Gridless Evolutionary Approach for Line Spectral Estimation With Unknown Model Order

Bai Yan, Qi Zhao, Jin Zhang, J. Andrew Zhang, Senior Member, IEEE, and Xin Yao, Fellow, IEEE

Abstract—Gridless methods show great superiority in line spectral estimation. These methods need to solve an atomic $\ell_0$ norm (i.e., the continuous analog of $\ell_0$ norm) minimization problem to estimate frequencies and model order. Since this problem is NP-hard to compute, relaxations of the atomic $\ell_0$ norm, such as the nuclear norm and reweighted atomic norm, have been employed for promoting sparsity. However, the relaxations give rise to a resolution limit, subsequently leading to biased model order and convergence error. To overcome the above shortcomings of relaxation, we propose a novel idea of simultaneously estimating the convergence error. To overcome the above shortcomings of relaxation, we propose a novel idea of simultaneously estimating the convergence error. To accomplish this idea, we build a multiobjective optimization model. The measurement error and the atomic $\ell_0$ norm are taken as the two optimization objectives. The proposed model directly exploits the model order via the atomic $\ell_0$ norm, thus breaking the resolution limit. We further design a variable-length evolutionary algorithm to solve the proposed model, which includes two innovations. One is a variable-length coding and search strategy. It flexibly codes and interactively searches diverse solutions with different model orders. These solutions act as steppingstones that helpfully exploring the variable and open-ended frequency search space and provide extensive potentials toward the optimas. Another innovation is a model-order pruning mechanism, which heuristically prunes less contributive frequencies within the solutions, thus significantly enhancing convergence and diversity. Simulation results confirm the superiority of our approach in both frequency estimation and model-order selection.

Index Terms—Gridless, line spectral estimation (LSE), multiobjective evolutionary algorithm, sparse representation.

I. INTRODUCTION

LINE spectral estimation (LSE) aims at frequency estimation and model-order selection from measurements collected as a superposition of complex sinusoids [1]. Here, the “model-order selection” means determining the number of frequencies. LSE has received significant attention as a major subject in signal processing fields. It has various applications, for example, direction-of-arrival estimation in radar and sonar [2], channel estimation in wireless communications, and simulation of atomic systems in molecular dynamics.

Many methods have been proposed for frequency estimation. Classical methods such as subspace methods [3]–[6] are based on sample statistics. Their performance highly relies on a large number of snapshots and uncorrelated components. Moreover, the model order is required as a priori. With the development of a compressive sensing theory [7]–[9], sparse methods have been presented for frequency estimation. These methods exhibit great advantages over subspace methods, such as robustness to noise, no requirement for model order, and low requirement for snapshots. Sparse methods can be divided into grid-based and gridless types. In the grid-based type [10]–[13], the continuous frequency domain must be discretized into a finite grid, and the frequencies are restricted to this grid. Then, the LSE problem is simplified to a sparse recovery problem. However, gridding gives rise to the well-known basis mismatch issue that limits the estimation accuracy. To avoid gridding, the so-called gridless type [14]–[16] has been proposed, which can directly operate in the continuous frequency domain.

Apart from frequency estimation, the model order should also be determined in LSE. For subspace-based methods, some popular choices based on information criterion [17]–[19] can be incorporated to estimate the model order. It may be challenging to derive the accurate model order in nonasymptotic regimes, such as limited snapshots or low signal-to-noise ratio (SNR) [15]. Differently, sparse methods adopt sparse penalties to exploit frequencies’ sparsity. Ideally, the $l_0$ norm or atomic $l_0$ norm is the best choice for sparse penalty, but they incur an NP-hard problem. To make this problem easily solvable, the $l_0$ norm or atomic $l_0$ norm is usually relaxed to other sparse metric, for example, $l_p$-norm ($p \in (0, 1)$) sparse penalty, atomic norm [20]; reweighted atomic-norm [21], [22]; Gaussian prior [23]; or Gaussian–Bernoulli prior [15], [24], [25]. Such
relaxation makes sparse methods suffer from a resolution limit [21], [26], that is, the true frequencies are required to be well separated for successful recovery. Consequently, the resolution limit may lead to biased model order and frequency error.

In order to not only jointly estimate the model order and frequencies but also break the resolution limit, in this article, we propose a multiobjective LSE model. The measurement error and atomic $l_0$ norm are taken as the two conflicting objectives. The multiobjective setting enables the frequencies and model order to be simultaneously estimated. Moreover, we exactly exploit the model order by the atomic $l_0$ norm without relaxation, thus breaking the resolution limit and providing the accurate model order.

The proposed multiobjective LSE model is an NP-hard problem. To solve this model, there are two issues to be addressed. One is how to handle the atomic $l_0$ norm without relaxation. Another is how to find the optima from the continuous (gridless) search space without knowing the true model order as a priori.

Aiming at the two issues, we design a multiobjective variable-length evolutionary search algorithm (MVESA) to solve the proposed model. Here, we design MVESA from the perspective of evolutionary algorithms because evolutionary algorithms have revealed their strong ability to handle NP-hard $l_0$ problems [27]–[29]. To deal with the unknown model-order problem, we introduce a variable-length coding and search strategy. This strategy flexibly codes diverse solutions with different lengths (i.e., different model orders). Then, it interactively searches diverse pathways (formed by solutions with different sizes) over the variable and open-ended frequency search space. These pathways act as steppingstones that helpfully exploring the search space and provide extensive potentials toward the optima. Furthermore, we develop a model-order pruning mechanism. This mechanism heuristically prunes less contributive frequencies within solutions. The pruning length is set at random. Hence, solutions’ convergence and diversity are both significantly improved.

4) We validate MVESA’s performance via simulations. Results confirm MVESA’s efficacy and better performance in terms of frequency estimation and model-order selection concerning state-of-the-art methods.

The remainder of this article is organized as follows. Section II provides background knowledge and related works. Sections III and IV present the proposed multiobjective atomic $l_0$ model and variable-length evolutionary search algorithm. Section V gives simulation results. Finally, Section VI concludes this article.

Notation: Boldface letters represent vectors and matrices. $\mathbb{R}$ and $\mathbb{C}$ denote the real domain and complex one, respectively. $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote the transpose, conjugate, and conjugate transpose of a vector or matrix, respectively.

II. BACKGROUND

In this section, we introduce the LSE problem first. Since our work falls into the gridless type and is closely related to evolutionary multiobjective optimization, we then review the existing gridless methods and background knowledge of evolutionary multiobjective optimization.

A. Line Spectral Estimation

In the LSE model [1], the measurements $\mathbf{Y} \in \mathbb{C}^{M \times L}$ are a sparse combination of $K^*$ complex sinusoids corrupted by the white Gaussian noise $\mathbf{N}$

$$\mathbf{Y} = \sum_{k=1}^{K^*} \mathbf{a}(\theta_k) \mathbf{s}_k^T + \mathbf{N} = \mathbf{A} \mathbf{S} + \mathbf{N} \quad (1)$$

where $\mathbf{a}(\theta_k) \triangleq [1, e^{j\pi \theta_k}, \ldots, e^{j(M-1)\pi \theta_k}]^T$ is the $k$th complex sinusoid, $\theta_k \in [-1, 1]$ and $\mathbf{s}_k \in \mathbb{C}^{L \times 1}$ denote the frequency and complex-valued amplitudes of the $k$th sinusoidal component. $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_{K^*}]$ denotes the frequency vector. $\mathbf{S}^T_k$ is the $k$th row of amplitude matrix $\mathbf{S}$. Note that the length of frequency vector $K^*$, also referred to as the “model order,” is unknown in this article. The goal of LSE is to estimate the frequencies $\theta$ and amplitude matrix $\mathbf{S}$, given the measurements $\mathbf{Y}$ and mapping $\mathbf{A}(\theta)$ (i.e., $\theta \rightarrow \mathbf{A}$). Therefore, LSE is an inverse problem.

B. Related Works

Gridless LSE methods do not need grid discretization and work directly in the continuous frequency domain. These methods need to solve an atomic $l_0$ norm (the continuous analog of $l_0$ norm) minimization problem. The atomic $l_0$ norm directly exploits sparsity and has no resolution limit, but it is NP-hard to compute. To make it tractable, earlier works switched to the convex atomic $l_1$ norm (also known as nuclear norm or atomic norm), including [31] and [32] for noiseless data and [33] and [34] for noisy data. Later, several works minimized a covariance matrix fitting criterion [14]. They had been proved to be equivalent to atomic norm-based methods.
However, due to the convex relaxation, the above methods suffer from a serious resolution limit, that is, the frequencies are required to be well separated for recovery.

To alleviate the resolution limit, the reweighted atomic-norm minimization [21], [22] and the reweighted covariance fitting criterion [35] were reported to approximate the atomic $l_0$ norm. They brought enhanced sparsity and resolution compared to convex atomic norm-based methods. Alternatively, alternating projections-based gridless methods [16], [36] directly solved the atomic $l_0$ norm minimization problem to pursue higher resolution. However, the convergence performance is not guaranteed due to unclosed or nonconvex sets [16]. Besides, by treating the frequencies as random variables, a few gridless LSE methods in the Bayesian framework [15], [24], [25] were also proposed to estimate frequencies.

Apart from frequency estimation, model order selection is also needed. Instead of using the atomic $l_0$ norm, atomic norm-based methods [21], [22], [31], [34] exploit the model order by relaxed sparse metrics. This relaxation manner suffers from a resolution limit, subsequently producing biased model order and large frequency error. For covariance fitting criterion methods [14], [35], the model order is usually identified by classic user-set threshold or information criterion methods [17]–[19] a posteriori. However, it is very challenging to derive accurate results due to inferior statistical properties in nonasymptotic regimes (e.g., limited snapshots or SNR) [15]. For alternating projection-based methods, the model order is required as a priori [16], [36]. In variable Bayesian methods [15], [24], [25], the Gaussian–Bernoulli prior is employed to promote sparsity. However, it is still not yet clear how to determine the optimal sparse distributions in Bayesian framework [26].

Except for the aforementioned optimized-based gridless methods, deep learning techniques have been also applied to estimate model order and large frequency error. For covariance fitting criterion methods [17]–[19] a posteriori. However, it is very challenging to derive accurate results due to inferior statistical properties in nonasymptotic regimes (e.g., limited snapshots or SNR) [15]. For alternating projection-based methods, the model order is required as a priori [16], [36].

III. PROPOSED MULTIOBJECTIVE LSE MODEL

To simultaneously estimate frequencies and model order, we naturally formulate the LSE model (1) as a MOP. The measurement error and the atomic $l_0$ norm are taken as two conflicting objectives. Our formulation holds two advantages: 1) frequencies and model order can be simultaneously estimated without adjusting any balancing parameter and 2) the model order is exactly exploited by the atomic $l_0$ norm without relaxations, hence breaking the resolution limit.

For clarity, we first profile the atomic $l_0$ norm before giving our proposed model. We follow the research [41] to define the atomic $l_0$ norm of measurements $\mathbf{Y}$. Specifically, defining an atomic set

$$A := \left\{ \mathbf{a}(\theta, \phi) := a(\theta)\phi : \theta \in [-1, 1], \phi \in \mathbb{C}^{1 \times L}, \|\phi\|_2 = 1 \right\}$$

it can be viewed as an infinite dictionary indexed by the continuous varying parameters $\theta$ and $\phi$. The atomic $l_0$ norm of measurements $\mathbf{Y}$, $\|\mathbf{Y}\|_{A,0}$, is defined as the minimum number of atoms in $A$ that synthesizes $\mathbf{Y}$

$$\|\mathbf{Y}\|_{A,0} = \inf_{\theta_k, \phi_k, c_k} \left\{ \kappa : \mathbf{Y} = \sum_{k=1}^{\kappa} \mathbf{a}(\theta_k, \phi_k)c_k, \theta \in [-1, 1] \right\}$$

$$\|\phi\|_2 = 1, c_k > 0 \right\}$$

$$= \inf_{\theta_k, s_k} \left\{ \kappa : \mathbf{Y} = \sum_{k=1}^{\kappa} \mathbf{a}(\theta_k)s_k, \theta \in [-1, 1] \right\}$$

where “inf” stands for infimum, $\phi_k = c_k^{-1}s_k$, and $c_k = \|s_k\|_2$.

By introducing the atomic $l_0$ norm (4), we formulate LSE as a multiobjective optimization model

$$\min_{f(\theta), S} \left( \|\mathbf{Y}\|_{A,0}, \|\mathbf{Y} - A\mathbf{S}\|^2_F \right)$$

where the decision variables include the frequency vector $\theta$ and amplitude matrix $\mathbf{S}$. The two objectives, $\|\mathbf{Y}\|_{A,0}$ and $\|\mathbf{Y} - A\mathbf{S}\|^2_F$, represent the atomic $l_0$ norm (equals the model order) and measurement error, respectively. The measurement error is a fidelity term, which aims at keeping the quality of decision variables. Thus, the measurement error is taken as an objective. However, the decision variables may contain redundant frequency components that barely affect the measurement error. These redundant frequencies are generally spurious and needs to be removed. To remove them,
we naturally consider the model order as the other objective. Here, the exact atomic $l_0$ norm is introduced to appropriately promote sparsity; thus, it does not suffer from a resolution limit compared to existing gridless methods [35], [36]. Hence, estimates of model order and frequencies can be more accurate. With the second objective, the estimated frequencies would be prompted to approach the ground-truth frequencies. Meanwhile, the model-order objective contributes to removing the redundant frequencies, which drives the estimated model order to be optimal. Therefore, the multiobjective model can achieve the goal of estimating frequencies and model order.

IV. PROPOSED VARIABLE-LENGTH EVOLUTIONARY SEARCH ALGORITHM

We design a variable-length evolutionary search algorithm to solve the proposed multiobjective LSE model (5). The designed algorithm includes two major innovations. One is a variable-length coding and search strategy. It flexibly codes and interactively searches diverse solutions with different model orders. These solutions act as steppingstones that helpfully exploring the variable and open-ended frequency search space and provide extensive potentials toward the optima. Another innovation is a model-order pruning mechanism. It heuristically prunes less contributive frequencies within the solutions. The pruning length is set at random. With this mechanism, solutions’ convergence and diversity are enhanced.

A. Overall Framework

The workflow of the proposed MVESA is shown in Algorithm 1. MVESA starts with initialization. A number of $N$ frequency vectors with different model orders are generated at random. Their corresponding amplitude matrices are recovered via the least square method (will be detailed in Section IV-B) and the polynomial mutation operator [43] are introduced to produce $N$ new frequency vectors with different model orders. New frequency vectors’ amplitude matrix are recovered by the least square method. The new frequency vectors and their amplitude matrices make up the offspring solution set $Q^G$. $Q^G$’s fitness is calculated by model (5).

Step 2 (Variable-Length Search): With obtained parents, a modified variable-length crossover (will be detailed in Section IV-B) and the polynomial mutation operator [43] are introduced to produce $N$ new frequency vectors with different model orders. New frequency vectors’ amplitude matrix are recovered by the least square method. The new frequency vectors and their amplitude matrices make up the offspring solution set $Q^G$. $Q^G$’s fitness is calculated by model (5).

Step 3 (Environmental Selection): The environmental selection operator of NSGA-II [44] is applied to select $N$ better (in terms of convergence and diversity) solutions from $P^G \cup Q$. These $N$ solutions reform $P^G$.

Step 4 (Archiving and Model-Order Pruning): We set an external archive $R^G$ to collect the best solutions with each possible model order found so far (i.e., from $\bigcup_{G=1}^{G} P^G$). This archive avoids missing optimal solutions during iterations. $R^G$ is first updated with elite solutions $P^G$. The solutions that newly join $R^G$ at the current generation are denoted as newcomers. We design a novel model-order pruning mechanism to improve the archive’s convergence and diversity significantly. Specifically, for each newcomer, this mechanism heuristically prunes its less contributive frequencies. Then, the pruned newcomer’s amplitude matrices are recovered by the least square method. With pruned newcomers’ frequency vectors and amplitude matrices, the fitness of pruned newcomers is obtained by model (5). Finally, archive $R^G$ and population $P^G$ are updated with the pruned newcomers.

Once the iterative generation terminates, we identify the knee solution from archive $R$ as the final solution. This is because the knee solution has the maximum marginal rates of return, that is, an improvement in one objective would lead to severe degradation in another. It provides an attractive tradeoff between the two objectives [27], [45], and the efficacy has been empirically confirmed in Section V-B1. Here, we employ the kink method [46] to identify the knee solution, by which the solution with the largest slope variance over the obtained PF is taken as the knee. Finally, MVESA returns the knee solution’s frequency vector as the output.

The core components of MVESA, that is, the variable-length coding and search, archiving and model-order pruning, are detailed as follows.

B. Variable-Length Coding and Search

The true model order is unknown in advance. Obtrusively using a predefined fixed-length (fixed model order) coding would lead to suboptimal, deteriorating LSE performance. To handle this, we naturally introduce a variable-length coding strategy to represent solutions with diverse possible model orders, offering tremendous flexibility. To the best of our knowledge, it is the first time that realizes direct LSE over the dynamic size of frequency search space.

In the variable-length coding strategy, we code each solution by a variable-length representation

\[
P = \{ (\theta_1, S_1), \ldots, (\theta_n, S_n), \ldots, (\theta_N, S_N) \}
\]

\[
\theta_n = [\theta_{n1}, \theta_{n2}, \ldots, \theta_{nd}] \in \mathbb{R}^{1 \times d_n}
\]

\[
S_n \in \mathbb{C}^{d_n \times L}
\]

(6)

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where the population $P$ consists of $N$ solutions, $(\theta_s, S_n)$ is the $n$th solution, and $d_n$ is the length (model order) of the $n$th solution. The variable length is reflected by solutions with different $d_n$s. The model order could be evolved toward the ground truth during evolutionary search. Fig. 1 gives an example of frequency vectors. Each row indicates a frequency vector, where frequencies are sorted in ascending order, and the number of lattices is the model order.

Once $\theta$ is obtained by the initialization or evolutionary search, we can employ a simple-yet-effective least square method to solve $\min_{S} \|Y - AS\|_2$ and acquire the corresponding amplitude matrix $S$

$$S = (A^T A)^{-1} A^T Y.$$  

(7)

Therefore, the task of LSE becomes to find the frequency vector as accurately as possible.

To find the optimal frequency vector, we should discretely design evolutionary search operators. Since mutation has no concern with solution’s lengths, we employ the classical polynomial mutation [43] to perturb solutions. However, traditional crossover operators cannot be incorporated because they are only designed for fixed-length coding. Recently, quite a few variable-length crossover operators have been proposed [47], for example, cut and splice, spatial, and similarity-based operators. Cut and splice operators are the most disruptive. Spatial operators are the least disruptive but can only be applied to specific problems with spatial components. Similarity-based operators are less disruptive by preserving common sequences and allowing only differences to be exchanged or removed. Thus, we modify a similarity-based operator, that is, the synapsing variable-length crossover [30] and incorporate it into our work.

We modify the synapsing variable-length crossover by randomly selecting the number of crossover points. This modification helps maintain good diversity of solutions. The process is exhibited in Fig. 2, which involves two steps as follows.

1) **Aligning Parents:** As displayed in Fig. 2(a), oblique black lines link the lattice of one parent to a counterpart of another parent, such that the linked two lattices are the most similar to each other. The similarity is measured by the Euclidean distance. Based on the links, we can align the two parents, as shown in Fig. 2(b).

2) **Executing $\bar{n}$-Point Crossover:** We randomly choose the number of crossover points $\bar{n}$ for maintaining diversity. Here, $\bar{n}$ is a random integer value between 1 and the length of the shorter parent. With $\bar{n}$ crossover points [red-dotted lines in Fig. 2(b)], the two parents are segmented into $\bar{n} + 1$ paired subsets. According to the principle of $\bar{n}$-point crossover, each even paired subsets exchange affiliations, respectively. For example, in Fig. 2(c), the second paired subsets $\{-0.31\}$ and $\{-0.2\}$ exchange their affiliations; so does the fourth paired subsets $\{0.6\}$ and $\{\}$. Finally, subsets are glued to produce two offsprings with different lengths, as depicted in Fig. 2(d).

C. Archiving and Model-Order Pruning Mechanism

We propose a novel archiving and model-order pruning mechanism, which tremendously enhances solutions’ convergence and diversity. Archiving refers to using the external archive $R$ to collect the best solutions with each possible model order found during iteration. It avoids missing optimal solutions. Furthermore, the convergence and diversity of archive solutions can be well maintained. Model-order pruning aims to prune less contributive frequencies within solutions, thus reducing solutions’ redundancy and enhancing the convergence performance.

The motivation of designing the model-order pruning step is as follows. The variable-length search is very likely to produce long solutions. These solutions may include both close-to-optimal frequencies and spurious ones. It is necessary to remove the spurious frequencies and push overlong solutions toward the optima. To do this, we tend to remove the frequencies with lower power. The intuition is, when the measurement matrix is well conditioned, that is, when the ground-truth frequencies are well separated with respect to the noise level, frequency recovery is stable [48]. In this case,
the measurement error is very small. Thus, pruning the frequencies with low power would be reasonable. Motivated by this, we design a model-order pruning mechanism to heuristically remove the frequencies with lower power, so that the frequency vector’s redundancy is greatly reduced and the resulting solution moves toward the optima.

The pseudocode of archiving and model-order pruning is shown in Algorithm 2. It includes three operations: archiving, model-order pruning, and update.

**Archiving (Lines 2–5 of Algorithm 2):** We aim to collect the best solutions with each possible model order so far and store them into archive \( R^G \). For clarity, we denote the Pareto nondominate solutions of population \( P^G \) as elite solutions. As depicted in Fig. 3(a), we put the archive solutions and elite solutions together. For each pair of archive and elite solutions with the same length, we replace the archive solution with the elite one only if this archive solution is dominated by the elite one. Thereafter, the solutions that newly join the archive are denoted as newcomers.

For example, in Fig. 3(a), the archive solution \( d \) and the elite solution \( c \) has the same model order. \( d \) is dominated by \( c \), so we replace \( d \) in archive by \( c \), enhancing the convergence performance. In this way, we determine all newcomers \([a, b, c]\). With the archiving mechanism, solutions with the best performance for each possible length can always be reserved; thus, the convergence and diversity of solutions are both improved.

**Model-Order Pruning (Lines 7–12 of Algorithm 2):** As exhibited in Fig. 3(b), we aim at pruning the frequency vectors of newcomers \([a, b, c]\) to create shorter solutions \([a’, b’, c’]\). Specifically, assuming \( x \) is a newcomer \((x \in [a, b, c])\), the decision variables of \( x \) are \((\theta, S)\), and the length of \( \theta \) is \( \bar{K} \), we prune the frequency vector \( \theta \) via three steps.

1) Calculate each frequency component’s power by averaging the signal amplitudes over multiple snapshots

\[
p_i = \sqrt{\sum_{l=1}^{L} |S_{il}|^2}, \quad i = 1, 2, \ldots, \bar{K}
\]  

where \( p_i \) is the \( i \)th frequency component’s power, and \( S_{il} \) is the \( i \)th row and the \( l \)th column element of \( S \).

2) Sort the \( \bar{K} \) frequencies components’ power in descending order

\[
p_{i_1} \geq p_{i_2} \geq \cdots \geq p_{i_{\bar{K}}}
\]  

where \( i_1, i_2, \ldots, i_{\bar{K}} \) is a permutation of \([1, 2, \ldots, \bar{K}]\).

3) For \( \theta \), remove a number of \( \bar{K}_{\text{cut}} \) frequencies with low power in priority. To maintain the length diversity in solutions, \( \bar{K}_{\text{cut}} \) is set to be a random value within \([1, \bar{K} - 1]\). After removing lower frequencies, the pruned frequency vector \( \theta_{\text{pruned}} \) is produced

\[
\theta_{\text{pruned}} = \left[ \theta_{i_1}, \theta_{i_2}, \ldots, \theta_{i_{\bar{K} - \bar{K}_{\text{cut}}}} \right].
\]
newcomers \{a, b, c\} give birth to their shorter versions \{a’, b’, c’\}.

Population and Archive Update (Line 14 of Algorithm 2): With pruned solutions \{a’, b’, c’\}, we update the archive \(R^G\) and population \(P^G\). The update rules include the following.

1) If the pruned solution is dominated by any archive solution with the same model order, the pruned solution is inferior and discarded.

2) If the pruned solution dominates any archive solution with the same length, the pruned solution substitutes this archive solution and is added into the population. If so, the convergence performance can be improved.

3) If no archive solution has the same model order as the pruned solution, the pruned solution is added into the archive and population. This action can maintain good solution diversity in model order.

Take Fig. 3(c) as an example of population and archive update. The pruned solution \(a’\) is dominated by the archive solution \(e\) with the same model order; hence, the first rule is satisfied. \(a’\) is an inferior solution and discarded. The pruned solution \(b’\) dominates the archive solution \(f\) with the same model order. Therefore, \(b’\) follows the second rule. It substitutes \(f\) and is added into the archive and population, improving the convergence performance. For the pruned solution \(c’\), its model order is different from all archive solutions’. Therefore, \(c’\) takes the third rule. \(c’\) is added to the archive and population to enhance the solution diversity in model order. Finally, it can be found that the inferior solutions \{a’, f\} are removed, and pruned solutions \{b’, c’\} are added into the archive. Compared to the old archive, the resulted archive (circled in the blue lines) obtains better convergence and diversity performance.

There may exist a problem that when the frequencies are closely spaced, the measurement matrix is ill conditioned. The solution of the intensity recovery is sensitive to noise, closely spaced, the measurement matrix is ill conditioned. The diversity prevents the solutions from being severely affected by spurious frequencies with high power. Finally, the mutation and crossover operators in the evolutionary algorithm enable the solutions to modify or eliminate the spurious frequencies. Therefore, in the cases with closely spaced frequencies, our algorithm can still have promising performance.

D. Convergence and Complexity Analysis

We now analyze the convergence and computational complexity of the proposed MVESA.

1) Convergence: The convergence of the proposed MVESA can be guaranteed.

Analysis: In MVESA, the archive \(R\) can be viewed as reserving the best solutions to a series of subproblems with different model orders. Specifically, for a subproblem associated with a specific model order, \(R\) always reserves a solution with lower measurement error to this subproblem. It implies that for each subproblem, the measurement error is nonincreasing after each iteration of MVESA. Since the measurement error is lower bounded for each subproblem, MVESA is guaranteed to converge.

2) Complexity: The main computational complexity of MVESA lies in the modified crossover and amplitude matrix recovery. At each generation, the modified crossover requires \(O(NM^2)\) computations, where \(N\) and \(M\) are the population size and the number of measurements, respectively. The computational complexity of recovering the amplitude matrix is \(O(2NM^3)\) in the worst case. Thus, the total complexity is \(O(G_{max}NM^3)\) with \(G_{max}\) denoting the maximum generation.

V. SIMULATION EXPERIMENTS

In this section, we conduct simulation experiments to evaluate the performance of the proposed MVESA compared to state-of-the-art algorithms under various scenarios.

A. Setup, Metrics, and Algorithms

Setup: According to the LSE model (1), a number of \(K\) frequencies are randomly generated within \([-1, 1]\). The amplitude matrix \(S\) are drawn from independent and identically distributed (i.i.d.) from \(CN(1, 0.1)\). In numerical simulations, we do not control the minimum frequency separation unless stated otherwise; thus, the frequencies may not be guaranteed to be recovered, even for large \(M\). The noise samples contaminating the measurements are independent and zero-mean complex Gaussian distributed.

Metrics: Since Bayesian-based methods and our algorithm do not output spatial spectrum, spatial spectrum will not be used for comparison. We employ two statistical measures, that is: 1) root mean square error (RMSE) and 2) success rate. RMSE is defined as

\[
RMSE(\hat{\theta}) = \sqrt{\frac{1}{\nu} \sum_{i=1}^{\nu} \|\hat{\theta} - \theta\|_2^2}
\]

(11)

where \(\theta\) is the true frequency vector. The averaging operation performs over the trials in which the estimated model number is greater than or equal to \(K\). Especially, \(\hat{\theta}\) denotes the estimated frequencies by keeping the associated largest \(K\) entries of the power estimate, which is a common operation in many works [14], [21]. The case that the estimated model order is less than \(K\) is ignored. Afterward, the assignment of the estimated frequencies to the true one is executed based on the Hungarian algorithm [51]. Finally, the RMSE can be obtained by averaging the Euclidean distance between \(\hat{\theta}\) and \(\theta\) over \(\nu\) Monte Carlo runs. The success rate is the empirical probability that the estimated model order \(\hat{K}\) is the same as the true value \(K\), denoted as \(Pr(K = \hat{K})\).

Algorithms: We conduct simulations to compare the performance of MVESA with the state-of-the-art gridless
algorithms, that is, SPA [14], RAM [21], APG [16], and VALSE_MMV [24]. The comparison algorithms are introduced as follows.

1) **SPA**: A gridless algorithm based on covariance fitting criteria and convex optimization. This algorithm can work without model order but cannot determine it accurately.

2) **RAM**: A gridless algorithm based on reweighted atomic-norm minimization for enhancing sparsity and resolution. The model order is exploited by the reweighted atomic norm.

3) **APG**: A gridless algorithm that directly solves the atomic $l_0$ norm minimization problem via alternating projections. But the model order needs to be known as a priori.

4) **VALSE_MMV**: A representative gridless sparse Bayesian inference-based algorithm that estimates the posterior probability density functions of frequencies. The model order is estimated using the Bernoulli–Gaussian distribution.

Comparison algorithms’ parameters are set in accordance with their original papers [14], [16], [21], and [24], respectively. For SPA, RAM, and APG, the model order is set to its possible maximum value, $M-1$. For the proposed MVESA, we set the population size $= 30$, mutation distribution index $= 20$, and mutation probability $= 1/k$, where $k$ is the model order of the current solution. To accelerate the search efficiency of MVESA, we generate the initial population in this way: a single solution with a maximum length $M-1$ is produced by the simple Capon method [52], and the remaining $N-1$ solutions are randomly generated.

For a fair comparison, all the algorithms stop running when the change of estimated measurements $\| \hat{Y}^G - \hat{Y}^{G-1} \|_F / \| \hat{Y}^{G-1} \|_F$ is less than $10^{-6}$ in three consecutive generations, or the total number of iterations for comparison methods and our method exceeds 5000 and 100, respectively. The reason for setting the number of iterations like this is that at each generation, no more than 50 solutions are explored in our method, and one solution is explored in comparison methods. As a result, all algorithms are allowed to explore at most 5000 solutions in total, providing a fair comparison. The total number of Monte Carlo runs is set to 200 for all algorithms.

B. Detailed Analysis of MVESA

In this section, the effectiveness of the two-objective function and the proposed archiving and model-order pruning mechanism is investigated to demonstrate the superiority of MVESA.

1) Study of Objective function: The two proposed objectives, including: 1) the model order and 2) measurement error, are conflicting with each other. This conflicting characteristic enables MVESA to determine the model order automatically. To validate it, we conduct a simulation with $K = 4$, $M = 15$, $T = 20$ and increase SNR from -5 to 15 dB. Apart from this, the noiseless case is involved as reference. Fig. 4(a) depicts the typical PF and Fig. 4(b) depicts the slope variance results of the final archive over 200 runs. Fig. 4(b) is obtained by computing the slope variance of the PF according to the kink method [46]. The knee solutions characterized by the maximum slope variance for different SNRs are identified and enframed within dotted line in Fig. 4(a). It can be seen that the identified knee solutions provide the best tradeoff between the two conflicting objectives and acquire the true model order. Thus, it indicates the efficiency of the two-objective function of MVESA.

2) Study of the Archiving and Model-Order Pruning Mechanism: To validate this mechanism’s efficacy, we compare the performance of three versions of MVESA, including: 1) the first version without archiving or model-order pruning; 2) the second version with archiving; and 3) the third version with archiving and model-order pruning. The simulation parameters are set to $M = 15$, $T = 10$ snapshots, SNR = 10 dB, and the number frequencies $K$ increases from 1 to 7. Fig. 5 plots the resultant RMSEs and success rates under different $K$s. The third version, that is, MVESA, is always superior to the other versions in terms of frequency error.
and success rate. This advantage can be explained as follows. Compared to the first version, the last two ones incorporate archive to collect the best solutions so far. Consequently, they can avoid missing the optimal solution, providing better estimation performance. The performance gap between the second version and MVESA is because the model-order pruning operation can fully explore the search space over different dimensionalities. This operation could skip suboptimal solutions, bringing enhanced convergence performance.

C. Comparison of BEA Against Other Methods

In this section, the algorithms’ ability to handle complete data and incomplete data is investigated, respectively. The complete data are referred to as all \( M \) measurements being available while incomplete data mean that some of the \( M \) measurements are missing. The missing data case may occur due to sensor failure, outliers, or other physical constraints. The time cost of all these algorithms is also compared.

1) Handling Complete Data: In Simulation 1, Monte Carlo trials are carried out to investigate the robustness to noise intensity. We set the parameters \( K = 4, M = 15, T = 30 \) snapshots, and vary SNRs from \(-6\) to \(15\) dB. Fig. 6 shows the RMSE and success rate and frequency errors versus SNRs. MVESA retains much lower frequency errors compared to other algorithms for all SNRs. In terms of success rate, SPA, RAM, and APG roughly fail to determine the model order. In contrast, VALSE_MMV and MVESA estimate the model order more accurately. Note that MVESA obtains the highest success rate at very low SNRs. The superior performance of MVESA is because it owns outstanding ability to explore the dimensionality-changing search space, which can skip suboptimal solutions and bring enhanced performance.

In Simulation 2, the capability of model-order selection in scenarios with varying model orders is investigated. The parameters are set to \( M = 15, T = 10 \) snapshots, SNR = 10 dB, and model order \( K \) increasing from 1 to 7. Results of RMSE and success rate under varying model orders are shown in Fig. 7. SPA, RAM, and APG still perform worse than VALSE_MMV and MVESA in terms of RMSE and success rate, because they lack the capability of model-order selection. MVESA shows an absolute advantage over VALSE_MMV in terms of the two evaluation metrics for most of \( K \)s. This advantage benefits from the atomic \( l_0 \) norm, which helps predict the model order more accurately and largely improve the estimation accuracy.
Simulation 3 studies the statistical performance of LSE of resolving two closely spaced frequencies. Assume the distance between two components vary from 0.02 to 0.26, \( M = 6 \), \( T = 10 \) snapshots, and \( \text{SNR} = 10 \) dB. Fig. 8 shows the results versus frequency separation. It can be observed that MVESA wins the best in 10 out of the 12 cases in terms of frequency error. When the two frequencies are located relatively closely (separation between 0.04 and 0.15), MVESA obtains a significantly high resolution, at the same time, other algorithms almost fail to work due to the resolution limit caused by suboptimal sparse metrics. The high resolution of MVESA validates the importance of the atomic \( l_0 \) norm for promoting sparsity.

2) Handling on Incomplete Data: Simulation 4 investigates the capability of handling incomplete sample data. We consider the estimation of \( M = 20 \), \( K = 3 \), \( T = 10 \) snapshots, and \( \text{SNR} = 10 \) dB. Particularly, we extract \( M_{\text{sel}} \leq M \) measurements from complete measurements \( Y \) with indices in \( M \subseteq \{0, \ldots, M-1\} \), \( |M|=M_{\text{sel}} \); thus, the resultant measurements data are incomplete. The RMSE and success rate results for incomplete data are shown in Fig. 9. It can be observed that for \( M_{\text{sel}} < 14 \), MVESA retains a better estimation performance in terms of RMSE and success rate. For \( 14 \leq M_{\text{sel}} < 20 \), MVESA achieves slightly worse frequency error than VALSE_MMV and RAM, and far surpasses SPA and APG. This phenomenon may be because the compared algorithms are quite sensitive to the setting of the upper bound of the \( L_2 \) norm of the Gaussian noise \[24\]. But with more complete data, the compared algorithms can obtain more accurate estimate of noise variance, such that the frequency estimation is more accurate. While our MVESA relies on the polynomial mutation to produce close-to-optimal frequencies, the randomness and efficiency of polynomial mutation may slightly affect the frequency estimation accuracy. To sum up, our method is more valuable for difficult cases with fewer incomplete data, while the compared algorithms almost fail to estimate frequencies in these cases. See the success rate under \( M_{\text{sel}} < 16 \) in Fig.9(b) for experimental proof.

3) Time Cost Analysis: Simulation 5 compares the time complexity of all algorithms under different \( M \)s. This simulation is implemented in MATLAB R2018b on a PC with Intel i7-7700 CPU and 32-GB RAM. Parametric settings are set the same as Fig. 9. The computational time is displayed in Table I. VALSE_MMV runs the fastest, following by MVESA and SPA, while RAM and APG are the most time consuming. Compared to VALSE_MMV, MVESA takes more running time, which is due to two reasons. On the one
Fig. 9. (a) RMSE(θ) and (b) success rate results of all algorithms under varying \( M \) for \( K = 3, T = 10, \) and \( \text{SNR} = 10 \) dB.

### TABLE I

| \( M \) | SPA   | RAM   | APG   | VALSE MMV | MVESA   |
|--------|-------|-------|-------|-----------|---------|
| 8      | 0.7609| 5.3165| 0.8966| 0.0400    | 0.6170  |
| 12     | 0.7846| 5.6782| 1.5747| 0.0603    | 0.7068  |
| 16     | 0.8237| 5.7789| 2.6863| 0.0913    | 0.8037  |
| 20     | 0.9202| 5.6872| 4.5730| 0.1133    | 0.9575  |
| 24     | 0.9233| 5.8040| 6.5828| 0.1443    | 1.0924  |

hand, VALSE MMV iteratively updates a single solution, while our algorithm is a population-based search algorithm that updates multiple solutions at each generation. On the other hand, recovering the signal intensity occupies the dominant running time. VALSE MMV performs intensity recovery by matrix multiplication with low complexity \( O(M^2 + ML) \). Comparatively, our algorithm employs matrix inversion to update the signal intensity, which has the complexity of \( O(M^3) \) for each solution. Therefore, our algorithm takes more time. Despite this, our algorithm achieves great improvements in RMSE and success rate over the state-of-the-art algorithms.

To sum up, our algorithm is more suitable for challenging scenarios, such as lower SNRs, more sources, closely spaced frequencies, or incomplete sample data with fewer measurements. The performance improvement of our algorithm in these cases can be clearly observed from Figs. 6–9. In these figures, our algorithm achieves lower RMSE values and higher success rates than comparison algorithms. Therefore, our algorithm is an excellent option when a high precision is desired in these challenging scenarios. For large-scale real-time applications, we suggest accelerating MVESA by parallel implementation [53], [54] to accelerate the implementation of our algorithm.

### VI. CONCLUSION

In this article, we have proposed a novel idea of simultaneously estimating the model order and frequencies by means of the atomic \( l_0 \) norm. To accomplish this, we have built a multiobjective optimization model, with the measurement error and atomic \( l_0 \) norm being the two objectives. The atomic \( l_0 \) norm directly exploits sparsity without relaxations, breaking the resolution limit and estimating the model order accurately.

To solve the resultant NP-hard problem, we have designed the MVESA with two innovations. One is the variable-length coding and search strategy, which provides a flexible representation of frequencies with different sizes, and implement full exploration over the variable and open-ended search space. Another innovation is the model-order pruning mechanism, which reduces the solutions’ redundancy by heuristically pruning less contributive frequencies. This mechanism highly improves the convergence and diversity performance. Experimental results have demonstrated the superiority of MVESA in terms of RMSE and success rate.

The proposed method involves matrix inversions, which is time consuming for large-scaled LSE problems. Therefore, in the future, we plan to design more computational efficient methods. We also expect to further improve MVESA to be statistically consistent so that the frequency estimation performance in cases with large \( M \) would be improved.

### REFERENCES

[1] P. Stoica and R. L. Moses, Spectral Analysis of Signals. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.
[2] T.-Q. Xia, “Joint diagonalization based DOD and DOA estimation for bistatic MIMO radar,” *Signal Process.*, vol. 108, pp. 159–166, Mar. 2015.
[3] B. D. Rao and K. V. S. Hari, “Performance analysis of root-music,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.
[4] S.-Y. Kung, K. S. Arun, and D. V. B. Rao, “State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem,” *J. Opt. Soc. Amer.*, vol. 73, no. 12, pp. 1799–1811, 1983.
[5] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IRE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[6] R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
[7] F. Liu et al., “Nonconvex compressed sensing by nature-inspired optimization algorithms,” *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1042–1053, May 2015.
[8] L. Lin, F. Liu, L. Jiao, S. Yang, and H. Hao, “The overcomplete dictionary-based directional estimation model and nonconvex reconstruction methods,” *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1042–1053, Mar. 2018.
[9] G. Mei, X. Wu, Y. Wang, M. Hu, J.-A. Lu, and G. Chen, “Compressive-sensing-based structure identification for multilayer networks,” *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 754–764, Feb. 2018.
Bai Yan received the Ph.D. degree in optical engineering from the Beijing University of Technology, Beijing, China, in 2019.

She was a joint Ph.D. student of Computational and Communication Engineering with the University of Technology Sydney, Ultimo, NSW, Australia, from 2017 to 2018. She was also a Postdoctoral Fellow of Computer Science and Technology with the Southern University of Science and Technology (SUSTech), Shenzhen, China, from 2019 to 2021, where she is currently a Research Assistant Professor with the Research Institute of Trustworthy Autonomous Systems, Guangdong Provincial Key Laboratory of Brain-Inspired Intelligent Information Processing, SUSTech. Her research interests include computational intelligence, signal processing for wireless communications, and reconfigurable intelligent software-aided communications.
Qi Zhao received the Ph.D. degree in management science and engineering from the Beijing University of Technology, Beijing, China, in 2019.

He was a joint Ph.D. student of Computer Science with the University of New South Wales, Canberra, ACT, Australia, from 2017 to 2018. He was a Postdoctoral Fellow of Computer Science and Technology with the Southern University of Science and Technology, Shenzhen, China, from 2019 to 2021, where he has been a Research Assistant Professor with the Department of Computer Science and Engineering since 2021. His research interests include evolutionary computation, automated algorithm design, and operations research.

Jin Zhang received the B.E. and M.E. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2004 and 2006, respectively, and the Ph.D. degree in computer science from the Hong Kong University of Science and Technology, Hong Kong, in 2009.

She is currently an Associate Professor with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. Her research interests are mainly in mobile healthcare and wearable computing, wireless communication and networks, network economics, cognitive radio networks, and dynamic spectrum management.

J. Andrew Zhang (Senior Member, IEEE) received the B.Sc. degree from Xi’an JiaoTong University, Xi’an, China, in 1996, the M.Sc. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1999, and the Ph.D. degree from the Australian National University, Canberra, ACT, Australia, in 2004.

He is currently an Associate Professor with the School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW, Australia. His research interests are in the area of signal processing for wireless communications and sensing, and autonomous vehicular networks. He has published more than 200 papers in leading international Journals and conference proceedings.

Dr. Zhang has won five Best Paper Awards for his work.

Xin Yao (Fellow, IEEE) received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, China, in 1982, the M.Sc. degree from the North China Institute of Computing Technology, Beijing, China, in 1985, and the Ph.D. degree from USTC in 1990.

He is a Chair Professor and the Founding Head of Computer Science and Engineering with the Southern University of Science and Technology, Shenzhen, China, and a part-time Professor of Computer Science with the University of Birmingham, Birmingham, U.K. His major research interests include evolutionary computation and ensemble machine learning. In particular, he is very interested in metaheuristics for dynamic optimization and capacitated arc routing problems.

Dr. Yao received the 2012 Royal Society Wolfson Research Merit Award, the 2013 IEEE Computational Intelligence Society (CIS) Evolutionary Computation Pioneer Award, and the 2020 IEEE Frank Rosenblatt Award. His work won the 2001 IEEE Donald G. Fink Prize Paper Award, the 2010, 2016, and 2017 IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION Outstanding Paper Awards, the 2011 IEEE Transactions on Neural Networks Outstanding Paper Award, the 2010 BT Gordon Radley Award for the Best Author of Innovation (Finalist), and several best paper awards at conferences. He was a Distinguished Lecturer of IEEE CIS. He served as the President of IEEE CIS from 2014 to 2015 and the Editor-in-Chief for IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION from 2003 to 2008.