Maximum Entropy Reconstruction

Jeffrey C. Hoch*
University of Connecticut Health Center, Farmington, CT, USA

Alan S. Stern
Rowland Institute at Harvard, Cambridge, MA, USA

&

Mehdi Mobli
The University of Queensland, St Lucia, Queensland, Australia

1 INTRODUCTION

The discrete Fourier transform (DFT) has played a seminal role in the development of modern NMR spectroscopy, but as a method for spectrum analysis of time series data, it has a number of well-known limitations.[1] Chief among them is the difficulty of obtaining high-resolution spectra from short data records. A host of non-Fourier methods of spectrum analysis have been developed that attempt to overcome the limitations of the DFT.[2–8] Several very different methods invoke the maximum entropy principle, which originates from the information-theoretic definition of entropy as a measure of missing information first formulated by Shannon. One of the most versatile approaches, called *maximum entropy reconstruction*, applies the maximum entropy principle in a Bayesian context to determine the spectrum having the highest entropy that is consistent with the measured data. Because it approaches spectrum analysis as an inverse problem, it is capable of performing stable deconvolution and processing data that are sampled at nonuniform intervals. The latter has found useful application in multidimensional NMR, where practical constraints on measuring time typically limit the attainable resolution along indirect dimensions. Here we describe the theory and application of maximum entropy reconstruction in NMR, and contrast it with other methods that invoke the maximum entropy principle.

*Update based on original article by Jeffrey C. Hoch, *Encyclopedia of Magnetic Resonance*, © 1996, John Wiley & Sons Ltd.

2 SPECTRUM ANALYSIS AS AN INVERSE PROBLEM

Inverse problems are those in which the properties of interest can only be viewed indirectly. A famous example is Plato's allegory of the cave,[5] in which the problem is to infer the characteristics of an object from the shadows it casts in the firelight on the walls of a cave. In NMR spectrum analysis, the inverse problem can be stated as that of recovering the spectrum f_n of an ensemble of spins when the only observations we have are the free induction decay (FID) at various times (d_k), contaminated by noise (ε_k):

$$\hat{d}_k = d_k + \varepsilon_k \tag{1}$$

We can approach this inverse problem by placing restrictions on the possible solutions f_n. The first restriction comes from the obvious requirement that f_n must be consistent with the experimental observations. The agreement can be quantified through the use of a constraint statistic that measures how well the reconstructed spectrum agrees with the available data. The statistic is determined by comparing a "mock" FID, given by the inverse DFT of the reconstructed spectrum, with the actual data. When the noise and experimental error in the measured data are normally distributed, an appropriate statistic is the (unweighted) χ^2 statistic

$$\chi^2 = \sum_{k=0}^{M-1} |m_k - \hat{d}_k|^2 \tag{2}$$

where M is the number of data samples collected and m_k is the mock FID. Other forms of the constraint statistic C are possible, but χ^2 is the one generally used because of its simplicity and the difficulty in justifying any other distribution of errors in the data. (A weighted χ^2 statistic can also be used, for example, if there is good reason to believe that some of the data points are more subject to error than others.) Whatever the form of C, the constraint that the reconstructed spectrum must be consistent with the measured data takes the form

$$C \leq C_0 \tag{3}$$

where C_0 is an upper bound on the allowed error. Given a prior estimate of the amount of noise in the data, C_0 should be comparable to the root mean square (RMS) noise level.

3 MAXIMUM ENTROPY RECONSTRUCTION

The constraint statistic alone does not provide sufficient restriction on the possible reconstructions to be of much use. In fact, any mock FID that matches the experimental FID to within C_0 for the first M points will satisfy the constraint. If we are to improve over the DFT, we need some additional criteria for regularizing the reconstructed spectrum that will effectively constrain the values of the mock FID beyond M. One such criterion is the maximum entropy principle, which says that a reasonable reconstruction should add no new information beyond that contained in the experimental data. The principle originated in the work of Claude Shannon on the information-carrying capacity of circuits,[9] where he showed that the entropy of a probability distribution p, which is defined by

$$S(p) = -\sum p_n \log p_n \tag{4}$$

eMagRes, Online © 2007 John Wiley & Sons, Ltd.

This article was previously published in the *Encyclopedia of Magnetic Resonance* in 2011 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470034590.emrstm0299.pub2

is a measure of the lack of information. There is a formal link between Shannon's entropy and the statistical entropy of an ensemble, and this link can be used to provide a constraint on possible spectrum reconstructions. On a more pragmatic level, the maximum entropy principle can be viewed simply as a means of obtaining a smooth, bounded reconstruction. The use of the maximum entropy principle to regularize the possible solutions to the spectrum analysis inverse problem is called *maximum entropy* (MaxEnt) *reconstruction*. (Maximum entropy reconstruction should be distinguished from the maximum entropy method (MEM) introduced by Burg,[10] which is more closely related to linear prediction (LP).) Entropy is not unique as a regularizer, however, and there are other functionals that have similar properties.[11]

An important distinction between MaxEnt reconstruction and methods of spectrum analysis (such as those based on LP) that implicitly assume Lorentzian lineshapes is that MaxEnt reconstruction per se makes no assumptions about the signal to be recovered. Yet, it is flexible enough to allow prior information to be incorporated into the reconstruction. A particularly simple means for incorporating prior knowledge is through a convolution kernel. Typical kernels might include exponential decay at some known rate, or J-modulation at a known frequency. The reconstruction then yields the *deconvolved* spectrum with the highest entropy.

4 ENTROPY FUNCTIONALS FOR NMR

For modern NMR experiments, we need to be able to reconstruct complex spectra containing both positive and negative components. The Shannon formula [equation (4)] clearly does not apply to complex-valued spectra or spectra containing negative components. Computing the entropy of the magnitude or power spectrum solves this problem, but suffers from other problems, including singular derivatives. Hore and colleagues[12] considered the entropy of an ensemble of spin-1/2 particles to derive the functional

$$S(f) = -\sum_{n=0}^{N-1} \frac{|f_n|}{def} \log\left(\frac{|f_n| \Big/ def + \sqrt{4 + |f_n|^2/def^2}}{2}\right) - \sqrt{4 + |f_n|^2/def^2} \qquad (5)$$

where *def* is a scaling parameter that is related to the sensitivity of the spectrometer and to the number of spins in the sample; however, it is more convenient to treat *def* as an adjustable parameter. This functional has the advantages that it is convex, the derivative is continuous, and the value is insensitive to phase of the signal.[11]

5 NUMERICAL ALGORITHMS

The next detail that we are faced with is how to solve the constrained optimization problem, that is, the problem of maximizing $S(f)$ subject to the constraint that $C \leq C_0$. An equivalent unconstrained optimization problem is to maximize the objective function

$$Q \quad S \quad \lambda C \qquad (6)$$

where λ is a Lagrange multiplier. The solution we seek corresponds to a critical point of Q, i.e., a point where $\nabla Q = 0$. In the general case there is no analytic solution to this problem, and we are forced to seek numerical solutions. For many problems of this type, a numerical search using the gradient of the objective function (steepest ascent or conjugate gradients) is a reasonably efficient method for finding the critical point. Unfortunately, an objective function constructed from two such dissimilar functionals as the χ^2 statistic and the entropy is quite difficult to maximize. Practical solution of the general MaxEnt reconstruction problem requires a fairly complex optimizer, and the most efficient algorithms require of the order of 100 (or more) times the computational effort of the fast Fourier transform (FFT).

Since the entropy (given by equation (5)) and the constraint statistic (equation (2)) are both everywhere convex, there is a unique, global solution to the constrained optimization problem. Furthermore, this maximum satisfies $C = C_0$ provided that the trivial solution $f_n = 0$ does not also satisfy the constraint.

There have been several algorithms published for computing MaxEnt reconstructions. A particularly robust and efficient algorithm was described by Skilling and Bryan.[13] This method uses multiple search directions and a variable metric to determine the size of the step at each iteration. We describe here a variant we have found useful in our laboratory.[14]

The algorithm begins with a flat trial spectrum equal to zero everywhere. At each iteration, a mock FID is computed from the current value of the trial spectrum. In the general case, this involves inverse Fourier transformation and multiplication by a decay kernel, which is specified by the user as an input parameter. A decay does not need to be applied, but when it is, the reconstructed spectrum is an approximation of what would have been observed had the decay not been present. The first M points of the N point mock FID are used to compute the value of C, according to equation (2). The algorithm constructs a small set of direction vectors, and computes a quadratic approximation to the entropy in the subspace spanned by these vectors. Since the constraint is itself quadratic, it is possible to maximize analytically the entropy approximation subject to the constraint in this subspace. This results in the trial spectrum for the next iteration.

The quadratic approximation to the entropy is accurate only for small step sizes, so when the value of C for the mock FID is far from the desired value C_0, it is better to solve the analytic maximization using a constraint value C', which is intermediate between C_0 and the current value of C, rather than attempt one large step. Consequently, the algorithm proceeds in two phases. In the first phase, C is larger than C_0 and the algorithm attempts to lower C. In the second phase, C_0 has been attained and the algorithm seeks to maximize the entropy.

This approach could be based on direction vectors that include ∇S, the gradient of the entropy (with respect to f). The problem is that the gradient can be dominated by small values of f, and the algorithm would spend more time adjusting values that are close to zero than adjusting signal peaks. The key insight of Skilling and Bryan was that the introduction of an "entropy metric" produces an algorithm that more evenly weights the contributions of large and small values of the spectrum. Using the entropy metric amounts to replacing ∇S and ∇C (the gradient of the constraint with respect to the trial spectrum) with $\mathbf{H}^{-1} S$ and $\mathbf{H}^{-1} C$, where $\mathbf{H}$ is ^2S, the

eMagRes, Online © 2007 John Wiley & Sons, Ltd.

This article was previously published in the *Encyclopedia of Magnetic Resonance* in 2011 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470034590.emrstm0299.pub2

negative of the Hessian (the matrix of second derivatives) of S. The matrix $\mathbf{H}$ is nearly diagonal, so its inverse is easy to compute. The search directions used during the first phase of the algorithm are

$$\nabla C, \mathbf{H}^{-1}\nabla C, \mathbf{H}^{-1}\nabla Q, \mathbf{H}^{-1}(\nabla^2 C)\mathbf{H}^{-1}\nabla Q, Q = S - \lambda C \tag{7}$$

and λ is set to $|\nabla S|/|\nabla C|$. During the second phase, the second search direction is replaced by ∇Q. In the original algorithm of Skilling and Bryan, neither ∇C nor ∇Q is used as a search direction. We found that including them speeds up the convergence. The major effort in computing these vectors stems from four DFTs (two each to compute ∇C and to apply $\nabla^2 C$) and multiplication by $\mathbf{H}^{-1}$.

With these direction vectors, a quadratic model for the entropy and the constraint can be constructed using a second-order Taylor expansion about the trial spectrum. A spectrum f' in the subspace spanned by the direction vectors can be represented by a vector of length four, $\bar{a}$; the correspondence is given by

$$f' = a_1\upsilon_1 + a_2\upsilon_2 + a_3\upsilon_3 + a_4\upsilon_4 + f_0 \tag{8}$$

where υ_1 through υ_4 are the direction vectors and f_0 is the current trial spectrum. The quadratic approximations then become

$$S(\bar{a}) \approx S(f_0) + \bar{B}\cdot\bar{a} + \frac{1}{2}\bar{a}\cdot\mathbf{M}\cdot\bar{a} \tag{9a}$$

and

$$C(\bar{a}) \approx C(f_0) + \bar{G}\cdot\bar{a} + \frac{1}{2}\bar{a}\cdot\mathbf{N}\cdot\bar{a} \tag{9b}$$

where vectors $\bar{B}$ and $\bar{G}$ are inner products of ∇C and ∇S with the direction vectors

$$\bar{B}_i = \nabla S\cdot\upsilon_i \tag{10a}$$

$$\bar{G}_i = \nabla C\cdot\upsilon_i \tag{10b}$$

Similarly, the 4×4 matrices $\mathbf{M}$ and $\mathbf{N}$ are the inner products of $\nabla^2 S$ and $\nabla^2 C$ with the direction vectors

$$M_{ij} = \upsilon_i\cdot\nabla^2 S\cdot\upsilon_j \tag{11a}$$

$$N_{ij} = \upsilon_i\cdot\nabla^2 C\cdot\upsilon_j \tag{11b}$$

Computing these values requires several more Fourier transformations and large matrix products. The Lagrange condition $\nabla S - \lambda\nabla C = 0$ for the maximum of the objective function in the subspace becomes

$$(\bar{B} + \mathbf{M}\cdot\bar{a}) - \lambda(\bar{G} + \mathbf{N}\cdot\bar{a}) = \bar{0} \tag{12}$$

By a suitable change of coordinates in the subspace, it is possible to simplify this equation by simultaneously diagonalizing $\mathbf{M}$ and $\mathbf{N}$. The solution then is given by

$$a_i = \frac{B_i - \lambda G_i}{\lambda N_{ii} - M_{ii}} \tag{13}$$

λ can be found by using a binary search to determine the value for which $C(\bar{a})$ equals the target value C'.

Convergence criteria for determining when the trial spectrum is sufficiently close to the MaxEnt reconstruction derive from the Lagrange condition and the requirement that $C \quad C$. The Lagrange condition implies that ∇C and ∇S are parallel, so convergence can be monitored by computing the value

$$Test = \left|\frac{\nabla S}{|\nabla S|} - \frac{\nabla C}{|\nabla C|}\right| \tag{14}$$

The algorithm terminates when $C \cong C_0$ and $Test \ll 1$. In practice, we stop the algorithm when $Test < 10^{-3}$, or when a preset maximum number of iterations has been executed.

The value of C_0 can be estimated from the data by examining a portion of an FID that is essentially free of signal, or alternatively, by examining a signal-free region of the DFT spectrum. For MaxEnt reconstructions, C_0 should generally be somewhat larger than the estimate of the RMS noise level. Very large values of C_0 result in overly smooth reconstructions in which weaker components are washed out.

The other adjustable parameter, *def*, is rather more difficult to prescribe. *def* represents the scale at which the nonlinear effects of MaxEnt reconstruction become significant. A large value of *def* results in nearly linear reconstructions. When *def* is large and C_0 is zero, the resulting reconstruction is essentially the same as the DFT of the zero-filled FID. However, very low values of *def* can give rise to spurious artefacts. As a rule of thumb, we use values of *def* somewhat lower than the noise level. Fortunately, the results are not overly sensitive to the choice of *def*. We know of no algorithmic procedure for determining the best value.

5.1 Constant-λ Algorithm

In principle, multidimensional MaxEnt spectra can be reconstructed while computing the overall entropy of the fully-dimensional spectrum, or by partitioning the reconstruction into a series of reconstructions computed for lower dimensional spectra. For example, a 2-D MaxEnt spectrum can be computed via a series of 1-D MaxEnt reconstructions in f_2 following Fourier processing of f_1. If the constraint statistic C_0 is kept constant between rows, it can lead to variation in the weighting (λ) of the constraint and the entropy. This is due to natural variations in signal and noise distribution, which affect the entropy. This scenario will introduce small changes in the reconstruction between rows and may have a significant effect on peak shapes. By using a constant value for the weighting λ, one can minimize the variation of the nonlinearity between rows. A good estimate of λ can be made by finding representative rows where the constraint statistic $C(f) = C_0$ is satisfied and using the value of λ found for these rows to perform the complete reconstruction. The same basic strategy can be applied to higher dimensions, e.g., a 3-D spectrum can be constructed as a series of 2-D plane reconstructions. The approach of using a fixed value of λ, rather than a fixed value of C_0, is called the constant-λ algorithm.[15] The advantage of this approach is that the memory requirements for intermediate storage are smaller, leading to more efficient computation.

5.2 Automated MaxEnt

While the formal derivation of the MaxEnt algorithm specifies criteria for determining the appropriate values of *def* and C (or λ), applying those criteria in practice is challenging

eMagRes, Online © 2007 John Wiley & Sons, Ltd.

This article was previously published in the *Encyclopedia of Magnetic Resonance* in 2011 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470034590.emrstm0299.pub2

and finding optimal values even more so. Fortunately, the results of MaxEnt reconstruction are not terribly dependent on the precise values of the parameters, over a wide range. Using empirical rules of thumb, Mobli *et al*. described an automated procedure for determining values for the adjustable parameters.[16,17] While they cannot be said to be optimal, they have proven to be useful estimates for most applications. The key to the approach is to evaluate the noise level of the acquired data in situ in order to estimate the value of C_0. Using Parseval's theorem, it is then possible to estimate the frequency domain *def* parameter. In practice, this approach tends to result in a very conservative estimate of *def* resulting in faster and more linear but noisier reconstructions. Alternatively, the value of *def* can be systematically reduced by some constant to produce smoother reconstructions. This procedure can also be implemented to use the constant-λ algorithm. In this approach, in addition to estimating the C_0 and *def* parameters, 10 frequencies with the largest signal components are identified. Dimensions orthogonal to these frequencies are then reconstructed using the estimated C_0 and *def* parameters. The resulting λ-values for these 10 reconstructions are evaluated, and the average λ value is used together with the previously estimated *def* parameter to reconstruct the entire dataset.

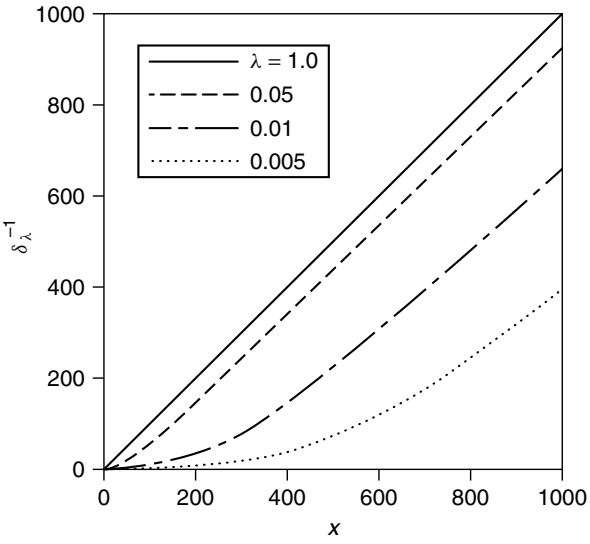

Figure 1 The nonlinear transformation $\delta_\lambda^{-1}(x)$ applied to the DFT spectrum that results in the MaxEnt reconstruction when $N = M$, for various values of the Lagrange multiplier λ

6 PROPERTIES

6.1 Analytic Solution

While numerical solution is required in the general case, there is a special case of MaxEnt reconstruction that has an analytical solution. Though unrealistic in several respects, examination of this special case can give some insights into how MaxEnt reconstruction works. When N (the number of points in the reconstructed spectrum) is equal to M (the number of experimental data points), and when the relationship between the trial spectrum and the mock FID is given simply by the inverse Fourier transform (i.e., we are not trying to deconvolve a decay), the solution can be found analytically. Parseval's theorem, which states that the sum of the squared magnitudes of a vector is equal to the sum of the squared magnitudes of its Fourier coefficients, permits the constraint statistic to be computed in the frequency domain. Under these circumstances, the Lagrange condition becomes

$$0 = \nabla Q = \nabla S - 2\lambda(f - F) \tag{15}$$

where F is the DFT of the data $\hat{d}$. The solution is given by

$$|f_n| = \delta_\lambda^{-1}(|F_n|) \tag{16a}$$

$$\text{phase}(f_n) = \text{phase}(F_n) \tag{16b}$$

where δ_λ is the function

$$\delta_\lambda(x) = x - s'(x)/2\lambda \tag{17}$$

and $s(x)$ is the contribution of a spectral component with magnitude x to the overall entropy S (see equation (5)). This result corresponds to a nonlinear transformation, applied point by point to the DFT of the time domain data. The transformation depends on the value of λ, and has the effect of scaling every point in the spectrum down, but points closer to the baseline are scaled down more than points far above the baseline. Figure 1 illustrates $\delta_\lambda^{-1}(x)$ for various values of λ. As λ increases, the relative weight given to the constraint term in the objective function increases, and the transformation becomes more nearly linear.

This helps to explain the characteristic of MaxEnt reconstructions that noise near the baseline is suppressed more than noise away from the baseline (for example, superimposed on a broad peak, Figure 2). A further point to note is that there is a very important distinction between the signal-to-noise ratio (SNR) and sensitivity. Sensitivity is the ability to distinguish signal from noise. Applying the same transformation to both the signal and the noise cannot improve this ability, since peaks that are comparable in height to the noise level will be reduced by the same amount as the noise. The ratio between the highest signal peaks and the noise may increase, but small peaks will be just as difficult to distinguish as before. For linear operations (such as apodization), an improvement in SNR necessarily implies an improvement in sensitivity; for nonlinear operations (such as MaxEnt reconstruction) this is not so. In this special case, gains in SNR in the MaxEnt reconstruction are purely cosmetic. In the more general case, for example when an approximately known decay is deconvolved from the reconstructed spectrum, there may be real sensitivity gains. However, a prudent investigator will always question whether gains in SNR really correspond to gains in sensitivity.[18]

6.2 SNR versus Sensitivity, and Quantification

The nonlinearity of MaxEnt reconstruction is an inherent characteristic, and is responsible for the method's ability to achieve noise suppression without sacrificing resolution to the extent associated with linear filtering (such as the windowed DFT). This nonlinearity has important implications in situations where quantification of peak intensities or volumes is required, such as nuclear Overhauser effect measurements or difference spectroscopy. For difference spectroscopy, it may

eMagRes, Online © 2007 John Wiley & Sons, Ltd.

This article was previously published in the *Encyclopedia of Magnetic Resonance* in 2011 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470034590.emrstm0299.pub2

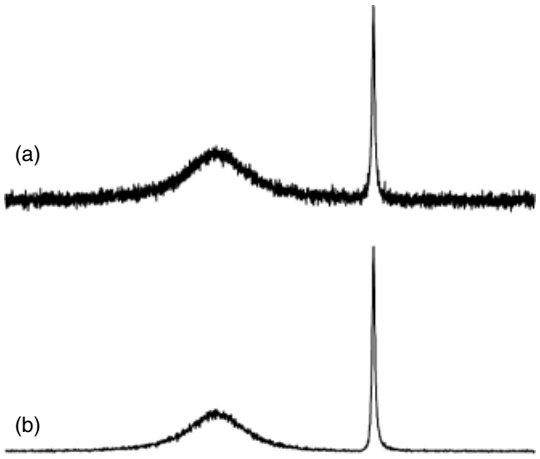

Figure 2 (a) DFT and (b) MaxEnt reconstruction for a synthetic noisy signal. Noise near the baseline is suppressed more effectively than noise superimposed on the broad peak

be sufficient to compute the difference using the time-domain data and compute the MaxEnt reconstruction of the difference. This ensures that the same nonlinearity applies to both experiments. When measuring nuclear Overhauser effects, or otherwise quantifying peak intensities, there are two possible approaches. One is to tightly constrain the reconstruction to match the data, which forces the reconstruction to be more nearly linear (although at the expense of noise suppression). Another is to add synthetic signals of known relative intensity into the time domain data prior to reconstruction. A calibration curve can then be constructed by quantifying the intensities of the known signals.[15]

7 APPLICATIONS

7.1 Deconvolution

An approach to deconvolution using linear methods of spectrum analysis that exploits the convolution theorem is to divide the time-domain data by a convolution kernel whose DFT is the function one wishes to deconvolve. This approach works reasonable well for window functions, for example that have values significantly greater than zero. For functions that have very small values or zero, this approach becomes unstable (or undefined). The inverse nature of MaxEnt reconstruction makes it amenable to performing stable deconvolution, even when the kernel has small or zero values. To perform deconvolution using MaxEnt, one modifies the constraint statistic used to defined consistency with the measured data [equation (2)] by including the kernel function

$$C(\mathbf{f}, \mathbf{d}) = \sum_{i=0}^{M-1} |m_i - d_i|^2 = \sum_{i=0}^{M-1} |k_i \cdot \mathrm{IDFT}(\mathbf{f})_i - d_i|^2 \qquad (18)$$

where k_i are the elements of the kernel function $\mathbf{k}$. When $\mathbf{k}$ is everywhere equal to 1, the mock data are simply the inverse discrete Fourier transform (IDFT) of the trial spectrum. Otherwise, the spectrum of the mock data is given by $\mathbf{f}$ convolved with the spectrum of $\mathbf{k}$, so $\mathbf{f}$ is the deconvolved spectrum.

A useful application of this approach is for virtual decoupling, in which a cosine modulation caused by spin–spin coupling is deconvolved. Shimba *et al*. showed that a single convolution kernel with a fixed frequency is sufficient to perform virtual decoupling so long as the variation in the couplings is comparable to or smaller than the linewidth of the resonances.[19,20] Virtual decoupling can be useful when radiofrequency (RF) decoupling is not feasible, and avoids sample heating or Bloch–Siegert shifts that can accompany RF decoupling.

7.2 Nonuniform Sampling (NUS)

The ability to handle data that have been collected at nonuniform time intervals is another consequence of the inverse nature of MaxEnt reconstruction, and this capability has been used to overcome limits on resolution in multidimensional experiments imposed by practical constraints on measuring time. Non-Fourier methods of spectrum analysis, including MaxEnt reconstruction, are typically used to compute spectra from data collected at nonuniform intervals. NUS in multidimensional NMR remains an active area of research, and is discussed elsewhere in this volume (see ***Rapid Multidimensional NMR: Decomposition Methods and their Applications***, ***Maximum Entropy: Multidimensional Methods***)

8 RELATED METHODS

8.1 Burg Maximum Entropy Method

An application of the maximum entropy principle to signal processing that preceded MaxEnt reconstruction was introduced by J. P. Burg.[10] Noting the Fourier transform relationship between the autocorrelation function of a time series and the power spectrum, Burg asked what constraints the positivity of the power spectrum places on the autocorrelation function beyond the measured interval. He showed that the answer is very similar to the result of extrapolating the autocorrelation function by maximizing the entropy, and devised a very simple procedure that amounts to fitting LP coefficients to the measured data, from which the MaxEnt (power) spectrum can then be derived. This "Burg maximum entropy method" bears a much closer resemblance to LP extrapolation and the linear prediction z-transform (LPZ) method[21] (see ***Fourier Transform and Linear Prediction Methods***) than to MaxEnt reconstruction. The use of the LP framework implicitly invokes a model of a sum of exponentially decaying sinusoids to describe the signals, in contrast to MaxEnt reconstruction, which does not assume the signals conform to any particular form.

8.2 Forward MaxEnt

The nonlinearity of MaxEnt reconstruction presents challenges to quantitative applications, such as determining nuclear Overhauser effects (NOEs), relaxation rates, or metabolomics. As mentioned above, one approach is to empirically determine

eMagRes, Online © 2007 John Wiley & Sons, Ltd.

This article was previously published in the *Encyclopedia of Magnetic Resonance* in 2011 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470034590.emrstm0299.pub2

a calibration curve, another is to perform reconstructions using large values of λ to minimize the nonlinearity. For applications involving NUS, Hyberts and Wagner[22] proposed taking this latter step to the extreme: by constraining the IDFT of the spectrum to exactly match the measured data, they use entropy maximization to determine the values of the data at intervals not sampled. The result, which they call forward MaxEnt (FM), amounts to the MaxEnt spectrum in the limit of infinite λ. As a statistical method, this would be described as overfitting, since the spectrum will not be consistent with an independent remeasurement of the data. On the other hand, the results are highly linear and can be quantified without resorting to calibration.

8.3 Compressed Sensing

In the broadest sense, MaxEnt reconstruction is a regularization method that uses the regularization functional to ensure smooth reconstructions that are consistent with the measured data. Functionals other than entropy have been proposed, including the total power, area, and absolute magnitude of the spectrum, and in practice many of them give highly similar results. The term *compressed sensing* has emerged to describe one such approach that uses the l_1-norm as the regularization functional applied to NUS data.[23] Although formal and computational arguments have been advanced for various functionals, we find that for NUS data the distribution of sampling times has a far greater impact on the resulting spectra than does the precise form of the regularization functional.

9 REFERENCES

1. R. R. Ernst, in 'Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy', eds J. C. Hoch, F. M. Poulsen, and C. Redfield, Plenum Press: New York, 1991, pp. 1–25.
2. S. J. Wernecke and L. R. D' Addario, *IEEE Trans. Comput.*, 1977, **26**, 351–364.
3. R. A. Chylla and J. L. Markley, *J. Biomol. NMR*, 1995, **5**, 245–258.
4. G. L. Bretthorst, *J. Magn. Reson.*, 1990, **88**, 533–551.
5. V. A. Mandelshtam, H. S. Taylor, and A. J. Shaka, *J. Magn. Reson.*, 1998, **133**, 304–312.
6. S. Kim and T. Szyperski, *J. Am. Chem. Soc.*, 2003, **125**, 1385–1393.
7. E. Kupče and R. Freeman, *J. Am. Chem. Soc.*, 2003, **125**, 13958–13959.
8. V. Y. Orekhov, I. V. Ibraghimov, and M. Billeter, *J. Biomol. NMR*, 2001, **20**, 49–60.
9. C. E. Shannon, *Bell Syst. Tech. J.*, 1948, **27**, 379–423.
10. J. P. Burg, 'Maximum Entropy Spectrum Analysis', Stanford University: Stanford, CA, 1975.
11. J. C. Hoch, A. S. Stern, D. L. Donoho, and I. M. Johnstone, *J. Magn. Reson.*, 1990, **86**, 236–246.
12. J. A. Jones and P. J. Hore, *J. Magn. Reson.*, 1991, **92**, 363–376.
13. J. Skilling and R. Bryan, *Mon. Not. R. Astron. Soc.*, 1984, **211**, 111–124.
14. J. C. Hoch and A. S. Stern, '*NMR Data Processing*', Wiley-Liss: New York, 1996.
15. P. Schmieder, A. S. Stern, G. Wagner, and J. C. Hoch, *J. Magn. Reson.*, 1997, **125**, 332–339.
16. M. Mobli, M. W. Maciejewski, M. R. Gryk, and J. C. Hoch, *J. Biomol. NMR*, 2007, **39**, 133–139.
17. M. Mobli, M. W. Maciejewski, M. R. Gryk, and J. C. Hoch, *Nat. Methods*, 2007, **4**, 467–468.
18. D. L. Donoho, I. M. Johnstone, A. S. Stern, and J. C. Hoch, *Proc. Natl. Acad. Sci. U.S.A.*, 1990, **87**, 5066–5068.
19. N. Shimba, H. Kovacs, A. S. Stern, A. M. Nomura, I. Shimada, J. C. Hoch, C. S. Craik, and V. Dotsch, *J. Biomol. NMR*, 2004, **30**, 175–179.
20. N. Shimba, A. S. Stern, C. S. Craik, J. C. Hoch, and V. Dotsch, *J. Am. Chem. Soc.*, 2003, **125**, 2382–2383.
21. J. Tang and J. R. Norris, *J. Magn. Reson.*, 1988, **78**, 23–30.
22. S. G. Hyberts, G. J. Heffron, N. G. Tarragona, K. Solanky, K. A. Edmonds, H. Luithardt, J. Fejzo, M. Chorev, H. Aktas, K. Colson, K. H. Falchuk, J. A. Halperin, and G. Wagner, *J. Am. Chem. Soc.*, 2007, **129**, 5108–5116.
23. M. Lustig, D. Donoho, and J. M. Pauly, *Magn. Reson. Med.*, 2007, **58**, 1182–1195.

Acknowledgments

We are grateful to David Donoho, Ian Johnstone, Peter Schmieder, Gerhard Wagner, Peter Connolly, and Mark Maciejewski for many stimulating discussions and continuing fruitful collaborations on aspects of maximum entropy reconstruction in NMR. We also thank John Osterhout and Marcela Oslin for providing the sample of α_1 B, and George Maalouf and Peter Connolly for collecting NMR data.

Biographical Sketches

Jeffrey C. Hoch. *b*. 1954. A.B. Boston University, 1976; A.M. Harvard, 1978; PhD Harvard, 1983. Completed PhD studies with M. Karplus and C. M. Dobson. Member of the Rowland Institute for Science, 1981–2003. Since 2003, Professor of Molecular, Microbial, and Structural Biology and Director of the Gregory P. Mullen NMR Structural Biology Facility at UConn Health Center, and Director of the UConn Partnership for Excellence in Structural Biology. Founding Chair of the Gordon Research Conference on Computational Aspects of Biomolecular NMR. Research interests: NMR data processing, analysis, and protein structure determination.

Alan S. Stern. *b*. 1958. A. B. Harvard, 1979; PhD U.C. Berkeley, 1984. Member of Rowland Institute since 1988. Research interests: signal processing and mathematical logic.

Mehdi Mobli. Graduation in chemical engineering, Chalmers, Sweden. Post graduation on the calculation of chemical shifts in organic molecules with Professor R. J. Abraham, University of Liverpool, UK, before starting to work on signal processing of nonuniformly sampled data in multidimensional NMR with Professor J. C. Hoch (UConn, USA). Currently Senior Research Officer at the Institute for Molecular Bioscience at the University of Queensland (UQ) in Australia. Currently developing automated methods for NMR structure determination of toxin peptides utilizing nonuniform sampling and maximum entropy reconstruction, a project led by Professor G. F. King at UQ. Research interests: NMR data processing, analysis, and protein structural biology.

DOI: 10.1002/9780470034590.emrstm0299.pub2