The regularized resolvent transform for processing DOSY spectra

Nikolaus Loening[†], James Keeler[†],

Chemistry Department University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

and

Geoffrey S. Armstrong[*], Joseph E. Curtis[*], A.J. Shaka[*] and Vladimir A. Mandelshtam[*]

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

E-mail: garmstro@uci.edu; curtisj@uci.edu; nloening@mit.edu; James.Keeler@ch.cam.ac.uk; ajshaka@uci.edu; mandelsh@uci.edu

We present a method, called iRRT, for processing Diffusion Ordered Spectroscopy (DOSY) data, that can produce results even in overlapping regions of the spectrum. From the results we obtain diffusion constants for camphene, geraniol, and quinine of $12 \times 10^{-6}\mathrm{cm^2s^{-1}}$, $9 \times 10^{-6}\mathrm{cm^2s^{-1}}$, and $6 \times 10^{-6}\mathrm{cm^2s^{-1}}$ respectively. We compare the results from this method to those obtained using a mono-exponential non-linear least squares fit. Both the 2D DOSY spectrum and the 1D spectra corresponding to each component of the mixture obtained from iRRT compare favorably to the spectra of the pure substances, and offer a vast improvement over the exponential fit results. This method is presented as a stable and reliable tool for solving the Inverse Laplace Transform (ILT) problem present in experiments such as DOSY.

Key Words: spectral estimation, diffusion ordered spectroscopy, inverse Laplace transform, regularized resolvent transform, filter diagonalization method.

1. INTRODUCTION

NMR spectroscopic methods have benefitted greatly from the development of multi-dimensional methods. Traditionally, these methods have focused on spin interactions, but recent advances have allowed the investigation of other properties. For example, the development of Pulsed Field Gradients (PFG), have allowed properties such as size and conformation to be investigated by NMR. As the gradients reveal spatial information about the sample, the translation of molecules in the NMR tube can be investigated[1, 2]. An example of this type of experiment is Diffusion Ordered Spectroscopy (DOSY) [3, 4, 5]. This experiment allows spectra of mixtures to be resolved into individual components, a feature that is very useful for examining impure samples, or biological fluid samples [6]. As the diffusion constant changes after binding, this method can also be used to separate promising drug candidates from a mixture of test compounds [7, 8]. Despite the advantages that DOSY offers, it's growth has been hampered by difficulties associated with the experiment. Until recently, DOSY experiments required expensive probes to afford the highest resolution. Also, pulse sequences needed to be developed to minimize the effects of eddy currents introduced by the strong gradients [9]. Finally, the problem of spectral analysis remains.

While both the Fourier Transform and the Inverse Fourier Transform exist and are stable for complex data analysis, they cannot be applied to the diffusion dimension of a DOSY experiment since the signals only decay allong this dimension without oscillations, so the Fourier spectral analysis does not reveal any useful information. With the correct scaling along the diffusion dimension the DOSY signal can be well described by an exponential decay. In this case an Inverse Laplace Transform (ILT) formally provides the relevant information, i.e., the positions of the peaks in the ILT spectrum give the decay constants (diffusion constants).

Unfortunately, the one-dimensional (1D) ILT corresponds to a very ill-defined inverse problem. In practice, it is often complicated by the fact that the data to be inverted is noisy and very short (decays quickly), i.e. it does not contain much information in the first place. Under such conditions it is not surprising that practically all existing numerical algorithms to perform the 1D ILT are very capricious. This explains the numerous attempts to solve

the problem [5, 10, 11, 12, 13, 14, 15]. The approach based on a mono-exponential fit [16] is most commonly used because of its relative stability and robustness. However, it gives poor results in the cases of mixtures, when the overlapping resonances lead to multiexponential decay. In theory the latter cases can be handled by a multi-exponential fitting. In practice it is very unstable and often produces spurious solutions, which can hardly be cured by regularization. There were attempts to develop techniques that could process the whole 2D DOSY data [15], that is, to exploit the correlations which exist in the 2D data and as such to provide more information than that contained in a 1D slice of a 2D array. However, to the best of our knowledge those are not commonly used. It is probably fair to claim that each of the existing methods have its own advantages and short-comings, at least we are sure in the latter. So far, a robust technique that performs well in a wide range of circumstances has not been developed. A combination of processing techniques is therefore necessary to ensure the "correct" fit has been made or at least to provide several alternatives. It was in the interest of finding one that this study began. It has proven to be a challenging problem that is not completely solved, but we are presenting an effective method that prevails even in the case of heavily overlapping resonances. The present technique is based on the Regularized Resolvent Transform (RRT) [26] which is in turn an extension of the Filter Diagonalization Method (FDM) [17, 18, 28] that have been used very successfully recently to process multidimensional NMR spectra [19, 20, 21, 24, 25, 27]. The two methods are similar although FDM was originally designed to solve the *harmonic inversion problem*, while the RRT solves the *spectral estimation problem*. The main advantage of FDM/RRT is that they are both true multidimensional methods, i.e., they process the whole data at once. They are relatively inexpensive compared to many other methods, especially those based on nonlinear optimization. In principle both approaches are valid, however, since the ILT problem is very ill-posed, the regularization aspect here is crucial [29, 22]. In FDM one has to regularize an ill-conditioned generalized eigenvalue problem, while in RRT, an ill-conditione linear system. It turns out that regularization of the former is very tricky, while regularization of linear systems is straightforward and well understood. So, our method of choice is RRT.

2. THEORY

Formulation of the spectral invertsion problem

In a DOSY experiment the PFG g_m is incremented rather than the delay (Δ') to decrease experiment time. The resulting signal is a 2D array $c(n,m) := c(n\tau, g_m)$ with $n = 0, .., N-1$ and $m = 0, ..., M-1$.

The 2D signal is assumed to be composed of a finite number, K, of components that oscillate along the acquisition time dimension (n) and decay exponentially in the diffusion dimension (m):

$$c(n,m) = \sum_{k=1}^{K} d_k u_k^n \lambda_k^m \ . \tag{1}$$

Here d_k are complex amplitudes that include the phase information. If we use the representations $u_k = e^{-i\tau\omega_k}$ and $\lambda_k = e^{-\alpha_k \beta^2 \Delta'}$, physically, ω_k will play the role of *complex frequencies* with Re ω_k corresponding to the resonance positions and Im ω_k, the resonance widths; α_k then correspond to the diffusion constants. In most cases, α_k are real and assume only a few different values (except for poly-disperse samples). However, numerically introducing any constraints on α_k may be problematic and is not done here. β is a fitting parameter and is described in the next section.

The 2D parameter estimation problem (1) is appealing but not used here explicitly. Instead the more convenient for deriving the working expressions (albeit less appealing) FDM assumption is made [19]:

$$c(n,m) = \Phi^{\mathrm{T}} U^n \Lambda^m \Phi \ , \tag{2}$$

where U and Λ are commuting $K \times K$ complex symmetric matrices:

$$U\Lambda = \Lambda U; \quad U^{\mathrm{T}} = U; \quad \Lambda^{\mathrm{T}} = \Lambda;$$

Φ is a column K-dimensional vector. It can be shown by resorting to the spectral representations of U and Λ that the two assumptions (1) and (2) are equivalent [19].

Our objective is to estimate a 2D spectrum as a function of two real parameters, the proton chemical shift ω and the diffusion constant α, for instance using

$$I(\omega, \alpha) = \sum_k d_k (1 - u_k/u)^{-1} (1 - \lambda_k/\lambda)^{-1} \tag{3}$$

with $u = e^{-i\tau\omega}$ and $\lambda = e^{\beta^2 \Delta' \alpha}$, while avoiding the solution for the parameters d_k, u_k and λ_k. Note that due to Eq. 1 the double sum

$$I(\omega, \alpha) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c(n,m) u^{-n} \lambda^{-m} \tag{4}$$

is a formal Taylor expansion of $I(\omega, \alpha)$. The summation over n is the usual discrete Fourier sum which is numerically well behaved. Unfortunately, the summation over m corresponds to the divergent ILT, so this expression cannot be used directly, even in a truncated form. However, one can utilize the equivalence of the forms (2) and [19] to obtain the resolvent formula [26]

$$I(\omega, \alpha) = \Phi^{\mathrm{T}} (1 - U/u)^{-1} (1 - \Lambda/\lambda)^{-1} \Phi \tag{5}$$

Note that Eq. 3 is the most natural and simple, but not the most useful spectral representation, as it does not

lead to double-absorption lineshapes. A much more useful representation that gives *quasi-absorption* lineshapes is [26]

$$A(\omega,\alpha)=\left|\sum_k d_k(1-u_k/u)^{-2}(1-\lambda_k/\lambda)^{-2}\right|. \quad (6)$$

The only drawback of this representation is that it distorts the peak amplitudes, i.e., scales them by a factor equal to the inverse widths, so sharp peaks are overemphasized relative to the broad ones. In terms of the resolvents Eq. 6 is written as

$$A(\omega,\alpha)=\left|\Phi^{\mathrm{T}}(1-U/u)^{-2}(1-\Lambda/\lambda)^{-2}\Phi\right| \quad (7)$$

Local spectral analysis using Fourier basis

Both expressions (5) and (7) are in principle well behaved, however, neither could be used directly for spectral estimation as the matrices U and Λ and the vector Φ are unknown. Fortunately, the corresponding resolvent matrix elements can be represented in terms of the available data by choosing a suitable basis. The simplest choice is given by

$$\Phi_{n,m}:=U^n\Lambda^m\Phi \quad (n=0,..,\tilde{N}-1;\quad m=0,..,\tilde{M}-1)$$

where, assuming both N and M being even integers, we have defined:

$$\tilde{N}:=N/2;\quad \tilde{M}:=M/2.$$

Evaluated in this basis the expressions in (5) and (7) will be given solely in terms of the available data $c(n,m)$. However, this would require solution of a linear system with $\tilde{N}\tilde{M}$ equations. A more appropriate choice is given by a small Fourier basis

$$\Psi_j=\sum_{n=0}^{\tilde{N}-1}\sum_{m=0}^{\tilde{M}-1}(U/z_j)^n\Lambda^m\Phi \quad (j=1,...,K_{\mathrm{win}}) \quad (8)$$

with a set of values on the unit circle $z_j=e^{-i\tau\varphi_j}$. Such a basis effectively represents only important contributions to the resolvent $(1-U/u)^{-1}$, if $z_j\sim u$ (or $\varphi_j\sim\omega$), and $(1-\Lambda/\lambda)^{-1}$ for real values of λ. Also note that, unlike the conventional case of a 2D Fourier spectral analysis, where each dimension is treated on the same footing, here the signal does not oscillate along the diffusion dimension and so the corersponding Fourier filter is much simpler and is designed to represent only the zero frequency contributions. A further simplification exists if the set of values $z_j\equiv e^{-i\tau\varphi_j}$ are chosen as K_{win} consequitive roots of unity of the $\tilde{N}$-th order[23]:

$$z_j^{\tilde{N}}=1.$$

This corresponds to choosing an equidistant set of real numbers $\varphi_j=2(j_0+j)\pi/\tilde{N}\tau$ $(j=1,...,K_{\mathrm{win}})$ in some a priori specified frequency window. The whole frequency range of interest is then split into small overlapping windows of equal size (see, e.g., ref. [33]). The spectra in each window are computed and then summed with an appropriate weighting to account for the overlap. Note that with this construction the rank of the original signal space K (see Eq. 1) is not a parameter of the method. It may seem though that K_{win} that defines the size of the local Fourier basis is an adjusting parameter of the method. However, the results are generally independent of K_{win} as long as it is sufficiently large. In our calculations we usually set it to $K_{\mathrm{win}}=50$, which is "sufficiently large" but does but make the calculations expensive.

iRRT: Working expressions

Numerical expressions to evaluate the spectra in a window basis (8) are given by (see, e.g., ref. [28] for derivation)

$$I(\omega,\alpha) = \mathbf{C}^{\mathrm{T}}\mathbf{R}_1^{-1}\mathbf{U}_0\mathbf{R}_2^{-1}\mathbf{C}, \quad (9)$$
$$A(\omega,\alpha) = \left|\mathbf{C}^{\mathrm{T}}\mathbf{R}_1^{-1}\mathbf{U}_0\mathbf{R}_1^{-1}\mathbf{U}_0\mathbf{R}_2^{-1}\mathbf{U}_0\mathbf{R}_2^{-1}\mathbf{C}\right|. \quad (10)$$

Here the elements of the column vector $\mathbf{C}$ are

$$[\mathbf{C}]_j=\sum_{n=0}^{\tilde{N}-1}\sum_{m=0}^{\tilde{M}-1}z_j^{-n}c(n,m).$$

The $K_{\mathrm{win}}\times K_{\mathrm{win}}$ matrix pencils are defined as

$$\mathbf{R}_1=\mathbf{U}_0-\mathbf{U}_1/u;\quad \mathbf{R}_2=\mathbf{U}_0-\mathbf{U}_2/\lambda.$$

The matrix elements of $\mathbf{U}_p$ $(p=0,1,2)$ are computed using:

$$[\mathbf{U}_p]_{jj'}=\begin{cases}\dfrac{z_{j'}x_p(z_j)-z_jx_p(z_{j'})}{z_{j'}-z_j}, & \text{if } j\neq j',\\ y_p(z_j), & \text{otherwise.}\end{cases} \quad (11)$$

where

$$x_p(z) = \sum_{n=0}^{\tilde{N}-1}z^{-n}(a_p(n)-a_p(n+\tilde{N}))$$
$$y_p(z) = \sum_{n=0}^{N-2}z^{-n}a_p(n)(\tilde{N}-|\tilde{N}-n-1|)$$

and the three arrays $a_p(n)$ computed by

$$a_p(n)=\sum_{m=0}^{M-2}c(n+n_p,m+m_p)(\tilde{M}-|\tilde{M}-m-1|)$$

with $(n_0,m_0)=(0,0)$; $(n_1,m_1)=(1,0)$; and $(n_2,m_2)=(0,1)$.

Even though the points $c(N,m)$ were formally included in the summation for $x_1(z)$, evaluation of the $\mathbf{U}_p$ matrices requires the knowledge of $c(n,m)$ only for $n=0,...,N-1=2\tilde{N}-1$ and $m=0,...,M-1=2\tilde{M}-1$: the final

result for $\mathbf{U}_1$ does not actually depend on $c(N,m)$ as the corresponding contributions in Eq. 11 cancel.

Note that $\mathbf{R}_1 = \mathbf{R}_1(\omega)$ is a function of the proton frequency ω and $\mathbf{R}_2 = \mathbf{R}_2(\alpha)$, a function of the diffusion constant α. So $\mathbf{R}_1$ must be inverted at each value of ω and $\mathbf{R}_2$, at each value of α where the spectrum $A(\omega,\alpha)$ is desired. Fortunately, with the Fourier basis the matrices are small ($K_{\text{win}} \leq 100$), so the multiple inversions are not time consuming. The primary issue here is regularization of $\mathbf{R}_1$ and $\mathbf{R}_2$.

Equations 9 and 10 are new and constitute one of the main results of this paper. They look though very similar to those derived for the 2D Fourier spectral estimation using the RRT [26]. A difference is that here one of the arguments, α, formally corresponds to the imaginary frequency ($\lambda = e^{\beta^2\Delta'\alpha}$ is real and not on the unit circle). Another difference is in the way the Fourier basis is set up (Eq. 8). To emphasize these differences we called the present method iRRT. Quite remarkably, the 2D spectral estimation is done by evaluating a direct transformation of the original data, avoiding any nonlinear least squares fit. Of course, this does not mean that numerical evaluation of Eqs. 9 and 10 is always straight-forward: no matter how good the method is, the original ILT problem is ill-posed, so one cannot avoid the problems associated with instability of the solution. The later is associated here with the ill-conditioned nature of the matrices to be inverted ($\mathbf{R}_1$ and $\mathbf{R}_2$). The very special choice of the Fourier basis (see above) was made, after extensive experimentation, in order to minimize the ill-conditioning. However, the matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ still remain ill-conditioned. After numerous attempts, we have discovered that a two-step regularization scheme is necessary to remove the instabilities in the iRRT spectra.

Two-step regularization

In principle regularization of $\mathbf{R}_1$ and $\mathbf{R}_2$ could be achieved using the Tikhonov regularization [29, 26], which is significantly faster than, e.g., the singular value decomposition (SVD). The problem with the former is that the whole procedure must be repeated for each value of the regularization parameter, which is usually unknown in advance, while the SVD needs to be applied only once, the generation of the spectra at different values of the regularization parameter requiring minimal extra work. For the latter reason we have found it advantageous to use a scheme based on the SVD:

$$\mathbf{R} = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^\dagger,$$

where $\mathbf{V}$ and $\mathbf{W}$ are unitary matrices, and $\mathbf{\Sigma} = \text{diag}(\sigma_i)$ is real diagonal, $\sigma_i > 0$. After performing the SVD the inverse $\mathbf{R}^{-1}$ can be readily calculated. However, since the diagonal elements σ_i can be small and in principle zero, their reciprocals will be very large leading to unphysical results. An effective regularization replaces the true inverse $\mathbf{R}^{-1} = \mathbf{W}\mathbf{\Sigma}^{-1}\mathbf{V}^\dagger$ by a pseudo-inverse, for example, using

$$\mathbf{R}_q^{-1} = \mathbf{W}\mathbf{\Sigma}_q^{-1}\mathbf{V}^\dagger, \tag{12}$$

where $\mathbf{\Sigma}_q = \text{diag}\left(\frac{\sigma_i^2+q}{\sigma_i}\right)$ and $q > 0$ is a regularization parameter, which is small, but generally unknown. As such several spectra $A_q(\omega,\alpha)$ using different values of q are generated and then analyzed to identify the optimal q that removes most of the artifacts associated with the ill-conditioned nature of the problem but retains the true peaks. Very often there may be no "optimal" q, for instance, an increas of q may remove the correct structure together with the artifacts, so a compromize must be found. For the present problem the SVD (as well as Tikhonov) regularization alone requires too much fiddling with q to reach a reasonable compromize, but, even more importantly, one has little control on how individual peaks are being "regularized". For instance, some peaks may be arbitrarily narrow in the diffusion dimension (in fact, the peak "width" in α is an ambiguous quantity) leading to both difficult to contour and difficult to read spectra. Much better results are obtained by implementing a two step regularization. In the first step the SVD regularization is used but with a smaller value of q than usual, i.e., applying a mild regularization that does not remove the correct structures but may leave some artifacts. The second step corresponds to a frequency correlated regularization, through Lorentz-Gauss convolution:

$$A_{q\sigma}(\omega,\alpha) = \int d\omega' d\alpha' A_q(\omega',\alpha') e^{-\left(\frac{\omega-\omega'}{2\sigma_\omega}\right)^2} e^{-\left(\frac{\alpha-\alpha'}{2\sigma_\alpha}\right)^2} \tag{13}$$

where σ_ω and σ_α are adjusting parameters. This has the effect of smoothing the spectrum and removing the remaining artifacts. Also, as the smoothing is applied after the spectrum is generated, it does not decrease the numerical efficiency of the method. Because we can still tolerate some artifacts left from the SVD regularization, it is not important that the parameter q is heavily optimized and as such it is possible to set it to some small value and thus almost eliminate one adjusting parameter. Since the actual value of q must be a function of the signal norm

$$\|c\| := NM \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} |c(n,m)|,$$

it is more convenient to use the scaled parameter

$$\tilde{q} = q\|c\|^{-1} \tag{14}$$

which is not sensitive to either scaling the signal or changing its size. The typical range for $\tilde{q}$ is then between 0.1 and 0.01. Although in general the smoothing along ω may be needed, here it was not neccessary and so we set $\sigma_\omega = 0$.

This two-fold method of regularization may seem redundant and not most efficient because of the need to generate $A_q(\omega,\alpha)$ at sufficiently fine grid of α values in order to

evaluate the integral over α in (13) accurately (there is no smoothing in ω). It may be possible to incorporate both regularization steps into a single regularization. So far we have not found a method to successfully implement this.

3. EXPERIMENTAL SETUP

A challenging system of 20 μl geraniol, 30 mg camphene, and 27 mg quinine in 500 μl deuteromethanol (Fig. 1) was chosen due to the overlapping aliphatic region of these molecules and the relatively similar diffusion constants. This system was first used by Barjat *et al.* [30] in a 3D experiment. The present method can give good results for this system even using a 2D experiment. DOSY experiments were run using a 500 MHz(^{1}H) Bruker Avance DRX500, at 300K. A modified DSTE[31, 32] pulse sequence was chosen (Fig. 2), with included homospoil gradients and spin-lock pulses. The gradients were shaped to half sine-bell, resulting in $\Delta' = \Delta - (\frac{9}{8} - \frac{\pi^2}{12})\delta$ and $g = 2\gamma G\delta/\pi$, with gyromagnetic ratio γ, physical gradient strength G, delay time $\Delta = 52$ms, and gradient time $\delta = 2$ms. 8192 points were collected in the direct dimension and 8 points in the diffusion dimension.

As mentioned previously, the sampling is crucial to the processing. In order to satisfy the assumptions of the method (Eq. 1), the signal must behave as sum of complex or real exponentials in all dimensions. For DOSY, the signal in the diffusion dimension decays as:

$$\frac{S}{S_0} = e^{-D\Delta' g^2} \ , \tag{15}$$

which is gaussian if g is sampled linearly. By sampling g non-linearly one can enforce the exponential behavior that is required. In this case g was chosen according to $g = \beta\sqrt{m}$, where β is adjusted to give the appropriate magnitude of the gradients. This non-uniform sampling technique gives the required form for the resulting signal [32]. We have also found that even with a linear gradient sampling the data may still be usable for the present technique after a correctly sampled data set is constructed by interpolation. As the decay is not oscillatory in the diffusion dimension, this practice does not degrade the signal appreciably.

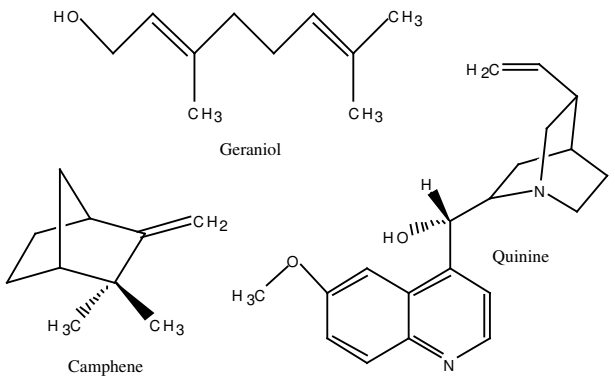

FIG. 1. Structures of camphene (136.24 Da), geraniol (154.25 Da), and quinine (324.43 Da). The sample was prepared with 30 mg camphene, 20 μl geraniol, and 27 mg quinine in 500 μl deuteromethanol.

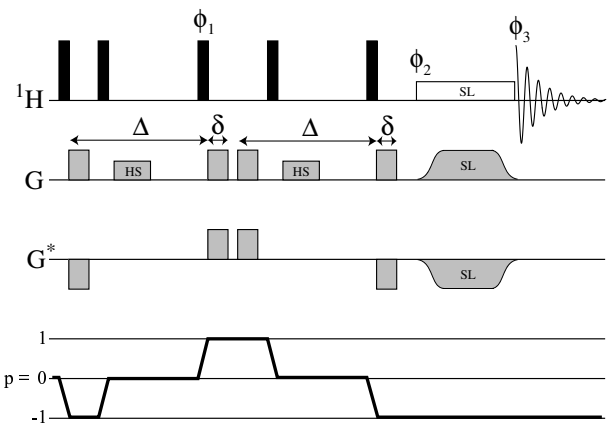

FIG. 2. DSTE pulse sequence timing diagram used for the DOSY experiments. Homospoil gradients are labeled "HS" and the spin-lock pulse is labeled "SL". All pulses have phase x except for $\phi_1(x,-x)$, $\phi_2(y)$, and $\phi_3(x,-x)$.

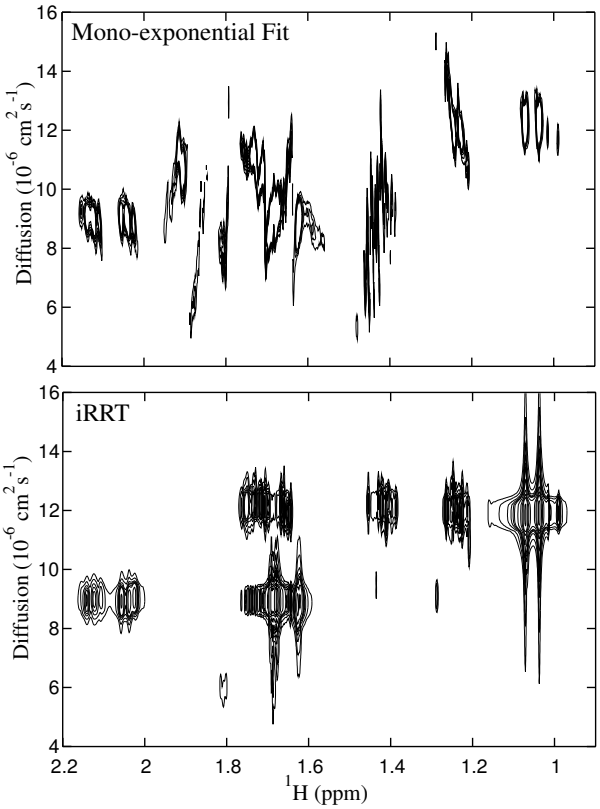

FIG. 3. Expanded aliphatic region of 2D DOSY spectra obtained from non-linear least squares mono-exponential fit and from iRRT. In this region it is difficult to use the information from the exponential fit, but the components are still resolved in the iRRT spectrum.

4. NUMERICAL RESULTS

The iRRT spectra presented here are all processed using Eqs. 10, 12 and 13 with $N = 6000$, $M = 8$, $K_{\text{win}} = 30$, $\tilde{q} = 0.1$, $\sigma_\omega = 0$ and $\sigma_\alpha = 1.6 \times 10^{-7}$.

For the comparison purposes we also process the data by the most conventional technique based on a mono-exponential fit, for which we used the Levenberg-Marquardt non-linear least squares algorithm[16]. The data is Fourier transformed in the proton dimension followed by the exponential fit for each frequency point. The slope of the log of the data is used as an initial guess for the fitting parameter, and the resulting diffusion constant, α_ω, and the standard deviation, σ_ω, are used to construct the spectrum as follows:

$$I(\omega, \alpha) = \frac{I(\omega, 0)}{\sqrt{2\pi\sigma_\omega^2}} e^{-\frac{(\alpha-\alpha_\omega)^2}{2\sigma_\omega^2}}.$$

This method results in gaussian lineshapes, whose widths reflect the uncertainty in the exponential fit, and is a standard processing technique for DOSY data [3, 4, 5]. Although a mono-exponential fit does not yield as much information as other methods can, it breaks down in a known way: it gives a peak at the weighted average of the diffusion constants of the components contributing at a single frequency. Other methods may yield cross peaks which can be mistaken for true peaks in the analysis.

In Fig. 4, we present the 2D DOSY spectra of the mixture. In well separated regions of the spectrum the mono-exponential fit is adequate, but in the crowded aliphatic region (Fig. 3), the results are extremely unstable and it is impossible to assign peaks to a particular component. In the iRRT spectrum, however, both the aliphatic protons and the remaining spectrum are well resolved. It is easy from the iRRT spectrum to assign diffusion constants to the three components: camphene $D = 12 \times 10^{-6} cm^2 s^{-1}$, geraniol $D = 9 \times 10^{-6} cm^2 s^{-1}$, and quinine $D = 6 \times 10^{-6} cm^2 s^{-1}$.

While the 2D spectrum is useful to see the spectral separation for different species and for obtaining the diffusion coefficients, it would be better to be able to make structural assignments of the individual components. This can be done by viewing the 1D "slices" of the 2D spectrum. Since there is some error in the diffusion constant of individual peaks this slice should be obtained by integrating over a small range in the diffusion dimension. This provides an average, and thus more stable 1D spectrum. The

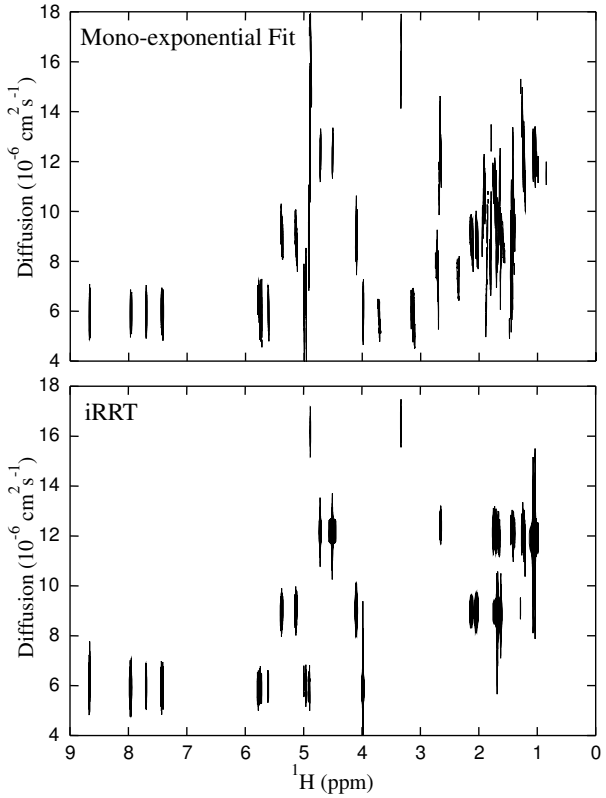

FIG. 4. Comparison of mono-exponential non-linear least squares fit with iRRT. Both methods are effective in the non-crowded region, but in the aliphatic region the fitting algorithm breaks down, giving very little usable information. The iRRT spectrum gives uniform results across the entire spectral range.

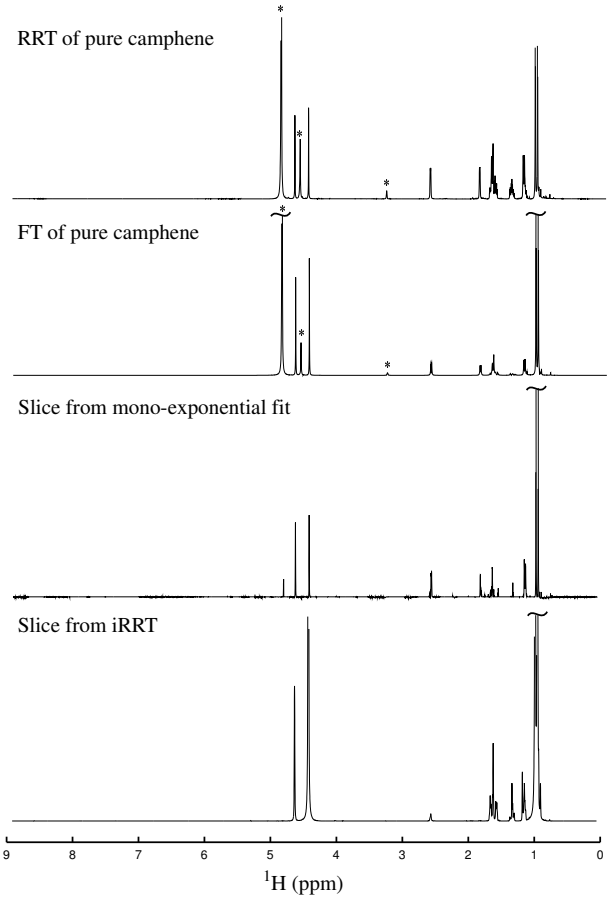

FIG. 5. 1D reconstructions for camphene. The solvent peaks in the pure spectrum are labeled with asterisks (*). All but one peak is reproduced by iRRT, but the structure in the aliphatic region is preserved much more in comparison to the exponential fit. Also note that one of the solvent peaks is still present in the exponential fit.

same integration range has been used to obtain the slices from both the non-linear least squares spectrum and the iRRT spectrum. The slices corresponding to camphene, geraniol, and quinine are presented in Figures 5, 6, and 7 respectively. Both the slices from the mono-exponential fit and iRRT are compared with the FT and RRT spectra of the pure substance. For the 1D RRT spectrum we purposely used the pseudo-absorption mode (that distorts the amplitudes) to make it more consistent with the 2D iRRT spectra. The solvent peaks, which are present in the pure spectrum are marked with an asterisk (*). For the camphene slice (Fig. 5), the iRRT is able to reproduce all but one of the resonances present in the pure substance without introducing any incorrect peaks. The mono-exponential fit is also able to produce most of the resonances, but in the crowded region the structure is almost completely lost. Also, one of the solvent peaks is not completely removed in the non-linear least squares slice. In the case of geraniol (Fig. 6), both methods reproduce all of the resonances in the pure spectrum. However, while the iRRT slice contains very few incorrect peaks, the mono-exponential fit contains numerous artifacts from other projections, and even contains solvent despite the large difference in diffusion constant between the two molecules. Finally, in the case of quinine (Fig. 7), the limitations of both methods are demonstrated. In the aliphatic region the intensity for both methods is severely reduced. This is due to the fact that the intensity is spread across two other components. Despite this, almost all of the resonances are reproduced by iRRT. In the uncrowded region of the spectrum, the iRRT reproduces the structure exactly, while the mono-exponential fit is unable to reproduce the resonances that overlap with the large solvent peak. In fact, in this situation iRRT reveals structure that is obscured by the solvent in the pure spectrum.

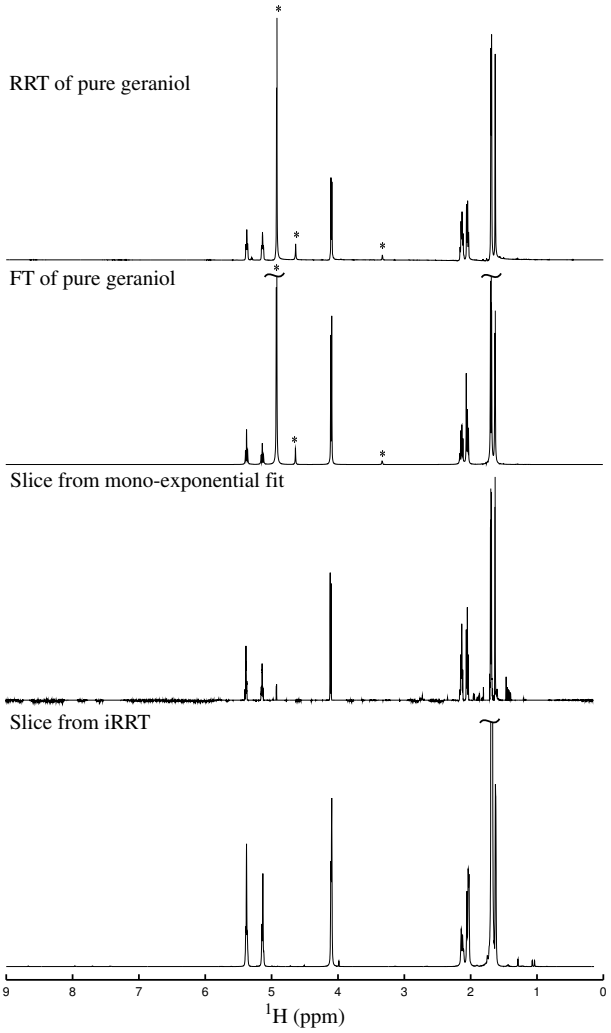

FIG. 6. 1D reconstructions for geraniol. The solvent peaks in the pure spectrum are labeled with asterisks (*). Both methods effectively resolve all the peaks, but the structure is much more prominent in the iRRT spectrum. In addition, solvent is not removed in the mono-exponential fit, and there are several peaks present that are not present in the pure spectrum.

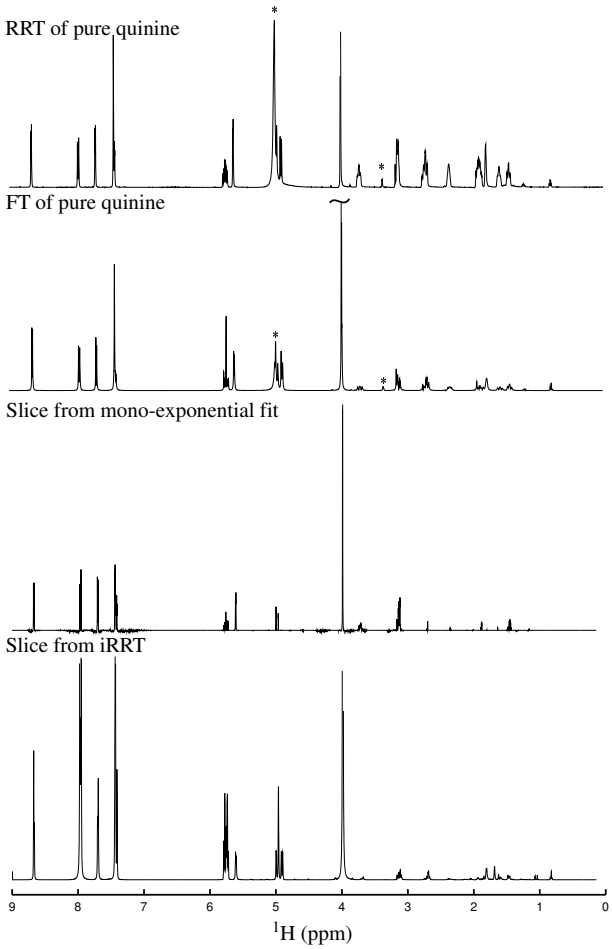

FIG. 7. 1D reconstructions for quinine. The solvent peaks in the pure spectrum are labeled with asterisks (*). Both methods suffer from reduced intensity in the aliphatic region, but the resonances are present. In this case, iRRT is able to reveal structure that was completely obscured by solvent, something that the mono-exponential fit fails at.

The limitations of both methods are well illustrated in the aliphatic region of the spectrum (Fig. 8). Neither method is able to obtain any useful information from the quinine slice due to the loss of intensity, so it is not examined further. In the case of geraniol, both methods produce the correct structure, but the mono-exponential fit contains numerous peaks that are not present in the pure spectrum. It should also be noted that both methyl peaks around 1.7 ppm are present in the iRRT spectrum, but are so intense that they are not seen using this scale. This is the disadvantage of using the pseudo-absorption representation. As this seems to primarily affect the methyl resonances, we still believe that this is the optimal representation. In the case of camphene, iRRT does a much better job of reproducing the structure present in the 1D spectrum, but the intensity of the peak at 1.9 ppm is severely reduced. This could be due to fast decay of the resonance in the time domain, leading to low intensity in the pseudo-absorption spectrum.

CONCLUSIONS

In this paper we adapted RRT for solving a 2D spectral estimation problem corresponding to the inverse Fourier transformation in the acquisition dimension, and the inverse Laplace transformation in the interferometric dimension. Due to the severely ill-conditioned nature of the problem, a two-fold regularization scheme was required to obtain good results. This involves adjusting the regularization parameter q, followed by Lorentz-Gauss convolution to smooth the remaining artifacts. By applying this method to DOSY significant improvements can be obtained over mono-exponential non-linear least squares fitting. With relative success we were also able to reproduce the 1D spectra of the pure substances from the 2D iRRT spectrum of the mixture. This can be very advantageous when trying to assign the structure of the compounds in a mixture. Because iRRT uses information from both dimensions, it is able to reliably produce spectra that have been previously unattainable without employing a 3D experiment. In our forthcomming papers the present technique will be generalized to 3D DOSY experiments.

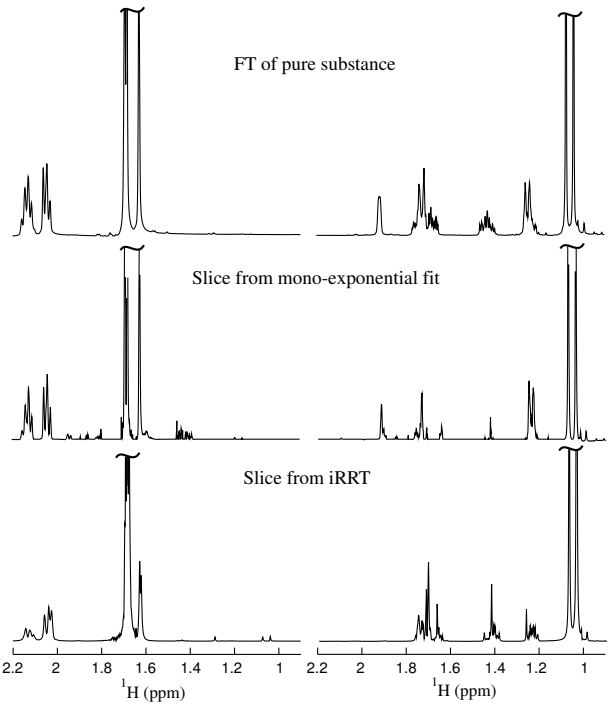

FIG. 8. Expanded aliphatic region of 1D slices for geraniol and camphene. For comparison, the FT spectra of the pure substances are also shown. Both methods the iRRT and mono-exponential fit suffer from problems in this region. Quite importantly though the iRRT spectra do not contain peaks that are not present in the pure substance.

ACKNOWLEDGMENTS.

V.A.M. acknowledges the NSF support, grant CHE-0108823. V.A.M. is Alfred P. Sloan research fellow. A.J.S. was supported by the NSF grant ... We would like to thank Dr. Jianhan Chen for his contributions to writing the original RRT code. Helpful discussions with Arnold Neumaier are also acknowledged.

REFERENCES

1. E. O. Stejskal and J. E. Tanner, spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient, *J. Chem. Phys.*, **42**, 288-292 (1965).
2. P. Stilbs, Fourier transformed pulsed-gradient spin-echo studies of molecular diffusion, *Prog. NMR Spectros.*, **19**, 1-45 (1987).
3. K. F. Morris and C. S. Johnson Jr., Diffusion-Ordered Two-Dimensional Nuclear Magnetic Resonance Spectroscopy, *J. Am. Chem. Soc.*, **114**, 3139-3141 (1992).
4. K. F. Morris and C. S. Johnson Jr., Resolution of Discrete and Continuous Molecular Size Distributions by Means of Diffusion-Ordered 2D NMR Spectroscopy, *J. Am. Chem. Soc.*, **115**, 4291-4299 (1993).
5. C. S. Johnson Jr., Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications, *Prog. Nuc. Magn. Reson. Spect.* **34**, 203-256 (1999).
6. T. L. James and G. G. McDonald, Measurement of the Self Diffusion Coefficient for Each Component in a Complex System Using Pulsed-Gradient Fourier Transform NMR, *J. Magn. Reson.*, **11**, 58-61 (1973).
7. M. F. Lin, M. J. Shapiro, and J. R. Wareing, Diffusion Edited NMR-Affinity NMR for Direct Observation of Molecular Interactions, *J. Am. Chem. Soc.*, **119**, 5249-5250 (1997).
8. M. F. Lin, M. J. Shapiro, and J. R. Wareing, Screening Mixtures by Affinity NMR, *J. Org. Chem.*, **62**, 8930-8931 (1997).
9. S. J. Gibbs and C. S. Johnson Jr., A PFG NMR Experiment for Accurate Diffusion and Flow Studies in the Presence of Eddy Currents, *J. Magn. Reson.*, **93**, 395-402 (1991).
10. D. N. Swingler, Differential Technique for the Fourier Transform Processing of Multicomponent Exponential Functions, *IEEE Trans. Biomed. Eng.*, **BME-24**, 408-410 (1977).
11. S. W. Provencher, An eigenfunction expansion method for the analysis of exponential decay curves, *J. Chem. Phys.*, **64**, 2772-2777 (1976).

12. S. W. Provencher and R. H. Vogel, Regularization Techniques for Inverse Problems in Molecular Biology, in "Numerical Treatment of Inverse Problems in Differential and Integral Equations" (P. Deuflhard and E. Hairer, Eds.), pp. 304-319, Birkhäuser, Boston (1983).

13. M. A. Delsuc, T. E. Malliavin, Maximum Entropy Processing of DOSY NMR Spectra, *Anal. Chem.*, **70**, 2146-2148 (1998).

14. W. Windig and B. Antalek, Direct Exponential Curve Resolution Algorithm (DECRA): A novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying contribution profiles, *Chemom. Intell. Lab. Syst.*, **37** 241-254 (1997).

15. P. Stilbs, K. Paulsen, P. C. Griffiths, Global Least-Squares Analysis of Large, Correlated Spectral Data Sets: Application to Component Resolved FT-PGSE NMR Spectroscopy, *J. Phys. Chem.*, **100**, 8180-8189 (1996).

16. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C, Second Edition," Cambridge University Press, Cambridge, (1992).

17. 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).

18. V. A. Mandelshtam, and H. S. Taylor, Harmonic inversion of time signals and its applications, *J. Chem. Phys.* **107**, 6756-6769 (1997).

19. V. A. Mandelshtam, and H. S. Taylor, Multidimensional harmonic inversion by filter-diagonalization, *J. Chem. Phys.* **108**, 9970-9977 (1998).

20. M.R. Wall, T. Dieckmann, J. Feigon, D. Neuhauser, Two-dimensional filter-diagonalization: spectral inversion of 2D NMR time-correlation signals including degeneracies, *Chem. Phys. Lett.* **291**, 465-470 (1998).

21. 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).

22. A. Neumaier, Solving ill-conditioned and singular linear systems: A tutorial on regularization, *SIAM Review* **40**, 636-666 (1998)

23. A. Neumaier, Harmonic inversion problem, preprint, unpublished.

24. V. A. Mandelshtam, The multidimensional filter diagonalization method. I. Theory and numerical implementation, *J. Magn. Reson.* **144**, 343-356 (2000).

25. A. A. De Angelis, H. Hu, V. A. Mandelshtam, and A.J. Shaka, The multidimensional filter diagonalization method. II. Applications to 2D, 3D and 4D NMR experiments, *J. Magn. Reson.* **144**, 357-366 (2000).

26. J. Chen, A.J. Shaka and V. A. Mandelshtam, RRT: The regularized resolvent transform for high resolution spectral estimation, *J. Magn. Reson.* **147**, 129-137 , (2000).

27. J. Chen, V. A. Mandelshtam and A.J. Shaka, Regularization of the filter diagonalization method: FDM2k, *J. Magn. Reson.* **146**, 363-368, (2000).

28. V. A. Mandelshtam, FDM: the filter diagonalization method for data processing in NMR experiments, *Prog. Nuc. Magn. Reson. Spect.* **38**, 159-196, (2001).

29. A. Tikhonov, Solution of incorrectly formulated problems and the regularization method, *Soviet Math. Dokl.* **4**, 1035-1038 (1963); A. Tikhonov and V. Arsenin, Solutions of ill-posed problems. Winston and Sons, Washington (1977).

30. H. Barjat, G.A. Morris and A.G. Swanson, A Three-Dimensional DOSY-HMQC Experiment for the High-Resolution Analysis of Complex Mixtures, *J. Magn. Reson.* **131**, 131-138 (1998).

31. A. Jershow and N. Müller, Suppression of Convection Artifacts in Stimulated-Echo Diffusion Experiments. Double-Stimulated-Echo Experiments., *J. Magn. Reson.*, **125**, 372-375 (1997).

32. J. E. Curtis, Spectral Processing Methods to Extract Imaginary Frequencies from Nuclear Magnetic Resonance Signals, PhD Thesis, University of California, Irvine, 2001.

33. 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).