REFINE

Users Guide

Version 5.99

written by

Reinhard Neder
Email: reinhard.neder@fau.de

http://tproffen.github.io/DiffuseCode

Document created: November 27, 2019

Preface

Disclaimer

The REFINE software described in this guide is provided without warranty of any kind. No liability is taken for any loss or damages, direct or indirect, that may result through the use of *DIFFEV*. No warranty is made with respect to this manual, or the program and functions therein. There are no warranties that the programs are free of error, or that they are consistent with any standard, or that they will meet the requirement for a particular application. The programs and the manual have been thoroughly checked. Nevertheless, it can not be guaranteed that the manual is correct and up to date in every detail. This manual and the REFINE program may be changed without notice.
REFINE is intended as a public domain program. It may be used free of charge. Any commercial use is, however, not allowed without permission of the authors.

Using REFINE

More information

This users guide can only provide program specific details. A broader discussion of simulation techniques and some REFINE examples and macro files can be found in our book

> NEDER, R.B. & PROFFEN, TH. "Diffuse Scattering and Defect Structure Simulations - A cook book using the programs DISCUS", *IUCr Texts on Crystallography*, Oxford University Press, 2007.

Contents

Chapter 1

Introduction

1.1 What is REFINE ?

REFINE is the direct Least Squares Refinement section of the DISCUS SUITE program. Its internal engine is a Levenberg-Marquardt type least squares fit that can be applied to 1D or 2D data.
See the DIFFEV section on a more general generic evolutionary refinement program that implements the differential evolutionary algorithm Price et al. (2005). Evolutionary or genetic refinement algorithms allow the refinement of models, functions, or more generally speaking the parameters of a cost function to obtain a good solution.
A least squares based refinement of a function $y = F(p_0, p_1, ..., p_n)$ requires the calculation of all partial derivatives $\partial y/\partial p_i$, either from an analytical or a numeric solution. The refinement that is described in this section is intended to work with disordered crystal structures as build by DISCUS . To achieve this REFINE uses a macro that is provided by the user to build a crystal and to calculate a diffraction pattern or a PDF. As the details of a calculation are hidden within the source code of the DISCUS and KUPLOT section, REFINE generally relies on a numerical calculation of the derivative.
Since this macro could calculate any cost function, REFINE is not limited to the refinement of a particular physical problem.

1.2 Getting started

After the program *DISCUS_SUITE* is installed properly and the environment variables are set, the program can be started by typing 'discus_suite' at the operating systems prompt.
The section uses the identical command language to interact with the user as is used throughout the DISCUS SUITE. The command `exit` terminates the section and returns control to the top level of the DISCUS SUITE. All commands of REFINE consist of a command verb, optionally followed by one or more parameters. All parameters must be separated from one another by a comma ",". There is no predefined need for any specific sequence of commands. REFINE is case sensitive, all commands and alphabetic parameters MUST be typed in lower case letters. If REFINE has been compiled using the `-DREADLINE` option (see installation files) basic line editing and recall of commands is possible. For more information refer to the reference man-

Symbol	Description
"text"	Text given in double quotes is to be understood as typed.
<text>	Text given in angled brackets is to be replaced by an appropriate value, if the corresponding line is used in DIFFEV. It could, for example be the actual name of a file, or a numerical value.
`text`	Text in single quotes exclusively refers to REFINE commands.
[text]	Text in square brackets describes an optional parameter or command. If omitted, a default value is used, else the complete text given in the square brackets is to be typed.
{text | text}	Text given in curly brackets is a list of alternative parameters. A vertical line separates two alternative, mutually exclusive parameters.

Table 1.1: Used symbols

Variable	Description
F_DATA	Kuplot data set that holds the experimental data.
F_SIGMA	Kuplot data set that holds the experimental uncertainties.
F_XMIN	Minimum 'x'-value of the experimental data set.
F_XMAX	Maximum 'x'-value of the experimental data set.
F_XSTP	Interval size along the 'x-axis of the experimental data set.
F_YMIN	Minimum 'y'-value of the experimental data set. xyz data only
F_YMAX	Maximum 'y'-value of the experimental data set. xyz data only
F_YSTP	Interval size along the 'y-axis of the experimental data set. xyz data only

Table 1.2: REFINE variables. Variables marked with * are read-only and cannot be altered.

ual or check the online help using (`help command input`). Names of input or output files are to be typed as they will be expected by the shell. If necessary include a path to the file. All commands may be abbreviated to the shortest unique possibility. At least a single space is needed between the command verb and the first parameter. No comma is to precede the first parameter. A line can be marked as comment by inserting a "#" as first character in the line.
The symbols used throughout this manual to describe commands, command parameters, or explicit text used by the program REFINE are listed in Table 1.1. There are several sources of information, first REFINE has a build in online help, which can be accessed by entering the command `help` or if help for a particular command <cmd> is wanted by `help <cmd>`. This manual describes background and principle functions of REFINE and should give some insight in the ways to use this program.
REFINE is distributed as part of the diffuse scattering simulation software DISCUS. However, REFINE can be used as general refinement program separate from the main purpose of the DISCUS program package.

1.3 Command language

The program includes a FORTRAN style interpreter that allows the user to program complex modifications. A detailed discussion about the command language, which is common to all DISCUS package programs can be found in the separate DISCUS package reference guide which is included with the package.
Table 1.2 shows a summary of REFINE specific variables. All variables are read/write.

Chapter 2

Least Squares Refinement

2.1 Refinement via least squares refinement

Every time we measure some physical effect and wish to understand how this effect works, we want to determine the parameters of a model function that will replicate the observations. The term refinement refers to the process by which the parameters of the function are tuned such as to give the best agreement between the observed and calculated values. The term *best agreement* merits careful definition, for right now it is sufficient to say that the sum over all squared differences between the observations and the calculations shall be minimized. Thus, refinement is but a special case of general optimization. A very different example for an optimization could be the task to place as many integrated circuits into a chip and simultaneously achieve the fastest computations. Quite well known is the traveling salesman problem. Here the optimization task requires to find the shortest route that visits a number of spots distributed in space.
By far the fastest refinement technique is a least squares algorithm. Such an algorithm can always be applied if we can describe the physical effect as a function of parameters:

$$y \; = \; F(p_0, p_1, ..., p_n), \tag{2.1}$$

and all the partial derivatives $\partial y / \partial p_i$ can be calculated, either analytically or numerically. For each observed value y_{obs}, we calculate a value y_{calc} and minimize the value of a weighted residual wR:

$$wR = \sqrt{\frac{\sum_i w_i (y_{obs}(i) - y_{calc}(i))^2}{\sum_i w_i y_{obs}(i)^2}} \tag{2.2}$$

Here each difference is multiplied by a weight w_i that reflects the uncertainties of the experimental values. In case of crystal structure analysis, the observed values would be the observed intensities in a diffraction pattern and the calculated values those intensities that were calculated based on a structural model. Model parameters will be the lattice parameters, the positions of the atoms in the unit cell, atomic displacement parameters etc. as well as experimental parameters, such as the background. Under the assumption that we have a periodic crystal, the partial derivatives of the intensity with respect to lattice parameters, atom positions etc., can all be derived analytically. This is the concept you will find within any single crystal structure refinement program or a Rietveld program.

For disordered structures, the situation becomes more complicated. Except for a few special cases like stacking faults or short-range order problems, no general analytical function straightforwardly links the disorder parameter to the intensity. The intensity can still be calculated from structural models. The simulation, however, usually involves the application of random choices to generate part or all of the atom positions, and the analytical derivative of the intensity with respect to the order parameter is no longer available. A numeric calculation of the derivatives involves the repeated simulation of a new model for each parameter and is fairly time consuming.

2.2 Algorithm in the REFINE section

The REFINE section uses a Least-Squares algorithm based on the software in the Numerical Recipes. This software is based on the Levenberg-Marquart algorithm. In any least-squares algorithm the derivatives are used to determine a new estimate for each parameter. While the refinement is far from the final solution, each parameter can be modified by a large step in order to quickly approach the global minimum. Close to the minimum the steps need to be smaller not to miss the minimum. Essentially the step size is roughly propertional to the derivative. The Levenberg-Marquardt algorithm, optimizes the steps to be taken.
Within the REFINE section of the DISCUS SUITE the derivatives are calculated numericaly. The program runs the simulation three times for each parameter at the current parameter value and at slightly larger and slightly smaller parameter values. The reulting R-values at each parameter value are analyzed to calculate the derivative for this parameter.
A refinement within REFINE typically consists of two user supplied macro. In the main macro the refinement parameters, the input data and convergence criteria are defined. The actual calculation of the model function, respectively the model structure and its diffraction pattern, are carried out by the second macro.
The REFINE section uses four parameters to determine is the refinement has reached convergence. These are based on the absolute value of χ^2, and on the change of χ^2, the relative parameter change and a confidence level.
Convergence is considered to have been reached if either the absolute value of χ^2 falls below a user defined leve, or if all three other criteria are met. These three criteria are:

- The value of χ^2 changes less than a user defined value between two cycles
- The largest relative parameter shift is less than a user defined value. The relative parameter shift is defined as the abolute value of the parameter change between two cycles divided by the parameter uncertainty.
- The confidence level reaches a user defined level.

The correct values of χ^2, of the parameter uncertainties and the confidence value depend on the correct values of the data uncertainties. Often the raw data do not provide an accurate estimate of the data uncertainties. In this case the best strategy to start with is to assign unit wheights to all data points. As this implies that the optimum χ^2 and the confidence level will be unknown, REFINE also requires the user to set a maximum number of refinement cycles.

The REFINE section defines the parameters that shall be refined as DISCUS SUITE variable names. These variable names and their current values are passed on to the user macro that is used to simulate the model structure. To tune the refinement, an initial value, a valid parameter range, and a status flag that indicates if this parameter is to be refined or to be kept fixed can be provided.

Chapter 3

Example refinements

3.1 Simple noisy data function

In this first example the refinement to a simple function is illustrated. The data were calculated as a simple polynomial, see Fig. 3.1:

$$y(x) = P_{const} + P_{lin} * x + P_{quad} * x^2 + P_{trip} * x^3 \tag{3.1}$$

A Gaussian distributed random error was added to each data point with a sigma equal to the absolute y-value of each data point. If y was close to zero, the minimum sigma was set to 0.001. As y depends linearly on each of the parameters P_i, we can expect that this refinement will run smoothly. Essentially arbitrary starting values could be used to perform the refinement.
Two macro files are needed for the refinement, the main refinement macro and the macro that caculates the theoretical function. The main refinement macro that was used for this simple introductory example is:

```
 1: refine
 2: data xy, DATA/triple.noise
 3: newparam, P_const, value:-8.01
 4: newparam, P_lin  , value:1.01
 5: newparam, P_quad , value:1.49
 6: newparam, P_trip , value:0.31
 7: set  cycle,   5
 8: set conv, dchi:0.50, pshift:0.005, conf:0.10, chi2:0.5
 9: run triple.mac
10: exit   ! back to SUITE
```

In line 1 we switch from the main DISCUS SUITE level to the REFINE section. The data are loaded in line 2 as a simple "xy" file. All formats that are supported by the KUPLOT `load` command are available. REFINE stores teh data set within its on memory. Thus no harm is done if KUPLOT resets the data sets during its calculations. During a refinement in which you need to calculate and average multiple powder or PDF data sets such a reset migh be the norm rather than the exception.
Lines 3 to 6 define the four parameters we wish to refine. For each parameter a suitable name needs to be defined. This name has to be a variable name that is valid within the DISCUS SUITE. If the variable name does not yet exist, it will be created. The command offers three

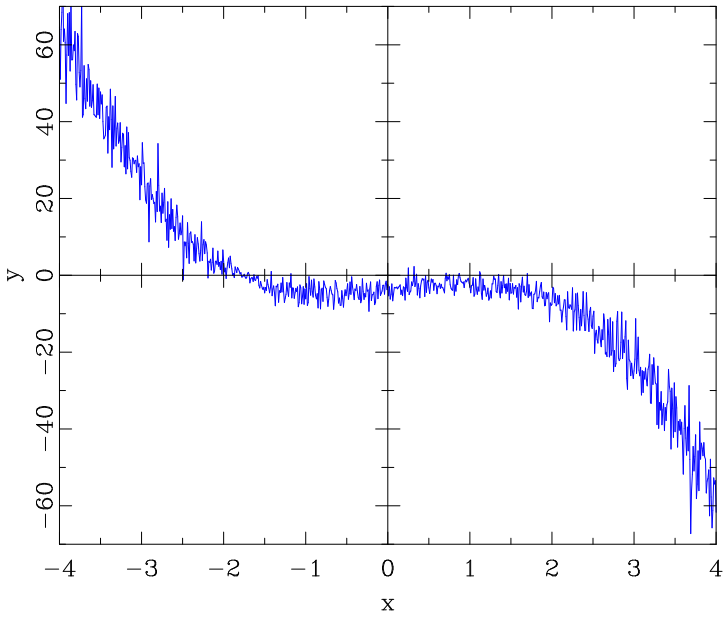

Figure 3.1: Experimental polynomial function

optional parameters, only the "value" parameter is used in this example. it initiates the parameter to the user supplied value, which can be a numerical expression. As further optional parameter you can limit the range over which a parameter is valid by the parameter "range:[<lower>,<upper>]. This will be helpful if you need to exclude negative numbers or if the physical range of a parameter is limited. The last optional parameter allows to change the status of a parameter with "status:fix" or "status:fixed" the parameter will be fixed. If a "value:" parameter is given, the parameter will be fixed to this value, otherwise it takes its current value. If the optional parameter "status:" is omitted or if its value is set to "refine" or "free", its value will be refined.
Line 7 defines the maximum number of refinement cycles that the REFINE section will perform. If the convergence criteria are met in an earlier cycle, the refinement will stop at that point. Several criteria are used to determine if the refinement has reached convergence. In this example they are all specified on line 8. The four criteria are:

- chi2:<value> Convergence has been reached if the value of chi2 drops below the user specified value.
- conf:<value> If the confidence level is larger than this value and
- pshift:<value> the largest parameter shift defined as the absolute value of the pareameter shift divided by the abolute value of the parameter uncertainty and
- dchi:<value> the change in chi2 is less than this value then convergence is assumed.

If either the `chi2` or all three other criteria `conf`, `pshift` and `dchi` are met, onvergence is reached and the refinement stops. Keep in mind that the value of chi2 and the parameter uncertainty wil depend on the uncertainties of the input data values.

The "run" command on line 9 will start the refinement using the macro "triple.mac" to calculate the theory data set.

```
 1: branch kuplot
 2: #
 3: if(F_DATA==0) then
 4:    reset
 5:    func (P_const)+(P_lin)*r[0]+(P_quad)*r[0]**2+(P_trip)*r[0]**3, F_XMIN, F_XMAX, F_XSTP
 6: else
 7:    func (P_const)+(P_lin)*r[0]+(P_quad)*r[0]**2+(P_trip)*r[0]**3, F_DATA
 8: endif
 9: exit
10: finished
```

The refinement begins within the REFINE section of the DISCUS SUITE. Since we want to calculate a simple x-y data set, the KUPLOT section is the best choice. With the `branch kuplot` command in line 1 REFINE switches to KUPLOT. In line 3 the macro tests if REFINE loaded the data into a KUPLOT data set with data set number `F_DATA`. The REFINE build in variable `F_DATA` defines the data set number within KUPLOT into which REFINE was able to load the data set with the `data` command, line 2 in the main fit macro. The actual limits and the step width that REFINE detects are stored in the REFINE variables `F_XMIN`, `F_XMAX` and `F_XSTP`. Thes variables are all Read/Write but change their values with caution.
In KUPLOT the `func` command will calculate a user supplied function over a user defined range. The first parameter here the string "(P_const)+(P_lin)*r[0]+(P_quad)*r[0]**2+(P_trip)*r[0]**3" defines the function to be calculated. If three parameter follow, they define x-min, xmax and the x-step for the function. The DISCUS SUITE variable r[0] takes the role of the functions x-coordinate. If only one parameter is supplied, it defines the data set number whose range and step width are to be used for the function calculation. In the current example macro lines 5 and 7 will thus calculate the theory function over the identical range.
A reason to deveate from the limits of the data set occures when you reifne against a powder diffraction data set. In order to take the effect of Bragg reflections slightly above the upper 2Theta limit into account, it is necessary to calculate the powder pattern to a larger 2Theta value. This will ensure that the low 2Theta tail of these Bragg reflections will b e present within the calculated powder patttern, even if the peak position of the Bragg reflections is not.
The user macro needs to ensure that the last data set loaded into KUPLOT is the final theory curve.
At line 9 the user macro returns to REFINE . Finally in line 10, the REFINE command `finished` tells REFINE that the user macro is finished.
For the current macro the output produced during the refinement will be the following lines:

```
Cyc Chi^2/(N-P)   MAX(dP/sig)   Conf          Lambda       wRvalue      Rexp
  0   1.0222      0.26239      0.34233      0.50000E-03  0.25018      0.24745
  1   1.0219      0.14187E-02  0.34404      0.25000E-03  0.25015      0.24745

Convergence reached

 Information about the fit :
   Chi^2      :  814.484          Chi^2/N :  1.01683
   Conf. level: 0.344038      Chi^2/(N-P) :  1.02194
   No.Data    :          801     No.Params:           4
   MRQ final  : 0.250000E-03
   wR value   : 0.250150          R exp   : 0.247451
```

```
Correlations larger than 0.8 :
  ** none **

  P_const          :   -3.92623      +-    0.107205
  P_lin            :    2.16236      +-    0.772162E-01
  P_quad           :   0.512809      +-    0.314855E-01
  P_trip           :   -1.11685      +-    0.134130E-01
```

Appendix A

REFINE commands

A.1 Summary

Here is a list and brief description of valid REFINE commands. Further help can be obtained by typing the corresponding command name at the help prompt.

```
data     ! defines the data set
newparam ! defines the parameters to be refined
fix      ! fixes a prameter, free with newparam
set      ! sets control values like cycles, convergence criteria
sigma    ! for 2D data defines a file that contains experimental uncertainties
News          : Information on program updates
```

A.2 Example

A simple fit might refine parameters of a straight line to observed data. Lets assume that the data are in a simple x/y file "observed.data". A suitable macro to refine is:

```
refine                          ! Switch from SUITE to REFINE section
data xy, observed.data          ! Load data set "observed.data"
                                ! The load command creates user variables
                                ! F_DATA = Number of Data set in Kuplot
                                ! F_XMIN, F_XMAX, F_XSTP
                                ! F_YMIN, F_YMAX, F_YSTP
                                ! That contain the data limits and step size.
newparam P_inter, value:1.0     ! Define y-axis intersept as first parameter
newparam P_slope, value:1.0     ! Define slope as second parameter
set cycle, 10                   ! Define maximum number of cycles
run fit_work.mac                ! Start the fit with user macro fit_work.mac
exit                            ! Back to SUITE
```

The user macro fit_run should be something like:

```
branch kuplot                   ! Step into the KUPLOT section
                                !   or for structures into DISCUS
func P_inter + P_slope*r[0], F_DATA
                                ! Calculate the function,
                                !   limits are set automatically from
                                !   data set F_data that was loaded by refine.
exit                            ! Back to REFINE
finished                        ! Special keyword signals end of user macro
```

A.3 data

```
data "kuplot", <number>
data <type>, <infile>
data "kuplot", <number>
```

The calculation will refine parameters versus the observed data that are stored in the KUPLOT data set number <number>

```
data <type>,<infile>
```

The calculation will refine parameters versus the observed data that are loaded from file <infile>.
See the ==> 'kuplot/load' command for details on proper ways to load a data set.
Refer to the KUPLOT section for instructions on loading data.

A.4 fix

```
fix <parname> [,"value:"<number>]
```

Fixes a parameter. Its value will remain at its current value or at <number>, if the optional parameter "value:" is used.
A fixed parameter can be freed with a ==> 'newparam' command.

A.5 newparam

```
newparam <parname> [,"value:"<start>] [,"status:"<flag>]
                   [,"range:["<lower>,<upper>"]"]
```

The command performs three tasks: - Defines a new parameter named <parname> - Frees a parametere <pname> that had previously been fixed by the ==> 'fix' command - Defines a constant <pname> that you may need to use in the user macro.
Defines a new parameter name <parname>. This should be any valid user defined variable name, limited to a length of 16 characters. The user variable is allowed to have been defined previously, and its current value will be used if the "value:" option is omitted.
See the general command line section for details on the definition of variables.
Optional parameters are: value:<start> The Parameter will be initialized to the <start> value. If omitted, the parameter value will take on its current value.
status:refine status:free status:fix status:fixed With the flags "refine" or "free" the parameter will be refined during the refinement cycles. With the flags "fixed" or "fix", the parameter will remain fixed at its current value. Be careful that the user macro does not change the parameter value! Default flag is "refine".
range:[<lower>, upper>] Defines a lower and upper boundary for the parameter. The fit will ensure that the parameter does not move outside the specified range. If the "range:" parameter is omitted, the default behavior is to assume no boundaries. In order to turn the boundaries off, simply state the 'newparam' command again for the refinement parameter without the "range:" option.

A.6 reset

```
reset
```

Buts the REFINE section back into the state at system start.

A.7 set

```
set cycle,<maxc>
set conver ["dchi:"<delta>] [,"pshift:"<max>] [,"conf:"<level>]
set relax  ["start:"<lamda_s>] [,"fail:"<lamda_f>]
           [,"success:"<lamda_g>]

set cycle,<maxc>
```

Sets the maximum number of refinement cycles

```
set conver [,"dchi:"<delta>] [,"pshift:"<max>] [,"conf:"<level>]
           [,"chi2:"<level>]
```

Allows the user to define convergence criteria.
dchi:<delta> If the value of Chi**2/(Ndata-Npara) decreases by less than <delta> convergence is reached. Defaults to 0.5 pshift:<max> If all refinement parameters change by less then |DeltaP/Sigma| convergence is reached. Defaults to 0.005 conf:<level> If the confidence level is greater than <level> convergence is reached. Defaults to 0.010 chi2:<level> If the value of Chi**2/(Ndata-Npara) falls below <level> convergence is reached. Defaults to 0.500
Convergence is reached, if: Either (dchi AND pshift AND conf) or chi2 are met.

```
set relax  ["start:"<lamda_s>] [,"fail:"<lamda_f>]
           [,"success:"<lamda_g>]
```

Allows to fine tune the relaxation behaviour. The initial Levenberg-Marquardt lamda parameter is defined with ["start:"<lamda_s>]. Default is 0.001
If a cycle was successful, the value of lamda is multiplied by <lamda_g>, default is 0.5.
If a cycle was not successful, the value of lamda is multiplied by <lamda_f>, default is 4.0.
A larger value of lambda implies smaller steps for the refinement parameters.

A.8 sigma

```
sigma "kuplot", <number>
sigma <type>,<infile>
sigma "kuplot", <number>
```

The calculation will refine parameters versus the observed data and use sigmas that are stored in the KUPLOT data set number <number>

```
sigma <type>,<infile>
```

The calculation will refine parameters versus the observed data and use sigmas that are loaded from file <infile>. See the ==> 'kuplot/load' command for details on proper ways to load a data set.
Refer to the KUPLOT section for instructions on loading data.

Bibliography

K. V. Price; R. M. Storn; J. A. Lampinen. *Differential Evolution; A Practical Approach to Global Optimization*. Springer, Berlin 2005.