A Robust and Statistically Efficient Maximum-Likelihood Method for DOA Estimation Using Sparse Linear Arrays

ZAI YANG, Senior Member, IEEE
XINYAO CHEN
XUNMENG WU, Senior Member, IEEE
Xi’an Jiaotong University, Xi’an, China

The recent trend of research on direction-of-arrival estimation is to localize more uncorrelated sources than sensors by using a proper sparse linear array (SLA) at the cost of robustness to source correlations even in the regime of less sources than sensors. This article is devoted to proposing one algorithm that can simultaneously tackle two challenging scenarios: 1) more uncorrelated sources than sensors and 2) highly correlated or coherent sources. In order to statistically efficiently localize a maximal number of uncorrelated sources, we use the stochastic maximum likelihood (SML) criterion and propose an effective algorithm based on elegant problem reformulations and the alternating direction method of multipliers (ADMM). Moreover, we prove that the SML is robust to source correlations under mild conditions, though it is derived under the assumption of uncorrelated sources. The proposed algorithm is usable for arbitrary SLAs (e.g., minimum redundancy arrays, nested arrays, and coprime arrays) and is named as maximum-likelihood estimation via sequential ADMM (MESA). Extensive numerical results are provided that collaborate our analysis and demonstrate the statistical efficiency and robustness of MESA against state-of-the-art algorithms. Our results also imply that it is possible to localize more sources than sensors in the presence of high source correlations.

I. INTRODUCTION

DIRECTION-OF-ARRIVAL (DOA) estimation is a fundamental problem in statistical and array signal processing. It refers to the problem of estimating the directions of a number of sources impinging on a sensor array given a series of snapshots of the output of the sensor array [3]. In this article, we consider DOA estimation for far-field narrowband sources using a uniform linear array (ULA) or sparse linear array (SLA), resulting in a DOA estimation problem equivalent to multiple-snapshot spectral analysis, a topic at the core of wireless channel estimation [4], radar signal processing [5], structural health monitoring [6], and fluorescence microscopy [7]. The SLA corresponds to the missing data case in the language of spectral analysis [8] or compressive data in the language of compressed sensing [9, 10] and brings new challenges to theoretical analysis and algorithm design.

The use of SLAs for DOA estimation dates back to [11] and has been extensively studied in the past two decades with an emphasis of localizing $O(M^2)$ sources using $M$ sensors only. The key to achieving such a goal is that under the assumption of uncorrelated sources, the data covariance matrix regarding a ULA becomes Toeplitz and, thus, can be determined by a few of its entries. With this in mind, different array geometries for SLAs, e.g., minimum redundancy arrays (MRAs) [11], nested arrays [12, 13, 14], and coprime arrays [15, 16], have been proposed to determine which entries (indexed by the coarray) are sampled to reconstruct the whole or a shrunken version of the Toeplitz covariance matrix. The assumption of uncorrelated sources is crucial to guarantee the Toeplitz covariance structure; however, it is not always satisfied. In fact, correlated and coherent (fully correlated) sources usually occur in practice, due to multipath propagations and other effects, and dealing with them has always been a central topic in DOA estimation (see, e.g., [17], [18], [19], [20], [21], and [22]). Consequently, the goal of localizing more uncorrelated sources than sensors by using the Toeplitz covariance structure seemingly contradicts with the one of robust localization of highly correlated and coherent sources due to the destruction of the Toeplitz structure, making the practical use of previous methods questionable in correlated environments. In this article, we make an attempt to propose one algorithm that can simultaneously handle the two aforementioned challenging scenarios: 1) more uncorrelated sources than sensors and 2) highly correlated or coherent sensors.

It is well known that the maximum likelihood (ML) method, if solvable, provides benchmark performance for DOA estimation. Under the assumption of uncorrelated sources, the stochastic ML (SML) method can be used to localize the maximal number of sources with statistical efficiency. Its asymptotic performance, in terms of the Cramér–Rao bound (CRB), has been well understood [23], [24]. However, few algorithms have been proposed for the SML method since it resorts to a highly nonconvex optimization problem. The challenges arise due to the nonlinearity with
respect to the DOAs, the nonconvex log-det term in the SML criterion function, and the source number constraint, and it becomes even worse in the SLA case. In this article, in order to accomplish the two aforementioned tasks using one algorithm, we present an effective algorithm for the SML and prove that the SML derived for uncorrelated sources is robust to source correlations under mild conditions. Our main contributions are summarized as follows.

1) We start with the specialized ULA case and formulate the SML optimization problem as a rank-constrained Toeplitz covariance estimation problem, which is further transformed as sequential rank-constrained semidefinite programs (SDPs) by applying a majorization–minimization (MM) technique [25]. An elegant reformulation of the rank-constrained SDP is derived to fit and solved using the alternating direction method of multipliers (ADMM) encouraged by its successes in solving nonconvex problems [26], [27]. The resulting algorithm is named as maximum-likelihood estimation via sequential ADMM (MESA) (see Section III).

2) In the general SLA case, we repeat the above derivations and show that the SML problem can be similarly solved, extending MESA to this case (see Section IV).

3) While the SML method concerned in this article is derived under the assumption of uncorrelated sources, we prove that it produces consistent estimates of the DOAs and the source powers (regardless of source correlations) as the noise vanishes in the presence of less sources than sensors, implying its robustness to correlated and coherent sources (see Section VI-D).

Besides the contributions above, numerical results are provided confirming the statistical efficiency of MESA for uncorrelated source localization, in cases when the source number is either less than or greater than the sensor number, and its robustness to correlated and coherent sources. Remarkably, our numerical results also show that MESA can localize more sources than sensors in the presence of source correlations, opening a future research direction (see Section VI-E).

A. Relations to Prior Art

The SML method is also known as unconditional ML and has a long history of research. Its asymptotic performance in the ULA case, in terms of the CRB, is well documented in the literature [28], [29]. To solve the SML optimization problem, expectation maximization and Newton-type algorithms have been proposed in earlier works [30], [31], but their performance heavily depends on the initialization step. Instead of solving for an exact ML estimator, great efforts have been made to develop algorithms that have the same asymptotic performance as the SML. Such examples include multiple signal classification (MUSIC) [32], method of direction estimation [33], and weighted subspace fitting [34]. Good reviews can be found in [35] and [36]. However, it is worth noting that these algorithms usually consider the ULA and some assume deterministic (as opposed to uncorrelated) sources, and thus, they cannot be used to localize more sources than sensors with an SLA.

Sparse optimization and compressed sensing methods [37], [38], [39], which have become popular since early of this century, do not use explicitly the array geometry and fit into the SLA case. The recent atomic norm and gridless compressed sensing methods [40], [41], [42], [43], [44] are remedies of earlier compressed sensing methods by working with continuous (as opposed to on-grid) DOAs and providing theoretical guarantees. These methods do not make statistical assumptions on the sources and are robust to source correlations. But correspondingly, they cannot localize more sources than sensors. Readers are referred to [45] for a review.

To localize more uncorrelated sources than sensors, a coarray-based averaging (CBA)/selection step is usually adopted to explicitly use the Toeplitz covariance structure and transform the sample covariance matrix regarding the SLA as an output regarding an enlarged virtual ULA, followed by a DOA estimation method for ULAs; see such two-step estimation approaches in [46], [47], [48], [49], [50], and [51], to name just a few. It is shown in [23] that CBA combined with spatial-smoothing (SS) MUSIC results in strictly nonefficient solutions. A state-of-the-art method is proposed in [52] that uses a weighted least square (WLS) criterion for the vectorized sample covariance and is shown to yield an asymptotically efficient estimator. An iterative algorithm is also proposed to solve the resulting nonconvex optimization problem. While these methods are tailored for uncorrelated sources, it is confirmed by numerical results in this article that they are indeed sensitive to highly correlated sources even in the regime of less sources than sensors. In contrast to this, MESA achieves statistical efficiency and robustness to source correlations simultaneously.

Several algorithms have been proposed to deal with a mixture of uncorrelated and coherent sources given the source coherence structure [20], [21], [22], [53]. Differently from these algorithms, MESA allows the sources to be correlated but noncoherent and needs only knowledge of the total number of sources (as opposed to the detailed coherence structure).

The SML method in the ULA case is closely related to structured (to be specific, Toeplitz) covariance estimation (see, e.g., [54], [55], [56], [57], and [58]) because the data covariance matrix is the sum of a low-rank Toeplitz covariance and the noise covariance, where the rank is specified by the source number and the DOAs are uniquely determined by the Toeplitz covariance matrix. While the low-rank constraint is a major challenge for Toeplitz covariance estimation, it is explicitly considered in [59] and [60]. In [59], the Toeplitz structure is relaxed initially and then used to obtain a Toeplitz approximation of an intermediate solution, which does not result in an exact SML estimator. In [60], the
noise variance is assumed known, and the Carathéodory–Fejér theorem [45, Th. 11.5] is invoked to approximate the Toeplitz covariance by a Vandermonde decomposition, in which the frequency nodes of the Vandermonde matrix are restricted on a fixed grid so that the original problem is transformed as one of nonnegative sparse vector recovery. In contrast to these methods, we make no approximations or relaxations regarding the Toeplitz covariance, and MESA solves the exact SML. Moreover, MESA is usable in the SLA case.

Since the difficulty in solving the SML problem partly comes from the source number constraint, which is known as signal sparsity in compressed sensing, it is relaxed in sparse Bayesian learning methods [61], [62], [63]. Similar relaxation techniques are also used in covariance fitting methods [64], [65], [66], [67], [68], which are approximate versions of the SML method by using convex surrogates for its criterion function. Interestingly, it has been empirically observed in [61], [65], and [66] that the resulting algorithms are robust to source correlations, though they are derived by assuming uncorrelated sources. In the recent work [69], the case of two correlated sources is considered, and it is shown that if the DOAs and the noise variance are known a priori, then the source powers can be stably estimated from the SML method. In contrast to this, our result on robustness is applicable to arbitrary source number and shows that the DOAs can be accurately estimated jointly with the source powers, at least in the high-signal-to-noise ratio (SNR) regime. It also partially explains the observations in [61], [65], and [66].

B. Notation

The sets of real and complex numbers are denoted by \( \mathbb{R} \) and \( \mathbb{C} \), respectively. For vector \( x \), \( \text{diag}(x) \) denotes a diagonal matrix with \( x \) on the diagonal. The \( j \)-th entry of vector \( x \) is \( x_j \). For matrix \( A \), \( A^T \), \( A^H \), \( |A| \), \( \text{rank}(A) \), tr\( (A) \), \( \lambda_{\min}(A) \), and \( ||A||_F \) denote the matrix transpose, conjugate transpose, determinant, inverse, rank, trace, minimum eigenvalue, and Frobenius norm of \( A \), respectively. Given vector \( t = [t_1, \ldots, t_{N-1}]^T \) satisfying \( t_j = t_{j+1} \) for \( j = 0, \ldots, N-1 \), \( T T = (t_{j-1} t_j, \ldots, t_{N-1} t_N) \) denotes an Hermitian Toeplitz matrix. The complex conjugate transpose of vector \( x \) is denoted by \( \bar{x} \). The real part of scalar \( x \) is denoted by \( \Re(x) \). The notation \( A \succeq 0 \) means that \( A \) is Hermitian positive semidefinite. For index set \( \Omega \) and matrix \( A, A_\Omega \) represents a submatrix of \( A \) obtained by keeping only the rows indexed by \( \Omega \) unless otherwise stated. The inner product of vectors or matrices \( A \) and \( B \) of the same dimension is denoted by \( \langle A, B \rangle = \Re(\text{tr}(A^H B)) \). For matrix \( Y \) and singular matrix \( C \succeq 0 \), we define \( \text{tr}(Y^H(C + \sigma I)^{-1}Y) \equiv \lim_{\sigma \rightarrow 0} \text{tr}(Y^H(C + \sigma I)^{-1}Y) \), where \( \sigma > 0 \) and \( I \) is an identity matrix. It follows immediately that

\[
\text{tr}(Y^H C^{-1}Y) = \min_x \text{tr}(X), \quad \text{subject to} \quad \begin{bmatrix} X & Y^H \\ Y & C \end{bmatrix} \succeq 0. 
\] (1)

The expectation of a random variable is denoted by \( \mathbb{E}[\cdot] \).

II. PRELIMINARIES

A. DOA Estimation Using SLAs

An \( M \)-element SLA of aperture \( N - 1 \) composes a subset of an \( N \)-element virtual ULA. Let the index set \( \Omega \subset \{1,\ldots,N\} \), of cardinality \( M \leq N \), denote the SLA. We first consider the special ULA case when \( \Omega = \{1,\ldots,N\} \) and \( M = N \). Assume that \( K \) far-field narrowband sources impinge on the ULA in which adjacent sensors are placed by half a wavelength apart. The output of the sensor array at each snapshot constitutes an \( N \times 1 \) complex vector \( y \) that can be modeled as [3], [36]

\[
y(l) = \sum_{k=1}^{K} a(f_k) s_k(l) + e(l), \quad l = 1, \ldots, L
\] (2)

where \( L \) is the number of snapshots, \( s_k(l) \) is the \( k \)-th (complex) source signal at the \( l \)-th snapshot, \( f_k \in [-\frac{1}{2}, \frac{1}{2}] \) has a one-to-one relation to the \( k \)-th DOA \( \theta_k \in [-90^\circ, 90^\circ] \) by \( f_k = \frac{\pi}{\lambda} \sin \theta_k \), \( a(f_k) \) denotes an \( N \times 1 \) steering vector given by

\[
a(f) = \left[ 1, e^{2\pi f}, \ldots, e^{2(N-1)\pi f} \right]^T
\] (3)

and \( e(l) \) is the vector of complex noise. It is seen that all the snapshots share the same parameters \( \{\theta_k\} \) and \( \{f_k\} \). By stacking \( \{s_k(l)\}, \{f_k\} \) into vectors \( s(l) \) and \( f \) and defining the steering matrix \( A(f) = [a(f_1), \ldots, a(f_K)] \) that is \( N \times K \) Vandermonde, the data model in (2) is written compactly as

\[
y(l) = A(f)s(l) + e(l), \quad l = 1, \ldots, L
\] (4)

In the general SLA case, the array output at one snapshot is a subvector of \( y(l) \), denoted by \( y_\Omega(l) \). The data model in (4), thus, becomes

\[
y_\Omega(l) = A_\Omega(f)s(l) + e_\Omega(l), \quad l = 1, \ldots, L
\] (5)

which subsumes (4) as a special case.

Our objective is to estimate the DOAs \( \{\theta_k\}_{k=1}^{K} \), or equivalently \( \{f_k\}_{k=1}^{K} \), given the multiple-snapshot data \( \{y_\Omega(l)\}_{l=1}^{L} \) under certain statistical assumptions on the source signals \( \{s(l)\} \) and noise \( \{e(l)\} \). Since each \( f_k \) is the frequency of a sinusoid, the DOA estimation problem that we concern is equivalent to multiple-snapshot spectral estimation with missing data. We focus on the estimation of \( \{f_k\}_{k=1}^{K} \) throughout this article.

B. SML Method for DOA Estimation

We make the following assumptions to derive the SML method for DOA estimation.

A1: The sources \( \{s(l)\} \) are spatially and temporally independent and follow a complex Gaussian distribution with zero mean and covariance \( P = \text{diag}(p_1, \ldots, p_K) \), where \( p_k > 0 \) denotes the \( k \)-th source power.

A2: The noises \( \{e(l)\} \) are spatially and temporally independent, and each entry follows a complex Gaussian distribution with zero mean and variance \( \sigma > 0 \).
A3: The sources and noises are independent.

It follows immediately that \( \{ y_i(l) \} \) is independent identically distributed Gaussian with zero mean and covariance 

\[
R_\Omega = A_\Omega(f) PA_\Omega^H(f) + \sigma I.  \tag{6}
\]

By maximizing the likelihood criterion, or equivalently minimizing the negative log-likelihood function, we obtain the SML optimization problem as 

\[
\min_{f, \sigma} \ln |R_\Omega| + \text{tr}(R_\Omega^{-1} \hat{R}_\Omega)
\]

where

\[
\hat{R}_\Omega = \frac{1}{L} \sum_{l=1}^{L} y(l)y(l)^H
\]

is the sample covariance matrix.

The SML method has good statistical properties. However, the SML problem in (7) is nonconvex and complicated to solve due to the log-det term and the nonlinearity of \( R_\Omega \) with respect to \( \{ f_k \}_{k=1}^K \). Moreover, the SML is derived under the assumption of uncorrelated sources, and its performance is unclear in the presence of source correlations.

C. ADMM Algorithm

The ADMM algorithm solves the following optimization problem:

\[
\min_{x \in D_1, q \in D_2} g(x) + h(q), \quad \text{subject to } Ax + Bq = e  \tag{9}
\]

where \( D_1 \) and \( D_2 \) define the feasible domains of \( x \) and \( q \), respectively, and \( A \) and \( B \) are linear operators. Write the augmented Lagrangian function as

\[
\mathcal{L}_\mu(x, q, \lambda) = g(x) + h(q) + \langle Ax + Bq - e, \lambda \rangle + \frac{\mu}{2} \| Ax + Bq - e \|^2_2
\]

\[
= g(x) + h(q) + \frac{\mu}{2} \| Ax + Bq - e + \mu^{-1} \lambda \|^2_2 + C  \tag{10}
\]

where \( \lambda \) is a Lagrangian multiplier, \( \mu > 0 \) is a penalty coefficient, and \( C \) is a constant independent of \( x \) and \( q \). The ADMM consists of the following iterations:

\[
x \leftarrow \arg \min_{x \in D_1} \mathcal{L}_\mu(x, q, \lambda)  \tag{11}
\]

\[
q \leftarrow \arg \min_{q \in D_2} \mathcal{L}_\mu(x, q, \lambda)  \tag{12}
\]

\[
\lambda \leftarrow \lambda + \mu (Ax + Bq - e)  \tag{13}
\]

where the latest values of the other variables are always used. The ADMM algorithm has been extensively studied and practically used due to its global optimality in solving convex problems, simplicity in dealing with nonsmooth functions, and good scalability for solving high-dimensional problems [26]. Good performance has also been frequently achieved for nonconvex problems [27], [70], [71], [72], [73], and [74]. It is worth noting that the key to using ADMM to solve a specific problem is to provide an elegant problem formulation within the ADMM framework so that the two subproblems in (11) and (12) can be simply and efficiently solved.

III. MESA in the ULA Case

In this section, we derive the MESA algorithm for the SML optimization problem in (14) in the specialized ULA case. In this case, we write the data and sample covariance matrices \( R_\Omega \) and \( \hat{R}_\Omega \) into \( R \) and \( \hat{R} \) for simplicity, and the problem to solve becomes

\[
\min_{f, \sigma} \ln |R| + \text{tr}(R^{-1} \hat{R})  \tag{14}
\]

where

\[
R = A(f) PA^H(f) + \sigma I.  \tag{15}
\]

The MESA algorithm consists of reparameterization, MM, problem reformulation, and ADMM steps, which are detailed as follows.

A. Reparameterization

The data covariance matrix \( R \) is a highly nonlinear function of \( \{ f_k \} \). To overcome such nonlinearity, a common scheme is to utilize the fact that the first term \( A(f) PA^H(f) \) in (15) is a Hermitian positive-semidefinite Toeplitz matrix of rank no greater than \( K \) and do the reparameterization [59], [60], [66].

\[
R = Tr + \sigma I, \quad Tr \succeq 0, \quad \text{rank} (Tr) \leq K.  \tag{16}
\]

It follows from the Carathéodory–Fejér theorem [45, Th. 11.5] that \( Tr \) above has a one-to-one connection to \( \{ f, p \} \) given \( K < N \). Consequently, the original SML problem (15) is transformed into a rank-constrained Toeplitz covariance estimation problem, in which \( R \) is a linear function of the variables \( \{ f, \sigma \} \). Once \( Tr \) is solved, the variables \( \{ f, \sigma \} \) can be computed from \( Tr \) by a subspace method, such as root-MUSIC [75].

B. Majorization Minimization

The objective function in (14) is nonconvex with respect to \( R \) since the log-det function \( \ln |R| \) is concave on the positive-semidefinite cone. A commonly used locally convergent method is the MM algorithm (see, e.g., [76]) that derives the objective function downhill by minimizing a simple surrogate function. At the \( j \)th iteration of MM, the SML objective function is linearized (and, thus, majorized) at the previous iterate \( R_{j-1} = Tr_{j-1} + \sigma_{j-1} I \), yielding the problem (by omitting constant terms)

\[
\min \text{tr}(R_{j-1}^{-1} R) + \text{tr}(R_{j-1}^{-1} \hat{R}) .  \tag{17}
\]

Substituting (16) into (17), we obtain the problem to solve at the \( j \)th iteration as

\[
\min_{t, \sigma \geq 0} \text{tr}(R_{j-1}^{-1}(Tr + \sigma I)) + \text{tr}((Tr + \sigma I)^{-1} \hat{R})
\]

\[
\text{subject to } Tr \in S^K_+  \tag{18}
\]

where \( S^K_+ \) is the set of positive-semidefinite matrices of rank no greater than \( K \).

YANG ET AL.: ROBUST AND STATISTICALLY EFFICIENT MAXIMUM-LIKELIHOOD METHOD FOR DOA ESTIMATION 6801
C. Problem Reformulation

Let \( W = R_{i,j}^{-1} \) and \( \hat{Y} \) be any matrix satisfying that
\[
\hat{R} = \hat{Y} \hat{Y}^H
\]
where \( \hat{Y} \) has at most \( \min(L, N) \) columns. The objective function in (18) then becomes
\[
\text{tr}(W(Tt + \sigma I)) + \text{tr}\left((\hat{Y}^H(Tt + \sigma I)^{-1})\hat{Y}\right). \tag{20}
\]

Making use of the following identity [77, Lemma 5]:
\[
\text{tr}(X^H(R_1 + R_2)^{-1} X) = \min_Z \text{tr}\left(Z^H R_1^{-1} Z + \text{tr}\left((X - Z)^H R_2^{-1} (X - Z)\right)\right) \tag{21}
\]
where \( R_1, R_2 \geq 0 \), the function in (20) becomes a function of \( t, \sigma \), and \( Z \)
\[
\text{tr}(W(Tt + \sigma I)) + \text{tr}(Z^H (Tt)^{-1} Z) + \sigma^2 \|\hat{Y} - Z\|_F^2. \tag{22}
\]
Since, in (22), the optimizer to \( \sigma \) is given in closed form by
\[
\sigma^* = \frac{1}{\sqrt{\text{tr}(W)}} \|\hat{Y} - Z\|_F \tag{23}
\]
the objective in (22) can be concentrated with respect to \( t \) and \( Z \), yielding the following problem to solve:
\[
\min_{t, \lambda} \text{tr}(W Tt) + \text{tr}(X) + 2\sqrt{\text{tr}(W)} \|\hat{Y} - Z\|_F \tag{24}
\]
subject to \( Z \in S^K \).

Next, we show that the problem in (25) is equivalent to the following:
\[
\min_{t, \lambda} \text{tr}(W Tt) + \text{tr}(X) + 2\sqrt{\text{tr}(W)} \|\hat{Y} - Z\|_F \tag{26}
\]
subject to \( X = \begin{bmatrix} X & Z^H \\ Z & Tt \end{bmatrix} \geq 0, \text{rank}(Tt) \leq K. \)

D. Using ADMM

We introduce an auxiliary matrix variable \( Q \) and rewrite (26) as
\[
\min_{t, x, z, q \in \mathbb{S}^K_+} \text{tr}(W Tt) + \text{tr}(X) + 2\sqrt{\text{tr}(W)} \|\hat{Y} - Z\|_F \tag{28}
\]
subject to \( Q = \begin{bmatrix} X & Z^H \\ Z & Tt \end{bmatrix} \)

which is exactly in the form of (9) by identifying that \( x = (t, X, Z) \) (\( x \), here, is a notation representing a set of variables), \( q = Q, Ax = \begin{bmatrix} X & Z^H \\ Z & Tt \end{bmatrix} B \) is an identity operator, \( g(x) = \text{tr}(W Tt) + \text{tr}(X) + 2\sqrt{\text{tr}(W)} \|\hat{Y} - Z\|_F, h(q) = 0, D_1 \) is the whole space, and \( D_2 = \mathbb{S}^K_+ \). Following the procedures of ADMM, we introduce the Hermitian Lagrangian multiplier \( A \) and write the augmented Lagrangian function as
\[
\mathcal{L}_\mu(t, X, Z, Q) = \text{tr}(W Tt) + \text{tr}(X) + 2\sqrt{\text{tr}(W)} \|\hat{Y} - Z\|_F + \text{tr}(Q - AX)^H A + \frac{\mu}{2} \|Q - AX\|_F^2 \tag{29}
\]

The remaining task is to solve the two subproblems in (11) and (12).

To solve (11), by partitioning \( Q = [\begin{bmatrix} Q_{1} & 0 \\ 0 & Q_2 \end{bmatrix} \] and \( A = [\begin{bmatrix} A_1 & A_2 \\ A_2^* & A_3 \end{bmatrix} \), we note that the objective function \( \mathcal{L}_\mu \) is separable in \( [t, X, Z] \), and thus, can be solved for separately. To solve for \( t \), let \( \mathcal{T}^* \) be the Hermitian adjoint of \( T \) that satisfies \( (Tt, A) = (t, \mathcal{T}^* A) \) for any \( t \) and Hermitian matrix \( A \) of proper dimension, to be specific, \( \mathcal{T}^* A = [a_{1-N}, \ldots, a_{N-1}]^T \), where \( a_0 = (tA) \) and \( a_i \) equals the sum of entries in the \( i \)th diagonal of \( A \). It follows immediately that \( \mathcal{T}^* Tt = Dt \), where \( D = \text{diag}[1, 2, \ldots, N-1, N, N-1, \ldots, 2, 1] \). Note by (29) that
\[
\mathcal{L}_\mu = \langle W, Tt \rangle + \frac{\mu}{2} \|Tt - Q_3 - \mu^{-1} A_3\|_F^2 + C_1
\]
\[
= \langle \mathcal{T}^* (W - \mu Q_3 - A_3), t \rangle + \frac{\mu}{2} \|\mathcal{T}^* Tt, t\|_F^2 + C_2 \tag{30}
\]

where \( C_1 \) and \( C_2 \) are constants independent of \( t \), and \( W \) is Hermitian. It follows that the derivative of \( \mathcal{L}_\mu \) with respect to \( t \) is given by [78, Sec. 4.1]
\[
\mathcal{T}^* (W - \mu Q_3 - A_3) + \mu \mathcal{T}^* Tt = \mathcal{T}^* (W - \mu Q_3 - A_3) + \mu Dt. \tag{31}
\]
By equating (31) to zero, we obtain the update
\[
t \leftarrow D^{-1} \mathcal{T}^* (Q_3 + \mu^{-1} A_3 - \mu^{-1} W). \tag{32}
\]
Similarly, the derivative of $L_\mu$ with respect to $X$ is given by $\mu X - \mu Q_1 - \Lambda_1 + I$, and we have the update

$$X \leftarrow Q_1 + \mu^{-1} \Lambda_1 - \mu^{-1} I.$$  \hfill (33)

For $Z$, the optimization problem to solve is given by

$$\min_{Z} \sqrt{tr(W)}||\hat{Y} - Z||_F + \frac{\mu}{2} ||Q_2 + \mu^{-1} \Lambda_2 - Z||_F^2. \hfill \text{(34)}$$

Let $L = \hat{Y} - Q_2 - \mu^{-1} \Lambda_2$. Then, we have the update

$$Z \leftarrow \begin{cases} \hat{Y}, & \text{if } L = 0 \\ \hat{Y} - \left(1 - \frac{\mu ||W||}{\mu ||L||_F}\right) + L, & \text{otherwise} \end{cases} \hfill \text{(35)}$$

where $(x)_+ \triangleq \max(x, 0)$. The detailed derivations of (35) are deferred to Appendix A.

Solving (12) results in the update [79]

$$Q \leftarrow \mathcal{P}_{S^k_+} \left( \begin{bmatrix} X & Z^H \\ Z & Tt \end{bmatrix} - \mu^{-1} \Lambda \right)$$  \hfill (36)

where $\mathcal{P}_{S^k_+}$ denotes the orthogonal projection onto $S^k_+$ that can be computed by the truncated eigendecomposition by keeping only the largest $K$ positive eigenvalues and associated eigenvectors (note that if the number of positive eigenvalues is less than $K$, then just keep those positive ones).

Finally, $\Lambda$ is updated according to (13) as

$$\Lambda \leftarrow \Lambda + \mu \left( Q - \begin{bmatrix} X & Z^H \\ Z & Tt \end{bmatrix} \right). \hfill \text{(37)}$$

The ADMM algorithm runs (32), (33), and (35)–(37) iteratively.

E. Convergence and Complexity of the Algorithm

The overall MESA algorithm consists of the outer MM loop and the inner ADMM loop. A theoretical proof of convergence of MESA requires the global optimality of each ADMM subalgorithm and cannot be accomplished currently. In fact, to the best of our knowledge, no results in the literature are available on the global optimality of ADMM for nonconvex optimization. It is shown in [70] and [71] that if the solution sequence produced by ADMM converges, then it converges to a stationary point. Better convergence results have been derived in [72], [73], and [74] in the case when the mapping $x \rightarrow Ax$ is surjective, an assumption that is not true for our problem. In contrast to the theoretical difficulties, good convergence and optimality of ADMM and the whole MESA algorithm will be demonstrated via extensive numerical simulations in Section VI, like many other successes of ADMM for nonconvex optimization [27].

The computations of MESA are dominated by the truncated eigendecomposition and the matrix inverse to compute $W$. Consequently, MESA has a computational complexity of $O(N^3)$ per inner iteration that is affordable in DOA estimation since the array aperture $N$ is usually small. The whole computational complexity equals the computational complexity per inner iteration multiplied by the whole number of inner iterations that will be reported in numerical simulations.

IV. MESA IN THE SLA CASE

In this section, we consider the general SLA case and derive the MESA algorithm by repeating the same steps as in the previous section.

A. Reparameterization

For a general SLA designated by $\Omega$, let $\Gamma \in \{0, 1\}^{M \times N}$ be the row-selection matrix satisfying that

$$y_\Omega = \Gamma y$$  \hfill (38)

for any $N \times 1$ vector $y$ and its subvector $y_\Omega$. Then, we have

$$R_\Omega = \mathbb{E}y_\Omega y_\Omega^H = \Gamma \Gamma^T$$  \hfill (39)

which is an $M \times M$ principal submatrix of $R$, defined in (15), and is reparameterized as a linear function of $(t, \sigma)$ following from (16).

B. Majorization Minimization

In this case, the objective function at the $j$th iteration becomes

$$\text{tr} \left( R_{\Omega, j-1}^{-1} R_\Omega \right) + \text{tr} \left( R_{\Omega, j-1}^{-1} \tilde{R}_\Omega \right)$$

= $\text{tr} \left( \Gamma^T R_{\Omega, j-1}^{-1} \Gamma R \right) + \text{tr} \left( R_{\Omega, j-1}^{-1} \tilde{Y}_\Omega \tilde{Y}_\Omega^H \right)$

= $\text{tr} (WR) + \text{tr} \left( \tilde{Y}_\Omega^H R_{\Omega, j-1}^{-1} \tilde{Y}_\Omega \right)$ \hfill (40)

where $R_{\Omega, j-1}^{-1}$ denotes the $(j-1)$th iterate of $R_\Omega$, $W = \Gamma^T R_{\Omega, j-1}^{-1} \Gamma$, and $\tilde{Y}_\Omega$ is any matrix satisfying that $\tilde{R}_\Omega = \tilde{Y}_\Omega \tilde{Y}_\Omega^H$ and has at most min$(L, M)$ columns.

C. Problem Reformulation

Note that the difference between (40) and (20) is in the second term. Fortunately, we have the following identity as a direct consequence of [77, Lemma 6]:

$$\text{tr} \left( \tilde{Y}_\Omega^H R_{\Omega, j-1}^{-1} \tilde{Y}_\Omega \right) = \min_{\tilde{Y}_\Omega} \text{tr} \left( \tilde{Y}_\Omega^H R_{\Omega, j-1}^{-1} \tilde{Y}_\Omega \right)$$  \hfill (41)

where the set $\Omega$ denotes the complement of $\Omega$, and $\tilde{Y}_\Omega$ and $\tilde{Y}_\Omega^H$ are nonoverlapping submatrices of $\tilde{Y}$. By substituting (16) and (41) into (40), the objective function becomes

$$\text{tr}(W(Tt + \sigma I)) + \text{tr} \left( \tilde{Y}_\Omega^H (Tt + \sigma I)^{-1} \tilde{Y}_\Omega \right)$$  \hfill (42)

with respect to $(t, \sigma, \tilde{Y}_\Omega^H)$, which is exactly in the form of (20). Consequently, the same derivations as in the ULA case can be applied, yielding the following optimization problem:

$$\min_{t, \sigma, \tilde{Y}_\Omega^H, Z} \text{tr}(WTT) + \text{tr}(X) + 2\sqrt{\text{tr}(W)}||\tilde{Y} - Z||_F$$

subject to $\begin{bmatrix} X & Z^H \\ Z & Tt \end{bmatrix} \in \mathbb{S}_+^K$  \hfill (43)
or equivalently
\[
\min_{t,X,Z} \text{tr}(WTT) + \text{tr}(X) + 2\sqrt{\text{tr}(W)}\|\hat{Y}_\Omega - Z_\Omega\|_F
\]
subject to
\[
\begin{bmatrix}
X
Z
\end{bmatrix}
\begin{bmatrix}
Z^H
T
\end{bmatrix} \in S_+^K
\] (44)
by noting that the solution to \(\hat{Y}_\Omega\) is exactly \(Z_\Omega\). In this process, as in (23), we have
\[
\sigma^* = \frac{1}{\sqrt{\text{tr}(W)}}\|\hat{Y}_\Omega - Z_\Omega\|_F.
\]
(45)

D. Using ADMM

The only difference between (44) and (26) is the inclusion of the index set \(\Omega\) in the term \(\|\hat{Y} - Z\|_F\). Consequently, the only difference in the ADMM algorithm occurs in the update of \(Z\). In particular, the objective function to minimize regarding \(Z\) changes from that in (34) to the following:
\[
\sqrt{\text{tr}(W)}\|\hat{Y}_\Omega - Z_\Omega\|_F + \frac{\mu}{2} \|Q_2 + \mu^{-1}A_2 - Z_\Omega\|_F^2
\]
\[
= \sqrt{\text{tr}(W)}\|\hat{Y}_\Omega - Z_\Omega\|_F + \frac{\mu}{2} \|[Q_2 + \mu^{-1}A_2] - Z_\Omega\|_F^2
\]
\[
+ \frac{\mu}{2} \|[Q_2 + \mu^{-1}A_2]_\Omega - Z_\Omega\|_F^2.
\] (46)

Therefore, it follows from (35) that
\[
Z_\Omega \leftarrow \hat{Y}_\Omega - \left(1 - \frac{\sqrt{\text{tr}(W)}}{\mu \|L_\Omega\|_F}\right) L_\Omega
\]
(47)
\[
Z_\Omega \leftarrow [Q_2 + \mu^{-1}A_2]_\Omega
\] (48)
where \(L\) is as defined above (35).

The ADMM algorithm in this case runs (32), (33), (47), (48), (36), and (37) iteratively. Again, the overall MESA algorithm consists of the outer MM loop and the inner ADMM loop, and it is illustrated in Algorithm 1. It has a computational complexity of \(\mathcal{O}(N^3)\) per inner iteration and degenerates into MESA in the previous section in the specialized ULA case.

V. ROBUSTNESS TO SOURCE CORRELATIONS

The assumption of uncorrelated sources is crucial to derive the SML method concerned in this article, for which the MESA algorithm is proposed. In this section, we show that the SML method is robust to source correlations in the regime of less sources than sensors, which implies the robustness of MESA.

We consider the SLA case that consists of the ULA case when \(M = N\). In order to show the robustness to source correlations, we will not use the statistical assumptions A1–A3 in Section II-B. Instead, we make the following (deterministic) assumptions, where \(f^*\) and \(S^0\) denote the true values of the parameters. The array outputs are given by
\[
Y_\Omega = A_\Omega(f^*)S^0 + E.
\]
Let \(\sigma^0 = \frac{\|E\|_F}{\|L\|_F}\) that denotes the noise level of random noise \(E\).

A4: \(f^*\) and \(S^0\) are uniquely identifiable from their product \(Y^*_\Omega = A_\Omega(f^*)S^0\).

\[\text{Algorithm 1: Maximum-Likelihood Estimation via Sequential ADMM.}\]

**Input:** SLA \(\Omega\), sample covariance \(\hat{R}_\Omega\), source number \(K\).

**Output:** Estimates of frequencies \(f\), source powers \(p\), and noise power \(\sigma\).

1. Calculate \(\hat{Y}_\Omega = \hat{R}_\Omega^{-1}\).
2. Initialize \(\hat{R}_\Omega, \hat{Q}, A, t, X, \) and \(Z\), and calculate \(W = \Gamma^T\hat{R}_\Omega^{-1}\Gamma\).
3. while not converged
   4. while not converged do
      5. Conduct the updates in (32), (33), (47), (48), (36), and (37) one after one;
   6. end while
7. Conduct the update for \(\sigma\) in (45);
8. Update \(R\) in (16) and \(W = \Gamma^T\hat{R}_\Omega^{-1}\Gamma\);
9. end while
10. Calculate the estimates of \(f\) and \(p\) by computing the decomposition \(TT = A(f)PA^H(f)\) using root-MUSIC.

A5: Both \(\sigma^0\) and \(\inf_{f,S} \|Y - A(f)S\|_F^2\) are strictly positive.

A6: \(A_\Omega(f)\) has full column rank in a neighborhood of \(f^*\).

Assumption A4 seems necessary if no statistical assumptions are made on the source signals in \(S^0\). Note that A4 implies \(K < M\). Given A4 and randomness of noise \(E\), A5 is trivial since, otherwise, the random outputs \(Y_\Omega\) must lie in a \(K\)-dimensional subspace, which happens with probability zero. A6 is a technical assumption ensuring that the estimates of source powers are consistent. The following proposition is a result of combining [41, Th. 1] and [80, Lemma 1].

**Proposition 1** Assumptions A4 and A6 hold true if

\[
K < \frac{\text{Spark}(\Omega) + \text{rank}(S^0) - 1}{2}
\] (49)

where \(\text{Spark}(\Omega)\) is defined as the smallest number of atoms in \([a_\Omega(f)]\) that are linearly dependent.

It follows from Proposition 1 that the assumptions can be satisfied if the number \(K\) of sources is below a threshold that depends on the array geometry \(\Omega\) and the rank of source signal matrix \(S^0\), though it might be difficult to verify the condition in (49) in practice.

Our main result is stated in the following theorem.

**Theorem 1** Under assumptions A4–A6 and letting \((f^*, p^*, \sigma^*)\) be the solution to the nominal SML optimization problem given by

\[
\min_{\{f_1, \ldots, f_M\}, p, \sigma} \left\{ \left| \ln|A_\Omega PA_\Omega^H + \sigma I| \right| + \frac{1}{L} \text{tr} \left( (A_\Omega PA_\Omega^H + \sigma I)^{-1} \hat{Y}_\Omega \right) \right\}
\] (50)
we have that $\sigma^* > 0$ and
\[
\lim_{\sigma^* \to 0} \sigma^* = 0 \quad \text{(51)}
\]
\[
\lim_{\sigma^* \to 0} f^* = f^o \quad \text{(52)}
\]
\[
\lim_{\sigma^* \to 0} p^*_k = \frac{1}{L} \| S^o_k \|^2 \quad \text{(53)}
\]
where $S^o_k$ denotes the $k$th row of $S^o$.

**Proof** See Appendix B.

It is shown in Theorem 1 that the SML method produces consistent estimates of the DOAs and source powers as the noise vanishes regardless of (spatial and temporal) source correlations, implying its robustness to (spatially) correlated or coherent sources, at least in the high SNR regime.

**Remark 1** Theorem 1 is related to [43, Th. 2], which is concerned with the problem
\[
\min_t |TT + \epsilon I| + \text{tr} \left( Z^T [TT]^{-1} Z \right) \text{ subject to } TT \succeq 0
\]
\[
(54)
\]
where $Z = A(f^o)S^o$ is noiseless and $\epsilon > 0$ is a fixed small constant. In contrast to this, Theorem 1 is on the noisy case, in which the noise variance $\sigma$ is a variable to optimize. Another difference is that the source number is explicitly given in Theorem 1, while there is no corresponding rank constraint on $TT$ in (54). All these differences arise due to the fact that the problem in (54) was introduced in [43] as a surrogate function for the spectral sparsity of $Z$, rather than a consequence of the SML as in this article.

**Remark 2** While the focus of this article is on DOA estimation using SLAs, we note that Theorem 1 is applicable to arbitrary linear arrays, for which the sensors are not necessarily located on a regular grid. Moreover, it can easily be extended to general joint sparse recovery problems [61], [80].

**VI. NUMERICAL RESULTS**

**A. Experimental Setup**

In this section, we present numerical results to illustrate the performance of the proposed MESA algorithm for DOA estimation using SLAs. In our implementation of MESA, $R_\Omega$ is initialized with $\bar{R}_\Omega$ plus a small scalar matrix (for proper regularization when computing the inverse) in general. However, if the ratio of the $k$th greatest eigenvalue and the smallest eigenvalue of $\bar{R}_\Omega$ is smaller than a threshold (set to 5), which happens in the presence of highly correlated sources, $R_\Omega$ is initialized as an identity matrix. As for the first ADMM loop, $\bar{Y}_\Omega = \bar{R}_\Omega^{-1}$ and $\bar{R}_\Omega$ are used to initialize $Z_\Omega$ and the corresponding principal submatrix of $Q_1$, respectively. Other variables are initialized with zero. The penalty parameter $\mu$ is initialized to 1 and varied as in [26, Sec. 3.4.1] to make performance less dependent on the initial choice and to accelerate convergence. The outer MM loop is terminated if the relative change of the negative log-likelihood function at two consecutive iterations is lower than $10^{-5}$ or the number of iterations reaches 20. The ADMM iteration is terminated if the relative and absolute errors are below $10^{-5}$ and $10^{-4}$, respectively (see [26, Sec. 3.3.1] for details), or a maximum number 1000 of iterations is reached. To better understand the performance of MESA, we also present its performance with a single MM iteration, termed as MESA-1. Note that the criterion of MESA-1 (when initializing $R_\Omega$ with $\bar{R}_\Omega$) has been used in [65] and [66], where the source number or the rank constraint is relaxed.

The methods that we use for comparison include SS-MUSIC [46], [47], WLS [52], multiple-snapshot Newtonized orthogonal matching pursuit (MNOMP) [44], [81], reweighted atomic-norm minimization (RAM) [43], and maximum-likelihood estimation of low-rank Toeplitz (MELT) [60]. SS-MUSIC is a popular CBA method and is implemented with forward–backward SS and root-MUSIC. WLS is the only asymptotically efficient algorithm prior to this article when more uncorrelated sources than sensors are present. It is initialized with SS-estimation of signal parameters via rotational invariance techniques (SS-ESPRIT) following from [52]. MNOMP is a greedy algorithm for the deterministic ML and does not require a complex initialization. RAM tries to minimize the number of sources subject to data fidelity. MNOMP and RAM do not make statistical assumptions on the sources and are expected to be robust to source correlations, but they cannot localize more sources than sensors. RAM requires the noise power rather than the source number. MELT solves the same SML problem as MESA but is usable only for ULAs. We also compare with the CRB that is computed following from [24] for uncorrelated sources. In the presence of correlated and uncorrelated sources, the CRB is computed based on a textbook routine with the prior knowledge on which sources are uncorrelated.

All the sources and noise are generated by using complex Gaussian distributions. All the sources have unit powers. The SNR is defined as the ratio of the source power to noise power. The root-mean-square error (RMSE) of the frequency estimates is computed as $\sqrt{\frac{1}{K} \| \tilde{f} - f^o \|^2}$ and then averaged over 1000 Monte Carlo runs, where $\tilde{f}$ is the vector of estimated frequencies. We consider three different types of SLAs consisting of a ten-element ULA, a six-element MRA, and an eight-element nested array given, respectively, by
\[
\Omega_{ULA} = \{1, 2, \ldots, 10\} \quad \text{(55)}
\]
\[
\Omega_{MRA} = \{1, 2, 7, 10, 12, 14\} \quad \text{(56)}
\]
\[
\Omega_{Nested} = \{1, 2, 3, 4, 5, 10, 15, 20\} \quad \text{(57)}
\]

**B. Convergence and Optimality**

In this subsection, we test the numerical performance of MESA in convergence and optimality by computing the negative log-likelihood function value at each outer MM iteration. To compare with MELT [60], we consider the ULA in (55). As in [60], the frequency domain $[-\frac{1}{2}, \frac{1}{2})$ is approximated by a set of $2N - 1$ uniform gridding points, and the true frequencies are selected from the gridding...
Experiment 1

In Experiment 1, $K = 3$ sources are generated with DOAs such that the frequencies are taken as the 5th, 7th, and 18th gridding points of MELT. We set SNR = 10 dB and the number of snapshots $L = 100$. The curve of the function value with respect to the index of the MM iteration of MESA is plotted in Fig. 1. Since the iterations of MELT are quite different from those of MESA, we plot only the final function value produced by MELT, indicated by a horizontal line as the ground truth. It is seen that the function value decreases monotonically and converges in nine iterations. MESA produces a function value smaller than the ground truth and MELT. MESA takes 534 inner iterations in total.

We tried a total number of 500 Monte Carlo runs, and MESA always converges and produces a function value smaller than the ground truth. It performs better than MELT in 498 out of 500 runs.

C. Statistical Efficiency for Uncorrelated Sources

In this subsection, we use the MRA in (56) and test the statistical efficiency of MESA for uncorrelated sources. RAM and MNOMP are considered only in the case of $K < M$.

In Experiment 2, we consider $K = 7$ sources with DOAs satisfying that the frequencies are taken in $\{-0.43, -0.28, -0.21, -0.05, 0.1, 0.26, 0.42\}$. We fix the number of snapshots $L = 200$ and vary the SNR from $-10$ to $20$ dB. Our simulation results are presented in Fig. 2. It is seen that WLS and MESA attain the CRB as SNR $\geq -5$ dB, while SS-MUSIC always produces an error greater than that of the CRB. It is interesting to note that the results of MESA-1 and MESA are almost indistinguishable. In this case, the sample covariance is a good estimate of the data covariance, and a single outer loop of MESA suffices to produce an accurate estimate.

In Experiment 3, we study the performance with respect to the number of snapshots. Two scenarios are considered: 1) $K = 3 < M$ sources with frequencies in $\{-0.2, -0.1, 0.2\}$, and 2) $K = 7 > M$ sources with the same frequencies as in Experiment 2. We fix SNR = 10 dB and vary the number of snapshots $L$ from 10 to 1280. Our simulation results are presented in Fig. 3. It is seen that MESA and WLS attain the CRB in both the scenarios with a modest number of snapshots. It is shown in Fig. 3(a) that MESA performs better than WLS in the regime of limited snapshots. Comparing Fig. 3(a) and (b), it is seen that a greater number of snapshots is required when more sources are present. In contrast to MESA and WLS, SS-MUSIC and RAM always produce an error greater than the CRB, and MNOMP fails in the first scenario.

In Experiment 4, we fix SNR = 10 dB and $L = 100$ and vary $K$ from 2 to $N - 1 = 13$. The $K$ sources are generated with $f_k = -0.44 + 0.98(k - 1)/K$, $k = 1, \ldots, K$. Our simulation results are presented in Fig. 4. Again, WLS and MESA (and MESA-1) attain the CRB or even better whenever the number of sources is smaller or greater than the number of sensors, while SS-MUSIC cannot (note that the ML estimator is usually biased and can possibly be better than the CRB [82]). MNOMP and RAM always produce an error greater than the CRB, and MNOMP fails in the first scenario.

In Experiment 5, we consider $K = 3$ sources with frequencies given by $\{-0.2, 0.1, 0.1 + \delta\}$ and vary $\delta \in \{0.001, 0.003, \ldots, 0.029\}$. We fix the number of snapshots $L = 100$ and SNR = 10 dB. It is seen in Fig. 5 that MESA attains the CRB for very closely located sources and, thus, has a higher resolution than that of the other methods. MNOMP fails to resolve the closely located sources. A gap is shown between MESA and MESA-1, implying that the MM iterations of MESA are useful to improve the resolution.
D. Robustness to Source Correlations

In Experiment 6, we consider \( K = 3 \) sources with frequencies in \([-0.2, -0.1, 0.2]\), where the first two sources are coherent with the correlation coefficient \( e^{i\pi/3} \) (that can be changed to any other value on the unit circle). We use the nested array in (57) for DOA estimation to make sure that the assumptions of Theorem 1 are satisfied, which can be verified according to Proposition 1 since the source matrix \( S \) has rank 2 and \( \text{Spark}(\Omega) \geq 6 \) by noting that the nested array contains a five-element ULA. We fix \( L = 100 \) and vary the SNR from \(-10\) to \(20\) dB. Our numerical results are presented in Fig. 6. It is seen that MESA has stable performance once the SNR is above \( 0 \) dB, which is consistent with Theorem 1. Remarkably, MESA attains the CRB as in the case of uncorrelated sources. In contrast to this, SS-MUSIC and WLS do not have the same robustness as MESA. Satisfactory performance is also obtained by MNOMP and RAM. In this case, MESA performs better than MESA-1 since the solution to \( R/Omega_1 \) is significantly different from the sample covariance, and an accurate initialization is unavailable.

In Fig. 7, we present results of one Monte Carlo run of the previous experiment (at SNR = 20 dB). It is seen that MESA can accurately estimate the frequencies/DOAs and the source power, validating Theorem 1. Interestingly, SS-MUSIC and WLS can accurately localize the two coherent sources but mislocate the third source that is uncorrelated with the other two. The reason underlies this behavior needs further investigation.

In Experiment 7, we repeat Experiment 6 by changing the nested array to the MRA in (56). In this case, it is difficult
to verify the assumptions of Theorem 1. We present our results in Fig. 8. It is seen that all the algorithms are affected to a larger extent by the source correlation as compared to the nested array case presented in Fig. 6. Differently from SS-MUSIC and WLS, MESA remains to be robust to source correlations. MNOMP has a poor performance in this case with a small array size.

E. More Sources Than Sensors With Correlations

In Experiment 8, by using the proposed MESA algorithm, we make a numerical study on the case of more sources than sensors with some of them being correlated, which to the best of our knowledge has not been investigated before. We repeat Experiment 2 by fixing SNR = 10 dB and letting the first and fourth sources be correlated with a correlation coefficient \( \rho = \pi/4 \), where the modulus \( \rho \) changes from 0 (uncorrelated) to 1 (coherent). MNOMP and RAM are not usable since more sources than sensors are present. It is seen in Fig. 9 that the performance of all the methods becomes worse as the correlation increases. In contrast to a steady performance loss of MESA, a sharp loss is shown for SS-MUSIC and WLS in the regime of highly correlated sources. As \( \rho = 1 \), in fact, SS-MUSIC and WLS mislocate at least one of the sources in over 10% of the Monte Carlo runs, while MESA always accurately localize the sources.

To sum up, we have shown by numerical results that MESA can statistically efficiently localize more uncorrelated sources than sensors and has robust performance in the presence of correlated and coherent sources. This makes it unique among the existing coarray-based methods tailored for uncorrelated sources and sparse methods usable for a small number of sources. MESA-1 is a good accelerated approximation of MESA in general, while the latter has improved resolution and robustness. More simulation results can be found in our conference papers [1], [2].

VII. CONCLUSION

In this article, we proposed one algorithm that can deal with two challenging scenarios of DOA estimation: 1) more uncorrelated sources than sensors and 2) highly correlated or correlated sources. This was realized by studying the robustness property of the ML method derived under the assumption of uncorrelated sources and proposing the MESA algorithm for the ML method based on elegant problem reformulations. Extensive numerical results were provided that validate our theoretical findings and demonstrate superior performance of MESA in terms of statistical efficiency, resolution, and robustness to highly correlated sources as compared to state-of-the-art algorithms.

It was shown in this article that MESA can localize more sources than sensors even in the presence of highly correlated or coherent sources. A theoretical understanding of this behavior, together with the parameter identifiability, should be investigated in future. Moreover, both the algorithm-dependent and algorithm-independent analyses are of interest to investigate how the number of localizable sources leverages with source correlations. For a particular algorithm, it is shown that using the assumption of uncorrelated sources does not necessarily contradict with its robustness to correlated sources. Therefore, it is of great interest to investigate their robustness for both the existing and future algorithms proposed with the uncorrelated setup. It is also shown that the array geometry is another factor affecting
the robustness to source correlations. It is interesting to take
the robustness into consideration for future array geometry
design.

APPENDIX

A. Proof of (35)

In the case when \( L = 0 \), we have \( \hat{Y} = Q_2 + \mu^{-1}A_2 \), and
thus, the optimization problem in (34) becomes

\[
\min_{\hat{Z}} \sqrt{\text{tr}(W)\|\hat{Y} - Z\|_F + \frac{\mu}{2}\|\hat{Y} - Z\|_F^2}.
\]

(58)

Evidently, the optimal solution to \( \hat{Z} \) is given by \( \hat{Y} \).

In the case when \( L \neq 0 \), to show (35), it suffices to show that

\[
\left(1 - \frac{\beta}{\|L\|_F}\right) L = \arg \min_{\hat{Z}} |\hat{Z}\|_F + \frac{1}{2}\|\hat{Z} - L\|_F^2
\]

(59)

by identifying that \( \hat{Z} = \hat{Y} - Z \) and \( \beta = \frac{\sqrt{\text{tr}(W)}}{\mu} \). To this end, observe that

\[
g(\hat{Z}) \triangleq \beta\|\hat{Z}\|_F^2 + \frac{1}{2}\|\hat{Z} - L\|_F^2 \\
\geq \beta\|\hat{Z}\|_F^2 + \frac{1}{2}\left(\|\hat{Z}\|_F^2 - \|L\|_F^2\right) \\
\geq \frac{1}{2}\left(\|L\|_F^2 - \beta\|L\|_F + \beta^2\right) + \beta\|L\|_F^2 - \beta^2
\]

(60)

where the equality is achieved if \( \hat{Z} \) has the sign of \( L \), with
\( \text{sgn}(L) \triangleq \frac{L}{\|L\|_F} \). Since the last expression is minimized if
\( \|\hat{Z}\|_F = (\|L\|_F - \beta)_+ \), the overall function \( g(\hat{Z}) \) is, therefore,
minimized at

\[
\hat{Z} = (\|L\|_F - \beta)_+ \text{sgn}(L) = \left(1 - \frac{\beta}{\|L\|_F}\right) L
\]

(61)

completing the proof.

B. Proof of Theorem 1

We first show the following lemma.

**LEMMA 1** Assume that \( A = [a_1, \ldots, a_K] \) and \( P \) is positive-
definite diagonal. Then, for any \( \sigma > 0 \) and \( k = 1, \ldots, K \), we have

\[
a_k^H (A^H + \sigma I)^{-1} a_k < p_k^{-1}.
\]

(62)

If, further, \( A \) has full column rank, then

\[
\lim_{\sigma \to 0} a_k^H (A^H + \sigma I)^{-1} a_k = p_k^{-1}.
\]

(63)

**PROOF** Note that the matrix

\[
\begin{bmatrix}
p_k^{-1} & a_k^H \\
a_k & A^H + \sigma I
\end{bmatrix}
\]

is positive definite since so is \( p_k^{-1} \) and its Schur complement

\[
(A^H + \sigma I) - p_k a_k a_k^H \geq \sigma I > 0.
\]

(64)

Consequently, the Schur complement regarding \( A^H + \sigma I \) is positive, yielding (62).

Without loss of generality, we next show (63) for \( k = 1 \). For any \( \epsilon > 0 \), \( \text{diag}(-\epsilon, p_2, \ldots, p_K) \) is indefinite and so is \( \text{Adiag}(-\epsilon, p_2, \ldots, p_K)A^H \) since by assumption \( A \) has full
column rank. It follows that for any

\[
0 < \sigma < -\lambda_{\min}(\text{Adiag}(-\epsilon, p_2, \ldots, p_K)A^H)
\]

(65)

the matrix

\[
A^H + \sigma I - (p_1 + \epsilon) a_1^H a_1
\]

\[
= \left[\text{Adiag}(-\epsilon, p_2, \ldots, p_K)A^H\right] + \sigma I
\]

is indefinite. Hence, the matrix

\[
(p_1 + \epsilon)^{-1} a_1^H a_1 (A^H + \sigma I)^{-1} a_1 < 0
\]

(67)

which combined with (62) yields

\[
(p_1 + \epsilon)^{-1} < a_1^H (A^H + \sigma I)^{-1} a_1 < p_1^{-1}
\]

(68)

of which a direct consequence is (63).

\[\Box\]

We are ready to prove Theorem 1. For notational simplicity, we omit the subscript \( \Omega \) in \( Y_\Omega \) and \( A_\Omega \) hereafter
and write them as \( Y \) and \( A \) without ambiguity. It follows from [77, Lemmas 4 and 5] that

\[
\text{tr}(Y^H (A^H + \sigma I)^{-1} Y) = \min_{\hat{Z}} \text{tr}(Z^H [A^H + \sigma I]^{-1} Z) + \sigma^{-1}\|Y - Z\|_F^2
\]

(69)

which can also be shown directly. Define

\[
\mathcal{L}(f, p, \sigma, S) = \ln |A^H + \sigma I| + \frac{1}{L\sigma} \text{tr}(S^H P^{-1} S)
\]

(70)

It follows that

\[
(f^*, p^*, \sigma^*, S^*) = \arg \min_{f, p, \sigma, S} \mathcal{L}(f, p, \sigma, S)
\]

(71)

and we let \( \mathcal{L}^* \) denote the optimal value.

We first show \( \sigma^* > 0 \). It suffices to show that

\[
\lim_{\sigma \to 0} \mathcal{L}(f, p, \sigma, S) = +\infty
\]

(72)

for any \( (f, p, S) \). To do so, note by (70) that

\[
\mathcal{L}(f, p, \sigma, S) \geq \ln |\sigma I| + \frac{1}{L\sigma} \|Y - AS\|_F^2
\]

(73)

and the only stationary point of the lower bound above regarding \( \sigma \), which is the global minimizer, is given by
\( \sigma = \frac{\|Y - AS\|_F^2}{\text{tr}(S^H P^{-1} S)} \) that is bounded from below by a positive
number for any \( (f, S) \) by Assumption A5. Consequently, the lower bound above always approaches infinity as \( \sigma \to 0 \),
resulting in (72).
Inserting the ground truth \((f^*, p^*, \sigma^*, S^*)\) into \(L\) and conditioning on \(\sigma^o \leq 1\), we have that
\[
L^* \leq L(f^o, S^o, p^o, \sigma^o)
= \ln |A^oP^oA^{oH} + \sigma^o I| + \frac{1}{L} \text{tr} (S^{oH} P^{o-1} S^o)
+ \frac{1}{L \sigma^o} \|Y - A^o S^o\|^2_F
\leq (M - K) \ln \sigma^o + \sum_{k=1}^K \ln (\lambda_k (A^oP^oA^{oH}) + 1)
+ K + M,
\] (74)
where \(\lambda_k\) denotes the \(k\)th greatest eigenvalue. Combining (74) and the inequality
\[
L^* \geq \ln |A^* P^* A^{*H} + \sigma^* I| \geq M \ln \sigma^*
\] (75)
yields
\[
\sigma^* \leq C \sigma^0 \frac{M - K}{M - K}
\] (76)
where \(C = e^{1+K/M} \prod_{k=1}^K (\sigma_k (A^oP^oA^{oH}) + 1)^{1/M} \) is a constant. A direct consequence of (76) is (51).

Inserting the stationary point \(\sigma = \frac{[Y - AS]}{NL}\) into the lower bound in (73) yields
\[
L(f, p, \sigma, S) \geq M \ln \sigma + \frac{1}{L \sigma} \|Y - AS\|^2_F
\geq M \ln \frac{\|Y - AS\|^2_F}{NL} + M.
\] (77)
Consequently
\[
L^* = L(f^*, p^*, \sigma^*, S^*) \geq M \ln \frac{\|Y - A^* S^*\|^2_F}{NL} + M.
\] (78)
Combining (78) and (74), we obtain
\[
\frac{\|Y - A^* S^*\|^2_F}{NL} \leq C e^{-1} \sigma^0 \frac{M - K}{M - K}.
\] (79)
Therefore
\[
\lim_{\sigma^* \to 0} A^* S^* = A^o S^o
\] (80)
implies the consistency of \((f^*, S^*)\) (up to permutations of entries) by Assumption A4.

Finally, note by (70) that
\[
p^* = \arg \min_p \ln |A^* P A^{*H} + \sigma^* I| + \frac{1}{L} \text{tr} (S^{*H} P^{-1} S^*)
\] (81)
where all the rows of \(S^*\) are nonzero and \(A^*\) has full column rank if \(\sigma^o\) is small enough due to their consistency as \(\sigma^o \to 0\) and Assumption A6. Note that the function value in (81) approaches infinity at any boundary point with \(p_k = 0\) by the fact that all the rows of \(S^*\) are nonzero. Consequently, the derivative of the above objective function with respect to \(p_k\) vanishes at \(p_k = p_k^*\), yielding
\[
a^H (f_k^*) (A^* P^* A^{*H} + \sigma^* I)^{-1} a (f_k^*) - \frac{\|S_k\|^2}{L p_k^2} = 0.
\] (82)
Applying Lemma 1, we then obtain \(\frac{\|S_k\|^2}{\sigma_k\pi} \leq \frac{1}{L} p_k^2\) and thus
\[
p_k^* > \frac{\|S_k\|^2}{L p_k^2},
\] (83)
is bounded from below by a universal positive number if \(\sigma^o\) is small enough. We further apply the second part of Lemma 1 to obtain by (51) and (82) that
\[
0 = \lim_{\sigma^* \to 0} a^H (f_k^*) (A^* P^* A^{*H} + \sigma^* I)^{-1} a (f_k^*) - \frac{\|S_k\|^2}{L p_k^2}
\]
(84)
which results in (53) and completes the proof.

Finally, it is worth noting that if multiple optimal solutions exist, then our proof and Theorem 1 hold for any of them.

ACKNOWLEDGMENT
The authors would like to thank Prof. Jiang Zhu for providing their code of MNOMP. They also thank Prof. Yimin D. Zhang, Prof. Wei Liu, and the anonymous reviewers for their valuable comments that have improved the quality of this article.

REFERENCES
[1] Z. Yang and X. Chen, “Maximum likelihood direction-of-arrival estimation via rank-constrained ADMM,” in Proc. CIE Int. Conf. Radar, 2021, pp. 2376–2380.
[2] X. Chen and Z. Yang, “Localizing more sources than sensors in presence of coherent sources,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2022, pp. 5013–5017.
[3] P. Stoica and R. L. Moses, Spectral Analysis of Signals. Upper Saddle River, NJ, USA: Pearson/Prentice-Hall, 2005.
[4] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, “Estimation of sparse MIMO channels with common support,” IEEE Trans. Commun., vol. 60, no. 12, pp. 3705–3716, Dec. 2012.
[5] J. Li and P. Stoica, “MIMO radar with colocated antennas,” IEEE Signal Process. Mag., vol. 24, no. 5, pp. 106–114, Sep. 2007.
[6] W. Heylen and P. Sas, Modal Analysis Theory and Testing. Leuven, Belgium:Katholieke Univ. Leuven, 2006.
[7] M. J. Rust, M. Bates, and X. Zhuang, “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),” Nature Methods, vol. 3, no. 10, pp. 793–796, 2006.
[8] Y. Wang, J. Li, and P. Stoica, “Spectral analysis of signals: The missing data case,” Synth. Lectures Signal Process. See., vol. 1, no. 1, pp. 1–102, 2006.
[9] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[10] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
[11] A. Moffet, “Minimum-redundancy linear arrays,” IEEE Trans. Antennas Propag., vol. AP-16, no. 2, pp. 172–175, Mar. 1968.
[12] P. Pal and P. Vaidyanathan, “Nested arrays: A novel approach to array processing with enhanced degrees of freedom,” IEEE Trans. Signal Process., vol. 58, no. 8, pp. 4167–4181, Aug. 2010.
[13] C.-L. Liu and P. Vaidyanathan, “Super nested arrays: Linear sparse arrays with reduced mutual coupling—Part I: Fundamentals,” IEEE Trans. Signal Process., vol. 64, no. 15, pp. 3997–4012, Aug. 2016.
[14] J. Shi, F. Wen, Y. Liu, Z. Liu, and P. Hu, “Enhanced and generalized coprime array for direction of arrival estimation,” *IEEE Trans. Aero. Electron. Syst.*, vol. 59, no. 2, pp. 1327–1339, Apr. 2023.

[15] P. P. Vaidyanathan and P. Pal, “Sparse sensing with co-prime samplers and arrays,” *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[16] S. Qin, Y. D. Zhang, and M. G. Amin, “Generalized coprime array configurations for direction-of-arrival estimation,” *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1377–1390, Mar. 2015.

[17] T.-J. Shan, M. Wax, and T. Kailath, “On spatial smoothing for direction-of-arrival estimation of coherent signals,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 806–811, Aug. 1985.

[18] S. U. Pillai and B. H. Kwon, “Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 8, pp. 1176–1189, Aug. 1989.

[19] K.-C. Tan and G.-L. Oh, “Estimating directions-of-arrival of coherent signals in unknown correlated noise via spatial smoothing,” *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 1087–1091, Apr. 1997.

[20] F. Liu, J. Wang, C. Sun, and R. Du, “Spatial differencing method for DOA estimation under the coexistence of both uncorrelated and coherent signals,” *IEEE Trans. Antennas Propag.*, vol. 60, no. 4, pp. 2052–2062, Apr. 2012.

[21] H. Tao, J. Xin, J. Wang, N. Zheng, and A. Sano, “Two-dimensional direction estimation for a mixture of noncoherent and coherent signals,” *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 318–333, Jan. 2015.

[22] S. Qin, Y. D. Zhang, and M. G. Amin, “DOA estimation of mixed coherent and uncorrelated targets exploiting coprime MIMO radar,” *Digit. Signal Process.*, vol. 61, pp. 26–34, 2017.

[23] M. Wang and A. Nehorai, “Cooarrays, MUSIC, and the Cramér-Rao bound,” *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb. 2016.

[24] C.-L. Liu and P. Vaidyanathan, “Cramér-Rao bounds for coprime and other sparse arrays, which find more sources than sensors,” *Digit. Signal Process.*, vol. 61, pp. 43–61, 2017.

[25] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[26] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[27] D. Diamond, R. Takapoui, and S. Boyd, “A general system for heuristic minimization of convex functions over non-convex sets,” *Optim. Methods Softw.*, vol. 33, no. 1, pp. 165–193, 2018.

[28] P. Stoica and A. Nehorai, “Performance study of conditional and unconditional direction-of-arrival estimation,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[29] P. Stoica, E. G. Larsson, and A. B. Gershman, “The stochastic CRB for array processing: A textbook derivation,” *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.

[30] M. Feder and E. Weinstein, “Parameter estimation of superimposed signals using the EM algorithm,” *IEEE Trans. Acoust., Speech Signal Process.*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[31] D. Storer and A. Nehorai, “Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise,” *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1528–1534, Jun. 1992.

[32] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[33] P. Stoica and K. C. Sharman, “Maximum likelihood methods for direction-of-arrival estimation,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.

[34] M. Viberg, B. Ottersten, and T. Kailath, “Detection and estimation in sensor arrays using weighted subspace fitting,” *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2436–2449, Nov. 1991.
Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer, “Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays. I. Fully augmentable arrays,” IEEE Trans. Signal Process., vol. 46, no. 9, pp. 2458–2471, Sep. 1998.

H. Li, P. Stoica, and J. Li, “Computationally efficient maximum likelihood estimation of structured covariance matrices,” IEEE Trans. Signal Process., vol. 47, no. 5, pp. 1314–1323, May 1999.

D. Romero, D. D. Ariyananda, Z. Tian, and G. Leus, “Compressive covariance sensing: Structure-based compressive sensing beyond sparsity,” IEEE Signal Process. Mag., vol. 33, no. 1, pp. 78–93, Jan. 2016.

B. Kang, V. Monga, and M. Rangaswamy, “Computationally efficient Toeplitz approximation of structured covariance under a rank constraint,” IEEE Trans. Aerosp. Electron. Syst., vol. 51, no. 1, pp. 775–785, Jan. 2015.

P. Babu, “MELT—Maximum-likelihood estimation of low-rank Toeplitz covariance matrix,” IEEE Signal Process. Lett., vol. 23, no. 11, pp. 1587–1591, Nov. 2016.

D. P. Wipf and B. D. Rao, “An empirical Bayesian strategy for solving the simultaneous sparse approximation problem,” IEEE Trans. Signal Process., vol. 55, no. 7, pp. 3704–3716, Jul. 2007.

Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, “An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning,” IEEE Trans. Wireless Commun., vol. 11, no. 10, pp. 1–11, Oct. 2012.

A. Das and T. J. Sejnowski, “Narrowband and wideband off-grid direction-of-arrival estimation via sparse Bayesian learning,” IEEE J. Ocean. Eng., vol. 43, no. 1, pp. 108–118, Jan. 2017.

B. Ottersten, P. Stoica, and R. Roy, “Covariance matching estimation techniques for array signal processing applications,” Digit. Signal Process., vol. 8, no. 3, pp. 185–210, 1998.

P. Stoica, P. Babu, and J. Li, “SPICE: A sparse covariance-based estimation method for array processing,” IEEE Trans. Signal Process., vol. 59, no. 2, pp. 629–638, Feb. 2011.

Z. Yang, L. Xie, and C. Zhang, “A discretization-free sparse and parametric approach for linear array signal processing,” IEEE Trans. Signal Process., vol. 62, no. 19, pp. 4959–4973, Oct. 2014.

H. Qiao and P. Pal, “Gridless line spectrum estimation and low-rank Toeplitz matrix compression using structured samplers: A regularization-free approach,” IEEE Trans. Signal Process., vol. 65, no. 9, pp. 2221–2236, May 2017.

X. Wu, W.-P. Zhu, and J. Yan, “A Toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation,” IEEE Trans. Veh. Technol., vol. 66, no. 9, pp. 8223–8237, Sep. 2017.

R. R. Poti and B. D. Rao, “Robustness of sparse Bayesian learning in correlated environments,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2020, pp. 9100–9104.

Y. Xu, W. Yin, Z. Wen, and Y. Zhang, “An alternating direction algorithm for matrix completion with nonnegative factors,” Front. Math. China, vol. 7, no. 2, pp. 365–384, 2012.

B. Jiang, S. Ma, and S. Zhang, “Alternating direction method of multipliers for real and complex polynomial optimization models,” Optimization, vol. 63, no. 6, pp. 883–908, 2014.

M. Hong, Z.-Q. Luo, and M. Razaviyayn, “Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems,” SIAM J. Optim., vol. 26, no. 1, pp. 337–364, 2016.

F. Wang, W. Cao, and Z. Xu, “Convergence of multi-block Bregman ADMM for nonconvex composite problems,” Sci. China Inf. Sci., vol. 61, no. 12, pp. 1–12, 2018.

Y. Wang, W. Yin, and J. Zeng, “Global convergence of ADMM in nonconvex nonsmooth optimization,” J. Sci. Comput., vol. 78, no. 1, pp. 29–63, 2019.

A. Barabell, “Improving the resolution performance of eigenstructure-based direction-finding algorithms,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1983, vol. 8, pp. 336–339.

M. Fazel, H. Hindi, and S. P. Boyd, “Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices,” in Proc. Amer. Control Conf., 2003, vol. 3, pp. 2156–2162.

Z. Yang and L. Xie, “On gridless sparse methods for line spectral estimation from complete and incomplete data,” IEEE Trans. Signal Process., vol. 63, no. 12, pp. 3139–3153, Jun. 2015.

K. B. Petersen and M. S. Pedersen, The Matrix Cookbook, vol. 7, Kongens Lyngby, Denmark: Tech. Univ. Denmark, 2008.

A. Dax, “Low-rank positive approximants of symmetric matrices,” Adv. Linear Algebra Matrix Theory, vol. 4, no. 3, 2014, Art. no. 50062.

M. E. Davies and Y. C. Eldar, “Rank awareness in joint sparse recovery,” IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.

B. Mamandipoor, D. Ramasamy, and U. Madhow, “Newtonized orthogonal matching pursuit: Frequency estimation over the continuum,” IEEE Trans. Signal Process., vol. 64, no. 19, pp. 5066–5081, Oct. 2016.

K. Mardia, H. Southworth, and C. Taylor, “On bias in maximum likelihood estimators,” J. Statist. Plan. Inference, vol. 76, nos. 1/2, pp. 31–39, 1999.