Chapter 9

Filter Diagonalization Methods for Time-Domain Signals

A. J. Shaka and Vladimir A. Mandelshtam

Chemistry Department, University of California at Irvine, 4134 Natural Sciences Building 1, Mail Code: 2025, Irvine, CA 92697, USA

9.1 INTRODUCTION

The *filter diagonalization method* (FDM) is a linear algebraic method of spectral analysis of time signals and, in particular, NMR signals, that is meant to complement or replace Fourier Transform (FT) spectral analysis. One-dimensional (1D) FDM and its multidimensional (mD) extensions exist. Related techniques that have emerged from the FDM are the *regularized resolvent transform* (RRT) and *extended Fourier transform* (XFT). Most results reported here can be found in Refs. 1–18. We attempt to present FDM critically, emphasizing both its advantages and drawbacks. FDM is not always advantageous, but for many kinds of spectra it can be. Some unsolved challenges, either conceptual or computational, associated mostly with the mD FDM, are identified. The first sections lay the mathematical underpinnings, while the latter sections show applications and highlight experimental considerations.

9.1.1 Important Notations

A linear operator is identified by a cap: $\widehat{A}$. In the expression $|b) = \widehat{A}|c)$ a vector $|c)$ from a linear space is mapped to a vector $|b)$ from the same space. $(c|$ defines a dual vector to $|c)$ and $(c|b) = (b|c)$, complex symmetric (i.e., no complex conjugation) inner product between the two vectors $|c)$ and $|b)$. Bold characters, such as $\boldsymbol{A}$ or $\boldsymbol{C}$, are used for matrix representations of linear operators or vectors. Their elements are then defined using the following notations: $[\boldsymbol{A}]_{mn}$ or $[\boldsymbol{C}]_n$. $\boldsymbol{A}^{\mathrm{T}}$ is a transpose of matrix $\boldsymbol{A}$, while $\boldsymbol{A}^{\dagger}$ is its adjoint (transposed and complex conjugated) matrix. Almost all variables are *complex* numbers, including ω_k, d_k, etc., although time variables t, τ, etc., are real.

9.2 FORMULATION OF THE PROBLEMS

9.2.1 1D Spectral Analysis

Consider a finite complex valued 1D time signal $c_n = c(n\tau)$, $(n = 0, \ldots, N-1)$, that has been sampled instantaneously on an equidistant discrete set of time points. By the *Fourier spectral estimation* problem we mean estimating the infinite time discrete Fourier transformation (DFT) of c_n:

$$I(s) := \sum_{n=0^+}^{\infty} c_n z^{-n} := \sum_{n=0}^{\infty} \left(1 - \frac{\delta_{n0}}{2}\right) c_n z^{-n} \quad (9.1)$$

using the finite data set. Here $z = e^{-i\tau s}$; the "shortcut" $\Sigma_{n=0^+}$, meaning that the first term in the sum is multiplied by 1/2, will be used throughout the chapter. The latter corrects an error due to the discretization of the half-line Fourier integral

$$I_0(s) := \int_0^{\infty} c(t) e^{its} \mathrm{d}t = \lim_{\tau \to 0} \tau I(s) \quad (9.2)$$

The 1D *parameter estimation* or *harmonic inversion problem* (HIP), by contrast, corresponds to the *fitting* of the finite data set c_n by the form

$$c_n = \sum_{k=1}^{K} d_k u_k^n \quad (9.3)$$

with $u_k = e^{-i\tau\omega_k}$. The unknown parameters are the complex poles u_k (or, alternatively, the frequencies ω_k) and complex amplitudes d_k. It is not obvious that a linear algebraic solution of the HIP exists as at first glance it requires a nonlinear least squares fit.

Once a *line list* $\{u_k, d_k\}$ is obtained, one can construct an *ersatz spectrum*

$$I(s) = \sum_k d_k \left\{ \frac{z}{z - u_k} - \frac{1}{2} \right\} \quad (9.4)$$

which estimates the DFT spectrum (9.1). Note that, although the infinite Fourier series (9.1) converges only if all poles u_k satisfy $|u_k| \leq 1$, a sensible and stable result can be obtained using equation (9.4), even when this is not the case. Moreover, the stability is maintained only if one does not exclude or modify the "unphysical" entries with u_k far outside the unit circle. It is recommended though to flip the poles with $|u_k| > 1$ which are close to the unit circle, i.e., $|u_k| \sim 1$, using $u_k \to u_k/|u_k|^2$ before inserting into equation (9.4). This eliminates sign problems caused by inaccurate width estimation for very narrow peaks when using short data records.

Note also that, if the sampling theorem is satisfied, there is no significant difference in the appearance of the DFT spectrum (9.1) and the half-line FT integral (9.2), except for very broad lines (i.e., with line widths $|\mathrm{Im}\ \omega_k|$ comparable to the spectral width $\mathrm{SW} = 1/\tau$). The dispersion-mode wings of these lines, in particular, alias in the DFT case. Practically speaking the form (9.4) is significantly more stable than its integral counterpart[16]

$$I_0(s) = \sum_k \frac{i d_k}{s - \omega_k} \quad (9.5)$$

and is what is used in practice.

9.2.2 mD Spectral Analysis

Consider a general complex valued mD time signal $c_{n_1, \ldots, n_\mathrm{D}} := c(n_1\tau_1, \ldots, n_\mathrm{D}\tau_\mathrm{D})$, defined on an equidistant rectangular time grid of size $N_\mathrm{total} = N_1 \times \cdots \times N_\mathrm{D}$.

The mD Fourier spectral estimation problem then becomes that of estimating the infinite sum,

$$I(s_1, \ldots, s_\mathrm{D}) = \sum_{n_1=0^+}^{\infty} z_1^{-n_1} \cdots \sum_{n_\mathrm{D}=0^+}^{\infty} z_\mathrm{D}^{-n_\mathrm{D}} c_{n_1, \ldots, n_\mathrm{D}} \quad (9.6)$$

with $z_l = e^{-in_l\tau_l s_l}$, $(l = 1, \ldots, \mathrm{D})$.

One possible and, probably, most desirable formulation of the mD HIP is

$$c_{n_1, \ldots, n_\mathrm{D}} = \sum_{k=1}^{K} d_k u_{1k}^{n_1} \cdots u_{\mathrm{D}k}^{n_\mathrm{D}} \quad (9.7)$$

where the unknown parameters are the complex amplitudes d_k and frequencies ω_{lk} with $u_{lk} = e^{-in_l\tau_l\omega_{lk}}$, $(l = 1, \ldots, \mathrm{D})$. These parameters, once computed, can be used to construct various representations of the mD spectrum, such as that defined by equation (9.6),

$$I(s_1, \ldots, s_\mathrm{D}) = \sum_k d_k \prod_{l=1}^{\mathrm{D}} \left\{ \frac{z_l}{z_l - u_{lk}} - \frac{1}{2} \right\} \quad (9.8)$$

or a more useful mD absorption-mode spectrum

$$A(s_1, \ldots, s_\mathrm{D}) = \mathrm{Re} \sum_k d_k \prod_{l=1}^{\mathrm{D}} \mathrm{Re} \left\{ \frac{z_l}{z_l - u_{lk}} - \frac{1}{2} \right\} \quad (9.9)$$

Unfortunately, the present versions of FDM can provide a sensible mD line list only when the data well approximates the form equation (9.7) with high signal-to-noise ratio (SNR). We will refer to such cases as the *well defined HIP*. However, for a poorly defined mD HIP various spectral representations, including absorption-mode spectra, can still be constructed in the FDM framework, but by carefully avoiding the solution of equation (9.7).

9.3 PARAMETER AND SPECTRAL ESTIMATION BY 1D FDM

The commonly used starting point in most linear algebraic approaches is the autoregression (AR) assumption, which in the 1D case reads

$$c_n = \sum_{p=1}^{K} a_p c_{n-p} \tag{9.10}$$

This assumption is equivalent to equation (9.3), but requires that one solve for the auxiliary prediction coefficients a_p. These are not, however, the ultimate quantities of interest. To recast the spectral analysis problem into the linear algebraic framework in FDM, one assumes the existence of an auxiliary quantum dynamical system with a complex symmetric evolution operator $\widehat{U}$ of rank K and a fictitious state Φ, so that the data may be represented in the form of the autocorrelation function[1]

$$c_n = (\Phi|\widehat{U}^n|\Phi) \tag{9.11}$$

It can be shown that the three assumptions (9.3), (9.10) and (9.11) are equivalent, although the latter leads directly to the solution of either HIP or the spectral estimation problem and is therefore more economical. For example, assuming $\widehat{U}$ is diagonalizable,

$$\widehat{U}\Upsilon_k = u_k \Upsilon_k, \quad (k = 1, \ldots, K) \tag{9.12}$$

we get equation (9.3) with

$$d_k = (\Phi|\Upsilon_k)^2 (\Upsilon_k|\Upsilon_k)^{-1} \tag{9.13}$$

which suggests that the HIP is equivalent to a matrix eigenvalue problem arising from equation (9.12). This can be shown by using $N = 2M$ and implementing the Krylov basis

$$\Phi_n := \widehat{U}^n \Phi, \quad (n = 0, \ldots, M-1)$$

We can now define the evolution operator matrix $\boldsymbol{U}_1$, the overlap matrix $\boldsymbol{U}_0$ and the signal vector $\boldsymbol{C}$ and evaluate them in terms of the signal points c_n utilizing the assumption (9.11):

$$[\boldsymbol{U}_p]_{nm} := (\Phi_n|\widehat{U}^p|\Phi_m) = c_{n+m+p}$$
$$[\boldsymbol{C}]_n := (\Phi|\Phi_n) = c_n$$

By expanding the eigenfunctions, $\Upsilon_k = \Sigma_k [\boldsymbol{B}_k]_n \Phi_n$, we obtain the generalized eigenvalue problem

$$\boldsymbol{U}_1 \boldsymbol{B}_k = u_k \boldsymbol{U}_0 \boldsymbol{B}_k \tag{9.14}$$

for the unknown eigenvalues u_k and eigenvectors $\boldsymbol{B}_k$. If $M \geq K$, with exact arithmetic equation (9.14) then yields the exact solution of the HIP (9.3) with

$$d_k = \left(\boldsymbol{C}^{\mathrm{T}} \boldsymbol{B}_k\right)^2 \left(\boldsymbol{B}_k^{\mathrm{T}} \boldsymbol{U}_0 \boldsymbol{B}_k\right)^{-1} \tag{9.15}$$

The latter result is derived from equation (9.13). This and similar expressions have been first derived in the FDM framework.[1,2] Note though that equation (9.14) was used previously[19] to estimate the parameters u_k.

Interestingly, once the data arrays $\boldsymbol{U}_0$, $\boldsymbol{U}_1$ and $\boldsymbol{C}$ are obtained in a suitable representation, the DFT spectrum $I(s)$ can be estimated *directly*, avoiding the use of the spectral parameters u_k and d_k and hence the solution of the generalized eigenvalue problem. This is done by first inserting equation (9.11) into equation (9.1) and evaluating the geometric series,

$$I(s) = (\Phi| \left(\frac{1 - \widehat{U}}{z} \right)^{-1} |\Phi) - \frac{c_0}{2}$$

and then using the above matrix representations

$$I(s) = \boldsymbol{C}^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{C} - \frac{c_0}{2}, \quad \boldsymbol{R} = \frac{\boldsymbol{U}_0 - \boldsymbol{U}_1}{z} \tag{9.16}$$

In the noiseless case for $M = K$ the spectral estimate (9.16) is exact. For $M \geq K$ the matrix pencil $\boldsymbol{R}$ is singular and the evaluation of the resolvent $\boldsymbol{R}^{-1}$ requires a *regularization*. For example, one can implement the *singular value decomposition* (SVD) of $\boldsymbol{R}$ and then evaluate the inverse in the range subspace of $\boldsymbol{R}$. This still gives the exact result for the noiseless case. In practice $\boldsymbol{R}$ is never exactly singular, but could be ill-conditioned, so that regularization is most definitely required. In this case truncated SVD (as described above) is not recommended. Instead, a rather more robust regularization scheme can be implemented,[14,17] such as the Tikhonov regularization,[20] that requires solution of a

Hermitian linear least squares problem at each value of frequency s:

$$I(s) \approx \boldsymbol{C}^{\mathrm{T}}\boldsymbol{X} - \frac{c_0}{2}, \quad \left(\boldsymbol{R}^{\dagger}\boldsymbol{R} + q^2\right)\boldsymbol{X} = \boldsymbol{R}^{\dagger}\boldsymbol{C} \quad (9.17)$$

where q is a real regularization parameter that controls the noise and artifact suppression. Bigger values of q lower the resolution, but lead to more uniform spectral estimates, rather like line broadening in FT analysis. The spectral estimate (9.17) was derived in Ref. 14 and named RRT (Regularized Resolvent Transform). In a related technique XFT the spectrum is estimated by partitioning the infinite DFT sum (9.1) into a finite DFT over the measured data length and a correction term, which is evaluated using an RRT-type expression.[17]

9.3.1 Fourier Subspace Filtering

Solution of the $M \times M$ generalized eigenvalue problem (9.14) or the matrix inversion in equation (9.17) has $\sim M^3$ numerical scaling and is thus impractical. A Fourier basis has proven to be very efficient for such problems.[1,2] A small Fourier subspace $\{\tilde{\Phi}_j\}$ of size $K_{\mathrm{win}} \ll M$ is constructed by choosing K_{win} complex numbers on the unit circle $y_j = e^{-i\tau\varphi_j}$, $(j = 1, \ldots, K_{\mathrm{win}})$ and using

$$\tilde{\Phi}_j = \sum_{n=0}^{M-1} \left(\frac{\widehat{U}}{y_j}\right)^n \Phi$$

where the tilde over vectors or matrices will denote the use of a Fourier basis. The results are not usually sensitive to the parameters of the basis as long as the points φ_j are chosen appropriately. In most cases an equidistant grid is effective, with spacing

$$\Delta\varphi = \frac{2\pi}{\aleph M \tau} \quad (9.18)$$

and adjustable parameter $\aleph \geq 1$ (e.g., one can use $\aleph = 1.2$). Since the basis is localized in the frequency domain, the diagonal elements of the matrices will dominate, and the off-diagonal elements will drop off as a sinc function. This allows an accurate solution of equation (9.14) or the spectral estimation in equation (9.17) in a small subspace, leading to numerically inexpensive linear algebra. A further gain may be achieved by using a multiscale Fourier basis[11] with a nonuniform set of φ_j values, containing information about the entire spectrum rather than just the spectral window. Note that filtering the *basis* is conceptually different from e.g., digital filtering of the *signal*. That is, the true signal is used to construct matrix elements, and then this matrix, which is potentially huge, is treated in a basis which allows diagonalization to proceed in a block-diagonal fashion, so that only a relatively small number of matrix elements need to be computed.

The matrix elements of the $K_{\mathrm{win}} \times K_{\mathrm{win}}$ square matrices $\tilde{\boldsymbol{U}}_1$ and $\tilde{\boldsymbol{U}}_0$ can be computed using[2]

$$\begin{aligned}\left[\tilde{\boldsymbol{U}}_p\right]_{jj'} &:= (\tilde{\Phi}_j | \widehat{U}^p | \tilde{\Phi}_{j'}) \\ &= \widehat{S} \sum_{\sigma=0,1} \frac{(-1)^\sigma (y_j / y_{j'})^M}{1 - y_j / y_{j'}} \\ &\quad \times \sum_{n=\sigma M}^{(\sigma+1)(M-1)} y_j^{-n} c(n+p)\end{aligned}$$

where $\widehat{S}$ defines symmetrization operator over the variables y_j and $y_{j'}$,

$$\widehat{S} g(y_j, y_{j'}) = g(y_j, y_{j'}) + g(y_{j'}, y_j)$$

For $j = j'$ we have

$$\left[\tilde{\boldsymbol{U}}_p\right]_{jj} = \sum_{n=0}^{2M-2} y_j^{-n} (M - |M - n - 1|) c(n+p)$$

The state vector Φ in the Fourier basis is a $1 \times K_{\mathrm{win}}$ column vector with elements

$$[\tilde{\boldsymbol{C}}]_j := (\tilde{\Phi}_j | \Phi) = \sum_{n=0}^{M-1} y_j^{-n} c_n$$

The generalized eigenvalue problem (9.14) in the Fourier basis now reads

$$\tilde{\boldsymbol{U}}_1 \tilde{\boldsymbol{B}}_k = u_k \tilde{\boldsymbol{U}}_0 \tilde{\boldsymbol{B}}_k \quad (9.19)$$

with the coefficients computed by

$$d_k = \left(\tilde{\boldsymbol{C}}^{\mathrm{T}} \tilde{\boldsymbol{B}}_k\right)^2 \left(\tilde{\boldsymbol{B}}_k^{\mathrm{T}} \tilde{\boldsymbol{U}}_0 \tilde{\boldsymbol{B}}_k\right)^{-1} \quad (9.20)$$

Defining the $K_{\mathrm{win}} \times K_{\mathrm{win}}$ matrix-pencil $\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{U}}_0 - \tilde{\boldsymbol{U}}_1 / z$ in the Fourier basis, the RRT spectral estimate becomes

$$I(s) \approx \tilde{\boldsymbol{C}}^{\mathrm{T}} \tilde{\boldsymbol{X}} - \frac{c_0}{2}, \quad \left(\tilde{\boldsymbol{R}}^{\dagger} \tilde{\boldsymbol{R}} + q^2\right) \tilde{\boldsymbol{X}} = \tilde{\boldsymbol{R}}^{\dagger} \tilde{\boldsymbol{C}} \quad (9.21)$$

Since a small subspace, rather than the complete basis $\{\Phi_n\}$, is used, equations (9.19), (9.20) and (9.21) must be viewed as an approximation to equations

(9.14), (9.15) and (9.17) with convergence parameter K_{win}. However, an acceptable convergence is usually achieved for sufficiently small sizes in the range $K_{\text{win}} \sim 10-100$. Moreover, in the Fourier basis, the matrices are much less ill-conditioned, which makes them easier to handle numerically.

9.4 mD FDM

Generalizing the 1D quantum ansatz (9.11) the mD signal is assumed to be generated by D commuting complex symmetric evolution operators $\widehat{U}_l$, ($l = 1, \ldots, \text{D}$):

$$c_{n_1,\ldots,n_\text{D}} = \left(\Phi \left| \prod_{l=1}^{\text{D}} \widehat{U}_l^{n_l} \right| \Phi \right) \tag{9.22}$$

Clearly, if all the operators $\widehat{U}_l$ have the same rank K with nondegenerate eigenvalues u_{lk} and the corresponding eigenfunctions Υ_k, equation (9.22) yields equation (9.7) with the amplitudes $d_k = (\Phi|\Upsilon_k)^2(\Upsilon_k|\Upsilon_k)^{-1}$. However, the assumption about the existence of a simultaneous basis $\{\Upsilon_k\}$ leads to numerical troubles.[8] It turns out that it can be avoided by keeping only the requirement that the $\widehat{U}_l$ commute with each other. That is, we will label the eigenfunctions of $\widehat{U}_l$ by l:

$$\widehat{U}_l \Upsilon_{lk} = u_{lk} \Upsilon_{lk}, \quad (l = 1, \ldots, \text{D}) \tag{9.23}$$

By inserting the quantum ansatz (9.22) into equation (9.6) and evaluating the product of D geometric sums, in analogy with equation (9.16), we get the resolvent representation of the mD DFT spectrum:

$$I(s_1, \ldots, s_\text{D}) = \left(\Phi \left| \prod_{l=1}^{\text{D}} \left\{ \frac{z_l}{z_l - \widehat{U}_l} - \frac{1}{2} \right\} \right| \Phi \right) \tag{9.24}$$

Now using $2M_l = N_l$ for $l = 1, \ldots, \text{D}$ and representing everything in the Krylov basis $\Phi_{n_1,\ldots,n_\text{D}} := \widehat{U}_1^{n_1}, \ldots, U_\text{D}^{n_\text{D}} \Phi$ with total size $M_{\text{total}} = M_1 \times \cdots \times M_\text{D}$, we can replace equation (9.23) by generalized eigenvalue problems with known data matrices $\boldsymbol{U}_l$ or, alternatively, obtain an mD RRT spectral estimate. However, as in the 1D case, it is advantageous to implement a Fourier basis. For the sake of simplicity, here we consider only the D = 2 case.

9.4.1 Fourier Basis: 2D Case

Let $\{\varphi_{1j}, \varphi_{2j}\}$ be a set of $K_{1\text{win}} \times K_{2\text{win}} = K_{\text{win}}$ grid points in a chosen 2D frequency window with spacings, given by equation (9.18), in each dimension. We can now define the Fourier basis using:

$$\tilde{\Phi}_j = \sum_{n_1=0}^{M_1-1} \sum_{n_2=0}^{M_2-1} \left(\frac{\widehat{U}_1}{y_{1j}} \right)^{n_1} \left(\frac{\widehat{U}_2}{y_{2j}} \right)^{n_2} \Phi$$

with $y_{lj} = e^{-in\tau_l\varphi_{lj}}$, ($l = 1, 2$). The square complex symmetric matrices of $\widehat{U}_1$, $\widehat{U}_2$ and the identity operator (the overlap matrix) in this basis are defined by $\tilde{\boldsymbol{U}}_1$, $\tilde{\boldsymbol{U}}_2$ and $\tilde{\boldsymbol{U}}_0$, accordingly. To consolidate the expressions, three new data sets are defined:

$$c^{(0)}_{n_1,n_2} := c_{n_1,n_2}; \quad c^{(1)}_{n_1,n_2} := c_{n_1+1,n_2};$$
$$c^{(2)}_{n_1,n_2} := c_{n_1,n_2+1}$$

The matrix elements of the U-matrices for both $y_{1j} \neq y_{1j'}$ and $y_{2j} \neq y_{2j'}$ can be computed by[3]

$$[\boldsymbol{U}_l]_{jj'} = \widehat{S}_1 \sum_{\sigma_1=0,1} \frac{(-1)^{\sigma_1}(y_{1j}/y_{1j'})^{M_1}}{1 - y_{1j}/y_{1j'}} \times \widehat{S}_2 \sum_{\sigma_2=0,1} \frac{(-1)^{\sigma_2}(y_{2j}/y_{2j'})^{M_2}}{1 - y_{2j}/y_{2j'}} \times \sum_{n_1=\sigma_1 M_1}^{(\sigma_1+1)(M_1-1)} \sum_{n_2=\sigma_2 M_2}^{(\sigma_2+1)(M_2-1)} y_{1j}^{-n_1} y_{2j}^{-n_2} c^{(l)}_{n_1 n_2}$$

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

$$\widehat{S}_1 g(y_{1j}, y_{1j'}) = g(y_{1j}, y_{1j'}) + g(y_{1j'}, y_{1j})$$

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

$$[\boldsymbol{U}_l]_{jj'} = \widehat{S}_2 \sum_{\sigma_2=0,1} \frac{(-1)^{\sigma_2}(y_{2j}/y_{2j'})^{M_2}}{1 - y_{2j}/y_{2j'}} \times \sum_{n_1=0}^{2M_1-2} \sum_{n_2=\sigma_2 M_2}^{(\sigma_2+1)(M_2-1)} y_{1j}^{-n_1} y_{2j}^{-n_2} c^{(l)}_{n_1 n_2} \times (M_1 - |M_1 - n_1 - 1|)$$

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

the U-matrices, we have

$$[\boldsymbol{U}_l]_{jj} = \sum_{n_1=0}^{2M_1-2}\sum_{n_2=0}^{2M_2-2} y_{1j}^{-n_1} y_{2j}^{-n_2} c_{n_1 n_2}^{(l)} \times (M_1 - |M_1 - n_1 - 1|) \times (M_2 - |M_2 - n_2 - 1|)$$

We will also need the $1 \times K_{\text{win}}$ column vector:

$$[\tilde{\boldsymbol{C}}]_j = \sum_{n_1=0}^{M_1-1}\sum_{n_2=0}^{M_2-1} y_{1j}^{-n_1} y_{2j}^{-n_2} c_{n_1 n_2}$$

9.4.2 Regularization of the mD FDM: FDM2K

Armed with the above expressions, we can replace the eigenvalue problems for the effective evolution operators (9.23) by generalized eigenvalue problems with known matrices and unknown eigenvalues u_{lk} and eigenvectors $\boldsymbol{B}_{lk}$:

$$\tilde{\boldsymbol{U}}_l \tilde{\boldsymbol{B}}_{lk} = u_{lk} \tilde{\boldsymbol{U}}_0 \tilde{\boldsymbol{B}}_{lk}, \quad (l = 1, \ldots, \mathrm{D}) \tag{9.25}$$

Unfortunately, unlike the 1D case, unless we are dealing with a noiseless well-defined mD HIP, a straightforward numerical solution of these equations gives meaningless results. A meaningful solution may be obtained by implementing a truncated SVD in which one defines the effective range space of $\boldsymbol{U}_0$, e.g., by keeping the largest P singular values, and then reevaluates the U-matrices in this subspace. However, this type of regularization, as in the RRT case, is usually quite sensitive to the choice of P and requires a lot of fiddling around. We have discovered empirically[15] a much more robust regularization procedure, in which equation (9.25) is replaced by

$$\tilde{\boldsymbol{U}}_0^{\dagger} \tilde{\boldsymbol{U}}_l \tilde{\boldsymbol{B}}_{lk} = u_{lk} \left(\tilde{\boldsymbol{U}}_0^{\dagger} \tilde{\boldsymbol{U}}_0 + q^2 \right) \tilde{\boldsymbol{B}}_{lk}$$

with regularization parameter q. The effect of the regularization is similar to that in RRT in that the results depend fairly smoothly on q, eliminating the need for extensive fiddling. Even though expression (9.20) to compute the coefficients holds, in principle, in the mD case, it is very tricky to use and is avoided here. A more correct approach is to use the eigenvalues u_{lk} and eigenvectors B_{lk} to construct the mD spectrum:[10,16]

$$I(s_1, \ldots, s_{\mathrm{D}}) = \sum_{k_1,\ldots,k_{\mathrm{D}}} d_{k_1,\ldots,k_{\mathrm{D}}} \prod_{l=1}^{\mathrm{D}} \left\{ \frac{z_l}{z_l - u_{lk_l}} - \frac{1}{2} \right\} \tag{9.26}$$

with

$$d_{k_1,\ldots,k_{\mathrm{D}}} = \frac{\tilde{\boldsymbol{C}}^{\mathrm{T}} \tilde{\boldsymbol{B}}_{1k_1} \tilde{\boldsymbol{B}}_{1k_1}^{\mathrm{T}}}{\tilde{\boldsymbol{B}}_{1k_1}^{\mathrm{T}} \tilde{\boldsymbol{U}}_0 \tilde{\boldsymbol{B}}_{1k_1}} \left(\prod_{l=2}^{\mathrm{D}} \frac{\tilde{\boldsymbol{U}}_0 \tilde{\boldsymbol{B}}_{lk_l} \tilde{\boldsymbol{B}}_{lk_l}^{\mathrm{T}}}{\tilde{\boldsymbol{B}}_{lk_l}^{\mathrm{T}} \tilde{\boldsymbol{U}}_0 \tilde{\boldsymbol{B}}_{lk_l}} \right) \boldsymbol{C}$$

Clearly, equation (9.26) has somewhat different structure from the form of equation (9.8) as each combination $(u_{1k_1}, \ldots, u_{\mathrm{D}k_{\mathrm{D}}})$ gives a peak with the amplitude $d_{k_1}, \ldots, k_{\mathrm{D}}$, i.e., many more than one would expect. This is the price for avoiding the use of a simultaneous eigenbasis in equation (9.23). Numerically, though, most cross-amplitudes $d_{k_1}, \ldots, k_{\mathrm{D}}$ will vanish. Expression (9.26) appears much more stable and easier to use numerically than equation (9.8) in the case of degenerate eigenvalues u_{lk} or low SNR.

By analogy with equation (9.9) an absorption-mode mD spectrum can be constructed, even from a single phase-modulated data set, by replacing the complex (phase-twist) Lorentzians by the absorption-mode lineshapes in equation (9.26):

$$A(s_1, \ldots, s_{\mathrm{D}}) = \mathrm{Re} \sum_{k_1,\ldots,k_{\mathrm{D}}} d_{k_1,\ldots,k_{\mathrm{D}}} \prod_{l=1}^{\mathrm{D}} \mathrm{Re} \left\{ \frac{z_l}{z_l - u_{lk_l}} - \frac{1}{2} \right\}$$

Of course the FT cannot construct an absorption-mode spectrum unless matched pairs of data sets are available (either N- and P- or sine and cosine) and in some cases this may save considerable instrument time, particularly with 3D or 4D spectra.

9.4.3 2D Spectral Estimation by RRT

As in the 1D case, the spectral estimation can be done directly by evaluating the resolvent matrix element in equation (9.24). For simplicity here we present only the 2D expression. Let $\tilde{\boldsymbol{R}}_l = \tilde{\boldsymbol{U}}_0 - \tilde{\boldsymbol{U}}_l / z_l$, $(l = 1, 2)$. The 2D RRT spectral estimate then becomes

$$I(s_1, s_2) \approx \tilde{\boldsymbol{X}}_1^{\mathrm{T}} \tilde{\boldsymbol{U}}_0 \tilde{\boldsymbol{X}}_2 - \tilde{\boldsymbol{C}}^{\mathrm{T}} \frac{(\tilde{\boldsymbol{X}}_1 + \tilde{\boldsymbol{X}}_2)}{2} + \frac{c_{00}}{4} \tag{9.27}$$

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

$$\left(\tilde{\boldsymbol{R}}_l^{\dagger}\tilde{\boldsymbol{R}}_l + q^2\right)\tilde{\boldsymbol{X}}_l = \tilde{\boldsymbol{R}}_l^{\dagger}\tilde{\boldsymbol{C}} \tag{9.28}$$

Note that the total number of the linear systems (9.28) 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. In practice, RRT is very fast indeed.

When only the complex DFT spectrum (9.6) is of interest, RRT is, arguably, the most consistent approach. Once again, as in the 1D case, using exact arithmetic for the noiseless case when $N_{\text{total}} \geq 4K$ (i.e., for the well defined mD HIP), the estimate (9.27) is essentially exact. In practice, however, there are several factors that affect the convergence; the regularization parameter q being an adjusting knob to choose between the high-resolution regime ($q \sim 0$), and noisy spectrum, and the low resolution regime (large q) with strong suppression of the artifacts.

Even though various spectral representations can be produced by RRT, unfortunately, it is difficult to compute a double-absorption spectrum by the 2D RRT using only a single purely phase modulated data set. Because a direct line list is not produced in RRT its scope is more limited. Thus FDM with its potential ability to generate a direct line list seems much more appealing.

9.5 PRACTICAL CONSIDERATIONS

The foregoing mathematical development, while of interest to those with training in alternative methods to the FT, will not be transparent to a typical experimentalist primarily interested in the question "What's in it for me?" In the following, a more pictorial and qualitative exposition will be used.

The first thing to keep in mind is that FDM is most useful when the data is likely to fit the model of discrete Lorentzian lines (9.7), and when the time-frequency uncertainty principle is limiting the FT resolution substantially. Very noisy data does not fit the model, as noise cannot be encapsulated by any parametric description with a limited number of parameters. Lines in solids are usually not purely Lorentzian, which may limit the application of FDM. Likewise, data that has been sampled essentially well into the noise cannot magically be processed to extract small splittings well within the instrumental linewidth. When using FDM the key question is this: If the data could be fully sampled in all dimensions, would the spectrum look like a discrete number of Lorentzians? If it would, then the same result can be obtained by FDM with much shorter data records. But if it would not, then FDM may not be useful. The presence of t_1-noise, which is definitely non-Lorentzian, must also be kept in mind. Generally speaking, FDM is a method to use on high-quality spectra that have been recorded expertly, using the best available technology. It is not a method to clean up bad data sets.

FDM is particularly advantageous in so-called Constant Time (CT) experiments, in which a fixed time, denoted $2T$, is used for chemical shift encoding by changing the position of a 180° refocusing pulse. These experiments are often employed in the study of uniformly labeled proteins. The interferograms from CT-HSQC experiments show essentially no decay in the indirect dimension(s) and so are, by definition, transform limited. The full-width at half-maximum (FWHM) of the sinc function, resulting from transformation of an unapodized CT signal is $0.6034/T$. Apodization by a cosine function, $\cos(\pi t/2T)$, reduces sidelobe amplitude to less than 5% of the main peak, but increases the FWHM to $0.8197/T$. Thus the transform-limited resolution is about 63 Hz, or 0.5 ppm at 125 MHz ^{13}C frequency, for a typical CT experiment with $2T = 26$ ms. For a hypothetical CT experiment with $2T = 4.7$ ms, lines no narrower than 348 Hz can be expected. To do better than this, some alternative to the FT must be used. Note that simply increasing $2T$ is not usually possible, as catastrophic loss of signal by transverse relaxation can ensue, especially for the important C_α resonances.

It might be tempting to assume that the FDM, as it fits the lines including their widths, could give essentially infinite resolution in a CT-HSQC experiment: after all, the lines should have nearly zero width in F_1. Very narrow lines can indeed be obtained, *but only if there are not too many of them, too closely spaced*. In other words, FDM cannot give infinite resolution, but only resolution commensurate with the local density of basis functions. This resolution is, as described above, dependent roughly on the *product* of acquisition times, so that some information in the acquisition dimension can be used to improve that in the shorter dimension, provided the signal does not decay too quickly.

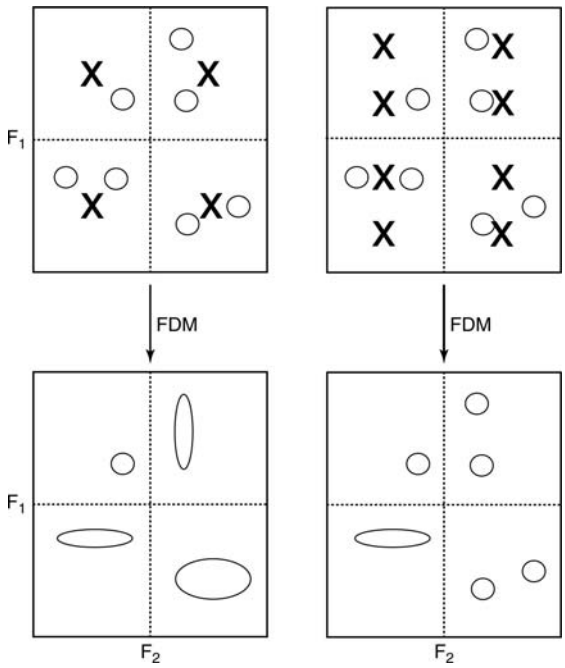

Figure 9.1. A schematic representation of how FDM treats a local spectral region. At the top, peaks are represented with circles and Fourier basis functions with crosses. At the bottom is the expected result after diagonalization. The left-hand panels show a situation in which four basis functions are available to try to fit the seven genuine peaks. A "best effort" fit leaves only one line converged, with the other three pairs fit as broader singlets. In the right-hand panels, eight basis functions are available because twice as much data has been sampled in the vertical dimension. Five of the seven peaks can be resolved, but the pair in the lower left-hand quadrant cannot, because they are degenerate in the vertical dimension. (Doubling the sampling in the horizontal dimension would, of course, resolve the two peaks.)

The situation is shown in Figure 9.1, with closed shapes showing hypothetical peak positions, and crosses showing the position of 2D Fourier basis functions. The top panels represent the initial matrix construction, and the lower panels are the FDM output. On the left panels, the local density of functions is four, but there are seven peaks present. Diagonalization does not change the size of the basis, so only four peaks can be used to fit seven. The result is shown in the lower panel, with three broad (and unstable) features masquerading as three sets of two lines, and one converged feature. In the right panel, the basis density is increased to eight by taking twice as many points in the vertical dimension. Now the vertical pair in the top right quadrant is resolved completely, as is the pair in the lower right quadrant, in which the F_1 frequencies are close, but not identical. However, the pair in the lower left quadrant is not resolved. The F_1 frequencies are nearly identical, so that increasing the sampling along F_1 does not help much: even though there are two basis functions in the vicinity, the two lines cannot be resolved. This kind of local convergence, in which more crowded spectral regions are the last to "settle down", is completely different from the slow but uniform convergence familiar from FT analysis. This behavior can be rationalized within a linear algebraic context. If two lines have equal frequencies (and linewidths), e.g., along F_1 (or if both of these are very close in practical terms), then $\widehat{U}_1$ will have two degenerate eigenvalues. In order to resolve these two peaks in the F_2 dimension, one needs a basis of at least two linearly-independent vectors containing the corresponding two-dimensional eigensubspace of $\widehat{U}_1$. However, those cannot be constructed using, for example, Krylov vectors $\boldsymbol{\Phi}_{n_1,n_2} = \widehat{U}_1^{n_1}\widehat{U}_2^{n_2}\boldsymbol{\Phi}$ with some fixed values of n_2 and any number of different n_1 values. One needs at least two vectors $\boldsymbol{\Phi}_{n_1,n_2}$ with different n_2. In practice, these considerations must be applied locally in the frequency domain, i.e., using the Fourier basis (the crosses), rather than Krylov basis, and have less obvious consequences because of several other factors. For instance, what happens to the "unused" basis functions? They end up fitting residual noise or non-Lorentzian components in the skirts of the peaks. How far can an initial basis function "move"? That depends on the SNR. The higher it is, the more freedom the crosses acquire to "lock on" to a signal peak. It is thus important to emphasize that the ideal kind of spectrum for FDM analysis is one in which the features are a fixed number of randomly distributed sharp Lorentzians that are well above the noise floor. This kind of signal can always be dominated by the available basis functions as more data is acquired. High noise levels effectively rob the algorithm of basis functions by pinning down many of them to fit noise locally. Spectra like COSY, in which many lines are absolutely degenerate, and there are locally dense regions within cross peak multiplets, cannot benefit much, if at all, from FDM. That is, it will still require many increments to resolve a COSY spectrum properly and, as new weak cross peaks

continuously emerge when the evolution time is incremented, convergence may never be achieved at all.

9.6 SELECTED APPLICATIONS

9.6.1 2D Constant-Time HSQC Spectra

With these cautions in mind, the results for protein NMR can be startlingly good. The three figures (Figures 9.2–9.4), show the C_α–H_α region of a CT-HSQC spectrum of uniformly labeled human ubiquitin (1 mM, 5 mm tube, 500 MHz) using a very short 4.7 ms CT. The FT spectrum shows the expected broad peaks in F_1, in accordance with the time-frequency uncertainty principle. Using Mirror-Image Linear Prediction in F_1[18] and FT in F_2 improves matters somewhat, but the spectrum is a distant second to the FDM2K result, in which nearly every resonance is fully resolved except for the predictable nonconvergence of the most congested region. The peaks are genuine, as can be verified by acquiring a much longer CT data set.[18]

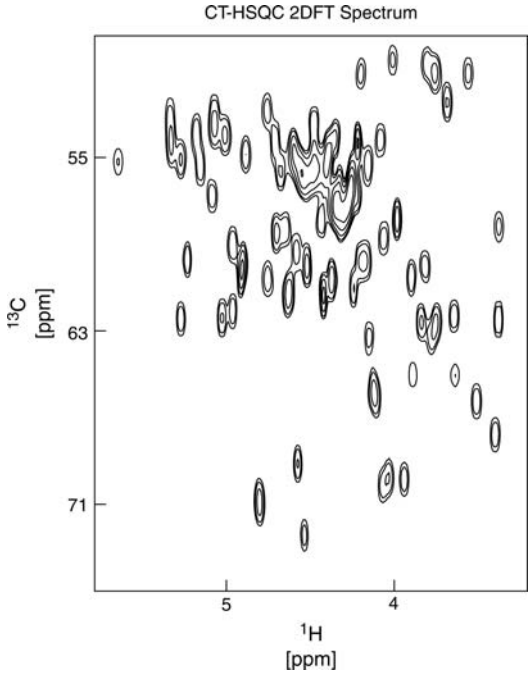

Figure 9.2. A small portion of the CT-HSQC spectrum of human ubiquitin (1 mM solution, 500 MHz proton frequency) obtained with a very short CT period of 4.7 ms and processed by 2D FT. The C_α–H_α region is displayed. The broadening in the vertical dimension is primarily due to the time-frequency uncertainty principle, so that, while sensitivity is quite good, resolution is sufficient to make only a small fraction of the assignments.

9.6.2 45° Projections of 2D *J* Spectra

The unique ability to create a pseudo-absorption spectrum from a single phase-twist FT spectrum is quite useful when there is only a single data set, as in 2D J-spectroscopy (see Chapter 11).[5,9,10] In this case, the improved resolution can be used to construct a proton-decoupled proton spectrum by computing the phase-sensitive 45° projection. This

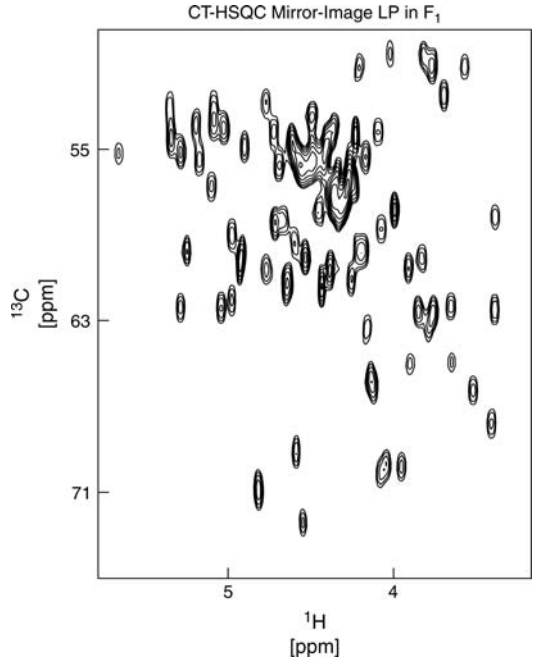

Figure 9.3. The same data set as in Figure 9.2, using mirror-image Linear Prediction in the vertical dimension to first extend the data length by a factor of two, followed by FT processing. Resolution is better than the 2D FT, but not as good as 2D FDM, as can be clearly seen by examining the clusters of peaks in the lower part of the figures.

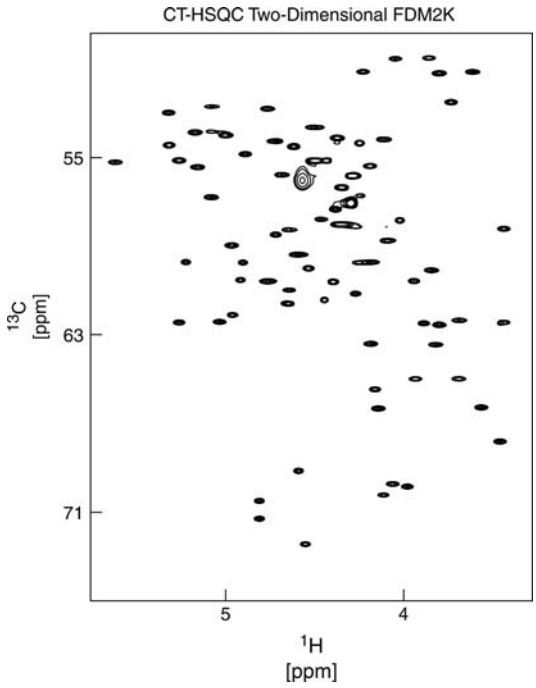

Figure 9.4. The same data set as in Figure 9.2, but processed by FDM2K. No FT was used to construct the spectrum: it came directly from the FDM2K line list. To improve the appearance of the spectrum (cosmetically) Gaussian line shapes have been substituted for the Lorentzian lines that are the natural form of the FDM spectrum. Resolution is dramatically enhanced. The spectrum is completely converged except for the most crowded region, which shows the kind of nonconvergence described when more local features are present than the number of local Fourier basis functions available to FDM.

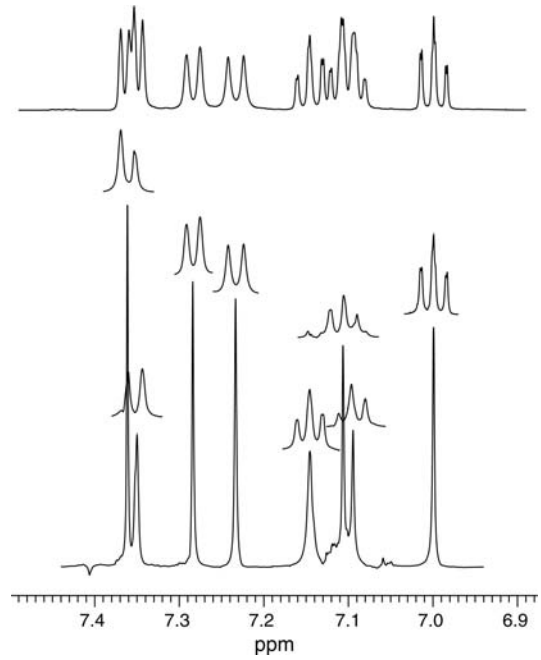

Figure 9.5. An example of direct calculation of a 1D 45° projection (the lowest trace) and multiplet cross sections by FDM. The upper trace corresponds to a conventional ^{1}H spectrum. The spectra were obtained using a purely phase modulated 2D-*J* signal of ditryptophan tripeptide[9] with just $N_1 = 4$ points along the *J*-dimension. $N_{\mathrm{FDM}} = 20$ calculations, with different number of points along the running time dimension in the range $N_2 = 11\,000$–$12\,000$, were added together to obtain artifact free spectra. Only a small part of the spectrum is shown, while the two spectral widths were $SW_1 = 80\,\mathrm{Hz}$ and $SW_2 = 8\,\mathrm{kHz}$.

projection vanishes in the FT spectrum because of the phase-twist line shape. Furthermore, this decoupling can be applied to other mD spectra, to condense proton multiplets into sharp singlets, thereby improving the resolution substantially. This combination of spectroscopic simplification with alternative methods of analysis is opening up new possibilities for future research. An example of a 1D 45° projection of a 2D *J* data set, with only a few increments in the *J* dimension, is shown in Figure 9.5. The results are quite encouraging, although more work needs to be done to make the method a turn-key experiment. Proper regularization of these projections is an area of current research.

9.7 REMAINING PROBLEMS

Both 1D FDM and RRT (its offspring) are essentially developed and well tested techniques that are generally as reliable as the FFT, sufficiently fast, and can often deliver resolution beyond the FT uncertainty relation if the data can be well represented by Lorentzians and is not very noisy.

FDM provides one with an effective evolution operator $\widehat{U}$ with eigenvalues and eigenvectors directly related to the spectral parameters. However, the difficulties associated with the construction of a meaningful line list for data of poor quality (i.e., not characterized by equation (9.3)) exist. These difficulties

are not associated with the lack of a reliable algorithm for selecting the "genuine" poles and throwing away the "noise" poles from the full list of complex eigenvalues of $\widehat{U}$, but are rather conceptual, and caused by the intrinsic ambiguity of the line list for a general data set that, a priori, does not fit any particular parametric form.

The mD case is considerably more difficult. For example, the multidimensional spectrum cannot generally be constructed from the multidimensional line list, as the latter can be very hard to obtain. Fortunately, various spectra can be obtained by avoiding the line list construction and using the resolvent expressions. The resolvent operator formalism appears to be very convenient as it allows one to construct various types of spectra, including absorption-mode spectra (nontrivial spectral projections discussed above). The main computational problem associated with the implementation of the resolvent formulas is that one typically deals with very ill-conditioned matrices, causing the spectrum to be very unstable with respect to both the FDM parameters, and small variations of the input data. Thus, unlike the 1D case, there are major problems to be solved in the multidimensional versions of both FDM and RRT. For instance, for a typical 2D NMR data set, even with relatively high SNR, one has difficulties in constructing two adequate commuting effective evolution operators $\widehat{U}_1$ and $\widehat{U}_2$ describing the 2D signal. This, in turn, makes it difficult to construct an adequate 2D line list corresponding to equation (9.7). We believe that this problem is, as in the 1D case, a consequence of the ill-defined nature of equation (9.7), although the additional requirement that the operators $\widehat{U}_l$ commute with each other makes the problem much worse than in 1D. Clearly, the key issue in mD FDM is to find a general computationally inexpensive and robust procedure that could be applied to regularize the FDM equations. At the present stage the problem is only partially solved. For example, the regularization procedure in RRT is very efficient, but, unfortunately, it can be used only as a spectral estimator and cannot easily overcome some of the DFT limitations. Regularization of RRT is not directly extendable to FDM, although several methods have already been developed, such as truncated SVD of $\boldsymbol{U}_0$ or FDM averaging.[10,12,16] Unfortunately, the former is not generic, while the latter is computationally very expensive and cannot be used to construct a multidimensional line list. Currently, the most promising regularization technique seems to be that associated with FDM2K.[15] Its implementation is computationally very simple and inexpensive, although, at this stage, it is not clear how general and reliable it is for various types of NMR data.

To conclude, the main avenues for the future research seem to be the following.

1. Given the matrix representations of the evolution operators, $\widehat{U}_1$, $\widehat{U}_2$, etc., in a non-orthonormal basis, which are possibly very ill-conditioned, find a fast and reliable method of evaluating various resolvents associated with these operators.
2. Given the matrix representations of the evolution operators, construct a set of commuting effective Hamiltonians $\widehat{\Omega}_1$, $\widehat{\Omega}_2$, etc., with eigenvalues and eigenvectors yielding the line list.

Both goals are associated with finding reliable and computationally inexpensive regularization techniques.

REFERENCES

1. M. R. Wall and D. Neuhauser, *J. Chem. Phys.*, 1995, **102**, 8011.
2. V. A. Mandelshtam and H. S. Taylor, *J. Chem. Phys.*, 1997, **107**, 6756.
3. V. A. Mandelshtam and H. S. Taylor, *J. Chem. Phys.*, 1998, **108**, 9970.
4. J. W. Pang, T. Dieckmann, J. Feigon, and D. Neuhauser, *J. Chem. Phys.*, 1998, **108**, 8360.
5. V. A. Mandelshtam, H. S. Taylor, and A. J. Shaka, *J. Magn. Reson.*, 1998, **133**, 304.
6. V. A. Mandelshtam, H. Hu, and A. J. Shaka, *Magn. Res. Chem.*, 1998, **36**, S17.
7. H. Hu, Q. N. Van, V. A. Mandelshtam, and A. J. Shaka, *J. Magn. Reson.*, 1998, **134**, 76.
8. M. R. Wall, T. Dieckmann, J. Feigon, and D. Neuhauser, *Chem. Phys. Lett.*, 1998, **291**, 465.
9. V. A. Mandelshtam, Q. N. Van, and A. J. Shaka, *J. Am. Chem. Soc.*, 1998, **120**, 12161.
10. V. A. Mandelshtam, N. D. Taylor, H. Hu, M. Smith, and A. J. Shaka, *Chem. Phys. Lett.*, 1999, **305**, 209.
11. J. Chen and V. A. Mandelshtam, *J. Chem. Phys.*, 2000, **112**, 4429.

12. V. A. Mandelshtam, *J. Magn. Reson.*, 2000, **144**, 343.

13. A. A. De Angelis, H. Hu, V. A. Mandelshtam, and A. J. Shaka, *J. Magn. Reson.*, 2000, **144**, 357.

14. J. Chen, A. J. Shaka, and V. A. Mandelshtam, *J. Magn. Reson.*, 2000, **147**, 129.

15. J. Chen, V. A. Mandelshtam, and A. J. Shaka, *J. Magn. Reson.*, 2000, **146**, 363.

16. V. A. Mandelshtam, *Prog. NMR Spectosc.*, 2001, **38**, 159.

17. G. S. Armstrong and V. A. Mandelshtam, *J. Magn. Reson.*, 2001, **153**, 22.

18. A. A. De Angelis, J. Chen, V. A. Mandelshtam, and A. J. Shaka, *J. Biomol. NMR*, 2003, **162**, 74.

19. Y. Hua and T. K. Sarkar, *IEEE Trans. ASSP*, 1990, **38**(5), 814.

20. A. Tikhonov, *Soviet Math. Dokl.*, 1963, **4**, 1035; A. Tikhonov and V. Arsenin, *Solutions of ill-posed problems*, Winston and Sons, Washington, 1977.