PcBc
NOVEL AUTOMATIC PHASE & BASELINE CORRECTION ALGORITHM

Ester Maria Vasini[1], Carlos Cobas[2] and Stanislav Sýkora[1]

[1] Extra Byte, Via Raffaello Sanzio 22/C, Castano Primo (Mi), Italy I-20022; em.vasini@extrabyte.eu, sykora@extrabyte.eu

[2] Mestrelab Research, Xosé Pasín 6, Santiago de Compostela, 15706 Spain; carlos@mestrec.com

DOI: 10.3247/SL7Nmr18.001

INTRODUCTION

Manually, phase (Pc) and baseline (Bc) corrections in 1D NMR spectra were always done as separate steps, carried out one after another. In addition, the Bc was always done only on the "real" part of the spectrum and not on the imaginary part. However, one often notices that there is a mutual interference between the two corrections, causing uncertainty about the best solution in the vicinity of an optimum. In addition, the fact that baseline correction is not done on the imaginary part implies that when one tries to iterate the whole process, any change in phase parameters brings back into the displayed real part the uncorrected baseline artifacts present in the imaginary part. Consequently, though a lot of work on the two (separate) corrections has been done over the years, further investigation is still called for. In particular, the problem becomes imperative when one needs a fully automatic and very robust algorithm to apply in industrial applications of NMR spectroscopy, especially those characterized by a very high sample throughput and/or a continuous or pseudo-continuous sampling characterizes, for example, a process-control setup.

Here we present recent advances in a novel iterative algorithm in which the 'quality function' is based on the amplitude histograms (real and imaginary) of the spectrum. Apart from the exploitation of some convenient features of histograms, the fully automatic algorithm handles BOTH corrections simultaneously (PcBc rather than Pc + Bc), and it applies the baseline correction to BOTH the real and the imaginary parts of the spectrum.

In summary, PcBc permits us to:

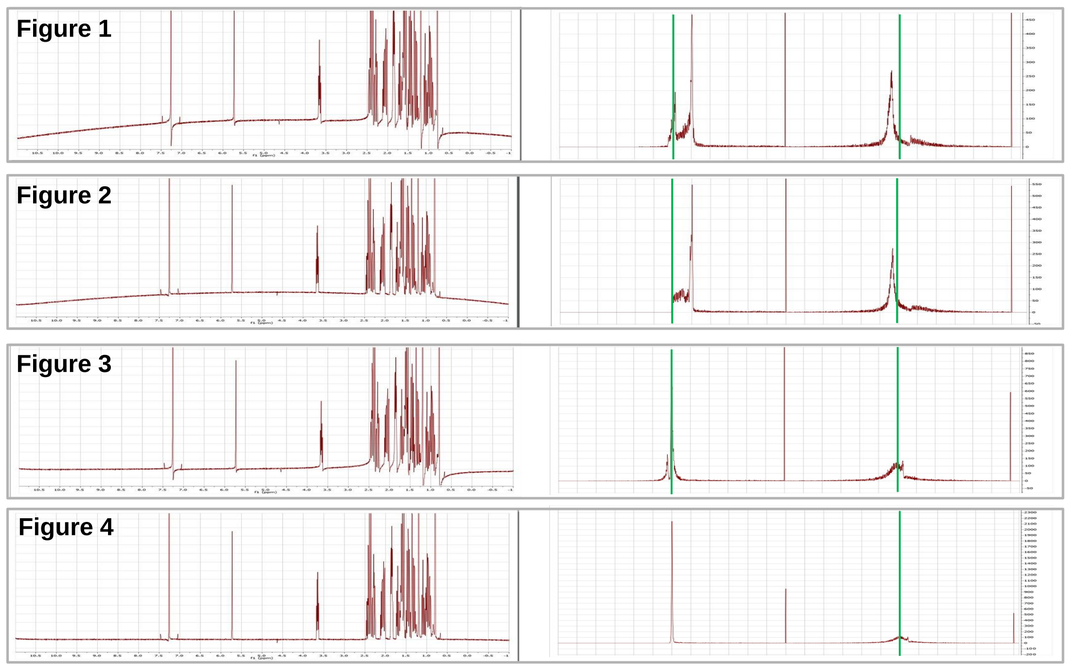
- Carry out the phase and baseline corrections *simultaneously;*
- Carry out both corrections (not just the phase) on *both* the in-phase and the out-of-phase parts of a spectrum;
- Enhance the *objectivity* of the corrections, especially considering that in practice one often encounters situations with multiple and/or ill-defined acceptable 'solutions'. Moreover, the manual corrections in current use (especially the phasing procedure) often differ depending of the personal habits of each spectroscopist.

Pilot idea: the histogram of the spectrum

The histogram H(ζ) is a diagram of data points counts when binned according to their height; we keep the size of the bins normalized so that there are about 5000 of them per Rms(h) average of the heights of all data points (the horizontal scales of all histogram plots appearing in this poster are normalized so that $\zeta = \pm 1.0$ corresponds to the height $\pm$ Rms(h).

When an NMR spectrum is properly conditioned, its absorption mode (real) part contains flat baseline stretches of noise with no peaks. In terms of histogram of such a data set this implies a sharp peak in the neighborhood of zero. Comparing the histograms of an experimental spectrum which is badly conditioned (Fig.1) with those of the same spectrum manually properly phased (Fig.2), or baseline corrected (Fig.3), or with both corrections (Fig.4), we see how their real parts tend to progressively peak around zero. In the final, properly conditioned spectrum, the histogram has practically no intensity at negative co-ordinates which exceed the noise, and just a small-intensity 'tail' at positive co-ordinates.

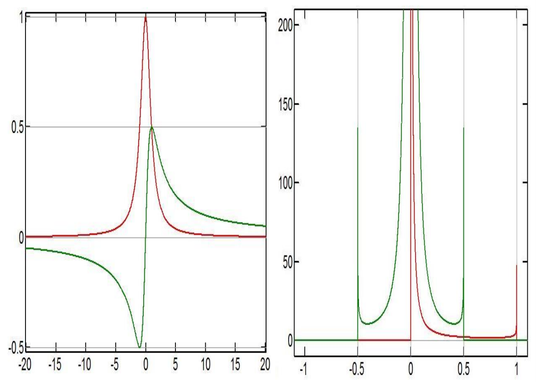
Observing the histograms of the imaginary part of a properly conditioned spectrum, it is possible to notice that it is symmetric with respect to zero and much smaller in maximum amplitude than the histogram of the real part.



Example of histograms:

Left column shows the spectra, while the dual right column shows the corresponding histograms of the real (left side) and imaginary (right side) part of the spectra. The zero position is highlighted by vertical green lines. For more information, see the text on the left.

From this point of view, it is interesting to compare this features with the ones of the histogram of an ideal Lorentzian line as shown on the right (with red traces showing the real parts and green traces the imaginary parts).
The upper figure shows a normalized Lorentzian peak, while the other one shows the respective theoretical histogram.
Note that the sharp asymptotes in the histogram become sharp peaks once an experimental noise is added (convolution with the noise probability function).



Modelling the baseline

For the fitting, apart from the two classical phase parameters ph0 (constant) and ph1 (linear) that need to be adjusted, we model the baseline correction by means of a linear combination of a pre-defined number (N) of low-frequency harmonic functions or, alternatively, low-indexed Chebyshev polynomials. In either case, this introduces 2*N fittable coefficients, because the baseline corrections for the real and the imaginary parts are considered totally independent.

'Anchoring' the baseline

Two points of the baseline, one close to the beginning and another close to the end, are forced to be zero. This decreases by 2 the dimensionality of the parameters space and helps the algorithm to ensure a flatter baseline. The 'anchors' are actually averages of 16 data points, strategically positioned to avoid the common smiley artifacts at the extremes of a spectrum.

The quality function

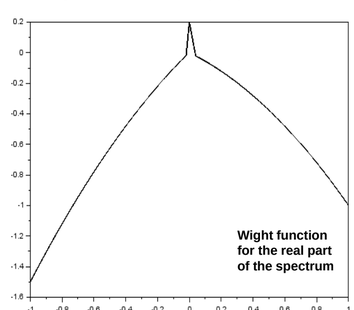
One then needs to construct a quality function Q(p) of all the 2*N+2 fittable parameters p which, when maximized, forces the histogram to possess the principle characteristics corresponding to well-conditioned spectra. We define Q(p) as *the integral of a properly weighted histogram H(ζ) of a spectrum corrected using the current parameters* p. Actually, two such integrals are computed, $Q_r(p)$ for the real part of the spectrum and $Q_i(p)$ for the imaginary part, with two different weight functions, $w_r(\zeta)$ and $w_i(\zeta)$, respectively, and then added together, $Q(p) = Q_r(p) + Q_i(p)$. The fact that this simple scheme is based on an integral over the histogram (and therefore over all experimental data points) is very important because it represents, as an extra bonus, a concurrent noise-suppressing filter.

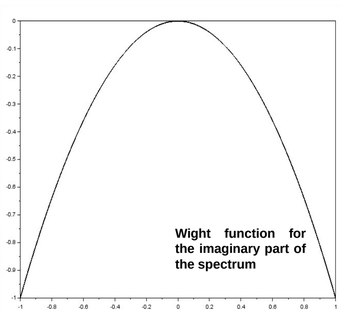
The weight functions

Modelling efficient weight functions suitable for our purposes is not an easy task because there is no rigorous mathematical theory which might help.

For the real part, it is evident that the area around $\zeta = 0$ should have the largest height, any negative values should be discouraged (negative penalty), and discouraged more than positive values for which the adjusted histograms still maintain some intensity, albeit small (see the figure on the left).

For the imaginary part, we are using a negative symmetric parabolic function, making sure that for perfectly symmetric histograms the result is zero, while any a-symmetric shape generates a negative value (see the figure on the right).





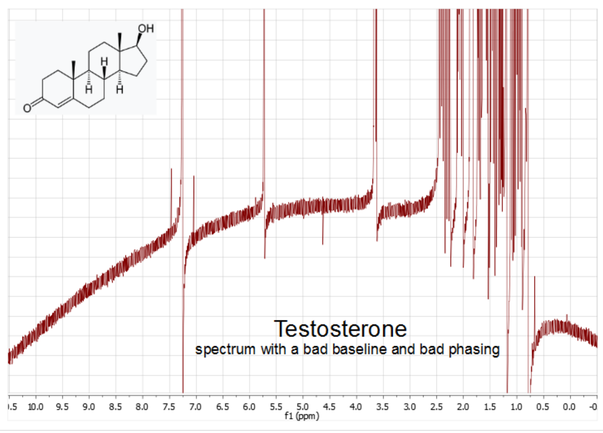
The Flowchart

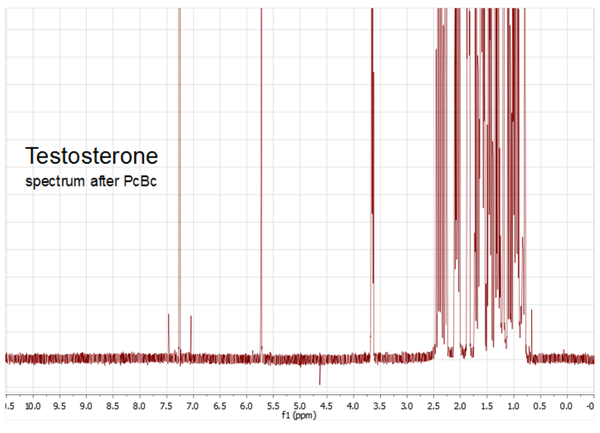
Hence, in extreme synthesis, the flowchart to compute the quality function Q(p) consists of these simple steps:

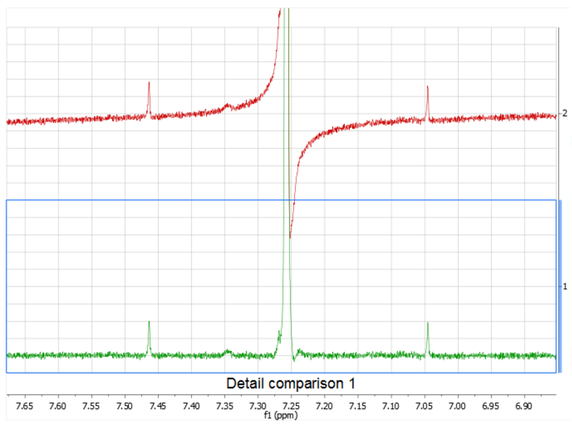
1. Take the experimental spectrum;
1. Apply Pc + Bc corrections defined by the parameters p;
2. For both Qr(p) and Qi(p): compute the histogram of the resulting spectrum;
3. For both Qr(p) and Qi(p): evaluate the weighted integral of the histogram.

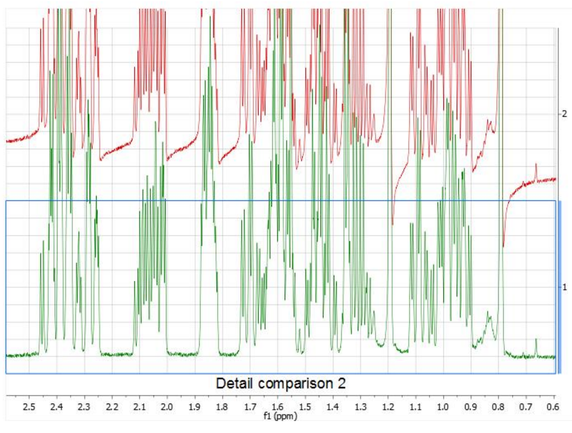
The values of the parameters p are modified in order to minimize Q(p) = Qr(p) + Qi(p) according to the down hill simplex algorithm.

Results using 5 harmonic coefficients for the baseline correction









References

1. Cobas C., Bernstein M.A, Martìn-Pastor M., Tahoces P.G., 2006. A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. Journal of Magnetic Resonance 183, 145-151;
2. Binczyk F., Tarnawski R., Polanska J., 2015. Strategies for optimizing the phase correction algorithms in Nuclear Magnetic Resonance spectroscopy. Biomedical Engineering On line 14 (S2);
3. Cobas C. and Sykora S., 2007. Extended DISPersion-Absorption (eDISPA) approach to Automatic Phasing of HR-NMR Spectra. 48th ENC Conference, 2007, DOI: 10.3247/SL2Nmr07.004.
4. Zorin V., Bernstein M.A. and Cobas C., 2017. a robust, general automatic phase correction algorithm for high-resolution NMR data, Magnetic Resonance in Chemistry 55, 738-746.

UNIÓN EUROPEA
Fondo Europeo de Desenvolvemento Rexional

XUNTA DE GALICIA
CONSELLERÍA DE INNOVACIÓN, INDUSTRIA E COMERCIO
Dirección Xeral de Investigación e Desenvolvemento

In collaboration with www.extrabyte.eu