Graph Spectral Compressed Sensing

Xiaofan Zhu and Michael Rabbat

Dept. Electrical and Computer Engineering

McGill University, Montréal, Québec

Email: xiaofan.zhu@mail.mcgill.ca

Technical Report TR-ECE

September 28, 2011

Abstract

This report focuses on the theoretical analysis of Compressed Sensing(CS) with the partial Graph Fourier ensemble sensing matrix. In this report, we prove that with the help of additional knowledge of a linear compressible signal, a stable recovery can still be guaranteed even if the entries of the orthogonal sensing matrix is not uniformly bounded. More specifically, we construct a sensing matrix by randomly select a certain number of the rows from the graph Laplacian eigenbasis matrix. It is shown in this report that if the signal is smooth on the graph which satisfies certain conditions, we can still utilize ℓ_1 decoding to reconstruct the original signal.

1 Introduction

Recently, due to the ground-breaking work by Donoho and Candès [1, 2, 3], Compressed Sensing (CS) has soon become a promising research area, which shows that we can find sparse solutions to an under-determined linear systems by taking random measurements. Candès[2, 4] and Rudelson[5] prove that if we construct the sensing matrix by randomly selecting a small portion of the rows from a certain uniformly bounded orthogonal matrix, the original signal can be perfectly reconstructed if it is sparse or compressible. In the technical report [6], we mention that the graph Laplacian eigenbasis can be regarded as the "Fourier" transform for signals supported on graph vertices. We call such basis as the Graph Fourier Transform (GFT) Thus, a natural question will be: can we construct a sensing matrix by randomly selecting a small portion of the rows from the Graph Fourier Transform basis?

In order to solve this problem, we need to first delve into the conventional CS literature: Candès [2] and Rudelson [5] discuss conditions that the structure of random matrices should satisfy to be valid CS sensing matrices:

1. The matrix should be orthogonal.
2. The entries of the normalized $N \times N$ matrix should be uniformly bounded by $O(\frac{1}{\sqrt{N}})$, i.e., the coherence of the sensing matrix $\mu = O(\frac{1}{\sqrt{N}})$, where $\mu = \max_{i,j} |\Phi_{i,j}|$.

The traditional Fourier basis is clearly a candidate fit for such criteria. If F is the Discrete Fourier Transform (DFT) basis and if Ω is a random subset of $\{1, 2, \cdots, N\}$ with dimension $|\Omega| = O(\text{poly}(K, \log N))$ and where K is the sparsity of x in the basis F, then we can reconstruct x by solving

$$\min_x \|x\|_1 \ s.t. \ y = F_\Omega x.$$

where F_Ω is a submatrix of F obtained by selecting the rows corresponding to Ω; i.e., F_Ω is the so-called "partial Fourier ensemble". Analogously, if U is the GFT basis, then we call U_Ω the partial Graph Fourier ensemble. In our scenario, it is straightforward to see that the GFT basis satisfies the first condition while the second one cannot be guaranteed. For example, circulant graphs will generate eigenbasis with uniformly bounded entries while more general graphs like KNN graphs or $\epsilon-$graphs fail with largest entries close to 1.

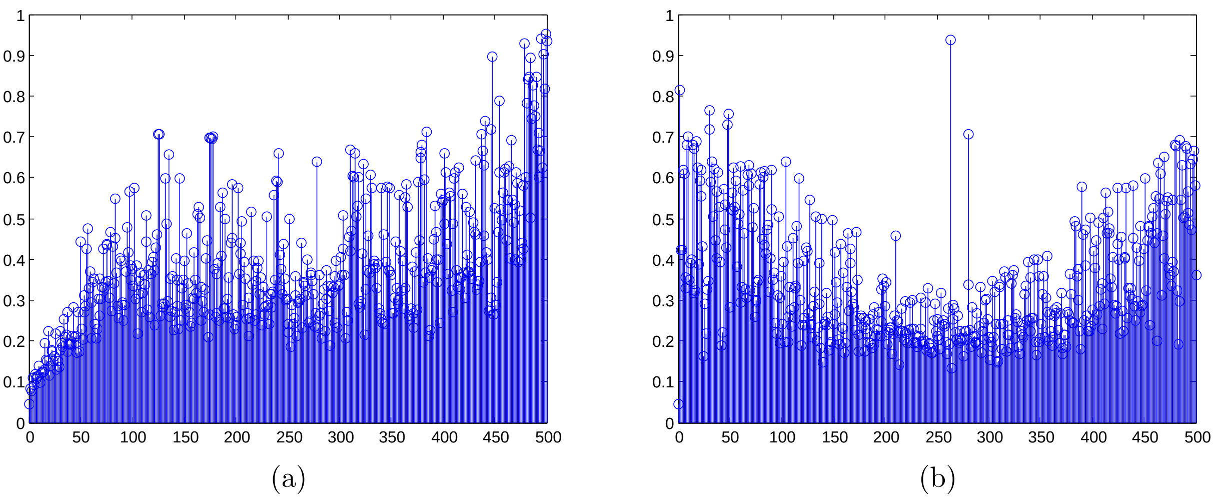

Figure 1: This figure plots the entry with largest magnitude of the entries in each eigenvector. The graph Fourier basis is generated by extracting the eigenbasis of a KNN graph. We denote k as the number of neighbors for a KNN graph. (a) k=5. (b) complete graph

Fig. 1(a) plots the largest components of each column for a graph Fourier basis. X-axis represents the corresponding eigenvectors and y-axis stands for the largest magnitude of the eigenvector entry. We formulate the graph Fourier basis by obtaining the Laplacian eigenvector of a KNN graph with $k = 5$. Clearly, the conventional coherence is close to 1 since this matrix is not uniformly bounded. In order to delve into more details about how the entries of the GFT basis are distributed, we generalize the definition of coherence as follows:

Definition 1. Define $\mu_\Phi(T) = \max_{i,j} |[\Phi_T]_{i,j}|$ to be the coherence of the matrix Φ_T, where T is a subset of $\{1, 2, \cdots, N\}$ and Φ_T is the submatrix obtained by selecting the columns of Φ corresponding to T. If $T = \{1, 2, \cdots, N\}$, then $\mu_\Phi(T)$ is equivalent to μ. In some part of this report, we abbreviate $\mu_\Phi(T)$ as $\mu(T)$.

However, it is worth noting that $\mu_U(T)$ is bounded when U_T corresponds to the eigenvectors whose associated eigenvalues are small, even if the coherence of the whole matrix is not bounded by $O(\frac{1}{\sqrt{N}})$. A natural question is: does such a trend exist for all the graphs? Or say, is $\mu_U(T)$ bounded by $O(\frac{1}{\sqrt{N}})$ for all kinds of graphs?

The answer to this question is negative. From the technique we develop in the technical report [6], we regard $\|f\|_G = f^T L f$ as the smoothness of signal f on a graph. If we also consider the eigenvector u_i as a graph signal and replace f with u_i, it's easy to check that $\|u_i\|_G = \lambda_i$, where λ_i is the ith eigenvalue corresponds to u_i, i.e. we can say that λ_i describes the smoothness of eigenvector u_i. If λ_i is small, then u_i will be smooth and its entries are likely to be bounded by small constant. Since $\lambda_0 = 0$, u_0 is perfectly smooth, or say, is the DC component of the graph Fourier basis. From Fig. 1, we can also see that the the largest entry of each eigenvector keeps increasing until i is around 50 and λ_i is large enough and can no longer affect the distribution of the eigenvector components. Based on the above discussion, we know that the graphs with adequately small eigenvalues in low frequencies can provide bounded $\mu_U(T)$. A counter example is the complete graph shown by Fig. 1(b) since the eigenvalues are equally very large. Thus, the coherence is still unbounded when $i < 50$. Note that this is a sufficient condition but not necessary.

In this report, we show that although entries of the graph Fourier aren't uniformly bounded, a stable recovery by solving ℓ_1 minimization problem is still possible for linear compressible graph signals.

2 Notation and Definition

Many of the natural signals are not strictly sparse but follow a power law decay. In our case, we consider the signals whose coefficients linearly decay, i.e. $x_i \leq G i^{-1/r}$. From the last report, we know that as long as the distribution of Laplacian eigenvalues satisfy certain properties, the signals on graphs can be well modeled as linear compressible signals. As discussed in [7], the compressible signals can be defined by its decaying behavior of non-linear approximation error. Since we are talking about linearly compressible signals, we can modulate the definition a little to fit our case:

Definition 2. The set of s-linear-compressible signals is defined as

$$\mathbb{L}_s = \{x \in \mathbb{R}^N : \epsilon_l(K, x) \leq S K^{-s}, 1 \leq K \leq N, S < \infty\},$$

where $\epsilon_l(K, x)$ is the K-term Linear Approximation Error.

Definition 3. The j^{th} set of the linear residual subspaces of size K is defined as: $L_{j,K} = \{u \in \mathbb{R}^N$ with support $T(L_{j,K}) = \{(j-1)K+1, (j-1)K+2, \cdots, jK-1\}$, i.e. $u_T \neq 0$ and $u_{T^c} = 0$. The last set may have less than K non-zero entries. We can abbreviate $T(L_{j,K})$ as T_j.

According to the definition, we can split a linear compressible signal into $\lceil N/K \rceil$ sets of the linear residual subspaces. For a linear compressible signal, the magnitude of coefficients in T_j decays when j becomes larger. Moreover, if T and Ω are subsets of $\{1, 2, \cdots, N\}$, we denote Φ_T as the submatrix of selecting the corresponding columns from matrix Φ and Φ_Ω as the submatrix of picking the corresponding rows and we call Φ_Ω partial graph Fourier ensemble.

Definition 4. An $m \times N$ matrix Φ has the $L_{j,K}$-restricted isometry property($L_{j,K}-$RIP) with constant δ_K if for all $x \in L_{j,K}$, we have

$$(1-\delta_K)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta_K)\|x\|_2^2$$

Accordingly, $L_{1,K}-$RIP deals with the signals whose coefficients are all zeros except the first K entries. Since most of the energy of a smooth graph signal lies in the lower frequencies, we can exploit $L_{1,K}-$RIP to guarantee a stable signal recovery. $L_{1,K}-$RIP is a special case of model based RIP since we know exactly how will the energy of signal decay and where the residual space locates. Likewise, we also need a certain tool to deal with how will the small non-zero entries outside $L_{1,K}$ behaves. Due to the fact that the coefficients outside $L_{1,K}$ have very small magnitudes, we can relax the conventional RIP to control their non-isometry property. The following definition of RAmP is a remedy counterpart of RIP in this case.

Definition 5. A matrix Φ has the $(\epsilon_K, r)-$ restricted amplification property(RAmP) for the linear residual subspaces $L_{j,K}$ if

$$\|\Phi u\|_2^2 \leq (1+\epsilon_K) j^{2r} \|u\|_2^2$$

for any $u \in L_{j,K}$ for each $1 \leq j \leq \lceil N/K \rceil$.

When $j = 1$, then it is just the upper bound in $L_{1,K}-$RIP. When j is larger, we utilize the parameter r to offset the impact of the decaying behavior of signal coefficients. The following lemma shows what conditions the partial graph Fourier sensing matrix should obey in order to satisfy $(\epsilon_K, r)-$RAmP.

3 Main Result

Theorem 3.1. *Let $x \in \mathbb{L}_s$ be an s-linear compressible signal. Let*

$$T_j = \{(j-1)K+1, (j-1)K+2, \cdots, jK-1\}.$$

If $\mu(T_j) \leq C \cdot j^{s-1}$ for all $j = 1, \cdots, \lceil N/K \rceil$ and some $C > 0$, and if the number of measurements M obeys $M \geq Const \cdot K \cdot \ln(\frac{K}{\delta})$ for some $\delta > 0$, then with probability $1-\delta$, the solution x^* obtained from ℓ_1 decoding satisfies

$$\|x - x^*\|_2 \leq C_0 \cdot \ln \left\lceil \frac{N}{K} \right\rceil SK^{-s} + 2\frac{\|x - x_K\|_1}{\sqrt{K}}. \quad (1)$$

where C_0 is a small constant.

The theorem claims that if the entries of the original signal decay quickly, we can still guarantee a stable recovery when the coherence $\mu(T_j)$ keeps increasing for larger j. Actually, we allow $\mu(T_j)$ to become unbounded if the entries of the original signals supported on T_j are small. The above theorem explains that why the partial Graph Fourier ensemble works as a sensing matrix for smooth signals supported on graphs. This is because smooth signals supported on graphs are linear compressible [6], i.e., most of the large Graph Fourier coefficients locate in the "low frequency component" while those components have relatively low coherence. The full proof of this result is given in the next section. The methodology of the proof is mainly based on [4], [7] and [8].

4 Proof

The following lemma is tailored from[4] to fit our need.

Lemma 4.1. *Let Φ be an $n \times n$ orthogonal matrix obeying $\Phi^T\Phi = nI$. Consider a fix set T and let Ω be a random set sampled using the Bernoulli model and $\mu_\Phi(T) = \underset{j \in T}{max}|\Phi_{i,j}|$. Denote $Y = \hat{\Phi}_{\Omega T}^T \hat{\Phi}_{\Omega T} - I$ where I is the identity matrix and $\hat{\Phi}_{\Omega T} = \frac{1}{\sqrt{m}}\Phi_{\Omega T}$ is the normalized matrix and $m = |\Omega|$. Then*

$$\mathbb{E}\|Y\| \leq C_R \cdot \mu(T)\frac{\sqrt{|T|log|T|}}{\sqrt{m}} \tag{2}$$

and

$$P(|\|Y\| - \mathbb{E}\|Y\|| > t) \leq 3exp(-\frac{t}{KB}log(1 + \frac{t}{1 + \mathbb{E}\|Y\|})), \tag{3}$$

where $B \leq \mu_\Phi^2(T)|T|/m$.

From this lemma, Candès further proves for $x \in \mathbb{R}^N$ be a sequence supported on a fixed set T, $\frac{m}{2}\|x\|_2^2 \leq \|\Phi_\Omega x\|_2^2 \leq \frac{3m}{2}\|x\|_2^2$. If we let $T = L_{j,K}$, then it is exactly the $L_{j,K}-$RIP. It is worth noting that there is one minor difference between the lemma here and the original work in [4]. Since in [4], the author discuss the case where T is fixed but arbitrary, hence they define the coherence $\mu = \max|\Phi_{i,j}|$. However, in the scenario we are interested in, we merely concern about $L_{1,K}-$RIP. Accordingly, the set T is not arbitrary but $T(L_{1,K})$. Correspondingly, we can replace μ with $\mu_\Phi(T)$ since only Φ_T is involved. This lemma can also be exploited as a useful tool for verifying the $(\epsilon_K, r)-$RAmP.

Theorem 4.2. *Let Φ be an $n \times n$ orthogonal matrix. Fix a subset T of the signal domain. Choose a subset Ω of the measurement domain of size $|\Omega| = m$. Suppose that the number of measurements m obeys:*

$$m \geq C_3|T|ln(\frac{|T|}{\delta})(\frac{\mu(T)}{j^r})^2 \tag{4}$$

for some positive constant C_3 and with proper choice of ϵ_K and r. Then, with probability $1 - \delta$, the *matrix* Φ_Ω *satisfies*

$$\|\Phi_\Omega u\|_2^2 \leq (1+\epsilon_K) j^{2r} \|u\|_2^2 \tag{5}$$

Proof.

$$P(\|\Phi u\|_2^2 > (1+\epsilon_K) j^{2r} \|u\|_2^2) \leq P(|\|\Phi u\|_2^2 - \|u\|_2^2| > [(1+\epsilon_K) j^{2r} - 1]\|u\|_2^2) \tag{6}$$

$$= P(\|\Phi_{\Omega T}^T \Phi_{\Omega T} - I\| > (1+\epsilon_K) j^{2r} - 1) \tag{7}$$

Denote $Y = \Phi_{\Omega T}^T \Phi_{\Omega T} - I$. Thus, the problem is now equivalent to bound $P(\|Y\| > (1+\epsilon_K) j^{2r} - 1)$. From Lemma 4.1, set $t = \frac{(1+\epsilon_K) j^{2r} - 1}{2}$ and pick m large enough to make $\mathbb{E}\|Y\| \leq t$, then (7) is bounded by the righthand of (3). Accordingly, we can obtain

$$m \geq \frac{4C_R^2 \mu^2(T)|T| log|T|}{[(1+\epsilon_K) j^{2r} - 1]^2} \tag{8}$$

which can be simplified as:

$$m \geq C_1 |T| log|T| (\frac{\mu(T)}{j^{2r}})^2 \tag{9}$$

where $C_1 = 4C_R^2/(1+\epsilon_K)^2$. Since $B \leq \mu^2(T)|T|/m$, (3) gives

$$P(\|Y\| > 2t) \leq 3exp(-\frac{mt \cdot log(1+\frac{t}{1+t})}{\mu^2(T)K}) \tag{10}$$

and let it be bounded by δ which provides the following:

$$m \geq ln(\frac{3}{\delta}) \frac{K\mu^2(T)|T|}{t \cdot log(\frac{1+2t}{1+t})} \tag{11}$$

where $t = \frac{(1+\epsilon_K)j^{2r}-1}{2}$. Since it is easy to pick proper ϵ_K and r to bound $log(\frac{1+2t}{1+t})$ away from zero, i.e. there exist a constant C such that $0 < C \leq log(\frac{1+2t}{1+t}) < log2$, we can further simplify (11) as:

$$m \geq C_2|T|ln(\frac{3}{\delta})(\frac{\mu(T)}{j^r})^2 \tag{12}$$

where $C_2 = \frac{K}{(1+\epsilon_K)log(\frac{1+2t}{1+t})}$. Combining (9) with (12) gives the conclusion. □

Corollary 4.3. *Let Φ be the same setting described in Theorem. 9. If $\mu(T_j) \leq C \cdot j^r$ for all $j = 1, \cdots, \lceil N/K \rceil$ and the number of measurements m obeys*

$$m \geq Const \cdot K \cdot ln(\frac{K}{\delta})$$

Then with probability $1-\delta$, the measurement matrix Φ satisfies the $(\epsilon_K, r)-RAmP$ for the linear residual subspaces $L_{j,K}$.

Corollary 4.4. *Let Φ be the same setting described in Theorem. 9. If $\mu(T_1) \leq C$ and the number of measurements m obeys*

$$m \geq Const \cdot K \cdot ln(\frac{K}{\delta})$$

Then with probability $1-\delta$, the measurement matrix Φ satisfies the $L_{1,K}-RIP$.

When we consider the situation $j = 1$ in the above theorem, the proof will directly give Corollary. 4.4.

Theorem. 4.2 along with Corollary. 4.3 provide us with the implication that the coherence of matrix Φ doesn't have to be uniformly bounded as discussed in the conventional compressed sensing literatures[5, 4, 2]. More specifically, if the coefficients of a signal that are supported on a certain residual space $L_{j,k}$ is quite small, we allow the corresponding coherence $\mu(T_j)$ larger. The next theorem shows that we can still guarantee a stable recovery for linear compressible signal without uniformly bounded coherence of sensing matrix. Since linear compressible signal is just one special case of model based compressible signa, the proof of this theorem is exactly the the same as that in [7].

Theorem 4.5. *Let $x \in \mathbb{L}_s$ be an s-linear compressible signal. If Φ has the $(\epsilon_K, r)-RAmP$ for the linear residual subspaces $L_{j,K}$ and $r = s - 1$, then we have*

$$\|\Phi(x - x_K)\|_2 \leq C_s\sqrt{1 + \epsilon_K}SK^{-s}ln\lceil\frac{N}{K}\rceil. \tag{13}$$

where $C_s = 2^s + 1$.

Theorem 4.6. *Let $x \in \mathbb{L}_s$ be an s-linear compressible signal. If Φ has the $L_{1,2K}-RIP$ and the $(\epsilon_K, r)-RAmP$ for the linear residual subspaces $L_{j,K}$ and $r = s - 1$, then the solution x^* obtained from L1 decoding satisfies*

$$\|x - x^*\|_2 \leq C_4SK^{-s} + 2\frac{\|x - x_K\|_1}{\sqrt{K}},$$

where $C_4 = C_sln\lceil\frac{N}{K}\rceil(\sqrt{\frac{1+\epsilon_K}{1-\delta_{2K}}} + \sqrt{\frac{1+\epsilon_K}{1-\delta_K}})$.

The technique used here is identical to [8]. We provide the proof here for the completeness of this report. It is worth noting that [7] just provide us with a performance bound for model based recovery algorithm while this report gives a detailed analysis of L1 decoding.

Proof. Let h denote the error $x^* - x$, thus it is straightforward to see that h lies in the null space of Φ. For the first step, we note that for $j \geq 2$,

$$\|h_{(T_1 \bigcup T_2)^C}\|_2 \leq \sum_{j\geq 3}\|h_{T_j}\|_2 \leq K^{-1/2}(\|h_{T_2}\|_1 + \|h_{T_3}\|_1 + \cdots) = K^{-1/2}\|h_{T_1^C}\|_1 \tag{14}$$

The second inequality is due to the fact that $\|h_{T_j}\|_2 \leq K^{1/2}\|h_{T_j}\|_\infty \leq K^{-1/2}\|h_{T_{j-1}}\|_1$. Since $\|x^*\|_1$ is minimized,

$$\|x\|_1 \geq \|x+h\|_1 = \|x_{T_1} + h_{T_1}\|_1 + \|x_{T_1^C} + h_{T_1^C}\|_1 \tag{15}$$

$$\geq \|x_{T_1}\|_1 - \|h_{T_1}\|_1 - \|x_{T_1^C}\|_1 + \|h_{T_1^C}\|_1 \tag{16}$$

Thus, $\|h_{T_1^c}\|_1 \leq 2\|x_{T_1^c}\|_1 + \|h_{T_1}\|_1$. Combined with (14) and the fact that $\|h_{T_1}\|_1 \leq K^{1/2}\|T_1\|_2$, we conclude that

$$\|h_{(T_1 \bigcup T_2)^c}\|_2 \leq K^{-1/2}\|h_{T_1^c}\|_1 \leq 2K^{-1/2}\|x_{T_1^c}\|_1 + \|h_{T_1}\|_2. \tag{17}$$

The second step bounds $\|h_{T_1 \bigcup T_2}\|$. Since Φ has the RIP for $T_1 \bigcup T_2$ and h lies in the null space of Φ

$$(1-\delta_{2K})\|h_{T_1 \bigcup T_2}\|_2^2 \leq \|\Phi h_{T_1 \bigcup T_2}\|_2^2 = \|\Phi h_{(T_1 \bigcup T_2)^C}\|_2^2 \tag{18}$$

Following the procedure of the proof of Theorem(13) in [7], we can conclude that the upper bound of $\|\Phi h_{T_1^c}\|_2^2$ also upper bounds $\|\Phi h_{(T_1 \bigcup T_2)^c}\|_2^2$. Accordingly,

$$\|h_{T_1 \bigcup T_2}\|_2 \leq C_s \sqrt{\frac{1+\epsilon_K}{1-\delta_{2K}}} S K^{-s} ln\lceil \frac{N}{K} \rceil \tag{19}$$

Similarly,

$$\|h_{T_1}\|_2 \leq C_s \sqrt{\frac{1+\epsilon_K}{1-\delta_K}} S K^{-s} ln\lceil \frac{N}{K} \rceil \tag{20}$$

And finally, combining (19)(20)(17)

$$\|h\|_2 \leq \|h_{T_1 \bigcup T_2}\|_2 + \|h_{(T_1 \bigcup T_2)^c}\|_2 \tag{21}$$

$$\leq C_s \sqrt{\frac{1+\epsilon_K}{1-\delta_{2K}}} SK^{-s} ln\lceil \frac{N}{K} \rceil + 2K^{-1/2}\|x_{T_1^c}\|_1 + C_s \sqrt{\frac{1+\epsilon_K}{1-\delta_K}} SK^{-s} ln\lceil \frac{N}{K} \rceil \tag{22}$$

□

Combining the above theorems and corollaries, we obtain Theorem. 3.1.

5 Conclusion

In this report, we show that a stable recovery can be guaranteed for linear compressible signals when we randomly select a small portion of the samples supported on graph vertices. Such application is a special case of model based compressive sensing. In order to draw the conclusion, we first prove that a non-uniform bounded orthogonal sensing matrix can satisfy the RAmP. Then combined with $L_{1,K}-$RIP, RAmP contribute to the upper bound of the error induced by ℓ_1 decoding. We believe the technique we develop in this report will have several potential applications, especially in the realm of sensor networks.

References

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[2] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, p. 969, 2007.

[5] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *Information Sciences and Systems, 2006 40th Annual Conference on*. IEEE, 2006, pp. 207–212.

[6] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," McGill University, Tech. Rep., 2011.

[7] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1982–2001, 2010.

[8] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.