Multiobjective Bilevel Evolutionary Approach for Off-Grid Direction-of-Arrival Estimation

Bai Yan[a,b], Qi Zhao[a,b], Jin Zhang[a,*], J. Andrew Zhang[c], Xin Yao[a]

[a] *Guangdong Provincial Key Laboratory of Brain-Inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*
[b] *School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China*
[c] *Global Big Data Technologies Center, University of Technology Sydney, NSW 2007, Australia*

Abstract

The source number identification is an essential step in direction-of-arrival (DOA) estimation. Existing methods may provide a wrong source number due to modeling errors caused by relaxing sparse penalties, especially in impulsive noise. This paper proposes a novel idea of simultaneous source number identification and DOA estimation to address this issue. We formulate a multiobjective off-grid DOA estimation model to realize this idea, by which the source number can be automatically identified together with DOA estimation. In particular, the source number is correctly exploited by the l_0 norm of impinging signals without relaxations, guaranteeing accuracy. We further design a multiobjective bilevel evolutionary algorithm to solve this model. The source number identification and sparse recovery are simultaneously optimized at the on-grid (lower) level. A forward search strategy is developed to further refine the grid at the off-grid (upper) level. This strategy does not need linear approximations and can eliminate the off-grid gap with low computational complexity. Simulation results demonstrate the outperformance of our method in terms of source number and root mean square error.

*Corresponding author
Email addresses: yanb@sustech.edu.cn (Bai Yan), zhaoq@sustech.edu.cn (Qi Zhao), zhangj4@sustech.edu.cn (Jin Zhang), Andrew.Zhang@uts.edu.au (J. Andrew Zhang), xiny@sustech.edu.cn (Xin Yao)

Keywords: Off-grid direction-of-arrival (DOA) estimation, evolutionary algorithm, multiobjective optimization, impulsive noise.

1. Introduction

Direction-of-arrival (DOA) estimation refers to finding the direction information of electromagnetic sources according to the outputs of receiving antennas that form an array. It is a crucial subject in array signal processing and is ubiquitous in radar, sonar, and wireless communications [1].

Many DOA estimation methods have been proposed. Classical subspace-type methods [2] such as MUSIC and ESPRIT, estimate DOAs by exploiting the eigenvalues decomposition on the sample covariance matrix [2]. They highly rely on equally-spaced measurements and the Gaussian noise assumption. However, non-Gaussian noise such as impulsive noise [3] may occur in practice. Unlike Gaussian noise, the probability density function of impulsive noise exhibits heavy-tailed property, i.e., the measurements contain outliers. As a result, the performance of subspace methods would degrade seriously.

Some variants of subspace methods are proposed to handle impulsive noise,. For example, maximum likelihood subspace methods [4][5] are developed for Gaussian mixture model-based noise. In [6][7], the fractional lower-order statistics instead of second-order covariance is adopted to cope with impulsive noise, but large samples are needed. Alternatively, the zero-memory nonlinear function [8] is applied to handle outliers. Inspired by the robust statistics, the l_p-MUSIC method considered the sample covariance matrix with l_p-norm minimization [9], but p is difficult to choose. In [10][11], an effective correntropy induced estimator is integrated into MUSIC to suppress outliers.

Recently, compressed sensing techniques proposed an alternative solution for DOA estimation, known as "sparse methods". Representatives include greedy algorithms [12], convex optimization approaches [13], and sparse Bayesian learning (SBL)-based methods [14][15]. Sparse methods [12][13][14] show many advantages over subspace-type, such as enhanced robustness to lower signal-to-

noise ratio (SNR), fewer snapshots, and highly correlated sources [16][17]. In particular, the angular domain is griddle into a finite set, and DOA estimation is simplified to select the right subsets. In practice, some sources may deviate from the preset grid, emerging the grid mismatch issue. This issue seriously degrades DOA estimation accuracy. To avoid this issue, gridless sparse methods [14][18][19] are developed by directly working in the continuous angular domain. However, they only work well with sufficiently separated sources and often need to solve a semidefinite programming problem by the interior point method, paying the price of high complexity [20]. Off-grid sparse methods are another type for fixing the grid gap [18][14][19]. They parameterize the gap into models, and estimate it by linear approximations [18][14], root-finding strategies [21][22], etc.

A few robust sparse methods [21][23][24] have been developed for impulsive noise. These methods model the measurement noise as a sum of sparse outlier noise and dense Gaussian noise to better learn the probability distribution of measurements. The off-grid SBL based method [21] falls into this type, but the maximum number of DOAs that can be estimated is reduced. In [23], the identify-and-reject strategy is presented to reject outliers, yet it works badly in cases with no outliers. Another robust method [24] identifies and corrects outliers during iterations, gaining improved performance. The three works highly rely on the assumption that outlier noise has a sparse structure. However, the outlier noise is not strictly sparse, leading to a performance loss [24].

In DOA estimation, identifying the source number is a crucial step. Particularly in subspace methods, the source number is an essential precondition for DOA estimation. Conventional techniques are based on information criterion such as Akaike information criteria (AIC), and minimum description length (MDL) [25][26], but they are only effective in Gaussian noise. By contrast, sparse methods do not need to know the source number. They use sparse penalties such as l_p-norm ($p \in (0, 1]$) sparse penalties, Laplace prior [14] or Gaussian scale mixtures [21] to approach the source number. These penalties are relaxations of the l_0 norm of impinging signals, which brings modeling error and may not perfectly represent the sparse distributions, leading to overestimated source

number [27]. Furthermore, how to identify the source number in impulsive noise via sparse methods has also not been investigated yet.

In summary, off-grid sparse methods exhibit many advantages compared to the subspace-based methods and gridless methods [16][17]. Nonetheless, they employ the suboptimal relaxed sparse penalties, which reportedly overestimates the source number and incurs a performance loss. In impulsive noise environments, the overestimation and performance loss would be even worse.

In this paper, we aim at simultaneous source number identification and DOA estimation in impulsive noise. We design a multiobjective bilevel evolutionary approach (MoBEA) with two significant innovations to realize this idea. Firstly, we build a multiobjective DOA estimation model, in which the source number and a robust fitting error are taken as two conflicting objectives. Unlike existing DOA estimation models, our model can automatically identify the source number together with DOA estimation. Moreover, our model perfectly exploits the source number via the original l_0-norm of impinging signals without relaxation, providing a more precise source number. Besides, different from existing models explicitly or implicitly working with the assumption of Gaussian noise, we incorporate the robustness metric "correntropy" [28] into the model, significantly improving the performance in impulsive noise.

Secondly, we design a multiobjective bilevel evolutionary algorithm to solve the proposed model. Our algorithm includes on-grid (lower) level and off-grid (upper) level optimization. At the on-grid level, we develop a population-based genetic algorithm to simultaneously identify the source number and recover the impinging signal matrix on a coarse grid. The diversified solutions in the population evolve various source numbers and communicate with each other, providing multiple pathways to the optima. Finally, the source number and active grid points are obtained after extracting the knee solution from on-hand solutions. At the off-grid level, we introduce a forward search strategy to refine the active grid points. This strategy does not need linear approximations and can efficiently eliminate the off-grid gap with low computational complexity. Overall, MoBEA's main contributions are:

- Multiobjective DOA estimation model. The proposed model can automatically identify the source number together with DOA estimation. Besides, it perfectly exploits the source number via the l_0-norm of impinging signals without relaxation, guaranteeing its accuracy.

- Multiobjective bilevel evolutionary DOA estimation algorithm. It solves the proposed model via alternate execution of two levels' optimization. A forward search strategy is introduced at the off-grid level, efficiently eliminating the off-grid gap with low computational complexity.

- Empirical validation of MoBEA's performance. Simulation results confirm the superiority of MoBEA in both source number identification and DOA estimation over state-of-the-art methods.

In this paper, bold-face letters represent vectors and matrices. $\mathbb{R}$ and $\mathbb{C}$ are the real and complex domain. $\mathbb{E}$ denotes the mathematical expectation. $\intercal$, $*$, and H denote transpose, conjugate, and conjugate transpose of a vector or matrix, respectively. $\mathbf{A}_{i,:}$ and $\mathbf{A}_{i,j}$ denote the i-th row and (i,j)-th element of matrix $\mathbf{A}$, respectively. $diag(\mathbf{w})$ is a diagonal matrix with $\mathbf{w}$ as the diagonal elements. In particular, $\mathbf{s}|_{\mathbf{e}}$ stands for the sub-vector of $\mathbf{s}$ with entries indexed by the set $I = \{i|\mathbf{e}_i = 1\}$. Similarly, $\mathbf{S}|_{\mathbf{e}}$ denotes the sub-matrix of $\mathbf{S}$ with rows indexed by the set $I = \{i|\mathbf{e}_i = 1\}$.

The rest of the paper includes the signal model in Section II, the proposed DOA estimation model and algorithm in Sections III and IV, simulation results in Section V, and conclusions in Section VI.

2. Signal Model

Consider K narrowband far-field sources $\dot{s}_1(t)$, $\dot{s}_2(t)$, ..., $\dot{s}_K(t)$ [1] impinging on a uniform linear array of M omnidirectional sensors from directions of $\dot{\boldsymbol{\theta}} =$

[1] The dot notation here is used to indicate that these variables are defined with the precondition "the true source number K is known".

$(\dot{\theta}_1, \dot{\theta}_2, ..., \dot{\theta}_K)^\intercal$. The array output [29] at time t is modeled as

$$\mathbf{y}(t) = \sum_{k=1}^{K} \mathbf{a}(\dot{\theta}_k)\dot{s}_k(t) + \mathbf{n}(t) = \mathbf{A}(\dot{\boldsymbol{\theta}})\dot{\mathbf{s}}(t) + \mathbf{n}(t), \tag{1}$$

where $\mathbf{y}(t) = (y_1(t), ..., y_M(t))^\intercal$ is the array output, $\dot{\mathbf{s}}(t) = (\dot{s}_1(t), ..., \dot{s}_K(t))^\intercal$ are the signal waveforms, $\mathbf{A}(\dot{\boldsymbol{\theta}}) = \left(\mathbf{a}(\dot{\theta}_1), ..., \mathbf{a}(\dot{\theta}_K)\right)$ is the array manifold matrix, $\mathbf{a}(\dot{\theta}_k) = \left(1, e^{-j2\pi d/\lambda \sin(\dot{\theta}_k)}, ..., e^{-j2\pi(M-1)d/\lambda \sin(\dot{\theta}_k)}\right)^\intercal$ contains the time delay of the k-th signal received at each sensor, λ is the wavelength of sources, d is the spacing distance between adjacent sensors, $\mathbf{n}(t) = (n_1(t), ..., n_M(t))^\intercal$ is unknown noise. When T snapshots are collected, the array output [29] is

$$\mathbf{Y} = \mathbf{A}(\dot{\boldsymbol{\theta}})\dot{\mathbf{S}} + \mathbf{N}, \tag{2}$$

where $\dot{\mathbf{S}} = (\dot{\mathbf{s}}(1), ..., \dot{\mathbf{s}}(T))$, $\mathbf{Y} = (\mathbf{y}(1), ..., \mathbf{y}(T))$, and $\mathbf{N} = (\mathbf{n}(1), ..., \mathbf{n}(T))$. Given $\mathbf{Y}$ and mapping $\dot{\boldsymbol{\theta}} \to \mathbf{A}(\dot{\boldsymbol{\theta}})$, the goal is to find DOAs with unknown K.

To accomplish this goal, existing grid-based sparse methods cast DOA estimation as a sparse recovery problem [12]. The angular range $[-\pi/2, \pi/2]$ is divided into an equi-spaced grid $\boldsymbol{\theta}_0 = (\theta_1, ..., \theta_N)^\intercal$ with the assumption that $N \gg K$, where N denotes the total number of grid points, and $r = \theta_{i+1} - \theta_i$ is the grid interval. If all the sources fall exactly on this grid, it would be an *on-grid* case. The array output $\mathbf{Y}$ for the on-grid model [12] is formulated as

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{S} + \mathbf{N}, \tag{3}$$

where the impinging signal matrix $\mathbf{S}$ is the extension of $\dot{\mathbf{S}}$ from $\dot{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$, with $\mathbf{S}$ being row-sparse, i.e., all columns of $\mathbf{S}$ are sparse and share the same support. However, some sources may deviate from the predefined grid (known as the *off-grid* case), as shown in Fig. 1. In this case, the sparse coefficient $\mathbf{S}$ is not sparse any more. Thus, the off-grid gap may lead to estimation degradation.

To alleviate the off-grid gap, the gap between true DOAs and the initial grid (denoted as grid mismatch $\boldsymbol{\zeta}$) is usually parameterized into the on-grid model (3). As a result, the array output for the off-grid model [13] is cast as

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta})\mathbf{S} + \mathbf{N}. \tag{4}$$

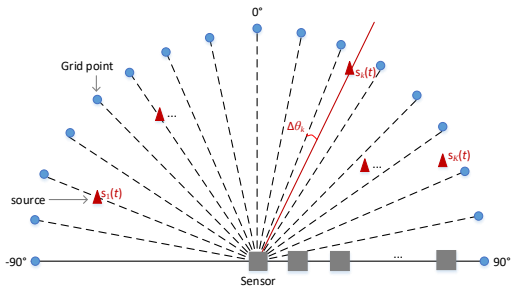


Figure 1: Off-grid case. Dotted lines stand for predefined grids. Red triangular symbols correspond to sources.

The grid mismatch and impinging signal matrix are estimated by solving the following regularization optimization problem [13]:

$$\min_{\boldsymbol{\zeta},\mathbf{S}} \ \rho\|\mathbf{S}\|_{2,0}+\frac{1}{2}\|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta}_0+\boldsymbol{\zeta})\mathbf{S}\|_F^2, s.t. \ -\frac{r}{2}<\zeta_i\leqslant\frac{r}{2}, \ i=1,...N, \quad \text{(P1)}$$

where ρ is a balancing parameter. $\|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta}_0+\boldsymbol{\zeta})\mathbf{S}\|_F^2$ is a fitting error term, denoted as "measurement error"; $\mathbf{S}$ is a row-sparse matrix whose i-th row corresponds to a possible source impinging on the array at $(\boldsymbol{\theta}_0+\boldsymbol{\zeta})_i$; $\|\mathbf{S}\|_{2,0}$ stands for the source number, which equals to the number of nonzero rows of $\mathbf{S}$:

$$\|\mathbf{S}\|_{2,0}=\|\left(\|\mathbf{S}_{1,:}\|_2,\|\mathbf{S}_{2,:}\|_2,...,\|\mathbf{S}_{N,:}\|_2\right)\|_0.$$

$\|\mathbf{S}\|_{2,0}$ is a l_0-norm penalty. It is the proper and exact sparsity-enforcing penalty, which plays a prominent role in identifying the source number.

3. Proposed Multiobjective DOA Estimation Model

Problem (P1) is NP-hard because of the l_0 norm[30]. Existing off-grid methods [12][14][15] relaxed the l_0 norm to other suboptimal penalties. Such relaxation brings modeling error, which may provide some spurious DOAs and degrade the localization performance[27]. Also, the performance is susceptible to the balancing parameter (i.e., ρ in (P1)), which is hard to tune.

To address the above issues, we naturally model the regularization optimization problem (P1) as a multiobjective optimization problem (MOP). Our

model holds two significant advantages: 1) it allows automatically identifying the source number together with DOA estimation, and 2) it perfectly exploits the source number via the l_0-norm of impinging signals without relaxation, guaranteeing accuracy. In detail, our model simultaneously minimizes the two conflicting objectives:

$$\begin{aligned} &\min_{\boldsymbol{\zeta},\mathbf{S}} \ \left(\|\mathbf{S}\|_{2,0}, \|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta}_0+\boldsymbol{\zeta})\mathbf{S}\|_F^2\right), \\ &s.t. \ -\frac{r}{2} < \zeta_i \leqslant \frac{r}{2}, \ i=1,...N, \end{aligned} \tag{P2}$$

where $\|\mathbf{S}\|_{2,0}$ and $\|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta}_0+\boldsymbol{\zeta})\mathbf{S}\|_F^2$ are the source number and measurement error, respectively.

The MOP (P2) inherently involves two problems. 0ne is the sparse recovery of the impinging signal matrix $\mathbf{S}$ with given grid, where $\|\mathbf{S}\|_{2,0}$ denotes the source number. The other is grid refinement with given $\mathbf{S}$. It should be noted that, compared to source number identification by sparse recovery, perceiving true DOAs is more critical. Therefore, the decision-maker of the grid refinement problem should have complete knowledge about sparse recovery, while the sparse recovery problem only observes the decisions of grid refinement. Therefore, we cast the MOP (P2) as a bilevel MOP:

$$\begin{aligned} &\min_{\boldsymbol{\zeta},\mathbf{S}} \ \|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta}_0+\boldsymbol{\zeta})\mathbf{S}\|_F^2, \\ &s.t. \ -\frac{r}{2} < \zeta_i \leqslant \frac{r}{2}, \ i=1,...N, \\ &\quad \mathbf{S} \in \arg\min_{\mathbf{S}} \ \left(\|\mathbf{S}\|_{2,0}, \|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta}_0+\boldsymbol{\zeta})\mathbf{S}\|_F^2\right), \end{aligned} \tag{P3}$$

where the first and the third rows are the off-grid (upper) level and the on-grid (lower) level problems, respectively. The on-grid level simultaneously identifies the source number and recovers $\mathbf{S}$ with a given grid. Based on the decisions of the on-grid level, the off-grid level minimizes the measurement error to refine the grid and update $\mathbf{S}$.

The bilevel MOP (P3) explicitly works in Gaussian noise. In practice, the impulsive noise may exist, bringing great challenges for accurate estimation. To cope with this, we incorporate a robust metric "*correntropy*" [28] to reduce the

detrimental effect of the impulsive noise. Correntropy is a local and nonlinear similarity measure in a feature space [28]. Given two arbitrary vectors $\mathbf{x} = (x_1, ..., x_M)^\intercal$ and $\mathbf{z} = (z_1, ..., z_M)^\intercal$, the correntropy is approximated as [28]

$$\begin{aligned} V(\mathbf{x}, \mathbf{z}) :=& \mathrm{E}[\kappa_\sigma(\mathbf{x} - \mathbf{z})] \approx \frac{1}{M} \sum_{i=1}^{M} \kappa_\sigma(x_i - z_i), \\ \kappa_\sigma(\gamma) =& \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\gamma\gamma^*}{2\sigma^2}\right), \end{aligned} \tag{5}$$

where $\kappa_\sigma(\gamma)$ is the Gaussian kernel function with a kernel size σ. Based on correntropy, correntropy-based loss function (CLF) [31] is defined as

$$V_{\mathrm{CLF}}(\mathbf{x}, \mathbf{z}) = 1 - \frac{1}{M} \sum_{i=1}^{M} \exp\left(-\frac{(x_i - z_i)(x_i - z_i)^*}{2\sigma^2}\right). \tag{6}$$

This function is related to Welsch's cost function [28]. Compared to the l_2-norm error term of the bilevel MOP (P3), V_{CLF} increases much slower and is bounded, thus large noise outliers have a limited effect on V_{CLF}. Hence, it has better robustness to impulsive noise. With incorporating V_{CLF}, we give the final bilevel multiobjective optimization model

$$\begin{aligned} &\min\ F(\mathbf{S}, \boldsymbol{\zeta}) = V_{\mathrm{CLF}}\left(\mathbf{Y}, \mathbf{A}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta})\mathbf{S}\right), \\ &s.t.\ -\frac{r}{2} < \zeta_i \leqslant \frac{r}{2},\ i = 1, ...N, \\ &\mathbf{S} \in \arg\min_{\mathbf{S}} \mathbf{f}(\mathbf{S}) = \left(\|\mathbf{S}\|_{2,0}, V_{\mathrm{CLF}}(\mathbf{Y}, \mathbf{A}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta})\mathbf{S})\right), \end{aligned} \tag{P4}$$

where the first and the third rows are the off-grid level and on-grid level problems, respectively. The second row is the grid mismatch-related constraint. Compared to the regularization problem (P1), no balancing parameter is needed. The sparsity of impinging signals is perfectly captured by the l_0 norm term without evoking relaxations. Hence the source number is accurately identified.

4. Proposed Multiobjective BiLevel Evolutionary DOA Estimation Algorithm

We design a multiobjective bilevel evolutionary algorithm to solve the proposed bilevel MOP (P4). The designed algorithm has three key features: 1)

it simultaneously identifies the source number and recovers the impinging signal matrix on a coarse grid at the on-grid level; 2) A forward search strategy is developed at the off-grid level, which eliminates the off-grid gap with lower computational complexity than other grid refinement strategies [18][19][22]; and 3) the population-based evolutionary algorithm evolves diversified search pathways to the optima, promoting the effectiveness and efficiency of the algorithm.

The workflow of the proposed algorithm is shown in Fig. 2, with the pseudocode and variable definitions exhibited in Algorithm 1 and Table 1, respectively. It starts with initialization, following by a bilevel optimization. At the on-grid level, an improved framework based on NSGA-II [32] is developed to simultaneously identify the source number and recover the impinging signal matrix $\mathbf{S}$. Instead of directly searching the complex-valued $\mathbf{S}$, we encode $\mathbf{S}$ into a binary vector for simpleness. After evolutionary search and selection, the knee solution is extracted to represent the active grid points, where "active" means a source located at that grid point. After that, a forward search strategy is developed and executed at the off-grid level for refining the grid represented by the knee solution. The population is then updated. When the stopping criterion is satisfied, the active grid points are the estimated DOAs. Key components of the proposed algorithm are detailed below.

4.1. Chromosome Encoding and Decoding

$(\mathbf{S}, \boldsymbol{\zeta})$ represents the solution of the proposed model (P4), but it is non-trivial to execute genetic operators on the complex-valued impinging signal matrix $\mathbf{S}$. Therefore, we search the locations of the nonzero rows of $\mathbf{S}$ instead and then recover these nonzero values. In detail, we encode $(\mathbf{S}, \boldsymbol{\zeta})$ as $(\mathbf{e}, \boldsymbol{\zeta})$, where $\mathbf{e}$ is a binary vector with "1" and "0" denoting the corresponding row of $\mathbf{S}$ being nonzero and zero entries, respectively. For unity, we denote $\mathbf{e}$ as "active set".

To recover $\mathbf{S}$ from $\mathbf{e}$, we develop a decoding method improved from the correntropy matching pursuit (CMP) algorithm [31]. For clarity, we briefly introduce CMP before describing the decoding process.

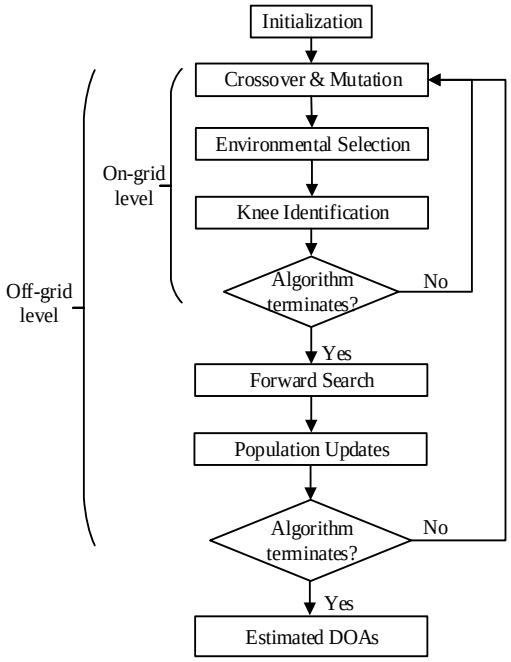


Figure 2: Framework of MoBEA.

4.1.1. Correntropy matching pursuit

CMP is a matching pursuit method for sparse recovery in impulsive noise [31]. In compressive sensing, let the measurement being $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}$ with the assumption that the dictionary $\mathbf{A}$ and measurements $\mathbf{y}$ are both known, CMP solves a correntropy loss minimization problem to recover $\mathbf{s}$

$$\mathbf{s} = \arg\min_{\mathbf{s}} V_{\mathrm{CLF}}(\mathbf{y}, \mathbf{A}\mathbf{s}) = 1 - \frac{1}{M}\sum_{i=1}^{M} \exp\left(-\frac{z_i z_i^*}{2\sigma^2}\right), \tag{7}$$

Algorithm 1 Pseudocode of MoBEA

Input: $\mathbf{A}$, $\mathbf{Y}$, $\boldsymbol{\theta}_0$, $\boldsymbol{\zeta}^0 \leftarrow 0^{N\times 1}$

Output: $\boldsymbol{\theta}$

$G = 0$;

$\left(\mathbf{P}^G, \mathfrak{S}^G, \ddot{\mathbf{e}}^G, \ddot{\mathbf{S}}^G\right)$ =*Initialization*$\left(\mathbf{A}, \mathbf{Y}, \boldsymbol{\zeta}^0\right)$;

while "the algorithm does not teminate" **do**

  $\left(\mathbf{P}^G, \mathfrak{S}^G, \ddot{\mathbf{e}}^G, \ddot{\mathbf{S}}^G\right)$ =*On_grid*$\left(\mathbf{P}^G, \mathfrak{S}^G, \ddot{\mathbf{e}}^G, \ddot{\mathbf{S}}^G\right)$;

  $\boldsymbol{\zeta}^{G+1}$ = *Forward_search* $\left(\mathbf{P}^G, \mathfrak{S}^G, \ddot{\mathbf{e}}^G, \ddot{\mathbf{S}}^G\right)$, where $\mathbf{P}^G = \left(\mathbf{E}^G, \boldsymbol{\zeta}^G\right)$;

  /*Population updates*/

  $\mathbf{P}^{G+1} \leftarrow \left(\mathbf{E}^G, \boldsymbol{\zeta}^{G+1}\right)$;

  $\mathfrak{S}^{G+1}$ =*Decoding*$\left(\mathbf{E}^G, \ddot{\mathbf{S}}^G, \boldsymbol{\zeta}^{G+1}, \sigma^G\right)$;

  $G = G + 1$;

end while

$\boldsymbol{\theta} = \left(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}^G\right)|_{\ddot{\mathbf{e}}^G}$.

where $\mathbf{z} = \mathbf{y} - \mathbf{A}\mathbf{s}$ is the residual. With the half-quadratic theory [33], problem (7) can be reformulated as a weighted least square problem and solved by [31]:

$$\sigma^{l+1} = \left(\frac{1}{2M}\|\mathbf{y} - \mathbf{A}\mathbf{s}^l\|_2^2\right)^{\frac{1}{2}}, \tag{8}$$

$$w_i^{l+1} = g_\sigma\left(y_i - A_{i,:}\mathbf{s}^l\right),\ i = 1, 2, ...M, \tag{9}$$

$$\begin{aligned}\mathbf{s}^{l+1}|_{\mathbf{e}} &= \underset{\mathrm{supp}(\mathbf{s})\in\{i|\mathbf{e}_i=1\}}{\arg\min} \|\sqrt{diag(\mathbf{w}^{l+1})}(\mathbf{y} - \mathbf{A}\mathbf{s})\|_2^2 \\ &= \left(\mathbf{A}^\intercal diag(\mathbf{w}^{l+1})\mathbf{A}\right)^{-1}\mathbf{A}^\intercal diag(\mathbf{w}^{l+1})\mathbf{y},\end{aligned} \tag{10}$$

where $\mathbf{s}^{l+1}|_{\mathbf{e}}$ is the sub-vector of $\mathbf{s}^{l+1}$ with entries indexed by $I = \{i|\mathbf{e}_i = 1\}$, $g_\sigma(p) = \exp\left(-\frac{pp^*}{2\sigma^2}\right)$, and $\mathrm{supp}(\mathbf{s})$ denotes the nonzero entries of $\mathbf{s}$. $\mathbf{w}$ is a weight vector. It indicates the importance of measurements, i.e., the small coefficients in $\mathbf{w}$ suppress the severely contaminated measurements, and the large coefficients enable to preserve the clean measurements. $\mathbf{w}$ can improve CMP to identify the true atoms, thus boosting the recovery performance [31].

Table 1: Notations in MoBEA.

Variable	Description
G	Outer generation number
lt	Inner generation number at the on-grid level
ut	Inner generation number in forward search
$\boldsymbol{\theta}_0$	Initial grid
$\boldsymbol{\zeta}$	Grid mismatch
$\mathbf{E}^G = \{\mathbf{e}_i^G\}_{i=1}^{\bar{N}}$	Active sets at G-th generation
$\mathfrak{S}^G = \{\mathbf{S}_i^G\}_{i=1}^{\bar{N}}$	Set of signal matrices decoded from $\mathbf{E}^G$
$\mathbf{P}^G = \left(\mathbf{E}^G, \boldsymbol{\zeta}^G\right)$	Population at G-th generation
$\ddot{\mathbf{e}}^G$	Active set of the knee solution at G-th generation
$\ddot{\mathbf{S}}^G$	Impinging signal matrix of the knee solution at G-th generation

4.1.2. Decoding

We extend the CMP to the case of multiple measurement vectors as the decoding process of the proposed algorithm (see Algorithm 2). In Algorithm 2, $\mathbf{W}$ is a weight matrix with each column being the weight vector at each snapshot. To decode the active set, we first compute the weight matrix based on the impinging signal matrix of the knee solution $\ddot{\mathbf{S}}$, where the knee solution will be detailed in Section 4.2. Then, the impinging signal matrix is recovered by solving a series of reweighted least square problems.

In the decoding process, the choice of kernel size σ is essential. To balance the recovery accuracy and convergence speed, we employ the kernel annealing method [28] instead of (8) due to its more excellent performance enhancement found in [34]. Specifically, the kernel size is defined by

$$\sigma(G) = \sigma_{max} \exp(-\nu G) + \sigma_{min}, \tag{11}$$

where G is the generation counter of the outer cycle, σ_{min}=0.03, σ_{max} is calculated by $0.5(abs(\mathbf{Y})_{0.875} - abs(\mathbf{Y})_{0.125}) - \sigma_{min}$ with $abs(\mathbf{Y})_l$ being the l-th

Algorithm 2 *Decoding*

Input: $\mathbf{A}$, $\mathbf{Y}$, $\mathbf{E}$, $\ddot{\mathbf{S}}$, $\boldsymbol{\zeta}$, σ, $\mathfrak{S} = \varnothing$

Output: $\mathfrak{S}$

1: $W_{i,j} = g_\sigma\left((\mathbf{Y} - \mathbf{A}\ddot{\mathbf{S}})_{i,j}\right)$, where $g_\sigma(p) = \exp\left(-\frac{pp^*}{2\sigma^2}\right)$, $i = 1, ..., M$, $j = 1, ..., T$;

2: **for** each active set $\mathbf{e} \in \mathbf{E}$ **do**

3: $\quad \mathbf{S} \leftarrow 0^{N \times T}$;

4: $\quad$ **for** $j = 1 : T$ **do**

5: $\quad\quad \mathbf{S}_{:,j}|_{\mathbf{e}} = \left(\mathbf{A}^\intercal diag(\mathbf{W}_{:,j})\mathbf{A}\right)^{-1} \mathbf{A}^\intercal diag(\mathbf{W}_{:,j})\mathbf{Y}_{:,j}$;

6: $\quad$ **end for**

7: $\quad$ put $\mathbf{S}$ into $\mathfrak{S}$;

8: **end for**

quantile of $\{|Y_{i,j}|\}^{M \times T}$, and $\nu = 2 \times 10^{-4}$ is the decay rate.

4.2. Initialization

We leverage the statistical correlation knowledge to provide a preliminary choice for the active set to enhance search efficiency. Since a uniform linear array can identify up to $M-1$ sources, the condition $K \leq M-1$ can be considered as a priori [35]. In noise-free cases, the sources can be roughly located by selecting the K atoms that are most correlated with the measurements. Motivated by this and consider the atoms' rank distortion caused by impulsive noise, we give priority to the atoms that have larger correlations, e.g., those who rank the top $2M$. In this way, some proper atoms are involved with great probability, thus improving the search efficiency. The process is shown in Line 1-5 of Algorithm 3. For the missing atoms, they can be further searched in the mutation step of the on-grid level.

After that, the knee solution is selected from the current population for preparing the subsequent decoding. In each decoding process, a solution is required to estimate the weight matrix. Here we choose the knee solution based on the following considerations. The knee solution is commonly an interesting

Algorithm 3 *Initialization*

Input: $\mathbf{A}$, $\mathbf{Y}$, $\boldsymbol{\zeta}^0$

Output: $\mathbf{P}^0$, $\mathfrak{S}^0$, $\ddot{\mathbf{e}}^0$, $\ddot{\mathbf{S}}^0$

1: $\tau = \sum_{t=1}^{T} | < \mathbf{y}(t), A_{:,j} > |$, $j = 1, ..., N$;

2: $\boldsymbol{\Lambda} \leftarrow$ save the indices of $2M$ highest correlation in τ;

3: $\mathbf{E}^0 \leftarrow$ randomly choose no more than $M-1$ indices from $\boldsymbol{\Lambda}$ to construct $\bar{N}$ active sets;

4: $\mathbf{P}^0 = \left(\mathbf{E}^0, \boldsymbol{\zeta}^0\right)$;

5: $\mathfrak{S}^0 = Decoding\left(\mathbf{E}^0, 0^{N\times T}, \boldsymbol{\zeta}^0, \sigma\right)$; // Algorithm 2

6: $\ddot{\mathbf{e}}^0, \ddot{\mathbf{S}}^0 \leftarrow Knee\_Identification\left(\mathbf{P}^0, \mathfrak{S}^0\right)$.

point along the Pareto front (PF), and it provides a good trade-off between the two objectives [36]. Therefore, it is expected to use the knee solution to obtain reasonable weights and bring performance enhancement. Even when the knee solution fails to be the most satisfactory, it is still a Pareto nondominate solution from the perspective of multiobjective optimization. Here the kink method [37] is employed to identify the knee solution, where the solution with the most significant slope variance over the PF is the knee solution.

4.3. On-grid Level Optimization

At the on-grid level, we develop a framework based on NSGA-II [32] to simultaneously identify the source number and recover the signals, where the grid mismatch acts as parameters. The procedure is exhibited in Algorithm 4. The active set $\mathbf{E}^{lt}$ firstly undergoes the one-point crossover, and the bitwise mutation operations [38] to create offspring solution set $\hat{\mathbf{E}}^{lt}$. $\hat{\mathbf{E}}^{lt}$ is then decoded according to Algorithm 2 to recover the corresponding signals $\mathfrak{S}^{lt}$. After that, the environmental selection operator of NSGA-II [32] is executed to select $\bar{N}$ elite solutions for the next generation. Lastly, the knee solution is identified. This solution will estimate the weight matrix in the next decoding and identify the active grid points for off-grid level refinement.

Algorithm 4 *On-Grid Level*

Input: $\mathbf{P}^{lt} = \left(\mathbf{E}^{lt}, \boldsymbol{\zeta}\right)$, $\mathfrak{S}^{lt}$, $\ddot{\mathbf{e}}^{lt}$, $\ddot{\mathbf{S}}^{lt}$

Output: $\mathbf{P}^{lt}$, $\mathfrak{S}^{lt}$, $\ddot{\mathbf{e}}^{lt}$, $\ddot{\mathbf{S}}^{lt}$

1: **for** "$lt = 1 : lt_{max}$" **do**

2: $\quad \hat{\mathbf{E}}^{lt} = Crossover\&Mutation(\mathbf{E}^{lt})$;

3: $\quad \hat{\mathfrak{S}}^{lt} = Decoding\left(\hat{\mathbf{E}}^{lt}, \ddot{\mathbf{S}}^{lt}, \boldsymbol{\zeta}, \sigma\right)$; // Algorithm 2

4: $\quad \mathbf{Q}^{lt} = \left(\hat{\mathbf{E}}^{lt}, \boldsymbol{\zeta}\right)$;

5: $\quad \mathbf{P}^{lt+1}, \mathfrak{S}^{lt+1} \leftarrow Environmental\_Selection\left(\mathbf{P}^{lt} \cup \mathbf{Q}^{lt}, \mathfrak{S}^{lt} \cup \hat{\mathfrak{S}}^{lt}\right)$;

6: $\quad \ddot{\mathbf{e}}^{lt+1}, \ddot{\mathbf{S}}^{lt+1} \leftarrow Knee\_Identification\left(\mathbf{P}^{lt+1}, \mathfrak{S}^{lt+1}\right)$;

7: **end for**

4.4. Forward Search

We propose a straightforward forward search strategy at the off-grid level for solving the grid mismatch. Unlike existing grid refinement strategies, this strategy does not require modeling approximation, achieving higher DOA estimation quality. To save computational cost, the grid points corresponding to the knee solution's nonzero entries are viewed as "active" and allowed to be refined. The perturbation direction of each active grid point where to decrease the correntropy-based loss function is obtained separately. Then all the active grid points are perturbed together along with the obtained directions.

The procedure of the forward search is shown in Algorithm 5. The indices of active grid points are saved into I according to the active set of the knee solution $\ddot{\mathbf{e}}$. Subsequently, the perturbation directions of the active grid points are detected and saved into β (Lines 3 to 10 of Algorithm 5): for each active grid point, a random perturbation direction is given, and it is accepted only if the corresponding correntropy-based loss function decreases; otherwise, the perturbation direction is switched to the opposite direction or zero. Then the active grid points are perturbed together along with β with a stepsize μ (Line 13), where μ equals $r/100$ to provide a high resolution. The perturbation is repeated several times until ut reaches its maximum value ut_{max} or the correntropy-based loss function no longer decreases.

Algorithm 5 *Forward Search*

Input: $\mathbf{P} = (\mathbf{E}, \boldsymbol{\zeta})$, $\mathfrak{S}$, $\ddot{\mathbf{e}}$, $\ddot{\mathbf{S}}$

Output: $\boldsymbol{\zeta}$

Find the indices of active grid points: $I = \{i | \ddot{\mathbf{e}}_i = 1\}$;
for each index $i \in I$, **do**
    give a random perturbation direction: $\beta_i = -1$ *or* $+1$;
    $\hat{\boldsymbol{\zeta}}_i = \boldsymbol{\zeta}_i + \mu\beta_i$;
    if $F\left(\ddot{\mathbf{S}}, \hat{\boldsymbol{\zeta}}\right) > F\left(\ddot{\mathbf{S}}, \boldsymbol{\zeta}\right)$ **then**
        $\beta_i = -\beta_i$;
    else if $F\left(\ddot{\mathbf{S}}, \hat{\boldsymbol{\zeta}}\right) = F\left(\ddot{\mathbf{S}}, \boldsymbol{\zeta}\right)$ **then**
        $\beta_i = 0$;
    end if
end for
$\boldsymbol{\zeta}^1 = \boldsymbol{\zeta}$;
for "$ut = 1 : ut_{max}$" **do**
    $\boldsymbol{\zeta}^{ut+1} = \boldsymbol{\zeta}^{ut} + \mu\beta$;
    if $F\left(\ddot{\mathbf{S}}, \boldsymbol{\zeta}^{ut+1}\right) \geqslant F\left(\ddot{\mathbf{S}}, \boldsymbol{\zeta}^{ut}\right)$ **then**
        $\boldsymbol{\zeta} \leftarrow \boldsymbol{\zeta}^{ut}$
        break;
    end if
end for

Remark 1. *For any active grid point, it is finally accepted if it is within the set of $-r/2 < \boldsymbol{\zeta}_i \leqslant r/2$. Otherwise, it is rejected, and the corresponding grid point remains unchanged.*

4.5. Computational Complexity

The main computational complexity of the proposed algorithm lies in the decoding process, i.e., Line 5 of algorithm 3, Line 3 of Algorithm 4 and Line 22 of Algorithm 5. For each solution, the decoding complexity is $T \times O(N^3)$ in the worst case, where T and N are the number of snapshots and the number of grid

points, respectively. Thus, the algorithm's complexity is $(lt+1)\bar{N}T \times O(N^3)$, where lt is the number of inner iterations of the on-grid level, $\bar{N}$ is the population size. It can be seen that the number of grid points has a significant impact on computational complexity. In practical applications, the decoding is suggested to be implemented in parallel to reduce the execution time.

5. Experimental Study

In this section, we first conduct simulations to investigate the effectiveness of the key operators in the proposed MoBEA. Then, we compare MoBEA with the robust state-of-the-art algorithms to analyze MoBEA's overall performance.

5.1. Experimental Settings

5.1.1. Comparing Algorithms

We compare MoBEA with four state-of-the-art robust algorithms, i.e., l_p-MUSIC [9], MCC-MUSIC [11], Bayes-optimal [21], and Fast-alternating [21]. They are introduced below.

- l_p-MUSIC: A representative robust subspace-based algorithm that adopts the l_p-norm of the residual fitting error matrix for subspace decomposition.
- MCC-MUSIC: A state-of-the-art robust subspace-based algorithm that estimates the signal subspace by solving an optimization problem under the maximum correntropy criterion.
- Bayes-optimal: A state-of-the-art off-grid SBL-based algorithm that models the measurement noise as the mix of Gaussian noise and outliers.
- Fast-alternating: A state-of-the-art off-grid SBL-based algorithm that is a fast execution version of the Bayes-optimal algorithm.

In simulations, we also compare the proposed forward search strategy to Taylor expansion [18], to investigate the effectiveness of our strategy in grid refinement. In Taylor expansion, the manifold matrix is approximated by the first

order Taylor approximation [18]: $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta}_0 + \boldsymbol{\zeta}) \approx \mathbf{A}(\boldsymbol{\theta}_0) + \mathbf{A}'(\boldsymbol{\theta}_0) diag(\boldsymbol{\zeta})$, where $\boldsymbol{\theta}_0$ and $\boldsymbol{\zeta}$ are the initial grid and grid mismatch, $\mathbf{A}'(\boldsymbol{\theta}_0)$ is the first derivative of $\mathbf{A}(\boldsymbol{\theta}_0)$ with respective to $\boldsymbol{\theta}_0$. Similar to [18], $\boldsymbol{\zeta}$ can be obtained by letting the derivative of $V_{\mathrm{CLF}}\left(\mathbf{Y}, (\mathbf{A}(\boldsymbol{\theta}_0) + \mathbf{A}'(\boldsymbol{\theta}_0) diag(\boldsymbol{\zeta}))\mathbf{S}\right)$ with respect to $\boldsymbol{\zeta}$ be zero.

5.1.2. Simulation Benchmarks

We use various simulations to testify algorithms' performance. The simulation benchmarks are detailed in Table 2, where T is the number of snapshots, c_2 is the probability parameter of noise outliers in Gaussian mixture model (GMM) [4], α is the characteristic exponent in symmetric α-stable distribution (SαS) noise [39], and M is the number of receiving antennas.

In all simulations, we consider a small number of uncorrelated sources impinging on a uniform linear array of sensors with an inter-sensor spacing of $d = \lambda/2$. The number of receiving sensors is $M = 8$. Two widely-used PDF models, i.e., GMM [4], and SαS [39], are considered to model the impulsive noise. All sources are assumed to possess equal power η_s^2.

The PDF of a two-term GMM noise $\mathbf{n}(t)$ can be described as

$$p_{\mathbf{n}}(x) = \sum_{i=1}^{2} \frac{c_i}{\pi \eta_i^2} \exp\left(-\frac{|x|^2}{\eta_i^2}\right), \tag{12}$$

where $0 \leq c_i \leq 1$ and η_i^2 are the probability and variance of the i-th term with $c_1 + c_2 = 1$, respectively. Assume $\eta_2^2 = 100\eta_1^2$ and $c1 > c2 > 0$, the SNR noise model can be viewed as the large noise outliers of variance η_2^2 with a smaller probability c_2 embedded into Gaussian noise of variance η_1^2 with a larger probability c_1. According to [21], the SNR is simplified as $\mathrm{SNR} = \eta_s^2/\eta_1^2$.

The symmetric SαS distribution with zero-location is used to model the noise. Its characteristic function is defined as

$$\varphi(x) = \exp(-\gamma^{\alpha}|x|^{\alpha}), \tag{13}$$

where $0 < \alpha \leq 2$ denotes the characteristic exponent that indicates the tail of the distribution, and γ is the scale. When $\alpha = 2$ and $\alpha = 1$, the SαS reduces to the Gaussian distribution and the Cauchy distribution, respectively. The

Table 2: Simulation benchmarks.

Name	Feature	Settings
Simulation 1	With different GSNRs in SαS noise	Three uncorrelated sources from $-9.7°$, $6.8°$, and $12.7°$. $T = 20$. $\alpha = 1.4$. GSNRs varies from -10dB to 15dB.
Simulation 2	With different grid intervals in GMM noise	Two uncorrelated sources from $-1.6°$ and $13.2°$. $T = 20$. SNR = 10dB, $c_2 = 0.1$. Grid interval varies from $2°$ to $10°$.
Simulation 3	With different SNRs or GSNRs	Three uncorrelated sources from $-2.7°$, $5.8°$, and $20.2°$. $T = 20$. $c_2 = 0.1$. $\alpha = 1.4$. SNR or GSNR varies from -10dB to 15dB.
Simulation 4	With different grid intervals in SαS noise	Three uncorrelated sources from $-19.7°$, $6.8°$, and $32.7°$. $T = 20$. $\alpha = 1.4$. GSNR = 10dB. Grid interval varies from $2°$ to $10°$.
Simulation 5	With different angular separations in GMM noise	Two uncorrelated sources, one from $-10.8°$, and another varies from $-8.8°$ to $-0.8°$. $T = 30$. $c_2 = 0.1$. SNR = 10dB.
Simulation 6	With different number of snapshots in SαS noise	Three uncorrelated sources randomly chosen from $-2.7°$, $5.7°$ and $20.2°$ impinging on the uniform linear array. $\alpha = 1.4$. GSNR = 10dB. T varies from 20 to 100.
Simulation 7	Under Gaussian noise	Three uncorrelated sources from $-2.7°$, $5.7°$, and $20.2°$ from a uniform linear array with $M = 8$. T varies from 1 to 50. SNR = 5dB.

smaller value of α, the more impulsive noise is. Since the closed-form PDF of SαS does not exist when $\alpha \neq 2$ and $\alpha \neq 1$ [39], the SNR becomes meaningless. Instead, the generalized SNR (GSNR) [9] is reformulated as GSNR $= \eta_s^2/\gamma^\alpha$.

5.1.3. Performance Metrics

Since the Bayesian-based algorithms and the proposed MoBEA do not output the spatial spectrum, the spatial spectrum is not used for comparison in this paper. We employ two statistical measures, i.e., the root mean square error (RMSE), and the average estimated source number.

Let $(\tilde{\theta}_k)_i$ stands for the estimate of $\dot{\theta}_k$ at the i-th Monte Carlo trial, the

RMSE can be formulated as

$$\text{RMSE} = \sqrt{\frac{1}{K\upsilon}\sum_{i=1}^{\upsilon}\sum_{k=1}^{K}\left((\tilde{\theta}_k)_i - \dot{\theta}_k\right)^2}, \tag{14}$$

where K is the true source number, and $\upsilon = 100$ is the total number of Monte Carlo trials. For a given simulation point, RMSE is obtained by averaging only the trials in which the estimated source number is greater than or equals to K.

The assignment of estimated source number to the true one is executed based on the Hungarian algorithm [40]. The average estimated source number is obtained by averaging the empirical source numbers of all Monte Carlo runs. These two metrics well reveal the identification capacity of the DOAs and source number.

5.1.4. Parameter Settings

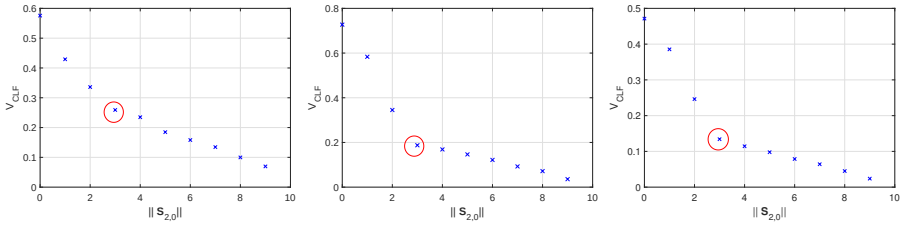
The comparing algorithms need the source number K as the input. Here we set $K = M - 1$ [2]. For l_p-MUSIC, we set $p = 1.1$. The grid interval is set to 0.1° for l_p-MUSIC and MCC-MUSIC, and 2° for the remaining methods if not stated. For MoBEA, the crossover probability and the mutation probability are set to 0.9 and $1/N$, respectively, following the practice of [32]. The population size $\overline{N}$ is 50. The on-grid level optimization terminates when its inner generation reaches 50, or the knee solution remains unchanged in five consecutive generations.

For a fair comparison, all methods stop running when the variance change of the impinging signal $\mathbf{S}^G$ is less than 10^{-6} in five consecutive generations, or the outer-cycle generation reaches 200; and the total number of Monte Carlo trials is 100 for all algorithms. All the experiments are implemented in MATLAB R2018b on a laptop with Intel i5-8265U CPU and 8GB RAM.

5.2. Investigation of Key Operators of MoBEA

This subsection investigates the efficacy of two key operators of MoBEA, i.e., the knee solution identification and the forward search scheme.

[2] A uniform linear array can identify up to $M - 1$ sources, the condition $K \leq M - 1$ can be considered as a priori while the exact value of K is unavailable [35].



(a) GSNR=0dB (b) GSNR=5dB (c) GSNR=10dB

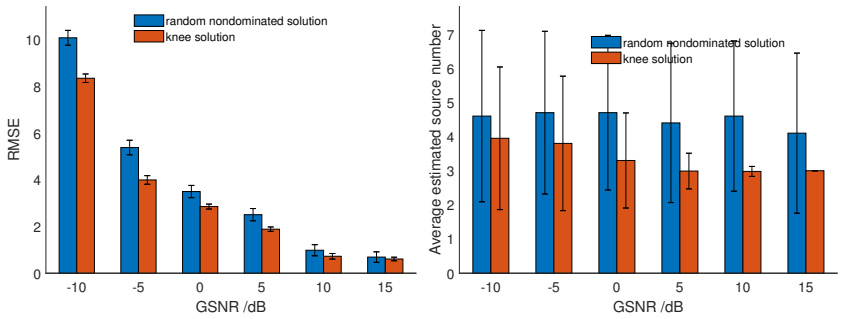
Figure 3: Final PFs obtained at on-grid level in Simulation 1. Solutions in red circles are knee solutions.

5.2.1. Efficacy of Knee Solution

At the on-grid level, the knee solution extracted from the PF is employed to represent the active grid points for grid refinement. It provides a promising trade-off between the source number and the estimation accuracy.

We first use Simulation 1 (detailed in Table 2) to investigate the relationship between the knee solution and the source number. Fig. 3 shows the typical final PF obtained by the on-grid level optimization of 100 trails. Since a uniform linear array can identify up to $M-1$ sources [35], it is expected to identify the knee solution from the non-dominated solutions whose source number is less than M (where $M = 8$). It can be observed that the knee solutions in the red circles of Fig. 3 correspond to the true source number $K = 3$.

To further exploit the impact of the knee solution on the accuracy of DOA estimate, the performance of MoBEA with the knee solution and a randomly selected nondominated solution entering the off-grid level optimization are compared using Simulation 1. Fig. 4 exhibits the RMSE and average estimated source number versus GSNR. It can be seen that the version with the knee solution achieves a smaller DOAs error compared to the random nondominated solution case, thanks to its more precise estimate of the source number. This result indicates that the knee solution is more valuable than other nondominated solutions regarding the source number and DOAs.



(a) RMSE (b) Average estimated source number

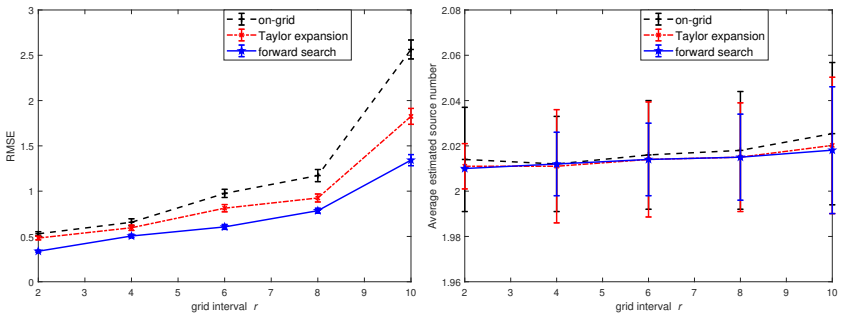
Figure 4: Results of DOA estimate corresponding to random nondominated solution and knee solution under various GSNRs in Simulation 1.

5.2.2. Efficacy of Forward Search

To validate the efficacy of the forward search, we compare the proposed MoBEA to two different versions , i.e., the "on-grid" MoBEA and the MoBEA with Taylor expansion (detailed in Section 5.1.1). Fig. 5 shows the results of DOA estimate versus grid interval in Simulation 2. From Fig. 5(a), the proposed forward search strategy obtains the best localization accuracy under different grid levels. The "on-grid" version and "Taylor expansion" versions suffer from localization performance loss due to ignorance of off-grid mismatch and large modeling error caused by Taylor expansion, respectively. From Fig. 5(b), we note that all three strategies estimate the source number rightly. These observations clearly illustrate the efficacy of forward search.

5.3. Comparison of MoBEA against State-of-the-Art Methods

In this subsection, we compare the proposed MoBEA with state-of-the-art methods, i.e., l_p-MUSIC [9], MCC-MUSIC([11]), Bayes-optimal [21], and Fast-alternating [21], in different kinds of simulations to testify the overall performance of MoBEA.



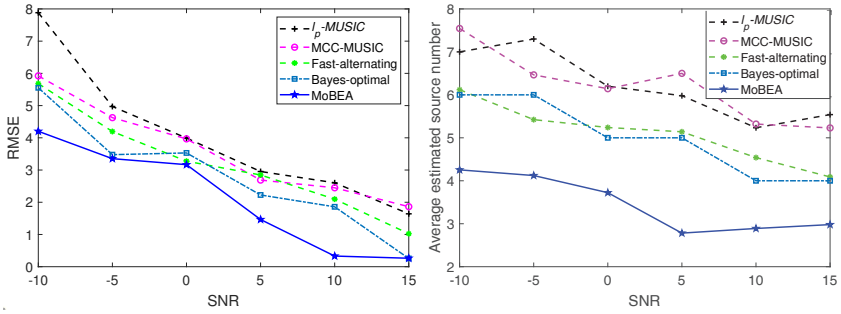
(a) RMSE (b) Average estimated source number

Figure 5: Results of DOA estimate versus grid intervals in Simulation 2.

5.3.1. Comparison Results

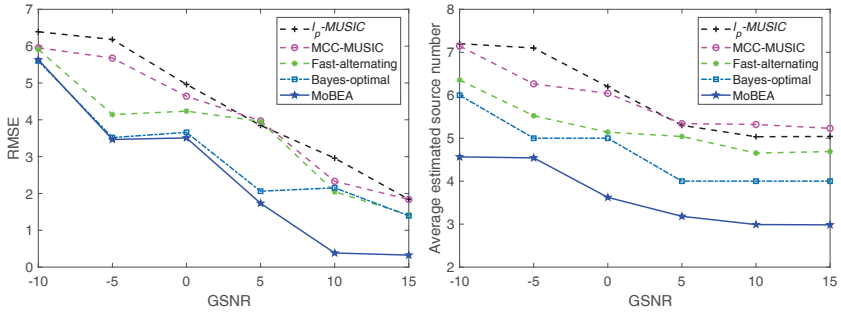
Results under various SNRs. Fig. 6 shows the RMSE and estimated source number versus SNR in Simulation 3 with GMM noise. It can be seen that the performance of all algorithms degrades obviously. l_p-MUSIC and MCC-MUSIC perform the worst under lower SNRs, which may be due to their high sensitivity to lower SNR and imprecise information of source number. The two Bayesian methods behave better in most cases, but they are inferior to MoBEA in localization accuracy. The reason is MoBEA estimating the source number more precisely, which avoids energy leakage on spurious DOAs. Fig. 7 further reports the performance in Simulation 3 with SαS noise. The two MUSIC-based methods still behave badly in impulsive noise. Compared to the SBL-based methods, MoBEA achieves higher or comparable DOA estimation accuracy but always predicts the source number more accurately.

Results under various grid intervals. Fig. 8 exhibits the RMSE and average estimated source number versus grid interval in Simulation 4 with SαS noise. MoBEA achieves comparable or smaller RMSE results than other algorithms under lower SNRs and performs much better with higher SNRs. The enhancement of MoBEA under lower SNRs is non-significant, probably because the localization accuracy is severely affected by the ignorance of the noise variance in the proposed model. It is worth noting that, MoBEA shows an absolute



(a) RMSE (b) Average estimated source number

Figure 6: Results of DOA estimate versus SNR (dB).



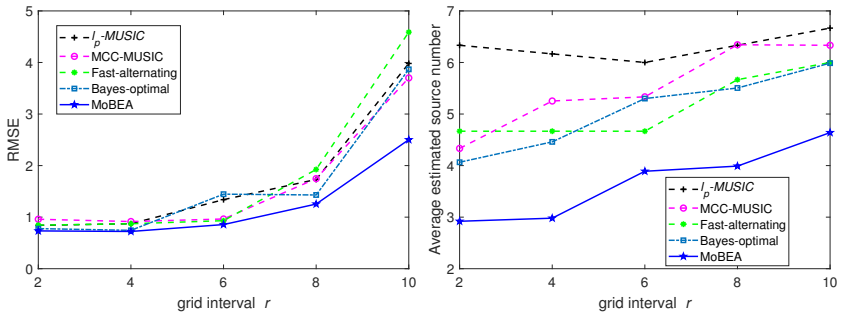
(a) RMSE (b) Average estimated source number

Figure 7: Results of DOA estimate versus GSNR (dB).

predominance in predicting the source number for all grid interval cases.

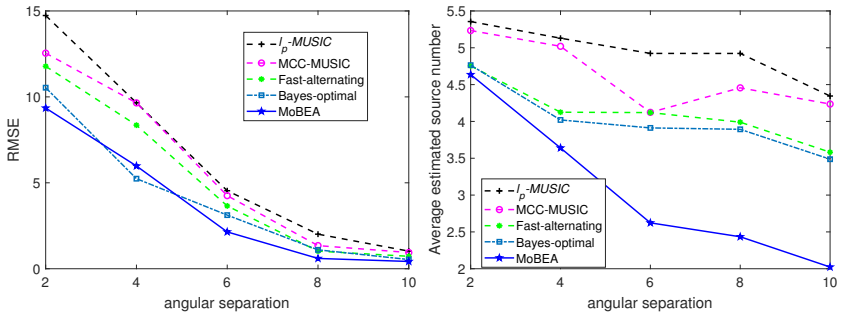
Results under various angular separations. This part examines algorithms' ability to identify two closely located sources. Fig. 9 displays the performance versus angular separations in Simulation 5 with GMM noise. The RMSE values of MoBEA retain the lowest in most cases. As the two sources separate from each other, MoBEA predicts the source number more precisely, which performs remarkably better than other algorithms. The results demonstrate that MoBEA has higher localization resolution when the sources are closely located.

Results under various numbers of snapshots. Fig. 10 shows the RMSE



(a) RMSE (b) Average estimated source number

Figure 8: Results of DOA estimate versus grid intervals.

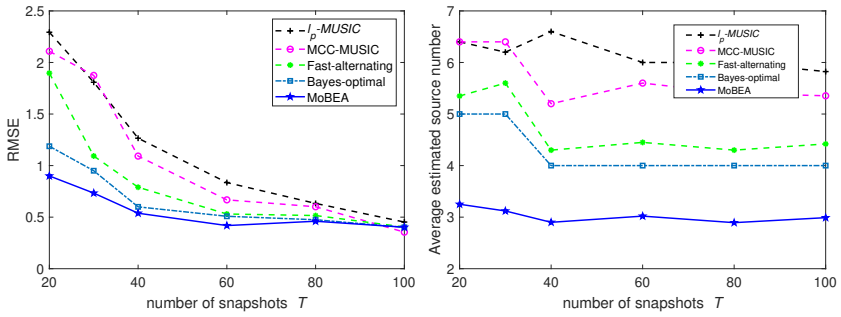


(a) RMSE (b) Average estimated source number

Figure 9: Results of DOA estimate versus angular separations.

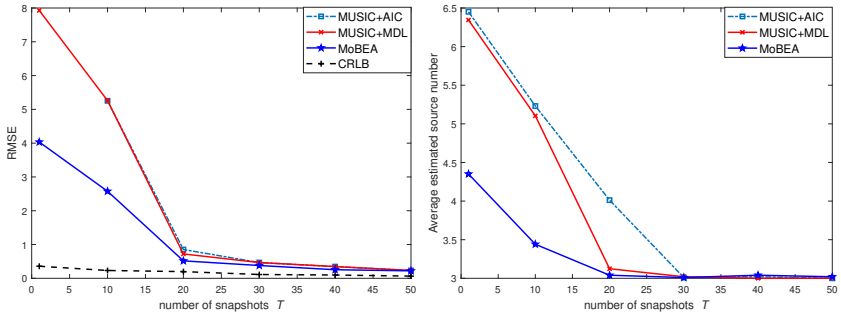
and estimated source number versus the number of snapshots in Simulation 6 with SαS noise. It can be observed that l_p-MUSIC and MCC-MUSIC algorithms gain the worst performance under fewer data samples. Fast-alternating and Bayes-optimal methods obtain significant enhancement in RMSE with fewer snapshots. As expected, MoBEA achieves the highest localization accuracy due to the accurate identification of the source number.

Results under Gaussian noise environment. The proposed MoBEA is compared to MUSIC, a classical DOA estimation method for handling Gaussian noise. MUSIC cannot work without the source number. So we consider using the AIC [25] and MDL [26] principles to estimate the source number for MUSIC.



(a) RMSE (b) Average estimated source number

Figure 10: Results of DOA estimate versus number of snapshots in SαS noise.



(a) RMSE (b) Average estimated source number

Figure 11: Results of DOA estimate versus number of snapshots in Gaussian noise.

We also present the Cramér-Rao Lower Bound (CRLB) to show the theoretical lower bound for comparison. Fig. 11 depicts the RMSE and estimated source number in Simulation 7 with Gaussian noise. Our method clearly shows a significant advantage over MUSICs in terms of RMSE and estimated source number with fewer snapshots (< 30). When the number of snapshots is larger, the performance of our method and MUSICs gets similar. Our method's RMSE value is closer to the CRLB compared to MUSIC.

5.3.2. Discussions on the Results

Significant findings from the comparing results are summarized below.

Firstly, according to the results reported in Figs. 6-11, the proposed MoBEA

achieves the most minor RMSE results in 26 out of the 34 test cases and obtains the source number most close to the ground truth in 30 out of the 34 test cases. These results demonstrate the overall superiority of the proposed MoBEA in DOA estimation and source number identification. In particular, MoBEA shows a strong ability to estimate the source number, especially in cases with fewer snapshots (Figs. 10 and 11). This is because MoBEA perfectly exploits the sparsity via the $l_{2,0}$ norm without relaxation. The source number can be automatically and effectively identified by the knee solution.

Secondly, two MUSICs, i.e., l_p-MUSIC and MCC-MUSIC, perform the worst in most of the 34 test cases. The reasons are two folds. On the one hand, the two MUSICs cannot estimate the source number but must preset the source number. If the preset source number deviates from the ground truth, their performance degrades significantly. On the other hand, fewer snapshots would also severely limit the estimation accuracy of MUSICs. In comparison, our method inherits the advantages of sparse methods, which does not require accurately preset source number and large samples. Moreover, our method can automatically estimate the source number, bringing excellent performance improvement.

Thirdly, the Bayesian-based methods (Fast-alternating and Bayes-optimal) basically achieve moderate performance compared to MUSICs and MoBEA. They employ the inherent Gaussian prior to promote sparse solutions. But the Gaussian prior is suboptimal, which often gives rise to overestimated source number, as reported in [27]. In addition, the Bayesian-based methods highly rely on the assumption that outlier noise has a sparse structure, so that they identify and reject the outliers for further DOA estimation. However, the outlier noise is not strictly sparse in practice, leading to some performance loss. By contrast, our method employs the original $l_{2,0}$ norm to exploit the sparsity without relaxation, thereby predicting more accurate source number. Moreover, MoBEA employs correntropy to handle impulsive noise, which does not need to assume that the outlier is sparse, thus achieving better performance.

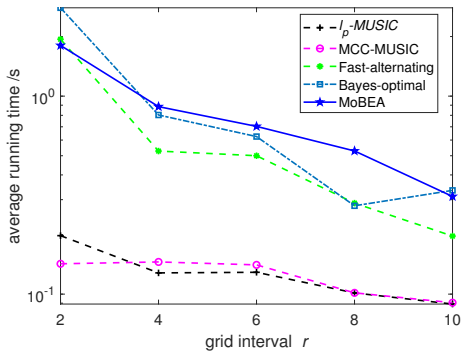


Figure 12: Average running time (seconds) of each algorithm versus grid intervals.

5.4. Efficiency on Running Time

Section 4.5 demonstrates that the number of grid points N has a remarkable impact on the computational complexity of MoBEA. To further quantify the effect, this part provides the average running time of MoBEA versus the grid interval compared to other algorithms, as shown in Fig. 12. The experimental settings are the same as Simulation 4. It can be observed that the computational time of all algorithms shows a decreasing trend with increasing grid intervals (i.e., fewer grid points). The l_p-MUSIC and MCC-MUSIC run much faster than other algorithms. MoBEA and Bayes-based algorithms are slower because they need to execute the time-consuming inversions. If there is a high demand for the faster running speed of MoBEA in practical applications, the decoding is suggested to be performed in parallel. Despite this, MoBEA achieves significant improvements in RMSE and estimated source number over the comparing algorithms in most scenarios.

6. Conclusion

This paper has proposed the MoBEA for DOA estimation. MoBEA involves two innovations. The first is the multiobjective DOA estimation model, in which the source number and a robust correntropy-based fitting error function are taken as two objectives. Unlike existing DOA estimation models, the

proposed model automatically identifies the source number together with DOA estimation. Besides, the original l_0-norm penalty is used to capture the sparsity of signal-of-interest, which avoids relaxing sparse-inducing penalties and enables source number identification more accurately. The second innovation is the multiobjective bilevel evolutionary DOA estimation algorithm for solving the proposed model. The on-grid level is for joint source number identification and sparse recovery, while the off-grid level works on grid refinement via the proposed forward search strategy. This strategy avoids linear approximation and enhances the localization accuracy. Thanks to the population-based search of the proposed algorithm, the solutions own different source numbers communicate to each other during the evolutionary search, which provides diverse search pathways to the optima.

Experimental results have shown that MoBEA is superior to l_p-MUSIC, MCC-MUSIC, Fast-alternating, and Bayes-optimal method in terms of RMSE and estimated source number in impulsive noise. It has also shown that MoBEA works well (close to CRLB) in the scenario of Gaussian noise.

In the future, parallel processing will be considered to accelerate our algorithm. A future study on reducing the computational complexity of MoBEA will also be explored.

References

[1] T. Xia, Joint diagonalization based dod and doa estimation for bistatic mimo radar, Signal Process. 108 (2015) 159–166.

[2] B. Rao, K. Hari, Performance analysis of root-music, IEEE Trans. Acoust. Speech Signal Process. 37 (12) (1989) 1939–1949.

[3] A. Zoubir, V. Koivunen, Y. Chakhchoukh, M. Muma, Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts, IEEE Signal Process. Mag. 29 (4) (2012) 61–80.

[4] R. J. Kozick, B. M. Sadler, Maximum-likelihood array processing in non-gaussian noise with gaussian mixtures, IEEE Trans. Signal Process. 48 (12) (2000) 3520–3535.

[5] J. Zhang, T. Qiu, P. Wang, S. Luan, A novel cauchy score function based doa estimation method under alpha-stable noise environments, Signal Process. 138 (2017) 98–105.

[6] T. Liu, J. M. Mendel, A subspace-based direction finding algorithm using fractional lower order statistics, IEEE Trans. Signal Process. 49 (8) (2001) 1605–1613.

[7] S. Visuri, H. Oja, V. Koivunen, Subspace-based direction-of-arrival estimation using nonparametric statistics, IEEE Trans. Signal Process. 49 (9) (2001) 2060–2073.

[8] A. Swami, B. M. Sadler, On some detection and estimation problems in heavy-tailed noise, Signal Process. 82 (12) (2002) 1829–1846.

[9] W. J. Zeng, H. C. So, H. Lei, l_p-music: Robust direction-of-arrival estimator for impulsive noise environments, IEEE Trans. Signal Process. 61 (17) (2013) 4296–4308.

[10] J. Zhang, T. Qiu, A. Song, H. Tang, A novel correntropy based doa estimation algorithm in impulsive noise environments, Signal Process. 104 (2014) 346–357.

[11] P. Wang, T. Qiu, F. Ren, A. Song, A robust doa estimator based on the correntropy in alpha-stable noise environments, Digit. Signal Process. 60 (2017) 242–251.

[12] J. W. Jhang, Y. H. Huang, A high-snr projection-based atom selection omp processor for compressive sensing, IEEE Trans. VLSI Syst. 24 (12) (2016) 3477–3488.

[13] Q. Liu, H. C. So, Y. Gu, Off-grid doa estimation with nonconvex regularization via joint sparse representation, Signal Process. 140 (2017) 171–176.

[14] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse bayesian inference, IEEE Trans. Signal Process. 61 (1) (2013) 38–43.

[15] Carlin, M, Rocca, P, Oliveri, G, Viani, F, Directions-of-arrival estimation through bayesian compressive sensing strategies, IEEE Trans. Antennas Propagat. 61 (7) (2013) 3828–3838.

[16] F. Chen, J. Dai, N. Hu, Z. Ye, Sparse bayesian learning for off-grid doa estimation with nested arrays, Digit. Signal Process. 82 (2018) 187–193.

[17] N. Hu, B. Sun, J. Wang, J. Dai, C. Chang, Source localization for sparse array using nonnegative sparse bayesian learning, Signal Process. 127 (2016) 37–43.

[18] X. Wu, W. Zhu, J. Yan, Z. Zhang, Two sparse-based methods for off-grid direction-of-arrival estimation, Signal Process. 142 (2018) 87–95.

[19] B. Qi, W. Wang, B. Wang, Off-grid compressive channel estimation for mm-wave massive mimo with hybrid precoding, IEEE Commun. Lett. 23 (1) (2018) 108–111.

[20] Q. Wang, Z. Zhao, Z. Chen, Z. Nie, Grid evolution method for doa estimation, IEEE Trans. Signal Process. (2018) 2374–2383.

[21] J. Dai, H. C. So, Sparse bayesian learning approach for outlier-resistant direction-of-arrival estimation, IEEE Trans. Signal Process. 66 (3) (2017) 744–756.

[22] Y. Zhang, Y. Yang, L. Yang, X. Guo, Root sparse asymptotic minimum variance for off-grid direction-of-arrival estimation, Signal Process. 163 (2019) 225–231.

[23] Q. Wan, H. Duan, J. Fang, H. Li, Z. Xing, Robust bayesian compressed sensing with outliers, Signal Process. 140 (2017) 104–109. doi:https://doi.org/10.1016/j.sigpro.2017.05.017.

[24] J. Dai, L. Zhou, C. Chang, W. Xu, Robust bayesian learning approach for massive mimo channel estimation, Signal Process. 168 (2020) 107345.

[25] S. Valaee, P. Kabal, An information theoretic approach to source enumeration in array signal processing, IEEE Trans. Signal Process. 52 (5) (2004) 1171–1178. doi:10.1109/TSP.2004.826168.

[26] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP) 33 (2) (1985) 387–392. doi:10.1109/TASSP.1985.1164557.

[27] J. Yang, Y. Yang, Sparse bayesian doa estimation using hierarchical synthesis lasso priors for off-grid signals, IEEE Trans. Signal Process. 68 (2020) 872–884.

[28] J. C. Principe, Information theoretic learning. Renyi's entropy and kernel perspectives, Springer Publishing Company, 2010.

[29] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94. doi:10.1109/79.526899.

[30] G. Davis, Adaptive nonlinear approximations, Ph.D. dissertation, Dept. Math., Courant Inst. of Math. Sci., New York Univ., New York, NY, USA, 1994.

[31] Y. Wang, Y. T. Yuan, L. Li, Correntropy matching pursuit with application to robust digit and face recognition, IEEE Trans. Cybern. 47 (6) (2016) 1–13.

[32] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: Nsga-ii, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.

[33] R. T. Rockafellar, Convex analysis, Vol. 28, Princeton university press, 1970.

[34] Y. He, F. Wang, S. Wang, J. Cao, B. Chen, Maximum correntropy adaptation approach for robust compressive sensing reconstruction, Inform. Sci. 480 (2019) 381–402.

[35] B. Ottersten, P. Stoica, R. Roy, Covariance matching estimation techniques for array signal processing applications, Digit. Signal Process. 8 (3) (1998) 185–210.

[36] L. Rachmawati, D. Srinivasan, Multiobjective evolutionary algorithm with controllable focus on the knees of the pareto front, IEEE Trans. Evol. Comput. 13 (4) (2009) 810–824.

[37] I. Mierswa, M. Wurst, Information preserving multi-objective feature selection for unsupervised learning, in: Proc. conf. Gene. Evol. Comput., ACM, 2006, pp. 1545–1552.

[38] R. Poli, W. B. Langdon, Schema theory for genetic programming with one-point crossover and point mutation, Evol. Comput. 6 (3) (1998) 231–252.

[39] C. L. Nikias, M. Shao, Signal processing with alpha-stable distributions and applications, Wiley-Interscience, 1995.

[40] J. Munkres, Algorithms for the assignment and transportation problems, J. Soc. Ind. Appl. Math. 5 (1) (1957) 32–38.