ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Magnetic Resonance 162 (2003) 74–89

JMR Journal of Magnetic Resonance

www.elsevier.com/locate/jmr

Progress on the two-dimensional filter diagonalization method. An efficient doubling scheme for two-dimensional constant-time NMR

Jianhan Chen, Anna A. De Angelis, Vladimir A. Mandelshtam,[1] and A.J. Shaka*

Chemistry Department, University of California at Irvine, Irvine, CA 92697-2025, USA

Received 10 June 2002; revised 17 December 2002

Abstract

An efficient way to treat two-dimensional (2D) constant-time (CT) NMR data using the filter diagonalization method (FDM) is presented. In this scheme a pair of N- and P-type data sets from a 2D CT NMR experiment are processed jointly by FDM as a single data set, twice as large, in which the signal effectively evolves in time for twice as long. This scheme is related to "mirror-image" linear prediction, but with the distinction that the data are directly used, without any preprocessing such as Fourier transformation along one dimension, or point-by-point reflection. As the signal has nearly perfect Lorentzian line shape in the CT dimension, it can be efficiently handled by the FDM approach. Applied to model and experimental signals, the scheme shows significant resolution improvement, and appears to tolerate noise reasonably well. Other complex aspects of multidimensional FDM are discussed and illustrated.

Keywords: Filter diagonalization method; FDM; FDM2K; Multidimensional NMR; Constant-time; Regularization; NMR data processing; Protein NMR; Ubiquitin

1. Introduction

Many protein backbone NMR experiments on uniformly labeled molecules employ a fixed, or constant time (CT) evolution [1] period, $2T$, to encode chemical shift information, as demonstrated by Vuister and Bax [2] for the CT ^{13}C–^{1}H two-dimensional heteronuclear single-quantum correlation (2D HSQC) experiment [3]. In some cases the chemical shift encoding can be carried out in a fixed delay that is already present for coherence transfer to another spin, for example in the CT-HNCO [4] experiment, thereby minimizing the total time between excitation and acquisition. In other cases the constant time period serves to decouple the homonuclear carbon–carbon couplings, leading to a sharp singlet instead of a broader multiplet [1,2]. The CT signals typically show almost no decay in the CT dimension with typical times $2T$, as the residual line width is determined by the quality of the magnetic field homogeneity, which, with modern shimming, is excellent. However, the Fourier time–frequency uncertainty principle places a strict bound on the achievable resolution: using Fourier transform (FT) analysis, the line width in the CT dimension cannot be narrower than that of a sine cardinal ("sinc") function line shape corresponding to acquisition over time T. Apodization of the CT signal, to reduce the sinc-function oscillations around the base of the peaks, inevitably increases this width further.

Because data acquisition is limited to the interval $[0, T]$, these CT experiments naturally attract the attention of alternative strategies aimed at overcoming the artificial line width imposed by Fourier transformation of data sharply truncated at time T. Previous methods have included random sampling in the $[0, T]$ interval in conjunction with maximum entropy reconstruction (MaxEnt) [5] and linear prediction (LP) extrapolation followed by Fourier transformation using a milder

* Corresponding author. Fax: 1-949-824-3168.
E-mail addresses: mandelsh@uci.edu (V.A. Mandelshtam), ajshaka@uci.edu (A.J. Shaka).
[1] Also corresponding author.

1090-7807/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1090-7807(03)00045-4

apodization function [6]. In practice, the LP extrapolation along t_1 is carried out in a trace-by-trace fashion after Fourier transformation along the running time, t_2, phasing, and isolation of the absorption-mode data in F_2. The expected phase behavior of the interferogram can be used to stabilize the LP coefficients using an artificial reflection of the data around $t_1 = 0$ [7], a method now designated "mirror image LP." Mirror image LP is quite routinely used in data work up [8], whereas nonuniform time grid sampling is incompatible with conventional FT analysis, limiting the popularity of the MaxEnt approach. In practice, one-dimensional time-domain extension of each interferogram to $[0, 2T]$, or sometimes $[0, 4T]$, is carried out before Fourier transformation along t_1. Evidently, the achievable resolution is still limited by the *actual length* of the data set submitted for FT processing, becoming twice, or four times as fine, respectively, in F_1. This estimate is overly optimistic, though. In actual fact, the true resolution enhancement delivered by mirror image LP is variable, depending on the filter length, the noise level, and other variables, and often falls well short of even this rather modest theoretical limit.

We present an alternative, and more aggressive, way to treat CT NMR data using the multidimensional filter diagonalization method (FDM) [9–11]. In particular, we show that it is possible to combine either N- and P-type phase modulated data sets, or sin- and cos-type amplitude modulated data sets, directly in the time domain. The N/P signals arise naturally when coherence transfer pathway selection by pulsed field gradients is employed, and are phase-modulated data that differ only in the apparent sign of the evolution frequency in the indirect dimension, i.e., $\exp(\pm i2\pi\delta_1 t_1)$. Amplitude modulation results from conventional phase cycling, and yields, as is well known, sets of signals like $\sin(2\pi\delta_1 t_1)$ and $\cos(2\pi\delta_1 t_1)$. By combining either of these pairs of data sets, as described later, a virtual signal of twice the length in t_1 results. This doubling scheme is related to the mirror image LP method, but is carried out directly in the time-domain so that the noise in the doubled signal is uncorrelated, rather than reflected. The FDM treatment is an intrinsic 2D analysis rather than a trace-by-trace 1D analysis, which has consequences for the separation of partially overlapping peaks. High quality 2D spectra can be obtained with a single calculation using the *regularized* FDM algorithm, FDM2K, described previously [12]. Application to CT-HSQC spectra of human ubiquitin shows that much shorter constant time periods can be used than typically employed in these experiments, while still accurately identifying the chemical shifts of the vast majority of the peaks. Local crowding in the 2D plane makes it problematical to identify individual peaks correctly using FDM2K, but there is good reason to believe that none of the alternative methods in the literature will succeed in identifying the true number of peaks, and their chemical shifts, under these circumstances. The behavior of the spectrum as a function of the number of available data points is studied using model signals, to show how convergence of the features evolves. Noisy data turns out to be a tougher case to quantify in the theory, but at least for 1 mM concentrations at 500 MHz using a conventional room temperature probe, it appears that sensitivity is more than adequate for FDM2K to be used routinely in CT experiments on uniformly labeled proteins. A series of increasingly noisy protein spectra is used to illustrate the degradation in performance that noise causes.

The discussion here is also meant to consolidate, clarify, and qualitatively describe the recent progress that we have made with multidimensional FDM. Perfectly plausible ways of treating the nD signal simply fail to give reasonable results. The failures are, nevertheless, quite important for a proper understanding of the potential of FDM, and have taken considerable effort to assemble into an organized picture.

2. Theory

FDM is a basis set method that makes assumptions about the form of the spectrum: it is assumed to consist of a finite number of Lorentzian lines. When the assumptions are closely enough satisfied and the basis is sufficiently large, essentially exact results can be obtained for peak positions, widths, amplitudes, and phases. There is an immediate analogy to quantum chemistry calculations [13], in which one can expand the basis set size indefinitely, if one is willing to contemplate an increasingly expensive computation, thereby improving accuracy for the energy. Unlike quantum chemistry, in FDM the basis set functions are not known per se. Rather it is assumed that matrix elements of some operators have effectively been measured by recording the free induction decay (FID). As it is a matrix representation of a "Hamiltonian" that is diagonalized to calculate "energies," it is not necessary to know the underlying functions if one has a matrix representation in a basis. Thus, there is no problem calculating frequencies of lines, assuming them to arise from some effective evolution operator, if the appropriate matrix representation is available. In FDM these matrix elements originate solely from the measured time-domain data. If the number of measured points is too few, then the basis size may be insufficient to calculate accurate eigenvalues. If the noise level is high or the lines depart severely from the Lorentzian assumption, then it becomes ever more difficult to characterize the resulting spectrum in terms of a finite number of Lorentzian features. This has the effect of making the basis *appear* to be too small, and lowers accuracy. Note, however,

that these limitations are different in nature than the time–frequency uncertainty principle, which draws a strict veil over line positions, independent of the type of signal or noise level. We can thus expect FDM to outperform FT when (i) the number of spectral features is sufficiently small, so that they are described adequately with the information content of the available data; (ii) the features are nearly Lorentzian; and (iii) the noise level is not too high. These are qualitative statements that will be made more quantitative in what follows.

2.1. Basis set methods, eigenvalue problems, and 1D FDM

In many elegant treatments of model quantum mechanical problems, the eigenkets of the Hamiltonian can be uncovered purely by operator techniques, in which the commutation relations of the operators dictate the number and kind of states. Thus, in the 1D simple harmonic oscillator, use of the creation and annihilation operators, $a^\dagger$ and a, the number operator $N = a^\dagger a$, and the commutation relations:

$$[a, a^\dagger] = 1, \tag{1}$$

$$[N, a^\dagger] = a^\dagger, \tag{2}$$

$$[N, a] = -a \tag{3}$$

allow the complete ladder of energy eigenstates $|n\rangle$ to be built up systematically by applying $a^\dagger$ iteratively to the ground state, $|0\rangle$ [14]. Likewise, in conventional treatments of angular momentum, the raising operator

$$J_+ = J_x + \mathrm{i}J_y \tag{4}$$

is used to build up the $(2j+1)$ joint eigenstates of J^2 and J_z, $|jm\rangle$, for a particle of spin j, and to show these states are bounded [15].

A very similar method is used to build up the primitive basis vectors for 1D FDM. Starting with an "initial state" Φ_0 we iteratively apply an operator U that translates states forward by one dwell time τ, to generate a collection of primitive basis functions that can be imagined as "time like" in nature

$$\Phi_n = U^n \Phi_0. \tag{5}$$

We also assume that the complex-valued FID, sampled instantaneously and discretely on a time grid, and written as a list of complex numbers $c_0, c_1, \ldots, c_{N-1}$ can itself be written as a time autocorrelation function [16], the complex symmetric inner product of the initial state with a time-shifted state

$$(\Phi_0|U^n\Phi_0) = (\Phi_0|\Phi_n) \equiv c_n, \tag{6}$$

where $(\Phi|\Psi) = (\Psi|\Phi)$ defines a (non-Hermitian) complex symmetric inner product between states Φ and Ψ. It is a simple matter to show that the requirement of a square matrix representation, with no missing entries, limits the size of the basis to $M = N/2$ for data length N [17,18]. As in the simpler cases of the harmonic oscillator and rigid rotor, the iterative construction of the basis allows the problem to be solved abstractly, but unlike those cases we cannot guarantee that the basis functions are orthogonal to each other, normalized, or complete. In addition, the "length" $(\Psi|\Psi)$ of a basis vector Ψ can be complex, a complication that is necessary to allow peaks to acquire width and arbitrary phase (frequencies and amplitudes are complex rather than real numbers). The complex symmetric "Hamiltonian" Ω underlying $U = \exp(-\mathrm{i}\Omega\tau)$ is also unknown, so that its eigenstates must be uncovered by purely numerical means.

With a non-orthonormal basis, the usual matrix eigenvalue problem becomes a generalized eigenvalue problem. If the M eigenfunctions of U are Y_k and the eigenvalues are u_k then the symbolic eigenvalue problem

$$UY_k = u_k Y_k \tag{7}$$

becomes, in the primitive basis, the generalized eigenvalue problem

$$\mathbf{U}^{(1)}\mathbf{B}_k = u_k \mathbf{U}^{(0)}\mathbf{B}_k, \tag{8}$$

where the complex symmetric matrices $\mathbf{U}$ have matrix elements $\mathbf{U}_{nm} = \mathbf{U}_{mn}$ that are simply the recorded data points [18]:

$$\mathbf{U}^{(0)}_{nm} = (\Phi_n|\Phi_m) = (U^n\Phi_0|U^m\Phi_0) = (\Phi_0|U^{m+n}\Phi_0) = c_{n+m}, \tag{9}$$

$$\mathbf{U}^{(1)}_{nm} = (\Phi_n|U\Phi_m) = (\Phi_n|\Phi_{m+1}) = c_{n+m+1} \tag{10}$$

and the last equalities result from the application of Eq. (6). Note that if the set $\{\Phi_k\}$ were orthonormal then $\mathbf{U}^{(0)}$ would be the identity matrix, and could be omitted from Eq. (8) altogether. With the eigenvalues and eigenvectors from Eq. (7) in hand, Eq. (6) can be expressed in the form

$$c_n = (\Phi_0|U^n\Phi_0) = \sum_k (\Phi_0|Y_k)u_k^n(Y_k|\Phi_0) = \sum_k (\Phi_0|Y_k)^2 u_k^n, \tag{11}$$

which *looks* exactly like a nonlinear *fit* of the FID in terms of complex sinusoids

$$c_n = \sum_k d_k \exp(-\mathrm{i}n\tau\omega_k) = \sum_k d_k u_k^n \tag{12}$$

once we make the identification

$$d_k = (\Phi_0|Y_k)^2 \tag{13}$$

between the eigenvectors and the line intensity and phase. For a noiseless model signal that satisfies the Lorentzian assumption exactly, and with a basis size that exceeds the true number of peaks, it can be shown in fact that FDM gives an exact solution to the fitting problem of Eq. (12). However, in all other cases, even though an exact fit of the N complex data points can

usually be obtained using $M = N/2$ basis functions, there is no guarantee that even a perfect fit of a given finite time segment of data makes the form obtained suitable for spectral estimation. In particular, if the basis is too small then the spectral features calculated by FDM can be in substantial error, even if the signal conforms exactly to the assumed model. Therefore, it is usually essential to analyze a spectral region several times, using varying basis size, to ensure that convergence has been achieved. For a general nD signal there is no proof of a general fit unless the Lorentzian assumption is satisfied exactly and the basis is sufficiently large.

2.2. Fourier bases

Oftentimes in quantum mechanics the basis functions that are most easily constructed in the first instance are not the best ones to solve a particular problem. For example, in a system of N coupled spin-1/2 particles the 2^N tensor product states $|\alpha_1 \cdots \alpha_N\rangle, \ldots, |\beta_1 \cdots \beta_N\rangle$ are the most natural to set up. But the fact that states with different total magnetic quantum number are not mixed by the high-resolution spin Hamiltonian can be used to diagonalize a number of smaller submatrices. The weak coupling approximation, when appropriate, results in further simplification. This kind of legerdemain lets NMR spin simulation programs handle a much larger number of spins. In an analogous way, the primitive basis functions Φ_n are not the best choice to obtain frequency positions of peaks, and linear combinations of them can be taken to improve the structure of the problem and limit the size of the matrix problem by neglecting unimportant far off-diagonal matrix elements.

Basically, it is not profitable to try to obtain a fit of even a 1D FID using the time-like basis functions Φ_n because the size of the $\mathbf{U}$ matrices becomes utterly enormous for any realistic data set [18]. Instead, a frequency is imposed by taking the Fourier transform of the Φ_n, picking some frequency position f_j within the spectral width:

$$\Psi_j = \sum_n \exp\left(-2\pi i f_j n\tau\right)\Phi_n. \tag{14}$$

To cover the whole Nyquist range, these frequencies f_j are laid down in an even grid across the entire spectrum, at about half the density that would result from a DFT (without zero filling). Very much in the same way that the intensity of a particular frequency point in the FT spectrum is usually dominated mostly by nearby peaks, so the "frequency-like" basis functions Ψ_j are dominated mostly by nearby eigenfunctions Y_k. Turning this around, one may assume that a given eigenfunction can be accurately represented by a small linear combination of nearby basis functions Ψ_j in a suitably small spectral window centered on the region of interest. This in turn allows the potentially giant generalized eigenvalue problem to be broken down into a series of much smaller problems that can be handled efficiently. The window is moved through the spectrum, usually with 50% overlap, and the calculated spectral regions are added up with some weighting, which we have chosen, after some experimentation, as a $\cos^2\theta$ function with $-\pi/2 \leqslant \theta \leqslant \pi/2$ across the window. These features are shown schematically in Fig. 1. This filtering of the basis

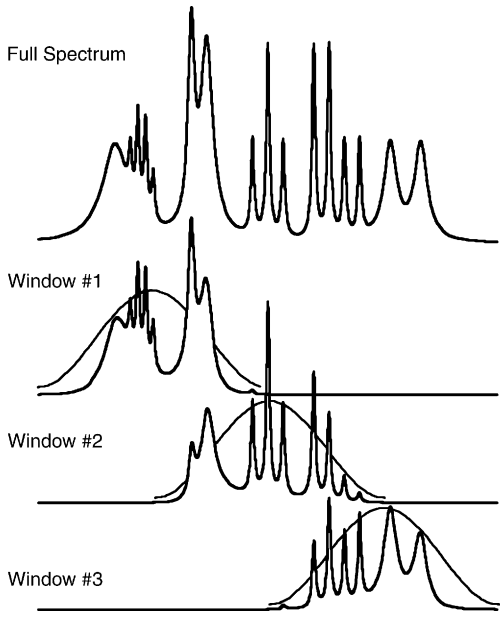

Fig. 1. Fitting a 1D spectral region with a series of three 1D windows. The narrowband window basis is set up over a grid of frequencies f_j that span the region of interest, windows #1, #2, and #3 in this example. Other basis functions are filtered out, or replaced with a much smaller number of diffuse functions [19]. An alternative way to take partial account of peaks outside the window is to increase, slightly, the basis density within the window, by about 10% or so. Diagonalization in the window #1 basis results in a list of eigenvalues and eigenvectors that are used to produce the FDM spectral estimate of the window #1 region. The individual entries may be a function of the exact window size, but the spectrum itself is fairly independent of the window details. Accuracy at the edges of the window is less, so the calculated spectrum is weighted by a $\cos^2\theta$ function, as shown, to smoothly attenuate these regions. The window is then shifted by half its width, to give window #2, and the whole procedure repeated. Summing the individual weighted windows produces the FDM spectral estimate. A fourth window, wrapping around the edges of the spectrum, would be needed to complete the estimate. Note that almost every line is calculated at least twice, so that the line list is redundant. It is tedious, however, to try to match and delete entries that are numerically close, but distinct, frequencies which may, or may not, refer to the same line. Broad or intense lines outside the window, whose spectral contribution is accurately reflected inside, would require some sort of special treatment. Their actual positions and widths are not particularly accurate except to produce tails within the current window. Occasionally they do not actually exist at all, being fictional lines produced by the limitation of the full matrix problem to the window basis. Very narrow features of small integral may refer to noise, so some very slight line broadening is always applied to limit the sharpness of lines before the spectral estimate is computed.

states with respect to frequency, followed by diagonalization in the smaller basis, is the origin of the name filter diagonalization [16].

Note that while the basis functions are filtered, the signal itself is not. If a large peak outside the window is contributing a significant baseline component to the region under consideration, it is still represented by the window basis, and the baseline within the window is well matched. That is, after the diagonalization at least one eigenvalue will refer to a frequency outside the window. It is possible to augment the window basis with a number of rather delocalized basis functions that can describe features outside the window more accurately, yet still give an overall eigenvalue problem that is not too large to handle. This multi-scale basis and its implementation have been described in detail previously [19]. It is unwise to tamper with the signal itself, by digital filtering for example [20], as the dispersion-mode line shape extends well away from the peak position. Its truncation in the frequency domain to try to generate a smaller pseudo-FID, by inverse Fourier transformation of just the window spectral region [20], leads to nonphysical contributions to the pseudo-signal. For the same reason, it is unwise to try to delete "spurious" entries from the FDM line list. A few may have frequencies outside the current window, and with large amplitudes and/or large widths and/or phases that seem to disagree with what one would expect from a genuine signal peak. However, they may be accurately characterizing the slow variation of the baseline or non-Lorentzian features within the window. A set of spectra that differ only slightly, say by different noise realizations, can be characterized by rather different lists of peaks, so that while the *spectrum* produced by FDM is a stable quantity, the individual eigenvalues and eigenvectors need not be. Producing a unique, reliable, and physically meaningful line list for an nD signal is apparently only possible in special cases, when it is already fairly obvious what the correct entries should be. Conventional peak picking of an FDM spectral region, after all the contributing windows have been processed, summed, and digitized with appropriate smoothing, is a usable workaround that avoids giving the specific realization of the line list any particular significance, but is in principle less accurate.

A profitable way to look at FDM is simply to consider other ways of attempting the fitting problem. There are many possibilities, but most are encompassed within two extremes. One approach could be a nonlinear least squares algorithm, directly in the frequency domain [21,22]. First compute the FT spectrum, with its resolution limited by the time–frequency uncertainty principle, and then consider a small region of the spectrum, attempting to fit it with some unknown linear combination of Lorentzians, convoluted with the sinc-function line shape due to the finite sampling time. For example, one could locally peak pick the spectrum and then try to adjust preliminary line positions, phases and intensities to obtain a good fit. This approach is hampered by tremendous problems with myriad local minima [22]. There is an apparently arbitrary choice of the number of lines used in the fit; the question of how to handle baseline roll, from peaks outside the area of consideration, is next on the list. Finally, the extent to which unresolved peaks can appear, for example, with huge intensity and opposite phases, etc., is hard to handle, so the whole approach is unlikely to be generally useful except in simple cases.

A second approach is to fit directly in the time domain using complex exponentials with damping. This is what FDM with the primitive basis functions attempts to do, and is also the goal of parametric LP. There is no problem in principle with local minima, because the formulation as a linear algebra problem guarantees, in exact arithmetic at least, a unique solution and a prescription to obtain it. Here the problems are numerical, namely the very large size of the linear algebra problem that is created, and the ill-conditioned matrices that have to be handled. This may create certain ambiguity due to the ambiguity in the method of handling the large and ill-posed numerical problem in finite arithmetic. In 1D FDM this ambiguity boils down to the choice of the frequency window which, while it may affect the line list (see above), hardly affects the estimated spectrum, which is the quantity of interest. In addition, LP has nettlesome problems with increasing exponentials (eigenvalues outside the unit circle in the complex plane). These exponentials must be arbitrarily deleted, or flipped to negative exponentials, before the linear least squares problem for the intensities of the lines can be solved. This leads to further problems that even the primitive formulation of FDM does not face. The intensities in FDM are not calculated in a separate step, and so errors in the frequency or width of one feature do not feed into the determination of the amplitude of another one like they can in parametric LP. When there are features like residual solvent signals, *etc.*, contaminating the time-domain data, a global fit of the FID, even were it computationally feasible, becomes distinctly unattractive.

Against this backdrop, FDM is an attempt to capture the best of both worlds, that is, *to fit the time domain data, but in the frequency domain.* Small linear algebra problems result, having unique solutions that can be obtained without any advance knowledge, or even initial guess, of the number of lines. Spectral windows containing residual solvent do not even need to be processed, so that many of the problems that would arise in a global fit of the FID do not surface at all. It is this conceptual sleight of hand that makes FDM particularly powerful, and appealing. It also underscores that it is the local density of basis functions that determines the

ability to resolve closely spaced lines. For example, oversampling the FID using a much larger spectral width produces many more data points, and so many more basis functions. However, the number of functions per unit frequency remains fixed, as does the approximate local resolving power when applying FDM.

2.3. Multidimensional bases, spectra, and regularization

The 2D extension of FDM is based on the assumption that there are two effective evolution operators U_1 and U_2 that commute. The iterative construction of the basis functions is generalized according to [23,24]

$$\Phi_{nm} = U_1^n U_2^m \Phi_{00} \tag{15}$$

with n and m ranging up to $\sim N_1/2$ and $\sim N_2/2$, respectively, showing that the number of basis functions depends on the *product* of the number of recorded data points. This is a key point of a multidimensional data analysis approach, in which 2D basis functions are used to describe 2D spectral features, 3D basis functions for 3D features, etc. Roughly speaking, the resolving power of 2D FDM depends on the *area* in the 2D time domain that has been measured rather than the individual linear extent of the two dimensions, as would be the case with FT analysis. This is a powerful argument in favor of an integrated nD time to nD frequency approach, and explains why Fourier transformation along the longer of the two time dimensions, as is routinely done in NMR processing, may prove to be disadvantageous. Fig. 2, for example, shows a simple case in which Fourier transformation in the long dimension followed by 1D FDM in the short dimension introduces line shape distortions, whereas 2D FDM uncovers the correct peaks.

How can one tell whether a method is truly multidimensional? One important key is whether *all the data* are required to start the calculation. In conventional FT spectroscopy, each t_1 increment can be transformed to an F_2 spectrum as it is acquired, that is, there is no requirement at all to have a complete set of interferograms in order to carry out the analysis of the data in the other time variable, t_2. In 2D FDM, the matrix elements of the evolution operators cannot be calculated using only part of the time-domain data, say, along one of the time axes. All the 2D time-domain data contributes to each matrix element of the evolution operators, and there is thus no factorization as in FT analysis.

While the potential resolving power of the 2D basis is greater, so are the technical and numerical stumbling blocks. Some of these have been discussed in some detail previously [9–12]. They break down into the following categories: (i) 2D FDM fits phase-modulated data to phase-twist line shapes, but most recorded data consist of complementary N- and P-type (or sin/cos) data sets; (ii) when data is even moderately noisy, the phases of the 2D peaks can be slightly in error; (iii) while the fit to phase-twist Lorentzian lines can be "converted" to absorption mode, the star-shaped contours that result are not as clean as the elliptical contours familiar from resolution-enhanced FT processing; (iv) there are apparently many more ways for 2D data to depart strongly from a small number of 2D phase-twists than there are for a 1D data set to fail to fit well to 1D Lorentzians; (v) it is no simple matter to identify the pair of frequencies that are supposed to correspond to a particular 2D feature—rather a list of *all* possible frequencies in each of the dimensions is the output of the diagonalizations; (vi) the **U** matrices can be extremely badly behaved numerically, much more so than in the 1D case, so that tiny perturbations of the input data result in large changes in the computed spectrum, a case requiring some sort of regularization; (vii) there could be some question about how small a spectral region can be isolated, i.e., how big the window basis should be for each calculation; (viii) there may be some more general kind of "uncertainty principle" when employing regulariza-

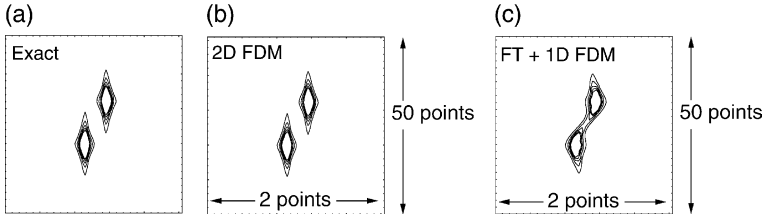

Fig. 2. Illustration of the difference between a direct 2D method and a 1D method applied to individual FT traces. (a) Two Lorentzian peaks. (b) The result of 2D FDM using 50 time points in the vertical dimension, and only 2 time points in the horizontal dimension. Both peaks are accurately obtained, as shown. (c) By contrast, Fourier transformation along the long dimension, followed by 1D FDM along the short dimension, produces this contour plot, in which new non-existent features have been created. The difference is that the 2D FDM approach used 25 2D basis functions to fit the region, whereas in the FT + 1D FDM approach only a single basis function was allowed for each of the 50 individual traces that result from the FT. A single basis function can result in only a single peak; the single frequency determined for each trace is a weighted mean of the two peaks, resulting in the strange ridging behavior. This kind of distortion is always present whenever a series of 1D methods (e.g., LP) is used if the number of points (or the filter length) in the short dimension is insufficient to characterize all the peaks on a given trace accurately. In this example, four points would result in a good fit, as there are at most two peaks on each horizontal trace. In general, the requirement to have a sufficient 1D basis density on the most crowded traces leads to a much larger data set than the 2D requirement to have enough local 2D basis functions.

tion, in that closely spaced lines that are degenerate in one of the frequency dimensions may tend to coalesce into a single feature while resolution in an orthogonal dimension improves.

We have dealt systematically with each of these aspects through a great deal of experimentation and trial-and-error. While it is premature to claim that all the issues have been resolved entirely there is, now, a clear way to deal with most of them. With regard to point (i), it seems that, for shorter data sets at least, a 2D gradient-selected spectrum with, say, $2N_1$ increments of either purely N- or P-type would be far superior to complementary N- and P-type data with only N_1 increments. For the shorter data sets, a calculation is carried out for the N-type data and P-type data *separately*, with the phase-twist results then folded in the usual way to obtain an absorption-mode spectrum. However, each individual calculation suffers from $\sqrt{2}$ worse signal-to-noise ratio (SNR) and, much more importantly, may fail to converge because of insufficient basis size; doubling the basis size in a single calculation may, however, lead to nearly quantitative results. Unfortunately, there is not yet any general scheme to process the two conventional data sets jointly and use them in concert to pin down peak positions more accurately. For CT data sets there is such a scheme, described later.

With regard to points (ii) and (iii), it is possible to artificially phase each peak by taking the real part of the complex amplitude, $\mathrm{Re}\{d_k\}$, assuming that the peaks have known phase by controlling the experimental conditions appropriately. All that survives in this approach is the algebraic sign of each peak. Likewise, it is possible to replace each Lorentzian by a corresponding Gaussian with the same full width at half maximum, and the same integral. The effect of these substitutions is shown in Fig. 3. Note that it is possible to plot a separate "absorption mode" spectrum using $\mathrm{Im}\{d_k\}$, to check for lines that might be badly out of phase. All these manipulations clearly lead to an illegitimate characterization of noise, other non-signal features, and overlapping peaks where the phase of individual components may be ill-defined, but we assume that these are not typically of great interest or, more candidly, that we cannot hope to extract them accurately. Furthermore, the regularization procedure that we employ may modify smaller features in any event.

With regard to point (iv), it seems that there is a significant and fundamental difference in complexity between 1D and nD signals. For example, a closely spaced doublet can usually be satisfactorily fit as a singlet at the mean frequency position in a 1D spectrum. However, a square doublet of doublets in a 2D spectrum cannot be fit well with, say, only three 2D peaks. One obtains an unstable fit that completely lacks the underlying symmetry of the data. Even with just two peaks of the same phase, the quality of the fit using a single peak

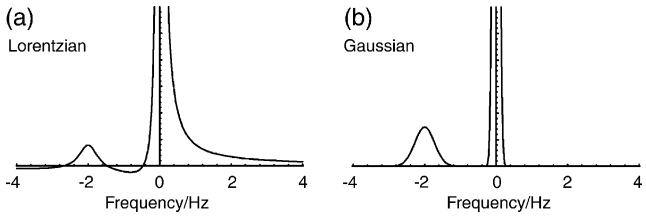

Fig. 3. Illustration of line shape transformation from slightly out of phase Lorentzian to purely absorptive Gaussian, using a 1D example. (a) Two Lorentzian peaks are shown. One is around 15° out of phase, as well as being 10 times more intense and four times narrower than the other one. The dip in the baseline makes the weaker peak hard to contour. (b) Transformation to absorption mode Gaussian line shape. Each Gaussian peak has the same full width at half maximum and the same integral as the corresponding in-phase Lorentzian peak. It is much easier to contour the weaker peak using the in-phase Gaussian representation because of the lack of any dispersive line shape and the narrow skirts of the Gaussian lines. Note, however, that noise and non-Lorentzian spectral features may be accurately fit as a linear combination of Lorentzian lines with various phases. In this case the Gaussian representation may be misleading, especially after the out-of-phase components are eliminated.

depends not only on their separation, but also on the relative orientation of the displacement vector connecting their frequency centers. When one considers that some 2D spectra contain significant "t_1-noise" that will not even remotely conform to the model of a fixed number of 2D phase-twist lines, it is clear that blind application of FDM can sometimes lead to unexpected results.

With regard to point (v), another complicating and unpleasant feature emerges in the multidimensional case. In 1D FDM the eigenvalue gives the position of the line and the eigenvector gives the intensity, and there is obviously no problem determining which eigenvalue matches which eigenvector. In 2D FDM, two eigenvalue problems arise but there is no straightforward simultaneous eigenvalue solver. Instead a list of eigenvalues and eigenvectors are obtained for each dimension separately. The eigenvectors along each dimension are, unfortunately, essentially completely different except in the case of a perfect model signal with vanishing noise. As a result, we may obtain 100 F_1 frequencies and 100 F_2 frequencies as 1D projections in a 2D frequency window, but no simple way of identifying each of the true frequency pairs. The solution to this problem is to formulate the spectrum in terms of the 2D resolvent, i.e., adopt a formula for the spectrum that does not ever require the identification of corresponding frequencies or joint eigenvectors. In the 2D case the absorption mode "aggressive spectrum" may be written [11],

$$A(f_1,f_2)=\sum_{k_1}\sum_{k_2}\mathrm{Re}\left[(\Phi_0|Y_{1k_1})(Y_{1k_1}|Y_{2k_2})(Y_{2k_2}|\Phi_0)\right]\times\mathrm{Im}\left[\frac{\mathrm{i}\tau_1}{1-u_{1k_1}\mathrm{e}^{2\pi\mathrm{i}\tau_1 f_1}}\right]\mathrm{Im}\left[\frac{\mathrm{i}\tau_2}{1-u_{2k_2}\mathrm{e}^{2\pi\mathrm{i}\tau_2 f_2}}\right] \tag{16}$$

and then the absorption Lorentzian lines converted into corresponding 2D Gaussian peaks. Instead of the expected M peaks allowed by the size of the basis, a total of M^2 peaks results from Eq. (16). Many of these are of vanishing intensity. However some are not, and are numerical artifacts that are the focus of point (vi). It is possible to compute a complex spectrum from Eq. (16) although it has, of course, phase-twist line shapes.

With regard to point (vi), the matrices involved in 2D FDM can be poorly behaved numerically, corresponding to a so-called ill-posed problem. Because the number of basis functions is large, it may greatly exceed the number of local peaks. This overcompleteness makes all $\mathbf{U}$ matrices close to singular. On the other hand, the restriction of the complete set of 2D Fourier basis functions to a local 2D window creates a potentially incomplete basis. So, the basis for the matrix representations may be both overcomplete and incomplete, and the true situation may vary from window to window. Add to this the confounding influence of noise, that cannot be efficiently characterized as 2D phase-twist Lorentzians, and the sensitivity to the exact parameters of the FDM calculation (window size, signal length used, etc.) emerges. Regularization is a partial cure for these kinds of problems, and consists of modifying the exact, and sensitive, solution to a more robust regular solution derived from a modified problem that employs some constraints on what constitutes an acceptable answer. The very simple approach that we have invented is to modify the generalized eigenvalue problem to the form [12]

$$\mathbf{U}^{(0)\dagger}\mathbf{U}^{(1)}\mathbf{B}_k = u_k\left\{\mathbf{U}^{(0)\dagger}\mathbf{U}^{(0)} + q^2\right\}\mathbf{B}_k \tag{17}$$

in which the singularity of the overlap matrix, and the joint null space of the two matrices, is effectively controlled by the non-negative regularization parameter q^2. There is a clear analogy to Tikhonov regularization [25], in which a similar structure is employed when solving nearly singular linear systems but, as far as we know, the formula in Eq. (17) has not been used before in generalized eigenvalue problems. It works in many situations, but is neither necessarily best nor most conservative. There may be other, more sophisticated approaches [26,27] to regularization of the generalized eigenvalue problem that are superior, but at this time they remain unproven for our application. Approximately speaking, the major effect of q^2 is to broaden all small signals, of the order of the noise, into oblivion. Significant peaks are also affected somewhat. The net result is a slightly broadened "theoretical" spectrum of the more significant features on a nearly noiseless base plane, making contouring particularly easy. If a range of stability can be found as q^2 is increased from zero, then the spectra are usually reliable. If no such range can be found then the basis may be too small, and acquiring more data may be the only option. Note that it is possible to get a rough estimate of the correct value of q^2 by processing a "blank" region of the 2D plane. The regularization parameter is increased until contouring the blank region shows only broad, smooth features of relatively small amplitude. More sophisticated ways of tuning q^2 can be envisioned, and are under consideration as part of a turnkey software package.

With regard to point (vii), some common sense must be used in defining the boundaries of the window. It makes no sense at all to consider a window smaller than, or even comparable to, typical line widths. Doing so will inevitably lead to errors in the computed spectrum. Likewise, if the number of points along one dimension is very small, it makes little sense to carve up the frequency range along this dimension. The assumption that accurate frequencies can be extracted from a small 2D window basis is only approximate, so that very small bases lead to noticeable error. Conversely, while the very largest generalized eigenvalue problem possible may therefore seem attractive, there are numerical limits to the reliability of the current library routines, and the cubic scaling of the latter mean that smaller is, up to a point, better both in terms of speed and reliability. Typically the $\mathbf{U}$ matrices range in size from 100×100 to 800×800 for a 2D calculation, depending on the number of time-domain data points involved. For reference, a 2D signal of size $N_1 \times N_2 = 128 \times 2048$ would have a full basis size of 65,536, leading to matrices with 4.29×10^9 matrix elements if a single calculation were attempted without using windows.

Finally, with regard to point (viii) it seems that assuming the 2D spectrum consists of an exact linear superposition of 2D phase-twist lines is equivalent to assuming that the operators U_1 and U_2 commute. This means that there is some compromise in the achievable resolution that can be obtained when these assumptions are untrue. It seems that a direct 2D method tends naturally toward a kind of "minimum uncertainty" representation, in which closely spaced lines along one dimension tend to merge together, while distinct shifts in the other dimension are teased apart. This observation may hinge on the particular regularization used in FDM2K, or it may be a more fundamental limitation: further research is needed.

2.4. Doubling scheme for CT experiments

As remarked earlier, the local basis density plays a key role when it comes to resolving spectral features. Assuming the acquisition time in the indirect dimension is relatively short, so that relaxation is not leading to large loss of signal, it would be advantageous to have a purely phase modulated signal of twice the length, rather than a set of N- and P-type data. However, the latter is an absolute must for FT processing, the most conservative method, because otherwise phase-twist line

shapes result. For most experiments, therefore, the optimum data for FDM and FT processing are different. Until FDM proves itself to have similar reliability to the FT, most experimentalists will not relish the all-or-nothing proposition of acquiring one phase modulated data set. Luckily, for the particular case of constant-time signals in the indirect dimension, no compromise is necessary. By simply concatenating the raw time-domain data sets, with the N-type data reversed in time, a continuous time-domain signal of twice the length results. Assuming zero phase at $t_1 = n_1\tau_1 = 0$, and $t_2 = n_2\tau_2 = 0$, which is readily achievable experimentally, the P- and N-type signals for a single peak may be written:

$$S_{\mathrm{P}}(n_1\tau_1, n_2\tau_2) = \exp[\mathrm{i}(\omega_2 n_2\tau_2 + \omega_1 n_1\tau_1)];$$
$$S_{\mathrm{N}}(n_1\tau_1, n_2\tau_2) = \exp[\mathrm{i}(\omega_2 n_2\tau_2 - \omega_1 n_1\tau_1)], \quad n_1\tau_1 = 0, \ldots, T, \tag{18}$$

where ω_2 is complex, but ω_1 is essentially real, as the imaginary part vanishes in the CT dimension. In a normal treatment the N-type frequencies appear to evolve forward in time but with opposite algebraic sign. Note, however, that we may consider them to evolve with the same algebraic sign if instead the time increment τ_1 is reversed in the second data set. That is, by preceding the P-type data set, in the time domain, with the N-type data in reverse chronological order, a pseudo-signal with t_1 ranging from $-T$ to T results

$$S_{\mathrm{NP}}(n_1\tau_1, n_2\tau_2) = \exp[\mathrm{i}(\omega_2 n_2\tau_2 + \omega_1 n_1\tau_1)], \quad n_1\tau_1 = -T, \ldots, T. \tag{19}$$

There are three advantages to this way of handling the data. First, it is completely compatible with conventional FT processing. Secondly, the CT data has, neglecting noise and spectrometer instability, nearly perfect Lorentzian line shape in the CT dimension. This means that the FDM model is very efficient. Finally, the doubling scheme doubles the number of basis functions and allows a single calculation, with the best chance of success, to be performed. Unfortunately, for real-time evolution, in which the ω_1 frequencies are also complex, reversing the ordering does *not* correspond to a reversal in time: the imaginary part of ω_1 would also have to change sign, and it does not do so. Note that no assumption that ω_1 is actually real is used in the FDM algorithm when the doubled data set is processed. Eigenvalues both inside and outside the unit circle are always produced when analyzing the data. The correct way to handle these to produce the best spectral estimate has already been described in great detail [11].

Only two modifications to existing FDM code need to be made to handle the doubled CT signals from conventional N- and P-type data. First, both data sets usually have a $t_1 = 0$ increment. These are averaged to arrive at the center point of the NP data set. The basis size is thus not quite doubled due to this redundancy. Secondly, it is important to take proper account of the true time origin when performing the Fourier transformation to the window basis matrix elements. Practically speaking, the matrix elements $\mathbf{U}^{(p)}_{nm}$ can be computed as previously described [9] and then a phase factor $\exp(\mathrm{i}\varphi_{nm}T)$ is used to take into account the shifted time origin.

2.5. Numerical experiments

Numerical studies on model signals plus noise have the advantage that one knows, exactly, the correct answer. As long as this information is not employed in any way by the algorithm, model data can be used as an objective benchmark by competing methods. We have made an attempt to objectify the performance of 2D FDM2K by examining model N- and P-type signals from a hypothetical constant time experiment, including various amounts of uncorrelated random noise. Fig. 4 shows a model spectrum meant to mimic a section of a CT-HSQC spectrum. There are 13 peaks of identical integral, with non-zero and similar line widths in F_2 and zero line width in F_1. The SNR is extremely high, around 1000:1 in the time domain. The exact spectrum (a) has been line broadened in F_1 to roughly the mean width in F_2 to allow meaningful contours to be drawn. The absorption mode DFT spectrum is shown in Fig. 4b, from a pair of N- and P-type data with size 10×10. Cosine apodization was used to limit the sinc-function wiggles and zero-filling to 512×512 points was used to obtain smooth interpolated contours. Some of the peaks are resolved, but others are too close to distinguish. In Fig. 4c the pseudo-absorption spectrum (512×512) from the 10×10 time-domain P-type data set is shown, using Eq. (16). Many of the peaks are obtained quantitatively, but local crowding results in several pairs of peaks being fit as a single broader peak. Note in addition that the apparent width of the single feature is less than the separation of the two genuine peaks. Convoluting this spectrum with the transform-limited line shape still gives a good match to the actual DFT spectrum (data not shown), but this is clearly not sufficient to guarantee that all the information has been obtained. Finally, in Fig. 4d the effect of processing a single "NP" data set is shown. The larger basis size (and possibly slightly better overall SNR) allows all the peaks to be obtained correctly. With experimental data, of course, one never knows in advance what the correct result is supposed to be! However, by doing several FDM2K calculations with different basis size, it is possible to identify stable recurring features, and also features that depend on the size of the basis. As long as a feature is stable with respect to basis size and the regularization parameter, our experience has been that it is quite accurate and reliable.

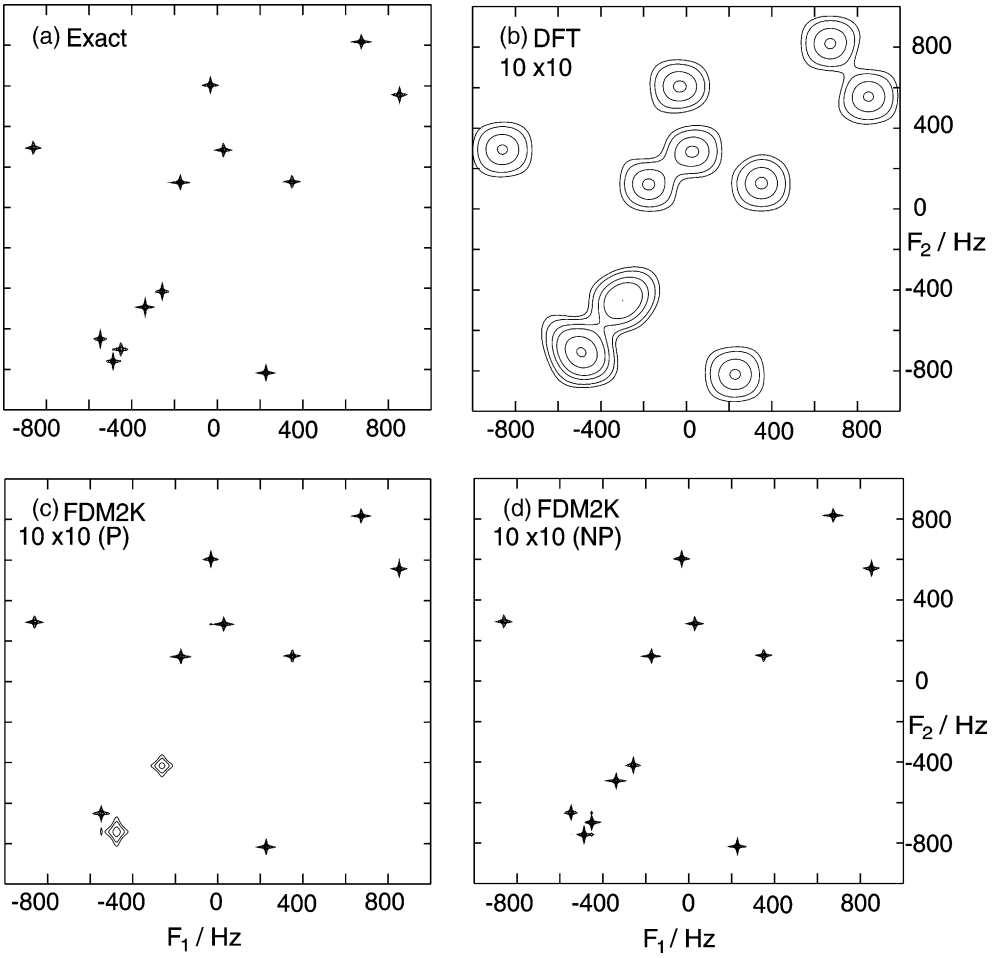

Fig. 4. A study of a model 2D signal with a high SNR of 1000:1 (first point) in the time domain. For such truncated data the true frequency domain SNR is somewhat ambiguous because there is essentially no baseline region in the FT spectrum. (a) The exact spectrum consists of 13 purely Lorentzian peaks with the positions shown and with line widths in the range 2.2 ± 0.2 Hz (at random) in F_2 and 0 Hz in F_1 modeling a constant time signal. The spectral region is 1 kHz by 1 kHz, so that the peaks are narrow enough that they are difficult to represent adequately with contours. The lines have been broadened by 2 Hz in F_1 to allow visualization of the 512×512 spectrum. (b) The absorption mode DFT spectrum of a 10×10 2D FID. Cosine apodization in both dimensions has been used to avoid very large truncation artifacts. The peaks show line widths in both dimensions of over 100 Hz, making it impossible to resolve all the features. Zero filling to a 512×512 data matrix has been employed to obtain smooth contours. (c) The FDM result with the same 10×10 data set, using only the P-type data and Eq. (16) to construct the 512×512 spectrum. A single window with 25 2D basis functions (a 5×5 grid) was used. Many of the lines are extracted quantitatively, but a few close pairs are fit as a single broader feature. (d) The FDM result with the same 10×10 data set, using the doubling scheme described to create the full NP signal. A single window with 45 2D basis functions (a 9×5 grid) was used. The FDM pseudo-absorption spectrum is nearly indistinguishable from the exact spectrum, 4(a), showing the improved resolution offered by the NP doubling scheme.

Fig. 5 is an attempt to be somewhat more realistic. The same features are present as in Fig. 4, but now the noise level is substantially higher, giving around 20:1 SNR in the time domain. A DFT of a pair of 18×18 time-domain data sets shows resolution of most, but not all, of the peaks in the absorption-mode spectrum, Fig. 5a. The FDM2K calculations all employ the doubling scheme, but now the results are less impressive. Using only a 10×10 signal there is now inadequate information for FDM2K to accurately identify all the peaks, as shown in Fig. 5b. Increasing the signal size to 14×14 in Fig. 5c sharpens up most of the features. But using the same 18×18 data set as used in the DFT calculation again allows all the features to be identified, as shown in Fig. 5d, although there are still some errors in the widths and intensities of the most crowded peaks. Further increase in the signal size results in only minor improvement to the spectrum, and is evidence that the calculation has converged onto the true features.

3. Experimental

All the ^{1}H–^{13}C CT-HSQC spectra shown were obtained using a 500 MHz Varian UnityPlus spectrometer, with a standard 5 mm HCN triple resonance probe. The sample was 700 μl of a 1 mM solution of 100% doubly labeled (^{13}C and ^{15}N) human ubiquitin in D_2O (VLI

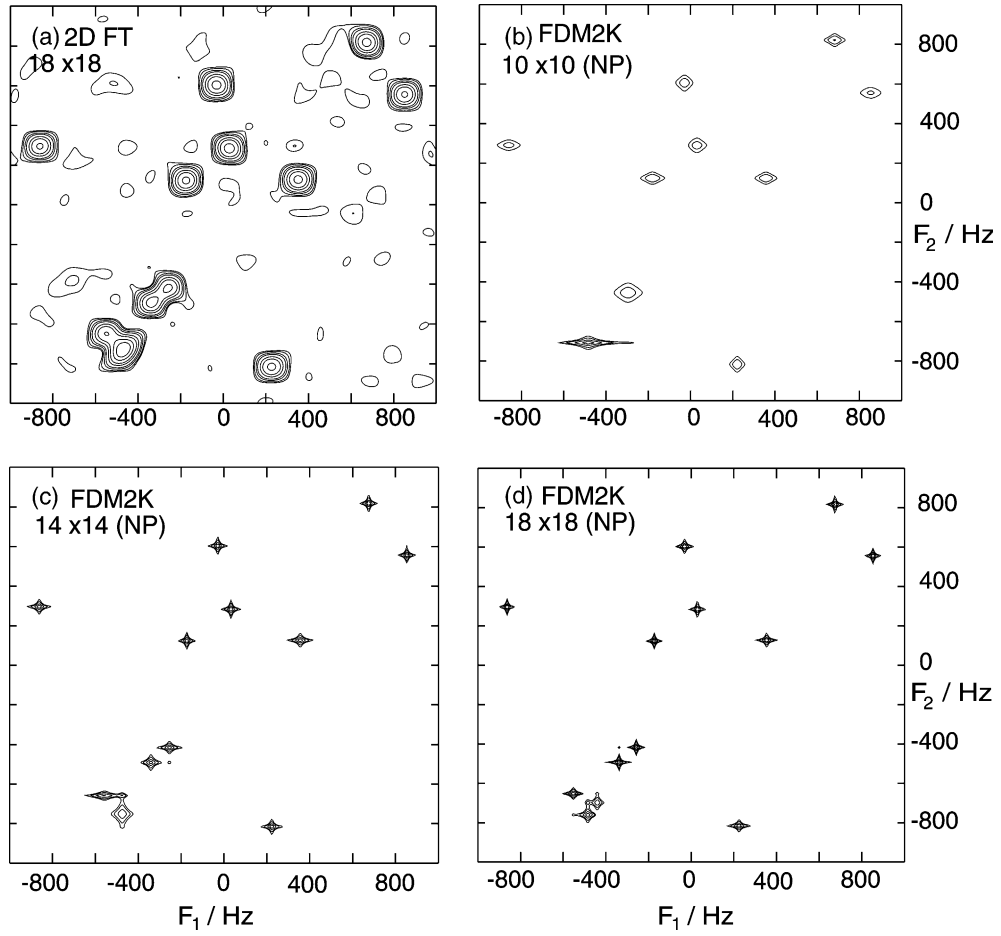

Fig. 5. A study of the same model 2D signal as Fig. 4 but with a low SNR of 20:1 (first point) in the time domain. (a) A contour plot of the 512×512 absorption mode DFT spectrum obtained from an 18×18 FID, with the contours drawn low enough to show some of the noise. (b) The FDM result with a 10×10 data set, using the NP doubling scheme. A single window with 45 2D basis functions (a 9×5 grid) was used. In contrast to Fig. 1d, the noisier data now requires larger regularization, resulting in broader peaks. Not all the peaks can be resolved, and the intensities are less reliable. (c) The FDM result with a 14×14 data set, using the NP doubling scheme. A single window with 91 2D basis functions (a 13×7 grid) was used. There is great improvement compared with (b) but not all the peaks are resolved. (d) The FDM2K result with the same 18×18 data set as (a), using the NP doubling scheme. A single window with 153 2D basis functions (a 17×9 grid) was used.

Research) at 24 °C. A slightly modified pulse sequence, shown in Fig. 6, was used to obtain the results. The partly selective *noco* and *co* pulses, of duration 200 μs, either invert the aliphatic region of the carbon-13 spectrum and not the carbonyls, or vice versa. They will be explicated in a related publication. Their only role here is to provide a spectrum with zero phase correction in F_1, with no Bloch-Siegert phase [28] shifts, and no loss of signal intensity over the slightly inhomogeneous radio-frequency field on the carbon-13 channel. Their short pulse width ensures that little of the $2T$ period is wasted.

As the brief gradients during the longitudinal states of the INEPT and reverse INEPT steps were enough to suppress the residual HDO resonance, amplitude modulated data sets were in fact acquired. The cosine and sine data sets required two scans each to complete the simplest difference phase cycle to select the ^{13}C-bearing protons. These two data sets were then combined to form pseudo-N- and P-type data sets, which were then processed as above. Calculations with actual gradient-selected data showed no significant difference as long as the experiments were done carefully and the same number of increments obtained. For ultra-short CT experiments, the loss of time during the actual pulsed field gradient is disadvantageous.

3.1. Noise performance

The numerical experiments of the last section carry over fairly well to actual CT-HSQC data. Fig. 7 shows a zoomed section, in a crowded portion of the C_α region, of a ^{1}H–^{13}C CT-HSQC spectrum. The spectra in Fig. 7 have been obtained using only two scans per increment for the N- and P-type data sets, and with a short constant time $2T = 7.5$ ms, well below the null in the CT transfer function at ~ 13 ms, and almost a factor of four

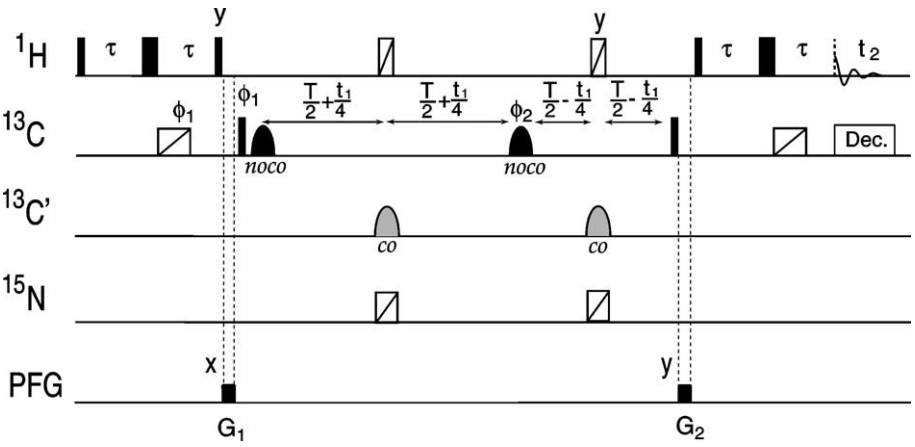

Fig. 6. Modified CT-HSQC pulse sequence to observe correlations between H_α and C_α spins in a protein. Aside from the minor addition of a pair of partly selective inversion pulses, denoted *noco* and *co*, and the broadband inversion pulses (BIPs), denoted with open scored icons, the pulse sequence is standard fare. Narrow filled icons are conventional 90° pulses, and wide filled icons are conventional 180° pulses. The 200 μs *noco* and *co* pulses have been developed for ultra-short CT experiments, in which longer pulses reduce the available number of increments. They work to invert the aliphatic region of the ^{13}C spectrum without inverting the carbonyl region (*noco*) or vice-versa (*co*). The BIPs cleanly invert the entire chemical shift range of the indicated nucleus. All these pulses show superb compensation for RF inhomogeneity. The proton 90° time was 10 μs and the carbon-13 90° time was 15 μs. The *co* and *noco* shaped pulses are applied with a maximum of 13 kHz; the other pulses are applied at full power. All the spectra shown result from two scans per increment, in which the phase φ_2 is 0° and 90°, and the receiver phase is $2\varphi_1$ The phase φ_1 of the first carbon-13 90° pulse is advanced inverted as t_1 is incremented, and the sin/cos data acquired according to the States-TPPI scheme. By using the pair of partly selective 180 s, all phase shifts are removed, leading to a spectrum with zero phase correction in the indirect dimension.

shorter than the usual time $2T = 1/{}^1J_{CC} \approx 26.4$ ms that is typically used in these experiments. The spectral width in F_1 covered the full carbon-13 chemical shift range, leading to 90 increments each of N- and P-type data. The spectra required about 7 min of instrument time to obtain. Lowering the power of the last proton 90° pulse has mimicked the effect of poorer sensitivity. The 2D FT spectra are on the top panels and the 2D FDM2K spectra from the same data are directly underneath. The FT spectrum from the FID of the first increment, with some blank downfield region included for a noise reference, is shown along the top of each of the series.

As the effective read pulse is reduced to 11° and finally 5°, corresponding to a relative SNR 5 and 10 times lower, respectively, it is clear from the first increment that the data becomes quite noisy. Nevertheless, FDM2K with the new doubling scheme behaves remarkably predictably, with the achievable resolution gradually diminishing, and weaker features disappearing along with the noise when the regularization is used. There is clearly no catastrophic loss of performance, and the resolution in many regions is always preferable to the FT result. The fine structure along the proton dimension, a known result of a flip angle effect [29] when the small flip angle read pulses are used, is blurred away by the regularization. The result is a set of nearly structureless ellipses centered on the correct chemical shifts. Particularly with the noisier spectra, attempts to resolve the proton multiplet structure, and simultaneously separate the close carbon-13 shifts, have not been successful.

3.2. Ultra-short constant time performance

Attenuation of C_α magnetization during the constant time period is a common problem in larger proteins, where the T_2 losses can be severe. However, using larger $\mathbf{B}_0$ fields in conjunction with FDM2K may allow very short constant time periods, greatly improving sensitivity without the need for deuteration. This is in contrast to the usual fixed value of $2T \approx 26.4$ ms, independent of the field strength. The difference is that by operating with shorter periods than the null condition at 13 ms, the delay can be shortened as much as we like, fixing only the number of increments obtained. For example, if a decent CT-HSQC spectrum can be obtained using $2T = 4.25$ ms at 500 MHz, then only 2.6 ms would be required at 800 MHz. The relative basis function density versus the number of local peaks is identical. However, there can be a great improvement in performance both on account of the superior intrinsic sensitivity of the higher field strength, and the reduction in relaxation losses. The local nature of the convergence of FDM means that many of the spectral regions can be essentially completely resolved, leaving a few dense regions that may require a longer experiment, some editing techniques, or a higher-dimensional experiment to make headway. Alternatively, the *same* 4.25 ms experiment may, at higher field, allow completely converged spectra to be obtained. The ability to achieve high resolution with very short constant time periods could establish FDM as a key component in successful NMR of larger proteins at high field.

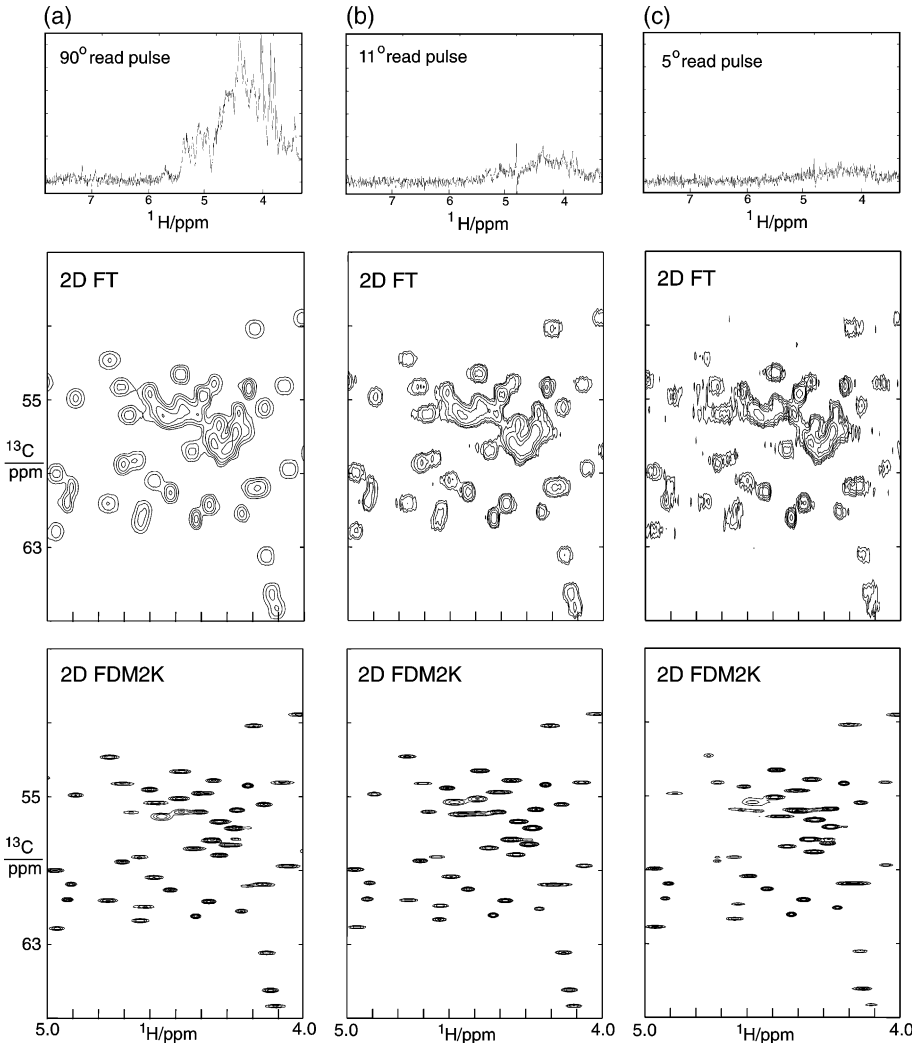

Fig. 7. Demonstration of the noise performance of FDM2K. The top traces show the FT of the FID from the first increment, with some downfield region included to show the noise level. The contour plots below show the 2D FT (top series) and 2D FDM2K (bottom series) spectra. The pulse sequence of Fig. 6 has been used with the power of the 10 μs read-out proton 90° pulse attenuated to achieve the desired flip angle. The 2D FT spectra are shown in the top series, and the 2D FDM2K spectra, from the doubled NP data set, are shown directly underneath. Regularization has been increased as the SNR decreases from left to right. Two scans per increment were recorded for the total of 90 increments, for the sine and cosine data sets: (a) 90° read pulse, (b) 11° read pulse (−18 dB), (c) 5° read pulse (−25 dB).

Human ubiquitin is *not* a rapidly relaxing protein, but can be studied simply to assess the degree of separation of the resonances that can be obtained with ultra-short constant times. Fig. 8a shows the conventional absorption-mode spectrum in the C_α region of the CT-HSQC spectrum using $2T = 26.4$ ms and the pulse sequence of Fig. 6. Two scans per increment for the N- and P-type data were used. This is our reference spectrum for the study. Many of the 2D peaks are clearly resolved, with a couple of denser regions. In the right-hand panel, Fig. 8b, the FT spectrum obtained with $2T = 4.3$ ms is shown. Many of the distinct peaks have coalesced because of the reduced resolution along the F_1 dimension. In Fig. 8c the result of processing the P-type data only is shown for the 4.25 ms data set, using the "aggressive" formula of Eq. (16) and Gaussian line shapes. A similar result was obtained with the N-type data. The spectrum shown is sensitive to changes in the regularization parameter, as well as other parameters of the FDM calculation (exact signal length, exact window size, etc.) and would not be considered reliable. Many regions of the spectrum remain unresolved, as the basis density is apparently too low to capture the local number of peaks. There are only eight basis functions along the C_α dimension in the region displayed. Finally in Fig. 8d the FDM2K result using the same data, but now with doubling, is displayed. Clearly the NP signal is a case where the whole is greater than the sum of the parts. Many of the regions display essentially identical resolution to the reference FT experiment. In the most crowded region the basis function density is still insufficient to characterize all the local 2D peaks, and so a

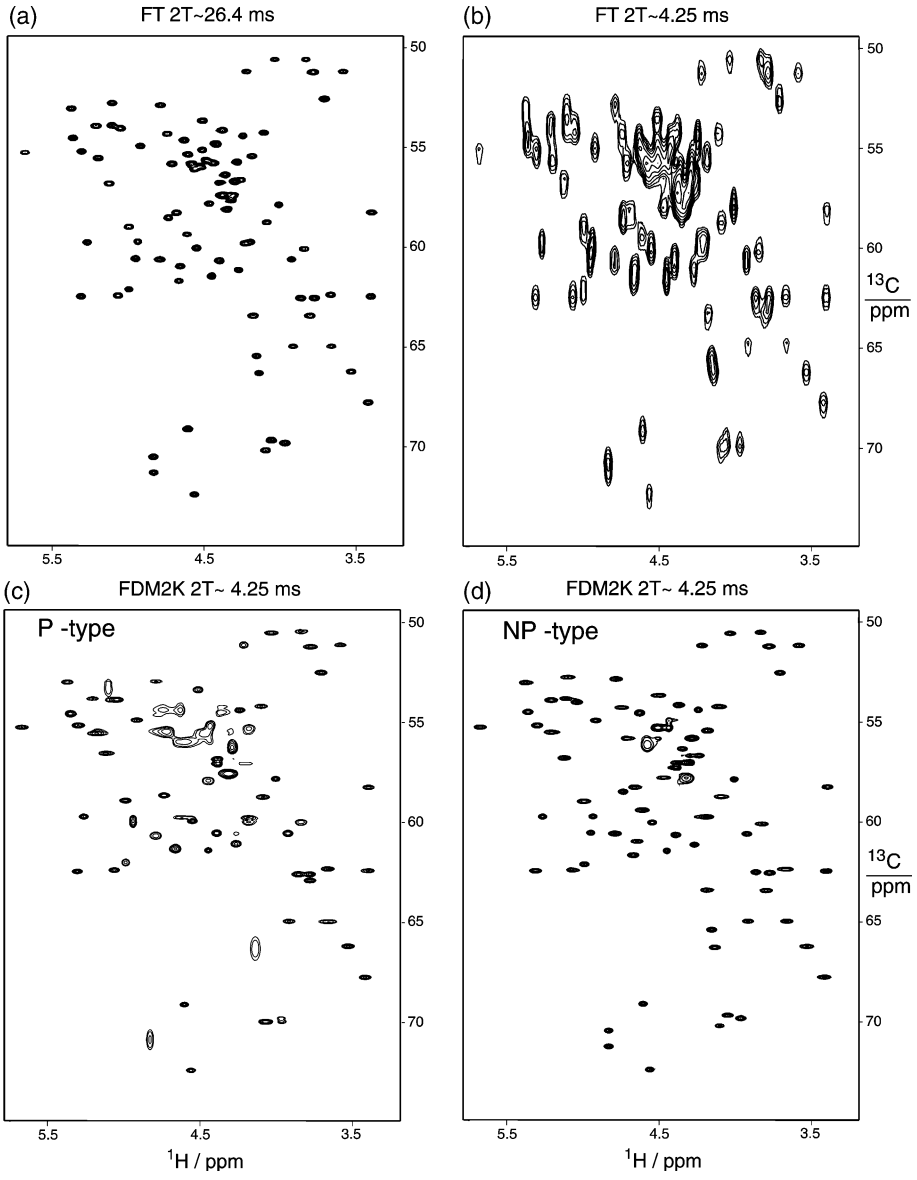

Fig. 8. Ultra-short constant time spectra compared with conventional measurement. (a) 2D FT spectrum using 26.4 ms constant time. This is the reference spectrum. (b) 2D FT spectrum using 4.25 ms constant time. There is a predictable loss of resolution along the indirect dimension, so that many of the peaks cannot be cleanly resolved. (c) 2D FDM2K spectrum using the same 4.25 ms data set as in (b), processing only the P-type data. This 2D spectrum is not stable, in many regions, with respect to changes in regularization parameter and would not be considered reliable. (d) Effect of the NP doubling scheme, again using the 4.25 ms data set. There is a vast improvement in performance compared with (c). Almost every region of this spectrum is stable, the exception being the most crowded area. Comparison of the reference spectrum, (a), shows that the peaks that are resolved are also very accurately obtained by FDM2K. The peaks superimpose almost perfectly on those from the much longer experiment that gives the FT spectrum. Note that all these spectra were obtained using only two scans per increment. This led to only 50 increments for the 4.25 ms experiment, making this an ultra-fast acquisition as well—less than 4 min.

distorted spectrum with some wide features is obtained. There are ways to conclusively show that regions like this are not converged, but no simple and reliable way to correct the inaccurate results. For example, a true resonance should have a line width of nearly zero along the F_1 dimension, with errors that depend on the noise level and quality of the data. The "unconverged" features have line widths that are quite large, and that are sensitive to the exact parameters of the FDM calculation, allowing them to be flagged as unreliable. Another approach is to use the FDM parameters to simulate the FT spectrum, and compare the result with the actual DFT of the data. Ideally, only featureless noise should remain after subtraction, if a perfect fit is obtained. In unconverged windows, a substantial residual may be used as evidence of untrustworthy local spectral features. We will describe these validation experiments in a further series of papers on practical aspects of FDM.

The calculation of the displayed regions using FDM2K is completed in a matter of a few minutes on either an AMD XP 1800+ Athlon PC running Linux, or an older 533 MHz DEC Alpha workstation. These are certainly inexpensive computers. Calculation of the entire 2D spectrum takes of the order of 2–3 h depending on the size of the window basis. As many regions are completely devoid of signals, however, a more realistic estimate is around 45 min for the relevant areas. With an appropriate parallel architecture, the computing time could be a few minutes for virtually any size data set, as each window is a separate job.

4. Conclusions

Direct multidimensional spectral estimation is a potentially useful adjunct to more conventional multidimensional FT spectral estimation. The resolution in a direct 2D method depends mostly on the total number of basis functions per unit area frequency, which in turn depends on the area in the time domain that has been measured. Long time acquisition in one dimension can thus be parlayed into better resolution along an orthogonal dimension for which it may be time-consuming, or impossible, to obtain a long data record. This is exactly the situation that pertains in heteronuclear multidimensional NMR experiments, especially those limited to a certain constant time period. When the number of 2D basis functions can dominate the true number of 2D peaks, a good fit can be obtained if the lines conform to the model of 2D phase-twist Lorentzians. For many proteins under typical conditions, the lines appear to be sufficiently well characterized in this way.

It is important to stress that amplitude modulated data are unfavorable for 2D FDM precisely because the number of spectral peaks is doubled compared with the corresponding phase modulated data set, resulting in poorer convergence. Thus, when amplitude modulated data are obtained, they should be first combined into N- and P-type data. It appears that processing the N- and P-type independent data sets, and then combining the results, is not greatly superior to just processing one of them. In particular, bad local regions remain unresolved. The SNR effect of processing separate N- and P-type data sets by FDM, followed by co-addition, is more complex than the $\sqrt{2}$ gain in sensitivity as in the FT case. The regularization step effectively suppresses the noise and all signals of about the same size, and this regularization is applied to the separate, and noisier, data sets before the results are combined. Line shapes and intensities for more significant peaks may improve somewhat, however.

Conversely, the constant time experiment is an especially favorable case for FDM because the basis size can actually be doubled while the number of peaks remains the same: in the NP data set all the information is put to use at once. There is a clear and telling improvement in resolution by the doubling of the basis size. In addition, the instrumental line shape in the indirect dimension is suppressed by the constant time evolution, resulting in peaks that match the Lorentzian model perfectly in at least one of the dimensions. This seems to be an important factor in obtaining the best results.

Optimization of nD NMR experiments has always presupposed that FT processing will be used to arrive at the nD spectrum. For example, 3D NMR is quite often employed because decent digital resolution can be obtained without extravagant experiment time. As such, a series of different 3D experiments may be used to correlate all the desired chemical shifts, and comparing planes from experiments taken at different times is usually mandatory to complete the assignment. Whether this is the most profitable use of human and instrument time, once nD FDM2K is reduced to practice, is certainly open to question. In FDM the digital resolution of individual dimensions has no direct relation to the ultimate spectral resolution. Further, we have shown here that the best data set is the one that gives the largest basis set. Two different 3D experiments do not lend themselves to this maximal basis set in the same way that one 4D experiment does. While extra delays may lead to some additional losses in the higher-dimensional experiment, there is also a gain in sensitivity by having all the data in one spectrum rather than divided up between several: this is important when regularization is employed. In 3D experiments where a set of nuclei are merely used to transfer magnetization, for example the CBCA(CO)NH pulse sequence [30], the delays are already present, and removing the parentheses to record the carbonyl chemical shift in a 4D experiment will be more sensitive and highly advantageous.

In closing, it is interesting to note that the time it takes to identify M random-frequency peaks accurately in an nD experiment depends on M but not n when FDM is employed. Indeed, as local spectral crowding is the only limiting factor once sufficient SNR is achieved, increasing the dimensionality of the experiment, so that the resonances are dispersed as widely as possible in frequency space, may result in a *shorter* overall experiment. For backbone assignment experiments on proteins with known primary structure, the expected value of M is known beforehand, so that the time required to obtain a converged high-dimensional FDM spectrum should be predictable, and scale with M. With the SNR of the first increment and the value of M in hand, it may be possible to semi-automate the data acquisition of even quite complex experiments. We intend to test this assertion in a future series of publications on the efficacy of FDM in 3D, 4D, and 5D experiments that have been specifically designed with the method in mind.

Acknowledgments

This work was supported by the National Science Foundation, CHE-9900422 and CHE-0108823. VAM is currently an Alfred P. Sloan Foundation research fellow.

References

[1] A. Bax, R. Freeman, Investigation of complex networks of spin–spin coupling by two-dimensional NMR, J. Magn. Reson. 44 (1981) 542–561.
[2] G.W. Vuister, A. Bax, Resolution enhancement and spectral editing of uniformly C-13-enriched proteins by homonuclear broadband C-13 decoupling, J. Magn. Reson. 98 (1992) 428–435.
[3] G. Bodenhausen, D.J. Ruben, Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy, Chem. Phys. Lett. 69 (1980) 185–189.
[4] R. Powers, A.M. Gronenborn, G.M. Clore, A. Bax, 3-dimensional triple-resonance NMR of C-13/N-15-enriched proteins using constant-time evolution, J. Magn. Reson. 94 (1991) 209–213.
[5] P. Schmeider, A.S. Stern, G. Wagner, J.C. Hoch, Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain, J. Biomol. NMR 4 (1994) 483–490.
[6] C.F. Tirendi, J.F. Martin, Fast linear prediction processing in two-dimensional NMR spectroscopy, J. Magn. Reson. 81 (1989) 577–585.
[7] G. Zhu, A. Bax, Improved linear prediction for truncated signals of known phase, J. Magn. Reson. 90 (1990) 405–410.
[8] 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.
[9] V.A. Mandelshtam, The multidimensional filter diagonalization method. I. Theory and numerical implementation, J. Magn. Reson. 144 (2000) 343–356.
[10] H. Hu, A.A. De Angelis, V.A. Mandelshtam, A.J. Shaka, The multidimensional filter diagonalization method. II. Application to 2D projections of 2D, 3D, and 4D NMR experiments, J. Magn. Reson. 144 (2000) 357–366.
[11] V.A. Mandelshtam, FDM: the filter diagonalization method for data processing in NMR experiments, Prog. NMR Spectrosc. 38 (2001) 159–196.
[12] J. Chen, V.A. Mandelshtam, A.J. Shaka, Regularization of the two-dimensional filter diagonalization method: FDM2K, J. Magn. Reson. 146 (2000) 363–368.
[13] W.J. Hehre, R. Ditchfield, J.A. Pople, Self-consistent molecular-orbital methods. 12. Further extensions of Gaussian-type basis sets for use in molecular-orbital studies of organic molecules, J. Chem. Phys. 56 (1972) 2257–2261.
[14] C. Cohen-Tannoudji, B. Diu, F. Laloë, Quantum Mechanics Volume One, Wiley, New York, 1977, Chapter V.
[15] A.R. Edmonds, Angular Momentum in Quantum Mechanics, Third Printing, Princeton University Press, 1974, Chapter 2.
[16] M.R. Wall, D. Neuhauser, Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode frequencies from a small number of residues or a short-time segment of a signal. I. Theory and application to a quantum-dynamics model, J. Chem. Phys. 102 (1995) 8011–8022.
[17] V.A. Mandelshtam, H.S. Taylor, Harmonic inversion of time signals and its applications, J. Chem. Phys. 107 (1997) 6756–6769.
[18] H. Hu, Q.N. Van, V.A. Mandelshtam, A.J. Shaka, Reference deconvolution, phase correction and line listing of NMR spectra by the 1D filter diagonalization method, J. Magn. Reson. 134 (1998) 76–87.
[19] J. Chen, V.A. Mandelshtam, Multi-scale filter diagonalization method for spectral analysis of noisy data with non-localized features, J. Chem. Phys. 112 (2000) 4429–4437.
[20] M. Deschamps, I. Burghardt, C. Derouet, G. Bodenhausen, D. Belkic, Nuclear magnetic resonance study of xenon-131 interacting with surfaces: effective Liouvillian and spectral analysis, J. Chem. Phys. 113 (2000) 1630–1640.
[21] F. Abildgaard, H. Gesmar, J.J. Led, Quantitative analysis of complicated nonideal Fourier transform NMR spectra, J. Magn. Reson. 79 (1988) 78–89.
[22] H. Gesmar, J.J. Led, F. Abildgaard, Improved methods for quantitative spectral analysis of NMR data, Prog. NMR Spectrosc. 22 (1990) 255–288.
[23] M.R. Wall, T. Dieckmann, J. Feigon, D. Neuhauser, Two-dimensional filter-diagonalization: spectral inversion of 2D NMR time-correlation signals including degeneracies, Chem. Phys. Lett. 291 (1998) 465–470.
[24] V.A. Mandelshtam, H.S. Taylor, Multidimensional harmonic inversion by filter-diagonalization, J. Chem. Phys. 108 (1998) 9970–9977.
[25] A.N. Tikhonov, Solutions of incorrectly formulated problems and the regularization method, Soviet Math. Dokl. 4 (1963) 1035–1038.
[26] J. Demmel, B. Kågström, The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. Part I: Theory and algorithms, ACM Trans. Math. Software 19 (1993) 160–174.
[27] J. Demmel, B. Kågström, The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. Part II: Software and applications, ACM Trans. Math. Software 19 (1993) 175–201.
[28] L. Emsley, G. Bodenhausen, Phase shifts induced by transient Bloch-Siegert effects in NMR, Chem. Phys. Lett. 168 (1990) 297–303.
[29] S. Schaublin, H. Hohener, R.R. Ernst, Fourier spectroscopy of nonequilibrium states, application to CIDNP, Overhauser experiments and relaxation time measurements, J. Magn. Reson. 13 (1974) 196–216.
[30] S. Grzesiek, A. Bax, An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins, J. Magn. Reson. 99 (1992) 201–207.