Iterative Algorithms for Channel Identification Using Superimposed Pilots

Angiras R. Varma, Lachlan L. H. Andrew, Chandra R. N. Athaudage, Jonathan H. Manton

***Abstract*— Channel identification of a time-varying channel is considered using superimposed training. A sequence of known symbols with lower power is arithmetically added to the information symbols before modulation and transmission. The channel estimation is done exploiting the known superimposed data in the transmitted signal. Two iterative algorithms are considered in this paper: recursive least squares (RLS) and the expectation maximization (EM). Performance of the proposed algorithms is compared with a simple avergaing scheme and the LMS algorihm. For short data blocks RLS outperforms EM, but with large blocks EM is superior.**

I. Introduction

In the conventional approach of channel identification using pilots, pilot symbols are time-multiplexed with information symbols before transmission. This technique wastes bandwidth, causing a reduction in data rate. For a time-varying channel, the conventional technique needs to transmit pilot data frequently enough to keep up with the channel variation. An alternative approach, studied in [1]-[7], is known as superimposed pilot training. The idea behind this approach is superimposing or arithmetically adding known pilot data at a lower power to the information data before modulation. Channel identification is done utilizing the superimposed data in transmitted symbols. The major advantage of this scheme is that the wastage of bandwidth is eliminated. In particular, a time-varying channel can be tracked using the superimposed pilots with out causing further wastage of resources. But the downside of this approach in a system with fixed transmit power is the allocation of power to superimposed pilots from the power budget for information data. This causes a reduction in the transmit power of information signals and effectively reduces the of signal to noise ratio (SNR), adversly affecting the bit error rate (BER). The optimal power allocation to achieve the balance between channel estimation accuracy and effective SNR is studied in [10]. Application of superimposed pilot training is found in multi-carrier systems such as orthogonal frequncy division multiplexing (OFDM) [12]. In OFDM systems, it is shown that the superimposed scheme is not only applicable for channel estimation/detection but also for peak to average power ratio (PAPR) reduction [11]. Application of this scheme is also extended to multi-input multi-output (MIMO) systems in [8] and [9].

The authors are with the ARC Special Research Center for Ultra-Broadband Information Networks (CUBIN), an affiliated program of National ICT Australia (NICTA), Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia. e-mail: {r.angiras,l.andrew,c.athaudage,j.manton}@ee.mu.oz.au

Different techniques for channel identification are considered in the literature [1]-[7]. Many of these techniques, [1]-[4], [7] obtain the channel estimate using first-order statistics. Within the first-order statistical approaches, periodicity of the pilots is explicitly used in [4], [7], whereas [4] also uses the cyclostationary nature of the received symbols. A superimposed pilot scheme for guard interval based systems is studied in [3]. The scheme in [5] uses first-order statistics to obtain the initial estimate of the channel and further improves this estimate using Deterministic Maximum Likelihood (DML) approach. The approach in [6] also starts with an initial estimate of the channel using first-order statistics and improves the estimate using decision feedback after symbol detection.

This paper considers two iterative algorithms namely RLS and EM algorithms for channel estimation. These two algorithms have not yet been considered for channel estimation in the context of superimposed pilot training in the literature. RLS is an iterative algorithm to obtain the least squares estimate [13]. In this scheme a small number of known pilot symbols are transmitted, just like the conventional time multiplexed scheme, to improve the initial convergence time. After this the RLS algorithm tracks the channel. As the second approach this paper derives an EM algorithm to obtain the maximum likelihood estimate of the channel. The EM algorithm iteratively converges to a local maximum of the likelihood function [14]. This algorithm make use of the statistical properties of transmitted sequences. Initial training with the known pilot symbols are used in the EM algorithm. Both RLS and EM are iterative algorithms but there is a clear distinction between two. RLS make use of the previous channel estimates right from the beginning of the algorithm, whereas the standard EM algorithm iterates over the same received block repeatedly and treats blocks independently. In this paper, we propose a variant of the EM algorithm which incorporates memory to allow noise to average out.

The organization of this paper is as follows. Section II describes the system model and the problem formulation. Section III explains the RLS algorithm and shows how to incorporate initial training. Derivation of the EM algorithm for channel identification is given in Section IV. Simulation results are presented in Section V followed by conclusions in Section VI.

II. System Model and Problem formulation

Consider a time-varying channel vector given by $\mathbf{h}(k) = \left[\begin{array}{c} h_0(k) \ldots\ldots h_{L-1}(k) \end{array}\right]^T$ at time index k. Let s_i be the ith symbol transmitted, and define the kth transmitted symbol vector $\mathbf{s}_k = \left[\begin{array}{c} s_k \ldots\ldots s_{k-L+1} \end{array}\right]^T$. Similarly, let c_i

be the ith superimposed pilot transmitted, and let $\mathbf{c}_k = \left[\begin{array}{c} c_k \ldots\ldots c_{k-L+1} \end{array}\right]^T$ at a time index k. The received symbol y_k in presence of the noise n_k is given by

$$y_k = \sum_{i=0}^{L-1} h_i(k) s_{k-i} + \sum_{i=0}^{L-1} h_i(k) c_{k-i} + n_k, \tag{1}$$

where n_k is Gaussian distributed, $\mathcal{N}(0, \sigma_n^2)$. This paper explores the options of estimating the channel vector $\mathbf{h}(k)$ given the received symbol y_j and the superimposed pilot symbol vector $\mathbf{c}_j$ for the time indices $1 \leq j \leq k$.

III. RLS Algorithm for channel identification

A. RLS Algorithm

The weighted least squares estimate of the channel vector is $\mathbf{h}$ that minimizes the error function

$$\xi_k = \sum_{i=1}^{k} w(k,i)|e_i|^2, \tag{2}$$

where $e_i = y_i - \mathbf{h}(k)^T \mathbf{c}_i$ is the symbol error and $w(k,i)$ is the weight function. The least squares estimate is given as

$$\hat{\mathbf{h}}(k) = \left[\sum_{i=1}^{k} w(k,i)\mathbf{c}_i \mathbf{c}_i^T\right]^{-1} \sum_{i=1}^{k} w(k,i)\mathbf{c}_i y_i, \tag{3}$$

where $\sum_{i=1}^{k} w(k,i)\mathbf{c}(i)\mathbf{c}(i)^T$ is the weighted correlation matrix of the superimposed data vectors. Estimation of the channel coefficient vector $\hat{\mathbf{h}}(k)$ using (3) is computationally costly. For each new k, inversion of the weighted correlation matrix of superimposed data is necessary. This can be avoided using the RLS algorithm, as it updates the correlation matrix with an increment of k without matrix inversion. This gives a benefit in terms of computational complexity. Consider a weighting function of the form $w(k,i) = A_i \lambda^{k-i}$, where λ^{k-i} is a forgetting factor that helps to forget the past history of a time-varying channel. A_i scales the ith error and scaling is used to emphasise or de-emphasise an error. With this specification of A_i, the RLS algorithm in [13] becomes

Initialization step, for $k = 0$

$$\begin{aligned} \mathbf{P}_0 &= \delta^{-1}\mathbf{I} \\ \hat{\mathbf{h}}(0) &= \mathbf{0} \end{aligned}$$

Repeat for $k = 1, 2, \ldots$

$$\begin{aligned} \mathbf{M}_k &= \frac{\lambda^{-1}\mathbf{P}_{k-1}\mathbf{c}_k}{{A_i}^{-1} + \lambda^{-1}\mathbf{c}_k^T \mathbf{P}_{k-1}\mathbf{c}_k} \\ \xi_k &= y_k - \hat{\mathbf{h}}_{k-1}^T \mathbf{c}_k \\ \hat{\mathbf{h}}(k) &= \hat{\mathbf{h}}(k-1) + \mathbf{M}_k \xi_k \\ \mathbf{P}_k &= \lambda^{-1}\mathbf{P}_{k-1} - \lambda^{-1}\mathbf{M}_k \mathbf{c}_k^T \mathbf{P}_{k-1}, \end{aligned} \tag{4}$$

where δ is a small positive constant, which is used for initialization. $\mathbf{P}_k$ is the inverse of the weighted correlation matrix and $\mathbf{M}_k = \mathbf{P}_k \mathbf{c}_k$.

B. Selection of Scaling Factor A_i

In the scheme of channel estimation using the RLS algorithm, a small number of pilot symbols with full power are transmitted for the initial training followed by the superimposed pilot symbols. The initial pilots have a better SNR than the superimposed ones and should be given more weight. In this scheme an approximate expression for the scaling factor is derived as follows. Let p be the number of full pilot symbols transmitted and $k-p$ be the number of superimposed pilot symbols transmitted. Treating the contribution of the transmitted symbols in the received symbols as noise, (1) is modified as

$$y_i = \mathbf{h}^T(i)\mathbf{c}_i + w_i, \tag{5}$$

where w_i is the noise from the information symbols and from the channel. The distribution of w_i is approximated to be white Gaussian $\mathcal{N}(0, \sigma_n^2)$ for $1 \leq i \leq p$ and $\mathcal{N}(0, \sigma_n^2 + \sigma_s^2)$ for $p+1 \leq i \leq k$, where σ_s^2 is the information symbol power and σ_n^2 is the channel noise power. The log likelihood function of received data given the channel vector $\log p(y_i \mid \mathbf{h}(i))$ is

$$\begin{aligned} \log p(y_i \mid \mathbf{h}(i)) &= K - \left[\frac{(y_i - \mathbf{h}(i)^T \mathbf{c}_i)^2}{2\sigma_n^2}\right], 1 \leq i \leq p \\ &= K - \left[\frac{(y_i - \mathbf{h}(i)^T \mathbf{c}_i)^2}{2(\sigma_n^2 + \sigma_s^2)}\right], p+1 \leq i \leq k, \end{aligned} \tag{6}$$

where K is the terms independent of y_i and $\mathbf{h}(i)$. Log likelihood function $\log p(y_1 \ldots y_k \mid \mathbf{h}(1) \ldots \mathbf{h}(k)) = \log p(y_1 \mid \mathbf{h}(1)) + \ldots + \log p(y_k \mid \mathbf{h}(k))$ is given by

$$\begin{aligned} &\log p(y_1 \ldots y_k \mid \mathbf{h}(1) \ldots \mathbf{h}(k)) = \\ &K - \left[\sum_{i=1}^{p} \frac{e_i^2}{2\sigma_n^2} + \sum_{i=p+1}^{k} \frac{e_i^2}{2(\sigma_n^2 + \sigma_s^2)}\right], \end{aligned} \tag{7}$$

where $e_i = y_i - \mathbf{h}(i)^T \mathbf{c}_i$ is the symbol error. The maximum likelihood estimate of the channel can be obtained from (7) as

$$\begin{aligned} \hat{\mathbf{h}} &= \arg\min_{\mathbf{h}} \sum_{i=1}^{p} \frac{e_i^2}{2\sigma_n^2} + \sum_{i=p+1}^{k} \frac{e_i^2}{2(\sigma_n^2 + \sigma_s^2)} \\ &= \arg\min_{\mathbf{h}} \frac{\sigma_n^2 + \sigma_s^2}{\sigma_n^2} \sum_{i=1}^{p} e_i^2 + \sum_{i=p+1}^{k} e_i^2. \end{aligned} \tag{8}$$

Comparing (2) with (8), scaling factor A_i for $\lambda = 1$ is

$$A_i = \begin{cases} \frac{\sigma_n^2 + \sigma_s^2}{\sigma_n^2} & 1 \leq i \leq p \\ 1 & p+1 \leq i \leq k \end{cases} \tag{9}$$

This weighting will be used in the results of Section V. It is worth noting that for a static channel, the least squares estimate (3) with a scaling factor of (9) in the presence of white Gaussian noise is the maximum likelihood (ML) estimate.

IV. Channel Identification using EM Algorithm

Although RLS gives the ML estimate when the symbols are independent, more powerful techniques such as EM are generally required. This section illustrates how the EM algorithm can make use of the embedded pilots. For expositional simplisity the case of independent symbols and white noise will be given, but the general case is similar.

A. Introduction to EM Algorithm

Let θ be the parameter to be estimated from a given observation y which is an outcome of a sample space Y. Then the maximum likelihood estimate of θ is

$$\hat{\theta} = \arg\max p(y \mid \theta) \tag{10}$$

where $p(y \mid \theta)$ is the conditional probability density function of the observed data given the unknown parameter. One way of computing ML estimate is by using the EM algorithm. The EM algorithm iteratively converges to a local maximum of the likelihood function. The framework of the EM algorithm is as follows.

Let x be an outcome of a sample space X such that x is not observed directly but only by means of y. Then x is termed "complete data" to indicate that there is a many to one mapping from X to Y, and the observation y is referred as the incomplete data. Let θ be the parameter to be estimated. Let $p(X \mid \theta)$ denote the probability density function of the complete data given the unknown parameter. The aim behind the EM algorithm is to find θ that maximizes $\log p(X \mid \theta)$ under the condition that x is not directly observable, given only an observation y. The EM algorithm does this job in two steps namely the Expectation step (E-step) and the Maximization step (M-step). For a given estimate of the unknown parameter $\theta^{[k]}$ and an observation y the E-step is given as

$$Q(\theta \mid \theta^{[k]}) = E\left[\log p(X \mid \theta) \mid y, \theta^{[k]}\right], \tag{11}$$

where k is the iteration count. The M-step $\theta^{[k+1]}$ to be the θ that maximizes $Q(\theta \mid \theta^{[k]})$:

$$\theta^{[k+1]} = \arg\max_{\theta} Q(\theta \mid \theta^{[k]}). \tag{12}$$

After initializing $\theta^{[0]}$, the EM algorithm applies (11) and (12) repeaatedly for increasing k until $\theta^{[k]}$ converges.

B. Channel Model

Grouping the transmitted symbols into blocks, system model (1) can be represented as follows. Consider a transmission scheme in which known pilot symbols $\mathbf{c} \in \mathbb{R}^N$ are superimposed on the transmitted data symbols $\mathbf{s} \in \mathbb{R}^N$. Considering a finite impulse response channel $\mathbf{H} \in \mathbb{R}^{M \times N}$ with zero mean additive white Gaussian noise $\mathbf{n} \in \mathbb{R}^M$ with covariance matrix $\sigma_n^2 \mathbf{I}_M$, the received symbol vector $\mathbf{y} \in \mathbb{R}^M$ is given by

$$\mathbf{y} = \mathbf{Hs} + \mathbf{Hc} + \mathbf{n}, \tag{13}$$

where the channel is assumed to be of known length $L = N - M + 1$, with channel coefficients $\mathbf{h} = \left[\begin{array}{c} h_0 \ldots\ldots h_{L-1} \end{array}\right]^T$. The channel matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ is constructed from the channel vector $\mathbf{h}$ and is a Toeplitz matrix given as

$$\left[\begin{array}{ccccccc} h_{L-1} & \ldots & h_0 & 0 & 0 & \ldots & 0 \\ 0 & h_{L-1} & \ldots & h_0 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & h_{L-1} & \ldots & h_0 \end{array}\right]. \tag{14}$$

The above channel model describes a block of symbols from a continuous transmission of data i.e. it is not block based transmission with guard intervals.

The aim of the EM algorithm is to estimate channel coefficient vector $\mathbf{h}$ from received symbols $\mathbf{y}$ by exploiting the presence of known superimposed pilots $\mathbf{c}$ in the transmitted data.

C. EM Algorithm for Channel Identification

This section derives an EM algorithm for the problem defined in Section IV-B. There exists constant matrices $\mathbf{J}_i = \partial \mathbf{H} / \partial h_i \in \mathbb{R}^{M \times N}, i = 0 \ldots L-1$ such that (13) becomes

$$\mathbf{y} = \left(\sum_{i=0}^{L-1} h_i \mathbf{J}_i\right) \mathbf{s} + \left(\sum_{i=0}^{L-1} h_i \mathbf{J}_i\right) \mathbf{c} + \mathbf{n}. \tag{15}$$

The framework for the EM algorithm is as follows. The complete data is $\mathbf{x} = \left[\begin{array}{cc} \mathbf{s}^T & \mathbf{y}^T \end{array}\right]^T$, which is the combination of information symbols and the received data. The received data $\mathbf{y}$ is the incomplete data. (Note that the complete and incomplete data are defined in the previous subsection). Similarly the unknown parameters to be estimated, θ, are the channel coefficients $\mathbf{h} = \left[\begin{array}{c} h_0 \ldots\ldots h_{L-1} \end{array}\right]^T$. The distribution of the noise vector $\mathbf{n}$ is defined in Section IV-B as $\mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$. We assume the distribution of transmitted symbol vector $\mathbf{s}$ is zero mean white Gaussian with a covariance matrix of $\sigma_s^2 \mathbf{I}_N$. In this framework, the E-step and the M-step are

$$\begin{aligned} Q(\mathbf{h} \mid \mathbf{h}^{[k]}) &= E\left[\log p(\mathbf{x} \mid \mathbf{h}) \mid \mathbf{y}, \mathbf{h}^{[k]}\right] \quad (16) \\ \mathbf{h}^{[k+1]} &= \arg\max_{\mathbf{h}} Q(\mathbf{h} \mid \mathbf{h}^{[k]}) \quad (17) \end{aligned}$$

To calculate (16), the conditional distribution of the complete data given the unknown parameter, $p(\mathbf{x} \mid \mathbf{h})$ is required. This distribution derived in Appendix I, is a multivariate Gaussian distribution. The mean vector $\mathbf{m}_\mathbf{x} = \left[\begin{array}{cc} \mathbf{0}^T & \mathbf{c}^T \mathbf{H}^T \end{array}\right]^T$, where $\mathbf{0}$ is a null vector of dimension N and the covariance matrix $\mathbf{R}_{\mathbf{xx}}$ is

$$\mathbf{R}_{\mathbf{xx}} = \left[\begin{array}{cc} \sigma_s^2 \mathbf{I}_N & \sigma_s^2 \mathbf{H}^T \\ \sigma_s^2 \mathbf{H} & \sigma_n^2 \mathbf{I}_M + \sigma_s^2 \mathbf{H}\mathbf{H}^T \end{array}\right] \tag{18}$$

with

$$\mathbf{R}_{\mathbf{xx}}^{-1} = \left[\begin{array}{cc} \frac{1}{\sigma_s^2} \mathbf{I}_N + \frac{1}{\sigma_n^2} \mathbf{H}^T \mathbf{H} & -\frac{1}{\sigma_n^2} \mathbf{H}^T \\ -\frac{1}{\sigma_n^2} \mathbf{H} & \frac{1}{\sigma_n^2} \mathbf{I}_M \end{array}\right]. \tag{19}$$

The logarithm of joint distribution of $\mathbf{y}$ and $\mathbf{s}$ given $\mathbf{h}$ is

$$\begin{aligned} \log p(\mathbf{x} \mid \mathbf{h}) = &-\frac{(M+N)}{2} \log 2\pi - \frac{1}{2} \log \det \mathbf{R}_{\mathbf{xx}} \\ &-\frac{1}{2} (\mathbf{x} - \mathbf{m}_\mathbf{x})^T \mathbf{R}_{\mathbf{xx}}^{-1} (\mathbf{x} - \mathbf{m}_\mathbf{x}). \end{aligned} \tag{20}$$

In the above expression the first term is independent of $\mathbf{H}$. Similarly $\det \mathbf{R}_{\mathbf{xx}} = \sigma_s^2 \sigma_n^2$, which is also independent of $\mathbf{H}$. Thus

$$\begin{aligned} \log p(\mathbf{s}, \mathbf{y} \mid \mathbf{h}) = K - \frac{1}{2\sigma_n^2} (&\mathbf{s}^T \mathbf{H}^T \mathbf{H} \mathbf{s} + \\ 2\mathbf{c}^T \mathbf{H}^T \mathbf{H} \mathbf{s} + \mathbf{c}^T \mathbf{H}^T \mathbf{H} \mathbf{c} &- 2\mathbf{y}^T \mathbf{H} \mathbf{s} - 2\mathbf{y}^T \mathbf{H} \mathbf{c}), \end{aligned} \tag{21}$$

where K contains the terms independent of $\mathbf{H}$. To find the conditional expectation given $\mathbf{y}$ and $\mathbf{h}^{[k]}$ in (16), the conditional probability density function of $\mathbf{s}$ given $\mathbf{y}$ and $\mathbf{h}^{[k]}$, $p(\mathbf{s} \mid \mathbf{y}, \mathbf{h}^{[k]})$

is required. It is the Gaussian distribution $\mathcal{N}(\mathbf{u}, \mathbf{B})$, with mean and variance shown in Appendix II to be

$$\mathbf{u}^{[k]} = \sigma_s^2 \left(\sigma_n^2 \mathbf{I}_N + \sigma_s^2 \mathbf{H}^{[k]^T} \mathbf{H}^{[k]}\right)^{-1} \mathbf{H}^{[k]^T} (\mathbf{y} - \mathbf{H}^{[k]}\mathbf{c})$$
$$\mathbf{B}^{[k]} = \left(\frac{1}{\sigma_s^2}\mathbf{I}_N + \frac{1}{\sigma_n^2}\mathbf{H}^{[k]^T}\mathbf{H}^{[k]}\right)^{-1}. \quad (22)$$

From (16) and (17), the combined iterative step to find the maximum likelihood estimate is

$$\mathbf{h}^{[k+1]} = \arg\max_{\mathbf{h}} E\left[\log p(\mathbf{s}, \mathbf{y} \mid \mathbf{h}) \mid \mathbf{y}, \mathbf{h}^{[k]}\right]. \quad (23)$$

This implies that $\mathbf{h}^{[k+1]}$ satisfies

$$E\left[\frac{\partial}{\partial h_i}\{\log p(\mathbf{s}, \mathbf{y} \mid \mathbf{h})\} \mid \mathbf{y}, \mathbf{h}^{[k]}\right] = 0,\ 0 \le i \le L-1. \quad (24)$$

Substituting (21) in (24) gives

$$E[\frac{1}{\sigma_n^2}(\mathbf{s}^T\mathbf{H}^T\mathbf{J}_i\mathbf{s} + \mathbf{c}^T\mathbf{H}^T\mathbf{J}_i\mathbf{s} + \mathbf{c}^T\mathbf{J}_i^T\mathbf{H}\mathbf{s} + \mathbf{c}^T\mathbf{H}^T\mathbf{J}_i\mathbf{c}$$
$$-\mathbf{y}^T\mathbf{J}_i\mathbf{s} - \mathbf{y}^T\mathbf{J}_i\mathbf{c}) \mid \mathbf{y}, \mathbf{h}^{[k]}] = 0,\ 0 \le i \le L-1. \quad (25)$$

Let $\mathbf{u}^{[k]}$ and $\mathbf{B}^{[k]}$ be the mean and covariance of $p(\mathbf{s} \mid \mathbf{y}, \mathbf{h})$ when $\mathbf{h} = \mathbf{h}^{[k]}$ and define the matrix $\mathbf{P}^{[k]} = \mathbf{B}^{[k]} + \mathbf{u}^{[k]}\mathbf{u}^{[k]^T}$. Then (25) can be rewritten

$$\mathrm{tr}(\mathbf{H}^T\mathbf{J}_i\mathbf{P}^{[k]}) + \mathbf{c}^T\mathbf{H}^T\mathbf{J}_i\mathbf{u}^{[k]} + \mathbf{c}^T\mathbf{J}_i^T\mathbf{H}\mathbf{u}^{[k]} + \mathbf{c}^T\mathbf{H}^T\mathbf{J}_i\mathbf{c}$$
$$-\mathbf{y}^T\mathbf{J}_i\mathbf{u}^{[k]} - \mathbf{y}^T\mathbf{J}_i\mathbf{c} = 0,\ 0 \le i \le L-1. \quad (26)$$

Substituting $\mathbf{H} = \sum_{j=0}^{L-1} h_j\mathbf{J}_j$ in (26) gives

$$\sum_{j=0}^{L-1} h_j\left[\mathrm{tr}(\mathbf{J}_j^T\mathbf{J}_i\mathbf{P}^{[k]}) + \mathbf{c}^T\mathbf{J}_j^T\mathbf{J}_i\mathbf{u}^{[k]} + \mathbf{c}^T\mathbf{J}_i^T\mathbf{J}_j\mathbf{u}^{[k]} + \mathbf{c}^T\mathbf{J}_j^T\mathbf{J}_i\mathbf{c}\right]$$
$$= \mathbf{y}^T\mathbf{J}_i\mathbf{u}^{[k]} + \mathbf{y}^T\mathbf{J}_i\mathbf{c},\ 0 \le i \le L-1. \quad (27)$$

This is a set of linear equations of the form

$$\mathbf{D}\mathbf{h} = \mathbf{d} \quad (28)$$

where the matrix $\mathbf{D} \in \mathbb{R}^{L\times L}$ and vector $\mathbf{d} \in \mathbb{R}^L$ are given by

$$\mathbf{D}_{ij} = \mathrm{tr}(\mathbf{J}_j^T\mathbf{J}_i\mathbf{P}^{[k]}) + \mathbf{c}^T\mathbf{J}_j^T\mathbf{J}_i\mathbf{u}^{[k]} + \mathbf{c}^T\mathbf{J}_i^T\mathbf{J}_j\mathbf{u}^{[k]} + \mathbf{c}^T\mathbf{J}_j^T\mathbf{J}_i\mathbf{c},$$

and

$$d_i = \mathbf{y}^T\mathbf{J}_i\mathbf{u}^{[k]} + \mathbf{y}^T\mathbf{J}_i\mathbf{c},\ 0 \le i, j \le L-1. \quad (29)$$

The EM algorithm for estimating the channel coefficients $\mathbf{h}$ can be summarized as follows. Obtain the constant matrices $\mathbf{J}_i = \partial\mathbf{H}/\partial h_i, i = 0, \ldots L-1$ and initialize the channel vector $\mathbf{h}^{[0]}$. At an iteration k (k = 0, 1, ...) use (22) to calculate $\mathbf{u}^{[k]}$ and $\mathbf{B}^{[k]}$. Determine $\mathbf{P}^{[k]}$ from the relation $\mathbf{P}^{[k]} = \mathbf{B}^{[k]} + \mathbf{u}^{[k]}\mathbf{u}^{[k]^T}$. Use (29) to calculate $\mathbf{d}$ and $\mathbf{D}$. Obtain the new estimate of the channel $\mathbf{h}^{[k+1]}$ by solving (28).

D. *Incorporating Memory to EM*

The standard EM algorithm, as described above, finds the ML estimate of the channel given a single block of data. However, if the block is short, then the significant uncertainty due to the data symbols can make even the optimal estimate poor. Moreover, if the channel varies slowly compared to the block length, then previous blocks provide significant information about the current channel. These observations motivate a heuristic modification to the EM algorithm.

Each step of the EM algorithm produces a channel estimate which reflects both the observation, y, and the initial estimate, $h^{[k]}$. If the initial estimate, $h^{[0]}$, is taken to be the estimate from the previous block, and only a small number of update steps are performed, then the correlation of the channel can be exploited; averaging is performed over a larger number of data symbols, resulting in a better channel estimate. The heuristic EM algorithm is then simply to start with an initial channel estimate $h^{[0]}(0)$, and for each block, i, perform a single update to obtain $h^{[1]}(i)$ from $h^{[0]}(i)$, and set $h^{[0]}(i+1) = h^{[1]}(i)$.

V. Simulation

BER results of the proposed algorithms are compared with those of a simple averaging scheme and the Least Mean Squares (LMS) algorithm. The channel is estimated using the given schemes and a minimum mean square error (MMSE) equalizer [15] is constructed for equalization. The channel is modeled as a time-varying channel using Jakes' model [16] with a Doppler frequency $f_d = 400$Hz and with a transmission rate $f_s = 4$Mbps. This Doppler frequency corresponds to a speed of 60km/hr. A baseband channel is considered, and so only the real part of Jakes' channel model is used. The length of the channel used is $L = 3$. The transmitted symbols are from a binary phase shift keying (BPSK) constellation. For all schemes initial training is performed with 8 pilot symbols followed by the superimposed training. The superimposed pilots are chosen as random binary data. For all schemes excluding EM with short symbol blocks, twenty percent of the power is allocated for pilots.

The RLS algorithm uses a forgetting factor, $\lambda = 0.996$ and the scaling factor is calculated using (9).

The simple averaging scheme is as follows. Treating transmitted information symbols as noise, the channel model (13) can be represented as

$$\mathbf{y} = \mathbf{H}\mathbf{c} + \mathbf{n}, \quad (30)$$

where $\mathbf{n}$ has both the channel noise and information symbol noise. Considering the specific structure of the channel matrix given in (14), (30) can be rewritten as

$$\mathbf{y} = \mathbf{C}\mathbf{h} + \mathbf{n}, \quad (31)$$

where $\mathbf{C} \in \mathbb{R}^{M\times L}$ is constructed from $\mathbf{c}$. The channel is estimated as $\hat{\mathbf{h}} = (1/M)\sum_{i=1}^{M}\mathbf{y}_i\mathbf{C}_i^\dagger$, where M is the number of symbol blocks transmitted and i is the index of the the ith block.

Considering the system model given in (31), the LMS algorithm estimates the channel recursively as

$$\hat{\mathbf{h}}_i = \lambda\hat{\mathbf{h}}_{i-1} + (1-\lambda)\mathbf{y}\mathbf{C}^\dagger. \quad (32)$$

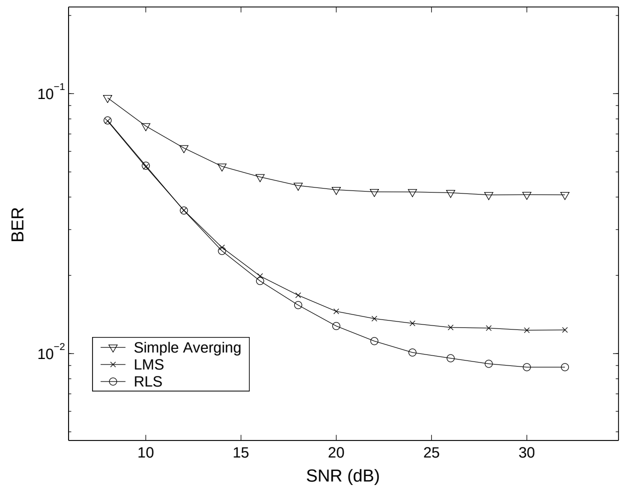

Fig. 1. BER performance comparison of RLS algorithm, LMS algorithm and Simple averaging scheme

The value of the $\hat{\mathbf{h}}$ is set to the channel estimate from initial training, and $\lambda = 0.996$.

Equalization is done using an MMSE equalizer with a tap length of 7. All the BER results are averaged over 100 realizations of the channel and for 300 transmitted symbols in each realization. The BER comparison is presented in Figure 1. The results show that the RLS algorithm performs better than the other schemes. It outperforms simple averaging due to the forgetting factor, λ. Since the LMS algorithm uses the same forgetting factor, it performs similarly to RLS at low SNR. However, since it converges more slowly, its BER is dominated by the high error rate before it converges; this gives rise to its worse performance at high SNR. The simple averaging scheme performs worst because it does not consider the channel variation. Figure 2 presents the convergence of the RLS algorithm with and without initial training. It can be seen that the initial training reduces convergence time and this helps the use of RLS algorithm from the beginning of data transmission.

The EM algorithm with and without memory uses a block length of 20 and fifty percent of the power is allocated to the pilot symbols. The stanadard EM algorithm iterates over a single block 200 times, whereas EM with memory does only 1 iteration. Figure 3 presents the comparison of the standard EM algorithm and the EM algorithm with memory. EM with memory significantly outperforms the standard EM algorithm. Figure 4 shows the BER performance comparison of the EM algorithm (without incorporating the strategy in Section IV-D) and the RLS algorithm for a block length of 400 symbols. The channel is assumed to be time-invariant over the block. EM is showing a better peformance towards higher SNR. This result indicates that EM needs more information about the channel for improved channel estimation.

All the schemes presented have a high error floor at high SNR. This is because all these schemes treat the information symbols as noise. Thus even if the channel SNR is infinite there will be an error in estimating the channel, which causes non-zero BER.

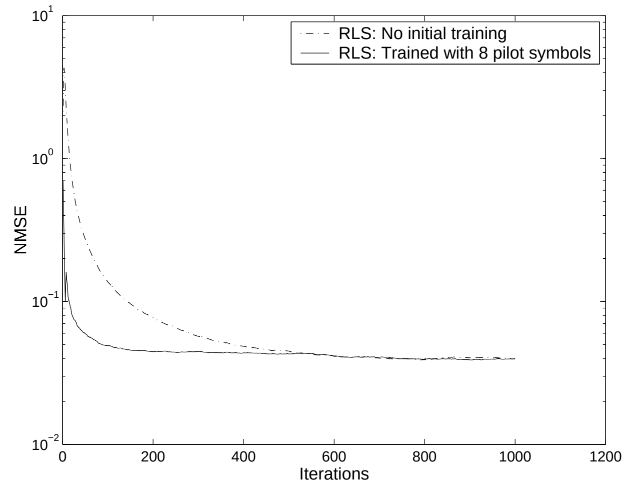

Fig. 2. Convergence of RLS algorithm with and without initial training at 10dB SNR.

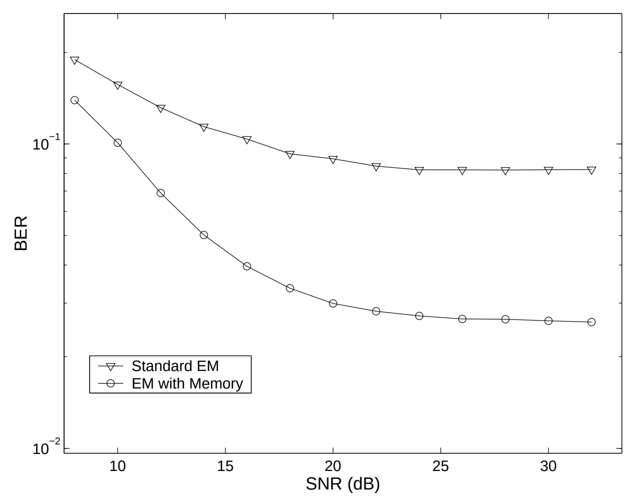

Fig. 3. BER performance comparison of standard EM algorithm and EM algorithm with memory.

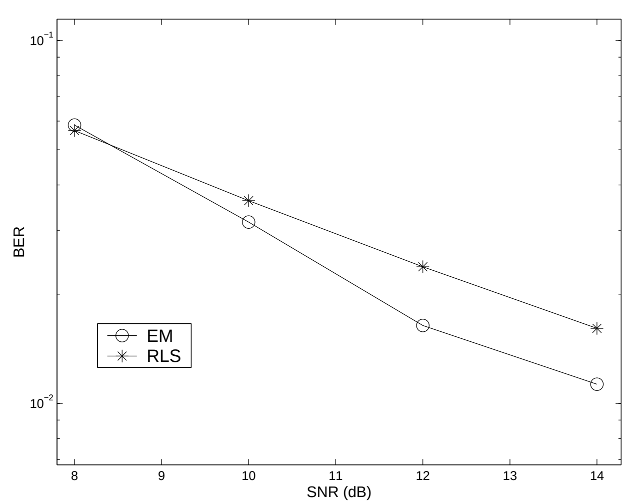

Fig. 4. BER performance comparison of standard EM with RLS for a block lenth of 400.

VI. Conclusion

The RLS algorithm has been shown to outperform a simple (time-invariant) averaging scheme and at high SNR it also outperforms LMS. The heuristic EM algorithm which incorporates memory from block to block greatly improves it's performance. EM shows a better performance than the RLS algorithm for large block sizes.

Appendix I

The distribution $P(\mathbf{y} \mid \mathbf{h})$ is Gaussian with mean and covariance calculated from (13) as

$$\begin{aligned} E[\mathbf{y} \mid \mathbf{h}] &= \mathbf{Hc} \\ \text{cov}[\mathbf{y} \mid \mathbf{h}] &= E\left[(\mathbf{Hs}+\mathbf{n})(\mathbf{Hs}+\mathbf{n})^T \mid \mathbf{h}\right] \\ &= E\left[\mathbf{Hss}^T\mathbf{H}^T + \mathbf{Hsn}^T + \mathbf{ns}^T\mathbf{H}^T + \mathbf{nn}^T \mid \mathbf{h}\right] \\ &= \sigma_s^2\mathbf{HH}^T + \sigma_n^2\mathbf{I}_M \end{aligned}$$

since $\mathbf{H}$ and $\mathbf{s}$ are independent. Moreover $p(\mathbf{s} \mid \mathbf{h}) = p(\mathbf{s})$ is Gaussian with $E[\mathbf{s}] = \mathbf{0}$ and $\text{cov}[\mathbf{s}] = \sigma_s^2\mathbf{I}_N$. Moreover

$$\begin{aligned} \text{cov}(\mathbf{y},\mathbf{s} \mid \mathbf{h}) &= E[(\mathbf{y} - E(\mathbf{y} \mid \mathbf{h}))(\mathbf{s} - E(\mathbf{s} \mid \mathbf{h}))^T \mid \mathbf{h}] \\ &= E[(\mathbf{Hs} + \mathbf{n})\mathbf{s}^T \mid \mathbf{h}] = \sigma_s^2\mathbf{H} \end{aligned}$$

$$\begin{aligned} \mathbf{R_{xx}} = \text{cov}(\mathbf{x},\mathbf{x} \mid \mathbf{h}) &= \begin{bmatrix} \text{cov}(\mathbf{s},\mathbf{s} \mid \mathbf{h}) & \text{cov}(\mathbf{s},\mathbf{y} \mid \mathbf{h}) \\ \text{cov}(\mathbf{y},\mathbf{s} \mid \mathbf{h}) & \text{cov}(\mathbf{y},\mathbf{y} \mid \mathbf{h}) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_s^2\mathbf{I}_N & \sigma_s^2\mathbf{H}^T \\ \sigma_s^2\mathbf{H} & \sigma_n^2\mathbf{I}_M + \sigma_s^2\mathbf{HH}^T \end{bmatrix} \end{aligned}$$

Appendix II

The conditional distribution $p\left(\mathbf{s} \mid \mathbf{y}, \mathbf{h}^{[k]}\right)$ can be obtained from (13) as follows:

$$p(\mathbf{s},\mathbf{n} \mid \mathbf{h}^{[k]},\mathbf{y}) = \frac{p(\mathbf{s})p(\mathbf{n})p(\mathbf{h}^{[k]} \mid \mathbf{n},\mathbf{s})p(\mathbf{y} \mid \mathbf{n},\mathbf{s},\mathbf{h}^{[k]})}{p(\mathbf{y},\mathbf{h}^{[k]})}$$

$$= \frac{p(\mathbf{s})p(\mathbf{n})p(\mathbf{h}^{[k]} \mid \mathbf{n},\mathbf{s})\delta(\mathbf{y} - \mathbf{H}^{[k]}\mathbf{s} - \mathbf{H}^{[k]}\mathbf{c} - \mathbf{n})}{p(\mathbf{y},\mathbf{h}^{[k]})}$$

$$\int p(\mathbf{s},\mathbf{n} \mid \mathbf{h}^{[k]},\mathbf{y})d\mathbf{n}$$

$$= \int \frac{p(\mathbf{s})p(\mathbf{n})p(\mathbf{h}^{[k]} \mid \mathbf{n},\mathbf{s})\delta(\mathbf{y} - \mathbf{H}^{[k]}\mathbf{s} - \mathbf{H}^{[k]}\mathbf{c} - \mathbf{n})}{p(\mathbf{y},\mathbf{h}^{[k]})}d\mathbf{n}$$

$$p(\mathbf{s} \mid \mathbf{h}^{[k]},\mathbf{y}) = p(\mathbf{s})p(\mathbf{y} - \mathbf{H}^{[k]}\mathbf{s} - \mathbf{H}^{[k]}\mathbf{c})K, \tag{33}$$

where K includes all the factors independent of $\mathbf{s}$. Denote $\mathbf{y} - \mathbf{H}^{[k]}\mathbf{c}$ by $\mathbf{y}'$. Equation (33) is expanded as

$$p(\mathbf{s} \mid \mathbf{h}^{[k]},\mathbf{y}) =$$
$$K' \exp\left(\tfrac{-1}{2\sigma_s^2\sigma_n^2}[\sigma_n^2\mathbf{s}^T\mathbf{s} + \sigma_s^2(\mathbf{y}' - \mathbf{H}^{[k]}\mathbf{s})^T(\mathbf{y}' - \mathbf{H}^{[k]}\mathbf{s})]\right),$$

where K' contains the terms independent of $\mathbf{s}$. Also

$$\sigma_n^2\mathbf{s}^T\mathbf{s} + \sigma_s^2(\mathbf{y}' - \mathbf{H}^{[k]}\mathbf{s})^T(\mathbf{y}' - \mathbf{H}^{[k]}\mathbf{s})$$
$$= \sigma_n^2\mathbf{s}^T\mathbf{s} + \sigma_s^2\mathbf{y}'^T\mathbf{y}' - 2\sigma_s^2\mathbf{y}'^T\mathbf{H}^{[k]}\mathbf{s} + \sigma_s^2\mathbf{s}^T\mathbf{H}^{[k]^T}\mathbf{H}^{[k]}\mathbf{s}$$
$$= (\mathbf{s} - \sigma_s^2\mathbf{A}^{-1}\mathbf{H}^{[k]^T}\mathbf{y}')^T\mathbf{A}(\mathbf{s} - \sigma_s^2\mathbf{A}^{-1}\mathbf{H}^{[k]^T}\mathbf{y}') - k',$$

where $\mathbf{A} = [\sigma_n^2\mathbf{I} + \sigma_s^2\mathbf{H}^{[k]^T}\mathbf{H}^{[k]}]$ and k' is terms independent of $\mathbf{s}$. Conditional distribution of transmitted symbols $\mathbf{s}$ given $\mathbf{h}^{[k]}$ and $\mathbf{y}$ is

$$p(\mathbf{s} \mid \mathbf{h}^{[k]},\mathbf{y}) =$$
$$K'' \exp\left(\tfrac{-1}{2\sigma_s^2\sigma_n^2}(\mathbf{s} - \sigma_s^2\mathbf{A}^{-1}\mathbf{H}^{[k]^T}\mathbf{y}')^T\mathbf{A}(\mathbf{s} - \sigma_s^2\mathbf{A}^{-1}\mathbf{H}^{[k]^T}\mathbf{y}')\right),$$

where K'' is terms independent of $\mathbf{s}$ and from this equation mean vector and covariance matrix are obtained as

$$\mathbf{u}^{[k]} = E(\mathbf{s} \mid \mathbf{y},\mathbf{h}^{[k]}) =$$
$$\sigma_s^2\left(\sigma_n^2\mathbf{I}_N + \sigma_s^2\mathbf{H}^{[k]^T}\mathbf{H}^{[k]}\right)^{-1}\mathbf{H}^{[k]^T}(\mathbf{y} - \mathbf{H}^{[k]}\mathbf{c})$$
$$\mathbf{B}^{[k]} = \text{cov}(\mathbf{s},\mathbf{s} \mid \mathbf{y},\mathbf{h}^{[k]}) = \left(\tfrac{1}{\sigma_s^2}\mathbf{I}_N + \tfrac{1}{\sigma_n^2}\mathbf{H}^{[k]^T}\mathbf{H}^{[k]}\right)^{-1}.$$

Acknowledgement

This work is supported by Australian Research Council (ARC).

References

[1] Franco Mazzenga, "Channel estimation an equalization for M-QAM transmission with a hidden pilot sequence," *IEEE Transactions on Broadcasting.*, vol. 46, no. 2, pp. 170–176, June 2000.

[2] G. Tong Zhou, M. Viberg, T. McKelvey, "A first-order statistical method for channel estimation," *IEEE Signal Processing Letters.*, vol. 10, no. 3, pp. 57–60, March 2003.

[3] D.H. Pham, J.H. Manton, "A first order statistical method for guard interval based channel identification and source recovery," *Proceedings of Fourth International Conference on Information, Communication and Signal Processing, Singapore.*, December 2003.

[4] J.K. Tugnait, W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Communication Letters.*, vol. 7, no. 9, pp. 413–415, September 2003

[5] X. Meng, J.K. Tugnait, "Semi-blind time-varying channel estimation using superimposed training," *Proceedings of ICASSP.*, 2004.

[6] X. Meng, J.K. Tugnait, "Semi-blind channel estimation and detection using superimposed training," *Proceedings of ICASSP.*, 2004.

[7] B. Farhang-Boroujeny, "Experimental study of semi-blind channel indetification/equalization through pilot signals," *Proceedings of ICSP.*, 1996.

[8] X. Meng, J.K. Tugnait, "MIMO channel estimation suing superimposed training," *Proceedings of ICC.*, 2004.

[9] H. Zhu, B. Farhang-Boroujey, C. Schegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication system," *IEEE Communication Letters.*, vol. 7, no. 1, pp. 30–32, January 2003.

[10] D.H. Pham, J.H. Manton, "An optimal pilot embedding scheme for channel identification and source recovery ," *Proceeedings of IEEE International Conference on Sympossium on Spread Spectrum Techniques and Applications.* , pp. 77–81, August 2004

[11] N. Chen, G.T. Zhou, "What is the price payed for superimposed training in OFDM? ," *Proceeedings of ICASSP.* 2004

[12] N. Chen, G.T. Zhou, "A superimposed periodic pilot scheme for semi-blind channel estimation of OFDM systems," *Proceeedings of Tenth IEEE DSP Workshop.* pp. 362–365, Pine Mountain GA, October 2003.

[13] S. Haykin, "Adaptive Filter Theory," Fourth Edition, Prentice Hall, Upper Saddle River, NJ 2002.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum likelihood from incomplete data via EM algorithm," *J.Roy.Statistic. Soc.* B, pp. 39:1–38, 1977.

[15] J. G. Proakis, "Digital Communications," Fourth Edition, McGrawHill 2001.

[16] W. C. Jakes," Microwave Mobile Communications. New York: Wiley", 1974.