EM-Type Algorithms for DOA Estimation in Unknown Nonuniform Noise

Ming-yan Gong∗ and Bin Lyu†

Abstract: The expectation–maximization (EM) algorithm updates all of the parameter estimates simultaneously, which is not applicable to direction of arrival (DOA) estimation in unknown nonuniform noise. In this work, we present several efficient EM-type algorithms, which update the parameter estimates sequentially, for solving both the deterministic and stochastic maximum–likelihood (ML) direction finding problems in unknown nonuniform noise. Specifically, we design a generalized EM (GEM) algorithm and a space-alternating generalized EM (SAGE) algorithm for computing the deterministic ML estimator. Simulation results show that the SAGE algorithm outperforms the GEM algorithm. Moreover, we design two SAGE algorithms for computing the stochastic ML estimator, in which the first updates the DOA estimates simultaneously while the second updates the DOA estimates sequentially. Simulation results show that the second SAGE algorithm outperforms the first one.

Keywords: DOA estimation; EM algorithm; Maximum likelihood estimation; Nonuniform noise

1 Introduction

Estimating the directions of arrival (DOAs) of narrowband far-field sources using sensor arrays is an important task in array signal processing and several types of estimation methods have been proposed in the literature [1], [2]. In particular, the classic maximum–likelihood (ML) method plays an essential role [3]–[5]. However, ML direction finding problems are non-convex and we are hard to derive their solutions in closed form.

For obtaining ML estimates in DOA estimation efficiently, the expectation–maximization (EM) algorithm in [6] has been adopted to solve ML direction finding problems [7], [8]. Further, the space-alternating generalized EM (SAGE) algorithm proposed in [9] has been also applied to DOA estimation in order to speed up the convergence of the EM algorithm [10]–[12]. However, these EM and SAGE algorithms are usually derived under the known or unknown uniform noise model, i.e., the sensor noise variances must be equal, which may be impractical in certain applications. Hence, the two algorithms should be developed for DOA estimation in the presence of other noise models.

∗M. Gong is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: jinyan00@outlook.com).
†B. Lyu is with the Key Laboratory of Ministry of Education in Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: blyu@njupt.edu.cn).
As a general case of the uniform noise model, nonuniform noise has attracted increasing attention. Nonuniform noise is with an arbitrary diagonal covariance matrix, i.e., the sensor noise variances can be unequal. Obviously, classical eigenstructure based subspace methods, e.g., the MUSIC algorithm, cannot be directly employed for DOA estimation in unknown nonuniform noise. For applying subspace methods to DOA estimation in nonuniform noise, a subspace separation approach is proposed in [13] while in [14], the noise covariance matrix is first estimated and the sensor data are then prewhitened. Moreover, the signal subspace can be determined by maximizing the stochastic log-likelihood function (LLF) or solving a least-squares problem [15]. In [16], the signal and noise subspaces are separated by means of the eigendecomposition of a reduced array covariance matrix when the sources are uncorrelated. The authors in [17] utilize the reduced array covariance matrix in [16] and propose a non-iterative two-phase subspace-based DOA estimation method. By analyzing the array covariance matrix in nonuniform noise, an optimization problem based on the signal subspace is formulated in [18], which leads to a new DOA estimator.

The ML method can enjoy excellent statistical properties [4], [5]. But, the analyses in [19] and [20] suggest that both the deterministic and stochastic ML direction finding problems in unknown nonuniform noise cannot be reduced to two problems with respect to only the DOAs because of the noise parameters and generally involve high-dimensional search algorithms. For efficiently computing both the deterministic and stochastic ML estimators, two alternating maximization (AM) algorithms are presented in [19] and [20], respectively. Unfortunately, the two AM algorithms include high-dimensional numerical search at every iteration and are thus computationally intensive, which motivates this work.

In this work, we first try to apply the EM algorithm, each iteration of which consists of an expectation step (E-step) and a maximization step (M-step). At the M-step, however, the EM algorithm updates all of the parameter estimates simultaneously, which requires high-dimensional numerical search due to the unknown noise parameters. Hence, we present several efficient EM-type algorithms, which update the parameter estimates sequentially and only require low-dimensional numerical search at every iteration, for solving both the deterministic and stochastic ML direction finding problems in unknown nonuniform noise. Specifically, we design a generalized EM (GEM) algorithm based on the expectation-conditional maximization (ECM) algorithm [21] and an SAGE algorithm, which updates the DOA estimates sequentially, for computing the deterministic ML estimator. Simulation results show that the SAGE algorithm outperforms the GEM algorithm, i.e., the SAGE algorithm converges faster and can avoid the convergence to an unwanted stationary point of the LLF more efficiently. Moreover, we design two SAGE algorithms for computing the stochastic ML estimator, in which the first updates the DOA estimates simultaneously while the second updates the DOA estimates sequentially. Simulation results show that the second SAGE algorithm outperforms the first one.

Notations: $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose, respectively. $0 = [0 \cdots 0]^T$ and $1 = [1 \cdots 1]^T$. $\|v\|$ and $[v]_i$ denote the Euclidean norm and $i$th element of a vector $v$, respectively. $M^{-1}$, Det$(M)$, Tr$(M)$, and $[M]_{i,i}$ denote the inversion, determinant, trace, $i$th diagonal element of a square matrix $M$, respectively. $I_N$ and $0_N$ are the $N \times N$ identity and zero matrices, respectively. $M \geq 0_N$ and $M > 0_N$ denote that the $N \times N$ square matrix $M$ is positive semi-definite and definite, respectively. $E\{\cdot\}$ and $\mathbb{D}\{\cdot\}$ denote expectation and covariance, respectively. $j$ is the imaginary unit.
2 Data Model and Problem Formulation

Let us consider an array of $N$ sensors, which receives the signals transmitted by $M$ narrowband far-field sources with the same known center wavelength $\lambda$. For simplicity, the array is assumed to be a uniform linear array of inter-sensor spacing $\lambda/2$. The DOA of the $m$th source is denoted by $\theta_m \in (0, \pi)$. Then, the signal vector received at this array can be written as [1], [2]

$$v(t) = \sum_{m=1}^{M} d(\theta_m)f_m(t) + z(t) = D(\theta)f(t) + z(t),$$

where

1. $d(\theta_m) = [1 \ e^{-j \pi \cos(\theta_m)} \ \ldots \ e^{-j(N-1)\pi \cos(\theta_m)}]^T$ is the steering vector of the $m$th source,
2. $f_m(t)$ is the $m$th source signal received at the 1st sensor and with power $P_m$,
3. $z(t) \sim \mathcal{C}\mathcal{N}(0, \Sigma)$ is a white complex Gaussian noise vector with covariance matrix $\Sigma = \text{diag}\{\sigma\}$ and $\sigma = [\sigma_1 \ \ldots \ \sigma_N]^T > 0$, i.e., the nonuniform noise model.

When $\sigma_1 = \ldots = \sigma_N = \sigma$, $\Sigma = \sigma I_N$ and the nonuniform noise model reduces to the uniform noise model. In (1), $\theta = [\theta_1 \ \ldots \ \theta_M]^T \in \Psi$ with $\Psi = (0, \pi)^M$, $D(\theta) = [d(\theta_1) \ \ldots \ d(\theta_M)]$, and $f(t) = [f_1(t) \ \ldots \ f_M(t)]^T$.

In the EM algorithm, the complete data need to be defined. We thus sample the received signal and design the samples (“snapshots”) or incomplete data by [7], [11]

$$v(t) = \sum_{m=1}^{M} [d(\theta_m)f_m(t) + z_m(t)] = \sum_{m=1}^{M} g_m(t), \quad t = 1, 2, \ldots, T,$$

where

1. $T$ is the total number of samples,
2. the $g_m(t)$’s are the underlying complete data,
3. the $z_m(t)$’s are mutually independent noise vectors and $z_m(t) \sim \mathcal{C}\mathcal{N}(0, \Sigma_m)$ with $\Sigma_m = \text{diag}\{\sigma_m\}$, $\sigma_m = [\sigma_{1,m} \ \ldots \ \sigma_{N,m}]^T > 0$, and $\Sigma = \sum_{m=1}^{M} \Sigma_m$.

When the noise is uniform, we set $\sigma_{1,m} = \ldots = \sigma_{N,m} = \delta_m$, i.e., $\Sigma_m = \delta_m I_N$ and $\sigma = \sum_{m=1}^{M} \delta_m$ [22]. Since both the incomplete- and complete-data LLFs must determine the statistical model of the $f_m(t)$’s, we consider the so-called deterministic and stochastic signal models separately.

2.1 Deterministic Signal Model

In the deterministic signal model [7], [8], [19], the $f_m(t)$’s are deterministic and unknown, so we have $g_m(t) \sim \mathcal{C}\mathcal{N}(d(\theta_m)f_m(t), \Sigma_m)$ and $v(t) \sim \mathcal{C}\mathcal{N}(D(\theta)f(t), \Sigma)$. Then, both the incomplete- and complete-data LLFs are, respectively, expressed by

$$L(\Phi, \sigma) = \sum_{t=1}^{T} \ln p(v(t); \theta, f(t), \sigma) = -TN \ln(\pi) - T \sum_{n=1}^{N} \ln(\sigma_n) - \sum_{t=1}^{T} [v(t) - D(\theta)f(t)]^H \Sigma^{-1} [v(t) - D(\theta)f(t)],$$

(3a)
\[ U(\Phi, \Omega) = \sum_{t=1}^{T} \sum_{m=1}^{M} \ln p(\mathbf{g}_m(t); \theta_m, \mathbf{f}_m(t), \sigma_m) = -TMN \ln(\pi) - T \sum_{m=1}^{M} \frac{\sum_{n=1}^{N} \ln(\sigma_{n,m})}{\sum_{t=1}^{T} \sum_{m=1}^{M} [\mathbf{g}_m(t) - \mathbf{d}(\theta_m)f_m(t)]' \Sigma^{-1}_m [\mathbf{g}_m(t) - \mathbf{d}(\theta_m)f_m(t)],} \]

where \( \mathbf{F} = [\mathbf{f}(1) \cdots \mathbf{f}(T)], \ \Phi = (\theta, \mathbf{F}) \) and \( \Omega = (\sigma_1, \ldots, \sigma_M) \) denote the signal and noise parameters, respectively. Based on (3a), the deterministic ML direction finding problem is constructed as

\[ \max_{\theta \in \Psi, \mathbf{F}, \sigma \geq 0} \mathcal{L}(\Phi, \sigma). \] (4)

### 2.2 Stochastic Signal Model

In the stochastic signal model, the source signals are modelled as \( f_m(t) \sim \mathcal{CN}(0, P_m) \) where \( P_m \) is the power of the \( m \)th source. For simplicity, we assume that the sources are independent of the noise and all of the \( f_m(t) \)'s are mutually independent \cite{7}, \cite{8}. Moreover, let \( \Sigma_m = \alpha_m \Sigma \) with \( \alpha = [\alpha_1 \cdots \alpha_M]' > 0 \) and \( 1^T \alpha = 1 \) known. Then, we have \( \mathbf{g}_m(t) \sim \mathcal{CN}(0, \mathbf{H}_m) \) and \( \mathbf{v}(t) \sim \mathcal{CN}(0, \mathbf{V}) \) where \( \mathbf{H}_m = P_m \mathbf{d}(\theta_m) \mathbf{d}'(\theta_m) + \alpha_m \Sigma > 0_N \) and \( \mathbf{V} = \sum_{m=1}^{M} \mathbf{H}_m = \sum_{m=1}^{M} P_m \mathbf{d}(\theta_m) \mathbf{d}'(\theta_m) + \Sigma > 0_N \).

Both the incomplete- and complete-data LLFs are, respectively, expressed by

\[ \mathcal{L}(\Phi, \sigma) = \sum_{t=1}^{T} \ln p(\mathbf{v}(t); \theta, \mathbf{P}, \sigma) = -T \left[ N \ln(\pi) + \ln \left( \det(\mathbf{H}_v) \right) + \text{Tr} \left( \mathbf{H}_v^{-1} \hat{\mathbf{R}}_v \right) \right], \] (5a)

\[ U(\Phi, \sigma) = \sum_{t=1}^{T} \sum_{m=1}^{M} \ln p(\mathbf{g}_m(t); \theta_m, \mathbf{P}_m, \sigma) = -T \sum_{m=1}^{M} \left[ N \ln(\pi) + \ln \left( \det(\mathbf{H}_m) \right) + \text{Tr} \left( \mathbf{H}_m^{-1} \hat{\mathbf{R}}_m \right) \right], \] (5b)

where \( \mathbf{P} = [P_1 \cdots P_M]' \), \( \Phi = (\theta, \mathbf{P}) \), \( \hat{\mathbf{R}}_v = (1/T) \sum_{t=1}^{T} \mathbf{v}(t) \mathbf{v}'(t) \), and \( \hat{\mathbf{R}}_m = (1/T) \sum_{t=1}^{T} \mathbf{g}_m(t) \mathbf{g}_m'(t) \).

Based on (5a), the stochastic ML direction finding problem is constructed as

\[ \max_{\theta \in \Psi, \mathbf{P} \geq 0, \sigma > 0} \mathcal{L}(\Phi, \sigma). \] (6)

**Remark 1** Although problems (4) and (6) can be further reduced to some problems with respect to fewer unknown parameters \cite{19}, \cite{20}, these problems are still non-convex and high-dimensional, i.e., applying conventional gradient-type search algorithms to solve these problems is computationally expensive. For efficiently computing both the deterministic and stochastic ML estimators of \( \theta \) in (4) and (6), we design some appropriate EM-type algorithms in the next two sections.

### 3 Deterministic Signal Model

In this section, we present a GEM algorithm and an SAGE algorithm for solving problem (4). For convenience, let \((\cdot)^{(b)}\) denote an iterative value at the \( b \)th iteration and \((\cdot)^{(0)}\) is an initial estimate.

#### 3.1 GEM Algorithm

We first try to apply the EM algorithm and the E- and M-steps at the \( b \)th iteration are introduced below
3.1.1 E-step

The EM algorithm calculates the conditional expectation of the complete-data LLF in (3b) [7], i.e.,

\[
\mathbb{E}\{\mathcal{U}(\Phi, \Omega) \mid V; \Phi^{(b-1)}, \Omega^{(b-1)}\} = -TMN \ln(\pi) - T \sum_{m=1}^{M} \sum_{n=1}^{N} \ln(\sigma_{n,m}) - \sum_{t=1}^{T} \sum_{m=1}^{M} \left( \text{Tr}(\Sigma_m^{-1} G_m^{(b)}) + [g_m^{(b)}(t) - d(\theta_m)f_m(t)]^H \Sigma_m^{-1} [g_m^{(b)}(t) - d(\theta_m)f_m(t)] \right),
\]

where \( \mathbf{V} = [v(1) \cdots v(T)] \), the conditional probability density function of \( g_m(t) \) can be derived from [23], and

\[
g_m^{(b)}(t) = \mathbb{E}\{g_m(t) \mid V; \Phi^{(b-1)}, \Omega^{(b-1)}\} = d(\theta_m^{(b-1)}) f_m^{(b-1)}(t) + \Sigma_m^{(b-1)} \Sigma_m^{-1} [v(t) - D(\theta_m^{(b-1)}) f_m^{(b-1)}(t)], \forall m, t, \tag{8a}
\]

\[
G_m^{(b)} = \mathbb{D}\{g_m(t) \mid V; \Phi^{(b-1)}, \Omega^{(b-1)}\} = \Sigma_m^{(b-1)} - \Sigma_m^{(b-1)} \Sigma_m^{-1} \Sigma_m^{(b-1)}, \forall m, t. \tag{8b}
\]

3.1.2 M-step

The EM algorithm estimates \( \Phi = (\theta, \mathbf{F}) \) and \( \Omega = (\sigma_1, \ldots, \sigma_M) \) by maximizing (7), which results in the \( M \) parallel subproblems:

\[
\min_{\theta_m \in (0, \pi), f_m, \sigma_m > 0} \sum_{n=1}^{N} \left[ \ln(\sigma_{n,m}) + c_m^{(b)}(\sigma_{n,m}) \right] + \frac{1}{T} \sum_{t=1}^{T} \left[ g_m^{(b)}(t) - d(\theta_m)f_m(t) \right]^H \Sigma_m^{-1} [g_m^{(b)}(t) - d(\theta_m)f_m(t)], \forall m, \tag{9}
\]

where \( f_m = [f_m(1) \cdots f_m(T)] \) and \( c_m^{(b)} = [G_m^{(b)}]_{n,n} = \sigma_n^{(b-1)} (1 - \sigma_n^{(b-1)} / \sigma_{n,m}^{(b-1)}) \) with \( \sigma_n^{(b-1)} = \sum_{m=1}^{M} \sigma_{n,m}^{(b-1)} \).

Subproblems (9) are hard to be reduced to parallel subproblems each with respect to only one exclusive DOA, we thus resort to the ECM algorithm [21] and replace the M-step with the following two conditional maximization steps (CM-steps), i.e., the EM algorithm becomes the ECM algorithm.

3.1.3 CM-steps

At the first CM-step, the ECM algorithm estimates \( \Phi \) but holds \( \Omega = \Omega^{(b-1)} \) fixed. Then, subproblems (9) are reduced to

\[
\min_{\theta_m \in (0, \pi), f_m} \frac{1}{T} \sum_{t=1}^{T} \left\| g_m^{(b)}(t) - d(\theta_m)f_m(t) \right\|^2, \forall m, \tag{10}
\]

where \( \Sigma_m^{-1/2} = \text{diag}\{1/\sqrt{\sigma_{1,m}}, \ldots, 1/\sqrt{\sigma_{N,m}}\} \), \( g_m^{(b)}(t) = [\Sigma_m^{(b-1)}]^{-1/2} g_m^{(b)}(t) \), and \( d(\theta_m) = [\Sigma_m^{(b-1)}]^{-1/2} d(\theta_m) \). Subproblems (10) can be solved in a separable manner and reduced to \( M \) parallel one-dimensional search subproblems each with respect to only one exclusive DOA [8]. Accordingly, the estimate of \( \Phi \) is updated by

\[
\theta_m^{(b)} = \arg \max_{\theta_m \in (0, \pi)} d_H(\theta_m) \tilde{R}_m(\theta_m), \forall m, \tag{11a}
\]
\[ f_m^{(b)}(t) = \tilde{d}^H(\theta_m^{(b)}) \tilde{g}_m^{(b)}(t) / q_m^{(b)}, \forall m, t, \]  

where \( \tilde{R}_m^{(b)} = (1/T) \sum_{t=1}^{T} \tilde{g}_m^{(b)}(t) \tilde{g}_m^{(b)}(t)^H \) and \( q_m^{(b)} = \tilde{d}^H(\theta_m^{(b)}) \tilde{d}(\theta_m^{(b)}) = \sum_{n=1}^{N} \sigma_n^{(b-1)} \).

At the second CM-step, the ECM algorithm estimates \( \Omega \) but holds \( \Phi = \Phi^{(b)} \) fixed. Then, subproblems (9) are reduced to

\[ \min_{\sigma_n^{(b)}>0} \sum_{n=1}^{N} \left[ \ln(\sigma_n,m) + \frac{c_n^{(b)} + d_n^{(b)}}{\sigma_n,m} \right], \forall m, \]  

where \(| \cdot |\) denotes the modulus of a complex number and

\[ d_n^{(b)} = \frac{1}{T} \sum_{t=1}^{T} \left[ |g_m^{(b)}(t) - d(\theta_m^{(b)}) f_m^{(b)}(t)| \right]^2 \geq 0, \forall n, m. \]

Thus, the estimate of \( \Omega \) is updated by

\[ \sigma_n^{(b)} = c_n^{(b)} + d_n^{(b)}, \forall n, m. \]  

Unfortunately, simulation results show that the operation of the algorithm may not be smooth when the estimate of \( \Omega \) is updated by (13). To ensure the stability, we decrease the difference between \( \sigma_n^{(b)} \) and \( \sigma_n^{(b-1)} \) by modifying (13) as

\[ \sigma_n^{(b)} = \beta \sigma_n^{(b-1)} + (1 - \beta)(c_n^{(b)} + d_n^{(b)}), \forall n, m, \]

where \( \beta \in [0, 1] \) is adjusted in simulation. Note that (14) guarantees the following monotonicity

\[ \ln(\sigma_n^{(b)}) + \frac{c_n^{(b)} + d_n^{(b)}}{\sigma_n^{(b)} - \sigma_n^{(b-1)}} \leq \ln(\sigma_n^{(b-1)}) + \frac{c_n^{(b)} + d_n^{(b)}}{\sigma_n^{(b)} - \sigma_n^{(b-1)}}, \forall n, m. \]

Moreover, if \( \sigma_n^{(b-1)} > 0, \forall m \), we can obtain \( c_n^{(b)} = \sigma_n^{(b-1)}(1 - \sigma_n^{(b-1)}/\sigma_n^{(b-1)}) > 0, \forall n, m, \) and \( \sigma_n^{(b)} \geq \beta \sigma_n^{(b-1)} + (1 - \beta)\sigma_n^{(b)} > 0, \forall n, m, \) i.e., \( \sigma_n^{(b)} > 0, \forall m. \)

Although the M-step of the EM algorithm is replaced with the above two CM-steps of the ECM algorithm, we can easily prove that the monotonicity of GEM algorithms still holds, i.e.,

\[ \mathcal{L}(\Phi^{(b)}, \sigma^{(b)}) \geq \mathcal{L}(\Phi^{(b-1)}, \sigma^{(b-1)}). \]

As stated in [21], the ECM algorithm is a GEM algorithm.

### 3.2 SAGE Algorithm

According to the above GEM algorithm, the corresponding SAGE algorithm is presented for speeding up the convergence. At every iteration, the SAGE algorithm sequentially updates the DOA estimates from \( \theta_1 \) to \( \theta_M \) using \( M \) cycles. For convenience, let \( (\cdot)^{(b,i)} \) denote an iterative value at the \( i \)th cycle of the \( b \)th iteration, \( (\cdot)^{(b,M)} = (\cdot)^{(b,1,0)} = (\cdot)^{(b)} \).

When the SAGE algorithm updates the estimate of \( \theta_i \) at the \( i \)th cycle of the \( b \)th iteration, all of the noise is first allocated to the \( i \)th source signal component by [9]–[11]

\[ g_m(t) = \begin{cases} 
\text{d}(\theta_m) f_m(t) + z(t) & m = i, \\
\text{d}(\theta_m) f_m(t) & m \neq i, 
\end{cases} \]

(15)
which indicates that $g_i(t) \sim \mathcal{CN}(d(\theta_i)f_i(t), \Sigma)$ and $g_m(t)$ is deterministic for $m \neq i$. The corresponding complete-data LLF is written as

$$ U(\theta_i, f_i, \sigma) = \sum_{t=1}^{T} \ln p(g_i(t); \theta_i, f_i(t), \sigma) = -TN \ln(\pi) - T \sum_{n=1}^{N} \ln(\sigma_n) - \sum_{t=1}^{T} [g_i(t) - d(\theta_i)f_i(t)]^H \Sigma^{-1} [g_i(t) - d(\theta_i)f_i(t)]. \quad (16) $$

The E- and CM-steps at the $i$th cycle of the $b$th iteration are introduced below.

### 3.2.1 E-step

Based on (16), the SAGE algorithm calculates the conditional expectation of $U(\theta_i, f_i, \sigma)$, i.e.,

$$ \mathbb{E} \{ U(\theta_i, f_i, \sigma) \mid V; \Phi^{(b,i-1)}, \sigma^{(b,i-1)} \} = -TN \ln(\pi) - T \sum_{n=1}^{N} \ln(\sigma_n) - \sum_{t=1}^{T} [g_i^{(b)}(t) - d(\theta_i)f_i(t)]^H \Sigma^{-1} [g_i^{(b)}(t) - d(\theta_i)f_i(t)], \quad (17) $$

where $\Phi = (\theta, F)$ and

$$ g_i^{(b)}(t) = \mathbb{E} \{ g_i(t) \mid V; \Phi^{(b,i-1)}, \sigma^{(b,i-1)} \} = d(\theta_i^{(b,i-1)})f_i^{(b,i-1)}(t) + [v(t) - D(\theta_i^{(b,i-1)})f_i^{(b,i-1)}(t)]$, \forall t, \quad (18a) $$

$$ \mathbb{D} \{ g_i(t) \mid V; \Phi^{(b,i-1)}, \sigma^{(b,i-1)} \} = 0_N, \forall t. \quad (18b) $$

### 3.2.2 CM-steps

The SAGE algorithm tries to estimate $\theta_i$, $f_i$, and $\sigma$ by maximizing (17), i.e.,

$$ \min_{\theta_i \in (0,\pi), f_i, \sigma > 0} \sum_{n=1}^{N} \ln(\sigma_n) + \frac{1}{T} \sum_{t=1}^{T} [g_i^{(b)}(t) - \tilde{d}(\theta_i)f_i(t)]^H \Sigma^{-1} [g_i^{(b)}(t) - d(\theta_i)f_i(t)]. \quad (19) $$

However, problem (19) is hard to be reduced to a one-dimensional search problem with respect to only $\theta_i$, we use two CM-steps based on the ECM algorithm [21].

At the first CM-step, the SAGE algorithm estimates $\theta_i$ and $f_i$ but holds $\sigma = \sigma^{(b,i-1)}$ fixed. Then, problem (19) is reduced to

$$ \min_{\theta_i \in (0,\pi), f_i} \frac{1}{T} \sum_{t=1}^{T} \| \tilde{g}_i^{(b)}(t) - \tilde{d}(\theta_i)f_i(t) \|^2, \quad (20) $$

where $\tilde{g}_i^{(b)}(t) = [\Sigma^{(b,i-1)}]^{-1/2} g_i^{(b)}(t)$ and $\tilde{d}(\theta_i) = [\Sigma^{(b,i-1)}]^{-1/2} d(\theta_i)$. Following (11a) and (11b), the estimates of $\theta_i$ and $f_i$ are updated by

$$ \theta_i^{(b)} = \theta_i^{(b,i)} = \arg \max_{\theta_i \in (0,\pi)} \tilde{d}^H(\theta_i) \tilde{R}_i(\theta_i), \quad (21a) $$

and
\[ f^{(b)}_i(t) = f^{(b,i)}_i(t) = \tilde{d}^H(\theta^{(b)}_i)\tilde{g}^{(b)}_i(t)/q^{(b,i)}, \forall t, \]  

(21b)

where \( \tilde{R}_i^{(b)} = (1/T)\sum_{t=1}^{T} \tilde{g}^{(b)}_i(t)\tilde{g}^{(b)}_i(t)^H \) and \( q^{(b,i)} = \tilde{d}^H(\theta^{(b)}_i)\tilde{d}(\theta^{(b)}_i) = \sum_{n=1}^{N} [\sigma_n^{(b,i-1)}]^{-1}. \)

At the second CM-step, the SAGE algorithm estimates \( \sigma \) but holds \( \theta_i = \theta_i^{(b)} \) and \( f_i = f_i^{(b)} \) fixed. Then, problem (19) is reduced to

\[
\min_{\sigma > 0} \sum_{n=1}^{N} \left[ \ln(\sigma_n) + \frac{d_n^{(b,i)}}{\sigma_n} \right],
\]

(22)

where \( d_n^{(b,i)} = (1/T)\sum_{t=1}^{T} \left| \left[ g^{(b)}_i(t) - d(\theta^{(b)}_i) f^{(b)}_i(t) \right]_n \right|^2 \geq 0, \forall n. \) Thus, the estimate of \( \sigma \) is updated by

\[
\sigma_n^{(b,i)} = d_n^{(b,i)}, \forall n.
\]

(23)

Unfortunately, the operation of the algorithm always is unstable in simulation when the estimate of \( \sigma \) is updated by (23). Like (14), we modify (23) as

\[
\sigma_n^{(b,i)} = \gamma \sigma_n^{(b,i-1)} + (1 - \gamma)d_n^{(b,i)}, \forall n,
\]

(24)

where \( \gamma \in (0,1] \) is adjusted in simulation and tends to be larger than \( \beta \). Note that in (24), \( \sigma_n^{(b,i)} \geq \gamma \sigma_n^{(b,i-1)} > 0 \) if \( \sigma_n^{(b,i-1)} > 0 \), which ensures that if \( \sigma_n^{(b,i-1)} > 0 \), we have \( \sigma_n^{(b,i)} > 0 \).

The other parameter estimates in \( (\Phi^{(b,i-1)}, \sigma^{(b,i-1)}) \) are not updated at this cycle and their iterative values are denoted by

\[
\dot{\theta}_m^{(b,i)} = \dot{\theta}_m^{(b,i-1)}, \forall m \neq i; \tag{25a}
\]

\[
f_m^{(b,i)}(t) = f_m^{(b,i-1)}(t), \forall m \neq i, t. \tag{25b}
\]

After this cycle, we have [9]

\[
L(\Phi^{(b,i)}, \sigma^{(b,i)}) \geq L(\Phi^{(b,i-1)}, \sigma^{(b,i-1)}).
\]

(26)

Thus, after the \( b \)th iteration, the monotonicity of the SAGE algorithm can be proved by

\[
L(\Phi^{(b)}, \sigma^{(b)}) = L(\Phi^{(b,M)}, \sigma^{(b,M)}) \geq \cdots \geq L(\Phi^{(b,0)}, \sigma^{(b,0)}) = L(\Phi^{(b-1)}, \sigma^{(b-1)}).
\]

(27)

### 4 Stochastic Signal Model

In this section, we present two SAGE algorithms for solving problem (6).

#### 4.1 First SAGE Algorithm

The first SAGE algorithm simultaneously updates the DOA estimates based on the EM algorithm. To this end, we first try to use the EM algorithm and the E- and M-steps at the \( b \)th iteration are introduced below.
4 STOCHASTIC SIGNAL MODEL

4.1.1 E-step

The EM algorithm calculates the conditional expectation of the complete-data LLF in (5b) [7], i.e.,

\[ \mathbb{E} \{ \mathcal{U}(\Phi, \sigma) | \mathbf{V}; \Phi^{(b-1)}, \sigma^{(b-1)} \} = -T M N \ln(\pi) - T \sum_{m=1}^{M} \left[ \ln \left( \text{Det}(\mathbf{H}_m) \right) + \text{Tr}(\mathbf{H}_m^{-1}\hat{\mathbf{R}}^{(b)}_m) \right] \tag{28} \]

with \( \Phi = (\theta, \mathbf{P}) \) and

\[ \hat{\mathbf{R}}^{(b)}_m = \mathbb{E} \{ \hat{\mathbf{R}}_m | \mathbf{V}; \Phi^{(b-1)}, \sigma^{(b-1)} \} \]
\[ = \mathbf{H}_m^{(b-1)}\mathbf{H}_m^{(b-1)} + \mathbf{H}_m^{(b-1)} - \mathbf{H}_m^{(b-1)} \mathbf{H}_m^{(b-1)}, \forall m, \tag{29} \]

where the conditional probability density function of \( g_m(t) \) can be derived from [23] and

\[ \mathbb{E} \{ g_m(t) | \mathbf{V}; \Phi^{(b-1)}, \sigma^{(b-1)} \} = \mathbf{H}_m^{(b-1)} \mathbf{H}_m^{(b-1)} v(t), \forall m, t, \tag{30a} \]
\[ \mathbb{D} \{ g_m(t) | \mathbf{V}; \Phi^{(b-1)}, \sigma^{(b-1)} \} = \mathbf{H}_m^{(b-1)} - \mathbf{H}_m^{(b-1)} \mathbf{H}_m^{(b-1)} \geq 0_N, \forall m, t. \tag{30b} \]

4.1.2 M-step

The EM algorithm estimates \( \Phi \) and \( \sigma \) by maximizing (28), i.e.,

\[ \min_{\theta \in \Phi, \mathbf{P} \geq 0, \sigma > 0} \sum_{m=1}^{M} \left[ \ln \left( \text{Det}(\mathbf{H}_m) \right) + \text{Tr}(\mathbf{H}_m^{-1}\hat{\mathbf{R}}^{(b)}_m) \right], \tag{31} \]

which is hard to be reduced to parallel subproblems since all the \( \mathbf{H}_m \)'s are related to \( \sigma \).

To proceed, we hold \( \sigma = \sigma^{(b-1)} \) fixed and reduce problem (31) to the \( M \) parallel subproblems:

\[ \min_{\theta_m \in (0, \pi), P_m \geq 0} \ln \left( \text{Det}(\mathbf{K}_m) \right) + \text{Tr}(\mathbf{K}_m^{-1}\hat{\mathbf{R}}^{(b)}_m), \forall m, \tag{32} \]

where \( \mathbf{H}_m = \left[ \Sigma^{(b-1)} \right]^{1/2} \mathbf{K}_m \left[ \Sigma^{(b-1)} \right]^{1/2}, \mathbf{K}_m = P_m \mathbf{d}(\theta_m)\mathbf{H}(\theta_m) + \alpha_m \mathbf{I}_N, \mathbf{d}(\theta_m) = \left[ \Sigma^{(b-1)} \right]^{-1/2} \mathbf{d}(\theta_m), \hat{\mathbf{R}}^{(b)}_m = \left[ \Sigma^{(b-1)} \right]^{-1/2} \mathbf{K}_m \left[ \Sigma^{(b-1)} \right]^{-1/2}. \]

Utilizing \( \text{Det}(\mathbf{K}_m) = (P_m q^{(b)} + \alpha_m)\alpha_m^{N-1} \) with \( q^{(b)} = \| \mathbf{d}(\theta_m) \|^2 = \sum_{n=1}^{M} [\sigma^{(b-1)}]^{-1} \) and

\[ \mathbf{K}_m^{-1} = \frac{1}{\alpha_m} \left( \mathbf{I}_N - \frac{P_m \mathbf{d}(\theta_m)\mathbf{H}(\theta_m)}{P_m q^{(b)} + \alpha_m} \right), \]

we further reduce subproblems (32) to

\[ \min_{\theta_m \in (0, \pi), P_m \geq 0} \ln \left( P_m q^{(b)} + \alpha_m \right) - \frac{P_m \mathbf{d}(\theta_m)\mathbf{H}(\theta_m)}{\alpha_m (P_m q^{(b)} + \alpha_m)}, \forall m, \tag{33} \]

which can be solved in a separable manner and reduced to \( M \) parallel one-dimensional search subproblems each with respect to only one exclusive DOA [8]. Accordingly, the estimate of \( \Phi \) is updated by

\[ \theta_m^{(b)} = \arg \max_{\theta_m \in (0, \pi)} \mathbf{d}(\theta_m)\mathbf{H}(\theta_m), \forall m, \tag{34a} \]
\[ \mathbf{P}_m^{(b)} = \max \left\{ \frac{1}{q^{(b)}} \left( \frac{\mathbf{d}(\theta_m)\mathbf{H}(\theta_m)}{q^{(b)}} - \alpha_m \right), 0 \right\}, \forall m, \tag{34b} \]

where \( \theta_m^{(b)} \) is indeterminate if \( \mathbf{P}_m^{(b)} = 0 \). Then, we have

\[ \mathcal{L}(\Phi^{(b)}, \sigma^{(b-1)}) \geq \mathcal{L}(\Phi^{(b-1)}, \sigma^{(b-1)}). \tag{35} \]
4.1.3 Additional E- and M-steps

The above M-step does not update the estimate of $\sigma$ due to the complexity in problem (31), we thus add novel E- and M-steps to obtain $\sigma^{(b)}$ easily, which leads to that the EM algorithm becomes the SAGE algorithm [9]. To this end, we use $\mathbf{F}$ and $\mathbf{Z} = [z(1) \cdots z(T)]$ as the complete data. The corresponding complete-data LLF is written as

$$\mathcal{U} (\mathbf{P},\mathbf{\sigma}) = \sum_{t=1}^{T} \left[ \ln p(f(t); \mathbf{P}) + \ln p(z(t); \mathbf{\sigma}) \right] = -TM \ln(\pi) - T \sum_{m=1}^{M} \left[ \ln(P_m) + \hat{P}_m / P_m \right] - TN \ln(\pi) - T \left[ \ln \left( \text{Det}(\Sigma) \right) + \text{Tr} (\Sigma^{-1} \hat{R}_z) \right],$$

(36)

where $\hat{P}_m = (1/T) \sum_{t=1}^{T} |f_m(t)|^2$ and $\hat{R}_z = (1/T) \sum_{t=1}^{T} z(t) z^H(t)$ are two complete-data sufficient statistics for $P_m$ and $\Sigma$, respectively. Based on (36), the additional E- and M-steps are introduced below.

At the additional E-step, the algorithm calculates the conditional expectation of $\mathcal{U} (\mathbf{P},\mathbf{\sigma})$, i.e.,

$$\mathbb{E} \{ \mathcal{U} (\mathbf{P},\mathbf{\sigma}) \mid \mathbf{V}; \Phi^{(b)}, \mathbf{\sigma}^{(b-1)} \} = -TM \ln(\pi) - T \sum_{m=1}^{M} \left[ \ln(P_m) + \hat{P}_m^{(b)}/P_m \right] - TN \ln(\pi) - T \left[ \ln \left( \text{Det}(\Sigma) \right) + \text{Tr} (\Sigma^{-1} \hat{R}_z^{(b)}) \right]$$

(37)

with

$$\hat{P}_m^{(b)} = \mathbb{E} \{ \hat{P}_m \mid \mathbf{V}; \Phi^{(b)}, \mathbf{\sigma}^{(b-1)} \} = [\mathbf{q}_m^{(b)}]^{H} \hat{\mathbf{R}}_v \mathbf{q}_m^{(b)} + P_m^{(b)} \left[ 1 - \mathbf{d}^H (\hat{\theta}_m^{(b)}) \mathbf{q}_m^{(b)} \right] \geq 0, \forall m,$$

(38a)

$$\hat{R}_z^{(b)} = \mathbb{E} \{ \hat{R}_z \mid \mathbf{V}; \Phi^{(b)}, \mathbf{\sigma}^{(b-1)} \} = \Sigma^{(b-1)} [\hat{\mathbf{H}}_v^{(b)}]^{-1} \hat{\mathbf{R}}_v [\hat{\mathbf{H}}_v^{(b)}]^{-1} \Sigma^{(b-1)} + \Sigma^{(b-1)} \Sigma^{(b-1)} [\hat{\mathbf{H}}_v^{(b)}]^{-1} \Sigma^{(b-1)} \geq 0_N,$$

(38b)

where $\hat{\mathbf{H}}_v^{(b)} = \sum_{m=1}^{M} P_m^{(b)} \mathbf{d}(\hat{\theta}_m^{(b)}) \mathbf{d}^H (\hat{\theta}_m^{(b)}) + \Sigma^{(b-1)}$, $\mathbf{d}_m^{(b)} = [\hat{\mathbf{R}}_v^{(b)}]^{-1} \mathbf{d}(\hat{\theta}_m^{(b)}) P_m^{(b)}$, and the conditional probability density functions of $f_m(t)$ and $z(t)$ can be derived from [23].

At the additional M-step, the algorithm estimates $\mathbf{P}$ and $\mathbf{\sigma}$ by maximizing (37), which results in the $M + N$ parallel subproblems:

$$\min_{P_m \geq 0} \ln(P_m) + \hat{P}_m^{(b)}/P_m, \forall m,$$

(39a)

$$\min_{\sigma_n > 0} \ln(\sigma_n) + [\hat{R}_z^{(b)}]_{n,n}/\sigma_n, \forall n.$$

(39b)

Thus, the estimates of $\mathbf{P}$ and $\mathbf{\sigma}$ are updated by

$$P_m^{(b)} = \hat{P}_m^{(b)}, \forall m,$$

(40a)

$$\sigma_n^{(b)} = [\hat{R}_z^{(b)}]_{n,n} \geq 0, \forall n,$$

(40b)

which indicate that the estimate of $\mathbf{P}$ is updated again at this iteration and $\sigma_n^{(b)} = 0$ is possible although its probability is very low. For example, if $M = 1$, $T = 1$, and $P_1^{(1)} = 0$ in (34b), we will have
\( \hat{H}_i^{(1)} = \Sigma^{(q)} > 0_N \) and \( \hat{R}_i^{(1)} = \mathbf{v}(1)\mathbf{v}^H(1) \) in (38b). Furthermore, if \( [\mathbf{v}(1)]_n = 0 \), we will obtain \( \sigma_n^{(1)} = 0 \) by (40b). To avoid \( \sigma_n^{(b)} = 0 \), we require that if \( \sigma_n^{(b)} = 0 \) in (40b), \( \sigma_n^{(b)} \) will be updated by
\[
\sigma_n^{(b)} = \zeta \sigma_n^{(b-1)} + (1 - \zeta)[\hat{R}_i^{(b)}]_{n,n},
\]
where \( \zeta \in (0, 1] \) and \( \sigma_n^{(b)} \geq \zeta \sigma_n^{(b-1)} > 0 \) if \( \sigma_n^{(b-1)} > 0 \), i.e., \( \sigma^{(b)} > 0 \) if \( \sigma^{(b-1)} > 0 \).

After the additional M-step, this iteration finishes and following (35), we can verify the monotonicity of the first SAGE algorithm by
\[
\mathcal{L}(\Phi^{(b)}, \sigma^{(b)}) \geq \mathcal{L}(\Phi^{(b)}, \sigma^{(b-1)}) \geq \mathcal{L}(\Phi^{(b-1)}, \sigma^{(b-1)}).
\]

### 4.2 Second SAGE Algorithm

The first SAGE algorithm updates the DOA estimates simultaneously in (34a), which leads to slow convergence. In order to speed up the convergence, we present the second SAGE algorithm, which updates the DOA estimates sequentially from \( \theta_1 \) to \( \theta_M \) like the SAGE algorithm in Subsection 3.2.

When the second SAGE algorithm updates the estimate of \( \theta_i \) at the \( i \)th cycle of the \( b \)th iteration, all of the noise is also allocated to the \( i \)th source signal component. According to (15), \( \mathbf{g}_i(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{H}_i) \) with \( \mathbf{H}_i = \mathbf{P}_i \mathbf{d}(\theta_i) \mathbf{d}^H(\theta_i) + \Sigma \) but for \( m \neq i \), the statistical distribution of \( \mathbf{g}_m(t) \) depends only on \( f_m(t) \).

The corresponding complete-data LLF is written as
\[
\mathcal{U}(\theta_i, \mathbf{P}, \sigma) = \sum_{t=1}^{T} \left[ \sum_{m \neq i} \ln(p(f_m(t); P_m)) + \ln(p(\mathbf{g}_i(t); \theta_i, P_i, \sigma)) \right]
= -T(M-1)\ln(\pi) - T \sum_{m \neq i} \ln(P_m) + \hat{P}_m/P_m - TN \ln(\pi) - T \left[ \ln \left( \text{Det}(\mathbf{H}_i) \right) + \text{Tr}(\mathbf{H}_i^{-1}\hat{\mathbf{R}}_i) \right],
\]
which is unsuitable for updating the estimate of \( \sigma \) due to the complexity of \( \ln \left( \text{Det}(\mathbf{H}_i) \right) + \text{Tr}(\mathbf{H}_i^{-1}\hat{\mathbf{R}}_i) \).

To proceed, we hold \( \sigma = \sigma^{(b-1)} \) fixed and let the algorithm only update the estimates of \( \theta_i \) and \( \mathbf{P} \) by the following E- and M-steps.

#### 4.2.1 E-step

The algorithm calculates the conditional expectation of \( \mathcal{U}(\theta_i, \mathbf{P}, \sigma) \), i.e.,
\[
\mathbb{E} \left\{ \mathcal{U}(\theta_i, \mathbf{P}, \sigma) | \mathbf{V}, \Phi^{(b,i-1)}, \sigma^{(b-1)} \right\} = -T(M-1)\ln(\pi) - T \sum_{m \neq i} \ln(P_m) + \hat{P}_m/P_m - T \left[ \ln \left( \text{Det}(\mathbf{H}_i) \right) + \text{Tr}(\mathbf{H}_i^{-1}\hat{\mathbf{R}}_i) \right],
\]
where
\[
\hat{P}_m^{(b,i)} = \mathbb{E} \left\{ \hat{P}_m | \mathbf{V}, \Phi^{(b,i-1)}, \sigma^{(b-1)} \right\}
= P_m^{(b,i-1)} \left[ 1 - \mathbf{d}^H(\theta_m^{(b,i-1)}) \mathbf{q}_m^{(b,i-1)} \right] + [\mathbf{q}_m^{(b,i-1)}]^H \hat{\mathbf{R}}_i \mathbf{q}_m^{(b,i-1)} \geq 0
\]
with
\[
\hat{\mathbf{H}}_i^{(b,i-1)} = \sum_{m=1}^{M} P_m^{(b,i-1)} \mathbf{d}(\theta_m^{(b,i-1)}) \mathbf{d}^H(\theta_m^{(b,i-1)}) + \Sigma^{(b-1)} \text{ and } \mathbf{q}_m^{(b,i-1)} = [\mathbf{H}_i^{(b,i-1)}]^{-1} \mathbf{d}(\theta_m^{(b,i-1)}) \]
and
\[
\hat{\mathbf{R}}_i^{(b,i)} = \mathbb{E} \left\{ \hat{\mathbf{R}}_i | \mathbf{V}, \Phi^{(b,i-1)}, \sigma^{(b-1)} \right\}
= \mathbf{H}_i^{(b,i-1)} [\mathbf{H}_i^{(b,i-1)}]^{-1} \hat{\mathbf{R}}_i [\mathbf{H}_i^{(b,i-1)}]^{-1} \hat{\mathbf{H}}_i^{(b,i-1)} + \hat{\mathbf{H}}_i^{(b,i-1)} - \mathbf{H}_i^{(b,i-1)} [\mathbf{H}_i^{(b,i-1)}]^{-1} \mathbf{H}_i^{(b,i-1)} \geq 0_N
\]
with
\[
\mathbf{H}_i^{(b,i-1)} = \mathbf{P}_i^{(b,i-1)} \mathbf{d}(\theta_i^{(b,i-1)}) \mathbf{d}^H(\theta_i^{(b,i-1)}) + \Sigma^{(b-1)}.\]
4.2.2 M-step

The algorithm estimates $\theta_i$ and $P$ by maximizing (43), i.e.,

$$
\min_{\theta_i \in (0, \pi), P \geq 0} \sum_{m \neq i} \left[ \ln(P_m) + \hat{P}^{(b,i)}_m / P_m \right] + \left[ \ln(\det(H_i)) + \text{Tr}(H_i^{-1}\hat{R}^{(b)}_i) \right],
$$

(46)

which leads to that

$$
P^{(b,i)}_m = \hat{P}^{(b,i)}_m, \forall m \neq i,
$$

(47)

while the estimates of $\theta_i$ and $P_i$ are updated by

$$
\min_{\theta_i \in (0, \pi), P_i \geq 0} \ln(\det(K_i)) + \text{Tr}(K_i^{-1}\hat{R}^{(b)}_i),
$$

(48)

where $H_i = \left[ \Sigma^{(b-1)} \right]^{1/2} K_i \left[ \Sigma^{(b-1)} \right]^{1/2}$, $K_i = P_i \tilde{d}(\theta_i) \tilde{d}^H(\theta_i) + I_N$, $\tilde{d}(\theta_i) = [\Sigma^{(b-1)}]^{-1/2} d(\theta_i)$, $\tilde{R}^{(b)}_i = (\Sigma^{(b-1)})^{-1/2} \hat{R}^{(b)}_i (\Sigma^{(b-1)})^{-1/2}$. Following (34a)–(34b), the estimates of $\theta_i$ and $P_i$ can be updated by

$$
\theta_i^{(b)} = \hat{\theta}_i^{(b,i)} = \arg \max_{\theta_i \in (0, \pi)} \tilde{d}^H(\theta_i) \tilde{R}^{(b)}_i \tilde{d}(\theta_i),
$$

(49a)

$$
P_i^{(b,i)} = \max \left\{ \frac{1}{q^{(b)}} \left( \frac{\tilde{d}^H(\theta_i^{(b)}) \tilde{R}^{(b)}_i \tilde{d}(\theta_i^{(b)})}{q^{(b)}} - 1 \right), 0 \right\}. 
$$

(49b)

The other signal parameter estimate(s) in $\Phi^{(b,i-1)} = (\theta^{(b,i-1)}, P^{(b,i-1)})$ is(are) not updated at this cycle and the iterative value(s) is(are) denoted by

$$
\theta_m^{(b,i)} = \theta_m^{(b,i-1)}, \forall m \neq i.
$$

(50)

After this cycle, we have

$$
\mathcal{L}(\Phi^{(b,i)}, \sigma^{(b-1)}) \geq \mathcal{L}(\Phi^{(b,i-1)}, \sigma^{(b-1)}).
$$

(51)

The above E- and M-steps (or cycle) are repeated until the estimate of $\theta_M$ is updated and we further have

$$
\mathcal{L}(\Phi^{(b)}, \sigma^{(b-1)}) = \mathcal{L}(\Phi^{(b,M)}, \sigma^{(b-1)}) \geq \cdots \geq \mathcal{L}(\Phi^{(b,0)}, \sigma^{(b-1)}) = \mathcal{L}(\Phi^{(b-1)}, \sigma^{(b-1)}).
$$

(52)

4.2.3 Additional E- and M-steps

The above procedure does not update the estimate of $\sigma$, so we also use the additional E- and M-steps 4.1.3 of the first SAGE algorithm to obtain $\sigma^{(b)}$ easily. After the additional E- and M-steps, the $b$th iteration finishes and following (52), we can verify the monotonicity of the second SAGE algorithm by

$$
\mathcal{L}(\Phi^{(b)}, \sigma^{(b)}) \geq \mathcal{L}(\Phi^{(b)}, \sigma^{(b-1)}) \geq \mathcal{L}(\Phi^{(b-1)}, \sigma^{(b-1))}.
$$

(53)
5 Convergence Properties of the EM-Type Algorithms

5.1 Convergence Point

It is easy to verify that the above EM-type algorithms satisfy standard regularity conditions and the two algorithms for the same signal model always converge to different stationary points or the same stationary point of $\mathcal{L}(\Phi, \sigma)$ [6], [9], [24].

It is well known that these algorithms, as “hill climbing” algorithms, require accurate initial points for obtaining their global maximum points. Moreover, note that processing the same samples, the two algorithms for the same signal model may converge to different stationary points of $\mathcal{L}(\Phi, \sigma)$ given the same initial point, which indicates that one of the two algorithm may be more efficient for avoiding the convergence to an unwanted stationary point of $\mathcal{L}(\Phi, \sigma)$ than the other. Hence, we compare convergence points of the EM-type algorithms given poor initial points in the next section.

5.2 Complexity and Stability

The computational complexities of the EM-type algorithms at the $b$th iteration mainly lie in the $M$ one-dimensional search problems:

$$\theta_m^{(b)} = \arg \max_{\theta_m \in (0, \pi)} \tilde{d}^H(\theta_m)\tilde{R}_m \tilde{d}(\theta_m), \forall m.$$  \hfill (54)

However, we find that when the powers of sources are unequal, the DOA estimates of multiple sources, updated by (54), tend to be consistent with the true DOA of the source with the largest power. As a consequence, the EM-type algorithms are unstable.

To address this issue, we reduce the difference between $\theta_m^{(b)}$ and $\theta_m^{(b-1)}$ and still use the method in our previous works [12], [22], i.e., choosing $\theta_m^{(b-1)}$ as the initial point of some gradient algorithm and then applying this gradient algorithm to search a local maximum point of $\tilde{d}^H(\theta_m)\tilde{R}_m \tilde{d}(\theta_m)$ as $\theta_m^{(b)}$, which leads to Algorithm 1 in the next section. Using this method, we still have

$$\tilde{d}^H(\theta_m^{(b)})\tilde{R}_m \tilde{d}(\theta_m^{(b)}) \geq \tilde{d}^H(\theta_m^{(b-1)})\tilde{R}_m \tilde{d}(\theta_m^{(b-1)}), \forall m,$$ \hfill (55)

which guarantees the monotonicity of the EM-type algorithms.

6 Simulation Results

Simulation results are given to illustrate the convergence of the EM-type algorithms. We adopt $M = 2$ and $\|\theta^{(b)} - \theta^{(b-1)}\| \leq 0.001^\circ$ as the stopping criterion of the EM-type algorithms. $\Phi$ in the deterministic signal model is also generated by the independent random numbers $f_m(t) \sim \mathcal{CN}(0, P_m)$. Algorithm 1 is used to search the $\theta_m^{(b)}$’s in (55). Moreover, $N = 10$ and $\sigma = [1.1 2.3 3.4 2.1 3.0 5.2 2.6 7.1 0].^T$.

6.1 Deterministic Signal Model

Fig. 1 plots the $\mathcal{L}(\Phi^{(b)}, \sigma^{(b)})$’s, $\theta_1^{(b)}$’s, and $\theta_2^{(b)}$’s obtained by the GEM and SAGE algorithms under one realization without loss of generality. Both algorithms share the same initial point and process the
Algorithm 1 Gradient Ascent Based Angle Estimation

1: \( h(\theta_m) = \hat{d}^H(\theta_m) \mathbf{R}^{(b)}_m \hat{d}(\theta_m) \), initialize \( \theta_m = \theta^{(b-1)}_m \in (0, \pi) \).

2: while \( |h'(\theta_m)| > 0.001 \) do

3: \( t = 0.1 \times \begin{cases} (\pi - \theta_m)/h'(\theta_m), & h'(\theta_m) > 0, \\ -\theta_m/h'(\theta_m), & h'(\theta_m) < 0. \end{cases} \)

4: while \( h(\theta_m + th'(\theta_m)) < h(\theta_m) + 0.3t|h'(\theta_m)|^2 \) do

5: \( t = 0.5t \).

6: end while

7: \( \theta_m = \theta_m + th'(\theta_m) \).

8: end while

9: \( \theta^{(b)}_m = \theta_m \).

It is straightforward to see that given a good initial point, both algorithms obtain consistent DOA estimates and the SAGE algorithm has faster convergence than the GEM algorithm.

Fig. 2 shows a scatter plot of the DOA estimates obtained by both algorithms under 100 independent realizations. Given the same initial point, the same samples of each realization are processed by both algorithms. In Fig. 2, the total numbers of wanted points from the GEM and SAGE algorithms are 8 and 100, respectively. Hence, we conclude that given a poor initial point, the SAGE algorithm can avoid the convergence to an unwanted stationary point of \( \mathcal{L}(\Phi, \sigma) \) more efficiently than the GEM algorithm.

According to Figs. 1 and 2, we can conclude that for the deterministic signal model, the SAGE algorithm outperforms the GEM algorithm. Thus, we only use the SAGE algorithm in Figs. 3 and 4.

Figs. 3 and 4 compare the root mean square error (RMSE) performances of DOA estimation obtained by the SAGE algorithm for the nonuniform and uniform noise models. The SAGE algorithm for unknown uniform noise is presented in [22]. The SAGE algorithm for each noise model is performed based on a good initial point for avoiding the convergence to an unwanted stationary point efficiently. Each RMSE in Figs. 3 and 4 is computed from 1000 independent realizations with the same \( \mathbf{F} \) and the corresponding Cramer-Rao lower bound (CRLB) is also provided [19]. From Figs. 3 and 4, we can observe that as expected, the SAGE algorithm for nonuniform noise yields smaller RMSEs than that for uniform noise. Moreover, we note that the SAGE algorithm for nonuniform noise cannot achieve the CRLB of \( \theta \) by increasing \( T \) or \( P \), which is consistent with a main conclusion in [4], i.e., the deterministic ML estimator of \( \theta \) is not statistically efficient if the number of sensors \( N \) is small.

6.2 Stochastic Signal Model

Fig. 5 plots the \( \mathcal{L}(\Phi^{(b)}, \sigma^{(b)}) \)'s, \( \theta^{(b)}_1 \)'s, and \( \theta^{(b)}_2 \)'s obtained by the first and second SAGE algorithms under one realization. Both algorithms share the same initial point and process the same samples. It is straightforward to observe that given a good initial point, both algorithms obtain consistent DOA estimates and the second SAGE algorithm has faster convergence than the first SAGE algorithm.

Fig. 6 shows a scatter plot of the DOA estimates obtained by both algorithms under 100 inde-
Figure 1: \( \mathcal{L}(\Phi^{(b)}, \sigma^{(b)}), \theta_1^{(b)}, \) and \( \theta_2^{(b)} \) comparison of the GEM and SAGE algorithms under one realization with \( \beta = 0.5, \gamma = 0.9, T = 500, \theta_1 = 40^\circ, \theta_2 = 80^\circ, P_1 = 6, P_2 = 8, \theta_1^{(0)} = 45^\circ, \theta_2^{(0)} = 85^\circ, \)
\( \mathbf{F}^{(0)} = [1 1]^T, \Omega^{(0)} = (1, 1)/2, \) and \( \sigma^{(0)} = 1. \)

pendent realizations. Given the same initial point, the same samples of each realization are processed by both algorithms. In Fig. 6, the total numbers of wanted points from the first and second SAGE algorithms are 90 and 100, respectively. Hence, we conclude that given a poor initial point, the second SAGE algorithm can avoid the convergence to an unwanted stationary point of \( \mathcal{L}(\Phi, \sigma) \) more efficiently than the first SAGE algorithm.

According to Figs. 5 and 6, we can conclude that for the stochastic signal model, the second SAGE algorithm outperforms the first SAGE algorithm. Thus, we only use the second SAGE algorithm in Figs. 7–9.

Figs. 7–9 compare the RMSE performances of DOA estimation obtained by the SAGE algorithm for the nonuniform and uniform noise models. The SAGE algorithm for unknown uniform noise is presented in [22]. The SAGE algorithm for each noise model is performed based on a good initial point for avoiding the convergence to an unwanted stationary point efficiently. Each RMSE in Figs. 7–9 is computed from 1000 independent realizations and the corresponding CRLB is provided [19]. From Figs. 7–9, we can observe that as expected, the SAGE algorithm for nonuniform noise yields smaller RMSEs than that for uniform noise. Moreover, we note that the SAGE algorithm for nonuniform noise can achieve the CRLB of \( \theta \) by increasing \( T \), which is consistent with a main conclusion in [5], i.e., the stochastic ML estimator of \( \theta \) asymptotically achieves the CRLB of \( \theta \) if the number of samples \( T \) is large.
Figure 2: Scatter plot of the DOA estimates obtained by the GEM and SAGE algorithms under 100 independent realizations with $\beta = 0.95$, $\gamma = 0.99$, $T = 150$, $\theta_1 = 60^\circ$, $\theta_2 = 90^\circ$, $P_1 = P_2 = 3$, $\theta^{(0)}_1 = 40^\circ$, $\theta^{(0)}_2 = 110^\circ$, $F^{(0)} = [1 \ 1]^T$, $\Omega^{(0)} = (1, 1)/2$, and $\sigma^{(0)} = 1$.

6.3 Deterministic and Stochastic Signal Models

The SAGE algorithm for the deterministic signal model can process samples from the stochastic signal model, so we compare it with the second SAGE algorithm for the stochastic signal model in this subsection. Note that the two algorithms estimate the same DOA parameter $\theta$, the stopping criterion $\|\theta^{(b)} - \theta^{(b-1)}\| \leq 0.001^\circ$ is suitable.

Fig. 10 shows a scatter plot of the DOA estimates obtained by both algorithms under 50 independent realizations. The same samples of each realization are processed by both algorithms and each algorithm is performed based on a good initial point for avoiding the convergence to an unwanted stationary point efficiently. From Fig. 10, we can observe that both algorithms tend to obtain inconsistent DOA estimates. Hence, we compare the RMSE performances of both algorithms in Figs. 11 and 12.

Figs. 11 and 12 compare the RMSE performances of DOA estimation obtained by the SAGE algorithm for the deterministic and stochastic signal models. The SAGE algorithm for each signal model is performed based on a good initial point for avoiding the convergence to an unwanted stationary point efficiently. Each RMSE in Figs. 11 and 12 is computed from 1000 independent realizations. From Figs. 11 and 12, we can observe that the SAGE algorithm for the stochastic signal model yields smaller RMSEs than that for the deterministic signal model, which is consistent with a main conclusion in [5], i.e., the stochastic ML estimator of $\theta$ is statistically more efficient than the deterministic ML estimator of $\theta$. 
7 Conclusion

We have presented several EM-type algorithms for efficiently computing both the deterministic and stochastic ML estimators in unknown nonuniform noise. Specifically, we design a GEM algorithm and an SAGE algorithm for computing the deterministic ML estimator. Simulation results show that the SAGE algorithm converges faster and is more efficient for avoiding the convergence to an unwanted stationary point of the LLF. Moreover, we design two SAGE algorithms for computing the stochastic ML estimator, in which the first updates the DOA estimates simultaneously while the second updates the DOA estimates sequentially. Simulation results show that the second SAGE algorithm converges faster and is more efficient for avoiding the convergence to an unwanted stationary point of the LLF.

References

[1] H. Krim and M. Viberg, “Two decades of array signal processing research: the parametric approach,” IEEE Signal Processing Magazine, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[2] L. C. Godara, “Application of antenna arrays to mobile communications. II. Beam-forming and direction-of-arrival considerations,” Proceeding of the IEEE, vol. 85, no. 8, pp. 1195–1245, Aug. 1997.
Figure 4: RMSEs of DOA estimation obtained by the SAGE algorithm for the nonuniform and uniform noise models with $\gamma = 0.99$, $\theta_1 = 80^\circ$, $\theta_2 = 140^\circ$, $T = 400$, $P_1 = P_2 = P$, $\theta_1^{(0)} = 75^\circ$, $\theta_2^{(0)} = 135^\circ$, $F^{(0)} = [1 \ 1]^T$, $\sigma^{(0)} = 1$, and $\sigma^{(0)} = 1$.

[3] I. Ziskind and M. Wax, “Maximum likelihood localization of multiple sources by alternating projection,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[4] P. Stoica and A. Nehorai, “MUSIC, maximum likelihood, and Cramer-Rao bound,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[5] P. Stoica and A. Nehorai, “Performance study of conditional and unconditional direction-of-arrival estimation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[7] M. Feder and E. Weinstein, “Parameter estimation of superimposed signals using the EM algorithm,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[8] M. I. Miller and D. R. Fuhrmann, “Maximum-likelihood narrow-band direction finding and the EM algorithm,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 9, pp. 1560–1577, Sep. 1990.
Figure 5: $L(b), \phi(b), \sigma(b)$ comparison of the first and second SAGE algorithms under one realization with $\alpha = [0.5 \ 0.5]^T$, $T = 500$, $\theta_1 = 40^\circ$, $\theta_2 = 80^\circ$, $P_1 = 6$, $P_2 = 8$, $\theta_1^{(0)} = 45^\circ$, $\theta_2^{(0)} = 85^\circ$, $P^{(0)} = 1$, and $\sigma^{(0)} = 1$.

[9] J. A. Fessler and A. O. Hero, “Space-alternating generalized expectation-maximization algorithm,” *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.

[10] P. Chung and J. F. Bohme, “Comparative convergence analysis of EM and SAGE algorithms in DOA estimation,” *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 2940–2949, Dec. 2001.

[11] P. Chung and J. F. Bohme, “DOA estimation using fast EM and SAGE algorithms,” *Signal Processing*, vol. 82, no. 11, pp. 1753–1762, Nov. 2002.

[12] M. Gong and B. Lyu, “Alternating maximization and the EM algorithm in maximum-likelihood direction finding,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 9634–9645, Oct. 2021.

[13] A. M. Zoubir and S. Aouada, “High resolution estimation of directions of arrival in nonuniform noise,” in *Proc. ICASSP*, Montreal, QC, Canada, May 2004.

[14] Y. Wu, C. Hou, G. Liao, and Q. Guo, “Direction-of-arrival estimation in the presence of unknown nonuniform noise fields,” *IEEE Journal of Oceanic Engineering*, vol. 31, no. 2, pp. 504–510, Apr. 2006.
Figure 6: Scatter plot of the DOA estimates obtained by the first and second SAGE algorithms under 100 independent realizations with $\alpha = [0.5 \ 0.5]^T$, $T = 150$, $\theta_1 = 60^\circ$, $\theta_2 = 90^\circ$, $P_1 = P_2 = 3$, $\theta_1^{(0)} = 40^\circ$, $\theta_2^{(0)} = 110^\circ$, $P^{(0)} = 1$, and $\sigma^{(0)} = 1$.

[15] B. Liao, S. Chan, L. Huang, and C. Guo, “Iterative methods for subspace and DOA estimation in nonuniform noise,” IEEE Transactions on Signal Processing, vol. 64, no. 12, pp. 3008–3020, Jun. 2016.

[16] B. Liao, L. Huang, C. Guo, and H. C. So, “New approaches to direction-of-arrival estimation with sensor arrays in unknown nonuniform noise,” IEEE Sensors Journal, vol. 16, no. 24, pp. 8982–8989, Dec. 2016.

[17] M. Esfandiari, S. A. Vorobyov, S. Alibani, and M. Karimi, “Non-iterative subspace-based DOA estimation in the presence of nonuniform noise,” IEEE Signal Processing Letters, vol. 26, no. 6, pp. 848–852, Jun. 2019.

[18] D. Madurasinghe, “A new DOA estimator in nonuniform noise,” IEEE Signal Processing Letters, vol. 12, no. 4, pp. 337–339, Apr. 2005.

[19] M. Pesavento and A. B. Gershman, “Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise,” IEEE Transactions on Signal Processing, vol. 49, no. 7, pp. 1310–1324, Jul. 2001.

[20] C. E. Chen, F. Lorenzelli, R. E. Hudson, and K. Yao, “Stochastic maximum-likelihood DOA estimation in the presence of unknown nonuniform noise,” IEEE Transactions on Signal Processing, vol. 56, no. 7, pp. 3038–3044, Jul. 2008.
Figure 7: RMSEs of DOA estimation obtained by the SAGE algorithm for the nonuniform and uniform noise models with $\theta_1 = 45^\circ$, $\theta_2 = 65^\circ$, $P_1 = P_2 = 3$, $\theta_1^{(0)} = 40^\circ$, $\theta_2^{(0)} = 70^\circ$, $P^{(0)} = 1$, $\sigma^{(0)} = 1$, and $\sigma^{(0)} = 1$.

[21] X. Meng and D. B. Rubin, “Maximum likelihood estimation via the ECM algorithm: A general framework,” *Biometrika*, vol. 80, no. 2, pp. 267–278, Jun. 1993.

[22] M. Gong and B. Lyu, “EM and SAGE algorithms for DOA estimation in the presence of unknown uniform noise.” https://arxiv.org/abs/2208.07510

[23] I. B. Rhodes, “A tutorial introduction to estimation and filtering,” *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 688–706, Dec. 1971.

[24] C. F. Jeff Wu, “On the convergence properties of the EM algorithm,” *Annals of Statistics*, vol. 11, no. 1, pp. 95–103, Mar. 1983.
Figure 8: RMSEs of DOA estimation obtained by the SAGE algorithm for the nonuniform and uniform noise models with $\theta_1 = 80^\circ$, $\theta_2 = 140^\circ$, $T = 100$, $P_1 = P_2 = P$, $\theta_1^{(0)} = 75^\circ$, $\theta_2^{(0)} = 135^\circ$, $\mathbf{P}^{(0)} = 1$, $\sigma^{(0)} = 1$, and $\sigma^{(0)} = 1$.

Figure 9: RMSEs of DOA estimation obtained by the SAGE algorithm for the nonuniform and uniform noise models with $\theta_1 = 80^\circ$, $\theta_2 = 140^\circ$, $T = 10$, $P_1 = P_2 = P$, $\theta_1^{(0)} = 75^\circ$, $\theta_2^{(0)} = 135^\circ$, $\mathbf{P}^{(0)} = 1$, $\sigma^{(0)} = 1$, and $\sigma^{(0)} = 1$. 
Figure 10: Scatter plot of the DOA estimates obtained by the SAGE algorithm for the deterministic and stochastic signal models under 50 independent realizations with $\gamma = 0.99$, $T = 100$, $\theta_1 = 50^\circ$, $\theta_2 = 100^\circ$, $P_1 = 4$, $P_2 = 6$, $\theta_1^{(0)} = 45^\circ$, $\theta_2^{(0)} = 105^\circ$, $\mathbf{F}^{(0)} = [1 \ 1]^T$, $\mathbf{P}^{(0)} = 1$, and $\sigma^{(0)} = 1$.

Figure 11: RMSEs of DOA estimation obtained by the SAGE algorithm for the deterministic and stochastic signal models with $\gamma = 0.99$, $\theta_1 = 40^\circ$, $\theta_2 = 70^\circ$, $P_1 = 5$, $P_2 = 3$, $\theta_1^{(0)} = 45^\circ$, $\theta_2^{(0)} = 65^\circ$, $\mathbf{F}^{(0)} = [1 \ 1]^T$, $\mathbf{P}^{(0)} = 1$, and $\sigma^{(0)} = 1$. 
Figure 12: RMSEs of DOA estimation obtained by the SAGE algorithm for the deterministic and stochastic signal models with $\gamma = 0.99$, $T = 200$, $\theta_1 = 135^\circ$, $\theta_2 = 75^\circ$, $P_1 = P_2 = P$, $\theta_1^{(0)} = 130^\circ$, $\theta_2^{(0)} = 70^\circ$, $F^{(0)} = [1 1]^T$, $P^{(0)} = I$, and $\sigma^{(0)} = 1$. 