Sparse Recovery with Very Sparse Compressed Counting

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

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

Tong Zhang
Department of Statistics & Biostatistics
Rutgers University
Piscataway, NJ 08854, USA	
tongz@rci.rutgers.edu

Abstract

Compressed\textsuperscript{1} sensing (sparse signal recovery) often encounters nonnegative data (e.g., images). Recently [11] developed the methodology of using (dense) Compressed Counting for recovering nonnegative $K$-sparse signals. In this paper, we adopt very sparse Compressed Counting for nonnegative signal recovery. Our design matrix is sampled from a maximally-skewed $\alpha$-stable distribution ($0 < \alpha < 1$), and we sparsify the design matrix so that on average $(1-\gamma)$-fraction of the entries become zero. The idea is related to very sparse stable random projections [9, 6], the prior work for estimating summary statistics of the data.

In our theoretical analysis, we show that, when $\alpha \rightarrow 0$, it suffices to use $M = \frac{K}{1-\gamma} \log N/\delta$ measurements, so that with probability $1 - \delta$, all coordinates can be recovered within $\epsilon$ additive precision, in one scan of the coordinates. If $\gamma = 1$ (i.e., dense design), then $M = K \log N/\delta$. If $\gamma = 1/K$ or $2/K$ (i.e., very sparse design), then $M = 1.58K \log N/\delta$ or $M = 1.16K \log N/\delta$. This means the design matrix can be indeed very sparse at only a minor inflation of the sample complexity.

Interestingly, as $\alpha \rightarrow 1$, the required number of measurements is essentially $M = eK \log N/\delta$ provided $\gamma = 1/K$. It turns out that this complexity $eK \log N/\delta$ (at $\gamma = 1/K$) is a general worst-case bound.

\textsuperscript{1}Part of the content of this paper was submitted to a conference in May 2013.
1 Introduction

In a recent paper [11], we developed a new framework for compressed sensing (sparse signal recovery) [4, 2], by focusing on nonnegative sparse signals, i.e., \( x \in \mathbb{R}^N \) and \( x_i \geq 0, \forall i \). Note that real-world signals are often nonnegative. The technique was based on Compressed Counting (CC) [8, 7, 10]. In that framework, entries of the (dense) design matrix are sampled i.i.d. from an \( \alpha \)-stable maximally-skewed distribution. In this paper, we integrate the idea of very sparse stable random projections [9, 6] into the procedure, to develop very sparse compressed counting for compressed sensing.

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

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

(1)

Here, \( s_{ij} \) is the \((i, j)\)-th entry of the design matrix with \( s_{ij} \sim S(\alpha, 1, 1) \) i.i.d, where \( S(\alpha, 1, 1) \) denotes an \( \alpha \)-stable maximally-skewed (i.e., skewness = 1) distribution with unit scale. 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.,

\[
r_{ij} = \begin{cases} 
1 & \text{with prob. } \gamma \\
0 & \text{with prob. } 1 - \gamma 
\end{cases} \quad \text{i.i.d.}
\]  

(2)

And any \( s_{ij} \) and \( r_{ij} \) are also independent.

In the decoding phase, our proposed estimator of the \( i \)-th coordinate \( x_i \) is simply

\[
\hat{x}_{i, \min, \gamma} = \min_{j \in T_i} \frac{y_j}{s_{ij}r_{ij}}
\]  

(3)

where \( T_i \) is the set of nonzero entries in the \( i \)-th row of the design matrix, i.e.,

\[
T_i = \{ j, 1 \leq j \leq M, r_{ij} = 1 \}
\]  

(4)

Note that the size of the set \(|T_i| \sim Binomial(M, \gamma)\).

To analyze the sample complexity (i.e., the required number of measurements), we need to study the following error probability

\[
\Pr (\hat{x}_{i, \min, \gamma} > x_i + \epsilon)
\]  

(5)

from which we can derive the sample complexity by using the following inequality

\[
N \Pr (\hat{x}_{i, \min, \gamma} > x_i + \epsilon) \leq \delta
\]  

(6)

so that any \( x_i \) can be estimated within \((x_i, x_i + \epsilon)\) with a probability (at least) \( 1 - \delta \).

Main Result 1: As \( \alpha \to 0^+ \), the required number of measurements is

\[
M = \frac{1}{- \log \left[ 1 - \frac{1}{K+1} (1 - (1 - \gamma)K+1) \right]} \log N/\delta
\]  

(7)

which can essentially be written as

\[
M = \frac{K}{1 - e^{-\gamma K}} \log N/\delta
\]  

(8)
If $\gamma = 1/K$, then the required $M$ is about $1.58K \log N/\delta$. If $\gamma = 2/K$, then $M$ is about $1.16K \log N/\delta$. In other words, we can use a very sparse design matrix and the required number of measurements will only be inflated slightly, if we choose to use a small $\alpha$.

Indeed, using $\alpha \to 0+$ achieves the smallest complexity. However, there will be a numerical issue if $\alpha$ is too small. To see this, consider the approximate mechanism for generating $S(\alpha, 1, 1)$ by using $1/U^{1/\alpha}$, where $U \sim \text{unif}(0, 1)$. If $\alpha = 0.05$, then we have to compute $\left(1/U\right)^{20}$, which may potentially create numerical problems. In our Matlab simulations, we do not notice obvious numerical issues with $\alpha = 0.05$ (or even smaller). However, if a device (e.g., camera or other hand-held device) has a limited precision and/or memory, then we expect that we must use a larger $\alpha$, away from 0.

**Main Result 2:** If $x_i > \epsilon$ whenever $x_i > 0$, then as $\alpha \to 1-$, the required number of measurements is

$$M = \frac{1}{-\log \left(1 - \frac{1}{K+1} \left(1 - \frac{1}{K+1}\right)^K\right)} \log N/\delta, \quad \text{with} \quad \gamma = \frac{1}{K+1}$$

This complexity bound can essentially be written as

$$M = eK \log N/\delta, \quad \text{with} \quad \gamma = \frac{1}{K}$$

Interestingly, this result $eK \log N/\delta$ (with $\gamma = 1/K$) is the general worse-case bound.

### 2 A Simulation Study

We consider two types of signals. To generate “binary signal”, we randomly select $K$ (out of $N$) coordinates to be 1. For “non-binary signal”, we assign the values of $K$ randomly selected nonzero coordinates according to $|N(0, 5^2)|$. The number of measurements is determined by

$$M = \nu K \log N/\delta$$

where $N \in \{10000, 100000\}$, $\delta = 0.01$ and $\nu \in \{1.2, 1.6, 2\}$. We report the normalized recovery errors:

$$\text{Normalized Error} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \text{estimated } x_i)^2}{\sum_{i=1}^{N} x_i^2}}$$

We experiment with all possible values of $1/\gamma \in \{1, 2, 3, ..., K\}$, although we only plot a few selected $\gamma$ values in Figures 1 to 4. For each combination $(\gamma, N, \nu)$, we conduct 100 simulations and report the median errors. The results confirm our theoretical analysis. When $\nu$ is small (i.e., less measurements), we need to choose a small $\alpha$ in order to achieve perfect recovery. When $\nu$ is large (i.e., more measurements), we can use a larger $\alpha$. Also, the simulations confirm that, in general, we can choose a very sparse design.

---

2This report does not include comparisons with the SMP algorithm [1, 5], as we can not run the code from [http://groups.csail.mit.edu/toc/sparse/wiki/index.php?title=Sparse_Recovery_Experiments](http://groups.csail.mit.edu/toc/sparse/wiki/index.php?title=Sparse_Recovery_Experiments). We will provide the comparisons after we are able to execute the code. We thank the communications with the author of [1, 5].
Figure 1: Normalized estimation errors (12) with $N = 10000$ and $K = 10$. 
Figure 2: Normalized estimation errors \([12]\) with \(N = 10000\) and \(K = 20\).
Figure 3: Normalized estimation errors with $N = 10000$ and $K = 100$. 
Figure 4: Normalized estimation errors (12) with $N = 100000$ and $\nu = 2$. 
3 Analysis

Recall, we collect our measurements as

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

(13)

where \( s_{ij} \sim S(\alpha, 1, 1) \) i.i.d. and

\[ r_{ij} = \begin{cases} 1 & \text{with prob. } \gamma \\ 0 & \text{with prob. } 1 - \gamma \end{cases} \text{ i.i.d.} \]  

(14)

And any \( s_{ij} \) and \( r_{ij} \) are also independent. Our proposed estimator is simply

\[ \hat{x}_{i,\min, \gamma} = \min_{j \in T_i} \frac{y_j}{s_{ij} r_{ij}} \]  

(15)

where \( T_i \) is the set of nonzero entries in the \( i \)-th row of \( S \), i.e.,

\[ T_i = \{ j, 1 \leq j \leq M, r_{ij} = 1 \} \]  

(16)

Conditional on \( r_{ij} = 1 \),

\[ \frac{y_j}{s_{ij} r_{ij}} \bigg| r_{ij} = 1 = \frac{\sum_{t=1}^{N} x_t s_{tj} r_{tj}}{s_{ij}} = x_i + \frac{\sum_{t \neq i}^{N} x_t s_{tj} r_{tj}}{s_{ij}} = x_i + (\eta_{ij})^{1/\alpha} \frac{S_2}{S_1} \]  

(17)

where \( S_1, S_2 \sim S(\alpha, 1, 1) \), i.i.d., and

\[ \eta_{ij} = \sum_{t \neq i}^{N} (x_t r_{tj})^\alpha = \sum_{t \neq i}^{N} x_t^\alpha r_{tj} \]  

(18)

Note that

\[ E(\eta_{ij}) = \gamma \sum_{t \neq i}^{N} x_t^\alpha \leq \gamma \sum_{t=1}^{N} x_t^\alpha, \quad \lim_{\alpha \to 0^+} E(\eta_{ij}) \leq \gamma K \]  

(19)

When the signals are binary, i.e., \( x_i \in \{0, 1\} \), we have

\[ \eta_{ij} \sim \begin{cases} \text{Binomial}(K, \gamma) & \text{if } x_i = 0 \\ \text{Binomial}(K - 1, \gamma) & \text{if } x_i = 1 \end{cases} \]  

(20)

The key in our theoretical analysis is the distribution of the ratio of two independent stable random variables. Here, we consider \( S_1, S_2 \sim S(\alpha, 1, 1) \), i.i.d., and define

\[ F_\alpha(t) = \text{Pr} \left( \left(\frac{S_2}{S_1}\right)^{\alpha/(1-\alpha)} \leq t \right), \quad t \geq 0 \]  

(21)

There is a standard procedure to sample from \( S(\alpha, 1, 1) \) [3]. We first generate an exponential random variable with mean 1, \( w \sim \exp(1) \), and a uniform random variable \( u \sim \text{unif}(0, \pi) \), and then compute

\[ \frac{\sin (\alpha u)}{[\sin u \cos (\alpha \pi/2)]^{1/\alpha}} \left[ \frac{\sin (u - \alpha u)}{w} \right]^{1/\alpha} \sim S(\alpha, 1, 1) \]  

(22)
Lemma 1 For any $t \geq 0$, $S_1, S_2 \sim S(\alpha, 1, 1)$, i.i.d.,

$$F_{\alpha}(t) = \Pr \left( \frac{(S_2/S_1)^{\alpha/(1-\alpha)}}{1+Q_{\alpha}/t} \leq t \right) = \frac{1}{\pi^2} \int_0^\pi \int_0^\pi \frac{1}{1+Q_{\alpha}/t} du_1 du_2$$

(23)

where

$$Q_{\alpha} = \left[ \frac{\sin(\alpha u_2)}{\sin(\alpha u_1)} \right]^{\alpha/(1-\alpha)} \left[ \frac{\sin u_1}{\sin u_2} \right]^{1-\alpha} \frac{\sin(u_2-\alpha u_2)}{\sin(u_1-\alpha u_1)}$$

(24)

In particular,

$$\lim_{\alpha \to 0^+} F_{\alpha}(t) = \frac{1}{1+1/t}, \quad F_{0.5}(t) = \frac{2}{\pi} \tan^{-1} \sqrt{t} \quad \square$$

(25)

3.1 Error Probability

The following Lemma derives the general formula (26) for the error probability in terms of an expectation, which in general does not have a close-form solution. Nevertheless, when $\alpha = 0^+$ and $\alpha = 0.5$, we can derive two convenient upper bounds, (28) and (30), respectively, which however are not tight.

Lemma 2

$$\Pr(\hat{x}_{i,min,\gamma} > x_i + \epsilon) = \left[ 1 - \gamma E \left\{ F_{\alpha} \left( \frac{\epsilon}{\eta_{ij}} \right)^{1/(1-\alpha)} \right\} \right]^M$$

(26)

When $\alpha \to 0^+$, we have

$$\Pr(\hat{x}_{i,min,\gamma} > x_i + \epsilon) \leq \left[ 1 - \frac{1}{1+1/(2\gamma+K+1)} \right]^M$$

(27)

$$\leq \left[ 1 - \frac{1}{1+\gamma + K} \right]^M$$

(28)

When $\alpha = 0.5$, we have

$$\Pr(\hat{x}_{i,min,\gamma} > x_i + \epsilon) \leq \left[ 1 - \gamma \frac{2}{\pi} \tan^{-1} \left( \frac{\sqrt{\epsilon}}{\gamma \sum_{t \neq i} x_t^{1/2}} \right) \right]^M$$

(29)

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

(30)

Proof: See Appendix A

□

It turns out, when $\alpha = 0^+$, we can precisely evaluate the expectation (26) and derive an accurate complexity bound (31) in Lemma 3.

Lemma 3 As $\alpha \to 0^+$, we have

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

(31)

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

(32)

$$\leq \left[ 1 - \frac{1}{1/\gamma + K} \right]^M$$

(33)

Proof: See Appendix B

□
3.2 Sample Complexity when $\alpha \to 0^+$

Based on the precise error probability (31) in Lemma 3, we can derive the sample complexity bound from

$$(N - K) \left[ 1 - \frac{1}{K+1} (1 - (1 - \gamma)^{K+1}) \right]^M + K \left[ 1 - \frac{1}{K} (1 - (1 - \gamma)^K) \right]^M \leq \delta$$  \hspace{1cm} (34)

Because $\left[ 1 - \frac{1}{K} (1 - (1 - \gamma)^K) \right]^M \leq \left[ 1 - \frac{1}{K+1} (1 - (1 - \gamma)^{K+1}) \right]^M$, it suffices to let

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

This immediately leads to the sample complexity result for $\alpha \to 0^+$ in Theorem 1.

**Theorem 1** As $\alpha \to 0^+$, the required number of measurements is

$$M = \frac{1}{-\log \left[ 1 - \frac{1}{K+1} (1 - (1 - \gamma)^{K+1}) \right]} \log \frac{N}{\delta}$$  \hspace{1cm} (35)

□

**Remark:** The required number of measurements (35) can essentially be written as

$$M = \frac{K}{1 - e^{-\gamma K}} \log \frac{N}{\delta}$$  \hspace{1cm} (36)

The difference between (35) and (36) is very small even when $K$ is small, as shown in Figure 5. Let $\lambda = \gamma K$. If $\lambda = 1$ (i.e., $\gamma = 1/K$), then the required $M$ is about $1.58K \log N/\delta$. If $\lambda = 2$ (i.e., $\gamma = 2/K$), then $M$ is about $1.16K \log N/\delta$. In other words, we can use a very sparse design matrix and the required number of measurements is only inflated slightly.

![Figure 5](image_url)

Figure 5: Solid curves: $-\log \left[ 1 - \frac{1}{K+1} (1 - (1 - \gamma)^{K+1}) \right]$. Dashed curves: $\frac{K}{1 - e^{-\gamma K}}$. The difference between (35) and (36) is very small even for small $K$. For large $K$, both terms approach $K$. 

3.3 Worst-Case Sample Complexity

**Theorem 2** If we choose \( \gamma = \frac{1}{K+1} \), then it suffices to choose the number of measurements by

\[
M = \frac{1}{- \log \left( 1 - \frac{1}{K+1} \left( 1 - \frac{1}{K+1} \right)^K \right)} \log N/\delta \tag{37}
\]

**Proof:** See Appendix C. \( \square \)

**Remark:** The worst-case complexity (37) can essentially be written as

\[
M = eK \log N/\delta, \quad \text{if} \quad \gamma = 1/K \tag{38}
\]

where \( e = 2.7183 \ldots \) The previous analysis of sample complexity for \( \alpha \to 0^+ \) says that if \( \gamma = 1/K \), it suffices to let \( M = 1.58K \log N/\delta \), and if \( \gamma = 2/K \), it suffices to let \( M = 1.15K \log N/\delta \). This means that the worst-case analysis is quite conservative and the choice \( \gamma = 1/K \) is not optimal for general \( \alpha \in (0, 1) \).

Interestingly, it turns out that the worst-case sample complexity is attained when \( \alpha \to 1^- \).

3.4 Sample Complexity when \( \alpha = 1^- \)

**Theorem 3** For a \( K \)-sparse signal whose nonzero coordinates are larger than \( \epsilon \), i.e., \( x_i > \epsilon \) if \( x_i > 0 \). If we choose \( \gamma = \frac{1}{K+1} \), as \( \alpha \to 1^- \), it suffices to choose the number of measurements by

\[
M = \frac{1}{- \log \left( 1 - \frac{1}{K+1} \left( 1 - \frac{1}{K+1} \right)^K \right)} \log N/\delta \tag{39}
\]

**Proof:** The proof can be directly inferred from the proof of Theorem 2 at \( \alpha = 1^- \). \( \square \)

**Remark:** Note that, if the assumption \( x_i > \epsilon \) whenever \( x_i > 0 \) does not hold, then the required number of measurements will be smaller.

3.5 Sample Complexity Analysis for Binary Signals

As this point, we know the precise sample complexities for \( \alpha = 0^+ \) and \( \alpha = 1^- \). And we also know the worst-case complexity. Nevertheless, it would be still interesting to study how the complexity varies as \( \alpha \) changes between 0 and 1. While a precise analysis is difficult, we can perform an accurate analysis at least for binary signals, i.e., \( \{0, 1\} \). For convenience, we first re-write the general error probability as

\[
\Pr(\hat{x}_{i,\min,\gamma} > x_i + \epsilon) = \left[ 1 - \frac{1}{K} (\gamma K) E \left\{ F_{\alpha} \left( \left( \frac{\epsilon}{\eta_{ij}} \right)^{(1/(1-\alpha))} \right) \right\} \right]^M \tag{40}
\]

For binary signals, we have \( \eta_{ij} \sim \text{Binomial}(K - 1 + 1_{x_i=0}, \gamma) \). Thus, if \( x_i = 0 \), then

\[
H = H(\gamma, K; \epsilon, \alpha) \triangleq (\gamma K) \sum_{k=0}^{K} F_{\alpha} \left( \left( \frac{\epsilon}{\eta_{ij}} \right)^{(1/(1-\alpha))} \right) \binom{K}{k} (1 - \gamma)^{K-k} \tag{41}
\]

The required number of measurements can be written as \( \frac{1}{- \log(1 - H/K)} \log N/\delta \), or essentially \( \frac{K}{H} \log N/\delta \). We can compute \( H(\gamma, K; \epsilon, \alpha) \) for given \( \gamma, K, \epsilon \), and \( \alpha \), at least by simulations.
4 Poisson Approximation for Complexity Analysis with Binary Signals

Again, the purpose is to study more precisely how the sample complexity varies with $\alpha \in (0, 1)$, at least for binary signals. In this case, when $x_i = 0$, we have $\eta_{ij} \sim \text{Binomial}(K, \gamma)$. Elementary statistics tells us that we can well approximate this binomial with a Poisson distribution with parameter $\lambda = \gamma K$ especially when $K$ is not small. Using the Poisson approximation, we can replace $H(\gamma, K; \epsilon, \alpha)$ in (41) by $h(\lambda; \epsilon, \alpha)$ and re-write the error probability as

$$
\Pr (\hat{x}_{i, \text{min}, \gamma} > x_i + \epsilon) = \left[ 1 - \frac{1}{K} h(\lambda; \epsilon, \alpha) \right]^M
$$

(42)

where

$$
h(\lambda; \epsilon, \alpha) = \lambda \sum_{k=0}^\infty F_\alpha \left( \left( \frac{\epsilon^\alpha}{k} \right)^{1/(1-\alpha)} \right) \frac{e^{-\lambda} \lambda^k}{k!}
$$

$$
= \lambda e^{-\lambda} + \lambda e^{-\lambda} \sum_{k=1}^\infty F_\alpha \left( \left( \frac{\epsilon^\alpha}{k} \right)^{1/(1-\alpha)} \right) \frac{\lambda^k}{k!}
$$

(43)

which can be computed numerically for any given $\lambda$ and $\epsilon$.

The required number of measurements can be computed from

$$
N \left[ 1 - \frac{1}{K} h(\lambda; \epsilon, \alpha) \right]^M = \delta \iff M = \frac{\log N/\delta}{-\log \left[ 1 - \frac{1}{K} h(\lambda; \epsilon, \alpha) \right]}
$$

(44)

for which it suffices to choose $M$ such that

$$
M = \frac{K}{h(\lambda; \epsilon, \alpha)} \log N/\delta
$$

(45)

Therefore, we hope $h(\lambda; \epsilon, \alpha)$ should be as large as possible.

4.1 Analysis for $\alpha = 0.5$

Before we demonstrate the results via Poisson approximation for general $0 < \alpha < 1$, we would like to illustrate the analysis particularly for $\alpha = 0.5$, which is a case readers can more easily verify.

Recall when $\alpha = 0.5$, the error probability can be written as

$$
\Pr (\hat{x}_{i, \text{min}, \gamma} > x_i + \epsilon) = \left[ 1 - \frac{1}{K} (\gamma K) E \left\{ \frac{2}{\pi} \tan^{-1} \left( \frac{\sqrt{\epsilon}}{\eta_{ij}} \right) \right\} \right]^M = \left[ 1 - \frac{1}{K} H(\gamma, K; \epsilon, 0.5) \right]^M
$$

where

$$
H(\gamma, K; \epsilon, 0.5) = (\gamma K) \frac{2}{\pi} \sum_{k=0}^K \tan^{-1} \left( \frac{\sqrt{\epsilon}}{k} \right) \frac{K^k}{k} \gamma^k (1-\gamma)^{K-k}
$$

(46)

From Lemma [2] in particular [20], we know there is a convenient lower bound of $H$:

$$
H(\gamma, K; \epsilon, 0.5) \geq H_{\text{lower}}(\gamma, K; \epsilon, 0.5) = (\gamma K) \left\{ \frac{2}{\pi} \tan^{-1} \left( \frac{\sqrt{\epsilon}}{K} \right) \right\} = \frac{2}{\pi} \tan^{-1} \left( \frac{\sqrt{\epsilon}}{\lambda} \right)
$$

(47)
We will compare the precise $H(\gamma, K; \epsilon, 0.5)$ with its lower bound $H^{lower}(\gamma, K; \epsilon, 0.5)$, along with the Poisson approximation:

$$H(\gamma, K; \epsilon, 0.5) \approx h(\lambda; \epsilon, 0.5) = \lambda e^{-\lambda} \frac{2}{\pi} \sum_{k=0}^{\infty} \tan^{-1}\left(\frac{\sqrt{\epsilon}}{k}\right) \frac{\lambda^k}{k!}$$ \hspace{1cm} (48)

Figure 6 confirms that the Poisson approximation is very accurate unless $K$ is very small, while the lower bound is conservative especially when $\gamma$ is around the optimal value. For small $\epsilon$, the optimal $\gamma$ is around $1/K$, which is consistent with the general worst-case complexity result.

4.2 Poisson Approximation for General $0 < \alpha < 1$

Once we are convinced that the Poisson approximation is reliable at least for $\alpha = 0.5$, we can use this tool to study for general $\alpha \in (0, 1)$. Again, assume the Poisson approximation, we have

$$\Pr(\hat{x}_{i,\min,\gamma} > x_i + \epsilon) = \left[1 - \frac{1}{K} h(\lambda; \epsilon, \alpha)\right]^M$$

where

$$h(\lambda; \epsilon, \alpha) = \lambda e^{-\lambda} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} F_\alpha\left(\left(\frac{\epsilon}{k}\right)^{1/(1-\alpha)}\right) \frac{\lambda^k}{k!}$$
The required number of measurements can be computed from \( M = \frac{K}{h(\lambda; \epsilon, \alpha)} \log N/\delta. \)

As shown in Figure 7 at fixed \( \epsilon \) and \( \lambda \), the optimal (highest) \( h \) is larger when \( \alpha \) is smaller. The optimal \( h \) occurs at larger \( \lambda \) when \( \alpha \) is closer to zero and at smaller \( \lambda \) when \( \alpha \) is closer to 1.

Figure 7: \( h(\lambda; \epsilon, \alpha) \) as defined in (43) for selected \( \alpha \) values ranging from 0.01 to 0.95. In each panel, each curve corresponds to an \( \epsilon \) value, where \( \epsilon \in \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\} \) (from bottom to top). In each panel, the curve for \( \epsilon = 0.01 \) is the lowest and the curve for \( \epsilon = 1 \) is the highest.
Figure 8 plots the optimal (smallest) $1/h(\lambda; \epsilon, \alpha)$ values (left panel) and the optimal $\lambda$ values (right panel) which achieve the optimal $h$.

Figure 8: Left Panel: $1/h(\lambda; \epsilon, \alpha)$ at the optimal $\lambda$ values. Right Panel: the optimal $\lambda$ values.

Figure 9 plots $1/h(\lambda; \epsilon, \alpha)$ for fixed $\lambda = 1$ (left panel) and $\lambda = 2$ (right panel), together with the optimal $1/h(\lambda; \epsilon, \alpha)$ values (dashed curves).

Figure 9: $1/h(\lambda; \epsilon, \alpha)$ at the fixed $\lambda = 1$ (left panel) and $\lambda = 2$ (right panel). The dashed curves correspond to $1/h(\lambda; \epsilon, \alpha)$ at the optimal $\lambda$ values.

4.3 Poisson Approximation for $\alpha \to 1$–

We now examine $h(\lambda; \epsilon, \alpha)$ closely at $\alpha = 1-$, i.e., $\frac{1}{1-\alpha} \to \infty$.

\[
h(\lambda; \epsilon, \alpha) = \lambda e^{-\lambda} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} F_{\alpha} \left( \frac{\epsilon^\alpha}{k} \right)^{1/(1-\alpha)} \frac{\lambda^k}{k!}
\]

Interestingly, when $\epsilon = 1$, only $k = 0$ and $k = 1$ will be useful, because otherwise $(\frac{\epsilon^\alpha}{k})^{1/(1-\alpha)} \to \infty$ as $\Delta = 1 - \alpha \to 0$. When $\epsilon < 1$, then only $k = 0$ is useful. Thus, we can write

\[
h(\lambda; \epsilon < 1, \alpha = 1-) = \lambda e^{-\lambda}
\]

(49)

\[
h(\lambda; \epsilon = 1, \alpha = 1-) = \lambda e^{-\lambda} + \lambda^2 e^{-\lambda} F_{\alpha} (1) = \lambda e^{-\lambda} + \lambda^2 e^{-\lambda}/2
\]

(50)
Notes that $F_{1-}(1) = 1/2$ due to symmetry.

This means, the maximum of $h(\lambda; \epsilon < 1, \alpha = 1-) = e^{-1}$ attained at $\lambda = 1$, and the maximum of $h(\lambda; \epsilon = 1, \alpha = 1-) = e^{-\sqrt{2}(1 + \sqrt{2})} = 0.5869$, attained at $\lambda = \sqrt{2}$, as confirmed by Figure 10 In other words, it suffices to choose the number of measurements to be

$$M = eK \log N/\delta \text{ if } \epsilon < 1, \quad M = 1.7038K \log N/\delta \text{ if } \epsilon = 1$$

(51)

\[\text{Figure 10: } h(\lambda; \epsilon, \alpha) \text{ as defined in (43) for } \alpha \text{ close to 1. As } \alpha \to 1-, \text{ the maximum of } h(\lambda; \epsilon, \alpha) \text{ approaches } e^{-1} \text{ attained at } \lambda = 1, \text{ for all } \epsilon < 1. \text{ When } \epsilon = 1, \text{ the maximum approaches 0.5869, attained at } \lambda = \sqrt{2}.}\]

5 Conclusion

In this paper, we extend the prior work on Compressed Counting meets Compressed Sensing [11] and very sparse stable random projections [9, 6] to the interesting problem of sparse recovery of nonnegative signals. The design matrix is highly sparse in that on average only $\gamma$-fraction of the entries are nonzero; and we sample the nonzero entries from an $\alpha$-stable maximally-skewed distribution where $\alpha \in (0, 1)$. Our theoretical analysis demonstrates that the design matrix can be extremely sparse, e.g., $\gamma = \frac{1}{K} \sim \frac{2}{K}$. In fact, when $\alpha$ is away from 0, it is much more preferable to use a very sparse design.
A Proof of Lemma 2

\[ \Pr (\hat{x}_{i,\min, \gamma} > x_i + \epsilon) = E \left( \Pr \left( \frac{y_{ij}}{s_{ij}} > x_i + \epsilon, j \in T_i | T_i \right) \right) \]

\[ = E \prod_{j \in T_i} \left[ 1 - F_\alpha \left( \frac{\epsilon}{\eta_{ij}} \right)^{1/(1-\alpha)} \right] \]

\[ = E \left\{ \left[ 1 - E \left\{ F_\alpha \left( \frac{\epsilon}{\eta_{ij}} \right)^{1/(1-\alpha)} \right\} \right]^{|T_i|} \right\} \]

\[ = \left[ 1 - \gamma + \gamma \left\{ 1 - E \left\{ F_\alpha \left( \frac{\epsilon}{\eta_{ij}} \right)^{\alpha/(1-\alpha)} \right\} \right\} \right]^M \]

When \( \alpha = 0.5 \), we have \( F_\alpha(t) = \frac{2}{\pi} \tan^{-1} \sqrt{t} \) and hence

\[ \Pr (\hat{x}_{i,\min, \gamma} > x_i + \epsilon) = \left[ 1 - \gamma + \gamma \left\{ 1 - E \left\{ F_\alpha \left( \frac{\epsilon}{\eta_{ij}} \right)^{0.5} \right\} \right\} \right]^M \]

\[ \leq \left[ 1 - \gamma \left\{ \frac{2}{\pi} \tan^{-1} \left( \frac{\sqrt{\epsilon}}{\eta_{ij}} \right) \right\} \right]^M \quad \text{(Jensen’s Inequality)} \]

\[ \leq \left[ 1 - \gamma \left\{ \frac{2}{\pi} \tan^{-1} \left( \frac{1}{\gamma \sum_{t \neq i} x_t^{1/2}} \right) \right\} \right]^M \]

17
When \( \alpha = 0^+ \), we have \( F_{0^+}(t) = \frac{1}{1+1/t} \) and hence

\[
\Pr(\hat{x}_{i,\min, \gamma} > x_i + \epsilon) = \lim_{\alpha \to 0^+} \left[ 1 - \gamma E \left\{ F_{0^+} \left( \frac{1}{\eta_{ij}} \right) \right\} \right]^M \\
= \lim_{\alpha \to 0^+} \left[ 1 - \gamma E \left\{ \left( \frac{1}{1 + \eta_{ij}} \right) \right\} \right]^M \\
\leq \lim_{\alpha \to 0^+} \left[ 1 - \gamma \left\{ \left( \frac{1}{1 + E\eta_{ij}} \right) \right\} \right]^M \\
\leq \lim_{\alpha \to 0^+} \left[ 1 - \gamma \frac{1}{1 + \gamma K} \right]^M \\
= \left[ 1 - \frac{1}{1/\gamma + K} \right]^M \\
\]

This completes the proof.

**B  Proof of Lemma 3**

**Proof:** When \( \alpha = 0^+ \), we have \( F_{0^+}(t) = \frac{1}{1+1/t} \) and hence

\[
\Pr(\hat{x}_{i,\min, \gamma} > x_i + \epsilon) = \lim_{\alpha \to 0^+} \left[ 1 - \gamma E \left\{ \left( \frac{1}{1 + \eta_{ij}} \right) \right\} \right]^M \\
\]

Suppose \( x_i = 0 \), then as \( \alpha \to 0^+ \), \( \eta_{ij} \sim \text{Binomial}(K, \gamma) \), and

\[
E \left( \frac{1}{1 + \eta_{ij}} \right) = \sum_{n=0}^{K} \frac{1}{1+n} \binom{K}{n} \gamma^n (1-\gamma)^{K-n} \\
= \sum_{n=0}^{K} \frac{1}{1+n} \frac{K!}{n!(K-n)!} \gamma^n (1-\gamma)^{K-n} \\
= \sum_{n=0}^{K} \frac{K!}{(n+1)!(K-n)!} \gamma^n (1-\gamma)^{K-n} \\
= \frac{1}{K+1} \gamma \sum_{n=0}^{K} \frac{(K+1)!}{(n+1)!(K+1-(n+1))!} \gamma^{n+1}(1-\gamma)(K+1)-(n+1) \\
= \frac{1}{K+1} \gamma \sum_{n=1}^{K+1} \frac{(K+1)!}{(n)!(K+1-(n))!} \gamma^n(1-\gamma)(K+1)-(n) \\
= \frac{1}{K+1} \gamma \left\{ \sum_{n=0}^{K+1} \frac{(K+1)!}{(n)!(K+1-(n))!} \gamma^n(1-\gamma)(K+1)-(n) - (1-\gamma)^{K+1} \right\} \\
= \frac{1}{K+1} \gamma \left\{ 1 - (1-\gamma)^{K+1} \right\} \\
\]

18
Similarly, suppose \( x_i > 0 \), we have
\[
E \left( \frac{1}{1 + \eta_{ij}} \right) = \frac{1}{K} \left\{ 1 - (1 - \gamma)^K \right\}
\]

Therefore, as \( \alpha \to 0^+ \), when \( x_i = 0 \), we have
\[
\Pr(\hat{x}_{i,\min,\gamma} > x_i + \epsilon) = \left[ 1 - \frac{1}{K + 1} (1 - (1 - \gamma)^{K+1}) \right]^M
\]
and when \( x_i > 0 \), we have
\[
\Pr(\hat{x}_{i,\min,\gamma} > x_i + \epsilon) = \left[ 1 - \frac{1}{K} (1 - (1 - \gamma)^K) \right]^M
\]

To conclude the proof, we need to show
\[
\left[ 1 - \frac{1}{K + 1} (1 - (1 - \gamma)^{K+1}) \right]^M \leq \left[ 1 - \frac{1}{1/\gamma + K} \right]^M
\]
\[
\iff \frac{1}{K + 1} (1 - (1 - \gamma)^{K+1}) \geq \frac{1}{1/\gamma + K}
\]
\[
\iff h(\gamma, K) = 1/\gamma - (1 - \gamma)^{K+1}/\gamma - K(1 - \gamma)^{K+1} - 1 \geq 0
\]

Note that \( 0 \leq \gamma \leq 1 \), \( h(0, K) = h(1, K) = h(\gamma, 1) = 0 \). Furthermore
\[
\frac{\partial h(\gamma, K)}{\partial K} = - (1 - \gamma)^{K+1} \log(1 - \gamma)/\gamma - (1 - \gamma)^{K+1} - K(1 - \gamma)^{K+1} \log(1 - \gamma)
\]
\[
= - (1 - \gamma)^{K+1} (\log(1 - \gamma)/\gamma + 1 + K \log(1 - \gamma)) \geq 0
\]
as \( \log(1 - \gamma)/\gamma < -1 \). Thus, \( h(\gamma, K) \) is a monotonically increasing function of \( K \) and this completes the proof.

## C Proof of Theorem 2

\[
\Pr(\hat{x}_{i,\min,\gamma} > x_i + \epsilon) = \left[ 1 - \gamma E \left\{ F_{\alpha} \left( \frac{e^{\alpha}}{\eta_{ij}} \right)^{1/(1-\alpha)} \right\} \right]^M
\]
\[
\geq \left[ 1 - \gamma \Pr(\eta_{ij} = 0) \right]^M
\]
\[
= \left[ 1 - \gamma (1 - \gamma)^{K-1+1_{x_i=0}} \right]^M
\]
\[
\geq \left[ 1 - \gamma (1 - \gamma)^K \right]^M
\]
The minimum of \( (1 - \gamma)^{K-1+1_{x_i=0}} \) is attained at \( \gamma = \frac{1}{K+1_{x_i=0}} \). If we choose \( \gamma^* = \frac{1}{K+1} \), then
\[
\Pr(\hat{x}_{i,\min,\gamma} > x_i + \epsilon) \geq \left[ 1 - \gamma^* (1 - \gamma^*)^K \right]^M
\]
\[
= \left[ 1 - \frac{1}{K + 1} \left( 1 - \frac{1}{K + 1} \right)^K \right]^M
\]
and it suffices to choose \( M \) so that
\[
M = \frac{1}{\log \left[ 1 - \frac{1}{K + 1} \left( 1 - \frac{1}{K + 1} \right)^K \right]} \log N/\delta
\]
This completes the proof.
References

[1] R. Berinde, P. Indyk, and M. Ruzic. Practical near-optimal sparse recovery in the l1 norm. In Communication, Control, and Computing, 2008 46th Annual Allerton Conference on, pages 198–205, 2008.

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

[3] John M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. Journal of the American Statistical Association, 71(354):340–344, 1976.

[4] David L. Donoho. Compressed sensing. IEEE Trans. Inform. Theory, 52(4):1289–1306, 2006.

[5] A. Gilbert and P. Indyk. Sparse recovery using sparse matrices. Proc. of the IEEE, 98(6):937–947, June 2010.

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

[7] Ping Li. Improving compressed counting. In UAI, Montreal, CA, 2009.

[8] Ping Li. Compressed counting. In SODA, New York, NY, 2009 [arXiv:0802.0802, arXiv:0808.1766].

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

[10] Ping Li and Cun-Hui Zhang. A new algorithm for compressed counting with applications in shannon entropy estimation in dynamic data. In COLT, 2011.

[11] Ping Li, Cun-Hui Zhang, and Tong Zhang. Compressed counting meets compressed sensing. Technical report, 2013.