ACCELERATED LINEAR EM-MAP ALGORITHM FOR OFDM CHANNEL ESTIMATION

Florence Alberge

Laboratoire des Signaux et Systèmes, Supélec,
Plateau de Moulon, 91190 Gif-sur-Yvette,
FRANCE
email : alberge@lss.supelec.fr

ABSTRACT

OFDM systems traditionally perform channel estimation relying on known training sequences. However in wireless systems, performance and mobility can be further enhanced by operating semi-blind channel estimation refinement between reference symbols. A channel tracking method based on a maximum a posteriori (MAP) version of the Expectation-Maximization (EM) algorithm and on a block representation of the channel variations has been proposed. It performs better than already existing recursive algorithms with the additional advantage of a linear arithmetic complexity. In this paper, we study the matrix governing the behaviour of this method. This leads to a practical rule for choosing optimally (with respect to the convergence rate) the unique parameter of the method. The proposed iterative procedure is applicable to both single carrier and OFDM systems. Simulations are presented in the context of 5 GHz WLANs, showing the practical interest of the method.

***Index Terms*—** Communication systems, Convergence of numerical methods

1. INTRODUCTION

Orthogonally Frequency Division Multiplexing (OFDM) has already been accepted for Wireless Local Area Network (WLAN) standards (IEEE 802.11a), High Performance Local Area Network Type 2 (HYPERLAN 2) and Japan's Mobile Multimedia Access Communication (MMAC) systems. In OFDM systems, the effect of the channel appear in the frequency domain as a simple scalar multiplication. Classical methods for estimating these coefficients are based on training sequences. In order to cope with Doppler effect due to the mobility of wireless systems, reference sequences must be repeated quite often resulting in a significant loss in the useful bit rate. An alternative is to estimate the channel based only on the observed data. A common way to design blind estimation algorithms is to use the Expectation Maximization (EM) algorithm. For likelihood functions with multiple maxima, the convergence point depends on the initial starting point and may be a local maximum. A semi-blind estimation method consist in tracking the channel variations by refining the channel coefficients blindly using the training sequence as an initialization for the estimator, then local convergence problems are avoided. EM-based blind or semi-blind channel estimation methods have already been proposed in the OFDM context [1],[2],[3] including time and frequency correlations or not. We work within the framework of slowly time-varying channels so that inter-channel interference (ICI) can be neglected. In this case, the channel can be considered constant during the transmission of a block of symbols and the size of the block is related to the mobility. Based on this idea, a maximum a posteriori (MAP) algorithm which takes into account a channel correlation model between the channel coefficients of successive blocks has been proposed in [4]. The application of a One Step Late (OSL) technique first presented in [5] permits to turn the arithmetic complexity from quadratic to linear [6]. The convergence of this new algorithm in our context is established and the stationary points are also stationary points of the EM-MAP [6]. This paper is the last step toward an efficient algorithm (EM-MAP) for channel estimation with linear arithmetic complexity and enhanced convergence rate. Indeed, we give here a closed-form expression for the matrix governing the behaviour of the OSL algorithm and we explain how to achieve the best convergence rate.

2. SYSTEM MODEL

2.1. Transmitter/Receiver structure

We consider a conventional cyclic prefix OFDM transceiver scheme. In this model, some side entries of the size P IFFT are zero and only Nc among the P available sub-carriers are effectively used for transmission of information data. The block of data $\mathbf{x} = [x_0, ..., x_{Nc-1}]^T$ is modulated in the time domain by IFFT. The channel is modelled by linear filtering. Some redundancy is introduced into the transmitted signal by cyclic prefix extension so that the overlapping introduced by the channel memory $[h'_0, ..., h'_{L-1}]^T$ corresponds to that of a circular convolution between $\mathbf{x}$ and the channel. Consequently, the channel is viewed in the frequency domain, after demodulation by the FFT, as parallel flat fading channels. Let $\mathbf{h}' = [h'_0, ..., h'_P - 1]^T$ denote the OFDM channel, the received signal $\mathbf{y} = [y_1, ..., y_{Nc}]^T$ can be modelled by the following equation:

$$\mathbf{y} = Diag(\mathbf{x})\mathbf{H}' + \mathbf{e} \tag{1}$$

where $\mathbf{H}' = SF\mathbf{h}'$, F is the $P \times P$ Fourier matrix, S is the $Nc \times P$ matrix selecting the Nc information sub-carriers, $S = [\mathbf{0}_{\mathbf{Nc},\frac{\mathbf{P-Nc}}{\mathbf{2}}} \;\; \mathbf{I}_{\mathbf{Nc}} \;\; \mathbf{0}_{\mathbf{Nc},\frac{\mathbf{P-Nc}}{\mathbf{2}}}]$, and $\mathbf{e}$ is a centered gaussian noise vector with variance σ_e^2
Note that $h'_i = 0$ for $i \geq L$, where L denote the cyclic prefix length. The OFDM systems are designed such that $L < P$ (in IEEE802.11a $L = P/4$). The taps are assumed independent and Rayleigh distributed.

2.2. Channel Model

The channel is considered **piecewise constant** over a block of length T (slow time-varying channel). Each channel coefficient is then computed from T observation symbols which improves the quality of the MAP estimator. It appears that $T = 20$ is a good compromise for Doppler speeds lower than $3m/s$ and $T = 10$ is a good choice for Doppler speeds greater than $3m/s$ [7]. Let H_k denote the channel coefficient of block k (on a given subcarrier), H_k is supposed constant over a block of length T. The time variations can be modelled using an autoregressive (AR) model of order 1, namely $H_k = \alpha H_{k-1} + \varepsilon_k$, where α is the time correlation coefficient and ε_k is an additive white gaussian noise of mean 0 and variance σ_ε^2.The parameter σ_ε^2 is unknown and will be estimated by the proposed procedure. On the contrary, α is assumed to be known since α can be perfectly determined thanks to Jakes' model [8]. Using matrix notations, we get for a given sub-carrier

$$\mathbf{y} = \mathbf{X}\tilde{\mathbf{H}} + \mathbf{e} \quad (2)$$

$$\mathbf{M}\mathbf{H} = \varepsilon + \mathbf{b} \quad (3)$$

where $\mathbf{b}$ is the contribution of the previous frame and $\mathbf{M}$ is a $q \times q$ ($q = N_c/T$) matrix given by

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\alpha & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\alpha & 1 \end{pmatrix}$$

We suppose that the hidden variables x_i, unknown at the receiver, are uniformly distributed and take values on an M-size constellation $\{s_1, ..., s_M\}$. Thus, the variable $\mathbf{X}$ takes one of the matrix values $\mathbf{S}_{\underline{\mathbf{m}}} = Diag(s_{m_1}, ..., s_{m_N})$ indexed by $\underline{\mathbf{m}} = [m_1, ..., m_N]^T \in \{1, ..., M\}^N$.
Finally, if $\mathbf{H} = [H_0, ..., H_{q-1}]^T$ denotes the vector of the channel coefficients to be estimated, then we define $\tilde{\mathbf{H}} = \mathbf{B}\mathbf{H}$, where $\mathbf{B}$ is the $N \times q$ matrix obtained from the tensorial product $\mathbf{I}_q \times [1, ..., 1]^T$. The joint probability of the observation and symbols reads

$$P(\mathbf{y}, \mathbf{S}_{\underline{m}} \mid \mathbf{H}, \frac{1}{\sigma_e^2}) = P(\mathbf{S}_{\underline{m}}) \frac{1}{\sigma_e^{2N}} exp(-\frac{1}{\sigma_e^2} \|\mathbf{y} - \mathbf{S}_{\underline{m}}\mathbf{B}\mathbf{H}\|^2) \quad (4)$$

The prior probability modelling the variation of the channel reads

$$P(\mathbf{H}, \sigma_\varepsilon^2) \propto \frac{1}{\sigma_\varepsilon^{2q}} exp(-\frac{1}{\sigma_\varepsilon^2}(\mathbf{H} - \mathbf{M}^{-1}\mathbf{b})^* \mathbf{C}^{-1}(\mathbf{H} - \mathbf{M}^{-1}\mathbf{b})) \quad (5)$$

where $\sigma_\varepsilon^2 \mathbf{C}$ is the covariance matrix of $\mathbf{H}$ and $\mathbf{C}^{-1} = \mathbf{M}^H\mathbf{M}$.

3. EM-MAP ALGORITHM WITH LINEAR COMPLEXITY

3.1. Review of the EM-MAP algorithm

In this section, we shall briefly review the EM-MAP algorithm. Classically, the corresponding auxiliary function reads [4]: $Q(\mathbf{H}, \sigma_e^2, \sigma_\varepsilon^2, \mathbf{H}^{(i)}, \sigma_e^{2(i)}, \sigma_\varepsilon^{2(i)}) = \sum_{\underline{m}} P(\mathbf{S}_{\underline{m}}|\mathbf{y}, \mathbf{H}^{(i)}, \sigma_e^{2(i)})[logP(\mathbf{y}|\mathbf{S}_{\underline{m}}, \mathbf{H}; \sigma_e^2) + logP(\mathbf{H}, \sigma_\varepsilon^2) + cst]$. Using (4) and (5) and setting to zero the derivative of $Q(\mathbf{H}, \sigma_e^2, \sigma_\varepsilon^2, \mathbf{H}^{(i)}, \sigma_e^{2(i)}, \sigma_\varepsilon^{2(i)})$ with respect to $\mathbf{H}^*$, we have

$$[\frac{1}{\sigma_e^{2(i)}}\frac{\mathbf{D}}{\Gamma} + \frac{1}{\sigma_\varepsilon^{2(i)}}\mathbf{C}^{-1}]\mathbf{H}^{(i+1)} = \frac{1}{\sigma_e^{2(i)}}\frac{\mathbf{V}}{\Gamma} + \frac{1}{\sigma_\varepsilon^{2(i)}}\mathbf{M}^*\mathbf{b} \quad (6)$$

where the matrix $\mathbf{D}/\Gamma$ is a diagonal matrix with entries

$$\frac{D_j}{\Gamma} = \sum_{l=jT}^{(j+1)T-1} \frac{\sum_{m=1}^M \|s_m\|^2 P(s_m|y_l, \mathbf{H}^{(i)}, \sigma_e^{2(i)})}{\sum_{m=1}^M P(s_m|y_l, \mathbf{H}^{(i)}, \sigma_e^{2(i)})} \quad (7)$$

and where the j^{th} tap of the vector $\frac{\mathbf{V}}{\Gamma}$ reads:

$$\frac{V_j}{\Gamma} = \sum_{l=jT}^{(j+1)T-1} \frac{\sum_{m=1}^M s_m^* y_l^2 P(s_m|y_l, \mathbf{H}^{(i)}, \sigma_e^{2(i)})}{\sum_{m=1}^M P(s_m|y_l, \mathbf{H}^{(i)}, \sigma_e^{2(i)})} \quad (8)$$

The current channel estimates are obtained by solving the linear system (6) by a Gauss method. This step has a quadratic arithmetic complexity.

By similar calculations, we obtain the update formulae of the channel noise variance σ_ε^2 and of the noise variance σ_e^2which have a linear arithmetic complexity [4].

3.2. A Ones Step Late EM-MAP algorithm

We proposed in [6], a simplified EM-MAP algorithm exhibiting a linear complexity. The simplification is based on the One Step Late technique first presented by Green in [5]. This method consists in computing the derivatives of the prior probability at the current value of the parameters rather than at the new value. An intuitive justification of this procedure is that if the algorithm converges slowly, the derivative computed at iteration i and at iteration $i+1$ will not be much different [5]. Moreover, it can straightforwardly be seen that this method has exactly the same fixed points than the original algorithm. In the update equation (6), the quadratic complexity comes from the non-diagonal matrix $\mathbf{C}^{-1}$. Following Green, we apply the OSL procedure on the part of the gradient involving $\mathbf{C}^{-1}$ to obtain a less computationally demanding algorithm. After some calculations, we obtain the new update equation for the channel coefficients

$$\begin{aligned}[\frac{1}{\sigma_{e(i)}^2}\frac{\mathbf{D}}{\Gamma} + \frac{\beta}{\sigma_\varepsilon^{2(i)}}\mathbf{I}]\mathbf{H}^{(i+1)} &= \tfrac{1}{\sigma_e^{2(i)}}\tfrac{\mathbf{V}}{\Gamma} + \tfrac{1}{\sigma_\varepsilon^{2(i)}}M^*\mathbf{b} \\ &\quad - \tfrac{1}{\sigma_\varepsilon^{2(i)}}\mathbf{C}^{-1}\mathbf{H}^{(i)} + \tfrac{\beta}{\sigma_\varepsilon^{2(i)}}\mathbf{H}^{(i)} \quad (9)\end{aligned}$$

where $\mathbf{I}$ denotes the identity matrix and β is a positive number chosen to ensure the convergence of the iterative process (see below). The update equations for the variances σ_e^2 and σ_ε^2 remain unchanged. Clearly, both algorithms have the same fixed points and the arithmetic complexity is linear.
Regarding the convergence of the method, we proved in [6] that the proposed OSL algorithm is a particular case of the proximal point algorithm (PPA). A generalized PPA is defined by the iterative process [9]

$$\hat{\Theta}^{(i+1)} = \arg\max_{\Theta}\{\epsilon(\Theta) - \gamma_i d(\Theta, \widehat{\Theta}^{(i)})\} \quad (10)$$

where γ_i is a sequence of positive number and $d(\Theta, \Theta^{(i)})$ is a penalty function which verifies

$$d(\Theta, \widehat{\Theta}^{(i)}) \geq 0 \quad \text{and} \quad d(\Theta, \widehat{\Theta}^{(i)}) = 0 \quad \text{iff } \Theta = \widehat{\Theta}^{(i)} \quad (11)$$

Then $\{\epsilon(\Theta^{(i)}), i = 0, 1, 2...\}$ is a nondecreasing sequence [9].
It turns out that the proposed OSL algorithm is a PPA with $\Theta = (\mathbf{H}, \sigma_e^2, \sigma_\varepsilon^2)$, $\epsilon(\Theta) = log\ p(y|\Theta) + log(\Theta$ (*ie* $\epsilon(\Theta)$ is the a posteriori log-probability), $\gamma_i = 1$ and $d(\Theta, \widehat{\Theta}^{(i)}) = \Big[p(x|y, \Theta^{(i)}) \parallel p(x|y, \Theta)\Big] + \frac{1}{\sigma_\varepsilon^{2(i)}}(\mathbf{H} - \mathbf{H}^{(i)})^*(\beta\mathbf{I} - \mathbf{C}^{-1})(\mathbf{H} - \mathbf{H}^{(i)})$. Since the first term of $d(\Theta, \widehat{\Theta}^{(i)})$ is a Kullback-Liebler divergence, a sufficient condition for (11) to be met is $\beta > \lambda_{max}(\mathbf{C}^{-1})$. It is well-known that $\lambda_{max}(\mathbf{C}^{-1}) > 1 + \alpha^2 + 2\alpha\cos(\frac{\pi}{q+1})$. Then a sufficient condition of convergence is $\beta \geq 1 + \alpha^2 + 2\alpha\cos(\frac{\pi}{q+1})$. Note that the positivity of $d(\Theta, \widehat{\Theta}^{(i)})$ is guaranteed because each term is a positive number. However, it may

exists lower values of β such that $d(\Theta, \widehat{\Theta}^{(i)})$ is a positive number even if the second term has a negative value.
At this point, we have proved that the proposed iterative procedure: *(i)* converges and increases the value of the a posteriori probability with the iterations, *(ii)* exhibits the same stationary point than the EM-MAP, *(iii)* has a linear arithmetic complexity.
The scope of this paper is to derive an accelerated procedure with the three nice properties listed above. This can be done by finding the parameter β which optimises the convergence speed. In the next section, we focus on the spectral radius of the matrix governing the convergence of the proposed OSL algorithm.

4. CONVERGENCE STUDY

4.1. Matrix $\mathbf{R}_\beta$ governing the convergence

Following [5], we obtain the matrix $\mathbf{R}_\beta$ governing the convergence behaviour of the iterative procedure in (9) near to the fixed point

$$\mathbf{R}_\beta(\mathbf{H}, \sigma_{\mathbf{e}}^{\mathbf{2}}, \sigma_\varepsilon^{\mathbf{2}}) = (\mathbf{B} + \mathbf{E} + \frac{\beta}{\sigma_\varepsilon^2}\mathbf{I})^{-1}(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta\mathbf{I}}{\sigma_\varepsilon^2}) \quad (12)$$

where $\mathbf{B} + \mathbf{E}$ is a diagonal matrix with entries

$$(\mathbf{B} + \mathbf{E})_j = \frac{1}{\sigma_e^2}\sum_{l=jT}^{(j+1)T-1}\sum_{m=1}^{M}\|s_m\|^2 P(s_m|y_l, \mathbf{H}, \sigma_e^2)$$

It turns out that $\mathbf{E}$ is also a diagonal matrix with entries :

$$E_j = \frac{1}{\sigma_e^2}\sum_{l=jT}^{(j+1)T-1}(\sum_{m=1}^{M}\|s_m\|^2\|y_l - s_m\mathbf{H}_j\|^2 P(s_m|y_l, \mathbf{H}, \sigma_e^2) - \|\sum_{m=1}^{M} s_m^*(y_l - s_m\mathbf{H}_l)P(s_m|y_l, \mathbf{H}, \sigma_e^2)\|^2)$$

Note that $(\mathbf{B} + \mathbf{E})$, $\mathbf{E}$ and $\frac{\beta}{\sigma_\varepsilon^2}\mathbf{I}$ are both diagonal matrices with non-negative entries. (The proofs are straightforward and are not reported here. They are based on the independence of the emitted symbols).

4.2. Evolution of the Spectral radius of $\mathbf{R}_\beta$ with β

The spectral radius of a matrix $\mathbf{A}$ is the positive number $\rho(\mathbf{A})$ defined as $\rho(\mathbf{A}) = max\{|\lambda_i(\mathbf{A})|\ ;\ 1 \le i \le q\}$ where $\lambda_i(\mathbf{A})$ stands for the i^{th} eigenvalue of $\mathbf{A}$ [10]. It is well-known that an iterative procedure is convergent if and only if the spectral radius of the matrix governing its behaviour is lower than 1. Moreover, small spectral radius leads to fast convergence. We then investigate the behaviour of $\rho(\mathbf{R}_\beta)$ with β for the matrix $\mathbf{R}_\beta$ given by (12).

Proposition 1 *Suppose that* $\mathbf{R}_\beta = (\mathbf{B} + \mathbf{E} + \frac{\beta}{\sigma_\varepsilon^2}\mathbf{I})^{-1}(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta\mathbf{I}}{\sigma_\varepsilon^2})$ *where* **B+E**, $\mathbf{E}$ *and* $\mathbf{C}^{-1}$ *are symmetric and non-negative definite matrices of the same order,* $\beta > 0$ *and* $\rho(\mathbf{R}_\beta) < 1$. *Then:*

$$\rho(\mathbf{R}_{\beta_1}) \le \rho(\mathbf{R}_{\beta_2})\ \ iff\ \ 0 < \beta_1 \le \beta_2$$

Proof. Let λ denote any eigenvalue of $\mathbf{R}_{\beta_1}$ then λ is solution of $det(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta_1\mathbf{I}}{\sigma_\varepsilon^2} - \lambda(\mathbf{B} + \mathbf{E} + \frac{\beta_1}{\sigma_\varepsilon^2}\mathbf{I})) = 0$. Introduce now β_2 in the expression $det((\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta_2\mathbf{I}}{\sigma_\varepsilon^2}) + \frac{(\beta_1-\beta_2)\mathbf{I}}{\sigma_\varepsilon^2} - \lambda(\mathbf{B} + \mathbf{E} + \frac{\beta_2}{\sigma_\varepsilon^2}\mathbf{I} + \frac{(\beta_1-\beta_2)\mathbf{I}}{\sigma_\varepsilon^2}) = 0$. Write $\mathbf{B} + \mathbf{E} + \frac{\beta_2}{\sigma_\varepsilon^2}\mathbf{I} = \mathbf{L}\mathbf{L}^T$ where $\mathbf{L}$ is square and non singular (Cholesky factorization [10]), then $det(\mathbf{L}^{-1}(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta_2\mathbf{I}}{\sigma_\varepsilon^2})\mathbf{L}^{-T} - \lambda\mathbf{I} + (1-\lambda)\frac{(\beta_1-\beta_2)}{\sigma_\varepsilon^2}\mathbf{L}^{-1}\mathbf{L}^{-T}) = 0$. So, there exists a non-zero vector $\mathbf{u}$ with $\mathbf{u}^T(\mathbf{L}^{-1}(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta_2\mathbf{I}}{\sigma_\varepsilon^2})\mathbf{L}^{-T})\mathbf{u} - \lambda\mathbf{u}^T\mathbf{u} = (1-\lambda)\frac{(\beta_2-\beta_1)}{\sigma_\varepsilon^2}\mathbf{u}^T\mathbf{L}^{-1}\mathbf{L}^{-T}\mathbf{u}$. If λ were greater than $\rho(\mathbf{R}_{\beta_2}) = \mathbf{L}^{-1}(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta_2\mathbf{I}}{\sigma_\varepsilon^2})\mathbf{L}^{-T}$ we would have $(1-\lambda)\frac{(\beta_2-\beta_1)}{\sigma_\varepsilon^2}\mathbf{u}^T\mathbf{L}^{-1}\mathbf{L}^{-T}\mathbf{u} < 0$ which is impossible since $1 - \lambda > 0$ and $\frac{(\beta_2-\beta_1)}{\sigma_\varepsilon^2}\mathbf{L}^{-1}\mathbf{L}^{-T}$ is non-negative definite. □
The matrix $\mathbf{R}_\beta$ defined in 4.1 satisfies the assumptions of Proposition 1. Then, if the algorithm converges (*ie* $\rho(\mathbf{R}_\beta) < 1$), the best choice for β in terms of convergence speed is the smallest value. At this point, we do not know the range of values for β such that $\rho(\mathbf{R}_\beta) < 1$. This is the scope of the next section.

4.3. Authorized range of values for β

The following proposition gives a necessary condition on β such that $\rho(\mathbf{R}_\beta) < 1$.

Proposition 2 *Suppose that* $\mathbf{R}_\beta = (\mathbf{B} + \mathbf{E} + \frac{\beta}{\sigma_\varepsilon^2}\mathbf{I})^{-1}(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta\mathbf{I}}{\sigma_\varepsilon^2})$ *where* $\beta > 0$, $\mathbf{C}^{-1}$ *is a symmetric and positive definite matrix and* **B+E** *and* $\mathbf{E}$ *are diagonal matrices with non-negative entries. Then,* $\rho(\mathbf{R}_\beta) < 1$ *if* $2\mathbf{E} + \mathbf{B} + \frac{2\beta\mathbf{I} - \mathbf{C}^{-1}}{\sigma_\varepsilon^2}$ *is positive definite.*

Proof. Let λ denote any eigenvalue of $\mathbf{R}_\beta$ then, there exists a non-zero vector $\mathbf{u}$ with $\mathbf{u}^T(\mathbf{E} - \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2} + \frac{\beta\mathbf{I}}{\sigma_\varepsilon^2})\mathbf{u} = \lambda\mathbf{u}^T(\mathbf{B} + \mathbf{E} + \frac{\beta}{\sigma_\varepsilon^2}\mathbf{I})\mathbf{u}$.
If λ could be greater than 1, it would exist a non-zero vector $\mathbf{u}$ with $\mathbf{u}^T(\mathbf{B} + \frac{\mathbf{C}^{-1}}{\sigma_\varepsilon^2})\mathbf{u} < 0$ which is impossible since in that case the EM-MAP of section 3.1 could diverge [6]
If λ could be smaller than -1, it would exist λ and a non-zero vector $\mathbf{u}$ with $\mathbf{u}^T(2\mathbf{E} + \mathbf{B} + \frac{2\beta\mathbf{I} - \mathbf{C}^{-1}}{\sigma_\varepsilon^2})\mathbf{u} < 0$ which is impossible provided that $2\mathbf{E} + \mathbf{B} + \frac{2\beta\mathbf{I} - \mathbf{C}^{-1}}{\sigma_\varepsilon^2}$ is positive definite. □
In consequence of this proposition, we see that the proposed OSL algorithm can be accelerated by choosing β such that $\beta > \frac{1}{2}\lambda_{max}(\mathbf{C}^{-1} - \sigma_\varepsilon^2(\mathbf{B} + 2\mathbf{E}))$. Remember that $\mathbf{B}$ and $\mathbf{E}$ are diagonal matrices whereas $\mathbf{C}^{-1}$ is tridiagonal meaning that we need to compute the highest eigenvalue of a tridiagonal matrix. This can be done using Givens (or bisection) method [10]. However, the matrices $\mathbf{E}$ and $\mathbf{B}$ are function of $\mathbf{H}$ and σ_e^2. Then β should be computed at each iteration. To avoid this extra computational cost, the lower bound on β is released in order to obtain a new bound independent of $\mathbf{H}$ and σ_e^2.
Since **B+E** and $\mathbf{E}$ are diagonal matrices with non-negative entries then 2**E+B** is also a diagonal matrix with non-negative entries. Then $\lambda_{max}(\mathbf{C}^{-1} - \sigma_\varepsilon^2(\mathbf{B} + 2\mathbf{E})) \le \lambda_{max}(\mathbf{C}^{-1})$. Finally, the optimal value of β in terms of computational cost and convergence rate is:

$$\beta = \frac{1 + \alpha^2 + 2\alpha\cos(\frac{\pi}{q+1})}{2}$$

We will see in the next section that choosing $\beta = \beta_{low} = \frac{1+\alpha^2+2\alpha\cos(\frac{\pi}{q+1})}{2}$ rather than the canonical choice $\beta = \beta_{high} = 1 + \alpha^2 + 2\alpha\cos(\frac{\pi}{q+1})$ leads to a significant gain in the number of iterations.

5. SIMULATIONS

In this section, we illustrate the behaviour of the proposed algorithm in the specific context of HIPERLAN/2 broadband wireless communication standard, which is similar to IEEE802.11a and MMAC.

EM-MAP	OSL (β_{high})	OSL (β_{low})
3.6057e-2	3,046e-2	3.5898e-2

Table 1. BER comparison - SNR=10dB - Doppler Speed: 3m/s

HIPERLAN/2. The cyclic prefix is 16 samples long and the number of carriers is $N_c = 64$. A rate $R = 1/2$, constraint length $l = 7$ Convolutional Code (CC) (171/133) is used before bit interleaving followed by 16-QAM mapping. Only 48 carriers are effectively used. Monte Carlo simulations are run and averaged over 5000 realizations of a BRAN C frequency selective channel in order to obtain BER curves. A classical Jake's Doppler spectrum and Rayleigh fading statistics are assumed for all taps. Speeds are supposed to be known from the receiver. Each frame processed contains 2 known training symbols, followed by 100 OFDM data symbols. The bit probabilities $P(b_l^m)$ estimates are performed in the E-step by two iterations of the turbo demodulation process. The estimation process is repeated until the mean square errors of the channel coefficients matrix and the channel noise variance are lower than 10^{-8} simultaneously. We compare the performance in term of channel estimation between the EM-Block algorithm which do not use any prior information, the EM-MAP algorithm and the OSL algorithm (the proposed one) with β set to β_{low} and β_{high}.
The OSL algorithm is expected to have the same stationary points than the EM-MAP algorithm for any β (as long as the OSL algorithm converges). This is confirmed by the results in Table 1 where we show the BER for a SNR of 10dB and a Doppler speed of 3m/s. Figure 1 depicts the mean square error as a function of SNR. We can see that the OSL algorithm performs better than the EM-Block algorithm and has the same performance than the initial EM-MAP algorithm.

This simulation confirm that the OSL algorithm has the same

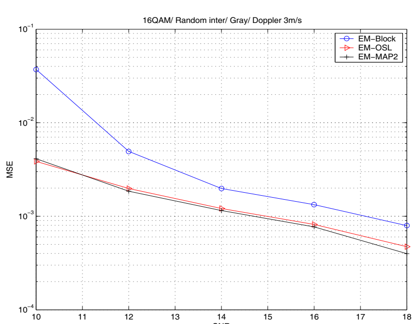

Fig. 1. MSE vs SNR - Comparison of the EM-Block, EM-MAP and OSL algorithms - Doppler speed: 3m/s

behaviour in term of estimation accuracy than the EM-MAP with the additional advantage of a linear computational complexity. In this paper, we have proved that the OSL algorithm converges more quickly with β_{low} than with any other greater value. The experiment described below quantify this gain. Let $N_{\beta_{low}}$ and $N_{\beta_{high}}$ denote the number of iterations until convergence ($MSE \leq 10^{-8}$) for the OSL algorithm respectively with β_{low} and β_{high}. The quantity I_{gain} is defined as $I_{gain} = \frac{N_{\beta_{high}} - N_{\beta_{low}}}{N_{\beta_{high}}}$. Figure 2 depicts I_{gain} versus SNR for $\alpha = 0.99$ and $\alpha = 0.997$. We observe in fig 2 that we can save up to 35% of the iterations thanks to an optimal choice of the parameter.

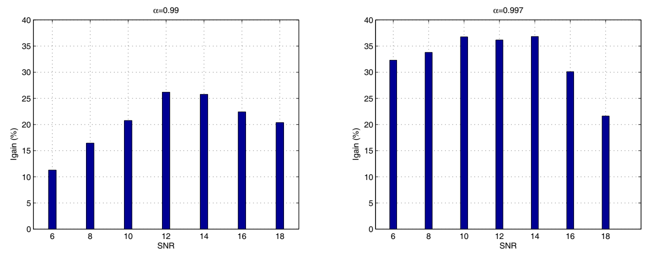

Fig. 2. Number of iterations saved with $\beta = \beta_{low}$ compared to $\beta = \beta_{high}$, $\alpha = 0.99$ (right), $\alpha = 0.997$ (left).

6. CONCLUSION

In this paper, we have studied the convergence behaviour of an OSL algorithm. We have shown that the convergence speed of this method could be enhanced by a judicious choice of the main parameter of the algorithm. This can lead to a saving of the iterations close to 35% compared with the canonical choice. The proposed method presents the following nice properties: an efficient channel estimation method, a low arithmetic complexity and an enhanced convergence speed.

7. REFERENCES

[1] E. Jaffrot and M. Siala. Turbo channel estimation for OFDM systems on higly time and frequency selective channels. In *ICASSP Proc, Istanbul, Turkey*, June 2000.

[2] X. Zhuang and F.W. Vook. Iterative channel estimation and decoding for a turbo-coded OFDM system via the EM algorithm. In *ICASSP Proc.*, pages 2337–2340, 2002.

[3] T. Y. Al-Naffouri, A. Bahai, and A. Paulraj. Semi-blind channel identification and equalization in OFDM: An Expectation-Maximization approach. In *VTC Proc., Vancouver, Canada*, 2002.

[4] J-M. Mamfoumbi-Ocloo, F. Alberge, and P. Duhamel. Semi-blind channel estimation for OFDM systems via an EM-MAP algorithm. In *SPAWC Proc., New York City, USA*, pages 605–609, 2005.

[5] P.J. Green. On the use of the EM algorithm for penalized likelihood estimation. *J.R. Statist. Soc. B*, pages 443–452, 1990.

[6] J-M. Mamfoumbi-Ocloo and F. Alberge. OFDM channel estimation by a linear EM-MAP algorithm. In *ICASSP Proc., Toulouse, France*, 2006.

[7] J-M. Mamfoumbi-Ocloo. *Estimation de canaux de transmission respectant un modèle d'évolution temps/fréquence*. Ph.D. dissertation, Université Paris-Sud XI, France, 2006.

[8] J. Proakis. *Digital Communications*. Mc Graw-Hill, 1989.

[9] S. Chrétien and A. O. Hero. Kullback proximal algorithms for maximum likelihood estimation. *IEEE Trans. on Inf. Theory*, 46(5):1800–1810, 2000.

[10] P.G. Ciarlet. *Introduction to numerical linear algebra and optimization*. Cambridge University Press, 1989.