Nonconvex Penalties with Analytical Solutions for One-bit Compressive Sensing

Xiaolin Huang[a], Ming Yan[b,*]

[a] *Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and MOE Key Laboratory of System Control and Information Processing, Shanghai 200240, P.R. China.*

[b] *Department of Computational Mathematics, Science and Engineering and Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA*

Abstract

One-bit measurements widely exist in the real world and can be used to recover sparse signals. This task is known as one-bit compressive sensing (1bit-CS). In this paper, we propose novel algorithms based on both convex and nonconvex sparsity-inducing penalties for robust 1bit-CS. We consider the dual problem, which has only one variable and provides a sufficient condition to verify whether a solution is globally optimal or not. For positive homogeneous penalties, a globally optimal solution can be obtained in two steps: a proximal operator and a normalization step. For other penalties, we solve the dual problem, and it needs to evaluate the proximal operators for many times. Then we provide fast algorithms for finding analytical solutions for three penalties: minimax concave penalty (MCP), ℓ_0 norm, and sorted ℓ_1 penalty. Specifically, our algorithm is more than 200 times faster than the existing algorithm for MCP. Its efficiency is comparable to the algorithm for the ℓ_1 penalty in time, while its performance is much better than ℓ_1. Among these penalties, sorted ℓ_1 is most robust to noise in different settings.

Keywords: one-bit compressed sensing, nonconvex penalty, analytical solutions

1. Introduction

Analog-to-digital converting (ADC) is a necessary process in digital processing, and the choice of the bit-depth is an important issue. The extreme case is to use one-bit measurements, which enjoy many advantages, e.g., they can be implemented by one low power comparator running at a high rate. Mathematically, one-bit compressive sensing (1bit-CS) is to recover a K-sparse vector $\mathbf{x} \in \mathbb{R}^n$ ($\|\mathbf{x}\|_0 \leq K$) from m one-bit quantized measurements

$$y_i = \mathrm{sgn}(\mathbf{u}_i^\top \mathbf{x} + \varepsilon_i), \tag{1}$$

where $\mathbf{u}_i \in \mathbb{R}^n$ is the ith sensing vector, ε_i is the noise in the measurement, and the function sgn returns 1 for a positive number and -1 otherwise. The sensing system and measurements are represented by $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m]$ and $\mathbf{y} = [y_1, y_2, \ldots, y_m]^\top$, respectively. Due to the low power and high sampling rate, one-bit measurements have been applied in the estimation of frequency, phase, and direction of arrival (DOA) [1, 2, 3]. For example, in the DOA estimation, a radar with one-bit measurements has a higher scan speed than others. One-bit measurements are also attractive in distributed networks [4, 5], where the use of one-bit measurements largely reduces the communication load.

If the underlying signal is sparse, then sparsity pursuit techniques can help signal recovery, which is similar to the regular compressive sensing. Therefore, since its proposal by [6], 1bit-CS has attracted much attention in both the signal processing society ([7, 8, 9, 10]) and the machine learning society ([11, 12, 13, 14]). Because the one-bit information has no capability to specify the magnitude of the original signal, we assume $\|\mathbf{x}\|_2 = 1$ without loss of generality (there is also some work on norm estimation, see, e.g., [15]), and 1bit-CS can be

*Corresponding author.
Email addresses: xiaolinhuang@sjtu.edu.cn (Xiaolin Huang), yanm@math.msu.edu (Ming Yan)

explained as finding the sparest vector on the unit sphere that coincides with the observed signs, i.e,

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \|\mathbf{x}\|_0, \\ \text{subject to} \quad & y_i = \text{sgn}(\mathbf{u}_i^\top \mathbf{x}), \quad \forall i = 1, 2, \ldots, m, \\ & \|\mathbf{x}\|_2 = 1. \end{aligned} \tag{2}$$

This is an NP-hard problem, and several algorithms are developed to approximately solve it or its variants [6, 7, 16, 17]. The constraint in (2) does not tolerate noise or sign flips, and it may exclude the real signal from the feasible set. Additionally, the feasible set may be empty, and there is no solution for (2). One way to deal with noise and sign flips is to replace the constraint $y_i = \text{sgn}(\mathbf{u}_i^\top \mathbf{x})$ by a loss function. For example, the one-sided ℓ_1 loss and the one-sided ℓ_2 loss are considered in [18] and [8]; the linear loss is used in [9] and [11]. It is reported that the linear loss generally outperforms the one-sided ℓ_1/ℓ_2 loss. Moreover, with proper regularization terms and constraints, the linear loss minimization can be solved analytically and enjoys great computational effectiveness.

In regular CS problems, nonconvex penalties have been insightfully investigated and widely applied to enhance sparsity. Similarly, those nonconvex techniques are applicable to 1bit-CS, and the recovery performance is expected to be improved. One obvious barrier is that nonconvex penalties lead to nonconvex problems, which are usually difficult to solve. An interesting result is recently represented in [12], which gives analytical solutions for two nonconvex penalties, namely the smoothly clipped absolute deviation (SCAD, [19]) and minimax concave penalty (MCP, [20]). Also [13] proposes an algorithm for 1bit-CS using the k-support norm. These nonconvex penalties are shown to obtain better results than convex ones in both theory and practice [12, 13] and, therefore, have been extended to other applications including the multi-label learning task [21].

In this paper, we discuss more convex and nonconvex penalties, for which analytical solutions can be obtained, and we provide fast algorithms for finding these solutions. These penalties include SCAD, MCP, ℓ_p-norm ($0 \leq p \leq +\infty$, [22]), ℓ_1-ℓ_2 norm [23], sorted ℓ_1 penalty [25, 26], and so on. The contributions of this paper can be summarized as follows.

- We analyze a generic model for 1bit-CS and provide a sufficient condition for the global optimality.

- For positive homogeneous penalties, we show that an optimal solution can be obtained in two steps: a proximal operator and a normalization step. For general penalties, we provide a generic algorithm by solving the dual problem.

- We provide algorithms for finding analytical solutions for three nonconvex penalties: MCP, ℓ_0 norm, and the sorted ℓ_1 penalty. These algorithms are much faster than the existing 1bit-CS algorithms for nonconvex penalties and even comparable to that for the convex ℓ_1 minimization problem, e.g., our algorithm is averagely 200 times faster than the algorithm given in [12] for MCP. In addition, we compare these nonconvex penalties with the convex ℓ_1 penalty and show that the sorted ℓ_1 performs the best in both performance and computational time.

The rest of this paper is organized as follows. Section 2 briefly reviews the existing related 1bit-CS algorithms. The main contributions, i.e., analytical solutions for different penalties and corresponding algorithms, are presented in Section 3. The numerical experiments are reported in Section 4. We end this paper with a brief conclusion.

2. Related Works

Model (2) for 1bit-CS has two main disadvantages: i) it is difficult to solve because of the ℓ_0 norm in the objective and the constraint $\|\mathbf{x}\|_2 = 1$; ii) the constraint $y_i = \operatorname{sgn}(\mathbf{u}_i^\top \mathbf{x})$ does not consider noisy sign measurements.

Several approaches are given to deal with both disadvantages. For the non-convexity, the ℓ_0 norm is replaced by the ℓ_1 norm, and the constraint $\|\mathbf{x}\|_2 = 1$ is replaced by other convex constraints. The first convex model [27] for 1bit-CS is

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \|\mathbf{x}\|_1, \\ \text{subject to} \quad & y_i(\mathbf{u}_i^\top \mathbf{x}) \geq 0, \ \forall i = 1, 2, \ldots, m, \\ & \|\mathbf{U}^\top \mathbf{x}\|_1 = r, \end{aligned} \tag{3}$$

where r is a given positive constant. In fact, the solutions for all positive r's have the same direction and the difference is only on the magnitudes of the reconstructed signals.

However, (3) still can not be applied when there are noisy measurements, because, it, same as (2), requires the sign consistence in the measurements. Noisy measurements come from both the noise during the acquisition before the quantization and sign flips during the transmission. To deal with noisy measurements, [18] introduces the following robust model using the one-sided ℓ_1 norm,

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \frac{1}{m}\sum_{i=1}^{m} \max\left\{0, -y_i(\mathbf{u}_i^\top \mathbf{x})\right\}, \\ \text{subject to} \quad & \|\mathbf{x}\|_2 = 1, \\ & \|\mathbf{x}\|_0 = K. \end{aligned}$$

The robust model using the one-sided ℓ_2 norm is also introduced. Several modifications are designed by [8], [28], and [29] to improve their robustness to sign flips and noise.

The linear loss for robost 1bit-CS attracts more attention because of its good performance and simplicity. Based on the linear loss, many results on sampling complexities are given recently [9, 11, 12, 13]. In [9], the first model using the linear loss for 1bit-CS is proposed and takes the following formulation,

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & -\frac{1}{m}\sum_{i=1}^{m} y_i(\mathbf{u}_i^\top \mathbf{x}), \\ \text{subject to} \quad & \|\mathbf{x}\|_2 \le 1, \\ & \|\mathbf{x}\|_1 \le s, \end{aligned} \tag{4}$$

where s is a given positive constant. One can also put the ℓ_1-norm in the objective instead of in the constraint, resulting in the problem given by [11],

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \mu\|\mathbf{x}\|_1 - \frac{1}{m}\sum_{i=1}^{m} y_i(\mathbf{u}_i^\top \mathbf{x}), \\ \text{subject to} \quad & \|\mathbf{x}\|_2 \le 1, \end{aligned} \tag{5}$$

where μ is the regularization parameter for the ℓ_1-norm. Note that the unit sphere constraint $\|\mathbf{x}\|_2 = 1$ is relaxed to the unit ball constraint $\|\mathbf{x}\|_2 \le 1$ in (4) and (5). As illustrated by [11], with proper parameters, this relaxation will not change the solution, which generally comes from the properties of the linear loss.

One attractive property for (5) over (4) is that there is a closed-form solution for (5), and thus, solving (5) is faster than (4), though both problems are equivalent in the sense that the solutions are the same for corresponding parameters s and μ. The convex penalty in (5) is replaced by several nonconvex penalties such as MCP [12] and k-support norm [13]. Better sampling complexities can be achieved for these nonconvex penalties and analytical solutions are obtained.

In this paper, we consider general penalties in (5) and derive efficient algorithms for many popular penalties by solving the dual problem. Even for many nonconvex penalties, our algorithms can find the global optimal solutions. These algorithms will help people investigate more theoretical properties for these nonconvex penalties such as the sampling complexity and consider better modifications such as adaptive sampling.

3. Analytical Solutions for 1bit-CS

In this section, we consider the following generic optimization problem for robust 1bit-CS,

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \qquad & f(\mathbf{x}) - \frac{1}{m}\sum_{i=1}^{m} y_i\langle \mathbf{u}_i, \mathbf{x}\rangle + \frac{\tau}{2}\|\mathbf{x}\|_2^2, \\ \text{subject to} \qquad & \|\mathbf{x}\|_2^2 \le 1, \end{aligned} \tag{6}$$

where $\tau \ge 0$ and $f(\mathbf{x})$ is the penalty. Most existing papers assume that $\tau = 0$, and the choice of $\tau > 0$ is introduced by [12]. We will show that there is no need to choose a positive τ because optimal solutions do not depend on τ when τ is small enough and optimal solutions are not on the unit sphere for a large τ. Let $\mathbf{v} = \frac{1}{m}\sum_{i=1}^{m} y_i\mathbf{u}_i$ and the objective function is

$$F(\mathbf{x}) = f(\mathbf{x}) - \langle \mathbf{v}, \mathbf{x}\rangle + \frac{\tau}{2}\|\mathbf{x}\|_2^2.$$

We make the following assumption for $f(\mathbf{x})$.

Assumption 3.1. $f(\mathbf{x}) \ge 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *and* $f(\mathbf{0}) = 0$.

The convexity of (6) depends on the penalty $f(\mathbf{x})$ and τ. For a convex $f(\mathbf{x})$, problem (6) is convex, and it could also be convex even if $f(\mathbf{x})$ is nonconvex when $\tau > 0$ is large enough. In order to find its global solution, we solve the corresponding dual problem and check whether the duality gap is zero, i.e., the optimal primal value is the same as the optimal dual value. We define the corresponding Lagrangian functional as

$$\mathcal{L}(\mathbf{x}, \mu) = F(\mathbf{x}) + \frac{\mu}{2}(\|\mathbf{x}\|_2^2 - 1), \tag{7}$$

and the following lemma gives a sufficient condition for a global optimal solution of problem (6).

Lemma 3.1. *[30, Theorem 6.2.5] If there exist* $(\mathbf{x}^*, \mu^*)$ *such that* $\|\mathbf{x}^*\|_2^2 \leq 1$, $\mu^* \geq 0$, $\mathcal{L}(\mathbf{x}^*, \mu^*) \leq \mathcal{L}(\mathbf{x}, \mu^*)$ *for all* $\mathbf{x}$*, and* $\mu^*(\|\mathbf{x}^*\|_2^2 - 1) = 0$*, then* $\mathbf{x}^*$ *and* μ^* *are optimal solutions to the primal and dual problems, respectively, with no duality gap.*

Proof. We have, for any $\mu \geq 0$,

$$\begin{aligned}\mathcal{L}(\mathbf{x}^*, \mu) =& F(\mathbf{x}^*) + \frac{\mu}{2}(\|\mathbf{x}^*\|_2^2 - 1) \leq F(\mathbf{x}^*) \\ =& F(\mathbf{x}^*) + \frac{\mu^*}{2}(\|\mathbf{x}^*\|_2^2 - 1) = \mathcal{L}(\mathbf{x}^*, \mu^*).\end{aligned}$$

The inequality arises from the fact $\|\mathbf{x}^*\|_2^2 \leq 1$, and the second equality holds because of $\mu^*(\|\mathbf{x}^*\|_2^2 - 1) = 0$. Thus, $(\mathbf{x}^*, \mu^*)$ is a saddle point of $\mathcal{L}(\mathbf{x}, \mu)$, i.e.,

$$\mathcal{L}(\mathbf{x}^*, \mu) \leq \mathcal{L}(\mathbf{x}^*, \mu^*) \leq \mathcal{L}(\mathbf{x}, \mu^*), \tag{8}$$

for any $\mu \geq 0$ and $\mathbf{x}$. Thus,

$$F(\mathbf{x}^*) = \mathcal{L}(\mathbf{x}^*, \mu^*) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mu^*).$$

The duality gap is zero, and $\mathbf{x}^*$ and μ^* are optimal solutions to the primal and dual problems, respectively. □

Based on the lemma, we first solve the dual problem because it is concave and easy to solve in many cases. Then, we find $\mathbf{x}^*$ and verify whether $\mu^*(\|\mathbf{x}^*\|_2^2 - 1) = 0$ is satisfied. If it is satisfied, then $\mathbf{x}^*$ is a global optimal solution of (6).

3.1. Positive Homogeneous Penalties

Assume that $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$ for any positive α, i.e., $f(\mathbf{x})$ is positive homogeneous. We can obtain a global solution in two steps: a proximal operator and a normalization step. Some positive homogeneous penalties are listed here.

- ℓ_p norm ($0 < p \leq +\infty$): e.g., ℓ_1 norm [11]; [1]
- ℓ_p norm minus ℓ_q norm: e.g., $\ell_1 - \ell_2$ norm [23, 24].

[1] ℓ_0 "norm" is not positive homogeneous and can not be applied here.

- 0 function: Lemma 4.1 from paper [12].

- sorted ℓ_1 penalty: nonconvex ones [31]; convex ones [26]; indicator function of ℓ_0 [13]; small magnitude penalized (SMAP) [32].

- One-sided norm [33].

- Gauges [34].

Note that for a given $\mathbf{x} \neq 0$, we have $f(\mathbf{x}) = \frac{df(\alpha\mathbf{x})}{d\alpha} = \langle \tilde{\nabla} f(\alpha\mathbf{x}), \mathbf{x} \rangle$, where $\tilde{\nabla} f(\alpha\mathbf{x})$ is a generalized subgradient of f at $\alpha\mathbf{x}$ [35, Definition 8.3]. Let $\alpha = 1$, and we have $f(\mathbf{x}) = \langle \tilde{\nabla} f(\mathbf{x}), \mathbf{x} \rangle$. Define the proximal operator of f as

$$\mathrm{Prox}_f(\mathbf{v}) := \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{v}\|_2^2,$$

and let $\mathbf{t}^* \in \mathrm{Prox}_f(\mathbf{v})$. We have the following lemma.

Lemma 3.2. *If $f(\mathbf{x})$ is positive homogeneous, we have that*

$$f(\mathbf{t}^*) - \langle \mathbf{t}^*, \mathbf{v} \rangle = -\|\mathbf{t}^*\|_2^2.$$

Proof. When $\mathbf{t}^* = \mathbf{0}$, we have $f(\mathbf{0}) = 0$ because of $f(\mathbf{x})$ being positive homogeneous, and the result is trivial. When $\mathbf{t}^* \neq \mathbf{0}$, we have $f(\mathbf{t}^*) = \langle \tilde{\nabla} f(\mathbf{t}^*), \mathbf{t}^* \rangle$. Therefore,

$$\begin{aligned} f(\mathbf{t}^*) - \langle \mathbf{t}^*, \mathbf{v} \rangle =& \langle \tilde{\nabla} f(\mathbf{t}^*), \mathbf{t}^* \rangle - \langle \mathbf{t}^*, \mathbf{v} \rangle \\ =& \langle \tilde{\nabla} f(\mathbf{t}^*) - \mathbf{v}, \mathbf{t}^* \rangle = -\|\mathbf{t}^*\|_2^2. \end{aligned}$$

The last equality is satisfied because $\mathbf{t}^* \in \mathrm{Prox}_f(\mathbf{v})$. ∎

Theorem 3.1. *If $f(\mathbf{x})$ is positive homogeneous and $\mathbf{t}^* = \mathrm{Prox}_f(\mathbf{v})$, then an optimal solution for (6) is*

$$\mathbf{x}^* = \begin{cases} \mathbf{t}^*/\tau(\text{or } \mathbf{0} \text{ if } \tau = 0) & \text{if } \|\mathbf{t}^*\|_2 \leq \tau, \\ \mathbf{t}^*/\|\mathbf{t}^*\|_2 & \text{if } \|\mathbf{t}^*\|_2 > \tau. \end{cases}$$

Proof. When $\tau > 0$, we have

$$\mathcal{L}(\mathbf{x}, \mu) = f(\mathbf{x}) + \frac{\tau + \mu}{2}\left\|\mathbf{x} - \frac{\mathbf{v}}{\tau + \mu}\right\|_2^2 - \frac{\|\mathbf{v}\|_2^2}{2(\tau + \mu)} - \frac{\mu}{2}.$$

Then $\mathbf{x}^* = \mathbf{t}^*/(\tau+\mu)$ is optimal for a given μ, and

$$\min_{\mathbf{x}} \mathcal{L}(\mathbf{x},\mu) = \frac{2f(\mathbf{t}^*) + \|\mathbf{t}^*\|_2^2 - 2\langle \mathbf{t}^*, \mathbf{v}\rangle}{2(\tau+\mu)} - \frac{\mu}{2} = \frac{-\|\mathbf{t}^*\|_2^2}{2(\tau+\mu)} - \frac{\mu}{2}.$$

The last equality comes from Lemma 3.2. Thus the dual problem is a concave function of μ and we can find μ^* as

$$\mu^* = \begin{cases} 0 & \text{if } \|\mathbf{t}^*\|_2 \le \tau, \\ \|\mathbf{t}^*\|_2 - \tau & \text{if } \|\mathbf{t}^*\|_2 > \tau. \end{cases} \tag{9}$$

Therefore we have

$$\mathbf{x}^* = \begin{cases} \mathbf{t}^*/\tau & \text{if } \|\mathbf{t}^*\|_2 \le \tau, \\ \mathbf{t}^*/\|\mathbf{t}^*\| & \text{if } \|\mathbf{t}^*\|_2 > \tau. \end{cases}$$

Thus $\mu^*(\|\mathbf{x}^*\|_2^2 - 1) = 0$ is satisfied, and Lemma 3.1 shows that $\mathbf{x}^*$ is a global optimal solution of (6).

Let $\tau = 0$. When $\mu > 0$, we have

$$\min_{\mathbf{x}} \mathcal{L}(\mathbf{x},\mu) = \frac{-\|\mathbf{t}^*\|_2^2}{2\mu} - \frac{\mu}{2} < 0.$$

Then we consider the case when $\mu = 0$. From Lemma 3.2, we have

$$\mathcal{L}(\mathbf{t}^*, 0) = f(\mathbf{t}^*) - \langle \mathbf{t}^*, \mathbf{v}\rangle = -\|\mathbf{t}^*\|_2^2.$$

Therefore, the positive homogeneity of f gives

$$\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, 0) = -\infty \text{ if } \mathbf{t}^* \neq \mathbf{0}.$$

When $\mathbf{t}^* = 0$, $\mathbf{t}^* \in \mathrm{Prox}_f(\mathbf{v})$ gives us that, for any $\mathbf{x}$,

$$\frac{1}{2}\|\mathbf{v}\|_2^2 \le f(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{v}\|_2^2,$$

which implies

$$-\frac{1}{2}\|\mathbf{x}\|^2 \le f(\mathbf{x}) - \langle \mathbf{x}, \mathbf{v}\rangle,$$

and the positive homogeneity of f shows that $\mathcal{L}(\mathbf{x}, 0) = f(\mathbf{x}) - \langle \mathbf{x}, \mathbf{0}\rangle \ge 0$ for all $\mathbf{x}$ and $\mathcal{L}(\mathbf{0}, 0) = 0$. Therefore,

$$\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, 0) = 0 \text{ if } \mathbf{t}^* = \mathbf{0}.$$

Together, we have $\mu^* = \|\mathbf{t}^*\|_2$ and $\mathbf{x}^* = \mathbf{t}^*/\mu = \mathbf{t}^*/\|\mathbf{t}^*\|_2$ if $\mathbf{t}^* \neq \mathbf{0}$. When $\mathbf{t}^* = \mathbf{0}$, we have $\mu^* = 0$ and $\mathbf{x}^* = \mathbf{0}$.

When $\tau = 0$, Lemma 3.1 tells us that

$$\mathbf{x}^* = \begin{cases} \mathbf{0} & \text{if } \|\mathbf{t}^*\|_2 = 0, \\ \mathbf{t}^*/\|\mathbf{t}^*\|_2 & \text{if } \|\mathbf{t}^*\|_2 > 0, \end{cases}$$

is a global optimal solution of (6). In fact, there may be multiple global solutions when $\mathbf{t}^* = \mathbf{0}$ (See [36] for examples). □

In sum, a globally optimal solution of (6) can be obtained in two steps: a proximal operator and a normalization step.

Algorithm 1: General Algorithm for Positive Homogeneous Penalties

Input: $\mathbf{v}$, f
Output: $\mathbf{x}$
$\mathbf{t}^* = \mathrm{Prox}_f(\mathbf{v})$
$\mathbf{x}^* = \mathbf{t}^*/\|\mathbf{t}^*\|_2$

Remark 3.1. *When $\tau < \|\mathbf{t}^*\|_2$, we have $\mathbf{x}^* = \mathbf{t}^*/\|\mathbf{t}^*\|_2$, i.e., $\mathbf{x}^*$ does not depend on τ. When $\tau > \|\mathbf{t}^*\|_2$, we have $\mathbf{x}^* = \mathbf{t}^*/\tau$, i.e., $\mathbf{x}^*$ is not on the unit sphere. Therefore, there is no need to choose a positive τ, and we let $\tau = 0$ in the numerical experiments. Note this result is consistent with Lemma 4.1 of [12] which shows that oracle estimators will not change when τ is small enough.*

3.2. General Penalties

For a general $f(\mathbf{x})$, we consider the dual function

$$G(\mu) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mu). \tag{10}$$

Given μ, let $\mathbf{x}^*(\mu)$ be an optimal solution of (10) defined as

$$\mathbf{x}^*(\mu) \in \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mu) = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{(\tau+\mu)}{2} \left\| \mathbf{x} - \frac{\mathbf{v}}{\tau+\mu} \right\|_2^2. \tag{11}$$

The following theorem provides the subdifferential of G.

Theorem 3.2. *Given $\mu \geq 0$, we have, for any $\tilde{\mu} \geq 0$,*

$$G(\tilde{\mu}) \leq G(\mu) + \frac{1}{2}(\|\mathbf{x}^*(\mu)\|^2 - 1)(\tilde{\mu} - \mu). \tag{12}$$

Proof. Using the definition of G in (10), we derive

$$\begin{aligned} G(\tilde{\mu}) =& \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \tilde{\mu}) \leq \mathcal{L}(\mathbf{x}^*(\mu), \tilde{\mu}) \\ =& \mathcal{L}(\mathbf{x}^*(\mu), \mu) + \frac{1}{2}(\|\mathbf{x}^*(\mu)\|_2^2 - 1)(\tilde{\mu} - \mu) \\ =& G(\mu) + \frac{1}{2}(\|\mathbf{x}^*(\mu)\|_2^2 - 1)(\tilde{\mu} - \mu), \end{aligned}$$

where the second equality follows from the definition of $\mathcal{L}$ in (7), and the last equality is valid because of (10) and (11). $\square$

The previous theorem shows that $(1 - \|\mathbf{x}^*(\mu)\|_2^2)/2 \in \partial(-G)(\mu)$, where $\partial(-G)$ is the subdifferential of $-G$. Note that when there are multiple optimal solutions of (10) for a given μ, the subdifferential of $-G$ is $[\min\{(1 - \|\mathbf{x}^*(\mu)\|_2^2)/2\}, \max\{(1 - \|\mathbf{x}^*(\mu)\|_2^2)/2\}]$. Then $G(\mu)$ being concave gives us a way to find the optimal μ^*. For general penalties, we turn to solve the dual problem, i.e., finding the maximizer of $G(\mu)$. The dual function is concave and has one variable, so many optimization methods can be applied. However, in the evaluation of the subgradient of G, a proximal operator is needed. Therefore, many evaluations of the proximal operator is needed.

When $\mu = +\infty$, we have $\mathbf{x}^*(+\infty) = \mathbf{0}$. If there exists an optimal solution of (10) such that $\|\mathbf{x}^*(0)\|_2^2 < 1$, then $G(\mu)$ is decreasing for $\mu \in [0, +\infty)$, and we have $\|\mathbf{x}^*(\mu)\|_2^2 < 1$ for all $\mu > 0$ and the optimal $\mu^* = 0$. Then $\mathbf{x}^*(0)$ is an optimal solution of (6) because of Lemma 3.1. Otherwise, we have to find μ^* such that $\|\mathbf{x}^*(\mu^*)\|_2^2 = 1$ or $0 \in \partial(-G)(\mu^*)$. If we find μ^* such that $\|\mathbf{x}^*(\mu^*)\|_2^2 = 1$ is satisfied, then $\mathbf{x}^*(\mu^*)$ is an optimal solution of (6) by Lemma 3.1, otherwise $\mu^*(\|\mathbf{x}^*\|_2^2 - 1) = 0$ is not satisfied and whether $\mathbf{x}^*(\mu^*)$ is an optimal solution of (6) is unknown.

Remark 3.2. *The following example shows that $\mathbf{x}^*(\mu^*)$ can still be optimal for (6) even when $\mu^*(\|\mathbf{x}^*\|_2^2 - 1) = 0$ is not satisfied for the optimal μ^*. Let $F(x) = \|x\|_0 - x/2$, then we have that the optimal solution is $x^* = 0$. The dual function of μ is*

$$G(\mu) = \min\left(-\frac{\mu}{2}, 1 - \frac{1}{8\mu} - \frac{\mu}{2}\right),$$

and the optimal $\mu^* = 1/8$. The optimal x^*'s for μ^* are 0 and 4. Thus, we can still find $x^*(\mu^*) = 0$ as a global optimal solution of $F(x)$. However, in this case, the primal-dual gap is not zero.

Remark 3.3. *If τ is small enough such that $\|\mathbf{x}^*(0)\|_2 > 1$ for all $\mathbf{x}^*(0)$, $\mathbf{x}^*$ does not depend on τ. When τ is large enough such that $\|\mathbf{x}^*(0)\|_2 < 1$ for some $\mathbf{x}^*(0)$, then $\mathbf{x}^*$ is not on the unit sphere. Therefore, there is no need to choose a positive τ, which is the same as in the case of positive homogeneous penalties, and we let $\tau = 0$ in the numerical experiments.*

In order to find a global optimal solution of (6) efficiently, we have to make sure that the proximal operator has an analytical solution, because the proximal operator is evaluated for multiple times. Some penalties that have analytical solutions are:

- MCP and its generalizations [20, 37, 38],
- SCAD [19],
- ℓ_0 norm,
- $\ell_{1/2}$ regularization [39],
- Partial regularization [40].

In the following subsections, we describe algorithms for MCP, ℓ_0 norm, and the nonconvex sorted ℓ_1. Our algorithm is different from that in [12] for MCP.

3.3. Minimax Concave Penalty

Let $f(\mathbf{x}) = \sum_{i=1}^n g_{\lambda,b}(x_i)$ and $g_{\lambda,b}$ be defined as

$$g_{\lambda,b}(x) = \begin{cases} \lambda|x| - x^2/(2b), & \text{if } |x| \le b\lambda, \\ b\lambda^2/2, & \text{if } |x| > b\lambda, \end{cases}$$

for fixed parameters $\lambda > 0$ and $b > 0$. The analytical solutions for (11) can be obtained.

When $\mu \leq 1/b$, we have

$$x^*(\mu) = \begin{cases} 0, & \text{if } v^2 \leq b\lambda^2\mu, \\ v/\mu, & \text{if } v^2 \geq b\lambda^2\mu, \end{cases} \tag{13}$$

and when $\mu > 1/b$, we have

$$x^*(\mu) = \begin{cases} 0, & \text{if } |v| \leq \lambda, \\ \frac{|v|-\lambda}{\mu-\frac{1}{b}}\mathrm{sgn}(v), & \text{if } \lambda < |v| \leq b\lambda\mu, \\ v/\mu, & \text{if } |v| \geq b\lambda\mu. \end{cases} \tag{14}$$

For some μ, we have two optimal solutions, as shown in the formulation. The resulting algorithm is shown in Alg. 2.

3.4. ℓ_0 norm

Let $f(\mathbf{x}) = \lambda\|\mathbf{x}\|_0$, and the analytical solutions for (11) is:

$$x^*(\mu) = \begin{cases} 0, & \text{if } v^2 \leq 2\lambda\mu, \\ v/\mu, & \text{if } v^2 \geq 2\lambda\mu. \end{cases}$$

The resulting algorithm is shown in Alg. 3.

3.5. Sorted ℓ_1 Penalty

Let $f(\mathbf{x}) = \lambda \sum_{i=1}^{n} w_i \left|x_{[i]}\right|$, where

$$x_{[1]}, x_{[2]}, \ldots, x_{[n]} = \mathrm{Sort}(|x_1|, |x_2|, \ldots, |x_n|)$$

is sorted by the absolute component values. Since weight w_i's are assigned according to the sort, this regularization is called sorted ℓ_1 penalty. When $w_1 \leq w_2 \leq \ldots \leq w_n$, it is convex [26]. Otherwise, it is nonconvex and can be used to enhance sparsity [31]. For the nonconvex case, a typical weight setting is

$$w_i = \begin{cases} 1, & i < n_1, \\ \exp(-5i/n_1), & i \geq n_1, \end{cases} \tag{15}$$

Algorithm 2: MCP

```
Input: λ, b
Output: μ
Initialize: μ = 1/b
v_[1], v_[2], ..., v_[n] = Sort(|v_1|, |v_2|, ..., |v_n|)
Find L such that v_[L] ≤ λ < v_[L+1]
d_2 = Σ_{i=L+1}^{n} v_[i]^2
d_max = b^2 d_2
if d_max > 1 then
  i = L + 1; d_1 = 0
  while d_max > 1 do
    μ = v_[i]/(bλ); d_max = d_1/(μ − 1/b)^2 + d_2/μ^2
    if d_max < 1 then
      Solve d_1/(μ − 1/b)^2 + d_2/μ^2 = 1; return
    end if
    d_1 = d_1 + (v_[i] − λ)^2; d_2 = d_2 − v_[i]^2
    d_max = d_1/(μ − 1/b)^2 + d_2/μ^2
    i = i + 1
  end while
else
  i = L;
  while d_max < 1 do
    μ = v_[i]^2/(bλ^2); d_max = d_2/μ^2
    if d_max > 1 then
      μ = √d_2; return
    end if
    d_2 = d_2 + v_[i]^2
    d_max = d_2/μ^2
    i = i − 1
  end while
end if
```

where n_1 is a parameter related to the sparsity. Since the signal in 1bit-CS is very sparse, $n_1 = 10$ is used in the numerical experiments. An optimal solution can be analytically given by

$$t^*_{[i]} = \max\left\{|v_{[i]}| - w_i\lambda, 0\right\} \operatorname{sgn}(v_{[i]}).$$

Because this penalty is positive homogeneous, we apply the proximal operator first and then a normalization step. The corresponding algorithm is given in

Algorithm 3: ℓ_0 norm

Input: λ
Output: μ
Initialize: $i = n$
$v_{[1]}, v_{[2]}, \ldots, v_{[n]} =$ Sort$(|v_1|, |v_2|, \ldots, |v_n|)$
$\mu = v_{[i]}^2/(2\lambda)$; $d = v_{[i]}^2$; $d_{\max} = d/\mu^2$
while $d_{\max} < 1$ **do**
  $i = i - 1$; $\mu = v_{[i]}^2/(2\lambda)$; $d_{\max} = d/\mu^2$
  if $d_{\max} > 1$ **then**
    $\mu = \sqrt{d}$; return
  else
    $d = d + v_{[i]}^2$; $d_{\max} = d/\mu^2$
  end if
end while

Alg. 4.

Algorithm 4: Sorted ℓ_1 Penalty

Input: λ, $\mathbf{w}$ (decreasing)
Output: μ
Initialize: $\mu = 0$
$v_{[1]}, v_{[2]}, \ldots, v_{[n]} =$ Sort$(|v_1|, |v_2|, \ldots, |v_n|)$
for $i = 1 : n$ **do**
  $t_{[i]} = \max\{|v_{[i]}| - w_i\lambda, 0\}\mathrm{sgn}(v_{[i]})$
end for
if $\|\mathbf{t}\| > 0$ **then**
  $\mu = \|\mathbf{t}\|$
end if

4. Numerical Experiments

In numerical experiments, we randomly choose K components from a n-dimensional signal, draw their values from the Gaussian distribution, and normalize the signal onto the unit ℓ_2-norm ball. Then, m sign observations are generated by (1), where ε is the Gaussian noise with noise level s_n, which stands for the ratio between the variances of the measurements and ε. We also consider sign flips with ratio 10%. All the experiments are done with Matlab 2014b on Core i5-3.10GHz and 8.0GB RAM.

Before considering the recovery accuracy, we compare the computational time of Alg. 2 and the algorithm in [12]. Both algorithms solve the same problem with the MCP penalty. By the concavity of the dual function and Theorem 3.2, the dual function is piecewise smooth and its subgradient is decreasing. So we can find the optimal μ^* or the interval that contains the optimal μ^*, and we find $\mathbf{x}^*$ from (13) or (14). Therefore, there is at most one single variate problem, i.e., $d_1/(\mu - 1/b)^2 + d_2/\mu^2 = 1$, to solve. While in [12], this problem is solved for $n - L$ times, and there are many redundant computation steps.

To numerically compare the computational time, several pairs of m and n are considered. For a fair comparison, we choose the same parameters for the MCP regularization $g_{\lambda,b}(x_i)$ as $\lambda = 0.1$ and $b = 3$. Then the average and the standard derivation of computational times over 100 trials are reported in Table 1, where computational time of the Passive algorithm [11] is given as well.

Table 1: Average Computational Time

m	n	Passive	Zhu's Alg.	Alg. 2
500	1000	3.6 ± 1.2 ms	1.51 ± 0.04 s	4.0 ± 0.8 ms
1000	1000	6.7 ± 1.4 ms	1.76 ± 0.05 s	8.9 ± 1.2 ms
1000	2000	14.8 ± 1.4 ms	6.95 ± 0.14 s	18.5 ± 1.3 ms
2000	2000	25.1 ± 1.7 ms	7.15 ± 0.16 s	30.2 ± 2.2 ms
5000	5000	148 ± 5.1 ms	43.6 ± 1.48 s	184 ± 17 ms

The above result illustrates that the proposed analytical solution based algorithm can significantly reduce the computational burden from Zhu's algorithm. Compared with the Passive algorithm, which solves the ℓ_1 minimization problem, Alg. 2, which solves the nonconvex MCP regularized problem, but the difference in computational time is minor. The comparison in performance is in the rest of this section.

As discussed previously, our analysis covers many possible nonconvex regularizations. Section 3 gives several such kinds of algorithms including

- Alg. 2 for minimizing the MCP penalty;
- Alg. 3 for ℓ_0 minimization;
- Alg. 4 for the nonconvex sorted ℓ_1 penalty.

It can be expected that those nonconvex regularizers can improve the recovery quality from the ℓ_1 norm when there are not plenty of measurements. In the following, we will report the performance of Alg. 2-4 and the passive algorithm.

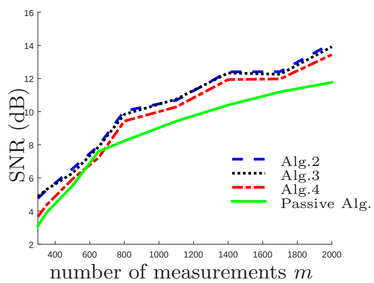
To select the parameters, we consider the following two method: i) choose the parameters based on the ℓ_2 distance to the real signal, which results in "ideal" parameters. With ideal parameters, we can evaluate the best performance of each algorithm in each data sets; ii) tune the parameters by cross-validation based on consistency, which is a practical method. The selected parameters are not necessarily the best, especially when there are only a few observations. The comparison between the ideal and the selected parameters also implies the robustness of these methods to different parameters.

First, we vary the number of measurements m from 300 to 2000 and report the recovered quality in Fig. 1, where the signal-to-noise ratio in dB, i.e.,

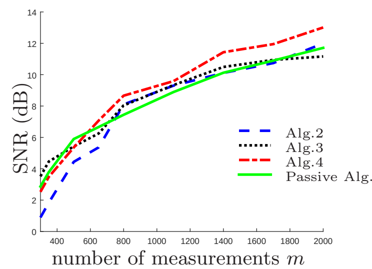
$$\mathrm{SNR}_{\mathrm{dB}}(\bar{\mathbf{x}}, \tilde{\mathbf{x}}) = 10 \log_{10} \left(\frac{\|\bar{\mathbf{x}}\|_2^2}{\|\bar{\mathbf{x}} - \tilde{\mathbf{x}}\|_2^2} \right),$$

is used to measure the quality of the recovered signal ($\bar{\mathbf{x}}$ is the real signal and $\tilde{\mathbf{x}}$ is the recovered one). Unless the amount of measurements is too small that no meaningful recovered signal can be obtained, the three nonconvex penalties can improve the performance from ℓ_1 minimization if the optimal parameters can be obtained. In this case, MCP and ℓ_0 achieve high recovery quality. The SNRs obtained by the sorted ℓ_1 minimization is a bit worse. When we select parameters by cross-validation, Alg. 4 performs better than Alg. 2 and 3. The comparison indicates that the sorted ℓ_1 is more stable to different parameters.

Notice that the four algorithms have no significant difference on computational time and, among these three nonconvex methods, Alg.4 is most efficient. Specifically in this experiment, when $n = 1000, m = 1000$, the average computational times are: 6.7ms (Passive), 8.9ms (Alg.2), 8.2ms (Alg.3), and 6.8ms (Alg.4). When $n = 10000, m = 5000$, the average computational times are: 284ms (Passive), 301ms (Alg.2), 295ms (Alg.3), 291ms (Alg.4).



(a)

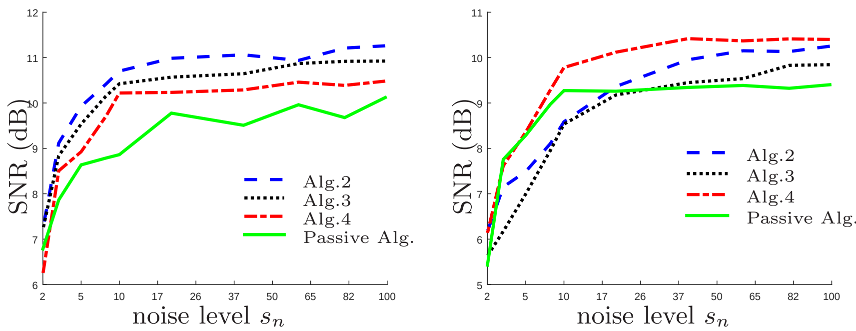


(b)

Figure 1: Recovery performance for different numbers of measurements: MCP minimization (blue dashed line), ℓ_0 minimization (black dotted line), sorted ℓ_1 penalty (red dash-dotted line), and ℓ_1 minimization (green solid line). In this experiment $n = 1000, K = 15, s_n = 10$, and sign flip ratio is 10%. (a) using the ideal parameters; (b) using parameters selected by 10-fold cross-validation.

Similar observations can be found in Fig. 2, where different noise levels are considered. Generally, the three algorithms can both tolerate the existence of noise and outliers (10% of the sign measurements are flipped). When the noise is not heavy, e.g., when ratio between the variance of the noise and that of the real measurements is below 0.1, the three proposed algorithms have good noise suppression.

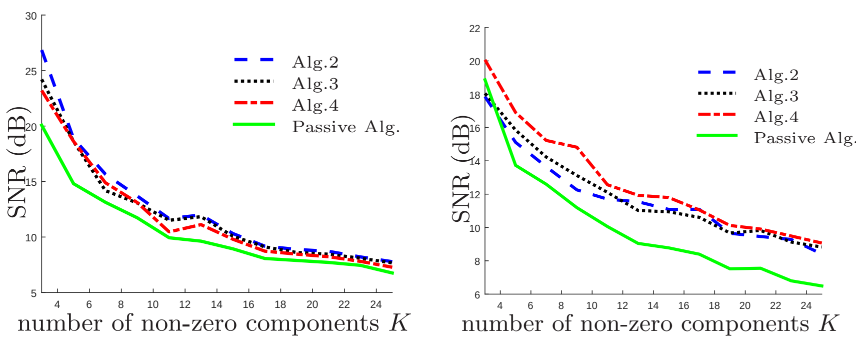
Last, we consider different numbers of non-zero components K with $n = 1000$, $m = 1000$, and $s_n = 10$. For the sorted ℓ_1 penalty, there is one parameter n_1 in its weight (15) that is related to the signal sparsity. In the previous experiments where K is fixed to be 15, we use $n_1 = 10$ without tuning. Though that value is not optimal, the performance of the sorted ℓ_1 penalty is generally satisfying. In this experiment, we select n_1 from $\{2, 4, ..., 16\}$ for different K's.



(a) (b)

Figure 2: Recovery performance for different noise levels: MCP minimization (blue dashed line), ℓ_0 minimization (black dotted line), sorted ℓ_1 penalty (red dash-dotted line), and ℓ_1 minimization (green solid line). In this experiment $n = 1000, K = 15, m = 1000$, and sign flip ratio is 10%. (a) using the ideal parameters; (b) using parameters selected by 10-fold cross-validation.

Totally, there are two parameters to tune for the sorted ℓ_1 penalty, the same as MCP. In Fig. 3(a) and 3(b), the average SNRs for ideal and selected parameters are displayed, respectively.



(a) (b)

Figure 3: Recovery performance for different sparsity levels: MCP minimization (blue dashed line), ℓ_0 minimization (black dotted line), sorted ℓ_1 penalty (red dash-dotted line), and ℓ_1 minimization (green solid line). In this experiment $n = 1000, m = 1000, s_n = 10$, and sign flip ratio is 10%. (a) using the ideal parameters; (b) using parameters selected by 10-fold cross-validation.

Besides SNR, there are also other signal recovery criteria including:

- angular error:

$$\mathrm{AE}(\bar{\mathbf{x}}, \tilde{\mathbf{x}}) = \frac{1}{\pi} \arccos\left(\frac{\bar{\mathbf{x}}^T \tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|_2}\right);$$

- inconsistency ratio:

$$\mathrm{INR}(\bar{\mathbf{x}}, \tilde{\mathbf{x}}) = \frac{\left|\{i : \mathrm{sgn}(\mathbf{u}_i^T \bar{\mathbf{x}}) \neq \mathrm{sgn}(\mathbf{u}_i^T \tilde{\mathbf{x}})\}\right|}{m};$$

- ratio of missing support:

$$\mathrm{FNR}(\bar{\mathbf{x}}, \tilde{\mathbf{x}}) = \frac{|\mathrm{supp}(\bar{\mathbf{x}}) \backslash \mathrm{supp}(\tilde{\mathbf{x}})|}{|\mathrm{supp}(\bar{\mathbf{x}})|},$$

  where $\mathrm{supp}(\mathbf{x})$ stands for the support set of $\mathbf{x}$; in our numerical experiments, a component being non-zero means that its absolute value is larger than 10^{-3};

- ratio of misidentified support:

$$\mathrm{FPR}(\bar{\mathbf{x}}, \tilde{\mathbf{x}}) = \frac{|\mathrm{supp}(\tilde{\mathbf{x}}) \backslash \mathrm{supp}(\bar{\mathbf{x}})|}{n - |\mathrm{supp}(\bar{\mathbf{x}})|}.$$

To give multiple views for the considered nonconvex regularizations, we also report the performance measured by these criteria. The following figures are the average results of 100 trials and the sub-figures (a-b), (c-d), (e-f), (g-h) correspond to AE, INR, FNR, and FPR, respectively. The performance for different numbers of measurements is reported in Fig. 4. Both the ideal and the selected parameters are used. Similarly, the performance for different noise levels (Fig. 5) and different sparsity levels (Fig. 6) are displayed. Together with SNRs shown before, we can have clear impression for the three proposed algorithms:

- Alg. 2 significantly reduces the computational time comparing to that given by [12] for MCP.

- The efficiency of the proposed algorithms are comparable to the Passive algorithm [11], and the recovery quality is improved.

- Alg. 4 takes the least computational time. Alg. 2 and 3 have better performance when the parameters can be properly selected, otherwise, Alg. 4 is better.

5. Conclusion

Applying nonconvex regularizations is promising in enhancing the sparsity for 1bit-CS. The major obstacle are that minimizing nonconvex regularizations usually requires long computational time and it is difficult to find the global optimal solution. In this paper, we developed fast algorithms for several nonconvex regularizations based on its analytical solutions. Our results extended the previous discussion on analytical solutions, which were limited to several specific regularizations, and also we significantly improved the computational efficiency for some problems. The proposed algorithms of several nonconvex regularizations are evaluated on numerical experiments and the computational time is comparable to the convex Passive algorithm, the currently fastest method for ℓ_1 minimization of 1bit-CS. In the future, we will consider the nonconvex penalties in norm estimation [15], robust losses [41], and adaptive thresholding [42, 43, 10]. These techniques are currently restricted to convex penalties, i.e., the ℓ_1-norm minimization. It is promising to enhance the sparsity without introducing too much computational burden by applying the discussed analytical solutions.

Acknowledgment

This work was supported by National Science Foundation Grant DMS-1621798 and National Natural Science Foundation of China Grant 61603248.

The authors are grateful to the anonymous reviewers for their helpful comments.

References

[1] Anders Host-Madsen and Knud Andersen. Lower bounds for estimation of frequency and phase of doppler signals. *Measurement Science and Technology*, 6(6):637, 1995.

[2] Anders Host-Madsen and Peter Handel. Effects of sampling and quantization on single-tone frequency estimation. *IEEE Transactions on Signal Processing*, 48(3):650–662, 2000.

[3] Ofer Bar-Shalom and Anthony J. Weiss. DOA estimation using one-bit quantized measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):868–884, 2002.

[4] Hao Chen and Pramod K. Varshney. Performance limit for distributed estimation systems with identical one-bit quantizers. *IEEE Transactions on Signal Processing*, 58(1):466–471, 2010.

[5] Hao Chen and Pramod K. Varshney. Nonparametric one-bit quantizers for distributed estimation. *IEEE Transactions on Signal Processing*, 58(7):3777–3787, 2010.

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

[7] Jason N. Laska, Zaiwen Wen, Wotao Yin, and Richard G. Baraniuk. Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements. *IEEE Transactions on Signal Processing*, 59(11):5289–5301, 2011.

[8] Ming Yan, Yi Yang, and Stanley Osher. Robust 1-bit compressive sensing using adaptive outlier pursuit. *IEEE Transactions on Signal Processing*, 60(7):3868–3875, 2012.

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

[10] Richard Baraniuk, Simon Foucart, Deanna Needell, Yaniv Plan, and Mary Wootters. Exponential decay of reconstruction error from binary measurements of sparse signals. *IEEE Transactions on Information Theory*, 2017.

[11] Lijun Zhang, Jinfeng Yi, and Rong Jin. Efficient algorithms for robust one-bit compressive sensing. In *Proceedings of the 31st International Conference on Machine Learning*, pages 820–828, 2014.

[12] Rongda Zhu and Quanquan Gu. Towards a lower sample complexity for robust one-bit compressed sensing. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 739–747, 2015.

[13] Sheng Chen and Arindam Banerjee. One-bit compressed sensing with the k-support norm. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pages 138–146, 2015.

[14] Pranjal Awasthi, Maria-Florina Balcan, Nika Haghtalab, and Hongyang Zhang. Learning and 1-bit compressed sensing under asymmetric noise. In *Proceedings of the 29th Annual Conference on Learning Theory*, 2016.

[15] Karin Knudson, Rayan Saab, and Rachel Ward. One-bit compressive sensing with norm estimation. *IEEE Transactions on Information Theory*, 62(5):2748–2758, 2016.

[16] Petros T. Boufounos. Greedy sparse signal reconstruction from sign measurements. In *Proceedings of the 43rd Asilomar Conference on Signals, Systems, and Computers*, pages 1305–1309, 2009.

[17] Yingying Xu, Yoshiyuki Kabashima, and Lenka Zdeborová. Bayesian signal reconstruction for 1-bit compressed sensing. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(11):P11015, 2014.

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

[19] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[20] Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.

[21] Shuang Qiu, Tingjin Luo, Jieping Ye, and Ming Lin. Nonconvex one-bit single-label multi-label learning. *arXiv preprint arXiv:1703.06104*, 2017.

[22] Rick Chartrand and Valentina Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24(3):035020, 2008.

[23] Penghang Yin, Yifei Lou, Qi He, and Jack Xin. Minimization of ℓ_{1-2} for compressed sensing. *SIAM Journal on Scientific Computing*, 37(1):A536–A563, 2015.

[24] Yifei Lou and Ming Yan. Fast L1-L2 minimization via a proximal operator. *Journal on Scientific Computing*, to appear.

[25] Xiaolin Huang, Yipeng Liu, Lei Shi, Sabine Van Huffel, and Johan A.K. Suykens. Two-level ℓ_1 minimization for compressed sensing. *Signal Processing*, 108(1):459–475, 2015.

[26] Małgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel J. Candès. SLOPE—Adaptive variable selection via convex optimization. *The Annals of Applied Statistics*, 9(3):1103–1140, 2015.

[27] Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66(8):1275–1297, 2013.

[28] Sohail Bahmani, Petros T Boufounos, and Bhiksha Raj. Robust 1-bit compressive sensing via gradient support pursuit. *arXiv preprint arXiv:1304.6627*, 2013.

[29] Dao-Qing Dai, Lixin Shen, Yuesheng Xu, and Na Zhang. Noisy 1-bit compressive sensing: models and algorithms. *Applied and Computational Harmonic Analysis*, 40(1):1–32, 2014.

[30] Mokhtar S. Bazaraa, Hanif D. Sherali, and Chitharanjan. M. Shetty. *Nonlinear programming: Theory and algorithms*. Wiley-Interscience, Hoboken, N.J, 3rd ed edition, 2006.

[31] Xiaolin Huang, Lei Shi, and Ming Yan. Nonconvex sorted ℓ_1 minimization for sparse approximation. *Journal of the Operations Research Society of China*, 3(2):207–229, 2015.

[32] Xiangrong Zeng and Mario A.T. Figueiredo. Decreasing weighted sorted ℓ_1 regularization. *IEEE Signal Processing Letters*, 21(10):1240–1244, 2014.

[33] Chengmin Yang. One-sided ℓ_1 norm and best approximation in one-sided ℓ_1 norm. *Numerical Functional Analysis and Optimization*, 28(3-4):503–518, 2007.

[34] Ralph Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1997.

[35] Ralph Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*. Fundamental Principles of Mathematical Sciences. Berlin: Springer-Verlag, 2005.

[36] Xiaolin Huang, Lei Shi, Ming Yan, and Johan A. K. Suykens. Pinball loss minimization for one-bit compressive sensing. *arXiv preprint arXiv:1505.03898*, 2015.

[37] Ankit Parekh and Ivan W. Selesnick. Enhanced low-rank matrix approximation. *IEEE Signal Processing Letters*, 23(4):493–497, 2016.

[38] Ivan Selesnick. Sparse regularization via convex analysis. *IEEE Transactions on Signal Processing*, 65(17):4481–4494, 2017.

[39] Zongben Xu, Xiangyu Chang, Fengmin Xu, and Hai Zhang. $l_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Transactions on Neural Networks and Learning Systems*, 23(7):1013–1027, 2012.

[40] Zhaosong Lu and Xiaorui Li. Sparse recovery via partial regularization: Models, theory and algorithms. *arXiv preprint arXiv:1511.07293*, 2015.

[41] Song Mei, Yu Bai, and Andrea Montanari. The landscape of empirical risk for non-convex losses. *arXiv preprint arXiv:1607.06534*, 2016.

[42] Cristian Rusu, Roi Mendez-Rial, Nuria Gonzalez-Prelcic, and Robert W. Heath. Adaptive one-bit compressive sensing with application to low-precision receivers at mmwave. In *Proceedings of IEEE Global Communications Conference*, pages 1–6. IEEE, 2015.

[43] Jun Fang, Yanning Shen, Linxiao Yang, and Hongbin Li. Adaptive one-bit quantization for compressed sensing. *Signal Processing*, 125:145–155, 2016.

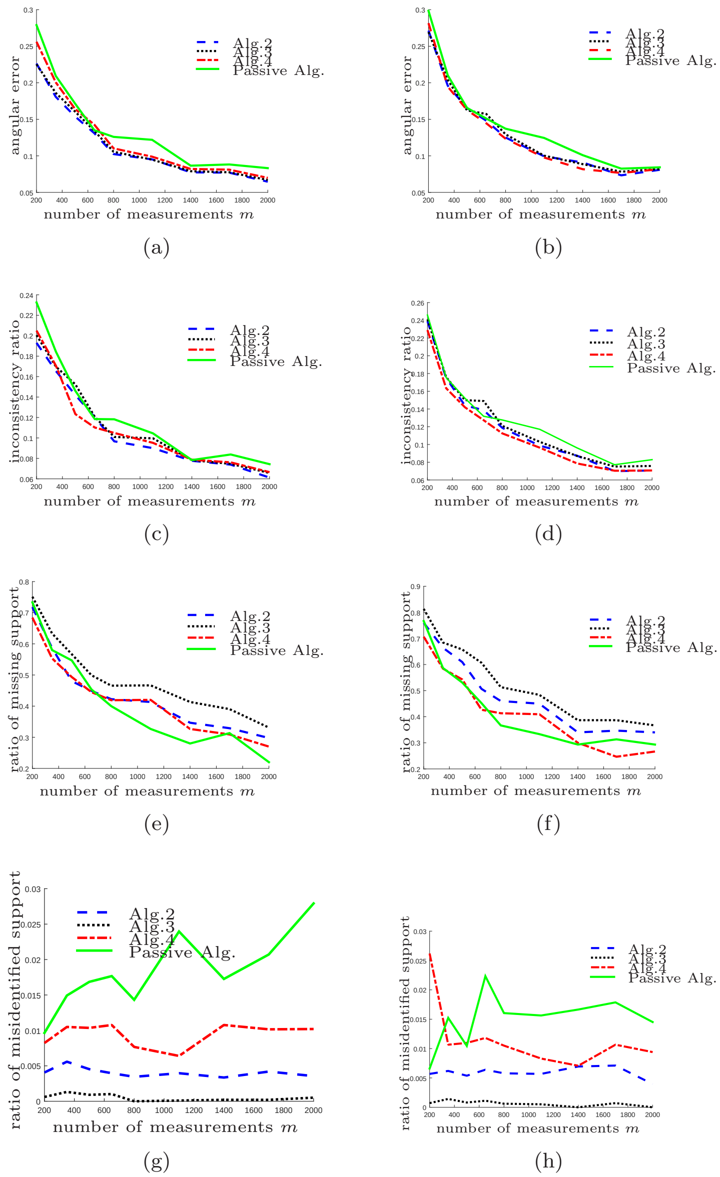
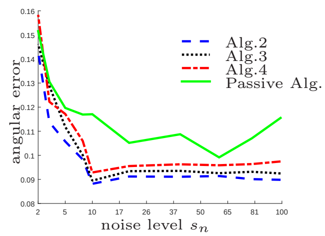
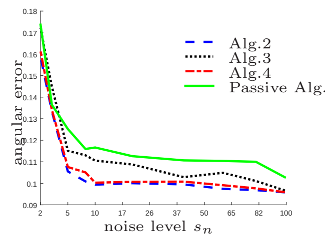


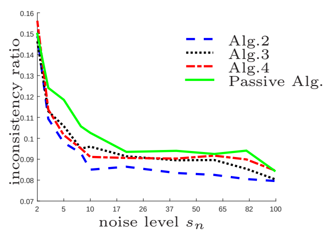
Figure 4: Recovery performance for different numbers of measurements. Three nonconvex regularizations are evaluated: MCP minimization (blue dashed line), ℓ_0 minimization (black dotted line), sorted ℓ_1 penalty (red dash-dotted line), and ℓ_1 minimization (green solid line). In this experiment $n = 1000, K = 15, s_n = 10$, and sign flip ratio is 10%. Left column: use ideal parameters; Right column: use parameters selected by cross-validation. (a-b) angular error; (c-d) inconsistency ratio; (e-f) ratio of missing support; (g-h) ratio of misidentified support.



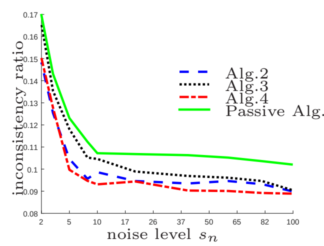
(a)



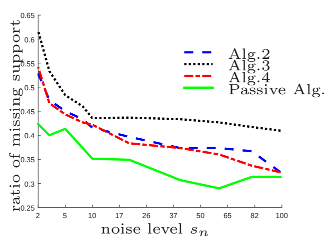
(b)



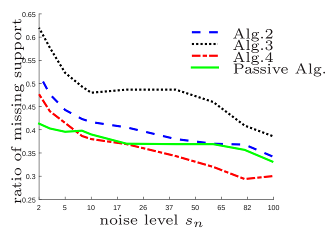
(c)



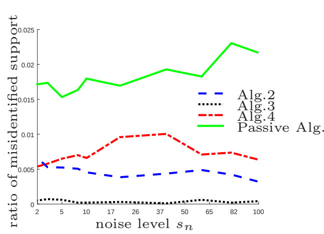
(d)



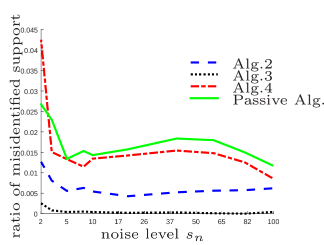
(e)



(f)



(g)



(h)

Figure 5: Recovery performance for different noise levels. Three nonconvex regularizations are evaluated: MCP minimization (blue dashed line), ℓ_0 minimization (black dotted line), sorted ℓ_1 penalty (red dash-dotted line), and ℓ_1 minimization (green solid line). In this experiment $n = 1000, K = 15, m = 1000$, and sign flip ratio is 10%. Left column: use ideal parameters; Right column: use parameters selected by cross-validation. (a-b) angular error; (c-d) inconsistency ratio; (e-f) ratio of missing support; (g-h) ratio of misidentified support.

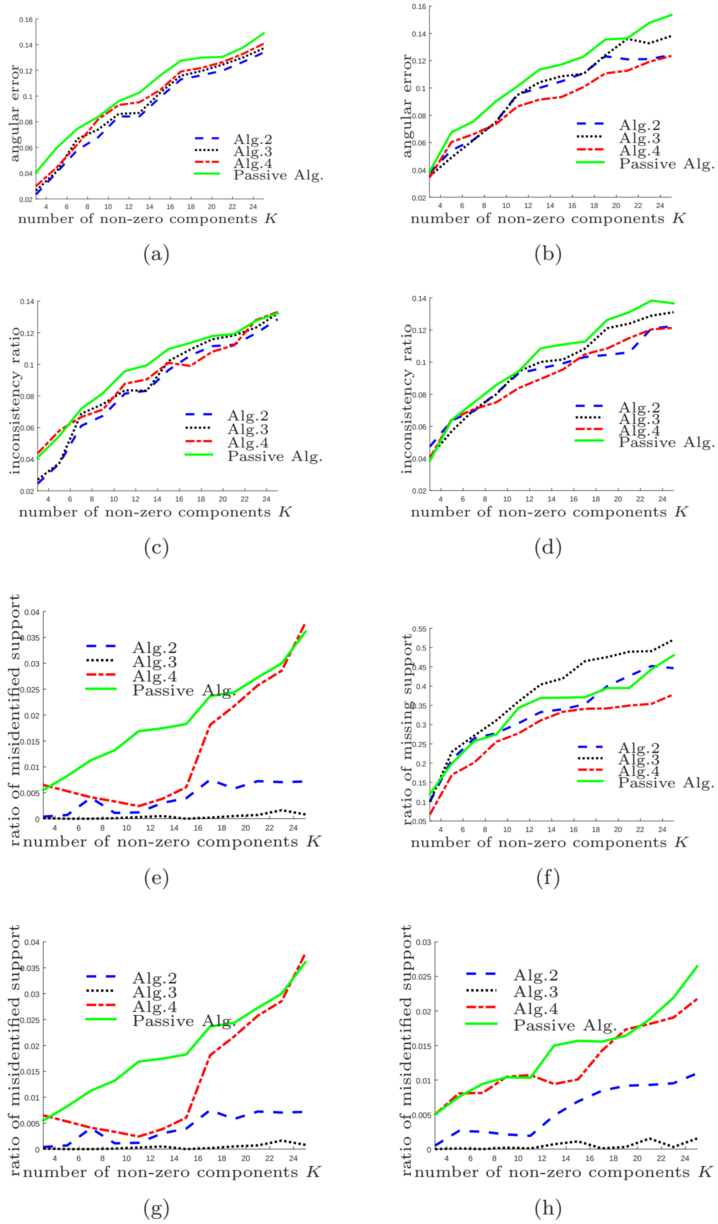


Figure 6: Recovery performance for different sparsity levels. Three nonconvex regularizations are evaluated: MCP minimization (blue dashed line), ℓ_0 minimization (black dotted line), sorted ℓ_1 penalty (red dash-dotted line), and ℓ_1 minimization (green solid line). In this experiment $n = 1000, m = 1000, s_n = 10$, and sign flip ratio is 10%. Left column: use ideal parameters; Right column: use parameters selected by cross-validation. (a-b) angular error; (c-d) inconsistency ratio; (e-f) ratio of missing support; (g-h) ratio of misidentified support.