APKM User manual

Contents

Chapter 1

Introduction

1.1 Author and contacts

This manual and program **APKM** is written by Maxim A. Dubinnyi. Information for contacts:

e-mail: maxim@nmr.ru

URL: http://www.model.nmr.ru/~maxim/APKM

ICQ: 176446633

Working phone: +7(095)335-27-33

Cell phone: +7(905)5115-77-1

1.2 Program version

This manual is written for program version 1.00, build 1216 (Release of December 2005). Starting from version 1070 the **APKM** is a module of the **Topspin** program, but not a stand-alone application.

1.3 Brief description

The program APKM uses symmetric isolated peaks, regions with positive/ negative signals and regions of flat baseline for automated phase correction of 1D NMR spectra. The automated phasing is performed by means of minimization of certain penalty function with four terms. The first term is responsible for phases of symmetric isolated peaks, the second accounts for regions with positive/negative signals, the third accounts for baseline regions, and the fourth gives additional penalty for large values of first-order phase correction parameter **PHC1**. The program is written as the module for the Topspin program.

All program parameters may be read either from configuration files, or specified as command line arguments.

1.4 Installation

Since the **Topspin** version **1.3** the **apkm** is the part of the **Topspin** distribution. To configure **apkm** for personal needs the user may edit the preset configuration files in the directory **$TOPSPIN_ROOT/classes/prop**. Each file with the name **apkm_<preset>.prop** contains some user-defined preset parameters for **apkm**, as described below.

1.5 Program run

1.5.1 Simple run

The simple execution of the APKM program from the Topspin command line shell is:

> **apkm** [<**preset**> |**0**|**1**]

All options are obligatory. Setting option **0** will execute automatic phasing for **PHC0** only, the value of **PHC1** is not changed. Option **1** will phase by changing **PHC1** but not **PHC0**. Option <**preset**> references to the existing file of the configuration parameters. The file with the name **apkm_<preset>.prop** in the directory **$TOPSPIN_ROOT/classes/prop** will be read to extract all configuration parameters of the program. The file **apkm_default.prop** will be used if no preset is selected. See the next section for description of the syntax of the **apkm** configuration files. Actual configuration parameters will be stored in the file **apkm.prop** of the current processing directory. Logging and debugging are not allowed if the **apkm** started via such a syntax.

1.5.2 Advanced run

The following syntax may be used to execute **apkm** module itself:

> **xcpr apkm [options...]**

Options may be skipped, in such a case the default values will be used. Any option should be either the value of the program parameter or the name of some configuration file with the list of program parameters:

option →(**Paramater_Name** = **Parameter_Value**) |
config_file_name

The space around equality sign '=' is not allowed in command line, but is admitted in configuration file. The path to configuration file may be absolute or relative, in the latter case the file is searched for in the following directories: processing directory of the current data set, the directory **$TOPSPIN_ROOT/exp/**

stan/nmr/lists/apkm and the directory **$TOPSPIN_ROOT/classes/prop**, in such order. The first occurence of the parameter either in command line or in some configuration file will have the presedence.

Parameter value is text string, real number or integer number:

Paramater_Value →**"String_Value"** |
Real_Value |
Integer_Value

In configuration file the strings with leading **'#'** are comments. Quotes around the string values are optional in command line, but obligatory in configuration file. Special characters in string values may be inserted after the 'at' letter (**@**). E.g., if quote is required inside the string value, it should be entered as **@"**, symbol **@** is types as **@@**.

The detailed description of the program parameters is given later in this manual.

1.6 Data input and output

The program **APKM** is the module of the program **Topspin**. It gets all required data from **Topspin**, makes automated phase correction and sends back the phased spectrum and values of parameters **PHC0** and **PHC1**. If requested, log file will be written to spectrum directory.

Prior to reading the data, current working directory is set to spectrum processing directory. By default, the program reads spectrum arrays from **1r** and **1i** files and spectrum parameters from **procs** file. Phased spectrum is written back to **1r** and **1i** files and new phasing parameters **PHC0** and **PHC1** are written to both **proc** and **procs** files. Program log level may be set to one of the values **0**, **1**, **2**, **3** or **4**. Values more then **2** should be used for debug purpouses only. Different log level may be selected for standard output and for output to log file. Log level **0** means silent execution with error messages only, **1** prints short program description and information about spectrum phases. Three values are printed for each phase **PHC0** ans **PHC1**: value befor phasing, value change suggested by **APKM** and new value written to **proc** files. The details of Input/Output related parameters are listed in table 1.1

Table 1.1: Input/Output related parameters of configuration file

Parameter	Description
Working_Dir	Working directory. Path to the acquisition directory of the spectrum to phase. By default, it is the acquisition directory of the current spectrum, as obtained from **Topspin**.
Bruk_Proc_Dir	Path to bruker processing directory relative to **Working_Dir**. Default value is **"pdata/1"**.
Bruk_File_Re	Input file for reading real part of the spectrum in bruker format. The file should be located in **Working_Dir/Bruk_Proc_Dir**. Default value is **"1r"**
Bruk_File_Im	Input file for reading imaginary part of the spectrum. Default value is **"1i"**
Bruk_File_Re_Out	Output file for writing real part part of the phased spectrum. Setting this value to an empty string (**""**) will suppress output of the phased spectrum. Default value is **"1r"**
Bruk_File_Im_Out	Output file for writing imaginary part part of the phased spectrum. Default value is **"1i"**
Bruk_File_Procs	Input/Output file for reading/writing processing parameters of the spectrum. The only written parameters are **PHC0** and **PHC1**. Default value is **"procs"**.
Bruk_File_Proc	Output file for writing **PHC0** and **PHC1** parameters of the phased spectrum. Default value is **"proc"**.
Debug_Level	Amount of programm messages appeared in flow-up windows of **Topspin**. Integer from **0** (error messages only) till **4** (huge amount of debug information). **1** prints only phases suggested by **APKM**, **2** prints detailed information about phasing progress. Larger values **3** and **4** should be used for debug pupouses only. Default value is **0**.
Log_File	Name of the file for duplication of **APKM** messages. File will be created in **Working_Dir/Bruk_Proc_Dir** unless **Log_Level**=0. Verbosity level for messages in this file is set by configuration parameter **Log_Level**. Default value is **"apkm.log"**.
Log_Level	Verbosity level for messages in the file **Log_File**. Possible values are the same as for parameter **Debug_Level**. Default value is **0** (don't generate **Log_File** at all).

Chapter 2

Principles of Automated Phase Correction in APKM

2.1 Penalty function

Spectrum phasing is performed by minimisation of penalty function $\boldsymbol{P}$ with four terms:

$$\boldsymbol{P} = \boldsymbol{P}_{S.P.} + \boldsymbol{P}_{B.R.} + \boldsymbol{P}_{S.R.} + \boldsymbol{P}_{\varphi_1} \tag{2.1}$$

Here the term $\boldsymbol{P}_{S.P.}$ accounts for simmetric isolated peaks in spectrum; the term $\boldsymbol{P}_{B.R.}$ accounts for regions of flat baseline, the term $\boldsymbol{P}_{S.R.}$ accounts for signal region in spectra and the term $\boldsymbol{P}_{\varphi_1}$ locks value of the first-order phasing parameter φ_1 in moderate region.

Each term in $\boldsymbol{P}$ is discussed later in appropriate section.

2.2 Phasing by symmetric isolated peaks

2.2.1 Recognition of symmetric isolated peaks in spectrum

Let us consider some NMR spectrum with real part $R_0(x)$ and imaginary one $I_0(x)$, where x denotes spectrum abscissa (e.g. frequency) in some convenient units. After zero-order phasing by φ degrees the new spectrum will have components:

$$\begin{aligned} R_1(x,\varphi) &= R_0(x)\cos(\varphi) - I_0(x)\sin(\varphi) \\ I_1(x,\varphi) &= R_0(x)\sin(\varphi) + I_0(x)\cos(\varphi) \end{aligned} \tag{2.2}$$

If this spectrum is phased correctly in region $[x_0 - W, x_0 + W]$ and has a local symmetry at the center of the region, then the following penalty function

$P(x,\varphi)$ should be relatively small for $x = x_0$:

$$P(x,\varphi) = \int_{-1}^{1} \big(R_1(x - Wt,\varphi) - R_1(x + Wt,\varphi)\big)^2 \rho(t)\, dt \tag{2.3}$$

Symmetric weighting function $\rho(t)$, $-1 \le t \le 1$ is introduced to give different penalty for points in window center and near it's borders. After substitution of (2.2) to (2.3) and some straightforward simplifications we obtain:

$$P(x,\varphi) = A(x) + B(x)\cos(2\varphi) + C(x)\sin(2\varphi) \tag{2.4}$$

Phase-independent coefficients $A(x)$, $B(x)$ and $C(x)$ may be calculated prior spectrum phasing as follows:

$$\begin{aligned} A(x) &= RR^+(x) + II^+(x) - RR^-(x) - II^-(x) \\ B(x) &= RR^+(x) - II^+(x) - RR^-(x) + II^-(x) \\ C(x) &= 2RI^-(x) - 2RI^+(x) \end{aligned} \tag{2.5}$$

Here new notation $XY^{\pm}(x)$ is introduced for:

$$XY^{\pm}(x) = \int_{-1}^{1} X(x + Wt)Y(x \pm Wt)\rho(t)dt \tag{2.6}$$

Symbols X and Y in (2.6) should be substituted by R and/or I from equation (2.5). It is obvious from equation (2.4) that:

$$\forall\varphi \quad A(x) - \sqrt{B(x)^2 + C(x)^2} \le P(x,\varphi) \le A(x) + \sqrt{B(x)^2 + C(x)^2} \tag{2.7}$$

Since all elements of equation (2.7) are nonnegative, we can extract square root from this equation and obtain:

$$\forall\varphi \quad \mathrm{LowLim}(x) \le \sqrt{P(x,\varphi)} \le \mathrm{UppLim}(x) \tag{2.8}$$

New functions $\mathrm{LowLim}(x)$ and $\mathrm{UppLim}(x)$ are defined as:

$$\begin{aligned} \mathrm{LowLim}(x) &= \sqrt{A(x) - \sqrt{B(x)^2 + C(x)^2}} \\ \mathrm{UppLim}(x) &= \sqrt{A(x) + \sqrt{B(x)^2 + C(x)^2}} \end{aligned} \tag{2.9}$$

Recall that $\mathrm{LowLim}(x)$ and $\mathrm{UppLim}(x)$ can be calculated *befor* spectrum phasing, so this two data sets may be used for analysis of local symmetry of *not phased spectrum*, and, consequently, for search of symmetric isolated peaks and regions of flat baseline in this spectrum. Evaluation of the functions $\mathrm{LowLim}(x)$ and $\mathrm{UppLim}(x)$ is time consuming due to numeric integration in (2.6), so a number of parameters are used to control such evaluations, see table 2.1.

Figure 2.1 shows functions $\mathrm{LowLim}(x)$ and $\mathrm{UppLim}(x)$ in the vicinity of symmetric peak in sample spectrum. Function $\mathrm{UppLim}(x)$ has a peak in the

Table 2.1: Parameters of configuration file, responsible for evaluation of LowLim and UppLim data.

Parameter	Notation	Description
Window_Width	W	Width of symmetrisation window, eqs. (2.3, 2.6, 2.10a, 2.10d, 2.10e). Default value is 20 Hz.
Window_Width_Units	—	Units of **Window_Width**, may be one of the following: **0**: Data points; **1**: Hz; **2**: ppm's. Default value is **1**.
Window_Function	$\rho(t)$	Window function, used in search of local symmetry, eqs. (2.3, 2.6). One of the following window functions may be selected by user: **0**: $\rho(t) = 1$; **1**: $\rho(t) = 1 - t^2$; **2**: $\rho(t) = 1 - t^4$; **3**: $\rho(t) = \cos(\pi t/2)$. Default value is **1**.
Window_Max_Points	—	Maximum allowed number of spectrum points per window. The spectrum will be scaled down prior to symmetry search provided that the **Window_Width** covers more then **Window_Max_Points** points of the spectrum. Default value is **24**.
Window_Sparce_Step	—	The step (in spectrum points) used for preliminary search of extrema in UppLim and LowLim data. Larger value gives faster search but may result in loose of some reference peaks for phasing. Default value is **3**.

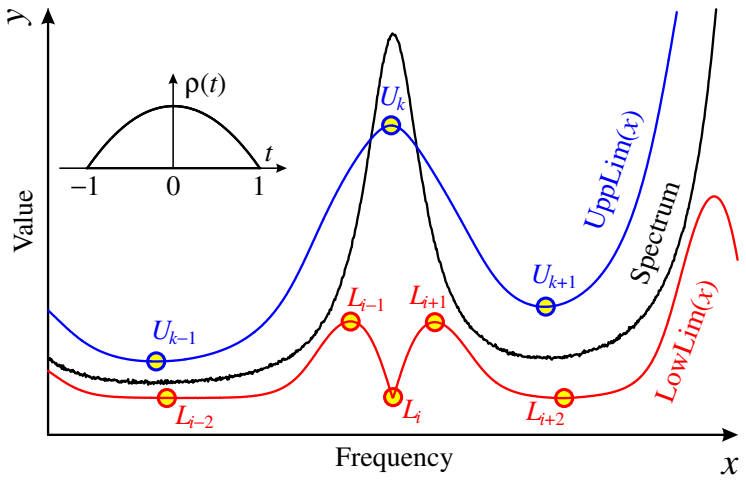

Figure 2.1: Principles of selection of symmetric isolated peaks in spectrum by means of two data sets LowLim(x) and UppLim(x). Inset shows window function $\rho(t) = 1 - t^2$, eqs. (2.3, 2.6) used in this example.

same position as original spectrum $R_0(x)$, while LowLim(x) looks like a hill with a deep pit. The more deep this pit is, the more symmetric peak may be obtained by spectrum phasing. To extract symmetric isolated peaks from *not phased* spectrum, LowLim and UppLim data lists are searched for maxima and minima as shown on figure 2.1, the found extrema will be denoted as L_i and U_k, $i, k \in \mathbb{Z}$ for LowLim and UppLim data, consequently. Let us consider minimum L_i in LowLim and neighboring extrema in LowLim and UppLim. This extremum is considered as corresponding to some symmetric isolated peak, if the following 5 criteria are passed:

$$\exists k: \; U_k \text{ is maximum and } \; |x(U_k) - x(L_i)| \le \beta_{UL} W \tag{2.10a}$$

Criteria (2.10a) checks that minimum L_i in LowLim has corresponding maximum U_k in UppLim and U_k is located close to L_i. Pit deepness criteria

$$\frac{\min\big(y(L_{i-1}), y(L_{i+1})\big)}{y(L_i)} \ge r_L \tag{2.10b}$$

Checks that pit in LowLim data is essentially deep. Relative height of maximum U_k for k from (2.10a) is checked by the ratio:

$$\frac{y(U_k)}{\max\big(y(U_{k-1}), y(U_{k+1})\big)} \ge r_U \tag{2.10c}$$

Peak isolation is checked by two separation criteria:

$$\max\big(\big|x(L_i) - x(L_{i-2})\big|, \big|x(L_i) - x(L_{i+2})\big|\big) \ge \beta_L W \tag{2.10d}$$

$$\max\big(\big|x(U_k) - x(U_{k-1})\big|, \big|x(U_k) - x(U_{k+1})\big|\big) \ge \beta_U W \tag{2.10e}$$

Table 2.2: Parameters of configuration file, responsible for selection of symmetric isolated peaks

Parameter	Notation	Description
LowL_UppL_sep_Max	β_{UL}	Maximum allowed separation of minimum in LowLim and maximum in UppLim, eq. (2.10a). The value is multiplier of W. Default value is **0.2**.
LowL_Ratio_Min	r_L	Dimensionless ratio in equation (2.10b). Default value is **4.0**.
UppL_Ratio_Min	r_U	Dimensionless ratio in equation (2.10c). Default value is **4.0**.
LowL_Sep_Min	β_L	Minimum separation of neigbour extrema in LowLim data, multiplier of W, see eq. (2.10d) and figure 2.1. Default value is **0.5**.
UppL_Sep_Min	β_U	Minimum separation of neigbour extrema in UppLim data, multiplier of W, see eq. (2.10e) and figure 2.1. Default value is **0.5**.

All minima in LowLim(x) which fulfills criteria (2.10) are considered as symmetric isolated peaks and used further as references for spectrum phasing. Relevant parameters of the program configuration file are listed in table 2.2.

2.2.2 Contribution of symmetric isolated peaks to penalty function

Each symmetric isolated peak L_i has it's own phase Φ_i

$$\begin{aligned} \Phi_i &= \begin{cases} \Phi_{i,1} & \text{if } R_1(x_i,\Phi_{i,1}) > R_1(x_i,\Phi_{i,2}), \text{ see eq. (2.2)} \\ \Phi_{i,2} & \text{otherwise} \end{cases} \\ \Phi_{i,1} &= \tfrac{1}{2}\arctan(-B_i,-C_i) \quad \text{see eq. (2.5)} \\ \Phi_{i,2} &= \Phi_{i,1}+\pi \\ x_i &= x(L_i), \qquad B_i = B(x_i), \qquad C_i = C(x_i) \end{aligned} \tag{2.11}$$

The phases of $N_{S.P.}$ symmetric isolated peaks are used as the first term in penalty function:

$$\boldsymbol{P}_{S.P.}(\varphi_0,\varphi_1) = w_{S.P.}\frac{1}{N_{S.P.}}\sum_{i=1}^{N_{S.P.}} D(\varphi_0+\varphi_1 x_i-\Phi_i) \tag{2.12}$$

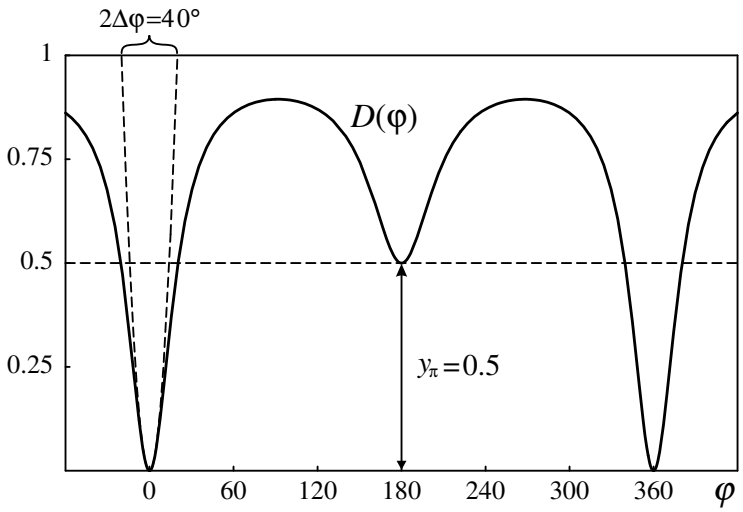

Figure 2.2: Plot of angular penalty function $D(\varphi)$, eq. (2.13) with default values of parameters $y_\pi = 0.5$ and $2\Delta\varphi = 40°$.

Here $w_{S.P.}$ gives the relative weight of the term, angular penalty function $D(\varphi)$ defines penalty value originated from odd phasing of the peak. The following angular penalty function is used in **APKM**:

$$D(\varphi) = \frac{1}{1 + \frac{1}{b(1-\cos(\varphi)} + \frac{1}{a+b(1+cos(\varphi)}}$$
$$a = \frac{y_\pi}{1 - y_\pi(1 + \frac{\Delta\varphi^2}{4})} \qquad b = \frac{2}{\Delta\varphi^2} \tag{2.13}$$

Figure 2.2 shows the plot of $D(\varphi)$ with default values of parameters y_π and $\Delta\varphi$. This function has values in range $0 \leq D(\varphi) \leq 1$, so the term $\boldsymbol{P}_{S.P.}$ (2.12) lies in the range $0 \leq D(\varphi) \leq w_{S.P.}$. Function $D(\varphi)$ has two minima, the main minimum is at $\varphi = 0$ with value 0 should catch most positive peaks of the spectrum. The second minimum at $\varphi = \pi$ and function value $D(\varphi) = y_\pi$ accounts for some probability of negative signals in the spectrum. Setting $y_\pi = 0$ makes positive and negative signals equivalent, $y_\pi = 1$ prohibits negative signals. Minimum width is regulated by the parameter $\Delta\varphi$, it should be close to the error in estimation of the peak phase. If an peak can't be phased with this accuracy then it is effectively ignored by the penalty function. Relevant parameters of the program configuration file are listed in table 2.3.

Table 2.3: Parameters of configuration file, which accounts for symmetric isolated peaks in spectrum phasing

Parameter	Notation	Description
Weight_Peaks	$w_{S.P.}$	Relative weight of the term $\boldsymbol{P}_{S.P.}$ in overall penalty function (2.1). This term accounts for symmetric isolated peaks in the spectrum, eq. (2.12). Default value is **1.0**.
Peak_Nagative_Score	y_π	This parameter reflects probability of finding negative peaks in spectrum, $0 \leq y_\pi \leq 1$. Setting $y_\pi = 0$ makes positive and negative signals equivalent, $y_\pi = 1$ prohibits negative signals. See eq. (2.13) and figure 2.2. Default value is **0.5**.
Peak_Phase_Delta	$\Delta\varphi$	This parameter should be close to error in estimation of the peak phase (in degrees). If an peak can't be phased with this accuracy then it is effectively ignored by the penalty function. See eq. (2.13) and figure 2.2. Default value is **20** °.

2.3 Phasing by baseline and signal regions

2.3.1 Recognition of baseline and signal regions

The LowLim(x) data list is used for detection of continuous baseline regions and regions with series of positive/negative signals. It can be easily observed, that LowLim(x) in baseline regions is close to small constant value. Using this observation, baseline regions are selected as follows:

1. Select all point of spectrum with LowLim(x) $\leq V_1 l_n$, where V_1 is selected threshold value and l_n is the noiselevel of LowLim(x) data. As the result the list of intervals in x-domain is obtained;
2. Expand each interval right and left by value Δx_E;
3. Fill small gaps between intervals. Small gap should (i) have length less then Δx_G and (ii) should not contain points with LowLim(x) $\geq V_2 l_n$, where $V_2 \geq V_1$ is another selected threshold value;
4. Remove intervals with length less then Δx_R;

5. To obtain the list of intervals with *signals*, inverse all intervals, to obtain the list of intervals with *baseline regions*, skip this step;

6. Optionally remove both regions of length Δx_B near spectrum borders and region of length Δx_C in spectrum center. These regions usually contain spectrum artifacts.

List $\{I_k\}$ of intervals in x-domain is obtained after step 6. This list is assumed to be the list of baseline regions or the list of signal regions (in dependence of step 5).

2.3.2 Contribution of baseline intervals to penalty function

All intervals in $\{I_k\}$ are grouped into sequential pairs $I_k \cup I_{k+1}$ and standard deviation of $R_1(x)$ over each pair of intervals contributes to second penalty term $P_{B.R.}$:

$$\boldsymbol{P}_{B.R.}(\varphi_0, \varphi_1) = \frac{w_{B.R.}}{P_{max}} \frac{1}{N_{B.R.} - 1} \sum_{k=1}^{N_{B.R.}-1} \operatorname{logp}\big(1, \boldsymbol{D}[Y_k]\big) \tag{2.14}$$

The notations used are: $w_{B.R.}$ – weight of the term of baseline regions in the overall penalty function; $N_{B.R.}$ – number of baseline regions found in the spectrum: P_{max} is defined later in equation (2.19); $\boldsymbol{D}[Y]$ – dispersion of the dataset Y, $\boldsymbol{D}[Y] = \langle Y^2 \rangle - \langle Y \rangle^2$. The function $\operatorname{logp}(a, y)$ and the data sets Y_k are defined as:

$$\operatorname{logp}(a, y) = \frac{1}{\frac{1}{y} + \frac{1}{a \log(1+y)}} \tag{2.15}$$

$$Y_k \stackrel{\text{def}}{\equiv} \{y\} : y = \frac{R_1(x, \varphi_0 + \varphi_1 x)}{y_n} \quad \& \quad x \in I_k \cup I_{k+1} \tag{2.16}$$

Here I_k is the k-th baseline interval and y_n is the spectrum noiselevel.

The term (2.14)will be relatively small if two conditions are fulfilled:

1. Ordinates of phased spectrum $R_1(x, \varphi_0 + \varphi_1 x)$ in each interval of baseline are scattered no more then noiselevel;

2. Sequential intervals of baseline have close mean values

Arbitrary spectrum should not contribute more then $w_{B.R.}$ to the baseline term of the penalty function, but an accurately phased spectrum should contribute much less. Relevant parameters of the configuration file are described in table 2.4.

Table 2.4: Parameters of configuration file relevant to baseline phasing

Parameter	Notation	Description
Weight_Baseln_Region	$w_{B.R.}$	Relative weight of the term $\boldsymbol{P}_{B.R.}$ in overall penalty function (2.1). This term accounts for baseline region in spectrum. See eq. (2.14). Default value is **1.0**.
Baseline_Data_Size	$N_{P.}$	Spectrum is scaled down to $N_{P.}$ points prior to baseline selection and building of baseline term in penalty function. Default value is **1024**.
Baseline_Cutoff_1	V_1	First cutoff parameter in baseline selection, see section 2.3.1, step 1. The parameter is multiplied by noiselevel to obtain actual cutoff threshold. Default value is **1.2**
Baseline_Cutoff_2	V_2	Second cutoff parameter in baseline selection. See section 2.3.1, step 3. The parameter is multiplied by noiselevel to obtain actual cutoff threshold. Default value is **7.0**
Baseline_Mrgn_Size	Δx_E	Value of expansion of each baseline interval in ppm's. See section 2.3.1, step 2. Default value is **0.0 ppm**
Baseline_Gap_Size	Δx_G	Value of negligeble gap size between sequential intervals of baseline in ppm's. See section 2.3.1, step 3. Default value is **0.05 ppm**
Baseline_Regn_Size	Δx_R	Minimal length of interval of baseline in ppm's. See section 2.3.1, step 4. Default value is **0.1 ppm**
Exclude_Margin	Δx_B	Length of intervals near spectrum borders excluded from baseline phasing in ppm's. See section 2.3.1, step 6. Default value is **0.1 ppm**
Exclude_Center	Δx_C	Length of interval in the spectrum center excluded from baseline phasing in ppm's. See section 2.3.1, step 6. Default value is **1.0 ppm**

2.3.3 Contribution of signal intervals to penalty function

Each signal region S_i contributes a term to the penalty function (2.1). The term $\boldsymbol{P}_{S.R.}$ is defined as:

$$\boldsymbol{P}_{S.R.}(\varphi_0, \varphi_1, y_0) = \frac{w_{S.R.}}{P_{max}} \frac{1}{N_{P.}} \sum_i \sum_{k:x_k \in S_i} U\left(\frac{R_1(x_k, \varphi_0 + \varphi_1 x_k) - y_0}{y_n} \right) \tag{2.17}$$

Notations are: $w_{S.R.}$ – relative weight of signal phasing term, $N_{P.}$ – number of points in the spectrum, y_0 – spectrum zero level, y_n – spectrum noise level. Asymmetric balance function $U(y)$ is a kind of heuristic and has the form:

$$U(y) = \begin{cases} \mathrm{logp}(1, y^2), & y > 0 \\ 0, & y = 0 \\ \mathrm{logp}(u_0, y^2), & y < 0 \end{cases} \tag{2.18}$$

The normalization factor P_{max} accounts for maximum allowed penalty $U(y)$ per data point:

$$P_{max} = \max\big(U(2y_{max}/y_n), U(-2y_{max}/y_n)\big) \tag{2.19}$$

and y_{max} is the maximum spectrum value.

The function $U(y)$ has the following properties. It is assimptotically quadratic at low y: $y \ll 1$, where the spectrum should contain mainly baseline points. It is logarithmic at high y: $y \gg 1$, where the spectrum should contain signal points. Balance parameter u_0 provides u_0-times large penalty for negative signals prior to positive ones. This function is a kind of "robust estimation". Figure 2.3 displays this function for default value of parameter $u_0 = 5$. Very large values of this parameter should be used for spectrum with positive signals only. Relevant parameters of configuration file are listed in table 2.5.

2.4 Locking of linear phasing in moderate region

The value of the coefficient of the linear phase correction **PHC1** $= \varphi_1$ does not have natural upper limit and may be set to arbitrary large value. But an experience of pahsing of ordinary experimantal spectra gives rise to moderate values of φ_1. This results from sertain efforts made while data acquisition and pulse sequence design to obtain preliminary correct spectrum phasing. On the other hand, the penalty function (2.1) may have inproper local minima for large values of φ_1. Trapping of this minima results in wrong spectrum phasing. To avoid large φ_1 values one should add the following term to penalty function:

$$\boldsymbol{P}_{\varphi_1}(\varphi_1) = w_{\varphi_1} \begin{cases} 0, & |\varphi_1| \le \varphi_1^{\max} \\ (\varphi_1 - \varphi_1^{\max})^2 & \text{otherwise} \end{cases} \tag{2.20}$$

Relevant parameters of configuration file are listed in table 2.6.

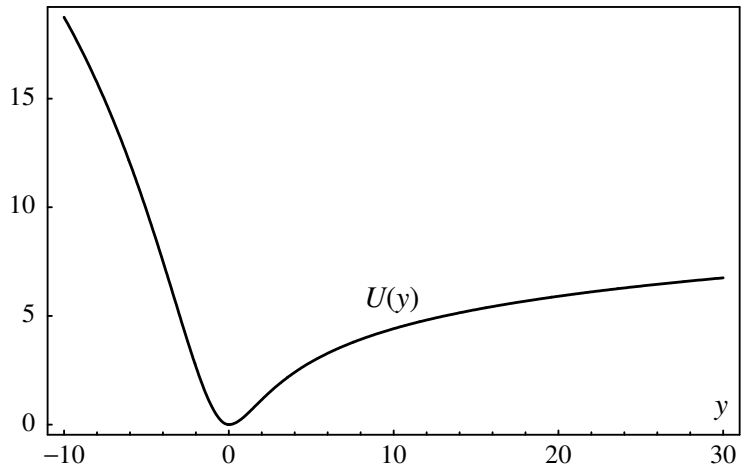

Figure 2.3: Plot of asymmetric balance function $U(y)$, eq. (2.18) with default value u_0=5.

Table 2.5: Parameters of configuration file relevant to signal phasing

Parameter	Notation	Description
Weight_Signal_Region	$w_{S.R.}$	Relative weight of the term $\boldsymbol{P}_{S.R.}$ in overall penalty function (2.1). This term accounts for signal regions in spectrum. See eq. (2.17). Default value is **1.0**.
Signal_Nagative_Ratio	u_0	Ratio of penalty for large positive/negative signals. See equation (2.18) and figure 2.3. Default value is **5.0**.

2.5 Minimization of penalty function

Penalty function (2.1) is minimized by conjugate gradient method [1]. The full penalty function $\boldsymbol{P}$ (2.1) has three arguments: the angle of zero-order phase correction φ_0, the angle of first-order phase correction φ_1 and spectrum zero level y_0 arised from the term $\boldsymbol{P}_{S.R.}$, eq. (2.17).

$$\boldsymbol{P} \equiv \boldsymbol{P}(\varphi_0, \varphi_1, y_0) \rightarrow \min \tag{2.21}$$

The program will vary all but explicitely freezed parameters to achive a minimum of the penalty function. Minimization routine is sequantially started from

Table 2.6: Parameters of configuration file relevant to PHC1 locking

Parameter	Notation	Description
Weight_PHC1_Lock	w_{φ_1}	Relative weight of the term $\boldsymbol{P}_{\varphi_1}$ in overall penalty function (2.1). This term prohibits large values of φ_1 while function minimisation. See eq. (2.20). Default value is **10.0**.
PHC1_Lock_Limit	$\varphi_1^{\max}$	Maximum value of φ_1 which is not penalized. See equation (2.20). Default value is **360.0**°.

a grid of seed values for φ_0, φ_1 supplied by the user:

$$\varphi_0^{(0,j)} = \varphi_0^{(0)} + j\Delta\varphi_0, \quad 0 \leq j \in \mathbb{Z}, \quad \varphi_0^{(0,j)} \leq \varphi_0^{(1)} \tag{2.22}$$

$$\varphi_1^{(0,k)} = \varphi_1^{(0)} + k\Delta\varphi_1, \quad 0 \leq k \in \mathbb{Z}, \quad \varphi_1^{(0,k)} \leq \varphi_1^{(1)} \tag{2.23}$$

The seed value for y_0 is always 0. For explicitely freezed angles the grid (2.22, 2.23) is not created. The following minimization protocol is executed starting from each seed value:

1. Freeze φ_1 to it's seed value $\varphi_1^{(0,k)}$. Account for $\boldsymbol{P}_{S.R.}$ term only. Find φ_0 and y_0 parameters by minimisation of $\boldsymbol{P}$.

2. Freeze y_0. Account for $\boldsymbol{P}_{S.P.}$ and $\boldsymbol{P}_{\varphi_1}$ terms only. Find parameters φ_0 and φ_1 by minimization of $\boldsymbol{P}$. Skip this step if no symmetric isolated peaks were found by the method described in section 2.2.1. If only one symmetric isolated peak was found, adjust φ_0 parameter only.

3. Perform full minimization with all terms in $\boldsymbol{P}$, vary all parameters but explicitely freezed by the user.

The attempt resulted in the least value of penalty function is accepted. Appropriate values of φ_0, φ_1 are used to perform spectrum phasing. The values are saved in processing files (see table 1.1 and section 1.6). Relevant parameters of configuration file are listed in table 2.7.

2.6 Example configuration files

There are example configuration files distributed with the program. They are suitable for sertain purpouses. The files should be placed to the directory **$TOPSPIN_ROOT/classes/prop** to be accessible from **Topspin**. Several files may be used together to combine their effects. These filese are:

Table 2.7: Parameters of configuration file relevant to minimization of the penalty function (2.1)

Parameter	Notation	Description
Max_Iterations		Maximum number of iterations performed by conjugate gradient method. Default is **50**.
Find_PHC0		Whether minimizer should vary φ_0?. Zero value freezes φ_0, other integer value allows parameter variation. Default is **1**.
Find_PHC1		Whether minimizer should vary φ_1?. Default is **1**.
Find_Baselevel		Whether minimizer should vary spectrum baselevel y_0?. Default is **1**.
PHC0_Grid_Start	$\varphi_0^{(0)}$	Origin of the grid of seed values for φ_0, see equation (2.22). Default value is $\mathbf{-120.0}^\circ$.
PHC0_Grid_Step	$\Delta\varphi_0$	Step of the grid of seed values for φ_0, see equation (2.22). Default value is $\mathbf{120.0}^\circ$.
PHC0_Grid_End	$\varphi_0^{(1)}$	End of the grid of seed values for φ_0, see equation (2.22). Default value is $\mathbf{120.0}^\circ$.
PHC1_Grid_Start	$\varphi_1^{(0)}$	The same grid parameter for φ_1, equation (2.23). Default value is $\mathbf{0.0}^\circ$.
PHC1_Grid_Step	$\Delta\varphi_1$	The same grid parameter for φ_1, equation (2.23). Default value is $\mathbf{0.0}^\circ$.
PHC1_Grid_End	$\varphi_1^{(1)}$	The same grid parameter for φ_1, equation (2.23). Default value is $\mathbf{0.0}^\circ$.

phc0	This file switches off the linear phasing, only **PHC0** but not **PHC1** is adjusted.
peaks	Use only symmetric isolated peaks for spectrum phasing. See the section 2.2 below for detailed description.
phc1grid	Use extensive grid search to find correct **PHC1** value. Generally takes a lot of time.
w40hz	Set large window width (40 Hz) for symmetry search. Results in more slow, but more robust phasing.

2.6.1 Configuration presets "phc0"

Configuration file **apkm_phc0.prop** switches off the linear phasing, only **PHC0** but not **PHC1** is adjusted. The presets of this file are activated by execution of the following command from the **Topspin** command line:

> **apkm phc0**

The contents of the configuration file is:

```
#  Example configuration file for APKM
#
#  This file switches off the linear phasing,
#  only PHC0 but not PHC1 is adjusted

Find_PHC0 = 1
Find_PHC1 = 0
```

Activation of this configuration file has the same effect as execution of the command:

> **apkm 0**

2.6.2 Configuration presets "peaks"

The configuration file **apkm_peaks.prop** is used to activate only phasing by symmetric isolated peaks of the spectrum. It is usefull for phasing of the spectra which are rich with such peaks or for testing of the phasing algorithm. The presets of this file are activated by execution of the following command from the **Topspin** command line:

> **apkm peaks**

The contents of the configuration file is:

```
#  Example configuration file for APKM
#
#  Use only symmetric isolated peaks for spectrum phasing

Weight_Peaks         = 1
Weight_Baseln_Region = 0
Weight_Signal_Region = 0
Weight_PHC1_Lock     = 1
```

This presets may fail on phasing of extremely overlapped spectra wich have no good reference peaks for phasing.

2.6.3 Configuration presets "phc1grid"

The configuration file **apkm_phc1grid.prop** is used for extensive grid search to find correct **PHC1** value. The phasing generally takes a lot of time. This preset is activated by the command:

> **apkm phc1grid**

The contents of the configuration file is:

```
#  Example configuration file for APKM
#
#  Use extensive grid search to find correct PHC1 value
#  Generally takes a lot of time

PHC1_Lock_Limit =  720.0
PHC1_Grid_Start = -720.0
PHC1_Grid_Step  =  180.0
PHC1_Grid_End   =  720.0
```

As can be seen from the configuration file, the program systematically starts minimisation from different **PHC1** values in the range -720 to $720°$.

2.6.4 Configuration presets "w40hz"

The configuration file **apkm_w40hz.prop** uses larger window width (40 Hz instead of default 20 Hz) for symmetry search. This is useful for spectra with broadened signals. The contents of the configuration file is:

```
#  Example configuration file for APKM
#
#  More slow, but more robust vershion of the phasing

Window_Width_Units=1
Window_Width=40
```

List of Tables

Bibliography

[1] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Conjugate Gradient Methods in Multidimensions*, chapter 10.6, pages 420–425. In [2], 1992.

[2] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.