Efficient Algorithms for Robust One-bit Compressive Sensing

Lijun Zhang ZHANGLJ@LAMDA.NJU.EDU.CN
National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

Jinfeng Yi JINFENGY@US.IBM.COM
IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

Rong Jin RONGJIN@CSE.MSU.EDU
Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

Abstract

While the conventional compressive sensing assumes measurements of infinite precision, one-bit compressive sensing considers an extreme setting where each measurement is quantized to just a single bit. In this paper, we study the vector recovery problem from noisy one-bit measurements, and develop two novel algorithms with formal theoretical guarantees. First, we propose a passive algorithm, which is very efficient in the sense it only needs to solve a convex optimization problem that has a *closed-form* solution. Despite the apparent simplicity, our theoretical analysis reveals that the proposed algorithm can recover both the *exactly* sparse and the *approximately* sparse vectors. In particular, for a sparse vector with s nonzero elements, the sample complexity is $O(s \log n/\epsilon^2)$, where n is the dimensionality and ϵ is the recovery error. This result improves significantly over the previously best known sample complexity in the noisy setting, which is $O(s \log n/\epsilon^4)$. Second, in the case that the noise model is known, we develop an *adaptive* algorithm based on the principle of active learning. The key idea is to solicit the sign information only when it cannot be inferred from the current estimator. Compared with the passive algorithm, the adaptive one has a lower sample complexity if a high-precision solution is desired.

1. Introduction

Compressive sensing is designed to recover a sparse signal from a small number of linear measurements (Donoho, 2006; Candes & Tao, 2006). A variant of compressive sensing, named one-bit compressive sensing, has attracted considerable interests over the past few years (Boufounos & Baraniuk, 2008). Unlike the conventional compressive sensing which relies on real-valued measurements, in one-bit compressive sensing, each mea-

Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

- Unlike previous studies of one-bit compressive sensing that require solving optimization problems (Plan & Vershynin, 2013b), the proposed algorithm has a closed-form solution, making it computationally attractive.
- Our analysis shows that in the case of noisy one-bit measure, the proposed algorithm improves the sample complexity from $O(s\log n/\epsilon^4)$ to $O(s\log n/\epsilon^2)$ when the target signal is an exactly s-sparse n-dimensional vector.
- We develop a novel *adaptive* algorithm to further reduce the number of one-bit measurements. When the noisy model is known, the proposed adaptive algorithm improves the sample complexity to $O(\min(s\log n/\epsilon^2, s\sqrt{n}\log n/\epsilon))$ if the target vector is exactly s-sparse and to $O(\min(s\log n/\epsilon^4, s\sqrt{n}\log n/\epsilon^3))$ if the target vector is approximately s-sparse (i.e., $\|\mathbf{x}_*\|_1/\|\mathbf{x}\|_2 \leq \sqrt{s}$).

2. Related Work

One-bit compressive sensing was first introduced in (Boufounos & Baraniuk, 2008), where only the *noiseless* one-bit measure is considered. Let $U = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{R}^{n\times m}$ be a known measurement matrix, and $\mathbf{y} = [y_1, \ldots, y_m]^\top$ be the m-dimensional one-bit measurement, where $y_i = \text{sign}(\mathbf{x}_*^\top \mathbf{u}_i)$. The authors propose to recover the direction of target signal $\mathbf{x}_*$ by solving the following optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{y}\circ(U^\top \mathbf{x}) \geq 0, \|\mathbf{x}\|_2 = 1 \quad (1)$$

where $\circ$ stands for the element-wise product between two vectors. One problem with (1) is that it requires solving a non-convex optimization problem. A provable optimization algorithm was proposed in (Laska et al., 2011) to find a stationary point of (1). However, none of these two works provide a formal guarantee on the sample complexity.

In (Jacques et al., 2013), the authors study a similar formulation by replacing $\|\mathbf{x}\|_1$ in (1) with $\|\mathbf{x}\|_0$, and show a sample complexity of $O(s\log n/\epsilon)$ for recovering the direction of a s-sparse vector. However, it remains unsolved as how to efficiently solve the corresponding non-convex optimization problem is unclear. Gopi et al. (2013) developed an efficient two-stage algorithm for one-bit compressive sensing that achieves a sample complexity of $O(s\log n/\epsilon)$.

The first convex formulation for one-bit compressive sensing was proposed in (Plan & Vershynin, 2013a). It solves the following linear programming problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{y}\circ(U^\top \mathbf{x}) \geq 0, \|U^\top \mathbf{x}\|_1 = m \quad (2)$$

An important property of the formulation in (2) is that it can recover not only the exactly sparse vector but also the approximately sparse vector (i.e., $\|\mathbf{x}_*\|_1/\|\mathbf{x}\|_2 \leq \sqrt{s}$). However, a major drawback of this study is the sample complexity, which is $O(s\log^2 n/\epsilon^5)$, exhibits a very high dependence on $1/\epsilon$.

So far, all the related work discussed above assume the one-bit measure to be perfect (i.e., $y_i = \text{sign}(\mathbf{x}_*^\top \mathbf{u}_i)$). Although several heuristic algorithms (Yan et al., 2012; Movahed et al., 2012; Jacques et al., 2013) were proposed to handle noise in one-bit measure, none of them has theoretical guarantees. The only provable recovery algorithm for robust compressive sensing is given in (Plan & Vershynin, 2013b), where the sparse vector is recovered by solving the following convex optimization problem

$$\max_{\mathbf{x}} \mathbf{x}^\top U\mathbf{y} \text{ s.t. } \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_1 \leq \sqrt{s} \quad (3)$$

Table 1. Sample Complexities of existing algorithms for one-bit compressive sensing.

	$\mathbf{x}_*$ is exactly sparse		$\mathbf{x}_*$ is approximately sparse	
	Sample Complexity	Reference	Sample Complexity	Reference
Noiseless	$O\left(\frac{s\log n}{\epsilon}\right)$	(Jacques et al., 2013) (Gopi et al., 2013)	$O\left(\frac{s\log^2 n}{\epsilon^5}\right)$	(Plan & Vershynin, 2013a)
	$O\left(\frac{s\log^2 n}{\epsilon^5}\right)$	(Plan & Vershynin, 2013a)		
Noisy	$O\left(\frac{s\log n}{\epsilon^4}\right)$	(Plan & Vershynin, 2013b)	$O\left(\frac{s\log n}{\epsilon^4}\right)$	(Plan & Vershynin, 2013b)
	$O\left(\frac{s\log n}{\epsilon^2}\right)$	(Our Passive Algorithm)		(Our Passive Algorithm)
	$O\left(\min\left(\frac{s\log n}{\epsilon^2}, \frac{s\sqrt{n}\log n}{\epsilon}\right)\right)$	(Our Adaptive Algorithm)	$O\left(\min\left(\frac{s\log n}{\epsilon^4}, \frac{s\sqrt{n}\log n}{\epsilon^3}\right)\right)$	(Our Adaptive Algorithm)

independently at random satisfying

$$\mathrm{E}[y_i|\mathbf{u}_i] = \theta(\mathbf{x}_*^\top \mathbf{u}_i), \quad i = 1, \ldots, m \tag{4}$$

where $\theta(z) : \mathbb{R} \mapsto [-1, +1]$ is some nonlinear function that can be unknown. In order to capture the relation between $\mathbf{u}_i$ and y_i, following (Plan & Vershynin, 2013b), we define λ for $\theta(z)$ as follows,

$$\lambda := \mathrm{E}_{g\sim\mathcal{N}(0,1)}[\theta(g)g] \tag{5}$$

where λ measures how well y_i is correlated with $\mathbf{x}_*^\top \mathbf{u}_i$. We assume $\lambda > 0$, implying that a positive correlation between the real-valued measurement and the binary output from $\theta(\cdot)$.

Since we only receive the sign information about the random measurements, it is impossible to recover the scale of $\mathbf{x}_*$. As a result, we will only consider the recovery of the direction of $\mathbf{x}_*$, and therefore assume $\|\mathbf{x}_*\|_2 = 1$.

3.2. Passive Algorithm for 1-bit CS

The proposed algorithm is inspired by the convex formulation in (3). Instead of having a constraint $\|\mathbf{x}\|_1 \le \sqrt{s}$ to ensure a sparse solution, we introduce a ℓ_1 regularizer in the objective function, leading to the following optimization problem

$$\min_{\|\mathbf{x}\|_2\le 1} -\frac{1}{m}\mathbf{x}^\top U\mathbf{y} + \gamma\|\mathbf{x}\|_1 \tag{6}$$

where $\gamma > 0$ is a regularization parameter, whose value will be discussed later. As shown below, the problem in (6) has a closed-form solution.

Define the soft-thresholding operator (Donoho, 1995; Duchi & Singer, 2009) as

$$P_\gamma(\alpha) = \begin{cases} 0, & \text{if } |\alpha| \le \gamma; \\ \mathrm{sign}(\alpha)(|\alpha| - \gamma), & \text{otherwise.} \end{cases} \tag{7}$$

We extend the operator $P_\gamma(\cdot)$ to vectors as

$$P_\gamma([\alpha_1, \ldots, \alpha_m]^\top) = [P_\gamma(\alpha_1), \ldots, P_\gamma(\alpha_m)]^\top.$$

Lemma 1. *Let $\hat{\mathbf{x}}$ be the optimal solution of (6). Then, we have*

$$\hat{\mathbf{x}} = \begin{cases} 0, & \text{if } \left\|\frac{1}{m}U\mathbf{y}\right\|_\infty \le \gamma; \\ \frac{1}{\left\|P_\gamma\left(\frac{1}{m}U\mathbf{y}\right)\right\|_2} P_\gamma\left(\frac{1}{m}U\mathbf{y}\right), & \text{otherwise.} \end{cases}$$

The proof can be found in the supplementary material.

The following theorem provides the recovery rate for the optimal solution to (6).

Theorem 1. *Assume*

$$\gamma = 2c\sqrt{\frac{t + \log n}{m}} \tag{8}$$

for some constant c. If $\mathbf{x}_*$ is exactly sparse, i.e., $\|\mathbf{x}_*\|_0 \le s$, with a probability at least $1 - e^{1-t}$, we have

$$\|\hat{\mathbf{x}} - \mathbf{x}_*\|_2 \le \frac{3\gamma}{\lambda}\sqrt{\|\mathbf{x}_*\|_0} = O\left(\sqrt{\frac{s\log n}{m}}\right).$$

If $\mathbf{x}_*$ is approximately sparse, i.e., $\|\mathbf{x}\|_1 \le \sqrt{s}$, with a probability at least $1 - e^{1-t}$, we have

$$\|\hat{\mathbf{x}} - \mathbf{x}_*\|_2 \le \sqrt{\frac{3\gamma}{\lambda}\|\mathbf{x}_*\|_1} = O\left(\sqrt[4]{\frac{s\log n}{m}}\right).$$

Remark Compared to the result in (Plan & Vershynin, 2013b), the proposed algorithm improves the sample complexity from $O(s\log n/\epsilon^4)$ to $O(s\log n/\epsilon^2)$ when recovering an exactly s-sparse vector from noisy one-bit measurements. In addition, the sample complexity of the proposed algorithm for one-bit compressive sensing matches the minimax rate of conventional compressive sensing (Raskutti et al., 2011) for both exactly sparse and approximately sparse vectors. We however emphasize that

Algorithm 1 An adaptive algorithm for One-bit Compressive Sensing

Input: the number of stages K, the initial sample size m_1, the initial regularizer γ_1, the step size $\eta \in \{2^{1/2}, 2^{1/4}\}$

Let $\mathbf{x}_1$ be any unit vector, $\delta_1 = 1$

for $k = 1$ to K **do**

  Randomly sample m_k Gaussian random vectors $\mathcal{G}_k = \{\mathbf{u}_1^k, \ldots, \mathbf{u}_{m_k}^k\}$.

  Divide $\mathcal{G}_k$ into $\mathcal{A}_k$ and $\mathcal{B}_k$ according to (11)

  For $\mathbf{u}_i^k \in \mathcal{A}_k$, generate the one-bit measurement y_i^k from $\text{sign}(\mathbf{x}_k^\top \mathbf{u}_i^k)$

  For $\mathbf{u}_i^k \in \mathcal{B}_k$, query the Oracle to obtain the one-bit measurement y_i^k

$$\mathbf{x}_{k+1} = \underset{\|\mathbf{x}\|_2 \le 1}{\text{argmin}} -\frac{1}{m_k}\sum_{i=1}^{m_k} y_i^k \mathbf{x}^\top \mathbf{u}_i^k + \gamma_k \|\mathbf{x}\|_1$$

  $m_{k+1} = 2m_k$, $\gamma_{k+1} = \gamma_k/\sqrt{2}$, $\delta_{k+1} = \delta_k/\eta$

end for

Output: the final solution $\mathbf{x}_{K+1}$

the guarantee for conventional compressive sensing algorithm does not directly apply to one-bit compressive sensing because $\mathrm{E}[y_i]$ is not proportional to $\mathbf{x}_*^\top \mathbf{u}_i$. We also note that this sample complexity is better than $O(n/\epsilon^2)$, which is the optimal rate for binary classification in the noisy setting (Anthony & Bartlett, 1999, Theorem 5.2).

3.3. An Adaptive Algorithm for 1-bit CS

The proposed algorithm aims to explore the idea of active learning (Dasgupta, 2011) to reduce the number of one-bit measurements. The key observation is that after observing certain number of one-bit measurements, we can obtain an intermediate solution $\hat{\mathbf{x}}$ that is reasonably close to the direction of the target vector. As a result, for the sequentially sampled random vector $\mathbf{u}$, we would expect $\text{sign}(\hat{\mathbf{x}}^\top \mathbf{u}) = \text{sign}(\mathbf{x}_*^\top \mathbf{u})$ if the direction of $\mathbf{u}$ is close to that of $\hat{\mathbf{x}}$ (or $-\hat{\mathbf{x}}$) and therefore do not need to ask for an one-bit measurement for $\mathbf{u}$. However, it is problematic to directly replace y, the one-bit measurement for $\mathbf{u}$, with $\text{sign}(\hat{\mathbf{x}}^\top \mathbf{u})$ since y is perturbed by random noise. A similar issue was also raised in (Yang & Hanneke, 2013), where the authors propose to re-noise the data to ensure all the measurements follow the same distribution. In this paper, for the sake of simplicity, we make the following assumption:

A1: *We assume that for a vector* $\mathbf{u}$*, if the value of* $\textit{sign}(\mathbf{x}_*^\top \mathbf{u})$ *is provided, we can generate the one-bit measurement* y *without querying the Oracle.*

One possible noise model is

$$y = \xi\,\text{sign}(\mathbf{x}_*^\top \mathbf{u}), \tag{9}$$

where ξ is a independent $\{-1, 1\}$ valued random variable with $\Pr(\xi = -1) = p$, representing random bit flips (Plan & Vershynin, 2013b). It is straightforward to generate the one-bit measurement y if both $\text{sign}(\mathbf{x}_*^\top \mathbf{u})$ and p are provided.

The complete procedure is provided in Algorithm 1. Our algorithm is closely related to the epoch gradient algorithm developed for stochastic optimization (Hazan & Kale, 2011). It divides the recovery process into K stages. At each stage $k > 1$, we assume that an approximate solution $\mathbf{x}_k$ is obtained from the previous stage with

$$\|\mathbf{x}_k\|_2 = 1, \text{ and } \|\mathbf{x}_k - \mathbf{x}_*\|_2 \le \delta_k. \tag{10}$$

Let $\mathcal{G}_k = \{\mathbf{u}_1^k, \ldots, \mathbf{u}_{m_k}^k\}$ be a set of m_k vectors that are independently sampled from Gaussian distribution. We divide the set $\mathcal{G}_k$ into two subsets:

$$\begin{aligned} \mathcal{A}_k &= \left\{\mathbf{u}_i^k : \left|\mathbf{x}_k^\top \frac{\mathbf{u}_i^k}{\|\mathbf{u}_i^k\|_2}\right| > \delta_k\right\}, \\ \mathcal{B}_k &= \left\{\mathbf{u}_i^k : \left|\mathbf{x}_k^\top \frac{\mathbf{u}_i^k}{\|\mathbf{u}_i^k\|_2}\right| \le \delta_k\right\}. \end{aligned} \tag{11}$$

where $\mathcal{A}_k$ includes random vectors whose directions are close to $\mathbf{x}_k$ or $-\mathbf{x}_k$ while $\mathcal{B}_k$ includes those that are far away from $\mathbf{x}_k$ and $-\mathbf{x}_k$. The following Lemma reveals an important property of $\mathcal{A}_k$

Lemma 2. *Under the condition in (10), we have*

$$\textit{sign}(\mathbf{x}_*^\top \mathbf{u}) = \textit{sign}(\mathbf{x}_k^\top \mathbf{u}), \ \forall \mathbf{u} \in \mathcal{A}_k.$$

Since for any $\mathbf{u} \in \mathcal{A}_k$, $\text{sign}(\mathbf{x}_*^\top \mathbf{u})$ can be inferred from $\text{sign}(\mathbf{x}_k^\top \mathbf{u})$, we can skip one-bit measurement for any $\mathbf{u} \in \mathcal{A}_k$ and reduce the number of one-bit measurements.

We now discuss the recover property of Algorithm 1. For the case that $\mathbf{x}_*$ is exactly sparse, we have the following theorem for the adaptive algorithm.

Theorem 2. *Suppose* $\mathbf{x}_*$ *is exactly sparse, i.e.,* $\|\mathbf{x}_*\|_0 \le s$*, and assumption* **A1** *holds. Let*

$$m_1 = \frac{72c^2 s(t + \log n)}{\lambda^2}, \ \gamma_1 = \frac{\lambda}{3\sqrt{2s}}, \ \eta = 2^{1/2}$$

where c *is the constant in Theorem 1. Then, with a probability at least* $1 - Ke^{1-t}$*, we have*

$$\|\mathbf{x}_{K+1} - \hat{\mathbf{x}}\|_2 \le \delta_{K+1} = \frac{1}{2^{K/2}}.$$

Furthermore, with a probability at least $1 - (e+1)(K-1)e^{-t}$*, the number of calls to the Oracle* $\sum_{k=1}^K |\mathcal{B}_k|$ *is bounded by*

$$\min\left\{2(K-1)t + (5\sqrt{n}2^{K/2} + 1)m_1, m_1 2^K\right\}.$$

The above theorem immediately implies the following corollary.

Corollary 1. *Under the condition in Theorem 2, the recovery rate of the adaptive algorithm is*

$$O\left(\min\left\{\sqrt{\frac{s\log n}{m}}, \frac{s\sqrt{n}\log n}{m}\right\}\right),$$

where $m = \sum_{k=1}^{K} |\mathcal{B}_k|$ is the total number of measurements. And thus the sample complexity is

$$O\left(\min\left\{\frac{s\log n}{\epsilon^2}, \frac{s\sqrt{n}\log n}{\epsilon}\right\}\right).$$

Remark As a result, the sample complexity of the adap-

Thus,

$$
\begin{aligned}
&\lambda(1-\boxtimes^\top \mathbf{x}_*) + \frac{\gamma}{2}\|\mathbf{P}_{\bar{\mathcal{S}}}(\boxtimes)\|_1 \\
\le\, &\gamma\|\mathbf{x}_*\|_1 - \gamma\|\mathbf{P}_{\mathcal{S}}(\boxtimes)\|_1 + \frac{\gamma}{2}\|\mathbf{P}_{\mathcal{S}}(\mathbf{x}_* - \boxtimes)\|_1 \\
\le\, &\frac{3\gamma}{2}\|\mathbf{P}_{\mathcal{S}}(\mathbf{x}_* - \boxtimes)\|_1 \le \frac{3\gamma}{2}\sqrt{\|\mathbf{x}_*\|_0}\|\mathbf{P}_{\mathcal{S}}(\mathbf{x}_* - \boxtimes)\|_2.
\end{aligned}
$$

Then, we have

$$
\|\mathbf{x}_* - \boxtimes\|_2^2 \le 2(1-\boxtimes^\top \mathbf{x}_*) \le \frac{3\gamma}{\lambda}\sqrt{\|\mathbf{x}_*\|_0}\|\mathbf{x}_* - \boxtimes\|_2
$$

which implies

$$
\|\mathbf{x}_* - \boxtimes\|_2 \le \frac{3\gamma}{\lambda}\sqrt{\|\mathbf{x}_*\|_0}.
$$

Next, we consider the case that $\mathbf{x}_*$ is approximately sparse, i.e., $\|\mathbf{x}_*\|_1 \le \sqrt{s}$. From (12), we have

$$
\begin{aligned}
&\lambda(1-\boxtimes^\top \mathbf{x}_*) \\
\le\, &\gamma\|\mathbf{x}_*\|_1 - \gamma\|\boxtimes\|_1 + \frac{\gamma}{2}\|\mathbf{x}_*\|_1 + \frac{\gamma}{2}\|\boxtimes\|_1 \le \frac{3\gamma}{2}\|\mathbf{x}_*\|_1.
\end{aligned}
$$

Thus,

$$
\|\mathbf{x}_* - \boxtimes\|_2^2 \le 2(1-\boxtimes^\top \mathbf{x}_*) \le \frac{3\gamma}{\lambda}\|\mathbf{x}_*\|_1.
$$

4.2. Proof of Theorem 2

From the updating rule in our algorithm, it is easy to check that

$$
\delta_k = \frac{1}{2^{(k-1)/2}}, \ \gamma_k = 2c\sqrt{\frac{t+\log n}{m_k}}, \ \forall k.
$$

So, the condition (8) in Theorem 1 is satisfied at each state.

We first consider the first stage. Since $\|\mathbf{x}_1\| = 1$ and $\delta_1 = 1$, the definitions in (11) ensures $\mathcal{B}_1 = \mathcal{G}_1$. And thus we will query the Oracle to obtain the one-bit measurements for all the elements in $\mathcal{G}_1$. As a result, we can apply Theorem 1 to bound the recovery error of $\mathbf{x}_2$. Specifically, with a probability at least $1 - e^{1-t}$, we have

$$
\|\mathbf{x}_2 - \boxtimes\|_2 \le \frac{3\gamma_1}{\lambda}\sqrt{s} = \frac{1}{\sqrt{2}} = \delta_2.
$$

Thus, the condition in (10) is true for $k = 2$. Based on Lemma 2, we can apply Theorem 1 again and get

$$
\|\mathbf{x}_3 - \boxtimes\|_2 \le \frac{3\gamma_2}{\lambda}\sqrt{s} = \frac{3\gamma_1}{\lambda}\sqrt{s}\frac{1}{\sqrt{2}} = \frac{\delta_2}{\sqrt{2}} = \delta_3.
$$

Repeating the above argument for all the stages, we obtain the first part of the theorem.

Now, we consider bounding the size of $\mathcal{B}_k$. Since $\mathcal{B}_1 = \mathcal{G}_1$, we have

$$
|\mathcal{B}_1| = m_1.
$$

For $k = 2$, we have with a probability at least $1-e^{1-t}$, (10) holds. We condition on the event that (10) is true, and proceed by analyzing the distribution of $\mathbf{x}_2^\top \mathbf{u}_i^2 / \|\mathbf{u}_i^2\|_2$ appears in the definition of $\mathcal{B}_2$. Since $\mathbf{u}_i^2$ is a Gaussian random vector, it is known that $\mathbf{u}_i^2/\|\mathbf{u}_i^2\|_2$ is uniformly distributed on the $n-11241001133500242335573$(o)013335]444042(a)044(t)11241135()0

4.3. Proof of Lemma 3

We need the following lemma on the expectation of $\mathbf{u}_i y_i$.

Lemma 4.

$$\mathrm{E}\left[\mathbf{u}_i y_i\right] = \lambda \mathbf{x}_*, \; i = 1, \ldots n.$$

Consider the j-th element of $\frac{1}{m} U \mathbf{y} - \lambda \mathbf{x}_*$, that is,

$$\left[\frac{1}{m} U \mathbf{y} - \lambda \mathbf{x}_*\right]_j = \frac{1}{m}\sum_{i=1}^{m} u_i^j y_i - \lambda x_*^j,$$

where u_i^j and x_*^j are the j-th element of $\mathbf{u}_i$ and $\mathbf{x}_*$, respectively.

Lemma 4 implies $\mathrm{E}[u_i^j y_i] = \lambda x_*^j$. From (Vershynin, 2012, Remark 5.18), we have

$$\left\| u_i^j y_i - \lambda x_*^j \right\|_{\psi_2} \leq 2 \left\| u_i^j y_i \right\|_{\psi_2} \tag{13}$$

where

$$\|X\|_{\psi_2} = \sup_{p \geq 1} p^{-1/2} (\mathrm{E}|X|^p)^{1/p}$$

is the sub-gaussian norm of random variable X (Vershynin, 2012, Definition 5.7). Since $y_i \in \{\pm 1\}$, we have

$$\left\| u_i^j y_i \right\|_{\psi_2} = \left\| u_i^j \right\|_{\psi_2} \leq c \tag{14}$$

where $c > 0$ is an absolute constant, and the last inequality follows from $u_i^j \sim \mathcal{N}(0, 1)$ and (Vershynin, 2012, Example 5.8).

We will use the Hoeffding-type inequality for sub-gaussian random variables given below.

Lemma 5. *(Vershynin, 2012, Proposition 5.10) Let $X_1, \ldots, X_N$ be independent centered sub-gaussian random variables, and let $K = \max_i \|X_i\|_{\psi_2}$. Then, for any $\alpha = [\alpha_1, \ldots, \alpha_N]^\top \in \mathbb{R}^N$ and every $t \geq 0$, we have*

$$\Pr\left(\left| \sum_{i=1}^{N} \alpha_i X_i \right| \geq t \right) \leq \exp\left(1 - \frac{ct^2}{K^2 |\alpha|_2^2} \right)$$

where $c > 0$ is an absolute constant.

Combining Lemma 5 with (13) and (14), we have with a probability at least $1 - e^{1-t}$,

$$\left| \frac{1}{m}\sum_{i=1}^{m} u_i^j y_i - \lambda x_*^j \right| \leq c\sqrt{\frac{t}{m}}$$

for some constant $c > 0$. We complete the proof by taking the union bound over $j = 1, \ldots, n$.

5. Experiments

In this section, we perform the recovery experiment to verify our theoretical claims. Due to space limitations, we only provide results for the exactly sparse vectors.

Table 2. Running time of each algorithm, when s=10, $n = 1000$, and $m = 1000$. For BIHT and BIHT-ℓ_2, there is no formal stopping criterion, and we report the running time after 100 iterations.

	PASSIVE	BIHT	BIHT-ℓ_2	CONVEX
TIME (S)	$1.1e{-3}$	1.7	1.7	0.72

Experimental Setup We generate the target vector $\mathbf{x}_* \in \mathbb{R}^n$ by drawing its nonzero elements from the standard Gaussian distribution, and then normalize it to have unit length. The locations of the s nonzero elements of $\mathbf{x}_*$ are randomly selected. The elements in the matrix matrix $U \in \mathbb{R}^{n \times m}$ are also drawn from the standard Gaussian distribution. To generate noisy measurements, we choose the observation model in (9), where the sign of $\mathbf{u}_i^\top \mathbf{x}_*$ is flipped with probability p

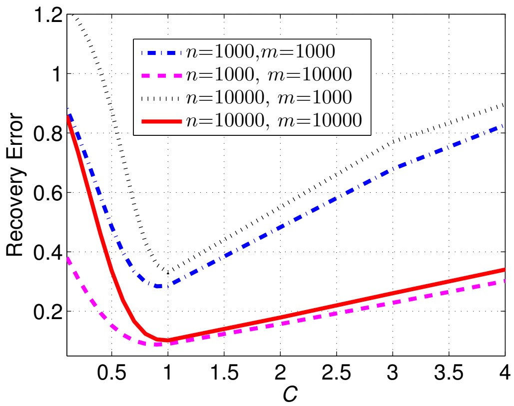


Figure 1. The recovery error of the passive algorithm versus to C, when $\gamma = C\sqrt{\frac{\log n}{m}}$, and $s = 10$.

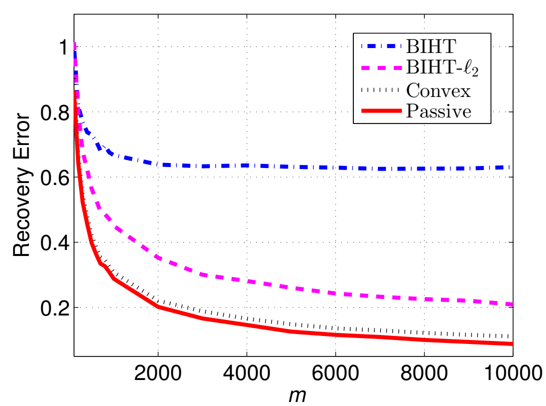
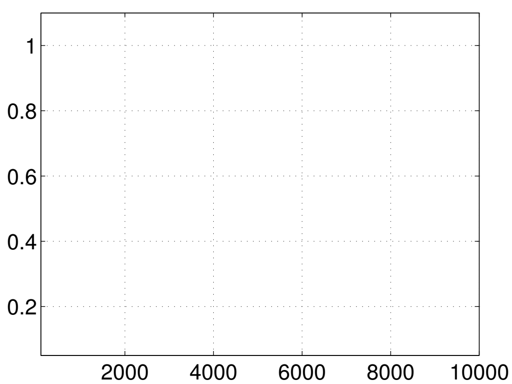


Figure 2. The recovery error of each algorithm versus the number of measurements m, when $s = 10$ and $n = 1000$.



References

Angluin, D. and Valiant, L.G. Fast probabilistic algorithms for hamiltonian circuits and matchings. *Journal of Computer and System Sciences*, 18(2):155–193, 1979.

Anthony, Martin and Bartlett, Peter L. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

Boufounos, Petros T. Reconstruction of sparse signals from distorted randomized measurements. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 3998–4001, 2010.

Boufounos, Petros T. and Baraniuk, Richard G. 1-bit compressive sensing. In *Proceedings of the 42nd Annual Conference on Information Sciences and Systems*, pp. 16–21, 2008.

Boyd, Stephen and Vandenberghe, Lieven. *Convex Optimization*. Cambridge University Press, 2004.

Candes, Emmanuel J. and Tao, Terence. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

Cho, Eungchun. Inner product of random vectors. *International Journal of Pure and Applied Mathematics*