COMPRESSIVE SENSING BOUNDS THROUGH A UNIFYING FRAMEWORK FOR SPARSE MODELS

Cem Aksoylar⋆ *George Atia*† *Venkatesh Saligrama*⋆

⋆Boston University, Boston, MA 02215
†University of Central Florida, Orlando, FL 32816

ABSTRACT

In this work we investigate the sample complexity of support recovery in sparse signal processing models, with special focus on two compressive sensing scenarios. In particular, we consider models where N covariates $X = (X_1, \ldots, X_N)$ along with outcome Y are observed, with the assumption that the outcome Y is conditionally independent of the other covariates given $K \ll N$ covariates. Using asymptotic information theoretic analyses, we establish sufficient conditions on the number of samples in order to successfully recover the K salient covariates. We apply our results to two variants of the compressive sensing (CS) problem: (1) compressive sensing with a measurement noise model, (2) 1-bit quantized compressive sensing. In both models we consider sensing with independent and correlated Gaussian sensing matrices. We show that the sufficiency bounds we obtain on the number of measurements in both cases are comparable to the best known bounds while providing a novel perspective for the theoretical analysis of such models. In addition, we quantify how the correlation between the sensing columns affects the number of measurements. Our findings for the CS models demonstrate the applicability and flexibility of our general results on the sample complexity in sparse signal processing models.

Index Terms— Sparse signal processing, compressive sensing, 1-bit compressive sensing, information theory

1. INTRODUCTION

Recent advances in sensing and storage systems have led to the proliferation of high-dimensional data such as images, video or genomic data. Such data cannot be processed efficiently using conventional signal processing methods due to their dimensionality. However, high-dimensional data often exhibit an inherent low-dimensional structure, so they can often be represented "sparsely" in some basis or domain. The discovery of an underlying sparse structure is important in order to compress the acquired data or to develop more robust and efficient processing algorithms.

In this paper, we are concerned with the asymptotic analysis of the sample complexity in problems where we aim to identify a set of *salient* covariates responsible for producing an outcome Y. In particular, we assume that among a set of N covariates $X = (X_1, \ldots, X_N)$, only K covariates (indexed by set S) are directly relevant to the outcome Y. We formulate this with the assumption that given $X_S = \{X_n\}_{n \in S}$, outcome Y is independent of other covariates $\{X_n\}_{n \notin S}$, i.e.,

$$P(Y|X) = P(Y|X_S). \quad (1)$$

This project was partially supported by NSF Grant 0932114, ONR grant N000141010477, NGA grant HM1582-09-1-0037, NSF grant CCF-0905541 and DHS grant 2008-ST-061-ED0001.

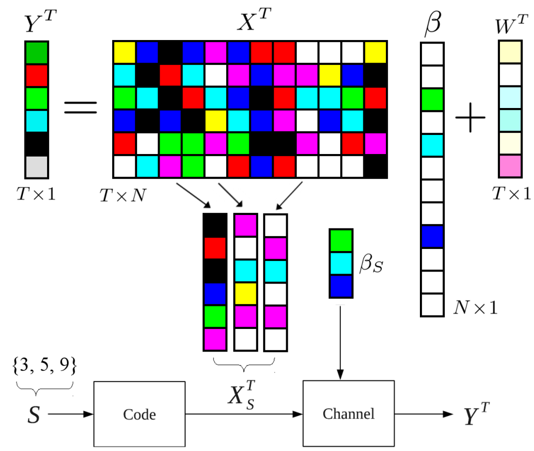

Fig. 1. Compressive sensing example and its mapping to the channel model.

We assume we are given T sample pairs (X, Y) and the problem is to identify the set of salient covariates, S, from these T samples given the knowledge of observation model $P(Y|X_S)$. Our analysis aims to establish sufficient conditions on T in order to recover the set S with an arbitrarily small error probability in terms of K, N, the observation model and other model parameters such as the signal-to-noise ratio.

In addition to the compressive sensing models that will be investigated in this paper, many problems from different areas of signal processing can be formulated in the general sparse recovery framework. Examples include problems such as group testing [1], array signal processing [2], sparse channel estimation [3] or graphical model selection [4]. Consequently, the sample complexity of support recovery in such problems can also be analyzed using the framework presented herein.

The analysis of the sample complexity is performed by posing this identification problem as an equivalent channel coding problem. The salient set S corresponds to the message transmitted through a channel. The set S is encoded by X_S^T of length T, which is the collection of codewords X_n^T for $n \in S$, from a codebook X^T. The coded message X_S^T is transmitted through a channel $P(Y|X_S)$ with output Y^T. As in channel coding, our aim is to identify which message S was transmitted given channel output Y^T and the codebook X^T.

The identification problem was first formulated in a channel coding framework in [1], where it was used to determine sufficient and necessary conditions on the number of tests in the group testing problem with independent and identically distributed (i.i.d.) test assignments. The sufficient condition was derived based on the analysis of a Maximum Likelihood (ML) decoder, while the necessary condition was derived using Fano's inequality [5]. The analysis was extended to general sparse signal processing models with i.i.d. covariates in [6]. The results were further extended to symmetric dependent covariates, utilizing a typicality idea for covariate realizations in [7].

In this paper, we focus on the applications of our results to two compressive sensing problems. First, we state results for general sparse models with an observation model in the form of $P(Y|X_S, \beta_S)$ where β_S is a latent random variable with prior $P(\beta_S)$. For instance, in the context of compressive sensing, β_S corresponds to the values of the non-zero coefficients of the support. This observation model extends our framework presented in [7] and allows for more flexibility in the analysis of sparse recovery problems. Then, using the bounds presented in this paper for general sparse models, we derive sufficient conditions for the two CS problems, with both independent and correlated Gaussian sensing matrices. The first CS problem we consider herein is linear CS with measurement noise [8]. The second one is the 1-bit CS problem [9], which is a non-linear model in contrast to the linear CS model.

Compressive sensing is a fairly well-studied problem. The conditions for recovery in linear CS with measurement noise has been described and studied extensively in the literature [8, 10, 11, 12, 13, 14, 15] through the analysis of properties such as the restricted isometry property [16], as well as using information-theoretic approaches. It has been established that $T = \Omega(K \log(N/K))$ is a sufficient condition for support recovery. While our analysis also relies on information theoretic techniques, the CS bounds are derived using our results for general sparse models and also account for sensing matrices with dependent columns.

1-bit CS [9] is interesting as the extreme case of CS models with quantized measurements, which are of practical importance in many real world applications. The conditions on the number of measurements have been studied for both noiseless [17] and noisy [18] models and $T = \Omega(K \log N)$ has been established as a sufficient condition for independent Gaussian sensing matrices.

In the next section, we provide a mathematical description of the considered problems. In Section 3, we state our sufficiency bounds for general models. Section 4 follows with the application of the bounds to CS models. Lastly, in Section 5 we provide some concluding remarks.

2. PROBLEM SETUP

We use upper case and lower case letters to distinguish between random quantities and their realizations. Subscripts and superscripts are used for column and row indexing, as explained in Table 1. Subscripting with a set S implies a selection of columns with indices in S. $\log$ is used to denote logarithm to the base 2 and the natural logarithm is denoted by $\ln$.

Table 1. Reference for notation used

	Random quant.	**Realization**
Covariates	$X_1, \ldots, X_N$	$x_1, \ldots, x_N$
$1 \times \lvert S \rvert$ random vector	X_S	x_S
$T \times N$ random matrix	X^T	x^T
t-th row of X^T	$X^{(t)}$	$x^{(t)}$
n-th column of X^T	X_n^T	x_n^T
n-th el. of t-th row	$X_n^{(t)}$	$x_n^{(t)}$
$T \times \lvert S \rvert$ sub-matrix	X_S^T	x_S^T
Outcome	Y	y
$T \times 1$ outcome vector	Y^T	y^T
t-th element of Y^T	$Y^{(t)}$	$y^{(t)}$

Let $X = (X_1, X_2, \ldots, X_N) \in \mathcal{X}^N$ denote a set of either i.i.d. or exchangeable random covariates with a joint probability distribution $P(X)$. An exchangeable sequence of random variables have the property that the joint probability distribution is invariant to any permutation of the random variables in the sequence, i.e., for any permutation mapping σ,

$$P_{X_1,\ldots,X_N}(x_1, \ldots, x_N) = P_{X_{\sigma(1)},\ldots,X_{\sigma(N)}}(x_1, \ldots, x_N).$$

This assumption implies that for any set $S \subset \{1, 2, \ldots, N\}$, the set of covariates $\{X_k\}_{k \in S}$ is also exchangeable. Hence, X_S is identically distributed for sets S that have the same cardinality. This assumption also implies that all dependencies between covariates are symmetric.

In view of the aforementioned symmetry, we use the compact notation $I_{i,j}$ to denote the mutual information [5] between two *disjoint* sets of i and j covariates, i.e., for continuous covariates,

$$I_{i,j} = \int_{\mathcal{X}^j} \int_{\mathcal{X}^i} P(x_{S_1}, x_{S_2}) \log \frac{P(x_{S_1}, x_{S_2})}{P(x_{S_1})P(x_{S_2})} \, dx_{S_1} \, dx_{S_2}.$$

This quantity is zero for all i, j, for i.i.d. covariates and identical for any two disjoint sets with i and j elements.

We let $Y \in \mathcal{Y}$ denote an observation or outcome, which depends only on a small subset of covariates $S \subset \{1, \ldots, N\}$ of known cardinality $|S| = K$ where $K \ll N$. In particular, Y is conditionally independent of the covariates given the subset of covariates indexed by the index set S, as in (1).

We consider an observation model with a latent random parameter β_S. We assume β_S is independent of covariates X and has a prior distribution $P(\beta_S)$. The outcomes depend on both X_S and β_S and are generated according to the model $P(Y|X_S, \beta_S)$. This latent variable corresponds to the non-zero coefficients of the K-sparse vector β in the CS framework (2). Note that (1) still holds in this model.

We observe the realizations (x^T, y^T) of T covariate-outcome pairs (X^T, Y^T). The covariates $X^{(t)}$ are distributed i.i.d. across $t = 1, \ldots, T$. Note that the outcomes $Y^{(t)}$ are independent for different t only when conditioned on β_S.

We define the sample mutual information for a matrix x^T and two sets S_1, S_2 as

$$\hat{I}_{S_1,S_2}(x^T) = \frac{1}{T} \sum_{t=1}^{T} \log \frac{P(x_{S_1}^{(t)}, x_{S_2}^{(t)})}{P(x_{S_1}^{(t)})P(x_{S_2}^{(t)})}.$$

We let $\hat{S}(X^T, Y^T)$ denote the estimate of the set S and $P(E)$ denote the average probability of error, averaged over all sets S of size K, all possible data samples X^T and outcomes Y^T, i.e.,

$$P(E) = P(\hat{S}(X^T, Y^T) \neq S).$$

For the compressive sensing with output (measurement) noise problem, we have the following normalized model [13],

$$Y^T = X^T \beta + W^T \tag{2}$$

where X^T is the $T \times N$ sensing matrix, β is a K-sparse vector of length N with support S, W^T is the measurement noise of length T and Y^T is the observation vector of length T. In particular, we assume that $X^{(t)}$ is a jointly Gaussian random vector and the vectors are independent across rows t. We assume that each element $X_i^{(t)}$ has zero mean, elements in different columns are correlated with correlation coefficient ρ and each element has variance $\frac{1}{T\rho_K}$, where $\rho_K = 1 + (K-1)\rho$. In other words, different columns have correlation $E[X_i^{(t)} X_j^{(t)}] = \frac{\rho}{T\rho_K}$ for any $i \neq j$, for all t. We assume that each element of W^T is distributed i.i.d. with $W \sim \mathcal{N}\left(0, \frac{1}{\text{SNR}}\right)$. For independent sensing columns $\rho = 0$ and $\rho_K = 1$, therefore the above formulation reduces to the model in [13].

For the 1-bit compressive sensing model, we have

$$Y^T = Q(X^T \beta) \tag{3}$$

where X^T and β are as before and $Q(\cdot)$ is a 1-bit quantizer which outputs 1 if the input is non-negative and 0 otherwise, for each element of the input vector.

In order to analyze the CS problems using the proposed sparse signal processing framework, it is important to observe how the CS model as defined above can be mapped to the general sparse model, as illustrated in Figure 1. In the case of CS, the elements in a row of the sensing matrix correspond to covariates $X_1, \ldots, X_N$ as defined in the beginning of this section. Each row of the sensing matrix is a realization of X and rows are generated i.i.d. to form X^T. It is easy to see that assumption (1) is satisfied in both models, since each measurement $Y^{(t)}$ depends only on the linear combination of the elements $X_S^{(t)}$ that correspond to the support of β. The coefficients of this combination are given by β_S, the values of the non-zero elements of β. β_S corresponds to the latent parameter of the observation model $P(Y|X_S, \beta_S)$. In the linear case, $P(Y|X_S, \beta_S)$ incorporates the measurement noise W. In the quantized case, the observation model can be defined as $P(Y|X_S, \beta_S) = \mathbb{1}_{\{(Y-\frac{1}{2})\langle X_S, \beta_S\rangle \geq 0\}}$, for $Y \in \{0, 1\}$.

3. CONDITIONS FOR RECOVERY

In this section we present our sufficiency conditions for the sample complexity of the support recovery problem with the latent variable observation model introduced in Section 2. Note that with β_S fixed and known, the bounds below reduce to the bounds in [7]. The proofs of the theorems stated in this section are omitted due to space constraints and we refer the reader to [19] for further details.

3.1. Sufficiency – IID covariates

In order to derive our sufficiency bounds, we consider the error probability of an ML decoder. The ML decoder goes through all $\binom{N}{K}$ sets of size K and outputs the set $S^\star$ such that

$$P_{Y^T|X_S^T}(y^T|x_{S^\star}^T) \geq P_{Y^T|X_S^T}(y^T|x_{\hat{S}}^T), \quad \forall \hat{S}, |\hat{S}| = K,$$

where $P_{Y^T|X_S^T}$ is the observation distribution conditioned on the true set S, averaged over values of β_S. We define the error event E to be the event $(S^\star \neq S)$ with average error probability $P(E)$. Note that the ML decoder requires the knowledge of the observation model $P(Y|X_S, \beta_S)$ and the prior $P(\beta_S)$.

Theorem 3.1 *Define $\Xi_S^{\{i\}}$ as the set of tuples $(\mathcal{S}^1, \mathcal{S}^2)$ partitioning the true set S into disjoint sets $\mathcal{S}^1$ and $\mathcal{S}^2$ with cardinalities i and $K-i$, respectively, i.e.,*

$$\Xi_S^{\{i\}} = \Big\{(\mathcal{S}^1, \mathcal{S}^2) : \mathcal{S}^1 \cap \mathcal{S}^2 = \emptyset, \\ \mathcal{S}^1 \cup \mathcal{S}^2 = S, |\mathcal{S}^1| = i, |\mathcal{S}^2| = K - i\Big\}.$$

If the number of samples T is such that

$$T > (1+\epsilon) \cdot \max_{\substack{i=1,\ldots,K, \\ (\mathcal{S}^1,\mathcal{S}^2) \in \Xi_S^{\{i\}}}} \frac{\log \binom{N-K}{i}\binom{K}{i}}{I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y|\beta_S)}, \tag{4}$$

then, asymptotically the average error probability approaches zero, i.e.,

$$\lim_{K\to\infty} \lim_{N\to\infty} P(E) = 0,$$

where $\epsilon > 0$ is an arbitrary constant independent of N and K and $I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y|\beta_S)$ is the mutual information [5] between $X_{\mathcal{S}^1}$ and $(X_{\mathcal{S}^2}, Y)$ conditioned on β_S. [1]

Intuitively, the bound in (4) can be explained as follows: For each i, the numerator is the number of bits required to represent all sets that differ from S in i elements. The denominator represents the information given by the subset $\mathcal{S}^2$ of $K-i$ true elements and the output variable Y about the remaining i covariates $\mathcal{S}^1$. Hence, the ratio represents the number of samples needed to control i support errors and the maximization accounts for all possible support errors.

3.2. Sufficiency – Dependent covariates

We define a notion of δ-*typicality* for realizations of the covariate matrix x^T. Intuitively, δ-typicality holds when the sample mutual information between a candidate set $\hat{S}$ and the real set S are δ-close to the true mutual information.

We now mathematically define the collection of δ-typical matrices G_δ. First, define

$$\mathcal{A}_i = \{\hat{S} : |\hat{S}| = K, |S \cap \hat{S}| = K - i\},$$

as the collection of sets that differ from set S in i elements, for $i = 1, \ldots, K$. For such sets, we define

$$G_{\delta,i,\hat{S}} = \Big\{x^T : \Big|\hat{I}_{\hat{S}\setminus S, S}(x^T) - I_{i,K}\Big| \leq \delta, \\ \Big|\hat{I}_{S\cap\hat{S}, S\setminus\hat{S}}(x^T) - I_{i,K-i}\Big| \leq \delta\Big\}, \quad \hat{S} \in \mathcal{A}_i,$$

where $\setminus$ is the set difference operator, i.e., $A \setminus B = A \cap B^c$. Finally, we define the collection of typical matrices as

$$G_\delta = \bigcap_{i=1}^{K} \bigcap_{\hat{S} \in \mathcal{A}_i} G_{\delta,i,\hat{S}}. \tag{5}$$

In order to analyze the probability of error, we first separate the error event into two events: the identification error with a typical covariate matrix and the error with an atypical covariate matrix, hence,

$$P(E) \leq P(E, X^T \in G_\delta) + P(X^T \notin G_\delta).$$

For the first type of error, we have the following sufficiency bound.

Theorem 3.2 *If the number of samples T is such that*

$$T > (1+\epsilon) \cdot \max_{\substack{i=1,\ldots,K, \\ (\mathcal{S}^1,\mathcal{S}^2) \in \Xi_S^{\{i\}}}} \frac{\log \binom{N-K}{i}\binom{K}{i}}{I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y|\beta_S) - I_{i,K} - 3\delta}, \tag{6}$$

then, the error probability $P(E, X^T \in G_\delta)$ asymptotically approaches zero, i.e.,

$$\lim_{K\to\infty} \lim_{N\to\infty} P(E, X^T \in G_\delta) = 0,$$

for any $\delta > 0$ for which the denominator is positive and an arbitrary constant $\epsilon > 0$ independent of N and K. Note that δ can be chosen to scale with K. [1]

[1] The results stated in this paper are valid for the case where the support size K is fixed with respect to the dimension N. We refer the reader to [19] for results for the more general regime $K = o(N)$, where K and N scale simultaneously.

The result we obtain is similar to the independent case, except for penalties in the denominator induced by the dependencies between covariates.

Although it is easy to see that the δ-typicality holds asymptotically for any $\delta > 0$ by the law of large numbers, its probability depends on the scaling of T with respect to K, N and δ. This relationship should be investigated to show that $P(X^T \notin G_\delta) \to 0$ and therefore $P(E) \to 0$, using Theorem 3.2.

The analysis of $P(X^T \notin G_\delta)$ depends on the covariate distribution $P(X)$. For instance we provide a sufficient condition for Gaussian covariates in Section 4 and Theorem 3.2 in [7] provides a sufficient condition that is useful for discrete covariates.

4. APPLICATIONS TO CS MODELS

In this section, utilizing the sufficiency bound of Theorem 3.2, we derive sufficient conditions on T to recover S for the linear CS and 1-bit CS problems. To simplify the analysis and exposition, we analyze the degenerate case of $\beta \in \{0,1\}^N$, i.e., the latent variable β_S is known and equal to the vector of 1's. However, the general case where β_S is random with a known prior distribution can also be analyzed using the condition given by Theorem 3.2.

In order to obtain the model-specific bounds, we analyze the conditions for δ-typicality with correlated Gaussian sensing columns and the terms $I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y|\beta_S)$ and $I_{i,K}$ for the two problems. For brevity, the proofs are deferred to [19].

First, we state a sufficient condition for δ-typicality for correlated Gaussian columns (as formulated in Section 2) in the following lemma.

Lemma 4.1 *For symmetrically dependent Gaussian sensing columns with correlation coefficient ρ, $T = \Omega(K \log N)$ samples are asymptotically sufficient to satisfy the δ-typicality (5) for $\delta = \Omega(\rho K)$, i.e., if $\delta = \Omega(\rho K)$ and $T = \Omega(K \log N)$, $P(G_\delta) \to 1$ as $K, N \to \infty$.*

For the following results, let $\alpha = \frac{i}{K}$ denote the support distortion, i.e., the ratio of misidentified elements of the support S. Note that $\frac{1}{K} \leq \alpha \leq 1$.

4.1. Compressive sensing with measurement noise

We obtain the following lemma by analyzing the CS model.

Lemma 4.2

$$I(X_{\mathcal{S}^1};X_{\mathcal{S}^2}, Y) - I_{i,K} = \frac{\ln 2}{2} \log \left(\left[1 + \alpha \frac{K\rho}{\rho_K} \right] \cdot \left[1 + \alpha \frac{1 - P\frac{K\rho}{\rho_K}}{P-1} \right] \right)$$

where $P = 1 + T/(K \cdot SNR)$. Note that the equality holds for any value of K, N, T, ρ and SNR.

Note that δ should satisfy that $\delta = O(I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y) - I_{i,K})$ for all α in view of Theorem 3.2, along with $\delta = \Omega(\rho K)$ to satisfy the δ-typicality condition.

Combining Lemma 4.1 and 4.2, considering all values of α for exact recovery and noting that the numerator $\log \binom{N-K}{i}\binom{K}{i} = \Theta(\alpha K \log N)$, we can readily state the following theorem.

Theorem 4.1 *For compressive sensing with independent or correlated Gaussian sensing columns with correlation coefficient $\rho = O(\frac{1}{K^2})$ and SNR $= \Omega(\log N)$ (which is a necessary condition for recovery [13]), $T = \Omega(K \log N)$ measurements are sufficient to recover S, the support of β, with an arbitrarily small error probability.*

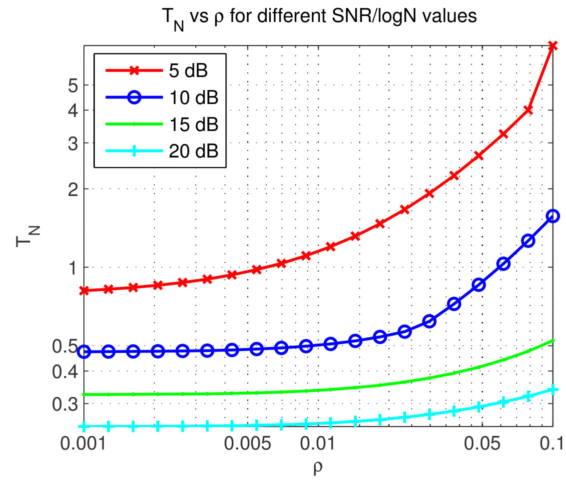

Fig. 2. Effects of ρ and SNR on the sufficient number of measurements given by Lemma 4.2 and Theorem 3.2, evaluated for $K = 50$ and $N = 10000$, where $T_N = T/(K \log(N/K))$.

Figure 2 illustrates our sufficiency bound for the number of measurements as a function of correlation coefficient ρ for different values of SNR, evaluated for finite K and N. The vertical axis of the figure represents the number of measurements normalized by $K \log(N/K)$. As expected, increasing the correlation increases the number of measurements. Similary, increasing SNR decreases the number of measurements.

4.2. 1-bit quantized compressive sensing

We obtain a similar result in the 1-bit CS problem, with a less stringent condition on the correlation coefficient ρ.

Theorem 4.2 *For 1-bit CS with independent or correlated Gaussian columns with correlation coefficient $\rho = O(\frac{1}{K\sqrt{K}})$, $T = \Omega(K \log N)$ measurements are sufficient to recover the support S with an arbitrarily small error probability.*

5. CONCLUSIONS

Based on the analysis of a general sparse model, we obtained a bound which is asymptotically identical to the best-known bound $T = \Omega(K \log(N/K))$ [13] for the linear CS problem with an independent Gaussian sensing matrix, in the sublinear sparsity regime. Similarly, for 1-bit CS we provide a sufficiency bound that matches [17] for independent covariates.

In addition, this analysis provides insight into how the dependence of the columns of the sensing matrix could affect the performance for the two problems. It is also shown that with a correlation coefficient that vanishes polynomially in K, the performance is identical to sensing with independent columns in both cases.

By leveraging bounds derived for a general sparse model, we established explicit sufficient conditions for the two variants of the CS framework, namely linear CS and 1-bit CS. Therefore we demonstrated that we can obtain tight and useful bounds for sparse recovery problems of interest, using the results for the general model presented in [7] and this paper. The flexibility of our formulation also allows the application of our results to many other sparse recovery problems that have not been mentioned in this paper.

6. REFERENCES

[1] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, March 2012.

[2] D. Malioutov, M. Çetin, and A.S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[3] S.F. Cotter and B.D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374 –377, March 2002.

[4] N. Santhanam and M.J. Wainwright, "Information-theoretic limits of graphical model selection in high dimensions," in *Proc. of the Int. Symp. on Information Theory (ISIT)*, July 2008, pp. 2136 –2140.

[5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: John Wiley and Sons, Inc., 1991.

[6] G. Atia and V. Saligrama, "A mutual information characterization for sparse signal processing," in *Proc. of Int. Colloq. on Automata, Languages and Programming (ICALP)*, Switzerland, July 2011.

[7] C. Aksoylar, G. Atia, and V. Saligrama, "Sample complexity of salient feature identification for sparse signal processing," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, Aug. 2012, pp. 329–332.

[8] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

[9] P.T. Boufounos and R.G. Baraniuk, "1-bit compressive sensing," in *Proc. of Conf. on Information Sciences and Systems (CISS)*, March 2008, pp. 16–21.

[10] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," in *Proc. of the Int. Symp. on Information Theory (ISIT)*, Nice, France, June 2007.

[11] A.K. Fletcher, S. Rangan, and V.K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, 2009.

[12] M.J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programs," in *Allerton Conf. on Communication, Control and Computing*, Monticello, IL, 2006.

[13] S. Aeron, M. Zhao, and V. Saligrama, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5111–5130, Oct. 2010.

[14] M. Akcakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan.

[15] Y. Wu and S. Verdu, "Optimal phase transitions in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6241–6263, Oct.

[16] E.J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 910, pp. 589–592, 2008.

[17] L. Jacques, J.N. Laska, P.T. Boufounos, and R.G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *arXiv preprint arXiv:1104.3160*, 2011.

[18] A. Gupta, R. Nowak, and B. Recht, "Sample complexity for 1-bit compressed sensing and sparse classification," in *Proc. of the Int. Symp. on Information Theory (ISIT)*, June 2010, pp. 1553–1557.

[19] C. Aksoylar, G. Atia, and V. Saligrama, "Supplementary notes," `http://people.bu.edu/aksoylar/icassp2013_extended.pdf`, March 2013.