

Published in final edited form as:

*J Magn Reson.* 2011 March ; 209(1): 75–78. doi:10.1016/j.jmr.2010.11.018.

## A Covariance NMR Toolbox for MATLAB and OCTAVE

Timothy Short<sup>a</sup>, Leigh Alzapiedi<sup>a</sup>, Rafael Brüschweiler<sup>b</sup>, and David Snyder<sup>a,\*</sup>

<sup>a</sup> Department of Chemistry, College of Science and Health, William Paterson University, 300 Pompton Road, Wayne, NJ 07470

<sup>b</sup> Chemical Sciences Laboratory, Department of Chemistry and Biochemistry, and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306

### Abstract

The Covariance NMR Toolbox is a new software suite that provides a streamlined implementation of covariance-based analysis of multi-dimensional NMR data. The Covariance NMR Toolbox uses the MATLAB or, alternatively, the freely available GNU OCTAVE computer language, providing a user-friendly environment in which to apply and explore covariance techniques. Covariance methods implemented in the toolbox described here include direct and indirect covariance processing, 4D covariance, generalized indirect covariance (GIC), and Z-matrix transform. In order to provide compatibility with a wide variety of spectrometer and spectral analysis platforms, the Covariance NMR Toolbox uses the NMRPipe format for both input and output files. Additionally, datasets small enough to fit in memory are stored as arrays that can be displayed and further manipulated in a versatile manner within MATLAB or OCTAVE.

### Keywords

Nuclear Magnetic Resonance; Resolution Enhancement; Sensitivity Enhancement; Generalized Indirect Covariance

### Introduction

Nuclear Magnetic Resonance (NMR) is a powerful technique for elucidating the connectivity, configuration, conformation, and dynamics of both small and large molecules. NMR experiments can probe through-bond and through-space interactions between a variety of nuclei, and multi-dimensional NMR can probe the interactions involving two, three, four or even five nuclei. However, NMR experiments are often limited by sensitivity and resolution. In particular, the acquisition of NMR data that correlate relatively insensitive nuclei or the collection of higher dimensional NMR spectra requires relatively long measurement times to achieve reasonable sensitivity and resolution along the indirect dimensions [i,ii].

A number of variations and extensions to the standard Fourier transform method to NMR data processing are available to leverage the information of a given collection of NMR data to obtain enhanced resolution or sensitivity [iii,iv,v,vi]. Covariance NMR [vii] takes

\*To whom correspondence should be addressed: Department of Chemistry, College of Science and Health, William Paterson University, 300 Pompton Road, Wayne, NJ 07470, U.S.A., snyderd@wpunj.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

advantage of the inherent symmetry present in many NMR experiments. Direct covariance [viii], which maps the high resolution direct (detection) dimension onto the (corresponding) indirect dimension, endows the “donor dimension(s)” with the same high resolution possessed by the “acceptor dimension(s)”. Indirect covariance NMR [ix] establishes new types of inherently symmetric correlations. The concept of covariance NMR can be further generalized by combining multiple NMR experiments to reconstruct data sets for which direct, experimental measurement would be unfeasible. This is accomplished by unsymmetric covariance NMR [x,xi], GIC [xii], and hyperdimensional NMR [xiii] and related methods [xiv,xv].

Much of the software currently available for the processing of NMR spectra using covariance techniques is geared toward particular covariance NMR applications and approaches. Here we present a Covariance NMR Toolbox – compatible with both the MATLAB and the freely available OCTAVE computing environments – that implements all of the major covariance techniques, including direct covariance [8], indirect covariance [9], GIC [12] and the Z-matrix transform [xvi]. The Covariance NMR Toolbox, which is available via the MATLAB Central File Exchange (<http://www.mathworks.com/matlabcentral/fileexchange/>) or by request from the authors, reads and writes data in the well-documented and spectrometer independent NMRPipe [xvii] format and uses a small number of easy to use and extensible scripts to manipulate frequency domain data. Moreover, datasets expected to fit in memory (which includes most multidimensional datasets except 4D covariance spectra) are stored as MATLAB/OCTAVE arrays, which can be displayed and further manipulated in a versatile manner within the MATLAB or OCTAVE computing environments.

## Results and Discussion

Table 1 lists the scripts comprising the Covariance NMR Toolbox. Scripts listed as “helper functions” will not usually be called by the user but are invoked by other scripts in the Covariance NMR Toolbox. The Covariance NMR Toolbox includes a user manual tabulating every option for each user function, and typing *help <function name>* at the MATLAB or OCTAVE prompt provides detailed assistance for each function in the toolbox. A typical session using the Covariance NMR Toolbox begins by calling the function *read\_nmrp* which reads an NMRPipe format file, creating an array to store the spectrum and representing the information found in the NMRPipe file header in tabular form. Donor dimensions can be distinguished from acceptor dimensions either as an argument to *read\_nmrp* or by using either one of the functions *reset\_donors* or *swap\_donor\_accp* once the spectrum is read. For example, following invocation of *A\_axes\_out = reset\_donors(A\_axes, 1)*, using *A\_axes\_out* as an argument to the function *covar* (*vide infra*), will perform indirect rather than direct covariance on the corresponding spectrum *A*.

The primary function for performing covariance NMR is the function *covar*. Invocation of the command *[C C\_axes] = covar(A, B, 'axes1', A\_axes, 'axes2', B\_axes, 'power', λ)* performs the covariance transformation  $[A*B]^λ$  (using the GIC notation described previously [12]) and, depending on the size of the resulting covariance spectrum, either stores that spectrum as a MATLAB array *C* (with the information required to write *C* out as an NMRPipe format file tabulated in *C\_axes*) or stores in *C* the name of a temporary file storing the covariance result as its singular value decomposition. Invocation of the command *[C C\_axes] = covar(A, B, 'axes1', A\_axes, 'axes2', B\_axes, 'power', λ, 'Z', 1)* produces a Z-matrix [16] instead of an unscaled covariance spectrum. Invocation of *covar* with only a single spectrum provided performs symmetric covariance processing, either direct or indirect, depending on which dimension(s) are identified as donor dimensions.

Invocation of the command `write_covar2pipe(C, C_axes, 'covar_out.pipe')` writes the covariance result to an NMRPipe format file `covar_out.pipe`: this syntax is used whether `C` is an MATLAB/OCTAVE array representation of the covariance spectrum or whether `C` points to a temporary file. Additional functionality of this toolbox includes masking, foot-printing [2] and regularization [xviii] prior to covariance and PCA analysis of peak-lists in GIC spectra [12].

Some of the covariance processing options of the Covariance NMR Toolbox are illustrated in Figures 1 and 2 for NMR spectra of the alanine-phenylalanine dipeptide (AF dipeptide) in D<sub>2</sub>O measured at 400 MHz field strength. Figure 1 demonstrates the resolution gain provided by application of direct covariance to 2D FT TOCSY spectra with (Fig. 1A,C) 512 and (Fig. 1B,D) 64 points in the indirect dimension. Figure 2 displays a novel correlation uncovered by generalized indirect covariance. Neither HMBC spectra [xix] (Fig. 2A, showing the expected 2 to 4 bond <sup>1</sup>H-<sup>13</sup>C correlations) nor TOCSY [xx] (which probes protons within the same “spin-system”) spectra can sensitively correlate distal parts of the benzene ring in phenylalanine to the phenylalanine  $\alpha$  proton. However, GIC between an HMBC spectrum and a TOCSY spectrum efficiently correlates an aromatic  $\epsilon$  carbon with its intra-residue  $\alpha$  proton (Fig. 2B). The magnitude only nature of the HMBC spectrum precludes interpretation of the Z-transform spectrum in terms of normally distributed Z-scores. Even in such cases, Z-transformation sometimes is useful to compress the dynamics range found in GIC spectra produced from Fourier transform spectra that differ greatly in intensity, although dynamic range is not a concern with the example dataset used here.

The scripts used to produce the results shown in Figures 1 and 2 are given in Figure 3, and each illustrates a typical session using the Covariance NMR Toolbox as described above. Calculation of the direct covariance result shown in Fig. 1A took 5.2 seconds (the direct covariance result shown in Fig. 1B with  $N_1 = 64 \ll 512$  took less than a second to calculate) while calculation of the GIC spectrum shown in Fig. 2B took 48.1 seconds.

While the results shown above provide examples of the functionality present in the Covariance NMR Toolbox, this toolbox can perform a broad range of covariance tasks, including covariance of 4D NMR [2,xxi]. Depending on the amount of memory available, larger data sets can present memory challenges, requiring the Covariance NMR Toolbox to store data in temporary files leading to longer processing times: in such cases, covariance processing of a 4D spectrum may take over an hour rather than a few seconds. Additionally, some memory intensive functions, such as the Z-matrix transform, may not work on systems with insufficient memory. Using OCTAVE rather than MATLAB also generally results in longer execution times, but even applying the Covariance NMR Toolbox to a 4D data set in OCTAVE is not a prohibitively time-consuming undertaking – a typical 4D data set ( $519 \times 72 \times 36 \times 101$  points) takes less than 10 hours to process (by way of comparison, the same dataset takes just over an hour to process in MATLAB), and thus 10-fold [ $(512 \times 72)/(36 \times 101)$ ] resolution enhancements of large datasets can readily be performed overnight using the Covariance NMR Toolbox.

## Methods

The Covariance NMR Toolbox is written using the MATLAB scripting language using standard MATLAB commands as well as commands from the Statistics Toolbox. All but one of the scripts, namely `Plot2DSpec.m`, of the Covariance NMR Toolbox (listed in Table 1) are also compatible with the OCTAVE computing environment. The toolbox performs Covariance NMR on frequency domain data using the SVD approach described previously [12,xxii]. Data is stored internally for use by these scripts either in the form of MATLAB/OCTAVE arrays (for small,  $\leq$  approximately 32 MB, datasets, e.g. 2D spectra with  $\leq 2048$

increments in each dimension) or in temporary files for larger datasets. The covariance NMR Toolbox is processor independent and runs on any platform for which OCTAVE or MATLAB is available. The example calculations reported here were performed on a Dell PowerEdge 2970 with a 2.0 GHz quad-core AMD Opteron processor and 8 GB of memory using MATLAB Version 7.9.0.529 (R2009b).

Example spectra, of a solution 0.1 M solution (in D<sub>2</sub>O) of AF dipeptide (Sigma) and recorded on a 400 MHz Bruker spectrometer, consisted of an <sup>1</sup>H-<sup>13</sup>C HMBC spectrum at natural abundance <sup>13</sup>C [19], and a <sup>1</sup>H-<sup>1</sup>H TOCSY spectrum recorded with the DIPSI-2 mixing sequence of 60 ms length [20]. The direct dimension of each spectrum had a sweep-width of 10.208 ppm with 1024 complex points recorded. The indirect dimension of the HMBC had 256 real points and a sweep-width of 222.095 ppm while the indirect dimension of the TOCSY had 256 complex points and a sweep-width of 10.202 ppm. Spectra were initially processed in NMRPipe, using standard-processing procedures for magnitude only (HMBC) and phase sensitive (TOCSY) data sets, prior to covariance analysis. Resonances for the AF dipeptide were assigned manually guided by the resonance assignments for free alanine and phenylalanine found in the BioMagResBank (BMRB) [23].

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

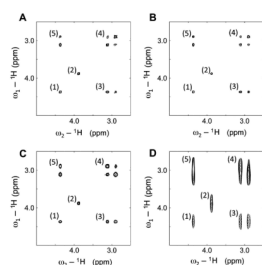
## Acknowledgments

The authors thank John Cain, Yanbin Chen, Megha Trambadiya, Pankaj Vekariya, Crystal Walcott and Fengli Zhang for stimulating discussions and advice concerning MATLAB scripting. The authors also thank Robin Rodriguez, Tom DePietro, Len Bogdon and the rest of the William Paterson University Information Services staff for assistance in setting up and for providing space for the William Paterson University Science computer cluster. This work was supported in part by the NIH (grant GM066041 to R.B.) as well as by the College of Science and Health (start up funds to D.A.S. and student worker funds to Timothy Short and Leigh Alzapiedi) and the Office of the Provost (Assigned Release Time for D.A.S.) of William Paterson University.

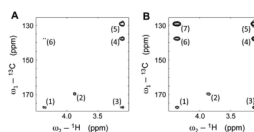
## References

1. Cavanagh, J.; Fairbrother, WJ.; Palmer, AG., III; Skelton, NJ. Protein NMR Spectroscopy: Principles and Practice. Academic Press; San Diego: 1996.
2. Snyder DA, Zhang F, Brüschweiler R. Covariance NMR in higher dimensions: application to 4D NOESY spectroscopy of proteins. *J Biomol NMR* 2007;39:165–175. [PubMed: 17876709]
3. Koehl P. Linear prediction spectral analysis of NMR data. *Prog Nucl Mag Res Sp* 1999;34:257–299.
4. Mobli M, Stern AS, Hoch JC. Spectral reconstruction methods in fast NMR: reduced dimensionality, random sampling and maximum entropy. *J Magn Reson* 2006;182:96–105. [PubMed: 16815055]
5. Xu Y, Long D, Yang D. Rapid Data Collection for Protein Structure Determination by NMR Spectroscopy. *J Am Chem Soc* 2007;129:7722–7723. [PubMed: 17536800]
6. Felli IC, Brutscher B. Recent advances in solution NMR: fast methods and heteronuclear direct detection. *Chem Phys Chem* 2009;10:1356–1368. [PubMed: 19462391]
7. Snyder, DA.; Bruschweiler, R. Encyclopedia of Magnetic Resonance. Sep. 2009 Multidimensional covariance spectroscopy by Covariance NMR.
8. Brüschweiler R, Zhang F. Covariance nuclear magnetic resonance spectroscopy. *J Chem Phys* 2004;120:5253–5260. [PubMed: 15267396]
9. Zhang F, Brüschweiler R. Indirect Covariance NMR Spectroscopy. *J Am Chem Soc* 2004;126:13180–13181. [PubMed: 15479045]
10. Blinov KA, Larin NI, Williams AJ, Mills KA, Martin GE. Using Unsymmetrical Indirect Covariance Processing to Calculate GHSQC-COSY spectra. *J Het Chem* 2006;43:163–166.

11. Blinov KA, Williams AJ, Hilton BD, Irish PA, Martin GE. The use of unsymmetrical indirect covariance NMR methods to obtain the equivalent of HSQC-NOESY data. *Magn Reson Chem* 2007;45:544–546. [PubMed: 17437315]
12. Snyder DA, Brüschweiler R. Generalized Indirect Covariance NMR Formalism for Establishment of Multidimensional Spin Correlations. *J Phys Chem A* 2009;113:12898–12903. [PubMed: 19810742]
13. Kupce E, Freeman R. Hyperdimensional NMR Spectroscopy. *Prog Nucl Mag Res Sp* 2008;52:22–30.
14. Lescop E, Brutscher B. Hyperdimensional Protein NMR Spectroscopy in Peptide-Sequence Space. *J Am Chem Soc* 2007;129:11916–11917. [PubMed: 17845051]
15. Benison G, Berkholz DS, Barbar E. Protein assignments without peak lists using higher-order spectra. *J Magn Reson* 2007;189:173–181. [PubMed: 17919953]
16. Snyder DA, Ghosh A, Zhang F, Szyperski T, Brüschweiler R. Z-Matrix formalism for quantitative noise assessment of covariance nuclear magnetic resonance spectra. *J Chem Phys* 2008;129:104511. [PubMed: 19044928]
17. Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–293. [PubMed: 8520220]
18. Chen Y, Zhang F, Snyder D, Gan Z, Bruschweiler-Li L, Brüschweiler R. Quantitative covariance NMR by regularization. *J Biomol NMR* 2007;38:73–77. [PubMed: 17351818]
19. Hadden CE, Martin GE, Krishnamurthy VV. Constant time inverse-detection gradient accordion rescaled heteronuclear multiple bond correlation spectroscopy: CIGAR-HMBC. *Magn Reson Chem* 2000;38:143–147.
20. Rucker SP, Shaka AJ. Broadband homonuclear cross polarization in 2D N.M.R. using DIPSI-2. *Mol Physics* 1989;68:509–517.
21. Snyder DA, Xu Y, Yang D, Brüschweiler R. Resolution-enhanced 4D  $^{15}\text{N}/^{13}\text{C}$  NOESY protein NMR spectroscopy by application of the covariance transform. *J Am Chem Soc* 2007;129:14126–14127. [PubMed: 17973386]
22. Trbovic N, Smirnov S, Zhang F, Brüschweiler R. Covariance NMR spectroscopy by singular value decomposition. *J Magn Reson* 2004;171:277–283. [PubMed: 15546754]
23. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL. BioMagResBank Nucleic Acids Res 2007;36:D402–D408.

**Fig. 1.**

Fourier transform and covariance TOCSY NMR spectra of the alanine-phenylalanine dipeptide (AF dipeptide). (A,C) with 512 points along the indirect dimension and (B,D) with 64 points along the indirect dimension. The covariance spectra (A,B) are similar (and both resolve the two phenylalanine H $\beta$  resonances) while the Fourier transform spectrum with only 64 points along the indirect dimension suffers from a loss of resolution and thus an inability to resolve the two H $\beta$  resonances. The displayed region of the spectrum includes the H $\alpha$ -H $\alpha$  diagonal peaks for both the (1) phenylalanine and (2) alanine residues, (3,5) the H $\alpha$ -H $\beta$  cross peaks for the phenylalanine and (4) the H $\beta$ -H $\beta$  cross and diagonal peaks.



**Fig. 2.**

(A) HMBC and (B) GIC [HMBC\*TOCSY]<sup>1/2</sup> spectra of the AF-dipeptide. The region displayed contains peaks correlating (1) phenylalanine C $\alpha$  and H $\alpha$ , (2) alanine C $\alpha$  and H $\alpha$ , phenylalanine (3) C $\alpha$  (4) C $\gamma$  and (5) C $\epsilon$  1 with one of the intra-residue H $\beta$  shifts and phenylalanine C $\gamma$  with H $\alpha$ . GIC (B) reveals (7) a phenylalanine C $\epsilon$  1-H $\alpha$  correlation not present in the HMBC spectrum.

```

A
tic
[fname,open_tocsy_fname]=read_array('tocsy.R2');
% TOCSY spectrum based on the data R2
[fname,open_tocsy_fname]=read_array('tocsy.R2.1');
% as in previous script, read in files for indirect covariance
% direct covariance with matrix square root requires no other validation parts
write_array([fname,open_tocsy_fname,'tocsy.R2.1'],tocsy);
% write output to tocsy.cov
tic

Plot2D(fnames{1},5,5,12,15,5,170,5,1,3,30);
% plot part of spectrum shown in Fig. 1A
print -dfig1 figure1A.pdf
Plot2D(fnames{2},5,5,12,15,5,170,5,1,3,30);
% plot part of spectrum shown in Fig. 1C
print -dfig1 figure1C.pdf

B
tic
[fname,open_tocsy_fname]=read_array('tocsy.R2.1');
% as in previous script, read in files for indirect covariance
[fname,open_tocsy_fname]=...
    covariance([fname,open_tocsy_fname,'tocsy.R2.1'],5);
% as performed in TOCSY+HMQC(1,2)
write_array([fname,open_tocsy_fname,'tocsy.R2.1'],tocsy);
tic

Plot2D(fnames{1},5,5,12,15,5,170,5,1,3,30);
% plot part of spectrum shown in Fig. 2A
print -dfig2 figure2A.pdf
Plot2D(fnames{2},5,5,12,15,5,170,5,1,3,30);
% plot part of spectrum shown in Fig. 2B
print -dfig2 figure2B.pdf

```

**Fig. 3.**

Scripts used to generate the plotted spectra of Figs. 1 and 2. Script (A) was used for Figs. 1A and 1C, script (B) for Figs. 2A and 2B. The script used to produce Figs. 1A and 1C was substantially the same as the script shown in panel (A) but it read a TOCSY spectrum with  $N_1 = 64$  rather than  $N_1 = 512$  with the run-time being reduced by an approximate factor of sixteen to about 0.3 seconds. The tic and toc statements, used to obtain the run-times indicated in the main text, bracket the primary covariance routines and thus are not effected by the time taken to plot images or write out those plots to pdf files. The axes labels and numbers shown in Figs. 1 and 2 were further modified for improved readability when the panels were compiled into two figures.



**Table 1**

<i>Script Name</i>	<b>Brief Description</b>
<i>read_nmrp</i>	Read in an NMR Pipe format file
<i>covar</i>	Covariance processing, e.g., of a spectrum uploaded using <i>read_nmrp</i>
<i>write_covar2pipe</i>	Write results of <i>covar</i> to a NMR Pipe format file
<i>Plot2DSpec</i>	Plot 2D spectrum (e.g. uploaded using <i>read_nmrp</i> or resulting from <i>covar</i> ) <sup>1</sup>
<i>apply_mask</i>	Masking and regularization of NMR datasets, e.g. uploaded using <i>read_nmrp</i>
<i>gic_pca</i>	Analyze a peak list using PCA and GIC
<i>swap_donor_accpt</i>	Swap dimensions which are identified as donor and which are identified as acceptor to toggle between indirect and direct covariance modes
<i>reset_donors</i>	Explicitly indicate which dimensions <i>covar</i> should consider to be “donor” dimensions subject to covariance
<i>get_header_data</i>	Parse NMR Pipe format header data (helper function)
<i>reset_header_data</i>	Reset NMR Pipe format header data (helper function)
<i>read_tmp_file</i>	Read temporary file written by <i>covar</i> when analyzing/writing out large data sets (helper function)
<i>inc2ppm</i>	Convert increments to ppm (helper function)
<i>ppm2inc</i>	Convert ppm to increments (helper function)

<sup>1</sup>This function (currently) assumes the input spectrum is a MATLAB array and not stored in a temporary file; this function also is not yet fully OCTAVE compatible