**Automated weight optimization**

In refinements that involve more than one dataset (including restraints), manual assignment of weights is challenging. In RMCProfile, this task can be accomplished using an automated procedure that assigns weights by analyzing statistical correlations between the changes in each residual term and the total residual. Weights are adjusted during fit to ensure comparable effects of all the datasets/restraints on the total residual while forcing all the residual components to decrease over time to their respective preset minimal values. The adjustment is performed automatically as the fit progresses. The user has an option to increase the weight of a specific dataset manually without stopping the fit. Also, the user must specify a desired fraction of accepted moves to be maintained during the fit and minimal values of a residual for each dataset/restraint. Experience demonstrates that using the optimization option greatly facilitates the conversion while significantly reducing the fitting time even for problems that include 2-3 datasets

The algorithm of automated weight adjustments works as follows. The total residual after the nth atomic move is

where Nds is the number of fitted datasets, including restraints, Ri(*n*) is the residual for a dataset *i* after the nth move; wi is the weight assigned to the *i*th dataset, and *T* is the global weight rescaling parameter.

According to the Metropolis minimization algorithm that underlies RMCProfile, the moves are accepted so that decrees over time. Therefore, the changes at each move, Δ = -, are mostly negative. The adjustment is performed after every *N* generated atomic moves, where *N* is specified by the user. During each update, the weight-adjustment routine evaluates sequences of Δ for each dataset/restraint collected since the last evaluation. A set of weights is sought that yields a positive Pearson correlation coefficient for each dataset i versus Δ. Only those datasets/restraints are included in this analysis that have their residual values greater than the minimal values preset by the user.

The optimization procedure is activated if a file optimization.dat is present in the working directory. This is a text file with the following content:

2000

0.2

.false.

.false.

.false.

.false.

1. The entry in the 1st line specifies the **number of generated moves *N* after which the weight adjustment is to be performed**. This number should be large enough to collect adequate statistics on Δ but not too large as otherwise the efficacy of the procedure will be reduced. Experience shows that **2000 to 3000 moves is optimal.**
2. The value in the 2nd line specifies a desired **fraction of accepted atomic moves**. According to the Metropolis algorithm all moves that result in negative Δ are accepted unconditionally. Thus, the fraction set by the user can only affect the probability of acceptance of unfavorable moves, which lead to positive Δ. If the weights of all the datasets/restraints are multiplied by a factor 1/T close to zero, nearly all unfavorable moves will be accepted. In contrast, scaling these weights by a large value of 1/T will result in nearly complete rejection of such “bad” moves. At each weight evaluation, the routine adjusts 1/T so, as to provide the acceptance rate equal or greater than the specified value. For example, in the early stages of a fit, inevitably, nearly all moves are accepted. If the minimal acceptance rate is set too high, the fit may not converge whereas if this value is too low, few moves will be accepted, and the fit would freeze. **A value of 0.2 is a reasonable choice** but it can be varied according to specific needs.
3. The false/true entries on the next 4 lines control the generation of diagnostic files. For normal use, these entries should be kept as “**false**” (no diagnostic files produced).

The parameters specified in the optimization.dat file can be changed by editing this file (any text editor) without interrupting the fit.

The user can also modify relative weights assigned to specific datasets/restrains by modifying the content of the file **adv\_opt.dat**, which is generated by the program at the beginning of the fit. Any text editor can be used, and changes can be performed without interrupting the fit.

This **adv\_opt.dat** file contains two columns of numbers followed by the third column which lists the corresponding datasets/restraints. The 1st column lists the weight scale factors for each dataset/restraint included in the fit. By default, these values are set to 1. The user can increase the effect of a dataset on the total residual by increasing the scale factor applied to its weight. In the example below, setting this scale factor for the EXAFS1 data to 1.3 while leaving the rest at 1.0 will cause the residual for this EXAFS dataset to decrease faster. The 2nd column specified the minimum residual values for each dataset. The default values are zero. In the absence of systematic errors, these minimum values would be determined by counting statistics. However, in most practical cases, systematic errors prevail. Forcing the residual to zero is likely to result in overfitting. At present, the extent of systematic errors is generally unknown. Therefore, sensible minimum values of the residual for each dataset are determined by the user empirically. Once the residual for a given dataset reached its minimum value, its weight for the next evaluation cycle is set to zero. Over time, this residual will oscillate about its preset minimum value.

Note: every time RMCProfile is restarted, the program will re-generate the **adv\_opt.dat** file with the default values: unity for the scale factors and zeros for the minimum residual values. The user will have to edit this file again to set these parameters to their desired values.

EXAMPLE of adv\_opt.dat file

Guideline for using the procedure:

1. Set all the weights in the stemname.dat file that controls the fit to 1.0.
2. Generate **the optimization.dat** file, like described above, and place it in the working directory.
3. Start the fit. The program will generate the **adv\_opt.dat** file and the **stemname.log** file with the numbers of accepted/generated moves and values of the residuals for each dataset/restraint.
4. Monitor the progress of the fit by examining the log file. If needed, modify the adv\_opt.dat file without stopping the fit.
5. If the fit had to be stopped and re-started, leave the initial weights in the stemname.dat file at 1.0. after re-starting, the program will generate new **stemname.log** and **adv\_opt.dat** files (the latter with the default values). Edit **adv\_opt.dat** file if needed.