bmatrix}
\begin{bmatrix}
-B^{-1} N \\
I_{p-m}
\end{bmatrix} \tag{126}
\]
and \( t \in \mathbb{R}^{p-m} \). From (123) we get the complete solution of (39):
\[
\alpha = \alpha_0 + Yt, \quad t \in \mathbb{R}^{p-m} \tag{127}
\]
where
\[
\alpha_0 = \Pi \alpha'_0 \tag{128}
\]
and
\[
Y = \Pi Y' \tag{129}
\]
It is practical to partition \( \Pi \) in column sections as follows:
\[
\Pi = \begin{bmatrix}
m & p-m \\
p & p-m
\end{bmatrix}
\begin{bmatrix}
\Pi_1 \\
\Pi_2
\end{bmatrix} \tag{130}
\]
Then (128) becomes
\[ \alpha_0 = \Pi_1 B^{-1} \gamma \]  
(131)

To express (129) we note that
\[ N = H \Pi_2 \]  
(132)

and so
\[ Y = (I_p - \Pi_1 B^{-1} H) \Pi_2 \]  
(133)

We have
\[ HY = H'Y' = 0 \]  
(134)

and the columns of \( Y \) form a basis of the null space or kernel of \( H \). Using (127) we can reformulate the constrained \( p \)-dimensional problem (119) to an unconstrained \( (p - m) \)-dimensional problem:
\[ \phi_{\min} = \min_{t \in \mathbb{R}^{p-m}} \{ ||W^{1/2}(y - P \alpha_0 - PYt)||^2 \} \]  
(135)

We see that this can be derived from (35) by substituting \( p - m \) for \( k \), \( y - P \alpha_0 \) for \( y, PY \) for \( P \), and \( t \) for \( b \). Thus we can immediately write down the normal equation system for (135) by making the corresponding substitutions in (36):
\[ (PY)^T W (PY) t = (PY)^T W (y - P \alpha_0) \]  
(136)

Next we shall derive an expression for the covariance matrix \( \Sigma(b) \) of the total parameter vector \( b = (\alpha, \beta) \) when the constraints (38) or (39) are included. Recalling that \( \Sigma(b) \) is independent of the actual fitting method, we can estimate it by perturbing the solution vector \( b \) at the end of iterations. From the normal equation system (136) we deduce in analogy with (42–53) that
\[ \Sigma(t) = \{(P_\alpha Y)^T W (P_\alpha Y)\}^{-1} \]  
(137)

We have written \( P_\alpha \) for \( P \) since we shall now reserve the notation \( P \) for the \( n \times k \) matrix containing derivatives \( \partial f_i/\partial p_j \) with respect to all the \( p+q \) components of \( b = (\alpha, \beta) \). Thus we shall write
\[ P = \begin{pmatrix} P_\alpha & P_\beta \end{pmatrix} \]  
(138)

We note that
\[ P = \begin{pmatrix} P_\alpha & P_\beta \end{pmatrix} = PZ \]  
(139)

where \( Z \) is given by
\[ Z = \begin{pmatrix} Z_{1} & \cdots & Z_{p-m} \\ \cdots & \cdots & \cdots \\ 0 & \cdots & 0 \end{pmatrix} \]  
(140)

This means that (137) can be extended from \( t \) to \( (t, \beta) \) as follows:
\[ \Sigma(t, \beta) = \{(PZ)^T W (PZ)\}^{-1} \]  
(141)

Furthermore, since
\[ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ 0 \end{pmatrix} + Z \begin{pmatrix} t \\ \beta \end{pmatrix} \]  
(142)

we obtain the result
\[ \Sigma(b) = Z \{(PZ)^T W (PZ)\}^{-1} Z^T \]  
(143)
4.7.3 Distribution of \( \phi_{\text{min}} \)

With \( b_0 = (b_0) \) defined in (41) and \( b = (b_j) \) being the solution of the normal equations (28), we obtain the following approximate expression of \( \phi_{\text{min}} \) from the linear expansion (42), which is valid for small \( \Delta b = (b_j - b_{0j}) \):

\[
\phi_{\text{min}} = \sum_{i=1}^{n} w_i \left( y_i - \eta_i - \sum_{j=1}^{k} p_{ij}(b_j - b_{0j}) \right)^2
\]

(144)

This can also be written

\[
\phi_{\text{min}} = ||W^{1/2}(\Delta y - P\Delta b)||^2 = (\Delta y - P\Delta b)^T W (\Delta y - P\Delta b)
\]

(145)

with \( \Delta y = (y_i - \eta_i) \) and \( P \) given by (29–30). By (45–46) \( \phi_{\text{min}} \) becomes a quadratic form in \( \Delta y \):

\[
\phi_{\text{min}} = \Delta y^T B \Delta y
\]

(146)

where \( B \) is found to

\[
B = W - WP(P^T WP)^{-1} P^T W
\]

(147)

Defining \( u_i = \Delta y_i / s_i \), we see that \( u_i \) becomes a standardized normal variable,

\[
u_i \in N(0,1)
\]

(148)

Then \( \phi_{\text{min}} \) can be expressed as a quadratic form in \( u = (u_i) \):

\[
\phi_{\text{min}} = u^T Cu
\]

(149)

with

\[
C = W^{-1/2} B W^{-1/2} = I_n - M
\]

(150)

where

\[
M = W^2 P(P^T WP)^{-1} P^T W^2
\]

(151)

Clearly the matrix \( M \) is of rank \( k \) and all its nonzero eigenvalues are unity, as is easily verified by premultiplying \( MX = \lambda x \) by \( P^T W^{1/2} \). Hence there is an orthogonal substitution \( u = Qx \) which transforms \( \phi_{\text{min}} \) into a sum of \( f = n - k \) squares:

\[
\phi_{\text{min}} = \sum_{i=1}^{n-k} \varepsilon_i^2
\]

(152)

where the \( \varepsilon_i \) are independent, and each \( \varepsilon_i \in N(0,1) \). This means that \( \phi_{\text{min}} \) has a \( \chi^2 \)-distribution with \( f \) degrees of freedom. If there are \( m \) independent linear constraints on the parameters, then the expression (135) demonstrates that the number of degrees of freedom is altered to \( f = n - (k - m) = n - k_{\text{free}} \).

4.7.4 Proof of parameter shift formula 

We shall here give a proof of the formula (66). In the following we consider \( \phi(b) \) with fixed spectrum \( (\eta_i) \). We fix \( \Delta b_1 \) and seek the conditional minimum when the other parameters vary. We shall use the expansion formula (73):

\[
\phi(b + \Delta b) = \phi(b) + \nabla \phi(b) \cdot \Delta b + \Delta b^T P^T WP \Delta b
\]

(153)

We introduce the vector \( z \) with components \( \Delta b_1, \ldots, \Delta b_k \). The gradient term in (153) can be written

\[
\nabla \phi(b) \cdot \Delta b = \frac{\partial \phi}{\partial b_1} \Delta b_1 + \nabla_z \phi(b) \cdot z
\]

(154)

where \( \nabla_z \phi(b) \) must be zero. Making the partition

\[
P^T WP = \begin{pmatrix} a_{11} & d^T \\ d & C \end{pmatrix}
\]

(155)
(153) can then be written

\[ \phi(b + \Delta b) = \phi(b) + \frac{\partial \phi}{\partial b_1} \Delta b_1 + a_{11} \Delta b_1^2 + 2 \Delta b_1 z^T d + z^T C z \]  

(156)

To minimize (156) we take the derivative with respect to \( z \). After equating the result to zero we deduce that

\[ z = -\Delta b_1 C^{-1} d \]  

(157)

Next we make the similar partitioning

\[ \Sigma(b) = \begin{pmatrix} \sigma_{11} & s^T \\ s & \Gamma \end{pmatrix} \]  

(158)

We assume statistical weighting which implies the identity \( P^T WP \Sigma(b) = I_k \), cf. (53). From this we infer that

\[ C^{-1} d = -\frac{1}{\sigma_{11}} s \]  

(159)

Inserting this in (157) yields

\[ z = \frac{\Delta b_1}{\sigma_{11}} s \]  

(160)

which proves formula (66) since \( s \) has components \( \sigma_{21}, \ldots, \sigma_{k1} \).

### 4.7.5 Scaling in separable NLLS

In order to facilitate the application of standard minimization software we found it appropriate to make a scaling of the problem formulation. We recast the original minimization problem (25) to

\[ \phi_{\text{min}} = \min_{b \in \mathbb{R}^n} \{ ||r(b)||^2 \} \]  

(161)

where \( r(b) \) is a (scaled) residual vector with components

\[ r_i = w_i^{1/2} (y_i - f_i(b)) \]  

(162)

This induces a number of vector and matrix transformations containing the matrix scaling factor \( W^{1/2} \):

\[ z = W^{1/2} y \]  

(163)

\[ e = W^{1/2} f \]  

(164)

\[ E = W^{1/2} F \]  

(165)

\[ G = W^{1/2} P \]  

(166)

Then the counterparts of (91–95) become:

\[ e(b) = E(\beta) \alpha \]  

(167)

\[ \min_{\alpha \in \mathbb{R}^p} \{ ||z - E(\beta) \alpha||^2 : \beta \text{ fixed} \} \]  

(168)

\[ E^T E \alpha = E^T z \]  

(169)

\[ \min_{\beta \in \mathbb{R}^q} \{ \phi(\beta) \equiv ||z - E(\beta) \alpha(\beta)||^2 \} \]  

(170)

\[ (G^T G + \lambda D) d = G^T (z - E \alpha) \]  

(171)

Moreover (97–98) are replaced by:

\[ \frac{\partial e}{\partial \beta_j} = \frac{\partial E}{\partial \beta_j} \alpha + E \frac{\partial \alpha}{\partial \beta_j} \]  

(172)

\[ E^T E \frac{\partial \alpha}{\partial \beta_j} = \frac{\partial E^T}{\partial \beta_j} (z - E \alpha) - E^T \frac{\partial E}{\partial \beta_j} \alpha \]  

(173)

Riso–R–1652-update
We see that the effect of these transformations is to “hide” the weights \( w_i \) entirely. To include the effect of the linear constraints on (172–173) we use (127) and find:

\[
\frac{\partial e}{\partial \beta_j} = \frac{\partial E}{\partial \beta_j} + EY \frac{\partial t}{\partial \beta_j} 
\]

\[
(EY)^T EY \frac{\partial t}{\partial \beta_j} = \frac{\partial (EY)^T}{\partial \beta_j} (z - E\alpha) - (EY)^T \frac{\partial E}{\partial \beta_j} \alpha
\]

(174)

(175)

4.7.6 QR decomposition

A direct solution of normal equations, even by Choleski decomposition, may present numerical difficulties inherent with the ill-conditioning of the positive-definite coefficient matrix, say \( E^T E \) in (169). Instead we use a procedure based on the so-called QR decomposition of \( E \), viz.

\[
E = QR
\]

(176)

where \( Q \) is an \( n \times p \) matrix with orthonormal columns and \( R \) is a \( p \times p \) upper triangular matrix (see, e.g., Chapter 9 in [42]). Using (176) the system (169) is reformulated to \( R\alpha = Q^T z \), which can be easily solved by back-substitution. The same procedure is used when solving (173) for \( \partial \alpha / \partial \beta_j \), with \( R \) being saved after the solution of (169).

The numerical computation of the covariance matrix \( \Sigma(b) \) can also be done by QR technique. We start with the expression (143) which includes the constraints and performs the usual scaling by defining \( T = W^{1/2} P \). Then we obtain

\[
\Sigma(b) = Z((TZ)^T TZ)^{-1} Z^T
\]

(177)

and can now make the decomposition (of course the factors are different from those in (176)),

\[
TZ = QR
\]

(178)

which leads to

\[
\Sigma(b) = ZR^{-1}(ZR^{-1})^T
\]

(179)

In some ill-conditioned problems the diagonal row of \( R \) may contain very small elements, which would render the evaluation of \( \Sigma(b) \) by (179) completely erratic. There exists a variant of the QR decomposition with column scaling and pivoting that admits a judicious discarding rule for insignificant elements in the \( R \)-diagonal \([42, 35]\). Following this idea, we shall replace (178) with

\[
TZ\Pi = QR
\]

(180)

where \( A \) is a diagonal scaling matrix, \( \Pi \) a permutation matrix, and the diagonal elements of \( R \) are in non-increasing order of magnitude. The entries in \( A \) are chosen as the inverse Euclidean norms of the column vectors of \( TZ \) and might be called “uncoupled standard deviations”. Instead of (179) we obtain

\[
\Sigma(b) = ZA\Pi R^{-1}(ZA\Pi R^{-1})^T
\]

(181)

The expression (181) is only used for the “significant” parameters which corresponds to the upper part of \( R \). The variance of the “insignificant” parameters are estimated by their uncoupled standard deviations, while the covariance calculation for such parameters are abandoned. (It is easier to grasp the essential features of this procedure if we simplify and replace \( TZ \) with \( P = QR \).)
5 Mathematical model details

In Sections 2.4 and 2.6 we gave a short presentation of the theoretical models used in PositronFit and ResolutionFit. Below we shall try to fill the gap between the rather brief description given there of the underlying mathematical models, and the least-squares theory in Chapter 4.

5.1 PositronFit

Writing formula (3) as

\[ f(t) = \sum_{j=1}^{k_n} \sum_{p=1}^{k_p} \omega_p (a_j * G_p)(t) + B \]  \hspace{1cm} (182)

we must evaluate the convolution integral

\[ (a_j * G_p)(t) = \int_{-\infty}^{\infty} a_j(v) G_p(t-v) dv \]  \hspace{1cm} (183)

where \( a_j \) and \( G_p \) were defined in (4) and (7), respectively. Henceforward, we prefer to describe the decay of a lifetime component in terms of the annihilation rate

\[ \lambda_j = 1/\tau_j \]  \hspace{1cm} (184)

instead of the lifetime \( \tau_j \) itself. It can be shown that

\[ (a_j * G_p)(t) = \frac{1}{2} A_j \phi(t - T_0 - \Delta t_p, \lambda_j, \sigma_p) \]  \hspace{1cm} (185)

Here the function \( \phi \) acts as a building block for our lifetime spectral model and is defined by

\[ \phi(u, \lambda, \sigma) = \exp \left( -\lambda u + \frac{1}{2} \lambda^2 \sigma^2 \right) \erfc \left( \frac{\lambda \sigma^2 - u}{\sqrt{2}\sigma} \right) \]  \hspace{1cm} (186)

where \( \erfc \) stands for the complementary error function

\[ \erfc(x) = 1 - \erf(x) \]  \hspace{1cm} (187)

and \( \erf \) in turn is defined by

\[ \erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \]  \hspace{1cm} (188)

Regarding the practical computation of (186), that formula may from a numerical point of view be dangerous to use as it stands. The difficulty arises when \( u \ll 0 \). Then \( \phi(u, \lambda, \sigma) \) itself is small; nevertheless, the first factor of (186) is large and may cause an overflow on the computer. At the same time, the second factor is very small and likely to underflow. A better alternative is to restate (186) as

\[ \phi(u, \lambda, \sigma) = \exp \left( -\frac{u^2}{2\sigma^2} \right) \erfc \left( \frac{\lambda \sigma^2 - u}{\sqrt{2}\sigma} \right) \]  \hspace{1cm} (189)

where \( \erfc \) stands for the normalized complementary error function

\[ \erfc(x) = \exp(x^2) \erf(x) \]  \hspace{1cm} (190)

It is not hard to develop robust and accurate numerical approximations for this slowly varying function, which is decreasing when \( x > 0 \) and behaves asymptotically as \( 1/(\sqrt{\pi} x) \). When \( x \) is negative we use the formula

\[ \erfc(-x) = 2 \exp(x^2) - \erfc(x) \]  \hspace{1cm} (191)
Inserting (185) in (182) we get

$$f(t) = \frac{1}{2} \sum_{j=1}^{k_0} A_j \sum_{p=1}^{k_0} \omega_p \phi(t - T_0 - \Delta_p, \lambda_j, \sigma_p) + B \quad \text{(192)}$$

Finally, we compute the integrated model-predicted count \( f_i \) defined by equation (11) in Section 2.4. We use the identity

$$\int \phi(u, \lambda, \sigma) \, du = -\frac{1}{\lambda} \left( \phi(u, \lambda, \sigma) + \text{erfc} \left( \frac{u}{\sqrt{2\sigma}} \right) \right) \quad \text{(193)}$$

and obtain

$$f_i = \sum_{j=1}^{k_0} F_i j + B = \sum_{j=1}^{k_0} \alpha_j f_i j + B \quad \text{(194)}$$

where

$$\alpha_j = \frac{1}{2} A_j / \lambda_j = \frac{1}{2} A_j \tau_j \quad \text{(195)}$$

is half the absolute intensity of lifetime component \( j \),

$$f_i j = \sum_{p=1}^{k_0} \omega_p \left\{ \phi(t_{i-1,j}, \lambda_j, \sigma_p) - \phi(t_{ip}, \lambda_j, \sigma_p) + \text{erfc} \left( \frac{t_{i-1,j} - t_{ip}}{\sqrt{2\sigma_p}} \right) - \text{erfc} \left( \frac{t_{ip}}{\sqrt{2\sigma_p}} \right) \right\} \quad \text{(196)}$$

and where we use the shorthand notation

$$t_{ip} = c_{\text{ch}}^{\text{min}} + i - 1 - T_0 - \Delta_p \quad \text{(197)}$$

By now we have arrived at the model expression \( f_i = f_i (b) \) entering the least-squares formulation of the fitting problem given in Chapter 4. We also see that (194) is separable as required; the parameter vector \( b \) splits into a “linear” parameter \( \alpha \) and a “nonlinear” one \( \beta \) given by

$$\alpha = (\alpha_1, \ldots, \alpha_{k_0}, B) \quad \text{(198)}$$

and

$$\beta = (\lambda_1, \ldots, \lambda_{k_0}, T_0) \quad \text{(199)}$$

Thus the separable fitting theory of Section 4.6 applies. To perform the computations outlined there, we must evaluate the derivatives of \( f_i j \) in (196) with respect to \( \lambda_j \) and \( T_0 \); this job is facilitated by the following two formulas:

$$\frac{\partial \phi}{\partial u} = -\lambda \phi(u, \lambda, \sigma) + \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \exp \left( -\frac{u^2}{2\sigma^2} \right) \quad \text{(200)}$$

and

$$\frac{\partial \phi}{\partial \lambda} = (\lambda \sigma^2 - u) \phi(u, \lambda, \sigma) + \sqrt{\frac{2}{\pi}} \sigma \exp \left( -\frac{u^2}{2\sigma^2} \right) \quad \text{(201)}$$

We introduce the abbreviations

$$\phi_{i,jp} = \phi(t_{ip}, \lambda_j, \sigma_p) \quad \text{(202)}$$

$$\delta \phi_{i,jp} = \phi_{i,jp} - \phi_{i-1,jp} \quad \text{(203)}$$

$$\sigma_{i,p}^\text{exp} = \exp \left( -\frac{t_{ip}^2}{2\sigma_p^2} \right) - \exp \left( -\frac{t_{i-1,p}^2}{2\sigma_p^2} \right) \quad \text{(204)}$$

and then obtain from (200–201)

$$\frac{\partial f_{i,j}}{\partial T_0} = -\lambda_j \sum_{p=1}^{k_0} \omega_p \delta \phi_{i,jp} \quad \text{(205)}$$

$$\frac{\partial f_{i,j}}{\partial \lambda_j} = \sum_{p=1}^{k_0} \omega_p \left\{ (\lambda_j \sigma_p^2 - t_{ip}) \delta \phi_{i,jp} + \sigma_p \sqrt{2/\pi} \sigma_{i,p}^\text{exp} \right\} \quad \text{(206)}$$
In Section 2.4 we mentioned the types of constraints which could be imposed on the parameters in *PositronFit*. Those constraints that fix one of the “primary” fitting parameters listed in (198) and (199) are realized by deleting the corresponding components from $\alpha$ or $\beta$. This may apply to $B$, $\lambda_j$, and $T_0$. Constraints of the type “fixed relative intensity” are not of this simple type because the relative intensities $\alpha_j / \sum \alpha_j$ are not primary parameters. But obviously such constraints are expressible as linear constraints on the linear parameters $\alpha_j$, i.e. relations of the form
\[
\sum h_{ij} \alpha_j = \gamma_i
\]
where $h_{ij}$ are known coefficients, cf. (38). The same holds good for constraints of the type “a linear combination of the relative intensities = 0”, as well as the total-area constraint (18).

### 5.2 ResolutionFit

Although the basic model in *ResolutionFit* is the same as in *PositronFit*, there are certain differences regarding which parameters enter as fitting parameters since the widths $\sigma_p$ and the shifts $\Delta t_p$ are fitting parameters in *ResolutionFit*. Hence (199) should be replaced by
\[
\beta = (\lambda_1, \ldots, \lambda_k, T_0, \sigma_1, \ldots, \sigma_k, \Delta t_1, \ldots, \Delta t_k)
\]
In Section 2.6 we mentioned the types of constraints which could be imposed on the parameters in *ResolutionFit*. Some of the parameters of (208) can be fixed and in that case should be deleted from $\beta$. This may apply to $\lambda_j$, $\sigma_p$, and $\Delta t_p$. Notice that $T_0$ in *ResolutionFit* is always a free parameter. As a consequence, we must require that at least one of the shifts $\Delta t_p$ be fixed.

In addition to (200–206) we shall need the following formulas:
\[
\frac{\partial \phi}{\partial \sigma} = \lambda^2 \sigma \phi(u, \lambda, \sigma) - \sqrt{\frac{n}{\pi}} \left( \lambda + \frac{n}{\sigma^2} \right) \exp \left( - \frac{\sigma^2}{2} \right)
\]
\[
\frac{\partial f_{ij}}{\partial \Delta t_p} = -\lambda_j \omega_p \delta \phi_{ijp}
\]
\[
\frac{\partial f_{ij}}{\partial \sigma_p} = \omega_p \{ -\lambda_j^2 \sigma_p \delta \phi_{ijp} + \lambda_j \sqrt{2/\pi} \sigma_p \}
\]
In *ResolutionFit* we compute shape parameters for the fitted resolution curve. This leads to nonlinear equations involving the $\phi$ function of (186). We use a Newton–Raphson procedure for the numerical solution of these.

### 6 Control files

When running PALSfit the program produces a control file, which contains all the input instructions for either a *PositronFit* or a *ResolutionFit* analysis. Normally the editing of this file is a task which is done automatically by using the PALSfit menus. Nevertheless there might be situations where an inspection or an external editing of the file is appropriate.

In Chapter 1 it was mentioned that in certain situations (batch processing or running under Linux) it may be useful to run the command-driven PATFIT programs *PositronFit* and *ResolutionFit* directly. In that case you will also need to know the structure of the input files [2]. Note that PATFIT and PALSfit are input-compatible.

Anyway, the knowledge of the structure of the control files may give the user a good overview of the capabilities of PALSfit. Therefore, in the following we shall describe the contents of
the control files for PositronFit and ResolutionFit in some detail. For convenience, we will in some cases use parameter names that occur in the programs.

The control file is composed of one or more control data sets. A control data set is partitioned into a number of data blocks, corresponding roughly to the menus in PALSfit. Each block is initiated by a so-called block header. For example, the first block header reads

```
POSITRONFIT DATA BLOCK: OUTPUT OPTIONS
```

in the case of PositronFit, and similarly for ResolutionFit.

### 6.1 PositronFit control file

A sample PALSfit control file for PositronFit with a single control data set is shown below.

```
POSITRONFIT DATA BLOCK: OUTPUT OPTIONS
0000
POSITRONFIT DATA BLOCK: SPECTRUM
512
(1017)
M1,Crystal spectra.dat
32800 CYCLOCTANE
1 0
32800 CYCLOOCTANE
```

```
32800 764 776 726 754 765 733 725 743 769
700 735 694 737 714 766 756 662 741 728
706 715 729 716 717 723 699 717 719 702
699 729 637 710 678 745 702 708 686 708
708 670 677 712 679 664 642 708 718 720
677 888 722 704 706 701 681 675 703 720
696 710 723 683 628 746 725 647 730 703
677 888 691 700 730 676 658 708 714 690
719 706 707 708 673 701 684 689 692 697
743 688 721 646 712 716 665 707 730 690
706 743 713 682 756 720 683 712 679 700
692 690 692 708 742 740 742 690 708 727
770 704 766 683 743 622 1049 1810 4146 11412
30066 71498 151210 277774 333259 561716 677357 706284 669401 596116
509460 426362 394314 2964562 2964562 206141 176625 151785 131162 114887
100962 90149 80988 73201 66518 60971 56076 52147 48976 45968
43147 40678 39126 36622 35059 33794 32103 30989 30125 29086
27905 26885 25805 24090 24441 23896 22792 22072 21308 21178
20044 19016 18569 18302 17973 17247 16784 16432 15678 15607
15241 14667 14085 13665 13469 12871 12622 12277 11916 11533
11407 11174 10807 10156 9994 9855 9690 9375 9142 8789
8694 8424 8249 7826 7537 7623 7368 7025 6916 6536
6748 6475 6307 6079 6018 5730 5628 5619 5577 5274
5190 5019 4929 4747 4660 4559 4541 4336 4194 4063
4007 3885 3855 3751 3701 3668 3545 3390 3263 3262
3218 3061 3007 2932 2786 2816 2759 2738 2722 2611
2631 2524 2661 2533 2463 2437 2380 2220 2214 2157
1774 1724 1747 1639 1728 1572 1601 1681 1679 1566
1612 1501 1466 1497 1480 1369 1404 1406 1336 1397
1337 1299 1268 1240 1281 1240 1203 1195 1217 1159
1117 1163 1079 1049 1077 1090 1038 1086 1020 1051
974 968 1039 996 1001 1018 967 967 964 927
975 963 885 942 921 915 911 935 859 851
880 873 945 863 944 827 852 806 822 838
830 802 826 825 810 774 772 784 796 801
803 799 815 790 796 778 763 752 733 760
719 744 765 796 729 717 719 688 730
744 742 672 733 756 719 689 749 686 754
696 693 716 630 716 705 666 743 753 734
710 715 737 705 670 712 650 758 705 764
778 746 688 728 662 716 707 697 733 660
667 768 704 680 644 730 692 691 694 675
```

38  
Riso-R-1652-update
Block 1 contains output options. Apart from the block header there is only one record. It contains four integer keys. Each key is either 0 or 1. The value 1 causes some output action to be taken, whereas 0 omits the action. The actions of the 4 keys are:

1. Write input echo to result file
2. Write each iteration output to result file
3. Write residual plot to result file
4. Write correlation matrix to result file

Regardless of the setting of these keys, PositionFit always produces the Main Output.

Block 2 contains the spectrum. The first record (after the block header) contains the integer NCH, which is the total number of channels in the spectrum. Next record contains a description of precisely how the spectrum values are “formatted” in the file—expressed as a so-called FORMAT in the programming language FORTRAN [44]; an asterisk * means free format. After this, two text records follow. In the first a name of a spectrum file is given. (Even when INSPEC = 1 (see below) this name should be present, but is in that case not
used by the program. In the other record an identification label of the spectrum is given.
The next record contains two integers, INSPEC and XXX, both taking a value of either
0 or 1. (XXX is specific for the PALS/fit/GUI and is not used by PATFIT.) INSPEC = 1
means that the spectrum is an intrinsic part of the present control file. In this case the next
record should be a text line with a description of the spectrum. The subsequent records are
supposed to hold the NCH spectrum values. On the other hand, INSPEC = 0 means that the
spectrum is expected to reside in an external file with the spectrum file name entered
above. The program tries to open this file (which may contain several spectra) and scans
it for a record whose start matches the identification label. After a successful match, the
matching (text) line and the spectrum itself is read from the subsequent records in exactly
the same way as in the case INSPEC = 1. If XXX = 1 all spectra in the spectrum file
will be analysed with the parameters defined in the present data set. If XXX = 0, only the
spectrum defined in the present data set will be analysed.

Block 3 contains information related to the measuring system. The first two records (after
the block header) contain two channel numbers ICHA1 and ICHA2. These numbers are
lower and upper bounds for the definition of a total area range. The next two records
contain also two channel numbers ICHMIN and ICHMAX. These define in the same way
the channel range which is used in the least-squares analysis. The next record contains the
channel width (in ns). The last two records in this block deal with $T_0$ (time=0 channel
number). First comes a constraint flag being either a G or an F. G stands for guessed (i.e.
free) $T_0$. F stands for fixed $T_0$. The other record contains the initial (guessed or fixed) value
of $T_0$.

Block 4 contains input for definition of the resolution function. The first record (after the
block header) contains the number $k_g$ of Gaussian components in the resolution function.
Each of the next three records contains $k_g$ numbers. In the first we have the full widths at
half maxima of the Gaussians (in ns), FWHM$_{j\gamma}$, $j = 1, \ldots, k_g$, in the second their relative
intensities (in percent) $\omega_j$, $j = 1, \ldots, k_g$, and in the third their peak displacements (in ns)
$\Delta t_j$, $j = 1, \ldots, k_g$.

Block 5 contains data for the lifetime components in the lifetime spectrum as well as
constraints on their relative intensities. The first record (after the block header) holds the
number $k_0$ of lifetime components assumed in the model. Each of the next two records
contains $k_0$ data. In the first we have the constraint flags (G = guessed, F = fixed) for the
lifetimes. The other record contains the initial values (guessed or fixed) of the $k_0$
lifetimes. After this comes a record with an integer $m$ telling the number and type of intensity
constraints, $|m|$ is equal to the number of constraints, but $m$ itself may be positive, negative,
or zero. If $m = 0$ there is no further input data in this block. If $m > 0$, $m$ of the relative
intensities are fixed. In this case the next data item is a pair of records with the numbers
$j_l$, $l = 1, \ldots, m$ and $I_{j_l}$, $l = 1, \ldots, m$; here $j_l$ is the term number (the succession agreeing
with the lifetimes on the previous record) associated with constraint number $l$, and $I_{j_l}$ is
the corresponding fixed relative intensity (in percent). If $m < 0$, $|m|$ linear combinations
of the intensities are equal to zero. In this case $|m|$ records follow, each containing the
$k_0$ coefficients $h_{lj}$, $j = 1, \ldots, k_0$ to the intensities for one of the linear combinations, cf.
equation (17) in Section 2.4.

Block 6 contains data related to the background. The first record (after the block header)
contains an integer indicator KB, assuming one of the values 0, 1, or 2. KB = 0 means a
free background; in this case no more data follows in this block. If KB = 1 the background
is fixed to the spectrum average from channel ICHBG1 to channel ICHBG2. These two
channel numbers follow on the next two records. If KB = 2, the background is fixed to an
input value which is entered on the next record.

Block 7 contains input for constraining the total area. The first record (after the block
header) holds an integer indicator KAR, assuming one of the values 0, 1, or 2. KAR =
0 means no area constraint; in this case no more data follows in this block. If KAR > 0, 
the area between two specified channel limits ICHBEG and ICHEND will be fixed, and 
these channel numbers follow on the next two records. If KAR = 1, the area is fixed to the 
measured spectrum, and no more input will be needed. If KAR = 2 the area is fixed to an 
input value which is entered on the next record.

Block 8 contains source correction data. The first record (after the block header) contains 
an integer k, denoting the number of components in the source correction spectrum. k = 0 
means no source correction, in which case the present block contains no more data. The 
next record contains the lifetimes $\tau_j^i, j = 1, \ldots, k$, and the following the relative intensities 
$I_j^i, j = 1, \ldots, k$, for the source correction terms. On the next record is the number $\alpha$ which 
is the percentage of positrons that annihilate in the source, cf. equation (20) in Section 2.4. 
Then there follows a record with an integer ISEC. When ISEC = 0 the new iteration cycle 
after the source correction starts from lifetime guesses equal to the converged values from 
the first (correction-free) cycle. ISEC = 1 tells that the second cycle starts from new input 
data. These 2nd-cycle input data are now entered in exactly the same way as the 1st- 
cycle data in Block 5. ISEC = 2 works as ISEC = 1, but with the additional possibility 
of changing the status of $T_0$; in this case two more records follow, the first containing the 
constraint flag (G = guessed, F = fixed) for $T_0$ and the second the value of $T_0$.

With the end of Block 8 the entire PositronFit control data set is completed. However, as 
previously mentioned, PALSfit accepts multiple control data sets in the same PositronFit 
control file.

### 6.2 ResolutionFit control file

A sample PALSfit control file for ResolutionFit with a single control data set is shown 
below.

RESOLUTIONFIT DATA BLOCK: OUTPUT OPTIONS
0000
RESOLUTIONFIT DATA BLOCK: SPECTRUM
1023
./metal_defects.spectra.dat
39699 CU-ANNEALED
0 0
RESOLUTIONFIT DATA BLOCK: CHANNEL RANGES, TIME SCALE, TIME-ZERO.
5
1000
140
500
0.013400
172.000
RESOLUTIONFIT DATA BLOCK: RESOLUTION FUNCTION
3
GGG
0.2600 0.3000 0.4000
77.000 19.000 4.000
PFG
0.0000 0.0228 -0.0462
RESOLUTIONFIT DATA BLOCK: LIFETIMES AND INTENSITY CONSTRAINTS
3
PFG
0.1100 0.1800 0.4000
0
RESOLUTIONFIT DATA BLOCK: BACKGROUND CONSTRAINTS
1
650
1000

Block 1 contains output options. It is identical to the corresponding block in the PositronFit...
control file (but of course the name RESOLUTIONFIT must appear in the block header).

Block 2 contains the spectrum. It is identical to the corresponding block in the PositronFit control file.

Block 3 contains information related to the measuring system. The first two records (after the block header) contain two channel numbers ICHA1 and ICHA2. These numbers are lower and upper bounds for the definition of a total area range. The next two records contain also two channel numbers ICHMIN and ICHMAX. These define in the same way the channel range which is used in the least-squares analysis. The next record contains the channel width (in ns). The last record in this block contains the initial (guessed) value of \( T_0 \).

Block 4 contains input for definition and initialization of the resolution function. The first record (after the block header) contains the number \( k_g \) of Gaussian components in the resolution function. Each of the next two records contains \( k_g \) data. In the first we have the constraint flags (G=guessed, F=fixed) for the Gaussian widths. The second contains the initial values (guessed or fixed) of the full widths at half maxima of the Gaussians (in ns), \( \text{FWHM}_j^{\text{int}}, j = 1, \ldots, k_g \). The next record contains the \( k_g \) Gaussian component intensities in percent, \( \omega_j, j = 1, \ldots, k_g \). The last two records in the block contain again \( k_g \) data each. First, we have the constraint flags (G=guessed, F=fixed) for the Gaussian shifts; notice that not all the shifts can be free. Next, we have the initial (guessed or fixed) peak displacements (in ns), \( \Delta_j^{\text{int}}, j = 1, \ldots, k_g \).

Block 5 contains data for the lifetime components in the lifetime spectrum as well as constraints on their relative intensities. It is identical to the corresponding block in the PositronFit control file.

Block 6 contains data related to the background. It is identical to the corresponding block in the PositronFit control file.

This completes the ResolutionFit control data set. Multiple data control data sets can be handled in the same way as for PositronFit.

### 6.3 Channel ranges in PALSfit

As we saw in the description of the control files for PositronFit (P) and ResolutionFit (R) in Sections 6.1 and 6.2, the data blocks contain a number of integers defining various kinds of channel ranges. Each range is an interval \([M, N]\) and is thus equal to the set of integers \( i \) satisfying \( M \leq i \leq N \). Using the same program acronyms for the channel limits as before, we can collect all the channel ranges in the following table:

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Program use</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total range</td>
<td>[1,NCH]</td>
<td>PR</td>
<td>( T )</td>
</tr>
<tr>
<td>Fit range</td>
<td>ICHMIN, ICHMAX</td>
<td>PR</td>
<td>( F )</td>
</tr>
<tr>
<td>Area range</td>
<td>ICHA1, ICHA2</td>
<td>PR</td>
<td>( A )</td>
</tr>
<tr>
<td>Background range</td>
<td>ICHBG1, ICHBG2</td>
<td>PR</td>
<td>( B )</td>
</tr>
<tr>
<td>Fixed-area range</td>
<td>ICHBEG, ICHEND</td>
<td>P</td>
<td>( A_f )</td>
</tr>
</tbody>
</table>

The first three ranges are always present while the existence of the two latter depends on optional constraints. There are certain restrictions on the ranges \( T, F, A, B, A_f \) (when present). All must be nonempty subsets of \( T \), and all the remaining must be subsets of \( A \). Thus in formal terms

\[
\emptyset \subset A \subseteq T \quad (212)
\]

\[
\emptyset \subset F \subseteq A \quad (213)
\]

42 Riso-R-1652-update
\[ \emptyset \subseteq B \subseteq A \quad (214) \]
\[ \emptyset \subseteq A_f \subseteq A \quad (215) \]

These restrictions are exhaustive. If a restriction is violated, PALSfit makes a suitable modification of the range in order to amend the problem and tries to continue. On the other hand, PATFIT refuses to make an analysis in such a case.

## 7 Simulation of lifetime spectra

In the following we shall very briefly describe the auxiliary programs POSGEN and PALSfit. Both use Monte Carlo simulation to generate spectra from given lifetime parameters. They admit a possibility of checking the PALSfit software, since we may subsequently ask the latter to try to retrieve the original parameters. POSGEN takes the formula apparatus in Section 5.1 as granted. Thus this program is primarily meant as an internal test and development tool.

In 1997 Hirade\(^1\) developed a simulation program PALGEN, written in the BASIC language. PALGEN uses Monte Carlo in its simplest version, the so-called direct simulation. This is close to the real-world physical setup and admits an independent assessment of the capability of PALSfit to recover correct lifetime values. Cheung et al. [45], also in 1997, describe a simulation tool which is equivalent to PALGEN; they used their program to study the merits of the PositionFit software as an analysis tool. At Riso DTU we have later created a modern FORTRAN-based version of PALGEN. Its functionality is incorporated in PALSfit.

The main sampling principle in PALGEN is to split the recorded annihilation time \( t \) as follows:

\[ t = T_0 + t_{\text{exp}} + t_{\text{gauss}} + t_{\text{shift}} \quad (216) \]

Here \( T_0 \) is the given time-zero, \( t_{\text{exp}} \) the true annihilation time, \( t_{\text{gauss}} \) the smear component, and \( t_{\text{shift}} \) a deterministic shift value associated with the latter. We begin with the sampling of \( t_{\text{exp}} \). First a random number determines the lifetime component from the given intensities. Then, with the lifetime \( \tau \), we have the exponential distribution function

\[ F(x) = 1 - e^{-x/\tau} \quad (217) \]

To sample \( x = t_{\text{exp}} \) we use the classical Monte Carlo formula,

\[ x = F^{-1}(\xi) \quad (218) \]

which is the inverse form of

\[ F(x) = \xi \quad (219) \]

with \( \xi \in (0, 1) \) being a uniform random number. Then, by replacing \( \xi \) with \( 1 - \xi \) we obtain the sampling formula

\[ t_{\text{exp}} = -\tau \ln \xi \quad (220) \]

Next we perform the sampling of \( t_{\text{gauss}} \). First we pick one of the gaussian components. Let the standard deviation of this be \( \sigma \). Then we use the well-known polar method of Box and Muller [46] to generate a normal variate \( \eta \) with variance 1, and obtain

\[ t_{\text{gauss}} = \eta \sigma \quad (221) \]

Finally the sampling of the background proceeds independently of the remaining spectrum and is accomplished by simple uniform multinomial sampling with bins = channels.

\(^1\)Personal communication, Tetsuya Hirade, Department of Materials Science, Japan Atomic Energy Research Institute.
Apart from statistical fluctuations the two programs POSGEN and PALGEN are found to be equivalent to each other. This provides (yet another) confirmation of the consistence of the PALSfit model.

References


