COMPRESSIVE SENSING WITH HIGHLY COHERENT DICTIONARIES

SARA COHEN

Abstract. Compressive sensing is an emerging field based on the discovery that sparse signals and images can be reconstructed from highly incomplete information. Conventional approaches follow Shannons theorem, which states that the sampling rate must be twice the maximum frequency present in the signal. In the case that the sensing matrix is highly coherent, which happens when signals are only sparse in a truly redundant dictionary, one must consider less traditional approaches to reconstruct the signals. Sensing matrices are highly coherent in imaging problems such as radar and medical imaging. This work compares existing methods to solve the coherence problem. The first method attempts recovery via an l_1-analysis optimization problem, the second uses l_1-minimization, and the third method uses algorithms based on band exclusion and local optimization. Detailed comparisons demonstrate the superiority of the l_1-minimization method which minimizes the distance of the nonzero entries in the sparse vector x.

1. Introduction

This paper is by no means an exhaustive survey of the literature on compressive sensing. It is merely an account of others own work and thinking in this area which includes a large number of references to other people's work. The following introduction is credited to the authors: Massimo Fornaier, Hoger Rauhut, Emmanuel J. Candes, Yonnia C. Eldar, Deanna Needell, Paige Randall, Albert Fannjiang, and Wenjing Liao.

The Nyquist/Shannon sampling theorem states that to avoid losing information when measuring a signal, one must sample twice as fast as the bandwidth of the signal. Similarly, the fundamental theorem of algebra suggests that the number of collected measurements of a discrete finite-dimensional signal should be at least as large as its length or dimension in order to ensure reconstruction. However, in some applications, increasing the sampling rate is either quite expensive or not feasible in the first place.

Date: May 23, 2012.

Compressive sensing is a rapidly growing field which presents a new method that allows signals to be captured and measured at a much lower sampling rate despite common wisdom. There are many applications of compressed sensing which range from medical imaging to radar and remote sensing to video electronics.

Compressive sensing depends on the empirical observation that many types of signals can be well-approximated by a sparse expansion, i.e only a small number of non-zero coefficients, in terms of a suitable basis. This is the key to many lossy compression techniques such as JPEG or MP3 [3]. Lossy compression refers to a data encoding method that compresses data by discarding, or losing some of it. A compression is obtained by storing only the largest basis coefficients. When reconstructing the signal, the non-stored coefficients are set to zero. This is a reasonable strategy when full information of the signal is available. However, when the signal has to be obtained by a costly, lengthy, or otherwise difficult sensing procedure, this strategy seems to be a waste of resources. The time and money is spent in order to obtain full measurements and then most of the information is thrown away during the compression stage. The goal would be to be able to obtain the compressed version of the signal more directly by taking a smaller number of measurements of the signal in the first place. It is not clear whether or not this is possible since measuring the large coefficients requires knowing their location beforehand. Nevertheless, compressive sensing provides a way to reconstruct a compressed version of the original signal by taking only a small amount of linear and non-adaptive measurements. The particular number of measurements is comparable to the compressed size of the signal. The measurements have to be suitably designed and surprisingly, all good measurement matrices designed thus far have been random.

In terms of compressive sensing, the interest is in the undersampled case, meaning there are fewer measurements than unknown signal values. There are countless numbers of applications in which this type of problem arises. For example, in radiology and biomedical imaging one is typically able to collect far fewer measurements about an image of interest than the number of unknown pixels. In wideband radio frequency signal analysis, one may be able to acquire a signal at a rate which is far lower than the Nyquist rate because of current limitations in Analog-to-Digital Converter technology. Lastly, gene expression studies also provide examples in that one would like

to infer the gene expression level of thousands of genes from a low number of observations.

It is another important feature of compressive sensing that useful reconstruction can be performed by using efficient algorithms. Since the attention is in the immensely undersampled case, the linear system describing the measurements is underdetermined and therefore has infinitely many solutions. The main idea is that the sparsity helps in isolating the original vector. The first naïve approach to a reconstruction algorithm entails searching for the sparsest vector that is consistent with the linear measurements. This leads to the combinatorial l_0-problem which is unfortunately NP-hard in general. There are essentially two approaches for alternative algorithms. The first is convex relaxation leading to l_1-minimization while the second constructs greedy algorithms. A greedy algorithm is an algorithm that follows the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding a global optimum. This paper will explain basic properties of the measurement matrix which ensure sparse recovery by l_1-minimization such as the null space property (NSP) and the restricted isometry property (RIP).

Compressed sensing suggests obtaining a signal $x \in \mathbb{R}^n$ by collecting m linear measurements of the form $y_k = \langle a_k, x\rangle + z_k$, $1 \leq k \leq m$, or in matrix notation

$$y = Ax + z, \tag{1.1}$$

where A is an $m \times n$ sensing matrix with m usually smaller than n by one or numerous orders of magnitude and z is an error term modeling measurement errors.

The matrix A is chosen independently of x. Under specific conditions of the matrix A, compressive sensing indicates that as long as the unknown signal x is reasonably sparse (contains mostly zeros) it is possible to recover x. The solution simplifies to

$$\min_{\tilde{x}\in\mathbb{R}^n} \|\tilde{x}\|_1 \text{ subject to } \|A\tilde{x} - y\|_2 \leq \epsilon, \tag{1.2}$$

where $\|\cdot\|_2$ denotes the standard Euclidean norm, $\|x\|_1 = \sum |x_i|$ is the l_1-norm and ϵ^2 is a likely upper bound on the noise power $\|z\|_2^2$.

Compressed sensing typically compares the quality of the reconstruction from the data y and the model $y = Ax + z$ with the s most significant entries of x. Let x_s denote the vector consisting of the s largest coefficients

of $x \in \mathbb{R}^n$ in magnitude

$$x_s = \arg\min_{\|\tilde{x}\|_0 \leq s} \|x - \tilde{x}\|_2, \tag{1.3}$$

where $\|x\|_0 = |\{i : x_i \neq 0\}|$. This vector x_s has s nonzero entries and is the best s-sparse approximation to the vector x. In other words, $x - x_s$ is the tail of the signal and consists of the smallest $n - s$ entries of x. It has been established by Romberg and Tao that (1.2) recovers a signal $\tilde{x}$ observing

$$\|\tilde{x} - x\|_2 \leq C_0 \frac{\|x - x_s\|_1}{\sqrt{s}} + C_1 \epsilon, \tag{1.4}$$

given that the $2s$-restricted isometry constant of A obeys $\delta_{2s} < 0.4652$. Since the recovery error from (1.2) is proportional to the measurement error and the tail of the signal, $x - x_s$, the approximation error of a nearly sparse signal is very small and the error completely vanishes for precisely sparse signals.

Definition 1.1. For an $m \times n$ measurement matrix A, the s-restricted isometry constant δ_s of A is the smallest quantity such that

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2, \tag{1.5}$$

holds for all s-sparse signals x. Then matrix A is said to satisfy the ***s*-restricted isometry property** with the s-restricted isometry constant δ_s. The RIP characterizes matrices which are almost orthonormal when operating on sparse vectors and it assures accurate recovery of signals that are nearly sparse in a highly overcomplete and coherent dictionary (frame). [1]

The condition (1.4) which $\tilde{x}$ must obey is quite natural since it prevents sparse signals from lying in the nullspace of the sensing matrix A. A matrix having a small restricted isometry constant means that every subset of s or fewer columns is nearly orthonormal. Many matrices with Gaussian, Bernoulli, or Fourier entries have small restricted isometry constants when the number of measurements m is on the order of $s \log(n/s)$.

Compressive sensing is based on the observation that many types of real-world signals and images have a sparse expansion in terms of a suitable basis. This means that the expansion has only a small number of significant terms, or in other words, that the coefficient vector can be well-approximated with one having only a small number of nonzero entries.

The **support** of a vector x is denoted $\text{supp}(x) = \{j : x_j \neq 0\}$, and

$$\|x\|_0 := |\mathrm{supp}(x)|.$$

It has become common to call $\|\cdot\|_0$ the l_0-norm, although it is not even a quasi-norm. A vector x is called ***s*-sparse** if $\|x\|_0 \leq s$. For $s \in \{1, 2, ..., N\}$,

$$\textstyle\sum_s := \{x \in \mathbb{C}^N : \|x\|_0 \leq s\}$$

denotes the set of s-sparse vectors. Furthermore, the best s-term approximation error of a vector $x \in \mathbb{C}^N$ in l_p is defined as

$$\sigma_s(x)_p = \inf_{z \in \sum_s} \|x - z\|_p.$$

If $\sigma_s(x)$ decays quickly in s then x is called **compressible**. In order to compress x one may simply store only the s largest entries. When reconstructing x from its compressed version the nonstored entries are simply set to zero, and the reconstruction error is $\sigma_s(x)_p$. It is emphasized at this point that the procedure of obtaining the compressed version of x is adaptive and nonlinear since it requires the search of the largest entries of x in absolute value. Specifically, the location of the non-zeros is a nonlinear type of information.

The best s-term approximation of x can be obtained using the nonincreasing rearrangement $r(x) = (|x_i|, \ldots, |x_{iN}|)_T$, where i_j denotes a permutation of the indices such that $|x_{i_{j+1}}| \leq |x_{i_j}|$ for $j = 1, \ldots, N-1$.

Then it is straightforward to check that

$$\sigma_s(x)_p := \left(\sum r_j(x)^p \right)^{1/p}, 0 < p < \infty$$

and the vector $x_{[s]}$ derived from x by setting to zero all the $N - s$ smallest entries in absolute value is the best s-term approximation,

$$x_{[s]} = \arg\min_{z \in \sum_s} \|x - z\|_p,$$

for any $0 < p \leq \infty$.

The next lemma states essentially that l_q-balls with small q (ideally $q \leq 1$) are good models for compressible vectors.

Lemma 1.2. *Let $0 < q < p \leq \infty$ and set $r = \frac{1}{q} - \frac{1}{p}$. Then*

$$\sigma_s(x)_p \leq s^{-r}, s = 1, 2, \ldots, N \text{ for all } x \in B_q^N.$$

Proof. Let T be the set of indices of the s-largest entries of x in absolute value. The non-increasing rearrangement satisfies $|r_s(x)| \leq |x_j|$ for all $j \in T$, and therefore

$$s r_s(x)^q \leq \textstyle\sum_{j \in T} |x_j|^q \leq \|x\|_q^q \leq 1.$$

Hence, $r_s(x) \leq s^{\frac{-1}{q}}$. Therefore

$$\sigma_s(x)_p^p = \sum_{j\notin T} |x_j|^p \leq \sum_{j\notin T} r_s(x)^{p-q}|x_j|^q \leq s^{-\frac{p-q}{q}} \|x\|_q^q \leq s^{-\frac{p-q}{q}},$$

which implies $\sigma_s(x)_p \leq s^{-r}$. □

The null space property is fundamental in the analysis of l_1-minimization.

Definition 1.3. A matrix $A \in \mathbb{C}^{m\times N}$ is said to satisfy the **null space property** (NSP) of order s with constant $\gamma \in (0,1)$ if

$$\|\eta_T\|_1 \leq \gamma \|\eta_{T^c}\| - 1, \tag{1.6}$$

for all sets $T \subset \{1,\dots,N\}, with\ \#T \leq s$ and for all $\eta \in \ker A$.

The following sparse recovery result is based on this notion.

Theorem 1.4. *Let $A \in \mathbb{C}^{m\times N}$ be a matrix that satisfies the NSP of order s with constant $\gamma \in (0,1)$. Let $x \in \mathbb{C}^N$ and $y = Ax$ and let x^* be a solution of the l_1-minimization problem. Then*

$$\|x - x^*\|_1 \leq \frac{2(1+\gamma)}{1-\gamma}\sigma_s(x)_1. \tag{1.7}$$

In particular, if x is s-sparse $x^* = x$.

Proof. Let $\eta = x^* - x$. Then $\eta \in \ker A$ and

$$\|x^*\|_1 \leq \|x\|_1,$$

because x^* is a solution of the l_1-minimization problem. Let T be the set of the s-largest entries of x in absolute value. One has

$$\|x_T^*\|_1 + \|x_{T^c}^*\|_1 \leq \|x_T\|_1 + \|x_T^c\|_1.$$

It follows immediately from the triangle equality that

$$\|x_T\|_1 - \|\eta_T\|_1 + \|\eta_{T^c}\|_1 - \|x_{T^c}\|_1 \leq \|x_T\|_1 + \|x_{T^c}.$$

Hence,

$$\|\eta_{T^c}\|_1 \leq \|\eta_T\|_1 + 2\|x_{T^c}\|_1 \leq \gamma\|\eta_{T^c}\|_1 + 2\sigma_s(x)_1,$$

Or, equivalently,

$$\|\eta_{T^c}\|_1 \leq \frac{2}{1-\gamma}\sigma_s(x)_1. \tag{1.8}$$

Finally,

$$\|x - x^*\|_1 = \|\eta_T\|_1 + \|\eta_{T^c}\|_1 \leq (\gamma+1)\|\eta_{T^c}\|_1 \leq \frac{2(1+\gamma)}{1-\gamma}\sigma_s(x)_1$$

and the proof is completed. □

One can also show that if all s-sparse x can be recovered from $y = Ax$ using l_1-minimization then necessarily A satisfies the NSP of order s with some constant $\gamma \in (0,1)$. Therefore, the NSP is actually equivalent to sparse l_1-recovery.

The RIP implies the NSP as shown in the following lemma.

Lemma 1.5. *Assume that $A \in \mathbb{C}^{m\times N}$ satisfies the RIP of order $S = s + h$ with constant $\delta_S \in (0,1)$. Then A has the NSP of order s with constant*

$$\gamma = \sqrt{\frac{s}{h}\frac{1+\delta_S}{1-\delta_S}}.$$

Proof. Let $\eta \in N = \ker A$ and $T \subset \{1,\dots,N\}, with\ \#T \le s$. Define $T_0 = T$ and $T_1, T_2, \dots, T_s$ to be disjoint sets of indexes of size at most h, associated to a non-increasing rearrangement of the entries of $\eta \in N$, i.e.,

$$|\eta_j| \le |\eta_i| \text{ for all } j \in T_l, i \in T_l, 1 \le l^{'} \le l. \tag{1.9}$$

Note that $A\eta = 0$ implies $A\eta_{T_0\cup T_1} = -\sum_{j=2}^{s} A\eta_{T_j}$. Then, from the Cauchy-Schwarz inequality, the RIP, and the triangle inequality, the following sequence is deduced,

$$\begin{aligned} \|\eta_T\|_1 &\le \sqrt{s}\|\eta_T\|_2 \le \sqrt{s}\|\eta_{T_0\cup T_1}\|_2 \\ &\le \sqrt{\frac{s}{1-\delta_S}}\|A\eta_{T_0\cup T_1}\|_2 = \sqrt{\frac{s}{1-\delta_s}}\|A\eta_{T_2\cup T_3\cup\dots T_s}\|_2 \\ &\le \sqrt{\frac{s}{1-\delta_S}}\sum_{j=2}^{s}\|A\eta_{T_j}\|_2 \le \sqrt{\frac{1+\delta_S}{1-\sigma_S}}\sqrt{s}\sum_{j=2}^{s}\|\eta_{T_j}\|_2. \end{aligned} \tag{1.10}$$

It follows from (1.9) that $|\eta_i| \le |\eta_l$ for all $i \in T_{j+1}$ and $l \in T_j$. Taking the sum over $l \in T_j$ first and then the l_2-norm over $i \in T_{j+1}$ yields

$$|\eta_i| \le h^{-1}\|\eta_{T_j}\|_1, \text{ and } \|\eta_{T_{j+1}}\|_2 \le h^{-1/2}\|\eta_{T_j}\|_1.$$

Using the latter estimates in (1.10) gives

$$\|\eta_T\|_1 \le \sqrt{\frac{1+\delta_S}{1-\delta_S}\frac{s}{h}}\sum_{j=1}^{s-1}\|\eta_{T_j}\|_1 \le \sqrt{\frac{1+\delta_S}{1-\delta_S}\frac{s}{h}}\|\eta_{T^c}\|_1, \tag{1.11}$$

and the proof is finished. □

Taking $h = 2s$ above shows that $\delta_{3s} < 1/3$ implies $\gamma < 1$. By Theorem 1.4 recovery of all s-sparse vectors by l_1-minimization is then guaranteed. Additionally, stability in l_1 is also ensured. The next theorem shows that RIP implies also a bound on the reconstruction error in l_2.

Theorem 1.6. *Assume $A \in \mathbb{C}^{m\times N}$ satisfies the RIP of order $3s$ with $\delta_{3s} < 1/3$. For $x \in \mathbb{C}^N$, let $y = Ax$ and x^* be the solution of the l_1-minimization problem. Then*

$$\|x - x^*\|_2 \leq C\frac{\sigma_s(x)_1}{\sqrt{s}},$$

with $C = \dfrac{2}{1-\gamma}\left(\dfrac{\gamma+1}{\sqrt{2}} + \gamma\right)$, *and* $\gamma = \sqrt{\dfrac{1+\delta_{3s}}{2(1-\delta_{3s})}}$.

Proof. Similarly as in the proof of Lemma 1.5, let $\eta = x^* - x \in N = \ker A$, $T_0 = T$ the set of the $2s$-largest entries of η in absolute value, and T_j's of size at most s corresponding to the non-increasing rearrangement of η. Then using (1.10) and (1.11) with $h = 2s$ of the previous proof,

$$\|\eta_T\|_2 \leq \sqrt{\frac{1+\delta_{3s}}{2(1-\delta_{3s})}}s^{-1/2}\|\eta_{T^c}\|_1.$$

From the assumption $\delta_{3s} < 1/3$ it follows that $\gamma := \sqrt{\dfrac{1+\delta_{3s}}{2(1-\delta_{3s})}} < 1$. Lemmas 1.2 and 1.5 yield

$$\begin{aligned}\|\eta_{T^c}\|_2 = \sigma_{2s}(\eta)_2 &\leq (2s)^{-\frac{1}{2}}\|\eta\|_1 = (2s)^{-1/2}(\|\eta_T\|_1 + \|\eta_{T^c}\|_1)\\ &\leq (2s)^{-1/2}(\gamma\|\eta_{T^c}\|_1 + \|\eta_{T^c}\|_1) \leq \frac{\gamma+1}{\sqrt{2}}s^{-1/2}\|\eta_{T^c}\|_1.\end{aligned} \tag{1.12}$$

Since T is the set of $2s$-largest entries of η in absolute value, it holds

$$\|\eta_{T^c}\|_1 \leq \|\eta_{(\mathrm{supp}x_{[2s]})^c}\|_1 \leq \|\eta_{(\mathrm{supp}x_{[s]})^c}\|_1, \tag{1.13}$$

where $x_{[s]}$ is the best s-term approximation to x. The use of this latter estimate, combined with the inequality 1.8, finally gives

$$\begin{aligned}\|x - x^*\|_2 &\leq \|\eta_T\|_2 + \|\eta_{T^c}\|_2\\ &\leq \Big(\frac{\gamma+1}{\sqrt{2}} + \gamma\Big)s^{-1/2}\|\eta_{T^c}\|_1\\ &\leq \frac{2}{1-\gamma}\Big(\frac{\gamma+1}{\sqrt{2}} + \gamma\Big)s^{-1/2}\sigma_s(x)_1.\end{aligned} \tag{1.14}$$

This concludes the proof. □

The compressive sensing techniques described above are used when the signals are sparse with respect to an orthonormal basis. However, there are often times when a signal is not sparse in an orthonormal basis. Sparsity if often expressed in terms of an overcomplete dictionary. An overcomplete dictionary refers to a dictionary or matrix, which has many more columns than rows. The use of overcomplete dictionaries is now widespread in the

field of compressed sensing. They are often used when working in situations in which no good orthonormal basis is known to exist. Additionally, overcomplete dictionaries provide benefits in certain applications such as deconvolution, tomography, and other signal-denoising problems.

In the overcomplete dictionary situation our signal $f \in \mathbb{R}^n$ is now expressed as $f = Dx$ where $D \in \mathbb{R}^{n \times d}$ is some overcomplete dictionary. Now, consider the case in which the sensing matrix A has Gaussian entries. If D is not a unitary matrix then the matrix AD will have correlated columns and thus would not satisfy the traditional requirements imposed by compressive sensing. This paper will discuss the potential of good recovery when the columns are highly correlated.

Traditional assumptions imposed by compressive sensing and sparse signal recovery say that the measurement matrix must have uncorrelated columns.

Definition 1.7. The **coherence** of a matrix B is defined as

$$\mu(B) = \max_{j<k} \frac{|\langle B_j, B_k \rangle|}{\|B_j\|_2 \|B_k\|_2} \tag{1.15}$$

where B_j and B_k denote columns in B. A dictionary is considered **incoherent** if μ is small. Traditional approaches to compressive sensing require that the measurement matrix satisfy a strict incoherence property such as the restricted isomtery property (RIP).

Applying Gershgorin's disc theorem to $B_T^* B_T - I$ with $\#T = s$ shows that

$$\delta_s \leq (s-1)\mu. \tag{1.16}$$

Several explicit examples of matrices are known which have small coherence $\mu = \mathcal{O}(1/\sqrt{m})$. A simple one is the concatenation $B = (I|F) \in \mathbb{C}^{m \times 2m}$ of the identity matrix and the unitary Fourier matrix $F \in \mathbb{C}^{m \times m}$ with entries $F_{j,s} = m^{-1/2} e^{2\pi i j s/m}$. It is easily seen that $\mu = 1/\sqrt{m}$ in this case. In these cases, $\delta_s \leq C \frac{s}{\sqrt{m}}$. Combining this estimate with the recovery results for l_1-minimization above shows that all s-sparse vectors x can be stably recovered from $y = Ax$ via l_1-minimization provided

$$m \geq C' s^2. \tag{1.17}$$

At first sight one might be satisfied with this condition since if s is very small compared to N then still m might be chosen smaller than N and all s-sparse vectors can be recovered from the undersampled measurements $y = Ax$. Although this is great news for a start, one might nevertheless hope

that (1.16) can be improved. In particular, one may actually expect that a linear scaling of m in s should be enough to guarantee sparse recovery by l_1-minimization. The existence of matrices, which provide recovery conditions of the form $Cs\log^{\alpha}(N) \leq m$ (or similar) with some $\alpha \geq 1$, is shown in the next section. Unfortunately, such results cannot be shown by simply using the coherence because of the general lower bound

$$\mu \geq \sqrt{\frac{N-m}{m(N-1)}} \sim \frac{1}{\sqrt{m}} (N \text{ sufficiently large}). \tag{1.18}$$

In particular, it is not possible to overcome the quadratic bottleneck in (1.17) by using Gershgorins theorem or Riesz-Thorin interpolation between $\|\cdot\|_{1\to 1}$ and $\|\cdot\|_{\infty\to\infty}$. In order to improve on (1.17) one has to take into account also cancellations in the Gramian $A_T^* A_T - I$, and this task seems to be quite difficult using deterministic methods. Therefore, it will not come as a surprise that the major breakthrough in compressive sensing was obtained with random matrices. It is indeed easier to deal with cancellations in the Gramian using probabilistic techniques.

Optimal estimates for the RIP constants in terms of the number m of measurement matrices can be obtained for Gaussian, Bernoulli, or more general subgaussian random matrices. Let X be a random variable. Then one defines a random matrix $A = A(\omega)$, $\omega \in \Omega$, as the matrix whose entries are independent realizations of X, where $(\Omega, \Sigma, \mathbb{P})$ is their common probability space. One assumes further that for any $x \in \mathbb{R}^N$ we have the identity $\mathbb{E}\|Ax\|_2^2 = \|x\|_2^2$, where $\mathbb{E}$ denotes expectation.

The starting point for the simple approach is a concentration inequality of the form

$$\mathbb{P}(|\|Ax\|_2^2 - \|x\|_2^2| \geq \delta \|x\|_2^2) \leq 2e^{-c_0 \delta^2 m}, 0 < \delta < 1, \tag{1.19}$$

where $c_0 > 0$ is some constant. The two most relevant examples of random matrices which satisfy the above concentration are Gaussian and Bernoulli matrices. Based on the concentration equality the following estimate on RIP constants can be shown.

Theorem 1.8. *Let $A \in \mathbb{R}^{m\times N}$ be a random matrix satisfying the concentration property. Then there exists a constant C depending only on c_0 such that the restricted isometry constant of A satisfies $\delta_k \leq \delta$ with probability exceeding $1-\epsilon$ provided*

$$m \geq C\delta_{-2}(k\log(N/m) + \log(\epsilon_{-1})).$$

Combining this RIP estimate with the recovery results for l_1-minimization shows that all s-sparse vectors $x \in \mathbb{C}^N$ can be stably recovered from a random draw of A satisfying (1.19) with high probability provided

$$m \geq Ck \log(N/m). \tag{1.20}$$

Up to the log-factor this provides the desired linear scaling of the number m of measurements with respect to the sparsity s. Furthermore, the above condition cannot be further improved; in particular, the log-factor cannot be removed.

Theorem 1.9. *If a system of linear equations $Ax = b$ has a solution obeying $\|x\|_0 < \frac{1}{2}(1 + 1/\mu(A))$, this solution is necessarily the sparsest possible.*

The coherence can never be smaller than $\frac{1}{\sqrt{n}/2}$, and therefore, the cardinality bound of the above theorem is never larger than $\frac{1}{\sqrt{n}}$.

Theorem 1.10. *For a system of linear equations $Ax = b$ ($A \in \mathbb{R}^{m \times n}$ full-rank with $n < m$), if a solution x exists obeying*

$$\|x\|_0 < \tfrac{1}{2}\left(1 + \tfrac{1}{\mu(A)}\right),$$

then l_1-minimization is guaranteed to find it exactly.

Next I would like to discuss how well one can estimate the response Ax where A is a matrix and x is an s-sparse vector. The generic s-sparse model is defined as follows:

(1) The support $I \subset \{1, \ldots, p\}$ of the s nonzero coefficients of A is selected uniformly at random.
(2) Conditionally on I, the signs of the nonzero entries of A are independent and equally likely to be -1 or 1.

No assumptions are made on the amplitudes. In some sense, this is the simplest statistical model. It says that all subsets of a given cardinality are equally likely, or in other words, one is not biased towards certain variables nor is there any reason to believe that a given coefficient is positive or negative.

On the other hand, suppose that our sparse vector x does not satisfy a minimum distance between its nonzero entries. To test this scenario I chose a 64×256 DFT and a sparse vector with 10 consecutive nonzero entries. Using l_1-minimization I attempted to recover the signal; however, as apparent in Figure 1 the reconstruction was not very accurate. There was a relative error of 1.214253256401029 which is quite high. Therefore signal recovery

via l_1-minimization fails unless the sensing matrix or the signal itself satisfies certain conditions.

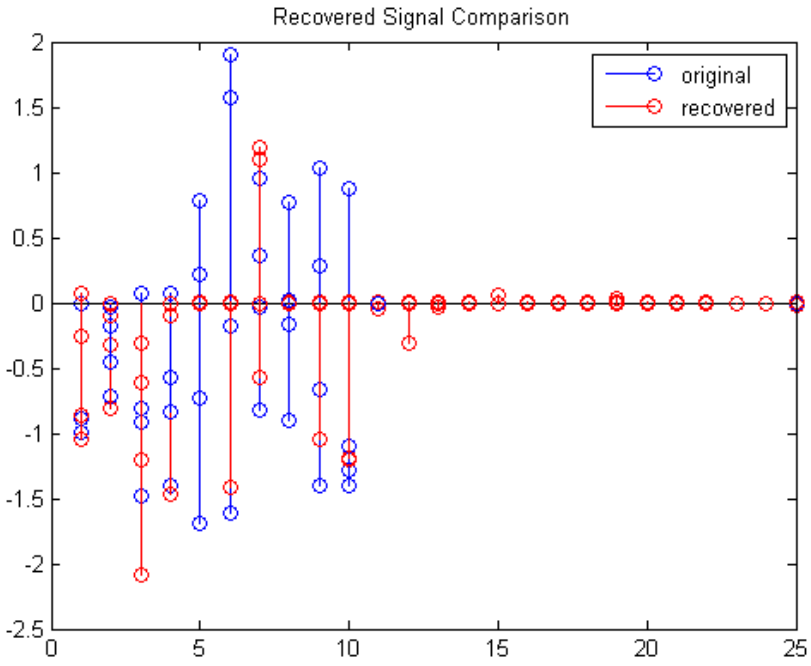

FIGURE 1

If two columns are highly correlated it would be nearly impossible to distinguish whether the signal comes from one or the other. For example, suppose we are not undersampling and that A is the identity matrix. We then observe $y = Dx$. Suppose that the first two columns are identical. Then it would not be possible to reconstruct a unique sparse signal x from measurements $y = ADx$. However, instead of recovering the coefficient vector x we are interested in the actual signal Dx. Thus the high correlation between the columns in D does not create a problem since differentiation between the coefficient vectors is not the goal. Therefore the low coherence of D may not be a necessary requirement for recovery.

To introduce my results, I will first discuss a concrete situation. I first assume that the sensing matrix or dictionary D has Gaussian entries with a decaying rate ϵ and shift 0.25. Next, I conduct the same tests using Fourier matrices. I am interested in recovering the actual signal $f = Dx$ instead of the sparse vector x.

The goal is to $min_{f\in\mathbb{R}}\|D^*f\|_1$ subject to $\|Af-y\|_2 \leq \epsilon$ given measurements y where $y = Af + z$ where z is noise. I want to find an f that fits the data up to ϵ which is related to the noise by $\|z\|_2 \approx \epsilon$.

I tried three different methods to solve this compressed sensing problem. The first method is l_1-analysis, the second is l_1-minimization, and the third method is band exclusion. The results show the superiority of l_1-minimization.

2. Methods

2.1. **l_1-analysis.** This section proposes a reconstruction from $y = Af + z$ by the method of l_1-analysis:

$$\tilde{f} = \underset{\tilde{f}\in\mathbb{R}^n}{\arg\min} \|D^*\tilde{f}\|_1 \text{ subject to } \|A\tilde{f} - y\|_2 \leq \epsilon \tag{2.1}$$

where again ϵ is a likely upper bound on the noise level $\|z\|_2$. Empirical studies have shown very promising results for the l_1-analysis problem. Its geometry has been studied as well as its applications to image restoration. However, there are no results in the literature about its performance in regard to the case where D is a redundant dictionary made of Gaussian functions. The solution to (2.1) is very accurate provided that D^*f has rapidly decreasing coefficients.

Theorem 2.1. *Let D be an arbitrary $n \times n$ tight frame and let A be a $m \times n$ Gaussian matrix with m on the order of $s\log(d/s)$. Then the solution $\tilde{f}$ to* (2.1) *obeys*

$$\|\tilde{f} - f\|_2 \leq C_0\epsilon + C_1\frac{\|D^*f - (D^*f)_s\|_1}{\sqrt{s}},$$

for some numerical constants C_0 and C_1, and where $(D^*f)_s$ is the vector consisting of the largest s entries of D^*f in magnitude as in (1.3).

2.2. **Band Exclusion.** This method relies on the importance of band exclusion. While many l_1-minimization algorithms require either incoherence or the Restricted Isometry Property to have good performances, this method does not. [2]

According to theory of optimal recovery, for time sampling in [0,1], the minimum resolvable length in the frequency domain is unity. This is the Rayleigh threshold and this length will be referred to as the Rayleigh length (RL). Thus, for the traditional inversion methods to work, it is essential that the grid spacing be no less than 1 RL. In the compressed sensing setting,

the Rayleigh threshold is closely related to the decay property of the mutual coherence. [2]

Without any prior information about the object support, the gridding error for the resolved grid, however, can be as large as the data itself, creating an unfavorable condition for sparse reconstruction. To reduce the gridding error, it is natural to consider the fractional grid

$$\mathbb{Z}/F = \{j/F : j \in \mathbb{Z}\}$$

with some large integer $F \in \mathbb{N}$ called the refinement factor. The relative gridding error is roughly inversely proportional to the refinement factor; however, the mutual coherence increases with F as the near-by columns of the sensing matrix become highly correlated.

The hope is that if the objects are sufficiently separated with respect to the coherence band, then the problem of a huge condition number associated with unresolved grids can be somehow circumvented and the object support can be approximately reconstructed.

The first technique that I will introduce to take advantage of the information that objects are widely separated is called Band Exclusion and it can be easily embedded in the greedy algorithm, Orthogonal Matching Pursuit (OMP). The following proposition is a standard performance guarantee for OMP.

Proposition 2.2. Suppose that the sparsity s of the signal vector x satisfies

$$\mu(A)(2s-1) + 2\frac{\|e\|_2}{x_{min}} < 1,$$

where $x_{min} = min_k|x_k| = |x_s|$. Let $\tilde{x}$ denote the output of OMP reconstruction. Then

$$supp(\tilde{x}) = supp(x),$$

where $supp(x)$ is the support of x. The ideal case where $e = 0$, reduces to

$$\mu(A) < \frac{1}{2s-1},$$

which is near the threshold of OMP's capacity for exact reconstruction of arbitrary objects of sparsity s.

Intuitively speaking, if the objects are not in each other's coherence band, then it should be possible to localize the objects approximately within their respective coherence bands, no matter how large the mutual coherence is.

Define the η-coherence band of the index k to be the set

$$B_\eta(k) = \{i \mid \mu(i,k) > \eta\},$$

and the η-coherence band of the index set S to be the set

$$B_\eta(S) = \bigcup_{k\in S} B_\eta(k).$$

Due to the symmetry $\mu(i,k) = \mu(k,i)$, for all $i,k \in B_\eta(k)$ if and only if $k \in B_\eta(i)$. Denote

$$B_\eta^{(2)}(k) := B_\eta(B_\eta(k)) = \cup_{j\in B_\eta(k)} B_\eta(j), \tag{2.2}$$

$$B_\eta^{(2)}(S) \equiv B_\eta(B_\eta(S)) = \cup_{k\in B_\eta(S)} B_\eta(k). \tag{2.3}$$

To embed BE into OMP, we make the following change to the matching step

$$i_{\max} = \arg\min_i |\langle r^{n-1}, a_i\rangle|,\ i \notin B_\eta^{(2)}(S^{n-1}), n = 1, 2, \dots$$

meaning that the double η-band of the estimated support in the previous iteration is avoided in the current search. This is natural if the sparsity pattern of the object is such that $B_\eta(j), j \in \mathrm{supp}(x)$ are pairwise disjoint. We call the modified algorithm the Band-excluded Orthogonal Matching Pursuit (BOMP) which is formally stated in the following Algorithm.

Algorithm 1 Band-excluded Orthogonal Matching Pursuit (BOMP)

Input: $A, b, \eta > 0$

Initialization: $x^0 = 0, r^0 = b, S^0 = \emptyset$

Iteration: For $n = 1, \dots, s$

(1) $i_{\max} = \arg\max_i |\langle r^{n-1}, a_i\rangle|,\ i \notin B_\eta^{(2)}(S^{n-1})$
(2) $S^n = S^{n-1} \cup \{i_{\max}\}$
(3) $x^n = \arg\min_z \|Az - b\|_2$ s.t. $\mathrm{supp}(z) \in S^n$
(4) $r^n = b - Ax^n$

Output: x^s

A main theoretical result of the present paper is the following performance guarantee for BOMP.

Theorem 2.3. *Let x be s-sparse. Let $\eta > 0$ be fixed. Suppose that*

$$B_\eta(i) \cap B_\eta^{(2)}(j) = \emptyset, \forall i, j \in supp(x), \tag{2.4}$$

and that

$$\eta(5s-4)\frac{x_{\max}}{x_{\min}} + \frac{5\|e\|_2}{2x_{\min}} < 1, \tag{2.5}$$

where

$$x_{\max} = \max_k |x_k|, x_{\min} = \min_k |x_k|.$$

Let $\tilde{x}$ be the BOMP reconstruction. Then $supp(\tilde{x}) \subseteq B_\eta(supp(x))$ and moreover every nonzero component of $\tilde{x}$ is in the η-coherence band of a unique component of x.

First, numerical evidence shows degradation in BOMPs performance with increased dynamic range consistent with the prediction of (2.5). Dynamic range of objects is clearly an essential factor determining the performance of recovery. This sensitivity to dynamic range can be drastically reduced by the local optimization technique which is introduced next. Secondly, condition (2.4) means that BOMP can resolve 3 RLs. Numerical experiments show that BOMP can resolve objects separated by close to 1 RL when the dynamic range is close to 1.

Numerical experiments show that the main shortcoming with BOMP is in its failure to perform even when the dynamic range is only moderate. To overcome this problem, we now introduce the second technique: the Local Optimization (LO). LO is a residual-reduction technique applied to the current estimate S^k of the object support. To this end, we minimize the residual $\|A\tilde{x}-b\|_2$ by varying one location at a time while all other locations are held fixed. In each step we consider a vector $\tilde{x}$ whose support differs from S^n by at most one index in the coherence band of S^n but whose amplitude is chosen to minimize the residual. The search is local in the sense that during the search in the coherence band of one nonzero component the locations of other nonzero components are fixed. The amplitudes of the improved estimate are carried out by solving the least squares problem. Because of the local nature of the LO step, the computation is not expensive.

Algorithm 2 Local Optimization (LO)

Input: $A, b, \eta > 0, S^0 = \{i_1, \ldots, i_k\}$

Iteration: For $n = 1, 2, \ldots, k$

(1) $x^n = \arg\min_z \|Az - b\|_2$,
$\quad \text{supp}(z) = (S^{n-1} \backslash \{i_n\}) \cup \{j_n\},\ j_n \in B_\eta(\{i_n\})$

(2) $S^n = \text{supp}(x^n)$

Output: S^k

Embedding LO in BOMP gives rise to the Band-excluded, Locally Optimized Orthogonal Matching Pursuit (BLOOMP).

We now give a condition under which LO does not spoil the BOMP reconstruction of Theorem 2.3.

Algorithm 3 Band-excluded, Locally Optimized Orthogonal Matching Pursuit (BLOOMP)

Input: $A, b, \eta > 0$ Initialization: $x^0 = 0, r^0 = b, S^0 = \emptyset$
Iteration: For $n = 1, \ldots, s$

(1) $i_{\max} = \arg\max_i |\langle r^{n-1}, a_i \rangle|,\ i \notin B_\eta^{(2)}(S^{n-1})$
(2) $S^n = LO(S^{n-1} \cup \{i_{\max}\})$ where LO is the output from Algorithm 2
(3) $x^n = \arg\min_z \|Az - b\|_2$ s.t. $\mathrm{supp}(z) \in S^n$
(4) $r^n = b - Ax^n$

Output: x^s

Theorem 2.4. *Let $\eta > 0$ and let x be a s-sparse vector such that (2.4) holds. Let S^0 and S^k be the input and output, respectively, of the LO algorithm. If*

$$x_{min} > (\epsilon + 2(s-1)\eta)\left(\frac{1}{1-\eta} + \sqrt{\frac{1}{(1-\eta)^2} + \frac{1}{1-\eta^2}}\right), \epsilon = \|e\| \tag{2.6}$$

and each element of S^0 is in the η-coherence band of a unique nonzero component of x, then each element of S^k remains in the η-coherence band of a unique nonzero component of x.

Corollary 2.5. Let $\tilde{x}$ be the output of BLOOMP. Under the assumptions of Theorems 2.3 and 2.4, $\mathrm{supp}(\tilde{x}) \subseteq B_\eta(\mathrm{supp}(x))$ and moreover every nonzero component of $\tilde{x}$ is in the η-coherence band of a unique nonzero component of x.

Even though we cannot improve the performance guarantee for BLOOMP, in practice the LO technique greatly enhances the success probability of recovery that BLOOMP has the best performance among all the algorithms tested with respect to noise stability and dynamic range. In particular, the LO step greatly enhances the performance of BOMP with respect to dynamic range. Moreover, whenever Corollary 2.5 holds, for all practical purposes we have the residual bound for the BLOOMP reconstruction $\tilde{x}$

$$\|b - A\tilde{x}\|_2 \leq c\|e\|_2,\ c \sim 1. \tag{2.7}$$

On the other hand, it is difficult to obtain bounds for the reconstruction error since $\|x - \tilde{x}\|_2$ is not a meaningful error metric without exact recovery of an overwhelming majority of the object support.

3. Numerics

3.1. **Method 1: l_1-analysis.** To test the accuracy for recovery using the l_1-analysis method I will start by defining the Gaussian matrix D.

I define dictionary D by the following MATLAB code. [5] The variable

$$
\begin{aligned}
&t = -\left((row/2) - 1\right) : (row/2);\\
&\text{for } k = -((column/2) - 1) : (column/2)\\
&\qquad D(:, k + (column/2)) = e^{(-(t-(0.25*k))^2/(\epsilon)^2)};\\
&\text{end}
\end{aligned}
\tag{3.1}
$$

ϵ is the decaying value of the Gaussian function. I have chosen to set this value to 3 due to the localization and specific coherence properties of the Gaussian function that I chose to use.

The next step is to create an s-spare vector x. The method I will use utilizes the Matlab command `randperm` to choose the entries of the vector which will contain the random nonzero coefficients [5]. The number of non-zero coefficients is denoted by sparsity and the length of the vector x is denoted by leng.

```
function [ vector ] = sparsevec( leng, sparsity )
index=randperm(leng);
index=index(1:round(sparsity));
vector=zeros(leng, 1);
vector(index)=randn(size(index));
```

To solve this l_1-analysis problem I have chosen to use `CVX`, which is is a Matlab-based modeling system for convex optimization. [4] [5] I was interested in testing this method for sparsity ranging from 1 to $round(row/\log(column))$ where row is the number of rows in the matrix D and $column$ is the number of columns. The Gaussian matrix I chose does not satisfy the requirements of Equation (1.20) thus I cannot claim that this upper bound holds; I can only use it as a guideline.

I tested the percent error of recovering $f = Dx$ for sparsities between 1 and $round(row/\log(column))$. For my simulation I used a row length of 64 and a column length of 256 which means that sparsity ranged from 1 to

The following code minimizes D^*f:

```
cvx_begin
    variable g(row)
    minimize(norm(D' * g, 1))
    subject to
    norm(A * g - y, 2) <= delta * norm(y, 2)
cvx_end
```

12. For each sparsity level I ran the code 10 times in order to compute an average error.

The following graphs in Figure 2 show the resulting percent errors for a range of sparsities. The results for the l_1-analysis method are disappointing. The resulting error when minimizing D^*x is higher than desired. The trend is somewhat random which would imply that the tail of $(D * f) - (D * f_s)$ is too large causing the l_1-analysis method to lack accuracy in reconstruction.

As we have noticed from the results of the Matlab test it is apparent that l_1-analysis is not the best method for recovery when using a dictionary that contains slowly decaying Gaussian matrices. [5]

Next, I would like to test the l_1-analysis method for Fourier matrices. I use the same code; however, I now define dictionary D by the Discrete Fourier Transform (DFT) matrix by the following Matlab code. [5]

(3.2)

```
for k = 1 : row
    xi = rand/20;
    D(k,:) = e^(-i*2*pi*xi*(0:column-1))/sqrt(row);
end
```

Similarly to the Gaussian case, a sparse vector x is created and `CVX` is used to solve the convex optimization problem. [4] Again, I was interested in analyzing the relative error in accurately recovering $f = Dx$ using a range of sparsities. The following graph in Figure 3 shows the resulting percent errors for a range of sparsities. The results, similar to the l_1-analysis method using Gaussian matrices, are discouraging. The resulting error when minimizing

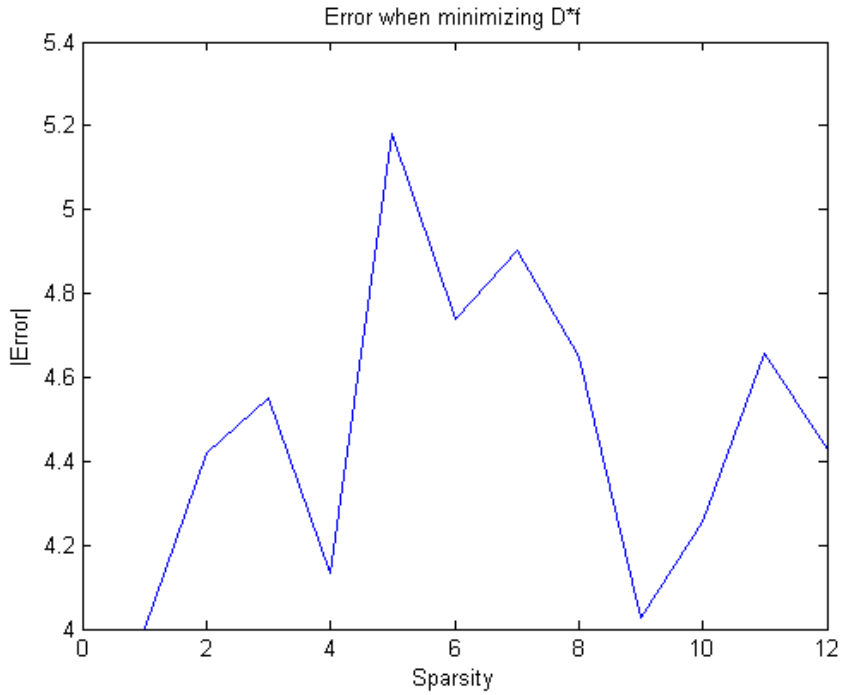

FIGURE 2

D^*x is higher than desired. There doesn't seem to be a straightforward trend with the increase in sparsity of the vector x and similar to the Gaussian case the trend is somewhat random which would imply that the tail of $(D*f)-(D*f_s)$ is too large causing the l_1-analysis method to lack accuracy in reconstruction.

As we have noticed from the results of the MATLAB test it is apparent that l_1-analysis is not the best method for recovery when using a dictionary that contains either Gaussian or Fourier matrices. [5]

3.2. **Method 2: l_1-minimization.** Similar to the l_1-analysis method I used the same Gaussian matrix D and tested the percent error of recovering $f = Dx$ for sparsities between 1 and $round(row/\log(column))$. For my simulation I used a row length of 64 and a column length of 256 which means that sparsity ranged from 1 to 12. For each sparsity level I ran the code 10 times in order to compute an average error.

The following graphs in Figure 4 show the resulting percent errors for a range of sparsities. The results for the l_1-minimization method where the

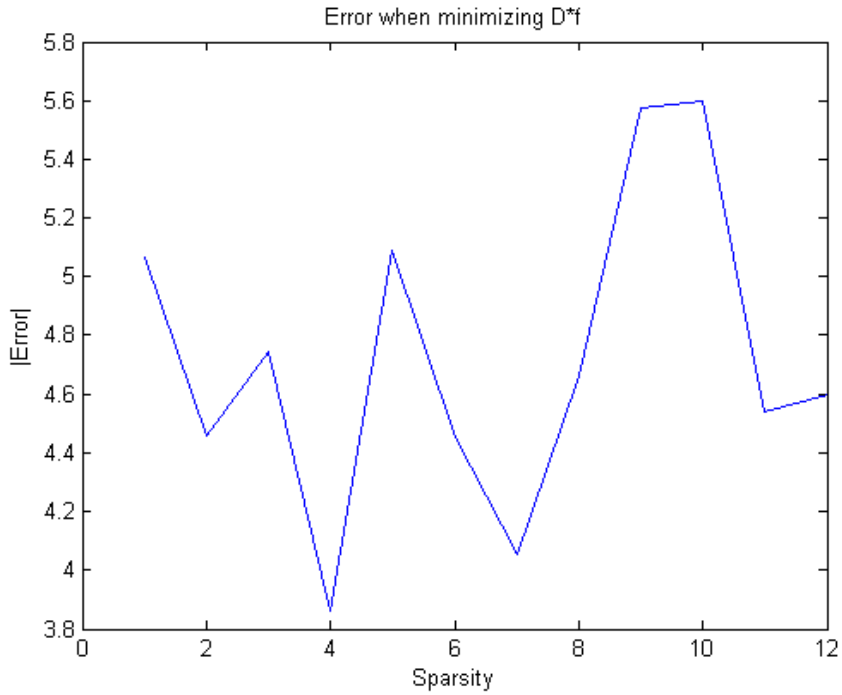

FIGURE 3

vector x is being minimized appear to show success. The errors for the range of sparsities are on an order of magnitude of 10^{-3}, which is relatively low.

Next, I tested the l_1-minimization method for Fourier matrices. I used the same Fourier matrix as I did in the l_1-analysis method and similarly tested the percent error of recovering $f = Dx$ for sparsities between 1 and $round(row/\log(column))$. The following graph in Figure 5 shows the resulting percent errors for a range of sparsities. The results, similar to the l_1-analysis using Gaussian matrices, appear to be quite accurate and are again on an order of magnitude of 10^{-8}.

In both cases, using Gaussian and Fourier matrices, the l_1-minimization method recovers the signal with great accuracy.

3.3. **Method 3: Band Exclusion.** To test the accuracy for recovery using the Band Exclusion method I considered multiple cases. I started off by using a dictionary D containing Gaussian matrices. I first considered the noiseless case and then I added noise using three different amounts: 5%, 10%, and 20% Gaussian noise.

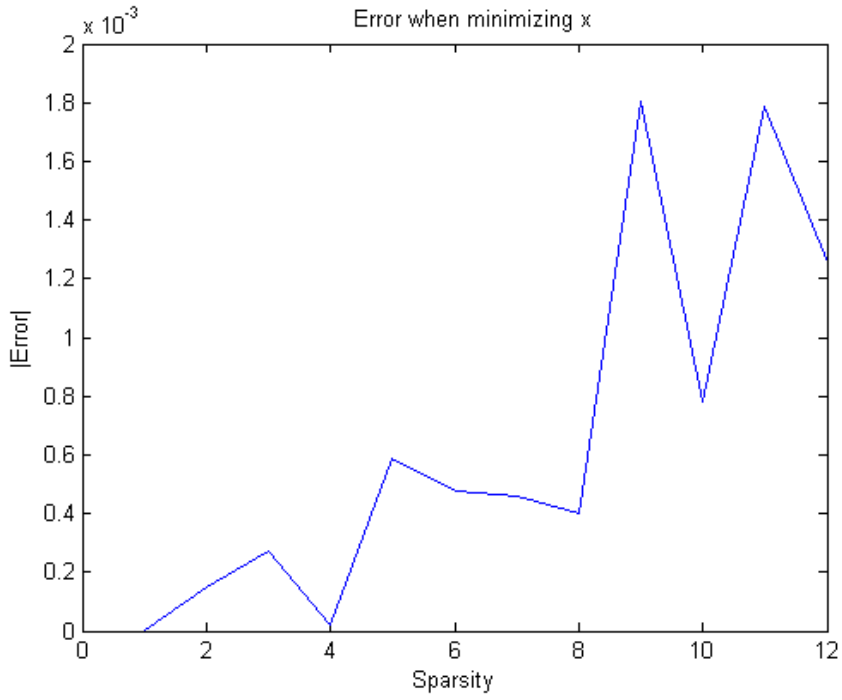

FIGURE 4

To create the code I used a refinement factor of 10. The larger the refinement factor is, the smaller the gridding error is, but more computations are involved. A value of 10 seemed to be a good balance between accuracy and computational complexity. Next I chose a 64×480 sensing matrix, which is significantly underdetermined. I chose the sensing vector to have sparsity equal to 8 due to the decaying properties of the function. Then I created the sensing matrix D with $N = 64, M = 480$, and $\epsilon = 6$ (the decaying variable of the Gaussian function) via the following MATLAB algorithm. [5]

(3.3)

$$
\begin{aligned}
&t = -((N) - 2) : 2 : (N) + 1; \\
&\text{for } k = -((M/2) - 1) : (M/2) \\
&\qquad D(:, k + (M/2)) = e^{-(t-(0.25*k)).^2/(\epsilon)^2}; \\
&\text{end}
\end{aligned}
$$

After defining the sensing matrix D, I needed to determine the band. The band is based off of the choice of the number of columns of the sensing

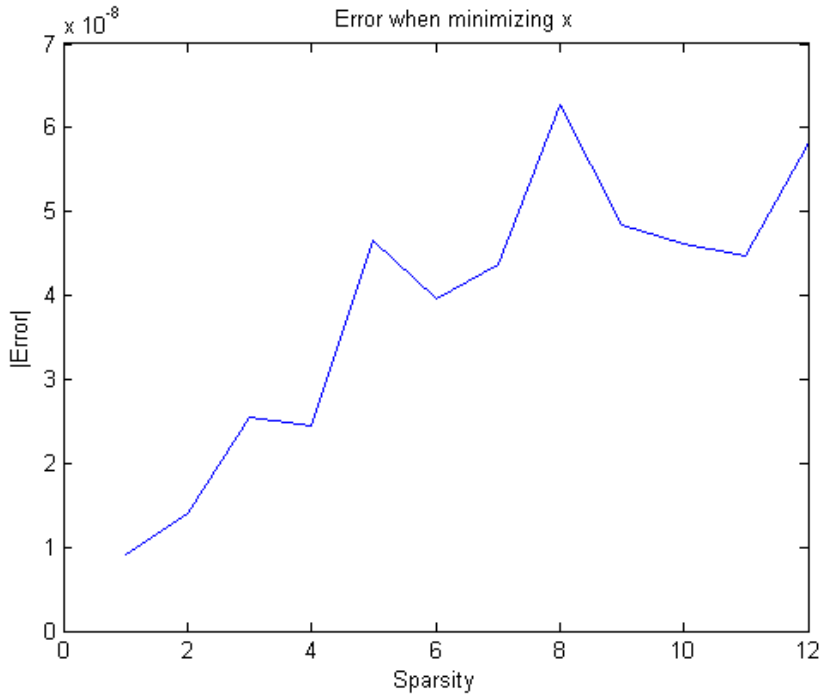

FIGURE 5

matrix as well as the shift of the Gaussian function. I wrote a function called `findband` which determines the length of the band by first comparing a Gaussian curve to the threshold $3.5/\sqrt{row\ length}$ and then determining the index where the Gaussian curve is below the threshold. The code is as follows:

---

function [radius] = findband(M, N, CoMatrix)
curve=abs(CoMatrix(M/2,:)/max(CoMatrix(M/2,:)));
diff=curve-(3.5/$\sqrt{N}$);
p=zeros(1,M);
t=p<diff;
ind=find(t,1,'last');
maxvalueind=find(curve==max(abs(CoMatrix(M/2,:)/max(CoMatrix(M/2,:)))));
radius=abs(maxvalueind-round(ind+(M/100)));

Figure 6 shows the coherence pattern of the 64×480 matrix with $F = 10$. The bright diagonal band represents a heightened correlation, or pairwise coherence, between a column vector and its neighbors on both sides.

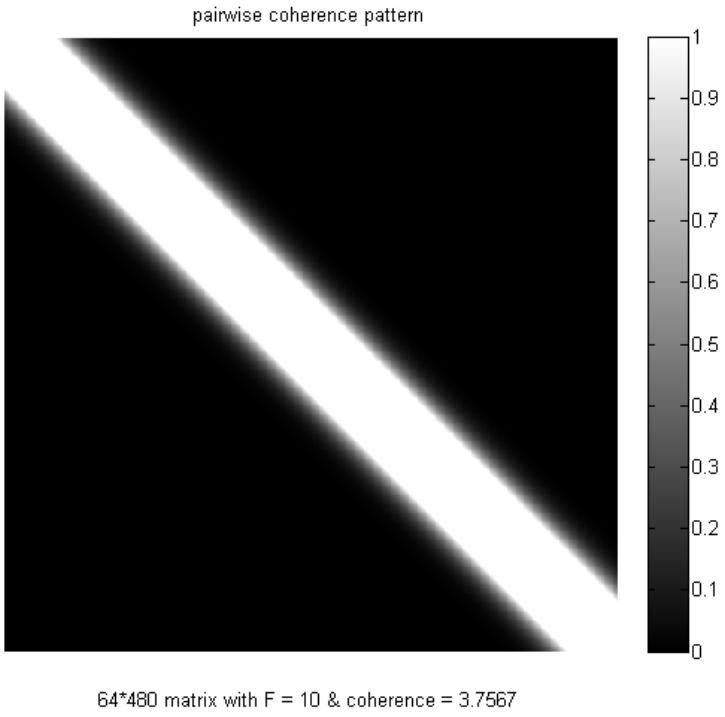

FIGURE 6

This information is then used to determine the relative error when using the OMP method, which is 0.42503 and can be seen in Figure 7.

Next, the the technique of band exclusion and local optimization is used in the BLOOMP algorithm. As you can see in Figure 8, the relative error is 0.0049354 which is significantly lower. When running the code 100 times, the average relative error when using the BLOOMP algorithm was 0.00334.

After testing the noiseless case I tested the algorithms with 5%, 10%, and 20% Gaussian noise. After conducting many experiments I found that the BLOOMP algorithm has a lower relative error in every case. Because of this I will only report the numerics when using the BLOOMP algorithm. I ran each algorithm 100 times and computed the average relative error. With

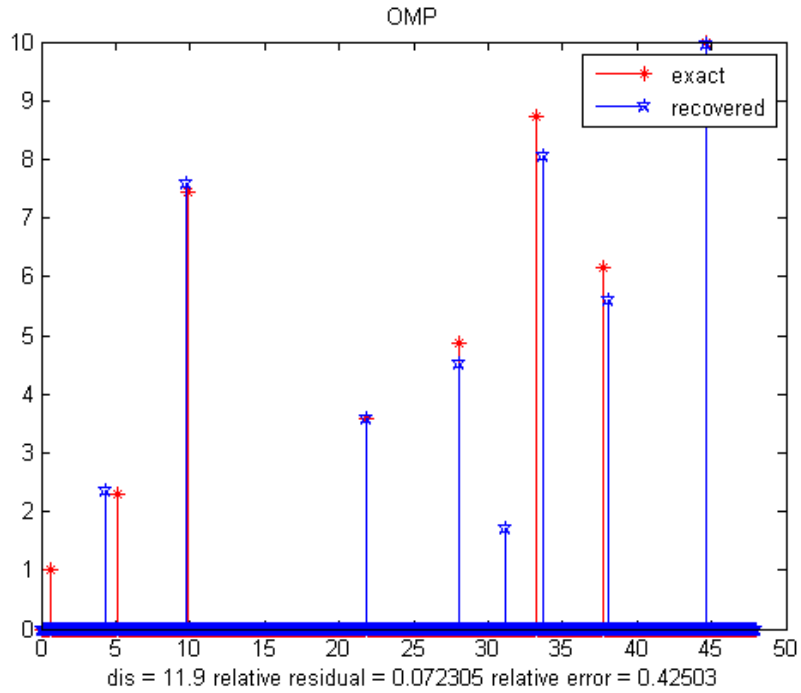

FIGURE 7

5% Gaussian noise the BLOOMP algorithm has a relative error of 0.218, with 10% Gaussian noise there is a relative error of 0.0474, and with 20% Gaussian noise there is a relative error of 0.181.

The threshold $3.5/\sqrt{row\ length}$ is larger than Equation (1.18) dictates and thus the band used when executing the algorithm is smaller than necessary. Despite the small band, the algorithm relatively accurately recovers the signal. Therefore the algorithm will work for a larger band.

Additionally, I tested the band exclusion method (both OMP and BLOOMP) for a dictionary made of a Fourier function. I used the same DFT as I did for the previous two methods. Similar to the Gaussian results, the BLOOMP algorithm had a lower relative error in every case so I will only report the results using the BLOOMP algorithm. Again, I was interested in testing the algorithms with no noise, 5%, 10%, and 20% Gaussian noise. I ran each algorithm 100 times and computed the average relative error.

Figure 9 shows the coherence pattern of the 64 x 480 matrix with F=10. The bright diagonal bands represent a heightened correlation, or pairwise coherence, between a column vector and its neighbors on both sides.

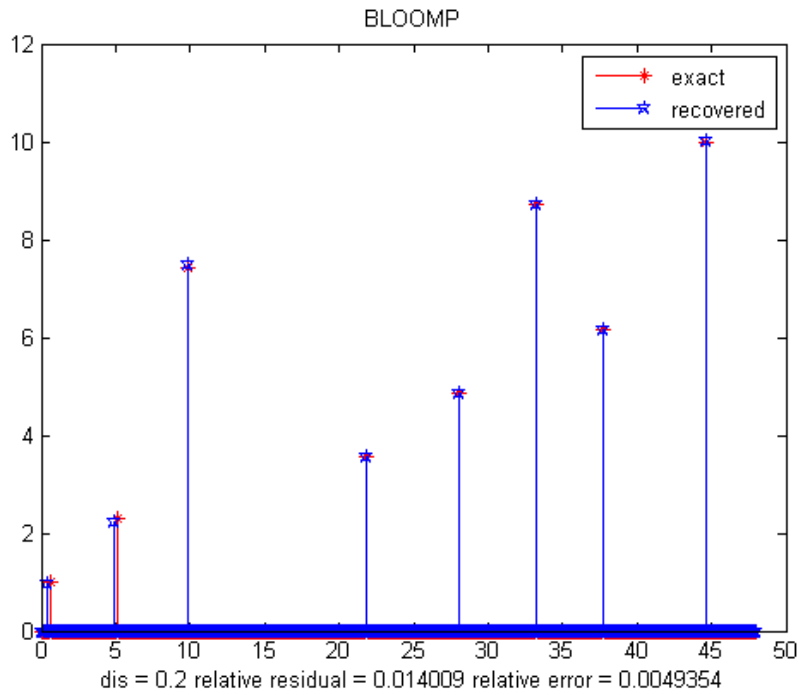

FIGURE 8

Figure 10 shows the relative error with no noise using the BLOOMP algorithm is 4.2715×10^{-16} which is significantly lower than the error using any of the other methods.

However, with 5% Gaussian noise the BLOOMP algorithm has a relative error of 0.0271 which is now larger than when using l_1-minimization. With 10% Gaussian noise there is a relative error of 0.0685 and with 20% Gaussian noise there is a relative error of 0.2623, both of these being higher than the relative errors determined using l_1-minimization.

As we can see from the above results, the method of band exclusion works reasonably well in reconstructing the signal. However; in all cases, with the exception of the noiseless Fourier case, l_1-minimization is a more accurate method for signal recovery.

3.4. **Imposed Minimum Distance Condition.** After testing the multiple cases above and seeing the success of the l_1-minimization method, I was lastly interested in determining how imposing a minimum distance between the nonzero entries in the sparse vector x would affect the accuracy

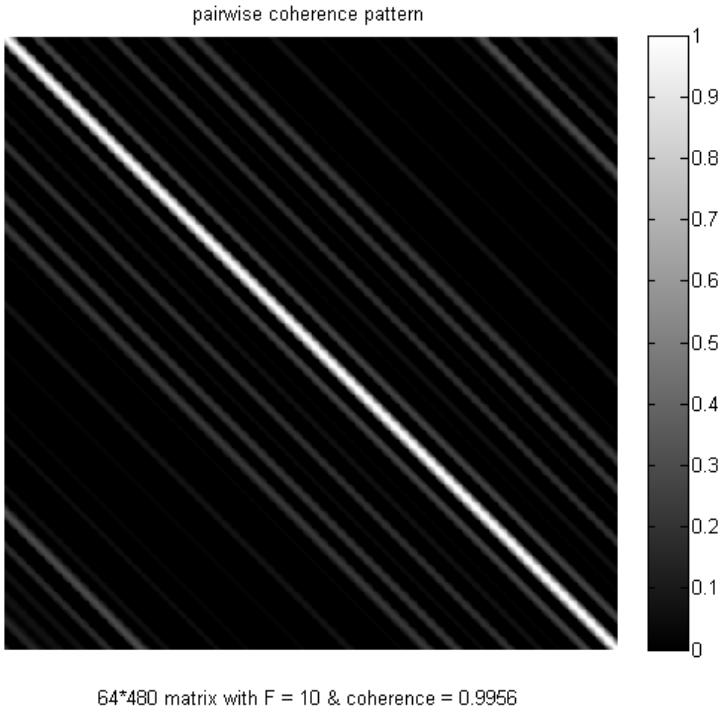

FIGURE 9

of recovery. I slowly increased the distance from 2 to 10 and noticed that the relative error for the l_1-minimization method decreased. For example, in Figure 11 I imposed a minimum distance of 2 between the nonzero entries of the vector x. The relative error when testing a range of sparsities did not exceed 7.4×10^{-8}. In Figure 12 I imposed a minimum distance of 10 where the relative error did not surpass 4.4×10^{-8}. It appears that increasing the distance between nonzero entries in the sparse vector causes the l_1-minimization method to more accurately recover the signal. As of now there is no theory in the relating literature as to why this is the case for coherent dictionaries. The next step in this area of compressive sensing would be to determine why l_1-minimization is most successful in accurate reconstruction and why the distance between nonzero entries in the sparse vector affects this recovery.

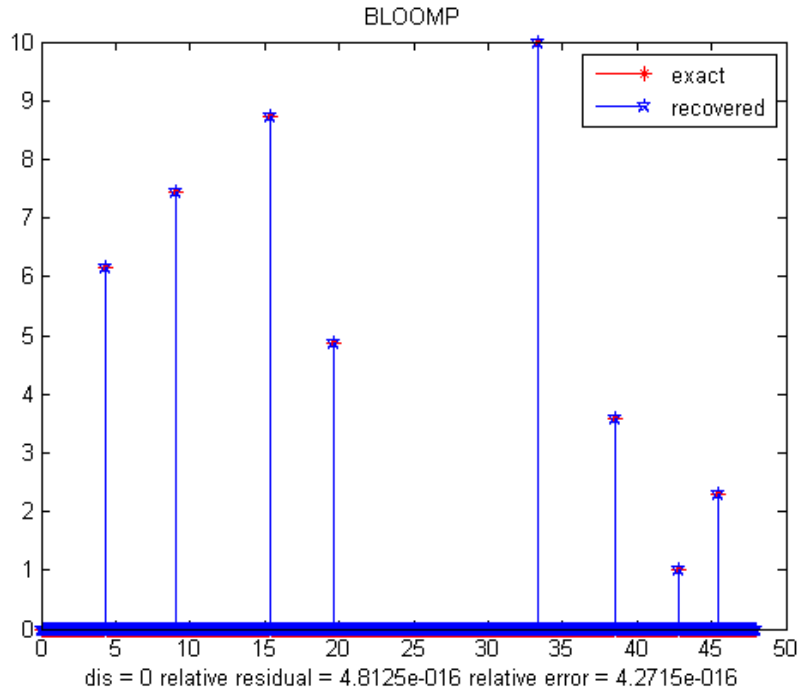

FIGURE 10

4. Conclusion

Compressive sensing is a developing field which asserts that one can recover certain signals and images from far fewer samples or measurements than traditional methods. When signals are only sparse in a truly redundant dictionary, the sensing matrix is highly coherent and methods such as l_1-analysis and other such nontraditional methods are used. This paper compares l_1-analysis, l_1-minimization, and band exclusion to see which method better solves the coherence problem. After computing the error in reconstructing signals with both Gaussian and Fourier dictionaries using the three methods I found that l_1-minimization is a superior method. It recovered the signal with the lowest relative error in every case with the exception of the noiseless Fourier case. Therefore, when trying to recover a signal in which a highly coherent Gaussian or Fourier dictionary is needed, the l_1-minimization method, as opposed to l_1-analysis and band exclusion,

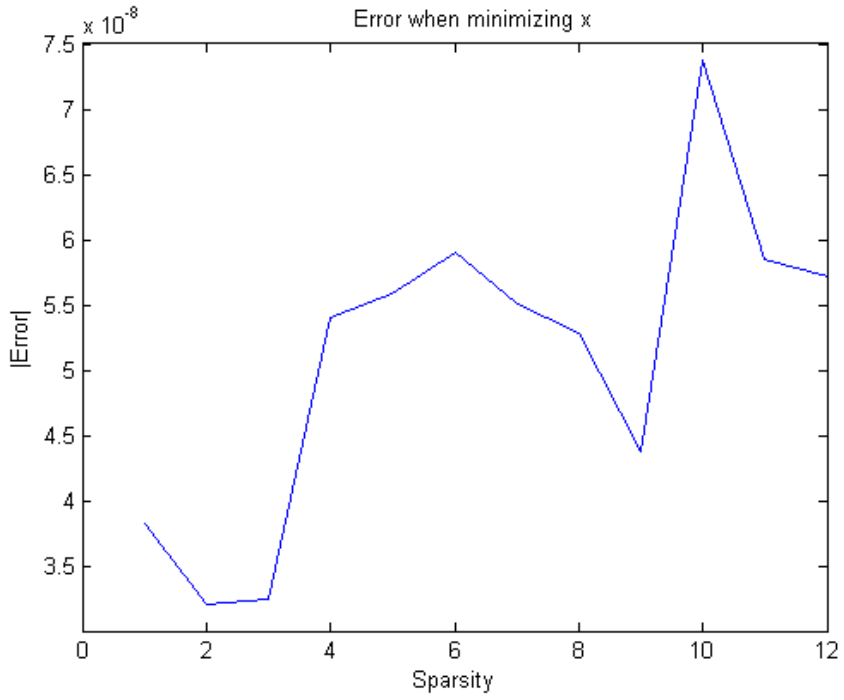

FIGURE 11

has proven to most accurately recover the signal. Thus, under certain specified conditions, I have established a link between the theory and practice of compressive sensing.

5. Acknowledgments

I would like to thank Professor Thomas Strohmer for guiding me through my undergraduate senior honors thesis during my final year at UC Davis. I am grateful for the encouragement and support I was given throughout my project and especially during the more challenging times. I would also like to thank Professor Steve Klee for helping me with technical aspects and editing. Last but not least, thanks to the Mathematics Department at UC Davis for this magnificent opportunity.

References

[1] Emmanuel J. Candes, Yonina C. Eldar, Deanna Needell, and Paige Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59 – 73, 2011.

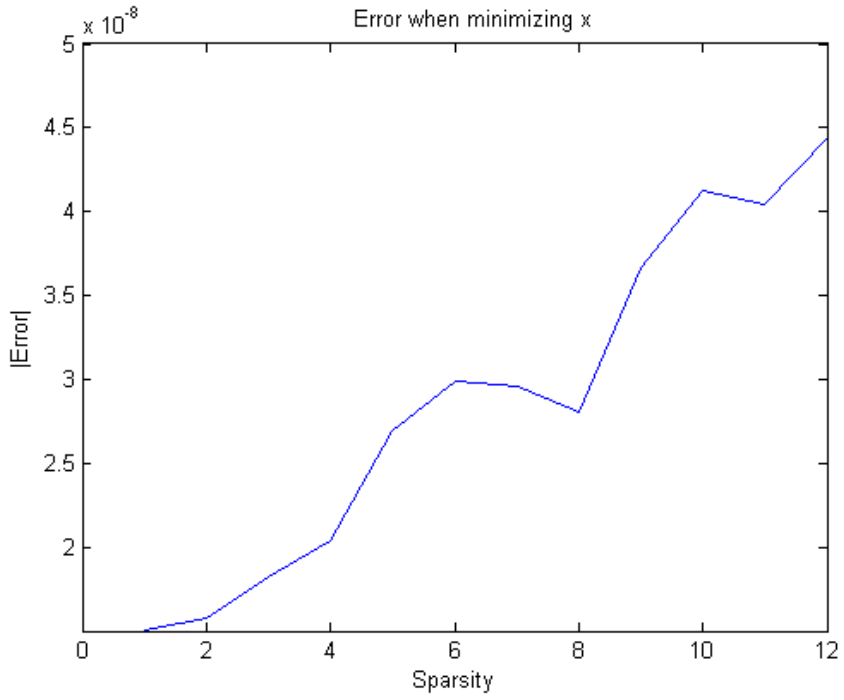

FIGURE 12

[2] Albert Fannjiang and Wenjing Liao. Coherence-pattern guided compressive sensing with unresolved grids. *CoRR*, abs/1106.5177, 2011.

[3] Massimo Fornasier and Holger Rauhut. Compressive sensing. 2010.

[4] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. `../../cvx`, April 2011.

[5] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

MATHEMATICAL SCIENCES BUILDING, ONE SHIELDS AVE., UNIVERSITY OF CALIFORNIA, DAVIS, CA 95616

E-mail address: `ucdcohen@ucdavis.edu`