

PAPER

HISTORY: An Efficient and Robust Algorithm for Noisy 1-Bit Compressed Sensing

Biao SUN^{†a)}, Member, Hui FENG[†], and Xinxin XU^{††}, Nonmembers

SUMMARY We consider the problem of sparse signal recovery from 1-bit measurements. Due to the noise present in the acquisition and transmission process, some quantized bits may be flipped to their opposite states. These sign flips may result in severe performance degradation. In this study, a novel algorithm, termed HISTORY, is proposed. It consists of Hamming support detection and coefficients recovery. The HISTORY algorithm has high recovery accuracy and is robust to strong measurement noise. Numerical results are provided to demonstrate the effectiveness and superiority of the proposed algorithm.

key words: 1-bit compressed sensing, sign flips, Hamming distance

1. Introduction

Compressed sensing, as introduced in [1]–[3], addresses the problem of estimating high dimensional signals from a set of relatively few linear measurements. It was demonstrated that a sparse signal can be reconstructed exactly if the measurement matrix satisfies the restricted isometric property (RIP) [4]. It was also shown that random matrices will satisfy the RIP with high probability if the entries are chosen according to independent and identically distributed (i.i.d.) Gaussian distribution.

In practical CS architectures, the measurements must be quantized to a finite number of bits. The extreme quantization setting where only the sign is acquired is known as 1-bit compressed sensing (1-bit CS) [5]. It has become increasingly popular due to its low computational cost and easy implementation for hardware [6]. In 1-bit CS, measurements of a signal $x \in \mathbb{R}^N$ are computed via

$$y = \text{sign}(\mathbf{A}x), \quad (1)$$

where $x \in \mathbb{R}^N$ is the signal, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the measurement matrix, $y \in \mathbb{R}^M$ is the set of 1-bit measurements, and function $\text{sign}(\cdot)$ maps the signal from \mathbb{R}^N to the Boolean cube $\mathcal{B}^M := \{-1, +1\}^M$. Since signs of real-valued measurements are used, one loses the ability to recover the magnitude of x and thus assumes that the signal has a unit norm, i.e., $\|x\|_2 = 1$. The 1-bit CS has been studied by many people and several algorithms have been developed to recover the

sparse signals [5], [7]–[12].

Despite the attractive attributes of 1-bit CS, the major disadvantage is that measurements are susceptible to noise during both acquisition and transmission [13]–[15]. In the noisy scenario, the output bit is randomly perturbed from the sign of the real-valued measurement, and the so-called *sign flips* seriously degrade recovery performance. The noise model is random sign-flip with probability ρ , i.e.,

$$y_i = b \cdot \text{sign}(\mathbf{A}^i x), \quad (2)$$

where b equals -1 with probability ρ , 1 with probability $1 - \rho$. y_i denotes the i^{th} element of y , and \mathbf{A}^i denotes the i^{th} row of \mathbf{A} . To date, researchers have developed numerous approaches for noisy 1-bit CS. Yan *et al.* [16] proposed a greedy method which detects the positions of sign flips iteratively, and recovers the signals using correct measurements. However, it requires the prior knowledge of noise level, which is often intractable in practical applications. Plan *et al.* [17] proposed a constrained optimization method with a linear objective. This convex formulation can work with a general notion of noise and achieve error for both exactly and approximately sparse signals. Ai *et al.* [18] extends [17] to sub-Gaussian measurements, and gets an irreducible component in the error and cannot be reduced by increasing the sample size or otherwise. However, they are computationally inefficient and difficult for hardware implementation. Recently, Zhang *et al.* [19] developed an efficient passive algorithm with closed-form solution, which improves the recovery performance for exactly K -sparse signals. Due to its high performance, robustness, and computational efficiency, they can be seen as the state-of-the-art algorithm for noisy 1-bit CS.

This study focuses on recovering EXACTLY K -sparse signals that have K nonzero coefficients in the noisy setting for 1-bit CS. We define Σ_K to be the set of all exactly K -sparse signals with unit norm as

$$\Sigma_K \stackrel{\text{def}}{=} \{v \in \mathbb{R}^N : \|v\|_0 = K, \|v\|_2 = 1\}. \quad (3)$$

A novel algorithm is proposed in this paper. Termed HISTORY, it consists of two key parts, namely *Hamming Support deTectioN*, and *cOefficients RecoverY*. The former aims to construct a candidate supports set by detecting possible supports of nonzero entries. The latter aims to calculate the coefficients belonging to the candidate supports set. Experimental results show that the proposed algorithm has high recovery performance than the state-of-the-art. Also, because

Manuscript received October 27, 2015.

Manuscript revised April 19, 2016.

Manuscript publicized July 6, 2016.

[†]The authors are with School of Electrical Engineering and Automation, Tianjin University, Tianjin, 300072, China.

^{††}The author is with Microsystems Technology Center, Information Science Academy of China Electronics Technology Group Corporation, Beijing, 100086, China.

a) E-mail: sunbiao@tju.edu.cn

DOI: 10.1587/transinf.2015EDP7438

containing no iterative step, it is computationally efficient and easy to implement.

2. HISTORY Algorithm

The main objective of this section is to characterize the HISTORY algorithm. Notations used throughout this paper are first described, then the two key parts of HISTORY are introduced in sequence.

2.1 Notations

Boldfaced capital letters such as \mathbf{A} are used for matrices. Italic capital letters such as S denote sets. For a matrix \mathbf{A} , the notations \mathbf{A}^i , \mathbf{A}_j , \mathbf{A}_j^i , \mathbf{A}^T , and \mathbf{A}_S denote its i^{th} row, j^{th} column, $i j^{\text{th}}$ element, transpose, and sub-matrix that contains the columns with indices in S , respectively. Small letters such as x are reserved for vectors and scalars. A vector x is called exactly K -sparse if K of its coefficients are nonzero. For a vector x , x_j , $\|x\|_p$, and x_S denote the j^{th} element of the vector, its p -norm, and sub-vector that contains the elements with indices in S , respectively. For two vectors $u \in \mathbb{R}^N$ and $v \in \mathbb{R}^N$, the notation $H(u, v)$ denotes the Hamming distance between them, which is defined as

$$H(u, v) \stackrel{\text{def}}{=} \#(u_j \neq v_j), \quad j \in 1, 2, \dots, N. \quad (4)$$

For an event E , the notation $\mathbb{P}(E)$ denotes its probability. For a random variable a , the notations $\mathbb{E}(a)$ and $\mathbb{D}(a)$ denote its expectation and variance, respectively.

2.2 Hamming Support Detection

To detect possible supports of nonzero coefficients from noisy 1-bit measurements, a *Hamming support detection* method is developed based on Angle Proportional Probability (APP), which is outlined as follows.

Theorem 1 (Angle Proportional Probability). *Let $x \in \Sigma_K$ be an exactly K -sparse signal with $\|x\|_2 = 1$. Let ϕ be a Gaussian random vector which is drawn uniformly from the unit ℓ_2 sphere in \mathbb{R}^N (i.e., each element of ϕ is firstly drawn i.i.d. from the standard Gaussian distribution $\mathcal{N}(0, 1)$). Define an event E to be*

$$E : \text{sign}(x^T \phi) \neq \text{sign}(\phi_j), \quad (5)$$

then it holds,

$$\mathbb{P}(E) = \frac{1}{\pi} \arccos(x_j). \quad (6)$$

The proof can be found in Appendix A. In particular, it shows that $\mathbb{P}(E)$ has a cosine function relationship with the j -th element of ϕ . Thus, x_j can be uniquely identified by $\mathbb{P}(E)$. In addition, the probability can be estimated from the instances of the random variable $\text{sign}(x^T \phi)$, which are exactly the 1-bit measurement vector y defined in (1). Therefore, y contains sufficient information to reconstruct x_j from

the estimation of $\mathbb{P}(E)$.

In the noisy setting, due to the fact that the signs of y are randomly perturbed, x_j cannot be computed directly from (5) and (6). However, given the noise level (sign flip ratio) as a prior knowledge, we have the following lemma.

Lemma 1. *Given an exactly K -sparse signal $x \in \Sigma_K$ with $\|x\|_2 = 1$, a standard Gaussian measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, and a 1-bit measurements vector $y = \text{sign}(\mathbf{A}x)$. In the noisy setting, suppose the sign flip ratio $\rho < 0.5$, define $P \in [0, 1]^N$ as a probability vector with P_j denoting its j -th element as*

$$P_j \stackrel{\text{def}}{=} \mathbb{P}(\text{sign}(y_i) \neq \text{sign}(\mathbf{A}_j^i)), \quad (7)$$

and it holds

$$P_j = \frac{1 - 2\rho}{\pi} \arccos(x_j) + \rho. \quad (8)$$

The proof can be found in Appendix B. From Lemma 1, we note that the Hamming distance between y and \mathbf{A}_j obeys the binomial distribution, i.e.,

$$H\{y, \mathbf{A}_j\} \sim B(M, P_j). \quad (9)$$

Moreover, by the definition of binomial distribution, we have

$$\mathbb{E}(H\{y, \mathbf{A}_j\}) = M \left(\frac{1 - 2\rho}{\pi} \arccos(x_j) + \rho \right). \quad (10)$$

Consequently, given the noise level and a relatively high measurement dimension, P_j can be well estimated by computing the Hamming distance, then x_j can be estimated accordingly. However, directly estimating x_j from (8) is intractable. For one thing, with the decrease of measurement dimension, the coefficients estimation performance degrades significantly. For another, (8) requires the sign flip ratio ρ as prior knowledge, which is often unknown in practical applications. To address the first problem, we only detect possible supports in current part and leave the coefficients estimation to the next one. To address the second problem, we propose the following lemma.

Lemma 2. *Given a finite measurement dimension M and a sign flip ratio $\rho < 0.5$, for any two different elements of x , denoted by $x_u, x_v, u \neq v$, if $x_u - x_v > \epsilon$, where ϵ is a small positive constant, we have*

$$\mathbb{P}(H\{y, \mathbf{A}_u\} < H\{y, \mathbf{A}_v\}) \geq 1 + C_1 - C_2 \epsilon^{-2}, \quad (11)$$

where C_1 and C_2 are constants and

$$C_1 = \frac{1}{4M}, \quad (12)$$

$$C_2 = \frac{\pi^2}{4M(1 - 2\rho)^2}.$$

The proof can be found in Appendix C. Note that with the increase of M , the probability $\mathbb{P}(H\{y, \mathbf{A}_u\} < H\{y, \mathbf{A}_v\})$ also increases, and when $M \rightarrow \infty$, we have

$$\mathbb{P}(H\{y, \mathbf{A}_u\} < H\{y, \mathbf{A}_v\}) \rightarrow 1, \quad \forall x_u - x_v > \epsilon. \quad (13)$$

From Lemma 2, it is easy to verify that despite the value of ρ , P_j in (8) is a monotone decreasing function with respect to x_j . The main point is that despite the noise level, the amplitude order of nonzero coefficients will maintain, while the dependencies in ρ will vanish in the corresponding Hamming distance. Therefore, we can set ρ to be an arbitrary value (e.g. $\rho = 0$) and compute approximate amplitudes of each coefficient via (8), then form the candidate supports set by selecting the supports with largest amplitudes.

2.3 Coefficients Recovery

Providing the candidate supports set, denoted by S , the next part is coefficients recovery, which aims to compute the amplitudes of nonzero coefficients. In this paper, we try to compute the coefficients vector c by solving the following constrained least squares problem,

$$c^* = \underset{c \in \mathbb{R}^{|S|}}{\text{minimize}} \|y - \mathbf{A}_S \cdot c\|_2 \quad \text{s.t.} \quad \|c\|_0 = K, \quad (14)$$

where $\|c\|_0$ denotes the 0-norm of c , i.e., counting the number of nonzero coefficients in c . Note that (14) is an overdetermined system when $|S| < M$. Thus, the sparsest solution to (14) is given by

$$c^* = \mathbf{A}_S \setminus y, \quad (15)$$

where “ \setminus ” denotes the *left matrix divide* operation. (15) can be solved via the QR decomposition [20] efficiently.

Based on the two parts described above, the HISTORY algorithm is fully summarized in Algorithm 1, where $\text{abs}(h)$ denotes the absolute value of each element of the vector h , $\text{FindSupp}(\text{abs}(h), \alpha K)$ returns the supports of the largest αK elements in $\text{abs}(h)$, and $\text{H}_K(\cdot)$ denotes the hard-thresholding operator who only preserves the largest K coefficients in magnitude and set others to 0. α is a parameter that controls the redundancy of support detection. For $\alpha > 1$, Algorithm 1 first selects more than K supports to form the candidate set. After computing the coefficients vector, the final K -sparse solution is obtained by hard-thresholding as in step 10. Note that when M is small, a high α is necessary to ensure the support detection accuracy. With the increase of M , a small α is sufficient to detect the supports accurately. In addition, a small α can decrease the computational complexity of (14), thus boost the whole algorithm. Based on the above analysis, we propose selecting α adaptively as

$$\alpha = 1 + \alpha_0 e^{-\tau \frac{M}{N}}, \quad (16)$$

where α_0 is the initial quantity and τ is the exponential decay constant.

It is worth noting that Algorithm 1 is a nearly-linear time algorithm, with its computational complexity to be $O(MN)$. Therefore, the proposed algorithm runs significantly faster than iterative algorithms.

Algorithm 1 HISTORY

Input: y, \mathbf{A}, K, α
1: **Initialize:** $x^* = \text{Zeros}(N)$
2: **for** each $j \in 1, \dots, N$ **do**
3: $P_j = H\{y, \mathbf{A}_j\}/M$
4: $h_j = \cos(\pi P_j)$
5: **end for**
6: $S = \text{FindSupp}(\text{abs}(h), \alpha K)$
7: $c^* = \mathbf{A}_S \setminus y$
8: $x_S^* = c^*$
9: **if** $\alpha > 1$ **then**
10: $x^* = \text{H}_K(x^*)$
11: **end if**
12: $x^* = x^* / \|x^*\|_2$
Output: recovered sparse signal x^*

3. Experiments

3.1 Experimental Setup

The target vector $x \in \mathbb{R}^N$ is generated by drawing its nonzero elements from the standard Gaussian distribution, and then normalized to have unit norm. The locations of the K nonzero coefficients of x are randomly selected. The elements in the measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ are also drawn from the standard Gaussian distribution. To generate sign flips, the measurement vector y is firstly acquired as in (1), then the sign of every element in y is flipped with probability ρ . For each setting of M , N , K , and ρ , the recovery experiment is repeated for 100 trials, and the average recovery error, denoted by $\|x - x^*\|_2 / \|x\|_2$, is reported. In all experiments, The parameter α is selected adaptively as in (16) with $\alpha_0 = 4$ and $\tau = 1$.

The HISTORY algorithm is compared with the following three algorithms,

- BIHT- ℓ_2 : a heuristic algorithm proposed in [14], which has been proved to have better performance than BIHT in the noisy setting. The maximum iterative number and step size are set to 200 and 1, respectively[†].
- Convex: a provable algorithm proposed in [17], which solves a convex optimization problem to recover the sparse signal^{††}.
- Passive: an efficient optimization algorithm with closed-form solution proposed in [19], experimental results illustrated that their passive algorithm outperforms other baselines. The regularization parameter γ is set to $\sqrt{\frac{\log N}{M}}$, which is the optimal choice in [19].

[†]A matlab implementation of BIHT- ℓ_2 algorithm can be downloaded from <http://perso.uclouvain.be/laurent.jacques/index.php/Main/BIHTDemo>.

^{††}The CVX package is used to solve this optimization problem. The package can be downloaded from <http://cvxr.com/cvx/>.

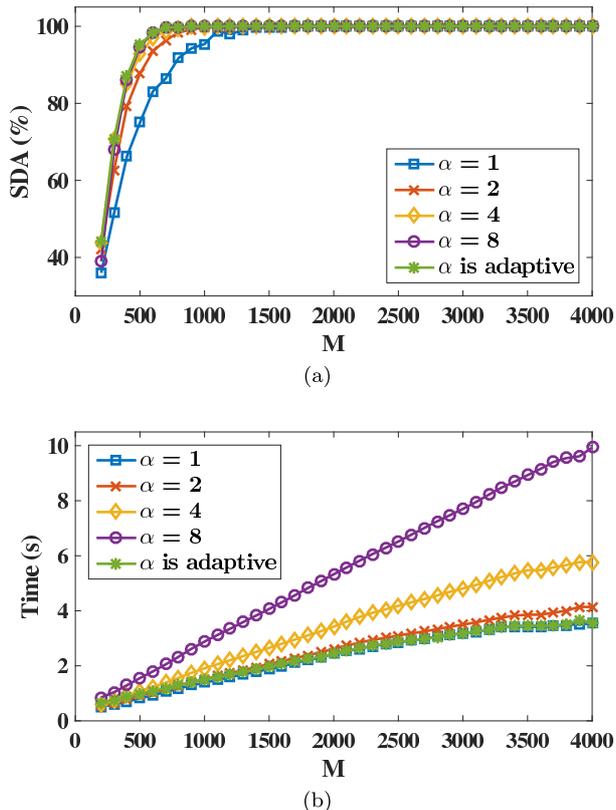


Fig. 1 (a) Evaluate support detection accuracy with different α , (b) evaluate CPU time with different α , when $N = 1000$, $K = 10$, $\rho = 0.1$, and M is varied from 200 to 4000.

3.2 Results

3.2.1 Support Detection Accuracy and Computational Efficiency of the Adaptive α

Firstly, the support detection accuracy with different α is studied. Parameters are set as $N = 1000$, $K = 10$, $\rho = 0.1$, and M is varied from 200 to 4000. The parameter α is selected adaptively as in (16) with parameters $\alpha_0 = 4$ and $\tau = 1$. α is also selected with fixed values as $\alpha = 1, 2, 4, 8$ for comparison. The support detection accuracy (SDA) is employed to quantify the percentage of detection success between the original supports and the reconstructed supports. The SDA is defined as

$$SDA = \frac{\#(\text{supp}(x^*) \cap \text{supp}(x))}{\#(\text{supp}(x))} \times 100\%. \quad (17)$$

The support detection accuracy curve with different α is shown in Fig. 1(a). It's observed that the adaptive α has the highest SDA. Although $\alpha = 8$ has better performance than $\alpha = 1, 2, 4$ with small M , all α values have same SDA after $M > 1500$. Therefore, a high α is not necessary when M is large.

To study the computational efficiency of the adaptive α , the CPU time of HISTORY with different α is evaluated.

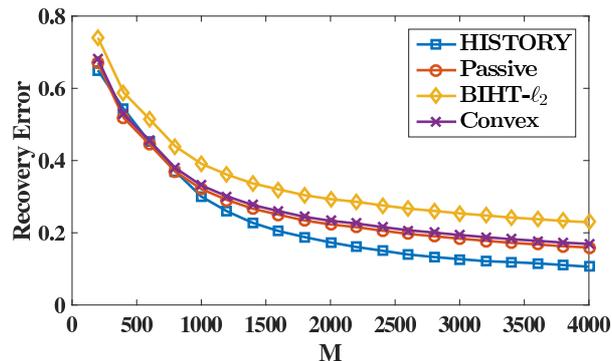


Fig. 2 Evaluate recovery error of each algorithm versus measurement dimension M , when $N = 1000$, $K = 10$, and $\rho = 0.1$.

For each point, the recovery experiment is repeated for 100 trials, and the total cpu time is reported in Fig. 1 (b). It is observed that larger α costs more computational resource. HISTORY with $\alpha = 8$ costs more than double CPU time of that with $\alpha = 1$ to recover the signals. The adaptive α costs least CPU time, i.e., almost same with $\alpha = 1$. Therefore, the adaptive α has better computational efficiency than fixed ones.

3.2.2 Recovery Error versus Measurement Dimension

Then the recovery error at different measurement dimension M is studied. Parameters are set as $N = 1000$, $K = 10$, $\rho = 0.1$, and M is varied from 200 to 4000. The recovery error curve is shown in Fig. 2. It is observed that with the increase of M , the recovery errors of all algorithms decrease. In particular, BIHT- ℓ_2 has the worst performance among these algorithms, that is because it is very sensitive to noise in the 1-bit measurements. In contrast, HISTORY has the best performance, especially when M is relatively large. The recovery errors of Convex and Passive are very similar.

3.2.3 Recovery Error versus Sparsity

Then the recovery error at different sparsity K is evaluated. Parameters are set as $N = 1000$, $\rho = 0.1$, $M = 4000$, and K is varied from 10 to 200. The recovery error curves are shown in Fig. 3. Results show that with the increase of K , the recovery errors of all algorithms increase. In particular, among these algorithms, HISTORY has the best performance while BIHT- ℓ_2 has the worst one. In addition, Passive and Convex almost have the same performance. Finally, we would like to emphasize that HISTORY increases its advantage with the increase of K , i.e., it is less sensitive to sparsity than other algorithms.

3.2.4 Recovery Error versus Sign Flip Ratio

Next, the recovery error at different sign flip ratio ρ is evaluated. Parameters are set as $N = 1000$, $K = 10$, $M = 4000$, and ρ is varied from 0 to 0.5. The recovery error curves

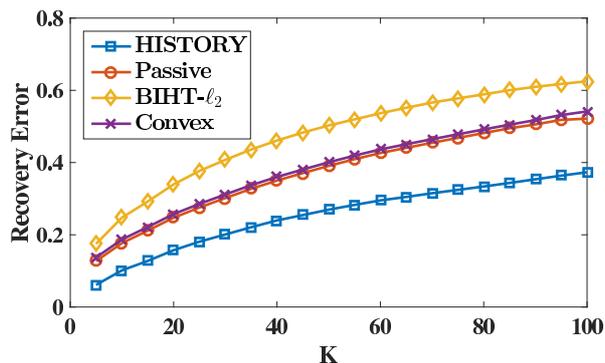


Fig. 3 Evaluate recovery error of each algorithm versus sparsity K , when $N = 1000$, $\rho = 0.1$, and $M = 4000$.

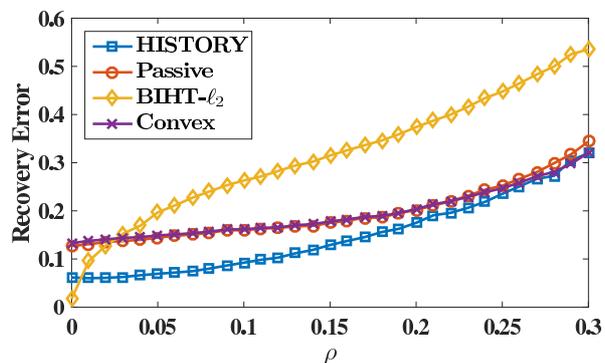


Fig. 4 Evaluate recovery error of each algorithm versus sign flip ratio ρ , when $N = 1000$, $K = 10$, and $M = 4000$.

are shown in Fig. 4. Though BIHT- ℓ_2 had the minimum recovery error when ρ is small, with the increase of ρ , its recovery error increased very quickly, making it be the worst algorithm at the high sign flip ratio. Passive and Convex had almost the same performance, which are better than that of BIHT- ℓ_2 . HISTORY has the best performance both at high and low sign flip ratio. Thus, HISTORY has the best noise robustness among these algorithms.

3.2.5 Recovery Error under Misspecified Model

Next, we study the error of each algorithm under the misspecified model, i.e., the sparsity of original signal is unknown. Parameters are set as $N = 1000$, $K = 10$, $M = 4000$, $\rho = 0.1$, and we select K_{select} from 1 to 20 to evaluate the algorithms. The recovery error curves are shown in Fig. 5. Results show that the recovery error of HISTORY sharply drops at the correct $K_{\text{select}} = K$. Moreover, HISTORY performs better than Passive and Convex in a neighborhood of K . Under misspecification with $K_{\text{select}} < K$, the recovery error is large since the error from unrecovered coefficients is large. For $K_{\text{select}} > K$, the nonzero coefficients are correctly recovered so that the corresponding error is small, but there is some additional error due to noise. To further improve the performance of HISTORY when K is unknown, many approaches can be used to estimate the sparsity level. For ex-

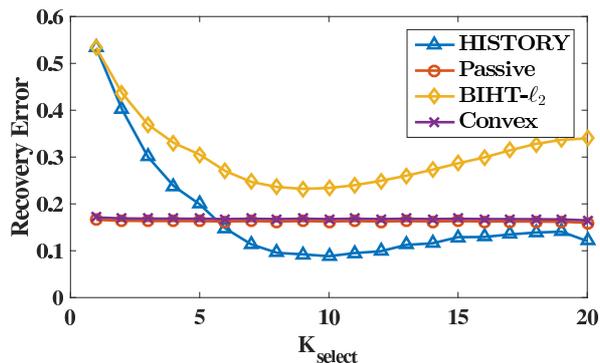


Fig. 5 Evaluate recovery error of each algorithm when K is unknown. Parameters are set to $N = 1000$, $K = 10$, $M = 4000$, and $\rho = 0.1$. K is selected from 1 to 20.

Table 1 Running time of each algorithm, when $N = 1000$, $K = 10$, $M = 4000$, and $\rho = 0.1$. For BIHT- ℓ_2 , there is no formal stopping criterion, and we report the running time after 100 iterations.

Algorithm	BIHT- ℓ_2	Convex	Passive	HISTORY
Time (s)	331.29	155.47	3.04	3.22

ample, we can select a regularization parameter γ first, and then use the thresholding method to estimate K as proposed in [19]. Other approaches such as 1-bit one scan [21] and sudocodes-based method [22] can also be used for sparsity estimation. We will leave this part for future research.

3.2.6 Computational Complexity

To evaluate the computational complexity of each algorithm, we study the running time of them. Parameters are set as $N = 1000$, $K = 10$, $M = 4000$, and $\rho = 0.1$. The running time of those algorithms can be found in Table 1. Results show that the running time of HISTORY and Passive are similar, while that of Convex and BIHT- ℓ_2 are significantly higher.

4. Conclusion

In this paper, we have developed an efficient and robust algorithm for noisy 1-bit compressive sensing. Compared with the existing methods, the proposed algorithm has several important advantages. It is robust to noise, it is computationally efficient, it has lower sample complexity, and it is easy to implement. Experimental results provide sound support to our theoretical development.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 61271321, 61473207 and 61401303, the Ph.D. Programs Foundation of the Ministry of Education of China under Grant 20120032110068, and Tianjin Key Technology Research and Development Program under Grant 14ZCZDS F00025.

References

[1] E.J. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?,” *IEEE Trans. Inform. Theory*, vol.52, no.12, pp.5406–5425, 2006.

[2] D.L. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol.52, no.4, pp.1289–1306, 2006.

[3] E.J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory*, vol.52, no.2, pp.489–509, 2006.

[4] E.J. Candès, “The restricted isometry property and its implications for compressed sensing,” *CR Math.*, vol.346, no.9, pp.589–592, 2008.

[5] P.T. Boufounos and R.G. Baraniuk, “1-bit compressive sensing,” *Proc. 42nd Annual Conf. Information Sciences and Systems (CISS)*, Princeton, USA, pp.16–21, 2008.

[6] P.T. Boufounos, “Reconstruction of sparse signals from distorted randomized measurements,” *Proc. IEEE International Conf. Acoustics Speech and Signal Processing (ICASSP)*, Dallas, USA, pp.3998–4001, 2010.

[7] P.T. Boufounos, “Greedy sparse signal reconstruction from sign measurements,” *Proc. 43rd Asilomar Conference on Signals, Systems and Computers*, Asilomar, USA, pp.1305–1309, 2009.

[8] A. Gupta, R. Nowak, and B. Recht, “Sample complexity for 1-bit compressed sensing and sparse classification,” *Proc. IEEE International Symposium on Information Theory (ISIT)*, Austin, USA, pp.1553–1557, 2010.

[9] J.N. Laska, Z. Wen, W. Yin, and R.G. Baraniuk, “Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements,” *IEEE Trans. Signal Process.*, vol.59, no.11, pp.5289–5301, 2011.

[10] T. Zhou and D. Tao, “1-bit hamming compressed sensing,” *Proc. IEEE International Symposium on Information Theory (ISIT)*, Cambridge, USA, pp.1862–1866, 2012.

[11] B. Sun and J. Jiang, “Investigation of sign spectrum sensing method,” *Acta Phys Sin-ch Ed*, vol.11, no.32, 110701, 2011.

[12] B. Sun, Q. Chen, X. Xu, L. Zhang, and J. Jiang, “A fast and accurate two-stage algorithm for 1-bit compressive sensing,” *IEICE Trans. Inf. & Syst.*, vol.E96.D, no.1, pp.120–123, 2013.

[13] S. Gopi, P. Netrapalli, P. Jain, and A. Nori, “One-bit compressed sensing: Provable support and vector recovery,” *Proc. 30th International Conf. Machine Learning (ICML)*, Atlanta, USA, pp.154–162, 2013.

[14] 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 Trans. Inform. Theory*, vol.59, no.4, pp.2082–2102, 2013.

[15] Y. Plan and R. Vershynin, “One-bit compressed sensing by linear programming,” *Commun. Pur. Appl. Math.*, vol.66, no.8, pp.1275–1297, 2013.

[16] M. Yan, Y. Yang, and S. Osher, “Robust 1-bit compressive sensing using adaptive outlier pursuit,” *IEEE Trans. Signal Process.*, vol.60, no.7, pp.3868–3875, 2012.

[17] Y. Plan and R. Vershynin, “Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach,” *IEEE Trans. Inform. Theory*, vol.59, no.1, pp.482–494, 2013.

[18] A. Ai, A. Lapanowski, Y. Plan, and R. Vershynin, “One-bit compressed sensing with non-gaussian measurements,” *Linear. Algebra. Appl.*, vol.441, pp.222–239, 2014.

[19] L. Zhang, J. Yi, and R. Jin, “Efficient algorithms for robust one-bit compressive sensing,” *Proc. 31st International Conf. Machine Learning (ICML)*, Beijing, China, pp.820–828, 2014.

[20] L.N. Trefethen and D. Bau, *Numerical linear algebra*, SIAM, Philadelphia, 1997.

[21] P. Li, “One scan 1-bit compressed sensing,” *19th International*

Conference on Artificial Intelligence and Statistics, Cadiz, Spain, pp.1517–1523, 2016.

[22] Y. Ma, D. Baron, D. Needell, “Two-Part Reconstruction with Noisy-Sudocodes,” *IEEE Trans. Signal Process.*, vol.62, no.23, pp.6323–6334, 2014.

Appendix A: Proof of Theorem 1

It is worth noting that

$$\begin{aligned} \mathbb{P}(\text{sign}(x^T \phi) = \text{sign}(\phi_j)) \\ = \mathbb{P}(x^T \phi > 0, \phi_j > 0) + \mathbb{P}(x^T \phi \leq 0, \phi_j \leq 0). \end{aligned} \tag{A.1}$$

We can divide $x^T \phi$ into two parts as

$$m = x^T \phi = m_j + m_c, \tag{A.2}$$

where

$$\begin{aligned} m_j &= \phi_j x_j, \\ m_c &= x^T \phi - \phi_j x_j. \end{aligned} \tag{A.3}$$

In addition, it can be easily verified both m_j and m_c satisfy Gaussian distribution, i.e.,

$$\begin{aligned} m_j &\sim \mathcal{N}(0, x_j^2), \\ m_c &\sim \mathcal{N}(0, 1 - x_j^2). \end{aligned} \tag{A.4}$$

Depending on x_j , we have three situations as follows, (1) when $x_j = 0$, we have

$$\begin{aligned} \mathbb{P}(x^T \phi > 0, \phi_j > 0) \\ = \mathbb{P}(m_c > 0, m_j > 0) \\ = \mathbb{P}(m_c > 0)P(m_j > 0) \\ = \frac{1}{4}. \end{aligned} \tag{A.5}$$

In the same way, we have

$$\mathbb{P}(x^T \phi \leq 0, \phi_j \leq 0) = \frac{1}{4}. \tag{A.6}$$

Therefore,

$$\mathbb{P}(\text{sign}(x^T \phi) = \text{sign}(\phi_j)) = \frac{1}{2}. \tag{A.7}$$

(2) when $x_j > 0$, we have

$$\mathbb{P}(x^T \phi > 0, \phi_j > 0) = \mathbb{P}(m_c + m_j > 0, m_j > 0) \tag{A.8}$$

The joint probability density function of m_c and m_j is

$$p(m_c, m_j) = \frac{1}{2\pi x_j \sqrt{1 - x_j^2}} \exp\left(-\frac{1}{2}\left(\frac{m_j^2}{x_j^2} + \frac{m_c^2}{1 - x_j^2}\right)\right) \tag{A.9}$$

Assume that

$$\begin{aligned} m_c &= r \cos \theta, \\ m_j &= r \sin \theta, \end{aligned} \quad (\text{A} \cdot 10)$$

then we have,

$$\begin{aligned} &\mathbb{P}(m_c + m_j > 0, m_j > 0) \\ &= \frac{1}{2\pi x_j \sqrt{1-x_j^2}} \int_0^{\frac{3}{4}\pi} d\theta \\ &\int_0^\infty \exp\left(-\frac{1}{2}\left(\frac{r^2 \cos^2 \theta}{1-x_j^2} + \frac{r^2 \sin^2 \theta}{x_j^2}\right)\right) r dr \\ &= \frac{1}{2} - \frac{1}{2\pi} \arccos(x_j). \end{aligned} \quad (\text{A} \cdot 11)$$

In the same way, we have

$$\begin{aligned} &\mathbb{P}(x^T \phi \leq 0, \phi_j \leq 0) \\ &= P(m_c + m_j \leq 0, m_j \leq 0) \\ &= \frac{1}{2} - \frac{1}{2\pi} \arccos(x_j). \end{aligned} \quad (\text{A} \cdot 12)$$

Therefore, we have

$$\mathbb{P}(\text{sign}(x^T \phi) = \text{sign}(\phi_j)) = 1 - \frac{1}{\pi} \arccos(x_j). \quad (\text{A} \cdot 13)$$

(3) when $x_j < 0$,

$$\begin{aligned} &\mathbb{P}(\text{sign}(x^T \phi) = \text{sign}(\phi_j)) \\ &= \mathbb{P}(x^T \phi > 0, \phi_j > 0) + \mathbb{P}(x^T \phi \leq 0, \phi_j \leq 0) \\ &= \mathbb{P}(m_c + m_j > 0, m_j < 0) + \mathbb{P}(m_c + m_j \leq 0, m_j \geq 0). \end{aligned} \quad (\text{A} \cdot 14)$$

The first part can be computed via

$$\begin{aligned} &\mathbb{P}(m_c + m_j > 0, m_j < 0) \\ &= \frac{1}{2\pi x_j \sqrt{1-x_j^2}} \int_{-\frac{1}{4}\pi}^0 d\theta \\ &\int_0^\infty \exp\left(-\frac{1}{2}\left(\frac{r^2}{1-x_j^2} \cos^2 \theta + \frac{r^2}{x_j^2} \sin^2 \theta\right)\right) r dr \\ &= \frac{1}{2} - \frac{1}{2\pi} \arccos(x_j). \end{aligned} \quad (\text{A} \cdot 15)$$

In the same way, we calculate the second part as

$$\mathbb{P}(m_c + m_j < 0, m_j > 0) = \frac{1}{2} - \frac{1}{2\pi} \arccos(x_j). \quad (\text{A} \cdot 16)$$

Therefore, we have

$$\mathbb{P}(\text{sign}(x^T \phi) = \text{sign}(\phi_j)) = 1 - \frac{1}{\pi} \arccos(x_j). \quad (\text{A} \cdot 17)$$

Synthesize the above three situations, we have

$$\mathbb{P}(\text{sign}(x^T \phi) = \text{sign}(\phi_j)) = 1 - \frac{1}{\pi} \arccos(x_j) \quad (\text{A} \cdot 18)$$

$$\mathbb{P}(\text{sign}(x^T \phi) \neq \text{sign}(\phi_j)) = \frac{1}{\pi} \arccos(x_j). \quad (\text{A} \cdot 19)$$

This concludes the proof.

Appendix B: Proof of Lemma 1

In the noiseless setting, we define an event $E1$ to be

$$E1 : \text{sign}(y_i) \neq \text{sign}(\mathbf{A}^i). \quad (\text{A} \cdot 20)$$

From Theorem 1, we have

$$\mathbb{P}(E1) = \frac{1}{\pi} \arccos(x_j). \quad (\text{A} \cdot 21)$$

In the noisy setting, we define an event $E2$ that y_i has its sign flipped, i.e.,

$$E2 : y_i = -1 \cdot \text{sign}(\mathbf{A}^i x), \quad (\text{A} \cdot 22)$$

and by the definition of sign flip ratio, we have

$$\mathbb{P}(E2) = \rho. \quad (\text{A} \cdot 23)$$

It is easy to verify that $E1$ and $E2$ are independent events, and we have

$$\begin{aligned} P_j &= \mathbb{P}(\overline{E1}E2) + \mathbb{P}(E1\overline{E2}) \\ &= \mathbb{P}(\overline{E1})\mathbb{P}(E2) + \mathbb{P}(E1)\mathbb{P}(\overline{E2}) \\ &= \left(1 - \frac{1}{\pi} \arccos(x_j)\right)\rho + \frac{1}{\pi} \arccos(x_j)(1 - \rho) \\ &= \frac{1 - 2\rho}{\pi} \arccos(x_j) + \rho. \end{aligned} \quad (\text{A} \cdot 24)$$

Then the proof completes.

Appendix C: Proof of Lemma 2

For any $x_u - x_v > \epsilon$, because P_j is continuous on the closed interval $[x_v, x_u]$ and differentiable on the open interval (x_v, x_u) , by using the Lagrange's mean value theorem, there exists a point x_c in (x_v, x_u) such that

$$\begin{aligned} P_v - P_u &= \frac{\partial P_j}{\partial x_j} \Big|_{x_j=x_c} (x_v - x_u) \\ &= -\frac{(1-2\rho)}{\pi \sqrt{1-x_c^2}} (x_v - x_u) \\ &> \frac{(1-2\rho)\epsilon}{\pi}. \end{aligned} \quad (\text{A} \cdot 25)$$

Because both $H\{y, \mathbf{A}_u\}$ and $H\{y, \mathbf{A}_v\}$ obey the binomial distribution, i.e.,

$$\begin{aligned} H\{y, \mathbf{A}_u\} &\sim B(M, P_u), \\ H\{y, \mathbf{A}_v\} &\sim B(M, P_v), \end{aligned} \quad (\text{A} \cdot 26)$$

therefore, the expectation and variance of $H\{y, \mathbf{A}_u\}$ and $H\{y, \mathbf{A}_v\}$ are given by

$$\begin{aligned} \mathbb{E}(H\{y, \mathbf{A}_u\}) &= MP_u, \\ \mathbb{E}(H\{y, \mathbf{A}_v\}) &= MP_v, \\ \mathbb{D}(H\{y, \mathbf{A}_u\}) &= MP_u(1 - P_u), \\ \mathbb{D}(H\{y, \mathbf{A}_v\}) &= MP_v(1 - P_v). \end{aligned} \quad (\text{A} \cdot 27)$$

Define a random variable z as

$$z \stackrel{\text{def}}{=} H\{y, \mathbf{A}_v\} - H\{y, \mathbf{A}_u\}. \quad (\text{A} \cdot 28)$$

Because $H\{y, \mathbf{A}_u\}$ and $H\{y, \mathbf{A}_v\}$ are independent, the expectation and the variance of z are given by

$$\begin{aligned} \mathbb{E}(z) &= \mathbb{E}(H\{y, \mathbf{A}_v\}) - \mathbb{E}(H\{y, \mathbf{A}_u\}) \\ &= M(P_v - P_u), \\ \mathbb{D}(z) &= \mathbb{D}(H\{y, \mathbf{A}_v\}) + \mathbb{D}(H\{y, \mathbf{A}_u\}) \\ &= M(P_v(1 - P_v) + P_u(1 - P_u)). \end{aligned} \quad (\text{A} \cdot 29)$$

In addition, note that the probability mass function of z is symmetrical with $\mathbb{E}(z)$. By using the Chebyshev's inequality, we have

$$\begin{aligned} \mathbb{P}(z > 0) &= 1 - \mathbb{P}(z \leq 0) \\ &= 1 - \frac{1}{2} \mathbb{P}(|z - \mathbb{E}(z)| \geq \mathbb{E}(z)) \\ &\geq 1 - \frac{\mathbb{D}(z)}{2\mathbb{E}^2(z)}. \end{aligned} \quad (\text{A} \cdot 30)$$

Let $f(z) = 1 - \frac{\mathbb{D}(z)}{2\mathbb{E}^2(z)}$, by substituting (A·25) and (A·29) into $f(z)$, we have

$$\begin{aligned} f(z) &= \\ 1 + \frac{1}{2M} + \frac{\pi^2 P_u(P_u - 1)}{M\epsilon^2(1 - 2\rho)^2} + \frac{\pi(P_u(2 - 4\rho) + 2\rho - 1)}{2M\epsilon(1 - 2\rho)^2} \end{aligned} \quad (\text{A} \cdot 31)$$

By computing the derivative of $f(z)$ with respect to P_u and set it to be 0, we compute the minimum value of $f(z)$ to be

$$f(z) \geq 1 - \frac{\pi^2 - \epsilon^2(1 - 2\rho)^2}{4M\epsilon^2(1 - 2\rho)^2}. \quad (\text{A} \cdot 32)$$

By combining (A·30) and (A·32), we have

$$\mathbb{P}(H\{y, \mathbf{A}_v\} > H\{y, \mathbf{A}_u\}) \geq 1 + \frac{1}{4M} - \frac{\pi^2}{4M(1 - 2\rho)^2} \epsilon^{-2}. \quad (\text{A} \cdot 33)$$

Then the proof completes.



Hui Feng is a master's degree in the Department of Electrical Engineering & Automation, Tianjin University, China. She is interested in compressed sensing.



Xinxin Xu received the Ph.D. degree in electrical science & technology from Huazhong University of Science and Technology in 2013. She is currently an engineer with the Microsystems Technology Center, Information Science Academy of China Electronics Technology Group Corporation. Her research interests include meta material, frequency selective surfaces, electromagnetic protection and electromagnetic measurement.



Biao Sun received the Ph.D. degree in electrical science & technology from Huazhong University of Science and Technology in 2013. He is currently an assistant professor with the Department of Electrical Engineering & Automation, Tianjin University, China. His research interests include neural information processing, compressed sensing, and machine learning.