One Scan 1-Bit Compressed Sensing

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

Abstract

Based on α-stable random projections with small α, we develop a simple algorithm for compressed sensing (sparse signal recovery) by utilizing only the signs (i.e., 1-bit) of the measurements. Using only 1-bit information of the measurements results in substantial cost reduction in collection, storage, communication, and decoding for compressed sensing. The proposed algorithm is efficient in that the decoding procedure requires only one scan of the coordinates. Our analysis can precisely show that, for a K-sparse signal of length N, $12.3K \log N/\delta$ measurements (where δ is the confidence) would be sufficient for recovering the support and the signs of the signal. While the method is highly robust against typical measurement noises, we also provide the analysis of the scheme under random flipping of the signs of measurements.

Compared to the well-known work on 1-bit marginal regression (which can also be viewed as a one-scan method), the proposed algorithm requires orders of magnitude fewer measurements. Compared to 1-bit Iterative Hard Thresholding (IHT) (which is not a one-scan algorithm), our method is still significantly more accurate. Furthermore, the proposed method is reasonably robust against random sign flipping while IHT is known to be very sensitive to this type of noise.

1 Introduction

Compressed sensing (CS) [7, 2] is a popular and important topic in mathematics and engineering, for recovering sparse signals from linear measurements. Here, we consider a K-sparse signal of length N, denoted by x_i, $i = 1$ to N, with $\sum_{i=1}^{N} 1\{x_i \neq 0\} = K$. In our scheme, the measurements are collected as follows

$$y_j = \sum_{i=1}^{N} x_i s_{ij}, \quad j = 1, 2, ..., M, \quad \text{where } s_{ij} \sim S(\alpha, 1)$$

where y_j's are the measurements and s_{ij} is the (i, j)-th entry of the design matrix sampled i.i.d. from an α-stable distribution with unit scale, denoted by $S(\alpha, 1)$. This is different from classical framework of compressed sensing. Classical algorithms of compressed sensing use Gaussian design (i.e., $\alpha = 2$ in the family of stable distribution) or Gaussian-like design (e.g., a distribution with finite variance), to recover signals via computationally intensive methods such as linear programming [5] or greedy methods such as orthogonal matching pursuit (OMP) [19, 16, 18, 23].

The recent works [14, 15] studied the use of α-stable random projections with $\alpha < 2$, for accurate one-scan compressed sensing. Basically, if $Z \sim S(\alpha, 1)$, then its characteristic function is $E\left(e^{\sqrt{-1}Zt}\right) = e^{-|t|^\alpha}$, where $0 < \alpha \leq 2$. Thus, both Gaussian ($\alpha = 2$) and Cauchy ($\alpha = 1$) distributions are special instances of α-stable distribution family. Inspired by [14, 15], we develop one scan 1-bit compressed sensing by using small α (e.g., $\alpha = 0.05$) and only the sign information (i.e., $sgn(y_j)$) of the measurements. Compared to alternatives, the proposed method is fast and accurate.

The problem of 1-bit compressed sensing has been studied in the literature of statistics, information theory and machine learning, e.g., [1, 11, 9, 20, 4, 22]. 1-bit compressed sensing has many advantages. When the measurements are collected, the hardware will anyway have to quantize the measurements. Also, using only the signs will potentially reduce the cost of storage and transmission (if the number of measurements does not have to increase too much).

Appearing in Proceedings of the 19^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2016, Cadiz, Spain. JMLR: W&CP volume 51.

In Section 6, our empirical comparisons with 1-bit marginal regression [20, 22] illustrate that the proposed method needs orders of magnitude fewer measurements. Compared to 1-bit Iterative Hard Thresholding (IHT) [11], our algorithm is still significantly more accurate. Furthermore, while our method is reasonable robust against random sign flipping, IHT is known to be very sensitive to that kind of noise.

A distinct advantage of our proposed method is that, largely due to the one-scan nature, we can precisely analyze the algorithm with or without random flipping noise; and we can provide the precise constants of the bounds. For example, even for a conservative version of our algorithm, the required number of measurements would be no more than $12.3K \log N/\delta$ (and the practical performance is even better). Here δ (e.g., 0.05) is the usual notation for confidence.

The method of Gaussian (i.e., $\alpha = 2$) random projections has become popular in machine learning and information theory (e.g., [8]). The use of α-stable random projections was previously studied in the context of estimating the l_α norms (e.g., $\sum_{i=1}^N |x_i|^\alpha$) of **data streams**, in the theory literature [10, 12] as well as in machine learning venue [13]. Consequently, our 1-bit CS algorithm also inherits the advantage when the data (signals) arrive as streams [17].

The recent work [14, 15] used α-stable projections with very small α to recover sparse signals, with many significant advantages: (i) the algorithm needs only one scan; (ii) the method is extremely robust against measurement noises (due to the heavy-tailed nature of the projections); and (iii) the recovery procedure is per coordinate in that even when there are no sufficient measurements, a significant portion of the nonzero coordinates can still be recovered. The major disadvantage of [14, 15] is that, since the measurements are also heavy-tailed, the required storage for the measurements might be substantial. Our proposed 1-bit algorithm provides one practical (and simple) solution.

2 The Proposed Algorithm

In our algorithm, the entries (i.e., s_{ij}) of the design matrix are sampled from i.i.d. α-stable with unit scale, denoted by $S(\alpha, 1)$. We can follow the classical procedure to generate samples [3] from $S(\alpha, 1)$. That is, we first sample independent exponential $w \sim exp(1)$ and uninform $u \sim unif(-\pi/2, \pi/2)$ variables, then

$$g(u, w; \alpha) = \frac{\sin(\alpha u)}{(\cos u)^{1/\alpha}} \left[\frac{\cos(u - \alpha u)}{w}\right]^{(1-\alpha)/\alpha} \sim S(\alpha, 1) \quad (1)$$

There are excellent books on stable distributions, e.g., [24, 21]. Basically, if $Z \sim S(\alpha, 1)$, then its characteristic function is $E\left(e^{\sqrt{-1}Zt}\right) = e^{-|t|^\alpha}$. However, closed-form expressions of the density exists only for $\alpha = 2$ (i.e., Gaussian), $\alpha = 1$ (i.e., Cauchy), or $\alpha = 0+$.

Algorithm 1 Stable measurement collection and the one scan 1-bit algorithm for sign recovery.

Input: K-sparse signal $\mathbf{x} \in \mathbb{R}^{1\times N}$, design matrix $\mathbf{S} \in \mathbb{R}^{N\times M}$ with entries sampled from $S(\alpha, 1)$ with small α (e.g., $\alpha = 0.05$). To generate the (i, j)-th entry s_{ij}, we sample $u_{ij} \sim uniform(-\pi/2,\ \pi/2)$ and $w_{ij} \sim exp(1)$ and compute $s_{ij} = g(u_{ij}, w_{ij}; \alpha)$ by (1).

Collect: measurements $y_j = \sum_{i=1}^N x_i s_{ij}$, $j = 1$ to M.
Compute: for each coordinate $i = 1$ to N, compute

$$Q_i^+ = \sum_{j=1}^M \log\left(1 + sgn(y_j)sgn(u_{ij})e^{-(K-1)w_{ij}}\right),$$

$$Q_i^- = \sum_{j=1}^M \log\left(1 - sgn(y_j)sgn(u_{ij})e^{-(K-1)w_{ij}}\right)$$

Output: for $i = 1$ to N, report the estimated sign:

$$\hat{sgn}(x_i) = \begin{cases} +1 & \text{if } Q_i^+ > 0 \\ -1 & \text{if } Q_i^- > 0 \\ 0 & \text{if } Q_i^+ < 0 \text{ and } Q_i^- < 0 \end{cases}$$

Alg. 1 summarizes our one-scan algorithm for recovering the signs of sparse signals. The central component is to compute Q_i^+ and Q_i^-, for $i = 1$ to N, where

$$Q_i^+ = \sum_{j=1}^M \log\left(1 + sgn(y_j)sgn(u_{ij})e^{-(K-1)w_{ij}}\right) \quad (2)$$

$$Q_i^- = \sum_{j=1}^M \log\left(1 - sgn(y_j)sgn(u_{ij})e^{-(K-1)w_{ij}}\right) \quad (3)$$

Later we will explain that it makes no essential difference if we replace $sgn(u_{ij})$ with $sgn(s_{ij})$ and w_{ij} with $1/|s_{ij}|^\alpha$. The parameter α should be reasonably small, e.g., $\alpha = 0.05$. In many prior studies of compressed sensing, K is often assumed to be known. When K is not known, we will need to develop efficient algorithms for the purpose of estimating K.

To make the theoretical analysis easier, Alg. 1 uses "0" as the threshold for estimating the sign:

$$\hat{sgn}(x_i) = \begin{cases} +1 & \text{if } Q_i^+ > 0 \\ -1 & \text{if } Q_i^- > 0 \\ 0 & \text{if } Q_i^+ < 0 \text{ and } Q_i^- < 0 \end{cases} \quad (4)$$

Later in the paper, Lemma 1 will show that at most one of Q_i^+ and Q_i^- can be positive. Using 0 as the threshold simplifies the analysis. As will be shown in our experiments, a practical variant will reduce the number of measurements predicted by the analysis.

Note that, unless the signal is ternary (i.e., $x_i \in \{-1, 0, 1\}$), we will need another procedure for estimating the values of the nonzero entries. A simple strategy is to do a least square on the reported coordinates, by collecting K additional measurements.

Next, we will present the intuition and theory for the proposed algorithm.

3 Intuition

Our proposed algorithm, through the use of Q_i^+ and Q_i^-, is based on the joint likelihood of $(sgn(y_j), s_{ij})$. Denote the density function of $S(\alpha, 1)$ by $f_S(s)$. Recall

$$y_j = \sum_{t=1}^{N} x_t s_{tj} = x_i s_{ij} + \sum_{t \neq i} x_t s_{tj} = x_i s_{ij} + \theta_i S_j$$

where $S_j \sim S(\alpha, 1)$ is independent of s_{ij} and $\theta_i = \left(\sum_{t \neq i} |x_t|^\alpha\right)^{1/\alpha}$. Using a conditional probability argument, the joint density of (y_j, s_{ij}) can be shown to be $\frac{1}{\theta_i} f_S(s_{ij}) f_S\left(\frac{y_j - x_i s_{ij}}{\theta_i}\right)$. Now, suppose we only use (store) the sign information of y_j. We have

$$\mathbf{Pr}\left(y_j > 0, s_{ij}\right) = \int_0^\infty \frac{1}{\theta_i} f_S(s_{ij}) f_S\left(\frac{y - x_i s_{ij}}{\theta_i}\right) dy$$

$$= f_S(s_{ij})\left(1 - F_S\left(\frac{-x_i s_{ij}}{\theta_i}\right)\right) = f_S(s_{ij}) F_S\left(\frac{x_i s_{ij}}{\theta_i}\right)$$

where F_S is the cumulative distribution function (cdf) of $S(\alpha, 1)$. Similarly, we also have

$$\mathbf{Pr}\left(y_j < 0, s_{ij}\right) = f_S(s_{ij}) F_S\left(-\frac{x_i s_{ij}}{\theta_i}\right)$$

which means the joint log-likelihood is proportional to $l(x_i, \theta_i) = \sum_{j=1}^{M} \log F_S\left(sgn(y_j)\frac{x_i s_{ij}}{\theta_i}\right)$.

As we use small α, we can take advantage of the limit at $\alpha = 0+$. Suppose $u \sim uniform(-\pi/2, \pi/2)$ and $w \sim exp(1)$. From (1), $Z = g(u, w; \alpha) \approx sgn(u)/w^{1/\alpha}$. In other words, as $\alpha \to 0+$, $1/|Z|^\alpha \sim exp(1)$. This fact was originally established by [6]. Thus, as $\alpha \to 0+$, we have $F_S(s) = \frac{1}{2} + sgn(s)\frac{1}{2} e^{-|s|^{-\alpha}}$, which leads to

$$l(x_i, \theta_i) = \sum_{j=1}^{M} \log\left(1 + sgn(s_{ij} x_i y_j) \exp\left(-\left|\frac{\theta_i}{x_i s_{ij}}\right|^\alpha\right)\right)$$

Clearly, if $x_i = 0$, then $l(x_i, \theta_i) = 0$. It is thus convenient to use 0 as the threshold. We can then use the following Q_i^+ and Q_i^- to determine if $x_i > 0$ or $x_i < 0$:

$$Q_i^+ = \sum_{j=1}^{M} \log\left(1 + sgn(s_{ij} y_j) \exp\left(-\frac{K-1}{|s_{ij}|^\alpha}\right)\right),$$

$$Q_i^- = \sum_{j=1}^{M} \log\left(1 - sgn(s_{ij} y_j) \exp\left(-\frac{K-1}{|s_{ij}|^\alpha}\right)\right)$$

As $\alpha \to 0+$, we have $\theta_i^\alpha = K - 1$ (if $x_i \neq 0$) or K (if $x_i = 0$). Also note that $|x_i|^\alpha = 0$ (if $x_i = 0$) or 1 (if $x_i \neq 0$). Because $sgn(s_{ij}) = sgn(u_{ij})$ and $\frac{1}{|s_{ij}|^\alpha}$ becomes w_{ij}, we can write them as

$$Q_i^+ = \sum_{j=1}^{M} \log\left(1 + sgn(y_j) sgn(u_{ij}) e^{-(K-1)w_{ij}}\right),$$

$$Q_i^- = \sum_{j=1}^{M} \log\left(1 - sgn(y_j) sgn(u_{ij}) e^{-(K-1)w_{ij}}\right)$$

This explains (2) and (3) in Alg. 1.

So far, we have explained the idea behind our proposed Alg. 1. Next we will conduct further theoretical analysis for the error probabilities and consequently the sample complexity bound.

4 Analysis

Our analysis will use the fact that $sgn(s_{ij} y_j) = sgn(y_j/s_{ij}) = sgn(x_i + \theta_i S_j/s_{ij})$, where $S_j \sim S(\alpha, 1)$ is independent of s_{ij} and $\theta_i = \left(\sum_{t \neq i} |x_t|^\alpha\right)^{1/\alpha}$. Note that both s_{ij} and y_j are symmetric random variables. Our first lemma says that at most one of Q_i^+ and Q_i^-, respectively defined in (2) and (3), can be positive.

Lemma 1 *If $Q_i^+ > 0$ then $Q_i^- < 0$. If $Q_i^- > 0$ then $Q_i^+ < 0$.*

Proof: *It is more convenient to examine $e^{Q_i^+}$ and $e^{Q_i^-}$. Let $z_j = e^{-(K-1)w_{ij}}$. Note that $0 < z_j < 1$. Now suppose $e^{Q_i^+} > 1$. We divide the coordinates, $j = 1$ to M, into two disjoint sets I and II, such that*

$$e^{Q_i^+} = \prod_{j \in I} |1 + z_j| \prod_{j \in II} |1 - z_j| > 1$$

As $\frac{1}{1-z_j} > 1 + z_j$ and $\frac{1}{1+z_j} > 1 - z_j$, we must have

$$\prod_{j \in I} \left|\frac{1}{1 - z_j}\right| \prod_{j \in II} \left|\frac{1}{1 + z_j}\right| > \prod_{j \in I} |1 + z_j| \prod_{j \in II} |1 - z_j| > 1$$

and

$$e^{Q_i^-} = \prod_{j \in I} |1 - z_j| \prod_{j \in II} |1 + z_j| < 1$$

This completes the proof. □

Although it is convenient to use 0 as the threshold, we will provide more general error bounds by comparing Q_i^+ and Q_i^- with $\epsilon M/K$. The following intuition might

be helpful to see why M/K is the right scale:

$$\begin{aligned}|Q_i^+| =& \left|\sum_{j=1}^{M} \log\left(1+sgn(y_j/s_{ij})\exp\left(-(K-1)w_{ij}\right)\right)\right| \\ \le& \sum_{j=1}^{M} \left|\log\left(1+sgn(y_j/s_{ij})\exp\left(-(K-1)w_{ij}\right)\right)\right| \\ \approx& \sum_{j=1}^{M} \exp\left(-(K-1)w_{ij}\right)\end{aligned}$$

By the moment generating function of the exponential distribution, we obtain

$$\begin{aligned}&E\left(\sum_{j=1}^{M}\exp\left(-(K-1)w_{ij}\right)\right) \\ =&\sum_{j=1}^{M} E\left(\exp\left(-(K-1)w_{ij}\right)\right) = \frac{M}{(1+K-1)} = \frac{M}{K}\end{aligned}$$

Lemma 2 concerns the error probability (i.e., false positive) when $x_i = 0$ and $\epsilon M/K$ is used as the threshold.

Lemma 2 *For any ϵ and any $t \ge 0$, we have*

$$\begin{aligned}&\mathbf{Pr}\left(Q_i^+ > \epsilon M/K, x_i = 0\right) = \mathbf{Pr}\left(Q_i^- > \epsilon M/K, x_i = 0\right) \\ \le& \exp\left\{-\frac{M}{K}H_1(t;\epsilon,K)\right\} \qquad (5)\end{aligned}$$

where

$$\begin{aligned}&H_1(t;\epsilon,K) \qquad (6)\\ =&\epsilon t - K\log\left(1+\sum_{n=2,4,6,\ldots}^{\infty}\frac{1}{nK-n+1}\prod_{l=0}^{n-1}\frac{t-l}{n-l}\right)\end{aligned}$$

In the limit as $K \to \infty$, we have

$$H_1(t;\epsilon,\infty) = \epsilon t - \sum_{n=2,4,6,\ldots}^{\infty}\frac{1}{n}\prod_{l=0}^{n-1}\frac{t-l}{n-l} \qquad (7)$$

Proof: *See Appendix A.* □

To minimize the error probability in Lemma 2, we need to seek the optimum (maximum) values of H_1 for given ϵ and K. Figure 1 plots the optimum values $t = t_1^*$ as well as the optimum values of H_1^* for $K = 5$ to 100. As expected, these optimum values are insensitive to K (in fact, no essential difference from the limiting case of $K \to \infty$). At $\epsilon = 0$, the value of $1/H_1^*$ is about 12.2. Note that to control the error probability to be $< \delta$, the required number of measurements will be $M \ge \frac{K}{H_1^*}\log N/\delta$. Thus we use a numerical number 12.3 for the bound of the sample complexity.

Next, Lemma 3 concerns the false negative error probability when $x_i \ne 0$.

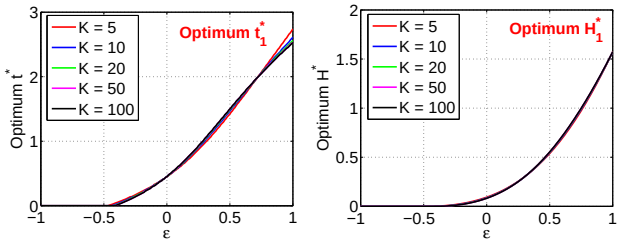

Figure 1: For Lemma 2, we plot the optimum $t = t_1^*$ values (left panel) which maximizes $H_1(t;\epsilon,K)$, as well as $H_1 = H_1^*$ at $t = t_1^*$ (right panel), for $K = 5$ to 100. The different curves essentially overlap. At the threshold $\epsilon = 0$, the value $1/H_1^*$ is about 12.2 (and smaller than 12.3).

Lemma 3 *For any ϵ, $0 < t < 1$, and $\alpha \to 0$, we have*

$$\begin{aligned}&\mathbf{Pr}\left(Q_i^+ < \epsilon M/K, x_i > 0\right) = \mathbf{Pr}\left(Q_i^- < \epsilon M/K, x_i < 0\right) \\ \le& \exp\left(-\frac{M}{K}H_2(t;\epsilon,K)\right) \qquad (8)\end{aligned}$$

where

$$H_2(t;\epsilon,K) = -\epsilon t - K\times\log[A] \qquad (9)$$

$$\begin{aligned}A =& 1+\sum_{n=2,4,6\ldots}^{\infty}\frac{1}{n(K-1)+1}\prod_{l=0}^{n-1}\frac{t-l}{n-l} \\ &-\sum_{n=1,3,5\ldots}^{\infty}\frac{1}{(n+1)(K-1)+1}\prod_{l=0}^{n-1}\frac{t-l}{n-l}\end{aligned}$$

and

$$\begin{aligned}H_2(t;\epsilon,\infty) =& -\epsilon t - \sum_{n=2,4,6\ldots}^{\infty}\frac{1}{n}\prod_{l=0}^{n-1}\frac{t-l}{n-l} \qquad (10)\\ &+\sum_{n=1,3,5\ldots}^{\infty}\frac{1}{(n+1)}\prod_{l=0}^{n-1}\frac{t-l}{n-l}\end{aligned}$$

Proof: *See Appendix B.* □

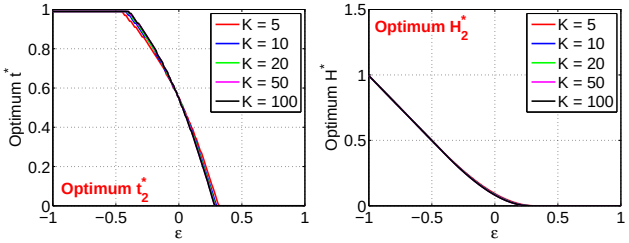

Figure 2: For Lemma 3, we plot the optimum $t = t_2^*$ values (left panel) which maximizes $H_2(t;\epsilon,K)$, as well as $H_2 = H_2^*$ at $t = t_2^*$ (right panel), for $K = 5$ to 100. The different curves essentially overlap. At $\epsilon = 0$, the value of $1/H_2^*$ is again about 12.2 (which is smaller than 12.3).

Figure 2 plots the optimum t_2^* values which maximize H_2, together with the optimum H_2^* values. Interestingly, when $\epsilon = 0$, the value of $1/H_2^*$ is also about 12.2

(smaller than 12.3). This is not surprising, because, for both $H_1(t;\epsilon,\infty)$ and $H_2(t;\epsilon,\infty)$, the leading term at $\epsilon = 0$ is $\frac{t(t-1)}{4}$.

Sample Complexity. Given K, N, ϵ, δ, the required number measurements can be computed from

$$(N-K) \times \mathbf{Pr}\left(Q_i^+ > \epsilon M/K, x_i = 0\right) + K \times \mathbf{Pr}\left(Q_i^+ < \epsilon M/K, x_i > 0\right) \leq \delta$$

When $\epsilon = 0$, because the constants of both error probabilities are upper bounded by 12.3, we have a convenient expression of complexity, as in Theorem 1.

Theorem 1 *Using Alg. 1, in order for the total error (for estimating the signs) of all the coordinates to be bounded by some $\delta > 0$, it suffices to use $M = \lceil 12.3K \log N/\delta \rceil$ measurements.*

5 Recovery Under Noise

We can add measurement noises: $y_j = \sum_{i=1}^N x_i s_{ij} + n_j$, where typically $n_j \sim N(0, \sigma^2)$ at some noise level σ. The framework of sparse recovery using α-stable random projections with small α is (boringly) robust against this type of measurement noises [15]. To make the study more interesting, we consider another common noise model for 1-bit compressed sensing by randomly flipping the signs of the measurements.

That is, we introduce independent variables r_j, $j = 1$ to M, so that $r_j = 1$ with probability $1-\gamma$ and $r_j = -1$ with probability γ. During recovery, we use $(r_j y_j)$ to replace the original y_j. For differentiation, we use $Q^+_{i,\gamma}$ and $Q^-_{i,\gamma}$, respectively, to replace Q_i^+ and Q_i^-.

Interestingly, Lemma 4 shows that random flipping does not affect the false positive probability.

Lemma 4 *For any ϵ and any $t \geq 0$, we have*

$$\mathbf{Pr}\left(Q^+_{i,\gamma} > \epsilon M/K, x_i = 0\right) = \mathbf{Pr}\left(Q^-_{i,\gamma} > \epsilon M/K, x_i = 0\right) \leq \exp\left\{-\frac{M}{K} H_1(t;\epsilon,K)\right\} \quad (11)$$

where $H_1(t;\epsilon,K)$ is the same as in Lemma 2.

Proof: *See Appendix C. The key is that $sgn(r_j u_{ij})$ and $sgn(u_{ij})$ have the same distribution.* □

On the other hand, as shown in the next lemma, this random flipping (with probability γ) does affect the false negative probability.

Lemma 5 *For any ϵ, $0 < t < 1$, and $\alpha \to 0$, we have*

$$\mathbf{Pr}\left(Q^+_{i,\gamma} < \epsilon M/K, x_i > 0\right) = \mathbf{Pr}\left(Q^-_{i,\gamma} < \epsilon M/K, x_i < 0\right) \leq \exp\left(-\frac{M}{K} H_4(t;\epsilon,K,\gamma)\right) \quad (12)$$

where

$$H_4(t;\epsilon,K,\gamma) = -\epsilon t - K \times \log[B] \quad (13)$$

$$B = 1 + \sum_{n=2,4,6...}^{\infty} \frac{1}{n(K-1)+1} \prod_{l=0}^{n-1} \frac{t-l}{n-l} - \sum_{n=1,3,5...}^{\infty} \frac{1-2\gamma}{(n+1)(K-1)+1} \prod_{l=0}^{n-1} \frac{t-l}{n-l}$$

$$H_4(t;\epsilon,\infty,\gamma) = -\epsilon t - \sum_{n=2,4,6...}^{\infty} \frac{1}{n} \prod_{l=0}^{n-1} \frac{t-l}{n-l} + \sum_{n=1,3,5...}^{\infty} \frac{1-2\gamma}{(n+1)} \prod_{l=0}^{n-1} \frac{t-l}{n-l} \quad (14)$$

Proof: *See Appendix D.* □

From Lemma 4 and Lemma 5, we can numerically compute the required number of measurements for any given N and K. We will also provide an empirical study in Section 6.

6 Experiments and Comparisons

In this section, we provide a series of experimental studies to verify the proposed algorithm. In the literature, the so-called 1-bit marginal regression [20, 22] can be viewed as a one-scan algorithm and hence it is the competitor we should compare our method with. As shown in the experiments, however, the proposed method needs orders of magnitude fewer measurements than 1-bit marginal regression (MR). Thus, to make the empirical study more interesting, we also compare the method with the well-known 1-bit Iterative Hard Thresholding (IHT) [11]. The comparison results show that the proposed algorithm is still significantly more accurate. Furthermore, our method is reasonably robust against random sign flipping, while IHT is known to be very sensitive to that kind of noise.

As mentioned earlier, the proposed method is boringly robust against additive Gaussian noise [15] and thus we focus on random sign flipping noise in the experiments. We should also mention that, in our experiments with 1-bit MR and 1-bit IHT, we always use Gaussian designs. Applying 1-bit MR and IHT with heavy-tailed designs would not lead to meaningful results.

6.1 Experiment Set-up

In our experiments, we generate signals based on the two parameters N and K. We choose $(N,K) \in \{(1000,20),(1000,50),(10000,20),(10000,50)\}$. For

each given N and K, we first randomly select K nonzero coordinates and then assign the values of the nonzero entries according to i.i.d. samples from $N(0, 5^2)$. We then apply the proposed algorithm to recover both the support and the signs of the signal. The number of measurements is set according to

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

where the confidence δ is set to be 0.01. We vary the parameter ζ from 2 to 15. Note that this choice of M is typically a small number compared to N. Recall that, in our analysis, the required number of measurements using criterion (4) is proved to be $12.3K \log N/\delta$, although the actual measurements needed will be smaller by using a practical variant of Alg. 1.

6.2 A Variant of Alg. 1

Although Alg. 1 is convenient for theoretical analysis, the practical performance can be improved by using a simple variant based on ranking, although the theoretical analysis would be more difficult.

Basically, after we have computed Q_i^+ and Q_i^- from (2) and (3), for $i = 1$ to N, instead of using 0 as the threshold, we choose the top-K coordinates ranked by $\max\{Q_i^+, Q_i^-\}$. Among the selected coordinates, if $Q_i^+ > Q_i^-$ (or $Q_i^- > Q_i^+$), then we estimate $sgn(x_i)$ to be positive (or negative). This procedure implicitly utilizes ϵ away from 0 and hence less conservative compared to vanilla Alg. 1 as validated by our experiments. The drawback is that it relies on knowing K. Nevertheless, in CS literature, reporting only the top-K recovered coordinates appears to be a common procedure. This is also used in our experiments for reporting the results of competing 1-bit CS algorithms.

Figure 3 reports the sign recovery errors defined as: $\sum_i^N |\widehat{sgn}(x_i) - sgn(x_i)|/K$, to confirm that the variant performs noticeably better than the original Alg. 1.

In the following experiments, we always only report the performance of this variant of Alg. 1.

6.3 Sign Recovery under Random Sign Flipping Noise

Figure 4 reports the sign recovery errors. In each panel, we report results for 3 different γ values ($\gamma = 0$, 0.1, and 0.2), where γ is the random sign flipping probability. The curves without label (red, if color is available) correspond to $\gamma = 0$ (i.e., no random sign flipping errors). The results in Figure 4 confirm that the proposed method works well as predicted by the theoretical analysis. Moreover, the method is fairly robust against random sign flipping noise.

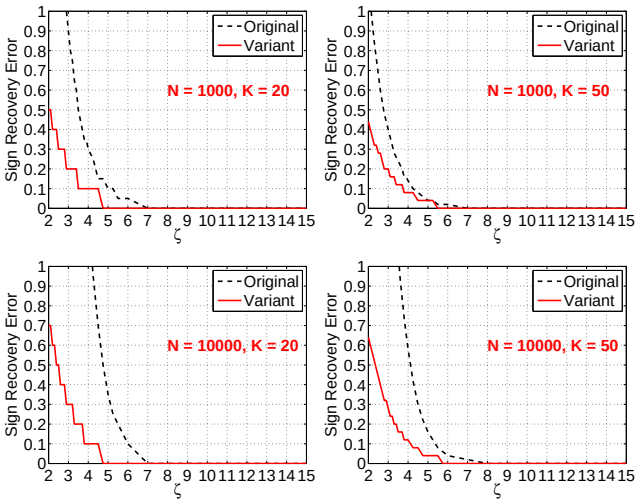

Figure 3: **Sign recovery errors for comparing Alg. 1 with a simple variant**. The number of measurements is chosen according to $\zeta K \log N/\delta$, for ζ ranging from 2 to 15. The recovery error is $\sum_i^N |\widehat{sgn}(x_i) - sgn(x_i)|/K$. We repeat each simulation 1000 times and report the median.

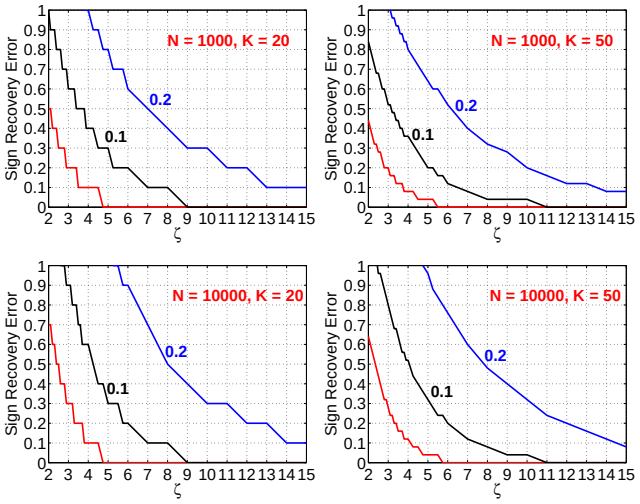

Figure 4: **Sign recovery under random sign flipping noise**. In each panel, the 3 curves correspond to 3 different random sign flipping probability γ, for $\gamma = 0$, 0.1, and 0.2, respectively. The curve without label (red, if color is available) is for $\gamma = 0$. We repeat each simulation 1000 times and report the median.

6.4 Support Recovery

We can generalize the practical variant of Alg. 1. That is, after we rank the coordinates according to $\max\{Q_i^+, Q_i^-\}$, we can choose top-βK coordinates for $\beta \geq 1$. We have used $\beta = 1$ in previous experiments. Figure 5 reports the recall values for support recovery:

$$recall = \#\{retrieved\ true\ nonzeros\}/K$$

for $\beta = 1$, 1.2, and 1.5. Note that in this case we just need to present the recalls, because *precision* =

$\#\{retrieved\ true\ nonzeros\}/(\beta K)$. As expected, using larger β values can reduce the required number of measurements. This experiment could be interesting for practitioners who care about this trade-off.

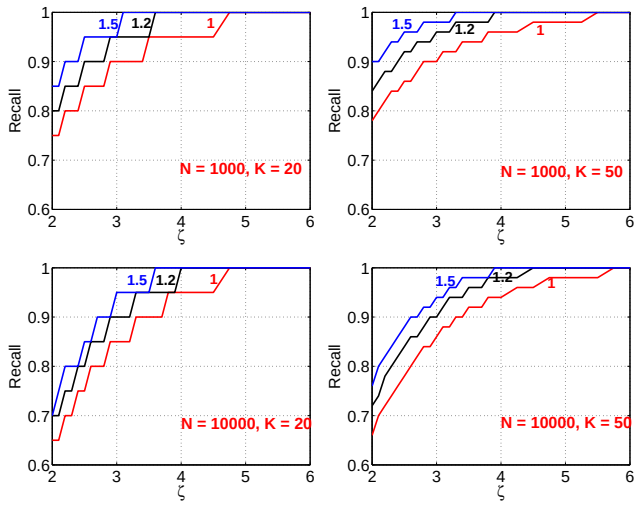

Figure 5: **Support recovery**. We report top-βK coordinates ranked by $\max\{Q_i^+, Q_i^-\}$, for $\beta \in \{1, 1.2, 1.5, 2\}$. We report the **recall** values, i.e., $\#\{retrieved\ true\ nonzeros\}/K$. As expected, using larger β will reduce the required number of measurements, which is set to be $\zeta K \log N/\delta$ (where $\delta = 0.01$).

6.5 Comparisons with 1-bit Marginal Regression

It is helpful to provide a comparison study with other 1-bit algorithms in the literature. Unfortunately, most of those available 1-bit algorithms are not one-scan methods. One exception is the 1-bit marginal regression [20, 22], which can be viewed as a one-scan algorithm. Thus, it is the target competitor we should compare our method with.

Figure 6 reports the sign recovery accuracy of 1-bit marginal regression in our experimental setting. That is, we also choose $M = \zeta K \log N/\delta$, although for this approach, we must enlarge ζ dramatically, compared to our method. We can see that even with $\zeta = 100$, the errors of 1-bit marginal regression are still large.

6.6 Comparisons with 1-bit Iterated Hard Thresholding (IHT)

We conclude this section by providing a comparison with the well-known 1-bit iterative hard thresholding (IHT) [11]. Even though 1-bit IHT is not a one-scan algorithm, we compare it with our method for completeness. As shown in Figure 7, the proposed algorithm is still significantly more accurate for sign recovery.

Note that Figure 7 does not include results of 1-bit IHT with random sign flipping noise. As previously shown, the proposed method is reasonably robust against this type of noise. However, we observe that 1-bit IHT is so sensitive to random sign flipping that the results are probably not presentable [1].

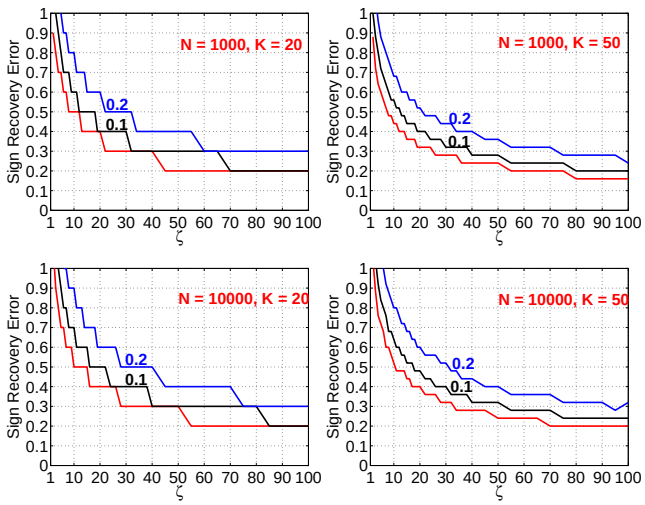

Figure 6: **Sign recovery with 1-bit marginal regression**. The errors are still very larger even with $\zeta = 100$, i.e., $M = 100K \log N/\delta$. Note that in each panel, the three curves correspond to three different random sign flipping probabilities: $\gamma = 0$, 0.1, and 0.2, respectively.

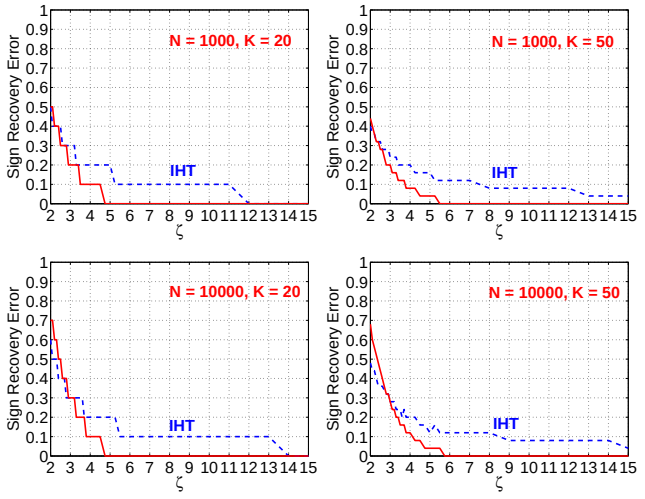

Figure 7: **Sign recovery with 1-bit iterative hard thresholding (IHT)**. The results of 1-bit IHT are presented as dashed (blue, if color is available) curves. For comparison, we also plot the results of the proposed method (solid and red if color is available).

[1] After consulting the author of [11], we decided not to present the random sign flipping experiment for 1-bit IHT.

7 Discussion: Estimation of K

In the theoretical analysis, we have assumed that K is known, like many prior studies in compressed sensing. The problem becomes more interesting when K is not known. In this case, we will need to use (say) m additional measurements for estimating K. One proposal is to use m full (i.e., infinite-bit) measurements such as the estimators developed in [12]. However, using full (infinite-bit) measurements for estimating K would be contradicting the original purpose of 1-bit CS. It is hence an interesting research problem to develop efficient bit-estimators of K. In fact, for any 1-bit CS algorithms, we must address the same issue.

8 Conclusion

1-bit compressed sensing (CS) is an important topic because the measurements are typically quantized (by hardware) and using only the sign information may potentially lead to cost reduction in collection, transmission, storage, and retrieval. Current methods for 1-bit CS are less satisfactory because they require a very large number of measurements and the decoding is typically not one-scan. Inspired by recent method of compressed sensing with heavy-tailed designs, we develop an algorithm for one-scan 1-bit CS, which is provably accurate and fast, as validated by experiments.

For sign recovery, our proposed one-scan 1-bit algorithm requires orders of magnitude fewer measurements compared to 1-bit marginal regression. Our method is still significantly more accurate than 1-bit Iterative Hard Thresholding (IHT), which is not one-scan. Moreover, unlike 1-bit IHT, the proposed algorithm is reasonably robust again random sign flipping noise (for which we also have a theoretical proof).

Acknowledgement

The work is partially supported by NSF-Bigdata-1419210, NSF-III-1360971, ONR-N00014-13-1-0764, and AFOSR-FA9550-13-1-0137. The work was first submitted to NIPS 2014.

References

[1] P. Boufounos and R. Baraniuk. 1-bit compressive sensing. In *Information Sciences and Systems, 2008.*, pages 16–21, March 2008.

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

[3] J. 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] S. Chen and A. Banerjee. One-bit compressed sensing with the k-support norm. In *AISTATS*, 2015.

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

[6] N. Cressie. A note on the behaviour of the stable distributions for small index. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31(1):61–64, 1975.

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

[8] Y. Freund, S. Dasgupta, M. Kabra, and N. Verma. Learning the structure of manifolds using random projections. In *NIPS*, Vancouver, BC, Canada, 2008.

[9] S. Gopi, P. Netrapalli, P. Jain, and A. Nori. One-bit compressed sensing: Provable support and vector recovery. In *ICML*, 2013.

[10] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of ACM*, 53(3):307–323, 2006.

[11] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, 2013.

[12] P. Li. Estimators and tail bounds for dimension reduction in l_α $(0 < \alpha \leq 2)$ using stable random projections. In *SODA*, pages 10 – 19, San Francisco, CA, 2008.

[13] P. Li and T. J. Hastie. A unified near-optimal estimator for dimension reduction in l_α $(0 < \alpha \leq 2)$ using stable random projections. In *NIPS*, Vancouver, BC, Canada, 2007.

[14] P. Li and C.-H. Zhang. Exact sparse recovery with L0 projections. In *KDD*, pages 302–310, 2013.

[15] P. Li, C.-H. Zhang, and T. Zhang. Compressed counting meets compressed sensing. In *COLT*, 2014.

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

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

[18] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[19] Y. 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.

[20] Y. Plan and R. Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1):482–494, 2013.

[21] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, 1994.

[22] M. Slawski and P. Li. b-bit marginal regression. In *NIPS*, Montreal, CA, 2015.

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

[24] V. M. Zolotarev. *One-dimensional Stable Distributions*. American Mathematical Society, Providence, RI, 1986.