A New Method For Increasing the Accuracy of EM-based Channel Estimation

Seyed Mohammad-Sajad SADOUGH, Mohammad MODARRESI

Cognitive Telecommunication Research Group, Department of Electrical Engineering,
Shahid Beheshti University G. C., 1983963113, Tehran, Iran

s_sadough@sbu.ac.ir

Abstract. *It was recently shown that the detection performance can be significantly improved if statistics of channel estimation errors are available and properly used at the receiver. Although in pilot-only channel estimation it is usually straightforward to characterize the statistics of channel estimation errors, this is not the case for the class of data-aided (semi-blind) channel estimation techniques. In this paper, we focus on the widely-used data-aided channel estimation techniques based on the expectation-maximization (EM) algorithm. This is achieved by a modified formulation of the EM algorithm which provides the receiver with the statistics of the estimation errors and properly using this additional information. Simulation results show that the proposed data-aided estimator outperforms its classical counterparts in terms of accuracy, without requiring additional complexity at the receiver.*

Keywords

EM algorithm, VBEM algorithm, joint iterative data detection and channel estimation, characterizing channel estimation errors, OFDM.

1. Introduction

It is well known that reliable coherent data detection is not possible unless an accurate estimate of the channel is available at the receiver. If the channel changes slowly, one can use pilot symbol assisted modulation (PSAM) [1] for obtaining channel state information (CSI) at the receiver. However, obtaining such an accurate estimate in highly mobile environments *only* through the use of pilots, would require inserting multiple training symbols into the data frame, which can result in a significant reduction of the spectral efficiency. Semi-blind channel estimation can enhance system performance by exploiting the unknown data symbols in addition to few pilots (usually one or two used for algorithm initialization) in the channel estimation process [2]. Recent works have reported promising results on the combination of channel estimation and data detection process. For instance, the use of the expectation-maximization (EM) algorithm for joint channel estimation and data detection is suggested in [3], [4]. More recently, the variational Bayesian (VB) inference [5], [6], which is closely related to mean-field methods in statistical physics, has been proposed as an effective method for tractable receiver design. However, regardless of the deployed technique (pilot-only or data-aided), channel estimation is an *imperfect* process and the poor quality of channel estimates degrades the detection performance at the receiver.

Very recently, assuming *pilot-only* channel estimation, the authors have shown that the *a posteriori* probability density function (pdf) of the channel conditioned on its estimate (which characterizes the statistics of the estimation errors), if exploited properly, has the potential to improve the receiver's performance in different transmission system (orthogonal frequency division multiplexing in [8], multiple-input-multiple-output in [9] and [11]), especially when the channel is estimated by a low number of *pilots*. Note that the key element for designing the improved receivers of [8], [9] and [11] was the availability of the the estimation error statistics, which is rather straightforward in *pilot-only* channel estimation. Our initial results provided in [10] indicated that there is a great potential to improve iterative semi-blind receivers if one efficiently derives and exploits the pdf of the channel conditioned on its estimate at each iteration. The main idea of [10] is generalized in this paper and related to more elaborated data-aided receivers as explained below.

In this paper, we focus on data-aided joint channel estimation and data detection where the derivation of the aforementioned a *posteriori* probability is not that straightforward. More precisely, we address data-aided methods based on the EM algorithm which is a widely adopted technique. More precisely, we propose modifications in the formulation of the EM estimator **i)** characterizing the estimation error statistics, and **ii)** using properly these statistics for deriving improved detection metrics robust to channel estimation errors. In this way, we propose an improved EM-based estimator that is capable of supplying the estimation error statistics at each iteration and hence reducing the impact of channel estimation inaccuracies at their respective receivers. We also formulate the VB estimation formalism adapted to our model. Then, we extend our initial results in [10] by deriving the relation that exists between the EM and the VBEM estimators. We will see that the proposed receiver becomes similar to the VBEM estimator for a particular choice of its parameters, and this observation will be confirmed by simulations. Results presented in this paper and the investigated similarity between the proposed EM algorithm and the VBEM algorithm open the way for searching other improved data-aided channel estimation techniques, that will be investigated in our future research.

Notational conventions are as follows. $\mathcal{D}_\mathbf{x}$ is a diagonal matrix with diagonal elements $\mathbf{x} = [x_1, \ldots, x_M]^T$, $\mathbb{E}_\mathbf{x}[.]$ or $\langle . \rangle_{p(\mathbf{x})}$ refer to expectation with respect to the random vector $\mathbf{x}$, $\mathbf{I}_M$ denotes an $(M \times M)$ identity matrix, $\propto$ denotes equality up to a normalization factor, $\mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$ denotes complex Gaussian vector distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}$; $|.|$ and $\|.\|$ denote absolute value and vector norm, respectively, $(.)^T, (.)^\dagger$ and $(.)^*$ denote vector transpose, Hermitian transpose and conjugation, respectively.

2. System Description and Channel Model

We consider a coded OFDM system with M subcarriers through a frequency-selective multipath fading channel[1].

We assume a block-fading channel model where each frame of size M_{frame} symbols corresponds to M_{block} independent fading blocks. Notice that $M_{\text{block}} = 1$ returns to the quasi-static channel model whereas $M_{\text{block}} = M_{\text{frame}}$ returns to the fast-fading channel model. Since the channel is assumed to be block-fading, for estimating the k-th complex channel frequency coefficient H_k, we receive $N = M_{\text{frame}}/M_{\text{block}}$ independent observations (see Fig. 1). At the receiver, after removing the cyclic prefix (CP), the signal corresponding to the k-th subcarrier in a given fading block writes

$$\mathbf{y}_k = H_k \mathbf{s}_k + \mathbf{z}_k \qquad \text{for } k = 1, \ldots, M \tag{1}$$

where the $(1 \times N)$ vector $\mathbf{y}_k = [y_{1,k}, \ldots, y_{N,k}]$, the entries of the noise vector $\mathbf{z}_k$ are assumed to be zero-mean circularly symmetric complex Gaussian (ZMCSCG), and the definition of $\mathbf{s}_k$ and $\mathbf{z}_k$ follow that of $\mathbf{y}_k$. For the sake of notational simplicity, we will not specify hereafter the subscript k in (1).

3. Improved EM-based Estimation

Considering the model (1), the problem is to estimate H from $\mathbf{y}$, where $\mathbf{s}$ is unknown. Thus, we have missing data and the estimate of H has no closed form. In such situations, the EM algorithm [13] is often used to maximize the expectation of the likelihood function over all possible missing and hidden variables. The ultimate aim in this Section is to derive the estimation error statistics in addition to the estimate of H. To this end, we propose to decompose the

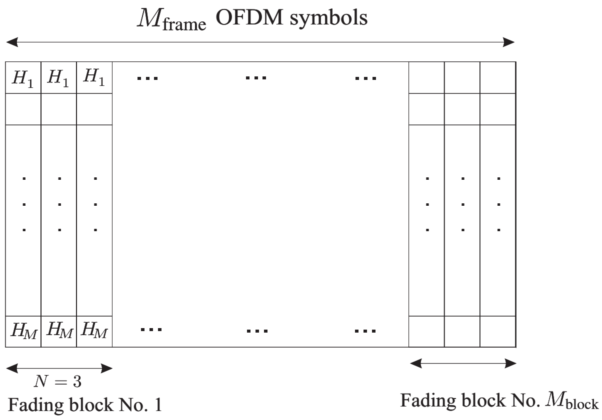

Fig. 1. The considered block-fading channel model with M_{block} independent fading blocks. In the above figure, we have $N = M_{\text{frame}}/M_{\text{block}} = 3$.

AWGN vector $\mathbf{z}$ in (1) into the sum of two independent Gaussian terms as

$$\mathbf{z} = \mathbf{s}\,\mathbf{E} + \mathbf{u} \tag{2}$$

where $\mathbf{E} = \mathcal{D}_\mathbf{w}$ with $\mathbf{w} = [w_1, \ldots, w_N]^T$ and $\mathbf{u} = [u_1, \ldots, u_N]$ are noise vectors such that $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_N - \alpha^2 \mathbf{F}\mathbf{F}^\dagger)$ and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \alpha^2 \mathbf{I}_N)$ with $\mathbf{F} = \mathcal{D}_\mathbf{s}$. Note that α^2 is the variance of each component in the noise vector $\mathbf{w}$ and denotes the part of the noise power allocated to $\mathbf{E}$ inside the noise decomposition model (2). More clearly, we define the positive design parameter $\rho = \alpha^2/\sigma_z^2$ as the proportion of noise that is assigned to $\mathbf{u}$.

The above noise decomposition allows us to write (1) in an equivalent form as

$$\begin{cases} \widetilde{\mathbf{H}} &= H\,\mathbf{I}_N + \mathbf{E} \\ \mathbf{y} &= \mathbf{s}\widetilde{\mathbf{H}} + \mathbf{u} \end{cases} \tag{3}$$

where $\widetilde{\mathbf{H}} = \mathcal{D}_\mathbf{R}$ with $\mathbf{R} = \left[\widetilde{H}_1, \ldots, \widetilde{H}_N\right]^T$.

This introduces the diagonal matrix $\widetilde{\mathbf{H}}$ which will let us derive the pdf $p(H|H^{(t)})$ in our subsequent developments, even if the two-stage observation model (3) is equivalent to (1).

Let $\mathbf{X} = \{\mathbf{y}, \mathbf{s}, \widetilde{\mathbf{H}}\}$ be the *complete data set* in the EM algorithm terminology. We are searching for H that maximizes $\log p(\mathbf{X}|H)$. After initialization by a short pilot sequence at the beginning of the frame, the EM algorithm alternates between the expectation (E) step and the maximization (M) step (until some stopping criterion) to produce a sequence of estimates $\{H^{(t)}, t = 0, 1, \ldots, t_{\max}\}$. When applied to (3), each step can be written as follows.

3.1 E-step: Computation of the Q-function

In the E-step, the conditional expectation of the complete log-likelihood given the observed vector and the current estimate up to time instant t, i.e., $H^{(t-1)}$ is computed.

[1] Although here we have considered the widely-used OFDM signal model, it is important to mention that the proposed approach can be extended to any transmission scenario.

This quantity is called the *auxiliary* or Q-function and is given by

$$Q\big(H,H^{(t-1)}\big) = \mathbb{E}_{\mathbf{s},\widetilde{\mathbf{H}}}\Big[\log p(\mathbf{y},\mathbf{s},\widetilde{\mathbf{H}}|H)\Big|\mathbf{y},H^{(t-1)}\Big]. \quad (4)$$

The complete-data likelihood required for the computation of the Q-function is:

$$p(\mathbf{y},\mathbf{s},\widetilde{\mathbf{H}}|H) = p(\mathbf{y}|\mathbf{s},\widetilde{\mathbf{H}})\,p(\widetilde{\mathbf{H}}|H)\,p(\mathbf{s}) \quad (5)$$

where we have used the fact that conditioned on $\widetilde{\mathbf{H}}$ and $\mathbf{s}$, $\mathbf{y}$ is independent of H. Besides, $\mathbf{s}$ which results from coding and interleaving of the bit sequence is independent of $\widetilde{\mathbf{H}}$ and H. Thus the Q-function is simplified to

$$Q(H,H^{(t-1)}) = \mathbb{E}_{\mathbf{s},\widetilde{\mathbf{H}}}\left[\log p(\widetilde{\mathbf{H}}|H)|\mathbf{y},H^{(t-1)}\right] + \text{cst.} \quad (6)$$

where cst. is a constant term gathering all terms that do not depend on H and $p(\widetilde{\mathbf{H}}|H)$ is given by

$$p(\widetilde{\mathbf{H}}|H) = \prod_{i=1}^{N} p(\widetilde{H}_i|H) = \prod_{i=1}^{N} \frac{1}{\pi\alpha^2}\exp\left\{-\frac{|\widetilde{H}_i - H|^2}{\alpha^2}\right\}. \quad (7)$$

It can be easily verified that the Q-function (6) is obtained as

$$Q(H,H^{(t-1)}) = -\sum_{i=1}^{N} \frac{|\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i} - H|^2}{\alpha^2} + \text{cst.} \quad (8)$$

where $\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i} \triangleq \mathbb{E}_{s_i,\widetilde{H}_i}[\widetilde{H}_i|y_i,H^{(t-1)}]$. It is obvious from (8) that the E-step requires only the computation of $\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i}$ as follows.

$$\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i} = \sum_{s_i} \mu_{\widetilde{H}_i}(s_i)\,p(s_i|y_i,H^{(t-1)}) \quad (9)$$

where the posterior mean $\mu_{\widetilde{H}_i}(s_i)$ of $\widetilde{H}_i$ is

$$\mu_{\widetilde{H}_i}(s_i) = \int_{\widetilde{H}_i} \widetilde{H}_i\,p(\widetilde{H}_i|y_i,s_i,H^{(t-1)})\,d\widetilde{H}_i. \quad (10)$$

After some algebra provided in the appendix, $\mu_{\widetilde{H}_i}(s_i)$ is shown to be given by

$$\mu_{\widetilde{H}_i}(s_i) = H^{(t-1)} + \rho\, s_i^*\left(y_i - s_i H^{(t-1)}\right). \quad (11)$$

We can now derive $\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i}$ of equation (9), by evaluating the expectation of (11) as

$$\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i} = H^{(t-1)} + \rho\, y_i\langle s_i^*\rangle_{s_i|y_i,H^{(t-1)}} - \rho H^{(t-1)}\langle|s_i|^2\rangle_{s_i|y_i,H^{(t-1)}} \quad (12)$$

where

$$\langle s_i\rangle_{s_i|y_i,H^{(t-1)}} = \sum_{s_i} s_i\,p(s_i|y_i,H^{(t-1)}). \quad (13)$$

3.2 M-step: Maximization of the Q-function

In this step, the estimated parameter H is updated according to:

$$H^{(t)} = \arg\max_{H}\left\{Q\big(H,H^{(t-1)}\big)\right\}. \quad (14)$$

By substituting (8) in (14) the channel update formula is given by

$$H^{(t)} = \frac{\sum_{i=1}^{N}\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i}}{N} \quad (15)$$

$$= H^{(t-1)}\left(1 - \frac{\rho}{N}\langle\|\mathbf{s}\|^2\rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}\right) + \frac{\rho}{N}\mathbf{y}\langle\mathbf{s}^\dagger\rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}. \quad (16)$$

3.3 Deriving the Estimation Error Statistics

From (8), we observe that $\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i}$ is no more than the maximum likelihood (ML) estimate of H from the observation model:

$$\langle\widetilde{H}_i\rangle_{s_i,\widetilde{H}_i} = H + w_i \quad \text{for} \quad i = 0,\cdots,N-1 \quad (17)$$

where $w_i \sim \mathcal{CN}(0,\alpha^2)$. Looking at (15) and (17), it is obvious that:

$$H^{(t)} = H + Z_1 \quad \text{where } Z_1 \sim \mathcal{CN}(0,\alpha^2/N). \quad (18)$$

Remember that the main motivation behind working with the model (3) is to characterize the estimation error statistics. This is achieved by using the equation (18) where the pdf $p(H^{(t)}|H)$ is available and equal to $\mathcal{CN}(H,\alpha^2/N)$. Assuming that the channel coefficient H is *a priori* distributed as $p(H) = \mathcal{CN}(0,\sigma_h^2)$, and using the Bayes formula, one can derive at each iteration of the EM algorithm the posterior distribution of the channel conditioned on its estimate as

$$p(H|H^{(t)}) = \mathcal{CN}(m_H^{(t)},\sigma_H^2) \quad (19)$$

where $m_H^{(t)} = \delta H^{(t)}$, $\sigma_H^2 = \delta\alpha^2/N$ and $\delta = \frac{\sigma_h^2}{\sigma_h^2+\alpha^2/N}$. The availability of the pdf (19) constitutes an interesting feature of the proposed EM estimator that will be exploited to improve the receiver's performance. Obviously, this is a direct consequence of the equivalent two-step modeling of equation (3). Details on the manner the receiver exploits the pdf (19) for performance improvement are provided in Section 5.

4. Relation Between the Improved EM and Variational Bayesian Estimation

4.1 Variational Bayesian Estimation Formalism

The optimal estimate of the symbol vector $\mathbf{s}$ in (1) by using the maximum *a posteriori* (MAP) rule is given by

$$\hat{\mathbf{s}}^{\text{MAP}} = \arg\max_{\mathbf{s}}\; p(\mathbf{s}|\mathbf{y}). \quad (20)$$

The objective function in (20) can be written as

$$p(\mathbf{s}|\mathbf{y}) = \int p(\mathbf{s},H|\mathbf{y})\,dH = \int p(\mathbf{s}|H,\mathbf{y})\,p(H|\mathbf{y})\,dH \quad (21)$$

where H is regarded as a nuisance parameter. Here, we assume that the channel is not known prior to data detection and thus the optimal solution is infeasible to obtain. The central idea of VB approximation [5] is to approximate the exact but intractable joint distribution $p(\mathbf{s},H|\mathbf{y})$ into a product of marginal probabilities $q(\mathbf{s})$ and $q(H)$. These two marginal probabilities are obtained as

$$\{q^\star(\mathbf{s}),q^\star(H)\} = \underset{q(\mathbf{s}),q(H)}{\arg\min}\,\mathrm{KL}\left[q(\mathbf{s})q(H)||p(\mathbf{s},H|\mathbf{y})\right], \quad (22)$$

subject to:

$$\begin{aligned} &q(\mathbf{s},H) = q(\mathbf{s})q(H),\\ &\int q(\mathbf{s})\,d\mathbf{s} = 1, q(\mathbf{s}) \geq 0\ \forall \mathbf{s},\\ &\int q(H)\,dH = 1, q(H) \geq 0\ \forall H \end{aligned}$$

where distributions $\{q^\star(\mathbf{s}),q^\star(H)\}$ are obtained as the result of the minimization problem (22) and

$$\mathrm{KL}\left[q(\mathbf{s})q(H)||p(\mathbf{s},H|\mathbf{y})\right] \triangleq \int q(\mathbf{s})q(H)\ln\frac{q(\mathbf{s})q(H)}{p(\mathbf{s},H|\mathbf{y})}d\mathbf{s}\,dH$$

is the Kullback-Leibler divergence [14].

By using the Lagrangian formalism, it can be seen that any solution of the optimization problem (22) is obtained by alternating between the VBE-step and the VBM-step as [7]

$$\mathrm{VBE}: q^{(t)}(\mathbf{s}) \propto \exp\{\langle \ln p(\mathbf{s},H,\mathbf{y})\rangle_{q^{(t-1)}(H)}\}, \quad (23)$$

$$\mathrm{VBM}: q^{(t)}(H) \propto \exp\{\langle \ln p(\mathbf{s},H,\mathbf{y})\rangle_{q^{(t)}(\mathbf{s})}\} \quad (24)$$

where the superscript (t) denotes the iteration index, due to the fact that the solution of (22) is not explicit since $q(\mathbf{s})$ and $q(H)$ depend on each other.

After simplification, (23) becomes

$$\begin{aligned} q^{(t)}(\mathbf{s}) &\propto \exp\left\{\langle \ln p(\mathbf{s},H,\mathbf{y})\rangle_{q(H)^{(t-1)}}\right\}\\ &\propto \exp\left\{\langle \ln[p(\mathbf{y}|H,\mathbf{s})p(H)p(\mathbf{s})]\rangle_{q^{(t-1)}(H)}\right\}\\ &\propto p(\mathbf{s})\,\exp\left\{\langle \ln p(\mathbf{y}|H,\mathbf{s})\rangle\right\}_{q^{(t-1)}(H)} \end{aligned} \quad (25)$$

where we have used the independence between $\mathbf{s}$ and H, and omitted all terms that do not depend on $\mathbf{s}$. Since the noise in (1) has a Gaussian distribution, (25) writes

$$q^{(t)}(\mathbf{s}) \propto p(\mathbf{s})\exp\left\{-\frac{\langle\|\mathbf{y}-H\mathbf{s}\|^2\rangle_{q^{(t-1)}(H)}}{\sigma_z^2}\right\}. \quad (26)$$

Let us now calculate $q^{(t)}(H)$. Starting from (24) we have

$$\begin{aligned} q^{(t)}(H) &\propto \exp\left\{\langle \ln p(\mathbf{s},H,\mathbf{y})\rangle_{q^{(t)}(\mathbf{s})}\right\}\\ &\propto \exp\left\{\langle \ln[p(\mathbf{y}|H,\mathbf{s})p(H)p(\mathbf{s})]\rangle_{q^{(t)}(\mathbf{s})}\right\}\\ &\propto p(H)\exp\left\{\langle \ln p(\mathbf{y}|\mathbf{s},H)\rangle_{q^{(t)}(\mathbf{s})}\right\}\\ &\propto p(H)\exp\left\{-\frac{\langle\|\mathbf{y}-H\mathbf{s}\|^2\rangle_{q^{(t)}(\mathbf{s})}}{\sigma_z^2}\right\} \end{aligned} \quad (27)$$

where we have omitted all terms that do not depend on H. Since $p(H) = \mathcal{CN}(0,\sigma_h^2)$, after simple calculus, (27) can be rewritten as

$$\begin{aligned} q^{(t)}(H) \propto \exp\Bigg\{&-\frac{|H|^2\langle\|\mathbf{s}\|^2\rangle_{q^{(t)}(\mathbf{s})}}{\sigma_z^2} - \frac{|H|^2}{\sigma_h^2}\\ &+2\,\mathbb{R}e\left[\frac{H^*\mathbf{y}\langle\mathbf{s}^\dagger\rangle_{q^{(t)}(\mathbf{s})}}{\sigma_z^2}\right]\Bigg\}. \end{aligned} \quad (28)$$

After some algebraic manipulations, we get

$$q^{(t)}(H) = \frac{1}{\pi\beta^{(t)}}\exp\left\{-\frac{\left|H-\mu^{(t)}\right|^2}{\beta^{(t)}}\right\} \quad (29)$$

with

$$\beta^{(t)} = \frac{\sigma_z^2\,\sigma_h^2}{\sigma_z^2+\sigma_h^2\langle\|\mathbf{s}\|^2\rangle_{q^{(t)}(\mathbf{s})}} \quad (30)$$

$$\mu^{(t)} = \beta^{(t)}\left[\frac{\mathbf{y}\langle\mathbf{s}^\dagger\rangle_{q^{(t)}(\mathbf{s})}}{\sigma_z^2}\right]. \quad (31)$$

Then new VBEM steps are written as follows

$$\mathrm{VBE}: q^{(t)}(\mathbf{s}) \propto p(\mathbf{s})\exp\left\{-\frac{\langle\|\mathbf{y}-H\mathbf{s}\|^2\rangle_{q^{(t-1)}(H)}}{\sigma_z^2}\right\} \quad (32)$$

$$\mathrm{VBM}: q^{(t)}(H) = \mathcal{CN}\left(\mu^{(t)},\beta^{(t)}\right) \quad (33)$$

where

$$\beta^{(t)} = \frac{\sigma_z^2\,\sigma_h^2}{\sigma_z^2+\sigma_h^2\langle\|\mathbf{s}\|^2\rangle_{q^{(t)}(\mathbf{s})}} \quad (34)$$

$$\mu^{(t)} = \beta^{(t)}\left[\frac{\mathbf{y}\,\langle\mathbf{s}^\dagger\rangle_{q^{(t)}(\mathbf{s})}}{\sigma_z^2}\right]. \quad (35)$$

4.2 Improved EM and its Relation to the VBEM Methodology

It is obvious that Equation (16) depends on the parameter ρ. Now if in (16), one substitutes ρ by $\frac{N}{\langle\|\mathbf{s}\|^2\rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}}$, we get

$$H^{(t)} = \frac{\mathbf{y}\langle\mathbf{s}^\dagger\rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}}{\langle\|\mathbf{s}\|^2\rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}}. \quad (36)$$

By inserting (36) and the aforementioned ρ in (19) and noting that $\alpha^2 = \rho\sigma_z^2$, we get

$$\sigma_H^{2\,(t)} = \frac{\sigma_z^2\sigma_h^2}{\sigma_z^2 + \sigma_h^2 \langle \|\mathbf{s}\|^2 \rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}}, \tag{37}$$

$$m_H^{(t)} = \sigma_H^{2\,(t)} \left[\frac{\mathbf{y}\,\langle \mathbf{s}^\dagger \rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}}{\sigma_z^2} \right]. \tag{38}$$

Comparing (34) and (35) to (37) and (38), it is obvious that when $\rho = \frac{N}{\langle \|\mathbf{s}\|^2 \rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}}$, the pdf $p(H|H^{(t)})$ derived from the improved EM algorithm is equal to the pdf $q^{(t)}(H)$ derived from the VBEM algorithm. For more clarity, we mention that the equivalent of $p(H|H^{(t)})$ in the VBEM algorithm is the distribution $q^{(t)}(H)$, with equality when ρ is adaptively selected and is equal to $\rho = \frac{N}{\langle \|\mathbf{s}\|^2 \rangle_{\mathbf{s}|\mathbf{y},H^{(t-1)}}}$.

5. Improved Iterative Detection and Decoding

Here, we explain how by a proper use of (19), the detection performance at the receiver can be improved. More precisely, we propose to use the distribution $p(H|H^{(t)})$ of (19) (improved EM) for deriving a modified likelihood function that is used at the receiver for data detection. By using the improved EM algorithm, we can evaluate, at each iteration, an averaged likelihood function as:

$$\begin{aligned} p(y_i|H^{(t)}, s_i) &= \int p(y_i|H, s_i)\; p(H|H^{(t)})\; \mathrm{d}H \\ &= \mathbb{E}_{H|H^{(t)}} \left[p(y_i|H, s_i) \,\big|\, H^{(t)} \right] \end{aligned} \tag{39}$$

where $p(H|H^{(t)})$ is the channel posterior distribution of (19). Then by using a theorem derived in the appendix of [8], we obtain

$$p(y_i|H^{(t)}, s_i) = \frac{1}{\pi\left(\sigma_z^2 + \sigma_H^2 |s_i|^2\right)} \exp\left\{ -\frac{\left| y_i - m_H^{(t)} s_i \right|^2}{\sigma_z^2 + \sigma_H^2 |s_i|^2} \right\}. \tag{40}$$

In this way, we can define:

$$\mathcal{D}_{\mathcal{M}}(s_i, y_i, H^{(t)}) \triangleq -\log p(y_i|H^{(t)}, s_i) \tag{41}$$

which is referred to as the *improved* ML decision metric that will be used under imperfect CSIR. Now, the receiver uses the modified likelihood $p(y_i|H^{(t)}, s_i)$ to evaluate by marginalization the bit metrics that are fed from the detector to the soft-input soft-output (SISO) decoder in our BICM reception scheme.

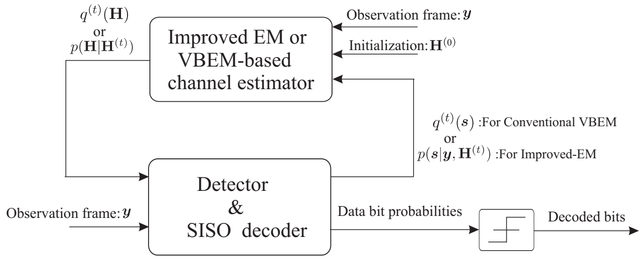

Fig. 2. EM and VBEM-based channel estimation combined with the decoding process.

At the receiver, we perform MAP symbol detection and channel decoding in an iterative manner. The block diagram of the receiver is shown in Fig. 2. Besides the channel estimation part, the rest of the receiver principally consists of the combination of two sub-blocks that exchange soft probabilistic information with each other. The first sub-block, referred to as soft demapper (also called detector), produces bit metrics (probabilities) from the input symbols and the second one is a SISO decoder. Using the modified likelihood function (39) (or equivalently the improved ML metric of (40)) has two main advantages for our robust receiver design, namely:

- in deriving improved bit metrics at the detector output,
- in deriving the *a posteriori* probability $p(s_i|y_i, H^{(t)})$ involved in (13) by taking into account channel estimation errors.

Each item is briefly explained in the following.

5.1 Derivation of Improved Bit Metrics

Let $d_k^{j,m}$ be the m-th ($m = 1, ..., B$) coded and interleaved bit corresponding to the constellation symbol s_k transmitted at the j-th time slot over the k-th subcarrier. We denote by $L(d_k^{j,m})$ the coded log-likelihood ratio (LLR) of the bit $d_k^{j,m}$ at the output of the detector. We partition the set $\mathcal{C}$ that contains all possibly-transmitted symbol s_k into two sets $\mathcal{C}_0^m$ and $\mathcal{C}_1^m$, for which the m-th bit of s_k equals "0" or "1", respectively. We have:

$$L(d_k^{j,m}) = \log \frac{p(y_k|d_k^{j,m} = 1, H_k^{(t)})}{p(y_k|d_k^{j,m} = 0, H_k^{(t)})} \tag{42}$$

$$= \log \frac{\sum\limits_{s_k \in \mathcal{C}_1^m} e^{-\mathcal{D}_{\mathcal{M}}(s_k, y_k, H_k^{(t)})} \prod\limits_{\substack{n=1 \\ n \neq m}}^{B} P_{\mathrm{dec}}^1\left(d_k^{j,n}\right)}{\sum\limits_{s_k \in \mathcal{C}_0^m} e^{-\mathcal{D}_{\mathcal{M}}(s_k, y_k, H_k^{(t)})} \prod\limits_{\substack{n=1 \\ n \neq m}}^{B} P_{\mathrm{dec}}^0\left(d_k^{j,n}\right)} \tag{43}$$

where $P_{\mathrm{dec}}^1(d_k^{j,n})$ and $P_{\mathrm{dec}}^0(d_k^{j,n})$ are *prior* probabilities coming from the SISO decoder. By doing so, the LLRs are adapted to the imperfect channel knowledge available at the receiver and consequently the impact of channel uncertainty on the SISO decoder performance is reduced. We refer to the latter approach as *improved* MAP detector.

5.2 Derivation of the *a posteriori* Probability $p(s_i|y_i, H^{(t)})$

The modified likelihood of (40) can also be used for the evaluation of $p(s_i|y_i, H^{(t)})$ which plays an important role in the convergence of the EM algorithm. At the t-th iteration, we have:

$$p(s_i|y_i, H^{(t)}) = \frac{p(y_i|H^{(t)}, s_i)p(s_i)}{\sum_{s_i} p(y_i|H^{(t)}, s_i)p(s_i)}. \tag{44}$$

Here, we use the averaged likelihood (40) for the evaluation of (44). More precisely, by using (40), channel uncertainties are taken into account for the evaluation of (44) at each iteration of the EM algorithm leading to controlled error propagation and faster convergence, as we will see in next Section.

6. Simulation Results

In this Section, we provide simulation results to compare the performance provided by the proposed improved EM-based estimator. We consider a bit-interleaved coded modulation (BICM) combined with OFDM where The parameters used throughout the simulations are as follows: One OFDM symbol is composed of $M = 40$ subcarriers. For channel coding, we consider the rate $1/3$ recursive systematic convolutional (RSC) code of constraint length 3 defined in octal form by $[5,7,7]_8$. The interleaver is pseudo-random and operates over the entire frame that contains 48 OFDM symbols. Data symbols belong to 16-QAM constellation with set-partition (SP) labeling. Performance evaluation is performed over the block-fading channel with parameters $M_{\text{block}} = 16$ (N=3) and $M_{\text{block}} = 3$ (N=16). Channel coefficients corresponding to different OFDM subcarriers are assumed uncorrelated and distributed according to the Rayleigh distribution. One OFDM pilot symbol is dedicated for initializing the channel in each fading block. The BICM receiver is an iterative one composed of a demapper (detector) and a SISO decoder as the main blocks. Moreover, we perform one SISO decoding iteration and 8 EM algorithm iterations.

Fig. 3 shows the BER performance versus E_b/N_0 (in dB) in the case of a block-fading channel with parameter $N = 3$. For comparison, we have also provided the BER obtained with perfect CSI as well as the BER obtained with pilot-only (obtained with one pilot symbol) and VBEM channel estimation. It can be seen that the improved EM provides a gain of about 1 dB at a BER of 10^{-4} with respect to the conventional EM. Moreover, the improved EM algorithm outperforms the VBEM algorithm. Similar plots are provided in Fig. 4 for parameter $N = 16$. In this case, since the number of observations for estimating each unknown parameter is increased, all investigated estimation schemes provide

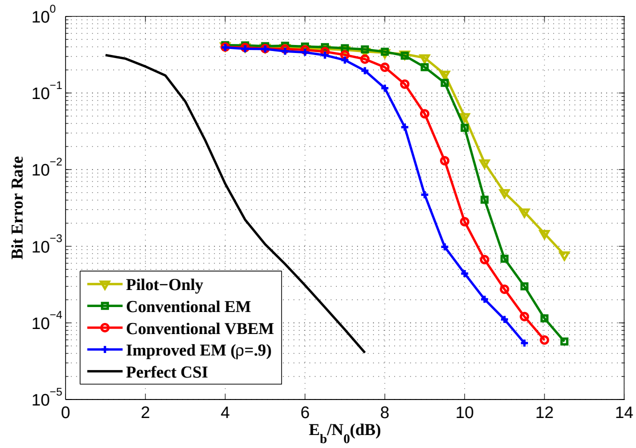

Fig. 3. BER performance versus E_b/N_0 in dB over a block-fading channel with $N = 3$.

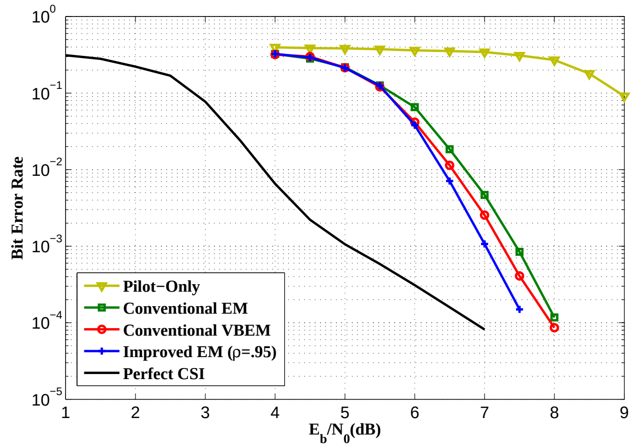

Fig. 4. BER performance versus E_b/N_0 in dB over a block-fading channel with $N = 16$.

lower BER performance. However, we observe that the proposed improved EM estimator still outperforms conventional EM and VBEM techniques.

Figs. 5 and 6 depict the BER versus the number of SISO iterations, at a fixed E_b/N_0 of 11 dB and 7.5 dB with $N = 3$ and $N = 16$, respectively. This allows us to evaluate the number of SISO iterations necessary to attain a target BER. From Fig. 5 we observe that the improved-EM detector requires 4 iterations to achieve a BER of 3×10^{-4} while the VBEM detector attains this BER after 6 iterations, and even more iterations are necessary for the conventional EM algorithm. A similar behavior is observed for the case with $N = 16$ in Fig. 6. Noting that each iteration involves the complicated forward-backward [15] decoding algorithm in addition to EM computations, reveals that the proposed detector is particularly useful for reducing the complexity at the receiver.

Moreover, it is worth mentioning that the aforementioned improvement provided by the proposed improved detector is obtained just by changing the metric of the EM algorithm, i.e., without requiring additional complexity at the receiver.

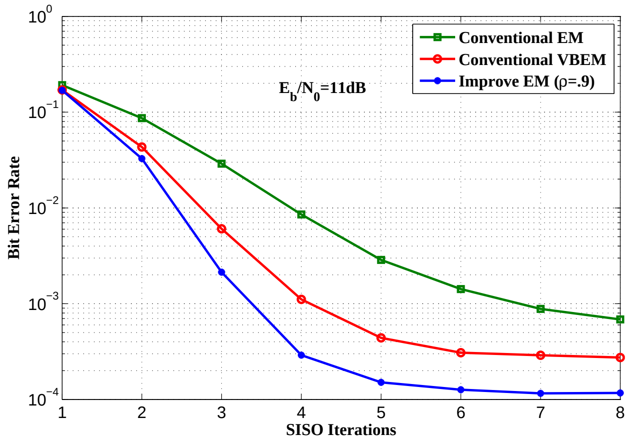

Fig. 5. BER performance versus SISO decoding iterations over a block-fading channel with $N = 3$, $E_b/N_0 = 11$ dB.

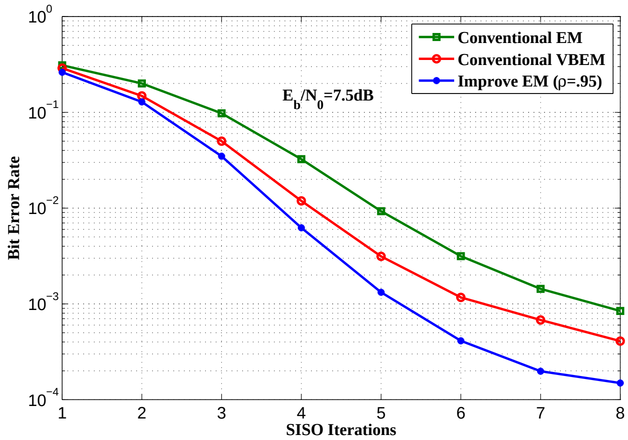

Fig. 6. BER performance versus SISO decoding iterations over a block-fading channel with $N = 16$, $E_b/N_0 = 7.5$ dB.

7. Conclusion

The problem of joint signal detection and data-aided (semi-blind) channel estimation based on the EM algorithm was investigated. We made some modifications in the formulation of the EM-estimator in order to derive the pdf characterizing the statistics of the estimation errors. This pdf was used at the receiver and led to a modification of the metrics used in the conventional EM iterative detector. Therefore, our approach does not increase the complexity at the receiver. We also formulated the VBEM estimator and then derived the relation that exists between the proposed improved EM algorithm and the more elaborated VBEM estimator. Our numerical results confirmed the superiority of the proposed detector in reducing the impact of channel estimation errors on the BER performance. Moreover, our proposed scheme reduces the number of decoding iterations necessary to achieve a target BER, compared to conventional EM and VBEM estimation techniques. Finally, we notice that the main idea and methodology used in this paper for improving the detector part for the EM algorithm can be used for improving the detection with the VBEM algorithm, that we will investigate in a future work.

Appendix: Derivation of the a Posteriori Mean (11)

In order to evaluate (11), we use the following theorem which is proposed in [16].

Theorem 1: Let $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}_x, \mathbf{R}_x)$ and $\mathbf{y} \sim \mathcal{CN}(\mathbf{m}_y, \mathbf{R}_y)$ be two complex Gaussian random vectors with the joint distribution

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{CN}\left(\begin{bmatrix} \mathbf{m}_x \\ \mathbf{m}_y \end{bmatrix}, \begin{bmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{bmatrix} \right)$$

where $\mathbf{R}_{xy} \triangleq \mathbb{E}[(\mathbf{x}-\mathbf{m}_x)(\mathbf{y}-\mathbf{m}_y)^*]$. Then, the conditional random vector $\mathbf{x}|\mathbf{y}$ is distributed as $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu} = \mathbf{m}_x + \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}(\mathbf{y}-\mathbf{m}_y)$ and covariance matrix $\boldsymbol{\Sigma} = \mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}$.

It is clear from (10) that we have to evaluate the mean of the posterior distribution $p(\widetilde{H}_i|y_i, s_i, H^{(t-1)})$ that we denote by $\mu_{\widetilde{H}_i}(s_i)$. To this end, we consider the above theorem and we set $\mathbf{x} = \widetilde{H}_i$ and $\mathbf{y} = y_i$. Note that $H^{(t-1)}$ and s_i are assumed to be known in the following. From (3), it is clear that $\widetilde{H}_i \sim \mathcal{CN}(\mathbf{m}_{\widetilde{H}_i}, \mathbf{R}_{\widetilde{H}_i})$ with $\mathbf{m}_{\widetilde{H}_i} = H$ and $\mathbf{R}_{\widetilde{H}_i} = \alpha^2$. Equivalently from (3), we have $y_i \sim \mathcal{CN}(\mathbf{m}_{y_i}, \mathbf{R}_{y_i y_i})$ with $\mathbf{m}_{y_i} = s_i H$ and $\mathbf{R}_{y_i y_i} = \sigma_z^2$. We also calculate $\mathbf{R}_{\widetilde{H}_i y_i}$ as

$$\begin{aligned} \mathbf{R}_{\widetilde{H}_i y_i} &= \mathbb{E}\left[\left(\widetilde{H}_i - \mathbf{m}_{\widetilde{H}_i}\right)\left(y_i - \mathbf{m}_{y_i}\right)^* \right] = \mathbb{E}\left[w_i \left(s_i w_i + u_i\right)^* \right] \\ &= \alpha^2 s_i^* . \end{aligned}$$

By using Theorem 1 we obtain

$$\mu_{\widetilde{H}_i}(s_i) = H^{(t-1)} + \frac{\alpha^2}{\sigma_z^2} s_i^* \left(y_i - s_i H^{(t-1)} \right). \qquad (45)$$

which is nothing but equation (11).

References

[1] CAVERS, J. K. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Transactions on Vehicular Technology*, 1991, vol. 40, no. 4, p. 686 - 693.

[2] CIRPAN, H. A., PANAYIRCI, E., DOGAN, H. Non-data-aided channel estimation for OFDM systems with space-frequency transmit diversity. *IEEE Transactions on Vehicular Technology*, 2006, vol. 55, no. 2, p. 449 - 457.

[3] MA, X., KOBAYASHI, H., SCHWARTZ, S. C. EM-based channel estimation algorithms for OFDM. *EURASIP Journal on Applied Signal Processing*, 2004, vol. 10, p. 1460 - 1477.

[4] LU, B., WANG, X., LI, Y. G. Iterative receivers for space-time block coded OFDM systems in dispersive fading channels. *IEEE Transactions on Wireless Communications*, 2002, vol. 1, no. 2, p. 213 - 225.

[5] CHRISTENSEN, L. P. B., LARSEN, J. On data and parameter estimation using the variational Bayesian EM-algorithm for block-fading frequency-selective MIMO channels. In*Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2006*. Toulouse (France), 2006, p. IV.

[6] LIN, D. D., LIM, T. J. A variational inference framework for soft-in soft-out detection in multiple-access channels. *IEEE Transactions on Information Theory*, 2009, vol. 55, no. 5, p. 2345 - 2363.

[7] STEFANATOS, S., KATSAGGELOS, A. K. Joint data detection and channel tracking for OFDM systems using the variational Bayes method. In *Proceedings of the IEEE Signal Processing Advances in Wireless Communications SPAWC 2007*. Helsinki (Finland), 2007, p. 1 - 5.

[8] SADOUGH, S. M. S., DUHAMEL, P. Improved iterative detection and achieved throughputs of OFDM systems under imperfect channel estimation. *IEEE Transactions on Wireless Communications*, 2008, vol. 7, no. 12., p. 5039 - 5050.

[9] SADOUGH, S. M. S., KHALIGHI, M. A., DUHAMEL, P. Improved iterative MIMO signal detection accounting for channel-estimation errors. *IEEE Transactions on Vehicular Technology*, 2009, vol. 58, no. 7, p. 3154 - 3167.

[10] SADOUGH, S. M. S. Robust iterative receiver design under imperfect channel estimation. In *Proceedings of 6th IEEE Conference on Wireless Advanced WiAD 2010*. London (UK), 2010, p. 1 - 5.

[11] SADOUGH, S., PIANTANIDA, P., DUHAMEL, P. MIMO-OFDM optimal decoding and achievable information rates under imperfect channel estimation. In *8th IEEE Workshop on Signal Processing Advances in Wireless Communications SPAWC 2007*. Helsinki (Finland), 2007, p. 1 - 5.

[12] PRASAD, R. *OFDM for Wireless Communications Systems*. Boston: Artech House, 2004.

[13] MCLACHLAN, G. J., KRISHNAN, T. *The EM Algorithm and Extensions*. New York: Wiley, 1997.

[14] KULLBACK, S. *Information Theory and Statistics* New York: Wiley, 1959.

[15] BAHL, L., COCKE, J., JELINEK, F., RAVIV, J. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Transactions on Information Theory*, 1974, vol. 20, no. 2, p. 284 - 287.

[16] SCHARF, L. *Statistical Signal Processing*. Reading (MA, USA): Addison and Wesley, 1991.

About Authors ...

Seyed Mohammad-Sajad SADOUGH was born in Paris in 1979. He received his B.Sc. degree in Electrical Engineering (electronics) from Shahid Beheshti University, Tehran, I.R. Iran in 2002 and the M.Sc. and his Ph.D. degrees in Electrical Engineering (telecommunication) from Paris-Sud 11 University, Orsay, France, in 2004 and 2008, respectively. From 2004 to 2007, he held a joint appointment with the National Engineering School in Advanced Techniques (ENSTA), Paris, France, and the Laboratory of Signals and Systems (LSS), at Supélec, Gif-sur-Yvette, France. He was a lecturer in the Department of Electronics and Computer Engineering (UEI), ENSTA, where his research activities were focused on improved reception schemes for ultra-wideband communication systems. From December 2007 to September 2008, he was a postdoctoral researcher with the LSS, Supélec-CNRS, where he was involved in the European research project TVMSL with Alcatel-Lucent France. Since October 2008, he has been a member of the Faculty of Electrical & Computer Engineering, Shahid Beheshti University, where he is currently an Assistant Professor in the Department of Telecommunication. Dr. Sadough's areas of research include signal processing, communication theory, and digital communication.

Mohammad MODARRESI received the B.Sc. degree in Electrical Engineering (electronics) from Shiraz University of Technology, Shiraz, Iran, in 2007. He is currently pursuing the M.Sc. degree in Electrical Engineering (telecommunications) at Shahid Beheshti University, Tehran, Iran. His research interests include turbo codes and iterative decoding, joint channel estimation/decoding and multicarrier systems.