Image Reconstruction from Multiple Sparse Representations

Robert Crandall
Advisor: Professor Ali Bilgin

University of Arizona
Program in Applied Mathematics
617 N. Santa Rita, Tucson, AZ 85719

Abstract

In this paper we introduce the emerging topic of compressed sensing. We will discuss the basics of compressed sensing and introduce several algorithms for solving practical problems. We will discuss the idea of combining multiple sparse representations of a signal to form a better reconstruction. Finally, we propose a new method for the reconstruction of compressible signals that combines ideas from two recently proposed methods, and show that this new method gives promising results.

1 Introduction to Compressed Sensing

Digital signal acquisition is the process of measuring some signal such as an image, a sound, or a time history of temperatures and storing it digitally in a computer or other device. Since computers have limited memory and can store only discrete values, we are limited to taking a finite set of measurements. Even if this number of measurements is large, we can never exactly store an arbitrary signal in a computer, since signals can be infinite dimensional. Still, digital signals are clearly capable of approximating real world data well; the prevalence of digital music and digital images is testament to this. The question is, then, just how many measurements must we take to get an accurate description of a signal?

The classical answer to this question is given by the Nyquist-Shannon sampling theorem. This celebrated result says that, if we know a priori that a signal of interest is bandlimited (meaning its Fourier transform has compact support), then we can reconstruct that signal exactly from discrete measurements taken at a sampling rate equal to at least twice the bandwidth of the signal.

Theorem 1. *Nyquist-Shannon Sampling Theorem*

Let $F(k)$ be the Fourier transform of a function $f(t)$. Suppose the support of F is contained in $(-2\pi W, 2\pi W)$ for some W. Then

$$f(t) = \sum_{m=-\infty}^{\infty} f\left(\frac{m}{2W}\right) \operatorname{sinc}(2Wt - m),$$

where $\operatorname{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$.

In other words, we can reconstruct a signal with bandwidth $< W$ exactly from samples taken at a rate $f_s = 2W$.

The so-called Nyquist sampling rate of $\frac{1}{2W}$ is important in nearly all current digital acquisition technologies; in many applications, signals are passed through a bandpass filter before being digitized to explicitly enforce the Nyquist-Shannon criterion. The important point is that we must, somehow or another, make an assumption that the signals of interest have limited information content that we can store at least approximately as a finite set of numbers in a computer. In the Nyquist-Shannon theory, this a priori assumption is that signals of interest have limited frequency content.

In compressed sensing, the assumption we make is that our signals are *sparse* or *compressible* (we will discuss the definitions of these terms in the next section). Such signals, like the band-limited signals of the Nyquist-Shannon framework, have limited information content that can be stored in a finite number of coefficients. Furthermore, as its name suggests, compressed sensing provides a way to acquire signals directly in a compressed form; this compressed signal can then be reconstructed after acquisition as a postprocessing step.

1.1 Reconstruction from Linear Measurements

In compressed sensing we assume that the measurement process can be modeled by a linear transformation $\Phi\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$. That is, given a signal $x \in \mathbb{R}^n$, we form a measurement $y = \Phi\, x$. This measurement y is the only information about the real world signal x that we can access directly, and our goal is to reconstruct x as closely as possible. Depending on the nature of the measurement matrix Φ and the signals of interest, we may be able to reconstruct the signal exactly, or we may have to settle for an approximation.

1.1.1 Complete Measurements

If $\Phi\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with $m \geqslant n$ and Φ is full rank, then we have a fully determined (or overdetermined) system $y = \Phi\, x$. In this case we can solve uniquely for $x = \Phi^{-1} y$ given any measurement y. This situation is of little interest mathematically; obviously if we take at least as many measurements as the size of the signal, then we can learn all there is to know about that signal. The interesting fact is that certain types of signals can be recovered exactly even from incomplete measurements, when $m < n$. In this case we are able to acquire a signal directly in a compressed form (requiring only m coeffiecients to store) and later reconstruct the full signal (containing the full n coefficients) in postprocessing.

1.1.2 Incomplete Measurements

If $m < n$, then the linear system $y = \Phi\, x$ is underdetermined. Since range(Φ) has smaller dimension then the domain of signals, there is no hope of being able to reconstruct every signal exactly. Since null(Φ) is necessarily nontrivial in this case, $\Phi(x + n) = \Phi(x)$ for any $x \in \mathbb{R}^n$ and any $n \in \text{null}(\Phi)$. This means that we can't determine x uniquely from the measurement y, but we can determine the affine space $x + \text{null}(\Phi)$ that maps to y; this is the set of solutions to the reconstruction problem. The goal of signal reconstruction is then to seek the "best" solution in this affine space, which in the compressed sensing paradigm usually means the sparsest solution. To illustrate this, suppose we have a 2×3 measurement matrix

$$\Phi = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix},$$

and we take a measurement $\Phi\, x = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$. The nullspace of Φ is the line parametrized by $\begin{pmatrix} t \\ -t \\ t \end{pmatrix}$, and the affine space of solutions is the line parametrized by $\begin{pmatrix} 1+t \\ 1-t \\ 1+t \end{pmatrix}$. The least squares solution is $\begin{pmatrix} 2/3 \\ 4/3 \\ 2/3 \end{pmatrix}$, and the sparsest solution is $\begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$.

1.1.3 Reconstruction by l^2 Minimization

Applying the pseudoinverse $\Phi^+ = \Phi^*(\Phi\, \Phi^*)^{-1}$ to y in an underdetermined system gives the *least squares solution*, which is the unique solution to the reconstruction problem with smallest l^2 norm. In many cases, when we expect the signal x to have small Euclidean norm, this can give a satisfactory answer. However, in compressed sensing the signals of interest are typically not well approximated by the least squares solution. Shown below is a least-squares reconstruction of an MRI image, which is measured using an undersampled Fourier matrix. The reconstruction contains excessive artifacts and blurring; we will see that we can find a much better reconstruction using compressed sensing methodologies.

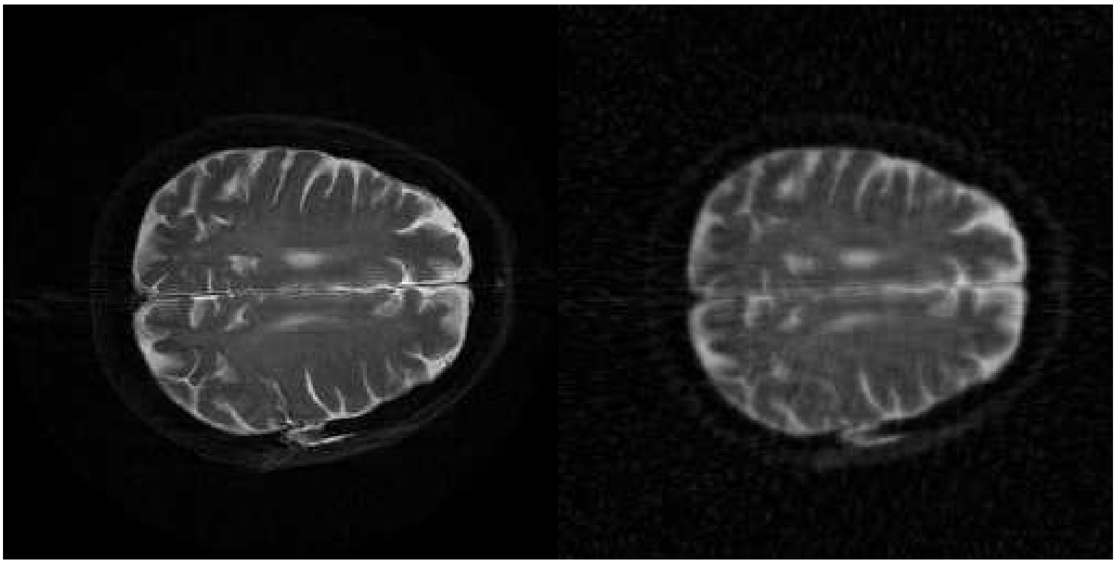

Figure 1. (left) An MRI scan of a human brain, and (right) the least-squares reconstruction from incomplete Fourier space measurements (64 radial lines)

1.2 Reconstruction of Sparse Signals by l^0 Minimization

In order to improve on the least squares reconstruction, we must have some a priori knowledge about the signal of interest. In compressed sensing, this a priori information takes the form of *sparsity* or *compressibility*. This means that the signals we are interested in are restricted to be in a low dimensional subspace of $\mathbb{R}^n$, or at least that most of their energy is contained in such a subspace.

Definition 2. *Sparsity and Sparsifying Transforms*
We say a signal $x \in \mathbb{R}^n$ is s-sparse *if at most s of its coefficients are nonzero. We say that x is s-sparse with respect to a transformation Ψ if Ψx is s-sparse. In this case Ψ is called a sparsifying transform.*

A *compressible* signal is one that is approximately sparse, in that it can be well approximated in the l^2 sense by a strictly sparse signal. For example, many images are compressible with respect to the discrete cosine transform; this is the principle behind the JPEG image compression standard. If we have an image x, we can apply a sparsifying DCT to obtain $y = \Psi x$. y can then be thresholded to obtain a strictly sparse signal; the image has then been compressed, since the sparse signal y takes less coefficients to store in memory than the original x. We can restore the image by applying the inverse Ψ^{-1} to the compressed signal y; the quality of the reconstruction will depend on how compressible the original signal was (i.e. how much of its energy was contained in the sparse representation).

1.2.1 Exact Reconstruction by l^0 Minimization

In compressed sensing we assume that all signals of interest are sparse or compressible with respect to some sparsity basis Ψ. This additional piece of information gives us, at least in principle, a way to determine the best reconstruction in the affine space $x + \text{null}(\Phi)$; we simply seek the sparsest solution we can find in this space. This amounts to solving the minimization problem

$$\text{argmin}_x \, \|\Psi x\|_0, \text{subject to}\, \Phi\, x = y.$$

The operator $\|\cdot\|_0$ counts the number of nonzero entries in a vector, i.e. $\|x\|_0 = \#\{n\, s.t. x_n \neq 0\}$.

The main advantage of compressed sensing is that we can often resolve a signal exactly from a number of measurements that is proportional to the sparsity of the signal, not the full signal length n. Consider first the case where x is sparse with respect to the standard basis, so $\|x\|_0$

itself is small. Clearly, not just any measurement basis will allow us to reconstruct x from incomplete measurements; for example, if Φ is a 4×10 measurement matrix whose rows are chosen from the 10×10 identity, then there is no hope of being able to reconstruct even a general 1-sparse signal from the four measurements Φx. However, if we take measurements using, for example, an undersampled Fourier matrix, then we can reconstruct sparse signals from less than the full set of measurements.

Proposition 3. *A Condition for Uniqueness of the Minimum-Sparsity Solution*

Suppose $\Phi: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear operator, and that any $2s$ columns of Φ are linearly independent. Then any s-sparse vector x can be reconstructed uniquely from the measurement $y = \Phi x$.

Proof. *Suppose x and z are two s-sparse vectors in $\mathbb{R}^n$ and $\Phi x = \Phi z$. Then $\Phi x - \Phi z = \Phi(x - z) = 0$. The vector $x - z$ is clearly $2s$-sparse, and since any $2s$ columns of Φ are linearly independent by assumption, this implies that $x - z = 0$.* □

Thus, we can reconstruct any s-sparse signal from its measurements Φx as long as any $2s$ columns of Φ are linearly independent. Suppose for example that Φ is a 4×10 matrix whose rows are chosen from a 10×10 Fourier matrix. Then any 4 columns of Φ are linearly independent; it follows that any 2-sparse vector x can be reconstructed from the measurements Φx. This means that we can acquire 2-sparse signals in $\mathbb{R}^{10}$ using only 4 Fourier space measurements.

Often, the signal of interest is sparse not in the standard basis, but rather under some sparsifying transform Ψ; i.e. $\alpha = \Psi x$ is s-sparse for some small s. In this case we must consider not only the properties of the measurement basis Φ, but also the sparsifying basis Ψ. By [prop number], if any $2s$ columns of $\Phi\Psi^{-1}$ are linearly independent, then we can reconstruct the s-sparse representation α uniquely from the measurement $y = \Phi x = \Phi\Psi^{-1}\alpha$. This property of $\Phi\Psi^{-1}$ is equivalent to the requirement that the measurement and sparsifying transforms be "incoherent," meaning essentially that no two columns of Φ and Ψ can be highly correlated (we will discuss incoherence further below). This is why we cannot reconstruct signals that are sparse in the standard basis by taking incomplete measurements in the standard basis; in this case the sparsifying transform Ψ is just the identity matrix, which is maximally coherent with itself and thus with the measurement matrix Φ. Thus, the full set of standard-basis measurements must be taken in order to reconstruct even signals with minimum sparsity. However, the Fourier basis is maximally incoherent with the standard basis, so we can reconstruct signals that are sparse in the standard basis from relatively few Fourier measurements. Similarly, we can reconstruct signals that are sparse in the frequency domain (i.e. with bandlimited Fourier transform) from relatively few time domain measurements; this is analogous to the Nyquist-Shannon theorem. This property of incoherence between Φ and Ψ is the key piece that allows compressed sensing to work.

We now have what we need to define a simple compressed sensing framework for strictly sparse signals. Given a sparsifying transform Ψ and a measurement basis Φ satisfying the conditions of Proposition 3, our system takes measurements $y = \Phi x$. If we can solve the minimization problem

$$\arg\min_x \|\Psi x\|_0, \text{subject to}\, \Phi x = y,$$

then we can reconstruct every s-sparse signal exactly in this way. Unfortunately, this l^0 minimization is not trivial; it is a non-convex optimization, and we usual cannot solve it in practice. As such, we are forced to use other methods to seek sparse solutions.

1.3 Reconstruction by l^1 Minimization

We have seen that l^2 minimization is easy to solve by standard least-squares methods, but gives poor results for compressed sensing problems; on the other hand, l^0 minimization gives ideal results for sparse signals, but is intractable in practice. It turns out that by taking something between these two extremes, namely l^1 minimization, it is possible to obtain useful results for real problems. In fact, in [3] and [4] it was shown that l^1 minimization works much better than one might expect; it can obtain exact solutions in many cases, an near-optimal solutions in many others. These papers began a flood of research into compressed sensing (hundreds of papers have been published on the topic since 2006), since they proved that it is indeed feasible to create practical devices that employ compressed sensing to decrease the number of necessary measurements.

1.3.1 The Role of Incoherence

We will now formalize the concept of incoherence that was mentioned above. Suppose we have measurement and sparsifying transforms Φ and Ψ (for simplicity, we will assume here that Φ and Ψ are both orthonormal bases for $\mathbb{R}^n$, although compressed sensing works for more general transforms). Then the coherence $\mu(\Phi, \Psi)$ between these pairs is defined as

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{1 \leqslant i, j \leqslant n} |\langle \phi_i, \psi_j \rangle|,$$

where ϕ_i and ψ_j are the ith and jth rows of Φ and Ψ respectively. Since Φ and Ψ are orthonormal, $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$; if $\mu = 1$ the pair is maximally incoherent, and if $\mu = \sqrt{n}$ the pair is maximally coherent. In compressed sensing applications we want our measurement and sparsity bases to be as incoherent as possible. Finding a suitable sparsifying transform for signals of interest and designing or selecting an incoherent measurement basis are paramount considerations in compressed sensing system design.

For example, the n-dimensional Fourier basis and the n-dimensional standard basis are maximally incoherent. An $n \times n$ Fourier matrix has entries

$$f_{kl} = \frac{1}{\sqrt{n}} e^{-2\pi i k l / n},$$

so $|\langle f_i, e_j \rangle| = \frac{1}{\sqrt{n}}$ for each i, j and the coherence is 1. This means that the Fourier and identity matrices are optimal pairs for use in compressed sensing; if our signals are sparse in the standard basis, we should use Fourier measurements, and if our signals are sparse in the frequency domain we should use the identity matrix to take measurements.

As another example, in magnetic resonance imaging, signals are acquired in the Fourier domain. The acquired images are typically sparse in various wavelet domains. The Fourier and wavelet bases are not maximally incoherent, but they are sufficiently incoherent to give good results and allow us to reconstruct MR images from significantly undersampled measurements. We will see examples of this later on.

1.3.2 Performance Guarantees for l^1 Minimization

The l^1 minimization problem for compressed sensing is stated formally as follows: given measurement and sparsity bases Φ and Ψ, a true signal x and a measurement $y = \Phi x$, we solve the minimization problem

$$\operatorname{argmin}_x \|\Psi x\|_1, \text{subject to } \Phi x = y.$$

This is a convex optimization problem, and it can be solved by standard linear programming methods.

We now reproduce a performance result presented in [1] for general measurement and sparsity bases with no noise is added to the measured signal. For any subset $\Omega \subset \{1, 2, ..., n\}$ and any orthonormal measurement matrix Φ, we can define an undersampled measurement matrix Φ_Ω defined by taking only the rows of Φ indexed by Ω. This is an $m \times n$ matrix where $m = \#\Omega < n$. The result states that for sparse enough signals and incoherent dictionaries, we can reconstruct signals exactly with high probability from a small number of uniform random measurements.

Theorem 4. *Performance Guarantees for Reconstruction by l^1 Minimization*
Suppose the number of measurements m satisfies

$$m \geqslant C\,(\mu(\Phi, \Psi))^2 \cdot S \cdot \log\left(\frac{n}{\delta}\right)$$

where C and δ are constants and S is the sparsity of the true signal. Suppose $\Omega \subset \{1, ..., n\}$ is a set of m measurements selected uniformly at random. Then the signal x can be reconstructed exactly from the measurements $\Phi_\Omega x$ by solving the l^1 minimization problem, with probability greater than $1-\delta$.

Thus, for highly incoherent bases we can expect to get exact recovery from $O(S \log n)$ randomly selected measurements; this can be significantly less than the full set of n as long as S/n is small. A general rule of thumb which is useful for many problems is that we can expect to need approximately $3s$ to $5s$ to reconstruct an s-sparse signal.

2 Other Practical Solution Methods

2.1 Iterative Hard Thresholding

In [5] an algorithm called iterative hard thresholding (IHTs) is shown to work well for the compressed sensing problem when the ratio of number of measurements to signal sparsity is large. In this algorithm, we select a desired sparsity s for our computed solution. We begin with $\alpha_0 = 0$, and update our solution using the iteration

$$\alpha_{n+1} = H_s(\alpha_n + \Psi\Phi^*(y - \Phi\Psi^*\alpha_n)).$$

Here H_s is the nonlinear hard thresholding operator that sets all but the s largest coefficients of its argument to zero, so the sparsity of the computed representation α is strictly enforced.

This algorithm boasts several useful properties. It is much faster than the OMP method, with the most computationally expensive step being the application of operator $\Psi\Phi^*$ and its transpose at each iteration. It gives near optimal performance guarantees, provided the sensing and sparsity bases are sufficiently incoherent. It converges quickly, with the number of iterations depending only on the signal to noise ratio of the signal. We will examine the performance of this algorithm further in a later section, when we compare it with a randomized version.

2.2 Orthogonal Matching Pursuit

Orthogonal Matching Pursuit, or OMP. OMP is a greedy algorithm that chooses, one basis vector at a time, a sparse representation α such that the error $\|y - \Phi\Psi^{-1}\alpha\|_2$ is small. Here y is the measurement of the original signal, possibly with noise: $y = \Phi\, x + v$ where v is some noise vector.

Define the "dictionary" for the problem to be the combined operator $D = \Phi\Psi^{-1}$. Let S be the support of our solution at any given step; e.g. if $S = \{1, 3\}$ then our solution is a linear combination of the colums d_1, d_3. We start with an initial, empty support $S = \{\}$, and define an initial residual $r = y - D\,\alpha = y$. At each step, we select a vector from the columnns of D to add to the support of α, by determining which column of D is most correlated with the current residual r. This will allow us to subtract as much energy as possible off of the residual at each step. Thus, we form the correlation $D^* r$, select the index j with maximal correlation, and add this index to our support S. We then find the least-squares solution to the problem $D_S \alpha = y$, where D_S is the matrix formed by selecting columns S from the original D. The least squares solution is found by applying the pseudoinverse $D_S^+ = (D_S^* D_S)^{-1} D_S^*$ to y. We repeat this iteration until the residual $r = y - D\,\alpha$ falls below some predetermined threshold, which is typically determined by our assumptions about measurement noise.

3 Fusing Multiple Representations

The two solution methods mentioned in the previous section are deterministic. This means that if our solution fails to find the optimal sparsest solution, we are stuck with what we get. Recent results have shown that better sparse representations can be obtained by somehow fusing a multitude of sparse representations together. In this way we are able to take advantage of the slightly different picture of the original signal that each individual representation represents, and obtain a solution with reduced noise and artifacts. One way to obtain multiple such representations is add some randomness to a basis pursuit algorithm, selecting not the optimal atom to add at each step but rather updating the solution support based on some probability distribution. We introduce here the RandOMP algorithm from [2]. This is a randomized version of the OMP method above, which can be used to generate multiple representations that can be fused together to give a better reconstruction than the standard OMP solution. We then apply a similar technique with Iterative Hard Thresholding to obtain a new algorithm that gives promising results.

3.1 RandOMP

The Orthogonal Matching Pursuit algorithm introduced above is deterministic; given the same input, the algorithm will always update the solution support in the same way at each iteration. This algorithm updates the support at each step in the best way possible given the greedy constraint that we can only add one coeffecient at a time; however, it often fails to find the global sparsest solution to the problem. In [2] it was shown that several sub-optimal representations, each less sparse than the OMP solution, can be averaged to form a better representation.

Elad and Yavneh [2] applied their RandOMP algorithm to the problem of denoising signals with sparse representations. The idea behind this algorithm is that multiple different sparse representations of the same signal each capture a good portion of the true signal, and each representation carries with it a unique representation of undesirable signal noise. By averaging these multiple solutions, the noise can be attenuated without losing the true signal.

RandOMP is identical to OMP, except that the support update step is randomized; instead of always selecting the column of the dictionary that is most correlated with the residual r, we select a column at random. In [2] the authors assume that the noise added to the signal is independent, identically distributed Gaussian noise drawn from $N(0,\,\sigma^2)$, while the nonzero coefficients of the signal's true sparse representation are distributed according to $N(0,\,\sigma_x^2)$. Given these assumptions, they derive a probability distribution used to update the solution support at each step such that averaging multiple representations gives optimal denoising performance. At each step, a new support index j is selected with probability proportional to

$$\exp\left(\frac{\sigma_x^2}{2\,\sigma^2(\sigma_x^2+\sigma^2)}\,\frac{|d_j^* r|^2}{\|d_j\|_2^2}\right).$$

Recall that d_j is the jth column of the dictionary, and r is the current residual at a given step; thus, the probability of selecting a given atom goes up exponentially with the square of the correlation of that atom with the residual.

To show that this RandOMP algorithm indeed works, we repeat some of the numerical experiments performed in [2]. We start by taking a random dictionary $D \in \mathbb{R}^{100\times 200}$, with entries chosen from $N(0,1)$, and l^2-normalize its columns. We create a random sparse representation α_0 with 10 nonzero entries also drawn from $N(0,\,1)$. We set $x = D\,\alpha$, and $y = x + v$ where v is a noise vector of 100 elements drawn from $N(0,1)$ as well (note that we are using a very low signal to noise ratio). We run RandOMP 1000 times with a threshold of $10 = \sigma\sqrt{n}$, and regular OMP once for comparison. For evalutation purposes, for a given representation we define the noise attenuation factor (we will call this the "denoising factor") by

$$\frac{\|D\,\alpha - x\|^2}{\|y - x\|^2}$$

where α is the sparse representation computed by the algorithm.

In our trial, the regular OMP method found a 5-sparse solution. This representation achieved a denoising factor of 0.4151, so over half the noise was attenuated. The OMP solution was the sparsest, with the RandOMP representations ranging from 6- to 27-sparse.

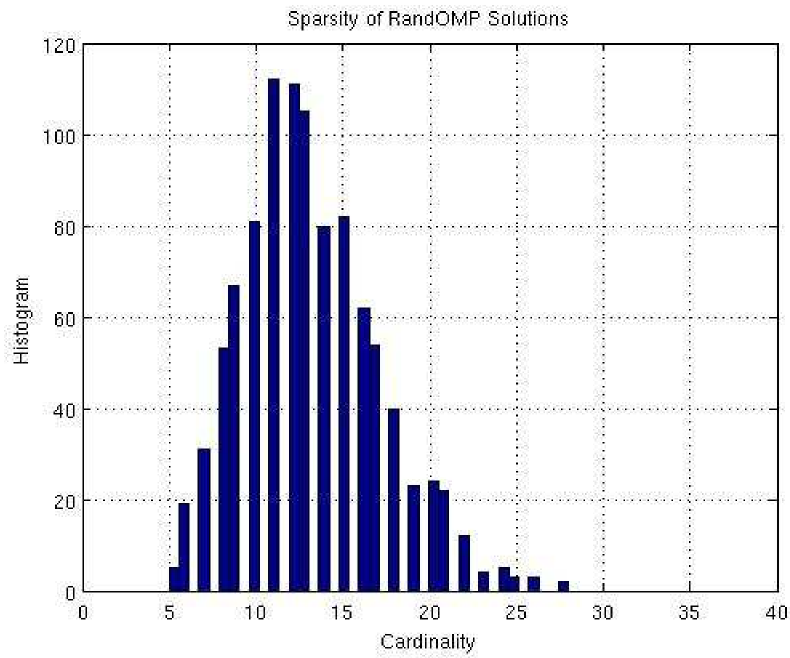

Figure 2. Histogram of cardinalities of RandOMP solutions

In terms of the denoising factor, just over half of the RandOMP solutions were better than the OMP solution, while the rest were worse.

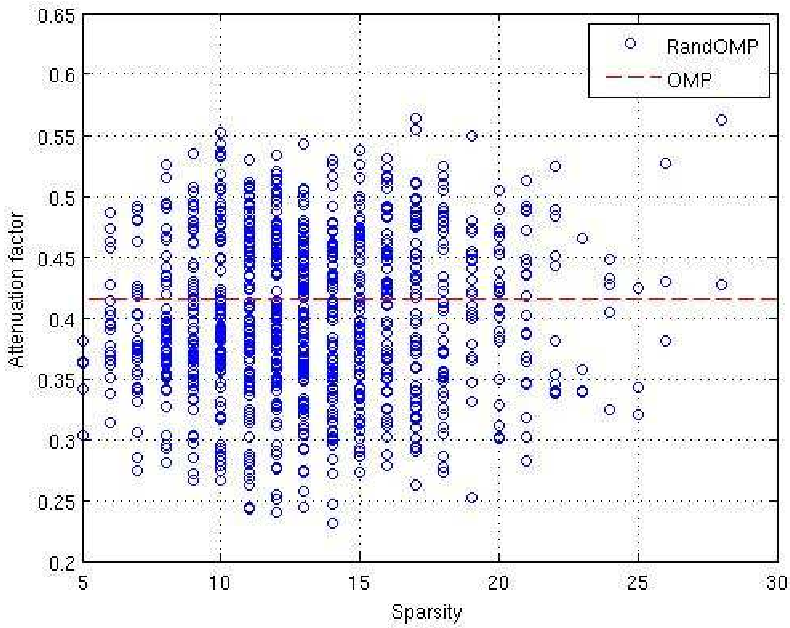

Figure 3. (blue dots) Denoising factor of RandOMP solutions vs. representation sparsity, (red line) Denoising factor of OMP solution

Indeed, as we hoped, averaging these 1000 randomized representations gives a solution that is superior to the OMP solution and each of the individual RandOMP solutions; the averaged representation achieved a denoising factor of 0.2204.

3.2 Randomizing IHTs

By analogy with the RandOMP method, we propose here a new method for compressed sensing reconstruction based on a randomized version of Iterative Hard Thresholding. The randomization is obtained by modifying the hard thresholding operator H_s. We replace this deterministic function with a randomized version which, instead of selecting the largest s coefficients of its input, selects s coefficients at random with probability proportional to

$$\exp(c \cdot |\alpha_i|),$$

where c is a constant and α_i is the ith coefficient in the current representation. This allows us to find, as in the RandOMP case, a set of unique representations that we can hopefully combine in some way to find a better solution than the solution found by deterministic IHTs. Preliminary results have shown that, indeed, we can see performance gains similar to those given by RandOMP by averaging multiple random solutions. The primary advantage of this method over OMP is that it is significantly faster, and thus is more suitable for use on higher dimensional problems such as image reconstruction.

3.2.1 Noisy Signal Reconstruction

To demonstrate the performance of the IHTs algorithm and its randomized counterpart, we generate a random dictionary $D \in \mathbb{R}^{128\times 256}$ by drawing its entries uniformly from the unit sphere. This dictionary should be thought of as the combined measurement/sparsity operator $\Phi\Psi^{-1}$, since we will try to reconstruct a sparse signal α from its measurement $y = D\,\alpha$. Since D is 128×256, this corresponds to 50% undersampling. Next, we generate a random 8-sparse representation $\alpha \in \mathbb{R}^{256}$ by selecting a random sparsity support and choosing the nonzero entries from $N(0, 1)$. We then create a noise vector $v \in \mathbb{R}^{128}$ with entries chosen from $N(0, 0.3/\sqrt{128})$, which corresponds to a signal to noise ratio $\frac{\|D\,\alpha\|_2}{\|v\|_2} \approx 10$, and form the noisy measurement

$$y = D\,\alpha + v.$$

We then compute a reconstruction $\hat{\alpha}$ from the noisy measurement using the IHTs algorithm.

In our trial the deterministic IHTs method converged in about 6 iterations to a solution with l^2 error ~ 0.08 (this is the error between the true, noise-free solution $D\,\alpha$ and the computed solution $D\,\hat{\alpha}$).

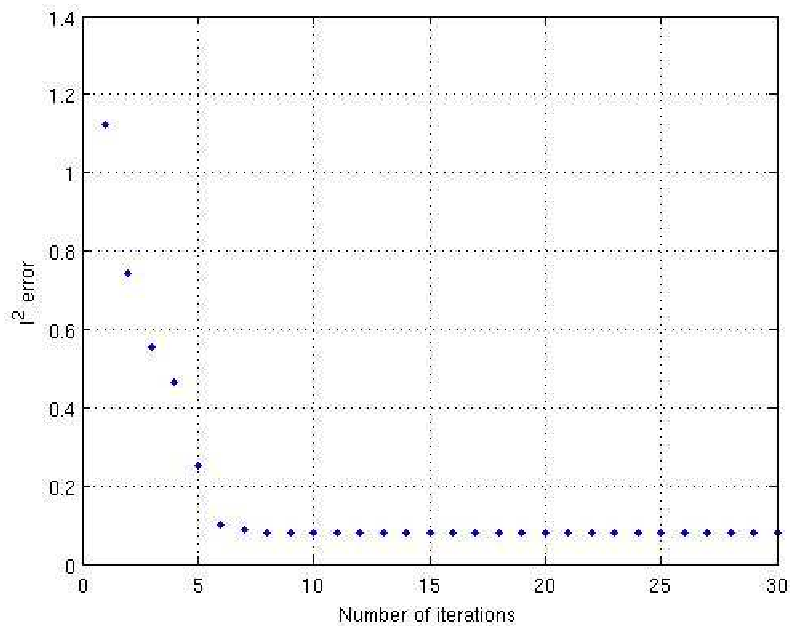

Figure 4. Convergence of IHTs solution

Next, we use the randomized version of IHTs to generate 100 distinct solutions and compute their l^2 errors. We then form a new solution by averaging the 100 random solutions. We find that, although most of the random solutions are inferior to the IHTs solution in terms of reconstruction error, the averaged solution performs better than the deterministic solution and any of the individual random solutions. The reconstruction error achieved by the averaged solution is about 20% less than the error for the standard IHTs solution. Note that the averaged solution is only 38-sparse, compared to the individual solutions which are 8-sparse.

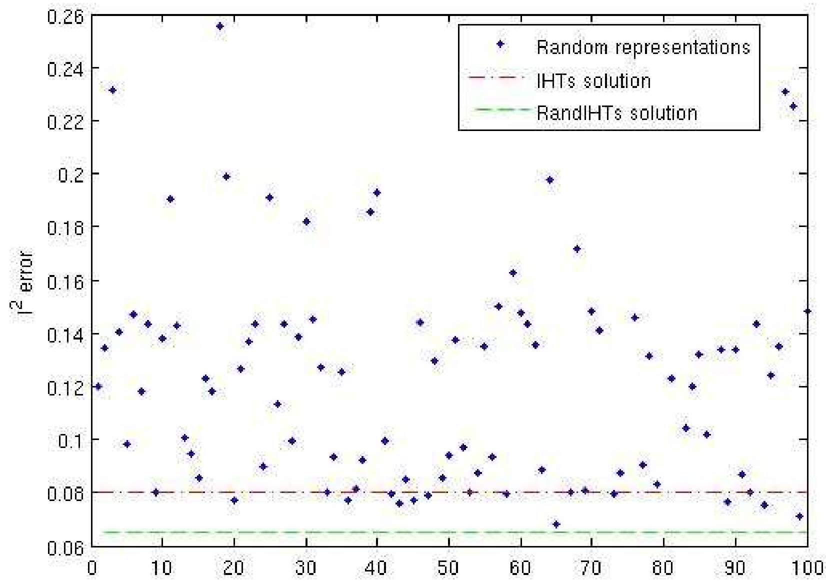

Figure 5. l^2 errors of randomly generated solutions (blue dots), the standard IHTs solution (red dash-dot line), and the averaged solution from 100 randomly generated solutions (green dashed line)

3.2.2 Image Reconstruction with Randomized IHTs

We will now show that IHTs can be used to reconstruct MRI images from undersampled measurements, and that the randomized IHTs can improve on the standard IHTs performance. We start with the MRI scan of a human brain shown in Figure 1. This image is then sampled along 64 radial lines to create the measurement $\hat{y}$ in the Fourier basis. We use an orthogonal wavelet transform (specificially a Daubechies 4 wavelet) as a sparsifying transform, and run the IHTs algorithm with a sparsity of 3000 to obtain a reconstruction. The IHTs reconstruction is a significant improvement over the least squares solution:

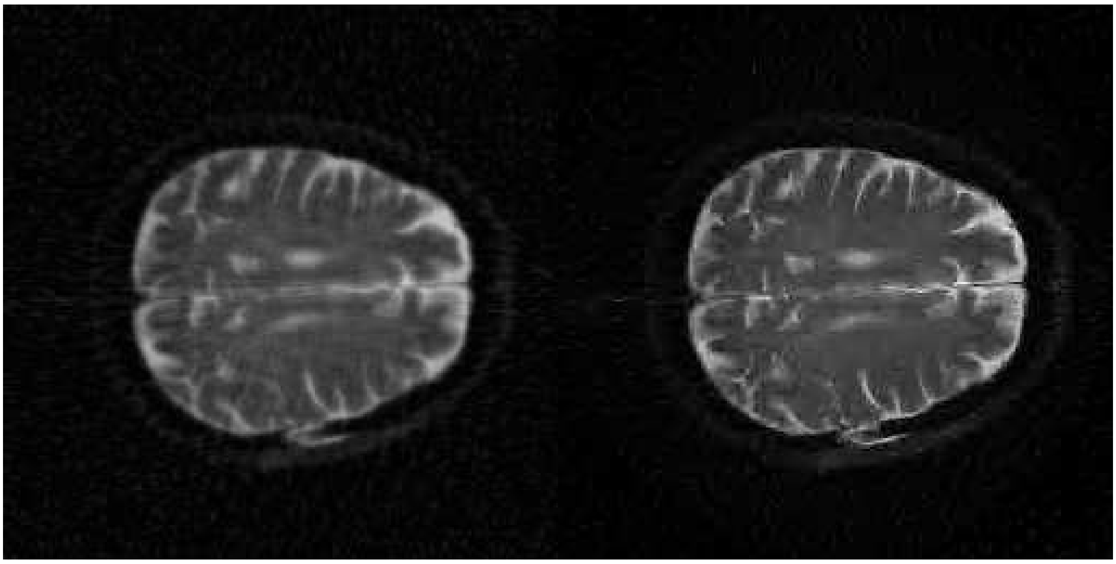

Figure 6. (left) the least squares reconstruction, (right) the IHTs solution with sparsity 3000

We then use the randomized IHTs algorithm to obtain 30 distinct solutions, and then average these 30 solutions together.

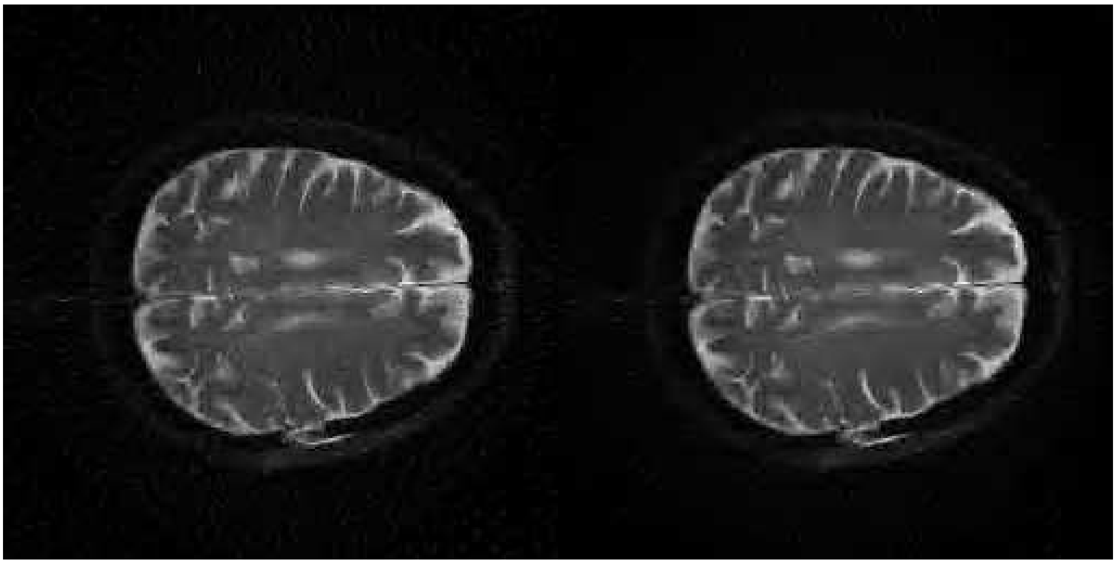

Figure 7. (left) the IHTs solution with sparsity 3000, (right) the solution obtained by averaging 30 RandIHTs representations

Averaging multiple random solutions does indeed give a better reconstruction than the standard IHTs method. We suspect that further improvements can be gained by a better selection of the probability distribution used to randomize the thresholding operator.

The improvement gained by averaging multiple RandOMP solutions becomes more apparent when the image is extra noisy, as we demonstrate by performing a sparse reconstruction of the following image from 50% undersampled measurements:

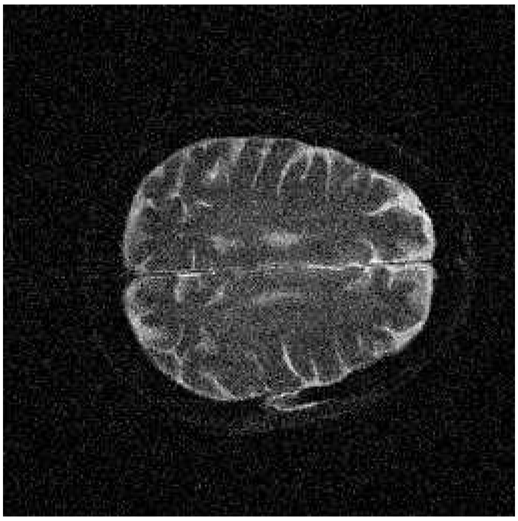

Figure 8. A very noisy image

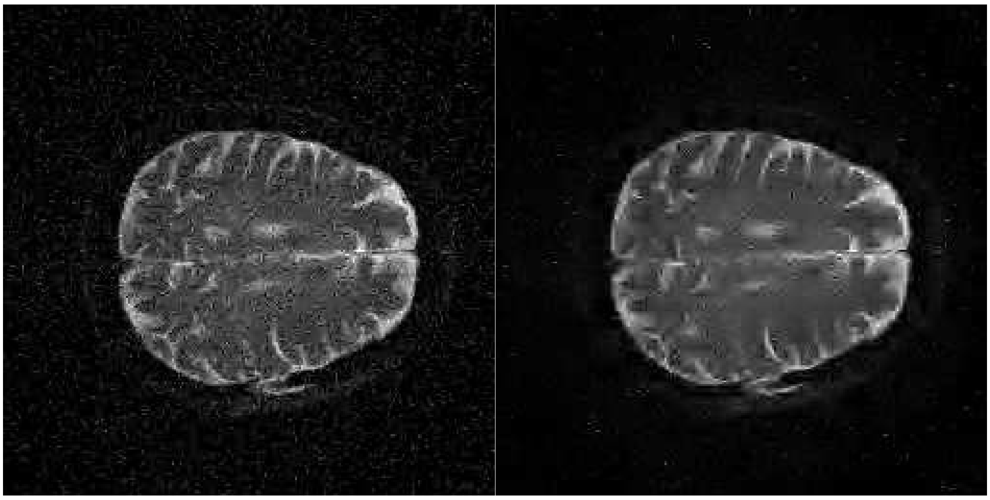

Figure 9. (left) IHTs reconstruction, (right) Reconstruction from averaged RandIHTs solutions

3.2.3 Future Work

We have shown that averaging solutions obtained by a randomized version of IHTs can give better results than the standard IHTs algorithm. We have not derived an optimal way to randomize the thresholding operator; rather, we simply chose a randomization scheme that tends to keep larger coefficients, and which gives random solutions that are only slightly inferior to the deterministic solution. In [2], the RandOMP method selects new atoms in such a way that by averaging RandOMP solutions, we obtain an approximation to the MMSE (...). It should be possible to obtain similar results for the randomized IHTs; we would need to choose an appropriate model for the distribution of the signal coefficients (analogous to σ_x in the RandOMP paper). The derivation of the optimal choice of probability distribution is expected to be similar to the derivation in [2]. Also, we have not examined rigorously the convergence of the randomized version of IHTs; in [5] it is proved that the deterministic version always converges under rather mild restrictions. It should be possible to derive a similar, probabilistic result for the randomized version that guarantees convergence with a certain probability depending on the distribution used to randomize the hard thresholding operator, the sparsity of the signal, and the noise distributions.

4 References

Most recent papers on compressed sensing are collected at the Rice university signal processing website: www.dsp.rice.edu/cs. The interested reader is advised to start with [1], which provides a good introduction to the subject as a whole.

[1] E. Candes and M. Wakin, *An introduction to compressive sampling*, IEEE Signal Pro cessing Magazine, vol. 25(2), pp. 21-30. 2008

[2] M. Elad and I. Yavneh, *A Plurality of Sparse Representations is Better than the Sparsest One Alone*, IEEE Transactions on Informatin Theory, vol. 55(10), pp. 4701-4714. 2009

[3] E. Candes, J. Romberg, T. Tau, *Robust uncertainty principles: Exact signal reconstruction from highly incomlpete frequency information*, IEEE Transactions on Information Theory, vol. 52(2), pp. 489-509. 2006

[4] D. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, vol 52(4), pp. 1289-1306. 2006

[5] T. Blumensath, M. Davies, *Iterative Hard Thresholding for Compressed Sensing*, Applied and Computational Harmonic Analysis, vol 27(3), pp.265-274. 2009