Recursive EM and SAGE-Inspired Algorithms With Application to DOA Estimation

Pei-Jung Chung and Johann F. Böhme, *Fellow, IEEE*

***Abstract*—This paper is concerned with recursive estimation using augmented data. We study two recursive procedures closely linked with the well-known expectation and maximization (EM) and space alternating generalized EM (SAGE) algorithms. Unlike iterative methods, the recursive EM and SAGE-inspired algorithms give a quick update on estimates given new data. Under mild conditions, estimates generated by these procedures are strongly consistent and asymptotically normally distributed. These mathematical properties are valid for a broad class of problems. When applied to direction of arrival (DOA) estimation, the recursive EM and SAGE-inspired algorithms lead to a very simple and fast implementation of the maximum-likelihood (ML) method. The most complicated computation in each recursion is inversion of the augmented information matrix. Through data augmentation, this matrix is diagonal and easy to invert. More importantly, there is no search in such recursive procedures. Consequently, the computational time is much less than that associated with existing numerical methods for finding ML estimates. This feature greatly increases the potential of the ML approach in real-time processing. Numerical experiments show that both algorithms provide good results with low computational cost.**

***Index Terms*—Array processing, DOA estimation, EM algorithm, recursive EM, recursive estimation, recursive SAGE, SAGE algorithm, stochastic approximation.**

I. Introduction

THE CENTRAL interest of this work is recursive parameter estimation using augmented data. The expectation and maximization (EM) [9] and space alternating generalized EM (SAGE) [12] algorithms are well-known iterative methods to locate modes of a likelihood function. If very large data sets are involved, numerical procedures can become very expensive. To overcome this problem, we propose two alternative procedures derived from the EM and SAGE algorithms in which the data are run through sequentially.

The first part of our paper is devoted to mathematical properties of the recursive EM and SAGE-inspired algorithms. The recursive EM and SAGE-inspired algorithms are stochastic approximation procedures with a specialized gain matrix derived from the augmented data. Choosing an appropriate gain matrix is particularly important if a good convergence rate is desired [2]. The algorithms presented in this paper provide a convenient way to design the gain matrix. In the pioneering paper [25], Titterington suggested the recursive EM algorithm and proved weak consistency and asymptotic normality for the univariate version. The analysis therein is based on classical methods of stochastic approximation [10], [22]. The results of [25] are extended to the multivariate case in [5]. Here, we consider a more general procedure in which constraint sets are allowed. We show that under proper conditions, the recursive EM algorithm leads to strong consistency and asymptotic normality. Rather than following the approach in [25], our investigation is based on the ordinary differential equation (ODE) method developed by Kushner and Clark [16] and Ljung [18], which characterizes the limit behavior of the algorithm by a mean limit ODE. The motivation for choosing the ODE method is that it simplifies the treatment of classical cases and provides a general approach to deal with complicated noise and dynamics.

In addition to the convergence properties of recursive EM, we also present a recursive procedure inspired by the SAGE algorithm. The recursive SAGE-inspired algorithm is derived for the specific case in which the parameter subsets in each cycle are disjoint. We use the term "SAGE-inspired" to emphasize the fact that rather than updating parameter subsets sequentially, the suggested recursion updates all elements of the parameter simultaneously. Under similar conditions, the recursive SAGE-inspired algorithm enjoys strong consistency and asymptotic normality as well. Although recursive EM and SAGE-inspired algorithms do not have optimal convergence rates, they are computationally preferable than the "optimal" stochastic approximation procedure. The augmented information matrix required by the recursive EM or SAGE-inspired algorithm is generally much easier to compute and invert than the Fisher information matrix required by the "optimal" recursive procedure.

The asymptotic results given in Section IV are applicable to a broad class of problems. In the second part, the recursive EM and SAGE-inspired algorithms are applied to array processing problems. Using recursive algorithms, there is no need to wait for a long time to collect the whole data set. Upon the arrival of new data, such procedures give a quick update of the estimate. This supports the use of the maximum-likelihood (ML) approach in real-time processing.

Based on recursive EM, algorithms for recursive direction of arrival (DOA) estimation were proposed in [13] and [19]. The methods in [13] are aimed at tracking multitargets. The signal waveforms were assumed to be known in the derivation.

Manuscript received February 19, 2003; revised July 13, 2004. This work was supported in part by the National Science Council (NSC), Taiwan, R.O.C., under Contact NSC-94-2213-E-009-050. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Constantinos B. Papadias.

P.-J. Chung is with the Department of Electronics Engineering, National Chiao Tung University, 300 Hsin Chu, Taiwan, R.O.C. (e-mail: peijung_chung@yahoo.com).

J. F. Böhme is with the Department of Electrical Engineering and Information Sciences, Ruhr-Universität Bochum, 44780 Bochum, Germany (e-mail: boehme@sth.ruhr-uni-bochum.de).

Digital Object Identifier 10.1109/TSP.2005.850339

In [19], both stochastic and conditional signal models are considered. Major differences between our algorithms and the previous work [19] include the following points.

1) The spectral parameters are updated by observed data rather than by augmented data to obtain better convergence rates and more stability.
2) The step size ϵ_n used in [19] is limited to an^{-1}.

In the current paper, we have a more flexible choice $an^{-\alpha}$. In most cases, a good choice of ϵ_n is essential for obtaining satisfactory convergence rates. With proper modifications, the proposed algorithms can also be used in tracking moving sources. Unlike subspace tracking methods [28], the recursive procedures based on the ML approach are not only applicable in the narrow band case but also in the wideband case. The first experimental and numerical results of recursive EM and SAGE-inspired algorithms are published in [7] and [8].

In the following, the recursive EM and SAGE-inspired algorithms are formulated in Sections II and III, respectively. Section IV deals with asymptotic properties of the proposed algorithms. Based on the recursive EM and SAGE-inspired algorithms, we develop recursive procedures for DOA estimation and discuss their convergence in Section V. Finally, numerical results are presented in Section VI.

II. Recursive EM Algorithm

Suppose $\underline{x}_1, \underline{x}_2, \ldots$ are independent observations, each with underlying density function $f_{\underline{X}}(\underline{x}|\underline{\vartheta})$. The augmented data specified by the EM algorithm $\underline{y}_1, \underline{y}_2, \ldots$ is independent with the density function $f_{\underline{Y}}(\underline{y}|\underline{\vartheta})$. The d-dimensional vector $\underline{\vartheta} \in \mathbb{R}^d$ represents the unknown parameter. According to [9], the augmented data $\underline{Y}$ is specified so that $\mathcal{M}(\underline{Y}) = \underline{X}$ is a many-to-one mapping. Let $\underline{\vartheta}^n$ denote the estimate after n observations. The following procedure

$$\underline{\vartheta}^{n+1} = \underline{\vartheta}^n + an^{-\alpha}\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n) \tag{1}$$

is aimed at finding the extremum $\underline{\vartheta}^*$ of $\log f_{\underline{X}}(\underline{x}|\underline{\vartheta})$ that would coincide with the ML estimator. The constants $a > 0$, $1/2 < \alpha \leq 1$ and

$$\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n) = \underline{\nabla}_{\underline{\vartheta}} \log f_{\underline{X}}(\underline{x}_n|\underline{\vartheta})\Big|_{\underline{\vartheta}=\underline{\vartheta}^n} \tag{2}$$

$$\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n) = \mathsf{E}\left[-\underline{\nabla}_{\underline{\vartheta}}\underline{\nabla}_{\underline{\vartheta}}^T \log f_{\underline{Y}}(\underline{y}|\underline{\vartheta})|\underline{x}_n, \underline{\vartheta}\right]\Big|_{\underline{\vartheta}=\underline{\vartheta}^n} \tag{3}$$

represent the gradient vector and the augmented information matrix calculated at $\underline{\vartheta}^n$, respectively. $\underline{\nabla}_{\underline{\vartheta}}$ is a column gradient operator with respect to $\underline{\vartheta}$. Recursion (1) is called a recursive EM algorithm because it is closely related to the EM algorithm.

In practice, the user will not allow the iterate to go beyond an upper or a lower bound. The iterate is confined to a bounded set Θ. Formally, the constrained algorithm can be expressed as

$$\underline{\vartheta}^{n+1} = \Pi_\Theta\left[\underline{\vartheta}^n + an^{-\alpha}\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right] \tag{4}$$

where Π_Θ is the projection onto the constraint set Θ. A simple example for Θ is a hyperrectangle $[a_1, b_1] \times \ldots \times [a_d, b_d]$ with $a_i < b_i$, $i = 1, \ldots, d$.

As pointed out by Titterington [25], recursion (1) has a strong link to the EM algorithm. By proper formulation of the EM algorithm, it can be shown that approximately, the augmented data log-likelihood is maximized by iterates generated by recursive EM.

III. Recursive SAGE-Inspired Algorithm

The SAGE algorithm [12] generalizes the idea of data augmentation to simplify computations of the EM algorithm. It preserves the stability of EM and can improve the convergence rate significantly in some settings. Instead of estimating all parameters at once, each iteration of SAGE consists of C cycles. The parameter subset associated with the cth cycle $\underline{\vartheta}_c$ is updated by maximizing the conditional expectation of the log-likelihood of the augmented data corresponding to this cycle. Like the EM algorithm, if the data sets are large, the required computational time may become long. Therefore, it could be advantageous to develop a recursive procedure finding the estimates generated by the SAGE algorithm.

For simplicity, we will only consider the case $\underline{\vartheta} = (\underline{\vartheta}_1, \ldots, \underline{\vartheta}_C)$ in which the parameter subsets are disjoint. Let $\underline{z}_{1c}, \underline{z}_{2c}, \ldots$ be the augmented data of the cth cycle corresponding to the observations $\underline{x}_1, \underline{x}_2, \ldots$. The characterizing density function is denoted by $f_{\underline{Z}_c}(\underline{z}_c|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)$, where $\underline{\tilde{\vartheta}}_c$ contains all components of $\underline{\vartheta}$, except those of $\underline{\vartheta}_c$. At the cth cycle, $\underline{\tilde{\vartheta}}_c$ is kept at a fixed value, and the conditional expectation of the augmented data log-likelihood is maximized with respect to $\underline{\vartheta}_c$. To find the maximizing point of $\log f_{\underline{X}}(\underline{x}|\underline{\vartheta})$, we suggest the following recursive procedure:

$$\underline{\vartheta}^{n+1} = \underline{\vartheta}^n + an^{-\alpha}\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n) \tag{5}$$

where the constants $a > 0$, $1/2 < \alpha \leq 1$, and $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$ is a block diagonal matrix

$$\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n) = \begin{pmatrix} \mathcal{I}_{\mathrm{SAGE}}^{[1]}(\underline{\vartheta}^n) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\mathrm{SAGE}}^{[2]}(\underline{\vartheta}^n) & \ldots & \mathbf{0} \\ \mathbf{0} & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \ldots & \mathcal{I}_{\mathrm{SAGE}}^{[C]}(\underline{\vartheta}^n) \end{pmatrix} \tag{6}$$

with the cth block

$$\mathcal{I}_{\mathrm{SAGE}}^{[c]}(\underline{\vartheta}^n) = \mathsf{E}\left[-\underline{\nabla}_{\underline{\vartheta}_c}\underline{\nabla}_{\underline{\vartheta}_c}^T \log f_{\underline{Z}_c}(\underline{z}_c|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)|\underline{x}_n, \underline{\vartheta}\right]\Big|_{\underline{\vartheta}=\underline{\vartheta}^n}. \tag{7}$$

If the constraint Θ is considered, recursion (5) can be written as

$$\underline{\vartheta}^{n+1} = \Pi_\Theta\left[\underline{\vartheta}^n + an^{-\alpha}\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right]. \tag{8}$$

It was mentioned in the previous section that there is a strong relationship between the EM algorithm and the recursive EM algorithm. Recursions (5) and (8), on the other hand, are similarly linked to the SAGE algorithm. Based on the whole data set $\{\underline{x}_1, \ldots, \underline{x}_n\}$, the SAGE algorithm updates the parameter subsets $(\underline{\vartheta}_1, \ldots, \underline{\vartheta}_C)$ sequentially in each iteration. Rather than giving a partial update on $\underline{\vartheta}$, the recursive procedure (5) [or (8)] generates a simultaneous update on all elements of $\underline{\vartheta}$ upon arrival of new data $\underline{x}_n$. The possibility of partial update on $\underline{\vartheta}$ and its influence on convergence properties are still under investigation.

The following derivation is based on Taylor expansion at the maximum point of each cycle. In the constrained case, we assume that the maximum point lies in the interior of Θ. Given the observations $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{n-1}$, the SAGE algorithm generates a sequence of estimates by repeating the following steps.

For $i = 1, 2, \ldots$
  Initialize $\underline{\vartheta}^{[i,0]}$

$$\underline{\vartheta}^{[i,0]} = \underline{\vartheta}^{[i-1,C]}.$$

  For $c = 1, \ldots, C$
    E-step: Evaluate

$$Q^{[c]}\left(\underline{\vartheta}_c, \underline{\vartheta}^{[i,c-1]}\right) = \mathsf{E}\left[\sum_{k=1}^{n-1} \log f_{\underline{Z}_c}(\underline{z}_{c,k}|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)| \underline{x}_1, \ldots, \underline{x}_{n-1}, \underline{\vartheta}^{[i,c-1]}\right]. \quad (9)$$

    M-step: Maximize $Q^{[c]}(\underline{\vartheta}_c, \underline{\vartheta}^{[i,c-1]})$ with respect to $\underline{\vartheta}_c$

$$\underline{\vartheta}_c^{[i,c]} = \arg\max_{\underline{\vartheta}_c} Q^{[c]}\left(\underline{\vartheta}_c, \underline{\vartheta}^{[i,c-1]}\right) \quad (10)$$

$$\underline{\vartheta}^{[i,c]} = \left[\underline{\vartheta}_1^{[i,c-1]}, \ldots, \underline{\vartheta}_{c-1}^{[i,c-1]}, \underline{\vartheta}_c^{[i,c]} \underline{\vartheta}_{c+1}^{[i,c-1]}, \ldots, \underline{\vartheta}_C^{[i,c-1]}\right]. \quad (11)$$

  end
end

Now, consider the recursive formulation of SAGE. At time instant n, define the "recursive" augmented likelihood as

$$L_{n+1}^{[c]}(\underline{\vartheta}_c) = L_n^{[c]}(\underline{\vartheta}_c) + \mathsf{E}\left[\log f_{\underline{Z}_c}\left(\underline{z}_{c,n}|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c^{\,n}\right)|\underline{x}_n, \underline{\vartheta}^n\right] \quad (c = 1, \ldots, C). \quad (12)$$

The new estimate $\underline{\vartheta}^{n+1}$ is obtained by the following procedure.

For $c = 1, \ldots, C$

$$\underline{\vartheta}_c^{n+1} = \arg\max_{\underline{\vartheta}_c} L_{n+1}^{[c]}(\underline{\vartheta}_c) \quad (13)$$

end

Finally,

$$\underline{\vartheta}^{n+1} = \left[\underline{\vartheta}_1^{n+1}, \ldots, \underline{\vartheta}_C^{n+1}\right]. \quad (14)$$

To obtain a proper approximation of $L_{n+1}^{[c]}(\underline{\vartheta}_c)$, we will consider the Taylor expansion of the first and second terms on the right-hand side of (12). Approximately

$$L_n^{[c]}(\underline{\vartheta}_c) = L_n^{[c]}(\underline{\vartheta}_c^n) + \frac{1}{2}(\underline{\vartheta}_c - \underline{\vartheta}_c^n)^T \times \left[\nabla_{\underline{\vartheta}_c} \nabla_{\underline{\vartheta}_c}^T L_n^{[c]}(\underline{\vartheta}_c)\right]_{\underline{\vartheta}=\underline{\vartheta}^n} (\underline{\vartheta}_c - \underline{\vartheta}_c^n). \quad (15)$$

The first-order term vanishes because $\nabla_{\underline{\vartheta}_c} L_n^{[c]}(\underline{\vartheta}_c)|_{\underline{\vartheta}=\underline{\vartheta}^n} = \underline{0}$. By the definition of $L_n^{[c]}(\underline{\vartheta}_c)$, $[\nabla_{\underline{\vartheta}_c} \nabla_{\underline{\vartheta}_c}^T L_n^{[c]}(\underline{\vartheta}_c)]_{\underline{\vartheta}=\underline{\vartheta}^n}$ is approximately given by

$$\left[\nabla_{\underline{\vartheta}_c} \nabla_{\underline{\vartheta}_c}^T L_n^{[c]}(\underline{\vartheta}_c)\right]_{\underline{\vartheta}=\underline{\vartheta}^n} \approx -(n-1)\mathcal{I}_{\mathrm{SAGE}}^{[c]}(\underline{\vartheta}^n). \quad (16)$$

Let $f_{\underline{Z}_c|\underline{X}}(\cdot)$ and $f_{\underline{X}|\underline{Z}_c}(\cdot)$ denote the conditional density functions. Using Bayes' law, the augmented data log-likelihood $\log f_{\underline{Z}_c}(\underline{z}_{c,n}|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)$ can be expressed as

$$\log f_{\underline{Z}_c}(\underline{z}_{c,n}|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c) = \log f_{\underline{X}}(\underline{x}_n|\underline{\vartheta}) + \log f_{\underline{Z}_c|\underline{X}} \times (\underline{z}_{c,n}|\underline{x}_n, \underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c) - \log f_{\underline{X}|\underline{Z}_c}(\underline{x}_n|\underline{z}_{c,n}, \underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c). \quad (17)$$

From [12], we know that the last term of (17) must be independent of $\underline{\vartheta}_c$ to assure the monotonic increase of the observed data log-likelihood. Thus, in the cth cycle, $\log f_{\underline{X}|\underline{Z}_c}(\underline{x}_n|\underline{z}_{c,n}, \underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)$ is regarded as a constant. Approximately, (17) is given by

$$\begin{aligned}
&\log f_{\underline{Z}_c}\left(\underline{z}_{c,n}|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c^{\,n}\right) \\
&\approx \text{const} + \log f_{\underline{Z}_c}\left(\underline{z}_{c,n}|\underline{\vartheta}_c^n, \underline{\tilde{\vartheta}}_c^{\,n}\right) \\
&\quad + \left[\nabla_{\underline{\vartheta}_c} \log f_{\underline{X}}(\underline{x}_n|\underline{\vartheta})\right]_{\underline{\vartheta}=\underline{\vartheta}^n}^T (\underline{\vartheta}_c - \underline{\vartheta}_c^n) \\
&\quad + \left[\nabla_{\underline{\vartheta}_c} \log f_{\underline{Z}_c|\underline{X}}(\underline{z}_{c,n}|\underline{x}_n, \underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)\right]_{\underline{\vartheta}=\underline{\vartheta}^n}^T (\underline{\vartheta}_c - \underline{\vartheta}_c^n) \\
&\quad + \frac{1}{2}(\underline{\vartheta}_c - \underline{\vartheta}_c^n)^T \left[\nabla_{\underline{\vartheta}_c} \nabla_{\underline{\vartheta}_c}^T \log f_{\underline{Z}_c}(\underline{z}_{c,n}|\underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)\right]_{\underline{\vartheta}=\underline{\vartheta}^n} \\
&\quad \times (\underline{\vartheta}_c - \underline{\vartheta}_c^n). \quad (18)
\end{aligned}$$

Given appropriate regularity

$$\mathsf{E}\left[\nabla_{\underline{\vartheta}_c} \log f_{\underline{Z}_c|\underline{X}}(\underline{z}_{c,n}|\underline{x}_n, \underline{\vartheta}_c, \underline{\tilde{\vartheta}}_c)|\underline{x}_n, \underline{\vartheta}^n\right] = \underline{0} \quad (19)$$

equations (15), (16), and (18) lead to the following expression:

$$\begin{aligned}
L_{n+1}^{[c]}(\underline{\vartheta}_c) \approx &\text{const} + L_n^{[c]}(\underline{\vartheta}_c^n) \\
&+ \mathsf{E}\left[\log f_{\underline{Z}_c}\left(\underline{z}_{c,n}|\underline{\vartheta}_c^n, \underline{\tilde{\vartheta}}_c^{\,n}\right)|\underline{x}_n, \underline{\vartheta}^n\right] \\
&+ \left[\nabla_{\underline{\vartheta}_c} \log f_{\underline{X}}(\underline{x}_n|\underline{\vartheta})\right]_{\underline{\vartheta}=\underline{\vartheta}^n}^T (\underline{\vartheta}_c - \underline{\vartheta}_c^n) \\
&- \frac{n}{2}(\underline{\vartheta}_c - \underline{\vartheta}_c^n)^T \mathcal{I}_{\mathrm{SAGE}}^{[c]}(\underline{\vartheta}^n)(\underline{\vartheta}_c - \underline{\vartheta}_c^n). \quad (20)
\end{aligned}$$

The maximizing parameter $\underline{\vartheta}_c^{n+1}$ is given by

$$\underline{\vartheta}_c^{n+1} = \underline{\vartheta}_c^n + \frac{1}{n}\mathcal{I}_{\mathrm{SAGE}}^{[c]}(\underline{\vartheta}^n)^{-1}\left[\nabla_{\underline{\vartheta}_c} \log f_{\underline{X}}(\underline{x}_n|\underline{\vartheta})\right]_{\underline{\vartheta}=\underline{\vartheta}^n}. \quad (21)$$

As the parameter subsets are disjoint, (21) can be implemented simultaneously for $c = 1, \ldots, C$. This implies

$$\underline{\vartheta}^{n+1} = \underline{\vartheta}^n + \frac{1}{n}\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n). \quad (22)$$

Thus, given regularity conditions, recursions (13) and (14) can be approximated by (22). The iterates of the recursive SAGE-inspired algorithm are given by (22). The recursive procedures (5) and (8) are more general forms of (22).

IV. Asymptotic Properties

Two questions of particular importance about the recursive EM and SAGE-inspired algorithms are whether the sequence $\{\underline{\vartheta}^n\}$ generated by (1), (4), (5), or (8) converges and how fast it converges. We will show that both algorithms are indeed stochastic approximation procedures with a specialized gain matrix for minimizing the Kullback–Leibler distance. The asymptotic behavior of the iterates is governed by a mean ordinary differential equation (ODE). Based on the stochastic approximation theory, we will show that under mild conditions, $\underline{\vartheta}^n$ is strongly consistent and asymptotically normally distributed.

To begin with, define

$$\underline{\Gamma}(\underline{\vartheta}) = \mathsf{E}_{\underline{\vartheta}^*}\left[\underline{\gamma}(\underline{x}, \underline{\vartheta})\right] \tag{23}$$

where the expectation is with respect to $f_{\underline{X}}(\underline{x}|\underline{\vartheta}^*)$ and $\underline{\gamma}(\underline{x}, \underline{\vartheta}) = \underline{\nabla}_{\underline{\vartheta}} \log f_{\underline{X}}(\underline{x}|\underline{\vartheta})$. We cite a theorem from [26].

Theorem 1: $\underline{\Gamma}(\underline{\vartheta})$ has the following properties: a) $\underline{\Gamma}(\underline{\vartheta}^*) = \underline{0}$. b) There exists Θ', a neighborhood of $\underline{\vartheta}^*$, such that $\sup_{\underline{\vartheta}}(\underline{\vartheta} - \underline{\vartheta}^*)^T \underline{\Gamma}(\underline{\vartheta}) < 0$ for $\underline{\vartheta} \in \Theta'$, $\underline{\vartheta} \neq \underline{\vartheta}^*$.

Proof: See Appendix A.

Now consider

$$\underline{g}_1(\underline{\vartheta}) = \mathcal{I}_{\mathrm{EM}}(\underline{\vartheta})^{-1}\underline{\Gamma}(\underline{\vartheta}) \quad \text{and} \quad \underline{g}_2(\underline{\vartheta}) = \mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta})^{-1}\underline{\Gamma}(\underline{\vartheta}). \tag{24}$$

Then

$$\underline{g}_1(\underline{\vartheta}^*) = \underline{0} \quad \text{and} \quad \underline{g}_2(\underline{\vartheta}^*) = \underline{0} \tag{25}$$

is a direct consequence of *Theorem 1*. It is clear that recursions (1), (4),(5), and (8) are procedures to find zeros of $\underline{g}_1(\underline{\vartheta})$ and $\underline{g}_2(\underline{\vartheta})$, respectively. Furthermore, the observation of $\underline{g}_1(\underline{\vartheta})$

$$\underline{Y}_n = \mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n) \tag{26}$$

can be decomposed as

$$\underline{Y}_n = \underline{g}_1(\underline{\vartheta}^n) + \delta\underline{M}_n \tag{27}$$

where $\delta\underline{M}_n$ is a martingale difference noise [17], [27]. Note that $\underline{Y}_n$ can be written as

$$\underline{Y}_n = \mathsf{E}[\underline{Y}_n|\underline{Y}_i, i < n] + (\underline{Y}_n - \mathsf{E}[\underline{Y}_n|\underline{Y}_i, i < n]). \tag{28}$$

The independence of $\underline{x}_1, \underline{x}_2, \ldots$ yields the following:

$$\mathsf{E}[\underline{Y}_n|\underline{Y}_i, i < n] = \underline{g}_1(\underline{\vartheta}^n) \tag{29}$$

and

$$\underline{M}_n = \sum_{j=1}^{n} \left(\underline{Y}_j - \mathsf{E}[\underline{Y}_j|\underline{Y}_i, i < j]\right) \tag{30}$$

where $\underline{M}_n$ is a martingale [27]. Then, (27) follows immediately. Similarly, in the recursive SAGE-inspired algorithm, $\underline{Y}_n = \mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ can be decomposed as

$$\underline{Y}_n = \underline{g}_2(\underline{\vartheta}^n) + \delta\underline{M}_n. \tag{31}$$

A. Convergence

From the previous discussion, we know that $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)^{-1}$ $\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ and $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ are unbiased observations of the functions $\underline{g}_1(\underline{\vartheta})$ and $\underline{g}_2(\underline{\vartheta})$ corrupted by a martingale difference noise. By probability inequalities for martingale sequences, it can be shown that the asymptotic behavior of the iterate $\underline{\vartheta}^n$ is determined effectively by that of a "mean" ODE. Since $\underline{\Gamma}(\underline{\vartheta})$ is a gradient and $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta})$ [or $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta})$] is positive definite, the stationary points of the ODE associated with the recursive EM and SAGE-inspired algorithms are asymptotically stable. If the initial estimate $\underline{\vartheta}^0$ is not too far from $\underline{\vartheta}^*$, we can expect that $\underline{\vartheta}^n$ converges to $\underline{\vartheta}^*$.

In order to describe the behavior of the constrained algorithm, it is necessary to define the projected ODE [17]

$$\dot{\underline{\vartheta}}(t) = \underline{g}_i\left(\underline{\vartheta}(t)\right) + \underline{z}(t), \quad i = 1, 2 \tag{32}$$

where $\underline{z}(t)$ is the projection term, the minimum term needed to keep $\underline{\vartheta}(t)$ in the constraint set Θ. $\underline{z}(t)$ is zero if $\underline{\vartheta}(t) \in \Theta$. Assume $\Theta = [a_1, b_1] \times \ldots \times [a_d, b_d]$. If $\underline{\vartheta}(t)$ is on the interior of a face of Θ and $\underline{g}_i(\underline{\vartheta}(t))$ points out of Θ, then $\underline{z}(t)$ points inward, orthogonal to the face. Note that $||\underline{z}(t)|| \leq ||\underline{g}_i(\underline{\vartheta}(t))||$ [17], where $\|\cdot\|$ denotes the Euclidean norm.

The probability one convergence of recursions (1), (4), (5), and (8) follows directly from convergence of a more general procedure

$$\underline{\vartheta}^{n+1} = \underline{\vartheta}^n + an^{-\alpha}\mathbf{K}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n) \tag{33}$$

in the unconstrained case, and

$$\underline{\vartheta}^{n+1} = \Pi_\Theta\left[\underline{\vartheta}^n + an^{-\alpha}\mathbf{K}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right] \tag{34}$$

in the constrained case. We assume $\mathbf{K}(\cdot)^{-1}$ is a positive definite matrix and continuous in $\underline{\vartheta}$. Applying the ODE method widely used in the stochastic approximation theory [2], [17], we obtain the following results.

Theorem 2: Suppose a) $\mathbf{0} < \mathbf{K}(\underline{\vartheta}^n) < \infty$ and b) $\mathsf{E}[||\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)||^2] < \infty$ hold for the recursions (33) and (34). Then, $\{\underline{\vartheta}^n\}$ generated by the constrained algorithm (4) converges with probability one to some invariant set of the ODE

$$\dot{\underline{\vartheta}}(t) = \underline{g}\left(\underline{\vartheta}(t)\right) + \underline{z}(t) \tag{35}$$

in Θ, where $\underline{g}(\underline{\vartheta}) = \mathbf{K}(\underline{\vartheta})^{-1}\underline{\Gamma}(\underline{\vartheta})$. For the unconstrained algorithm (1), we assume additionally that $\{\underline{\vartheta}^n\}$ is bounded with probability one. Then, $\{\underline{\vartheta}^n\}$ converges with probability one to the invariant set of the ODE

$$\dot{\underline{\vartheta}}(t) = \underline{g}\left(\underline{\vartheta}(t)\right). \tag{36}$$

Furthermore, $\{\underline{\vartheta}^n\}$ converges with probability one to the set of stationary points of the corresponding ODE.

Proof: See Appendix B.

Remark: The result of *Theorem 2* holds for the recursive EM and SAGE-inspired algorithms if $\mathbf{K}(\underline{\vartheta}^n)$ is replaced by $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ or $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$, and $\underline{g}(\underline{\vartheta})$ is replaced by $\underline{g}_1(\underline{\vartheta})$ or $\underline{g}_2(\underline{\vartheta})$.

The previous theorem shows us that under proper conditions, $\underline{\vartheta}^n$ converges with probability one to the stationary points. For the unconstrained algorithm, the stationary points are points such that $\underline{g}(\underline{\vartheta}) = \underline{0}$. For the constrained algorithm, if the stationary points lie in Θ, the stationary condition is just $\underline{g}(\underline{\vartheta}) = \underline{0}$. On the boundary, the condition is $\underline{g}(\underline{\vartheta})+\underline{z} = \underline{0}$. As pointed out in [17], the constraint Θ can give rise to spurious stationary points on the boundary, but this is the only type of singular points that can be introduced by the constraint.

Now, assume that the constraint set is large enough, so that the stationary points are inside the Θ. From *Theorem 1*, we know the true parameter vector $\underline{\vartheta}^*$ satisfies the stationary condition and is one of the limit points of $\underline{\vartheta}^n$. In view of the well-known multiple maxima that are possible on likelihood surfaces, one could, of course, not expect consistency of the recursions irrespective of the starting point [25]. In other words, the iterate $\underline{\vartheta}^n$ converges to $\underline{\vartheta}^*$ if the initial estimate $\underline{\vartheta}^0$ is in the domain of attraction of $\underline{\vartheta}^*$.

B. Asymptotic Distribution

In this subsection, we will be concerned with the convergence speed of the algorithm. The convergence speed is measured by the convergence rate, defined as the normalized error $(\underline{\vartheta}^n - \bar{\underline{\vartheta}})/\sqrt{\epsilon_n}$ around the limit point $\bar{\underline{\vartheta}}$ with $\epsilon_n = an^{-\alpha}$. Under assumptions made in *Theorem 2*, it can be shown that $(\underline{\vartheta}^n - \bar{\underline{\vartheta}})/\sqrt{\epsilon_n}$ is normally distributed with zero mean and covariance matrix $\mathbf{V}$. To avoid mathematical difficulties, the limit point $\bar{\underline{\vartheta}}$ is supposed to be interior to the constrained set Θ in the constrained algorithm.

First, we consider the following approximate expression.

Lemma 1: The function $\underline{g}(\underline{\vartheta})$ around the limit point $\bar{\underline{\vartheta}}$ can be approximated by

$$\underline{g}(\underline{\vartheta}) = \mathbf{g}(\bar{\underline{\vartheta}})(\underline{\vartheta} - \bar{\underline{\vartheta}}) + o\left(\|\underline{\vartheta} - \bar{\underline{\vartheta}}\|\right) \tag{37}$$

where $\mathbf{g}(\bar{\underline{\vartheta}}) = -\mathbf{K}(\bar{\underline{\vartheta}})^{-1}\mathcal{I}(\bar{\underline{\vartheta}})$ is a stable matrix. A stable matrix is a square matrix, all of whose eigenvalues have negative real parts. The information matrix of the observed data is $\mathcal{I}(\underline{\vartheta}) = \mathsf{E}[-\underline{\nabla}_{\underline{\vartheta}}\underline{\nabla}_{\underline{\vartheta}}^T \log f_{\underline{X}}(\underline{x}|\underline{\vartheta})]$.

Proof: See Appendix C.

Exploiting results of stochastic approximation theory [2], [17], [20], one can obtain the asymptotic normality of the recursive EM and SAGE-inspired algorithms. The following theorem shows how the rate of convergence depends on the choice of α and the object function $\underline{g}(\underline{\vartheta})$.

Theorem 3: Consider the recursions (33) and (34). Suppose a) $\mathbf{0} < \mathbf{K}(\underline{\vartheta}^n) < \infty$, b) $\mathsf{E}[\|\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\|^2] < \infty$, and c) $\underline{\vartheta}^n$ converges to an isolated stable point $\bar{\underline{\vartheta}}$ of the ODE $\dot{\underline{\vartheta}}(t) = \underline{g}(\underline{\vartheta}(t))$.

Then, i) if $\alpha = 1$ and $\mathbf{D} = (1/2)\mathbf{I} - a\mathbf{K}(\bar{\underline{\vartheta}})^{-1}\mathcal{I}(\bar{\underline{\vartheta}})$ is a stable matrix, $n^{1/2}(\underline{\vartheta}^n - \bar{\underline{\vartheta}})$ has asymptotic distribution $\mathcal{N}(\underline{0}, \mathbf{V})$, where $\mathbf{V}$ is the solution of

$$\left(a\mathbf{A} + \frac{1}{2}\mathbf{I}\right)\mathbf{V} + \mathbf{V}\left(a\mathbf{A} + \frac{1}{2}\mathbf{I}\right)^T = -a^2\mathbf{C} \tag{38}$$

where $\mathbf{A} = -\mathbf{K}(\bar{\underline{\vartheta}})^{-1}\mathcal{I}(\bar{\underline{\vartheta}})$, $\mathbf{C} = \mathbf{K}(\bar{\underline{\vartheta}})^{-1}\mathcal{I}(\bar{\underline{\vartheta}})\mathbf{K}(\bar{\underline{\vartheta}})^{-1}$.

ii) If $1/2 < \alpha < 1$, $n^{\alpha/2}(\underline{\vartheta}^n - \bar{\underline{\vartheta}})$ has asymptotic distribution $\mathcal{N}(\underline{0}, \mathbf{V})$, where $\mathbf{V}$ is the solution of

$$\mathbf{AV} + \mathbf{VA} = -a\mathbf{C}. \tag{39}$$

Proof: See Appendix D.

Remark: The results of *Theorem 3* hold for the recursive EM and SAGE-inspired algorithms if $\mathbf{K}(\bar{\underline{\vartheta}})$ is replaced by $\mathcal{I}_{\mathrm{EM}}(\bar{\underline{\vartheta}})$ or $\mathcal{I}_{\mathrm{SAGE}}(\bar{\underline{\vartheta}})$.

From *Theorem 3*, we know that as $n \to \infty$, the expectation of $\|\underline{\vartheta}^n - \bar{\underline{\vartheta}}\|^2$ decreases with the order $O(n^{-1})$ for $\alpha = 1$ and $O(n^{-\alpha})$ for $1/2 < \alpha < 1$. Consequently, the best choice is $\alpha = 1$. However, in practice, selecting $\alpha = 1$ may cause too small step sizes at the initial stages, and the algorithm may not provide sufficient changes in $\underline{\vartheta}^n$ to approach the true parameter. Furthermore, when the gain matrix $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)^{-1}$ or $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}$ is replaced by $\mathcal{I}(\underline{\vartheta}^n)^{-1}$, the recursion

$$\underline{\vartheta}^{n+1} = \Pi_\Theta\left[\underline{\vartheta}^n + n^{-1}\mathcal{I}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right] \tag{40}$$

yields the asymptotically optimal covariance $\mathbf{V} = \mathcal{I}(\underline{\vartheta})^{-1}$. Thus, the estimates generated by (40) is asymptotically efficient in the sense that the covariance matrix approaches the lower bound given by the asymptotic Cramér–Rao inequality [11]. However, the augmented information matrices $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta})$ and $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta})$ are, in general, much easier to calculate and invert than $\mathcal{I}(\underline{\vartheta})$. Besides, an optimal rate of convergence can also be achieved by the recursive EM or SAGE-inspired algorithms if an additional averaging step is undertaken

$$\bar{\underline{\vartheta}}^n = \frac{1}{n}\sum_{i=1}^{n}\underline{\vartheta}^i \tag{41}$$

which is known as the Polyak–Ruppert procedure [21]. More details about this method can be found in [17] and [20].

V. Application to DOA Estimation

The recursive EM and SAGE-inspired algorithms introduced previously are applied to the source localization problem. A narrow-band signal model is used in the derivation. As the proposed approach is based on the likelihood function, extension to the wideband case is straightforward if one applies the asymptotic properties of Fourier transformation to the data [3], [4]. The algorithms presented here are aimed to find a fixed parameter. However, if a properly chosen small constant step size is used, the same procedures can be also used to track slowly varying parameters [2]. In the following, we give a brief description of the signal model and develop recursive methods for estimating DOA parameters. Also, we will show that the proposed algorithms are closely related to maximization of the concentrated likelihood function.

A. Signal Model

Consider an array of L sensors receiving signals generated by M far-field narrow-band sources. The nth snapshot of the array output $\underline{X}(n)$ is given by

$$\underline{X}(n) = \mathbf{H}(\underline{\theta})\underline{S}(n) + \underline{U}(n), \quad n = 1, 2, \ldots \tag{42}$$

where $\underline{\theta} = [\theta_1, \ldots, \theta_M]^T$ summarizes the DOA parameters, $\mathbf{H}(\underline{\theta}) = [\underline{d}(\theta_1), \ldots, \underline{d}(\theta_M)] \in \mathbb{C}^{L \times M}$ contains the steering vectors $\underline{d}(\theta_m)$, $(m = 1, \ldots, M)$, $\underline{S}(n) = [S_1(n), \ldots, S_M(n)]^T \in \mathbb{C}^{M \times 1}$, and $\underline{U}(n) \in \mathbb{C}^{L \times 1}$ denote the signal waveform and noise vector, respectively. We assume 1) $\underline{S}(n)$ is unknown and deterministic, 2) $\underline{U}(n)$ is independent, identically complex normally distributed with zero mean and covariance matrix $\nu\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. Consequently, $\underline{X}(n)$ is independent, complex normally distributed with mean $\mathbf{H}(\underline{\theta})\underline{S}(n)$ and covariance matrix $\nu\mathbf{I}$. The log-likelihood function of $\underline{x}_n = \underline{X}(n)$ conditioned on the signal waveform $\underline{S}(n)$ can be written as

$$\log f_{\underline{X}}(\underline{x}_n|\underline{\vartheta}) = -\Big[N\log\pi + N\log\nu + \frac{1}{\nu}\left(\underline{X}(n) - \mathbf{H}(\underline{\theta})\underline{S}(n)\right)^H \left(\underline{X}(n) - \mathbf{H}(\underline{\theta})\underline{S}(n)\right)\Big] \quad (43)$$

where $\underline{\vartheta} = [\underline{\theta}^T, \underline{S}(n)^T, \nu]^T$. The problem of interest is to estimate $\underline{\theta}$ sequentially.

B. Recursive Estimation of DOA Parameters

Based on the recursive EM and SAGE-inspired algorithms presented in Sections II and III, we develop a recursive procedure for estimating DOA parameters. The nth snapshot $\underline{X}(n)$ corresponds to $\underline{x}_n$, the observed data at time n. According to recursion (1), (4), (5), or (8), all elements in $\underline{\vartheta}$ should be updated simultaneously. However, to avoid a complicated gain matrix $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ or $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$, these procedures will only be applied to $\underline{\theta}$. The signal and noise parameters, denoted by $\underline{S}^n = [S_1^n, \ldots, S_M^n]$ and ν^n, are updated by computing their ML estimates once the new DOA parameters $\underline{\theta}^n$ are available. Thus, the gradient vector $\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ and the augmented information matrices $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n), \mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$ are calculated with respect to $\underline{\theta}$. Note that $\underline{\vartheta}^n = [\underline{\theta}^n, \underline{S}^n, \nu^n]$.

Taking the first derivative of the log-likelihood (43) with respect to θ_m, we can easily obtain the mth element of the gradient vector $\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ at time instant n

$$\left[\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right]_m = \frac{2}{\nu^n}\mathrm{Re}\left[\left(\underline{X}(n) - \mathbf{H}(\underline{\theta}^n)\underline{S}^n\right)^H \left(\underline{d}'(\theta_m^n)\, S_m^n\right)\right] \quad (44)$$

where $\underline{d}'(\theta_m) = \partial\underline{d}(\theta_m)/\partial\theta_m$.

The data augmentation scheme used for computing $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ is obtained by decomposing the array output into its signal and noise parts. Formally, it can be expressed as

$$\underline{Y}(n) = \left[\underline{Y}_1(n)^T, \ldots, \underline{Y}_m(n)^T, \ldots, \underline{Y}_M(n)^T\right]^T. \quad (45)$$

The augmented data associated with the mth signal

$$\underline{Y}_m(n) = \underline{d}(\theta_m)S_m(n) + \underline{U}_m(n) \quad (46)$$

is complex normally distributed with mean $\underline{d}(\theta_m)S_m(n)$ and covariance matrix $\nu_m\mathbf{I}$ with the constraint $\sum_{m=1}^M \nu_m = \nu$. For simplicity, we choose $\nu_m = \nu/M$. The log-likelihood associated with this augmentation scheme is given by

$$\log f_{\underline{Y}}(\underline{y}|\underline{\vartheta}) = -\sum_{m=1}^{M}\Big[N\log\pi + N\log\Big(\frac{\nu}{M}\Big) + \frac{M}{\nu}\times\left(\underline{Y}_m(n) - \underline{d}(\theta_m)S_m(n)\right)^H\left(\underline{Y}_m(n) - \underline{d}(\theta_m)S_m(n)\right)\Big]. \quad (47)$$

Since the signals are decoupled through the augmentation scheme (45), $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ is a $M \times M$ diagonal matrix when we only consider the DOA parameters $\underline{\theta}$. By definition (3), the mth diagonal element of $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ is the conditional expectation of the second derivative of the augmented log-likelihood

$$\left[\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)\right]_{mm} = \mathsf{E}\left[-\frac{\partial^2}{\partial\theta_m^2}\log f_{\underline{Y}}(\underline{y}_n|\underline{\vartheta})|\underline{x}_n, \underline{\vartheta}^n\right] \quad (48)$$

which is given by

$$\left[\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)\right]_{mm} = \frac{2}{\nu^n}\mathrm{Re}\Big[-\left(\underline{d}''(\theta_m^n)\, S_m^n\right)^H \times \left(\underline{X}(n) - \mathbf{H}(\underline{\theta}^n)\underline{S}^n\right) + M\left\|\underline{d}'(\theta_m^n)\, S_m^n\right\|^2\Big] \quad (49)$$

where $\underline{d}''(\theta_m) = \partial^2\underline{d}(\theta_m)/\partial\theta_m^2$.

In the recursive SAGE-inspired algorithm, the augmented data of the mth cycle is specified as

$$\underline{Z}_m(n) = \underline{d}(\theta_m)S_m(n) + \underline{U}(n), \quad (m = 1, \ldots, M). \quad (50)$$

At each cycle, we consider one signal and the total noise component. The associated log-likelihood is given by

$$\log f_{\underline{Z}_m}(\underline{z}_m|\underline{\vartheta}) = -\Big[N\log\pi + N\log\nu + \frac{1}{\nu}\times\left(\underline{Z}_m(n) - \underline{d}(\theta_m)S_m(n)\right)^H\left(\underline{Z}_m(n) - \underline{d}(\theta_m)S_m(n)\right)\Big]. \quad (51)$$

As the recursive SAGE is only applied to the DOA parameter, $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$ is a diagonal matrix. By (6) and (51), $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$, the mth diagonal element can be expressed as

$$\left[\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)\right]_{mm} = \frac{2}{\nu^n}\mathrm{Re}\Big[-\left(\underline{d}''(\theta_m^n)\, S_m^n\right)^H \times \left(\underline{X}(n) - \mathbf{H}(\underline{\theta}^n)\underline{S}^n\right) + \left\|\underline{d}'(\theta_m^n)\, S_m^n\right\|^2\Big]. \quad (52)$$

More details about calculating $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ and $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$ can be found in [6].

For both algorithms, once the estimate $\underline{\theta}^n$ is available, the signal and noise parameters are obtained by computing their ML estimates for given $\underline{\theta}^n$ and $\underline{X}(n)$ as follows:

$$\underline{S}^n = \mathbf{H}(\underline{\theta}^n)^{\#}\underline{X}(n) \quad (53)$$

$$\nu^n = \frac{1}{L}\mathrm{tr}\left[\mathbf{P}^{\perp}(\underline{\theta}^n)\widehat{\mathbf{C}}_{\underline{X}}(n)\right] \quad (54)$$

where $\mathbf{H}(\underline{\theta}^n)^{\#}$ is the generalized left inverse of the matrix $\mathbf{H}(\underline{\theta}^n)$, $\mathbf{P}^{\perp}(\underline{\theta}^n) = \mathbf{I} - \mathbf{P}(\underline{\theta}^n)$ is the orthogonal complement of the projection matrix $\mathbf{P}(\underline{\theta}^n) = \mathbf{H}(\underline{\theta}^n)\mathbf{H}(\underline{\theta}^n)^{\#}$, and $\widehat{\mathbf{C}}_{\underline{X}}(n) = \underline{X}(n)\underline{X}(n)^H$.

At beginning of the recursion, the initial estimate $\underline{\theta}^0$ may be obtained by beamforming or subspace methods. The signal and noise parameters $\underline{S}^0$ and ν^0 can be initialized by (53) and (54), respectively. In summary, the recursive EM (or SAGE-inspired) algorithm proceeds as follows.

Initialize $\underline{\vartheta}^0$
For $n = 1, 2, \ldots$

Calculate gradient vector and augmented information matrix by equations (44) and (49) [or (52)].
Update the DOA parameters by

$$\underline{\theta}^{n+1} = \underline{\theta}^n + an^{-\alpha}\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n) \qquad (55)$$

or

$$\underline{\theta}^{n+1} = \underline{\theta}^n + an^{-\alpha}\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n) \qquad (56)$$

in the unconstrained case. Use the projection operator Π_{Θ} in the constrained case.
Update the signal and noise parameters by equations (53) and (54).

end

Because of the simple structure of the augmented data, $\mathcal{I}_{\mathrm{EM}}(\underline{\theta}^n)$ and $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$ are $M \times M$ diagonal matrices. The associated inverse matrices can be easily obtained by inverting the diagonal elements of $\mathcal{I}_{\mathrm{EM}}(\underline{\theta}^n)$ and $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$. Compared to the optimal procedure (40) with the gain matrix

$$\mathcal{I}(\underline{\vartheta}^n)^{-1} = \left[\frac{2}{\nu^n}\mathrm{Re}\left(\mathbf{S}(n)^H\mathbf{D}(\underline{\theta}^n)^H\mathbf{D}(\underline{\theta}^n)\mathbf{S}(n)\right)\right]^{-1} \qquad (57)$$

where $\mathbf{S}(n)$ is a diagonal matrix with elements of $\underline{S}^n$ and $\mathbf{D}(\underline{\theta}^n) = [\underline{d}(\theta_1)', \ldots, \underline{d}(\theta_M)]$ [23], recursions (55) and (56) are much easier to implement. Although the algorithms presented in this subsection are developed for a narrow band signal, they can be easily extended to the wide band case. From the asymptotic theory of Fourier transform, we know that the Fourier transformed data at each frequency bin is independent from each other [3], [4]. In this case, the contribution of each frequency bin can be calculated by (44), (49), and (52). The gradient vector and the augmented matrix are obtained by summing up contributions over frequencies of interest.

Comparing (49) and (52), $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ differs from $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$ only in the coefficient of the second term. When the iterate $\underline{\theta}^n$ is close to the true parameter $\underline{\theta}^*$, the first term can be neglected. The gain matrix of the recursive SAGE-inspired algorithm is M times that of the recursive EM algorithm. It can be expected that both algorithms have the same convergence behavior if the step size of the recursive EM algorithm is chosen to be M times that of the recursive SAGE-inspired algorithm.

C. Relation to the Concentrated Likelihood Function

One may notice that under the conditional signal model, the likelihood function (43) varies from snapshot to snapshot. The assumption made by the recursive EM and SAGE-inspired algorithms that all observations are i.i.d. is no longer satisfied. Do the procedures developed in the previous subsection really lead to the true DOA parameters? We shall show that the recursive EM and SAGE-inspired algorithms for DOA estimation do lead to the desired true parameters $\underline{\theta}^*$.

It is well known in the literature [3] that the log-likelihood function (43) can be concentrated with respect to the signal and noise parameters. For *one* snapshot, the concentrated likelihood is given by

$$L_{\underline{X}}(\underline{x}_n, \underline{\theta}) = -\log\mathrm{tr}\left[\mathbf{P}^{\perp}(\underline{\theta})\widehat{\mathbf{C}}_{\underline{X}}(n)\right] \qquad (58)$$

where

$$\hat{\mathbf{C}}_{\underline{X}}(n) = \underline{X}(n)\underline{X}(n)^H. \qquad (59)$$

Let

$$\underline{\eta}(\underline{x}_n, \underline{\theta}) = \nabla L_{\underline{X}}(\underline{x}_n, \underline{\theta}) \qquad (60)$$

be the gradient vector of $L_{\underline{X}}(\underline{x}_n, \underline{\theta})$. It will be shown in the following that the gradient vector $\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$, defined by (44), has the same direction as $\underline{\eta}(\underline{x}_n, \underline{\theta}^n)$.

Result 1: Let $\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ and $\underline{\eta}(\underline{x}_n, \underline{\theta})$ be defined by (44) and (60), respectively. Then

$$\underline{\eta}(\underline{x}_n, \underline{\vartheta}^n) = \frac{1}{L}\underline{\gamma}(\underline{x}_n, \underline{\theta}^n). \qquad (61)$$

Proof: See Appendix E.

To investigate the convergence behavior, we need to define a function closely related to the concentrated likelihood function

$$\mathcal{L}_{\underline{X}}(\underline{\theta}) = -\log\mathrm{tr}\left[\mathbf{P}(\underline{\theta})^{\perp}\mathbf{C}_{\underline{X}}\right] \qquad (62)$$

where

$$\mathbf{C}_{\underline{X}} = \mathsf{E}\left[\underline{X}(n)\underline{X}(n)^H\right] = \mathbf{H}(\underline{\theta}^*)\mathbf{C}_{\underline{S}}\mathbf{H}(\underline{\theta}^*)^H + \nu\mathbf{I}. \qquad (63)$$

Note that $\mathbf{C}_{\underline{X}}$ is the covariance matrix of array outputs when $\underline{S}(n)$ is assumed to be a zero-mean stationary process with covariance matrix $\mathbf{C}_{\underline{S}}$. The notation $\underline{\theta}^*$ in (63) is used to emphasize that $\mathbf{C}_{\underline{X}}$ is the true covariance matrix. Although the recursive DOA estimation algorithms were derived under the deterministic signal model, the dynamics of the algorithms are governed by $\mathcal{L}_{\underline{X}}(\underline{\theta})$, in which $\hat{\mathbf{C}}_{\underline{X}}(n)$ is replaced by $\mathbf{C}_{\underline{X}}$. Using (63) does not constitute conflicts with the deterministic signal model used in derivation. The source signals are a stationary process. When $\underline{S}(n)$ is seen as a realization, we have the deterministic signal model.

Result 2: The recursive EM and SAGE-inspired algorithms for DOA estimation are stochastic approximation procedures for finding zeros of

$$\underline{g}(\underline{\theta}) = \nabla_{\underline{\theta}}\mathcal{L}_{\underline{X}}(\underline{\theta}) \qquad (64)$$

which may coincide with the true parameters $\underline{\theta}^*$.

Proof: See Appendix F.

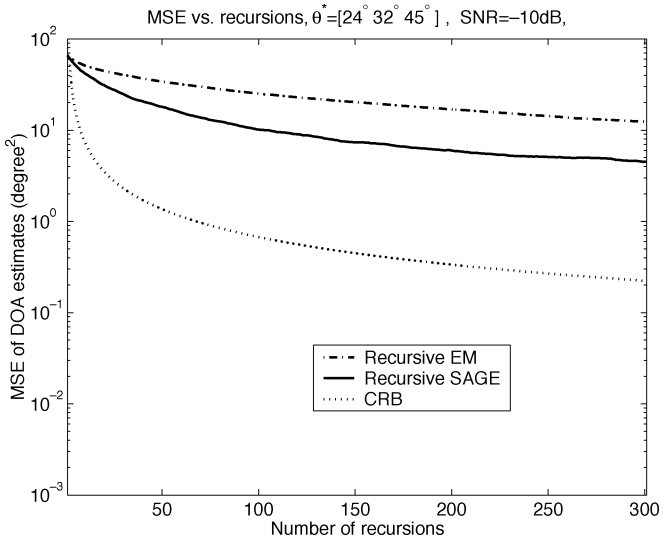

Fig. 1. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. SNR $= -10$ dB. $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

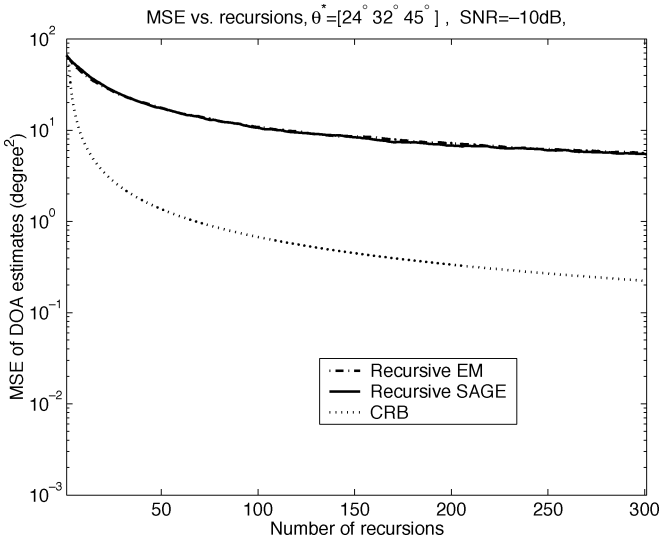

Fig. 2. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. SNR $= -10$ dB. $\epsilon_n^{\mathrm{REM}} = 3n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

VI. Simulations

We study the performance of the recursive EM and SAGE-inspired algorithms for DOA estimation by numerical experiments. A linear array of 15 sensors with equal interelement spacings of half a wavelength is used. Results from three numerical experiments will be presented.

In the first experiment, three sources of equal power located at $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$ are used to generate signals. The procedure starts from the initial value $\underline{\theta}^0 = [19^\circ\ 36^\circ\ 50^\circ]$. In the second experiment, a more critical situation $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$ with two closely located sources is considered. The initial estimate is given by $\underline{\theta}^0 = [19^\circ\ 32^\circ\ 50^\circ]$. The signal-to-noise ratio (SNR), defined as the ratio between the signal power of each source and sensor noise $10\log[|S_m(n)|^2/\nu]$, is chosen to be -10, 0, 10 dB. The maximum number of snapshots is 300. The step size for the recursive SAGE-inspired algorithm is chosen to be $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$. The recursive EM algorithm uses two step sizes, $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$ or $\epsilon_n^{\mathrm{REM}} = 3n^{-0.6}$. In the third experiment, we consider the scenario with $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$ and SNR $= 0$ dB. In addition to the recursive EM and SAGE-inspired algorithms, a procedure using the gain matrix $\mathcal{I}(\underline{\vartheta}^n)^{-1}$ is applied to the same data.

The number of Monte Carlo trials performed in each experiment is 500. The estimation accuracy is measured by the mean-squared error (MSE), which is defined as $\|\underline{\theta}^n - \underline{\theta}^*\|^2$. The results are compared to the Cramér–Rao bound (CRB) for stochastic signal model, which is the lowest bound for the ML estimator [24] and does not depend on realizations of the signals.

In order to avoid outliers, the recursion (55) or (56) is carried out when the change δ in each DOA parameter at each stage is less than a fixed value $\bar{\delta}$. In our simulation, $\bar{\delta}$ is chosen to be 2°. If $|\delta| > \bar{\delta}$, we set the change in the DOA parameter to be 0.001 with the sign of the corresponding element of the gradient vector. Using such a mechanism is equivalent to replacing $[\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)]_{mm}$ [or $[\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)]_{mm}$] obtained by (49) or (52) by a large positive number. The condition for probability one convergence $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n) < \infty$ [or $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n) < \infty$] is still satisfied. Also, the iterate $\underline{\theta}^n$ is kept bounded.

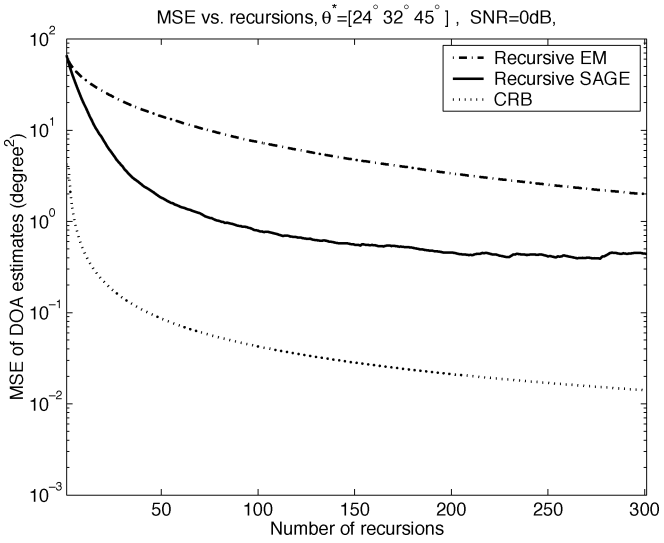

Fig. 3. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. SNR $= 0$ dB. $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

Results of the first experiment are plotted in Figs. 1–6. The MSEs of both algorithms decrease with increasing number of recursions. Figs. 1, 3, and 5 show that when the same step sizes are used, the recursive SAGE-inspired algorithm converges faster than the recursive EM algorithm. When the step size of the recursive EM algorithm is three times as much as that of the recursive SAGE-inspired algorithm, their convergence rates are almost identical (see Figs. 2, 4, and 6). At SNR $= 0$, 10 dB, the recursive EM algorithm has slightly lower MSE than the recursive SAGE-inspired algorithm. The results presented in Figs. 2, 4, and 6 are not surprising. Since the number of sources $M = 3$, when $\epsilon_n^{\mathrm{REM}} = M\epsilon_n^{\mathrm{RSAGE}}$, both algorithms should have similar convergence behavior as predicted by (49) and (52). Comparing MSEs at SNR $= -10$, 0, 10 dB, we can easily see that MSEs become smaller at higher SNRs.

Figs. 7–12 show the performance of both algorithms in a more critical scenario $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$. The distance of the initial estimate to the true parameters $\|\underline{\theta}^0 - \underline{\theta}^*\|^2$ is the same as that

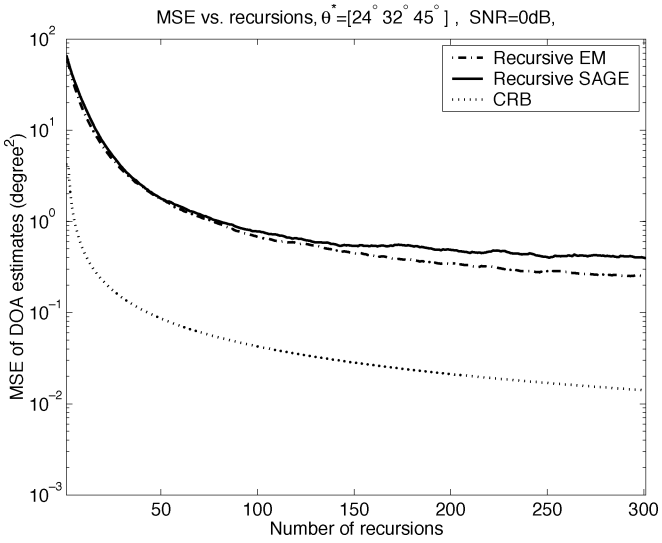

Fig. 4. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. SNR $= 0$ dB. $\epsilon_n^{\mathrm{REM}} = 3n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

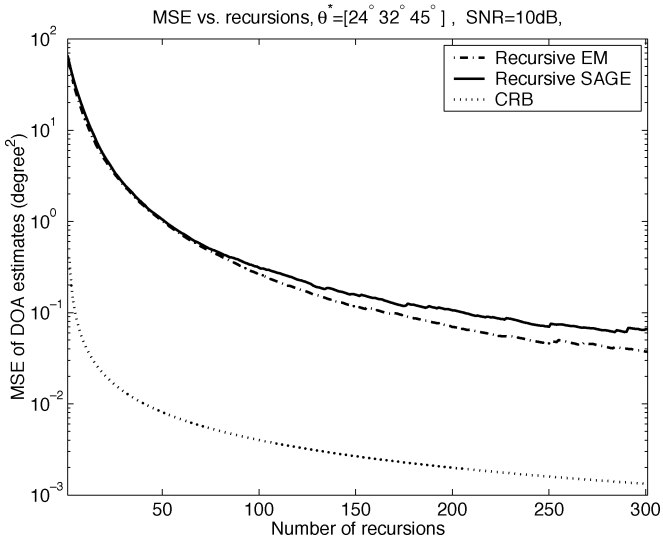

Fig. 6. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. SNR $= 10$ dB. $\epsilon_n^{\mathrm{REM}} = 3n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

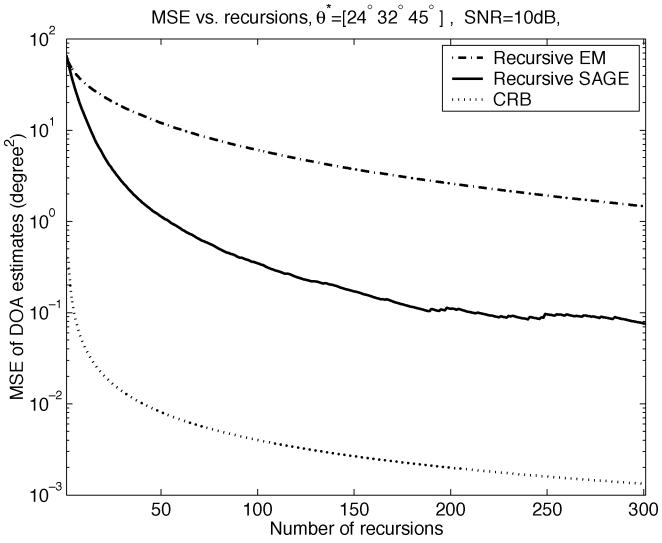

Fig. 5. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. SNR $= 10$ dB. $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

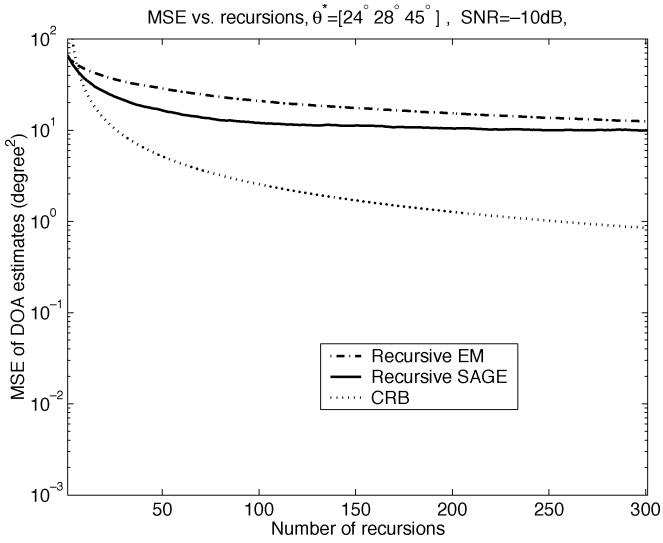

Fig. 7. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$. SNR $= -10$ dB. $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

in the first experiment. Both procedures lead to convergent iterates at all SNRs. At higher SNRs, the estimation accuracy is better than that at lower SNRs. In general, both algorithms perform worse than in the first experiment. It can be also observed that the MSEs are larger, and thus, the convergence rates are slower. Comparing Figs. 7, 9, and 11 with Figs. 1, 3, and 5, we can observe that the influence of the closely located sources become more significant at higher SNRs. In Figs. 8, 10, and 12, it can also be observed that both algorithms have a similar convergence behavior when $\epsilon_n^{\mathrm{REM}} = M\epsilon_n^{\mathrm{RSAGE}}$.

Three stochastic approximation procedures with gain matrices $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)^{-1}$, $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)^{-1}$, and $\mathcal{I}(\underline{\vartheta}^n)^{-1}$ are compared in the third experiment. The initial estimates are given by $\underline{\theta}^0 = [19^\circ\ 36^\circ\ 50^\circ]$ and $\underline{\theta}^0 = [22^\circ\ 36^\circ\ 47^\circ]$. From Fig. 13, we can observe that the algorithm with the same gain matrix as the optimal procedure (40) has a better convergence rate after the iterates are close enough to the true parameter. Note that using the gain matrix $\mathcal{I}(\underline{\vartheta}^n)^{-1}$ does not mean the optimal procedure because a step size of $n^{-0.6}$ rather than n^{-1} is used. So, we refer to this algorithm as the $\mathcal{I}$-procedure. As pointed out in [17], n^{-1} is the best step size for $n \to \infty$. In practice, using fast decreasing step size may not provide sufficient changes to approach to the true parameter. Fig. 14 presents results obtained by using a better initial estimate. The MSEs are, in general, lower than those in Fig. 13 for the same number of snapshots. This is not surprising since the algorithm requires fewer recursions to enter the neighborhood of the true parameter.

In summary, the estimates generated by the recursive EM and SAGE-inspired algorithms converge to the true parameters. Using the same step sizes, the recursive SAGE-inspired algorithm converges faster than the recursive EM algorithm. When $\epsilon_n^{\mathrm{REM}} = M\epsilon_n^{\mathrm{RSAGE}}$, they have similar convergence behavior. Both methods perform better at higher SNRs or in a situation with widely separated sources. The convergence rate can be improved by using the gain matrix $\mathcal{I}(\underline{\vartheta}^n)^{-1}$ at the expense of higher computational cost. The CRB computed for a stochastic

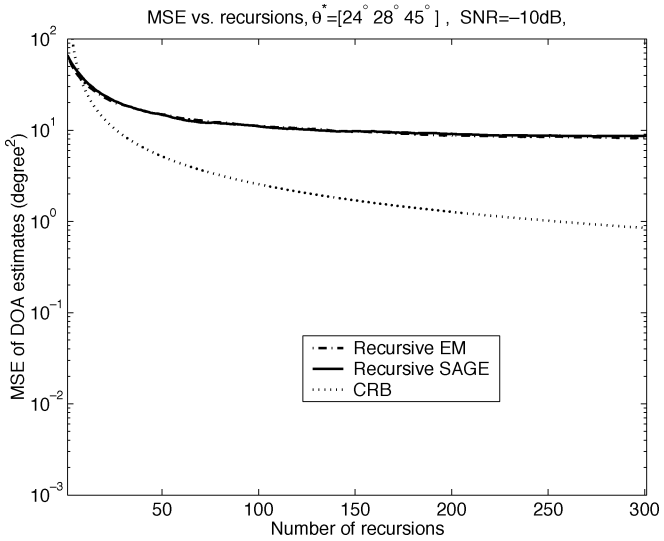

Fig. 8. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$. SNR $= -10$ dB. $\epsilon_n^{\mathrm{REM}} = 3n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

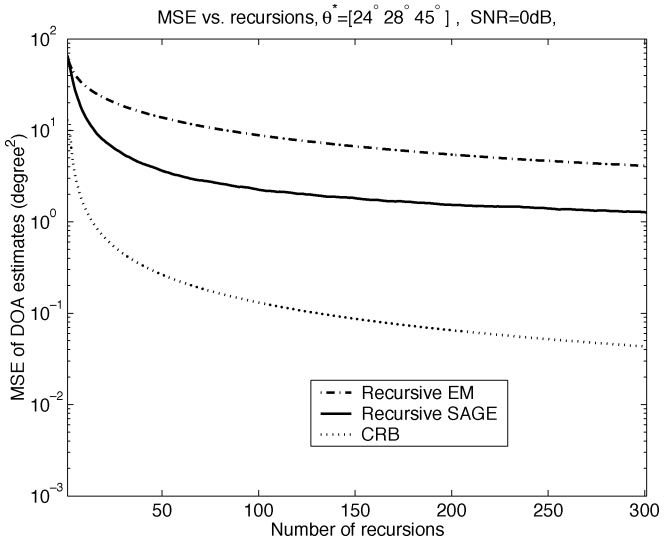

Fig. 9. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$. SNR $= 0$ dB. $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

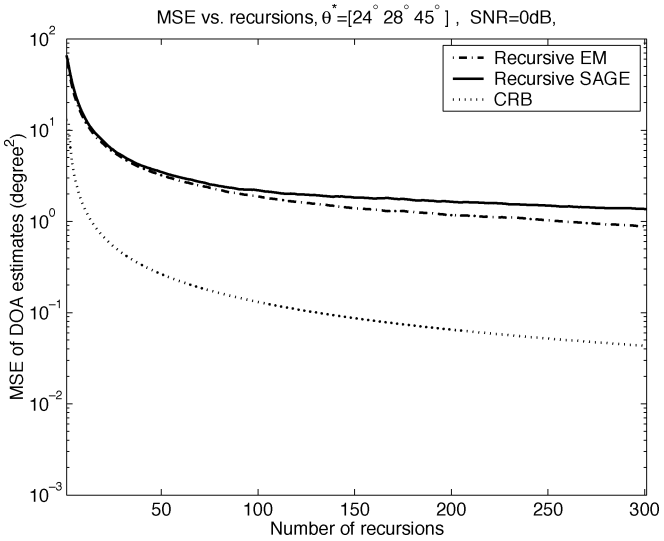

Fig. 10. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$. SNR $= 0$ dB. $\epsilon_n^{\mathrm{REM}} = 3n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

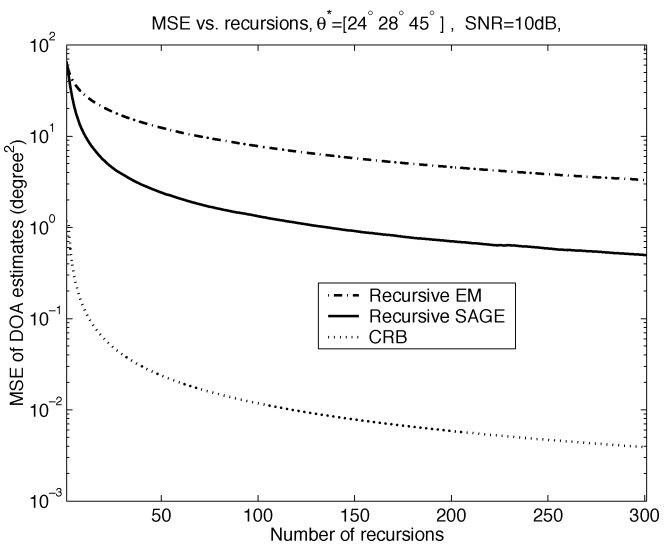

Fig. 11. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$. SNR $= 10$ dB. $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

signal model is not achieved because 1) the algorithms presented in this paper are derived under the deterministic signal model, 2) the recursion procedure is only applied to the DOA parameter, 3) the convergence rates given in Section IV-B are asymptotic results, and 4) a step size of $n^{-0.6}$ rather than n^{-1} is chosen because of a practical convergence rate. Compared to the batch processing, the recursive procedure is fast and computationally efficient but has a poorer estimation performance.

VII. Conclusion

This paper is concerned with recursive parameter estimation using augmented data. The recursive EM and SAGE-inspired algorithms are formulated in a very general form. It was shown that the recursive SAGE-inspired algorithm is closely related to the SAGE algorithm. Under mild conditions, the sequence of the estimates generated by the recursive EM and SAGE-inspired algorithms converge with probability one to a stationary point of the likelihood function. The normalized error vector is asymptotically normal distributed with zero mean and a covariance matrix that can be obtained by solving a matrix equation. These results are valid for a broad class of problems.

Based on the recursive EM and SAGE-inspired algorithms, we developed recursive procedures for estimating DOA parameters. Because of the simple structure of the augmented data, the recursive procedures have a very simple implementation. It was proved that convergence behavior of the proposed algorithms is governed by a function closely related to the concentrated likelihood function. Simulations showed that estimates generated by the recursive EM and SAGE-inspired algorithms achieve satisfactory accuracy over a wide range of SNRs. In general, higher SNR and better initial estimates lead to faster convergence. Both algorithms performed better in a scenario with widely separated signal sources than with closely located signal sources. The application to the direction-finding problem demonstrated that the

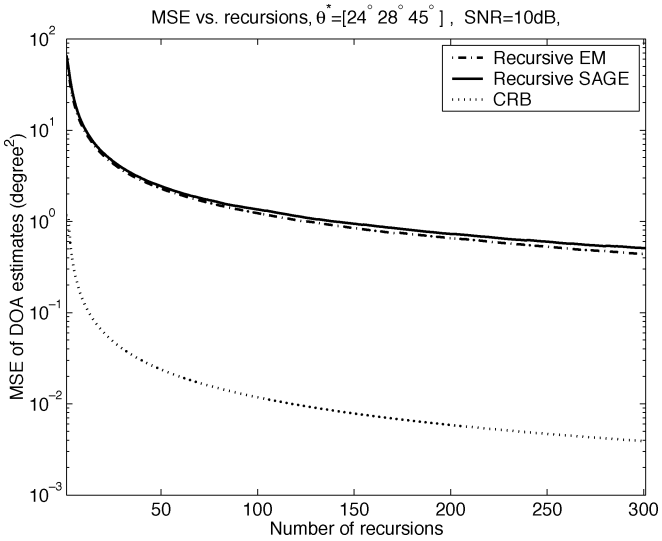

Fig. 12. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 28^\circ\ 45^\circ]$. SNR $= 10$ dB. $\epsilon_n^{\mathrm{REM}} = 3n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

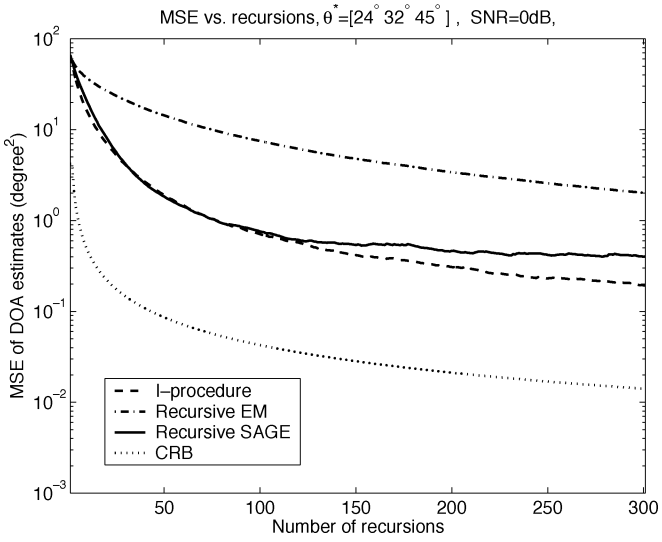

Fig. 13. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. $\underline{\theta}^0 = [19^\circ\ 36^\circ\ 50^\circ]$. SNR $= 0$ dB. $\epsilon_n^{\mathcal{I}} = n^{-0.6}$, $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

recursive EM and SAGE-inspired algorithms provide a computationally efficient method to find ML estimates.

Appendix A
Proof of Theorem 1

Note that

$$\underline{\Gamma}(\underline{\vartheta}) = -\underline{\nabla}_{\underline{\vartheta}} J(\underline{\vartheta}, \underline{\vartheta}^*) \tag{65}$$

where

$$J(\underline{\vartheta}, \underline{\vartheta}^*) = \int \log\left(\frac{f_{\underline{X}}(\underline{x}|\underline{\vartheta}^*)}{f_{\underline{X}}(\underline{x}|\underline{\vartheta})}\right) f_{\underline{X}}(\underline{x}|\underline{\vartheta}^*)\mathrm{d}\underline{x} \tag{66}$$

is the Kullback–Leibler distance between $f_{\underline{X}}(\underline{x}|\underline{\vartheta}^*)$ and $f_{\underline{X}}(\underline{x}|\underline{\vartheta})$. a) and b) follow immediately since it is well known

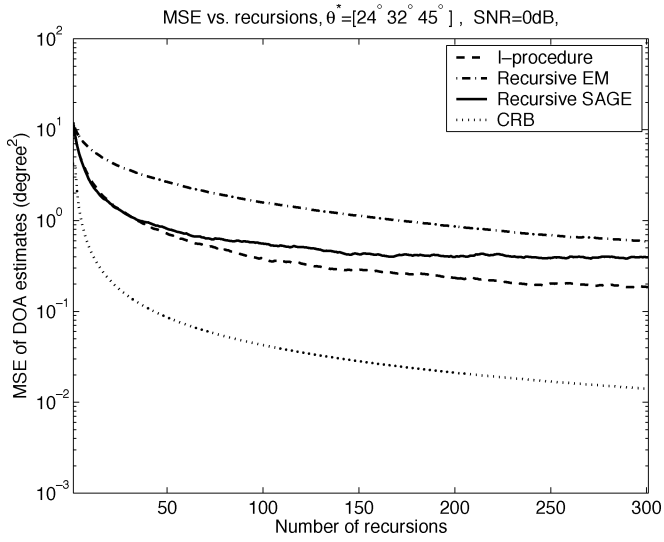

Fig. 14. MSE versus recursions. $\underline{\theta}^* = [24^\circ\ 32^\circ\ 45^\circ]$. $\underline{\theta}^0 = [22^\circ\ 34^\circ\ 47^\circ]$. SNR $= 0$ dB. $\epsilon_n^{\mathcal{I}} = n^{-0.6}$, $\epsilon_n^{\mathrm{REM}} = n^{-0.6}$, $\epsilon_n^{\mathrm{RSAGE}} = n^{-0.6}$.

that $J(\underline{\vartheta}, \underline{\vartheta}^*) \geq 0$, with equality if and only if $\underline{\vartheta} = \underline{\vartheta}^*$ for an identifiable model $f_{\underline{X}}(\underline{x}|\underline{\vartheta})$. □

Appendix B
Proof of Theorem 2

1) The convergence of $\underline{\vartheta}^n$ to the invariant set of (35) or (36) is guaranteed by decreasing step sizes and conditions a) and b).

The step sizes are specified as $\epsilon_n = an^{-\alpha}$, $a > 0$, $1/2 < \alpha \leq 1$, and $n \geq 0$. Therefore

$$\epsilon_n > 0, \quad \epsilon_n \to 0, \quad \sum_{n=0}^{\infty} \epsilon_n = \infty, \quad \sum_{n=0}^{\infty} \epsilon_n^2 < \infty. \tag{67}$$

Assumptions a) and b) yield the following inequality:

$$\begin{aligned} &\mathsf{E}\left[\left\|\mathbf{K}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right\|^2\right] \\ &\quad \leq \mathsf{E}\left[\left\|\mathbf{K}(\underline{\vartheta}^n)^{-1}\right\|^2 \left\|\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right\|^2\right] \\ &\quad = \mathsf{E}\left[\left\|\mathbf{K}(\underline{\vartheta}^n)^{-1}\right\|^2\right] \mathsf{E}\left[\left\|\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)\right\|^2\right] < \infty. \end{aligned} \tag{68}$$

Note that similar to (27), $\underline{Y}_n = \mathbf{K}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ can be decomposed as

$$\underline{Y}_n = \underline{g}(\underline{\vartheta}^n) + \delta\underline{M}_n. \tag{69}$$

By (67)–(69), and [17, Th. 2.1, p. 95], we conclude that asymptotically, $\underline{\vartheta}^n$ follows the trajectory of the solution $\underline{\vartheta}(t)$ to the ODE (35) in the constrained case, (36) in the unconstrained case.

2) The stationary points of (35) satisfy the condition $\underline{g}(\underline{\vartheta}) + \underline{z} = 0$. Recall that $\underline{g}(\underline{\vartheta}) = \mathbf{K}(\underline{\vartheta})^{-1}\underline{\Gamma}(\underline{\vartheta})$. As $\underline{\Gamma}(\underline{\vartheta}) = \underline{\nabla}_{\underline{\vartheta}} J(\underline{\vartheta}, \underline{\vartheta}^*)$ is the gradient of the Kullback–Leibler distance and $\mathbf{K}(\underline{\vartheta})^{-1}$ is positive definite, we can show that

the stationary points are asymptotically stable in the Liapunov sense. Thus, $\underline{\vartheta}^n$ converges to the set of stationary points.

More precisely, using $J(\underline{\vartheta}, \underline{\vartheta}^*)$ as the Liapunov function, the derivative of $J(\underline{\vartheta}, \underline{\vartheta}^*)$ along the solution can be expressed as

$$\begin{aligned}&\left[\underline{\nabla}_{\underline{\vartheta}} J(\underline{\vartheta}, \underline{\vartheta}^*)\right]^T \left[-\mathbf{K}(\underline{\vartheta})^{-1} \underline{\nabla}_{\underline{\vartheta}} J(\underline{\vartheta}, \underline{\vartheta}^*) + \underline{z}\right] \\ &\quad \le -\left[\underline{\nabla}_{\underline{\vartheta}} J(\underline{\vartheta}, \underline{\vartheta}^*)\right]^T \left[\mathbf{K}(\underline{\vartheta})^{-1} \underline{\nabla}_{\underline{\vartheta}} J(\underline{\vartheta}, \underline{\vartheta}^*)\right] \\ &\qquad + \left\|\underline{\nabla}_{\underline{\vartheta}} J(\underline{\vartheta}, \underline{\vartheta}^*)\right\| \|\underline{z}\| \\ &\quad \le 0. \end{aligned} \tag{70}$$

The inequality on the right-hand side results from the fact that $\|\underline{z}(t)\| \le \|\underline{g}(t)\|$.

The stationary points of (36) satisfy the condition $\underline{g}(\underline{\vartheta}) = \underline{0}$. Using $J(\underline{\vartheta}, \underline{\vartheta}^*)$ as a Liapunov function, it can be also shown that the stationary points are asymptotically stable. □

Appendix C
Proof of Lemma 1

Consider the first-order Taylor expansion

$$\underline{g}(\underline{\vartheta}) = \underline{g}(\bar{\underline{\vartheta}}) + \mathbf{g}(\bar{\underline{\vartheta}})(\underline{\vartheta} - \bar{\underline{\vartheta}}) + o\left(\|\underline{\vartheta} - \bar{\underline{\vartheta}}\|\right) \tag{71}$$

where the Jacobi matrix of $\underline{g}(\underline{\vartheta})$ is given by

$$\mathbf{g}(\bar{\underline{\vartheta}}) = \underline{\nabla}_{\underline{\vartheta}}^T \left[\mathbf{K}(\underline{\vartheta})^{-1} \underline{\Gamma}(\underline{\vartheta})\right]_{\underline{\vartheta}=\bar{\underline{\vartheta}}}. \tag{72}$$

Then, the lth row of $\mathbf{g}(\bar{\underline{\vartheta}})$ is given by

$$\begin{aligned}\left(\mathbf{g}(\bar{\underline{\vartheta}})\right)_l = \Bigg\{&\underline{\Gamma}(\underline{\vartheta})^T \left[\underline{\nabla}_{\underline{\vartheta}} \left(\mathbf{K}(\underline{\vartheta})^{-1}\right)_l\right]^T \\ &+ \left(\mathbf{K}(\underline{\vartheta})^{-1}\right)_l \left[\underline{\nabla}_{\underline{\vartheta}}^T \underline{\Gamma}(\underline{\vartheta})\right]\Bigg\}_{\underline{\vartheta}=\bar{\underline{\vartheta}}}\end{aligned} \tag{73}$$

where $(\mathbf{K}(\underline{\vartheta})^{-1})_l$ denotes the lth row of $\mathbf{K}(\underline{\vartheta})^{-1}$. Since $\underline{\Gamma}(\bar{\underline{\vartheta}}) = \underline{0}$, (73) can be simplified to

$$\begin{aligned}\left(\mathbf{g}(\bar{\underline{\vartheta}})\right)_l &= \left\{\left(\mathbf{K}(\underline{\vartheta})^{-1}\right)_l \left[\underline{\nabla}_{\underline{\vartheta}}^T \underline{\Gamma}(\underline{\vartheta})\right]\right\}_{\underline{\vartheta}=\bar{\underline{\vartheta}}} \\ &= -\left\{\left(\mathbf{K}(\underline{\vartheta})^{-1}\right)_l \mathcal{I}(\underline{\vartheta})\right\}_{\underline{\vartheta}=\bar{\underline{\vartheta}}}.\end{aligned} \tag{74}$$

The equation above implies

$$\mathbf{g}(\bar{\underline{\vartheta}}) = -\mathbf{K}(\bar{\underline{\vartheta}})^{-1} \mathcal{I}(\bar{\underline{\vartheta}}). \tag{75}$$

Because $\mathbf{K}(\bar{\underline{\vartheta}})^{-1}$ and $\mathcal{I}(\bar{\underline{\vartheta}})$ are positive definite, $\mathbf{g}(\bar{\underline{\vartheta}})$ is negative definite; thus, it is a stable matrix. □

Appendix D
Proof of Theorem 3

We will show that under assumptions a), b), and c), recursions (33) and (34) satisfy the following conditions.

1) There are constants k_1, k_2 such that $k_1\|\underline{\vartheta} - \bar{\underline{\vartheta}}\|^2 \le \underline{g}(\underline{\vartheta})^t(\underline{\vartheta} - \bar{\underline{\vartheta}}) \le k_2\|\underline{\vartheta} - \bar{\underline{\vartheta}}\|^2$.
2) There are constants k_3, k_4 such that $\|\underline{g}(\underline{\vartheta})\| \le k_3\|\underline{\vartheta}\| + k_4$.
3) $\mathsf{E}\|\underline{Y}_n - \underline{g}(\underline{\vartheta})\|^2 < \infty$.
4) $\underline{g}(\underline{\vartheta}) = \mathbf{A}(\underline{\vartheta} - \bar{\underline{\vartheta}}) + o(\underline{\vartheta} - \bar{\underline{\vartheta}})$ as $\underline{\vartheta} \to \bar{\underline{\vartheta}}$, where $\mathbf{A}$ is a stable matrix.
5) The matrix $\mathbf{C}$ is defined as $\mathsf{E}[(\underline{Y}_n - \underline{g}(\bar{\underline{\vartheta}}))(\underline{Y}_n - \underline{g}(\bar{\underline{\vartheta}}))^T]$ and then, apply results from [20].

Conditions 1), 2), and 4) can be easily verified by *Lemma 1*. Using (37)

$$\underline{g}(\underline{\vartheta})^T(\underline{\vartheta} - \bar{\underline{\vartheta}}) \approx (\underline{\vartheta} - \bar{\underline{\vartheta}})^T \mathbf{g}(\bar{\underline{\vartheta}})(\underline{\vartheta} - \bar{\underline{\vartheta}}) \tag{76}$$

where $\mathbf{g}(\bar{\underline{\vartheta}}) = -\mathbf{K}(\bar{\underline{\vartheta}})^{-1}\mathcal{I}(\bar{\underline{\vartheta}})$ is a bounded matrix under assumptions a) and b). Thus, 1) is satisfied. Condition 2) can be also verified by (37). Condition 4) is a direct consequence of *Lemma 1*.

By definition, $\underline{Y}_n = \mathbf{K}(\underline{\vartheta}^n)^{-1}\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$. With assumptions a) and b), we can verify 3) by the following inequality:

$$\mathsf{E}\left[\left\|\underline{Y}_n - \underline{g}(\underline{\vartheta}^n)\right\|^2\right] = \mathsf{E}\left[\|\underline{Y}_n\|^2\right] - \mathsf{E}\left[\left\|\underline{g}(\underline{\vartheta}^n)\right\|^2\right] < \infty. \tag{77}$$

The matrix $\mathbf{C}$ is the covariance of $\underline{Y}_n$ at the limit point $\bar{\underline{\vartheta}}$. Since $\underline{g}(\bar{\underline{\vartheta}}) = \underline{0}$

$$\begin{aligned}\mathbf{C} &= \mathsf{E}\left[\underline{Y}_n \underline{Y}_n^T\right] \\ &= \mathsf{E}\left[\mathbf{K}(\underline{\vartheta})^{-1}\underline{\gamma}(\underline{x}, \underline{\vartheta})\underline{\gamma}(\underline{x}, \underline{\vartheta})^T \mathbf{K}(\underline{\vartheta})^{-1}\right]_{\underline{\vartheta}=\bar{\underline{\vartheta}}} \\ &= \mathbf{K}(\bar{\underline{\vartheta}})^{-1}\mathcal{I}(\bar{\underline{\vartheta}})\mathbf{K}(\bar{\underline{\vartheta}})^{-1}.\end{aligned} \tag{78}$$

We have shown that conditions 1)–5) are satisfied by recursions (1) and (4). The asymptotic normality of $n^{\alpha/2}(\underline{\vartheta}^n - \bar{\underline{\vartheta}})$ follows from [20, Th. 5.8 and 5.10, pp. 291–293]. □

Appendix E
Proof of Result 1

Substituting (58) in (60), the mth element of $\underline{\eta}(\underline{x}_n, \underline{\theta}^n)$ can be written as

$$\begin{aligned}\left.\frac{\partial L_{\underline{X}}(\underline{x}_n, \underline{\theta})}{\partial \theta_m}\right|_{\underline{\theta}=\underline{\theta}^n} = &\frac{1}{\mathrm{tr}\left[\mathbf{P}^{\perp}(\underline{\theta}^n)\widehat{\mathbf{C}}_{\underline{X}}(n)\right]} \\ &\times \mathrm{tr}\left[\frac{\partial \mathbf{P}(\underline{\theta})}{\partial \theta_m}\widehat{\mathbf{C}}_{\underline{X}}(n)\right]_{\underline{\theta}=\underline{\theta}^n}.\end{aligned} \tag{79}$$

Note that $(1/L)\mathrm{tr}[\mathbf{P}^{\perp}(\underline{\theta}^n)\widehat{\mathbf{C}}_{\underline{X}}(n)]$ is the ML estimate ν^n for the noise parameter at $\underline{\theta} = \underline{\theta}^n$. By the fact $\widehat{\mathbf{C}}_{\underline{X}}(n) = \underline{X}(n)\underline{X}(n)^H$ and the identity [15]

$$\frac{\partial \mathbf{P}(\underline{\theta})}{\partial \theta_m} = \mathbf{P}^{\perp}(\underline{\theta})\frac{\partial \mathbf{H}(\underline{\theta})}{\partial \theta_m}\mathbf{H}(\underline{\theta})^{\#} + \mathbf{H}(\underline{\theta})^{\#H}\frac{\partial \mathbf{H}(\underline{\theta})^H}{\partial \theta_m}\mathbf{P}^{\perp}(\underline{\theta}) \tag{80}$$

(79) can be further simplified to

$$\frac{2}{L\nu^n}\mathrm{Re}\left[\underline{X}(n)^H \mathbf{P}^{\perp}(\underline{\theta}^n)\frac{\partial \mathbf{H}(\underline{\theta}^n)}{\partial \theta_m}\mathbf{H}(\underline{\theta}^n)^{\#}\underline{X}(n)\right]. \tag{81}$$

Note that $\mathbf{H}(\underline{\theta}^n)^{\#}\underline{X}(n)$ is the ML estimate $\underline{S}^n$ for the signal parameters at $\underline{\theta} = \underline{\theta}^n$. The first derivative of $\mathbf{H}(\underline{\theta})$ with respect to θ_m has only nonzero elements in the mth column

$$\frac{\partial \mathbf{H}(\underline{\theta}^n)}{\partial \theta_m} = \left[\underline{0} \quad \cdots \quad \underline{d}'(\theta_m) \quad \cdots \quad \underline{0}\right]. \tag{82}$$

These observations lead to

$$\frac{\partial \mathbf{H}(\underline{\theta}^n)}{\partial \theta_m}\mathbf{H}(\underline{\theta}^n)^{\#}\underline{X}(n) = \underline{d}'\left(\theta_m^n\right) S_m^n. \tag{83}$$

Furthermore

$$\begin{aligned}\mathbf{P}^{\perp}(\underline{\theta}^n)\underline{X}(n) &= \underline{X}(n) - \mathbf{H}(\underline{\theta}^n)\mathbf{H}(\underline{\theta}^n)^{\#}\underline{X}(n)\\ &= \underline{X}(n) - \mathbf{H}(\underline{\theta}^n)\underline{S}^n.\end{aligned} \tag{84}$$

Equations (81), (83),and (84) yield the following:

$$\begin{aligned}\left.\frac{\partial L_{\underline{X}}(\underline{x}_n, \underline{\theta})}{\partial \theta_m}\right|_{\underline{\vartheta}=\underline{\vartheta}^n} &= \frac{2}{L\nu^n}\mathrm{Re}\\ &\times \left[\left(\underline{X}(n) - \mathbf{H}(\underline{\theta}^n)\underline{S}^n\right)^H \left(\underline{d}'\left(\theta_m^n\right) S_m^n\right)\right].\end{aligned} \tag{85}$$

Comparing the above result with (44), it can be concluded that

$$\underline{\eta}(\underline{x}_n, \underline{\vartheta}^n) = \frac{1}{L}\underline{\gamma}(\underline{x}_n, \underline{\theta}^n). \tag{86}$$

□

Appendix F
Proof of Result 2

Equation (62) can be rewritten as

$$\begin{aligned}\mathcal{L}_{\underline{X}}(\underline{\theta}) &= -\log\left[\mathrm{tr}\left[\mathbf{P}(\underline{\theta})^{\perp}\mathbf{H}(\underline{\theta}^*)\mathbf{C}_{\underline{S}}\mathbf{H}(\underline{\theta}^*)\mathbf{P}(\underline{\theta})^{\perp}\right]\right.\\ &\qquad \left.+\mathrm{tr}\left[\mathbf{P}(\underline{\theta})^{\perp}(\nu\mathbf{I})\right]\right]\\ &= -\log\left[\mathrm{tr}\left[\mathbf{P}(\underline{\theta})^{\perp}\mathbf{H}(\underline{\theta}^*)\mathbf{C}_{\underline{S}}\mathbf{H}(\underline{\theta}^*)\mathbf{P}(\underline{\theta})^{\perp}\right]\right.\\ &\qquad \left.+\nu(L-M)\right].\end{aligned} \tag{87}$$

Obviously, $\mathcal{L}_{\underline{X}}(\underline{\theta})$ has a maximum at $\underline{\theta}^*$. Consequently

$$\underline{g}(\underline{\theta}^*) = \underline{0}. \tag{88}$$

Comparing functions (58) and (62), the gradient vector $\underline{\eta}(\underline{x}_n, \underline{\vartheta}^n)$ can be seen as a noisy observation of $\underline{g}(\underline{\theta}^n)$. By *Result* 1, we know $\underline{\gamma}(\underline{x}_n, \underline{\vartheta}^n)$ and $\underline{\eta}(\underline{x}_n, \underline{\vartheta}^n)$ are equivalent, except for a constant, which can be absorbed into the step size. The inverses of $\mathcal{I}_{\mathrm{EM}}(\underline{\vartheta}^n)$ and $\mathcal{I}_{\mathrm{SAGE}}(\underline{\vartheta}^n)$ can be regarded as gain matrices that improve the convergence rate of the algorithm. Thus, the recursive EM and SAGE-inspired algorithms derived under a deterministic signal model are indeed stochastic approximation procedures for finding zeros of $\underline{g}(\underline{\theta})$. For a good initial estimate $\underline{\theta}^0$, the recursive procedures converge to the desired parameters $\underline{\theta}^*$. □

Acknowledgment

The authors would like to thank A. O. Hero for a fruitful discussion. They also would like to thank Associate Editor C. B. Papadias and the anonymous reviewers for their constructive comments that significantly improved the manuscript.

References

[1] T. M. Apostol, *Mathematical Analysis*, 2nd ed. Reading, MA: Addison-Wesley, 1975.
[2] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag, 1990.
[3] J. F. Böhme, "Array processing," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 1–63.
[4] D. R. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden-Day, 1981.
[5] P. J. Chung and J. F. Böhme, "Recursive EM and SAGE algorithms," in *Proc. IEEE Workshop Statist. Signal Process.*, Singapore, Aug. 6–8, 2001, pp. 540–542.
[6] ——, "Comparative convergence analysis of EM and SAGE algorithms in DOA estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 2940–2949, Dec. 2001.
[7] ——, "DOA estimation of multiple moving sources using recursive EM algorithms," in *Proc. Sensor Array Multichannel Signal Process. Workshop*, Washington, DC, Aug. 4–6, 2002, pp. 323–326.
[8] ——, "EM and SAGE algorithms for towed array data," in *The Applications of Space-Time Adaptive Processing*, R. Klemm, Ed. London, U.K.: IEE, 2004.
[9] A. P. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. B39, pp. 1–38, 1977.
[10] V. Fabian, "On asymptotic normality in stochastic approximation," *Ann. Math. Statist.*, vol. 39, pp. 854–866, 1968.
[11] ——, "On asymptotically efficient recursive estimation," *Ann. Statist.*, vol. 6, no. 4, pp. 854–866, 1978.
[12] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
[13] L. Frenkel and M. Feder, "Recursive Expectation-Maximization (EM) algorithms for time-varying parameters with applications to multiple target tracking," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 306–320, Feb. 1999.
[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[15] D. Kraus, "Approximative Maximum-Likelihood-Schätzung und verwandte Verfahren zur Ortung und Signalschätzung mit Sensorgruppen," Dr.-Ing. dissertation, Faculty Elect. Eng., Ruhr-Universität Bochum, Shaker Verlag, Aachen, Germany, 1993.
[16] H. J. Kushner and D. S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*. New York: Springer-Verlag, 1978.
[17] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.
[18] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Autom. Control*, vol. AC-22, no. 4, pp. 551–575, Aug. 1977.
[19] D. Maiwald, "Breitbandverfahren zur Signalentdeckung und-ortung mit Sensorgruppen in Seismik- und Sonaranwendungen," Dr.-Ing. dissertation, Faculty Elect. Eng., Ruhr-Universität Bochum, Shaker Verlag, Aachen, Germany, 1995.
[20] G. Pflug, *Optimization of Stochastic Models*. Norwell, MA: Kluwer, 1996.
[21] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM J. Control Optim.*, vol. 30, pp. 838–855, 1992.
[22] J. Sacks, "Asymptotic distribution of stochastic approximation procedures," *Ann. Math. Statist.*, vol. 29, pp. 373–405, 1958.
[23] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
[24] ——, "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.
[25] D. M. Titterington, "Recursive parameter estimation using incomplete data," *J. R. Statist. Soc. B*, vol. 46, no. 2, pp. 257–267, 1984.
[26] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
[27] D. Williams, *Probability With Martingales*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[28] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

Pei-Jung Chung received the Dr.-Ing. degree in 2002 from Ruhr-Universität Bochum, Bochum, Germany.

From 1998 to 2002, she was with Signal Theory Group, Department of Electrical Engineering and Information Sciences, Ruhr-Universität Bochum. From May 2002 to January 2004, she held a postdoctoral position at Carnegie Mellon University, Pittsburgh, PA. From February 2004 to July 2004, she was a postdoctoral researcher at University of Michigan, Ann Arbor. She is now an Assistant Professor with the Department of Electronics Engineering, National Chiao Tung University, Hsin Chu, Taiwan, R.O.C. Her research interests include array processing, statistical signal processing, and its applications.

Johann F. Böhme (M'74–SM'88–F'90) received the Diplom degree in mathematics in 1966, the Dr.-Ing. degree in 1970, and the Habilitation degree in 1977, both in computer science, from the Technical University of Hannover, Hannover, Germany, the University of Erlangen, Erlangen, Germany, and the University of Bonn, Bonn, Germany, respectively.

From 1967 to 1974, he was with the sonar-research laboratory of Krupp Atlas Elektronik, Bremen, Germany. He then joined the University of Bonn until 1978 and the FGAN in Wachtberg-Werthhoven, Germany. Since 1980, he has been Professor of signal theory with the Department of Electrical Engineering and Information Sciences, Ruhr-Universität Bochum. His research interests are in the domain of statistical signal processing and its applications.

Dr. Böhme was the recipient of the 2003 IEEE Signal Processing Society Technical Achievement Award.