Nuclear Magnetic Resonance Data Processing Methods

A thesis submitted for the degree of

Doctor of Philosophy

Jonathan Alcwyn Jones

St John's College, Oxford

Trinity Term 1992

Abstract

Nuclear Magnetic Resonance Data Processing Methods

A thesis submitted for the degree of Doctor of Philosophy

Jonathan Alcwyn Jones
St John's College, Oxford
Trinity Term 1992

This thesis describes the application of a wide variety of data processing methods, in particular the Maximum Entropy Method (MEM), to data from Nuclear Magnetic Resonance (NMR) experiments. Chapter 1 provides a brief introduction to NMR and to data processing, which is developed in chapter 2. NMR is described in terms of the classical model due to Bloch, and the principles of conventional (Fourier transform) data processing developed. This is followed by a description of less conventional techniques. The MEM is derived on several grounds, and related to both Bayesian reasoning and Shannon information theory.

Chapter 3 describes several methods of evaluating the quality of NMR spectra obtained by a variety of data processing techniques; the simple criterion of spectral appearance is shown to be completely unsatisfactory. A Monte Carlo method is described which allows several different techniques to be compared, and the relative advantages of Fourier transformation and the MEM are assessed.

Chapter 4 describes *in vivo* NMR, particularly the application of the MEM to data from Phase Modulated Rotating Frame Imaging (PMRFI) experiments. In this case the conventional data processing is highly unsatisfactory, and MEM processing results in much clearer spectra.

Chapter 5 describes the application of a range of techniques to the estimation and removal of splittings from NMR spectra. The various techniques are discussed using simple examples, and then applied to data from the amino acid iso-leucine.

The thesis ends with five appendices which contain historical and philosophical notes, detailed calculations pertaining to PMRFI spectra, and a listing of the MEM computer program.

Τοις ἁγιοις τοις οὐσιν και πιστοις ἐν Χριστῳ Ἰησου.

Acknowledgements

It is my duty and my joy to begin by thanking my supervisor, Peter Hore. His guidance has been a constant source of encouragement and inspiration during the past seven years.

Next I would like to thank my co-workers in the field of data processing. Foremost among these is Geoff Daniell of Southampton University, who provided Maximum Entropy Method software and gave much helpful advice on other matters. Also noteworthy is my predecessor, David Grainger, whose insistence that our research was a branch of chemistry, and not a branch of mathematical philosophy, helped to curb my worst excesses; in deference to him I have restricted my musings to the (copious) appendices. My successor, Paul Hodgkinson, with his persistent refusal to believe my explanations and insistence on asking questions I couldn't answer, has also been extremely helpful.

This is a thesis about data processing; consequently it has little, if anything, to say about the important topic of data acquisition. I have adopted a particularly simple approach: using other people's data. The data used in chapter 4 was recorded by Ronald Ouwerkerk and Ruth Dixon of the John Radcliffe Hospital; Chris Relf gave assistance in transferring it down the hill, and advice on the best methods of displaying it. I am also grateful to Peter Styles for introducing me to *in vivo* NMR, and co-supervising my work in this field. The data used in chapter 5 was recorded by David Grainger, adding further to my debt of gratitude.

Modern data processing means using computers, and I am grateful to all those who have helped me work with them. I would particularly like to thank David Hunter, whose belief (against all the evidence) that it *was* possible to use the Norsk resulted in feats of programming ingenuity. I would also like to thank Pete Biggs for making the new Sun network so easy and pleasant to use.

There are, of course, many other people who have helped me in my research. Among these are Bill "swamp monster" Broadhurst, who taught me quantum mechanics, Steve Wimperis, who contributed several useful discussions, and Gooitzen Zwannenburg, who provided regular bulletins on life in Ambridge.

Life is more than Chemistry, and I am particularly grateful to those people who continue to remind me of this. Worthy of particular note are the Thursday Tea Group (Elizabeth, Matthew, Emily, Hugh, Jackie, Bill, Larry and Sally), the St John's College Ecumenical Christian Fellowship (too numerous to name) and my godson Andrew.

Last, but never least, I would like to thank my family: Ruth, Alan, David and Tracey.

Contents

Chapter 1

Introduction

> O stands for Oxford. Hail salubrious seat
> Of learning! Academical Retreat!
> Home of my Middle Age! Malarial Spot!
> Which People call Medeeval (though it's not).
> The marshes in the neighbourhood can vie
> With Cambridge, but the town itself is dry,
> And serves to make a kind of Fold or Pen
> Wherein to herd a lot of Learned Men.
>
> Hilaire Belloc, "A Moral Alphabet" (1899)

From its simple beginnings[1, 2] in 1948, Nuclear Magnetic Resonance (NMR)[3] has rapidly risen to become one of the most important techniques available to scientists. It is one of the cornerstones of analytical chemistry[4], a valuable probe of biochemical reactions *in vivo*[5], and the basis of many medical and industrial imaging techniques[6]. This rise has occurred in parallel with, and partly as a result of, increases in the power and decreases in the cost of computers. Computers now play a central role in modern NMR: firstly in experimental control, secondly in data processing, and thirdly in the assignment and analysis of spectra. This thesis is concerned with the application of new data processing methods to NMR, and in particular application of the Maximum Entropy Method (MEM)[7].

This chapter contains a simple introduction to data processing, to NMR and to the MEM. Chapter 2 extends and develops the concepts of chapter 1, and discusses the application of the MEM to NMR; other "non-conventional" data processing

techniques are also discussed. Chapter 3 describes a number of techniques for evaluating the MEM and other methods as applied to NMR, and indicates some of the strengths and weaknesses of each method. Chapters 4 and 5 describe more complex applications of the MEM to NMR: chapter 4 considers *in vivo* spectroscopy, while chapter 5 concerns the measurement and removal of spectral splittings.

1.1 Data processing

During any experiment measurements are made and conclusions obtained. Most experiments are quantitative, that is both the measurements and the conclusions can be expressed as one or more numbers; the numbers making up the conclusions may be a small number of parameters, or they may describe a digitised spectrum or image. In general the measurements can be written as a vector $\boldsymbol{d}$ of P elements and the conclusions as a vector $\boldsymbol{s}$ of N elements. For simplicity these may be referred to as the "data" and the "spectrum" respectively[1].

In order to interpret the data, some sort of theory is required. Generally this theory contains a number of unknowns (the conclusions or spectrum) and gives a means of calculating the expected data for any set of values of these unknowns (strictly speaking the theory should also contain a model of systematic and random errors present in any measurement). Given some experimental data, the theory must be inverted in order to estimate the conclusions; this inversion step is called data processing.

Data processing problems may be divided into three groups depending on the relative sizes of N and P[8].

1. $N < P$; the conclusions are over-determined by the data. In general the data is inconsistent, that is there is no set of N conclusions consistent with all P measurements, as a result of experimental error.

2. $N = P$; the conclusions are exactly determined by the data. There is one

[1]Throughout this thesis "data" is used as a convenient abbreviation for "data vector", and is consequently a *singular* noun.

consistent set of conclusions which can be determined mathematically.

3. $N > P$; the conclusions are under-determined by the data. There is no unique set of conclusions consistent with the data; instead there is an infinite number of them.

Note that in the above it is assumed that all the elements of $\boldsymbol{s}$ and $\boldsymbol{d}$ are independent, so that N and P are the number of independent conclusions and measurements.

Problems of the first and second kind may be readily solved by conventional methods (whether or not this is the best way to solve such problems will be discussed later), but problems of the third kind are much more difficult. There are three broad approaches to this problem.

1. The theory is developed, or assumptions are made, so that the data may be fitted by a simple model; this converts the problem to one of the $N < P$ kind. The model constrains the form of the conclusions quite sharply, and a poor model can result in distortions.

2. Sufficient data values are invented to convert the problem to one of the $N = P$ kind, which is then solved by conventional methods; this is a frequent practice in Fourier methods, in which unmeasured Fourier components are assumed to be zero. Unsurprisingly this can result in artefacts and distortions.

3. Assumptions about the conclusions are made so that they are no longer independent and the problem becomes soluble. Clearly this is related to case (1), but the assumptions made are of a general kind, and do not sharply constrain the form of the conclusions. Examples of this approach include constrained optimisation, the use of regularization functions, and the MEM.

Examples of all of these techniques will be described in subsequent chapters.

1.2 Nuclear Magnetic Resonance

1.2.1 Nuclear spin

Many atomic nuclei possess an intrinsic angular momentum called spin[9]. A nucleus of spin quantum number I (which is integral or half-integral) has an angular momentum $\boldsymbol{I}$ with magnitude $\sqrt{I(I+1)}\hbar$. Associated with this is a magnetic moment $\boldsymbol{\mu} = \gamma \boldsymbol{I}$, where γ is a constant characteristic of the nucleus (the gyromagnetic ratio). In the presence of a magnetic field $\boldsymbol{B}_0$ the energy of the nucleus is given by

$$E = -\boldsymbol{\mu} \cdot \boldsymbol{B}_0. \tag{1.1}$$

Conventionally the magnetic field points along the z-axis, and so

$$E = -\mu_z B_0 \tag{1.2}$$

where μ_z is the component of $\boldsymbol{\mu}$ along the z-axis. For a nucleus of non-zero spin I there are $2I+1$ possible orientations of $\boldsymbol{\mu}$ with respect to the field with $\mu_z = m_I \gamma \hbar$, where m_I, the magnetic quantum number, can take values from $-I$ to $+I$ in integral steps. Each orientation has a different energy

$$E_{m_I} = -m_I \gamma \hbar B_0 \tag{1.3}$$

so there are $2I + 1$ energy levels.

This thesis is concerned solely with spin-$\frac{1}{2}$ nuclei, such as ^{1}H and ^{31}P. In this case there are two energy levels with $m_I = \pm\frac{1}{2}$, separated by an energy gap $\Delta E = \gamma \hbar B_0$. Transitions between these levels are detected in NMR experiments. For typical fields (of the order of 10 Tesla) this gap corresponds to radiofrequency radiation.

1.2.2 Chemical shift

Originally NMR was used as a method of measuring gyromagnetic ratios, but in 1950 it was discovered that the ^{14}N NMR spectrum of ammonium nitrate contained two lines[10]. These two lines correspond to nitrogen nuclei in two different

chemical environments, and so the effect was referred to as the chemical shift. In modern NMR experiments the chemical shift is used to distinguish between nuclei in chemically distinct environments.

The chemical shift arises from local magnetic fields induced in the sample. These local fields are due to motion of the surrounding electrons induced by the applied field; they may either reinforce or oppose the applied field, and are proportional to it. The total field experienced by a nucleus is simply the sum of these fields; this sum may be written as $B = B_0(1 - \sigma)$ where σ is called the screening constant.

The screening constant is an anisotropic quantity, that is the value of σ for a particular nucleus depends on the orientation of the molecule containing the nucleus with respect to the applied field. In liquids and solutions rapid molecular tumbling averages out the anisotropy, resulting in a single average value for σ called the isotropic screening constant.

Nuclei in different chemical environments will be surrounded by different electron clouds, and so will have different screening constants; it is this variation in σ which constitutes the chemical shift. The chemical shift range for most nuclei is small, and is conveniently measured in parts per million (ppm); for ^{1}H the typical chemical shift range is around 10 ppm. In order to measure these small shifts accurately the applied magnetic field must be extremely homogeneous, typically to 1 part in 10^{10}.

1.2.3 Couplings

In addition to structure due to chemical shifts, most NMR spectra also contain fine structure due to spin-spin coupling. This results in groups of lines called multiplets. Coupling between two spin-$\frac{1}{2}$ nuclei results in the line due to each nucleus being split into a doublet; the splitting in each doublet is the same and is called the spin-spin coupling constant, J. In larger systems each nucleus may be coupled to several nuclei, giving rise to doublets of doublets and other more complex patterns; the value of J may be different for each coupling.

Spin-spin coupling can occur through two mechanisms: dipole-dipole coupling and scalar coupling. Dipole-dipole coupling is caused by the direct interaction between two magnetic dipoles due to a pair of spins. This is a large effect, but it is strictly anisotropic (that is it is averaged to zero by molecular tumbling); consequently the splittings observed in spectra of liquid samples cannot be due to this effect.

Scalar coupling is a hyperfine effect, which occurs as a result of the Fermi contact interaction. Nuclei are of finite size, and consequently their magnetic field is not that of a pure point dipole; at very small distances (of the order of 10^{-14} m) it is significantly different. This is irrelevant for other nuclei (at a distance of around 10^{-10} m), but is important for electrons which approach the nucleus much more closely. Bonding electrons can interact with two or more nuclei simultaneously, and so can mediate an internuclear interaction; the presence of a scalar coupling indicates that two nuclei are connected by a chemical bond. Since the interaction is transmitted via electrons rather than directly through space, it is isotropic and so is not removed by tumbling.

1.2.4 Relaxation

As described above, a nucleus with spin $I = \frac{1}{2}$ in a magnetic field has two states of different energy; at thermal equilibrium nuclei will be found in both states with a Boltzmann distribution. There will be more nuclei in the lower state than the upper state, and so the sample will have a small total magnetic moment along the field direction, called longitudinal magnetization. At equilibrium there will be no phase-coherence between nuclei, and so the sample will have no transverse magnetization.

If the sample is perturbed, such that the populations of the spin states are changed or phase coherence is created, the spin system will relax back to its equilibrium state. Longitudinal relaxation involves transitions between the upper and lower energy states, and so involves the transfer of energy between the spin system and its surroundings. Transverse relaxation involves the loss of phase coher-

ence: this can occur through exchange of energy between the spin system and the surroundings, or through energy exchange within the spin system as a result of simultaneous transitions by two nuclei.

Unlike many other spectroscopic techniques, NMR relaxation does not occur by spontaneous emission: because of the low energies involved, spontaneous emission is negligibly slow. The transitions involved in relaxation must be stimulated by a fluctuating magnetic field. Fluctuations in the shielding constant of a nucleus and in dipole-dipole coupling with other nuclei, as a result of molecular tumbling, provide a suitable fluctuating field.

In practical experiments transverse magnetization can also decay as a result of magnetic field inhomogeneity; this, however, is not relaxation since the decay can be reversed, giving rise to a spin echo[11].

1.3 The Maximum Entropy Method

The maximum entropy method is a general technique for solving inverse linear problems, that is problems where the expected measurements are a linear function of the conclusions. It was originally developed in optics[12] and astronomy[13], but is now used in many fields including NMR[14–54]. The method may be justified in many different ways, but discussion of the underlying principles is deferred to chapter 2; a simple outline of the method is given below.

For any problem there is a set of all possible spectra, $\{\boldsymbol{s}\}$. Most of these are inconsistent with the observed data $\boldsymbol{d}$. For any trial spectrum $\boldsymbol{s}$, corresponding trial data $\boldsymbol{t}$ can be calculated by

$$\boldsymbol{t} = \boldsymbol{R}\boldsymbol{s} \tag{1.4}$$

where $\boldsymbol{R}$ is a matrix encoding the linear relationship between conclusions and measurements. This trial data can be compared with the experimental data using a χ^2 criterion

$$\chi^2 = \sum_{i=1}^{P} (d_i - t_i)^2 / \sigma_i^2 \tag{1.5}$$

where σ_i is the estimated error in the corresponding data point. A trial spectrum

is consistent with the data if $\chi^2 \leq P$. There are, however, an infinite number of spectra which satisfy this consistency criterion; all these spectra match the data equally well, and it is impossible to select any one of them on the basis of the data. Any such selection must be based on prejudice about the desired solution.

The MEM chooses the spectrum of highest spectral entropy. This is a non-linear pointwise convex function of the elements of $\boldsymbol{s}$; its exact form will be discussed later. The maximum entropy criterion is sufficient to select a single spectrum, which is called the MEM spectrum.

Chapter 2

Detailed theory

> Mathematicians are like Frenchmen: whenever you say something to them, they translate it into their own language, and at once it is something entirely different.
>
> Johann Wolfgang von Goethe, "Maxims and Reflections" (1829).

In this chapter the theory of NMR and of the MEM will be developed from the simple explanations of the previous chapter. The first section deals with NMR and describes the classical framework for understanding it developed by Bloch (this classical model is sufficient for this thesis, and the more complete quantum mechanical description will not be used). Subsequent sections develop the theory of the MEM and other data processing methods, describing both their theoretical justification and their practical application.

2.1 Theory of NMR

As described in chapter 1, a nucleus of non-zero spin I, with angular momentum $\boldsymbol{I}$, has an associated magnetic moment $\boldsymbol{\mu} = \gamma \boldsymbol{I}$. The motion of this moment in a magnetic field can be described classically; combined with a phenomenological treatment of relaxation, suggested by Bloch[55], this is sufficient to describe many NMR experiments.

2.1.1 Bloch vector model

Motion of a free spin

According to classical electromagnetic theory, a magnetic moment $\boldsymbol{M}$ in a field $\boldsymbol{B}$ experiences a torque $\boldsymbol{C} = \boldsymbol{M} \wedge \boldsymbol{B}$, resulting in a change in the angular momentum of the system. A nucleus of non-zero spin I has angular momentum $\boldsymbol{I}$ of magnitude $\sqrt{I(I+1)}\,\hbar$ and a magnetic moment $\boldsymbol{\mu} = \gamma \boldsymbol{I}$; hence in a field $\boldsymbol{B}_0$ its classical behaviour[56] is described by

$$\frac{d\boldsymbol{I}}{dt} = \boldsymbol{\mu} \wedge \boldsymbol{B}_0 \tag{2.1}$$

or

$$\frac{d\boldsymbol{\mu}}{dt} = \gamma \boldsymbol{\mu} \wedge \boldsymbol{B}_0. \tag{2.2}$$

The assumption of classical behaviour might seem cavalier, but it can be easily justified. Within the Heisenberg representation[57], the evolution of the spin angular momentum is

$$\frac{d\boldsymbol{I}}{dt} = \frac{i}{\hbar}[\mathcal{H}, \boldsymbol{I}] \tag{2.3}$$

where $\mathcal{H} = -\gamma \boldsymbol{B}_0 \cdot \boldsymbol{I}$ is the Hamiltonian describing the interaction between the spin and the magnetic field. The evolution of the z-component is[56]

$$\begin{aligned} \frac{dI_z}{dt} &= -\frac{i\gamma}{\hbar}\left(B_{0x}[I_x, I_z] + B_{0y}[I_y, I_z] + B_{0z}[I_z, I_z]\right) \\ &= \gamma\left(I_x B_{0y} - I_y B_{0x}\right) \end{aligned} \tag{2.4}$$

and the evolution of the x and y-components may be obtained by cyclic permutation. Combining these three

$$\frac{d\boldsymbol{I}}{dt} = \gamma \boldsymbol{I} \wedge \boldsymbol{B}_0, \tag{2.5}$$

which is identical to the classical result.

An NMR sample contains a very large number of nuclei; fortunately it is not necessary to consider them all individually. The individual magnetic moments may be summed to give a sample magnetic moment, $\boldsymbol{M}$. This behaves like the individual moments, and so

$$\frac{d\boldsymbol{M}}{dt} = \gamma \boldsymbol{M} \wedge \boldsymbol{B}_0. \tag{2.6}$$

Equation 2.6 corresponds to precession of $\boldsymbol{M}$ around $\boldsymbol{B}$ at an angular frequency $\omega_0 = -\gamma B_0$ called the Larmor frequency; this frequency corresponds to the gap between energy levels described in chapter 1.

Relaxation

As described in chapter 1, the equilibrium sample magnetic moment lies along the z-axis, that is $\boldsymbol{M} = M_0\boldsymbol{k}$. If it is perturbed, it will evolve according to equation 2.6, but in addition it will relax back to its equilibrium value. Phenomenologically this relaxation can be described by two processes[55]:

1. relaxation of the z-magnetization towards its equilibrium value M_0, with a time constant T_1;

2. relaxation of the x and y-magnetizations towards their equilibrium values of 0, with a time constant T_2.

Combining this with equation 2.6 gives

$$\frac{d\boldsymbol{M}}{dt} = \gamma\boldsymbol{M} \wedge \boldsymbol{B}_0 - \frac{M_x\boldsymbol{i} + M_y\boldsymbol{j}}{T_2} - \frac{M_z - M_0}{T_1}\boldsymbol{k} \tag{2.7}$$

where $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ are unit vectors along the x, y and z-axes.

The rotating frame

Equation 2.7 refers to the motion of $\boldsymbol{M}$ in the laboratory frame. In many cases it is more convenient to consider its motion in a rotating frame[56], with a rotation vector $\boldsymbol{\omega}$. For any time-dependent vector $\boldsymbol{A}(t)$ the time derivatives in the laboratory and rotating frames are related by

$$\left(\frac{d\boldsymbol{A}}{dt}\right)_{lab} = \left(\frac{d\boldsymbol{A}}{dt}\right)_{rot} + \boldsymbol{\omega} \wedge \boldsymbol{A}. \tag{2.8}$$

Hence for the sample magnetic moment

$$\left(\frac{d\boldsymbol{M}}{dt}\right)_{rot} = \gamma\boldsymbol{M} \wedge \left(\boldsymbol{B}_0 + \frac{\boldsymbol{\omega}}{\gamma}\right) - \frac{M_x\boldsymbol{i} + M_y\boldsymbol{j}}{T_2} - \frac{M_z - M_0}{T_1}\boldsymbol{k}. \tag{2.9}$$

This equation has the same form as equation 2.7 if the applied magnetic field $\boldsymbol{B}_0$ is replaced by an effective field $\boldsymbol{B}_e = \boldsymbol{B}_0 + (\boldsymbol{\omega}/\gamma)$.

Because of the great utility of the rotating frame, and the simplicity of conversion between laboratory and rotating frames, it will be widely used in the following sections. In general the subscripts *lab* and *rot* will be dropped; it will be obvious from the context which frame is implied. Similarly $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ will be used for unit vectors in both frames without distinction.

Continuous wave NMR

Consider a spin system at thermal equilibrium in a magnetic field $\boldsymbol{B}_0$. Suppose that an additional field $\boldsymbol{B}_1$ is applied which is perpendicular to $\boldsymbol{B}_0$ and rotating around it at a frequency ω. Viewed in a frame rotating at the same frequency, $\boldsymbol{B}_1$ is static, and can be taken to lie along $\boldsymbol{i}$. The effective field is now

$$\boldsymbol{B}_e = \left(B_0 + \frac{\omega}{\gamma}\right)\boldsymbol{k} + B_1\boldsymbol{i} \tag{2.10}$$

and the Bloch equations in the rotating frame are

$$\frac{d\boldsymbol{M}}{dt} = \boldsymbol{M} \wedge (\Delta\omega\boldsymbol{k} - \omega_1\boldsymbol{i}) - \frac{M_x\boldsymbol{i} + M_y\boldsymbol{j}}{T_2} - \frac{M_z - M_0}{T_1}\boldsymbol{k} \tag{2.11}$$

where $\Delta\omega = \omega - \omega_0$ and $\omega_1 = -\gamma B_1$ are defined for convenience. The steady state solution of equation 2.11 (obtained by setting $d\boldsymbol{M}/dt = 0$) is

$$\begin{aligned} M_x &= \frac{-\Delta\omega\,\omega_1\,T_2^2}{1 + (T_2\Delta\omega)^2 + \omega_1^2 T_1 T_2} M_0 \\ M_y &= \frac{-\omega_1\,T_2}{1 + (T_2\Delta\omega)^2 + \omega_1^2 T_1 T_2} M_0 \\ M_z &= \frac{1 + (T_2\Delta\omega)^2}{1 + (T_2\Delta\omega)^2 + \omega_1^2 T_1 T_2} M_0. \end{aligned} \tag{2.12}$$

In the laboratory frame these magnetization components precess at a frequency ω; this will induce an oscillating current in a coil which can be readily detected. Unsurprisingly the signal intensity is proportional to the initial polarization M_0.

If the rotating field is weak ($\omega_1^2 T_1 T_2 \ll 1$) these equations can be approximated by

$$\begin{aligned} M_x &= \frac{-\Delta\omega\,\omega_1\,T_2^2}{1 + (T_2\Delta\omega)^2} M_0 \\ M_y &= \frac{-\omega_1\,T_2}{1 + (T_2\Delta\omega)^2} M_0 \\ M_z &= M_0 \end{aligned} \tag{2.13}$$

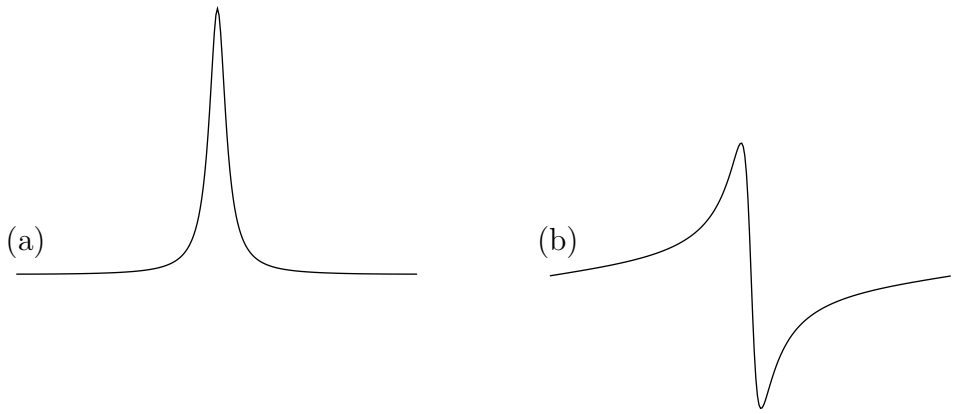

Figure 2.1: Lorentzian lineshapes: (a) absorption, (b) dispersion.

indicating that this situation corresponds to negligible saturation. In this case $M_y(\Delta\omega)$ is proportional to the normalized Lorentzian absorption lineshape

$$A_{T_2}(\Delta\omega) = \frac{1}{\pi}\frac{1/T_2}{(1/T_2)^2 + (\Delta\omega)^2} \tag{2.14}$$

which has a half width at half maximum height of $1/T_2$. Similarly $M_x(\Delta\omega)$ is proportional to the dispersion lineshape

$$D_{T_2}(\Delta\omega) = \frac{1}{\pi}\frac{\Delta\omega}{(1/T_2)^2 + (\Delta\omega)^2}. \tag{2.15}$$

These lineshapes are depicted in figure 2.1. In practice the phase of the detector must be adjusted so that the absorption component is detected.

If the intensity of the rotating field is increased the absorptive resonance curve broadens but retains its shape. The signal intensity initially increases, since the numerators are proportional to ω_1, reaches at maximum at $\omega_1^2 T_1 T_2 = 1$, and then decreases as a result of the broadening. By contrast the intensity of the dispersion component increases with ω_1 to an asymptotic value[56].

These equations form the basis of continuous wave NMR spectroscopy. The sample is placed in a static magnetic field and a linearly oscillating field applied. This linear oscillation can be considered as the vector sum of two counter-rotating fields: the component rotating in the appropriate sense will produce transverse magnetization when the system is at or near resonance; the component rotating

in the opposite sense can be largely ignored (accurate treatment shows that it results in a small shift in resonance frequency, called the Bloch–Siegert shift[58]). A spectrum is acquired by measuring the transverse magnetization while the static field strength is slowly altered to bring successive signals into resonance (this is more convenient than changing the frequency of the oscillating field).

2.1.2 Fourier transform NMR

Continuous wave NMR has now been largely superseded by Fourier transform NMR[59], also known as pulsed NMR. Rather than excite each resonance in turn with a weak field, the entire system is excited simultaneously by a short intense pulse and its response recorded. This response, the free induction decay or FID, consists of a superposition of damped oscillations which is then analysed, conventionally by Fourier transformation[60, 61].

Excitation

The system is excited with a short intense pulse of radiofrequency radiation at a frequency ω close to the Larmor frequency. The oscillating magnetic field can once more be decomposed into two counter-rotating fields, of which only the appropriate one is considered. Viewed in a rotating frame, rotating at the same frequency, the field $\boldsymbol{B}_1$ is static in the transverse plane; for convenience it can be taken to lie along $\boldsymbol{i}$. Since the pulse is short relaxation can be ignored: in this case equation 2.11 simplifies to

$$\frac{d\boldsymbol{M}}{dt} = \boldsymbol{M} \wedge (\Delta\omega \boldsymbol{k} - \omega_1 \boldsymbol{i}). \tag{2.16}$$

If the $\boldsymbol{B}_1$ field is intense compared to the residual field $\gamma\Delta\omega\boldsymbol{k}$ (that is $\omega_1 \gg \Delta\omega$), this corresponds to precession in the yz-plane at a frequency ω_1. If the residual field is significant, the situation is more complex; an example of this is discussed in chapter 4.

Evolution

In the simplest NMR experiment the $\boldsymbol{B}_1$ field is applied for a time t_{90} such that the sample magnetization is rotated through an angle of 90° into the xy-plane; this is called a 90° pulse. The field is then switched off and the magnetization allowed to evolve. Equation 2.11 now simplifies to

$$\frac{d\boldsymbol{M}}{dt} = \boldsymbol{M} \wedge \Delta\omega \boldsymbol{k} - \frac{M_x \boldsymbol{i} + M_y \boldsymbol{j}}{T_2} - \frac{M_z - M_0}{T_1}\boldsymbol{k}. \tag{2.17}$$

The transverse magnetization precesses at a frequency $-\Delta\omega$, and decays with a time constant T_2.

In the laboratory frame the transverse magnetization rotates at a frequency $\omega - \Delta\omega$; this can induce a current in a coil and so can be detected.

Detection

The induced current can be detected using a phase sensitive detector; this removes the carrier frequency (ω) to give signals oscillating at their offsets $\Delta\omega$. In practice it is convenient to use two detectors in quadrature, that is with their reference phases 90° out of phase, so that the sign of $\Delta\omega$ can be determined as well as its magnitude. The two signals can be combined to give a complex signal, or FID,

$$s(t) = Ae^{i\Delta\omega t}e^{-t/T_2} \tag{2.18}$$

for a single species with amplitude A, frequency offset $\Delta\omega$ and relaxation time T_2. Note that this formula only applies for positive times (there is no signal at times before the pulse). For a more complex system the signal will be given by the sum of a number of such terms.

Analysis

The FID from a single species can be analysed directly, but a FID containing several different damped oscillations is more difficult to analyse. Conventionally this problem has been solved using the Fourier transform (FT).

The one-sided inverse Fourier transform of equation 2.18 is given by

$$\begin{aligned} S(\omega) &= \int_0^\infty s(t)e^{-i\omega t}dt \\ &= \int_0^\infty Ae^{i\delta\omega t}e^{-t/T_2}dt \\ &= A\frac{1/T_2 + i\delta\omega}{(1/T_2)^2 + (\delta\omega)^2} \end{aligned} \tag{2.19}$$

where $\delta\omega = \Delta\omega - \omega$. Comparison with equations 2.14 and 2.15 indicates that the real and imaginary parts of $S(\omega)$ are Lorentzian absorption and dispersion lineshapes centred at $\Delta\omega$.

Since the Fourier transform is a linear operation, the FT of the sum of several different damped oscillations will simply be the sum of several different Lorentzian terms, one for each oscillation. $S(\omega)$ is called the spectrum corresponding to a FID $s(t)$; this is reasonable since its form is identical to that of the continuous wave spectrum described above (in the absence of saturation effects in the latter).

Sampling

The description above, in which the FID is recorded as a continuous function for an infinite length of time is, of course, idealised. In practice the FID is digitized, that is it is recorded to finite precision at discrete intervals of time, and the recording is stopped after a finite period. The Fourier transform is performed numerically using an algorithm for discrete data.

In many cases this is unimportant. In particular if the FID is sampled sufficiently finely for a sufficiently long time the discrete spectrum will be very similar to the continuous spectrum. Here *sufficiently long* means that the FID must be sampled until it has effectively decayed to zero (or more commonly to the noise level), while *sufficiently frequently* means that any frequency appearing in the FID must lie below a critical frequency called the Nyquist frequency.

The Nyquist frequency ($\mathcal{N}$) is the highest frequency which can be represented by a set of samples at a given spacing. If a single detector is used it is determined by the condition that there are exactly two sample points for each cycle of the oscillation; if quadrature detection is used one complex point per cycle is sufficient.

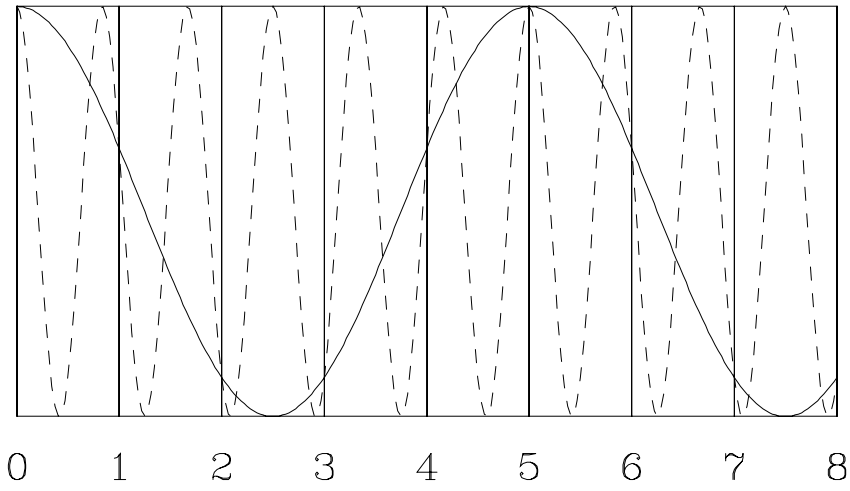

Figure 2.2: The real part of complex datasets corresponding to frequencies of $0.2\mathcal{N}$ (solid line) and $1.2\mathcal{N}$ (dashed line). At the sample points, indicated by grid lines, these two datasets are indistinguishable.

Frequencies above $\mathcal{N}$ are "aliased" by the sampling process, so that they appear to be offset by sufficient integer multiples of $\mathcal{N}$ to bring them into range (see figure 2.2).

Advantages of FT NMR

The principle advantage of Fourier transform NMR over continuous wave NMR is the Fellgett (multiplex) advantage[62], common to all Fourier techniques. Since all the resonances are excited and detected simultaneously, rather than one by one, the time required to obtain a spectrum is enormously decreased. If desired the time saved can be used in signal averaging (see below), thus increasing the signal-to-noise ratio and allowing the detection of much weaker signals. This is even more important in multidimensional experiments, in which correlations between different nuclei are detected.

Signal averaging

A simple NMR experiment need only consist of a single pulse, data acquisition and processing, but if the signal is weak the signal-to-noise ratio in the spectrum may be

insufficient for reliable analysis. In this case it is possible to repeat the experiment several times, adding together the datasets acquired from each measurement. The signal will be identical in each experiment, but the noise will vary: summing N datasets will increase the signal by a factor of N, while the noise will only be increased by a factor of $\sqrt{N}$ as a result of partial cancellation[1]. Consequently the signal-to-noise ratio will be increased by a factor of $\sqrt{N}$.

2.1.3 Response theory

The previous sections have indicated the equivalence of Fourier transform and continuous wave NMR, but as described this equivalence might appear fortuitous. In fact their equivalence is a consequence of properties underlying all spectroscopic techniques; these properties are described by response theory[3].

Many systems can be characterized in terms of their input-output relations as follows: when the system is perturbed by any input signal $x(t)$ it produces a corresponding response $y(t)$; these are related by

$$y(t) = \Phi\{x(t)\} \tag{2.20}$$

where Φ, called the system operator, is time independent. The system is said to be linear if Φ is linear, that is

$$\Phi\{x_1(t) + x_2(t)\} = \Phi\{x_1(t)\} + \Phi\{x_2(t)\}. \tag{2.21}$$

If the system is linear an arbitrary input $x(t)$ may be represented by an integral over a set of basis functions

$$x(t) = \int X(p)g(p,t)dp \tag{2.22}$$

and the output represented by an integral over the corresponding response functions

$$y(t) = \int X(p)\Phi\{g(p,t)\}dp. \tag{2.23}$$

[1]This is only true for Gaussian noise. For other noise distributions the noise intensity will increase more rapidly, but will tend towards $\sqrt{N}$ as N increases; this is a result of the central limit theorem.

A particularly useful basis set is the set of Dirac δ functions

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau \tag{2.24}$$

since in this case calculation of the response only requires knowledge of the impulse response function of the system

$$h(t) = \Phi\{\delta(t)\}. \tag{2.25}$$

Combining these equations gives

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \\ &= h(t) * x(t) \end{aligned} \tag{2.26}$$

showing that the response of the system to an arbitrary input is equal to the convolution of the input signal and the impulse response of the system. The impulse response completely characterizes the system.

Another useful set of basis functions are the eigenfunctions of the operator. It can be shown that any exponential function e^{pt} (where p is an arbitrary complex number) is an eigenfunction of any time independent linear operator[2]:

$$y(t) = \Phi\{e^{pt}\} = f(p)e^{pt}. \tag{2.27}$$

An input oscillation reappears as the same oscillation multiplied by a complex eigenvalue $f(p)$ which describes changes in amplitude and phase caused by the system. For the special case $p = i\omega$ this may be conveniently written as $H(\omega)$. The response to a harmonic input is given by equation 2.26 as

$$y(t) = \int_{-\infty}^{\infty} h(\tau)e^{i\omega t} \cdot e^{-i\omega\tau}d\tau = H(\omega)e^{i\omega t} \tag{2.28}$$

(since the excitation is an eigenfunction) and so

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-i\omega t}dt. \tag{2.29}$$

Hence the impulse response $h(t)$ and the frequency response $H(\omega)$ form a Fourier transform pair. In pulsed NMR the impulse response can be identified with the FID and the frequency response can be identified with the complex spectrum.

[2]For any time independent linear operator $(\partial/\partial t)\Phi\{x(t)\} = \Phi\{\partial x(t)/\partial t\}$. Substituting $x(t) = e^{pt}$ gives $\Phi\{x(t)\} = Ae^{pt}$; hence e^{pt} is an eigenfunction.

Causality

When discussing physical systems, there are two further restrictions (in addition to time independence and linearity) on the form of the system operator: the response must be causal, and the response to a finite excitation must be finite (here "causal" simply means that the response cannot precede the excitation). These four restrictions lead to the Kramers–Kronig relations[3, 56], which express the fact that the real and imaginary parts of the response are related to one another by a Hilbert transform

$$H(\omega) = \frac{-i}{\pi} \int_{-\infty}^{\infty} \frac{H(\omega')}{\omega - \omega'} d\omega'. \tag{2.30}$$

In NMR this means that it is possible to calculate a spectrum of arbitrary phase given only the real or the imaginary part.

It is useful to note here that the dispersion lineshapes in equation 2.19 are a result of the causal nature of the NMR response. If the response were not causal it would not be necessary to use a one sided Fourier transform, and the dispersion component would disappear.

2.1.4 Conventional data processing

As described in the previous two sections the sampled FID and spectrum are related by a discrete Fourier transform, and so Fourier transformation lies at the heart of conventional data processing. There are, however, three other stages: zero-filling, the application of weighting functions, and rephasing of the spectrum[63].

Zero-filling

This is simply the process of appending zeros to the recorded FID to increase its length. There are three reasons why this might be done.

1. To extend the data to a convenient length. The discrete FT is an order (N^2) process if performed directly, and so can be very slow for large datasets. Fortunately there exist a number of algorithms for speeding up the calculation.

In general these rely on factoring the FT matrix, and so require N to be factorizable. The most commonly used algorithm[61, 64, 65] (an order $(N \lg_2 N)$ process) requires that N be a power of two; consequently acquired data must be padded up to the next power of two.

2. To make the best use of the information in the data. As described above the real and imaginary parts of a continuous spectrum are related by a Hilbert transform, and so contain exactly the same information. This is not true of a discrete spectrum. A FID of N complex points contains $2N$ independent items of information, and so a spectrum of N complex points calculated from it must also contain $2N$ independent items of information; consequently the real and imaginary parts must be independent. However, if the FID is zero-filled to twice its length before transformation, the real and imaginary parts of the resulting spectrum (of $2N$ complex points) *are* related by a discrete Hilbert transform[66]; the two parts contain the same information and the imaginary part can be discarded.

   Another way of looking at this problem is that the real and imaginary parts of the spectrum contain equivalent signal but different noise. Zero-filling before Fourier transformation is equivalent to adding the Hilbert transform of the imaginary part of the spectrum to the real part and vice versa. The signal components will add, and so the signal intensity will be doubled, but the noise components will partially cancel so that the noise intensity is only increased by $\sqrt{2}$; consequently the signal-to-noise ratio is increased by a factor of $\sqrt{2}$.

3. To improve spectral digitization. If a spectrum is poorly digitized it may be difficult to see fine details in it. Appending zeros before transformation results in a more finely digitized spectrum, which may be easier to interpret. The additional points are effectively interpolated between the original points, and so no new information is added, but information may be greatly clarified[63].

Zero-filling is clearly appropriate if the FID has effectively decayed to zero by the end of the detection period: any further data points would simply be zero. If, however, the decay is seriously truncated there is a discontinuity at the transition from data to zeros, and this discontinuity results in "sinc-wiggle" artefacts in the spectrum. This is best understood in terms of convolutions[67].

The convolution theorem states that for any functions a and b

$$\mathcal{F}(a * b) = \mathcal{F}(a) \cdot \mathcal{F}(b) \tag{2.31}$$

(where $\mathcal{F}$ is the Fourier transform operator), so convolution in one domain may be replaced by multiplication in the corresponding Fourier domain (this relation holds for both continuous and discrete Fourier transforms). Truncation can be considered as the multiplication of the ideal untruncated data by a step function which has value unity upto the truncation point and zero from there on. The Fourier transform of a step function is a sinc function, and so truncation is equivalent to convolution with a sinc function[3]. This results in line broadening and wiggles around the lines (see figure 2.3).

Weighting functions

It is common practice in spectroscopy to convolve the spectrum with one or more functions, in order to reduce the effects of artefacts, to change lineshapes, or to optimise the signal-to-noise ratio[11]. This is particularly simple in Fourier techniques since the convolution may be replaced by a multiplication as described above. Use of the convolution theorem also clarifies why certain techniques work.

For example sinc-wiggles can be attenuated by multiplying the data by a decaying function prior to transformation. This weakens the discontinuity, and so reduces the sinc-wiggles; alternatively multiplication by a decaying function is equivalent to convolving with a broadening function, and so the wiggles are smoothed out (at the expense of further broadening the lines).

Multiplication by a decaying function also reduces the contribution from noise in the tail of the FID, and so can improve the signal-to-noise ratio. There is,

[3] $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$

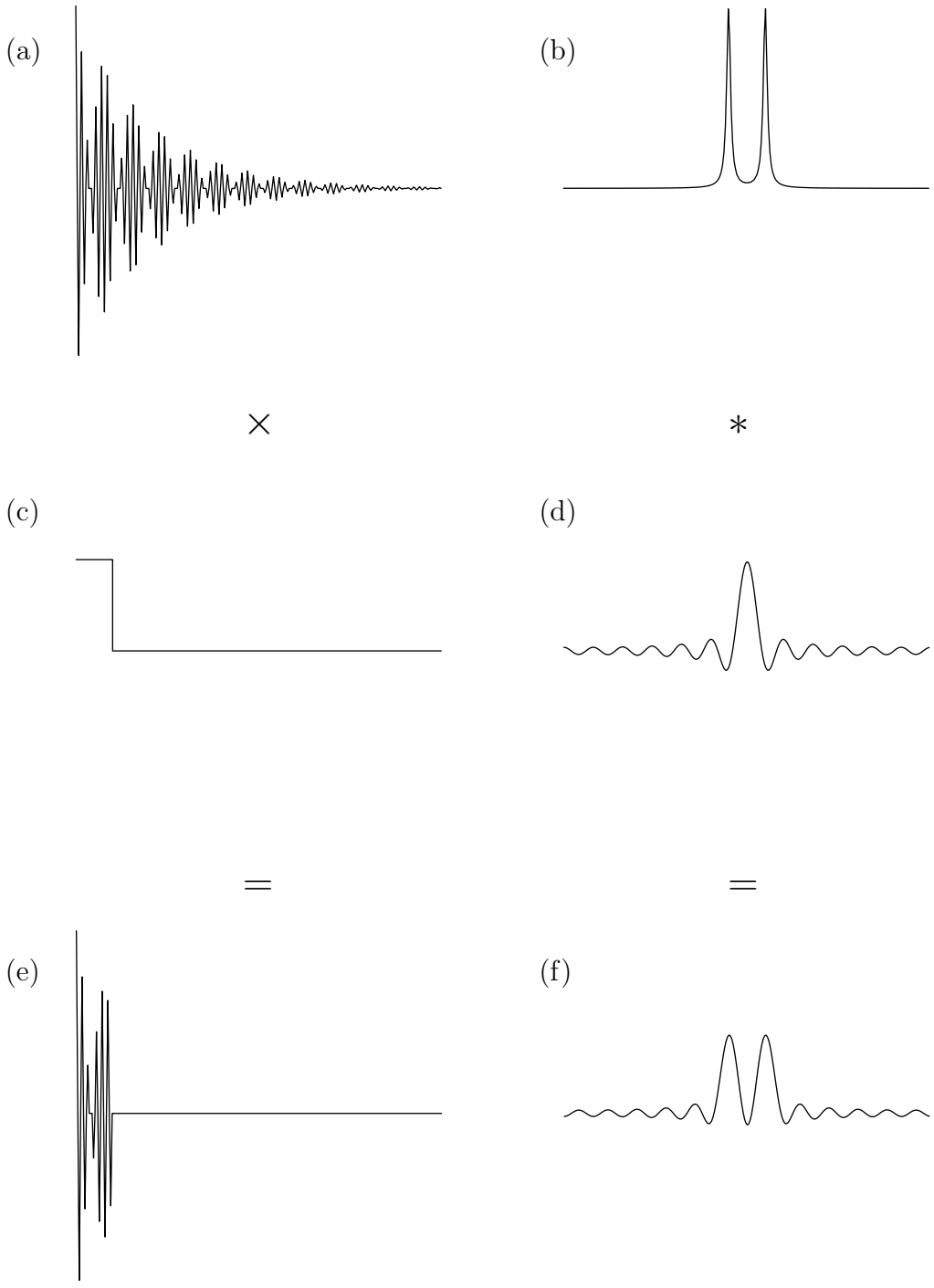

Figure 2.3: Demonstration of truncation effects. Multiplying (a) and (c) gives (e), while (f) is the convolution of (b) and (d); (b), (d) and (f) are the inverse Fourier transforms of (a), (c) and (e) respectively.

however, a limit beyond which signal is damped more rapidly than noise, and so the signal-to-noise ratio begins to decrease again. The optimum occurs when the decaying function exactly matches the decay of the FID (matched filtration); this is equivalent to convolving the spectrum with the spectral lineshape, thus selecting signal in preference to noise (see figure 2.4). Clearly it is only possible to do this perfectly if all the lines have the same shape and width.

Spectral lines can also be narrowed by multiplying the data by a rising function, but this has the effect of emphasising the noise in the tail, and so decreasing the signal-to-noise ratio. If the data is zero-filled prior to transformation, this will also enhance the sinc-wiggles. Consequently resolution enhancement is best achieved by multiplying the data by a function which first rises and then falls, thus reducing noise magnification. This also changes the lineshape which can be useful in its own right. For example it is often advantageous to replace the natural Lorentzian lineshape by a Gaussian lineshape, since Gaussian lines are narrower at the base than Lorentzian lines with the same half-height linewidth (see figure 2.4).

Phase corrections

NMR spectra are complex, but the two parts (real and imaginary) are not independent, and it is only necessary to display one part. Similarly NMR signals can be in either absorption or dispersion mode, or in some combination, and these two modes are equivalent. Nevertheless it is convenient to view a spectrum in absorption mode, since absorption lines are narrower than dispersion lines and have their maximum intensity at the line centre. It is desirable to adjust the spectrum so that all the lines appear in absorption mode in the real part of the spectrum; the imaginary part (containing dispersion mode signals) is discarded. This process is called phase correction, and three types of effect must be considered.

1. Zero order effects. These arise from differences between the excitation and detection phases; they may be readily corrected by multiplying the entire spectrum by $e^{i\phi_0}$, where ϕ_0 is called the zero order phase correction.

2. First order effects. These arise from unavoidable delays between the end of

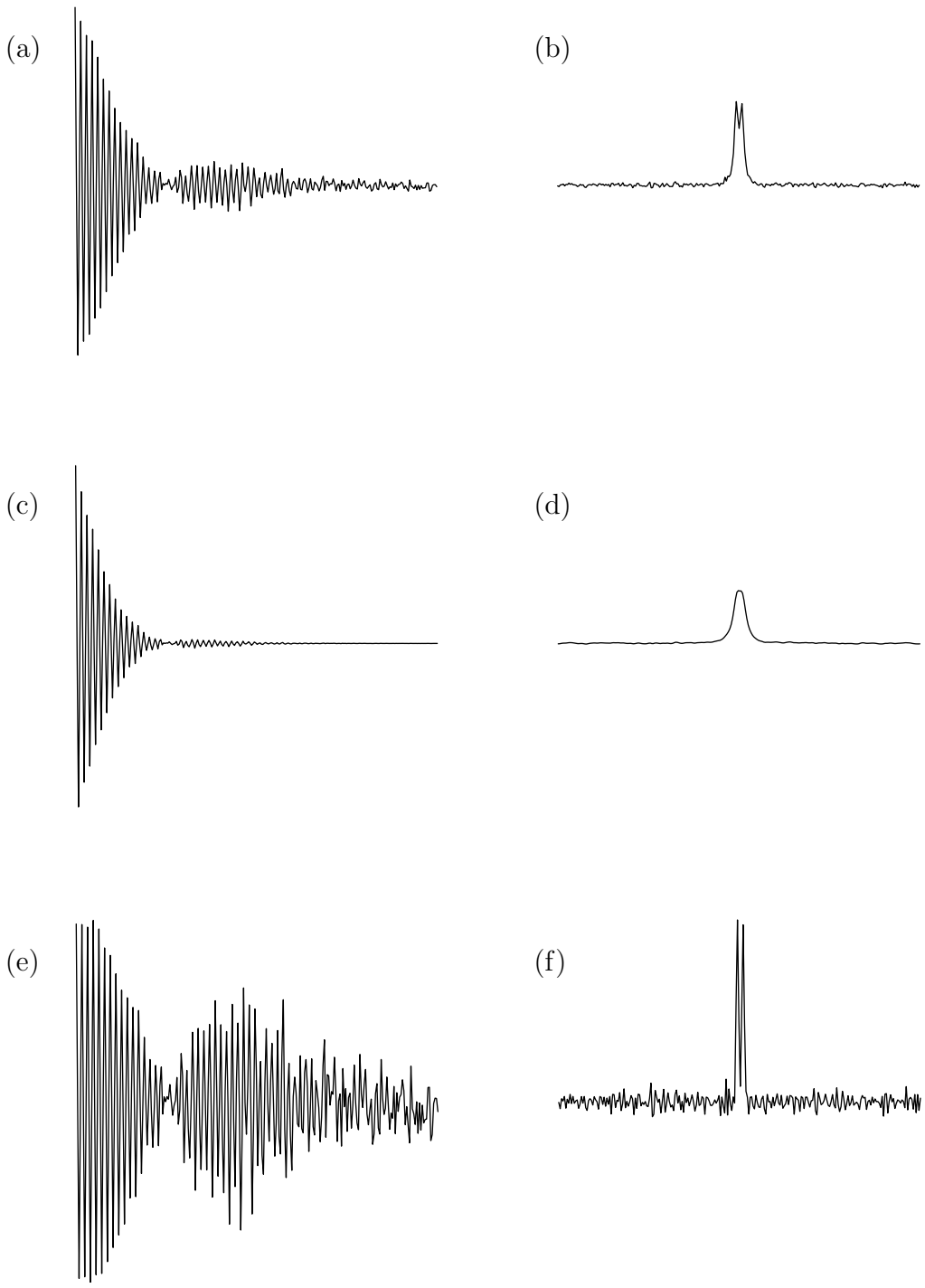

Figure 2.4: The effects of weighting functions. The original, unweighted, FID is shown in (a) and its Fourier transform in (b). The result of applying a matched filter is shown in (c) and (d): note that the signal-to-noise ratio has been increased at the cost of decreased resolution. The results of resolution enhancement using Lorentzian to Gaussian lineshape transformation are shown in (e) and (f): the resolution has been improved at the cost of a decrease in signal-to-noise ratio. Signal-to-noise ratios are 66, 145 and 29 for (b), (d) and (f) respectively.

the excitation and the beginning of detection; this causes each line to have an additional phase error $e^{i\Delta\omega\tau}$, where τ is the time delay. If all the lines have negligible width it can be corrected by multiplying the spectrum by $e^{i\phi}$, where ϕ varies linearly between $\pm\phi_1$ across the spectrum; ϕ_1 is called the first order phase correction. If a spectral line has significant width the correction will be imperfect as the angle ϕ will vary across the width of the line; ideally the entire line should be phase corrected by an amount determined by the position of the line centre.

3. Specific effects. The treatment above assumes that all the signals will have the same intrinsic phase, with differences in apparent phases being due to experimental difficulties. This is true in simple experiments, but is not true in general: multipulse NMR experiments can give rise to signals of arbitrary intrinsic phase. There is no general method for correcting these specific effects.

The first two effects are present in all NMR experiments, and are readily corrected. Specific effects must be treated individually; it may be convenient to display the same spectrum phased in several different ways. It is possible to display the spectrum in absolute value mode, that is to display the modulus of the complex spectrum (equation 2.19)

$$|S(\omega)| = A\frac{\sqrt{(1/T_2)^2 + (\delta\omega^2)}}{(1/T_2)^2 + (\delta\omega)^2}. \tag{2.32}$$

This might seem ideal, but has the disadvantage that the phasing information is lost, and that the lines are broader than the corresponding absorption peaks.

2.2 Theory of the MEM

The maximum entropy method is a very general method, and its use can be justified on many different grounds. The first part of this section presents three different derivations of the MEM equations, while the second part discusses the practicalities of actually implementing the MEM. This is followed by a discussion of the

application of the MEM to NMR data, while the final section returns to theory in presenting some analytical results pertinent to MEM processing.

2.2.1 Derivations of the MEM

In the following the MEM will be derived as an example of Bayesian reasoning, on the basis of information theory, and in a purely pragmatic fashion. The first two derivations are closely related, and serve to indicate the underlying basis of the method; the final derivation indicates the relation of the MEM to other, more conventional, methods.

Bayesian reasoning

Bayesian reasoning is reasoning based of the use of Bayes' theorem[68, 69]. This theorem describes the manner in which estimates of the probability of some set of conclusion should be changed when new data is obtained. Given some prior knowledge (K) and some new data (D) the probability of a set of conclusions (C) is given by

$$\mathcal{P}(C|D,K) = \frac{\mathcal{P}(D|C,K)\cdot\mathcal{P}(C|K)}{\mathcal{P}(D|K)} \tag{2.33}$$

where $\mathcal{P}(X|Y)$ denotes the probability of X given Y. In words the probability of the conclusions given the data is equal to the product of the prior probability of the conclusions and the ratio of the probability of observing the data if the conclusions are true to the prior probability of observing the data.

The crucial point here is that Bayes' theorem is only a prescription for updating probabilities in the light of new data: every Bayesian calculation must start from an assumed prior probability for the conclusions. If enough data is collected the final result can be largely independent of the assumptions used, but if the data is marginal the assigned probabilities are strongly affected by them. This is often stated as being a weakness of Bayesian methods, but supporters retort that the same problem exists in all other theories: it is simply particularly clearly displayed in the Bayesian formalism. Further discussion of this point, together with the issue of the meaning of "probability" is deferred to appendix C. It is also important to

note that Bayes' theorem makes universal claims: Bayes' method can be used to tackle all problems. In the language of chapter 1, it is applicable to problems where $N < P$, $N = P$ and $N > P$. Many conventional methods of data processing for the $N < P$ case can be seen as special cases of the general Bayesian solution.

Given some prior knowledge and some data, Bayes' theorem allows one to calculate a probability distribution for the set of all possible sets of conclusions. In general this probability distribution is difficult to visualise; consequently it is common to choose one set of conclusions, corresponding to the maximum probability, and to display this. Since the denominator in equation 2.33 is independent of the conclusions, it is sufficient to maximise the numerator. For notational convenience the set of conclusions will be referred to as a spectrum, but the following is not limited to cases where the set of conclusions is a spectrum.

The first term depends on the data, our model of the experiment, and our model of the experimental noise. From any spectrum $\boldsymbol{s}$ it is possible to calculate the data which would have been observed (in the absence of noise) if this were the correct spectrum: this is the trial data $\boldsymbol{t}$. The trial data can be subtracted from the experimental data $\boldsymbol{d}$ to give a set of residuals $\boldsymbol{r}$, which are non zero as a consequence of noise. If the noise is assumed to have a Gaussian distribution, as is generally the case, $\mathcal{P}(\boldsymbol{d}|\boldsymbol{s})$ is proportional to

$$\prod_i \exp\left(-\frac{r_i^2}{2\sigma_i^2}\right) = \exp\left(-\tfrac{1}{2}\sum_i \frac{r_i^2}{\sigma_i^2}\right) = \exp(-\tfrac{1}{2}\chi^2). \tag{2.34}$$

where σ_i is the estimated error in d_i.

The second term depends on our model of the experiment and our prior knowledge about the spectrum. In many cases the spectrum $\boldsymbol{s}$ represents a digitized function which is known to be everywhere positive. The problem may be further simplified by quantizing the elements of $\boldsymbol{s}$, so that $s_i = n_i\delta$, where the quantum δ is so small that this is a negligible approximation; the function is now represented by the vector of integers $\boldsymbol{n}$. Given M quanta it is possible to distribute them in a number of ways to get vectors $\boldsymbol{n}$, where $M = \sum n_i$. The probability of obtaining a

particular vector is given by

$$\mathcal{P}(\boldsymbol{n}) = 2^{-M} \frac{M!}{n_1! n_2! \cdots}. \tag{2.35}$$

Since δ is small, the n_i are large and Stirling's approximation

$$\log N! \approx x \log x - x \tag{2.36}$$

can be applied; hence

$$\mathcal{P}(\boldsymbol{n}) = 2^{-M} \exp\left(M \log M - M - \sum_i (n_i \log n_i - n_i) \right), \tag{2.37}$$

or

$$\mathcal{P}(\boldsymbol{n}) \propto \exp(\alpha S) \tag{2.38}$$

where

$$S = -\sum_i s_i \log(s_i/b) \tag{2.39}$$

and α and b are constants depending on M and δ. S is called the entropy of $\boldsymbol{s}$.

Combining these two terms gives

$$\mathcal{P}(\boldsymbol{s}|\boldsymbol{d}) \propto \exp(\alpha S - \tfrac{1}{2}\chi^2). \tag{2.40}$$

Maximising this is equivalent to maximising $\alpha S - \frac{1}{2}\chi^2$. The constants α and b must be chosen in some way: α is generally chosen to make χ^2 equal to the number of data points P, while b can be considered as defining the baseline level.

Information theory

Information theory was invented by Shannon as a way of defining the information content of a message[70], and hence the capacity of a message system to transmit information. The core of his idea is that a completely uniform signal carries no information, and that information is carried by deviations from uniformity. He developed an equation for the information content of a distribution on the basis of consistency arguments. For a distribution $\boldsymbol{p}$ of N elements p_i such that $\sum p_i = 1$ the information content is given by

$$\mathcal{I}(\boldsymbol{p}) = \sum_i p_i \log p_i. \tag{2.41}$$

Clearly $\mathcal{I}$ takes a minimum value of $-\log N$ for a uniform distribution ($p_i = 1/N$).

The information theory approach starts off by considering the set of all possible spectra $\{\boldsymbol{s}\}$. As before, every spectrum $\boldsymbol{s}$ has corresponding trial data $\boldsymbol{t}$. The trial data may be compared with the experimental data by means of a χ^2 statistic

$$\chi^2 = \sum_i \frac{(d_i - t_i)^2}{\sigma_i^2}. \tag{2.42}$$

A spectrum $\boldsymbol{s}$ is said to be inconsistent with the data if the corresponding value of χ^2 is too large; a typical criterion of consistency is that $\chi^2 \leq P$ where P is the number of data points. This leaves a very large number of consistent spectra, from which one must be chosen. This cannot be done on the basis of the data; it must be done on the basis of some prior preference.

One sensible approach is to choose the spectrum of minimum information content: this should be the safest choice, since all other choices introduce additional information which is not required by the data. Minimising equation 2.41 is clearly equivalent to maximising the entropy expression

$$S = -\sum_i s_i \log(s_i/b) \tag{2.43}$$

where b is now a normalisation constant rather than a baseline estimate.

Combining these results, the "best" spectrum is chosen by maximising S subject to the constraint $\chi^2 \leq P$. S is a convex function with a unique maximum, and the χ^2 constraint defines a hyper-ellipse in spectrum space; if the global maximum of S lies outside this hyper-ellipse, the constrained maximum will lie at its surface. The problem is now to maximise S subject to the constraint that $\chi^2 = P$; using Lagrange multipliers this is equivalent to maximising

$$S - \lambda\chi^2 \tag{2.44}$$

where λ is a Lagrange multiplier determined by $\chi^2 = P$. The analogy with equation 2.40 is obvious.

A pragmatic approach

The previous two derivations indicate the theoretical basis of the MEM in Bayesian reasoning and information theory. This final derivation is not a derivation in the

true sense of the word: the entropy function is plucked out of the air and used without justification. Nevertheless this derivation indicates important similarities between the MEM and other more conventional methods[71].

Many conventional data processing methods are equivalent to minimising a χ^2 parameter. As usual any spectrum $\boldsymbol{s}$ has corresponding trial data $\boldsymbol{t}$, which is compared with the experimental data. The solution is obtained by minimising χ^2 (equation 2.42). In many cases this equation does not possess a unique minimum, and the method breaks down; in other cases instability can result in noise being massively amplified. The conventional solution to this is to modify χ^2 by the addition of a regularization function, which adds a penalty for large deviations of the spectrum from zero. The most common form is a quadratic penalty, so that χ^2 is replaced by

$$\xi^2 = \sum_i \frac{(d_i - t_i)^2}{\sigma_i^2} + \alpha \sum_i s_i^2, \tag{2.45}$$

but other penalties are used, in particular the linear penalty $\alpha \sum_i |s_i|$.

It would be possible to use the entropy formula S as a regularization function by minimising

$$\xi^2 = \sum_i \frac{(d_i - t_i)^2}{\sigma_i^2} - \alpha S; \tag{2.46}$$

this is clearly equivalent to maximising equations 2.40 and 2.44. Once again there are parameters (α and b) which must be fixed in some way.

2.2.2 Implementation

Before discussing the actual implementation of the MEM, it is useful to narrow the range of problems being discussed. Firstly only spectroscopic and imaging problems will be discussed, that is the set of conclusions $\boldsymbol{s}$ is a spectrum or image (of arbitrary dimension) which has been sampled in some simple, regular fashion; for simplicity $\boldsymbol{s}$ will be always referred to as a spectrum. Secondly only linear problems will be discussed, that is problems where the data is a linear function of the spectrum. In this case we can write

$$t_i = \sum_j R_{ij} s_j, \tag{2.47}$$

or in more compact matrix notation $\boldsymbol{t} = \boldsymbol{R}\boldsymbol{s}$.

However the MEM is derived, the problem reduces to maximising

$$Q = S - \lambda\chi^2 \tag{2.48}$$

which is a constrained nonlinear optimisation problem. Because of the large sizes of the vectors $\boldsymbol{s}$ and $\boldsymbol{d}$ it is necessary to choose an efficient method of solving the problem. A practical algorithm has been developed by Skilling and Bryan[72]; this works by an iterative search using $\nabla S = \{\partial S/\partial s_j\}$, $\nabla\chi^2 = \{\partial\chi^2/\partial s_j\}$ and a weighted combination of these as search vectors.

The gradient of the entropy is simply a pointwise function of $\boldsymbol{s}$, but the gradient of χ^2 is more complicated:

$$\begin{aligned}
\frac{\partial\chi^2}{\partial s_j} &= \sum_i \left(\frac{\partial\chi^2}{\partial t_i}\right)\left(\frac{\partial t_i}{\partial s_j}\right) \\
&= \sum_i \frac{2\,(t_i - d_i)}{\sigma_i^2}\,R_{ij} \\
&= \sum_i 2\,R_{ij}\,\frac{t_i - d_i}{\sigma_i^2}
\end{aligned} \tag{2.49}$$

Note that the summation is over the *first* index of $\boldsymbol{R}$; feedback occurs via the transpose of $\boldsymbol{R}$, not its inverse. It is never necessary to calculate $\boldsymbol{R}^{-1}$; indeed it need not even exist.

Since the search is iterative it is necessary to have appropriate stopping criteria. Two criteria are necessary: firstly the vectors ∇S and $\nabla\chi^2$ must be close to antiparallel, and secondly χ^2 must be close to its target value. In practice the tests

$$\sqrt{\tfrac{1}{2}\left(1 + \frac{\nabla S\cdot\nabla\chi^2}{|\nabla S||\nabla\chi^2|}\right)} \leq 0.02 \tag{2.50}$$

and that χ^2 is within 1% of its desired value are adequate.

2.2.3 Application to NMR

The form of the MEM described above can only be used for spectra which are real and strictly positive; this is because the entropy expression is not defined for negative or complex intensities. This criterion is met by most images, and by

some spectra, but is not naturally met by NMR spectra; these are intrinsically complex, and the real part can contain both positive and negative intensity. There are three main approaches to this problem: (i) using the real part of the spectrum and enforcing positivity, (ii) using a number of real positive sub-spectra which are linearly combined and (iii) using a different entropy expression defined on the complex spectrum.

Enforcing positivity

Although an NMR spectrum is intrinsically complex, the real and imaginary parts are not independent, but are related by the Hilbert transform. Since all the information is contained in the real (or imaginary) part, it is sufficient to fix attention on the real part. In simple cases it is possible to correct the phase of the spectrum such that all the signals are in positive absorption mode in the real part; the real part is then strictly positive (neglecting the effects of noise). The phase corrected spectrum can be back transformed to give corresponding phase corrected data to which the MEM may be applied.

The problem with this approach is that it is restricted to simple cases. In many more complex NMR experiments it is impossible to phase the spectrum in this way. For example in inversion recovery experiments[4] and CIDNP experiments[73] signals may appear in emission (negative absorption), and in coherence transfer experiments signals may appear in dispersion. In such cases the method cannot really be applied. For strictly positive spectra, however, it works better than other methods and is used almost universally throughout this thesis.

Using sub-spectra

In some cases it is possible to divide the signals in an NMR spectrum into a small number of groups, each of which has its own intrinsic phase. For example, spectra from CIDNP experiments may contain emissive and absorptive lines. The emissive lines are simply in negative absorption mode; hence the spectrum can be described

by

$$s_i = s_i^p - s_i^n \tag{2.51}$$

where $\boldsymbol{s}^p$ is a sub-spectrum containing all the absorptive signals, and $\boldsymbol{s}^n$ is a sub-spectrum containing all the emissive (negative absorptive) lines. The sub-spectra $\boldsymbol{s}^p$ and $\boldsymbol{s}^n$ are themselves positive, but the spectrum may contain both positive and negative intensity. The entropy of the spectrum can be defined as the sum of the entropies of the sub-spectra[31], while χ^2 is calculated directly from the spectrum $\boldsymbol{s}$. At the end of the calculation it may be convenient to display either the spectrum, or the individual sub-spectra.

In simple cases this method can work reasonably well. This two sub-spectrum method can also be applied to spectra containing lines of arbitrary phase, but it performs poorly and has no clear theoretical basis. It is possible to extend the method by using more sub-spectra; in particular it is possible to describe many spectra using four sub-spectra corresponding to positive and negative absorption and dispersion modes. This approach is not used in this thesis.

Defining entropy on the complex spectrum

In many ways this is the most satisfactory approach since it allows spectra of arbitrary phase to be processed. Several different entropy expressions[32, 40, 41] have been suggested, on a variety of grounds, the most successful of which is due to Daniell and Hore[40]. Rather than considering the spectrum directly, they returned to the underlying physics of NMR.

After excitation a classical spin system can be completely described by the distribution of magnetization vectors $\mathcal{N}(\theta, \phi, \omega)$ where θ and ϕ are the initial orientation of a magnetization vector and ω is its precession frequency. From this distribution the transverse magnetization can be calculated at any time (neglecting relaxation); hence this distribution contains all the significant information about the system. Since it contains all this information, and since it is undeniably a probability distribution, it is possible to calculate the entropy over $\mathcal{N}$. It can be shown that this is equivalent to applying a different formula to the corresponding

complex spectrum; this formula can be called the spectral entropy as a convenient shorthand.

It is also possible to treat a quantum mechanical spin system by this means: the distribution $\mathcal{N}(\theta, \phi, \omega)$ is replaced by the corresponding density matrix. The overall result is similar, but depends on the spin quantum number I. In the limit that $I \to \infty$ the classical result is reproduced; for $I = \frac{1}{2}$ the entropy is given by

$$S_{1/2} = -\sum_i \left(z_i \log\left(z_i/b + \sqrt{1 + z_i^2/b^2} \right) - \sqrt{1 + z_i^2/b^2} \right) \tag{2.52}$$

where z_i is the modulus of the complex spectrum at point i and b is a baseline estimate.

This approach will be used in several places in this thesis.

2.2.4 Advantages of MEM processing

For some authors the philosophical basis of the MEM is sufficient to justify its use[74–77], but this opinion is not widely shared. The Fourier transform is simple, convenient and well understood, and any replacement must offer some real improvement. There are several areas in which the MEM might be useful.

Truncation artefacts

Many NMR signals are truncated, that is the FID has not effectively decayed to zero by the end of the acquisition period. Conventional data processing, involving zero-filling and Fourier transformation results in line broadening and wiggles around the lines. These artefacts arise because unmeasured Fourier components have been assumed to be zero. Conventional apodisation techniques can only smooth out the wiggles, at the expense of yet more line broadening; they cannot remove the artefacts entirely.

By contrast the MEM makes no simple assumptions about missing data points; instead it attempts to estimate them. Sinc wiggles apparently introduce additional structure into the spectrum, and so the entropy constraint should work against them. Similarly the positivity constraint (if present) will help suppress the wiggles,

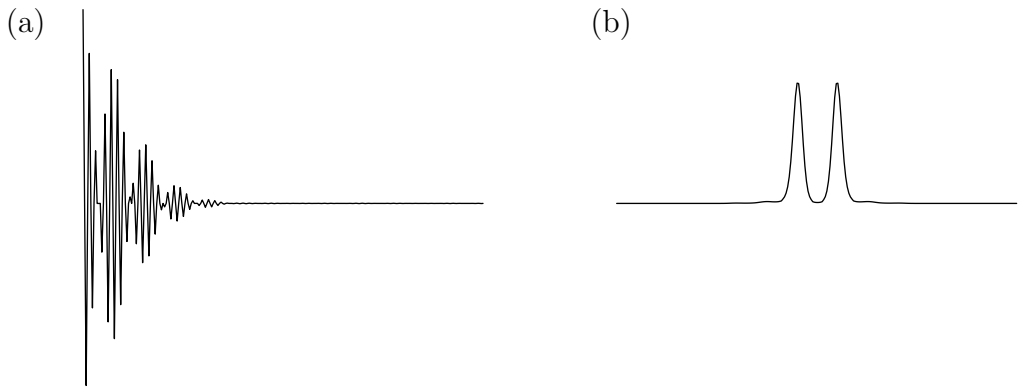

Figure 2.5: Suppression of truncation artefacts by the Maximum Entropy Method. The truncated data shown in figure 2.3e was processed using the MEM to give the spectrum (b); the corresponding trial data is shown in (a). Since the data used was noiseless, a value of 0.1% was arbitrarily assumed for the noise level.

since they give rise to negative regions.

Overall the MEM should go some way towards suppressing truncation artefacts, resulting in spectra containing narrower lines and fewer wiggles. This is demonstrated in figure 2.5, which should be compared with figure 2.3. In fact, while the wiggles are well suppressed, significant residual broadening remains.

Dispersion lineshapes

Dispersion lineshapes arise because NMR data is one sided, that is it only exists for times after excitation. This is, of course, a consequence of the causality of physical systems. It can, however, be considered as a form of truncation: if the MEM could in some way estimate this missing data, the dispersion lineshapes might be removed.

In the simplest form of the MEM, in which positivity is enforced, this is achieved in a trivial fashion. The MEM spectrum is real, not complex, and so its Fourier transform is conjugate symmetric: the data before the pulse is the complex conjugate of the data after the pulse. This is entirely appropriate if the phase of all the signals is zero.

The Daniell–Hore version of the MEM is more interesting. The spectrum is now complex, and the data before the pulse is not related to that after the pulse

in any simple way. Instead the method attempts to predict the missing data, and so dispersion lineshapes are partially removed. In practice, however, the method is not entirely successful, and some dispersion mode signal remains[47].

Resolution enhancement

Conventionally resolution enhancement is achieved by multiplying the data by a rising function, in order to attenuate the natural decay. This also has the adverse effects of enhancing noise in the tail of the FID and magnifying sinc-wiggles. Once again these can be considered as artefacts arising from altering the data.

The MEM approach to resolution enhancement is quite different. The trial spectrum is Fourier transformed to give trial data, which is then multiplied by a decaying function before being compared with the experimental data (figure 2.6); this allows lines to be narrowed without increasing the noise in the spectrum.

2.2.5 Transfer functions

Because of the non-linearity of the entropy expression, it is difficult to predict exactly what the MEM will do in any particular case. Nevertheless it is possible to analyse the method in a few simple cases[41, 45, 50, 71].

In many cases the noise is constant throughout the data, ie $\sigma_j = \sigma$. In this case the MEM solution lies at the global maximum of

$$Q = -\sum_{i=1}^{N} s_i \log(s_i/b) - \lambda \sum_{j=1}^{P} \left(d_j - \textstyle\sum_k R_{jk} s_k\right)^2 \tag{2.53}$$

where the factor of σ^2 has been absorbed into the λ. The simplest possible problem is when $\boldsymbol{R}$ is the identity matrix, ie $R_{jk} = \delta_{jk}$, so that

$$Q = -\sum_{i=1}^{N} \left(s_i \log(s_i/b) + \lambda (d_i - s_i)^2 \right). \tag{2.54}$$

The solution occurs at $(\partial Q/\partial s_i) = 0$ for each element, or

$$-1 - \log(s_i/b) - 2\lambda(d_i - s_i) = 0, \tag{2.55}$$

which can be solved directly. Clearly $\boldsymbol{s}$ can be obtained from $\boldsymbol{d}$ by pointwise non-linear scaling; the amount of non-linearity depends on λ, and hence on the

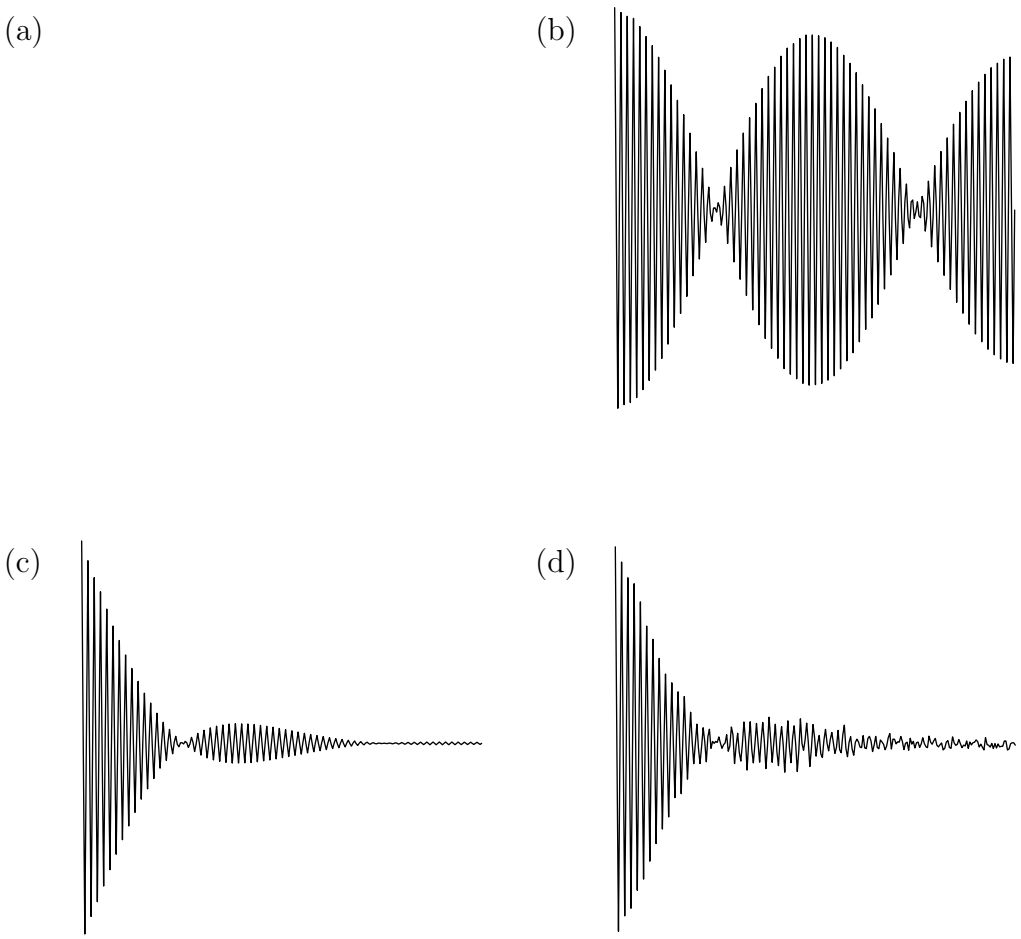

Figure 2.6: Resolution enhancement by the Maximum Entropy Method. A trial spectrum of narrow lines (a) has slowly decaying trial data (b); this is multiplied by a decaying exponential to give rapidly decaying trial data (c), which is then compared with the rapidly decaying experimental data (d). The data is taken from figure 2.4.

signal-to-noise ratio. In other words, there is a transfer function between the data and the MEM solution.

This result might seem trivial, but it demonstrates one case in which the method's behaviour is questionable. The problem described above is the problem of estimating an image given a noisy estimate of the image. The conventional solution would be take the noisy estimate itself as the best possible estimate of the underlying image; by contrast the MEM chooses to systematically distort this estimate.

The result can be extended to a larger class of problems, namely those where $\boldsymbol{R}$ is orthogonal. In this case $\boldsymbol{R}^{-1}$ exists (it is equal to $\tilde{\boldsymbol{R}}$) and so there is a conventional spectrum $\boldsymbol{c} = \boldsymbol{R}^{-1}\boldsymbol{d}$. Furthermore $\boldsymbol{R}$ will preserve Euclidean distances, so that χ^2 can be calculated between $\boldsymbol{c}$ and $\boldsymbol{s}$, instead of $\boldsymbol{d}$ and $\boldsymbol{t}$. Now

$$Q = -\sum_{i=1}^{N} \left(s_i \log(s_i/b) + \lambda(c_i - s_i)^2 \right), \tag{2.56}$$

which is the same problem as that in equation 2.54. Note that the orthogonality of $\boldsymbol{R}$ requires that $N = P$.

A number of important problems can be described in terms of orthogonal matrices: in particular the Fourier transform matrix is orthogonal. Consequently when the MEM is used as an alternative to the Fourier transform, transfer function behaviour can occur. If, however, the MEM is used to remove truncation artefacts or as a method of resolution enhancement, the matrix $\boldsymbol{R}$ is not a simple Fourier transform, and more complex behaviour can occur.

2.3 Other methods

In addition to the methods described above, there are many other data processing methods in use. Some of these are restricted to particular problems; others are of wider applicability. Two further methods will be considered here. Curve fitting, also called model fitting, is an extremely important technique in many fields, and is relatively well understood. By contrast, constrained minimisation (which includes

linear programming) is not widely applied to scientific problems, although it is important in economics.

2.3.1 Curve fitting

In many cases a conventional spectrum contains far more information than is actually required. If the spectrum contains a small number of sharp lines, the experimenter is generally interested not in the spectral intensity at all frequencies, but rather in the positions, intensities and widths of these lines.

A FID containing signals at M different frequencies can be described by

$$d(t) = \sum_{j=1}^{M} A_j e^{i\omega_j t} e^{-t/T_j} + n(t) \tag{2.57}$$

which contains $3M$ parameters (amplitude A_j, frequency ω_j and decay constant T_j for $j = 1, \ldots, M$) plus a noise term $n(t)$. A recorded FID of P points only contains $3M$ significant pieces of information: the remaining $P - 3M$ pieces of information simply describe the noise. It might be useful to extract the $3M$ interesting parameters directly from the data.

This can be done by curve fitting. For any set of parameters trial data can be calculated, which is compared with the experimental data by means of a χ^2 statistic. The set of parameters which minimises χ^2 is taken as the best estimate of the true parameters; if $3M < P$ this criterion is generally sufficient to define the parameters uniquely.

The major disadvantage of curve fitting is that a model must be assumed. If this model is incorrect, the results will reflect the inadequacies of the model as well as the information in the data. In extreme cases it can be difficult to formulate any sort of model at all. In general curve fitting is used as a second round of data processing: the data is initially processed to produce a spectrum from which a model and estimates of the relevant parameters are obtained.

It is also important to note that the model should ideally be fitted directly to the data, not to some spectrum obtained from it. If the spectrum is obtained by an orthogonal transformation the two methods are equivalent, but this is not true

in general.

Implementation

Since curve fitting is achieved by minimising χ^2, any general purpose minimisation routine can be used as the basis of a curve fitting routine. Since the parameter space may be of high dimension, it is important that an efficient routine is used. Several well known algorithms exist: the calculations in this thesis were performed using the SPIRAL algorithm[78].

While the global minimum is generally unique, the χ^2 surface may contain one or more local minima. The presence of these local minima is a major problem for any algorithm: there is no general solution for avoiding them. The problem can be minimised by the use of good starting values, obtained by some other means. In some cases it may be possible to use an exhaustive grid search, or a technique such as simulated annealing, to obtain them.

Linear prediction

Linear prediction[14, 65] is a data processing method which has been applied to NMR in recent years; it has been suggested as an alternative to the MEM for dealing with seriously truncated data sets. It exists in many variations, and because of the complexity of the technique it will not be discussed in this thesis. Broadly speaking it may be considered as a form of automatic curve fitting, in which the number of significant signals is estimated directly from the data. The major disadvantage of linear prediction, as in curve fitting, is the problem of using an inappropriate model.

Cramér–Rao theory

Despite the reservations expressed above, curve fitting is in many ways an ideal method of data processing. If the assumed model is really correct, then this information should be used in analysing the data. Parameters extracted by this means should be better estimates of the true parameters than those extracted in-

directly. Knowledge of the noise distribution might also allow the reliability of these estimates to be determined.

This is the basis of Cramér–Rao theory[79–81], which may be applied to any parametric technique. This theory gives a lower bound on the accuracy of any estimate of a parameter. It is only applicable if the assumed model function and noise distribution are correct, and so must be applied with care. Nevertheless in certain cases it provides a useful measure of reliability.

Suppose a data set $\boldsymbol{d}$ can be exactly represented by a model function $\boldsymbol{f}$ and noise $\boldsymbol{n}$ of known distribution, that is $\boldsymbol{d} = \boldsymbol{f} + \boldsymbol{n}$. These vectors will be assumed to be real (problems where the vectors are complex can generally be recast in terms of real vectors). If the noise has a Gaussian distribution, the probability distribution of a particular element is given by

$$\mathcal{P}(n_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{n_i^2}{2\sigma^2}\right), \tag{2.58}$$

but since the model is assumed to be exact n_i can be replaced by $d_i - f_i$. The probability distribution for the vector $\boldsymbol{n}$ is simply the product of the component probability distributions, and so

$$\mathcal{P}(\boldsymbol{d} - \boldsymbol{f}) = \prod_{i=1}^{P} \mathcal{P}(d_i - f_i). \tag{2.59}$$

This equation gives the probability of measuring data $\boldsymbol{d}$ given the parameters corresponding to $\boldsymbol{f}$. In order to estimate the parameters, it is reasonable to choose those for which this probability is a maximum; in this context equation 2.59 is called the likelihood function, and the parameter estimates are called the maximum likelihood estimates.

Maximising the likelihood function is the same as maximising the log-likelihood function $\mathcal{L}(\boldsymbol{d} - \boldsymbol{f}) = \log(\mathcal{P}(\boldsymbol{d} - \boldsymbol{f}))$; for Gaussian noise this is simpler since

$$\mathcal{L}(\boldsymbol{d} - \boldsymbol{f}) = -\frac{P}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{P}(d_i - f_i)^2. \tag{2.60}$$

Since the first term is constant, maximising equation 2.60 is equivalent to minimising the second term, which equals $\frac{1}{2}\chi^2$. In this case maximum likelihood estimation

is identical to the conventional approach. Of course, if the noise distribution is not Gaussian, the result will be different.

The theory of maximum likelihood estimation also leads to lower bounds on the errors for the parameter estimates; irrespective of the method used to estimate the parameters, these lower bounds cannot be exceeded. The following results will be stated without proofs, which may be found elsewhere[81].

Given a log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of model parameters (not to be confused with the model data $\boldsymbol{f}$), the Fisher information matrix $\boldsymbol{I}$ is defined by

$$I_{jk} = -\mathcal{E}\left(\frac{\partial^2 \mathcal{L}}{\partial\theta_j \partial\theta_k}\right) \tag{2.61}$$

where $\mathcal{E}$ indicates the expectation value, obtained by averaging over all possible values of the noise. For Gaussian noise this simplifies to

$$I_{jk} = \frac{1}{\sigma^2}\sum_{i=1}^{P}\left(\frac{\partial f_i}{\partial\theta_j}\frac{\partial f_i}{\partial\theta_k}\right), \tag{2.62}$$

which may be readily evaluated. Under certain (fairly unrestrictive) conditions the Cramér–Rao lower bounds on the standard deviation of parameter estimates are given by

$$\mathrm{SD}(\theta_j) \geq \sqrt{(I^{-1})_{jj}}. \tag{2.63}$$

In simple cases it is possible to obtain an analytical expression for these standard deviations, but in general they must be evaluated by computer. One simple result is that for Gaussian noise the standard deviations increase in proportion to the noise level σ, exactly as expected.

The utility of the Cramér–Rao result will be demonstrated in the following chapter, where the reliabilities of parameter estimates obtained by a variety of methods will be compared with the theoretical ideal.

2.3.2 Constrained Minimisation

Constrained minimisation is simply the minimisation of a function subject to certain constraints on its parameters: these constraints are typically linear equality or inequality constraints. When the function to be minimised is also a linear function

of the parameters, the problem is particularly simple. This technique, called linear programming, is of great importance in economics[82]. Algorithms exist for several other special cases; in the general case the problem may be solved by a normal minimisation algorithm, with special care taken to remain within the constraints. The calculations in this thesis were performed using the NAG routine `E04JAF`[83]; this is a general purpose routine, based on an iterative, quasi-Newton, search[65].

Constrained minimisation may be used in two ways in data processing. Firstly it may be used with a model fitting technique to constrain the parameters to have reasonable values. Secondly it may be used for spectral reconstruction, where (for example) the spectrum is constrained to be positive. In both cases the function minimised is usually χ^2, but other choices are possible.

An example of constrained spectral reconstruction will be found in chapter 5.

Chapter 3

Evaluation of the MEM

> "It's much better than mine," said Pooh admiringly, and he really thought it was.
> "Well," explained Eeyore modestly, "it was meant to be."
>
> A. A. Milne, "The House at Pooh Corner" (1928)

3.1 Criteria of spectral quality

In recent years there has been a great deal of interest in the MEM as a data processing method. Many papers have been published claiming spectacular improvements in spectral quality using maximum entropy techniques[16, 17, 30, 38, 84]. In general the criteria of spectral quality are not clearly defined; indeed they may amount to little more than the general appearance of the spectrum. These claims must be assessed with care.

3.1.1 Signal-to-noise ratio

In conventional NMR data processing, the most important criterion used is the signal-to-noise (S/N) ratio of a spectrum, defined as the ratio of the peak absorption mode signal to the root-mean-square (rms) value of the noise[4, 11]. The rms value of the noise is estimated using a region of baseline free of spectral features.

Using this criterion the MEM can undoubtedly improve spectral quality. Figure 3.1 shows the FT and MEM spectra calculated from a synthesized FID. These

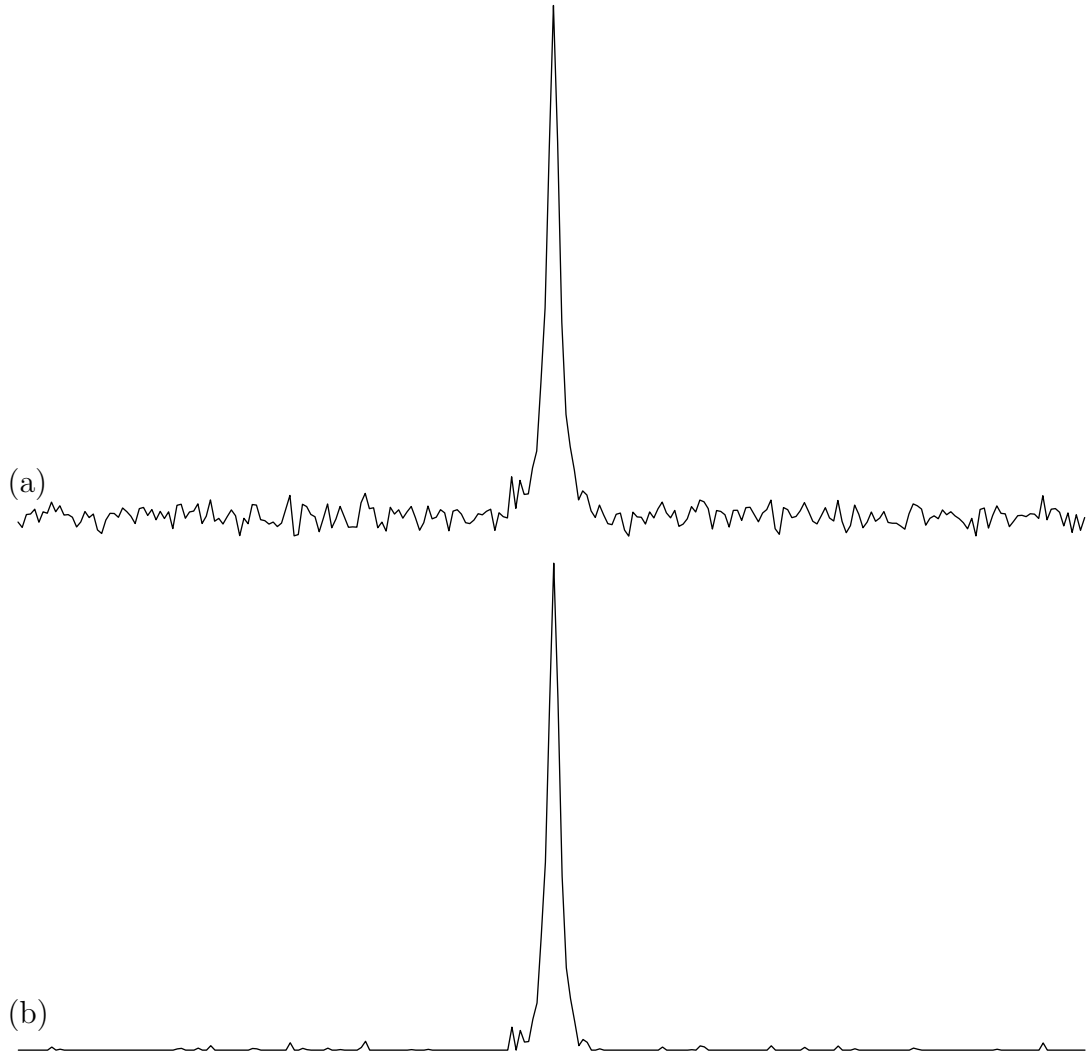

Figure 3.1: FT (a) and MEM (b) spectra obtained from a synthesized FID.

have S/N ratios of 63 and 463 for the FT and MEM spectra respectively. Clearly the MEM is capable of greatly reducing noise on the baseline, and so increasing the S/N ratio. It is unclear, however, whether this improvement is useful or merely cosmetic.

This point can be investigated by plotting the intensity of each point in the MEM spectrum against the intensity of the corresponding point in the FT spectrum, as is shown in figure 3.2. All the points lie on a single smooth curve, indicating that there is a transfer function between the MEM and FT spectra, as described in chapter 2. The MEM spectrum can be obtained directly from the FT spectrum by suitable non-linear amplification, and so the increase in S/N ratio is

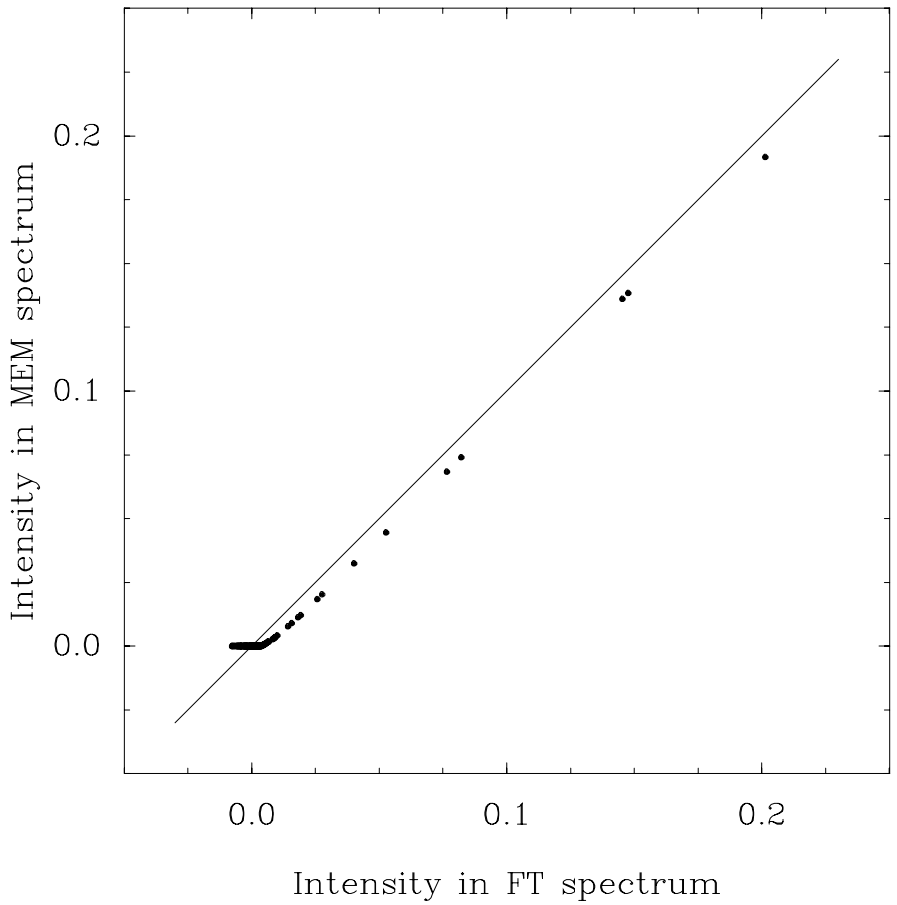

Figure 3.2: Plot of the intensity of each point in the MEM spectrum (figure 3.1b) against the intensity of the corresponding point in the FT spectrum (figure 3.1a). All the points lie on a smooth curve, indicating transfer function behaviour. At large intensities, this curve lies close to a straight line of gradient one.

purely cosmetic.

This example indicates the danger of using conventional concepts, such as signal-to-noise ratio, when dealing with unconventional data processing techniques. The Fourier transform is a linear technique: hence it is possible to consider signal and noise separately, combining the results at the end. One consequence of this is that noise on the baseline and noise on top of peaks will be treated identically, and so any technique that suppresses noise on the baseline will also suppress noise on top of peaks. By contrast the MEM is a non-linear technique: it is not possible to consider signal and noise separately. This means that suppression of noise on the baseline need not go hand in hand with suppression of noise on top of peaks[11].

3.1.2 Resemblance to the ideal spectrum

In the example above, the data used was synthetic, but this was not important to the argument. There are, however, a number of techniques for estimating spectral quality which can only be used with synthetic data. These techniques rely on knowledge of the ideal spectrum.

From any spectrum it is possible to calculate corresponding data. This data may then be corrupted in some desired manner, for example by the addition of noise. The corrupted data can be processed by a variety of methods to give a variety of spectra; these may then be compared to the ideal spectrum, which is simply the original spectrum with which the process started.

Obviously this comparison can be performed in a variety of ways. Lieu *et al* suggested a way of comparing two data processing methods[85]. An ideal spectrum $\boldsymbol{s}$ is used as a source of data from which two spectra are calculated: the conventional method is used to calculate $\boldsymbol{c}$, while an alternative method is used to calculate $\boldsymbol{a}$. They defined

$$A = \frac{\|\boldsymbol{c} - \boldsymbol{s}\|_2^2 - \|\boldsymbol{a} - \boldsymbol{s}\|_2^2}{\|\boldsymbol{c} - \boldsymbol{s}\|_2^2}, \tag{3.1}$$

where $\|\boldsymbol{x}\|_2$ is the l_2-norm of $\boldsymbol{x}$ defined by

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_i x_i^2}. \tag{3.2}$$

Clearly alternative data processing methods can be classified (relative to the conventional method) according to the sign of A: a value of $A > 0$ indicates that the alternative method reproduces the ideal spectrum more closely than the conventional method, while a value of $A < 0$ indicates the opposite.

This technique can be applied to the spectra shown in figure 3.1, taking the FT as the conventional method and the MEM as the alternative method. Calculated over the whole spectrum $A = 0.62$, indicating that the MEM has reproduced the ideal spectrum more accurately than the FT has. However this improvement is not quite as good as it seems. Figure 3.3 shows the variation in the A value when it is calculated over different parts of the spectrum (in each case the calculation range is centred on the line). Clearly the value is strongly dependent on the calculation

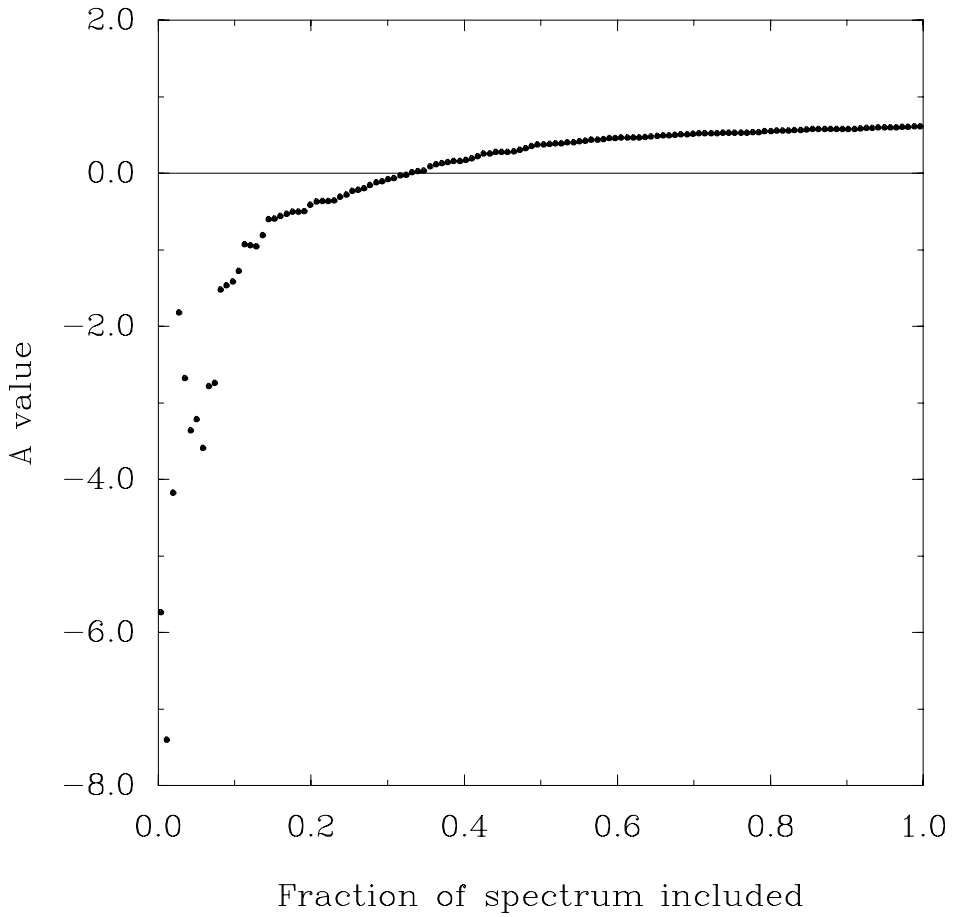

Figure 3.3: Plot of the Lieu A value for the spectra in figure 3.1 as a function of the calculation range.

range; in particular A increases as the range is increased.

This is easily understood. The transfer function in figure 3.2 shows that the MEM shrinks small intensities towards zero: noise on the baseline is well suppressed and so the flat baseline of the ideal spectrum is modelled extremely well. Consequently the A value is positive when large regions of baseline are included. By contrast the same shrinkage applied to points on the spectral lines distorts them away from the ideal, resulting in a negative A value.

This behaviour is not confined to the MEM: it has been shown[71] that many data processing methods can give improvements in A values as long as the spectrum is "nearly black", that is as long as it consists mostly of baseline with narrow isolated features. When the spectrum is not nearly black, that is it contains many features or the features are broad, the performance of the MEM and of these other

methods is greatly impaired. This fact is of considerable importance in NMR spectroscopy: many experimental spectra contain baseline distortions, which may be considered as extremely broad spectral features.

3.2 Reliability of integrals

The section above indicates that simple measures of spectral quality must be applied with great caution: the quality may depend as much on the specific spectrum studied as on the data processing method used, and apparent improvements may be merely apparent. A good measure of spectral quality is one sensitive to the types of information the experiment is actually supposed to measure.

In many NMR experiments the integral of a peak or group of peaks is measured; consequently the reliability with which these integrals may be determined is of prime importance. Here reliability means the ratio of a parameter to the uncertainty in its value. For a linear data processing method reliabilities can be estimated directly: for example the reliability of a line height is equal to its signal-to-noise ratio. For non-linear methods, however, this is impossible. The noise on the peaks cannot easily be measured, and so the uncertainty in peak height is unknown.

3.2.1 A Monte Carlo approach

While the noise on the peaks cannot be measured directly, it can be estimated by a statistical Monte Carlo approach. Starting from a single ideal spectrum it is possible to calculate many datasets which are corrupted by different noise. Each of these datasets may be processed to give a spectrum, and the parameters of interest (for example integrals) extracted. The mean of these measurements gives the best estimate of the parameter, while their standard deviation gives an estimate of the uncertainty in the parameter. As before, the ratio of these values gives the reliability[49, 51, 54, 86].

This technique was used to compare the MEM with Fourier transformation.

	mean	SD	reliability
FT	0.939	0.0185	50.7
MEM	0.785	0.0203	38.6

Table 3.1: Mean, standard devation and reliability of integrals.

A FID of 128 complex points was synthesized using $s(t) = e^{i\pi t}e^{-t/25.6}$; this corresponds to a single line at half the Nyquist frequency, with an initial intensity of unity and a final intensity of e^{-5}. This was used to form 256 noisy datasets, each containing different pseudorandom noise, taken from a near Gaussian distribution[1] with mean zero and standard deviation 0.04. Each dataset was processed by Fourier transformation and by the MEM; figure 3.1 shows typical spectra.

Integrals were calculated for each spectrum by the trapezium rule; the integration range was 16 points (approximately 4 linewidths) on each side of the line centre. The results are listed in table 3.1.

In both cases the mean integral is too low (the ideal value is unity). For integrals determined from FT spectra, this is explained by the limited integration range: Lorentzian lineshapes fall off very slowly at large distances from the line centre, and so contain large amounts of intensity in the wings. For integrals determined from MEM spectra this effect is reinforced by the tendency of the method to shrink intensities towards the baseline.

The standard deviations are also interesting: integrals determined from MEM spectra are more variable than those determined from the corresponding FT spectra. Rather than suppressing noise, the MEM is actually increasing it.

3.2.2 Other Methods

The Monte Carlo technique described above is entirely general, and may be used to investigate any data processing method. Many variations on Fourier transform and MEM processing have been suggested; of these six were investigated by this

[1]Pseudorandom noise with a rectangular $[0 \cdots 1)$ distribution is easily obtained; the sum of 12 samples has a near Gaussian distribution of mean six and standard deviation unity.

technique[51]. In each case the dataset described above was processed to give a spectrum of 256 points. The six methods investigated were as follows (the abbreviation used to identify each method in the subsequent discussion is given in parentheses).

1. *Fourier transformation (FT).* Each FID was zero-filled to 256 complex points, Fourier transformed, and the imaginary part discarded.

2. *Fourier transformation with matched filtration (MF).* As FT, except that each FID was multiplied by a matched exponential decay before Fourier transformation.

3. *Strictly positive maximum entropy (S).* This is the MEM in its simplest form. The spectrum is constrained to be real and strictly positive, allowing equation 2.39 to be used for the spectral entropy.

4. *Complex maximum entropy ($S_{1/2}$).* This is the more general form of the MEM, where the spectrum is complex and equation 2.52 is used for the spectral entropy.

5. *Strictly positive maximum entropy with resolution enhancement (S-RE).* Like method S, except that resolution enhancement is applied.

6. *Complex maximum entropy with resolution enhancement ($S_{1/2}$-RE).* Like method $S_{1/2}$ except that resolution enhancement is applied.

A typical spectrum for each method is shown in figure 3.4.

Results

Each spectrum was analysed as before, except that the integration range used was ± 32 points for MF, ± 16 points for FT, S, and $S_{1/2}$, and ± 4 points for S-RE and $S_{1/2}$-RE; in each case this corresponds to approximately four linewidths on each side of the line. In addition the FT and MF spectra were analysed using curve fitting to a single Lorentzian line. This was done in two ways: first with the

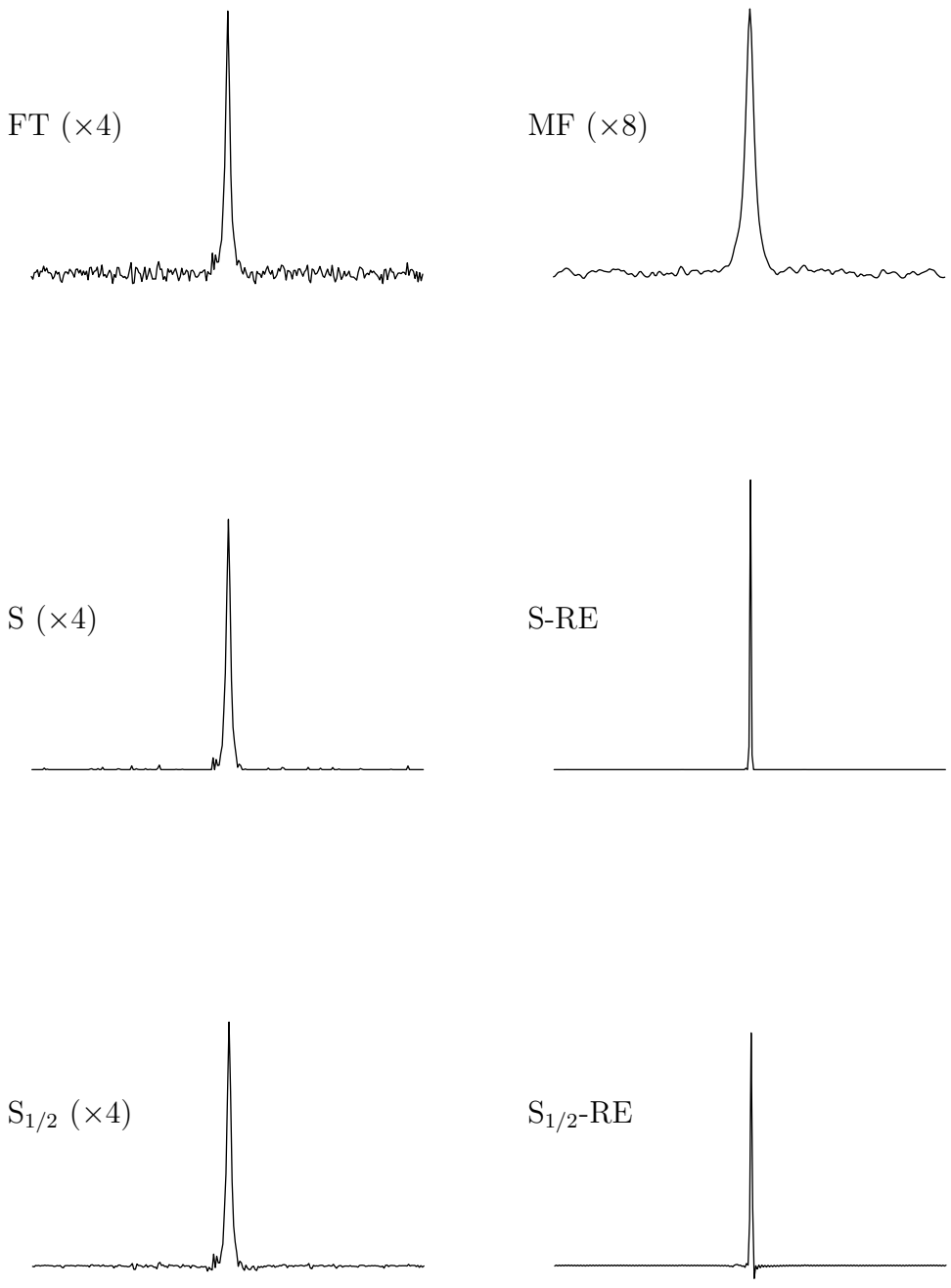

Figure 3.4: Typical spectra from each of the six methods described in the text; each spectrum is labelled by its abbreviation. Note that some of the spectra have been expanded vertically by a factor of 4 or 8 as indicated.

method	mean	SD
FT	0.939	0.0185
MF	0.941	0.0235
S	0.785	0.0203
S-RE	0.981	0.0168
$S_{1/2}$	0.713	0.0151
$S_{1/2}$-RE	0.970	0.0155
FT-CF	1.001	0.0147
FT-CFW	1.001	0.0110
MF-CF	1.003	0.0169
MF-CFW	1.001	0.0118

Table 3.2: Mean and standard deviation of line integral from data processed by a variety of methods.

position, height and width of the line as free variables, and second with the width fixed at its correct value and only the position and height left free. These estimates are denoted FT-CF (curve fitting) and FT-CFW (curve fitting, width fixed) for the Fourier transform spectra, and MF-CF and MF-CFW for the matched filtered spectra. No attempt was made to analyse the other four spectra by curve fitting, as their lineshapes were unpredictable. The results are listed in table 3.2 and plotted in figure 3.5.

Discussion

Several points emerge from figure 3.5.

1. As described above, the simple maximum entropy spectra (S) are of lower quality than the Fourier transform spectra (FT). The MEM underestimates the line integral *and* gives a greater standard deviation.

2. Somewhat surprisingly the integrals determined from matched filtered spectra (MF) have a higher standard deviation than those determined from unfiltered spectra (FT), even though the matched filtered spectra have lower baseline noise. This is a consequence of the larger integration range needed for the broader matched filtered lines[87]. Similarly integrals determined by curve fitting are more variable when the matched filtered spectra are used

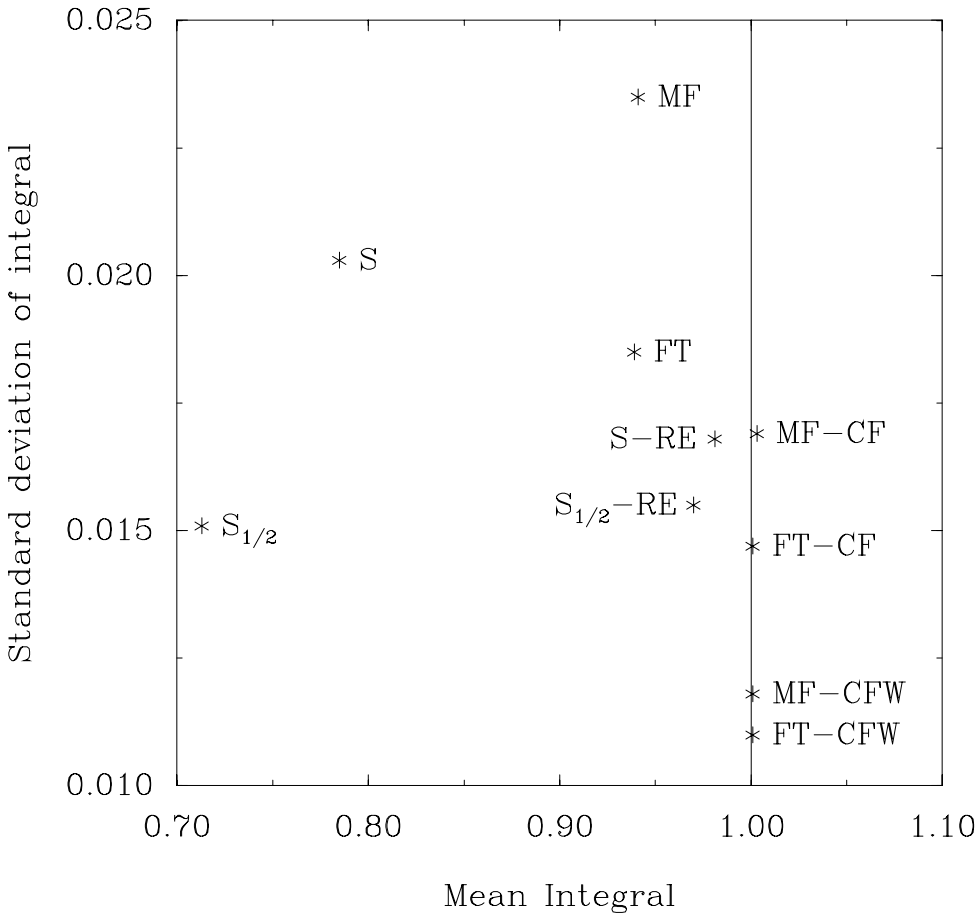

Figure 3.5: Plot of standard deviation against mean for line integrals from data processed by a variety of methods (see table 3.2). The vertical line indicates the true integral (1.0); the best methods give points lying close to this line and towards the bottom of the graph.

(MF-CF and MF-CFW) than when the unfiltered spectra are used (FT-CF and FT-CFW).

3. Complex maximum entropy ($S_{1/2}$) gives less variable integrals than strictly positive maximum entropy (S); indeed the integrals are less variable than for the Fourier transform. The integrals are, however, extremely low.

4. Maximum entropy processing performs better (the integrals are higher and their standard deviation lower) when resolution enhancement is used: compare S-RE with S and $S_{1/2}$-RE with $S_{1/2}$. Indeed these methods do better than any other method except for curve fitting. This must be a result of including information on the lineshape and linewidth into the calculation:

improvements in spectral quality can *only* come from the inclusion of prior knowledge.

5. Curve fitting performs the best of all; it is also the method in which the most prior knowledge is used. Fixing the linewidth at its correct value (FT-CFW), effectively including more information, reduces the standard deviation of the integral compared to letting the linewidth run free (FT-CF).

For comparison the Cramér–Rao bound on the standard deviation of the line integral is 0.0152 when the linewidth is unknown and 0.0110 when it is fixed, showing that curve fitting to the unfiltered spectrum effectively reaches the Cramér–Rao bound[2].

3.2.3 Truncation

In order to investigate the effects of truncation, the calculations described above were repeated using only the first 26 complex points from each dataset, so that the final signal intensity was $1/e$; spectra of 256 points were obtained from each dataset. The resulting Fourier transform spectra are severely corrupted by sinc wiggles, and so curve fitting was performed using the convolution of a Lorentzian lineshape and the appropriate sinc function.

Typical spectra from each method are plotted in figure 3.6. As expected the simple Fourier transform spectrum (FT) has severe sinc-wiggle artefacts; these are partially suppressed by matched filtration (MF). All the maximum entropy methods perform well at suppressing the wiggles. Strictly positive maximum entropy succeeds as a method of resolution enhancement (S-RE) but complex maximum entropy performs much less well ($S_{1/2}$-RE).

Results and discussion

The Monte Carlo results are listed in table 3.3 and plotted in figure 3.7. Three

[2]Remember that the values in table 3.2 are only estimates: they are accurate to about one part in $2\sqrt{N}$ where N is the number of datasets used. For $N = 256$ this corresponds to an error of approximately $\pm 3\%$.

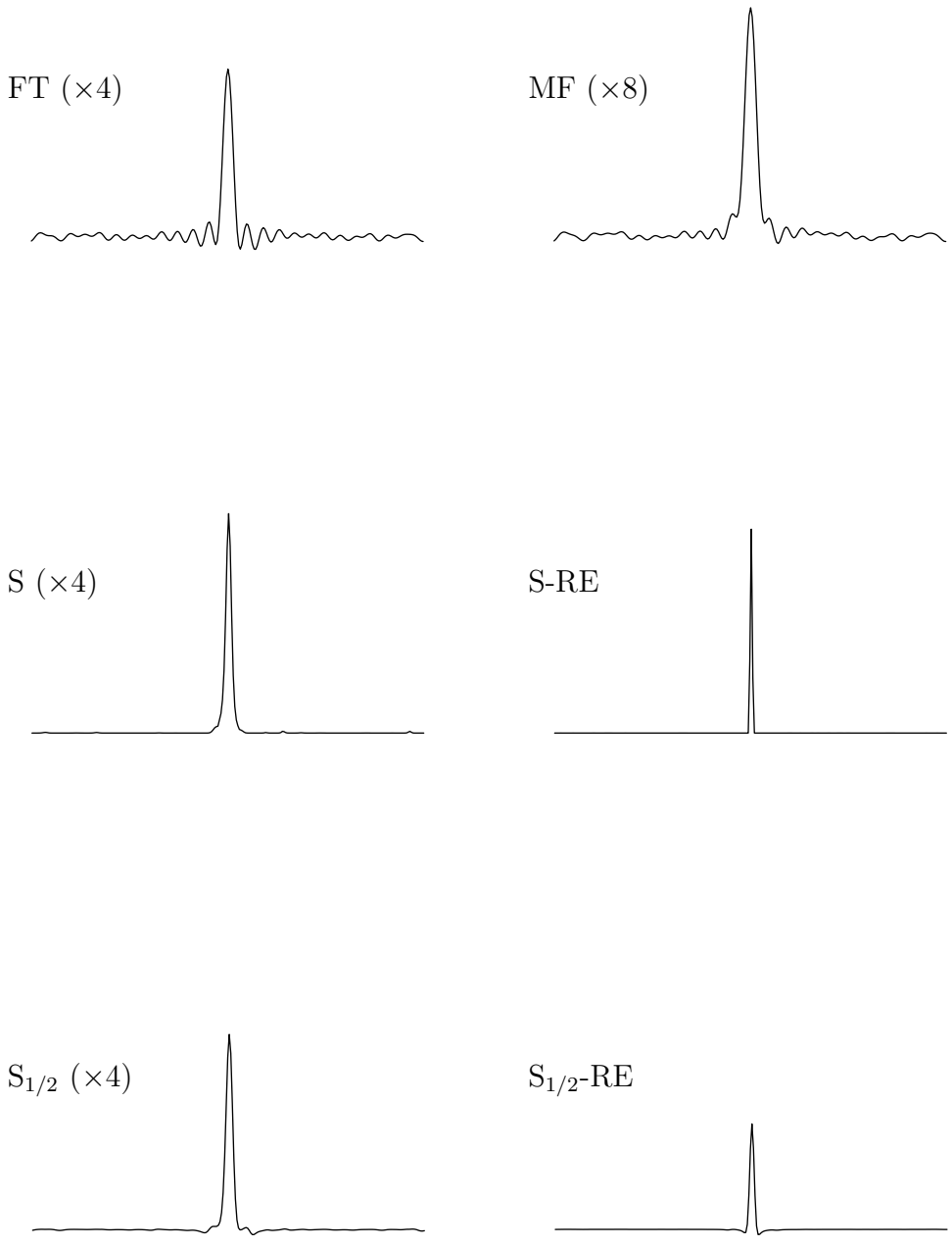

Figure 3.6: Typical spectra from truncated data for each of the six methods described in the text; each spectrum is labelled by its abbreviation. Note that some of the spectra have been expanded vertically by a factor of 4 or 8 as indicated.

method	mean	SD
FT	0.954	0.0183
MF	0.940	0.0235
S	0.855	0.0212
S-RE	0.992	0.0144
$S_{1/2}$	0.807	0.0178
$S_{1/2}$-RE	1.002	0.0148
FT-CF	1.011	0.0176
FT-CFW	0.997	0.0117
MF-CF	1.006	0.0182
MF-CFW	0.983	0.0118

Table 3.3: Mean and standard deviation of line integral from truncated data processed by a variety of methods.

points emerge from a comparison with figure 3.5.

1. The broad pattern of results is similar for truncated and untruncated data.

2. In general the integrals are similar or slightly better (that is, closer to the true value with a smaller standard deviation). This is because the shorter FIDs contain a higher proportion of signal relative to noise: in the long FIDs the signal has effectively decayed to zero well before the end of the FID, but the noise remains at the same level.

3. The major exception to the above is for methods MF-CF and FT-CF. These methods use information on the lineshape but not the linewidth. This result probably reflects the difficulty of estimating linewidths from severely truncated data.

3.3 Reliability of line heights

While line integrals are generally more useful than line heights when analysing data, line heights are immediately obvious from the spectrum and so are very important in the initial stages of analysis. The reliability of line heights might be expected to follow the reliability of line integrals, but this is not so. Reliabilities for untruncated data are listed in table 3.4 and plotted in figure 3.8. The "linewidth"

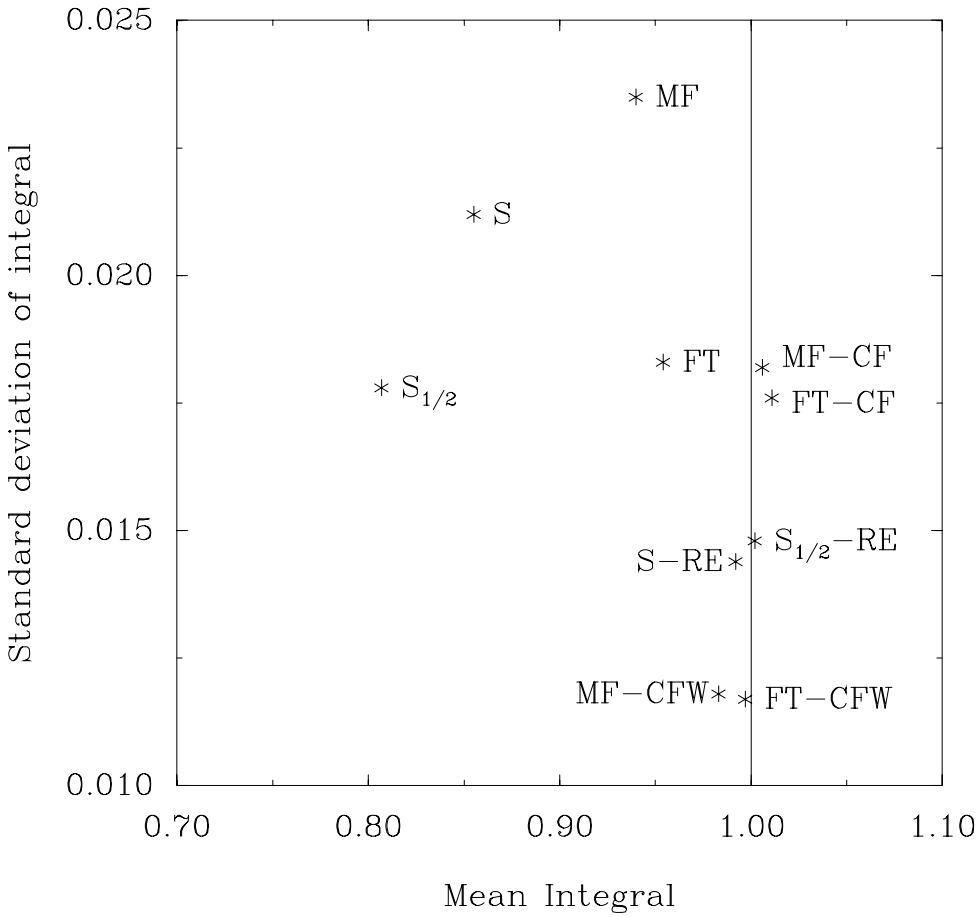

Figure 3.7: Plot of standard deviation against mean for line integrals from truncated data processed by a variety of methods (see table 3.3). The vertical line indicates the true integral (1.0); the best methods give points lying close to this line and towards the bottom of the graph.

parameter used in the table and figure is defined as the ratio of the mean integral to the mean height for the method. This is a robust definition of linewidth, which may be applied to lines of any shape.

The major trend visible in this data is for the reliability of the line height to decrease as the linewidth is decreased: even though method S-RE gives very reliable integrals, the corresponding heights are very unreliable. This is a result of a relatively large uncertainty in the width of very narrow lines.

Truncation

The results for truncated data are listed in table 3.5 and plotted in figure 3.9. Once more the results are broadly similar to those for untruncated data: the reliability of

method	width	reliability
FT	4.73	57.2
MF	9.40	91.1
S	4.13	52.1
S-RE	1.14	12.2
$S_{1/2}$	3.87	53.0
$S_{1/2}$-RE	1.35	16.3
FT-CF	5.00	62.5
FT-CFW	5.00	91.1
MF-CF	10.02	80.0
MF-CFW	10.00	84.7

Table 3.4: Reliability of line height and linewidth parameter from data processed by a variety of methods.

method	width	reliability
FT	7.56	79.2
MF	10.87	84.9
S	5.28	18.6
S-RE	1.88	7.0
$S_{1/2}$	5.44	41.7
$S_{1/2}$-RE	3.02	21.2
FT-CF	5.20	32.9
FT-CFW	5.00	84.8
MF-CF	10.11	64.8
MF-CFW	10.00	83.1

Table 3.5: Reliability of line height and linewidth parameter from truncated data processed by a variety of methods.

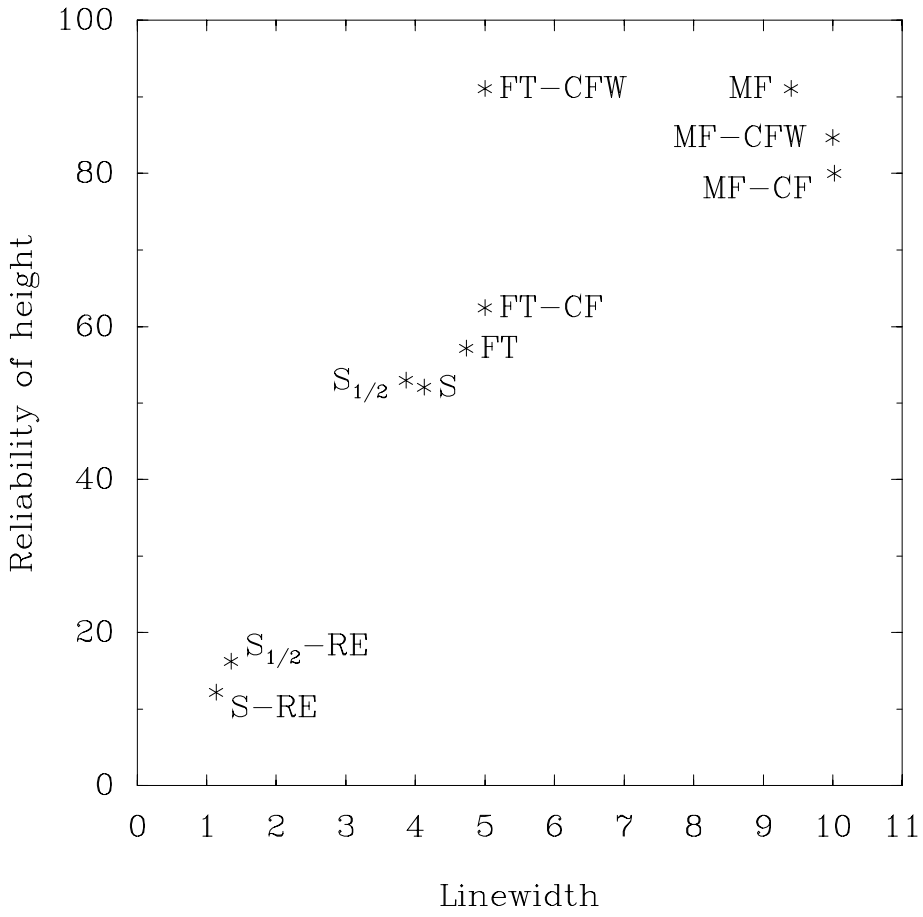

Figure 3.8: Plot of reliability of line height against linewidth for data processed by a variety of methods (see table 3.4).

the line height tends to decrease as the linewidth is reduced. The major difference is that the lines are broader, reflecting imperfect suppression of truncation artefacts.

3.4 Resolution enhancement

3.4.1 Partial resolution enhancement

The results above show that the use of resolution enhancement greatly increases the ability of the MEM to extract integrals from NMR data. This does, however, require the lineshape and linewidth to be known. With genuine data the linewidth will not be known accurately, and the lineshape need not be a perfect Lorentzian. Furthermore different lines may well have different linewidths, and it may not be possible to choose one resolution enhancement parameter appropriate to the whole

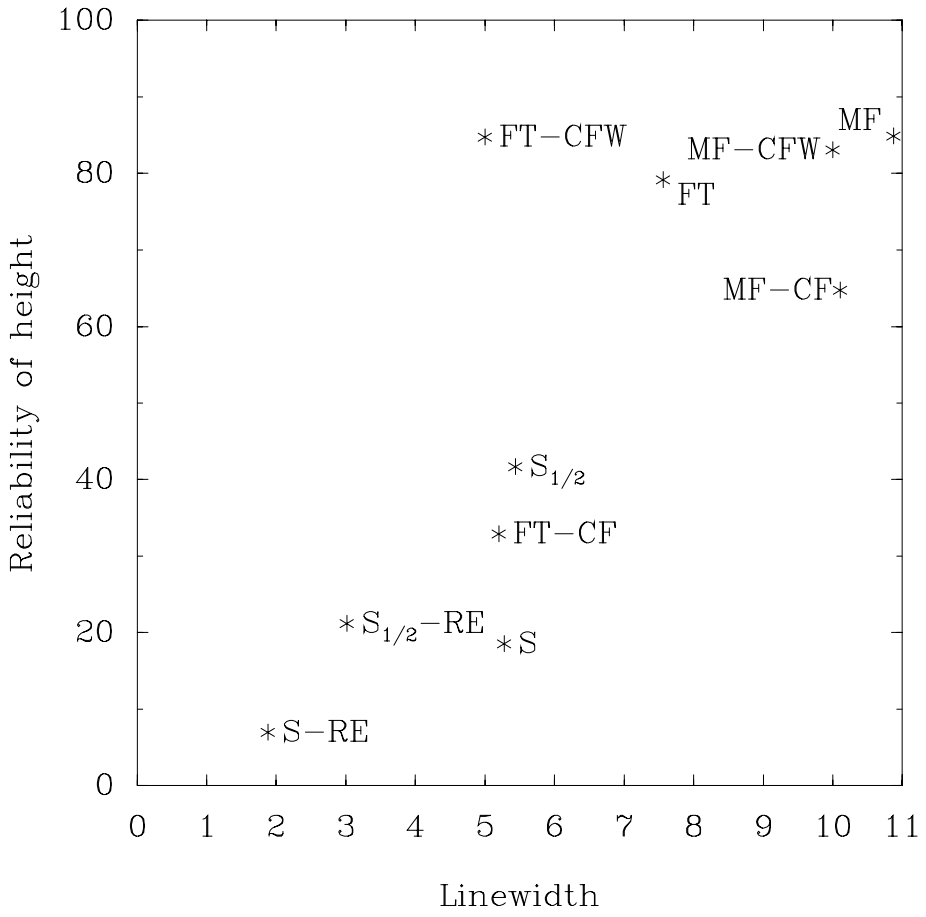

Figure 3.9: Plot of reliability of line height against linewidth for truncated data processed by a variety of methods (see table 3.5).

spectrum. Consequently it is important to study the effect of partial resolution enhancement. It is also important to compare resolution enhancement by the MEM with more conventional resolution enhancement methods.

In order to examine these points the untruncated datasets were processed by three different methods:

1. Fourier transformation with resolution enhancement using a rising exponential as the weighting function;

2. Fourier transformation with resolution enhancement using a Lorentzian to Gaussian transformation as the weighting function;

3. Maximum entropy processing.

In each case four different weighting functions were used, in order to remove ap-

method	width	reliability
FT-E	4.73	57.2
	3.61	35.9
	2.57	19.8
	1.01	6.2
FT-LG	4.13	76.0
	3.36	65.8
	2.59	46.1
	1.15	7.8
S-RE	4.14	51.9
	3.31	34.1
	2.44	21.9
	1.15	15.3

Table 3.6: Reliability of line height and linewidth parameter from data processed by a variety of methods. FT-E indicates Fourier transformation using a rising exponential weighting function, FT-LG indicates Fourier transformation using a Lorentzian to Gaussian transform as weighting function and S-RE indicates the MEM with resolution enhancement.

proximately 0, 25, 50 and 100% of the linewidth. Typical spectra are shown in figure 3.10.

When the spectrum is resolution enhanced by conventional methods, the baseline noise is greatly magnified resulting in a decreased signal-to-noise ratio, but when the spectrum is resolution enhanced by the MEM, baseline noise is actually suppressed as a result of the inclusion of information about the lineshape in the processing: the spectrum with the narrowest lines also has the lowest baseline noise level. As usual this suppression of baseline noise need not indicate that line heights may be determined more reliably by this method; it is necessary to perform a Monte Carlo study to examine this.

The results of such a study are listed in table 3.6 and plotted in figure 3.11. For all three methods the reliability of the line height decreases as the linewidth is reduced. The MEM (S-RE) follows simple resolution enhancement (FT-E) quite closely while the Lorentzian to Gaussian transform (FT-LG) gives more reliable heights than either of the other two when the linewidths are only slightly reduced. The MEM gives the best results when complete resolution enhancement is attempted.

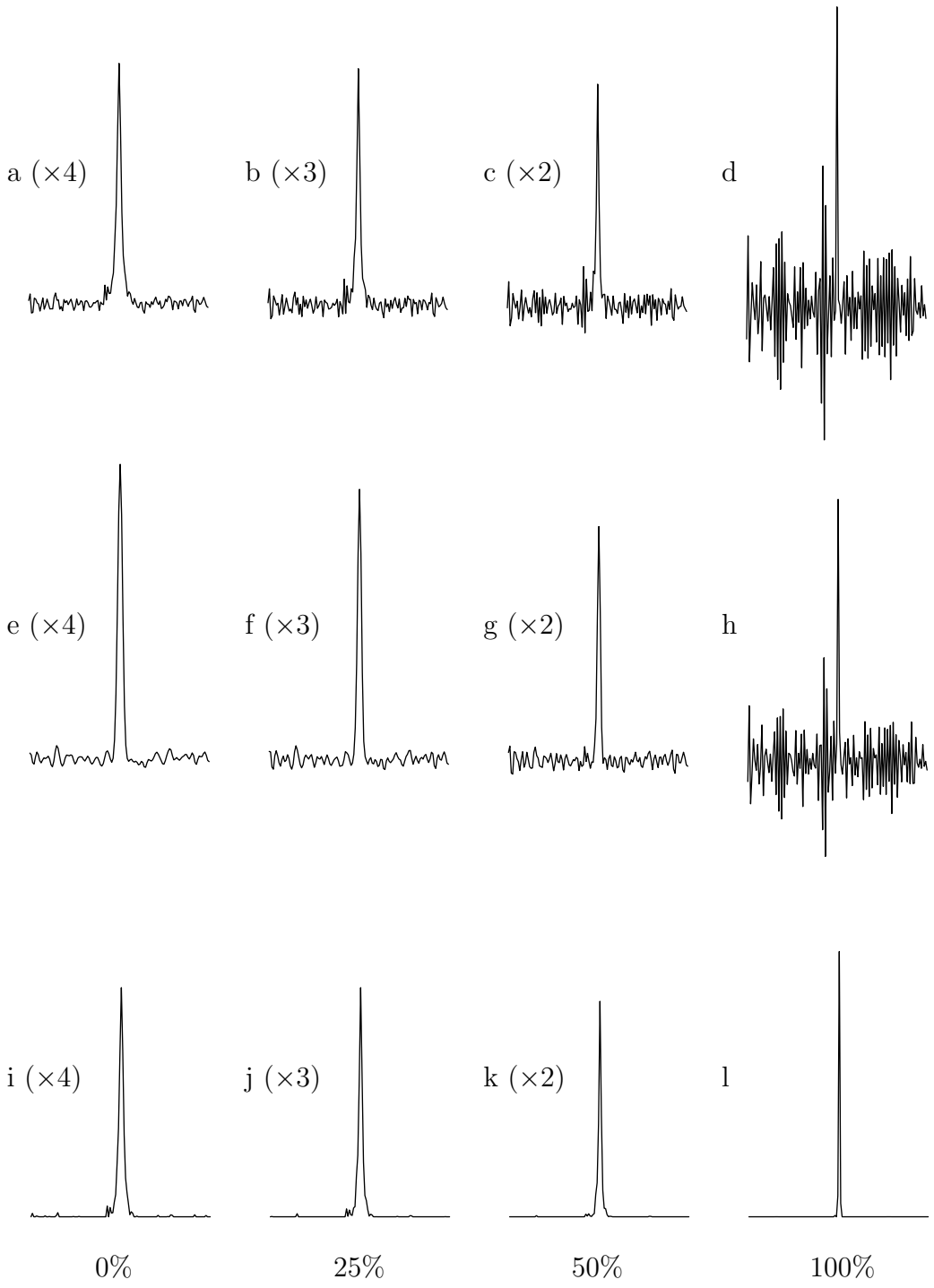

Figure 3.10: Typical resolution enhanced spectra obtained as described in the text; note that only the central half of each spectrum is shown and that some of the spectra have been expanded vertically by factors of 2, 3 and 4 as indicated. Spectra (a)–(d) were obtained by Fourier transformation with a rising exponential as weighting function, (e)–(h) were obtained by Fourier transformation with a Lorentzian to Gaussian transform, and (i)–(l) were obtained using the MEM. Nominal linewidth reductions are indicated at the bottom of each column.

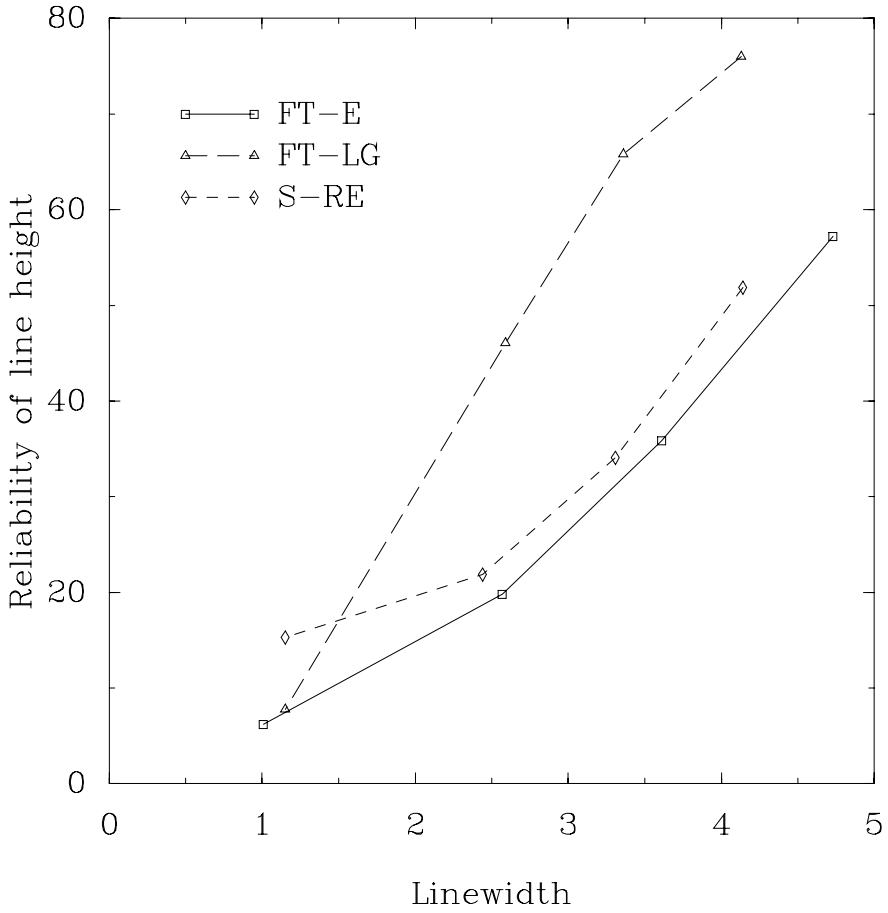

Figure 3.11: Plot of reliability of line height against linewidth for truncated data processed by a variety of methods (see table 3.6).

3.4.2 Two line spectra

Strictly speaking the results above have not actually demonstrated resolution enhancement; they have only demonstrated linewidth reduction. Linewidth is a property of a single line, while resolution refers to the ability to distinguish between intensity arising from two or more closely spaced lines. Just as a non-linear method can suppress baseline noise without increasing the accuracy with which integrals can be measured, such a method can also decrease linewidths without necessarily increasing resolution. For example consider a spectrum containing two lines so close together that they are completely unresolved. Appropriate non-linear amplification of such a spectrum can decrease the width of the combined line by any desired amount, without increasing the resolution at all.

In order to examine this point it is necessary to study a dataset containing at

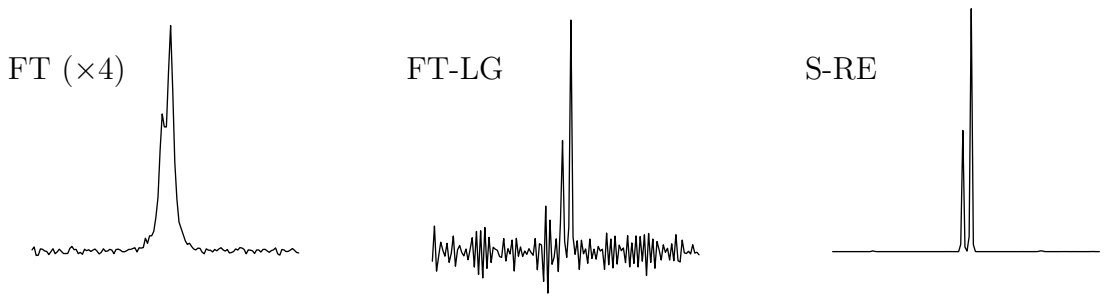

Figure 3.12: Typical spectra from a dataset corresponding to two closely spaced lines processed by various methods. Note that only the central half of each spectrum is shown, and that spectrum FT has been expanded vertically by a factor of 4.

least two lines. 256 datasets were synthesised corresponding to two lines which are barely resolved; these lines have the same width and have intensities in the ratio 2 to 1. Each dataset was processed by simple Fourier transformation, Fourier transformation with a Lorentzian to Gaussian transform, and the MEM with resolution enhancement. Typical spectra are shown in figure 3.12. Both Lorentzian to Gaussian transformation and the MEM result in improved resolution, and the corresponding spectra can be integrated directly. The simple Fourier transform spectrum cannot be integrated directly because of the extensive overlap. It can however be analysed by curve fitting. As usual curve fitting was performed both with the widths allowed to vary freely and with the widths fixed at their correct values; in addition curve fitting was also performed with the linewidths constrained to have the same value (this value being allowed to vary freely). This estimate is denoted FT-CFS. The results are listed in table 3.7.

The means of the integrals obtained from the MEM spectra are systematically low, as found for the single line spectra discussed above, but the ratio of the two means (2.02) is close to the theoretical value. The standard deviations increase in the order FT-CFW $<$ FT-CFS $<$ S-RE $<$ FT-CF $\ll$ FT-LG. The results for methods S-RE and FT-CFW are close to the corresponding results for single line spectra, but method FT-CF performs much less well. Comparison with method FT-CFS suggests that this arises from difficulties in determining the linewidths of

method	mean	SD
FT-CF	1.002	0.0300
	2.000	0.0296
FT-CFW	1.002	0.0129
	2.000	0.0132
FT-CFS	1.001	0.0130
	2.006	0.0155
FT-LG	1.025	0.0810
	2.022	0.0803
S-RE	0.984	0.0150
	1.989	0.0158

Table 3.7: Mean and standard deviation of the integrals of a two line dataset (see figure 3.12).

severely overlapped lines. Unsurprisingly Lorentzian to Gaussian transformation of sufficient severity to resolve the two lines down to the baseline results in large noise amplification and poorly determined integrals.

3.5 Conclusions

The results above have demonstrated the need for careful studies of a non-linear data processing technique, such as the MEM. It is extremely dangerous to rely on the appearance of a single spectrum: with non-linear methods increases in signal-to-noise need not go hand-in-hand with increases in the reliability of line heights and line integrals, and decreases in linewidths do not necessarily result in genuine increases in resolution. When the performance of the technique cannot be analysed directly, Monte Carlo methods provide an extremely general and robust approach, providing information on both the reliabilities of various parameters and on systematic distortions in these parameters.

In general a spectrum can only be more accurately quantified if additional prior information is brought into the analysis. Knowledge of lineshapes, particularly if linewidths are also known, can give dramatic improvements. Any other information which can help distinguish between signal and noise should give further improvements. Such information must, of course, be reliable.

When lineshapes are known curve fitting is often the best method for quantifying a spectrum. If the model used is correct then parameters obtained by curve fitting are reliable; furthermore they are not systematically distorted. If, however, the model is incorrect serious distortions will occur. In particular it is necessary to know the number of spectral lines. If the data is severely truncated, so that the simple FT spectrum is corrupted by sinc-wiggles, this may be difficult to determine.

While the MEM is generally less suitable than curve fitting for quantitation, it does provide a spectrum which can be readily analysed by eye. If appropriate resolution enhancement is used, or if the data is severely truncated, this spectrum may be a more reliable depiction of the information contained in the data than the corresponding FT spectrum. If quantitation is desired this spectrum may be used to obtain initial estimates of spectral parameters which are then refined by curve fitting.

Chapter 4

Phase Modulated Rotating Frame Imaging

> . . . images serve for none other purpose but as to be books of unlearned men that cannot know letters . . .
>
> Thomas Crumwell, Vicar General of Canterbury, 1536

4.1 Introduction to *in vivo* NMR

Conventional NMR[4] is performed *in vitro*. A small quantity of the sample, which is often a pure compound, is dissolved in some convenient solvent to give a volume of around 1 ml and placed in a small glass tube. The sample is then placed in the centre of a homogeneous magnetic field, and surrounded by a coil capable of generating a homogeneous radiofrequency field. This spatial homogeneity, both of the fields and of the sample, greatly simplifies the description of the system.

When applied *in vivo* many or all of these simplifying assumptions break down. The sample is neither physically nor chemically homogeneous, and it is extremely difficult to generate homogeneous fields over the large sample volumes that may be involved. Several approaches[5, 6] have been developed to tackle these problems; in particular methods have been developed which use field inhomogeneity to obtain spectra from a small region of the sample, of homogeneous chemical composition.

In addition to these spectroscopic experiments *in vivo* NMR is used as an imaging technique, that is as a method for obtaining information about the physical structure of the sample[6]. This is achieved by determining the spatial distribution of some property which can be measured by NMR. The simplest possibility is to measure the number density of the nucleus of interest, but other properties are also used.

Phase Modulated Rotating Frame Imaging (PMRFI)[88, 89] is not, as its name suggests, a simple imaging experiment; instead it combines ideas from both imaging and spectroscopy to obtain spatially resolved spectra from the sample.

4.1.1 Spectroscopy

Conventional NMR spectroscopy is usually performed on pure compounds in order to obtain chemical information about the compound. By contrast *in vivo* NMR spectroscopy is generally used as an analytical technique, in order to obtain information on the concentrations of various compounds present in the sample. The compound dependent parameters, such as chemical shift and coupling patterns, are used as molecular fingerprints, allowing signals from different compounds to be distinguished.

Most biologically interesting compounds contain both hydrogen and carbon, and these might seem to be the obvious nuclei for spectroscopic study. In fact they present several problems, while phosphorus has proved extremely useful.

Hydrogen has three stable isotopes, of which ^{1}H is by far the most abundant (99.98%). ^{1}H nuclei have the dual advantages of sensitivity[1] and high natural abundance, but the vast majority of ^{1}H nuclei present are in the form of water, and so *in vivo* spectra are dominated by an enormous water peak; furthermore the ubiquity of hydrogen atoms in biological compounds means that the spectrum is crowded with signals arising from a vast number of different compounds. The consequence of these two problems is that while ^{1}H NMR is well suited to imaging experiments (described below) it is not generally useful for spectroscopy.

[1] Only tritium (^{3}H) nuclei are more sensitive.

Carbon might seem better than hydrogen since there is no troublesome water peak. Unfortunately the most common isotope of carbon (^{12}C, 98.89%) has nuclear spin $I = 0$, and so cannot be observed by NMR. The less common isotope (^{13}C, 1.11%) can be observed, but it is approximately forty times less sensitive than ^{1}H; combined with the low abundance this means that carbon spectra are around four thousand times less intense than corresponding hydrogen spectra. It is possible greatly to increase the sensitivity by isotopic enrichment, a technique which has proved useful in tracer work, but this is not a general solution.

Phosphorus has only one stable isotope, ^{31}P, which is about ten times less sensitive than ^{1}H. While biological systems contain phosphorus in a wide variety of environments, most of these give rise to extremely broad signals[2]. Superimposed on these are signals arising from a small number of phosphorus containing compounds such as adenosine tri-phosphate (ATP) which are extremely important metabolic intermediates. The majority of ^{31}P studies are performed in order to measure the relative concentrations of these compounds.

In addition to the three nuclei described above many other nuclei have been studied by *in vivo* NMR, but these three are dominant. In particular ^{1}H is the dominant nucleus for imaging studies, while ^{31}P is the dominant nucleus for spectroscopic studies.

4.1.2 Localised Spectroscopy

Since biological subjects are not chemically uniform, it is frequently useful to obtain NMR spectra from small regions of the sample which *are* chemically uniform. This can be done in four ways:

1. by using a small coil, if necessary surgically implanted, to surround the region of interest;

[2]Most of the phosphorus in the body is found in bones, in phospholipid membranes, and in DNA. Nuclei in these environments are almost stationary on the NMR timescale, and so the internuclear dipolar couplings (described in chapter 1) remain significant. Since each nucleus is coupled to a very large number of other nuclei, and the distribution of molecular orientations is random, these couplings result in broadening rather than resolved splittings. Small molecules in solution undergo rapid tumbling, averaging out the dipolar coupling and removing the broadening.

2. by using the inhomogeneity of the B_0 field to select a region of interest;

3. by using the inhomogeneity of the B_1 field to select a region of interest;

4. by combing the spectroscopic study with an imaging experiment.

Method (i), while of historical interest, is invasive and is obviously not generally applicable. Further discussion of method (iv) must be deferred until after the discussion of imaging experiments, but methods (ii) and (iii) can be described in more detail now.

Using an inhomogeneous B_0 field

This technique, known as Topical Magnetic Resonance (TMR), uses an inhomogeneous B_0 field to smear out signals from all parts of the sample except the region of interest[90]. Since the resonance frequency of a nucleus is proportional to the applied magnetic field, a group of otherwise identical nuclei experiencing different applied fields will have different resonance frequencies, and so give rise to a broad line. TMR uses a B_0 field which is extremely inhomogeneous except in a small region; signals from this region will give sharp lines superimposed on extremely broad signals from the rest of the sample. The broad lines can be suppressed by a number of data processing techniques, leaving only signals from the region of interest. This region can be changed by moving the region of field homogeneity or, more commonly, by moving the sample. The method is rarely used nowadays.

Using an inhomogeneous B_1 field

Rather than smearing out extraneous signals, this method works by only exciting nuclei in the region of interest[91]. In *in vitro* NMR, solenoidal or saddle coils are used to generate a homogeneous B_1 field throughout the sample, but other coil designs, such a surface coils, can be used to generate extremely inhomogeneous B_1 fields. This field inhomogeneity acts to select a region of the sample in two separate ways.

Firstly only nuclei in the region of interest are excited. The intensity of an NMR signal is proportional to the transverse magnetization produced by the initial excitation. This magnetization is proportional to $\sin(\gamma B_1 \tau)$, where τ is the length of the excitation pulse. Where B_1 is small the transverse magnetization is small, and so the NMR signal is weak.

Secondly only nuclei in the region of interest are detected efficiently. In the simplest case the NMR signal is detected using the same coil as was used for the excitation pulse; in more complex cases another coil is used, but this second coil is generally similar to the first one. Since detection involves the conversion of an oscillating magnetic field to an oscillating current (the reverse of the first stage of the excitation process), the detection efficiency is proportional to the B_1 field strength.

Combining these two effects the intensity of the detected signal is proportional to $B_1 \sin(\gamma B_1 \tau)$. Consequently any coil capable of generating an inhomogeneous B_1 field will provide a technique for obtaining localised spectra. Obviously this will only be useful if the sensitive region is conveniently shaped and positioned; fortunately surface coils possess a very convenient sensitive region, positioned close to the coil.

4.1.3 Imaging

As described above, NMR imaging refers to the determination of the spatial distribution of any property of the sample measurable by NMR. The following discussion will be limited to the simplest case, that is the determination of number densities[3].

All imaging techniques rely on making the Larmor frequency of the nucleus of interest depend on its position in space, usually by the application of field gradients. The underlying variations in Larmor frequency, due to chemical shifts, are a problem, but this can be side-stepped by making the spatial variation very large compared with the chemical shift variation. In the following the problem of

[3]While simple to measure, number density is not very useful since (for ^{1}H) it shows very little variation in most biological specimens. In general the number densities are weighted, for example by the local relaxation rate, in order to increase the image contrast.

chemical shifts will be largely ignored.

Field Gradients

Consider a sample containing nuclei whose number density $N(x, y, z)$ varies in space. In a spectroscopic experiment the sample is placed in a homogeneous magnetic field, excited to produce transverse magnetization, and the precession of this magnetization recorded. The resulting signal is

$$s(t) = Ce^{i\omega_0 t}e^{-t/T_2}, \tag{4.1}$$

where

$$C \propto \iiint N(x, y, z)\, dx\, dy\, dz, \tag{4.2}$$

$\omega_0 = -\gamma B_0$, and exponential relaxation has been assumed. The corresponding spectrum, obtained by Fourier transformation, is

$$S(\omega) = A(\omega_0) + iD(\omega_0) \tag{4.3}$$

where $A(\omega_0)$ and $D(\omega_0)$ are absorption and dispersion Lorentzian lineshapes centred at ω_0 (see chapter 2). Note that all information about $N(x, y, z)$, other than its total integral, has been lost. Since the applied fields are homogeneous there is no distinction between nuclei in different spatial positions.

In order to obtain spatial discrimination one of the fields must be made inhomogeneous. The simplest possibility is to keep both B_1 and B_0 homogeneous during the excitation period, and then cause B_0 to vary linearly along some axis, for example the x-axis. In this case the precession frequency of a given nucleus depends on its x-coordinate:

$$\omega_x = -\gamma B(x) = -\gamma(B_0 + Gx) \tag{4.4}$$

where G is the gradient strength. The signal arising from any given x-plane of thickness dx is

$$s_x(t) = C(x)e^{i\omega_x t}e^{-t/T_2}dx, \tag{4.5}$$

where

$$C(x) \propto \iint N(x, y, z)\, dy\, dz \tag{4.6}$$

is the projection of the number density onto the x-axis. Fourier transformation gives

$$S_x(\omega) = A(\omega_x) + iD(\omega_x), \tag{4.7}$$

so the corresponding spectrum is a Lorentzian line of intensity $C(x)\,dx$ centred at ω_x. The total signal is a sum of terms like equation 4.5 and so the complete spectrum is a sum of terms like equation 4.7. This corresponds to $C(x)$ convolved with an appropriate Lorentzian lineshape. Similarly if the sample contains nuclei with different chemical shifts this spectral structure will also be convolved with $C(x)$. Clearly it is desirable to minimise these effects: this can be achieved by making G as large as possible.

Multi-dimensional Imaging

The experiment described above can only give information about the spatial structure of the sample along one axis; structure along the other axes is lost as a result of the integration in equation 4.6. Information about this structure can be obtained using a "multi-dimensional" experiment, in which data is measured as a function of two or more different time variables[4].

For example it is possible to obtain information about structure along the x and y-axes as follows. The sample is placed in a homogeneous field and excited to produce transverse magnetization. This is allowed to precess in a field whose strength varies linearly along x for a time period t_1; the x-gradient is then replaced by a similar gradient along the y-axis. The precessing magnetization is recorded as a function of t_2, the time after the change in field gradients. The signal arising from any xy-column of cross-section $dx\,dy$ is

$$s_{xy}(t_1, t_2) = C(x, y)e^{i\omega_x t_1}e^{-t_1/T_2}e^{i\omega_y t_2}e^{-t_2/T_2}dx\,dy, \tag{4.8}$$

where

$$C(x, y) \propto \int N(x, y, z)\,dz \tag{4.9}$$

[4]There is another method of obtaining multi-dimensional images, called Projection Reconstruction Imaging[6], which works by measuring projections of the number density onto various axes.

is the projection of the number density onto the xy-plane. The experiment is repeated for many different values of t_1, giving a two-dimensional dataset.

Fourier transformation of equation 4.8 with respect to t_2 gives

$$S_{xy}(t_1, \omega_2) = C(x, y)e^{i\omega_x t_1}e^{-t_1/T_2}[A_2(\omega_y) + iD_2(\omega_y)], \tag{4.10}$$

where A_2 and D_2 and Lorentzian lineshapes in ω_2. Fourier transformation with respect to t_1 then gives

$$\begin{aligned} S_{xy}(\omega_1, \omega_2) &= [A_1(\omega_x) + iD_1(\omega_x)] \times [A_2(\omega_y) + iD_2(\omega_y)] \\ &= [A_1(\omega_x)A_2(\omega_y) - D_1(\omega_x)D_2(\omega_y)] \\ &\quad +i[A_1(\omega_x)D_2(\omega_y) + D_1(\omega_x)A_2(\omega_y)] \end{aligned} \tag{4.11}$$

corresponding to a line centred at (ω_x, ω_y). Once more the overall two-dimensional spectrum corresponds to $C(x, y)$ convolved with an appropriate two-dimensional Lorentzian lineshape. Unfortunately the real part of the spectrum contains not simply the convenient absorption-absorption lineshape, but the difference between this and the dispersion-dispersion lineshape (these various lineshapes are depicted in figure 4.1). This combination, known as the phase-twist lineshape, is much broader than the absorption-absorption lineshape, and also contains regions of negative intensity. The negative regions can be removed by displaying the absolute value of the complex spectrum, but this results in further broadening.

If this broadening is a serious problem, it may be desirable to obtain pure absorption-absorption lineshapes. This can be achieved in several ways, all of which require each dataset to be acquired in two different ways. For example it is possible to acquire one dataset as described above, and one with the x-gradient reversed. The data from a single xy-column is now

$$\begin{aligned} s^1_{xy}(t_1, t_2) &= C(x, y)e^{i\omega_x t_1}e^{-t_1/T_2}e^{i\omega_y t_2}e^{-t_2/T_2}dx\,dy, \\ s^2_{xy}(t_1, t_2) &= C(x, y)e^{-i\omega_x t_1}e^{-t_1/T_2}e^{i\omega_y t_2}e^{-t_2/T_2}dx\,dy. \end{aligned} \tag{4.12}$$

The real parts of the corresponding spectra are

$$\begin{aligned} S^1_{xy}(\omega_1, \omega_2) &= [A_1(\omega_x)A_2(\omega_y) - D_1(\omega_x)D_2(\omega_y)] \\ S^2_{xy}(\omega_1, \omega_2) &= [A_1(-\omega_x)A_2(\omega_y) - D_1(-\omega_x)D_2(\omega_y)]. \end{aligned} \tag{4.13}$$

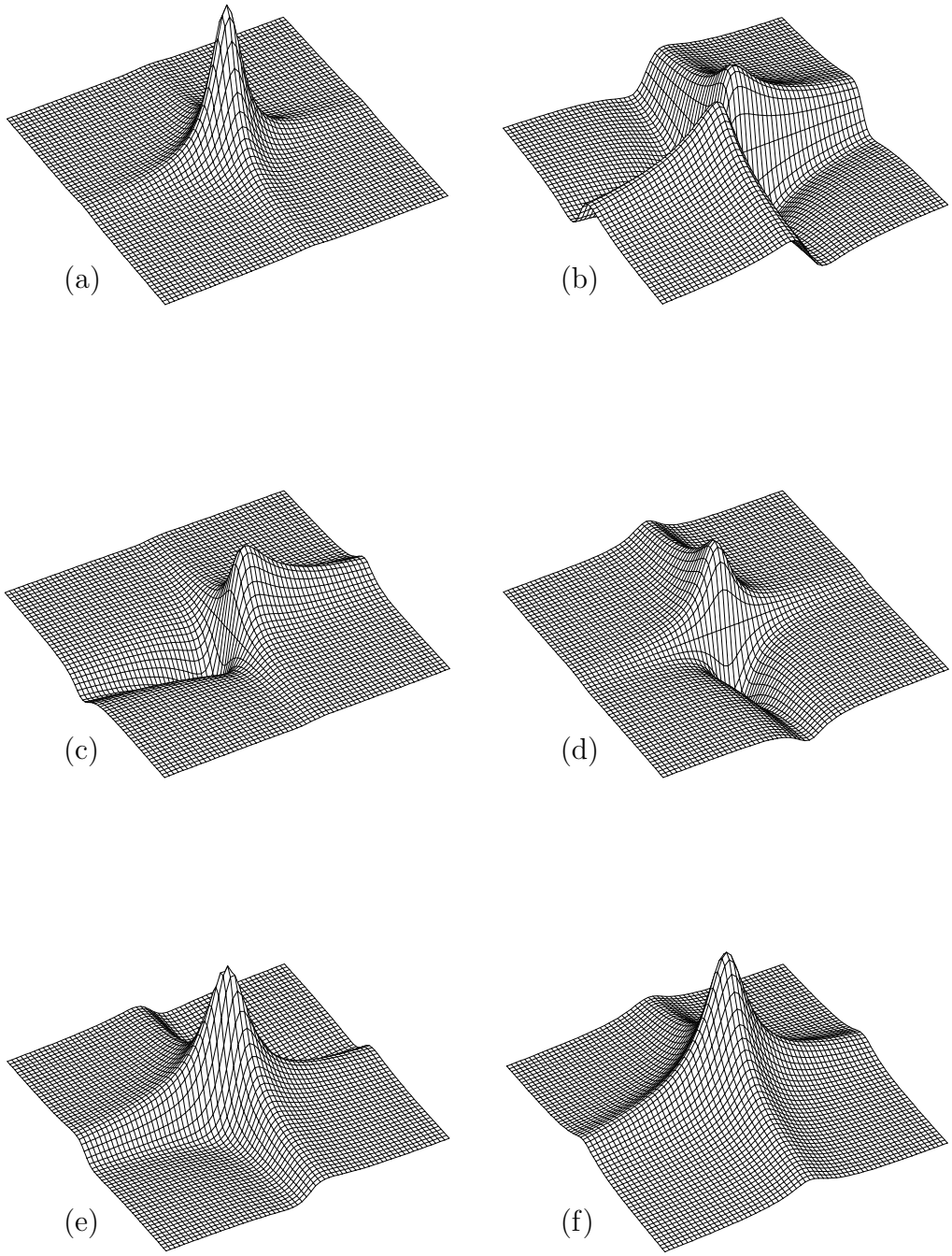

Figure 4.1: Lorentzian lineshapes in two-dimensional spectra: (a) absorption-absorption A_1A_2; (b) dispersion-dispersion D_1D_2; (c) absorption-dispersion A_1D_2; (d) dispersion-absorption D_1A_2; (e) phase-twist lineshape $A_1A_2 - D_1D_2$; (f) absolute value lineshape $[(A_1A_2 - D_1D_2)^2 + (A_1D_2 + D_1A_2)^2]^{1/2}$.

It is possible to reverse all the ω_1-frequencies in spectrum S^2, so that the lines occur in the same positions as in S^1. The symmetric absorption lineshape is unaffected by this process, but the antisymmetric dispersion lineshape is inverted, so that

$$S^{2'}_{xy}(\omega_1, \omega_2) = [A_1(\omega_x)A_2(\omega_y) + D_1(\omega_x)D_2(\omega_y)]. \tag{4.14}$$

Adding this to the first spectrum gives pure absorption lineshapes

$$S^1_{xy} + S^{2'}_{xy} = 2A_1(\omega_x)A_2(\omega_y). \tag{4.15}$$

This technique is used in the PMRFI experiment, where absorption lineshapes are highly desirable.

The experiment described above can give information about the spatial structure along two axes, but in general information about all three axes is required. This can be achieved by adding a third time period, t_3, during which the magnetization precesses in a field with a z-gradient; the signal is recorded as a function of t_3 for many different values of t_1 and t_2. Because of the large number of experiments required complete three-dimensional images can take a very long time to acquire. In many cases it is sufficient to obtain a two-dimensional image of one or more planes, confining the initial excitation to the plane of interest. This can be accomplished by performing the excitation in a z-gradient, so that nuclei outside the desired plane are far from resonance, and are not significantly excited[5].

Rotating Frame Imaging

The imaging techniques described above rely on gradients in the static field B_0 to achieve spatial discrimination. Rotating frame imaging[6, 88, 89, 92, 93] uses gradients in the radiofrequency field B_1 to achieve the same end. Viewed in an appropriate rotating frame B_1 is static, and during the excitation period the sample magnetization precesses around it at a frequency $\omega_1 = -\gamma B_1$. If the B_1 field strength varies along some axis, then so does ω_1, and this variation can be detected.

For example two-dimensional rotating frame imaging can be achieved as follows. Consider a sample in a homogeneous static field B_0 which is excited by a B_1 field

[5]In practice "shaped" excitation pulses must be used to ensure reasonable selectivity.

whose intensity varies linearly along the x-axis, so that $B_1(x) = Gx$. The sample magnetization precesses around B_1 at a frequency $\omega_x = -\gamma Gx$, and the resulting transverse magnetization after a time t_1 depends on $\sin \omega_x t_1$. After excitation a y-gradient is applied to the B_0 field, and the transverse magnetization precesses around B_0 at a frequency $\omega_y = -\gamma B_0(y)$. This precession is recorded as a function of t_2, the time after excitation. The resulting signal is

$$\begin{aligned} s_{xy}(t_1, t_2) &= C(x, y) \sin(\omega_x t_1) e^{-t_1/T_2} e^{i\omega_y t_2} e^{-t_2/T_2} dx\, dy \\ &= \frac{C(x, y)}{2i} [e^{i\omega_x t_1} - e^{-i\omega_x t_1}] e^{-t_1/T_2} e^{i\omega_y t_2} e^{-t_2/T_2} dx\, dy, \end{aligned} \tag{4.16}$$

where exponential relaxation has been assumed during t_1 and t_2. Fourier transformation gives a spectrum containing two phase-twist lineshapes, centred at (ω_x, ω_y) and $(-\omega_x, \omega_y)$.

Many variations of this experiment exist; for example it is possible to remove the mirror image peaks, and to remedy the phase-twist lineshape.

4.1.4 Spectroscopic Imaging

Spectroscopic imaging experiments combine the spatial discrimination of an imaging experiment with the chemical shift discrimination of a spectroscopic experiment. The imaging experiments described above can all be converted into spectroscopic imaging experiments by the addition of another time dimension, during which the sample magnetization precesses freely in a homogeneous field. For example, if an x-gradient is applied to the static field during the first stage of a two-dimensional experiment, and no gradient is applied during the second stage, the signal resulting from a single plane has the form

$$s_x(t_1, t_2) = C(x) e^{i\omega_x t_1} e^{-t_1/T_2} e^{i\omega t_2} e^{-t_2/T_2} dx, \tag{4.17}$$

where ω is the precession frequency in the homogeneous field. The corresponding spectrum has a line centred at (ω_x, ω), where ω_x depends on spatial position and ω depends on chemical shifts.

4.2 Surface coils

A surface coil is simply a circular loop of wire placed on or near the surface of the sample. The field from such a coil is inhomogeneous, and may be readily calculated on account of the simplicity of the system. This inhomogeneous field may be used for localised spectroscopy, or as part of a rotating frame imaging experiment.

4.2.1 Field Distribution

Consider a circular coil, of radius a, aligned along the x-axis as shown in figure 4.2[91]. It is convenient to work in cylindrical coordinates (x, ρ, ϕ), as defined in the figure. The magnetic field at any point due to a current I flowing in the loop has axial and radial components B_x and B_ρ given by[94]

$$B_x = \frac{\mu I}{2\pi} \frac{1}{[(a+\rho)^2 + x^2]^{1/2}} \left[K + \frac{a^2 - \rho^2 - x^2}{(a-\rho)^2 + x^2} E \right] \tag{4.18}$$

$$B_\rho = \frac{\mu I}{2\pi} \frac{x}{\rho[(a+\rho)^2 + x^2]^{1/2}} \left[-K + \frac{a^2 + \rho^2 + x^2}{(a-\rho)^2 + x^2} E \right] \tag{4.19}$$

where K and E are complete elliptic integrals of the first and second kind[65, 95], defined by

$$K = \int_0^{\pi/2} \frac{1}{\sqrt{1 - k^2 \sin^2\theta}} \, d\theta \tag{4.20}$$

$$E = \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2\theta} \, d\theta \tag{4.21}$$

with

$$k^2 = \frac{4a\rho}{(a+\rho)^2 + x^2}. \tag{4.22}$$

Note that the field is independent of ϕ; this is a consequence of the cylindrical symmetry of the system.

In any NMR experiment it is of course necessary to consider the static magnetic field B_0, as well as the oscillating field B_1. Clearly one effect of the B_0 field is to remove the cylindrical symmetry (unless it is directed parallel to the axis of the circular coil). In the following sections it will be assumed that B_0 is directed along the z-axis, perpendicular to the axis of the coil.

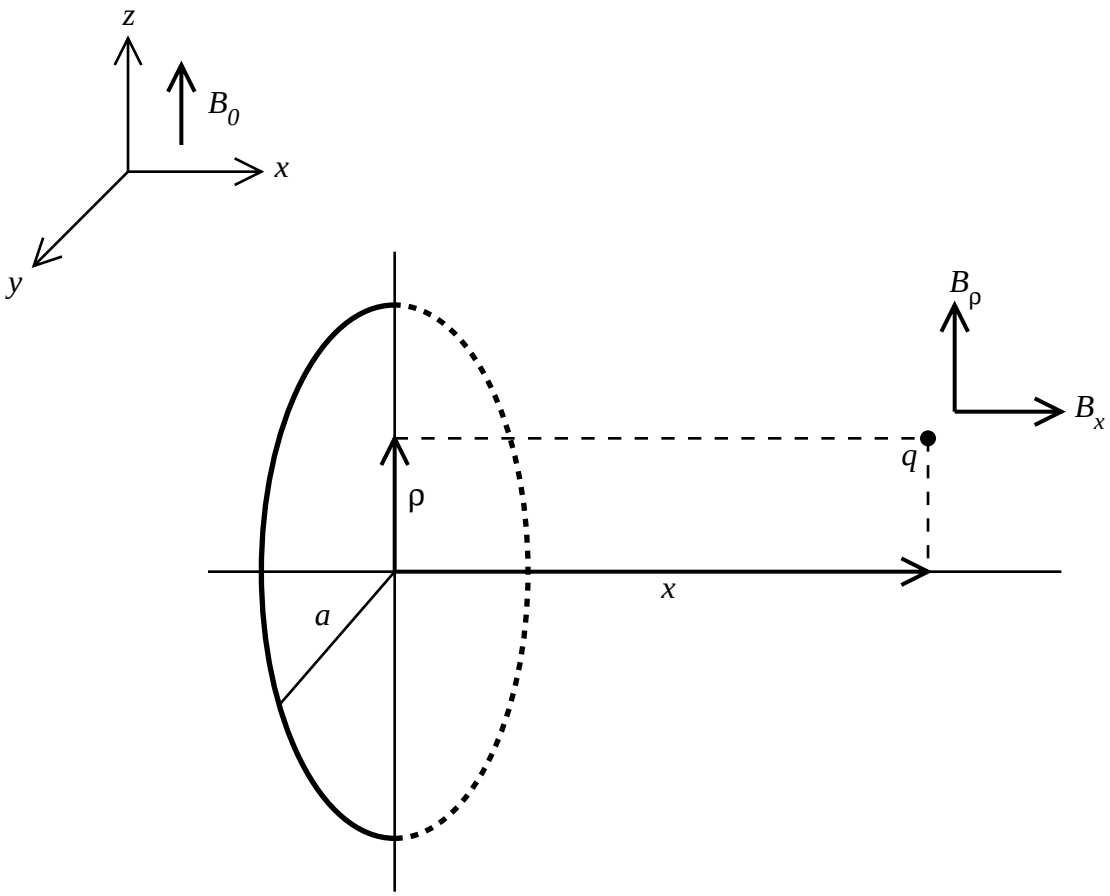

Figure 4.2: Coordinate system used with a surface coil. The laboratory axis system is indicated in the upper left hand corner. The coil, of radius a, has its axis directed along the laboratory x-axis. The point q is described using cylindrical coordinates (x, ρ, ϕ) as follows: x is the distance of q from the coil centre measured along the coil axis; ρ is the perpendicular distance of q from the coil axis; ϕ is the angle between the radial vector $\boldsymbol{\rho}$ and the laboratory z-axis measured in the laboratory yz-plane. The B_0 field is directed along the laboratory z-axis; the B_1 field has an axial component B_x and a radial component B_ρ directed parallel to the vectors $\boldsymbol{x}$ and $\boldsymbol{\rho}$ respectively.

4.2.2 Localised Spectroscopy

As described above, the inhomogeneous field due to a surface coil provides a simple method of obtaining localised spectra. Since nuclei are excited by the B_1 field, signals will only arise from regions of the sample where this field is significant, that is regions close to the coil; what "close" means in this context can be calculated from equations 4.18 and 4.19. Similarly the detection efficiency will vary with the B_1 field strength.

Since it is magnetization transverse to the applied magnetic field which is excited and detected, only the transverse component of the B_1 field is important. The coil axis is perpendicular to the applied field, and so the axial component, B_x, is always entirely transverse. The radial component, B_ρ, is more complicated; it has a transverse component of $B_\rho \sin\phi$ where ϕ is the angle between the directions of B_ρ and B_0. For simplicity the following discussion will be limited to points in the xz-plane, for which the radial contribution to the transverse field is zero.

First consider the field strength along the coil axis ($\rho = 0$). In this case $K = E = \pi/2$, and the equations simplify to

$$\begin{aligned} B_x &= \frac{\mu a^2 I}{2(a^2+x^2)^{3/2}} \\ &= \frac{1}{[1+(x/a)^2]^{3/2}} B_x^0 \end{aligned} \tag{4.23}$$

where B_x^0 is the axial field at the centre of the coil. The field strength falls off with distance from the coil, tending towards a inverse third power dependence at large distances.

Away from the axis the situation is more complex. At small axial distances the field strength initially increases with radial distance, but when $x \geq \frac{1}{2}a$ the field falls off monotonically with both axial and radial distance. The variation of B_x with both x and ρ is plotted in figure 4.3; the field is largely confined to regions near the coil.

The overall sensitivity depends on the length of the excitation pulse, τ, as well as on the field distribution. One possibility is to choose τ such that $\gamma B_x^0 \tau = \pi/2$; the result of this choice is plotted in figure 4.4. The sensitive region is approximately

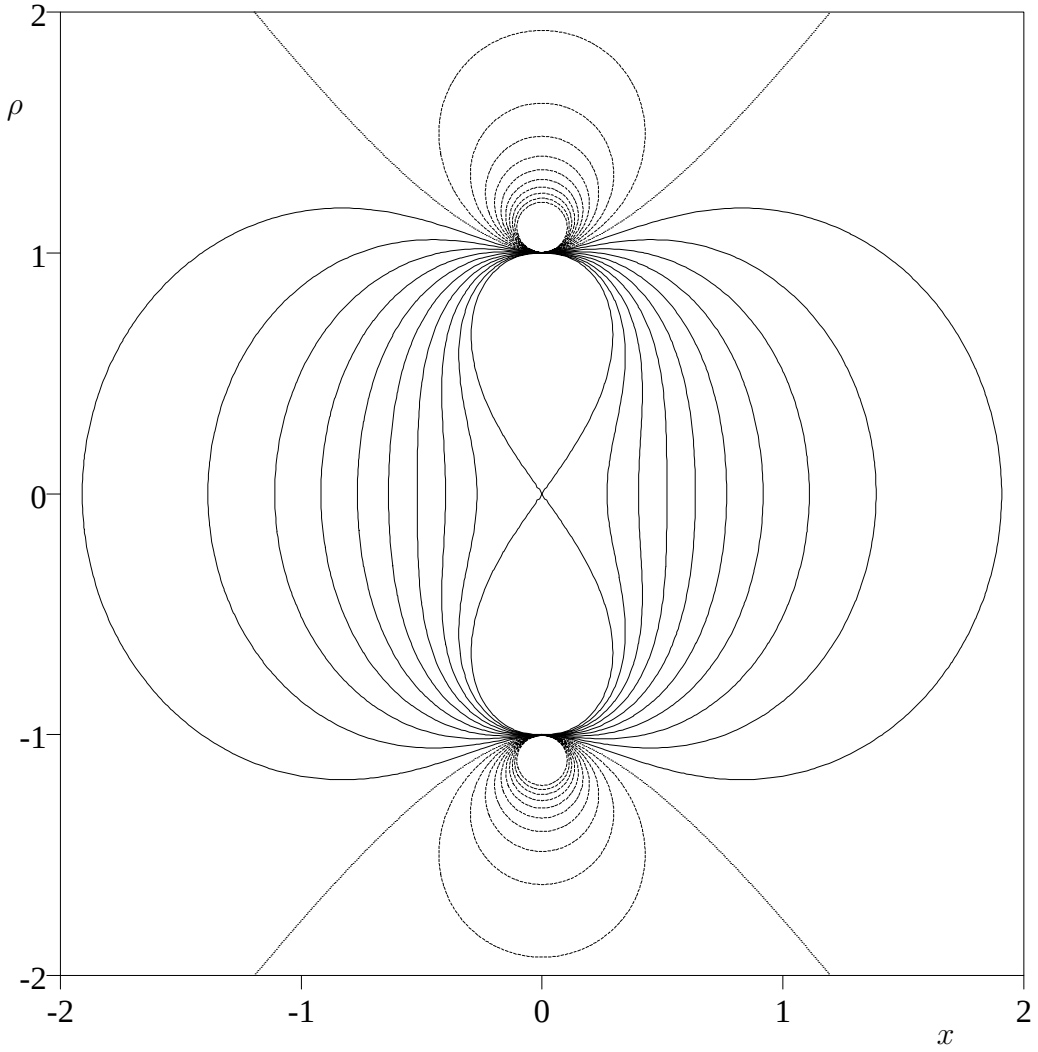

Figure 4.3: Plot of the axial field from a surface coil. Axial distance (x) and radial distance (ρ) are measured in multiples of the coil radius so that the coil wires are perpendicular to the page at ($x = 0, \rho = \pm 1$). Contours are drawn at 0, ± 10, ..., $\pm 100\%$ of the field strength at the centre of the coil; negative contours are dotted. Note that the axial field is infinite at the wires themselves.

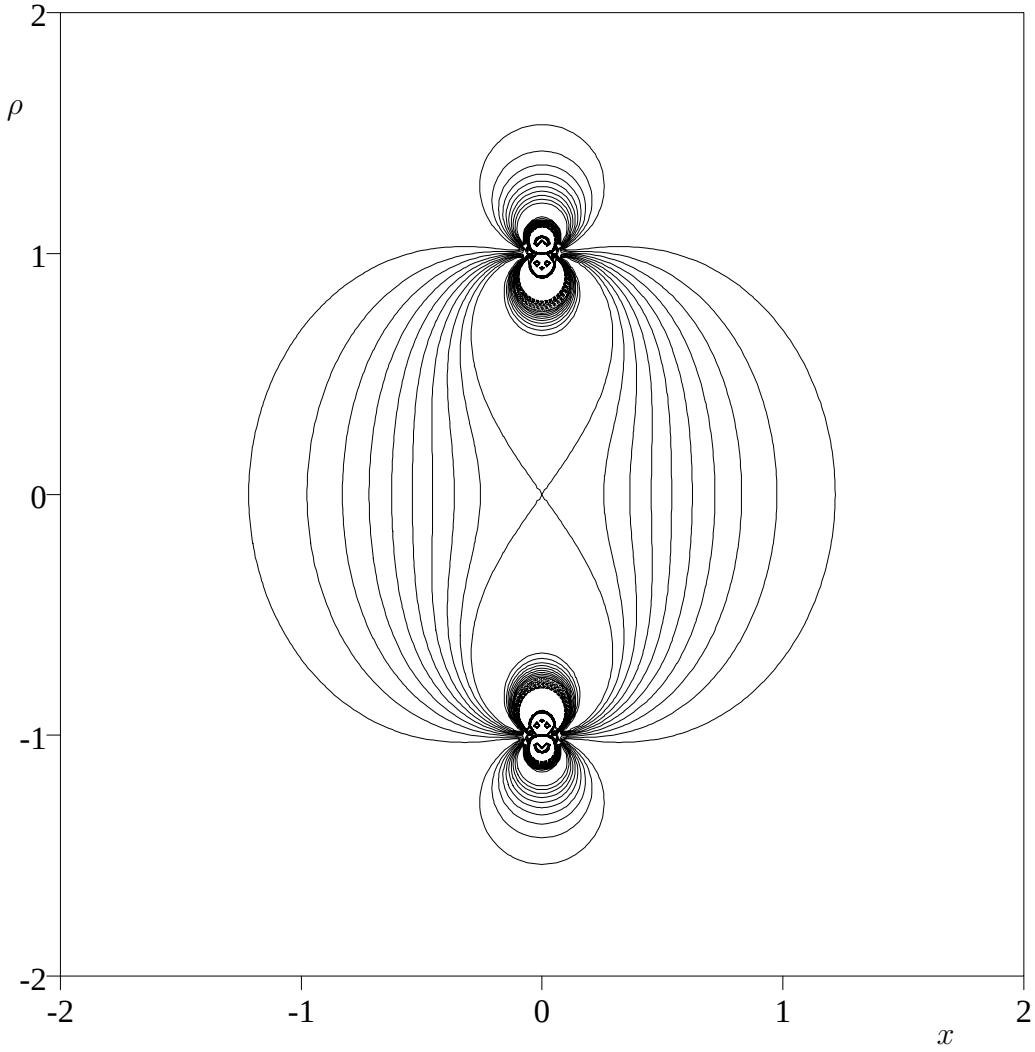

Figure 4.4: Plot of the overall sensitivity of a surface coil with $\gamma B_x^0 \tau = \pi/2$. Axial distance ($x$) and radial distance ($\rho$) are measured in multiples of the coil radius so that the coil wires are perpendicular to the page at ($x = 0, \rho = \pm 1$). Contours are drawn at 0, ± 10, ..., $\pm 100\%$ of the sensitivity at the centre of the coil; negative contours are dotted. The axial field is very large close to the wires, and the sensitivity varies very rapidly in this region.

cylindrical, with radius equal to the coil radius, and extending approximately one radius to each side of the coil.

Away from the xz-plane the situation is even more complex, since the transverse component of the radial field is important. This will not be considered in detail; the sensitive region is broadly similar but is larger and not so conveniently shaped.

4.3 PMRFI

PMRFI is a rotating frame spectroscopic imaging experiment, which provides a convenient method for obtaining spectra from biological samples using a surface coil. At first sight a surface coil seems poorly suited to an imaging experiment; ideally imaging fields should vary linearly along one axis and be constant in other directions; by contrast the field from a surface coil varies in a complex, non-linear fashion. The field does, however, tend to decrease with distance from the coil, and so can provide a crude method of discrimination between nuclei at different distances.

Crude as it is, this discrimination is sufficient to be useful in some cases. For example it may be desirable to obtain a spectrum from a human liver. A surface coil placed on the abdomen will detect signals from the liver, but will also detect signals from the muscle layers above and below. These signals can be well separated using the distance discrimination of PMRFI.

4.3.1 Simple Theory

The PMRFI pulse sequence[89]

$$\theta_{\pm x}(t_1) - \lambda_y - acquire(t_2) \tag{4.24}$$

consists of a variable length pulse θ of duration t_1, a phase-encoding pulse λ and a period of free precession t_2 during which the NMR signal is detected. The idealised operation of the experiment is easily visualised in the rotating frame. Magnetization, initially directed along the z-axis, precesses during t_1 in the yz-plane at frequency $\omega_1 = \mp\gamma B_1$, where B_1 is the local radiofrequency field strength,

which varies with distance from the surface coil. The sense of precession depends on whether the θ pulse is applied along the $+x$ or $-x$ axis. The pulse length, t_1, is varied linearly with an increment Δt_1, chosen so that all the spins undergo rotations in the range 0°–180° (that $0 < \gamma B_1 \Delta t_1 < \pi$). The second pulse, ideally a 90° rotation for all spins, flips the magnetization from the yz-plane into the xy-plane, where it precesses at the offset frequency $\omega_2{=}\gamma\Delta B_0$ during the acquisition period t_2. Hence the ideal signal takes the form

$$s_{\pm}(t_1, t_2) = Ce^{\mp i\gamma B_1 t_1} e^{i\gamma \Delta B_0 t_2}, \tag{4.25}$$

where C is a real intensity factor, relaxation has been neglected, and the $\mp$ sign refers to the sense of precession during t_1. This data has the same form as that in equation 4.12: Fourier transformation of each dataset gives a spectrum containing phase-twist lineshapes, but the dispersion components can be cancelled to give simple absorption-absorption lines at $(\gamma B_1, \gamma \Delta B_0)$. Hence information about distance from the coil is displayed along ω_1, while information about chemical shifts is displayed along ω_2.

Unfortunately this description is highly idealised; practical experiments differ in several important ways.

Firstly the phase-encoding pulse λ is usually applied using the same coil as the variable length pulse θ, and so varies across the sample. In practice the pulse length is chosen to be Δt_1, so that the rotation angle is 90° at the centre of the region of interest, less at greater depth, and rising to 180° close to the coil.

Secondly the B_1 field power is limited by financial and safety considerations, which can result in severe off-resonance effects. Even if the power is adequate close to the coil, it will always be low at large distances from the coil.

Thirdly the limited patience of subjects means that the number of t_1 increments that may be acquired is quite restricted, so that the dataset is severely truncated in t_1.

The effect of truncation is obvious: the resulting spectrum will suffer from sinc wiggles in ω_1 or, if the dataset is apodised, severe broadening in ω_1. The effects of the other two defects require more careful consideration.

Phase-encoding Errors

The purpose of the phase-encoding pulse, λ, is to convert amplitude modulation to phase modulation. After the θ pulse the magnetization vector lies in the yz-plane, and has a transverse component of $\sin \omega_1 t_1$; Fourier transformation of this with respect to t_1 results in two peaks of height $\frac{1}{2}$ at $\pm\omega_1$. This is called an amplitude modulated dataset. The effect of a perfect phase-encoding pulse is to flip the magnetization into the xy-plane, so that the transverse component is $e^{i\omega_1 t_1}$; Fourier transformation of this gives a single peak of height 1 at ω_1. This is called a phase modulated dataset.

An imperfect phase-encoding pulse causes only partial transfer of magnetization from the yz-plane to the xy-plane, resulting in a mixture of phase and amplitude modulation. When λ is 0° or 180° the dataset is purely amplitude modulated, when λ is 90° the dataset is purely phase modulated, and at other values of λ the modulation is mixed. Consequently there is a partial transfer of intensity from the main peak at ω_1 to a shadow peak at $-\omega_1$. This shadow peak is not itself important since all genuine peaks have positive frequencies in ω_1 and only the positive half of the spectrum is displayed, but the loss of intensity from the main peak is a problem (the intensity of the main peak varies as $\frac{1}{2}[1 + \sin \lambda]$).

Off-resonance Effects

The descriptions of NMR given above and in chapter 2 have assumed that all B_1 pulses are approximately on-resonance, so that their effect can be approximated by rotation at a frequency ω_1 around the pulse axis (in the rotating frame). If off-resonance effects are significant, that is ω_1 is not much greater than $\Delta\omega$, the situation is more complicated. The effective $\boldsymbol{B}_1$ field is given by the vector sum of the nominal $\boldsymbol{B}_1$ field and the residual $\boldsymbol{B}_0$ field[3]; this field has strength

$$B_1^e = \sqrt{B_1^2 + \Delta B_0^2} = B_1 \sec \alpha \tag{4.26}$$

and is tilted out of the xy-plane by an angle

$$\alpha = \tan^{-1}(\Delta B_0 / B_1). \tag{4.27}$$

The magnetization precesses around this effective field at frequency $\gamma B_1 \sec\alpha$.

Off-resonance effects, which occur during both the θ and λ pulses, have four consequences for the Fourier transform spectrum: mixed phase (absorption and dispersion) lineshapes in both frequency dimensions, "axial peaks" at $\omega_1 = 0$, transfer of intensity from main peaks to shadow peaks (at negative ω_1), and a shift of lines in the ω_1 dimension from γB_1 to $\gamma B_1 \sec\alpha$. This final effect means that the position of a peak in ω_1 reflects not simply the position of the nucleus in space, but also its chemical shift.

4.3.2 Detailed Theory

In order to develop the theory of PMRFI in more detail, it is necessary to use an accurate description of the effects of pulses upon the spin system, including off-resonance effects. As described above, the effect of any pulse on the magnetization remains a rotation, but the rotation axis and angle will differ from their nominal values.

Any NMR pulse can be described by the general rotation matrix[52, 96]

$$P(\beta_\phi, \alpha) = \begin{bmatrix} \sin^2\phi\cos\beta' & -\sin\alpha\sin\beta' & \sin\phi\cos\alpha\sin\beta' \\ +\cos^2\phi(\cos^2\alpha & +\cos^2\alpha\sin\phi\cos\phi & +\cos\phi\sin\alpha \\ +\sin^2\alpha\cos\beta') & \times(1-\cos\beta') & \times\cos\alpha(1-\cos\beta') \\ \\ \sin\alpha\sin\beta' & \cos^2\phi\cos\beta' & -\cos\phi\cos\alpha\sin\beta' \\ +\cos^2\alpha\sin\phi\cos\phi & +\sin^2\phi(\cos^2\alpha & +\sin\phi\sin\alpha \\ \times(1-\cos\beta') & +\sin^2\alpha\cos\beta') & \times\cos\alpha(1-\cos\beta') \\ \\ -\sin\phi\cos\alpha\sin\beta' & \cos\phi\cos\alpha\sin\beta' & \\ +\cos\phi\sin\alpha & +\sin\phi\sin\alpha & \sin^2\alpha+\cos^2\alpha\cos\beta' \\ \times\cos\alpha(1-\cos\beta') & \times\cos\alpha(1-\cos\beta') & \end{bmatrix} \quad (4.28)$$

which depends on three angles: β, the nominal rotation angle, ϕ, the phase of the pulse (that is the axis in the xy-plane along which the pulse is nominally applied), and α, the off-resonance angle, which is defined in equation 4.27; β' is a convenient notation for $\beta\sec\alpha$.

The PMRFI experiment begins with initial magnetization equal to the equilibrium magnetization,

$$\boldsymbol{I}^i = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} I_0, \tag{4.29}$$

and operates on this with two pulses, θ and λ, to give a final magnetization

$$\boldsymbol{I}^f_{\pm} = P(\lambda_y, \alpha)P(\theta_{\pm x}, \alpha)\boldsymbol{I}^i. \tag{4.30}$$

This magnetization precesses in the xy-plane during t_2, so that the final signal is of the form

$$s_{\pm}(t_1, t_2) = F_{\pm}(t_1)e^{i\omega_2 t_2} I_0, \tag{4.31}$$

with

$$\begin{aligned}
\Re[F_{\pm}] &= \pm\cos\lambda'\sin\alpha\cos\alpha(1-\cos\theta') \pm \sin\alpha\cos\alpha\sin\lambda'\sin\theta' \\
&\quad + \cos\alpha\sin\lambda'\sin^2\alpha + \cos^3\alpha\cos\theta'\sin\lambda', \\
\Im[F_{\pm}] &= \mp\cos^3\alpha\sin\theta' \mp \sin^2\alpha\cos\alpha\sin\lambda'(\cos\theta'-1) \mp \sin^2\alpha\cos\lambda'\cos\alpha\sin\theta' \\
&\quad + \sin^3\alpha\cos\alpha + \sin\alpha\cos^3\alpha\cos\theta' - \sin^3\alpha\cos\alpha\cos\lambda' \\
&\quad - \sin\alpha\cos^3\alpha\cos\lambda'\cos\theta',
\end{aligned} \tag{4.32}$$

where the real and imaginary parts of $F_{\pm}$ are taken from the x and y components of $\boldsymbol{I}^f_{\pm}$. Note that for an ideal experiment ($\alpha = 0$, $\lambda = \pi/2$) this reduces to

$$F_{\pm} = \cos\theta \mp i\sin\theta = e^{\mp i\theta} = e^{\mp i\omega_1 t_1} \tag{4.33}$$

as expected.

4.3.3 Artefacts

The additional terms in equation 4.32 result in artefacts when the data is processed by the Fourier transform. Details of the calculation of these artefacts may be found in appendix D; the overall result is

$$S(\omega_1, \omega_2) = A_1^+ A_2 \quad [+\cos^3\alpha(\sin\lambda'+1) + \sin^2\alpha\cos\lambda'\cos\alpha]$$

$$
\begin{aligned}
&+ \quad D_1^+ D_2 \quad [-\cos\lambda' \sin\alpha \cos\alpha] \\
&+ \quad A_1^+ D_2 \quad [+\sin\alpha \cos^3\alpha(\cos\lambda' - 1) - \sin\alpha \cos\alpha \sin\lambda'] \\
&+ \quad D_1^+ A_2 \quad [-\sin^2\alpha \cos\alpha \sin\lambda'] \\
&+ \quad A_1^- A_2 \quad [+\cos^3\alpha(\sin\lambda' - 1) - \sin^2\alpha \cos\lambda' \cos\alpha] \\
&+ \quad D_1^- D_2 \quad [-\cos\lambda' \sin\alpha \cos\alpha] \\
&+ \quad A_1^- D_2 \quad [+\sin\alpha \cos^3\alpha(\cos\lambda' - 1) + \sin\alpha \cos\alpha \sin\lambda'] \\
&+ \quad D_1^- A_2 \quad [-\sin^2\alpha \cos\alpha \sin\lambda'] \\
&+ \quad A_1^0 A_2 \quad [+2\cos\alpha \sin\lambda' \sin^2\alpha] \\
&+ \quad D_1^0 D_2 \quad [+2\cos\lambda' \sin\alpha \cos\alpha] \\
&+ \quad A_1^0 D_2 \quad [+2\sin^3\alpha \cos\alpha(\cos\lambda' - 1)] \\
&+ \quad D_1^0 A_2 \quad [-2\sin^2\alpha \cos\alpha \sin\lambda'].
\end{aligned} \tag{4.34}
$$

where the notation is similar to that used before, except that $X_1^+ Y_2$ indicates a line centred at $(+\omega_1, \omega_2)$, $X_1^- Y_2$ indicates a line centred at $(-\omega_1, \omega_2)$ and $X_1^0 Y_2$ indicates a line centred at $(0, \omega_2)$.

The twelve components of the spectrum can be described as follows.

$A_1^+ A_2$: absorption-absorption lineshape at (ω_1, ω_2); this corresponds to the desired signal except that its position in ω_1 is shifted from γB_1 to $\gamma B_1 \sec\alpha$.

$D_1^+ D_2$, $A_1^+ D_2$, $D_1^+ A_2$: lines at the same position as $A_1^+ A_2$, but with different lineshapes; these make the line appear improperly phased.

$A_1^- A_2$: absorption-absorption lineshape at $(-\omega_1, \omega_2)$; such mirror image peaks are called "quadrature image" peaks.

$D_1^- D_2$, $A_1^- D_2$, $D_1^- A_2$: further quadrature image peaks with various lineshapes; normally these quadrature image peaks are not seen as the region of the spectrum corresponding to negative ω_1 is not displayed.

$A_1^0 A_2$: absorption-absorption lineshape at $(0, \omega_2)$; such zero-frequency peaks are called axial peaks.

$D_1^0\,D_2$, $A_1^0\,D_2$, $D_1^0\,A_2$: further axial peaks with various lineshapes; these peaks will be partly visible, distorting other peaks at low ω_1 (large distances from the coil).

In addition to these peaks, truncation errors (usually only significant in t_1) will give rise to sinc wiggle artefacts, which may appear as additional peaks.

All these effects are demonstrated in figure 4.5, which shows the result of processing a dataset synthesised according to equations 4.31 and 4.32. The parameters used were $\gamma B_1 \Delta t_1 = 0.75\pi$, $\gamma \Delta B_0 \Delta t_2 = 0.5\pi$, with $\Delta t_1 = 350\mu s$ and $\Delta t_2 = 500\mu s$; these parameters are similar to those used in real experiments. The ideal spectrum would have a single peak at $(0.75\pi, 0.5\pi)$, but since the experiment is far from ideal ($\lambda = 135°$, $\alpha = 25°$), serious artefacts are expected. The Fourier transform spectrum shows five main groups of peaks: (1) is the main peak, (2) is a sinc wiggle, (3) is an axial peak, (4) is the quadrature image of (2), and (5) is the quadrature image of (1). All five peaks have distorted lineshapes, and the main peak is shifted from its expected position towards higher frequencies. This peak is expected to move by a factor of $\sec\alpha$, bringing it to $\omega_1 = 0.83\pi$, but it appears to have moved further than expected; this is a consequence of the distorted lineshape.

By contrast the MEM spectrum (whose calculation is described below) contains a single peak at the expected position.

4.3.4 Application of the MEM to PMRFI

The application of the MEM to PMRFI data[52] is conceptually very simple. A trial spectrum, containing intensity as a function of nominal frequency, is used to calculate two datasets using equations 4.31 and 4.32; these two datasets are the compared with the experimental data, and the trial spectrum adjusted as before. Since exact equations are used to calculate trial data no artefacts are introduced; similarly since no assumptions are made about missing data there should be no problems with sinc wiggles.

The derivations above have largely ignored relaxation effects; in fact the equations only hold when relaxation is extremely slow, so that linewidths are negligible.

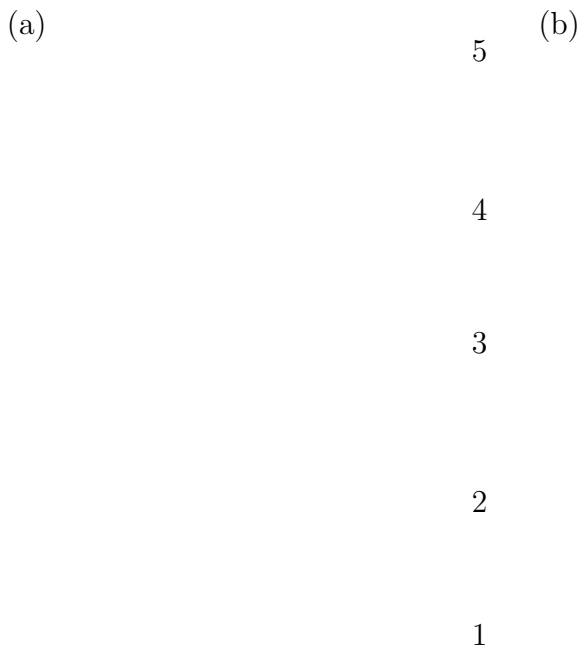

Figure 4.5: Artefacts in a PMRFI spectrum. A dataset was synthesised using equations 4.31 and 4.32, with parameters as described in the main text. The Fourier transform spectrum is plotted in (a), and the MEM spectrum in (b). Note that the spectra are plotted with large ω_1 at the bottom; this is the conventional method for displaying PMRFI spectra, since peaks at the back (top) of the spectrum correspond to nuclei at large distances from the coil. The axes are labelled in fractions of the Nyquist frequency, and only the region of the spectrum corresponding to positive ω_2 is shown. Contours are drawn at ± 5, 10, 20, 40, 80% of the maximum intensity in each spectrum; negative contours are dashed. Peak labels are explained in the text.

This is not true in practice, and is not true for the synthetic data used to calculate figure 4.5; this effect can be seen as the slightly unusual lineshape in the MEM spectrum (b).

4.4 Results

The section above indicates that the MEM can cope well with processing synthesised "data" from PMRFI experiments, but this may provide an unrealistic picture of the method's ability to handle real data.

In order to investigate this point, three genuine datasets were processed. The first two datasets were obtained from simple samples, commonly called phantoms[6]; this allows particular features of MEM processing to be studied. The final dataset was obtained from the abdomen of a human subject[7].

4.4.1 Truncation Artefacts

The first phantom contained two cylindrical compartments, each of diameter 10 mm and thickness 6 mm, separated by a distance of 6 mm, arranged along the axis of the coil, so that the whole system was cylindrically symmetric. The compartments contained dilute phosphoric acid and the remainder of the phantom was empty. In the absence of processing artefacts the spectrum should contain two peaks, one from each compartment. The data were recorded with the B_1 field resonant with the phosphoric acid signal (so that $\alpha = 0$) and with the compartments near the centre of the sensitive region (so that $\lambda \approx 90°$); consequently the only significant artefacts should be due to truncation.

The effects of truncation are shown in figure 4.6. The dataset was collected using 32 increments in t_1, which is more increments than is normally possible, but the FT spectrum (a) still shows some evidence of sinc wiggles. If only the first 16 increments are used, with the missing data replaced by zero-filling, the FT

[6]This data was recorded by Dr R. Ouwerkerk, MRC Biochemical and Clinical Magnetic Resonance Unit, John Radcliffe Hospital, Oxford.

[7]This data was recorded by Dr R. Dixon, Department of Biochemistry, University of Oxford.

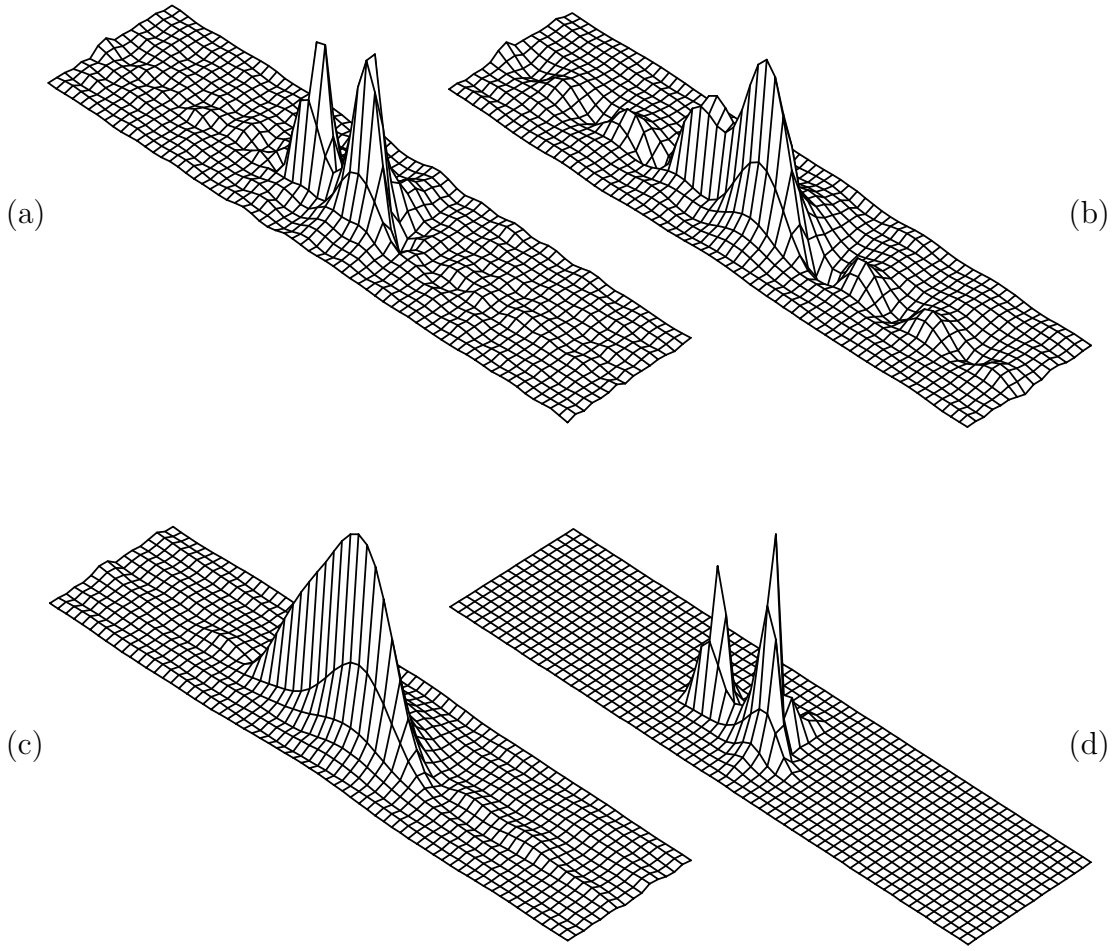

Figure 4.6: Truncation artefacts in a PMRFI spectrum. The data was obtained from a simple two compartment phantom, as described in the text. Only part of each spectrum is shown, corresponding to positive frequencies in ω_1, and to the region immediately around the signal in ω_2. The longer axis corresponds to the spatial dimension, with distant nuclei appearing at the back (top) of each spectrum; the shorter axis corresponds to chemical shift. The dataset contained 16 or 32 increments in t_1, which was zero-filled to give 128 points in ω_1 (of which 64 are displayed). Even with 32 increments in t_1 the FT spectrum (a) shows small sinc wiggles, and if the number of increments is reduced to 16 there are large wiggles in the FT spectrum (b). These wiggles can be suppressed using an apodisation function (c), but this results in additional linebroadening and complete loss of spatial resolution. The MEM spectrum, obtained from 16 increments, is shown in (d); the truncation artefacts are well suppressed.

spectrum shows extensive sinc wiggles (b); these can be largely suppressed using a severe apodisation function in t_1, but this results in serious line broadening (c). By contrast the MEM spectrum (d), obtained using the first 16 increments, shows no evidence of truncation artefacts.

Typically PMRFI datasets are obtained using between 8 and 16 increments in t_1, and so truncation artefacts are a serious problem. Their suppression by the MEM may prove useful in some cases.

4.4.2 Off Resonance Artefacts

The second phantom also contained two compartments. In this case they were 1 cm thick, 2 cm apart, and filled with 150 mM ATP solution. ATP contains three phosphate groups (labelled α, β and γ), each of which will give rise to a signal, and so the ideal spectrum should contain two groups of three lines. This dataset was acquired using typical parameters, which result in the γ and α-ATP peaks being close to resonance, but the β-ATP peak being far from resonance. Both off resonance effects and pulse length errors should be important. The dataset contained 32 increments in t_1, and so truncation effects should be relatively unimportant.

The results of Fourier transform and Maximum Entropy processing are shown in figure 4.7; each spectrum is displayed as both a stack plot and a contour plot.

The FT spectrum shows that the β-ATP peaks (which are well off resonance) appear artefactually close to the front of the spectrum. The shift is more pronounced for the signal from the deeper compartment, where the B_1 field is less intense and the off resonance effects consequently more severe. All three signals give rise to substantial axial peaks, and the front β-ATP peak has a large quadrature image peak.

By contrast the MEM spectrum is entirely free of artefact peaks, and the six genuine peaks are all at their correct positions. The β-ATP peaks show some evidence of additional broadening in ω_1, but otherwise the spectrum is exactly as expected. This clearly demonstrates the ability of the MEM to cope with off resonance effects and pulse length errors.

(a) (c)

(b) (d)

Figure 4.7: PMRFI spectra from a two compartment phantom containing ATP. The FT spectrum is shown in (a) and (b), while the MEM spectrum is shown in (c) and (d). The positions of the signals from the α, β and γ phosphate groups of ATP are indicated by arrows. In order to illustrate all the artefacts both positive and negative frequencies in ω_1 are shown. In (b) AX denotes axial peaks and QI denotes quadrature image peaks. Contours are drawn at $\pm$5,10, 20, 40, 80% of the maximum intensity in each spectrum; negative contours are dashed.

4.4.3 Experimental Data

The final dataset was collected from the abdomen of a human subject, and is typical of the use of PMRFI for medical studies. Fourier transform and Maximum Entropy spectra are shown in figure 4.8. The FT spectrum is difficult to interpret, although with experience it is possible to analyse spectra such as this. By contrast the MEM spectrum is extremely clear, containing seven well resolved peaks. Pi is a signal from phosphate ions (frequently described as "inorganic phosphate"), PDE is due to signals from a range of phosphodiesters, PCr comes from phosphocreatine, and the three peaks labelled α, β and γ are due to ATP as before. The peak labelled "phantom" is a calibration signal, arising from a small bulb built into the coil assembly. The peaks in the MEM spectrum are largely undistorted and occur in their expected positions. In particular it is possible to distinguish three separate spatial layers: (i) the coil assembly containing only the phantom; (ii) a superficial layer corresponding to intercostal muscle, which contains inorganic phosphate, phosphocreatine and ATP; (iii) a deeper layer, corresponding to the liver, which contains no phosphocreatine but does contain inorganic phosphate, ATP, and a range of phosphodiesters.

(a) (c)

(b) (d)

Figure 4.8: PMRFI spectra obtained from the abdomen of a human subject; the surface coil was placed adjacent to a region where ribs and intercostal muscle lie over the liver. The acquisition parameters were typical for such an experiment, and 16 increments were acquired in t_1. The FT spectrum, depicted in (a) and (b), was calculated using apodisation in t_1 followed by zero-filling to give 128 points in ω_1 of which 64 (corresponding to positive ω_1) are displayed; matched filtration was applied in t_2. The MEM spectrum is depicted in (c) and (d). Only a limited range of frequencies in ω_2 are shown; zero frequency coincides with the peak labelled PCr. Peak labels are described in the text.

Chapter 5

J-deconvolution

Essentia non sunt multiplicanda praeter necessitatem.

John Ponce's summary of the ideas of William of Ockham

J-deconvolution is a convenient term for a variety of data processing techniques which aim to remove splittings from NMR spectra. As described in chapter 1 these splittings are due to spin-spin coupling, also called J-coupling. While these splittings can be extremely useful (they provide information about interactions between two or more nuclei), they can also act to obscure other sorts of valuable information, and their removal can simplify spectra in a useful fashion. In particular the removal of one or more splittings from a complicated multiplet may simplify the analysis of other splittings in the multiplet.

The first section of this chapter introduces the COSY experiment, in which such spectral simplifications are extremely useful, and describes the process of direct deconvolution. This is in general extremely unstable, and of little or no practical use. The next section describes a number of alternative methods which attempt to side-step the problems of direct deconvolution in various ways. The third section describes the application of these methods to several real examples, based on the COSY spectrum of the amino acid iso-leucine. The final section summarises the results of the preceding sections and indicates the relative merits of the various deconvolution methods described.

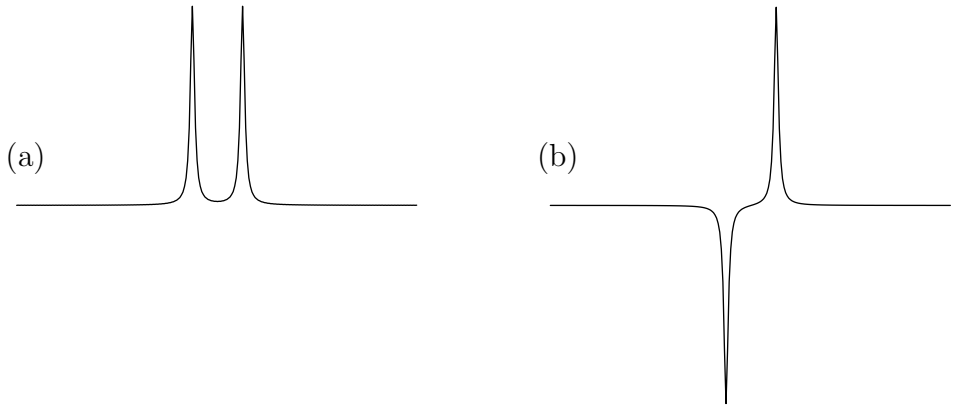

Figure 5.1: Spectral splittings: (a) inphase, (b) antiphase.

5.1 Introduction

Spectral splittings can be conveniently divided into two groups, called inphase and antiphase splittings[1]; the form of these is indicated in figure 5.1. Simple NMR spectra only contain inphase splittings, but antiphase splittings occur in spectra from a number of more complex experiments, such as COSY.

COSY (COrrelation SpectroscopY)[3, 4] is a two dimensional spectroscopic experiment of enormous importance for the detection and analysis of J-couplings. The appearance of a COSY spectrum can only be explained using a full quantum mechanical description of the experiment, but it can be described using the ideas developed in previous chapters. Consider a system containing two nuclei, A and B, with Larmor frequencies ω_A, ω_B, and a spin-spin coupling J. The corresponding COSY spectrum contains peaks in four regions as follows:

1. peaks around (ω_A, ω_A) and (ω_B, ω_B), called diagonal peaks;

2. peaks around (ω_A, ω_B) and (ω_B, ω_A), called cross peaks.

The presence of cross peaks is indicative of the coupling between A and B, and so COSY spectra are a useful method for detecting couplings. Each region contains

[1]This simple description is only appropriate to "weakly coupled" systems, in which the coupling constant, $|J|$, is much smaller than the difference in the Larmor frequencies of the coupled nuclei, $|\Delta\omega|$. If $|J|$ is comparable to $|\Delta\omega|$, the system is described as "strongly coupled", and more complex intensity patterns can occur[97].

Peak type	Components
Diagonal	$+D_1(\omega_A - J/2)D_2(\omega_A - J/2)$
	$+D_1(\omega_A - J/2)D_2(\omega_A + J/2)$
	$+D_1(\omega_A + J/2)D_2(\omega_A - J/2)$
	$+D_1(\omega_A + J/2)D_2(\omega_A + J/2)$
	$+D_1(\omega_B - J/2)D_2(\omega_B - J/2)$
	$+D_1(\omega_B - J/2)D_2(\omega_B + J/2)$
	$+D_1(\omega_B + J/2)D_2(\omega_B - J/2)$
	$+D_1(\omega_B + J/2)D_2(\omega_B + J/2)$
Cross	$+A_1(\omega_A - J/2)A_2(\omega_B - J/2)$
	$-A_1(\omega_A - J/2)A_2(\omega_B + J/2)$
	$-A_1(\omega_A + J/2)A_2(\omega_B - J/2)$
	$+A_1(\omega_A + J/2)A_2(\omega_B + J/2)$
	$+A_1(\omega_B - J/2)A_2(\omega_A - J/2)$
	$-A_1(\omega_B - J/2)A_2(\omega_A + J/2)$
	$-A_1(\omega_B + J/2)A_2(\omega_A - J/2)$
	$+A_1(\omega_B + J/2)A_2(\omega_A + J/2)$

Table 5.1: COSY spectrum from a system containing two nuclei, A and B, with a spin-spin coupling J. The spectrum contains 16 peaks, which can be divided into four groups of four. Peaks in two of these groups have $\omega_1 \approx \omega_2$, and are called diagonal peaks; peaks in the other two groups are called cross peaks.

four peaks, centred at $(\omega_1 \pm J/2, \omega_2 \pm J/2)$; conventionally the data is processed so that the diagonal peaks are in dispersion-dispersion mode and the cross peaks are in absorption-absorption mode. The splittings between the diagonal peaks are inphase, while those between the cross peaks are antiphase. The overall spectrum is described in table 5.1 and depicted in figure 5.2. Note that the spectrum is completely symmetrical around the diagonal.

The system described above is, of course, extremely simple, but COSY spectra from more complex systems are similar. Each distinct nucleus gives rise to a group of diagonal peaks, and each pair of coupled nuclei gives rise to two groups of cross peaks. Additional couplings in the system appear as additional splittings on the diagonal and cross peaks. These additional couplings are called passive couplings, and give rise to inphase splittings, while the coupling causing the cross peak is called the active coupling and gives rise to an antiphase splitting. As before the spectrum is symmetrical about the diagonal. This is shown in figure 5.3 for the case of three coupled spins. Cross sections through this spectrum, at the positions

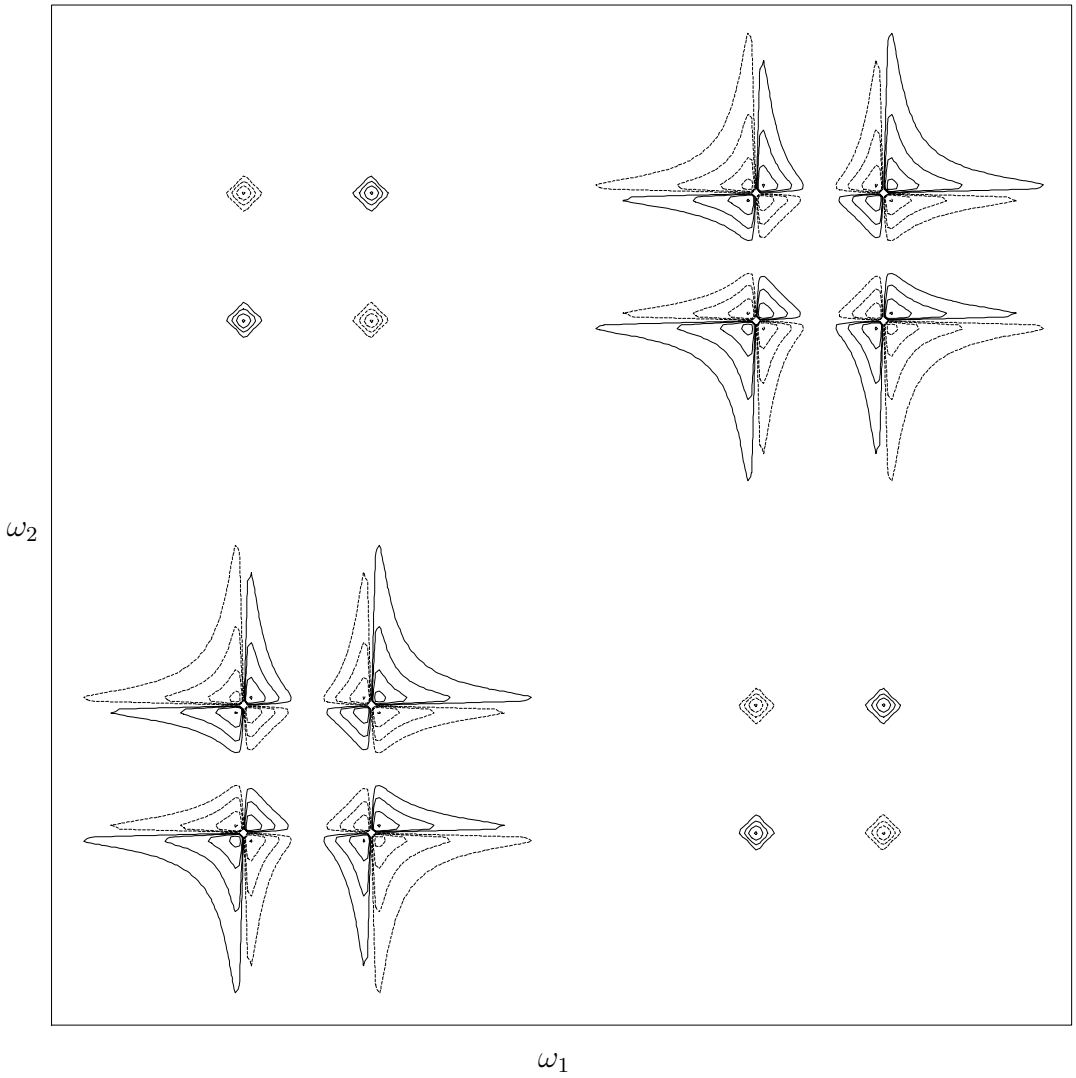

Figure 5.2: Schematic COSY spectrum from a pair of coupled nuclei. Weak coupling has been assumed, even though the coupling constant is comparable to the difference in Larmor frequencies. The diagonal peaks are in dispersion-dispersion mode and the cross peaks in absorption-absorption mode (see figure 4.1); interference between the diagonal peaks results in the characteristic "angel" peak shape.

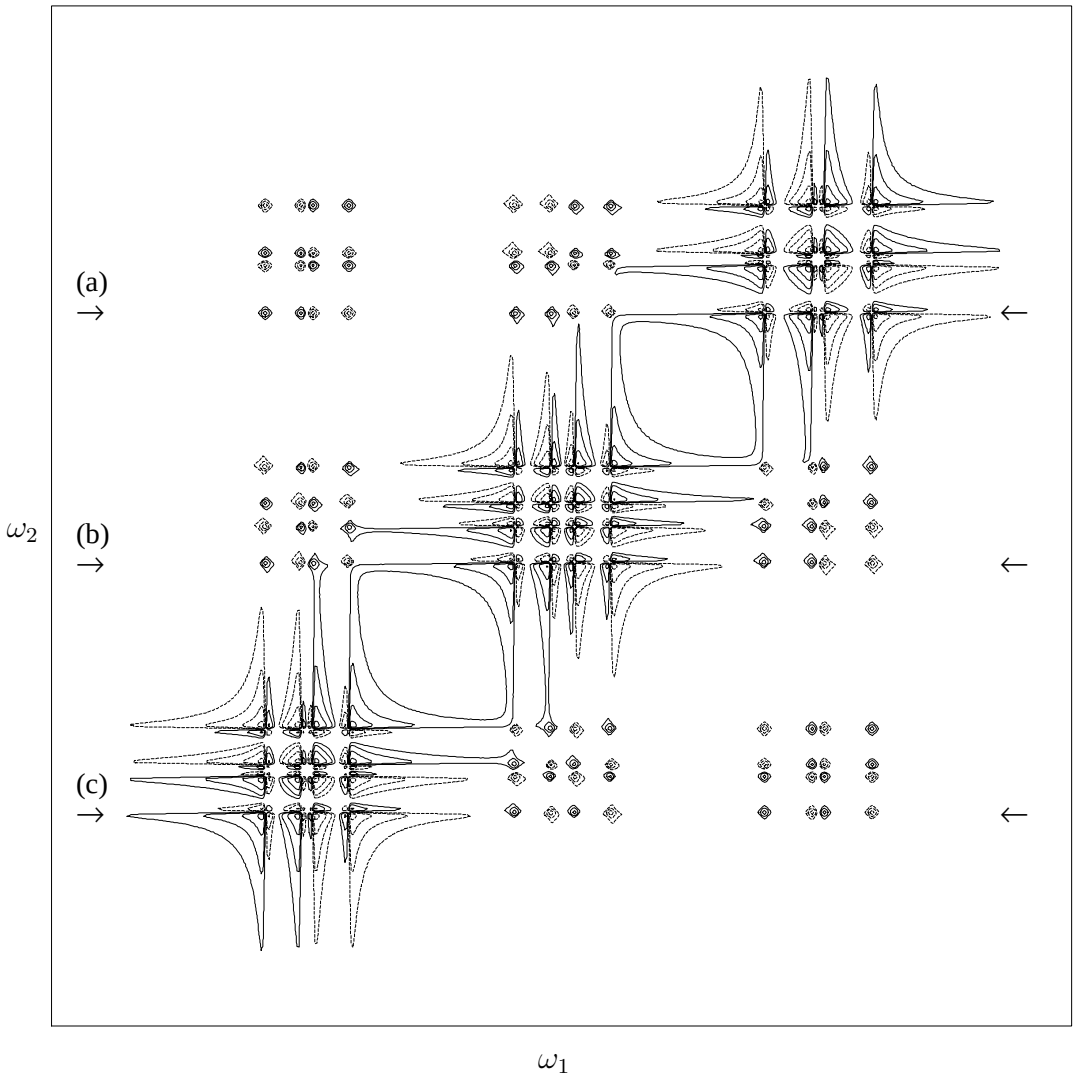

Figure 5.3: Schematic COSY spectrum from three coupled nuclei. Weak coupling has been assumed, even though the coupling constants are comparable to the differences in Larmor frequencies. The diagonal peaks are in dispersion-dispersion mode, but interference between them results in complex shapes. The cross peaks are in absorption-absorption mode. Note that each cross peak group shows one antiphase splitting (due to the active coupling), which has the same size in each dimension; additional passive couplings result in inphase splittings which are different in the two dimensions. Cross sections through the spectrum at the positions marked by arrows are shown in figure 5.4.

marked by arrows, are shown in figure 5.4; the pattern of antiphase and inphase splittings is clearly displayed.

Each cross peak group contains exactly one active (antiphase) splitting, but can contain many passive (inphase) splittings, resulting in complex patterns. The antiphase splitting can result in partial cancellations between two or more lines, further complicating the appearance of the spectrum. Removing the antiphase splitting removes the cancellation problems, and will greatly simplify the spectrum.

Rather than using the whole spectrum, it is usually sufficient to extract the cross peak of interest and work with this. Furthermore it may be sufficient to work with one dimensional cross sections taken from the two dimensional spectrum, rather than using the entire cross peak. All the examples in this chapter are based on such cross sections, which are called multiplets by analogy with multiplets in simple one dimensional NMR spectra. In general it is more convenient to use time domain data, and so the multiplet is back Fourier transformed to give a "pseudo-FID".

5.1.1 Direct Deconvolution

The introduction of a splitting into a spectrum can be conveniently described as the convolution of the spectrum with an appropriate pattern of δ-functions; for inphase and antiphase splittings these correspond to inphase and antiphase "stick-doublets", as depicted in figure 5.5. Using the convolution theorem (see chapter 2) convolution in the frequency domain can be replaced by multiplication of the corresponding Fourier transform pairs: convolution with an inphase doublet corresponds to multiplication by $\cos(\pi Jt)$, while convolution with an antiphase doublet corresponds to multiplication by $i\sin(\pi Jt)$.

Deconvolution can be achieved by *division* of the corresponding Fourier transform pairs. For example the time domain signal corresponding to a multiplet containing one antiphase splitting is

$$s(t) = s_p(t) i \sin(\pi J_a t) \tag{5.1}$$

where J_a is the antiphase (active) splitting, and s_p describes the remaining (passive)

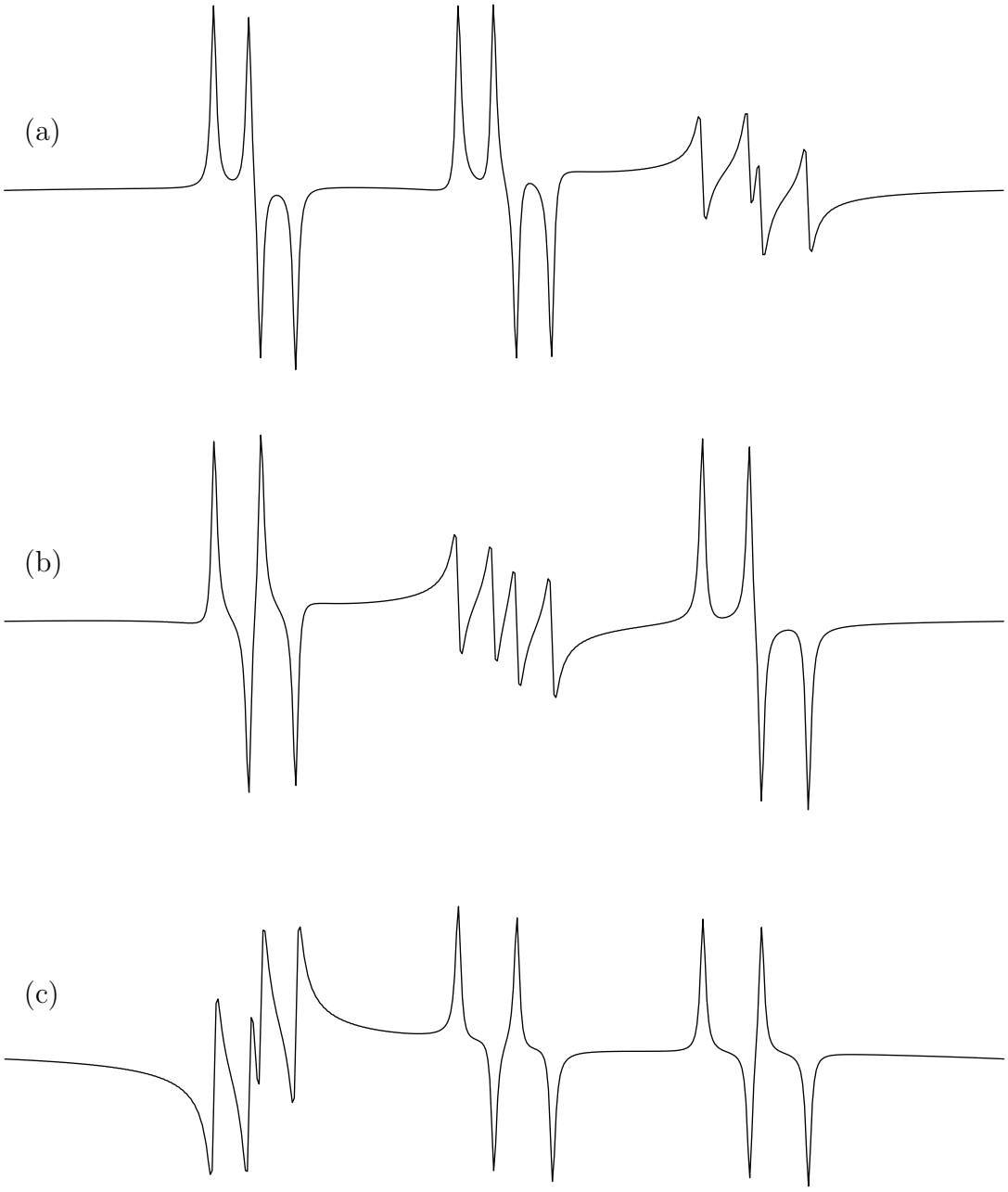

Figure 5.4: Cross sections through the COSY spectrum shown in figure 5.3 at the positions marked by arrows.

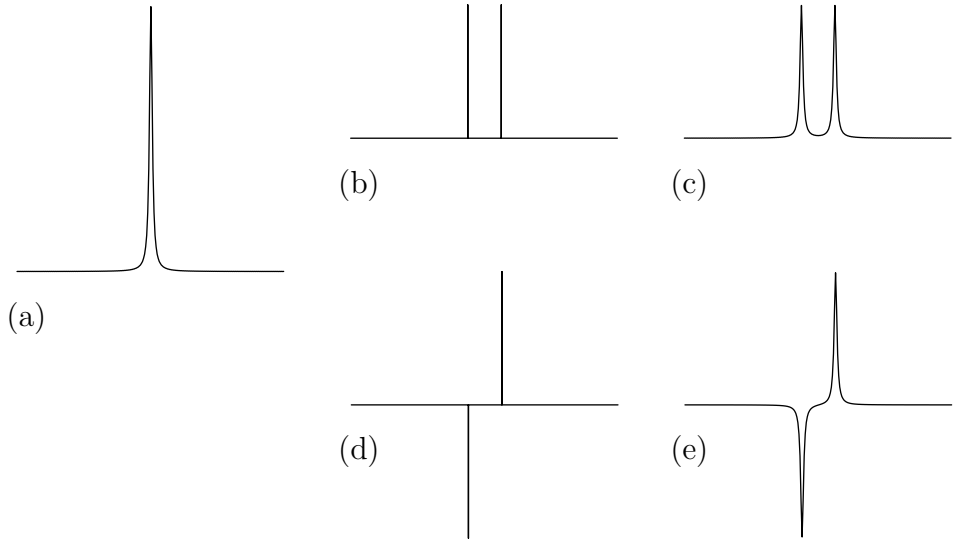

Figure 5.5: Convolution of a singlet (a) with an inphase stick doublet (b) gives an inphase doublet (c); convolution of (a) with an antiphase stick doublet (d) gives an antiphase doublet (e).

structure. Obviously

$$\frac{s(t)}{i\sin(\pi J_a t)} = s_p(t), \tag{5.2}$$

and Fourier transformation gives a spectrum containing only the passive structure. Unfortunately this division is very unstable: $\cos(\pi J_p t)$ and $\sin(\pi J_a t)$ are oscillatory functions, and are liable to contain exact or near zero-crossings; indeed $\sin(\pi J t)$ always contains at least one zero-crossing (at $t = 0$). Exact zero-crossings make direct division impossible, while near zero-crossings will result in massive noise amplification. Some method of side-stepping the zero-crossings or of stabilising the division must be found.

5.1.2 Determining J

Even if some method of stabilising the direct deconvolution is found, one further problem remains: determining the value of J. Clearly no form of deconvolution is possible unless this value is known. In simple cases it is possible to estimate J by measuring the distance between the centres of two peaks, but if the two peaks are poorly resolved this estimate can be badly wrong. Peak overlap reduces the apparent size of an inphase splitting, and increases the apparent size of an

antiphase splitting[98]. When several splittings are present, cancellation effects may render direct measurement of the various splittings impossible.

In general it should be possible to determine the value of J by trial and error[2]. If a multiplet is deconvolved using an inappropriate value of J, the resulting spectrum will contain a variety of artefactual peaks; minimising these artefacts should provide a reasonable criterion for estimating J.

Since cross sections through COSY cross peaks only contain one antiphase splitting, but may contain many different inphase splittings, the determination of the antiphase splitting is generally much simpler than the determination of the various inphase splittings. Furthermore, antiphase splittings obscure the multiplet structure more that inphase splittings do (as a result of cancellation between different regions of the multiplet). Consequently the sections below will concentrate on the removal of antiphase splittings, and will only discuss inphase splittings where their treatment differs significantly.

5.2 Methods

5.2.1 Stabilised Deconvolution

As described above, the simplest method of J-deconvolution is direct division of the FID by $\cos(\pi J_p t)$ or $i\sin(\pi J_a t)$ as appropriate, but this direct method is impossible in the presence of zero-crossings. In favourable cases this is not a problem, since exact zero-crossings are rare. The data is recorded at discrete points in time, and a zero-crossing is only significant if it coincides with one of these points. For antiphase splittings this occurs when $Jt = n$ (where n is an integer), and for inphase splittings this occurs when $Jt = n + \frac{1}{2}$. Of course antiphase splittings always give rise to a zero-crossing at $t = 0$, but this point only determines the integral of the

[2]The methods described in this chapter should not be confused with the distantly related method of Titman and Keeler[99], which estimates J using both a COSY spectrum and a TOCSY spectrum. The TOCSY spectrum[100] is very similar to the COSY spectrum, except that the active coupling appears as an inphase splitting. The COSY multiplet is convolved with an inphase stick doublet of splitting J_a^*, and the TOCSY multiplet is convolved with an antiphase stick doublet of the same size; when $J_a^* = J_a$ the resulting multiplets should be identical. The method works by varying J_a^* to minimise the χ^2 difference between the two convolved multiplets.

spectrum, and can generally be adjusted by hand to give a reasonable baseline. However the problem of near zero-crossings remains serious; these can give rise to massive amplification of noise components.

Bothner-By and Dadok have suggested a remedy to this problem, in which the envelope of the original FID is used as cut-off to limit the size of noise spikes introduced by the division[101]. This can certainly reduce the size of the problem, but it is only a partial solution, and is difficult to implement.

A more elegant method, proposed by Le Parco, McIntyre and Freeman, avoids the zero-crossings completely by changing the sampling times[102]. The data is first extensively interpolated, and then resampled at times well away from zero-crossings. In practice this method works extremely well with synthetic data and high quality experimental data, but works less well with low quality experimental data. In particular it is necessary to flatten the baseline in such multiplets in order to avoid the introduction of serious artefacts[53].

This method can also be used to determine J by trial and error. If an incorrect value is used the resulting spectrum contains a host of sidebands, which can be readily detected by eye. It is also possible to automate this process by minimising the integral of the modulus of the deconvolved spectrum[53, 102].

While stabilised J-deconvolution can work quite well, the closely related technique of J-doubling gives very similar results and is much simpler to implement; consequently stabilised J-deconvolution will not be used in the remainder of this chapter.

5.2.2 J-doubling

The method of J-doubling[103], suggested by McIntyre and Freeman, solves the overlap problem not by collapsing the two halves of the multiplet onto one another, but by "dragging" them apart. In order to deal with an antiphase splitting the data is multiplied by

$$\prod_{m=0}^{n-1} \cos(\pi 2^m J_a t). \tag{5.3}$$

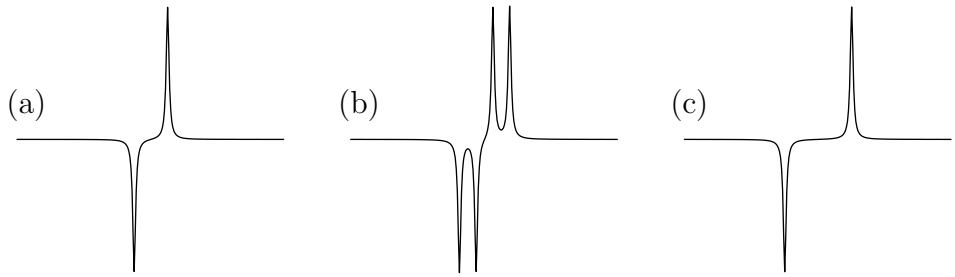

Figure 5.6: J-doubling of the antiphase doublet (a). Convolution of (a) with an inphase stick doublet of splitting $J_a^* \neq J_a$ results in a spectrum containing four lines (b); when $J_a^* = J_a$ the two central lines cancel, resulting in an antiphase doublet of splitting $2J_a$ (c).

Since $\sin(2\theta) \equiv 2\sin\theta\cos\theta$, this multiplication converts a modulation of $i\sin(\pi J_a t)$ to one of $i\sin(\pi 2^n J_a t)$. Fourier transformation of the new dataset gives a multiplet with the antiphase splitting scaled up by a factor of 2^n; the value of n can be chosen so that the two halves of the multiplet are completely separated. This is depicted in figure 5.6 for the case of $n = 1$. Similarly an inphase splitting can be dealt with by multiplying the data by

$$i\sin(\pi J_p t)\prod_{m=1}^{n-1}\cos(\pi 2^m J_p t); \tag{5.4}$$

this converts an inphase splitting to a scaled antiphase splitting.

J-doubling gives very similar results to stabilised J-deconvolution, but is much simpler to implement as it is not necessary to interpolate the data. Once again it is necessary to flatten the baseline of the multiplet before doubling the splittings. J-doubling can also be used to estimate J in the same way as for J-deconvolution: doubling is performed with a variety of trial splittings J_a^* and the area of the modulus spectrum is minimised with respect to J_a^*. This is demonstrated in figure 5.7 for the data shown in figure 5.6a. The result depends on the value of n (see equation 5.3): for $n = 1$ there is a single minimum at $J_a^* = J_a$; for $n = 2$ this minimum is sharper and deeper, but small false minima occur at fractions of the correct splitting. With more complex multiplets, false minima also occur at sums and differences of the antiphase and the various inphase splittings[53, 103].

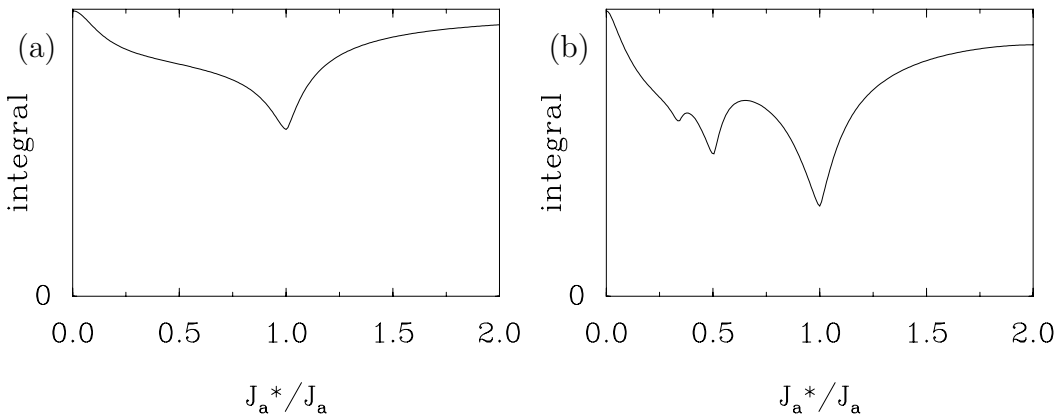

Figure 5.7: Estimation of J_a by J-doubling for the data shown in figure 5.6a. The modulus area of the J-doubled spectrum is plotted as a function of the trial splitting, J_a^*, for (a) $n = 1$, and (b) $n = 2$ (see equation 5.3). In each case a minimum occurs at $J_a^* = J_a$; for $n = 2$ subsidiary minima are also visible at $J_a/2$ and $J_a/3$.

5.2.3 Deconvolution by the MEM

J-deconvolution can be achieved by the MEM[29, 47, 49, 53, 104] using the same technique as was used for resolution enhancement in chapter 2. A trial spectrum is postulated and back transformed to give trial data; this is then multiplied by $i\sin(\pi J_a t)$ before being compared with the experimental data. It is possible to combine J-deconvolution with resolution enhancement by multiplying the trial data by both an oscillatory and a decaying function. The process is depicted in figure 5.8. Similarly it is possible to use the MEM to suppress truncation artefacts at the same time as performing J-deconvolution. In particular removal of the antiphase splitting allows the MEM to be used with cross sections from COSY cross peaks; without deconvolution this is not possible as the multiplets contain regions of negative intensity[3].

As usual J-deconvolution by the MEM only works when the value of J is known in advance. It might appear possible to determine J by trial and error, but this is difficult. If an incorrect value of J is used, the method will generally not converge.

[3]It is, of course, possible to use one of the variations of the MEM which allows spectra to contain regions of negative intensity, but these methods are relatively poor at the suppression of truncation artefacts.

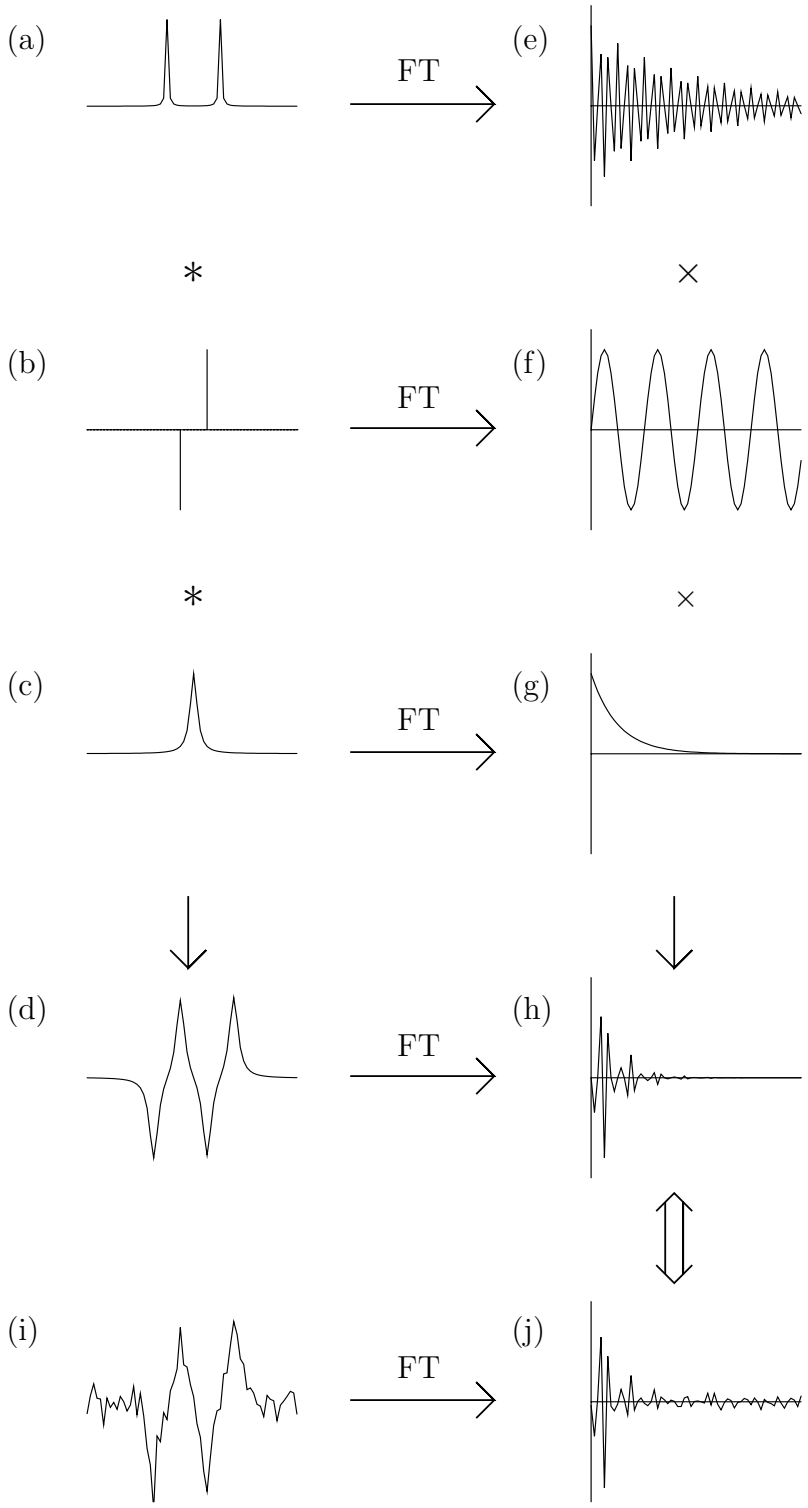

Figure 5.8: J-deconvolution by the Maximum Entropy Method. (a) is the trial spectrum, (b) is an antiphase stick doublet, (c) is a lineshape function and (d) is the convolution of (a) with (b) and (c); (e)–(h) are Fourier transforms of (a)–(d) respectively. (h), the trial data, is the product of (e), (f) and (g); this is compared with the experimental data, (j), which is the Fourier transform of the original multiplet (i).

Nonconvergence could be used as a criterion for eliminating incorrect values of J, but it is poorly defined. With complex multiplets the method may converge at several incorrect values of J, corresponding to sums and differences between splittings, and distinguishing these from the correct value is difficult.

An extension of the method has been suggested by Delsuc and Levy[29] as a way of determining J_a. The method starts with many trial spectra instead of just one; these are Fourier transformed, multiplied by $i\sin(\pi J_a^* t)$, with a different value of J_a^* for each spectrum, and then added together to give the trial data. The trial data is then matched to the experimental data as before. This results in a two-dimensional maximum entropy spectrum which should contain the deconvolved multiplet in the sub-spectrum that corresponds to the true splitting ($J_a^* = J_a$), and no intensity in the other sub-spectra (corresponding to other, incorrect, splittings). In fact with complex multiplets the method places intensity in many different sub-spectra[53], so that it is difficult or impossible to determine J_a.

5.2.4 Constrained χ^2-Minimisation

Constrained χ^2-minimisation is a simple, but relatively unknown, method for performing deconvolutions[53]. Like the MEM, the method starts off by postulating a trial spectrum; this is Fourier transformed and weighted to give trial data which is compared with the experimental data by means of a χ^2 parameter. The trial spectrum is then varied to minimise χ^2. In its simplest form the method is unstable: points which correspond to zero-crossings in the weighting function make no contribution to χ^2, and so are free to take any value. Any function whose Fourier transform is non-zero only at these points can be added to the trial spectrum without affecting χ^2, allowing the introduction of massive artefacts. This is demonstrated in figure 5.9 for the case of inphase J-deconvolution. In order to determine these points, and so discriminate against the artefacts, further constraints must be imposed. These constraints are determined by prior knowledge of the form of the desired spectrum.

For example it is commonly expected that the deconvolved spectrum will be

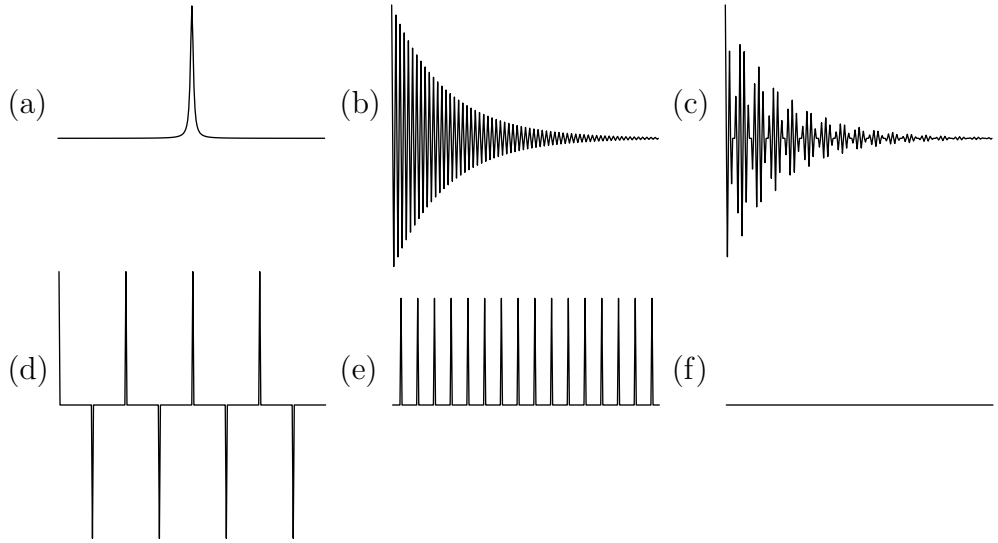

Figure 5.9: Inphase J-deconvolution by χ^2-minimisation. The trial spectrum (a) has corresponding time domain data (b); this is multiplied by the weighting function $\cos(\pi J_p t)$ to give trial data (c) which is then matched to the experimental data. The spectrum (d) has time domain data (e), which is non-zero only at the zero-crossings of the weighting function, and so its trial data (f) is everywhere zero. Consequently spectra like (d) can be added to the trial spectrum without affecting the value of χ^2.

everywhere positive; in this case the techniques described in chapter 2 can be used. This constraint is sufficient to stabilise the deconvolution of an inphase doublet. In this case the weighting function, $\cos(\pi J_p t)$, is zero when $J_p t = n + \frac{1}{2}$ (where n is an integer), and any function whose Fourier transform is an appropriate "comb" of δ-functions can be added to the trial spectrum. These functions are oscillatory, and so contain negative regions. If the multiplet is surrounded by a region of flat baseline, it is impossible to add such functions to the trial spectrum without introducing negative regions; hence the positivity constraint is sufficient to prevent the introduction of artefacts. If the baseline is distorted this method will be less successful, but this can be tackled using the method for antiphase splittings outlined below.

With antiphase splittings the simple positivity constraint is not sufficient to prevent the introduction of artefacts. The weighting function is now $i\sin(\pi J_a t)$, which is zero when $J_a t = n$; consequently the first point of the trial data ($t = 0$)

is undetermined. Once again a range of oscillatory functions can be added to the trial spectrum without affecting χ^2, but in addition any constant offset can also be added (the time domain data corresponding to a constant offset is zero everywhere except at the first point). Hence it is possible to add any oscillatory function to the trial spectrum, as long as a sufficiently large offset is added at the same time. The solution to this is to constrain the wings of the trial spectrum, on either side of the deconvolved multiplet, to have zero intensity. This second constraint is reasonable: neglecting noise terms the deconvolved spectrum is expected to have no intensity outside the multiplet itself.

Constrained χ^2 minimisation also lends itself to the determination of J. As usual deconvolution is performed using a range of trial values, and the J-value which results in the smallest value of χ^2 is chosen. Incorrect J-values require the trial spectrum to contain negative intensity if the experimental data is to be exactly matched; since the method forbids such negative intensity the resulting value of χ^2 will be high.

As described in chapter 2, the constrained minimisations were performed using the NAG routine `E04JAF`[83].

5.2.5 Curve Fitting

Curve fitting is not properly speaking a method of deconvolution, since it depends on prior knowledge of the approximate form of the deconvolved spectrum. It does, however, provide a very convenient method for obtaining spectral parameters, such as coupling constants. Once a multiplet has been partially deconvolved by some other method, in order to obtain initial estimates of these parameters, curve fitting can be used to refine their values. As usual the initial values must be fairly good, to prevent the fitting routine becoming lost in local extrema.

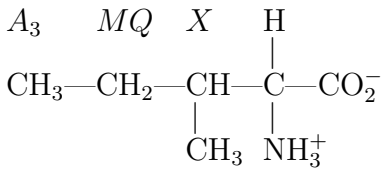

Figure 5.10: Structure of the amino acid iso-leucine. The letters in italics are labels for the three identical methyl protons (A_3), the two distinct methylene protons (MQ) and the single methine proton (X) referred to in the text.

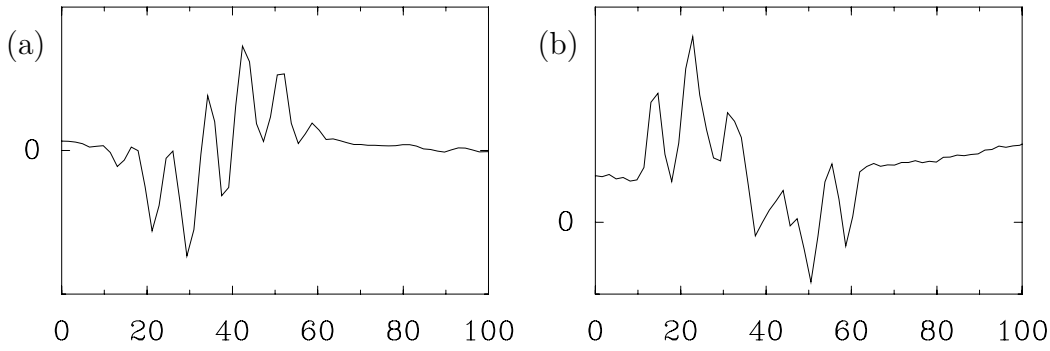

Figure 5.11: Cross sections through cross peaks from the COSY spectrum of iso-leucine: (a) the MQ cross peak, (b) the AM cross peak. The horizontal (frequency) axis is in Hz, the vertical (intensity) axis is in arbitrary units. The severe baseline distortion in (b) arises from the wings of a diagonal peak.

5.3 Applications

The various J-deconvolution methods outlined above were applied to two multiplets taken from the COSY spectrum of the amino acid iso-leucine[4] (see figure 5.10). The COSY spectrum was obtained by conventional processing, involving zero-filling and Lorentzian to Gaussian lineshape transformation. Cross sections through the MQ and AM cross peaks were extracted, and Fourier transformed to give data for subsequent processing. These cross sections are shown in figure 5.11; note that they show the multiplet structure of proton M in each case. Each multiplet contains one antiphase splitting and several passive splittings, and in both cases overlap obscures the underlying structure.

[4]This data was recorded by Dr D. S. Grainger, Physical Chemistry Laboratory, University of Oxford.

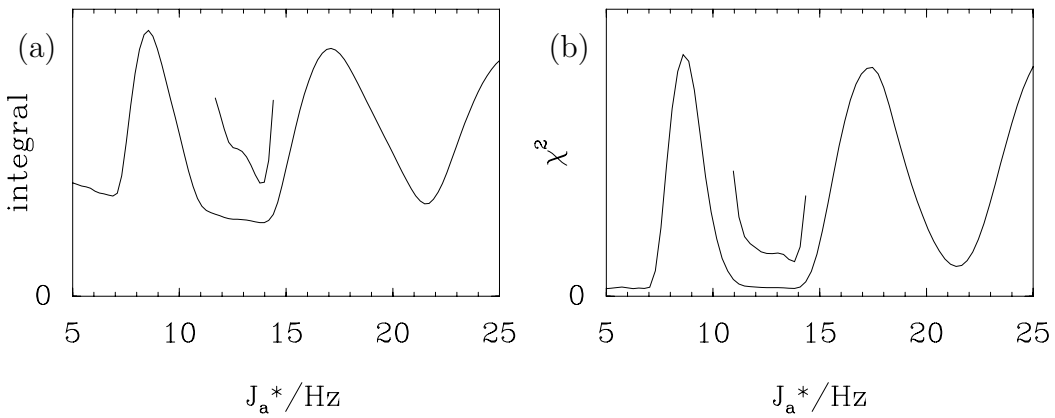

Figure 5.12: Estimation of J_a for the MQ cross peak by J-doubling and constrained χ^2-minimisation: (a) the integral of the modulus of the J-doubled spectrum ($n = 5$) as a function of J_a^*; (b) the minimised value of χ^2 as a function of J_a^*. Part of each plot is drawn with the vertical scale expanded to emphasise the positions of the minima.

5.3.1 The MQ Cross Peak

The value of the antiphase splitting, J_a, was estimated using J-doubling and constrained χ^2-minimisation; the results are shown in figure 5.12. J-doubling gives a very clear indication of the active coupling constant (13.8 Hz), with subsidiary minima around 6 Hz and 22 Hz. χ^2-minimisation is slightly more ambiguous: the minima occur in similar places, but the minimum around 6 Hz is slightly lower than that at 13.8 Hz. The subsidiary minima occur around fractions of the true antiphase splitting and sums and differences of the antiphase splitting with the various passive splittings (about 8 Hz).

Deconvolved spectra obtained using J_a=13.8 Hz are shown in figure 5.13. The spectra obtained by J-doubling (a) and χ^2-minimisation (b) contain a quintet of lines with approximate relative intensities 1:4:6:4:1. Attempts to resolve further splittings, for example by additional resolution enhancement of the multiplets prior to deconvolution or direct resolution enhancement of the deconvolved spectra, were unsuccessful. J-deconvolution gave very similar results to J-doubling.

The spectrum obtained by the MEM with resolution enhancement (c) shows clear evidence for further splittings. The quintet is resolved into a quartet of doublets, with approximate coupling constants of 7.6 Hz for the quartet (J_{AM}) and

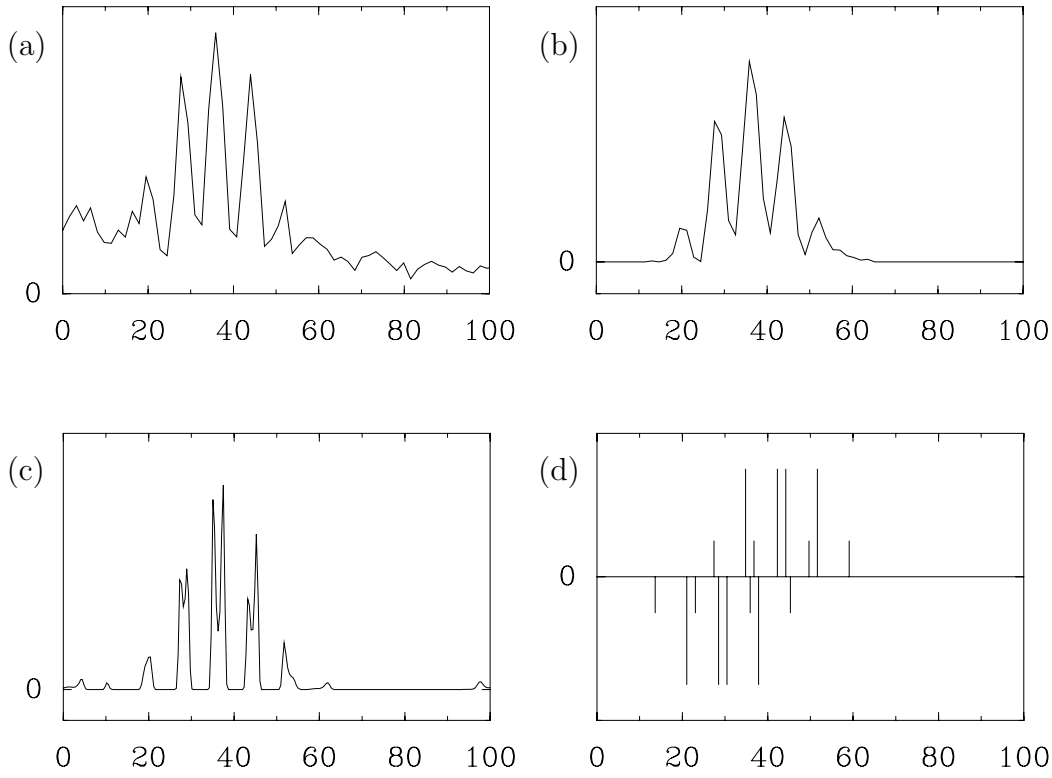

Figure 5.13: Spectra obtained from figure 5.11a after removal of an antiphase splitting with J_a=13.8 Hz: (a) spectrum obtained by J-doubling using $n = 5$ (see equation 5.3), only one of the two antiphase halves is shown; (b) spectrum obtained by χ^2-minimisation, with the regions 0–13 Hz and 65–100 Hz constrained to be zero; (c) spectrum obtained by the MEM with resolution enhancement. A stick spectrum constructed using the refined values of the splittings (obtained by curve fitting) is drawn as (d).

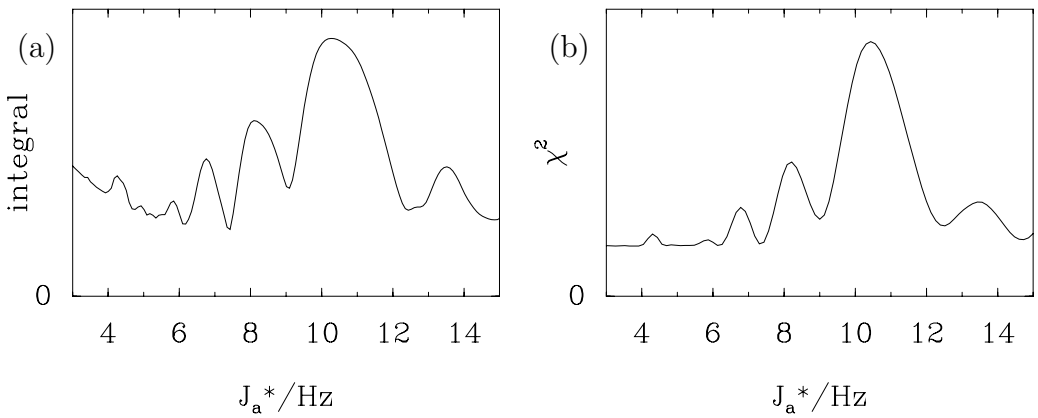

Figure 5.14: Estimation of J_a for the AM cross peak by J-doubling and constrained χ^2-minimisation: (a) the integral of the modulus of the J-doubled spectrum ($n = 5$) as a function of J_a^*; (b) the minimised value of χ^2 as a function of J_a^*.

9.4 Hz for the doublet (J_{MX}). These splittings were then refined by curve fitting to the original multiplet, giving values of $J_{MQ} = J_a = 13.76$ Hz, $J_{AM} = 7.42$ Hz and $J_{MX} = 9.40$ Hz. A stick spectrum constructed from these refined values (d) shows clearly the degree of overlap and cancellation due to the antiphase splitting.

5.3.2 The AM Cross Peak

The AM cross peak is a much more challenging test, with more severe overlap and cancellation and a badly distorted baseline (due to the wings of a nearby diagonal peak). The baseline distortion was partially corrected by setting the first data point to zero. This correction was vital to the success of J-doubling. Once again the value of the antiphase splitting, J_a, was estimated using J-doubling and constrained χ^2-minimisation; the results are shown in figure 5.14. As before the plots are rather similar, but while J-doubling gives an unambiguous value of $J_a = 7.3$ Hz, χ^2-minimisation gives very similar minima around 3.6, 5.1, 6.1 and 7.3 Hz; in both cases there are a host of minima at higher values of J_a^*. These false minima correspond to fractions of the true splitting and to sum and difference splittings.

Deconvolved spectra obtained using J_a=7.3 Hz are shown in figure 5.15. Spectra obtained by J-doubling (a), χ^2-minimisation (b) and the MEM (c) contain complex

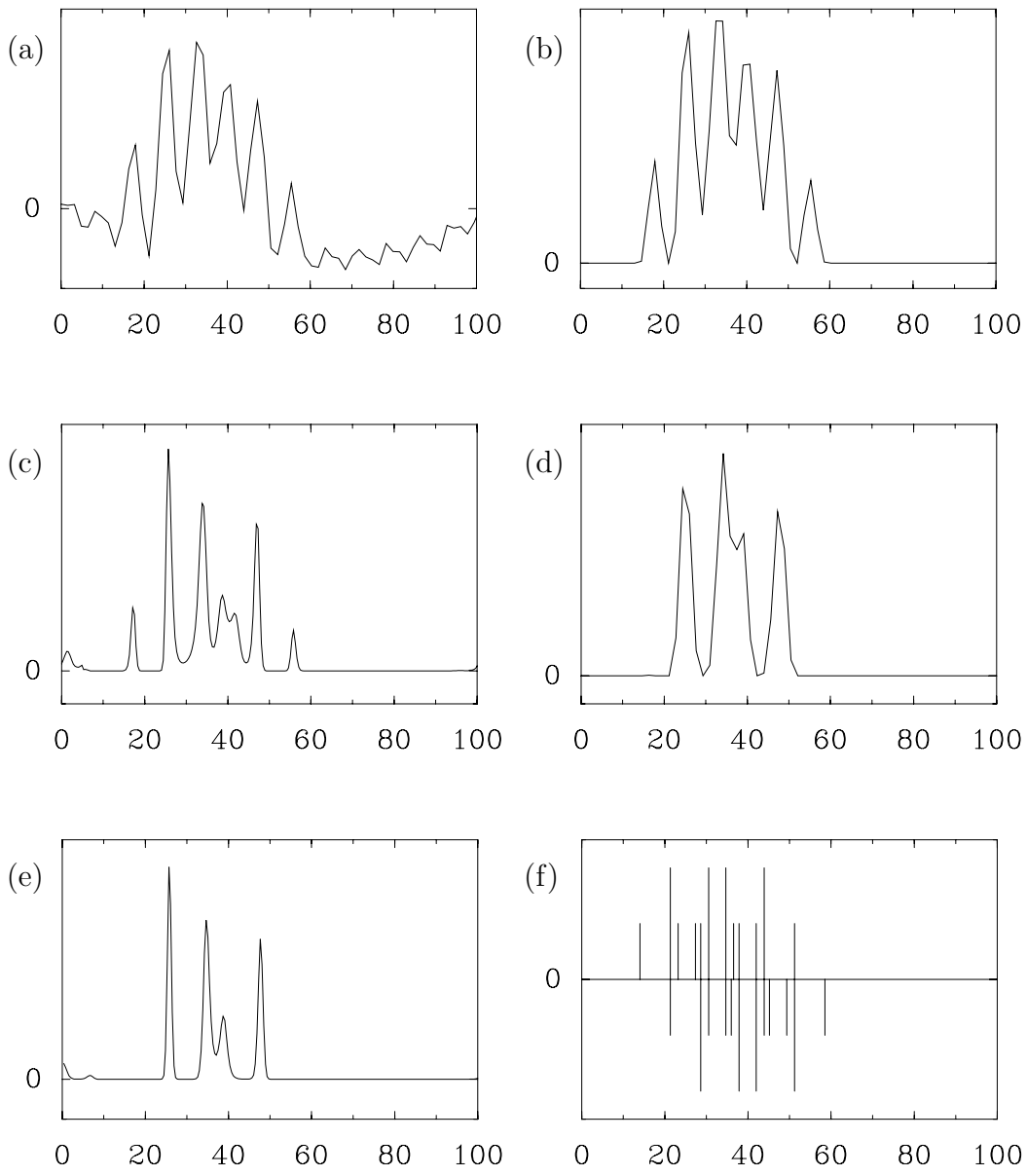

Figure 5.15: Spectra obtained from figure 5.11b after removal of an antiphase splitting with J_a=7.3 Hz: (a) spectrum obtained by J-doubling using $n = 5$ (see equation 5.3), only one of the two antiphase halves is shown; (b) spectrum obtained by χ^2-minimisation, with the regions 0–13 Hz and 65–100 Hz constrained to be zero; (c) spectrum obtained by the MEM with resolution enhancement. Spectra (d) and (e) were obtained in the same way as (b) and (c), except that an antiphase quartet was removed instead of an antiphase doublet. A stick spectrum constructed using the refined values of the splittings (obtained by curve fitting) is drawn as (f).

multiplets; in no case is the underlying structure readily apparent. The original multiplet (figure 5.11b) is too complex, and of too low quality, to yield a deconvolved spectrum which can be immediately recognised. It is necessary to use additional prior knowledge to solve the problem.

Fortunately such prior knowledge is available. The active coupling is to a methyl group, containing three equivalent protons; consequently there are two passive couplings with the same coupling constant as the active coupling. This results in a modulation of $i\sin(\pi J_a t)\cos^2(\pi J_a t)$ which can be removed by χ^2-minimisation or the MEM. The resulting spectra are shown in (d) and (e). Both spectra contain four lines, which clearly constitute a doublet of doublets, with approximate coupling constants of 9.0 Hz and 13.8 Hz.

Curve fitting was used to refine these estimates, resulting in values of $J_{AM} = 7.32\,\text{Hz}$, $J_{MQ} = 13.38\,\text{Hz}$ and $J_{MX} = 9.21\,\text{Hz}$. A stick spectrum constructed from these values is shown in (f); this indicates the enormous amount of overlap and cancellation in the original multiplet.

A Cramér–Rao analysis of the multiplets suggests that the precision of the coupling constants is about $\pm 0.1\,\text{Hz}$; this is consistent with the variation between the values determined from the MQ and AM cross peaks.

5.3.3 Determination of J_a by the MEM

As described above, Delsuc and Levy have suggested a method of estimating the value of J_a using the MEM[29]. They only applied the method to simple antiphase doublets, with which the method is fairly successful. With complex multiplets, however, the method is much less successful[49, 53].

The method was applied to the MQ and AM multiplets discussed above; the results are shown in figure 5.16. Instead of concentrating intensity in the subspectrum corresponding to $J_a^* = J_a$, as naïvely expected, there are peaks at values of J_a^* equal to sums and differences of J_a^* with the passive splittings. In fact any pair of antiphase lines in the original multiplet gives rise to a peak at the appropriate value of J_a^*.

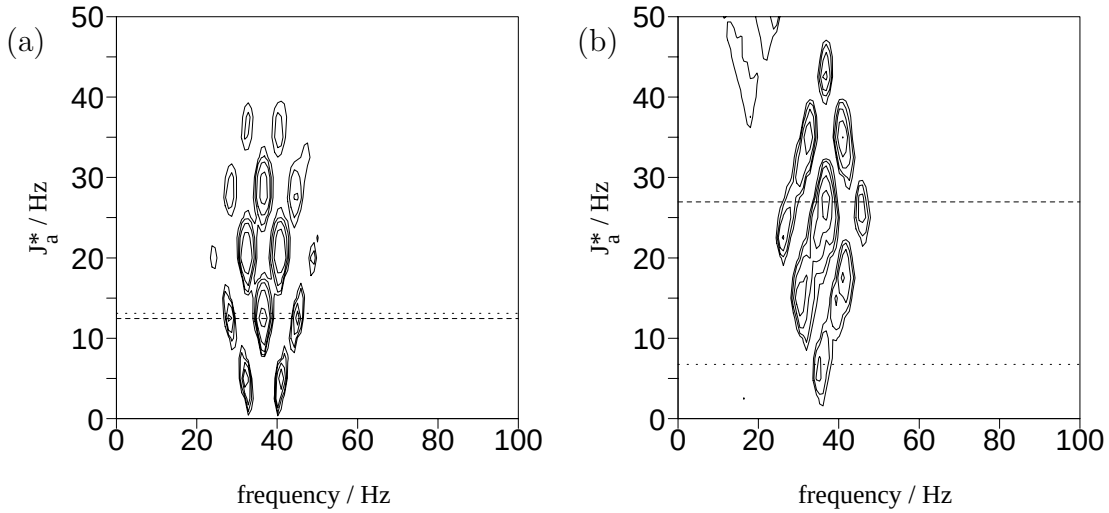

Figure 5.16: Estimation of J_a by the MEM using the method of Delsuc and Levy: (a) the MQ multiplet (see figure 5.11a), (b) the AM multiplet (see figure 5.11b). The dotted lines indicate the value of J_a determined by J-doubling, while the dashed lines indicate the value of J_a^* corresponding to the maximum intensity in each spectrum.

In the case of the MQ multiplet (a), the maximum intensity (indicated by a dashed line) does lie close to the value of J_a, suggesting that this might provide an appropriate criterion. However the spectrum from the AM multiplet (b) shows that the maximum intensity can occur at a quite different splitting.

Thus the Delsuc–Levy technique has little or nothing to recommend it as a method of determining J_a; the results are unambiguous only when the multiplet is so simple that the value can be estimated directly. The reason for this failure is easy to see. The maximum entropy criterion encourages the algorithm to spread intensity as evenly as possible consistent with matching the experimental data and ensuring positivity; consequently the method will prefer to match the data by placing intensity in sub-spectra corresponding to several different splittings, rather than concentrating it in a single sub-spectrum.

5.3.4 Truncated Data

J-deconvolution of severely truncated data is potentially extremely useful since it allows the use of the MEM to suppress truncation artefacts. It is also a very challenging problem, since truncation artefacts result in extensive overlap and can-

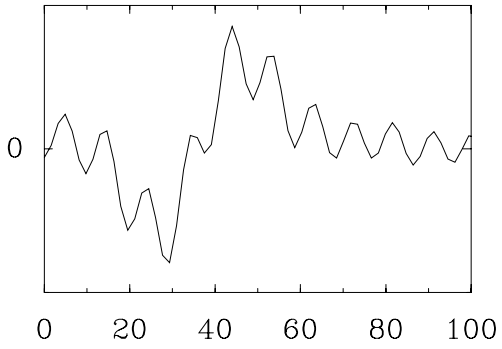

Figure 5.17: The MQ cross peak from the COSY spectrum of iso-leucine with severe truncation effects. The sinc wiggles completely obscure the underlying structure.

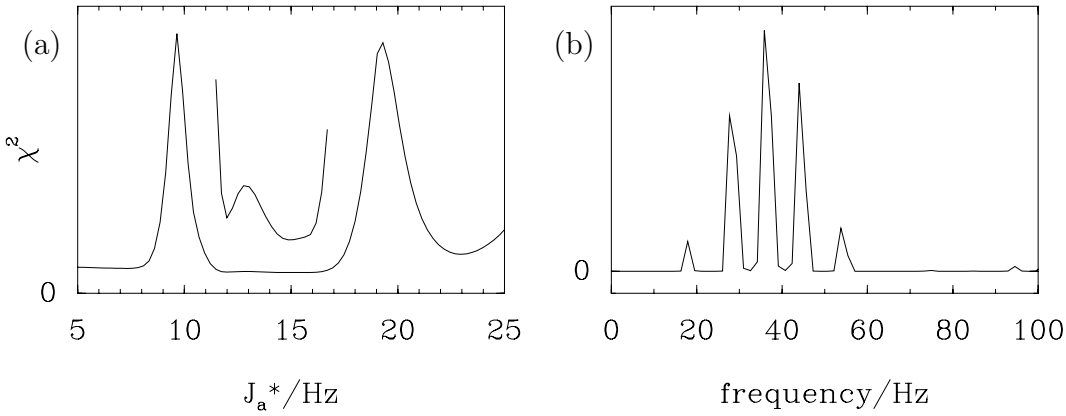

Figure 5.18: J-deconvolution of the severely truncated data depicted in figure 5.17: (a) estimation of J_a using χ^2-minimisation, part of the plot is drawn with an expanded vertical scale to emphasise the position of the minimum; (b) J-deconvolution by the MEM with resolution enhancement using $J_a = 14.8\,\mathrm{Hz}$.

cellation in the original multiplet, rendering some methods ineffective. However both the MEM and χ^2-minimisation should be capable of coping with these artefacts.

Figure 5.17 shows the multiplet MQ with severe truncation: the sinc wiggles result in the underlying structure being almost completely obscured (compare this with figure 5.11a which shows the multiplet in the absence of truncation effects). Attempts to estimate J_a by J-doubling were unsuccessful, as a result of the sinc wiggles. χ^2-minimisation (figure 5.18a) gives a broad, but clear, minimum around 14.8 Hz. This value was used to deconvolve the splitting using the MEM

(figure 5.18b); the resulting spectrum contains a quintet. Attempts at further resolution enhancement, in order to distinguish between the similar passive splittings, were unsuccessful.

This shows that J-deconvolution *can* be performed, even in the presence of truncation effects. This is important since COSY spectra are frequently severely truncated. In such cases, however, it may be difficult to estimate J_a accurately, and the passive structure may be only partially recovered.

5.4 Discussion

The examples above demonstrate that J-deconvolution provides an effective way of simplifying complex multiplets from COSY spectra. When the data is severely truncated, simultaneous J-deconvolution and suppression of truncation artefacts by the MEM, can allow coupling information to be extracted from apparently hopeless multiplets.

The various J-deconvolution techniques discussed above vary substantially in their power and ease of use. In general it is wise to tackle a problem using several methods simultaneously. Nevertheless some general recommendations are possible.

1. J-doubling and stabilised J-deconvolution give very similar results, but J-doubling is much simpler to implement, and is consequently preferable.

2. J-doubling (and stabilised J-deconvolution) are very sensitive to baseline distortions, which are common in COSY spectra. It is vital that the baseline be corrected if these methods are to be used successfully.

3. With untruncated data J-doubling is probably the best way of estimating J_a, but with severely truncated data constrained χ^2-minimisation is much more successful.

4. The MEM is generally the best method for performing J-deconvolution once the value of J_a is known, especially when combined with resolution enhancement or the suppression of truncation artefacts. Furthermore the MEM is

relatively insensitive to baseline distortions.

5. In complex cases where the multiplet contains one or more passive splittings identical to the active splitting the inclusion of this prior knowledge greatly simplifies the analysis process.

6. Once estimates of the splittings have been obtained, they can be readily refined by curve fitting to the original multiplet.

Appendix A

Historical notes

> The reader, at this point, will have realized for some time now that this is not a chemical treatise: my presumption does not reach so far—"*ma voix est foible, et même un peu profane*." Nor is it an autobiography, save in the partial and symbolic limits in which every piece of writing is autobiographical, indeed every human work; but it is in some fashion a history.
>
> "Carbon", from "The Periodic Table" by Primo Levi (translated by R. Rosenthal).

Bayes, Thomas (1702?–1761). The date of Thomas Bayes' birth is not completely certain, but it is generally fixed in 1702. He was born in London, the eldest son of Joshua and Anne Bayes. His father was one of the first Nonconformist ministers to be publicly ordained in England and a Fellow of the Royal Society. He was educated privately, and it has been suggested that he learned mathematics from De Moivre, although there is no direct evidence of this. He was eventually ordained, and began his ministry working with his father at the Presbyterian meeting house in Holborn; later he took up the pulpit of the Presbyterian chapel in Tunbridge Wells. In 1731 he published a tract entitled *"Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government Is the Happiness of His Creatures"*, in opposition to the works of Balguy and Grove who, respectively, maintained that the principle action of the Deity was Rectitude and Wisdom. This was followed in 1736 by a tract entitled *"An Introduction to*

the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of the Analyst"; although published anonymously this work is universally ascribed to Bayes. In it he defended Newton's theory of fluxions (calculus) against the metaphysical objections of the Analyst (Bishop Berkeley). This tract may have been the cause of Bayes' election in 1742 to Fellowship of the Royal Society.

Bayes partially retired from his ministry in 1749, and was succeeded in 1752 by the Rev. William Johnson, but he continued to live in Tunbridge Wells until his death on 17 April 1761. His will, executed on 12 December 1760, left the sum of £100 to "Richard Price now I suppose preacher at Newington Green"; perhaps surprisingly he also left many of his papers to this man, whose whereabouts he was unsure of. Fortunately the Rev. Richard Price, a Unitarian minister indeed working in Newington Green, realised the importance of these papers, and on 10 November 1763 sent an edited version of Bayes' *"An Essay towards solving a Problem in the Doctrine of Chances"* to John Canton, FRS, thus ensuring its publication[68, 69, 105–109].

Bayes' Theorem. Bayes' theorem can be found in essentially its final form in his famous *"Essay towards solving a Problem in the Doctrine of Chances"*, but neither he nor his contemporaries realised the importance of it. Bayes himself thought the theorem worthy of proof and discussion, but only as a stepping stone on the way to solving a specific problem. Bayes' essay was almost completely ignored for several decades, until Laplace realised the great importance of the theorem and gave it its present name.

The theorem may be found in two propositions of the *Essay*: proposition 3 and proposition 5. These are

Proposition 3 The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st and the probability of the 2d on supposition the 1st happens.

In modern terminology

$$\mathcal{P}(AB) = \mathcal{P}(A)\,\mathcal{P}(B|A).$$

Proposition 5 If there be two subsequent events, the probability of the 2d b/N and the probability of both together P/N, and it being 1st discovered that the 2d event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is P/b. In modern terminology

$$\mathcal{P}(A|B) = \mathcal{P}(AB)\,/\,\mathcal{P}(B).$$

The difference between propositions 3 and 5 is subtle but significant; for further discussion of this point see appendix C. Combining the two propositions gives

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(B|A)\,\mathcal{P}(A)}{\mathcal{P}(B)}$$

which is Bayes' theorem in its modern form[68, 69, 107, 110].

Bloch, Felix (1905–1982). Born in Zürich, Switzerland, Felix Bloch took up an academic post at Stanford University in California. His pioneering work on NMR undertaken there led to the award of the Nobel prize, joint with Purcell, in 1952[111].

Fourier, (Jean-Baptiste-)Joseph, Baron (1768–1830). Born in Auxerre in 1768, Fourier was the son of a tailor. He was educated at a school run by Benedictine monks, and was so skilled at mathematics that he soon became a teacher of mathematics at the same school. He joined the École Normale in 1794 and moved to the École Polytechnique in 1795. In 1798 he went to Egypt, where he undertook research into antiquities, and advised on engineering and diplomatic problems. For three years he was secretary to the Institut d'Égypte in Cairo. On his return to France in 1801 he was placed in charge of the publication of the great 21 volume *"Description de l'Égypte"*. In 1802 he was appointed prefect of the département of Isère, and took up

residence in Grenoble. In 1809 he was made a baron. After Napoleon's fall from power (1815) Fourier left politics and returned to academic life, being appointed director of the Statistical Bureau of the Seine. In 1817 he was elected to the Académie des Sciences, and in 1822 took up the post of permanent secretary. In 1826 his research into Egyptology was rewarded by election to the Académie Française.

The work for which Fourier is now chiefly remembered came late in his life. While working as prefect in Grenoble he began work on his *"Théorie analytique de la chaleur"* (The Analytical Theory of Heat), which was finally published in 1822, in which he introduced the use of Fourier series for the solution of problems involving the conduction of heat through solid bodies. His series, treated with some scepticism by mathematicians of the day, have since been applied to a huge variety of problems. He died on 16 May 1830 in Paris[105].

Insufficient Reason, Principle of. The principle of Insufficient Reason, also referred to as the Principle of Indifference, was first articulated by Bernoulli in his *"Ars Conjectandi"* but lies at the heart of all early work on probability. The principle simply states that if reason is insufficient to determine the relative probabilities of two events, they must be assigned the same probability. When applied to discrete events the principle is fairly simple. For example if a coin is tossed there are two possible outcomes: heads and tails. If nothing else is known about the coin or the toss, these outcomes must be assigned equal probabilities. When applied to experiments whose outcome is a continuous variable, however, the problem is much more contentious. Naïvely it might seem that if the result of the experiment is x, assigning constant probability density $\mathcal{P}(x)$ is appropriate; however it is also possible to label the result of the experiment by some arbitrary function $f(x)$, and the same arguments lead to the assertion that $\mathcal{P}(f(x))$ is constant. One well known example of this is Bertrand's paradox, in which three different ways of assigning the probability density, which are apparently equally reasonable, lead

to three contradictory results.

Bayes has been frequently accused of using the principle in his *Essay*, but this is quite false. In fact Richard Price's introduction to the essay (see appendix B) makes clear that Bayes was aware of the controversy and saw a solution to it. This is the purpose of the *scholium* section of the *Essay*, in which he cleverly redefines the meaning of "complete ignorance". Bayes' paper deals with the problem of determining a number whose value is continuous by means of experiments whose outcomes are discrete; consequently he can define ignorance in terms of ignorance of the results of the discrete experiments, rather than ignorance of the value itself.

A second trick, due to Jaynes, uses the theory of transformation groups to avoid the problem. This remarkable method uses ignorance of certain details of the problem as the basis of determining the solution. If some parameter is left unstated in a problem and a solution exists, then the solution *must* be independent of the value of that parameter; this independence results in restrictions on the possible form of the solution. Using this technique Jaynes has succeeded in solving Bertrand's paradox.

Some authors are worried by the philosophical implications of the principle: in 1850 the mathematician and philosopher Leslie Ellis wrote "... mere ignorance is no ground for any inference whatever. *Ex nihilo nihil.* It cannot be that because we are ignorant of the matter we know something about it." In the case of transformation groups, the solution to this is quite simple: inferences about the solution are based not on ignorance of some parameter, but on the assertion that a general solution, independent of this parameter, actually exists. As is often the case in probability theory, the exact meaning of phrases like "complete ignorance" is extremely important[107, 112–116].

Ockham, William of (1285?–1349?). William was born in Ockham, Surrey, and is chiefly known by this name today (the variation "Occam" is a later latinization). He was educated in a Franciscan convent, where he specialised

in the subject of logic. Later he went to Oxford, where from 1317–1319 he lectured on the *"Sentences"* of Peter Lombard, which was the basic theology text of the day. His lectures on book 1 of the *Sentences* were written down to form Ockham's *"Ordinatio"*. His opinions aroused such strong opposition that he was forced to leave Oxford without obtaining his master's degree, and so he remained an undergraduate, or "inceptor" for the rest of his life; this adds a peculiar charm to the title "Venerabilis Inceptor" (the venerable seeker) by which he is sometimes known. Ockham's expulsion was no rare event: a few years later John Luttrell, the Chancellor of Oxford, was deposed at the request of the same theologians. Ockham continued his theological career, living in a variety of English convents, until 1324, when he was called to the Avignon court by Pope John XXII (Jacques of Cahors). Soon after his arrival he met John Luttrell, who denounced Ockham to the Pope, submitting a list of 56 serious errors in the *Ordinatio*. Ockham accepted some of Luttrell's points, and re-issued the *Ordinatio* with some modifications, but he remained adamant on several points. An attempt was made to have him formally condemned, but for reasons that are not entirely clear this did not occur.

While at Avignon Ockham became embroiled in the dispute between the Pope and the Franciscan order about the poverty of the mendicant orders. In 1328 he publicly endorsed the position of Michael of Cesena and Bonagratia of Bergamo, and was forced to flee with them to the court of Emperor Louis IV the Bavarian (Louis had recently been excommunicated and deposed by John, and was sympathetic to the Franciscan cause). From there he wrote fervently against the papacy and in defence of the imperial claims. When John issued three papal bulls on the subject of poverty, Ockham condemned him as a heretic and declared him deposed; this view was confirmed by John's developing views on Purgatory.

Excommunicated by John in 1328, Ockham maintained his position even after John's death in 1334. Only the death of Emperor Louis, with the

consequent loss of imperial protection, forced Ockham to discuss submission to Clement VI. He died in Munich, probably of the Black Death, in 1349[105, 117, 118].

Ockham's Razor. Ockham's razor is in fact nothing more than the old scholastic "Law of Economy". This was first invoked by Durand de Saint-Pourçain (French bishop and theologian, 1270–1334, most noted for his opposition to the work of Thomas Aquinas), but was so frequently used by Ockham that it became attached to his name. Ockham used the principle to dispose of many philosophical ideas, claiming that they were unnecessary abstractions: for example he claimed that "motion" has no separate existence in itself, but is simply the disappearance of an object from one place and its reappearance in another. Ockham is generally considered the founder of the school of thought called Nominalism, which denies that universal concepts have any reality apart from the individual things signified by the universal term.

The phrase "Essentia non sunt multiplicanda praeter necessitatem", frequently ascribed to Ockham, is in fact due to John Ponce of Cork, a 17th century follower of John Duns Scotus and Ockham. The term "Ockham's razor" first appears in the 19th century, in the writings of Sir William Hamilton[105, 119, 120].

Purcell, Edward Mills (1912–). Born in Taylorville, Illinois, Edward Purcell was educated at Purdue and Harvard Universities. His pioneering work on NMR undertaken at Harvard led to the award of the Nobel prize, joint with Bloch, in 1952. He is currently Emmeritus Professor of Physics at Harvard[111, 121].

Shannon, Claude Elwood (1916–). Born in Gaylord, Michigan, Claude Shannon was educated at the University of Michigan and MIT. In his master's thesis he described the application of Boolean algebra to the description of electronic switches and relays, thus providing a firm theoretical description of computational devices. After obtaining his Ph. D. in 1941 he worked as a re-

search mathematician at Bell Laboratories; initially working on cryptography he soon turned to the study of the transmission of information down telegraph lines. This led, in 1948, to his classic paper on information theory. His introduction of the concept of the "entropy" of a message had consequences as profound as they were controversial (Shannon introduced the term on the advice of von Neumann, who claimed out that since no one really knew what the term meant, Shannon would have an advantage in any debates about his theory).

Shannon left Bell Laboratories in 1956 to take up a professorship at MIT; he remained there until his formal retirement in 1978. A childhood obsession with mechanical toys has remained with him throughout his life, and he is currently devoted to the mathematical analysis of juggling, and the construction of juggling machines. Despite constructing a unified theory of juggling (if B denotes the number of balls, H the number of hands, D the time each ball spends in a hand, F the flight time of each ball and E the time each hand is empty, then $B/H = [D+F]/[D+E]$) he has yet to improve on his personal record of four balls. He is the proud owner of a two seater unicycle[121, 122].

Appendix B

Towards Solving a Problem in the Doctrine of Chances

> One of the principal Uses to which this *Doctrine of Chances* may be apply'd, is the discovering of some Truths, which cannot fail of pleasing the Mind, by their Generality and Simplicity; the admirable Connexion of its Consequences will increase the Pleasure of the Discovery; and the seeming Paradoxes wherewith it abounds, will afford very great matter of Surprize and Entertainment to the Inquisitive.
>
> From the preface to "The Doctrine of Chances" by A. De Moivre

On the death of Thomas Bayes a large quantity of his mathematical papers were left to Richard Price. Among these was an essay addressing a new problem in probability theory (at that time referred to as "the doctrine of chances"). On the tenth of November 1763 Richard Price wrote to John Canton, FRS, enclosing this essay together with an appendix (written by himself) and a long covering letter. The essay was read at the Royal Society on the twenty third of December and printed in volume LIII of the *Philosophical Transactions* under the title "An Essay towards solving a Problem in the Doctrine of Chances".

Richard Price's covering letter provides an excellent introduction to the issues discussed in this paper, which lie at the heart of Bayesian reasoning, and as such it is printed below, in full and without further comment.

Dear Sir,

I now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illustrious Society, and was much esteemed by many in it as a very able mathematician. In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times. He adds, that he soon perceived that it would not be very difficult to do this, provided some rule could be found according to which we ought to estimate the chance that the probability for the happening of an event perfectly unknown, should lie between any two named degrees of probability, antecedently to any experiments made about it; and that it appeared to him that the rule must be to suppose the chance the same that it should lie between any two equidifferent degrees; which, if it were allowed, all the rest might be easily calculated in the common method of proceeding in the doctrine of chances. Accordingly, I find among his papers a very ingenious solution of this problem in this way. But he afterwards considered, that the *postulate* on which he had argued might not perhaps be looked upon by all as reasonable; and therefore he chose to lay down in another form the proposition in which he thought the solution of the problem is contained, and in a *scholium* to subjoin the reasons why he thought so, rather than to take into his mathematical reasoning any thing that might admit dispute. This, you will observe, is the method which he has pursued in this essay.

Every judicious person will be sensible that the problem now mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to

be solved in order to a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter. Common sense is indeed sufficient to shew us that, from the observation of what has in former instances been the consequence of a certain cause or action, one may make a judgement what is likely to be the consequence of it another time, and that the larger the number of experiments we have to support a conclusion, so much the more reason we have to take it for granted. But it is certain that we cannot determine, at least not to any nicety, in what degree repeated experiments confirm a conclusion, without the particular discussion of the beforementioned problem; which, therefore, is necessary to be considered by any one who would give a clear account of the strength of *analogical* or *inductive reasoning*; concerning, which at present, we seem to know little more than that it does sometimes in fact convince us, and at other times not; and that, as it is the means of cquainting us with many truths, of which otherwise we must have been ignorant; so it is, in all probability, the source of many errors, which perhaps might in some measure be avoided, if the force that this sort of reasoning ought to have with us were more distinctly and clearly understood.

These observations prove that the problem enquired after in this essay is no less important than it is curious. It may be safely added, I fancy, that it is also a problem that has never before been solved. Mr. De Moivre, indeed, the great improver of this part of mathematics, has in his *Laws of chance**, after Bernoulli, and to a greater degree of exactness, given rules to find the probability there is, that if a very great number of trials be made concerning any event, the proportion of the number of times it will happen, to the number of times it will fail in those trials, should differ less than by small assigned limits from the proportion of the probability of its happening to the probability of its failing in one single trial. But I know of no person who has shewn how to deduce the solution of the converse problem to this; namely, "the number of times an unknown event has happened and failed being given, to find the chance that the probability of its happening should lie

*See Mr. De Moivre's *Doctrine of Chances,* p. 243, &c. He has omitted the demonstrations of his rules, but these have been since supplied by Mr. Simpson at the conclusion of his treatise on *The Nature and Laws of Chance.*

somewhere between any two named degrees of probability." What Mr. De Moivre has done therefore cannot be thought sufficient to make the consideration of this point unnecessary: especially, as the rules he has given are not pretended to be rigorously exact, except on supposition that the number of trials made are infinite; from whence it is not obvious how large the number of trials must be in order to make them exact enough to be depended upon in practice.

Mr. De Moivre calls the problem he has thus solved, the hardest that can be proposed on the subject of chance. His solution he has applied to a very important purpose, and thereby shewn that those are much mistaken who have insinuated that the Doctrine of Chances in mathematics is of trivial consequence, and cannot have any place in any serious enquiry*. The purpose I mean is, to shew what reason we have for believing that there are in the constitution of things fixt laws according to which events happen, and that, therefore, the frame of the world must be the effect of the wisdom and power of an intelligent cause; and thus to confirm the argument taken from final causes for the existence of the Deity. It will be easy to see that the converse problem solved in this essay is more directly applicable to this purpose; for it shews us, with distinctness and precision, in every case of any particular order or recurrency of events, what reason there is to think that such recurrency or order is derived from stable causes or regulations in nature, and not from any of the irregularities of chance.

The two last rules in this essay are given without the deductions of them. I have chosen to do this because these deductions, taking up a good deal of room, would swell the essay too much; and also because these rules, though of considerable use, do not answer the purpose for which they are given as perfectly as could be wished. They are however ready to be produced, if a communication of them should be thought proper. I have in some places writ short notes, and to the whole I have added an application of the rules in the essay to some particular cases, in order to convey a clearer idea of the nature of the problem, and to shew how far the solution of it has been carried.

*See his Doctrine of Chances, p. 252, &c.

I am sensible that your time is so much taken up that I cannot reasonably expect that you should minutely examine every part of what I now send you. Some of the calculations, particularly in the Appendix, no one can make without a good deal of labour. I have taken so much care about them, that I believe there can be no material error in any of them; but should there be any such errors, I am the only person who ought to be considered as answerable for them.

Mr. Bayes has thought fit to begin his work with a brief demonstration of the general laws of chance. His reason for doing this, as he says in his introduction, was not merely that his reader might not have the trouble of searching elsewhere for the principles on which he has argued, but because he did not know whither to refer him for a clear demonstration of them. He has also made an apology for the peculiar definition he has given of the word *chance* or *probability.* His design herein was to cut off all dispute about the meaning of the word, which in common language is used in different senses by persons of different opinions, and according as it is applied to *past* or *future* facts. But whatever different senses it may have, all (he observes) will allow that an expectation depending on the truth of any *past* fact, or the happening of any *future* event, ought to be estimated so much the more valuable as the fact is more likely to be true, or the event more likely to happen. Instead therefore, of the proper sense of the word *probability,* he has given that which all will allow to be its proper measure in every case where the word is used. But it is time to conclude this letter. Experimental philosophy is indebted to you for several discoveries and improvements; and, therefore, I cannot help thinking that there is a peculiar propriety in directing to you the following essay and appendix. That your enquiries may be rewarded with many further successes, and that you may enjoy every valuable blessing, is the sincere wish of,
Sir,

your very humble servant,

Richard Price.

Appendix C

Probability

> The brilliant Cerebron, attacking the problem analytically, discovered three distint kinds of dragon: the mythical, the chimerical, and the purely hypothetical. They were all, one might say, nonexistent, but each nonexisted in an entirely different way.
>
> "The Dragons of Probability", from "The Cyberiad" by Stanislaw Lem (1975).

The theory of probability is one of the most contentious areas in mathematics. Just as Cerebron distinguished three varieties of dragon, it is possible to distinguish three quite separate meanings of the word probability[110, 123].

The first of these, due to Kolmogoroff, refers to the fundamental theory of probability as a branch of pure mathematics: this is concerned with a space Ω with a measure $\mathcal{P}$ such that $\mathcal{P}(\Omega) = 1$. A random variable X is a function defined on Ω, with mean value given by the integral of X over Ω. This is not controversial; what *is* controversial is the application of this theory. There are two main schools of thought as to how this should be done.

The older, and still the more popular, approach is that probabilities ultimately reflect fractional frequencies. The origin of this view lies in the analysis of gambling due to De Moivre and others[109, 112, 124–127]. To them the statement "the probability of throwing a six with a die is one sixth" meant that if the die were thrown a very large number of times, the proportion of sixes among the throws would approach one sixth. This is sometimes described as an "objective" view of probability, since the probability of an event is believed to be an objective,

ultimately measurable, property of the system being described.

The alternative approach to probability also has its origin in gambling: according to this view probabilities represent the odds at which a rational gambler would bet. This view is first found in the writings of Bayes[68][1], and is forcefully argued by Jaynes[115]. Suppose a rational gambler is playing a game where he must pay a stake s before throwing a die, and he collects winnings w if he throws a six; in this case the statement "the probability of throwing a six with a die is one sixth" means that the gambler will only play the game if $w \geq 6s$. This is sometimes described as a "subjective" view of probability, since the probability of an event is not a property of the system, but a property of the gambler's knowledge about the system[2].

The first view described above seems so natural that it has been adopted by many people without question. This state of affairs has been greatly strengthened by the polemical use of the words "objective" and "subjective". However, when speaking loosely, most people mean by probability something much closer to the second definition. In order to justify this assertion I will consider three problems, and indicate how the two schools describe them. In order to avoid polemicism, the two schools will be referred to as "frequentist" and "Bayesian" respectively.

All three problems will be based on the problem of throwing a die discussed above. For simple questions about this system, the two schools give very similar answers, but for more complex questions the answers can be quite different.

The first problem is a classic example in probability theory:

I am about to throw a die. What is the probability that I throw a six?

[1]Huygens' treatise on probability contains the statement[125]

> Ones Hazard or Expectation to gain anything is worth so much, as, if he had it, he could purchase the like Hazard or Expectation again in a just and equal game.

which might appear to be an earlier statement of this view; however Huygens only uses this statement to justify his method of calculating expectations, and it is unlikely that he had any more radical ideas in mind.

[2]Earlier authors did make statements similar to this, but they only considered the reverse case. For example Huygens[124, 125] calculated the probability of throwing a six with a single throw of a die as one sixth, and so advised his readers not to bet that they would throw a six in these circumstances unless the winnings were at least six times the stake. Bayes' radical departure was his reversal of the conventional statement, allowing him to *define* probabilities in terms of rational gambling.

Note that I am assuming here that the die has never been thrown before; the more complex case where the results of previous throws are known will be dealt with later.

A frequentist would like to estimate the probability of throwing a six by the fractional frequency in a set of trials. Since he has no such data available to him, he must construct an imaginary ensemble of throws and estimate the probability by the fractional frequency in this ensemble. Clearly the ensemble must be carefully constructed so that the throws differ subtly: if all the throws were identical, the same result would occur in each case, and the fractional frequency of throwing a six would be either 0 or 1. Fortunately it is not necessary to formally construct the ensemble, and no frequentist would actually do this. At some stage he will invoke the principle of insufficient reason (see appendix A) to deduce that no particular outcome is favoured, and so assign a fractional frequency (and hence a probability) of one sixth.

The Bayesian's approach is much simpler, since his probabilities are not a property of the die, but of his knowledge about it. He knows that it has six sides, and he has no reason to believe that any particular side is favoured. Any assignment of probabilities other than an equal probability to each side is irrational, and so the Bayesian immediately assigns a probability of one sixth.

In this case both schools have come to the same answer, using methods that are ultimately quite similar (the frequentist's application of the principle of insufficient reason and the Bayesian's assumption of equiprobability). The main difference is that the frequentist must invoke an imaginary ensemble; the significance of this will soon be made clear.

The second problem is only slightly different:

> I have just thrown a die, but have not looked at the result. What is the probability that I have just thrown a six?

This differs from the first problem only in the time at which the question is asked; this apparently trivial change has profound consequences for the frequentist.

As far as the frequentist is concerned the probability of a past event is either

1 (if it happened) or 0 (if it didn't). The answer is determined when the trial is performed, that is when the die is thrown. It is, of course, possible to form an ensemble of throws, but they must all be identical to one another and to the throw that actually occurred. Strict frequentism rules out all discussion of the probability of an event in the past, since it is impossible to form an appropriate ensemble.

The Bayesian's approach is totally different: as far as he is concerned this problem is completely identical to the first. Probabilities are a property of his state of knowledge about the system, and throwing the die (without looking at it) has not changed his knowledge at all. Hence the probability is one sixth, exactly as before.

This inability to cope with past events is a powerful argument against strict frequentism, and this is one of the cases which causes many frequentists to shy away. The second question makes no less sense than the first, so why should the treatment of it be so different? Furthermore the separation of events into "past" and "future" is difficult in a world which is at least partially deterministic—exactly when does the result of the throw become determined?

The Bayesian understanding is much more natural. What matters is not whether or not the trial has occurred, but whether or not he knows the result of it. This neatly sidesteps issues of determinism, and corresponds much more closely to the colloquial use of the word.

Bayes' own approach to this problem is indicated by propositions 3 and 5 of his essay. These are

> Proposition 3: The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st and the probability of the 2d on supposition the 1st happens.

and

> Proposition 5: If there be two subsequent events, the probability of the 2d b/N and the probability of both together P/N, and it being 1st discovered that the 2d event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is P/b.

These propositions indicate that Bayes himself was happy to talk about the probability of a past event, but that he was aware that his readers might not be.

Expressed in modern notation these propositions are

$$\mathcal{P}(AB) = \mathcal{P}(A)\,\mathcal{P}(B|A),$$

and

$$\mathcal{P}(A|B) = \mathcal{P}(AB)\,/\,\mathcal{P}(B),$$

where A is the first event and B is the second. Clearly the second proposition can be derived from the first by simply exchanging A and B, and a modern Bayesian would be happy to do this. In order to avoid controversy Bayes chose to prove the second proposition directly, by showing that any other possibility led to a contradiction. He also shied away from talking directly about the probability of an event in the past, preferring to talk about the probability that he is right if he assumes that the event has occurred.

Bayes' struggle to find a language suitable for discussing the probability of both past and future events led him to a remarkable redefinition of the word:

> The probability of an event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon it's happening.

This is exactly the definition used earlier, in which the concept of a "rational gambler" was introduced. Bayes was aware that this was an unusual definition (see appendix B), but he also knew that it was the only definition that would serve his purpose.

My third die problem is quite different from the other two:

> Suupose that N throws of my die have resulted in P sixes. What is the probability that the limiting fractional frequency of sixes for this die, θ, lies between two values α and β?

This is a further radical shift in the meaning of the word "probability". The previous problems have referred to the probability of an event occurring, while this one refers to the probability of a statement ($\alpha < \theta < \beta$) being true. To a strict frequentist this is utter nonsense: the statement is either true or it isn't. He

cannot imagine an ensemble of worlds, in some of which the statement is true and in some of which it is false.

The Bayesian, by contrast, has no difficulty, since the problem is essentially identical to Bayes' problem[68]

> *Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

Bayes' solution to the die problem is

$$\mathcal{P}(\alpha < \theta < \beta \,|\, P \text{ sixes}) = \frac{\int_\alpha^\beta \binom{N}{P} \theta^P (1-\theta)^{N-P}\, d\theta}{\int_0^1 \binom{N}{P} \theta^P (1-\theta)^{N-P}\, d\theta}$$

which can be evaluated numerically[3].

This last problem may seem peculiar, but in fact it is a typical problem in inverse probability theory, and as such is essentially identical to the sort of problem which most scientists tackle most of the time. Given a series of measurements, a scientist will attempt to estimate one or more parameters, and to say something about the reliability of his estimates. Typicaly his results will be stated in the language of probabilities: "There is a 95% chance that x lies between y and z". To a strict frequentist this is inadmissible, but this does not prevent many professing frequentists from making this kind of statement.

In summary, while the frequentist view of probability is initially attractive, it has many hidden problems. In particular it cannot be used to discuss the probability of events in the past, or the probability of statements being true. When interpreted strictly, the view is so restrictive that it cannot be applied to many interesting problems. By contrast the Bayesian viewpoint can be easily applied.

It should be emphasised that when dealing with problems in simple probability theory (such as the first die problem), frequentist and Bayesian calculations will give the same results. With problems from inverse probability theory, however, the situation is very different, since the frequentist viewpoint cannot be used in any

[3]The calculation of this probability involves evaluation of the incomplete beta function. This problem had not been solved at the time, and the second half of Bayes' essay consists of (largely unsuccesful) attempts to approximate the integral.

satisfactory manner. While a number of non-Bayesian methods for tackling problems in inverse probability do exist, none of these conform strictly to frequentist philosophical principles.

Appendix D

Artefacts in PMRFI spectra

> Their images are but wind and confusion.
>
> Isaiah 41 : 29b

The form of a conventionally processed PMRFI spectrum can be calculated by Fourier transforming equation 4.31. To simplify the calculation it is convenient to rewrite equation 4.32 as

$$F_{\pm} = C_1^{\pm}\cos\theta' + C_2^{\pm}\sin\theta' + C_3^{\pm} + iC_4^{\pm}\cos\theta' + iC_5^{\pm}\sin\theta' + iC_6^{\pm}, \qquad \text{(D.1)}$$

with coefficients as follows:

$$\begin{aligned}
C_1^{\pm} &= \mp\cos\lambda'\sin\alpha\cos\alpha + \cos^3\alpha\sin\lambda', \\
C_2^{\pm} &= \pm\sin\alpha\cos\alpha\sin\lambda', \\
C_3^{\pm} &= \pm\cos\lambda'\sin\alpha\cos\alpha + \cos\alpha\sin\lambda'\sin^2\alpha, \\
C_4^{\pm} &= \mp\sin^2\alpha\cos\alpha\sin\lambda' + \sin\alpha\cos^3\alpha - \sin\alpha\cos^3\alpha\cos\lambda', \\
C_5^{\pm} &= \mp\cos^3\alpha \mp \sin^2\alpha\cos\lambda'\cos\alpha, \\
C_6^{\pm} &= \pm\sin^2\alpha\cos\alpha\sin\lambda' + \sin^3\alpha\cos\alpha - \sin^3\alpha\cos\alpha\cos\lambda'. \qquad \text{(D.2)}
\end{aligned}$$

These coefficients determine the relative intensities of six different groups of lines; the nature of each group is described below. Using the notation $s \Longleftrightarrow S$ to indicate that s and S form a Fourier transform pair, the calculations of chapter 2 give

$$e^{i\omega_1 t_1}e^{i\omega_2 t_2} \iff [A_1^+A_2 - D_1^+D_2] + i[A_1^+D_2 + D_1^+A_2]$$

$$
\begin{aligned}
e^{-i\omega_1 t_1} e^{i\omega_2 t_2} &\iff [A_1^- A_2 - D_1^- D_2] + i[A_1^- D_2 + D_1^- A_2] \\
e^{i\omega_2 t_2} &\iff [A_1^0 A_2 - D_1^0 D_2] + i[A_1^0 D_2 + D_1^0 A_2]
\end{aligned} \tag{D.3}
$$

where $A_1^+ A_2$ indicates a line at (ω_1, ω_2), $A_1^- A_2$ indicates a line at $(-\omega_1, \omega_2)$ and $A_1^0 A_2$ indicates a line at $(0, \omega_2)$. Taking appropriate linear combinations yields

$$
\begin{aligned}
\cos(\omega_1 t_1) e^{i\omega_2 t_2} &\overset{\Re}{\Longrightarrow} \tfrac{1}{2}[+A_1^+ A_2 - D_1^+ D_2 + A_1^- A_2 - D_1^- D_2] \\
\sin(\omega_1 t_1) e^{i\omega_2 t_2} &\overset{\Re}{\Longrightarrow} \tfrac{1}{2}[+A_1^+ D_2 + D_1^+ A_2 - A_1^- D_2 - D_1^- A_2] \\
e^{i\omega_2 t_2} &\overset{\Re}{\Longrightarrow} +A_1^0 A_2 - D_1^0 D_2 \\
i\cos(\omega_1 t_1) e^{i\omega_2 t_2} &\overset{\Re}{\Longrightarrow} \tfrac{1}{2}[-A_1^+ D_2 - D_1^+ A_2 - A_1^- D_2 - D_1^- A_2] \\
i\sin(\omega_1 t_1) e^{i\omega_2 t_2} &\overset{\Re}{\Longrightarrow} \tfrac{1}{2}[+A_1^+ A_2 - D_1^+ D_2 - A_1^- A_2 + D_1^- D_2] \\
ie^{i\omega_2 t_2} &\overset{\Re}{\Longrightarrow} -A_1^0 D_2 - D_1^0 A_2
\end{aligned} \tag{D.4}
$$

where $s \overset{\Re}{\Longrightarrow} S$ indicates that S is the real part of the Fourier transform pair of s. The six functions on the right hand side may be conveniently labelled S_1, S_2, ..., S_6, allowing the real part of the Fourier transform of equation 4.31 to be written as

$$
S_\pm(\omega_1, \omega_2) = C_1^\pm S_1 + C_2^\pm S_2 + C_3^\pm S_3 + C_4^\pm S_4 + C_5^\pm S_5 + C_6^\pm S_6 \tag{D.5}
$$

where the factor of I_0 has been dropped. This time the S_+ spectrum is frequency reversed along ω_1 before being added to S_-. This frequency reversal inverts the dispersive components but not the absorptive components, so that terms in A_1^- are converted to terms in A_1^+, but terms in D_1^- are converted to terms in $-D_1^+$; similarly terms in A_1^0 are unchanged, while terms in D_1^0 are converted to terms in $-D_1^0$. Collecting all the above gives

$$
\begin{aligned}
S'_+(\omega_1, \omega_2) &= +\tfrac{1}{2}(C_1^+ + C_5^+)(+A_1^- A_2 + D_1^- D_2) + \tfrac{1}{2}(C_1^+ - C_5^+)(+A_1^+ A_2 + D_1^+ D_2) \\
&\quad +\tfrac{1}{2}(C_2^+ + C_4^+)(-A_1^+ D_2 + D_1^+ A_2) + \tfrac{1}{2}(C_2^+ - C_4^+)(+A_1^- D_2 - D_1^- A_2) \\
&\quad +\tfrac{1}{2}C_3^+(+A_1^0 A_2 + D_1^0 D_2) + \tfrac{1}{2}C_6^+(-A_1^0 D_2 + D_1^0 A_2) \qquad \text{(D.6)} \\
S_-(\omega_1, \omega_2) &= +\tfrac{1}{2}(C_1^- + C_5^-)(+A_1^+ A_2 - D_1^+ D_2) + \tfrac{1}{2}(C_1^- - C_5^-)(+A_1^- A_2 - D_1^- D_2) \\
&\quad +\tfrac{1}{2}(C_2^- + C_4^-)(-A_1^- D_2 - D_1^- A_2) + \tfrac{1}{2}(C_2^- - C_4^-)(+A_1^+ D_2 + D_1^+ A_2) \\
&\quad +\tfrac{1}{2}C_3^-(+A_1^0 A_2 - D_1^0 D_2) + \tfrac{1}{2}C_6^-(-A_1^0 D_2 - D_1^0 A_2) \qquad \text{(D.7)}
\end{aligned}
$$

which must be summed to give the overall spectrum.

The final spectrum can be conveniently written as the sum of twelve terms:

$$\begin{aligned}
S(\omega_1,\omega_2) &= A_1^+ A_2 \quad [\tfrac{1}{2}(+C_1^+ + C_1^- - C_5^+ + C_5^-)] \\
&+ D_1^+ D_2 \quad [\tfrac{1}{2}(+C_1^+ - C_1^- - C_5^+ - C_5^-)] \\
&+ A_1^+ D_2 \quad [\tfrac{1}{2}(-C_2^+ + C_2^- - C_4^+ - C_4^-)] \\
&+ D_1^+ A_2 \quad [\tfrac{1}{2}(+C_2^+ + C_2^- + C_4^+ - C_4^-)] \\
&+ A_1^- A_2 \quad [\tfrac{1}{2}(+C_1^+ + C_1^- + C_5^+ + C_5^-)] \\
&+ D_1^- D_2 \quad [\tfrac{1}{2}(+C_1^+ - C_1^- + C_5^+ + C_5^-)] \\
&+ A_1^- D_2 \quad [\tfrac{1}{2}(+C_2^+ - C_2^- - C_4^+ - C_4^-)] \\
&+ D_1^- A_2 \quad [\tfrac{1}{2}(-C_2^+ - C_2^- + C_4^+ - C_4^-)] \\
&+ A_1^0 A_2 \quad [+C_3^+ + C_3^-] \\
&+ D_1^0 D_2 \quad [+C_3^+ - C_3^-] \\
&+ A_1^0 D_2 \quad [-C_6^+ - C_6^-] \\
&+ D_1^0 A_2 \quad [+C_6^+ - C_6^-]. \qquad \text{(D.8)}
\end{aligned}$$

Substituting for these coefficients from equation D.2 gives

$$\begin{aligned}
S(\omega_1,\omega_2) &= A_1^+ A_2 \quad [+\cos^3\alpha(\sin\lambda'+1) + \sin^2\alpha\cos\lambda'\cos\alpha] \\
&+ D_1^+ D_2 \quad [-\cos\lambda'\sin\alpha\cos\alpha] \\
&+ A_1^+ D_2 \quad [+\sin\alpha\cos^3\alpha(\cos\lambda'-1) - \sin\alpha\cos\alpha\sin\lambda'] \\
&+ D_1^+ A_2 \quad [-\sin^2\alpha\cos\alpha\sin\lambda'] \\
&+ A_1^- A_2 \quad [+\cos^3\alpha(\sin\lambda'-1) - \sin^2\alpha\cos\lambda'\cos\alpha] \\
&+ D_1^- D_2 \quad [-\cos\lambda'\sin\alpha\cos\alpha] \\
&+ A_1^- D_2 \quad [+\sin\alpha\cos^3\alpha(\cos\lambda'-1) + \sin\alpha\cos\alpha\sin\lambda'] \\
&+ D_1^- A_2 \quad [-\sin^2\alpha\cos\alpha\sin\lambda'] \\
&+ A_1^0 A_2 \quad [+2\cos\alpha\sin\lambda'\sin^2\alpha] \\
&+ D_1^0 D_2 \quad [+2\cos\lambda'\sin\alpha\cos\alpha] \\
&+ A_1^0 D_2 \quad [+2\sin^3\alpha\cos\alpha(\cos\lambda'-1)] \\
&+ D_1^0 A_2 \quad [-2\sin^2\alpha\cos\alpha\sin\lambda']. \qquad \text{(D.9)}
\end{aligned}$$

Appendix E

Listings

> C, noun: A programming language that is sort of like Pascal except more like assembly except that it isn't very much like either one, or anything else. It is either the best language available to the art today, or it isn't.
>
> Ray Simard (found in a "Fortune Cookie").

The MEM algorithm of Skilling and Bryan[72] has been implemented as a `FORTRAN` program by G. J. Daniell of Southampton University; for use in this thesis, the program was translated into `C`. A version of this program suitable for processing PMRFI data is presented below.

The program may be divided into two parts:

1. Subroutines from `maxent` inclusive. This is the implementation of the Skilling and Bryan algorithm, and is based on Daniell's program. This portion of the program is independent of the problem being solved.

2. Subroutines before `maxent`. These depend on the problem being solved, in this case processing PMRFI data. Of particular importance are the routines `opus` and `tropus`; these define operations equivalent to the matrices $\boldsymbol{R}$ and $\bar{\boldsymbol{R}}^{\mathrm{T}}$ respectively.

The program is followed by two short files containing details of the input/output routines used by the program. The Fourier transform routine `dfft` is simply a double precision implementation of the Numerical Recipes routine `four1`[65].

```
/* Maximum Entropy Method in SUN C.  Based on fortran 77
   version by G.J.Daniell.

   Program to process PMRFI data.  Assumes that both the
   incremental pulse and the phase encoding pulse are between
   0 and 180 degrees along the sample.  Cosequently while w2
   goes from 0 to 2pi, w1 only goes from 0 to pi.

   J. A. Jones (1991).
*/

#include <stdio.h>
#include <math.h>

FILE *eopen();
float getfloat();

#define PI 3.14159265358979323846
#define TWOPI (PI*2.0)
#define PIBY2 (PI/2.0)
#define HUGENUM 1.0e20

/* Major global variables.  The experimental data consists
   of two sets of NT1 fids each of NT2 complex points, while
   the resulting spectrum contains NF1 by NF2 real points.
   P and N are the total sizes (measured in real points)
   of the data and the spectrum respectively.  M is the number
   of search vectors used (dimensionality of the search
   subspace).
*/

#define NT1 16
#define NT2 128
#define NF1 64
#define NF2 256
#define P 4*NT1*NT2
#define N NF1*NF2
#define M 3

/* datum holds the experimental data, sigma holds the estimated
   error for each point, f contains the spectrum, base is a
   baseline parameter, blank is related to base and chizer is
   the target value of chi-squared (typically P).  max is the
   maximum number of iterations before giving up.
*/

double datum[P],sigma[P],f[N],base[N],blank,chizer;
int max;

/* Minor global variables.  These are used within the maxent
   function but are not accessed or set except by it and its
   subfunctions.
 */

double chisqr,chitarg,xsum,beta[M];
double s1[M],s2[M][M],c1[M],c2[M][M];

/* Variable referring to experimental details.  These are only used
   in calcmat.
*/

double t90,sw;

/* Opus and tropus calculate linear transformations between the
   data and image spaces.  Formally these involve a matrix O and
   its transpose OT.  In normal NMR processing O and OT are
   replaced by FFT routines; these are much faster (~N ln N
   instead of ~N^2) and do not require O to be written out.  For
```

```
   PMRFI data we cannot use an FFT in both dimensions, but it is
   fortunately not necessary to write out the full version of
   O (which contains P*N = 4*NT1*NT2*NF1*NF2 elements).  The
   transformation in F2 can be achieved by an FFT, that in F1
   must be done by direct multiplication by a matrix I call g.
   Because of off resonance effects F depends on F2 (but not on
   T2) and so contains 4*NT1*NF1*NF2 elements.

   It is convenient to think of the real and imaginary parts of
   the theta+x and theta-x experiments as forming four parts of
   a "hypercomplex" number, and to consider them all together.
   Hence the structure definition below.
*/

struct hc {
    double rp;
    double ip;
    double rm;
    double im;
} g[NT1][NF2][NF1];

/* If desired weightings can be included in opus and tropus to achieve
   lineshape deconvolution.  gw holds the weighting applied along t2.
*/

double gw[2*NT2];

main()
{
    getdata();
    calcmat();
    maxent();
    putdata();
}

getdata()
{
    float sig,b,w,temp[2*NT2];
    int i,j,k,l,nignore;
    FILE *fp;

    printf("\n\n\nMEM processing of PMRFI data.\n\n\n");
    printf("Data assumed to consist of 2 sets of %d sets ",NT1);
    printf("of fids\neach containing %d complex points ",NT2);
    printf("making a\ntotal of %d real points.\n\n",P);

    fp=eopen("Data file? ","r");

/* Read in data one fid at a time into temp and then rearrange
   it into datum.  datum is NT1*2*NT2*2, while data in file is
   2*NT1*NT2*2.
*/

    for (i=0;i<2;i++) {
        for (j=0;j<NT1;j++) {
            rfloat(fp,temp,2*NT2);
            for (k=0;k<NT2;k++) {
                l=2*NT2*i+4*NT2*j+2*k;
                datum[l]=(double)temp[2*k];
                datum[l+1]=(double)temp[2*k+1];
            }
        }
    }

    fclose(fp);

/* Now get parameters */
```

```
    sig=getfloat("Sigma ? ");
    b=getfloat("Blank ? ");
    nignore=getint("Number of complex points to ignore in t2 ? ");
    blank=(double) b;
    max=getint("Maximum number of iterations? ");
    for (i=0;i<2*NT1;i++) {
        for (j=0;j<2*nignore;j++)
            sigma[2*NT2*i+j]=HUGENUM;
        for (j=2*nignore;j<2*NT2;j++)
            sigma[2*NT2*i+j]=(double) sig;
    }
    chizer=(double) (4*NT1*(NT2-nignore));

    sw=getfloat("Spectral width (Hz) ? ");
    t90=getfloat("Nominal 90 pulse length (s) ? ");
    w=getfloat("Exponential constant in t2 (0=none) ? ");
    if (w==0.0)
        w=1.0e20;
    for (i=0;i<NT2;i++)
        gw[2*i]=gw[2*i+1]=exp(-(double)i/w);
}

putdata()
{
    FILE *fp;
    float temp[NF2];
    int i,j;

    fp=eopen("Output file? ","w");
    for (i=0;i<NF1;i++) {
        for (j=0;j<NF2;j++)
            temp[j]=(float)f[i*NF2+j];
        wfloat(fp,temp,NF2);
    }
    fclose(fp);
}

calcmat()

/* Function to calculate the matrix g*/

{
    double w,df,t,t1,l,l1,a,ct,st,cl,sl,ca,c3,sa,s2,s3,b1;
    int i,j,k;

    for (i=0;i<NT1;i++) {
        for (j=0;j<NF1;j++) {
            w=PI*j/NF1;
            t=w*i;
            b1=(2.0*j)/NF1;
            if (b1<1e-10)
                b1=1e-10;
            l=PIBY2*b1;
            for (k=0;k<NF2;k++) {
                df=sw*((double)k/NF2-0.5);
                a=atan(4*df*t90/b1);
                t1=t/cos(a);
                l1=l/cos(a);
                ct=cos(t1);
                st=sin(t1);
                cl=cos(l1);
                sl=sin(l1);
                ca=cos(a);
                c3=ca*ca*ca;
                sa=sin(a);
                s2=sa*sa;
                s3=s2*sa;
                g[i][k][j].rp =sl*c3*ct-cl*sa*ca*ct;
```

```
                g[i][k][j].rp+=sl*sa*ca*st;
                g[i][k][j].rp+=cl*sa*ca+sl*s2*ca;
                g[i][k][j].ip =sl*s2*ca*ct-sa*c3*ct+cl*sa*c3*ct;
                g[i][k][j].ip+=c3*st+cl*s2*ca*st;
                g[i][k][j].ip+=cl*s3*ca-sl*s2*ca-s3*ca;
                g[i][k][j].rm =sl*c3*ct+cl*sa*ca*ct;
                g[i][k][j].rm-=sl*sa*ca*st;
                g[i][k][j].rm-=cl*sa*ca-sl*s2*ca;
                g[i][k][j].im =cl*sa*c3*ct-sl*s2*ca*ct-sa*c3*ct;
                g[i][k][j].im-=c3*st+cl*s2*ca*st;
                g[i][k][j].im+=cl*s3*ca+sl*s2*ca-s3*ca;
            }
        }
    }
}

opus(x,ox)
double x[],ox[];
{
    double z,temp[NT1][2][NF2*2];
    int i,j,k,lr,li;

    for (i=0;i<NT1;i++) {
        for (j=0;j<NF2;j++) {
            lr=2*j;
            li=2*j+1;
            temp[i][0][lr]=temp[i][0][li]=0.0;
            temp[i][1][lr]=temp[i][1][li]=0.0;
            for (k=0;k<NF1;k++) {
                z=x[NF2*k+j];
                temp[i][0][lr]+=z*g[i][j][k].rp;
                temp[i][0][li]+=z*g[i][j][k].ip;
                temp[i][1][lr]+=z*g[i][j][k].rm;
                temp[i][1][li]+=z*g[i][j][k].im;
            }
        }
        dfft(temp[i][0],NF2,1);
        dfft(temp[i][1],NF2,1);
        lr=NT2*i*4;
        li=lr+2*NT2;
        for (j=0;j<2*NT2;j++) {
            ox[lr+j]=temp[i][0][j]*gw[j];
            ox[li+j]=temp[i][1][j]*gw[j];
        }
    }
}

tropus(ox,x)
double ox[],x[];
{
    double temp[NT1][2][NF2*2];
    int i,j,k,lr,li,m;

    for (i=0;i<N;i++)
        x[i]=0.0;
    for (i=0;i<NT1;i++) {
        lr=NT2*i*4;
        li=lr+NT2*2;
        for (j=0;j<2*NT2;j++) {
            temp[i][0][j]=ox[lr+j]*gw[j];
            temp[i][1][j]=ox[li+j]*gw[j];
        }
        for (j=2*NT2;j<2*NF2;j++)
            temp[i][0][j]=temp[i][1][j]=0.0;
        dfft(temp[i][0],NF2,-1);
        dfft(temp[i][1],NF2,-1);
        for (j=0;j<NF2;j++) {
            for (k=0;k<NF1;k++) {
```

```
                m=NF2*k+j;
                lr=2*j;
                li=lr+1;
                x[m]+=temp[i][0][lr]*g[i][j][k].rp;
                x[m]+=temp[i][0][li]*g[i][j][k].ip;
                x[m]+=temp[i][1][lr]*g[i][j][k].rm;
                x[m]+=temp[i][1][li]*g[i][j][k].im;
            }
        }
    }
}

/************************************************************
*This is the actual Maximum Entropy Subroutine.  The original*
*was written by G.J.Daniell, Dept. Theoretical Physics,      *
*Southampton based on the algorithm of Bryan and Skilling in *
*Mon. Not. Roy. Astron. Soc. 211 (1984) 111.                 *
************************************************************/

maxent()
{
    double e,snorm,cnorm,tnorm,a,b,c,s,test;
    double ox[P],eta[M][P];
    double cgrad[N],sgrad[N],xi[M][N];
    int i,j,k,iter,converged;

/* initialisation section */
    e=2.7182818284;
    iter=0;
    chitarg=chizer;
    if (blank>0.0)
        for (i=0;i<N;i++)
            base[i]=blank;
    else {
        blank=0;
        for (i=0;i<N;i++)
            blank+=base[i];
        blank/=N;
    }
    for (i=0;i<N;i++)
        f[i]=base[i];

/* start of main loop */
    do {
        opus(f,ox);
        chisqr=0.0;
        for (i=0;i<P;i++) {
            a=ox[i]-datum[i];
            chisqr+=(a*a)/(sigma[i]*sigma[i]);
            ox[i]=2.0*a/(sigma[i]*sigma[i]);
        }
        tropus(ox,cgrad);
        xsum=test=snorm=cnorm=tnorm=test=0.0;
        for (i=0;i<N;i++) {
            xsum+=f[i];
            sgrad[i]=-log(2.0*f[i]/base[i])/blank;
            snorm+=sgrad[i]*sgrad[i]*f[i];
            cnorm+=cgrad[i]*cgrad[i]*f[i];
            tnorm+=sgrad[i]*cgrad[i]*f[i];
        }
        snorm=sqrt(snorm);
        cnorm=sqrt(cnorm);
        a=1.0;
        b=1.0/cnorm;
        c=b;
        if (iter!=0) {
            test=sqrt(0.5*(1.0-tnorm/(snorm*cnorm)));
            a=1.0/(snorm*2.0*test);
```

```
    b=1.0/(cnorm*2.0*test);
}
for (i=0;i<N;i++) {
    xi[0][i]=f[i]*c*cgrad[i];
    xi[1][i]=f[i]*(a*sgrad[i]-b*cgrad[i]);
}
opus(xi[0],eta[0]);
opus(xi[1],eta[1]);
for (i=0;i<P;i++)
    ox[i]=eta[1][i]/(sigma[i]*sigma[i]);
tropus(ox,xi[2]);
a=0.0;
for (i=0;i<N;i++) {
    b=f[i]*xi[2][i];
    a+=b*xi[2][i];
    xi[2][i]=b;
}
a=1.0/(sqrt(a));
for (i=0;i<N;i++)
    xi[2][i]*=a;
opus(xi[2],eta[2]);
for (i=0;i<M;i++) {
    s1[i]=c1[i]=0.0;
    for (j=0;j<N;j++) {
        s1[i]+=xi[i][j]*sgrad[j];
        c1[i]+=xi[i][j]*cgrad[j];
    }
    c1[i]/=chisqr;
}
for (i=0;i<M;i++) {
    for (j=0;j<=i;j++) {
        s2[i][j]=c2[i][j]=0.0;
        for (k=0;k<N;k++)
            s2[i][j]-=xi[i][k]*xi[j][k]/f[k];
        for (k=0;k<P;k++) {
            double sig;
            sig=sigma[k]*sigma[k];
            c2[i][j]+=eta[i][k]*eta[j][k]/sig;
        }
        s2[i][j]/=blank;
        c2[i][j]*=2.0/chisqr;
    }
}
c2[0][1]=c2[1][0];
c2[0][2]=c2[2][0];
c2[1][2]=c2[2][1];
s2[0][1]=s2[1][0];
s2[0][2]=s2[2][0];
s2[1][2]=s2[2][1];
s=0.0;
for (i=0;i<N;i++)
    s-=f[i]*log(2.0*f[i]/(base[i]*e));
a=s/xsum;
s/=blank*e;
printf("%3d %11.4e %11.4e ",iter,test,s);
printf("%11.4e %11.4e %11.4e %11.4e\n",chitarg,chisqr,xsum,a);
beta[0]=-0.5*c1[0]/c2[0][0];
beta[1]=beta[2]=0.0;
if (iter!=0)
    move();
for (i=0;i<N;i++) {
    for (j=0;j<M;j++)
        f[i]+=beta[j]*xi[j][i];
    if (f[i]<0.0)
        f[i]=blank/1000.0;
}
iter++;
converged=(test<0.02 && fabs(chisqr/chizer-1.0)<0.01);
```

```
    } while (iter<=max && !converged);
}

move()
{
    double dist(),chinow();
    double cmin,ctarg,a1,a2,anew,f1,f2,fx,w;
    int i;

    cmin=chinow(0.0);
    if (cmin*chisqr>chizer)
        ctarg=0.5*(cmin+1.0);
    else
        ctarg=chizer/chisqr;
    a1=0.0;
    a2=1.0;
    f1=cmin-ctarg;
    f2=chinow(1.0)-ctarg;
    do {
        anew=0.5*(a1+a2);
        fx=chinow(anew)-ctarg;
        if (f1*fx>0.0) {
            a1=anew;
            f1=fx;
        }
        if (f2*fx>0.0) {
            a2=anew;
            f2=fx;
        }
    } while (fabs(fx)>=1.0e-3);
    w=dist();
    if (w>0.1*xsum/blank)
        for (i=0;i<M;i++)
            beta[i]*=sqrt(0.1*xsum/(blank*w));
    chitarg=ctarg*chisqr;
}

double dist()
{
    double w,z;
    int i,j;

    w=0.0;
    for (i=0;i<M;i++) {
        z=0.0;
        for (j=0;j<M;j++)
            z-=s2[i][j]*beta[j];
        w+=beta[i]*z;
    }
    return w;
}

double chinow(ax)
double ax;
{
    double bx,a[M][M],b[M],w,z;
    int i,j;

    bx=1.0-ax;
    for (i=0;i<M;i++) {
        for (j=0;j<M;j++)
            a[i][j]=bx*c2[i][j]-ax*s2[i][j];
        b[i]=-(bx*c1[i]-ax*s1[i]);
    }
    chosol(a,b);
    w=0.0;
    for (i=0;i<M;i++) {
        z=0.0;
```

```
        for (j=0;j<M;j++)
            z+=c2[i][j]*beta[j];
        w+=beta[i]*(c1[i]+0.5*z);
    }
    return (w+1.0);
}

chosol(a,b)
double a[M][M],b[M];
{
    double l[M][M],bl[M],z;
    int i,j,k;

    for (i=0;i<M;i++)
        for (j=0;j<M;j++)
            l[i][j]=0.0;

    l[0][0]=sqrt(a[0][0]);
    for (i=1;i<M;i++) {
        l[i][0]=a[i][0]/l[0][0];
        for (j=1;j<=i;j++) {
            z=0.0;
            for (k=0;k<j;k++)
                z+=l[i][k]*l[j][k];
            z=a[i][j]-z;
            if (j==i)
                l[i][j]=sqrt(z);
            else
                l[i][j]=z/l[j][j];
        }
    }
    bl[0]=b[0]/l[0][0];
    for (i=1;i<M;i++) {
        z=0.0;
        for (j=0;j<i;j++)
            z+=l[i][j]*bl[j];
        bl[i]=(b[i]-z)/l[i][i];
    }
    beta[M-1]=bl[M-1]/l[M-1][M-1];
    for (i=0;i<M-1;i++) {
        k=M-i-2;
        z=0.0;
        for (j=k+1;j<M;j++)
            z+=l[j][k]*beta[j];
        beta[k]=(bl[k]-z)/l[k][k];
    }
}
```

```
/* Various utilities to make file access on the Sun slightly
   more friendly; eg error messages on open.
   Depends on screenio.

   J. A. Jones (1991).
*/

#include <stdio.h>

/* Open a file; if not openable ask for another filename */

FILE *eopen(message,mode)
char message[],mode[];
{
    char fname[80];
    FILE *fp;

    do {
        getstring(message,fname,80);
        fp=fopen(fname,mode);
        if (fp==NULL)
            printf("Cannot open %s, mode=%s\n",fname,mode);
    } while (fp==NULL);
    return fp;
}

/* Write/Read n floats to/from file *fp from/to array a */

wfloat(fp,a,n)
FILE *fp;
float a[];
int n;
{
    int i,j;

    for (i=0;i<n;i++) {
        j=fprintf(fp,"%e\n",a[i]);
        if (j<=0) {
            printf("Error writing line %d\n",i);
            exit(1);
        }
    }
}

rfloat(fp,a,n)
FILE *fp;
float a[];
int n;
{
    int i,j;

    for (i=0;i<n;i++) {
        j=fscanf(fp,"%f",&a[i]);
        if (j!=1) {
            printf("Error reading line %d\n",i);
            exit(1);
        }
    }
}
```

```
/* Various utilities to make screen/terminal i/o on the Sun slightly
   more friendly.

   J. A. Jones (1991).  Based on FORTRAN routines by J. Keeler (1984).
*/

#include <stdio.h>
#include <string.h>
#define MAXLINE 80

getint(s)
char s[];
{
    char line[MAXLINE];
    int value,n;

    do {
        printf("%s",s);
        getline(line,MAXLINE);
        n=sscanf(line,"%d",&value);
        if (n!=1)
            printf("Please enter an integer.\n");
    } while (n!=1);
    return value;
}

float getfloat(s)
char s[];
{
    char line [MAXLINE];
    int n;
    float value;

    do {
        printf("%s",s);
        getline(line,MAXLINE);
        n=sscanf(line,"%g",&value);
        if (n!=1)
            printf("Please enter a float.\n");
    } while (n!=1);
    return value;
}

getstring(s,t,max)
char s[],t[];
int max;
{
    int n;

    printf("%s",s);
    n=getline(t,max);
    if (t[n-1]=='\n')
        t[--n]='\0';
}

getline(line,max)                   /* see K&R 165 */
char line[];
int max;
{
    if (fgets(line,max,stdin)==NULL) {
        printf("EOF\n");
        exit(1);
    }
    else
        return strlen(line);
}
```

Bibliography

[1] E. M. Purcell, H. C. Torrey and R. V. Pound, *Phys. Rev.* **69,** 37 (1946).

[2] F. Bloch, W. W. Hansen and M. E. Packard, *Phys. Rev.* **69,** 127 (1946).

[3] R. R. Ernst, G. Bodenhausen and A. Wokaun, *"Principles of Nuclear Magnetic Resonance in One and Two Dimensions"*, Oxford University Press (1987).

[4] A. E. Derome, *"Modern NMR Techniques for Chemistry Research"*, Pergamon Press (1987).

[5] D. G. Gadian, *"Nuclear magnetic resonance and its application to living systems"*, Oxford University Press (1982).

[6] P. G. Morris, *"Nuclear Magnetic Resonance Imaging in Medicine and Biology"*, Oxford University Press (1986).

[7] B. Buck and V. A. Macaulay (Eds.), *"Maximum Entropy in Action"*, Oxford University Press (1991).

[8] G. J. Daniell, *"Of maps and monkeys"*, in reference [7].

[9] P. W. Atkins, *"Molecular Quantum Mechanics"*, Oxford University Press, second edition (1983).

[10] W. G. Proctor and F. C. Yu, *Phys. Rev.* **77,** 717 (1950).

[11] R. Freeman, *"A handbook of nuclear magnetic resonance"*, Longman (1987).

[12] B. R. Frieden, *J. Opt. Soc. Am.* **62,** 511 (1972).

[13] S. F. Gull and G. J. Daniell, *Nature (London)* **272,** 686 (1978).

[14] D. S. Stephenson, *Prog. NMR Spectrosc.* **20,** 515 (1988).

[15] S. Sibisi, *Nature (London)* **301,** 134 (1983).

[16] S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue and J. Staunton, *Nature (London)* **311,** 446 (1984).

[17] F. Ni, G. C. Levy and H. Scheraga, *J. Magn. Reson.* **66,** 385 (1986).

[18] A. R. Mazzeo and G. C. Levy, *Comput. Enhanced. Spectrosc.* **3,** 165 (1986).

[19] P. J. Hore, *J. Magn. Reson.* **62,** 561 (1985).

[20] J. C. Hoch, *J. Magn. Reson.* **64,** 436 (1985).

[21] J. F. Martin, *J. Magn. Reson.* **65,** 291 (1985).

[22] E. D. Laue, M. R. Mayger, J. Skilling, and J. Staunton, *J. Magn. Reson.* **68,** 14 (1986).

[23] E. D. Laue, K. O. B. Pollard, J. Skilling, J. Staunton and A. C. Sutkowski, *J. Magn. Reson.* **72,** 493 (1987).

[24] A. Heuer and U. Haeberlen, *J. Magn. Reson.* **85,** 79 (1989).

[25] S. J. Davies, C. Bauer, P. J. Hore and R. Freeman, *J. Magn. Reson.* **76,** 476 (1988).

[26] P. J. Hore and G. J. Daniell, *J. Magn. Reson.* **69,** 386 (1986).

[27] J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling and S. J. P. Worrall, *J. Magn. Reson.* **73,** 69 (1987).

[28] J. C. J. Barna and E.D. Laue, *J. Magn. Reson.* **75,** 384 (1987).

[29] M. A. Delsuc and G. C. Levy, *J. Magn. Reson.* **76,** 306 (1988).

[30] E. D. Laue, J. Skilling, J. Staunton, S. Sibisi and R. G. Brereton, *J. Magn. Reson.* **62,** 437 (1985).

[31] E. D. Laue, J. Skilling and J. Staunton, *J. Magn. Reson.* **63,** 418 (1985).

[32] K. M. Wright and P. S. Belton, *Mol. Phys.* **58,** 485 (1986).

[33] J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling and S. J. P. Worrall, *Biochem. Soc. Trans.* **14,** 1262 (1986).

[34] M. A. Delsuc, F. Ni and G. C. Levy, *J. Magn. Reson.* **73,** 548 (1987).

[35] J. C. J. Barna and E. D. Laue, *Lab. Practice* **36,** 102 (1987).

[36] M. L. Waller and P. S. Tofts, *Magn. Reson. Med.* **4,** 385 (1987).

[37] R. H. Newman, *J. Magn. Reson.* **79,** 448 (1988).

[38] A. R. Mazzeo, M. A. Delsuc, A. Kumar and G. C. Levy, *J. Magn. Reson.* **81,** 512 (1989).

[39] T. N. Huckerby, I. A. Nieduszynski, G. H. Cockin, J. M. Dickenson, H. Morris, P. N. Sanderson and D. J. Thornton, *Eur. Polym. J.* **25,** 861 (1989).

[40] G. J. Daniell and P. J. Hore, *J. Magn. Reson.* **84,** 515 (1989).

[41] J. C. Hoch, A. S. Stern, D. L. Donoho and I. M. Johnstone, *J. Magn. Reson.* **86,** 236 (1990).

[42] P. J. Hore, D. S. Grainger, S. Wimperis and G. J. Daniell, *J. Magn. Reson.* **89,** 415 (1990).

[43] P. J. Hore, *"Maximum entropy and nuclear magnetic resonance"*, in reference [7].

[44] S. Davies, K. J. Packer, A. Baruya and A. J. Grant, *"Enhanced information recovery in spectroscopy"*, in reference [7].

[45] D. L. Donoho, I. M. Johnstone, A. S. Stern and J. C. Hoch, *Proc. Natn. Acad. Sci. USA* **87,** 5066 (1990).

[46] A. S. Stern and J. C. Hoch, *J. Magn. Reson.* **97,** 255 (1992).

[47] D. S. Grainger, D. Phil. Thesis, Oxford University (1991).

[48] P. Hodgkinson, Chemistry Part II Thesis, Oxford University (1992).

[49] J. A. Jones, Chemistry Part II Thesis, Oxford University (1989).

[50] J. A. Jones and P. J. Hore, *J. Magn. Reson.* **92,** 276 (1991).

[51] J. A. Jones and P. J. Hore, *J. Magn. Reson.* **92,** 363 (1991).

[52] J. A. Jones, P. J. Hore, C. P. Relf, R. Ouwerkerk and P. Styles, *J. Magn. Reson.* **98,** 73 (1991).

[53] J. A. Jones, P. J. Hore, D. S. Grainger and G. J. Daniell, *J. Magn. Reson.* (in press).

[54] J. A. Jones, discussion section of reference [71].

[55] F. Bloch, *Phys. Rev.* **70,** 460 (1946).

[56] A. Abragam, *"Principles of Nuclear Magnetism"*, Oxford University Press (1961).

[57] P. A. M. Dirac, *"The Principles of Quantum Mechanics"*, 4th Edition, Oxford University Press (1958).

[58] F. Bloch and A. Siegert, *Phys. Rev.* **57,** 522 (1940).

[59] R. R. Ernst and W. A. Anderson, *Rev. Sci. Instrum.* **107,** 46 (1957).

[60] R. N. Bracewell, *"The Fourier Transform and its Applications"*, 2nd Edition, McGraw–Hill (1985).

[61] E. O. Brigham, *"The Fast Fourier Transform"*, Prentice–Hall (1974).

[62] P. R. Griffiths, *"Transform Techniques in Chemistry"*, Plenum (1978).

[63] J. C. Lindon and A. G. Ferrige, *Prog. NMR Spectrosc.* **14,** 27 (1980).

[64] J. W. Cooley and J. W. Tukey, *Math. Comput.* **19,** 297 (1965).

[65] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *"Numerical Recipes"*, Cambridge University Press (1986).

[66] E. Bartholdi and R. R. Ernst, *J. Magn. Reson.* **11,** 9 (1973).

[67] P. A. Jansson (Ed.), *"Deconvolution With Applications in Spectroscopy"*, Academic Press (1984).

[68] T. Bayes, *Phil. Trans. Roy. Soc. Lond.* **53,** 370 (1763).

[69] T. Bayes, *Phil. Trans. Roy. Soc. Lond.* **54,** 296 (1764).

[70] C. E. Shannon, *Bell System Tech. J.* **27,** 379 & 623 (1948).

[71] D. L. Donoho, I. M. Johnstone, J. C. Hoch and A. S. Stern, *J. R. Statist. Soc. B* **54,** 41 (1992).

[72] J. Skilling and R. K. Bryan, *Mon. Not. R. Astron. Soc.* **211,** 111 (1984).

[73] L. T. Muus, P. W. Atkins, K. A. McLauchlan and J. B. Pedersen, *"Chemically Induced Magnetic Polarization"*, Nato Advanced Study Institutes C.34, Reidel (1977).

[74] S. F. Gull and J. Skilling, *IEEE Proc. F* **131,** 646 (1984).

[75] J. Skilling, *Nature (London)* **309,** 748 (1984).

[76] J. Skilling, *Nature (London)* **312,** 382 (1984).

[77] S. F. Gull, discussion section of reference [71].

[78] A. Jones, *The Computer Journal* **13,** 301 (1970).

[79] W. W. F. Pijnappel, Ph. D. Thesis, University of Delft (1991).

[80] A. van den Bos in *"Handbook of Measurement Science"*, (P. H. Sydenham, Ed.), Wiley (1982).

[81] J. P. Norton, *"An introduction to identification"*, Academic Press (1986).

[82] G. B. Dantzig, *"Linear Programming and Extensions"*, Princeton University Press (1963).

[83] The Numerical Algorithms Group Ltd, *"The NAG Fortran Library—Mark 13"*, (1988).

[84] A. J. Redfearn, *The Times (London)*, 17th October 1984.

[85] R. Lieu, R. B. Hicks and C. J. Bland, *Mon. Not. R. Astron. Soc.* **229,** 49p (1987).

[86] H. Barkhuijsen, R. de Beer and D. van Ormondt, *J. Magn. Reson.* **67,** 371 (1986).

[87] R. Nadjari and J.-Ph. Grivet, *J. Magn. Reson.* **91,** 351 (1991).

[88] D. I. Hoult, C.-N. Chen and L. K. Hedges, *Ann. N.Y. Acad. Sci.* **508,** 366 (1987).

[89] M. J. Blackledge, B. Rajagopalan, R. D. Oberhaensli, N. M. Bolas, P. Styles and G. K. Radda, *Proc. Natl. Acad. Sci. USA* **84,** 4283 (1987).

[90] R. E. Gordon, P. E. Hanley, D. Shaw, D. G. Gadian, G. K. Radda, P. Styles, P. J. Bore and L. Chan, *Nature (London)* **287,** 736 (1980).

[91] J. J. H. Ackerman, T. H. Grove, G. G. Wong, D. G. Gadian and G. K. Radda, *Nature (London)* **283,** 167 (1980).

[92] D. I. Hoult, *J. Magn. Reson.* **33,** 183 (1979).

[93] S. J. Cox and P. Styles, *J. Magn. Reson.* **40,** 209 (1981).

[94] W. R. Smythe, *"Static and Dynamic Electricity"*, 3rd Edition, McGraw–Hill (1968).

[95] W. H. Beyer (Ed.), *"CRC Standard Mathematical Tables"*, 25th Edition, CRC Press (1979).

[96] M. J. Blackledge and P. Styles, *J. Magn. Reson.* **77,** 203 (1988).

[97] M. Goldman, *"Quantum Description of High-Resolution NMR in Liquids"*, Oxford University Press (1988).

[98] D. Neuhaus, G. Wagner, M. Vasak, J. H. R. Kägi and K. Wütrich, *Eur. J. Biochem.* **151,** 257 (1985).

[99] J. J. Titman and J. Keeler, *J. Magn. Reson.* **89,** 640 (1990).

[100] L. Braunschweiler and R. R. Ernst, *J. Magn. Reson.* **53,** 521 (1983).

[101] A. A. Bothner-By and J. Dadok, *J. Magn. Reson.* **72,** 540 (1987).

[102] J.-M. Le Parco, L. McIntyre and R. Freeman, *J. Magn. Reson.* **97,** 553 (1992).

[103] L. McIntyre and R. Freeman, *J. Magn. reson.* **96,** 425 (1992).

[104] R. A. Jackson, *J. Magn. Reson.* **75,** 174 (1987).

[105] *"The New Encyclopædia Britannica"*, 15th Edition, University of Chicago (1974).

[106] G. A. Barnard, *Biometrika* **45,** 293 (1958).

[107] S. M. Stigler, *"The History of Statistics: The Measurement of Uncertainty before 1900"*, Harvard University Press (1986).

[108] L. E. Maistrov, *"Probability Theory, a historical sketch"*, translated by S. Kotz, Academic Press (1974).

[109] I. Todhunter, *"A History of the Mathematical Theory of Probability, from the time of Pascal to that of Laplace"*, MacMillan (1865).

[110] E. J. G. Pitman, *"Some Remarks on Statistical Inference"*, in *"Bernoulli (1713), Bayes (1763), Laplace (1813)"*, (J. Neyman and L. M. Le Cam, Eds.), Springer–Verlag (1965).

[111] W. Kemp, *"NMR in Chemistry: A Multinuclear Introduction"*, MacMillan (1986).

[112] J. Bernoulli, *"Ars Conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola Gallicè scripta de ludo pilæ reticularis"*, Basil (1713).

[113] E. T. Jaynes, *Foundations of Physics* **3,** 477 (1973).

[114] J. S. Rowlinson, *Nature (London)* **225,** 1196, (1970).

[115] E. T. Jaynes, *"Where do we stand on Maximum Entropy?"*, in *"Papers on probability, statistics and statistical physics"*, (R. D. Rosenkrantz, Ed.), Reidel (1983).

[116] R. L. Ellis, *Phil. Mag. S. 3* **37,** 322 (1850).

[117] U. Eco, *"The Name of the Rose"*, translated by W. Weaver, Picador (1984).

[118] J. N. D. Kelly, *"The Oxford Dictionary of Popes"*, Oxford University Press (1986).

[119] W. of Ockham, *"Exposito in Libros Physicorum Aristotelsis"*, in *"Guillelmi de Ockham, Opera Philosophica et Theologica"*, St Bonaventure University (1985).

[120] A. Garrett, *Physics World* **4.5,** 39 (1991).

[121] *"American Men & Women of Science"*, 18th Edition, R. R. Bowker (1992).

[122] J. Horgan, *Sci. Am.* **262,** 16 (1990).

[123] A. J. M. Garrett, Appendix A of *"Macroirreversibility and microreversibility reconciled: the second law"*, in reference [7].

[124] C. Huygens, *"Christiani Hugenii Libellus de Ratiociniis in Ludo Aleae. Or, the Value of all Chances in Games of Fortune; Cards, Dice, Wagers, Lotteries, etc. Mathematically Demonstrated"*, translated by W. Browne, London (1714).

[125] C. Huygens, *"Of the Laws of Chance, or, A Method of Calculating the Hazards of Game, Plainly Demonstrated, And applied to Games at present most in Use, which may be easily extended to the most intricate Cases of Chance imaginable"*, translated by J. Arbuthnot, London (1692).

[126] A. De Moivre, *"The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play"*, 2nd Edition, London (1738). Facsimile edition by Frank Cass & Co. Ltd (1967).

[127] A. De Moivre, *Phil. Trans. Roy. Soc. Lond.* **27,** 214 (1711).