Published in final edited form as:

Appl Magn Reson. 2010 January 1; 37(1-4): 865–880. doi:10.1007/s00723-009-0102-7.

Magnetic Resonance Spectra and Statistical Geometry

Keith A. Earle¹, Laxman Mainali¹, Indra Dev Sahu¹, and David J. Schneider³

¹ Physics Department, University at Albany, 1400 Washington Ave, Albany NY 12222

³ USDA Agricultural Research Service and Department of Plant Pathology, Cornell University, Ithaca NY 14853

Abstract

Methods of statistical geometry are introduced which allow one to estimate, on the basis of computable criteria, the conditions under which maximally informative data may be collected. We note the important role of constraints which introduce curvature into parameter space and discuss the appropriate mathematical tools for treating curvature effects. Channel capacity, a term from communication theory, is suggested as a useful figure of merit for estimating the information content of spectra in the presence of noise. The tools introduced here are applied to the case of a model nitroxide system as a concrete example, but we stress that the methods described here are of general utility.

1 Introduction

When designing an experiment to test a model, it is useful to have a criterion for determining what experimental conditions will yield the most informative data. Then, once data are in hand, a reliable means of analysis is necessary to extract optimum sets of model parameters on the basis of some fitting criterion. In automated parameter fitting, a computable means of determining an appropriate direction and step size in parameter space is needed. The purpose of this paper is to draw attention to a class of methods and concepts that can inform each stage of the process from experimental design to parameter estimation. These methods go by a number of names, but they are all based on applying geometrical concepts to problems in statistical inference. Statistical geometry is a useful way to characterize these methods.[1,2]

In most forms of spectroscopy, measuring the linear response of some system to an external perturbation is still the most common experimental arrangement.[3] From the perspective of communication theory, the desired signal, in this case the linear response absorption, is sent down a noisy channel, and the original signal must be reconstructed from the received message which is partially ambiguous due to noise in the channel.[4] A model chosen for fitting purposes should be as faithful as possible to the known characteristics of the original signal. For example, the linear response of a system is often described in terms of a susceptibility. On the basis of very general considerations, susceptibilities are known to satisfy various constraints.[5] Any fitting model that is chosen should have these *a priori* constraints built-in, otherwise the returned fit parameters will be sub-optimal. We note that the approach given here is an attempt to combine insights and methods from Statistics, Differential Geometry, and Communication Theory, informed by the practical experience of working spectroscopists. Given the disparate threads that wind through this work, we provide the following road map.

We begin with a standard fitting model for a Gaussian lineshape in Section 2 and note how failure to properly account for constraints can lead to inefficiencies and inaccuracies in parameter estimation. We then broaden the scope of the sorts of lineshapes that may be treated in Section 3. The constraints imposed by susceptibility sum rules introduce a notion of curvature in parameter space which we explore in Section 4. Differential entropy is introduced in Section 5 as a way of summarizing the information content of an entire spectrum. The expected signal strength is evaluated under quite general conditions in Section 6. Section 7 provides a general method for handling finite signal to noise ratios based on channel capacity. Finally, Section 8 summarizes the main conclusions and suggests lines of inquiry for subsequent work.

2 Rational Parameter Selection and Estimation

Consider a simple case of a lineshape that is known to be Gaussian on the basis of prior information, such as previous experience with similar systems. In general there will be extrinsic properties of the system such as the signal to noise ratio that depend on concentration or other external conditions, and properties that may more properly be considered as intrinsic to the model, such as the mean value and the linewidth. In this work we will be primarily concerned with intrinsic quantities. We discuss the important role of the signal to noise ratio in Section 7. In order to perform parameter estimation, an unnormalized trial function of the following form is chosen all too often

$$\hat{p}(\omega, \theta) = a \exp(-(\omega - \mu)^2 / 2\sigma^2), \quad (1)$$

where θ is the vector of parameters (a, μ, σ) and the hat on $\hat{p}(\omega, \theta)$ indicates that the trial function is not normalized. This function can be scaled to match the peak height, and the mean μ and variance σ may be progressively optimized to yield an optimum set of parameters. The model in Equation 1 is in fact deficient because it does not build in the constraint that the integral of the susceptibility over all frequencies is invariant to changes in the mean or the variance, i.e.,

$$\int d\omega p(\omega, \theta) = 1 \quad (2)$$

We shall call the invariance property specified by Equation 2 the area preserving or normalization constraint.

In order to quantify the change in lineshape as parameters are varied, it is useful to specify the relative change in p as the parameter vector θ is varied. In statistics, this relative change is known as the score

$$\ell_i(\omega, \theta) \equiv \frac{1}{p(\omega, \theta)} \frac{\partial p(\omega, \theta)}{\partial \theta^i}. \quad (3)$$

The score may also be written as the logarithmic derivative of $p(\omega, \theta)$.

$$\ell_i(\omega, \theta) \equiv \frac{\partial}{\partial \theta^i} \ln(p(\omega, \theta)) \quad (4)$$

We note that the position of the index i is not arbitrary but indicates how the indexed object behaves under coordinate transformations. For example,

$$\bar{\theta}^j = \frac{\partial \bar{\theta}^j}{\partial \theta^i} \theta^i \quad (5)$$

is an example of a contravariant quantity while

$$\frac{\partial}{\partial \bar{\theta}^j} = \left(\frac{\partial \bar{\theta}^j}{\partial \theta^i} \right)^{-1} \frac{\partial}{\partial \theta^i} \quad (6)$$

is an example of a covariant quantity. The quantity $\partial \bar{\theta}^j / \partial \theta^i$ is known as the Jacobian of the transformation. It characterizes smooth coordinate or parameter transformations and is an important ingredient for defining a notion of distance, as discussed in Section 4. For a more thorough discussion of covariance and contravariance, Kühnel provides a readable, contemporary account emphasizing the invariance properties of smooth coordinate or parameter transformations.[6] A more traditional account which may be more accessible to readers with conventional training in physics or chemistry is given by Synge and Schild.[7]

For this problem, the scores are simple to compute:[2,8]

$$\ell = \ln a - \frac{(\omega - \mu)^2}{2\sigma^2} \quad (7)$$

$$\ell_a = \frac{1}{a} \quad (8)$$

$$\ell_\mu = \frac{\omega - \mu}{\sigma^2} \quad (9)$$

$$\ell_\sigma = \frac{(\omega - \mu)^2}{\sigma^3} \quad (10)$$

It is important to note that the scores are only defined when the underlying function p is specified. Note that the scores provide local information about the sensitivity of the spectrum to changes in parameters. A useful way to summarize the sensitivity of the function p to parameter changes is to compute auto and cross correlations of the scores with

respect to the function p . See Section 4 for more details. This procedure defines a matrix which can be written as a two-index covariant quantity

$$I_{ij}(\theta) \equiv \int d\omega p(\omega, \theta) \ell_i(\omega, \theta) \ell_j(\omega, \theta) \quad (11)$$

When $p(\omega, \theta)$ is a probability distribution function, $I_{ij}(\theta)$ is known as the Fisher information matrix. It will play a fundamental role in what follows. For the trial function $\hat{p}$, the matrix in Equation 11 is proportional to the Fisher information matrix, accounting for the observation that we are using an unnormalized trial function.

Returning to our model, we can write down the matrix elements of the Fisher information matrix

$$\hat{I}_{ij}(\theta) = \sqrt{2\pi} a \sigma \begin{bmatrix} 1/a^2 & 0 & 1/a\sigma \\ 0 & 1/\sigma^2 & 0 \\ 1/a\sigma & 0 & 2/\sigma^2 \end{bmatrix}, \quad (12)$$

where the hat on $\hat{I}_{ij}(\theta)$ indicates that we are using the unnormalized trial function given in Equation 1. As expected, we see that a and σ are strongly correlated via the off-diagonal matrix elements of Equation 12. Anyone who has fit a Gaussian lineshape with a fitting function of the form given in Equation 1 will know from experience that fitting a , μ and σ simultaneously will lead to large correlations between a and σ as well as large uncertainties. This is due to the fact that the model of Equation 1 mixes intrinsic and extrinsic parameters. A better model would use the observation that the spectral lineshape satisfies the normalization criterion. For example

$$\bar{p}(\omega, \theta) \equiv \frac{a}{\sqrt{2\pi}\sigma^2} \exp(-(\omega - \mu)^2/2\sigma^2) \quad (13)$$

is a suitable model. The bar on $\bar{p}(\omega, \theta)$ indicates that this function is normalized according to Equation 2 when $a = 1$. Note that in this parameterization of the model that a and σ are uncorrelated (cf. Equation 18). For this model, the scores are

$$\ell = \ln a - \ln \sigma - \frac{(x - \mu)^2}{2\sigma^2} - (1/2) \ln(2\pi) \quad (14)$$

$$\ell_a = \frac{1}{a} \quad (15)$$

$$\ell_\mu = \frac{x - \mu}{\sigma^2} \quad (16)$$

$$\ell_{\sigma} = -\frac{1}{\sigma} + \frac{(x - \mu)^2}{\sigma^3} \quad (17)$$

An easy calculation gives the (unnormalized) Fisher information matrix

$$\bar{I}_{ij}(\theta) = a \begin{bmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/\sigma^2 & 0 \\ 0 & 0 & 2/\sigma^2 \end{bmatrix} \quad (18)$$

When $a = 1$, $\bar{I}_{ij}(\theta)$ is a properly normalized Fisher information matrix. We note that the parameters now have no intrinsic cross correlations arising from off-diagonal components of the Fisher information matrix. In the presence of noise, what might be called extrinsic correlations will arise due to spectral information that is obscured by transmission via a noisy channel. The effect of a finite signal to noise ratio is discussed in Section 7. The point is that proper incorporation of the constraints has allowed us to choose a functional form for fitting that is the optimum for obtaining reliable results.

3 An Exactly Solvable Model System

Given that not all physical processes produce Gaussian lineshapes, it is useful to have a model that can be used to describe wider classes of lineshapes. The randomly modulated oscillator is an example of an important theoretical model that produces lineshapes that vary smoothly between Gaussian and Lorentzian forms.[3] Although the randomly modulated oscillator can be solved exactly, the closed form expressions that one derives are not particularly illuminating. Student's t distribution with ν degrees of freedom is an example of a probability density function (PDF) that has the same functional form as a Gaussian or a Lorentzian line shape as $\nu \rightarrow \infty$ or $\nu = 1$ respectively and leads to tractable expressions for the scores for $0 < \nu < \infty$. [4,8] Student's t distribution has the following functional form which may interpreted as the susceptibility of the physical system

$$J_{\nu}(\omega) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} \left(1 + \frac{\omega^2}{\nu}\right)^{-(\nu+1)/2}, \quad (19)$$

for ν degrees of freedom. Some insight into the form of this equation may be obtained by considering its Fourier transform, which defines the correlation function.[9]

$$G_{\nu}(t) = \frac{2^n \sqrt{\nu}^{\nu}}{\Gamma(\nu)} \int_0^{\infty} \exp(-\sqrt{\nu}(2x+|t|))(x(x+|t|))^{(\nu-1)/2} dx \quad (20)$$

Note that when $\nu = 1$, the correlation function

$$G_1(t) \propto \exp(-|t|) \quad (21)$$

corresponding to a Lorentzian line shape. When $\nu \rightarrow \infty$,

$$G_{\infty}(t) \propto \exp(-t^2/2) \quad (22)$$

corresponding to a Gaussian lineshape. For values of ν between 1 and ∞ the spectrum of Student's t distribution interpolates smoothly between the Lorentzian and Gaussian line shapes, in a manner similar to but distinct from, the Voigt lineshape. Note that ν need not be an integer in Equation 19 since $J_{\nu}(\omega)$ is expressed in terms of Euler's Gamma function. As Student's t distribution is a PDF, it is normalized and thus parameter variations are area preserving. We shall use a generalization of Student's t distribution and express it in terms of a location scale family where $f(\omega) \rightarrow f((\omega - \mu)/\sigma)$. [8]

Upon making the following labor-saving substitutions, $a = 1/\nu$, $b = (\nu + 1)/2$ and $c_{\nu} = \Gamma((\nu+1)/2) / \sqrt{\pi\nu}\Gamma(\nu/2)$, [8] the scores may be evaluated as follows

$$\ell = -b \ln(1 + a(\omega - \mu)^2/\sigma^2) + \ln c_{\nu} - \ln \sigma \quad (23)$$

$$\ell_{\mu} = \frac{2\sqrt{ab}}{\sigma} (1 + a(\omega - \mu)^2/\sigma^2)^{-1} \left(\frac{\sqrt{a}(\omega - \mu)}{\sigma} \right) \quad (24)$$

$$\ell_{\sigma} = -\frac{1}{\sigma} + \frac{2b}{\sigma} (1 + a(\omega - \mu)^2/\sigma^2)^{-1} \left(\frac{a(\omega - \mu)^2}{\sigma^2} \right), \quad (25)$$

The Fisher information matrix may be evaluated as follows for $0 < \nu < \infty$: [8,10]

$$I_{ij}(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} \frac{\nu+1}{\nu+3} & 0 \\ 0 & \frac{2\nu}{\nu+3} \end{bmatrix} \quad (26)$$

Note that when $\nu \rightarrow \infty$ we recover the Fisher information matrix for the Gaussian lineshape given in Equation 18 under the assumption that the scale parameter a is fixed at $a = 1$. The Lorentzian case is recovered when $\nu = 1$, in which case the diagonal elements of the Fisher information matrix are equal. We note that it is a straightforward exercise to extend this model to the case of variable ν , but we do not do so here.

Given that $I_{ij}(\theta)$ in Equation 26 is a function of ν , one expects that the scores, which quantify local information about parameter variation will depend in characteristic ways on ν . Figure 1 shows the scores weighted by the probability function. Using the definition of the score in terms of the logarithmic derivative (cf Equation 3) we see that the weighted scores are nothing more than the parametric derivatives. Note that the parametric derivatives with respect to the scale parameter σ are particularly sensitive to the value of ν . As is well-known, one must go far out into the wings of a Lorentzian in order to obtain accurate estimates of the widths. The σ score is a useful way of quantifying this consideration.

4 The Geometry of Parameter Space

In the example of the Gaussian lineshape discussed in Section 2 we noted that proper incorporation of constraints led to parameter sets that were a more faithful representation of the physical problem. If one imagines that the parameter space characterizing the spectral lineshape is a Euclidean space then distances would be measured by some generalization of the Pythagorean theorem. For example, if a , μ and σ are made dimensionless, then distance would be characterized by some equation of the form $s^2 = a^2 + \mu^2 + \sigma^2$. In this model, s^2 can vary from 0 to ∞ . In Euclidean space, we can see how constraints modify our notion of distance. For the familiar case of the surface of a sphere, the distance s^2 is fixed and the three coordinate directions in parameter space are no longer independent. In practice, one specifies the surface of the sphere in terms of embedding equations. In the familiar r , ϑ , φ parameterization, the equation of constraint leads to an induced metric of the form

$$g_{\text{sph}} = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2 \vartheta \end{bmatrix} \quad (27)$$

for unit radius.[6,7] The square root of the determinant of the metric is the well-known weight function or Jacobian for integrating over the surface of the unit sphere. When coordinate transformations are performed in parameter space, the determinant of the Fisher matrix serves the same function as the Jacobian in coordinate space.

Continuing the analogy, in statistical analysis, the Fisher information matrix serves as a metric in parameter space.[4,8] In this interpretation, the scores are the basis vectors defining the tangent plane at the point parameterized by θ which enter into the definition of the metric.[2,6,7,8] From Figure 1, it is clear that the scores depend in a characteristic way on the parameter vector θ . The unit vectors pointing along azimuths and longitudes defined on the surface of a sphere provide a more familiar example of an equivalent phenomenon. The Fisher information matrix defines a notion of distance in parameter space measured by associated changes in the susceptibility parameterized by different θ vectors.

Using the Fisher information matrix as a metric in parameter space for the lineshape problem is only valid to the extent that the lineshape function satisfies the criteria for a PDF, which provides the justification for the averaging process introduced in Section 2. Following Amari,[1] we will say that a PDF can be defined on a set of points, such as the set of points at which the resonance absorption is measured, Ω . A PDF is a mapping from the set Ω to the real numbers $\mathbb{R}$. This can be expressed symbolically as $p: \Omega \rightarrow \mathbb{R}$. The PDF must be non-negative for every ω in Ω , a condition which the linear response absorption satisfies. We also require that the mapping or PDF be normalizable:

$$\int p(\omega, \theta) d\omega = 1, \quad (28)$$

which the linear response absorption also satisfies (cf Equation 2). Physically, we can interpret the PDF as the absorption cross section as a function of ω (cf Section 6).[5]

With the interpretation of the linear response susceptibility as a PDF, we may exploit the rich body of results of statistical geometry in order to learn more about the intrinsic geometry of parameter space. We will extend the use of Student's t distribution in Section 3 as a case study of this process. Given that the lineshape has an area preserving constraint, we expect that the Fisher information matrix, interpreted as a metric, will describe a curved

surface. In such a space, the basis vectors corresponding to the parameters μ and σ will twist and turn in order to satisfy the equal area constraint as discussed above. Christoffel symbols are the analytical objects that are used to quantify this variation.[6,7] For the Fisher information matrix, the Christoffel symbols may be computed as follows

$$\Gamma_{ijk} \equiv \frac{1}{2}(I_{ik,j}(\theta) + I_{jk,i}(\theta) - I_{ij,k}(\theta)). \quad (29)$$

The notation $f_{,k} \equiv \partial f / \partial \theta^k$. Equation 29 defines what are known as Christoffel symbols of the first kind. Note that all of its indices are covariant (cf Section 2). One defines Christoffel symbols of the second kind as follows

$$\Gamma_{ij}^k \equiv I(\theta)^{kl} \frac{1}{2}(I_{il,j}(\theta) + I_{jl,i}(\theta) - I_{il,k}(\theta)), \quad (30)$$

where $I(\theta)^{kl}$ is defined as the matrix inverse of the Fisher information matrix. Note that $I(\theta)^{kl}$ has contravariant indices (cf Section 2). We also introduce the Einstein summation convention that repeated indices in different quantities with one index up and one index down are summed over.

In practice, Christoffel symbols are used to provide curvature corrections for partial derivatives.[6,7] For a scalar quantity φ , such as the squared deviations of the model function and the data as a function of ω , no curvature corrections are necessary.[6,7] One finds

$$\varphi_{,i} \equiv \frac{\partial \varphi}{\partial \theta^i}. \quad (31)$$

If derivatives of $\varphi_{,i}$ are desired, then curvature corrections are necessary and one finds

$$\varphi_{;ij} \equiv \frac{\partial^2 \varphi}{\partial \theta^i \partial \theta^j} + \Gamma_{ij}^k \frac{\partial \varphi}{\partial \theta^k}. \quad (32)$$

The semi-colon notation is a standard way of indicating that curvature corrections are being explicitly included. If the function φ is chosen to be the line-shape function, then Equation 32 can be thought of as a curvature-corrected Hessian. Such objects are a current area of active research.[11] We suggest how these curvature corrections may be used in practical applications in Section 8.

5 Differential Entropy and Spectroscopy

One might expect that narrow distributions with a small σ parameter would be more informative than broad distributions with a large σ parameter. On the other hand, the σ scores plotted in Figure 1 indicate that it is the behavior in the wings of the spectrum that give detailed information about the scale parameter σ as a function of ν . In the context of Information theory, the differential entropy[4]

$$h(\theta) \equiv -\int d\omega p(\omega, \theta) \ell(\omega, \theta) \quad (33)$$

is a useful way to estimate how informative a PDF is. For Student's t distribution, the differential entropy is usually specified with respect to the number of degrees of freedom ν . [4]

$$h_\nu = \frac{n+1}{2} \psi\left(\frac{n+1}{2}\right) - \psi\left(\frac{n}{2}\right) + \ln \left[\sqrt{(\nu)B}\left(\frac{1}{2}, \frac{n}{2}\right) \sigma \right]. \quad (34)$$

This result generalizes the expression given in Cover and Thomas to a translation scale family of PDF's.[4] Here ψ is the digamma function and B is the Euler beta function. The Lorentzian and Gaussian lineshape are simple limiting cases. If the lineshape is a Lorentzian $h_1 = \ln(4\pi\sigma)$. For a Gaussian, one has $h_\infty = \ln(\sqrt{2\pi e}\sigma)$. Note that for sufficiently small values of σ the differential entropy can be negative.[4] A related quantity, the Kulback-Leibler divergence, is defined with respect to an underlying distribution and does not suffer from this defect.[4] We will use the differential entropy, as we will always consider broad distributions for which negative differential entropy does not arise.

From the perspective of information content, large values of h correspond to more informative experiments. Given that the differential entropy is a monotonically increasing function of σ for the model of Section 3, this suggests that manipulating physical conditions to disperse overlapping peaks will enhance the entropy. For the magnetic resonance spectrum, this generally corresponds to higher magnetic fields, where one has improved resolution. We note that a PDF that consists of sums of individual PDF's is known as a mixture family.[4]

6 Estimating Signal Strength

In order to assess whether the mathematical criteria we have developed are reasonable, it is necessary to display the signal strength as a relative change in some physical quantity. Figure 2 shows a lumped equivalent circuit for a transmission mode spectrometer.[12] Expressions for the power dissipated in the load impedance Z_0 can be found in standard treatises.[13] Although this material is well-known, we will review it carefully in order to make our notation clear. In order to compute the power dissipated in the load impedance Z_0 , we can solve the coupled circuit equations appropriate for Figure 2 for the load current, i_L . An easy calculation gives

$$i_L = \frac{V_0 n_1 n_2}{Z_0 n_1^2 + R + j(\omega L - 1/\omega C) + Z_0 n_2^2}, \quad (35)$$

where $j \equiv \sqrt{-1}$. On microwave resonance the reactive term in the denominator vanishes. As the system passes through the magnetic resonance condition there will be dispersive and absorptive contributions. The resonance absorption adds a resistive term to the denominator which is parallel to $R + Z_0 n_1^2 + Z_0 n_2^2$ in a phasor representation and thus gives a first order correction to the denominator in Equation 35.[13] The dispersive term is reactive and thus orthogonal to the steady state term $R + Z_0 n_1^2 + Z_0 n_2^2$ in a phasor representation. It contributes in

second order to the impedance and thus we will ignore it in the small signal limit. The steady state current in the load away from magnetic resonance i_0 is therefore

$$i_0 = \frac{V_0 n_1 n_2}{Z_0 n_1^2 + R + Z_0 n_2^2} \quad (36)$$

Introducing the coupling parameters $\beta_i = n_i^2 Z_0 / R$, this can be rewritten

$$i_0 = \frac{(V_0/Z_0) \sqrt{\beta_1 \beta_2}}{1 + \beta_1 + \beta_2} \quad (37)$$

We assume that the coupling is symmetric and will take the reference power level to correspond to $\beta_1 = \beta_2 = 1/2$. The available power from the generator is defined as

$P_0 = V_0^2 / 4Z_0$. For the reference coupling chosen here, $P_0/2$ is dissipated in the resonator, $P_0/4$ is reflected back to the generator from the input coupling, and $P_0/4$ is transmitted to the load. The change in the load current as the system passes through magnetic resonance Δi_0 is

$$\Delta i_0 = - \frac{(V_0/Z_0) \sqrt{\beta_1 \beta_2}}{(1 + \beta_1 + \beta_2)^2} \left(\frac{\Delta R}{R} \right). \quad (38)$$

Note that, just as in the case of the scores discussed in Section 2, it is the fractional change in the resonator resistance ($\Delta R/R$) that is the important quantity. This fractional change is just $\chi'' \eta Q_U$, where χ'' is the rf susceptibility, η is the filling factor and Q_U is the unloaded resonator quality factor.[13] The change in the power dissipated in the load is therefore

$$\Delta P_L = 2i_0 \Delta i_0 Z_0 = - \frac{8\beta^2}{(1+2\beta)^3} P_0 \chi'' \eta Q_U, \quad (39)$$

where we have chosen symmetric coupling $\beta_1 = \beta_2 = \beta$. The maximum signal occurs when $\beta = 1$ which implies

$$\left(\frac{\Delta P_L}{P_0} \right)_{\max} = - \frac{8}{27} \chi'' \eta Q_u \quad (40)$$

for a transmission spectrometer.

This result provides the connection between the experimentally measured quantity, a relative change in power level dissipated in a load, and the susceptibility which we have been treating as a PDF on the basis of our discussion in Section 4.

7 Finite Signal to Noise

Until now we have considered signal propagation in the absence of noise. The relevant ratio entering any estimate of the information content of an experiment is P_S/P_N where $P_S = \Delta P_L$ is the signal power due to passage through magnetic resonance computed in Section 6 and P_N is the noise power. Note that P_S is a function of frequency and P_N is assumed to be independent of frequency, at least over the signal bandwidth of interest. On the basis of very general considerations, Shannon was able to show that there is a maximum rate at which one may transmit messages down a noisy transmission line, the channel capacity.[4] We suggest that the channel capacity is a useful Figure of Merit (FOM) for estimating the optimum conditions for performing informative experiments. A high channel capacity is consistent with a high signal to noise ratio over a large bandwidth. A low channel capacity is consistent with low signal to noise over a noise-limited bandwidth. For frequency dependent signal and noise power, the Shannon-Hartley theorem may be expressed as follows

$$C(\theta) = \int d\omega \log_2 \left(1 + \frac{P_S(\omega, \theta)}{P_N(\omega)} \right), \quad (41)$$

where $C(\theta)$ is the channel capacity expressed in bits as a function of the parameter vector θ . [4] Note that when the signal power $P_S(\omega, \theta)$ is much less than the noise power $P_N(\omega)$ that the integrand goes to zero, as $\log_2(1) = 0$. Thus one may extend the limits of integration to $\pm\infty$ without introducing serious error, as long as the noise bandwidth is significantly larger than the signal bandwidth. It will be useful to define a dimensionless parameter χ_c which characterizes the signal to noise ratio. Then

$$C(\theta) = \int d\omega \log_2 \left(1 + \frac{\chi''(\omega, \theta)}{\chi_c} \right). \quad (42)$$

We use $\chi''(\omega, \theta)$ here instead of $p(\omega, \theta)$ to emphasize that we are now working with experimentally observable quantities. When $\chi''(\omega) \ll \chi_c$, for all ω in the signal bandwidth, $C \rightarrow 0$ which has the correct behavior for a useful FOM. If there is a spectral band, or bands, for which $\chi''(\omega) \gg \chi_c$ then those regions will make a significant contribution to the channel capacity.

Note that experiments at different frequencies can be characterized by frequency dependent signal to noise ratios which depend on the available source power through P_S and the available noise powers P_N , whose behavior depends on the noise figure of the detection scheme. There are reports in the literature of sensitivities that have a wide variety of frequency dependencies depending on the experimental conditions.[13,15] The advantage of equations such as Equation 42 is that the possible information gain from different experiments under different conditions may be estimated *a priori* for a given set of model parameters.

In order to make some of these abstract notions more tangible, we consider a specific model relevant for spin-label ESR spectroscopy. Figure 3 shows a set of simulated absorption spectra over two decades of frequency, corresponding to 2.4 GHz (S band), 9.4 GHz (X band), 24 GHz (K band), 94 GHz (W band) and 240 GHz (G band). We used the EPRLL spectral simulation software to compute field-swept spectral simulations that were subsequently transformed into the frequency-swept susceptibilities shown in Figure 3. The EPRLL software is available from the ACERT web site.[14] Treating the susceptibility as a

PDF, as we have advocated, allows one to compute the differential entropy in the spectrum from Equation 33. Regardless of the detailed behavior of the spectra in Figure 3, we note that the spectral extent increases monotonically from S through G band. Reasoning by analogy, we expect from our study of the model of Section 3 that the differential entropy should also increase monotonically as the observation frequency is increased. Figure 4 shows that this reasonable hypothesis is borne out. One point to make is that, qualitatively, the entropy or information in a spectrum scales with the spectral extent and that this scaling depends on the rotational regime. Note that for the chosen model parameters, the entropy is similar for both rotational diffusion rates.

When noise is present, we have argued that channel capacity is a better FOM for doing experimental design. Figure 5 shows the channel capacity as a function of ESR frequency for three different noise models. The first noise model, represented by the * symbols, corresponds to a model where the signal to noise ratio is the same for all ESR frequencies. In this case, the channel capacity has similar behavior to the entropy shown in Figure 4. The noise model for the + symbols assumes that the signal to noise ratio decreases at 3dB per decade. Such a model is consistent with the generally higher noise figures of rf components as the frequency increases as well as the lower available power of millimeter wave sources. Note that for the spectral line shapes used to compute the channel capacity, corresponding to the top row of Figure 3, the X and K band spectra are less informative than the W and G band spectra. We note that the best estimators of the channel capacity are the spectral extent and the ratio of the available signal power to the noise power. Finally, the $\odot$ symbols correspond to a model where the signal to noise ratio decreases as 6dB per decade. The poor high frequency performance of such a model would provide a strong disincentive for performing high frequency experiments, unless a means could be found to improve the signal to noise ratio. The moral of Figure 5 is that experimental conditions should be chosen to maximize the spectral dispersion and the signal to noise ratio in order to perform the most informative experiments. We note that increasing the spectral extent will reduce the signal to noise ratio due to the normalization constraint. We repeat that Figure 5 makes clear that in order for high frequency experiments to be competitive with lower frequency experiments, it is necessary to have as high a signal to noise ratio as possible.

8 Discussion

In Section 7 we introduced criteria for estimating experimental conditions that would lead to the most informative data. The conclusions implied by the calculations accord with practical experience, although we emphasize that the advantage of the current approach is that information theory provides a computable FOM within the context of a specific model for the spectrum and the noise power density. Figure 5 makes very clear that there is a trade-off between resolution and information in the presence of finite noise. It is important to note that the channel capacity and the entropy are integrated quantities which represent a smoothing of the spectral details. Careful study of Figure 3 and Figure 4 show that, indeed, the details of the spectra are less important than the spectral extent. In some sense, then, the entropy and channel capacity are an abstraction of the entire information content of a spectrum and represent a ‘global’ measure of variations in the absorption across the frequency range of interest. The detailed structure of the scores of the line shape function which played such an important role in our study of the model of Section 3 reports on the ‘local’ behavior of the model and is relevant for estimating parameter sensitivity as distinct from the ‘global’ spectral information properties. We expect that both the ‘global’ and ‘local’ perspectives will inform the parameter estimation problem in important ways, and we intend to explore the available insights from these approaches in subsequent publications.

In particular, we expect that the curvature-corrected Hessian will be a useful tool for refining parameter searches. In order to assess the importance of curvature corrections in the context of the model of Section 3, we need the non-vanishing Christoffel symbols of the second kind which read:

$$\Gamma_{\sigma\sigma}^{\sigma} = -\frac{1}{\sigma} \quad (43)$$

$$\Gamma_{\mu\mu}^{\sigma} = \frac{1}{\sigma} \frac{\nu+1}{2\nu} = \frac{2ab}{\sigma} \quad (44)$$

$$\Gamma_{\mu\sigma}^{\mu} = \Gamma_{\sigma\mu}^{\mu} = -\frac{1}{\sigma}. \quad (45)$$

Note that as σ increases, curvature effects diminish in importance. In terms of the scores we find after an easy calculation

$$p_{;\mu\mu} = p \left[\ell_{\mu\mu} + \ell_{\mu} \ell_{\mu} + \frac{2ab}{\sigma} \ell_{\sigma} \right] \quad (46)$$

$$p_{;\mu\sigma} = p \left[\ell_{\mu\sigma} + \ell_{\mu} \ell_{\sigma} - \frac{1}{\sigma} \ell_{\mu} \right] \quad (47)$$

$$p_{;\sigma\sigma} = p \left[\ell_{\sigma\sigma} + \ell_{\sigma} \ell_{\sigma} - \frac{1}{\sigma} \ell_{\sigma} \right] \quad (48)$$

Expressions for the second derivative terms for the model of Section 3 are given in Kass and Vos.[8] We note that the first and second terms on the right hand side of Equations 46–48 enter into the integrand for the Fisher information matrix, Equation 12. The rightmost terms on the right hand sides of Equations 46–48 are the curvature correction terms, which are seen to be proportional to the plotted terms in the second and third rows of Figure 1. Since these correction terms can be written in terms of the scores we expect that incorporating these corrections into existing parameter optimization codes should be a straightforward proposition, and we are actively exploring this intriguing possibility.

Although the methods introduced here may be applied to a wide variety of spectroscopies, it is particularly useful when the fitting function has analytical derivatives that may be computed efficiently. For the ESR line-shape problem, where rotational diffusion provides the relaxation mechanism, the spectrum is computed operationally from the inverse of a large sparse matrix.[16] Recent unpublished work has shown that it is possible to write down closed form expressions for the needed derivatives which may be evaluated subsequently. This is work in progress and will be the subject of a forthcoming paper.

Acknowledgments

KAE thanks ACERT (P41 RR016292) for the use of its computational resources. This research was supported in part by a Faculty Research Awards Program grant from the University at Albany (SUNY). He also notes that he will be exactly 49/60 the age of Wolfgang Lubitz on 23 July 2009.

References

1. Amari S, Nagaoka H. Methods of Information Geometry. Translations of Mathematical Monographs 2000;191:2.
2. Murray, MK.; Rice, JW. Differential Geometry and Statistics. Chapman and Hall/CRC; Boca Raton: 1993.
3. Toda, M.; Kubo, R.; Saito, M.; Hashitsume, N. Statistical Physics: Nonequilibrium statistical mechanics. 2. Springer; New York: 1992.
4. Cover, TM.; Thomas, JA. Elements of Information Theory. Wiley; New York: 1991.
5. Forster, D. Advanced Book Classics. Addison-Wesley; New York: 1995. Hydrodynamic fluctuations, broken symmetry, and correlation functions.
6. Kühnel, W. Differential Geometry: Curves–Surfaces–Manifolds. 2. AMS; New York: 2006.
7. Synge, JL.; Schild, A. Tensor Calculus. Dover; New York: 1978.
8. Kass, RE.; Vos, PW. Geometrical Foundations of Asymptotic Inference. Wiley; New York: 1997.
9. Dreier I, Kotz S. A note on the characteristic function of the t distribution. Statistics and Probability Letters 2002;57:221–224.
10. Abramowitz, M.; Stegun, I. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover; New York: 1964.
11. Shima, H. The Geometry of Hessian Structures. World Scientific; Singapore: 2007.
12. Budil DE, Ding Z, Smith GR, Earle KA. Jones Matrix Formalism for Quasioptical EPR. J Magn Reson 2000;144:20–34. [PubMed: 10783270]
13. Poole, CP, Jr. Electron Spin Resonance: A Comprehensive Treatise on Experimental Techniques. 2. Dover; New York: 1983.
14. <http://www.acert.cornell.edu>
15. Rinard GA, Quine RW, Harbridge JR, Song R, Eaton GR, Eaton SS. Frequency Dependence of EPR Signal-to-Noise. J Magn Reson 1999;140:218–227. [PubMed: 10479565]
16. Schneider, DJ.; Freed, JH. Spin Labeling: Theory and Applications, Vol. III Biological Magnetic Resonance. Vol. 8. Plenum, NY: 1939. Calculating Slow Motional Magnetic Resonance Spectra: A User's Guide; p. 1-76.

Figure 1.

Model dependence of spectral derivatives. The first row contains the lineshape. The second row contains derivatives of the lineshape with respect to μ . The third row contains derivatives with respect to σ . The first column is a Gaussian lineshape. The third column is a Lorentzian lineshape. The second column corresponds to a lineshape intermediate between a Lorentzian and a Gaussian ($\nu = 3$).

Figure 2.

Equivalent circuit for a transmission mode spectrometer. Here Z_0 is the source impedance and equal to the load impedance. $V_0^2/2Z_0$ is the available power of the generator. The quantities L , C and R characterize the resonant structure in which the sample is placed. Coupling is modeled by ideal transformers at the input (turns ratio $n_1:1$) and the output (turns ratio $1:n_2$) of the resonant structure.

Figure 3.

Representative 'generic nitroxide' EPR absorption spectra for a model of isotropic rotational diffusion. The upper traces correspond to a rotational tumbling rate of $10^7 \text{ rad}^2/\text{sec}$. The lower traces correspond to a rotational tumbling rate of $10^8 \text{ rad}^2/\text{sec}$. The dashed bar below the traces in the lower row corresponds to a spectral extent of 100 G in an equivalent field swept experiment

Figure 4.

Differential entropy as a function of ESR frequency for different rotational diffusion rates. The y-axis has units of nats.[4] The x-axis has units of angular frequency. The + symbols correspond to $R = 10^7 \text{ rad}^2/\text{sec}$. The $\odot$ symbols refer to a rotational diffusion rate of $10^8 \text{ rad}^2/\text{sec}$. The differential entropy was computed numerically from the spectra shown in Figure 3.

Figure 5.

Channel capacity as a function of ESR frequency for three different noise models. The y-axis has units of bits.[4] The x-axis has units of angular frequency. The * symbols correspond to a model where the noise power density is the same for all ESR frequencies. The + symbols correspond to a model where the signal to noise ratio decreases at a rate of 3 dB per decade. The ○ symbols correspond to a model where the signal to noise ratio decreases at 6dB per decade