J. Mag. Res., **93,** pp.369-394 1991

Bayesian Analysis. IV. Noise And Computing Time Considerations [1]

G. LARRY BRETTHORST
Washington University,
Department of Chemistry
Campus Box 1134
1 Brookings Drive,
St. Louis, Missouri 63130-4899

ABSTRACT. Probability theory, when interpreted as logic, enables one to ask many questions not possible with the frequency interpretation of probability theory. Often, answering these questions can be computationally intensive. If these techniques are to find their way into general use in NMR, a way that allows one to calculate the probability for the frequencies, amplitudes, and decay rate constants quickly and easily must be found. In this paper, a procedure that allows one to compute the posterior probability for the frequencies, amplitudes, and decay rate constants from a series of zero-padded discrete Fourier transforms of the complex FID data when the data have been multiplied by a decaying exponential is described. Additionally, the calculation is modified to include prior information about the noise, and it is shown that obtaining a sample of the noise is almost as important as obtaining a signal sample, because it allows one to investigate complicated spectra using simple models. Three examples are given to illustrate the calculations.

Introduction

In previous works (*1–3*) probability theory, when interpreted as logic, was applied to the combined problems of NMR parameter estimation, model selection, and signal detection. The formalism used in those papers enables one to understand, on an intuitive level, what frequencies and decay rate constants will be estimated and why. But the formalism used is not the best (in the sense of computationally most efficient) way to program one's computer and, perhaps more importantly, the formalism hides the fact that for sinusoidal models with exponential decay, all of the relevant statistics for parameter estimation may be computed from a zero-padded discrete Fourier transform of the complex FID data when the data have been multiplied by a decaying exponential. If one is willing to accept the digital resolution implicit in a zero-padded discrete Fourier transform, then essentially all parameter estimation problems in high-resolution NMR may be done rapidly. For more on Bayesian probability theory and applications to NMR, see Refs. [1] through [15].

This paper is organized as follows: first, the general class of models used in the calculation is discussed. Then the parameter estimation calculation is modified to include information about the noise. After this generalization is completed the calculation is then organized in such a way that the resulting numerical calculation is computationally efficient. Last, three examples of the calculation are given to illustrate the effect of including various types of prior information.

In the first example, the single frequency case is examined for both stationary and exponentially decaying sinusoids. In this example, the posterior probability for a single stationary sinusoidal frequency independent of the amplitude, phase, and noise variance is used to illustrate the effect of including information about the noise. This calculation is then generalized to include decay, and the posterior probability for the frequency is computed independent of the amplitude, phase, variance of the noise, and decay rate constant.

[1] Copies of all computer codes used in the examples are available from the author on $5\frac{1}{4}$ inch floppies.

In the second example, the posterior probability for two frequencies of oscillation is computed independent of the amplitudes, phases, variance of the noise, and decay rate constants. This calculation illustrates how probability theory can be used to "see" inside a single peak in the discrete Fourier transform and illustrates how including even a modest amount of prior information about the signal can improve the resolution of the frequencies.

In the third example, the joint posterior probability for the center frequency and the J coupling constant of an arbitrary multiplet is computed independent of the amplitude, phase, variance of the noise, and decay rate constant. This example is then specialized to a triplet model and used to illustrate that probability theory can still determine the J coupling constant, and center frequency of a multiplet, even when no multiplet structure is visible in the traditional absorption spectrum. Last, the calculations are applied to more complex multiplets, a 1:1:1 triplet and a triplet of doublets.

The Model Signal

The calculation described here concerns parameter estimation for quadrature detected, high resolution NMR FID data. As was previously discussed [1] for quadrature detected FID data, there are actually two data sets, the so-called real data (0° phase) and the so-called imaginary data (90° phase). Here it is assumed that the data may be written as the sum of a "signal" plus an unknown "randomly" varying component referred to as "noise." For the real channel the data may be written as

$$d_R(t_i) = f_R(t_i) + e_R(t_i), \tag{1}$$

and similarly for the imaginary channel

$$d_I(t_i) = f_I(t_i) + e_I(t_i), \tag{2}$$

where $f_R(t_i)$ and $f_I(t_i)$ represent the signal in the real and imaginary channels, and $e_R(t_i)$ and $e_I(t_i)$ represent the noise. The signals f_R and f_I are assumed to be identical except for a 90° phase shift. The variance of the noise is assumed to be the same in both data sets, but the actual samples of noise realized in the two data sets are assumed to be different. In addition to the data $D \equiv \{d_R(t_1), \ldots, d_R(t_N), d_I(t_1), \ldots, d_I(t_N)\}$, which have been sampled at discrete times t_i $(1 \le i \le N)$, it is assumed that a sample of the noise, $D_\sigma \equiv \{e_R(t'_1), \ldots, e_R(t'_{N_\sigma}), e_I(t'_1), \ldots, e_I(t'_{N_\sigma})\}$, has been obtained. This noise FID was sampled at discrete times t'_i $(1 \le i \le N_\sigma)$, and has a mean square given by

$$\overline{d^2_\sigma} \equiv \frac{1}{2N_\sigma}\left[e_R \cdot e_R + e_I \cdot e_I\right], \tag{3}$$

where " $\cdot$ " means sum over discrete times, e.g.,

$$e_R \cdot e_R \equiv \sum_{i=1}^{N_\sigma} e_R(t'_i) e_R(t'_i). \tag{4}$$

The most general model signal that is considered in this paper is written as

$$f_R(t_i) = \sum_{j=1}^{m} B_j U_j(t_i) \tag{5}$$

for the real channel and

$$f_I(t_i) = \sum_{j=1}^{m} B_j V_j(t_i) \tag{6}$$

for the imaginary channel, where B_j is the amplitude of the jth signal function. The signal functions $U_j(t_i)$ and $V_j(t_i)$ are assumed to be of the form

$$U_j(t_i) = \sum_{k=1}^{n_j} \mathrm{w}_{jk} G_{jk}(t_i) \mathrm{e}^{-\alpha_{jk} t_i} \tag{7}$$

and

$$V_j(t_i) = \sum_{k=1}^{n_j} \mathrm{w}_{jk} H_{jk}(t_i) \mathrm{e}^{-\alpha_{jk} t_i}, \tag{8}$$

where w_{jk} is a weight, α_{jk} is an exponential decay rate constant, and n_j is the number of terms used in the jth signal function. The weights w_{jk} are not assumed to be constants (they could be a function of a parameter). For example, they could be functions of a global phase. The $G_{jk}(t_i)$ functions are either a cosine or sine,

$$G_{jk}(t_i) \equiv \begin{cases} \cos(\omega_{jk} t_i) \\ \sin(\omega_{jk} t_i), \end{cases} \tag{9}$$

and the functions $H_{jk}(t_i)$ are just the $G_{jk}(t_i)$, but shifted by a 90° phase shift. They are defined as

$$H_{jk}(t_i) \equiv \begin{cases} -\sin(\omega_{jk} t_i) & \text{if } G_{jk} = \cos(\omega_{jk} t_i) \\ \cos(\omega_{jk} t_i) & \text{if } G_{jk} = \sin(\omega_{jk} t_i), \end{cases} \tag{10}$$

where ω_{jk} is the frequency of oscillation. The model signals f_R and f_I essentially define what is meant by an arbitrary exponentially decaying sinusoid.

Neither the decay rate constants nor the frequencies are assumed to be distinct; that is, α_{jk} could be equal to α_{lm} and ω_{jk} could be equal to ω_{lm}. For example, suppose that one wishes to investigate a triplet, and one wants to compute the posterior probability for the center frequency, ω_c, and the J coupling, Δ. In the weak coupling regime, for a spin-$\frac{1}{2}$ system, a triplet should have frequencies located at $\omega_c - \Delta$, ω_c, and $\omega_c + \Delta$. In addition, the amplitudes should be in a ratio of 1:2:1, and there should be a common phase and decay rate constant. There are two signal functions in the real channel and two in the imaginary channel. The model signal f_R which incorporates all of this prior information is given by

$$\begin{aligned} f_R(t_i) &= B_1 \left\{ \cos([\omega_c - \Delta]t_i) + 2\cos(\omega_c t_i) + \cos([\omega_c + \Delta]t_i) \right\} \mathrm{e}^{-\alpha t_i} \\ &+ B_2 \left\{ \sin([\omega_c - \Delta]t_i) + 2\sin(\omega_c t_i) + \sin([\omega_c + \Delta]t_i) \right\} \mathrm{e}^{-\alpha t_i}. \end{aligned} \tag{11}$$

There is a similar model signal f_I in the imaginary channel:

$$\begin{aligned} f_I(t_i) &= -B_1 \left\{ \sin([\omega_c - \Delta]t_i) + 2\sin(\omega_c t_i) + \sin([\omega_c + \Delta]t_i) \right\} \mathrm{e}^{-\alpha t_i} \\ &+ B_2 \left\{ \cos([\omega_c - \Delta]t_i) + 2\cos(\omega_c t_i) + \cos([\omega_c + \Delta]t_i) \right\} \mathrm{e}^{-\alpha t_i}. \end{aligned} \tag{12}$$

In the notation introduced earlier, the frequencies $\omega_{1j} = \omega_{2j}$ for $j = \{1, 2, 3\}$, and the frequencies are located at $\omega_{11} = \omega_c - \Delta$, $\omega_{12} = \omega_c$, and $\omega_{13} = \omega_c + \Delta$. Similarly, the weights $\mathrm{w}_{1j} = \mathrm{w}_{2j}$ for $j = \{1, 2, 3\}$, and the weights are given by $\mathrm{w}_{11} = 1$, $\mathrm{w}_{12} = 2$, and $\mathrm{w}_{13} = 1$. There is only a single decay rate constant, $\alpha = \alpha_{jk}$, for all j and k. There are three nonlinear parameters: ω_c, the center frequency of the multiplet; Δ, the J coupling constant; and α, the exponential decay rate constant. Use of this model is demonstrated later, where the joint posterior probability for the center frequency ω_c, and the J coupling constant Δ are computed independent of the amplitude, phase, variance of the noise, and exponential decay rate constant.

Including Information About The Noise

The parameters of interest are the frequencies and perhaps the decay rate constants. All of the parameters which appear in the model equations in a nonlinear fashion are labeled $\mathbf{\Theta} \equiv \{\Theta_1, \ldots, \Theta_r\}$, where r is the total number of nonlinear $\mathbf{\Theta}$ parameters; i.e., for the triplet model $\Theta_1 = \omega_c$, $\Theta_2 = \Delta$, and $\Theta_3 = \alpha$. The general model represented by Eqs. [5–10] includes most models of interest in high resolution NMR.

In this section prior information about the noise is included in the calculation. Including prior information about the noise proves to be very important, because it sets a scale against which small effects in the data may be tested, thus allowing complex spectra to be investigated using relatively simple models. When prior information about the noise is not included in the calculation [1], this is more difficult because use of the Student t distribution requires all systematic effects in the data to be accounted for in the model.

Including information about the noise modifies the calculation in a straightforward way. Here the calculation begins after the amplitudes have been removed as nuisances, but before the variance of the noise is removed [1]. At this point in the calculation one has

$$\begin{aligned} P(\mathbf{\Theta}|D_\sigma, D, I) &\propto \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} \int d\sigma P(\sigma|I) P(D_\sigma|\sigma, I) \\ &\times (2\pi\sigma^2)^{\frac{m-2N}{2}} \exp\left\{-\frac{2N\overline{d^2} - m\overline{h^2}}{2\sigma^2}\right\}, \end{aligned} \tag{13}$$

where $\overline{d^2}$ is the mean-square data value (Eq. [17]), $\overline{h^2}$ summarizes all of the information in the data relevant to the estimation of nonlinear $\mathbf{\Theta}$ parameters and plays the role of a sufficient statistic; it is defined in Eq. [21]. The posterior probability for the nonlinear $\mathbf{\Theta}$ parameters, $P(\mathbf{\Theta}|D_\sigma, D, I)$, is the quantity that is computed for the general model, Eqs. [5–10]. The factor $\lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}}$ is the Jacobian of the g_{jl} matrix (Eq. [24]). The exponential term is essentially the marginal likelihood of the nonlinear $\mathbf{\Theta}$ parameters and represents what was learned about the nonlinear $\mathbf{\Theta}$ parameters from the data.

The three factors, $P(\sigma|I)$, $P(D_\sigma|\sigma, I)$, and the exponential, represent everything that is known about the noise. The first term, $P(\sigma|I)$, is the prior probability for the standard deviation of the noise. It represents what was known about the noise before either the data or the noise sample was obtained. The second term, $P(D_\sigma|\sigma, I)$, is the probability for the noise sample given that one knows the standard deviation of the noise. It represents what was learned about the standard deviation of the noise from the noise sample. The third term, the exponential, represents what was learned about the noise from the data D.

In previous work (*1–3*), no noise sample was gathered. In those works a Jeffreys prior was assigned to $P(\sigma|I)$, and that is done here also – see Refs. [4] and [5] for more on the Jeffreys prior. The Jeffreys prior is simply $1/\sigma$. This prior is not strictly speaking a probability density function at all; rather it is the limiting form of a sequence of proper probability density functions in the limit of infinite uncertainty in the standard deviation. This prior could be bounded and the calculation done more carefully, but in parameter estimation problems, bounding this prior and taking limits is unnecessary – the results are unchanged from using the improper prior directly.

The second term, $P(D_\sigma|\sigma, I)$, is the probability of the noise sample given that one knows the standard deviation of the noise. When the marginal posterior probability for the nonlinear $\mathbf{\Theta}$ parameters was derived a Gaussian prior probability was assigned for the noise prior. For consistency a Gaussian must be used here. This gives

$$P(D_\sigma|\sigma, I) = (2\pi\sigma^2)^{-N_\sigma/2} \exp\left\{-\frac{2N_\sigma \overline{d_\sigma^2}}{2\sigma^2}\right\} \tag{14}$$

as the direct probability for the noise sample, D_σ.

Multiplying the Jeffreys prior, $1/\sigma$, and the direct probability for the noise sample (Eq. [14]) and inserting these into the posterior probability for the nonlinear $\mathbf{\Theta}$ parameters (Eq. [13]) gives

$$\begin{aligned} P(\mathbf{\Theta}|D_\sigma, D, I) &\propto \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} \int d\sigma \sigma^{-1} (2\pi\sigma^2)^{-N_\sigma/2} \exp\left\{-\frac{2N_\sigma \overline{d_\sigma^2}}{2\sigma^2}\right\} \\ &\times (2\pi\sigma^2)^{\frac{m-2N}{2}} \exp\left\{-\frac{2N\overline{d^2} - m\overline{h^2}}{2\sigma^2}\right\}. \end{aligned} \tag{15}$$

By evaluating the integral over the standard deviation, σ, one obtains

$$\begin{aligned} P(\mathbf{\Theta}|D_\sigma, D, I) &\propto \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} \left[1 - \frac{m\overline{h^2}}{2N_\sigma \overline{d_\sigma^2} + 2N\overline{d^2}}\right]^{\frac{m-2N-2N_\sigma}{2}} \\ &\approx \left[1 - \frac{m\overline{h^2}}{2N_\sigma \overline{d_\sigma^2} + 2N\overline{d^2}}\right]^{\frac{m-2N-2N_\sigma}{2}}. \end{aligned} \tag{16}$$

For typical NMR models the Jacobian is a slowly varying function of the nonlinear $\mathbf{\Theta}$ parameters. Because it is slowly varying it will be nearly a constant over the region where the posterior probability is sharply peaked and will cancel when the distribution is normalized. Additionally, its size (for sinusoidal models) is typically of the order of $\sqrt{N}$. Since the calculation is only accurate to $\sqrt{N}$, ignoring this term will not introduce any additional uncertainties into the calculation. The mean-square data value, $\overline{d^2}$, is defined as

$$\overline{d^2} \equiv \frac{1}{2N} \left[d_R \cdot d_R + d_I \cdot d_I\right] \tag{17}$$

with

$$d_R \cdot d_R \equiv \sum_{i=1}^{N} d_R(t_i) d_R(t_i) \tag{18}$$

and

$$d_I \cdot d_I \equiv \sum_{i=1}^{N} d_I(t_i) d_I(t_i). \tag{19}$$

When no noise sample is available, $N_\sigma = 0$; these equations reduce to those found in Ref. [1]. This formulation of the problem represents a generalization of the previous work.

The term $2N_\sigma \overline{d_\sigma^2}$ places a floor on the minimum value obtainable in the denominator and effectively guarantees that the Student t distribution, Eq. [16], can never become singular. This floor sets a scale against which very small effects in the data can be tested: having a good determination of the noise will allow one to investigate complex spectra by modeling the features of interest. For small noise samples, $N_\sigma \overline{d_\sigma^2} \ll N\overline{d^2}$, the denominator is effectively unchanged and the height of the logarithm of the posterior probability density scales like the number of data values plus the number of noise values – so gathering a noise sample the same size as the data sample effectively doubles the logarithm of the posterior probability for the nonlinear $\mathbf{\Theta}$ parameters and improves the resolution of the parameters when simple models are used. In the other limit, $N_\sigma \overline{d_\sigma^2} \gg N\overline{d^2}$, the posterior probability density goes smoothly into

$$P(\mathbf{\Theta}|D_\sigma, D, I) \propto \exp\left\{\frac{m\overline{h^2}}{2\overline{d_\sigma^2}}\right\}, \tag{20}$$

where $N_\sigma + N - 1/N_\sigma \approx 1$ was assumed. This probability density function is what one would have obtained if the standard deviation of the noise were known from prior information. Thus the Student t distribution, Eq. [16], smoothly interpolates from the case where the noise variance is completely unknown to the case where the noise variance is completely determined from prior information.

Organizing the Numerical Calculation

In this section three things are done: first, the calculation of the sufficient statistic are expressed in a way that is computationally more stable than the formulation given in [1]. Second, the g_{jl} matrix is expressed in closed form. Third, the calculation is organized in a way that allows one to compute the sufficient statistics from a series of zero-padded discrete Fourier transforms of the complex data when the data have been multiplied by a decaying exponential.

In the formulation introduced earlier [1], the sufficient statistic was computed from the eigenvalues and eigenvectors of the g_{jl} matrix (Eq. [24]). Here an alternate formulation that is computationally more stable and more efficient is used. In this formulation, the sufficient statistic is given by

$$\overline{h^2} \equiv \frac{1}{m} \sum_{l=1}^{m} \hat{B}_l T_l, \tag{21}$$

where the values for the $\hat{B}_l$ are given by the solution to the linear system

$$\sum_{j=1}^{m} \hat{B}_j g_{jl} = T_l, \tag{22}$$

and the projections of the data onto the model, the T_l, are given by

$$T_l \equiv d_R \cdot U_l + d_I \cdot V_l = \sum_{i=1}^{N} d_R(t_i) U_l(t_i) + d_I(t_i) V_l(t_i). \tag{23}$$

The matrix g_{jl} is defined as

$$g_{jl} \equiv \epsilon \delta_{jl} + U_j \cdot U_l + V_j \cdot V_l = \epsilon \delta_{jl} + \sum_{i=1}^{N} U_j(t_i) U_l(t_i) + V_j(t_i) V_l(t_i), \tag{24}$$

where ϵ is a small positive number (small compared to $U_j \cdot U_j + V_j \cdot V_j$), and δ_{jl} is the Kronecker delta function.

The term $\epsilon \delta_{jl}$ did not appear in the previous formulation [1]. It was obtained here by using an informative prior probability when the amplitudes are removed as nuisance parameters. This prior,

$$P(\mathbf{B}|I) = \left[\frac{2\pi\sigma^2}{\epsilon}\right]^{-\frac{m}{2}} \exp\left\{-\frac{\epsilon}{2\sigma^2} \sum_{j=1}^{m} B_j^2\right\}, \tag{25}$$

effectively places a limit on the size of the amplitudes. The parameter ϵ is similar to a fractional variance; small values of ϵ correspond to large uncertainties in the amplitudes. This additional information is what is needed to stabilize the numerical calculation: it ensures that the g_{jl} matrix is never singular. The value of ϵ is assumed to be known, but even if ϵ is unknown, any value of ϵ small compared to the diagonal will work to stabilize the numerical calculation.

For uniformly sampled data, the sums over discrete times appearing in the g_{jl} may be done explicitly. To see this, the model signal, Eqs. [5]–[10], are substituted into the definition of g_{jl},

Eq. [24], to obtain

$$\begin{aligned} g_{jl} &= \epsilon\delta_{jl} + \sum_{i=1}^{N} U_j(t_i)U_l(t_i) + V_j(t_i)V_l(t_i) \\ &= \epsilon\delta_{jl} + \sum_{k=1}^{n_j}\sum_{s=1}^{n_l} \mathrm{w}_{jk}\mathrm{w}_{ls} \sum_{i=1}^{N} [G_{jk}(t_i)G_{ls}(t_i) + H_{jk}(t_i)H_{ls}(t_i)]\, \mathrm{e}^{-(\alpha_{jk}+\alpha_{ls})t_i} \\ &= \epsilon\delta_{jl} + \sum_{k=1}^{n_j}\sum_{s=1}^{n_l} \mathrm{w}_{jk}\mathrm{w}_{ls} W_{jlks}, \end{aligned} \tag{26}$$

where

$$W_{jlks} \equiv \begin{cases} C(\omega_{jk}-\omega_{ls}, \alpha_{jk}+\alpha_{ls}), & \text{if } G_{jk} = \cos(\omega_{jk}t) \text{ and } G_{ls} = \cos(\omega_{ls}t), \\ -S(\omega_{jk}-\omega_{ls}, \alpha_{jk}+\alpha_{ls}), & \text{if } G_{jk} = \cos(\omega_{jk}t) \text{ and } G_{ls} = \sin(\omega_{ls}t), \\ S(\omega_{jk}-\omega_{ls}, \alpha_{jk}+\alpha_{ls}), & \text{if } G_{jk} = \sin(\omega_{jk}t) \text{ and } G_{ls} = \cos(\omega_{ls}t), \\ C(\omega_{jk}-\omega_{ls}, \alpha_{jk}+\alpha_{ls}), & \text{if } G_{jk} = \sin(\omega_{jk}t) \text{ and } G_{ls} = \sin(\omega_{ls}t), \end{cases} \tag{27}$$

and

$$\begin{aligned} C(\omega,\alpha) &\equiv \sum_{i=1}^{N} \cos(\omega t_i)\mathrm{e}^{-\alpha t_i} \\ &= \frac{1-\cos(\omega)\mathrm{e}^{-\alpha} - \cos(N\omega)\mathrm{e}^{-N\alpha} + \cos[(N-1)\omega]\mathrm{e}^{-(N+1)\alpha}}{1-2\cos(\omega)\mathrm{e}^{-\alpha}+\mathrm{e}^{-2\alpha}}, \end{aligned} \tag{28}$$

$$\begin{aligned} S(\omega,\alpha) &\equiv \sum_{i=1}^{N} \sin(\omega t_i)\mathrm{e}^{-\alpha t_i} \\ &= \frac{\sin(\omega)\mathrm{e}^{-\alpha} - \sin(N\omega)\mathrm{e}^{-N\alpha} + \sin[(N-1)\omega]\mathrm{e}^{-(N+1)\alpha}}{1-2\cos(\omega)\mathrm{e}^{-\alpha}+\mathrm{e}^{-2\alpha}}. \end{aligned} \tag{29}$$

The sums, in the definitions of $C(\omega,\alpha)$ and $S(\omega,\alpha)$, were done by first changing the sine or cosine to exponential form. These sums are of the form $\sum_{k=0}^{N-1} x^k$, which may be summed explicitly using $1/(1-x) = 1 + x + x^2 \cdots$. Because the g_{jl} may be obtained in a closed form, its inverse may be computed and the sufficient statistic $\overline{h^2}$ may often be obtained in closed form.

Dimensionless units have been used. In these units, the times, t_i, are $t_i = \{0, 1, 2, \ldots, N-1\}$, and the frequencies and decay rate constants are measured in radians. The conversion to more conventional dimensional units is given by

$$f = \frac{\omega}{2\pi\Delta T} \text{ hertz} \quad \text{and} \quad \beta = \frac{\alpha}{\pi\Delta T} \text{ hertz}, \tag{30}$$

where f is the frequency in hertz, β is the decay rate constant in hertz, and ΔT is the time interval between data values in one channel. On many spectrometers $1/\Delta T$ is referred to as the sweep width and is the number of complex samples per second.

Obtaining the g_{jl} matrix in closed form is useful, because it allows closed form solutions to be obtained for the sufficient statistic; but it will not significantly reduce the computation time of the sufficient statistics. To do this, the time required to project the data onto the model must be reduced. But the projection of the model onto the data may be done using a fast Fourier transform.

To demonstrate this the model function, Eqs. [5] and [6] are substituted into the projections of the data onto the model, Eq. [23], to obtain

$$\begin{aligned} T_l &= d_R \cdot U_l + d_I \cdot V_l \\ &= \sum_{i=1}^{N} d_R(t_i) U_l(t_i) + d_I(t_i) V_l(t_i) \\ &= \sum_{i=1}^{N} d_R(t_i) \left[\sum_{j=1}^{n_l} \mathrm{w}_{lj} G_{lj}(t_i) \right] + \sum_{i=1}^{N} d_I(t_i) \left[\sum_{j=1}^{n_l} \mathrm{w}_{lj} H_{lj}(t_i) \right] \\ &= \sum_{j=1}^{n_l} \mathrm{w}_{lj} \left[d_R \cdot G_{lj} + d_I \cdot H_{lj} \right]. \end{aligned} \tag{31}$$

There are two cases: (i) when G_{lj} is a cosine and (ii) when $G_{lj}(t_i)$ is a sine. When $G_{lj}(t_i)$ is a cosine, the quantity $d_R \cdot G_{lj} + d_I \cdot H_{lj}$ is given by

$$\begin{aligned} \sum_{i=1}^{N} d_R(t_i) G_{lj}(t_i) \quad + \quad & d_I(t_i) H_{lj}(t_i) = \\ & \sum_{i=1}^{N} [d_R(t_i) \cos(\omega_{lj} t_i) - d_I(t_i) \sin(\omega_{lj} t_i)] \mathrm{e}^{-\alpha_{lj} t_i}. \end{aligned} \tag{32}$$

In the second case, when $G_{lj}(t_i)$ is a sine, then the quantity $d_R \cdot G_{lj} + d_I \cdot H_{lj}$ becomes

$$\begin{aligned} \sum_{i=1}^{N} d_R(t_i) G_{lj}(t_i) \quad + \quad & d_I(t_i) H_{lj}(t_i) = \\ & \sum_{i=1}^{N} [d_R(t_i) \sin(\omega_{lj} t_i) + d_I(t_i) \cos(\omega_{lj} t_i)] \mathrm{e}^{-\alpha_{lj} t_i}. \end{aligned} \tag{33}$$

These two quantities may be computed from a discrete Fourier transform. The discrete Fourier transform of the complex FID data when the data have been multiplied by a decaying exponential of decay rate constant α is given by

$$\begin{aligned} \mathcal{F}\left([d_R(t_i) + \mathrm{i} d_I(t_i)] \mathrm{e}^{-\alpha t_i}\right) &= \sum_{i=1}^{N} [d_R(t_i) + \mathrm{i} d_I(t_i)] \mathrm{e}^{-\alpha t_i} \mathrm{e}^{\mathrm{i}\omega_i t_i} \\ &= \sum_{i=1}^{N} [d_R(t_i) + \mathrm{i} d_I(t_i)] \mathrm{e}^{-\alpha t_i} [\cos(\omega t_i) + \mathrm{i} \sin(\omega t_i)] \\ &= \sum_{i=1}^{N} [d_R(t_i) \cos(\omega t_i) - d_I(t_i) \sin(\omega t_i)] \mathrm{e}^{-\alpha t_i} \\ &+ \mathrm{i} \sum_{i=1}^{N} [d_R(t_i) \sin(\omega t_i) + d_I(t_i) \cos(\omega t_i)] \mathrm{e}^{-\alpha t_i}. \end{aligned} \tag{34}$$

If one defines the functions $R(\omega, \alpha)$ as the real part of the discrete Fourier transform,

$$R(\omega, \alpha) = \sum_{i=1}^{N} [d_R(t_i) \cos(\omega t_i) - d_I(t_i) \sin(\omega t_i)] \mathrm{e}^{-\alpha t_i}, \tag{35}$$

and $I(\omega, \alpha)$ as the imaginary part

$$I(\omega, \alpha) = \sum_{i=1}^{N} [d_R(t_i) \sin(\omega t_i) + d_I(t_i) \cos(\omega t_i)] \mathrm{e}^{-\alpha t_i}, \tag{36}$$

then the projections of the data onto the model, the T_l, are given by

$$T_l = \sum_{j=1}^{n_l} \mathrm{w}_{lj} \begin{cases} R(\omega_{lj}, \alpha_{lj}) & \text{if } G_{lj} \text{ is a cosine,} \\ I(\omega_{lj}, \alpha_{lj}) & \text{if } G_{lj} \text{ is a sine.} \end{cases} \tag{37}$$

If one is willing to restrict the accuracy of the frequency estimates to the digital accuracy implicit in a zero-padded discrete Fourier transform of the complex FID data multiplied by an exponential, and if one is willing to compute several of these transforms (corresponding to different values of the decay rate constants), then essentially all posterior probability density functions associated with parameter estimation using NMR high resolution FID data may be computed with little additional computing time above the cost of the discrete Fourier transforms.

Example 1: Estimating One Frequency With Decay

In this example, the effect of obtaining a finite noise sample is demonstrated. To do this, the posterior probability for a single stationary frequency is computed without a noise sample, and then the number of noise samples is gradually increased. The maximum "height" of the posterior probability is monitored as a function of the number of noise samples.

In all of the examples (with the exception of Fig. 3), the calculations were computed from a set of 21 discrete Fourier transforms. Each of these transforms were generated by zero-padding the FID (adding zeros to the end of the FID to make a new FID of length 32K). This new FID was multiplied by a decaying exponential function. The decay rate constants for the 21 discrete Fourier transforms started at zero and went to 0.01 by steps of 0.0005 in dimensionless units. A decay rate constant of 0.01 corresponds to the signal decaying to $\exp\{-5.12\} \approx 0.006$ of the original value in the first 512 data values. These discrete Fourier transforms were done using the fast Fourier transform routine [16]. The total CPU time for this step was less than 3 min on a VAXstation 3100 M38.

Suppose that the absorption spectrum looks like that shown in Fig. 1. This spectrum was computed from a simulated FID. The three lines shown consist of an unresolved triplet (far right), two unresolved and unrelated close lines (center), and an isolated single line (far left). There are $N = 512$ data values per channel. The data were generated from

$$\begin{aligned} f_R(t_i) &= [10 \cos(0.29i) + 20 \cos(0.3i) + 10 \cos(0.31i)] \mathrm{e}^{-0.004i} \\ &+ 10 \cos(-0.5i) \mathrm{e}^{-0.004i} + 7 \cos(-0.51i) \mathrm{e}^{-0.006i} \\ &+ 12 \cos(-0.2i) \mathrm{e}^{-0.01i} + e_i \qquad (0 \le i \le 511), \end{aligned} \tag{38}$$

and similarly for the imaginary channel, where e_i represents a random noise component of unit variance. In addition to the data, the noise was sampled 100,000 times. The standard deviation of the noise was $\overline{d_\sigma^2} = 0.99961$, with $N_\sigma = 100,000$. The routine [16] was used to generate the Gaussian distributed random numbers used in this process.

In previous work [3, 8], it was demonstrated that when the sinusoids are well separated and stationary, the "best" estimate of the frequencies is given by a zero-padded, discrete Fourier transform *power spectrum*. For well separated sinusoids, the presence of exponential decay does not alter the location of the maxima [8]. Thus, the posterior probability for the frequency of a single stationary sinusoid represents an essential starting point for all problems of frequency estimation. If one computes the posterior probability for a single stationary sinusoidal frequency, independent of the

Figure 1: Absorption Spectrum: Computer Simulated Data

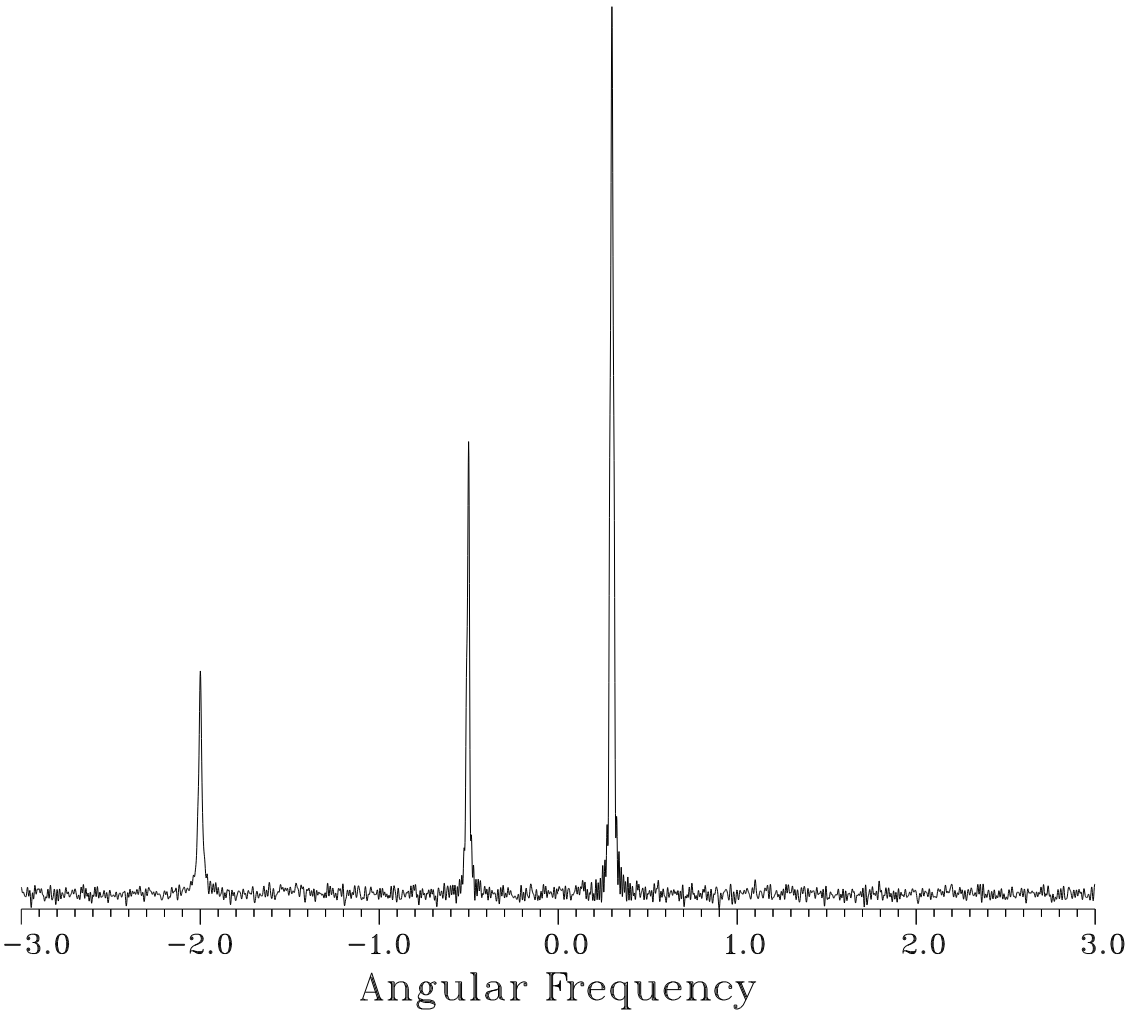

Fig. 1. This spectrum is of computer-simulated data (see text for details of data preparation). The three lines shown are respectively a triplet (far right), a single line (far left), and a pair of close unresolved and unrelated lines (center).

amplitude, phase, and variance of the noise, given the data and the sample of the noise, one finds

$$P(\omega|D_\sigma, D, I) \propto \left[1 - \frac{R(\omega,0)^2 + I(\omega,0)^2}{2N(N_\sigma \overline{d_\sigma^2} + N\overline{d^2})}\right]^{1-N-N_\sigma}. \tag{39}$$

The base 10 logarithm of $P(\omega|D_\sigma, D, I)$ has been plotted in Fig. 2A, with $N_\sigma = 0$. This probability density function was computed from the series of 21 discrete Fourier transforms that were done initially. This particular probability density function can be computed from the first discrete Fourier transform of the complex data. The total CPU time required to run this computation was on the order of 45 CPU seconds on a VAXstation 3100 M38 (most of this time was I/O).

To illustrate the effect of obtaining a sample of the noise, the height of the peak at $\omega = 0.3$ was monitored as the number of noise samples was increased. This plot, Fig. 3, rises almost linearly and then begins to approach an asymptote. The asymptote is approached when the total length of the noise vector is much greater than the total length of the data vector. Asymptotically the posterior probability for the frequency, $P(\omega|D_\sigma, D, I)$, goes into

$$P(\omega|D_\sigma, D, I) \propto \exp\left\{\frac{R(\omega,0)^2 + I(\omega,0)^2}{2N\overline{d_\sigma^2}}\right\}. \tag{40}$$

When $N_\sigma \overline{d_\sigma^2} \ll N\overline{d^2}$, the denominator in Eq. [39] is nearly unchanged, and the logarithm of the posterior probability increases approximately linearly with the number of noise samples. Thus gathering even a modest sample of the noise can greatly improve the resolution of the frequencies and decay rate constants, when simple models are used. In this example, $N_\sigma = 100,000$ and $\overline{d_\sigma^2} = 0.99961$. The posterior probability for a single stationary sinusoidal frequency, Fig. 2B, is almost $10,000$ orders of magnitude higher, and accurate estimates of the parameters will be possible, when the appropriate simple models are used to examine each of the three regions.

But in this example the lines are not stationary. Including decay in the calculation is almost a trivial extension. The model signal, for an exponentially decaying sinusoid, is given by

$$f_R(t_i) = [B_1 \cos(\omega t_i) + B_2 \sin(\omega t_i)]\mathrm{e}^{-\alpha t_i} \tag{41}$$

for the real channel, and similarly

$$f_I(t_i) = [B_2 \cos(\omega t_i) - B_1 \sin(\omega t_i)]\mathrm{e}^{-\alpha t_i} \tag{42}$$

for the imaginary channel, where B_1 and B_2 are the amplitudes of the sinusoids, ω is the frequency of oscillation, and the exponential decay rate constant, α, is treated as a nuisance parameter (a parameter that is uninteresting and the desired posterior probability density function is to be formulated independent of its value).

The joint posterior probability for the frequency, ω, and the exponential decay rate constant, α, is given by the Student t distribution, Eq. [16]. This probability density function is independent of the amplitudes, phases, and variance of the noise and is given by

$$P(\omega, \alpha|D_\sigma, D, I) \propto \left[1 - \frac{R(\omega,\alpha)^2 + I(\omega,\alpha)^2}{C(0,2\alpha)(2N_\sigma\overline{d_\sigma^2} + 2N\overline{d^2})}\right]^{1-N-N_\sigma}. \tag{43}$$

The posterior probability for the frequency, ω, independent of the decay rate constant, α, is the desired quantity. To obtain this, the product rule is applied to Eq. [43], followed by the sum rule. This gives

$$P(\omega|D_\sigma, D, I) \propto \int d\alpha P(\alpha|I)\left[1 - \frac{R(\omega,\alpha)^2 + I(\omega,\alpha)^2}{C(0,2\alpha)(2N_\sigma\overline{d_\sigma^2} + 2N\overline{d^2})}\right]^{1-N-N_\sigma} \tag{44}$$

as the posterior probability for the frequency, where $P(\alpha|I)$ is the prior probability for the decay rate constant.

Figure 2: The $\log_{10}[P(\omega|D_\sigma, D, I)]$

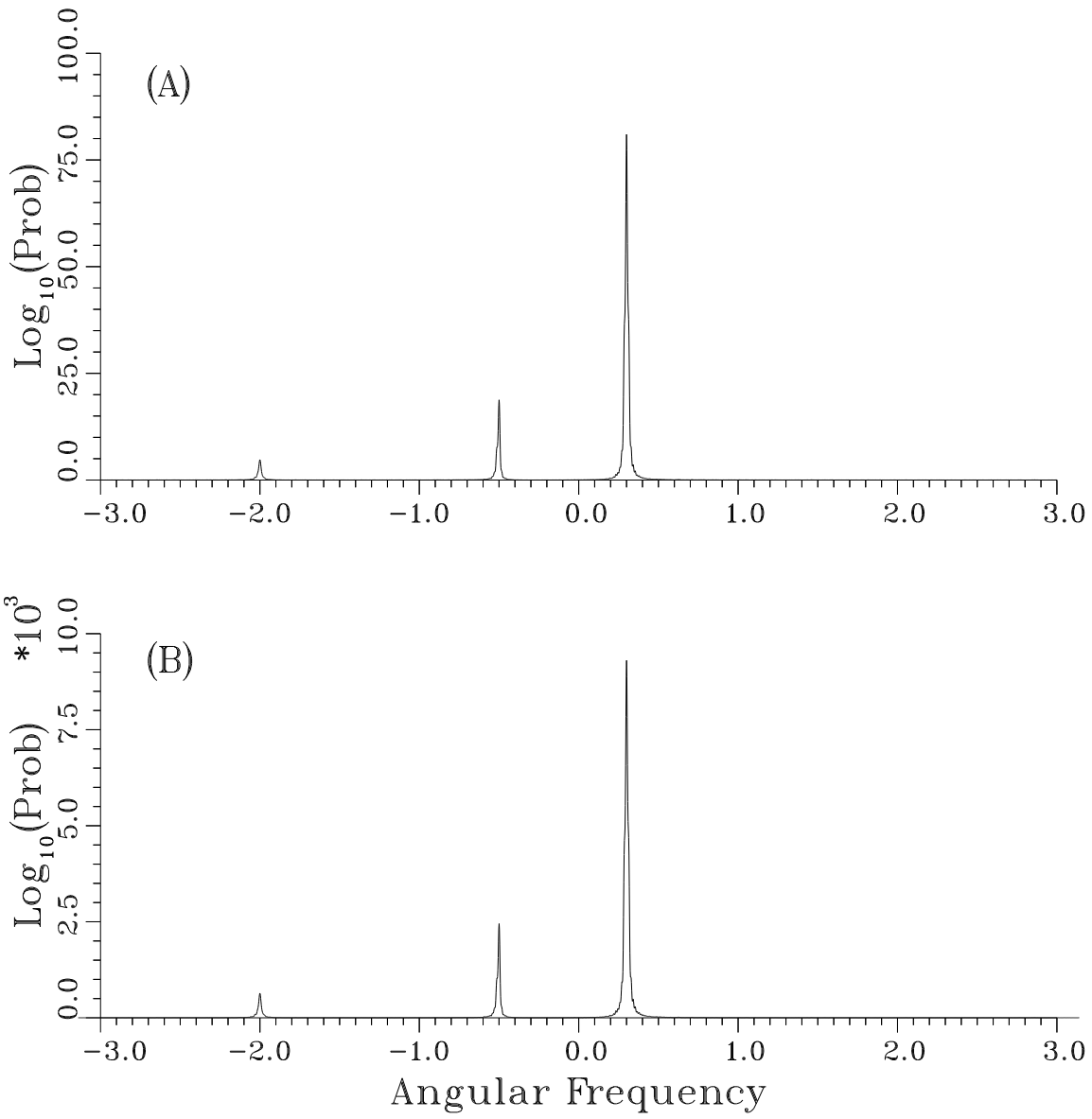

Fig. 2(A). Plot of the base 10 logarithm of the posterior probability for a single stationary sinusoidal frequency. There are $N = 512$ data points per channel, and $N_\sigma = 0$. If one knows that there are three well-separated sinusoids present, then the optimum answer to the frequency estimation problem is essentially to take the three largest peaks from the discrete Fourier transform power spectrum. Note that even the small peak at -2 is some 5 orders of magnitude above the noise, so all of these frequencies have been well determined. (2B) The same plot with $N_\sigma = 100,000$ and $\overline{d_\sigma^2} = 0.99961$. Here the posterior probability is some 10,000 orders of magnitude higher than that in (A).

Figure 3: The Effect of Sampling the Noise

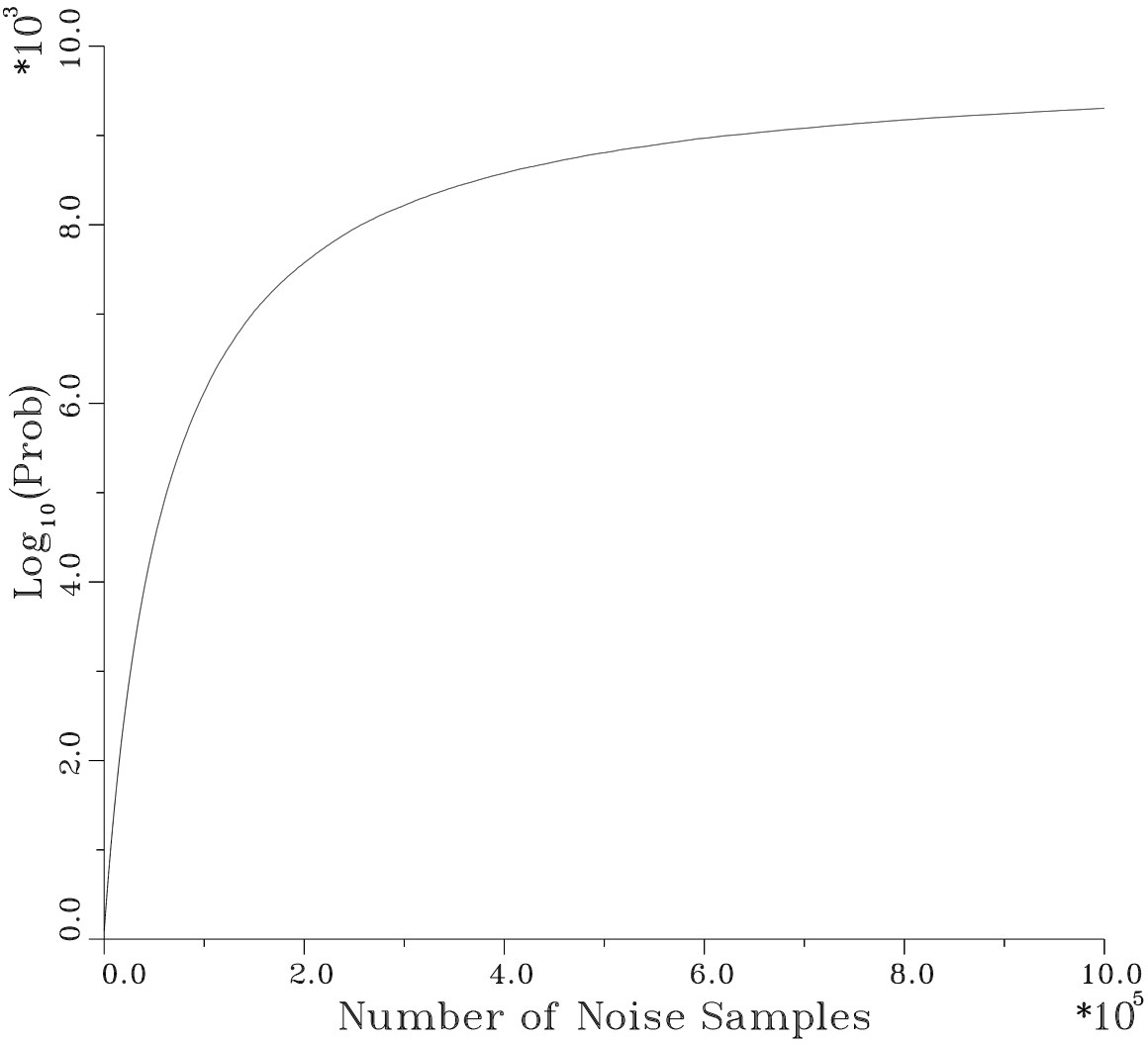

Fig. 3. Plot of the logarithm of peak value of the posterior probability for a single stationary frequency as a function of the number of noise samples N_σ. The peak at $\omega = 0.3$ rad was used in this demonstration. Note that for small N_σ, the height of the posterior probability rises almost linearly, and then as the length of the noise vector becomes large compared to the length of the data vector, the posterior probability goes asymptotically into the case where the variance of the noise is known from prior information. Because the early growth is linear, obtaining even a modest sample of the noise is very important when simple models are used.

If, as indicated earlier, a series of discrete Fourier transforms of the complex FID data, when the data have been multiplied by a decaying exponential, are used to obtain the functions $R(\omega,\alpha)$ and $I(\omega,\alpha)$, then α will be computed only on a finite set of points. Because α will be computed on a coarse grid, the joint posterior probability for both the frequency and decay rate constant will change significantly when moving one step on this grid. Unless the prior information about α is highly specific, the likelihood factor (essentially Eq. [43]) will completely dominate the prior. The decay rate constant is a scale parameter and the completely uninformative prior for a scale parameter is the Jeffreys prior $1/\alpha$. But the likelihood factor will be so sharply peaked as a function of α that the Jeffreys prior is essentially a constant over all reasonable values of α. Consequently a uniform prior will be used for no other reason than to clean up the equations. Thus the posterior probability for the frequency independent of the decay rate constant is given approximately by

$$P(\omega|D_\sigma,D,I) \propto \sum_{j=1}^{max}\left[1-\frac{R(\omega,\alpha_j)^2+I(\omega,\alpha_j)^2}{C(0,2\alpha_j)(2N_\sigma\overline{d_\sigma^2}+2N\overline{d^2})}\right]^{1-N-N_\sigma}, \tag{45}$$

where max is the number of discrete Fourier transforms used in the analysis. In these examples, $max = 21$.

The sufficient statistic for the single exponentially decaying sinusoidal model is given by

$$\begin{aligned}\overline{h^2} &= \frac{R(\omega,\alpha)^2+I(\omega,\alpha)^2}{C(0,2\alpha)} \\ &= \frac{1}{C(0,2\alpha)}\sum_{i=1}^{N}\left|\left\{d_R(t_i)\mathrm{e}^{-\alpha t_i}+id_I(t_i)\mathrm{e}^{-\alpha t_i}\right\}\mathrm{e}^{\mathrm{i}\omega t_i}\right|^2.\end{aligned} \tag{46}$$

Up to the factor $C(0,2\alpha)$, this is the *power spectrum* of the complex data, when the FID has been multiplied by a decaying exponential. If the decay rate constant α were taken to be equal to the natural linewidth of the signal, then the sufficient statistic is essentially a matched filter. But when the "true" linewidth is not known, probability theory takes a weighted sum over all possible values of α. Moreover, unlike the matched filter, probability theory will effectively change the filter bandwidth for each value of the frequency. When there are multiple well-separated sinusoids, it will automatically choose the matched filter appropriate for each frequency.

Figure 4 is a plot of the base 10 logarithm of the posterior probability for the frequency, Eq. [45], independent of the amplitude, phase, variance of the noise, and exponential decay rate constant. If one looks only at the peaks and not at the scale, one would conclude incorrectly that including decay has caused poorer resolution. If one examines the scale, it becomes obvious that this is not the case. Including decay has raised the height of these peaks by some 15,000 orders of magnitude. Including information about the decay is more important than including information about the noise; but not much more important.

As with the previous examples, this probability density function, Eq. [45], was computed from the 21 discrete Fourier transforms done initially and, as with the other examples, $N_\sigma = 100,000$ with $\overline{d_\sigma^2} = 0.99961$. Because this probability density function requires one to integrate with respect to the decay rate constant one would expect this posterior probability density function to be much more time consuming to compute than the posterior probability for a single stationary frequency. Indeed this was the case. The computation over all 32K points in frequency space required approximately four CPU minutes on a VAXstation 3100 M38.

Example 2: Estimating Two Frequencies With Decay

As was emphasized earlier [3], using a model that does not incorporate all of the relevant prior information can lead to very precise estimates of quantities that are not appropriate. This effect is

Figure 4: The Effect of Including and Removing Decay

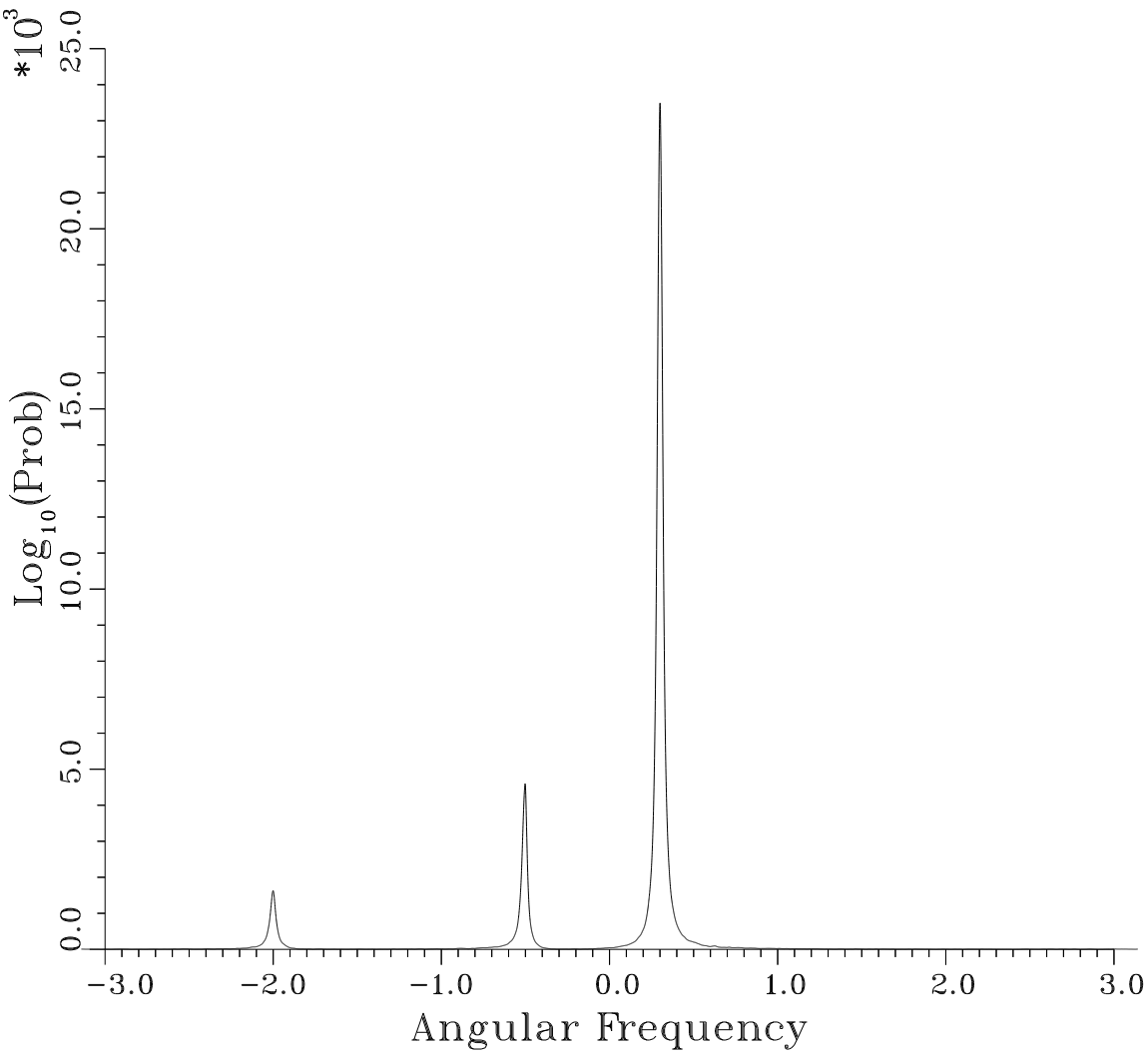

Fig. 4. Plot of the base 10 logarithm of the posterior probability for the frequency ω, independent of the amplitude, phase, variance of the noise, and the decay rate constant. Including decay has increased the height of the main peak by roughly 15,000 orders of magnitude.

evident here. The highest peak on Fig. 4 is at $\omega = 0.3$. This peak is not an isolated sinusoid; rather it is a triplet. The model employed assumes that the signal can be represented by a series of very well-separated exponentially decaying sinusoids. When this condition is met (e.g., the isolated line at -2.0 rad), Eq. [45] will give nearly optimal estimates of the frequency. But the line near 0.3 rad is a triplet, and because of the symmetries in the signal, Eq. [45] will essentially estimate the center position of the triplet. The region around -.5 rad contains two close but unrelated sinusoids, and the frequency estimated by Eq. [45] will be a complicated weighted average, not the true frequency. To estimate the value of the frequencies around $-.5$ rad one should use a model that contains two exponentially decaying sinusoids.

In this example, the posterior probability for two frequencies independent of the amplitudes, phases, variance of the noise, and decay rate constants is computed to demonstrate that probability theory can estimate the value of both frequencies even when the lines completely overlap in the absorption spectrum. The model signal which contains two exponentially decaying sinusoids is given by

$$\begin{aligned} f_R(t_i) &= [B_1\cos(\omega_1 t_i) + B_2\sin(\omega_1 t_i)]\mathrm{e}^{-\alpha_1 t_i} \\ &+ [B_3\cos(\omega_2 t_i) + B_4\sin(\omega_2 t_i)]\mathrm{e}^{-\alpha_2 t_i} \end{aligned} \tag{47}$$

for the real channel, and the model signal for the imaginary channel is given by

$$\begin{aligned} f_I(t_i) &= [B_2\cos(\omega_1 t_i) - B_1\sin(\omega_1 t_i)]\mathrm{e}^{-\alpha_1 t_i} \\ &+ [B_4\cos(\omega_2 t_i) - B_3\sin(\omega_2 t_i)]\mathrm{e}^{-\alpha_2 t_i}. \end{aligned} \tag{48}$$

Application of the general formalism of the joint posterior probability for all of the nonlinear parameters, including the decay rate constants, is given by Eq. [16]. For the two frequency model, Eq. [16] becomes

$$P(\omega_1,\omega_2,\alpha_1,\alpha_2|D_\sigma,D,I) \propto \left[1 - \frac{2\overline{h^2}(\omega_1,\omega_2,\alpha_1,\alpha_2)}{(N_\sigma\overline{d_\sigma^2} + N\overline{d^2})}\right]^{2-N-N_\sigma}. \tag{49}$$

The sufficient statistic $\overline{h^2}$ is given by

$$\overline{h^2}(\omega_1,\omega_2,\alpha_1,\alpha_2) = \frac{1}{4}[\hat{B}_1 R(\omega_1,\alpha_1) + \hat{B}_2 I(\omega_1,\alpha_1) + \hat{B}_3 R(\omega_2,\alpha_2) + \hat{B}_4 I(\omega_2,\alpha_2)], \tag{50}$$

where

$$\begin{aligned} \hat{B}_1 &= \frac{-E_2 R(\omega_1,\alpha_1) + C(\Delta,\beta)R(\omega_2,\alpha_2) + S(\Delta,\beta)I(\omega_2,\alpha_2)}{C(\Delta,\beta)^2 + S(\Delta,\beta)^2 - E_1E_2}, \\ \hat{B}_2 &= \frac{-E_2 I(\omega_1,\alpha_1) - S(\Delta,\beta)R(\omega_2,\alpha_2) + C(\Delta,\beta)I(\omega_2,\alpha_2)}{C(\Delta,\beta)^2 + S(\Delta,\beta)^2 - E_1E_2}, \\ \hat{B}_3 &= \frac{C(\Delta,\beta)R(\omega_1,\alpha_1) - S(\Delta,\beta)I(\omega_2,\alpha_2) - E_1 R(\omega_2,\alpha_2)}{C(\Delta,\beta)^2 + S(\Delta,\beta)^2 - E_1E_2}, \\ \hat{B}_4 &= \frac{S(\Delta,\beta)R(\omega_1,\alpha_1) + C(\Delta,\beta)I(\omega_2,\alpha_2) - E_1 I(\omega_2,\alpha_2)}{C(\Delta,\beta)^2 + S(\Delta,\beta)^2 - E_1E_2}, \end{aligned} \tag{51}$$

and $\Delta \equiv \omega_1 - \omega_2$, $\beta \equiv \alpha_1 + \alpha_2$, $E_1 \equiv C(0, 2\alpha_1) + 1$, and $E_2 \equiv C(0, 2\alpha_2) + 1$.

In this example the parameter ϵ was set to one. The parameter ϵ represents what was known about the uncertainty in the amplitudes before the data were obtained. In setting the value of ϵ, the only requirement is that the prior information be uninformative compared to the information content in the FID. This implies that ϵ must be small compared to the diagonal elements of the g_{jl}

matrix (Eq. [24]). The diagonal elements of this matrix are given approximately by

$$g_{11} = g_{22} = \sum_{i=0}^{N-1} \mathrm{e}^{-2\alpha_1 i} \approx \int_0^N \mathrm{e}^{-2\alpha_1 x} dx = \frac{1}{2\alpha_1} \tag{52}$$

and

$$g_{33} = g_{44} = \sum_{i=0}^{N-1} \mathrm{e}^{-2\alpha_2 i} \approx \int_0^N \mathrm{e}^{-2\alpha_2 x} dx = \frac{1}{2\alpha_2}. \tag{53}$$

If there are $N = 512$ data values per channel, and if the signal is sampled to $3T_2^*$, then in dimensionless units

$$\alpha_1 \approx \alpha_2 \approx \frac{3}{512} \approx 0.00586 \tag{54}$$

and the diagonal elements will be on the order of 85. Any number small compared to 85 will work; provided one does not choose a value that is computationally insignificant.

The posterior probability for the frequencies independent of the decay rate constants is the desired quantity. When removing the decay rate constants one must sum over each of the decay rate constants. If one designates these discrete values as α_{1j} and α_{2j}, and the number of discrete values as max, then the posterior probability for the frequencies independent of the amplitudes, phases, variance of the noise, and decay rate constants is given by

$$P(\omega_1, \omega_2 | D_\sigma, D, I) \propto \sum_{j=1}^{max} \sum_{k=1}^{max} \left[1 - \frac{2\overline{h^2}(\omega_1, \omega_2, \alpha_{1j}, \alpha_{2k})}{N_\sigma \overline{d_\sigma^2} + N\overline{d^2}} \right]^{2-N-N_\sigma}, \tag{55}$$

where a uniform prior probability was used for the decay rate constant for the same reasons discussed earlier.

Note that this statistic is more complex than the statistic for a single exponentially decaying frequency, but it may still be computed from a series of zero-padded discrete Fourier transforms of the complex FID data. As in the previous example, the series of 21 transforms mentioned earlier were used and the same noise sample was considered: $N_\sigma = 100,000$ and $\overline{d_\sigma^2} = 0.99961$. This posterior probability density function is plotted as a contour plot in Fig. 5. This plot is of the base 10 logarithm of the posterior probability for the frequencies of oscillation of two exponentially decaying sinusoids (Eq. [55]). The posterior probability for two frequencies has a line of symmetry along $\omega_1 = \omega_2$. If the posterior probability increases as you move away from this line, it indicates evidence in favor of two nonequal frequencies. In this plot the evidence increases by approximately 100 orders of magnitude, indicating strong evidence for the two frequencies. The closely spaced contours inside the contour labeled 1000 are dropping by 1 from the maximum. The region inside of the second small circle contains 99% of the total posterior probability. If the probability density function had been plotted, instead of its logarithm, essentially all of the posterior probability would be concentrated inside of these two circles.

There are two nuisance decay rate constants to be removed, and one would expect this routine to run fairly slowly; indeed, this is the case, requiring 18 CPU minutes on a VAXstation 3100 M38. But to put this number in proper perspective, consider that this was run on a 91×91 grid, and an older version of this program that was used previously took 2 CPU hours on the same machine. Using a smaller grid around the peak would reduce the computation time significantly. For example, using a 25×25 grid around the maximum would reduce the total computation time to about a minute and a half.

The small tail in the region of the third contour down from the maximum is caused by the coarseness of the decay rate discretization: taking smaller steps and doing more discrete Fourier transforms will remove this computational artifact, but at the cost of increased computing time.

Last, note that this statistic could be run on all of the peaks in the discrete Fourier transform. If this were done and the peak is an isolated frequency (like the one at -2 rad) then this statistic

Figure 5: The Probability For Two Close Frequencies

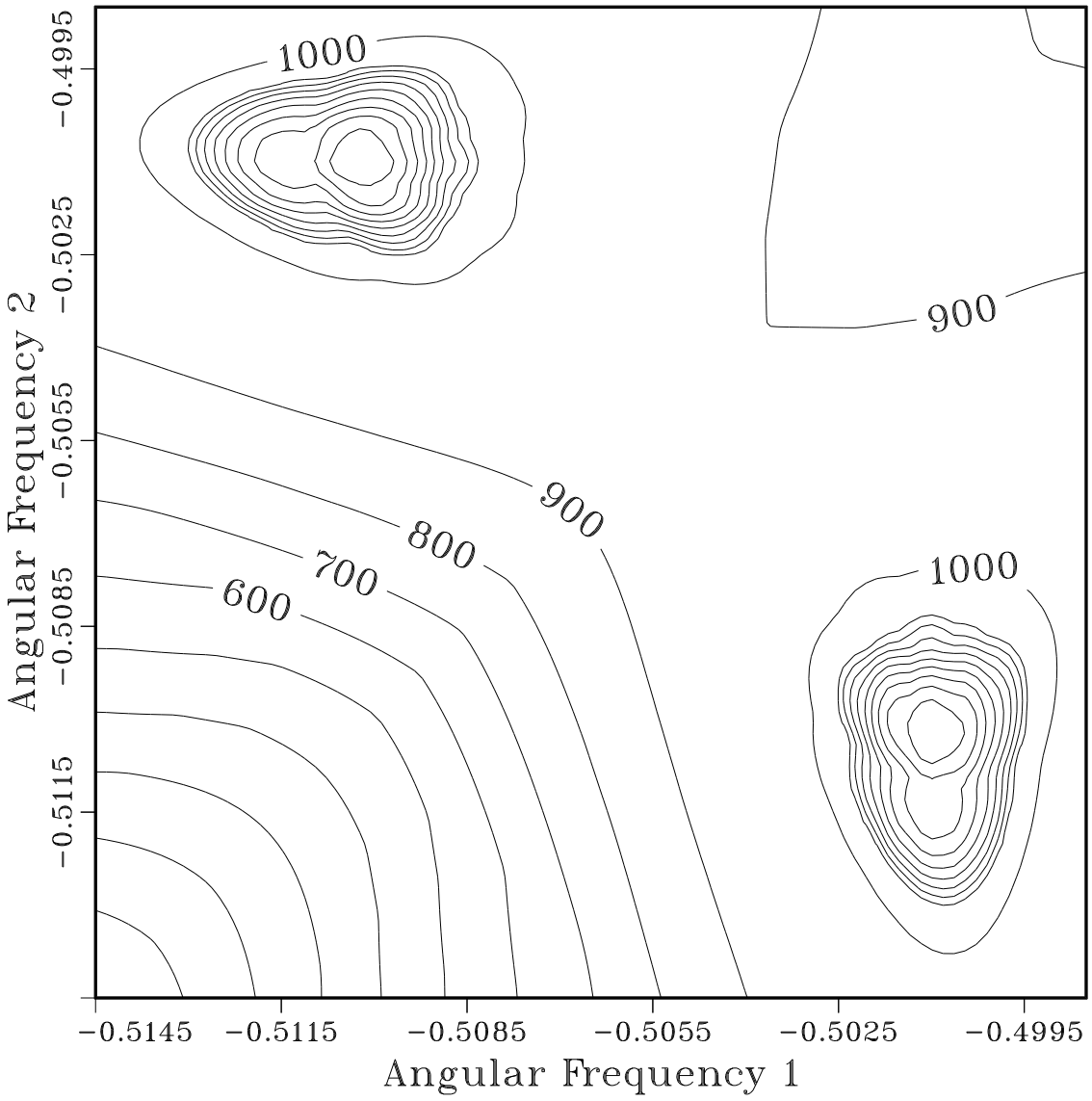

Fig. 5. Plot of the base 10 logarithm of the posterior probability for two frequencies independent of the decay rate constants, phases, amplitudes, and variance of the noise. The closely spaced contours inside of the contours labeled 1000 are decreasing by 1 from the maximum. The region inside the second circle contains 99% of the total posterior probability, indicating that the two frequencies have been accurately determined.

will not have a true maximum; rather there will be two ridges at right angles to each other. These ridges will intersect at $\omega_1 = -2$ and $\omega_2 = -2$, and there will not be a maximum at this location; rather the intersection of these ridges is more like two walls intersecting. This type of behavior is understandable in terms of the symmetries in the model. There is only one frequency in the data; but the model insists that there are two, probability theory is simply saying one or both of the frequencies is -2. On the other hand, the region around 0.3 rad contains a triplet and one would expect this statistic to have peaks indicating the presence of these frequencies provided they are not so close together that they cannot be resolved by this two-frequency model.

Example 3: Estimating The Center Frequency And J Coupling Constant Of A Triplet

In this third example, the posterior probability for an arbitrary multiplet, independent of the amplitude, phase, variance of the noise, and decay rate constant, is computed and then specialized to the triplet model. When discussing the triplet model the sufficient statistic for a 1:2:1 triplet is derived and applied in the example. In addition the sufficient statistic for a 1:1:1 triplet and for a triplet of doublets is derived to illustrate how to include more complex prior information in the calculation.

What distinguishes a multiplet from an arbitrary set of lines is that some type of prior information about the structure of the lines is known. It is assumed that this prior information includes the weights associated with a frequency, how the frequencies are spaced (for example, uniformly at some unknown interval), whether or not the frequencies in the multiplet are decaying with the same exponential decay rate constants, and last, whether the sinusoids all have the same phase. The model signal for a multiplet that incorporates all of this prior information is given by

$$f_R(t) = B_1 U_1(t) + B_2 U_2(t) \tag{56}$$

for the real channel and

$$f_I(t) = B_1 V_1(t) + B_2 V_2(t) \tag{57}$$

for the imaginary channel, where the amplitudes B_1 and B_2 express the phase coherence prior information.

The signal functions U_1, U_2, V_1, and V_2 contain the prior information about the frequencies, the weights, and the decay rate constants. The signal functions, for the real channel, are written as

$$\begin{aligned} U_1(t_i) &= \sum_{j=1}^{n} \mathrm{w}_j \cos(\omega_j t_i) \mathrm{e}^{-\alpha_j t_i}, \\ U_2(t_i) &= \sum_{j=1}^{n} \mathrm{w}_j \sin(\omega_j t_i) \mathrm{e}^{-\alpha_j t_i}, \end{aligned} \tag{58}$$

and for the imaginary channel

$$\begin{aligned} V_1(t_i) &= -\sum_{j=1}^{n} \mathrm{w}_j \sin(\omega_j t_i) \mathrm{e}^{-\alpha_j t_i}, \\ V_2(t_i) &= \sum_{j=1}^{n} \mathrm{w}_j \cos(\omega_j t_i) \mathrm{e}^{-\alpha_j t_i}, \end{aligned} \tag{59}$$

where the weights w_j are assumed to be known from prior information. Because the weights are the same for the two signal functions, the first subscript from the general formalism has been dropped: $\mathrm{w}_{1j} = \mathrm{w}_{2j} \equiv \mathrm{w}_j$. The number of signal components, n_j, is the same in both signal functions, $n_1 = n_2 \equiv n$. This number is the order of the multiplet (e.g., singlet, doublet, triplet, etc.).

Some prior information about the frequencies is assumed to be known. For example, in the case of a triplet in the loose coupling regime, the frequencies would be equally spaced, so the three frequencies could be expressed in terms of two new parameters; the center frequency of the multiplet, ω_c, and the J coupling constant, Δ. For multiplets with uniformly spaced frequencies, all of the frequencies may be written in terms of the center frequency and the coupling constant, then the jth frequency, ω_j, is given by

$$\omega_j = \omega_c - \frac{n+1-2j}{2}\Delta. \tag{60}$$

The formalism does not require the spacing to be uniform, only that some information about the spacing be known. Later an example where the frequencies are not uniformly spaced is given.

In addition to prior information about the weights and the frequencies, one often knows whether the electronic environment of the spins is the same – so it is often known that the decay rate constants will be the same on the various peaks. It is not required that one know this, only that if it is known, it should be put into the model and the number of nonlinear parameters should be reduced.

To derive the sufficient statistic for an arbitrary multiplet, first the g_{jl} matrix is derived. The g_{jl} matrix for this multiplet model is diagonal and given by

$$g_{jl} = \begin{pmatrix} \sum_{k=1}^{n}\sum_{m=1}^{n} \mathrm{w}_k \mathrm{w}_m C(\omega_k - \omega_m, \alpha_k + \alpha_m) & 0 \\ 0 & \sum_{k=1}^{n}\sum_{m=1}^{n} \mathrm{w}_k \mathrm{w}_m C(\omega_k - \omega_m, \alpha_k + \alpha_m) \end{pmatrix}. \tag{61}$$

The sufficient statistic is then given by

$$\overline{h^2}(\omega_1, \ldots, \omega_n, \alpha_1, \ldots, \alpha_n) = \frac{\left[\sum_{j=1}^{n} \mathrm{w}_j R(\omega_j, \alpha_j)\right]^2 + \left[\sum_{j=1}^{n} \mathrm{w}_j I(\omega_j, \alpha_j)\right]^2}{2\sum_{j=1}^{n}\sum_{k=1}^{n} \mathrm{w}_j \mathrm{w}_k C(\omega_j - \omega_k, \alpha_j + \alpha_k)}, \tag{62}$$

where w_j is the weight of each "peak" in the multiplet, and $R(\omega_j, \alpha_j)$ and $I(\omega_j, \alpha_j)$ are the real and imaginary parts of the discrete Fourier transform when the data have been multiplied by a decaying exponential.

The joint posterior probability for the frequencies and the decay rate constants is then given by

$$P(\omega_1, \ldots, \omega_n, \alpha_1, \ldots, \alpha_n | D_\sigma, D, I) \propto \left[1 - \frac{\overline{h^2}(\omega_1, \ldots, \omega_n, \alpha_1, \ldots, \alpha_n)}{N\overline{d^2} + N_\sigma \overline{d_\sigma^2}}\right]^{1-N-N_\sigma}. \tag{63}$$

To remove the dependence on the decay rate constants one must multiply by the appropriate prior probabilities and marginalize them from the probability density function. However, in many multiplets the electrons all see the same electronic environment, so the spins in the multiplet all decay at the same exponential rate. Making this assumption reduces the number of decay rate constants from n to one and the posterior probability for the frequencies independent of the decay rate is given by

$$P(\omega_1, \ldots, \omega_n | D_\sigma, D, I) \propto \sum_{j=1}^{max} \left[1 - \frac{\overline{h^2}(\omega_1, \ldots, \omega_n, \alpha_j)}{N\overline{d^2} + N_\sigma \overline{d_\sigma^2}}\right]^{1-N-N_\sigma}. \tag{64}$$

This assumption does not have to be made, but then one must sum over each decay rate constant that is to be removed from the joint posterior probability density function (Eq. [64]).

Table 1: $\overline{h^2}$ for Multiplets

n	$2\overline{h^2}$
1	$\frac{R_1^2 + I_1^2}{C(0,2\alpha)}$
2	$\frac{(R_1+R_2)^2 + (I_1+I_2)^2}{2C(0,2\alpha)+2C(\Delta,2\alpha)}$
3	$\frac{(R_1+2R_2+R_3)^2 + (I_1+2I_2+I_3)^2}{6C(0,2\alpha)+8C(\Delta,2\alpha)+2C(2\Delta,2\alpha)}$
4	$\frac{(R_1+3R_2+3R_3+R_4)^2 + (I_1+3I_2+3I_3+I_4)^2}{20C(0,2\alpha)+30C(\Delta,2\alpha)+12C(2\Delta,\alpha)+2C(3\Delta,2\alpha)}$
5	$\frac{(R_1+4R_2+6R_3+4R_4+R_5)^2 + (I_1+4I_2+6I_3+4I_4+I_5)^2}{70C(0,2\alpha)+112C(\Delta,2\alpha)+56C(2\Delta,2\alpha)+16C(3\Delta,2\alpha)+2C(4\Delta,2\alpha)}$

The sufficient statistics for the first five spin-$\frac{1}{2}$ multiplets are shown. The notation $R_j \equiv R(\omega_j, \alpha)$ and $I_j \equiv I(\omega_j, \alpha)$ has been used, where ω_j is given by Eq. [60]. Note that for $n = 1$, the sufficient statistic is essentially a power spectrum of the complex FID data multiplied by a decaying exponential. As n increases the power spectrum is generalized to a more complex statistic, which includes all of the prior information about the multiplet structure.

For spin-$\frac{1}{2}$ systems, in the weak coupling regime, the weights are just the numbers from Pascal's triangle. The sufficient statistics for $n = 1$ through $n = 5$ are shown in Table 1. Note that for $n = 1$, the sufficient statistic is essentially a discrete Fourier transform *power spectrum* when the data have been multiplied by a decaying exponential. As n increases, probability theory essentially generalizes the power spectrum to a new statistic that looks for more complex structures in the data.

In the simulated data given earlier (Fig. 1) the region around 0.3 rad is a triplet, and in this example the base 10 logarithm of the joint posterior probability for the center frequency, ω_c, and the J coupling constant, Δ, is computed independent of the amplitude, phase, variance of the noise, and decay rate constant. The triplet is taken in the weak coupling regime. The model equations for a spin-$\frac{1}{2}$ triplet were given earlier (Eqs. [11] and [12]). By substituting these model equations into the sufficient statistic (Eq. [62]) and substituting the sufficient statistic into the posterior probability for the center frequency and the coupling constant (Eq. [64]), one obtains

$$P(\omega_c, \Delta | D_\sigma, D, I) \propto \sum_{j=1}^{max} \left[1 - \frac{\overline{h^2}(\omega_c, \Delta, \alpha_j)}{N\overline{d^2} + N_\sigma \overline{d_\sigma^2}} \right]^{1-N-N_\sigma}, \tag{65}$$

where

$$\overline{h^2}(\omega_c, \Delta, \alpha_j) \equiv \frac{1}{2} \left[\frac{T_1(\omega_c, \Delta, \alpha_j)^2 + T_2(\omega_c, \Delta, \alpha_j)^2}{6C(0,2\alpha_j) + 8C(\Delta, 2\alpha_j) + 2C(2\Delta, 2\alpha_j)} \right], \tag{66}$$

and

$$\begin{aligned} T_1(\omega_c, \Delta, \alpha_j) &= R(\omega_c - \Delta, \alpha_j) + 2R(\omega_c, \alpha_j) + R(\omega_c + \Delta, \alpha_j), \\ T_2(\omega_c, \Delta, \alpha_j) &= I(\omega_c - \Delta, \alpha_j) + 2I(\omega_c, \alpha_j) + I(\omega_c + \Delta, \alpha_j). \end{aligned} \tag{67}$$

This probability density function incorporates everything that is known about an isolated spin-$\frac{1}{2}$ triplet. Because the sufficient statistic, $\overline{h^2}$, involves six terms from the discrete Fourier transform,

it will be much more sharply peaked than the power spectrum. The corresponding probability density function will be able to resolve structures, even when the discrete Fourier transform shows no evidence for them.

As with all multiplets, this probability density function can be computed from a zero-padded discrete Fourier transform of the complex FID data. In this example, the probability density function was computed from the 21 discrete Fourier transforms done initially. As with the other examples $N_\sigma = 100,000$ and $\overline{d_\sigma^2} = 0.99961$ were used. Because the joint posterior probability density function estimates the value of two parameters (the center frequency, ω_c, and the J coupling constant, Δ) the base 10 logarithm is displayed as a contour plot (Fig. 6). There are several noteworthy features of this plot.

First, the linear-shaped series of very closely spaced contours in the upper right-hand corner is an area where this probability density function drops almost 6,000 orders of magnitude. Above this line there are no parameter values consistent with a triplet model. Below this line the contours of log probability still increase rapidly, but this increase is very slow compared to that of the region in the upper right-hand corner.

Second, the jagged appearance of some of the contour lines is an artifact of the digital resolution implicit in a zero-padded discrete Fourier transform. These artifacts could be removed, but at the cost of zero-padding the discrete Fourier transform so much that it would be faster to do the slow transform over the regions needed. This illustrates one of the major drawbacks of these multiplet statistics. In typical data, the higher the multiplet order, the more sharply peaked the statistic; for a triplet, the statistic already concentrates most of the posterior probability on a single frequency in the discrete Fourier transform. There were $N = 512$ data values per channel and the FID was zero-padded to 32K. As the multiplet order increases, the problem becomes worse, and these statistics can peak so sharply that one may have trouble locating the maximum.

Third, the maximum of this probability density function is located inside the closed, closely spaced contours just above the center frequency value of 0.3 rad. These tightly-spaced contours are dropping by one from the maximum. If the posterior probability density had been plotted instead of its log, 99% of the total posterior probability density would have been contained within the second contour. So an accurate determination of the center frequency of the multiplet and the J coupling constant has been made, even though the discrete Fourier transform absorption spectrum shows no visible evidence for the multiplet structures.

The discussion so far has proceeded as if the equations for an arbitrary multiple were applicable only to spin-$\frac{1}{2}$ systems. But the formalism is completely general and can be applied to any multiplet. Here the sufficient statistics for a 1:1:1 triplet and a triplet of doublets are derived. In the case of a 1:1:1 triplet all of the weights are equal. By applying Eq. [62] one obtains

$$\overline{h^2}(\omega_c, \Delta, \alpha) \equiv \frac{[R(\omega_1,\alpha) + R(\omega_2,\alpha) + R(\omega_3,\alpha)]^2 + [I(\omega_1,\alpha) + I(\omega_2,\alpha) + I(\omega_3,\alpha)]^2}{12C(0,2\alpha) + 8C(\Delta,2\alpha) + 4C(2\Delta,2\alpha)} \tag{68}$$

as the sufficient statistic, where the frequencies are given by Eq. [60]. This statistic, when used in the posterior probability for the frequencies in a multiplet, will give the best estimate of the center frequency and coupling constant possible for a 1:1:1 triplet.

But even more complex multiplets occur in high resolution NMR. Often one sees multiplets of multiplets. There are so many different possibilities that no attempt was made to do these in general. Instead the sufficient statistic for a triplet of doublets is derived to illustrate how to apply the formalism. For this multiplet there are two coupling constants, Δ_T denotes the J coupling constant for the triplet and Δ_D for the doublet. A triplet of doublets has weights given by $(1,1,2,2,1,1)$ and the frequencies ω_1 through ω_6 are given by $\omega_1 = \omega_c - \Delta_T - \Delta_D$, $\omega_2 = \omega_c - \Delta_T + \Delta_D$, $\omega_3 = \omega_c - \Delta_D$, $\omega_4 = \omega_c + \Delta_D$, $\omega_5 = \omega_c + \Delta_T - \Delta_D$, $\omega_6 = \omega_c + \Delta_T + \Delta_D$, respectively. By substituting this information into the equation for the sufficient statistic, Eq. [66], one obtains

$$\overline{h^2}(\omega_c, \Delta_D, \Delta_T, \alpha) = \frac{T_1^2 + T_2^2}{2Z}, \tag{69}$$

Figure 6: The Joint Probability for the Center Position of a Triplet, and the J Coupling Constant

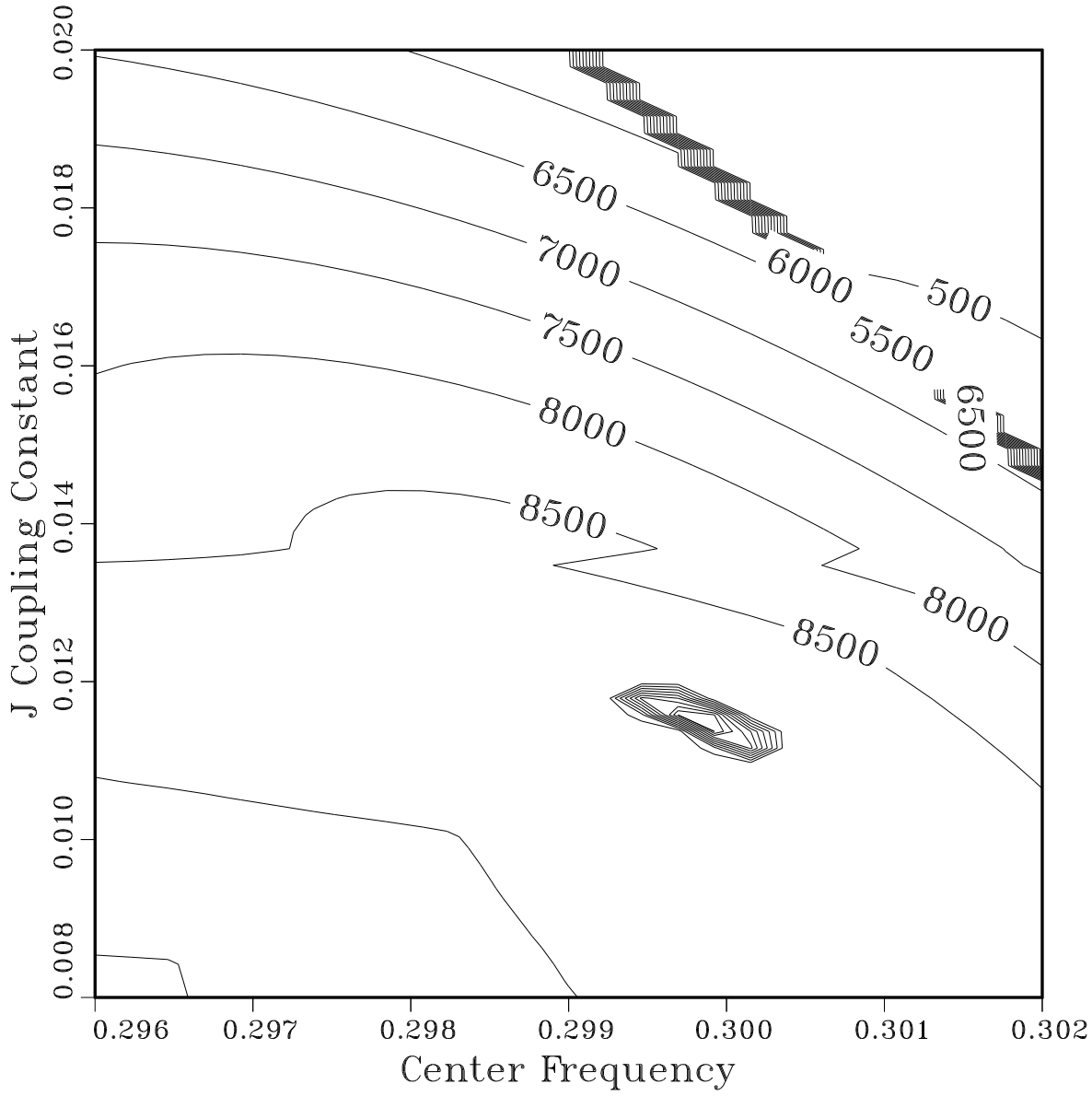

Fig. 6. Plot of base 10 logarithm of the joint posterior probability for the center position of a triplet and the J coupling constant independent of the decay rate constants, phases, amplitudes, and variance of the noise. The unlabeled, closed, and closely spaced contours are decreasing by 1 from the maximum. The region inside the second contour contains 99% of the total posterior probability. Indicating that both the center frequency, ω_c, and the J coupling constant, Δ, have been accurately determined.

where

$$\begin{aligned} T_1 &= R(\omega_c - \Delta_T - \Delta_D, \alpha) + R(\omega_c - \Delta_T + \Delta_D, \alpha) + 2R(\omega_c - \Delta_D, \alpha) \\ &+ 2R(\omega_c + \Delta_D, \alpha) + R(\omega_c + \Delta_T - \Delta_D, \alpha) + R(\omega_c + \Delta_T + \Delta_D, \alpha) \end{aligned} \tag{70}$$

$$\begin{aligned} T_2 &= I(\omega_c - \Delta_T - \Delta_D, \alpha) + I(\omega_c - \Delta_T + \Delta_D, \alpha) + 2I(\omega_c - \Delta_D, \alpha) \\ &+ 2I(\omega_c + \Delta_D, \alpha) + I(\omega_c + \Delta_T - \Delta_D, \alpha) + I(\omega_c + \Delta_T + \Delta_D, \alpha) \end{aligned} \tag{71}$$

and the normalization constant Z is given by

$$\begin{aligned} Z &= 2C(2\Delta_T + 2\Delta_D, 2\alpha) + 2C(2\Delta_T - 2\Delta_D, 2\alpha) + 8C(\Delta_T + 2\Delta_D, 2\alpha) \\ &+ 8C(\Delta_T - 2\Delta_D, 2\alpha) + 4C(2\Delta_T, 2\alpha) + 16C(\Delta_T, 2\alpha) \\ &+ 12C(2\Delta_D, 2\alpha) + 12C(0, 2\alpha). \end{aligned} \tag{72}$$

While this statistic is more complicated, it will give the best estimate of the two coupling constants and the center frequency possible from the data and one's prior information. The only real difficulty is that there are now three parameters to be estimated: the center frequency ω_c, the triplet coupling constant Δ_T, and the coupling constant for the doublet Δ_D. This increases the computational complexity and costs time. Furthermore, it makes presenting the results difficult. The best way to proceed would be to generate the posterior probability as a function of all three parameters and then use the sum and product rules from probability theory to eliminate any two of them. This would give three probability density functions to view: one for each of the three parameters.

Summary and Conclusions

The calculations presented here generalized the parameter estimation calculation [1, 8] to include prior information about the noise. Using examples, the effect of obtaining information about the noise was demonstrated and it was shown that obtaining information about the noise is almost as important as obtaining information about the signal. This should not be taken to imply that one must always obtain information about the noise; rather one should determine the noise level of the spectrometer and then use that information in the analysis. This is fairly trivial to do in an NMR experiment, where accurate noise files can be taken readily as part of most experimental protocols.

The three examples given would have been much more difficult had information about the noise not been available. To perform them without noise prior information, one must account for all systematic effects in the data. These systematic effects contain parameters which would have to be removed as nuisances. There are so many of these parameters that the integrals over these parameters would not have been tractable.

An alternate way to proceed would be to test various possible models. This is possible using model-selection procedures [2, 8]. Indeed those procedures and the ones described here are complementary. The procedures described here are parameter estimation procedures, and as such, they are very useful in exploratory work and in estimating the parameters when the true model is known. But they will answer specific questions. These questions are always of the form "what is the best estimate of the parameters that may be obtained, given the functional form of the model?" They never question the functional form of the model. The model selection procedures developed [2, 8] can be used to test possible alternatives and derive the best model. The procedures given here will aid in choosing what models should be tested, and often in conjunction with good common sense will serve adequately by themselves.

An alternate formulation of the posterior probability was given. This formulation is computationally more stable than the singular valued decomposition given in Ref. [1] and can be computed from a series of discrete Fourier transforms of the complex data when the data have been multiplied by a decaying exponential. Use of the discrete Fourier transform in computing these statistics makes it possible to eliminate the decay rate constants from the probability density functions. This allows one to examine what probability theory has to say about the frequencies alone.

Finally, by using these calculations and information about the noise it will be possible to implement many important Bayesian calculations on most existing spectrometers. These new tools should include calculations for the amplitudes, frequencies, decay rate constants, and J coupling constants. Some of these statistics will be suitable for real-time processing and some will not. These tools should enable experimenters to resolve frequencies, amplitudes, decay rate constants, and coupling constants to an unprecedented accuracy. But with these new tools comes new responsibilities. Probability theory will answer the questions asked of it, and it will do so optimally, but it will not volunteer information; you must ask, and you must understand what question is being asked.

Acknowledgments

The encouragement of Professor J. J. H. Ackerman is greatly appreciated, as are extensive conversations with Professor E. T. Jaynes. This work was partially supported by a gift from the Monsanto Company and NIH grant GM30331, of which J. J. H. Ackerman principal investigator.

References

[1] Bretthorst, G. Larry, "Bayesian Analysis. I. Parameter Estimation Using Quadrature NMR Models," *J. Magn. Reson.* **88**, pp. 533-551 (1990).

[2] Bretthorst, G. Larry, "Bayesian Analysis. II. Model Selection," *J. Magn. Reson.* **88**, pp. 552-570 (1990).

[3] Bretthorst, G. Larry, "Bayesian Analysis. III. Applications to NMR Signal Detection, Model Selection and Parameter Estimation," *J. Magn. Reson.* **88**, pp. 571-595 (1990).

[4] Jeffreys, H., "Theory of Probability," Oxford Univ. Press, London, 1939; Later editions, 1948, 1961.

[5] Jaynes, E. T., in "Papers on Probability, Statistics and Statistical Physics" (R. D. Rosenkrantz *Ed.*), Reidel, Dordrecht, The Netherlands, 1983.

[6] Jaynes, E. T., *Bayesian Spectrum and Chirp Analysis,* in "Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems" (C. R. Smith and G. J. Erickson *Eds.*), p. 1, Reidel, Dordrecht, The Netherlands, 1987.

[7] Bretthorst, G. Larry, *An Introduction to Parameter Estimation Using Bayesian Probability Theory,* in "Maximum Entropy and Bayesian Methods" (P. Fougere *Ed.*), p. 53, Kluwer Academic, Dordrecht, The Netherlands, 1989.

[8] Bretthorst, G. Larry, *Bayesian Spectrum Analysis and Parameter Estimation,* in "Lecture Notes in Statistics" Vol. **48,** Springer-Verlag, New York, 1988.

[9] Gull, S. F., *Bayesian Inductive Inference and Maximum Entropy,* in "Maximum Entropy and Bayesian Methods in Science and Engineering" (G. J. Erickson and C. R. Smith *Eds.*), Vol. **1,** pp. 53-75, Kluwer Academic, Dordrecht, The Netherlands, 1988.

[10] Jaynes, E. T., *How Does the Brain do Plausible Reasoning?* unpublished Stanford University Microwave Laboratory Report No. 421 (1957); reprinted in "Maximum-Entropy and Bayesian Methods in Science and Engineering" (G. J. Erickson and C. R. Smith *Eds.*), Vol. **1,** pp. 1-24, Kluwer Academic, Dordrecht, The Netherlands, 1988.

[11] Tribus, M., "Rational Descriptions, Decisions and Designs," Pergamon Press, Oxford, 1969.

[12] Zellner, A., "An Introduction to Bayesian Inference in Econometrics," Wiley, New York, 1971.

[13] Jaynes, E. T., "Probability Theory – The Logic of Science," in preparation. Copies of the completed parts of this manuscript are available from E. T. Jaynes, Washington University, Dept. of Physics, St. Louis, MO 63130.

[14] Bretthorst, G. Larry, "^{31}P NMR Bayesian Spectral Analysis of Rat Brain *in Vivo*" *Magn. Reson. Med.* **9**, pp. 282-287 (1989).

[15] Bretthorst, G. Larry, "Bayesian Analysis of Time-Domain Magnetic Resonance Signals," *J. Magn. Reson.* **79**, pp. 369-376 (1988).

[16] Press, William H., Brian P. Flannery, Saul A. Teukolsky, AND William T. Vettering, *Numerical Recipes: The Art of Scientific Computing,* Cambridge Univ. Press, Cambridge 1986.