

Published in final edited form as:

*J Magn Reson.* 2007 October ; 188(2): 295–300. doi:10.1016/j.jmr.2007.07.008.

## NMR Data Processing using Iterative Thresholding and Minimum $l_1$ -norm Reconstruction

Alan S. Stern<sup>\*</sup>, David L. Donoho<sup>†</sup>, and Jeffrey C. Hoch<sup>§</sup>

<sup>\*</sup>Rowland Institute at Harvard 100 Edwin H. Land Blvd. Cambridge, MA 02142 USA

<sup>†</sup>Department of Statistics Stanford University Stanford, CA 94305 USA

<sup>§</sup> Department of Molecular, Microbial, and Structural Biology University of Connecticut Health Center 263 Farmington Ave. Farmington, CT 06030-3305

### Abstract

Iterative thresholding algorithms have a long history of application to signal processing. Although they are intuitive and easy to implement, their development was heuristic and mainly ad hoc. Using a special form of the thresholding operation, called soft thresholding, we show that the fixed point of iterative thresholding is equivalent to minimum  $l_1$ -norm reconstruction. We illustrate the method for spectrum analysis of a time series. This result helps to explain the success of these methods and illuminates connections with maximum entropy and minimum area methods, while also showing that there are more efficient routes to the same result. The power of the  $l_1$ -norm and related functionals as regularizers of solutions to underdetermined systems will likely find numerous useful applications in NMR.

### Introduction

The computation of NMR spectra from short, noisy data records has long been a challenging problem. Procedures adopted from fields outside of NMR have proven to be superior to the discrete Fourier transform (DFT)[1, 2]; examples include maximum entropy (MaxEnt) [3, 4] and minimum-area[5] reconstruction, maximum likelihood reconstruction (MLM)[6, 7], and matrix methods such as LPSVD[8] and HSVD[9]. A method that is conceptually simpler and easier to implement than these methods is iterative thresholding[10-12]. This and related thresholding algorithms are widely used in the fields of image processing and fMRI. Iterative thresholding is a fixed-point technique; an operation is repeatedly applied until it no longer results in a change. The repeated operation has a particularly simple form: it involves setting all values below a threshold  $\tau$  (in absolute value) to zero, while leaving values above  $\tau$  unchanged.

Iterative thresholding algorithms applied to spectrum analysis have a strong heuristic appeal. Consider a free induction decay (FID) containing several exponentially decaying sinusoids with similar amplitudes, in which the length of the FID is short compared to the decay time of the sinusoids. A high-resolution estimate of the spectrum can be obtained by zero-filling

© 2007 Elsevier Inc. All rights reserved.

Correspondence to: Jeffrey C. Hoch.

Address correspondence to this author at hoch@uchc.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

the FID prior to discrete Fourier transformation, but the DFT spectrum will contain truncation artifacts. Choose a threshold value  $\tau$  that is smaller than the peak maxima, but larger than any truncation artifacts, and set every point in the DFT spectrum that is below  $\tau$  to zero, leaving the others unchanged. The inverse DFT of the thresholded spectrum will not agree very well with the input FID. However, if we consider only the part that extends the measured data, we may find that it is a more realistic extension than extending the data with zeros. Thus a better spectral estimate may be obtained by extending the FID with values from the inverse DFT of the thresholded spectrum, or equivalently, replacing the initial part of the inverse DFT with the original FID. This process of inverse Fourier transformation, replacement, Fourier transformation, and thresholding (Fig. 1) is repeated until there is no change in the spectrum, that is, until the fixed-point is reached.

We refer to the thresholding operation in which values above  $\tau$  are unchanged as *hard* thresholding. By contrast, in the *soft* thresholding operation, data with absolute values above  $\tau$  are reduced by  $\tau$  (with no change in complex phase). We show in the Appendix (Supplementary Material) that the fixed-point of iterative soft thresholding (IST) is also a minimum  $l_1$ -norm reconstruction (defined below). This equivalence has several implications. (1) Formal results on the properties of minimum  $l_1$ -norm reconstructions have been derived; no comparable results on the properties of the fixed-points of iterative thresholding procedures are available; (2) Logan's theorem [13] shows that under certain circumstances  $l_1$ -norm reconstruction is able to perfectly reconstruct the spectrum of a noisy signal; (3) The equivalence demonstrates how IST can be generalized to incorporate deconvolution and modified to improve convergence; (4) The equivalence illuminates similarities between IST and methods such as MaxEnt and minimum-area reconstruction. The power of the  $l_1$ -norm for regularizing reconstructions from sparse or noisy data has received considerable attention in the statistical and applied mathematics communities[13-16], and helps to explain the success of iterative thresholding methods.

We also note that fixed thresholding has been applied to the wavelet domain as a means for “denoising” NMR spectra[17-19]. While we consider thresholding in the frequency or Fourier domain in this work, similar results (the equivalence of iterative soft thresholding and  $l_1$ -norm regularization) apply to wavelet thresholding.

## Minimum $l_1$ -norm reconstruction

The  $l_1$ -norm of a spectrum  $\mathbf{f}$  is defined as

$$L(\mathbf{f}) = \|\mathbf{f}\|_1 \equiv \sum_{\omega=0}^{N-1} |f_{\omega}| \quad [1]$$

where  $N$  is the number of points in the complex vector  $\mathbf{f}$ . The goal of minimum  $l_1$ -norm reconstruction is to find the spectrum  $\mathbf{f}$  which minimizes  $L(\mathbf{f})$  subject to the constraint that  $\mathbf{f}$  is consistent with the experimental data. This constraint is expressed by the formula

$$C(\mathbf{f}) \equiv \frac{1}{2} \sum_{k=0}^{M-1} |\text{IDFT}(\mathbf{f})_k - d_k|^2 \leq C_0 \quad [2]$$

where  $M$  is the number of points in the complex FID  $\mathbf{d}$  and  $C_0$  is an estimate of the experimental error; IDFT is the inverse DFT. Minimum  $l_1$ -norm reconstruction is similar to minimum area reconstruction, proposed by Newman[5], the only difference being that

Newman's “area” amounts to  $\sum_{\omega} |\operatorname{Re}(f_{\omega})| + |\operatorname{Im}(f_{\omega})|$ , which is not invariant under changes of phase. As we shall see later, minimum  $l_1$ -norm also bears a resemblance to MaxEnt.

The properties of minimum  $l_1$ -norm reconstruction were extensively studied by Logan[20]. One of his most striking results is that under certain conditions, having to do with the relative sparsity of the noise in the FID and the peaks in the spectrum, minimum  $l_1$ -norm reconstruction can result in a *perfect* spectrum, with no residual noise. Unfortunately, Logan's conditions do not apply to real NMR data, but this result indicates the potential power of the technique.

The problem of determining the minimum  $l_1$ -norm reconstruction can be converted to an unconstrained optimization problem by introducing a Lagrange multiplier  $\tau$ . Let the objective function  $Q(\mathbf{f})$  be given by

$$Q(\mathbf{f}) \equiv \tau L(\mathbf{f}) + C(\mathbf{f}). \quad [3]$$

As we show in the Appendix (Supplementary Material), if  $\mathbf{f}$  is a minimum  $l_1$ -norm reconstruction, then  $\mathbf{f}$  is a minimum of  $Q$ . One technique for finding this minimum is to perform a gradient search. The gradient of  $Q$  is

$$\nabla Q(\mathbf{f}) = \tau \nabla L(\mathbf{f}) + \nabla C(\mathbf{f}). \quad [4]$$

Computing the gradient of  $L$  is straightforward. Expanding Eq. [1], and writing  $f_{\omega}^r$  and  $f_{\omega}^i$  for the real and imaginary parts of  $f_{\omega}$ , we obtain

$$L(f) = \sum_{\omega=0}^{N-1} |f_{\omega}|, \quad [5]$$

$$\frac{\partial L}{\partial f_{\omega}^r} = \frac{f_{\omega}^r}{|f_{\omega}|}, \quad \frac{\partial L}{\partial f_{\omega}^i} = \frac{f_{\omega}^i}{|f_{\omega}|} \quad [6]$$

We will adopt the convention that  $\partial L / \partial f_{\omega}$  is  $\partial L / \partial f_{\omega}^r + i \partial L / \partial f_{\omega}^i$ , which yields

$$\frac{\partial L}{\partial f_{\omega}} = \nabla L(\mathbf{f})_{\omega} = \frac{f_{\omega}}{|f_{\omega}|}. \quad [7]$$

The derivation of  $\nabla C(\mathbf{f})$  is more laborious; we simply state the result (details are given in (14)). Let  $\mathbf{F}$  be the  $(N \times N)$  unitary matrix corresponding to the DFT, and let  $\mathbf{K}$  be the  $(M \times N)$  projection matrix which shortens an  $N$ -element vector to its first  $M$  elements. Then

$$C(\mathbf{f}) = \frac{1}{2} \|\mathbf{K}\mathbf{F}^{\dagger}\mathbf{f} - \mathbf{d}\|^2 \quad [8]$$

and

$$\nabla C(\mathbf{f}) = \mathbf{F}\mathbf{K}^{\dagger} (\mathbf{K}\mathbf{F}^{\dagger}\mathbf{f} - \mathbf{d}), \quad [9]$$

where  $\dagger$  denotes the Hermitian transpose;  $\mathbf{F}^\dagger$  corresponds to the IDFT and  $\mathbf{K}^\dagger$  corresponds to zero-filling.

## Relationship of IST to minimum $l_1$ -norm reconstruction

The somewhat opaque expression in Eq. [9] belies the simplicity of the underlying operations. Re-writing Eq. [9] in operator notation,

$$\nabla C(\mathbf{f}) = \text{DFT} \{ \text{zerofill}(\text{shorten}(\text{IDFT}(\mathbf{f})) - \mathbf{d}) \} = \text{DFT} \{ \text{trunc}(\text{IDFT}(\mathbf{f})) - \text{zerofill}(\mathbf{d}) \} \quad [10]$$

where  $\text{trunc}(\mathbf{x}) = \text{zerofill}(\text{shorten}(\mathbf{x}))$  is the operation of setting the elements  $x_M, \dots, x_{N-1}$  to zero. Now we are in a position to see the unexpected relationship between the gradient of  $Q$  and the operations of IST. Soft thresholding by  $\tau$  is expressed by

$$\text{thresh}_\tau(\mathbf{f})_\omega = \begin{cases} f_\omega - \tau f_\omega / |f_\omega|, & \text{if } |f_\omega| > \tau \\ 0, & \text{if } |f_\omega| \leq \tau. \end{cases} \quad [11]$$

So for indices  $\omega$  at which  $|f_\omega| > \tau$ , soft thresholding is the same as subtracting  $\tau \nabla L$ . The replacement step of IST is expressed by

$$\mathbf{f} \rightarrow \text{DFT} \{ [\text{IDFT}(\mathbf{f}) - \text{trunc}(\text{IDFT}(\mathbf{f}))] + \text{zerofill}(\mathbf{d}) \} = \mathbf{f} - \text{DFT} \{ \text{trunc}(\text{IDFT}(\mathbf{f})) - \text{zerofill}(\mathbf{d}) \}. \quad [12]$$

Comparison with Eq. [10] shows that replacement is the same as subtracting  $\nabla C$ . So combining the replacement operation and the thresholding operation, we see that one iteration of IST corresponds to motion opposite the gradient of  $Q$ , and a fixed point of IST corresponds to a minimum of  $Q$ .

This description is not quite complete, since  $\nabla L$  is not defined for  $f_\omega = 0$  and soft thresholding is not the same as subtracting  $\tau f_\omega / |f_\omega|$  if  $|f_\omega| < \tau$ . Nevertheless, the description does suggest the relationship between IST and minimum  $l_1$ -norm reconstruction; the formal proof of their equivalence is given in the Appendix.

This equivalence shows how IST can be generalized to perform deconvolution. Deconvolving a decay  $\mathbf{w}$  involves reconstructing a spectrum whose IDFT, when weighted by  $\mathbf{w}$ , agrees with the experimental FID. This weighting can be incorporated into Eq. [2] by setting

$$C(\mathbf{f}) \equiv \frac{1}{2} \sum_{k=0}^{M-1} |w_k \text{IDFT}(\mathbf{f})_k - d_k|^2. \quad [13]$$

The corresponding modification to Eqs. [8] and [9] simply involve setting the diagonal elements of the matrix  $\mathbf{K}$  equal to  $\mathbf{w}$ . The only change needed in IST is to adjust the replacement operation so that it coincides with motion opposite the gradient of the revised constraint. In operator notation, this becomes

$$\mathbf{f} \rightarrow \mathbf{f} + \text{DFT} \{ \text{zerofill} [ \text{weight} [ \mathbf{d} - \text{weight}(\text{shrink}(\text{IDFT}(\mathbf{f}))) ] ] \} \quad [14]$$

where the weight operator corresponds to pointwise multiplication by the decay  $\mathbf{w}$ .

IST is capable of reconstructing spectra for non-linearly sampled data (15). Indeed, this can be viewed as a special case of deconvolution: the weights for points not sampled are simply set to zero.

## Relationship of IST to MaxEnt

The equivalence of IST and minimum  $l_1$ -norm reconstruction also makes clear the similarity to MaxEnt. The basic aim of MaxEnt reconstruction is the same as that of minimum  $l_1$ -norm reconstruction, except the regularization functional is the entropy, rather than the  $l_1$ -norm. The entropy functional  $S$  is approximately given by

$$S(\mathbf{f}) = - \sum_{\omega=0}^{N-1} |f_{\omega}| \log |f_{\omega}|. \quad [15]$$

The main difference between the entropy and the  $l_1$ -norm is the opposite sign; consequently maximizing the entropy is very similar to minimizing the  $l_1$ -norm. It has been shown that under certain circumstances, the action of MaxEnt reconstruction is equivalent to a nonlinear scaling of the DFT spectrum[21]. The nonlinear scaling reduces the absolute value of each point in the spectrum; small values are scaled down proportionally more than large values. The similarity to soft thresholding is clear.

## Convergence of IST

While the formal results presented in the appendix show that the fixed-point of IST is the minimum  $l_1$ -norm reconstruction, they say nothing about how quickly IST converges. Indeed, the finite precision of computer arithmetic makes it entirely possible that the step size may reach machine zero — a value smaller than the smallest non-zero number that can be represented — well before the minimum of  $Q$  has been reached. Should this happen, the algorithm would appear to have converged, but it would not produce the correct result. In testing for convergence, therefore, it is important that we monitor not only the step size, defined by

$$Stepsize = \left[ \frac{1}{N} \sum_{\omega=0}^{N-1} |f_{\omega}(i+1) - f_{\omega}(i)|^2 \right]^{1/2}, \quad [16]$$

where  $f_{\omega}(i)$  is the  $\omega^{\text{th}}$  element of  $\mathbf{f}$  at iteration  $i$ , but also the quantity

$$Test = \left| \frac{\nabla L}{|\nabla L|} \right| - \left| \frac{\nabla C}{|\nabla C|} \right|, \quad [17]$$

which is equal to zero only at a minimum of  $Q$ . (At indices  $\omega$  for which  $f_{\omega}=0$ , the gradient of  $L$  is not defined; these indices can simply be ignored in calculating  $Test$ .)

Figure 2 shows various spectrum reconstructions of a  $f_1$  column from a NOESY data set (following processing in  $f_2$ ) in which the diagonal resonance is about 10 times more intense than the off-diagonal resonances. Row A is the unapodized, zero-filled DFT. The result of IST applied for 800 steps is shown in row B;  $Stepsize$  became negligibly small, but not zero. Row C shows the results of a modified algorithm in which the change during each iteration is  $-\beta(i)\nabla Q$ , instead of  $-\nabla Q$ , where  $\beta(i)$  is a scale factor chosen to minimize  $Q$  at iteration  $i$ ;

it can be found by a simple line search (we refer to this algorithm as line-search IST, as opposed to simple IST).

Figure 3 shows  $Q$ ,  $Test$ , and  $Stepsize$  as a function of  $i$ . Since IST moves opposite  $\nabla Q$ , the value of  $Q$  always decreases monotonically, in contrast to  $Test$  and  $Stepsize$ . The non-monotonic behavior of  $Stepsize$  and  $Test$  for line-search IST reflects well-known deficiencies of gradient descent[22]. Figure 3 also shows that merely monitoring  $Stepsize$  is not a safe way to test for convergence, since it becomes very small for simple IST even while  $Q$  is still changing.  $Test$  has not converged to zero for simple IST even after 800 iterations. In contrast,  $Test$  converges to zero for line-search IST at about the time that  $Stepsize$  becomes very small. The final values for  $Q$  are 14830 and 14665 for simple IST and line-search IST, respectively.

## Robustness of Minimum $l_1$ -norm Reconstruction

Figure 2 illustrates that IST is capable of suppressing the truncation artifacts typical of zero-filled DFT spectra. For comparison, row D of Figure 2 shows a MaxEnt reconstruction. The most striking difference is that minimum  $l_1$ -norm reconstruction apparently “resolves” fine structure in the peaks that is not evident at all in the MaxEnt reconstruction. A simple test demonstrates that the additional structure in the minimum  $l_1$ -norm reconstruction is not correct. Figure 4 contains IST and MaxEnt reconstructions for the same data as Figure 3, except that a single synthetic decaying sinusoid has been added to the time domain data prior to reconstruction. Row A shows that IST yields an artifactual split line, while MaxEnt (row B) correctly yields a single peak. The reconstructions agree equally well with the time domain data: they have identical values of  $C(\mathbf{f})$ .

In addition to proving more robust, MaxEnt reconstruction is more efficient. The modified Cambridge algorithm used to compute the MaxEnt reconstruction (14, 18) involves 8 discrete Fourier transformations during each iteration (except the first), plus several other matrix operations. Line-search IST requires two DFT's per iteration. IST typically requires on the order of ten times as many iterations as MaxEnt; the reconstructions in Fig. 4 used 800 iterations of line-search IST and 40 iterations of MaxEnt, and the MaxEnt processing required approximately one-sixth the computer time of IST. The difference in efficiency can be attributed mainly to the sophisticated optimizer used in the Cambridge algorithm, which incorporates elements of conjugate-gradient and variable metric optimizers, in contrast to the simple gradient-descent of line-search IST. In principle, computation of minimum  $l_1$ -norm reconstructions could be made comparably efficient to MaxEnt reconstruction by the use of more sophisticated search techniques.

## Concluding Remarks

Iterative thresholding algorithms have proven to be popular because of their simplicity. We have shown that iterated soft thresholding leads to the computation of the minimum  $l_1$ -norm reconstruction, and is closely related to MaxEnt and minimum-area reconstruction. Although simple implementations of minimum  $l_1$ -norm reconstruction and MaxEnt reconstruction based on more powerful optimization techniques can yield strikingly different results, the differences appear to be due to the relative robustness of the optimizers employed, and not the objective functions. The relationship among thresholding, minimum  $l_1$ -norm reconstruction, and MaxEnt reconstruction demonstrated here provides an avenue for unifying these different approaches to the problem of spectrum reconstruction and deconvolution.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

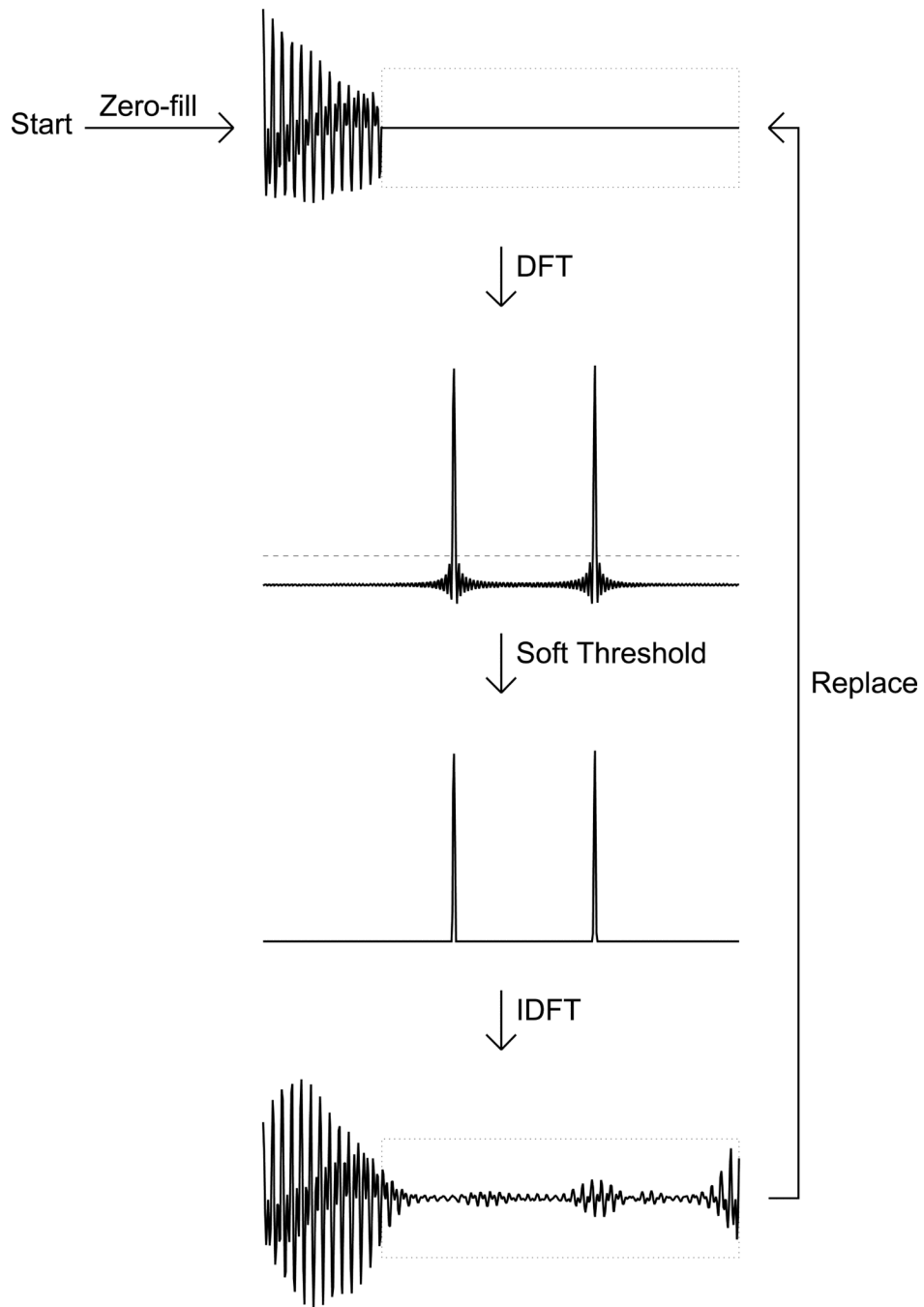
This is a contribution from the NMRA Consortium. We thank Peter Connolly for supplying experimental data. This work was supported by the Rowland Institute at Harvard, and by grants from NIH (GM-47467, RR-20125, and GM-72000).

## References

1. Stephenson DS. Linear Prediction and Maximum Entropy Methods in NMR Spectroscopy. *Prog. NMR Spec.* 1988; 20:515–626.
2. Hoch JC. Modern Spectrum Analysis in NMR: Alternatives to FT. *Meth. Enzym.* 1989; 176:216–241. [PubMed: 2811688]
3. Gull SF, Daniell GJ. Image reconstruction from incomplete and noisy data. *Nature.* 1978; 272:686–690.
4. Hoch, JC.; Stern, AS. Maximum Entropy Reconstruction in NMR. In: Grant, DM.; Harris, RK., editors. *Encyclopedia of NMR.* John Wiley & Sons; New York: 1996. p. 2980-2988.
5. Newman RH. Maximization of entropy and minimization of area as criteria for NMR signal processing. *J. Magn. Reson.* 1988; 79:448–460.
6. Bretthorst GL. Bayesian Analysis I. Parameter estimation using quadrature NMR models. *J. Magn. Reson.* 1990; 88:533–551.
7. Chylla RA, Markley JL. Improved frequency resolution in multidimensional constant-time experiments by multidimensional Bayesian analysis. *J. Biomol. NMR.* 1993; 3:515–33. [PubMed: 8219738]
8. Barkhuijsen H, et al. Application of linear prediction and singular value decomposition (LPSVD) to determine NMR frequencies and intensities from the FID. *Magn. Reson. Med.* 1985; 2:86–9. [PubMed: 3831681]
9. Barkhuijsen H, de Beer R, van Ormondt D. Improved Algorithm for Noniterative Time-Domain Model Fitting to Exponentially Damped Magnetic Resonance Signals. *J. Magn. Reson.* 1987; 73:533–557.
10. Papoulis A. A new algorithm in spectral analysis and band-limited extrapolation. *IEEE Trans. Circ. Syst.* 1975; CAS-22:735–742.
11. Jansson PA, Hunt RH, Pyle EK. Resolution enhancement of spectra. *J. Opt. Soc. Am.* 1970; 60:596–599.
12. van Cittert PH. Zum Einfluss der Spaltbreite auf die Intensitätsverteilung in Spektrallinien. *Z. Physik.* 1931; 69:298–308.
13. Donoho DL, Logan BF. Signal Recovery and the Large Sieve. *SIAM J. Appl. Math.* 1992; 52:577–591.
14. Donoho DL, Huo X. Uncertainty Principles and Ideal Atomic Decomposition. *IEEE Trans. Info. Theory.* 2001; 47:2845–2862.
15. Donoho DL, et al. Maximum Entropy and the Nearly Black Object (with discussion). *J. Roy Stat Soc B.* 1992; 54:41–81.
16. Donoho DL, Stark PB. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.* 1989; 49:906–931.
17. Ahmed OA. New denoising scheme for magnetic resonance spectroscopy signals. *IEEE Trans Med Imaging.* 2005; 24:809–16. [PubMed: 15957602]
18. DANCEA F, GUNTHER U. Automated protein NMR structure determination using wavelet de-noised NOESY spectra. *J. Biomol NMR.* 2005; 33:139–52. [PubMed: 16331419]
19. Trbovic N, et al. Using wavelet de-noised spectra in NMR screening. *J Magn Reson.* 2005; 173:280–7. [PubMed: 15780919]

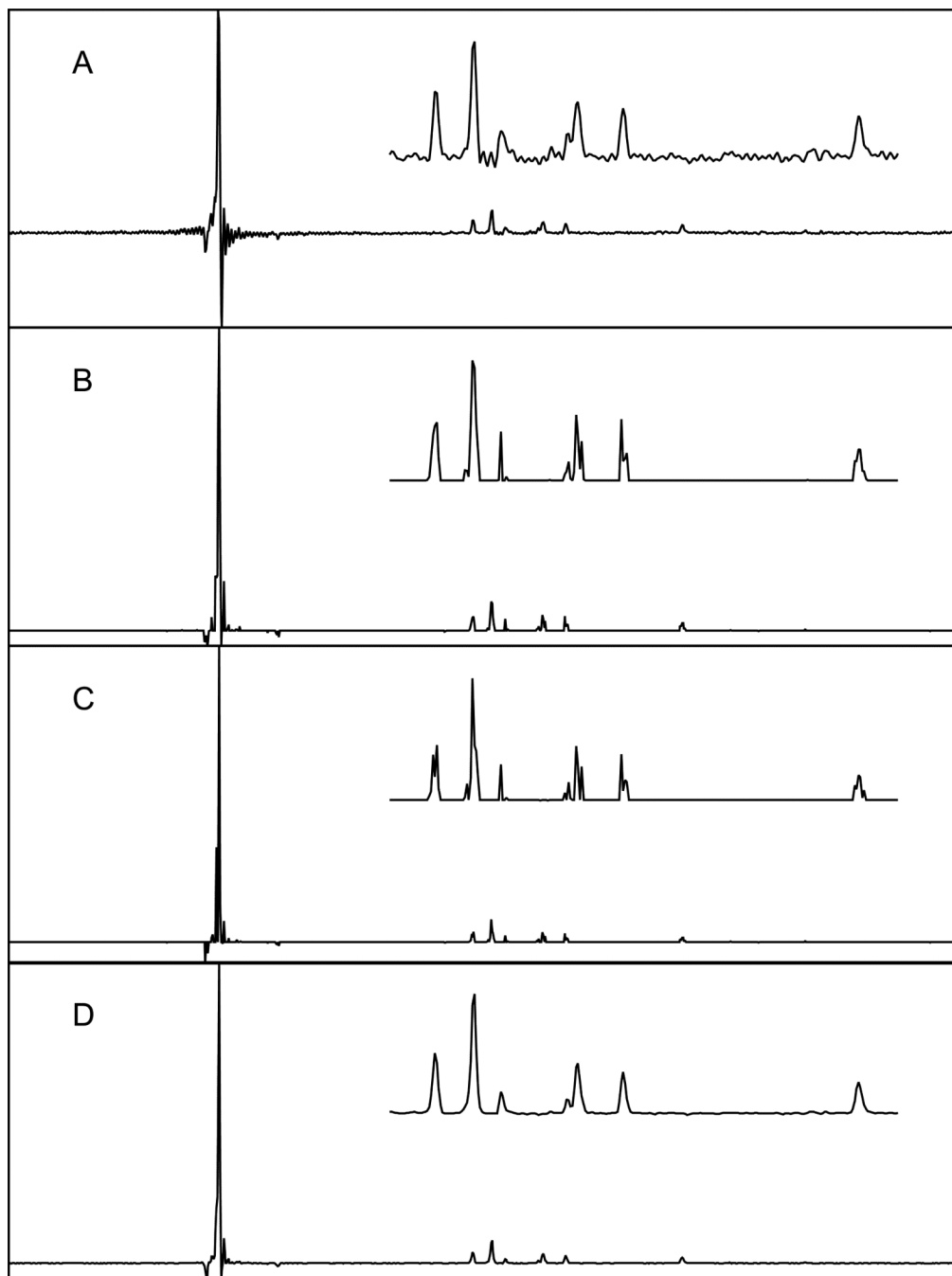
20. Logan, BF. Ph.D. thesis. Department of Electrical Engineering, Columbia University; 1965. Properties of high-pass signals.
21. Donoho DL, et al. Does the maximum entropy method improve sensitivity? Proc. Natl. Acad. Sci. USA. 1990; 87:5066–5068. [PubMed: 11607089]
22. Press, WH., et al. Numerical Recipes in Fortran. Cambridge University Press; Cambridge: 1992.



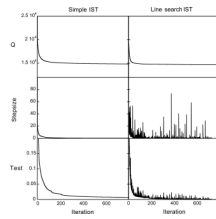


**1.**

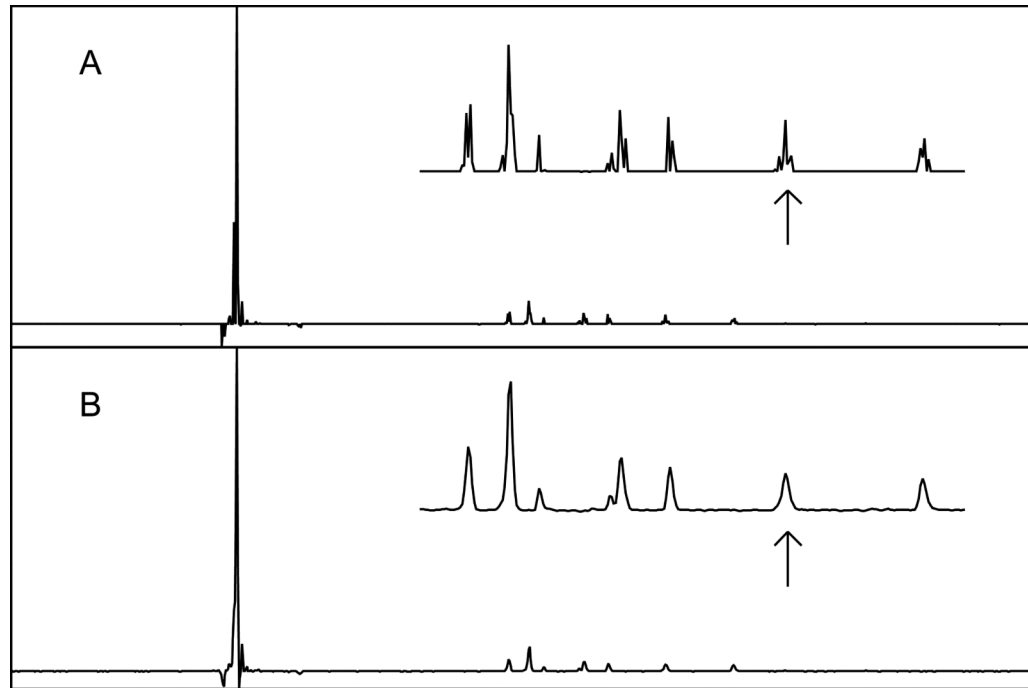
A schematic diagram of IST. The initial trial spectrum is the DFT of the zero-filled data. Subsequent trial spectra are computed by applying the soft thresholding operation: setting points below a threshold to zero, and subtracting the threshold from all other points. The result is inverse Fourier transformed, the “tail” is used to extend the measured data, and the augmented data is forward Fourier transformed and thresholded.



2.  
Spectral reconstructions of an  $f_1$  column from NOESY data for a 66-residue protein. Row A is the unapodized zero-filled DFT spectrum; rows B and C show the results of 800 iterations of simple and line-search IST, respectively. D is the result of 40 iterations of MaxEnt reconstruction.



3. Testing for convergence of IST by plotting the values of the objective function  $Q$ ,  $Test$ , and  $Stepsize$ . Note that for simple IST,  $Q$  and  $Test$  continue to decrease even after  $Stepsize$  has approached zero.



4. Reconstructions of the same data shown in Figure 2, except that a single decaying sinusoid was added to the time domain data. The artificial peak is indicated by the arrow. A) 800 iterations of IST. B) MaxEnt reconstruction.