NMR Data Evaluation: Review of Covariance Applications

Juan Carlos Cobas Gómez, Mestrelab Research
First published in February 2014

Covariance NMR has attracted considerable interest in recent years and new applications of this technique aimed at paving the way forward to structure elucidation are appearing frequently in the scientific literature. For example, Zhang and coworkers used Covariance TOCSY NMR to improve the resolution of 2D spectra and simplify the task of mixtures analysis.[1] The objectives of this technique are manifold, ranging from the resolution enhancement of homonuclear 2D spectra to the generation of new spectra that would require longer acquisition times by "combining" two spectra with a lower time "cost".

Whilst the mathematical foundations and chemical applications are very well established (although growing) and described in the specialized literature in great detail, chemists without a solid background on NMR spectroscopy might have difficulties in understanding the particular jargon of Covariance NMR. It is the goal of this article to provide a very high level description of what Covariance NMR entails as well as giving practical hands-on advice on how to use the different Covariance NMR methods within Mnova NMR software.

Keywords: review, NMR, spectroscopy, data, evaluation, covariance NMR, software, method

Contents

Introduction

Under the more generic name of Covariance NMR, it is convenient to consider four different methods according to the number and type of the spectra involved in the particular application: Direct, Indirect, Unsymmetrical (and Generalized) and doubly Indirect Covariance NMR. These methods are summarized in Table 1, below.

Looking back at the origins of this technique, statistical Covariance based methods were originally proposed by Noda in the fields of optical spectroscopy, first in Infrared[2] and then in Raman, Ultraviolet and other spectral types.[3,4,5] It was then applied to NMR spectra and termed **Generalized 2D NMR (GEN2D-NMR)** showing its potential in the analysis of Pulse Field Gradient experiments (i.e. DOSY)[6] and in solid state NMR as a method capable of reducing the experimental acquisition time for 2D homonuclear NMR spectra - in particular, 13C-13C J-correlation experiments and 1H-1H NOESY.[7]

Cobas C., *NMR Data Evaluation: Review of Covariance Applications*, DOI: 10.3247/SL5Nmr14.001

GEN2D was further extended in the context of NMR to establish spin correlations through magnetization or coherence transfer as an alternative route to 2D FT-NMR, and its name was changed to *Covariance NMR*. The categories of Covariance (Table 1) will be discussed in the following sections.

Table 1: NMR Covariance methods

Method	Equation	Description	Application
Direct Covariance	$C_{direct} = (F^T \cdot F)^{1/2}$	Mapping of the detected dimension (*donor*) onto the indirect dimension (*acceptor*)	Resolution enhancement and reduction of experimental time in 2D homonuclear experiments (e.g. COSY, TOCSY, NOESY, ROESY)
Indirect Covariance	$C_{indirect} = (F \cdot F^T)^{1/2}$	Mapping of the indirect dimension (*donor*) onto the direct dimension (*acceptor*)	Generation of symmetric homonuclear spectra (e.g., 13C-13C TOCSY) from non-symmetric heteronuclear experiments (e.g. 1H-13C-HSQC-COSY/TOCSY)
Unsymmetrical Indirect Covariance (UIC)	$C_{indirect} = F \cdot G^T$	Calculation of a new spectrum from pair of spectra, F and G, with a shared common frequency domain	Reconstruction of the equivalent of low sensitivity, hyphenated NMR experiments by combining data from a pair of higher sensitivity experiments (e.g. COSY + HSQC, HSQC+HMBC)
Generalized Indirect Covariance NMR (GIC)	$S = [(F \cdot G)(F \cdot G)^T]^{\lambda}$	Same as UIC	Same as UIC, but with improved artifact reduction
Doubly Indirect Covariance	$C = H \cdot Y \cdot H^T$	Combination of homonuclear (Y) and heteronuclear (H) NMR spectra to produce a 13C-13C correlation spectrum.	Construction of a 2D-INADEQUATE-like experiments that correlates two carbon resonances whose directly attached protons display a cross-peak in the COSY spectrum but at much higher sensitivity of the 13C-1H HSQC experiment

Direct Covariance NMR

Direct Covariance NMR, developed by the group of Rafael Brüchweiler[8], is mathematically analogous to GEN2D-NMR and equally its main application is to increase the resolution of the indirect dimension of 2D homonuclear experiments like COSY, TOCSY or NOESY, as well as reducing the experimental time.

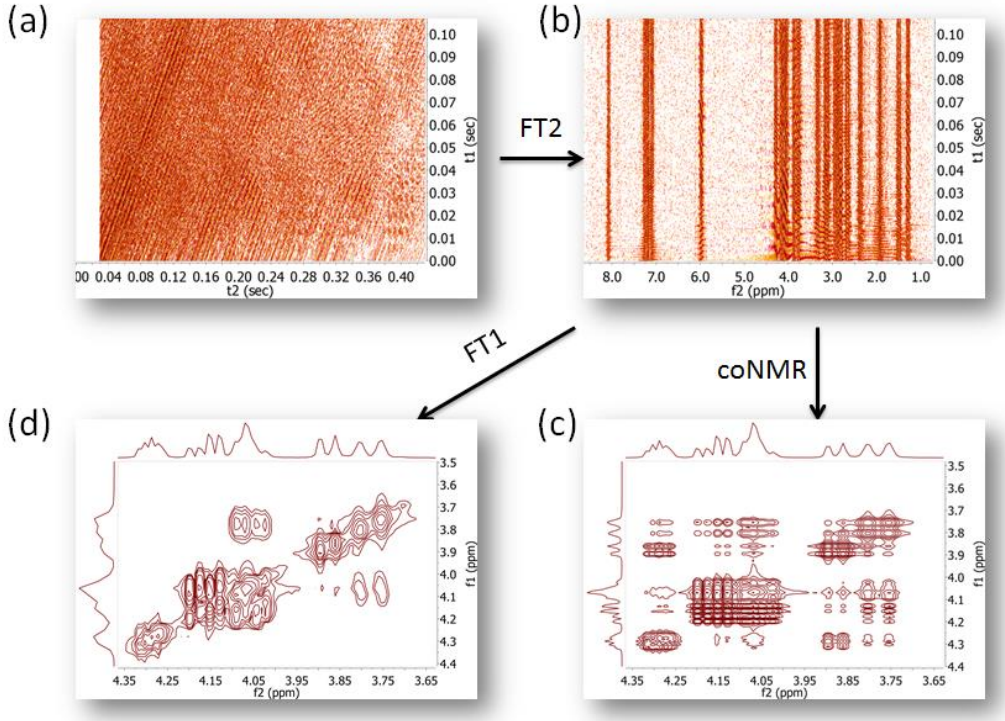
Initially, it was formulated as an alternative to applying a second Fourier Transform. The basic idea was that after transforming a 2D spectrum along the direct time dimension (t2) to yield a mixed time-frequency matrix (denoted by **S**), instead of performing the traditional second FT along the indirect or evolution dimension (t1), the so-called Covariance matrix C^2 is computed by multiplying the transpose of S by itself:

[1] $C^2 = S^T S$

where T denotes the matrix transpose and S is the (f2,t1), mixed time-frequency matrix (i.e. the data matrix resulting after applying the first Fourier Transform along the acquired dimension).

Cobas C., *NMR Data Evaluation: Review of Covariance Applications*, DOI: 10.3247/SL5Nmr14.001

The overall process is illustrated in Figure 1. Figure 1a shows the acquired FIDs having 1024 complex points along t2 and 256 t1 increments. After Fourier Transforming along t2, a new data matrix, S, in the mixed frequency-time domain with 1024 x 256 (N2 x N1) points is obtained (Figure 1b). Next, in order to obtain the final 2D spectrum, two possibilities are available: Apply a Fourier Transform along the



indirect dimension to yield a spectrum with 1024 x 256 data points (Figure 1d) or Covariance NMR using equation [1], shown in Figure 1c. Considering equation [1], it should be obvious that the resulting matrix C^2 will have 1024x1024 data points, whereas the conventional second FT will yield a 2D data matrix having 1/4 as many data points.

Figure 1: The full time domain 2D experiment (a) is transformed using FT along the indirect dimension to yield a mixed time-frequency matrix S with N2 and N1 data points. FT along the indirect dimension will yield the traditional 2D spectrum (d) with N2 and N1 data points (provided that no zero filling is applied in the evolution dimension). (c) shows the Covariance NMR spectrum with N2 x N1 data points. Note the improved digital resolution in the indirect dimension of the Covariance spectrum (1c) compared with the spectrum that had the traditional double FT (1d).

The important point to consider here is that application of [1] is equivalent (although not identical - more about this in a moment) to applying the second FT, that is, the result will be similar to a 2D correlation spectrum obtained via traditional FT but with one very important difference: [9] the digital resolution in the indirect dimension will be exactly the same as in the direct dimension. From a very high level point of view, we can understand this by considering that by multiplying the transpose of S by itself, we are transferring the resolution of the direct dimension to the indirect dimension. Following this reasoning, we could call the direct dimension the ***donor*** dimension and the indirect the ***acceptor*** dimension. On the other hand, it is worth noting that similar results in terms of resolution enhancement could be achieved by zero filling in the indirect dimension followed by symmetrization.

In practice, there is no need to apply Covariance NMR in the mixed time-frequency domain spectrum. It can be shown (e.g. by using Parseval's theorem) that applying Covariance NMR in the mixed time domain spectrum is equivalent to applying it in the final 2D-FT spectrum:

[2] $C^2 \approx S^T S \approx F^T F$.

Cobas C., *NMR Data Evaluation: Review of Covariance Applications*, DOI: 10.3247/SL5Nmr14.001

In order to approximate the intensities of the covariance spectrum C^2 to those of the idealized 2D FT spectrum, the square root of C^2 should be taken. Root squaring may also reduce false correlations which may be present in C^2 due to resonance overlaps.

Summary: direct Covariance NMR is used in homonuclear 2D spectra to improve resolution along the indirect dimension.

This is done in Mnova by issuing the command *Process/Covariance/Direct Covariance*. It includes an efficient algorithm to calculate the square root of the covariance matrix[10] as well as the possibility to regularize the spectrum by adding a properly scaled unit matrix.[11]

Indirect Covariance NMR

As described in the previous section, direct Covariance NMR involves multiplying the transpose of a homonuclear spectrum by itself. This results in a resolution transfer from the direct dimension to the indirect one. For convenience, let's consider again the equation for direct Covariance NMR:

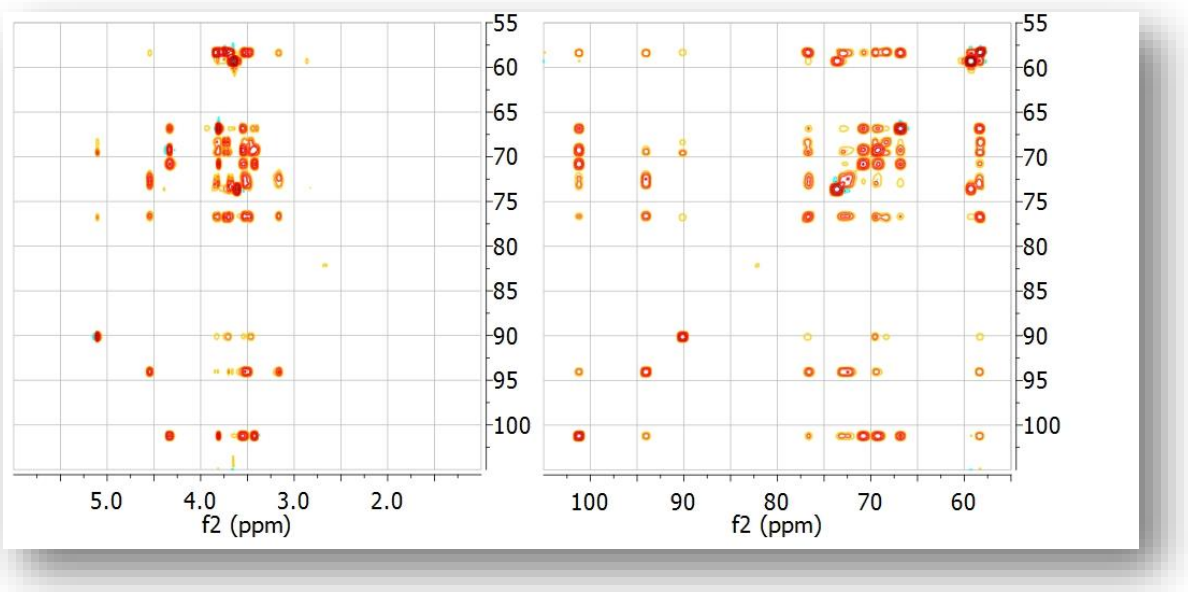
[3] $C_{direct} = (F^T F)^{1/2}$.

What would it happen if instead of using a homonuclear spectrum we use a heteronuclear experiment? For example, let's consider a 2D 13C-1H HSQC-TOCSY experiment. From the above definition of direct Covariance NMR we can simply swap the position of F with F^T leading to the new indirect Covariance NMR:

[4] $C_{indirect} = (FF^T)^{1/2}$.

In analogy with direct Covariance NMR (although working in the opposite direction) we are now *transferring* the information from the indirect dimension (which becomes now the *donor*) to the direct one (*acceptor*), in such a way that the spectral resolution along both frequency axes is now determined by the sampling along the evolution t1 time and consequently, there is a drop in resolution as opposed to what occurs in direct Covariance NMR. Now we start with a N2xN1 (e.g. 2048 x 256) data matrix and as a result of indirect Covariance NMR we obtain a symmetric N1xN1 spectrum (e.g. 256 x 256).

However, the advantage is obvious: Indirect Covariance NMR yields the information contained in an information-rich 13C-13C correlation spectrum between contiguous protonated carbons (Figure 2) without having to detect the 13C nucleus directly. In other words, it brings in a sensitivity increase of 8-fold over a 13C detected experiment.



Cobas C., *NMR Data Evaluation: Review of Covariance Applications*, DOI: 10.3247/SL5Nmr14.001

Figure 2: HSQC-TOCSY spectrum of sucrose (left) and its resulting indirect covariance counterpart (right) which contains essentially the same spin-connectivity information as a 13C-13C TOCSY with direct 13C detection

Indirect Covariance NMR was developed by the group of Rafael Brüchweiler[12] and was the seed for the development of Unsymmetrical Covariance NMR that will be covered in the next section.

It is worth noting that indirect Covariance NMR, as in the case of the direct counterpart, has the propensity of generating artifacts due, for example, to proton-resonance overlap.[13]

Summary: indirect Covariance NMR is used with indirectly detected heteronuclear 2D experiments to yield a homonuclear experiment (i.e. protonated carbon to protonated carbon correlation). This is done in Mnova by issuing the command *Process/Covariance/Indirect Covariance*.

Unsymmetrical Covariance NMR

Up to this point all Covariance NMR methods operated with one individual 2D spectrum. As first demonstrated by G. E. Martin and coworkers,[14,15] a straight extension of the indirect Covariance NMR method would involve introducing in the covariance equation two spectra that share a common frequency axis.

For example, a very practical 2D experiment is the HSQC-COSY 2D NMR spectrum which correlates carbons and protons that are two bonds away (i.e. similar to the H2BC spectrum). Unfortunately, the sensitivity of this experiment is quite poor, usually requiring several hours of acquisition time. On the other hand, the acquisition of the COSY and HSQC experiments separately is much faster, typically in the order of a few minutes. The goal of Unsymmetrical Indirect Covariance NMR is to combine the two individually acquired NMR spectra (e.g. COSY and HSQC) to produce a hyphenated 2D experiment, such as HSQC-COSY. Mathematically this is formulated as in Eq. 5:

[5] $C_{unsymet} = FG^T$.

Where F could be a 1H-13C-HSQC experiment and G a 1H-1H-COSY. Many other combinations are possible such as[16] 1H-13C-HSQC and 1H-1H-NOESY, or 1H-13-HSQC and 1H-13C-HMBC, yielding, in the latter case, a 13C-13C correlation on the basis of J connectivities representing an effective alternative to the direct measurement of a homonuclear 13C-detected spectrum however with greater sensitivity and with considerably less spectrometer time.

Once again, it is important to highlight the fact that in cases of peaks overlap in the proton dimension, artifact peaks can be generated during unsymmetrical indirect covariance processing, so this method works best when applied to small molecules systems.

Summary: Unsymmetrical Covariance NMR methods can be used to mathematically calculate the equivalent of low sensitivity, hyphenated NMR experiments by combining data from a pair of higher sensitivity experiments. Unsymmetrical Covariance NMR has not been available in Mnova. Instead, Generalized Indirect Covariance NMR was implemented in Mnova as it has the same applications but with a superior treatment of potential artifacts (see next section)

Generalized Indirect Covariance NMR

Both direct and indirect Covariance produce symmetric spectral matrices which makes it possible to calculate the square root (the square root of a matrix is properly defined only for symmetric and positive semi-definite matrices). This operation is important, not only to get a spectrum that closely resembles the standard 2D FT one (with the enhanced resolution along the indirect dimension), but also because it helps to minimize artifacts due to relay effects and chemical shift (near) degeneracy.

On the other hand, Unsymmetrical Covariance NMR, according to eq. [5] produces unsymmetric matrices which are therefore not amenable for square rooting. To overcome this limitation Snyder and Brüschweiler developed a new mathematical framework that they termed Generalized Indirect Covariance NMR.[17] The fundamental idea is that starting from the spectra to be co-processed, an intermediate symmetric covariance matrix is generated, which is now amenable to general matrix operations such as square rooting.

The general process can be described as follows using as an example, an HSQC (H) and a COSY (C) spectrum to produce a hyphenated HSQC-COSY spectrum.

First, the two spectra are combined to produce a new stacked spectrum S:

[6] $$S = \begin{bmatrix} H \\ C \end{bmatrix}.$$

Next, S is multiplied by its transpose to yield a generalized covariance matrix C:

[7] $$C = SS^T = \begin{bmatrix} H \\ C \end{bmatrix} [H^T C^T].$$

The key point in here is that matrix C is now symmetric and semipositive definite. This constitutes the main difference with the Unsymmetrical Covariance NMR as now it is possible to apply any arbitrary matrix calculation, including matrix roots (λ) which permits reduction of relay and pseudo-relay effects:

[8] $$S = C^{\lambda} = [(HC)(HC)^T]^{\lambda}.$$

The Generalized Indirect Covariance spectrum S can be efficiently computed using Singular Value Decomposition (SVD) and several values of λ can be employed, although the most common value is λ = 0.5. In fact, using different values of λ can provide a helpful method to detect potential artifact responses, but this is beyond the scope of this article.

GIC can be used in the same applications as UIC.

Recently, G. E. Martin and co-workers have used GIC to create hyphenated HSQC-1,1-ADEQUATE which pinpoints neighboring 13C-13C correlations from the 1,1-ADEQUATE spectrum.[18]

Summary: Generalized Indirect Covariance NMR is equivalent to Unsymmetrical Indirect Covariance but with a more sound mathematical framework.

Doubly Indirect Covariance NMR

A further extension of Covariance NMR is the so-called **Doubly Indirect Covariance** method where HSQC and COSY spectra are co-processed to yield a 2D-INADEQUATE-like experiment that correlates two carbon resonances whose directly attached protons display a cross-peak in the COSY spectrum but at much higher sensitivity of the 13C-13C experiment.

If H is a 1H-13C HSQC and Y is a 1H-1H COSY, Doubly Indirect Covariance is defined as:

[6] $$C = H \cdot Y \cdot H^T.$$

Intuitively, this could be understood as a two-step process: In analogy with UIC or GIC, multiplication of the HSQC (represented by H) by a COSY (represented by Y) would result in a hyphenated HSQC-COSY which in turn is multiplied by the same HSQC (H^T)to produce the final 13C-13C correlation spectrum. Of course, correlations between non protonated carbons will not be observed, although they could be established by replacing one of the HSQC experiments in [9] by an HMBC spectrum.

An example of the application of this technique is shown in Figure 3 in which a COSY and an HSQC are combined via equation [9] to yield the 13C-13C correlation spectrum.

Cobas C., *NMR Data Evaluation: Review of Covariance Applications*, DOI: 10.3247/SL5Nmr14.001

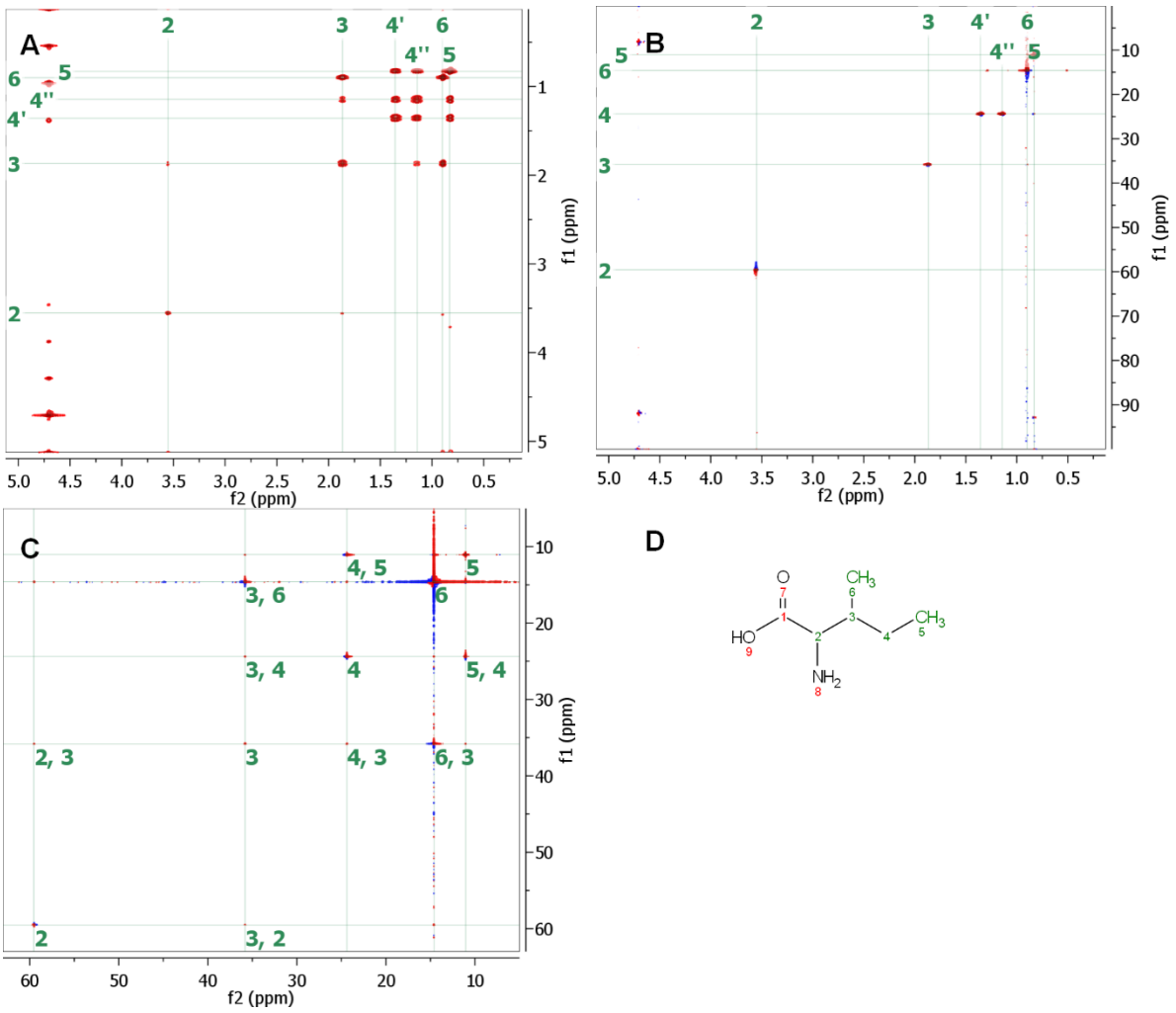


Figure 3. (A) COSY spectrum. (B) HSQC spectrum. Doubly indirect covariance experiment (C) allows carbon-carbon connectivity of isoleucine to be derived (D). Assigned ^{13}C-^{13}C cross-peaks in C show the connections between vicinal carbons with protons attached of isoleucine carbons 2 to 6. (COSY and HSQC spectra courtesy of R. Brüschweiler and F. Zhang)

Summary: Doubly indirect covariance NMR is a generalization of the indirect covariance method that can be used to generate 13C-13C correlation spectrum from a 2D HSQC spectrum and a COSY spectrum.

Conclusion

In this article we have covered the most basic elements and main applications of Covariance NMR. Further details can be found in the excellent reviews by D.A. Snyder and R. Brüschweiler [19] and G.A. Martin.[20]. All these processing algorithms are available in Mnova NMR.

Acknowledgement

The author thanks R. Brüschweiler for the spectra of isoleucine and G. A. Martin for helpful discussions and for kindly providing NMR data sets used to develop and test the covariance algorithms implemented in Mnova. Michael A. Bernstein is also acknowledged for his helpful feedback.

[1] Fengli Zhang, Aaron T. Dossey, Cherian Zachariah, Arthur S. Edison, and Rafael Brüschweiler, Strategy for Automated Analysis of Dynamic Metabolic Mixtures by NMR. Application to an Insect Venom, Analytical Chemistry 2007 79 (20), 7748-7752.
[2] Noda I., Two-dimensional infrared spectroscopy, J. Am. Chem. Soc. 111 (1989) 8116–8118.
[3] Noda I., Generalized 2-dimensional correlation method applicable to infrared, Raman, and other types of spectroscopy, Appl. Spectrosc.47 (1993) 1329–1336.
[4] Noda I., Determination of two-dimensional correlation spectra using the Hilbert transform, Appl. Spectrosc. 54 (2000) 994–999.
[5] Noda I., Dowrey A.E., Marcott C., Story G.M., Ozaki Y., Generalized two-dimensional correlation spectroscopy, Appl. Spectrosc. 54 (2000) 236A–248A.
[6] Eads C.D., Noda I., Generalized correlation NMR spectroscopy, J. Am. Chem. Soc. 124 (2002) 1111–1118.
[7] Hu B.-W., Zhou P., Noda I., Zhao G.-Z., An NMR approach applicable to biomolecular structure characterization, Analytical Chemistry, 77(23), (2005) 7534–8.
[8] Brüschweiler R., Zhang F., Covariance nuclear magnetic resonance spectroscopy, J. Chem. Phys. 120 (2004) 5253.
[9] There are other advantages to the Covariance spectrum such as no phase correction being required along the indirect dimension, but this will not be covered here in any detail.
[10] Trbovic Nikola, Smirnov Segre, Zhang Fengli, Brüschweiler Rafael, Covariance NMR spectroscopy by singular value decomposition, J. Magn. Reson. 171 (2004) 277–283.
[11] Chen Y., Zhang F., Snyder D. A., Gan Z., Brüschweiler-Li L., Brüschweiler R., Quantitative covariance NMR by regularization, J. Biomol. NMR 38 (2007) 73–77.
[12] Zhang F., Brüschweiler R., Indirect covariance NMR spectroscopy, J. Am. Chem. Soc. 126(41), (2004) 13180–1.
[13] Blinov K.A., Larin N.I., Kvasha M.P., Moser A., Williams A.J., Martin G. E., Analysis and elimination of artifacts in indirect covariance NMR spectra via unsymmetrical processing. Magn. Reson. Chem. 43(12), (2005) 999–1007.
[14] Blinov K.A., Larin N.I., Williams A.J., Mills K.A., Martin G.E., Unsymmetrical covariance processing of COSY or TOCSY and HSQC NMR data to obtain the equivalent of HSQC-COSY or HSQC-TOCSY spectra, J. Heterocyclic Chem. 43 (2006) 163–166.
[15] Blinov K.A., Larin N.I., Williams A.J., Zell M., Martin G.E., Long-range carbon–carbon connectivity via unsymmetrical indirect covariance processing of HSQC and HMBC NMR data, Magn. Reson. Chem. 44 (2006) 107–109.
[16] Blinov K.A., Williams A.J., Hilton B.D., Irish P.A., Martin G.E., The use of unsymmetrical indirect covariance NMR methods to obtain the equivalent of HSQC-NOESY data, Magn. Reson. Chem. 45 (2007) 544–546.
[17] Snyder D.A., Brüschweiler R., Generalized indirect covariance NMR formalism for establishment of multidimensional spin correlations, J. Phys. Chem. A, 113(46), (2009) 12898–903.
[18] Snyder D.A., R. Brüschweiler, Generalized Indirect Covariance NMR Formalism for Establishment of Multidimensional Spin Correlations, J. Chem. Phys. A. 113 (2009) 12898–12903.
[19] Snyder D.A., Brüschweiler R., Multidimensional correlation spectroscopy by covariance NMR, in: Multidimensional NMR Methods in the Solution State, G. A. Morris and J. W. Emsley (eds.), John Wiley and Sons, Ltd., Chichester, 2010,pp. 97–105.
[20] G.A. Martin, Using 1,1- and 1,n-ADEQUATE 2D NMR Data in Structure Elucidation Protocols, Annual Report on NMR Spectroscopy, G.A. Webb, Ed., vol. 74, 2011, pp 215-291.