Joint Channel and Frequency Offset Estimation in Distributed MIMO Flat-Fading Channels

The-Hanh Pham, A. Nallanathan, *Senior Member, IEEE,* and Ying-Chang Liang, *Senior Member, IEEE*

***Abstract*— In this paper, we consider the problem of joint channel and frequency offset estimation in a flat-fading multi-input multi-output (MIMO) system. We assume that each pair of transmit and receive antennas has a different frequency offset. We present two computationally efficient iterative algorithms based on expectation conditional maximization (ECM) and space-alternating generalized expectation-maximization (SAGE) algorithms. The mean-square-error (MSE) performance of the interested parameters for the proposed algorithms is compared with the Cramér-Rao Bound (CRB). Simulation results show that the proposed iterative algorithms achieve the CRB and overcome the drawbacks of existing algorithms.**

***Index Terms*— Channel estimation, flat-fading channel, frequency offset estimation, multi-input multi-ouput (MIMO).**

I. Introduction

MULTI-ANTENNA transmission over multi-input multi-output (MIMO) channels has been proved to be effective in combating multipath fading, as well as increasing the channel capacity [1], [2]. In practice, coherent detection requires accurate channel and frequency offset information, thus channel and frequency offset estimation has become a critical task in modern wireless communication systems. In conventional MIMO systems, the transmit/receive antennas are colocated, thus they usually share one oscillator, and it is usually assumed that there is only one frequency offset within the system [3]–[9]. Recently, there is an increasing interest in the research of the distributed MIMO systems [10]–[12] where each of the transmit antennas is utilized by one user and the receive antennas are distributed in various locations in order to compensate for long-term shadowing fading. One typical distributed MIMO scenario is the cellular systems where several cell edge users communicate with several base stations. In this case, each transmit/receive antenna is equipped with its own oscillator, thus different transmit-receiver pair may have different frequency offset.

Maximum-Likelihood (ML) estimation of channel coefficients and frequency offsets in a flat fading MIMO channels with the assumption of having different frequency offsets between different transmit and receive antenna pairs has been studied in [13]. In this work, the authors stated that the ML estimation is a multi-dimensional minimization problem and thus has a very high computational complexity. Then, they introduced two computationally efficient algorithms which require that when one transmit antenna transmits a training symbol, the others do not transmit anything (i.e., other antennas are off). Consequently, it increases the dynamic range of the power amplifier [14]. Furthermore, it was pointed out that there exists numerical problems when the frequency offsets are estimated using the popular training sequences, e.g., sequence consisting of all ones, or if the frequency offsets are close to each other. This is due to the fact that the involved matrices are rank-deficient. To overcome this drawback, in [15], the authors proposed a correlation-based method for frequency offset estimation. This method, however, introduces an error floor in the MSE performance due to the existence of the interference in a multi-antenna system. In addition, if the frequency offset is too large, this method does not perform well.

Recently, an iterative method has been introduced in [14], where the frequency offsets are obtained using the method of [15]. Based on these values, coarse estimates of channel coefficients are obtained using least-square method. Then, the interference from unintended transmit antennas is subtracted from the received signal and the process is repeated. This method does not have the error floor as compared to in [15]. However, the performance does not reach the CRB. Furthermore, the estimation of frequency offsets is inherited from [15], therefore, whenever a large frequency offset estimation range is required, the performance becomes worse.

In this paper, we propose two iterative algorithms to estimate the channel coefficients and frequency offsets in a flat-fading MIMO channel. We assume that each pair of transmit-receive antenna has a distinct frequency offset value. Our proposed algorithms decouple the multi-dimensional optimization problem into many one-dimensional optimization problems where the channel coefficient and frequency offset of each pair of transmit-receive antenna can be determined separately. Simulations show that the proposed methods reach the CRB while avoiding the drawbacks of methods in [13]–[15].

The organization of this paper is as follows. Section II presents an overview of expectation maximization (EM) and EM-type [1] algorithms. Section III presents the system model and ML estimation of channel and frequency offsets. Section IV describes the two proposed iterative algorithms. Simulation results are presented in Section V. Finally, conclusions

Manuscript received August 20, 2006; revised February 13, 2007, May 11, 2007, and July 6, 2007; accepted July 20, 2007. The associate editor coordinating the review of this paper and approving it for publication was Y. Ma.

T.-H. Pham, and A. Nallanathan are with the Department of Electrical & Computer Engineering, National University of Singapore, 4, Engineering Drive 3, Block E4, 119260 (e-mail: g0301029@nus.edu.sg; elena@nus.edu.sg).

Y.-C. Liang is with the Institute for Infocomm Research (I^{2}R), 21 Heng Mui Keng Terrace, Singapore 119613 (e-mail: ycliang@i2r.a-star.edu.sg).

Digital Object Identifier 10.1109/TWC.2008.060605.

[1]In this paper, we refer the classical EM algorithm simply as EM algorithm. EM-type algorithms refer to ECM and SAGE.

are drawn in Section VI.

Notations: The capital bold letters denote matrices and the small bold letters denote row/column vectors; transpose and Hermitian of a vector/matrix are denoted by $(\cdot)^T$ and $(\cdot)^H$, respectively; $\Re\{a\}$, $\Im\{a\}$, $|a|$ and a^* denote the real part, imaginary part, absolute value and conjugate of a complex number a, respectively; $\odot$ denotes the element-wise product of two vectors/matrices. $\boldsymbol{I}_N$ denotes the identity matrix of size N.

II. Overview of EM and EM-type Algorithms

In this section, we give an overview of EM and EM-type algorithms which are used to find the ML estimate with a lower complexity. Let $\boldsymbol{\theta}$ denote a possibly vector-valued parameter to be estimated from a possibly vector-valued observation $\boldsymbol{y}$ with probability density $f(\boldsymbol{y}|\boldsymbol{\theta})$. The ML estimate of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; f(\boldsymbol{y}|\boldsymbol{\theta})$. This ML estimate usually has a very high complexity. Therefore, EM [16]–[18] and EM-type algorithms are proposed to find the ML estimate in an iterative manner.

A. EM Algorithm

The derivation of EM algorithm relies on the concept of a hypothesis, so-called *complete data space* $\boldsymbol{x}$. The observed random variable $\boldsymbol{y}$, which is referred to as *incomplete data space*, is related to $\boldsymbol{x}$ by a mapping $\boldsymbol{y} = g(\boldsymbol{x})$. The function g is a many-to-one transformation. Since $\boldsymbol{x}$ is not observable, at the m^{th} iteration, the EM algorithm computes its first step, called expectation step (E-step), which gives

$$Q\big(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[m]}\big) = E\big\{\log f(\boldsymbol{x}|\boldsymbol{\theta})|\boldsymbol{y}, \hat{\boldsymbol{\theta}}^{[m]}\big\}. \tag{1}$$

In the second step, called maximization step (M-step), the parameter vector is updated according to

$$\hat{\boldsymbol{\theta}}^{[m+1]} = \arg\max_{\boldsymbol{\theta}} \; Q\big(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[m]}\big). \tag{2}$$

The ability of the EM algorithm to find a global maximum depends on the initialization $\hat{\boldsymbol{\theta}}^{[0]}$. The convergence rate of the EM algorithm is inversely related to the conditional Fisher information matrix of $\boldsymbol{x}$, $\boldsymbol{y}$ [19]. This rate is very slow when the dimension of the complete data is large.

B. Expectation Conditional Maximization (ECM) Algorithm

In some cases, when the M-step of EM algorithm is too complicated, the ECM algorithm can be used to simplify the computation. The ECM algorithm [20] replaces the complicated M-step of EM algorithm by a series of smaller and less complicated steps. Specifically, if the parameter vector $\boldsymbol{\theta}$ can be divided into M groups of $\boldsymbol{\theta}_l$, $l = 1, 2, \cdots, M$, then the M-step of EM algorithm at the m^{th} iteration can be performed by M smaller steps in which $\boldsymbol{\theta}_l$ is updated at the l^{th} step, $l = 1, 2, \cdots, M$, while $\boldsymbol{\theta}_v$'s, $v \neq l$ are fixed at their most updated values. The l^{th} step consists of:

$$\begin{aligned}\text{Finding}: \; & \hat{\boldsymbol{\theta}}_l^{[m+1]} = \arg\max_{\boldsymbol{\theta}_l} \; Q\big(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[m]}\big)\big|_{\boldsymbol{\theta}_v = \hat{\boldsymbol{\theta}}_v^{[m]},\; v\neq l} \\ \text{Updating}: \; & \hat{\boldsymbol{\theta}}_l^{[m]} = \hat{\boldsymbol{\theta}}_l^{[m+1]}\end{aligned} \tag{3}$$

C. Space-Alternating Generalized Expectation-Maximization (SAGE) Algorithm

Despite the versatility of EM algorithm, it has slow convergence rate. To overcome this problem, the SAGE algorithm [19] has been proposed. The idea of the SAGE algorithm is to divide the parameter vector into smaller groups and update them sequentially. While updating one group of parameters, the other groups remain unchanged. In SAGE algorithm, instead of one large complete data, an *admissible hidden-data space* is introduced for each group of parameters. Suppose S is an index set [19], $\boldsymbol{\theta}_S$ denotes a vector consists of element(s) of $\boldsymbol{\theta}$ indexed by the members of S, $\boldsymbol{\theta}_{\tilde{S}}$ is the vector consists of all remaining element(s) of $\boldsymbol{\theta}$. Let $\boldsymbol{x}^S$ be the admissible hidden-data space for $\boldsymbol{\theta}_S$. SAGE algorithm, to update $\boldsymbol{\theta}_S$, consists of two steps:

E-step: Determine the conditional expectation of the log-likelihood of

$$Q_S\big(\boldsymbol{\theta_S}|\hat{\boldsymbol{\theta}}^{[m]}\big) = E\big\{\log f\big(\boldsymbol{x}^S|\boldsymbol{\theta}_S, \hat{\boldsymbol{\theta}}_{\tilde{S}}^{[m]}\big)|\boldsymbol{y}, \hat{\boldsymbol{\theta}}^{[m]}\big\}. \tag{4}$$

M-step: Maximize (4) to find

$$\hat{\boldsymbol{\theta}}_S^{[m+1]} = \arg\max_{\boldsymbol{\theta}_S} \; Q_S\big(\boldsymbol{\theta_S}|\hat{\boldsymbol{\theta}}^{[m]}\big). \tag{5}$$

III. System Model and ML Estimation

Consider a MIMO system with N_T transmit and N_R receive antennas operating under a flat-fading environment. We assume that each transmit-receive antenna pair has a different frequency offset. The received signal at the k^{th} receive antenna at time t can be written as [13]

$$y_k(t) = \sum_{l=1}^{N_T} h_{k,l} e^{j w_{k,l} t} s_l(t) + n_k(t), \;\; t = 1, 2, \cdots, N \tag{6}$$

where

- $\{s_l(t)\}_{t=1}^N$ is the sequence of symbols transmitted from the l^{th} transmit antenna.
- $h_{k,l}$ and $w_{k,l}$ are the channel coefficient and frequency offset between the l^{th} transmit antenna and the k^{th} receive antenna, respectively. $h_{k,l}$ and $w_{k,l}$ are assumed to be unknown and unchanged over the interval of $t = 1, 2, \cdots, N$. Here, we make a standard assumption that $h_{k,l}$'s are statistically independent and complex-valued Gaussian random variables with zero-mean and variance of 1.
- $\{n_k(t)\}_{t=1}^N$ is a sequence of zero-mean, independent and identically distributed complex-valued Gaussian random variables having variance of σ^2. Noise sequences at N_R receive antennas are statistically independent.

For the purpose of joint channel and frequency offset estimations, the training sequence $\{s_l(t)\}_{t=1}^N$ is assumed to be known.

If we define:

$$\boldsymbol{y}_k = [y_k(1) \; y_k(2) \; \cdots \; y_k(N)]^T \tag{7}$$

$$\boldsymbol{h}_k = [h_{k,1} \; h_{k,2} \; \cdots \; h_{k,N_T}]^T \tag{8}$$

$$\boldsymbol{w}_k = [w_{k,1} \; w_{k,2} \; \cdots \; w_{k,N_T}]^T \tag{9}$$

$$\boldsymbol{S}_{\boldsymbol{w}_k} =$$

$$\begin{bmatrix} s_1(1)e^{jw_{k,1}} & s_2(1)e^{jw_{k,2}} & \cdots & s_{N_T}(1)e^{jw_{k,N_T}} \\ s_1(2)e^{j2w_{k,1}} & s_2(2)e^{j2w_{k,2}} & \cdots & s_{N_T}(2)e^{j2w_{k,N_T}} \\ \vdots & \vdots & \ddots & \vdots \\ s_1(N)e^{jNw_{k,1}} & s_2(N)e^{jNw_{k,2}} & \cdots & s_{N_T}(N)e^{jNw_{k,N_T}} \end{bmatrix} \tag{10}$$

Then the ML estimation of $\{h_{k,l}\}$ and $\{w_{k,l}\}$ is achieved through minimizing of the following log-likelihood function [13]:

$$\Lambda = \sum_{k=1}^{N_R} \|\boldsymbol{y}_k - \boldsymbol{S}_{\boldsymbol{w}_k}\boldsymbol{h}_k\|^2. \tag{11}$$

Since the term in the sum of (11) depends only on $\boldsymbol{w}_k$ and $\boldsymbol{h}_k$, each set of $\boldsymbol{h}_k$ and $\boldsymbol{w}_k$ can be derived by minimizing separate functions. Specifically, the ML estimation of $\boldsymbol{h}_k$ and $\boldsymbol{w}_k$ amounts to the minimization of

$$\Lambda_k = \|\boldsymbol{y}_k - \boldsymbol{S}_{\boldsymbol{w}_k}\boldsymbol{h}_k\|^2. \tag{12}$$

For a given value of $\boldsymbol{w}_k$, the $\boldsymbol{h}_k$ that minimizes (12) is given by

$$\hat{\boldsymbol{h}}_k = (\boldsymbol{S}_{\boldsymbol{w}_k}^H \boldsymbol{S}_{\boldsymbol{w}_k})^{-1} \boldsymbol{S}_{\boldsymbol{w}_k}^H \boldsymbol{y}_k. \tag{13}$$

Substituting (13) back to (12), the frequency offset estimation is as follows

$$\hat{\boldsymbol{w}}_k = \arg\max_{\boldsymbol{w}_k} \ \boldsymbol{y}_k^H \boldsymbol{S}_{\boldsymbol{w}_k} (\boldsymbol{S}_{\boldsymbol{w}_k}^H \boldsymbol{S}_{\boldsymbol{w}_k})^{-1} \boldsymbol{S}_{\boldsymbol{w}_k}^H \boldsymbol{y}_k. \tag{14}$$

Equation (14) is a multi-dimensional minimization problem which has a high computational complexity.

IV. Proposed Iterative Joint Channel and Frequency Offsets Estimators

It is evident that the channel coefficient and frequency offset estimation problem for a MIMO system can be considered as N_R independent estimation problems for N_R MISO (Multi-Input Single-Output) systems. Hence, in this section, without loss of generality, we only present the algorithms to estimate the channel coefficients and frequency offsets from all transmit antennas to the k^{th} receive antenna using the received signal at the k^{th} receive antenna and the information of training sequences from all transmit antennas.

A. Algorithm 1: ECM Based Approach

If we define $\boldsymbol{s}_l = [s_l(1)\ s_l(2)\ \cdots\ s_l(N)]^T$ as the training signal vector transmitted from the l^{th} transmit antenna and $\boldsymbol{e}_{k,l} = [e^{jw_{k,l}}\ e^{j2w_{k,l}}\ \ldots\ e^{jNw_{k,l}}]^T$, then the received signal vector $\boldsymbol{y}_k$ in (7) can be written as

$$\boldsymbol{y}_k = \sum_{l=1}^{N_T} (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l}) h_{k,l} + \boldsymbol{n}_k, \tag{15}$$

where $\boldsymbol{n}_k = [n_k(1)\ n_k(2)\ \cdots\ n_k(N)]^T$ and $\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$. In the EM terminology, the observed signal vector $\boldsymbol{y}_k$ is the incomplete data space. The parameter to be estimated is $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T\ \cdots\ \boldsymbol{\theta}_l^T\ \cdots\ \boldsymbol{\theta}_{N_T}^T]^T$, where $\boldsymbol{\theta}_l = [w_{k,l}\ h_{k,l}]^T$ is the two parameters corresponding to the pair of the l^{th} transmit antenna and the k^{th} receive antenna. Following [21], we define the complete data space as $\boldsymbol{z}_k = [\boldsymbol{z}_{k,1}\ \boldsymbol{z}_{k,2}\ \cdots\ \boldsymbol{z}_{k,N_T}]^T$ where

$$\boldsymbol{z}_{k,l} \triangleq (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l}) h_{k,l} + \boldsymbol{n}_{k,l},\ \ l = 1, 2, \cdots, N_T. \tag{16}$$

Thus, the relation between the complete data space $\boldsymbol{z}_l$ and incomplete data $\boldsymbol{y}_k$ is given by

$$\sum_{l=1}^{N_T} \boldsymbol{z}_{k,l} = \boldsymbol{y}_k. \tag{17}$$

In (16), $\boldsymbol{n}_{k,l}$'s are obtained by decomposing the total noise vector $\boldsymbol{n}_k$ into N_T components such that

$$\sum_{l=1}^{N_T} \boldsymbol{n}_{k,l} = \boldsymbol{n}_k \tag{18}$$

and $\boldsymbol{n}_{k,l}$'s are statistically independent, zero-mean Gaussian random vectors with covariance matrix of $\beta_l \sigma^2 \boldsymbol{I}_N$. The β_l's are non-zero real-valued numbers such that

$$\sum_{l=1}^{N_T} \beta_l = 1,\ \ \beta_l > 0. \tag{19}$$

There is no available strategy to choose the optimum set of $\{\beta_l\}_{l=1}^{N_T}$. In practice, however, β_l's are usually chosen to be equal, i.e., $\beta_l = \frac{1}{N_T}$ for all l.

To process further, we denote $\hat{\boldsymbol{\theta}}^{[m]} = \big[\hat{\boldsymbol{\theta}}_1^{[m]}\ \hat{\boldsymbol{\theta}}_2^{[m]}\ \cdots \hat{\boldsymbol{\theta}}_{N_T}^{[m]}\big]^T$ as the estimated value of $\boldsymbol{\theta}$ obtained after the $(m-1)^{th}$ iteration and forwarded to the m^{th} iteration in which $\hat{\boldsymbol{\theta}}_l^{[m]} = \big[\hat{w}_{k,l}^{[m]}\ \hat{h}_{k,l}^{[m]}\big]^T,\ l = 1, 2, \cdots, N_T$.

The proposed Algorithm 1 at the m^{th} iteration contains the E-step and M-step as follows.

1) E-step: In this step, we compute the expectation of the complete data space log-likelihood function given the parameter $\boldsymbol{\theta}$ and conditioned upon the incomplete data and the current estimated value of $\hat{\boldsymbol{\theta}}^{[m]}$, that is:

$$Q\big(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[m]}\big) \triangleq E\big\{\log f(\boldsymbol{z}_k|\boldsymbol{\theta})|\boldsymbol{y}_k, \hat{\boldsymbol{\theta}}^{[m]}\big\}. \tag{20}$$

Due to the statistical independence among $\boldsymbol{n}_{k,l}$'s, the probability density function of $\boldsymbol{z}_k$ as a function of $\boldsymbol{\theta}$ is:

$$\begin{aligned} f(\boldsymbol{z}_k|\boldsymbol{\theta}) &= \prod_{l=1}^{N_T} f(\boldsymbol{z}_{k,l}|\boldsymbol{\theta}_l) \\ &= \prod_{l=1}^{N_T} \frac{1}{(\pi\beta_l\sigma^2)^N} \exp\left\{-\frac{\|\boldsymbol{z}_{k,l} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l})h_{k,l}\|^2}{\beta_l\sigma^2}\right\}. \end{aligned} \tag{21}$$

Substituting (21) into (20), we obtain

$$\begin{aligned} Q\big(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[m]}\big) &= C_1 - \\ & E\Big\{\sum_{l=1}^{N_T} \frac{1}{\beta_l\sigma^2} \|\boldsymbol{z}_{k,l} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l})h_{k,l}\|^2 \big| \boldsymbol{y}_k, \hat{\boldsymbol{\theta}}^{[m]}\Big\} \\ &= C_2 - \sum_{l=1}^{N_T} \frac{1}{\beta_l\sigma^2} \|\hat{\boldsymbol{z}}_{k,l}^{[m]} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l})h_{k,l}\|^2, \end{aligned} \tag{22}$$

where

$$\hat{\boldsymbol{z}}_{k,l}^{[m]} \triangleq E\big\{\boldsymbol{z}_{k,l}\big|\boldsymbol{y}_k, \hat{\boldsymbol{\theta}}^{[m]}\big\}. \tag{23}$$

and C_1 and C_2 are two constants independent of $\boldsymbol{\theta}$.

Since $\boldsymbol{z}_{k,l}$ and $\boldsymbol{y}_k$ are jointly Gaussian distributed and satisfy (17), it is easy to have

$$\hat{\boldsymbol{z}}_{k,l}^{[m]} = \left(\boldsymbol{s}_l \odot \hat{\boldsymbol{e}}_{k,l}^{[m]}\right)\hat{h}_{k,l}^{[m]} + \beta_l\Big(\boldsymbol{y}_k - \sum_{v=1}^{N_T}\left(\boldsymbol{s}_v \odot \hat{\boldsymbol{e}}_{k,v}^{[m]}\right)\hat{h}_{k,v}^{[m]}\Big) \tag{24}$$

where $\hat{\boldsymbol{e}}_{k,v}^{[m]} = \left[e^{j\hat{w}_{k,v}^{[m]}}\ e^{j2\hat{w}_{k,v}^{[m]}}\ \cdots\ e^{jN\hat{w}_{k,v}^{[m]}}\right]^T$.

2) M-step: In this step, the updated value of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^{[m+1]}$, is determined as

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{[m+1]} &= \arg\max_{\boldsymbol{\theta}}\ Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{[m]}\right)\\ &= \arg\min_{\boldsymbol{\theta}}\ \sum_{l=1}^{N_T}\|\hat{\boldsymbol{z}}_{k,l}^{[m]} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l})h_{k,l}\|^2.\end{aligned} \tag{25}$$

We can easily see from (25) that the updating process of $\boldsymbol{\theta}$ can be decoupled into N_T updating processes of $\boldsymbol{\theta}_l$ for $l = 1, 2, \cdots, N_T$. Hence, we have the equation to determine $\hat{\boldsymbol{\theta}}_l^{[m+1]}$ as

$$\hat{\boldsymbol{\theta}}_l^{[m+1]} = \arg\min_{\boldsymbol{\theta}_l}\ \|\hat{\boldsymbol{z}}_{k,l}^{[m]} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l})h_{k,l}\|^2. \tag{26}$$

In contrast to the EM algorithm, where the updating process of parameters is taken place simultaneously, the ECM algorithm minimizes (26) in two steps. In the first step, (26) is minimized with respect to (w.r.t.) one of $(w_{k,l}, h_{k,l})$ while the others are kept at their most updated values. We denote $\hat{\boldsymbol{\theta}}_l^{[m+c/2]}$ as the estimate of $\boldsymbol{\theta}_l$ at c^{th} step of m^{th} iteration of the ECM algorithm, $c = 1, 2$. Then, the M-step of the ECM algorithm consists of the following two smaller steps.

a) Step 1: In this step, we determine the updated value of $w_{k,l}$ while $h_{k,l}$ is fixed at $\hat{h}_{k,l}^{[m]}$, i.e., we determine $\hat{\boldsymbol{\theta}}_l^{[m+1/2]} = \left[\hat{w}_{k,l}^{[m+1]}\ \hat{h}_{k,l}^{[m]}\right]^T$ where

$$\begin{aligned}\hat{w}_{k,l}^{[m+1]} &= \arg\min_{w_{k,l}}\ \|\hat{\boldsymbol{z}}_{k,l}^{[m]} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l})h_{k,l}\|^2\big|_{h_{k,l}=\hat{h}_{k,l}^{[m]}}\\ &= \arg\min_{w_{k,l}}\ \sum_{t=1}^{N}|\hat{z}_{k,l}^{[m]}(t) - s_l(t)e^{jw_{k,l}t}\hat{h}_{k,l}^{[m]}|^2\\ &= \arg\max_{w_{k,l}}\ \sum_{t=1}^{N}\Re\Big\{\left(\hat{z}_{k,l}^{[m]}(t)\right)^* s_l(t)\hat{h}_{k,l}^{[m]}e^{jw_{k,l}t}\Big\}.\end{aligned} \tag{27}$$

where $\hat{z}_{k,l}^{[m]}(t)$ is the t^{th} element of $\hat{\boldsymbol{z}}^{[m]}$, $t = 1, 2, \cdots, N$.

To handle the nonlinearity of (27), we can resort to Taylor's series expansion of $e^{jw_{k,l}t}$ around $\hat{w}_{k,l}^{[m]}$ to the second-order term as:

$$\begin{aligned}e^{jw_{k,l}t} \approx\ & e^{j\hat{w}_{k,l}^{[m]}t} + (w_{k,l} - \hat{w}_{k,l}^{[m]})(jt)e^{j\hat{w}_{k,l}^{[m]}t}\\ &+ \frac{1}{2}(w_{k,l} - \hat{w}_{k,l}^{[m]})^2(jt)^2e^{j\hat{w}_{k,l}^{[m]}t}.\end{aligned} \tag{28}$$

Therefore, (27) can be written after dropping some terms that do not relate to $w_{k,l}$ as in (29) on the top of next page.

Differentiating the function inside $\{\cdot\}$ of (29) w.r.t. $w_{k,l}$ and equating the result to zero we get the updated value $\hat{w}_{k,l}^{[m+1]}$ as

$$\hat{w}_{k,l}^{[m+1]} = \hat{w}_{k,l}^{[m]} - \frac{\sum_{t=1}^{N} t\Im\Big\{\left(\hat{z}_{k,l}^{[m]}(t)\right)^* s_l(t)\hat{h}_{k,l}^{[m]}e^{j\hat{w}_{k,l}^{[m]}t}\Big\}}{\sum_{t=1}^{N} t^2\Re\Big\{\left(\hat{z}_{k,l}^{[m]}(t)\right)^* s_l(t)\hat{h}_{k,l}^{[m]}e^{j\hat{w}_{k,l}^{[m]}t}\Big\}}. \tag{30}$$

TABLE I

SUMMARY OF ALGORITHM 1 FOR THE k^{th} RECEIVE ANTENNA

1. **Initialization**
   Obtain $\hat{w}_{k,l}^{[0]}$ and $\hat{h}_{k,l}^{[0]}$ for $l = 1, 2, \cdots, N_T$.
2. **ECM**
   For $m = 0, 1, \cdots$
   - For $l = 1, 2, \cdots, N_T$
     Compute:
     $\hat{\boldsymbol{z}}_{k,l}^{[m]} = (\boldsymbol{s}_l \odot \hat{\boldsymbol{e}}_{k,l}^{[m]})\hat{h}_{k,l}^{[m]} + \beta_l\Big(\boldsymbol{y}_k - \sum_{v=1}^{N_T}(\boldsymbol{s}_v \odot \hat{\boldsymbol{e}}_{k,v}^{[m]})\hat{h}_{k,v}^{[m]}\Big)$.
   - For $l = 1, 2, \cdots, N_T$
     Compute:
     $\hat{w}_{k,l}^{[m+1]} = \hat{w}_{k,l}^{[m]} - \frac{\sum_{t=1}^{N} t\Im\left\{\left(\hat{z}_{k,l}^{[m]}(t)\right)^* s_l(t)\hat{h}_{k,l}^{[m]}e^{j\hat{w}_{k,l}^{[m]}t}\right\}}{\sum_{t=1}^{N} t^2\Re\left\{\left(\hat{z}_{k,l}^{[m]}(t)\right)^* s_l(t)\hat{h}_{k,l}^{[m]}e^{j\hat{w}_{k,l}^{[m]}t}\right\}}$.
     Compute:
     $\hat{h}_{k,l}^{[m+1]} = \frac{1}{\sum_{t=1}^{N}|s_l(t)|^2}\sum_{t=1}^{N}\frac{\hat{z}_{k,l}^{[m]}(t)s_l^*(t)}{e^{j\hat{w}_{k,l}^{[m+1]}t}}$.

Equation (30) is valid only when the function inside $\{\cdot\}$ in (29) is convex. It is hard to prove this property analytically from (29) because it consists of estimated parameters which change from iteration to iteration. However, from our simulations, we observe that the function is always convex; hence, (30) provides satisfied results as will be illustrated in the Section V.

b) Step 2: In this step, the updated value of $h_{k,l}$, $\hat{h}_{k,l}^{[m+1]}$ is determined, where $w_{k,l}$ is set at its newest value of $\hat{w}_{k,l}^{[m+1]}$. Hence, we have $\hat{\boldsymbol{\theta}}_l^{[m+1]} = \left[\hat{w}_{k,l}^{[m+1]}\ \hat{h}_{k,l}^{[m+1]}\right]^T$ where

$$\begin{aligned}\hat{h}_{k,l}^{[m+1]} &= \arg\min_{h_{k,l}}\ \|\hat{\boldsymbol{z}}_{k,l}^{[m]} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l})h_{k,l}\|^2\big|_{w_{k,l}=\hat{w}_{k,l}^{[m+1]}}\\ &= \arg\min_{h_{k,l}}\ \sum_{t=1}^{N}|\hat{z}_{k,l}^{[m]}(t) - s_l(t)e^{j\hat{w}_{k,l}^{[m+1]}t}h_{k,l}|^2\end{aligned} \tag{31}$$

which can be solved by

$$\hat{h}_{k,l}^{[m+1]} = \frac{1}{\sum_{t=1}^{N}|s_l(t)|^2}\sum_{t=1}^{N}\frac{\hat{z}_{k,l}^{[m]}(t)s_l^*(t)}{e^{j\hat{w}_{k,l}^{[m+1]}t}}. \tag{32}$$

Algorithm 1 for the k^{th} receive antenna is summarized in the Table 1. By changing k from 1 to N_R (or equivalently, by repeating the Algorithm 1 N_R times), we obtain the channel coefficients and frequency offsets for the MIMO system.

B. Algorithm 2: SAGE-ECM Based Approach

In EM-type algorithm, the convergence rate is inversely proportional to the Fisher information of its complete data [19]. In the above algorithm, the noise variance is distributed over $\boldsymbol{z}_{k,l}$ for all value of l. Therefore the Fisher information of $\boldsymbol{z}_{k,l}$ is relatively large. To improve the convergence rate, we use SAGE algorithm in which the parameter $\boldsymbol{\theta}$ is divided into N_T groups of $\boldsymbol{\theta}_l$, $l = 1, 2, \cdots, N_T$. For each group, a hidden data space must be chosen [19]. The update process of any group is taken place while the others are fixed at their latest updated values.

$$\hat{w}_{k,l}^{[m+1]} = \arg\max_{w_{k,l}} \Big\{ -(w_{k,l} - \hat{w}_{k,l}^{[m]}) \sum_{t=1}^{N} t \Im\Big\{ \big(\hat{z}_{k,l}^{[m]}(t)\big)^* s_l(t) \hat{h}_{k,l}^{[m]} e^{j\hat{w}_{k,l}^{[m]} t} \Big\} - \frac{1}{2}(w_{k,l} - \hat{w}_{k,l}^{[m]})^2 \sum_{t=1}^{N} t^2 \Re\Big\{ \big(\hat{z}_{k,l}^{[m]}(t)\big)^* s_l(t) \hat{h}_{k,l}^{[m]} e^{j\hat{w}_{k,l}^{[m]} t} \Big\} \Big\}. \tag{29}$$

Specifically, for the group of $\boldsymbol{\theta}_l$, l belongs to the set $\{1, 2, \cdots, N_T\}$. The hidden data space is defined as

$$\boldsymbol{x}_{k,l} \triangleq (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l}) h_{k,l} + \boldsymbol{n}. \tag{33}$$

Note that we associate all the noise to the hidden data space of $\boldsymbol{\theta}_l$ to reduce its Fisher information to increase the convergence rate.

Following is the update process of $\boldsymbol{\theta}_l$ at the m^{th} iteration which also consists of E-step and M-step. Note that there is a total of N_T times of these two steps at the m^{th} iteration to update all $\boldsymbol{\theta}_l$, $l = 1, 2, \cdots, N_T$.

1) E-step: In this step, we determine the expectation of the hidden data space log-likelihood function given the parameter $\boldsymbol{\theta}_l$ while $\boldsymbol{\theta}_v$'s, $v \neq$, are kept at $\hat{\boldsymbol{\theta}}_v^{[m]}$ and conditioned upon the observed vector $\boldsymbol{y}_k$ as well as current estimate of $\hat{\boldsymbol{\theta}}^{[m]}$ as:

$$Q(\boldsymbol{\theta}_l | \hat{\boldsymbol{\theta}}^{[m]}) = E\Big\{ \log f\big(\boldsymbol{x}_{k,l} | \boldsymbol{\theta}_l, \{\hat{\boldsymbol{\theta}}_v^{[m]}\}_{v \neq l}\big) \big| \boldsymbol{y}_k, \hat{\boldsymbol{\theta}}^{[m]} \Big\}. \tag{34}$$

We have

$$\begin{aligned} & f\big(\boldsymbol{x}_{k,l} | \boldsymbol{\theta}_l, \{\hat{\boldsymbol{\theta}}_v^{[m]}\}_{v \neq l}\big) = f(\boldsymbol{x}_{k,l} | \boldsymbol{\theta}_l) \\ &= \frac{1}{(\pi\sigma^2)^N} \exp\Big\{ -\frac{1}{\sigma^2} \| \boldsymbol{x}_{k,l} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l}) h_{k,l} \|^2 \Big\}. \end{aligned} \tag{35}$$

After replacing (35) into (34), we obtain

$$\begin{aligned} Q(\boldsymbol{\theta}_l | \hat{\boldsymbol{\theta}}^{[m]}) &= C_3 - \frac{1}{\sigma^2} E\Big\{ \| \boldsymbol{x}_{k,l} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l}) h_{k,l} \|^2 \big| \boldsymbol{y}_k, \hat{\boldsymbol{\theta}}^{[m]} \Big\} \\ &= C_4 - \frac{1}{\sigma^2} \| \hat{\boldsymbol{x}}_{k,l}^{[m]} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l}) h_{k,l} \|^2, \end{aligned} \tag{36}$$

where

$$\begin{aligned} \hat{\boldsymbol{x}}_{k,l}^{[m]} &\triangleq E\big\{ \boldsymbol{x}_{k,l} | \boldsymbol{y}_k, \hat{\boldsymbol{\theta}}^{[m]} \big\} \\ &= \big(\boldsymbol{s}_l \odot \hat{\boldsymbol{e}}_{k,l}^{[m]}\big) \hat{h}_{k,l}^{[m]} + \Big(\boldsymbol{y}_k - \sum_{v=1}^{N_T} \big(\boldsymbol{s}_v \odot \hat{\boldsymbol{e}}_{k,v}^{[m]}\big) \hat{h}_{k,v}^{[m]} \Big) \\ &= \boldsymbol{y}_k - \sum_{v=1, v \neq l}^{N_T} \big(\boldsymbol{s}_v \odot \hat{\boldsymbol{e}}_{k,v}^{[m]}\big) \hat{h}_{k,v}^{[m]}. \end{aligned} \tag{37}$$

and C_3 and C_4 are two constants independent of $\boldsymbol{\theta}_l$.

2) M-step: In this step, the updated value of $\boldsymbol{\theta}_l$, $\hat{\boldsymbol{\theta}}_l^{[m+1]}$, is determined by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_l^{[m+1]} &= \arg\max_{\boldsymbol{\theta}_l} \; Q(\boldsymbol{\theta}_l | \hat{\boldsymbol{\theta}}_l^{[m]}) \\ &= \arg\min_{\boldsymbol{\theta}_l} \; \| \hat{\boldsymbol{x}}_{k,l}^{[m]} - (\boldsymbol{s}_l \odot \boldsymbol{e}_{k,l}) h_{k,l} \|^2. \end{aligned} \tag{38}$$

This equation can be solved like the M-step of the previous algorithm where ECM is deployed, i.e., it consists of two small steps and elements of $\boldsymbol{\theta}_l = [w_{k,l} \; h_{k,l}]^T$ are updated sequentially. Here, we only state the results.

TABLE II

SUMMARY OF ALGORITHM 2 FOR THE k^{th} RECEIVE ANTENNA

1. **Initialization**
   Obtain $\hat{w}_{k,l}^{[0]}$ and $\hat{h}_{k,l}^{[0]}$ for $l = 1, 2, \cdots, N_T$.
2. **SAGE-ECM**
   For $m = 0, 1, \cdots$
   - For $l = 1, 2, \cdots, N_T$
     Compute:
     $\hat{\boldsymbol{x}}_{k,l}^{[m]} = (\boldsymbol{s}_l \odot \hat{\boldsymbol{e}}_{k,l}^{[m]}) \hat{h}_{k,l}^{[m]} + \Big(\boldsymbol{y}_k - \sum_{v=1}^{N_T} (\boldsymbol{s}_v \odot \hat{\boldsymbol{e}}_{k,v}^{[m]}) \hat{h}_{k,v}^{[m]} \Big)$.
     Compute:
     $\hat{w}_{k,l}^{[m+1]} = \hat{w}_{k,l}^{[m]} - \frac{\sum_{t=1}^{N} t \Im\left\{ \left(\hat{x}_{k,l}^{[m]}(t)\right)^* s_l(t) \hat{h}_{k,l}^{[m]} e^{j\hat{w}_{k,l}^{[m]} t} \right\}}{\sum_{t=1}^{N} t^2 \Re\left\{ \left(\hat{x}_{k,l}^{[m]}(t)\right)^* s_l(t) \hat{h}_{k,l}^{[m]} e^{j\hat{w}_{k,l}^{[m]} t} \right\}}$.
     Compute:
     $\hat{h}_{k,l}^{[m+1]} = \frac{1}{\sum_{t=1}^{N} |s_l(t)|^2} \sum_{t=1}^{N} \frac{\hat{x}_{k,l}^{[m]}(t) s_l^*(t)}{e^{j\hat{w}_{k,l}^{[m+1]} t}}$.
     Update : $\hat{w}_{k,l}^{[m]} = \hat{w}_{k,l}^{[m+1]}$, $\hat{h}_{k,l}^{[m]} = \hat{h}_{k,l}^{[m+1]}$.

a) Step 1: In this step, $\hat{w}_{k,l}^{[m+1]}$ is determined by:

$$\hat{w}_{k,l}^{[m+1]} = \hat{w}_{k,l}^{[m]} - \frac{\sum_{t=1}^{N} t \Im\Big\{ \big(\hat{x}_{k,l}^{[m]}(t)\big)^* s_l(t) \hat{h}_{k,l}^{[m]} e^{j\hat{w}_{k,l}^{[m]} t} \Big\}}{\sum_{t=1}^{N} t^2 \Re\Big\{ \big(\hat{x}_{k,l}^{[m]}(t)\big)^* s_l(t) \hat{h}_{k,l}^{[m]} e^{j\hat{w}_{k,l}^{[m]} t} \Big\}}. \tag{39}$$

where $\hat{x}_{k,l}^{[m]}(t)$ is the t^{th} element of $\hat{\boldsymbol{x}}_{k,l}^{[m]}$ in (37), $t = 1, 2, \cdots, N$.

b) Step 2: Here, the updated value $\hat{h}_{k,l}^{[m+1]}$ is calculated as:

$$\hat{h}_{k,l}^{[m+1]} = \frac{1}{\sum_{t=1}^{N} |s_l(t)|^2} \sum_{t=1}^{N} \frac{\hat{x}_{k,l}^{[m]}(t) s_l^*(t)}{e^{j\hat{w}_{k,l}^{[m+1]} t}}. \tag{40}$$

Algorithm 2 is named as SAGE-ECM since it is derived by combining ECM and SAGE. The algorithm is summarized in the Table II. By repeating it N_R times (by changing k from 1 to N_R), we obtain the desired parameters for MIMO system.

V. SIMULATION RESULTS

Computer simulations have been carried out to evaluate the performance of the proposed algorithms in some system setups. The performance of the proposed algorithms in terms of MSE and bit error rate (BER) is also compared with the existing methods.

A. Example 1: 2×1 *System With Fixed Channel and Fixed Offset*

In this subsection, we consider a system with $N_T = 2$ transmit antennas and $N_R = 1$ receive antenna. Following [13], the frequency offsets between the receive antenna

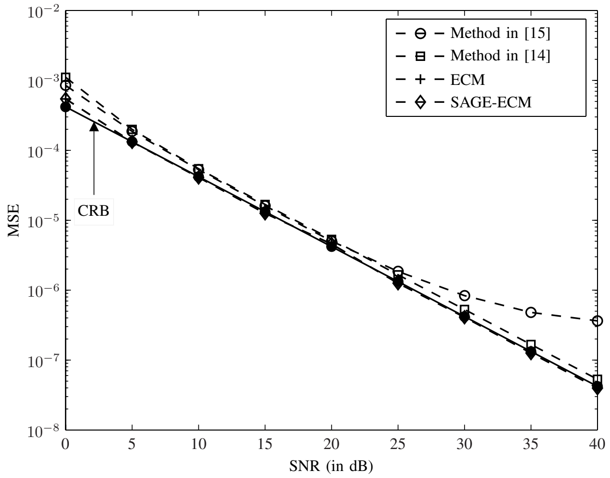

Fig. 1. Comparison of MSE performances of $w_{1,2}$ of [15], [14], ECM and SAGE-ECM algorithms.

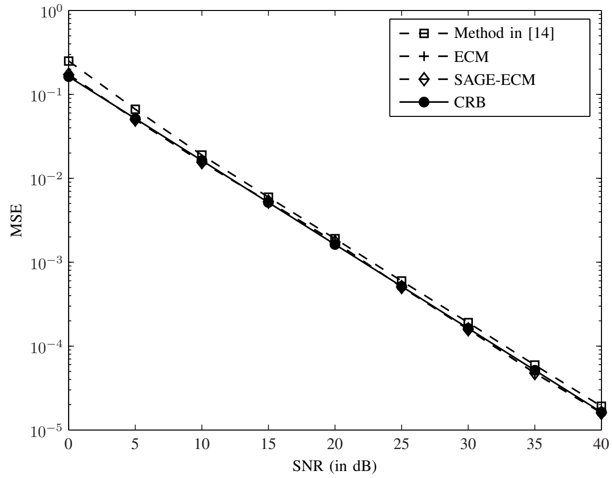

Fig. 2. Comparison of MSE performances of $h_{1,2}$ of [14], ECM and SAGE-ECM algorithms.

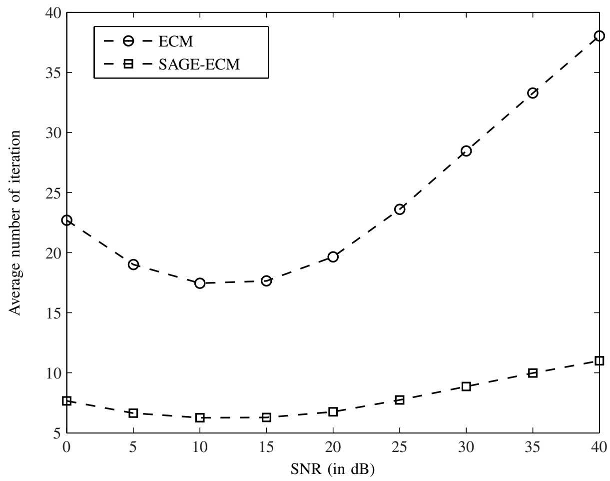

Fig. 3. Average number of iterations of ECM and SAGE-ECM algorithms.

and the two transmit antennas are $\boldsymbol{w} = [w_{1,1}\ \ w_{1,2}]^T = 2\pi[0.01\ \ 0.015]^T$. Note that these two frequency offsets are quite close to each other and the ML estimation [13] should encounter the numerical problems. A specific channel is used for all simulations in this case [13]: $\boldsymbol{h}_1 = [h_{1,1}\ \ h_{1,2}]^T = [0.2929 + 0.5169i\ 0.1074 - 0.9303i]^T$. During the evaluation, the length of pilot sequences from the transmit antennas is $N = 32$. Each training sequence is taken from a row of a Hadamard matrix with appropriate size. Hence, we can consider that it consists of P repetitive blocks, P is called the correlator length in [15] and the length controls the range of estimated frequency offsets. Correlator length is taken as 8. Our proposed algorithms stop when the difference between log-likelihood function of the two consecutive iterations is less than 0.001.

Firstly, we compare the performances of our proposed ECM and SAGE-ECM based algorithms. We present the results for the pair of the second transmit antenna and the receive antenna. We use the method in [15] to have the initialization values for frequency offsets. After having these values, (13) is used to get the initial estimates for channel coefficients. In Fig. 1 and Fig. 2, the MSE of $w_{1,2}$ and $h_{1,2}$ are illustrated, respectively.

We can see that the performance of our proposed algorithms reach the CRB. Readers are referred to [13] for detailed derivations and results of CRB. In order to have the above results, the required average number of iterations for both algorithms is presented in Fig. 3. We can see that the SAGE-ECM based algorithm greatly reduces the complexity compared to the ECM based algorithm. Note that, this is a fair comparison because each iteration of both algorithms has roughly the same complexity. It can be seen that the required average number of iterations become high at low- and high-SNR regions. This is because the performance of initialization is far away from CRB for the above SNR regions than the mid-SNR one. The comparisons our algorithms with [15] and [14] are also presented in Fig. 1 and Fig. 2. Specifically, Fig. 1 presents the obtained MSE of $w_{1,2}$ for different algorithms. The algorithm of [14] overcomes the error floor of [15] at only high-SNR region. However, it is still around 1 dB worse than our proposed algorithms. Fig. 2 illustrates the result for $h_{1,2}$. The same observations can be obtained.

Finally, we compare the performance of different algorithms for different values of correlator length. Here, two values, 4 and 2, are examined. The method in [15] gives large range of estimated frequency offsets with small value of P. However, the estimation error increases. This phenomenon is illustrated in Fig. 4 for $w_{1,2}$. The method in [14] can resolve the error floor but the performance becomes worse with the decreasing value of P. This is because, basically, the process of estimating frequency offsets is taken from [15]. In the mean time, our proposed algorithms still reach the CRB from low SNR values. The same observation can be found in Fig. 2 for the performance of $h_{1,2}$. In Fig. 6, the average number of iterations for SAGE-ECM based algorithm for different values of P is illustrated. We can see that when SNR is high, the required iterations for two cases become the same. This is because when SNR is high, the performances of [15] (initialization for our proposed algorithms) for $P = 4$ and $P = 2$ are approximately the same. It again proves the importance of initialization for EM-type algorithms.

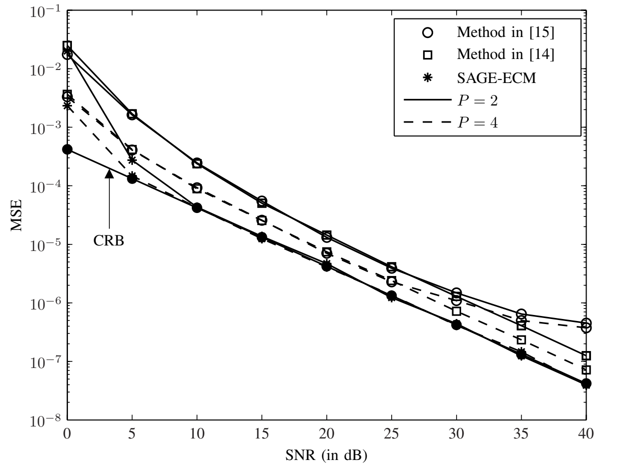

Fig. 4. Comparison of MSE performances of $w_{1,2}$ of [15], [14] and SAGE-ECM algorithm for different values of P.

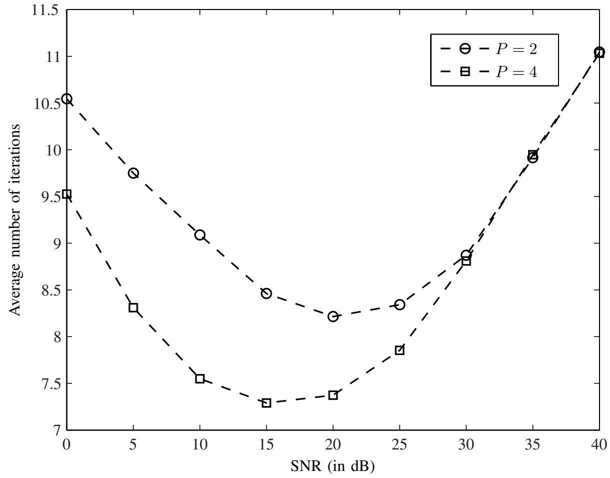

Fig. 6. Comparison of average number of iterations of SAGE-ECM algorithm for different values of P.

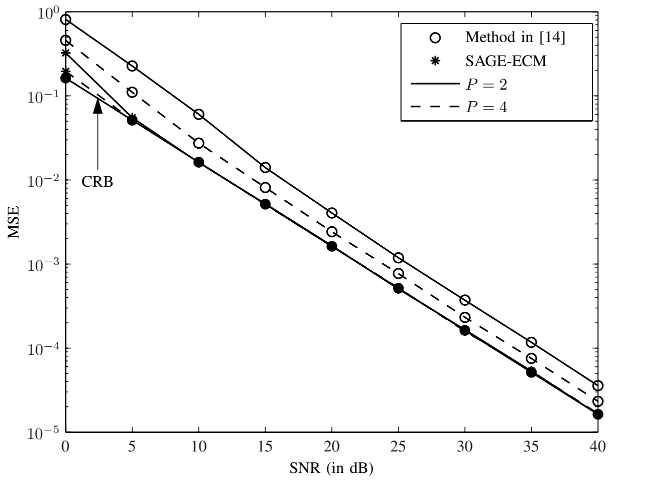

Fig. 5. Comparison of MSE performances of $h_{1,2}$ of [14] and SAGE-ECM algorithm for different values of P.

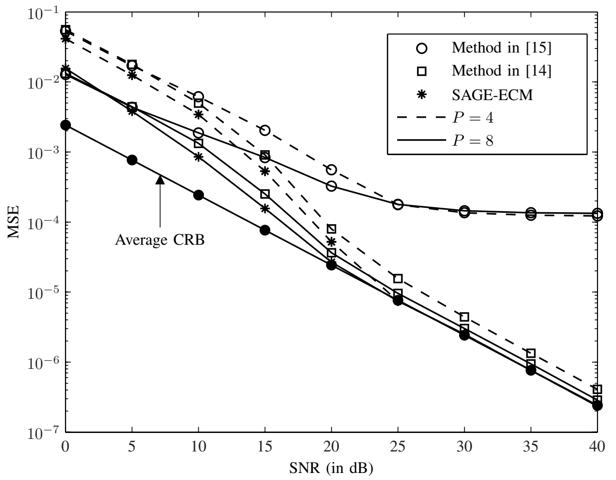

Fig. 7. Comparison of MSE performances of $w_{1,2}$ of [15], [14], and SAGE-ECM algorithms for 4 transmit antennas system.

B. Example 2: 4×1 System, Fading Channel and Fixed Offset

For the next case we consider a system with $N_T = 4$ transmit antennas and $N_R = 1$ receive antenna. The frequency offset values from the transmit antennas to the receive antenna are $\boldsymbol{w} = 2\pi[0.01\ 0.015\ 0.02\ 0.025]^T$ as in [15]. Fig. 7 and Fig. 8 present the MSE performance of $w_{1,2}$ and $h_{1,2}$ in the Rayleigh fading channels, respectively.

Here, the pilot length is $N = 32$ and we use the correlator length P of 4 and 8. We can see that the proposed SAGE-ECM can reach the CRB when SNR is large enough; however, the frequency offset estimation performance of methods in [15] and [14] and channel estimation performance of method in [14] are worse.

C. Example 3: Fading Channel and Random Offset

In this subsection, we assume that the frequency offset from any transmit antenna to the/any receive antenna in any evaluation is a realization of a uniform random variable over the interval of $2\pi(-0.1\ 0.1]$. Moreover, the system is operating under the Rayleigh fading environment.

1) 2×1 System, MSE Performance: We come back to the case of $N_T = 2$ transmit antennas and $N_R = 1$ receive antenna. We consider the pilot length of $N = 32$ and correlator length P of 2 and 4. The MSE of $w_{1,2}$ and $h_{1,2}$ are illustrated in Fig. 9 and Fig. 10, respectively. Once again, the proposed SAGE-ECM reaches the CRB while the other methods are worse.

2) 2×2 System, BER Performance: To see the impact of the estimated parameters on the BER performance of the system, we present in Fig. 11 the BER using parameters derived from different algorithms. We consider the system having $N_T = 2$ transmit antennas and $N_R = 2$ receive antennas. Each frame from each transmit antenna consists of two portions: the pilot portion and data portion. The pilot portion is with length of $N = 32$, which constitutes 10% of total frame length. The BPSK constellation is used and the coherent detection method is used at the receiver to detect the signal. We observe that at the level of $\mathrm{BER} = 10^{-3}$, the receiver using SAGE-ECM can obtain a 3 dB gain as compared with [14], for the case of $P = 4$. When $P = 2$, this gain increases to around 5 dB.

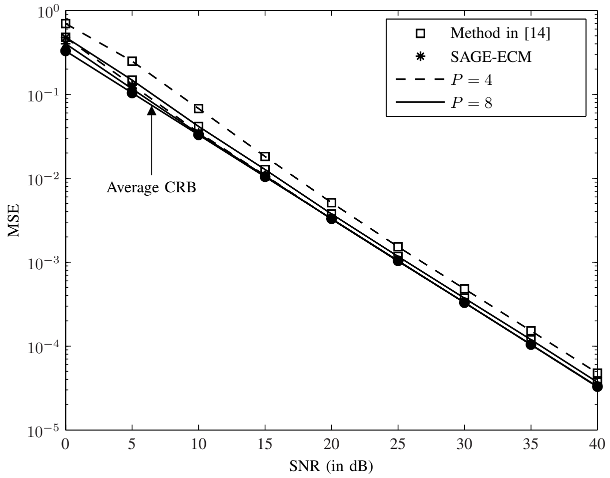

Fig. 8. Comparison of MSE performances of $h_{1,2}$ of [14] and SAGE-ECM algorithms for 4 transmit antennas system.

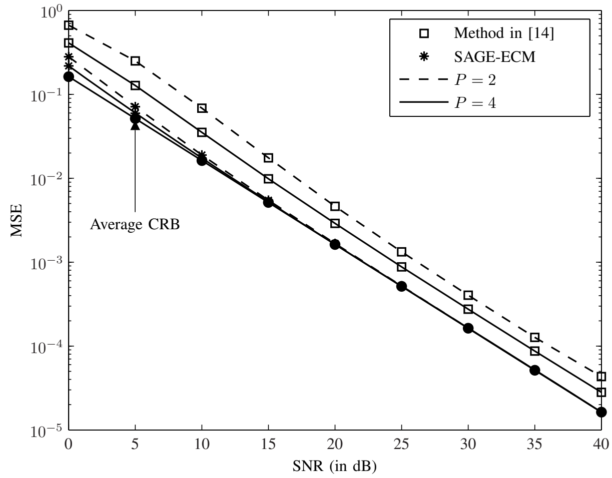

Fig. 10. Comparison of MSE performances of $h_{1,2}$ of [14] and SAGE-ECM algorithms for 2 transmit antennas system.

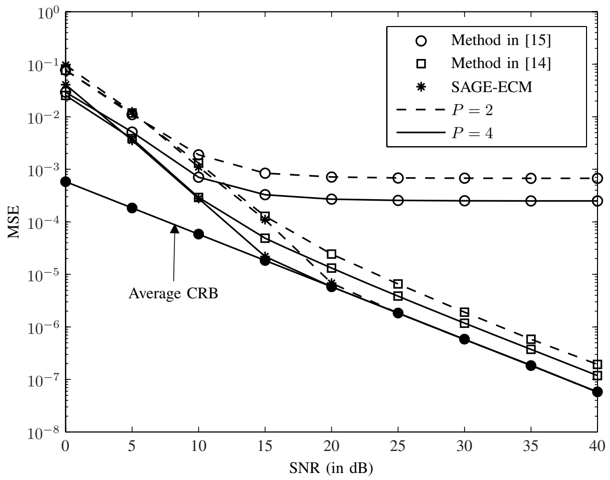

Fig. 9. Comparison of MSE performances of $w_{1,2}$ of [15], [14], and SAGE-ECM algorithms for 2 transmit antennas system.

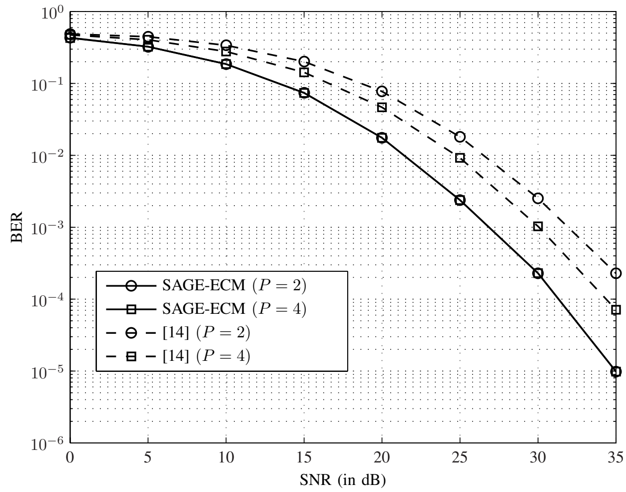

Fig. 11. Comparison of BER performances of [14] and SAGE-ECM algorithms for 2×2 system.

VI. Conclusions

In this paper, we investigated the problem of estimating the channel coefficients and frequency offsets based on training sequences for a MIMO system operating under a flat-fading environment. We proposed two iterative algorithms to the above estimation problem. The first algorithm is based on ECM algorithm. The performance of this algorithm achieves the CRB for both channel and frequency offsets estimations. It overcomes the error floor in MSE performance of frequency offsets estimation in [15] and also outperforms [14]. This algorithm requires no special design on training sequences. However, the complexity in terms of required number of iterations is still high. Hence, we proposed the second algorithm based on SAGE and ECM algorithms.

Finally, we point out that the proposed algorithms can also be extended to frequency selective fading channels. In this scenario, the transmit-receive channel model (6) needs to be modified to take care of the multiple delayed paths for each transmit-receive antenna pair, and the proposed algorithms need to estimate a larger number of unknown parameters in each step.

Acknowledgement

The authors would like to thank the anonymous reviewers for their critical comments that greatly improved this paper.

References

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Labs Intern. Tech. Memo*, July 1995.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[3] Z. Liu, G. Giannakis, and B. Hughes, "Double-differential space-time block coding for time-selective fading channels," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1529–1539, Sep. 2001.

[4] D. Hong, Y. Lee, D. Hong, and C. Kang, "Robust frequency offset estimation for pilot symbol assisted packet CDMA with MIMO antenna systems," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 262–264, June 2002.

[5] F. Simoens and M. Moeneclaey, "Reduced complexity data-aided and code-aided frequency offset estimation for flat-fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1558–1567, June 2006.

[6] ——, "Computationally efficient frequency offset estimation for flat-fading MIMO channels: Performance analysis and training sequence design," in *Proc. IEEE Global Telecommun. Conf.*, 2004, vol. 4, pp. 2460–2464.

[7] T. Cui and C. Tellambura, "Joint channel and frequency offset estimation and training sequence design for MIMO systems over frequency selective channels," in *Proc. IEEE Global Telecommun. Conf.*, 2004, vol. 4, pp. 2344–2348.
[8] S. Shahbazpanahi, A. B. Gershman, and G. B. Giannakis, "Joint blind channel and carrier frequency offset estimation in orthogonally space-time block coded MIMO systems," in *Proc. IEEE Workshop Signal Processing Advances Wireless Commun.*, 2005, pp. 363–367.
[9] M. R. Bhatnagar, R. Vishwanath, and M. K. Arti, "On blind estimation of frequency offsets in time varying MIMO channels," in *Proc. IEEE IFIP International Conf. Wireless Optical Commun. Networks*, vol. 4, 2006, pp. IV–73–IV–76.
[10] Wonil Rho and A. Paulraj, "MIMO channel capacity for the distributed antenna systems," in *Proc. IEEE Veh. Technol. Conf.*, 2006, vol. 2, pp. 706–709.
[11] Z. Ni and D. Li, "Impact of fading correlation and power allocation on capacity of distributed MIMO," in *Proc. IEEE CAS Emerging Technol.*, 2004, vol. 2, pp. 697–700.
[12] Z. Li, D. Qu, and G. Zhu, "An equalization technique for distributed STBC-OFDM system with multiple carrier frequency offsets," in *Proc. IEEE Wireless Commun. Networking Conf.*, 2006, vol. 2, pp. 839–843.
[13] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 602–613, Mar. 2003.
[14] Z. Lu, J. Li, L. Zhao, and J. Pang, "Iterative parameter estimation in MIMO flat-fading channels with frequency offsets," in *Proc. IEEE International Conf. Advanced Information Networking Applications*, vol. 2, 2006.
[15] Y. Yao and T. Ng, "Correlation-based frequency offset estimation in MIMO system," in *Proc. IEEE Veh. Technol. Conf.*, 2003, vol. 1, pp. 438–442.
[16] P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Royal Stat. Soc. Series B*, vol. 39, no. 1, pp. 447–450, Oct. 1993.
[17] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 13, pp. 47–60, Nov. 1996.
[18] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extension*. New York: Wiley, 1997.
[19] T. A. Fesler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2664–2677, Oct. 1994.
[20] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, June 1993.
[21] M. Feder and E. Weistein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

The-Hanh Pham received the B.Eng. degree in Electronics and Telecommunications from Hanoi University of Technology, Hanoi, Vietnam, in 2001, and is currently working toward the Ph.D. degree at the Department of Electrical Engineering, National University of Singapore. His research interests are in signal processing for communications, MIMO systems.

Arumugam Nallanathan (S'97-M'00-SM'05) received the B.Sc. with honors from the University of Peradeniya, Sri-Lanka, in 1991, the CPGS from Cambridge University, United Kingdom, in 1994 and the Ph.D. from the University of Hong Kong, Hong Kong, in 2000, all in Electrical Engineering. Since then, he has been an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. His research interests include OFDM systems, ultra-wide bandwidth (UWB) communication and localization, MIMO systems, and cooperative diversity techniques. In these areas, he has published over 100 journal and conference papers. He is a co-recipient of the Best Paper Award presented at 2007 IEEE International Conference on Ultra-Wideband.

He currently serves on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *John-Wiley's Wireless Communications and Mobile Computing* and the *EURASIP Journal of Wireless Communications and Networking* as an Associate Editor. He served as a Guest Editor for the *EURASIP Journal of Wireless Communications and Networking: Special Issue on UWB Communication Systems-Technology and Applications*. He also served as a technical program committee member for more than 25 IEEE international conferences. He currently serves as the General Track Chair for IEEE VTC'2008-Spring, Co-Chair for the IEEE GLOBECOM'2008 Signal Processing for Communications Symposium, and the IEEE ICC'2009 Wireless Communications Symposium.

Ying-Chang Liang (SM'00) received the PhD degree in Electrical Engineering in 1993. He is now a Senior Scientist at the Institute for Infocomm Research (I2R), Singapore. He also holds adjunct associate professorship positions at Nanyang Technological University (NTU) and the National University of Singapore (NUS), both in Singapore, and an adjunct professorship position for the University of Electronic Science and Technology of China (UESTC). From Dec. 2002 to Dec. 2003, Dr Liang was a visiting scholar with the Department of Electrical Engineering, Stanford University. He has been teaching graduate courses at NUS since 2004. In I2R, he has been leading the research activities in the area of cognitive radio, cooperative communications, and standardization activities for IEEE 802.22 wireless regional networks (WRAN), for which his team has made fundamental contributions in physical layer, MAC layer, and spectrum sensing. His research interest includes cognitive radio, dynamic spectrum access and reconfigurable signal processing for broadband communications, space-time wireless communications, information theory, and wireless networks and statistical signal processing, in which he has published over 150 international journal and conference papers.

Dr Liang received the Best Paper Awards from IEEE VTC-Fall'1999 and IEEE PIMRC'2005. He served as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2002 to 2005, and is serving as Lead Guest-Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, SPECIAL ISSUE ON COGNITIVE RADIO: THEORY AND APPLICATIONS. Dr Liang has served for various IEEE conferences as a technical program committee (TPC) member. He was Publication Chair for the 2001 IEEE Workshop on Statistical Signal Processing, TPC Co-Chair for IEEE ICCS'2006, Panel Co-Chair for IEEE VTC'2008-Spring, and Co-Chair, Thematic Program on Random Matrix Theory and its Applications in Statistics and Wireless Communications, the Institute for Mathematical Sciences, National University of Singapore, 2006. Dr Liang is a Senior Member of IEEE.