

## Research Article

# Fast $\ell_1$ Minimization by Iterative Thresholding for Multidimensional NMR Spectroscopy

Iddo Drori

*Department of Statistics, Stanford University, Stanford, CA 94305-4065, USA*

Received 18 September 2006; Revised 5 April 2007; Accepted 28 August 2007

Recommended by Sabine Van Huffel

Fast multidimensional NMR is important in chemical shift assignment and for studying structures of large proteins. We present the first method which takes advantage of the sparsity of the wavelet representation of the NMR spectra and reconstructs the spectra from partial random measurements of its free induction decay (FID) by solving the following optimization problem:  $\min \|x\|_1$  subject to  $\|y - SF^T W^T x\|_2 \leq \epsilon$ , where  $y$  is a given  $n \times 1$  observation vector,  $S$  a random sampling operator,  $F$  denotes the Fourier transform, and  $W$  an orthogonal 2D wavelet transform. The matrix  $A = SF^T W^T$  is a given  $n \times p$  matrix such that  $n < p$ . This problem can be solved by general-purpose solvers; however, these can be prohibitively expensive in large-scale applications. In the settings of interest, the underlying solution is sparse with a few nonzeros. We show here that for large practical systems, a good approximation to the sparsest solution is obtained by iterative thresholding algorithms running much more rapidly than general solvers. We demonstrate the applicability of our approach to fast multidimensional NMR spectroscopy. Our main practical result estimates a four-fold reduction in sampling and experiment time without loss of resolution while maintaining sensitivity for a wide range of existing settings. Our results maintain the quality of the peak list of the reconstructed signal which is the key deliverable used in protein structure determination.

Copyright © 2007 Iddo Drori. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

High-dimensional NMR spectroscopy is time consuming, requiring between days and weeks of acquisition time [1]. Numerous researchers are interested in reconstructing NMR spectra from undersampled signals [2–5]. For example, strategies include sampling along random lines in the indirect dimension and along radial lines. In such methods there are several observations per degree of freedom in the underlying object. Simple linear methods for reconstruction such as backprojection lead to artifacts when used in an undersampled setting. More elaborate methods such as maximum entropy suppress artifacts in a nonuniform fashion.

In this work we bring two new elements into play. The first is the idea that there is a rigorously valid sense in which NMR spectra are reconstructed from undersampled data, which includes two components: (i) sparsity: representing spectra in an orthogonal wavelet basis which results in a sparse representation of the signal; and (ii) nonlinear reconstruction: application of an optimization criterion which produces a sparse set of representation coefficients. The limitation of this approach is computational complexity; there-

fore, the second idea is that for solving these massive optimization problems, there are fast iterative algorithms based on thresholding which are used for reconstruction. In this paper, we combine these ideas and demonstrate effective reconstruction of 2D NMR spectra using significantly undersampled data. We experiment with various sampling patterns and perform a detailed comparison of different reconstruction methods.

The wavelet transform is well known and efficiently used for signal compression. It transforms the signal into a domain in which only a few significant coefficients are required for reconstruction. Considering that a signal to be acquired is sparse under some transform, the idea of this work is to utilize the sparse representation of the NMR signal in some domain to sample in advance only a subset of the signal, thus reconstructing the original signal from a small set of measurements.

Our work is motivated by fast methods for iterative image reconstruction from random samples [6] and operations performed on a sparse set of wavelet coefficients [7]. More recently, the notions of compressed sensing and compressive sampling [8–10] have been formulated—a signal which is

compressible in a known basis (e.g., wavelet or Fourier) can be reconstructed from fewer measurements than the nominal sampling density, provided that the samples are made on a specially transformed version of the signal. Roughly speaking, the samples measure linear functionals which look like random linear combinations of the basis functions and reconstruction involves quadratic programming.

In this work, we use the sparsity of the NMR signal inherent in the transform domain to sample the signal and reconstruct it. The original signal is recovered by using a nonlinear reconstruction scheme—an approximation to  $\ell_1$  minimization of the transform.

## 2. $\ell_1$ MINIMIZATION

Formally, the problem we wish to solve is finding the sparsest solution to an underdetermined system of equations:

$$(P_0) \min \|x\|_0 \text{ subject to } y = Ax, \quad (1)$$

where  $y$  is observed data,  $A$  is a known  $n \times p$  matrix,  $n < p$ , and  $x$  is an unknown vector in  $\mathbf{R}^p$ . Here  $\|x\|_0$  represents the number of nonzeros in  $x$  and is not a norm. This is a nonconvex combinatorial optimization problem. In general, finding the sparsest solution is NP hard [11]. Traditionally, there are three approaches to getting around NP hard problems: if the inputs are very small, an algorithm with exponential time may be satisfactory although not elegant; we may be able to isolate special cases that are feasible in polynomial time; or it may be possible to find a near optimal solution in polynomial time using an approximation algorithm. Therefore, we solve the problem for the  $\ell_1$  norm [12]:

$$(P_1) \min \|x\|_1 \text{ subject to } y = Ax, \quad (2)$$

which is convex and tractable. In addition, when the solution is sufficiently sparse for certain levels of underdeterminicity there exists equivalence between the  $\ell_1$  and sparsest solutions [13].

The notion of using  $\ell_1$ -penalization as a sparsifying constraint has been around for decades. In the context of statistical estimation, this has been made explicit by Efron and Tibshirani [14, 15] who suggested the following  $\ell_1$ -constrained minimization:

$$(L_t) \min_x \|y - Ax\|_2^2 \text{ subject to } \|x\|_1 \leq t. \quad (3)$$

This problem can also be written in the augmented Lagrangian form,

$$(D_\lambda) \min_x \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1. \quad (4)$$

Problems  $(L_t)$  and  $(D_\lambda)$  are equivalent. Indeed, it is easy to verify that for  $x_t^*$ , a solution of  $(L_t)$  for some  $t \geq 0$ , there exists a  $\lambda \geq 0$  such that  $x_t^*$  solves  $(D_\lambda)$ .

Problem  $(D_\lambda)$  has been introduced to the signal processing community by Chen et al. [12] with the name basis pursuit denoising (BPDN) [16, 17]. This is equivalent to a linear objective perturbed by a quadratic term, yielding thus a

quadratic program, retaining structure similar to linear programming:

$$\min c^T x + \frac{1}{2} \|p\|^2 \text{ subject to } \Phi x + p = y, \quad x \geq 0, \quad (5)$$

where  $c^T x$  is the inner product of two vectors with  $c$  a vector of ones,  $\Phi x$  is a matrix-vector product such that  $\Phi = (A, -A)$ , and  $x \geq 0$  means that each entry of the vector  $x$  must be nonnegative.

If  $A$  is orthogonal, then  $\|y - Ax\|_2 = \|\hat{y} - x\|_2$ , for  $\hat{y} = A^T y$ , and problem  $(D_\lambda)$  is equivalent to

$$\min_x \frac{1}{2} \sum_i (\hat{y}_i - x_i)^2 + \lambda \sum_i |x_i| \quad (6)$$

written coordinate-wise, and therefore we get the soft thresholding nonlinearity which solves the scalar minimization problem:

$$\delta_\lambda(y) = \frac{1}{2} \arg \min_x (y - x)^2 + \lambda |x| \quad (7)$$

and is the soft thresholding operator:

$$(\delta_\lambda(y))_i = \text{sgn}(y_i) (|y_i| - \lambda)_+. \quad (8)$$

It is worth noting that  $(L_t)$  has been studied by Tibshirani and others in the case  $n > p$ , that is  $Ax = y$  is an overdetermined system of equations, whereas  $(D_\lambda)$  was introduced [12] and analyzed [18] for  $n < p$ .

In most practical settings, we observe noisy data and would like to solve the problem in the presence of noise [16]:

$$(P_{1,\varepsilon}) \min \|x\|_1 \text{ subject to } \|y - Ax\| \leq \varepsilon, \quad (9)$$

which is equivalent to the optimization problem  $(D_\lambda)$ . The optimization problems  $(P_1)$  and  $(P_{1,\varepsilon})$  can be cast as a perturbed linear program which is a quadratic program, and therefore computationally tractable, and solved efficiently using general purpose solvers such as simplex and interior point methods [19, 20]. The simplex algorithm moves along the exterior of the feasible region maintaining a solution which is a vertex of the simplex, often solving linear programs quickly in practice, however the algorithm can require exponential time for specific inputs. In contrast, interior point methods move along the interior of the feasible region maintaining intermediate solutions which are not necessarily vertices of the simplex, and run in polynomial time.

However, for many signal and image processing applications as well as for NMR applications, the problem sizes are quite large, hence computational complexity is a driving concern for these applications. Standard general-purpose solvers such as the simplex and interior point methods can be used; however, general-purpose solvers ultimately require solution of “full” linear systems which may require cubic computation of the order  $O(np^2)$  which is prohibitive for large-scale applications. General-purpose optimizers, while extremely useful for establishing the validity of  $\ell_1$ -based methods, must give way in actual applications to specialized optimizers. In this work, we describe such specialized efficient algorithms which solve these problems in quadratic time of the input size.

Recently,  $\ell_1$ -norm minimization problems have been applied to a range of important practical applications, particularly in conjunction with sparse representation [10, 12, 17]. The  $\ell_1$ -norm minimization is a way to attempt to regularize the solution. Applications have been proposed in the context of time-frequency representation [12], overcomplete signal representation [17], compressed sensing [9, 10], MRI [21–23], removing impulsive noise [24], error-correcting codes [25, 26], and genome-wide analysis of mRNA lengths [27].

### 3. APPLICATION

An important motivation for multidimensional nuclear magnetic resonance (NMR) spectroscopy is to determine the structure of proteins. Briefly, NMR occurs when atomic nuclei immersed in a magnetic field are exposed to an oscillating magnetic field. When a substance is placed in a magnetic field, some nuclei have orientations called spins. Each spin corresponds to a different energy level and the spins jump between levels when excited by radio waves whose frequency matches the energy spacing, which is called resonance. By changing frequency, the energy level spacings are measured, and a signal is created when the spins flip at resonance. The NMR spectrum shows the magnitude of the signal as a function of frequency.

Numerous processing and analysis steps are performed: beginning with the initial recording of the free induction decay (FID), computation of the spectrum, structure interpretation and assignment. Several of the key desired properties in this chain of operations are as follows: (i) feasible acquisition time for high-dimensional experiments required for handling large proteins, (ii) identifying exact peak positions, and (iii) accurately determining structure from a peak list and additional information. In this work, we focus on the first point which is the most critical bottleneck in practical settings.

Our approach is a general method for fast multidimensional NMR by random sampling and fast  $\ell_1$  norm reconstruction by iterative methods. Our main practical result estimates a four-fold reduction in sampling and experiment time without loss of resolution while maintaining sensitivity for a wide range of existing settings. This is important for chemical shift assignment and for studying protein structure.

Recent fast multidimensional NMR methods [28, 29] include parallel acquisition schemes, selective recording of output spectra, reconstruction from a small set of projections, and reconstruction from random and nonuniform measurements.

Replacing serial acquisition in the indirect dimension [30] with parallel acquisition by applying gradients along the  $z$ -axis results in a method for obtaining 2D data in a single scan. This high throughput hardware solution is extended to reconstruction of 3D and 4D spectra by varying gradients of additional axes in encoding. GFT-NMR [1] obtains a full high-dimensional spectrum from a set of low-dimensional spectra by coupling evolution of nuclei thereby reducing the number of indirect dimensions. The approach recovers one  $d$ -dimensional spectrum from  $2^{d-k+1}k$ -dimensional spectra, with  $d > k$ , and involves a least squares fit to an overdeter-

mined system of equations. The application of GFT-NMR as demonstrated for the protein ubiquitin results in a reduction of experiment time by an order of magnitude for each reduced indirect dimension.

Similar to reconstruction methods used in tomography, a full multidimensional spectrum can be reconstructed from a small set of projections [31, 32], utilizing the Fourier projection-slice theorem. More specifically, recent experiments include reconstruction of 3D spectra of the small ubiquitin protein and long protein HasA, in which a spectrum is reconstructed from  $(t_1, t_3)$ ,  $(t_2, t_3)$  projections as well as 30, 60 degree projections with respect to the  $t_1$  axis.

In Hadamard spectroscopy, the evolution periods of the pulse sequences are replaced by arrays which cover a narrow range of the full frequency range. This setting reduces experiment time for exploring specific sites. For example, an experiment for recovering a subset of peaks in a spectrum of the ubiquitin protein results in a speedup by two orders of magnitude compared with full acquisition.

Multiway decomposition [33] represents the spectrum as a multilinear form consisting of line shapes of peaks. For example, in the 2D case, the spectrum is represented as  $XAY^T$ , where  $A$  is a matrix of amplitudes, and  $X, Y$  are shape matrices with various constraints such as nonnegativity, symmetry, and orthogonality. This global parametric representation results in filling in a decomposed sparsely sampled signal.

Maximum entropy is often used for reconstruction [2–5] of a randomly sampled subset of FID measurements in the indirect dimension. As expected, recovery of 3D protein spectra from 20–33 percent sampling of the  $(t_1, t_2)$  plane reduces experiment time by factors of 3–4. Both maximum entropy and  $\ell_1$  norm reconstruction use similar sampling patterns; however, our approach is the first to utilize the underlying sparsity of the NMR spectra in the wavelet domain. An FID has a known mathematical formulation involving a few parameters, however this representation is nonlinear. The representation in the wavelet domain has the advantage of being a linear representation of the signal. The wavelet transform has been previously used for smoothing and denoising NMR spectra [30]. In this work, the wavelet transform is important for obtaining a sparse representation of the signal, and the specific transform for a rapid decay of the wavelet coefficients.

### 4. NUMERICAL SCHEME

In order to reconstruct an NMR spectrum from partial random measurements of its free induction decay, we solve the following optimization problem:

$$\min \|x\|_1 \text{ subject to } \|y - SF^T W^T x\| \leq \epsilon, \quad (10)$$

where  $y$  is the observation,  $S$  is a random sampling operator,  $F$  denotes the 2D Fourier transform, and  $W$  an orthogonal 2D wavelet transform.

We apply iterative soft thresholding by the iteration

$$x^{k+1} = x^k + \delta_{t_k} (A^* (y - A(x^k))), \quad (11)$$

where  $x^k$  is the  $k$ th approximate solution,  $\delta_t$  is soft thresholding at amplitude  $t$  and the threshold  $t_k$  decreases with

**Input:**  $n \times p$  matrix  $A$ , observation vector  $y$ ,  $\varepsilon$ .  
**Output:** solution of  $(P_{1,\varepsilon})$ .  
**Algorithm:**  
Initialization:  
Residual  $r_0 = y$ .  
Correlation  $c_0 = A^T r_0$ .  
Threshold  $t_0 = \max(|c_0|)$ .  
Solution  $x_0 = 0$  of length  $n$ .  
while  $\|r_k\|_2 \geq \varepsilon \|y\|_2$ :  
  update **solution**  $x_{k+1} = x_k + \delta_{t_k}(c_k)$ .  
  compute **residual**  $r_{k+1} = y - Ax_{k+1}$ .  
  compute **correlation**  $c_{k+1} = A^T r_{k+1}$ .  
  update **threshold**  $t_{k+1} = \mu t_k$ .  
end while

ALGORITHM 1: Iterative soft thresholding pseudocode.

increasing iteration count. The “fast operator” composing random sampling, Fourier transform, and the wavelet transform is denoted by  $A$  and the conjugate by  $A^*$ . More specifically, we use an orthonormal 2D wavelet transform with 8 vanishing moments.

As mentioned, there exist general purpose solvers for finding the solution of the problem  $(P_{1,\varepsilon})$ ; thus, for large-scale applications, we introduce efficient special purpose solvers. In addition, recent work [34] has focused on finding good sparse approximations for underdetermined linear systems of equations for typical/random matrices; whereas in the next section, we describe in detail fast iterative thresholding algorithms suitable for large-scale practical applications.

#### 4.1. Iterative soft thresholding

Iterative soft thresholding, in various guises, has been used for years [30, 35, 36]. A set of thorough analysis works [37–41] offer algorithms and proof of their convergence to the global minima of the problems. While not all of these are exactly comparable to the our setting, the family resemblance should be clear. In our work, the process of iterative soft thresholding approximately solves  $(P_{1,\varepsilon})$  when the solution is sufficiently sparse and  $p$  and  $n$  are sufficiently large.

A fast solution to the  $(P_{1,\varepsilon})$  optimization problem is obtained by a simple iterative soft thresholding (IST) algorithm. Let  $\delta_t(y)$  denote the soft thresholding operator:

$$(\delta_t(y))_i = \text{sgn}(y_i)(|y_i| - t)_+. \quad (12)$$

Consider the iteration

$$x^{k+1} = x^k + \rho \delta_{t_k}(A^T(y - Ax^k)). \quad (13)$$

Here  $0 < \rho \leq 1$ , we start this iteration from  $x^0 = 0$ , and  $t_k$  decreases from iteration to iteration. Each iteration requires applications of  $A$  and  $A^T$ ; if these can be performed rapidly, then each step of IST will run rapidly.

**Input:**  $n \times p$  matrix  $A$  and observation vector  $y$ ,  $\varepsilon$ .  
**Output:** solution of  $(P_{1,\varepsilon})$ .  
**Algorithm:**  
Initialization:  
Residual  $r_0 = y$ .  
Correlation  $c_0 = A^T r_0$ .  
Threshold  $t_0 = \max(|c_0|)$ .  
Solution  $x_0 = 0$  of length  $n$ .  
while  $\|r_k\|_2 \geq \varepsilon \|y\|_2$ :  
  compute **support**  $I = \{i : |c_k| \geq t_k - \varepsilon\}$ .  
  **solve** by least squares  $r_k = A_I dx_I$ .  
  update **solution**  $x_{k+1} = x_k + \gamma dx$ .  
  compute **residual**  $r_{k+1} = y - Ax_{k+1}$ .  
  compute **correlation**  $c_{k+1} = A^T r_{k+1}$ .  
  update **threshold**  $t_{k+1} = \mu t_k$ .  
end while

ALGORITHM 2: Iterative thresholding by least squares pseudocode.

We view the sequence of iterates  $x^k$  as approximately following the LARS [14] path at threshold  $t_k$ . Indeed, this is a sort of fixed point iteration for the solution  $\tilde{x}_t$  which satisfies

$$0 = \delta_t(A^T(y - A\tilde{x}_t)) \quad (14)$$

which is related to the normal equations  $A^T A x = A^T y$ , and so, starting from a hypothetical point on the LARS path itself, the iteration would produce no change. Moreover, small perturbations away from the LARS path produce countervailing adjustments.

Algorithm 1 describes the pseudocode of the algorithm. In iterative soft thresholding,  $A$  and  $A^T$  are either applied directly or as fast operators. It is an iterative algorithm involving applications of  $A$  and  $A^T$  which converges rapidly. A variation on this is a block solution which accelerates the basic soft threshold iteration. The matrix  $A$  is partitioned into blocks  $A = [B_1, B_2 \dots B_j]$  by taking random disjoint columns. Then the least-squares projection of  $B_j$  is applied in computing the correlations:

$$x^{k+1,j} = x^{k,j} + \rho \delta_{t_k}((B_j^T B_j)^{-1} B_j^T (y - Ax^k)). \quad (15)$$

#### 4.2. Iterative thresholding by least squares

A variant of iterative thresholding, differing in two lines of code, begins with an empty model, an initial estimate of the solution as zeros, and the residual as the observation. We proceed in an iterative fashion, finding the correlations above a threshold and solve a least-squares problem in each iteration. This means that in each iteration, as the threshold is decreased, the algorithm solves a least-squares problem using an iterative conjugate-gradient-like method [42] with an increasing number of elements on a larger space. Here,  $dx_j$  is defined as the least-squares solution of a linear system, such that the elements of the vector  $dx$  outside the support  $I$  are zeros. We have empirically found the choice of  $\mu + \gamma = 1$  to be suitable for our purposes. Algorithm 2 describes the pseudocode of the algorithm.

### 4.3. Computational complexity

In this section, we describe the time and space complexity of our algorithms in various settings. In the case where  $A$  is an explicit  $n \times p$  matrix such that  $n < p$ , iterative soft thresholding described in Algorithm 1 computes at each iteration  $k$  the residual  $r = y - Ax$  and correlation vectors  $c = A^T r$ , and applies a soft thresholding operation  $\delta_t(c)$ . The number of iterations is typically a small constant depending on the thresholding schedule, and the computational complexity is  $O(np)$ .

The iterative method described in Algorithm 2 computes the residual and correlation vectors as before, and applies a small constant number of conjugate gradient steps to approximately solve the system  $A_I^T A_I d_I(I) = c_I(I)$ , with  $I$  denoting the current active set. Each iteration involves application of  $A_I$  and  $A_I^T$  in the order of  $O(n^2)$ .

We compare this setting to the performance of one specific interior-point method, namely a primal-dual barrier method for convex optimization (PDCO) [19], for solving the linear program arising from  $(P_1)$ . PDCO solves a standard form linear program by appending a log-barrier term to the objective to replace the nonnegativity constraints and forming the KKT matrix to solve for the Newton update direction. The number of iterations is a small constant, and the overall complexity is  $O(np^2)$ . The key point in this case is that the space complexity which is linear in the input dimension is impractical for large-scale problems such as processing multidimensional signals.

We therefore consider the case where  $A$  is represented as a linear operator which is applied to a vector, for which we coin the term “fast operator”. The Fourier transform, Hadamard transform, and wavelet transform all belong to this class of operators. Such operators are important in large-scale applications, where storing the explicit matrix  $A$  is only suitable for relatively small-scale problems. As described, the main problem with representing the operator as an explicit matrix is that the space complexity is  $O(np)$ . In large-scale applications it is not feasible to store the matrix in memory. In such a case, when using a “fast operator”, the iterative algorithms presented and the specific interior method both have modest space requirements which make them suitable for large-scale applications. For example, the iterative method only requires storing the current estimate  $x$  and residual vector  $r$ . Similarly, an efficient implementation of PDCO stores in each stage the current primal and dual solution estimates, the next primal and dual steps, and residual. Such space requirements make these methods superior to LARS [14] or OMP [43, 44] in large-scale applications. Similarly, the algorithms in  $\ell_1$ -magic, a collection of methods for solving convex optimization programs, do not provide a method for solving the problem  $(D_\lambda)$  in which the matrix is accessed only through matrix-vector multiplications involving  $A$  and  $A^T$  or their fast operators. In this case, the specific interior-point method discussed above requires computing in each iteration the solution of a system which is prohibitively large, in the order of the data size. When  $A$  is given as a fast linear operator, such a system may be solved using a conjugate-gradient-type solver which is dom-

inated by two matrix-vector multiplications per iteration of PDCO.

Implementing the operation of the  $n \times p$  matrix  $A$  as a linear operator may take fewer than  $n \cdot p$  operation to apply, and is dependent on the specific operator involved in each application. For example, in the application described in the next section the operator involves random sampling, Fourier transform which is computed by applying the 1-dimensional transform on each dimension in turn in any order in  $O(p \log p)$  independent of dimension  $d$ , and wavelet transform, and therefore time complexity is equal to the time complexity of applying a given operator.

## 5. SIMULATIONS

In one dimension, a characteristic NMR signal is simulated by a sum of exponentially decaying sinusoids [30] with additive noise as a function of the number of peaks, amplitude, phase, decay time, and frequency:

$$\sum_{j=1}^L (A_j e^{i\varphi_j}) e^{-k\Delta t/\tau_j} e^{2\pi i k \Delta t w_j} + z. \quad (16)$$

Here  $L$  denotes the number of sinusoids, each corresponding to a single nuclear resonance. For each sinusoid  $j$ ,  $A_j$  denotes the amplitude,  $\varphi_j$  its phase in radians, and  $\tau_j$  the decay time.  $k$  denotes translation,  $\Delta t$  the sampling rate, and  $z \sim N(0, 1)$  the additive noise. This results in a spectrum which is the sum of Lorentzian peaks at the frequencies  $w_j$ . In our simulation, the amplitude, phase, and frequencies are random uniformly distributed, with constant decay time. A two-dimensional NMR signal is simulated by extending the characteristic one-dimensional equation such that there are separate decays, phases, and frequencies for each dimension  $d = 1, 2$ . The components in each dimension are of the same amplitude, and the projection onto each dimension is given by

$$\frac{e^{i\varphi_{d,j}} e^{-t_{d,j}}}{\tau_{d,j} e^{2\pi i t_{d,j} f_{d,j}}}. \quad (17)$$

Two-dimensional NMR was first proposed in 1971 by Jeener and became practical in experiments by Ernst and Freeman. In a normal pulsed NMR experiment, an excitation pulse is followed immediately by data acquisition or detection in which the FID is recorded and data stored. The basic principle of a 2D NMR experiment is that the first variable  $t_1$  is the period during which the system evolves, which can be of the order of milliseconds to seconds. This is followed by a constant mixing time which depends on the experiment, in which the spin states are allowed to mix, and finally by a detection phase at time  $t_2$  which is the second variable. Numerous experiments are performed, each with increasing values of  $t_1$ . When the interferograms (FIDs) from all these experiments undergo Fourier transform with respect to  $t_2$  we obtain multiple frequency domain spectra  $f_2$ . If decoupling is applied during the whole experiment, all of the  $f_2$  spectra are identical except for a decrease in intensity due to relaxation during time  $t_1$ . However, if the decoupler is on only during

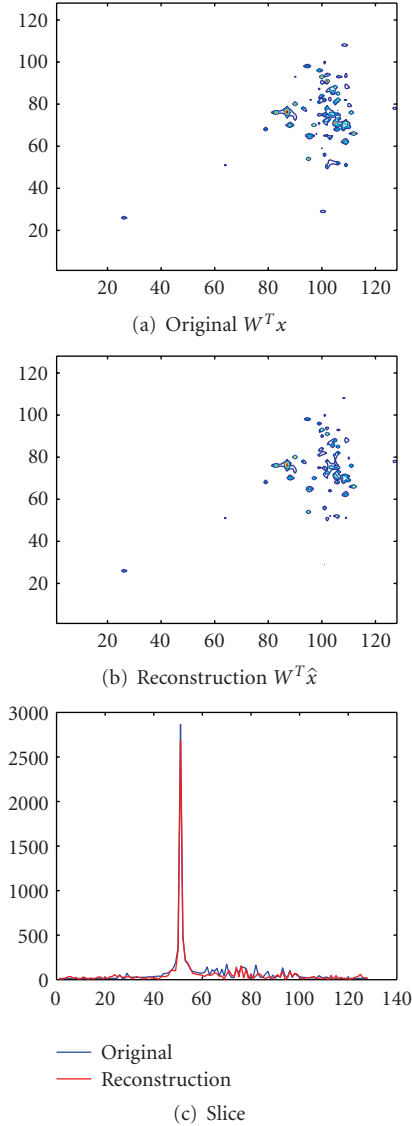


FIGURE 1: (a) Original spectrum. (b)  $\ell_1$  norm reconstruction. (c) Slice of original and reconstructed result.

data acquisition, the nuclei and protons are coupled during the evolution phase. In this work, we perform a realistic simulation of a broad range of 2D NMR spectra by using a data set containing 35 complex signals with varying parameters, including the rate of decay, noise, and peak crowdedness. The data set is used in an experiment management system for biomolecular NMR [45].

### 5.1. Random sampling

In the application of NMR spectroscopy, random sampling in the indirect dimension of the signal translates into reducing the experiment and actual measurement time. We reconstruct the spectra from the sampled signal by solving an optimization problem of the form  $(P_{1,\epsilon})$ . We set up a system in which the observation is the signal in scanline order, breaking up its real and imaginary parts. For example, Figure 1

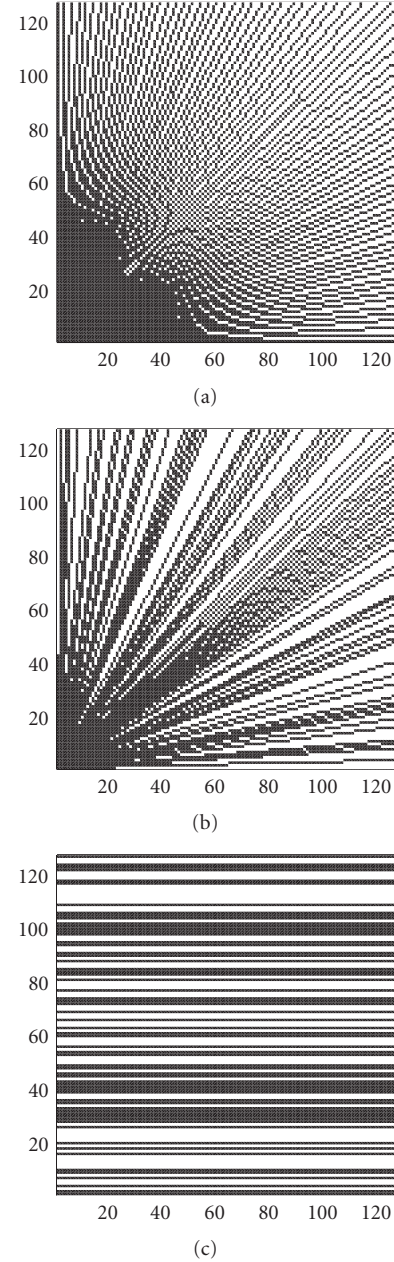


FIGURE 2: Different sampling patterns with a comparable number of samples: (a) equispaced radial lines. (b) Random radial lines. (c) Uniform random horizontal lines.

shows an original spectrum in Figure 1(a), the corresponding  $\ell_1$  norm reconstruction from random line samples in Figure 1(b), and a slice of the original and reconstructed result in Figure 1(c).

We experiment with several different sampling patterns as shown in Figure 2: Figure 2(a) equispaced radial lines, Figure 2(b) random radial lines, and Figure 2(c) uniform and nonuniform random horizontal lines; as well as randomly jittered radial lines and curves, to predict the number of measurements required to reconstruct with good

TABLE 1: Comparable undersampling factors for different sampling patterns.

Complex points	Random lines	Radial lines	Random radial lines
$128^2$	64	115	114.2
$256^2$	128	133	125.15
$512^2$	256	165	119.6
$1024^2$	512	187	129.35

TABLE 2: Dimensions, number of peaks, iterations, reconstruction accuracy, and running times for spectra.

Signal	Dimensions	Peaks	Time (s)
<i>Characteristic</i>	$128^2$	128	42.75
<i>Characteristic</i>	$256^2$	256	146.6
<i>Characteristic</i>	$512^2$	512	427.5
<i>Synthetic</i>	$128^2$	128	45.9
<i>Synthetic</i>	$128^2$	128	46

accuracy. We study the difference between sampling schemes and compare the results empirically.

The mutual coherence  $M(A)$  of a matrix  $A$  is the maximal off-diagonal entry of the Gram matrix  $A^T A$ . Given a sampling pattern as a binary matrix  $S$  and a basis matrix  $B$ , we form the corresponding Gram matrix  $G = A^T A$ , where  $A = SB$ . We compare the quantities  $M(S, B) = \max_{i \neq j} \|G_{i,j}\|/w^2$  averaged over 100 trials, where  $B$  is the Fourier basis,  $S$  the various sampling patterns, and  $w$  a normalizing scalar. We experiment with different sampling patterns illustrated in Figure 2, while keeping the same undersampling factor as shown in Table 1. As a numerical example, for a given synthetic  $128^2$  FID with a 2 : 1 sampling ratio, the mean reconstruction error for radial sampling is 0.28 and for random lines 0.67, whereas for a 4 : 1 sampling ratio the mean reconstruction error for radial sampling is 0.87 and for random lines 1.3. This demonstrates improved accuracy in radial sampling compared to random sampling. However, radial sampling commonly results in artifacts. We therefore introduce randomness into the sampling to diminish artifacts while maintaining accuracy.

## 5.2. Reconstruction time

The undersampling factor is important for estimating the reduction in actual NMR experiment time by simulation using characteristic and synthetic signals. There is a classical trade-off here, which is that reconstructing the signal requires solving a large optimization problem—our motivation for developing rapid iterative solvers. Therefore the computational efficiency of the nonlinear reconstruction becomes relevant. Reconstruction running times for undersampling by a factor of 2 are summarized in Table 2, on a standard PC running Matlab. The algorithm converges consistently within RMSE of 0.5 in under 50 iterations with a tolerance less than  $1e-5$  used in the stopping criterion. For a time comparison, solving an instance of the same optimization problem, ensur-

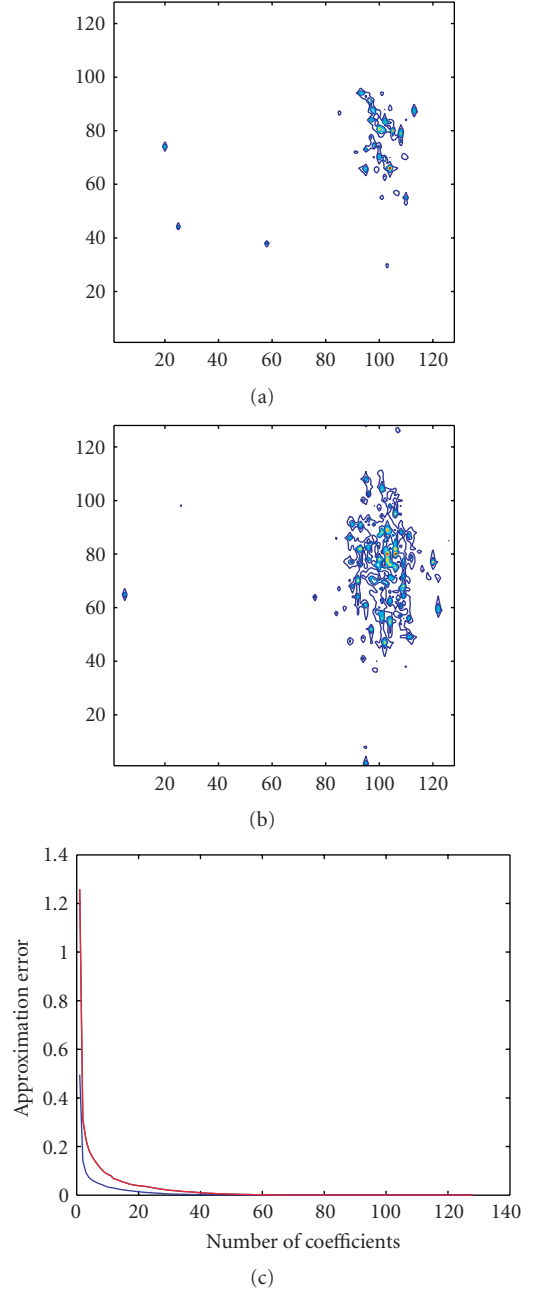


FIGURE 3: (a) Uncrowded and (b) crowded spectra; (c) rapid decay of approximation error for symmlet with 8 vanishing moments.

ing the same tolerance, for a  $256^2$  complex FID using PDCO takes more than three weeks of computation time.

## 5.3. Analysis of reconstruction results

We use the wavelet transform to represent NMR spectra. In order to measure the compressibility of the NMR spectra, we measure the decay of its wavelet coefficients. Figure 3 shows the rapid decay of mean approximation error for crowded (red) and uncrowded (blue) spectra using symmlet [46] with 8 vanishing moments. Figure 4 shows the decay of wavelet

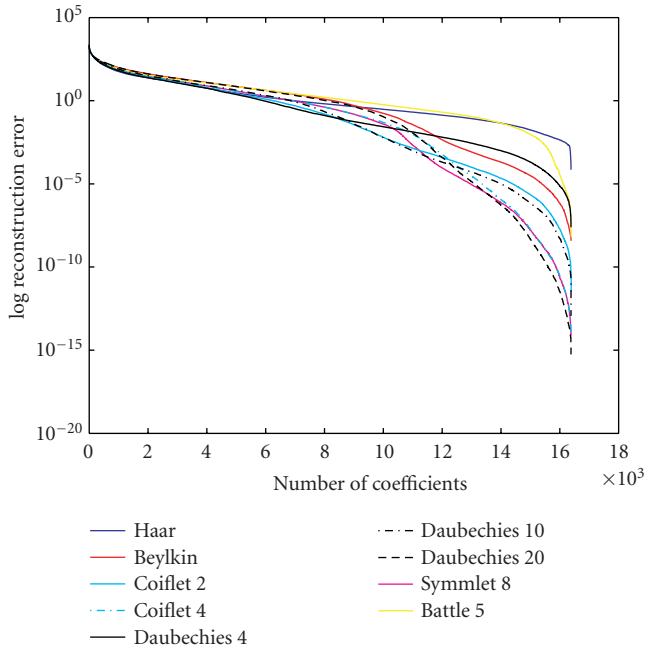


FIGURE 4: Decay of wavelet coefficients of real part on a semilog scale for various wavelets and vanishing moments.

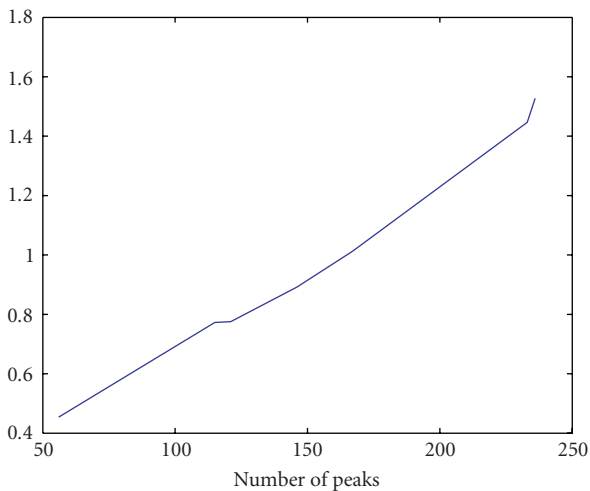


FIGURE 5: Mean reconstruction error as a function of the number of peaks for an increasing degree of peak crowdedness.

coefficients of the real part of spectra on a semilog scale for different types of wavelets with varying number of vanishing moments.

The peak crowdedness defines the number of peaks tightly clustered together, and defines the degree of overlap between peaks. We have experimented with reconstruction using a random sampling schedule in the indirect dimension for varying levels of peak crowdedness. Figure 5 shows the linear behavior of the mean reconstruction error as a function of the number of peaks for an increasing degree of peak crowdedness.

## 6. CONCLUSIONS

In conclusion, our methodology provides efficient iterative thresholding algorithms for finding sparse solutions to underdetermined systems of linear equations for large-scale practical applications. We demonstrate the applicability of our approach to NMR spectra by utilizing the sparsity of the wavelet transform and provide a valuable tool for speeding-up multidimensional NMR. We would like to compare the accuracy and sensitivity of our approach with maximum entropy reconstruction of high-dimensional spectra [2, 3]. Theoretically, it is well known that there exists an equivalence between maximum entropy and iterative soft thresholding up to a set of parameters. An important contribution of our work compared with maximum entropy spectra reconstruction is our use of the wavelet transform to obtain a sparse NMR representation.

## ACKNOWLEDGMENTS

The author thanks David Donoho, Yaakov Tsaig, and Jonathan Kaplan for useful discussions. This work was supported in part by grants from NIH Project R01GM072000-03 and NMRA: New Mathematical Methodology for NMR Spectroscopy.

## REFERENCES

- [1] S. Kim and T. Szyperski, "GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information," *Journal of the American Chemical Society*, vol. 125, no. 5, pp. 1385–1393, 2003.
- [2] M. Mobli, A. S. Stern, and J. C. Hoch, "Spectral reconstruction methods in fast NMR: reduced dimensionality, random sampling and maximum entropy," *Journal of Magnetic Resonance*, vol. 182, no. 1, pp. 96–105, 2006.
- [3] D. Rovnyak, D. P. Frueh, M. Sastry, et al., "Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction," *Journal of Magnetic Resonance*, vol. 170, no. 1, pp. 15–21, 2004.
- [4] P. Schmieder, A. S. Stern, G. Wagner, and J. C. Hoch, "Application of nonlinear sampling schemes to COSY-type spectra," *Journal of Biomolecular NMR*, vol. 3, no. 5, pp. 569–576, 1993.
- [5] P. Schmieder, A. S. Stern, G. Wagner, and J. C. Hoch, "Improved resolution in triple-resonance spectra by nonlinear sampling in the constant-time domain," *Journal of Biomolecular NMR*, vol. 4, no. 4, pp. 483–490, 1994.
- [6] I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-based image completion," *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, pp. 303–312, 2003.
- [7] I. Drori and D. Lischinski, "Fast multiresolution image operations in the wavelet domain," *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, no. 3, pp. 395–411, 2003.
- [8] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, "Near-optimal sparse Fourier representations via sampling," in *Proceedings on 34th Annual ACM Symposium on Theory of Computing*, pp. 152–161, Montreal, Quebec, Canada, May 2002.
- [9] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

- [10] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [11] E. Berlekamp, R. McEliece, and H. van Tilborg, "On the inherent intractability of certain coding problems," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 384–386, 1978.
- [12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [13] D. L. Donoho, "For most large underdetermined systems of equations, the minimal  $\ell_1$ -norm near-solution approximates the sparsest near-solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 7, pp. 907–934, 2006.
- [14] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.
- [16] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [17] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
- [18] D. L. Donoho and Y. Tsaig, "Fast solution of  $\ell_1$ -norm minimization problems when the solution may be sparse," Tech. Rep. 2006-18, Department of Statistics, Stanford University, Stanford, Calif, USA, 2006.
- [19] M. A. Saunders and B. Kim, "PDCO: primal-dual interior method for convex objectives," <http://www.stanford.edu/group/SOL/software/pdco.html>.
- [20] Y. Ye, *Interior Point Algorithms: Theory and Analysis*, John Wiley & Sons, New York, NY, USA, 1997.
- [21] M. Lustig, D. L. Donoho, and J. M. Pauly, "Rapid MR imaging with compressed sensing and randomly under-sampled 3DFT trajectories," in *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM '06)*, Seattle, Wash, USA, May 2006.
- [22] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "Rapid MR angiography with randomly under-sampled 3DFT trajectories and non-linear reconstruction," in *Proceedings of the 9th Annual Scientific Sessions of the Society for Cardiovascular Magnetic Resonance (SCMR '06)*, Miami, Fla, USA, January 2006.
- [23] M. Lustig, J. M. Santos, J.-H. Lee, D. L. Donoho, and J. M. Pauly, "Compressed sensing for rapid MR imaging," in *Proceedings of the Signal Processing with Adaptive Sparse Structured Representations (SPARS '05)*, Rennes, France, November 2005.
- [24] D. L. Donoho and B. F. Logan, "Signal recovery and the large sieve," *SIAM Journal on Applied Mathematics*, vol. 52, no. 2, pp. 577–591, 1992.
- [25] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [26] M. Rudelson and R. Vershynin, "Geometric approach to error-correcting codes and reconstruction of signals," *International Mathematics Research Notices*, vol. 2005, pp. 4019–4041, 2005.
- [27] I. Drori, V. C. Stodden, and E. H. Hurowitz, "Fast  $\ell_1$  minimization for genome-wide analysis of mRNA lengths," in *Proceedings of the IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS '06)*, pp. 19–20, College Station, Tex, USA, May 2006.
- [28] R. Freeman and E. Kupče, "New methods for fast multidimensional NMR," *Journal of Biomolecular NMR*, vol. 27, no. 2, pp. 101–114, 2003.
- [29] D. Malmodin and M. Billeter, "High-throughput analysis of protein NMR spectra," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 46, no. 2-3, pp. 109–129, 2005.
- [30] J. C. Hoch and A. S. Stern, *NMR Data Processing*, John Wiley & Sons, New York, NY, USA, 1996.
- [31] S. Hiller, F. Fiorito, K. Wüthrich, and G. Wider, "Automated projection spectroscopy (APSY)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 31, pp. 10876–10881, 2005.
- [32] E. Kupče and R. Freeman, "Fast reconstruction of four-dimensional NMR spectra from plane projections," *Journal of Biomolecular NMR*, vol. 28, no. 4, pp. 391–395, 2004.
- [33] V. Y. Orekhov, I. Ibraghimov, and M. Billeter, "Optimizing resolution in multidimensional NMR by three-way decomposition," *Journal of Biomolecular NMR*, vol. 27, no. 2, pp. 165–173, 2003.
- [34] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Tech. Rep. 2006-02, Department of Statistics, Stanford University, Stanford, Calif, USA, 2006.
- [35] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 361–379, 2000.
- [36] S. Sardy, A. G. Bruce, and P. Tseng, "Robust wavelet denoising," *IEEE Transactions on Signal Processing*, vol. 49, no. 6, pp. 1146–1152, 2001.
- [37] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [38] M. Elad, "Why simple shrinkage is still relevant for redundant representations?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5559–5569, 2006.
- [39] M. Elad, B. Matalon, and M. Zibulevsky, "Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization," *Journal on Applied and Computational Harmonic Analysis*, 2006.
- [40] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.
- [41] M. A. T. Figueiredo and R. D. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '05)*, vol. 2, pp. 782–785, Genova, Italy, September 2005.
- [42] C. C. Paige and M. A. Saunders, "LSQR: an algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software (TOMS)*, vol. 8, no. 1, pp. 43–71, 1982.
- [43] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Asilomar Conference on Signals, Systems & Computers*, vol. 1, pp. 40–44, Pacific Grove, Calif, USA, November 1993.
- [44] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

- [45] P. T. Lee, J. Li, M. R. Chapman, et al., “SESAME: an experiment management system for biomolecular NMR,” <http://www.sesame.wisc.edu>.
- [46] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, Pa, USA, 1994.

---

**Iddo Drori** received the B.S. degree in computer science and mathematics in the Amirim program for excellence from the Hebrew University of Jerusalem in 1998 and received the M.S. degree in computer science with honors from the Hebrew University in 2000. He received his Ph.D., in computer science from Tel-Aviv University in 2004 and is the receipt of the Maus prize in CS. From 2004 till 2006, he was a post-doc at the Department of Statistics at Stanford University with David Donoho, focusing on sparse solution of underdetermined linear systems, multiscale representations of manifold valued data, NMR spectroscopy, and computational biology. Dr. Drori's research interests include applied statistic, scientific computing, computer graphics and vision, sound, image and video processing, and applied computational geometry. Most recently he has been working on performance capture, articulation, and geometry processing. For more information see <http://www.cs.tau.ac.il/~idrori>.

