SEMI-BLIND CHANNEL IDENTIFICATION AND EQUALIZATION IN OFDM: AN EXPECTATION-MAXIMIZATION APPROACH

Tareq Y. Al-Naffouri[1], Ahmad Bahai[1,2], *and* Arogyaswami Paulraj[1]

[1] Electrical Engineering Department, Stanford University, CA 94305
[2] National Semiconductor Santa Clara, CA 95052

ABSTRACT

This work proposes an expectation-maximization approach to channel identification and equalization in OFDM. The algorithm exploits the natural constraints imposed by the channel (sparsity, maximum delay spread, and a priori statistical information) and those imposed by the transmitter (pilots, cyclic prefix, and the finite alphabet constraint). These constraints are used to reduce the number of pilots needed for channel and data recovery and also to perform this task within one packet.

1. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) is an effective technique for high bit rate transmission. It has found widespread applications and is already part of many standards including digital audio and video broadcasting in Europe and high speed transmission over digital subscriber line (DSL) in the United States. It has been proposed recently for local area mobile wireless standards including IEEE 802.11a and HIPERLAN/2. [1]

For proper operation, the OFDM receiver needs accurate channel state information. The receiver can go around this by employing differential modulation at the cost of a 3 to 4 dB degradation in SNR. Otherwise, the receiver needs to jointly recover the input and channel information. For rapidly time varying channels, the receiver faces the additional challenge of performing the recovery within the same packet. Many techniques have been proposed in literature to achieve this (see, e.g., [1] and the references therein). Thus, pilots were employed in [2] and [3] to perform channel estimation. The redundancy due to the presence of the cyclic prefix (CP) was utilized in [4] and [5] to perform blind channel identification and in [6] for channel tracking. The a priori information of time and frequency correlation was used in [7] and [8] to estimate the channel's frequency response. At a higher level of abstraction, each of these methods utilizes one or more constraints on the input or channel to perform channel and/or data recovery. None of them, however, makes a *collective* use of the channel and data constraints so to as to improve the quality of the channel estimate and/or to reduce the overhead necessary to achieve this task.

The work of T. Y. Al-Naffouri was partially supported by a fellowship from King Fahd University of Petroleum and Minerals, Saudi Arabia.

In this paper, we propose an expectation-maximization (EM) approach for semi-blind channel identification and equalization. Specifically, we exploit the natural channel constraints and those imposed by the transmitter to perform channel and data recovery within the same packet and to reduce the number of pilots that are eventually needed. These constraints include

- Maximum delay spread and sparsity constraints on the channel, which mean that there are only a few active taps with known location
- A priori channel statistics like the channel mean and covariance
- Redundancy of the input in the form of cyclic prefix (redundancy due to the presence of a real code is considered in [9])
- The finite alphabet constraint on the data
- The artificial constraint of pilots whose number we are able to reduce by building upon the aforementioned more natural constraints

As we shall soon see, the algorithm can make use of these constraints collectively and the channel identification step always boils down to a (regularized) least-squares LS problem. While we don't pursue that here, this least-squares formulation can be easily generalized to incorporate time correlation information.

1.1. Notation

We denote scalars with small-case letters (e.g. x) , vectors with small-case boldface letters (e.g. $\boldsymbol{x}$), and matrices with uppercase boldface letters (e.g. $\boldsymbol{X}$). Caligraphic notation (e.g. $\boldsymbol{\mathcal{X}}$) is reserved for vectors in the frequency domain. The dependence on time is indicated by a time index i that appears as a subscript (e.g. $x_i, \boldsymbol{x}_i, \boldsymbol{X}_i, \boldsymbol{\mathcal{X}}_i$).

Now consider a length-N vector $\boldsymbol{x}_i$. We deal with three derivatives associated with this vector. The first two are obtained by partitioning $\boldsymbol{x}_i$ into an upper (prefix) vector $\underline{\boldsymbol{x}}_i$ and a lower (usually longer) vector $\tilde{\boldsymbol{x}}_i$. The third derivative, $\overline{\boldsymbol{x}}_i$, is created by concatenating $\boldsymbol{x}_i$ with a copy of its prefix $\underline{\boldsymbol{x}}_i$. The relation among $\boldsymbol{x}_i$ and its derivatives is summarized by

$$\overline{\boldsymbol{x}}_i = \begin{bmatrix} \boldsymbol{x}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} \tag{1}$$

This notational convention will be extended to matrices as well. Thus, a matrix $\boldsymbol{Q}$ with N rows can be partitioned as

$$\boldsymbol{Q} = \begin{bmatrix} \underline{\boldsymbol{Q}} \\ \tilde{\boldsymbol{Q}} \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{Q}}_P \\ \tilde{\boldsymbol{Q}}_{N-P} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q}_{\underline{I}} \\ \boldsymbol{Q}_{\tilde{I}} \end{bmatrix} \tag{2}$$

The subscripts stand for the number of rows in the submatrix (when these rows are understood) or for the index set of the rows of the mother matrix that belong to the submatrix.

2. ESSENTIAL ELEMENTS OF OFDM

Consider an OFDM transmission system. The length-N data packet $\boldsymbol{\mathcal{X}}_i$ undergoes an IDFT operation to produce the time-domain packet $\boldsymbol{x}_i$ ($\boldsymbol{x}_i = \boldsymbol{Q}\boldsymbol{\mathcal{X}}_i$). A cyclic prefix $\underline{\boldsymbol{x}}_i$ of length P is appended to $\boldsymbol{x}_i$ resulting in the larger packet

$$\overline{\boldsymbol{x}}_i = \begin{bmatrix} \boldsymbol{x}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} \tag{3}$$

When passed through a channel $\underline{\boldsymbol{h}}$ of (maximum) length $P+1$, $\overline{\boldsymbol{x}}_i$ produces the length $N+P$ packet $\overline{\boldsymbol{y}}_i$ at the output. Just as in (3), we split the output packet $\overline{\boldsymbol{y}}_i$ as

$$\overline{\boldsymbol{y}}_i = \begin{bmatrix} \boldsymbol{y}_i \\ \underline{\boldsymbol{y}}_i \end{bmatrix} \tag{4}$$

We can show that the output prefix $\underline{\boldsymbol{y}}_i$ absorbs all ISI that takes place between $\boldsymbol{x}_{i-1}$ and $\boldsymbol{x}_i$ and that $\boldsymbol{y}_i$ is totally dependent on $\boldsymbol{x}_i$. The total OFDM channel can thus be decomposed into two constituent channels.

2.1. Circular Channel

We can show that $\boldsymbol{y}_i$ is related to $\boldsymbol{h} \triangleq [\ \underline{\boldsymbol{h}}^T \quad \boldsymbol{0}^T_{N-P-1}\ ]^T$ and $\boldsymbol{x}_i$ through circular convolution

$$\boxed{\boldsymbol{y}_i = \boldsymbol{h} \circledast \boldsymbol{x}_i + \boldsymbol{n}_i} \tag{5}$$

where $\boldsymbol{n}_i$ is the output noise which we take to be white Gaussian with variance σ_n^2. In the frequency domain, (5) reduces to an element-by-element operation

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}} \odot \boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{N}}_i} \tag{6}$$

where $\boldsymbol{\mathcal{Y}}_i$, $\boldsymbol{\mathcal{H}}$, $\boldsymbol{\mathcal{X}}_i$, and $\boldsymbol{\mathcal{N}}_i$ are the DFT's of $\boldsymbol{y}_i$, $\boldsymbol{h}$, $\boldsymbol{x}_i$, and $\boldsymbol{n}_i$, respectively. It will be useful to rewrite (6) in terms of the time-domain channel $\underline{\boldsymbol{h}}$. To do this, we simply replace $\boldsymbol{\mathcal{H}}$ by the partial FFT relationship

$$\boldsymbol{\mathcal{H}} = \underline{\boldsymbol{Q}}^*_{P+1}\underline{\boldsymbol{h}}$$

where, as per our notation, $\underline{\boldsymbol{Q}}_{P+1}$ consists of the first $P+1$ rows of $\boldsymbol{Q}$. We can thus write

$$\boldsymbol{\mathcal{Y}}_i = \text{diag}(\boldsymbol{\mathcal{X}}_i)\underline{\boldsymbol{Q}}^*_{P+1}\underline{\boldsymbol{h}} + \boldsymbol{\mathcal{N}}_i \tag{7}$$

Estimaing the channel from (7) instead of (6) makes use of the finite delay spread property, reducing the number of parameters to be estimated from N to $P+1$.

2.2. Linear Channel

Similarly, the input cyclic prefix sequence is related to the output prefix sequence through convolution with the channel [6]

$$\boxed{\underline{y}(i) = \underline{h}(i) * \underline{x}(i) + \underline{n}(i)}$$

We can write this in matrix form as

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i\underline{\boldsymbol{h}} + \underline{\boldsymbol{n}}_i \tag{8}$$

where $\underline{\boldsymbol{X}}_i$ is a $P \times (P+1)$ Toeplitz matrix created from the cyclic prefixes $\underline{\boldsymbol{x}}_{i-1}$ and $\underline{\boldsymbol{x}}_i$.

2.3. Total Channel

Naturally, the sequences $\overline{\boldsymbol{x}}_i$ and $\overline{\boldsymbol{y}}_i$ are related by linear convolution with the channel, i.e.

$$\overline{y}(i) = \underline{h} * \overline{x}(i) + \overline{n}(i)$$

Alternatively, with $\overline{\boldsymbol{h}} \triangleq [\ \underline{\boldsymbol{h}}^T \quad \boldsymbol{0}^T_{N-1}\ ]^T$, we can write

$$\boxed{\overline{y}(i) = \overline{h}(i) * \overline{x}(i) + \overline{n}(i)} \tag{9}$$

By combining (7) and (8), we can put the total convolution in the following more useful (matrix) form

$$\boxed{\overline{\boldsymbol{\mathcal{Y}}}_i = \overline{\boldsymbol{X}}_i\underline{\boldsymbol{h}} + \overline{\boldsymbol{\mathcal{N}}}_i} \tag{10}$$

where, in line with the notational convention (3),

$$\overline{\boldsymbol{X}}_i \triangleq \begin{bmatrix} \text{diag}(\boldsymbol{\mathcal{X}}_i)\underline{\boldsymbol{Q}}^*_{P+1} \\ \underline{\boldsymbol{X}}_i \end{bmatrix}, \quad \overline{\boldsymbol{\mathcal{Y}}}_i = \begin{bmatrix} \boldsymbol{\mathcal{Y}}_i \\ \underline{\boldsymbol{y}}_i \end{bmatrix}, \quad \overline{\boldsymbol{\mathcal{N}}}_i = \begin{bmatrix} \boldsymbol{\mathcal{N}}_i \\ \underline{\boldsymbol{n}}_i \end{bmatrix}$$

2.4. Channel and Data Recovery

The input/output relationships (6) and (10) are all that is needed to perform channel and data recovery. Specifically, equation (6) can be used for optimal (MMSE) data recovery. Assuming that $\mathcal{X}_i(l)$ takes on its values from the alphabet $A = \{A_1, A_2, \ldots, A_{|A|}\}$ with equal probability, we can show that the MMSE estimate is given by [1]

$$\hat{\mathcal{X}}_i^{\mathsf{MMSE}}(l) = E[\mathcal{X}_i(l)|\boldsymbol{\mathcal{Y}}_i] = \frac{\sum_{j=1}^{j=|A|} A_j e^{-\frac{|\mathcal{Y}_i(l)-\mathcal{H}(l)A_j|^2}{\sigma_n^2}}}{\sum_{j=1}^{|A|} e^{-\frac{|\mathcal{Y}_i(l)-\mathcal{H}(l)A_j|^2}{\sigma_n^2}}} \tag{11}$$

Similarly, we can calculate the second-order moment of the data as

$$E[|\mathcal{X}_i(l)|^2|\boldsymbol{\mathcal{Y}}_i] = \frac{\sum_{j=1}^{j=|A|} |A_j|^2 e^{-\frac{|\mathcal{Y}_i(l)-\mathcal{H}(l)A_j|^2}{\sigma_n^2}}}{\sum_{j=1}^{|A|} e^{-\frac{|\mathcal{Y}_i(l)-\mathcal{H}(l)A_j|^2}{\sigma_n^2}}} \tag{12}$$

[1] This should outperform the linear estimate

$$\hat{\mathcal{X}}_i^{\mathsf{MMSE}}(l) = \frac{\sqrt{\mathcal{E}_X}\mathcal{H}^*(l)}{\mathcal{E}_X||\mathcal{H}(l)||^2 + \sigma_n^2}\mathcal{Y}_i(l)$$

that is usually employed in literature and which is not globally optimum in the finite alphabet case.

Equation (10) can be used for the dual job of channel estimation. Thus, given an estimate of the ith packet $\boldsymbol{\mathcal{X}}_i$ and the cyclic prefix of the previous packet $\underline{\boldsymbol{x}}_{i-1}$, we can construct the matrix $\overline{\boldsymbol{X}}_i$ and perform channel estimation by solving (10) in the LS sense

$$\min_{\underline{h}} \|\overline{\boldsymbol{y}}_i - \overline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}\|^2 \tag{13}$$

By iterating between (11) and (13), we can perform joint channel and data recovery from the output data (see [10]). The iterative procedure is best articulated as an expectation-maximization (EM) iterative algorithm. The algorithm is kick-started from an initial channel estimate obtained from a set of pilots, as described below.

2.5. Pilots for Initial Channel Estimation

With a maximum of $P+1$ active channel taps, we need an equal number of pilots to identify the channel uniquely. We can, however, capitalize on other natural constraints and reduce the number of pilots necessary to initialize the estimation process. Clearly, only the cyclic (diagonal) channel can make use of the pilot information. Now let $I_{\mathsf{p}} = \{i_1, i_2, \ldots, i_{L_{\mathsf{p}}}\}$ denote the index set of the pilot bins. Starting from (7), the pilots induce the following subsystem of equations

$$\boldsymbol{\mathcal{Y}}_{i\, I_{\mathsf{p}}} = \operatorname{diag}\left(\boldsymbol{\mathcal{X}}_{i\, I_{\mathsf{p}}}\right) \underline{\boldsymbol{Q}}^*_{P+1} \underline{\boldsymbol{h}} + \boldsymbol{\mathcal{N}}_{i_{I_{\mathsf{p}}}} \tag{14}$$

The subscript I_{p}, e.g. in $\boldsymbol{\mathcal{Y}}_{i_{I_{\mathsf{p}}}}$, acts as an the indicator set of the rows in the vector or matrix. As we would like to reduce the number of pilots as much as possible, (14) is usually underdetermined ($L_{\mathsf{p}} < P+1$) and hence must be solved in the regularized LS sense

$$\min_{\underline{h}} \alpha\|\underline{\boldsymbol{h}}\|^2 + \|\boldsymbol{\mathcal{Y}}_{i\, I_{\mathsf{p}}} - \operatorname{diag}\left(\boldsymbol{\mathcal{X}}_{i\, I_{\mathsf{p}}}\right) \underline{\boldsymbol{Q}}^*_{P+1} \underline{\boldsymbol{h}}\|^2_W \tag{15}$$

3. THE EM ALGORITHM FOR JOINT CHANNEL AND DATA RECOVERY

3.1. The EM Algorithm

Ideally, we identify $\underline{\boldsymbol{h}}$ by maximizing the log-likelihood function

$$\hat{\underline{\boldsymbol{h}}}^{\mathsf{ML}} = \arg\max_{\underline{h}} \ln p\left(\underline{\boldsymbol{h}} | \boldsymbol{\mathcal{X}}_i, \overline{\boldsymbol{\mathcal{Y}}}_i\right)$$

Since the input $\boldsymbol{\mathcal{X}}_i$ is unobserved, we maximize instead an averaged form of the likelihood function using the expectation-maximization (EM) algorithm. To this end, we split the available variables into three classes

- The parameters to be identified which in our case consist of the channel impulse response $\underline{\boldsymbol{h}}$
- The observed data $\overline{\boldsymbol{\mathcal{Y}}}_i = \left[\begin{array}{cc} \boldsymbol{\mathcal{Y}}_i^T & \underline{\boldsymbol{y}}_i^T \end{array}\right]^T$ consisting of the output of both channels
- The unobserved or hidden data $\boldsymbol{\mathcal{X}}_i$

The EM algorithm attempts to identify $\underline{\boldsymbol{h}}$ by solving the following problem iteratively

$$\hat{\underline{\boldsymbol{h}}}^{(j+1)} = \arg\max_{\underline{h}} E_{\mathcal{X}_i | \overline{\mathcal{Y}}_i, \hat{\underline{h}}^{(j)}} \left[\ln p(\underline{\boldsymbol{h}} | \boldsymbol{\mathcal{X}}_i, \overline{\boldsymbol{\mathcal{Y}}}_i)\right] \tag{16}$$

Thus, each iteration involves expectation and maximization steps. As we will now show, this boils down to minimizing a regularized quadratic expression in $\underline{\boldsymbol{h}}$. To this end, consider the input/output equation (10) of the total channel and note that

$$\begin{aligned} \ln p(\underline{\boldsymbol{h}} | \boldsymbol{\mathcal{X}}_i, \overline{\boldsymbol{\mathcal{Y}}}_i) &= \ln p(\boldsymbol{\mathcal{X}}_i, \overline{\boldsymbol{\mathcal{Y}}}_i | \underline{\boldsymbol{h}}) \\ &= \ln p(\boldsymbol{\mathcal{X}}_i | \underline{\boldsymbol{h}}) + \ln p(\overline{\boldsymbol{\mathcal{Y}}}_i | \boldsymbol{\mathcal{X}}_i, \underline{\boldsymbol{h}}) \\ &= \ln p(\boldsymbol{\mathcal{X}}_i) + \ln p(\overline{\boldsymbol{\mathcal{Y}}}_i | \boldsymbol{\mathcal{X}}_i, \underline{\boldsymbol{h}}) \end{aligned}$$

where we assumed that the channel response is deterministic and hence independent of the input. Since the noise is Gaussian, we can further write

$$\ln p(\overline{\boldsymbol{\mathcal{Y}}}_i | \boldsymbol{\mathcal{X}}_i, \underline{\boldsymbol{h}}) = -(N+P)\ln(\sigma_n^2) - \frac{1}{\sigma_n^2}\left\|\overline{\boldsymbol{\mathcal{Y}}}_i - \overline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}\right\|^2 \tag{17}$$

Thus, the iterative relation (16) for the EM algorithm becomes

$$\hat{\underline{\boldsymbol{h}}}^{(j+1)} = \arg\max_{\underline{h}} -E_{\mathcal{X}_i | \overline{\mathcal{Y}}_i, \hat{\underline{h}}^{(j)}} \left\|\overline{\boldsymbol{\mathcal{Y}}}_i - \overline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}\right\|^2 \tag{18}$$

It now remains to evaluate the expectation in (18).

3.2. The Expectation Step

In the absence of the expectation operator, the objective function (18) becomes quadratic in $\underline{\boldsymbol{h}}$ and is subsequently easy to optimize. The effect of the expectation operator is to replace $\overline{\boldsymbol{X}}_i$ in (18) by its expectation and to add a regularizing term, thus transforming the LS into a regularized form

$$\hat{\underline{\boldsymbol{h}}}^{(j+1)} = \arg\max_{\underline{h}} -\left\{\left\|\overline{\boldsymbol{\mathcal{Y}}}_i + E\left[\overline{\boldsymbol{X}}_i\right]\underline{\boldsymbol{h}}\right\|^2 + \|\underline{\boldsymbol{h}}\|^2_{\operatorname{Cov}[\overline{X}_i^*]}\right\} \tag{19}$$

where $\operatorname{Cov}(\overline{\boldsymbol{X}}_i^*)$ is the covariance matrix of $\overline{\boldsymbol{X}}_i^*$

$$\operatorname{Cov}(\overline{\boldsymbol{X}}_i^*) = E\left[\overline{\boldsymbol{X}}_i^* \overline{\boldsymbol{X}}_i\right] - E\left[\overline{\boldsymbol{X}}_i^*\right] E\left[\overline{\boldsymbol{X}}_i\right] \tag{20}$$

The first and second moments of $\overline{\boldsymbol{X}}_i$ in (19) can be calculated from the mean and second moment of the data packet $\boldsymbol{\mathcal{X}}_i$, which are already evaluated in (11) and (12), respectively.

3.3. The Maximization Step

The maximization step is now straightforward to carry out. Specifically, we have

$$\begin{aligned} \hat{\underline{\boldsymbol{h}}}^{(j+1)} &= \left(\operatorname{Cov}[\overline{\boldsymbol{X}}_i^*] + E\left[\overline{\boldsymbol{X}}_i\right]^* E\left[\overline{\boldsymbol{X}}_i\right]\right)^{-1} E\left[\overline{\boldsymbol{X}}_i\right]^* \overline{\boldsymbol{\mathcal{Y}}}_i \\ &= \left(E\left[\overline{\boldsymbol{X}}_i^* \overline{\boldsymbol{X}}_i\right]\right)^{-1} E\left[\overline{\boldsymbol{X}}_i\right]^* \overline{\boldsymbol{\mathcal{Y}}}_i \end{aligned} \tag{21}$$

If the noise variance is not known, the EM algorithm can also be used to estimate it. We can show that the noise estimate is related to the (EM-based) channel estimate by

$$\sigma_n^{2\,(j+1)} = \frac{1}{N+P}\left(\left\|\overline{\boldsymbol{\mathcal{Y}}}_i - E\left[\overline{\boldsymbol{X}}_i\right]\hat{\underline{\boldsymbol{h}}}^{(j+1)}\right\|^2 + \left\|\hat{\underline{\boldsymbol{h}}}^{(j+1)}\right\|^2_{\operatorname{Cov}[\overline{X}_i^*]}\right)$$

4. UTILIZING A PRIORI STATISTICAL INFORMATION

At the receiver, we usually have additional statistical information about the channel. This information can be used to enhance channel estimation. Specifically, wireless channels are usually Gaussian distributed with certain mean (zero or non-zero mean depending on whether fading is Rayleigh or Rician) and certain covariance (also known as the power delay profile).

Thus, assuming that the channel follows a normal distribution $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Pi})$, its pdf takes the form

$$p(\underline{\boldsymbol{h}}) = \frac{1}{((2\pi)^{P+1}|\boldsymbol{\Pi}|)^{1/2}} e^{-\|\underline{\boldsymbol{h}}-\boldsymbol{m}\|^2_{\Pi^{-1}}} \tag{22}$$

which corresponds to the log-likelihood function (discarding a constant term)

$$\ln p(\underline{\boldsymbol{h}}) = -\|\underline{\boldsymbol{h}} - \boldsymbol{m}\|^2_{\Pi^{-1}} \tag{23}$$

With this additional information, the EM-based channel estimate (16) takes the form

$$\begin{aligned}\hat{\underline{\boldsymbol{h}}}^{(j+1)} &= \arg\max_{\underline{h}} E_{\mathcal{X}_i|\overline{\mathcal{Y}}_i,\hat{\underline{h}}^{(j)}} \ln p(\boldsymbol{\mathcal{X}}_i, \overline{\boldsymbol{\mathcal{Y}}}_i|\underline{\boldsymbol{h}}) + \ln p(\underline{\boldsymbol{h}}) \\ &= \begin{array}{l}\arg\max_{\underline{h}} \Big\{ -\frac{1}{\sigma_n^2} \left\|\overline{\boldsymbol{\mathcal{Y}}}_i - E[\overline{\boldsymbol{X}}_i]\underline{\boldsymbol{h}}\right\|^2 \\ -\frac{1}{\sigma_n^2} \|\underline{\boldsymbol{h}}\|^2_{\mathrm{Cov}[\overline{X}_i^*]} \quad - \|\underline{\boldsymbol{h}} - \boldsymbol{m}\|^2_{\Pi^{-1}} \Big\}\end{array}\end{aligned}$$

Alternatively, we can write the objective function as a single Euclidean norm

$$\underline{\boldsymbol{h}}^{(j+1)} = \arg\max_{\underline{h}} \; -\|\boldsymbol{b} - \boldsymbol{A}\underline{\boldsymbol{h}}\|^2 \tag{24}$$

where

$$\boldsymbol{b} = \begin{bmatrix} \frac{1}{\sigma_n}\overline{\boldsymbol{\mathcal{Y}}}_i \\ \boldsymbol{\Pi}^{-1/2}\boldsymbol{m} \\ \boldsymbol{O} \end{bmatrix} \quad \text{and} \quad \boldsymbol{A} = \begin{bmatrix} \frac{1}{\sigma_n} E\left[\overline{\boldsymbol{X}}_i\right] \\ \boldsymbol{\Pi}^{-1/2} \\ \frac{1}{\sigma_n}\mathrm{Cov}\left[\overline{\boldsymbol{X}}_i^*\right]^{1/2} \end{bmatrix} \tag{25}$$

It is easy to show that the corresponding channel estimate is given by

$$\begin{aligned}\hat{\underline{\boldsymbol{h}}}_i^{(j+1)} = &\left(\frac{1}{\sigma_n^2} E\left[\overline{\boldsymbol{X}}_i^* \overline{\boldsymbol{X}}_i\right] + \boldsymbol{\Pi}^{-1}\right)^{-1} \\ &\left(\frac{1}{\sigma_n^2} E\left[\overline{\boldsymbol{X}}_i\right]^* \overline{\boldsymbol{\mathcal{Y}}}_i + \boldsymbol{\Pi}^{-1}\boldsymbol{m}\right)\end{aligned} \tag{26}$$

Remarks

- The semi-blind algorithm described above uses the constraints imposed by the finite alphabet nature of the input and pilots. It also makes use of the redundancy exhibited by the cyclic prefix. We can also utilize redundancy due the presence of a real code, something that is relegated to [9].
- We also utilized the finite delay spread nature of the channel and the a priori statistical information. It is straight forward to extend the above developments to the case where the channel has a sparse nature when the active taps locations are known a priori. Utilizing these constraints always boils down to solving a (regularized) LS problem as reflected by (15), (21), and (26).
- We have assumed that the channel remains constant over one OFDM packet. When channel variation from one packet to the next follows a state-space structure, we can use the previous channel estimate together with the Kalman filter to initialize channel recovery for the new OFDM packet, thereby reducing the number of pilots that are eventually needed.

5. SIMULATIONS

We consider an OFDM transmission system with packet length $N = 128$ and cyclic prefix length $P = 15$. The channel is assumed to have 16 active complex taps. The input is taken to be 4-QAM transmitted at an SNR = 15 dB. Channel identification is performed within the same packet using a number of pilots and the mean-square error of the channel estimate is plotted vs the number of iterations. We demonstrate the performance of the EM algorithm under various constraints. Fig. 1 shows the learning curves for the algorithm using 8,13, and 16 pilots without using any statistical information. As expected, the accuracy of channel estimation improves as the number of pilots increases. We also note that convergence is achieved within 6 iterations.

We next demonstrate the improvement in the performance of the EM algorithm when the additional statistical information (frequency correlation) is available. We perform the same set of simulations for a channel with an exponentially decaying power profile and run the EM algorithm both when this information is available to the receiver and when it is not. We note (see Fig. 2) that channel estimation with 5 pilots and a priori statistical knowledge outperforms estimation with 14 pilots and no statistical information. When only 4 pilots are available (in addition to a priori statistical information), the algorithm attains the same accuracy as that for the 14 pilot case at the cost of increased number of iterations. Finally, Fig. 3 depicts the gain that the EM iterations provide vs SNR. Thus, the upper curve in this figure represents the channel estimation error using 4 pilots in addition to frequency correlation information without invoking the EM iterations. The bottom curve is the final estimation error after the EM iterations are completed.

6. CONCLUSION

In this paper, we considered the problem of semi-blind channel and data recovery in OFDM. Specifically, we designed an EM receiver that makes use of the data and channel constraints to perform this recovery with zero latency and with small pilot overhead. The receiver uses the pilots to kick start channel estimation and subsequently iterates between that and data recovery. The receiver utilizes the data constraints (which include the cyclic prefix as well as the finite

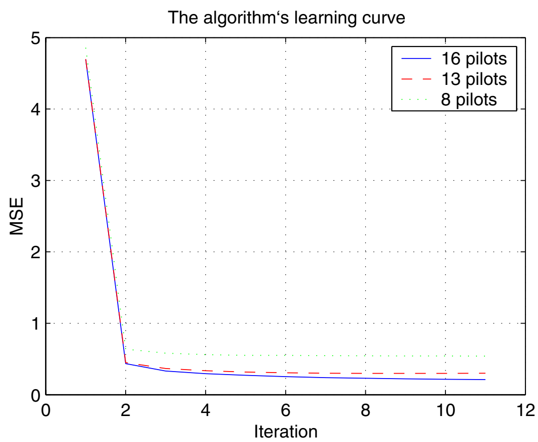

Figure 1: Reduction in MSE due to pilot overhead– EM learning curves for different number of pilots

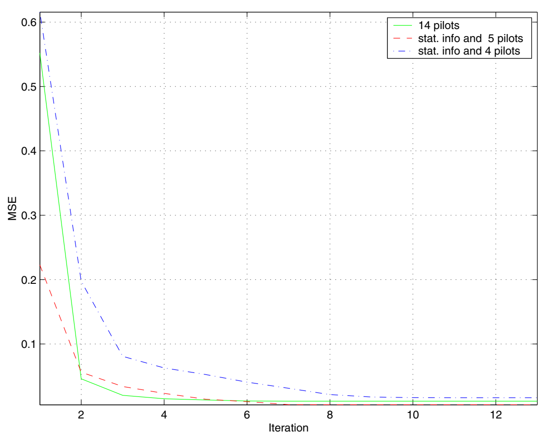

Figure 2: Reduction in MSE due to a priori statistical information– EM learning curves in the presence and absence of a priori statistical information

alphabet nature of the data) and employs the data estimates in soft format. The receiver also makes use of the various constraints on the channel (which includes sparsity, finite delay spread information, and frequency correlation). Thanks to the presence of the cyclic prefix, optimal data recovery is always done on element by element basis. Channel recovery always boils down to solving a regularized least-squares problem.

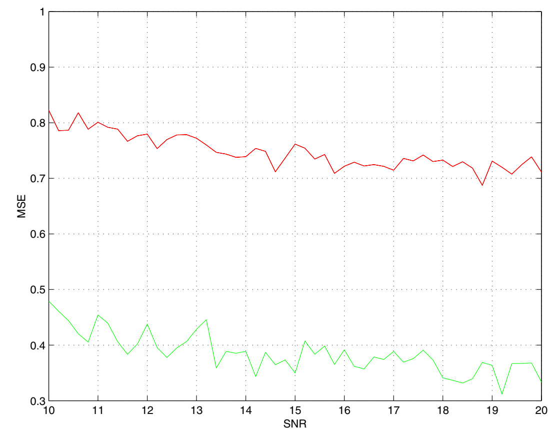

Figure 3: Reduction in MSE due to EM iterations– steady-state MSE vs. SNR before and after invoking the EM iterations

7. REFERENCES

[1] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Process. Mag.,* vol. 17, no. 3, pp. 29-48, May 2000.

[2] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electr.,* vol. 44, no. 3 , pp. 1122–1128, Aug. 1998.

[3] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *ICASSP 2001*, vol. 4, pp. 2389–2392.

[4] R. W. Heath Jr. and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Process.,* vol. 47, pp. 774-780, Mar. 1999.

[5] B. Muquet, M. de Courville, P. Duhamel, and V. Buzenae, "A subspace based blind and semi-blind channel identification method for OFDM systems," *Proc. IEEE Workshop for Signal Process. Advances in Wireless Comm.,* Annapolis, MD, pp. 243-246, May 1999.

[6] X. Wang and R. Liu, "Adaptive channel estimation using cyclic prefix in multicarrier modulation system," *IEEE Comm. Lett.,* vol. 3, no. 10, Oct. 1999.

[7] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.,* vol. 46, no. 7 , pp. 902–915, Jul. 1998.

[8] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Communications,* vol. 46, no. 7, pp. 931–939, Jul. 1998.

[9] G. Al-Rawi, T. Y. Al-Nafouri, A. Bahai, and J. Cioffi, "Exploiting error-control coding and cyclic prefix in channel estimation for coded OFDM systems," *to appear in GLOBECOM 2002.*

[10] T. Y. Al-Naffouri, G. Al-Rawi, A. Bahai, and A. Paulraj, "A least-/ mean- squares approach to channel identification and equalization in OFDM," *ICASSP 2002*, vol. 3, pp. 2577 -2580