Journal of Magnetic Resonance **146,** 363–368 (2000)
doi:10.1006/jmre.2000.2155, available online at http://www.idealibrary.com on IDEAL®

Regularization of the Two-Dimensional Filter Diagonalization Method: FDM2K

Jianhan Chen, Vladimir A. Mandelshtam, and A. J. Shaka[1]

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

Received March 22, 2000; revised June 26, 2000

We outline an important advance in the problem of obtaining a two-dimensional (2D) line list of the most prominent features in a 2D high-resolution NMR spectrum in the presence of noise, when using the Filter Diagonalization Method (FDM) to sidestep limitations of conventional FFT processing. Although respectable absorption-mode spectra have been obtained previously by the artifice of "averaging" several FDM calculations, no 2D line list could be directly obtained from the averaged spectrum, and each calculation produced numerical artifacts that were demonstrably inconsistent with the measured data, but which could not be removed *a posteriori.* **By regularizing the intrinsically ill-defined generalized eigenvalue problem that FDM poses, in a particular quite plausible way, features that are weak or stem from numerical problems are attenuated, allowing better characterization of the dominant spectral features. We call the new algorithm FDM2K.**

Key Words: **filter diagonalization method; FDM; FDM2K; multidimensional NMR; regularization; generalized eigenvalue problem; linear algebra; protein NMR; rubredoxin.**

We have previously described how the Filter Diagonalization Method (FDM) (*1, 2*) can be generalized to the multidimensional time signals that arise in NMR experiments (*3–6*). Good quality two-dimensional (2D) spectra could be obtained, but it was essential to "average" several calculations in order to obtain them. It therefore seemed that n-dimensional (nD) time signals ($n > 1$) were rather different than 1D signals, where averaging was not typically required. Apparently, *regularization* of multi-dimensional FDM is a key issue (see, for example, Golub and van Loan (*7*) where some aspects of regularization are discussed for "incorrectly posed" problems). It turns out that a slight modification of the previously published algorithm allows us to produce regularized nD spectral parameters in a single calculation, making a direct nD line list a distinct possibility. We call the modified algorithm FDM2K, in honor of the new millennium. To establish notation, and describe the previous difficulties, the important points are recapped below.

In FDM we postulate that the digitized measured complex-valued signal c_n that we would like to fit as a sum of damped sinusoids,

$$c_n = \sum_{k=1}^{K} d_k \exp(-in\omega_k\tau), \qquad [1]$$

can be represented as the time autocorrelation function of a fictitious dynamical system with non-Hermitian but symmetric Hamiltonian $\hat{\Omega}$,

$$c_n = (\Phi_0|e^{-in\tau\hat{\Omega}}\Phi_0) \equiv (\Phi_0|\hat{U}^n\Phi_0), \qquad [2]$$

so that the highly nonlinear fitting problem of Eq. [1] is reduced to that of diagonalizing $\hat{U}$, the evolution operator over a single time step. Reduction to a linear algebra problem guarantees the existence and uniqueness of the solution, although the possible over- and/or under-determined nature of the parametric fitting problem, and the non-Hermiticity of the matrices involved, makes numerical stability a potential issue.

Neither the explicit form of the Hamiltonian $\hat{\Omega}$ nor the "initial state" Φ_0 need be known, as only a matrix representation in some basis is required for a numerical solution. The primitive basis is iteratively derived by letting $\hat{U}$ act on Φ_0,

$$\Phi_n = \hat{U}^n\Phi_0, \qquad [3]$$

so that the overlap matrix elements $\mathbf{U}^{(0)}_{nm}$ and the matrix elements $\mathbf{U}^{(1)}_{nm}$ of $\hat{U}$ are given by the measured data

$$\mathbf{U}^{(0)}_{nm} = (\Phi_n|\Phi_m) = c_{n+m};\ \mathbf{U}^{(1)}_{nm} = (\Phi_n|\hat{U}\Phi_m) = c_{n+m+1}. \qquad [4]$$

The extraction of the eigenvalues, $u_k = \exp(-i\omega_k\tau)$, that determine line position and width, and the eigenvectors, $\mathbf{B}_k$, that determine amplitude and phase, then proceeds by solving a generalized eigenvalue problem of the form

$$\mathbf{U}^{(1)}\mathbf{B}_k = u_k\mathbf{U}^{(0)}\mathbf{B}_k. \qquad [5]$$

[1] To whom correspondence should be addressed. E-mail: ajshaka@uci.edu.

1090-7807/00 $35.00

FDM would be very easy to understand if Eq. [5] could be directly applied. Unfortunately the huge size and notoriously ill-conditioned matrices that enter into Eq. [5] would make this straightforward approach both numerically expensive and very unreliable for any realistic data sets.

By changing to a Fourier basis (*1*, *2*)

$$\Psi_j = \sum_k \exp(-ik\varphi_j)\Phi_k \qquad [6]$$

and then *filtering* this basis over a small window of frequencies, the size of the **U** matrices can be controlled. The entire spectrum is effectively broken down into frequency windows, and Eq. [5] is solved for each window. The results are then concatenated to give the entire spectral representation. For noisy data with broad features of greater width than the window, a multi-scale basis may be used (*8*). While the Fourier basis results in a smaller matrix problem, it does not *necessarily* improve the condition number of $\mathbf{U}^{(0)}$ or $\mathbf{U}^{(1)}$, which may be nearly singular in some cases. For the 1D case $\mathbf{U}^{(0)}$ is diagonally dominant in the Fourier basis, but this is not so for higher dimensionality. By using the QZ algorithm (*9*) to solve the generalized eigenvalue problem, numerically exact eigenvalues and eigenvectors can be obtained even in these pathological cases. But the fundamental problem, namely that the parameters extracted from the data can become incredibly sensitive to small input perturbations of the original data, is not solved. In 2D and 3D NMR, $\mathbf{U}^{(0)}$ often has this bad structure. That problem is now addressed here.

What is needed is a way to regularize the problem. One could imagine solving the generalized eigenvalue problem by inverting $\mathbf{U}^{(0)}$, multiplying both sides by $[\mathbf{U}^{(0)}]^{-1}$ and then diagonalizing $[\mathbf{U}^{(0)}]^{-1}\mathbf{U}^{(1)}$. This is, in fact, a very poor way to tackle the problem, but points out that the difficulty may lie in the inversion. A sensible, tried-and-true way to proceed is to make a singular value decomposition (SVD) (*10*) of $\mathbf{U}^{(0)}$ (*3*) to obtain

$$\mathbf{U}^{(0)} = \mathbf{V}\Sigma\mathbf{W}^{\dagger}, \qquad [7]$$

where **V** and **W** are unitary and Σ is the diagonal matrix of singular values. One then hopes for the situation illustrated in Fig. 1a. There are a certain number of large singular values and then a break followed by much smaller values. Truncated SVD consists of choosing a suitable cutoff, so that only the significant features, each of which we hope (and pray) will correspond to a particular line, are retained. Numerically quite large but meaningless eigenvectors are also kept out of all the subsequent manipulations. These are precisely the eigenvectors

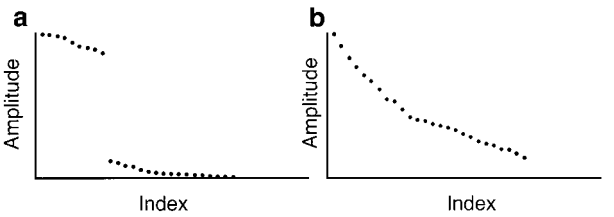

FIG. 1. Two possibilities for the distribution of singular values obtained from an SVD of the matrix $\mathbf{U}^{(0)}$. In case (a) there is a dramatic break in the distribution, signaling that the remaining space lies in the null space of $\mathbf{U}^{(0)}$ and so should be discarded before solving the generalized eigenvalue problem. This case is perfect for SVD. In case (b), which is more typical for measured NMR data, there is no clean break and so choosing a cutoff value becomes highly subjective.

mostly within the null space of $\mathbf{U}^{(0)}$ which if retained will, after the normalization

$$\mathbf{B}_k^T\mathbf{U}^{(0)}\mathbf{B}_k = 1, \qquad [8]$$

become disastrously large. The computation of the amplitude d_k then allows these vectors to amplify random noise to give a false amplitude d_k, while the eigenvalue u_k may or may not fall by chance within the specified frequency window and be retained. This squares with the computational observation of artifacts in 2DFDM that become invisible to the eye after sufficient averaging over slightly different input data.

The truncated SVD procedure is supposed to ameliorate these problems (*11*). After pruning the basis by SVD, diagonalization in the reduced basis then delivers the desired positions, widths, amplitudes, and phases (u_k and d_k) for the selected number of features. Unfortunately, we have found that typical NMR data has no clean break, and may display a singular value profile like Fig. 1b. Choosing any cutoff in this situation can be very risky, as genuine peaks may be thrown out in the truncated SVD regularization. Furthermore, imposing some desired number of features on the data gives the operator a tremendous chance to bias the experimental results to support a particular viewpoint and is dangerous. For these reasons, we eventually abandoned our nascent implementation (*3*) of SVD. However, keeping the entire space is also not good, as numerical artifacts are invariably produced in the spectrum. Because these artifacts are, by their nature, very sensitive to the input, it is enough to slightly change the latter, by increasing the data length slightly, or even by adding small pseudo-random numerical noise to the **U** matrices (*5*). Coaddition of all the spectra produced yields a far better result, which is now *insensitive* to the exact input. However, the artifacts are not completely eliminated; they are just reduced enough that they are not problematical. Furthermore, no direct *n*D line list is obtained because the total number of features

depends on the number of averaged spectra and generally exceeds that allowed by information considerations.

The regularization we propose here is much less drastic than truncated SVD in its action and is reminiscent of Tikhonov regularization (*12*) (also known as ridge regression in the statistics literature (*7*)) that is used in the solution of linear systems. In that case the linear system

$$\mathbf{Ax} = \mathbf{b} \qquad [9]$$

is replaced by the regularized system

$$(\mathbf{A}^\dagger\mathbf{A} + q^2)\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}, \qquad [10]$$

thereby avoiding potential problems revolving around the singularity of the matrix **A**. In fact, we have exploited exactly this trick to construct a regularized *spectrum* from the **U** matrices using only solutions of linear systems (*13*). Such a direct regularized transformation does not, however, deliver an explicit line list like FDM does. In order to do that, it appears that the generalized eigenvalue problem must be tackled, and the generalized eigenvalue problem does not, to the best of our knowledge, allow such a straightforward regularization as Eq. [10]. Instead, we proceeded by analogy to try to regularize Eq. [5]. There are two steps. First, multiply each side of Eq. [5] on the left by $[\mathbf{U}^{(0)}]^\dagger$. Second, add a diagonal matrix q^2 to the right-hand side, giving the FDM2K regularized formula

$$[\mathbf{U}^{(0)}]^\dagger\mathbf{U}^{(1)}\mathbf{B}_k = u_k\{[\mathbf{U}^{(0)}]^\dagger\mathbf{U}^{(0)} + q^2\}\mathbf{B}_k. \qquad [11]$$

The first step alone is bad in, e.g., least squares problems, as it *squares* the condition number of the resulting matrix (*10*). Luckily, we are rescued by q^2, which automatically imposes a minimum singular value. That is, the right-hand side is now a positive definite Hermitian matrix whose condition number is controlled by q^2. These two features reduce potential problems arising from the structure of $\mathbf{U}^{(0)}$, and Eq. [11] lies at the heart of the improved algorithm.

We have found Eq. [11] to be mostly irrelevant in 1D FDM, but quite important in 2D FDM calculations, so that it is here that FDM2K is an improvement. In 2D FDM we solve two generalized eigenvalue problems that correspond to shifting the data set by a single point along each of the time dimensions, e.g.,

$$\mathbf{U}^{(1)}\mathbf{B}_{1k} = u_{1k}\mathbf{U}^{(0)}\mathbf{B}_{1k};\ \mathbf{U}^{(2)}\mathbf{B}_{2k} = u_{2k}\mathbf{U}^{(0)}\mathbf{B}_{2k} \qquad [12]$$

which, in regularized form, become

$$[\mathbf{U}^{(0)}]^\dagger\mathbf{U}^{(1)}\mathbf{B}_{1k} = u_{1k}\{[\mathbf{U}^{(0)}]^\dagger\mathbf{U}^{(0)} + q_1^2\}\mathbf{B}_{1k} \qquad [13]$$

$$[\mathbf{U}^{(0)}]^\dagger\mathbf{U}^{(2)}\mathbf{B}_{2k} = u_{2k}\{[\mathbf{U}^{(0)}]^\dagger\mathbf{U}^{(0)} + q_2^2\}\mathbf{B}_{2k} \qquad [14]$$

and the spectral representation is formed as described in (*4, 5*) using the eigenvalues and eigenvectors obtained. The explicit expressions for all the matrix elements needed in Eqs. [13] and [14] have been derived previously (*4, 5*) and are not presented again here. The bottom line is that the naïve solution of Eq. [12] contains spurious entries, leading to a host of artifacts that are sensitive to any changes in the parameters of the calculation (*5, 6*) and that necessitate either averaging or regularization. We have typically used $q_1 = q_2 \equiv q$ to simplify things, although in principle they could differ. Note also that *any* matrix of the form $\mathbf{Q}^\dagger\mathbf{Q}$ could be used to regularize the equations, an aspect that remains unexplored at this time.

It may not be obvious what distortions regularization may cause, so a simple example is in order, one that was previously used to illustrate FDM (*14*). This example is only meant to give an idea of possible distortions and is not comprehensive. The system in question also obviously does *not* require any regularization. Consider a single line whose FID we have discretely sampled at $t = 0$ and $t = \tau$. Then, in the absence of noise, the two-point FID is

$$c_0 = d_k;\ c_1 = d_k\exp(-i\omega_k\tau) \equiv d_k\exp(-i2\pi f_k\tau)\exp(-\gamma_k\tau), \qquad [15]$$

where $\omega_k = 2\pi f_k - i\gamma_k$ is the complex frequency and Eq. [11] becomes a 1×1 matrix problem

$$c_0^* c_1 b = u_k(c_0^* c_0 + q^2)b \qquad [16]$$

which yields the eigenvalue

$$u_k = \left(\frac{|d_k|^2}{|d_k|^2 + q^2}\right)\exp(-i\omega_k\tau) \equiv \exp(-i(\omega_k - i\epsilon_k)\tau), \qquad [17]$$

where

$$\exp(-\tau\epsilon_k) = \left(\frac{|d_k|^2}{|d_k|^2 + q^2}\right) \qquad [18]$$

showing that the line frequency remains unchanged but the width γ_k has been increased by ϵ_k. In the limit that $q^2 \ll |d_k|^2$ we have

$$\tau\epsilon_k \approx \frac{q^2}{|d_k|^2}, \qquad [19]$$

so that regularization causes an insignificant increase in the linewidth, with no change at all in line position, phase, or integral. On the other hand, if $q^2 \approx |d_k|^2$, then there will be a catastrophic change in linewidth, because the second data point will effectively be attenuated and will so produce a peak that

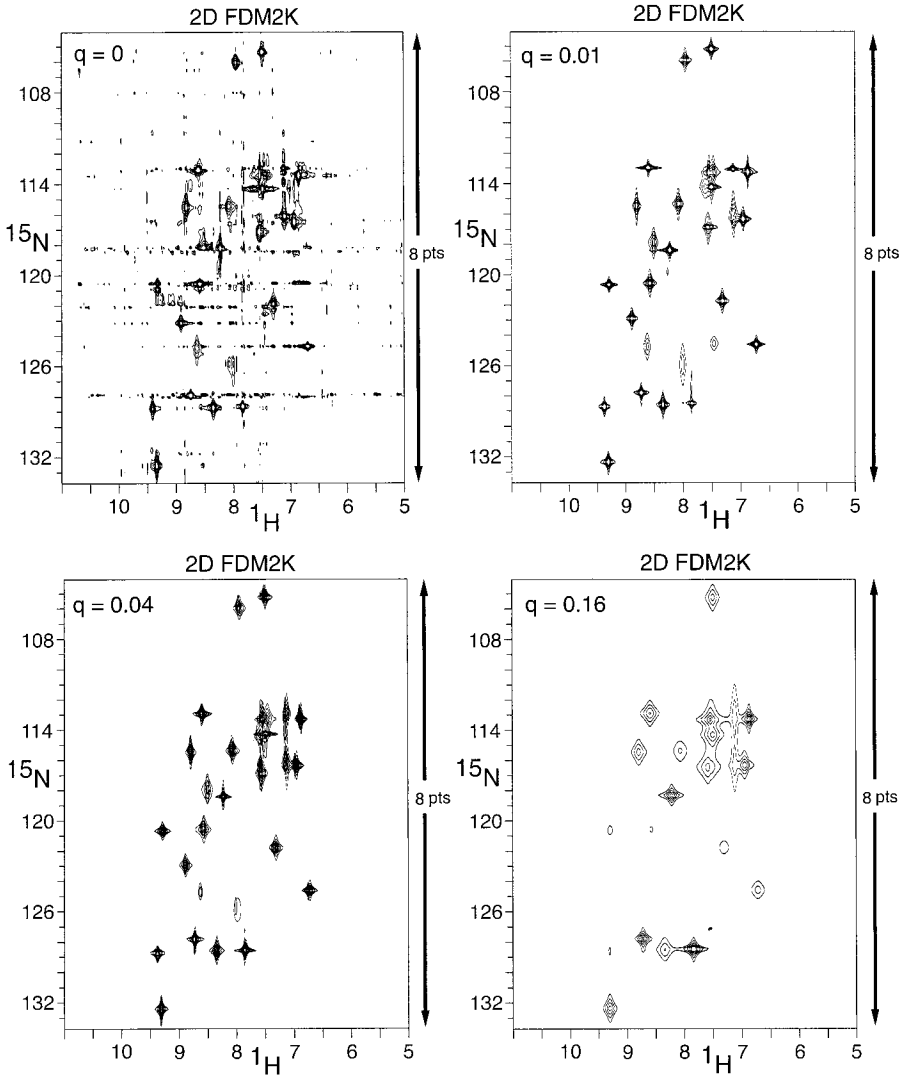

FIG. 2. ^{15}N–^{1}H 2D spectra of nitrogen-15 labeled (100%) rubredoxin, a 4.5 kDa metalloprotein, obtained in 90% H_2O/10% D_2O at 25°C. All the FDM spectra are obtained from an 8×300 phase-modulated data set, from which an absorption-mode spectrum can be obtained as described in (*4*). The top left panel shows the FDM spectrum obtained with $q = 0$, illustrating the complex nature of the artifacts produced in a single calculation. By scanning q through a range it is possible to obtain beautiful results that are not too sensitive to q and which require no assumption about the "true" number of peaks. The spectra have been aggressively contoured to reveal the weaker features.

appears only as a broad baseline feature. Note that it does not matter whether a feature is signal or noise: if its integral is small compared to q then it will be smoothed away, although the position and integral will remain intact. We thus see that the regularization corresponds to a nonlinear shift of the ω_k values along the negative imaginary axis in the complex plane. It also, however, improves the ability to detect and quantify the remaining features, as the resulting matrix problem becomes better conditioned. In this sense it is conceptually different from merely post-smoothing the spectrum by $\omega_k \rightarrow \omega_k - i\Gamma$, even if Γ were made feature-dependent. When more than one line is involved, however, one must be careful not to jump to conclusions, as the amplitude d_k only emerges *after* the generalized eigenvalue problem is solved, and the regularization changes this solution. It is thus quite possible both for more complex distortions to be introduced, and for *large* artifacts to be completely eliminated. In the 2D case, both sets of eigenvalues will be altered by the introduction of q^2.

Clearly, the FDM2K procedure regularizes the ill-conditioning caused by the over-determined problem when the number of "true" spectral lines is dominated by the rank of the matrices. This can be understood as follows. The parameter K in Eq. [1] is ill-defined once $K < (N/2)$ for an N-point complex FID, as any generic perturbation of c_n will require an increase in K. Thus, a more sensible approach is to merely assume $K = (N/2)$ and allow terms with small d_k to enter into Eq. [1].

Regularization then eliminates these terms in a controlled way according to Eq. [17].

It is extremely tempting to associate $(\mathbf{U}^{(0)})^{+}\mathbf{U}^{(0)}$ with the "signal + noise power" and q^2 with the "noise power" whereupon the regularization acquires some superficial aspects of the Maximum Entropy Method (*11*) of fitting. In the latter, small features, for which there is little evidence in the data, are discarded, and as smooth a spectrum as is compatible with the data is produced. For now we resist any concrete identification and treat q^2 simply as a parameter of the FDM2K algorithm. In real applications, we can scan q^2 from small to large values, and stop at any point that looks useful, much like an apodization or filter function can be optimized in FT spectra. Setting $q = 0$ in Eq. [11] reverts to conventional FDM, albeit with the caveat on the condition number mentioned earlier. As q is increased, a smoother, less featured spectrum generally emerges. In the limit of very large q an essentially flat spectrum with the correct integral should be produced.

Figure 2 shows four 2D ^{15}N–^{1}H shift correlation spectra obtained from a sample of nitrogen-15 labeled rubredoxin, a small metalloprotein (*15*). Only 8 increments have been used in the ^{15}N dimension, with 300 complex points along the ^{1}H dimension. The upper left panel shows the results obtained with $q = 0$, showing a forest of artifacts along with the true peaks, some of which are significantly distorted. The remaining panels show the effect of scanning q. There is a large range of stability, with smaller features eventually disappearing as q is increased. The spectra with optimum q are demonstrably superior to those obtained with averaging using FDM (*6*) and, as a bonus, FDM2K is 50–100 times faster, as only a single calculation is required, assuming that an optimal value of q is known. Each 2D spectral calculation, including construction of the highly digitized spectrum, took about a minute on a 533 MHz Dec Alpha workstation under the LINUX operating system. The code used has not been optimized for speed.

Figure 3 shows a vertical trace at the proton shift of 7.5 ppm from each of the 2D spectra. Even with only 8 points along this dimension, quite high resolution is achieved. (The FFT spectrum shows completely unusable resolution.) The large unstable artifacts with $q = 0$ are effectively attenuated by increasing q, with *true* peaks following suit as q becomes too large. Note that q^2 spans a range of 256 over the three spectra, so that it is not surprising that significant changes are observed in the results.

There are many interesting avenues to pursue at this point, and they are actively under intensive investigation. It is clear that FDM2K will not improve sensitivity. In fact we have shown quite clearly that the opposite is the case. It is therefore of interest to study the rich dynamics arising from the interplay of noise level and q^2 with the degree of truncation and the desired resolution. For example, one could independently vary the value of q for each of the spectral dimensions, as the number of features projected each way might be radically different. In protein NMR the ability to remove, selectively via the FDM2K line list, entire nearby peaks from other spectral planes in a 3D experiment will facilitate visual inspection of 2D planes and assignment by minimizing problems with interference from large peaks. Finally, the promise of a direct, intrinsically high-resolution numerical summary of the data that can be imported into an assignment/structure software suite is bound to be of interest. Compared with digging through a low-resolution FFT spectrum to pick peaks, FDM2K offers a rational controllable way to select the features to study without having any omniscient knowledge in advance. The impact of this new algorithm could thus be substantial for multidimensional NMR processing applications.

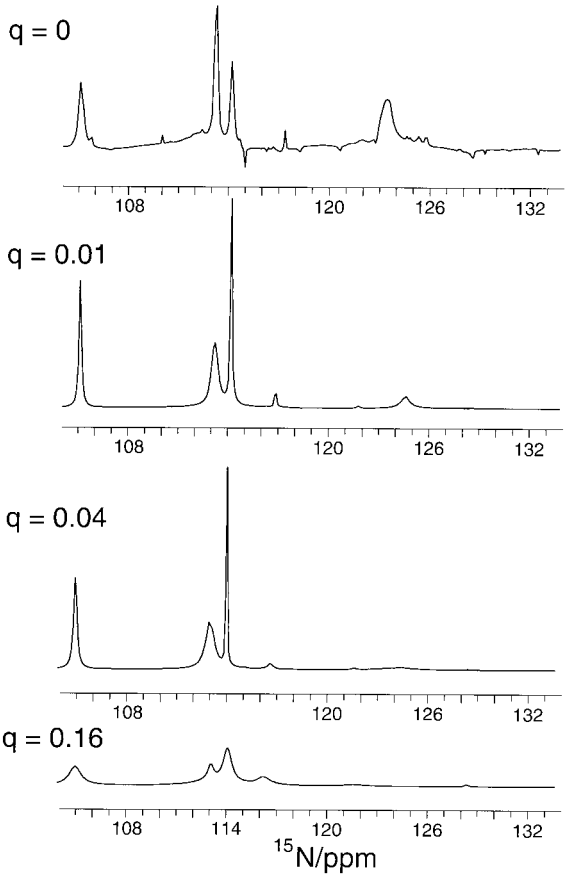

FIG. 3. A trace from each of the four 2D spectra in Fig. 2, at the proton shift of 7.5 ppm. The top spectrum illustrates that the unwanted peaks from a single FDM calculation need not be small, and can in fact dominate the true signals at some frequencies. The lower series show the progressive effect of the regularization, from reasonable to much too severe. An FFT spectrum of the 8-point interferogram at this frequency gives a very broad nearly useless spectrum (data not shown).

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, CHE-9900422 and CHE-9807229. The authors acknowledge useful conversations with Dr. J. C. Hoch and Dr. A. S. Stern at the ENC 2000.

REFERENCES

1. M. R. Wall and 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,** 8011–8022 (1995).
2. V. A. Mandelshtam and H. S. Taylor, Harmonic inversion of time signals and its applications, *J. Chem. Phys.* **107,** 6756–6769 (1997).
3. V. A. Mandelshtam, H. Hu, and A. J. Shaka, Two-dimensional HSQC NMR spectra obtained using a self-compensating double pulsed field gradient and processed using the filter diagonalization method, *Magn. Reson. Chem.* **36,** S17–S28 (1998).
4. V. A. Mandelshtam, H. Hu, N. D. Taylor, M. Smith, and A. J. Shaka, Highly resolved double absorption 2D NMR spectra from complex severely truncated 2D phase-modulated signals by filter-diagonalization-averaging method, *Chem. Phys. Lett.* **305,** 209–216 (1999).
5. V. A. Mandelshtam, The multidimensional filter diagonalization method. I. Theory and numerical implementation, *J. Magn. Reson.* **144,** 343–356 (2000).
6. H. Hu, A. A. De Angelis, V. A. Mandelshtam, and A. J. Shaka, The multidimensional filter diagonalization method. II. Application to 2D projections of 2D, 3D, and 4D NMR experiments, *J. Magn. Reson.* **144,** 357–366 (2000).
7. G. H. Golub and C. F. van Loan, "Matrix Computations," Johns Hopkins Univ. Press, Baltimore (1989).
8. J. Chen and V. A. Mandelshtam, Multi-scale filter diagonalization method for spectral analysis of noisy data with non-localized features, *J. Chem. Phys.* **112,** 4429–4437 (2000).
9. C. B. Moler and G. W. Stewart, *SIAM J. Numer. Anal.* **10,** 241–256 (1973).
10. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical recipes in C, Second Edition," Chap. 2. Cambridge Univ. Press, Cambridge, UK (1996).
11. J. C. Hoch and A. S. Stern, "NMR Data Processing," Chap. 4, 5, Wiley, New York (1996).
12. A. N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, *Soviet Math. Dokl.* **4,** 1035–1038 (1963).
13. J. Chen, A. J. Shaka, and V. A. Mandelshtam, RRT: The regularized resolvent transform for high-resolution spectral estimation, submitted for publication.
14. H. Hu, Q. N. Van, V. A. Mandelshtam, and A. J. Shaka, Reference deconvolution, phase correction and line listing of NMR spectra by the 1D filter diagonalization method, *J. Magn. Reson.* **134,** 76–87 (1998).
15. B. F. Volkman, A. M. Prantner, S. J. Wilkens, B. Xia, and J. L. Markley, Assignment of H-1, C-13, N-15 signals of oxidized Clostridium pasteurianum rubredoxin, *J. Biomol. NMR* **10,** 409–410 (1997).