

Article

Data-Driven Method for Robust Recovery in 1-Bit Compressive Sensing with the Minimax Concave Penalty

Cui Jia ^{1,2}  and Li Zhu ^{1,*}

¹ School of Statistics and Data Science, Ningbo University of Technology, No. 201, Fenghua Road, Jiangbei District, Ningbo 315211, China; cja80@whu.edu.cn

² School of mathematics and Statistics, Wuhan University, Wuhan 430072, China

* Correspondence: lzhu-ts@nbut.edu.cn

Abstract: With the advent of large-scale data, the demand for information is increasing, which makes signal sampling technology and digital processing methods particularly important. The utilization of 1-bit compressive sensing in sparse recovery has garnered significant attention due to its cost-effectiveness in hardware implementation and storage. In this paper, we first leverage the minimax concave penalty equipped with the least squares to recover a high-dimensional true signal $\mathbf{x} \in \mathbb{R}^p$ with k -sparse from n -dimensional 1-bit measurements and discuss the regularization by combing the nonconvex sparsity-inducing penalties. Moreover, we give an analysis of the complexity of the method with minimax concave penalty in certain conditions and derive the general theory for the model equipped with the family of sparsity-inducing nonconvex functions. Then, our approach employs a data-driven Newton-type method with stagewise steps to solve the proposed method. Numerical experiments on the synthesized and real data verify the competitiveness of the proposed method.

Keywords: 1-bit compressive sensing; sparsity-inducing nonconvex function; data driven

MSC: 65J20; 47J06; 65F22



Citation: Jia, C.; Zhu, L. Data-Driven Method for Robust Recovery in 1-Bit Compressive Sensing with the Minimax Concave Penalty. *Mathematics* **2024**, *12*, 2168. <https://doi.org/10.3390/math12142168>

Academic Editor: Ke-Lin Du

Received: 12 June 2024

Revised: 6 July 2024

Accepted: 8 July 2024

Published: 10 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the advent of big data, people's demand for signal processing is increasing, which drives the need for simpler signal processing technology. In the past decade or two, compressive sensing (CS), as a high-efficiency tool in signal processing, has garnered a lot of interest. However, classical CS faces substantial computing challenges when dealing with larger-dimensional signals due to storage bottlenecks, high transmission bandwidth requirements, and algorithmic inefficiencies. To mitigate these computing costs, it is often effective to quantize measurements by using a finite number of bits. A quantized signal collection method for high-dimensional sparse data processing and storage is 1-bit compressive sensing [1], which transforms the measurements to a single bit, i.e., $\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x})$ with $\mathbf{y} \in \mathbb{R}^n$, "sign" captures the sign of $\mathbf{A}\mathbf{x}^l \in \mathbb{R}^p$, $d^l \in \mathbb{R}^p$, and the element y_i is 1 or -1 , $y_i \in \{-1, 1\}$, $\mathbf{A} \in \mathbb{R}^{n \times p}$, $\mathbf{x} \in \mathbb{R}^p$, n and p are positive integers. Such 1-bit compressive sensing lowers transmission costs, resulting in a significant decrease for hardware implementation. A 1-bit quantizer only needs the sign of the measurements, and thus, is admissible to nonlinear distortion. Furthermore, 1-bit CS can be used effectively in computation due to the binary measurements. The 1-bit CS is suitable for various fields including wireless sensing network and radar analysis [2,3], and in domains like machine learning research, scientific studies, and industrial application. Researchers need to deal with quantized data for producing portable devices, which are necessary in modern life.

In their pioneering work [1], Boufounos and Baraniuk firstly took into account 1-bit noiseless measurements and studied the ℓ_1 -norm optimization problem to recover \mathbf{x} with two constraints $\mathbf{y} \odot \mathbf{A}\mathbf{x} \geq 0$ and $\|\mathbf{x}\|_2 = 1$, where the sign " \geq " represents the element-wise

inequality and the Hadamard product is represented by \odot . Jacques et al. [4] considered ℓ_0 -norm instead of ℓ_1 -norm to enhance sparsity, and established the binary iterative hard thresholding (BIHT or BIHT- ℓ_2) algorithm for 1-bit noisy measurements $\mathbf{y} = \text{sign}(A\mathbf{x} + \mathbf{b})$, $\mathbf{b} \in \mathbb{R}^n$. Matsumoto et al. [5] studied a noisy version with iterative hard thresholding, where a fraction of the measurements can be flipped. Yan et al. [6] considered sign flips in 1-bit measurements by using an adaptive outlier pursuit (AOP) algorithm that was a modified version of BIHT. Plan et al. [7] discussed a combination between the linear loss $(\mathbf{y}, A\mathbf{x})$ and ℓ_1 -constraint, just as $\min_{\mathbf{x}} (\mathbf{y}, A\mathbf{x})/n, s.t. \|\mathbf{x}\|_1 \leq \sqrt{k}$ and $\|\mathbf{x}\|_2 \leq 1$. Ref. [8] studied the Lagrange optimization of the linear loss and ℓ_1 -norm and attained the analysis solution $\min_{\|\mathbf{x}\|_2 \leq 1} (\mathbf{y}, A\mathbf{x})/n + \lambda \|\mathbf{x}\|_1$. Dai et al. [9] considered the one-sided ℓ_0 objective model for the noisy 1-bit CS model. Subsequently, Ref. [10] introduced the pitball loss (it involves both ℓ_1 -norm and ℓ_2 -norms), and proposed the PBAOP algorithm to decode from binary noisy data. Aiming at solutions with better sparsity, Ref. [11] introduced the nonconvex minimax concave penalty (MCP) with linear loss. Refs. [12,13] extended this to the family of nonconvex sparsity-inducing penalties with analytical solutions. However, Refs. [12,13] did not give the estimation error. Ref. [14] considered projected least squares and proposed the LinProj algorithm to decode unknown signals. Huang et al. [15] studied ℓ_1 -regularized least squares to decode from 1-bit noisy compressive measurements, and, respectively, proposed PDASC algorithms. Ref. [16] considered nonconvex least-squares loss with ℓ_q constrained with a higher sparsity level. To attain the target signal with more sparsity, Ref. [17] used the double sparsity of the unknown signal and the measurement to solve 1-bit CS optimization. The above works use the assumption of a sparse signal. In practice, there are other signals such as the gradient sparse signal or block sparse signal. Hou and Liu [18] studied the L1-TV model to solve signals possessing both element sparse and gradient sparse properties. Zhong et al. [19] proposed total variation minimization in the binary setting. Inspired by [20,21], the family of nonconvex sparsity reducing penalties has been shown to enjoy good performance when it approximates ℓ_0 -norm in conventional CS. In this paper, we try to study the least squares with minimax concave penalty (MCP) and other nonconvex sparsity-inducing penalties to decode from 1-bit CS measurements.

We end this section by giving the structure of this paper as follows. Section 2 gives some notation and preliminaries. Section 3 proposes the least-squares loss equipped with MCP and other nonconvex sparsity-inducing penalties, and discusses the estimation of the non-asymptotical error and the complexity. It then utilizes a Newton-type algorithm with a data-driven parameter voting rule to implement the proposed approach. In Section 4, experiments on synthetic and real datasets are implemented to verify the superiority of the proposed model, and illustrate the robustness and high efficiency of our method. Section 5 gives a short conclusion.

2. Preliminaries

In our work, the 1-bit CS measurement is generally modeled as [15,22]

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

where $A \in \mathbb{R}^{n \times p}$ is the Gaussian measurement matrix, $\mathbf{x} \in \mathbb{R}^p$ is the unknown signal to recover, the measurement $\mathbf{y} \in \{-1, 1\}^n$, $\mathbf{b} \in \mathbb{R}^n$ is the random Gauss noise sampled from $\mathcal{N}(0, \epsilon^2 I_n)$ with noise level ϵ , $\mathbf{c} \in \mathbb{R}^n$ is the sign flip, and \odot is the Hadamard product. Due to binary nature of the sign function, (1) is nonlinear, and so it is challenging to decode the unknown signal from the noisy and sign-flipped data.

We then introduce assumptions and notation in the sequel. The unknown recovered signal is assumed up to a positive constant. Since only the signs of real-valued measurements are preserved, scaling \mathbf{x} will not make changes to the measurements, i.e., $\text{sign}(A\mathbf{x} + \mathbf{b}) = \text{sign}(A\alpha\mathbf{x} + \alpha\mathbf{b})$ with $\forall \alpha > 0$, which indicates that the best way is to recover \mathbf{x} up to a scale factor. The measurement matrix $A = (A_1^t, A_2^t, \dots, A_n^t)^t \in \mathbb{R}^{n \times p}$ is a random matrix, whose rows $A_i \in \mathbb{R}^{p \times 1}$, $i = 1, 2, \dots, n$ is independently distributed

(i.i.d.) from $\mathcal{N}(\mathbf{0}, \Sigma)$, where Σ is the unknown covariance matrix. For $\mathbf{x} \in \mathbb{R}^p$, the ℓ_r -norm is defined as $\|\mathbf{x}\|_r = (\sum_{i=1}^r |x_i|^r)^{1/r}$, with $r \in (1, +\infty)$. $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq p} |x_i|$ is the infinity norm. Let \mathbf{x}^* be the target signal with k sparse level. And the true \mathbf{x} is given. We use $[p]$ to denote the integer set, i.e., $[p] = \{1, 2, \dots, p\}$. For a subset $\mathcal{U} \subseteq [p]$ with the cardinality of $|\mathcal{U}|$, \mathcal{V} is the supplementary set of \mathcal{U} , i.e., $\mathcal{V} = \bar{\mathcal{U}}$. $\mathbf{x}_{\mathcal{U}} = \{\mathbf{x}_i, i \in \mathcal{U}\} \in \mathbb{R}^{|\mathcal{U}|}$ and $A_{\mathcal{U}} = \{J_j, j \in \mathcal{U}\} \in \mathbb{R}^{n \times |\mathcal{U}|}$, where $J_j \in \mathbb{R}^{n \times 1}$ is the j th column of the measurement matrix A . For the entries of sign flips \mathbf{b} , it is assumed that the probability of positive signs is η and that of negative signs is $1 - \eta$. A , \mathbf{b} , and \mathbf{c} are mutually independent. $\mathcal{O}(\cdot)$ denotes that a positive numerical constant is ignored. Next, we give the definition and the theoretical basis.

Definition 1 ([20,23]). *The mapping $F : D \subseteq X \rightarrow Z$ is called Newton differentiable in the open subset $U \subseteq D$ if there exists a family of mappings $G : U \rightarrow \mathcal{L}(X, Z)$ such that*

$$\lim_{\|\mathbf{h}\| \rightarrow 0} \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - G(\mathbf{x} + \mathbf{h})\mathbf{h}\|}{\|\mathbf{h}\|} = 0, \forall \mathbf{x} \in U.$$

The mapping G is called a Newton derivative for F in U .

This definition can be used to analyze and solve the proposed model.

Lemma 1 ([20]). *Let $\tilde{\mathbf{y}}_i = \tilde{\mathbf{c}}_i \odot \text{sign}(\tilde{A}_i^t \mathbf{x} + \tilde{\mathbf{b}}_i)$ be the 1-bit model (1) at the population level, where the elements \mathbf{y}_i , \mathbf{c}_i , A_i , and \mathbf{b}_i in model (1) are random variables with the same distributions as those of $\tilde{\mathbf{y}}_i$, $\tilde{\mathbf{c}}_i$, \tilde{A}_i , and $\tilde{\mathbf{b}}_i$, respectively. Let $P[\tilde{\mathbf{c}} = 1] = \eta \neq \frac{1}{2}$, $A_i \in \mathcal{N}(\mathbf{0}, \Sigma)$, $\mathbf{b} \in \mathcal{N}(\mathbf{0}, \epsilon^2)$. It follows that*

$$\Sigma^{-1} \mathbb{E}[\tilde{\mathbf{y}} \tilde{A}_i] / \alpha = \mathbf{x},$$

where $\alpha = (2\eta - 1) \sqrt{\frac{2}{\pi(\epsilon^2 + 1)}}$ and \mathbf{x} is the true signal.

Based on this lemma, \mathbf{x} is a multiple of $\Sigma^{-1} \mathbb{E}[\tilde{\mathbf{y}} \tilde{A}_i] / \alpha$. Note that

$$\mathbb{E}[A^t A / n] = \mathbb{E}\left[\sum_{i=1}^n A_i A_i^t\right] / n = \Sigma,$$

and

$$\mathbb{E}[A^t \mathbf{y} / n] = \mathbb{E}\left[\sum_{i=1}^n A_i \mathbf{y}\right] / n = \mathbb{E}[\tilde{\mathbf{y}} \tilde{A}].$$

Meanwhile, if $A^t A / n$ is invertible, though \mathbf{y} consists of a flipped signal,

$$\mathbf{x}_{ls} = (A^t A / n)^{-1} A^t \mathbf{y} / n.$$

where \mathbf{x}_{ls} indicates the least-squares loss can be used as a loss function if $\Sigma \neq \mathbf{I}_n$.

3. The Least-Squares Approach with MCP Regularized

3.1. Combination of Least Squares and MCP

In this subsection, we try to recover the sparse signal \mathbf{x} from the 1-bit CS model (1) with $n < p$. L_0 -norm is the best choice to use for sparsity recovery. Nonetheless, the L_0 -norm optimization problem is generally difficult because of the nonconvexity and nonsmoothness. MCP regression models have the oracle property and produce unbiased solutions [24]. Based on these advantages, we combine MCP with least-squares loss to decode the unknown signal from the binary measurements as follows:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^p} \frac{1}{2n} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \mathcal{M}_{\lambda, \zeta}(\mathbf{x}), \tag{2}$$

where $\hat{\mathbf{x}} \in \mathbb{R}^p$ is the minimum of model (2), the element \mathbf{x} is denoted by \mathbf{x}_j , $j = 1, 2, \dots, p$, $\mathcal{M}_{\lambda, \zeta}(\mathbf{x}) = \sum_{j=1}^n m_{\lambda, \zeta}(\mathbf{x}_j)$ is defined by

$$m_{\lambda, \zeta}(\mathbf{x}_j) = \lambda \int_0^{\mathbf{x}_j} (1 - t / (\lambda \zeta))_+ dt$$

with a regularization parameter $\zeta > 1$, and $(t)_+ = \max\{t, 0\}$, where t is a scalar. The form of MCP after integration is generally used, namely,

$$m_{\lambda, \zeta}(t) = \begin{cases} \lambda(|t| - \frac{t^2}{2\lambda\zeta}), & |t| < \lambda\zeta, \\ \frac{\lambda^2\zeta}{2}, & |t| \geq \lambda\zeta. \end{cases} \tag{3}$$

Then, $m_{\lambda, \zeta}(t)$ can be formulated as the sum of the ℓ_1 -penalty and a concave part $m_{\lambda, \zeta}(t) = \lambda|t| + h_{\lambda, \zeta}(t)$ for all $|t| < \lambda\zeta$. Even though $m_{\lambda, \zeta}(t)$ is nonconvex and nonsmooth, $h_{\lambda, \zeta}(t)$ is differentiable. MCP satisfies the following unbiasedness, selection features, and bounded properties:

- (a) $m'_{\lambda, \zeta}(t) = 0, \forall t \geq \lambda\zeta \geq 0$;
- (b) $m'_{\lambda, \zeta}(0^+) = \lambda$, where 0^+ denotes that the variable approaches 0 from the right;
- (c) $|h'_{\lambda, \zeta}(t)| \leq \lambda, \forall |t| < \lambda\zeta$.

When $\zeta = \infty$, MCP coincides with the ℓ_1 -penalty; when $\zeta \rightarrow 0^+$, MCP approaches the L_0 -penalty. Further, these properties are suitable for the SCAD function. As we know, this is the first time the least-squares loss has been combined with MCP regularization in a 1-bit CS setting.

To solve the nonlinear model problem (2), we adopt the concentration of measure theorem and use the linear model to approximate the measurement data. First, we give one assumption.

Assumption 1 (ν -null consistency condition). Assume that $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ and $\nu \in (0, 1)$, the least-squares model with the nonconvex sparsity-inducing $P_\lambda(\mathbf{x})$ regularization, i.e.,

$$\min_{\mathbf{x} \in \mathbb{R}^p} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 / (2n) + \|P_\lambda(\mathbf{x})\|_1$$

has the ν -null consistency condition if it satisfies

$$\min_{\mathbf{x} \in \mathbb{R}^p} \|\mathbf{b}/\nu - \mathbf{A}\mathbf{x}\|_2^2 / (2n) + \|P_\lambda(\mathbf{x})\|_1 = \|\mathbf{b}/\nu\|_2^2 / (2n).$$

This condition guarantees that the global minimizer by the model is achievable at $\hat{\mathbf{x}} = 0$ if the target signal $\mathbf{x} = 0$. The null consistency condition is used to attain the ℓ_2 -regularity condition [25] for large-scale optimization.

Lemma 2. Assume that \mathbf{x}_{mcp} is a global minimizer of (2). Let $\tilde{\mathbf{x}} = \alpha\mathbf{x}^*$, where \mathbf{x}^* is the given true signal with k sparse level. $\mathbf{R} = \mathbf{x}_{mcp} - \tilde{\mathbf{x}}$; $\Delta = \mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}$; $\zeta = (1 + \nu) / (1 - \nu)$; and $\mathcal{S} = \text{supp}(\mathbf{x})$. If Assumption (1) holds, then it has

$$\frac{1}{2n} \|\mathbf{A}\mathbf{R}\|_2^2 + \|\mathcal{M}_{\lambda, \zeta}(\mathbf{R}_{\bar{\mathcal{S}}})\|_1 \leq \zeta \|\mathcal{M}_{\lambda, \zeta}(\mathbf{R}_{\mathcal{S}})\|_1. \tag{4}$$

Proof. Since \mathbf{x}_{mcp} is the minimizer of the problem (2), \mathbf{x}_{mcp} satisfies the following inequality

$$\frac{1}{2n} \|\mathbf{A}\mathbf{x}_{mcp} - \mathbf{y}\|_2^2 + \|\mathcal{M}_{\lambda, \zeta}(\mathbf{x}_{mcp})\|_1 \leq \frac{1}{2n} \|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{y}\|_2^2 + \|\mathcal{M}_{\lambda, \zeta}(\tilde{\mathbf{x}})\|_1. \tag{5}$$

We put $y = \Delta + A\bar{x}$ and $R = \mathbf{x}_{mcp} - \bar{x}$ into the above inequality (5), and obtain

$$\begin{aligned} \frac{1}{2n} \|AR - \Delta\|_2^2 + \|\mathcal{M}_{\lambda,\zeta}(\mathbf{x}_{mcp})\|_1 &\leq \frac{1}{2n} \|\Delta\|_2^2 + \|\mathcal{M}_{\lambda,\zeta}(\bar{x})\|_1. \\ \frac{1}{2n} \|AR\|_2^2 + \|\mathcal{M}_{\lambda,\zeta}(\mathbf{x}_{mcp})\|_1 &\leq \langle \mathbf{R}, A^t \Delta / m \rangle + \|\mathcal{M}_{\lambda,\zeta}(\bar{x})\|_1. \end{aligned}$$

Next, applying ν -null consistency, we have

$$\langle \mathbf{R}, A^t \Delta / n \rangle \leq \frac{\nu}{2n} \|AR\|_2^2 + \nu \|\mathcal{M}_{\lambda,\zeta}(\mathbf{R})\|_1. \tag{6}$$

Using $\|\mathbf{x} + \mathbf{y}\|_1 \leq \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1$ and $\bar{x}_{\bar{S}} = 0$ and (6), we have

$$(1 - \nu) \|AR\|_2^2 / (2n) \leq (\nu + 1) \|\mathcal{M}_{\lambda,\zeta}(\mathbf{R}_{\bar{S}})\|_1 + (\nu - 1) \|\mathcal{M}_{\lambda,\zeta}(\mathbf{R}_S)\|_1.$$

Finally, by $\zeta = (1 + \nu) / (1 - \nu)$, this gives the conclusion. \square

Due to Lemma 2, with $\mathbf{z} \in \mathbb{R}^p$ the error of the model (2) satisfies

$$\mathbf{R} \in \{\mathbf{z} : \|\mathcal{M}_{\lambda,\zeta}(\mathbf{z}_{\bar{S}})\|_1 \leq \zeta \|\mathcal{M}_{\lambda,\zeta}(\mathbf{z}_S)\|_1\}. \tag{7}$$

In order to give the bound of the random measurement matrix, we find the property of the measurement matrix. Here, an ℓ_2 -regularization condition is needed, which is used for concave penalty functions in other works.

Definition 2. Let $T \subseteq [p], \zeta > 0, q \geq 1$; a cone invertibility factor (CIF) is defined by [25]

$$CIF_q(\zeta, T) := \inf \left\{ \frac{|T|^{1/q} \|A^t A \mathbf{u}\|_\infty}{n \|\mathbf{u}\|_q} : \|\mathbf{u}_{\bar{T}}\|_1 \leq \zeta \|\mathbf{u}_T\|_1 \right\}. \tag{8}$$

Definition 3. Let $T \subseteq [p], \zeta > 0, q \geq 1$; $P_\lambda(\mathbf{x})$ denotes any penalty function; the restricted invertibility factor is defined as [25]

$$RIF_q(\zeta, T) := \inf \left\{ \frac{|T|^{1/q} \|A^t A \mathbf{u}\|_\infty}{n \|\mathbf{u}\|_q} : \|P_\lambda(\mathbf{u}_{\bar{T}})\|_1 \leq \zeta \|P_\lambda(\mathbf{u}_T)\|_1 \right\}. \tag{9}$$

Comparing the definitions of the above invertibility factors, the $RIF_q(\zeta, T)$ factor is more general than the $CIF_q(\zeta, T)$ factor. When P_λ is specifically the ℓ_1 -norm penalty function, $RIF_q = CIF_q$. For the MCP function $\mathcal{M}_{\lambda,\zeta}(t)$, there is a relationship between RIF_q and CIF_q described in the following proposition.

Lemma 3. If $t/m_{\lambda,\zeta}(t)$ is increasing when $t \in (0, \infty)$, then it holds that

$$RIF_q(\zeta, T) \geq \inf_{|B|=|T|} CIF_q(\zeta, B).$$

Hence, when $1 \leq q \leq 2$, $RIF_q(\zeta, T) \asymp 1$ is an ℓ_2 regularity condition for the measure matrix A , where the notation $f \asymp 1$ denotes f asymptotically approaching 1. Since $m_{\lambda,\zeta}(t)$ is concave in $t \in [0, \infty)$, then $t/m_{\lambda,\zeta}(t)$ is increasing in t . Lemma 3 is applicable to MCP, SCAD, and other nonconvex sparsity-inducing functions.

From this proposition, we know that if the lower bound of CIF_q exists, then the lower bound of RIF_q exists, and therefore, the upper bound of $1/RIF_q$ exists. When $1 \leq q \leq 2$, the sparse eigenvalue of the measurement is uniformly bound from below in terms of CIF_q [26]. Thus, there exists a constant C_0 with respect to the sparse eigenvalue of the given measurements, such that

$$1/RIF_q(\zeta, T) \leq C_0.$$

Let $q = 2$ and $T = \mathcal{S}$, with (7); for $\forall \mathbf{R} \in \{\mathbf{z} \in \mathbb{R}^p : \|\mathcal{M}_{\lambda, \zeta}(\mathbf{z}_{\mathcal{I}})\|_1 \leq \zeta \|\mathbf{z}_{\mathcal{S}}\|_1\}$, we attain the property of the measurement matrix as follows:

$$m \|\mathbf{R}\|_{2\text{RIF}_2(\zeta, \mathcal{S})} \leq |\mathcal{S}|^{1/2} \|A^t \mathbf{A} \mathbf{R}\|_{\infty}, \tag{10}$$

Then, we give the non-asymptotic error estimates of the model (2).

Theorem 1. *Let \mathbf{x}^* be the true signal with k -sparse level and $C_3 > \|\mathbf{x}\|_1$ and $C_0 \geq 1/\text{RIF}_2(\zeta, \mathcal{S})$. If the v -consistency condition (1) holds, as $n > \frac{4C_1}{C_2} \log p$, then with probability at least $1 - 2/p^3 - 2/p^2$, it holds that*

$$\|\mathbf{x}_{mcp}/\alpha - \mathbf{x}^*\|_2 \leq \frac{6(1 + |\alpha|C_3)}{C_0|\alpha|\sqrt{C_1}} \sqrt{\frac{k \log p}{n}}, \tag{11}$$

where C_1 and C_2 are generic positive constants depending on the maximum sub-Gaussian norm of rows of A , respectively, (see details in Lemma B.2 of [15]) and $\alpha = (2\eta - 1)\sqrt{2/(\pi(\epsilon^2 + 1))}$.

Proof. As \mathbf{x}_{mcp} is a minimizer of the model (2), from the optimization conditions of vector extremum problems, \mathbf{x}_{mcp} satisfies the following inequality:

$$\frac{1}{2n} \|A\mathbf{x}_{mcp} - \mathbf{y}\|_2^2 + \|\mathcal{M}_{\lambda, \zeta}(\mathbf{x}_{mcp})\|_1 \leq \frac{1}{2n} \|A\tilde{\mathbf{x}} - \mathbf{y}\|_2^2 + \|\mathcal{M}_{\lambda, \zeta}(\mathbf{x}_{mcp})\|_1.$$

From the KKT optimal condition, it has $\frac{1}{n} A^t(A\mathbf{x}_{mcp} - \mathbf{y}) + \partial \mathcal{M}_{\lambda, \zeta}(\mathbf{x}_{mcp}) = 0$. And we define $\mathbf{d}_{mcp} = \partial \mathcal{M}_{\lambda, \zeta}(\tilde{\mathbf{x}})$. Moreover, due to $|h'_{\lambda, \zeta}(t)| \leq \lambda$, we obtain

$$\|\mathbf{d}_{mcp}\|_{\infty} \leq 2\lambda. \tag{12}$$

Due to (10) and Lemma 2, it holds that

$$\begin{aligned} \|\mathbf{R}\|_{2\text{RIF}_2(\zeta, \mathcal{S})} &\leq \sqrt{k} \|A^t \mathbf{A} \mathbf{R}\|_{\infty} / n \\ &\leq \sqrt{k} \|A^t(A\mathbf{x}_{mcp} - \mathbf{y}) + A^t(\mathbf{y} - A\tilde{\mathbf{x}})\|_{\infty} / n \\ &\leq \sqrt{k} \left(\|A^t(A\mathbf{x}_{mcp} - \mathbf{y})\|_{\infty} / n + \|A^t(\mathbf{y} - A\tilde{\mathbf{x}})\|_{\infty} / n \right) \\ &\leq \sqrt{k} (\|\mathbf{d}_{mcp}\|_{\infty} + \|A^t \Delta\|_{\infty} / n) \\ &\leq \sqrt{k} (2\lambda + \lambda) \\ &\leq 3\lambda \sqrt{k}, \end{aligned}$$

where the third inequality uses the trigonometric inequality for the infinite norms, the right item of the fourth inequality uses Lemma D.2. in [15], and the fifth inequality applies Formula (12). Let $\lambda = \frac{2(1+|\alpha|C_3)}{\sqrt{C_1}} \sqrt{\frac{\log p}{n}}$, and then,

$$|\alpha| \|\mathbf{x}_{mcp}/\alpha - \mathbf{x}^*\|_2 \leq \frac{3\lambda}{\text{RIF}_2(\zeta, \mathcal{S})}.$$

Since $C_0 \geq 1/\text{RIF}_2(\zeta, \mathcal{S})$, we have

$$\|\mathbf{x}_{mcp}/\alpha - \mathbf{x}^*\|_2 \leq \frac{3\lambda}{|\alpha|C_0},$$

i.e.,

$$\|\mathbf{x}_{mcp}/\alpha - \mathbf{x}\|_2 \leq \frac{6(1 + |\alpha|C_3)}{C_0|\alpha|\sqrt{C_1}} \sqrt{\frac{k \log p}{n}}$$

This gives a conclusion. \square

Remark 1. Through Theorem 1, we know that for any $\delta > 0$, if $n \geq \mathcal{O}(\frac{k \log p}{\delta^2})$, the estimated signal decoded from (2) \mathbf{x}_{mcp} belongs to the δ -neighborhood of $\alpha \mathbf{x}^*$. Based on Theorem 1, we attain the same order of sample complexity $n = \mathcal{O}(\frac{k \log p}{\delta^2})$ with [8,15,17]. The least-squares model based on the MCP penalty function has lower sample complexity than that of the ℓ_q -constrained least-squares model [16], which coincides with less computational time in experiments.

Remark 2. MCP belongs to a class of nonconvex penalty functions that can reduce the sparsity, which are proved to satisfy the ν -null consistency (see details of [25]). The SCAD function also has the special property Section 3.1. Hence, under the ν -null consistency assumption and the definition of RIF for the least squares based on the SCAD penalty function, we can arrive at a similar upper bound for the reconstruction error.

Finally, we obtain the estimated error of a one-class sparsity-inducing nonconvex penalty. Here, as a convenience, we concentrate all regularization parameters in the penalty function on one parameter λ . For the 1-bit noisy and sign-flipped compressive sensing measurement, the least-squares model based on the sparse nonconvex penalty function is written as

$$\min_{\mathbf{x} \in \mathbb{R}^p} \frac{1}{2n} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \|P_\lambda(\mathbf{x})\|_1. \tag{13}$$

Theorem 2. Let $\mathcal{S} = \text{supp}(\mathbf{x})$ and $\nu \in (0, 1)$, $\zeta = (1 + \nu)/(1 - \nu)$. Suppose that $\hat{\mathbf{x}}$ is a minimum of model (13). If the ν -null consistency condition 1 holds, it has

$$\|\hat{\mathbf{x}}/\alpha - \mathbf{x}\|_2 \leq C_0(1 + \nu)\lambda k^{1/2}/\alpha, \tag{14}$$

where $\lambda := \inf_{t>0} \{t/2 + \phi_\lambda(t)/t\}$, $\alpha = (2\eta - 1)\sqrt{2/(\pi(\epsilon^2 + 1))}$, and $C_0 \geq 1/\text{RIF}_2(\zeta, \mathcal{S})$.

Proof. Let $\mathbf{R} = \hat{\mathbf{x}} - \alpha \mathbf{x}$ and $\Delta = \mathbf{y} - \mathbf{A}(\alpha \mathbf{x})$. Due to the given conditions, we have that if Assumption 1 holds, then it has

$$\min_{\mathbf{x} \in \mathbb{R}^p} (\|\Delta/\nu - \mathbf{Ax}\|_2^2/(2n) + \|P_\lambda(\mathbf{x})\|_1) = \|\Delta/\nu\|_2^2/(2n).$$

Therefore, it holds that

$$P_\lambda(\mathbf{R}_{\overline{\mathcal{S}}})\|_1 \leq \zeta \|P_\lambda(\mathbf{R}_{\mathcal{S}})\|_1,$$

and

$$\|A^t \Delta/n\|_\infty \leq \nu \lambda.$$

Since the lower bound of $\text{RIF}_2(\zeta, \mathcal{S})$ exists, we obtain $C_0 \geq 1/\text{RIF}_2(\zeta, \mathcal{S})$, and

$$\|A^t \mathbf{AR}\|_\infty \leq (1 + \nu)\lambda.$$

So, we give a conclusion about the reconstruction error that

$$\|\hat{\mathbf{x}}/\alpha \mathbf{x} - \mathbf{x}^*\|_2 \leq C_0(1 + \nu)\lambda k^{1/2}/\alpha.$$

□

Remark 3. The above theorem gives the estimated error under the general cases. That is, there exists a upper bound for the least-squares loss with a class of sparsity-inducing nonconvex penalty functions. Theorem 1 gives a more precise calculation for the reconstruction error. We can extend the result of the linear measurement model to the nonlinear 1-bit measurement model.

For the comparison, Table 1 is reported.

Table 1. The complexity of 1-bit compressive sensing for many methods.

1-Bit CS	x Is Exactly Sparse		x Is Approximately Sparse		Others
	Sample Complexity	Reference	Sample Complexity	Reference	Reference
Noiseless	$\mathcal{O}(\frac{s \log n}{\delta})$	[4,27]	$\mathcal{O}(\frac{s \log^2 n}{\delta^5})$	[22]	-
	$\mathcal{O}(\frac{s \log n}{\delta^5})$	[22]			
Noisy	$\mathcal{O}(\frac{s \log n}{\delta^4})$	[7]	$\mathcal{O}(\frac{s \log n}{\delta^4})$	[8,22]	$ x_i \geq \nu, \mathcal{O}(s/\delta^2)$ [11]
	$\mathcal{O}(\frac{s \log n}{\delta^2})$	[8,15] (2)			$ x_i \leq \nu, \mathcal{O}(s \log n / \delta^2)$ [11]
	$\mathcal{O}(\min(\frac{s \log n}{\delta^2}, \frac{s \sqrt{n} \log n}{\delta}))$	[8]	$\mathcal{O}(\min(\frac{s \log n}{\delta^4}, \frac{s \sqrt{n} \log n}{\delta^5}))$	[8]	-
	$\mathcal{O}(\frac{s^{1/q-1/2} \log n}{\delta^2})$	[16]	-	-	-

3.2. Algorithm for (2)

For the proposed method, MCP and other penalties are nonconvex. This is not easy to compute effectively. Fortunately, these nonconvex sparsity-inducing functions equipped with the least-squares loss obtain the thresholding operator [20]. We first give the description of the thresholding operator and derive the Karush–Kuhn–Tucker (KKT) equations of the minimizer solution and necessary condition for the global minimizer of (2). Specifically, we show that the global minimizer satisfies the Karush–Kuhn–Tucker (KKT) equations.

Theorem 3. Let \mathbf{x}_{mcp} be the global minimizer of (2). Then,

$$\begin{cases} \hat{\mathbf{d}} = A^t(\mathbf{y} - A\mathbf{x}_{mcp})/m, \\ \mathbf{x}_{mcp} = \mathcal{T}_{\lambda,\xi}(\hat{\mathbf{d}} + \mathbf{x}_{mcp}), \end{cases} \tag{15}$$

where $\mathcal{T}_{\lambda,\xi}(t)$ is the thresholding operator of the MCP function, defined as

$$\mathcal{T}_{\lambda,\xi}(t) = \begin{cases} 0, & |t| \leq \lambda, \\ \text{sgn}(t) \frac{\xi|t| - \lambda}{\xi - 1}, & |t| < \lambda\xi, \\ t, & |t| > \lambda\xi. \end{cases} \tag{16}$$

Proof. The proof refers to Lemma 3.4 of [20]. □

Then, based on the KKT system, we let $J_a = A^t A/n$, $\hat{\mathbf{y}} = A^t \mathbf{y}/n$ and introduce a nonlinear operator $F : \mathbb{R}^{2p} \rightarrow \mathbb{R}^{2p}$ as

$$F(\mathbf{x}, \mathbf{d}) = \begin{bmatrix} J_a \mathbf{x} + \mathbf{d} - \hat{\mathbf{y}} \\ \mathbf{x} - \mathcal{T}_{\lambda,\xi}(\mathbf{d} + \mathbf{x}) \end{bmatrix}$$

Let $\mathbf{z} = (\mathbf{x}; \mathbf{d})$ be the root of $F(\mathbf{x}, \mathbf{d}) = 0$. We use the Newton-type to find the root. Denote the derivative function of $F(\mathbf{x}, \mathbf{d})$ as $\nabla F = G(\mathbf{x}, \mathbf{d})$ by

$$G(\mathbf{x}, \mathbf{d}) = \begin{bmatrix} J_a & I \\ I_S^> & -I_S \end{bmatrix}$$

From (15), we can derive that

$$|(\mathbf{x}_{mcp} + \hat{\mathbf{d}})_i| < \lambda \Rightarrow (\mathbf{x}_{mcp})_i \neq 0. \tag{17}$$

For an initial $\mathbf{x}^0 \in \mathbb{R}^p$ and $\mathbf{d}^0 \in \mathbb{R}^p$ given [20], we implement the following procedure to update the l -th iterated value $\mathbf{x}^{l+1} \in \mathbb{R}^p, \mathbf{d}^{l+1} \in \mathbb{R}^p$. Let an active set $\mathcal{U}_l = \{i : |\mathbf{x}_i^l + \mathbf{d}_i^l| > \lambda\}$ and an inactive set $\mathcal{V}_l = \overline{\mathcal{U}_l}$. Then, we use the Newton update as

$$\begin{bmatrix} \mathbf{x}^{l+1} \\ \mathbf{d}^{l+1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^l \\ \mathbf{d}^l \end{bmatrix} - G(\mathbf{x}^l, \mathbf{d}^l)^{-1} F(\mathbf{x}^l, \mathbf{d}^l)$$

which, after some algebra computation, becomes

$$\begin{cases} \mathbf{x}_{\mathcal{V}_l}^l = \mathbf{0}, \\ A_{\mathcal{U}_l}^t A_{\mathcal{U}_l} \mathbf{x}_{\mathcal{U}_l}^l = A_{\mathcal{U}_l}^t \mathbf{y} - n \mathbf{v}_{\mathcal{U}_l}^l, \\ \mathbf{d}^l = A^t(\mathbf{y} - A\mathbf{x}^l)/n \end{cases} \tag{18}$$

with

$$\forall i \in \mathcal{U}_l, \quad \mathbf{v}_i^l = \begin{cases} \frac{1}{\xi}(\text{sign}(\mathbf{x}_i^l + \mathbf{d}_i^l)\lambda\xi - \mathbf{x}_i^l), & \text{if } \lambda \leq |\mathbf{x}_i^l + \mathbf{d}_i^l| \leq \lambda\xi \text{ and } |\mathbf{x}_i^l \mathbf{d}_i^l| \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

To obtain the effective solution, we introduce the weight $w \in (0, 1)$ into the active set, i.e., $\mathcal{U}_l = \{i : |w\mathbf{x}_i^l + (1 - w)\mathbf{d}_i^l| > \lambda\}$.

The method is a Newton-type method, and therefore, it heavily relies on the input value and the Lagrange multiplier λ to obtain a meaningful result. So, the continuation technique [15,20] is used to attain the initial value and a proper tuning parameter is simultaneously selected by using a decreasing sequence and majority-vote strategy. Specifically, from a decreasing sequence $\{\lambda_l\}$, the output of the λ_l -problem is used as the initial value of the λ_{l+1} -problem. Let $\lambda_0 \geq \|A^t \mathbf{y}/n\|$, since $\mathbf{0}$ is a minimizer to the model (2), and $\lambda_l = \lambda_0 \beta^l, \beta \in (0, 1), l = 1, \dots$. By using the series $\{\lambda_l\}_k$ (a scalar), we conduct the algorithm until the stopping criterion is satisfied, i.e., $\|\mathbf{x}_{\lambda_l}\|_0 > \lfloor \frac{m}{\log n} \rfloor$, where $\mathbf{x}_{\lambda_l} \in \mathbb{R}^p, \mathbf{d}^l \in \mathbb{R}^p$ and the symbol $\lfloor \cdot \rfloor$ represents the largest integer smaller than itself. The algorithm on λ_l is conducted to obtain a solution path and select a λ by using the majority-vote rule [28], i.e.,

$$\hat{\lambda} = \max_i \{\Lambda_i\} \text{ and } \bar{i} = \arg \max_i \{|\Lambda_i|\}. \tag{19}$$

Then, we solve the proposed model by Algorithm 1 below, for simplicity termed MCPWP.

Algorithm 1 MCPWP

- 1: Input: $y, A, b, w, L, \alpha, \lambda_0 = \|A^t \mathbf{y}\|_\infty, \mathbf{x}_{\lambda_0} = \mathbf{0}$.
 - 2: **for** $l = 1, 2, \dots, L$ **do**
 - 3: $\lambda_l = \lambda_0 \beta^{l-1}, \mathbf{x}^0 = \mathbf{x}_{\lambda_{l-1}}, \mathbf{d}^0 = A^t(\mathbf{y} - A\mathbf{x}^0)/n$;
 - 4: **for** $j = 1, 2, \dots, J$ **do**
 - 5: update the active sets \mathcal{U}_j and inactive sets \mathcal{V}_j ,
 - 6: $\mathcal{U}_j = \{i : |w\mathbf{x}_i^{j-1} + (1 - w)\mathbf{d}_i^{j-1}| > \lambda_j\}$ and $\mathcal{V}_j = \overline{\mathcal{A}_j}$;
 - 7: update \mathbf{x}_j and \mathbf{d}_j by (18)
 - 8: Check stopping criterion: $\mathcal{U}_j = \mathcal{U}_{j-1}$;
 - 9: **end for**
 - 10: Stop if $\|\mathbf{x}_{\lambda_l}\|_0 > \lfloor \frac{n}{\log p} \rfloor$;
 - 11: **end for**
 - 12: Output: Solution path $\mathbf{x}_{\lambda_k}, k = 1, 2, \dots$;
 - 13: Select $\hat{\lambda}$ by majority vote rule (19);
 - 14: Output $\mathbf{x}_{\hat{\lambda}}$.
-

4. Experiments

We conduct all experiments on MATLAB 2018a running on a PC with Intel(R) Core (TM)i5-8265 CPU (1.60 GHz) and RAM with 8 GB. For the code of the proposed algorithm, refer to “<https://github.com/cjia80/mcpwp>” (accessed on 12 June 2024).

4.1. Testing Examples

Example (correlated covariance [15]). The rows of a random Gaussian measurement matrix $A \in \mathbb{R}^{n \times p}$ are independently and identically distributed (i.i.d.) from $\mathcal{N}(\mathbf{0}, \Sigma)$ with $\Sigma_{l,s} = \mu^{|s-l|}, 1 \leq l, s \leq n, \mu \in (0, 1)$. The nonzero true signal vector \mathbf{x} is generated from the normal Gaussian distribution. To avoid negligible entries of the true signal, we modify $\mathbf{x} = \mathbf{x} + \text{sign}(\mathbf{x})$, while ensuring that the number of nonzero elements is k . The noise term b is sampled from the Gaussian distribution, i.e., $b \sim \mathcal{N}(0, \epsilon^2 \mathbf{I}_m)$, where ϵ represents the noise level. The sign flips with the probability η are also given. Then, the measurement data y are generated by $\mathbf{y} = \mathbf{c} \odot \text{sign}(A\mathbf{x} + \mathbf{b})$. Additionally, for the data-driven selection method, we divide the interval $[\lambda_L, \lambda_0]$ into L subintervals, i.e., $[\lambda_l, \lambda_{l+1}], l = 0, \dots, L, L = 200$ to choose the parameter λ and obtain a good initial value of our Newton-type algorithm. All the 1-bit recovered signals need to be normalized for comparison. Furthermore, based on previous studies [6,8,15], we set all experiments with 100 epochs, and manually adjust values of ζ and w in subsequent epochs. For the comparison of the recovered performance, we choose the computational (CPU) time, the exact recovery probability (PrE; defined as the frequency of the number of trials where the estimated support coincides with the true support), and the signal-to-noise ratio (SNR; denoted by $-20 * \log_{10} \|\mathbf{x}^* - \mathbf{x}_{rec}\|$) in dB (where \mathbf{x}_{rec} is the recovered signal) or estimated error ($\|\mathbf{x}^* - \mathbf{x}_{rec}\|$).

4.2. Implementation and the Effect of Parameter Selection

We firstly consider the effect of the probability of the sparsity k , the noise level ϵ , and the sign flip probability η when using the MCPWP algorithm to solve the model (2). We run MCPWP under the following three cases: $p = 1000, n = 500, \mu = 0.3$, with various sparsity k , noise level ϵ , and probability of sign flips η . The experiment results are plotted accordingly in Figure 1. Let us take a close look at the results. From Figure 1a, the proposed method has an accuracy of 99% PrE if the target is very sparse and is above 91% if $k < 10$. In Figure 1b, the recovery probability changes very little if the noise level is below 0.8. The same result is shown in terms of various sign flip rates in Figure 1c. This figure implies that our method is robust to the sparsity, noise level, and sign flips.

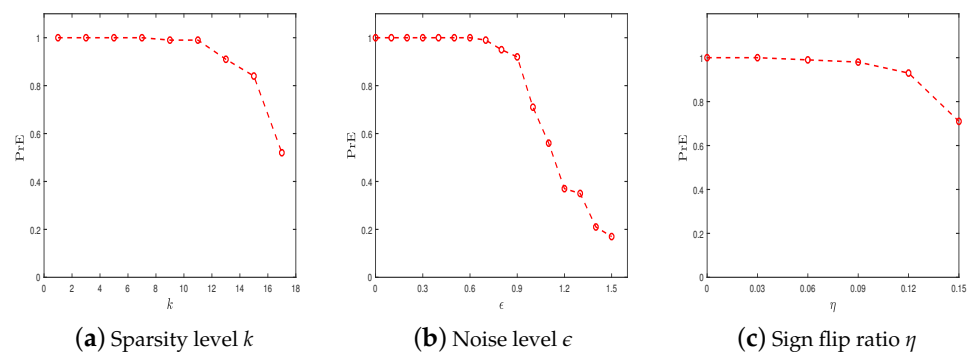


Figure 1. The robustness of MCPWP for various k, ϵ, η under $p = 1000, n = 500, \mu = 0.3$. (a) $k = 1:2:20, \epsilon = 0.01, \eta = 0.05$; (b) $k = 5, \epsilon = 0 : 0.1 : 1, \eta = 0.05$; (c) $k = 5, \mu = 0.3, \epsilon = 0.01, \eta = 0:0.03:0.15$.

4.3. Benchmark Methods

In this test, we try to verify the superiority of our method in comparison with other methods with respect to SNR, the exact probability (PrE(%)), and average computational time (CPU time (s)). The benchmark methods consist of BIHT [4], AOP [6], PBAOP [10], Lin-Proj [29,30], PDASC [15], MCP [12], WPDASC [16], and [17]. Although GPSP is an effective

method to improve sparsity in synthesis experiments, GPSP needs the two priors of sparsity. And the first four algorithms and MCP require the prior knowledge of sparsity. In addition, the sign flip rate is specified in AOP and PBAOP. Our method, PDASC, and WPDASC do not need priors. The performance is illustrated in Figures 2 and 3 through the three indexes SNR, PrE, and CPU time when the experiments are conducted at various sparsity level settings. Figure 2 shows that MCPWP and WPDASC exhibit superior performance, while GPSP, MCP, and BIHT are more likely to fail as the sparsity increases. In Figure 3a, MCPWP demonstrates a higher exact recovery probability compared to all other methods except for WPDASC. Meanwhile, it is evident that MCPWP requires less CPU time than WPDASC from Figure 3b. These findings highlight that our method MCPWP is an optimal choice in terms of both performance metrics here. Then, the performance of different methods on the various noise levels with others' fixed is depicted in Figures 4 and 5. As is plotted, MCPWP generally exhibits a higher SNR than other methods across most scenarios. Notably, for the three noise levels, MCPWP demonstrates the highest SNR, which reveals the robustness of the proposed method. When the sign flip rate becomes larger, our method has a higher SNR and less CPU time, as illustrated in Figures 6 and 7. In addition, with high noise ϵ and a high sign flip rate η , MCPWP has a higher SNR than the benchmark methods. Additionally, as the sparsity k increases, MCPWP surpasses the other approaches in CPU time except for LinProj (due to its lower computational time as a convex approach). Thus, the above experiments demonstrate that MCPWP is robust against noise, sign flips, and sparsity level, and also outperforms other methods. More higher-dimensional experiments are reported in Table 2. By comparison with other methods, our model is superior.

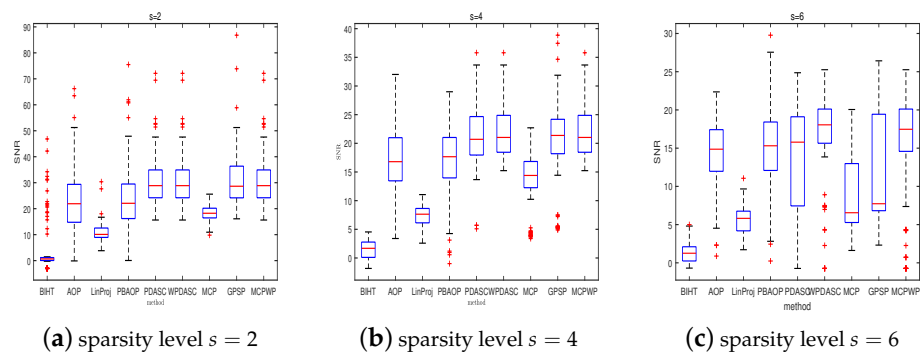


Figure 2. The SNR for different methods with different sparsities. $p = 1000$, $n = 500$, $k = 2:2:6$, $\mu = 0.1$, $\epsilon = 0.05$, $\eta = 0.01$.

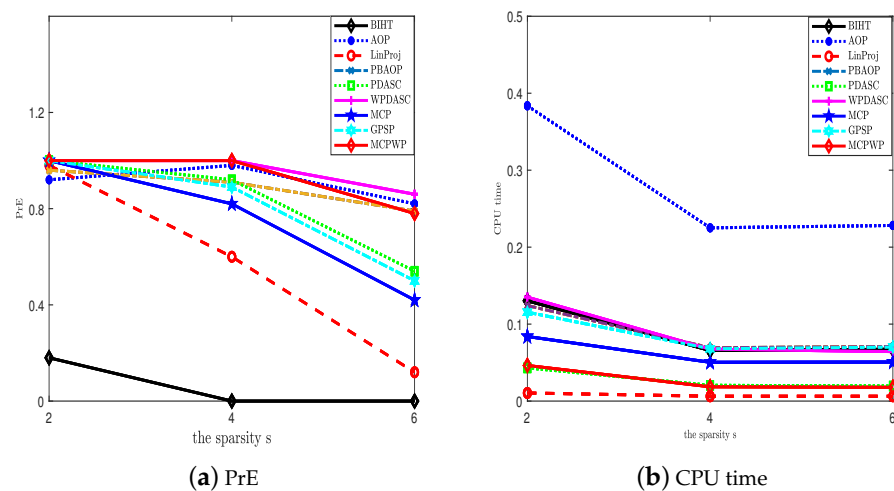


Figure 3. The exact probability and CPU time for different methods with different sparsities. $p = 1000$, $n = 500$, $k = 2:2:6$, $\mu = 0.1$, $\epsilon = 0.05$, $\eta = 0.01$.

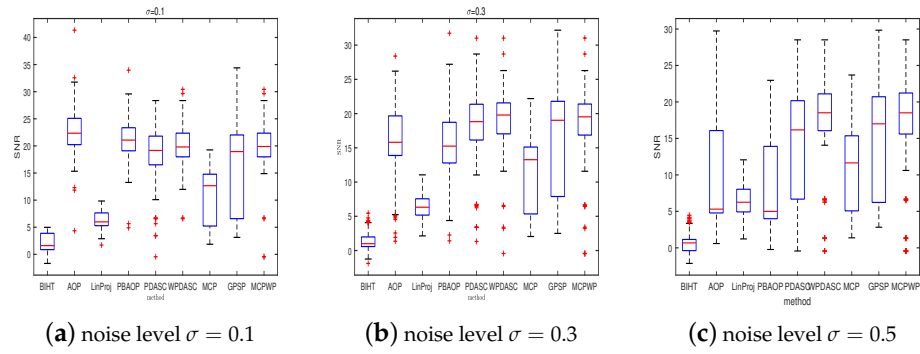


Figure 4. The SNR for different methods with different sparsities. $p = 1000, n = 500, k = 5, \mu = 0.1, \epsilon = 0.1 : 0.2 : 0.05, \eta = 0.01$.

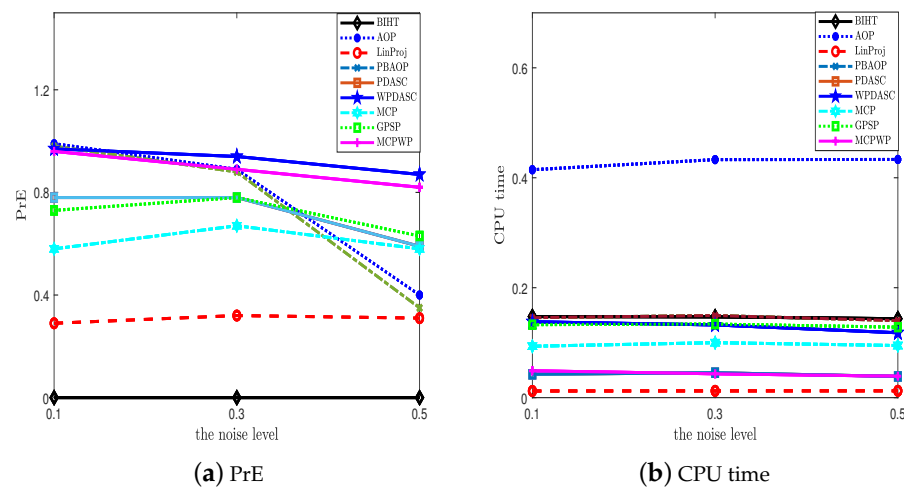


Figure 5. The exact probability and CPU time for different methods with different sparsities. $p = 1000, n = 500, k = 5, \mu = 0.1, \epsilon = 0.1 : 0.2 : 0.5, \eta = 0.01$.

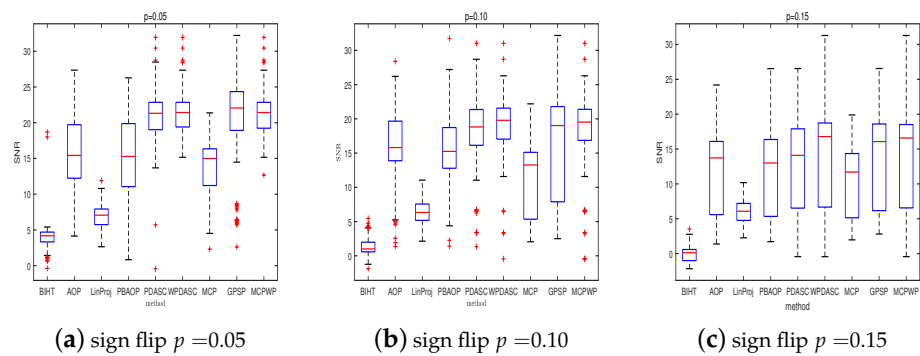


Figure 6. The SNR for different methods with different sparsities. $p = 1000, n = 500, k = 5, \mu = 0.1, \epsilon = 0.05, \eta = 0.05:0.05:0.15$.

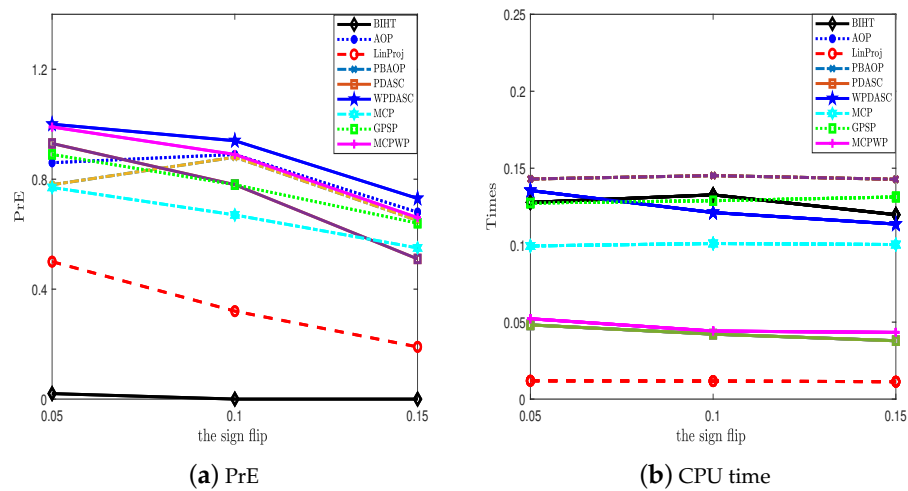


Figure 7. the exact probability and CPU time for different methods on the different sparsity. $p = 1000, n = 500, k = 5, \mu = 0.1, \epsilon = 0.05, \eta = 0.05:0.05:0.15$.

Table 2. Comparison of many methods under higher-dimension situations.

Method	(a) $p = 10,000, n = 2000, k = 30$ $\mu = 0.3, \epsilon = 0.3, \eta = 0.05$			(b) $p = 10,000, n = 2000, k = 30$ $\mu = 0.3, \epsilon = 0.5, \eta = 0.1$			(c) $p = 30,000, n = 7000, k = 20$ $\mu = 0.3, \epsilon = 0.3, \eta = 0.1$		
	Time (s)	ℓ_2 -error	PrE (%)	Time (s)	ℓ_2 -error	PrE (%)	Time (s)	ℓ_2 -error	PrE (%)
BIHT	4.28	5.55×10^{-1}	84	3.82	7.12×10^{-1}	0	8.98×10^{-2}	7.06×10^{-1}	0
AOP	1.10×10^1	1.75×10^{-1}	100	9.25	3.19×10^{-1}	0	1.10×10^{-1}	1.95×10^{-1}	90
LinProj	2.16×10^{-1}	4.96×10^{-1}	61	1.65×10^{-1}	5.12×10^{-1}	0	4.00×10^{-4}	3.59×10^{-1}	90
PBAOP	4.21	3.15×10^{-1}	100	3.68	2.98×10^{-1}	0	3.68×10^{-2}	1.68×10^{-1}	90
PDASC	1.17	1.19×10^{-1}	97	8.85×10^{-1}	3.37×10^{-1}	0	1.05×10^{-2}	6.28×10^{-2}	100
GPSP	3.11×10^1	2.56×10^{-1}	89	6.81×10^2	1.10	0	6.17	6.28×10^2	100
MCP	8.02×10^{-1}	3.24×10^{-1}	89	6.34×10^{-1}	4.22×10^{-1}	0	1.19×10^{-1}	6.28×10^{-2}	100
WPDASC	1.80	1.19×10^{-1}	100	1.41	4.22×10^{-1}	100	3.36×10^{-1}	6.28×10^{-2}	100
MCPWP	6.43×10^{-1}	1.58×10^{-1}	100	5.89×10^{-1}	1.44×10^{-1}	100	5.10×10^{-3}	6.28×10^{-2}	100

Method	(a) $p = 25,000, n = 5000, k = 20$ $\mu = 0.3, \epsilon = 0.3, \eta = 0.01$			(b) $p = 25,000, n = 5000, k = 20$ $\mu = 0.3, \epsilon = 0.5, \eta = 0.1$			(c) $p = 30,000, n = 7000, k = 20$ $\mu = 0.3, \epsilon = 0.5, \eta = 0.1$		
	Time (s)	ℓ_2 -error	PrE (%)	Time (s)	ℓ_2 -error	PrE (%)	Time (s)	ℓ_2 -error	PrE (%)
BIHT	2.37×10^1	2.20×10^{-1}	70	2.16×10^1	6.80×10^{-1}	0	3.65×10^2	7.91×10^{-1}	0
AOP	5.83×10^1	1.62×10^{-1}	80	5.25×10^1	1.91×10^{-1}	90	1.09×10^2	2.35×10^{-1}	70
LinProj	3.36×10^{-1}	3.89×10^{-1}	90	3.11×10^1	3.89×10^{-1}	100	4.00	3.61×10^{-1}	100
PBAOP	2.18×10^1	3.61×10^{-1}	70	2.15×10^1	2.04×10^{-1}	100	3.49×10^1	3.59×10^{-1}	60
PDASC	7.22	4.93×10^{-2}	100	6.16×10^1	7.46×10^{-2}	90	9.80	7.20×10^{-2}	100
GPSP	4.19×10^2	4.93×10^2	100	4.23×10^2	7.46×10^{-2}	100	1.44×10^3	1.03×10^{-1}	80
MCP	6.49	4.93×10^{-2}	100	5.79	7.46×10^{-2}	100	1.12×10^1	7.20×10^{-2}	100
WPDASC	9.31	4.93×10^{-2}	100	6.02	7.46×10^{-2}	100	5.44	7.20×10^{-2}	100
MCPWP	3.67	4.93×10^{-2}	100	3.00	7.47×10^{-2}	100	5.20	7.20×10^{-2}	100

4.4. Numerical Comparisons on 1-D Real Data

In this section, we aim to decode a 1-dimensional signal from real-world data with $k = 15$ under the wavelet basis “Db1” [31]. There is a random Gaussian and an inverse wavelet transform in measurement matrix $A \in \mathbb{R}^{n \times p}$ with the noise level $\epsilon = 0.5$ and sign flip ratio $\eta = 0.04$. It should be noted that since GPSP is not suitable for real data, it has been omitted from our analysis. The behaviors of the aforementioned methods are reported in Table 3 and Figure 8. Among three indexes, the peak signal-to-noise ratio (PSNR) is defined as $10 \log \frac{M^2}{\|x_{rec} - x\|^2}$, in which M is the absolute value of the maximum in the true vector. In Table 3, as shown in our findings, MCPWP has both higher PSNR values and smaller recovery errors compared to other methods while maintaining a faster CPU time. These results demonstrate the superiority of our decoder.

Table 3. Comparison of methods for 1-D real signal. $p = 8000, n = 2500, \epsilon = 0.5, \eta = 0.04$.

Method	BIHT	AOP	LinProj	PBAOP	PDASC	WPDASC	MCP	MCPWP
PSNR	28.82	32.65	32.81	30.62	36.44	44.89	35.82	46.32
Err	0.764	0.491	0.482	0.621	0.318	0.120	0.341	0.102
Time (s)	3.299	3.410	0.133	3.493	2.320	2.874	0.358	1.825

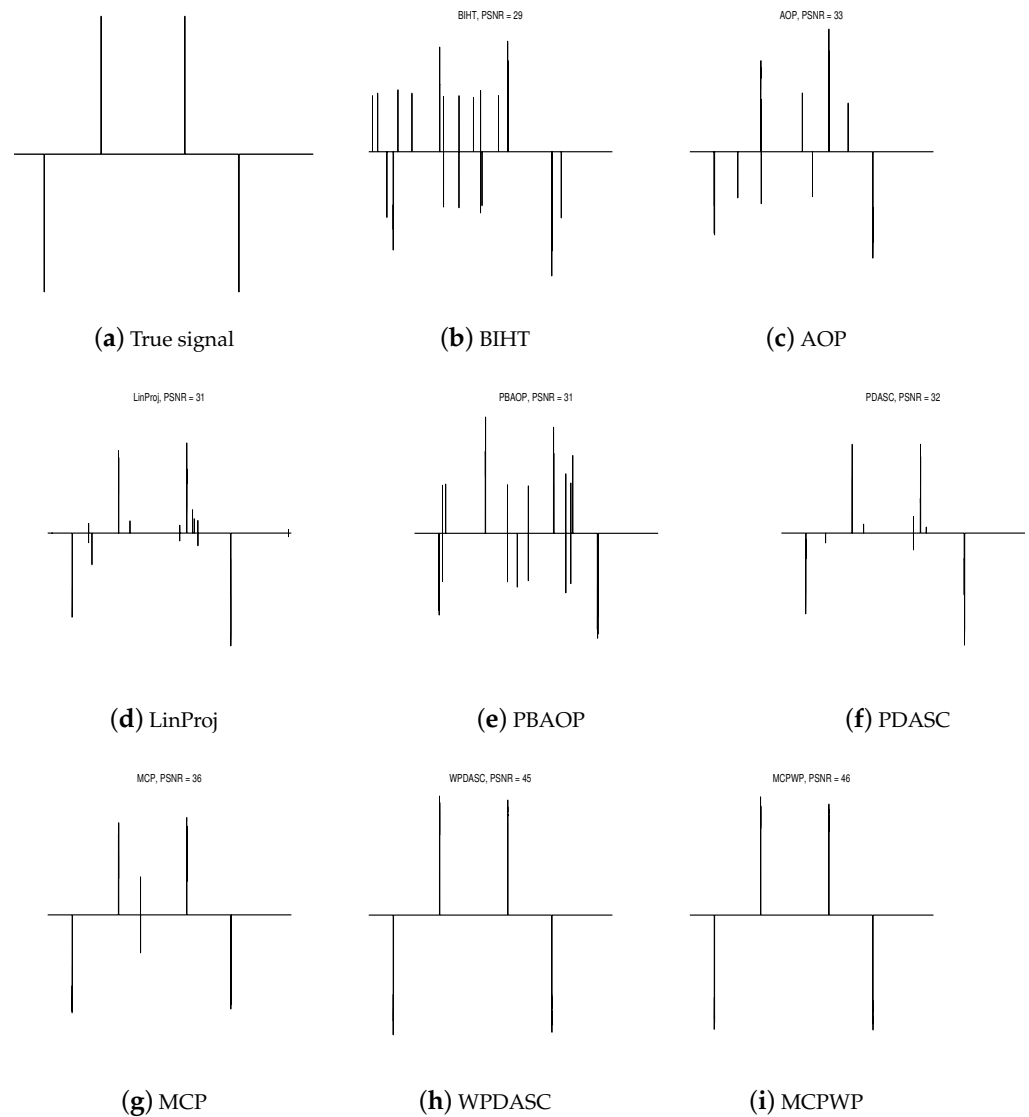


Figure 8. Comparison of different methods under 1-D real signal. $p = 8000, n = 2500, k = 15, \mu = 0.1, \epsilon = 0.5, \eta = 0.04$.

Finally, we conduct an investigation into the performance of the sparsity-inducing nonconvex regularization function, which contains ℓ_q , MCP, SCAD, and Capped- ℓ_1 . We utilize the data-driven semi-smooth Newton algorithm equipped with continuity described in [20]. The recovery performance with different regularization functions is presented in Table 4. It is worth noting that the CPU time is very similar across all cases, indicating that the least-squares method with sparsity-inducing nonconvex functions proves to be efficient for 1-bit compressive sensing.

Table 4. Comparison of penalty functions under different situations.

Method	(a) $p = 1000, n = 500, k = 5$ $\mu = 0.3, \epsilon = 0.3, \eta = 0.05$			(b) $p = 5000, n = 1000, k = 15$ $\mu = 0.5, \epsilon = 0.3, \eta = 0.05$			(c) $p = 20,000, n = 5000, k = 50$ $\mu = 0, \epsilon = 0.3, \eta = 0.05$		
	Time (s)	ℓ_2 -error	PrE (%)	Time (s)	ℓ_2 -error	PrE (%)	Time (s)	ℓ_2 -error	PrE (%)
WPDASC	1.74×10^{-2}	8.43×10^{-2}	100	2.69×10^{-1}	1.44×10^{-1}	92	5.62	1.03×10^{-1}	100
MCPWP	2.10×10^{-2}	8.43×10^{-2}	100	1.84×10^{-1}	1.48×10^{-1}	90	4.02	1.03×10^{-1}	99
SCAD	2.30×10^{-2}	8.43×10^{-2}	100	1.84×10^{-1}	1.53×10^{-1}	88	3.95	1.03×10^{-1}	99
Capped- ℓ_1	1.78×10^{-2}	8.43×10^{-2}	100	1.82×10^{-1}	1.45×10^{-1}	91	1.00	1.03×10^{-1}	99

5. Conclusions

In this work, we have analyzed the least-squares regularization approach with MCP and other nonconvex sparsity-inducing functions to reconstruct the signal from 1-bit CS data with added noise and sign flips. We have obtained the competitive complexity of the proposed method, i.e., $m \geq \mathcal{O}(\frac{s \log n}{\delta^2})$ with the accuracy δ as long as $\eta \neq 1/2$. A data-driven Newton type with an active set and majority-vote selection rule are utilized to implement the unknown signal. Through a lot of numerical simulations, we validate that the proposed model surpasses other state-of-the-art methods.

Author Contributions: Conceptualization, C.J. and L.Z.; methodology, C.J.; validation, C.J.; writing—original draft preparation, C.J.; writing—review and editing, L.Z.; funding acquisition, C.J. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by OF FUNDER, grant number Y202147627, Provincial Department of Education/Research Program.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Boufounos, P.T.; Baraniuk, R.G. 1-bit compressive sensing. In Proceedings of the 2008 42nd Annual Conference on Information Sciences and Systems, Princeton, NJ, USA, 19–21 March 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 16–21.
- Fan, X.; Wang, Y.; Huo, Y.; Tian, Z. 1-bit compressive sensing for efficient federated learning over the air. *IEEE Trans. Wirel. Commun.* **2022**, *22*, 2139–2155. [\[CrossRef\]](#)
- Qing, C.; Ye, Q.; Cai, B.; Liu, W.; Wang, J. Deep learning for 1-bit compressed sensing-based superimposed CSI feedback. *PLoS ONE* **2022**, *17*, e0265109. [\[CrossRef\]](#)
- Laurent, J.; Jason N., L.; Petros T., B.; Richard G., B. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Trans. Inf. Theory* **2013**, *59*, 2082–2102.
- Matsumoto, N.; Mazumdar, A. Robust 1-bit Compressed Sensing with Iterative Hard Thresholding. In Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, Alexandria, VA, USA, 7–10 January 2024; pp. 2941–2979.
- Yan, M.; Yang, Y.; Osher, S. Robust 1-bit compressive sensing using adaptive outlier pursuit. *IEEE Trans. Signal Process.* **2012**, *60*, 3868–3875. [\[CrossRef\]](#)
- Plan, Y.; Vershynin, R. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Trans. Inf. Theory* **2012**, *59*, 482–494. [\[CrossRef\]](#)
- Zhang, L.; Yi, J.; Jin, R. Efficient algorithms for robust one-bit compressive sensing. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 820–828.
- Dai, D.Q.; Shen, L.; Xu, Y.; Zhang, N. Noisy 1-bit compressive sensing: Models and algorithms. *Appl. Comput. Harmon. Anal.* **2016**, *40*, 1–32. [\[CrossRef\]](#)
- Huang, X.; Shi, L.; Yan, M.; Suykens, J.A. Pinball loss minimization for one-bit compressive sensing: Convex models and algorithms. *Neurocomputing* **2018**, *314*, 275–283. [\[CrossRef\]](#)
- Zhu, R.; Gu, Q. Towards a lower sample complexity for robust one-bit compressed sensing. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 739–747.
- Huang, X.; Yan, M. Nonconvex penalties with analytical solutions for one-bit compressive sensing. *Signal Process.* **2018**, *144*, 341–351. [\[CrossRef\]](#)
- Xiao, P.; Liao, B.; Li, J. One-bit compressive sensing via Schur-concave function minimization. *IEEE Trans. Signal Process.* **2019**, *67*, 4139–4151. [\[CrossRef\]](#)
- Plan, Y.; Vershynin, R. The generalized lasso with non-linear observations. *IEEE Trans. Inf. Theory* **2016**, *62*, 1528–1537. [\[CrossRef\]](#)
- Huang, J.; Jiao, Y.; Lu, X.; Zhu, L. Robust Decoding from 1-Bit Compressive Sampling with Ordinary and Regularized Least Squares. *SIAM J. Sci. Comput.* **2018**, *40*, A2062–A2086. [\[CrossRef\]](#)

16. Fan, Q.; Jia, C.; Liu, J.; Luo, Y. Robust recovery in 1-bit compressive sensing via L_q -constrained least squares. *Signal Process.* **2021**, *179*, 107822. [[CrossRef](#)]
17. Zhou, S.; Luo, Z.; Xiu, N.; Li, G.Y. Computing One-bit Compressive Sensing via Double-Sparsity Constrained Optimization. *IEEE Trans. Signal Process.* **2013**, *61*, 5777–5788. [[CrossRef](#)]
18. Hou, J.; Liu, X. 1-bit compressed sensing via an l_1 -tv regularization method. *IEEE Access* **2022**, *10*, 116473–116484. [[CrossRef](#)]
19. Zhong, Y.; Xu, C.; Zhang, B.; Hou, J.; Wang, J. One-bit compressed sensing via total variation minimization method. *Signal Process.* **2023**, *207*, 108939. [[CrossRef](#)]
20. Huang, J.; Jiao, Y.; Jin, B.; Liu, J.; Lu, X.; Yang, C. A unified primal dual active set algorithm for nonconvex sparse recovery. *Statist. Sci.* **2021**, *36*, 215–238. [[CrossRef](#)]
21. Shen, L.; Suter, B.W.; Tripp, E.E. Structured sparsity promoting functions. *J. Optim. Theory Appl.* **2019**, *183*, 386–421. [[CrossRef](#)]
22. Plan, Y.; Vershynin, R. One-Bit Compressed Sensing by Linear Programming. *Commun. Pure Appl. Math.* **2013**, *66*, 1275–1297. [[CrossRef](#)]
23. Hintermüller, M.; Ito, K.; Kunisch, K. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.* **2002**, *13*, 865–888. [[CrossRef](#)]
24. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [[CrossRef](#)]
25. Zhang, C.H.; Zhang, T. A general theory of concave regularization for high-dimensional sparse estimation problems. *Stat. Sci.* **2012**, *27*, 576–593. [[CrossRef](#)]
26. Ye, Y.; Zhang, C.H. Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **2010**, *11*, 3519–3540.
27. Gopi, S.; Netrapalli, P.; Jain, P.; Nori, A. One-bit compressed sensing: Provable support and vector recovery. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013; pp. 154–162.
28. Kim, Y.; Kwon, S.; Choi, H. Consistent model selection criteria on high dimensions. *J. Mach. Learn. Res.* **2012**, *13*, 1037–1057.
29. Plan, Y.; Vershynin, R.; Yudovina, E. High-dimensional estimation with geometric constraints. *Inf. Inference A J. IMA* **2016**, *6*, 1–40.
30. Vershynin, R. Estimation in high dimensions: A geometric perspective. In *Sampling Theory, a Renaissance*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 3–66.
31. Mallat, S. *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed.; AP Professional: London, UK, 2009.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.