PcBc - A NOVEL AUTOMATIC PHASE & BASELINE CORRECTION ALGORITHM

A. Ester Maria Vasini [1, 2], B. Stanislav Sykora [2], C. Carlos Cobas [3]

[1] DICAM, University of Bologna, Italy

[2] Extra Byte, Castano Primo, Italy

[3] Mestrelab Research, Santiago de Compostela, Spain

E-mail: ester@ebyte.it

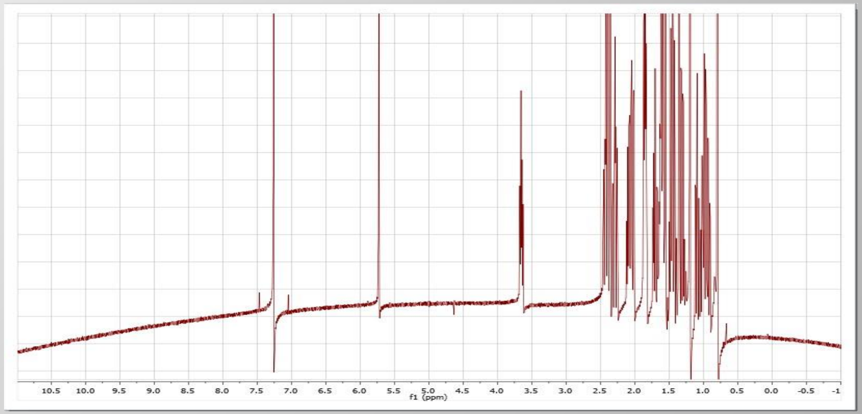
DOI: 10.3247/SL6Nmr17.005

Contents

Introduction

The **algorithmic problem** of **phase correction** (Pc), and that of baseline correction (Bc), of 1D NMR spectra have been both tackled many times by many authors over the last half a century.

- There are many algorithms which emulate manual procedures:
  - optimization (fit) of the parameters which describe either the phase (ph0, ph1) or the baseline (according to various parametric models);
  - Maximization of some quality assessment of the corrected spectrum.
- Historically, the employed quality functions included features such as:
  - peak heights,
  - negative peak lobes,
  - symmetry,
  - DISPA patterns,
  - selected baseline points,
  - peak ablation,
  - and others.



Introduction

A different type of "quality function" Q to be optimized, based on the histograms of the spectrum (real and imaginary parts) which turn out to be very sensitive to both phase and baseline distortions.

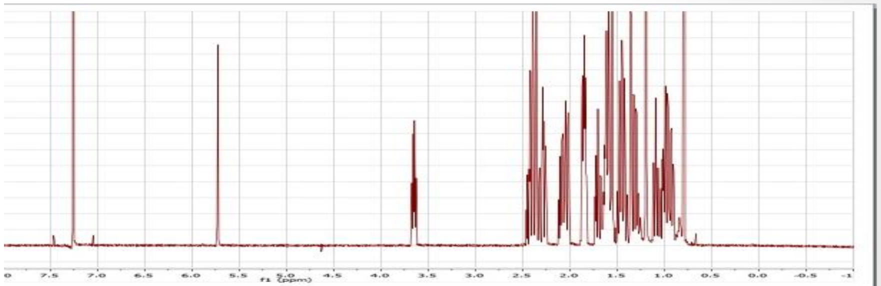
This permits us to:

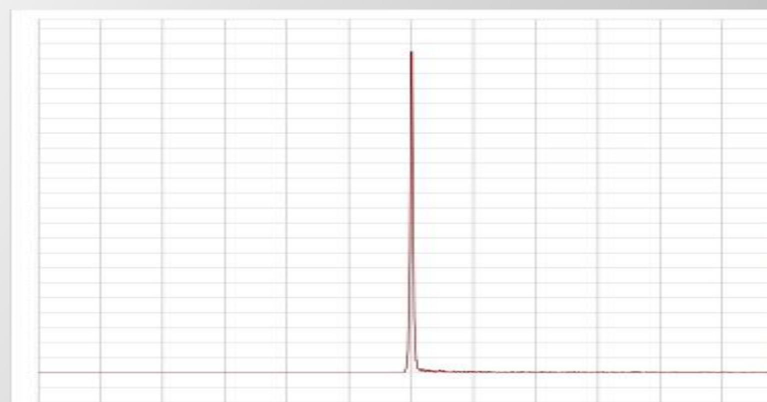
- Carry out the phase and baseline corrections **simultaneously**;
- Carry out the corrections (not just phase) on **both** the **in-phase** and the **out-of-phase** parts of a spectrum;
- Enhance the **objectivity** of the corrections;

PcBc is a novel algorithm developed for the stated purpose.

Pilot idea:
the histogram of the spectrum

When an NMR spectrum is well conditioned, its absorption mode (real) part contains flat baseline stretches of noise with no peaks.

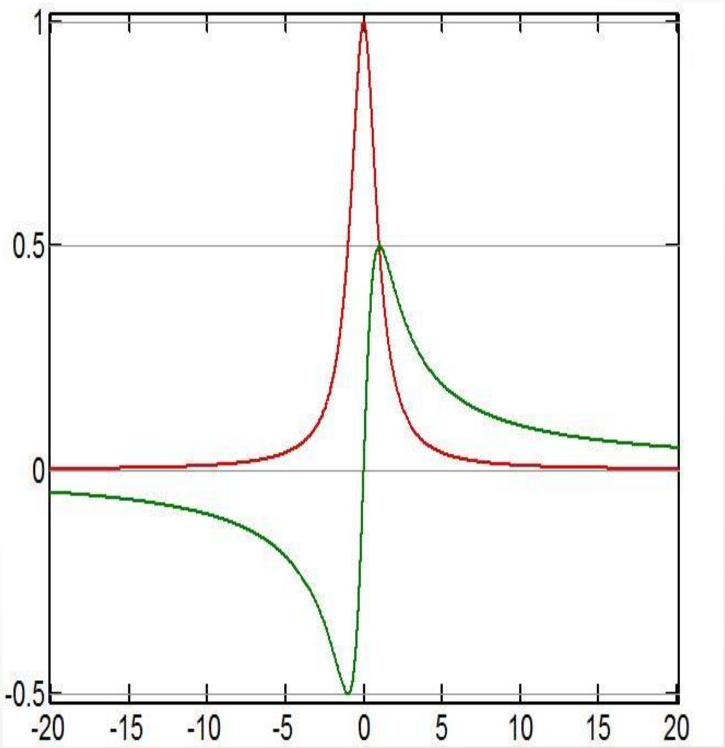


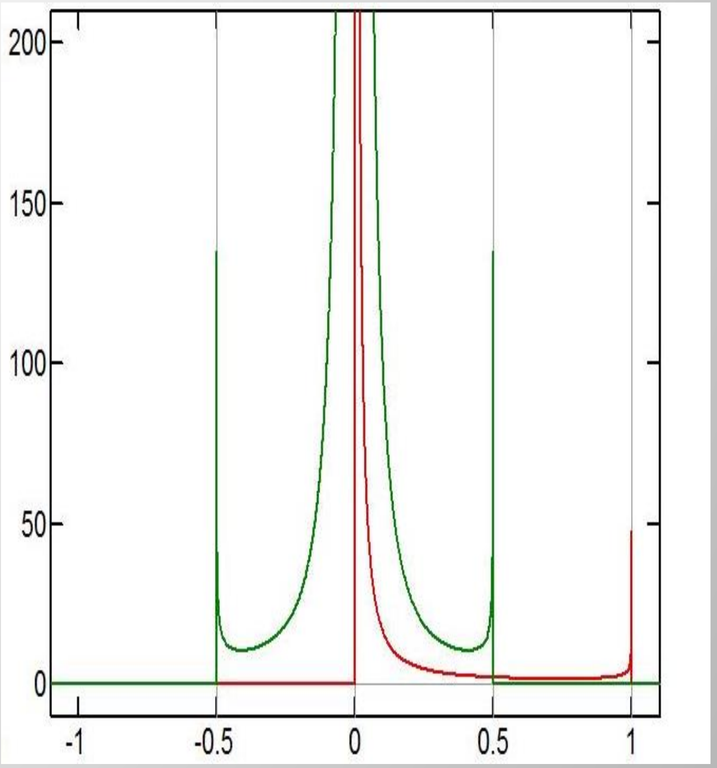


- In a histogram of such a data set this implies a sharp peak in the neighborhood of zero

(note: histogram H(z) 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 z = ±1.0 corresponds to the height ± Rms(h).

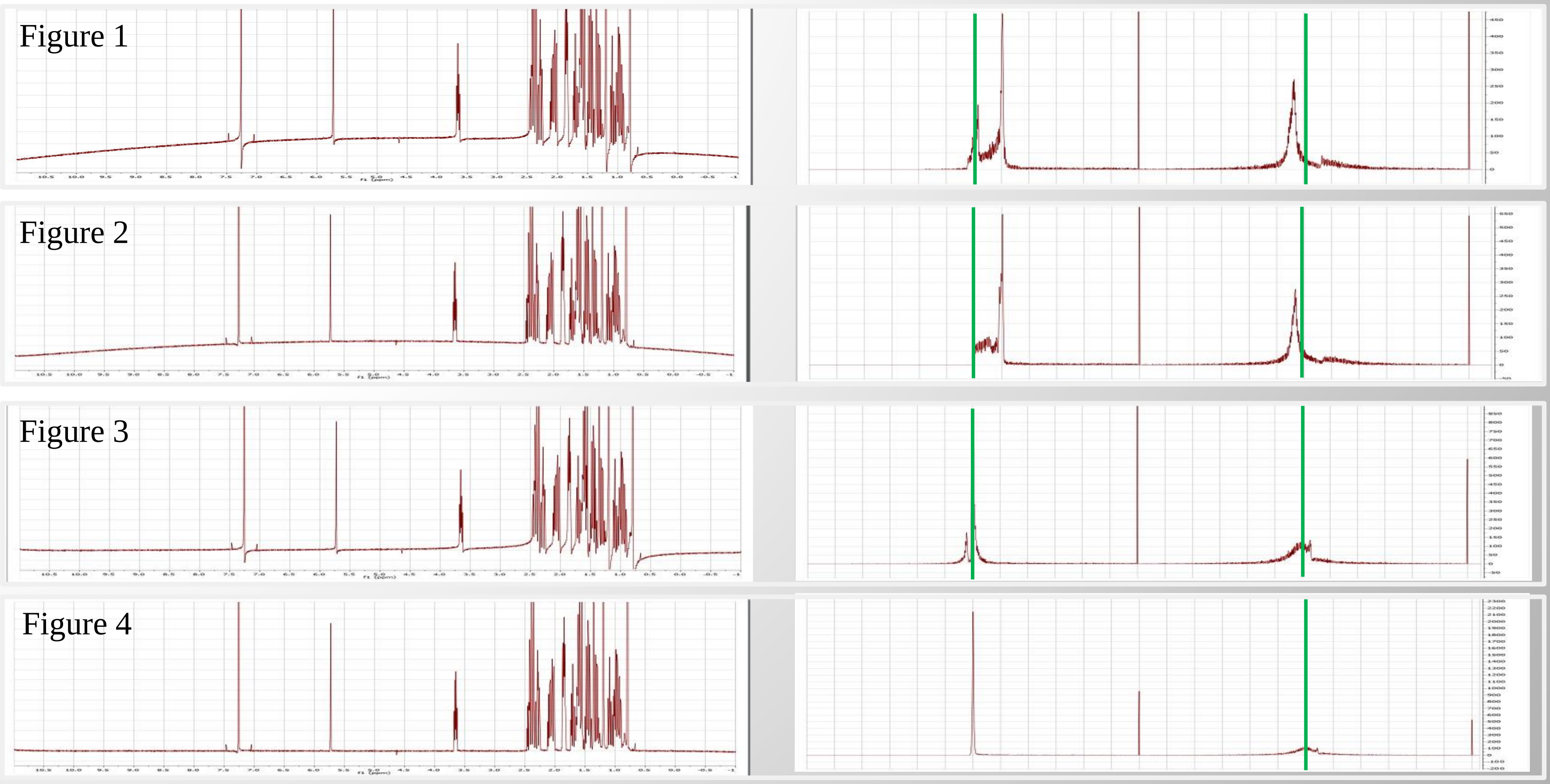
Pilot idea:
the histogram of the spectrum





- It is also interesting to compare this with the histogram of an ideal Lorentzian line as shown on the left (with red traces showing the real parts and green traces the imaginary parts. The left plot 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).

Pilot idea:
the histogram of the spectrum



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.

Pilot idea:
the histogram of the spectrum

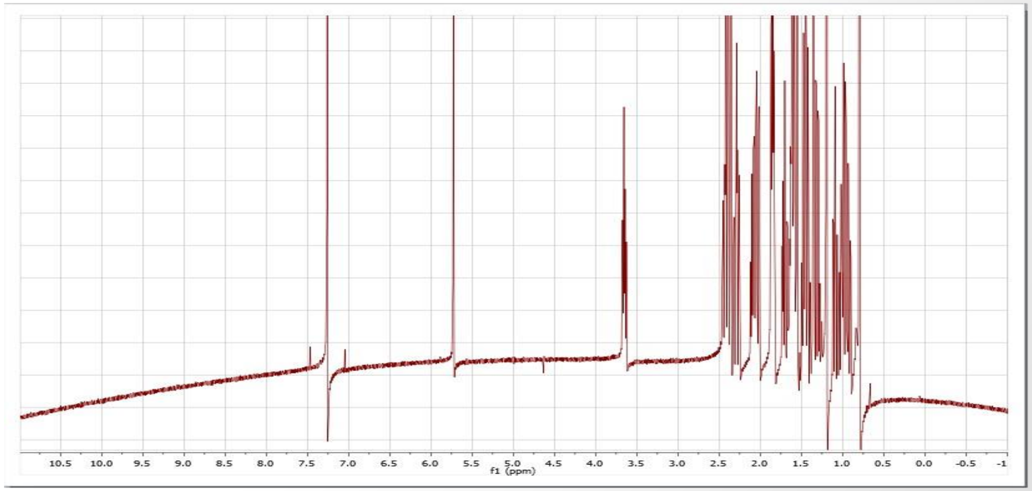
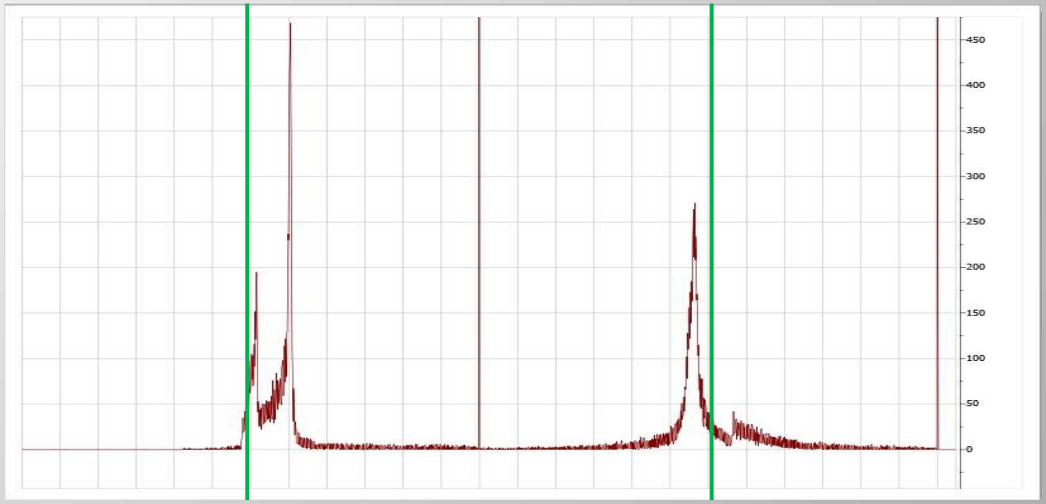


Figure 1:
Bad conditioned spectrum (up-sx) and the corresponding histogram (down-dx) of the real (left side) and imaginary (right side) part.



Pilot idea:
the histogram of the spectrum

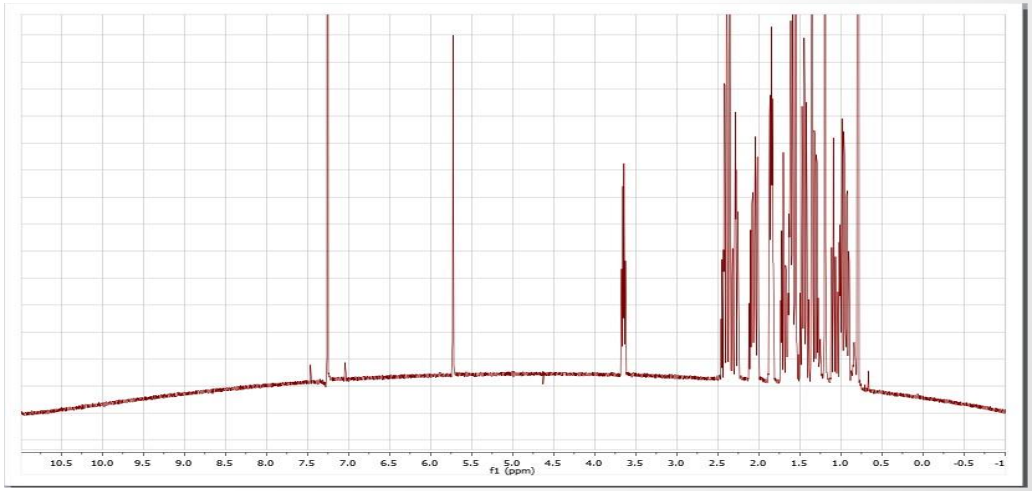
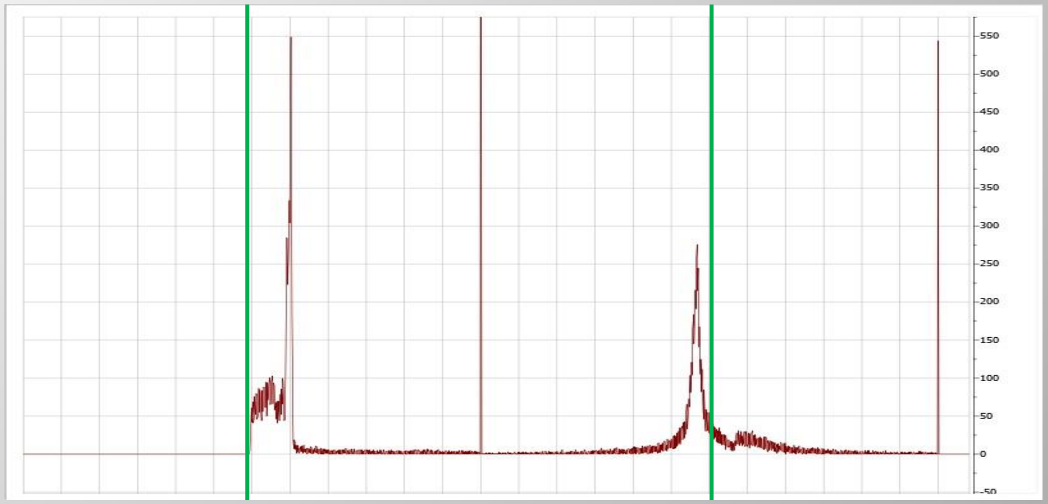


Figure 2:
Manually phase corrected spectrum (up-sx) and the corresponding histogram (down-dx) of the real (left side) and imaginary (right side) part.



Pilot idea:
the histogram of the spectrum

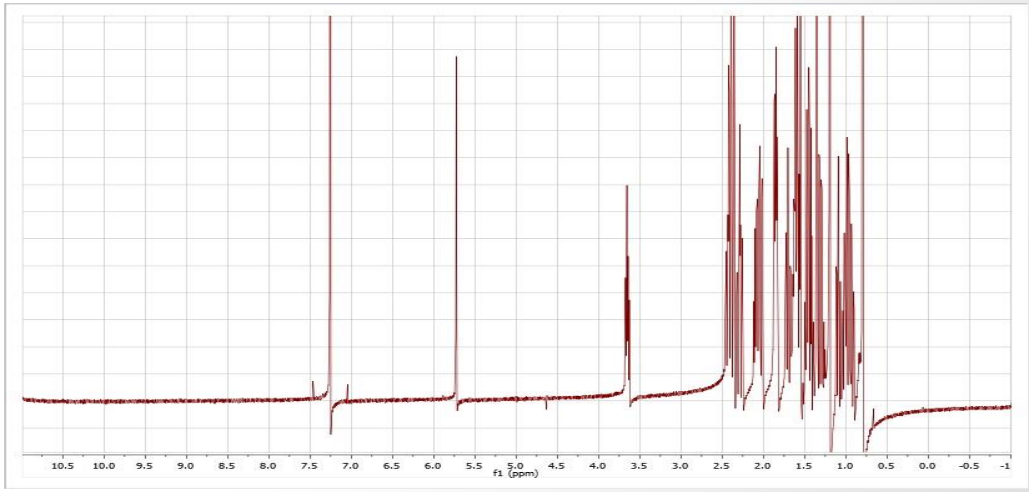
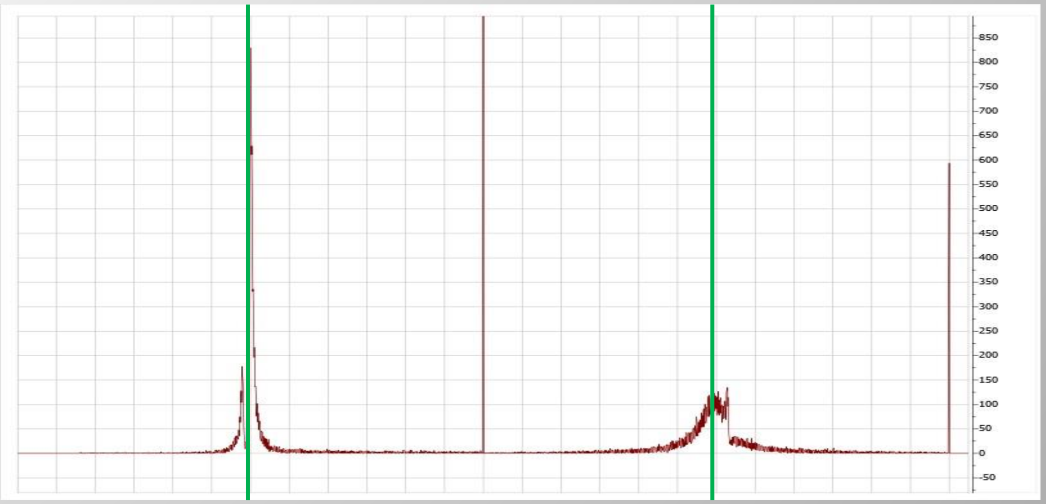


Figure 3:
Baseline corrected spectrum (up-sx) and the corresponding histogram (down-dx) of the real (left side) and imaginary (right side) part.



Pilot idea:
the histogram of the spectrum

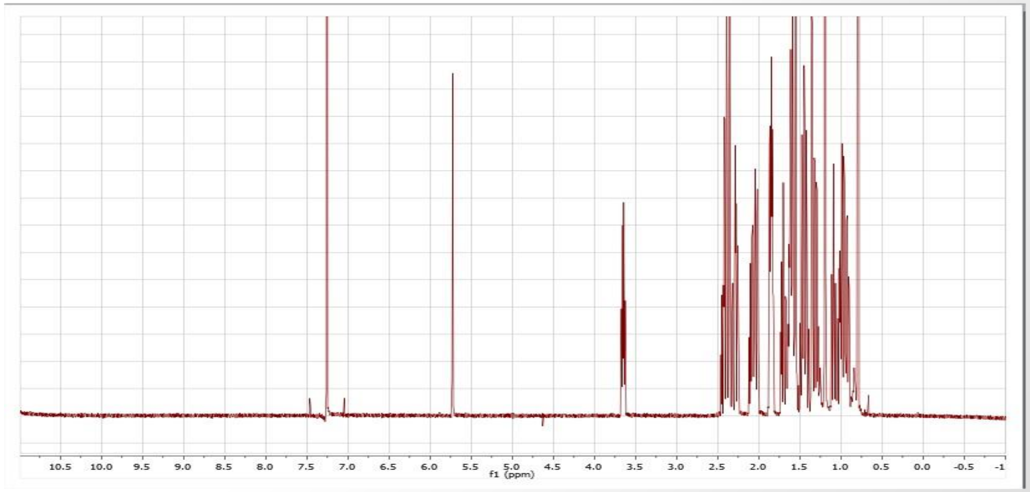
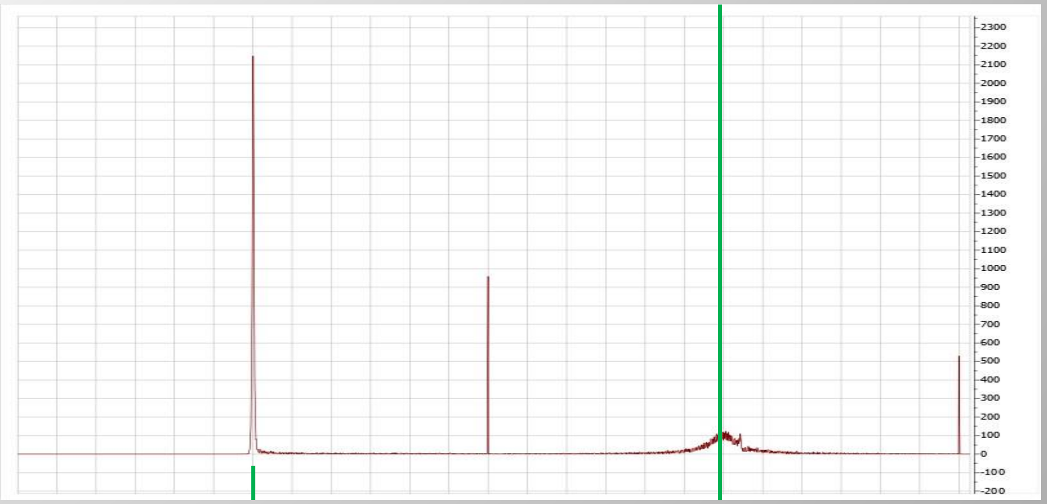


Figure 4:
Phase and Baseline corrected spectrum (up-sx) and the corresponding histogram (down-dx) of the real (left side) and imaginary (right side) part.



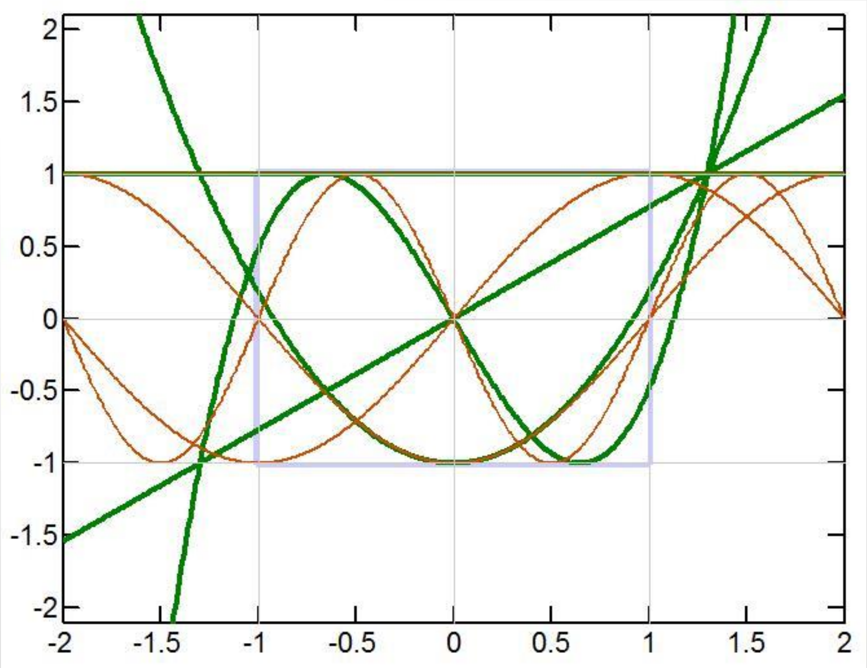
Modelling the baseline

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.

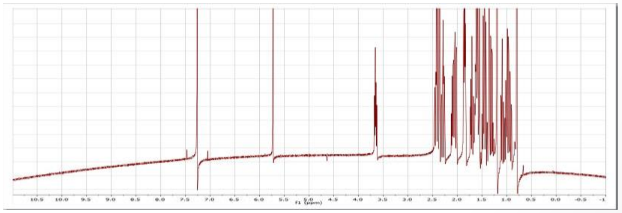


→ 2*N fittable coefficients,

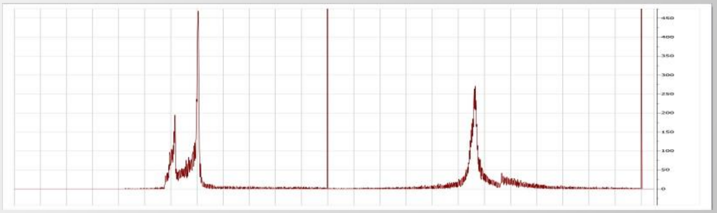
because the baseline corrections for the real and the imaginary parts are considered independent.

The quality function

Original spectrum



Original spectrum histogram

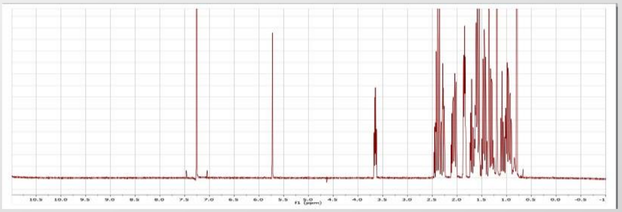


Q(p) → integral of a properly weighted histogram of a spectrum corrected using the current parameters p (2*N+2 parameters)

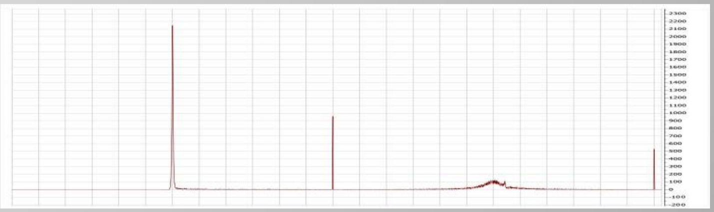
(integral over the histogram
→ concurrent noise suppressing filter)

$$Q(p) = Q_r(p) + Q_i(p)$$

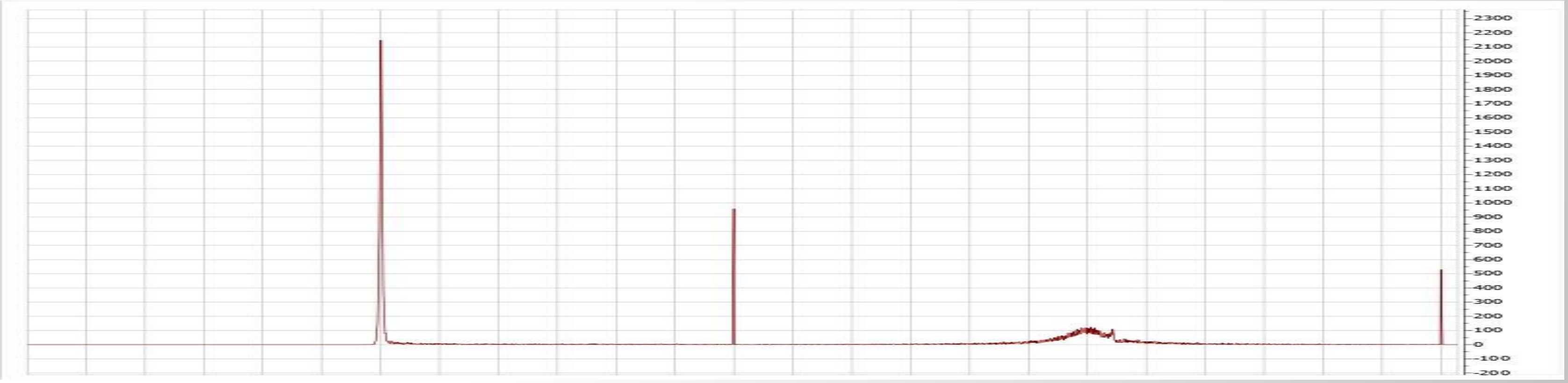
PcBc processed spectrum



PcBc processed spectrum histogram



The weight function

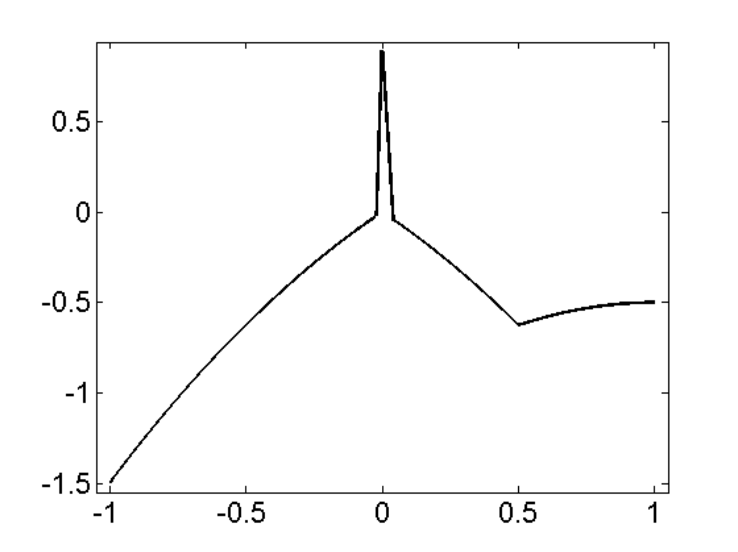


Imaginary part:

- Check for asymmetries;

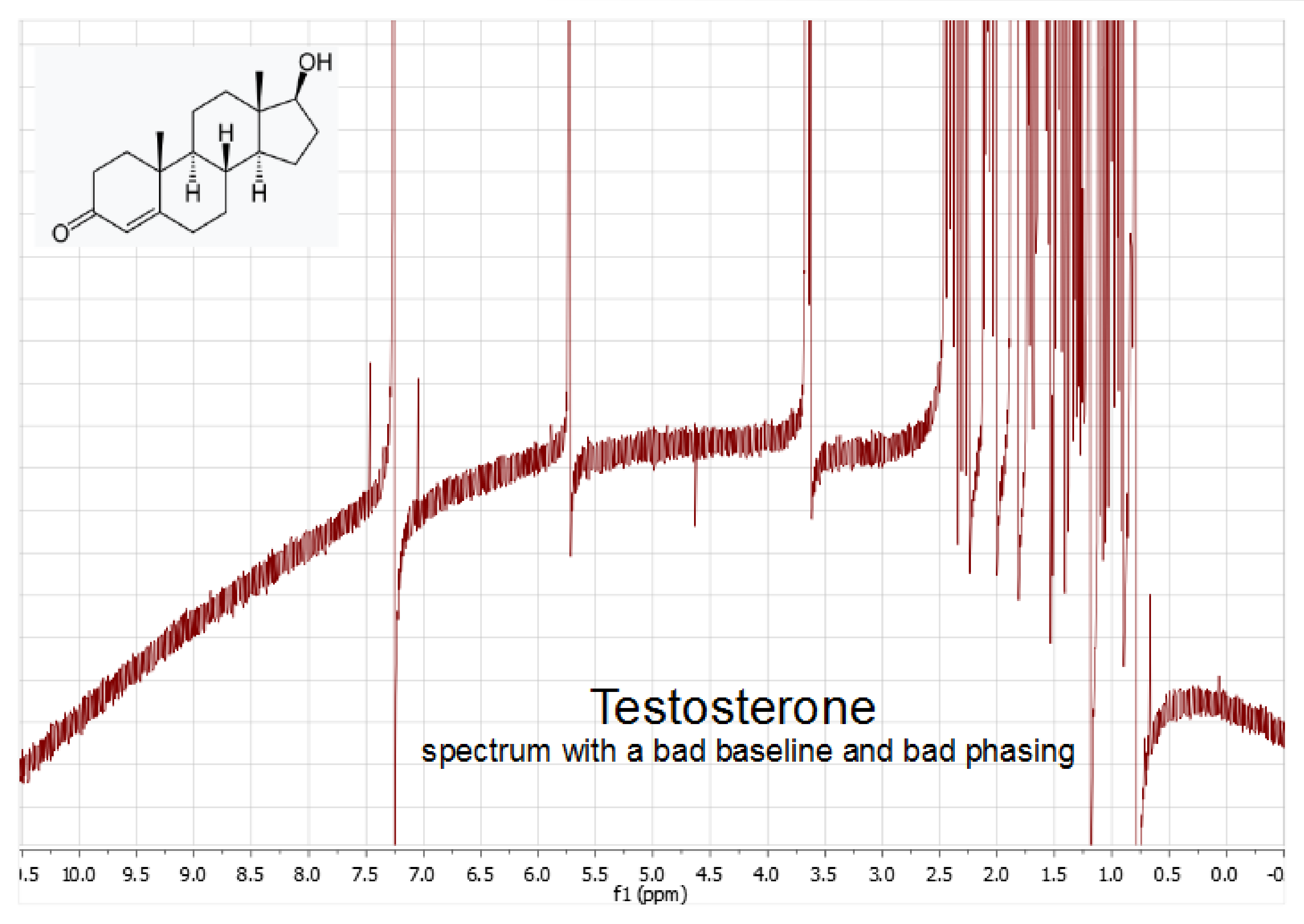
Real Part:

- largest height for the area around z =0;
- Discouraging any negative values (negative penalty);
- and the negative values are discouraged more than positive values for which the adjusted histograms still maintain some intensity, albeit small;

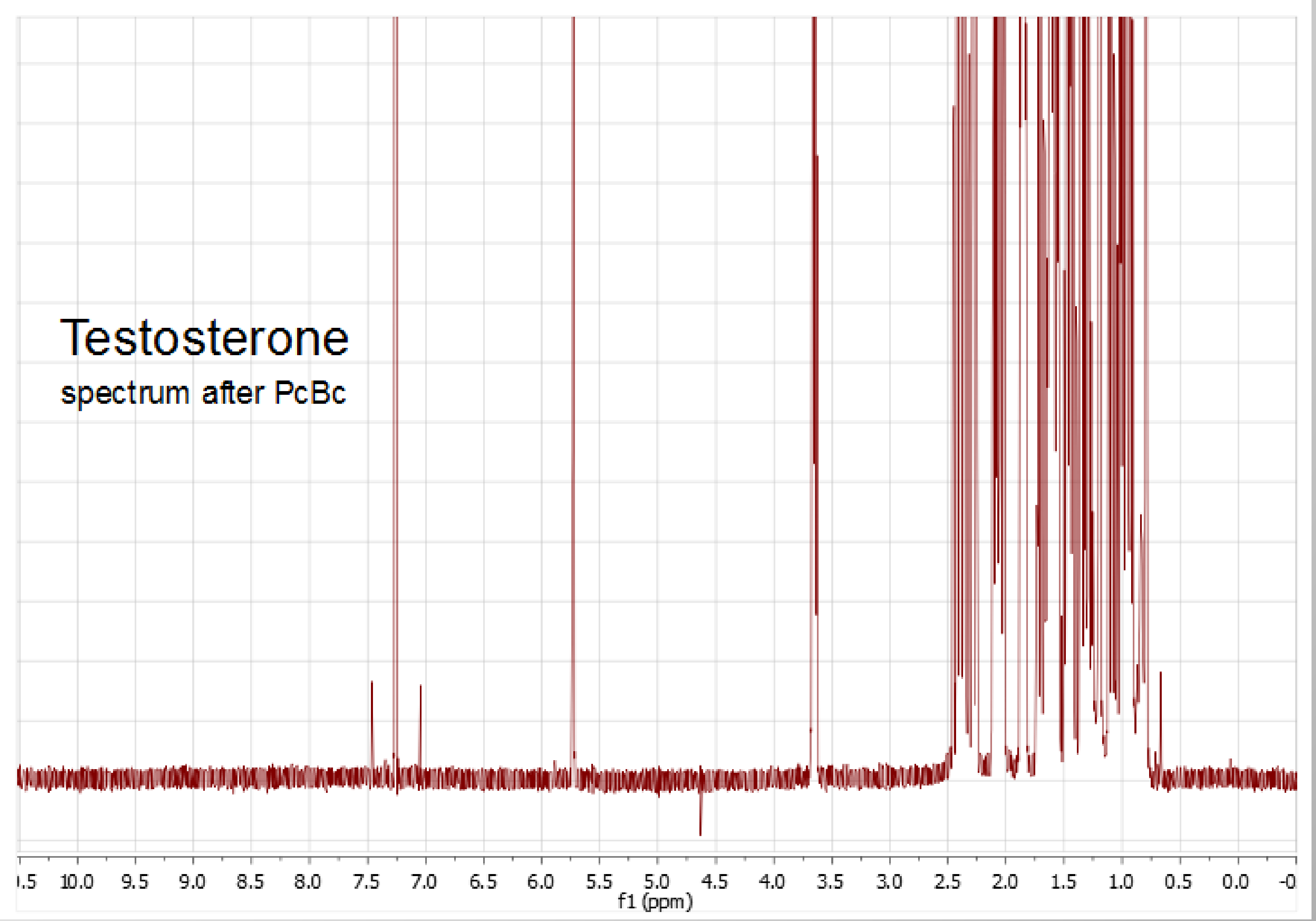


Up to now the weight function is characterized by **seven parameters**.

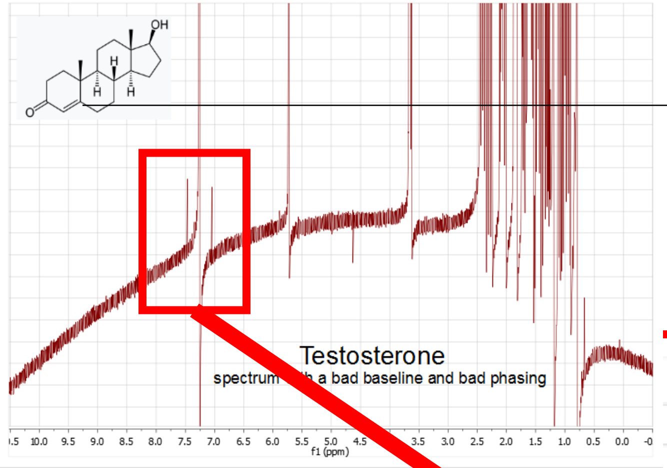
Result

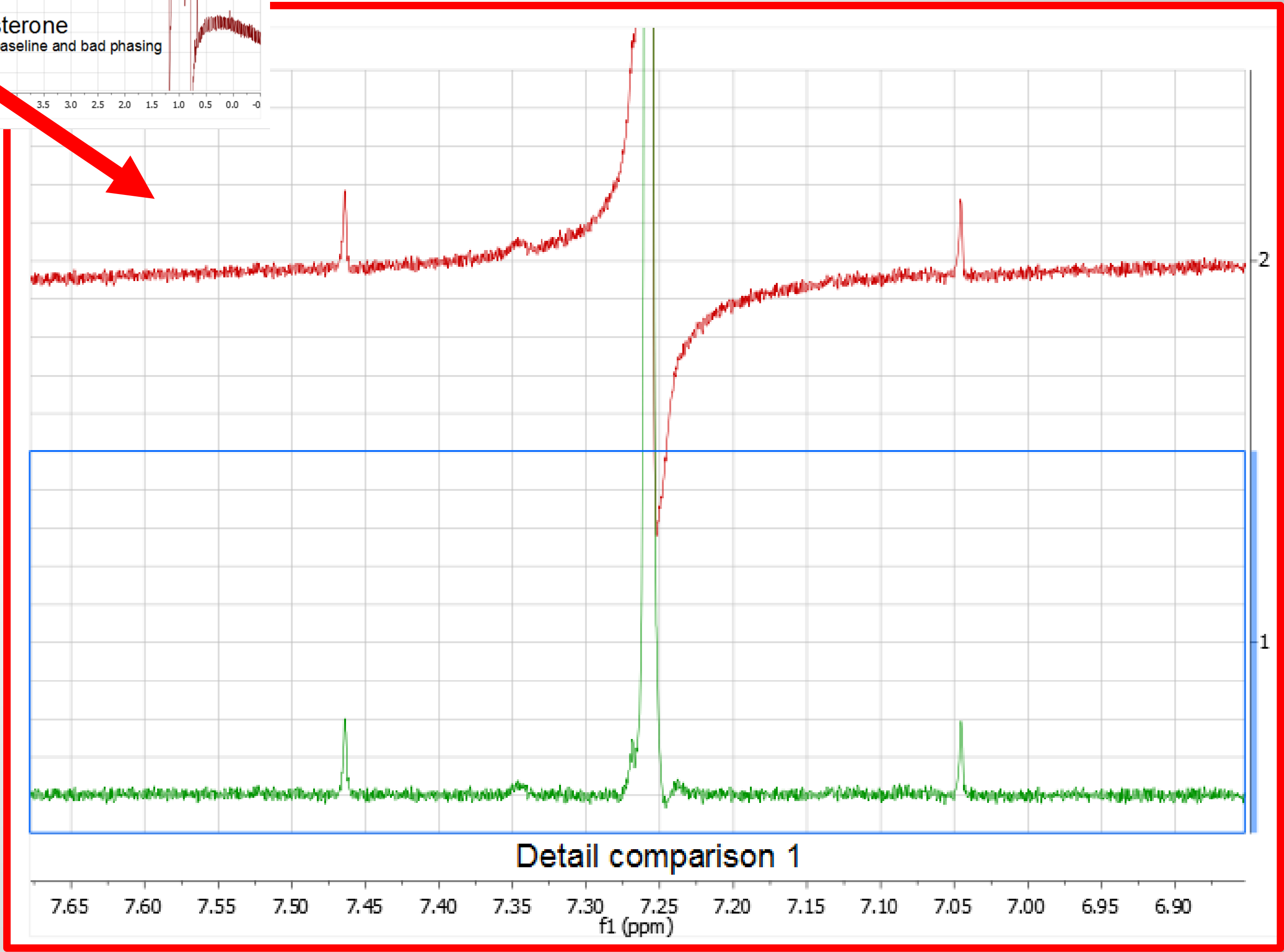

OH
H
H
H
O
Testosterone
spectrum with a bad baseline and bad phasing
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

Result


Testosterone
spectrum after PcBc
.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0
f1 (ppm)

Result


OH
Testosterone
spectrum with a bad baseline and bad phasing
f1 (ppm)


2
1
Detail comparison 1
7.65 7.60 7.55 7.50 7.45 7.40 7.35 7.30 7.25 7.20 7.15 7.10 7.05 7.00 6.95 6.90
f1 (ppm)

Conclusions

- The results obtained are in good agreement with the expectations;
- There is more work to do related to:
  - The weight function (calibration of its parameters);
  - The histogram of the imaginary part of the spectrum;
  - The optimization algorithm;
- Test of the procedure on a large sample of spectra.

Thanks for the attention!

Contact info:

Ester Maria Vasini

ester@ebyte.it

estermaria.vasini2@unibo.it

Stanislav Sykora

sykora@ebyte.it

Carlos Cobas

carlos@mestrelab.com