Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 171 (2004) 277–283
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Covariance NMR spectroscopy by singular value decomposition

Nikola Trbovic, Serge Smirnov, Fengli Zhang, Rafael Brüschweiler*

Carlson School of Chemistry and Biochemistry, Clark University, Worcester, MA 01610, USA

Received 8 June 2004; revised 6 August 2004
Available online 2 October 2004

Abstract

Covariance NMR is demonstrated for homonuclear 2D NMR data collected using the hypercomplex and TPPI methods. Absorption mode 2D spectra are obtained by application of the square-root operation to the covariance matrices. The resulting spectra closely resemble the 2D Fourier transformation spectra, except that they are fully symmetric with the spectral resolution along both dimensions determined by the favorable resolution achievable along ω_2. An efficient method is introduced for the calculation of the square root of the covariance spectrum by applying a singular value decomposition (SVD) directly to the mixed time-frequency domain data matrix. Applications are shown for 2D NOESY and 2QF-COSY data sets and computational benchmarks are given for data matrix dimensions typically encountered in practice. The SVD implementation makes covariance NMR amenable to routine applications.

Keywords: Covariance spectroscopy; Homonuclear 2D NMR spectroscopy; Singular value decomposition; Principal component analysis; Hypercomplex data; States; TPPI; COSY; NOESY

1. Introduction

The recently introduced covariance NMR spectroscopy method [1] serves as an alternative to homonuclear 2D Fourier transform (FT) NMR. It produces symmetric spectra with the spectral resolution solely determined by the sampling, apodization, and processing along the detection dimension t_2. The naturally high resolution along the indirect dimension allows a reduction in the number of t_1 increments and thereby the saving of NMR measurement time. It was previously demonstrated [1] how covariance spectra can be obtained from standard 2D NMR time-domain data sets $s(t_1,t_2)$ [2] recorded using time-proportional phase incrementation (TPPI) [3] along the indirect time domain t_1. It is shown here how covariance spectroscopy can be extended to 2D NMR data sets recorded using the hypercomplex method [4]. Moreover, a computationally efficient implementation is introduced, which is based on singular value decomposition (SVD) [5].

In the hypercomplex method by States et al. [4], two 2D data sets are acquired, one with cosine modulation, $s_c(t_1,t_2)$, and one with sine modulation, $s_s(t_1,t_2)$, along t_1. 2D FT processing involves separate complex FT of $s_c(t_1,t_2)$ and $s_s(t_1,t_2)$ along t_2, followed by phase correction and discarding of the imaginary (dispersive) part: $s_{c,s}(t_1,\omega_2) = \mathrm{Re}\int_0^\infty \mathrm{d}t_2 \exp(-\mathrm{i}\omega_2 t_2) s_{c,s}(t_1,t_2)$. A complex signal $A(t_1,\omega_2) = s_c(t_1,\omega_2) + \mathrm{i}s_s(t_1,\omega_2)$ is then constructed and subjected to a complex Fourier transformation along t_1, followed by phase correction along ω_1

$$F(\omega_1,\omega_2) = \int_0^\infty \mathrm{d}t \exp(-\mathrm{i}\omega_1 t_1) A(t_1,\omega_2). \quad (1)$$

The real part of $F(\omega_1,\omega_2)$ exhibits the desired pure absorption features along ω_1 [4].

* Corresponding author. Fax: +1 508 793 8861.
E-mail address: bruschweiler@nmr.clarku.edu (R. Brüschweiler).

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.08.007

2. Covariance NMR

Covariance NMR typically operates on the mixed time-frequency domain data set $A(t_1,\omega_2) = s_c(t_1,\omega_2) + is_s(t_1,\omega_2)$. To reflect the discrete nature of the t_1 and ω_2 parameters in experimental 2D NMR data, we use the following discretization scheme: $t_1 = (k-1)\cdot\Delta t_1$ $(k = 1,\ldots,N_1)$, where N_1 is the total number of t_1 increments and $\omega_2(l) = 2\pi\cdot\nu_2(l) = 2\pi\{-SW/2 + (SW/N_2)\cdot l\}$ $(l = 1,\ldots,N_2)$ where N_2 is the total number of frequency points along ω_2. $\Delta t_1 = (SW)^{-1}$ is the t_1 increment and SW is the spectral width in Hz, $A(t_1,\omega_2)$ can be represented in practice by a complex $N_2\times N_1$ matrix $\boldsymbol{A}$ with elements $A(k,l) = A((k-1)\cdot\Delta t_1,\omega_2(l)) = s_c(k,l) + is_s(k,l)$. The complex covariance matrix $\boldsymbol{C}$ is then given by

$$C_{ij} = \frac{1}{N_1-1}\sum_{k=1}^{N_1}(A(k,i) - \langle A(i)\rangle)^*(A(k,j) - \langle A(j)\rangle) \quad \text{or}$$

$$\boldsymbol{C} = \frac{1}{N_1-1}\tilde{\boldsymbol{A}}^{\dagger}\cdot\tilde{\boldsymbol{A}}, \tag{2}$$

where $\langle A(j)\rangle = N_1^{-1}\sum_{k=1}^{N_1}A(k,j)$ is the value of the average 1D spectrum at frequency $\omega_2(j)$ and $\tilde{\boldsymbol{A}}$ is obtained from $\boldsymbol{A}$ by making it offset free

$$\tilde{A}_{ij} = A_{ij} - \langle A(j)\rangle. \tag{3}$$

Matrix element C_{ij} is the mathematical covariance between the complex amplitudes $s_c(k,i) + is_s(k,i)$ and $s_c(k,j) + is_s(k,j)$ at frequencies $\omega_2(i)$ and $\omega_2(j)$ of the N_1 1D spectra k. This spin correlation information is obtained without the use of a second Fourier transformation along the indirect time domain t_1. The covariance matrix $\boldsymbol{C}$ is Hermitian and has thereby along the second dimension, ω_2', the same high spectral resolution as along the direct dimension ω_2. In contrast to 2D FT, a zeroth order phase correction along t_1, $A'(k,i) = e^{i\varphi_1}A(k,i)$, does not affect $\boldsymbol{C}$ and can therefore be omitted. Depending on the type of 2D NMR experiment and the sample, the covariances C_{ij} can take positive or negative values. The covariance spectrum $\boldsymbol{C}$ therefore qualitatively differs from an absolute value or a power spectrum.

It is shown in the Appendix that for a large class of 2D NMR experiments, including NOESY, TOCSY, and COSY experiments, the real part of $\boldsymbol{C}$ contains all the relevant spectral information, while the imaginary part is essentially zero and contains mainly noise. The real part of $\boldsymbol{C}$ is given by

$$\mathrm{Re}\{\boldsymbol{C}\} = \mathrm{cov}(s_c) + \mathrm{cov}(s_s), \tag{4}$$

where $\mathrm{cov}(s_{c,s}) = \frac{1}{N_1-1}\sum_{k=1}^{N_1}(s_{c,s}(k,i) - \langle s_{c,s}(i)\rangle)(s_{c,s}(k,j) - \langle s_{c,s}(j)\rangle)$.

A real covariance matrix is also obtained when using the TPPI acquisition scheme along t_1 [1]. Both the $A(t_1,\omega_2) = s_{\mathrm{TPPI}}(t_1,\omega_2)$ matrix and the corresponding covariance matrix $\boldsymbol{C}$, calculated using Eq. (4), are real. Using Parseval's theorem it has been shown [6] that for a large class of 2D NMR experiments the real covariance matrix $\boldsymbol{C}$ obtained from Eq. (4) or from a TPPI data set is related to the real part of the 2D FT spectrum $\boldsymbol{F}$ (Eq. (1)) via

$$\boldsymbol{C} \propto \boldsymbol{F}^{\mathrm{T}}\cdot\boldsymbol{F}. \tag{5}$$

By taking the square root of the real part of $\boldsymbol{C}$, a symmetric matrix $\boldsymbol{S} = \boldsymbol{C}^{1/2}$ is obtained, which has high resolution along both dimensions and which otherwise closely resembles the 2D FT spectrum $\boldsymbol{F}$ (see Section 4). The covariance NMR method is complementary to other resolution enhancement methods; it can be combined, for example, with the time-domain linear prediction method [7] by applying linear prediction prior to the covariance processing.

3. SVD formulation

The square root of the covariance matrix, $\boldsymbol{S} = \boldsymbol{C}^{1/2}$, can be calculated by first diagonalizing $\boldsymbol{C}$:

$$\boldsymbol{C}\boldsymbol{v}_j = \lambda_j\boldsymbol{v}_j \quad \text{or} \quad \boldsymbol{C} = \sum_j \lambda_j\boldsymbol{v}_j\cdot\boldsymbol{v}_j^{\dagger}, \tag{6}$$

where λ_j is the real eigenvalue to eigenvector $\boldsymbol{v}_j$. It follows for $\boldsymbol{S}$

$$\boldsymbol{S} = \boldsymbol{C}^{1/2} = \sum_i \sqrt{\lambda_j}\boldsymbol{v}_j\cdot\boldsymbol{v}_j^{\dagger} = \boldsymbol{Q}\cdot\boldsymbol{D}^{1/2}\cdot\boldsymbol{Q}^{\dagger}, \tag{7}$$

where $\boldsymbol{Q}$ contains the eigenvectors $\boldsymbol{v}_j$ as columns and $\boldsymbol{D}$ is a diagonal matrix, which contains the eigenvalues λ_j of $\boldsymbol{C}$ as its diagonal elements. Because $\boldsymbol{C}$ is a covariance matrix, all eigenvalues fulfill $\lambda_j \geqslant 0$ and a maximum of N_1 eigenvalues are not equal to zero. In what follows, Eqs. (6) and (7) are referred to as the "standard method."

The computation of the covariance matrix requires $O(N_1N_2^2/2)$ floating point operations, whereas the square-root operation based on diagonalization of $\boldsymbol{C}$ requires $O(N_2^3)$ floating point operations. For some of the larger 2D NMR data matrices encountered in practice, the computational effort can be quite significant even on a modern computer workstation (see Tables 1 and 2). Therefore, a more efficient computational approach is desirable. In practice, the number of t_1 increments, N_1, is often much smaller than N_2. In this case, a significant speed-up for the computation of $\boldsymbol{S}$ can be obtained by applying a singular value decomposition (SVD) [5] to the transposed real mixed time-frequency matrix $\tilde{\boldsymbol{A}}^{\mathrm{T}}$:

$$\tilde{\boldsymbol{A}}^{\mathrm{T}} = \boldsymbol{U}\cdot W\cdot\boldsymbol{V}^{\mathrm{T}}, \tag{8}$$

where $\boldsymbol{U}$ is a $N_2\times N_1$ matrix, $\boldsymbol{V}$ is a $N_1\times N_1$ matrix provided that $N_1 \leqslant N_2$. $\boldsymbol{W}$ is a diagonal $N_1\times N_1$ matrix with real singular values $w_i \geqslant 0$ as its diagonal elements. Matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are both orthogonal ($\boldsymbol{V}^{\mathrm{T}}\cdot\boldsymbol{V} = \boldsymbol{1}$,

Table 1
Computational performance for TPPI data in seconds

Method	N_2(real)	256	512	1024				2048					4096
	N_1 (real)	256	512	128	256	512	1024	128	256	512	1024	2048	1024
Standard covariance method	Covariance[a]	<0.1	0.4	0.3	0.9	1.9	3.7	1.9	3.9	7.7	15.0	29.7	60.4
	Diag[b]	0.2	1.5	12.8	11.7	10.3	12.9	106	101	91.3	85.6	88.2	857
	Reconstruct[c]	0.1	0.7	0.7	1.4	2.7	5.5	2.9	5.6	11.1	21.9	45.0	88.4
	Total[d]	0.3	2.7	13.9	14.0	15.2	22.6	111	111	111	124	165	1009
Direct SVD method	SVD[e]	0.2	2.2	0.2	1.2	6.3	20.0	0.7	2.53	11.5	55.2	132	98.7
	Reconstruct[f]	0.1	0.7	0.6	1.3	2.7	5.3	2.8	5.4	10.8	21.3	43.3	83.6
	Total[d]	0.3	3.0	0.9	2.6	9.3	25.8	3.6	8.2	22.9	77.9	178	184

[a] Time for construction of covariance matrix.
[b] Time for diagonalization of covariance matrix.
[c] Time for reconstruction of square root of covariance matrix from eigenvectors and eigenvalues.
[d] Total computational time (including reading and writing of input and output files).
[e] Time for diagonalization of input mixed time-frequency data matrix via SVD (Eq. (10)).
[f] Time for reconstruction of spectrum (from eigenvectors and singular values).

Table 2
Computational performance for hypercomplex data in seconds

Method	N_2(real)	256	512	1024				2048					4096
	N_1(complex)	128	256	64	128	256	512	64	128	256	512	1024	512
Standard covariance method	Covariance[a]	<0.1	0.3	0.3	0.6	1.3	2.5	1.5	2.8	5.3	10.3	20.1	42.3
	Diag[b]	0.2	1.6	12.7	11.6	10.4	13.6	106	103	92.3	86.2	113	853
	Reconstruct[c]	<0.1	0.7	0.7	1.4	2.7	5.5	2.9	5.7	11.1	22.1	45.6	88.8
	Total[d]	0.3	2.7	13.7	13.7	14.7	22.2	111	112	109	120	181	988
Direct SVD method	SVD[e]	0.1	1.1	0.1	0.5	2.5	13.2	0.3	1.4	5.1	22.9	113	44.9
	Reconstruct[f]	0.1	0.6	0.7	1.3	2.7	5.4	2.9	5.8	11.3	22.0	43.8	87.1
	Total[d]	0.2	1.8	0.8	1.9	5.3	19.0	3.3	7.4	16.8	45.8	158	134

[a] Time for construction of complex covariance matrix.
[b] Time for diagonalization of real part of complex covariance matrix.
[c] Time for reconstruction of square root of real part of complex covariance matrix from eigenvectors and eigenvalues.
[d] Total computational time (including reading and writing of input and output files).
[e] Time for diagonalization of input mixed time-frequency data matrices via SVD (Eq. (10)).
[f] Time for reconstruction of spectrum (from eigenvectors and singular values).

$\boldsymbol{U}^{\mathrm{T}} \cdot \boldsymbol{U} = \boldsymbol{I}$) and $\boldsymbol{U}$ contains as its columns the eigenvectors $\boldsymbol{v}_j$ of $\boldsymbol{C}$, as can be seen from the following identity:

$$\boldsymbol{C} = \tilde{\boldsymbol{A}}^{\mathrm{T}} \cdot \tilde{\boldsymbol{A}} = \boldsymbol{U} \cdot W \cdot \boldsymbol{V}^{\mathrm{T}} \cdot \boldsymbol{V} \cdot W \cdot \boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U} \cdot W^2 \cdot \boldsymbol{U}^{\mathrm{T}}. \quad (9)$$

It further follows that the squares of the diagonal elements of W, w_i^2, correspond to the (non-zero) eigenvalues of $\boldsymbol{C}$. For the square root of $\boldsymbol{C}$ one simply obtains

$$\boldsymbol{S} = (\tilde{\boldsymbol{A}}^{\mathrm{T}} \cdot \tilde{\boldsymbol{A}})^{1/2} = \boldsymbol{U} \cdot W \cdot \boldsymbol{U}^{\mathrm{T}}. \quad (10)$$

Eq. (10) provides a straightforward recipe for the efficient calculation of the square root of the real covariance spectrum without requiring the explicit computation of the covariance matrix of Eq. (2) in the first place. Moreover, Eq. (9) corresponds to a principal component analysis (PCA) with the eigenvectors of $\boldsymbol{C}$ containing useful information, for example for the spectral deconvolution of chemical mixtures [8]. Eq. (10) can be generalized for complex (i.e., hypercomplex or States-type) data by replacing the transpose by the adjoint operation. The SVD of a complex mixed time-frequency matrix yields the square root of the complex covariance matrix, which is equal to the square root of the real part if the imaginary part is zero. However, due to the presence of noise experimental NMR data possess a non-zero imaginary part, which affects the square root of the complex covariance matrix resulting in small spectral artifacts. It is found that the square root of the real part of the complex covariance matrix is closely related to the sum of the two individual square roots of the covariance matrices derived from the cosine and sine modulated data sets. Those square roots can then be efficiently obtained by the application of separate SVDs to the two data sets. This treatment proves to be an excellent approximation, which is void of the artifacts exhibited by the square root of the complex covariance matrix. The key advantage of the SVD method over the diagonalization method is its computational efficiency since it requires only $O(N_1^2 N_2)$ operations. Therefore, it is considerably faster whenever N_1 is sufficiently

smaller than N_2, a situation that is often encountered in practice.

The sequential steps of the covariance processing of a hypercomplex time-domain data set consisting of $s_c(t_1,t_2)$ and $s_s(t_1,t_2)$ can be summarized as follows:

1. After complex Fourier transformation of $s_c(t_1,t_2)$ and $s_s(t_1,t_2)$ with respect to t_2 and phase and baseline correction in ω_2, the real parts $s_c(t_1,\omega_2) = \mathrm{Re}\{\mathrm{FT}_{t_2}(s_c(t_1,t_2))\}$ and $s_s(t_1,\omega_2) = \mathrm{Re}\{\mathrm{FT}_{t_2}(s_s(t_1,t_2))\}$ are retained.
2. $s_c(t_1,\omega_2)$ and $s_s(t_1,\omega_2)$ are made offset free (Eq. (3)) before they are separately subjected to the SVD and the matrix square-root construction operation (Eqs. (9) and (10)), yielding $S_c(\omega_2',\omega_2)$ and $S_s(\omega_2',\omega_2)$.
3. The final 2D covariance spectrum is then calculated as the sum $S(\omega_2',\omega_2) = S_c(\omega_2',\omega_2) + S_s(\omega_2',\omega_2)$.

For a TPPI-type time-domain data set $s(t_1,t_2) = s_{\mathrm{TPPI}}(t_1,t_2)$, processing is slightly simpler:

1. After complex Fourier transformation of $s(t_1,t_2)$ with respect to t_2 and phase and baseline correction in ω_2, the real part $s(t_1,\omega_2) = \mathrm{Re}\{\mathrm{FT}_{t_2}(s(t_1,t_2))\}$ is retained.
2. $s(t_1,\omega_2)$ is made offset free (Eq. (3)) before it is subjected to the SVD and the matrix square-root construction operation (Eqs. (9) and (10)), yielding the final covariance spectrum $S(\omega_2',\omega_2)$.

4. Results and discussion

The direct SVD method is compared with the standard covariance method by applying it to 2D 2QF-COSY [9] and NOESY data [10–12] recorded on a 8 mM sample of human ubiquitin at pH 6.5 in D_2O purchased from Sigma–Aldrich (St. Louis, MO). The 2QF-COSY experiment was collected in hypercomplex States-TPPI mode [3,4] with $N_2 = 1024$ complex points along t_2 and $N_1 = 512$ complex points along t_1. The NOESY experiment was collected with a mixing time of 200 ms with $N_2 = 1024$ complex points along t_2 and $N_1 = 1024$ real points along t_1 for the TPPI case [3] and $N_1 = 512$ complex points in the States-TPPI case [3,4]. All experiments were collected at 298 K at 600 MHz proton frequency. The time-domain data were zero-filled to 2048 complex points along the t_2 dimension, apodized by a cosine bell function, and Fourier transformed. For 2D FT processing, apodization along the t_1 dimension was applied using a cosine bell function. Spectra with a data matrix size smaller than 2048×1024 (real + imaginary) points were generated by truncation of the original data matrix to the final size. Spectra with a data matrix size larger than 2048×1024 (real + imaginary) points were generated by zero filling of the original data matrix to the final size. Apodization, phase correction, and Fourier transformation was performed using NMRPipe [13]. 2D NMR spectra were visualized using SPARKY 3 (T.D. Goddard, D.G. Kneller, University of California, San Francisco).

Computational benchmarks were obtained for different data matrix dimensions for the standard covariance method using Eq. (7) and the direct SVD method using Eq. (10) for TPPI (Table 1) and States-TPPI data (Table 2). In the case of the standard method covariance matrices were diagonalized using the CLAPACK [14] subroutine optimized for symmetric matrices, whereas in the case of the direct SVD method the asymmetric mixed time-frequency matrices were processed by the CLAPACK SVD subroutine. All calculations were performed on an AMD Opteron 64 bit 3 GHz computer with 2 GBytes of RAM.

In the standard method the computationally most expensive part is the matrix diagonalization, which in terms of computational time is followed by the reconstruction of the square-root spectrum and by the calculation of the covariance matrix itself. For example, for a 2048×512 hypercomplex mixed time-frequency input matrix, computation of the covariance matrix takes 10.3 s while matrix diagonalization takes 86.2 s. The direct SVD method, on the other hand, avoids the explicit construction of the covariance matrix and takes advantage of the improved efficiency of SVD over diagonalization when $N_1 < N_2$. SVD of the same data matrix, i.e., the two separate SVDs of the cosine and sine modulated matrices, takes only 22.9 s and the total computational time is 45.8 s (Table 2), which represents a 2.6-fold speed-up over the standard method. In the case of $N_1 = 256$ and $N_2 = 2048$ the SVD takes 5.1 s resulting in a total computational time of 16.8 s, which corresponds to a 6.5-fold speed-up over the standard covariance method taking a total computational time of 109 s. Although 2D FT processing is typically by a factor 5–20 faster, the speed-up achieved by the SVD procedure turns covariance NMR spectroscopy into a computationally affordable general purpose NMR processing method.

While the standard diagonalization method performs equally well for TPPI and hypercomplex input matrices of equivalent size, for States-TPPI data the direct SVD method is faster than for TPPI data. The difference results from the fact that for a 2048×512 hypercomplex input matrix two real SVDs of 2048×512 matrices are performed, while the TPPI case involves one real SVD of a 2048×1024 input matrix. Since the SVD scales with N_1^2, the SVD in the hypercomplex case saves a factor of 2 in computational time.

A comparison of 2D spectra obtained by the covariance method and by 2D FT is given in Figs. 1 and 2. Fig. 1 shows the same section of the 2QF-COSY experiment of ubiquitin processed by different methods. Panel A shows the 2D FT spectrum, panel B depicts the square root of the covariance spectrum calculated from the 2D FT spectrum (Eq. (5)), and panel C shows the covariance

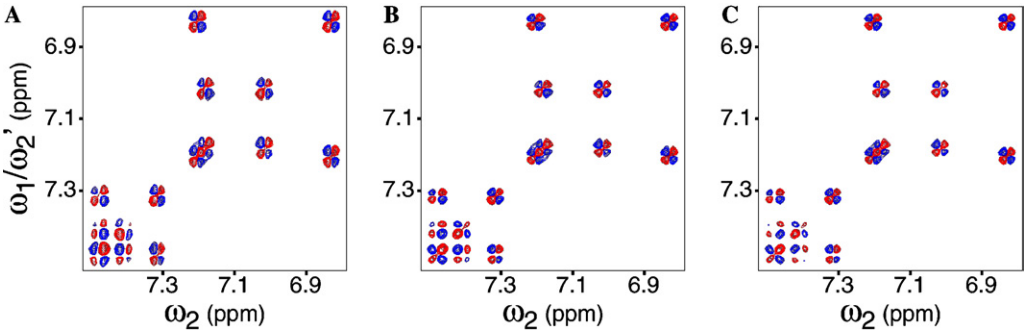

Fig. 1. Comparison of aromatic sub-region of a 2D 2QF-COSY experiment of ubiquitin processed by 2D FT and covariance methods, respectively. Panel A was processed using the standard 2D FT method, panel B depicts the covariance spectrum calculated from the 2D FT spectrum using Eq. (5) followed by the matrix square-root operation, and panel C shows the covariance spectrum calculated from the mixed time-frequency data (Eq. (10)).

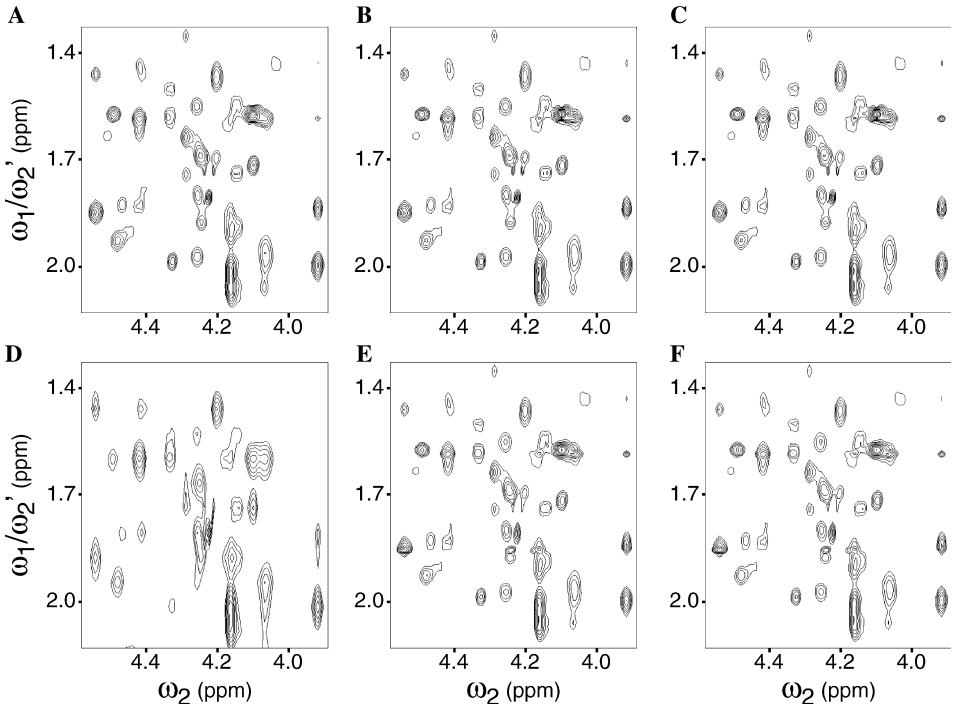

Fig. 2. Comparison of spectral H^N–H^α region of a 2D NOESY experiment of ubiquitin with 200 ms mixing time processed by different methods. Panels A and D were processed using the standard 2D FT method, panels B and E depict the covariance spectrum calculated by the standard covariance method, and panels C and F were processed using the direct SVD method (Eq. (10)). Panels A, B, and C were calculated using 512 complex t_1 increments, whereas panels D, E, and F were processed using only the first 192 complex t_1 increments.

spectrum computed from the mixed time-frequency data (Eq. (10)). The three spectra are very similar, as is expected from the theory, displaying phase sensitivity along both dimensions for both diagonal and cross-peaks.

Fig. 2 shows the same H^N–H^α region of the 2D NOESY experiment collected using States-TPPI processed by different methods. Panel A was processed by 2D FT, panel B by the standard covariance method (Eq. (7)), and panel C by the SVD-based covariance method (Eq. (10)). Again, all three spectra are very similar. The improvement in resolution along ω_2'/ω_1 provided by the covariance method becomes apparent in panels D, E, and F, which show the same spectral region processed using only the first $N_1 = 192$ t_1 increments.

5. Conclusion

The extension of covariance NMR spectroscopy to hypercomplex 2D NMR data makes this method accessible to a wide range of experimental NMR data. The square-root operation applied to the covariance matrix produces spectra that closely resemble the 2D FT spectra, except that they benefit from resolution enhancement along ω_2' without requiring apodization or phase correction along this dimension. The direct SVD method presented here is an efficient tool for covariance NMR processing applicable to both hypercomplex and TPPI-type data. It substantially reduces the computational time for rectangular 2D NMR input matrices typ-

ically encountered in practice. The SVD implementation turns covariance NMR into a viable alternative to 2D FT NMR in the liquid and in the solid state that is suitable for routine applications. Covariance NMR spectroscopy along with its SVD implementation is readily applicable to homonuclear (but generally not to heteronuclear) 2D planes in homo- and heteronuclear 3D and 4D NMR experiments. A SVD-based subroutine for covariance NMR processing will be made available on our website.

Acknowledgments

N.T. acknowledges a short-term fellowship of the Studienstiftung (Germany). This work was supported by Grants GM066041 (NIH) and MRI-0320875 (NSF).

Appendix A

In this appendix, the covariance NMR spectrum is explicitly calculated for a generic homonuclear 2D NMR data set collected using the following pulse scheme [2]:

$$90^\circ_{y,x} - t_1 - 90^\circ_{-y} - \text{mixing} - 90^\circ_{y} - t_2, \qquad \text{(Scheme A.1)}$$

where frequency discrimination along t_1 is achieved by the hypercomplex method [4] applied to the excitation pulse. During the mixing period magnetization is exchanged between the Zeeman magnetizations of spins 1/2 with individual resonance frequencies ω_k ($k = 1, \ldots, N$) according to a symmetric transfer matrix $\boldsymbol{T}$ with elements T_{lk} that denote the amount of magnetization transferred from spin k to spin l during the mixing process. The transfer can be due to, for example, isotropic mixing (TOCSY) or cross-relaxation (NOESY). After the second 90° pulse, the longitudinal magnetizations are, depending on whether the excitation pulse phase is x or y,

$$\sigma_c(t_1) = \sum_{k=1}^{N} \cos(\omega_k t_1) I_{kz} \quad \text{or} \quad \sigma_s(t_1) = \sum_{k=1}^{N} \sin(\omega_k t_1) I_{kz}, \tag{A.1}$$

respectively. After the mixing period, the density operators are $\sigma_c(t_1,\tau_m) = \sum_{k,l=1}^{N} T_{lk} \cos(\omega_k t_1) I_{lz}$ and $\sigma_s(t_1,\tau_m) = \sum_{k,l=1}^{N} T_{lk} \sin(\omega_k t_1) I_{lz}$, respectively. The final 90°_y pulse converts the I_{kz} operators into transverse I_{kx} operators, which precess during the detection time t_2 yielding signal matrices according to

$$\begin{aligned} s_c(t_1,\tau_m,t_2) &= \sum_{k,l=1}^{N} T_{lk} \cos(\omega_k t_1) e^{i\omega_l t_2} \quad \text{and} \\ s_s(t_1,\tau_m,t_2) &= \sum_{k,l=1}^{N} T_{lk} \sin(\omega_k t_1) e^{i\omega_l t_2}, \end{aligned} \tag{A.2}$$

where transverse T_2 relaxation has been neglected. Fourier transformation along t_2, phase correction, and discarding of the imaginary part yields

$$\begin{aligned} s_c(t_1,\tau_m,\omega_2) &= \pi \sum_{k,l=1}^{N} T_{lk} \cos(\omega_k t_1) \delta(\omega_2 - \omega_e), \\ s_s(t_1,\tau_m,\omega_2) &= \pi \sum_{k,l=1}^{N} T_{lk} \sin(\omega_k t_1) \delta(\omega_2 - \omega_e), \end{aligned} \tag{A.3}$$

which can be combined in form of the complex data set

$$s(t_1,\tau_m,\omega_2) = s_c(t_1,\tau_m,\omega_2) + i s_s(t_1,\tau_m,\omega_2). \tag{A.4}$$

2D FT processing proceeds with a complex FT along t_1, which yields the 2D spectrum

$$F(\omega_1,\tau_m,\omega_2) = \pi^2 \sum_{k,l=1}^{N} T_{lk} \delta(\omega_1 - \omega_k) \delta(\omega_2 - \omega_e). \tag{A.5}$$

By contrast, in covariance spectroscopy the covariances between pairs of t_1 columns to frequencies ω_2 and ω_2' are given by Eq. (2)

$$\begin{aligned} C(\omega_2',\omega_2) = \big\langle \big(& s(t_1,\tau_m,\omega_2') \\ & - \langle s(t_1,\tau_m,\omega_2') \rangle \big)^{\dagger} \big(s(t_1,\tau_m,\omega_2) \\ & - \langle s(t_1,\tau_m,\omega_2) \rangle \big) \big\rangle, \end{aligned} \tag{A.6}$$

where the angular brackets indicate averaging over t_1. Insertion of Eq. (A.4) into Eq. (A.6) yields

$$\begin{aligned} C(\omega_2',\omega_2) = & \langle s_c(t_1,\tau_m,\omega_2') s_c(t_1,\tau_m,\omega_2) \rangle \\ & - \langle s_c(t_1,\tau_m,\omega_2') \rangle \langle s_c(t_1,\tau_m,\omega_2) \rangle \\ & + \langle s_s(t_1,\tau_m,\omega_2') s_s(t_1,\tau_m,\omega_2) \rangle \\ & - \langle s_s(t_1,\tau_m,\omega_2') \rangle \langle s_s(t_1,\tau_m,\omega_2) \rangle \\ & + i \langle s_c(t_1,\tau_m,\omega_2') s_s(t_1,\tau_m,\omega_2) \rangle \\ & - i \langle s_c(t_1,\tau_m,\omega_2') \rangle \langle s_s(t_1,\tau_m,\omega_2) \rangle \\ & - i \langle s_s(t_1,\tau_m,\omega_2') s_c(t_1,\tau_m,\omega_2) \rangle \\ & + i \langle s_s(t_1,\tau_m,\omega_2') \rangle \langle s_c(t_1,\tau_m,\omega_2) \rangle. \end{aligned} \tag{A.7}$$

After insertion of Eq. (A.3) into Eq. (A.7) one obtains

$$\begin{aligned} \langle s_c(t_1,\tau_m,\omega_2') s_c(t_1,\tau_m,\omega_2) \rangle &= \pi^2 \sum_{lkl'k'} T_{lk} T_{l'k'} \delta(\omega_2 - \omega_l) \\ &\quad \times \delta(\omega_2' - \omega_{l'}) \tfrac{1}{2} (\delta_{\omega_k \omega_{k'}} + \delta_{\omega_k - \omega_{k'}}), \\ \langle s_s(t_1,\tau_m,\omega_2') s_s(t_1,\tau_m,\omega_2) \rangle &= \pi^2 \sum_{lkl'k'} T_{lk} T_{l'k'} \delta(\omega_2 - \omega_l) \\ &\quad \times \delta(\omega_2' - \omega_{l'}) \tfrac{1}{2} (\delta_{\omega_k \omega_{k'}} - \delta_{\omega_k - \omega_{k'}}), \end{aligned} \tag{A.8}$$

where $\langle \cos(\omega_k t_1) \cos(\omega_{k'} t_1) \rangle \cong \frac{1}{2} (\delta_{\omega_k \omega_{k'}} + \delta_{\omega_k - \omega_{k'}})$ and $\langle \sin(\omega_k t_1) \sin(\omega_{k'} t_1) \rangle \cong \frac{1}{2} (\delta_{\omega_k \omega_{k'}} - \delta_{\omega_k - \omega_{k'}})$ was used and $\delta_{\omega_k \omega_{k'}}$ is the Kronecker delta. Using $\langle \cos(\omega_k t_1) \rangle \cong \langle \sin(\omega_k t_1) \rangle \cong \langle \cos(\omega_k t_1) \sin(\omega_{k'} t_1) \rangle \cong 0$, one then obtains

$$C(\omega_2', \omega_2) \cong \langle s_c(t_1, \tau_m, \omega_2') s_c(t_1, \tau_m, \omega_2) \rangle + \langle s_s(t_1, \tau_m, \omega_2') s_s(t_1, \tau_m, \omega_2) \rangle = \pi^2 \sum_{ll'k} T_{lk} T_{l'k} \delta(\omega_2 - \omega_l) \delta(\omega_2' - \omega_{l'}), \quad \text{(A.9)}$$

reflecting that the covariance cross-peak intensity at position $(\omega_2', \omega_2) = (\omega_{l'}, \omega_l)$ is $\pi^2 \sum_k T_{lk} T_{kl'}$. Therefore, the real part of the covariance spectrum $C(\omega_2', \omega_2)$ is related to the 2D FT spectrum $\boldsymbol{F}$ (Eq. (A.5)) via Eq. (5)

$$\boldsymbol{C} = \frac{1}{\pi^2} \boldsymbol{F}^\dagger \boldsymbol{F}. \quad \text{(A.10)}$$

In analogy to the TPPI case, the square root of the real part of the covariance spectrum of hypercomplex data closely resembles the 2D FT spectrum. The square-root operation can again be performed either by diagonalization of the real part of the covariance matrix or by SVD of the mixed time-frequency data. It can be seen from the above theoretical treatment that the imaginary part of the covariance matrix is zero for 2D NMR experiments following (Scheme A.1).

References

[1] R. Brüschweiler, F. Zhang, Covariance nuclear magnetic resonance spectroscopy, J. Chem. Phys. 120 (2004) 5253–5260.

[2] R.R. Ernst, G. Bodenhausen, A. Wokaun, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon, Oxford, 1987.

[3] D. Marion, K. Wüthrich, Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H–^{1}H spin–spin coupling constants in proteins, Biochem. Biophys. Res. Commun. 113 (1983) 967–974.

[4] D.J. States, R.A. Haberkorn, D.J. Ruben, A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants, J. Magn. Reson. 48 (1982) 286–292.

[5] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C, Cambridge University Press, Cambridge, 1988.

[6] R. Brüschweiler, Theory of covariance NMR spectroscopy, J. Chem. Phys. 121 (2004) 409–414.

[7] H. Barkhuijsen, R. de Beer, W.M.M.J. Bovee, D. van Ormondt, Retrieval of frequencies, amplitudes, damping factors and phases from time-domain signals, using a linear least-squares procedure, J. Magn. Reson. 61 (1985) 465–481.

[8] F. Zhang, R. Brüschweiler, Spectral deconvolution by covariance NMR spectroscopy, ChemPhysChem 5 (2004) 794–796.

[9] M. Rance, O.W. Sorensen, G. Bodenhausen, G. Wagner, R.R. Ernst, K. Wüthrich, Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering, Biochem. Biophys. Res. Commun. 117 (1983) 479–485.

[10] J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst, Investigation of exchange processes by two-dimensional NMR spectroscopy, J. Chem. Phys. 71 (1979) 4546–4553.

[11] S. Macura, R.R. Ernst, Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy, Mol. Phys. 41 (1980) 95–117.

[12] A. Kumar, R.R. Ernst, K. Wüthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules, Biochem. Biophys. Res. Commun. 95 (1980) 1.

[13] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J. Biomol. NMR 6 (1995) 277–293.

[14] E. Anderson, LAPACK Users Guide, third ed., Society for Industrial and Applied Mathematics, Philadelphia, 1999.