PAPER

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

Biao SUN[†], **Hui FENG**[†], ***Nonmembers, and*** **Xinxin XU**[††], ***Member***

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

$$y = \text{sign}(\mathbf{A}x + n), \tag{1}$$

where $x \in \mathbb{R}^N$ is the signal, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the measurement matrix, $n \in \mathbb{R}^M$ is the measurement noise, $y \in \mathbb{R}^M$ 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 susceptive to noise during both acquisition and transmission [13]–[15]. In 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. To date, researchers have developed numerous approaches for noisy 1-bit CS. Yan *et al.* [16] proposed a greedy method which detect the positions of sign flips iteratively, and recover 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 computational 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 signal. 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 signal in the noisy setting for 1-bit CS. A novel algorithm is proposed. Termed HISTORY, it consists of two key parts, namely *HammIng Support deTection*, and *cOefficients RecoverY*. The former aims to construct a candidate support set by detecting possible supports of nonzero entries. The latter aims to calculate the coefficients belonging to the candidate support set. Experimental results show that the proposed algorithm has high recovery performance than the state-of-the-art. In addition, due to the fact that containing no iterative step, it is computationally efficient and easy to implement.

Manuscript received October 23, 2015.
Manuscript revised October 23, 2015.

[†]The author is with School of Electrical Engineering and Automation, Tianjin University, Tianjin, 300072, China. Email: sunbiao@tju.edu.cn

[††]The author is with Microsystems Technology Center, Information Science Academy of China Electronics Technology Group Corporation, Beijing, 100086, China

DOI: 10.1587/transinf.E98.D.1

2. HISTORY Algorithm

The main objective of this section is to characterize the HISTORY algorithm. Notations used in 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 $\mathbf{A}$, $\mathbf{A}_j$, $\mathbf{A}_{i,j}$, $\mathbf{A}^{\mathrm{T}}$, and $\mathbf{A}_S$ denote its j^{th} column, ij^{th} element, transpose, and sub-matrix which contains the columns with indices in S, respectively. Small letters such as x are reserved for vectors and scalars. A vector x is called K-sparse if at most 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 which 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

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

The notation $\mathbb{P}$ denotes probability of an event.

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 a 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*

$$E : \operatorname{sign}(x^{\mathrm{T}}\phi) \neq \operatorname{sign}(\phi_j), \tag{3}$$

then it holds,

$$\mathbb{P}(E) = \frac{1}{\pi}\arccos(x_j). \tag{4}$$

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 $\operatorname{sign}(x^{\mathrm{T}}\phi)$, 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 noisy setting, due to the fact that the signs of y are randomly perturbed, x_j cannot computed directly from (3) and (4). However, given the noise level (sign flip ratio) as a prior knowledge, we have the following lemma,

Lemma 1. *Given a K-sparse signal x with $\|x\|_2 = 1$, a standard Gaussian measurement matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$, and a 1-bit measurements vector $y = \operatorname{sign}(\mathbf{A}x)$. In 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*

$$P_j \stackrel{\text{def}}{=} \mathbb{P}\big(\operatorname{sign}(y_i) \neq \operatorname{sign}(\mathbf{A}_{ij})\big), \tag{5}$$

and it holds

$$P_j = \frac{1-2\rho}{\pi}\arccos(x_j) + \rho. \tag{6}$$

The proof can be found in Appendix B. In the limit of large systems as the measurements dimension $M \to \infty$, P_j is an asymptotically approximation with the *Hamming Distance* between y and $\mathbf{A}_j$, i.e., $H\{y, \mathbf{A}_j\}/M = P_j$. Consequently, given the noise level and a relatively high measurements 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 (6) is intractable. For one thing, with the decrease of measurements dimension, the coefficients estimation performance degrades significantly. For another, (6) 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, it is easy to verify that despite the value of ρ, P_j in (6) is a monotone decreasing function with respect to x_j, i.e., when $M \to \infty$, we have

$$H\{y, \mathbf{A}_u\} > H\{y, \mathbf{A}_v\}, \quad \forall\, x_u < x_v. \tag{7}$$

The main point is that despite the noise level, amplitude order of nonzero coefficients will be maintained while dependencies in ρ vanish in the corresponding Hamming distance. Therefore, we can set ρ to be an arbitrarily value (e.g. $\rho = 0$) and compute approximate amplitudes of each coefficients via (6), then form the candidate support set by selecting the supports with largest amplitudes.

2.3 Coefficients recovery

Providing the candidate support 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,

$$c^* = \underset{c\in\mathbb{R}^{\|S\|_0}}{\operatorname{minimize}} \|y - \mathbf{A}_S \cdot c\|_2 \quad \text{s.t.} \quad \|c\|_0 \leq K, \tag{8}$$

where $\|c\|_0$ denotes the 0-norm of c, i.e., counting the number of nonzero coefficients in c. It's worth noting that is a overdetermined system when $|S|_0 < M$. Thus, the sparsest solution to (8) is given by

$$c^* = \mathbf{A}_{\mathcal{S}} \setminus y, \tag{9}$$

where $\setminus$ denotes the *left matrix divide* operation. (9) 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 FindSupp($|h|, \alpha K$) returns the supports of the largest αK elements in $|h|$, and $\mathrm{H_K}(\cdot)$ denotes the hard thresholding operator which only preserve the largest K coefficients and set others to 0. It's 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: $y, \mathbf{A}, K, \alpha$
Initialize: $x^* = \mathrm{Zeros}(N)$
for each $j \in 1, \ldots, N$ **do**
  $P_j = H\{y, \mathbf{A}_j\}/M$
  $h_j = \cos(\pi P_j)$
end for
$S = \mathrm{FindSupp}(|h|, \alpha K)$
$c^* = \mathbf{A}_{\mathcal{S}} \setminus y$
$x^*_{\mathcal{S}} = c^*$
if $\alpha > 1$ **then**
  $x^* = \mathrm{H_K}(|x^*|)$
end if
$x^* = x^*/\|x^*\|_2$
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 in Algorithm 1 is set to 2.

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 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].

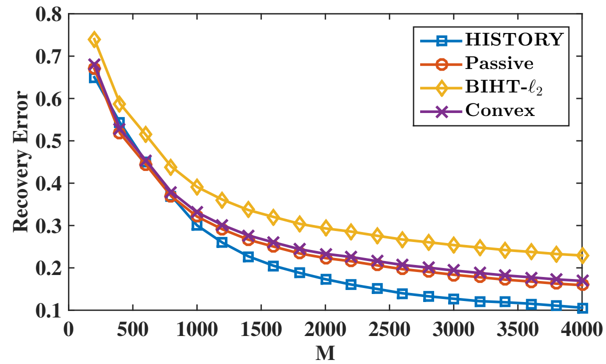

Fig. 1: Evaluate recovery error of each algorithm versus measurement dimension M, when $N = 1000$, $K = 10$, and $\rho = 0.1$.

3.2 Results

3.2.1 Recovery error versus measurement dimension

First 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. 1. It's observed that with the increase of M, the recovery error of all algorithms decrease. In particular, BIHT-ℓ_2 has worst performance among these algorithms, that is because it is very sensitive to noise in the 1-bit measurements. In contrast, HISTORY has best performance, especially when M is relatively large. The recovery error of Convex is similar to Passive as in noiseless scenario.

3.2.2 Recovery error versus sparsity

Then the recovery error at different sparsity K is eval-

†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/.

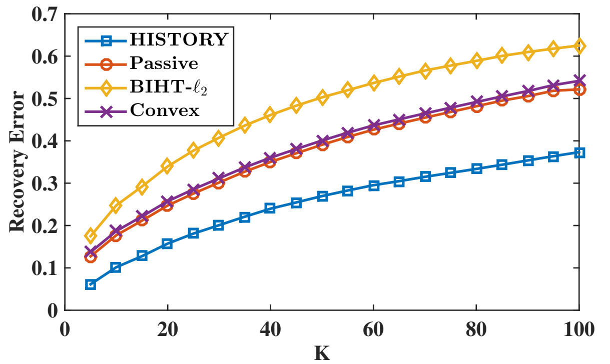

Fig. 2: Evaluate recovery error of each algorithm versus sparsity K, when $N = 1000$, $\rho = 0.1$, and $M = 4000$.

uated. 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. 2. Results show that with the increase of K, the recovery error of all algorithms increase. In particular, among these algorithms, HISTORY has best performance while BIHT-ℓ_2 has worst one. In addition, Passive and Convex almost have the same performance. Finally, we would like to emphasize that HISTORY increase its advantage with the increase of K, i.e., it is less sensitive to sparsity than other algorithms.

3.2.3 Recovery error versus noise level

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 are shown in Fig. 3. 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 to be the worst algorithm at high noise level. Passive and Convex had almost the same performance, which are better than that of BIHT-ℓ_2. HISTORY has best performance both at high and low noise levels. Thus, HISTORY has best noise robustness among these algorithms.

3.2.4 Recovery error under misspecified model

Next we study the error of each algorithm under 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. 4. 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 mis-specification 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

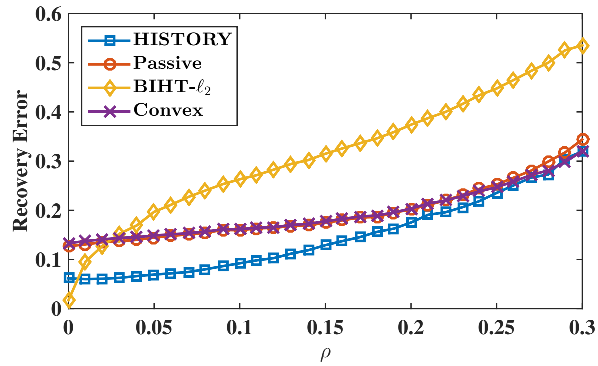

Fig. 3: Evaluate recovery error of each algorithm versus sign flip ratio ρ, when $N = 1000$, $K = 10$, and $M = 4000$.

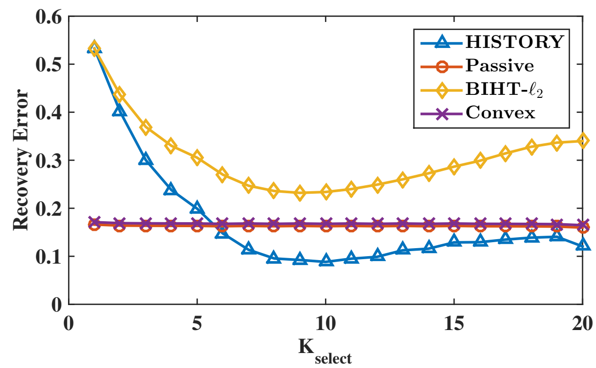

Fig. 4: Evaluate recovery error of each algorithm when K is unknown. Parameters are set to $N = 1000$, $K = 10$, $M = 4000$, and $\rho = 0.1$. K is selected from 1 to 20.

correctly recovered so that the corresponding error is small, but there is some additional error due to noise.

3.2.5 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.

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 stoping criterion, and we report the running time after 100 iterations.

Algorithm	BIHT-ℓ_2	Convex	Passive	HISTORY
Time (s)	325.82	163.54	2.95	3.51

4. Conclusion

In this paper, we develop a efficient and robust algorithm for noisy 1-bit compressive sensing. Compared with the existing methods, the proposed algorithm have several important advantages: it is robust to noise, it is computationally efficient, it has lower sample complexity, and it is easy to be 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 T. Inform Theory, vol.52, no.12, pp. 5406-5425, 2006.
[2] D. L. Donoho, "Compressed sensing," IEEE T. 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 T. 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. on 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. on 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 T. Signal Proces, 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, pp. 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. 96, 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. on 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 T. 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 T. Signal Proces, 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 T. 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. on Machine Learning (ICML), Beijing, China, pp. 820-828, 2014.
[20] L. N. Trefethen and D. Bau, Numerical linear algebra, SIAM, Philadelphia, 1997.

Appendix A: Proof of Theorem 1

It's worth noting that

$$\begin{aligned}&\mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) = \mathrm{sign}(\phi_j)) \\ &= \mathbb{P}(x^{\mathrm{T}}\phi > 0, \phi_j > 0) + \mathbb{P}(x^{\mathrm{T}}\phi \le 0, \phi_j \le 0).\end{aligned}$$

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

$$m = x^{\mathrm{T}}\phi = m_j + m_c,$$

where

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

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

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

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

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

In the same way, we have

$$\mathbb{P}(x^{\mathrm{T}}\phi \leqslant 0, \phi_j \leqslant 0) = \frac{1}{4}.$$

Therefore,

$$\mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) = \mathrm{sign}(\phi_j)) = \frac{1}{2}.$$

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

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

The joint probability density function of m_c and m_j is

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

Assume that

$$m_c = r\cos\theta,$$
$$m_j = r\sin\theta,$$

then we have,

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

In the same way, we have

$$\begin{aligned}&\mathbb{P}(x^{\mathrm{T}}\phi \le 0, \phi_j \le 0)\\&= P(m_c + m_j \le 0, m_j \le 0)\\&= \frac{1}{2} - \frac{1}{2\pi}\arccos(x_j).\end{aligned}$$

Therefore, we have

$$\mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) = \mathrm{sign}(\phi_j)) = 1 - \frac{1}{\pi}\arccos(x_j).$$

(3) when $x_j < 0$,

$$\begin{aligned}&\mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) = \mathrm{sign}(\phi_j))\\&= \mathbb{P}(x^{\mathrm{T}}\phi > 0, \phi_j > 0) + \mathbb{P}(x^{\mathrm{T}}\phi \le 0, \phi_j \le 0)\\&= \mathbb{P}(m_c + m_j > 0, m_j < 0) + \mathbb{P}(m_c + m_j \le 0, m_j \ge 0).\end{aligned}$$

The first part can be computed via

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

In the same way, we calculate the second part as

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

Therefore, we have

$$\mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) = \mathrm{sign}(\phi_j)) = 1 - \frac{1}{\pi}\arccos(x_j).$$

Synthesize the above three situations, we have

$$\mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) = \mathrm{sign}(\phi_j)) = 1 - \frac{1}{\pi}\arccos(x_j)$$
$$\mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) \neq \mathrm{sign}(\phi_j)) = \frac{1}{\pi}\arccos(x_j).$$

This concludes the proof.

Appendix B: Proof of Lemma 1

Providing $M \to \infty$ with sign flip ratio ρ, the Hamming distance between y and ϕ_j can be computed via

$$\begin{aligned}&H\{y, \phi_j\}\\&= (1-\rho)M \cdot \mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) \neq \mathrm{sign}(\phi_j))\\&+ \rho M \cdot \mathbb{P}(\mathrm{sign}(x^{\mathrm{T}}\phi) = \mathrm{sign}(\phi_j))\\&= \frac{(1-\rho)M}{\pi}\arccos(x_j) + \rho M\left(1 - \frac{1}{\pi}\arccos(x_j)\right)\\&= \left(\frac{1-2\rho}{\pi}\arccos(x_j) + \rho\right)M,\end{aligned}$$

which yields

$$\frac{H\{y, \phi_j\}}{M} = \frac{1-2\rho}{\pi}\arccos(x_j) + \rho.$$

Then the proof completes.

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.

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.