Application of multiplet structure deconvolution to extract scalar coupling constants from 1D NMR spectra

Damien Jeannerat[1,2], Carlos Cobas[2]

[1]NMRprocess.ch, Geneva, 1200, Switzerland
5   [2]Mestrelab Research S.L. Santiago de Compostela, A Coruña, 15706 Spain

Correspondence to: Damien Jeannerat (damien.jeannerat@protonmail.com)

Abstract. Multiplet structure deconvolution provides a robust method to determine the values of the coupling constants in first-order 1D NMR spectra. Functions simplifying the coupling structure for partner with spin larger than 1/2 and for doublet with unequal amplitudes were introduced. The chemical shifts of the coupling partners causing mild second-order effects can,
10   in favourable cases, be calculated from the slopes measured in doublet structures. Illustrations demonstrate that deconvolution can straightforwardly analyse multiplet posing difficulties to humans and, in some cases, extract coupling constants from unresolved multiplets.

1 Introduction

When studying organic compounds, 1D ^{1}H NMR spectra are often the only, and sufficient, analytical method engaged. All
15   chemist known how to describe simple multiplet structures. It consists of determining their position (chemical shift), integral and, when possible, their coupling structure by identifying doublets (*d*), double doublets (*dd*), *etc.* Any serious analysis also includes the values of the scalar coupling constants. Indeed, the multiplet structures provide important topological and structural information on organic compounds and natural products. For example, the presence of a methyl group on a carbon bearing a proton will produce a quartet structure, the conformation of a double bond has clearly distinct geminal coupling
20   constants, dihedral angle influences vicinal coupling constants (Karplus, 1963), *etc*.
These NMR parameters are not only helping researchers to identify their products, but also allow reviewers to assess the validity of the argument supporting their identification and benefit the community by providing very precious information when structurally similar compounds are encountered. It is therefore of fundamental importance to provide chemists with the most powerful tools to analyse multiplets in particular in the cases where the multiplet structure is too complex to be deciphered
25   by visual inspection because of high degeneracy (complex structures such as "*dqd*"), partial overlap of multiplet structures or second-order effects.
The measurement of scalar coupling constants has been the subject of very intense academic work, either for direct application to 1D spectra (McIntyre and Freeman, 1992;del Río Portilla and Freeman, 1993) or involving the development of NMR pulse sequences producing spectra where coupling constants can be measured easily (Marquez et al., 2001) even in situations of
30   extensive overlap (Prasch et al., 1998;Kiraly et al., 2020;Berger, 2018). The sad observation is that the impact of these developments is extremely limited because the broad community of chemists almost exclusively rely on basic 1D ^{1}H spectra to extract coupling constants. Even if users of NMR were aware of them, they would be reluctant to use experiments going against the common practice in their fields except if they were producing key results - something NMR coupling constant alone, rarely do.
35   We shall report here on a significant improvement of the analysis of 1D ^{1}H spectra based on multiplet-structure deconvolution which will be introduced in a future release of the Mestrelab's Mnova software (Mnova NMR version 14.3.0, 2021). This method was originally developed during the '90s to automatize the analysis of multiplet structure from spectra generated by one of the forgotten pulse sequences: the soft-COSY experiment. (Emsley et al., 1990) Multiplet deconvolution shown to be applicable to the more "standard" DQF-COSY experiment (Jeannerat, 2000) after simply increasing the direct acquisition time
40   to provide higher resolution - something which was problematic because of the data storage available at that time, but is no

longer justifiable. We shall demonstrate the power of the application of multiplet deconvolution outside the world of 2D correlation spectroscopy, where positive and negative peaks coexist and cause specific challenges (Jeannerat and Bodenhausen, 1999) and apply it to the common – not to call it mundane – multiplet structure present in standard 1D NMR spectra.

1.1 Multiplet structure in 1D spectra

45   For weakly coupled spin systems, the multiplet structure observed in 1D spectra can be seen as the result of the combination (see Fig. 1) of the resonance frequency, the effects of the exponential relaxation, B_0-field inhomogeneity and scalar coupling (J) interactions. (Metz et al., 2000)
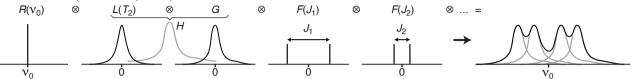

Figure 1: Multiplet structure of a 1D NMR spectrum expressed as the convolution product of the contributions of the Larmor
50   **frequency of the nucleus $R(v_0)$, the Lorentzian line shape $L(T_2)$, the B_0-field inhomogeneity $G(inhom.)$ and the scalar coupling constants $F(J_i)$.**

In the frequency domain, the $\otimes$ symbol stands for the "convolution product" which correspond, in the time domain, to pointwise multiplication. The detected FID is, indeed, the product of the chemical shift evolution, the exponential decay (producing a Lorentzian shape $L(v)$), a function reflecting the effect of B_0 inhomogeneity (often modelled as Gaussian shapes
55   in both domains) and cosine modulations caused by the first-order J-coupling interactions (resulting to doublet in the 1D spectrum).

In liquid state NMR and in isotropic media a doublet $F(J)$ is expressed, mathematically, by a pair of so-called δ functions, *i.e.* a function returning zero values everywhere except for the values of $v = -J/2$ and $+J/2$, where J is the scalar coupling constant. When taking the convolution product of any set of δ functions with any other function, typically a spectral lineshape
60   H, the position and amplitude of the δ functions can be understood as to indicate where the convoluted function H, should be duplicated before summation (see the grey components in the right part of Fig. 1).

The deconvolution consists in reverting the effect of the convolution of a given function. The symbols $\oslash F$ is sometimes used to represent the process reversing the effect of $\otimes F$. But what seems to be a simple division is misleading because, in reality, the $\otimes$ symbol stands for an integral which has no general inverse. One should, instead, look for an inverse of the F function
65   and replace the deconvolution operator with the convolution with an inverse function $\otimes 1/F(J)$. An alternative is to operate in the time domain but the simplicity of the division is compensated by the difficulty of dealing with division by zero.

Whether it is applied in the time or frequency domain, deconvolution is a difficult process prone to produce noise and artifacts. The low efficiency of image deblurring is a reminder of this fundamental difficulty. In the field of NMR, methods improving the spectral lineshape (relaxation and field inhomogeneity) were developed in the group of Gareth A. Morris under the general
70   framework known as *Reference Deconvolution*, (Morris et al., 1997;Morris, 2002) but the difficulty to automate them makes deconvolution disappointingly disregarded. We shall not discuss, here, the deconvolution of spectral lineshape but concentrate on the deconvolution of the effect of coupling interactions called "multiplet-structure deconvolution". Following developments in the time domain (Le Parco et al., 1992;Bothner-By and Dadok, 1987;Prost et al., 2006) the frequency domain became more popular to avoid the back-and-forth Fourier transformation. This branch of research started in the '80s in the group of R.
75   Freeman (del Río Portilla et al., 1994) continued in the '90s in G. Bodenhausen's group (Huber and Bodenhausen, 1993a, b;Jeannerat, 2000;Jeannerat and Bodenhausen, 1999) and continued to resonate when the ACCA method (Cobas et al., 2005) was developed for Mestrelab's software Mnova.

This paper presents a follow-up of one published in 1999 (Jeannerat and Bodenhausen, 1999) which showed how to effectively use a set of inverse function of F to obtain, through a recursive procedure (Novič and Bodenhausen, 2002), the list of the
80   coupling constants and the lineshape G. The inverse function $M(J^*)$, reminiscent of the set of δ functions seen in the '90s (del Río Portilla et al., 1994) is discussed in the next paragraph. While earlier work focused on the simplification of simple doublet, we will present applications to the deconvolution of structures originating from the coupling with partners with S > 1/2, and deal with partially overlapping multiplet structures. It is generally limited to the analysis of first-order multiplets but a method

2

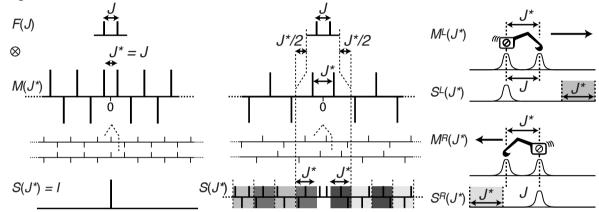to accommodate some "roof effect" is discussed (*i.e.* deal with the unequal amplitude of the two lines of doublets caused by second-order effects, see §1.7) and turn it into an advantage to determine, in favorable cases, the chemical shifts of the coupling partners.

1.2 Deconvolution of doublet structures

The process we shall use to reverse the effect of a doublet is the convolution product of the starting multiplet with the simplification function $M(J^*)$ (see Fig. 2). The deconvoluted function will usually be an experimental multiplet with a possibly complex structure, but the top of Fig. 2 shows how it applies to the model function of a doublet $F^{1/2}(J)$ to provide the mathematical demonstration that M simplifies any doublet structure. The function $M(J^*)$ is a doubly infinite series of δ functions separated by J^*. Their signs are positive for the pair located at $\pm J^*$ and alternate signs when running away from this core along the two directions of the abscissa.
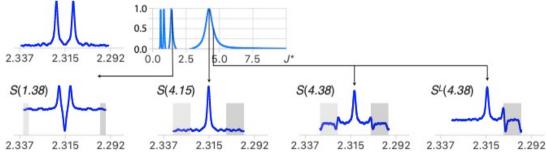

Figure 2: (left) The function $M(J^*)$ cancels the doublet of a weakly coupled partner with a spin 1/2 when $J^* = J$. (middle left) Duplicating M and shifting one copy by the distance J indeed cancels all except the central peak resulting in the expected simplification (bottom left). (centre) When $J^* \neq J$, the convolution product results in the complex set of δ functions S (middle bottom). (right) Illustration of the algorithm starting at one side of the spectrum. It can be seen as walking from one boundary of the multiplet, in the manner of an integrator, and subtracting at the distance J^*, the amplitude measure locally.

The simplification of the doublet structure into the singlet I is obtained when the distance between the δ functions of $M(J^*)$ corresponds to the true coupling (*i.e.* when $J^* = J$, bottom left in Fig. 2). Otherwise $S = F(J) \otimes M(J^*)$ consists in a complex set of δ functions that we call deconvolution artifacts (bottom center in Fig. 2). Determining the coupling constants consists in measuring the degree of simplification of the result S as a function of J^*. In principle, S should be infinite arrays, but beyond the boundaries of the experimental multiplet structures (see the section framed by dotted lines in the center of Fig. 2) two patterns (in four shades of grey) repeats themselves with alternating signs rendering the useful part of the S series finite. In the perspective of speeding up calculations, it has been demonstrated (Jeannerat, 1997;Bothner-By, 1995) that the result of the application of the arrays of δ functions $M(J^*)$ is equivalent to a recursive process starting at either side of the multiplet illustrated on the right part of Fig. 2. It shows that the calculation can be understood as the result of moving a cursor (represented by the excavator) over the vector of signal amplitudes and subtract, at the arm's distance, representing the tested value of J^*, the value measured at its current position. After chopping off the margins which should be empty (in grey) to align the sub-multiplets, the sum of S^L and S^R is numerically equal to S. This approach has the advantage of being computationally quite effective: each point in the deconvoluted spectrum is calculated as the sum of only two points instead of $[W/J^*]$ where W is the width of the multiplet region. When $J^* = J$, the remote subtraction eliminates the second occurrence of the substructure split by J. A key property of this algorithm is that artifacts and the residuals of sub-multiplet subtraction increase from the starting boundary to the other as the number of operations increases. It is tempting to reduce artifacts by taking only the most favorable half of the simplified multiplets, but it is not generally recommended because it often produces a small discontinuity in the middle of the resulting multiplet. The most significant advantage of the side-to-side process is to facilitate

3

the measure of the similarity of S^L and S^R to test the success of the simplification process. It takes advantage of presence of deconvolution artifacts indicating that $J^* \neq J$ only in the right, respectively left marginal regions (in grey in Fig. 2) of S^L and

120   S^R. Instead of the classical Chi-squared test, we favored a scalar product (Huber and Bodenhausen, 1993a) as measure of similarity but this should only be a matter of personal preference. Finally, the presence of two sub-multiplets makes it possible to simply discard one of them when an artifact (solvent singlet, spike, *etc.*) or partial overlap makes the complementarity fail. (see §1.8)

1.3 Recursive simplification of multiplet structure

125   As mentioned earlier, deconvolution produces artifacts because the components of multiplets are never perfectly identical. This causes imperfect multiplet subtraction due to the presence of random noise, *etc*. These artifacts tend to be duplicated for each δ function of M and their presence is often a limiting factor for the analysis of complex structures. Reducing them is therefore decisive to make multiplet deconvolution reliable. Deconvolution with large values of J^* generate fewer artifacts because they contain less δ function per unit distance. This is one of the reasons to aim first at the largest coupling present in

130   any multiplet. Note that it is the opposite to what a human would do based on the knowledge that only the smallest coupling can be readily measured between the first and second outermost lines of any multiplet. The other reason to start with the largest coupling constants is that values of $J^* = J/(2n+1)$ (where $n \in \mathbb{N}^*$) also produces multiplet structure with no marginal signal (see the extremum at $J^* = J/3$ in Fig. 3) making it unreliable to use symmetry as a criterion for the determination of values smaller than one third of the largest coupling of the multiplet. This problem becomes irrelevant when using a recursive process

135   where the result of the simplification of the largest coupling constants is used as the starting point of the next step. In order to limit the accumulation of artifacts, applying symmetry at each step or taking the best part of the sub-multiplet S^L and S^R, is generally recommended.

140   **Figure 3: (top right) Quality function of the deconvolution of an experimental doublet (top left). When starting with large values of J^*, the first extremum points to the largest coupling constant of the multiplet structure ($J^* = 4.15$ Hz). (bottom) Structures obtained for deconvolution with $J^* = 1.38$ ($= J/3$), 4.15 ($= J$) and 4.38 ($= J + 0.23$) Hz and M^L for 4.38 Hz. When $J^* \neq J$, the presence of deconvolution artefacts (in the grey area) breaks the symmetry of the structure and causes a drop of the quality function except when $J^* = J/(2n+1)$, where $n \in \mathbb{N}^*$. (See text for more details.)**

145   An alternative to the measure of symmetry consists in testing the sum of the absolute values of the resulting multiplets. It should reach half the value of the starting multiplet when it is simplified. It requires only M^L and M^R (instead of both for the comparison using symmetry) but we used it solely when only one side was available, for example when analysing partially overlapping multiplets (see §1.8).

Having a good starting multiplet is also important and requires the spectrum to present a reasonable baseline and the

150   identification and subtraction of solvent and other artifact peaks by the processing software prior to the deconvolution process increases the chances of success of multiplet deconvolution. On the other hand, random noise, should only impact on the confidence level of the identification of each coupling.

1.4 Multiplet lineshape

An interesting property of multiplet deconvolution is to make no assumption on the underlying signals lineshape. It can be any
mixed Lorentz/Gauss function, transformed by any type of apodization or lineshape deconvolution. Multiplet deconvolution
only needs a pattern, of any shape, to be equally split. Figures 4 and S1 illustrate this point for a classical phase distortion, a
distorted lineshape, and the results of a resolution enhancement replacing the multiplet with its second derivative (Wahab et
al., 2018). This being said, the above-mentioned symmetrization procedure applied at each step of the simplification process
is generally beneficial (compare Figs. S2 and S3). It averages out noise, artifacts and asymmetrical coupling structures caused
by mild second-order effects (roof effects) but it usually does not work properly when signals do not have a pure absorption
lineshape (*i.e.* when the phase is incorrect) or when the B_0 inhomogeneity produces a severely asymmetrical lineshape.
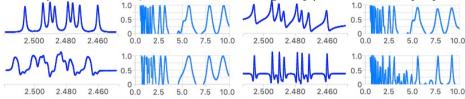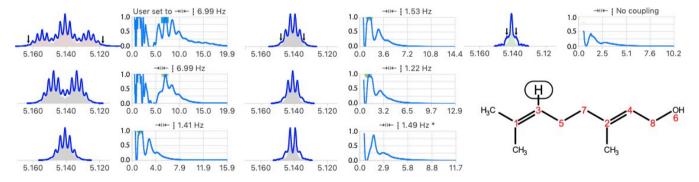

**Figure 4: Comparison of the first step of the analysis of reference multiplet before (top left) and after introduction of a phase error
(top right), a distorted line-shape (bottom left) and using its second derivative (bottom right). The measured coupling constants are
the same in all cases (not shown) as the position of extrema in the four error functions shows. The narrower extrema of the error
functions in the latter indicate a higher potential to identify small coupling constants. See Fig. S1 for a more detailed analysis.**

1.5 Spin system degeneracy

When spin systems are *n*-fold degenerated, the convolution with $M(J^*)$ is simply applied *n* times before measuring the degree
of simplification. A trivial example of quartet structure is shown in Fig. S4. A user-controlled specification of degeneracy
simplifies the structure in one step and avoids the possible errors of the automatic identification which takes as *degenerate* the
coupling constants differing by less than 0.5 Hz.
Another difficulty arises, in highly degenerate systems, when the automatic analysis misses the outermost signals (see Fig. 5).
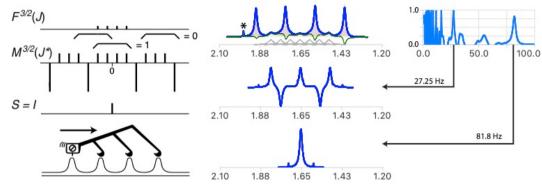

**Figure 5: Example of automatic analysis failing to identify a triplet structure after six steps of simplification. Because the outermost
lines (see arrows) were too small. It identified a problem because the reconstructed multiplet (triple quintet filled in red) was not
perfectly matching (the scalar product was < 0.99) the starting multiplet (in blue). The post-analysis procedure found a satisfactory
match (not shown) for a triplet (7.0 Hz) caused by methylene (C5) of heptet (^4J = 1.3 Hz) caused by the coupling with two methyl
groups bound to C1 of geraniol.**

A post-processing procedure using multiplet simulations (see §2.3) to test if higher levels of degeneracy match better the
starting multiplet turned out to be useful, including in the case illustrated in Fig. 5.

5

1.6 Coupling partner with S > 1/2

When compounds contain deuterium ($S = 1$, 1:1:1 structure), boron (^{11}B, $S = 3/2$, 1:1:1:1 structures) and other atoms with $S > 1/2$ isotopes, the simplification of the multiplet structure with n equal-amplitude δ function is necessary when quadrupolar relaxation is not effectively hiding the coupling structure. The general inverse function $M(J^*)$, valid for $S \geq 1/2$, consists in a core of n δ functions with unit values producing the main singlet. It is flanked, at both sides by a repeat of blocks made of one negative δ function with intensity equal to $-(n-1)$ and $n-1$ lines with unit intensities. The function $M^{3/2}(J^*)$ is shown in Fig. 6.

Figure 6: (top right) Model multiplet structure for coupling with a $S = 3/2$ partners. (middle right) Simplifying series of δ functions $M^{3/2}(J^*)$. The braces show examples of integration regions and illustrate that it is zero for all positions except the central one. The side-to-side process is quite similar to the case of $S = 1/2$ except that subtraction of the cursor position has to be made to the $n-1$ positions located J^* further in the data array. (center) Analysis of the proton spectrum of BH$_4^+$. Specifying the spin of the partner to be $3/2$ replaces the wrong assignment dd ($J = 163.6$, 81.6 Hz, not shown) with the correct 1:1:1:1 quartet with $J = 81.6$ Hz (bottom center). The extremum of the quality function (right) for $J^* = J/3 = 27.25$ is due to non-simplified but symmetric structure (bottom center). The presence of the minor multiplet for the ^{10}B isotope (20% natural abundance, $S = 3$, 7 transitions) with a γ one third that of ^{11}B is highlighted in grey in the top subplot. Note that the artefact (probably traces of acetone highlighted by a star symbol) was scaled down by symmetrization.

The convolution product of $F^{3/2}(J)$ and $M^{3/2}(J^* = J)$ is indeed I because the sum of any set of four consecutive δ functions ($\sum_{i=m}^{m+n} \delta_i$) is equal to zero for all positions except the middle one. The side-to-side recursive functions $M^{L,S>1/2}(J^*)$ and $M^{R,S>1/2}(J^*)$ also exist (see M^L at the bottom left of Figure 6).

For a given coupling constant, the density of δ functions in the simplifying function is the same as for spin 1/2, but the fact that the average of the intensities of the M functions increase with spin order and that these multiplets are more extended should generally make the simplification of complex multiplets more prone to artifacts.

The analysis of the main 1:1:1:1 quartet of the proton of borohydride produced by the coupling with ^{11}B is straightforward (see Fig. 6). The process is not disturbed by the presence of the signal of the 20% abundance of the ^{10}B isotopologue. Note that a refinement of the parameters involving a fit of the experimental spectrum with spectral simulation taking into account the presence of isotopologues, could provide values for the isotopic shift and the line widths of both multiplets.

The analysis of a system including the deuterium atoms of the CHD$_2$ residual signal of DMSO-D6 is shown in Fig. S5.

1.7 Mild second-order effects

Because second-order effects are often observed in otherwise well-behaved multiplet (*i.e.* when only the peak intensities are uneven with no significant shift or additional second-order transitions) we introduced, in this work, the function deconvoluting doublets with non-equal amplitudes. Mild second-order effects are usually not impairing the extraction of coupling constants, especially when symmetrizing multiplet structures, but the validation of the results matching simulation with the experimental multiplet (see §2.3) may fail because of the mismatch of the peak amplitudes. Accounting for the roof effect restores high confidence in the results of the analysis. The simplification function $M^{1/2}(J^*, \theta)$ (see Fig. 7) includes the strength of the

6

coupling as a parameter expressed a $\theta = \tan^{-1}(J/\Delta\delta)$ where $\Delta\delta$ is the difference in the chemical shift of the two coupling partners expressed in Hz. It produces the identity function when taking the convolution product with the second-order doublet $F^{1/2}(J, \theta)$ where the ratio of peak amplitudes

$$r = \frac{1 - \sin(\theta)}{1 + \sin(\theta)}. \tag{1}$$

Note that the stronger the coupling, the more artifacts will be amplified at one side of the resulting multiplet. In order to avoid having two independent variables to adjust, θ can be optimized only for values of J^* corresponding to an extremum of the values obtained for $\theta = 0$.
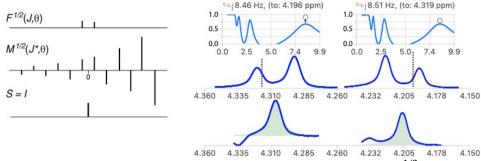

Figure 7: (left) Illustration of the model function of a doublet with non-equal amplitude $F^{1/2}(J, \theta)$ and its inverse function $M^{1/2}(J^*, \theta)$. The amplitude ratio of the absolute values of two consecutive δ in M is the inverse of the ratio in $F^{1/2}(J, \theta)$. (right) Analysis of an two multiplet structures of an AB spin system (300 MHz spectrum of a taxol sample). $\Delta\delta/J = 6.6$, $2\theta = 8.6$ deg. and the ratio of signal amplitudes $r = 0.74$. The error in the determination of partner chemical shift was about 0.01 ppm. The open circles in the quality functions indicate the level of the measure of the quality function after optimization of θ.

Besides the fact that taking into account the roof effects increases the level of confidence that any measure splitting is correct, exploiting them provides, when it can be measured with sufficient precision, the chemical shift of the relevant coupling partner. (Sykora, 2008) This information normally requires a 2D COSY spectrum, but extracting this information for the 1D spectrum has the additional advantage to identify which of the coupling observed in a given multiplet corresponds to the designated coupling partner. In other words, instead of only qualitatively pointing to the side of the spectrum where one should look for the coupling partner of a signal showing roof effects, one can actually point to the chemical shift of the partner according to:

$$\delta_{partner} = \delta_{ref.} + \frac{J}{\tan(\theta)}. \tag{2}$$

The determination of the coupling partner position is expected to provide useful values for a relatively narrow range of second-order effects (say $3 < \Delta\delta/J < 20$). In the example of Fig.7, taken from the ^{1}H spectrum of a sample of taxol (Peng et al., 1997), $\Delta\delta/J = 6.6$. Indeed, the second-order effect should not be too strong to avoid the minor line of the doublet to be too small to be measured accurately. Similarly, if the effect is too weak, the difference of intensity of the two lines of the doublet becomes too small to be significant and generate large errors because $1/\tan(\theta)$ becomes quite large when θ is small (see the example in Fig. S6). Taking into account the error on the partner position by standard error propagation calculation is therefore essential to avoid misinterpretation. Obviously, the chances of success of simply picking the nearest multiplet about the expected chemical shift as true partner depends on the complexity of the spectra.

1.8 Partially overlapping multiplets

The side-to-side deconvolution algorithm allows to deal with multiplets that are partially superposed. Figure 8 shows that signal overlap over a distance smaller than the largest coupling present in the multiplet can be analysed successfully. When attempting to simplify the left multiplet, the deconvolution process should be run from left to right and the deconvolution process stopped at the position where the integral reaches 1/2 in the reconstructed multiplet.
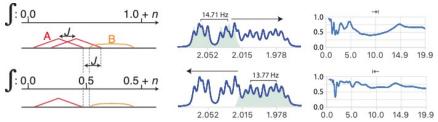
7

Figure 8: Analysis of the multiplet A in the presence of a partially overlapping multiplet B with relative integral of 1 and n respectively. When deconvolution is successful, the left-to-right deconvolution process eliminates the second occurrence of the sub-multiplet A which effectively eliminates overlap (bottom left). (right) Example of separation of two slightly overlapping multiplet structures from a sample of artemicin dissolved in CDCl$_3$. The coupling constants J = 14.71, 4.77, 2.89 and 13.77, 6.15, 3.91, 2.93 Hz were extracted from the left and right multiplets respectively and used to reconstruct the matching green multiplets. Note the slight mismatch due to second-order effects which could be ignored in this analysis. See Fig. S7 for the detailed simplification.

When testing different values of J^*, the measure of the success of the simplification cannot be applied on the whole spectral region because the segment where the sub-multiplet subtraction occurs may be occupied by multiplet B. But successful deconvolution produces a minimal and predictable integral when $J^* = J$. This multiplet separation method requires to know the integral of the analysed multiplet relative to the total integral of the region of interest. This is usually not problematic as integration is generally straightforward. An example of analysis resulting to a sub-multiplet separation is shown on the right part of Figs. 8 and S7.

Following the analysis of A, the multiplet B can be analysed independently (from right to left) only if its largest coupling is also large enough, as in the example on the right of Fig. 8. Otherwise, a recursive process analysing and subtracting multiplet structures sequentially is necessary. This would impose the quality of the subtraction of each multiplet to be high enough to leave no significant residual signals on the remaining, but allows, in principle, to deal with more than two overlapping multiplets.

For more severely overlapping structures, the methodology developed to separate 2D multiplets (Jeannerat and Bodenhausen, 1996) could be adapted to the processing of 1D spectra.

2 Implementation

2.1 Deconvolution parameters

As mentioned above, the simplification of structure is a recursive process where the noise and artifacts tend to increase along the course of the analysis and depend on the quality of the multiplet (complexity, base line distortion, presence of spurious peaks, *etc.*)

In order to increase the robustness of multiplet deconvolution, different set of parameters driving the deconvolution process are tested and their results validated (see next paragraph). These parameters are: the threshold for considering a deconvolution to be successful, the threshold for considering the recursive process to be completed, the range of values tested for the roof effect, whether symmetry is applied before starting the analysis, whether symmetry is applied after each simplification, whether the baseline offset if corrected, whether the crude spectrum is used instead of the GSD-based synthetic spectrum (Cobas et al., 2008;Schoenberger et al., 2016), *etc*.

2.2 Limit of deconvolution analysis

Obviously, the recursive simplification stops when the result is a singlet. But the ability of multiplet deconvolution to find unresolved coupling makes that simply testing the presence of only one extremum in the result of a simplification step could miss unresolved coupling constants (see §3.2). Testing values down to 1 Hz seems to be a good compromise for standard 1D ^{1}H spectra. Note that sub-linewidth coupling constants should be used with great care and only to facilitate the process of a

8

structure elucidation process. They should not be used as sole argument for any key interpretation. Indeed, the behavior of the deconvolution has not been studied sufficiently to provide a general and safe limit as to how far out it can be carried. A rigorous analysis of unresolved structures would involve a process taking, among other elements, the lineshape found in other multiplet into account. When a candidate structure is available, predictable long-range coupling constants should be considered. Depending on the priority of the application, one may favor safety over ambition and decide to ignore any unresolved structure in a "fully automated" mode. Pushing the limit of multiplet analysis could be a user-triggered operation providing visualization tools to assess the validity of the results and correct them if necessary.

2.3 Validation and post-processing

The validation and ranking of the results of the different methods consists in the measure of the similarity of the starting multiplet with a reconstructed multiplet build using the extracted data. The latter are obtained by starting with the final singlet shape and reintroducing the measured coupling constants including, when relevant the tilt caused by second-order effect, and testing the similarity with the initial multiplet. In this work, the starting lineshapes were synthetic Lorentzian but allowing more degrees of freedom (*i.e.* using Voigt, Generalized Lorentzian, (Schoenberger et al., 2016) *etc*. lineshape) should improve the match with the shape obtained at the end of the deconvolution.

Finally, a refinement of all parameters: signal intensity, chemical shift, coupling constants, but also roof effect, signal phase, baseline level, and lineshape parameters could be used to further improve the quality of the validation and take into account context-dependent information such as the degeneracy of the coupling constants.

3 Results and discussion

Multiplet deconvolution has been applied to numerous test spectra. Many multiplet structures which could not be analysed using the GSD peaks-based multiplet analysis (reported as "*m*") were correctly deciphered. It also has a lower occurrence of false positive – the situation where coupling constants are automatically extracted but turn out to be incorrect – thanks to the strict validation process. We shall only present here two typical examples where multiplet deconvolution succeeds while manual analysis would be difficult.

3.1 Human unfriendly multiplets

Figure 9 shows a typical example of multiplet that cause difficulties to human analysis. The partial overlap of three transitions (see arrows) makes the structure difficult to interpret. The destiny of such a structure is to be called the bad name "*m*" for a lack of courage and proper tool to identify it. Multiplet deconvolution straightforwardly identify the largest coupling as 9.9 Hz. Similarly, the seemingly triplet of the second step (bottom left of Fig. 9) would have been assigned to a *t* with two equal coupling constants of 5.27 Hz - possibly causing confusion if nothing explains this apparent degeneracy. Here again, deconvolution has no difficulties to resolve overlap and identifies a double doublet with 6.32 and 4.22 Hz - almost 2 Hz difference for a final line width of 2.7 Hz.
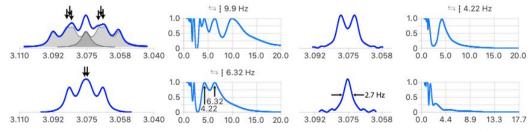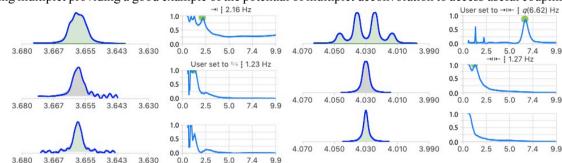

Figure 9: Example of a *ddd* structure analysed straightforwardly by multiplet deconvolution. The ⦙ symbol indicate that symmetrization was applied before deconvolution. The ⇆ symbol indicates that the results of the deconvolution running from both sides were added (instead of taking the best part of both of each or only the result of one direction).

9

An example of successful analysis of "*dqdd*" multiplet structure which was too complex to be analysed manually is shown in Fig. S8. The six coupling interactions, among which three were degenerate, produced 32 lines with no visible splitting patterns.

3.2 Unresolved couplings

330 In carbohydrate chemistry, geminal coupling constants are often too small to be resolved making it difficult to assign their signals. The analysis of the multiplet structures of Fig. 10, identified a common coupling of 1.2 Hz which increases the confidence in their suspected geminal relationship. In the signal at 3.655 ppm, neither the 2.2 nor the 1.2 Hz were resolved in the starting multiplet providing a good example of the potential of multiplet deconvolution to access useful coupling constants.

335 **Figure 10: Extraction of unresolved coupling constants of the proton signals of C4 (left) and C5 (right) of a 6-deoxy-pyranose unit. The ⇥ symbol indicates that symmetrization was not applied and deconvolution run from left to right.**

Conclusion

The recursive analysis of NMR spectra by multiplet deconvolution demonstrated its ability to extract coupling constants in a robust and fully automated manner. The results are validated when simulations based on the extracted data match the
340 experimental multiplet structures. Additional features including the analysis of structures produced by coupling partners with S > 1/2 and of regions presenting signal overlap are proposed with the option of using a graphical interface providing a full user control on the stepwise analysis of multiplet structures.

By reducing the number of the unsuccessfully analysed multiplet, multiplet deconvolution increases the information content of NMR spectra at a much lower cost than increasing the field strength of NMR spectrometers. Combined with other efforts
345 aiming at automatizing the analysis of NMR spectra, these methods can significantly increase the quantity and quality of NMR data available to the community. But this will only occur if researchers make their NMR spectra and extracted data available as supplementary data deposited on public databases instead of providing only crude images of spectra and other non-computer readable information.

Data Availability

350 The software presented here is proprietary of Mestrelab but the code for a basic recursive deconvolution of first-order multiplet with spin 1/2 partners is copyright free and available as a JavaScript node modules on https://cheminfo.github.io/multiplet-analysis. The JCAMP-DX files of the spectra presented in this paper are available on Zenodo doi: 10.5281/zenodo.4616665. They can be visualized with the open-source NMR-displayer (also known as NMRium) accessible from the https://github.com/cheminfo/nmr-dataset3 and analysed by the basic implementation of deconvolution by clicking on the
355 *Ranges Picking* button and select a multiplet region with the mouse, while pressing the shift key.

Acknowledgements

References

Berger, S.: A quarter of a century of SERF: The progress of an NMR pulse sequence and its application, Prog Nucl Magn Reson Spectrosc, 108, 74-114, 10.1016/j.pnmrs.2018.10.001, 2018.

Bothner-By, A. A., and Dadok, J.: Useful manipulations of the free induction decay, Journal of Magnetic Resonance (1969), 72, 540-543, 10.1016/0022-2364(87)90158-2, 1987.

Bothner-By, A. A.: Private communication, 1995.

Cobas, C., Seoane, F., and Sykora, S.: Global Spectral Deconvolution (GSD) of 1D-NMR spectra. Stan's Library, Volume II. 10.3247/sl2nmr08.011, 10.3247/sl2nmr08.011, 2008.

Cobas, J. C., Constantino-Castillo, V., Martin-Pastor, M., and del Rio-Portilla, F.: A two-stage approach to automatic determination of ^{1}H NMR coupling constants, Magn Reson Chem, 43, 843-848, 10.1002/mrc.1623, 2005.

del Río Portilla, F., and Freeman, R.: Measurement of Spin Coupling Constants by Decoupling and Reconvolution, J. Magn. Reson., Ser. A, 104, 358-362, 10.1006/jmra.1993.1236, 1993.

del Río Portilla, F., Blechta, V., and Freeman, R.: Measurement of Poorly Resolved Splittings by J doubling in the Frequency Domain, J. Magn. Reson., Ser. A, 111, 132-135, 10.1006/jmra.1994.1238, 1994.

Emsley, L., Huber, P., and Bodenhausen, G.: Unraveling Overlapping Multiplets in Two-Dimensional NMR Correlation Spectra by Selective Inversion of Coupling Partners, Angew. Chem. Int. Ed. Engl., 29, 517-520, 10.1002/anie.199005171, 1990.

Huber, P., and Bodenhausen, G.: Simplification of Multiplets by Deconvolution in One- and Two-Dimensional NMR Spectra, Journal of Magnetic Resonance, Series A, 102, 81-89, 10.1006/jmra.1993.1070, 1993a.

Huber, P., and Bodenhausen, G.: Bidirectional Deconvolution, an Improvement for the Stepwise Simplification of Multiplets, Journal of Magnetic Resonance, Series A, 104, 96-102, 10.1006/jmra.1993.1194, 1993b.

Jeannerat, D., and Bodenhausen, G.: Separation of overlapping multiplets and contraction of substructures within multiplets using symmetry properties, J. Magn. Reson., Ser. A, 119, 139-144, 1996.

Jeannerat, D.: Progress in cross-peak structure analysis in two-dimensional nuclear magnetic resonance spectroscopy, PhD, University of Lausanne, Lausanne, 133 pp., 1997.

Jeannerat, D., and Bodenhausen, G.: Determination of Coupling Constants by Deconvolution of Multiplets in NMR, J. Magn. Reson., 141, 133-140, 1999.

Jeannerat, D.: Application of frequency domain deconvolution to DOF-COSY spectra for the determination of coupling constants as a tool for signal assignment based on coupling networks, Magn. Reson. Chem., 38, 156-164, 10.1002/(SICI)1097-458X(200003)38:3<156::AID-MRC610>3.0.CO;2-R, 2000.

Karplus, M.: J. Am. Chem. Soc., 85, 2870, 10.1021/ja00901a059, 1963.

Kiraly, P., Kern, N., Plesniak, M. P., Nilsson, M., Procter, D. J., Morris, G. A., and Adams, R. W.: Single-Scan Selective Excitation of Individual NMR Signals in Overlapping Multiplets, Angewandte Chemie, 133, 676-679, 10.1002/ange.202011642, 2020.

Le Parco, J.-M., McIntyre, L., and Freeman, R.: Accurate coupling constants from two-dimensional correlation spectra by "J deconvolution", Journal of Magnetic Resonance (1969), 97, 553-567, 10.1016/0022-2364(92)90035-6, 1992.

Marquez, B. L., Gerwick, W. H., and Thomas Williamson, R.: Survey of NMR experiments for the determination of nJ(C,H) heteronuclear coupling constants in small molecules, Magnetic Resonance in Chemistry, 39, 499-530, 10.1002/mrc.902, 2001.

McIntyre, L., and Freeman, R.: Accurate Measurement of Coupling Constants by J Doubling, J. Magn. Reson., 96, 425-431, 10.1016/0022-2364(92)90098-R, 1992.

Metz, K. R., Lam, M. M., and Webb, A. G.: Reference deconvolution: A simple and effective method for resolution enhancement in nuclear magnetic resonance spectroscopy, Concepts in Magnetic Resonance, 12, 21-42, 10.1002/(sici)1099-0534(2000)12:1<21::Aid-cmr4>3.0.Co;2-r, 2000.

Morris, G. A., Barjat, H., and Horne, T. J.: Reference deconvolution methods, Progress in Nuclear Magnetic Resonance Spectroscopy, 31, 197-257, 1997.

Morris, G. A.: Reference deconvolution, Encyclopedia of Nuclear Magnetic Resonance, 9, 125-131, 2002.

11

Novič, M., and Bodenhausen, G.: Symmetry mapping and recursive reduction of multiplet patterns in nuclear magnetic resonance spectra obtained with two-dimensional correlation and multiple-quantum methods, Analytical Chemistry, 60, 582-591, 10.1021/ac00157a018, 2002.

Peng, C., Jeannerat, D., and Bodenhausen, G.: Determination of homonuclear scalar coupling constants by combining selective two-dimensional NMR spectroscopy with convolution and deconvolution: applications to paclitaxel (taxol), Magn. Reson. in Chem., 35, 91-99, 10.1002/(SICI)1097-458X(199702)35:2<91::AID-OMR22>3.0.CO;2-B, 1997.

Prasch, T., Gröschke, P., and Glaser, S. J.: SIAM, a Novel NMR Experiment for the Determination of Homonuclear Coupling Constants, Angew. Chem. Int. Ed. Engl., 37, 802-806, 10.1002/(sici)1521-3773(19980403)37:6<802::Aid-anie802>3.0.Co;2-m, 1998.

Prost, É., Bourg, S., and Nuzillard, J.-M.: Automatic first-order multiplet analysis in liquid-state NMR, Comptes Rendus Chimie, 9, 498-502, 10.1016/j.crci.2005.05.012, 2006.

Schoenberger, T., Menges, S., Bernstein, M. A., Perez, M., Seoane, F., Sykora, S., and Cobas, C.: Improving the Performance of High-Precision qNMR Measurements by a Double Integration Procedure in Practical Cases, Anal Chem, 88, 3836-3843, 10.1021/acs.analchem.5b04911, 2016.

Sykora, S.: Stan's Library, Vol.II, 2008.

Wahab, M. F., Gritti, F., O'Haver, T. C., Hellinghausen, G., and Armstrong, D. W.: Power Law Approach as a Convenient Protocol for Improving Peak Shapes and Recovering Areas from Partially Resolved Peaks, Chromatographia, 82, 211-220, 10.1007/s10337-018-3607-0, 2018.