Robust Single-bit Direction-of-Arrival Estimation with Magnitude Recovery

Xiao-Peng Li, *Member, IEEE*, Zhang-Lei Shi, Lei Huang, *Senior Member, IEEE*, Yu-Hang Xiao *Member, IEEE*, Hing Cheung So, *Fellow, IEEE*

***Abstract*—Single-bit direction-of-arrival (DoA) estimation utilizes the sign information of received signals to estimate the DoAs of interesting targets. This approach enables the use of single-bit analog-to-digital converters (ADCs) for quantization, reducing cost and hardware complexity significantly. However, the single-bit ADCs conceal the signal magnitude, leading to a limited dynamic range in radar systems. Moreover, the existing methods for single-bit DoA estimation typically require the prior knowledge of the target number or necessitate a computationally expensive grid search to determine an auxiliary parameter. In this work, we propose a new single-bit quantization model by incorporating a non-zero threshold for the received signal. Then, we employ the capped Frobenius norm and $\ell_{2,0}$-norm to formulate the single-bit DoA estimation problem, where the former measures the fitting error and the latter is used for estimating the target number. Subsequently, the resultant task is solved by combining proximal alternating minimization and forward-backward splitting. In our algorithm, an adaptive strategy is introduced to determine the regularization parameter associated with the $\ell_{2,0}$-norm, thus allowing for automatic estimation of the target number. Additionally, we analyze the algorithm convergence behavior, proving that both the objective function value and the variable sequence are convergent. Simulation results demonstrate that the devised method can estimate both the target number and signal magnitude with high probability. Furthermore, it outperforms state-of-the-art single-bit approaches in terms of DoA estimation accuracy.**

***Index Terms*—Robust algorithm, single bit, direction-of-arrival estimation, non-convex optimization.**

I. Introduction

Single-bit direction-of-arrival (DoA) estimation refers to the process in which an array radar system utilizes single-bit analog-to-digital converters (ADCs) to acquire the sign information of the received signals. Then, the DoAs of the signal sources are estimated based on this information [1]–[3]. The single-bit ADC can be considered as the simplest converter, which avoids high-precision quantization and reduces the number of data bits. This, in turn, decreases the burden on data storage, transmission, and processing. From the perspective of hardware design, a single-bit ADC is just a comparator, which dramatically increases the speed of data acquisition. Due to its simple system design, low implementation cost, and minimal resource requirements, single-bit DoA estimation is considered an effective way to simplify radar processing systems.

The frameworks for DoA estimation using single-bit measurements can be broadly classified into two categories, namely, subspace-based methods [4], [5] and sparse recovery-based algorithms [6]. The former first reconstructs the original (unquantized) covariance matrix using the single-bit covariance matrix. Then, the signal subspace is estimated from the reconstructed covariance matrix, after which algorithms such as multiple signal classification (MUSIC) [7], [8] are used to search for the DoAs. The latter utilizes the row-sparsity characteristic of the signal matrix in the spatial domain, and applies the sparse recovery theory to estimate the DoAs.

Bar-Shalom and Weiss are the first to propose reconstructing the original covariance matrix using the arcsine law and compensate for the performance loss caused by single-bit quantization [9]. They derive the Cramér-Rao bound (CRB) for single-bit DoA estimation based on two array elements and also analyze the accuracy limitation based on the bound. Later, Huang and Liao approximate the single-bit covariance matrix by a weighted sum of the original covariance matrix and the identity matrix [10]. They also demonstrate that the estimation performance is not significantly sensitive to the choice of the weighting coefficient. This finding implies that the signal subspace can be estimated directly using the single-bit covariance matrix. Guo *et al.* analyze the performance of single-bit DoA estimation in the presence of mutual coupling among receiving antennas, and propose a covariance matrix reconstruction algorithm with low computational complexity [11]. To further increase the accuracy of covariance matrix reconstruction, Liu and Lin employ single-bit ADC with a fixed nonzero threshold, achieving effective and consistent covariance matrix estimation [12]. Subsequently, Xiao *et al.* utilize the single-bit ADC with a variable threshold to further improve reconstruction performance, showing that the statistical property of the variable threshold is more conducive to reconstruct the original covariance matrix compared to the fixed threshold [13]. Beyond the studies on uniform linear array (ULA) mentioned above, several algorithms are proposed for other array configurations, including sparse array [14]–[16], uniform rectangular array [17], and coprime array [18].

Compared to subspace-based single-bit DoA estimation

This work was supported in part by the National Natural Science Foundation of China under Grant 62401373 and 62306337, in part by the Young Innovative Talents Projectof Guangdong Provincial Department of Education (Natural Science), China under Grant 2023KQNCX063, in part by the Shandong Provincial Natural Science Foundation under Grant ZR2024QF071, and in part by the Project of Shenzhen Peacock Plan Teams under Grant KQTD20210811090051046. (*Corresponding author: Lei Huang*)

Xiao-Peng Li, Lei Huang, and Yu-Hang Xiao are with the State Key Laboratory of Radio Frequency Heterogeneous Integration, Shenzhen University, Shenzhen, 518060, China. (e-mail: x.p.li@szu.edu.cn; lhuang@szu.edu.cn; yuhangxiao@szu.edu.cn).

Zhang-Lei Shi is with College of Science, China University of Petroleum (East China), Qingdao, 266580, China (e-mail: zlshi@upc.edu.cn).

Hing Cheung So is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China (e-mail: hcso@ee.cityu.edu.hk).

methods, sparse recovery-based approaches achieve superior performance in scenarios with few snapshots, low signal-to-noise ratio (SNR), and even in the presence of highly correlated or coherent signals [6]. To attain accurate estimation, the angular search interval typically needs to be discretized with small grid intervals. This results in a large number of candidate directions that is much larger than the true source number. Consequently, the signal matrix possesses a row-sparsity characteristic in the spatial domain. Stockle *et al.* propose using the complex-valued binary iterative hard thresholding (CBIHT) algorithm for single-bit DoA estimation problem [19]. Subsequently, Gao *et al.* establish a connection between single-bit DoA estimation and machine learning classification and then adopt the support vector machine (SVM) as the solver [20]. Meng and Zhu observe that the real and imaginary parts of the signal matrix exhibit joint sparsity [21]. Thus, they utilize generalized sparse Bayesian learning to address the estimation problem. To reduce the computational complexity, Chen *et al.* employ orthogonal matching pursuit (OMP) [22]. It is worth mentioning that the aforementioned algorithms require the prior knowledge of the source number. To eliminate the dependence on this information, Huang *et al.* [23] use the regularization theory to convert the source number constraint to an l_1-norm regularization term and subsequently relax the l_1-norm to the convex l_p-norm ($p > 1$). Following [23], Xiao *et al.* expand this l_1-norm-based approach into a neural network and train the network with a large dataset to determine its parameters. In doing so, both the estimation accuracy and speed of the single-bit DoA algorithm are enhanced in high SNR scenarios [24]. Additionally, Feng *et al.* adopt the improved maximum $a\ posteriori$ probability (MAP) criterion and an approximate function of the ℓ_0-norm to formulate the single-bit DoA estimation problem [25], which is then solved by the block successive upper-bound method. Chen *et al.* combine maximum likelihood (ML) estimation and sparse recovery techniques to further improve the performance of single-bit DoA estimation [26]. Li *et al.* propose a new quantization model and then design a robust method termed robust single-bit compressed sensing (ROSC) to improve estimation accuracy in the presence of noise [27].

It should be pointed out that the above sparse regularization-based methods do not provide a specific strategy for selecting the regularization parameter. Instead, they necessitate a grid search to determine an appropriate parameter, ensuring accurate estimation of the source number. On the other hand, the dynamic range determines the capacity of radar system to detect and differentiate signals across a wide range of power levels. However, existing approaches are unable to accurately estimate the signal magnitudes, resulting in a limited dynamic range in radar systems.

In this paper, we devise a robust single-bit DoA method to address the two existing issues. The signal after single-bit quantization is first reformulated using a non-zero threshold. Then, the sign function is approximated by a smooth hyperbolic tangent (tanh) function. Based on this new model, we formulate the single-bit DoA problem using the the capped Frobenius norm (CFN) and $\ell_{2,0}$-norm. Herein, the CFN is used to minimize the fitting error, while the $\ell_{2,0}$-norm is adopted as a row-sparsity penalty for determining the target number. To tackle the resultant non-convex and non-smooth problem, the half-quadratic optimization is employed to convert the intractable optimization into a manageable task, which is then tackled by the proximal alternating minimization (PAM) [28], [29] and forward–backward splitting (FBS) [30] methods. Additionally, a strategy is suggested to adaptively determine the penalty parameter associated with the $\ell_{2,0}$-norm regularizer, enabling automatic estimation of the target number. Furthermore, theoretical analysis of the proposed method is conducted, including computational complexity, convergence behaviors of the loss function value and the variable sequence.

Our main contributions are summarized as:

1) The quantized signal is re-expressed using a non-zero threshold and smooth tanh function, while the characteristic of the additive noise after quantization is analyzed. The new model facilitates the signal magnitude estimation and demonstrates the impulsive property of the noise.
2) Based on the devised model, the single-bit DoA problem is formulated using the CFN norm and $\ell_{2,0}$-norm. The former is employed to resist impulsive noise, while the latter is designed as a penalty term for determining the signal number. As a result, this approach does not require the prior information on the target number.
3) To estimate signal number using the spatial spectrum, the peaks corresponding to the estimated DoAs are considered as outliers relative to the other points. Although the proposed method contains a regularization parameter, an adaptive method is proposed to determine it, enabling automatic estimation of the target number. Besides, the convergence behavior of our algorithm is analyzed, demonstrating that the objective function value converges and the variable sequence is guaranteed to converge to a critical point.
4) Simulation results demonstrate that the proposed single-bit DoA method can estimate both the target number and signal magnitude. Furthermore, it outperforms conventional methods in terms of DoA estimation accuracy.

The rest of this paper is organized as follows. In Section II, notations are introduced, and relevant contemporary models are reviewed. Section III presents the proposed model and algorithm, along with the analysis of convergence behavior and computational complexity. In Section IV, simulations are conducted to evaluate the proposed method through comparison with state-of-the-art (SOTA) approaches. Finally, Section V provides concluding remarks.

II. Background

In this section, notation and signal model are presented, while relevant works are reviewed.

A. Notation

Italic, bold lower-case, and bold upper-case letters denote scalars, vectors, and matrices, respectively. The sign function is denoted by $\mathrm{sgn}(\cdot)$ and $\mathrm{csgn}(\cdot)$ for real and complex values, respectively, while the entry-wise tanh function is represented as $\tanh_c(x) = \tanh(cx) = (e^{cx} - e^{-cx})/(e^{cx} + e^{-cx})$

with c being a positive constant. For a column vector $\mathbf{a}$, its ℓ_0-norm is defined as $\|\mathbf{a}\|_0$, corresponding to the number of non-zero entries, and the operator $\mathrm{diag}(\mathbf{a})$ returns a square diagonal matrix $\mathbf{D} \in \mathbb{C}^{M\times M}$ with the elements of $\mathbf{a}$ on the main diagonal. Given a matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$, $\mathbf{A}^\top$ and $\mathbf{A}^\mathrm{H}$ are its transpose and Hermitian transpose, respectively. Additionally, its ith row and jth column are represented as $\mathbf{a}_{i,:}$ and $\mathbf{a}_{:,j}$, respectively. Besides, its Frobenius norm is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} a_{i,j}^2}$, while its $\ell_{2,0}$-norm is $\|\mathbf{A}\|_{2,0} = \|[\|\mathbf{a}_{1,:}\|_2, \|\mathbf{a}_{2,:}\|_2, \cdots, \|\mathbf{a}_{M,:}\|_2]^\mathrm{T}\|_0$, representing the row sparsity. Given $\mathbf{a} \in \mathbb{R}^M$, the operator $[\mathbf{a}]_-$ denotes the negative function, such that $[a_i]_- = a_i$ if $a_i < 0$ and $[a_i]_- = 0$ otherwise, while $[\mathbf{a}]_+ := [\max(a_1, 0), \max(a_2, 0), \ldots, \max(a_M, 0)]^\top$. Furthermore, $\mathbf{a} \geq 0$ indicates that all entries of $\mathbf{a}$ are non-negative. Given a scalar a, $|a|$ is its absolute value.

B. Signal Model

Consider a ULA equipped with M sensors, where inter-element spacing is d. It is essential that the spacing satisfies $d \leq \lambda/2$ to avert phase ambiguity, where λ is the wavelength of the incoming signal. Assume that $\bar{S}$ far-field, uncorrelated narrow-band signals impinge on the array from distinct directions $\bar{\boldsymbol{\theta}} = \{\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_{\bar{S}}\}$. The discrete-time complex baseband signal received by the mth sensor at time instant q is modeled as [10]

$$\bar{y}_m(q) = \sum_{s=1}^{S} x_s(q) e^{\mathrm{j}2\pi(m-1)\sin(\bar{\theta}_s)\frac{d}{\lambda}} + n_m(q), \tag{1}$$

where $x_s(q)$ denotes the sth source signal, $\mathrm{j} = \sqrt{-1}$ is the imaginary unit, and $n_m(q)$ represents the additive noise.

Furthermore, the output signal of the M sensors can be grouped into the vector form:

$$\bar{\mathbf{y}}_q = \bar{\mathbf{A}}\mathbf{x}_q + \mathbf{n}_q, \tag{2}$$

where $\bar{\mathbf{y}}_q = [\bar{y}_1(q), \ldots, \bar{y}_M(q)]^\top \in \mathbb{C}^M$ is the received signal vector, $\mathbf{x}_q = [x_1(q), \ldots, x_{\bar{S}}(q)]^\top \in \mathbb{C}^{\bar{S}}$ is the source vector, $\mathbf{n}_q = [n_1(q), \ldots, n_M(q)]^\top \in \mathbb{C}^M$ is the noise vector, and $\bar{\mathbf{A}} \in \mathbb{C}^{M\times\bar{S}}$ is the array manifold matrix, given by

$$\bar{\mathbf{A}} = [\mathbf{a}(\bar{\theta}_1), \mathbf{a}(\bar{\theta}_2), \ldots, \mathbf{a}(\bar{\theta}_{\bar{S}})], \tag{3}$$

where $\mathbf{a}(\bar{\theta}_s)$ is the steering vector, defined as

$$\mathbf{a}(\bar{\theta}_s) = \left[1, e^{\mathrm{j}2\pi\sin(\bar{\theta}_s)\frac{d}{\lambda}}, \ldots, e^{\mathrm{j}2\pi(M-1)\sin(\bar{\theta}_s)\frac{d}{\lambda}}\right]^\top. \tag{4}$$

Collecting Q snapshots, the received signal can be constructed as the matrix form:

$$\begin{aligned}\bar{\mathbf{Y}} &= [\bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_Q] \\ &= \bar{\mathbf{A}}\bar{\mathbf{X}} + \mathbf{N},\end{aligned} \tag{5}$$

where $\bar{\mathbf{X}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_Q] \in \mathbb{C}^{\bar{S}\times Q}$ and $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_Q] \in \mathbb{C}^{M\times Q}$. Then, the quantized signal with single-bit ADC can be expressed as

$$\mathbf{Y} = \mathrm{csgn}(\bar{\mathbf{Y}}) = \mathrm{sgn}(\Re(\bar{\mathbf{Y}})) + \mathrm{jsgn}(\Im(\bar{\mathbf{Y}})), \tag{6}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary parts of a complex number.

C. Related Work

To estimate the DoAs of signal sources based on $\mathbf{Y}$, a popular approach is subspace estimation. Huang and Liao [10] have shown that the single-bit covariance matrix can be approximated as a combination of a scaled unquantized covariance matrix and a scaled identity matrix under mild conditions. This allows MUSIC [7] to be directly applied for single-bit DoA estimation. Specifically, the covariance matrix of the quantized received signal is given by $\widehat{\mathbf{C}}_x = \frac{1}{Q}\mathbf{Y}\mathbf{Y}^\mathrm{H}$. Since the signal and noise are uncorrelated, the covariance matrix of the received signal is given by:

$$\mathbf{C}_x = \mathbf{U}_s\boldsymbol{\Lambda}_s\mathbf{U}_s^\mathrm{H} + \sigma^2\mathbf{U}_n\mathbf{U}_n^\mathrm{H}, \tag{7}$$

where $\mathbf{U}_s \in \mathbb{C}^{M\times\bar{S}}$ is the signal subspace, $\boldsymbol{\Lambda}_s \in \mathbb{C}^{\bar{S}\times\bar{S}}$ is a diagonal matrix containing the corresponding eigenvalues sorted in descending order, and $\mathbf{U}_n \in \mathbb{C}^{M\times(Q-\bar{S})}$ is called the noise subspace. Subsequently, the spatial spectrum is constructed as:

$$P(\theta) = \frac{1}{\mathbf{a}^\mathrm{H}(\theta)\mathbf{U}_n\mathbf{U}_n^\mathrm{H}\mathbf{a}(\theta)}. \tag{8}$$

Finally, the DoA estimates can be obtained via finding the spectrum peaks.

On the other hand, sparse recovery can be used for single-bit DoA estimation by availing the sparsity property of single-bit signals in the spatial domain. First, it requires discretizing the potential DoA range (i.e. $(-90°, 90°)$) into S grids:

$$\bar{\theta} \subset \theta = \{\theta_1, \theta_2, \ldots, \theta_S\}, \tag{9}$$

where $S \gg \bar{S}$. Then, the quantized signal is re-expressed as

$$\mathbf{Y} = \mathrm{csgn}(\mathbf{A}\mathbf{X} + \mathbf{N}), \tag{10}$$

where $\mathbf{A} \in \mathbb{C}^{M\times S}$ is the extended array manifold, defined as

$$\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2)\ldots, \mathbf{a}(\theta_S)], \tag{11}$$

and $\mathbf{X} \in \mathbb{C}^{S\times Q}$ is the extended signal matrix, which is a row-sparse matrix. This means that $\mathbf{X}$ includes lots of zero rows. Specifically, the ith row of $\mathbf{X}$ equals the jth row of $\bar{\mathbf{X}}$ if $\theta_i = \bar{\theta}_j$. Otherwise, the ith row of $\mathbf{X}$ is zero.

To seek for $\bar{\theta}$, Stöckle *et al.* [19] formulate the single-bit DoA estimation as the following constrained optimization problem:

$$\begin{aligned}&\min_{\mathbf{X}} \left\| \left[\begin{bmatrix}\Re(\mathbf{Y})\\ \Im(\mathbf{Y})\end{bmatrix} \odot \begin{bmatrix}\Re(\mathbf{A}\mathbf{X})\\ \Im(\mathbf{A}\mathbf{X})\end{bmatrix} \right]_- \right\|_F^2, \\ &\text{s.t.} \quad \|\mathbf{X}\|_{2,0} \leq \bar{S}, \|\mathbf{x}(i,:)\|_2 = 1,\end{aligned} \tag{12}$$

where $\odot$ denotes the the element-wise product. Then, the binary iterative hard thresholding approach (BIHT) [31] is adopted as the solver.

To improve the performance in noisy scenarios, Li *et al.* formulate single-bit DoA estimation as [27]

$$\begin{aligned}&\min_{\mathbf{X},\mathbf{O}_r} \left\| \begin{bmatrix}\Re(\mathbf{Y})\\ \Im(\mathbf{Y})\end{bmatrix} - \tanh_\mathrm{c}\left(\begin{bmatrix}\Re(\mathbf{A}\mathbf{X})\\ \Im(\mathbf{A}\mathbf{X})\end{bmatrix} \right) - \mathbf{O}_r \right\|_F^2 + \xi\|\mathbf{O}_r\|_0, \\ &\text{s.t.} \ \|\mathbf{X}\|_{2,0} \leq \bar{S},\end{aligned} \tag{13}$$

where $\mathbf{O}_r \in \mathbb{R}^{2M\times Q}$ and $\xi > 0$ is to control the sparsity of $\mathbf{O}_r$. Nevertheless, (12) and (13) require the prior information of target number.

Furthermore, Chen *et al.* combine the sparse recovery and maximum likelihood techniques to formulate the task as [26]

$$\min_{\mathbf{X}} \mathcal{L}(\mathbf{X}) + \tau\|\mathbf{X}\|_{2,1}, \tag{14}$$

where $\tau > 0$ is a penalty parameter to control the row-sparsity of $\mathbf{X}$ and $\mathcal{L}(\mathbf{X})$ is the likelihood term, defined as

$$\begin{aligned}\mathcal{L}(\mathbf{X}) = \sum_{i=1}^{M}\sum_{j=1}^{Q} &- \ln\left(\Phi\left(\Re(y_{i,j})\frac{\sqrt{2}\Re(\mathbf{a}_{i,:}\mathbf{x}_{:,j})}{\sigma}\right)\right) \\ &- \ln\left(\Phi\left(\Im(y_{i,j})\frac{\sqrt{2}\Im(\mathbf{a}_{i,:}\mathbf{x}_{:,j})}{\sigma}\right)\right).\end{aligned} \tag{15}$$

Here, $\Phi(x) = 1/\sqrt{2\pi}\int_{-\infty}^{x} e^{-t^2/2}\mathrm{d}t$. It is worth mentioning that $\|\cdot\|_{2,1}$ is a convex approximation of $\|\mathbf{X}\|_{2,0}$, generating an approximately row-sparse solution. Here, the approximate row-sparsity means that a row has non-zero components with extremely small magnitudes. In practice, the choice of τ requires a time-consuming grid search to ensure accurate estimation.

D. Capped Frobenius Norm

In (6), the real part or imaginary part of complex-valued $\mathbf{Y}$ might yield incorrect quantization due to the presence of additive noise. The truncated least squares loss (TLS) function [32] can be adopted as a robust measurement metric, which is defined as:

$$\begin{aligned}\|\mathbf{E}\|_{\mathrm{TLS}} &= \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\min(|e_{i,j}|^2, \epsilon^2)} \\ &= \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\min(\Re(e_{i,j})^2 + \Im(e_{i,j})^2, \epsilon^2)},\end{aligned} \tag{16}$$

where $\epsilon > 0$ is a cap threshold. However, even when only the real part of $e_{i,j}$ is affected by an outlier, the TLS still considers that the imaginary part is also corrupted by impulsive noise. As a result, the TLS function may not accurately calculate the error for complex values.

To address this limitation, we introduce CFN for handling complex-valued data, defined as

$$\begin{aligned}&\|\mathbf{E}\|_{\mathrm{CF}} \\ &= \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\Big(\min(\Re(e_{i,j})^2, \epsilon^2) + \min(\Im(e_{i,j})^2, \epsilon^2)\Big)}.\end{aligned} \tag{17}$$

The CFN applies a threshold separately to both the real and imaginary components, making it a more accurate metric compared to the TLS. Additionally, as $\epsilon \to +\infty$, the CFN is equivalent to the Frobenius norm.

E. Half-Quadratic Optimization

The CFN is non-convex, which makes the CFN based optimization problem non-convex. Thereby, we introduce the half-quadratic optimization technique to address the CFN-based optimization task.

The half-quadratic optimization simplifies a non-convex problem by transforming it into a series of convex subproblems [33]. Given a function $\phi(\cdot)$, it establishes the following relationship [34]:

$$\phi(x) = \inf_{y} Q(x, y) + \varphi(y), \tag{18}$$

where $Q(\cdot,\cdot)$ is a half-quadratic function and $\varphi(\cdot)$ is the dual potential function of $\phi(\cdot)$. Furthermore, if $\phi(\cdot)$ makes $x^2/2 - \phi(x)$ convex, it can be further re-expressed as:

$$\phi(x) = \inf_{y} \frac{(x-y)^2}{2} + \varphi(y), \tag{19}$$

where $\varphi(y)$ makes $y^2/2 + \varphi(y)$ convex. The above equation corresponds to the additive form of half-quadratic optimization. That is, for non-convex $\phi(\cdot)$, the half-quadratic minimization converts it to a convex function with respect to (w.r.t.) x. For example, He *et al.* [34] exploit the half-quadratic minimization to transform the non-convex Huber function, given by

$$\phi_H(x) = \begin{cases}\frac{x^2}{2}, & |x| \le \lambda, \\ \lambda|x| - \frac{x^2}{2}, & |x| > \lambda,\end{cases} \tag{20}$$

into the following form:

$$\phi_H(x) = \min_{y} \frac{(x-y)^2}{2} + \varphi_H(y), \tag{21}$$

where $\varphi_H(y)$ is the associated regularizer and its expression is typically unknown.

As the CFN is exploited to formulate our estimation problem, we show its half-quadratic expression:

$$\min_{\mathbf{Y}} \|\mathbf{X} - \mathbf{Y}\|_F^2 + \varphi_{\mathrm{CF}}(\mathbf{Y}), \tag{22}$$

where $\varphi_{\mathrm{CF}}(\mathbf{Y})$ is an element-wise function, defined as

$$\varphi_{\mathrm{CF}}(\mathbf{Y}) = \begin{cases}-(\epsilon - |y_{i,j}|)^2 + \epsilon^2, & |y_{i,j}| < \epsilon, \\ \epsilon^2, & |y_{i,j}| \ge \epsilon.\end{cases} \tag{23}$$

The detailed derivation is provided in Section III.

III. Algorithm Development

In this section, we first present a new quantization model for single-bit DoA estimation and then devise an effective algorithm with convergence guarantee.

A. Proposed Model

Signal magnitude is important information in array radar systems to distinguish strong and weak targets. However, the single-bit quantization conceals the signal magnitude, resulting in a minimal dynamic range.

To increase the dynamic range of the single-bit DoA radar, we introduce a non-zero threshold for single-bit quantization, leading to the following expression:

$$\mathbf{Y} = \mathrm{csgn}(\mathbf{AX} + \mathbf{N} + \mathbf{T}), \tag{24}$$

where $\mathbf{T} \in \mathbb{R}^{M\times Q}$ is the known threshold for quantization.

In general, determining a proper loss function requires the prior knowledge of the noise distribution. For example, the Frobenius norm is optimal for additive Gaussian white noise, while the ℓ_1-norm is more suitable for additive Laplacian noise. However, single-bit quantization obscures the noise distribution, making it challenging to select an appropriate loss function to ensure reliable estimation performance. To handle this issue, we further modify (24) as

$$\mathbf{Y} = \mathrm{csgn}(\mathbf{AX} + \mathbf{T}) + \mathbf{O}, \tag{25}$$

where $\mathbf{O} \in \mathbb{C}^{M\times Q}$ with $\Re(o_{i,j}) \in \{-2, 0, 2\}$ and $\Im(o_{i,j}) \in \{-2, 0, 2\}$. It is observed that (25) shifts the noise component from inside the sign function to the outside. Moreover, this model implies that, regardless of the distribution of $\mathbf{N}$, $\mathbf{O}$ exhibits the same structure, in which the non-zero elements are either 2 or -2. Different distributions or intensities of $\mathbf{N}$ only affect the number of non-zero entries in $\mathbf{O}$.

Furthermore, as the sign function is discontinuous and not lower semicontinuous, it brings challenges to analyze the convergence behavior of the resultant algorithm. To address this issue, we approximate the sign function using the tanh function, which is continuous and definable in an o-minimal structure. This smoothness property allows us to establish the convergence guarantee for our method. Thereby, the single-bit quantization model is reformulated as

$$\mathbf{Y} = \mathrm{ctanh_c}(\mathbf{AX} + \mathbf{T}) + \widetilde{\mathbf{N}} + \mathbf{O}, \tag{26}$$

where $\widetilde{\mathbf{N}} \in \mathbb{C}^{M\times Q}$ is the approximation error and $\mathrm{ctanh_c}(a) = \tanh_{\mathrm{c}}(\Re(a)) + \mathrm{j}\tanh_{\mathrm{c}}(\Im(a))$. Apparently, as $c \to +\infty$, $\tanh_c(\cdot)$ becomes $\mathrm{sgn}(\cdot)$. In practice, while c cannot be infinite, it can be chosen to be sufficiently large to reduce the approximation error. Consequently, the magnitude of $\widetilde{\mathbf{N}}$ is much smaller than that of $\mathbf{O}$. Additionally, $\widetilde{\mathbf{N}} + \mathbf{O}$ can be considered as an additive noise, where $\widetilde{\mathbf{N}}$ is dense with low power, whereas $\mathbf{O}$ is sparse with high power. Thereby, $\mathbf{O}$ can be treated as the outlier component in the mixture of $\widetilde{\mathbf{N}}$ and $\mathbf{O}$. Thereby, robust signal processing techniques can be employed for single-bit DoA estimation, including Huber function [35], ℓ_p-norm ($0 < p < 2$) [36], and Welsch function [37], among others.

In this work, we propose utilizing CFN to mitigate the effect of $\widetilde{\mathbf{N}} + \mathbf{O}$ and thus formulating the single-bit DoA estimation problem as

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathrm{ctanh_c}(\mathbf{AX} + \mathbf{T})\|_{\mathrm{CF}}^2 + \tau\|\mathbf{X}\|_{2,0}, \tag{27}$$

where $\tau > 0$ is a regularization parameter that controls the row sparsity of $\mathbf{X}$. For CFN, its auxiliary parameter ϵ enables it to resist the outliers in the mixed noise. In addition, the regularization term of $\|\mathbf{X}\|_{2,0}$ in (27) eliminates the need for the prior information about the target number. However, the regularization introduces a challenge, as the estimation performance heavily depends on the choice of τ. In the next subsection, we will present a strategy to automatically update τ to achieve satisfactory performance in different noise environments. Furthermore, compared with (14), $\ell_{2,0}$-norm results in a row-sparse solution, indicating that the peak number in the spatial spectrum corresponds to the target number.

B. Proposed Algorithm

For (27), directly solving it does not allow us to determine the magnitude of $\mathbf{X}$. This is because the non-zero threshold provides prior information regarding the magnitude of $\mathbf{AX}$. To address this, we recast the quantized signal as an augmented form:

$$\begin{aligned}\mathbf{Y} &= \mathrm{ctanh_c}(\mathbf{AX} + \mathbf{T}) + \widetilde{\mathbf{N}} + \mathbf{O} \\ &= \mathrm{ctanh_c}(\mathbf{A}_a\mathbf{X}_a) + \widetilde{\mathbf{N}} + \mathbf{O},\end{aligned} \tag{28}$$

where $\mathbf{A}_a = [\mathbf{A}, \mathbf{T}] \in \mathbb{C}^{M\times(S+Q)}$ and $\mathbf{X}_a = [\mathbf{X}^\top, \mathbf{I}]^\top \in \mathbb{C}^{(S+Q)\times Q}$ with $\mathbf{I}$ being an identity matrix. Since $\mathbf{I}$ serves as the prior information for $\mathbf{X}_a$, the magnitude of $\mathbf{X}_a$ can be estimated after $\mathbf{X}_a$ is obtained. Then, the single-bit DoA estimation task is re-expressed as

$$\begin{aligned}\min_{\mathbf{X}_a} \|\mathbf{Y} - \mathrm{ctanh_c}(\mathbf{A}_a\mathbf{X}_a)\|_{\mathrm{CF}}^2 + \tau\|\mathbf{X}\|_{2,0} \\ \text{s.t. } \mathbf{X}_a(1:S,:) = \mathbf{X}.\end{aligned} \tag{29}$$

Given that CFN is non-convex and non-smooth, we employ the half-quadratic technique to convert (29) into a tractable form:

$$\begin{aligned}\min_{\mathbf{X}_a,\mathbf{O}} \|\mathbf{Y} - \mathrm{ctanh_c}(\mathbf{A}_a\mathbf{X}_a) - \mathbf{O}\|_F^2 + \varphi_{\mathrm{CF}}(\mathbf{O}) + \tau\|\mathbf{X}\|_{2,0} \\ \text{s.t. } \mathbf{X}_a(1:S,:) = \mathbf{X},\end{aligned} \tag{30}$$

where $\varphi_{\mathrm{CF}}(\cdot)$ is an element-wise function that operates independently on the real and imaginary components. The reason is the same to that of the CFN.

We utilize the Legendre-Fenchel transform to derive the exact expression for $\varphi_{\mathrm{CF}}(\cdot)$. The conjugate function of $x^2 - \min(x^2, \epsilon^2)$ is given by:

$$f^*(y) = \sup_x xy - (x^2 - \min(x^2, \epsilon^2)) \tag{31a}$$

$$= \sup_x(-(x-y)^2 + \min(x^2, \epsilon^2)) + y^2 \tag{31b}$$

$$= \varphi(y) + y^2. \tag{31c}$$

Besides, the conjugate function of $f^*(y)$ is

$$f^{**}(x) = \sup_y xy - f^*(y) \tag{32a}$$

$$= \sup_y -(x-y)^2 - \varphi(y) + x^2. \tag{32b}$$

Since $x^2 - \min(x^2, \epsilon^2)$ is convex, the following equality holds:

$$f^{**}(x) = x^2 - \min(x^2, \epsilon^2). \tag{33}$$

Combing (32b) and (33), we have

$$\min(x^2, \epsilon^2) = \inf_y (x-y)^2 + \varphi(y) \tag{34a}$$

$$= \min_y (x-y)^2 + \varphi(y). \tag{34b}$$

Therefore, $\varphi_{\mathrm{CF}}(\cdot)$ for a real value is defined as

$$\varphi_{\mathrm{CF}}(y) = \sup_x(-(x-y)^2+\min(x^2,\epsilon^2)) \tag{35a}$$

$$= \begin{cases} \sup\limits_x\ (-(x-y)^2+x^2), & |x|<\epsilon, \\ \sup\limits_x\ (-(x-y)^2+\epsilon^2), & |x|\geq\epsilon, \end{cases}$$

$$= \begin{cases} -(\epsilon-|y|)^2+\epsilon^2, & |y|<\epsilon, \\ \epsilon^2, & |y|\geq\epsilon. \end{cases} \tag{35b}$$

Next, directly solving (30) may not result in real-valued matrix $\mathbf{X}_a(S+1:S+Q,:)$, which causes the solution to deviate from the ground truth. A common approach is to convert the complex-valued matrices to real-valued ones and then handle the problem as a real-valued optimization. However, this conversion increases the dimensions of the target variables, leading to higher computational cost.

In this work, we propose a two-stage solution to tackle (30), involving DoA estimation and magnitude estimation. In the former, we solve (30) in its complex-valued form, while, in the latter, we first reduce the dimensions of target variables and then handle the optimization problem as a real-valued one.

1) DoA Estimation:

As (30) involves two target variables, we employ the PAM concept [28], [29] to address it, leading to the following iterative procedure:

$$\mathbf{X}_a^k = \arg\min_{\mathbf{X}_a} \|\mathbf{Y}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}^{k-1}\|_F^2 + \tau\|\mathbf{X}_a(1:S,:)\|_{2,0}+\rho\|\mathbf{X}_a-\mathbf{X}_a^{k-1}\|_F^2, \tag{36a}$$

$$\mathbf{O}^k = \arg\min_{\mathbf{O}} \|\mathbf{Y}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a^k)-\mathbf{O}\|_F^2 + \varphi_{\mathrm{CF}}(\mathbf{O})+\rho\|\mathbf{O}-\mathbf{O}^{k-1}\|_F^2, \tag{36b}$$

where $\rho>0$ is a pre-defined proximal parameter and the last terms in (36a) and (36b) represent proximal regularization. It is worth mentioning that ρ is used to ensure convergence and can be set to a very small value. When $\rho=0$, PAM reduces to the general alternating minimization approach [38]. Note that (36a) does not include the equality constraint from (29), as this constraint is converted into a regularization term, consistent with (27). Specifically, the row-sparsity constraint applies only to the first S rows for DoA estimation, while the remaining rows are unrestricted.

As there is no closed-form solution to (36a), we adopt FBS [30] as the solver, leading to the following iterative update:

$$\mathbf{Z} = (\mathbf{X}_a)_{p-1}^{k-1} - \eta\nabla g((\mathbf{X}_a)_{p-1}^{k-1}, \mathbf{O}^{k-1}), \tag{37a}$$

$$(\mathbf{X}_a)_p^{k-1}(i,:) = \begin{cases} \mathbf{Z}(i,:), & \text{if } \|\mathbf{Z}(i,:)\|_2 > \sqrt{\tau} \\ 0, & \text{otherwise} \end{cases} \text{ for } i\in[1,S], \tag{37b}$$

$$(\mathbf{X}_a)_p^{k-1}(i,:) = \mathbf{Z}(i,:),\ \text{for } i\in[S+1,S+Q], \tag{37c}$$

where $\nabla g(\mathbf{X}_a,\mathbf{O}^{k-1})$ is the gradient of $\|\mathbf{Y}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}^{k-1}\|_F^2+\rho\|\mathbf{X}_a-\mathbf{X}_a^{k-1}\|_F^2$ w.r.t. $\mathbf{X}_a$, given by:

$$-2c\mathbf{A}_a^{\mathrm{H}}\Big((\mathbf{S}^{k-1}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a))\odot(\mathbf{1}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)^2)\Big) + 2\rho(\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}), \tag{38}$$

where $\mathbf{S}^{k-1}=\mathbf{Y}-\mathbf{O}^{k-1}$. Herein, superscript $(\cdot)^k$ indicates the iteration number in the PAM process, while subscript $(\cdot)_p$ refers to the pth iteration in the FBS procedure. For the initialization of $(\mathbf{X}_a)_0^{k-1}$, we set $(\mathbf{X}_a)_0^{k-1}=(\mathbf{X}_a)^{k-1}$.

To derive the solution to (36b), we first re-express (36b) as:

$$\mathbf{O}^k = \arg\min_{\mathbf{O}}\|\mathbf{E}^k-\mathbf{O}\|_F^2+\varphi_{\mathrm{CF}}(\mathbf{O})+\rho\|\mathbf{O}-\mathbf{O}^{k-1}\|_F^2, \tag{39}$$

where $\mathbf{E}^k=\mathbf{Y}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a^k)$. Notably, each element $o_{i,j}^k$ depends only on $e_{i,j}^k$ and $o_{i,j}^{k-1}$, allowing (39) to be solved in scalar form:

$$o^k = \arg\min_o (e^k-o)^2+\varphi_{\mathrm{CF}}(o)+\rho(o-o^{k-1})^2, \tag{40}$$

where subscript $(\cdot)_{i,j}$ is omitted for simplicity.

As $\varphi_{\mathrm{CF}}(\cdot)$ operates separately on the real and imaginary parts, (40) can be split into the following two subproblems:

$$\Re(o^k) = \arg\min_{\Re(o)}(\Re(e^k-o))^2+\varphi_{\mathrm{CF}}(\Re(o)), \tag{41a}$$

$$\Im(o^k) = \arg\min_{\Im(o)}(\Im(e^k-o))^2+\varphi_{\mathrm{CF}}(\Im(o)). \tag{41b}$$

An optimal solution to (41a) is derived as (see Appendix A):

$$\Re(o^k)=\begin{cases}\Re(e^k), & |\Re(e^k)|>\epsilon,\\ 0, & |\Re(e^k)|\leq\epsilon.\end{cases} \tag{42}$$

Similarly, $\Im(o^k)$ can be determined. Subsequently, $\mathbf{O}^k$ is obtained

$$\mathbf{O}^k(i,j) = \Re(o_{i,j}^k)+\mathrm{j}\Im(o_{i,j}^k). \tag{43}$$

After K iterations, we obtain the solution of $\mathbf{X}_a^K$ and then calculate the spatial spectrum as follows:

$$\mathbf{X}^K = \mathbf{X}_a^K(1:S,:), \tag{44}$$

$$\mathbf{p}(i) = \|\mathbf{X}_a^K(i,:)\|_2,\ \text{for } i\in[1,S]. \tag{45}$$

Finally, the DoAs of the source signals can be determined by identifying the peaks in $\mathbf{p}$. The locations of these peaks indicate the directions from which the signals originate, and the number of peaks corresponds to the estimated source number.

It is important to note that the row-sparsity level of $\mathbf{X}^K$ may not exactly match the target number. This is because, in the spatial spectrum, the regions adjacent to a peak are typically non-zero. Additionally, when the row-sparsity is formulated as a constraint, as shown in (12), the upper bound for $\|\mathbf{X}\|_{2,0}$ should be set larger than the actual target number.

2) Magnitude Estimation:

Let the estimated DoAs be $\widetilde{\theta}=\{\widetilde{\theta}_1,\widetilde{\theta}_2,\cdots,\widetilde{\theta}_{\widetilde{S}}\}$. Then, we extract columns from $\mathbf{A}$ with $\theta_i=\widetilde{\theta}_j$ to construct $\mathbf{A}_c\in\mathbb{C}^{M\times\widetilde{S}}$. Similarly, we extract the corresponding rows from $\mathbf{X}$ to yield $\mathbf{X}_c\in\mathbb{C}^{\widetilde{S}\times Q}$. This leads to:

$$\mathbf{Y} = \mathrm{ctanh}_{\mathrm{c}}(\mathbf{A}_c\mathbf{X}_c)+\widetilde{\mathbf{N}}+\mathbf{O}. \tag{46}$$

Furthermore, (46) is rewritten in real-valued form as:

$$\mathbf{Y}_r = \mathrm{ctanh}_{\mathrm{c}}(\mathbf{A}_r\mathbf{X}_r)+\widetilde{\mathbf{N}}_r+\mathbf{O}_r, \tag{47}$$

where

$$\mathbf{Y}_r = [\Re(\mathbf{Y}); \Im(\mathbf{Y})], \tag{48a}$$

$$\mathbf{A}_r = [\Re(\mathbf{A}_c), -\Im(\mathbf{A}_c), \Re(\mathbf{T}); \Im(\mathbf{A}_c), \Re(\mathbf{A}_c), \Im(\mathbf{T})], \tag{48b}$$

$$\mathbf{X}_r = [\Re(\mathbf{X}_c); \Im(\mathbf{X}_c); \mathbf{I}], \tag{48c}$$

$$\mathbf{N}_r = [\Re(\mathbf{N}); \Im(\mathbf{N})], \tag{48d}$$

$$\mathbf{O}_r = [\Re(\mathbf{O}); \Im(\mathbf{O})]. \tag{48e}$$

Subsequently, the magnitude estimation task is formulated as:

$$\min_{\mathbf{X}_r, \mathbf{O}_r} \|\mathbf{Y}_r - \mathrm{ctanh_c}(\mathbf{A}_r\mathbf{X}_r) - \mathbf{O}_r\|_F^2 + \varphi_{CF}(\mathbf{O}_r) \quad \text{s.t. } \mathbf{X}_r \in \mathrm{C}, \tag{49}$$

where C is a convex set, defined as

$$(\mathbf{X}_r)_{i,j} = 0, \text{ if } i \neq j, \quad i, j \in [2\widetilde{S} + 1, 2\widetilde{S} + Q], \tag{50}$$

indicating that $\mathbf{X}_r(2\widetilde{S} + 1 : 2\widetilde{S} + Q, :)$ is a diagonal matrix.

As (49) is similar to (30), we adopt PAM as the solver, resulting in the following iterative procedure:

$$\mathbf{X}_r^k = \arg\min_{\mathbf{X}_r \in \mathrm{C}} \|\mathbf{Y}_r - \mathrm{ctanh}_c(\mathbf{A}_r\mathbf{X}_r) - \mathbf{O}_r^{k-1}\|_F^2 + \rho\|\mathbf{X}_r - \mathbf{X}_r^{k-1}\|_F^2, \tag{51a}$$

$$\mathbf{O}_r^k = \arg\min_{\mathbf{O}_r} \|\mathbf{Y}_r - \mathrm{ctanh}_c(\mathbf{A}_r\mathbf{X}_r^k) - \mathbf{O}_r\|_F^2 + \varphi_{\mathrm{CF}}(\mathbf{O}_r) + \rho\|\mathbf{O}_r - \mathbf{O}_r^{k-1}\|_F^2, \tag{51b}$$

Similar to the derivation procedure for solving (36a), we employ FBS to find a solution to (51a), leading to:

$$\mathbf{Z} = (\mathbf{X}_r)_{p-1}^{k-1} - \eta\nabla h((\mathbf{X}_r)_{p-1}^{k-1}, \mathbf{O}^{k-1}), \tag{52a}$$

$$(\mathbf{X}_r)_p^{k-1}(i,j) = \begin{cases} 0, & \text{if } i \neq j, \\ & i, j \in [2\widetilde{S} + 1, 2\widetilde{S} + Q], \\ \mathbf{Z}(i,j), & \text{otherwise}, \end{cases} \tag{52b}$$

where $\nabla h((\mathbf{X}_r)_{p-1}^{k-1}, \mathbf{O}_r^{k-1})$ is the gradient of $h(\mathbf{X}_r, \mathbf{O}_r^{k-1})$ w.r.t. $\mathbf{X}_r$, which is similar to (38). It is seen that the procedure for handling (51a) follows the same pattern as (36a), involving a gradient descent step followed by projection. Both projection operations involve setting specific entries to zero while retaining the remaining elements.

For (51b), it is analogous to (36b) and can thus be addressed using (43).

After obtaining $\mathbf{X}_r$, we normalize the portion of $\mathbf{X}_r(2\widetilde{S}+1 : 2\widetilde{S} + Q, :)$ to form an identity matrix, which allows for the magnitude estimation. The complex-valued signal can then be reconstructed as:

$$\mathbf{X}_c = \mathbf{X}_r(1 : \widetilde{S}, :) + \mathrm{j}\mathbf{X}_r(\widetilde{S} + 1 : 2\widetilde{S}, :). \tag{53}$$

The proposed approach is termed **ro**bust **s**ingle-bit (RoS) DoA, and its steps are summarized in Algorithm 1. It has two stopping criteria. One is to reach the maximum iteration number $K_{\max}$. Our experiments suggest that $K_{\max} = 20$ is sufficient to ensure convergence. In addition, the algorithm will terminate when it satisfies the following condition

$$\frac{\|(\boldsymbol{X}_a)^k - (\boldsymbol{X}_a)^{k-1}\|_F^2}{\|(\boldsymbol{X}_a)^k\|_F^2} < 10^{-6}. \tag{54}$$

Algorithm 1 RoS

Input: $\mathbf{Y} \in \mathbb{C}^M$, $\mathbf{A} \in \mathbb{C}^{M \times N}$, $\rho = 10^{-8}$, $\eta = 10^{-5}$, $\epsilon = 1.9$, $c = 10^4$, $P_{\max} = 50$, and $K_{\max} = 20$
Initialize: Randomize $\mathbf{X}^1 = \mathbf{A}^{\mathrm{H}}\mathbf{Y}$ and $\mathbf{O}^0 = \mathbf{0}$
(1) DoA Estimation
for $k = 1, 2, \ldots, K_{\max}$ **do**
  1) Determine τ
  2) $(\mathbf{X}_a)_p^{k-1} = (\mathbf{X}_a)^{k-1}$
  for $p = 1, \ldots, P_{\max}$ **do**
    3) Update $(\mathbf{X}_a)_p^{k-1}$ via (37a)-(37c)
  end for
  4) $(\mathbf{X}_a)^k = (\mathbf{X}_a)_p^{k-1}$
  5) Update $\mathbf{O}^k$ via (43)
  Stop if stopping criterion is met.
end for
(2) Magnitude Estimation
for $k = 1, 2, \ldots, K_{\max}$ **do**
  1) $(\mathbf{X}_r)_p^{k-1} = (\mathbf{X}_r)^{k-1}$
  for $p = 1, \ldots, P_{\max}$ **do**
    2) Update $(\mathbf{X}_r)_p^{k-1}$ via (52a) and (52b)
  end for
  3) $(\mathbf{X}_r)^k = (\mathbf{X}_r)_p^{k-1}$
  4) Update $\mathbf{O}_r^k$ via (43)
  Stop if stopping criterion is met.
end for
$\mathbf{X}_c = \mathbf{X}_r(1 : \widetilde{S} :) + \mathrm{j}\mathbf{X}_r(\widetilde{S} + 1 : 2\widetilde{S}, :)$,
Output: $\widetilde{\boldsymbol{\theta}}$ and $\mathbf{X}_c$

C. Parameter Selection

1) Selection of τ:

Prior to analyzing how to determine an appropriate value for τ, we first provide a perspective on its selection. In the spatial spectrum, the peaks corresponding to the estimated DoAs can be considered as outliers relative to the other points. As illustrated in Fig. 1, the power of the target signals is much higher than that of the remaining points. By choosing an appropriate threshold (e.g., the red line in the figure), the correct peaks can be effectively identified. From (37b), we know that $\sqrt{\tau}$ serves as this threshold. To ensure satisfactory performance, we use the normalized median absolute deviation (MAD) [35] to adaptively determine τ through the following steps:

1. Compute the vector $\mathbf{z} = [\|\mathbf{z}_{1,:}\|_2, \|\mathbf{z}_{2,:}\|_2, \cdots, \|\mathbf{z}_{S,:}\|_2]^\top$.
2. Remove top 1% largest values from $\mathbf{z}$ and compute the mean, denoted as μ.
3. Calculate the standard deviation of $\mathbf{z}$ using $\sigma = 1.4826 \times \mathrm{Med}(|\mathbf{z} - \mathrm{Med}(\mathbf{z})|)$ [39], where 1.4826 is determined according to the MAD of a standard normal random variable satisfying $1/1.4826 \approx 0.6745$ and $\mathrm{Med}(\cdot)$ is the operator to calculate the median value [39].

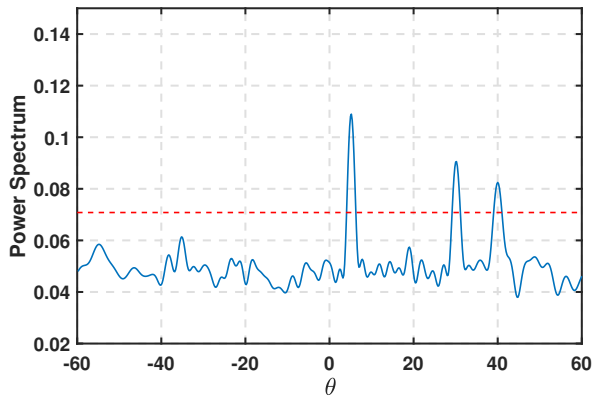

Fig. 1: Spatial spectrum of three signals.

4. Set $\tau = (\mu+\zeta\sigma)^2$ with ζ being a predefined constant value. That is, the appropriate value of τ depends on ζ. In the next section, we demonstrate that the estimation performance is not sensitive to the choice of ζ.

2) Selection of ϵ:

We know that $\mathbf{E}^k = \mathbf{Y} - \operatorname{ctanh}_c(\mathbf{A}_a\mathbf{X}_a^k)$ and thus the entries of $\mathbf{E}^k$ fulfill

$$|\Re(e_{i,j})| \in [0,2], \tag{55a}$$
$$|\Im(e_{i,j})| \in [0,2], \tag{55b}$$

where $|\Re(e_{i,j})|$ and $|\Im(e_{i,j})|$ approaching zero indicate that $y_{i,j}$ and the corresponding entry in $\operatorname{csgn}_c(\mathbf{A}_a\mathbf{X}_a)$ have the same sign. In this case, $y_{i,j}$ can be considered to have the correct sign. Conversely, when $|\Re(e_{i,j})|$ and $|\Im(e_{i,j})|$ are close to 2, signifying that $y_{i,j}$ is influenced by noise and has potentially been flipped. Therefore, the appropriate range for the threshold ϵ is $(0,2]$, as it serves to differentiate between correct and incorrect measurements based on the fitting error. In practice, ϵ should be set close to 2 to accommodate noisy conditions.

D. Convergence Analysis

In this subsection, we analyze the convergence behavior of the proposed method. It is important to note that the original optimization task is solved in two stages, namely, DoA estimation and magnitude estimation. The former involves a complex-valued optimization, whereas the latter is a real-valued problem. Both stages are handled using PAM and FBS. Consequently, the convergence analysis for both stages can be conducted in a similar manner.

To prove the convergence, we first introduce the following lemma:

Lemma 1. *[40] Let $F\colon \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ be a proper lower semicontinuous function that is lower bounded and satisfies the Kuradyka–Łojasiewicz (KŁ) property. Suppose that F takes the form*

$$F = H + J, \tag{56}$$

where $H:\mathbb{R}^n\to\mathbb{R}$ is finite valued, differentiable, and has a L-Lipschitz gradient, and J is a proper lower semicontinuous function or indicator function. Then, $\min_{\mathbf{X}} F(\mathbf{X})$ can be tackled using the FBS method. Moreover, the iterative procedure satisfies the following property:

When the step-size in gradient descent meets $0<\eta<1/L$, $F(\mathbf{X}^k)+\frac{\rho}{2}\|\mathbf{X}^k-\mathbf{X}_{k-1}\|_F^2 \le F(\mathbf{X}_{k-1})$ holds for some $\rho > L$.

Based on Lemma 1, we present the convergence behaviors of the devised method in the following theorems. The first one exhibits the convergence of the objective function value.

Theorem 1. *Let $\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)$ be the loss function value generated by the proposed DoA estimation algorithm. Then, the following properties hold:*

(i) $\hat{\mathcal{L}}(\mathbf{X}_a) = \mathcal{L}(\mathbf{X}_a,\mathbf{O}^{k-1}) + \rho\|\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}\|_F^2 = \|\mathbf{Y}-\operatorname{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}^{k-1}\|_F^2+\tau\|\mathbf{X}_a(1\colon S,:)\|_{2,0}+\rho\|\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}\|_F^2$ is a proper lower semicontinuous function with a lower bound and satisfies the KŁ property. In addition, $\|\mathbf{Y}-\operatorname{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}^{k-1}\|_F^2+\rho\|\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}\|_F^2$ w.r.t. $\mathbf{X}_a$ has a Lipschitz gradient.

(ii) When $0<\eta<1/L$ with L being the Lipschitz constant of the gradient of $\|\mathbf{Y}-\operatorname{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}^{k-1}\|_F^2+\rho\|\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}\|_F^2$ w.r.t. $\mathbf{X}_a$, the sequence $\{\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ is nonincreasing.

(iii) $\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)$ is lower bounded.

As a result, $\{\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ is convergent.

The proof is provided in Appendix B.

Then, we analyze the sequence behavior in Theorem 2.

Theorem 2. *Let $\{(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ be the sequence generated by the proposed DoA estimation algorithm. Additionally, $\{(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ is considered to be bounded. Then, the following properties hold:*

(i)

$$\lim_{K_{\max}\to+\infty}\sum_{k=1}^{K_{\max}}(\|\mathbf{X}_a^k-\mathbf{X}_a^{k-1}\|_F^2+\|\mathbf{O}^k-\mathbf{O}^{k-1}\|_F^2) < +\infty. \tag{57}$$

Hence, we obtain

$$\lim_{K_{\max}\to+\infty}(\|\mathbf{X}_a^k-\mathbf{X}_a^{k-1}\|_F+\|\mathbf{O}^k-\mathbf{O}^{k-1}\|_F)=0. \tag{58}$$

(ii) The objective function $\mathcal{L}$ is bounded below and has the KŁ property, and the function $(\mathbf{X}_a,\mathbf{O})\mapsto\|\mathbf{Y}-\operatorname{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}\|_F^2$ has a Lipschitz gradient. Then, the sequence $\{(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ converges to a critical point.

(iii) The suggested DoA estimation method converges with at least a sublinear rate.

The proof is provided in Appendix C.

E. Computational Complexity

In this subsection, we study the computational complexity of the proposed algorithm. The complexity of updating $\mathbf{X}_a^k$ is $\mathcal{O}(P_{\max}(MQ(S+Q)))$. In addition, computing

$\mathbf{O}^k$ has a complexity of $\mathcal{O}(MQ(S+Q))$. As a result, the overall computational requirement of the DoA estimation method is $\mathcal{O}(K_{\max}P_{\max}(MQ(S+Q))$. For the magnitude estimation approach, its computational complexity is $\mathcal{O}(K_{\max}P_{\max}(2MQ(2\widetilde{S}+Q))$. Generally, the order of magnitude S is 10^5, while that of $\widetilde{S}$ is 10^0. As $\widetilde{S} \ll S$, the complexity of the magnitude estimation method is much lower than that of DoA estimation approach.

TABLE I: Complexity comparison of different algorithms in per iteration under the condition of $S \gg M \approx Q$.

Method	Computational complexity
Proposed	$\mathcal{O}(P_{\max}M^2S)$
CBIHT-ℓ_2	$\mathcal{O}(8M^2S)$
1-bit OGIR	$\mathcal{O}((2M^2S+MS^2))$
1-bit ML	$\mathcal{O}((2M^2S))$
1-bit APG	$\mathcal{O}(5M^3+MS^2+S^3)$

Table I tabulates the computational requirement of five algorithms, where the conditions of $S \gg M$ and $M \approx Q$ facilitate comparison. According to $S \gg M$, the suggested method, CBIHT-ℓ_2, and 1-bit ML have a comparable computational complexity that is less than that of 1-bit OGIR and 1-bit APG.

F. Comparison with ROCS

Although the proposed signal model and that in ROCS share similarities in terms of noise modeling and smooth surrogate function, there are key differences in the quantization strategy and the optimization method. Specifically, ROCS employs a zero quantization threshold, whereas the proposed model utilizes a non-zero threshold. This non-zero threshold introduces the prior information about the signal magnitude, which enhances the dynamic range in radar systems and improves estimation performance. In addition, ROCS assigns the $\ell_{2,0}$-norm as a constraint and thus requires the prior information on target number. In contrast, the suggested method treats the $\ell_{2,0}$-norm as a penalty term and incorporates a strategy for adaptively determining the corresponding penalty parameter. As a result, the proposed scheme does not require the target number that will be adaptively estimated.

IV. Experimental Results

All simulations are conducted using MATLAB (R2019b) on a computer with Intel(R) Core(TM) i7- 14900 3.2GHz CPU and 64 GB memory.

A. Simulation Setting

Unless stated otherwise, the signals are coherent sources with Gaussian distribution. The ULA is composed of $M = 60$ sensors, each with an inter-element spacing of $d = 0.5\lambda$. The number of snapshots is $Q = 64$, and the DoAs of the three signals are $\theta_1 = 40.2°$, $\theta_2 = -30.5°$, and $\theta_3 = 5.1°$, with respective magnitudes of 1.0, 1.4, and 1.8. For the quantization threshold, there are two common distributions, namely, the Gaussian distribution and uniform distribution. In our work, we adopt the zero-mean Gaussian distribution to generate the quantization threshold. Zero-mean white Gaussian noise is added to the clean signal, and the signal-to-noise ratio (SNR) is defined as:

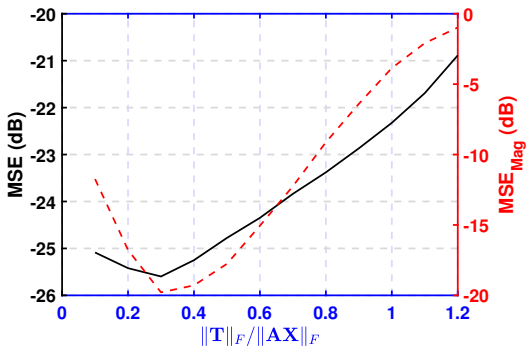

Fig. 2: Performance illustration under different thresholds with 3 targets, 60 antennas, 64 snapshot, and SNR = 0dB.

$$\text{SNR} = 10\log_{10}\Big(\frac{\|\mathbf{AX}\|_F^2}{\sigma^2}\Big), \tag{59}$$

where σ^2 is the noise variance. The estimation performance is evaluated using the mean square error (MSE) in dB, defined as:

$$\text{MSE} = 10\log_{10}\Big(\frac{1}{M_c}\sum_{m=1}^{M_c}\sum_{s=1}^{\widetilde{S}}\frac{(\widehat{\theta}_{s,m}-\theta_s)^2}{\widetilde{S}}\Big), \tag{60}$$

$$\text{MSE}_{\text{Mag}} = 10\log_{10}\Big(\frac{1}{M_c}\sum_{m=1}^{M_c}\sum_{s=1}^{\widetilde{S}}\frac{(\widehat{\varrho}_{s,m}-\varrho_s)^2}{\widetilde{S}}\Big), \tag{61}$$

where $M_c = 100$ is the number of Monte Carlo trials, while $\widehat{\theta}_{s,m}$ and $\widehat{\varrho}_{s,m}$ are the sth DoA estimate and magnitude estimate at the mth trial.

B. Investigation of Quantization Threshold

We use empirical results to demonstrate the impact of the quantization threshold on estimation performance. The results are plotted in Fig. 2, where the quantization threshold obeys the zero-mean Gaussian distribution. The results indicate that the errors in both DoA estimation and magnitude recovery are minimized when the ratio is approximately 0.3. When the ratio increases beyond this value, the performance degrades.

C. Investigation of ζ

In Section III.C, we have dissected the penalty parameter τ, that is, its determination is equivalent to the choice of ζ. In this subsection, we examine the effect of ζ on the estimation performance of target number. The results are plotted in Fig. 3, where SNR varies from -6dB to 6dB. We observe that larger values of ζ achieve a high estimation probability, e.g., 100%, at high SNRs, whereas smaller values of ζ perform better at low SNRs. Furthermore, the estimation probability is less sensitive to ζ under low noise conditions but becomes more sensitive as the noise level increases. Overall, setting $\zeta = 5.5$ maintains an estimation probability of at least 99% across the entire SNR range from -6 dB to 6 dB. Therefore, we choose $\zeta = 5.5$ for the following simulations.

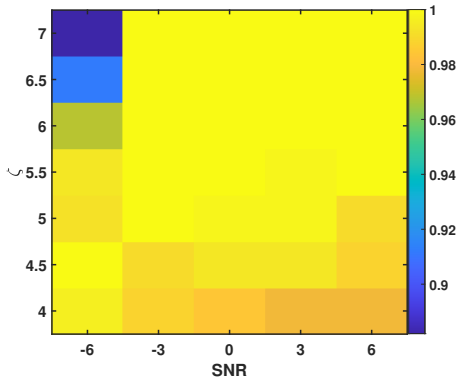

Fig. 3: Target number estimation probability versus ζ and SNR with 3 targets, 60 antennas, and 64 snapshots.

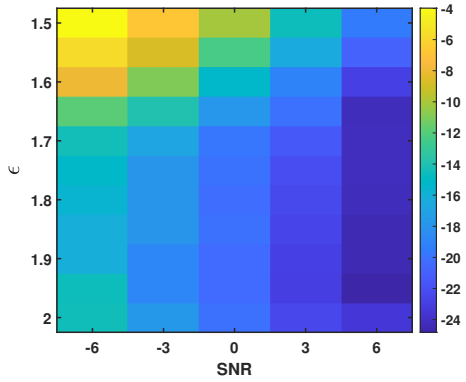

Fig. 4: Signal magnitude estimation performance versus ϵ and SNR with 3 targets, 60 antennas, and 64 snapshots.

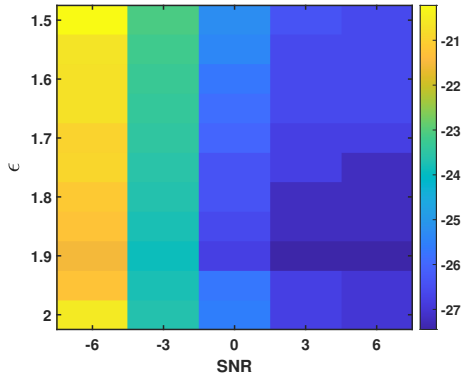

Fig. 5: DoA estimation performance versus ϵ and SNR with 3 targets, 60 antennas, and 64 snapshots.

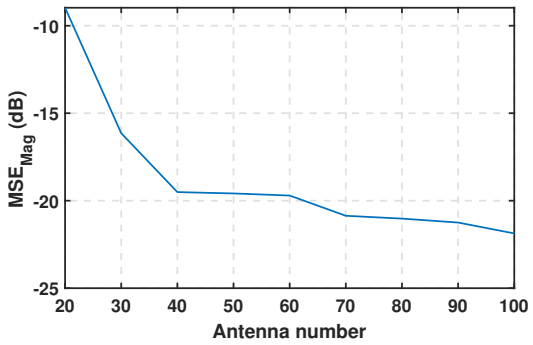

Fig. 6: Signal magnitude estimation performance versus antenna number with 3 targets, 64 snapshots, and 0 dB Gaussian noise.

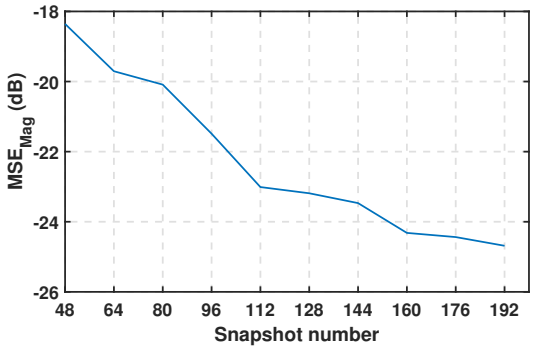

Fig. 7: Signal magnitude estimation performance versus snapshot number with 3 targets, 60 antennas, and 0dB Gaussian noise.

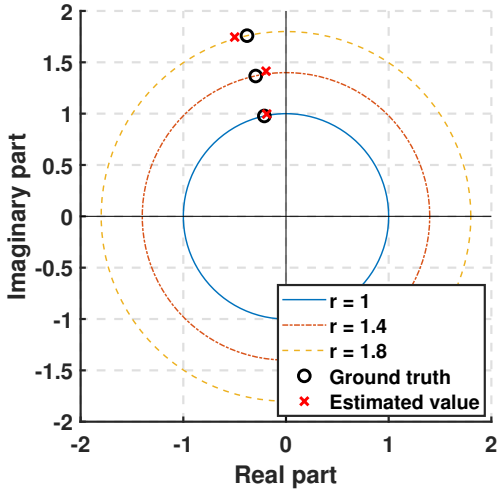

Fig. 8: Illustration of magnitude recovery with 3 targets, 60 antennas, 64 snapshot, and SNR = 0dB.

D. Investigation of ϵ

We have analyzed the potential values of ϵ in Section III.C. In this subsection, we investigate the impact of ϵ on recovery performance through numerical experiments. The results are presented in Figs. 4 and 5, which correspond to signal magnitude and DoA estimation performance, respectively. For the magnitude estimation, we observe that the $\mathrm{MSE}_{\mathrm{Mag}}$ remains largely unchanged with $\mathrm{SNR} \geq 3$, indicating that the estimation performance is not sensitive to ϵ at low noise environments. However, under high noise conditions, selecting an appropriate value for ϵ is critical to achieving satisfactory performance. Specifically, $\mathrm{MSE}_{\mathrm{Mag}}$ shows significant variation as ϵ increases from 1.6 to 2.0, reaching a minimum at $\epsilon = 1.9$. These findings are consistent with our analysis in Section III.C. For DoA estimation, similar behavior is observed, as depicted in Fig. 5.

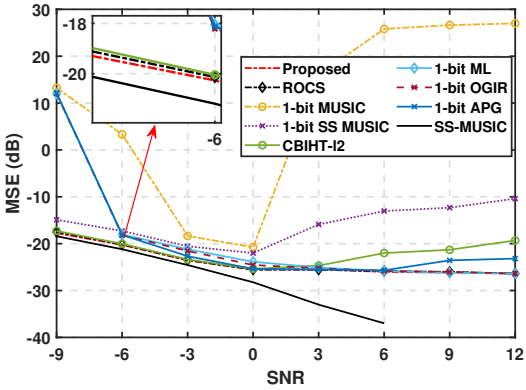

Fig. 9: MSE versus SNR with 3 targets, 60 antennas, and 64 snapshots.

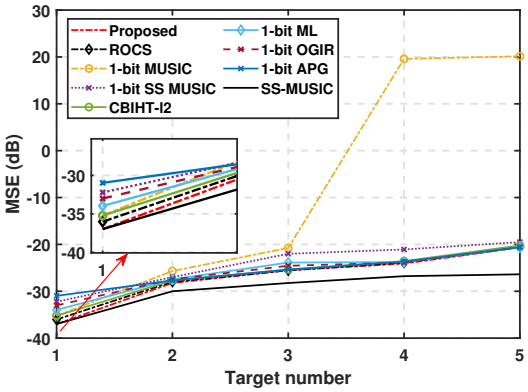

Fig. 10: MSE versus target number with 60 antennas, 64 snapshots, and SNR = 0dB.

E. Investigation of Antenna and Snapshot Numbers

Furthermore, we examine the impact of the number of antennas and snapshots on the signal magnitude estimation. The corresponding results are plotted in Figs. 6 and 7, respectively. It is seen that the magnitude estimation error decreases as the number of antennas increases. Notably, when the number of antennas increases from 20 to 40, there is a significant improvement in performance. Beyond this point, further increases continue to enhance performance, but the effect becomes less pronounced. Similarly, as illustrated in Fig. 7, increasing the number of snapshots significantly reduces the magnitude estimation error. Moreover, we illustrate the performance of magnitude estimation in Fig. 8. It is seen that the proposed method is able to accurately recover the signal magnitude.

F. Performance Comparison

The proposed method is compared with six popular algorithms, namely, 1-bit multiple signal classification (1-bit MUSIC) [10], 1-bit spatial smoothing MUSIC (1-bit SS-MUSIC) [41], complex-valued BIHT based on ℓ_2-norm (CBIHT-ℓ_2) [19], 1-bit maximum likelihood (1-bit ML) [26], 1-bit-off-grid iterative reweighted (1-bit OGIR) [25], and accelerated proximal gradient algorithm for single-bit DoA estimation (1-bit APG) [42]. For 1-bit ML, the parameter

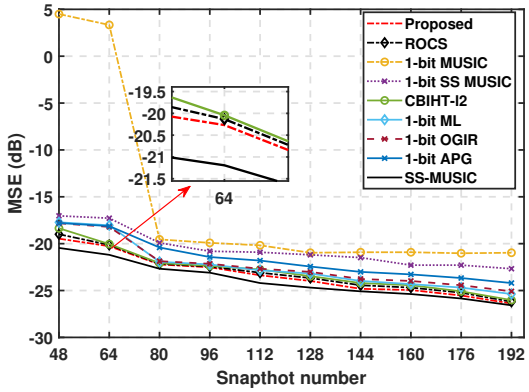

Fig. 11: MSE versus snapshot number with 3 targets, 60 antennas, and SNR = -6dB.

λ requires tuning for accurate target number estimation, depending on the number of antennas and snapshots. As such, λ should be determined using grid-search for different scenarios in general. In addition, the default step-size of CBIHT-ℓ_2 is $0.1/M$, where M is the number of antennas. Moreover, 1-bit OGIR sets the reweighting exponent to 10^{-4}. Since the source signals are correlated, the SS-MUSIC algorithm is also tested using unquantized received signals, which serves as a baseline.

The performance comparison under different SNRs is first studied, and the results are plotted in Fig. 9. It is seen that SS-MUSIC achieves the smallest MSEs across all SNR levels due to the absence of quantization error in its received signals. The proposed method, along with 1-bit ML and 1-bit OGIR, demonstrates comparable estimation performance at high SNRs. However, the proposed method outperforms the others in low-SNR scenarios, showing enhanced robustness. In contrast, 1-bit APG, CBIHT-ℓ_2, 1-bit SS-MUSIC, and 1-bit MUSIC exhibit a trend where their estimation performance deteriorates as SNR increases. Overall, the proposed method consistently demonstrates superiority over the other 1-bit algorithms.

It is seen that several algorithms have a performance degradation when SNR increases. Strong noise may destroy signal coherence, while weak noise may retain this characteristic, resulting in that 1-bit MUSIC has a severely degraded performance in high SNRs. For 1-bit SS-MUSIC, CBITH-L2, 1-bit APG, their performance has a slight reduction at high SNRs. This behavior can be explained by the introduction of appropriate random noise, which spreads the quantized values between +1 and -1. This added noise makes the DoA estimation process easier because it introduces more statistical variance in the quantized signal. However, at low noise levels, the quantized signal is deterministic, providing little information about the actual signal variations, making it harder to accurately estimate the true direction.

Then, the estimation accuracy of nine methods under different targets numbers is investigated in Fig. 10, where the DoAs of the fourth and fifth signals are $\theta_4 = -20.3°$, and $\theta_5 = 20.7°$, respectively. It is observed that SS-MUSIC based on unquantized data still achieves the best performance, with its superiority becoming more pronounced as the num-

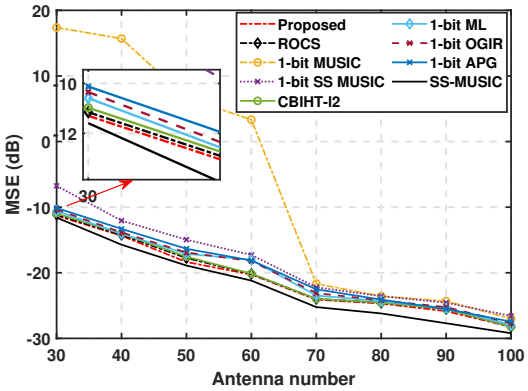

Fig. 12: MSE versus antenna number with 3 targets, 64 snapshots, and SNR = -6dB.

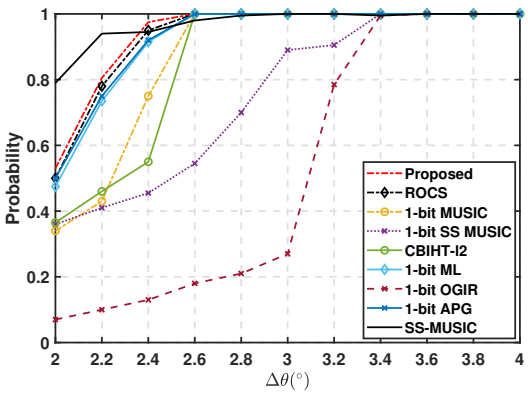

Fig. 13: Probability of resolution versus $\Delta\theta$ with 3 targets, 60 antennas 64 snapshot, and SNR = -6dB.

ber of targets increases. The proposed method demonstrates comparable MSE to SS-MUSIC, significantly outperforming other competing approaches. However, as the number of targets increases, the performance of the proposed method gradually converges towards that of 1-bit SS-MUSIC, 1-bit APG, CBIHT-ℓ_2, 1-bit ML, and 1-bit OGIR. Notably, 1-bit MUSIC exhibits a marked performance degradation as the number of targets increases, indicating its limitations in handling scenarios with a large number of targets. Overall, the proposed method remains highly competitive across different target numbers.

Moreover, we compare all methods under different numbers of snapshots and the results are depicted in Fig. 11. When the snapshot number increases, their estimation accuracy improves, but the MSE of 1-bit MUSIC becomes steady. In scenarios with a small number of snapshots, the proposed method, SS-MUSIC, and CBIHT-ℓ_2 achieve significantly lower MSEs compared to the other algorithms. As the number of snapshots grows, the proposed method, SS-MUSIC, CBIHT-ℓ_2, 1-bit OGIR, and 1-bit ML attain comparable performance, outperforming 1-bit APG, 1-bit MUSIC, and 1-bit SS-MUSIC. Notably, as the number of snapshots increases, the performance of the proposed method approaches that of SS-MUSIC.

The estimation performance of nine approaches under different antenna numbers is illustrated in Fig. 12. We see that increasing the antenna number from 30 to 70 is able to significantly improve the performance of 1-bit MUSIC. For the other methods, a larger number of antennas consistently leads to lower MSEs. With a smaller number of antennas, the proposed method, CBIHT-ℓ_2, 1-bit ML, 1-bit OGIR, and 1-bit APG are comparable to SS-MUSIC. However, with a larger number of antennas, the superiority of SS-MUSIC over the 1-bit data-based algorithms becomes more evident. Despite this, the proposed method consistently ranks the first place among all the single-bit approaches.

Furthermore, the angular resolution of different methods is investigated. The results are plotted in Fig. 13, where if the magnitude of the DOA estimation error is less than 0.5° for each target, then the corresponding source is considered to be successfully resolved. It is seen that SS-MUSIC is better than all single-bit algorithms when $\Delta\theta$ is less than 2.4°. However, when $\Delta\theta = 2.6^\circ$, our method, ROCS, 1-bit APG, 1-bit ML, 1-bit MUSIC, and CBIHT-ℓ_2 outperform SS-MUSIC, but the probabilities of 1-bit SS MUSIC and 1-bit OGIR are still less than 60%. Among all single-bit approaches, the proposed method is superior to its competitors when $\Delta\theta$ varies from 2° to 2.4°. When $\Delta\theta = 2.6^\circ$, single-bit methods, except for 1-bit SS MUSIC and 1-bit OGIR, achieve a 100% probability of successful resolution.

V. Conclusion

In this article, the quantized signal model for single-bit DoA estimation is established by leveraging a non-zero threshold to enable signal magnitude estimation. Then, we construct the optimization problem using the capped Frobenius norm and the $\ell_{2,0}$-norm, where the former restricts the fitting error, and the latter serves as a penalty term for target number estimation. To address the resultant optimization task, we employ proximal alternating minimization and forward-backward splitting. Although the proposed algorithm includes an auxiliary parameter for the target number estimation, we develop a strategy to automatically determine this parameter. The convergence properties of the proposed method are analyzed, demonstrating that the objective function value converges and the variable sequence converges to a critical point. Simulation results show that the proposed algorithm can accurately estimate both the target number and the signal magnitude. Furthermore, our scheme outperforms SOTA algorithms in terms of DoA estimation accuracy.

As part of our future work, we will explore the optimal determination of the quantization threshold. In the present study, the threshold follows a zero-mean Gaussian distribution. For deterministic signals, the optimized threshold design may further enhance the performance of single-bit DoA estimation. In addition, as our objective function involves tanh function and $\ell_{2,0}$-norm, the proposed method cannot guarantee the global optimum. Thereby, we will explore other technologies to establish global optimality of the sequence convergence for the non-convex and non-smooth objective function in the future work.

APPENDIX A
DERIVATION OF SOLUTION TO (41a)

Based on the definition of $\varphi_{\mathrm{CF}}(\cdot)$, we have

$$
\begin{aligned}
o^* &= \begin{cases} \arg\min\limits_{o,|o|\le\epsilon} (e-o)^2-(\epsilon-|o|)^2+\epsilon^2+\rho(o-o^{k-1})^2, \\ \arg\min\limits_{o,|o|>\epsilon} (e-o)^2+\epsilon^2+\rho(o-o^{k-1})^2, \end{cases} \\
&= \begin{cases} \arg\min\limits_{o}\ e^2-2eo+2\epsilon|o|+\rho(o-o^{k-1})^2, |o|\le\epsilon, \\ \arg\min\limits_{o}\ (e-o)^2+\epsilon^2+\rho(o-o^{k-1})^2, |o|>\epsilon, \end{cases} \\
&= \begin{cases} \arg\min\limits_{o}\ e^2-2o(e-\epsilon)+\rho(o-o^{k-1})^2, 0\le o\le\epsilon, \\ \arg\min\limits_{o}\ e^2-2o(e+\epsilon)+\rho(o-o^{k-1})^2, -\epsilon\le o\le 0, \\ \arg\min\limits_{o}\ (e-o)^2+\epsilon^2+\rho(o-o^{k-1})^2, |o|>\epsilon, \end{cases} \\
&= \begin{cases} \max\left(0, \min\left(o^{k-1}+\dfrac{e-\epsilon}{\rho},\ \epsilon\right)\right), & 0\le o\le\epsilon, \\ & |e|\le\epsilon \\ \max\left(-\epsilon, \min\left(o^{k-1}+\dfrac{e+\epsilon}{\rho},\ 0\right)\right), & -\epsilon\le o\le 0, \\ & |e|\le\epsilon, \\ \mathrm{sgn}(\frac{e+\rho o^{k-1}}{1+\rho})\cdot\max\left(\epsilon,\ |\frac{e+\rho o^{k-1}}{1+\rho}|\right), & |o|>\epsilon, |e|>\epsilon. \end{cases}
\end{aligned} \tag{62}
$$

When ρ is small enough, for example $\rho=10^{-8}$, the following equalities hold

$$o^{k-1}+\frac{e-\epsilon}{\rho}<0,\ 0\le o\le\epsilon, \tag{63a}$$

$$o^{k-1}+\frac{e+\epsilon}{\rho}>0,\ -\epsilon\le o\le 0, \tag{63b}$$

$$\frac{e+\rho o^{k-1}}{1+\rho}=e. \tag{63c}$$

Therefore, the optimal solution o^* can be simplified as

$$o^*=\begin{cases} 0, & |e|\le\epsilon, \\ e, & |e|>\epsilon. \end{cases} \tag{64}$$

The proof is complete. ■

APPENDIX B
PROOF OF THEOREM 1

A. Property (i)

We have:

$$
\begin{aligned}
\hat{\mathcal{L}}(\mathbf{X}_a) &= \mathcal{L}(\mathbf{X}_a,\mathbf{O}^{k-1})+\rho\|\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}\|_F^2 && (65a)\\
&= \|\mathbf{Y}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}^{k-1}\|_F^2+\tau\|\mathbf{X}_a(1\!:\!S,:)\|_{2,0} \\
&\qquad +\rho\|\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}\|_F^2 && (65b)\\
&= \tilde{\mathcal{L}}(\mathbf{X}_a,\mathbf{O}^{k-1})+\tau\|\mathbf{X}_a(1\!:\!S,:)\|_{2,0} && (65c)
\end{aligned}
$$

It is clear that $\hat{\mathcal{L}}(\mathbf{X}_a)$ is lower semicontinuous and lower bounded. Besides, it is known that the $\tanh_c(\cdot)$ function is definable in an o-minimal structure [28], and $\|\mathbf{X}_a(1\!:\!S,:)\|_{2,0}$ is a proper lower semicontinuous function [40]. Therefore, $\hat{\mathcal{L}}(\mathbf{X}_a)$ has the KŁ property [28].

Then, we prove that $\tilde{\mathcal{L}}(\mathbf{X}_a,\mathbf{O}^{k-1})$ has a Lipschitz gradient. Its first and second derivatives are:

$$
\begin{aligned}
&\nabla_{\mathbf{X}_a}\tilde{\mathcal{L}}(\mathbf{X}_a,\mathbf{O}^{k-1}) \\
&=-2c\mathbf{A}_a^{\mathrm{H}}\Big((\mathbf{S}-\tanh_c(\mathbf{A}_a\mathbf{X}_a))\odot(\mathbf{1}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)^2)\Big) \\
&+2\rho(\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}), && (66a)\\
&\nabla^2_{\mathbf{X}_a}\tilde{\mathcal{L}}(\mathbf{X}_a,\mathbf{O}^{k-1}) \\
&=2c^2(\mathbf{D}_1\mathbf{A}_a)^{\mathrm{H}}\mathbf{A}_a+4c^2\mathbf{A}_a^{\top}\mathbf{D}_2\mathbf{A}_a+2\rho\mathbf{I}, && (66b)
\end{aligned}
$$

where $\mathbf{S}=\mathbf{Y}-\mathbf{O}^{k-1}$, $\mathbf{D}_1=\Big((\mathbf{1}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)^2)^2\Big)$, and $\mathbf{D}_2=\Big((\mathbf{S}-\tanh_c(\mathbf{A}_a\mathbf{X}_a))\odot\tanh_c(\mathbf{A}_a\mathbf{X}_a)\odot(\mathbf{1}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)^2)\Big)$. As $-1\le\tanh_c(\mathbf{A}_a(i,:)\mathbf{X}_a(:,j))\le 1$, we obtain

$$|\tanh_c(\mathbf{A}_a(i,:)\mathbf{X}_a(:,j))(1-\tanh_c(\mathbf{A}_a(i,:)\mathbf{X}_a(:,j))^2)|<\frac{1}{2}, \tag{67a}$$

$$\|\mathbf{D}_1\|_2\le\sigma_{\max 1},\text{with }\mathbf{D}_1(i,j)\in[0,1], \tag{67b}$$

$$\|\mathbf{D}_2\|_2\le\frac{1}{2}\|\mathbf{S}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)\|_2\le\frac{\sigma_{\max 2}}{2}, \tag{67c}$$

where $\sigma_{\max 1}$ and $\sigma_{\max 2}$ are the largest singular value of $\mathbf{D}_1$ and $\mathbf{S}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)$, respectively. Moreover, we attain

$$
\begin{aligned}
&\|\nabla^2_{\mathbf{X}_a}\tilde{\mathcal{L}}(\mathbf{X}_a,\mathbf{O}^{k-1})\|_2 \\
&=\|2c^2(\mathbf{D}_1\mathbf{A})'\mathbf{A}+4c^2(\mathbf{D}_2\mathbf{A})'\mathbf{A}+2\rho\mathbf{I}\|_2 && (68a)\\
&\le 2c^2\|(\mathbf{D}_1\mathbf{A})'\mathbf{A}\|_2+4c^2\|(\mathbf{D}_2\mathbf{A})'\mathbf{A}\|_2+2\rho && (68b)\\
&\le 2c^2\|\mathbf{D}_1\|_2\|\mathbf{A}\|_2^2+4c^2\|\mathbf{D}_2\|_2\|\mathbf{A}\|_2^2+2\rho && (68c)\\
&\le 2c^2(\sigma_{\max 1}+\sigma_{\max 2})\|\mathbf{A}\|_2^2+2\rho. && (68d)
\end{aligned}
$$

When $\|\mathbf{A}\|_F$ is finite, $\|\mathbf{A}\|_2^2\le\|\mathbf{A}\|_F^2<+\infty$ must hold, resulting in

$$2c^2(\sigma_{\max 1}+\sigma_{\max 2})\|\mathbf{A}\|_2^2+2\rho<+\infty. \tag{69}$$

As a result, $\tilde{\mathcal{L}}(\mathbf{X}_a,\mathbf{O}^{k-1})$ has a Lipschitz gradient. The proof is complete. ■

B. Property (ii)

For (36a), it is equivalent to $\hat{\mathcal{L}}(\mathbf{X}_a)$. Since it is solved using forward–backward splitting and thus based on Lemma 1, we have

$$\hat{\mathcal{L}}(\mathbf{X}_p^{k-1})-\hat{\mathcal{L}}(\mathbf{X}_{p-1}^{k-1})\le-\frac{\rho_{\hat{\mathcal{L}}}}{2}\|\mathbf{X}_p^{k-1}-\mathbf{X}_{p-1}^{k-1}\|_F^2, \tag{70}$$

where $\rho_{\hat{\mathcal{L}}}>2c^2(\sigma_{\max 1}+\sigma_{\max 2})\|\mathbf{A}\|_2^2+2\rho$.

By induction on p, we further obtain

$$
\begin{aligned}
&\hat{\mathcal{L}}(\mathbf{X}^k)-\hat{\mathcal{L}}(\mathbf{X}^{k-1})\le\sum_{p=1}^{P_{\max}}-\frac{\rho_{\hat{L}}}{2}\|\mathbf{X}_p^{k-1}-\mathbf{X}_{p-1}^{k-1}\|_F^2 && (71a)\\
&\Leftrightarrow\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^{k-1})-\mathcal{L}(\mathbf{X}_a^{k-1},\mathbf{O}^{k-1}) \\
&\le-\rho\|\mathbf{X}^k-\mathbf{X}^{k-1}\|_F^2-\sum_{p=1}^{P_{\max}}\frac{\rho_{\hat{\mathcal{L}}}}{2}\|\mathbf{X}_p^{k-1}-\mathbf{X}_{p-1}^{k-1}\|_F^2, && (71b)
\end{aligned}
$$

where $\mathbf{X}_0^{k-1}=\mathbf{X}^{k-1}$ and $\mathbf{X}^k=\mathbf{X}_{P_{\max}}^{k-1}$. We see that updating $\mathbf{X}$ does not increase the objective function value.

On the other hand, as (43) is an optimal solution of (36b), we have

$$\begin{aligned}&\|\mathbf{Y}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a^k)-\mathbf{O}^k\|_F^2+\varphi_{\mathrm{CF}}(\mathbf{O}^k)+\rho\|\mathbf{O}^k-\mathbf{O}^{k-1}\|_F^2\\&\le\|\mathbf{Y}-\mathrm{ctanh}_c(\mathbf{A}_a\mathbf{X}_a^k)-\mathbf{O}^{k-1}\|_F^2+\varphi_{\mathrm{CF}}(\mathbf{O}^{k-1})\quad(72a)\\&\Leftrightarrow\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)-\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^{k-1})\le-\rho\|\mathbf{O}^k-\mathbf{O}^{k-1}\|_F^2\quad(72b)\end{aligned}$$

Combining (71b) and (72b) results in

$$\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)-\mathcal{L}(\mathbf{X}_a^{k-1},\mathbf{O}^{k-1})\le 0,\tag{73}$$

which indicates that $\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)$ is nonincreasing as all variables update. The proof is complete. ■

C. Property (iii)

Based on the definition of $\varphi_{\mathrm{CF}}(\cdot)$, we have

$$\begin{aligned}&(e-o)^2+\varphi_{\mathrm{CF}}(o)>0,\text{if }|e|>\epsilon\quad(74a)\\&(e-o)^2+\varphi_{\mathrm{CF}}(o)\\&=(e-o)^2+(2\epsilon-|o|)|o|>0,\text{if }|o|\le\epsilon\quad(74b)\end{aligned}$$

Besides, the $\ell_{2,0}$-norm in $\mathcal{L}(\mathbf{X}_a,\mathbf{O})$ is no smaller than zero. Therefore, $\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)\ge 0$ must hold.

As a result, based on Properties (ii) and (iii), the convergence of $\{\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ is guaranteed. The proof is complete. ■

Appendix C
Proof of Theorem 2

A. Property (i)

From (71b), we get

$$\mathcal{L}(\mathbf{X}_a^{k-1},\mathbf{O}^{k-1})-\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^{k-1})\ge\rho\|\mathbf{X}^k-\mathbf{X}^{k-1}\|_F^2.\tag{75}$$

Combining (75) and (72b) yields:

$$\begin{aligned}&\|\mathbf{O}^k-\mathbf{O}^{k-1}\|_F^2+\|\mathbf{X}^k-\mathbf{X}^{k-1}\|_F^2\\&\qquad\le\frac{\mathcal{L}(\mathbf{X}_a^{k-1},\mathbf{O}^{k-1})-\mathcal{L}(\mathbf{X}_a^k,\mathbf{O}^k)}{\rho}.\end{aligned}\tag{76}$$

By induction on k, we obtain

$$\begin{aligned}&\lim_{K_{\max}\to+\infty}\sum_{k=1}^{K_{\max}}(\|\mathbf{O}^k-\mathbf{O}^{k-1}\|_F^2+\|\mathbf{X}^k-\mathbf{X}^{k-1}\|_F^2)\\&\qquad\le\lim_{K_{\max}\to+\infty}\frac{\mathcal{L}(\mathbf{X}_a^0,\mathbf{O}^0)-\mathcal{L}(\mathbf{X}_a^{K_{\max}},\mathbf{O}^{K_{\max}})}{\rho}\\&\qquad<+\infty.\end{aligned}\tag{77}$$

Therefore, under the assumption that $\{(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ is bounded, we have

$$\lim_{K_{\max}\to+\infty}(\|\mathbf{O}^k-\mathbf{O}^{k-1}\|_F+\|\mathbf{X}^k-\mathbf{X}^{k-1}\|_F)=0.\tag{78}$$

The proof is complete. ■

B. Property (ii)

It is obvious that $\mathcal{L}$ is bounded below. In addition, as the $\ell_{2,0}$-norm is semi-algebraic, it is also definable in an o-minimal structure [43]. Combining Property (i) in Appendix B, we conclude that $\mathcal{L}$ has the KŁ property.

On the other hand, we have proved the following function has a Lipschitz gradient:

$$\begin{aligned}\tilde{\mathcal{L}}(\mathbf{X}_a,\mathbf{O}^{k-1})=&\|\mathbf{Y}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}^{k-1}\|_F^2\\&+\rho\|\mathbf{X}_a-(\mathbf{X}_a)_{p-1}^{k-1}\|_F^2\end{aligned}\tag{79}$$

Therefore, the function $(\mathbf{X}_a,\mathbf{O})\mapsto\|\mathbf{Y}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}\|_F^2$ has a Lipschitz gradient w.r.t. $\mathbf{X}_a$. Then, we prove that it has a Lipschitz gradient w.r.t. $\mathbf{O}$:

$$\left\|\frac{\partial^2\|\mathbf{Y}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}\|_F^2}{\partial^2\mathbf{O}}\right\|_2=\|2\mathbf{I}\|_2\le 2,\ \forall\mathbf{O}\in\mathbb{C}^M.\tag{80}$$

As a result, the function $(\mathbf{X}_a,\mathbf{O})\mapsto\|\mathbf{Y}-\tanh_c(\mathbf{A}_a\mathbf{X}_a)-\mathbf{O}\|_F^2$ has a Lipschitz gradient.

Furthermore, as $\min_{\mathbf{X}_a,\mathbf{O}}\mathcal{L}(\mathbf{X}_a,\mathbf{O})$ is solved using the PAM and FBS, and $(\mathbf{X}_a^k,\mathbf{O}^k)_{k\in\mathbb{N}}$ is bounded, $\{(\mathbf{X}_a^k,\mathbf{O}^k)\}_{k\in\mathbb{N}}$ converges to a critical point based on Theorem 6.2 in [40]. The proof is complete. ■

C. Property (iii)

As $\mathcal{L}$ has the KŁ property, based on Theorem 1 in [43], our method converges at least sublinearly. The proof is complete. ■

References

[1] M. Deng, Z. Cheng, L. Wu, B. Shankar, and Z. He, "One-bit ADCs/DACs based MIMO radar: Performance analysis and joint design," *IEEE Trans. Signal Process.*, vol. 70, pp. 2609–2624, May 2022.

[2] M. Chen, Q. Li, L. Huang, L. Feng, and M. Rihan, "One-bit Cramér–Rao bound of direction of arrival estimation for deterministic signals," *IEEE Trans. Circuits Syst. Express Briefs*, vol. 71, no. 2, pp. 957–961, Feb. 2023.

[3] Y.-H. Xiao, D. Ramírez, L. Huang, X. P. Li, and H. C. So, "One-bit target detection in collocated MIMO radar with colored background noise," *IEEE Trans. Signal Process.*, vol. 72, pp. 5274–5290, Nov. 2024.

[4] A. Eamaz, F. Yeganegi, and M. Soltanalian, "Covariance recovery for one-bit sampled stationary signals with time-varying sampling thresholds," *Signal Process.*, vol. 206, p. 108899, May 2023.

[5] ——, "Covariance recovery for one-bit sampled non-stationary signals with time-varying sampling thresholds," *IEEE Trans. Signal Process.*, vol. 70, pp. 5222–5236, Oct. 2022.

[6] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*, 2018, vol. 7, pp. 509–581.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[8] X. P. Li, Z. Liu, Z.-L. Shi, and H. C. So, "MUSIC with capped Frobenius norm: Efficient robust direction-of-arrival estimator," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 6, pp. 8090–8103, Dec. 2023.

[9] O. Bar-Shalom and A. J. Weiss, "DOA estimation using one-bit quantized measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 868–884, Jul. 2002.

[10] X. Huang and B. Liao, "One-bit MUSIC," *IEEE Signal Process Lett.*, vol. 26, no. 7, pp. 961–965, Jul. 2019.

[11] N. Guo, Z. Zheng, and W.-Q. Wang, "Robust DOA estimator against mutual coupling using one-bit sampling," *Circuits Systems Signal Process.*, vol. 43, no. 4, pp. 2626–2638, Jan. 2024.

[12] C.-L. Liu and Z.-M. Lin, "One-bit autocorrelation estimation with non-zero thresholds," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Toronto, Ontario, Canada, Jun. 2021, pp. 4520–4524.

[13] Y.-H. Xiao, L. Huang, D. Ramírez, C. Qian, and H. C. So, "Covariance matrix recovery from one-bit data with non-zero quantization thresholds: Algorithm and performance analysis," *IEEE Trans. Signal Process.*, vol. 71, pp. 4060–4076, Nov. 2023.

[14] C.-L. Liu and P. Vaidyanathan, "One-bit sparse array DOA estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, New Orleans, USA, Mar. 2017, pp. 3126–3130.

[15] S. Sedighi, B. Shankar, M. Soltanalian, and B. Ottersten, "One-bit DOA estimation via sparse linear arrays," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2020, pp. 9135–9139.

[16] S. Sedighi, M. Soltanalian, and B. Ottersten, "On the performance of one-bit DOA estimation via sparse linear arrays," *IEEE Trans. Signal Process.*, vol. 69, pp. 6165–6182, Oct. 2021.

[17] Y. Xiong, Z. Li, and F. Wen, "2D DOA estimation for uniform rectangular array with one-bit measurement," in *Proc. IEEE Sens. Array Multichannel Signal Process. Workshop*, Hangzhou, China, Jun. 2020, pp. 1–5.

[18] C. Zhou, Y. Gu, Z. Shi, and M. Haardt, "Direction-of-arrival estimation for coprime arrays via coarray correlation reconstruction: A one-bit perspective," in *Proc. IEEE Sens. Array Multichannel Signal Process. Workshop*, Hangzhou, China, Jun. 2020, pp. 1–4.

[19] C. Stöckle, J. Munir, A. Mezghani, and J. A. Nossek, "1-bit direction of arrival estimation based on compressed sensing," in *Proc. IEEE 16th Int.Workshop Signal Process. Adv. Wireless Commun.*, Stochholm, Sweden, Jun. 2015, pp. 246–250.

[20] Y. Gao, D. Hu, Y. Chen, and Y. Ma, "Gridless 1-bit DOA estimation exploiting SVM approach," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2210–2213, Oct. 2017.

[21] X. Meng and J. Zhu, "A generalized sparse Bayesian learning algorithm for 1-bit DOA estimation," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1414–1417, Jul. 2018.

[22] D. Chen, L. Tian, and C. Xu, "MMV-based OMP for DOA estimation with 1-bit measurement," in *Journal of Physics: Conference Series*, vol. 1550, no. 3, 2020, p. 032150.

[23] X. Huang, P. Xiao, and B. Liao, "One-bit direction of arrival estimation with an improved fixed-point continuation algorithm," in *Proc. Wirel. Commun. Signal Process.*, Hangzhou, China, Dec. 2018, pp. 1–4.

[24] P. Xiao, B. Liao, and N. Deligiannis, "Deepfpc: A deep unfolded network for sparse signal recovery from 1-bit measurements with application to DOA estimation," *Signal Process.*, vol. 176, p. 107699, Nov. 2020.

[25] L. Feng, L. Huang, Q. Li, Z.-Q. He, and M. Chen, "An off-grid iterative reweighted approach to one-bit direction of arrival estimation," *IEEE Trans. Veh. Technol.*, vol. 7, no. 6, pp. 8134–8139, Jan. 2023.

[26] M. Chen, Q. Li, X. P. Li, L. Huang, and M. Rihan, "One-bit DOA estimation for deterministic signals based on $\ell_{2,1}$-norm minimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 2, pp. 2438–2444, Apr. 2024.

[27] X.-P. Li, Z.-L. Shi, L. Huang, A. M.-C. So, and H. C. So, "ROCS: Robust one-bit compressed sensing with application to direction of arrival," *IEEE Trans. Signal Process.*, vol. 72, pp. 2407 – 2420, Apr. 2024.

[28] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, Apr. 2010.

[29] F. Wen, R. Ying, P. Liu, and T.-K. Truong, "Nonconvex regularized robust PCA using the proximal block coordinate descent algorithm," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5402–5416, Oct. 2019.

[30] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, pp. 629–654, Sep. 2008.

[31] 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 Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, Apr. 2013.

[32] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT press, 1987.

[33] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.

[34] R. He, W.-S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 261–275, Feb. 2014.

[35] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge Univ. Press, 2018.

[36] W.-J. Zeng, H.-C. So, and L. Huang, "ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, Sep. 2013.

[37] Y. He, F. Wang, Y. Li, J. Qin, and B. Chen, "Robust matrix completion via maximum correntropy criterion and half-quadratic optimization," *IEEE Trans. Signal Process.*, vol. 68, pp. 181–195, Nov. 2019.

[38] A. Beck, "On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes," *SIAM J. Optim.*, vol. 25, no. 1, pp. 185–209, Jan. 2015.

[39] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust Statistics: Theory and Methods (with R)*. John Wiley & Sons, 2019.

[40] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Math. Program.*, vol. 137, no. 1-2, pp. 91–129, Aug. 2013.

[41] D. A. Linebarger, R. D. DeGroat, and E. M. Dowling, "Efficient direction-finding methods employing forward/backward averaging," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2136–2145, Aug. 1994.

[42] W.-G. Tang, H. Jiang, and Q. Zhang, "One-bit gridless DOA estimation with multiple measurements exploiting accelerated proximal gradient algorithm," *Circuits Systems Signal Process.*, vol. 41, no. 2, pp. 1100–1114, Aug. 2021.

[43] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Programming*, vol. 146, no. 1-2, pp. 459–494, Jul. 2014.