Journal of Magnetic Resonance 207 (2010) 17–23

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Filter diagonalization using a "sensitivity-enhanced basis": Improved performance for noisy NMR spectra

Hasan Celik [1], A.J. Shaka *

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

ARTICLE INFO

Article history:
Received 2 April 2010
Revised 17 July 2010
Available online 7 August 2010

Keywords:
Filter diagonalization
Noise
Generalized eigenvalue problem
Data fitting
2D NMR
Resolution enhancement
Sensitivity

ABSTRACT

The Filter Diagonalization Method (FDM) has been used to process NMR data in liquids and can be advantageous when the spectrum is sparse enough, the lines are sharp and Lorentzian, raw sensitivity is adequate, and the measured time-domain data is short, so that the Fourier Transform spectrum exhibits distorted line shapes. Noise can adversely impact resolution and/or frequency accuracy in FDM spectral estimates. Paradoxically, more complete data can lead to worse FDM spectra if there is appreciable noise. However, by modifying the numerical method, the FDM noise performance improves significantly, without apparently losing any of the existing advantages. The two key modifications are to adjust the FDM basis functions so that matrix elements computed from them have less noise contribution on average, and to regularize each dimension of a multidimensional spectrum *independently*. The modifications can be recommended for general-purpose use in the case of somewhat noisy, incomplete data.

1. Introduction

Introduced by Wall and Neuhauser in 1995 [1], the Filter Diagonalization Method (FDM) is a general way to fit a signal $C(n\tau)$ recorded on a regular time grid as a linear superposition of complex sinusoids (Eq. (1).

$$C(n\tau) = \sum_{k=1}^{K} d_k e^{in\omega_k\tau}, \quad n = 0, \ldots, 2M-1, \tag{1}$$

$$\omega_k = 2\pi f_k + i\gamma_k. \tag{2}$$

It was refined [2] and improved [3], but the core of FDM has remained mostly unchanged. Although originally designed to treat neither spectroscopic nor multidimensional data, FDM was adapted to process NMR spectra of both large and small molecules in liquids in 1D [4], 2D [5–7], 3D [8,9], and 4D [10] spectra, with the multidimensional cases requiring important modifications [6]. FDM outperformed the discrete Fourier Transform (DFT) whenever sensitivity was good, the spectrum was sufficiently sparse, the lines were sharp and Lorentzian, and the extent of the measured time-domain data was short, so that the FT spectrum exhibited line shapes that were distorted by convolution with a sinc-function. If the desired NMR spectrum can be cast efficiently as a linear combination of Lorentzian peaks with variable position and width (ω_k), and amplitude and phase (d_k), FDM should be efficient. As NMR resonances are well modeled by this form, FDM expedites data collection whenever spectra have sharp lines and little noise. Unfortunately, the fidelity of the FDM spectral estimate degrades as the noise level increases. Noise is not captured by a small number of Lorentzian peaks, so a decaying signal may not improve the spectral estimate at all, instead introducing spurious "peaks". More data points may also degrade the accuracy of the parameters describing the desired signal peaks, making it advantageous in some cases to truncate the time signal artificially. The tipping point can be well before the free-induction decay (FID) has visually fallen into the noise, and 2D or higher spectra are more delicate. In fact, FDM was used primarily for constant-time (CT) spectra where *no* signal decay occurred except in the directly-detected (running-time) dimension. These spectra gave superior results compared to the real-time multidimensional spectra [7–10]. By contrast, FT spectra almost always benefit from a longer signal as long as the signal has not decayed into the noise. Noise appears as low amplitude incoherent features that are only rarely misinterpreted as true resonances.

The sensitivity of the FDM spectral estimate to the data length adds unwelcome complexity to the method: one should try different data lengths, compute the spectrum, and *then* try to hone in on the optimum length to use. This optimization is open to bias. In 2D applications, where data sets may be highly asymmetrical (long in F_2 and short in F_1), artificially shortening the data used in F_2, to compensate for sensitivity to noise, can lower resolution in F_2 even as the resolution improves in F_1. As the signal-to-noise ratio (SNR) of a 2D time signal $C(t_1,t_2)$ degrades rapidly as t_1 increases [11], the overall contribution of noise can result in artifactual peaks or loss

* Corresponding author. Fax: +1 949 821 9920.
E-mail addresses: hcelik@uci.edu (H. Celik), ajshaka@uci.edu (A.J. Shaka).
[1] Fax: +1 949 821 9920.

1090-7807/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2010.07.019

of resolution in F_2, depending on the degree of regularization (see below) employed. It would thus be desirable to reformulate FDM so the results are less degraded by noise for complete data, but still show superior resolution for incomplete data. In this way, essentially all the data can be used each time, and just one calculation conducted. We present a way to move toward this goal.

2. Theory

2.1. The sensitivity-enhanced basis

Given an N-point 1D time signal, the solution to Eq. (1) is obtained by representing the time domain signal as a discrete time autocorrelation function[2]:

$$c_n = (\Phi_0|\widehat{U}^n|\Phi_0), \tag{3}$$

so that extracting the eigenvalues and eigenvectors of $\widehat{U}$ solves Eq. (1) [2]. No explicit information about the "Hamiltonian" of this fictitious system, nor its "initial state" Φ_0, is required [1] to obtain a numerical solution if a matrix representation of the time evolution operator $\widehat{U}$ is available in some basis. The matrix elements $\mathbf{U_{nm}}$ can be expressed by constructing an iterative basis from Φ_0:

$$\begin{aligned} &|\Phi_1) = \widehat{U}|\Phi_0), \quad |\Phi_2) = \widehat{U}|\Phi_1) = \widehat{U}^2|\Phi_0), \ldots, \\ &|\Phi_n) = \widehat{U}^n|\Phi_0); \quad n = 0, 1, 2, \ldots, M-1, \;\; M = N/2. \end{aligned} \tag{4}$$

Assuming Eq. (3) is satisfied, the matrix elements of $\widehat{U}^p$ in this basis are simply the measured data shifted forward by p time steps:

$$\mathbf{U^{(p)}_{nm}} = (\Phi_n|\widehat{U}^p|\Phi_m) = (\Phi_0|\widehat{U}^{n+m+p}|\Phi_0) = c_{n+m+p} \tag{5}$$

which gives a generalized eigenvalue problem

$$\mathbf{U^{(1)}B_k} = u_\mathbf{k}\mathbf{U^{(0)}B_k} \tag{6}$$

as the way to fit the data. Unfortunately, the size of these matrices is absolutely huge for most data sets [2], making this approach impractical. Changing to a Fourier basis [1,2],

$$|\Psi_j) = \sum_{n=0}^{M-1} e^{-2\pi i n \tau f_j}|\Phi_n), \tag{7}$$

allows a far more efficient divide-and-conquer strategy [2,12] where localized eigenvalues can be extracted in a frequency "window" of interest, $f_{min} < f_j < f_{min} + \Delta f$ and $\Delta f \ll 1/\tau = SW$ where SW is the full spectral width. The matrix elements $(\Psi_{j'}|\widehat{U}^p|\Psi_j)$ in the Fourier basis can be evaluated straightforwardly

$$\begin{aligned} \mathbf{U^{(p)}_{j'j}} &= (\Psi_{j'}|\widehat{U}^p|\Psi_j) = \sum_{m=0}^{M-1}\sum_{n=0}^{M-1} e^{-2\pi i m \tau f_{j'}} e^{-2\pi i n \tau f_j} (\Phi_m|\widehat{U}^p|\Phi_n) \\ &= \sum_{m=0}^{M-1}\sum_{n=0}^{M-1} z_{j'}^{-m} z_j^{-n} c_{n+m+p}, \quad z_j = e^{(2\pi i f_j \tau)} \end{aligned} \tag{8}$$

and were simplified by rearranging Eq. (8) and then analytically evaluating the sums [2,12], giving

$$\begin{aligned} \mathbf{U^{(p)}_{j'j}} = &\frac{1}{z_j - z_{j'}} \\ &\times \left(z_j \sum_{l=0}^{M-1} z_{j'}^{-l} c_{l+p} - z_{j'} \sum_{l=0}^{M-1} z_j^{-l} c_{l+p} - z_j^{-(M-1)} \sum_{l=M}^{2M-2} z_{j'}^{-l} c_{l+p} + z_{j'}^{-(M-1)} \sum_{l=M}^{2M-2} z_j^{-l} c_{l+p} \right) \\ &\times z_j \neq z_{j'} \end{aligned} \tag{9}$$

$$\mathbf{U^{(p)}_{jj}} = \sum_{l=0}^{2M-2} (M - |M - l - 1|) z_j^{-l} c_{l+p}, \;\; z_j = z_{j'}. \tag{10}$$

[2] The inner product is complex symmetric rather than Hermitian, hence the use of rounded parentheses rather than the more common Dirac bra-ket notation.

Aside from the numerical efficiency, Eq. (10) also clarifies the nature of the matrix elements showing that for $p = 0$ or 1 they amount to an $(N-1)$-point FT of the data with a severe triangular weighting function that peaks in the midpoint of time domain. The information content of the data points c_n is thus apparently emphasized in a rather undemocratic way in the matrices used to fit the data, with the center points contributing most to the matrix elements, while the first and last points contribute hardly at all. This triangular weighting follows on inexorably as a consequence of how the matrix elements in the Fourier basis are tied to the measured data points. As more data is collected, the weighting also becomes relatively more severe. This apparent emphasis has no consequence in the case of a noiseless model signal, where quantitative results are invariably obtained to many digits of accuracy. But because the midpoint will usually have reduced SNR for any decaying experimental noisy signal, the consequence for longer experimental signals with finite SNR may be quite different. The presence of noise becomes especially problematic when real-time evolution periods are used in multidimensional NMR [13]. The triangular weighting harks back to the "pseudo-echo" function used to convert absolute-value NMR line shapes into absorption-like line shapes, albeit with a heavy sensitivity penalty [14]. Formulas analogous to Eq. (10) exist for 2D, 3D, 4D, etc. cases [12,15], where the triangular weighting of the data points is independently applied in each dimension, giving a highly-peaked direct product weighting in the matrix elements between nD Fourier type basis functions [2]. Clearly, the nD midpoint of all the data may not be the best choice to emphasize unless the data is short in all dimensions, in which case the weighting is less severe and has not decayed much.

Careful experimental study of this noise sensitivity, in conjunction with a more formal theoretical analysis of some simpler cases [16] has led us to the conclusion that it is nearly always a problem whenever multidimensional data sets are sampled, in more than one dimension, to times of the order of the inverse line width, i.e. for data sets with more than one dimension with decent digital resolution. Luckily, early FT NMR papers suggest a work around. Natural abundance ^{13}C NMR suffers from reduced sensitivity but improved resolution compared with ^{1}H. Thus, since the earliest days of FT NMR, a matched exponentially decaying function has been routinely applied to ^{13}C FIDs before Fourier transformation to optimize sensitivity at the expense of resolution. The same could be done to the raw FIDs submitted to FDM analysis, if the resultant line broadening were acceptable. However, the idea behind FDM is to improve the resolution, and broadening the lines defeats its *raison d'être* as broad lines are handled perfectly in FT spectra. A better strategy therefore is to weight the *basis* rather than the *data*,

$$|\chi(f,\gamma)) = \sum_{n=0}^{M-1} e^{-n\tau\gamma} e^{-2i\pi n\tau f}|\Phi_n). \tag{11}$$

Weighting the basis will not change final computed line widths appreciably, as the actual underlying data remains unchanged. Indeed, in a hypothetical single-window calculation encompassing the entire spectral width, the FDM result, in exact arithmetic, should be *invariant* to any rank-preserving linear transformation of the basis $\{\Phi_n\}$. Linear combinations of the basis functions are used by the diagonalization routine to obtain the eigenfunctions, and hence forming other linear combinations up front has no effect. By contrast, in a limited window the basis is incomplete, and incorporating different weighting can improve the final results. What basis weighting can do, in particular, is to partially compensate for the potential noise amplification discussed above. While it is not possible to achieve equal representation of each and every data point in the resultant matrix elements, it *is* possible to compensate for the very large emphasis of the central part of the data, the amount of compensation being chosen in accord with the intrinsic SNR of

the data set. An exponential weighting of the Φ_n can be accomplished with the least change to the efficient formulas Eqs. (9) and (10), making it a good first choice. Including a small imaginary part to the frequency, $2\pi f \rightarrow 2\pi f - i\gamma$ gives Eq. (11), accomplishing the weighting. Note that the sensitivity of the computed spectrum to errors in the input data varies non-uniformly in related methods like linear prediction, and may also become an issue in other non-linear methods that try to circumvent the time-frequency uncertainty principle. We will explore these issues in a more comprehensive investigation of noise sensitivity in future.

For now, the treatment in Eq. (11) was applied to each dimension of a 2D experiment, and the value of γ chosen, for example, so that the midpoint of the data was weighted *equally* to the first point. This seems to be a reasonable choice for many data sets we examined, especially when the time dimension in question was sampled so that appreciable decay of the signal occurred. In a very short dimension, in which the SNR of each increment was similar, the weighting need not be this severe. The effect of γ is shown in Fig. 1. The weighting deemphasized the noisier portions, and thus suggesting the name *sensitivity-enhanced basis*. Note, however, that the actual profile still emphasizes data points beyond the first one, because the result of the double sum is more complex than a simple weighting of an FID.

2.2. Variable regularization for multidimensional spectra

For *n*D NMR spectra the naive application of FDM using an *n*D Fourier basis led to large noise-like spikes and other irregularities and artifacts [6]. By modifying the generalized eigenvalue problem

$$\mathbf{U}_l^{(1)}\mathbf{B}_{\mathbf{lk}} = u_{lk}\mathbf{U}^{(0)}\mathbf{B}_{\mathbf{lk}} \tag{12}$$

to

$$\mathbf{U}^{(0)\dagger}\mathbf{U}_l^{(1)}\mathbf{B}_{\mathbf{lk}} = u_{lk}\{\mathbf{U}^{(0)\dagger}\mathbf{U}^{(0)} + q^2\}\mathbf{B}_{\mathbf{lk}},\ l = 1,2\ldots \tag{13}$$

with the parameter q^2 assuring that the right-hand matrix is non-singular, far better results were obtained [6]. Regularization is a blunt instrument and the minimum amount that avoids artifactual peaks is to be preferred. Its effect on isolated features is primarily to broaden small peaks in a non-linear way while leaving the peak integral nearly constant. In more complex spectra, strong regularization can lead to nearby peaks coalescing into simpler features and hence significant potential loss of information. Whenever q^2 is on the order of the peak integral, the peak is smoothed into a broad baseline feature. A proper adjustment of q^2 resulted in FDM spectra with negligible artifacts while stronger peaks were widened to some extent. By processing a "blank" spectral region and ensuring that no large artifacts were produced in the spectral estimate, or, more conservatively, that all eigenvalues lay within the unit circle in the complex plane, an approximate value of the regularization parameter could be obtained.

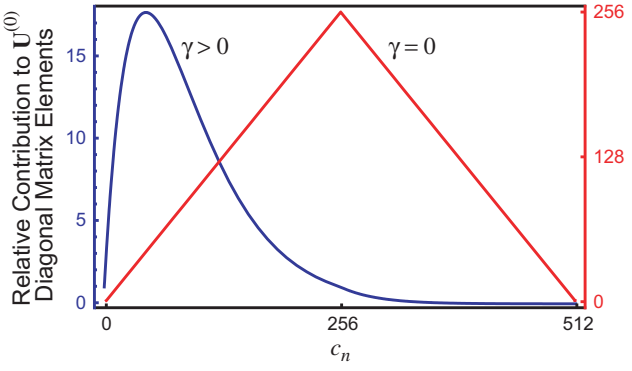

Fig. 1. A plot comparing the relative contribution of data point c_n to the diagonal matrix elements of the $\mathbf{U}^{(0)}$ matrix in the case of 1D time signal with 512 data points. The right hand scale refers to unmodified Fourier basis, and the left hand scale refers to a weighted basis in which γ has been chosen to give equal weighting to the midpoint relative to the first point.

Note that the right-hand matrix in Eq. (13) is independent of the particular dimension l under analysis. To the extent that regularization was meant only to correct defects in $\mathbf{U}^{(0)}$, the amount of regularization needed would not seem to depend on the dimension l, and in all published FDM spectra q^2 was independent of l. However, this was an oversimplification, and it seems that each dimension benefits from individual treatment, *i.e.*

$$\mathbf{U}^{(0)\dagger}\mathbf{U}_l^{(1)\mathbf{B}_{\mathbf{lk}}} = u_{lk}\{\mathbf{U}^{(0)\dagger}U^{(0)} + q_l^2\}\mathbf{B}_{\mathbf{lk}},\quad l = 1,2\ldots \tag{14}$$

This subtlety emerged when considering highly asymmetric 2D data sets with few points in t_1 but many points in t_2. The sensitivity-enhanced basis, by mitigating the effect of noise, allowed less total regularization to be employed and in turn led to the discovery that each dimension should be regularized independently for best performance. The new method maintained high resolution in each dimension without any obvious trade-off as more data was gathered. The prior implementations of FDM did not perform in this way: the most stable results for moderately noisy data were obtained often when fine multiplet structure in the long dimension was collapsed completely by rather aggressive regularization, then allowing significant line-narrowing in the indirect dimension. The sensitivity-enhanced basis better preserves fine multiplet structure (see Results section). While the extent of regularization was dimension dependent, it should not be window dependent. The FDM spectral estimate, obtained by a series of frequency windows with 50% overlap and smoothly summed [8], meant that broadening lines differently in two adjacent windows could lead to a confusing superposition of wide and narrow peaks in the final spectral estimate, and was avoided.

Dimension-dependent regularization could be time consuming without a reasonable guess for q_l. A useful general method, proposed by Hoch and Stern [17], namely an *in situ* approach was useful here: two (or more) peaks were artificially injected into an "empty" region of the spectrum, each with similar intensity, width, etc. to typical multiplet components. The results of the method could then be compared with the known input to ascertain any systematic distortion of the data, the absolute accuracy of peak positions, and sensitivity to SNR. It may also indicate what peak separation would likely be resolved, the smallest detectable peak, and the optimum range of q_l. In principle, appropriate software can always undertake this kind of analysis on its own, providing a detailed summary, and freeing the operator from any decisions regarding unknown, and potentially confusing, parameters of the method.

3. Results and discussion

The sensitivity-enhanced basis was first tested using model noisy data with known peak parameters, and next by a number of 2D experiments. Although it had been customary to convert the purely Lorentzian FDM line shape to an equivalent Gaussian line shape, no such line shape manipulation was used in this study, in order to isolate the difference between weighted and unweighted bases more clearly. The parameter γ was chosen such that the relative weighting of the first point and middle points of the signal along the diagonal elements of the $\mathbf{U}^{(0)}$ matrix were equal, i.e.

$$e^{-\gamma_2(M-1)} = \frac{1}{(M-1)} \tag{15}$$

in F_2. In F_1 there was less signal decay, and hence a less severe value γ_1 was used as determined by the *in situ* analysis.

We also investigated identical and dimension-dependent regularization. The real part of the FDM spectral representations were created using an aggressive representation of the infinite DFT of the signal obtained analytically from the extracted 2D FDM parameters [12],

$$S(f_1,f_2) = \tau_1\tau_2\left\{\sum_{k_1k_2} Re\{d_{k_1k_2}\}Re\left(\frac{1}{(1-u_{k_1}e^{-2i\pi f_1\tau_1})}-\frac{1}{2}\right)\right.$$
$$\left.\times Re\left(\frac{1}{(1-u_{k_2}e^{-2i\pi f_2\tau_2})}-\frac{1}{2}\right)\right\} \quad (16)$$

Eq. (16) artificially "phases" the spectrum to absorption, so that the absolute phase correction is required beforehand. It has some deficiencies, but is always a well-behaved function, whether $|u_{k_l}| > 1$ or not, although the apparent algebraic sign of the peak switches with the sign of the line width of the peak [18]. Regularization was chosen to bring all eigenvalues within the unit circle, for physical (positive) line widths. This simplified determining the algebraic sign of the peak, which in some very important practical applications may be either positive or negative [19]. Furthermore, only eigenvalues with real frequencies within the chosen 2D frequency window were used to create the spectral representation. Eigenvalues with frequency coordinates well outside the window are often inaccurate, and do not much influence the spectral estimate itself, as they mostly constitute fictitious features that simply model broad baseline roll within the window, or the bleed-in of the skirts of a genuine large peak outside the window. Thus, they might require a higher degree of regularization, degrading resolution for genuine peaks within the window. The correct way to handle baseline roll, wide lines of the order of the FDM frequency window, and out-of-phase peaks will be deferred to a future study. For now, we note that Eq. (16) has been used previously with good results, proving adequate as long as the 2D (FDM) spectra were sufficiently well resolved. For highly congested regions, in which individual 2D peaks may be badly out of phase, Eq. (16) may be inadequate, although such highly congested regions resisted analysis whether they were accurately represented in the FDM spectral estimate or not.

Preliminary values for regularization parameters for experimental signals were obtained by the aforementioned *in situ* analysis. Once determined, the same absolute regularization was used for each window, all of which were of identical size and contained the same number of 2D basis functions. The amount of regularization was conveniently characterized as a percentage of the norm of $\mathbf{U}^{(\mathbf{0})}$,

$$\|\mathbf{U}^{(\mathbf{0})}\| = Tr\left\{\mathbf{U}^{(\mathbf{0})\dagger}\mathbf{U}^{(\mathbf{0})}\right\} \quad (17)$$

to be able to compare the weighted and unweighted results objectively.

3.1. Accuracy of FDM parameters

The FDM results of a single window calculation over the entire Nyquist range using all the data (in cases where the data set was small enough to allow this approach) were invariant to any sensible choice of γ. The case of interest was thus a small window, the

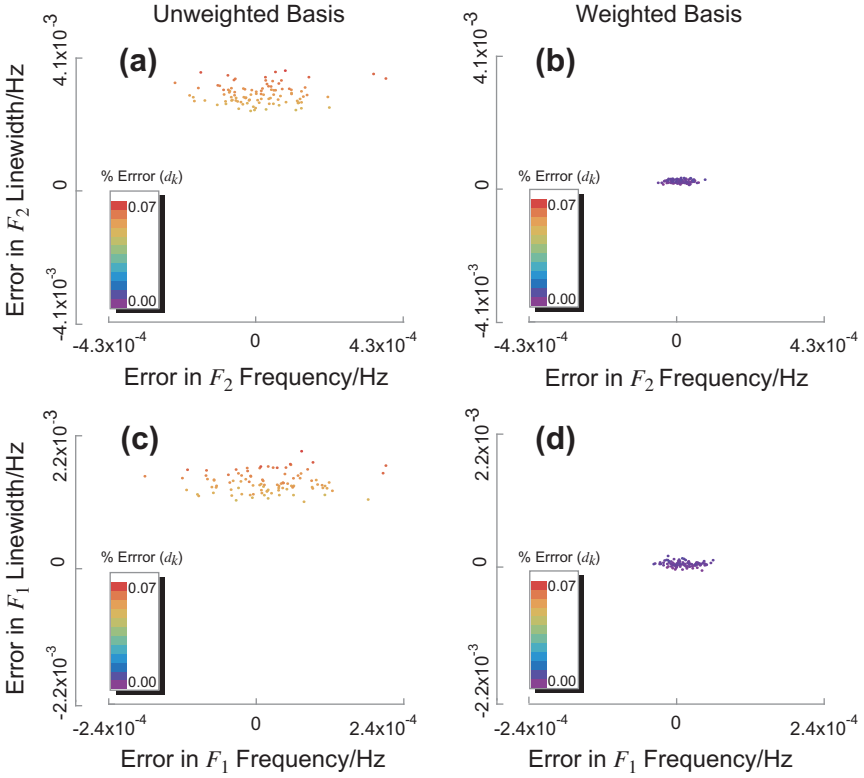

Fig. 2. Monte-Carlo error analysis of a 2D model time signal 512 × 512 in the time domain. Error in F_2 frequency and line width (a) in the unweighted basis and (b) in the weighted basis. Error in F_1 frequency and line width (c) in the unweighted basis and (d) in the weighted basis. A 1 kHz spectral width in each dimension gave a digital resolution of around 2 Hz. There was a single peak at F_2 = 123.7 Hz F_1 = 321.7 Hz, and with full widths at half maximum of 2 Hz (F_1) and 4 Hz (F_2); the line shape was Lorentzian in each dimension. The model 2D data set was purely phase modulated. FDM using a single 2D basis function in the unweighted (γ = 0) and weighted (γ = 0.02175) case was used on 100 different realizations of added Gaussian noise. The abscissae show the error in peak position in the respective dimension, and the ordinates display the error in line width. The colored legend indicates the percentage error in the calculated amplitude $|d_k|$ for the peak, ignoring phase. Results from the (unweighted) Fourier basis (a) and (c) showed significantly larger scatter than weighted (*sensitivity-enhanced*) Fourier basis results (b) and (d). The accuracy of all determined parameters, especially the calculated amplitude $|d_k|$, was noticeably improved using the weighted Fourier basis.

smallest possible "window" consisting of just a single basis function. As the 2D case was of interest, we concentrated on calculations to detect a single 2D peak with known frequency coordinates and line widths in each dimension, and a known amplitude (see Electronic Supplementary Information). To this model FID, random Gaussian noise was added and then an FDM calculation performed using a single 2D basis function, with frequencies chosen to coincide with the actual peak position. One hundred 2D model signals with different random noise realizations were synthesized and analyzed. The results (Fig. 2) indicate that notable improvement was achieved in the accuracy of FDM by introducing the sensitivity-enhanced basis, in this extreme case. In passing, we note that the selection of a limited frequency window is, in itself, a kind of regularization, as it limits the number of peaks that can interact, and allows the sensitivity-enhanced basis to have a larger effect. Selecting a window is thus numerically sound, as well as being numerically efficient.

3.2. Two-dimensional experiments

A simple 2D ^{13}C–^{1}H HSQC spectrum provided a suitable test case. Using a large enough spectral width in F_2 gave plenty of 2D baseline containing essentially only noise. These blank regions were best-case scenario, as actual 2D spectra show so-called "t_1-noise" that results from longer term spectrometer/magnet instability and is proportional to the intensity of the signal itself at any given F_2 frequency. Thus, actual regularization in F_1 may need to be increased somewhat from the preliminary value given by the *in situ* analysis, until $|u_{lk}| < 1$ for all the peaks.

Variations of FDM with different basis weighting and regularization were tested on the ^{13}C–^{1}H HSQC spectrum of 5 mg of phenanthridinone, a molecule studied previously in this laboratory [20] in 750 μl DMSO-d_6 at 500 MHz using a standard Varian triple resonance HCN triple resonance proton-observe probe with tri-axial pulsed field gradients. The pulse sequence employed was previously described [21] and compensates for both B_1 inhomogeneity and resonance offset effects of imperfect pulses in both dimensions. As each carbon atom was bonded to no more than one hydrogen atom, a gain in sensitivity was possible in analogy to the more common case of ^{15}N–^{1}H HSQC [22] in proteins. The data set was heavily truncated in t_1 (8 points) such that FDM should greatly improve resolution. All the data in t_2 (2048 points) were used. The conventional absorption-mode 2D FT spectrum is shown in Fig.3. The very short 8-point interferograms required strong apodization in t_1, and the contours were set just high enough to miss the residual sinc-function "wiggles" occurring around each multiplet, a possibility in this case as the 2D peaks have similar intensity.

Using FDM with the unweighted Fourier basis, we probed the effect of identical regularization in both dimensions, choosing the larger value of q_1^2 for the F_2 dimension as well. Using such strong symmetric regularization immediately led to a stable spectrum in which the resolution in the short F_1 dimension was much better than that in the FT spectrum, but in which the resolution in the long F_2 dimension was *worse*, (Fig. 4). The same symmetric regularization, but with the sensitivity-enhanced basis, resulted in improved resolution (see Electronic Supplementary Information), but was insufficient in the direct dimension to clearly resolve the

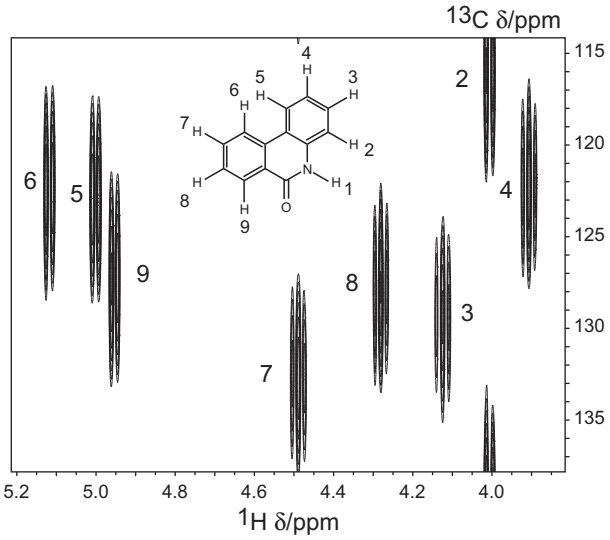

Fig. 3. An absorption mode 2D FT spectrum of 6(5 H)-phenanthridinone with the structure and assignment indicated. A pair of 8 × 2048 N-and P-type data sets were acquired with spectral widths of 3.0 kHz and 4.5 kHz in F_1 and F_2, respectively, of which a selected region is displayed. Cosine-squared apodization in each time dimension was used prior to Fourier transformation, to attenuate truncation artifacts. Although the natural line shape was Lorentzian in both dimensions, the apodization and sinc-function character combined with contour levels high enough to avoid baseline undulations gave each peak more elliptical 2D contours, as opposed to the star-shaped contours expected for a 2D absorption-mode Lorentzian peak. The raw digital resolution was 375 Hz in F_1 and 2.2 Hz in F_2. The data set was zero-filled to give a 256 × 4096 frequency-domain spectrum.

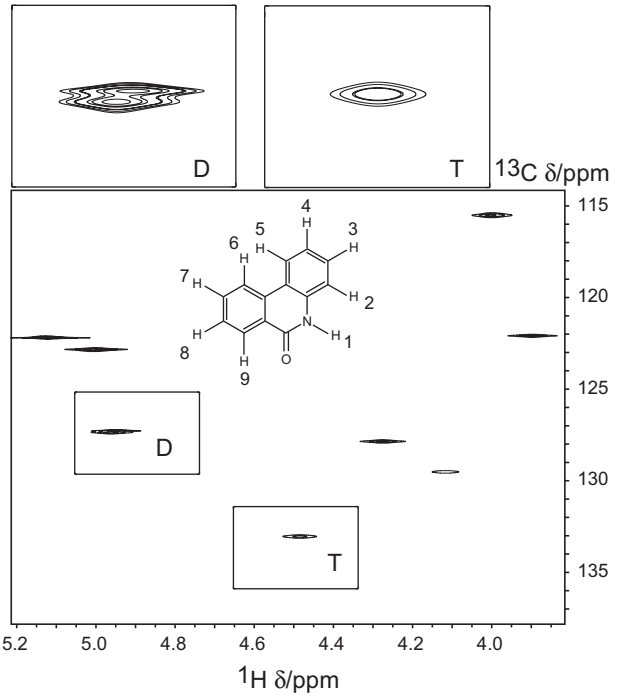

Fig. 4. The FDM spectral estimate using the Fourier basis and symmetric regularization with the value of q^2 increased until all eigenvalues within the frequency window in each dimension satisfied $|u_k| < 1$. With reference to Eq. (17) the regularization was 0.259%. Each window employed 63 × 4 2D basis functions over a frequency range of 3 kHz (the entire range) in F_1 and 250 Hz in F_2. Windows were overlapped by 50% and the spectral estimates coadded after weighting by a cos^2 function going to zero at the edges of the window. The contours in this figure and those that follow were chosen by introducing a synthetic *in situ* singlet of equal intensity to a representative true peak in the spectrum, submitting the synthetic peak to analysis in the same window shown, determining the maximum intensity of the synthetic peak, and then setting logarithmic contours at $1/2^k$ fractions, $k = 1,2,\ldots,6$ of this intensity. This method assures one that excessive broadening does not result in higher relative contour levels giving an artificially good appearance. No smoothing or any other line shape modification was used in any of these experimental contour plots. The zoomed regions show a triplet (which has collapsed to a featureless singlet, and a doublet, which was barely resolved but which exhibits slight frequency shifts in F_1. Although the doublet was resolved, the frequency splitting projected onto the F_2 axis was inaccurate.

couplings, even though these couplings were evident in the conventional 2D FT spectrum (Fig. 3). The problem was that the very short indirect dimension, in which there was little actual signal decay, and hence a high probability for noise to shift eigenvalues outside the unit circle, and in which there was extra t_1-noise from instrumental instabilities, required a value of q_1^2 that, when also applied to the direct dimension (in which the signal had decayed appreciably), resulted in too much peak broadening.

Next, the effects of dimension-dependent regularization were probed. The unweighted basis was able to reveal general features of the multiplets when individual regularization parameters were used for F_1 and F_2. The resolution was greatly improved in F_1 compared to the FT spectrum, and there was no loss of resolution in F_2. However this came with a price in stability and absolute accuracy of the peak positions (see Electronic Supplementary Information). The unweighted basis displayed distortion from Lorentzian line shape by the instability in the phases of peaks interacting with the formula Eq. (16) and a perceptible frequency shift between the peaks of individual multiplets which should be exactly degenerate in F_1.

The best results were obtained when the dimension-dependent regularization was combined with sensitivity enhanced basis. While the resolution in F_1 was superior to the FT result, the peaks have degenerate frequencies in F_1 for all the multiplets (Fig. 5). It was also remarkable that the sensitivity-enhanced basis turned up a pair of small cross peaks between two proton multiplets with a frequency separation in F_2 close to $^1J_{CH}/2$ for these aromatic ring protons. It is known that this condition, in which just one of the two carbon-13 satellites is strongly coupled to a vicinal neighbor proton, leads to parasitic magnetization transfer by the mixing of the spin eigenstates under the application of a 180° pulse [23]. This is one case in which the improved FDM spectral estimate revealed an unexpected feature which was stubbornly stable because, although not expected in the weak coupling limit, it was genuinely present in the data. The weak anti-phase doublets are designated with asterisks in the figure.

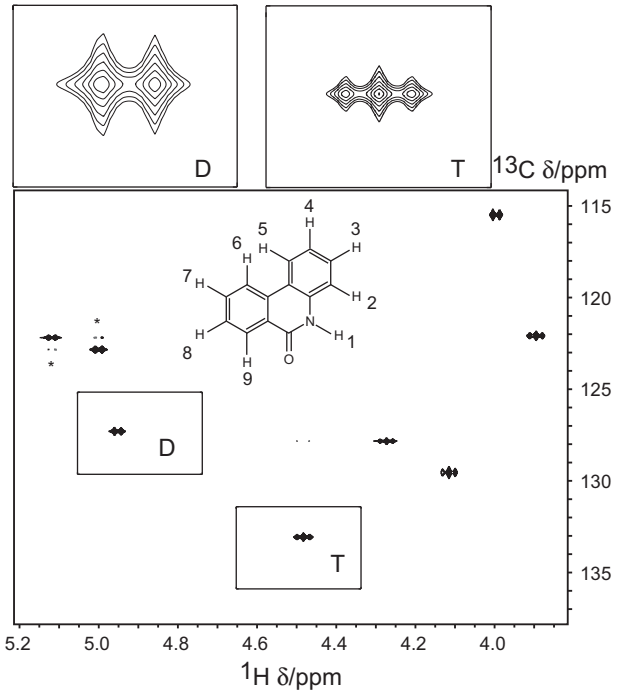

Fig. 5. The same data as in Fig. 4 but using sensitivity enhanced basis and differential regularization. The regularization was 4.21×10^{-1}% in F_1 and 7.02×10^{-7}% in F_2. The resolution in F_2 was greatly improved and the 2D frequency coordinates with the sensitivity-enhanced basis were more accurate, with excellent frequency registration in F_1. In addition, a weak pair of anti-phase cross peaks (top left) indicated with asterisks became observable. Previously, they were regularized into the baseline. These strong coupling artifacts are rarely seen in HSQC spectra, but are to be expected whenever two carbon-bearing protons are coupled to each other and separated by a frequency difference of about 1/2 $^1J_{CH}$ in chemical shift at the magnetic field employed in the experiment.

4. Conclusion

Although FDM improved resolution greatly under the right conditions, noise had the effect of introducing a host of "peaks" some of which refer to negative line widths, i.e. increasing exponential functions of time. Regularization to bring all eigenvalues within the unit circle in the complex plane flattened and smoothed the noise, and widened true peaks somewhat, causing loss of resolution and information. However, by adjusting the Fourier basis to what we have termed a sensitivity-enhanced basis, less noise was included in the matrix elements of the **U** matrices used to obtain the FDM spectral estimate. As a consequence, less regularization was needed to obtain stable, reliable spectra and it became beneficial to optimize the regularization parameter q^2 individually for each dimension. After the application of these two key changes to the FDM algorithm, improved accuracy and resolution resulted. For simplicity, the case of a simple exponential weighting of the basis was demonstrated, using a plausible choice of the parameter γ to attenuate the contribution of noisier data points to the FDM matrix elements. However, the universe of functions is essentially as large as the choice of weighting and apodization functions in conventional FT spectroscopy. Other possibilities will be explored in a more thorough analysis of all aspects of FDM as it is applied to genuine and highly imperfect NMR data sets.

Acknowledgments

This work was supported by a grant from the National Science Foundation, CHE-0703182. The authors continue to be indebted to Prof. V.A. Mandelshtam for insightful comments about various aspects of FDM.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2010.07.019.

References

[1] 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. 1. Theory and application to a quantum-dynamics model, J. Chem. Phys. 102 (20) (1995) 8011–8022.
[2] V.A. Mandelshtam, H.S. Taylor, Harmonic inversion of time signals and its applications, J. Chem. Phys. 107 (17) (1997) 6756–6769.
[3] J.H. Chen, V.A. Mandelshtam, Multiscale filter diagonalization method for spectral analysis of noisy data with nonlocalized features, J. Chem. Phys. 112 (10) (2000) 4429–4437.
[4] 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 (1) (1998) 76–87.
[5] 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 (2) (2000) 357–366.
[6] J.H. Chen, V.A. Mandelshtam, A.J. Shaka, Regularization of the two-dimensional filter diagonalization method: FDM2K, J. Magn. Reson. 146 (2) (2000) 363–368.
[7] J.H. Chen, A.A. De Angelis, V.A. Mandelshtam, A.J. Shaka, Progress on the two-dimensional filter diagonalization method. An efficient doubling scheme for two-dimensional constant-time NMR, J. Magn. Reson. 162 (1) (2003) 74–89.
[8] J.H. Chen, D. Nietlispach, A.J. Shaka, V.A. Mandelshtam, Ultra-high resolution 3D NMR spectra from limited-size data sets, J. Magn. Reson. 169 (2) (2004) 215–224.

[9] G.S. Armstrong, K.E. Cano, V.A. Mandelshtam, A.J. Shaka, B. Bendiak, Rapid 3D NMR using the filter diagonalization method: application to oligosaccharides derivatized with C-13-labeled acetyl groups, J. Magn. Reson. 170 (1) (2004) 156–163.
[10] G.S. Armstrong, V.A. Mandelshtam, A.J. Shaka, B. Bendiak, Rapid high-resolution four-dimensional NMR spectroscopy using the filter diagonalization method and its advantages for detailed structural elucidation of oligosaccharides, J. Magn. Reson. 173 (1) (2005) 160–168.
[11] M.H. Levitt, G. Bodenhausen, R.R. Ernst, Sensitivity of two-dimensional spectra, J. Magn. Reson. 58 (3) (1984) 462–472.
[12] V.A. Mandelshtam, FDM: the filter diagonalization method for data processing in NMR experiments, Prog. Nucl. Mag. Res. Spectrosc. 38 (2) (2001) 159–196.
[13] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Prog. Nucl. Mag. Res. Spectrosc. 34 (2) (1999) 93–158.
[14] A. Bax, R. Freeman, G.A. Morris, A simple method for suppressing dispersion-mode contributions in NMR-spectra – the pseudo echo, J. Magn. Reson. 43 (2) (1981) 333–338.
[15] V.A. Mandelshtam, The multidimensional filter diagonalization method -I. Theory and numerical implementation, J. Magn. Reson. 144 (2) (2000) 343–356.
[16] H. Celik, A.J. Shaka, V.A. Mandelshtam, Sensitivity analysis of solutions of the harmonic inversion problem: are all data points created equal?, J Magn. Reson. (2010) 2–8. doi: 10.1016-j.jmr210.06.015.
[17] J.C. Hoch, A.S. Stern, NMR Data Processing, Wiley, Liss, 2005.
[18] C.D. Ridge, A.J. Shaka, "Ersatz" and "hybrid" NMR spectral estimates using the filter diagonalization method, J. Phys. Chem. A 113 (10) (2009) 2036–2052.
[19] M. Wittekind, L. Mueller, HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonance in proteins, J. Magn. Reson. Ser. B 101 (1993) 201–205.
[20] K.E. Cano, M.J. Thrippleton, J. Keeler, A.J. Shaka, Cascaded z-filters for efficient single-scan suppression of zero-quantum coherence, J. Magn. Reson. 167 (2) (2004) 291–297.
[21] S. Keppetipola, W. Kudlicki, B.D. Nguyen, X. Meng, K.J. Donovan, A.J. Shaka, From gene to HSQC in under five hours: high-throughput NMR proteomics, J. Am. Chem. Soc. 128 (14) (2006) 4508–4509.
[22] L. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, J. Am. Chem. Soc. 114 (1992) 10663–10665.
[23] J.H. Keeler, Oxford University Part 2 Thesis, Oxford Bodelian Library, 1981.