NAME

nlinLS − Nonlinear Spectral Lineshape Modeling

SYNOPSIS

nlinLS −in inTable −out outTable −data datafile **|** −list listName [−w widthList] [−mod modList] [−mult mType] [−apod apodListName] [−nots] [−fix parmList] [−delta deltaList] [−delta2 delta2List] [−limit limitList] [−tol funcTol] [−con confid] [−integ integLW] [−vol volScale] [−sum] [−iter maxIter] [−maxf maxFnEval] [−dig goodDigits] [−noise RMS] [−ppm] [−norm] [−verb] [−help]

DESCRIPTION

NLinLS is a general purpose system for modeling multidimensional signals. The models are constructed as the products of one−dimensional lineshape profiles, one per spectral dimension. The 1D profiles may be common spectral lineshapes (e.g. Gaussian or Lorentzian), relaxation expressions (e.g. exponentials) or time−domain modulation terms (sinusoids).

NLinLS attempts to minimize the least−squares error between regions in the target spectrum and the selected lineshape model by adjusting the model parameters. Confidence limits derived from the standard errors of the model parameters are also estimated.

NLinLS requires spectral data as part of its input. The spectral data can be stored in a single file (e.g. a 2D spectrum), or as a series of related files (e.g. 3D experiment, or a 2D series).

NLinLS also requires a peak table as its input; the peak table includes initial estimates for all model parameters, such as peak position and linewidth. It also includes information about which peaks must be modeled simultaneously because of overlap, in the form of cluster ID numbers. Cluster ID numbers are recorded in the peak table as CLUSTID parameters, as described below.

In addition to peak table input, NLinLS will generally also require specification of the size of a typical peak region, in data points, so that peak regions of the correct size can be extracted from a spectrum for modeling. Estimates of the RMS noise in the spectrum will also be required for generating error statistics.

Therefore, in a typical application, the following steps will be taken:

1. Use a peak picker to establish initial peak table.
2. Estimate spectral noise content.
3. Estimate size of typical peak region.
4. Identify clusters of overlapping peaks.
5. Select a model and invoke NLinLS.

NLinLS produces two peak table outputs. One lists the refined parameter estimates derived from the fitting procedure. The other, which is written to standard output, lists the estimated confidence limits in the model parameters, along with related statistics and diagnostics.

PEAK TABLES

The peak tables used by NLinLS allow for User−Specified content and formats. The general table format includes a VARS line, which identifies the columns of the peak table, and a FORMAT line, which suggests C−style output formats for the data in each column. The number of parameters listed in the VARS line must match the number of output formats in the FORMAT line. The C−style formats are limited to types %d (integer) %f (float) %s (text) and %e (scientific notation), and should not include spaces.

Each peak is described by a single line in the table, with a set of space−delimited parameters. The order of the parameters is determined by the definitions in the VARS line, and all fields must include a value. Blank lines, extra spaces, and REMARK lines are all permitted. Line length is limited to 255 characters.

For example, a 2D peak will typically be described by two positions, two linewidths, and a height. In this case, a initial peak table might look like this:

```
REMARK Sample Three−Peak 2D Peak table
VARS   INDEX CLUSTID X_AXIS Y_AXIS XW YW HEIGHT ASSIGN
FORMAT %4d   %4d     %5.2f  %5.2f  %5.3f %5.3f %+8e %s
```

```
1  1  122.00   76.00 2.50 3.75 −9.4545e+6 ALA81
2  1  129.00   70.00 2.20 3.55 +9.3995e+6 GLY66
3  2  455.00  121.00 2.10 3.25 +3.8210e+6 SER26
```

Note that the parameters can be given in any order, as long as peak entries are consistent with the parameter order specified in the VARS list. Note also that parameters unrelated to the model, like the ASSIGN parameter above, can be included in the input peak table for convenience; they will be reproduced in the output peak table.

Depending on the model selected, NLinLS will require that specific parameters be included as initial estimates in the input peak table. An error message will be generated if required parameters are missing. In addition, the input peak table can include optional parameters such as VOL, RMS, etc, listed below. These optional parameters allow for useful auxiliary information to be included in the output.

The first letter in a parameter name will generally indicate which axis of the spectrum it pertains to; for example, peak positions will be listed as X_AXIS, Y_AXIS, Z_AXIS, A_AXIS, etc. Note that NLinLS operates without regard to the transpose state of the data. So, by default, X_ parameters pertain to rows in the current dimension of the data; this may be either the F2 or F1 dimension, depending on the transpose state. The relative meaning of X_, Y_, and Z_ can be adjusted with the −axis switch.

A list of peak table parameter names recognized by NLinLS is given here. Note that the first letter of X_ parameter names listed below can be adjusted to correspond to the appropriate dimension:

HEIGHT
(Required Input) Peak height or overall model amplitude. All models must include exactly one HEIGHT parameter.

X_AXIS
Peak position in points, origin at 1.

XW Peak full linewidth at half−height, points.

X_FREQ
Sinusoidal frequency, as SIN(X_FREQ∗TAU). Used for time−domain modulation models.

X_DAMP
Exponential dampening, as EXP(X_DAMP∗TAU). Used for time−domain modulation models.

X_P0 Adjustable Zero−Order Phase, Degrees. Requires TIME1D GTIME1D or CTIME1D profiles.

X_P1 Adjustable First−Order Phase, Degrees. Requires TIME1D GTIME1D or CTIME1D profiles.

J1 J−Coupling value equivalent, in points; also J2, J3, J4.

X_A A generic coefficient. Exponential coefficient for relaxation profiles, as EXP(X_A∗TAU). Also, SCALE1D profile scaling coefficients, X_A1, X_A2, X_A3, etc.

X_B A generic coefficient. Exponential scaling coefficient for biexponential relaxation profiles, as EXP(X_A1∗TAU)+X_B∗EXP(X_A2∗TAU).

INDEX (Required Input) A unique integer identifying a peak in the table.

CLUSTID
(Required Input) An integer identifying a cluster, e.g. a group of peaks which should be fit simultaneously. Only peaks in the same cluster have the same CLUSTID value. For example, consider the sample peak table above; peaks 1 and 2 have a cluster ID of 1, indicating that they overlap with each other. Peak 3 has a cluster ID of 2, indicating that it does not overlap with either peak 1 or 2. Note that the utility "clustTab" can be used to add CLUSTID information to existing peak tables.

VOL (Output) The estimated peak volume, derived numerically, of a region around the peak center; the region size is determined as a +/− number of linewidths from the peak center, via the −integ flag. Currently, only profiles GAUSS1D LORENTZ1D and TIME1D will have volume estimates. Volume estimates will be included in the output only if the input peak table also contains a VOL parameter.

TROUBLE
(Output) An integer indicating that problems occurred during the modeling procedure; a value of zero means all went well.

CHI2 (Output) Results of the chi−squared test, which indicates whether the model residual can be accounted for by spectral noise. A high CHI2 probability suggests that the model residual is consistent with the spectral noise. Note that the CHI2 value depends on accurate noise estimates, as specified by the −rms flag.

PNORM
(Output) Results of the chi−squared test comparing the shape of the model residual to a normal distribution. A high PNORM probability means that the model residual is not substantially different from a normal distribution.

EXPL (Output) The percentage of variance in the data explained by the model.

RMS (Output) The RMS residual per point of the model.

OPTIONS

−in inTable
Specifies the input peak table identifying the peaks to fit. All parameters needed for the model must be included as estimates in this table.

−out outTable
Specifies the output peak table, which will contain the refined model parameters found by NLinLS. The format of the output file will match the format indicated in the input peak table.

−data datafileName
Name of the file containing the spectral data to analyze. Should not be used with the −list flag.

−list listName
Name of the text file listing the spectral data files in an analysis series, such as a relaxation experiment. The names are listed one per line. This option should not be used with the −data flag.

−w widthList [2 2 ...]
A list of the +/− widths of isolated peak regions in each dimension, given in points. The number of widths given should match the number of dimensions in the model. In the case that a given dimension does not have a peak region width associated with it (such as the build−up dimension of an NOE series) use the length of that dimension in place of the width. Note that it is VERY important to specify these widths correctly for successful use of NLinLS.

−mod modelList [GAUSS1D GAUSS1D]
List of the model profile types, one per dimension, in X,Y,Z order. Valid profile names and their associated parameters include:

```
GAUSS1D   Gaussian lineshape.
        X_AXIS XW
LORENTZ1D Lorentzian lineshape.
        X_AXIS XW
TIME1D    FFT of apodized, exp-damped sinusoid (+).
        X_AXIS XW
GTIME1D   FFT of apodized, gauss-damped sinusoid (+).
        X_AXIS XW
CTIME1D   FFT of apodized, un−damped sinusoid (+).
        X_AXIS
COS1D     Cosine modulation (∗).
        X_FREQ
SIN1D     Sine modulation (∗).
```

X_FREQ

DCOS1D Damped cosine modulation (∗).
X_DAMP X_FREQ

DSIN1D Damped sine modulation (∗).
X_DAMP X_FREQ

EXP1A1D Single Exponential (∗).
X_A

EXP2A1D Double exponential (∗).
X_A1 X_A2

EXP2AB1D Double Exponential, mixed scaling (∗).
X_A1 X_A2 X_B

SCALE1D Amplitude scaling vector.
X_A1 X_A2 X_A3 . . .

NULL1D Null profile; all values equal one.
No parameters.

Notes (+) Requires apodization function files via −apod.
(∗) Requires tau values recorded with the spectrum.
All models require exactly one HEIGHT parameter.

−mult multName [PEAK]

Describes the multiplet structure to use for the spectral model. A multiplet is described as a single entry in the peak table, which includes appropriate coupling values. The following multiplet types are recognized:

PEAK A single peak.
JX Two peaks, separated in X by J1.
JY Two peaks, separated in Y by J1.
JXY Two peaks, separated in X by J1, and in Y by J2.
AX Four peak antiphase multiplet, separated by J1.

−fix P1 P2 ... [None]

Allows the named model parameters P to be fixed during the optimization; normally, all model parameters will be adjusted. The values for the parameters named here will be kept at the original values listed in the input peak table.

−delta P1 A1 ... [None]

Constrains the named model parameters P to stay within the bounds of their initial values, as P-A:P+A.

−delta2 P1 A1 B1 ... [None]

Constrains the named model parameters P to stay within the bounds of their initial values, as P+A:P+B.

−limit P1 A1 B1 ... [None]

Constrains the named model parameters P to stay within the bounds explicitly specified, as P=A:P=B.

−apod apodListName

The name of a text file listing the names of apodization files, one per line. An apodization file is a 1D datafile in the same format as the spectrum, which contains the window function used on a given dimension of the spectrum. The size of the apodization file should reflect zero filling performed on the data; it should also contain the correct original time−domain size in the header information. Apodization files must be given when using TIME1D GTIME1D or CTIME1D profiles.

−nots This flag will suppress all scaling of time-domain models; by default the models are scaled according to the integrated decay and window, and according to the number of data points. Use this flag to measure absolute time-domain amplitudes.

–axis axisOrder [XYZABC]
A text string indicating the order of axes in the spectral data (e.g. the transpose state of the spectral data). Use this to exchange the meaning of X_, Y_, and Z_ peak parameters. For example, "–axis YX" would indicate a transposed 2D file.

–integ integLW [3.0]
Defines the size of peak integration regions, as +/– linewidths from the peak center. The integration result is reported as the VOL output peak table parameter.

–ppm When this flag is used, values for peak table positions and linewidths recorded Hz and ppm will be updated (X_PPM X_HZ XW_HZ etc.) after the fitting procedure. If this flag is not used, only the values recorded in points (X_AXIS XW etc.) will be changed.

–sum Allows measurement of volumes via summation of experimental intensities rather than via the theoretical lineshapes. The experimental data points within the integration linewidths of the peak center will be added to form the volume. The number of linewidths used can be specified by the –integ argument.

–vol volScale [1.0]
The reported volumes will be divided by the scale factor given here. This argument can be used to normalize volumes.

–tol relFnTol [1.0e–08]
Estimated relative residual tolerance for convergence.

–dig goodDig [5]
Estimated good digits in model residual.

–con confidVal [0.95]
Desired confidence limits to use for error estimates.

–noise noiseRMS [0.0]
Specifies the RMS noise per point in the spectral data. Required for correct error statistic PCHI2 produced by the –norm switch.

–iter maxIter [–1]
Sets the maximum iterations allowed per region. Values of 50–100 are common.

–maxf maxFnEval [–1]
Sets the maximum number of model function evaluations allowed per interation.

–norm Enables chi–squared testing of residual shape and magnitude. Generally used along with noise estimates given by the –noise switch.

–verb Turns verbose mode on. Verbose mode is highly verbose, since it will list all intensities in each region modeled.

–help Prints a brief list of flags and arguments.

FILES

2D Demonstration data is included in directory nlsdemo.

EXAMPLES

Fitting 2D data with peaksizes of about 9x11 points to gaussians:

```
nlinLS –in test.tab –data test.ft2 –mod GAUSS1D GAUSS1D \
   –w 4 5 –out test.nlin
```

Add residual statistics to the above model, given that the RMS spectral noise is 4600/point:

```
nlinLS –in test.tab –data test.ft2 –mod GAUSS1D GAUSS1D \
   –w 4 5 –out test.nlin –norm –noise 4600
```

Perform the fit as above, with peak positions constrained to within +/- 3 points of their initial values:

```
nlinLS –in test.tab –data test.ft2 –mod GAUSS1D GAUSS1D \
    –w 4 5 –out test.nlin –norm –noise 4600 \
    –delta X_AXIS 3 Y_AXIS 3
```

Perform the fit as above, with linewidths constrained to the range 3 to 6 points:

```
nlinLS –in test.tab –data test.ft2 –mod GAUSS1D GAUSS1D \
    –w 4 5 –out test.nlin –norm –noise 4600 \
    –limit XW 3 6 YW 3 6
```

Fitting 2D data with gaussians of fixed linewidth:

```
nlinLS -fix XW YW –in test.tabf –data test.ft2 –w 4 5 \
    –mod GAUSS1D GAUSS1D –out test.fix
```

Fitting a 2D NOE buildup series of 10 2D files:

```
nlinLS –in test.tab –list noe.list –w 4 5 10 \
    –mod GAUSS1D GAUSS1D EXP1D –out noe.nlin
```

Find lists of 2D volumes in buildup series above:

```
nlinLS –in test.tab –list noe.list  –w 4 5 10 \
    –mod GAUSS1D GAUSS1D SCALE1D –out noe2d.nlin
```

Fitting a 3D spectrum using time–domain based lineshapes:

```
nlinLS –in test3d.tab –list 3d.list –w 3 3 3 \
    –mod TIME1D TIME1D TIME1D –apod apod.list \
    –out 3d.nlin
```

SEE ALSO

Peak table inputs for NLinLS must include cluster information, in the form of CLUSTID parameters, which indicates peaks that should be modeled simultaneously. The utility "clustTab" can be used to add this information to an existing peak table.

The utilities "simSpecND" and "addNoise" can be used to generate synthetic spectra from NlinLS peak table output files.

DIAGNOSTICS

NLinLS will report the names of required parameters which are missing in the input peak table.

NLinLS will report diagnostic messages along with the error estimate output. If the input peak table includes a TROUBLE variable, the output peak table will also indicate which peaks and clusters caused errors.

BUGS

In some cases, NLinLS may fail to converge in the default number of iterations. This problem can be avoided by increasing the maximum allowable iterations or function evaluations using the –iter and –maxf flags.

NLinLS will often fail to converge when fitting groups of more than five peaks. In addition, very complicated or ill-determined models or poor initial parameter estimates may cause NLinLS to crash. Note also that large complicated regions may often take a very long time (hours) to converge or fail; adjusting maximum iteration and function evaluation counts may help.

These problems can sometimes be avoided by adjusting initial parameter estimates. They can also sometimes be avoided by reducing the complexity of the model by defining peak clusters to be as small and simple as possible.

Confidence limits supplied by NLinLS assume correct model solutions and normally distributed measurement errors. If these assumptions are not valid, the confidence limits should not be used.

Volumes reported for models with TIME1D profiles will not match actual numerical volumes from the spectral data, since the TIME1D volumes are calculated without adjustment for apodization.