Non-convex Penalties with Analytical Solutions for One-bit Compressive Sensing

Xiaolin Huang, Ming Yan

Abstract—One-bit measurements widely exist in the real world, and they can be used to recover sparse signals. This task is known as the problem of learning halfspaces in learning theory and one-bit compressive sensing (1bit-CS) in signal processing. In this paper, we propose novel algorithms based on both convex and nonconvex sparsity-inducing penalties for robust 1bit-CS. We provide a sufficient condition to verify whether a solution is globally optimal or not. Then we show that the globally optimal solution for positive homogeneous penalties can be obtained in two steps: a proximal operator and a normalization step. For several nonconvex penalties, including minimax concave penalty (MCP), ℓ0 norm, and sorted ℓ1 penalty, we provide fast algorithms for finding the analytical solutions by solving the dual problem. Specifically, our algorithm is more than 200 times faster than the existing algorithm for MCP. Its efficiency is comparable to the algorithm for the ℓ₁ penalty in time, while its performance is much better. Among these penalties, the sorted ℓ₁ penalty is most robust to noise in different settings.

Index Terms—optimization, one-bit compressive sensing, non-convex penalty, analytical solutions

I. INTRODUCTION

ANALOG-to-digital converting (ADC) is a necessary process in digital processing, and the choice of the bit-depth in ADC is an important issue. The extreme case is to use one-bit measurements, which enjoys many advantages, e.g., it 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 x ∈ ℝⁿ (∥x∥₀ ≤ K) from m one-bit quantized measurements

\[ y_i = \text{sgn}(u_i^\top x + \varepsilon_i), \]

where uᵢ ∈ ℝⁿ is the i-th sensing vector, εᵢ is the noise in the measurement, and sgn returns 1 for a positive number and −1 otherwise. The sensing system is represented by \( U = [u_1, u_2, \ldots, u_m] \text{ and } Y = [y_1, y_2, \ldots, y_m]^\top. \)

Since its proposal by [1], 1bit-CS has attracted much attention in both the signal processing society ([2], [3], [4], [5]) and the machine learning society ([6], [7], [8], [9]). Since the one-bit information has no capability to specify the magnitude of the original signal, we assume \( \|x\|_2 = 1 \) without loss of generality (there is also some work on norm estimation, see, e.g., [10]), and 1bit-CS can be explained as finding the sparsest vector on the unit sphere that coincides with the observed signs, i.e.,

\[
\begin{align*}
\text{minimize} & \quad \|x\|_0, \\
\text{subject to} & \quad y_i = \text{sgn}(u_i^\top x), \quad \forall i = 1, 2, \ldots, m, \\
& \quad \|x\|_2 = 1.
\end{align*}
\]

This is an NP-hard problem, and several algorithms are developed to approximately solve it or its variants, which can be found in [1], [2], [11], [12].

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 \( y_i = \text{sgn}(u_i^\top x) \) by a loss function. For example, the one-sided ℓ₁ loss and the one-sided ℓ₂ loss are considered in [13] and [3]; the linear loss is used in [4] and [6]. It is reported that the linear loss generally outperforms the one-sided ℓ₁/ℓ₂ 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 [7], which gives analytical solutions for two nonconvex penalties, namely the Smoothly Clipped Absolute Deviation (SCAD, [13]) and Minimax Concave Penalty (MCP, [15]). Also, [8] 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 [7, 8] and, therefore, have been extended to other applications including the multi-label learning task [16].

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, ℓᵰ norm (0 ≤ p ≤ +∞, [17]), ℓ₁-ℓ₂ norm [18], sorted ℓ₁ penalty [19], [20], and so on. The contributions of this paper can be summarized as follows:

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

- For positive homogeneous penalties, we show that the optimal solution can be obtained by two steps: a proximal operator and a normalization step. For a general penalty, we provide a generic algorithm by solving the dual problem.
We provide algorithms for finding analytical solutions for several nonconvex penalties including MCP, \( \ell_0 \) norm, and the sorted \( \ell_1 \) penalty. These algorithms are much faster than the existing 1bit-CS algorithms for nonconvex penalties and even comparable to that for the convex \( \ell_1 \) minimization problem, e.g., our algorithm is averagely 200 times faster than the algorithm given in [7] for MCP. In addition, we compare three nonconvex penalties, i.e., MCP, \( \ell_0 \) norm, and the sorted \( \ell_1 \) penalty, and show that the sorted \( \ell_1 \) performs the best in both performance and computational time.

The rest of this paper is organized as follows. Section II 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 III. The numerical experiments are reported in Section IV. We end this paper with a brief conclusion.

### II. RELATED WORKS

Model (2) for 1bit-CS has two main disadvantages: i) it is difficult to solve because of the \( \ell_0 \) norm in the objective and the constraint \( \|x\|_2 = 1 \); ii) the constraint \( y_i = \text{sgn}(u_i^\top x) \) does not consider noisy sign measurements.

Several approaches are given to deal with both disadvantages. For the nonconvexity, the \( \ell_0 \) norm is replaced by the \( \ell_1 \) norm, and the constraint \( \|x\|_2 = 1 \) is replaced by other convex constraints. The first convex model [21] for 1bit-CS is

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1, \\
\text{subject to} & \quad y_i(u_i^\top x) \geq 0, \ \forall i = 1, 2, \ldots, m, \\
& \quad \|U^\top x\|_1 = r,
\end{align*}
\]

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, [13] introduces the following robust model using the one-sided \( \ell_1 \) norm:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{m} \sum_{i=1}^m \max \{0, -y_i(u_i^\top x)\}, \\
\text{subject to} & \quad \|x\|_2 = 1, \\
& \quad \|x\|_0 = K.
\end{align*}
\]

The robust model using the one-sided \( \ell_2 \) norm is also introduced. Several modifications are designed by [3], [22], and [23] to improve their robustness to sign flips and noise.

The linear loss for robust 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 [4], [6], [7], [8]. In [4], the first model using the linear loss for 1bit-CS is proposed and takes the following formulation,

\[
\begin{align*}
\text{minimize} & \quad -\frac{1}{m} \sum_{i=1}^m y_i(u_i^\top x), \\
\text{subject to} & \quad \|x\|_2 \leq 1, \\
& \quad \|x\|_1 \leq s,
\end{align*}
\]

where \( s \) is a given positive constant. One can also put the \( \ell_1 \)-norm in the objective instead of in the constraint, resulting in the problem given by [6]:

\[
\begin{align*}
\text{minimize} & \quad \mu \|x\|_1 - \frac{1}{m} \sum_{i=1}^m y_i(u_i^\top x), \\
\text{subject to} & \quad \|x\|_2 \leq 1,
\end{align*}
\]

where \( \mu \) is the regularization parameter for the \( \ell_1 \)-norm. Note that the unit ball constraint \( \|x\|_2 = 1 \) is relaxed to the unit ball constraint \( \|x\|_2 \leq 1 \) in (4) and (5). As illustrated by [6], 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 \( \mu \). The convex penalty in (5) is replaced by several nonconvex penalties such as MCP [7] and \( k \)-support norm [8]. Better sampling complexities can be achieved for these nonconvex penalties and analytical solutions are obtained.

In this paper, we consider general penalties in (3) 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.

### III. ANALYTICAL SOLUTIONS FOR 1BIT-CS

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

\[
\begin{align*}
\text{minimize} & \quad f(x) - \frac{1}{m} \sum_{i=1}^m y_i(u_i, x) + \frac{\tau}{2} \|x\|_2^2, \\
\text{subject to} & \quad \|x\|_2^2 \leq 1,
\end{align*}
\]

where \( \tau \geq 0 \) and \( f(x) \) is the penalty. Most existing papers assume that \( \tau = 0 \), and the choice of \( \tau > 0 \) is introduced by [7]. We will show that there is no need to choose a positive \( \tau \) because the optimal solution does not depend on \( \tau \) when \( \tau \) is small enough and the optimal solution is not on the unit sphere for a large \( \tau \). Let \( v = \frac{1}{m} \sum_{i=1}^m y_i u_i \) and

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

We make the following assumption for \( f(x) \).

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

The convexity of (6) depends on the penalty \( f(x) \) and \( \tau \). For a convex \( f(x) \), problem (6) is convex, and it could also be nonconvex when \( f(x) \) is nonconvex. In order to find its global solution, we solve the corresponding dual problem and verify
that the duality gap is zero, i.e., the optimal primal value is the same as the optimal dual value. Define the corresponding Lagrangian functional as

\[ \mathcal{L}(x, \mu) = F(x) + \frac{\mu}{2} (\|x\|^2 - 1), \]

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

**Lemma 3.1:** [24] Theorem 6.2.5] If there exist \((x^*, \mu^*)\) such that \(\|x^*\|^2 \leq 1\), \(\mu^* \geq 0\), \(\mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(x, \mu)\) for all \(x\), and \(\mu^*(\|x^*\|^2 - 1) = 0\), then \(x^*\) and \(\mu^*\) are optimal solutions to the primal and dual problems, respectively, with no duality gap.

**Proof:** Because \(\mu^*(\|x^*\|^2 - 1) = 0\) and \(\|x^*\|^2 \leq 1\), we have, for any \(\mu \geq 0\),

\[ \mathcal{L}(x^*, \mu) = F(x^*) + \frac{\mu}{2} (\|x^*\|^2 - 1) \leq F(x^*), \]

Therefore, \((x^*, \mu^*)\) is a saddle point of \(\mathcal{L}(x, \mu)\), i.e.,

\[ \mathcal{L}(x^*, \mu) = \mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(x, \mu^*), \tag{7} \]

for any \(\mu \geq 0\) and \(x\). Thus,

\[ F(x^*) = \mathcal{L}(x^*, \mu^*) = \min_x \mathcal{L}(x, \mu^*). \]

The duality gap is zero, and \(x^*\) and \(\mu^*\) 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 \(x^*\) and verify whether \(\mu^*(\|x^*\|^2 - 1) = 0\) is satisfied. If it is satisfied, then \(x^*\) is a global optimal solution of (6).

**A. Positive Homogeneous Penalties**

Assume that \(f(\alpha x) = \alpha f(x)\) for any positive \(\alpha\), i.e., \(f(x)\) is positive homogeneous. We can obtain the global solution in two steps: a proximal operator and a normalization step. Some examples of positive homogeneous penalties are:

- \(\ell_p\) norm \((0 < p \leq +\infty)\): e.g., \(\ell_1\) norm \([6]\);
- \(\ell_p\) norm minus \(\ell_q\) norm: e.g., \(\ell_1 - \ell_2\) norm \([18]\);
- 0 function: Lemma 4.1 from paper \([7]\);
- sorted \(\ell_1\) penalty: nonconvex ones \([23]\); convex ones \([20]\);
- indicator function of \(\ell_0\) \([8]\); small magnitude penalized (SMAP) \([26]\);
- One-sided norm \([27]\);
- Gauges \([28]\).

Define the proximal operator of \(f\) as

\[ \text{Prox}_f(v) := \arg \min_x f(x) + \frac{1}{2}\|x - v\|^2, \]

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

**Lemma 3.2:** If \(f(x)\) is positive homogeneous, we have that

\[ f(t^*) - \langle t^*, v \rangle = -\|t^*\|^2. \]

**Proof:** When \(t^* = 0\), we have \(f(0) = 0\) because of \(f(x)\) being positive homogeneous, and the result is trivial.

When \(t^* \neq 0\), we have \(f(t^*) = \langle \nabla f(t^*), t^* \rangle\), where \(\nabla f(t^*)\) is a generalized subgradient of \(f\) at \(t^*\) \([29]\) Definition 8.3]. Therefore,

\[ f(t^*) - \langle t^*, v \rangle = \langle \nabla f(t^*), t^* \rangle - \langle t^*, v \rangle \]

\[ = \langle \nabla f(t^*) - v, t^* \rangle = -\|t^*\|^2. \]

The last equality is satisfied because \(t^* \in \text{Prox}_f(v)\).

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

\[ x^* = \begin{cases} t^*/\tau & \text{if } \|t^*\|_2 \leq \tau, \\ \frac{f(t^*)}{\|t^*\|^2} & \text{if } \|t^*\|_2 > \tau. \end{cases} \]

**Proof:** When \(\tau > 0\), we have

\[ \mathcal{L}(x, \mu) = f(x) + \frac{\tau + \mu}{2} \|x - v/\tau + \mu\|_2^2 - \frac{\|v\|^2_2}{2(\tau + \mu)} - \frac{\mu}{2}. \]

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

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

The last equality comes from Lemma \([3.2]\) Thus the dual problem is a concave function of \(\mu\) and we can find \(\mu^*\) as

\[ \mu^* = \begin{cases} 0 & \text{if } \|t^*\|_2 \leq \tau, \\ \|t^*\|_2^2 - \tau & \text{if } \|t^*\|_2 > \tau. \end{cases} \tag{8} \]

Therefore we have

\[ x^* = \begin{cases} t^*/\tau & \text{if } \|t^*\|_2 \leq \tau, \\ \frac{t^*}{\|t^*\|^2} & \text{if } \|t^*\|_2 > \tau. \end{cases} \]

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

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

\[ \min_x \mathcal{L}(x, \mu) = -\|t^*\|^2_2 - \frac{\mu}{2} < 0. \]

Then we consider the case when \(\mu = 0\). From Lemma \([3.2]\) we have

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

Therefore, the positive homogeneity of \(f\) gives

\[ \min_x \mathcal{L}(x, 0) = -\infty \text{ if } t^* \neq 0. \]

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

\[ \frac{1}{2}\|v\|^2_2 \leq f(x) + \frac{1}{2}\|x - v\|^2_2, \]

which implies

\[ -\frac{1}{2}\|x\|^2 \leq f(x) - \langle x, v \rangle, \]

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

\[ \min_x \mathcal{L}(x, 0) = 0 \text{ if } t^* = 0. \]
Together, we have $\mu^* = \|t^*\|_2$ and $x^* = t^*/\mu = t^*/\|t^*\|_2$ if $t^* \neq 0$. When $t^* = 0$, we have $\mu^* = 0$ and $x^* = 0$.

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

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

is a global optimal solution of (9). In fact, there may be multiple global solutions when $t^* = 0$ (See 130 for examples).

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

**Remark 3.1:** When $\tau < \|t^*\|_2$, we have $x^* = t^*/\|t^*\|_2$, i.e., the optimal $x^*$ does not depend on $\tau$. When $\tau > \|t^*\|_2$, we have $x^* = t^*/\tau$, i.e., the optimal $x^*$ is not on the unit sphere. Therefore, there is no need to choose a positive $\tau$, and we let $\tau = 0$ in the numerical experiments.

**B. General Penalties**

For a general $f(x)$, we consider the dual function

$$G(\mu) = \min_x L(x, \mu). \quad (9)$$

Given $\mu$, let $x^*(\mu)$ be an optimal solution of (9) defined as

$$x^*(\mu) \in \arg \min_x L(x, \mu) = \arg \min_x f(x) + \frac{(\tau + \mu)}{2} \|x - \frac{v}{\tau + \mu}\|_2^2. \quad (10)$$

The following theorem provides the subdifferential of $G$.

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

$$G(\tilde{\mu}) \leq G(\mu) + \frac{1}{2} (\|x^*(\mu)\|^2 - 1)(\tilde{\mu} - \mu). \quad (11)$$

**Proof:** We have

$$G(\tilde{\mu}) = \min_x L(x, \mu) \leq L(x^*(\mu), \mu) \leq L(G(\mu), \mu)$$

$$= L(G(\mu), \mu) + \frac{1}{2} (\|x^*(\mu)\|^2 - 1)(\tilde{\mu} - \mu)$$

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

The last equality is valid because of (9).

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

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

and the optimal $\mu^* = 1/8$. The optimal $x^*$'s for $\mu^*$ 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 $\tau$ is small enough such that $\|x^*(0)\|_2 > 1$ for all $x^*(0)$, the optimal $x^*$ does not depend on $\tau$. When $\tau$ is large enough such that $\|x^*(0)\|_2 < 1$ for some $x^*(0)$, then the optimal $x^*$ is not on the unit sphere. Therefore, there is no need to choose a positive $\tau$, 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 an global optimal solution of (6) efficiently, we have to make sure that the proximal operator has an analytical solution, because the proximal operator was evaluated for multiple times. Some penalties that have analytical solutions are:

- MCP [15],
- SCAD [14],
- $\ell_0$ norm,
- $\ell_{1/2}$ regularization [31],
- Partial regularization [32].

In the following subsections, we describe algorithms for MCP, $\ell_0$ norm, and the nonconvex sorted $\ell_1$. Our algorithm is different from that in [7] for MCP. We let $\tau = 0$.

**C. Minimax Concave Penalty**

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

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

for fixed parameters $\lambda > 0$ and $b > 0$. The analytical solutions for (10) can be obtained.

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

$$x^*(\mu) = \begin{cases} 0, & \text{if } v^2 \leq b\lambda^2\mu, \\ \frac{v}{\mu}, & \text{if } v^2 \geq b\lambda^2\mu, \end{cases} \quad (12)$$

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

$$x^*(\mu) = \begin{cases} 0, & \text{if } |v| \leq \lambda, \\ \frac{|v| - \lambda}{\mu - \tau} \text{sgn}(v), & \text{if } \lambda < |v| \leq b\lambda\mu, \\ \frac{v}{\mu}, & \text{if } |v| \geq b\lambda\mu. \end{cases} \quad (13)$$

For some $\mu$, we have two optimal solutions, as shown in the formulation. The resulting algorithm is shown in Alg. 1.

**D. $\ell_0$ norm**

Let $f(x) = \lambda \|x\|_0$, and the analytical solutions for (10) is:

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

The resulting algorithm is shown in Alg. 2.
Algorithm 1: MCP

Input: $\lambda, b$
Output: $\mu$
Initialize: $\mu = 1/b$
$v[1], v[2], \ldots, v[n] = \text{Sort}(|v_1|, |v_2|, \ldots, |v_n|)$
Find $L$ such that $v[L] \leq \lambda < v[L+1]$
$d_2 = \sum_{i=L+1}^{n} v[i]$
$d_{max} = b^2d_2$
if $d_{max} > 1$ then
  $i = L + 1; d_1 = 0$
  while $d_{max} > 1$ do
    $\mu = v[i]/(b\lambda); d_{max} = d_1/(\mu - 1/b)^2 + d_2/\mu^2$
    if $d_{max} < 1$ then
      Solve $d_1/(\mu - 1/b)^2 + d_2/\mu^2 = 1$; return
    end if
    $d_1 = d_1 + (v[i] - \lambda)^2; d_2 = d_2 - v[i]$
    $d_{max} = d_1/(\mu - 1/b)^2 + d_2/\mu^2$
    $i = i + 1$
  end while
else
  $i = L$
  while $d_{max} < 1$ do
    $\mu = v[i]^{2}/(b\lambda); d_{max} = d_2/\mu^2$
    if $d_{max} > 1$ then
      $\mu = \sqrt{d_2};$ return
    end if
    $d_2 = d_2 + v[i]^2$
    $d_{max} = d_2/\mu^2$
    $i = i - 1$
  end while
end if

Algorithm 2: $\ell_0$ norm

Input: $\lambda$
Output: $\mu$
Initialize: $i = n$
$v[1], v[2], \ldots, v[n] = \text{Sort}(|v_1|, |v_2|, \ldots, |v_n|)$
$\mu = v[i]^{2}/(2\lambda); d = v[i]^2; d_{max} = d/\mu^2$
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
  end if
  $d = d + v[i]^2; d_{max} = d/\mu^2$
end while

E. Sorted $\ell_1$ Penalty

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

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

is sorted by the absolute component values. Since weight $w_i$’s are assigned according to the sort, this regularization is called sorted $\ell_1$ penalty. When $w_1 \leq w_2 \leq \ldots \leq w_n$, it is convex [20]. Otherwise, it is nonconvex and can be used to enhance sparsity [25]. For the nonconvex case, a typical weight setting is

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

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. The optimal solution can be analytically given by

$$t^*_i = \max \{|v[i]| - w_i\lambda, 0\} \text{sgn}(v[i]).$$

Because this penalty is positive homogeneous, we apply the proximal operator first and then a normalization step. The corresponding algorithm is given in Alg. 3.

Algorithm 3: Sorted $\ell_1$ Penalty

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

IV. 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 $\ell_2$-norm ball. Then, $m$ sign observations are generated by (1), where $\varepsilon$ is the Gaussian noise with noise level $s_n$, which stands for the ratio between the variances of the measurements and $\varepsilon$. 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. 1 and the algorithm in [7]. Both algorithms solve the same problem with the MCP penalty. By the concavity of the dual function and Theorem 5.2, the dual function is piecewise smooth and its subgradient is decreasing. So we can find the optimal $\mu^*$ or the interval that contains the optimal $\mu^*$, and we find $x^*$ from (12) or (13). 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 [7], 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 computational times over 100 trials are reported in Table 1 where the computational time for the $\ell_1$ minimization by the Passive algorithm [6] is given as well.
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 $\ell_1$ minimization problem, Alg. 1 which solves the MCP regularized problem, takes a longer time but the recovery quality can be improved, as theoretically analyzed and numerically confirmed by [7].

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

- Alg. 1 for minimizing the MCP penalty;
- Alg. 2 for $\ell_0$ minimization;
- Alg. 3 for the nonconvex sorted $\ell_1$ penalty.

With properly selected parameters, nonconvex regularizations improve the recovery quality from the $\ell_1$ norm. To observe the improvements, we set $n = 10000$, $m = 5000$, $K = 15$, $s_n = 10$, and use the proposed algorithms to recover the signal. Suppose $\bar{x}$ is the real signal and $\tilde{x}$ is the recovered one. Then the signal-to-noise ratio in dB, i.e.,

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

is used to measure the quality of the recovered signal.

We first use the “ideal” parameters for each method to show the best performance of each algorithm. Here, the best parameters are chosen based on the $\ell_2$ distance to the real signal. We repeat the process for 100 trials and display the box plots in Fig. 1, which illustrates the improvement from the Passive algorithm. The computational time for nonconvex penalties is comparable to the Passive algorithm.

In practice, it is difficult to find the best parameters. Instead, we can tune the parameters by cross-validation based on consistency. But the selected parameters are not the best for specific data, especially when there are only a few observations. In the following, we reduce the problem size to $n = 1000$, $m = 1000$ and use 10-fold cross-validation to tune the parameters. The corresponding SNRs are compared in Fig. 2. As shown in the figure, $\ell_0$ minimization may give good results. But it is not stable and may also give some bad results. By contrast, Alg. 3 is more stable. Note that the experiments here only provide some hints, it is still not easy to predict which algorithm is the best for a specific problem. For the computational time, Alg. 3 is the most efficient among the three nonconvex regularization algorithms, since it only additionally requires a sort operation based on the Passive algorithm.

### Table 1: Average Computational Time

| $m$   | $n$   | Passive | Zhu’s Alg. | Alg. 1 |
|-------|-------|---------|------------|--------|
| 500   | 1000  | 3.6 ms  | 1.51 s     | 4.0 ms |
| 1000  | 1000  | 6.7 ms  | 1.76 s     | 8.9 ms |
| 1000  | 2000  | 14.8 ms | 6.95 s     | 18.5 ms|
| 2000  | 2000  | 25.1 ms | 7.15 s     | 30.2 ms|
| 5000  | 5000  | 148 ms  | 43.6 s     | 184 ms |

### Fig. 1
The recovered SNR with the ideal parameters for the case $n = 10000$, $m = 5000$, $K = 15$, $s_n = 10$, and sign flip ratio 10%. The average computational times are: 284ms (Passive), 301ms (Alg. 1), 295ms (Alg. 2), 291ms (Alg. 3).

### Fig. 2
The recovered SNRs with 10-fold cross-validation for the case $n = 1000$, $m = 1000$, $K = 15$, $s_n = 10$, and sign flip ratio 10%. The average computational times are: 6.7ms (Passive), 8.9ms (Alg. 1), 8.2ms (Alg. 2), and 6.8ms (Alg. 3).

Since the comparison between MCP and the $\ell_1$ minimization has been fully considered by [7], we, in the following, focus on the three nonconvex regularizations. We will vary the number of measurements $m$ from 300 to 2000 and report the recovered quality in Fig. 3. Same as in the previous experiments, the performance for both ideal parameters and parameters selected by cross-validation are considered. The comparison in Fig. 3 indicates that if the optimal parameters can be obtained, then both MCP and $\ell_0$ achieve high recovery quality. The SNRs obtained by the sorted $\ell_1$ minimization is a bit worse in this case. But since Alg. 3 takes less time than the other two algorithms, it is still a promising choice.

When we select parameters by cross-validation, Alg. 3 performs better than Alg. 1 and 2, which coincides with our observations in Fig. 1 and Fig. 2. The comparison indicates that the sorted $\ell_1$ is more stable to different parameters.

Similar observations can be found in Fig. 4, 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.
we select \( n \) tuning. Though that value is not optimal, the performance of \( n \) mini-
imization (black dotted line), and sorted \( n \) penalty (red solid line). In this experi-
ment \( n = 1000, K = 15, s_n = 10, \) and sign flip ratio is 10\%. (a) use
the ideal parameters; (b) use parameters selected by 10-fold cross-validation.

Fig. 3. Recovery performance of nonconvex regularizations for different
numbers of measurements: MCP minimization (blue dashed line), \( \ell_0 \) mini-
mization (black dotted line), and sorted \( \ell_1 \) penalty (red solid line). In this
experiment \( n = 1000, K = 15, s_n = 10, \) and sign flip ratio is 10\%. (a) use
the ideal parameters; (b) use parameters selected by 10-fold cross-validation.

Last, we consider different numbers of non-zero compo-
nents \( K \) with \( n = 1000, m = 1000, \) and \( s_n = 10. \) For the
sorted \( \ell_1 \) penalty, there is one parameter \( n_1 \) in its weight (14)
that is related to the signal sparsity. In the previous experi-
ments 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 \( \ell_1 \) penalty is generally satisfying. In this experiment,
we select \( n_1 \) from \( \{2, 4, ..., 16\} \) for different \( K \)’s. Totally, there
are two parameters to tune for the sorted \( \ell_1 \) penalty, the same
as MCP. In Fig. 5(a) and 5(b) the average SNRs for ideal and
selected parameters are displayed, respectively.

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

• angular error:

\[
AE(\hat{x}, \tilde{x}) = \frac{1}{\pi} \arccos \left( \frac{\hat{x}^T \tilde{x}}{||\tilde{x}||_2} \right);
\]

• inconsistency ratio:

\[
INR(\hat{x}, \tilde{x}) = \frac{|\{i : \text{sgn}(u_i^T \tilde{x}) \neq \text{sgn}(u_i^T \hat{x})\}|}{m};
\]

• ratio of missing support:

\[
FNR(\hat{x}, \tilde{x}) = \frac{|\text{supp}(\hat{x}) \setminus \text{supp}(\tilde{x})|}{|\text{supp}(\hat{x})|},
\]

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

• ratio of misidentified support:

\[
FPR(\hat{x}, \tilde{x}) = \frac{|\text{supp}(\tilde{x}) \setminus \text{supp}(\hat{x})|}{n - |\text{supp}(\hat{x})|}.
\]

To give multi-view for the considered nonconvex regular-
izations, 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. 6. Both
the ideal and the selected parameters are used. Similarly, the
performance for different noise levels (Fig. 7) and different
sparsity levels (Fig. 8) are displayed. Together with SNRs
shown before, we can have clear impression for the three
proposed algorithms:

• Alg. 1 significantly reduces the computational time com-
paring to that given by [7] for MCP.

• The efficiency of the proposed algorithms are comparable
to the Passive algorithm [6], and the recovery quality is
improved.

• Alg. 3 takes the least computational time. Alg. 1 and 2
V. CONCLUSION

Applying nonconvex regularizations is promising in enhancing the sparsity for 1-bit 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 Passive algorithm, the currently fastest method for $\ell_1$ minimization of 1bit-CS. In the future, we will consider the nonconvex penalties in norm estimation [10], robust losses [33], and adaptive thresholding [34, 35, 5]. These techniques are currently restricted to convex penalties, i.e., the $\ell_1$-norm minimization. It is promising to enhance the sparsity without introducing too much computational burden by applying the discussed analytical solutions.

REFERENCES

[1] P. T. Boufounos and R. G. Baraniuk, “1-bit compressive sensing,” in Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on, 2008, pp. 16–21.

[2] 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.
Inconsistency ratio

| Noise level $s_n$ | Alg.1 | Alg.2 | Alg.3 |
|-------------------|-------|-------|-------|
| 0.005             | 0.00  | 0.00  | 0.00  |
| 0.010             | 0.01  | 0.01  | 0.01  |
| 0.015             | 0.02  | 0.02  | 0.02  |
| 0.020             | 0.03  | 0.03  | 0.03  |
| 0.025             | 0.04  | 0.04  | 0.04  |
| 0.030             | 0.05  | 0.05  | 0.05  |
| 0.035             | 0.06  | 0.06  | 0.06  |
| 0.040             | 0.07  | 0.07  | 0.07  |

Angular error

| Noise level $s_n$ | Alg.1 | Alg.2 | Alg.3 |
|-------------------|-------|-------|-------|
| 0.25              | 0.30  | 0.30  | 0.30  |
| 0.30              | 0.35  | 0.35  | 0.35  |
| 0.35              | 0.40  | 0.40  | 0.40  |
| 0.40              | 0.45  | 0.45  | 0.45  |
| 0.45              | 0.50  | 0.50  | 0.50  |
| 0.50              | 0.55  | 0.55  | 0.55  |
| 0.55              | 0.60  | 0.60  | 0.60  |
| 0.60              | 0.65  | 0.65  | 0.65  |
| 0.65              | 0.70  | 0.70  | 0.70  |

Ratio of misidentified support

| Noise level $s_n$ | Alg.1 | Alg.2 | Alg.3 |
|-------------------|-------|-------|-------|
| 0.00              | 0.00  | 0.00  | 0.00  |
| 0.01              | 0.01  | 0.01  | 0.01  |
| 0.02              | 0.02  | 0.02  | 0.02  |
| 0.03              | 0.03  | 0.03  | 0.03  |
| 0.04              | 0.04  | 0.04  | 0.04  |
| 0.05              | 0.05  | 0.05  | 0.05  |
| 0.06              | 0.06  | 0.06  | 0.06  |
| 0.07              | 0.07  | 0.07  | 0.07  |
| 0.08              | 0.08  | 0.08  | 0.08  |

Number of non-zero components

| Alg.1 | Alg.2 | Alg.3 |
|-------|-------|-------|
| 1     | 2     | 3     |
| 4     | 5     | 6     |
| 7     | 8     | 9     |
| 10    | 11    | 12    |
| 13    | 14    | 15    |
| 16    | 17    | 18    |
| 19    | 20    | 21    |
| 22    | 23    | 24    |

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

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

[5] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, “Exponential decay of reconstruction error from binary measurements of sparse signals,” IEEE Transactions on Information Theory, 2017.

[6] L. Zhang, J. Yi, and R. Jin, “Efficient algorithms for robust one-bit compressive sensing,” in Proceedings of the 31st International Conference on Machine Learning, 2014, pp. 820–828.

[7] R. Zhu and Q. Gu, “Towards a lower sample complexity for robust one-bit compressed sensing,” in Proceedings of The 32nd International Conference on Machine Learning, 2015, pp. 739–747.

[8] S. Chen and A. Banerjee, “One-bit compressed sensing with the k-support norm,” in Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, 2015, pp. 138–146.

[9] P. Awasthi, M.-F. Balcan, N. Haghhtalab, and H. Zhang, “Learning and 1-bit compressed sensing under asymmetric noise,” in Proceedings of the 29th Annual Conference on Learning Theory, 2016.

[10] K. Knudson, R. Saab, and R. Ward, “One-bit compressive sensing with norm estimation,” IEEE Transactions on Information Theory, vol. 62, no. 5, pp. 2748–2758, 2016.

[11] P. T. Boufounos, “Greedy sparse signal reconstruction from sign measurements,” in Proceedings of the 43rd Asilomar Conference on Signals,
[12] Y. Xu, Y. Kabashima, and L. Zdeborová, “Bayesian signal reconstruction for 1-bit compressed sensing,” Journal of Statistical Mechanics: Theory and Experiment, vol. 2014, no. 11, p. P11015, 2014.

[13] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, “Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors,” IEEE Transactions on Information Theory, vol. 59, no. 4, pp. 2082–2102, 2013.

[14] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,” Journal of the American Statistical Association, vol. 96, no. 456, pp. 1348–1360, 2001.

[15] C.-H. Zhang, “Nearly unbiased variable selection under minimax concave penalty,” The Annals of Statistics, vol. 38, no. 2, pp. 894–942, 2010.

[16] S. Qiu, T. Luo, J. Ye, and M. Lin, “Nonconvex one-bit single-label multi-label learning,” arXiv preprint arXiv:1703.06104, 2017.

[17] R. Chartrand and V. Staneva, “Restricted isometry properties and nonconvex compressive sensing,” Inverse Problems, vol. 24, no. 3, p. 035020, 2008.

[18] P. Yin, Y. Lou, Q. He, and J. Xin, “Minimization of $\ell_{1-2}$ for compressed sensing,” SIAM Journal on Scientific Computing, vol. 37, no. 1, pp. A536–A563, 2015.

[19] X. Huang, L. Shi, S. Van Huffel, and J. A. Suykens, “Two-level $\ell_1$ minimization for compressed sensing,” Signal Processing, vol. 108, no. 1, pp. 459–475, 2015.

[20] M. Bogdan, E. van den Berg, C. Sabatti, W. Su, and E. J. Candès, “SLOPE—Adaptive variable selection via convex optimization,” The Annals of Applied Statistics, vol. 9, no. 3, pp. 1103–1140, 2015.

[21] Y. Plan and R. Vershynin, “One-bit compressed sensing by linear programming,” Communications on Pure and Applied Mathematics, vol. 66, no. 8, pp. 1275–1297, 2013.

[22] D.-Q. Dai, L. Shen, Y. Xu, and N. Zhang, “Noisy 1-bit compressive sensing: models and algorithms,” Applied and Computational Harmonic Analysis, vol. 40, no. 1, pp. 1–32, 2014.

[23] R. T. Rockafellar, Convex Analysis. Princeton University Press, 1997.

[24] R. T. Rockafellar and R. J.-B. Wets, Variational Analysis, third corrected printing ed., ser. Fundamental Principles of Mathematical Sciences. Berlin: Springer-Verlag, 2005.

[25] X. Huang, L. Shi, M. Yan, and J. A. K. Suykens, “Pinball loss minimization for one-bit compressive sensing,” CoRR, vol. abs/1505.03898, 2015.

[26] Z. Lu and X. Li, “Sparse recovery via partial regularization: models, theory and algorithms,” ArXiv e-prints, Nov. 2015.