Journal of Biomolecular NMR, 3 (1993) 569–576
ESCOM

J-Bio NMR 134

Application of nonlinear sampling schemes to COSY-type spectra

Peter Schmieder[a], Alan S. Stern[b], Gerhard Wagner[a] and Jeffrey C. Hoch[b,*]

[a]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, U.S.A.*

[b]*Rowland Institute for Science, 100 Edwin H. Land Boulevard, Cambridge, MA 02142, U.S.A.*

Received 10 March 1993
Accepted 14 May 1993

Keywords: Nonlinear sampling; COSY-type spectra; Maximum entropy reconstruction; Resolution enhancement; Sensitivity enhancement

SUMMARY

Nonlinear sampling along the t_1 dimension is applied to COSY-type spectra. The sine dependence of the time domain signals for the cross peaks is matched by a nonlinear sampling scheme that samples most densely around the maximum of the sine function. Data are processed by maximum entropy reconstruction, using a modified implementation of the 'Cambridge' algorithm of Skilling and Bryan. The procedure is demonstrated for P.E.COSY spectra recorded on a cyclic hexapeptide and on a 126-residue domain of the protein villin. The number of t_1 values in the nonlinearly sampled experiments was reduced by a factor of four compared to linear sampling. The sensitivity and resolution of the resulting spectra are comparable to those achieved by conventional methods. The method described can thus significantly reduce the measuring time for COSY-type spectra.

INTRODUCTION

The twin limitations in protein NMR are sensitivity and resolution. Improvements in hardware and higher magnetic fields contribute to higher sensitivity, while multidimensional experiments and the use of isotopically labeled molecules improve resolution. Sensitivity and resolution are intimately related, so it is often possible to trade one for the other by adjusting the number of sampled time intervals and the amount of signal averaging for each sampled interval, although the ability to do so may be limited by the requirement for phase cycling. Continued improvements in both sensitivity and resolution will enable the study of even larger molecules.

Spectral resolution is essentially determined by the length of the sampling period. While the

*To whom correspondence should be addressed.

Abbreviations: COSY, two-dimensional correlation spectroscopy; E.COSY, exclusive COSY; P.E.COSY, primitive E.COSY; TPPI, time proportional phase incrementation; FID, free induction decay.

discrete Fourier transform (DFT) requires samples at uniformly spaced intervals, some alternative methods do not, allowing comparable resolution to be achieved by collecting fewer samples spanning the same sampling period. One such method is maximum entropy reconstruction (MaxEnt) (Sibisi, 1983; Laue et al., 1985, 1986; Hoch, 1989; Jones and Hore, 1991). MaxEnt has other desirable properties, such as better ability to handle truncated data records without introducing artefacts. MaxEnt has not yet enjoyed widespread use, however, mainly due to its increased computational cost compared to the DFT. Recent dramatic developments in computer performance can be expected to enable wider use of MaxEnt for processing two- and higher-dimensional data.

NONLINEAR SAMPLING

Nonlinear sampling (Barna et al., 1987) offers two advantages. One is that fewer samples are required to achieve a given resolution. The other is that sampling can be tailored to match the intensity of the signal, improving the sensitivity for a given measuring time. These advantages are not realized in the directly detected dimension, since the number of data samples in that dimension hardly affects the overall measuring time. In the indirectly detected dimensions of two- and higher-dimensional spectra, however, alternative sampling schemes can be used to save time or to allow for lower sample concentrations, more extensive phase cycling, or higher resolution in a given period of time.

Previous applications of nonlinear sampling dealt with detection of in-phase (cosine modulated) magnetization, using a simple exponential sampling schedule (Robin et al., 1991). In the present study we generalize the sampling schedule for optimal detection of antiphase (sine modulated) magnetization, such as arises in COSY-type experiments, and to explicitly take into account the relaxation properties of the signal. We illustrate the significant time savings that can be achieved using this method with P.E.COSY experiments for a peptide and a protein.

SELECTION OF SAMPLE POINTS

The procedure used to determine the sampling schedule is based on the one given by Barna et al. (1987). First, a continuous density function is defined. In the case of in-phase magnetization the cosine term is omitted for simplicity and only a term representing the exponential decay is retained:

$$D(t) = A\exp(-kt) \tag{1}$$

with

$$\int_0^{N\Delta t} D(t)\, dt = n \tag{2}$$

(Here n is the number of exponentially sampled points to be distributed through a time interval consisting of N uniform sample periods, each of length Δt.) This density function differs by the factor A from that given by Barna et al., allowing adjustment of the parameter k to the actual T_2 of the molecule. Solving the integral leads to

$$A = \frac{nk}{1-\exp(-kN\Delta t)} \tag{3}$$

Antiphase magnetization requires a density function in which the sine term occurs explicitly:

$$D(t) = A\exp(-kt)\sin(t\pi/N\Delta t) \tag{4}$$

In this treatment it is assumed that half a period of the sine function is sampled; however, the function can easily be modified to allow other possibilities, e.g., if the relaxation properties restrict the sampling time. Substituting this density function in Eq. 2 leads to

$$A = \frac{n(N\Delta t/\pi)(k^2 + (\pi/N\Delta t)^2)}{1 + \exp(-kN\Delta t)} \tag{5}$$

We then calculate the integrals

$$I(T) = \int_0^T D(t)\,dt, \text{ for } T = \Delta t,\ldots, N\Delta t, \tag{6}$$

and for each integer $j = 1,\ldots,n$, the delay time t_j for the j-th data point is taken to be the least integer multiple of Δt greater than t_{j-1} such that $I(t_j) \geq j$. The resulting sampling schedule has n points, is adjusted to the relaxation and modulation properties of the magnetization, and extends as long as a linearly sampled N-point data set. (Note that this procedure is inappropriate for generating sampling schedules for constant-time experiments. Since there is no decay in these experiments, the resulting schedules would consist simply of uniformly spaced delays, effectively equivalent to a linear sampling scheme with a reduced spectral width.)

Because MaxEnt is inherently nonlinear, practical guidelines for the use of nonlinear sampling schedules are mainly empirical. The resolution is determined primarily by the longest sampling interval, the choice of the sampling distribution is based on the relaxation times and J values, and the sampling schedule together with the signal-to-noise ratio helps determine the spectral quality. In practice we find that within these constraints, at least one-quarter of the number of linearly spaced intervals between the beginning of the FID and the longest sampling interval should be collected.

SPECTRUM RECONSTRUCTION

Maximum entropy reconstructions were performed using a modified version of the 'Cambridge' algorithm (Skilling and Bryan, 1984). The algorithm was simplified by eliminating the 'P-chop'. The complex entropy functional was given by the S_2 form described by Hoch et al. (1990) (or equivalently, the $S_{1/2}$ form described by Jones and Hore (1991)):

$$S = -\sum_{\omega}\left[|f_\omega|\log\frac{|f_\omega| + \sqrt{4def^2 + |f_\omega|^2}}{2def} - \sqrt{4def^2 + |f_\omega|^2}\right] \tag{7}$$

where f_ω is the complex value of the spectrum at frequency ω and def is a user-specified parameter of the reconstruction. The reconstructed spectrum is the one having the maximum value of S subject to the constraint

$$\sum_{t \in SS} |d_t - D_t|^2 \leq E \tag{8}$$

where E is the expected experimental error, SS is the set of points in the sampling schedule, D_t is the value of the experimental data at time t, and d_t is the value of mock data given by

$$d_t = \frac{1}{\sqrt{N}} \sum_{\omega} f_\omega \exp(2\pi i t\omega/N) \exp(-\pi L t) \tag{9}$$

where L is an estimated linewidth. Finally, the search directions are different from the ones given by Skilling and Bryan. Following the notation in (Skilling and Bryan, 1984), the algorithm proceeds initially by using the search directions

$$\nabla C,\ f(\nabla C),\ |\nabla S|^{-1} f(\nabla S) - |\nabla C|^{-1} f(\nabla C), \text{ and } |\nabla S|^{-1} f(\nabla\nabla C) f(\nabla S) - |\nabla C|^{-1} f(\nabla\nabla C) f(\nabla C) \tag{10}$$

until the desired value of E is reached. It then switches to using

$$|\nabla S|^{-1} \nabla S - |\nabla C|^{-1} \nabla C \tag{11}$$

as the second search direction, to maximize the entropy S while maintaining the constraint C. We find that these search directions improve the convergence properties of the algorithm for both nonlinearly and linearly sampled data.

MATERIALS AND METHODS

A P.E.COSY (Müller, 1987) experiment was recorded with a sample of cyclo-(-D-Ala-Phe-Trp-Lys-Val-Phe-), 30 mM in d_6-DMSO at 300 K. A P.E.COSY experiment was also recorded with a sample of a 14-kDa domain of villin 14T (Bazari et al., 1988), 3.5 mM in D_2O at 298 K. All spectra were recorded nonspinning on a Bruker AMX 600 spectrometer.

The P.E.COSY of the cyclic hexapeptide was recorded with 16 scans, 2048 FIDs with 2048 complex points each recorded, using the TPPI-States procedure (Marion et al., 1989) to achieve quadrature detection in F1. The spectral window was 8333.3 Hz in both dimensions, and the mixing pulse was 35°. A reference spectrum was recorded following the procedure outlined by Marion and Bax (1988) and subtracted prior to processing to reduce the dispersive diagonal.

The P.E.COSY of villin 14T was recorded with nonlinear sampling. Following the procedure outlined above, 256 t_1 values were selected, assuming a T_2 of 20 ms and using N equal to 1024. The FIDs were recorded with 104 scans and 1024 complex points each. The spectral window was 8333.3 Hz in both dimensions, the mixing pulse was 30°, and the total measuring time was 20 h.

The first P.E.COSY was processed in the F2 dimension with a 60°-shifted squared sinebell. Next, three different data sets were constructed starting from this one. For the first two sets, the data were processed in F1, using a 60°-shifted squared sinebell, zero-filled to 2048 points and Fourier transformed, resulting in a data matrix of 2048 by 2048 real points. In one case all 2048 FIDs were used, in the other only 512, corresponding to 1024 and 256 complex points, respectively. The third data set contained only 256 complex points in t_1 assembled by extracting the FIDs belonging to a sampling schedule assuming a T_2 of 50 ms. MaxEnt reconstruction then led to a

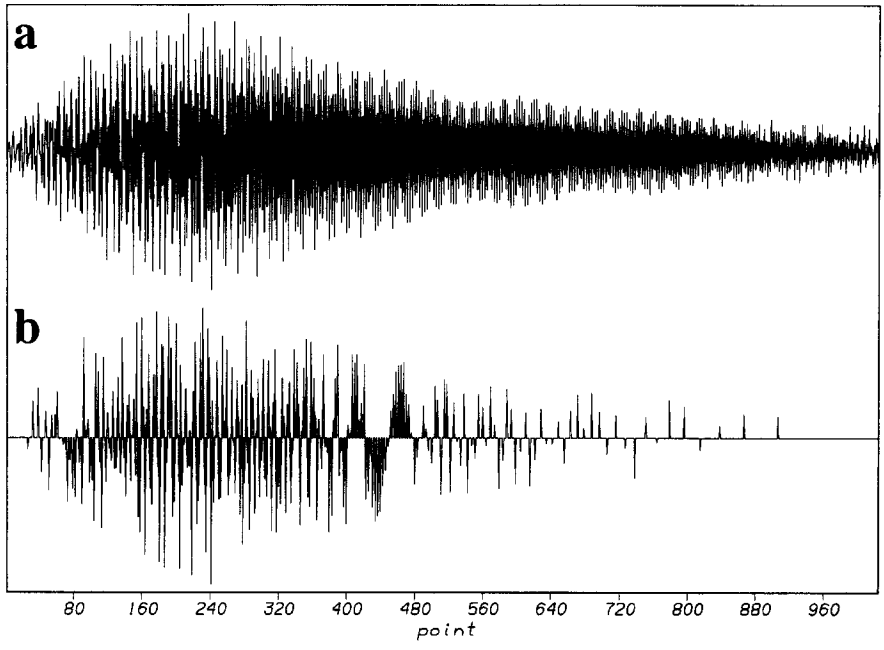

Fig. 1. Linearly sampled COSY data (a) and nonlinearly sampled data (b). The data in (b) are the same as in (a), except that points not belonging to the sampling schedule are set to zero.

data matrix of 2048 by 2048 real points. The 1-Hz linewidth parameter used in the reconstruction was chosen to give a sharpening comparable to that caused by the application of a shifted squared sinebell.

The P.E.COSY of villin 14T was zero-filled and apodized with a 60°-shifted squared sinebell in F2 prior to Fourier transformation. MaxEnt reconstruction was applied in F1 with a linewidth parameter of 1 Hz, again leading to a 2048 by 2048 real matrix.

Conventional Fourier processing was done using the Felix program (Hare Research) on a Sun Sparc SLC. The MaxEnt reconstructions were done on a Silicon Graphics 4D/480 computer running on eight processors with 64 Mb of memory, using the algorithm described above. The reconstruction took less than 30 min. MaxEnt reconstruction was also performed on a Digital Equipment Corporation DECmpp 12000 computer equipped with 8192 processors and 512 Mb of memory. Here, the reconstruction took 15 min. The apparent discrepancy in the speed-up using the massively parallel computer is due to its less powerful processors and communication overhead, among other factors.

RESULTS AND DISCUSSION

Figure 1 illustrates the sampling schedule used for the P.E.COSY experiment. Figure 1a shows a conventional, linearly sampled FID of a methyl group of the hexapeptide. In Fig. 1b, points not included in the sampling schedule are set to zero. Close inspection shows that the sampling density is highest close to the maximum of the sine-modulated intensity envelope.

Figure 2 compares the effects of conventional processing and nonlinear sampling with MaxEnt

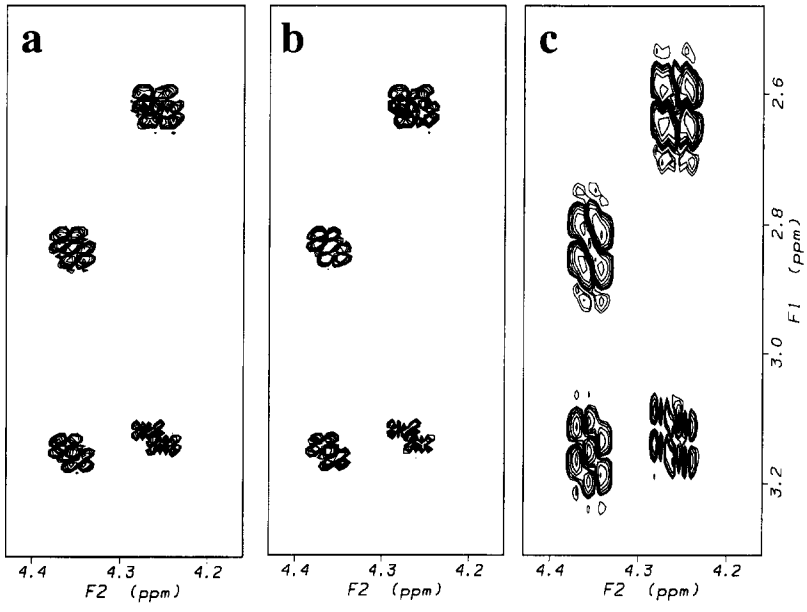

Fig. 2. Phenylalanine H^{α}–H^{β} cross peaks from P.E.COSY spectra of cyclo-(-D-Ala-Phe-Trp-Lys-Val-Phe-). Processing in F1 was performed by (a) DFT, using 2048 linear t_1 samples, (b) MaxEnt reconstruction using 512 nonlinear t_1 samples, and (c) DFT, using 512 linear t_1 samples.

processing on the P.E.COSY spectrum of cyclo-(-D-Ala-Phe-Trp-Lys-Val-Phe-). The DFT of the complete data set (1024 complex points, Fig. 2a) and the MaxEnt reconstruction of the nonlinearly sampled data set (256 complex points, Fig. 2b) are nearly indistinguishable, whereas the DFT using only 256 complex points (Fig. 2c) shows a marked reduction in resolution. The figure clearly demonstrates that nonlinear sampling schemes can yield results comparable to conventional methods in as little as one-quarter the experiment time.

In a final application, P.E.COSY data for the protein villin 14T were acquired using nonlinear sampling; the spectrum is shown in Fig. 3. COSY-type spectra of proteins are particularly challenging due to the cancellation of cross-peak intensity as the linewidth approaches the magnitude of the coupling constant. In this experiment, the time savings afforded by nonlinear sampling permitted additional signal averaging without increasing the overall experiment time. The resolution of the resulting spectrum is comparable to what could be achieved in the same time by conventional methods, but there is a substantial increase in sensitivity.

In P.E.COSY experiments used to measure small coupling constants, resolution is especially important to achieve sufficient displacement of the E.COSY cross peaks. Spectral folding is often used to increase digital resolution; this technique can be used with nonlinear sampling and MaxEnt reconstruction, although it may no longer be necessary. In heteronuclear 3D spectra, nonlinear sampling and MaxEnt reconstruction can achieve a similar resolution without sacrificing spectral simplicity as one does with folding.

It should be emphasized that nonlinear sampling inherently involves a trade-off: sampling time and spectral resolution are improved at the cost of sensitivity and spectral quality. Frequently this trade-off is quite acceptable, since many experiments have adequate signal-to-noise ratio and the

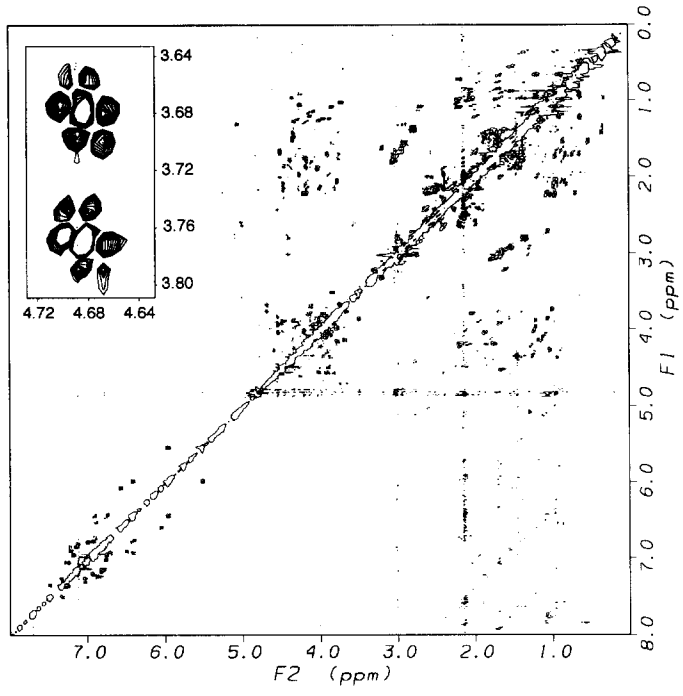

Fig. 3. P.E.COSY spectrum of villin 14T, obtained by nonlinear sampling (256 t_1 values), and MaxEnt reconstruction in F1. The inset shows the H^{α}–H^{β} cross peaks of Ser^{58}.

primary goal is to shorten the experiment time or increase resolution. However, the use of a nonlinear sampling schedule introduces aliasing artefacts, which can be particularly troublesome in spectra exhibiting high dynamic range. These artefacts can be ameliorated by appropriate choice of the sampling schedule. Alternatively, there are instances where the dynamic range can be reduced prior to MaxEnt reconstruction without loss of information. For example, homonuclear spectra can be preprocessed to suppress the strong diagonal signals. A detailed description of these considerations will be presented elsewhere; for the present we note that the benefits greatly outweigh the problems. (No effort was made to correct for these effects in the spectrum shown in Fig. 3; they manifest themselves as shifted images of the diagonal. The t_1 noise apparent in the spectrum is genuine.)

The use of nonlinear sampling in conjunction with MaxEnt is not a general technique to be applied routinely. However, there are circumstances that occur reasonably often and warrant the additional computational effort, for example, when sensitivity or resolution are severely limited by sample concentration or lifetime of the molecule under investigation. In addition, when a great many survey spectra are being collected, for example experiments to determine pH or temperature dependence, the time savings afforded by nonlinear sampling can be substantial. Relaxation measurements appear to be well suited to nonlinear sampling; however, the ability to quantitate reliably features in spectra reconstructed from nonlinearly sampled data has not yet been demonstrated.

The results shown here clearly illustrate the advantages of nonlinear sampling schemes applied

to the collection of COSY-type NMR data. These schemes are applicable to the indirectly detected dimensions of two- and higher-dimensional experiments. The technique can be used to improve resolution or sensitivity, or to reduce overall experiment time. Implementation of nonlinear sampling is not difficult on modern spectrometers, and with the advent of more powerful computers, the computational cost is rapidly becoming less burdensome. As a result, nonlinear sampling and MaxEnt reconstruction can be added to the repertoire of techniques used to enhance the sensitivity and resolution of protein NMR experiments.

ACKNOWLEDGEMENTS

We thank Dr. M. Kurz and Prof. Dr. H. Kessler for the sample of the cyclic hexapeptide and Michelle Markus, Tomoko Nakayama, and Paul Matsudaira for the sample of villin 14T. P.S. thanks the Deutsche Forschungsgemeinschaft for a fellowship. We are grateful to the Digital Equipment Corporation for the loan of a DECmpp 12000 computer. This work was supported by NIH grant GM47467 and by The Rowland Institute for Science.

REFERENCES

Barna, J.C.J., Laue, E.D., Mayger, M.R., Skilling, J. and Worrall, S.J.P. (1987) *J. Magn. Reson.*, **73**, 69–77.

Bazari, W.L., Masudaira, P., Wallek, M., Smeal, T., Jakes, R. and Ahmed, Y. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4986–4990.

Hoch, J.C. (1989) *Methods Enzymol.*, **176**, 216–241.

Hoch, J.C., Stern, A.S., Donoho, D.L. and Johnstone, I.M. (1990) *J. Magn. Reson.*, **86**, 236–246.

Jones, J.A. and Hore, P.J. (1991) *J. Magn. Reson.*, **92**, 276–292.

Laue, E.D., Skilling, J., Staunton, J., Sibisi, S. and Bereton, R.G. (1985) *J. Magn. Reson.*, **62**, 437–452.

Laue, E.D., Mayger, M.R., Skilling, J. and Staunton, J. (1986) *J. Magn. Reson.*, **68**, 14–29.

Marion, D. and Bax, A. (1988) *J. Magn. Reson.*, **80**, 528–532.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Comm.*, **113**, 967–974.

Müller, L. (1987) *J. Magn. Reson.*, **72**, 191–197.

Robin, M., Delsuc, M.-A., Guittet, E. and Lallemand, J.-Y. (1991) *J. Magn. Reson.*, **92**, 645–650.

Sibisi, S. (1983) *Nature*, **301**, 134–136.

Skilling, J. and Bryan, R.K. (1984) *Mon. Not. R. Astron. Soc.*, **211**, 111–124.