Line Search Based Iterative Joint Estimation of Channels and Frequency Offsets for Uplink OFDMA Systems

Yanxin Na, *Member, IEEE,* and Hlaing Minn, *Senior Member, IEEE*

***Abstract*— This paper proposes a novel iterative joint estimation of channels and frequency offsets for uplink OFDMA systems. The proposed method is based on the iterative line search algorithm with a trust region. The objective function minimizes the mean-square distance between the received vector and the reconstructed received vector based on the estimated parameters. In each iteration, the parameters are updated based on the minimum mean square error criterion. We also present a stop criterion for the iterative algorithm. The simulation results show the fast convergence rate of the proposed algorithm with asymptotic performance approaching the average CRBs (averaged over the random frequency offsets and fading channels). Our proposed method achieves a much lower computational complexity and better bit error rate performance than the existing SAGE method.**

***Index Terms*— CRB, joint estimation, line search, OFDMA.**

I. Introduction

ORTHOGONAL Frequency Division Multiple Access (OFDMA) has been proposed for several wireless communication systems (e.g., IEEE 802.16a, 802.16e, 802.20) and has recently attracted more attention [1]–[7] as a promising candidate for the fourth generation (4G) systems. In OFDMA, multiple users simultaneously transmit orthogonal frequency division multiplexing (OFDM) symbols through groups of assigned subcarriers. OFDMA systems have several advantages such as adaptability according to the users' channel conditions (multiuser diversity, dynamic subcarrier assignment, adaptive bit loading/modulation/power allocation), scalability in data rate and spectral occupancy, modularity, and low complexity equalization in dispersive channels. However, the asynchronous time and frequency offsets among different uplink users perturb the orthogonality among the sub-carriers and yield significant interference in channel estimation which in turn degrades the system error performance significantly. Through downlink synchronization or initial ranging process [6], timing offsets can be limited within the interference-free interval of the the cyclic prefix. But reliable estimation of the multiuser residual frequency offsets and the channel coefficients remains a challenging task and is critical for the data detection in OFDMA uplink.

Several methods for estimating carrier frequency offsets and/or channel coefficients in OFDMA systems have been recently presented in [2]–[5], [8], [9]. For OFDMA systems where users are assigned to sub-bands of adjacent subcarriers, [2] addressed the mobile's uplink time and frequency offset estimation by using the cyclic prefix redundancy. In [3], based on multiple OFDM symbols, a time and frequency offset estimation of a new OFDMA uplink user is presented for the situation where all other users have already been synchronized perfectly. A signal structure-based deterministic estimation algorithm is proposed in [4] and [5] for simultaneously estimating the carrier frequency offsets of all OFDMA uplink users with a particular interleaved subcarrier assignment. In [8] and [9], space-alternating generalized expectation-maximization (SAGE) algorithm and alternating-projection frequency estimator (APFE), both of which require grid search, are implemented as a joint frequency offsets and channels estimation in OFDMA uplink.

The above existing methods demand one or more of the following requirements: specific subcarrier assignment, multiple training symbols, ideal system situation, and very large complexity. In this paper, for joint estimation of multiuser frequency offsets and channels in the OFDMA uplink with arbitrary subcarrier assignment using a training symbol, we propose a simpler iterative algorithm than the existing counterparts. Our proposed method is based on the line search with a trust region. In each iteration, updating the parameters is based on the minimum mean square error criterion between the received vector and the trial received vector reconstructed from the trial estimated parameters. We also present a stop criterion for our line search (LS) joint estimation algorithm. Our algorithm has a much simpler computation complexity than the existing SAGE or APFE algorithms. Due to the downlink frequency offset estimation and correction at each user terminal, the frequency offset in the uplink will be small and hence our algorithm considers the frequency offset estimation range of [-0.5, 0.5] times the subcarrier spacing. We show the Cramer-Rao bounds for the joint estimation of multiuser frequency offsets and channels in OFDMA uplink. The simulation results show that our proposed algorithm converges faster than the existing algorithms, and its estimation performance is close to that of APFE algorithm and is better than that of SAGE algorithm. Our proposed algorithm gives

Manuscript received April 6, 2006; revised July 8, 2007 and August 7, 2007; accepted August 14, 2007. The associate editor coordinating the review of this paper and approving it for publication was H. Xu. This work was supported by Project Emmitt, the University of Texas at Dallas, Texas, USA.

Y. Na is with Navini Networks, 2240 Campbell Creek Boulevard, Richardson, Texas 75082 USA (e-mail: yna@navini.com).

H. Minn is with the University of Texas at Dallas, 2601 N. Floyd Road, Richardson, TX USA (e-mail: hlaing.minn@utdallas.edu).

Digital Object Identifier 10.1109/TWC.2007.060147.

about 2 dB SNR advantage over the existing SAGE method at the uncoded bit error rate (BER) of 0.01.

The rest of this paper is organized as follows. Section II presents the signal model for an OFDMA uplink. The proposed method for the joint estimation of channels and frequency offsets in OFDMA uplink is described in Section III. In Section IV, we discuss the convergence of our proposed algorithm and derive the Cramer-Rao bound of the considered joint parameter estimation. Simulation results are presented in Section V and the paper is concluded in Section VI.

Notations: An upper (lower) case boldfont letter represents a matrix (column vector). The superscripts T and H denote the transpose and the Hermitian transpose, respectively. $[\mathbf{X}]_{i,j}$ denotes the $(i,j)^{\underline{\text{th}}}$ element of matrix $\mathbf{X}$. The term $y(n)$ denotes the $n^{\underline{\text{th}}}$ element of vector $\mathbf{y}$. $Tr(\ )$, $E(\ )$, $\Re(\ )$ and $\Im(\ )$ stand for the matrix trace, the expectation, the real part and the imaginary part, respectively. $|\mathbf{y}|$ denotes the Euclidean norm of $\mathbf{y}$. diag$(x_1, x_2, \ldots, x_N)$ represents a diagonal matrix whose $l^{\underline{\text{th}}}$ diagonal element is x_l while diag($\mathbf{X}$) denotes a vector of the diagonal elements of $\mathbf{X}$. $\mathbf{I}_N$ is the identity matrix of size $N \times N$.

II. Signal Model

We consider an uplink OFDMA system with N subcarriers and K active users. Let $\mathbf{d}_{k,n} = [d_{k,n}(0),\ d_{k,n}(1),\ \ldots\ ,\ d_{k,n}(N-1)]^T$ denote the $n^{\underline{\text{th}}}$ block of frequency domain symbols transmitted by the $k^{\underline{\text{th}}}$ user, where $E[\mathbf{d}_{k,n}^H \mathbf{d}_{k,n}] = N_k$ and $d_{k,n}(m)$ is nonzero only if the $m^{\underline{\text{th}}}$ subcarrier is one of the N_k subcarriers assigned to the user k. The corresponding low-pass equivalent discrete-time transmitted signal (sampled at N times the subcarrier spacing) is given by

$$\begin{aligned} \mathbf{s}_{k,n} &= [s_{k,n}(0), s_{k,n}(1), \ldots, s_{k,n}(N-1)]^T \\ &= \sqrt{P_k}\mathbf{F}^H \mathbf{d}_{k,n}\,, \end{aligned} \tag{1}$$

where $\mathbf{F}$ is the $N \times N$ discrete Fourier transform matrix with $[\mathbf{F}]_{i,m} = \frac{1}{\sqrt{N}}\exp(\frac{-j2\pi i m}{N})$, for $i = 0, \ldots, N-1$, $m = 0, \ldots, N-1$, $j = \sqrt{-1}$; and P_k is the transmitted subcarrier power of the user k. For an inter-block interference free reception, N_G cyclic prefix samples ($s_{k,n}(i) = s_{k,n}(i+N)$, $-N_G \le i \le -1$) are inserted. Then the resulting signal is transmitted over a frequency-selective fading channel. For simplicity, we omit the OFDM symbol block index n. Let $\mathbf{h}_k = [h_k(0), h_k(1), \ldots, h_k(L-1)]^T$ denote the discrete-time equivalent channel impulse response (CIR) of the $k^{\underline{\text{th}}}$ user which includes the effects of the transmit and receive filters, the channel medium, the carrier phase offset and the timing error[1]. We use ϵ_k to denote the $k^{\underline{\text{th}}}$ user's carrier frequency offset normalized by the subcarrier spacing. Then the low-pass equivalent uplink received vector $\mathbf{y}$ after the cyclic prefix removal is given by

$$\mathbf{y} = \sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{C}_{\epsilon_\mathbf{k}} \mathbf{S}_\mathbf{k} \mathbf{h}_\mathbf{k} + \mathbf{v}\,, \tag{2}$$

where $\mathbf{C}_{\epsilon_k} = \text{diag}(1, e^{\frac{j2\pi\epsilon_k}{N}}, \ldots, e^{\frac{j2\pi(N-1)\epsilon_k}{N}})$, $[\mathbf{S}_k]_{l,m} = s_k(l-m)$ for $l = 0, 1, \ldots, N-1$, $m = 0, 1, \ldots, L-1$, and $\mathbf{v}$ is a vector of circularly-symmetric complex Gaussian noise samples with mean zero and covariance matrix $\sigma^2 \mathbf{I}_N$. The signal to noise power ratio (SNR) of the $k^{\underline{\text{th}}}$ user is defined as $\text{SNR}_k = \dfrac{P_k}{\sigma^2}$.

[1]Due to the absorption of the timing error into the CIR, some channel coefficients at the beginning and the end of $\mathbf{h}_k$ can be just zeros.

III. Proposed Joint Estimation of Channels and Frequency Offsets

The optimal maximum likelihood (ML) joint estimation of channels and frequency offsets in the OFDMA uplink is computationally prohibitive for practical implementation. To alleviate the complexity barrier of the ML method, several variants [8], [9] of the expectation-maximization (EM) approach have been proposed but their computation complexities are still very high especially when the number of users is large. In this work, we propose a simpler recursive joint estimation algorithm without sacrificing the estimation performance. The basic strategy relies on the iterative line search as will be described below.

A. Joint Estimation Algorithm

The $(K+LK)$ variables to be estimated are defined as the following vector

$$\mathbf{x} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_K, \mathbf{h}_1^T, \mathbf{h}_2^T, \ldots, \mathbf{h}_K^T]^T\,, \tag{3}$$

and the corresponding estimates are denoted as

$$\hat{\mathbf{x}} = [\hat{\epsilon}_1, \hat{\epsilon}_2, \ldots, \hat{\epsilon}_K, \hat{\mathbf{h}}_1^T, \hat{\mathbf{h}}_2^T, \ldots, \hat{\mathbf{h}}_K^T]^T. \tag{4}$$

Let the superscript (i) denote the iteration index. The estimated vector $\hat{\mathbf{x}}$ at the $i^{\underline{\text{th}}}$ iteration step is $\hat{\mathbf{x}}^{(i)} = [\hat{\epsilon}_1^{(i)}, \hat{\epsilon}_2^{(i)}, \ldots, \hat{\epsilon}_K^{(i)}, (\hat{\mathbf{h}}_1^{(i)})^T, (\hat{\mathbf{h}}_2^{(i)})^T, \ldots, (\hat{\mathbf{h}}_K^{(i)})^T]^T$. We can reconstruct the received signal based on the parameter vector $\tilde{\mathbf{x}}$ as

$$\hat{\mathbf{y}} \triangleq \hat{\mathbf{y}}(\tilde{\mathbf{x}}) = \sum_{k=1}^{K} \mathbf{C}_{\tilde{\epsilon}_k} \mathbf{S}_k \tilde{\mathbf{h}}_k \tag{5}$$

and at the $i^{\underline{\text{th}}}$ iteration, the corresponding reconstructed received signal $\hat{\mathbf{y}}^{(i)}$ based on the estimated vector $\hat{\mathbf{x}}^{(i)}$ is

$$\hat{\mathbf{y}}^{(i)} \triangleq \hat{\mathbf{y}}^{(i)}(\hat{\mathbf{x}}^{(i)}). \tag{6}$$

From (2) and (6), we obtain the error signal $\mathbf{z}^{(i)}$ at the $i^{\underline{\text{th}}}$ iteration as

$$\mathbf{z}^{(i)} = \mathbf{y} - \hat{\mathbf{y}}^{(i)}\,. \tag{7}$$

In the $i^{\underline{\text{th}}}$ iteration, $\hat{\mathbf{y}}^{(i)}$ is updated based on minimizing the objection function $\{E|\mathbf{z}^{(i)}|^2\}$. The strategy of our proposed joint estimation algorithm is as follows. We start with an initial point $\hat{\mathbf{y}}^{(0)}$, follow the gradient direction, iteratively obtain the point $\hat{\mathbf{y}}^{(i)}$ close to the received signal $\mathbf{y}$ in the minimum mean square error (MMSE) sense. It implies that the objective function of our proposed algorithm is to minimize $\{E|\mathbf{y} - \hat{\mathbf{y}}|^\mathbf{2}\}$.

The signal model $\mathbf{y}$ (or $\hat{\mathbf{y}}$) is characterized by a nonlinear function of $\mathbf{x}$ (or $\hat{\mathbf{x}}$), and hence the gradient and the curvature of the function $\mathbf{y}$ (or $\hat{\mathbf{y}}$) have a significant impact on identifying the convergent point. Using the gradient or the curvature (Hessian, the second derivative of the objective function) is a trade-off between the simplicity and the convergent rate. For implementation simplicity, we only use the gradient of the function in this paper and analyze the performance. In other words, we approximate that the function $\mathbf{y}$ is linear in a small region which enables us to use the line search in our proposed joint estimation algorithm. By applying a first-order Taylor series expansion, the function $\mathbf{y}$ at the $i^{\underline{\text{th}}}$ iteration step can be approximated as

$$\mathbf{y} \approx \hat{\mathbf{y}}^{(i)} + \mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i)}) + \mathbf{v}\,, \tag{8}$$

where the gradient matrix $\mathbf{G}^{(i)}$ is given by

$$\begin{aligned}\mathbf{G}^{(i)} &\triangleq \left[\frac{\partial \hat{\mathbf{y}}}{\partial \tilde{\epsilon_1}}, \ldots, \frac{\partial \hat{\mathbf{y}}}{\partial \tilde{\epsilon_K}}, \frac{\partial \hat{\mathbf{y}}}{\partial (\tilde{\mathbf{h_1}})^T}, \ldots, \frac{\partial \hat{\mathbf{y}}}{\partial (\tilde{\mathbf{h_K}})^T}\right]_{\tilde{\mathbf{x}}=\hat{\mathbf{x}}^{(i)}} \\ &= [\mathbf{\Lambda C}_{\hat{\epsilon}_1^{(i)}}\mathbf{S}_1\hat{\mathbf{h}}_1^{(i)}, \ldots, \mathbf{\Lambda C}_{\hat{\epsilon}_K^{(i)}}\mathbf{S}_K\hat{\mathbf{h}}_K^{(i)}, \mathbf{C}_{\hat{\epsilon}_1^{(i)}}\mathbf{S}_1, \ldots, \mathbf{C}_{\hat{\epsilon}_K^{(i)}}\mathbf{S}_K]\,,\end{aligned} \tag{9}$$

where the diagonal matrix $\mathbf{\Lambda} = \text{diag}(0, \frac{j2\pi}{N}, \ldots, \frac{j2\pi(N-1)}{N})$. In the following we use the Jacobian matrix $\mathbf{G}^{(i)}$ to update the estimated variables.

By using (2), (6) and (8), the objective function at the $(i+1)^{\underline{\text{th}}}$ iteration can be written as

$$\begin{aligned} &\min\{E(|\mathbf{y} - \mathbf{y}^{(i+1)}|^2)\} \\ \approx\ &\min\{E(|\mathbf{y}^{(i)} + \mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i)}) + \mathbf{v} - \mathbf{y}^{(i)} \\ &-\mathbf{G}^{(i)}(\hat{\mathbf{x}}^{(i+1)} - \hat{\mathbf{x}}^{(i)})|^2)\} \\ =\ &\min\{E(|\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)}) + \mathbf{v}|^2)\}\,. \end{aligned} \tag{10}$$

We use the line search strategy in minimizing the new objective function in (10). We choose a direction and search along this direction from the current point $\hat{\mathbf{x}}^{(i)}$ to a new point $\hat{\mathbf{x}}^{(i+1)}$ (with a new objective function value) as

$$\hat{\mathbf{x}}^{(i+1)} = \hat{\mathbf{x}}^{(i)} + \mathcal{A}^{(i)}\mathbf{z}^{(i)}\,, \tag{11}$$

where $\mathbf{z}^{(i)}$ is defined in (7) and the optimal updated gain $\mathcal{A}^{(i)}$ based on the MMSE criterion is given by

$$\begin{aligned}\mathcal{A}^{(i)} \;=\; &\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{G}^{(i)})^H \\ &[\mathbf{G}^{(i)}\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{G}^{(i)})^H + \sigma^2\mathbf{I}_N]^{-1}\,,\end{aligned} \tag{12}$$

or

$$\mathcal{A}^{(i)} = [(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)} + \sigma^2(\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H)^{-1}]^{-1}(\mathbf{G}^{(i)})^H\,, \tag{13}$$

where $\mathbf{b}^{(i)} = (\mathbf{x} - \hat{\mathbf{x}}^{(i)})$. The derivation of $\mathcal{A}^{(i)}$ is given in the Appendix-A.

Note that the residual error vector $\mathbf{b}^{(i)}$ is unknown and hence (12) or (13) is not directly applicable. The line search algorithm generates a limited number of trial step lengths until it converges to the vicinity of the desired value. An exact minimization is not necessary in the $i^{\underline{\text{th}}}$ iteration and a small step length is preferable for the convergence. Based on these facts, we introduce a scaling parameter $\beta^{(i)}$ $(0 \le \beta^{(i)} \le 1)$ and approximate $\mathcal{A}^{(i)}$ by considering $E(\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H)$ in replace of $(\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H)$ and by assuming [2] $E(\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H) = \gamma\mathbf{I}$ where γ is a scalar. Then our updated gain becomes

$$\mathcal{A}^{(i)} = \beta^{(i)}(\mathbf{G}^{(i)})^H[\mathbf{G}^{(i)}(\mathbf{G}^{(i)})^H + (\sigma^2/\gamma)\mathbf{I}_N]^{-1} \tag{14}$$

[2] $\mathbf{b}$ is the residual error vector of the estimation of the parameter vector $\mathbf{x}$ (frequency offsets and channel coefficients). We do not use channel statistics in the estimation. In practical implementation, the exact estimation error covariance matrix is intractable. Here for sake of practical implementation, we assume the frequency offsets and channel coefficients to different users or different taps are independent. For the joint estimation, it is reasonable to assume that the estimation errors of elements of $\mathbf{x}$ are uncorrelated. Hence, a reasonable and practical choice for $E[b(i)b(i)^H]$ is γI .

or

$$\mathcal{A}^{(i)} = \beta^{(i)}[(\mathbf{G}^{(i)})^H(\mathbf{G})^{(i)} + (\sigma^2/\gamma)\mathbf{I}]^{-1}(\mathbf{G}^{(i)})^H \tag{15}$$

where σ^2/γ is replaced by a fixed design value in the implementation. The direction and the radius of the trust-region at the $i^{\underline{\text{th}}}$ iteration are respectively given by

$$(\mathbf{G}^{(i)})^H[\mathbf{G}^{(i)}(\mathbf{G}^{(i)})^H + (\sigma^2/\gamma)\mathbf{I}_N]^{-1/2} \tag{16}$$

$$\text{and} \qquad \beta^{(i)}[\mathbf{G}^{(i)}(\mathbf{G}^{(i)})^H + (\sigma^2/\gamma)\mathbf{I}_N]^{-1/2}\mathbf{z}^{(i)}\,. \tag{17}$$

In our proposed algorithm, how to set up the value of $\beta^{(i)}$ is critical to the convergence of our algorithm. Since the sufficient condition for the convergence is $|\mathbf{y} - \mathbf{y}^{(i+1)}| \le |\mathbf{y} - \mathbf{y}^{(i)}|$ [see Appendix-b], we will set up the value of $\beta^{(i)}$ to satisfy the sufficient condition. First we set the initial value $\beta^{(0)}$ to 1. In the iteration $i+1$, we first set $\beta_1^{(i+1)} = \beta^{(i)}$ and $\gamma = 1$ to get $\mathbf{y}^{(i+1)}$. If $|\mathbf{y} - \mathbf{y}^{(i+1)}| > |\mathbf{y} - \mathbf{y}^{(i)}|$, the value of $\beta_n^{(i+1)}$ is repeatedly reduced to $\beta_{n+1}^{(i+1)} = \rho\beta_n^{(i+1)}$ until the condition $|\mathbf{y} - \mathbf{y}^{(i+1)}| \le |\mathbf{y} - \mathbf{y}^{(i)}|$ is satisfied where n is the repeated number and $0 < \rho < 1$. Based on our simulation results, we suggest the following setting: the value of ρ should be in the range of $0.5 < \rho < 1$, the value of γ is 1, and if $\beta_n^{(i+1)} < 10^{-10}$, set $\gamma = \gamma/10$ and $\beta_1^{(i+1)} = 1$, then repeat the above process on $\beta^{(i+1)}$ until $|\mathbf{y} - \mathbf{y}^{(i+1)}| \le |\mathbf{y} - \mathbf{y}^{(i)}|$ is satisfied. In practical implementation, we could also set β value as $\beta_{n+1}^{(i+1)} = -\rho\beta_n^{(i+1)}$ without updating γ and the matrix inversion.

B. Initialization of the Estimation Algorithm

It is well known that a good initial value will not only influence the convergence rate of an optimization algorithm but also decide if it has a high probability to converge to the global maximum point. If relatively reliable coarse estimates of the parameters are available from previous transmission (in slow-varying channels), we can use them as our initial values of the parameters. For a fair comparison with SAGE and APFE, in the paper we use the same initial estimates in all methods. The initial frequency offsets trial values are set to zeros, and the estimated channel coefficients for each user are initialized as

$$\hat{\mathbf{h}}_{\mathbf{k}}^{(\mathbf{0})} = \{[\mathbf{C}_{\epsilon_\mathbf{k}}\mathbf{S}_\mathbf{k}]^\mathbf{H}[\mathbf{C}_{\epsilon_\mathbf{k}}\mathbf{S}_\mathbf{k}]\}^{-\mathbf{1}}[\mathbf{C}_{\epsilon_\mathbf{k}}\mathbf{S}_\mathbf{k}]^\mathbf{H}\mathbf{y}\,, \tag{18}$$

This setting reduces the complexity and improves the stability of the proposed algorithm when comparing with simply setting all initial values to zero.

C. Stop Criterion

When our iterative algorithm should stop is a trade-off between the computation complexity and the convergence (or accuracy). Most of the mathematical optimization algorithms stop the iterations when $\mathbf{e}_x^{(i+1)} = |\hat{\mathbf{x}}^{(i+1)} - \hat{\mathbf{x}}^{(i)}| < \delta_1$ where δ_1 is chosen according to the desired convergent accuracy. For our estimation problem, the converged value of $\hat{\mathbf{x}}$ is not always the exact value of $\mathbf{x}$ due to the noise contamination. Using the above stop criterion for our estimation algorithm is associated with a disadvantage that $\mathbf{e}_x^{(i+1)}$ is sensitive to SNR and hence the value of δ_1 needs to be changed according to the SNR to get the same estimation performance.

Note that even if the estimate of $\mathbf{x}$ is perfect, we have $E\{|(\hat{\mathbf{y}}^{(i+1)} - \mathbf{y})|^2\} = N\sigma^2$ or $E\{|(\hat{\mathbf{y}}^{(i+1)} - \mathbf{y})|\} = \sqrt{2}\frac{\Gamma((N+1)/2)}{\Gamma(N/2)}\sigma$. This fact implies that the converged estimated vector would not be exactly the same as the actual parameter vector due to the noise, and high level of convergent accuracy where the fluctuation of $\hat{\mathbf{y}}^{(i+1)}$ becomes much smaller than the noise variance σ^2 is unnecessary. Hence, we propose the following stop criterion: $\mathbf{e}_y^{(i+1)} = |\hat{\mathbf{y}}^{(i+1)} - \mathbf{y}| - |\hat{\mathbf{y}}^{(i)} - \mathbf{y}| < \delta$ where δ can be set to a value between 0.01σ and 0.1σ. Note that we have applied the fact that if $\mathbf{x}$ is convergent, then so is $\hat{\mathbf{y}}$ or $(\hat{\mathbf{y}} - \mathbf{y})$. In our stop criterion, value of δ needs not be changed for different SNR values.

IV. Performance Measure

We use an approximate signal model in (8) and apply a gradient search based line search approach. In each step, the gain $\mathcal{A}^{(i)}$ is updated based on MMSE criterion. In minimizing MSE in (10), we use the approximate signal model, and hence it is an approximate MMSE. So, it may be considered as an approximate LMS algorithm, or a modified Kalman filter. The LMS algorithm is well-known to have a slow convergence rate. But our proposed algorithm shows a fast convergence rate. As the simulation results show, the proposed algorithm converges in about 3 iterations. The equation (14) and (15) resemble the structure of Newton optimization algorithm, which may be a reason for the fast convergence of the proposed algorithm. We would like to mention that to the best of our knowledge, the joint estimation of multiuser CFO and channel estimation has not been tackled by an LMS type approach which deserves to be investigated especially in the context of OFDMA uplink synchronization and channel estimation.

We use the computation complexity and the estimation MSE as our performance measure in comparing several iterative joint estimation methods. The computation complexity of each method depends on the corresponding number of iterations and each iteration's computation complexity. The number of iterations to achieve a certain accuracy is related to the convergence rate of the algorithm. The estimation MSE is related to the convergence of the algorithm. The convergence and the convergent rate of our proposed algorithm will be discussed in the following. To judge the MSE performance of our proposed method, we also derive the Cramer-Rao bound (CRB) of the considered joint parameter estimation.

A. Convergence and Convergence Rate

Our proposed joint estimation algorithm is based on line search strategy ([10]–[13]). Since $|\mathbf{y} - \hat{\mathbf{y}}^{(i)}|$ as the function of $\mathbf{x}$ is continuously differentiable and bounded below with $|\mathbf{y} - \hat{\mathbf{y}}^{(i)}| \geq 0$, in a descent direction at $\mathbf{x}^{(i)}$ for $i > 0$, there exits a range of step lengths satisfying the Wolfe conditions which guarantee the convergence of an algorithm [11], [12].

We use the steepest descent direction, the gradient of the current point, since among all the directions we could move from the current point it is the one along which the objective function decreases most rapidly. In general, any descent direction that makes an angle of strictly less than $\pi/2$ radians with the gradient direction is guaranteed to produce a decrease in the objective function. We can verify it by using Taylor's theorem ([11], Theorem 2.1).

A small step length is preferred for the convergence while a larger step length is desirable for a faster convergence rate. This trade-off problem of choosing the step length is typically addressed by a trust-region method. Trust-region methods define a region around the current iterate point within which they trust the model to be an adequate representation of the objective function, and then choose the step to be the approximate minimizer of the model in this trust region. The method chooses the direction and the step length simultaneously. If a step is not acceptable, they reduce the size of the region and find a new optimal point. The size of the trust region is critical to the effectiveness of each iteration step. If the size of region is too small, the algorithm misses an opportunity to take a substantial step that will move it much closer to the optimal point. If it is too large, the next new state point may be far from the minimum point of the objective function, thus we may have to reduce the size of the region and try again. In our algorithm, the size of the trust region depends on $\beta^{(i)}$. For a faster convergence rate, we start with $\beta^{(0)} = 1$ but for the convergence, we maintain the sufficient condition for convergence $|\mathbf{y} - \hat{\mathbf{y}}^{(i+1)}| \leq |\mathbf{y} - \hat{\mathbf{y}}^{(i)}|$ by reducing $\beta^{(i)}$ if necessary.

Being a line search algorithm, our proposed method inherits the linear convergent rate. Let $\hat{\mathbf{x}}^{(i)}$ be a sequence that converges to $\bar{\mathbf{x}}$, or $\bar{\mathbf{x}}$ is defined as the convergent point of the sequence $\hat{\mathbf{x}}^{(i)}$. The traditional definition of the linear convergent rate r_x is

$$r_x \geq \frac{|\hat{\mathbf{x}}^{(i+1)} - \bar{\mathbf{x}}|}{|\hat{\mathbf{x}}^{(i)} - \bar{\mathbf{x}}|} \tag{19}$$

for all sufficiently large i. Since our iterative algorithm is for a statistical estimation, we do not need accuracy as high as in a numerical optimization and we can stop the algorithm when a certain convergence accuracy (defined by δ of our stop criterion) is achieved. Accordingly, we consider the convergent rate of the steps before our proposed algorithm stops based on our stop criteria. We define the linear convergence rate of our proposed algorithm at the $(i+1)^{\underline{\text{th}}}$ iteration $r_x^{(i+1)}$ as

$$r_x^{(i+1)} = \frac{|\hat{\mathbf{x}}^{(i+1)} - \bar{\mathbf{x}}|}{|\hat{\mathbf{x}}^{(i)} - \bar{\mathbf{x}}|}, \tag{20}$$

for $0 < i < N_{\text{stop}}$, where N_{stop} is the iteration index where the algorithm is stopped. As will be shown in the simulation results section, the average convergent rate r_x of our proposed joint estimation algorithm is fast. Since $\hat{\mathbf{y}}^{(i)}$ is also convergent, we will also present in the simulation section the corresponding average convergent rate of $\hat{\mathbf{y}}^{(i)}$ at the $(i+1)^{\underline{\text{th}}}$ iteration defined by $r_y^{(i+1)} = \frac{|\hat{\mathbf{y}}^{(i+1)} - \bar{\mathbf{y}}|}{|\hat{\mathbf{y}}^{(i)} - \bar{\mathbf{y}}|}$, where $\bar{\mathbf{y}}$ is defined as the convergent point of the sequence $\hat{\mathbf{y}}^{(i)}$.

B. Computation Complexity

Based on the measurement of the numbers of addition, multiplication and inversions operations, we analyze the computation complexity of every iteration step in our proposed LS algorithm, SAGE and APFE algorithms respectively in the following.

Except diagonal matrix or some special matrix, in general, calculation of an $N \times M$ matrix $\mathbf{X}$ multiplying an $M \times L$ matrix $\mathbf{Y}$ needs NML complex number multiplications and $N(M-1)L$ complex number additions. In our proposed LS algorithm, we need to operate $(2KNL+KN)$ complex number multiplications and $KN(L-1)$ complex number additions to calculate the gradient matrix $\mathbf{G}^{(i)}$. Calculating the estimated received signal $\mathbf{y}^{(i+1)}$ in the equation (5) needs $N(K-1)$ complex number additions; the error signal $\mathbf{z}^{(i)}$ in the equation (7) needs N complex number additions. The updated gain $\mathcal{A}^{(i)}$ in the equation (15) needs $2N(K+KL)^2$ complex number multiplications, $(K+KL)[(2N-1)(K+KL)-N+1]$ complex number additions and an inversion of a $(K+KL)\times(K+KL)$ matrix. The new estimated variables vector $\hat{\mathbf{x}}^{(i+1)}$ in the equation (11) needs $(K+KL)N$ complex number multiplications and $N+(K+KL)(N-1)$ complex number additions. Thus summing up the computation complexity of $\mathbf{G}^{(i)}$, $\mathbf{y}^{(i+1)}$, $\mathbf{z}^{(i)}$, $\mathcal{A}^{(i)}$ and $\hat{\mathbf{x}}^{(i+1)}$, we obtain the basic computation complexity of our proposed LS algorithm in one iteration as $2K^2(1+L)^2N+3KLN+2KN$ complex number multiplications, $K^2(1+L)^2(2N-1)+KLN+N$ complex number additions and an inversion of a $(K+KL)\times(K+KL)$ matrix. If we update a new γ value, we will calculate the $\mathbf{y}^{(i+1)}$, $\mathbf{z}^{(i)}$, $\mathcal{A}^{(i)}$ and $\hat{\mathbf{x}}^{(i+1)}$ once. If we update a new β value, we need to calculate $\mathbf{y}^{(i+1)}$, $\mathbf{z}^{(i)}$ and $\hat{\mathbf{x}}^{(i+1)}$ once.

In SAGE or APFE algorithms, if we use one coarse grid searches and Δ fine grid search with Θ grids in each search and we update all users estimated variables in each iteration, we obtain the computation complexity of SAGE as $KLN(2L+2\Theta(\Delta+1)+3)$ complex number multiplications, $KL(2LN+\Theta(\Delta+1)N+N-L-1)$ complex number addition and K inversions of $L\times L$ matrices. We also get that APFE needs $K(\Theta(\Delta+1)LN^2+\Theta(\Delta+1)N^2+4K^2L^2N+3\Theta(\Delta+1)L^2N+3L^2N-6KL^2N+KLN+K^2L^2)$ complex number multiplications, $K(\Theta(\Delta+1)LN^2+\Theta(\Delta+1)N^2+3\Theta(\Delta+1)L^2N+3L^2N+4K^2L^2N-6KL^2N-KLN-3\Theta(\Delta+1)NL+2NL-\Theta(\Delta+1)N+N-K^2L^2+2KL^2-\Theta(\Delta+1)L^2-L^2-2KL)$ complex number additions, K inversion of $(K-1)L\times(K-1)L$ matrices, K inversions of $L\times L$ matrices, and K inversions of $KL\times KL$ matrices. For SAGE and APFE algorithms, we set the initial estimated frequency offsets to zeros and the initial channel coefficients are obtained by regarding all other users's signals as noise [8]. The initialization in SAGE and APFE algorithms needs $KLN(2L+3)$ complex multiplications, $K(2L^2N-L^2+LN-L+N)$ complex additions and K inversions of $L\times L$ matrices. In practical environment, the number of subcarriers N is always much larger than the number of channel paths L or the number of users K, hence even with frequency offset grid search accuracy of 10^{-2}, and together with the parabolic interpolation [14] to 3 adjacent frequency offset points to get a peak value as the estimated frequency offset, we find the computation complexity in APFE is much larger than SAGE algorithm or our proposed LS algorithm in each iteration step, and LS algorithm has less complexity than SAGE algorithm.

C. The Cramer-Rao Bound

The CRB for the covariance matrix of the estimation parameter vector is given by the inverse of the Fisher information matrix and the CRB for the variance of $n^{\underline{\text{th}}}$ variable of the parameter vector is given by the $n^{\underline{\text{th}}}$ diagonal element of the CRB for the covariance matrix [15], [16]. Define

$$\begin{aligned}\mathbf{U} &= [\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_K], &(21)\\ \mathbf{Q} &= [\mathbf{q}_1,\mathbf{q}_2,\cdots,\mathbf{q}_K], &(22)\\ \mathbf{u_k} &= \mathbf{C}_{\epsilon_k}\mathbf{S}_k, &(23)\\ \mathbf{q}_k &= \mathbf{M}\mathbf{C}_{\epsilon_k}\mathbf{S}_k\mathbf{h}_k, &(24)\\ \mathbf{M} &= \mathrm{diag}(0,1,\cdots,N-1). &(25)\end{aligned}$$

Then, by straight-forward calculation, we obtain the CRB for the frequency offset vector as

$$\mathrm{CRB}(\boldsymbol{\epsilon}) = \mathrm{diag}[(\mathbf{Z}_{11}-\mathbf{Z}_{12}\mathbf{Z}_{22}^{-1}\mathbf{Z}_{21})^{-1}], \quad (26)$$

and the CRB for the channel coefficient vector as

$$\mathrm{CRB}([\Re(\mathbf{h}^T),\Im(\mathbf{h}^T)]^T) = \mathrm{diag}[(\mathbf{Z}_{22}-\mathbf{Z}_{21}\mathbf{Z}_{11}^{-1}\mathbf{Z}_{12})^{-1}], \quad (27)$$

where

$$\begin{aligned}\mathbf{Z}_{11} &= \left[(\tfrac{2\pi}{N})^2\Re(\mathbf{Q}^H\mathbf{Q})\right],\\ \mathbf{Z}_{12} &= \left[\tfrac{2\pi}{N}\Im(\mathbf{Q}^H\mathbf{U}) \quad \tfrac{2\pi}{N}\Re(\mathbf{Q}^H\mathbf{U})\right],\\ \mathbf{Z}_{21} &= \begin{bmatrix}-\tfrac{2\pi}{N}\Im(\mathbf{U}^H\mathbf{Q})\\ \tfrac{2\pi}{N}\Re(\mathbf{U}^H\mathbf{Q})\end{bmatrix}, \quad (28)\\ \mathbf{Z}_{22} &= \begin{bmatrix}\Re(\mathbf{U}^H\mathbf{U}) & -\Im(\mathbf{U}^H\mathbf{U})\\ \Im(\mathbf{U}^H\mathbf{U}) & \Re(\mathbf{U}^H\mathbf{U})\end{bmatrix}.\end{aligned}$$

Note that the above CRBs are only for a particular realization of the parameter vector. We obtain the average CRBs by averaging the above CRBs over the random frequency offsets and the random channel gains.

V. Simulation Results and Discussions

We consider an OFDMA system with $N=128$ subcarriers and $N_G=16$ CP samples, operated in the 5 GHz frequency band. For simplicity, BPSK modulation is used and the data rate is 2.5 Mbps. The preamble consists of random BPSK modulated pilots on all subcarriers. During data transmission, we use interleaved carrier assignment where subcarriers $k-1$, $K+k-1$, $2K+k-1$,... , are assigned to user k. The normalized carrier frequency offsets are modeled as independent uniformly-distributed continuous random variables in the range of $[-0.5, 0.5]$ for all users. Timing offsets are modeled as independent discrete uniform random variables with the range $[0, 10]$ samples for all users. We use the IEEE 802.16 SUI-3 frequency-selective fading channel model together with a spectral raised cosine-filter for the combined transmit and recieve filter. The total CIR length after incorporating timing offsets is $L=N_G$.

For comparison, existing joint estimation methods (SAGE and APFE from [8] and [9]) are also evaluated. The same stop criterion is used for all methods. Note that both reference methods represent iterative implementations of maximum likelihood algorithm and hence they give very good estimation performance. In the SAGE and APFE algorithms we set carrier frequency offset grid search accuracy to 10^{-2}, and then applying the parabolic interpolation [14] to 3 adjacent frequency offset points to get a peak value as the estimated frequency offset. In our proposed method, we use (15) for the updated gain, and $\rho=5/6$ in the step length setting.

The convergence behaviors and the convergence rates of all considered iterative methods are presented in Figs. 1-5. The convergence behaviors of the proposed LS method and the reference SAGE method with $K = 2$ users are shown in Fig. 1 in terms of the average normalized error vector length $d_y = \mathbf{E}\{\frac{|\hat{\mathbf{y}}-\mathbf{y}|}{|\mathbf{y}|}\}$ versus the number of iterations. Our proposed LS method converges after a few (2 or 3) iterations for all considered SNR values. The SAGE algorithm needs 8, 13, and 15 iterations for convergence at $\mathrm{SNR} = 5$, 15 and 25 dB, respectively.

The average numbers of iterations for the three algorithms are shown in the Fig. 2. The average number of iterations for our proposed LS algorithm is around 3 to 4 for the number of users up to 4 and it remains almost the same for all considered SNRs. A larger SNR, a larger number of users, or/and a smaller (better) value of stop accuracy increase the average number of iterations for the SAGE method which ranges from 6 to 26. Hence, the complexity of the SAGE method is more sensitive to the number of users than the LS method. The APFE method requires only 2 or 3 iterations on the average at a higher computational complexity per iteration than the proposed LS method. The overall complexity depends on the number of iterations and the complexity in each iteration. The APFE requires much larger complexity in each iteration than the LS method. To illustrate this, we plot the relative MATLAB run-times of all methods with $K = 2$ and $\delta = 0.1\sigma$ in Fig. 3. In the figure, LS stands for the LS algorithm using (15). The computation complexity of the SAGE method is around 3 to 4 times and that of the APFE method is around 10 to 40 times the complexity of the proposed LS method.

In the Fig. 4 and 5, we plot the average convergence rate r_x and r_y as a function of the number of iterations for $K = 2$. We assume that the estimated values after 45 iterations is the convergent values. The convergence rate is faster at higher SNR. For the SAGE method, the convergence rate at each iteration remains almost the same. For the proposed LS method, the convergence rates at iteration 1, 2 and 3 are much faster than other iterations. This fact explains that the proposed LS method converges faster than the SAGE method and requires on the average 3 iterations to reach the convergent point.

Figs. 6 and 7 present the comparison of the proposed LS method's MSE performances with the average CRBs for the frequency offset estimation and the channel estimation, respectively. The proposed LS method's MSE performances are very close to the CRBs for high SNR. A smaller stop accuracy gives MSE performance closer to the CRB.

In Figs. 8 and 9, we compare the MSE performances of the proposed LS method and the reference methods for frequency offset estimation and channel estimation, respectively. In frequency offset estimation, for the case of $\delta = 0.1\sigma$, the LS method achieves 4 to 5 dB improvement over the SAGE method and almost the same MSE performance as the APFE method. In channel estimation, the MSE performances of the LS method and the SAGE method are almost the same.

As an overall performance evaluation, we compare the uncoded BER performance of the system using the proposed LS method and the SAGE method in Fig. 10. A frame is composed of a preamble symbol followed by 5 data symbols.

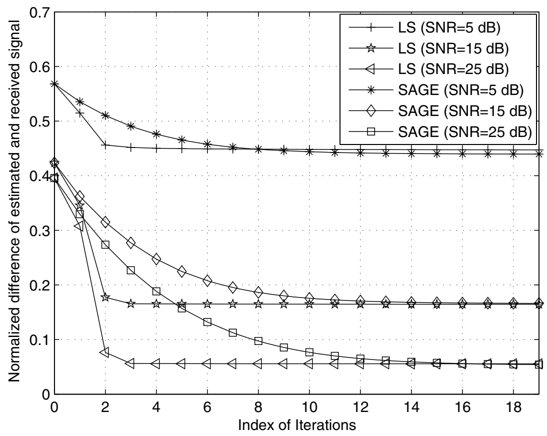

Fig. 1. Average normalized error vector length $d_y = E\{\frac{|\hat{\mathbf{y}}-\mathbf{y}|}{|\mathbf{y}|}\}$ as a function of the number of iterations in an OFDMA system with $K = 2$ users.

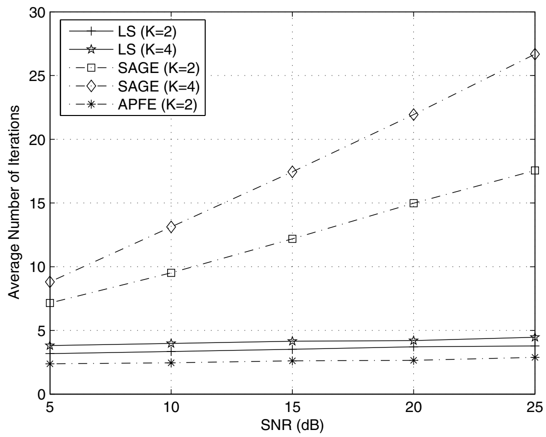

Fig. 2. The average number of iterations as a function of SNR for the stop accuracy $\delta = 0.1\sigma$.

We apply a least square detection. The received signal can also be written as

$$\mathbf{y} = \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{K}} \mathbf{C}_{\epsilon_\mathbf{k}} \mathbf{H}_\mathbf{k} \mathbf{F}^\mathbf{H} \mathbf{D}_\mathbf{k} + \mathbf{v} = \mathbf{\Lambda} \mathbf{D} + \mathbf{v}\,, \tag{29}$$

where D is the unknown data in frequency domain, and H is the circulant matrix of time-domain channel coefficients. We detect the data in the simulation by

$$\hat{D} = sign((\Lambda^H \Lambda)^{-1} \Lambda^H \mathbf{y})\,, \tag{30}$$

The BER obtained with the perfect knowledge of the frequency offsets and the channels (denoted by Per.) is also included as a lower bound. The proposed LS method achieves about 2 dB SNR advantage at the uncoded BER of 0.01 over the SAGE method.

VI. Conclusions

Joint estimation of frequency offsets and channel coefficients in OFDMA uplink is a challenging task requiring very high complexity. By applying a linearized approximate signal

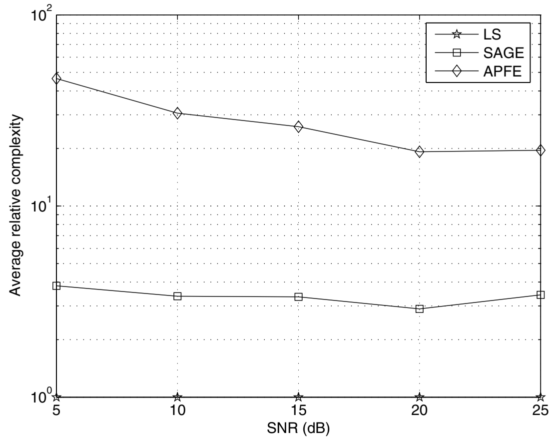

Fig. 3. The average relative complexity as a function of SNR for an OFDMA system with $K = 2$ users and the stop accuracy $\delta = 0.1\sigma$.

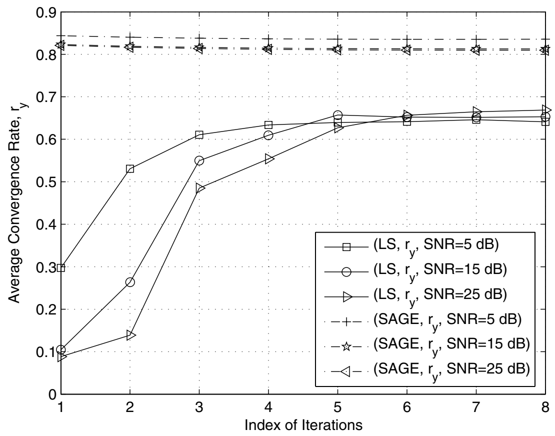

Fig. 5. Average convergence rate r_y as a function of the number of iterations for an OFDMA system with $K = 2$ users.

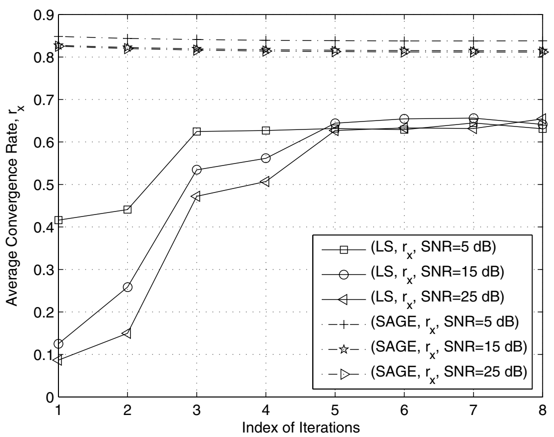

Fig. 4. Average convergence rate r_x as a function of the number of iterations for an OFDMA system with $K = 2$ users.

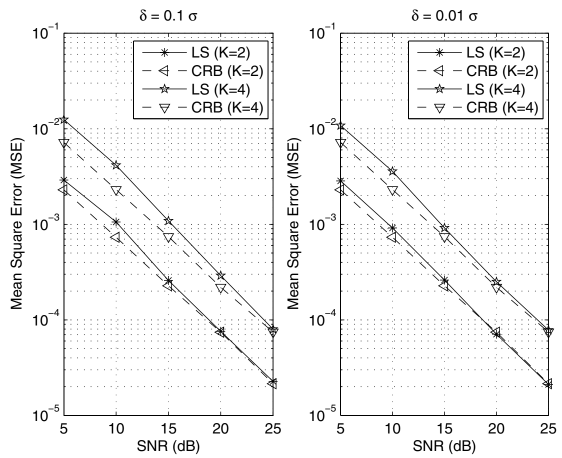

Fig. 6. MSE of the proposed carrier frequency offset estimation as a function of SNR for the stop accuracy $\delta = 0.1\sigma$ or $\delta = 0.01\sigma$.

model, the proposed line search based iterative joint estimation algorithm yields a significant complexity reduction over the existing iterative algorithms. Moreover, the proposed method achieves the same channel estimation performance as and better frequency offset estimation and better BER performance than the existing SAGE method.

APPENDIX

a) Derivation of (12) and (13): We want to get new state of the estimated variable at the $(i+1)^{\underline{\text{th}}}$ iteration based on the MMSE (minimum mean square error) criterion. Thus we want to obtain $\mathcal{A}^{(\mathbf{i})}$ by the MMSE criterion as

$$\begin{aligned}&\mathcal{A}^{(i)} = \arg\min\nolimits_{\mathcal{A}^{(i)}} E(|\mathbf{y} - \hat{\mathbf{y}}^{(i+1)}|^2)\\&= \arg\min\nolimits_{\mathcal{A}^{(i)}} E[(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})^H(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})].\end{aligned} \tag{31}$$

Combining the equations (11), (7) and (31), we obtain

$$\begin{aligned}&E[(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})^H(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})]\\&= E[(\mathbf{x} - \hat{\mathbf{x}}^{(i)} - \mathcal{A}^{(i)}(\mathbf{y} - \hat{\mathbf{y}}^{(i)}))^H(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}\\&(\mathbf{x} - \hat{\mathbf{x}}^{(i)} - \mathcal{A}^{(i)}(\mathbf{y} - \hat{\mathbf{y}}^{(i)}))].\end{aligned} \tag{32}$$

By applying the equation (8) to the equation (32), the equation (32) can be approximated as

$$\begin{aligned}&E[(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})^H(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})]\\&\approx E[(\mathbf{x} - \hat{\mathbf{x}}^{(i)} - \mathcal{A}^{(i)}\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i)}) - \mathcal{A}^{(i)}\mathbf{v})^H(\mathbf{G}^{(i)})^H\\&\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i)} - \mathcal{A}^{(i)}\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i)}) - \mathcal{A}^{(i)}\mathbf{v})].\end{aligned} \tag{33}$$

After defining $\mathbf{b}^{(i)} \overset{\triangle}{=} (\mathbf{x} - \hat{\mathbf{x}}^{(i)})$, the equation (33) can be rewritten as

$$\begin{aligned}&E[(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})^H(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}(\mathbf{x} - \hat{\mathbf{x}}^{(i+1)})]\\&\approx E[((\mathbf{I} - \mathcal{A}^{(i)}\mathbf{G}^{(i)})\mathbf{b}^{(i)} - \mathcal{A}^{(i)}\mathbf{v})^H(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}\\&((\mathbf{I} - \mathcal{A}^{(i)}\mathbf{G}^{(i)})\mathbf{b}^{(i)} - \mathcal{A}^{(i)}\mathbf{v}),\end{aligned} \tag{34}$$

where $\mathbf{I}$ is an identity matrix. Taking the derivative on the value of mean square error with respect to $\mathcal{A}^{(i)}$ and using $\frac{\partial}{\partial \mathbf{A}}Tr(\mathbf{YAX}) = \mathbf{Y}^T\mathbf{X}^T$ [17], we obtain the equation (35) which is on the top of the next page. Equating the resultant equation (35) to zero, we get the optimum updated gain $\mathcal{A}^{(i)}$ as

$$\begin{aligned}\mathcal{A}^{(i)} \;=\; &\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{G}^{(i)})^H(\mathbf{G}^{(i)}\mathbf{b}^{(i)}\\&(\mathbf{b}^{(i)})^H(\mathbf{G}^{(i)})^H + \sigma^2\mathbf{I}_N)^{-1}.\end{aligned} \tag{36}$$

$$\begin{aligned}&\frac{\partial E[((\mathbf{I}-\mathcal{A}^{(i)}\mathbf{G}^{(i)})\mathbf{b}^{(i)}-\mathcal{A}^{(i)}\mathbf{v})^H(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}((\mathbf{I}-\mathcal{A}^{(i)}\mathbf{G}^{(i)})\mathbf{b}^{(i)}-\mathcal{A}^{(i)}\mathbf{v})]}{\partial \mathcal{A}^{(i)}}\\&=\frac{\partial[Tr(\mathbf{G}^{(i)}(\mathbf{I}-\mathcal{A}^{(i)}\mathbf{G}^{(i)})\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{I}-\mathcal{A}^{(i)}\mathbf{G}^{(i)})^H(\mathbf{G}^{(i)})^H+\sigma^2\mathbf{G}^{(i)}\mathcal{A}^{(i)}(\mathcal{A}^{(i)})^H(\mathbf{G}^{(i)})^H)]}{\partial \mathcal{A}^{(i)}}\\&=-(\mathbf{G}^{(i)})^T((\mathbf{G}^{(i)})^H)^T(\mathbf{G}^{(i)}\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{I}-\mathcal{A}^{(i)}\mathbf{G}^{(i)})^H-\sigma^2(\mathcal{A}^{(i)})^H)^T.\end{aligned} \tag{35}$$

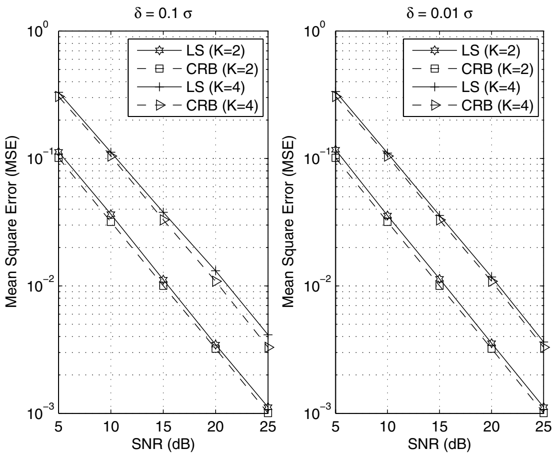

Fig. 7. MSE of the proposed channel estimation as a function of SNR for the stop accuracy $\delta = 0.1\sigma$ or $\delta = 0.01\sigma$.

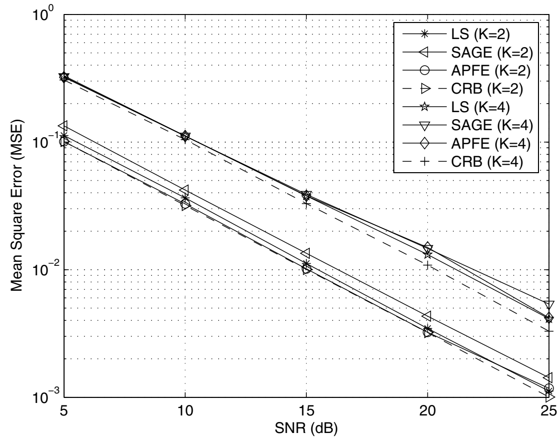

Fig. 9. Channel estimation MSE comparison among several iterative methods with the stop accuracy $\delta = 0.1\sigma$.

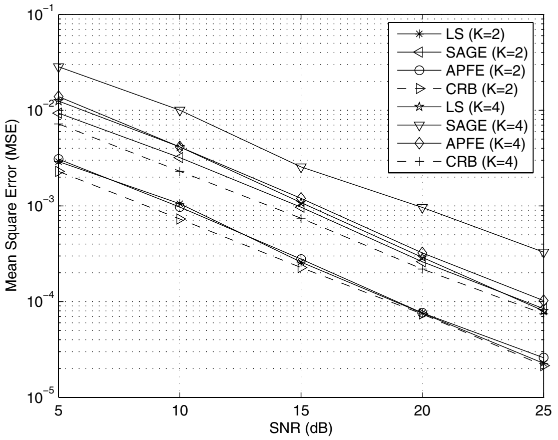

Fig. 8. Frequency offset estimation MSE comparison among several iterative methods with the stop accuracy $\delta = 0.1\sigma$.

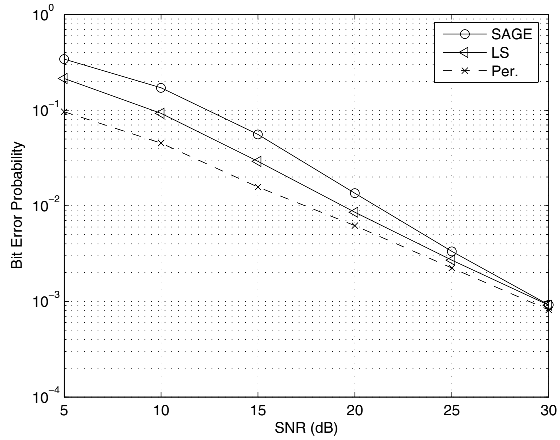

Fig. 10. Uncoded BER comparison of the proposed LS method and the reference SAGE method with the stop accuracy $\delta = 0.1\sigma$ in an OFDMA system with $K = 2$ users.

Due to the fact

$$\begin{aligned}&(\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{G}^{(i)})^H+\sigma^2(\mathbf{G}^{(i)})^H\\&=(\mathbf{G}^{(i)})^H(\mathbf{G}^{(i)}\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{G}^{(i)})^H+\sigma^2\mathbf{I})\\&=((\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}+\sigma^2(\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H)^{-1})\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H(\mathbf{G}^{(i)})^H,\end{aligned}$$

using the matrix inversion lemma, we obtain an alternative form of $\mathcal{A}^{(i)}$ as

$$\mathcal{A}^{(i)}=((\mathbf{G}^{(i)})^H\mathbf{G}^{(i)}+\sigma^2(\mathbf{b}^{(i)}(\mathbf{b}^{(i)})^H)^{-1})^{-1}(\mathbf{G}^{(i)})^H \tag{37}$$

which gives a complexity advantage over (36).

b) Sufficient condition for the convergence: Let $\mathbf{X}$ be the set of $\hat{\mathbf{x}}$, possible estimated coefficients $\mathbf{x}$, $(\mathbf{Y},\mu)$ be a metric space where $\mathbf{y}$ and $\hat{\mathbf{y}}$ are in the space of $\mathbf{Y}$, the metric μ be the Euclidean distance, and $\{\mathbf{\Omega}_i\}$ with $\mathbf{\Omega}_i = |\mathbf{y}-\hat{\mathbf{y}}^{(i)}|$ be a sequence of functions from $\mathbf{X}$ to $\mathbf{Y}$. For each value of $\mathbf{x}$ and for any scalar $\xi > 0$, if there exists an integer Π such that $\mu(\mathbf{\Omega}_m, \bar{\mathbf{\Omega}}) < \xi$ for all $m > \Pi$, we would say that $\mathbf{\Omega}_i$ converges pointwise to $\bar{\mathbf{\Omega}}$. Or based on Cauchy convergence criterion (sufficient and necessary condition), for any scalar $\xi > 0$, if there exists an integer Π such that for all $n, m > \Pi$, $\mu(\mathbf{\Omega}_m, \mathbf{\Omega}_n) < \xi$ holds, then $\mathbf{\Omega}_i$ converges pointwisely. In the following, we will prove that the sequence $\{\mathbf{\Omega}_i\}$ converges pointwisely based on the Cauchy convergence criterion.

Since $\mathbf{y}$ is the received signal, it is bounded as $|\mathbf{y}| < c_1$. We also have $|\hat{\mathbf{y}}^{(0)}| < c_2$. We obtain the value of $|\mathbf{y}-\hat{\mathbf{y}}^{(0)}|$ is bounded as $|\mathbf{y}-\hat{\mathbf{y}}^{(0)}| \le |\mathbf{y}|+|\hat{\mathbf{y}}^{(0)}| = c_1+c_2 \stackrel{\triangle}{=} c$. Consider the following condition

$$|\mathbf{y}-\hat{\mathbf{y}}^{(i+1)}| \le |\mathbf{y}-\hat{\mathbf{y}}^{(i)}|. \tag{38}$$

Then for any $m \geq 1$, we also have

$$|\mathbf{y} - \hat{\mathbf{y}}^{(0)}| - |\mathbf{y} - \hat{\mathbf{y}}^{(m)}| \leq c \tag{39}$$

$$|\mathbf{y} - \hat{\mathbf{y}}^{(i+1)}| - |\mathbf{y} - \hat{\mathbf{y}}^{(i+m)}| \leq |\mathbf{y} - \hat{\mathbf{y}}^{(i)}| - |\mathbf{y} - \hat{\mathbf{y}}^{(i+m)}|. \tag{40}$$

Next we make an assumption that the sequence $\{\mathbf{\Omega}_i\}$ is not pointwise convergent. Then the Cauchy convergence criterion is not satisfied and for each value of $\mathbf{x}$, there exists a scalar $\xi > 0$ and an integer Π_1, such that $\mu(\mathbf{\Omega}_0, \mathbf{\Omega}_{\Pi_1}) \geq \xi$ or $|\mathbf{y} - \hat{\mathbf{y}}^{(0)}| - |\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_1)}| \geq \xi$, and there exist integers Π_2, Π_3, ..., Π_n such that $|\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_m)}| - |\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_{m+1})}| \geq \xi$ where $0 \leq m \leq n-1$. For $n > \lceil \frac{c}{\xi} + 1 \rceil$, we obtain

$$\begin{aligned} &|\mathbf{y} - \hat{\mathbf{y}}^{(0)}| - |\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_1)}| + \cdots + |\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_m)}| \\ &-|\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_{m+1})}| + \cdots + |\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_{n-1})}| - |\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_n)}| \\ &= |\mathbf{y} - \hat{\mathbf{y}}^{(0)}| - |\mathbf{y} - \hat{\mathbf{y}}^{(\Pi_n)}| \geq n\xi > c\,. \end{aligned} \tag{41}$$

Since (39) contradicts (41), our assumption of non-convergence is wrong, and the condition in (38) is sufficient for the sequence $\{\mathbf{\Omega}_i\}$ to be pointwise convergent.

On the another hand, minimizing the value of $|y - y^{(i+1)}|^2$ is equivalent to maximizing the log-likelihood. Thus, the update criterion $|y - y^{(i+1)}| <= |y - y^{(i)}|$ increases the log-likelihood after each iteration, which guarantees the convergence of the proposed update process [9].

REFERENCES

[1] H. Sari and G. Karam, "Orthogonal frequency-division multiple access and its application to CATV networks," *Eur. Trans. Telecommun.*, vol. 45, pp. 507–516, Dec. 1998.

[2] J.-J. van de Beek, P. Borjesson, M.-L. Boucheret, D. Landstrom, J. Arenas, P. Odling, C. Ostberg, M. Wahlqvist, and S. Wilson, "A time and frequency synchronization scheme for multiuser OFDM," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1900–1914, Nov. 1999.

[3] M. Morelli, "Timing and frequency synchronization for the uplink of an OFDMA system," *IEEE Trans. Commun.*, vol. 52, pp. 296–306, Feb. 2004.

[4] Z. Cao, U. Tureli, and Y.-D. Yao, "Deterministic multiuser carrier-frequency offset estimation for interleaved OFDMA uplink," *IEEE Trans. Commun.*, vol. 52, pp. 1585–1594, Sep. 2004.

[5] Z. Cao, U. Tureli, Y.-D. Yao, and P. Honan, "Frequency synchronization for generalized OFDMA uplink," in *Proc. IEEE Global Telecommun. Conf.*, Nov.–Dec. 2004, vol. 2, pp. 1071–1075.

[6] H. Minn and X. Fu, "A new ranging method for OFDMA systems," in *Proc. IEEE GLOBECOM*, 2005, pp. 1435–1440.

[7] N. Balamurali and D. Jalihal, "An efficient algorithm for joint carrier frequency offset and channel estimation in IEEE 802.16 OFDM systems," in *Proc. IEEE Wireless Communication Syst. Conf.*, Sep. 2004, pp. 428–432.

[8] M.-O. Pun, S.-H. Tsai, and C.-C. Kuo, "An EM-based joint maximum likelihood estimation of carrier frequency offset and channel for uplink OFDMA systems," in *Proc. IEEE VTC*, Sep. 2004, vol. 1, pp. 598–602.

[9] ——, "Joint maximum likelihood estimation of carrier frequency offset and channel in uplink OFDMA systems," in *Proc. IEEE GLOBECOM*, Nov.–Dec. 2004, vol. 6, pp. 3748–3752.

[10] R. E. Miller, *Optimization, Foundations and Applications*. New York: John Wiley and Sons, Inc. 2000.

[11] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, Inc. 1993.

[12] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. New York: Academic Press, Inc, 1981.

[13] P. Venkataraman, *Applied Optimization With Matlab Programming*. New York: John Wiley and Sons, Inc, 2001.

[14] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory,*, vol. 20, pp. 591–598, Sep. 1974.

[15] S. Sezginer and P. Bianchi, "Joint frequency offset and channel estimation in the OFDMA uplink: Cramer-Rao bound and training sequence design," in *Proc. IEEE PAWC*, June 2005, pp. 585–589.

[16] M. O. Pun, M. Morelli, and C. C. Kuo, "Maximum likelihood synchronization and channel estimation for OFDMA uplink transmissions," *IEEE Trans. Commun.*, vol. 54, pp. 726–736, Apr. 2006.

[17] G. Stewart, *Matrix Algorithms*. Heidelberg, Germany: Library of Congress Cataloging-in-Publication Data, Inc., 1998.

Yanxin Na (M'05) was born in Shenyang, China. She received the B.S. and M.S. degrees from the Northeast University of China in 1993 and 1996, respectively, and the M.S. and Ph.D. degrees in electrical engineering from The University of Texas at Dallas in 2001 and 2005, respectively. Her doctoral thesis focused on the tradeoff study between diversity orders and channel estimation errors in broadband wireless communication systems.

From 1996 to 1999, she was with the Chinese Communication Technology Institution, Beijing, China, and was involved in the development of the WLL SCDMA system. Currently, she is working on algorithms (patents) design for WiMAX systems at Navini Networks. Her several patents about synchronization or beam-forming are on processing, and some of her algorithms are running in Navini's current WiMAX products. Her research interests include SDMA, TDD/FDD beam-forming, synchronization, channel or frequency offset estimation, multiuser detection, ultrawide band (UWB) communications, MIMO, and OFDM/OFDMA communications.

Dr. Na was granted the best doctoral dissertation award from the Department of Electrical Engineering, The University of Texas at Dallas, 2005.

Hlaing Minn (S'99-M'01-SM'07) received his B.E. degree in Electronics from Yangon Institute of Technology, Yangon, Myanmar, in 1995, M.Eng. degree in Telecommunications from Asian Institute of Technology (AIT), Pathumthani, Thailand, in 1997 and Ph.D. degree in Electrical Engineering from the University of Victoria, Victoria, BC, Canada, in 2001.

He was with the Telecommunications Program in AIT as a laboratory supervisor during 1998. He was a research assistant from 1999 to 2001 and a post-doctoral research fellow during 2002 in the Department of Electrical and Computer Engineering at the University of Victoria. Since September 2002, he has been with the Erik Jonsson School of Engineering and Computer Science, the University of Texas at Dallas, USA, as an Assistant Professor. His research interests include wireless communications, statistical signal processing, error control, detection, estimation, synchronization, signal design, cross-layer design, and cognitive radios. He is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.