

RRT: The Regularized Resolvent Transform for High-Resolution Spectral Estimation

Jianhan Chen, A. J. Shaka, and Vladimir A. Mandelshtam

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

E-mail: jianhanc@uci.edu; ajshaka@uci.edu; mandelsh@uci.edu

Received June 7, 2000; revised July 25, 2000

A new numerical expression, called the regularized resolvent transform (RRT), is presented. RRT is a direct transformation of the truncated time-domain data into a frequency-domain spectrum and is suitable for high-resolution spectral estimation of multidimensional time signals. One of its forms, under the condition that the signal consists only of a finite sum of damped sinusoids, turns out to be equivalent to the exact infinite time discrete Fourier transformation. RRT naturally emerges from the filter diagonalization method, although no diagonalization is required. In RRT the spectrum at each frequency s is expressed in terms of the resolvent $R(s)^{-1}$ of a small data matrix $R(s)$, that is constructed from the time signal. Generally, R is singular, which requires certain regularization. In particular, the Tikhonov regularization, $R^{-1} \approx [R^\dagger R + q^2]^{-1} R^\dagger$ with regularization parameter q , appears to be computationally both efficient and very stable. Numerical implementation of RRT is very inexpensive because even for extremely large data sets the matrices involved are small. RRT is demonstrated using model 1D and experimental 2D NMR signals.

© 2000 Academic Press

1. INTRODUCTION

There is a certain attraction to the idea of a *transform*, that converts data from one representation to another, more useful, representation. In the common case of Fourier transformation (FT) a continuous function of time $c(t)$ is transformed to a function of frequency $I(s)$, by performing, for instance, a Fourier integral. For each value of s and signal $c(t)$ we can obtain a complex number that gives the local amplitude $I(s)$. The simplicity and transparency of the transform make it appealing even when, as in the discrete FT (DFT), the digital spectrum may deviate appreciably from the true integral representation (I).

By contrast, parametric methods that rely on fitting the data to a functional form are rather more complex in nature. They typically rely on adjustable parameters to obtain satisfactory results and can be less user-friendly in operation. In this paper we bridge the parameter estimation approach to the transform approach by introducing a new transform, the regularized resolvent transform (RRT). The RRT approach converts

damped sinusoids in the time domain into Lorentzian lines in the frequency domain, but *without* any intermediate line list. It can be implemented efficiently and maintains the spirit of local spectral analysis that can be lost in the quest to obtain a giant parametric fit of an entire data set.

2. 1D RRT

In this section we consider a 1D spectral analysis problem which has been previously treated by a variety of methods and is well understood. However, the expressions derived here are generalized to the less well explored multidimensional case in the next section.

Given a discrete equidistant time signal $c(n\tau)$, our goal here is to estimate its infinite time DFT,

$$I(s) = \sum_{n=0}^{\infty} c(n\tau) e^{in\tau s} \left(1 - \frac{\delta_{n0}}{2}\right), \quad [1]$$

but using only a finite portion of the data with $n = 0, 1, \dots, N - 1$. The term $(1 - \delta_{n0}/2)$ in Eq. [1] multiplies $c(0)$ by $\frac{1}{2}$ to correct the error introduced by the discrete sum approximation of the continuous half-line Fourier integral.

The assumption used here to derive the linear algebraic expressions is due to Wall and Neuhauser (2) in which $c(n\tau)$ is associated with a time autocorrelation function of a fictitious quantum system for some fictitious initial state Φ_0 ,

$$c(n\tau) = (\Phi_0 | \hat{U}^n \Phi_0), \quad [2]$$

where we used the round brackets for the symmetric (not Hermitian) inner product, $(\Psi | \Phi) = (\Phi | \Psi)$. The effective evolution operator \hat{U} may be nonunitary but is assumed to be symmetric with respect to the inner product, i.e., $(\hat{U}\Psi | \Phi) = (\Psi | \hat{U}\Phi)$.

The assumption that the evolution operator \hat{U} has a finite rank K turns to be equivalent to assuming that the signal of

question can be represented as a sum of exactly K complex sinusoids (2),

$$c(n\tau) = \sum_{k=1}^K d_k e^{-in\tau\omega_k}, \quad [3]$$

with nondegenerate and generally complex frequencies ω_k and complex coefficients d_k . In Refs. (2, 3) the filter diagonalization method (FDM) was developed to compute those frequencies and coefficients by essentially diagonalizing \hat{U} in a special basis. It was, in particular, noted (3) that a signal with total size $N = 2K$ yields exactly K pairs (ω_k, d_k) . These can then be used to evaluate the spectrum $I(s)$. We do not recast the spectral estimation problem into one of calculating the spectral parameters (2, 3), not because the latter fails, but rather because the present approach leads to very elegant and numerically efficient direct signal transformation which may be advantageous in certain contexts. It is also interesting that a spectrum, very similar to that obtained with FDM, may be constructed without ever determining any spectral parameters.

Substituting Eq. [2] into Eq. [1] and evaluating the geometric sum analytically we obtain

$$I(s) = \left(\Phi_0 \left| \left\{ \frac{1}{1 - e^{i\tau s} \hat{U}} - \frac{1}{2} \right\} \Phi_0 \right. \right). \quad [4]$$

Equation [4] cannot be used directly for calculating $I(s)$ as we still need to obtain a matrix representation of the auxiliary objects \hat{U} and Φ_0 in terms of the known data $c(n\tau)$. To do so, as in Ref. (3), we introduce an auxiliary (Krylov) basis,

$$\Phi_n = \hat{U}^n \Phi_0, \quad n = 0, 1, \dots, M-1. \quad [5]$$

According to our assumption, the rank of \hat{U} is K , so that the $M \leq K$ Krylov vectors Φ_n are generally linearly independent and we can rewrite Eq. [4] by evaluating everything in this basis:

$$I(s) \approx \mathbf{C}^T \mathbf{R}(s)^{-1} \mathbf{C} - \frac{c(0)}{2}, \quad [6]$$

with $\mathbf{R}(s) = \mathbf{U}_0 - e^{i\tau s} \mathbf{U}_1$, where the evolution operator and the overlap matrix elements are defined as, respectively, $[\mathbf{U}_1]_{nn'} = (\Phi_n | \hat{U} \Phi_{n'})$, $[\mathbf{U}_0]_{nn'} = (\Phi_n | \Phi_{n'})$, and the coefficients of the $1 \times M$ column vector \mathbf{C} are $[\mathbf{C}]_n = (\Phi_n | \Phi_0)$. It is not hard to see (3) that \mathbf{U}_1 , \mathbf{U}_0 , and \mathbf{C} are all representable in terms of the signal data points $c(n\tau)$ as

$$\begin{aligned} [\mathbf{U}_p]_{nn'} &= c[(n + n' + p)\tau], \quad p = 0, 1 \\ [\mathbf{C}]_n &= c(n\tau). \end{aligned} \quad [7]$$

Rather surprisingly, Eqs. [6] and [7] are working expressions. In the case that Eq. [3] is exactly satisfied, Eq. [6] yields the exact infinite time Fourier spectrum,

$$I(s) = \sum_{k=1}^K \frac{d_k}{1 - e^{i\tau(s - \omega_k)}} - \frac{c(0)}{2}, \quad [8]$$

if we choose $M = K$, even though only a finite part of the signal $c(n\tau)$ of size $N = 2M$ is used and the spectral parameters ω_k and d_k are not computed. The result is also exact if $M > K$, although in this case the set of vectors Φ_n is linearly dependent, requiring evaluation of a pseudo-inverse of the singular $M \times M$ matrix $\mathbf{R}(s)$. For example, the singular value decomposition (SVD) of $\mathbf{R}(s)$ could be used. In practice measured data contains some noise, so the matrices are not exactly singular, although they could still be very ill-conditioned, implying that some kind of regularization will often be advantageous. We will revisit this issue later in the paper.

Our result is also interesting because it is reminiscent of the method referred to as MUSIC (multiple signal classification). In MUSIC, a pseudo-spectrum is constructed by evaluating a certain pseudo-inverse of a data matrix (4). Note that unlike the present approach, the MUSIC spectrum is only useful for finding the unknown complex frequencies ω_k and so is not useful for NMR applications. Another approach, LPZ (linear prediction z -transform) (5), is a variant of the LP method in which the infinite-time Fourier spectrum is estimated directly from the data matrix inversion (more precisely, from the LP coefficients). However, LPZ has an unfavorable N^3 numerical scaling with respect to the data size N and is not commonly used. Just like LPZ, Eq. [6] is numerically also very expensive for realistic data sets. Fortunately, this problem can be circumvented by performing the spectral analysis locally in the frequency domain. The FDM approach (2, 3) is similar in spirit to that commonly used in the engineering literature and known as ‘‘beamspacing’’ (see, e.g., Ref. (6)). Also see Ref. (7) related to the present context). Here a basis localized in the frequency domain is obtained by Fourier transformation of the Krylov basis $\{\Phi_n\}$,

$$\tilde{\Phi}_j = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} \Phi_n, \quad j = 1, 2, \dots, K_{\text{win}}, \quad [9]$$

where here and throughout the rest of the paper the tilde identifies the use of the Fourier basis. The values φ_j could be evenly spaced using

$$\Delta\varphi = \frac{2\pi}{\mathfrak{N}M\tau}. \quad [10]$$

For $\mathfrak{N} = 1$ and $K_{\text{win}} = M$ the transformation from the Krylov

basis $\{\Phi_n\}$, to the Fourier basis $\{\tilde{\Phi}_j\}$ is unitary. However, as argued first in Ref. (2) and later in (3), it is advantageous to use a local basis of small size $K_{\text{win}} \ll M$ with $\mathfrak{N} > 1$ (e.g., $\mathfrak{N} = 1.1$). It could be even more efficient to use *multiscale basis* (8) corresponding to a nonuniform distribution of φ_j 's with j -dependent Fourier length $M = M_j$ in Eq. [9], which describes a very narrow frequency window in high resolution and the rest of the spectrum, in low resolution, while having minimal overall size.

One can now use the new Fourier basis, Eq. [9], to reevaluate the matrices in Eq. [7], which after some manipulations (3) leads for $j \neq j'$ to

$$[\tilde{\mathbf{U}}_p]_{jj'} = \hat{S} \sum_{\sigma=0,1} \frac{e^{i\sigma[\tau M(\varphi_{j'} - \varphi_j) + \pi]}}{1 - e^{i\tau(\varphi_{j'} - \varphi_j)}} \times \sum_{n=\sigma M}^{(\sigma+1)(M-1)} e^{in\tau\varphi_j} c[(n+p)\tau], \quad [11]$$

where \hat{S} defines symmetrization operator over the variables φ_j and $\varphi_{j'}$,

$$\hat{S}g(\varphi_j, \varphi_{j'}) = g(\varphi_j, \varphi_{j'}) + g(\varphi_{j'}, \varphi_j). \quad [12]$$

For $j = j'$ one has

$$[\tilde{\mathbf{U}}_p]_{jj} = \sum_{n=0}^{2M-2} e^{in\tau\varphi_j} (M - |M - n - 1|) c[(n+p)\tau]. \quad [13]$$

Now by evaluating \mathbf{C} in the Fourier basis,

$$[\tilde{\mathbf{C}}]_j = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} c(n\tau), \quad [14]$$

we can rewrite Eq. [6] as

$$I(s) \approx \tilde{\mathbf{C}}^T \tilde{\mathbf{R}}(s)^{-1} \tilde{\mathbf{C}} - \frac{c(0)}{2}, \quad [15]$$

with $\tilde{\mathbf{R}}(s) = \tilde{\mathbf{U}}_0 - e^{i\tau s} \tilde{\mathbf{U}}_1$.

Due to the very nature of the Fourier transformation the spectral properties around some frequency s are completely defined by a very small subspace $\{\tilde{\Phi}_j\}$ of size K_{win} (e.g., $K_{\text{win}} = 10$) with $\varphi_j \sim s$. Therefore, only a small $K_{\text{win}} \times K_{\text{win}}$ matrix $\tilde{\mathbf{R}}(s)$ has to be inverted in Eq. [15] to yield a well-converged spectrum $I(s)$.

Clearly, evaluation of Eq. [15] can be done for all values of s within a chosen frequency window, once the corresponding generalized eigenvalue problem,

$$\tilde{\mathbf{U}}_1 \tilde{\mathbf{B}}_k = u_k \tilde{\mathbf{U}}_0 \tilde{\mathbf{B}}_k, \quad [16]$$

is solved for the eigenvalues u_k and eigenvectors $\tilde{\mathbf{B}}_k$. Although this method might seem preferable to any other alternative, we note that Eq. [15] can also be evaluated directly, for example, by solving the associated linear system,

$$\tilde{\mathbf{R}}(s) \tilde{\mathbf{X}}(s) = \tilde{\mathbf{C}}, \quad [17]$$

and then using

$$I(s) \approx \tilde{\mathbf{C}}^T \tilde{\mathbf{X}}(s) - \frac{c(0)}{2}. \quad [18]$$

In the 1D case the latter approach, Eqs. [17] and [18], may actually appear computationally efficient if the spectrum is to be evaluated at relatively few values of s . This will be even more so in the case of $D = 2$ and higher dimensionality.

A less obvious issue is the stability and robustness of the algorithm. The matrix $\tilde{\mathbf{R}}(s)$ may be very ill-conditioned, so its inversion or use in Eq. [17] requires some kind of regularization. One possibility is to use SVD of $\tilde{\mathbf{R}}(s)$ to calculate a pseudo-inverse by discarding the singular subspace. However, SVD, if applied for each value of s , would be quite computationally expensive. A much less expensive regularization of the resolvent can be obtained using the Tikhonov regularization (9),

$$I(s) \approx \tilde{\mathbf{C}}^T (\tilde{\mathbf{R}}(s)^\dagger \tilde{\mathbf{R}}(s) + q^2)^{-1} \tilde{\mathbf{R}}(s)^\dagger \tilde{\mathbf{C}} - \frac{c(0)}{2}, \quad [19]$$

where the dagger denotes the Hermitian conjugate, and q is a real number. With such a regularization the singularity in the denominator is removed as $(\tilde{\mathbf{R}}^\dagger \tilde{\mathbf{R}} + q^2)$ is a Hermitian and positive definite matrix.

Equation [19] can be evaluated by solving the regularized Hermitian least squares problem,

$$(\tilde{\mathbf{R}}^\dagger(s) \tilde{\mathbf{R}}(s) + q^2) \tilde{\mathbf{X}}(s) = \tilde{\mathbf{R}}^\dagger \tilde{\mathbf{C}}, \quad [20]$$

and then using Eq. [18].

It should be noted that the expensive matrix–matrix multiplication $\tilde{\mathbf{R}}(s)^\dagger \tilde{\mathbf{R}}(s)$ in Eq. [20], which is a K_{win}^3 process, does not have to be performed at each value of s . Significant numerical saving can be achieved by using

$$\tilde{\mathbf{R}}(s)^\dagger \tilde{\mathbf{R}}(s) = \tilde{\mathbf{U}}_0^\dagger \tilde{\mathbf{U}}_0 + \tilde{\mathbf{U}}_1^\dagger \tilde{\mathbf{U}}_1 - e^{-i\tau s} \tilde{\mathbf{U}}_1^\dagger \tilde{\mathbf{U}}_0 - e^{i\tau s} \tilde{\mathbf{U}}_0^\dagger \tilde{\mathbf{U}}_1. \quad [21]$$

Generally, solution of a linear system scales also as K_{win}^3 , however, one can solve Eq. [20] iteratively at each s by, e.g.,

a conjugate gradient method, which may have much better scaling, thus significantly reducing the overall numerical effort. These issues will be explored in more detail in our forthcoming publications.

The spectral estimation given by Eq. [19] (or [18, 20]) is one of the main results of this paper. Operationally it has a status of “transform” (like FFT), while a “method,” e.g., the filter diagonalization method, would refer to a procedure that would generally be less obvious to use. More precisely, Eq. [19] corresponds to a direct nonlinear transformation, here called the *regularized resolvent transform*, of the time signal to the frequency domain. Unlike most other nonlinear high-resolution spectral estimators, RRT is very stable, computationally very inexpensive, and has adjusting parameters that are very straightforward to use. These parameters are K_{win} and \mathfrak{N} defining the size and spacing of the Fourier basis in the frequency domain and the regularization parameter q . Note that K_{win} could in principle be as small as 3, although a larger K_{win} generally improves the resolution, while increasing the cpu time according to the cubic scaling of a linear solver. For sufficiently large K_{win} , which is usually less than 100, the results do not change noticeably. Thus, K_{win} can be chosen according to how long one would like to wait for the spectrum to be computed. The choice for the basis density parameter, \mathfrak{N} , between 1.1 and 1.2 usually works well if a single-scale basis (as opposed to a multiscale basis (8)) is used.

The spectrum $I(s)$ computed by RRT with $q = 0$ should generally be indistinguishable from that computed by a previous version of FDM (3) based on solving the generalized eigenvalue problem, Eq. [16]. However, for $q > 0$ they will differ as demonstrated in the following model example.

The example, called “Jacob’s ladder” (Fig. 1), was borrowed from Ref. (10) in which the signal was synthesized by Eq. [3] using a set of 50 triplets with both the peak widths and the splittings gradually decreasing from the right to the left, allowing examination of the breakdown of the resolution for any fixed signal size. The assumed spectral width was 5 kHz with the right-most triplet at 2500 Hz and the left-most at 14 Hz. The signal was normalized to an initial value $c(0) = 16,384$ and then integerized by taking the closest lower integer for each $c(n\tau)$ in order to mimic the ADC. The spectral region shown has very small splittings and so is hard to resolve. The upper trace is the exact spectrum that could be obtained by RRT using, for example, $N = 48,000$ data points. The FFT spectrum using $N = 64,000$ cannot resolve all the triplets in this region, while using FFT with $N = 32,000$ does not resolve any of them. In addition, one can see some baseline ringing that is not completely removed by apodization of the FID. The RRT result using $N = 32,000$ and $q = 0$ is very close to the exact result, although some amplitudes are slightly inaccurate. Interestingly, one can control the appearance of the RRT spectrum by tuning q . An increase of q makes the spectral estimation more uniform while decreasing the resolution.

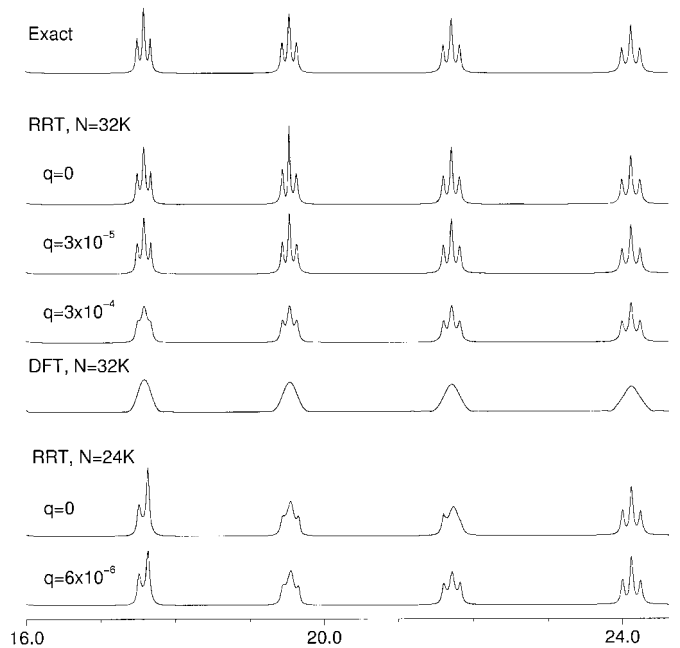


FIG. 1. “Jacob’s ladder” absorption spectra obtained by processing a model time signal (see text) by the regularized resolvent transform (RRT) and FFT using different signal sizes, $N = 32,000$ and $N = 24,000$. The RRT spectra have higher resolution than the FFT spectrum with $N = 32,000$, but the nonregularized spectra ($q = 0$) may be somewhat nonuniform. An increase of the regularization parameter q generally leads to a more uniform spectral estimate and gradual decrease of resolution, although the best result is often obtained using RRT with some small $q > 0$.

(Note that the scale for q is not related to the frequency scale, but rather to the normalization of the U -matrices.)

The lower two traces show RRT spectra using $N = 24,000$, i.e., in the regime where the method fails to resolve the narrowest triplets in the spectral region because the signal length is too short. The resolution failure for $q = 0$ manifests itself in a quite nonuniform appearance of the peaks in the multiplets. In particular, only two peaks appear in the left-most triplet with the smallest spacings. By increasing q one can make them more uniform and sometimes improve the resolution (e.g., note the triplet at ~ 22 Hz). That is, for short data sets the RRT “fails” in a controllable fashion (as is the case for FFT), while typically providing a higher resolution than FFT for narrow Lorentzian lines.

It is important to note here a subtlety present in all linear algebraic algorithms, in particular, in FDM and RRT, related to the fit of the time signal by damped sinusoids, Eq. [3]. For example, in RRT the derivation of Eq. [4] assumes convergence of the infinite geometric series $\sum_{n=0}^{\infty} (e^{i\tau s} \hat{U})^n$. However, this assumption may be ambiguous since only a finite part of the data is available. Numerically, when evaluated in a finite basis \hat{U} may have eigenvalues $u_k = e^{-i\tau\omega_k}$ outside the unit circle, corresponding to the negative “width” $\gamma_k = -\text{Im}\{\omega_k\}$. If, at the same time, the d_k is small there is no problem.

Now consider two dangerous cases.

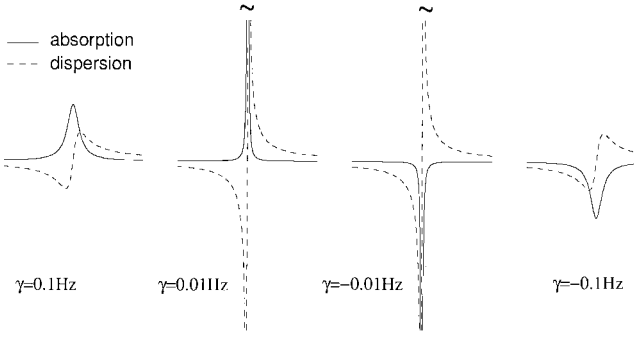


FIG. 2. The real (absorption) and imaginary (dispersion) parts of the RRT spectrum $I(s) = d_k/(1 - e^{i\tau(s-\omega_k)})$ for a single line as a function of its “width” $\gamma = -\text{Im}\{\omega_k\}$. The sign of the absorption peak is flipped when γ changes its sign, while the appearance of the dispersion peak only depends on the absolute value of γ .

(i) Both d_k and $\gamma_k < 0$ are large. Even though the signal $c(n\tau)$ is expected to decay at large enough times, this problem can arise as an artifact of a local spectral analysis caused by a narrowband Fourier basis that in turn does not describe the spectral features outside the chosen spectral region. (Note that with multiscale basis (8) such poles do not occur.) Surprisingly, even though they have nothing to do with the true spectral parameters, the poles with both large positive imaginary parts and large amplitudes correctly describe the background spectrum in the chosen interval if Eq. [15] is used. In Eq. [15] the large contributions from broad, interfering poles cancel each other.

(ii) The true width of a peak is very small, or zero as in a constant-time experiment, making it likely (because of numerical errors or noise) to have $\gamma_k < 0$. This in turn results in a flip of the sign of the absorption part of the peak as shown in Fig. 2. Note, though, that the sign of the dispersion part is unaffected. This property of the RRT lineshapes can be easily understood by considering the behavior of $I(s)$ (see Eq. [8]) near an eigenfrequency $\omega = \omega_r - i\gamma$, i.e., assuming $\tau|s - \omega| \ll \pi$ and then extracting the real and imaginary parts of the complex Lorentzian as

$$\begin{aligned} I(s) &\approx \frac{d_k}{1 - e^{i\tau(s-\omega)}} \\ &\approx \frac{id_k}{\tau} \frac{1}{s - \omega} \\ &= \frac{id_k}{\tau} \left[\frac{(s - \omega_r)}{(s - \omega_r)^2 + \gamma^2} + \frac{i\gamma}{(s - \omega_r)^2 + \gamma^2} \right]. \quad [22] \end{aligned}$$

In FDM and other parameter estimators the negative sign of γ can simply be flipped to lead to the correct appearance of the peak in the phase-sensitive spectrum. Since in RRT the eigenvalues are not computed, one cannot manipulate these in any simple fashion. One way to circumvent this problem in RRT is

to shift the argument of $I(s)$ by $i\Gamma$ with $\Gamma > -\gamma$, i.e., construct $I(s + i\Gamma)$ which effectively flips the negative width γ . The result of such a shift is demonstrated in the third trace of Fig. 3 where all the peaks have the correct sign, but are slightly broadened.

3. 2D RRT

The generalization of the 1D RRT to the multidimensional case is reminiscent to the derivations in Refs. (11, 12), except here we completely avoid even mentioning the Hamiltonian operators. Since there is only a minor difference between the 2D RRT and $D > 2$ RRT, for the sake of simplicity we restrict our presentation to the $D = 2$ case.

To this end, given a 2D time signal $c(n_1\tau_1, n_2\tau_2)$ defined on an equidistant rectangular time grid, our goal is to estimate the 2D (infinite time) discrete Fourier sum,

$$\begin{aligned} I(s_1, s_2) &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} e^{in_1\tau_1 s_1} e^{in_2\tau_2 s_2} \\ &\times c(n_1\tau_1, n_2\tau_2) \left(1 - \frac{\delta_{n_1 0}}{2}\right) \left(1 - \frac{\delta_{n_2 0}}{2}\right), \quad [23] \end{aligned}$$

using only the finite $N_1 \times N_2$ part of the signal.

The 1D assumption of Eq. [2] is generalized (13, 14) by using two commuting complex symmetric evolution operators \hat{U}_1 and \hat{U}_2 ,

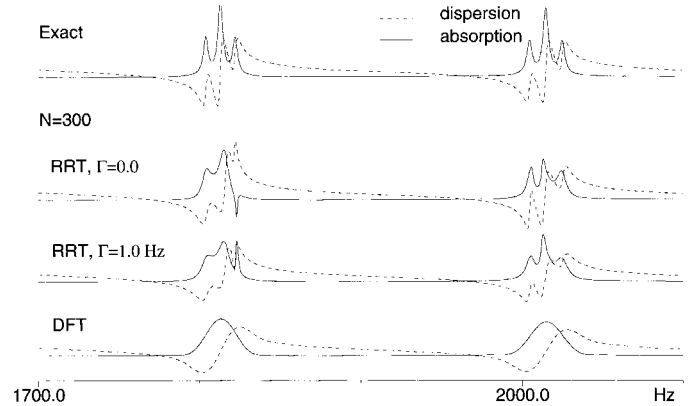


FIG. 3. An example of a failure of the RRT spectral representation with no shift and no regularization implemented (second trace, both $q = 0$ and $\Gamma = 0$). Because the signal used to construct the spectrum is too short in length, both the amplitudes and the widths of the peaks are quite inaccurate. In particular, one peak appears with the wrong sign (note that the appearance of all the dispersion-mode peaks is correct). The wrong absorption peak is flipped by plotting $I(s + i\Gamma)$ (the lowest trace) instead of $I(s)$ with $\Gamma = 1$ Hz. Note that this procedure replaces the width $\gamma_k = -\text{Im}\{\omega_k\}$ of each peak by $\gamma_k + \Gamma$, i.e., effectively fixes the signs of all peaks with $-\Gamma < \gamma_k < 0$ and broadens all the peaks with the (correct) positive γ_k .

$$c(n_1\tau, n_2\tau_2) = (\Phi_0 | \hat{U}_1^{n_1} \hat{U}_2^{n_2} \Phi_0). \quad [24]$$

Inserting Eq. [24] into Eq. [23] and evaluating the geometric sums analytically gives

$$I(s_1, s_2) = \left(\Phi_0 \left| \left\{ \frac{1}{1 - e^{i\tau_1 s_1} \hat{U}_1} - \frac{1}{2} \right\} \times \left\{ \frac{1}{1 - e^{i\tau_2 s_2} \hat{U}_2} - \frac{1}{2} \right\} \right| \Phi_0 \right). \quad [25]$$

The 2D Krylov basis is

$$\Phi_{n_1, n_2} = \hat{U}_1^{n_1} \hat{U}_2^{n_2} \Phi_0, \quad [26]$$

with $n_1 = 0, 1, \dots, M_1 - 1$, and $n_2 = 0, 1, \dots, M_2 - 1$, and the total size $M_1 \times M_2 = N_1/2 \times N_2/2$ defined by the signal size.

A 2D Fourier basis is introduced by defining a 2D grid in the frequency domain $(\varphi_{1j}, \varphi_{2j})$, $j = 1, 2, \dots, K_{\text{win}}$, that spans only a small region around the reference point (s_1, s_2) . A convenient choice corresponds to a direct product grid with $K_{\text{win}} = K_{1\text{win}} \times K_{2\text{win}}$ in which the spacing between the grid points in an l th dimension is defined according to Eq. [10].

Thus, we have

$$\tilde{\Phi}_j = \sum_{n_1=0}^{M_1-1} \sum_{n_2=0}^{M_2-1} e^{in_1\tau_1\varphi_{1j}} e^{in_2\tau_2\varphi_{2j}} \Phi_{n_1, n_2}. \quad [27]$$

Now evaluating Eq. [25] in the 2D Fourier basis we obtain the resolvent transform analogous to Eq. [15],

$$I(s_1, s_2) \approx \tilde{\mathbf{C}}^T \left\{ \tilde{\mathbf{R}}_1(s_1)^{-1} \tilde{\mathbf{U}}_0 \tilde{\mathbf{R}}_2(s_2)^{-1} - \frac{\tilde{\mathbf{R}}_1(s_1)^{-1} + \tilde{\mathbf{R}}_2(s_2)^{-1}}{2} \right\} \tilde{\mathbf{C}} + \frac{c(0, 0)}{4}, \quad [28]$$

with

$$[\tilde{\mathbf{C}}]_j = \sum_{n_1=0}^{M_1-1} \sum_{n_2=0}^{M_2-1} e^{in_1\tau_1\varphi_{1j}} e^{in_2\tau_2\varphi_{2j}} c(n_1\tau_1, n_2\tau_2), \quad [29]$$

and

$$\tilde{\mathbf{R}}_l(s_l) = \tilde{\mathbf{U}}_0 - e^{i\tau_l s_l} \tilde{\mathbf{U}}_l, \quad l = 1, 2. \quad [30]$$

The expressions to evaluate the overlap matrix, $\tilde{\mathbf{U}}_0$, and the matrix representations of the evolution operators, $\tilde{\mathbf{U}}_1$ and $\tilde{\mathbf{U}}_2$, were derived previously (13) and are given below without derivation.

For convenience define three auxiliary data sets,

$$\begin{aligned} c_0(n_1, n_2) &\equiv c[n_1\tau_1, n_2\tau_2], \\ c_1(n_1, n_2) &\equiv c[(n_1 + 1)\tau_1, n_2\tau_2], \\ c_2(n_1, n_2) &\equiv c[n_1\tau_1, (n_2 + 1)\tau_2]. \end{aligned} \quad [31]$$

The matrix elements of the U -matrices for both $\varphi_{1j} \neq \varphi_{1j'}$ and $\varphi_{2j} \neq \varphi_{2j'}$ can be computed by

$$\begin{aligned} [\mathbf{U}_l]_{jj'} &= \hat{S}_1 \sum_{\sigma_1=0,1} \frac{e^{i\sigma_1[M_1\tau_1(\varphi_{1j'} - \varphi_{1j}) + \pi]}}{1 - e^{i\tau_1(\varphi_{1j'} - \varphi_{1j})}} \\ &\times \hat{S}_2 \sum_{\sigma_2=0,1} \frac{e^{i\sigma_2[M_2\tau_2(\varphi_{2j'} - \varphi_{2j}) + \pi]}}{1 - e^{i\tau_2(\varphi_{2j'} - \varphi_{2j})}} \\ &\times \sum_{n_1=\sigma_1 M_1}^{(\sigma_1+1)(M_1-1)} \sum_{n_2=\sigma_2 M_2}^{(\sigma_2+1)(M_2-1)} e^{in_1\varphi_{1j}} e^{in_2\varphi_{2j}} c_l(n_1, n_2), \end{aligned} \quad [32]$$

where \hat{S}_1 and \hat{S}_2 define the symmetrization operators over the corresponding pairs of variables as, e.g.,

$$\hat{S}_1 g(\varphi_{1j}, \varphi_{1j'}) = g(\varphi_{1j}, \varphi_{1j'}) + g(\varphi_{1j'}, \varphi_{1j}). \quad [33]$$

For $\varphi_{1j} = \varphi_{1j'}$ and $\varphi_{2j} \neq \varphi_{2j'}$ we have

$$\begin{aligned} [\mathbf{U}_l]_{jj'} &= \hat{S}_2 \sum_{\sigma_2=0,1} \frac{e^{i\sigma_2[M_2\tau_2(\varphi_{2j'} - \varphi_{2j}) + \pi]}}{1 - e^{i\tau_2(\varphi_{2j'} - \varphi_{2j})}} \\ &\times \sum_{n_1=0}^{2M_1-2} \sum_{n_2=\sigma_2 M_2}^{(\sigma_2+1)(M_2-1)} e^{in_1\varphi_{1j}} e^{in_2\varphi_{2j}} c_l(n_1, n_2) \\ &\times (M_1 - |M_1 - n_1 - 1|), \end{aligned} \quad [34]$$

which can trivially be rewritten for the symmetric case of $\varphi_{1j} \neq \varphi_{1j'}$ and $\varphi_{2j} = \varphi_{2j'}$. For the case of both $\varphi_{1j} = \varphi_{1j'}$ and $\varphi_{2j} = \varphi_{2j'}$, i.e., the diagonal elements of the U -matrices, we have

$$\begin{aligned} [\mathbf{U}_l]_{jj} &= \sum_{n_1=0}^{2M_1-2} \sum_{n_2=0}^{2M_2-2} e^{in_1\varphi_{1j}} e^{in_2\varphi_{2j}} c_l(n_1, n_2) \\ &\times (M_1 - |M_1 - n_1 - 1|)(M_2 - |M_2 - n_2 - 1|). \end{aligned} \quad [35]$$

A 2D RRT can be obtained by regularizing the two resolvents in Eq. [28], leading to

$$I(s_1, s_2) \approx \tilde{\mathbf{X}}(s_1)^T \tilde{\mathbf{U}}_0 \tilde{\mathbf{X}}(s_1) + \frac{c(0, 0)}{4} - \tilde{\mathbf{C}}^T \frac{\tilde{\mathbf{X}}(s_1) + \tilde{\mathbf{X}}(s_2)}{2}, \quad [36]$$

with the two frequency-dependent vectors $\tilde{\mathbf{X}}_l(s_l)$, $l = 1, 2$, computed by solving the regularized Hermitian least squares problems,

$$(\tilde{\mathbf{R}}_l(s_l)^\dagger \tilde{\mathbf{R}}_l(s_l) + q^2) \tilde{\mathbf{X}}_l(s_l) = \tilde{\mathbf{R}}_l(s_l)^\dagger \tilde{\mathbf{C}}. \quad [37]$$

Note that the total number of the linear systems to be solved for each 2D frequency window is equal to $N_{s_1} + N_{s_2}$, where N_{s_1} and N_{s_2} are the numbers of the frequency grid points, s_1 and s_2 , to plot the spectrum in the window.

Derivation of the 2D RRT is the main result of this paper. Its potential advantage compared to the 2D DFT is, similarly to the 2D FDM, in the ability to process the whole 2D data set simultaneously, thus converting the time domain information into the frequency domain with minimal loss. Unlike the 2D FDM, in RRT we do not directly refer to the 2D harmonic inversion problem,

$$c(n_1\tau_1, n_2\tau_2) = \sum_{k=1}^K d_k e^{-in_1\tau_1\omega_{1k}} e^{-in_2\tau_2\omega_{2k}}, \quad [38]$$

although the two methods are very closely related. In particular, if the form of (Eq. [38]) is satisfied, the RRT spectrum for a sufficiently large data set processed should converge to

$$I(s_1, s_2) = \frac{c(0, 0)}{4} + \sum_{k=1}^K \frac{d_k}{2} \times \left\{ \frac{2}{[1 - e^{i\tau_1(s_1 - \omega_{1k})}][1 - e^{i\tau_2(s_2 - \omega_{2k})}]} - \frac{1}{[1 - e^{i\tau_1(s_1 - \omega_{1k})}]} - \frac{1}{[1 - e^{i\tau_2(s_2 - \omega_{2k})}]} \right\}. \quad [39]$$

Ideally, the results should fully converge if the total size of the data set satisfies the condition $N_{\text{total}} = N_1 \times N_2 \geq 4K$, no matter whether N_2 or N_1 are small. This, of course, does not take into account noise, degeneracies, and round-off errors which could affect the convergence conditions significantly.

Due to its numerical stability, robustness, and ease of implementation, RRT has an advantage over FDM, once only *complex* 1D or 2D spectra, as defined by Eqs. [1] and [23], are of interest. However, an advantage of FDM, especially in the multidimensional case, lies in the ability to

construct various nonanalytic spectral representations, such as a double-absorption spectrum (11) using a single purely phase modulated 2D signal due to the availability of the eigenfrequencies ω_{1k} and ω_{2k} . Furthermore, in FDM inaccuracies in determining the imaginary parts of the computed frequencies, in particular, their signs, can be corrected (see above). In RRT no spectral parameters are produced and, therefore, the artifacts are removed by regularization with a finite value of q and/or by shifting the spectrum using $I(s_1 + i\Gamma_1, s_2 + i\Gamma_2)$ with shifting parameters Γ_1 and Γ_2 so that the spectrum is always smoothed somewhat.

In the next example we apply 2D RRT to a frequently employed 2D experiment to correlate proton and nitrogen-15 chemical shifts. The 2D shift correlation spectra of the nitrogen-15-labeled metalloprotein rubredoxin (15) shown in Fig. 4 are particularly favorable for FFT analysis because a phase-sensitive spectrum can be computed and the peaks are of similar intensity, so that a poorly resolved intense peak does not mask a weaker feature nearby. Nevertheless, RRT yields a welcome improvement in resolution or a decrease in the experiment time for a similar resolution. As in FDM, the gain in resolution hinges on the spectrum having sufficient signal-to-noise ratio. It is neither possible to improve the resolution of noisy spectra, nor to magically divine the presence of peaks buried in noise.

3.1. Other Spectral Representations using RRT

When severely truncated data sets are used in RRT, just like in FDM, the positions of the peaks converge best, while the phases of the amplitudes are least accurate. This may, in turn, lead to incorrect lineshapes in phase-sensitive spectra. Moreover, in certain situations only a single purely phase modulated data set is available (e.g., 2D J experiments). In such cases an absolute-value RRT spectrum may be an option, although the resolution of the latter may often be unacceptable due to the contribution of the dispersion lineshapes. Note, though, the existence of a variety of other spectral representations, in which the dispersion contributions are eliminated, and which, therefore, lead to much higher resolution. An example of such a representation in 2D spectroscopy is

$$A(s_1, s_2) = \text{Re}\{I(s_1, s_2 + i\Gamma_2) - I(s_1, s_2 + i\Gamma'_2)\}. \quad [40]$$

The dispersion contributions along s_2 are eliminated by the subtraction of the two complex spectra using different shifts $\Gamma_2 \neq \Gamma'_2$ (with possibly positive Γ_2 and negative Γ'_2), while taking the real part leads to the absorption lineshapes along s_1 . The disadvantage of Eq. [40] is the need to fiddle with too many adjusting parameters.

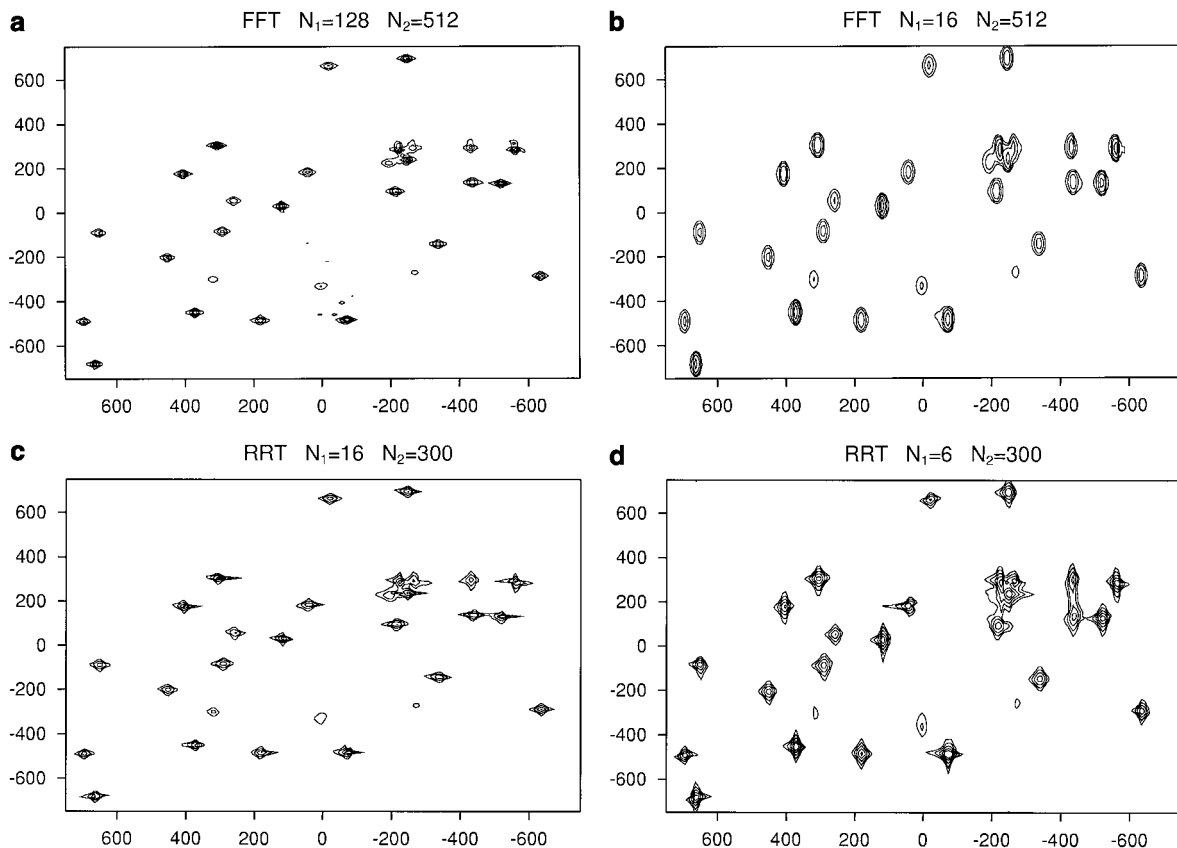


FIG. 4. Chemical shift correlation spectra of the metalloprotein rubredoxin obtained with gradient selection of the coherence transfer pathway. Each of the double-absorption spectra were generated using the conventional procedure that combines the complex N - and P -type spectra.

Another possible spectral representation is given by

$$I^{(2)}(s_1, s_2) = \left(\Phi_0 \left| \frac{1}{[1 - e^{i\tau_1 s_1} \hat{U}_1]^2 [1 - e^{i\tau_2 s_2} \hat{U}_2]^2} \Phi_0 \right| \right) \\ = \sum_{k=1}^K \frac{d_k}{[1 - e^{i\tau_1(s_1 - \omega_{1k})}]^2 [1 - e^{i\tau_2(s_2 - \omega_{2k})}]^2}, \quad [41]$$

in which the peaks are narrow and have quasi-absorption lineshapes if the absolute value spectrum $|I^{(2)}(s_1, s_2)|$ is plotted. That is, $|I^{(2)}(s_1, s_2)|$ will have a great resolution advantage over $I(s_1, s_2)$. Note, also, that an amplitude d_k is not squared, although the peak integral will effectively be divided by the factor $\text{Im}\{\omega_{1k}\}\text{Im}\{\omega_{2k}\}$, defined by the two widths. That is, $I^{(2)}(s_1, s_2)$ will favor the narrow peaks against the broad ones. As before, these effects can be reduced by shifting the spectrum using $|I^{(2)}(s_1 + i\Gamma_1, s_2 + i\Gamma_2)|$ which effectively increases the widths of all peaks, making them more uniform.

While $I^{(2)}(s_1, s_2)$ cannot be constructed by FT, it is representable in terms of the U -matrices in complete analogy with Eq. [28]. Of course, the corresponding resolvents should be regularized accordingly. In Fig. 5 we show the RRT spectra

using $|I^{(2)}(s_1 + i\Gamma_1, s_2 + i\Gamma_2)|$ with $\Gamma_1 = 1$ Hz and $\Gamma_2 = 10$ Hz applied to the same 2D signal as in Fig. 4. The peaks do have absorption lineshapes, although the heights are somewhat distorted when compared to the correct spectra of Fig. 4. Clearly, many other spectral representations of this type are

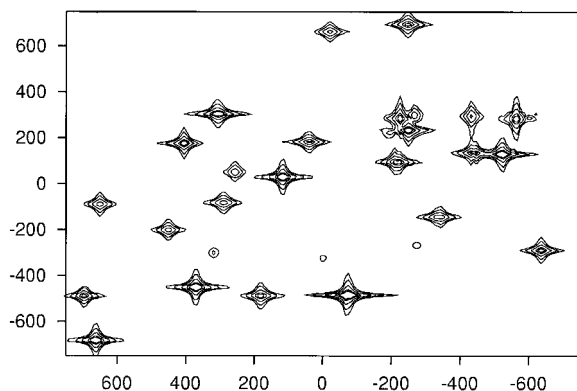


FIG. 5. RRT pseudo-absorption spectrum $|I^{(2)}(s_1 + i\Gamma_1, s_2 + i\Gamma_2)|$ (see Eq. [41]) constructed using only the N -type (i.e., the purely phase modulated) data of Fig. 4 with $N_1 = 16$, $N_2 = 300$, $q = 0.0003$, $\Gamma_1 = 2$ Hz, and $\Gamma_2 = 1$ Hz.

available, each of which can have their own advantages and disadvantages.

4. SUMMARY

We have presented new numerical expressions for high-resolution spectral estimation of multidimensional time signals. Under the condition that the data exactly fits the Lorentzian assumption, RRT possesses the optimal property of being able to produce the exact infinite time DFT spectra using only a finite and small portion of the time domain data.

The main advantages of RRT are the following. (i) It is a direct transformation of the time domain data into the frequency domain spectrum. (ii) Unlike most other nonlinear spectral estimators, it is easy to use. There is no need to fiddle much with the adjusting parameters as the results only smoothly depend on them. (iii) Unlike the multidimensional versions of FDM, in RRT the regularization can be implemented in a very straightforward fashion. The regularization parameter q unambiguously controls the spectral appearance: a larger value of q systematically suppresses the artifacts and small spectral features, providing a more uniform spectral estimate, albeit with decreasing resolution.

Although FDM and RRT are very closely related, RRT, after the regularization, loses the structure that was present originally in FDM, which allows one to obtain a parametric description, i.e., a direct line list. The inability to provide a line list seems to be the main disadvantage of RRT. This, in particular, makes it difficult to obtain spectral representations (e.g., a double-absorption spectrum from a purely phase modulated data) that are different from the conventional DFT. Therefore, FDM has, in principle, fewer limitations both on the type of data to be processed and on the spectra to be constructed. These issues will be explored in more detail in our forthcoming publications.

ACKNOWLEDGMENTS

This work was supported by the NSF. J.C. and V.A.M. acknowledge Grant CHE-9807229 and A.J.S. Grant CHE-9900422.

REFERENCES

1. E. O. Brigham, "The Fast Fourier Transform," Prentice Hall, Englewood Cliffs, NJ, 1974.

2. 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).
3. V. A. Mandelshtam and H. S. Taylor, Harmonic inversion of time signals and its applications, *J. Chem. Phys.* **107**, 6756–6769 (1997).
4. R. O. Schmidt, Multiple emitter location and signal parameter estimation, Proc. RADC Spectrum Estimation Workshop, pp. 243–258, Griffiss AFB, New York, 1979.
5. J. Tang and J. R. Norris, LPZ spectral analysis using linear prediction and the z-transform, *J. Chem. Phys.* **84**, 5210–5211 (1986).
6. S. D. Silverstein and M. D. Zoltowski, The mathematical basis for element and Fourier beamspace MUSIC and root-MUSIC algorithms, *Dig. Signal Process.* **1**, 161–175 (1991).
7. D. Belkic, P. A. Dando, H. S. Taylor, and J. Main, Decimated signal diagonalization for obtaining the complete eigenspectra of large matrices, *Chem. Phys. Lett.* **315**, 135–139 (1999).
8. J. Chen and V. A. Mandelshtam, Multiscale filter diagonalization method for spectral analysis of noisy data with nonlocalized features, *J. Chem. Phys.* **112**, 4429–4437 (2000).
9. A. Tikhonov, Solution of incorrectly formulated problems and the regularization method, *Soviet Math. Dokl.* **4**, 1035–1038 (1963); G. H. Golub and C. F. van Loan, "Matrix Computations," Johns Hopkins Univ. Press, Baltimore, 1989.
10. 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).
11. V. A. Mandelshtam, N. D. Taylor, H. Hu, 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).
12. V. A. Mandelshtam, The multidimensional filter diagonalization method. I. Theory and numerical implementation, *J. Magn. Reson.* **144**, 343–356 (2000).
13. V. A. Mandelshtam and H. S. Taylor, Multidimensional harmonic inversion by filter-diagonalization, *J. Chem. Phys.* **108**, 9970–9977 (1998).
14. M. R. Wall, T. Dieckmann, J. Feigon, and D. Neuhauser, Two-dimensional filter-diagonalization: Spectral inversion of 2D NMR time-correlation signals including degeneracies, *Chem. Phys. Lett.* **291**, 465–470 (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).