COMPARATIVE CONVERGENCE ANALYSIS OF EM AND SAGE ALGORITHMS IN DOA ESTIMATION

Pei Jung Chung and Johann F. Böhme

Dept. of Electrical Engineering and Information Science
Ruhr-Universität Bochum, 44780 Bochum, Germany
e-mail: pjc@sth.ruhr-uni-bochum.de

ABSTRACT

In this work, the convergence rates of direction of arrival (DOA) estimates using Expectation-Maximization (EM) and Space Alternating Generalized EM (SAGE) algorithms are investigated. EM algorithm is a well known recursive method for locating modes of a likelihood function which is characterized by simple implementation and stability. Unfortunately the slow convergence associated with EM makes it less attractive. The recently proposed SAGE algorithm, based on the same idea of data augmentation, preserves the advantage of simple implementation and has the potential to speed up convergence. Theoretical analysis shows that SAGE has faster convergence rate than EM under certain conditions. This conclusion is also supported by numerical experiments carried out over a wide range of SNRs and different numbers of snapshots.

1. INTRODUCTION

The problem of estimating the direction of arrival (DOA) of plane waves sampled by an array of sensors has been extensively investigated in last decades. Among all estimation methods Maximum Likelihood (ML) has asymptotically best statistical performance and is in some sense robust against small sample number and coherent source signals. Unfortunately the high computational complexity associated with standard implementation of ML makes it less attractive in practice. It is an important and challenging task to find faster, more efficient implementations of ML.

One popular approach to overcome this computational difficulty is the Expectation–Maximization (EM) algorithm [4]. EM algorithm enjoys two favorable properties: simple implementation and stability. However, in many cases it suffers from slow convergence. Based on the same idea of data augmentation, the Alternating Generalized EM (SAGE) algorithm developed by Fessler and Hero [6] can improve the convergence rate significantly.

The EM algorithm was firstly applied to the array processing problem in [5] for the deterministic signal model assuming known signal waveforms and noise covariance. The SAGE algorithm was applied to estimate DOA of wideband signals in [1] using deterministic signal models with known noise covariance structure. It was observed in simulation results that SAGE algorithm requires less iterations to reach a stationary point than EM algorithm.

In this work we study the convergence rates of EM and SAGE algorithms for direction finding problem under deterministic signal models, Gaussian noise with known structure. The rate matrices derived herein have similar structure as those derived for the problem of estimating linear superimposed signals in Gaussian noise [6]. It has been demonstrated that SAGE algorithm converges faster than EM algorithm without introducing additional conditions. However, a more careful analysis shows that the faster convergence of SAGE algorithm can be only guaranteed if certain conditions on observed and augmented information matrices are satisfied.

Furthermore, computer simulations are carried out to investigate convergence behavior of EM and SAGE algorithms under different SNRs and numbers of snapshots. It is observed that the SAGE has significantly faster convergence than EM but is less stable for low SNR and small number of snapshots.

This paper is organized as follows. The signal model is described in section 2. In section 3 and 4 we develop EM and SAGE algorithms for the direction finding problem under deterministic signal model and Gaussian noise assumption. Convergence analysis is presented in section 5. Finally simulation results are discussed in section 6.

2. SIGNAL MODEL

Consider an array of N sensors receiving signals generated by M far field narrowband sources. The array output $\underline{X}(t) \in \mathbb{C}^{N\times 1}$ is sampled at time instances $t = 1, \ldots, T$. For signals arriving from $\underline{\theta} = (\theta_1, \ldots, \theta_M)$ the array output $\underline{X}(t)$ can be described as

$$\underline{X}(t) = \mathbf{H}(\underline{\theta})\, \underline{s}(t) + \underline{U}(t), \tag{1}$$

where $\mathbf{H}(\underline{\theta}) = [\underline{d}(\theta_1), \ldots, \underline{d}(\theta_M)] \in \mathbb{C}^{N\times M}$ contains M steering vectors $\underline{d}(\theta_m) \in \mathbb{C}^{N\times 1}$ $(m = 1, \ldots, M)$, $\underline{s}(t) = [s_1(t) \ldots s_M(t)]^T \in \mathbb{C}^{M\times 1}$, $\underline{U}(t) \in \mathbb{C}^{N\times 1}$ denote signal waveforms, noise vector, respectively. We assume the signal $\underline{s}(t)$ to be unknown, deterministic and the noise $\underline{U}(t)$ to be complex normally distributed with zero mean and covariance matrix $\nu\mathbf{I}$, where ν is the noise spectral parameter and $\mathbf{I}$ is the identity matrix. Furthermore, the number of sources, M, is known. The problem is to estimate $\underline{\theta}$ using the observed data $\underline{X} = \{\underline{X}(t) : 1 \le t \le T\}$.

3. EM ALGORITHM

EM algorithm is a well known iterative method for locating modes of likelihood function. The basic idea is: rather than maximizing the likelihood function of observed data $\underline{X}$ which may be complicated and intractable, one specifies an augmented data $\underline{Y}$ so that $\underline{X} = \mathcal{M}(\underline{Y})$ is a many–to–one mapping and performs a series of

This work has been supported by German Science Foundation.

simple maximizations [4] . More specifically, each iteration consists of two steps: the E–step (expectation) which approximates the augmented data by conditional expectation and the M–step (maximization) which maximizes the augmented data likelihood.

We construct augmented data by exploiting the superposition property of array outputs [5]. Decomposing $\underline{X}(t)$ into its signal plus noise components the augmented data

$$\underline{Y}(t) = [\underline{Y}_1(t)^T \dots \underline{Y}_m(t)^T \dots \underline{Y}_M(t)^T]^T \tag{2}$$

where

$$\underline{Y}_m(t) = \underline{d}(\theta_m) s_m(t) + \underline{U}_m(t). \tag{3}$$

The noise processes $\underline{U}_m(t)$ are independent from each other, complex Gaussian distributed with zero mean and covariance matrices $\nu_m \mathbf{I}$ $(m = 1, \dots, M)$ under the constraint $\sum_{m=1}^{M} \nu_m = \nu, (0 < \nu_m < \nu)$. A convenient choice is $\nu_m = (1/M)\nu$. In the following we consider $\{\nu_m\}_{m=1}^{M}$ as known parameters so that EM and SAGE algorithms have the same number of unknown parameters. The unknown parameter vector is now given by $\underline{\vartheta} = (\underline{\vartheta}_1, \dots, \underline{\vartheta}_m, \dots, \underline{\vartheta}_M)$ with $\underline{\vartheta}_m = (\theta_m, s_m(1), \dots, s_m(T))$.

Let $(\cdot)^{[i]}$ and $(\cdot)^{[i+1]}$ denote the estimates from i–th and $(i+1)$–th iteration respectively. Making use of the property of conditional Gaussian distribution the E– and M– step of the $(i+1)$–th iteration can be described as:

E–step: Calculate $(m = 1, \dots, M)$

$$\begin{aligned}
\hat{\underline{Y}}_m(t, \underline{\vartheta}^{[i]}) &= \mathsf{E}\left[\underline{Y}_m(t) \mid \underline{X}, \underline{\vartheta}^{[i]}\right] \\
&= \underline{d}(\theta_m^{[m]}) s_m^{[i]}(t) + \frac{\nu_m}{\nu} \left(\underline{X}(t) - \mathbf{H}(\underline{\theta}^{[i]}) \underline{s}^{[i]}(t)\right), \\
\widehat{\mathbf{R}}_{\underline{Y}_m}(\underline{\vartheta}^{[i]}) &= \mathsf{E}\left[\underline{Y}_m \underline{Y}_m^H \mid \underline{X}, \underline{\vartheta}^{[i]}\right] \\
&= \frac{1}{T} \sum_{t=1}^{T} \hat{\underline{Y}}_m(t, \underline{\vartheta}^{[i]}) \hat{\underline{Y}}_m(t, \underline{\vartheta}^{[i]})^H + \frac{\nu_m^2}{\nu} \mathbf{I}.
\end{aligned}$$

M–step: Update $\underline{\vartheta}_m$ $(m = 1, \dots, M)$

$$\theta_m^{[i+1]} = \arg\max_{\theta_m} \underline{d}(\theta_m)^H \widehat{\mathbf{R}}_{\underline{Y}_m}(\underline{\vartheta}^{[i]}) \underline{d}(\theta_m), \tag{4}$$

$$s_m^{[i+1]}(t) = \frac{1}{N} \underline{d}(\theta_m^{[i+1]})^H \hat{\underline{Y}}_m(t, \underline{\vartheta}^{[i]}), (t = 1, \dots, T). \tag{5}$$

Note that $\underline{\vartheta}_m (m = 1, \dots, M)$ are updated in parallel. The optimization of (4) requires only one dimensional search.

4. SAGE ALGORITHM

The space alternating generalized EM (SAGE) algorithm [6] generalizes the idea of data augmentation to simplify computations of EM algorithm and can improve the convergence rate in some settings. Instead of estimating all parameters at once, SAGE breaks up the problem into several smaller ones and uses EM to update the parameter subset associated with each reduced problem. More specifically, each iteration consists of several cycles. The parameter subset associated with the m–th cycle $\underline{\vartheta}_m$ is updated by maximizing the conditionally expected likelihood of augmented data $\underline{Z}_m$ corresponding to this cycle.

We divide the parameter vector $\underline{\vartheta}$ into M subsets $\{\underline{\vartheta}_m = (\theta_m, \underline{s}_m)\}_{m=1}^{M}$ with $\underline{s}_m = (s_m(1), \dots, s_m(T))$. Each of them corresponds to one source signal. Accordingly the augmented data associated with the m–th cycle, $\{\underline{Z}_m(t)\}_{t=1}^{T}$ is given by

$$\underline{Z}_m(t) = \underline{d}(\theta_m) s_m(t) + \underline{U}(t). \tag{6}$$

The parameter subsets $\{\underline{\vartheta}_m : 1 \le m \le M\}$ are updated sequentially within one iteration. After M cycles all elements of $\underline{\vartheta}$ are updated once. Let $(\cdot)^{[i,m]}$ denote the estimate from m–th cycle of the i–th iteration. The i–th iteration can be expressed as:

$$\underline{\vartheta}^{[i,0]} = \underline{\vartheta}^{[i-1,M]}. \tag{7}$$

For $m = 1, \dots, M$

E–step Calculate

$$\begin{aligned}
\widehat{\underline{Z}}_m(t, \underline{\vartheta}^{[i,m-1]}) &= \mathsf{E}\left[\underline{Z}_m(t) \mid \underline{X}, \underline{\vartheta}^{[i,m-1]}\right] = \\
\underline{d}(\theta_m^{[i,m-1]}) s_m^{[i,m-1]}(t) &+ \underline{X}(t) - \mathbf{H}(\underline{\vartheta}^{[i,m-1]}) \underline{S}^{[i,m-1]}(t), \\
\widehat{\mathbf{R}}_{\underline{Z}_m}(\underline{\vartheta}^{[i,m-1]}) &= \mathsf{E}\left[\underline{Z}_m \underline{Z}_m^H \mid \underline{X}, \underline{\vartheta}^{[i,m-1]}\right] = \\
\nu \mathbf{I} &+ \frac{1}{T} \sum_{t=1}^{T} \hat{\underline{Z}}_m(t, \underline{\vartheta}^{[i,m-1]}) \hat{\underline{Z}}_m(t, \underline{\vartheta}^{[i,m-1]})^H.
\end{aligned}$$

M–step Update $\underline{\vartheta}_m$

$$\begin{aligned}
\theta_m^{[i,m]} &= \arg\max_{\theta_m} \underline{d}(\theta_m)^H \widehat{\mathbf{R}}_{\underline{Z}_m}(\underline{\vartheta}^{[i,m-1]}) \underline{d}(\theta_m), \\
s_m^{[i,m]}(t) &= \frac{1}{N} \underline{d}(\theta_m^{[i,m]})^H \hat{\underline{Z}}_m(t, \underline{\vartheta}^{[i,m-1]}), (t = 1, \dots, T), \\
\underline{\vartheta}^{[i,m]} &= (\underline{\vartheta}_1^{[i,m-1]}, \dots, \underline{\vartheta}_{m-1}^{[i,m-1]}, \underline{\vartheta}_m^{[i,m]}, \underline{\vartheta}_{m+1}^{[i,m-1]}, \dots, \underline{\vartheta}_M^{[i,m-1]}).
\end{aligned}$$

Comparing the maximization steps of SAGE and EM algorithm we note that both algorithms have the same computational complexity at each iteration. However the faster convergence of SAGE algorithm can reduce the number of necessary iterations and then needs less computations than EM algorithm.

5. CONVERGENCE ANALYSIS

It is well known that under standard regularity conditions the sequence of estimates generated by EM algorithm converges to a stationary point of the likelihood function and the likelihood never decreases during the iterations [4],[2]. Analog results for SAGE algorithm have been developed in [6],[10]. In this section we investigate the convergence properties of the algorithms derived before.

Based on the augmentation scheme introduced in section 3, 4, we derive the information matrices of augmented and observed data and then compare the convergence rates associated with the EM and SAGE algorithms. We assume that 1) the regularity conditions are satisfied, 2) the maximum point of the likelihood $\underline{\vartheta}^*$ is an interior point of the parameter space.

Let $f_{\underline{X}}(\underline{\vartheta})$, $f_{\underline{Y}}(\underline{\vartheta})$, $f_{\underline{Z}_m}(\underline{\vartheta}_m)$ denote the density function of $\underline{X}, \underline{Y}, \underline{Z}_m$ respectively. From [4] we know that the rate matrix for EM algorithm is given by:

$$\mathbf{D}\mathbf{M}^{\mathrm{EM}} = \mathbf{I} - \mathcal{I}_{\mathrm{EM}}^{-1} \mathcal{I}_{\mathrm{obs}}, \tag{8}$$

where $\mathbf{I}$ denotes the identity matrix,

$$\mathcal{I}_{\rm obs} = -\underline{\nabla}_{\underline{\vartheta}} \underline{\nabla}_{\underline{\vartheta}}^T \log f_{\underline{X}}(\underline{\vartheta})|_{\underline{\vartheta}=\underline{\vartheta}^*} \quad (9)$$

is the observed information matrix and

$$\mathcal{I}_{\rm EM} = \mathsf{E}\left[-\underline{\nabla}_{\underline{\vartheta}} \underline{\nabla}_{\underline{\vartheta}}^T \log f_{\underline{Y}}(\underline{\vartheta})|\, \underline{X}, \underline{\vartheta}\right]|_{\underline{\vartheta}=\underline{\vartheta}^*} \quad (10)$$

is the expected augmented information matrix. $\underline{\nabla}_{\underline{\vartheta}}$ is a column gradient operator with respect to the parameter vector $\underline{\vartheta}$.

The rate matrix for the SAGE algorithm is given by [6]:

$$\mathbf{DM}^{\rm SAGE} = \mathbf{I} - \tilde{\mathcal{I}}_{\rm SAGE}^{-1} \mathcal{I}_{\rm obs}\,, \quad (11)$$

with

$$\tilde{\mathcal{I}}_{\rm SAGE} = \mathcal{I}_{\rm SAGE} + \mathcal{L}\,. \quad (12)$$

$\mathcal{I}_{\rm SAGE}$ is a block diagonal matrix with the m–th block given by

$$\mathcal{I}_{\rm SAGE}^{[m]} = \mathsf{E}\left[-\underline{\nabla}_{\underline{\vartheta}_m} \underline{\nabla}_{\underline{\vartheta}_m}^T \log f_{\underline{Z}_m}(\underline{\vartheta}_m)|\, \underline{X}, \underline{\vartheta}\right]|_{\underline{\vartheta}=\underline{\vartheta}^*}\,. \quad (13)$$

The matrix $\mathcal{L}$ is defined by splitting $\mathcal{I}_{\rm obs}$ as following

$$\mathcal{I}_{\rm obs} = \mathcal{D} + \mathcal{L} + \mathcal{L}^T\,, \quad (14)$$

where $\mathcal{D}$, $\mathcal{L}$ represent the block diagonal, strict lower triangular block parts of $\mathcal{I}_{\rm obs}$, respectively. The m–th block of $\mathcal{D}$ corresponds to $\underline{\vartheta}_m$. Note that $\mathcal{I}_{\rm obs}$ is positive semidefinite at the stationary point. To simplify our analysis $\mathcal{I}_{\rm obs}$ is assumed to be positive definite. The convergence rate of the algorithm is determined by the spectral radius $\rho(\mathbf{DM})$ of the rate matrix $\mathbf{DM}$ [9]. Larger $\rho(\mathbf{DM})$ leads to slower convergence speed [10] .

Since computations of $\mathcal{I}_{\rm EM}$, $\mathcal{I}_{\rm SAGE}$, $\mathcal{I}_{\rm obs}$ are tedious but not difficult, we only present the most relevant results about their structures. The elements of $\mathcal{I}_{\rm EM}, \mathcal{I}_{\rm SAGE}, \mathcal{I}_{\rm obs}$ are real and depend on $(\underline{X} - \mathbf{H}(\underline{\theta})\underline{S})$, $\underline{d}(\theta_m)$, first and second derivatives of $\underline{d}(\theta_m)$. Other details can be found in [3].

Result 1 $\mathcal{I}_{\rm EM}$ is a block diagonal matrix and

$$\mathcal{I}_{\rm EM} = \mathcal{I}_{\rm SAGE} + \mathbf{\Delta}, \quad (15)$$

where $\mathbf{\Delta}$ is positive semidefinite.

Result 2

$$\mathcal{I}_{\rm SAGE} = \mathcal{D}. \quad (16)$$

By means of **Result 1** and **Result 2** the rate matrices defined in (8), (11) can be expressed as

$$\mathbf{DM}^{\rm EM} = \mathbf{I} - (\mathcal{D} + \mathbf{\Delta})^{-1}\mathcal{I}_{\rm obs}, \quad (17)$$

$$\mathbf{DM}^{\rm SAGE} = \mathbf{I} - (\mathcal{D} + \mathcal{L})^{-1}\mathcal{I}_{\rm obs}. \quad (18)$$

Interestingly, $\mathbf{DM}^{\rm SAGE}$ has the same structure as the iteration matrix of the block Gauss-Seidel method [8],[7]. The convergence of SAGE algorithm follows directly from convergence of the block Gauss-Seidel method. In the following theorem we show for $T=1$ that SAGE algorithm converges faster than EM algorithm under certain conditions. A similar proof for $T > 1$ can be found in [3].

Theorem 1 *For DOA estimation problem with $T = 1$, deterministic signals and the data augmentation schemes specified in (2), (6), the SAGE algorithm converges faster than EM algorithm, i.e.*

$$\rho(\mathbf{DM}^{\rm SAGE}) < \rho(\mathbf{DM}^{\rm EM}) \quad (19)$$

if

$$(\mathcal{D} + \mathbf{\Delta}) > \frac{1}{c}\mathcal{D} \quad \text{where} \quad c = 1 - \frac{1}{\sqrt{5}} \quad (20)$$

and

$$\rho(\ \mathcal{D}^{-1}\mathcal{L}\mathcal{D}^{-1}\mathcal{L}^H\ ) < 1/4. \quad (21)$$

Proof: Let γ and Γ denote the smallest eigenvalues of $\mathcal{D}^{-1}\mathcal{I}_{\rm obs}$ and $(\mathcal{D} + \mathbf{\Delta})^{-1}\mathcal{I}_{\rm obs}$, respectively. Note that $\gamma \in [0\ 1)$. Taking inverse of both sides of (20) we obtain the following inequality

$$(\mathcal{D} + \mathbf{\Delta})^{-1} < c\ \mathcal{D}^{-1}. \quad (22)$$

The above relation is preserved under congruence transformation with $\mathcal{I}_{\rm obs}^{1/2}$

$$\mathcal{I}_{\rm obs}^{1/2}\ (\mathcal{D} + \mathbf{\Delta})^{-1}\ \mathcal{I}_{\rm obs}^{1/2} < c\ \mathcal{I}_{\rm obs}^{1/2}\ \mathcal{D}^{-1}\ \mathcal{I}_{\rm obs}^{1/2}. \quad (23)$$

Because $(\mathcal{D} + \mathbf{\Delta})^{-1}\mathcal{I}_{\rm obs}$ and $\mathcal{I}_{\rm obs}^{1/2}\ (\mathcal{D} + \mathbf{\Delta})^{-1}\ \mathcal{I}_{\rm obs}^{1/2}$, as well as $\mathcal{D}^{-1}\mathcal{I}_{\rm obs}$ and $\mathcal{I}_{\rm obs}^{1/2}\ \mathcal{D}^{-1}\ \mathcal{I}_{\rm obs}^{1/2}$ are similar, we get

$$0 < \Gamma < c\gamma < 1 \quad (24)$$

from (23). Therefore

$$\rho(\mathbf{DM}^{\rm EM}) = 1 - \Gamma > 1 - c\gamma. \quad (25)$$

From (21) and [8] we have

$$\|\mathbf{DM}^{\rm SAGE}\|_{\mathcal{I}_{\rm obs}} < (1 + 4\gamma)^{-1/2}, \quad (26)$$

where $\|\mathbf{DM}^{\rm SAGE}\|_{\mathcal{I}_{\rm obs}} = \|{\mathcal{I}_{\rm obs}}^{1/2}\mathbf{DM}^{\rm SAGE}{\mathcal{I}_{\rm obs}}^{-1/2}\|_2$ with $\|\cdot\|_2$ representing the induced matrix 2 norm. In addition, for $f(\gamma) = (1 + 4\gamma)^{-1/2}$, $\gamma \in [0\ 1)$, $c = 1 - \frac{1}{\sqrt{5}}$,

$$f(\gamma) < 1 - c\gamma. \quad (27)$$

From (18),(25),(26) and (27),

$$\rho(\mathbf{DM}^{\rm SAGE}) \le \|\mathbf{DM}^{\rm SAGE}\|_{\mathcal{I}_{\rm obs}} < \rho(\mathbf{DM}^{\rm EM}). \quad (28)$$

Thus we have proved that SAGE algorithm converges faster than EM algorithm if conditions (20),(21) are satisfied. □

The problem of estimating linear superimposed signals in Gaussian noise discussed in [6] has similar rate convergence matrices for EM and SAGE algorithms as in this work. It has been shown that SAGE algorithm converges faster than EM algorithm without introducing additional conditions. This could be a little bit misleading because the proof therein is based on the assumption that the eigenvalue λ of $\mathbf{DM}^{\rm SAGE}$ with the largest magnitude $|\lambda| = \rho(\mathbf{DM}^{\rm SAGE})$ is positive. But the rate convergence matrix $\mathbf{DM}^{\rm SAGE}$ in (18) is not symmetric and must not have a positive spectrum.

6. NUMERICAL EXPERIMENTS

In this section we investigate the convergence properties of EM and SAGE algorithms under various SNRs and numbers of snapshots by numerical experiments. The narrow band wave signals are generated by three sources of equal power located at $\underline{\theta}_{\rm true} = [24^\circ\ 27^\circ\ 45^\circ]$. The SNR is defined as $10\log[|s_m(t)|^2/\nu]$. A uniformly linear array of 15 sensors with inter-element spacings of half a wavelength is used. The arriving angles are measured from broadside of the array. The initial estimate for DOA is chosen to

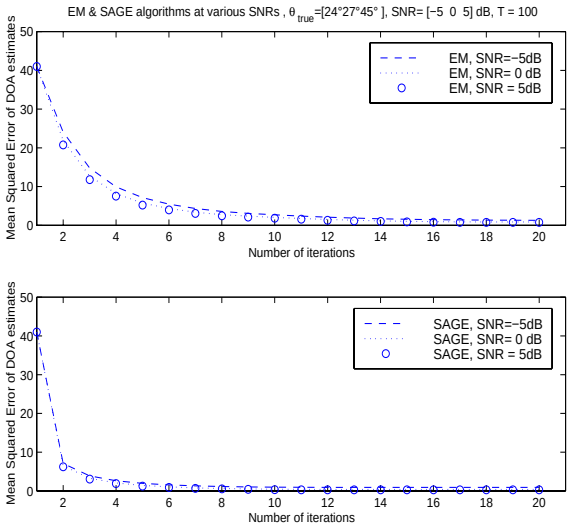

Fig. 1. Convergence rates at different SNRs.

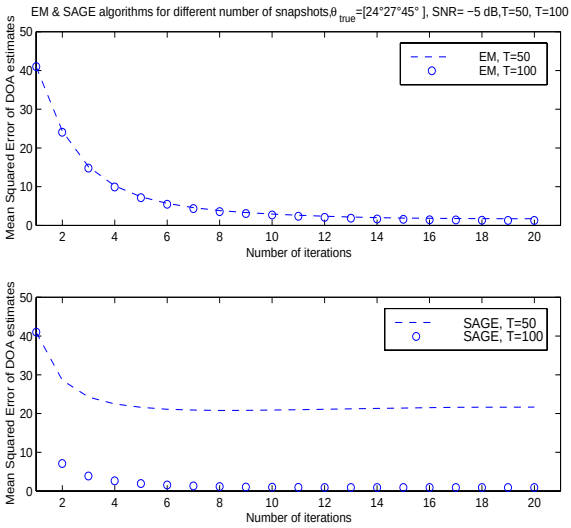

Fig. 2. Convergence rates for $T = 50$ and $T = 100$ at SNR=-5 dB.

be $\underline{\theta}^{[0]} = [20^\circ \ 30^\circ \ 41^\circ]$. The maximum number of iterations is set to 30. Each experiment is run through 50 Monte Carlo simulations. The mean squared error (MSE) at i–th iteration is defined as the mean of $|\underline{\theta}^{[i]} - \underline{\theta}_{\text{true}}|^2$.

In the first experiment the number of snapshots is $T = 100$, SNR varies from -5dB to 5dB at a 5 dB step. Figure (1) shows that SAGE converges faster than EM algorithm. Both algorithms improves convergence speed slightly with increasing SNR. However, comparing of the upper and lower parts of figure (1), it can be observed that larger SNR causes more change in convergence rate for EM than for SAGE algorithm.

In the second experiment SNR is fixed at -5dB, $T = 50, 100$. Figure 2 shows that while EM algorithm converges at almost the same rate, SAGE algorithm fails to converge to the true parameter for $T = 50$. It implies that the initial estimate must be even closer to the true parameter to get the correct final estimate. As discussed in [6] the monotone convergence region is reduced through the use of less informative augmented data. This is the price paid for the faster convergence rate.

7. CONCLUSION

In this work we investigate the convergence properties of EM and SAGE algorithms in the application to DOA estimation. The algorithms are derived for deterministic signal models, Gaussian noise with known noise structure. The rate matrices of EM and SAGE algorithms are calculated and their spectral radii are compared with each other. Our analysis shows that under certain conditions SAGE algorithm converges faster than EM algorithm. Numerical experiments demonstrate that SAGE algorithm can improve convergence rates significantly. A main drawback of SAGE algorithm is that it may become unstable for low SNR and small number of snapshots. However, the flexibility in choosing parameter sets and fast convergence suggest that SAGE algorithm is an useful numerical tool to find ML estimates.

8. ACKNOWLEDGEMENT

The authors are very grateful to Ivor Nissen, FWG Kiel, Germany for valuable suggestions.

9. REFERENCES

[1] Nail Cadalli and Orhan Arikan. Wideband maximum likelihood direction finding and signal parameter estimation by using tree–structured EM algorithm. *IEEE Trans. Signal Processing*, 47(1):201–206, January 1999.

[2] C.F.J.Wu. On the convergence properties of the EM algorithm. *Ann. Stat.*, 11:95–103, 1983.

[3] Pei Jung Chung and Johann F. Böhme. Comparative convergence analysis of EM and SAGE algorithms in DOA estimation. Submitted to *IEEE Trans. Signal Processing*, 2000.

[4] A. P. Dempster, N. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B39:1–38, 1977.

[5] Meir Feder and Ehud Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. ASSP*, 36(4):477–489, April 1988.

[6] Jeffrey A. Fessler and Alfred O. Hero. Space–alternating generalized expectation–maximization algorithm. *IEEE Trans. Signal Processing*, 42(10):2664–2677, October 1994.

[7] G. H. Golub and C. F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, 1989.

[8] Wolfgang Hackbusch. *Iterative Solution of Large Sparse Systems of Equations*. Springer, 1994.

[9] Roger A. Horn and Charles R.Johnson. *Matrix Analysis*. 1985, Cambridge University Press.

[10] X.L. Meng and D. van Dyk. The EM algorithm – an old folk song sung to the fast tune. *Journal of the Royal Statistical Society, Ser. B.*, 59:511–567, 1997.