Compressed Sensing with Very Sparse Gaussian Random Projections

Ping Li
Department of Statistics and Biostatistics
Department of Computer Science
Rutgers University
Piscataway, NJ 08854, USA
`pingli@stat.rutgers.edu`

Cun-Hui Zhang
Department of Statistics and Biostatistics
Rutgers University
Piscataway, NJ 08854, USA
`cunhui@stat.rutgers.edu`

Abstract

We study the use of *very sparse random projections* [12, 11] for compressed sensing (sparse signal recovery) when the nonzero coordinates of signals can be either positive or negative. In our setting, the entries of a Gaussian design matrix are randomly sparsified so that only a very small fraction of entries are nonzero. Our proposed decoding algorithm is simple and efficient in that the major cost is one linear scan of the coordinates. Using our proposed ***"tie estimator"***, we are able to recover a K-sparse signal of length N using $1.551eK \log K/\delta$ measurements (where $\delta \leq 0.05$ is the confidence) in one scan. The practical performance of our method, however, can be substantially better than this bound. The Gaussian design assumption is not essential although it simplifies the analysis.

Prior studies have shown that existing one-scan (or roughly one-scan) recovery algorithms using sparse matrices would require substantially (e.g., one order of magnitude) more measurements than L1 decoding by linear programming, when the nonzero coordinates of signals can be either negative or positive. In this paper, following a well-known experimental setup [1], we show that, at the same number of measurements, the recovery accuracies of our proposed method are similar to the standard L1 decoding.

1 Introduction

Compressed Sensing (CS) [8, 3] has become an important and popular topic in several fields, including Computer Science, Engineering, Applied Mathematics, and Statistics. The goal of compressed sensing is to recover a sparse signal $\mathbf{x} \in \mathbb{R}^{1 \times N}$ from a small number of non-adaptive linear measurements $\mathbf{y} = \mathbf{x}\mathbf{S}$, where $\mathbf{S} \in \mathbb{R}^{N \times M}$ is the "design" matrix (or "sensing" matrix). Typically, the signal $\mathbf{x}$ is assumed to be K-sparse (i.e., K nonzero entries) and neither the magnitudes nor locations of the nonzero coordinates are known. Many streaming/database applications can be naturally formulated as compressed sensing problems [4, 6, 16]. The idea of compressed sensing may be traced back to many prior papers, for example [9, 7].

In the literature of compressed sensing, entries of the design matrix $\mathbf{S}$ are often sampled i.i.d. from a Gaussian distribution (or Gaussian-like distribution, e.g., a distribution with a finite second moment). Well-known recovery algorithms are often based on linear programming (LP) (e.g., *basis pursuit* [5] or L1 decoding) or greedy methods such as orthogonal matching pursuit (OMP) [18, 15, 20, 19]. In general, L1 decoding is computationally expensive. OMP is often more efficient than L1 decoding but it can still be expensive especially when K is large.

1.1 Compressed Sensing with Very Sparse Random Projections

The process of collecting measurements, i.e., $\mathbf{y} = \mathbf{x}\mathbf{S}$, is often called "random projections". [12, 11] studied the idea of "very sparse random projections" by randomly sparsifying the sensing matrix $\mathbf{S}$ so that only a very small fraction of the entries can be nonzero. In this paper, we will continue to investigate the idea of very sparse random projections in the context of compressed sensing.

Our work is related to "sparse recovery with sparse matrices" [2, 10, 17], for example, the SMP (*Sparse Matching Pursuit*) algorithm [2]. There is a well-known wiki page [1], which summarizes the comparisons of L1 decoding with count-min sketch [6] and SMP. Their results have shown that, in order to achieve similar recovery accuracies, count-min sketch needs about 10 to 15 times more mea-

surements than L1 decoding and SMP needs about half of the measurements of count-min sketch.

In comparison, our experimental section (e.g., Figure 2) demonstrates that the proposed method can be as accurate as L1 decoding, at the same number of measurements. The major cost of our method is one linear scan of the coordinates, like count-min sketch.

Our work is also inspired by the prior work on *very sparse compressed counting* [14], which was designed for recovering nonnegative sparse signals, based on maximally-skewed stable random projections. The method proposed in [14] is much simpler than our proposed estimator in this paper. In addition, the use of heavy-tailed design makes the method in [14] very resilient to measurement noise. Nevertheless, the work of [14] is restricted to nonnegative signals.

1.2 Linear Measurements from Sparse Projections

In this paper, our procedure for compressed sensing first collects M non-adaptive linear measurements

$$y_j = \sum_{i=1}^{N} x_i [s_{ij} r_{ij}], \qquad j = 1, 2, ..., M \quad (1)$$

Here, s_{ij} is the (i, j)-th entry of the design matrix with $s_{ij} \sim N(0, 1)$ i.i.d. Instead of using a dense design matrix, we randomly sparsify $(1 - \gamma)$-fraction of the entries of the design matrix to be zero, i.e.,

$$r_{ij} = \begin{cases} 1 & \text{with prob. } \gamma \\ 0 & \text{with prob. } 1 - \gamma \end{cases} \quad i.i.d. \quad (2)$$

Note that any s_{ij} and r_{ij} are also independent.

Our proposed decoding scheme utilizes two simple estimators: (i) the *tie estimator* and (ii) the *absolute minimum estimator*. For convenience, we will theoretically analyze them separately. In practice, these two estimators should be combined to form a powerful decoding framework.

1.3 The Tie Estimator

The tie estimator is developed according to the following interesting observation on the *ratio statistics* $\frac{y_j}{s_{ij} r_{ij}}$. Conditional on $r_{ij} = 1$, we can write

$$\left. \frac{y_j}{s_{ij} r_{ij}} \right|_{r_{ij}=1} = \frac{\sum_{t=1}^{N} x_t s_{tj} r_{tj}}{s_{ij}} \quad (3)$$

$$= x_i + \frac{\sum_{t \neq i}^{N} x_t s_{tj} r_{tj}}{s_{ij}} = x_i + (\eta_{ij})^{1/2} \frac{S_2}{S_1}$$

where $S_1, S_2 \sim N(0, 1)$, i.i.d., and

$$\eta_{ij} = \sum_{t \neq i}^{N} |x_t r_{tj}|^2 = \sum_{t \neq i}^{N} |x_t|^2 r_{tj} \quad (4)$$

Note that η_{ij} has certain probability of being zero. If $\eta_{ij} = 0$, then $\left. \frac{y_j}{s_{ij} r_{ij}} \right|_{r_{ij}=1} = x_i$. Thus, given M measurements, if $\eta_{ij} = 0$ happens (at least) **twice** (i.e., a **tie** occurs), we can exactly identify the value x_i. This key observation motivates our proposal of the tie estimator.

Another key observation is that, if $x_i = 0$, then we will not see a nonzero tie (i.e., the probability of nonzero tie is 0). This is due to the fact that we use a Gaussian design matrix, which excludes unwanted ties. It is also clear that the Gaussian assumption is not needed, as long as s_{ij} follows a continuous distribution. In this paper we focus on Gaussian design because it simplifies some analysis.

To an extent, the tie estimator studied in this paper might be viewed as a special case of the *gap estimator* [13], i.e., the gap estimator with zero gap. The work of [13] was based on extremely heavy-tailed stable random projections. In our study, we focus on Gaussian (or Gaussian-like) designs which might be more commonly used.

1.4 The Absolute Minimum Estimator

It turns out that, if we just need to detect whether $x_i = 0$, the task is easier than estimating the value of x_i, for a particular coordinate i. Given M measurements, if $\eta_{ij} = 0$ happens (at least) **once**, we will be able to determine whether $x_i = 0$. Note that unlike the tie estimator, this estimator will generate "false positives". In other words, if we cannot be certain that $x_i = 0$, then it is still possible that $x_i = 0$ indeed.

From the practical perspective, at a particular coordinate i, it is preferable to first detect whether $x_i = 0$ because that would require fewer measurements than using the tie estimator. Later in the paper, we can see that the performance can be potentially further improved by a more general estimator, i.e., the so-called *absolute minimum estimator*:

$$\hat{x}_{i,min,\gamma} = z_{i,t}, \quad \text{where } t = \underset{1 \leq j \leq M}{\operatorname{argmin}} |z_{i,j}|, \quad z_{ij} = \frac{y_j}{s_{ij} r_{ij}} \quad (5)$$

We will also introduce a threshold ϵ and provide a theoretical analysis of the event $\hat{x}_{i,min,\gamma} \geq \epsilon$. When $\epsilon = 0$, it becomes the "zero-detection" algorithm. Our analysis will show that by using $\epsilon > 0$ we can exploit the prior knowledge we have about the signal and improve the accuracy.

1.5 Combining Two Estimators

For the convenience of theoretical analysis, we will separately analyze the tie estimator and the absolute minimum estimator. However, we recommend a mixed procedure. That is, we first run the absolute minimum estimator in one scan of the coordinates, $i = 1$ to N. Then we run the tie estimator only on those coordinates which are possibly not

zero. Recall that the absolute minimum estimator may generate false positives.

As an option, we can iterate this process for several rounds. After one iteration (i.e., the absolute minimum estimator followed by the tie estimator), there might be a set of coordinates for which we cannot decide their values. We can compute the residuals and use them as measurements for the next iteration. Typically, a few (e.g., 3 or 4) iterations are sufficient and the major computational cost is computing the absolute minimum estimator in the first iteration.

2 Analysis of the Absolute Minimum Estimator

The important task is to analyze the false positive probability: $\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0\right)$ for some chosen threshold $\epsilon > 0$. Later we will see that ϵ is irrelevant if we only care about the worst case.

Recall that, conditional on $r_{ij} = 1$, we can express $\frac{y_j}{s_{ij}r_{ij}} = x_i + (\eta_{ij})^{1/2}\frac{S_2}{S_1}$, where $S_1, S_2 \sim N(0,1)$ i.i.d. and η_{ij} is defined in (4). It is known that S_2/S_1 follows the standard Cauchy distribution. Therefore,

$$\mathbf{Pr}\left(\left|\frac{S_2}{S_1}\right| \le t\right) = \frac{2}{\pi}\tan^{-1}(t), \ t > 0 \qquad (6)$$

We are ready to present the Lemma about the false positive probability, including a practically useful data-dependent bound, as well as a data-independent bound (which is convenient for worst-case analysis). The proof heavily utilizes (6) and also follows some ideas in [14].

2.1 The False Positive Probability

Lemma 1 *Data-dependent bound:*

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0\right)$$

$$= \left[1 - \gamma E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M \qquad (7)$$

$$\le \left[1 - \gamma\left\{\frac{2}{\pi}\tan^{-1}\left\{\frac{\epsilon}{\sqrt{\gamma\sum_t x_t^2}}\right\}\right\}\right]^M \qquad (8)$$

Data-independent (worst case) bound:

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0\right) \le \left[1 - \gamma(1-\gamma)^K\right]^M \qquad (9)$$

Remark: The data-dependent bound (7) and (8) can be numerically evaluated if we have information about the data. The bound will help us understand why empirically the performance of our proposed algorithm is substantially better than the worst-case bound. On the other hand, the

worst case bound (9) is convenient for theoretical analysis. In fact, it directly leads to the $eK\log N$ complexity bound.

Proof of Lemma 1: For convenience, we define the set $T_i = \{j, 1 \le j \le M, r_{ij} = 1\}$.

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0\right)$$

$$= E\left(\mathbf{Pr}\left(\left|\frac{y_j}{s_{ij}}\right| > \epsilon, x_i = 0, j \in T_i | T_i\right)\right)$$

$$= E\prod_{j \in T_i}\left[\mathbf{Pr}\left(\left|\frac{S_2}{S_1}\right| > \frac{\epsilon}{\eta_{ij}^{1/2}}, x_i = 0\right)\right]$$

$$= E\prod_{j \in T_i}\left[1 - \frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right]$$

$$= E\left\{\left[1 - E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^{|T_i|}\right\}$$

$$= \left[1 - \gamma + \gamma\left\{1 - E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right\}\right]^M$$

$$= \left[1 - \gamma E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M$$

By noticing that $f(x) = \tan^{-1}\frac{a}{\sqrt{x}}$, (where $a > 0$), is a convex function of $x > 0$, we obtain an upper bound by Jensen's inequality.

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0\right)$$

$$= \left[1 - \gamma E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M$$

$$\le \left[1 - \gamma\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{(E\eta_{ij})^{1/2}}\right)\right\}\right]^M$$

$$= \left[1 - \gamma\left\{\frac{2}{\pi}\tan^{-1}\left\{\frac{\epsilon}{\left(\gamma\sum_{t \ne i}x_t^2\right)^{1/2}}\right\}\right\}\right]^M$$

$$= \left[1 - \gamma\left\{\frac{2}{\pi}\tan^{-1}\left\{\frac{\epsilon}{\sqrt{\gamma\sum_t x_t^2}}\right\}\right\}\right]^M$$

We can further obtain a worst case bound as follows. Note that η_{ij} has some mass at 0.

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0\right)$$

$$= \left[1 - \gamma E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M$$

$$\le \left[1 - \gamma\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{0}\right)\right\}\mathbf{Pr}\left(\eta_{ij} = 0\right)\right]^M$$

$$= \left[1 - \gamma(1-\gamma)^K\right]^M \qquad \square$$

2.2 The False Negative Probability

It is also necessary to control the false negative probability: $\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| \leq \epsilon,\ x_i \neq 0\right)$. See the proof in Appendix.

Lemma 2

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| \leq \epsilon,\ x_i \neq 0\right) = 1- \qquad (10)$$

$$\left[1 - \gamma E\left\{\frac{1}{\pi}\tan^{-1}\left(\frac{\epsilon + x_i}{\eta_{ij}^{1/2}}\right) - \frac{1}{\pi}\tan^{-1}\left(\frac{x_i - \epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M$$

$$\leq 1 - \left[1 - \frac{2}{\pi}\gamma\tan^{-1}\epsilon\right]^M \qquad (11)$$

Remark: Again, if we know information about the data, we might be able to numerically evaluate the exact false negative probability (10). The (loose) upper bound (11) is also insightful because it means this probability $\rightarrow 0$ if $\epsilon \rightarrow 0$. Note that in Lemma 1, the worst case bound is actually independent of ϵ. This implies that, if we only care about the worst case, we do not have to worry about the false positive probability since we can choose $\epsilon \rightarrow 0$.

2.3 The Worst Case Complexity Bound

From the worst-case false positive probability bound: $\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0\right) \leq \left[1 - \gamma\left(1-\gamma\right)^K\right]^M$, by choosing $\gamma = 1/K$ (and $\epsilon \rightarrow 0$), we can easily obtain the following Theorem regarding the sample complexity of only using the absolute minimum estimator, by solving for M from $N\left[1 - \gamma\left(1-\gamma\right)^K\right]^M \leq \delta$.

Theorem 1 *Using the absolute minimum estimator and $\gamma = 1/K$, for perfect support recovery (with probability $> 1 - \delta$), it suffices to use*

$$M \geq \frac{\log N/\delta}{\log \frac{1}{1 - \frac{1}{K}\left(1 - \frac{1}{K}\right)^K}} \qquad (12)$$

$$\approx eK\log N/\delta \qquad (13)$$

measurements.

Remark: The term $\frac{1}{K}/\log\frac{1}{1 - \frac{1}{K}\left(1 - \frac{1}{K}\right)^K}$ approaches $e = 2.7183...$ very quickly. For example, the difference is only 0.1 when $K = 10$.

3 Analysis of the Absolute Minimum Estimator on Ternary Signals

Although the complexity result in Theorem 1 can be theoretically exciting, we would like to better understand why empirically we only need substantially fewer measurements. In this section, for convenience, we consider the

special case of "ternary" signals, i.e., $x_i \in \{-1, 0, 1\}$. The exact expectation (7), i.e.,

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon,\ x_i = 0\right)$$

$$= \left[1 - \gamma E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M$$

which, in the case of ternary data, becomes

$$\eta_{ij} = \sum_{i=1}^N |x_t|^2 r_{tj} \sim Binomial(K, \gamma) \qquad (14)$$

For convenience, we write

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon,\ x_i = 0\right)$$

$$= \left[1 - \frac{1}{K}\left(\gamma K\right) E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M$$

$$= \left[1 - \frac{1}{K}H(\epsilon, K, \gamma)\right]^M \qquad (15)$$

where

$$H(\epsilon, K, \gamma) = (\gamma K) E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\sqrt{Z}}\right)\right\}, \qquad (16)$$

$$Z \sim Binomial(K, \gamma)$$

which can be easily computed numerically for given γ, K, and M. In order for $\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon,\ x_i = 0\right) \leq \delta$ for all i, we should have

$$M \geq \frac{K}{H(\epsilon, K, \gamma)}\log N/\delta \qquad (17)$$

It would be much more convenient if we do not have to worry about all combinations of γ and K. In fact, we can resort to the well-studied *poisson approximation* by considering $\lambda = \gamma K$ and defining $Z \sim Poisson(\lambda)$ and

$$h(\epsilon, \lambda)$$

$$= \lambda E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\sqrt{Z}}\right)\right\}, \quad Z \sim Poisson(\lambda) \qquad (18)$$

$$= \lambda\sum_{k=0}^\infty \left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\sqrt{k}}\right)\right\}\frac{e^{-\lambda}\lambda^k}{k!}$$

$$= \lambda e^{-\lambda} + \lambda e^{-\lambda}\sum_{k=1}^\infty \left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\sqrt{k}}\right)\right\}\frac{\lambda^k}{k!} \qquad (19)$$

Figure 1 plots $\frac{1}{H(\epsilon, K, \gamma)}$ and $\frac{1}{h(\epsilon, \lambda)}$ to confirm that the Poisson approximation is very accurate (as one would expect). At $\gamma = 1/K$ (i.e., $\lambda = 1$), the two terms $\frac{1}{H(\epsilon, K, \gamma)}$ and $\frac{1}{h(\epsilon, \lambda)}$ are upper bounded by e. However, when ϵ is not too small, the constant e can be conservative. Basically, the choice of ϵ reflects the level of prior information about the

signal. If the signals are significantly away from 0, then we can choose a larger ϵ and hence the algorithm would require less measurements. For example, if we know the signals are ternary, we can perhaps choose $\epsilon = 0.5$ or larger. Also, we can notice that $\gamma = 1/K$ is not necessarily the optimum choice for a given ϵ. In general, the performance is not too sensitive to the choice $\gamma = \lambda/K$ as long as ϵ is not too small and the λ is reasonably large. This might be good news for practitioners.
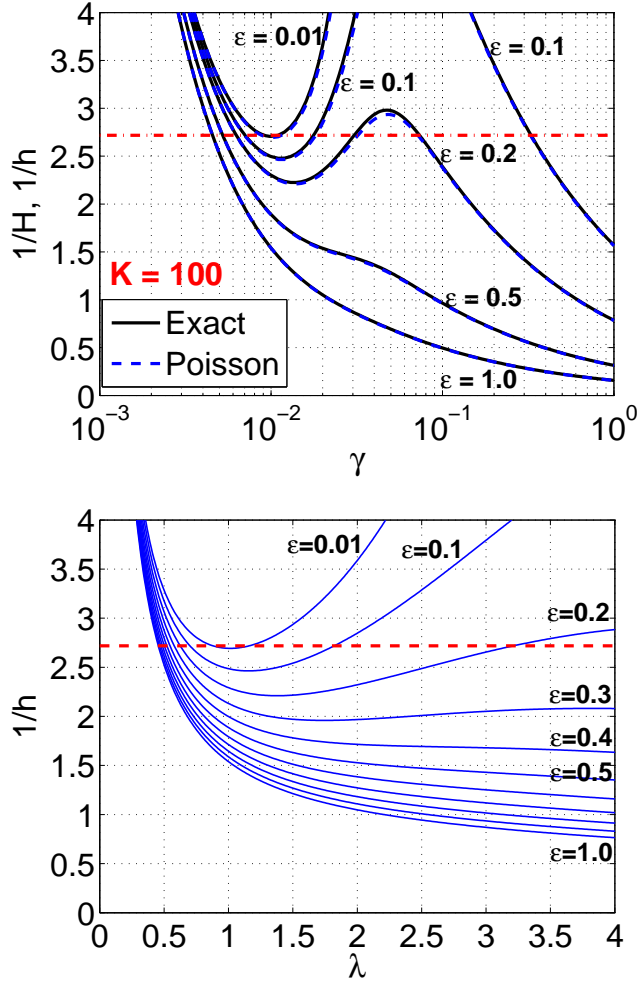

Figure 1: **Upper Panel**: $\frac{1}{H(\epsilon,K,\gamma)}$ (solid) and $\frac{1}{h(\epsilon,\lambda)}$ (dashed), for $K = 100$ and $\epsilon \in \{0.01, 0.1, 0.2, 0.5, 1.0\}$. This plot confirms that the Poisson approximation is indeed very accurate (as expected).
Bottom Panel: Poisson approximation $\frac{1}{h(\epsilon,\lambda)}$ for $\epsilon \in \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. In both panels, we use the horizontal line to indicate $e = 2.7183....$ When $\gamma = 1/K$, i.e., $\lambda = 1$, both $\frac{1}{H(\epsilon,K,\gamma)}$ and $\frac{1}{h(\epsilon,\lambda)}$ are upper bounded by e.

4 Analysis of the Absolute Minimum Estimator with Measurement Noise

We can also analyze the absolute minimum estimator when measurement noise is present, i.e.,

$$\tilde{y}_j = y_j + n_j = \sum_{i=1}^{N} x_i \left[s_{ij} r_{ij} \right] + n_j, \tag{20}$$

where $n_j \sim N(0, \sigma^2), \qquad j = 1, 2, ..., M$

Again, we compute the ratio statistic

$$\left. \frac{y_j + n_j}{s_{ij} r_{ij}} \right|_{r_{ij}=1} = \frac{\sum_{t=1}^{N} x_t s_{tj} r_{tj} + n_j}{s_{ij}}$$

$$= x_i + \frac{\sum_{t \neq i}^{N} x_t s_{tj} r_{tj} + n_j}{s_{ij}} = x_i + (\tilde{\eta}_{ij})^{1/2} \frac{S_2}{S_1} \tag{21}$$

where $S_1, S_2 \sim N(0, 1)$, i.i.d., and

$$\tilde{\eta}_{ij} = \sum_{t \neq i}^{N} |x_t r_{tj}|^2 + \sigma^2 = \sum_{t \neq i}^{N} |x_t|^2 r_{tj} + \sigma^2 \tag{22}$$

Lemma 3 *Data-dependent bound:*

$$\mathbf{Pr} \left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0 \right)$$

$$= \left[1 - \gamma E \left\{ \frac{2}{\pi} \tan^{-1} \left(\frac{\epsilon}{\tilde{\eta}_{ij}^{1/2}} \right) \right\} \right]^{M} \tag{23}$$

$$\leq \left[1 - \gamma \left\{ \frac{2}{\pi} \tan^{-1} \left\{ \frac{\epsilon}{(\sigma^2 + \gamma \sum_t x_t^2)^{1/2}} \right\} \right\} \right]^{M} \tag{24}$$

Data-independent bound:

$$\mathbf{Pr} \left(|\hat{x}_{i,min,\gamma}| > \epsilon, \ x_i = 0 \right)$$

$$\leq \left[1 - \gamma \left\{ \frac{2}{\pi} \tan^{-1} \left(\frac{\epsilon}{\sigma} \right) \right\} (1 - \gamma)^K \right]^{M} \tag{25}$$

Data-independent complexity bound: With $\gamma = 1/K$, in order to achieve $\mathbf{Pr} \left(|\hat{x}_{i,min,\gamma}| > \epsilon, \ x_i = 0 \right) \leq \delta$ for all i, it suffices to use

$$M \geq e \left\{ \frac{2}{\pi} \tan^{-1} \left(\frac{\epsilon}{\sigma} \right) \right\} K \log N/\delta \tag{26}$$

measurements.

Proof of Lemma 3:

$$\mathbf{Pr} \left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i = 0 \right)$$

$$= \left[1 - \gamma E \left\{ \frac{2}{\pi} \tan^{-1} \left(\frac{\epsilon}{\tilde{\eta}_{ij}^{1/2}} \right) \right\} \right]^{M}$$

$$\leq \left[1 - \gamma \left\{ \frac{2}{\pi} \tan^{-1} \left(\frac{\epsilon}{(E\tilde{\eta}_{ij})^{1/2}} \right) \right\} \right]^{M}$$

$$= \left[1 - \gamma \left\{ \frac{2}{\pi} \tan^{-1} \left\{ \frac{\epsilon}{(\sigma^2 + \gamma \sum_t x_t^2)^{1/2}} \right\} \right\} \right]^{M}$$

which is still expressed in terms of the summary of the signal. To obtain a data-independent bound, we have

$$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, \ x_i = 0\right)$$

$$= \left[1 - \gamma E\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\tilde{\eta}_{ij}^{1/2}}\right)\right\}\right]^M$$

$$\leq \left[1 - \gamma\left\{\frac{2}{\pi}\tan^{-1}\left(\frac{\epsilon}{\sigma}\right)\right\}(1-\gamma)^K\right]^M$$

$\square$

5 Analysis of the Tie Estimator

To construct the *tie estimator*, we first compute $z_{ij} = \frac{y_j}{s_{ij}r_{ij}}$ which is anyway needed for the absolute minimum estimator. At each i of interest, we sort those M z_{ij} values and examine the order statistics, $z_{i,(1)} \leq z_{i,(2)} \leq ... \leq z_{i,(M)}$, and their consecutive differences, $z_{i,(j+1)} - z_{i,(j)}$ for $j = 1, 2, ..., M-1$. Then we have

$$\hat{x}_{i,tie,\gamma} = z_{i,(j_i)}, \text{if } z_{i,(j_i+1)} - z_{i,(j_i)} = 0 \text{ and } |z_{i,(j_i)}| \neq \infty$$

Recall

$$\left.\frac{y_j}{s_{ij}r_{ij}}\right|_{r_{ij}=1} = \frac{\sum_{t=1}^N x_t s_{tj} r_{tj}}{s_{ij}} = x_i + \frac{\sum_{t\neq i}^N x_t s_{tj} r_{tj}}{s_{ij}}$$

$$= x_i + (\eta_{ij})^{1/2}\frac{S_2}{S_1}$$

where $S_1, S_2 \sim N(0,1)$, i.i.d., and $\eta_{ij} = \sum_{t\neq i}^N |x_t|^2 r_{tj}$, which has a certain probability of being zero. If $\eta_{ij} = 0$, then $\left.\frac{y_j}{s_{ij}r_{ij}}\right|_{r_{ij}=1} = x_i$. To reliably estimate the magnitude of x_i, we need $\eta_{ij} = 0$ to happen more than once, i.e., there should be a tie. Note that

$$\mathbf{Pr}\left(\eta_{ij} = 0, r_{ij} = 1\right) = \begin{cases} \gamma(1-\gamma)^K & \text{if } x_i = 0 \\ \gamma(1-\gamma)^{K-1} & \text{if } x_i \neq 0 \end{cases}$$

$$(27)$$

For a given nonzero coordinate i, we would like to have $\eta_{ij} = 0$ more than once among M measurements. This is a binomial problem, and the error probability is simply

$$\left[1 - \gamma(1-\gamma)^{K-1}\right]^M \qquad (28)$$

$$+ M\left(\gamma(1-\gamma)^{K-1}\right)\left[1 - \gamma(1-\gamma)^{K-1}\right]^{M-1}$$

Suppose we use $\gamma = 1/K$. To ensure this error is smaller than δ for all K nonzero coordinates, it suffices to choose M so that

$$K\left[1 - \gamma(1-\gamma)^{K-1}\right]^M \qquad (29)$$

$$+ KM\left(\gamma(1-\gamma)^{K-1}\right)\left[1 - \gamma(1-\gamma)^{K-1}\right]^{M-1} \leq \delta$$

It is easy to see that this choice of M suffices for recovering the entire signal, not just the nonzero entries. This is due to the nice property of the tie estimator, which has no false positives. That is, if there is a tie, we know for sure that it reveals the true value of the coordinate. For any zero coordinate, either there is no tie or the tie is zero. Therefore, it suffices to choose M to ensure all the nonzero coordinates are recovered.

Theorem 2 *Using the tie estimator and $\gamma = \frac{1}{K}$, for perfect signal recovery (with probability $> 1 - \delta$), it suffices to choose the number of measurements to be*

$$M \geq 1.551 eK \log K/\delta, \qquad \delta \leq 0.05 \qquad (30)$$

Proof of Theorem 2: The recovery task is trivial when $K = 1$. Consider $K \geq 2$ and $p = \frac{1}{K}\left(1 - \frac{1}{K}\right)^{K-1}$, i.e., $p \leq 1/4$. We need to choose M such that $K\left((1-p)^M + Mp(1-p)^{M-1}\right) \leq \delta$. Let M_1 be such that $K(1-p)^{M_1} = \delta$, i.e., $M_1 = \frac{\log \delta/K}{\log(1-p)} = \frac{\log K/\delta}{\log\frac{1}{1-p}}$. Suppose we choose $M = (1+\alpha)M_1$. Then.

$$K\left((1-p)^{(1+\alpha)M_1} + (1+\alpha)M_1 p(1-p)^{(1+\alpha)M_1-1}\right)$$

$$= \delta\left((\delta/K)^\alpha + (1+\alpha)\frac{\log K/\delta}{\log\frac{1}{1-p}}\frac{(\delta/K)^\alpha}{1-p}p\right)$$

Therefore, we need to find the α so that

$$T(\delta, K, \alpha) = (\delta/K)^\alpha + \frac{(1+\alpha)\log(K/\delta)(\delta/K)^\alpha}{\log(1-p)(1-1/p)} \leq 1$$

Since $p \leq 1/4$, we have $\frac{\partial}{\partial p}\log(1-p)(1-1/p) = (\log(1-p) + p)/p^2 < 0$. Because p is decreasing in K, we know that $\frac{1}{\log(1-p)(1-1/p)}$ is decreasing in K. Also, note that

$$\frac{\partial}{\partial K}\left[\log(K/\delta)(\delta/K)^\alpha\right] = (\delta/K)^\alpha/K\left(1 - \alpha\log K/\delta\right)$$

$$\frac{\partial}{\partial \delta}\left[\log(K/\delta)(\delta/K)^\alpha\right] = (\delta/K)^\alpha/\delta\left(-1 + \alpha\log K/\delta\right)$$

As we consider $K \geq 2$ and $\delta \leq 0.05$, we know that, as long as $\alpha \geq 1/\log\frac{2}{0.05} = 1/\log 40$, the term $\log K/\delta(\delta/K)^\alpha$ is increasing in δ and decreasing in K. Combining the calculations, we know that $T(\delta, K, \alpha)$ is decreasing in K and increasing in δ, for $\alpha > 1/\log 40$. It is thus suffices to consider $\delta = 0.05$ and $K = 2$. Because $T(0.05, 2, \alpha)$ is decreasing in α, we only need to numerically find the α so that $T(0.05, 2, \alpha) = 1$, which happens to be $0.5508...$

Therefore, it suffices to choose $M = 1.551 M_1 = 1.551\frac{\log K/\delta}{\log\frac{1}{1-\frac{1}{K}\left(1-\frac{1}{K}\right)^{K-1}}}$ measurements. It remains to show that $\frac{1}{K\log\frac{1}{1-\frac{1}{K}\left(1-\frac{1}{K}\right)^{K-1}}} \leq e$. Due to $\log\frac{1}{1-x} \geq x$,

$\forall\, 0 < x < 1$, we have

$$\frac{1}{K \log \frac{1}{1 - \frac{1}{K}\left(1 - \frac{1}{K}\right)^{K-1}}}$$

$$\leq \frac{1}{K} \frac{1}{\frac{1}{K}\left(1 - \frac{1}{K}\right)^{K-1}}$$

$$= \frac{1}{\left(1 - \frac{1}{K}\right)^{K-1}} = \left(1 + \frac{1}{K-1}\right)^{K-1} \leq e$$

$\square$

Readers probably have noticed that the tie estimator could be sensitive to measurement noises. It appears that certain important applications such as databases and networks mainly focus on the noiseless case. It is nevertheless an interesting problem to study estimators which are more robust to measurement noises. If possible, we can use the heavy-tailed design [13]. If we must use Gaussian (or Gaussian-like) designs, we might consider modifying the tie estimator by allowing some (nonzero) gaps. We leave it for future work to systematically analyze this more generalized estimator.

6   An Experimental Study

Compressed sensing is an important problem of broad interest, and it is crucial to experimentally verify that the proposed method performs well as predicted by our theoretical analysis. In this study, we closely follow the experimental setting as in the well-known wiki page (see [1]), which compared count-min sketch, SMP, and L1 decoding, on ternary (i.e., $\{-1, 0, 1\}$) signals. In particular, the results for $N = 20000$ are available for all three algorithms. Their results have shown that, in order to achieve similar recovery accuracies, count-min sketch needs about 10 to 15 times more measurements than L1 decoding and SMP only needs about half of the measurements of count-min sketch.

As shown in the success probability contour plot in Figure 2 (for $\gamma = 1/K$), the accuracy of our proposed method is similar to the accuracy of L1 decoding (based on [1]). This should be exciting because, at the same number of measurements, the decoding cost of our proposed algorithm is roughly the same as count-min sketch.

7   Conclusion

Compressed sensing has become a popular and important research topic. Using a sparse design matrix has a significant advantage over dense design. For example, in sensing networks, we can replace a dense constellation of sensors by a randomly sparsified one, which may result in substantially saving of sensing hardware and labor costs. In this paper, we show another advantage from the
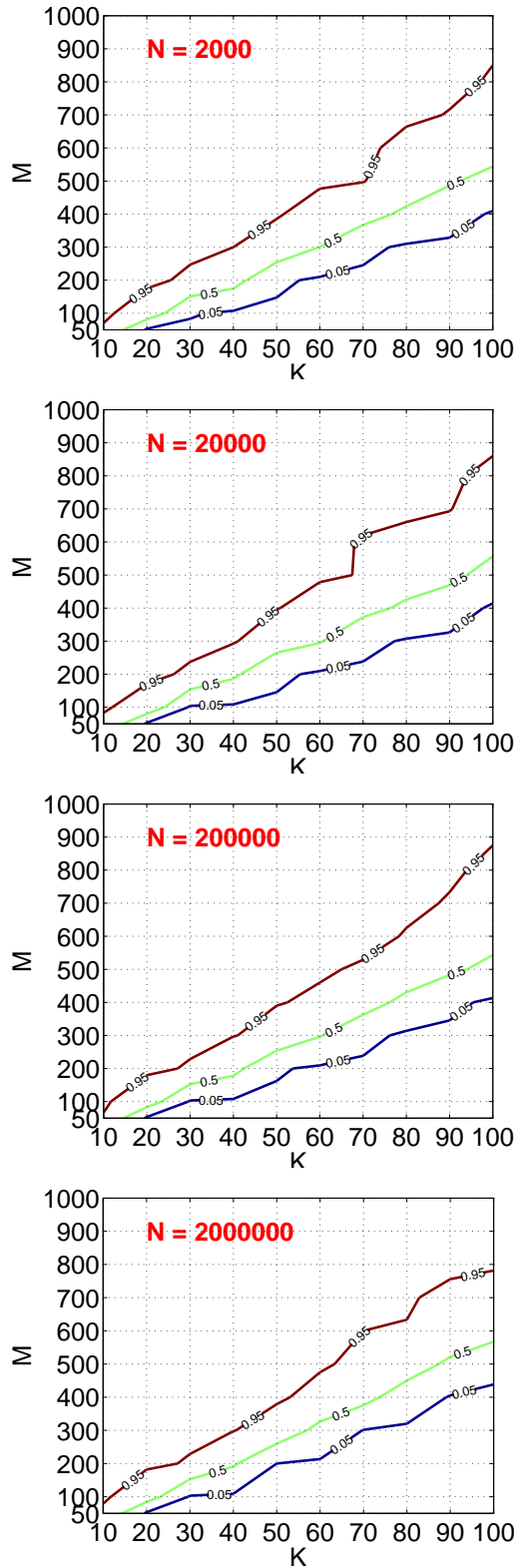

Figure 2: Contour plot of the empirical success probabilities of our proposed method, for $N = 2000$, 20000, 200000, and 2000000. For each combination (N, M, K), we repeated the simulation 100 times. For $N = 20000$, we can see from the wiki page [1] that our prosed method provides accurate recovery results compared to L1 decoding.

computational perspective of the decoding step. It turns out that using a very sparse design matrix can lead to a computationally very efficient recovery algorithm without losing accuracies (compared to L1 decoding).

Acknowledgement

Appendix   Proof of Lemma 2:

$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| \leq \epsilon, x_i \neq 0\right)$

$=1 - \mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| > \epsilon, x_i \neq 0\right)$

$=1 - E\left(\mathbf{Pr}\left(\left|\dfrac{y_j}{s_{ij}}\right| > \epsilon, \ x_i \neq 0, \ j \in T_i | T_i\right)\right)$

$=1 - E\prod_{j \in T_i}\left[\mathbf{Pr}\left(\left|x_i + \eta_{ij}^{1/2}\dfrac{S_2}{S_1}\right| > \epsilon, x_i \neq 0\right)\right]$

$=1 - E\prod_{j \in T_i}\left[1 - \dfrac{1}{\pi}\tan^{-1}\left(\dfrac{\epsilon - x_i}{\eta_{ij}^{1/2}}\right) - \dfrac{1}{\pi}\tan^{-1}\left(\dfrac{\epsilon + x_i}{\eta_{ij}^{1/2}}\right)\right]$

$=1 - E\left\{[1 - E\{A_{ij}\}]^{|T_i|}\right\}$

$=1 - [1 - \gamma + \gamma\{1 - E\{A_{ij}\}\}]^M$

$=1 - [1 - \gamma E\{A_{ij}\}]^M$

were $A_{ij} = \frac{1}{\pi}\tan^{-1}\left(\frac{\epsilon + x_i}{\eta_{ij}^{1/2}}\right) - \frac{1}{\pi}\tan^{-1}\left(\frac{x_i - \epsilon}{\eta_{ij}^{1/2}}\right)$.

Note that $\tan^{-1}(z + \epsilon) - \tan^{-1}(z - \epsilon) \leq 2\tan^{-1}\epsilon \leq 2\epsilon$, for $\epsilon \geq 0$. Therefore,

$\mathbf{Pr}\left(|\hat{x}_{i,min,\gamma}| \leq \epsilon, x_i \neq 0\right) = 1-$

$\left[1 - \gamma E\left\{\dfrac{1}{\pi}\tan^{-1}\left(\dfrac{\epsilon + x_i}{\eta_{ij}^{1/2}}\right) - \dfrac{1}{\pi}\tan^{-1}\left(\dfrac{x_i - \epsilon}{\eta_{ij}^{1/2}}\right)\right\}\right]^M$

$\leq 1 - \left[1 - \dfrac{2}{\pi}\gamma\tan^{-1}\epsilon\right]^M$

which approaches zero as $\epsilon \to 0$. □

References

[1] Sparse recovery experiments with sparse matrices. http://groups.csail.mit.edu/toc/sparse/wiki/index.php?title=Sparse_Recovery_Experiments.

[2] R. Berinde, P. Indyk, and M. Ruzic. Practical near-optimal sparse recovery in the L1 norm. In *Commu-nication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 198 –205, Sept 2008.

[3] Emmanuel Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.

[4] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. *Theor. Comput. Sci.*, 312(1):3–15, 2004.

[5] Scott Shaobing Chen, David L. Donoho, Michael, and A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1998.

[6] Graham Cormode and S. Muthukrishnan. An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithm*, 55(1):58–75, 2005.

[7] Daivd L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 40(7):2845–2862, nov. 2001.

[8] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[9] David L. Donoho and Philip B. Stark. Uncertainty principles and signal recovery. *SIAM Journal of Applied Mathematics*, 49(3):906–931, 1989.

[10] Anna Gilbert and Piotr. Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937 –947, june 2010.

[11] Ping Li. Very sparse stable random projections for dimension reduction in l_α ($0 < \alpha \leq 2$) norm. In *KDD*, San Jose, CA, 2007.

[12] Ping Li, Trevor J. Hastie, and Kenneth W. Church. Very sparse random projections. In *KDD*, pages 287–296, Philadelphia, PA, 2006.

[13] Ping Li and Cun-Hui Zhang. Exact sparse recovery with L0 projections. In *KDD*, pages 302–310, 2013.

[14] Ping Li, Cun-Hui Zhang, and Tong Zhang. Sparse recovery with very sparse compressed counting. Technical report, 2013.

[15] S.G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397 –3415, 1993.

[16] S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends in Theoretical Computer Science*, 1:117–236, 2 2005.

[17] Dapo Omidiran and Martin J. Wainwright. High-dimensional variable selection with sparse random projections: Measurement sparsity and statistical efficiency. *Journal of Machine Learning Research*, 11:2361–2386, 2010.

[18] Y.C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44 vol.1, Nov 1993.

[19] J.A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231 – 2242, oct. 2004.

[20] Tong Zhang. Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Transactions on Information Theory*, 57(9):6215 –6221, sept. 2011.