UNIVERSITY OF CALIFORNIA,
IRVINE

# Nonlinear Methods for High Resolution Spectral Analysis and Their Applications in Nuclear Magnetic Resonance Experiments

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry

by

Jianhan Chen

Dissertation Committee:
Professor Vladimir A. Mandelshtam, Chair
Professor Craig C. Martens
Professor Steven R. White

2002

The dissertation of Jianhan Chen
is approved and is acceptable in quality
and form for publication on microfilm:

Committee Chair

University of California, Irvine
2002

To my grandparents, my parents,

and particularly, my wife, Xueying

**献给我的祖父母，父母，以及我的妻子。**

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGEMENTS

First, I would like to thank my thesis advisors, Professors Vladimir A. Mandelshtam and A. J. Shaka, for their support, encouragement, and inspiration during the last four years. Particularly, I thank Professor Mandelshtam for his patience, and the many long hours he spent talking with me about the research projects. I thank both advisors for being great mentors. It has been a great honor to be their student.

Special thanks go to Professors V. Ara Apkarian, Craig C. Martens and Douglas J. Tobias for for their enthusiastic support and encouragement. I am also grateful to Professors P. Taborek, S. R. White, D. A. Brant, and R. B. Gerber for their support during my study in UC-Irvine. Dr. Anna A. De Angelis and Dr. Daniel Nietlispach are specially acknowledged for great collaborations and helpful discussions.

Also many thanks to current and former group members of Shaka and Mandelshtam groups: Dr. Haitao Hu, Dr. Que N. Van, Dr. Anna A. De Angelis, Dr. Joseph E. Curtis, Dr. Mari A. Smith. Geoff S. Armstrong, and Kristin E. Cano. I am especially indebted to Dr. I-Feng William Kuo for helping to make my experience with computers enjoyable, to Dr. Joseph E. Curtis for providing an excellent template for this manuscript, and to Geoff S. Armstrong for proofreading this manuscript.

Most importantly, I would like to thank my family for their love and support during the whole journey of coming to US to pursue this degree. I am especially indebted to my wife, Xueying Qin, for her love, sacrifice and companionship.

Finally, I would like to thank the Department of Chemistry at UC-Irvine for the supports that make my study in USA possible.

# CURRICULUM VITAE

# JIANHAN CHEN

## Education

2002 Ph.D. in Chemistry, University of California at Irvine
Concentration: Chemical and Materials Physics
Research Advisors: V. A. Mandelshtam and A. J. Shaka
Dissertation title: "Nonlinear Methods for High-Resolution Spectral Analysis and Their Applications in Nuclear Magnetic Resonance Experiments".

2000 M.S. in Chemistry, University of California at Irvine
Research Advisors: V. A. Mandelshtam and A. J. Shaka
Dissertation title: "Filter Diagonalization Method as a High Resolution Spectral Estimator".

1998 B.Sc. in Chemistry, University of Science & Technology of China
Research Advisors: Kaibin Tang and Yitai Qian
Thesis title: "Synthesis and Characterization of New 123-Series Cuprates".

## Experience

1999-2002 Graduate Research Assistant, University of California at Irvine

1996-1998 Reasearch Assistant, University of Science & Technology of China

## Awards

2002 The E. K. C. Lee Fellowship

2002 Regents' Dissertation Fellowship

2001 American Chemical Society Travel Fellowship

2001 Physical Science Faculty Endowed Fellowship

2001 42$^{nd}$ ENC Student Travel Award

1998 Lian Xiang Scholarship

1993 Xu A'Quong Scholarship

## Publications

1. J. Chen and V. A. Mandelshtam, "Multi-Scale Filter Diagonalization Method for spectral analysis of noisy data with nonlocalized features", J. Chem. Phys. **112**, 4429 (2000).
2. J. Chen, V. A. Mandelshtam and A. J. Shaka, "Regularization of the Filter Diagonalization Method: FDM2K", J. Magn. Reson. **146**, 368 (2000).

3. J. Chen, A. J. Shaka, V. A. Mandelshtam, "RRT: The Regularized Resolvent Transform for high resolution spectral estimation", J. Magn. Reson. **147**, 129 (2000).
4. A. A. De Angelis, J. Chen, V. A. Mandelshtam and A. J. Shaka, "A Method to Obtain High Resolution 13C CT-HSQC Spectra of Proteins with a Short Constant-Time Period", J. Biomol. NMR (submitted).
5. J. Chen, A. A. De Angelis, V. A. Mandelshtam and A. J. Shaka, "Progress on Two-Dimensional Filter Diagonalization Method. An Efficient Doubling Scheme for Two-Dimensional Constant-Time NMR", J. Mag. Reson. (submitted).
6. M. A. Smith, J. Chen, and A. J. Shaka, "Small Molecule NOE Difference Spectroscopy for High-Resolution NMR", J. Am. Chem. Soc. (submitted).
7. J. Chen, V. Mandelshtam and D. Neuhauser, "Calculating excited vibrational levels of polyatomic molecules by diffusion Monte Carlo", J. Chem. Phys., in preparation.
8. J. Chen, D. Nietlispach, V. A. Mandelshtam, and A. J. Shaka, "Ultra-high quality HNCO spectra with very short constant times", in preparation.

## Selected Presentations

1. "A faster Filter Diagonalization Method with fully automated multi-windowing", Poster Presentation, 41st Experimental Nuclear Magnetic Resonance Conference, April 2000.
2. "RRT: The Regularized Resolvent Transform for high resolution spectral estimation", Poster Presentation, 42nd Experimental Nuclear Magnetic Resonance Conference, March 2001.
3. "The Regularized Resolvent Transform and its Applications", Invited Talk, ACS Annual Meeting, August 2001.
4. "The Filter Diagonalization Method for Spectral Analysis of NMR Data", Invited Talk, Washington Area NMR Group, February 2002.
5. "High Resolution Double Constant-Time 3D NMR Spectra by Filter Diagonalization Method", Poster Presentation, 43nd Experimental Nuclear Magnetic Resonance Conference, April 2002.

## ABSTRACT OF THE DISSERTATION


Nonlinear Methods for High Resolution Spectral Analysis and Their Applications in Nuclear Magnetic Resonance Experiments

by

Jianhan Chen

Doctor of Philosophy in Chemistry

University of California at Irvine, 2002

Professor Vladimir A. Mandelshtam, Chair

Novel nonlinear methods for high resolution spectral analysis of time domain signals, namely, the Filter Diagonalization Method (FDM) and Regularized Resolvent Transform (RRT), have been developed and applied to the signal processing problems in Nuclear Magnetic Resonance (NMR) experiments. Several important breakthroughs that significantly improve the efficiency and stability of these algorithms are discussed. They include *multi-scale* Fourier-type basis, regularization of multi-dimensional FDM, and a doubling scheme for Constant-Time NMR signals. Applied to several 2D and 3D protein NMR experiments, substantial resolution enhancement is demonstrated. Several qualitative and semi-quantitative ways of studying the reliability and sensitivity of FDM and RRT are discussed and various issues specific to NMR spectral analysis are also addressed.

# Chapter 1

# Introduction to Spectral Analysis

## 1.1 Definition and Significance

Spectral analysis is the process of obtaining the spectral information, such as characteristic frequencies, amplitudes and phases, from one or more time-varying signals. Spectral analysis plays an important role in many fields of science and engineering. Time domains signals are often acquired in many modern experiments including Nuclear Magnetic Resonance (NMR), Ion Cyclotron Resonance (ICR) and others. Time signals also occur in numerical simulations of dynamics systems such as in unimolecular reactions [1], dissociation of vibrationally excited van der Waals complexes [2], and many other chemical processes [3]. Spectral analysis is the key to revealing the information encoded in these signals. The goal of spectral analysis is to extract maximum information from the available time domain data, or, to minimize the required experimental or computational time for extracting the same amount of information. The conventional Fourier Transform (FT) [4] has played an essential role in spectral analysis and was crucial to the development of many modern spectroscopies including NMR, optical and mass spectrometry [5]. FT is a linear transformation that is stable, reliable and fast if implemented as the Fast Fourier Transform (FFT) [6, 7]. In addition, FT does not make any assumption about the functional form of the signal and can be applied to processing any type of signals. Accordingly, FT is a very powerful method and has been widely used for spectral analysis in many fields.

However, FT also has some well-known limitations. The FT resolution is limited by the FT time-frequency uncertainty principle. Truncation of the signal leads to sinc-like oscillative artifacts in the baseline of the FT spectrum. Another limitation of FT is that addition information about the signal, if exists, can not be efficiently incorporated and utilized. Furthermore, multi-dimensional FT is simply a series of 1D FT applied to each dimension independently. Thus FT is essentially a 1D method for spectral analysis and can not utilize the important information that evolutions in all dimensions are actually correlated. Therefore, spectral analysis remains an active research area [8]. A lot of efforts have been devoted to developing various alternative methods to the FT spectral analysis, seeking to make use of some *a priori* knowledge and to extract information more efficiently from the time signals.

The purpose of this chapter is to introduce the existing methods for spectral analysis, with mainly the application of NMR spectroscopy in mind. We will first discuss the basic properties and major limitations of the conventional FT spectral analysis in Section 1.2, then briefly review some existing alternative methods in Section 1.3. Finally, some conclusions and remarks are given in Section 1.4.

## 1.2 Fourier Transform Spectral Analysis

A physical process can be either described by a time domain function, $c(t)$, where $t$ is the *time* variable in *seconds*, or a frequency domain function, $I(\omega)$, where $\omega$ is the *angular frequency* in *radians per second.* To be general, we consider both functions being complex. These two functions form a "Fourier pair" as they are connected by the *Fourier transform* equations:

$$\begin{aligned} I(\omega) &= \int_{-\infty}^{\infty} c(t)\, e^{i\omega t}\, dt\,, \\ c(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} I(\omega)\, e^{-i\omega t}\, d\omega\,. \end{aligned} \tag{1.1}$$

$I(\omega)$ is also called a *spectrum*, or a *spectral representation* of the time signal. The transformation from $c(t)$ to $I(\omega)$ is often called the (forward) Fourier transform, and its inverse is often called the inverse Fourier transform.

From Eq. 1.1, it is obvious that the Fourier transformations are linear operations. There is no information gain or lost by transforming from one representation to the other: the two functions contain exactly the same information but in different formats. The important difference is that features that are delocalized in one function become localized in the other one. Interestingly, this becomes very important when it comes to interpreting the information encoded in these functions. In a typical time signal, contributions from different constituents, or resonances, are highly delocalized and strongly interfere with each other, making direct interpretation of the signal very difficult. For example, Figure 1.1 plots two simple time signals: one with single resonance, trace (a), and the other one with two resonances of equal intensities and decay rates, trace (b). When there is only a single resonance, it is possible to estimate

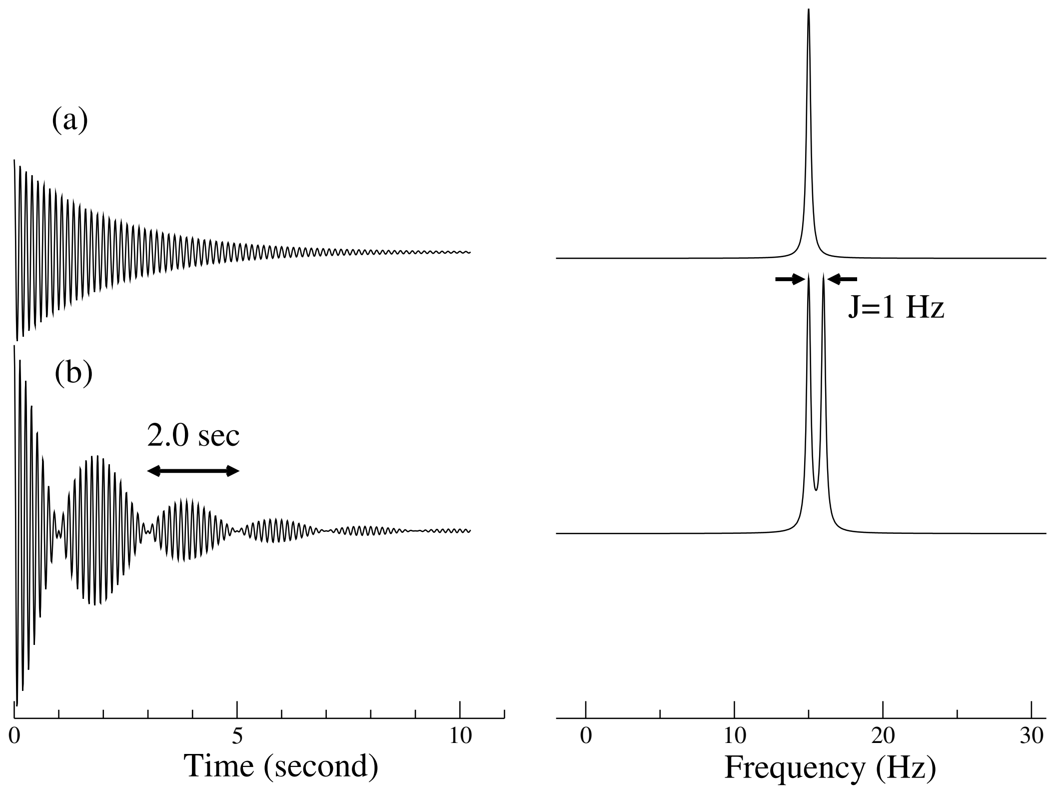



Figure 1.1: Two simple time signals and corresponding Fourier transform spectra (only real parts of the complex functions are shown). While contributions from different components are delocalized and interfere with each other in the time domain, they are highly localized and separated according to the characteristic frequencies in the frequency domain.

the frequency by counting the zero-crossing and the linewidth from the decay rate [8]. When there are two components with equal intensity present, it is possible to using the "beating", a consequence of interference, to estimate the separation of two frequencies. But it becomes difficult to tell much about the characteristics by looking at the time signal, other than that one or more components are narrow if the signal decays slowly, and other vague judgments. When more than two components with different intensities and decaying rates are present, it becomes impossible to interpret the time domain signals directly. On the contrary, features are highly localized and separated according to the characteristic frequencies in the frequency domain, making the interpretation of them much easier.

Now it is clear that spectral analysis, the process of obtaining a spectral representation from the time domain signal, is a key step in making the information interpretable, even though such a transformation does not magically increase the information content by any means. FT is a natural choice for performing such a transformation. However, as it will become clear later, it is not the only way, neither the optimal one, for obtaining the spectrum in many cases. One of the main focuses of this dissertation is to exploit alternative methods for extracting more (interpretable) information from the available data set, when the signal satisfies certain known constraints.

### 1.2.1 Fourier Transform of Discrete Signals

In the most common situations, the time signal $c(t)$ is sampled (i.e., its value is recorded) on an equidistant time grid, $t_n = n\tau = 0, \tau, \ldots, (N-1)\tau$, resulting in a discrete signal $c(n) \equiv c(n\tau)$. The interval $\tau$ is also called *sampling rate*, or *dwell time* in NMR terminology. The fact that the signal is discrete instead of continuous has profound implications on its Fourier transform.

#### Sampling Theorem and Aliasing

For any sampling rate $\tau$, there is a special frequency $\omega_c$, called the *Nyquist critical frequency*, given as,

$$\omega_c \equiv \frac{\pi}{\tau} \tag{1.2}$$

The Nyquist critical frequency is important for two reasons. The first one is known as the *sampling theorem* [9]: "If a continuous function $c(t)$, sampled at an interval $\tau$, happens to be *bandwidth limited* to the frequencies in magnitude smaller than $\omega_c$, i.e., if $I(\omega) = 0$ for all $|\omega| \geq \omega_c$, the the function $c(t)$ is *completely determined* by its samples $c(n)$." The explicit formula for $c(t)$, given in terms of $c(n)$, is,

$$c(t) = \tau \sum_{n=-\infty}^{\infty} c(n) \frac{\exp[\omega_c(t - n\tau)]}{\pi(t - n\tau)} \,. \tag{1.3}$$

Therefore, even though generally speaking, the information content of a discrete signal is infinitely smaller than that of a continuous signal, entire information content of the bandwidth limited signal can be recorded in a discrete signal, sampled at at rate $\tau^{-1} \geq 2\,\omega_{\max}$. Fairly often, this is the case. For example, NMR signals are always filtered and amplified in a finite frequency range before sampled.

Another consequences of discrete sampling is a phenomenon called *aliasing* for signals that are *not* bandwidth limited to less than the Nyquist critical frequency: any frequency component outside of the Nyquist frequency range, $(-\omega_c, \omega_c)$, is *aliased* (or falsely translated) into that range. In another word, frequencies that differ by multiples of $2\,\omega_c$ are completely indistinguishable due to the discrete sampling. Aliasing causes ambiguity in determining the true frequencies and should be avoided in most circumstances. In some special cases, aliasing can be used on purpose to reduce the spectral range and thus increase the digital resolution of the FT spectrum [10, 11].

### Discrete Fourier Transform

Given a finite discrete time signal $c(n)$, available from $n = 0, 1, \ldots, N-1$, we can estimate its Fourier transform on the same numbers of discrete frequency points,

$$I(k) = I(\omega_k) = \int_{-\infty}^{\infty} c(t)\, e^{i\omega_k t}\, dt \approx \tau \sum_{n=0}^{N-1} c(n)\, e^{in\omega_k \tau}\,, \tag{1.4}$$

where $\omega_k = 2\pi k/N\tau, k = -N/2, \ldots, N/2$. The summation of Eq. 1.4 is called the *Discrete Fourier Transform* (DFT) of the N data points $c(n)$. The inverse discrete Fourier transform to recover $c(n)$ from $I(k)$ is:

$$c(n) = \frac{1}{N} \sum_{k=0}^{N-1} I(k)\, e^{-in\omega_k \tau}\,. \tag{1.5}$$

DFT is a linear transform of the time domain data, and is thus stable and reliable. However, direct calculation of DFT is an order of $N^2$ process. A more efficient implementation is called the Fast Fourier Transform (FFT) [6, 7], which requires only $O(N \ln N)$ operations. FFT requires $N$ to be the power of two[1]. For data sets of arbitrary sizes, the Fastest Fourier Transform in the West (FFTW) [12] can be used.

### Resolution and Sensitivity

The two most important measurements of the quality of a spectrum are *resolution* and *sensitivity*. Resolution refers to the ability to distinguish components of the signal that are close in frequency. It can be intrinsically determined by the signal itself, or, in turn, by the underlying physical system and experimental techniques used to probe and detect the signal. It is also often affected by the numerical method used to obtain

[1]There are also FFT algorithms for data sets of length $N$ that are not a power of two. But they still require $N$ being divisible by some other prime numbers [9].

the spectral presentation. The obtainable resolution of a spectral analysis method can be roughly measured by the Full-Width-at-Half-Height (FWHH) of the resulting lineshape of an arbitrarily narrow line. The DFT resolution for a finite signal is limited by the so called *FT time-frequency uncertainty principle*,

$$\delta\omega \sim \frac{1}{N\tau}\,. \tag{1.6}$$

Thus, the resolution of DFT converges linearly with respect to the signal length. This is one of the limitations of DFT. We will discuss its consequences in more details in Section 1.2.4.

Sensitivity refers to the ability to distinguish signal from noise. A commonly used measurement for sensitivity is the *Signal-to-Noise Ratio* (SNR) of the spectrum. SNR can be defined as the ratio between the height of the highest signal peak to the standard deviation of the noise. As pointed out by Hoch and Stern [8], while SNR is often used to quantify the sensitivity, they are not the same. When nonlinear methods like maximum entropy reconstruction or others are used to compute the spectrum, SNR is not a good measurement of the sensitivity any more. In these cases, sensitivity should be accessed explicitly, rather than inferred from the SNR of the resulting spectrum.

### 1.2.2 Data Manipulations

Direct Fourier transform of the time signal is often unsatisfactory, mainly due to many potential imperfections of the data. It is routine practice to “pre-process” the data set before it is Fourier transformed. In this section, we will discuss several common data manipulations, especially in NMR data processing.

### Zero- and First-Order Phase Correction

In order to understand the concept of phase corrections, we need to first discuss two different lineshapes in NMR. Nuclear magnetic resonances in liquids can be modeled as damped harmonic oscillators. Let's consider a signal that has a single component with frequency $\omega_0$ and decay constant $\gamma_0 > 0$. It can be written as,

$$s(t) = \exp[-i\omega_0 t - \gamma_0 t] = \exp[-i(\omega_0 - i\gamma_0)t]\,, \tag{1.7}$$

for $t \geq 0$; $s(t) = 0$ for $t < 0$. Such a signal is also called a *damped sinusoid.* Its Fourier transform is then,

$$\begin{aligned} I(\omega) &= \int_0^\infty s(t)\, \exp(i\omega t)\, dt = \frac{1}{i(\omega - \omega_0) + \gamma_0} \\ &= \frac{\gamma_0}{(\omega - \omega_0)^2 + \gamma_0^2} + i\frac{(\omega - \omega_0)}{(\omega - \omega_0)^2 + \gamma_0^2} = A(\omega) + iD(\omega)\,, \end{aligned} \tag{1.8}$$

where $A(\omega)$ and $D(\omega)$ are called *absorption* and *dispersion* Lorentzian lineshapes, illustrated in Figure 1.2. For large frequency offsets, the absorption Lorentzian lineshape decays as $1/(\omega - \omega_0)^2$, while the decay of dispersion Lorentzian lineshape is proportional to $1/(\omega - \omega_0)$. Accordingly, absorptive Lorentzian lineshape is more localized and thus preferred for high-resolution NMR. In addition, the zero crossing at the center of the dispersive Lorentzian lineshape is also unfavorable.

Experimental signals are not always as perfect as Eq. 1.7. The signal might not have zero phase at $t = 0$ and it might not even start at $t = 0$. A more general model for the signal is,

$$s(t) = \exp[i(\phi_0 + \omega_0\, t_0)]\, \exp[-i(\omega_0 - i\gamma_0)t]\,, \tag{1.9}$$

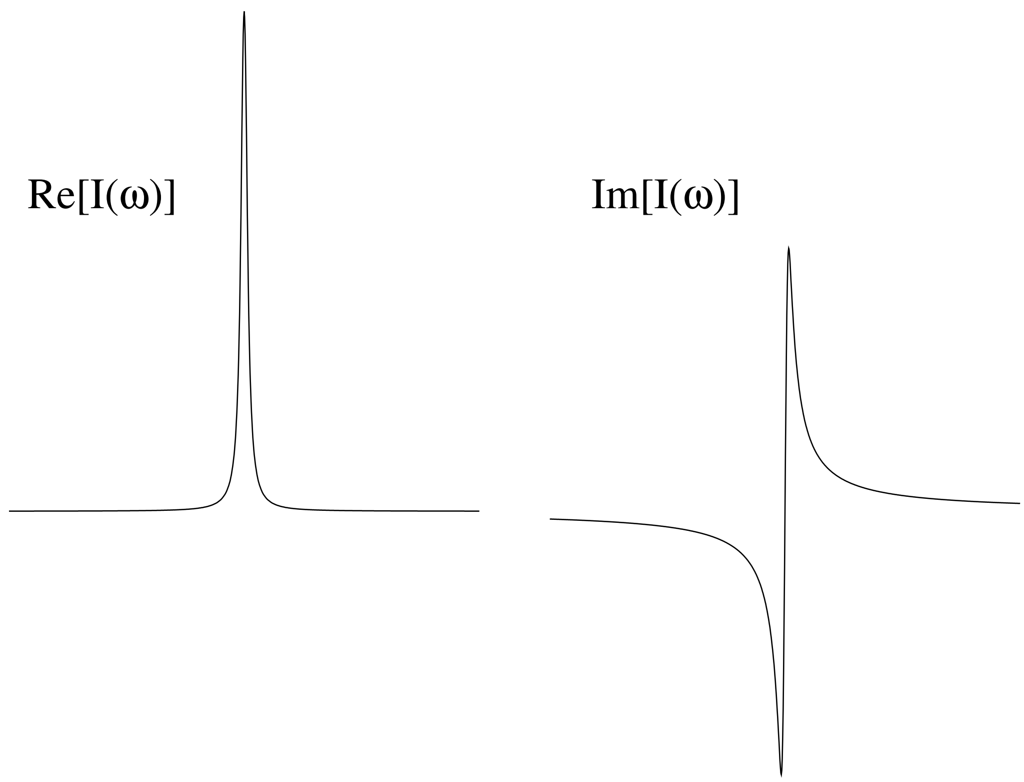



Figure 1.2: Real and imaginary part of the Fourier transform of a damped sinusoid with zero phase at $t = 0$: (a) absorption Lorentzian lineshape; (b) dispersion Lorentzian linshape.

where $\phi_0$ is called the *zero-order phase*, and $\omega_0\, t_0$ the *first-order phase*. The first-order phase is due to the delay $t_0$ and also called the *linear phase* as it has linear dependence on the frequency. According to the linear property of FT, the Fourier transform of $s(t)$ is given as,

$$I(\omega) = \exp[i(\phi_0 + \omega_0\, t_0)]\, [A(\omega) + iD(\omega)]\,. \tag{1.10}$$

Note that the real part of $I(\omega)$ automatically yields absorption lineshapes only when both zero- and first-order phase are zero, or more precisely, $\phi_0 + \omega_0\, t_0 = 0$. Otherwise, the real part of the complex spectrum will contain unfavorable dispersion lineshapes, requiring the *zero- and first-order phase corrections.* The first-order phase correction is also called *linear phase correction.* The zero-order phase can be reliably removed simply by multiplying $I(\omega)$ by $\exp(-i\phi_0)$. However, the first-order phase can only be approximately corrected in FT. The reason is that exact first-order phase correction requires knowledge of the frequencies of all components, which is not available in FT spectral analysis. In practice, we simply multiply $I(\omega)$ by $\exp(-i\omega\, t_0)$ to remove the first order phase. Such a point-by-point correction will be exact only at the center of each resonance. A large linear phase correction leads to severe phase roll in the baseline. Such an example is given in Figure 3.1, trace (c). The linear phase occurs very often in NMR due to many experimental limitations including limited transmitter response time, stabilization delay after field gradient pulses and others. One of the main goals of developing alternative methods to FT is actually to find a more consistent way of correcting the linear phase.

**Zero-Filling**

The digital resolution of the DFT spectrum is given as $\delta\omega = 1/N\tau$ and thus limited by the number of data points in the time domain. We can append a sequence of zeros to the available data set to arbitrarily increase the total number of data points and thus improve the digital resolution of the resulting DFT spectrum. Such a process is called *zero filling.* Zero-filling is equivalent to some interpolation in the frequency domain. At the frequency points they have in common $(n/N\tau)$, the DFT spectra of the zero-filled and original data will agree.

NMR signals obeys the *causality* principle as $c(t) = 0$ for all $t < 0$. As a consequence, the real and imaginary parts of the complex FT spectrum have a deterministic relationship according to the Kramers-Kronig relations [13, 14]. However, as noted by Bartholdi and Ernst, the Kramers-Kronig relations do not hold for discretely sampled NMR signal unless the signal is extended by a factor of two by zero-filling, because the periodicity of DFT makes the real and imaginary parts independent [15]. It was proved that real improvement in the information content was obtained by zero-filling the signal by a factor of two. Further zero-filling results in only cosmetic interpolation between data points in the frequency domain and no additional information is obtained.

### Apodization

*Apodization* is the process of modifying the original data by multiplying it with a *apodization* function before zero-filling and Fourier transform:

$$c'(n) = c(n)\, w(n)\,.$$

$w(n)$ is also called *windowing function* or *filter function.* The original DFT can be treated as a DFT with a *rectangular* windowing function, $w(n) = 1$. The purpose of apodization is to improve the quality of certain aspect of the spectrum or to obtain a desired lineshape. For example, a *matched filter*, an apodization function that matches the envelope of the signal, is often used to enhance the sensitivity of the resulting spectrum [16], with the side effect of broadening the lines. Or, an increasing exponential apodization function can be used to improve the resolution, but on the price of reducing the sensitivity. For truncated signal, it is often necessary to use some apodization to reduce the amplitude of the signal smoothly to zero in order to suppress the truncation artifacts[2], which also inevitably broadens the lines. In practice, one has to compromise among resolution, sensitivity, lineshape, amount of truncation artifacts in finding an "optimal" apodization function. A comprehensive overview of various commonly used apodization functions in NMR can be found in Hoch and Stern's seminar book on NMR data processing [8].

### 1.2.3 Multi-Dimensional DFT

It is the ability to perform multidimensional NMR experiments that makes NMR such a powerful method as it is today. By introducing an additional dimensions, we are not only able to disperse the potentially heavily overlapped features in the one-dimensional (1D) spectrum into the addition dimensions and thus have a much better separation, but also able to obtain important information about the chemical and/or dynamic correlations between resonances. It is fair to say that without the multi-dimensional NMR experiments, many of the applications that fueled the explosive growth of NMR in the last two decades would not be possible [8].

A D-dimensional complex signal, sampled discretely on a D-dimensional time grid, can be written as,

$$c(\vec{n}) \equiv c(n_1\tau_1, \ldots, n_D\tau_D)\,, \qquad n_l = 0, 1, \ldots, N_l - 1\,,$$

where $\tau_l$ is the sampling rate along $l$-th dimension. In NMR, the last time variable $t_D = n_D\tau_D$ is the *acquisition time dimension* by convention. The other time dimensions are called *indirect dimensions*, because they can only be incremented indirectly

[2]See Section 1.2.4 for more details on truncation artifacts.

as varying delays during the experiments. Figure 1.3 shows a general scheme [11] for recording two-dimensional NMR signal. There are two time dimension, $t_1$ and $t_2$. Only the acquisition time $t_2$ is the real time. $t_1$ is an indirect time dimension, implemented as a varying delay. For each value of $t_1 = n_1\tau_1, n_1 = 0, 1, \ldots, N_1 - 1$, the whole experiment is repeated and a 1D signal, $c(n_1\tau_1, t_2)$, is recorded. After a total of $N_1$ 1D experiments, a 2D signal results. It is important to notice that the

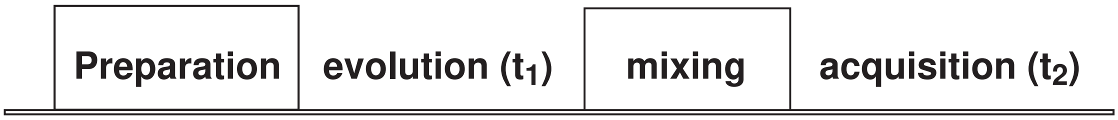



Figure 1.3: General scheme for 2D NMR spectroscopy. It includes four segments: preparation of the states, evolution in the indirect dimension ($t_1$), mixing period, and acquisition time ($t_2$). In the mixing period, a correlation is established between the two time dimensions. Only $t_2$ is the real time.

acquisition time dimension is the only inexpensive dimension, sampled in real time at a rate of one data point per $ms$. Each increment in the indirect dimension requires repeating the 1D experiment over and is thus time consuming. By analogy, for a 3D experiment, each increment in the second indirect dimension requires repeating the whole 2D experiment over, and so on. The total experimental of a D-dimensional NMR experiment is thus proportional to the total number of increments the indirect dimensions , $N_1 \times N_2 \times \ldots \times N_{\mathrm{D}-1}$. This is the reason why multidimensional NMR signals are typically truncated in the indirect dimensions.

Surprisingly, despite all of its importance, processing of multidimensional time signals by FT requires few new concepts beyond those involved in the 1D FT spectral analysis. The complex D-dimensional DFT, defined as,

$$I(\omega_1, \ldots, \omega_D) = \prod_{l=1}^{D} \tau_l \sum_{n_1=0}^{N_1-1} \cdots \sum_{n_D=0}^{N_D-1} e^{in_1\omega_1\tau_1} \ldots e^{in_D\omega_D\tau_D}\, c(n_1\tau_1, \ldots, n_D\tau_D)\,, \quad (1.11)$$

where $\omega_l$ is a $N_l$-point frequency grid specified as $\omega_{lk} = 2\pi k/N_l\tau_l, k = -N_l/2, \ldots, N_l/2$, can be simply treated as a series of 1D DFT, applied to each dimension successively:

$$\mathbf{c(t_1, \ldots, t_D)} \xrightarrow{\mathbf{FT}} \mathbf{c(t_1, \ldots, \omega_D)} \xrightarrow{\mathbf{FT}} \mathbf{\ldots} \xrightarrow{\mathbf{FT}} \mathbf{I(\omega_1, \ldots, \omega_D)}$$

All the data manipulations such as phase corrections, zero-filling and apodizations can be applied to each dimension independently.

### Absorption Mode DFT Spectrum Using Multiple Data Sets

One of the few complications of multidimensional DFT is that the absorption mode lineshape can not be obtained simply by taking the real part of the appropriately phased D-dimensional complex DFT spectrum. For the sake of simplicity, let's only consider the 2D case. Following Eq. 1.8, the 2D complex DFT spectrum computed from a single 2D complex signal can be written as,

$$\begin{aligned} I(\omega_1, \omega_2) &= (A_1 + iD_1)(A_2 + iD_2) \\ &= (A_1A_2 - D_1D_2) + i(A_1D_2 + D_1A_2) \end{aligned} \quad (1.12)$$

with convenient notations $A_l = A(\omega_l), D_l = D(\omega_l),\ l = 1, 2$. Therefore, the the real part of the 2D complex spectrum is a supposition of the double-absorptive, $A_1A_2$, and double-dispersive, $D_1D_2$, 2D Lorentzian lineshapes. The resulting lineshape is *phase-twisted* and extremely undesirable in high-resolution NMR spectroscopy, as the dispersive contributions tail off very slowly and degrade the resolution significantly. Figure 1.4 compares the (a) double-absorption, (b) double-dispersion, and (c) phase-twisted 2D Lorentzian lineshapes.

In order to obtain the desired double-absorption lineshape, FT requires two independent 2D complex data sets, acquired from the same experiment. In NMR, the evolution in the indirect dimension, $t_1$, can modulate either the amplitude or the phase of the signal recorded during $t_2$[3]. In the former case, two *amplitude modulated* signals, *cosine* and *sine modulated* signals, are acquired, which, in case of a signal resonance, can be written as,

$$
\begin{aligned}
S_c(t_1, t_2) &= \cos(\omega_{10} t_1) \exp(-\gamma_{10} t_1)\, \exp[-i(\omega_{20} - i\gamma_{20}) t_2]\,, && (1.13)\\
S_c(t_1, t_2) &= \sin(\omega_{10} t_1) \exp(-\gamma_{10} t_1)\, \exp[-i(\omega_{20} - i\gamma_{20}) t_2]\,. && (1.14)
\end{aligned}
$$

In the later case, N- and P-type *phase modulated* signals are acquired,

$$
\begin{aligned}
S_N(t_1, t_2) &= \exp(-i\omega_{10} t_1 - \gamma_{10} t_1)\, \exp[-i(\omega_{20} - i\gamma_{20}) t_2]\,, && (1.15)\\
S_P(t_1, t_2) &= \exp(i\omega_{10} t_1 - \gamma_{10} t_1)\, \exp[-i(\omega_{20} - i\gamma_{20}) t_2]\,. && (1.16)
\end{aligned}
$$

Note that the signs of the frequency in $t_1$ are opposite for N- and P-signals. These two data formats can be converted into each other simply by linear combinations. For example,

$$
\begin{aligned}
S_c(t_1, t_2) &= [S_P(t_1, t_2) + S_N(t_1, t_2)]/2\,, && (1.17)\\
S_s(t_1, t_2) &= [S_P(t_1, t_2) - S_N(t_1, t_2)]/2i\,. && (1.18)
\end{aligned}
$$

Using linear combination of the 2D DFT of the two 2D data sets from the same

[3]In derivation of Eq. 1.12, a purely phase modulated signal was actually assumed.

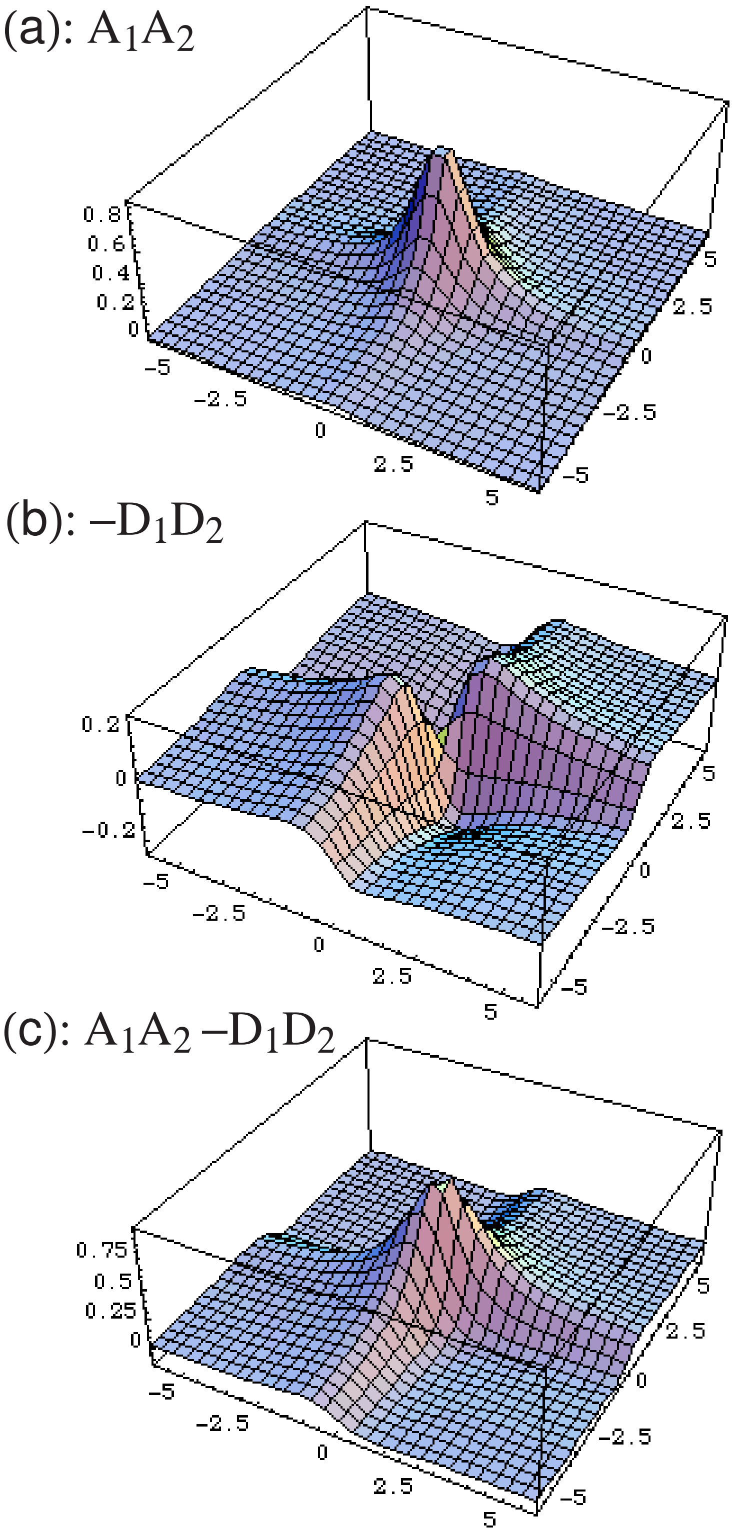



Figure 1.4: Comparison of (a) double-absorption, (b) double-dispersion, and (c) phase-twisted 2D Lorentzian lineshapes, for a 2D Lorentzian line at frequency $(\omega_1, \omega_2) = (0, 0)$ Hz with linewidth being 1 Hz in both dimensions.

experiment, double absorption lineshape can be constructed. Here, a procedure for obtaining absorption mode spectrum from two amplitude modulated data sets, assuming that both data sets are correctly phased, is outlined:

(i). Apply DFT in the $t_2$ dimension for both data sets:

$$S_c(t_1, \omega_2) = \cos(\omega_{10} t_1) \exp(-\gamma_{10} t_1)(A_2 + iD_2)\,, \tag{1.19}$$

$$S_s(t_1, \omega_2) = \sin(\omega_{10} t_1) \exp(-\gamma_{10} t_1)(A_2 + iD_2)\,. \tag{1.20}$$

(ii). Construct an intermediate signal according to:

$$S_{\text{new}}(t_1, \omega_2) = \text{Re}[S_c(t_1, \omega_2)] - i\,\text{Re}[S_s(t_1, \omega_2)] = A_2 \exp[-i(\omega_{10} - i\gamma_{10})t_1]\,. \tag{1.21}$$

(iii). Apply DFT to the intermediate signal in the $t_1$ dimension. The real part of the result gives the desired absorption mode 2D spectrum:

$$A(\omega_1, \omega_2) = \text{Re}[I(\omega_1, \omega_2)] = \text{Re}[A_2(A_1 + iD_1)] = A_1 A_2\,. \tag{1.22}$$

**First Point Correction**

Another complication of multidimensional FT is also related to the problem of obtaining an absorption mode spectrum. The approximation of DFT to FT integral given in Eq. 1.4 has a systematic error: the first point should be weighted by 1/2 before transformed. While this only causes a constant vertical offset in 1D DFT, the amount of such an offset will vary for each row in the indirect dimension, causing so called $t_1$-*ridges* in the 2D FT spectrum. These ridges deteriorate the resolution of an absorption mode spectrum significantly and should be removed, simply by incorporating the *first point corrections* into the multi-dimensional DFT summation,

$$
\begin{aligned}
I(\omega_1, \ldots, \omega_D) &= \prod_{l=1}^{D} \tau_l \sum_{n_1=0}^{N_1-1} \cdots \sum_{n_D=0}^{N_D-1} e^{in_1\omega_1\tau_1} \ldots e^{in_D\omega_D\tau_D} \\
&\times \; (1 - \frac{\delta_{n_1,0}}{2}) \ldots (1 - \frac{\delta_{n_D,0}}{2})\, c(n_1\tau_1, \ldots, n_D\tau_D)\,.
\end{aligned}
\tag{1.23}
$$

### 1.2.4 Limitations of FT

The Fourier transform is fast and numerically stable, and can produce frequency-domain spectra in a convenient representation. It is the method of choice for spectral analysis in many applications including NMR. Nonetheless, FT also has some well known limitations, leaving plenty of room for improvement.

#### Truncation Artifacts and FT Uncertainty Principle

One of the most severe limitations of the FT spectral analysis occurs when the signal is truncated before it fully decays to zero, or, more realistically, below noise level. The effect of truncation can be modeled as a step function, $r(t) = 1$ for $t \leq T_{\max}$, and $r(t) = 0$ for $t > T_{\max}$. The acquired signal $c(t)$ can be then represented as the product of the signal (extending to $t = \infty$), $s(t)$, and the step function $r(t)$: $c(t) = s(t)r(t)$. According to the convolution theorem, the resulting finite FT spectrum is given by,

$$
I(\omega) = \int_{-\infty}^{\infty} s(t)r(t)dt = S(\omega) * sinc(T_{\max}\omega)\,, \tag{1.24}
$$

where the sinc function is defined as $sinc(x) = \sin(x)/x$. Convolution of the sinc function has two direct consequences on the finite time FT spectrum: first, it produces severe oscillating *truncation artifacts* in the baseline. The artifacts are often called

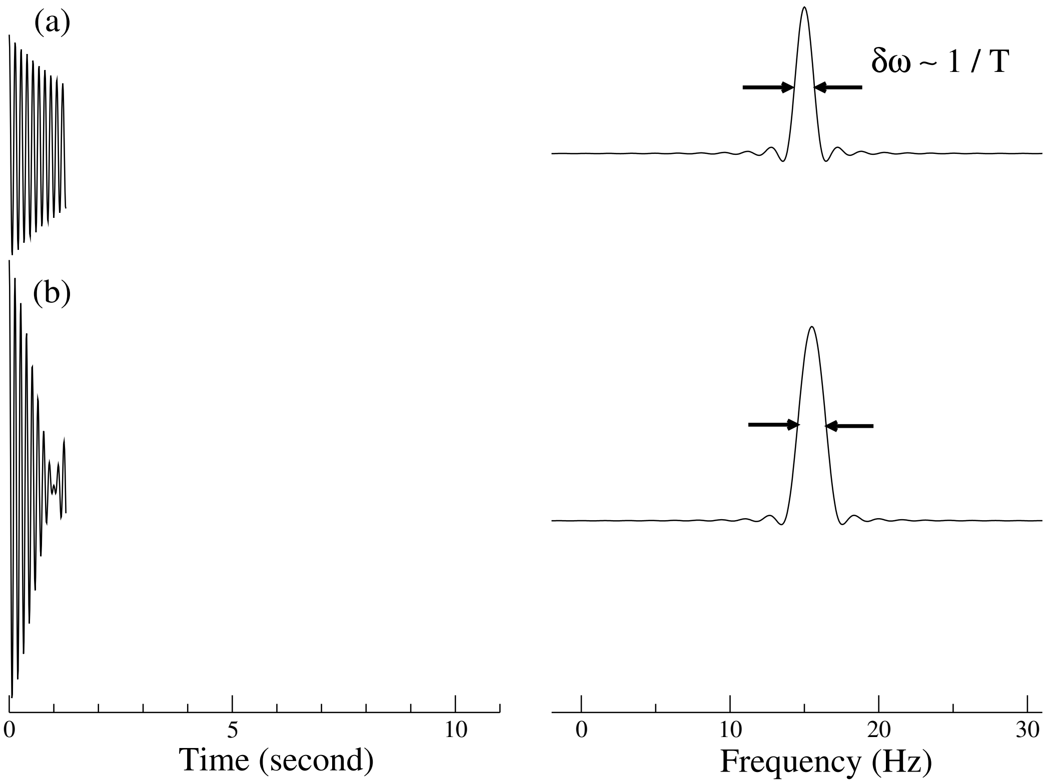



Figure 1.5: Effects of truncation on the FT spectrum: (1) sinc-wiggles in the baseline; (2) all lines are broadened to a minimum width of $\delta\omega \sim 1/T_{\max}$. Note that the doublet of (b) is now unresolved due to the truncation of the signal.

*sinc-wiggles* or *Gibbs oscillations*. Second, it broaden all the peaks to be at least the width of center envelope of the sinc function, which is roughly inverse of the signal length. The later is often called the *FT time-frequency uncertainty principle*. Figure 1.5 illustrates these two truncation effects using the same time signals used in Figure 1.1, but truncated to 1/8 of the original sizes.

The sinc-wiggles are very unfavorable in high-resolution NMR as they obscure the weak peaks around a strong peak. Apodization can be used to suppress the truncation artifacts, which will further broaden the lines. Note that the resolution of finite FT, determined by the FT uncertainty principle, improves only linearly with respect to the signal length. While it does not look as bad in 1D, this slow convergence property becomes a severe limitation of the multi-dimensional FT spectral analysis.

### 1D Spectral Analysis

Multidimensional Fourier transform is simply a series of 1D FT, applied to each dimension sequentially. The FT uncertainty principle applies to each dimensions independently. Accordingly, in order to obtain high resolutions in all dimensions of the FT spectrum, one has acquire a signal that is long in all dimensions. However, as we discussed in the beginning of Section 1.2.3, the indirect dimensions of a multi-dimensional NMR experiment are expensive and thus always truncated, or severely truncated in the case of three- and four-dimensional experiments, leading to poor FT resolutions in corresponding spectral dimensions. This is one of the major limitations of multidimensional FT-NMR.

The major limitation of 1D spectral analysis is that correlations among multiple dimensions are not explicitly exploited. The fact that evolutions in all dimensions are correlated is completely ignored in the FT spectral analysis. A true multi-dimensional method for spectral analysis, which can efficiently use this information, can have substantial advantages over any 1D methods including FT.

### Modelless Spectral Analysis

FT does not make any assumption about the signal. On one hand, this is a big advantage as such FT can be used to analyze any type of signals. On the other hand, this can be a big disadvantage when the signal does satisfy certain models. For example, a NMR signal in liquids can be well modeled by a sum of finite number of damped sinusoids. Unfortunately, this useful information can not utilized and thus totally wasted in the FT spectral analysis.

## 1.3 Alternative Methods for Spectral Analysis

Due the limitations of FT discussed in previous section, extensive efforts have been made to develop alternative methods. One common goal of all these methods is to produce resolutions beyond the FT uncertainty principle, especially in the case of truncated signals. We thus call these methods *high-resolution methods* for spectral analysis. In this section, we will first discuss the basis principle of high resolution spectral analysis and then briefly review major high-resolutions methods that have been applied to NMR spectral analysis.

### 1.3.1 Principle of High-Resolution Spectral Analysis

A high-resolution method for spectral analysis is a method that can obtain a spectral resolution beyond FT time-frequency uncertainty principle. In order to obtain high-resolution, *additional information* must be used.

Let's consider a simple example where we know that the signal contains a single Lorentzian line and thus satisfies following model:

$$c(t) = d_0 \exp(-i\omega_0 t)\,, \tag{1.25}$$

where the complex amplitude $d_0 = |d_0| \exp(i\phi_0)$ and complex frequency $\omega_0 = 2\pi f_0 - i\gamma_0$ are two unknown parameters. Note that we extend the notation $\omega_0$ be complex so that both the line position and linewidth can be represent by the a single parameter. The magnitude of the complex amplitude, $|d_0|$, gives the peak integral, and $\phi_0$ defines the phase of the peak. As shown in Figure 1.1 (a) and 1.5 (a), FT is not able to use this information and can not provide a fully converged spectrum[4] unless the signal fully decays (Figure 1.1 (a)). Alternatively, armed with the model of Eq. 1.25, we know that only two unknowns, $d_0$ and $\omega_0$, need to be found. In turn, it requires only two data points, so that we can set up the two equations, from which both parameters can be solved accurately:

$$\begin{cases} c(0) = d_0 \\ c(\tau) = d_0 \; \exp(-i\,\omega_0 \tau) \end{cases} \Longrightarrow \begin{cases} d_0 = c(0) \\ \omega_0 = i \ln\left[c(\tau)/c(0)\right]/\tau \end{cases}$$

By analogy, for the slightly more complicated case of two Lorentzian lines (Figure 1.1 (b) and 1.5 (b)), as few as *four* complex data points are required to solve for four unknowns: two pairs of complex amplitudes and complex frequencies, providing a fully converged spectrum.

We just demonstrate that it is possible to obtain resolutions beyond the FT uncertainty principle by using some addition information that is not utilized in FT. However, in practice, how we model the signal, how the model is used to reformulate the problem and how the problem is solved numerically are all very important limiting factors. Experimental signals like NMR signals are often both large and complex, and contain unknown numbers of peaks. In addition experimental signals do not satisfy the model perfectly due to the noise and other imperfections. As a result, one can easily end up solving a large, ill-conditioned linear or nonlinear problem, which

[4] A fully converged spectrum should show a peak at the right position with the true linewidth.

is computationally expensive and numerically unstable. In the rest of this section, we will briefly review existing high-resolution methods, with the main focus on their main advantages and remaining difficulties.

The existing high-resolution methods can be grouped into two main categories: parametric methods and nonparametric methods, based on whether a *parametric representation*, a list of complex characteristic frequencies and amplitudes, is involved. FT is a nonparametric method, as the spectral representation is directly obtained without computing an intermediate parametric representation of the signal.

### 1.3.2 Parametric Methods

Practical methods for estimating the spectral parameters can be traced all the way back to 1795, when Baron de Prony published his method for fitting the data to a sum of complex sinusoids [17]. The idea of using linear algebra for identifying the parameters of multi-exponential decay of a time signal is still used in more recent parametric methods including Linear Prediction (LP) [18, 19, 20, 8], the matrix pencil method [21, 22, 23], and the Filter Diagonalization Method (FDM) [24, 25]. The first two methods will be described below, and FDM will be introduced later in Chapter 2. Other parametric methods rely on nonlinear optimization of complex score functions, based on either probability theories such as in Bayesian and maximum likelihood techniques [26, 27, 28, 29], or simple minimum variation principle like in Least Squares (LSQ) fitting in the frequency domain [30] and the time domain [31].

All these methods model the NMR signal as a sum of Lorentzian lines:

$$c(n) = \sum_k d_k\, e^{-in\omega_k\tau} \, + \xi(n) = \sum_k d_k\, u_k^n \, + \xi(n)\,, \tag{1.26}$$

where $d_k$ and $\omega_k$ are complex amplitudes and complex frequencies, and $\xi(n)$ is noise. The line list $\{\omega_k, d_k\}$ is called a *parametric representation* of the signal $c(n)$. Each pair of $(\omega_k, d_k)$ represents a *peak*, or, a *pole*. This model has been proved to be a good model for most NMR signals in liquids. It is the additional information that all these methods utilize, in one way or another. However, how this model is used to reformulate the problem and the numerical approach used to solve for the unknown spectral parameters vary from one method to another, and determine the efficiency and stability of each method.

**Linear Prediction**

Linear prediction is one of the most popular high-resolution alternatives for NMR spectral analysis. Even though linear prediction is typically only use to extrapolate the signal, LP is intrinsically a parametric method. It is based on the concept of *autoregressive modeling* (AR) [17, 16]:

$$c(n) = \sum_{k=1}^{K} a_k\, c(n-k) \, + \epsilon(n)\,, \quad n = K, \ldots, N-1\,, \tag{1.27}$$

where the coefficients $a_k$ are called *LP coefficients*, $K$ is the *prediction order*, and $\epsilon(n)$ is the prediction error. It can be proved that the AR model is equivalent to the Lorentzian model of Eq. 1.26 with $K$ damped sinusoids. Once the prediction coefficients are determined, the spectral representation can be directly computed, or more commonly, the AR expression is used to extrapolate the signal, typically, by a factor of 2, then the FT of the extended signal is computed.

There exist several LP algorithms for finding the prediction coefficients including LP-SVD [32], LP-QRD [33] and LP-TLS [34]. Detailed descriptions can be also found in Refs. [8] and [20]. The basic idea is to find a set of coefficients $a_k$ that minimizes the prediction error by solving a linear least squares problem,

$$\begin{bmatrix} c(K-1) & c(K-2) & \cdots & c(0) \\ c(K) & c(K-1) & \cdots & c(1) \\ \vdots & \vdots & \ddots & \vdots \\ c(N-2) & c(N-3) & \cdots & c(N-K-1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix} = \begin{bmatrix} c(K) \\ c(K+1) \\ \vdots \\ c(N-1) \end{bmatrix} + \begin{bmatrix} \epsilon(K) \\ \epsilon(K+1) \\ \vdots \\ \epsilon(N-1) \end{bmatrix},$$

or in a compact form,

$$\mathrm{H}\,\mathrm{x} = \mathrm{b} + \epsilon\,. \tag{1.28}$$

The complex frequencies $\omega_k$ of all poles can be then computed by rooting the *characteristic polynomial*, defined as,

$$P(u) = \sum_{k=1}^{K} a_k\, u^{K-k}\,, \tag{1.29}$$

where the roots $u_k = \exp(-i\omega_k\tau)$. The complex amplitudes $d_k$ are determined by solving another linear least squares problem based on Eq. 1.26:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_1 & u_2 & \cdots & u_K \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{N-1} & u_2^{N-1} & \cdots & u_K^{N-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_K \end{bmatrix} = \begin{bmatrix} c(0) \\ c(1) \\ \vdots \\ c(N-1) \end{bmatrix} + \begin{bmatrix} \xi(0) \\ \xi(1) \\ \vdots \\ \xi(N-1) \end{bmatrix}. \tag{1.30}$$

Eqs. 1.29 and 1.30 might seem necessary only when LP is used as a parametric estimator, while for LP extrapolation, only Eq. 1.28 needs to be solved. This is not true. Some roots of the characteristic polynomial $P(u)$ might fall outside of the unit circle on the complex plane, i.e., $|u_k| > 1$, which correspond to increasing exponentials. In this case, direct use of the prediction coefficients to extrapolate the signal is unstable and can result in numerical overflow. In practice, all the roots of $P(u)$ are first found, then some procedures have to be used to eliminate unwanted roots. For example, those $u_k$ lying outside of the unit circle are either discarded or replaced by their images reflected about the circle: $u_k \to [u_k^{-1}]^*$. Finally, a new set of prediction coefficients are computed by reconstructing the characteristic polynomial using the "corrected" roots:

$$P'(u) = (u - u_1)(u - u_2) \ldots (u - u_K) \,. \tag{1.31}$$

The main advantage of LP is that the highly nonlinear problem of fitting the signal to the summation of damped sinusoids is solved by a *linear* fitting of the signal to the AR model, and thus avoids the notoriously finicky nonlinear fitting procedures. However, there are also several severe limitations:

(i). Filter order is a complete part of the AR model, and finding its correct value is a crucial issue. Even though extensive efforts have been made in developing methods for order selection [35, 36, 37, 38], no universal solution exits [20].

(ii). LP attempts to fit the whole 1D signal in one shot. The size of the least squares problem is determined by the signal size and the filter order. Thus LP can be prohibitively expensive for very large data sets. In addition, rooting high degree polynomial requires significant higher algorithm complexity necessary to avoid numerical instability, round-off errors and overflow.

(iii) When parametric LP is used, it requires a second least squares calculation for computing the complex amplitudes. Any errors in the rooting step might be amplified, making the parametric LP very unreliable.

(iv) In principle, it is possible to use LP for complete analysis of multidimensional data sets. However, it is not in common practice due to the expensive computational cost and numerical instabilities. In stead, hybrid LP and FT spectral analysis is typically used. For example, in 2D, LP is used to extrapolate each $t_1$ trace independently after FT is applied to the $t_2$ dimension. Such a hybrid method is also essentially a 1D method and does not have the advantages of true multidimensional data processing.

(v) Even though the prediction coefficients are optimized by various tricks, the noisy tail part of the signal is always used for extrapolation. Thus the extrapolation is doomed to be noisy and unreliable.

### The Matrix Pencil Method

The matrix pencil method was developed independently by Hua and Sarkar [21, 22] and by Kailath and coworkers [39]. Instead of rooting a polynomial to obtain the complex frequencies, they are computed directly from the data matrices by solving a generalized eigenvalue problem. Its formulation exploits of the property of the *matrix pencil* constructed from the underlying exponentially damped sinusoids (Eq. 1.26). The matrix pencil refers to linear combinations of two matrices defined on a common domain (i.e., same linear space). For example, given two matrices **L** and **R**, the set of all matrices of form $\mathbf{L} - \lambda\mathbf{R}$ is said to be a matrix pencil.

Given a discrete time signal $c(n)$ with $N$ data points that satisfies Eq. 1.26, and ignoring the noise term $\xi(n)$, the $(N-L)\times L$ data matrices $\mathbf{X}_0$ and $\mathbf{X}_1$, which are defined as:

$$\mathbf{X}_l = \begin{bmatrix} c(L-1+l) & c(L-2+l) & \cdots & c(l) \\ c(L+l) & c(L-1+l) & \cdots & c(1+l) \\ \vdots & \vdots & \ddots & \vdots \\ c(N-2+l) & c(N-3+l) & \cdots & c(N-L-1+l) \end{bmatrix}, \tag{1.32}$$

$l = 0, 1$, with $L$ being the *pencil parameter*, can can be decomposed as [21],

$$\mathbf{X}_0 = \mathbf{Z}_L \mathbf{B} \mathbf{Z}_R\,, \quad \mathbf{X}_1 = \mathbf{Z}_L \mathbf{B} \mathbf{Z} \mathbf{Z}_R\,, \tag{1.33}$$

where

$$\mathbf{Z}_L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_1 & u_2 & \cdots & u_K \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{N-L-1} & u_2^{N-L-1} & \cdots & u_K^{N-L-1} \end{bmatrix}, \ \mathbf{Z}_R = \begin{bmatrix} u_1^{L-1} & u_1^{L-2} & \cdots & 1 \\ u_2^{L-1} & u_2^{L-2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ u_K^{L-1} & u_K^{L-2} & \cdots & 1 \end{bmatrix}, \tag{1.34}$$

and $\mathbf{B} = diag\{b_1, b_2, \ldots, b_K\}$, $\mathbf{Z} = diag\{u_1, u_2, \ldots, u_K\}$. If $K \le L \le N-K$, each of $\{u_k : k = 1, 2, \ldots, K\}$ is the rank reducing number of the matrix pencil $\mathbf{X}_1 - u\mathbf{X}_0$:

$$\mathbf{X}_1 - u\mathbf{X}_0 = \mathbf{Z}_L \mathbf{B} \begin{bmatrix} u_1 - u & 0 & \cdots & 0 \\ 0 & u_2 - u & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_K - u \end{bmatrix} \mathbf{Z}_R\,. \tag{1.35}$$

Therefore, by definition, $\{u_k\}$ are the $K$ nonzero generalized eigenvalues of the matrix pair $(\mathbf{X}_1, \mathbf{X}_0)$,

$$\mathbf{X}_1\, \mathbf{q}_k = u_k \mathbf{X}_0\, \mathbf{q}_i\ , \tag{1.36}$$

where $\mathbf{q}_k$ is the eigenvectors associated with the eigenvalue $u_k$. Note that Eq. 1.36 can be solved by first multiplying both sides by a pseudo-inverse of $\mathbf{X}_0$, $\mathbf{X}_0^+$, and then using standard eigenvalue solvers to find the $K$ nonzero eigenvalues of $L \times L$ matrix product $\mathbf{X}_0^+\mathbf{X}_1$:

$$\mathbf{X}_0^+\mathbf{X}_1\, \mathbf{q}_k = u_k\, \mathbf{q}_k\ . \tag{1.37}$$

Finally, the complex amplitudes $d_k$ can be computed by solving the linear least squares problem of Eq. 1.30. Note that spurious poles with $|u_k| > 1$ can lead to numerical instabilities in solving Eq. 1.30, and should be eliminated in similar ways to those employed in linear prediction.

While sharing the advantages of being a linear algebraic method with LP, the matrix pencil method has several advantages over LP. First, The pencil parameter $L$ will affect the results, in a similar way that LP is affected by the prediction order. However, the performance of matrix pencil method is less sensitive to the choice of $L$ [21]. It was empirically found that the optimal value for $L$ ranges from $L = N/3$ for noisy signals to $L = N/2$ for signals with high SNR. Second, in matrix pencil method, all the poles are computed directly from the data matrices in one step, while LP requires first solving a linear least squares problem and then rooting the resulting polynomial. Accordingly, the matrix pencil method is more efficient and more accurate. It was also found [21, 23] that matrix pencil method was less sensitive to noise than the polynomial method.

Nevertheless, the matrix pencil method still has some severe limitations, most of which are shared by linear prediction. They are listed below:

(i). Estimating the pencil parameter $L$ is still a problem, even though the results are less sensitive to its value than in linear prediction.

(ii). The matrix pencil method still attempts to fit the whole signal in a single shot. Thus for large data sets, it is computationally very expensive and has problems with numerical instabilities, round-off error and overflow.

(iii). The amplitudes $d_k$ are computed by solving a separate linear least squares problem. Errors that are introduced in the diagonalization step might lead to disastrous results in $d_k$, making its application to analyzing NMR signals problematic[5].

(iv). There are significant difficulties in generalizing the matrix pencil method to processing multidimensional signals. Specifically, for example, in 2D, there are problems in identifying the frequency pairs [22].

### The Nonlinear Least Squares Analysis

Several least squares analysis of the DFT NMR spectra have appeared in the literature (see ref. [30] and references therein). Some of them are based on fitting the DFT spectrum to the theoretical lineshape, Eq. 1.8, obtained from evaluating the infinite time FT integral. However, as pointed out by Abildgaard and coworkers [40, 30], there are several severe discrepancies between the DFT spectrum and the theoretical spectrum, which in many cases hamper these analyses. Instead, a more realistic

[5]The matrix pencil method was initially introduced to process acoustic and radar imaging signals, where only the complex frequencies were of general interest.

lineshape that corresponds to finite DFT of damped sinusoids should be used:

$$I^{\tau}(\omega) = \tau \sum_{n=0}^{N-1} c(n)\, e^{in\omega\tau} = \tau \sum_{k} d_k \frac{1-(u_k/u)^N}{1-u_k/u}\,, \tag{1.38}$$

where $u \equiv \exp(-i\omega\tau)$. The object of least squares fitting in the frequency domain is to minimize the variance between the estimated spectrum of Eq. 1.38 and the actual DFT. Note that no zero-filling or apodization should be used in obtaining the actual DFT spectrum. Least squares analysis can be also directly carried out in the time domain by simply minimizing following loss function [31],

$$\chi^2(\mathbf{f}) = \chi^2(K, \{\omega_k, d_k : k = 1, 2, \ldots, K\}) = \sum_{n=0}^{N-1} \left| c(n) - \sum_{k=1}^{K} d_k\, e^{-in\omega_k\tau} \right|^2 . \tag{1.39}$$

It was claimed that the complete least squares analysis was the optimum spectral estimator and maximum spectral information could be extracted [40]. While the validity of such a statement is still under debate, there are other serious drawbacks of the LSQ estimator. Firstly, LSQ is a nonlinear optimization method. Both an accurate estimate of the number of pole and a good set of initial values for the spectral parameters are required and essential for convergence. In the case of complicated spectra with heavily overlapped features, such a requirement can not be fulfilled. In addition, LSQ also attempts to fit the whole spectrum together in a single shot and is thus computationally very expensive for large data sets. Finally, in principle, LSQ can be applied to analyze multidimensional signals. However, due to the large size and high complexity of multidimensional spectra, general application of LSQ to multidimensional spectral analysis is totally infeasible.

### Bayesian and Maximum Likelihood Methods

Bayesian analysis is a powerful statistical technique for estimating the degree to which an hypothesis is confirmed by experimental data. It was introduced to NMR spectral analysis in 1990 by Bretthorst [26, 27, 28] and Sibisi [41]. It has also been applied to processing multidimensional NMR signals [29]. Only the basic principles are described here. The details are quite complicated and can be found in the original references or Hoch and Stern's book on NMR data processing [8].

In essence, it is based on the probability theory: given the data $D$ and a particular model $M$, according to *Bayes's rule*, we have,

$$P(D)P(M|D) = P(D,M) = P(M)P(D|M) \tag{1.40}$$

where $P(D)$ and $P(M)$ are the individual probabilities, $P(D|M)$ and $P(M|D)$ are the conditional probabilities, and $P(D,M)$ is the joint probability[6]. Accordingly, we can calculate the conditional probability of the model for a given data set as

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)} \sim P(D|M)P(M)\,. \tag{1.41}$$

$P(M|D)$ is also called the *posterior* of the model; $P(D|M)$ is also called the *likelihood* of the data (i.e., for a given model $M$, how likely the experimental data would be obtained); and $P(M)$ is also called the *prior* of the model (i.e., empirical or theoretical estimation of the probability of the model $M$ being true). $P(D)$ is independent of the model, and thus can be dropped out. The prior is only part the human *a priori* knowledge can bias the results. It can be set to a flat function (such as a infinitely

[6] The conditional probability $P(A|B)$ is the probability of $A$ being true given that $B$ is true. The joint probability $P(A,B)$ is the probability of both $A$ and $B$ being true.

broad Gaussian function) so that no model is preferred. In this case, $P(M)$ term can be also dropped in the analysis.

Assuming that 1D NMR signals can be modeled as a sum of damped sinusoids[7] of Eq. 1.26, the parameters of the model are then the number of peaks, $K$, and 2K complex frequencies and complex amplitudes, $\{\omega_k, d_k\}$. Thus,

$$P(M|D) \sim P(D|M)P(M) = P(D|K, \{\omega_k, d_k : k = 1, 2, \dots, K\})P(M)\,. \qquad (1.42)$$

The calculation of the likelihood $P(D|M)$ can be quite complicated. Details can found in the original references. In order to find the best (most likely) model, i.e, the number of poles and corresponding spectral parameters, for a given experimental data, one needs to maximize the posterior, or equivalently, the likelihood in the case of a flat prior function, by solving a high-dimensional optimization problem.

Despite of the neat probability theory behind them, Bayesian and maximum likelihood techniques are hampered by the resulting nonlinear optimization problem. Similar to LSQ method, finding *a* reasonable solution requires good initial set of values for all the parameters, which is not available for complex signals. In addition, constructing a good the likelihood funtion, $P(D|M)$, is essential and can be tricky. Finally, even though, in principle, these techniques can be apply to analyze any dimensional signals, the fast growing computational cost and resulting high dimensional optimization problem make them practically infeasible.

[7]This assumption is equivalent to use a prior function $P(M) = 1$ for all Lorentzian models and $P(M) = 0$ for others.

### 1.3.3 Nonparametric Methods

The parametric approaches for high-resolution spectral analysis suffer from a common drawback: it is essential for the time signal to sufficiently satisfy the model employed. In the contrary, nonparametric methods like FT do not make any assumption about the signal's functional form, and are thus insensitive to the noise and lineshape distortions of the signal. While there exist other nonparametric methods [42, 43, 44], the most popular one in NMR spectroscopy is the Maximum Entropy Reconstruction (MaxEnt).

#### Maximum Entropy Reconstruction

The maximum entropy reconstruction for NMR spectral analysis was first introduced by Sibisi in 1983 [45]. It can provide high resolution spectral estimation from truncated data sets, and can be even used analyze signals that are nonuniformly sampled. MaxEnt has generated considerable interest, and also some controversy on the claims of simultaneous noise suppression and resolution enhancement. The details of the method has been described extensively [46, 47, 19, 8]. Here we merely summarize the principles.

MaxEnt reconstructs the frequency domain spectrum, $\mathbf{f}(\omega)$, directly by solving an optimization problem similar to those in nonlinear least squares methods, with an additional constraint of maximizing some entropy function, $S(\mathbf{f})$. It is claimed that the minimum variance principle along (such as Eq. 1.39) does not provide enough information on the possible spectral reconstructions to be of much use. There exists infinite number of spectral reconstructions that satisfy the statistic constraint,

$$\chi^2(\mathbf{f}) < \sum_{n=0}^{N-1} |\xi(n)|^2 = C_0\,,$$

where $\xi(n)$ is the noise in the data, and $\chi^2(\mathbf{f})$ is defined as,

$$\chi^2(\mathbf{f}) = \sum_{n=0}^{N-1} \Big|[\mathbf{iDFT}(\mathbf{f})]_n - c(n)\Big|^2\,, \tag{1.43}$$

where $\mathbf{iDFT}(\mathbf{f})$ is the *"mock"* time signal computed by inverse DFT of the spectrum $\mathbf{f}(\omega)$. If we are to improve over DFT, we have to include some additional criterion for choosing an optimal reconstruction, such as the maximum entropy principle. What maximum entropy principle implies is that a reasonable reconstruction should not add any new information beyond what is contained in the experimental data [46, 8]. The entropy the spectrum $\mathbf{f}$ can be expressed in a general form,

$$S(\mathbf{f}) = -\sum_{n=1}^{N-1} R(f_n)\,, \tag{1.44}$$

where $R(f_n)$ is the negative of the contribution of each spectral point to the overall entropy. The functional form of $R(f_n)$ resembles $|f_n|\log(f_n)$. Their actual forms can be very complicated [47, 19, 8], but irrelevant to the present discussions. The constrained optimization problem can be converted to an unconstrained one of maximizing the following objective function,

$$Q(\mathbf{f}) = S(\mathbf{f}) - \lambda\,\chi^2(\mathbf{f})\,, \tag{1.45}$$

where $\lambda$ is a Lagrange multiplier that determines the relative weight of the entropy and $\chi^2$-statistic constraint. The maximization problem of Eq. 1.45 does not have an analytical solution and has to be solved numerically. Many algorithms have been proposed [48, 8]. They typically rely on iteratively improving a trial guess of the spectrum and simultaneously adjusting certain parameters such as $\lambda$ to arrive at a final solution.

There are certain attractions to MaxEnt. First, it does not make any assumption about the signal to be recovered, while being flexible enough to incorporate some prior information into reconstruction when such kind of knowledge is available. Second, a good initial spectrum is not required for convergence, nor an accurate estimation of the number of poles present in the signal. Third, it can be proved that there exists a unique, global solution. Fourth, data sets that are not uniformly sampled can be used, providing considerable flexibility to balance among sensitivity, resolution and experimental time. However, the claim that MaxEnt can produce simultaneous noise suppression and resolution enhancement [46] was proved to be largely unsubstantiated due to the confusions between the sensitivity and SNR of the spectrum and between the resolution enhancement and linewidth reduction [49, 8].

Nevertheless, there are also some severe limitations of MaxEnt:

(i). The efficiency and accuracy of MaxEnt hinge on the techniques used to solve the high dimensional optimization problem. While theoretically a unique solution exists, in practice, it relies on adjusting parameters such as the noise power $C_0$, and there is little guarantee that such a solution could be actually found.

(ii). MaxEnt is computationally very expensive, attempting to reconstruct the whole signal in a single shot. Similar to all the previously described high-resolution methods, it is infeasible to use MaxEnt to analyze very large data sets.

(iii). As a result of its expensive computational cost, MaxEnt is typically only applied in a trace-by-trace fashion for reconstruction of multidimensional spectra, losing the advantages of true multi-dimensional spectral analysis.

(iv). Nonlinearity is intrinsic to MaxEnt, making it hard to use MaxEnt for quantitative purpose. Some techniques such as *in situ* calibration [48] were proposed to overcome this limitation but only with limited efficiency.

(v). Even though additional information can be incorporated in MaxEnt, it is not used in the optimal way. For example, when the signal sufficiently satisfies the Lorentzian model, more aggressive methods like LP or the matrix pencil method can use this information more efficiently and provide larger resolution enhancements.

### 1.3.4 Subspace Methods

Another common limitation of all the high-resolution alternatives we just discussed is the expensive computational cost. They all attempt to analyze the whole signal in a single shot, i.e., in the *full space*, and can be prohibitively expensive and very unstable for large data sets. In this section, we will discuss an example of *subspace* methods, LP-Zoom, that can avoid this limitation. Even though due to other remaining difficulties, it is not actually successful enough for general spectral analysis in NMR, the idea of converting a large problem into small problems in the subspace is very interesting and one of the key reasons for the success of to the filter diagonalization method.

**LP-Zoom**

LP-Zoom [50] is based on LP theory but revised to analyze the local spectral structures in contrast the global structure, and thus avoid the demands for large computational time and memory of ordinary LP methods. The idea is as simple as applying the z-transform to both sides of the AR model of Eq. 1.27,

$$g_0(z) = \sum_{k=1}^{K'} a_k g_k(z) \,, \tag{1.46}$$

where

$$g_k(z) = \sum_{n=K'}^{N-1} c(n-k) z^n \,, \tag{1.47}$$

with $z = \exp(i\omega\tau)$, and $K'$ is a new prediction order. The essence of LP-Zoom is that Eq. 1.46 allows the selective evaluation in a certain (small) frequency range of interest. Because the zoomed area might contain only a few peaks, even if the whole spectrum is very complicated, the required prediction order, $K'$, is much smaller than that of conventional LP. Selectively analyzing areas that contain peaks rather than fitting the whole spectrum can significantly reduce the computational time and can also yield a more accurate result.

However, other major difficulties of LP remain unsolved: 1). one still needs to estimate the order of prediction filter, which is further complicated by the local spectral analysis as an "effective rank" could be even more difficult to define. 2). the amplitudes are still computed by solving another least square problem after all the roots are solved. The errors in the rooting step could be disastrous for the second step and are problematic to correct. 3). True multidimensional extensions still have severe difficulties.

## 1.4 Summary and Remarks

Spectral analysis is an important step in extracting the spectral information from the time signal. The conventional Fourier transform spectral analysis is a linear method that is stable, reliable and fast if implemented as fast Fourier transform. However, FT also has some well known limitations such as the FT time-frequency uncertainty principle and sinc-like truncation artifacts. Moreover, FT is essentially a one-dimensional spectral analysis method that is not able to use the important information that evolutions in all dimensions are correlated. Finally, additional information about the signal, if exists, can not be efficiently incorporated in FT.

Extensive efforts have been made to develop high-resolution alternatives to the FT spectral analysis. By utilizing some additional information about the signal, it is possible to obtain resolutions beyond the FT uncertainty principle. However, due to the large size and high complexity of NMR signals, none of these methods have been completely successful. They all suffer from one or more of the following limitations:

(i). Rely on adjusting parameters to obtain satisfactory results.

(ii). Computationally very expensive and numerically unstable.

(iii). Rely on nonlinear optimization of some complex score functions with hundreds of degrees of freedom.

(iv). Significant difficulties or limitations in their multidimensional extensions.

The Filter Diagonalization Method (FDM) is a high-resolution method that was introduced to NMR spectral analysis only very recently. It seems to be free of all the difficulties mentioned above, and shows the promise to be a final working method.

In the next two chapters, we will describe the theory of FDM and its application to NMR spectral analysis in details. Here, a preview of FDM is provided:

(i). FDM is a parametric method based on the Lorentzian model of Eq. 1.26.

(ii). FDM solves the highly nonlinear fitting problem by converting it into pure linear algebraic problems of diagonalizing some *small* data matrices in the frequency subspace. It is thus intrinsically both computational efficient and numerically stable.

(iii). In FDM, both the complex frequencies and complex amplitudes are obtained simultaneously from the same eigenvalue problem.

(iv). Multi-dimensional extensions of FDM exist and have been successfully applied to analyze realistic NMR experimental signals. The whole multidimensional data set is used by FDM to characterize the intrinsically multidimensional features. Astonishing resolution enhancements have been demonstrated.

# Chapter 2

# Theory of the Filter Diagonalization Method

## 2.1 Historical Overview

The filter diagonalization was first introduced by Neuhauser in 1990 to extract eigenvalues and eigenstates of a potentially large system in any energy range of interest [51]. In essence, a Fourier filter is applied to remove the correlation between distant eigenstates, then, by diagonalizing the Hamiltonian matrix in the subspace, the eigenstates in a small energy range can be computed very accurately. There are two advantages of filter diagonalization: first, by filtering, the size of the Hamiltonian matrix is reduced; second, by diagonalization, it is possible to obtain accuracy beyond the Heisenberg uncertainty principle, $\delta E \sim \hbar/t$, where $t$ is the length of the time propagation. One of the most successful early applications of filter diagonalization was the accurate calculation of quantum scattering resonance of the $HO_2$

radical [52], which is still very difficult to be reproduced by other methods. An important breakthrough was made in 1995 by Wall and Neuhauser [24], where the generation of a quantum time correlation function and its spectral analysis were split into two independent steps. It was discovered that as long as the time correlation function was available, the Hamiltonian matrix could be computed and diagonalized to provide the eigenstates. In this new formulation, FDM became suitable for spectral analysis of a general time signal, simply by ignoring the signal generation step. However, this perspective remained only an interesting observation until its efficiency was significantly improved by Mandelshtam and Taylor [25, 53]. The new version, named the Filter Diagonalization Method (FDM), used a rectangular filter that was computationally more efficient, and was reformulated for the conventional problem of analyzing discrete time signals defined on an equidistant time grid. It was shown to be an optimal method for fitting the signal to the summation of Lorentzian lines in several points of view including convergence, computational efficiency and numerical stability. FDM was then extensively exploited and applied to many problems in both theoretical and experimental physics and chemistry, where the harmonic inversion was essential [54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64]. In particular, FDM started to be applied to processing Nuclear Magnetic Resonance (NMR) signals [65, 66, 67], which turned out to be one of the most interesting applications of FDM.

From the spectral analysis point of view, 1D NMR does not really have limitations as very long signals are typically available and therefore converged FT spectra[1] can be obtained. However, multi-dimensional NMR signals are often truncated in the in-

[1] A converged spectrum should show peaks at right frequencies with true linewidths.

direct dimensions due to many practical limitations, leading to poor FT resolution in corresponding frequency dimensions. Therefore there is plenty of room for resolution enhancement. Interestingly, the true power of FDM actually lies in its multidimensional extensions. FDM is a true multidimensional method that is able to process the whole multi-dimensional data set to characterize the intrinsically multi-dimensional features. The obtainable resolutions in all dimensions are determined together by the total information content of the signal, instead of by the signal sizes along individual dimensions. In another word, it is possible for FDM to use the information encoded in the long dimensions to enhance the resolutions in the short dimensions. Early implementations of multi-dimensional FDM [68, 69, 70] require simultaneous diagonalization of multiple data matrices, which is very difficult, if not impossible, for noisy signals with degenerate resonances. A major breakthrough was made by Mandelshtam and coworkers [71], when a Green's function approach was introduced to compute multidimensional spectral estimations directly using the eigenvalues and eigenvectors of solving several independent 1D generalized eigenvalue problems. The difficult problem of finding a unique set of eigenvectors to diagonalize multiple data matrices is thus avoided. However, applied to noisy experimental signals, the spectra computed from a single FDM calculation are often be contaminated by many artifacts including spurious poles and distortions of genuine features, which are very sensitive to any change in the input data and FDM processing parameters. The source of the artifacts was not fully understood at first, but the sensitivity was exploited by various averaging methods to suppress the artifacts [72, 71, 73]. An obvious drawback of these averaging methods was that they were very time consuming and the convergence was slow. It was later realized that the averaging was actually an inefficient way of regularization for removing the ill-conditioning of FDM. More efficient methods of regularization were exploited, among which FDM2K [74], proposed in year 2000 and thus named FDM2K to honor the new millennium, was particular efficient and effective. With regularization implemented, stable high-resolution spectral estimations can be computed using single FDM calculations, provided that the signal sufficiently satisfies the Lorentzian model and is sufficiently long. FDM2K has since been applied to several Constant-Time (CT) NMR experiments [75, 76] and demonstrated significant resolution enhancements. There are some special properties associated with processing CT-NMR signals, which makes them particularly suitable for FDM data processing. Very short constant-time periods can be used without sacrificing the resolution, leading to better sensitivities, especially for fast relaxing proteins.

It was also realized that various spectral representations could be directly computed without diagonalizing the data matrices [77, 78]. The new expression was named Regularized Resolvent Transform (RRT). RRT corresponds a direct, nonlinear transformation of time domain signal into frequency domain spectrum. It is a very efficient method with appealing transparency and simplicity. RRT is also quite flexible and can be used to construct many non-FT types of spectral estimations.

In the rest of the chapter, the theory of 1D FDM will be first described in Section 2.2, then extended to a multi-dimensional case in Section 2.3. Various ways to regularize FDM will be discussed in Section 2.4, followed by detailed descriptions of RRT in Section 2.5. Several ways for checking the validity of FDM/RRT results will be discussed in Section 2.6. Finally, some conclusions are given in Section 2.7.

## 2.2 One-Dimensional FDM

In this section, we first review the theory of 1D FDM with emphasis on its two most important properties: solving the nonlinear fitting problem by pure linear algebra and local spectral analysis by using a Fourier-type basis. FDM is able to fit both large and complicated signals to the summation of Lorentzian lines in an efficient and stable way. In the case of of noisy data, the original FDM algorithm might yield a line list that contains unphysical poles with both large amplitude and negative linewidth. By using a more sophisticated *Multi-Scale* Fourier basis [79], this kind of poles can be eliminated and a line list more consistent with the decaying signal can be obtained.

### 2.2.1 Nonlinear Fitting by Linear Algebra

The basic object of 1D FDM is to fit a given complex time signal $c(n) = c(t_n)$, defined on an equidistant time grid, $t_n = n\tau, n = 0, 1, 2, \ldots, N-1$, to the sum of exponentially damped sinusoids (i.e., Lorentzian lines),

$$c(n) = \sum_{k=1}^{K} |d_k| e^{i\theta} e^{-in2\pi\tau f_k} e^{-n\tau\gamma_k} = \sum_{k=1}^{K} d_k e^{-in\tau\omega_k}\,, \tag{2.1}$$

with a total of $2K$ unknowns: the $K$ *complex amplitudes* $d_k = |d_k|e^{i\theta}$ and $K$ *complex frequencies* $\omega_k = 2\pi f_k - i\gamma_k$. The real part of complex frequency, $2\pi f_k$, gives the line position and negative of the imaginary part, $\gamma_k$, gives the line width. The absolute value of complex amplitude, $|d_k|$, corresponds to the integral of the peak and the phase angle, $\theta$, defines the phase of the peak. The line list $\{\omega_k, d_k\}$ corresponds to a *parametric representation* of the time signal, $c(n)$, of which each pair of $(\omega_k, d_k)$ defines a *peak*, or, a *pole*.

Although the fitting problem of Eq. 2.1 is highly nonlinear, its solutions can be obtained by pure linear algebra. In FDM, we assume that the complex time signal, $c(n)$, is associated with a time autocorrelation function of a dissipative dynamical quantum system that is described by an effective non-Hermitian Hamiltonian operator $\hat{\Omega}$ and some initial state $\Phi_0$ [24],

$$c(n) = \left(\Phi_0 \middle| \Phi_n\right) = \left(\Phi_0 \middle| e^{-in\tau\hat{\Omega}} \middle| \Phi_0\right) \ . \tag{2.2}$$

Here a complex symmetric inner product[2] is used, $(a|b) = (b|a)$ without complex conjugation. We use the round brackets to distinguish it from the Hermitian inner product, $< a|b >=< b|a >^*$. $\hat{\Omega}$ is a complex symmetric operator,

$$\{(\Psi|\}\{\hat{\Omega}|\Phi)\} = \{(\Psi|\hat{\Omega}\}\{|\Phi)\} = (\Psi|\hat{\Omega}|\Phi) \ . \tag{2.3}$$

Assuming that $\hat{\Omega}$ is diagonalizable, it can be written in its spectral representation,

$$\hat{\Omega} = \sum_k \omega_k |\Upsilon_k)(\Upsilon_k| \ , \tag{2.4}$$

where $\omega_k$ are the eigenvalues and $\Upsilon_k$, the corresponding eigenvectors. $\{\Upsilon_k\}$ are orthonormal with respect to the complex symmetric inner product, i.e.,

$$(\Upsilon_k|\Upsilon_{k'}) = \delta_{kk'} \ . \tag{2.5}$$

Inserting Eq. 2.4 into Eq. 2.2, we can recover the HIP equation of Eq. 2.1 with,

$$d_k^{1/2} = (\Upsilon_k|\Phi_0) \ . \tag{2.6}$$

Therefore the harmonic inversion problem of Eq. 2.1 becomes equivalent to the linear algebraic problem of diagonalizing the "Hamiltonian" $\hat{\Omega}$. Or equivalently, we can

[2]See Appendix at the end of this chapter for a brief overview of non-Hermitian quantum mechanics for describing dissipative systems.

diagonalize the evolution operator $\hat{U} \equiv e^{-i\tau\hat{\Omega}}$:

$$\hat{U}|\Upsilon_k) = u_k|\Upsilon_k) \ , \tag{2.7}$$

with eigenvalues $u_k = e^{-i\tau\omega_k}$. From the eigenvalues of $\hat{U}$, we can determine $\{\omega_k\}$, *line positions* and *widths*, and from the eigenvectors, $\{d_k\}$, *line integrals* and *phases*.

Even though neither $\hat{U}$ nor $\Phi_0$ is explicitly available, their matrix representation in an appropriately chosen basis set can be computed purely from the time signal $c(n)$. The basis can be chosen in many different ways. The total number of basis functions is determined by the size of the available signal. For a signal of $N$ complex points, the maximum basis size is N/2: each Lorentzian line requires two complex parameters to specify it, so that at most $M = N/2$ lines can be uniquely fitted to a signal of length $N$.

**Krylov Basis**

The simplest basis would correspond to a set of Krylov vectors $|\Phi_n), n = 0, 1, \ldots, M-1$, with $M = N/2$, generated by propagating the initial state, $|\Phi_0)$, using the effective evolution operator $\hat{U}$,

$$|\Phi_n) = \hat{U}^n|\Phi_0) = \exp(-in\tau\hat{\Omega})|\Phi_0). \tag{2.8}$$

The matrix elements of $\hat{U}$ in this basis are trivial to obtain as

$$U^{(1)}_{nm} = \left(\Phi_n|\hat{U}|\Phi_m\right) = \left(\Phi_0|U^n U U^{m+1}|\Phi_0\right) = (\Phi_0|\Phi_{n+m+1}) = c(n+m+1) \ . \tag{2.9}$$

Since the Krylov basis is not orthonormal, the overlap matrix

$$U^{(0)}_{nm} = (\Phi_n|\Phi_m) = (\Phi_0|\Phi_{n+m}) = c(n+m), \tag{2.10}$$

must also be computed. Here convenient notations $U^{(1)}$and $U^{(0)}$ are adopted to represent the evolution matrix and the overlap matrix. Note that both matrices are symmetric, which is a direct result of the complex symmetric inner product formalism. The fitting problem of Eq. 2.1 is then cast into a complex Generalized Eigenvalue Problem (GEP),

$$\mathbf{U}^{(1)}\mathbf{B}_k = u_k \mathbf{U}^{(0)}\mathbf{B}_k \ . \tag{2.11}$$

Note that the eigenvectors $\mathbf{B}_k$ are normalized with respect $\mathbf{U}^{(0)}$ as,

$$\mathbf{B}_k^{\mathrm{T}}\mathbf{U}^{(0)}\mathbf{B}_{k'} = \delta_{kk'}. \tag{2.12}$$

The complex frequencies and complex amplitudes are then,

$$\begin{aligned} \omega_k &= \frac{i}{\tau}\ln(u_k) \\ d_k^{1/2} &= \sum_{n=0}^{M-1} \left[\mathbf{B}_k\right]_n c(n) \ , \end{aligned} \tag{2.13}$$

where Eq. 2.13 follows from Eq. 2.6 by substituting

$$\left|\Upsilon_k\right) = \sum_{n=0}^{M-1} \left[\mathbf{B}_k\right]_n \left|\Phi_n\right) \ . \tag{2.14}$$

Note that the maximum basis size $M = N/2$ is always used. There are several advantages to this. First, we can thus avoid the problem of estimating the number of resonances present in the signal, which is problematic for most realistic signals. Second, all the available data are used to construct the evolution matrix and overlap matrix. Therefore, the resulting data matrices contain all the information that we can use. The rest of the problem becomes a pure mathematical problem of solving the generalized eigenvalue problem. When the true number of resonances is less than

$M$, the matrices become singular, or near singular in the case of noisy data. In these cases, eigenvalue solvers that are capable of handling ill-conditioned matrices should be used, such as QZ [80] and GUPTRI [81, 82]. Numerical experiments showed that the U matrices in 1D were actually well-behaviour enough so that stable solutions of GEP could be readily obtained by most standard eigenvalue solvers such as those from the LAPACK [83] library.

Armed with Eq. 2.9, Eq. 2.10 and Eq. 2.11, we have a simple method for solving the nonlinear fitting problem of Eq. 2.1 by pure linear algebra of solving a generalized eigenvalue problem. The latter is a very well studied problem and an unique solution is always guaranteed. Unfortunately, while the problem is certainly easy to set up, it could be very difficult to solve in practice. The size of the data matrices is determined by the signal size. For large signals, solving the generalized eigenvalue problem becomes very demanding on the computational power as well as the computer memory for storage. Large matrices might also lead to numerical problems such as round-off error and overflow. In addition, if the signal happens to contain far less than $M = N/2$ peaks ($K \ll M$), then the basis becomes over-complete and the matrices are ill-conditioned. As a result, formulated in this way, we end up solving a large and ill-conditioned linear system, of which stable solutions are very difficult, if not impossible, to obtain numerically. This is also one of the main drawbacks of all full space method such as LP and the matrix pencil method we discussed in Chapter 1. In conclusion, the simple formalism with the Krylov basis has severe limitations and cannot be applied on a regular basis to signals of size more than, say, a few thousand data points.

### 2.2.2 Local Spectral Analysis: Fourier-Type Basis

The primitive Krylov basis leads to large and ill-conditioned linear systems for realistic experimental signals, and therefore is not a practical choice. There is a way to avoid diagonalizing the huge matrices in single step. By applying a special unitary transformation to the data matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(0)}$, the diagonalization can be carried out in multiple steps. In other words, instead of trying to analyze the whole spectrum at one time, we can divide it into small spectral windows and analyze them separately. This local spectral analysis is the second important property of FDM.

Any linear combinations of the primitive Krylov basis functions can also serve as a basis. Among them, a good choice is a Fourier transform of the Krylov basis, of which a particularly simple and efficient variant is the rectangular window Fourier basis [53], defined as [3]:

$$|\Psi_j) = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} |\Phi_n) \ , \tag{2.15}$$

with $\{\varphi_j\}$ being a set of equidistant values taken inside a small frequency window of interest. For this choice the transformation from the Krylov basis $\{\Phi_n\}$ , $n = 0, 1, 2, ..., M-1$, to the Fourier basis $\{\Psi_j\}$ , $j = 1, 2, ..., M$, is unitary. More importantly, due to the Fourier transform, each basis function $|\Psi_j)$ is localized in the frequency domain, i.e., it is a linear combination of only those eigenfunctions $|\Upsilon_k)$ of $\hat{\Omega}$ for which $\omega_k \sim \varphi_j$. This implies that we can consider a small subset of, say, $K_{\text{win}} \ll M$ values $\varphi_j$ in the frequency region $[\omega_{\min}, \omega_{\max}]$ where the eigenfunctions

[3] Other window functions can also be used in constructing the Fourier basis. However, they are mathematically equivalent to the case of no window function, and practically lead to little numerical difference. See Appendix III for more details.

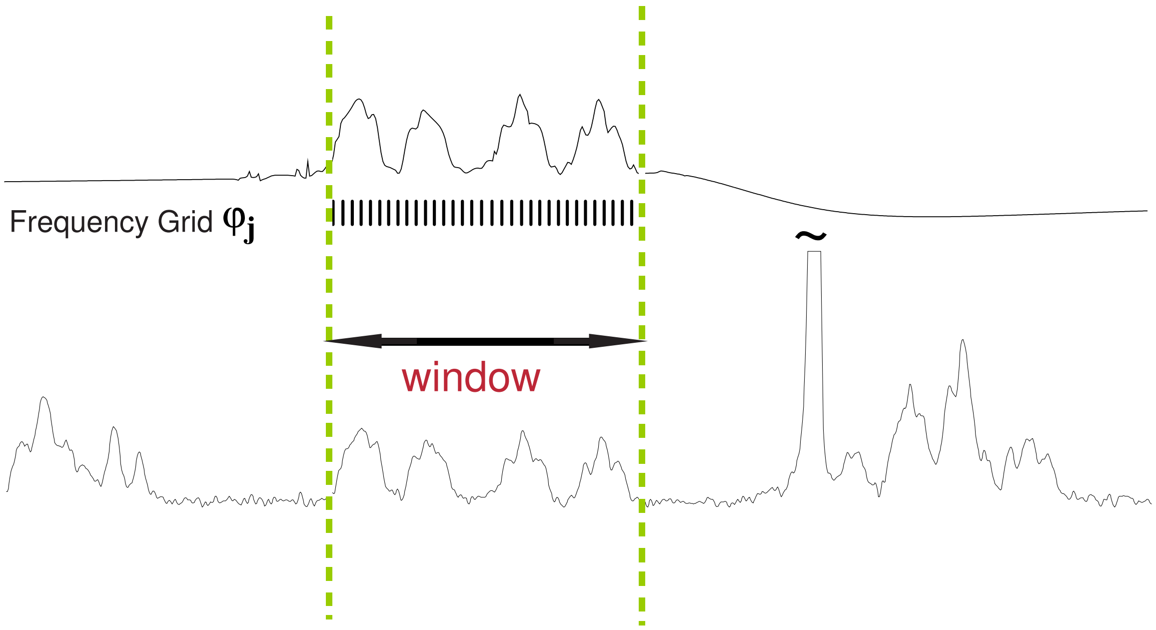



Figure 2.1: An example of FDM single window calculations. Using a small set of Fourier-Type basis functions inside the specified frequency range, the resulting data matrices contain mostly the spectral information of the local spectral window. The signal is a 1D NMR signal of human progesterone with N=2000 points and spectral width of 4 ppm.

can be sufficiently represented by,

$$|\Upsilon_k) = \sum_{\omega_k \sim \varphi_j} \left[\tilde{\mathbf{B}}_k\right]_j |\Phi_j) \ . \tag{2.16}$$

Note that a tilde ($\sim$) is added to notation $\mathbf{B}$ because $|\Upsilon_k)$ is now expanded in a Fourier-type basis. It will also be added to notation $\mathbf{U}$, the matrix representation of $\hat{U}$ operator in the Fourier basis.

Eq. 2.16 means that the operator $\hat{U}$ can be diagonalized in the Fourier subspace that corresponds to some pre-specified frequency window to yield the eigenvalues and eigenvectors, and subsequently, the spectral parameters for that window. Figure 2.1 illustrates such a window calculation. Note that the extraction of the complex frequencies $\omega_k$ to a high precision only requires that the *local completeness condition*,

$$\rho(\varphi_j) = \frac{M\tau}{2\pi} \geq \rho(\omega_k) \ , \tag{2.17}$$

be satisfied for the densities of the grid points $\rho(\varphi_j)$ and that of the eigenfrequencies $\rho(\omega_k)$. In other words, the number $K_{\rm win}$ of the basis vectors $\Phi_j$ should be larger than the number of the eigenvalues in the small interval $[\omega_{\rm min}, \omega_{\rm max}]$. Moreover, the local spectral analysis to certain extent is insensitive to the spectral properties outside the chosen small spectral domain. Therefore, FDM has a *local convergence* property. Even when the signal is so short that some crowded regions can not be fully resolved, accurate results can still be obtained for spectral regions that are less crowded. This property will be demonstrated with more examples when we discuss 2D FDM ant its applications.

The idea of local spectral analysis was not first invented in FDM. The idea of making high-resolution spectral analysis local, called "beamspacing" [84] in digital signal processing literature, has been known for a long time and widely exploited in various algorithms such as MUSIC and ESPRIT [3]. In NMR data processing literature, there is also a variant of the beamspace idea that is called LP-ZOOM [50]. LP-Zoom is a modified linear prediction algorithm which uses $z$-transform to zoom in on a small spectral region in the frequency domain and thus reduces the size of the problem. In principle, it shares the many advantages of local spectral analysis with 1D FDM. However, other difficulties of LP, such as estimating the prediction order and accurate computation of the amplitudes (see Section 1.3.4 for more details), remains unaddressed.

The matrix elements of the evolution operator in the Fourier basis can also be evaluated purely in terms of the time signal $c(n)$:

$$
\begin{aligned}
\left[\tilde{\mathbf{U}}^{(p)}\right]_{jj'} &= \left(\Psi_j\middle|\hat{U}^{(p)}\middle|\Psi_{j'}\right) \\
&= \sum_{n=0}^{M-1}\sum_{n'=0}^{M-1} e^{in\tau\varphi_j} e^{in'\tau\varphi_{j'}} \left(\Phi_n\middle|\hat{U}^{(p)}\middle|\Phi_{n'}\right) \\
&= \sum_{n=0}^{M-1}\sum_{n'=0}^{M-1} e^{in\tau\varphi_j} e^{in'\tau\varphi_{j'}}\, c(n+n'+p) \qquad (2.18)\\
&= \sum_{n=0}^{M-1}\sum_{n'=0}^{M-1} e^{i(n+n')\tau\varphi_j} e^{in'\tau(\varphi_{j'}-\varphi_j)}\, c(n+n'+p)\ . \qquad (2.19)
\end{aligned}
$$

The double sum in Eq. 2.18 can be simplified to a single sum by first changing the variables from $(n, n')$ to $(l = n + n', n')$, Eq. 2.19, and breaking it into two terms,

$$
\sum_{n=0}^{M-1}\sum_{n'=0}^{M-1} = \sum_{l=0}^{M-1}\sum_{n'=0}^{l} + \sum_{l=M}^{2(M-1)}\sum_{n'=l-(M-1)}^{M-1}\ ,
$$

then analytically evaluating the summations over $n'$, which gives,

$$
\begin{aligned}
\left[\mathbf{U}^{(p)}\right]_{jj'} &= \frac{1}{1-e^{i\tau(\varphi_{j'}-\varphi_j)}}\left\{\sum_{n=0}^{M-1} e^{in\tau\varphi_j} c(n+p)\right.\\
&- e^{i\tau(\varphi_{j'}-\varphi_j)}\sum_{n=0}^{M-1} e^{in\tau\varphi_{j'}} c(n+p)\\
&+ e^{-i(M-1)\tau(\varphi_{j'}-\varphi_j)}\sum_{n=M}^{2(M-1)} e^{in\tau\varphi_{j'}} c(n+p)\\
&- \left. e^{iM\tau(\varphi_{j'}-\varphi_j)}\sum_{n=M}^{2(M-1)} e^{in\tau\varphi_j} c(n+p)\right\}\ . \qquad (2.20)
\end{aligned}
$$

Note that the final result Eq. 2.20 is symmetric with respect to the interchange of indices $j$ and $j'$ as it should be. We can rewrite Eq. 2.20 in a compact form that can be easily generalized to the multi-dimensional cases,

$$
\begin{aligned}
\left[\tilde{\mathbf{U}}^{(p)}\right]_{jj'} &= \hat{S}\sum_{\sigma=0,1}\frac{e^{i\sigma[\tau M(\varphi_{j'}-\varphi_j)+\pi]}}{1-e^{i\tau(\varphi_{j'}-\varphi_j)}} \qquad (2.21)\\
&\times \sum_{n=\sigma M}^{(\sigma+1)(M-1)} e^{in\tau\varphi_j} c(n+p)\ ,
\end{aligned}
$$

where $\hat{S}$ defines symmetrization operator over the indices $j$ and $j'$:

$$\hat{S}\ g_{jj'} = g_{jj'} + g_{j'j}\ . \tag{2.22}$$

In principle, Eq. 2.21 is correct for all choices of $\varphi$ and $\varphi_{j'}$ except for the singularity arising at $\varphi_j = \varphi_{j'}$. To obtain a numerically practical expression for this singular case, we evaluate the $\varphi_j \to \varphi_{j'}$ limit, which leads to,

$$\left[\tilde{\mathbf{U}}^{(p)}\right]_{jj} = \sum_{n=0}^{2(M-1)} e^{in\tau\varphi_j}(M - |M-n-1|)c(n+p)\ . \tag{2.23}$$

Notably and quite importantly, the resulting matrices $\tilde{\mathbf{U}}^{(p)}$ have a sinc-like structure with generally large diagonal and decaying off-diagonal terms. The latter become much smaller than the former once $M\tau|\varphi_j - \varphi_{j'}| \gg 2\pi$. It is this structure that justifies the possibility to perform the eigenvalue calculation in a small $K_{\text{win}} \times K_{\text{win}}$ block fashion for possibly large $M \times M$ matrices $\tilde{\mathbf{U}}^{(p)}$.

After the new matrix representations are computed, the same generalized complex eigenvalue equation of Eq. 2.11 can be solved. From the eigenvalues and eigenvectors, we can compute the complex frequencies and complex amplitudes. According to Eqs. 2.6 and 2.15 the amplitudes $d_k$ are given as[4],

$$d_k^{1/2} = \tilde{\mathbf{B}}_k^{\mathrm{T}}\tilde{\mathbf{C}}\,, \tag{2.24}$$

where $\tilde{\mathbf{B}}_k^{\mathrm{T}}$ is the transpose of eigenvectors, and the $1 \times K_{\text{win}}$ column vector $\tilde{\mathbf{C}}$ is the FT of the original $1 \times M$ signal array $\mathbf{C}$:

$$\left[\tilde{\mathbf{C}}\right]_j = \sum_{n=0}^{M-1} e^{in\tau\varphi_j}c(n)\ ,\quad j = 1, 2, ..., K_{\text{win}}\ . \tag{2.25}$$

[4]There exists an alternative expression for computing the amplitudes, which is more accurate for narrow poles in the case of high signal to noise ratio. See Appendix II for details.

### 2.2.3 FDM as a High-Resolution Spectral Estimator

Up to this point, FDM seems to be just a parameter estimator which can provide the parameter representation $\{w_k, d_k\}$ for a given FID by fitting it into Eq. 2.1. However, FDM can also be used as a spectral estimator. Once the parameters are available, one can easily construct the spectral representation $I(w)$, formally defined as the infinite time discrete Fourier Transform (DFT) summation,

$$I(\omega) = \tau \sum_{n=0}^{\infty} \left(1 - \frac{\delta_{n0}}{2}\right) C(n) e^{i\omega n\tau} , \tag{2.26}$$

where the term $\left(1 - \frac{\delta_{n0}}{2}\right)$ multiplies $c_0$ by $\frac{1}{2}$ to correct the error introduced by the discrete sum approximation of the continuous half-line Fourier integral (so called "first point correction"). By substituting the damped sinusoids model Eq. 2.1 into Eq. 2.26 and evaluating the geometric summation analytically, we can obtain the corresponding spectral estimate, which we called the FDM *ersatz spectrum*,

$$I^{\tau}(\omega) = \tau \sum_{k=1}^{K} d_k \left( \frac{1}{1 - e^{i\tau(\omega - \omega_k)}} - \frac{1}{2} \right) . \tag{2.27}$$

In FDM, some poles with extremely small imaginary part may occur (either due to the present of noise or real peaks), which leads to high spikes in the *ersatz* spectrum producing unfavorable results. Therefore, it is often useful to include a smoothing parameter $\Gamma > 0$ to improve the appearance of the spectra,

$$I^{\tau}(\omega) = \tau \sum_{k=1}^{K} d_k \left( \frac{1}{1 - e^{i\tau(\omega - \omega_k + i\Gamma)}} - \frac{1}{2} \right) . \tag{2.28}$$

The absorption mode *ersatz spectrum* then corresponds to,

$$A^{\tau}(\omega) = \mathrm{Re}[I^{\tau}(\omega)] = \tau \sum_{k=1}^{K} \mathrm{Re} \left[ d_k \left( \frac{1}{1 - e^{i\tau(\omega - \omega_k + i\Gamma)}} - \frac{1}{2} \right) \right] . \tag{2.29}$$

In a single calculation, FDM can only obtain results for a small frequency window. The construction of the spectral estimation for a larger spectral range requires combining the results from multiple window calculations. To reduce the inaccuracies at the window edges, the adjacent windows overlap with each other, typically by 50%. We first construct segments of the spectrum for each windows, then sum them up with appropriate weighting functions, such as cosine square functions, that add up to one,

$$I^{\tau}(\omega) = \sum_{c} g_c(\omega) I_c^{\tau}(\omega) \ , \tag{2.30}$$

where the subscript $c$ represents different spectral windows, and the weighting functions $g_c(\omega)$ satisfy $\sum_c g_c(\omega) = 1$. Figure 2.2 illustrates such a multi-window implementation of FDM. Cosine square functions (shown as dotted lines) are used to weight the segments before they are concatenated together, so that the center parts of each window segments will contribute more to final spectrum than the edges. Note that for windows on the two edges the large spectral range of interest, only half of the corresponding segments of the spectrum are retained.

### Practical Issues on Spectral Estimation via FDM

There are several additional complications involved in constructing a stable spectral representations using the spectral parameters computed by FDM.

1. $I^{\tau}(\omega)$ or $I(\omega)$?

For a very long time after the invention of FDM, a spectral representation $I(w)$ that corresponds to the infinite time Fourier integral had been used. Substituting Eq. 2.1 into the infinite time FT integral (see Eq. 1.1) and evaluating the integration

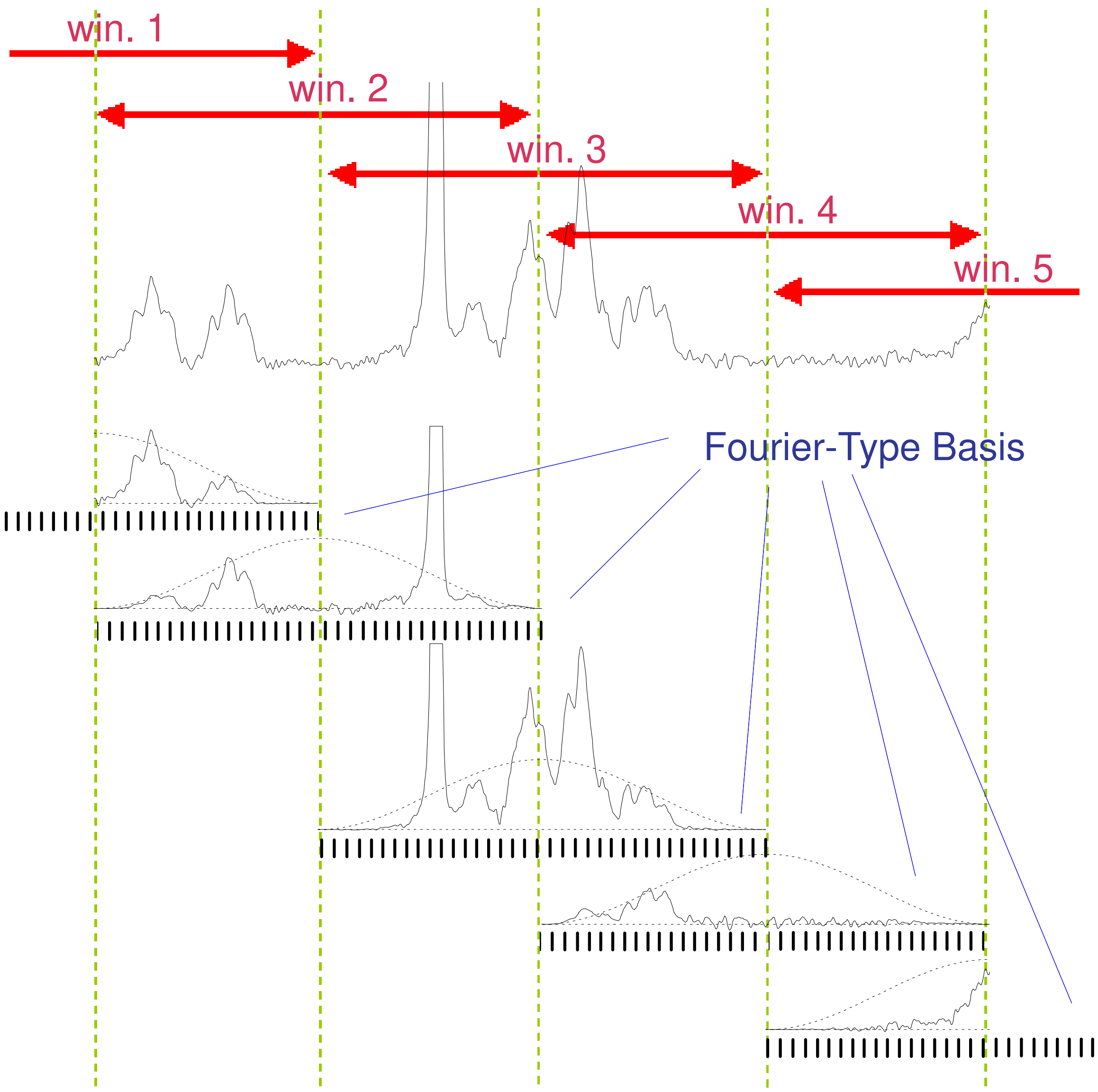



Figure 2.2: A multi-window implementation of FDM. Cosine-square functions, shown as dotted lines, are used for weighting the segments of the spectrum before they are concatenated together.

analytically yields,

$$I(\omega) = i \sum_{k=1}^{K} \frac{d_k}{\omega - \omega_k} \ . \tag{2.31}$$

Generally, for small $\tau$ we have $I(\omega) \approx I^{\tau}(\omega)$, but unlike $I^{\tau}(\omega)$, which is periodic in $\omega$ with the period equal to the Nyquist width $2\pi/\tau$, $I(\omega)$ is not periodic. However, as we are dealing with discrete time signals, the resulting spectra should be periodic. When there are some broad poles such that $\gamma_k \tau << 1$ is not satisfied, $I(\omega)$ becomes a bad approximation to DFT spectrum. The mutual interferences and cancellations of the poles are not conserved, leading to some instabilities in the baseline. Numerical experiments showed that broad poles occurred frequently when FDM was applied to analyze noisy signals and/or signals with a strong background. For example, Figure. 2.3 compares the FDM *ersatz* spectra obtained using Eq. 2.31, trace (b), and Eq. 2.27, trance (c). While the $I^{\tau}(\omega)$ spectrum is very stable, the $I(\omega)$ spectrum shows significant baseline distortions. This demonstrates that Eq. 2.31 is dangerous to use if there are interfering poles $\omega_k \equiv \nu_k - i\gamma_k$ with both large widths $\gamma_k$ and large amplitudes $d_k$. With the slightly wrong formula (Eq. 2.31 rather than Eq. 2.27) the negative contributions are not correctly canceled by the positive ones, leading to the baseline distortions. Similar discussions were also independently given before in the nonlinear least squares fitting literature [30]. In conclusion, $I^{\tau}(\omega)$ should always be constructed.

2. What should we do with unphysical poles with negative $\gamma_k$?

Due to the imperfections of the signals such as noise and lineshape distortions, and certain limitations of FDM, poles with negative $\gamma_k$ (or negative linewidth) often appear in the line list. This formally corresponds to unphysical exponential increasing signals. Now we are faced with a dilemma on what to do with such kind of poles. There are a few options.

a. Do nothing.

b. Simply flip the sign of all negative $\gamma_k$.

c. Throw the unphysical poles away.

This dilemma also occurs in linear prediction. In LP, the amplitudes $d_k$ are computed by solving a least squares problem after the frequencies $\omega_k$ are calculated by rooting a polynomial. An $\omega_k$ with negative $\gamma_k$ can blow up all the amplitudes. Thus, generally, a combination of (b) and (c) has to used: the $\omega_k$ with small $\gamma_k$ are retained with the sign of $\gamma_k$ flipped, while the $\omega_k$ with large and negative $\gamma_k$ values are rejected. Even when LP is only used to extrapolate the signal, it is always necessary to first root the characteristic polynomial and then eliminate roots corresponding to exponentially growing signals, in order to obtain a stable, meaningful extrapolation.

In FDM the situation is quite different as the amplitudes $d_k$ are computed simultaneously with the frequencies $\omega_k$ using the eigenvectors and eigenvalues from the same generalized eigenvalue problem. Therefore, $\gamma_k < 0$ does not necessarily lead to any numerical instability in FDM. In the early applications of FDM [53, 67], $I(\omega)$ was always constructed. A narrow peak with negative $\gamma_k$ will appear upside-down in the absorption spectrum, which can be easily fixed by flipping the sign of $\gamma_k$. However, because of the local nature of the Fourier basis, which is manifestly incomplete, some computed pairs $(\omega_k, d_k)$ may have large and negative $\gamma_k$ and large $d_k$ values. When $I(\omega)$ is constructed, these entries would often result in a noticeable baseline distortion, as demonstrated by trace (b) of Figure. 2.3. Neither throwing away such a "spurious" entry nor flipping the $\gamma_k$ would fix the baseline.

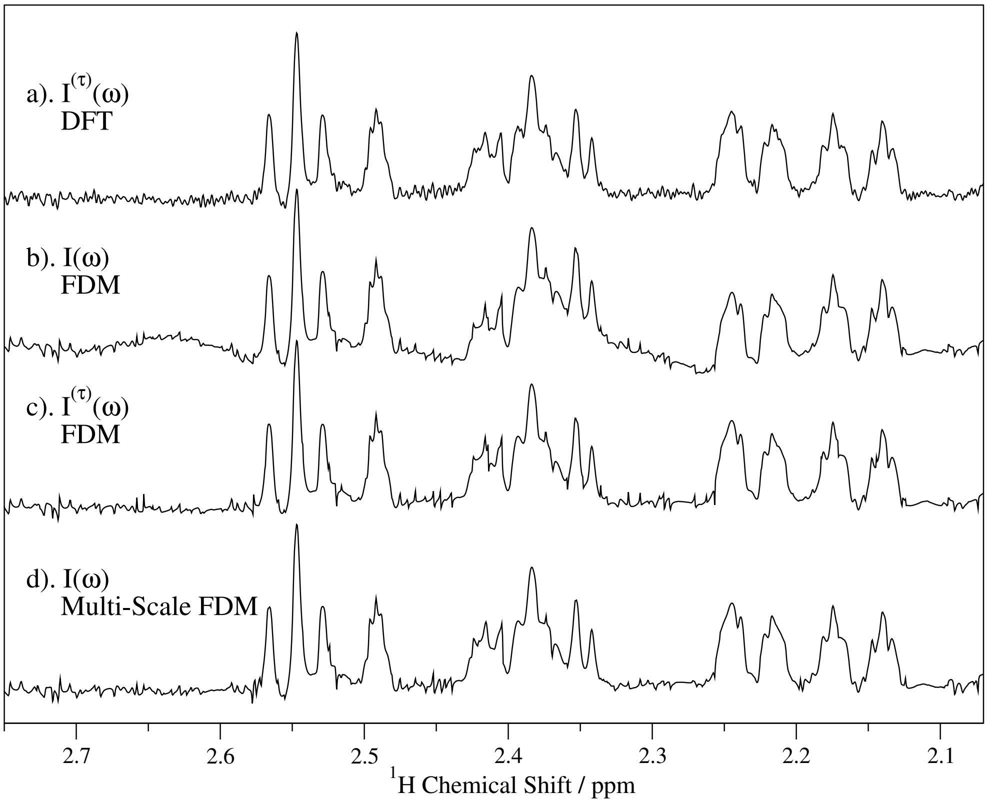



Figure 2.3: 1D proton NMR absorption spectra of progesterone for a representative spectral region of a noisy signal. The signal length is $N = 2000$ and spectral width is $SW = 4$ ppm (2 KHz). All the entries returned by FDM were used to reconstruct the *ersatz spectrum* without any attempt to identify and/or throw away any spurious poles. From the top to the bottom, the traces are,
(a). DFT with appropriate apodization function.
(b). FDM *ersatz spectrum* computed using Eq. 2.31 with single-scale FDM.
(c). FDM *ersatz spectrum* computed with Eq. 2.27 with single scale FDM.
(d). FDM *ersatz spectrum* computed using Eq. 2.31 Eq. 2.31 with *Multi-Scale* FDM.

Trace (c) of Figure 2.3 was a little unexpected in that $I^{\tau}(\omega)$ constructed with the line list $\{\omega_k, d_k\}$ containing some "spurious" entries leads to the correct spectrum with undistorted baseline. It can be explained by the fact that it is the $I^{\tau}(\omega)$, not

$I(\omega)$, which is invariant to the basis set representation. Due to the existence of very delicate mutual interferences and cancellations of the contributions all poles in Eq. 2.27, the instabilities do not occur if one keeps all the entries produced by FDM. This also explains the stability of the RRT (see Section 2.5) which relies on the same DFT based expression. Clearly, if $I^{\tau}(\omega)$ gives the correct baseline in the presence of such poles, $I(\omega)$ must fail as the two expressions are quite different for large $\gamma_k$. Accordingly, when possible,

- *All the poles with significant amplitudes $d_k$ must be retained in the sum.*

- *Only narrow poles with $\gamma_k < 0$ should be flipped: $\gamma_k \to -\gamma_k$.*

Note that even though the spectrum $I^{\tau}(\omega)$ looks correct, the underlining broad poles with negative $\gamma_k$ and large $d_k$ do not correctly describe the time domain data. Therefore, for obtaining a line list consistent with the decaying time signal, and with minimized cancellation effects, the use of *Multi-Scale* FDM is advantageous, which will be described in the next section.

### A Numerical Example

To demonstrate the ability of FDM to identify and resolve Lorentzian lines with a wide range range of linewidths and splittings, a model signal named "Jacob's Ladder" [67] is constructed. The signal contains 50 triplets with progressively smaller linewidths and splittings from high frequency to low frequency,

$$\begin{aligned} c(n) = \sum_{k=0}^{50} \quad & e^{-i2\pi\times(0.001n)\times(500.0-1.0i)\,0.9^k} + \\ & 2\,e^{-i2\pi\times(0.001n)\times(497.5-1.0i)\,0.9^k} + e^{-i2\pi\times(0.001n)\times(495.0-1.0i)\,0.9^k} \end{aligned} \tag{2.32}$$

A total of $N = 64,000$ complex data points were produced. Figure 2.4 shows the spectrum of the model signal. Due to the wide range of linewidths and splittings, accumulation of lines at low frequency and large size of the signal make it a very chal-

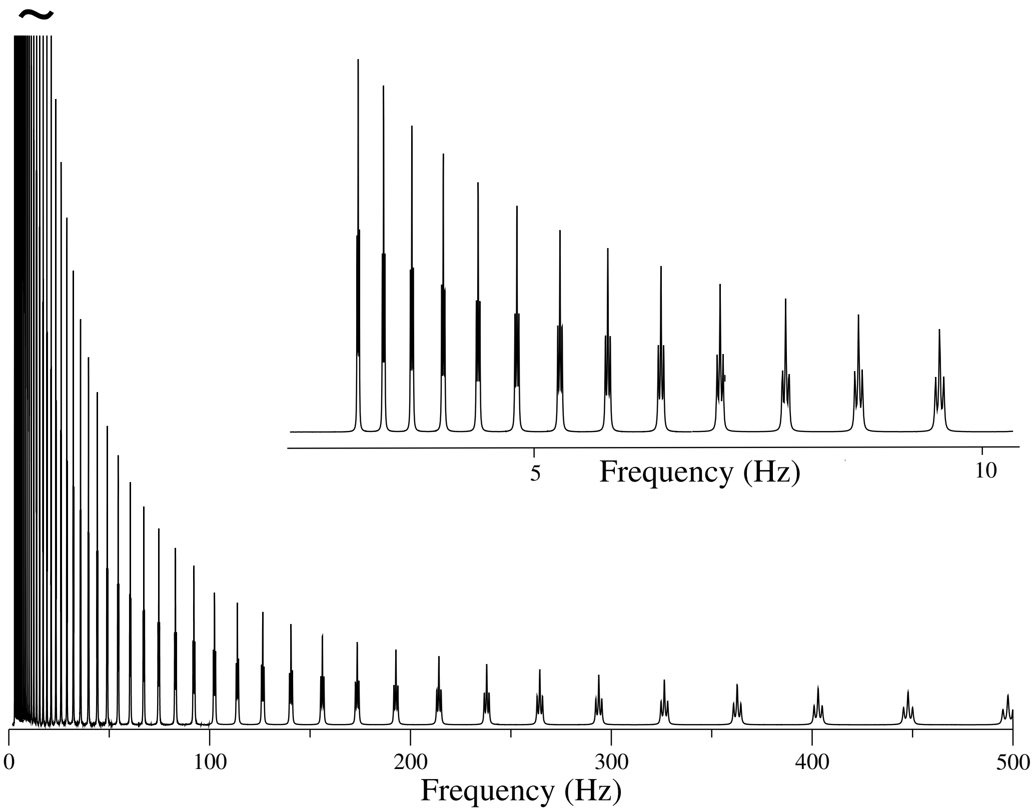



Figure 2.4: A total of 50 triplets: beginning with 1.0 Hz linewidth and 2.5 Hz coupling centered at 497.5 Hz and finishing at 5.73 mHz linewidth and 14.3 mHz coupling centered at 2.82 Hz. The spectral width is 1 KHz. The time signal contains 64,000 data points. The wide range of linewidth and splittings, accumulation of lines near zero frequency, and large size make this signal challenging for non FT methods.

lenging signal for any other non-FT methods such as Linear Prediction and Maximum Entropy Reconstruction.

Figure 2.5 compares the DFT and FDM on the densest part of the spectrum of the test signal, using 32K and 64K data points. It demonstrates the ability of FDM to fit both large and complicated time domain signals to the summation of Lorentzians, and provide resolution beyond the FT uncertainty principle. Due to the local spectral analysis, it took only a few seconds to obtain these results on a small PC workstation.

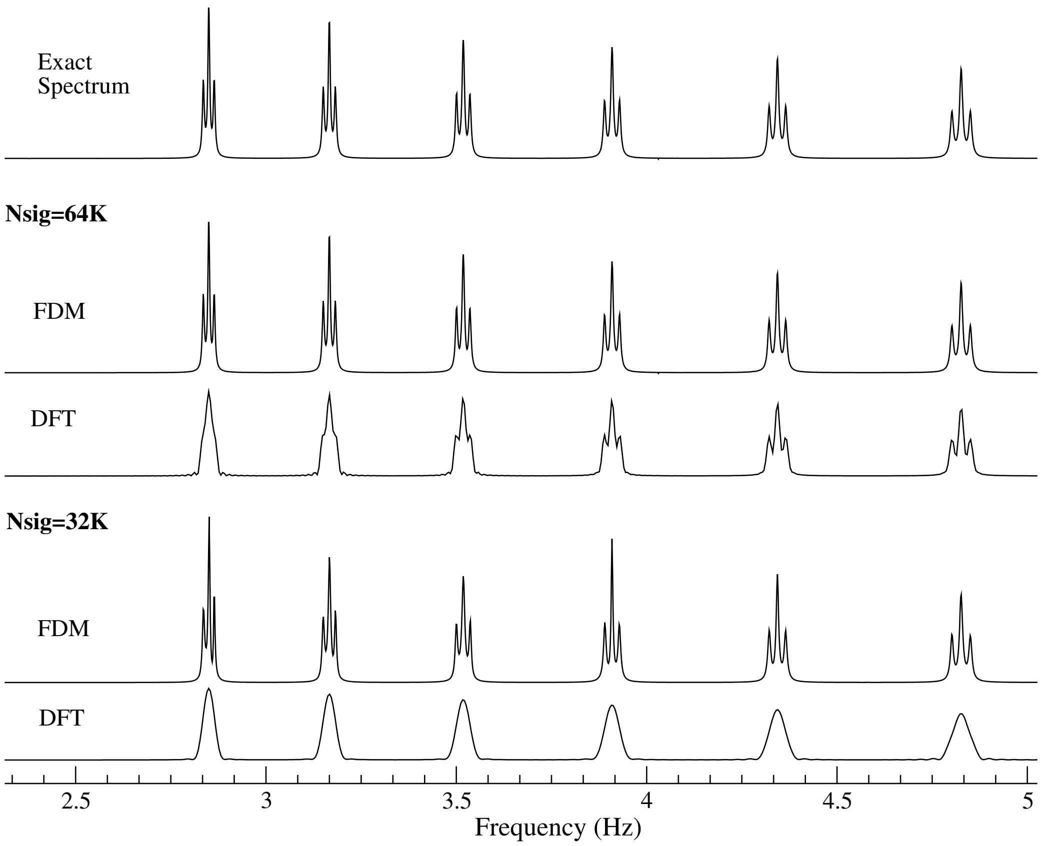



Figure 2.5: A comparison of FT and FDM on the densest part of the signal "Jacob's Ladder" (Figure 2.4). In FT calculations, the signal was first zero-filled to 128K data points. FDM resolves all the multiplets with signal length of 32K points, while the left most triplets are not resolved by FT even with 64K data points.

### A Case Where FT Beats FDM

We have demonstrated that FDM is able to effectively use the information that the signal contains purely Lorentzian lines to provide resolution beyond the FT uncertainty principle. However, this does not mean that FDM will provide better resolution for all Lorentzian signals. It is most advantageous for FDM when the distribution of the peaks is highly non-uniform, and highly disadvantageous for FDM when the peaks are uniformly distributed. To demonstrate the point, a special model signal that contains 21 peaks evenly distributed in the center part of the spectrum is constructed.

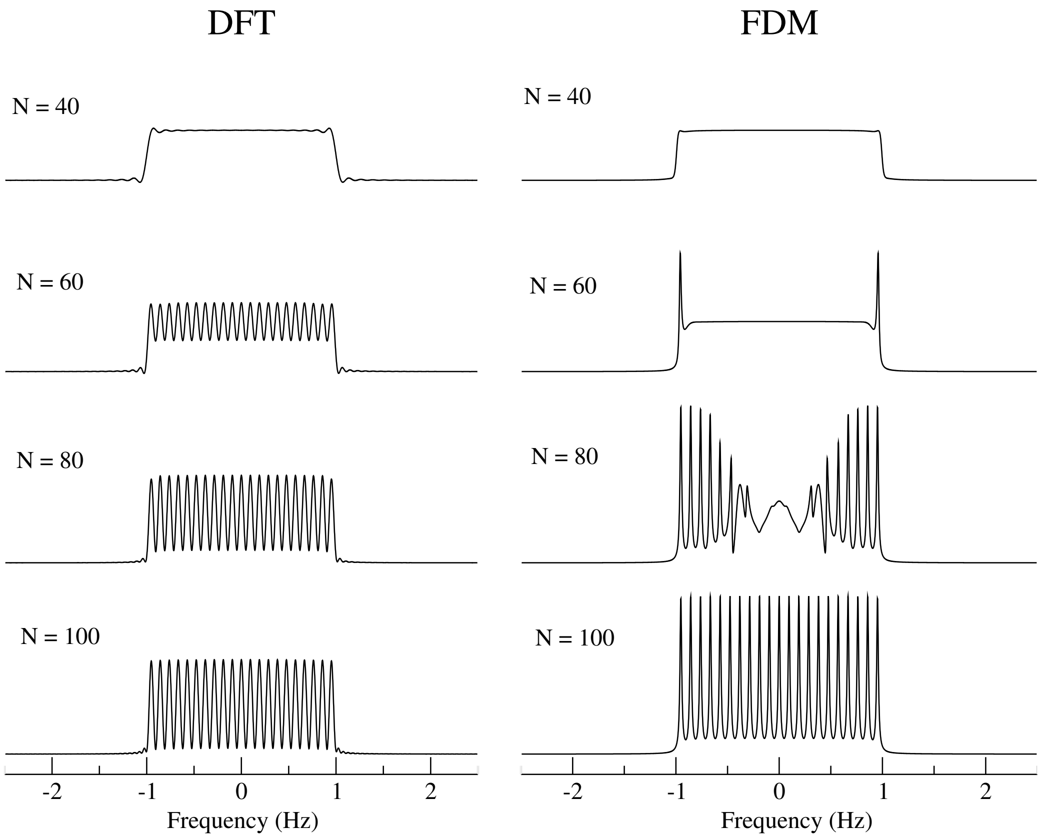



Figure 2.6: A special case where DFT can beat FDM even though the signal contains purely Lorentzian lines. The model signal is integerized with the first point $c_0 = 2048$.

The signal is integerized to mimic ADC with the first point being $c_0 = 2048$. Even though this signal contains purely Lorentzian lines, it is optimal for DFT processing, because DFT provides uniform resolution and thus works the best when the features are uniformly distributed. Figure 2.6 compares the performance of DFT and FDM on this signal. It shows that DFT is able to resolve all 21 peaks using as few as $N = 60$ data points, while FDM does not converge until $N \geq 100$. However, note that even though DFT is able to resolve all 21 peaks for $N \geq 60$, the linewidths converge very slowly, while FDM is able recover all parameters very accurately once the size of the signal is above certain threshold.

### Summary of the FDM Algorithm

The practical numerical procedure for implementing the FDM algorithm can be summarized as the following steps:

1. Choose a spectral window: specify a small enough frequency range of interest, $[\omega_{min}, \omega_{max}]$, inside the Nyquist range for a given time signal of length $N$, $c_n = c(n\tau), n = 0, 1, \ldots, N-1$. "Small enough" is operationally defined by the resulting size of the linear algebraic problem, which could depend on the type of the signal and computational power available.

2. Set up the Fourier basis: set up an evenly spaced angular frequency grid inside the specified window, $\omega_{min} < \varphi_j < \omega_{max}, j = 1, 2, \ldots, K_{win}$, where the basis size $K_{win}$ is determined by $K_{win} = \rho\, M(\omega_{max} - \omega_{min})\,\tau/2\pi$. $M = N/2$ is the maximum size of basis. $\rho \geq 1$ is the basis density, defined in Eq. 2.33. Typically $\rho = 1.1 \sim 1.2$.

3. Compute the matrix elements using Eqs. 2.21 and 2.23. The resulting matrices, $\tilde{\mathbf{U}}^{(1)}$ and $\tilde{\mathbf{U}}^{(0)}$, are complex symmetric and of size $K_{win} \times K_{win}$.

4. Solve the generalized eigenvalue problem of Eq. 2.11 using any standard eigenvalue solvers such as CG [83] and QZ [80] routines.

5. Compute the complex amplitudes according to Eq. 2.24.

6. Use the line list $\{\omega_k, d_k\}$ as an input for the spectral estimation (optional).

7. If of interest, go to step (1) choosing the next frequency window.

Note that in the multi-window implementation, large spectral range can be specified, which is then automatically divided into multiple smaller windows.

**Computational Cost of 1D FDM**

The most time consuming steps are the matrix construction, step (3), and diagonalization, steps (4). Close inspection of Equations 2.21 and 2.23 shows that for an arbitrary choice of 1D grid of size $K_{win}$, the evaluation of all $\tilde{\mathbf{U}}^{(p)}$ elements scales as $K_{win} \times N$: linear with respect to the size of the time signal and the basis size. The cost of the diagonalization step scales as $K_{win}^3$. Therefore the overall computational cost of FDM is $a\,K_{win} \times N + b\,K_{win}^3$, where $a$ and $b$ are scaling factors. In the case of small spectral windows, $K_{win}^2 << N$, the speed determining step is the U matrices evaluation step, which is often the case for 1D FDM. In order to process the whole spectral range, approximately $N/K_{win}$ window calculations are needed. Thus, the total cost of processing a 1D signal roughly scales as $(N/K_{win}) \times (K_{win}N) = N^2$, if small windows are used. Therefore, the computational cost of FDM is comparable to that of Discrete FT, although still more expensive than $N \log(N)$ of FFT. For example, for a signal of 10000 complex data points, FDM takes a few seconds to analyze the whole spectral range using a basis size of 50 on single AMD XP 1800+ workstation, which cost less than \$500 in May of 2002.

### 2.2.4 *Multi-Scale* Fourier-Type Basis

As mentioned above, unphysical poles with large and negative linewidths and large amplitudes can occur in FDM, manifested as the baseline instabilities of the

$I(\omega)$ spectrum. This can be due to the narrow band Fourier basis functions, Eq. 2.15, which are highly localized in the frequency domain. On one hand, localization is a big advantage of using the Fourier basis, on the other hand, in the presence of noise, a narrow band basis is inadequate to represent a broad spectral feature that is not localized in a single window, as each basis function $|\Psi_j)$ is "locked" on the noise peaks that appear close to $\varphi_j$ and cannot "see" the peaks that are far away. For example, if there exists a broad peak that centers outside of the current spectral window but tails into the window, the narrow band Fourier basis would not be enough to accurately represent the contribution of this broad peak. Thus, with the narrow band Fourier basis only, there might be some ambiguity in reproducing the broad spectral features or the baseline, and this is the source of the instability. Obviously, it can sometimes be reduced by using larger windows. However, increasing the basis size does not always work, and even when it does, it is impractical due to the unfavorable scaling of the computational cost for an eigenvalue solver. In this section, we will introduce a more sophisticated way of local spectral analysis using a *Multi-Scale* Fourier-type basis. The improved algorithm, named *Multi-Scale* FDM, is more robust and reliable, and at the same time computationally efficient.

The idea is illustrated in Figure 2.7. In addition to the narrow band Fourier basis functions inside the window, we add some *coarse* (i.e., delocalized or broad band) basis functions outside of the window, so that the global behaviour of the spectrum can be captured, and its interference on the window calculations minimized. Since these global features such as broad peaks and background do not require high resolution we do not need many of such coarse basis functions. We also want to minimize the

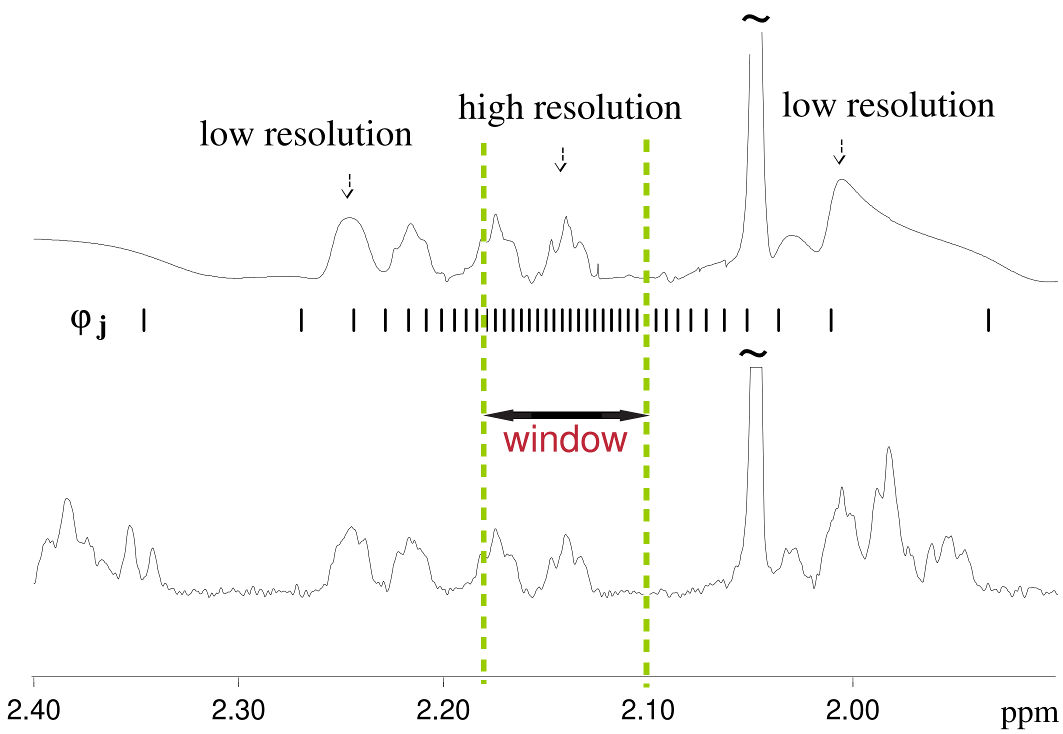



Figure 2.7: An example of a multi-scale basis set and the spectrum obtained using this basis for the same signal used in Fig. 2.1. $K_{\text{win}} = 10$ narrow band and $K_c = 20$ coarse basis functions (indicated by an impulse at each $\varphi_j$) were used. The coarse functions are distributed non-uniformly according to the displacement from the edges of the window (see text).

number of coarse basis functions with the condition that the non-localized features that may affect the local spectral analysis are represented adequately, which can be achieved by choosing the most efficient coarse basis distributions.

Given that the farther away from the current window, the less significant the effect of a broad spectral feature to the local analysis, the coarse basis could be chosen such that the spacing between the adjacent coarse basis functions and, accordingly, their bandwidth, monotonically increase with respect to the distance between the center of the basis function ($\varphi_j$) and the spectral window of interest. Thus only the features which are broad enough to affect the current window will be captured by the coarse basis. Here the local density of the basis functions is defined as

$$\rho_j = \frac{2\Delta\varphi_{\min}}{|\varphi_{j+1} - \varphi_{j-1}|} \tag{2.33}$$

where $\Delta\varphi_{\min} = 2\pi/\tau M_{\max}$ is determined by the spacing between the narrow band Fourier basis functions. The Fourier length should be scaled accordingly to the local density using $M_j = \rho_j M_{\max}$, so that the bandwidth of the basis functions is consistent with the basis density. For a given small spectral window we can construct a multi-scale basis, which contains $K_{\text{win}}$ narrow band Fourier basis functions localized in the window and $K_{\text{c}}$ coarse basis functions spread over a wide spectral range:

$$|\Psi_j) = \sum_{n=0}^{M_j-1} e^{in\tau\varphi_j}|\Phi_n), \;\; j = 1, 2, \ldots, K = K_{\text{win}} + K_{\text{c}} \, , \tag{2.34}$$

with the Fourier length $M_j$ depending on $j$: $M_j = M_{\max}$ for the narrow band, and $M_j = M_{\max}\,\rho_j$, for the coarse basis functions.

The elements of $\mathbf{U}$ matrices can be then evaluated in terms of the time signal. Using the definition of multi-scale FT basis functions $|\Psi_j)$ and $|\Psi_{j'})$, Eq. 2.34, and the assumption of Eq. 2.2, we obtain,

$$\begin{aligned} \left[\mathbf{U}^{(p)}\right]_{jj'} &= \sum_{n=0}^{M_j-1}\sum_{n'=0}^{M_{j'}-1} e^{in\tau\varphi_j} e^{in'\tau\varphi_{j'}} c(n+n'+p) \\ &= \sum_{n=0}^{M_j-1}\sum_{n'=0}^{M_{j'}-1} e^{i(n+n')\tau\varphi_j} e^{in'\tau(\varphi_{j'}-\varphi_j)} c(n+n'+p) \; . \end{aligned} \tag{2.35}$$

Unlike previous case of single-scale Fourier basis, here $M_j$ and $M_{j'}$ do not always equal to each other, making the expression a little more complicated. However, using a similar strategy used to derive the efficient expression of Eq. 2.21, we can still reduce the numerically expensive double sum to several single summations. First, by substituting $l = n + n'$, and assuming $M_j < M_{j'}$, we break it into three terms,

$$\sum_{n=0}^{M_j-1} \sum_{n'=0}^{M_{j'}-1} = \sum_{l=0}^{M_j-1} \sum_{n'=0}^{l} + \sum_{l=M_j}^{M_{j'}-1} \sum_{n'=l-(M_j-1)}^{l} + \sum_{l=M_{j'}}^{M_{j'}+M_j-2} \sum_{n'=l-(M_j-1)}^{M_{j'}-1} .$$

Then by evaluating analytically the summations over $n'$, we can replace each term by single summations. After some manipulations, an efficient expression for off-diagonal matrix elements results, which is given in a compact form here,

$$\left[\mathbf{U}^{(p)}\right]_{jj'} = \hat{S} \sum_{\sigma=0,1} \frac{e^{i\sigma[M_{j'}\tau(\varphi_{j'}-\varphi_j)+\pi]}}{1 - e^{i\tau(\varphi_{j'}-\varphi_j)}} \times \sum_{n=\sigma M_{j'}}^{\sigma(M_{j'}-1)+M_j-1} e^{in\tau\varphi_j} c(n+p) \ , \tag{2.36}$$

where $\hat{S}$ is the symmetrization operator over the subscripts $j$ and $j'$, as defined in Eq. 2.22. For the diagonal matrix elements, $\varphi_j = \varphi_{j'}$, we have,

$$\left[\mathbf{U}^{(p)}\right]_{jj} = \sum_{n=0}^{2M_j-2} (M_j - |M_j - n - 1|) e^{in\tau\varphi_j} c(n+p) \ . \tag{2.37}$$

Note that when $M_j = M_{j'}$, Eqs 2.36 and 2.37 boil down to the formulas previously derived for the case of single-scale Fourier basis, Eqs 2.21 and 2.23.

Once the U-matrices are available in this new basis, the generalized eigenvalue problem Eq. 2.11 can be solved. Due to Eqs. 2.13 and 2.34 the complex amplitudes are computed as,

$$\sqrt{d_k} = \sum_{j=1}^{K} [\mathbf{B}_k]_j \sum_{n=0}^{M_j-1} e^{in\tau\varphi_j} c(n) \, . \tag{2.38}$$

Figure. 2.7 shows an example of such a *multi-scale* basis set for a particular window together with the FDM ersatz spectrum obtained using this basis and plotted, intentionally, in a wide spectral range. As expected, fine features are captured inside the window where dense and narrow band basis functions are used, and only coarsely resolved features appear outside this window. Furthermore, the spectral resolution decreases smoothly in the directions away from the window, so there are no edge effects associated with the local spectral analysis. This makes it easier to combine the results of different windows to construct the overall spectrum. More importantly, with the presence of coarse basis, the line list is free from unphysical poles with large and negative $\gamma_k$ and large amplitudes, although poles with small and negative $\gamma_k$ might still occur. The latter can be easily fixed simply by flipping the sign. With *Multi-Scale* basis, the spectral representation $I(\omega)$, defined in Eq. 2.31, will also be stable and baseline distortion free. An example is shown in Figure 2.3, trace (d). However, $I^{\tau}(\omega)$ is still preferred and should be used.

### Some Aspects of Numerical Implementations

A comment must be made on how to compute the discrete Fourier sums for all $j$ and $j'$ in Eq. 2.36 efficiently. The first two sums need to be computed only once for every $j$. The other two sums depend on both indices $j$ and $j'$, and might seem expensive. Here we describe an algorithm which scales as $M_{\max} \times K_{\text{win}}$ for all the matrix elements of $\mathbf{U}^{(p)}$ rather than $M_{\max} K_{\text{win}}^2$ as one might think. Thus the computational cost of evaluating the U-matrices in a multi-scale Fourier basis is similar to that of the original version of FDM.

For convenience, we introduce the notation:

$$g_j^{(p)}(M) \equiv \sum_{n=M+1}^{M+M_j-1} e^{in\tau\varphi_j} c(n+p) \ .$$

Since $g_j^{(p)}(M_{\max}-1)$ only depends on $\varphi_j$, it can be evaluated once and stored in an array for later use. Now given $g_j^{(p)}(M_{\max})$, one can obtain $g_j^{(p)}(M_{j'})$ recursively for all

$M_{j'} < M_{\max}$ according to

$$g_j^{(p)}(M-1) = g_j^{(p)}(M) \;+\; e^{iM\tau\varphi_j} c(M+p) \;-\; e^{i(M+M_j-1)\tau\varphi_j} c(M+M_j-1+p) \; .$$

Finally, note that there is no need to re-evaluate the Fourier sums in Eq. 2.36 for $p = 1$, once they have been computed for $p = 0$. There exists a simple recursive relation between these two,

$$\sum_{n=0}^{M} e^{in\tau\varphi} c(n+1) = e^{-i\tau\varphi} \left[ \sum_{n=0}^{M} e^{in\tau\varphi} c(n) - c(0) \right] e^{iM\tau\varphi} c(M+1) \; .$$

### A Double-Scale Fourier Basis

We would also like to find out the most efficient coarse basis distribution that requires minimum computational efforts to obtain similarly reliable results. We found, again by numerical experiment, that for the simplest realization of a multi-scale Fourier basis, one can consider just two scales with $M_j = M_{\max}$, for the narrow band window basis, and $M_j = M_c \ll M_{\max}$, for the coarse basis. This corresponds to having two equidistant grids with spacings, $\Delta\varphi_{\min} = 2\pi/\tau M_{\max}$ and $\Delta\varphi_c = 2\pi/\tau M_c$, respectively. At first look, the double scale distribution may not sound as effective as the real "multi-scale" basis with smoothly decaying basis density. However, numerical experiments showed that such a basis distribution is sufficient to obtain "good" results for most circumstances, while there may always exist some cases that require more complicated basis distributions.

Unlike the example shown in Figure. 2.7 which might seem a little complicated, the double-scale basis implementation simplifies the calculation of the U-matrix elements. Furthermore, the size of coarse basis can be further reduced by putting coarse basis functions $|\Psi_j)$ only in the regions with significant peaks in a low resolution picture. Numerical implementation of this "adaptive coarse basis" requires pre-applying DFT or FDM to a short signal to obtain a low resolution picture in order to decide where the coarse basis functions should be placed. The examples shown the next section were all obtained using the double-scale FDM (but without the adaptive basis implementation).

### Multi-Scale Fourier Basis Using Complex Frequency Grid

There is an even simpler implementation of the multi-scale FDM which does not require any of the new formulas derived earlier in this section. Instead of using a real frequency grid $\{\varphi_j\}$, a list of complex values $\varphi_j^{(c)} = \varphi_j + i\gamma_j$ is used,

$$|\Psi_j) = \sum_{n=0}^{M_{\max}-1} e^{in\tau\varphi_j}\, e^{-n\tau\,\gamma_j}\, |\Phi_n) \ , \tag{2.39}$$

Note that if $\gamma_j = 0$ this new definition is the same as Eq. 2.15. When $\gamma_j > 0$, the exponential $e^{-n\tau\,\gamma_j}$ acts as a FT weighting function, and changes the effective FT length. Adjusting the imaginary part of the complex grid points $\varphi_j^{(c)}$ will have the same effect as directly adjusting the FT length, $M_j$ in Eq. 2.34. The relation between $\gamma_j$ and $M_j$ is given as,

$$\sum_{n=0}^{M_{\max}-1} e^{-n\tau\,\mathrm{Im}[\varphi_j^{(c)}]} = M_j \ . \tag{2.40}$$

This equation can be solved analytically by replacing the summation with a integral and solving the resulting simplified equation. The solution can then be expanded in Taylor series and approximated by,

$$\tau\,\mathrm{Im}[\varphi_j^{(c)}] = \begin{cases} \frac{1}{t} + x - x^2 + \frac{3}{2}x^3 - \frac{8}{3}x^4 + \frac{125}{24}x^5 + O(x^6)\ , & \text{if } t < 0.685 \\ 2.82(1-t)\,, & \text{if } t > 0.685 \end{cases}\ , \qquad (2.41)$$

where $t = M_j/M_{\max}$, $x = -\exp(-1/t)/t$. Note that $M_j$ has the same dependence on basis distribution as previously described. A plot of $\gamma_j$ vs. $t$ is shown in Figure. 2.8. With the complex definition of the multi-scale FT basis, the formulas of the single

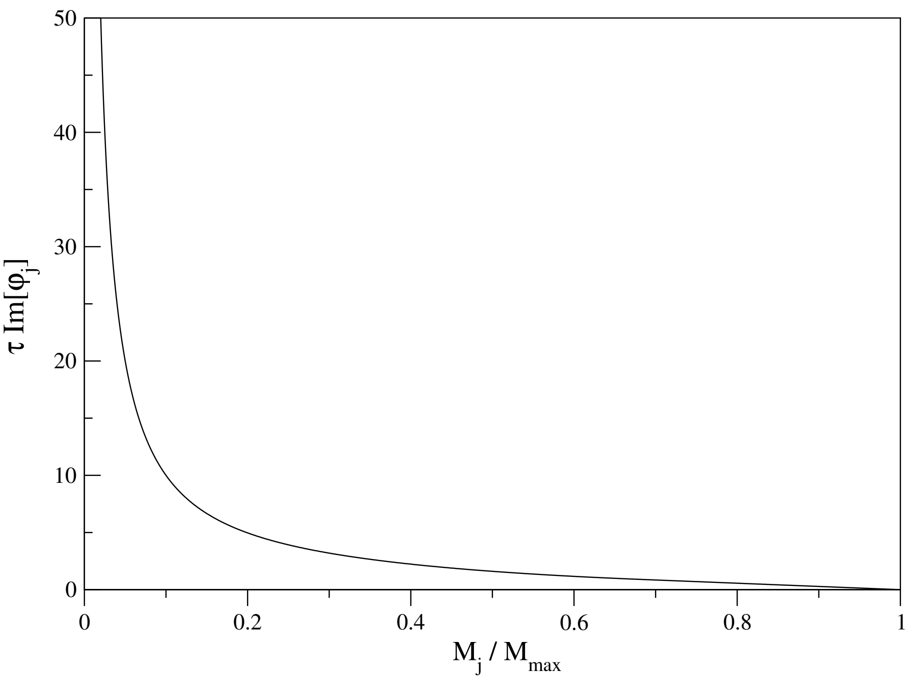



Figure 2.8: A plot of $\tau\,\mathrm{Im}[\varphi_j^{(c)}]$ vs. $M_j/M_{\max}$ for complex frequency grid implementation of *Multi-Scale* Fourier basis.

scale FDM by used by simply replacing the real grid $\varphi_j$ by the complex one. This can greatly simplify the implementation of *multi-scale* Fourier basis in 2D or higher. Numerical experiments showed that the performance of the *multi-scale* FDM with complex grid is similar to that of the original implementation.

**Numerical Examples**

The example of Fig. 2.9 demonstrates the reliability and robustness of the multi-

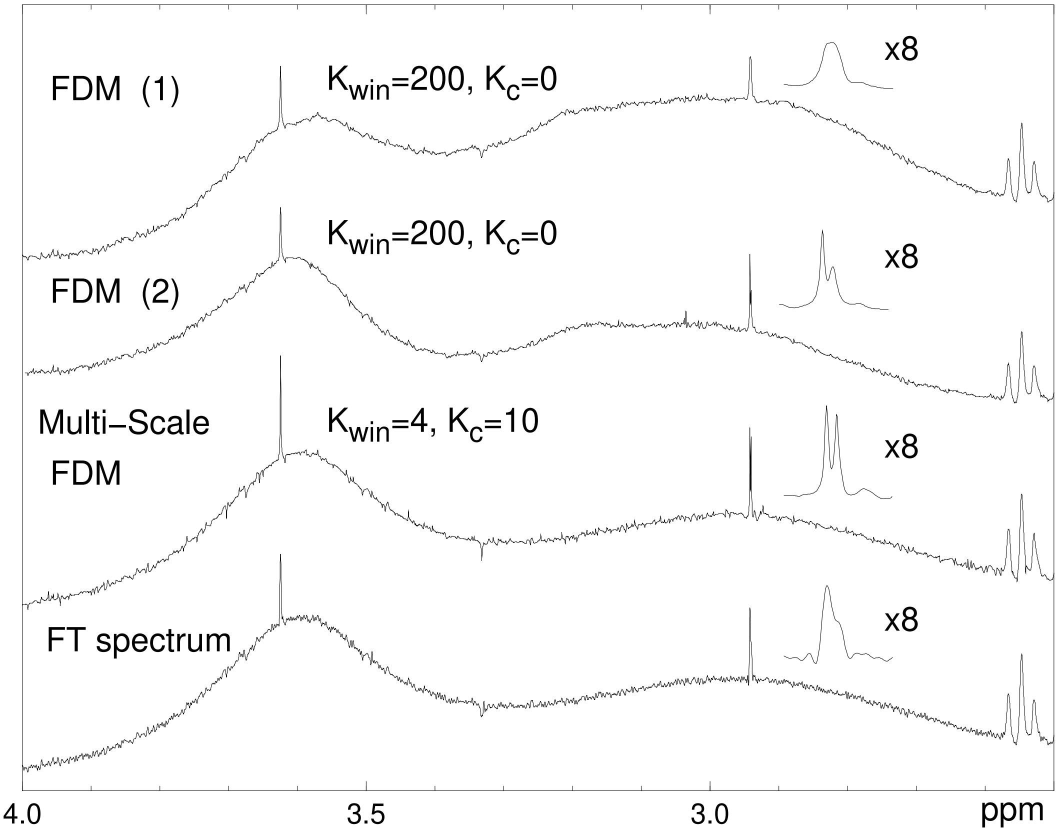



Figure 2.9: This example shows the spectra of an artificially made signal (see text), processed by three different methods: single-scale FDM, *multi-scale* FDM and DFT. The spectrum marked with FDM (2) was obtained with the same parameters as FDM (1) but the position of the windows was slightly shifted. The instability in both representing the background spectrum and resolving the fine features (the doublet) occurs even with $K_{\text{win}}$ as large as 200 basis functions (corresponding to 0.88 ppm spectral window). The doublet is not resolved in the FT spectrum either due to the FT uncertainty principle, although the spectrum envelop is reproduced correctly. The result obtained by the multi-scale FDM with just $K_{\text{win}} + K_c = 4 + 10 = 14$ basis functions is superb in all respects and requires minimal computational effort.

scale FDM. Even under extreme conditions it is still as stable as the Fourier spectrum but can deliver a higher resolution than both the Fourier spectrum and the previous single-scale version of FDM. We considered the same NMR signal as that used in

Figure. 2.1 and 2.7 to which, in the region with no genuine NMR lines, we added artificially some very broad Lorentzian lines to simulate a huge background, and some very narrow lines. $I(\omega)$ instead of ${}^{\tau}(\omega)$ was used to reveal the instabilities due to unphysical poles. For this extreme case, the single-scale FDM does not work well even when a quite big window basis is implemented. For example, for signal length $N \sim 2000$ and $K_{\mathrm{win}} \sim 200$ window basis functions the spectrum envelop is not reproduced correctly. Moreover, in this case the results appeared to be quite sensitive to the input parameters, such as a slight shift in the window position. While reproducing the background spectrum is not a problem for the Fourier spectrum, it cannot resolve the doublet made out of two narrow, equal in height and closely spaced Lorentzian lines. Quite surprisingly, the multi-scale FDM with just $K_{\mathrm{win}} + K_c = 4 + 10 = 14$ basis functions per window reproduces all the relevant spectral features quite accurately. Moreover, the doublet is now much better resolved than in both the single-scale FDM and FT spectra.

Finally, Fig. 2.10 presents an IR spectrum. The interferogram contained 4744 data points that were processed by both FT and multi-scale FDM to generate the absolute value spectra $|I(\omega)|$. The FT spectrum is very hard to interpret and, probably, hard to quantify by conventional means, as the peaks are not quite narrow and both the overlapping effects and the interference with the background are significant. Unlike the FT case, the FDM peaks are generally much sharper. Notably and most importantly, the FDM spectrum is fit by the form of Eq. 2.1, so the parameters of the peaks (such as the positions, widths and amplitudes) are known, while fitting the absolute spectrum $|I(\omega)|$ by Lorentzians would be a very challenging project. Also note that the overall shape of the spectrum is reproduced well by the *multi-scale* FDM, while the single-scale version would be very unstable for this signal because of the strong background.

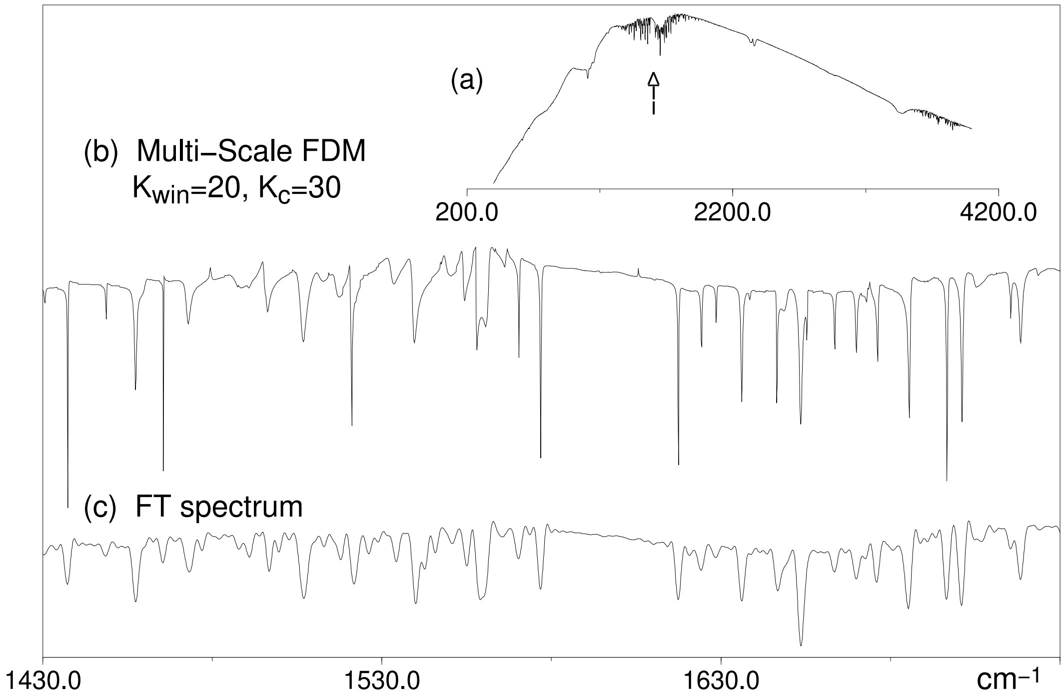



Figure 2.10: FT-IR spectrum (a) and an interesting part of the same spectrum processed by the Multi-Scale FDM with $K_{\text{win}} + K_c = 20 + 30 = 50$ (b) and by FT (c). The interferogram contained 4744 data points.

### 2.2.5 Summary

1D FDM is an efficient and stable algorithm for fitting both large and complicated time domain signals to the summation of Lorentzian lines. It solves the highly nonlinear fitting problem by recasting it into pure linear algebraic problems of diagonalizing some small data matrices. FDM can be used as a parametric estimator as well as a spectral estimator. Provided that signal sufficiently satisfies the Lorentzian model, FDM can deliver resolutions beyond FT time-frequency uncertainty principle. While the spectral estimation, $I^{\tau}(\omega)$, given by FDM is always very stable, unphysical poles with large and negative linewidths and large amplitudes can occur when the signal is very noisy and/or contains nonlocalized features such as strong background and broad peaks. A *multi-scale* Fourier-type basis is introduced to improve the efficiency of local spectral analysis and suppress these unphysical poles. With a *Multi-Scale* Fourier basis, it is possible to use much smaller spectral windows and still obtain reliable results even for very noisy signals.

Currently the 1D FDM algorithm is already a well-developed method which is very stable, sufficiently fast, and can provide reliable high-resolution spectral estimations, given that the signals sufficiently satisfy the Lorentzian model. However, obtaining a compact true line list is still problematic. In the case of heavy overlapped features and/or very noisy data, the lines obtained by FDM do not necessarily all represent the true resonances in the time signal, even though the spectrum computed from such a line list is a reliable estimate of the true infinite DFT spectrum. Any attempt for non-trivial use of the line list, such as large linear phase corrections and line narrowing, might break the mutual interference and cancellation of the lines and result in unstable spectral estimations.

## 2.3 Multi-Dimensional Extensions of FDM

In this section, we describe multi-dimensional extensions of the FDM algorithm (MD-FDM). As we discussed in Chapter 1, multi-dimensional spectral analysis is much more challenging, and the conventional FT spectral analysis has major limitations. Interestingly, the true power of FDM actually lies in its multi-dimensional extensions, in the ability to process the whole multidimensional data set together to obtain the maximum possible resolution in all dimensions. We will start from explaining the concept of true multi-dimensional spectral analysis in Section 2.3.1, then describe a "naive" version of multi-dimensional FDM, which requires simultaneous diagonalization of multiple matrices, in Section 2.3.3. In Section 2.3.4, we will describe a more robust approach that uses the Green's functions to compute the multi-dimensional spectral estimations directly using the results of solving several generalized eigenvalue problems. However, due to the ill-conditioning of the problem, spectra obtained from single FDM calculations are often contaminated with artifacts that are sensitive to any change of the FDM parameters and/or input signal. Several primary averaging procedures which make use of this sensitivity to suppress these artifacts will then be described in Section 2.3.5. Finally, a summary will be given in Section 2.3.6.

For the sake of simplicity, only 2D FDM will be explicitly discussed. Extension of 2D FDM to higher dimensionality is straightforward and does not involve any new fundamental difficulty.

### 2.3.1 What Is a True Multi-Dimensional Method?

Before we describe the theory of multi-dimensional FDM, let's first define the concept of true multi-dimensional spectral analysis. As we discussed in Chapter 1, multi-dimensional FT is essentially a 1D spectral analysis method. The FT time-frequency principle applies to all dimensions independently. In order to obtain high resolution in all dimensions, FT requires long signals in all dimensions. Due to many practical limitations such as limited experimental time and limited long-term instrumental stability, the multi-dimensional NMR signals are typically truncated in the indirect dimensions, leading to poor FT resolutions in the corresponding dimensions. This is one of the major limitations of multi-dimensional FT-NMR.

On the contrary, for a true multi-dimensional method, the obtainable resolutions in all dimensions are determined together by the *total information content* of the signal, which can be roughly measured by the total size of the signal, $N_1 \times N_2 \times \ldots \times N_D$, where $N_l$ is the number of data points along the $l$th dimension. Figure 2.11 shows a schematic illustration of a true two-dimensional spectral analysis method. As long as the total size of the 2D signal is sufficiently large, whether the signal is long or short in a particular dimension has little effect on the obtainable resolution for a true 2D spectral analysis method. In another word, a long time dimension can be used to enhance the resolution along an orthogonal dimension for which it may be time-consuming, or impossible, to obtain a long signal. This property is particularly useful for NMR data processing. In multidimensional NMR experiments, acquisition dimension is cheap and long signals are always available, while the indirect dimensions are expensive and typically truncated. Therefore a multi-dimensional NMR signal is almost always very large and might contain enough information to pin down all the multi-dimensional features, if we have a true-multidimensional spectral analysis method capable of using the information effectively. Multi-dimensional FDM is such a method, as it will become clear later in this section.

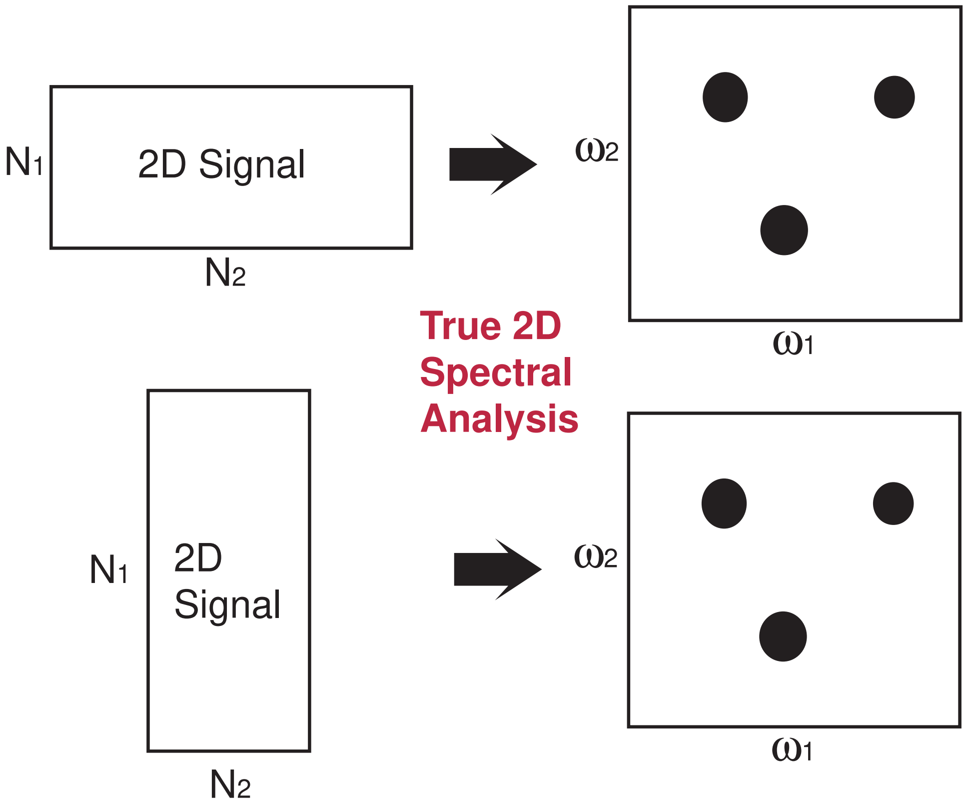



Figure 2.11: A schematic illustration of a true two-dimensional spectral analysis method. As long as the total area of the signal, defined as $N_1 \times N_2$, is sufficiently large, a true 2D spectral analysis method should be able to obtain similar 2D spectral estimations no matter whether one dimension is longer or shorter than the other.

### 2.3.2 2D Harmonic Inversion Problem

Given a 2D complex valued time signal $c(\vec{n}) \equiv c(n_1\tau_1, n_2\tau_2)$, with $n_1 = 0, 1, ..., N_1 - 1$, $n_2 = 0, 1, ..., N_2 - 1$, where $\vec{n}$ is the time vector[5] defined on a 2D equidistant time grid, we want to fit it to the summation of 2D Lorentzian lines,

$$c_{\vec{n}} = \sum_{k=1}^{K} d_k e^{-i\vec{n}\vec{\omega}_k} \equiv \sum_{k=1}^{K} d_k \exp\left(-in_1\tau_1\omega_{1k} - in_2\tau_2\omega_{2k}\right) \ , \tag{2.42}$$

where $\vec{\omega}_k \equiv (\omega_{1k}, \omega_{2k})$ are vectors of unknown complex frequencies, $\omega_{lk} = 2\pi f_{lk} - i\gamma_{lk}$, and $d_k$ are unknown complex amplitudes. This is a 2D *Harmonic Inversion Problem* (2D HIP). The total number of unknown complex parameters in the 2D line list $\{\vec{\omega}_k, d_k\}$ with $K$ entries is $3K$. Note that this formalism of 2D HIP is similar to but different from those used in 2D LP [16] and the 2D matrix pencil method [22]. In the latter cases, models with a direct product set of frequencies, $\{\omega_{1k}, \omega_{2k'}, d_{kk'}\}$, $k = 1, 2, \ldots, K_1$, $k' = 1, 2, \ldots, K_2$, are used, so that the total number of unknowns is $K_1 + K_2 + K_1 \times K_2$. However, a typical 2D NMR spectrum does not have completely direct product pattern. Even for spectra that occur in COSY [85], NOESY [86] and TOCSY [87] types of experiments, the direct product patterns are more or less localized. The non-direct product model of Eq. 2.42 is actually a more general formalism that allows a direct-product output, while leading to a much more compact representation for spectra without direct-product patterns.

It should also be pointed out that a true multi-dimensional method will have the maximal advantage over 1D spectral analysis when the spectra contain minimal direct product patterns. For spectra with a lot of direct product peak patterns such as

[5]Even though only the 2D case is considered, the vector notations make the extension to the case of more than 2D straightforward.

COSY, NOESY, and TOCSY spectra, processing the whole data set together does not lead to a big advantage compared to trace-by-trace 1D processing. In these cases, MD FDM could only help to improve resolution by being an efficient and stable nonlinear parametric method. Therefore, the most interesting applications of MD FDM in NMR are the multi-dimensional experiments that are routinely used to obtain the chemical shift assignment such as HSQC, HNCO(A) and others. Several examples will be given in Chapter 3.

### 2.3.3 A "Naive" Version of 2D FDM

Here we describe a straightforward 2D extension of FDM that was used in Ref. [68, 69, 70]. The object of 2D FDM is to fit a given 2D discrete time signal, defined on a 2D equidistant time grid, to the summation of Lorentzian lines, i.e., solve the 2D HIP of Eq. 2.42. starting point. Similar to the 1D case, this highly nonlinear fitting problem can be cast into a linear algebraic problem, or more precisely, a family of generalized eigenvalue problems, by associating the 2D time signal to a double-time correlation function of a fictitious dynamic system with two commuting non-Hermitian symmetric Hamiltonians $\hat{\Omega}_1$ and $\hat{\Omega}_2$ [68, 70],

$$c(\vec{n}) = \left(\Phi_0\middle|e^{-i\vec{n}\vec{\Omega}}\Phi_0\right) \equiv \left(\Phi_0\middle|e^{-in_1\tau_1\hat{\Omega}_1-in_2\tau_2\hat{\Omega}_2}\Phi_0\right) . \tag{2.43}$$

The Hamiltonians are diagonalizable by the same set of eigenvectors,

$$\hat{\Omega}_l\left|\Upsilon_k\right) = \omega_{lk}\left|\Upsilon_k\right) \ , \quad l = 1, 2 \ , \tag{2.44}$$

where the eigenvectors are orthonormal according to the symmetric complex inner product, $(\Upsilon_k|\Upsilon_{k'}) = \delta_{k,k'}$. Inserting their spectral representations

$$\hat{\Omega}_l = \sum_k \omega_{lk} \left|\Upsilon_k\right) \left(\Upsilon_k\right| \tag{2.45}$$

into Eq. 2.43, we can recover the 2D HIP equation 2.42 with

$$d_k = (\Phi_0|\Upsilon_k)^2 \,. \tag{2.46}$$

Therefore, the nonlinear fitting problem of Eq. 2.42 is recast into a pure linear algebraic problem of diagonalizing the Hamiltonian operators, or equivalently, the evolution operators, $\hat{U}_l \equiv e^{-i\tau_l\hat{\Omega}_l}$,

$$\hat{U}_l \left|\Upsilon_k\right) = u_{lk} \left|\Upsilon_k\right) \ , \quad l = 1, 2 \ . \tag{2.47}$$

The matrix representations of the evolution operators in an appropriate basis can be computed from the 2D time signal. The simplest choice is a set of time like Krylov basis functions created by propagating the "initial" state using the evolution operators,

$$|\Phi_{\vec{n}}) \equiv \hat{U}(\vec{n})|\Phi_0) = \hat{U}_1^{n_1}\hat{U}_2^{n_2}|\Phi_0) \,. \tag{2.48}$$

The elements of the evolution matrices in Krylov basis are simply given by the 2D signal points,

$$[\mathbf{U}^{(\vec{p})}]_{\vec{n},\vec{n}'} = (\Phi_{\vec{n}}|\hat{U}^{(\vec{p})}\Phi_{\vec{n}'}) = c(\vec{n} + \vec{n}' + \vec{p}) \,, \tag{2.49}$$

for any time vector $\vec{p} = (p_1, p_2)$, among which for $\vec{p} = (0,0), (1,0)$, and $(0,1)$ $\mathbf{U}^{(\vec{p})}$ corresponds to the overlap matrix $\mathbf{U}_0$ and evolution matrices in two time dimensions, $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively. The requirement of no missing entries in these matrices limits the size of the Krylov basis to $M_1 \times M_2 = N_1/2 \times N_2/2$. Similar to 1D FDM, the maximum basis size is always used to avoid the problem of estimating the number of poles and to maximize the information content of the $\mathbf{U}$ matrices. As a result, in the Krylov basis, one has to diagonalize potentially both large and ill-conditioned matrices, which is not feasible in practice. For example, a normal 2D NMR signal with $1024 \times 256$ complex data points will lead to $\mathbf{U}$ matrices of dimension 65536 by 65536, which can hardly be handled even by modern supercomputers.

Instead of trying the solve the 2D HIP in the full space, a more efficient approach is to break it into many small problems in the subspace. This can be achieved by using a 2D Fourier-type localized basis, among which the rectangular-window Fourier basis[6] is particularly efficient [68],

$$|\Psi_j) = \sum_{n_1=0}^{M_1-1} \sum_{n_2=0}^{M_2-1} e^{i\vec{n}(\vec{\varphi}_j - \vec{\Omega})} |\Phi_0) \ , \tag{2.50}$$

where $\vec{\varphi}_j \equiv (\varphi_{1j}, \varphi_{2j}),\ j = 1, 2, \ldots, K_{\text{win}}$, is a 2D frequency grid within a small pre-specified 2D spectral window $[\omega_{\text{min1}}, \omega_{\text{max1}}] \times [\omega_{\text{min2}}, \omega_{\text{max2}}]$. The 2D Fourier basis is locally complete, so that small data matrices can be constructed and then diagonalized to obtain accurate spectral estimations within the local spectral window. In principle, any type of 2D frequency grid can be used as long as the basis distribution is sufficiently uniform and the total number of basis functions $K_{\text{win}}$ satisfies,

$$K_{\text{win}} = \rho \, \frac{\Delta\omega_1 \tau_1 \times \Delta\omega_2 \tau_2}{4\pi^2} \, M_1 M_2 \, , \tag{2.51}$$

with $\rho \geq 1.0$ being the density of basis functions, and $\Delta\omega_l = \omega_{\text{maxl}} - \omega_{\text{minl}}, l = 1, 2$. A particularly efficient setup corresponds to having a uniform direct-product 2D grids: $\vec{\varphi}_j = (\varphi_{j_1}, \varphi_{j_2}), j_1 = 1, 2, \ldots, K_{\text{win1}}, j_2 = 1, 2, \ldots, K_{\text{win2}}$, where $K_{\text{win1}} =$

[6] In principle, a 2D *multi-scale* Fourier basis can be used [79]. However, as it will become clear later, there are other problems in MD-FDM, making the necessity of using such a sophisticated local basis less obvious.

$\rho_1 \Delta\omega_1 \tau_1 M_1/2\pi$ and $K_{\text{win2}} = \rho_2 \Delta\omega_2 \tau_2 M_2/2\pi$. Thus total basis size $K_{\text{win}} = K_{\text{win1}} \times K_{\text{win2}}$, and overall basis density $\rho = \rho_1\rho_2$. The **U** matrix elements in such a 2D Fourier basis can be efficiently computed using a procedure similar to the 1D case (see Section 2.2). The final expressions in a compact form can be directly deduced from Eq. 2.21 and Eq. 2.23. They are given here[7],

$$\begin{aligned}\left[\tilde{\mathbf{U}}^{(\vec{p})}\right]_{jj'} &= \hat{S}_1\hat{S}_2 \sum_{\sigma_1=0,1} \frac{e^{i\sigma_1[M_1\tau_1(\varphi_{1j'}-\varphi_{1j})+\pi]}}{1-e^{i\tau_1(\varphi_{1j'}-\varphi_{1j})}} \sum_{n_1=\sigma_1 M_1}^{(\sigma_1+1)(M_1-1)} \\ &\times \sum_{\sigma_2=0,1} \frac{e^{i\sigma_2[M_2\tau_2(\varphi_{2j'}-\varphi_{2j})+\pi]}}{1-e^{i\tau_2(\varphi_{2j'}-\varphi_{2j})}} \sum_{n_2=\sigma_2 M_2}^{(\sigma_2+1)(M_2-1)} e^{i\vec{n}\vec{\varphi}_j} c(\vec{n}+\vec{p}) \ ,\end{aligned} \tag{2.52}$$

where $\hat{S}_l$ defines the symmetrization operator over the subscripts $lj$ and $lj'$ as in Eq. 2.22. When $\varphi_{lj} = \varphi_{lj'}$ Eq. 2.52 is rewritten according to Eq. 2.23. For example, for $\varphi_{1j} = \varphi_{1j'}$ we have

$$\begin{aligned}\left[\tilde{\mathbf{U}}^{(\vec{p})}\right]_{jj'} &= \hat{S}_2 \sum_{\sigma_2=0,1} \frac{e^{i\sigma_2 M_2\tau_2(\varphi_{2j'}-\varphi_{2j})+\pi]}}{1-e^{i\tau_2(\varphi_{2j'}-\varphi_{2j})}} \sum_{n_2=\sigma_2 M_2}^{(\sigma_2+1)(M_2-1)} \\ &\times \sum_{n_1=0}^{2(M_1-1)} \left(M_1 - |M_1-1-n_1|\right) e^{i\vec{n}\vec{\varphi}_j} c(\vec{n}+\vec{p})\end{aligned} \tag{2.53}$$

which can be trivially rewritten for the symmetric case of $\varphi_{2j} = \varphi_{2j'}$. For the case $\vec{\varphi}_j = \vec{\varphi}_{j'}$,

$$\begin{aligned}\left[\tilde{\mathbf{U}}^{(\vec{p})}\right]_{jj} &= \sum_{n_1=0}^{2(M_1-1)} \sum_{n_2=0}^{2(M_2-1)} e^{i\vec{n}\vec{\varphi}_j} c(\vec{n}+\vec{p}) \\ &\times \ (M_1 - |M_1-1-n_1|)(M_2 - |M_2-1-n_2|) \ .\end{aligned} \tag{2.54}$$

Once the **U** matrices are computed, one then needs to solve the 2D generalization of the 1D generalized eigenvalue problem,

$$\tilde{\mathbf{U}}_l \tilde{\mathbf{B}}_k = u_{lk} \tilde{\mathbf{U}}_0 \tilde{\mathbf{B}}_k \ , \quad l = 1, 2 \ . \tag{2.55}$$

[7] Similar to the 1D case, tilde $\sim$ is added the notations **B**, **U** and others to indicate the use of the Fourier basis.

The amplitudes $d_k$ are then obtained from the eigenvectors as

$$d_k^{1/2} = \tilde{\mathbf{B}}_k^{\mathrm{T}} \tilde{\mathbf{C}} \,, \tag{2.56}$$

where the coefficients of the $1 \times K_{\text{win}}$ column vector $\tilde{\mathbf{C}}$ are computed using the following 2D FT of the original signal array $c(\vec{n})$:

$$\left[\tilde{\mathbf{C}}\right]_j = \sum_{n_1=0}^{M_1-1} \sum_{n_2=0}^{M_2-1} e^{i\vec{n}\vec{\varphi}_j} c(\vec{n}) \;, \quad j = 1, 2, ..., K_{\text{win}} \;. \tag{2.57}$$

After the parameter list $\{\omega_{1k}, \omega_{2k}, d_k\}$ is computed, 2D spectral estimations can be constructed. The 2D complex spectrum that corresponds to infinite time 2D discrete Fourier summation,

$$I(\omega_1, \omega_2) = \tau_1 \tau_2 \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} e^{in_1\tau_1\omega_1} e^{in_2\tau_2\omega_2} \left(1 - \frac{\delta_{n_1 0}}{2}\right) \left(1 - \frac{\delta_{n_2 0}}{2}\right) c(n_1\tau_1, n_2\tau_2) \,, \tag{2.58}$$

can be obtained by substituting Eq. 2.42 into Eq. 2.58 and evaluating the summations analytically:

$$I(\omega_1, \omega_2) = \tau_1 \tau_2 \sum_k d_k \left[\frac{1}{1 - e^{i\tau_1(\omega_1 - \omega_{1k})}} - \frac{1}{2}\right] \left[\frac{1}{1 - e^{i\tau_2(\omega_2 - \omega_{2k})}} - \frac{1}{2}\right] . \tag{2.59}$$

In many applications, a double absorption spectrum is desirable. 2D FT of purely phase modulated signals gives rise to "phase-twisted" lineshape [88]. Neither the real nor the imaginary part of $I(\omega_1, \omega_2)$ yields double absorption line shape. A double absorption spectrum can only be obtained by using both cosine- and sine-modulated or N- and P-type data sets (see Section 1.2.3). However, in FDM, the parametric representation $\{\omega_{1k}, \omega_{2k}, d_k\}$ is computed using a single purely phase-modulated data set, thus a double absorption type of spectrum can be constructed, for example, using,

$$A(\omega_1, \omega_2) \approx \tau_1 \tau_2 \sum_k \mathrm{Re}[d_k] \,\mathrm{Re}\left[\frac{1}{1 - e^{i\tau_1(\omega_1 - \omega_{1k})}} - \frac{1}{2}\right] \mathrm{Re}\left[\frac{1}{1 - e^{i\tau_2(\omega_2 - \omega_{2k})}} - \frac{1}{2}\right] , \tag{2.60}$$

where we assume that the signal is correctly phased. However, this expression is not a unique representation, and is meaningful only when the different peaks with absorption shape are not overlapping too much, i.e., the interference effects are not significant. Alternatively, one can use use an expression where the interference effects in the second dimension are conserved,

$$A(\omega_1,\omega_2) \approx \tau_1\tau_2 \sum_k \mathrm{Re}\left[\frac{1}{1-e^{i\tau_1(\omega_1-\omega_{1k})}} - \frac{1}{2}\right] \mathrm{Re}\left[d_k\left(\frac{1}{1-e^{i\tau_2(\omega_2-\omega_{2k})}} - \frac{1}{2}\right)\right] . \quad (2.61)$$

Equations 2.52 through 2.57 outline an efficient way of solving the 2D HIP of Eq. 2.42 using pure linear algebra in the frequency subspace. The two most important properties of 1D FDM are preserved. Therefore, 2D FDM is also intrinsically superior in efficiency and numerical stability. However, there is a new problem. The essential step of 2D FDM is solving the 2D generalized eigenvalue problem of Eq. 2.55, which requires finding a unique set of eigenvectors that simultaneously diagonalizes both evolution matrices. As there is no existing 2D generalized eigenvalue problem solver, in practice 1D generalized eigenvalue solver is used twice to diagonalize $\tilde{\mathbf{U}}_1$ and $\tilde{\mathbf{U}}_2$ separately,

$$\tilde{\mathbf{U}}_l\tilde{\mathbf{B}}_{lk} = u_{lk}\tilde{\mathbf{U}}_0\tilde{\mathbf{B}}_{lk} \ , \quad l = 1,2 \ . \quad (2.62)$$

Two sets of eigenvectors $\tilde{\mathbf{B}}_{lk}, l = 1,2$ are obtained and then need to be matched. The matching step is typically problematic due to several factors: (i) the presence of noise and other imperfections; (ii) use of the Fourier subspace; (iii) degeneracy of eigenvalues. As a result, even though the Hamiltonians $\Omega_1$ and $\Omega_2$ are assumed to commute with each other, the $\mathbf{U}$ matrices constructed in the Fourier basis for noisy signals typically do not commute anymore. Therefore, an unique set of eigenvectors does not necessarily exist. There are various attempts at trying to find some unique set of eigenvectors by reformulating the 2D generalized eigenvalue problems [66, 68, 69] or by "simultaneous diagonalization" [70]. However, they are only applicable to signals with very high SNR and fail for realistic signals.

Nevertheless, applied to model signals, even such a "naive" version of FDM already shows the power of true multi-dimensional data processing. Figure 2.12 is an interesting demonstration of 2D FDM using a nearly noiseless model signal. The signal consists of 32 non-degenerate 2D Lorentzian lines with frequency coordinates created by a random number generator. The average line width in both dimension is 0.02 Hz and all the peaks have the same integral. The spectral width is 2 Hz in both dimensions. The complete data set contains $128 \times 128$ complex data points. In the 2D DFT calculations, cosine weighting function was always used in both dimensions. Double absorption FT spectra were obtained using both N-type and P-type signals. The FDM calculation only used the N-type signal as double-absorption type of spectra can be computed from a single purely phase-modulated signal. The left column shows the 2D DFT spectra obtained by using different number of data points. We can clearly see the effects of FT uncertainty principle. Whenever the signal is short in a time dimensions, the FT resolution is low in the corresponding frequency dimension (panel a and b). It is only when the signal is long in both time dimensions that FT can obtain high resolution in both frequency dimensions (panel c). The 2D FDM spectra shown in the right column tell a totally different story. As long as the total size of the signal is sufficiently large, whether it is $32 \times 8$ (panel d) or $8 \times 32$ (panel e) has little effect on the final results: FDM can always provide a fully resolved 2D

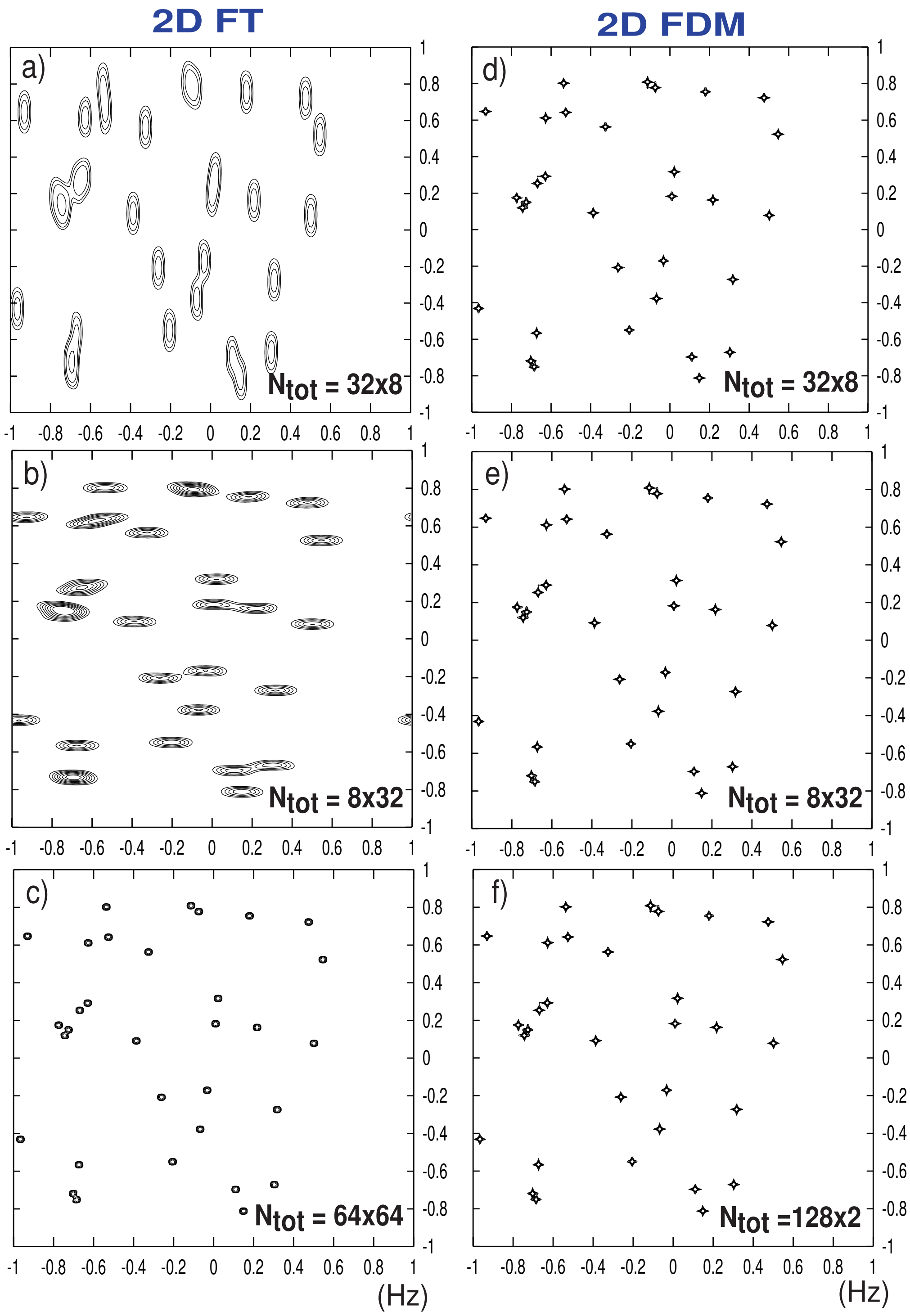



Figure 2.12: A demonstration of FDM using a nearly noiseless model signal that contains 32 2D Lorentzian lines. 2D DFT can provide high resolutions in both dimensions only when the signal is long both ways. However, for FDM, as long as the area of the 2D signal is sufficiently large, fully resolved spectra can always be obtained, even in the extreme case where the signal has only 2 increments in one dimension.

spectrum. This is exactly what we expect from a true 2D spectral analysis method, as illustrated in Figure 2.11. Even in the extreme case where there are 128 data points in the first dimension and only 2 increments in the second dimension (panel f), FDM still fully resolves all 32 peaks, since the total size is sufficiently large.

### Computational Cost of 2D FDM

There are two steps that are time consuming in 2D FDM: construction of the $\mathbf{U}$ matrices and diagonalization of these matrices. By using a direct product frequency grid, one can use the efficient expressions of Eqs. 2.52, 2.53 and 2.54. The computational cost of the double summations in Eq. 2.52 scales as $N_1 \times N_2$. We only need to evaluate these double summations once for each of $K_{\rm win}$ basis functions. Accordingly, the total cost of computing the $\mathbf{U}$ matrices scales as $K_{\rm win} \times N_1 \times N_2$. The cost of diagonalization step is of the order of $K_{\rm win}^3$. Thus the total computational cost of each FDM single window calculation is $K_{\rm win}(a\,K_{\rm win}^2 + b\,N_1 N_2)$, with $a$ and $b$ being scaling factors ($a > b$ in our case).

To analyze the whole spectral range, one needs to carry out the order of $N_1 N_2 / K_{\rm win}$ window calculations, leading to a total cost of $N_1 N_2(a\,K_{\rm win}^2 + b\,N_1 N_2)$. In the limit of small windows, where $K_{\rm win}^2 \ll N_1 N_2$, the total numerical effort of analyzing the whole spectral range using FDM scales as $N_1^2 \times N_2^2$, which is of the same order as that of 2D DFT. In the case of short signals, where $K_{\rm win}^2$ is comparable to or greater than $N_1 \times N_2$, the computational cost will have a strong dependence on the window size. For example, to compute the 2D spectral estimation over the whole spectral range for a 2D signal with $1024 \times 256$ complex data point using basis size $K_{\rm win} = 400$, the total processing time is about 10 hours on a small AMD XP 1800+ CPU Linux workstation. In most cases, only a small portion of the spectral range is of interest. Therefore the actual processing time is often less than one hour for most 2D NMR applications.

### 2.3.4 Green's Function Approach

When the signal is noisy and/or contains degenerate poles, the "naive" version of 2D FDM simply fails due to the difficulty of finding an unique set of eigenvectors that simultaneously diagonalizes both Hamiltonians. In this subsection, we are going to show that there is a way to avoid this problem and construct 2D spectral estimations directly from the results of two 1D generalized eigenvalue problems in terms of Green's functions [71]. A new derivation that is very similar to that of the Regularized Resolvent Transform [77] is given here.

First, let's redefine the object of 2D FDM. Given a finite 2D signal $c(\vec{n}) \equiv c(n_1\tau_1, n_2\tau_2)$ defined on an equidistant 2D time grid, instead of trying to obtain a compact parametric representation $\{\omega_{1k}, \omega_{2k}, d_k\}$, the new object is to estimate the infinite time 2D DFT spectrum, defined in Eq. 2.58, using the finite signal. Inserting the quantum ansatz of Eq. 2.43 into Eq. 2.58 and evaluating the geometric summations analytically gives

$$\begin{aligned} I(\omega_1, \omega_2) &= \left(\Phi_0 \middle| \hat{G}_1(\omega_1)\, \hat{G}_2(\omega_2) \middle| \Phi_0\right) \\ &= \left(\Phi_0 \middle| \left\{ \frac{1}{1 - e^{i\tau_1\omega_1}\hat{U}_1} - \frac{1}{2} \right\} \left\{ \frac{1}{1 - e^{i\tau_2\omega_2}\hat{U}_2} - \frac{1}{2} \right\} \middle| \Phi_0 \right) , \end{aligned} \tag{2.63}$$

where $\hat{G}_l$ are resolvent operators given in terms of Green's functions. Substituting the spectral representation of the evolution operators,

$$\hat{U}_l = \sum_k u_{lk} \left|\Upsilon_{lk}\right) \left(\Upsilon_{lk}\right| , \quad l = 1, 2 , \tag{2.64}$$

into Eq. 2.63, we can obtain an expression for computing the 2D complex spectrum directly from the eigenvectors and eigenvalues,

$$I(\omega_1, \omega_2) = \tau_1 \tau_2 \sum_{k,k'} D_{k,k'} \left[ \frac{1}{1 - e^{i\tau_1(\omega_1 - \omega_{1k})}} - \frac{1}{2} \right] \left[ \frac{1}{1 - e^{i\tau_2(\omega_2 - \omega_{2k'})}} - \frac{1}{2} \right] , \tag{2.65}$$

with cross-amplitudes $D_{k,k'}$ defined as,

$$D_{k,k'} = b_{1k} T_{k,k'} b_{2k'} = (\Phi_0|\Upsilon_{1k})(\Upsilon_{1k}|\Upsilon_{2k'})(\Upsilon_{2k'}|\Phi_0) . \tag{2.66}$$

In addition, one can easily calculate the 1D projections in both dimensions,

$$I_l(\omega_l) = \left(\Phi_0 \,\middle|\, \hat{G}_l(\omega_l)\ \Phi_0\right) = \sum_{lk} b_{lk}^2 \left[ \frac{1}{1 - e^{i\tau_2(\omega_2 - \omega_{2k'})}} - \frac{1}{2} \right] , \tag{2.67}$$

where $b_{lk} \equiv (\Upsilon_{lk}|\Phi_0)$ as defined in Eq. 2.66. Finally, a double-absorption type of spectral estimation can be computed as,

$$A(\omega_1, \omega_2) \approx \tau_1 \tau_2 \sum_{k,k'} \mathrm{Re}[D_{k,k'}] \mathrm{Re} \left[ \frac{1}{1 - e^{i\tau_1(\omega_1 - \omega_{1k})}} - \frac{1}{2} \right] \mathrm{Re} \left[ \frac{1}{1 - e^{i\tau_2(\omega_2 - \omega_{2k'})}} - \frac{1}{2} \right] , \tag{2.68}$$

where we assume that the signal is correctly phased and that different peaks with absorption shape are not overlapping too much so that the interference effects are not significant.

Evaluated in the 2D Fourier basis, defined in Eq. 2.50, the operator expressions of Eqs. 2.65, 2.67 and 2.68 become working formula for computing corresponding spectral estimations with,

$$b_{lk} = \tilde{\mathbf{B}}_{lk}^{\mathrm{T}} \tilde{\mathbf{C}} , \tag{2.69}$$

$$T_{k,k'} = \tilde{\mathbf{B}}_{1k}^{\mathrm{T}} \tilde{\mathbf{U}}_0 \tilde{\mathbf{B}}_{2k'} , \tag{2.70}$$

where the vector $\tilde{\mathbf{C}}$ is defined in Eq. 2.57, and eigenvectors $\tilde{\mathbf{B}}$ are obtained from solving the two 1D GEPs of Eq. 2.62.

Note that cross matrix $T_{k,k'} = (\Upsilon_{1k}|\Upsilon_{2k'})$ is not necessarily diagonal; nor can it necessarily be reduced to diagonal form by permutations. However, if we let $(\Upsilon_{1k}|\Upsilon_{2k'}) = \delta_{k,k'}$, Eq. 2.65 boils down to the ideal representation of Eq. 2.59. Thus adopting representation of Eq. 2.65 covers the ideal case when two evolution operators can be simultaneously diagonalized, and at the same time avoids the necessity of generating a unique set of eigenvectors $\{\Upsilon_k\}$. As a result, it more robust and applicable to processing general noisy experimental signals.

### 2.3.5 Averaging Approaches to Suppress The Artifacts

In principle, most entries in $T_{k,k'}$ should be nearly zero for signals that give rise to spectra without excessive direct-product patterns. Unfortunately, in practice, this is not necessary the case, which manifest itself as the spurious spikes in the 2D spectral estimations, even for the signals with a reasonable signal to noise ratio. For example, Figure 2.13 shows several double absorption type of 2D FT and FDM spectra, computed from a $^1$H-$^{15}$N chemical shift correlation NMR signal of $^{15}$N labeled metalloprotein rubredoxin [89]. Only 8 increments have been used in $^{15}$N dimension with 300 complex data points along the proton dimension. Even though the FDM spectra seem to show some high-resolution characters, they are contaminated by various artifacts such as spurious spikes randomly distributed all over the spectrum and poorly converged genuine poles. Moreover, the artifacts are very sensitive to both

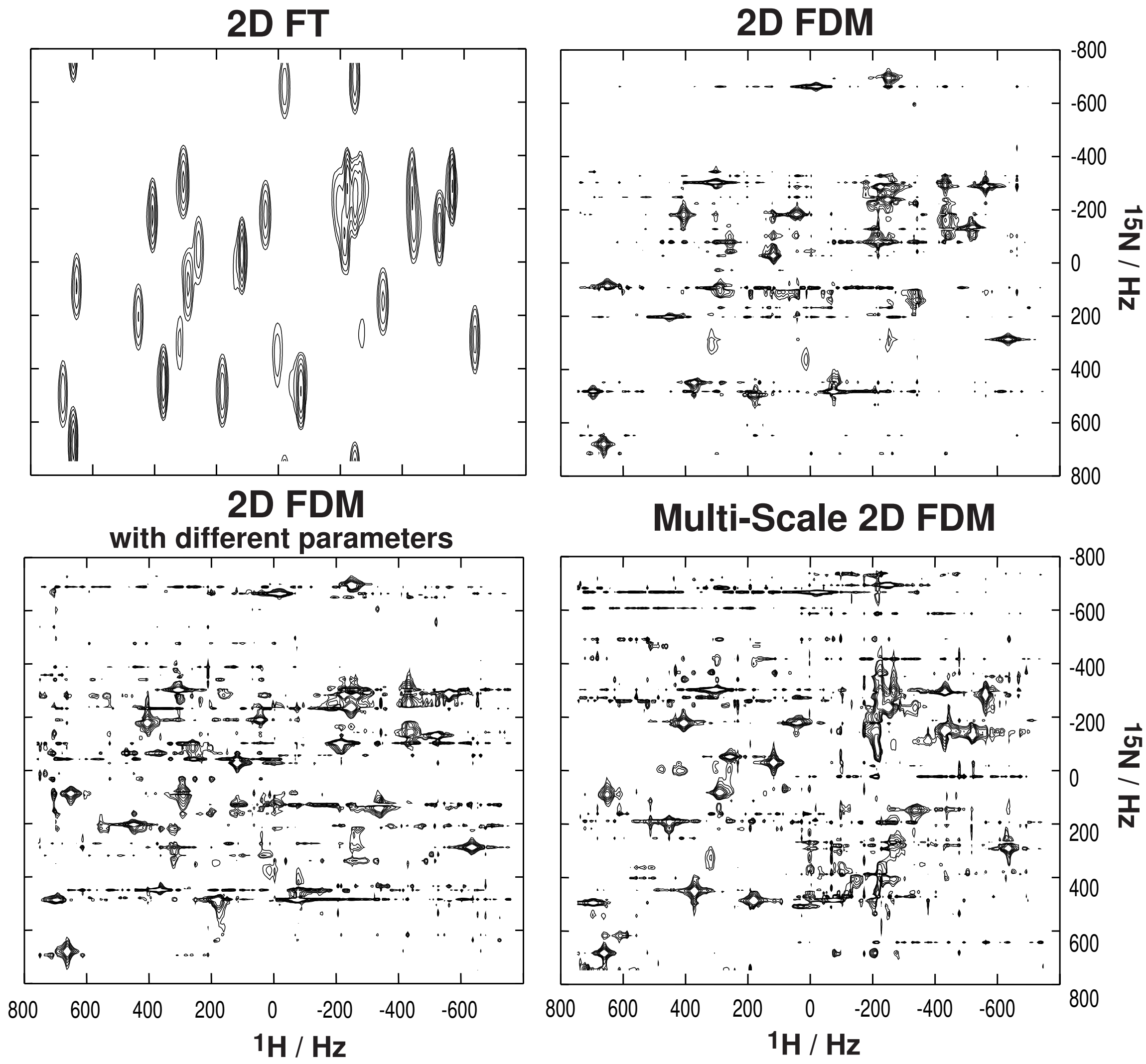



Figure 2.13: 2D FDM and 2D FT double absorption spectra of a $^{1}$H-$^{15}$N chemical shift correlation 2D NMR experiment of a $^{15}$N labeled metalloprotein, rubredoxin. The 2D time signal contains $300 \times 8$ complex points. While showing some high-resolution features, the spectra obtained from a single FDM calculation are contaminated with artifacts such as various spurious spikes randomly distributed throughout the whole spectral window and lineshape distortions of genuine peaks. These artifacts are highly sensitive to any change of the input signal and FDM parameters (e.g, compare the two 2D FDM spectra computed using same data set but with slightly different parameters). Using a *Multi-Scale* Fourier basis does not help to suppress the artifacts (lower right panel).

the small variations in the input data and the parameters of the FDM calculation. *Multi-Scale* Fourier basis does not help to suppress these artifacts (see lower right panel), indicating that this is a different kind of instability. One might suspect that this is due to the fact that the eigenvectors are computed in a non-variational way and therefore less accurate than the eigenvalues. However, there are more essential reasons for it. How to improve the structure of $T_{k,k'}$ is one of most important problems in 2D FDM. These will be discussed in details next section.

Despite the complexity of the problem, there are several straightforward averaging procedures for suppressing the artifacts, which simply make use of their high sensitivity to any change of input data and FDM processing parameters. These procedures were successfully used in early applications of multi-dimensional FDM [69, 72, 73, 90].

The first averaging procedure is *signal-length averaging*. It is based on multiple applications of FDM, applied to the nested subsets of the same signal, using a progressively larger total size in successive calculations. As most artifacts are very sensitive to any change of FDM parameters (while the true features are more stable), they can be averaged out by summing many ersatz spectra computed from different sizes of subsets of the signal. In principle, any FDM parameters such as basis size, basis density and window positions can be used in the same context. However, it was found that varying the signal size is particularly robust. In addition, typical 2D NMR signals have many data points in acquisition dimension, providing a large range for changing the signal size to achieve sufficient averaging. Examples of the *signal-length averaging* procedure can be found in Ref. [69, 72].

The second procedure, *pseudo-noise averaging*, is a little more elegant and does not require a long dimension to provide sufficiently large room for varying the signal length. In stead, it exploits the great sensitivity of FDM ersatz spectrum to small perturbations of the time signal. A small change $\delta(\vec{n})$ in the time signal call lead to a large variation of the ersatz spectrum. Similar to the case of varying the signal length, the artifacts are more sensitive than the true features. Therefore, by averaging over sufficiently many realizations of pseudo-noise perturbation one can suppress the artifacts in the ersatz spectrum. Interestingly, the pseudo-noise perturbation can be implemented right before solving the generalized eigenvalue problems of Eq. 2.62: by adding pseudo-random noise to the $\mathbf{U}$ matrices computed from the original signal, one can save the computation time spent on re-computing the $\mathbf{U}$ matrices for each perturbation. The qualitative justification is that $\mathbf{U}$ matrices are linear functions of the input time signal and therefore a variation of the time signal transfers linearly into the variation of the $\mathbf{U}$ matrices [91]:

$$\tilde{\mathbf{U}}_l \to \tilde{\mathbf{U}}_l + r\tilde{\mathbf{Y}} \ , \tag{2.71}$$

where $\tilde{\mathbf{Y}}$ is a complex symmetric matrix with independent random elements satisfying $\langle [\tilde{\mathbf{Y}}]_{jj'} \rangle = 0$, $\langle |[\tilde{\mathbf{Y}}]_{jj'}|^2 \rangle = 1$. $r$ is a real number and defines the level of perturbation, which, in practice, can be calculated as,

$$r = \frac{q}{K_{\mathrm{win}}^2} \sum_{j=1}^{K_{\mathrm{win}}} \sum_{j'=1}^{K_{\mathrm{win}}} \left| [\tilde{\mathbf{U}}_l]_{j,j'} \right| , \tag{2.72}$$

where $q$ is a real number. Typical value of $q$ ranges from 0.01 to 0.05 depending on how persistent' the artifacts are with respect to the perturbations. 20 to 100 perturbations are usually sufficient to provide stable averaged 2D spectra that are free from most artifacts. Figure 2.14 shows a stable, clean and high-resolution 2D double absorption spectrum obtained by FDM *pseudo-noise averaging*, applied to the same signal used to in Figure 2.13.

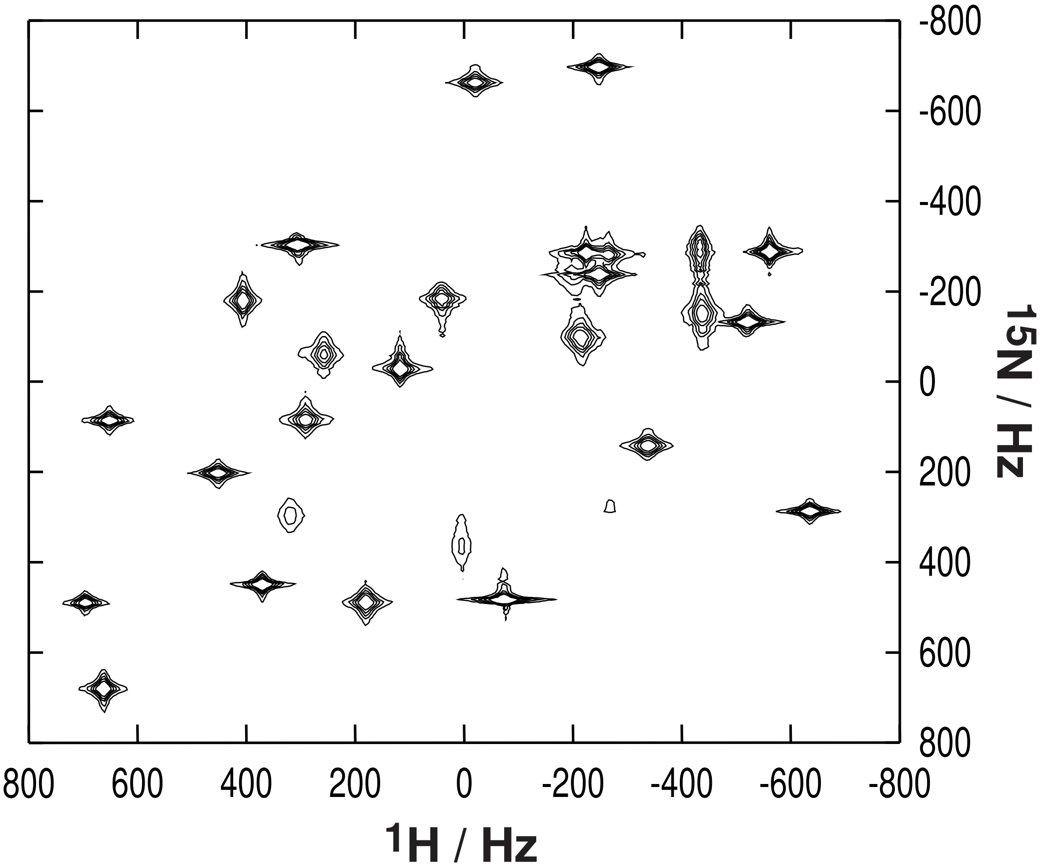



Figure 2.14: 2D FDM double absorption spectrum computed from the same signal used in Fig.2.13. Only 300×8 complex points were used for the purpose of demonstrating the high-resolution nature of FDM. Results of 50 FDM calculations were co-added and averaged with the pseudo-noise perturbation level of $q = 0.02$ (see text). The final spectrum is well resolved (see Fig. 2.13 and Fig.2.17 for 2D FT spectra using $300 \times 8$ and $300 \times 128$ data points respectively), and free of most artifacts.

An efficient implementation of the pseudo-noise averaging was given in Ref. [91]. The expression for computing a 2D double-absorption spectral estimation with *pseudo-noise averaging* is given here. We simply rewrite Eq. 2.68 as

$$A^{\tau}(\omega_1, \omega_2) = \tilde{\mathbf{X}}_{1q}(\omega_1)^{\mathrm{T}} \tilde{\mathbf{U}}_0 \tilde{\mathbf{X}}_{2q}(\omega_2) \ , \tag{2.73}$$

where the absorption-mode vectors $\tilde{\mathbf{X}}_{lq}(\omega_l)$ are computed independently for each $l = 1, 2$ using the eigenfrequencies and eigenvectors of the corresponding perturbed generalized eigenvalue problems,

$$\left(\tilde{\mathbf{U}}_l + r\tilde{\mathbf{Y}}\right)\tilde{\mathbf{B}}_{lk} = u_{lk}\tilde{\mathbf{U}}_0\tilde{\mathbf{B}}_{lk} \ , \tag{2.74}$$

and then averaged over a sufficient number, $N_{\mathrm{FDM}}$, of realizations of the random perturbation $\tilde{\mathbf{Y}}$ as

$$\tilde{\mathbf{X}}_{lq}(\omega_l) = \tau_l \left\langle \sum_k \mathrm{Re}\left[\frac{1}{1 - e^{i\tau_l(\omega_l - \omega_{lk})}} - \frac{1}{2}\right] \tilde{\mathbf{B}}_{lk}\tilde{\mathbf{B}}_{lk}^{\mathrm{T}}\tilde{\mathbf{C}} \right\rangle_{\tilde{\mathbf{Y}}} \ . \tag{2.75}$$

The obvious drawback of both averaging procedures is that the computational cost is multiplied the number of FDM averaging, $N_{\mathrm{FDM}}$. The result improves roughly as $\sqrt{N_{\mathrm{FDM}}}$, assuming the artifacts are completely random for different calculations. It was realized later that the averaging procedure was actually a general but primitive realization of regularization for solving ill-conditioned problems. More efficient regularization techniques such as Singular Value Decomposition (SVD) [9, 92] and Tikhonov Regularization [93] were then introduced to FDM, where only a single FDM calculation is required for obtaining stable and clean 2D spectral estimations. They will be discussed in the next section.

### 2.3.6 Summary

The 1D FDM algorithm was successfully extended to processing multi-dimensional signals. The two most important properties of 1D FDM, namely, solving the non-linear fitting problem by pure linear algebra and local spectral analysis by using a

Fourier-type localized basis, both apply to MD-FDM. Thus MD-FDM is intrinsically efficient and stable method for fitting the time signal to the summation of multi-dimensional Lorentzian lines.

One of the most interesting properties of MD-FDM is that it is a true multi-dimensional spectral analysis method, where the obtainable resolutions in all dimensions are determined together by the total information content of the signal. MD-FDM is able to process the whole data set to characterize the intrinsically multi-dimensional features and provide high-resolution in all dimensions, given that the signal sufficiently satisfies the Lorentzian model and is sufficiently long.

There are also some new problems associated with processing multi-dimensional time signals. Straightforward MD extensions of FDM require finding a unique set of eigenvectors to simultaneously diagonalize multiple data matrices. Such a solution might not exist at all in the cases of noisy signals with potential degeneracies. It can be avoided by using a Green's function approach, in which 2D spectral estimations can be directly computed from the results of solving several 1D generalized eigenvalue problems. With Green's function formalism, FDM can be applied to analyzing general noisy time signals. However a meaningful, compact parametric representation is typically no longer available.

Finally, spectra obtained by a single FDM calculation are still contaminated with artifacts which are very sensitive to any change of input signal and FDM processing parameters. This problem is very different from what we have seen in 1D case, and can not be solved by using a *multi-scale* Fourier basis. Some primitive averaging procedures that simply make use of high sensitivity are introduced to suppress the artifacts. While stable high-resolution spectra can be obtained, the averaging significantly increases the computational cost of MD-FDM and therefore is not the optimal solution. More efficient methods for suppressing the artifacts will be discussed in the next section.

## 2.4 Regularization of Multi-Dimensional FDM

In this section, we will further address the problem of ill-conditioning in multi-dimensional FDM. The averaging procedures we discussed in the previous section are effective but computationally expensive and converge slowly. Further study of the problems helps us to understand the sources of the instability better and makes it possible for us to develop more efficient methods of regularizing FDM. Again, we use 2D FDM as an example to illustrate the principles.

### 2.4.1 Ill-conditioning of 2D FDM

The source of the instability can be traced back to the 2D HIP. In the 1D case, given $N = 2M$ complex data points in the time domain, there exists an exact and unique set of $M$ pairs $\{\omega_k, d_k\}$ that satisfies the 1D HIP equation Eq. 2.1, if there is a solution[8], even though there is no guarantee that such a fit is a "good" fit in the sense of all poles being physical and truly present in the time signal. However, the 2D HIP is ill-posed. Given a 2D signal of the size of $N_1 \times N_2$, the maximum number of unknowns that we can obtain from solving the 2D HIP of Eq. 2.42 is only

[8] In the case of degeneracy, there might be no solution to the 1D HIP.

$3K = 3M_1M_2 = 3N_1N_2/4 < N_1 \times N_2$. Therefore, the 2D HIP is posed as an over-determined problem[9]. For a general 2D signal, there is no exact solution unless the Lorentzian assumption is strictly satisfied and the size the signal is sufficiently large to recover all the poles. In practice, we almost always end up with ill-conditioned generalized eigenvalue problems due to the following reasons:

i) Realistic signals always contain noise which does not satisfy the 2D Lorentzian model. In particular, multi-dimensional NMR signals are especially noisy due to lower sensitivity, $t_1$ noise, and limited long term instrumental stability.

ii) The peaks do not have perfect Lorentzian lineshapes due to many experimental limitations, such as imperfect decoupling, long range couplings, inhomogeneity of magnetic field and many other factors. However, in some special cases such as Constant-Time experiments [75, 76], it is possible to enforce the Lorentzian lineshape in the indirect dimensions and thus improve the performance of FDM significantly.

iii) A Fourier-type localized basis has to be used to avoid the problem of solving huge and ill-conditioned systems. However, the Fourier filter is not perfect. Interferences from nearby features and nonlocalized features (such as very broad poles that define the baseline) act effectively as some additional "noise" to the local spectral analysis. One might try to use a more sophisticated basis such as a 2D *Multi-Scale* Fourier basis to minimize the window effects. However, numerical experiments seem to suggest that the 2D *multi-scale* Fourier basis does not make a significant difference. This could indicate that 2D *Mulit-Scale* Fourier basis is not as efficient as the

[9] The 2D HIP can be also formulated using direct product poles $\{\omega_{1k}, \omega_{2k'}, d_{kk'}\}$ with $2K + K^2 = N_1N_2/2 + N_1^2N_2^2/16$ unknowns. It is then posed as an under-determined problem except for the case of very small signals ($N_1N_2 \leq 8$), which in general, has infinite number of solutions and are thus very unstable numerically.

1D version and that there are still other limiting factors.

iv) The total size of the 2D signal is typically much greater than the total number of peaks present. For example, let's consider a $^{1}$H-$^{13}$N chemical shift correlation experiment of a medium-size protein of 200 residues. A typical signal could contain $1024 \times 256$ data points or even more, from which a maximum number of $512 \times 128 = 65536$ peaks could be determined, much greater than the number of possible peaks, which is of the order of a few hundred. Therefore, in practice, the 2D HIP problem is even more over-determined.

As a result, the data matrices that occur in 2D FDM are almost always nearly singular. The condition numbers[10] of the $\mathbf{U}$ matrices are typically greater than $10^{10}$ in 2D FDM, while in 1D FDM the condition number rarely exceeds $10^{5}$. Exact solution to such an ill-conditioned problem is very unstable and often meaningless. This is exactly what we have seen previously in Figure 2.13. So called *regularization* techniques are required in order to obtain a stable and meaningful solution. Simply put, regularization is the numerical procedure of removing the ill-conditioning of the problem. For example, regularization can be implemented by adding some "smoothness" constraint [94],

$$||\mathrm{R}\,\mathrm{X} - \mathrm{C}|| + q\,||\mathrm{X}|| = \mathrm{min.}, \tag{2.76}$$

with $q$ being the regularization parameter, when solving an ill-posed linear system $\mathrm{R}\,\mathrm{X} = \mathrm{C}$. However, unlike the case of solving a linear system, it is not obvious how the conventional regularization techniques, such as SVD and Tikhonov regularization,

[10] The condition number is a numerical measurement of the conditioning of a matrix. It can be defined as the ratio of largest and smallest singular values of the matrix.

can be applied to the generalized eigenvalue problems encountered in FDM. Further complications include the lack of 2D generalized eigenvalue solvers and the difficulty of finding a unique set of eigenvectors that simultaneously diagonalizes both evolution matrices. We have to use the Green's function approach, where two 1D generalized eigenvalue problems are solved independently. It is not clear how we can regularize both eigenvalue problems in a systematic, consistent fashion. Further more, as the eigenvectors are obtained in a nonvariational way for typical eigenvalue solvers, the numerical errors in each set of the eigenvectors are larger (compared to the eigenvalues), which is another source of many non-zero entries in the cross matrix $T_{k,k'}$, defined in Eq. 2.69.

In the rest of this section, we will discuss three regularization methods that have been applied to 2D FDM. Of these, the FDM2K algorithm [74] is particularly efficient and is used currently as the method of regularization for FDM.

### 2.4.2 Singular Value Decomposition

Singular Value Decomposition, or, SVD, is a very powerful technique for diagnosing the matrices that are either singular or numerically close to singular. In certain cases, SVD can also help to solve the problem by isolating the sources of instability, even though one should be careful in analyzing the "answer" from SVD. Any $N \times N$ square[11] matrix $\mathbf{R}$ can be decomposed into the product of three unique matrices,

$$\mathbf{R} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{V}^{\dagger}, \tag{2.77}$$

[11]Note that SVD can also be applied to a rectangular matrix.

where $\mathbf{W}$ and $\mathbf{V}$ are $N \times N$ unitary matrices, and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_i\}$, a real diagonal matrix with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$ being singular values. The superscript † denotes "conjugate transpose". The columns of matrix $\mathbf{W}$, or equivalently, those of matrix $\mathbf{V}$, form an orthonormal set of basis vectors which covers both the nullspace[12] (if any) and range of matrix $\mathbf{R}$. All the singularity or ill-conditioning of $\mathbf{R}$ is contained in the diagonal matrix $\mathbf{\Lambda}$. By examining the singular values, it is possible to determine the rank and separate the nullspace of $\mathbf{R}$. Specifically, the columns of $\mathbf{W}$, whose same-numbered singular values $\lambda_i$ are *nonzero*, form an orthonormal set of basis vectors that span the range; the columns of $\mathbf{V}$, whose same-numbered singular values $\lambda_i$ are *zero*, form an orthonormal basis for the nullspace [9]. Unfortunately, reals situations are more complicated than this ideal assumption. SVD of the $\mathbf{U}$ matrices computed from a real signal with noise will contain only nonzero singular values. There are only 'large' and 'small' singular values. One can then hope that plotting $\lambda_j$ vs. $j$ shows some abrupt magnitude change so that a zero threshold can be unambiguously picked. This might be true in some cases but often not in NMR applications, making choosing a zero threshold problematic.

There are several ways SVD can be used to regularize FDM. They can be divided into two categories: hard and soft regularization.

**1. Hard Regularization: truncated SVD.**

In this implementation, we apply SVD to $\tilde{\mathbf{U}}_0$ and assume that it is possible to determine the rank by choosing some zero threshold. Once the rank is identified, we

[12]For a singular matrix $\mathbf{R}$, there exists some vectors $\mathbf{x}$ in some subspace, called nullspace, that satisfy $\mathbf{R} \cdot \mathbf{x} = 0$. The dimension of the nullspace, defined by the number of linearly independent vectors $\mathbf{x}$, is called the *nullity* of $\mathbf{R}$.

can re-evaluate the $\mathbf{U}$ matrices in the reduced basis formed by columns of $\mathbf{W}$ that correspond to the singular values greater than the zero threshold,

$$[\tilde{\mathbf{U}}_l^{red}]_{ij} = \mathbf{W}_i^{\mathrm{T}} \tilde{\mathbf{U}}_l \mathbf{W}_j \, . \tag{2.78}$$

The column vector $\tilde{\mathbf{C}}$ defined in Eq. 2.57 should be also re-evaluated in the new basis,

$$[\tilde{\mathbf{C}}^{red}]_i = \mathbf{W}_i^{\mathrm{T}} \tilde{\mathbf{C}} \, . \tag{2.79}$$

Note that SVD of $\tilde{\mathbf{U}}^{(1)}$ and $\tilde{\mathbf{U}}^{(2)}$ should give similar results. However, $\tilde{\mathbf{U}}_0$ is the least noisy and thus preferred. Also note that the transpose instead of the conjugate transpose of $\mathbf{W}_i$ is used, as the complex inner product instead of the Hermitian inner product is always used in FDM for describing dissipative systems.

As pointed out by Moler and Stewart [80], even when the matrices on both sides of the GEP have a common null space (as is the case in FDM), it is not recommended to use SVD to get rid of the null subspace. The reason is that the eigenvalues and eigenvectors become very sensitive to the assumed rank of range subspace. On the other hand, it was argued that the QZ algorithm [80] provided accurate eigenvalues and eigenvectors in terms of two numbers, $u_k = \alpha_k/\beta_k$, the accuracy of which was not affected by ill-conditioning of the matrices. It was also argued that "unreliable" eigenvalues could be identified by smallness of both $\alpha_k$ and $\beta_k$. Our experience with 2D FDM is somewhat contradictory to these recommendations. Spectra constructed from the accurate eigenvalues and eigenvectors computed by QZ, applied to the original $\mathbf{U}$ matrices, are contaminated by artifacts, while the truncated SVD of $\mathbf{U}_0$ can help in removing the artifacts, as demonstrated in Figure 2.15. Despite this, the ambiguity in deciding how many basis vectors to retain remains a major problem.

Figure 2.15 shows four 2D $^{15}N - {}^{1}H$ chemical shift correlation spectra obtained from a sample of $^{15}N$ labeled rubredoxin, a small metalloprotein [89]. Only 16 increments are used in the $^{15}N$ dimension, with 200 complex points along the $^{1}H$ dimension. The initial basis size is then $50 \times 8 = 400$. Plotting the singular values (top panel) clearly shows an abrupt magnitude change at around $i \sim 30$. Setting a cutoff at this point will give a very clean, well converged spectrum, panel (c). We do not actually need to throw away so many basis vectors. Panel (b) shows that even throwing away only 23 basis vectors corresponding to the smallest 23 singular values already reduces most of the artifacts. Further reduction of the basis size to around 150 will basically yield the same spectrum as shown in panel (c) (data not shown). While it is safer to keep a few more basis vectors,it is extremely dangerous to throw away more vectors than necessary, which could lead to missing genuine peaks, e.g., panel (d). This is a big disadvantage of such an aggressive procedure. Even worse, typical NMR data may not have such a clean break in singular values. Then choosing any cutoff can be very risky and give the operator a chance to bias the experimental results to support some particular viewpoints. Therefore, such a procedure is not generally applicable. It is better to be more conservative, applying "soft cutoff" instead of the "hard cutoff".

**2. Soft Regularization: Pseudo-Inverse of $\mathbf{U}_0$:** $1/\lambda_j \to \lambda_j/(\lambda_j^2 + q^2)$

In this implementation, we solve the GEP of equation Eq. 2.62 in two steps: first we apply SVD to $\tilde{\mathbf{U}}_0$ and compute a pseudo inverse,

$$\tilde{\mathbf{U}}_0^{-1}(q) = \mathbf{V}\,[\mathrm{diag}(1/\lambda_j)]_q\,\mathbf{W}^\dagger, \tag{2.80}$$

which is dependent on a real regularization parameter $q$. Then, we solve the normal

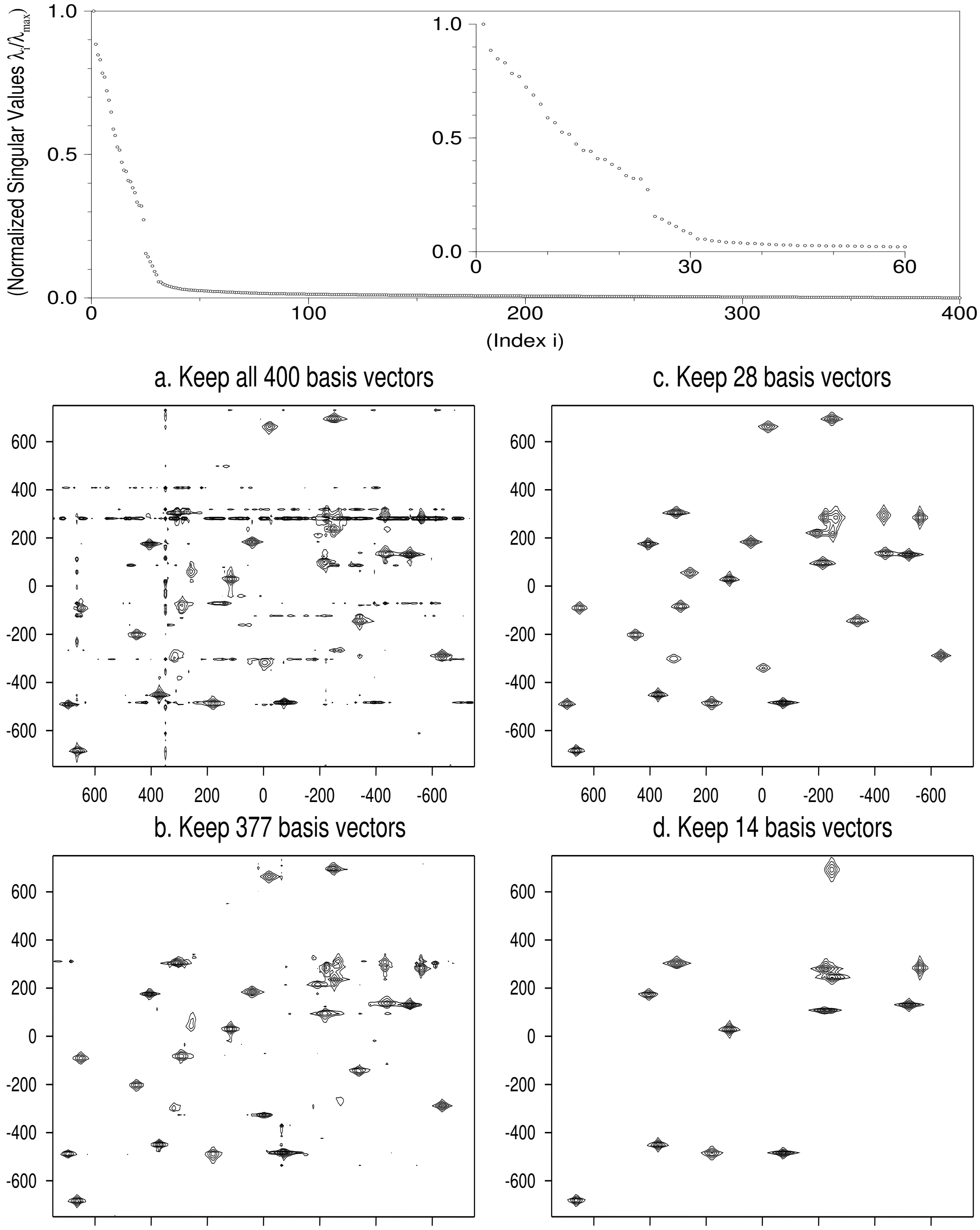



Figure 2.15: Chemical shift correlation spectra obtained by FDM with hard regularization via truncated SVD (see text). The total size of the signal used is $N_1 \times N_2 = 200 \times 16$, leading to a total basis size is $\mathrm{Nb}_{\mathrm{total}} = 50 \times 8 = 400$ (a single window was used). The top panel shows the distribution of singular values of the overlap matrix $\mathbf{U}_0$. The four spectra (a,b,c,d) at the bottom are obtained by retaining various numbers of basis vectors.

eigenvalue problems,

$$\tilde{\mathbf{U}}_0^{-1}(q)\,\tilde{\mathbf{U}}_l\,\tilde{\mathbf{B}}_{lk} = u_{lk}\tilde{\mathbf{B}}_{lk} \quad l = 1, 2\,. \tag{2.81}$$

There are several possible ways of evaluating the pseudo-inverse matrix $\tilde{\mathbf{U}}_0^{-1}(q)$. One can simply use the conventional truncated SVD fashion pseudo inverse: replace all the $1/\lambda_j$ by zero if $\lambda_j$ is smaller than some zero threshold $q$. Then the results will be also sensitive to the choice of zero threshold just like the hard regularization. A less aggressive way is to replace all $1/\lambda_j$ with $\lambda_j/(\lambda_j^2 + q^2)$ with $q$ being a real regularization parameter. The advantage of this approach is that the regularization is "softer" and small deviation of $q$ from the optimal value will not lead to dramatic deterioration of the results. There is usually a stable region where changing $q^2$ up to one order of magnitude leads to little change in the 2D spectral estimation, provided that the signal contains enough information for FDM to obtain a stable estimation. In practice, the optimal level of regularization can by found by carrying out multiple calculations with different values of $q^2$.

Shown in Figure 2.16 are several 2D double absorption spectra obtained by single FDM calculations with different levels of regularization, applied to the same signal as that used in Figure 2.15. When the regularization is too small, the spectra is still contaminated with artifacts, e.g., panel (a); further increasing the regularization to a sufficiently large level results in a stable and high-resolution 2D spectra, shown in panel (b). There is a region where the spectra obtained are relatively stable according to the change of $q$. Finally, when the regularization is too large, the FDM spectrum is furthered smoothed and small features are suppressed down to the baseline, e.g.,

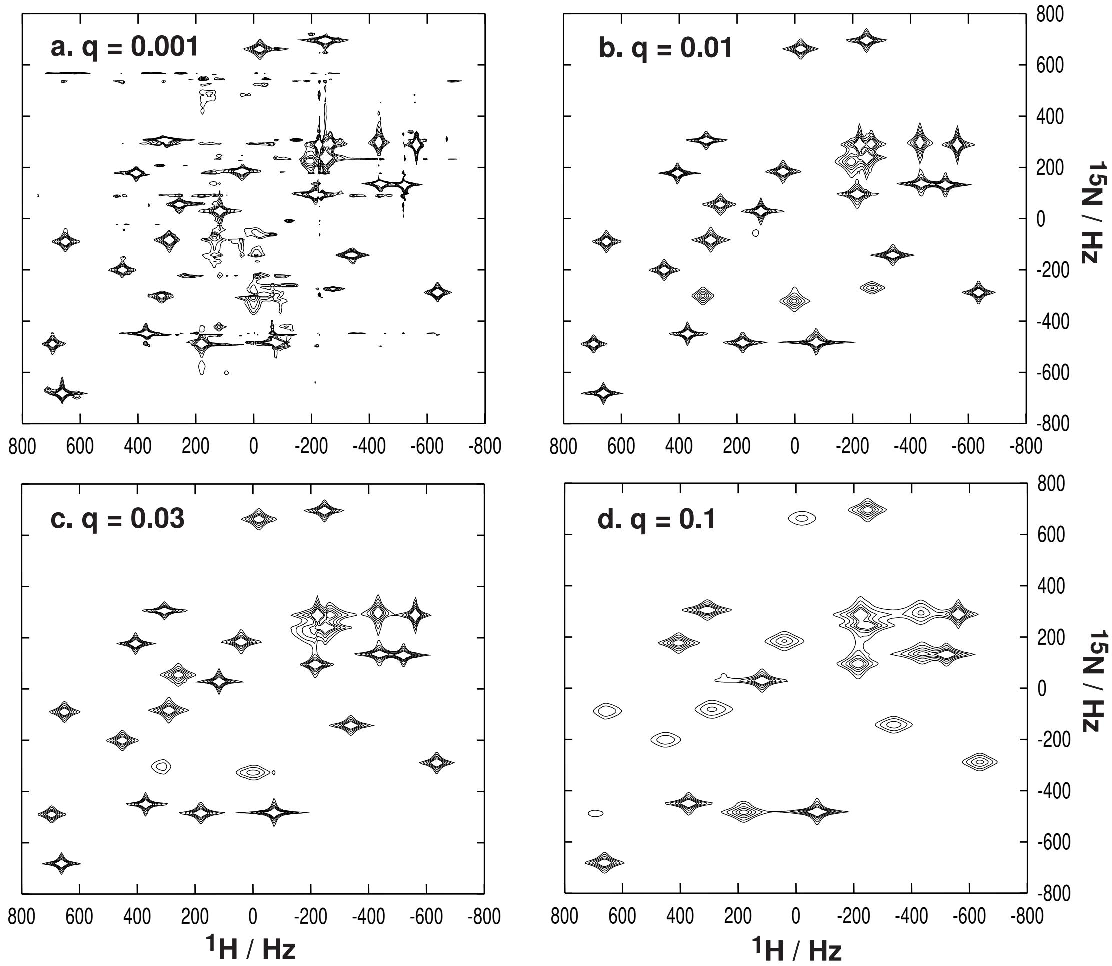



Figure 2.16: Chemical shift correlation spectra obtained by FDM using soft regularization via SVD (see text). The signal used is the same as that used in Figure 2.15. The signal length processed is $N_1 \times N_2 = 200 \times 16$ and total basis size is $\text{Nb}_{\text{total}} = 50 \times 8 = 400$. It shows that with appropriate level of regularization, stable and high-resolution spectra can be obtained using single FDM calculation (panel b: $q = 0.01$). In addition, the appearance of the spectra has a smooth dependence on the regularization parameter (panel b, c and d). When the regularization is too large, small features are suppressed to baseline (panel d). Note that the overall spectrum is smoothed in a very non-uniform way. The values of $q$ shown are ratios of real $q$ to the largest singular value ($\lambda_1$).

panel (d). Moreover, the spectrum is smoothed in a non-uniform way. We can pretend to understand the effect of the regularization simply by saying that increasing $q$ de-emphasizes the contribution of small singular values, which are responsible for describing the noise, small features and fine structures. However, exactly how a particular peak might be affected by the regularization is not totally clear.

### 2.4.3 FDM2K

The soft regularization via pseudo inverse of $\mathbf{U}_0$ is a very efficient method in suppressing the artifacts. However, SVD is expensive and therefore suboptimal computationally. There is another very effective but cheaper way of regularizing FDM, which was invented in the year of 2000 and therefore named FDM2K. In FDM2K, we simply rewrite the original GEP of Eq. 2.62 as,

$$\mathbf{U}_0^\dagger \mathbf{U}_l \mathbf{B}_{lk} = u_{lk} \left( \mathbf{U}_0^\dagger \mathbf{U}_0 + q^2 \right) \mathbf{B}_{lk} \ , \tag{2.82}$$

by introducing a real, positive regularization parameter $q$. The new generalized eigenvalue problem now has a positive definite Hermitian matrix on the right hand side. It can be proven that $q$ imposes minimum singular value for the new right matrix and therefore effectively controls its condition number. Similar to the case of soft regularization, the resulting spectra will have a relatively smooth dependence on the regularization level. Figure 2.17 demonstrate the FDM2K algorithm using the same $^{15}$N-$^{1}$H HSQC signal of rubredoxin used in previous examples. When signals are sufficiently long and have reasonably high SNR[13], there will be a stable region where the

[13]This qualitative statement will become more quantitative in Chapter3, where we will study the dependence of performance of FDM2K on SNR and signal sizes.

spectra obtained change very little while the regularization changes by up to an order of magnitude. When the regularization is too high, small features are smoothed out first while the strong peaks are almost unaffected, and the overall smoothing effect is non-uniform.

It is very tempting to treat $q$ as the "noise power", just like in the Maximum Entropy applications [8]. Indeed, it can be proved semiquantitatively [74], that the spectral features with amplitudes $d_k$ of order of $q$ and below are smoothed out, while the stronger peaks remain essentially unaffected. Based on these observations, the actual procedure implemented for computing the regularization level in current version of FDM2K is as following,

$$q^2_{\mathrm{Act.}} = q^2\, M_1^2 M_2^2 \sum_{\vec{n}} |c(\vec{n})|^2\,, \tag{2.83}$$

where $q^2_{\mathrm{Act.}}$ is the actual value of regularization used in Eq. 2.82 and $q^2$ is the relative scale value specified by the user. Implemented this way, the optimal $q^2$ for typical experimental signals falls into a small range of from $10^{-5}$ to $10^{-3}$ with $10^{-4}$ being optimal for most NMR experimental signals acquired under normal conditions: $\geq$ 1mM sample concentration, normal probe, 500 MHz Spectrometer, $2-16$ scans per increment, room temperature, reasonably long $T_1$ and $T_2$, etc. Optimal $q^2$ might be lower for signals acquired under better conditions with higher SNR, and vice versa.

Several important aspects of FDM2K need to be discussed here. First, exactly how particularly peaks are regularized by $q$ is still not fully understood, due to the highly nonlinear property of Eq. 2.82. In the original reference of FDM2K [74], we demonstrated that in the simple case of a single peak, the peak position, intensity

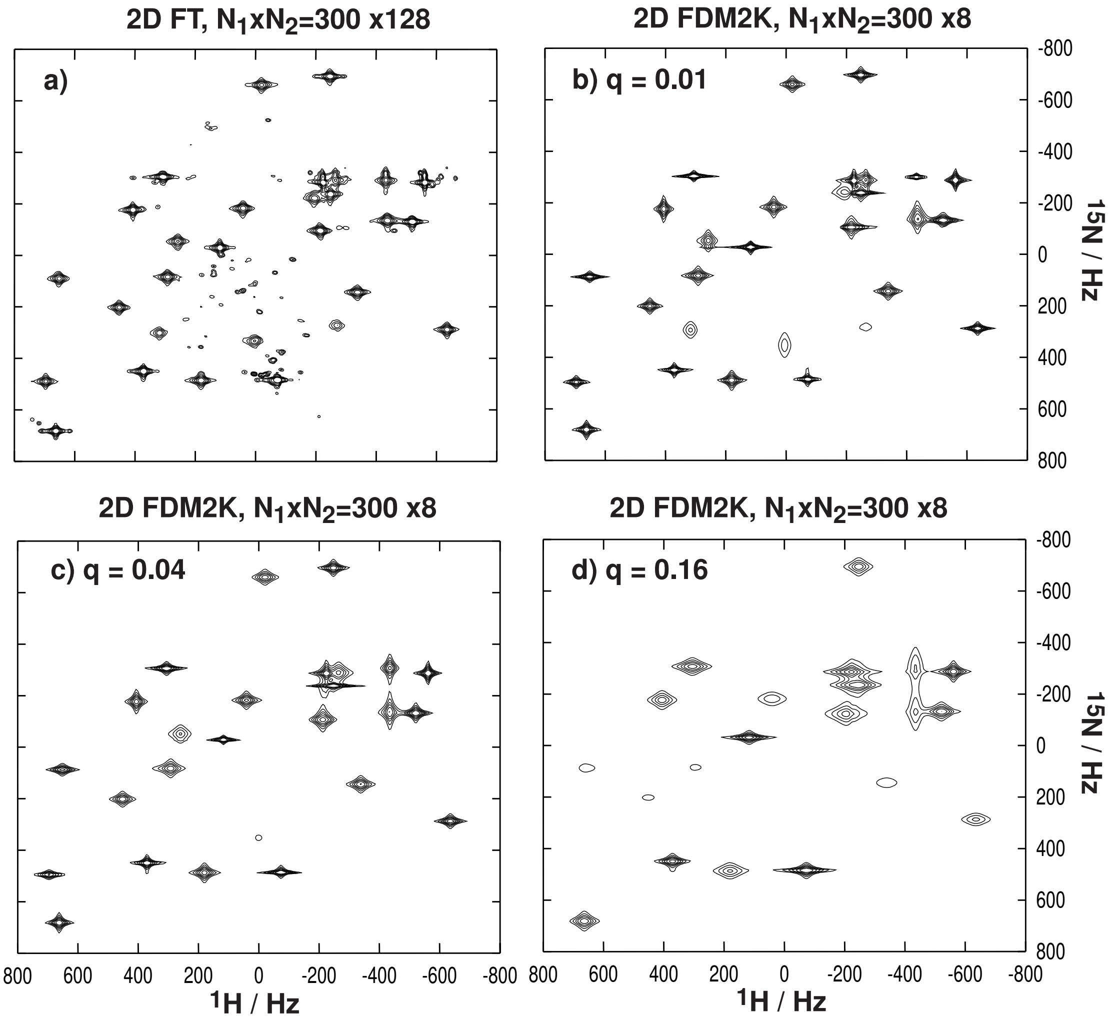



Figure 2.17: Chemical shift correlation spectra obtained by 2D FT and FDM2K (see text). The signal used is the same as that used in Figure 2.15. 2D FT spectrum was obtained by using both N- and P-type data sets, each with $N_1 \times N_2 = 300 \times 128$ complex data points. Cosine apodization function was used in both dimensions. The signal length used in FDM2K calculations is $N_1 \times N_2 = 300 \times 8$. It shows that with appropriate level of regularization, stable and high-resolution spectra can be obtained using single FDM calculation (panel b). In addition, the behaviour of the spectra does have a smooth dependence on the regularization parameter (panel b, c and d). When the regularization is too large, the whole spectrum is further smoothed and small features are suppressed to baseline (panel d). However note that it also happens in a very non-uniform way. Also note that the values of $q$ shown are relative values (see text).

and phase were not affected by the regularization while the regularized linewidth was bigger than the true value. The change of linewidth is determined by the relative amplitude of peak intensity $d_k$ and regularization level $q$: in the limit of $q^2 \ll |d_k^2|$, the change is insignificant; in the other limit, the peak will be significantly broadened. In the general cases of many resonances, how a particular peak is regularized seems to depend not only on the peak intensity, but also on its environment such as close-by peaks, degeneracy/near degeneracy with other peaks, and others. Second, there is definitely some similarity between FDM2K and the conventional Tikhonov Regularization. However, the regularization in FDM2K is efficient only when matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ are "bigger" than $\mathbf{U}_0$[14] . This is usually true as signals decay in the time domain. However, in the case of noisy signals, some components of $\mathbf{U}_1$ or $\mathbf{U}_2$ may be larger than corresponding counterparts of $\mathbf{U}_0$ so that large $q$ must be used in order to suppress all artifacts, resulting in lower resolution. Figure 2.18 compares the effects of Tikhonov regularization (implemented as RRT; see Section 2.5 for details) and FDM2K using a 1D model signal. While the RRT spectra are regularized in a smooth and predictable fashion, the FDM2K spectra show some unfavorable non-uniform distortions. Third, matrices involved in Eq. 2.82 is asymmetric when $q \neq 0$. To solve it exactly, both left and right eigenvectors should be computed. It is usually accurate enough to calculate the right eigenvectors only, when $q$ is small compared

[14] To understand this statement, one needs to write out the corresponding resolvent for Eq. 2.82 for 2D spectral estimation, which is

$$\mathbf{R}(\omega_l) \sim (\mathbf{U}_0^\dagger \mathbf{U}_0 + q^2) - e^{i\tau_l \omega_l} \mathbf{U}_0^\dagger \mathbf{U}_l.$$

Therefore, $q^2$ can guarantee that the resolvent is non-singular only when all eigen-components of $\mathbf{U}_0$ are greater than corresponding counterparts of $\mathbf{U}_l$.

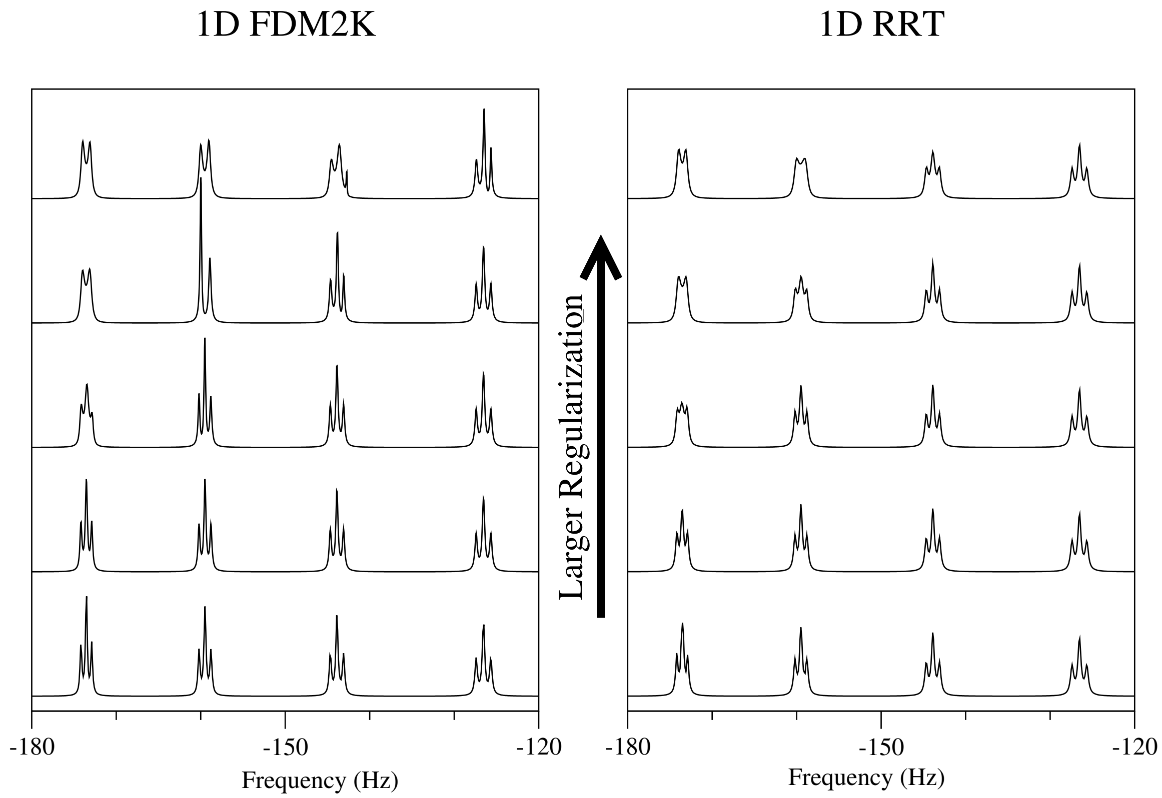



Figure 2.18: Comparison of Tikhonov Regularization (implemented as RRT) and FDM2K using a “Jacob’s Ladder” model signal.

to the norm of the matrices and the eigenvectors are used only for calculating the amplitudes. However, in other cases such as using the eigenvectors to construct some nontrivial projection operators (e.g., J-projections; see Section 3.3), one might need to compute both eigenvectors to avoid introducing extra artifacts.

### 2.4.4 Snap-Shot FDM: Self-Regularization.

There is another very interesting way of regularizing FDM where no additional regularization parameters such as $q$ is necessary. Instead, it requires experimentally acquire multiple signals that consist of the same genuine peaks but uncorrelated noise. In NMR experiment, multiple scans, $c^{(i)}(n),\ i = 1, 2, \ldots, N_s$, are routinely acquired and then co-added to obtain an averaged signal, $\bar{c}(n) = \frac{1}{Ns}\sum_{i=1}^{N_s} c^{(i)}(n)$, which has a better signal to noise ratio. These scans can be conveniently stored separately before they are co-added, even though this requires more disk space for data storage. The second approach leads to little benefit if the Fourier transform is to be used. The reason is that FT is a linear method. Whether FT is applied to each scan separately and then the results are averaged or the signals are averaged before being transformed does not change the resulting spectrum. However, when a nonlinear method like FDM is used, storing each scan separately might be advantageous. It is possible to co-process all scans together to distinguish "signal" and "noise" components, based on the assumption that "signal" should be consistent from scan to scan while "noise" is random and uncorrelated. The FDM algorithm can be modified to use the addition information encoded in multiple scans, or "Snap-Shots". This idea discussed below was first suggested by Neumaier [95] and then exploited by Curtis [96]. We will use 1D as an example for the sake of clarity. 2D or higher dimensional cases are more complicated but the basic idea is the same.

In the normal FDM, the problem of identifying the spectral parameters is reformulated as a generalized eigenvalue problem of Eq. 2.11 where the right and left matrices are computed using the averaged signal. To process $N_s$ signals that consist of same set of genuine damped sinusoids, a super generalized eigenvalue problem can be solved,

$$\begin{bmatrix} U_1^{(1)} \\ \vdots \\ U_1^{(N_s)} \end{bmatrix} B_k = u_k \begin{bmatrix} U_0^{(1)} \\ \vdots \\ U_0^{(N_s)} \end{bmatrix} B_k \tag{2.84}$$

where $U_l^{(i)}, l = 0, 1$ are the evolution and overlap matrices constructed using $i$th scan, $c^{(i)}(n)$. By solving all $N_s$ GEPs in parallel for a single common set of eigenvalues and eigenvectors, components present in all data sets are emphasized while the random (noise) components are suppressed. This technique has also been used in radar imaging and acoustic signal processing [22, 3]. Similar techniques have also been used to extract quantum tunneling splittings from semi-classical real-time cross-correlation functions [97], or multiple excited vibrational states from imaginary-time cross-correlation functions [78, 98].

The dimension of the rectangular matrices on both sides is $N_s K_{\text{win}} \times K_{\text{win}}$, but there will be only $K_{\text{win}}$ eigenvalues and corresponding eigenvectors. It is reasonable and efficient to first reduce both matrices to $K_{\text{win}} \times K_{\text{win}}$ square matrices. There are quite a few choices here. For example, one can simply apply SVD to right (or left) matrix, then re-evaluate both matrices in the reduced space formed by $K_{\text{win}}$ basis vectors that correspond to the $K_{\text{win}}$ largest singular values (see Section 2.4.2 for details). Alternatively, one can use the FDM2K approach and recast Eq. 2.84 into a normal GEP that only involves square matrices,

$$U_r^\dagger \, U_l \, B_k = u_k \, U_r^\dagger \, U_r \, B_k \tag{2.85}$$

where $U_r$ and $U_l$ denote right and left rectangular matrices respectively. The second procedure is numerically more efficient and generally yields similar results to the truncated SVD approach.

By recasting the super GEP into a normal GEP, we can compute both eigenvalues and eigenvectors using standard routines such as QZ or CG. The eigenvalues can be more accurate than those computed either by using a single scan signal $c^i(n)$ or by using the averaged signal. In particular, the complex frequencies that correspond to "noise" tend to have a large and positive imaginary part, which means that the noise spikes are effectively suppressed. However, it is less obvious how we can compute the complex amplitudes as accurately. Remember that we first need to normalize the eigenvectors with respect to $B_k^{\mathrm{T}} U_0 B_{k'} = \delta_{kk'}$, where $U_0$ is the overlap matrix, then compute the amplitudes as $\sqrt{d_k} = B_k^{\mathrm{T}} \tilde{\mathbf{C}}$. It is not clear how we can use $N_s$ copies of overlap matrices $U_0^{(i)}$ and FT vectors $\tilde{\mathbf{C}}^{(i)}$ jointly in the most efficient way. There are several formulas proposed and tested [96]. An expression that corresponds to the average of all $d_k^{(i)}$ was found to be optimal by numerical experiments,

$$\sqrt{d_k} = \frac{1}{N_s} \sum_{i=1}^{N_s} d_k^{(i)} = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[ \frac{\sum_{n=1}^{K_{\text{win}}} [\mathbf{B}_k]_n \tilde{\mathbf{C}}_n^{(i)}}{\sqrt{\mathbf{B}_k^{\top} \mathbf{U}_0^{(i)} \mathbf{B}_k}} \right] \tag{2.86}$$

There is another potential problem here. Compared to the averaged signal, the SNR of each individual scan is $\sqrt{N_s}$ lower. Thus the matrices in the super GEP are noisier, making the Snap-Shot FDM approach less attractive. Unlike the problem of the ambiguity of computing the amplitudes, here we have a good solution, which was also suggested by Neumaier [95]. Instead of using the noisy snap-shots $c^{(i)}(n)$ directly, we first reconstruct a new set of snap-shots according to,

$$c_{\text{new}}^{(i)}(n) = \bar{c}(n) + \frac{1}{N_s} \left( c^{(i)}(n) - \bar{c}(n) \right) . \tag{2.87}$$

The new signals have similar SNR to the average signal, while the random and uncorrelated noise is preserved. It has been shown by numerical experiments that this procedure does improve the results [96].

Tested with simple 1D experimental NMR signals, Snap-Shot FDM seems to be capable of separating "signal" from "noise". The complex frequencies computed typically consist of certain number of genuine poles with relative small (true) line width and a bunch of "noise" poles with much larger linewidth. The FDM ersatz spectrum constructed tends to show a clean "signal" spectrum with a fat baseline (no noise). Thus, the idea of "self-regularization" seems to work in these cases. However, when the signal gets more complicated and requires local spectral analysis, the results were less satisfactory. The self-regularization seems to be insufficient. When extended to the most interesting and most demanding case of multi-dimensional spectral analysis, the self-regularization seems to over-regularize the problem. Even genuine peaks are significantly broadened. Amplitudes are also quite inaccurate due to the ambiguity of the optimal formula. Nevertheless, the self-regularization is a very interesting idea, deserving of further studies, which might have profound impact on deepening our understanding of the optimal way of acquiring and processing time signals.

### 2.4.5 Summary

In this section, we discussed the ill-conditioning of multi-dimensional FDM and investigate various regularization techniques for suppressing the artifacts, among which the FDM2K algorithm seems to be particularly efficient and effective, even though by no means perfect or the best method for regularizing FDM. With regularization implemented, it is possible to obtain stable, high-resolution spectral estimations in a single FDM calculation, given that the signal sufficiently satisfies the Lorentzian model and is long enough. The resulting spectrum has a smooth dependence on the regularization level. The noise and small features are suppressed first. There is a range of regularization level where the spectrum changes very little while the regularization parameter $q^2$ is changed by an order of magnitude. When the regularization is too large, all features are significantly broadened but in a non-uniform way. How a particular peak is affected by the regularization depends on its intensity as well as it environments such as nearby peaks and (near) degeneracy with other peaks. The problem of regularization of FDM is only partially solved at present. More studies are needed for developing the ultimate method for regularizing FDM.

## 2.5 Regularized Resolvent Transform

There is a certain attraction to the idea of a *transform*: converting data from one representation to another, often more useful, representation. For example, Fourier Transformation converts a continuous function of time $c(t)$ to a function of frequency $I(\omega)$ by Fourier integration. For each frequency $\omega$, we can obtain a complex number that gives the local spectral amplitude $I(\omega)$. The simplicity and transparency of the transform make it appealing even when, as in the discrete Fourier Transform (DFT), the digital spectrum may deviate appreciably from the true integral representation [7]. By contrast, parametric methods that rely on fitting the data to a functional form are rather more complex in nature. There are typically adjustable parameters which need to be tuned to obtain satisfactory results. In section, we bridge the parameter estimation approach to the transform approach by introducing a new transform, the Regularized Resolvent Transform (RRT): spectral representations are computed directly using the same data matrices as those in FDM without the intermediate step of computing the spectral parameters. RRT maintains the spirit of local spectral analysis and can be implemented efficiently.

### 2.5.1 One-Dimensional RRT

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

Given a discrete equidistant time signal $c(n) \equiv c(n\tau)$, the goal is to estimate its infinite time DFT, defined in Eq. 2.26, but using only a finite portion of the data with $n = 0,\ 1,\ ...,\ N-1$. The assumption used here to derive the linear algebraic expressions is the same as the quantum ansatz used in FDM: $c(n)$ is associated with a time autocorrelation function of a fictitious dissipative quantum system defined by an effective evolution operator $\hat{U}$ and some initial state $\Phi_0$ [24],

$$c(n) = \left(\Phi_0|\hat{U}^n\Phi_0\right) . \tag{2.88}$$

The assumption that the evolution operator $\hat{U}$ has a finite rank is equivalent to assuming that the signal of question can be represented as a sum of damped sinusoids. Substituting Eq. 2.88 into Eq. 2.26 and evaluating the geometric summation analytically, we obtain:

$$\begin{aligned} I(\omega) &= \tau\left(\Phi_0\middle|\left\{\sum_{n=0}^{\infty} e^{in\tau\omega}\hat{U}^n - \frac{1}{2}\right\}\Phi_0\right) \\ &= \tau\left(\Phi_0\middle|\left\{\frac{1}{1-e^{i\tau\omega}\hat{U}} - \frac{1}{2}\right\}\Phi_0\right) . \end{aligned} \tag{2.89}$$

Note that $\hat{U}^\infty$ vanishes as the system is assumed to be dissipative.

Evaluated in an appropriate basis $\{|\Psi_j)\}$, the operator expression of Eq. 2.89 becomes a working formula for directly calculating $I(\omega)$,

$$I(\omega) = \tau\left[\mathbf{C}^{\mathrm{T}}\mathbf{R}(\omega)^{-1}\mathbf{C} - \frac{c(0)}{2}\right] , \tag{2.90}$$

with resolvent matrix $\mathbf{R}(\omega) = \mathbf{U}_0 - e^{i\tau\omega}\mathbf{U}_1$. The elements of the evolution and overlap matrices are defined respectively as, $[\mathbf{U}_1]_{jj'} = \left(\Psi_j|\hat{U}\Psi_{j'}\right)$ , $[\mathbf{U}_0]_{jj'} = \left(\Psi_j|\Psi_{j'}\right)$ , and the coefficients of the column vector $\mathbf{C}$ are $[\mathbf{C}]_j = (\Psi_j|\Phi_0)$. In particular, when a Fourier-type basis defined in Eq. 2.15 is used, the spectral properties around some frequency $\omega$ are completely defined by a very small subspace $\{|\Psi_j)\}$ of size $K_{\text{win}}$ with $\varphi_j \sim \omega$.

Therefore, only a small $K_{\text{win}} \times K_{\text{win}}$ matrix $\tilde{\mathbf{R}}(\omega)$ has to be inverted in Eq. 2.90 to yield an accurate spectral estimation, $I(\omega)$. The expressions for computing matrix elements are already given in Section 2.2 (Eq. 2.36 and Eq. 2.37). The column vector $\tilde{\mathbf{C}}$ is a FT of the original signal array $c(n)$,

$$[\tilde{\mathbf{C}}]_j = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} c(n) \,. \tag{2.91}$$

Clearly, evaluation of Eq. 2.90 can be done for all values of $\omega$ within a chosen frequency window by solving the corresponding generalized eigenvalue problem to obtain the eigenvalues $u_k$ and eigenvectors $\tilde{\mathbf{B}}_k$, which is the approach used in FDM. Although, this method might seem preferable to any other alternative, we note that Eq. 2.90 can also be evaluated directly, for example, by solving the associated linear system,

$$\tilde{\mathbf{R}}(\omega)\tilde{\mathbf{X}}(\omega) = \tilde{\mathbf{C}} \,, \tag{2.92}$$

and then using

$$I(\omega) \approx \tau \left[ \tilde{\mathbf{C}}^{\mathrm{T}} \tilde{\mathbf{X}}(\omega) - \frac{c(0)}{2} \right] \,. \tag{2.93}$$

Mathematically, both approaches should provide exactly the same results. However, numerically, the second approach might have advantages due to its simplicity and transparency. In addition, Eqs. 2.92 and 2.93, may actually appear computationally efficient if the spectrum is to be evaluated at relatively few values of $\omega$. This will be even more the case for the multi-dimensional spectral estimation.

A less obvious issue is how stable and robust the algorithm is. Apparently, the matrix $\tilde{\mathbf{R}}(\omega)$ may be very ill-conditioned, so its inversion or use in Eq. 2.92 requires some kind of regularization. One possibility is to use SVD of $\tilde{\mathbf{R}}(\omega)$ to calculate a pseudo inverse (see Section 2.4.2: soft regularization). However, SVD, if applied for each value of $\omega$, would be quite expensive. A much less expensive regularization of the resolvent can be obtained using the Tikhonov Regularization [93, 99],

$$I(\omega) \approx \tau \left[ \tilde{\mathbf{C}}^{\mathrm{T}} \left( \tilde{\mathbf{R}}(\omega)^{\dagger} \tilde{\mathbf{R}}(\omega) + q^2 \right)^{-1} \tilde{\mathbf{R}}(\omega)^{\dagger} \tilde{\mathbf{C}} - \frac{c(0)}{2} \right] , \tag{2.94}$$

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

Eq. 2.94 can be evaluated by solving the regularized Hermitian least squares problem,

$$\left( \tilde{\mathbf{R}}^{\dagger}(\omega) \tilde{\mathbf{R}}(\omega) + q^2 \right) \tilde{\mathbf{X}}(\omega) = \tilde{\mathbf{R}}^{\dagger} \tilde{\mathbf{C}} , \tag{2.95}$$

and then using Eq. 2.93.

It should be noted that the expensive matrix-matrix multiplication $\tilde{\mathbf{R}}(\omega)^{\dagger} \tilde{\mathbf{R}}(\omega)$ in Eq. 2.95, which is a $K_{\mathrm{win}}^3$ process, does not have to be performed at each value of $\omega$. Significant numerical saving could be achieved by using

$$\tilde{\mathbf{R}}(\omega)^{\dagger} \tilde{\mathbf{R}}(\omega) = \tilde{\mathbf{U}}_0^{\dagger} \tilde{\mathbf{U}}_0 + \tilde{\mathbf{U}}_1^{\dagger} \tilde{\mathbf{U}}_1 - e^{-i\tau\omega} \tilde{\mathbf{U}}_1^{\dagger} \tilde{\mathbf{U}}_0 - e^{i\tau\omega} \tilde{\mathbf{U}}_0^{\dagger} \tilde{\mathbf{U}}_1 .$$

The spectral estimation given by Eq. 2.94 is one of the main results of this section. Operationally it has a status of "transform" (like DFT), 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. 2.94 corresponds to a direct nonlinear transformation, here called the *Regularized Resolvent Transform* (RRT), of the time signal to the frequency domain spectrum. Unlike most other nonlinear high resolution spectral estimators, RRT is very stable, computationally inexpensive, and has adjusting parameters that are straightforward to use. These parameters are the size, $K_{\text{win}}$, and density, $\rho$, of the Fourier basis in the frequency domain, and the regularization parameter $q$. Note that $K_{\text{win}}$ could in principle be as small as 3, although a larger $K_{\text{win}}$ generally improves the resolution, while increasing the CPU-time according to the cubic scaling of a linear solver. For sufficiently large $K_{\text{win}}$, which is usually less than 100, the results do not change noticeably. Thus, $K_{\text{win}}$ 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, $\rho$, between 1.1 and 1.2 usually works well if a single-scale basis (as opposed to the *multi-scale* one [79]) is used.

The spectrum $I(\omega)$ computed by RRT with $q = 0$ should generally be indistinguishable from that computed by the FDM algorithm based on solving the generalized eigenvalue problem, Eq. 2.11, and using Eq. 2.27 for spectral estimation. However, for $q > 0$ they will differ as demonstrated in Figure 2.19 using the model signal, "Jacob's Ladder", described in Section 2.2. The signal contains 50 triplets of which both the peak widths and the splittings gradually decrease from the right to the left, allowing examination the breakdown of the resolution for any fixed signal size. The spectral region shown has very small spacings and is therefore hard to resolve. The upper trace is the exact spectrum that could be obtained by RRT using, for example, $N = 48K$ data points. The DFT spectrum using $N = 64K$ cannot resolve all the triplets in this region, while using DFT with $N = 32K$ does not resolve any of them. In addition, one can see some baseline ringing that is not completely removed by apodization of the signal. The RRT result using $N = 32K$ and $q = 0$ is very close to the exact result, although some linewidths and amplitudes are slightly inaccurate. Interestingly, one

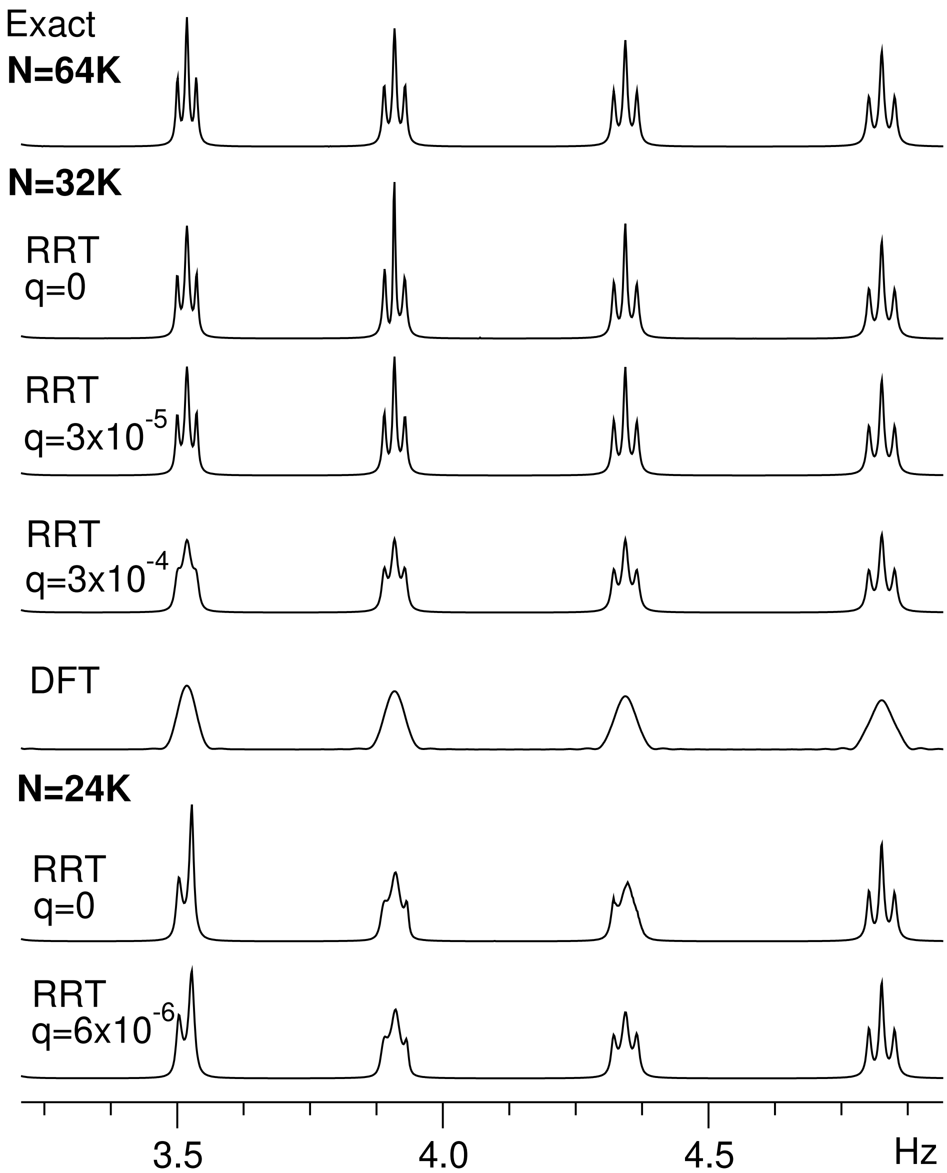



Figure 2.19: Absorption spectra obtained by processing a model time signal, "Jacob's ladder", by the *Regularized Resolvent Transform* (RRT) and DFT using different sizes of the signal, $N = 32,000$ and $N = 24,000$. The RRT spectra have much higher resolution than the DFT spectrum with $N = 32,000$. The non-regularized spectra ($q = 0$) may be quite non-uniform. An increase of the regularization parameter $q$ generally leads to a more uniform spectral estimate and gradual decrease of resolution, however the best result is often obtained using RRT with some small $q > 0$. The $q$ values shown are relative values. The actual value used in Eq. 2.94 is computed as $q^2_{\text{act}} = q^2 < [R^\dagger R]_{ii} >_i$.

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. Note that the scale for $q$ is related to the normalization of the U-matrices.

The lower two traces show RRT spectra using $N = 24K$, which is too short for the method to resolve the narrowest multiplets in this spectral region. The resolution failure for $q = 0$ manifests itself in a quite non-uniform 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 $\sim$ 22 Hz). That is, for short data sets the RRT "fails" in a controllable fashion (as is the case for DFT), while typically providing a higher resolution than DFT 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. 2.1. For example, in RRT the derivation of Eq. 2.89 assumes convergence of the infinite geometric series $\sum_{n=0}^{\infty}\left(e^{i\tau\omega}\hat{U}\right)^n$. 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 linewidth $\gamma_k$. Section 2.2.3 contains a detailed discussion of how to deal with these "spurious" poles. However, in RRT, the spectral parameters are not explicitly computed, making direct manipulation of individual poles impossible. This has both advantages and disadvantages. One one hand, we can conserve the mutual interferences and cancellations to the maximum extend; on the other hand, when the true peak width is very small or zero as in Constant-Time NMR experiment (see

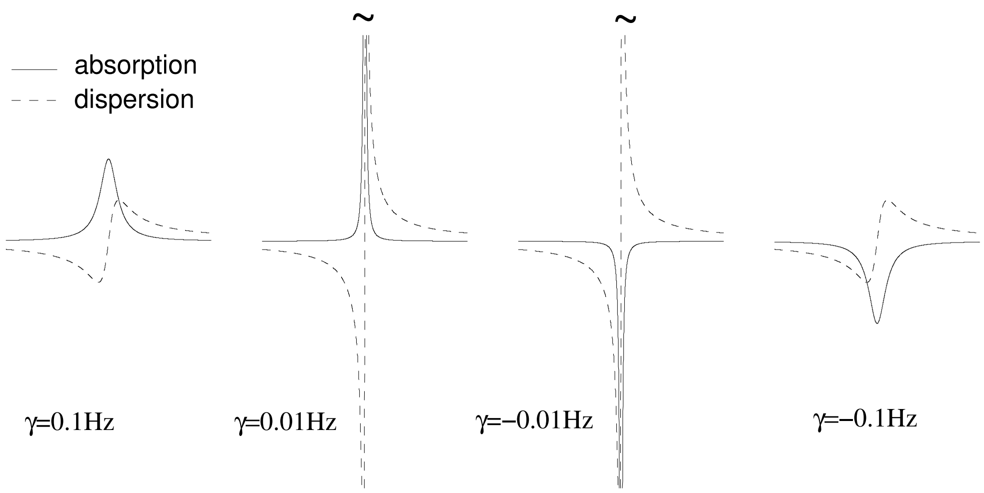



Figure 2.20: The real (absorption) and imaginary (dispersion) parts of the RRT spectrum $I(\omega) = \tau d_k/(1 - e^{i\tau(\omega-\omega_k)})$ for a single line as a function of its "width" $\gamma = -\mathrm{Im}\{\omega_k\}$. The sign of the absorption peak is flipped when $\gamma$ changes its sign, while the appearance of the dispersion peak only depends on the absolute value of $\gamma$.

Chapter 3 for more details), it is likely to have $\gamma_k < 0$ due to noise and/or numerical errors. This in turn results in a flip of the sign of the absorption part of the peak as shown in Figure 2.20. 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 $I(\omega)$ (see Eq. 2.27) near an eigenfrequency $\omega_k = \omega_r - i\gamma$, i.e., assuming $\tau|\omega - \omega_k| \ll \pi$ and then extracting the real and imaginary parts of the complex Lorentzian as

$$I(\omega) \approx \frac{\tau d_k}{1 - e^{i\tau(\omega-\omega_k)}} \approx \frac{id_k}{\omega - \omega_k} = \frac{d_k\gamma}{(\omega-\omega_r)^2 + \gamma^2} + i\frac{d_k(\omega-\omega_r)}{(\omega-\omega_r)^2 + \gamma^2}\,. \qquad (2.96)$$

Since in RRT the eigenvalues are not computed, one cannot manipulate with those in a simple fashion like in FDM. One way to circumvent this problem in RRT is to shift the argument of $I(\omega)$ by $i\Gamma$ with $\Gamma > -\gamma_k$ , i.e, construct $I(\omega + i\Gamma)$. It replaces $\gamma_k$ of all peaks by $\gamma_k + \Gamma$ and effectively flips the negative $\gamma_k$. The result of such a shift is demonstrated in the third trace of Fig. 2.21 where all the peaks have the correct sign, but are slightly broadened. Clearly, this procedure may not be always well defined and is one drawback of RRT. For example, peaks with $\gamma_k < -\Gamma$ are effectively narrowed and might result in artifacts.

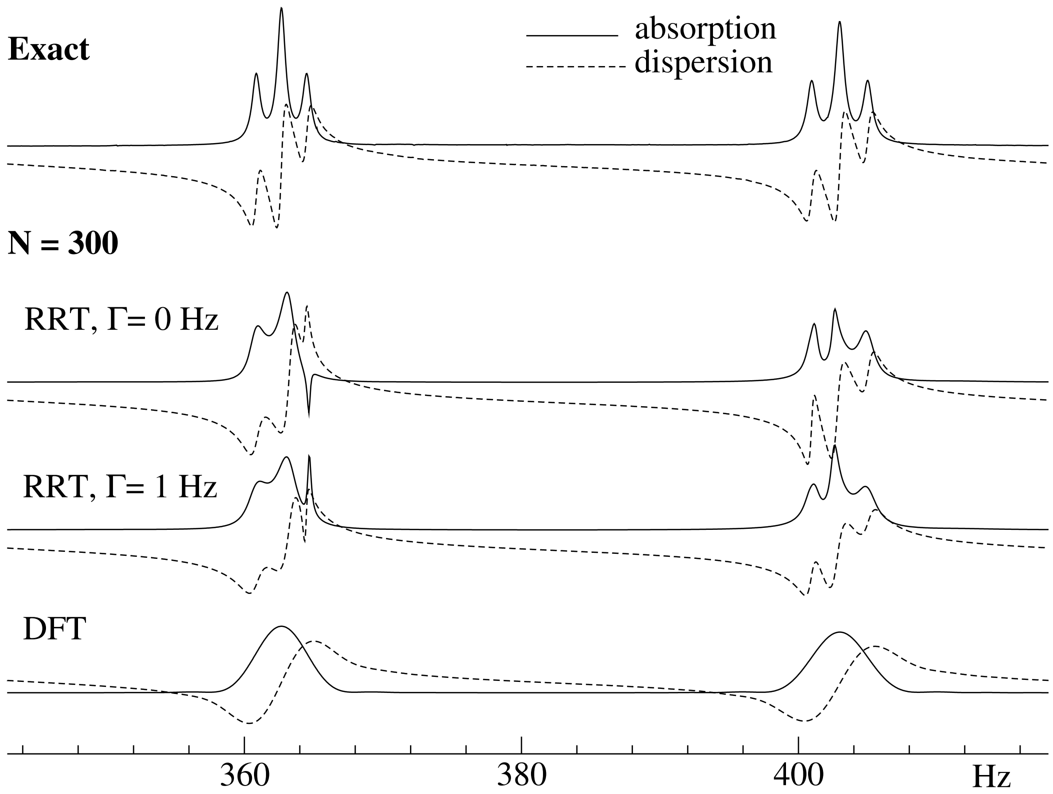



Figure 2.21: An example of a failure of the RRT spectral representation with no shift and no regularization implemented (second trace, both $q = 0$ and $\Gamma = 0$). Because of a too short signal used to construct the spectrum both the amplitudes and widths of the peaks are quite inaccurate, in particular, one peak appears with the wrong sign (note that the appearance of the dispersion peaks is correct). The wrong absorption peak is flipped by plotting $I(\omega + i\Gamma)$ (the lowest trace) instead of $I(\omega)$ with $\Gamma = 1$ Hz. Note, that this procedure replaces $\gamma_k$ of all peaks by $\gamma_k + \Gamma$, i.e., effectively fixes the signs of all peaks with $\gamma_k > -\Gamma$ and broadens all the peaks with the correct $\gamma_k > 0$.

### 2.5.2 Multi-Dimensional RRT

Generalization of RRT to a multidimensional case is essentially equivalent to how the 1D FDM is generalized. 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 $c(\vec{n}) \equiv c(n_1\tau_1, n_2\tau_2)$ defined on an equidistant rectangular time grid, our goal is to estimate the 2D (infinite time) discrete Fourier sum, defined in Eq. 2.58, using only the finite $N_1 \times N_2$ part of the signal. The 1D quantum ansatz of Eq. 2.88 is generalized by using two commuting complex symmetric evolution operators $\hat{U}_1$ and $\hat{U}_2$,

$$c(\vec{n}) \equiv c(n_1\tau, n_2\tau_2) = \left(\Phi_0|\hat{U}_1^{n_1}\hat{U}_2^{n_2}\,\Phi_0\right) \ . \tag{2.97}$$

Inserting Eq. 2.97 into Eq. 2.58 and evaluating the geometric sums analytically gives

$$I(\omega_1,\omega_2) = \tau_1\tau_2\left(\Phi_0\left|\left\{\frac{1}{1-e^{i\tau_1\omega_1}\hat{U}_1}-\frac{1}{2}\right\}\times\left\{\frac{1}{1-e^{i\tau_2\omega_2}\hat{U}_2}-\frac{1}{2}\right\}\Phi_0\right.\right) \ . \tag{2.98}$$

Evaluated in the 2D Fourier basis, defined in Eq. 2.50, the operator equation of Eq. 2.98 becomes the working resolvent transform expression for estimating complex infinite time 2D DFT spectrum,

$$\begin{aligned} I(\omega_1,\omega_2) \ &\approx \ \tau_1\tau_2 \\ &\times \ \left\{\tilde{\mathbf{C}}^{\mathrm{T}}\left[\tilde{\mathbf{R}}_1(\omega_1)^{-1}\tilde{\mathbf{U}}_0\,\tilde{\mathbf{R}}_2(\omega_2)^{-1}-\frac{\tilde{\mathbf{R}}_1(\omega_1)^{-1}-\tilde{\mathbf{R}}_2(\omega_2)^{-1}}{2}\right]\tilde{\mathbf{C}}+\frac{c(0,0)}{4}\right\} \ , \end{aligned} \tag{2.99}$$

with resolvent matrices

$$\tilde{\mathbf{R}}_l(\omega_l) = \tilde{\mathbf{U}}_0 - e^{i\tau_l\omega_l}\tilde{\mathbf{U}}_l \ , \quad l = 1, 2 \ , \tag{2.100}$$

where $\tilde{\mathbf{U}}_l$ matrices and the Fourier vector $\tilde{\mathbf{C}}$ can be evaluated using the expressions derived previously in Section 2.3 (Eqs. 2.52, 2.53, 2.54, and 2.57).

A 2D RRT can be obtained by regularizing the two resolvents in Eq. 2.99, leading to

$$I(\omega_1,\omega_2) \approx \tau_1\tau_2 \left\{ \tilde{\mathbf{X}}(\omega_1)^{\mathrm{T}}\tilde{\mathbf{U}}_0\tilde{\mathbf{X}}(\omega_2) - \frac{\tilde{\mathbf{X}}(\omega_1)^{\mathrm{T}}\tilde{\mathbf{C}} + \tilde{\mathbf{C}}^{\mathrm{T}}\tilde{\mathbf{X}}(\omega_2)}{2} + \frac{c(0,0)}{4} \right\} , \quad (2.101)$$

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

$$\left(\tilde{\mathbf{R}}_l(\omega_l)^{\dagger}\tilde{\mathbf{R}}_l(\omega_l) + q^2\right)\tilde{\mathbf{X}}_l(\omega_l) = \tilde{\mathbf{R}}_l(\omega_l)^{\dagger}\tilde{\mathbf{C}} \, . \quad (2.102)$$

Note that the total number of the linear systems to be solved for each 2D frequency window is equal to $N_{\omega_1} + N_{\omega_2}$ instead of $N_{\omega_1} \times N_{\omega_2}$, where $N_{\omega_1}$ and $N_{\omega_2}$ are the numbers of the frequency grid points, $\omega_1$ and $\omega_2$, used to plot the 2D spectrum in the window.

Derivation of the 2D RRT is another main result of this section. Similar to the 2D FDM, its substantial advantage as compared to the 2D DFT is 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, although the two methods are very closely related. In particular, if the form of Eq. 2.42 is satisfied, the RRT spectrum for a sufficiently large data set processed should converge to

$$\begin{aligned} I(\omega_1,\omega_2) &= \tau_1\tau_2 \sum_{k=1}^{K} \frac{d_k}{2} \left\{ \frac{2}{\left[1-e^{i\tau_1(\omega_1-\omega_{1k})}\right]\left[1-e^{i\tau_2(\omega_2-\omega_{2k})}\right]} \right. \quad (2.103) \\ &- \left. \frac{1}{\left[1-e^{i\tau_1(\omega_1-\omega_{1k})}\right]} - \frac{1}{\left[1-e^{i\tau_1(\omega_2-\omega_{2k})}\right]} \right\} + \frac{\tau_1\tau_2\, c(0,0)}{4} \, . \end{aligned}$$

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 $N_1$ is small. This, of course, does not take into account noise, degeneracies and roundoff errors which could affect the convergence conditions significantly.

Due to its numerical stability, robustness and ease implementation, the RRT has an advantage over FDM, when only complex 1D or 2D FT type spectra, as defined by Eqs. 2.26 and 2.58, are of interest. Note, yet, an advantage of FDM, especially, in the multidimensional case, lies in its ability to construct various nonanalytic spectral representations, such as a double-absorption spectrum using a single purely phase modulated 2D data due to the availability of the spectral parameters (Eq. 2.68). Furthermore, in FDM inaccuracies in determining the imaginary parts of the computed frequencies, in particular, their signs, might be corrected. In RRT the spectral parameters are not produced and therefore the artifacts are removed by regularization with a finite value of $q$ and/or by shifting the spectrum using $I(\omega_1 + i\Gamma_1, \omega_2 + i\Gamma_2)$ with shifting parameters $\Gamma_1$ and $\Gamma_2$.

In the next example we apply 2D RRT to the process the same $^{15}$N-$^{1}$H HSQC signal of rubredoexin used in previous examples. Figure 2.22 compares the RRT and DFT spectra obtained using various signal sizes. Even though this signal is particular favorable for DFT as it contains pure singlet of similar intensity, RRT still provides a welcome improvement in resolution or a decrease in experimental time for given resolution. As in FDM, the gain in resolution hinges on the total information content of the signal, which depends on signal size, signal to noise ratio, amount of various imperfections and others. It is neither possible to improve the resolution of very noisy spectra, nor to magically divine the presence of peaks buried in noise.

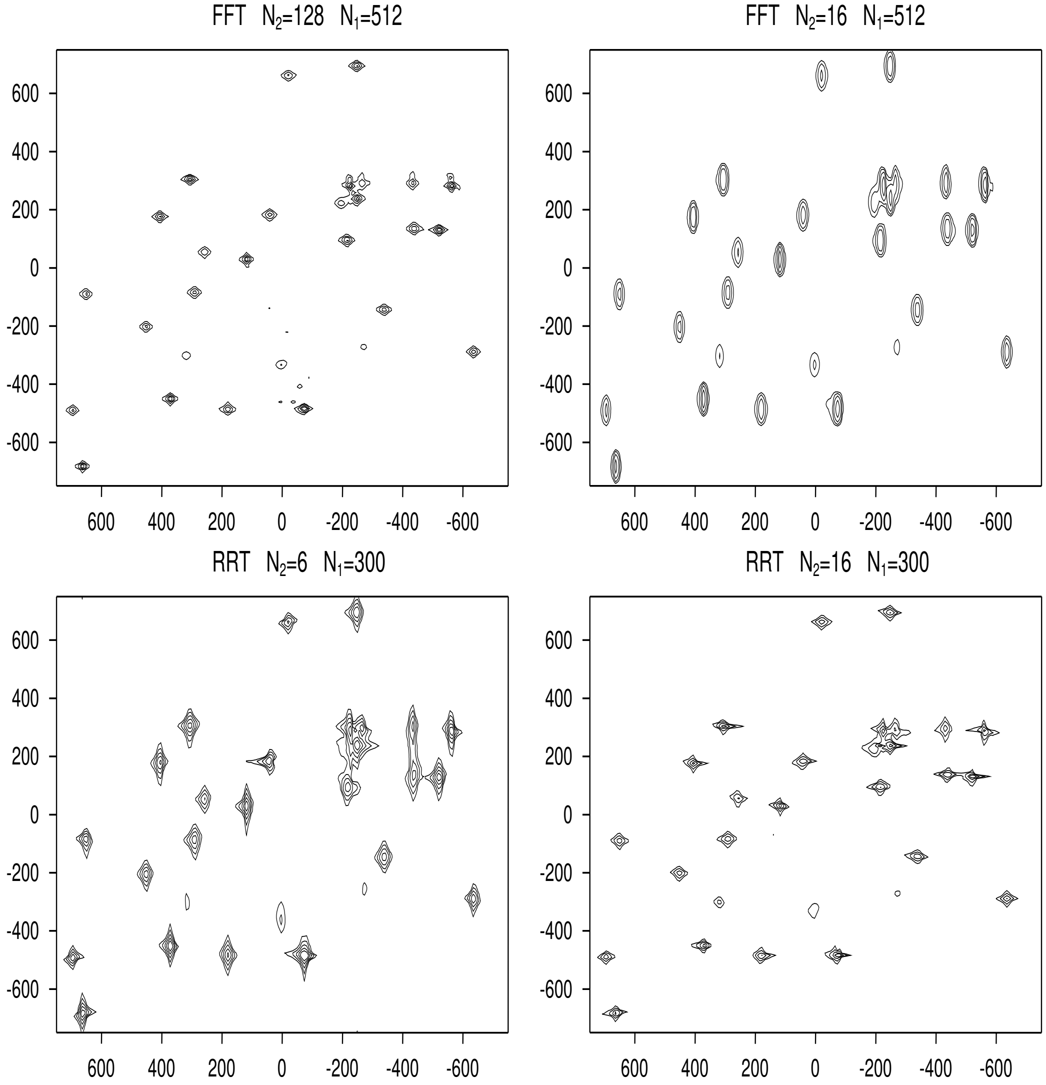



Figure 2.22: Chemical shift correlation spectra of the metalloprotein rubredoxin obtained by DFT and RRT using various signal sizes. All the double-absorption spectra are generated using the conventional procedure which combines the complex N- and P-type complex spectra. Regularization parameters are optimized for both RRT calculations. The shifting parameters $\Gamma_1 = \Gamma_2 = 1\,Hz$.

### 2.5.3 Other Spectral Representations Using RRT

The resolvent formulas of Eq. 2.90 and Eq. 2.99 provide us a way to estimate the infinite time DFT spectra using only finite data sets. In this section, I would like to demonstrate that RRT is a more powerful method than just nonlinear FT-type spectral estimator. With minor modifications, the resolvent formulas can be used to construct many other types of useful spectral representations, revealing more information about the signal.

**1. Inverse Laplace transform via RRT: iRRT**

Any complex variable can be used in RRT. In particular, when the frequency variable is purely imaginary, RRT becomes an efficient expression for computing inverse Laplace transform without exponential instabilities. It has been shown that it is possible to recover up to 4 exponential decaying constants with significant gap between the fastest and lowest decaying components in the Quantum Monte Carlo (QMC) calculations of vibrational states of polyatomic systems [98]. The RRT expression with imaginary frequency variable is also called *iRRT*. iRRT has been also applied to processing the Diffusion Ordered Spectroscopy (DOSY) [100, 101] NMR experimental signals, showing the promise to identifying multiple diffusion constants even in the heavily overlapped spectral regions [96]. However, the overall results have not been totally satisfactory considering the reliability of the results and accuracy of intensities. There are several practical difficulties: first, accurate characterization of both fast and slow decaying components requires signals with very high signal to noise ratio, while typical DOSY experimental data sets are very noisy due to the limited quality of magnetic field gradient and many other experimental limitations [102]. Second, fitting data to multiple exponential decaying components has been proved to be a notoriously unreliable procedure. Even though the iRRT expression seems to be very stable, there are still uncertainties such as the effect of regularization on the accuracy of computing the decaying constants. It is necessary to study the results as a function of the signal length (or other parameters such as basis size and regularization) to check the validity of the results. Finally, the intensity computed by iRRT is not accurate enough for quantitative usage and it is not obvious how it can be improved.

**2. Pseudo-2D RRT spectra of both real and imaginary frequencies**

There is another an interesting spectral representation that can be constructed using RRT with complex variables. By computing the spectral intensity $I(\omega + i\Gamma)$ as a function of both the real frequency $\omega$ and imaginary frequency $\Gamma$, a *pseudo*-2D spectrum results. As $I(\omega + i\Gamma)$ diverges at $\omega + i\Gamma = \omega_k$ on the complex plane, we can identify both the peak positions and the linewidths simply by looking at the singular points on the 2D contour plot. In addition, it is also possible to separate peaks that are severely overlapped according to their linewidths. Figure 2.23 shows a contour plot of such a *pseudo*-2D spectrum, applied to a 1D NMR signal. Note that a slice along $\Gamma = 0$ Hz corresponds to the normal 1D FT-type absorption spectrum with no smoothing, which is shown on top of the 2D plot for direct comparison. The 2D plot reals more information about the signal. For example, the 2D plot shows that there are possibly two peaks, **A** and **B**, at around $\omega = -5600$ Hz (indicated by two arrows), while this information is totally hidden in the conventional 1D FT spectrum due to the heavy overlapping.

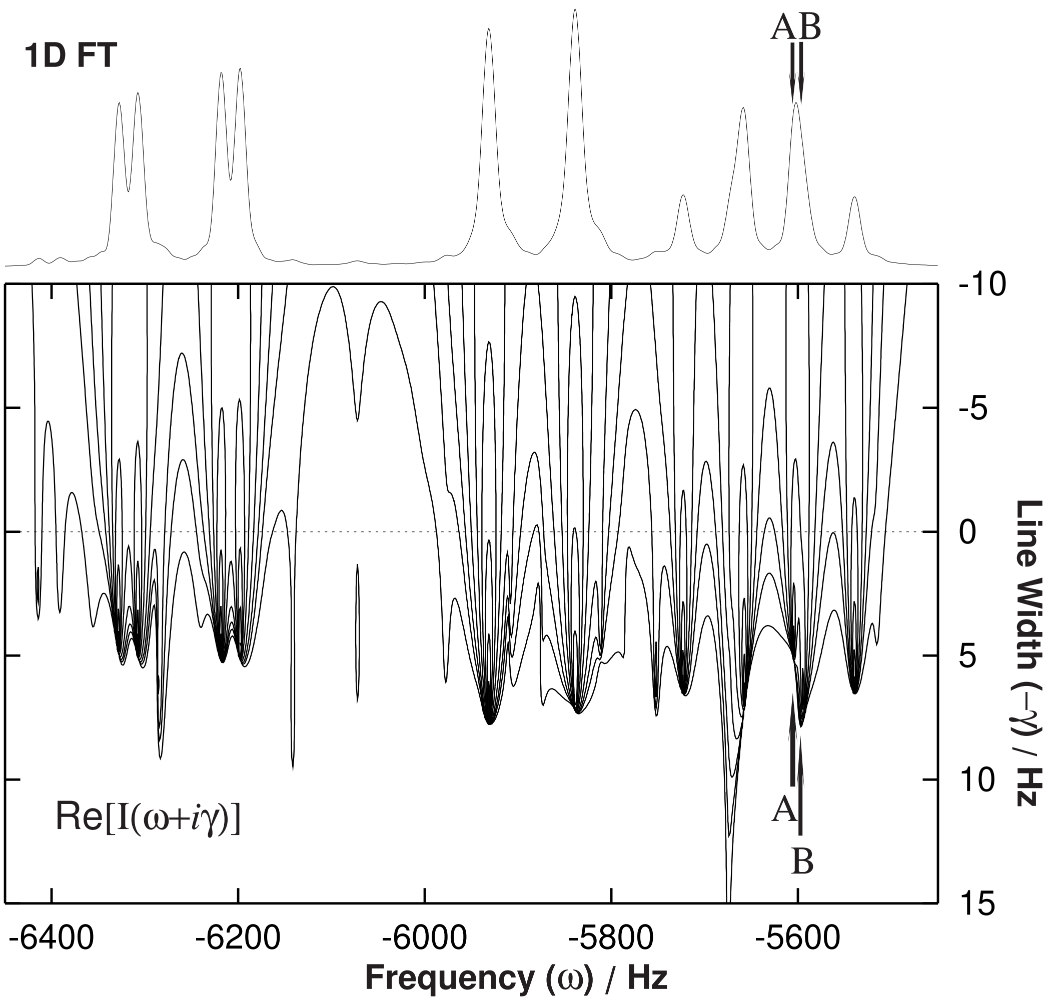



Figure 2.23: Contour plot of $I(\omega + i\Gamma)$ on the complex plane. Peaks are spread out along the imaginary frequency axis according to the line width, providing more information than the conventional 1D FT spectrum (shown on the top). For example, the pseudo-2D plot indicates that there are possibly two peaks, noted as A and B, around -5600 Hz, while from the 1D FT spectrum this information is hidden due to the heavy overlapping of the two peaks. Note that a slice at $\gamma = 0$ of the pseudo-2D spectrum corresponds to the 1D FT-type spectrum without any linebroadening.

**2. Pseudo-absorption 2D RRT spectral estimations**

When severely truncated data sets are used in RRT, just like in FDM, while the positions of the peaks converge best, the phases of the complex amplitudes are least accurate, which manifest themselves as the incorrect phases of corresponding peaks in the double-absorption spectrum obtained by combining N- and P-type complex spectra. For example, close examination of the RRT spectra shown in Figure 2.22 reveals several peaks slightly out of phase, e.g., the peak at around (-600 Hz, 350 Hz). In principle, increasing regularization will suppresses the phase instability. However, numerical experiments showed that the phase stability improved slowly with regard to RRT regularization, especially for truncated signals. Moreover, in certain situations, only a single purely phase modulated data set is available (e.g., in 2D J experiments). In such cases an absolute value RRT spectrum may be an appealing and/or the only option, although the resolution of the latter are unacceptable due to the contribution of the dispersion lineshapes. Note, yet, there exists a variety of other spectral representations, in which the dispersion contributions are eliminated, therefore leading to much higher resolution. An example of such a representation in 2D, which is also used in 2D FT spectral analysis, is

$$A(\omega_1, \omega_2) = \mathrm{Re}\{I(\omega_1, \omega_2 + i\Gamma_2) - I(\omega_1, \omega_2 + i\Gamma_2')\} \ . \tag{2.104}$$

The dispersion contributions along $\omega_2$ are eliminated by the subtraction of the two complex spectra using different shifts $\Gamma_2 \neq \Gamma_2'$ (with possibly positive $\Gamma_2$ and negative $\Gamma_2'$), while taking the real part leads to the absorption lineshapes along $\omega_1$. The disadvantage of Eq. 2.104 is the need to fiddle with too many adjusting parameters.

Another very useful spectral representation is given by

$$I^{(2)}(\omega_1, \omega_2) \quad = \quad \left( \Phi_0 \left| \frac{\tau_1 \tau_2}{\left[1 - e^{i\tau_1\omega_1}\hat{U}_1\right]^2 \left[1 - e^{i\tau_2\omega_2}\hat{U}_2\right]^2} \Phi_0 \right. \right)$$

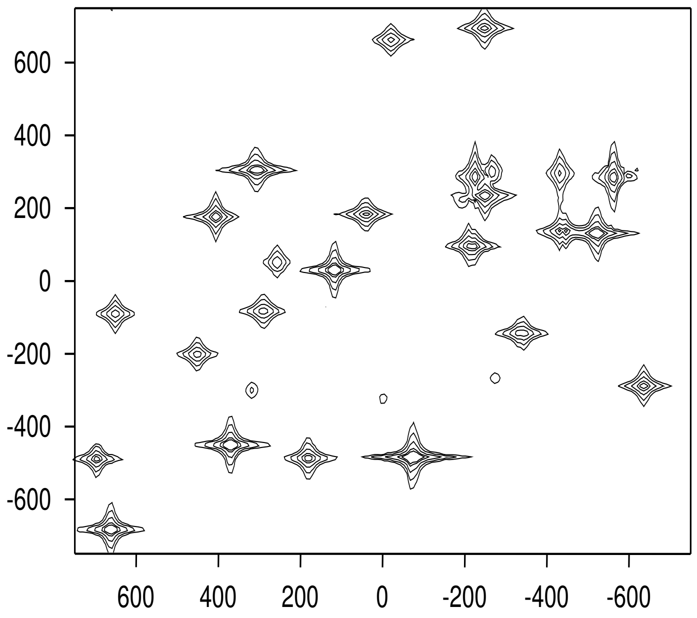



Figure 2.24: RRT pseudo-absorption spectrum $\left|I^{(2)}(\omega_1 + i\Gamma_1, \omega_2 + i\Gamma_2)\right|$ (see Eq. 2.105) constructed using only the N-type (i.e., the purely phase modulated) data of Figure 2.22 with $N_1 = 16$, $N_2 = 300$, $q = 0.0003$, $\Gamma_1 = 2$ Hz and $\Gamma_2 = 1$ Hz.

$$= \sum_{k=1}^{K} \frac{\tau_1 \tau_2 d_k}{\left[1 - e^{i\tau_1(\omega_1 - \omega_{1k})}\right]^2 \left[1 - e^{i\tau_2(\omega_2 - \omega_{2k})}\right]^2} . \tag{2.105}$$

The absolute value spectrum $\left|I^{(2)}(\omega_1, \omega_2)\right|$ will have a great resolution advantage over $|I(\omega_1, \omega_2)|$. The peaks are narrow and have quasi-absorption lineshapes in $\left|I^{(2)}(\omega_1, \omega_2)\right|$. Note that the amplitude $d_k$ is not squared. Thus the peak integral will effectively be divided by the product of the linewidths in two dimensions, $\gamma_{1k}\gamma_{2k}$. Accordingly, $I^{(2)}(\omega_1, \omega_2)$ will favor the narrow peaks against the broad ones. This non-uniform distortion can be reduced by shifting the spectrum using $\left|I^{(2)}(\omega_1 + i\Gamma_1, \omega_2 + i\Gamma_2)\right|$ which effectively increases the widths of all peaks making them more uniform. While $I^{(2)}(\omega_1, \omega_2)$ cannot be constructed by FT, it is representable in terms of the U-matrices in complete analogy with Eq. 2.99. Of course, the corresponding resolvents should be regularized accordingly. In Figure 2.24 we show the RRT spectra using $\left|I^{(2)}(\omega_1 + i\Gamma_1, \omega_2 + i\Gamma_2)\right|$ with $\Gamma_1 = 1$ Hz and $\Gamma_2 = 10$ Hz using the same 2D signal as in Figure 2.22. The peaks do have absorption lineshapes, although the heights are somewhat distorted when compared to the correct spectra of Figure 2.22. Higher order pseudo-absorption spectra can also be easily constructed by using higher power of the resolvent matrices. In the later cases, SVD is preferred, as once decomposition of the resolvent matrix $\mathbf{R}(\omega_l)$ is available, it is trivial to construct regularized pseudo-inverse matrix $\mathbf{R}_q^{-n}(\omega_l)$ with same computational efforts for any integer $n$.

### 2.5.4 Extended Fourier Transform

As we discussed before, the phase computed by RRT seems to be the least stable. Especially when the signal is severely truncated in all dimensions (even though it rarely so in NMR experiments), the RRT spectra could be significantly distorted. Increasing the regularization does not help much here unless the regularization level is so high that most of the features are smoothed out. This is a disadvantage of RRT. On the other hand, even though FT converges slowly with regard to the signal size, it does not fail even in the severely truncated cases. The eXtended Fourier Transform (XFT), developed by Armstrong and Mandelshtam [103], is a hybrid method of DFT and RRT designed to combine the high resolving power of RRT and the stability of DFT. The basic idea is to use RRT only to estimate the remainder of finite DFT of the signal. The details of XFT can be found in the original reference [103]. Here we give a brief derivation of 1D XFT for the illustration purpose.

Given a finite time signal $c(n)$ defined on an equidistant time grid of $N$ points, we want to estimate the infinite time DFT spectrum by breaking it into two parts[15],

$$I(\omega) = \tau \sum_{n=0}^{\infty} c(n) e^{i\omega n\tau} = \tau \left\{ \sum_{n=0}^{N-1} c(n) e^{i\omega n\tau} + \sum_{n=N}^{\infty} c(n) e^{i\omega n\tau} \right\} . \tag{2.106}$$

Note that the first summation is simply the conventional finite DFT summation. Due to the truncation of the signal, the second term can not be directly computed, which is the source of FT time-resolution uncertainty principle and the sinc-like truncation artifacts (see Chapter 1). However, it can be estimated using the RRT expression,

$$\begin{aligned}
\Delta_N(\omega) &= \sum_{n=N}^{\infty} c(n) e^{i\omega n\tau} = \sum_{n=N}^{\infty} e^{i\omega n\tau} \left( \Phi_0 \middle| \hat{U}^n \middle| \Phi_0 \right) \\
&= \sum_{n=N}^{\infty} e^{i\omega n\tau} \left( \Phi_{N/2} \middle| \hat{U}^{n-N} \middle| \Phi_{N/2} \right) \\
&= e^{i\omega N\tau} \sum_{n'=0}^{\infty} \left( \Phi_{N/2} \middle| \hat{U}^{n'} e^{i\omega n'\tau} \middle| \Phi_{N/2} \right) \\
&= e^{i\omega N\tau} \left( \Phi_{N/2} \middle| \frac{1}{1 - e^{i\tau\omega}\hat{U}} \middle| \Phi_{N/2} \right) .
\end{aligned} \tag{2.107}$$

Evaluated in the Fourier basis defined in Eq. 2.15, we can obtain the numerical expression with regularization for estimating the remainder of DFT,

$$\Delta_N(\omega) = e^{i\omega N\tau} \, \tilde{\mathbf{C}}_{N/2}^{\mathrm{T}} \, \tilde{\mathbf{R}}_q^{-1}(\omega) \, \tilde{\mathbf{C}}_{N/2} \, , \tag{2.108}$$

with the same resolvent matrix $\mathbf{R}(\omega) = \mathbf{U}_0 - e^{i\tau\omega}\mathbf{U}_1$ and a shifted FT vector,

$$[\tilde{\mathbf{C}}_{N/2}]_j = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} \left( \Phi_n | \Phi_{N/2} \right) = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} c(n + N/2) \, . \tag{2.109}$$

Therefore, XFT is essentially a straightforward extension of RRT. We simply need to construct the shifted FT vector $\mathbf{C}$, evaluate the RRT estimation of the DFT remainder, then add the correction to the DFT summation. Figure 2.25 is an interesting

[15] The zero correction term is dropped out here for the sake of simplicity.

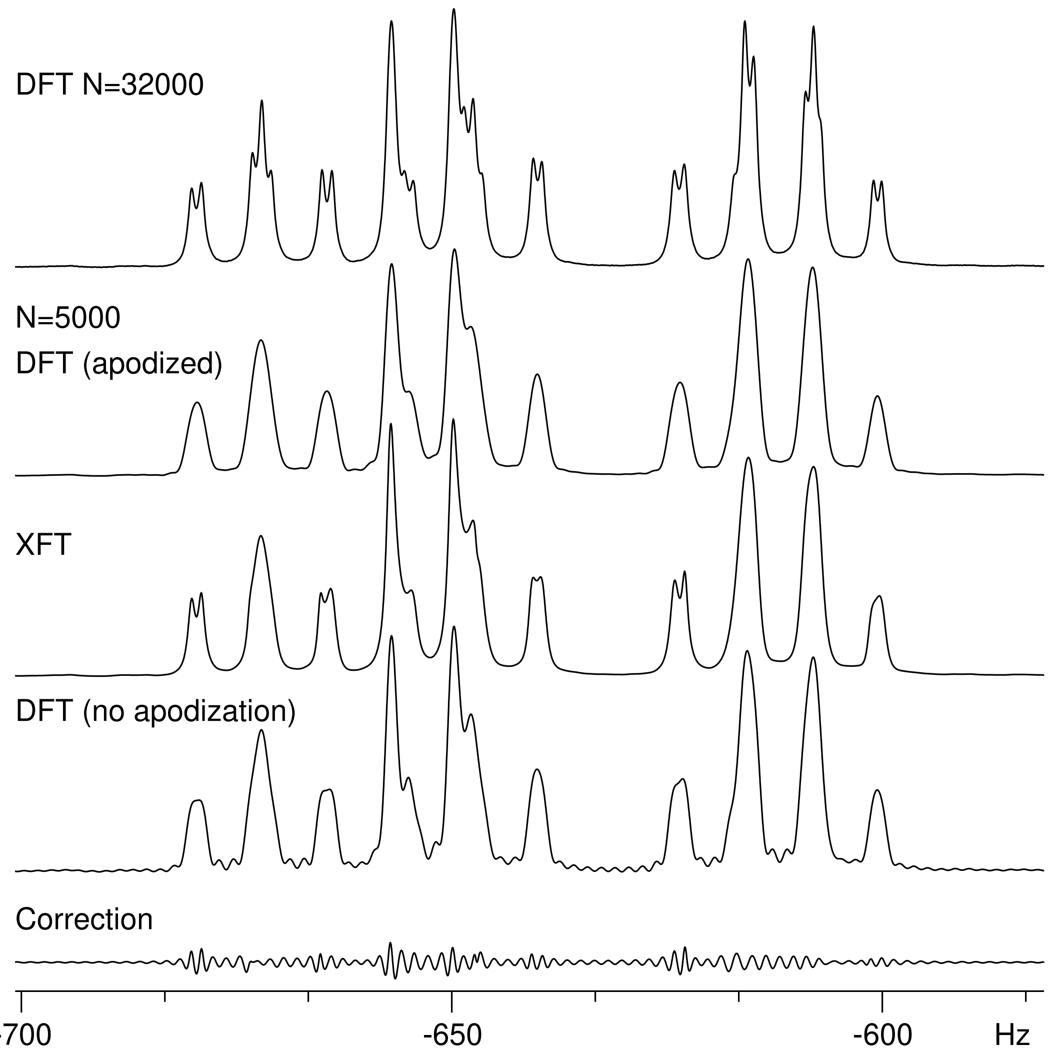



Figure 2.25: An interesting region of the 1D NMR spectrum of 12$H$-pyrido[1,2-$a$:3,4-$b$]diinole, processed by the XFT using a truncated (N=5000) data set. The figure illustrates the behavior of each term involved in the calculation of the XFT. The XFT is obtained by adding together the unapodized DFT term, and the correction term calculated with a regularization parameter of $q$ that is optimized for this case. The correction term exactly cancels the sinc-like oscillations present in the unapodized DFT. It also adds small terms to increase the resolution of some peaks, allowing some doublet structures to be identified. For comparison the apodized DFT spectra for N=32000 and N=5000 are presented. (This figure is courtesy of G. S. Armstrong [103]).

illustration of how XFT works. The XFT of an experimental NMR signal is used to examine the roles of both terms in Eq. 2.106. The lower three traces were computed with only the first 5000 data points. The DFT trace shows several unresolved multiplets and sinc-oscillations in the baseline. The correction trace has oscillations equal and opposite to those in DFT trace. When added together, the oscillations cancel, giving the correct baseline in the XFT trace. Furthermore, the correction contains some high resolution features, revealing some splittings that are not resolved in the DFT spectrum.

It is less obvious but can be proved that when the signal satisfies the Lorentzian model and is sufficiently large, the XFT spectrum, Eq. 2.106, gives the exact infinite time DFT of the signal. By separating the infinite DFT summation into two parts and evaluating them independently provides us a way to combine the stability of DFT and high-resolution of RRT together: on one hand, one can use a large regularization to suppress the contribution of RRT correction, providing a more conservative, finite DFT like spectrum; on the other hand, one can optimize the regularization for maximum resolution enhancement. Thus, XFT does not fail even when the signal is too short to obtain meaningful high-resolution spectrum. This is a very attractive property, especially considering that in FDM and RRT, it is either "everything" or "nothing" for certain cases.

XFT can be also extended to processing multi-dimensional data sets. The idea is also to use RRT only for estimating the remainder of 2D DFT. Details can be found in Ref. [103]. Applied to experimental 2D NMR signals, it has been shown that even for signals severely truncated in both dimensions, 2D XFT can provide very stable

2D spectral estimation with resolution similar to conventional 2D DFT but *without* any sinc-like truncation artifacts. In the case of large data sets, 2D XFT is able to provide higher resolution beyond FT uncertainty principle. However, the phase instability seen in 2D RRT is still present in high-resolution the 2D XFT spectra, which is a little disappointing but reasonable as RRT is responsible for the high-resolution features. In addition, the shift FT vector $\mathbf{C}_{N/2}$ is computed using the tail of the signal and therefore more noisy than the FT vector used in RRT, making the phase instability potentially more severe in the XFT high-resolution estimation.

### 2.5.5 Conclusions

In this section, we derived a new numerical expression, the Regularized Resolvent Transform (RRT), which corresponded to a direct transformation of the time-domain data to a frequency-domain spectrum. RRT 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 number of damped sinusoids, turns out to be equivalent to the exact infinite time discrete Fourier transformation. RRT naturally emerges from the Filter Diagonalization Method (FDM), but no diagonalization is required. The spectral intensity at each frequency $\omega$ is expressed in terms of the resolvent $\mathbf{R}(\omega)^{-1}$ of a small data matrix $\mathbf{R}(\omega)$ constructed from the time signal. Generally $\mathbf{R}$ is singular, and thus its inversion requires certain regularization. Among various possible regularizations of $\mathbf{R}^{-1}$, the Tikhonov regularization appears to be computationally both efficient and stable. Numerical implementation of RRT is very inexpensive because even for extremely large data sets the matrices involved are small.

We also showed that RRT could be used to constructed various non-FT spectral representations. In particular, using a purely imaginary frequency variable, RRT becomes an expression for estimating inverse Laplace transform without exponential instabilities. RRT can also be used to compute a *pseudo*-2D spectrum, $I(\omega+i\Gamma)$, with respect to both the real and the imaginary frequencies, revealing more information than the conventional FT-type spectral estimation. Finally, pseudo-absorption type spectra can be also constructed by RRT when an absolute value spectrum is desirable.

When the signal is severely truncated in all dimensions, 2D RRT spectra might be highly distorted due to the phase instability. Even though it can be improved by constructing a pseudo-absorption spectrum, there exists a more elegant alternative. The eXtended Fourier Transform, or XFT, attempts to combine the stability of DFT and high-resolution of RRT, by using RRT only to estimate the remainder of finite time DFT. XFT can provide a stable high-resolution estimation when the data set is sufficiently large and does not fail even the data set is extremely small.

## 2.6 Reliability and Sensitivity of FDM/RRT

In previous sections, we have been focusing on how to extract maximal information from the available data. In this section, we will exploit another important aspect of signal processing—-checking the reliability of the results, which has been ignored in previous discussions. Unfortunately, this is no easy task due to the high nonlinearity and complexity of FDM and RRT algorithms. What makes the analysis even more complicated are various practical factors including noise, lineshape distortions and other NMR experimental artifacts[16], which are very difficult to model and quantify. As a result, we will only be able to check the reliability of FDM/RRT results in various qualitative and/or semi-quantitative ways. Instead of providing some rigorous, conclusive statements, we will focus on how to double check the results and make some qualitative judgments. A semi-quantitative way for studying the sensitivity of FDM/RRT spectral estimations with respect to various processing parameters will also be described.

### 2.6.1 Reliability of 1D FDM and RRT

Reliability of 1D FDM and RRT is relatively easy to check. As we known, both FDM and RRT can be used as a high-resolution spectral estimator. One of the simplest ways to check the validity of the result is to inverse DFT the FDM/RRT spectrum to recover the time domain signal and compare it with original one. Alternatively, when FDM is used as a parametric method, one can plug the parameter

[16]Such as $t_1$ noise, pulsing artifacts and phase-cycling/gradient residual artifacts

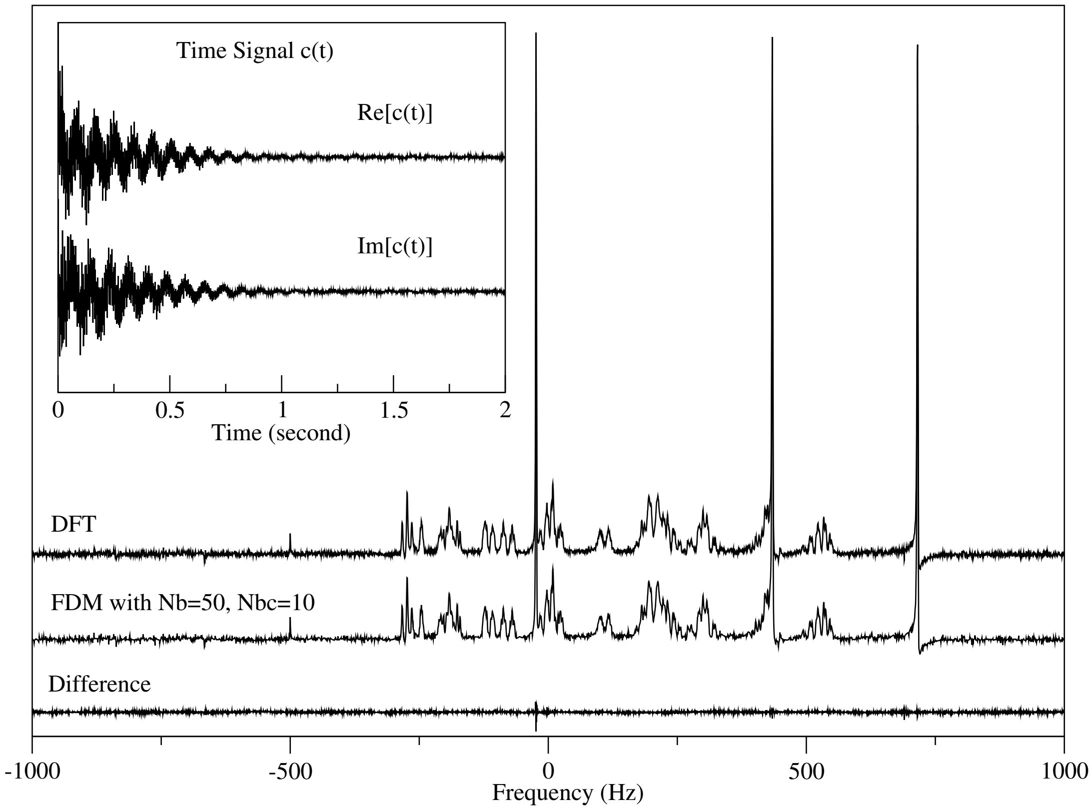



Figure 2.26: The spectral estimation obtained by FDM matches almost perfectly with the DFT spectrum. The difference is within the noise level except for one of the three most intensive peaks. The parameters used in FDM calculation were set to typical values and not optimized.

list $\{\omega_k, d_k\}$ into Eq. 2.1 to reconstruct the time signal and compare it with the original one. In the examples given in Section 2.2 and 2.5.1, we already compared the FDM/RRT spectra with DFT spectra and showed that they matched very well. Here we present a reliability study where FDM is used both as a parametric method and as a spectral estimator to analyze a noisy experimental NMR signal. Figure 2.26 compares the FT and FDM spectra, with the inner panel showing the real and imaginary part of the time domain signal. Even though the signal is very noisy and contains many heavily overlapped features, the difference between DFT and FDM spectra is

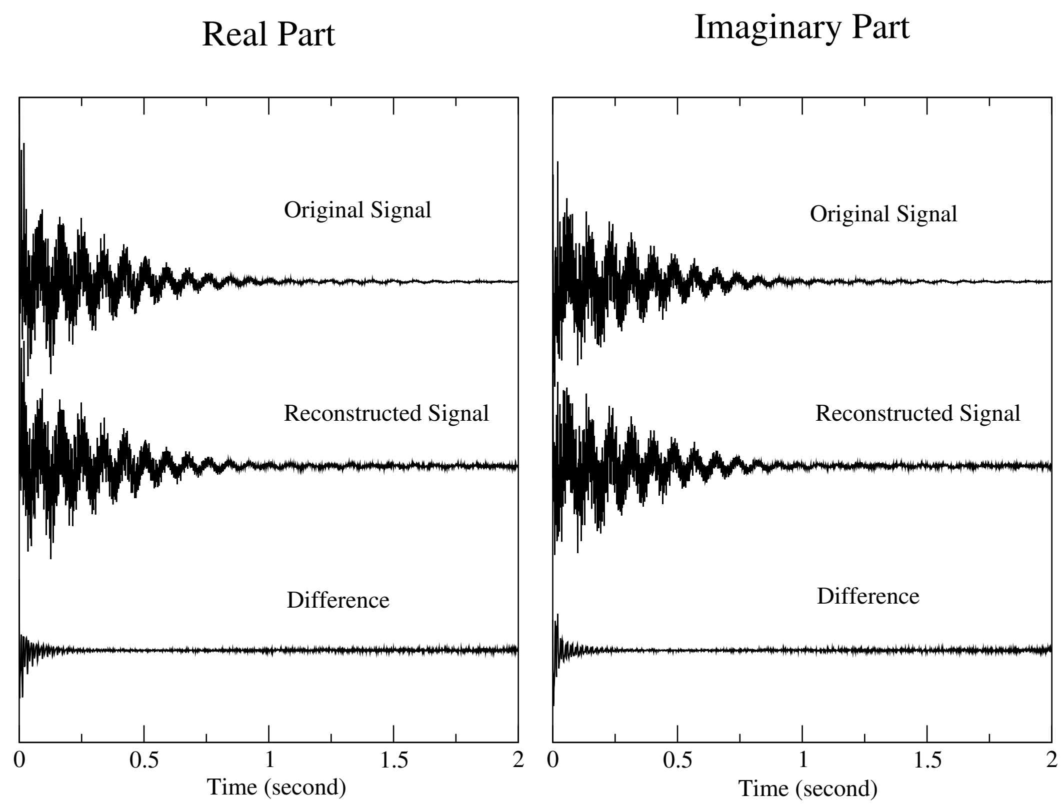



Figure 2.27: Comparison of the original signal and that reconstructed from the spectral parameters computed by FDM. Due to the use of *multi-scale* basis and multi-windowing, only an approximate complete line list that primarily describes narrow features can be obtained (see text), and used here to reconstruct the time signal. The initial part of the signal is not well reproduced. However, the rest of this signal, from $t = 0.2s$ up to $t = 2s$ (maximum), mainly defined by narrow poles, is very well reproduced. The difference between original and reconstructed signals is within the noise level.

within the noise level throughout the spectral range except a region around one of the three most intensive peaks ($\sim$-50 Hz). Similar comparison can be made between RRT and DFT spectra (data not shown). FDM/RRT have also been tested using many other 1D NMR experimental signals and proved to be very stable and reliable as high-resolution spectral estimators.

It is less straightforward to check the spectral parameters computed by FDM for experimental signals. One of the reasons is that the "true" spectral parameters are not available for direct comparison. In addition, the spectral parameters are the most accurate when a *multi-scale* Fourier basis is used, where the background is recomputed for each window calculations. It is not obvious how to combine results from all window calculations to construct an effective line list for describing the background (very broad poles). Furthermore, the adjacent spectral windows overlap with each other by 50%. The spectral parameters for the same overlapped region from two adjacent window calculations are similar but slightly different, especially for those that describe noise. Instead of trying to match two lists to remove the redundancy, which could be very tricky and unreliable, a simplified approach is used here: for each window calculations, only the parameters for describing the poles that locate inside the center half of the spectral window is retained. By concatenating parameters from all window, we can obtain an approximate "complete" line list for the whole spectral range. Figure 2.27 compares the signal reconstructed using such a line list to the original one. It is not surprising that the initial part of the signal, which is mostly defined by the broad features, is not reproduced well, due to the lack of entries for the broad features. The rest of the signal is reproduced very well. The difference between the original and reconstructed signals is about the noise level. This proves that most, if not all, the narrow features are parametrized quite accurately. However, this observation should not be extended to the conclusion that *all* the poles computed by FDM are genuine. It is highly possible that multiple Lorentzian lines are responsible for describing a single genuine peak, especially in the case of noisy signals with heavily overlapped features and/or lineshape distortions.

### 2.6.2 Possible Ways for Checking Individual Peaks

Figure 2.27 is an example where we check the overall validity of the parametric fit provided by FDM. In many cases, we are also interested in knowing whether a particular entry in the FDM line list is genuine or not, and if yes, what the uncertainty (error bar) is. These are very difficult problems where the ultimate solutions are still to be found. However, there are a few qualitative and semiquantitative methods for estimating the error bar and reliability of individuals peak computed by FDM.

Method 1: There existed a method for estimating the error of each eigenvalue since FDM was first introduced [24, 67], which was simply based on

$$\epsilon(n) = ||(U^{(n)} - u_k^n U^{(0)})B_k||, \quad n \geq 2\,,$$

where $U^{(n)}$ is the matrix representation of n-step evolution operator $\hat{\mathbf{U}}^n = \exp(-in\tau\hat{\Omega})$. Typically n=2 is used. The spurious poles usually have a much larger estimated error and this had been used to filter out spurious poles in the early days of FDM. It was later discovered that it should only be used as a qualitative estimation, instead of a rigorous criteria. Broader features, genuine as well as spurious, tend to have larger estimated errors, and some narrow spurious poles might have small estimated errors by accident. However, it is relatively safe to treat those narrow poles with large estimated errors as spurious poles.

Method 2: This method relies on estimating the "velocity" of eigenvalues numerically by small random noise perturbation. It is assumed that eigenvalues corresponding to genuine poles tend to be less sensitive to small perturbations while the spurious eigenvalues behave oppositely. Numerically, one can first run a FDM calculation using the original signal, then add small pseudo-random noise perturbation ($\sim 1\%$) to the signal and re-run the FDM calculation. Then one needs to match up two sets of eigenvalues and estimate the velocity of eigenvalues numerically by $v(\omega_k) \sim \delta\omega_k/|noise|$. The eigenvalue matching step might be tricky, but numerically experiments showed that it could usually be achieved as long as the noise perturbation was sufficiently small. The velocity estimation is typically consistent with the error estimation described in previous paragraph. This estimation is also only qualitatively correct. There is no guarantee that *all* spurious peaks can be detected, nor that *all* eigenvalues with large velocities are spurious.

Method 3: This method involves injecting a narrow Lorentzian pole close to the peak(s) of interest, rerunning the FDM calculation and comparing the results. This is also based on the assumption that genuine peaks are more stable than the spurious ones. If the peak is genuine, it is unlikely that injecting a small Lorentzian pole close-by will have a large effect on it. However, if the peak corresponds to some poorly converged features or noise, any change of the close-by environment might significantly change the local results. A systematic scheme for carrying out this "probe-and-test" procedure requires more studies on what kind of probe pole (e.g., peak intensity and linewidth) should be used and where to inject it.

Method 4: Finally, one can always study the convergence of a particular peak as a function of FDM/RRT processing parameters such as basis size, basis density, signal length used, and regularization level (for RRT). If a feature is stable or becomes stable with respect to the change of these parameters, we can have more confidence saying that it is a genuine peak. This procedure is more time consuming and might seem unnecessary in 1D case. However, in 2D or higher dimensional applications, it is almost always necessary to run the FDM/RRT calculations multiple times, particularly using various signal sizes, before one can be confident that the results have successfully converged.

### 2.6.3 A Semi-Quantitative Way for Estimating the Sensitivity of FDM/RRT Spectral Estimations

Here we describe a semi-quantitative way for estimating the sensitivity of FDM/RRT spectral estimations. The idea is very simple. The spectral estimation $I(\omega)$ computed by FDM/RRT has a periodic dependence on the spectral window position. If one computes the spectral intensity at a particular frequency $\omega_0$ multiple times using different spectral window conditions, the results will differ, even though only by small amount. The degree of variance provides us a semi-quantitative way of measuring the sensitivity (or uncertainty) of FDM/RRT spectral estimation. For example, Figure 2.28 illustrates this uncertainty using a special setup called *Continuous Windowing* FDM/RRT, or CW-FDM/RRT. In CW-FDM/RRT, for each frequency point $\omega$, a spectral window centered at $\omega$ is first set up and the $\mathbf{U}$ matrices constructed. Then these matrices are either diagonalized or used in RRT to compute the spectral intensity *only* at $\omega$. Same procedure is repeated for each frequency grid. Obviously this is more time consuming than the normal FDM/RRT algorithm where the same $\mathbf{U}$ matrices are used to compute spectral estimation for *all* frequency points inside the spectral window. However, this setup reveals the uncertainty of FDM/RRT spectral

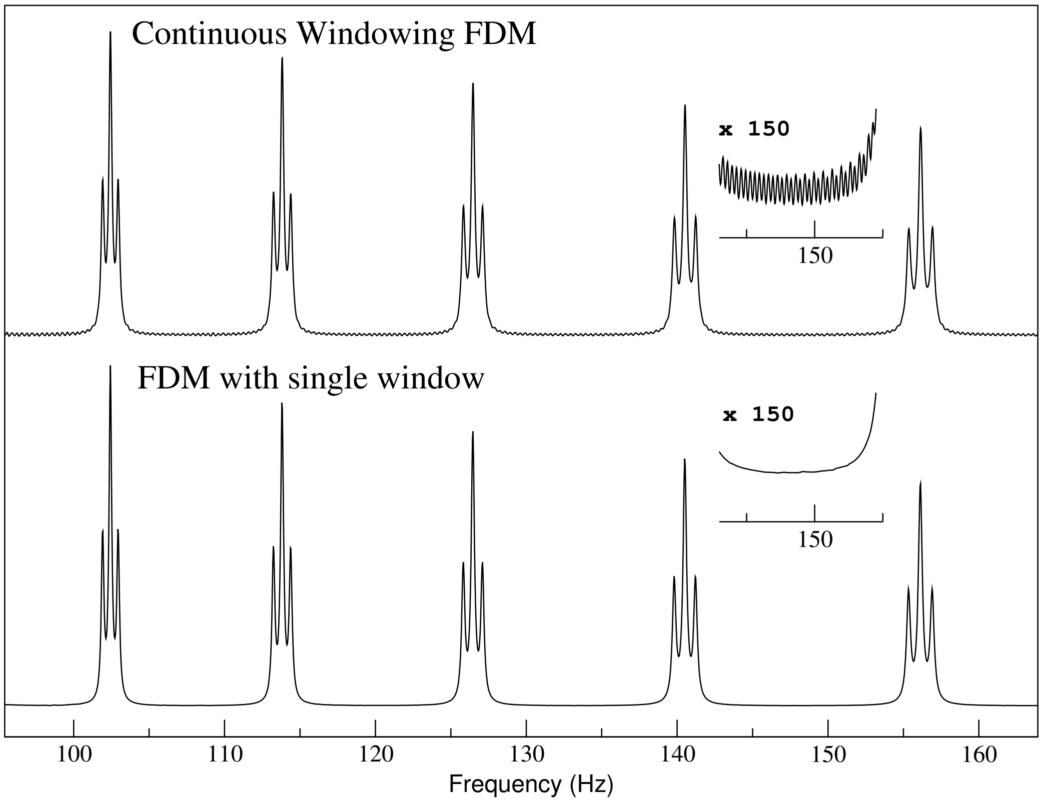



Figure 2.28: Sensitivity of FDM/RRT calculations illustrated by using the *continuous windowing* FDM. While the baseline of normal FDM calculation is perfectly flat, it is highly oscillative in CW-FDM spectrum. The amplitude of the oscillation provides useful information about the sensitivity (or uncertainty) of FDM/RRT spectral estimations.

estimations that is hidden in the normal FDM/RRT calculations. For example, while the baseline of the bottom trace of Figure 2.28 is flat, highly oscillative behaviour is observed in the top trace, computed by CW-FDM (CW-RRT spectrum is similar and not shown). The period of the oscillation is purely determined bye the signal length, but its amplitude depends on all the FDM/RRT parameters such as the basis density, basis size and regularization level (for RRT). It is possible to use the amplitude of this oscillation as a semi-quantitative measurement of the uncertainty of FDM/RRT spectral estimations. One can simply use the amplitude of the oscillation as the error

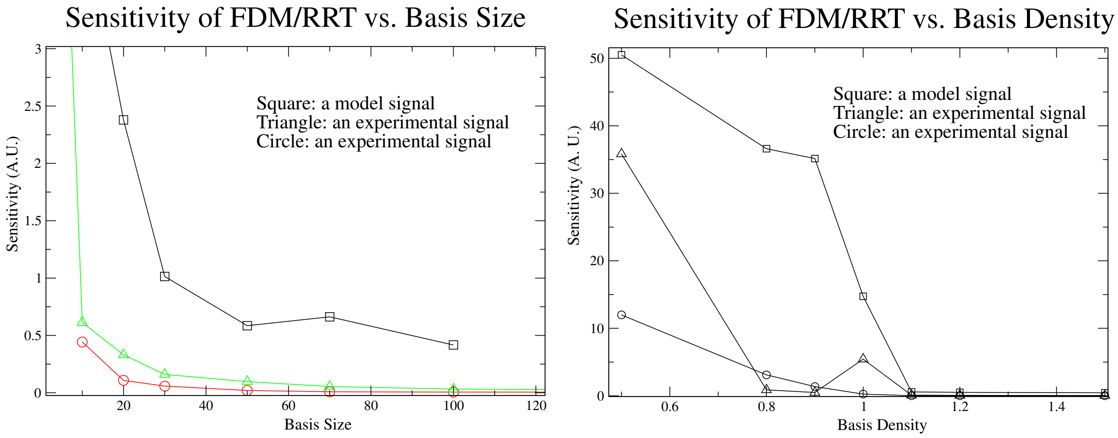



Figure 2.29: Sensitivity of FDM/RRT spectral estimation as functions of basis size and basis density using two experimental NMR signal and a model signal with integerization noise.

bar for the spectral estimation. Another, more interesting application is to use it to study the sensitivity as a function of all FDM/RRT parameters and find out the optimal processing parameters for typical signals. For example, Figure 2.29 plots the sensitivity of FDM/RRT spectral estimations as functions of the basis size and basis density using two experimental NMR signals and a model signal with integerization noise. We can see that basis size of $Nb = 50$ and basis density of $\rho = 1.1$ are the most efficient values for obtaining stable FDM/RRT spectral estimations for these three signals. Previous numerical experiments also suggest that these are indeed the most efficient optimal conditions for processing typical 1D signals. It is very interesting that such a semi-quantitative method is able to re-predict the optimal processing parameters in a straightforward, explicit way.

### 2.6.4 Reliability of 2D FDM

When 2D FDM and RRT are used as spectral estimators, one can also use inverse FT to recover the time domain signal and compare to the original one. However, this method is less convenient in 2D or higher dimensional cases due to the following reasons: first, to compare the signals in the time domain, spectra over the whole spectral range must be computed, while we are typically only interested in a few local regions; second, multi-dimensional NMR signals typically contain a water signal, which is often much more significant than all the other signals. Thus it is suboptimal to include the water signal in checking the fitting quality of the small features of interest. Instead, it is more efficient and more reliable to compare the results with FT of the original signal locally in the frequency domain.

Infinite 2D FT spectrum is typically estimated by 2D FDM. In order to make direct comparison, we need to estimate the finite 2D FT spectrum using the same spectral parameters. This can be done by replacing the double-absorption expression of Eq. 2.68 with the following expression, obtained the analytically evaluating the finite 2D FT summation[17]

$$\begin{aligned} A^{N_1,N_2}(\omega_1,\omega_2) \quad \approx \quad & \tau_1\tau_2 \sum_{k,k'} \mathrm{Re}[D_{k,k'}] \\ & \mathrm{Re}\left[\frac{1-e^{iN_1\tau_1(\omega_1-\omega_{1k})}}{1-e^{i\tau_1(\omega_1-\omega_{1k})}} - \frac{1}{2}\right] \mathrm{Re}\left[\frac{1-e^{iN_2\tau_2(\omega_2-\omega_{2k'})}}{1-e^{i\tau_2(\omega_2-\omega_{2k'})}} - \frac{1}{2}\right]. \end{aligned} \tag{2.110}$$

By computing the difference of the FDM estimation of 2D DFT with the actual 2D DFT of original signal, one can have a rough measurement of the fitting quality.

[17]Note that FT with certain apodizations such as exponential and trigonometric functions can be evaluated analytically and incorporated.

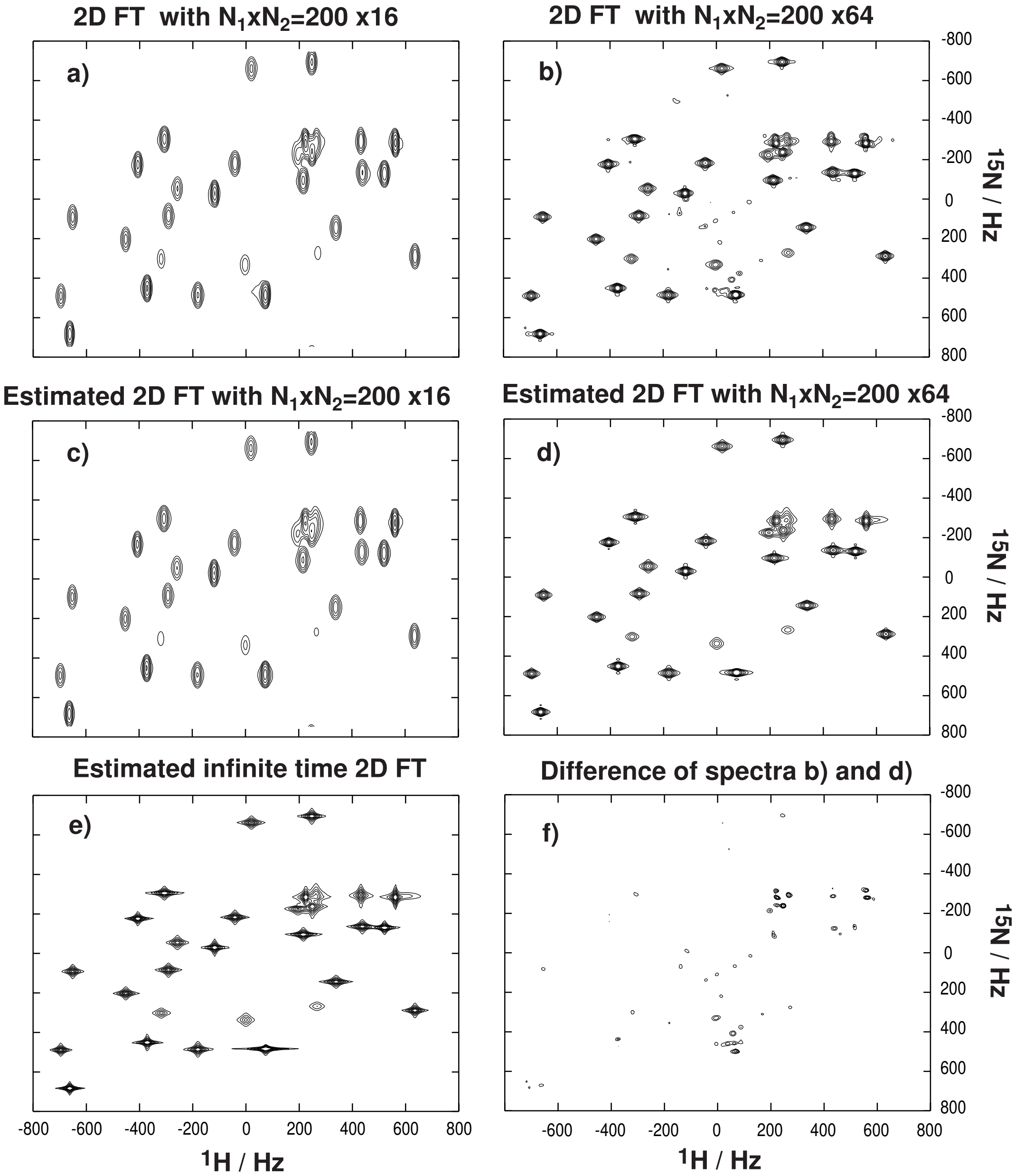



Figure 2.30: An example where the spectral parameters obtained by FDM are used to estimate finite 2D FT spectra for direct comparison with the actual 2D FT spectra. The signal used is the same $^{1}$H-$^{15}$N HSQC signal of rubredoxin used in previous 2D examples. The FDM spectral parameters were obtained using 200×16 N-type complex data points. The actual 2D FT spectra (panel a and b) were obtained using both N- and P-type data sets with cosine weighting functions in both dimensions. It shows that the estimated finite 2D FT spectra match with the actual 2D FT spectra very well. The mismatch between spectra (a) and (c) is only about 1.9% (see text). Even when the same parameters obtained using only 200×16 data points were used to estimate the finite 2D FT with length of $N_1 \times N_2 = 200 \times 64$, the mismatch between spectra (b) and (d) is only 2.9%. Note that the noise that is only present in long 2D FT spectra (panel b) can be accounted for a large percentage of the larger mismatch.

Figure 2.30 is a demonstration of such a checking procedure using the same 2D HSQC NMR experimental signal of rubredoxin used in previous 2D examples. Actual 2D FT spectra were computed by using both N- and P-type data set with cosine weighting functions in both dimensions. Both a short 2D FT spectrum (panel a, $200 \times 16$) and a long one (panel b, $200 \times 64$) are shown. Note that the long 2D FT spectrum is noisier. Spectral parameters $\{\omega_{1k}, \omega_{2k'}, D_{k,k'}\}$ were computed by FDM2K only using $200 \times 16$ complex data points. These parameters were then used to estimated two finite time 2D FT spectra with signal length $N_1 \times N_2$ being $200 \times 16$ (panel c) and $200 \times 64$ (panel d) respectively[18], and the infinite time 2D FT spectrum (panel e). One can simply stare at these spectra (such as panel a and c) and see that they match very well. Quantitatively, one can compute a relative "matching error" based on,

$$\epsilon = \frac{\sum |A^{N_1,N_2}(\omega_1, \omega_2) - FT^{N_1,N_2}(\omega_1, \omega_2)|^2}{\sum |FT^{N_1,N_2}(\omega_1, \omega_2)|^2}, \tag{2.111}$$

where $FT^{N_1,N_2}(\omega_1, \omega_2)$ denotes the actual 2D FT spectrum. The as calculated matching error between spectra (a) and (c) is only 1.9%, which is within the noise level. This is might not be surprising as we did use $200 \times 16$ data points to compute these parameters in FDM and a good match is expected. We can also use the same spectral parameters to estimate a long finite time 2D FT spectra with $N_1 \times N_2 = 200 \times 64$, four times longer than the signal size used in FDM calculation. Again, the estimated 2D FT spectrum (panel d) matches with the actual 2D FT (panel b) very well, with the matching error being only slightly larger, 2.9%. Note that a significant portion of the increased mismatch can be due to the noise that is only present in the long 2D

[18] Note that cosine weighting function is evaluated analytically, resulting a formula that is slightly different from Eq. 2.110.

FT spectrum. This provides a very strong evidence that FDM is capable of providing a good, reliable fit of the time signal to the Lorentzian model.

When the signal does not satisfy the Lorentzian model as well, the fitting error will be larger. More examples will be given in Chapter 3, where FDM2K is applied to analyzing 2D and 3D protein NMR signals. It has been shown that FDM is capable of providing reliable fit of both large and complex multidimensional NMR signals in most cases. The reliability and accuracy of individual peaks can be also be estimated, using the techniques listed in previous section. In practice, these approximate error bars do not always provide much information about the authenticity of the peak. Instead, it is more useful and necessary to study the behaviour of particular peak(s) as a function of processing parameters, especially the signal length used. More discussions can be found in Chapter 3.

## 2.7 Summary and Remarks

Two nonlinear methods, namely the Filter Diagonalization Method (FDM) and Regularized Resolvent Transform (RRT), for high-resolution spectral analysis of time domain signals have been presented.

FDM solves the highly nonlinear fitting problem of Harmonic Inversion Problem (HIP) by pure linear algebra of diagonalizing some small data matrices in the frequency domain. FDM is intrinsically stable and efficient due to its two most important properties: solving the nonlinear fitting problem via linear algebra and local spectral analysis via Fourier-type basis. Applied to both model and experimental signals, it has been shown that FDM is able to deliver resolutions beyond the FT time-frequency uncertainty principle, given that the signal sufficiently satisfies the Lorentzian model and is long enough (but typically not long enough for FT to fully resolve all the features). The actual power of FDM lies in its multi-dimensional extensions (MD-FDM). MD-FDM is a true multi-dimensional spectral analysis method, which is able to process the whole data set to characterize the multi-dimensional features. The achievable resolutions in all dimensions are determined together by the total information content of the signal, instead of the signal sizes along individual dimensions. MD-FDM is ill-conditioned due to several fundamental and practical limitations. Regularization techniques must be used in order to obtain stable and meaningful results with single FDM calculations. Among all the regularization methods exploited, FDM2K and SVD are particularly efficient. However, the ultimate optimal regularization is yet to be found. In conclusion, 1D FDM is basically a well-developed method, that is stable, efficient and capable of provide resolutions higher than FT spectral analysis. Stable, high-resolution multi-dimensional spectral estimations can be obtained efficiently by MD-FDM with regularizations, but the problems associated with processing multi-dimensional signals have only been partially solved. Even though the results we have obtained are already superior compared to those given by FT or other high-resolution methods such as LP and MaxEnt, there is still plenty of room for further improvements.

RRT emerges naturally from FDM. The same data matrices are used but spectral estimations are directly computed without the diagonalization step. Implementation of RRT is straightforward and numerically efficient. Conventional techniques such as Tikhonov regularization or SVD can be efficiently use to regularize RRT. RRT can be also use to construct non-FT types of spectral representations such as inverse Laplace transform and various pseudo-absorption spectra. The advantage of RRT lies in its simplicity and transparency, and the drawback is the lack of the freedom to directly manipulate individual peaks. Both 1D and MD RRT are well-developed methods that are stable, reliable and efficient. The problems associated with RRT are mostly understood and significant improvements over current version of RRT are not much expected. The major limitation of RRT lies in the phase instability of the peaks, manifesting it as the lineshape distortions in the final absorption spectra, especially when the signal is truncated in all dimensions. While the hybrid method eXtended Fourier Transform (XFT) intends to combine the stability of FT with high-resolving power of RRT, the results are not totally satisfactory. In the high-resolution XFT spectra where the contribution from RRT correction is significant, the phase instability is still present.

Finally, a limitation shared by RRT and FDM is the lack of a consistent way for verifying the results. While it is possible to qualitatively check whether a particular feature is likely to be genuine or not, there is not a rigorous, quantitative way to make some solid, conclusive judgments. False positives as well as false negatives can be costly for those unaware of the underlying uncertainty. One of the main focuses for future research should be developing some systematic schemes for detecting false positives and false negatives. Before such a scheme becomes available, the question of "I see something great, but is it real?" will always be haunting us and hamper FDM/RRT from being widely accepted and used.

# Appendix I: Non-Hermitian Quantum Mechanics

Here we briefly review the linear algebra of Non-Hermitian Quantum Mechanics [91]. The conventional quantum mechanics considers only Hermitian operators acting in a Hilbert space. The Hilbert space is characterized by a Hermitian inner product $\langle\Psi|\Phi\rangle = \langle\Phi|\Psi\rangle^*$ with the asterisk defining the complex conjugate. Any vector in Hilbert space has is a well defined the norm, $\langle\Psi|\Psi\rangle$, which is always real and non-negative. The eigenvalues of a Hermitian operator are always real. However, Hermitian quantum mechanics is not sufficient to describe dissipative dynamic systems, as we need non-Hermitian operators with complex eigenvalues representing the frequencies of decaying sinusoids.

Consider an abstract non-Hilbert linear vector space $A$. To distinguish between the Hermitian and complex symmetric inner product for the latter we use the round brackets, $(\Psi|\Phi) = (\Phi|\Psi)$. Note that $(\Psi|\Psi)$ is not necessarily real, i.e., the norm of $\Psi$ is not necessarily defined in our non-Hilbert space. Moreover, $(\Psi|\Psi)$ can even vanish for a nonzero vector $\Psi$. Although numerically this is unlikely to happen, it is clearly an indication of possible problems (e.g., instability) in the numerical algorithms involving the non-Hermitian inner products.

We will always identify linear operators, $\hat{U}$, $\hat{\Omega}$, etc., that act on vectors in $A$ by a cap. By the complex symmetric operator $\hat{\Omega}$ we mean that it satisfies the following relationships:

$$\{(\Psi|\}\{\hat{\Omega}|\Phi)\} = \{(\Psi|\hat{\Omega}\}\{|\Phi)\} = (\Psi|\hat{\Omega}|\Phi) \ , \tag{2.112}$$

for any two vectors $|\Psi)$ and $|\Phi)$ from $A$. In words, it does not matter whether we first operate with $\hat{\Omega}$ on $|\Phi)$ and then evaluate the inner product with $(\Psi|$ or vice versa.

An operator $\hat{\Omega}$ is diagonalizable, if it has a set of eigenvalues $\omega_k$ and eigenvectors $|\omega_k)$ satisfying

$$\hat{\Omega}|\omega_k) = \omega_k|\omega_k) \ , \tag{2.113}$$

where the eigenvectors are orthonormalized with respect to the complex symmetric inner product, i.e.,

$$(\omega_k|\omega_{k'}) = \delta_{kk'} \ . \tag{2.114}$$

We will also assume implicitly that our operators are not pathological, which, in particular, means that the eigenvectors form a complete basis and one can use the resolution of identity,

$$\hat{I} = \sum_k |\omega_k)(\omega_k| \ . \tag{2.115}$$

This also implies that $\hat{\Omega}$ can be expressed using the spectral representation,

$$\hat{\Omega} = \sum_k \omega_k|\omega_k)(\omega_k| \ . \tag{2.116}$$

The spectral representation becomes very useful when we want to obtain an expression for a function $f(\hat{\Omega})$ of an operator $\hat{\Omega}$, whose eigenvalues and eigenvectors are known:

$$f(\hat{\Omega}) = \sum_k f(\omega_k)|\omega_k)(\omega_k| \ . \tag{2.117}$$

Note that $f(\hat{\Omega})$ is also an operator with the eigenvalues $f_k = f(\omega_k)$ and the same eigenvectors $|\omega_k)$.

# Appendix II: An alternative expression for computing the amplitudes

Here we derive an alternative expression for computing the amplitudes, which is more accurate for narrow poles. As for normal eigenvalue solvers, the eigenvalues are obtained variationally, while the eigenvectors are obtained in a non-variational way. Thus according to Eq. 2.24 the amplitudes $d_k$ are generally much less accurate than the frequencies $\omega_k$. In addition, given an eigenvector $\tilde{\mathbf{B}}_k$, Eq. 2.24 only makes use of the first half ($M = N/2$) of the time signal available and is not always the most accurate expression for the coefficient $d_k$ that exists, especially for a narrow pole $\omega_k$. As such a more accurate formula is derived for the narrow poles. First let us rewrite Eq. 2.6 in a more general form:

$$\begin{aligned} d_k^{1/2} &\equiv (\Upsilon_k|\Phi_0) = \left(\left[\hat{U}/u_k\right]^{n'} \Upsilon_k|\Phi_0\right) \\ &= u_k^{-n'} \left(\Upsilon_k|\hat{U}^{n'}\Phi_0\right) = u_k^{-n'} \sum_j \left[\tilde{\mathbf{B}}_k\right]_j (\Psi_j|\Phi_{n'}) \\ &= e^{n'\gamma} \sum_j \left[\tilde{\mathbf{B}}_k\right]_j \sum_{n=0}^{M-1} e^{in'\tau(\omega_k+i\gamma)} e^{in\tau\varphi_j} c_{n+n'} \ , \end{aligned}$$

where $\gamma$ and $n'$ are free parameters. We can now average the above expression over $n' = 0, 1, \ldots, M_{\mathrm{aver}} - 1$ for an arbitrary $M_{\mathrm{aver}}$ between 1 and $M$,

$$\begin{aligned} d_k^{1/2} &= \frac{1 - e^{-\tau\gamma}}{1 - e^{-M_{\mathrm{aver}}\tau\gamma}} \sum_j \left[\tilde{\mathbf{B}}_k\right]_j \\ &\times \sum_{n'=0}^{M_{\mathrm{aver}}-1} \sum_{n=0}^{M-1} e^{in'\tau(\omega_k+i\gamma)} e^{in\tau\varphi_j} c_{n+n'} \ , \end{aligned} \tag{2.118}$$

where the averaging was done by weighting each term with $e^{-n'\gamma}$ (to eliminate the prefactor) and then normalizing the final result by $\left[\sum_{n'=0}^{M_{\mathrm{aver}}-1} e^{-n'\gamma}\right]^{-1}$. Just like in

Eq. 2.18 one of the two sums in Eq. 2.118 can be evaluated analytically (see Section 2.2.4). To eliminate the ambiguity in the choice of the free parameters in Eq. 2.118 for narrow poles $\omega_k$ (for which Eq. 2.118 is relevant), it suffices to stick with $M_{\text{aver}} = M$ and

$$\gamma = \begin{cases} -\mathrm{Im}\{\omega_k\} & , \ \mathrm{Im}\{\omega_k\} < 0 \\ 0 & , \ \mathrm{Im}\{\omega_k\} > 0 \end{cases} .$$

For model signal or real signals with very high SNR, Eq. 2.118 does yield the amplitudes with more accuracy. However, it will lead to an ambiguous result if $M\tau|\mathrm{Im}\{\omega_k\}| \gg 1$, in which case the signal corresponding to these "broad" peaks should decay away very quickly and thus including longer signal to calculate their amplitudes would only deteriorate the accuracy. In these cases, Eq. 2.24 should be used. A final word on whether Eq. 2.24 or Eq. 2.118 should be used is as following: unless the signal has very high SNR and one is really interested in computing the amplitudes of some narrow poles as accurately as possible, Eq. 2.24 should always be used.

# Appendix III: Other Fourier-Type Basis

The rectangular-window Fourier basis defined in Eq. 2.15 is very simple and easy to implement. However, experience with Fourier Transform spectral analysis might suggest that it might not the most efficient one, for example, considering the highly oscillative sinc-like structure of off-diagonal matrix elements. Here we discuss a more general definition of Fourier basis by including a filter function $g_n$,

$$|\Psi(\varphi_j)) = \sum_{n=0}^{M-1} g_n \, e^{in\tau\varphi_j} |\Phi_n) \ . \tag{2.119}$$

Mathematically, any non-vanishing filter functions $g_n$ are exactly identical in the full space. However, the point of using a Fourier-type localized basis is to reduce the problem to a small subspace. It is possible, but not obvious, that a good filter function might lead to more accurate and stable numerical approach.

In general, the efficient expressions we derived for computing the $\mathbf{U}$ matrix elements (Eq. 2.20 and Eq. 2.23) can not be used. Chen and Guo [62] have derived an alternative expression to compute the $\mathbf{U}$ matrix elements in a Fourier basis with a general filter. This expression requires $K_{rmwin}$ times more CPU time compared to the simple case of rectangular-window Fourier basis of Eq. 2.15. However, there are several other type of filter functions that can be implemented with similar computational efficiency of rectangular window function [67]. For example, a exponential filter function,

$$g_n = \exp(-\gamma n) \, , \tag{2.120}$$

corresponds to replacing the real frequency grid $\varphi_j$ by a complex one, $\varphi_j + i\gamma$, and expressions Eq. 2.20 and Eq. 2.23 can still be used.

Trigonometric type of filter functions can also be efficiently implemented. For example, a cosine filter function,

$$g_n = \cos\left(\frac{\pi}{2}\frac{n}{M+1}\right) , \tag{2.121}$$

is often used in Fourier Transform and leads to a good balance between the suppression-of sinc wiggles and efficient sampling of initial part of the signals. In this case, Eq. 2.119 can be rewritten as in terms of rectangular-window Fourier basis,

$$|\Psi(\varphi_j))_{\cos} = \frac{1}{2}(\,|\Psi(\varphi_j+\alpha)) + |\Psi(\varphi_j-\alpha))\,) , \tag{2.122}$$

where $\alpha = \pi/(2(M+1))$. Then the $\mathbf{U}$ matrix elements can be computed using the same expressions, Eq. 2.20 and Eq. 2.23.

To our disappointment, numerical experiments showed that introducing these window functions had not led to a significant numerical difference. This indicates that a rectangular-window Fourier basis is already very efficient in carrying out the local spectral analysis. The use of different filter functions only affect the initial structure of $\mathbf{U}$ matrices and have minor effect on the whole subspace formed by $K_{\text{win}}$ window basis functions.

# Chapter 3

# Application of FDM for NMR Spectral Analysis

## 3.1 Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the most powerful tools for probing the structure of the physical matters. It is one of most commonly used non-destructive analytical tools for studying the structure of chemical compounds; it is one of the only two methods for determining the 3D structure of biological macromolecules, and the only one that works in solution. In addition, NMR is a powerful tool for studying the dynamics and interactions between molecules, providing unique information for understanding how biomolecules such as proteins and DNA function in their natural environments.

NMR is based on the principle of nuclear spin angular momentum, an intrinsically quantum mechanical property that is analogous to electron spin angular momentum.

It was first detected in bulk matter by Bloch [104] and Purcell [105] in 1946, who then shared the Nobel Prize for Physics in 1952 for this discovery. In the molecule, the nuclei are surrounded by electrons, which act like a shield to the external magnetic field. Thus the actual strength of magnetic field experienced by the nucleus has dependence on the local electronic environment. The observed nuclear magnetic resonance frequencies will be slightly different for nuclei in different chemical environments. This frequency shift is thus called the "chemical shift". The chemical shift was first predicted by Dharmatti and then demonstrated by the famous three-line proton NMR spectrum of pure ethanol [106]. Even though the existence of chemical shift is a big disappointment for physicists [107], it provides a way for chemists to study the structure of chemical compounds. At that time, NMR spectroscopy was called continuous wave NMR (CW-NMR) as it relied on sweeping either the magnetic field or the radio-frequency to detect difference resonances. Due to its low sensitivity and long experimental time, NMR did not actually gain much popularity in organic chemistry, until Ernst and Anderson developed modern Fourier-Transform NMR (FT-NMR) [108] in 1966. Short bursts of RF field, or pulses, were applied perpendicular to the static magnetic field to excite all the resonances at once. A Free Induction Decay (FID) time domain signal was then acquired and Fourier Transformed to provide the frequency spectrum. FT-NMR speeded up the experiment by hundreds of times and improved the signal to noise ratio (SNR) enormously for the data obtained in the same amount of time. With the development of the Fast Fourier Transform (FFT) algorithm by Cooley and Tukey [6] and decreasing cost of computers in the late 1960s, FT-NMR grew explosively and almost completely replaced the CW-NMR.

Another milestone in the history of NMR spectroscopy was the introduction of two-dimensional NMR experiments by Ernst [109] in 1976, based on the idea proposed by Jeener [110] in 1971. By introducing a second dimension, not only could the potentially heavily overlapped resonances be dispersed into the second dimension, but useful information about the chemical and/or dynamical connections between spins could be obtained. With the aid of a quantum-mechanical representation known as the *density matrix formalism* [111, 112], which provides a complete description of the state and dynamics of the spin system, numerous novel 2D and higher dimensional NMR experiments were developed. In 1991, Ernst was warded the Nobel Prize for Chemistry for his pioneering work on modern NMR methodology. Nowadays, 1D and multidimensional NMR experiments are performed on a routine basis in many fields including chemistry, solid state physics, molecular biology, and medicine [113, 11].

The effectiveness of NMR hinges on the ability to obtain a resolved spectrum. While the sensitivity of NMR has been steadily improved by new advances in NMR probe and high-field magnet technologies over the last two decades, the fundamental resolution of NMR solely depends on the obtainable magnetic field strength and has only seen limited improvement [114]. The resolution that can actually be achieved is further limited due several severe drawbacks of the conventional FT spectral analysis. In particular, the FT resolution is limited by the so-called FT time-frequency uncertainty principle, which means a long time signal has to be acquired in order to obtain a high spectral resolution. What makes it worse is that the FT uncertainty principle applies to all dimensions independently in the multidimensional FT spectral analysis. Due to various practical limitations, multi-dimensional NMR signals are typically truncated or severely truncated in the indirect time dimensions, leading to poor FT resolution in the corresponding frequency dimensions. This is one of the major limitations of the multi-dimensional NMR experiments. As we discussed in Chapter 1, considerable efforts have been spent on developing various high-resolution alternatives to the FT spectral analysis. However, due the high complexity and large scale of the problem, these methods are typically both computationally expensive and numerically unstable. Accordingly, FT remains the method of choice in most cases for NMR spectral analysis.

The FDM and RRT algorithms introduced in Chapter 2 have proved to be superior in computational efficiency and numerical stability. It shows promise as a working method for high-resolution spectral analysis. Multi-dimensional FDM and RRT are true multidimensional methods. The whole data set are processed together to provide resolution enhancements in all dimensions, according to the total information content of the signal. FDM and RRT are thus particularly suitable for NMR spectral analysis. In this chapter, we will discuss various issues that are specific to analyzing NMR experimental signals. The rest of the chapter is organized as following: we will first discuss the linear phase correction via FDM/RRT for 1D NMR signals in Section 3.2, then in Section 3.3, we will discuss the construction of non-trivial projections and its application to a *singlet*-TOCSY experiment. In Section 3.4, an efficient scheme for processing multi-dimensional Constant-Time (CT) NMR signals is introduced and then applied to 2D CT-HSQC and 3D CT-HNCO NMR experiments of proteins. Finally, a summary is given in Section 3.5.

## 3.2 Application of FDM/RRT for 1D NMR

From the spectral analysis point of view, 1D NMR does not really have limitations, simply because long signals are typically available and high resolution FT spectra can be thus obtained. The resolution enhancement that can be achieved is marginal, even for methods like FDM or RRT. However, FDM and RRT can provide additional useful information besides the high-resolution spectra. For example, as discussed in Section 2.5.3, RRT can be used to generate a pseudo-2D spectrum $I(\omega + i\Gamma)$ as a function of both real and imaginary frequencies and separate potentially heavily overlapped resonances. FDM can provide a parameter list which is generally a reasonable fit of the time signal, as was demonstrated in Section 2.6.1. Even though multiple entries might be used to represent a single spectral feature and any nontrivial use of line list should be done with cautions, the line list provides a much better starting point for peak picking, peak integration and coupling constant measurement, while the conventional peak searching methods are sensitive to noise, baseline distortion, and closeby strong features. In this section, we will focus on correcting the linear phase by FDM and RRT, where the conventional FT spectral analysis has problems.

### 3.2.1 Linear Phase Correction

Problems with the first few data points of NMR signal can arise from many sources including the transient response of the audio filter and probe ring-down. Linear phase delay can be also caused by the stabilization delay following the magnetic field gradient pulses and some special pulse sequences such as 1-3-3-1 water suppression sequence [115]. In these cases, the signal $c(t)$ is only available from some time $t_0 > 0$ instead of $t_0 = 0$. In 1D NMR, all resonances are typically controlled to have the same phase at $t = 0$, giving rise to a nice absorption spectrum with all peaks "in phase". It is thus desirable to produce such a "in phase" spectrum even when the signal does not start from $t = 0$, requiring a time shifting or so called "linear phase correction" in FT,

$$I(\omega) = \int_{t_0}^{T} C(t)e^{i\omega t}dt = e^{i\omega t_0}\int_{0}^{T-t_0} c(t')e^{i\omega t'}dt' = e^{i\omega t_0}\,\mathrm{FT}[c(t')]\,, \qquad (3.1)$$

where the term $e^{i\omega t_0}$ corresponds to a phase correction that is linear with respect to the frequency, $\omega$. The FT linear phase correction works well only when the delay is short, so that the phase roll over the typical linewidth is small. A large linear phase correction can cause severe phase rolls in the baseline. For example, trace (c) of Figure 3.1 shows that, with the delay $t_0 = 17ms$, the baseline is severely distorted after the FT linear phase correction. The source of the baseline distortion is follows: FT corrects the linear phase point by point on the frequency axis, which is only accurate at the center of the peaks. In another word, the phase roll is due to the fact that FT is not able to recognize individual peaks and correct their phases accordingly. For example, considering a signal that contains an isolated resonance at zero frequency with a linewidth of 5 Hz, and assuming a delay of 10 ms, the phase corrections for spectral points at -5 Hz and +5 Hz differ by $\Delta\phi = 10 \times 0.010 \times 360 = 36^o$, while the phase correction should be the same as across the same peak (at zero frequency). FDM and RRT fit the signal to a sum of Lorentzian lines and thus can potentially

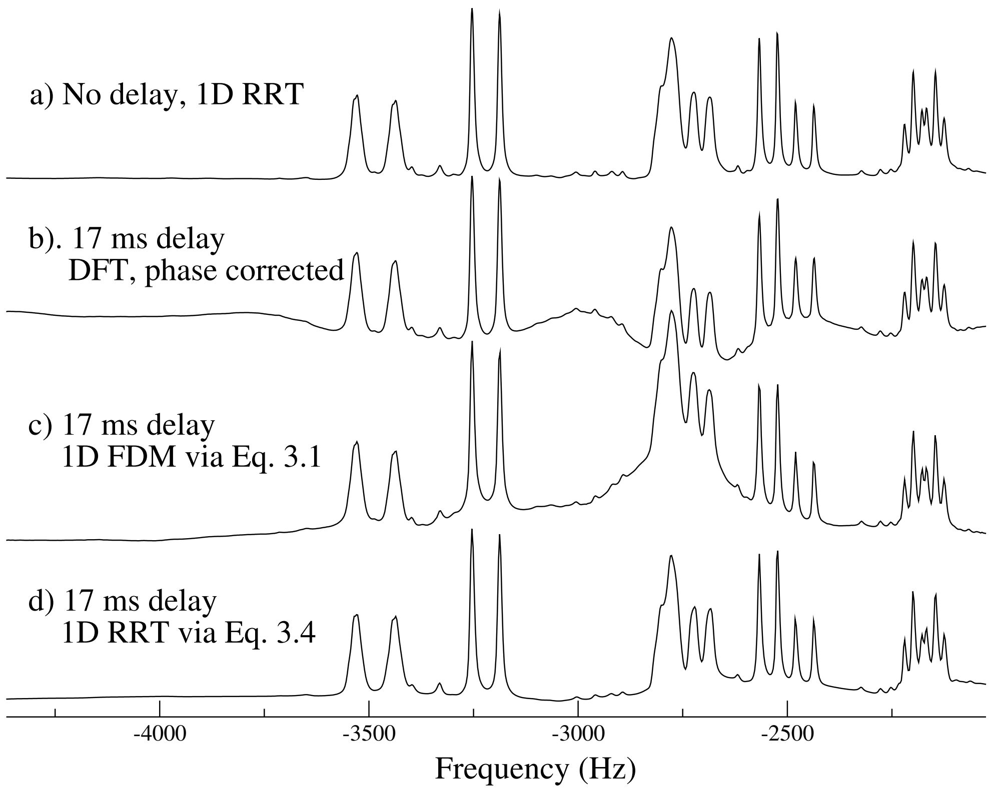



Figure 3.1: Linear phase correction via DFT and FDM/RRT, applied to a NMR experimental 1D signal with 5K complex data points. The FT linear correction introduces severe phase rolls in the baseline (trace b). The FDM linear phase correction via Eq. 3.2 is numerically unstable and can also introduce baseline distortions (trace c). By incorporating the linear phase correction into the RRT expression or the generalized eigenvalue problem, the numerical instabilities introduced by Eq. 3.2 can be eliminated (trace d). However, there are still some small instabilities due to the existence of broad poles and noise, which can be further suppressed by FDM/RRT averaging.

recognize all peaks and correct the phases of all peaks in a more consistent way.

In FDM is used, we have all the spectral parameters $\{\omega_k, d_k(t_0)\}$, where $d_k(t_0)$ denotes the complex amplitude at $t = t_0 > 0$. They can be then used to estimate the amplitudes at $t = 0$ simply by

$$d_k(0) = d_k(t_0)\exp(i\omega_k t_0)\,. \tag{3.2}$$

However, such a correction is only reliable if both the frequencies $\omega_k$ and the amplitudes $d_k(t_0)$ are accurately extracted. This is often true for the narrow peaks when the signal sufficiently satisfies the Lorentzian model (as we demonstrated in Section 2.6.1). However, for broad features, the parameters extracted are less accurate, and such a correction might introduce a large error and result in baseline distortions. In addition, multiple entries might often be responsible for describe a single spectral feature. The mutual cancellation and interference might not be conserved after the linear phase correction, resulting in distorted lineshapes. Finally, the line list contains entries that describe noise. Linear phase corrections for these entries do not make sense and might amplify the noise. Unfortunately, there is no reliable method for separating noise spikes from genuine features. As a result, Eq. 3.2 can introduce numerical instabilities. For example, trace (b) of Figure 3.1 shows an example where Eq. 3.2 is directly used to correct the linear phase caused by a delay of $17ms$ for an experimental signal that contains noise and broad features. Even though it seems that the correction is quite good for narrow peaks and there is no phase roll in the baseline, there is a big baseline distortion due to some broad poles around $\omega \sim -2750Hz$. An averaging method that is similar the stabilization scheme proposed by Mandelshtam [116] can be used to suppress the instabilities. The idea simply relies on the fact that the errors in the poles are random and thus can be reduced by averaging over many FDM calculations with varying signal length. A successful example was given in Ref. [67].

Here we present an alternative scheme where the linear phase correction is incorporated into the generalized eigenvalue problem and can significantly reduce the instabilities. RRT expressions can also be used to obtain the linear phase corrected spectral estimation directly from the signal in this scheme.

Given a discrete time signal $c(n) \equiv c(n\tau)$, available only from $n = n_0, n_0 + 1, \ldots, n_0 + N - 1$, the goal is to estimated the infinite time DFT summation[1],

$$I(\omega) = \tau \sum_{n=0}^{\infty} c(n) e^{i\omega n\tau} \ . \tag{3.3}$$

Substituting the FDM quantum ansatz, $c(n) = \left(\Phi_0 | \hat{U}^n \Phi_0\right)$, into Eq. 3.3 and evaluating the geometric summation analytically leads to[2],

$$\begin{aligned} I(\omega) &= \tau \left( \Phi_0 \left| \frac{1}{1 - e^{i\tau\omega}\hat{U}} \right| \Phi_0 \right) \\ &= \tau \left( \Phi_{n_0} \left| \hat{U}^{-n_0} \frac{1}{1 - e^{i\tau\omega}\,\hat{U}} \hat{U}^{-n_0} \right| \Phi_{n_0} \right) \\ &= \tau \left( \Phi_{n_0} \left| \frac{1}{\hat{U}^{2n_0} - e^{i\tau\omega}\,\hat{U}^{2n_0+1}} \right| \Phi_{n_0} \right) \ . \end{aligned} \tag{3.4}$$

Evaluated in the Fourier basis defined in Eq. 2.15, the operator expression of Eq. 3.4 becomes a numerical expression for estimating the linear phase corrected spectra directly using the delayed finite time signal:

$$I(\omega) = \tau\, \tilde{\mathbf{C}}_{n_0}^{\mathrm{T}} \left[ \tilde{\mathbf{U}}^{(2n_0)} - e^{i\tau\omega} \tilde{\mathbf{U}}^{(2n_0+1)} \right]^{-1} \tilde{\mathbf{C}}_{n_0} = \tau\, \tilde{\mathbf{C}}_{n_0}^{\mathrm{T}}\, \tilde{\mathbf{R}}^{(2n_0)}(\omega)^{-1}\, \tilde{\mathbf{C}}_{n_0} \,, \tag{3.5}$$

where the shifted FT column vector $\tilde{\mathbf{C}}_{n_0}$ is computed as,

$$[\tilde{\mathbf{C}}_{n_0}]_j = (\Phi_{n_0} | \Psi_j) = \sum_{n=0}^{M-1} e^{in\tau\varphi_j} c(n + n_0) \,,$$

and the delayed evolution matrices $\mathbf{U}^{(2n_0+l)}, l = 1, 2$ can be computed using Eq. 2.20 and Eq. 2.23 with $p = 2n_0$ and $p = 2n_0 + 1$ respectively. Note that $M = (N - n_0)/2$ due to the condition that $2(M - 1) + 2n_0 + 1 = N + n_0 - 1$.

[1] The zero point correction term is dropped here for the sake of simplicity.

[2] Assume that for dissipative systems, $\hat{U}^{\infty} \to 0$.

Obviously, Eq. 3.5 can be directly evaluated by computing a pseudo-inverse of the resolvent matrix via either SVD or Tikhonov Regularization. Alternatively, we can solve a generalized eigenvalue problem,

$$\tilde{\mathbf{U}}^{(2n_0+1)}\tilde{\mathbf{B}}_k = u_k \tilde{\mathbf{U}}^{(2n_0)}\tilde{\mathbf{B}}_k \,, \tag{3.6}$$

compute the amplitude (already linear phase corrected) as

$$d_k^{1/2}(0) = \tilde{\mathbf{C}}_{n_0}^{\mathrm{T}}\tilde{\mathbf{B}}_k \,, \tag{3.7}$$

and then use Eq. 2.27 to construct the linear phase corrected FDM ersatz spectrum. In principle, both approaches should give essentially the same result. However, as a large linear phase correction is ill-conditioned for signals with noise, broad poles and/or heavily overlapped features, the RRT approach is more suitable for obtaining a stable spectral estimation.

By incorporating the linear phase correction into the RRT expression or generalization eigenvalue problem, we totally avoid the possibility of amplifying the numerical errors in the spectral parameters by using a separate linear phase correction step of Eq. 3.2. Expression Eq. 3.5 is numerically much more stable and can potentially used to correct much larger time delays. For example, Figure 3.2 demonstrates the robustness of Eq. 3.5 using the model signal "Jacob's Ladder" that has been used in previous 1D examples. It shows that Eq. 3.5 can be used to apply the linear phase correction for virtually any length of delays without running into numerical instability problems. Applied to the experimental signal, Eq. 3.5 also shows significant improvement in baseline stability compared to both DFT and Eq. 3.2, demonstrated

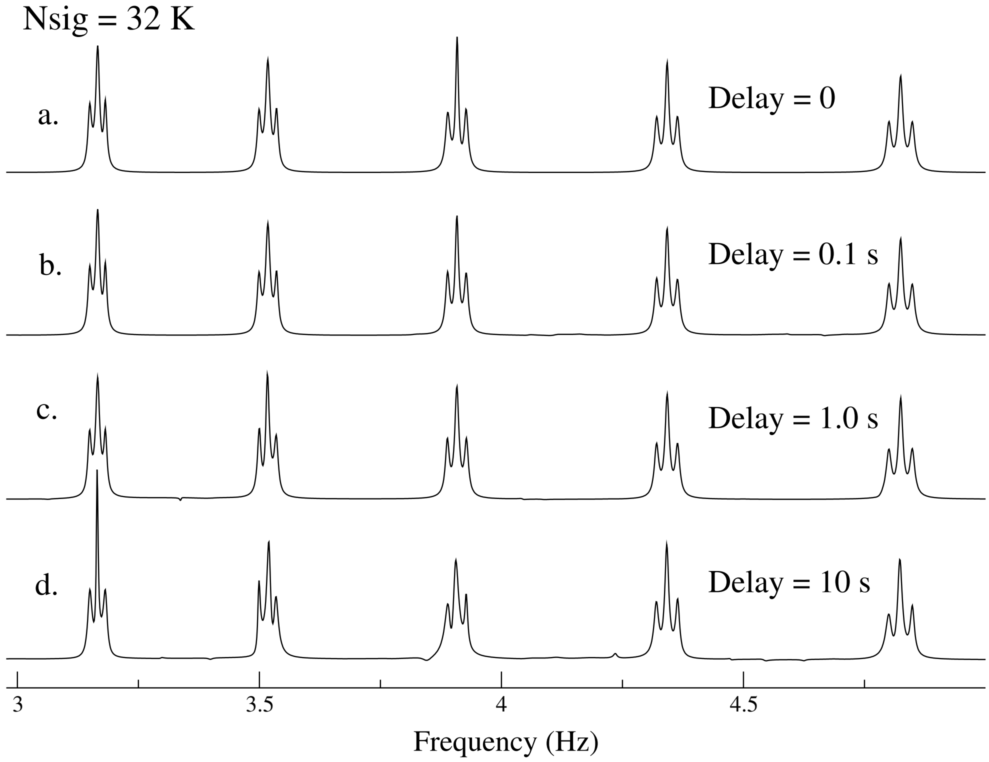



Figure 3.2: A demonstration of the robustness of linear phase correction via Eq. 3.5 using the model signal "Jacob's Ladder". Shown here is a dense region of the spectrum. Only 32 K data points were used and thus the multiplets were not totally converged, even though sufficiently resolved. Trace (b) through (d) show the spectra obtained by first skipping 100 up to 10000 data points before reading in 32K data points from the rest of the signal, corresponding to time delays of $t_0 = 0.1$ second up to $t_0 = 10$ seconds. Even with a delay of 10 seconds, a reasonably good spectrum can still be obtained. However, note that small artifacts start to appear in the baseline, and the phases of several peaks are not perfect.

in trace (d) of Figure 3.1. However, there is still some small distortions in the baseline. The reason is that even though we eliminate the numerical instability introduced by Eq. 3.2, instability due to the ill-conditioning of large linear phase corrections for noisy signals with broad peaks is still present. FDM averaging procedures can be used if one wish to further suppress the instabilities.

A last comment on linear phase correction via FDM and RRT is that the *multi-*

*scale* Fourier basis should not be used. Even though it can improve the accuracy of the parameters for the narrow features, this is achieved by using broad poles to describe the background and features outside the spectral window in lower resolution. These broad poles are often some effective representation of more complicated fine structure of actual features. Linear phase correction of these "unphysical" poles is meaningless and likely to introduce severe baseline distortions.

## 3.3 Nontrivial Projections by FDM

In addition to the trivial projections such as the conventional 1D projections (Eq. 2.67), FDM can be also used to construct some non-trivial projections. One of the most useful projections for NMR is the $45^o$ projection for $J$-experiment [72, 71, 73, 90]. It was demonstrated in that highly resolved absorption-mode $45^o$ projection can be obtained by FDM, while for FT construction of absorption-mode $45^o$ is impossible and only skew $45^o$ projection of an absolute value 2D-J spectrum can be obtained [118]. In this section, we will first outline the mathematical approach for constructing a general nontrivial projection along an arbitrary time vector, then focus on its application to obtain highly resolved homonuclear correction spectra using a few TOCSY data set. Numerical difficulties related to constructing stable nontrivial projections will also be addressed.

Let $\vec{p} = (p_1, \ldots, p_D)$ be a general N-dimensional time vector along which we want to construct a spectral projection. $p_l$ can be any real or even complex numbers (such as in 2D-J DOSY [96]). For example, in 2D case, $\vec{p} = (1, 0)$ and $(0, 1)$ correspond to the normal 1D projections along the first and second time dimensions; $\vec{p} = (1, -1)$ corresponds to the non-trivial $45^o$ projection. Given an N-dimensional line list $\{d_k, \vec{\omega}_k\}$, the $\vec{p}$-projection of the frequencies is simply defined as,

$$\omega_{\vec{p}k} = \vec{p}\,\vec{\omega}_k = \sum_{l=1}^{D} p_l\,\omega_{lk}\,. \tag{3.8}$$

The complex spectral $\vec{p}$-projection is then[3],

$$I_{\vec{p}}(\omega) = i\sum_k \frac{d_k}{\omega - \omega_{\vec{p}k}}\,. \tag{3.9}$$

An absorption-mode projection can then be obtained by taking the real part of the complex spectrum. Therefore, if we can obtain a parametric representation of the signal, it is trivial to compute any trivial or non-trivial projections. In practice, this procedure has difficulties due to the lack of a reliable N-dimensional FDM line list. Instead, it is less demanding to first construct the $\vec{p}$-projection Hamiltonian [73, 91],

$$\hat{\Omega}_{\vec{p}} = \vec{p}\,\vec{\Omega} = \sum_{l=1}^{D} p_l\,\hat{\Omega}_l \tag{3.10}$$

and then solve the eigenvalue problem,

$$\hat{\Omega}_{\vec{p}}\,|\Upsilon_{\vec{p}k}) = \omega_{\vec{p}k}\,|\Upsilon_{\vec{p}k})\,, \tag{3.11}$$

to obtain the $\vec{p}$-projection of the frequencies. The corresponding resolvent for estimating complex spectral $\vec{p}$-projection is then,

$$I_{\vec{p}}(\omega) = i\left(\Phi_0\left|\frac{1}{\omega - \hat{\Omega}_{\vec{p}k}}\right|\Phi_0\right) = i\sum_k \frac{(\Phi_0|\Upsilon_{\vec{p}k})^2}{\omega - \omega_{\vec{p}k}} = i\sum_k \frac{d_{\vec{p}k}}{\omega - \omega_{\vec{p}k}}\,. \tag{3.12}$$

[3]Note we use the integral Fourier spectral representation instead of the discrete one as the latter is unlikely to be advantageous for non-trivial projections

Note that for certain types of projections such as the $45^o$ projection of $J$-experiments, $\omega_{\vec{p}k}$ has virtually zero linewidth. Some smoothing should be used to improve the appearance of the spectra (see Eq. 2.28).

Evaluated in Fourier basis, the operator equations of Eqs. 3.10-3.12 become numerical expressions for computing general $\vec{p}$-projections. Given the matrix representations $\tilde{\mathbf{U}}_l$ for $l = 0, ..., \mathrm{D}$ in the Fourier basis, $\{|\varphi_j)\}$, a numerical procedure of computing a general $\vec{p}$-projection is given as following,

(i) Solve D number of 1D generalized eigenvalue problems (Eq. 2.62) independently to obtain the eigenvalues $u_{lk} \equiv e^{-i\tau_l \omega_{lk}}$ and eigenvectors $\tilde{\mathbf{B}}_{lk}$.

(ii) Use $\omega_{lk}$ and $\tilde{\mathbf{B}}_{lk}$ to construct a matrix representation of $\hat{\Omega}_{\vec{p}}$ in the basis of $\{|\varphi_j)\}$,

$$\tilde{\mathbf{\Omega}}_{\vec{p}} = \sum_{l=1}^{\mathrm{D}} p_l \, \tilde{\mathbf{\Omega}}_l = \sum_{l=1}^{\mathrm{D}} p_l \, \sum_k \omega_{lk} \tilde{\mathbf{U}}_0 \tilde{\mathbf{B}}_{lk} \tilde{\mathbf{B}}_{lk}^{\mathrm{T}} \tilde{\mathbf{U}}_0, \tag{3.13}$$

where $\tilde{\mathbf{\Omega}}_l$ is the corresponding matrix representation of $\hat{\Omega}_l$ in the Fourier basis.

(iii) Solve another generalized eigenvalue problem

$$\tilde{\mathbf{\Omega}}_{\vec{p}} \tilde{\mathbf{B}}_{\vec{p}k} = \omega_{\vec{p}k} \tilde{\mathbf{U}}_0 \tilde{\mathbf{B}}_{\vec{p}k} \ ; \quad \tilde{\mathbf{B}}_{\vec{p}k}^{\mathrm{T}} \tilde{\mathbf{U}}_0 \tilde{\mathbf{B}}_{\vec{p}k} = 1 \ . \tag{3.14}$$

(iv) Compute the frequencies $\omega_{\vec{p}k}$, and the amplitudes $d_{\vec{p}k}$ from the eigenvectors $\tilde{\mathbf{B}}_{\vec{p}k}$ (Eq. 2.56), then use Eq. 3.12 to compute the complex spectral projection $I_{\vec{p}}(\omega)$.

Initially, FDM averaging was used to obtain an artifacts free spectrum [73]. Here we implement the more efficient FDM2K regularization, simply by reformulating all the GEPs involved according to the FDM2K algorithm (Eq. 2.82). For non-zero $q^2$, the matrices on both sides of Eq. 2.82 are asymmetric. Thus both left and right eigenvectors need to be computed when diagonalizing $\tilde{\mathbf{U}}_l$, especially for large regularizations:

$$\tilde{\mathbf{U}}_0^\dagger \tilde{\mathbf{U}}_l \tilde{\mathbf{R}}_{lk} = u_{lk} \left( \tilde{\mathbf{U}}_0^\dagger \tilde{\mathbf{U}}_0 + q^2 \right) \tilde{\mathbf{R}}_{lk} \,, \tag{3.15}$$

$$\tilde{\mathbf{L}}_{lk}^{\mathrm{T}} \tilde{\mathbf{U}}_0^\dagger \tilde{\mathbf{U}}_l = u_{lk} \tilde{\mathbf{L}}_{lk}^{\mathrm{T}} \left( \tilde{\mathbf{U}}_0^\dagger \tilde{\mathbf{U}}_0 + q^2 \right) \,, \tag{3.16}$$

where $\tilde{\mathbf{R}}_{lk}$ and $\tilde{\mathbf{L}}_{lk}$ denote the right and left eigenvectors, normalized according to

$$\tilde{\mathbf{L}}_{lk}^{\mathrm{T}} \tilde{\mathbf{U}}_0 \tilde{\mathbf{R}}_{lk'} = \delta_{k,k'} \,.$$

Eq. 3.13 then becomes,

$$\tilde{\mathbf{\Omega}}_{\vec{p}} = \sum_{l=1}^{\mathrm{D}} p_l \sum_k \omega_{lk} \tilde{\mathbf{U}}_0 \tilde{\mathbf{R}}_{lk} \tilde{\mathbf{L}}_{lk}^{\mathrm{T}} \tilde{\mathbf{U}}_0 \,. \tag{3.17}$$

### 3.3.1 $45^o$-Projections of $J$-Experiments.

One of the most interesting non-trivial projects for NMR spectroscopy is the $45^o$ projections of various $J$-experiments. $J$-experiment was first proposed by Ernst and coworkers to obtain broad-band proton-decoupled proton NMR spectrum [118]. It is hampered by the poor resolution of phase-twisted line shape, because in $J$-experiments only a single purely phase-modulated data set can be acquired. The conventional FT spectral analysis requires many $J$ increments and can only provide absolute value mode $45^o$-projections, which are unimpressive in resolution. The $J$-spectroscopy has thus not been widely used. However, combined with FDM data processing, $J$-experiments can be very useful in simplifying complicated spectra with heavily overlapped multiplets, by first dispersing the multiplets into an additional $J$-dimension and then reducing them into sharp singlets in the $45^o$ projection.

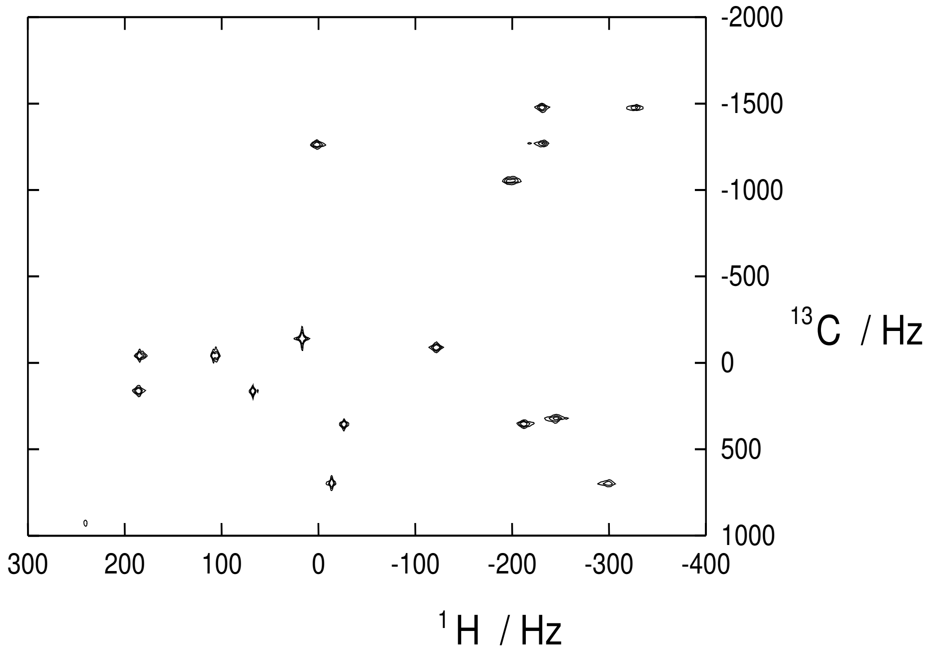



Figure 3.3: The *singlet*-HSQC spectrum of progesterone obtained using a HSQC-J pulse sequence given in Ref. [90]. All the proton multiplets are collapsed into singlets! The 3D purely phase modulated signal consists of $N_1 \times N_2 \times N_3 = 600 \times 64 \times 2$ points in the proton, carbon-13, and proton J dimensions respectively. FDM signal-length averaging was used to obtain an artifact free spectrum.

FDM has been successfully applied to various $J$-experiments including 2D-J [72, 71] and 3D HSQC-J [73, 90]. In 2D-J experiments, $\vec{p}$ is set to $(1, -1)$ to compute the $45^o$ projection. Highly resolved absorption mode homonuclear decoupled 1D proton NMR spectrum was obtained using as few as four $J$-increments [72]. In addition, for each singlet, an estimate of the corresponding multiplet can be computed. In 3D HSQC-J experiments, first two dimensions are proton and carbon chemical shifts and the third dimension is the proton homonuclear J dimension. $\vec{p} = (1, 0, -1)$ corresponds to the proton-decoupled $45^o$ projection. Correlated with the carbon chemical shifts, we can obtain a so called *singlet*-HSQC 2D double-absorption spectrum where the proton multiplets of the conventional HSQC spectrum are collapsed into singlets at the proton chemical shifts,

$$A(\omega_{\vec{p}}, \omega_2) = \sum_{k,k'} \mathrm{Re}[D_{k,k'}]\mathrm{Re}\left[\frac{1}{1 - e^{i\tau_2(\omega_2 - \omega_{2k})}} - \frac{1}{2}\right] \mathrm{Re}\left[\frac{i}{\omega_{\vec{p}} - \omega_{\vec{p}k'}}\right], \qquad (3.18)$$

with cross amplitudes $D_{k,k'}$ computed as Eq. 2.66. It was demonstrated that the $J$-dimensions did not have to be fully resolved in order to obtain a good $45^o$-projection. As few as one additional increment in each $J$-dimension might be sufficient in many cases. An example of *singlet*-HSQC spectrum is shown in Figure 3.3.

### 3.3.2 *Singlet*-TOCSY Spectra via FDM

Another interesting application of nontrivial FDM projections is the *singlet*-TOCSY experiment, where two $J$-dimensions, combined with FDM data processing, are used to decouple the homonuclear coupling in both proton dimensions. The closely spaced rectangular multiplet of conventional TOCSY is reduced to a sharp and intense singlet at the chemical shifts of two correlated protons, leading to a substantial resolution improvement. The simplified 2D spectra are thus called *singlet*-TOCSY spectra.

A J-TOCSY-J pulse sequence, developed by De Angelis [119], is given in Figure 3.4 in comparison with that of the conventional TOCSY experiment. The double spin echoes preceding each of the traditional time variable, $t_1$ and $t_2$, add two homonuclear $J$-dimensions, $t_3$ and $t_4$, leading to 4D time signals. Detailed descriptions of the NMR experimental parameters can be found in Refs. [90, 119]. Surprisingly, as few as one additional increment along each $J$-dimension is sufficient. Finally, only three TOCSY data sets, with $(t_3, t_4) = (0, 0)$, $(\tau_3, 0)$ and $(0, \tau_4)$, are required to obtain the 2D *singlet*-TOCSY spectrum.

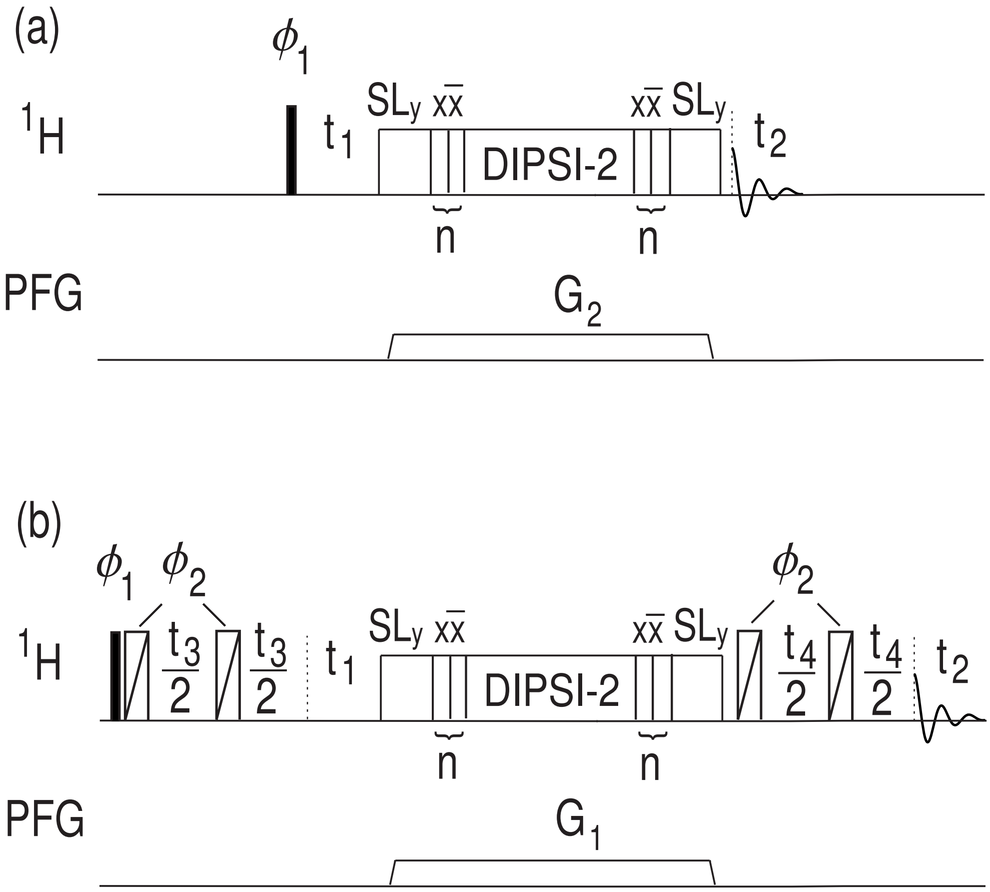



Figure 3.4: Timing diagrams of the conventional TOCSY, (a), and J-TOCSY-J experiments (b). The black solid icons represent $90^o$ pulses. Scored icons represent constant amplitude FM $180^o$ pulses. A spin lock and short Tr-ROESY purging pulse [120] surround the isotropic mixing sequence DIPSI-2 [121, 122]. The 4-step phase cycling $\phi_1 = (x, y, -x, -y); \phi_2 = 4(x)$; receiver= $(x, y, -x, -y)$ is combined with CYCLOPS [123] for a total of 16 scans per increment.

A 2D double $45^o$ projection of the 4D time signal can be constructed by setting $\vec{p}_1 = (1, 0, -1, 0)$ and $\vec{p}_2 = (0, 1, 0, -1)$ and computing the double-absorption mode 2D spectral estimation as,

$$A(\omega_{\vec{p}_1}, \omega_{\vec{p}_2}) = \sum_{k,k'} \mathrm{Re}[D_{k,k'}]\mathrm{Re}\left[\frac{i}{\omega_{\vec{p}_1} - \omega_{\vec{p}_1 k}}\right] \mathrm{Re}\left[\frac{i}{\omega_{\vec{p}_2} - \omega_{\vec{p}_2 k'}}\right]. \tag{3.19}$$

Note that 2D spectral estimation with a single homonuclear decoupled proton dimension can also be obtained simply by replacing either $\vec{p}_1$ or $\vec{p}_2$ with the conventional projection vectors. For example, setting $\vec{p}_1 = (1, 0, 0, 0)$ and $\vec{p}_2 = (0, 1, 0, -1)$ leads to a 2D spectrum homonuclear decoupled only in the second proton dimension.

Figure 3.5 demonstrates the *singlet*-TOCSY using a simple spin system, 2,3-dibromopropionic acid, with three weakly coupled protons. The pulse sequence shown in Figure 3.4 was used with a $40.0ms$ spin lock, $35.0ms$ Tr-ROESY purging sequence, and $53.0ms$ DIPSI-2 mixing time. The spectral widths are $25Hz$ for both $J$-dimensions, and $700Hz$ for both proton dimensions. For each $(t_3, t_4)$ increment, $200 \times 80$ complex data points were acquired. The regularizations were optimized for all FDM calculations. Even though one additional increment in each $J$-dimension is not sufficient to fully resolve the multiplets in $J$-dimension, clean and sharp singlets can be obtained in the $45^o$-projections. Figure 3.6 compares the corresponding vertical cross sections of the central part of the central triplet in the conventional TOCSY and *singlet*-TOCSY spectra. The multiplets are not fully resolved in the indirect dimension ($t_1$), shown in the top trace, while nice and intense singlets are obtained in the *singlet*-TOCSY, shown in the bottom trace.

The simple example of Figure 3.5 demonstrates the possibility of using FDM to obtain homonuclear decoupled TOCSY spectra using small J-TOCSY-J 4D data sets. However, its performance on more complex systems still needs further investigation. There are several difficulties associated with computing 2D *singlet*-TOCSY spectra. First, TOCSY spectra contain a lot of multiplets that are degenerate in both dimensions, are unfavorable for FDM. Second, the scaling step of Eq. 3.13 is sensitive to any numerical errors in the eigenvalues and eigenvectors. It is not clear how regularization will effect the accuracy of projection operators as constructed. Third, it

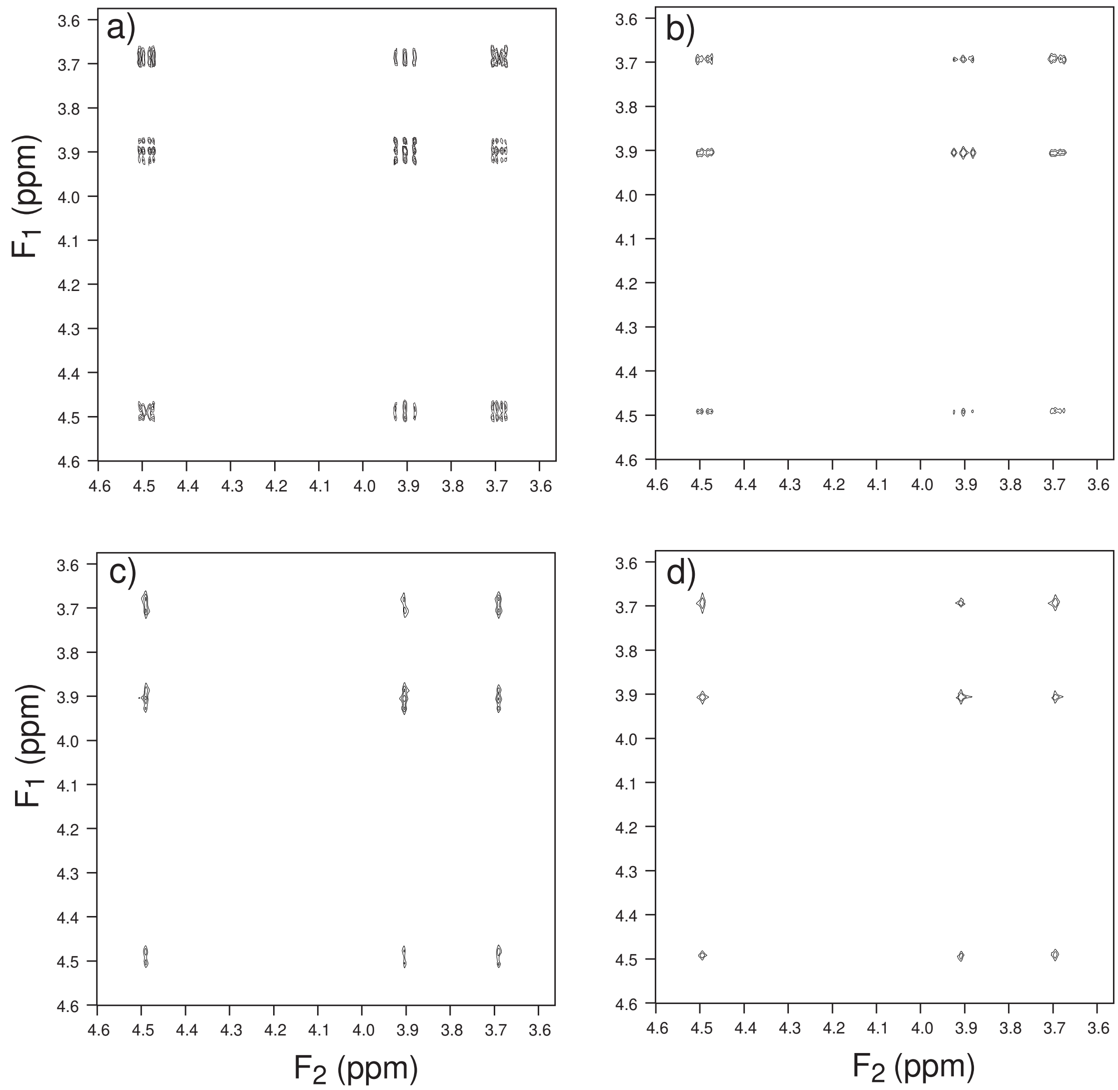



Figure 3.5: A demonstration of successive simplification of TOCSY spectra using a weakly coupled three-spin system. The 4D J-TOCSY-J signal consists of three 2D TOCSY data sets, each of which consists of 200 × 80 complex data points. Panel (a) through (d) show the conventional, homonuclear decoupling in F1 only, homonuclear decoupling in F2 only, and homonuclear decoupling in both F1 and F2 TOCSY spectra. The projection vectors used are: (a) $\vec{p}_1 = (1,0,0,0), \vec{p}_2 = (0,1,0,0)$; (b) $\vec{p}_1 = (1,0,0,0), \vec{p}_2 = (0,1,0,-1)$; (c) $\vec{p}_1 = (1,0,-1,0), \vec{p}_2 = (0,1,0,0)$; (d) $\vec{p}_1 = (1,0,-1,0), \vec{p}_2 = (0,1,0,-1)$.

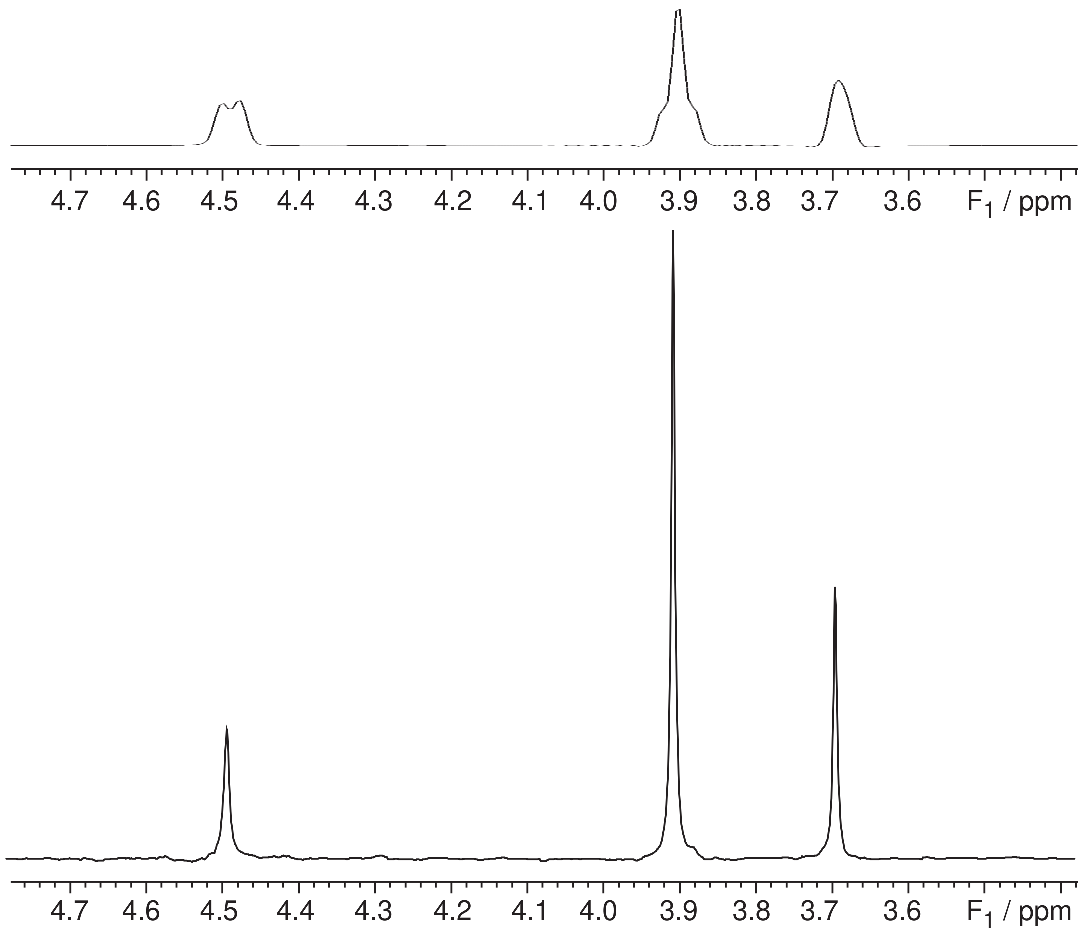



Figure 3.6: Vertical traces along the center peak of the central triplet of conventional TOCSY (panel (a) of Figure 3.5) and *singlet*-TOCSY spectra (panel (d) of Figure 3.5).

is not clear whether and how we should regularize the second eigenvalue problem of Eq. 3.14. Until we have solutions for overcoming these difficulties, application of *singlet*-TOCSY to simplify complex systems with many heavily overlapped multiplets will remain problematic.

## 3.4 Constant-Time NMR Signals Processed by FDM

In conventional 2D NMR experiments, two data sets, either N-/P-type phase modulated signals or cosine/sine amplitude modulated signals, are acquired[4]. This is an absolute must for FT to obtain a double absorption spectrum. However, FDM only requires a single purely phase modulated signal, either N- or P-type, to compute a double absorption mode 2D spectral estimation. It appears that processing N- and P-type independent data sets, and then combining the results does not greatly improve the FDM results. In addition, FDM might fail due to insufficient signal size of a single phase-modulated signal (either N- or P-type). Instead, a single, twice as long N-type (or P-type) data set would be far superior for FDM data processing, while such as a data set will be useless for FT data processing, because only absolute value spectra can be obtained. Therefore, it seems that the optimal data formats for FT and FDM are different and incompatible with each other. Fortunately, for the special case of Constant-Time (CT) NMR experiments, there exists a doubling scheme for FDM to co-process both N- and P-type signals together as a single signal, but twice as large. With the doubling scheme, optimal data formats for FDM and FT become compatible. In addition, there are some special properties of CT signals, making them particularly suitable for the FDM spectral analysis.

In this section, we will first briefly review relevant properties of CT NMR experi-

[4]These two types of data sets can be converted into each other simply by taking linear combinations. For example, given cosine and sine modulated time signal, $S_c(t_1, t_2)$ and $S_s(t_1, t_2)$, purely phase modulated N- and P-type signals can be obtained as,

$$S_P(t_1, t_2) = S_c(t_1, t_2) + iS_s(t_1, t_2)\,, S_N(t_1, t_2) = S_c(t_1, t_2) - iS_s(t_1, t_2)\,.$$

ments, and then describe the FDM doubling scheme for processing CT signals. Both numerical and experimental 2D and 3D examples will be given to demonstrate the efficiency of the doubling scheme.

### 3.4.1 Special Properties of CT NMR Signals

Many protein backbone NMR experiments on uniformly-labelled molecules employ a fixed, or constant-time evolution period [124, 125], $2T$, to encode chemical shift information. In some cases, the constant-time period is used to decouple the homonuclear carbon-carbon couplings, leading to sharp and intense singlets instead of broader multiplet, as demonstrated in the CT-HSQC experiment [126]. In other cases, constant-time evolution is encoded within a fixed delay that is already present for coherence transfer between spins, for example, in the CT-HNCO experiment [127], thereby minimizing the total time between excitation and acquisition.

The CT signals almost show no decay in the CT dimension, with the residual linewidth purely determined by the quality of the magnetic field homogeneity, which is typically excellent with modern shimming. In the FT spectral analysis, the linewidth in the CT dimension is governed by the FT time-frequency uncertainty principle. The maximum signal length is limited by half of the total constant period, the obtainable FT resolution is thus limited by $1/T$. Apodization of the CT signal, which is to reduce the sinc-like truncation artifacts, further broadens the lines. Therefore, longer constant-time period is preferred from the resolution point of view, if FT is to be used. However, there are other factors that prohibit the use of very long constant-

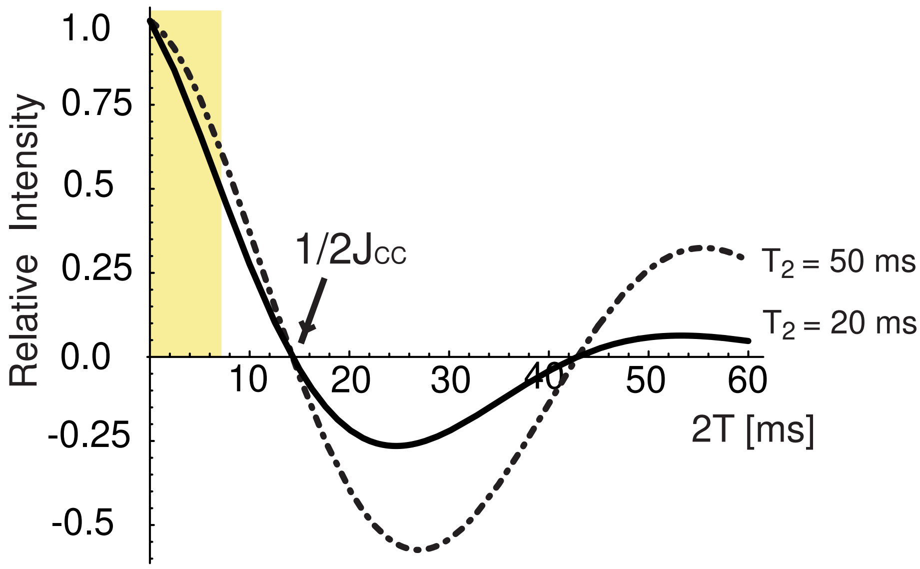



Figure 3.7: The sensitivity of CT experiments as a function of the constant-time period for $C_\alpha$ with a 35 Hz one-bond coupling to a single $C_\beta$.

time period. The observed signal intensity decreases exponentially with respect the constant-time period, $S(2T) \sim e^{-2T/T_2}$, with $T_2$ being the traverse relaxation time constant. Thus to from the sensitivity point of view, a short constant-time period is preferred. Accordingly, when FT is used, one has to compromise between resolution and sensitivity and this is one of the major limitations of CT experiments.

The actual observed intensity of the aliphatic carbon is also modulated by the one-bond coupling between $C_\alpha$ and $C_\beta$, $J_{cc} \sim 35Hz$, further limiting the usable constant-time period for carbon:

$$S(2T) = e^{-2T/T_2} \cos(2\pi J_{cc} T) \,. \tag{3.20}$$

Figure 3.7 plots the relative intensity of $C_\alpha$ as a function of the constant-time period for two different relaxation time constants. To obtain a sufficient FT resolution without losing most of the signal, $2T$ is typically chosen to be either $26 - 28ms$ or $52 - 56ms$, avoiding to the nulls at $1/J_{cc}$ and $2/J_{cc}$.

FDM is a true multi-dimensional method in which the obtainable resolutions in all dimensions are determined together by the total information content of the data set. It is not necessary for the signal to be long in all dimensions in order to obtain high spectral resolutions in all dimensions. In particular, it is possible for FDM to use the information encoded in a long dimension to enhance the resolutions in the short dimensions. Thus, combined with FDM data processing, one does not have to compromise between sensitivity and resolution in CT experiments. Short constant-time periods can be used to improve the sensitivity without sacrificing the resolution in a CT dimension. An additional bonus of using FDM data processing is that the better sensitivity, as a result of using shorter constant-time period, can be effectively translated into a better resolution in the FDM spectra.

Last, but not the least, CT signals have two special properties that make them particularly suitable for FDM data processing. First, CT signals have virtually zero linewidth and thus perfect Lorentzian lineshape in the CT dimensions. Therefore, CT signals can be handled by FDM very efficiently. Second, there is a time reverse symmetry between N- and P-type CT data sets. It is possible for FDM to process them together as a single data set, but twice as large. Due to the sudden convergence property of nonlinear methods, doubling the signal size can often lead to a dramatic improvement in the obtainable resolution. This doubling scheme will be first described in details and then demonstrated using both numerical and experimental signals.

### 3.4.2 A Doubling Scheme for Processing CT NMR Signals

Assuming zero phase at $t_1 = n_1\tau_1 = 0, t_2 = n_2\tau_2 = 0$, which is readily achievable experimentally, the N- and P-type signals for a single peak may be written as,

$$\begin{aligned} S_N(n_1\tau_1, n_2\tau_2) &= \exp[i(\omega_2 n_2\tau_2 - \omega_1 n_1\tau_1)]\,, \\ S_P(n_1\tau_1, n_2\tau_2) &= \exp[i(\omega_2 n_2\tau_2 + \omega_1 n_1\tau_1)]\,, \quad n_1\tau_1 = 0, \tau_1, \ldots, T\,, \end{aligned} \tag{3.21}$$

where $t_1$ corresponds to the CT dimension, and $t_2$ the acquisition time dimension. $\omega_2$ is complex, but $\omega_1$ is essentially real[5], as the imaginary part vanishes in the CT dimension. In a conventional picture, the N-type signal appears to evolve forward in time but with "negative frequency". Alternatively, we can consider it to evolve with the same "positive frequency" as P-type signal, but with negative time step: $n_1(-\tau_1) = 0, -\tau_1, \ldots, -T$. Therefore, by preceding the P-type data set, in the time domain, with the N-type data in reverse chronological order, a pseudo-signal that evolves effectively almost twice as long in CT dimension results:

$$S_{NP}(n_1\tau_1, n_2\tau_2) = \exp[i(\omega_2 n_2\tau_2 + \omega_1 n_1\tau_1)]\,, \quad n_1\tau_1 = -T, \ldots, T\,. \tag{3.22}$$

Note that this is achievable only if there is no decay in the CT dimension. Otherwise, $S_{NP}$ will have a Golden Bridge shape and is inconsistent before and after $t_1 = 0$.

Even though the combined signal $S_{NP}$ is twice as long, due to the time symmetry at $t_1 = 0$, the FT resolution in the CT dimension is still determined by $1/T$ instead of $1/2T$. FT of such the combined signal automatically produces absorption lineshape in the $t_1$ dimension, which is not surprising since the purpose of acquiring both N-

[5]No assumption that $\omega_1$ should be real is used in FDM when CT signals are analyzed.

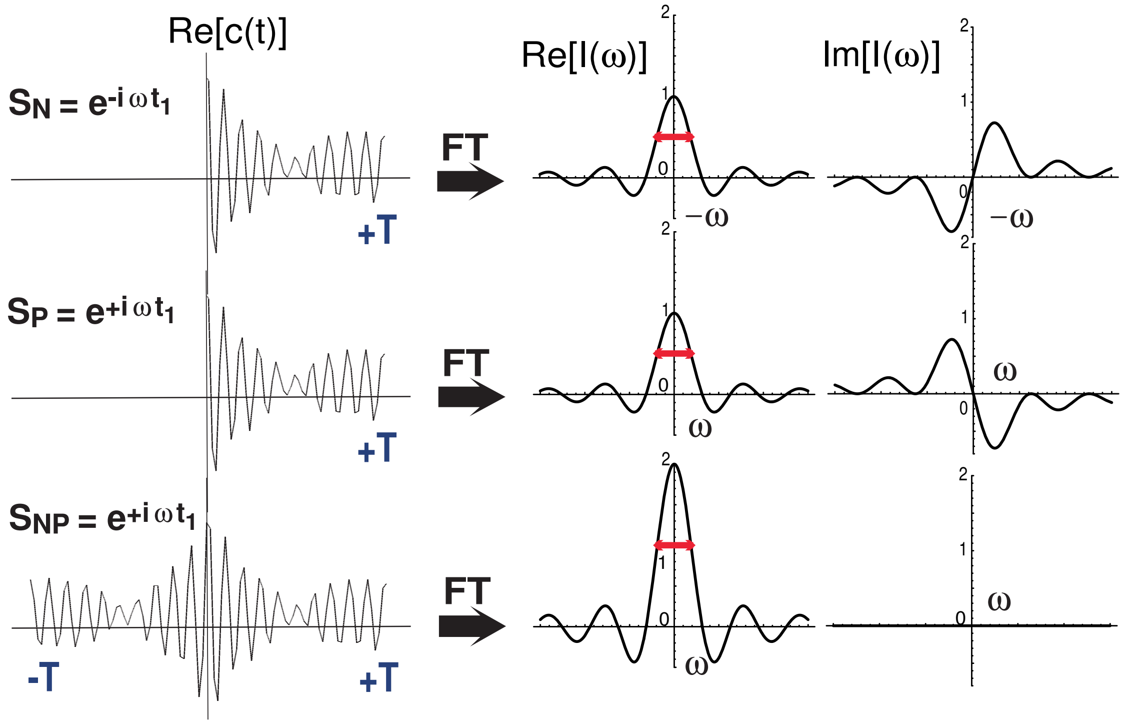



Figure 3.8: Time symmetry between N- and P-type CT signals and Fourier Transforms of them. It shows that due to the time symmetry between N- and P-type CT signals, FT of the combined signal, $S_{NP}$, automatically gives an absorption spectrum with vanishing imaginary part. The peak is twice as intense but the linewidth is the same as those in absorption mode FT spectra of N- or P-type signals (indicated by the double arrows.)

and P-type signals is to obtain an absorption lineshape in the indirect dimension. Interestingly, this property was actually utilized, long ago, to obtain absorption-mode lineshape in 2D NMR experiments [124, 128]. Figure 3.8 illustrates the time reverse symmetry between N- and P-type signals and its effects on the lineshape and resolution of the FT of the combined signal.

However, this time symmetry is totally irrelevant for FDM data processing. FDM is a nonlinear method that is capable of producing double absorption type spectra from a single purely phase-modulated signal. The signals effectively evolve twice as long and thus the combined signal contains roughly twice as much information that can be utilized by FDM. Due to the "sudden convergence" property of any nonlinear method, doubling the signal size can lead to a substantial resolution enhancement.

Our procedure is clearly reminiscent of the method called "mirror-image" linear prediction [129, 127, 130]. It has been noticed [127] that, as the CT signal does not decay, twice as many points can be used to extrapolated the signal by LP if the data is "reflected" in $F_1$. Thus twice as many LP coefficients can be computed and then used to extrapolate the data, usually be a factor of two. Once the longer signal is obtained, the "mirror-image" part is discarded and spectrum is obtained by taking the FT of the new extended data set. This procedure can improve the resolution in the $F_1$ dimension, although the final spectral resolution can not exceed that expected from a hypothetical experiment in which data over [0,2T] is collected and transformed. There are further limitations. Reflecting the signal also symmetrizes the noise, which could introduce some avoidable bias.e An improved mirror-image LP will use $S_{NP}$, to compute the LP coefficients and then use them to both forward and backward extrapolate the signal by a factor of two to obtain an extended signal , available from $t_1 = -2T, \ldots, 2T$, from which extended N-type and P-type signals can be obtained by simply keeping the negative or positive halves. A more severe limitation of the mirror-image LP procedure, which is shared by the Fourier transform, lies in the fact that it is essentially a 1D procedure and unable to utilize the important information that evolutions in two dimensions are correlated.

In order to process the signal $S_{NP}$, only minor changes to the previous FDM2K algorithm are required. Let's redefine the 2D discrete time signal $c(\vec{n}) = c(n_1\tau_1, n_2\tau_2)$ to be available from $n_1 = -(N_1 - 1), \ldots, 0, \ldots, N_1 - 1$ and $n_2 = 0, \ldots, N_2 - 1$. The new Fourier basis then reads,

$$\begin{aligned}\left|\Psi_j^{NP}\right) &= \sum_{n_1=-(M_1-1)}^{M_1-1} \sum_{n_2=0}^{M_2-1} e^{in_1\tau_1\varphi_{1j}} e^{in_2\tau_2\varphi_{2j}} \left|\Phi_{n_1,n_2}\right) &(3.23)\\ &= e^{-i(M_1-1)\varphi_{1j}} \sum_{n_1=0}^{2(M_1-1)} \sum_{n_2=0}^{M_2-1} e^{in_1\tau_1\varphi_{1j}} e^{in_2\tau_2\varphi_{2j}} \left|\Phi_{n_1-(M_1-1),\,n_2}\right) \\ &= e^{-i(M_1-1)\varphi_{1j}} \left|\Psi_j(M_1-1)\right), &(3.24)\end{aligned}$$

where $M_1 \equiv N_1/2, M_2 \equiv N_2/2$, and $|\Psi_j(M_1-1))$ denotes the normal 2D Fourier basis defined in Eq. 2.50 but with shifted Krylov basis $|\Phi_{n_1-(M_1-1),\,n_2})$. $j = 1, 2, ..., K_{\text{win}}$.

The $\mathbf{U}$ matrices in the new Fourier basis can be then computed as,

$$\begin{aligned}\left[\mathbf{U}^{(\vec{p})}\right]_{jj'} &= \left(\Psi_j^{NP}\right| \hat{U}^{(\vec{p})} \left|\Psi_{j'}^{NP}\right) \\ &= e^{-i(M_1-1)(\varphi_{1j}+\varphi_{1j'})} \left(\Psi_j(M_1-1)\right| \hat{U}^{(\vec{p})} \left|\Psi_{j'}(M_1-1)\right)\ , &(3.25)\end{aligned}$$

where the $(\Psi_j(M_1-1)|\,\hat{U}^{(\vec{p})}\,|\Psi_{j'}(M_1-1))$ part can be computed efficiently using the formulas derived previously (Eqs. 2.52-2.54), but with shifted data matrices, $c(\vec{n}' + \vec{p}) = c(n_1 - 2(M_1-1) + p_1, n_2 + p_2)$. $e^{-i(M_1-1)(\varphi_{1j}+\varphi_{1j'})}$ is simply a “phase factor” that can be included after the $(\Psi_j(M_1-1)|\,\hat{U}^{(\vec{p})}\,|\Psi_{j'}(M_1-1))$ term is computed.

Once the $\mathbf{U}$ matrices are available, one can solve the generalized eigenvalue problems of Eq. 2.62 and then use the Green's Function formulas, such as Eq. 2.68, to construct various 2D spectral estimations. The last minor modification is to redefine the FT column vector $\tilde{\mathbf{C}}$ (Eq. 2.57) accordingly,

$$\left[\tilde{\mathbf{C}}\right]_j = \sum_{n_1=-(M_1-1)}^{M_1-1} \sum_{n_2=0}^{M_2-1} e^{i\vec{n}\vec{\varphi}_j} c(\vec{n})\ ,\quad j = 1, 2, ..., K_{\text{win}}\ . \qquad (3.26)$$

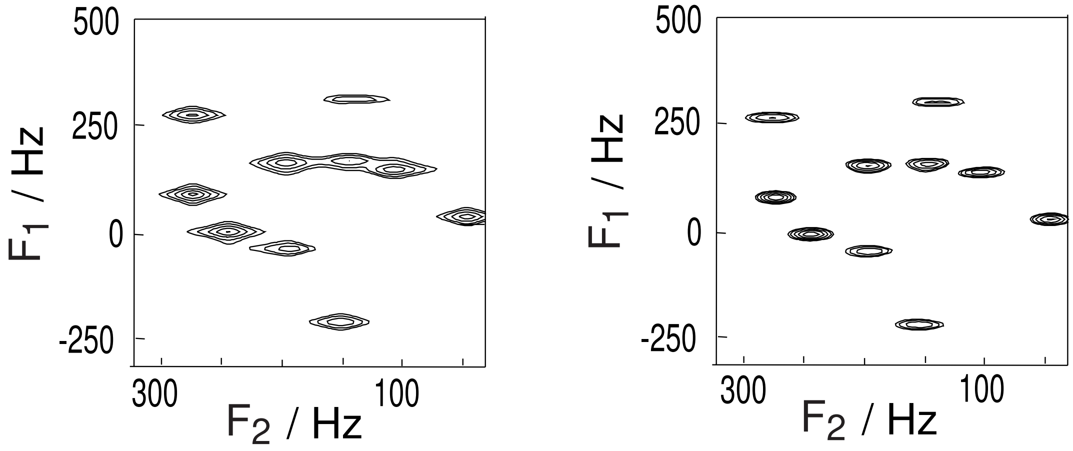



Figure 3.9: A comparison between 2D Lorentzian (left panel) and Gaussian lineshape (right panel), using a selected region of a 2D $^{13}$C Constant-Time HSQC of uniformly labelled human ubiquitin.

### 3.4.3 Lineshape Modification

Although the output of FDM is expressed in terms of Lorentzians lines, it is possible to convert them into any other lineshapes such as a Gaussian lineshape. The most conservative way is to employ the standard Lorentzian to Gaussian deconvolution [131]. A rather more aggressive procedure, and the one we have employed here, is simply to replace each Lorentzian with a Gaussian of the same full-width-at-half-magnitude (FWHM) and equal integral,

$$I_G(\omega_1, \omega_2) = \frac{4}{\pi} \sum_{k,k'} \frac{\mathrm{Re}\,[d_{k,k'}]}{\gamma_{1k}\gamma_{2k'}}\, e^{-(\omega_1 - \nu_{1k})^2/\gamma_{lk}^2}\; e^{-(\omega_2 - \nu_{2k'})^2/\gamma_{2k'}^2}\,, \tag{3.27}$$

where $\nu_{lk} = \mathrm{Re}[\omega_{lk}]$ and $\gamma_{lk} = -\mathrm{Im}[\omega_{lk}]$ are the line positions and linewidths computed by FDM. The well-known star-shaped contours of the 2D Lorentzian peaks become elliptical contours that are more easily discerned, as demonstrated in Figure 3.9. This line shape modification is purely cosmetic in the sense that the information on the peaks is already entirely encapsulated by the parametric description of the spectrum. The downside of this more aggressive procedure is that when broad lines interfere by overlapping Lorentzian tails, the sharper cutoff of the Gaussian may introduce "new" features that are misleading. Practically, well-resolved spectra are readily obtainable by FDM and additional artifact due to Gaussian formula is often at minimum. When the method fails to resolve all the peaks, instability might occur locally in the unresolved region. This will be further discussed later on in this section when we demonstrate this doubling scheme using a experimental NMR signal and study its sensitivity to the noise level.

### 3.4.4 Numerical Experiments

Numerical studies on model signals plus noise have the advantage that the correct answer is known exactly. As long as this information is not employed in any way by the algorithm, model data can be used as a powerful tool for studying and demonstrating new computation methods. We have made an attempt to objectify the performance of 2D FDM2K by examining model N- and P-type signals from a hypothetical constant time experiment, including various amounts of uncorrelated random noise. Figure 3.10 shows a model spectrum meant to mimic a section of a CT-HSQC spectrum. There are 13 peaks of identical integral, with similar line widths in $F_2$ in the range of $2.2 \pm 0.2$ Hz (at random) and zero line width in $F_1$ (mimicking CT dimension). The spectral width is 1 KHz in both dimensions. The SNR is extremely high, around 1000:1 in the time domain. The exact spectrum, panel (a), has been line broadened in $F_1$ to roughly the mean width in $F_2$ to allow meaningful contours to be drawn. The absorption mode DFT spectrum is shown in panel (b), computed from a pair of N- and P-type data with size $10 \times 10$. Cosine apodization was used to limit the sinc-function wiggles and zero-filling to $512 \times 512$ points was used to obtain same digital resolution as the FDM spectra. Some of the peaks are resolved, but others are too close to distinguish. In panel (c) the FDM double-absorption spectrum from the $10 \times 10$ time-domain P-type data set is shown. Many of the peaks are obtained quantitatively, but local crowding results in several pairs of peaks being fit as a single broader peak. Note in addition that the apparent width of the single feature is less than the separation of the two genuine peaks. Convoluting this spectrum with the transform-limited line shape still gives a good match to the actual DFT spectrum (data not shown), but this is clearly not sufficient to guarantee that all the information has been obtained. Finally, in panel (d) the effect of processing a single "NP" data set, $S_{NP}$, is shown. The larger basis size (and possibly slightly better overall SNR) allows all the peaks to be obtained correctly. With experimental data, of course, one never knows in advance what the correct result is supposed to be! However, by doing several FDM2K calculations with different basis size, it is possible to identify stable recurring features, and also features that depend on the size of the basis. As long as a feature is stable with respect to basis size and the regularization parameter, our experience has been that it is quite accurate and reliable.

Figure 3.11 is an attempt to be somewhat more realistic. The same features are present as in Figure 3.10, but now the noise level is substantially higher, giving around 20:1 SNR in the time domain. A DFT of a pair of $18 \times 18$ time-domain data sets shows resolution of most, but not all, of the peaks in the absorption-mode spectrum, panel

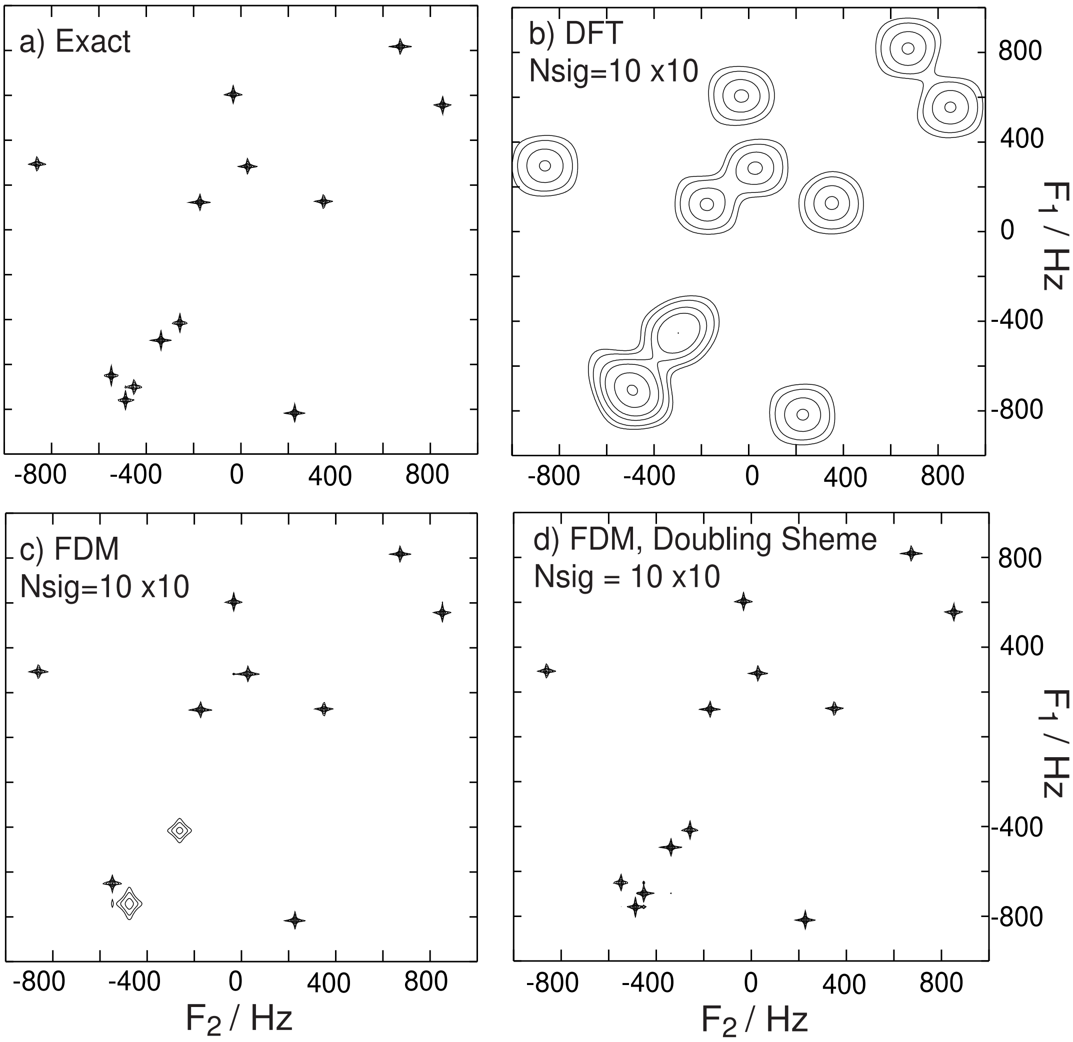



Figure 3.10: Demonstration of FDM doubling scheme for analyzing CT NMR signals, using a model signal with a high SNR of 1000:1 in the time domain (first point, see text for details). In all FDM spectra, the linewidth in $F_1$ (CT dimension) is smoothed to the average linewidth of $2Hz$ in $F_2$. With $Nsig = 10 \times 10$, while neither DFT nor normal FDM is able to fully resolve the most crowded region, FDM with doubling scheme is able to efficiently use the information from both N- and P-type data sets together and provide a fully resolved spectral estimation, shown in panel (d). Note that 2D DFT of the combined signal $S_{NP}$ will be identical to the DFT spectrum shown in (b), given that the factor of 2 in intensity is correctly taken into account (data not shown).

(a). The FDM2K calculations all employ the doubling scheme, but now the results are less impressive. Using only a $10 \times 10$ signal there is now inadequate information for FDM2K to accurately identify all the peaks, as shown in panel (b). Increasing the signal size to $14 \times 14$ in panel (c) sharpens up most of the features. But only when using the same $18 \times 18$ data set as used in the DFT calculation again allows all the features to be identified, as shown in panel (d), although there are still some errors in the widths and intensities of the most crowded peaks. Further increasing the signal size only results in minor improvements and is evident that the calculation has converged to the true features.

Another interesting observation of the numerical experiments of Figure 3.11 is the different behaviours of peaks in crowded regions and isolated ones, due to the local convergence property of FDM. Isolated features tend to converge first and stay stable with respect to the change of signal size, while crowded regions are more difficult to converge and sensitive to the change of signal size until fully resolved. By computing a series of FDM spectra with varying signal size, one can quite reliably decide whether a particular feature is genuine or not. In addition, sensitivity of unconverged features is not limited to varying signal size. Other FDM parameters such as basis size, basis density and regularization can also be used to probe the different stabilities between converged and unconverged features. Again, in practice when we use FDM to analyze experimental NMR signals, it is always necessary to produce several, if not many, FDM spectra with different signal size and computing parameters before we want to draw some reliable conclusions on the results as obtained.

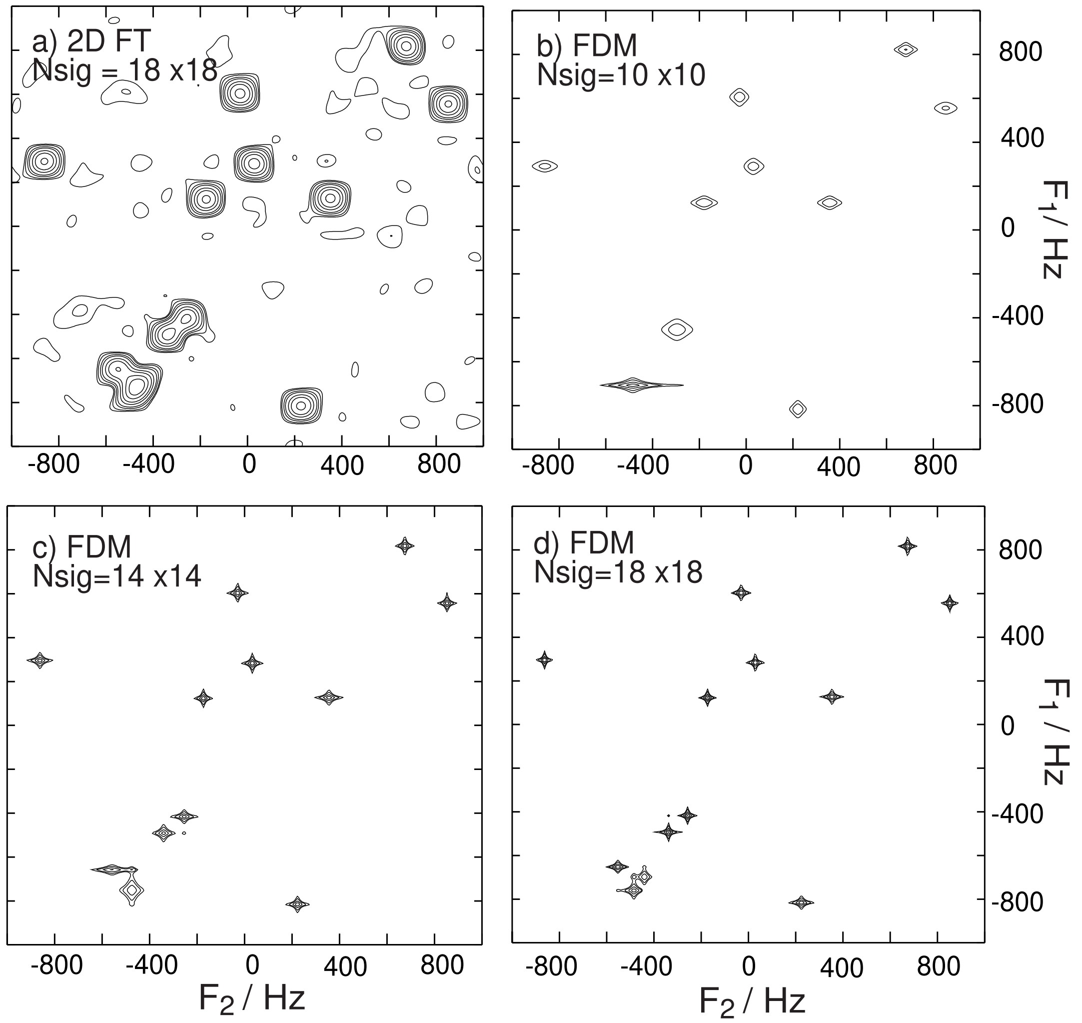



Figure 3.11: Demonstration of FDM doubling scheme for analyzing CT NMR signals, using the same model signal as Figure 3.10 but with a low SNR of 20:1 in the time domain. Panel (a) shows a contour plot of absorption mode DFT spectrum obtained using $18 \times 18$ complex data points (both N- and P-type signals), with the contours drawn low enough t show some of the noise. Panel (b) through (d) show the FDM spectra obtained using successively larger data sets, from $10 \times 10$ up to $18 \times 18$. The doubling scheme was used in all calculations. A continuous improvement in both intensity and resolution is seen with increasing signal size. With $18 \times 18$, all peaks are essentially resolved by FDM. Further increase of signal size will slightly improve the peak lineshapes and intensities, but the resolution remains similar (not shown). Also note the different behaviours of the peaks in crowded regions and isolated ones, which is an important way of checking the convergence of FDM calculations (see text).

### 3.4.5 Application to a CT-HSQC Experiment

The 2D Heteronuclear Single-Quantum Correlation (HSQC) experiment [132] is one of most basic and most useful NMR experiments. HSQC correlates heteronuclear resonances with proton resonances by transferring the polarization between heteronuclear spins, typically $^{1}$H and $^{13}$C or $^{15}$N. It is routinely used to obtain the backbone assignments of proteins due to the high sensitivity. Together with 2D Heteronuclear Multiple-Quantum Correlation (HMQC) [133], they are integral components of all heteronuclear 3D and 4D experiments [11].

Constant-Time HSQC (CT-HSQC) [126, 11] is often used to simplify the spectra by collapsing the multiplets along the carbon dimension into sharp and intense singlets, making it easier to resolve and assign all the resonances. Here we demonstrate the power of FDM with doubling scheme using a CT-HSQC experiment of a model protein, human ubiquitin. All the $^{1}$H-$^{13}$C CT-HSQC spectra shown were obtained using a 500 MHz Varian UnityPlus spectrometer equipped with a standard 5 mm HCN triple resonance probe. The sample was 700 ml of a 1 mM solution of 100% doubly labeled ($^{13}$C and $^{15}$N) human ubiquitin in $D_2O$ (VLI Research) at 24 $^oC$. A slightly modified pulse sequence, shown in Figure 3.12, was used to acquire the signals. The partly selective *noco* and *co* pulses [136], of duration 200 ms, either invert the aliphatic region of the carbon-13 spectrum and not the carbonyls, or vice-versa. Their only role here is to provide a spectrum with zero phase correction in $F_1$ , with no Bloch-Siegert phase shifts [137], and no loss of signal intensity over the slightly inhomogeneous radio-frequency field on the carbon-13 channel. Their short pulse

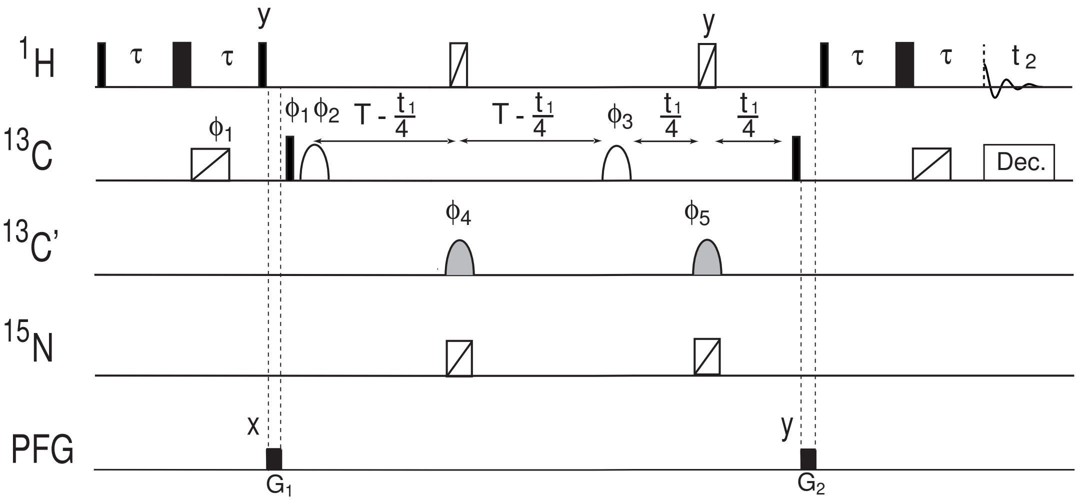



Figure 3.12: Modified CT-HSQC pulse sequence to observe correlations between $H_\alpha$ and $C_\alpha$ spins in a protein. Aside from the minor addition of a pair of partly selective pulses, denoted *noco* and *co*, and the broadband inversion pulses (BIPs) [134], denoted with scored icons, the pulse sequence is standard. The 200 ms *noco* and *co* pulses have been developed for ultra-short CT experiments, in which longer pulses reduce the available number of increments. They work to invert the aliphatic region of the $^{13}$C spectrum without inverting the carbonyl region (*noco*) or vice-versa (*co*). The BIPs cleanly invert the entire chemical shift range of the indicated nucleus. All these pulses show superb compensation for RF inhomogeneity. By using the pair of partly selective 180s, all phase shifts are removed, leading to a spectrum with zero phase correction in the indirect dimension. The proton $90^o$ time was 10 $\mu s$ and the carbon-13 $90^o$ time was 15 $\mu s$. The *co* and *noco* shaped pulses are applied with a maximum of 10 kHz; the other pulses are applied at full power. All the spectra shown result from two scans per increment, in which the phase $\phi_3 = x, y$; receiver$=x, -x$. $\phi_1$ and the receiver phase are inverted for each $t_1$ increment. Cosine/sine modulated data are acquired according to the States-TPPI scheme [135].

width ensures maximal use of 2T period for data acquisition.

As the brief gradients during the longitudinal states of the INEPT and reverse INEPT steps were enough to suppress the residual HDO resonance, amplitude modulated data sets were in fact acquired. The cosine and sine data sets required two scans each to complete the simplest difference phase cycle to select the $^{13}$C-bearing protons. These two data sets were then combined to form pseudo N- and P-type data sets, which were then processed as above. Calculations with actual gradient-selected data showed no significant difference as long as the experiments were done carefully and the same number of increments obtained. For ultra-short CT experiments, the loss of time during the actual pulsed field gradient is disadvantageous.

### Noise Performance

The numerical experiments of the last section carry over fairly well to actual CT-HSQC data. Figure 3.13 shows a zoomed section, in a crowded portion of the $C_\alpha$ region, of a $^1$H-$^{13}$C CT-HSQC spectrum. A signal with $1024 \times 90$ complex data points (for both N- and P-type data sets) was acquired with a short constant time of $2T = 7.5$ ms, well below the null in the CT transfer function at $\sim 13$ ms, and almost a factor of four shorter than the typical time of $2T = 1/J_{cc} \approx 26.4$ ms. The total experimental time is about 7 minutes. Lowering the power of the last proton 90 pulse has mimicked the effect of poorer sensitivity. The 2D FT spectra are on the top panels and the 2D FDM2K spectra from the same data are directly underneath. The FT spectrum from the FID of the first increment, with some blank down-field region included for a noise reference, is shown along the top of each of the series.

As the effective read pulse is reduced to $11^o$ and finally $5^o$, corresponding to a relative SNR five and ten times lower, respectively, it is clear from the first increment that the data becomes quite noisy. Nevertheless, FDM2K with the new doubling scheme behaves remarkably predictably, with the achievable resolution gradually diminishing, and weaker features disappearing along with the noise when the regularization is used. There is clearly no catastrophic loss of performance, and the resolution in many regions is always preferable to the FT result. The fine structure along the

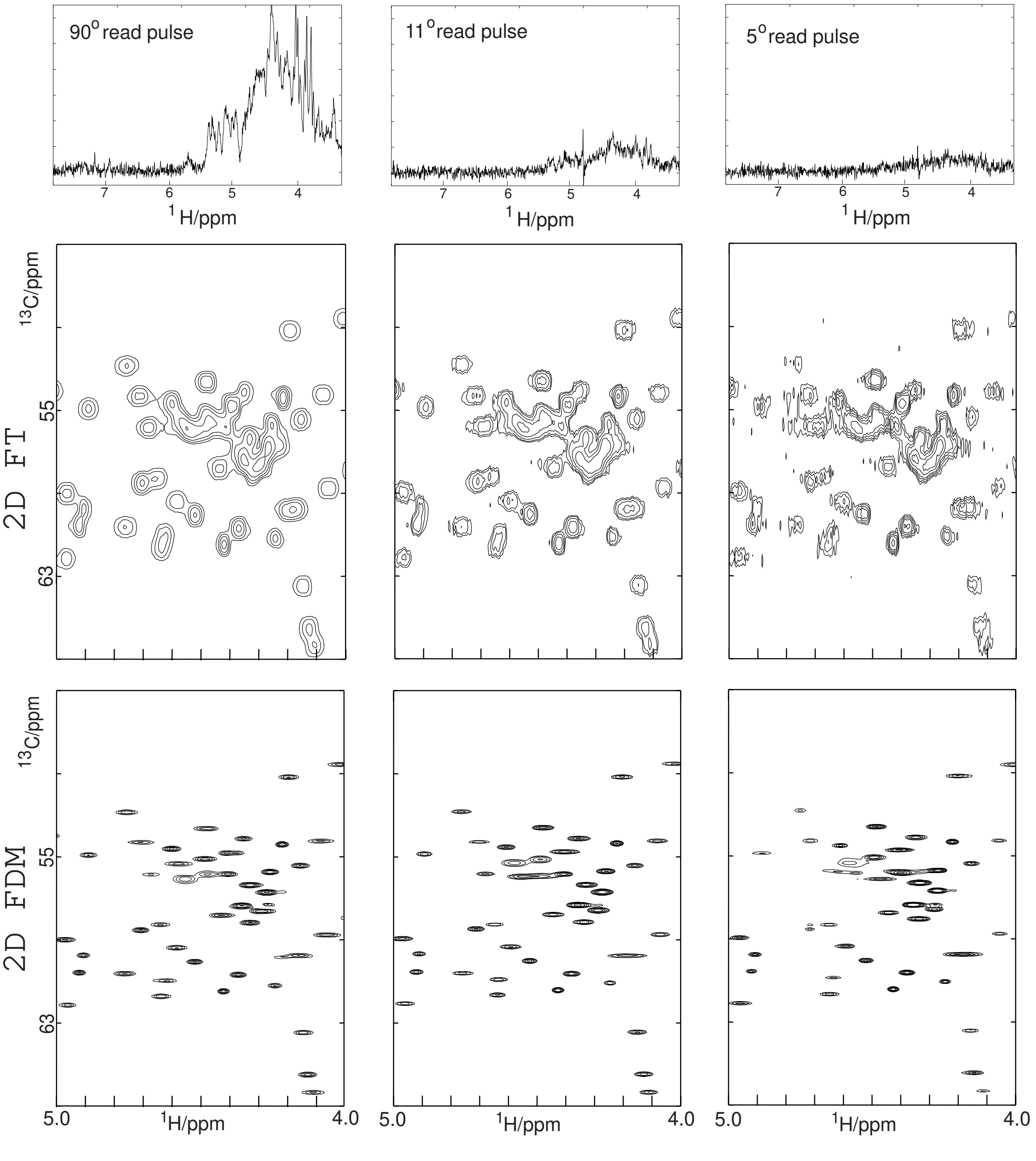



Figure 3.13: Demonstration of the noise performance of FDM2K. The top traces show the FT of the FID from the first increment, with some down-field region included to show the noise level. The contour plots below show the 2D FT (top series) and 2D FDM2K (bottom series) spectra. The pulse sequence of Figure 3.12 has been used with the power of the 10 ms read-out proton $90^o$ pulse attenuated to achieve the desired flip angle. The 2D FT spectra are shown in the top series, and the 2D FDM2K spectra, from the doubled NP data set, are shown directly underneath. Regularization has been increased as the SNR decreases from left to right. Two scans per increment were recorded for the total of 90 increments, for the sine and cosine data sets. (a) $90^o$ read pulse. (b) $11^o$ read pulse (-18 dB). (c) $5^o$ read pulse (-25 dB).

proton dimension, a known result of a flip angle effect [138] when the small flip angle read pulses are used, is blurred away by the regularization. The result is a set of nearly structureless ellipses centered on the correct chemical shifts. Particularly with the noisier spectra, attempts to resolve the proton multiplet structure, and simultaneously separate the close carbon-13 shifts, have not been successful.

### Ultra-short Constant Time Performance

Attenuation of $C_\alpha$ magnetization during the constant time period is a common problem in larger proteins, where the $T_2$ losses can be severe. However, using larger $\mathbf{B}_0$ fields in conjunction with FDM2K may allow very short constant time periods, greatly improving sensitivity without the need for deuteration. This is in contrast to the usual fixed value of $2T \approx 26.4$ ms, independent of the field strength. The difference is that by operating with (much) shorter periods than the null condition at 13 ms, shown as the shade area in Figure 3.7, the delay can be shortened as much as we like, fixing only the number of increments obtained. For example, if a decent CT-HSQC spectrum can be obtained using $2T = 4.25$ ms at 500 MHz, then only 2.6 ms would be required at 800 MHz. The relative basis function density versus the number of local peaks is identical. However, there can be a great improvement in performance both on account of the superior intrinsic sensitivity of the higher field strength, and the reduction in relaxation losses. The local nature of the convergence of FDM means that many of the spectral regions can be essentially completely resolved, leaving a few dense regions that may require a longer experiment, some editing techniques, or a higher-dimensional experiment to make headway. Alternatively, the same 4.25 ms experiment may, at higher field, allow completely converged spectra to be obtained. The ability to achieve high resolution with very short constant time periods could establish FDM as a key component in successful NMR of larger proteins at high field.

Human ubiquitin is not a rapidly-relaxing protein, but can be studied simply to assess the degree of separation of the resonances that can be obtained with ultra-short constant times. Figure 3.14 panel (a) shows the conventional absorption-mode spectrum in the $C_\alpha$ region of the CT-HSQC spectrum using 26.4 ms and the pulse sequence of Figure 3.12. Two scans per increment for the N- and P-type data were used. This is our reference spectrum for the study. Many of the 2D peaks are clearly resolved, with a couple of denser regions. In the right-hand panel, (b), the FT spectrum obtained with $2T = 4.25$ ms is shown. Many of the distinct peaks have coalesced because of the reduced resolution along the dimension. In panel (c) FDM absorption mode spectrum of processing the P-type data only is shown for the same 4.25 ms data set. A similar result was obtained with the N-type data (not shown). The spectrum shown is sensitive to changes in FDM processing parameters, and would not be considered reliable. Many regions of the spectrum remain unresolved, as the basis density is apparently too low to capture the local number of peaks. There are only eight basis functions along the $C_\alpha$ dimension in the region displayed. Finally in panel (d) the FDM2K result using the same data, but now with doubling scheme, is displayed. Clearly the combined signal $S_{NP}$ is a case where the whole is greater than the sum of the parts. Many of the regions display essentially identical resolution to the reference FT spectrum, panel (a). In the most crowded region the basis function density is still insufficient to characterize all the local 2D peaks, and so a distorted

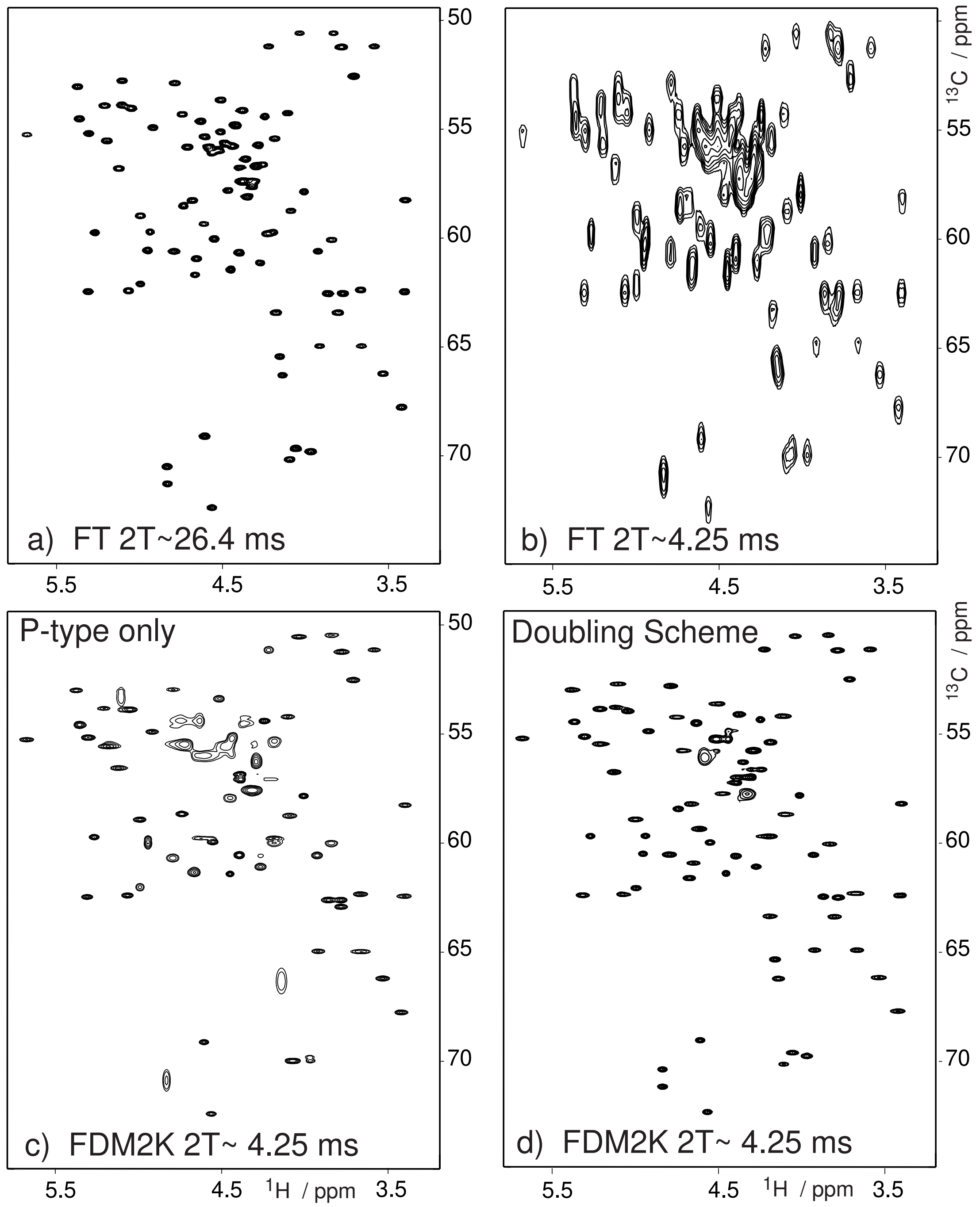



Figure 3.14: Ultra-short constant time spectra compared with conventional measurement. (a) Reference 2D FT spectrum using $2T = 26.4$ ms. (b) 2D FT spectrum using $2T = 4.25$ ms. (c) 2D FDM2K spectrum using the same 4.25 ms data set as in (b), processing only the P-type data. This 2D spectrum is not stable in many regions and would not be considered reliable. (d) Result of the FDM with doubling scheme applied to the same 4.25 ms data set. There is a vast improvement in both resolution and stability. Comparison of reference spectrum, panel (a), shows that peaks that are resolved are also very accurately recovered by FDM2K.

spectrum with some wide features is obtained. There are ways to conclusively show that regions like this are not converged, but no simple and reliable way to correct the inaccurate results. For example, a true resonance should have a line width of nearly zero along the $C_\alpha$ dimension, with errors that depend on the noise level and quality of the data. The "unconverged" features have line widths that are quite large, and that are sensitive to the exact parameters of the FDM calculation, allowing them to be flagged as unreliable. Another approach is to use the FDM parameters to simulate the FT spectrum, and compare the result with the actual DFT of the data. Ideally, only featureless noise should remain after subtraction, if a perfect fit is obtained. In unconverged windows, a substantial residual may be used as evidence of untrustworthy local spectral features. More details can be found in Section 2.6.

Note that there are only 50 increments for the 4.25 ms experiment, making this an ultra-fast acquisition as well - less than four minutes. The calculation of the displayed regions using FDM2K is completed in a matter of a few minutes on either an AMD XP 1800+ Athlon PC running Linux, or an older 533 MHz DEC Alpha workstation. Calculation of the entire 2D spectrum takes of the order of 2-3 hours depending on the size of the window basis. As many regions are completely devoid of signals, however, a more realistic estimate is around 45 minutes for the relevant areas. With an appropriate parallel architecture, the computing time could be a few minutes for virtually any size data set, as each window is a separate job.

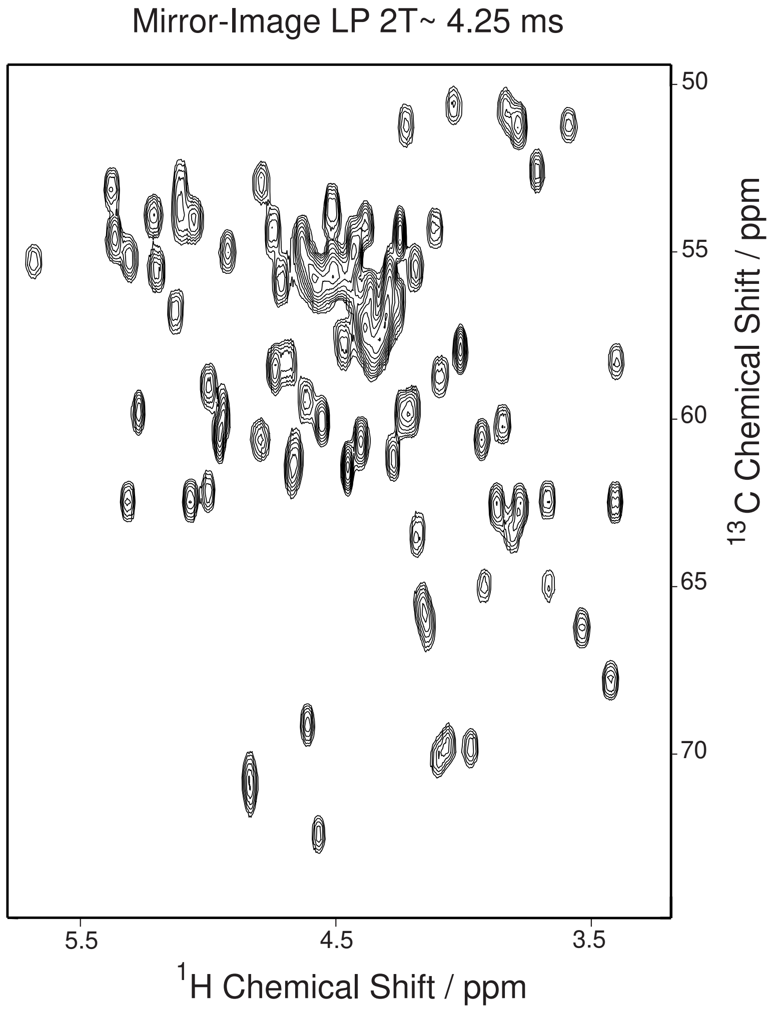



Figure 3.15: A absorption mode spectrum obtained by mirror-image LP, applied to the same $2T = 4.25$ ms ultra-short CT signal used in previous example. There is essentially no resolution enhancement compared to the 2D FT spectrum show in Figure 3.14, panel (b). This is mainly due to the fact that mirror-image LP employs trace-by-trace 1D data processing and totally misses the important information that two dimensions are actually correlated. The filter length was optimized by trial and error.

### Comparison with Mirror-Image LP

There are many factors that work together to make the result shown in Figure 3.14 possible, including the two special properties of CT signals and intrinsic high-resolving power of FDM. In particular, it is essential to process the whole two-dimensional together and make use of the important information that the evolutions in two dimensions are correlated. To demonstrate the point, here we applied the more commonly employed mirror-image LP to the same ultra-short CT data set used in Figure 3.14. The mirror-image LP implemented in NMRPipe [139] software package was used to obtained the spectrum shown in Figure 3.15. The filter order was optimized by trial and error. Because mirror-image LP is implememnted as 1D LP extrapolations of the $F_1$ interferograms, rather than a true 2D method, it is not able to use the information that both dimensions are correlated. In this case, there are very few increments in the indirect dimension (effectively 8 data points for the region shown, thus a maximum of 4 poles can be used), making it impossible for LP to efficiently model the evolution along $F_1$.

### Performance on Larger Proteins

Human ubiquitin is a relatively small and well-behaved protein that is known to give some of the best spectra ever seen. It is not obvious how FDM will perform on larger and less well-behaved proteins, until we try the same experimental and data processing scheme on such kind of systems. Unfortunately, highly pure doubly-labeled proteins are not readily accessible except for molecular biologists, making extensive test of FDM currently infeasible. Here we present some preliminary results where FDM was applied to analyze CT-HSQC experimental signals of MTH1598, a 16 KDa protein with 141 residues. MTH1598 is one of the non-membrane protein of *methanalbacterium theromautotrophicum* $\delta H$, courtesy of Dr. Cheryl Arrowsmith of the Ontario Cancer Institute. Figure 3.16 compares the conventional DFT spectrum of a long CT-HSQC experiment with $2T = 26.4$ ms, with the FT and FDM spectra of a much shorter experiment with $2T = 7.1$ ms. It clearly shows that the performance of FDM, shown in the center panel, is just as impressive. The resolution of the FDM spectrum, obtained using 7.1 ms constant time, is even higher than that of the FT spectrum of a much longer experiment, shown in the top panel. There is another interesting observation. The FDM spectrum shows a weak peak in the right bottom corner, highlighted by a rectangular box, which is missing in high-resolution FT spectrum. For comparison, the bottom panel shows the FT spectrum using 7.1 ms constant time. The peak is now present, proving that it is genuine. However, the resolution is much worse. For example, the two peaks highlighted by the square box are not resolved in low-resolution FT spectrum, but well resolved in both the FDM and high-resolution FT spectra. This is exactly one of the major limitation of CT experiments that we discussed in the beginning of this section: if FT is used, one has to compromise between resolution and sensitivity. Therefore, FDM might even be more useful for analyzing larger proteins, especially those that relax very fast. In this case, FDM data processing, combined with short or ultra-short constant time experiments, is the optimal scheme with both good sensitivity and high resolution.

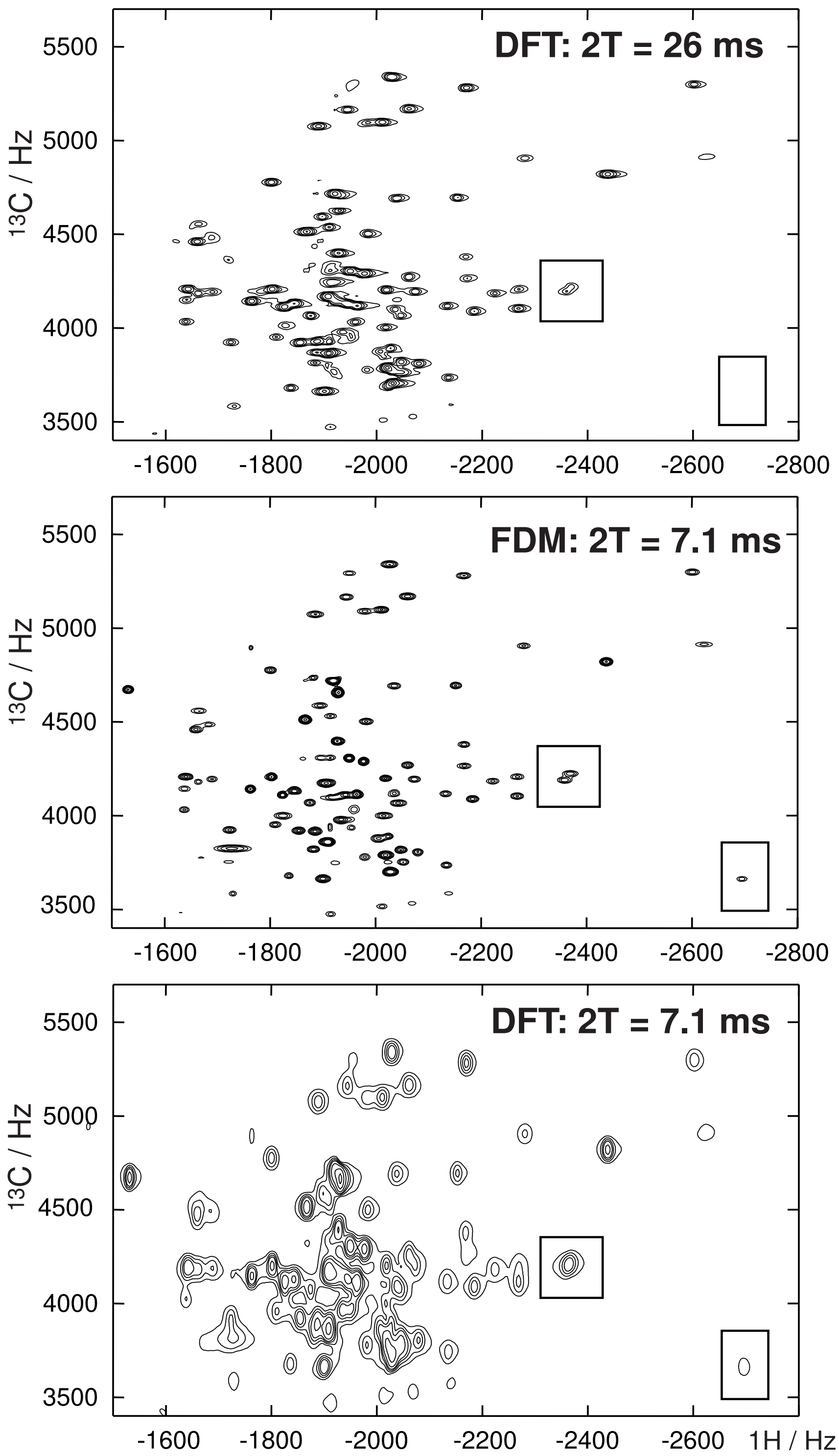



Figure 3.16: Comparison of CT-HSQC spectra of MTH1598 obtained with 26.4 ms constant time and processed by FT (top), with those obtained with 7.1 ms constant time and processed by: FDM with doubling scheme (center) and 2D FT (bottom). Short constant time, combined with FDM data processing, seems to optimal and can provide both high resolution and good sensitivity.

### 3.4.6 A Quadrupling Scheme for 3D NMR Experiments with Two Constant-Time Dimensions

We have demonstrated that FDM, as a true two-dimensional method, can have a substantial advantage over the FT spectral analysis or any other 1D methods like mirror-image LP. The advantages of true multidimensional spectral analysis will be even more enormous when more dimensions are involved. In this section, we extend the FDM doubling scheme for 2D CT signals to a *quadrupling scheme* for 3D NMR experiments with CT evolution implemented in both indirect dimensions. The achievable resolution enhancement will be really astonishing.

Extension of 2D FDM to the 3D case is straightforward. A third evolution operator is introduced and the formulas derived in Section 2.3 for the 2D case can be easily modified accordingly. Additional details can be found in Refs. [73] and [91]. For a 3D NMR signal with two CT indirect dimensions, all the four 3D data sets[6] can be combined together and processed by FDM as a single data set, but four times as large. The doubling scheme of 2D thus becomes a *quadruple scheme* in 3D. Same minor changes as those in the 2D case need to be applied to the original 3D FDM formulas. Mathematical derivations are very similar and thus not given again here.

The example used to demonstrate the quadrupling scheme is a double constant-time 3D HNCO experiment of human ubiquitin. HNCO is also one of the most basic 3D heteronuclear experiments for obtaining the protein backbone chemical shift

[6] In 3D NMR, in order to obtain absorption lineshapes in all three dimensions in the 3D FT spectra, either cosine/sine or N-/P-type signals are acquired for both indirect dimensions, leading to $2 \times 2 = 4$ 3D data sets in total.

assignments. It correlates the chemical shifts of $(^{15}\mathrm{N})\mathrm{H}$, $^{15}\mathrm{N}$ and $^{13}\mathrm{C(O)}$ of previous residue, leading to nice singlets at the chemical shifts of $^{1}\mathrm{H}_i^N$-$^{15}\mathrm{N}_i$-$^{13}\mathrm{C(O)}_{i-1}$. HNCO experiments have good sensitivity and are thus chosen for this demonstration. The standard (single) CT-HNCO [127, 11] pulse sequence was slightly modified [76] by implementing the CT evolution to encode the CO chemical shift information. The signals were all acquired on a 500 MHz spectrometer equipped with a normal probe at room temperature. 8 scans were acquired per increment. The sample was 1 mM uniformly labeled human ubiquitin. States-TPPI was implemented in the CO dimension and gradient select was used in the $^{15}\mathrm{N}$ dimension. The constant-time period was fixed at the standard value of $2T = 48$ ms for the $^{15}\mathrm{N}$ dimension to maximize the polarization transfer. A very short constant time period of $2T = 10$ ms was used in the $^{13}\mathrm{C(O)}$ dimension, to minimize the intensity lost due to the transverse relaxation. A maximum of 6 $^{13}\mathrm{C(O)}$ increments can be acquired with a spectral width of 1 KHz. The spectral width was 2 KHz in the $^{15}\mathrm{N}$ dimension, and, in principle, up to 48 increments could be acquired. However, only 8 increments were actually acquired in the $^{15}\mathrm{N}$ dimension to minimize the experimental time, and also to study the low limit of signal size required for FDM to fully resolve all the resonances for the system under study. The spectral width in proton dimension was 10 KHz and 1024 complex points were acquired. At the end, a 3D signal of size $(H, C, N) = 1024 \times 6 \times 8$ (for each of four data sets) was acquired in about 25 minutes.

The 3D FT spectrum were obtained using cosine weighting functions along all dimensions. The signal was zero-filled to $128 \times 128$ in the $^{13}\mathrm{CO}$ and $^{15}\mathrm{N}$ dimensions to provide sufficient digital resolution for smooth contours to be drawn. The FDM

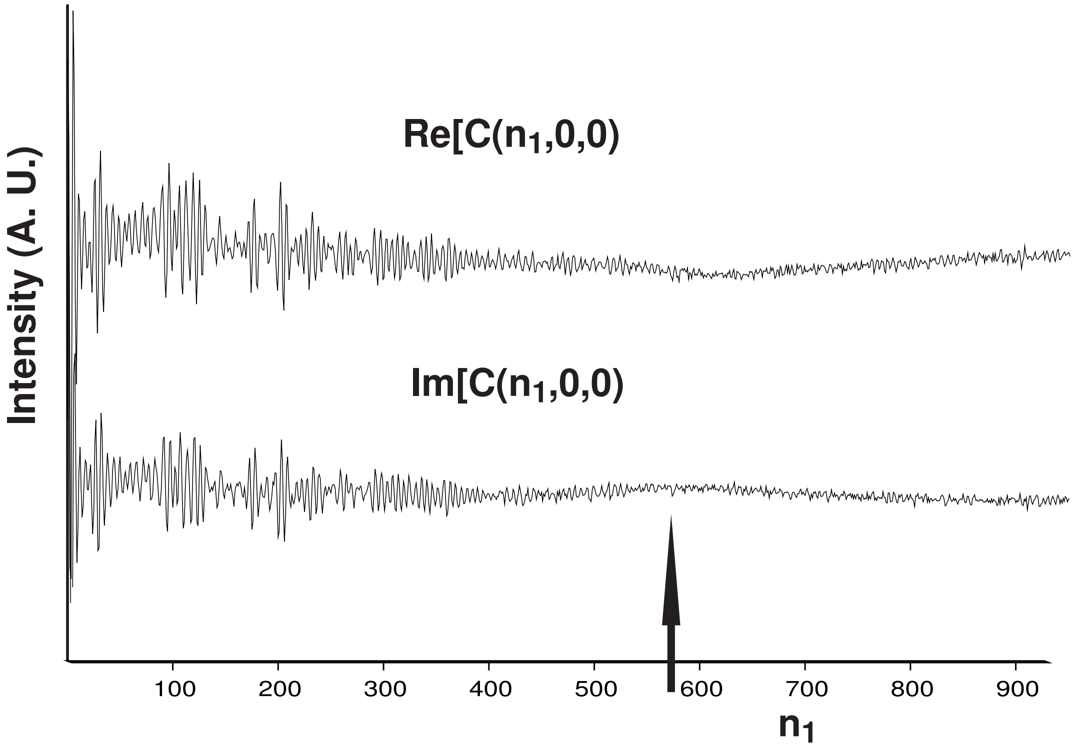


Figure 3.17: The real and imaginary parts of the first FID of the 3D HNCO signal used in Figures 3.18-3.20. Note that the signal already decays to below noise level by about half of the full length. The tail part of the signal contains mainly noise and should not be include in FDM calculation.

spectra were obtained using the FDM2K algorithm with the *quadruple scheme*. Regularization was optimized, with the final regularization level $q^2 = 10^{-4}$ (see Eq. 2.83). As the signal was very short in the indirect dimensions, the 3D spectral windows were chosen to cover the whole spectral ranges, leading to a basis size of $5 \times 7$ in the (C,N) dimensions[7]. In the proton dimension, fully decayed signals were acquired and thus the tail parts were mainly noise, as shown in Figure 3.17. Acquiring extra points in the time domain helps to improve the digital resolution without sacrificing sensitivity for FT with proper apodizations. However, the noisy tail part of the sig-

[7]With doubling scheme implemented in both (C,N) dimensions, a combined signal with $11 \times 15$ (C,N) increments results. Only even number of increments can be used by FDM, $N_l = 2\,M_l$. The maximum basis size is thus $M_2 \times M_3 = 5 \times 7$.

nal does not contain any meaningful information and should not be included in FDM calculations. There are several severe drawbacks for including the noisy tail in the FDM calculations. First, it introduces more noise to the $\mathbf{U}$ matrices, making them more ill-conditioned. Second, it increases the total basis size that has to be used, making the problem even more overdetermined. Third, it increases computational cost. Therefore, in practice, it is necessary to plot out, typically, the first FID for choosing the optimal signal size to be used in the proton dimension. For example, simply by looking at the plot of Figure 3.17, one can easily estimate that 500-600 data points should be using in proton dimension. When the total size the multi-dimensional signal is sufficiently large, the FDM results are typically insensitive to the actual number of proton increments used, as long as it is in a reasonable range. This can be used as an indication of the convergence of FDM calculations. The 2D projections shown in Figure 3.18-3.20 were computed using 500 increments in proton dimension. Thus the maximum basis size in proton dimension for the region shown was $Nb \sim 500/2 \times 2000Hz/SW = 50$. Four overlapping windows with 20 basis functions per window were actually used. The total basis size for each window was $20 \times 5 \times 7 = e700$. The whole calculation took about 40 minutes on a small PC Linux workstation with an AMD XP 1800+ CPU. Calculations were also carried out using slightly different numbers of proton increments between 400 and 600 data points. They produced very similar spectral estimations, showing that FDM already converged. As there is no direct way to plot the 3D spectra, the 2D projections along three planes, namely, (H,C), (H,N), and (C,N), are shown instead in Figures 3.18-3.20. All the FDM projections shown are properly smoothed in the CT dimensions for

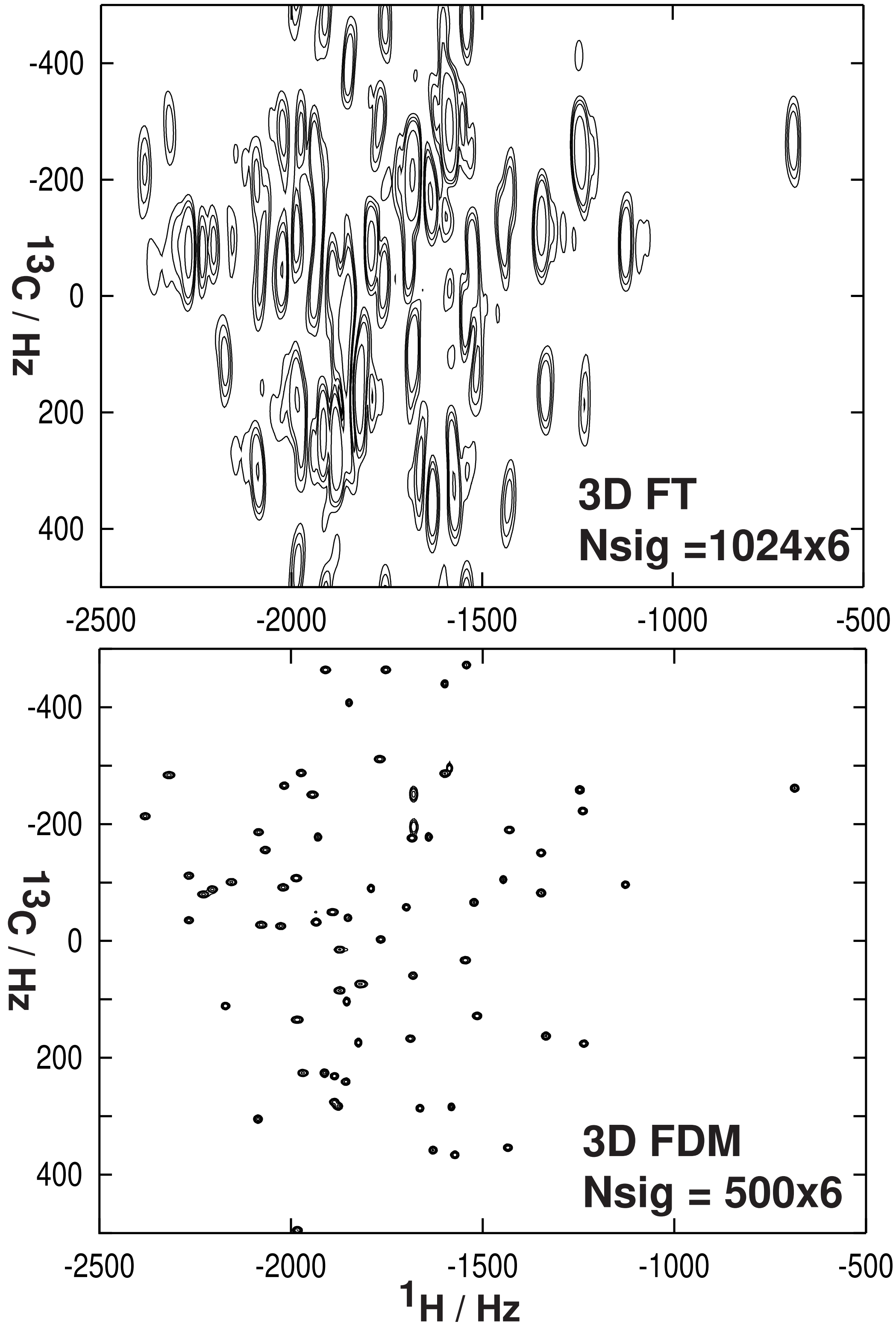



Figure 3.18: Comparison of 2D projections of the 3D FT and 3D FDM spectra in the HC plane, obtained from a short 3D double constant-time HNCO signal (see text for details about the experiment and data processing).

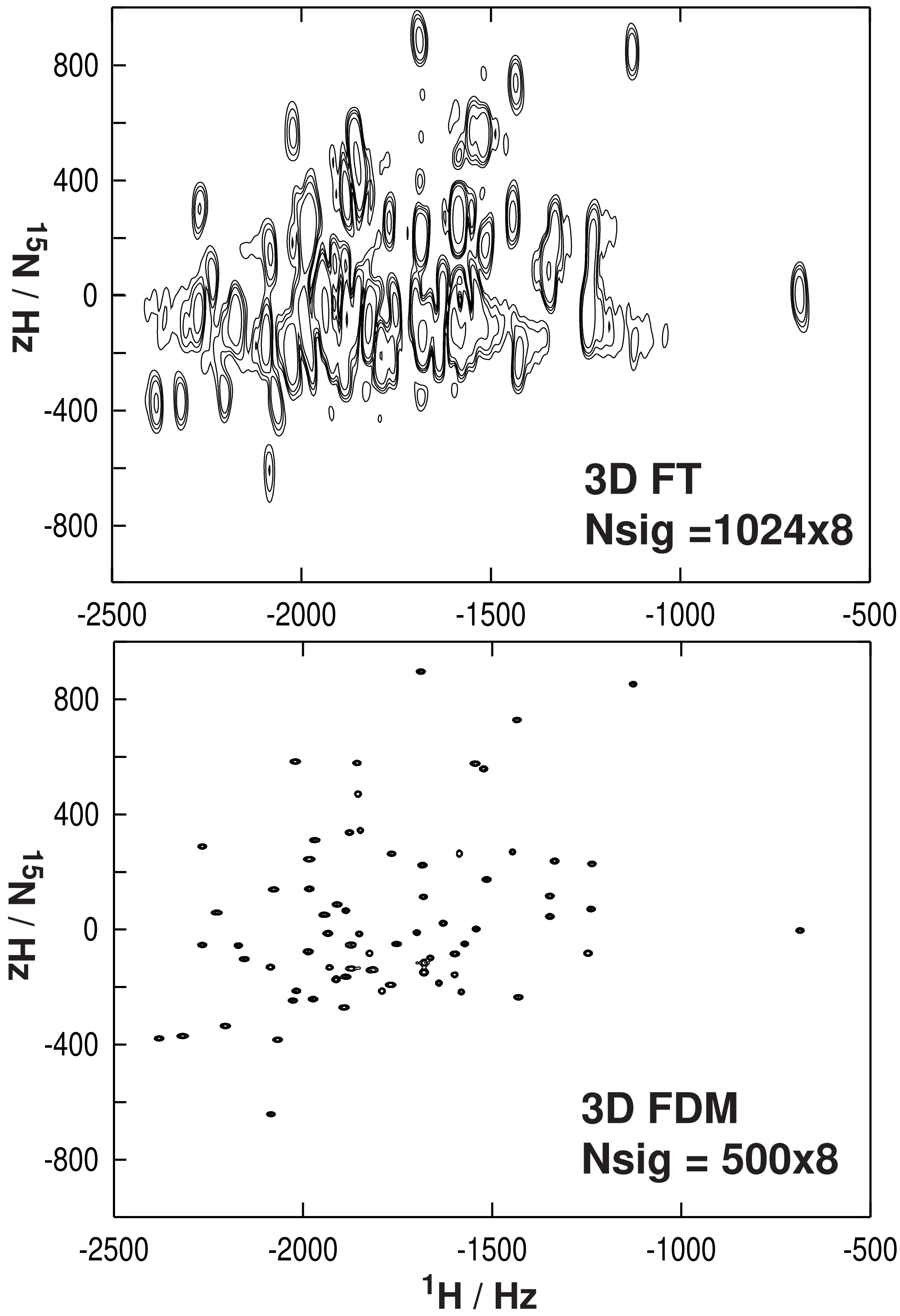



Figure 3.19: Comparison of 2D projections of the 3D FT and 3D FDM spectra in the HN plane, obtained from a short 3D double constant-time HNCO signal.

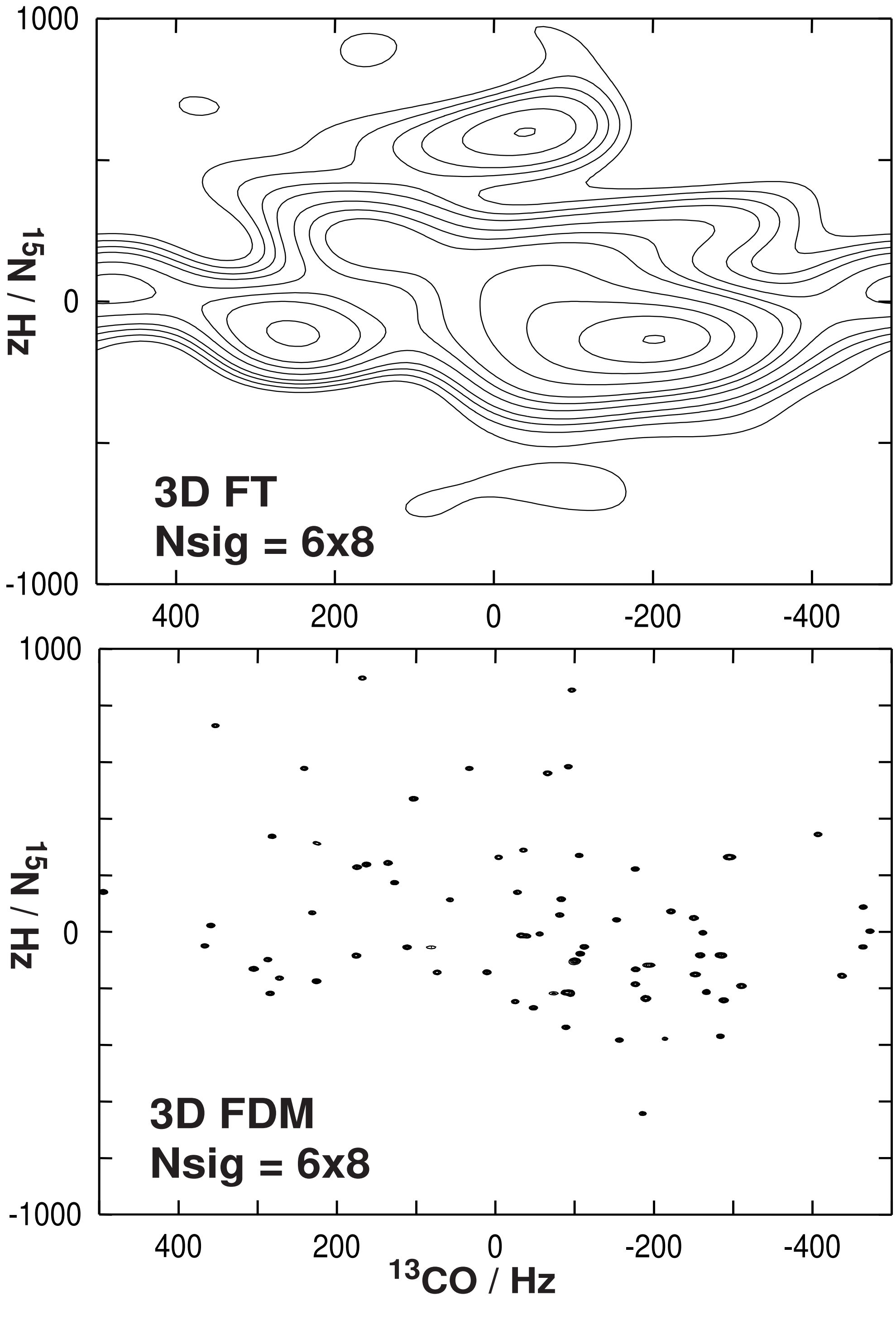



Figure 3.20: Comparison of 2D projections of the 3D FT and 3D FDM spectra in the CN plane, obtained from a short 3D double constant-time HNCO signal. There are only $6 \times 8$ points in this plane, making it the most troublesome plane for FT. However, FDM, as a true 3D method, is able to use the information from the whole 3D data set and provide good resolutions in all dimensions, including these two extremely short dimensions. The resolution enhancement is truly astonishing!

better plotting. The Lorentzian lineshapes were converted to the Gaussian lineshapes according to the procedure described previously (Eq. 3.27).

Figure 3.18 compares the 2D FT and FDM projections in the (H,C) plane. We can clearly see the effects of the FT time-frequency uncertainty principle applying to each dimensions independently: the resolution is high in the proton dimension where long signal exists, while the carbon dimension is poorly resolved as there are only 6 increments. The overall resolution of FT projection is very poor, making it useless for obtaining chemical shift assignments. However, the FDM projection, obtained using the same data set, shows extremely high resolutions in both dimensions. All the peaks are fully resolved and can be easily assigned. This is possible because FDM is a multidimensional method. The resolutions in all dimensions are determined together by the total information content of the multidimensional signal. Even though there are only 6 increments in the $^{13}$C(O) dimension, the total size of the signal is sufficiently large for FDM to pin down all the 3D features and thus very high resolutions can be obtained in both the proton and the $^{13}$C(O) dimensions. Similar pictures are seen in Figure 3.19, where we compare the FT and FDM projections in the (H,N) plane. What is truly astonishing, at least to those who are used to the FT spectral analysis, is the FDM projection in the (C,N) plane, shown in Figure 3.20. In this plane, there are only $6 \times 8$ increments. There is almost no usable information in the FT projection. Surprisingly, and not surprisingly, the FDM projection in (C,N) plane are just as highly resolved as projections along the other two planes, demonstrating the substantial advantage of true multidimensional data processing. It is predictable that any methods that fail to utilize this advantage could not possibly obtain comparable results to what has been shown here. What should also be emphasized are the two other essential properties of FDM: solving the nonlinear fitting problem by linear algebra and local spectral analysis via Fourier basis. Without these two properties, the scale and complexity of the problem would make any methods practically infeasible. This is one of the main reasons why methods like, Maximum Entropy Reconstruction, even though a multi-dimension method, relying on nonlinear optimizations and attempting to fit the whole signal in a single shot, could not be used to analyze realistic data sets. It should be pointed out that 3D FDM is the first true three-dimensional method that has ever been successfully applied to analyze experimental NMR signals and produce meaningful results.

**Validity of the FDM Results**

Unfortunately, it is difficult to directly check the validity of the FDM results, partially due to the lack of comparable references. However, there are various ways to validate the results indirectly, as discussed in Section 2.6. We have already checked that the results were stable with respect to varying signal size in the proton dimension, and other FDM parameters such as basis size, basis density and regularization level. Here we present some additional results using the checking mechanism described and demonstrated in Section 2.6.4.

After all the spectral parameters $\{D_{k,k',k''}, \omega_{1k}, \omega_{2k''}, \omega_{3k''}\}$ are computed, in addition to estimating the infinite time DFT spectra, which have been shown in Figure 3.18-3.20, we can also estimate various finite time DFT spectra using,

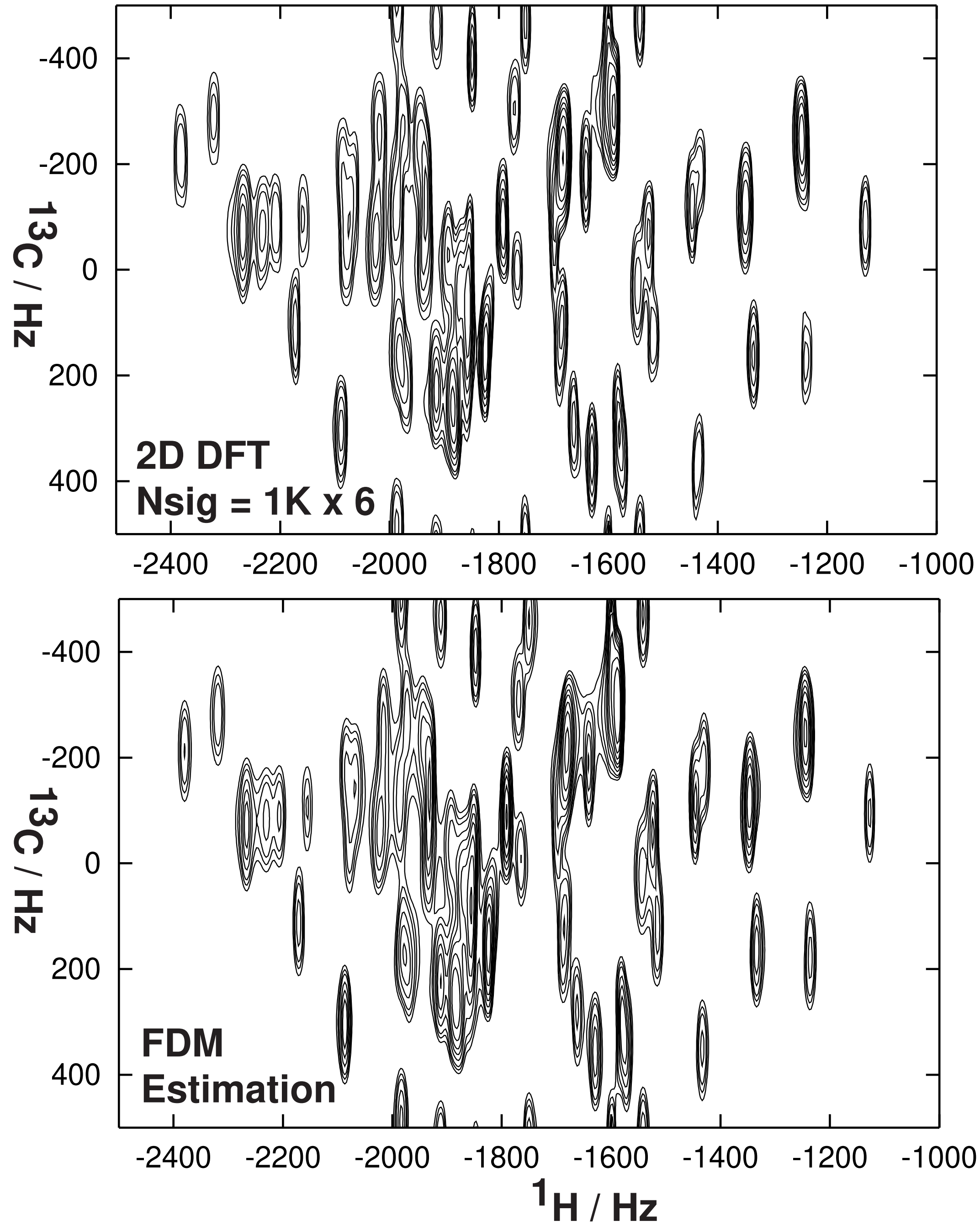



Figure 3.21: Comparison of the FDM estimation of the finite (H,C) FT projection with the actual (H,C) FT projection obtained from the original data. FDM estimation was computed using Eq. 3.28 with $(\alpha, \beta) = (1, 2)$; $N_1 = 1024, N_2 = 6$ and the spectral parameters $\{D_{k,k',k''}, \omega_{1k}, \omega_{2k''}, \omega_{3k''}\}$ computed from the short constant-time 3D HNCO signal (see text for details). The FDM estimation agrees with the actual FT projection very well, proving that FDM does provide a reasonably good fit of the 3D signal. Similar comparison was also made along (H,N) plane, which also shows a very good match (data not shown).

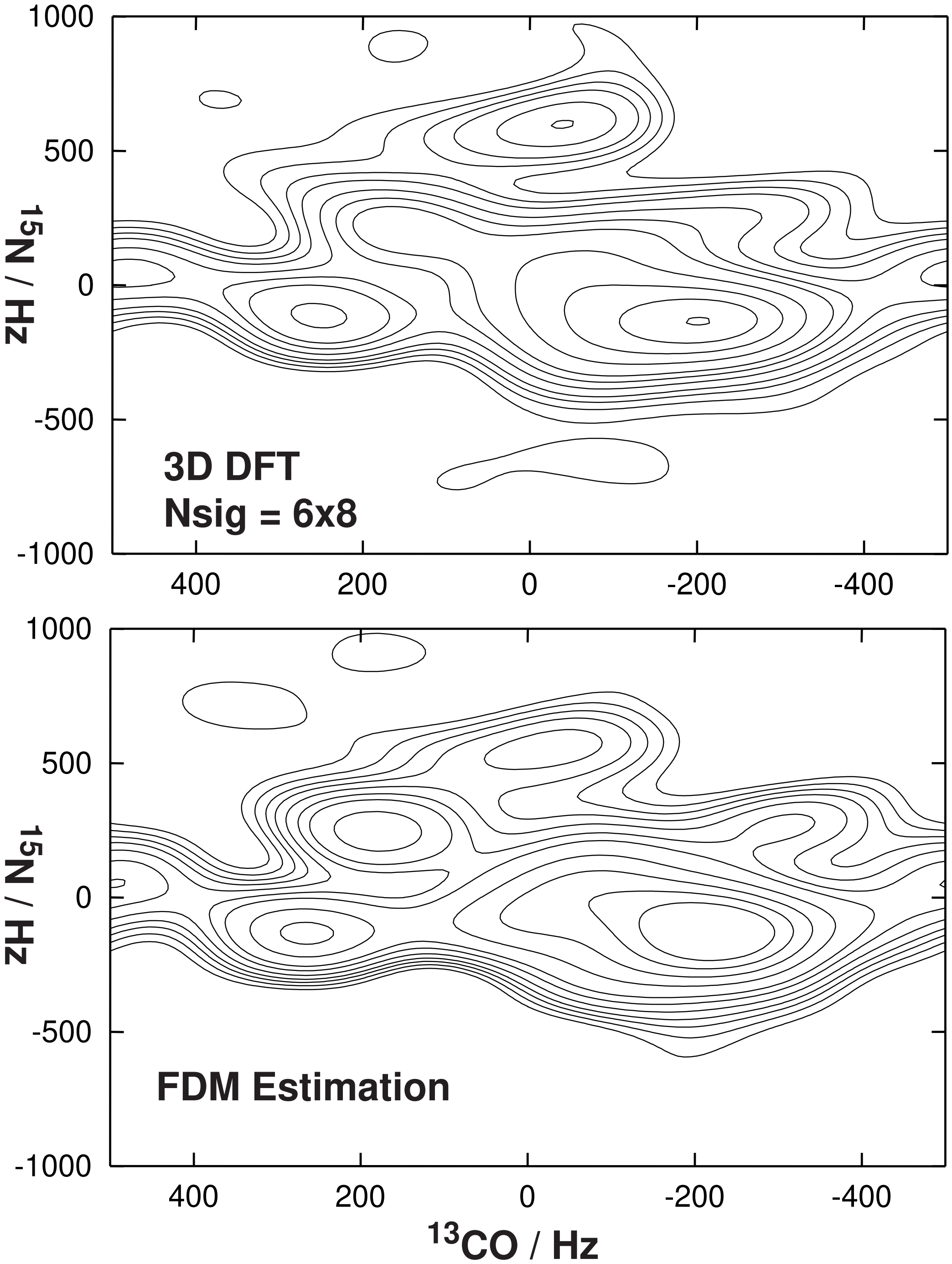
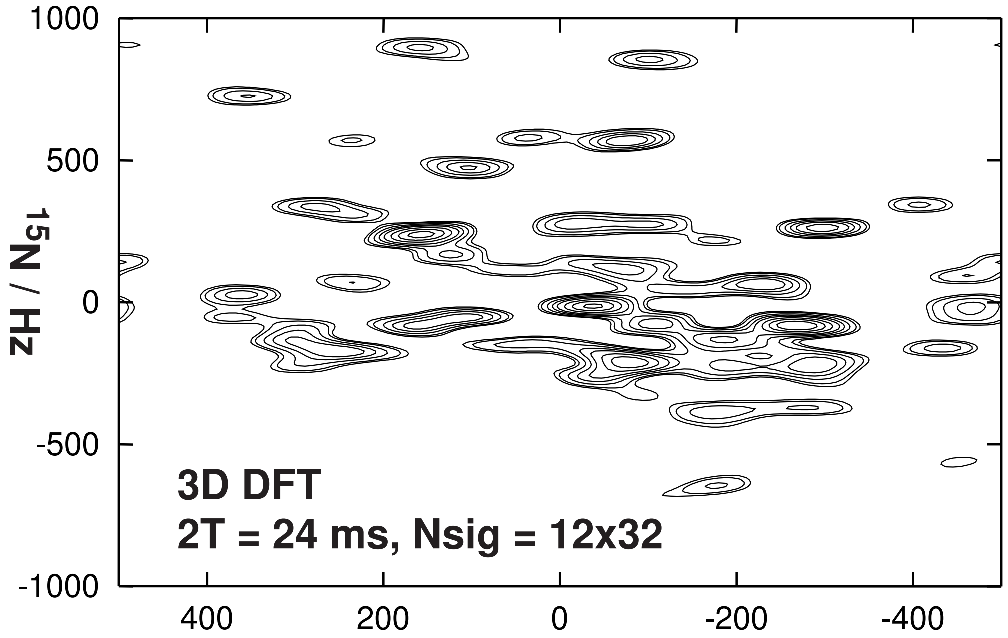



Figure 3.22: Comparison of the FDM estimation of the finite (C,N) FT projection with the actual FT projection obtained from the original data. FDM estimation was computed using Eq. 3.28 with $(\alpha, \beta) = (2, 3)$; $N_2 = 6, N_3 = 8$ and using the spectral parameters computed from the short constant-time 3D HNCO signal. It appears that the matching is reasonable but not as good as in (H,C) plane. This is because water signal leaks into all (C,N) planes in the FT spectra, while such a leak is totally eliminated in FDM (see text).



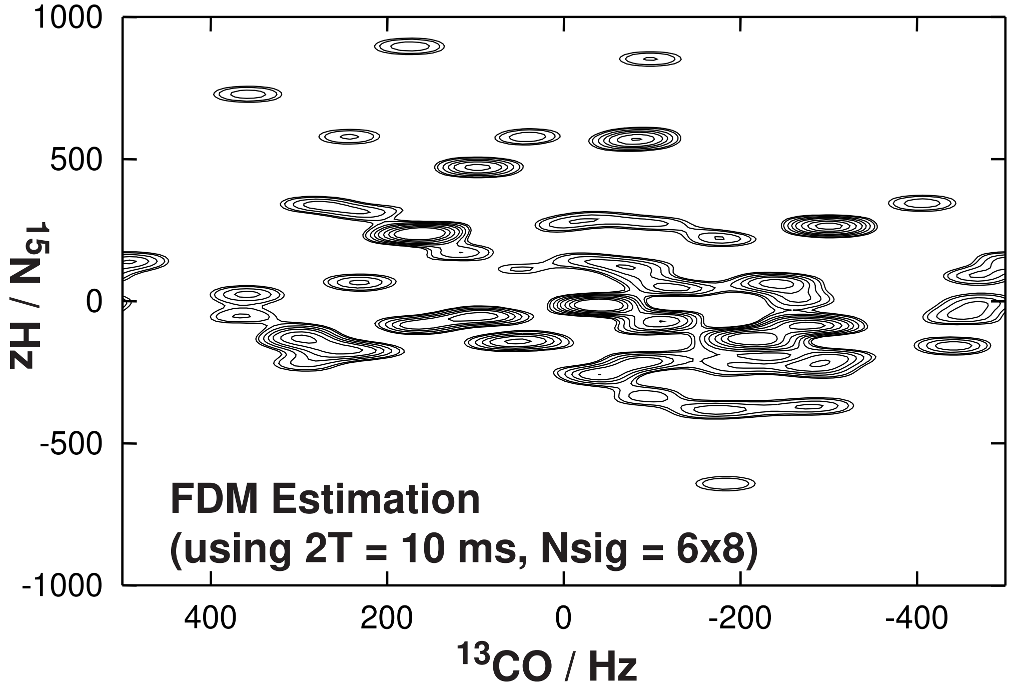




Figure 3.23: Comparison of the estimated "high-resolution" FT projection with the actual FT projection obtained from a long experimental signal with $12 \times 32$ increments in (C,N) dimensions. FDM estimation was computed using expression Eq. 3.28 with $(\alpha, \beta) = (2, 3); N_2 = 12, N_3 = 32$ but using the spectral parameter that were computed from the short signal with only $6 \times 8$ (C,N) increments. The matching is still very well, further proving that FDM does provide a very accurate and reliable fit of the 3D signal. Note that the intensities in the actual FT projection is lower due to the relaxation during the longer CO constant time ($2T = 24ms$ vs. $2T = 10ms$).

$$
\begin{aligned}
A^{N_\alpha,N_\beta}(\omega_\alpha,\omega_\beta) \quad &\approx \quad \tau_1\tau_2\tau_3 \sum_{k,k',k''} \mathrm{Re}[D_{k,k',k''}] \\
&\mathrm{Re}\left[\frac{1-e^{iN_\alpha\tau_\alpha(\omega_\alpha-\omega_{\alpha k})}}{1-e^{i\tau_\alpha(\omega_\alpha-\omega_{\alpha k})}}-\frac{1}{2}\right]\mathrm{Re}\left[\frac{1-e^{iN_\beta\tau_\beta(\omega_\beta-\omega_{\beta k'})}}{1-e^{i\tau_\beta(\omega_\beta-\omega_{\beta k'})}}-\frac{1}{2}\right],
\end{aligned}
\tag{3.28}
$$

where $(\alpha, \beta)$ can be set to (1,2), (1,3), and (2,3) to compute the estimated finite FT projections along (H,C), (H,N) and (C,N) planes respectively. Apodization functions such as exponential and trigonometric functions can be also included analytically into Eq. 3.28. All the results shown in Figures 3.21-3.23) were obtained with cosine window functions in all dimensions, implemented either numerically for the actual DFT calculations or analytically for computing the FDM estimations.

Figure 3.21 compares the FDM estimation of the finite FT projection with $1024\times6$ increments in (H,C) plane, to the actual FT projection obtained from the experimental data of the same size. The FDM estimation was computed using the spectral parameters obtained from the short HNCO signal with the same processing parameters used to produce Figures 3.18-3.20. The estimated and actual projections match with each other very well, proving that FDM does provide a good fit of the experimental signal. Similar comparison was also done in (H,N) plane, and the matching quality was similar (data not shown). Quantitatively, the relative matching errors, as defined in Eq. 2.111, are only 1.8% and 1.7% respectively, which are below the noise level. Figure 3.22 compares the FDM estimation of the finite FT projection along (C,N) plane with $6\times8$ increments to the actual FT projection. The mismatch is much higher than those in the other two planes, with relative matching error being 7.5%. However, this increase of mismatch does not mean that the parameters in (C,N) dimensions are less accurate. Instead, most of the increased mismatch is due to the water signal that leaks to every (C,N) plane in 3D FT spectrum[8], while for FDM, the use of the Fourier basis followed by diagonalization is very efficient in suppressing the interference of the features outside of the spectral window. To further support the quality of FDM fit, we used the same spectral parameters, that were computed from the short HNCO signal with $6 \times 8$ (C,N) increments, to estimate a "high-resolution" FT projection corresponding $12 \times 32$ (C,N) increments, and compared it with the actual FT projection obtained from a much longer experiments, where the constant time period in $^{13}$C(O) dimension was $2T = 24ms$ and $12 \times 32$ (C,N) increments were actually acquired. The results are shown in Figure 3.23. Due to the transverse relaxation, the intensities in the actual FT projection are lower. Otherwise, the two projections match with each other very well, providing a strong evidence that the fit computed by FDM is an accurate and reliable fit of the 3D signal.

## 3.5 Summary and Remarks

In this chapter, we applied the FDM and RRT algorithms to processing in 1D and multidimensional NMR signals.

Long time signals are typically available in 1D NMR and high resolution FT spectra can be obtained, leaving little room for resolution enhancement. However, the conventional FT spectral analysis can still have problems in certain cases. For example, large linear phase corrections give rise to severe phase rolls in the baseline

[8]The water signal can be actually seen in th plot of first FID as low frequency oscillations, as shown in Figure 3.17.

of the FT spectra, because of the point-by-point fashion of linear phase correction in FT. FDM and RRT can recognize the individual peaks and correct the linear phase in a more consistent way. The phase roll due to point-by-point phase correction is completely removed. For noisy signal with heavily overlapped features and strong background, there are still some small instabilities problems in the FDM/RRT linear phase corrected spectra, which can be further suppressed by averaging.

Multidimensional NMR signals are typically truncated in the indirect dimensions due to practical limitations including expensive experimental time and limited long-term instrumental stability, resulting in poor FT resolutions in the corresponding frequency dimensions. FDM is a true multidimensional method and can have substantial advantages over FT in these cases. FDM is capable of analyzing the whole data set to pin down the intrinsically multidimensional features. The obtainable resolutions in all dimensions are determined together by the total information content, or, roughly speaking, the total size, of the signal. Multidimensional NMR signals typically have a large total size due to the long acquisition time dimension and thus contain a lot of information that can be used by multidimensional methods like FDM and RRT to significantly enhance the resolutions in the truncated dimensions.

Nontrivial projections can be constructed by FDM, among which the $45^o$ projections of $J$-experiments are particularly useful. By constructing $45^o$ projections, homonuclear coupling are decoupled and intensive singlets result. FDM has been successfully applied to various $J$-experiments including 2D-J, 3D HSQC-J and 4D J-TOCSY-J. In J-TOCSY-J, double $45^o$ projections are used to decouple the homonuclear coupling in both proton dimensions. So called *singlet*-TOCSY spectra can be obtained using as few as one additional increment in each $J$-dimension. However, application of the method to simplifying highly complex spectra has difficulties with degeneracy and others, and needs more investigations.

In this chapter, we have been focusing on applying FDM to analyzing a special type of NMR experiments that utilized Constant-Time (CT) evolutions to encode the chemical shift information in the indirect dimensions. There are some special properties of CT signals that make them particularly suitable for FDM data processing. Firstly, there exists an efficient doubling scheme for CT signals. A pair of N- and P-type data sets from the same 2D CT NMR experiment are processed jointly by FDM as a single data set, twice as large, in which the signal effectively evolves in time for twice as long. Secondly, the signal has nearly perfect Lorentzian lineshape in the CT dimension, thus can be efficiently handled by FDM. Applied to both model and experimental signals, FDM with doubling scheme shows significant resolution improvement and appears to tolerate the noise well. Very short constant time period can be used without sacrificing the resolution, leading to higher sensitivity, especially for proteins with short transverse relaxation time constants.

The doubling scheme can be applied to each CT dimension in higher dimensional NMR experiments with multiple CT dimensions. When CT evolution is implemented in all indirect dimensions, all the $2^{N-1}$ N-dimensional data sets from the same N-dimensional experiment can be processed jointly as a single data set by FDM to extract maximum information. This was demonstrated by a quadruple scheme for a 3D HNCO experiment with two CT dimensions. By processing all four 3D data sets from the same experiment together, FDM was able to provide a 3D HNCO spectrum of protein human ubiquitin that are fully resolved in all dimensions, even when there are only as few as $6 \times 8$ increments in the $^{13}$C(O) and $^{15}$N dimensions. The results were validated extensively by various approaches including: (1) repeating the FDM calculations with a wide range of different parameters such as signal size, basis size, basis density and regularization level; (2) comparing the FDM estimations of low-resolution and high-resolution finite FT spectra, using the spectral parameters computed from the short signal, to the actual FT spectra obtained from both short and long experimental signals. They all showed that the FDM spectrum already converged and that an accurate, reliable fit of the 3D signal was obtained by FDM.

Optimization of N-dimensional NMR experiments has always presupposed that FT will be used to obtain the $N$-dimensional spectrum. For example, 3D NMR is quite often employed because decent digital resolution can be obtained without extravagant experiment time. As such, a series of different 3D experiments may be used to correlate all the desired chemical shifts, and comparing planes from experiments taken at different times is usually mandatory to complete the assignment. Whether this is the most profitable use of human and instrument time, once FDM is reduced to practice, is certainly open to question. In FDM the digital resolution of individual dimensions has no direct relation to the ultimate spectral resolution. Further, we have shown here that the best calculation is the one with the largest available basis set that can be obtained from the measured signal. Two different 3D experiments do not lend themselves to this maximal basis set in the same way that one 4D experiment does. While extra delays may lead to some additional losses in the higher-dimensional experiment, there is also a gain in sensitivity by having all the data in one spectrum rather than divided up between several. In 3D experiments where a set of nuclei are merely used to transfer magnetization, for example the HN(CO)CA pulse sequence [140], the delays are already present, and removing the parentheses to record the carbonyl chemical shift in a 4D experiment will be more sensitive and highly advantageous, if FDM is to be used.

In closing, it is interesting to note that the time it takes to identify M random-frequency peaks accurately in an N-dimensional experiment depends on M but not N when FDM is employed. Indeed, as local spectral crowding is the only limiting factor once sufficient SNR is achieved, increasing the dimensionality of the experiment, so that the resonances are dispersed as widely as possible in frequency space, may result in a shorter overall experiment. For backbone assignment experiments on proteins with known primary structure, the expected value of M is known beforehand, so that the time required to obtain a converged high-dimensional FDM spectrum should be predictable, and scale with M. With the SNR of the first increment and the value of M in hand, it may be possible to semi-automate the data acquisition of even quite complex experiments. This assertion should be tested in a future series of 3D, 4D, and 5D experiments that are specifically designed with the method in mind.

# Bibliography


[1] P. J. Robinson and K. A. Holbrook. *Unimolecular Reactions.* Wiley, New York, 1972.

[2] D. H. Levy. Laser spectroscopy of cold gas-phase molecules. *Annu. Rev. Phys. Chem.*, 31:1971–1999, 1980.

[3] R. Roy, B. G. Sumpter, G. A. Preffer, S. K. Gray, and D. W. Noid. Novel methods for spectral analysis. *Physics Reports*, 3:109–152, 1991.

[4] J. B. J. Fourier. *Theorie analytique de la chaleur.* Chez Firmin Didot, Paris, France, 1822.

[5] A. G. Marshall and F. R. Verdun. *Fourier Transforms in NMR, Optical, and Mass Spectrometry.* Elsevier, New York, 1990.

[6] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19:297–301, 1965.

[7] E. O. Brigham. *The Fast Fourier Transform.* Prentice Hall, Englewood Cliffs, New Jersey, 1974.

[8] J. C. Hoch and A. S. Stern. *NMR Data Processing.* Wiley-Liss, New York, 1996.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipies in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, UK, second edition, 1992.

[10] A. Bax, M. Ikura, L. E. Kay, and G. Zhu. Removal of F1-base-line distortion and optimization of folding in multidimensional NMR-spectra. *J. Magn. Reson.*, 91:174–178, 1991.

[11] J. Cavanagh, W. J. Faribrother, A. G. Palmer III, and M. J. Skelton. *Protein NMR Spectroscopy: principles and practice.* Academic Press, San Diego, 1996.

[12] M. Frigo and S. G. Johnson. FFTW: an adaptive software architechure for the FFT. In *ICASSP Conference Proceedings*, volume 3, pages 1381–1384, 1998. WWW: http://www.fftw.org.

[13] H. A. Kramers. *Nature*, 117:775, 1926.

[14] R. De L. Kronig. *J. Opt. Soc. Am.*, 12:547, 1926.

[15] E. Bartholdi and R. R. Ernst. Fourier spectroscopy and causality principle. *J. Magn. Reson.*, 11:9–19, 1973.

[16] Marple S. L. *Digital Signal Processing.* Prentice Hall, Englewood Cliffs, New Jersey, 1987.

[17] Baron Gaspard Riche de Prony. Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, á différentes températures. *Journal de l'École Polytechnique*, 1(cahier 22):24–76, 1795.

[18] H. Barkhuijsen, D. Debeer, W. M. M. J. Bovee, and D. Van Ormondt. Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure. *J. Magn. Reson.*, 61:465–481, 1985.

[19] D. S. Stephenson. Linear prediction and maximum-entropy in NMR-spectroscopy. *Prog. NMR. Spec.*, 20:515–626, 1988.

[20] P. Koehl. Linear prediction spectral analysis of NMR data. *Prog. NMR. Spec.*, 34:257–299, 1999.

[21] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped undamped sinusoids in noise. *IEEE Trans. ACOUST. SPEECH*, 38:814–824, 1990.

[22] Y. Hua and T. K. Sarkar. Estimating tow-dimensional frequencies by matrix enhancement and matrix pencil. *IEEE Trans. Signal Process.*, 40:2267–2280, 1992.

[23] Y. Y. Lin, P. Hodgkinson, M. Ernst, and A. Pines. A novel detection-estimation scheme for noisy NMR signals: applications to delayed acquisition data. *J. Magn. Reson.*, 128:30–41, 1997.

[24] 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.

[25] V. A. Mandelshtam and H. S. Taylor. A low-storage filter diagonalization method for quantum eigenenergy calculation or for spectral analysis of time signals. *J. Chem. Phys.*, 106:5085–5090, 1997.

[26] G. L. Bretthorst. Bayesian-analysis .1. parameter-estimation using quadrature NMR models. *J. Magn. Reson.*, 88:533–551, 1990.

[27] G. L. Bretthorst. Bayesian-analysis .2. signal-detection and model selection. *J. Magn. Reson.*, 88:552–570, 1990.

[28] G. L. Bretthorst. Bayesian-analysis .3. applications to nmr signal-detection, model selection, and parameter-estimation. *J. Magn. Reson.*, 88:571–595, 1990.

[29] R. A. Chylla and J. L. Markley. Theory and application of the maximum-likelihood principle to NMR parameter-estimation of multidimensional NMR data. *J. Biomol. NMR*, 5:245–258, 1995.

[30] H. Gesmar, J.J. Led, and F. Abildgaard. Improved methods for quantitative spectral-analysis of NMR data. *Prog. NMR. Spec.*, 22:255–288, 1990.

[31] J. W. C. van der Veen, R. de Beer, P. P. Luyten, and D. van Ormondt. Accurate quantification of *in vivo* $^{31}$p NMR signals using the variable projection method and prior knowledge. *Magn. Reson. Med.*, 6:92–98, 1988.

[32] H. Barkhuijsen, R. Debeer, W. M. M. J. Bovee, J. H. N. Creyghton, and D. Van Ormondt. Application of linear prediction and singular value decomposition (LPSVD) to determine nmr frequencies and intensities from the FID. *Magn. Reson. Med.*, 2:86–89, 1985.

[33] J. Tang, C. P. Lin, M. K. Bowman, and J. R. Norris. An alternative to fourier-transform spectral-analysis with improved resolution. *J. Magn. Reson.*, 62:167–171, 1985.

[34] C. F. Tirendi and J. F. Martin. Quantitative-analysis of NMR-spectra by linear prediction and total least-squares. *J. Magn. Reson.*, 85:162–169, 1989.

[35] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, AC-19:667–674, 1974.

[36] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust. Speech*, 33:387–392, 1985.

[37] M. Wax. Oder selection for AR models by predictive least squares. *IEEE Trans. Acoust. Speech*, 36:581–588, 1988.

[38] M. Wax and I. Ziskind. Detection of the number of coherent signals by the MDL principle. *IEEE Trans. Acoust. Speech*, 37:1190–1196, 1989.

[39] R. Roy, A. Paulraj, and T. Kailath. ESPRIT - a subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE Trans. ACOUST. SPEECH*, 34:1340–1342, 1986.

[40] F. Abildgarrd, H. Gesmar, and J. J. Led. Quantitative-analysis of complicated nonideal fourier-transform NMR-spectra. *J. Magn. Reson.*, 79:78–89, 1988.

[41] S. Sibisi. *"Quantified MaxEnt: An NMR application", in Maximum Entropy and Baysian Methods*, pages 351–358. Kluwer Academic, Dordrecht, 1990.

[42] J. Capon. High resolution frequency wave number spectrum analysis. *Proc. IEEE*, 57:1408–1418, 1969.

[43] P. Stoica and R. Moses. *Introduction to spectral analysis*. Prentice Hall, Uper Sadder River, New Jersey, 1997.

[44] J. Li and P. Stoica. Adaptive filtering approach to spectral estimation and SAR imaging. *IEEE Trans. Signal Process.*, 44:1469–1484, 1996.

[45] S. Sibisi. Two-dimensional reconstructions from one-dimensional data by maximum-entropy. *Nature*, 30:134–136, 1983.

[46] S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue, and J. Staunton. Maximum-entropy signal-processing in practical NMR-spectroscopy. *Nature*, 311:446–447, 1984.

[47] E. D. Laue, J. Skilling, J. Staunton, S. Sibisi, and R. G. Brereton. Maximum-entropy method in nuclear magnetic-resonance spectroscopy. *J. Magn. Reson.*, 62:437–452, 1985.

[48] P. Schmieder, A. S. Stern, G. Wagner, and J. C. Hoch. Quantification of maximum-entropy spectrum reconstructions. *J. Magn. Reson.*, 125:332–339, 1997.

[49] J. A. Jones and P. J. Hore. The maximum entropy method. appearance and reality. *J. Magn. Reson.*, 92:363–367, 1991.

[50] J. Tang and J. Norris. LP-ZOOM, a linear prediction method for local spectral analysis of nmr signals. *J. Magn. Reson.*, 79:190–196, 1988.

[51] D. Neuhauser. Bound state eigenfunctions from wave packets - time-energy resolution. *J. Chem. Phys.*, 93:2611–2616, 1990.

[52] V. A. Mandelshtam, T. P. Grozdanov, and H. S. Taylor. Bound states and resonances of the hydroperoxyl radical $ho_2$. an accurate quantum mechanical calculation using filter diagonalization. *J. Chem. Phys.*, 103:10074–10084, 1995.

[53] V. A. Mandelshtam and H. S. Taylor. Harmonic inversion of time signals and it's applications. *J. Chem. Phys.*, 107:6756–6769, 1997.

[54] J. Main, V. A. Mandelshtam, and H. S. Taylor. High resolution recurrence spectra: Beyond the uncertainty principle. *Phys. Rev. Lett.*, 78(4351–4354), 1997.

[55] J. Main, V. A. Mandelshtam, and H. S. Taylor. Periodic orbit quantization by harmonic inversion. *Phys. Rev. Lett.*, 79(825–828), 1997.

[56] E. Narevicius, D. Neuhauser, H.J. Korsch, and N. Moiseyev. Resonances from short time complex-scaled cross-correlation probability amplitudes by the filter-diagonalization method. *Chem. Phys. Lett.*, 276:250–254, 1997.

[57] F. Grossmann, V. A. Mandelshtam, H. S. Taylor, and J.S. Briggs. Harmonic inversion of semiclassical short time signals. *Chem. Phys. Lett.*, 279:355–360, 1997.

[58] V. A. Mandelshtam. Harmonic inversion of time cross-correlation functions. the optimal way to perform quantum or semiclassical dynamics calculations. *J. Chem. Phys.*, 108:9999–10007, 1998.

[59] V. A. Mandelshtam and M. Ovchinnikov. Extraction of tunneling splittings from a real-time semiclassical propagation. *J. Chem. Phys.*, 108:9206–9209, 1998.

[60] G. Jeschke, V.A. Mandelshtam, and A.J. Shaka. Pure absorption electron spin echo envelope modulation spectra by using the filter-diagonalization method for harmonic inversion. *J. Magn. Res.*, 137:221–230, 1998.

[61] M.H. Beck and H.-D. Meyer. Extracting accurate bound-state spectra from approximate wave packet propagation using the filter-diagonalization method. *J. Chem. Phys.*, 109:3730–3741, 1998.

[62] R. Chen and H. Guo. Calculation of matrix elements in filter diagonalization: a generalized method based on fourier transform. *Chem. Phys. Lett.*, 279:252–258, 1997.

[63] R. Chen and H. Guo. Symmetry-enhanced spectral analysis via the spectral method and filter diagonalization. *Phys. Rev. E*, 57:7288–7293, 1998.

[64] A.J.R. Da Silva, J. W. Pang, and E. A. Carter. Anharmonic vibrations via filter diagonalization of ab initio dynamics trajectories. *J. Phys. Chem. A*, 102:881–885, 1998.

[65] J.W. Pang, T. Dieckmann, J. Feigon, and D. Neuhauser. Extraction of spectral information from a short-time signal using filter-diagonalization: Recent developments and applications to semiclassical reaction dynamics and nuclear magnetic resonance signals. *J. Chem. Phys.*, 108:8360–8368, 1998.

[66] V. A. Mandelshtam, H. S. Taylor, and A. J. Shaka. Application of the filter diagonalization method to one- and two-dimensional NMR spectra. *J. Magn. Reson.*, 133:304–312, 1998.

[67] 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.

[68] V. A. Mandelshtam and H. S. Taylor. Multidimensional harmonic inversion by filter-diagonalization. *J. Chem. Phys.*, 108:9970–9977, 1998.

[69] V. A. Mandelshtam, H. Hu, and A. J. Shaka. Two-dimensional HSQC NMR spectra obtained using a self-compensating double pulsed field gradient and processed using the filter digonalization method. *Magn. Reson. Chem.*, 36:S17–S28, 1998.

[70] 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.

[71] V. A. Mandelshtam, H. Hu, N. D. Taylor, M. Smith, and A. J. Shaka. Highly resolved double absorption 2D NMR spectra from complex severely truncated 2D phase-modulated signals by filter-diagonalization-averaging method. *Chem. Phys. Lett.*, 305:209–216, 1999.

[72] V. A. Mandelshtam, Q. N. Van, and A. J. Shaka. Obtaining proton chemical shifts and multiplets from several 1D NMR signals. *J. Am. Chem. Soc.*, 120:12161, 1998.

[73] V. A. Mandelshtam. The multidimensional filter diagonalization method. I. theory and numerical implementation. *J. Magn. Reson.*, 144:343–356, 2000.

[74] J. Chen, V. A. Mandelshtam, and A.J. Shaka. Regularization of the filter diagonalization method: FDM2K. *J. Magn. Reson.*, 146:363–368, 2000.

[75] J. Chen, A. A. De Angelis, V. A. Mandelshtam, and A. J. Shaka. Progress on two-dimensional filter diagonalization method. an efficient doubling scheme for two-dimensional constant-time NMR. *J. Magn. Reson.*, 2002. submitted.

[76] J. Chen, D. Nietlispach, V. A. Mandelshtam, and A. J. Shaka. Ultra-high quality HNCO spectra with very short constant times. *in preparation*, 2002.

[77] 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.

[78] V. A. Mandelshtam. The regularized resolvent transform for quantum dynamics calculations. *J. Chem. Phys. A*, 105:2764–2769, 2001.

[79] 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.

[80] C. B. Moler and G. W. Stewart. An algorithm for generalized matrix eigenvalue problems. *SIAM J. Num. Anal.*, 10:241–256, 1973.

[81] J. Demmel and B. Kagstrom. The generalized schur decomposition of an arbitrary pencil a - zb: robust software with error bounds and applications. part i: theory and algorithms. *ACM Trans. Math. Softw.*, 19(2):160–174, 1993.

[82] J. Demmel and B. Kagstrom. The generalized schur decomposition of an arbitrary pencil a - zb: robust software with error bounds and applications. part ii: software and applications. *ACM Trans. Math. Softw.*, 19(2):175–201, 1993.

[83] LAPACK: http://www.netlib.org/lapack.

[84] S. D. Silverstein and M.D. Zoltowski. The mathematical basis for element and fourier beamspace MUSIC and root-MUSIC algorithm. *Digital Signal Process.*, 1:161–175, 1991.

[85] W. P. Aue, E. Bartholdi, and P. R. Ernst. Two-dimensional spectroscopy. application to nuclear magnetic resonance. *J. Chem. Phys.*, 64:2229–2246, 1976.

[86] J. Jeener, B. H. Meier, P. Bachmann, and R. R. Ernst. Investigation of exchange process by 2-dimensional nmr-spectroscopy. *J. Chem. Phys.*, 71:4546–4553, 1979.

[87] Braunschweiler and R. R. Ernst. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.*, 53:521–528, 1983.

[88] G. Bodenhausen, R. Freeman, R. Niedermeyer, and D. L. Turne. Double fourier transformation in high-resolution NMR. *J. Magn. Reson.*, 26:133–164, 1977.

[89] B. F. Volkman, A. M. Prantner, S. J. Wilkens, B. Xia, and J. L. Markeley. Assignment of h-1, c-13, n-15 signals of oxidized clostridum pasteurianum rubredoxin. *J. Biomol. NMR*, 10:409–410, 1997.

[90] H. Hu, A. A. De Angelis, V. A. Mandelshtam, and A. J. Shaka. The multidimensional filter diagonalization method. II. applications to 2D projections of 2D, 3D, and 4D NMR experiments. *J. Magn. Reson.*, 144:357–366, 2000.

[91] V. A. Mandelshtam. FDM: The filter diagonalization method for data processing in NMR experiments. *Prog. NMR. Spec.*, 38:159–196, 2001.

[92] G. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, third edition, 1996.

[93] A. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4:1035–1038, 1963.

[94] A. Neumaier. Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM Review*, 40:636–666, 1998. WWW: http://www.mat.univie.ac.at/∼neum/papers.html#regtutorial.

[95] A. Neumaier. The overdetermined eigenvalue problem. private communication.

[96] J. E. Curtis. *Spectral Processing Methods to Extract Imaginary Frequencies from Nuclear Magnetic Resonance Signals*. PhD thesis, University of California at Irvine, 2001.

[97] V. A. Mandelshtam and M. Ovchinnikov. Extraction of tunneling splittings from a real-time semiclassical propagation. *J. Chem. Phys.*, 108:9206–9209, 1998.

[98] J. Chen, V. A. Mandelshtam, J. Ka, and D. Neuhauser. Calculating excited vibrational levels of polyatomic molecules by diffusion monte carlo. *J. Chem. Phys.*, in preparation.

[99] A. Tikhonov and V. Arsenin. *Solutions of ill-posed problems.* Winston and Sons, Washington, 1977.

[100] K. F. Morris and C. S. Johnson Jr. Diffusion ordered two-dimensional nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.*, 114:3139–3142, 1992.

[101] K. F. Morris and C. S. Johnson Jr. Resolution of discrete and continuous molecular size distributions by means of diffusion-ordered 2D NMR. *J. Am. Chem. Soc.*, 115:4291–4299, 1993.

[102] C. S. Johnson Jr. Diffusion ordered nuclear magnetic resonance spectroscopy: Principles and applications. *Prog. NMR. Spec.*, 34:203–256, 1999.

[103] G. S. Armstrong and V. A. Mandelshtam. The extended fourier transform for 2D spectral estimation. *J. Magn. Reson.*, 153:22–31, 2001.

[104] F. Bloch, W. W. Hansen, and M. Packard. Nuclear induction. *Phys. Rev.*, 69:127–127, 1946.

[105] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69:37–38, 1946.

[106] J. T. Arnold, S. S. Dharmatii, and M. E. Packard. Chemical effects on nuclear induction signals from organic compounds. *J. Chem. Phys.*, 19:507, 1951.

[107] R. Freeman. Pioneers of high-resolution NMR. *Concepts of Magnetic Resonance*, 11:61–70, 1999.

[108] R. R. Ernst and W. A. Anderson. Application of fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instrum.*, 37:93, 1966.

[109] W. P. Aue, E. Bartholdi, and R. R. Ernst. Two-dimensional spectroscopy. application to nuclear magnetic resonance. *J. Chem. Phys.*, 64:2229–2246, 1976.

[110] J. Jeener. Ampére international summer school. Basko Polje, Yugoslavia, 1971.

[111] K. Blum. *Density Matrix Theory and Applications.* Plenum Press, New York, 1981.

[112] R. R. Ernst, G. Bodenhausen, and A. Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Oxford University Press, Oxford, UK, 1990.

[113] R. R. Ernst. Nuclear magnetic resonance fourier-transform spectroscopy (nobel lecture). *Angew. Chem., Int. Eng. Ed.*, 31:805–930, 1992.

[114] H. Hu. *Applications of Composite Pulsed Field Gradients and semi-selective pulses to high resolution multidimensional NMR spectroscopy.* PhD thesis, University of California at Irvine, 1999.

[115] P. J. Hore. Solvent suppression in fourier transform nuclear magnetic resonance. *J. Magn. Reson.*, 55:283–300, 1983.

[116] V. A. Mandelshtam, T. R. Ravuri, and H. S. Taylor. Calculation of the density of resonance states using the stabilization method. *Phys. Rev. Lett.*, 70(1932–1935), 1993.

[117] G. A. Morris. Compensation of instrumental imperfections by deconvolution using an internal reference signal. *J. Magn. Reson.*, 80:547–552, 1988.

[118] W.P. Aue, J. Karhan, and R.R. Ernst. Homonuclear broad band decoupling and two-dimensional j-resolved nmr spectroscopy. *J. Chem. Phys.*, 64:4226–4227, 1976.

[119] A. A. De Angelis. *New Methods to Improve Spectral Resolution in Solution NMR.* PhD thesis, University of California at Irvine, 2001.

[120] T. L. Hwang and A. J. Shaka. Cross relaxation without TOCSY: Transverse rotating-frame overhauser effect spectroscopy. *J. Am. Chem. Soc.*, 114:3157–3159, 1992.

[121] A. J. Shaka, C. J. Lee, and A. Pines. Iterative schemes for bilinear operators-application to spin decoupling. *J. Magn. Reson.*, 77:274–293, 1988.

[122] S. P. Rucker and A. J. Shaka. Broad-band homonuclear cross polarization in 2D NMR using DIPSI-2. *Mol. Phys.*, 68:509–517, 1989.

[123] D. I. Hoult and R. E. Richards. Critical factors in design sensitive high-resolution nuclear magnetic resonace spectrometer. *Proc. Roy. Soc. Lond. A Mat.*, 334(311–340), 1975.

[124] A. Bax, A. F. Mehlkopf, and J. Smidt. Homonuclear broadband-decoupled absorption-spectra, with linewidths which are independent of the transverse relaxiation time. *J. Magn. Reson.*, 35:373–377, 1979.

[125] A. Bax and R. Freeman. Investigation of complex networks of spin-spin coupling by two-dimensional NMR. *J. Magn. Reson.*, 44:542–561, 1981.

[126] G. W. Vuister and A. Bax. Resolution enhancement and spectral editing of uniformly C-13-enriched proteins by homonuclear broadband C-13 decoupling. *J. Magn. Reson.*, 98:428–435, 1992.

[127] P. Powers, A. M. Gronenborn, G. M. Clore, and A. Bax. 3-dimensional triple resonance NMR of C-13/N-15-enriched proteins using constant-time evolution. *J. Magn. Reson.*, 94:209–213, 1991.

[128] A. Bax, A. F. Mehlkopf, and J. Smidt. Absorption-spectra from phase-modulated spin echoes. *J. Magn. Reson.*, 35:167–169, 1979.

[129] G. Zhu and A. Bax. Improved linear prediction for truncated signals of known phase. *J. Magn. Reson.*, 90:405–410, 1990.

[130] G. Zhu and A. Bax. Two-dimensional linear prediction for signals truncated in both dimensions. *J. Magn. Reson.*, 98:192–199, 1992.

[131] J. C. Lindon and A. G. Ferrige. Digitization and data-processing in fourier-transform NMR. *Prog. NMR Spec.*, 14:27–66, 1980.

[132] G. Bodenhausen and D. J. Ruben. Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, 69:185–189, 1980.

[133] L. Muller. Sensitivity ehanced detection of weak nuclei using heteronuclear multiple quantum coherence. *J. Am. Chem. Soc.*, 101:4481–4484, 1979.

[134] M. A. Smith, H. Hu, and A. J. Shaka. Improved broadband inversion performance for nmr in liquids. *J. Magn. Reson.*, 151:269–283, 2001.

[135] J. Keeler and D. Neuhaus. Comparison and evaluation of methods for two-dimensional NMR-spectra with absorption-mode lineshapes. *J. Magn. Reson.*, 63:454–472, 1985.

[136] A. A. De Angelis, H. Hu, and A. J. Shaka. Partly selective radiofrequency pulses. *J. Magn. Reson.*, (in preparation).

[137] L. Emsley and G. Bodenhausen. Phase shifts induced by transient Bloch-Siegert effects in nmr. *Chem. Phys. Lett.*, 168:297–303, 1990.

[138] S. Schaublin, H. Hohener, and P. R. Ernst. Fourier spectroscopy of nonequilibrium states, applicatioin to CIDNP, Overhauser exepriments, and relaxation time measurements. *J. Magn. Reson.*, 13:196–216, 1974.

[139] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMR-Pipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.*, 6:277–293, 1995.

[140] S. Grzesiek and A. Bax. An efficient experiment for sequencial backbone assignement of medium-size isotopically enriched proteins. *J. Magn. Reson.*, 99:201–207, 1992.