Robust 1-bit Compressive Sensing using Adaptive Outlier Pursuit

Ming Yan, Yi Yang and Stanley Osher

***Abstract*—In compressive sensing (CS), the goal is to recover signals at reduced sample rate compared to the classic Shannon-Nyquist rate. However, the classic CS theory assumes the measurements to be real-valued and have infinite bit precision. The quantization of CS measurements has been studied recently and it has been shown that accurate and stable signal acquisition is possible even when each measurement is quantized to only one single bit. There are many algorithms proposed for 1-bit compressive sensing and they work well when there is no noise in the measurements, e.g., there are no sign flips, while the performance is worsened when there are a lot of sign flips in the measurements. In this paper, we propose a robust method for recovering signals from 1-bit measurements using adaptive outlier pursuit. This method will detect the positions where sign flips happen and recover the signals using "correct" measurements. Numerical experiments show the accuracy of sign flips detection and high performance of signal recovery for our algorithms compared with other algorithms.**

***Index Terms*—1-bit compressive sensing, adaptive outlier pursuit**

I. INTRODUCTION

THE theory of compressive sensing (CS) enables reconstruction of sparse or compressible signals from a small number of linear measurements relative to the dimension of the signal space [1], [2], [3], [4], [5]. In this setting, we have

$$y = \Phi x, \tag{1}$$

where $x \in \mathbf{R}^N$ is the signal, $\Phi \in \mathbf{R}^{M\times N}$ with $M < N$ is an underdetermined measurement system, and $y \in \mathbf{R}^M$ is the set of linear measurements. It was demonstrated that K-sparse signals, i.e., $x \in \Sigma_K$ where $\Sigma_K := \{x \in \mathbf{R}^N : \|x\|_0 := |\text{supp}(x)| \le K\}$, can be reconstructed exactly if Φ satisfies the restricted isometry property (RIP) [6]. 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.

Classic compressive sensing assumes that the measurements are real valued and have infinite bit precision. However, in practice, CS measurements must be quantized, i.e., each measurement has to be mapped from a real value (over a potentially infinite range) to a discrete value over some finite range, which will induce error on the measurements. The quantization of CS measurements has been studied recently and several new algorithms are proposed [7], [8], [9], [10], [11], [12].

Furthermore, for some real world problems, severe quantization may be inherent or preferred. For example, in analog-to-digital conversion (ADC), the acquisition of 1-bit measurements of an analog signal only requires a comparator to zero, which is an inexpensive and fast piece of hardware that is robust to amplification of the signal and other errors, as long as they preserve the signs of the measurements, see [13], [14]. In this paper, we will focus on the CS problem when 1-bit quantizer is used.

The 1-bit compressive sensing framework proposed in [13] is as follows. Measurements of a signal $x \in \mathbf{R}^N$ are computed via

$$y = A(x) := \text{sign}(\Phi x). \tag{2}$$

Therefore, the measurement operator $A(\cdot)$ is a mapping from $\mathbf{R}^N$ to the Boolean cube [1] $\mathcal{B}^M := \{-1, 1\}^M$. We have to recover signal $x \in \sum_K^* := \{x \in S^{N-1} : \|x\|_0 \le K\}$ where $S^{N-1} := \{x \in \mathbf{R}^N : \|x\|_2 = 1\}$ is the unit hyper-sphere of dimension N. Since the scale of the signal is lost during the quantization process, we can restrict the sparse signals to be on the unit hyper-sphere. Jacques et al. provided two flavors of results for the 1-bit CS framework [15]: 1) a lower bound is provided on the best achievable performance of this 1-bit CS framework, and if the elements of Φ are drawn randomly from i.i.d. Gaussian distribution or its rows are drawn uniformly from the unit sphere, then the solution will have bounded error on the order of the optimal lower bound. 2) A condition on the mapping A, binary ϵ-stable embedding (BϵSE), that enables stable reconstruction is given to characterize the reconstruction performance even when some of the measurement signs have changed (e.g., due to noise in the measurements).

Since this problem was introduced and studied by Boufounos and Baraniuk in 2008 [13], it has been studied by many people and several algorithms have been developed [13], [15], [16], [17], [18], [19], [20]. Binary iterative hard thresholding (BIHT) [15] is shown to perform better than other algorithms such as matching sign pursuit (MSP) [16] and restricted-step shrinkage (RSS) [18] in both reconstruction error as well as consistency, see [15] for more details. The experiment in [15] shows that the one-sided ℓ_1 objective

This work is supported by NSF Grant DMS-0714945, DMS-0914561, Center for Domain-Specific Computing (CDSC) under the NSF Expeditions in Computing Award CCF-0926127, ONR Grant N00014-11-1-0719, N00014-08-1-119 and an ARO MURI subcontract from Rice University.

M. Yan, Y. Yang and S. Osher are with the Department of Mathematics, University of California, Los Angeles, CA, 90095 USA. E-mail: yanm@math.ucla.edu; yyang@math.ucla.edu; sjo@math.ucla.edu. Phone: (310)-825-1758. Fax: (310)-206-6673.

[1] Generally, the M-dimensional Boolean cube is defined as $\{0, 1\}^M$. Without loss of generality, we use $\{-1, 1\}^M$ instead.

(BIHT) performs better when there are only a few errors, and the one-sided ℓ_2 objective (BIHT-ℓ_2) performs better when there are significantly more errors, which implies that BIHT-ℓ_2 is useful when the measurements contain significant noise that might cause a large number of sign flips.

In practice, there will always be noise in the measurements during acquisition and transmission, therefore, a robust algorithm for 1-bit compressive sensing when the measurements flip their signs is strongly needed. One possible way to build this robust algorithm is to introduce an outlier detection technique.

There are many applications where the accurate detection of outliers is needed. For example, when an image is corrupted by random-valued impulse noise, the corrupted pixels are useless in image denoising. There are some methods (e.g., adaptive center-weighted median filter (ACWMF) [21]) for detecting the damaged pixels. But these methods will miss quit a lot of real noise and false-hit some noise-free pixels when the noise level is high. In [22], we proposed a method to adaptively detect the noise pixels and restore the image with ℓ_0 minimization. Instead of detecting the damaged pixels before recovering the image, we iteratively restore the image and detect the damaged pixels. This idea works really well for impulse noise removal. In this 1-bit compressive sensing framework, when there is a sign flip in one measurement, this measurement will worsen the reconstruction performance. If we can detect all the measurements with sign flips, then we can change the signs for these measurements and improve the reconstruction performance a lot. However, it is much more difficult than detecting impulse noise and there is no method for detecting sign flips, but we can still utilize the idea in [22] to adaptively find the sign flips. In this paper, we will introduce a method for robust 1-bit compressive sensing which can detect the sign flips and reconstruct the signals with very high accuracy even when there are a large number of sign flips.

This paper is organized as follows. We will introduce several algorithms for reconstructing the signal and detecting the sign flips in section II. Section III studies the case when the noise information is not given. The performance of these algorithms is illustrated in section IV with comparison to BIHT and BIHT-ℓ_2. We will end this work by a short conclusion.

II. Robust 1-bit Compressive Sensing using Adaptive Outlier Pursuit

Binary iterative hard thresholding (BIHT or BIHT-ℓ_2) in [15] is the algorithm for solving

$$\begin{array}{rl} \min\limits_{x} & \sum\limits_{i=1}^{M} \phi(y_i,(\Phi x)_i) \\ \text{subject to:} & \|x\|_2=1, \quad \|x\|_0 \le K, \end{array} \tag{3}$$

where ϕ is the one-sided ℓ_1 (or ℓ_2) objective:

$$\phi(x,y)=\left\{\begin{array}{ll} 0, & \text{if } x\cdot y>0, \\ |x\cdot y| \ (\text{or } |x\cdot y|^2/2), & \text{otherwise.} \end{array}\right. \tag{4}$$

The high performance of BIHT is demonstrated when all the measurements are noise-free. However when there are a lot of sign flips, the performance of BIHT and BIHT-ℓ_2 is worsened by the noisy measurements. There is no method to detect the sign flips in the measurements, but adaptively finding the sign flips and reconstructing the signals can be combined together as in [22] to obtain better performance.

Let us assume firstly that the noise level (the ratio of the number of sign flips over the number of measurements for 1-bit compressive sensing) is provided. Based on this information, we can choose a proper integer L such that at most L elements of the total measurements are wrongly detected (having sign flips). For measurements $y \in \{-1,1\}^M$, $\Lambda \in \mathbf{R}^M$ is a binary vector denoting the "correct" data:

$$\Lambda_i=\left\{\begin{array}{ll} 1, & \text{if } y_i \text{ is "correct"}, \\ 0, & \text{otherwise.} \end{array}\right. \tag{5}$$

According to the assumption, we have $\sum\limits_{i=1}^{M}(1-\Lambda_i) \le L$.

Introducing Λ into the old problem solved by BIHT, we have the following new problem with unknown variable x and Λ:

$$\begin{array}{rl} \min\limits_{x,\Lambda} & \sum\limits_{i=1}^{M} \Lambda_i \phi(y_i,(\Phi x)_i) \\ \text{s.t.} & \sum\limits_{i=1}^{M}(1-\Lambda_i) \le L, \\ & \Lambda_i \in \{0,1\} \quad i=1,2,\cdots,M, \\ & \|x\|_2=1, \quad \|x\|_0 \le K. \end{array} \tag{6}$$

The above model can also be interpreted in the following way. Let us consider the noisy measurement y as the sign of Φx with additive unknown noise n, i.e. $y=\text{sign}(\Phi x+n)$. Though the binary measurement is robust to noise as long as the sign does not change, there exist some n_i's such that the corresponding measurements change. In our problem, only a few measurements are corrupted, and only these corresponding n_i's are important. Therefore, n can be considered as sparse noise with nonzero entries at these locations, and we have to recover the signal x from sparse corrupted measurements [23], [24], even when the measurements is acquired by taking the sign of $\Phi x+n$. This equivalent problem is

$$\begin{array}{rl} \min\limits_{x,n} & \sum\limits_{i=1}^{M} \phi(y_i,(\Phi x)_i+n_i) \\ \text{s.t.} & \|n\|_0 \le L, \\ & \|x\|_2=1, \quad \|x\|_0 \le K. \end{array} \tag{7}$$

The equivalence is described in the appendix.

The problem defined in (6) is non-convex and has both continuous and discrete variables. It is difficult to solve it together, thus we use alternating minimization method, which separates the energy minimization over x and Λ into two steps:

- Fix Λ and solve for x:

$$\begin{array}{rl} \min\limits_{x} & \sum\limits_{i=1}^{M} \Lambda_i \phi(y_i,(\Phi x)_i) \\ \text{s.t.} & \|x\|_2=1, \quad \|x\|_0 \le K. \end{array} \tag{8}$$

  This is the same as (3) with revised Φ and y. We only need to use the ith rows of Φ and y where $\Lambda_i=1$.

- Fix x and update Λ:

$$\begin{aligned} \min_{\Lambda} \quad & \sum_{i=1}^{M} \Lambda_i \phi(y_i, (\Phi x)_i) \\ \text{s.t.} \quad & \sum_{i=1}^{M} (1-\Lambda_i) \leq L, \\ & \Lambda_i \in \{0,1\} \quad i = 1, 2, \cdots, M. \end{aligned} \tag{9}$$

This problem is to choose $M-L$ elements with least sum from M elements $\{\phi(y_i, (\Phi x)_i)\}_{i=1}^{M}$. Given an x estimated from (8), we can update Λ in one step:

$$\Lambda_i = \begin{cases} 0, & \text{if } \phi(y_i, (\Phi x)_i) \geq \tau, \\ 1, & \text{otherwise.} \end{cases} \tag{10}$$

where τ is the L^{th} largest term of $\{\phi(y_i, (\Phi x)_i)\}_{i=1}^{M}$. If the L^{th} and $(L+1)^{th}$ larges terms are equal, then we can choose any Λ such that $\sum_{i=1}^{M} \Lambda_i = M - L$ and

$$\min_{i, \Lambda_i = 0} \phi(y_i, (\Phi x)_i) \geq \max_{i, \Lambda_i = 1} \phi(y_i, (\Phi x)_i).$$

Since for each step, the updated Λ identifies the outliers, this method is named as adaptive outlier pursuit (AOP). When $L = 0$, this is exactly the BIHT proposed in [15]. Our algorithm is as follows:

Algorithm 1 AOP

Input: $\Phi \in \mathbf{R}^{M \times N}$, $y \in \{-1, 1\}^M$, $K > 0$, $L \geq 0$, $\alpha > 0$, Miter > 0
Initialization: $x^0 = \Phi^T y / \|\Phi^T y\|$, $k = 0$, $\Lambda = \mathbf{1} \in \mathbf{R}^M$, Loc = $1:M$, tol $=$ inf, TOL $=$ inf.
while $k \leq$ Miter and $L \leq$ tol **do**
  Compute $\beta^{k+1} = x^k + \alpha \Phi(\text{Loc}, :)^T (y(\text{Loc}) - \text{sign}(\Phi(\text{Loc}, :) x^k))$.
  Update $x^{k+1} = \eta_K(\beta^{k+1})$,
  Set tol $= \|y - A(x^{k+1})\|_0$.
  if tol $\leq$TOL **then**,
    Compute Λ with (10).
    Update Loc to be the location of 1-entries of Λ.
    Set TOL = tol.
  end if
  $k = k + 1$.
end while
return $x^k / \|x^k\|$.

$\eta_K(v)$ computes the best K-term approximation of v by thresholding. Since $y_i \in \{-1,\ 1\}$, once we find the locations of the errors, instead of deleting these data, we can also "correct" them by flipping their signs. Hence x can also be updated with Φ and this new measurements. This algorithm with changing signs is called AOP with flips.

Remark: Similar to BIHT-ℓ_2, we can also choose the one-sided ℓ_2 objective instead of ℓ_1 objective and obtain two other algorithms.

III. The case with L unknown

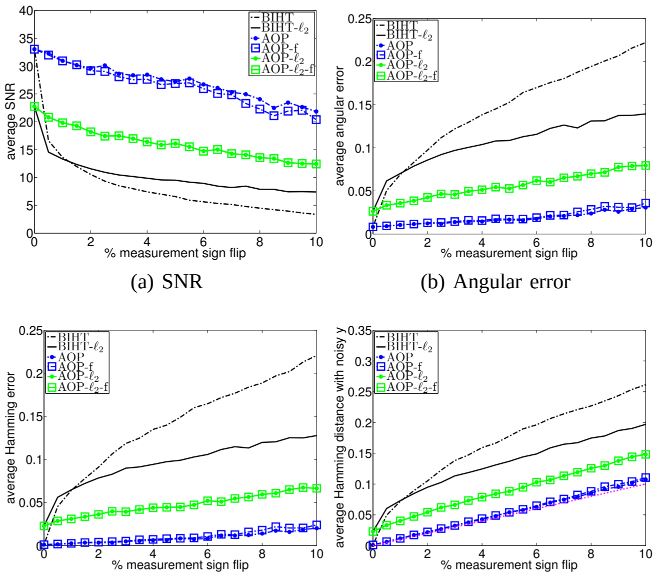
In previous section, we assume that L, the number of corrupted measurements, is known in advance. However in real world applications there are cases when no pre-knowledge about the noise is given. If L is chosen smaller or larger than the true value, the performance of these algorithms will get worse. As shown numerically in section IV, when L is less than the true value, even if the L detections are completely correct, some sign flips still stay in the measurements. On the other hand, some correct measurements will be lost if L is too large, and the problem will have more solutions if the number of total measurements is not large enough, which will affect the accuracy of the algorithm. Therefore, in this scenario we have to apply an L detection skill to find an L which is not far from the true value.

When no noise information is given, the following procedure can be applied to predict L. The first-phase preparation is to do extensive experiments on simulated data with known L and record the Hamming distances between $A(x)$ and noisy y of BIHT-ℓ_2 and AOP. Here we can simply use the results in our first experiment in section IV. The average of the results describes nicely the behavior of these two algorithms at different noise levels. Hence a formula can be derived to predict the Hamming distance of AOP based on the results obtained by BIHT-ℓ_2. This could be a fair initial guess for the noise level, and we can derive an L based on the result, labeled as L_0. Then we calculate $L_t = \|A(x) - y\|_0$ using the result x gained by AOP with L_0 as the input for L. If L_t is greater than L_0, which means that L_0 is too small while L_t is too large, we set L_t as the upper bound $L_{\max}$ and L_0 as the lower bound $L_{\min}$. Otherwise, if L_t is smaller than or equal to L_0, which means L_0 may be too large, we use μL_0 $(0 < \mu < 1)$ as the new L_0 to look for new L_t. We will keep doing this until L_t is greater than L_0. Then the previous L_0 is defined as the upper bound $L_{\max}$ and the new L_0 is defined as the lower bound $L_{\min}$. This is just one method for finding lower and upper bounds for L, and there are certainly other possible ways to decide the bounds. Then we use the bisection method to find a better L. The mean of $L_{\max}$ and $L_{\min}$ (L_{mean}) is then used as input to derive L_t with AOP. If L_t is greater than L_{mean}, we update $L_{\min}$ with L_{mean}. Otherwise, L_{mean} is set as $L_{\max}$. This bisection method is applied to update these two bounds until $L_{\max} - L_{\min} \leq 1$. The final $L_{\min}$ is our input L.

IV. Numerical Results

In this section we use several numerical experiments to demonstrate the effectiveness of AOP algorithms. Here AOP is implemented in the following four ways: 1) AOP with one-sided ℓ_1 objective (AOP); 2) AOP with flips and one-sided ℓ_1 objective (AOP-f); 3) AOP with one-sided ℓ_2 objective (AOP-ℓ_2); 4) AOP with flips and one-sided ℓ_2 objective (AOP-ℓ_2-f). The four algorithms, together with BIHT and BIHT-ℓ_2, are studied and compared in the following experiments.

The setup for our experiments is as follows. We first generate a matrix $\Phi \in \mathbf{R}^{M \times N}$ whose elements follow i.i.d. Gaussian distribution. Then we generate the original K-sparse signal $x^* \in \mathbf{R}^N$. Its non-zero entries are drawn from standard Gaussian distribution and then normalized to have norm 1. $y^* \in \{-1, 1\}^M$ is computed by $A(x^*)$.



(a) SNR

(b) Angular error

(c) Hamming error between $A(x)$ and $A(x^*)$

(d) Hamming distance between $A(x)$ and y

Fig. 1: Algorithm comparison on corrupted data with different noise levels. (a) average SNR vs noise level, (b) average angular error vs noise level, (c) average Hamming error between $A(x)$ and $A(x^*)$ vs noise level, (d) average Hamming distance between $A(x)$ and noisy measurements y vs noise level. AOP proves to be more robust to measurement sign flips compared with BIHT.

A. noise levels test

In our first experiment, we set $M = N = 1000$, $K = 10$, and examine the performance of these algorithms on data with different noise levels. Here in each test, we choose a few measurements at random and flip their signs. The noise level is between 0% and 10% and we assume it is known in advance. For each level, we perform 100 trials and record the average signal-to-noise ratio (SNR), average reconstruction angular error for each reconstructed signal x with respect to x^*, average Hamming error between $A(x)$ and $A(x^*)$ and average Hamming distance between $A(x)$ and the noisy measurements y. Here SNR is denoted by $10\log_{10}(\|x\|^2/\|x-x^*\|^2)$, angular error is defined as $\arccos\langle x, x^*\rangle/\pi$, Hamming error stands for $\|A(x)-A(x^*)\|_0/M$ and the Hamming distance between $A(x)$ and y, defined as $\|A(x)-y\|_0/M$, is used to measure the difference between $A(x)$ and the noisy measurements y. The results are depicted in Figure 1. The plots demonstrate that in these comparisons four AOP algorithms outperform BIHT and BIHT-ℓ_2 for all noise levels, significantly so when more than 2% of the measurements are corrupted. Compared with BIHT, BIHT-ℓ_2 tends to give worse results when there are only a few sign flips in y and better results if we have high noise level. This has been shown and studied in [15]. Of all the AOP series, AOP and AOP-f give better results compared with AOP-ℓ_2 and AOP-ℓ_2-f. We can also see that there is a lot of overlap between the results obtained by AOP and the ones acquired by AOP with flips, especially when one-sided ℓ_2 objective is used, the results are almost the same. Figure 1(d) compares the average Hamming distances between $A(x)$ and the noisy measurements y for all algorithms. If the sign flips can be found correctly, then the Hamming distance between $A(x)$ and y should be equal to noise level. The result shows that average Hamming distances for AOP and AOP-f are slightly above the noise levels, which means that AOP with one-sided ℓ_1 objective performs better in consistency than other algorithms in noisy cases.

In order to show that our algorithms can find the positions of sign flips with high accuracy, we measure the probabilities of correct detections of sign flips in the noisy measurements for different noise levels from 0.5% to 10% in Figure 2 ($M = N = 1000$, $K = 10$). The exact number of sign flips is used as L in the algorithms and we compare the exact locations of sign flips in measurements y with those detected from the algorithms for all 100 trials, then the average probabilities of correct detections are shown for different algorithms at different noise levels. From this figure, we can see that all four algorithms have high accuracy in detecting the sign flips. When the noise level is low ($\leq$4%), the accuracy of AOP and AOP-f can be as high as 95%, even when the noise level is high (e.g., 10%), the accuracy of AOP and AOP-f is still above 90%. Comparing to algorithms with one-sided ℓ_1 objective, algorithms with one-sided ℓ_2 objective have lower accuracy. The accuracy for AOP-ℓ_2 and AOP-ℓ_2-f is around 80%.

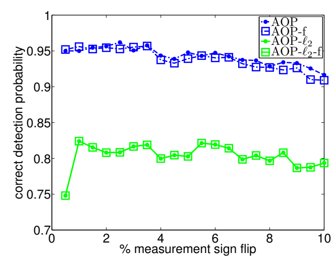
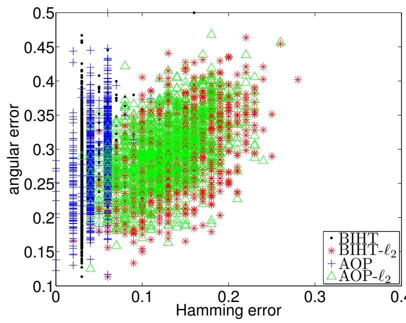


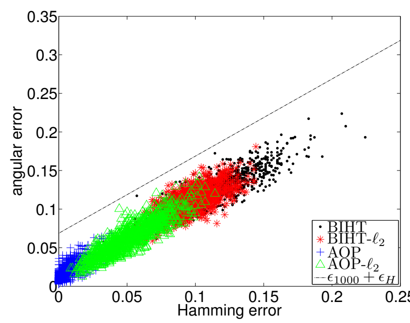
Fig. 2: The probabilities of correct detections of sign flips for different noise levels ranging from 0.5% to 10%. AOP and AOP-f have very high accuracy (great than 90%) in detecting the sign flips, while AOP-ℓ_2 and AOP-ℓ_2-f have relatively lower accuracy (around 80%).

B. M/N test

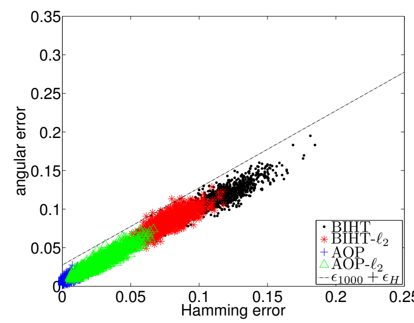
In the second experiment, $N = 1000$, $K = 10$ and the noise level 3% are fixed, and we change M/N within the range $(0, 2]$. 40 different M/N are considered and we perform 300 tests for each value. The results are displayed in five different ways: the average SNR, average angular error, average Hamming error between $A(x)$ and $A(x^*)$, average Hamming distance between $A(x)$ and y and average percentage of coefficient "misses". Here "misses" stands for the coefficients where $x_i^* \neq 0$ while $x_i = 0$. According to Figure 3, although all the algorithms show the same trend as M/N increases, AOP and AOP-f always obtain a much smaller angular error (higher SNR) than BIHT and BIHT-ℓ_2. There are also fewer coefficient misses in the results acquired by AOP series. Furthermore, we see that even when 3% of the measurements are corrupted, AOP can still recover a



(a) M=0.1N

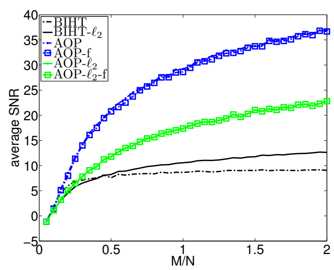


(b) M=0.7N

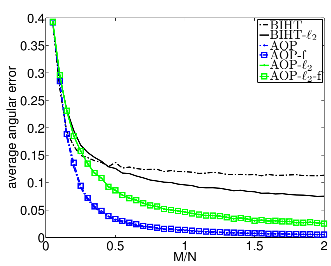


(c) M=1.5N

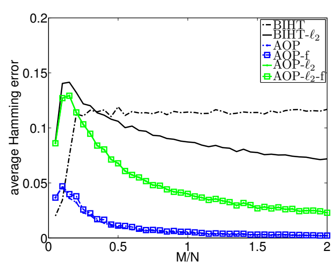
Fig. 4: Hamming error vs angular error with different M. AOP gives the most consistent results for $M = 0.7N$ and $M = 1.5N$. In these two cases we can see a linear relationship $\epsilon_{\text{sim}} \approx C+\epsilon_H$ between the average angular error ϵ_{sim} and average Hamming error ϵ_H, where C is constant. For really small M ($M = 0.1N$) BIHT returns almost the same results as AOP since AOP may fail to find the exact sign flips in the noisy measurements. The dashed line $\epsilon_{1000} + \epsilon_H$ is a upper bound for 1000 trials.



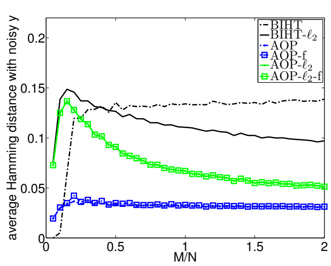
(a) SNR



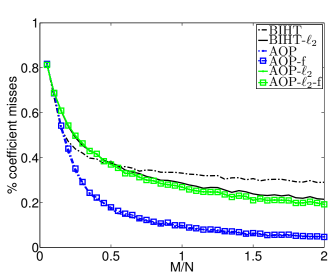
(b) Angular error



(c) Hamming distance between $A(x)$ and $A(x^*)$



(d) Hamming error between $A(x)$ and y



(e) % coefficient misses

Fig. 3: Algorithm comparison on corrupted data with different M/N. (a) average SNR vs M/N, (b) average angular error vs M/N, (c) average Hamming error between $A(x)$ and $A(x^*)$ vs M/N, (d) average Hamming distance between $A(x)$ and y vs M/N, (e) average percentage of coefficient misses vs M/N. AOP yields a remarkable improvement in reducing the Hamming and angular error and achieving higher SNR.

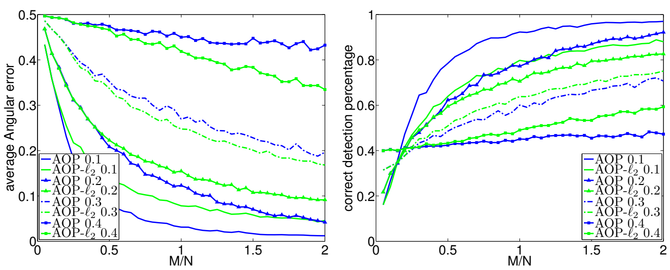
signal with SNR greater than 20 using less than 0.5 bits per coefficient of x^*. In Hamming error comparison, AOP and AOP-f beat other algorithms significantly when $M/N > 0.15$. Moreover, we see that the average Hamming error of AOP and AOP-f is extremely close to zero when $M/N > 0.5$. When $M/N < 0.15$, the seemingly failure of AOP and AOP-f compared with BIHT is due to the fact that there are usually more than one solution to (6) for really small M, and with high probability our method will return one solution with L sign flips, which may not be the actual ones. Hence we may not be able to detect the actual errors in the measurements.

We also try to explore the relationship between the Hamming error between $A(x)$ and $A(x^*)$ and the reconstruction angular error. With $N = 1000$, $K = 10$ and the noise level 3% fixed, we plot the Hamming error vs angular error for three different M in Figure 4. Since AOP and AOP with flips tend to return almost the same results if we use the same objective (one-sided ℓ_1 or one-sided ℓ_2) for x update, we only compare the results acquired by BIHT, BIHT-ℓ_2, AOP and AOP-ℓ_2. We can see clearly that almost all the blue (+) points stay in the lower left part of the graph for $M = 0.7N$ and $M = 1.5N$, which proves that AOP gives more consistent results compared with other three algorithms. For these two M, the average angular error is close to a linear function of average Hamming error, which is predicted by BϵSE property in [15]. We also plot an empirical upper bound for AOP of $\epsilon_{1000} + \epsilon_H$ defined in [15], where ϵ_{1000} is the largest angular error of AOP and ϵ_H is the Hamming distance. For especially "under-sampled" case like $M = 0.1N$, none of these algorithms is able to return consistent reconstructions, as we can see the points scatter almost randomly over the domain. In this case the results obtained by BIHT stay really close to those gained by AOP. As mentioned above, this is because AOP may not be able to detect the exact sign flips in the noisy measurements when M is too small.

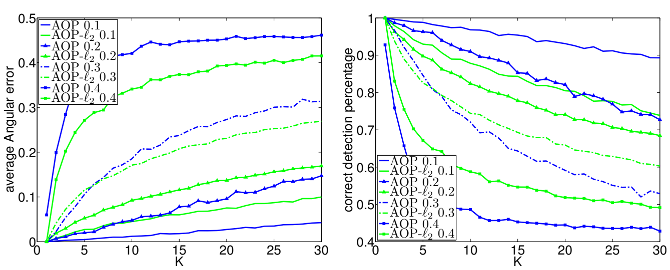
C. high noise levels

In this subsection, we study the performance of AOP and AOP-ℓ_2 when a large number of measurements are corrupted.

Two settings are considered. In the first experiment, we fix $N = 1000$, $K = 10$, and change the M/N ratio between 0.05 and 2. Four different noise levels are considered from 0.1 to 0.4 and we record the average angular error and correct detection probability from 100 tests. In the second setting, we fix $M = 2000$, $N = 1000$ and change K from 1 to 30. Still, four noise levels are considered and the mean results from 100 tests are recorded. From Figure 5 (a) and (b), we can see similar trend for the behavior of angular error and correct detection probability as we have discovered in Figure 3. According to (c), (d), for all the noise levels the performance of these two algorithms tends to get worse as K increases. We also have another interesting discovery that when the noise level is greater than 0.2, AOP-ℓ_2 turns out to be a better choice than AOP. This is because when the noise level is extremely high, even with outlier detection technique, lots of sign flips remain in the recovered measurements, and this new "noise level" is still relatively high. According to [15], BIHT-ℓ_2 outperforms BIHT when the measurements contain lots of sign flips. Therefore, when the noise level is high enough, AOP-ℓ_2 is considered as a better choice compared with AOP.



(a) Angular Error vs M/N (b) Correct Detection vs M/N

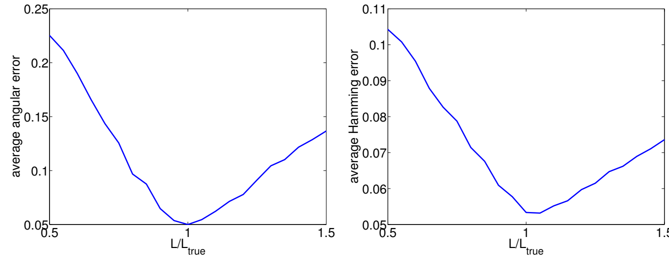


(c) Angular Error vs K (d) Correct Detection vs K

Fig. 5: AOP and AOP-ℓ_2 performance under different noise levels. (a) average angular error vs M/N with different noise levels, (b) correct detection percentage vs M/N with different noise levels, (c) average angular error vs K with different noise levels, (d) correct detection percentage vs K with different noise levels. The performance gets better when we increase M/N or decrease K.

D. L mismatch

In Figure 6, we analyze the influence of incorrect selection of L on AOP. Here we choose $M = N = 1000$, $K = 10$, noise level 5%, and change the input value from $0.5L$ to $1.5L$. 100 tests are conducted and the mean results are recorded. It is easily seen that the error will become larger when the input L digresses from its true value. According to this plot, we know that in order to obtain good performance for our method, we should choose a proper L as input.

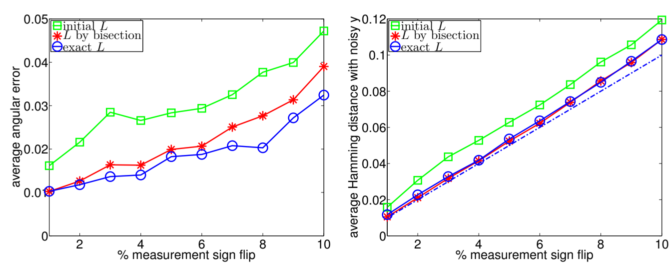


(a) Angular Error (b) Hamming Error

Fig. 6: AOP performance with different L inputs. L has to stay close to its true value in order to get good performance.

E. unknown L

To show that our method works even when L is not given, we use the method described in Section III to find an approximation of L, and compare the results of AOP with different L. Here $M = N = 1000$, $K = 10$ are fixed, and 10 different noise levels (from 1% to 10%) are tested. Three inputs for L: the initial L_0 predicted from the result of BIHT-ℓ_2, L obtained from bisection method, exact L, are used in AOP to obtain the results. The following figure 7 is depicted with the average results from 100 trials. Even with the initial L_0, the results are comparable to those with exact L, and bisection method can provide a better approximation for L with longer time for predicting L.



(a) Angular error (b) Hamming distance with noisy y

Fig. 7: Comparison of results by different L at different noise levels from 1% to 10%. (a) average angular error vs noise level, (b) average Hamming distance between $A(x)$ and noisy y vs noise level. By choosing appropriate L as the input, we can still obtain the results comparable to those with exact L.

V. Conclusion

In this paper, we propose a method based on adaptive outlier pursuit for robust 1-bit compressive sensing. By iteratively detecting the sign flips in measurements and recovering the signals from "correct" measurements, this method can obtain better results in both finding the noisy measurements and recovering the signals, even when there are a lot of sign flips in the measurements. Four algorithms (AOP, AOP-f, AOP-ℓ_2 and AOP-ℓ_2-f) are given based on this method, and the performances of these four algorithms are shown in the numerical experiments. The algorithms based on one-sided ℓ_1 objective (AOP and AOP-f) have better performance compared to the other two algorithm (AOP-ℓ_2 and AOP-ℓ_2-f), which are based on one-sided ℓ_2 objective when the noise level is not high (less than 20%), when the noise level is extremely high, AOP-ℓ_2 is a better choice compared with AOP. In addition, we proposed a simple method to find a candidate for the number of sign flips L when L is unknown and the numerical experiments show that the performance of AOP with this inexact input L is comparable with that of exact L.

APPENDIX

In this appendix, we show the equivalence of problem (6) and (7). If (x, n) satisfies the constraints of problem (7), we can define

$$\Lambda_i = \begin{cases} 1, & \text{if } n_i = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

then we have $\phi(y_i, (\Phi x)_i + n_i) = 0$ if $\Lambda_i = 0$, since we can always find n_i such that $\phi(y_i, (\Phi x)_i + n_i) = 0$ for fixed x. If $\Lambda_i = 1$, we have $n_i = 0$, thus $\phi(y_i, (\Phi x)_i + n_i) = \phi(y_i, (\Phi x)_i)$. Therefore, problem (7) is equivalent to

$$\begin{array}{ll} \min\limits_{x,n} & \sum\limits_{i=1}^{M} \Lambda_i \phi(y_i, (\Phi x)_i) \\ \text{s.t.} & \|n\|_0 \le L, \\ & \|x\|_2 = 1, \quad \|x\|_0 \le K. \end{array} \tag{12}$$

From the relation of Λ and n in (11), we know the constraint $\|n\|_0 \le L$ in the above problem can be replaced with the constraints on Λ defined in (6). Therefore, problem (6) and (7) are equivalent.

REFERENCES

[1] E. Candès, "Compressive sampling," in *Int. Congress of Mathematics*, vol. 3, Madrid, Spain, 2006, pp. 1433–1452.

[2] E. J. Candès, J. K. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[7] J. Z. Sun and V. K. Goyal, "Quantization for Compressed Sensing Reconstruction," in *SAMPTA'09, International Conference on Sampling Theory and Applications*, L. Fesquet and B. Torrésani, Eds., Marseille, France, 2009, p. Special session on sampling and quantization.

[8] W. Dai, H. V. Pham, and O. Milenkovic, "Distortion-rate functions for quantized compressive sensing," in *IEEE Information Theory Workshop on Networking and Information Theory*, 2009, pp. 171–175.

[9] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 149–152, 2010.

[10] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429–443, 2011.

[11] L. Jacques, D. K. Hammond, and M.-J. Fadili, "Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine." *IEEE Transactions on Information Theory*, pp. 559–571, 2011.

[12] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *ArXiv preprint arXiv:1202.1212*, 2012.

[13] P. T. Boufounos and R. G. Baraniuk, "One-bit compressive sensing," in *Conference on Information Sciences and Systems (CISS)*, Princeton, March 2008.

[14] J. N. Laska and R. G. Baraniuk, "Regime change: Bit-depth versus measurement-rate in compressive sensing," *ArXiv preprint arXiv:1110.3450*, 2011.

[15] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *ArXiv preprint arXiv:1104.3160*, 2011.

[16] P. T. Boufounos, "Greedy sparse signal reconstruction from sign measurements," in *Proceedings of the 43rd Asilomar conference on Signals, systems and computers*, ser. Asilomar'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1305–1309.

[17] A. Gupta, R. Nowak, and B. Recht, "Sample complexity for 1-bit compressed sensing and sparse classification," in *Proceedings of the IEEE International Symposium on Information Theory*, 2010, pp. 1553–1557.

[18] 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 Transactions on Signal Processing*, vol. 59, no. 11, pp. 5289–5301, 2011.

[19] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *ArXiv preprint arXiv:1109.4299*, 2011.

[20] T. Zhou and D. Tao, "Hamming compressed sensing," *ArXiv preprint arXiv:1110.0073*, 2011.

[21] T. Chen and H. R. Wu, "Adaptive impulse dectection using center-weighted median filters," *IEEE Signal Processing Letters*, vol. 8, no. 1, pp. 1–3, 2001.

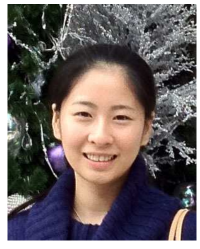
[22] M. Yan, "Restoration of images corrupted by impulse noise using blind inpainting and ℓ_0 norm," *UCLA CAM report 11-72*, 2011.

[23] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Proceedings of the 43rd Asilomar conference on Signals, systems and computers*, ser. Asilomar'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1556–1560.

[24] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Transactions on Information Theory*, to appear.

Ming Yan received the B.S. and M.S. degrees in computational Mathematics from University of Science and Technology of China, Hefei, China, in 2005 and 2008, respectively. He is currently pursuing the Ph.D. degree from the Department of Mathematics, University of California, Los Angeles.

His current research interests include variational and optimization methods for image and signal processing.



Yi Yang received the B.S. degrees in Mathematics from University of Science and Technology of China, Hefei, China, in 2009. She is currently pursuing the Ph.D. degree from the Department of Mathematics, University of California, Los Angeles.

Her current research interests include optimization and its applications to image and signal processing.

Stanley Osher received the Ph.D. degree in mathematics from New York University's Courant Institute of Mathematical Sciences.

He is a Professor of Mathematics, Computer Science and Electrical Engineering at UCLA. He is also the Director of Special Projects of the NSF funded Institute for Pure and Applied Mathematics. He is a member of the National Academy of Sciences, the American Academy of Arts and Sciences and is one of the top 25 most highly cited researchers in both mathematics and computer sciences.

He has received numerous academic honors and has co-founded three successful companies, each based largely on his own (joint) research. His current interests mainly involve information science which includes graphics, image processing, compressed sensing and machine learning.

He has co-invented and/or co-developed the following widely used algorithms: (1) Essentially nonoscillatory (ENO), weighted essentially nonoscillatory (WENO) and other shock capturing schemes for hyperbolic systems of conservation laws and their analogues for Hamilton-Jacobi equations. (2) The level set method for capturing dynamic surface evolution. (3) Total variation and other partial differential based methods for image processing. (4) Bregman iterative methods for L1 and related regularized problems which arrive in compressive sensing, matrix completion, imaging and elsewhere. (5) Diffusion generated motion by mean curvature and other threshhold dynamics methods.

His website is http//www.math.ucla.edu/∼sjo. His recent papers are on the UCLA Math Dept's CAM website:
http://www.math.ucla.edu/applied/cam/index.shtml.