Novel Data Evaluation Algorithms: The 1D and 2D Resolution Booster™ (RB)

Carlos Cobas [1], Nikolay Larin [1], Isaac Iglesias[1], Felipe Seoane[1], Stanislav Sýkora [2]

[1] Mestrelab Research, Xosé Pasín 6, Santiago de Compostela, 15706 Spain; carlos@mestrec.com
[2] Extra Byte, Via Raffaello Sanzio 22/C, Castano Primo (Mi), Italy I-20022; sykora@ebyte.it

Introduction

Resolution is a key concept in high resolution NMR spectroscopy and considerable hardware and software efforts have been devoted in the past to optimize it. It is certainly not necessary to list here all aspects of NMR spectroscopy which benefit from even a modest increase in resolution.

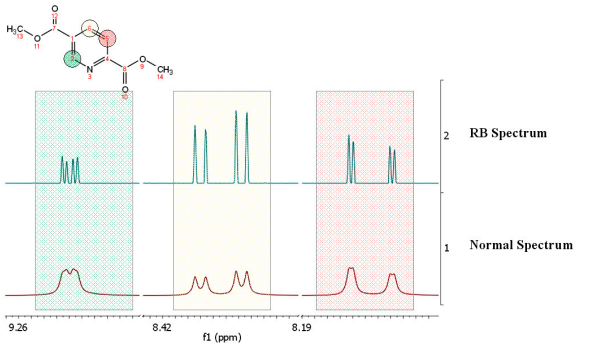
In this work we will propose a new data processing algorithm, the **Resolution Booster™** (**RB**) which:

Provides a **resolution enhancement** exceeding than of any existing method

1. Can be cast into a **completely automatic**, **black-box** form
2. Has only a **modest effect on experimental noise**
3. For purposes such as peak recognition, it is convenient to **discard the noise** altogether
4. Can **discern hidden peaks** about as well, if not better, as a trained human eye
5. Implicitly **corrects line-position errors** due to line overlap in spectral multiplets
6. Automatically **removes any baseline drift**

RB at a glance: Improved elucidation of coupling patterns

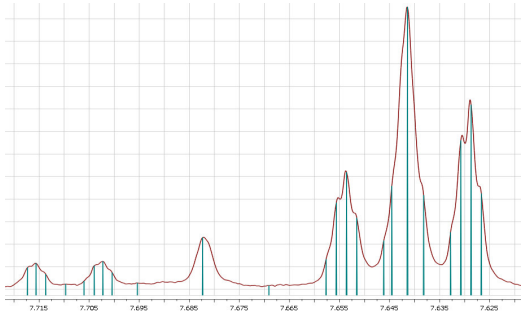
The following Figure shows the results of applying RB to the ^{1}H spectrum of *dimethyl pyridine-2,5-dicarboxylate* acquired at 250 MHz. In the untreated spectrum (lower trace), the long range coupling $^5J_{5\text{-}2}$ cannot be estimated because of the lack of resolution.



Applying RB to the same data (upper trace), spectral resolution is increased by over 200%, thus making it possible to estimate the value of the long range coupling $^5J_{5\text{-}2}$, equal to 0.82 Hz.

RB at a glance: Improved recognition of spectral lines

The Figure below shows a very small fraction of a **metabolomic spectrum** to which we have applied an automated RB procedure and, subsequently, a peak-picking routine. In this way, over 900 peaks were identified in the whole spectrum and the positions of their centers were marked, all in a single command (to simplify the visualization, we show only the peak marks here, not the whole RB peaks).



Notice that, thanks to RB, **barely resolved shoulders are correctly identified** as peaks. To human eye, some of the peak markers appear slightly off-center but we will explain below that the RB line center estimates are actually remarkably correct.

The Math behind RB

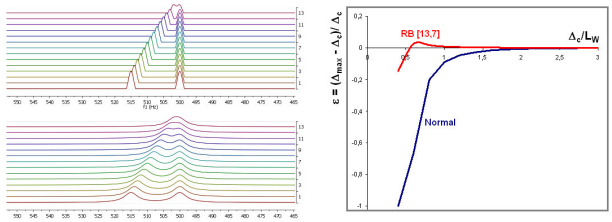
Considering the performance of **RB**, the underlying math is relatively simple. It is based on the recognition that negative 2nd derivative of a Lorentzian (as well as Gaussian) spectral line has a single, narrow positive peak and broader negative "lobes". **RB** is the result of combining a suitable algorithm for the extraction of the 2nd derivative of a noisy data set (in this case the Savitsky-Golay algorithm) with a non-linear cut-off of the negative peaks.

Those who have some experience with this kind of numeric routines probably recognize that this is much easier said than done, especially if all the involved parameters are to be selected and optimized automatically by the algorithm itself. One must set properly the Savitzky-Golay parameters (the order and the number of points in the convolution pattern) and the noise cut-off threshold. This implies an automatic estimation of the mean linewidth in terms of digital spectral steps and a reliable estimation of the noise level both before and after the Savitsky-Golay convolution.

We have overcome these obstacles by means of independent innovative algorithms (to be published).

Resolving Power

This Graph illustrates an important feature of the **Resolution Booster™** which is its capability to correct coalescence errors in the positions of overlapping lines. In the simplest case, when two equally intense lines of width L_w are placed so that the distance between their centers is Δ_c, the two maxima (as long as they are resolved) appear separated by a distance $\Delta_m < \Delta_c$. Plotting the relative apparent separation error $(\Delta_m-\Delta_c)/\Delta_c$ against the ratio Δ_c/L_w, one obtains the blue error curve shown in the graph. A similar error curve can be drawn for any resolution enhancement method and, subsequently, the curves can be used to objectively compare different methods. In this example we have chosen to carry out such a comparison empirically for the specific case of an RB algorithm using 7th order Savitsky-Golay 2nd derivative filter operating on 13 points (a typical choice). We have simulated a series of 13 spectra, each containing two lines with a fixed width of 5 Hz but with different separations between their true centers. We have then measured their apparent separation (that between their maxima) both in the untreated spectra (bottom set, blue line) and in the spectra passed through the RB (upper set, red line).

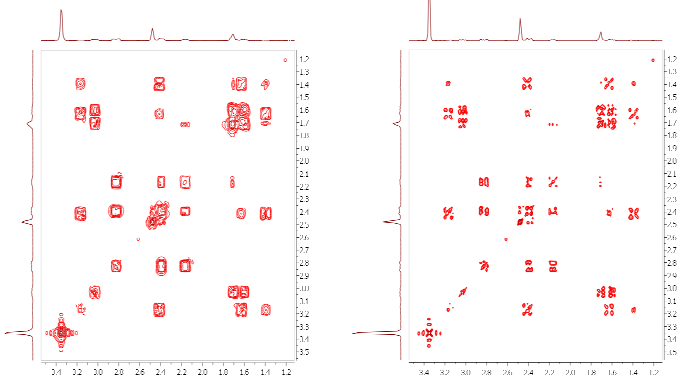


Clearly, the RB has the effect of efficiently correcting the coalescence error and shifting the coalescence point where the two lines are no longer resolved (intercept of the curve with the $\varepsilon = -1$ line) to much lower values of Δ_c/L_w. This behaviour quite general also in complex multiplets. Moreover, it is not observed in alternative resolution-enhancement methods such as those based on Lorentz-Gauss FID apodization.

2D Resolution Booster

Any 1D resolution enhancement method can be easily extended to two of more dimensions by applying it separately to each row and each column. Actually, the RB algorithm offers a bit more sophisticated possibilities with genuine 2D extensions, but this is not the place to open this topic.

To show at least one practical example of the application of RB to a 2D spectrum, we have chosen a very simple case – that of the COSY spectrum of Qu



Here, the original spectrum is on the left, while the one treated with RB is on the right. The enhanced resolution is evident and remarkable.

Conclusions

The limited space of a poster does not permit a sufficiently detailed analysis of all the features of a sophisticated data processing technique such as the **Resolution Booster™**. In particular, we have omitted to discuss its surprising tolerance to experimental noise. Overall, the behavior of the **Resolution Booster™** is very satisfactory in the sense that it enhances the resolution as much as possible but not beyond, while keeping experimental noise in check (the latter two points are anything but trivial). It operates directly on frequency domain data and can be easily applied either to a complete spectrum or to an arbitrary cutoff thereof. It is likely that it will play an important role in:

- **Processing of 2D and 3D spectra**
- **Verification and elucidation of molecular structures** with large, coupled spin systems
- **NMR applications in metabolomics**
- **Proton NMR of proteins** and other **natural macromolecules** with overcrowded spectra
- Processing of data sets originating in **techniques other than NMR** (MS, HPLC, etc)

Acknowledgments

Financial Support from the Xunta de Galicia, Spanish Ministry of Education (Torres Quevedo program) and the European Social Fund are greatly acknowledged.

DOI: 10.3247/sl2nmr08.008