Information-theoretic limits on sparse support recovery: Dense versus sparse measurements

Wei Wang, Martin J. Wainwright, and Kannan Ramchandran
Department of Electrical Engineering
and Computer Sciences, UC Berkeley
Berkeley, CA 94720
{wangwei, wainwrig, kannanr}@eecs.berkeley.edu

***Abstract*— We study the information-theoretic limits of exactly recovering the support of a sparse signal using noisy projections defined by various classes of measurement matrices. Our analysis is high-dimensional in nature, in which the number of observations n, the ambient signal dimension p, and the signal sparsity k are all allowed to tend to infinity in a general manner. This paper makes two novel contributions. First, we provide sharper necessary conditions for exact support recovery using general (non-Gaussian) dense measurement matrices. Combined with previously known sufficient conditions, this result yields a sharp characterization of when the optimal decoder can recover a signal with linear sparsity ($k = \Theta(p)$) using a linear scaling of observations ($n = \Theta(p)$) in the presence of noise. Our second contribution is to prove necessary conditions on the number of observations n required for asymptotically reliable recovery using a class of γ-sparsified measurement matrices, where the measurement sparsity $\gamma(n, p, k) \in (0, 1]$ corresponds to the fraction of non-zero entries per row. Our analysis allows general scaling of the quadruplet (n, p, k, γ), and reveals three different regimes, corresponding to whether measurement sparsity has no effect, a minor effect, or a dramatic effect on the information-theoretic limits of the subset recovery problem.**

I. Introduction

The problem of estimating a k-sparse vector $\beta \in \mathbb{R}^p$ based on a set of n noisy linear observations is of broad interest, arising in subset selection in regression, graphical model selection, group testing, signal denoising, sparse approximation, and compressive sensing. A large body of recent work (e.g., [6], [3], [14]) has analyzed the use of ℓ_1-relaxation methods for estimating high-dimensional sparse signals, and established conditions (on signal sparsity and the choice of measurement matrices) under which they succeed with high probability.

Of complementary interest are the information-theoretic limits of the sparsity recovery problem, which apply to the performance of any procedure regardless of its computational complexity. Such analysis has two purposes: first, to demonstrate where known polynomial-time methods achieve the information-theoretic bounds, and second, to reveal situations in which current methods are sub-optimal. With this motivation, this paper makes two contributions. First, we derive sharper necessary conditions for exact support recovery, applicable to a general class of dense measurement matrices (including non-Gaussian ensembles). In conjunction with the sufficient conditions from previous work [15], this analysis provides a sharp characterization of necessary and sufficient conditions for exact support recovery in the regime of linear signal sparsity ($k = \Theta(p)$) using a linear fraction of observations ($n = \Theta(p)$). Our second contribution is to address the effect of *measurement sparsity*, meaning the fraction γ of non-zeros per row in the matrices used to collect measurements. We derive lower bounds on the number of observations required for exact sparsity recovery, as a function of the signal dimension p, signal sparsity k, and measurement sparsity $\gamma \in (0, 1]$. This analysis highlights a trade-off between the statistical efficiency of a measurement ensemble and the computational complexity associated with storing and manipulating it. Given space constraints, this paper only provides statements and high-level proof ideas; we refer the reader to the technical report [17] for details.

A. Problem formulation

In this paper, we consider a deterministic signal model, in which $\beta \in \mathbb{R}^p$ is a fixed but unknown vector with exactly k non-zero entries. We refer to k as the *signal sparsity* and p as the *signal dimension*, and define the support set of β as

$$S \quad := \quad \{i \in \{1, \ldots, p\} | \beta_i \neq 0\}. \tag{1}$$

Note that there are $N = \binom{p}{k}$ possible support sets, corresponding to the N possible k-dimensional subspaces in which β can lie. We are given a vector of n noisy observations $Y \in \mathbb{R}^n$, of the form

$$Y \quad = \quad X\beta + W, \tag{2}$$

where $X \in \mathbb{R}^{n \times p}$ is the measurement matrix, and $W \sim N(0, \sigma^2 I_{n\times n})$ is additive Gaussian noise. Throughout this paper, we assume without loss of generality that $\sigma^2 = 1$, since any scaling of σ can be accounted for in the scaling of β.

Our goal is to perform exact recovery of the support set S, which corresponds to a standard model selection error criterion. More precisely, we measure the error between the estimate $\widehat{\beta}$ and the true signal β using the $\{0, 1\}$-valued loss function:

$$\rho(\widehat{\beta}, \beta) := \mathbb{I}\left[\{\widehat{\beta}_i \neq 0,\ \forall i \in S\} \cap \{\widehat{\beta}_j = 0,\ \forall j \notin S\}\right]. \tag{3}$$

The results of this paper apply to arbitrary decoders. Any decoder is a mapping g from the observations Y to an estimated subset $\widehat{S} = g(Y)$. Let $\mathbb{P}[g(Y) \neq S \mid S]$ be the conditional probability of error given that the true support is S. Assuming that β has support S chosen uniformly at random over the N possible subsets of size k, the average probability of error is given by

$$p_{err} = \frac{1}{\binom{p}{k}} \sum_{S} \mathbb{P}[g(Y) \neq S \mid S]. \quad (4)$$

We say that sparsity recovery is asymptotically reliable if $p_{err} \to 0$ as $n \to \infty$. Since we are trying to recover the support exactly from noisy measurements, our results necessarily involve the minimum value of β on its support,

$$\beta_{min} := \min_{i \in S} |\beta_i|. \quad (5)$$

In particular, our results apply to decoders that operate over the signal class

$$\mathcal{C}(\beta_{min}) := \{\beta \in \mathbb{R}^p \mid |\beta_i| \geq \beta_{min} \; \forall i \in S\}. \quad (6)$$

With this set-up, our goal is to find necessary conditions on the quadruple (n, p, k, γ) that any decoder, regardless of its computational complexity, must satisfy for asymptotically reliable recovery to be possible. We are interested in such converse bounds in sparse regimes where both the signal sparsity k and the measurement sparsity γ are allowed to scale with the signal dimension p. As our analysis shows, the appropriate notion of rate for this problem is $R = \frac{\log \binom{p}{k}}{n}$.

B. Related work and our contributions

A body of past work [8], [12], [1] has focused on the information-theoretic limits of sparse estimation under ℓ_2 and other distortion metrics, using power-based SNR measures of the form

$$\text{SNR} := \frac{\mathbb{E}[\|X\beta\|_2^2]}{\mathbb{E}[\|W\|_2^2]} = \|\beta\|_2^2. \quad (7)$$

Note that the second equality assumes that the measurement matrix is standardized, with each element X_{ij} having zero-mean and variance one. The minimum value is related to this power-based measure by the inequality $k\beta_{min}^2 \leq \text{SNR}$. It is important to note that the power-based SNR (7), though appropriate for ℓ_2-distortion, is not suitable for the support recovery problem. For the ensemble of signals $\mathcal{C}(\beta_{min})$ in (6) (with $k > 1$), the SNR measure (7) can be made arbitrarily large, while still having one coefficient β_i equal to the minimum value. Consequently, as our results show, it is possible to generate problem instances for which support recovery is arbitrarily difficult (by sending $\beta_{min} \to 0$ at a rapid rate), even as the SNR (7) becomes arbitrarily large.

The paper [15] was the first to consider the information-theoretic limits of exact subset recovery using dense Gaussian measurement ensembles, explicitly tracking the minimum value β_{min}. This analysis yielded necessary and sufficient conditions on general quadruples (n, p, k, β_{min}) for asymptotically reliable recovery. Subsequent work [11], [2] has extended this type of analysis to the criterion of partial support recovery.

In this paper, we consider only exact support recovery, but provide results for general dense measurement ensembles, thereby generalizing previous results. In conjunction with known sufficient conditions [15], one consequence of our first main result (Theorem 1, below) is a set of sharp necessary and sufficient conditions for the optimal decoder to recover the support of a signal with linear sparsity ($k = \Theta(p)$), using only a linear fraction of observations ($n = \Theta(p)$). Moreover, for the special case of the standard Gaussian ensemble, Theorem 1 also recovers results independently obtained in concurrent work [11], [7].

Our second main result (Theorem 2) addresses the effect of measurement sparsity, which we assess in terms of the fraction $\gamma \in (0, 1]$ of non-zeros per row of the the measurement matrix X. To appreciate the importance of this issue, dense measurement matrices may lead to prohibitively high computational complexity and storage requirements. Sparse matrices can reduce this complexity, and also lower communication cost and latency in distributed network and streaming applications. With this motivation, some past work in compressive sensing has addressed the issue of measurement sparsity, including work inspired by coding theory [18], [13], sparse random projections [16], and group testing [4], [9]. All of this work deals with the noiseless observation model, in contrast to the noisy observation model (2) considered here. The paper [1] provides results on sparse measurements for noisy problems and distortion-type error metrics, using a Bayesian signal model and power-based SNR that is not appropriate for the subset recovery problem. Finally, concurrent work [10] provides sufficient conditions for support recovery using the Lasso (ℓ_1-constrained quadratic programming) for appropriately sparsified ensembles; these results can be viewed as complementary to the information-theoretic analysis of this paper.

Measurement sparsity, though certainly beneficial from the computational standpoint, may have a negative impact on statistical efficiency. Therefore, an important issue is to characterize the trade-off between measurement sparsity and statistical efficiency. With this motivation, our second main result (Theorem 2 below) provides necessary conditions for exact support recovery, using γ-sparsified Gaussian measurement matrices (see equation (8)), for general scalings of the parameters $(n, p, k, \beta_{min}, \gamma)$. This analysis reveals three regimes of interest, corresponding to whether measurement sparsity has no effect, a small effect, or a significant effect on the number of measurements necessary for recovery. Thus, there exist regimes in which measurement sparsity fundamentally alters the ability of any method to decode.

II. MAIN RESULTS AND CONSEQUENCES

In this section, we state our main results, and discuss some of their consequences. Our analysis applies to random ensembles of measurement matrices $X \in \mathbb{R}^{n \times p}$, where each entry X_{ij} is drawn i.i.d. from some underlying distribution. The most commonly studied random ensemble is the standard Gaussian case, in which each $X_{ij} \sim N(0, 1)$. Note that this choice generates a highly dense measurement matrix X, with np non-zero entries. Theorem 1 applies to more general ensembles that satisfy the moment conditions $\mathbb{E}[X_{ij}] = 0$ and $\text{var}(X_{ij}) = 1$, which allows for a variety of non-Gaussian distributions (e.g., uniform, Bernoulli etc.). In addition, we also derive results for γ-sparsified matrices X, in which each entry X_{ij} is i.i.d. drawn according to

$$X_{ij} = \begin{cases} N(0, \frac{1}{\gamma}) & \text{w.p. } \gamma \\ 0 & \text{w.p. } 1-\gamma \end{cases}. \tag{8}$$

Note that when $\gamma = 1$, X is exactly the standard Gaussian ensemble. We refer to the sparsification parameter $0 \le \gamma \le 1$ as the *measurement sparsity*. Our analysis allows this parameter to vary as a function of (n, p, k).

A. Tighter bounds on dense ensembles

We begin by noting an analogy to the Gaussian channel coding problem that yields a straightforward but loose set of necessary conditions. Support recovery can be viewed as a channel coding problem, in which there are $N = \binom{p}{k}$ possible support sets of β, corresponding to messages to be sent over a Gaussian channel with noise variance 1. The effective code rate is then $R = \frac{\log \binom{p}{k}}{n}$. If each support set S is encoded as the codeword $c(S) = X\beta$, where X has i.i.d. Gaussian entries, then by standard Gaussian channel capacity results, we immediately obtain a lower bound on the number of observations n necessary for asymptotically reliable recovery,

$$n \quad > \quad \frac{\log \binom{p}{k}}{\frac{1}{2}\log\left(1 + \|\beta\|_2^2\right)}. \tag{9}$$

This bound is tight for $k = 1$ and Gaussian measurements, but loose in general. As Theorem 1 clarifies, there are additional elements in the support recovery problem that distinguish it from a standard Gaussian coding problem: first, the signal power $\|\beta\|_2^2$ does not capture the inherent problem difficulty for $k > 1$, and second, there is overlap between support sets for $k > 1$. The following result provides sharper conditions on subset recovery.

Theorem 1 (General ensembles)**.** *Let the measurement matrix $X \in \mathbb{R}^{n\times p}$ be drawn with i.i.d. elements from any distribution with zero-mean and variance one. Then a necessary condition for asymptotically reliable recovery over the signal class $\mathcal{C}(\beta_{min})$ is*

$$n \quad > \quad \max\{f_1(p,k,\beta_{min}),\ f_2(p,k,\beta_{min}),\ k-1\}, \tag{10}$$

where

$$f_1(p,k,\beta_{min}) \quad := \quad \frac{\log\binom{p}{k} - 1}{\frac{1}{2}\log\left(1 + k\beta_{min}^2\left(1 - \frac{k}{p}\right)\right)} \tag{11a}$$

$$f_2(p,k,\beta_{min}) \quad := \quad \frac{\log(p-k+1) - 1}{\frac{1}{2}\log(1 + \beta_{min}^2(1 - \frac{1}{p-k+1}))} \tag{11b}$$

In addition to the standard Gaussian ensemble ($X_{ij} \sim N(0,1)$), this result also covers matrices from other common ensembles (e.g., Bernoulli $X_{ij} \in \{-1,+1\}$). It generalizes and strengthens earlier results on subset recovery [15]. Note that $\|\beta\|_2^2 \ge k\beta_{min}^2$ (with equality in the case when $|\beta_i| = \beta_{min}$ for all indices $i \in S$), so that this bound is strictly tighter than the intuitive bound (9). Moreover, by fixing the value of β at $(k-1)$ indices to β_{min} and allowing the last component of β to tend to infinity, we can drive the power $\|\beta\|_2^2$ to infinity, while still having the minimum enter the lower bound.

	Necessary conditions (Theorem 1)	Sufficient conditions (Wainwright [15])
$k = \Theta(p)$, $\beta_{min}^2 = \Theta(\frac{1}{k})$	$\Theta(p\log p)$	$\Theta(p\log p)$
$k = o(p)$, $\beta_{min}^2 = \Theta(\frac{1}{k})$	$\Theta(k\log(p-k))$	$\Theta(k\log(p-k))$
$k = \Theta(p)$, $\beta_{min}^2 = \Theta(\frac{\log k}{k})$	$\Theta(p)$	$\Theta(p)$
$k = o(p)$, $\beta_{min}^2 = \Theta(\frac{\log k}{k})$	$\Theta\left(\frac{k\log\frac{p}{k}}{\log\log k}\right)$	$\Theta\left(k\log\frac{p}{k}\right)$
$k = \Theta(p)$, $\beta_{min}^2 = \Theta(1)$	$\Theta(p)$	$\Theta(p)$
$k = o(p)$, $\beta_{min}^2 = \Theta(1)$	$\Theta\left(\frac{k\log\frac{p}{k}}{\log k}\right)$	$\Theta\left(k\log\frac{p}{k}\right)$

TABLE I

TIGHT NECESSARY AND SUFFICIENT CONDITIONS ARE OBTAINED IN SEVERAL REGIMES OF INTEREST.

Theorem 1 has some consequences related to results proved in concurrent work. Reeves and Gastpar [11] have shown that in the regime of linear sparsity $k/p = \alpha > 0$, if any decoder is given only a linear fraction sample size (meaning that $n = \Theta(p)$), then in order to recover the support exactly, one must have $k\beta_{min}^2 \to +\infty$. This result is one corollary of Theorem 1, since if $\beta_{min}^2 = \Theta(1/k)$, then we have

$$n \;>\; \frac{\log(p-k+1)-1}{\frac{1}{2}\log(1+\Theta(1/k))} \;=\; \Omega(k\log(p-k)) \;\gg \Theta(p),$$

so that the scaling $n = \Theta(p)$ is precluded. In other concurrent work, Fletcher et al. [7] used direct methods to show that for the special case of the standard Gaussian ensemble, the number of observations must satisfy $n > \Omega\left(\frac{\log(p-k)}{\beta_{min}^2}\right)$. This bound is a consequence of our lower bound $f_2(p,k,\beta_{min})$; moreover, Theorem 1 implies the same lower bound for general (non-Gaussian) ensembles as well.

In the regime of linear sparsity, Wainwright [15] showed, by direct analysis of the optimal decoder, that the scaling $\beta_{min}^2 = \Omega(\log(k)/k)$ is sufficient for exact support recovery using a linear fraction $n = \Theta(p)$ of observations. Combined with the necessary condition in Theorem 1, we obtain the following corollary that provides a sharp characterization of the linear-linear regime:

Corollary 1 (Sharp conditions for linear-linear recovery)**.** *Consider the regime of linear sparsity, meaning that $k/p = \alpha \in (0,1)$, and suppose that a linear fraction $n = \Theta(p)$ of observations are made. Then the optimal decoder can recover the support exactly if and only if $\beta_{min}^2 = \Omega(\log k/k)$.*

B. Effect of measurement sparsity

We now turn to the effect of measurement sparsity on recovery, considering in particular the γ-sparsified ensemble (8).

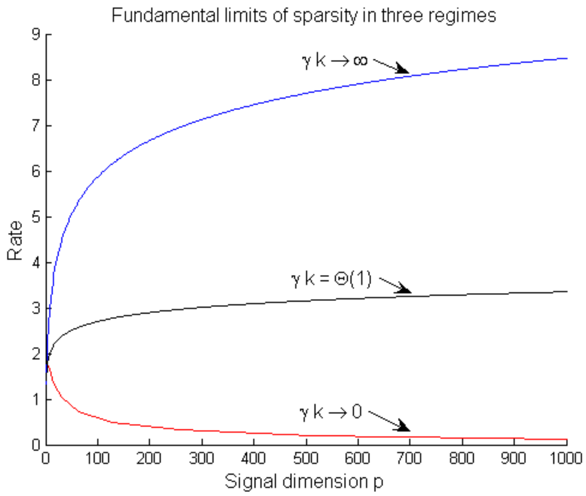

Fig. 1. The rate $R = \frac{\log \binom{p}{k}}{n}$ is plotted using equation (12) in three regimes, depending on how the quantity γk scales, where $\gamma \in [0,1]$ denotes the measurement sparsification parameter and $k < p$ denotes the signal sparsity.

Even though the average signal-to-noise ratio of our channel remains the same (since $\mathrm{var}(X_{ij}) = 1$ for all choices of γ by construction), the Gaussian channel coding bound (9) is clearly not tight for sparse X, even in the case of $k = 1$. The loss in statistical efficiency is due to the fact that we are constraining our codebook to have a sparse structure, which may be far from a capacity-achieving code. Theorem 1 applies to any ensemble in which the components are zero-mean and unit variance. However, if we apply it to the γ-sparsified ensemble, it yields lower bounds that are independent of γ. Intuitively, it is clear that the procedure of γ-sparsification should cause deterioration in support recovery. Indeed, the following result provides refined bounds that capture the effects of γ-sparsification. Let $L \sim \mathrm{Bin}(k, \gamma)$ denote a binomial variate, and let $H(\cdot)$ denote the entropy function, and for any scalar γ, let $H_{binary}(\gamma)$ denote the entropy of an $\mathrm{Ber}(\gamma)$ variate. With this notation, we have the following result.

Theorem 2 (Sparse ensembles)**.** *Let the measurement matrix $X \in \mathbb{R}^{n \times p}$ be drawn with i.i.d. elements from the γ-sparsified Gaussian ensemble* (8)*. Then a necessary condition for asymptotically reliable recovery over the signal class $\mathcal{C}(\beta_{min})$ is*

$$n \quad > \quad \max\left\{g_1(p,k,\beta_{min},\gamma), \quad g_2(p,k,\beta_{min},\gamma)\right\}, \tag{12}$$

where

$$g_1 \quad := \quad \frac{\log \binom{p}{k} - 1}{\mathbb{E}_L\left[\frac{1}{2}\log\left(1 + \frac{L\beta_{min}^2}{\gamma}\right)\right] + H(L)} \tag{13a}$$

$$g_2 \quad := \quad \frac{\log(p-k+1) - 1}{\frac{\gamma}{2}\log(1 + \frac{\beta_{min}^2}{\gamma}) + H_{binary}(\gamma)}. \tag{13b}$$

The bound in Theorem 2 is plotted in Figure 1, showing distinct regimes of behavior depending on how the quantity γk scales, where $\gamma \in [0,1]$ is the measurement sparsification parameter and k is the signal sparsity index. For the support recovery problem, even sparse measurements that do not hit the non-zero signal values give information about their locations. Corollary 2 below precisely characterizes how much measurement sparsity the decoder can tolerate before performance degrades.

Corollary 2 (Three regimes)**.** *The first necessary condition g_1 in Theorem 2 can be simplified in the following three cases.*

1) *If $\gamma k > 1$, then*

$$n \quad > \quad \frac{\log \binom{p}{k} - 1}{\frac{1}{2}\log\left(1 + k\beta_{min}^2\right)}. \tag{14}$$

2) *If $\gamma k = \alpha$ for some constant α, then*

$$n \quad > \quad \frac{\log \binom{p}{k} - 1}{\frac{1}{2}\alpha e^{\alpha}\log\left(1 + \frac{k\beta_{min}^2}{\alpha}\right) + H(L)}, \tag{15}$$

where $H(L) \le \frac{1}{2}\log(2\pi e(\alpha + \frac{1}{12}))$.

3) *If $\gamma k \le 1$, then*

$$n \quad > \quad \frac{\log \binom{p}{k} - 1}{\frac{1}{2}e\gamma k\log\left(1 + \frac{\beta_{min}^2}{\gamma}\right) + kH_{binary}(\gamma)}, \tag{16}$$

where $H_{binary}(\gamma)$ is the binary entropy function.

Corollary 2 reveals three regimes of behavior, defined by the scaling of the measurement sparsity γ and the signal sparsity k. If $\gamma k \to \infty$ as $p \to \infty$, then the recovery threshold (14) is of the same order as the threshold for dense measurement ensembles. In this regime, sparsifying the measurement ensemble has no asymptotic effect on performance. In sharp contrast, if $\gamma k \to 0$ as $p \to \infty$, then the recovery threshold (16) changes fundamentally compared to the dense case. In particular, if $\gamma = o(\frac{1}{k \log k})$ and the minimum value β_{min}^2 does not increase with k, then the denominator in (16) goes to zero. Hence, the number of measurements that any decoder needs in order to recover reliably increases dramatically in this regime. Finally, if $\gamma k = \Theta(1)$, then the recovery threshold (15) transitions between the two extremes. In this middle regime, measurement sparsification only affects performance by a constant factor.

III. Proof Sketches

In this section, we provide the basic intuition underlying the proofs of Theorems 1 and 2. Establishing necessary conditions for exact sparsity recovery amounts to finding conditions on (n, p, k, β_{min}) (and possibly γ) under which the probability of error of any recovery method stays bounded away from zero as $n \to \infty$. At a high-level, our general approach is quite simple: we consider restricted problems in which the decoder has been given some additional side information, and then apply Fano's inequality [5] to lower bound the probability of error. We consider two classes of restricted ensembles: one which captures the bulk effect of having many competing subsets at large distances, and the other which captures the effect of a smaller number of subsets at very close distances. In all cases, we assume that the support S of β is chosen randomly and uniformly over all $\binom{p}{k}$ possible support sets.

A. Restricted problem A: Bulk effects

In the first restricted problem, also exploited in previous work [15], we assume that while the support set S is unknown, the decoder knows *a priori* that $\beta_j = \beta_{min}$ for all $j \in S$. In other words, the decoder knows the value of β on its support, but it does not know the locations of the non-zeros. If a decoder can recover the support of any p-dimensional k-sparse vector β, then it must be able to recover a k-sparse vector that is constant on its support. Furthermore, having knowledge of the value β_{min} at the decoder cannot increase the probability of error. Finally, we assume that $\beta_j = \beta_{min}$ for all $j \in S$ to construct the most difficult possible instance within our ensemble. Thus, we can apply Fano's inequality to lower bound the probability of error in the restricted problem, and so obtain a lower bound on the probability of error for the general problem. This procedure yields the lower bounds $f_1(p, k, \beta_{min})$ and $g_1(p, k, \beta_{min}, \gamma)$ in Theorems 1 and 2 respectively.

B. Restricted ensemble B: Near-by subsets

The second restricted ensemble is designed to capture the confusability effects of the relatively small number $(p-k+1)$ of very close-by subsets. This restricted ensemble is defined as follows. Suppose that the decoder is given the locations of all but the smallest non-zero value of the vector β, as well as the values of β on its support. More precisely, let $j^\star$ denote the unknown location of the smallest non-zero value of β, which we assume achieves the minimum (i.e., $\beta_{j^\star} = \beta_{min}$), and let $T = S \setminus \{j^\star\}$. Given knowledge of $(T, \beta_T, \beta_{min})$, the decoder may simply subtract $X_T \beta_T = \sum_{j \in T} X_j \beta_j$ from Y, so that it is left with the modified n-vector of observations

$$\widetilde{Y} \quad := \quad X_{j^\star} \beta_{min} + W. \tag{17}$$

By re-ordering indices as need be, we may assume without loss of generality that $T = \{p-k+2, \ldots, p\}$, so that $j^\star \in \{1, \ldots, p-k+1\}$. The remaining sub-problem is to determine, given the observations $\widetilde{Y}$, the location of the single non-zero. Note that when we assume that the support of β is uniformly chosen over all $N := \binom{p}{k}$ possible subsets of size k, then given T, the location of the remaining non-zero is uniformly distributed over $\{1, \ldots, p-k+1\}$. It is clear that the probability of error of this restricted problem gives us a lower bound on the probability of error in the original problem, since the decoder has been provided side information which cannot be harmful. As before, we can apply Fano's inequality to lower bound the probability of error in this restricted problem, thereby obtaining the lower bounds $f_2(p, k, \beta_{min})$ and $g_2(p, k, \beta_{min}, \gamma)$ in Theorems 1 and 2 respectively.

IV. Discussion

In this paper, we have studied the information-theoretic limits of exact support recovery for general scalings of the quadruplet (n, p, k, γ). Our first result (Theorem 1) applies generally to measurement matrices with zero-mean and unit variance entries. It strengthens previously known bounds, and combined with known sufficient conditions [15], yields a sharp characterization of recovering signals with linear sparsity with a linear fraction of observations (Corollary 2). Our second result (Theorem 2) applies to γ-sparsified Gaussian measurement ensembles, and reveals three different regimes of measurement sparsity, depending on how significantly they impair statistical efficiency. For linear signal sparsity, Theorem 2 is not a sharp result (by comparison to Theorem 1 in the dense case); however, its tightness for sublinear signal sparsity is an interesting open problem.

Acknowledgment

The work of WW and KR was supported by NSF grant CCF-0635114. The work of MJW was supported by NSF grants CAREER-CCF-0545862 and DMS-0605165.

References

[1] S. Aeron, M. Zhao, and S. Venkatesh. Information-theoretic bounds to sensing capacity of sensor networks under fixed snr. In *Information Theory Workshop*, September 2007.

[2] M. Akcakaya and V. Tarokh. Shannon theoretic limits on noisy compressive sampling. Technical Report arXiv:cs.IT:0711.0366, November 2007.

[3] E. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Info Theory*, 51(12):4203–4215, December 2005.

[4] G. Commode and S. Muthukrishnan. Towards an algorithmic theory of compressed sensing. Technical report, Rutgers University, July 2005.

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

[6] D. Donoho. Compressed sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, April 2006.

[7] A. K. Fletcher, S. Rangan, and V. K. Goyal. Necessary and sufficient conditions on sparsity pattern recovery. Technical Report arXiv:cs.IT:0804.1839, April 2008.

[8] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran. Denoising by sparse approximation: Error bounds based on rate-distortion theory. *Journal on Applied Signal Processing*, 10:1–19, 2006.

[9] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin. Algorithmic linear dimension reduction in the ℓ_1-norm for sparse vectors. In *Proc. Allerton Conference on Communication, Control and Computing*, Allerton, IL, September 2006.

[10] D. Omidiran and M. J. Wainwright. High-dimensional subset recovery in noise: Sparsified measurements without loss of statistical efficiency. Technical report, Department of Statistics, UC Berkeley, April 2008. Short version presented at Int. Symp. Info. Theory, July 2008.

[11] G. Reeves. Sparse signal sampling using noisy linear projections. Master's thesis, UC Berkeley, December 2007.

[12] S. Sarvotham, D. Baron, and R. G. Baraniuk. Measurements versus bits: Compressed sensing meets information theory. In *Proc. Allerton Conference on Control, Communication and Computing*, September 2006.

[13] S. Sarvotham, D. Baron, and R. G. Baraniuk. Sudocodes: Fast measurement and reconstruction of sparse signals. In *Int. Symposium on Information Theory*, Seattle, WA, July 2006.

[14] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy recovery of sparsity using using ℓ_1-constrained quadratic programs. Technical Report 709, Department of Statistics, UC Berkeley, 2006.

[15] M. J. Wainwright. Information-theoretic bounds for sparsity recovery in the high-dimensional and noisy setting. Technical Report 725, Department of Statistics, UC Berkeley, January 2007. Presented at International Symposium on Information Theory, June 2007.

[16] W. Wang, M. Garofalakis, and K. Ramchandran. Distributed sparse random projections for refinable approximation. In *Proc. International Conference on Information Processing in Sensor Networks*, Nashville, TN, April 2007.

[17] W. Wang, M. J. Wainwright, and K. Ramchandran. Information-theoretic limits on sparse support recovery: Dense versus sparse measurements. Technical report, Department of Statistics, UC Berkeley, April 2008.

[18] W. Xu and B. Hassibi. Efficient compressed sensing with deterministic guarantees using expander graphs. In *Information Theory Workshop (ITW)*, September 2007.