Robust Support Recovery Using Sparse Compressive Sensing Matrices

Jarvis Haupt* and Richard Baraniuk†
*University of Minnesota, Minneapolis MN
†Rice University, Houston TX

***Abstract*—This paper considers the task of recovering the support of a sparse, high-dimensional vector from a small number of measurements. The procedure proposed here, which we call the *Sign-Sketch* procedure, is shown to be a robust recovery method in settings where the measurements are corrupted by various forms of uncertainty, including additive Gaussian noise and (possibly unbounded) outliers, and even subsequent quantization of the measurements to a single bit. The Sign-Sketch procedure employs sparse random measurement matrices, and utilizes a computationally efficient support recovery procedure that is a variation of a technique from the sketching literature. We show here that $O(\max\{k\log(n-k), k\log k\})$ non-adaptive linear measurements suffice to recover the support of any unknown n-dimensional vector having no more than k nonzero entries, and that our proposed procedure requires at most $O(n\log n)$ total operations for both acquisition and inference.**

***Index Terms*—Support recovery, sparsity pattern recovery, model selection, feature selection, sparse recovery, robust inference, sketching, compressive sensing.**

I. Introduction

A. Motivation

This paper considers a fundamental problem in sparse estimation known as the *support recovery* problem. The goal in support recovery (also called *model selection* or *feature selection*) is to identify the set of locations corresponding to nonzero components of a sparse high-dimensional vector from a collection of linear measurements that may be corrupted by some form of measurement uncertainty. This task arises in a wide variety of application domains, including subset selection in linear regression [1], multiple hypothesis testing [2], signal denoising [3], and the emerging field of *compressive sensing* (CS) [4], [5] (see also the tutorial articles [6]–[8] and the references therein).

In so-called underdetermined settings—which are the primary focus in CS—the number of measurements obtained is much less than the ambient dimension of the unknown vector, which is assumed to be sparse. Formally, let $x \in \mathbb{R}^n$ denote our signal of interest, and suppose that only k of its entries are nonzero. Suppose that we are able to obtain a total of m measurements ($k \leq m \leq n$) by nominally observing x through the action of an $m \times n$ matrix A that we may design and specify. Motivated by uncertainties present in practical systems, it is typically assumed that these linear measurements are corrupted by some form of uncertainty, or "noise." A common choice in the literature is to model the uncertainty by zero-mean additive white Gaussian noise, giving rise to an observation model of the form $y = Ax + w$, where $w \sim \mathcal{N}(0, \sigma^2 I_{m\times m})$; several existing works have examined support recovery procedures in such settings [9]–[15].

However, other forms of *non-Gaussian* measurement uncertainties may also be present in scenarios of practical interest. For example, constraints on measurement precision (which could be due to acquisition hardware limitations) may be modeled as measurement quantization. Several existing works have examined the general effects of quantization in CS [16]–[20], some focusing exclusively on the case where the measurements are highly quantized to a single bit [21]–[23]. Other non-Gaussian measurement uncertainty models may include corruption by a moderate number of large-valued "outliers." Robust sparse recovery in the presence of outliers was examined in [24]–[27].

The problem of support recovery from compressive measurements corrupted by non-Gaussian uncertainties has received little attention to date in the CS literature. Indeed, the majority of the work examining CS under non-Gaussian corruption models has focused primarily on *estimation* of the unknown signal x, treating the non-Gaussian uncertainty as noise [16], explicitly accounting for it with additional sampling or algorithmic constraints [18]–[22], [24]–[27], or explicitly designing it in order to yield optimal estimation error performance using a particular recovery technique [17]. One exception is the work in [23], which considered the problem of support recovery from measurements corrupted by Gaussian noise and subsequently quantized to a single bit.

Here, we address the problem of support recovery in CS when the measurements are corrupted by possibly unbounded outliers in addition to additive white Gaussian noise, and even subsequently quantized, perhaps to a single bit. Our main results here establish that a computationally efficient procedure, which we call *Sign-Sketch*, is a provably robust approach for support recovery in these settings. We begin with a formal specification of the problem of interest.

B. Problem Specification

Let $x \in \mathbb{R}^n$ denote our object of interest, which we will assume to be fixed but unknown. The support of x, denoted

This work was supported by the DARPA Analog-to-Information (A2I) Program, DARPA/ONR Grant N66001-10-1-4090, AFOSR Grant FA9550-09-1-0432, and NSF Grant CCF-0926127.

$\mathcal{S} = \mathcal{S}(x)$, is defined to be the set of locations where x is nonzero: $\mathcal{S} := \{i \in \{1, 2, \ldots, n\} \ : \ x_i \neq 0\}$. We will say that x is k-sparse when $|\mathcal{S}| = k$. We acquire m measurements of x, nominally of the form Ax, where A is an $m \times n$ matrix that we may design. We examine two scenarios, corresponding to different forms of measurement uncertainty:

1) *Corruption by Gaussian Noise and Outliers*
   Measurements are of the form
$$y = Ax + w + o, \tag{1}$$
   where $w \sim \mathcal{N}(0, \sigma^2 I_m)$ and $o \in \mathbb{R}^m$ is a sparse vector of outliers whose nonzero entries take unspecified (and possibly large or unbounded) values.
2) *Highly Quantized Noisy Measurements*
   Measurements are of the form
$$y = \text{sgn}(Ax + w + o), \tag{2}$$
   where w and o are as above, and $\text{sgn}(\cdot)$ denotes the entry-wise application of the scalar sign function which, for $z \in \mathbb{R}$, is given by
$$\text{sgn}(z) := \begin{cases} -1, & z < 0 \\ 0, & z = 0 \\ +1, & z > 0 \end{cases} .$$

Our goal in each case will be to obtain an estimate $\widehat{\mathcal{S}} = \widehat{\mathcal{S}}(y, A)$ (which is a function of the measurements y and the measurement matrix A) that is an accurate estimate of the true unknown signal support $\mathcal{S}$.

C. Our Contributions

The main contributions of this paper come in the form of conditions under which accurate support recovery can be achieved in each of the two aforementioned scenarios. Our analysis is constructive, pertaining specifically to cases where measurements are obtained using a certain class of structured random matrices. We perform support estimation using a simple procedure that we call *Sign-Sketch*, which traces its origins to hashing techniques employed in the sketching literature, where similar constructions have been utilized to obtain approximations of sparse signals from uncorrupted measurements [28]. The Sign-Sketch procedure inherits some of the virtues of existing sketch-based approximation techniques, including low sample complexity and low computational complexity. In particular, our main results establish that the support of a k-sparse vector $x \in \mathbb{R}^n$ can be exactly recovered using the Sign-Sketch procedure (i.e., $\widehat{\mathcal{S}}(y, A) = \mathcal{S}(x)$) with high probability using only $O(\max\{k \log(n-k), k \log k\})$ measurements[1] under each of the two measurement uncertainty models described above, and the total number of operations required by the procedure for both measurement and inference is at most $O(n \log n)$.

[1]Here and throughout we will employ the standard asymptotic notation: we say $f(x) = O(g(x))$ when $\exists c_0 > 0$ and x_0 such that $\forall x \geq x_0$, $f(x) \leq c_0 g(x)$. We say $f(x) = \Omega(g(x))$ when $\exists c_1 > 0$ and x_1 such that $\forall x \geq x_1$, $f(x) \geq c_1 g(x)$. Finally, we write $f(x) = \Theta(g(x))$ if $f(x) = O(g(x))$ and $f(x) = \Omega(g(x))$.

The results established here also address a previously open question posed in [23]. That work considered the problem of support recovery from measurements corrupted by additive Gaussian noise and quantized to a single bit and proposed a procedure based on non-adaptive measurements requiring $O(Dk \log n)$ measurements for exact recovery, where $D := \max_{i,j \in \mathcal{S}} |x_i|/|x_j|$ is the dynamic range of the signal. The dependence on dynamic range was eliminated via an adaptive measurement procedure, prompting a question as to whether there exists a non-adaptive measurement procedure that can achieve the information-theoretically optimal sample complexity for support recovery from 1-bit measurements. The Sign-Sketch procedure proposed here provides an affirmative answer to that question.

D. Outline

The remainder of the paper is organized as follows. A brief review of the Count-Sketch technique for sparse estimation [29], which will be pertinent to our approach here, appears in Section II. We describe the Sign-Sketch procedure and state our main results in Section III. Section IV provides a discussion of our contributions in the broader context of existing work on support recovery and sparse inference. Finally, a discussion of the results is provided along with some possible extensions in Section V. Proofs of the main results are provided in the Appendix.

II. Review: Sketching for Sparse Estimation

The measurement matrices that we will employ here are *structured random matrices*, similar to those utilized by the Count-Sketch procedure proposed in [29]. The distributions that generate these matrices are parameterized by values $R, T \in \mathbb{N}$ and $\alpha > 0$, which may be specified, and (implicitly) by the ambient dimension n. Formally, we will denote by $\mathcal{A}(R, T, n, \alpha)$ a particular distribution over matrices having RT rows and n columns (to be described below), and we will assume that measurement matrices are drawn from this distribution.

Matrices $A \sim \mathcal{A}(R, T, n, \alpha)$ are composed of the vertical concatenation of T individual random matrices, denoted A_t for $t = 1, \ldots, T$, each having R rows and n columns. Each A_t is a sparse matrix containing exactly n nonzero values—one per column—where the location of the nonzero component in each column is chosen uniformly at random (with replacement) from the set $\{1, 2, \ldots, R\}$, and the nonzero component takes the value $\pm\alpha$ with probability $1/2$. For a given realization, we let $h_{t,1}, \ldots, h_{t,n} \in \{1, \ldots, R\}$ denote which entry of the corresponding column of A_t is nonzero, and we let $s_{t,1}, \ldots, s_{t,n} \in \{-\alpha, +\alpha\}$ be the corresponding values, for $t = 1, \ldots, T$. We note that when constructing each A_t, all random quantities are assumed independent.

Suppose that measurements are obtained according to the noise-free model $y = Ax$, where $A \sim \mathcal{A}(R, T, n, 1)$. For $t = 1, \ldots, T$, let y_t denote the subvector of y corresponding to observations obtained via the submatrix A_t—that is, $y_t = A_t x$.

The Count-Sketch procedure is implemented as follows. Form the "estimates" $\widetilde{x}_t \in \mathbb{R}^n$, where $\widetilde{x}_{t,i} = s_{t,i}\ y_{t,h_{t,i}}$, and then form a signal estimate $\widehat{x}$ whose entries are given by $\widehat{x}_i = \text{median}\,\{\widetilde{x}_{t,i}\}_{t=1}^T$. In other words, each entry of the signal estimate is obtained as the median of the corresponding entries of the "estimates" $\widetilde{x}_t$, each of which is formed simply by indexing and scaling the entries of the corresponding observations y_t. As shown in [29], one may choose $R = O(k)$ and $T = O(\log n)$ (for a total of $O(k \log n)$ measurements), and the estimate $\widehat{x}$ formed as above satisfies $\|\widehat{x} - x\|_2^2 \leq C\|x - x^k\|_2^2$ with high probability, where $C > 0$ is a specified constant and x^k denotes the best k-term approximation to x.

The rationale for forming the signal estimate using the median (instead of the mean, for example) becomes clear upon consideration of a given $\widetilde{x}_t$. For the sake of illustration, consider the case where x is k-sparse. For a given location $i \in \mathcal{S}$, notice that $\widetilde{x}_{t,i}$ will be exactly equal to x_i whenever $h_{t,i}$ is distinct from $h_{t,j}$, for all $j \in \mathcal{S} \setminus i$. To determine the probability of this event, consider a particular t and condition on the realization of the submatrix of A_t formed by the columns indexed by elements of the set $\mathcal{S} \setminus i$ (denoted here by $A_{t,\mathcal{S}\setminus i}$). For any such realization of $A_{t,\mathcal{S}\setminus i}$, there are at most $R - (k-1)$ allowable choices for the location of the nonzero entry of $A_{t,i}$ out of a total of R which will ensure $\widetilde{x}_{t,i} = x_i$, and these choices are equally likely since the locations of the nonzeros are drawn uniformly at random. Further, all realizations of $A_{t,\mathcal{S}\setminus i}$ are also equally likely by construction, and so we have for a given i, that

$$\Pr(\widetilde{x}_{t,i} = x_i) \geq \frac{R - (k-1)}{R},$$

implying, for example, that $\Pr(\widetilde{x}_{t,i} = x_i) \geq 2/3$ whenever $R \geq 3k$. A similar analysis establishes that, given this choice of R, each entry of $\widetilde{x}_t$ indexed by an element in the set $\mathcal{S}^c = \{1, 2, \ldots, n\} \setminus \mathcal{S}$ is equal to zero with probability at least $2/3$. The final result follows from applying the union bound to ensure that every entry of $\widehat{x}$ is equal to the corresponding entry of x, leading to the a requirement that $T = \Omega(\log n)$.

III. The Sign-Sketch Procedure

The approach we propose here is a variation of the Count-Sketch procedure. Instead of estimating the unknown signal entries using the median of the $\widetilde{x}_{t,i}$, our procedure leverages the observation that for the two uncertainty models described above, the majority of the $\{\widetilde{x}_{t,i}\}_{t=1}^T$ may have the same *sign* as x_i for indices $i \in \mathcal{S}$, and their signs will otherwise be equally likely for $i \in \mathcal{S}^c$. Our procedure, which we call *Sign-Sketch*, entails first forming the estimate

$$\widehat{x} = \frac{1}{T} \sum_{t=1}^{T} \text{sgn}(\widetilde{x}_t), \tag{3}$$

where the $\widetilde{x}$ are as above, and then obtaining a support estimate from it, according to

$$\widehat{\mathcal{S}} = \{i \in \{1, 2, \ldots, n\} : |\widehat{x}_i| > \tau\}, \tag{4}$$

for a specified threshold τ. The main results below describe the performance of this approach for support recovery for the two uncertainty models we consider.

A. Main Results

Our first result ensures that accurate support recovery is possible in the case where observations are corrupted by Gaussian noise and outliers. The proof is provided in the appendix.

Theorem 1: Suppose that measurements of x are obtained according to the model

$$y = Ax + w + o,$$

where $A \sim \mathcal{A}(R, T, n, \alpha)$ for some specified $R, T \in \mathbb{N}$ and $\alpha > 0$, $w \sim \mathcal{N}(0, \sigma^2 I_m)$, and $o \in \mathbb{R}^m$ is a vector of outliers whose entries take some unspecified (and possibly large) value independently with probability q, and 0 otherwise.

Let the number of nonzero entries of x be k, and let $x_{\min}$ denote the minimum amplitude of the nonzero components of x (i.e., $x_{\min} = \min_{i \in \mathcal{S}} |x_i|$). Define the quantity

$$\widetilde{p} := \frac{(k-1)}{R} + \frac{1}{2}\exp\left(-\frac{\alpha^2 x_{\min}^2}{2\sigma^2}\right) + q.$$

If the following are true:

- $\widetilde{p} < 1/2$,
- *τ is chosen so that $0 < \tau < 1 - 2\widetilde{p}$,*
- *T satisfies*

$$T \geq \max\left\{ \frac{2}{(\tau - (1 - 2\widetilde{p}))^2} \log\left(4k(n-k)^{\lambda}\right), \frac{2}{\tau^2} \log\left(4(n-k)^{\lambda+1}\right) \right\}$$

for any $\lambda > 0$,

then the estimate $\widehat{\mathcal{S}}$ formed according to (4) *satisfies* $\Pr\left(\widehat{\mathcal{S}} \neq \mathcal{S}\right) \leq (n-k)^{-\lambda}$.

A few salient points are immediately evident upon examination of this result. First, the requirement that $\widetilde{p}$ be less than $1/2$ implies that the following are strictly necessary conditions: i) the number of rows in each matrix A_t must satisfy $R > 2(k-1)$; ii) the minimum signal amplitude $x_{\min}$ must be $\Omega(\sigma/\alpha)$; and iii) the probability of outliers must satisfy $q < 1/2$. Second, the region of parameter values for which the procedure succeeds at recovering the support allows significant flexibility to adjust certain parameters to offset the effects of others. For example, provided that the parameters remain within the allowable ranges (so that $\widetilde{p}$ remains less than $1/2$), doubling R, and thus the total number of measurements, offsets the effect of a doubling of the outlier probability q, consistent with intuition. Also, increasing R can permit recovery of signals with weaker features (i.e., smaller values of $x_{\min}$), and so on. Finally, note that for a given q and

$x_{\min}$ such that

$$\frac{1}{2}\exp\left(-\frac{\alpha^2 x_{\min}^2}{2\sigma^2}\right) + q \leq \varrho < 1/2,$$

the procedure will succeed provided $R > (k-1)/(1/2 - \varrho)$, implying the total number of measurements $RT = O(\max\{k\log(n-k), k\log k\})$.

Our second result, stated below as a corollary, demonstrates the effectiveness of the proposed procedure in the case where measurements are corrupted by Gaussian noise and outliers and then quantized to a single bit.

Corollary 1: The results of Theorem 1 hold also in the case where measurements are quantized to a single bit; i.e., the observations are given by $y = \mathrm{sgn}(Ax + w + o)$.

Note that the Sign-Sketch procedure, in a sense, derives its robustness properties from the fact that it utilizes only the sign of each measurement. Thus, the results of Theorem 1 also hold in the case where the measurements are comprised of sign information only, a result that follows almost immediately from the proof of Theorem 1. In fact, the same results are valid in the case where y is quantized—perhaps randomly or nonuniformly—*to any number of levels*, provided only that the quantization retains the sign information of the underlying observations.

B. Computational Complexity Analysis

We briefly comment on the computational complexity of the Sign-Sketch procedure. First, for each t, the matrix A_t has only n nonzero entries, which implies that the corresponding observations can be formed using $O(n)$ operations. Accounting for each of the T observation steps, we conclude that the acquisition process can be performed using $O(\max\{n\log(n-k), n\log k\})$ operations.

Likewise, each "estimate" $\widetilde{x}_t$ can be formed using $O(n)$ operations (each entry must be multiplied by its unique signed amplitude value $s_{t,i}$), and there are a total of T such estimates. Finally, the thresholding step requires $O(n)$ operations, one per component. Thus, the overall computational complexity of the recovery procedure is also $O(\max\{n\log(n-k), n\log k\})$. Taken together, and considering the worst-case for the maximum (which occurs when $k = \Theta(n)$), we have that the overall computational complexity of the proposed procedure is at most $O(n\log n)$ operations for both acquisition and recovery.

IV. Connections with Existing Work

In addition to the connection with existing work in support recovery mentioned above [9]–[15], [23], the results established here also compliment a growing body of work in sparse recovery using sparse measurement matrices. A comprehensive overview of recent work in this field is provided in [30]; here we provide a brief (and necessarily incomplete) list of a few related works.

As discussed above, our approach traces its origins to the sketching literature, where similar approaches have been utilized to estimate sparse or nearly sparse vectors in noise-free settings. The Count-Sketch procedure was proposed in [29]; a related effort obtained estimation error bounds for a similar procedure using structured random measurement matrices similar to those employed here, but whose entries are either 0 or 1 [31]. Sparse recovery in noise-free settings using structured random binary matrices was also examined in [32], and an extension of that work to the recovery of approximately sparse vectors was proposed in [33].

A number of recent works have examined sparse recovery using measurement matrices formed as the adjacency matrices of certain expander graphs. The performance of convex optimizations for sparse estimation in this setting were examined in [34]. Greedy and other ad-hoc procedures for signal estimation in such settings were examined in [35]–[38].

The information-theoretic limits of support recovery using any random measurement matrices with i.i.d. entries was examined in [13]. There it was shown that sparsifying the measurement matrix can only increase the number of measurements needed for support recovery. Interestingly, while the results obtained here are not directly comparable with the results in [13] (since our approach utilizes structured random matrices, whose entries are not all statistically independent), the sample complexity of our approach in the case where measurements are corrupted only by Gaussian noise matches the information-theoretically optimal sample complexity in two of the settings examined there. Namely, upon renormalizing, the results in [13] show that in a regime where $x_{\min} = \Theta(1)$, the number of measurements necessary for support recovery by any procedure whatsoever (including perhaps combinatorial optimizations) is $\Omega(k\log(n-k))$ when $k = o(n)$ and $\Omega(n\log n)$ when $k = \Theta(n)$. This matches the sufficient conditions identified for the Sign-Sketch procedure in both cases.

V. Discussion

Note that although the Sign-Sketch procedure was specified under the assumption that the nonzero entries of x were real-valued, the approach outlined here could also be extended to recover sparse $x \in \mathbb{C}^n$. One way to do this would be to replace the estimate $\widehat{x}$ in (3) with the two estimates $\widehat{x}^{\mathrm{re}} = \sum_{t=1}^{T}\mathrm{sgn}(\mathrm{Re}(\widetilde{x}_t))$ and $\widehat{x}^{\mathrm{im}} = \sum_{t=1}^{T}\mathrm{sgn}(\mathrm{Im}(\widetilde{x}_t))$ corresponding to counting the signs of the real and imaginary parts of $\widetilde{x}_t$, respectively. The support estimate could then be formed by

$$\widehat{\mathcal{S}} = \left\{i \in \{1,2,\ldots,n\} : \max\left\{|\widehat{x}_i^{\mathrm{re}}|, |\widehat{x}_i^{\mathrm{im}}|\right\} > \tau\right\},$$

for an appropriately chosen τ. The results would qualitatively be the same as those obtained here—the main difference would come in an increase in some of the constants to account for additional union bounding. It follows that the Sign-Sketch procedure may also be used to recover the support of any vector x having sparse representation (with possibly complex coefficients) in some orthonormal basis.

We also mention that there is nothing inherently unique about the particular noise models examined here. Generally speaking, this approach could be easily generalized to other noise specifications, including heavy-tailed additive noise (in which case the sufficient conditions on $x_{\min}$ would change according to the particular noise distribution), as well as other non-additive noise models, and even to the "missing data" case where some fraction of the observations are randomly deleted. Expanding on this last point, we note that the net effect of missing data will manifest itself in a manner analogous to that of the outliers noise model considered here. Thus, drawing upon the intuition provided by the results here we conclude that a sufficient condition for accurate support recovery in the missing data scenario is that the probability of a given measurement being deleted does not exceed $1/2$. And, as above, the "erasure" probability may also be traded-off against the other problem parameters provided they all lie within the allowable ranges.

Finally, it is worth commenting that while our approach utilizes a particular form of structured random matrix, the density of nonzero entries in the sensing matrices A constructed as described above is $1/R$. This suggests that a (slightly modified) procedure using matrices A whose entries are generated independently according to

$$A_{i,j} \overset{iid}{\sim} \begin{cases} +\alpha & \text{w.p. } \frac{1}{2R} \\ 0 & \text{w.p. } 1-\frac{1}{R} \\ -\alpha & \text{w.p. } \frac{1}{2R} \end{cases}$$

may exhibit similar performance for the support recovery task. If this were the case, then the information-theoretic optimality of such an approach could be obtained by direct comparison with the results in [13]. A formal treatment of this idea will be left for future work.

Appendix

A. Proof of Theorem 1

Recall that our observations of x are of the form $y = Ax + w + o$, where $A \sim \mathcal{A}(R,T,n,\alpha)$ for some specified $R, T \in \mathbb{N}$ and $\alpha > 0$, $w \sim \mathcal{N}(0, \sigma^2 I_m)$, and $o \in \mathbb{R}^m$ is a vector of outliers whose entries take some unspecified (and possibly large) value independently with probability q, and 0 otherwise. The support estimate (4) is formed by thresholding the entries of (3). Assume that x has no more than k nonzero entries and that the minimum amplitude of the nonzero entries of x is denoted by $x_{\min}$.

Our analysis proceeds by analyzing the entries of any particular "estimate" $\widetilde{x}_t$. Recall from Section II that for any $i = 1, 2, \ldots, n$, we have $\widetilde{x}_{t,i} = s_{t,i}\ y_{t,h_{t,i}}$, where $s_{t,i}$ and $h_{t,i}$ are chosen independently, uniformly at random from the sets $\{-\alpha, +\alpha\}$ and $\{1, 2, \ldots, R\}$, respectively. Under this observation model, this implies that

$$\widetilde{x}_{t,i} = s_{t,i}\ \left((A_t x)_{h_{t,i}} + w_{h_{t,i}} + o_{h_{t,i}}\right),$$

where $(A_t x)_{h_{t,i}}$ denotes the $h_{t,i}$-th entry of the $R \times 1$ vector $A_t x$. Our first step will be to determine conditions under which the sign of $\widetilde{x}_{t,i}$ matches that of x_i for $i \in \mathcal{S}$. We will obtain a probabilistic statement of this result by iterated conditioning.

Fix any $i \in \mathcal{S}$, and condition on the submatrix of A_t formed by the columns indexed by elements of the set $\mathcal{S} \setminus i$ (denoted $A_{t,\mathcal{S}\setminus i}$). In this event, following the reasoning in Section II, we have that $\widetilde{x}_{t,i} = \alpha^2 x_i + s_{t,i}\ \left(w_{h_{t,i}} + o_{h_{t,i}}\right)$ except for an event of probability no greater $(k-1)/R$. Now, conditionally on this event, we have that $\widetilde{x}_{t,i} = \alpha^2 x_i + s_{t,i}\ w_{h_{t,i}}$ except for an event of probability q, which follows from the assumption on the distribution of the outliers. Finally, we may condition on this last event to determine that $\text{sgn}(\widetilde{x}_{t,i}) = \text{sgn}(x_i)$ except for an event of probability $(1/2)\exp\left(-\frac{\alpha^2 x_{\min}}{2\sigma^2}\right)$. This follows from the use of a standard Gaussian tail bound and the fact that the random quantity $s_{t,i}\ w_{h_{t,i}}$ is distributed as $\mathcal{N}(0, \alpha^2\sigma^2)$, which is easy to verify by computing its moment generating function. Putting the results together, and using the fact that each realization of $A_{t,\mathcal{S}\setminus i}$ is equally likely, we have that $P\left(\text{sgn}(\widetilde{x}_{t,i}) \neq \text{sgn}(x_i)\right) \leq \widetilde{p}$, where

$$\widetilde{p} := \frac{1}{2}\exp\left(-\frac{\alpha^2 x_{\min}}{2\sigma^2}\right) + q + \frac{(k-1)}{R}.$$

This bound holds for each $i \in \mathcal{S}$.

Next we consider the entries of $\widetilde{x}_{t,i}$ for $i \notin \mathcal{S}$. In this case it is easy to see that

$$P\left(\ \text{sgn}(\widetilde{x}_{t,i}) = +1\ \right) = P\left(\ \text{sgn}(\widetilde{x}_{t,i}) = -1\ \right) = 1/2.$$

This follows directly from the fact that the distribution of $\widetilde{x}_{t,i} = s_{t,i}\ \left((A_t x)_{h_{t,i}} + w_{h_{t,i}} + o_{h_{t,i}}\right)$ is symmetric and has zero probability mass at zero. The last point is a consequence of the presence of the Gaussian term in $\left((A_t x)_{h_{t,i}} + w_{h_{t,i}} + o_{h_{t,i}}\right)$, implying that this quantity is nonzero with probability 1.

Now, we turn our attention to the entries of $\widehat{x} = \frac{1}{T}\sum_{t=1}^{T} \text{sgn}(\widetilde{x}_t)$. Note that for each $i \in \mathcal{S}$, we have $|\mathbb{E}[\widehat{x}_i]\,| \geq 1 - 2\widetilde{p}$. Select a threshold τ that satisfies $0 < \tau < 1 - 2\widetilde{p}$. It is clear by inspection that $|\widehat{x}_i| > \tau$ whenever the condition $|\widehat{x}_i - \mathbb{E}[\widehat{x}_i]\,| < |\tau - (|\mathbb{E}[\widehat{x}_i]\,|)|$ holds. Using Hoeffding's inequality, we have

$$\begin{aligned} &P\left(|\widehat{x}_i - \mathbb{E}[\widehat{x}_i]\,| > |\tau - (|\mathbb{E}[\widehat{x}_i]\,|)|\right) \\ &\qquad \leq \quad 2\exp\left(-\frac{T(\tau - |\mathbb{E}[\widehat{x}_i]\,|)^2}{2}\right). \end{aligned}$$

Applying a union bound, we have that $|\widehat{x}_i| > \tau$ for all $i \in \mathcal{S}$ except for an event of probability

$$\sum_{i \in \mathcal{S}} 2\exp\left(-\frac{T(\tau - |\mathbb{E}[\widehat{x}_i]\,|)^2}{2}\right),$$

which is no greater than

$$2k\exp\left(-\frac{T(\tau - (1-2\widetilde{p}))^2}{2}\right). \tag{5}$$

For the entries of $\widehat{x}$ corresponding to locations where no signal component is present, we can again apply Hoeffding's inequality, followed by a union bound, to establish that $|\widehat{x}_i| < \tau$ for all indices $i \in \mathcal{S}^c$, except in an event of probability

$$2(n-k)\exp\left(-\frac{T\tau^2}{2}\right). \tag{6}$$

Now, note that choosing $T > 2\log\left(4(n-k)^{\lambda+1}\right)/\tau^2$ for any $\lambda > 0$ ensures that (6) does not exceed $(n-k)^{-\lambda}/2$. Similarly, we have that (5) does not exceed $(n-k)^{-\lambda}/2$

whenever $T \geq 2\log\left(4k(n-k)^\lambda\right)/(\tau-(1-2\widetilde{p}))^2$. It follows that the proposed procedure results in correct support recovery (ie, $|\widehat{x}_i| > \tau$ for all $i \in \mathcal{S}$, and $|\widehat{x}_i| < \tau$ for all $i \in \mathcal{S}^c$) provided

$$T \geq \max\left\{\frac{2}{(\tau-(1-2\widetilde{p}))^2}\log\left(4k(n-k)^\lambda\right), \frac{2}{\tau^2}\log\left(4(n-k)^{\lambda+1}\right)\right\},$$

as claimed.

References

[1] A. J. Miller, *Subset Selection in Regression*, 2nd ed. New York: Chapman & Hall/CRC, 2002.

[2] Y. Ingster and I. Suslina, *Nonparametric Goodness-of-Fit Testing under Gaussian Models*, ser. Lecture Notes in Statistics. Springer, 2003, vol. 169.

[3] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Academic Press, Dec 2008.

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

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

[6] E. Candes, "Compressive sampling," in *Proceedings of the International Congress of Mathematics*, Madrid, Spain, 2006, pp. 1433–1452.

[7] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, July 2007.

[8] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[9] P. Zhao and B. Yu, "On model selection consistency of lasso," *Journal of Machine Learning Research*, vol. 7, pp. 2541–2563, 2006.

[10] M. Wainwright, "Sharp thresholds for noisy and high-dimensional recovery of sparsity using ℓ_1-constrained quadratic programming (lasso)," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[11] ——, "Information-theoretic limitations on sparsity recovery in the high-dimensional and noisy setting," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5728–5741, December 2009.

[12] A. Fletcher, S. Rangan, and V. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," 2008, submitted.

[13] W. Wang, M. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2967–2979, June 2010.

[14] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," in *Proc. IEEE International Symposium on Information Theory*, July 2008.

[15] W. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," 2010, submitted.

[16] V. Goyal, A. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Processing Magazine*, vol. 25, pp. 48–56, March 2008.

[17] J. Sun and V. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE International Symposium on Information Theory*, 2009.

[18] W. Dai, H. Pham, and O. Milenkovic, "A comparative study of quantized compressive sensing schemes," in *Proc. IEEE International Symposium on Information Theory*, 2009.

[19] L. Jacques, D. Hammond, and M. Fadili, "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine," *IEEE Transactions on Information Theory*, vol. 56, no. 11, 2010.

[20] J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," 2010, submitted.

[21] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," in *Proc. 42nd Annual Conference on Information Sciences and Systems*, Princeton, NJ, March 2008.

[22] P. Boufounos, "Greedy sparse signal reconstruction from sign measurements," in *Proc. Asilomar Conference on Signals Systems and Computers*, Pacific Grove, CA, Nov 2009.

[23] A. Gupta, R. Nowak, and B. Recht, "Sample complexity for 1-bit compressed sensing and sparse classification," in *Proc. IEEE International Symposium on Information Theory*, 2010.

[24] R. Carrillo, K. Barner, and T. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE Journal of Special Topics in Signal Processing*, vol. 4, no. 2, pp. 392–408, April 2010.

[25] J. Wright and Y. Ma, "Dense error correction via ℓ^1-minimization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3540–3560, July 2010.

[26] J. Laska, M. Davenport, and R. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Proc. Asilomar Conference on Signals Systems and Computers*, Pacific Grove, CA, Nov 2009.

[27] G. Giannakis, G. Mateos, S. Farahmand, and H. Zhu, "USPACOR: Universal sparsity-controlling outlier rejection," manuscript.

[28] S. Muthukrishnan, "Data streams: Algorithms and applications," *Foundations and trends in theoretical computer science*, vol. 1, no. 2, 2005.

[29] M. Charikar, K. Chen, and M. Farach-Colton, "Finding frequent items in data streams," in *Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, 2002.

[30] P. Indyk and A. Gilbert, "Sparse recovery using sparse matrices," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 937–947, June 2010.

[31] G. Cormode and S. Muthukrishnan, "Improved data stream summaries: The count-min sketch and its applications," *Journal of Algorithms*, vol. 55, no. 1, pp. 58–75, 2005.

[32] S. Sarvotham, D. Baron, and R. Baraniuk, "Sudocodes – Fast measurement and reconstruction of sparse signals," in *Proc. International Symposium on Information Theory*, Seattle, WA, July 2006, pp. 2804–2808.

[33] ——, "Compressed sensing reconstruction via belief propagation," *Rice University Technical Report ECE-06-01*, 2006.

[34] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *Proc. 46th Annual Allerton Conf. on Communications, Control, and Computing*, 2008, pp. 798–805.

[35] S. Jafarpour, W. Xu, B. Hassibi, and A. R. Calderbank, "Efficient and robust compressed sensing using high-quality expander graphs," 2008, submitted.

[36] M. Khajehnejad, A. Dimakis, W. Xu, and B. Hassibi, "Sparse recovery of positive signals with minimal expansion," 2008, submitted.

[37] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the l1 norm," in *Proc. 49th Annual Symposium on Foundations of Computer Science*, 2008, pp. 199–207.

[38] R. Berinde, P. Indyk, and M. Ruzic, "Practical near-optimal sparse recovery in the l1 norm," in *Proc. 46th Annual Allerton Conference on Communications, Control, and Computing*, 2008, pp. 198–205.