TRAINING BASED MAXIMUM LIKELIHOOD CHANNEL IDENTIFICATION*

Olivier Rousseaux[1], *Geert Leus*[1†], *Petre Stoica*[2] *and Marc Moonen*[1]

[1] K.U.Leuven - ESAT
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
Email: {orousso, leus, moonen}@esat.kuleuven.ac.be
[2] Uppsala University - Department of Systems and Control
P.O. Box 337 SE-751 05, Uppsala Sweden
Email: ps@syscon.uu.se

ABSTRACT

In this paper, we address the problem of identifying convolutive channels in a Maximum Likelihood (ML) fashion when a constant training sequences is periodically inserted in the transmitted signal. We consider the case where the channel is quasi-static (i.e. the sampling period is several orders of magnitude below the coherence time of the channel). There are no requirements on the length of the training sequence and all the received symbols that contain contributions from the training symbols are exploited. We first propose an iterative method that converges to the ML estimate of the channel. We then derive a closed form expression of the ML channel estimate.

1. INTRODUCTION

The market need for high data rate communications has driven the research in telecommunications during the last years. In order to increase data rates when transmitting data over wireless channels, it is often needed to use broadband communication systems. A major impediment of such systems is that the sampling period can get smaller than the delay spread of the channel, especially in multipath scenarios. This results in ISI (Inter Symbol Interference), a phenomenon that needs to be combatted at the receiver in order to restore the transmitted information. This is usually done using serial or block equalization techniques. Channel State Information (CSI) is needed at the receiver in order to design the equalizer and combat the ISI in an efficient way.

The CSI is obtained through the use of channel identification algorithms. These can be divided in two families: blind or training-based. The blind algorithms estimate the channel based on properties of the transmitted signals (finite alphabet properties, higher order statistics, ...) Training-based techniques assume that known symbols (training sequences or pilot symbols) are inserted in the transmitted signals. It is then possible to identify the channel at the receiver exploiting the knowledge of these training sequences.

In this paper, we focus on the family of training-based channel identification algorithms. The problem of optimally identifying the channel at the receiver when training sequences are inserted in the transmitted signals has been widely discussed in the literature (see e. g. [1]). When the training sequences are long enough, some of the received symbols only contain contributions from the known training sequences (as opposed to the received symbols that contain contributions from both the unknown data symbols and the training symbols or from the data symbols only). The problem of performing ML channel identification when only these received symbols are used is equivalent to a least squares problem. However, this approach is sub-optimal since not all the received symbols that contain contributions from the training sequences are considered.

We present here a new method that does not require a minimal length for the training sequences and performs ML channel estimation exploiting all the received symbols that contain contributions from the training sequences. We consider a transmission scheme where a constant training sequence is repeatedly inserted between blocks of data symbols. We consider quasi-static channels (the channel stays constant during the transmission of several blocks of data). The case where the training sequence changes from block to block is discussed elsewhere.

Note that there are several existing transmission schemes to which this new channel estimation method can be applied. Examples are for instance Known Symbol Padding (KSP) block transmission [2], [3] or Pilot Symbol Assisted Modulation (PSAM) [4].

Notation: We use upper (lower) case bold face letters to denote matrices (column vectors). $\mathbf{I}_N$ is the identity matrix of size $N \times N$ and $\mathbf{O}_{M\times N}$ is the all-zero matrix of size $M \times N$. The operator $(.)^*$ denotes the complex conjugate, $(.)^T$ the transpose of a matrix and $(.)^H$ its complex conjugate transpose. Finally, $tr(\mathbf{A})$ denotes the trace of a matrix $\mathbf{A}$, and $|\mathbf{A}|$ its determinant.

* This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Inter-university Poles of Attraction Programme (2002-2007) - IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modeling') and P5/11 ('Mobile multimedia communication systems and networks'), the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government, Research Project FWO nr.G.0196.02 ('Design of efficient communication techniques for wireless time-dispersive multi-user MIMO systems'). The scientific responsibility is assumed by its authors.

† Postdoctoral Fellow of the F.W.O. Vlaanderen.

2. DATA MODEL

We consider stationary Finite Impulse Response (FIR) convolutive channels of order L: $\mathbf{h} = [h[0], h[1], \cdots, h[L]]^T$. A sequence $x[n]$ of symbols is transmitted over the channel. The received sequence $y[n]$ is the linear convolution of the transmitted sequence with the channel impulse response:

$$y[n] = \sum_{i=0}^{L} h[i]x[n-i] + \eta[n],$$

where $\eta[n]$ is the Additive White Gaussian Noise (AWGN) at the receiver.

As mentioned in the introduction, we consider a transmission scheme where a constant training sequence $\mathbf{t} = [t[1], \ldots, t[n_t]]^T$ is inserted between blocks of data symbols. A total number of K such training sequences is inserted, the kth training sequence starts at position n_k: $[x[n_k], \ldots, x[n_k + n_t - 1]]^T = \mathbf{t}$.

Define the vector $\mathbf{u}_k$ of received symbols that contain a contribution from the kth transmitted training sequence:

$$\mathbf{u}_k = [y[n_k], \ldots, y[n_k + n_t + L - 1]]^T .$$

The vector $\mathbf{u}_k$ contains a contribution from the training sequence $\mathbf{t}$ plus an additional term that collects the contributions from both the unknown surrounding data symbols and the noise. We can thus describe $\mathbf{u}_k$ as the sum of a deterministic and a stochastic term:

$$\mathbf{u}_k = \mathbf{T}\mathbf{h} + \boldsymbol{\epsilon}_k, \tag{1}$$

where $\mathbf{T}$ is an $(n_t + L) \times (L + 1)$ tall Toeplitz matrix with $\left[\mathbf{t}^T, 0, \ldots, 0\right]^T$ as first column and $[t[1], 0, \ldots, 0]$ as first row. $\mathbf{T}\mathbf{h}$ is the deterministic term; the stochastic term, $\boldsymbol{\epsilon}_k$, is described as follows:

$$\boldsymbol{\epsilon}_k = \underbrace{\left[\begin{array}{ccc|ccc} h_L & \cdots & h_1 & & & \mathbf{0} \\ & \ddots & \vdots & h_0 & & \\ & & h_L & \vdots & \ddots & \\ \mathbf{0} & & & h_{L-1} & \cdots & h_0 \end{array}\right]}_{\mathbf{H}_s (n_t + L) \times (2L)} \mathbf{s}_k + \boldsymbol{\eta}_k, \tag{2}$$

where $\mathbf{s}_k = [x[n_k - L], \ldots, x[n_k - 1], x[n_k + n_t], \ldots, x[n_k + n_t + L]]^T$ is the vector of surrounding data symbols, and $\boldsymbol{\eta}_k$ is the corresponding AWGN term. Assuming that both the noise and the data are white and zero-mean ($E\{s[n]s^*[l]\} = E\{\eta[n]\eta^*[l]\} = 0, \ \forall n \neq l$, and $E\{s[n]\} = E\{\eta[n]\} = 0$), we can say that $\boldsymbol{\epsilon}_k$ is zero-mean. Defining the signal and noise variances as $\lambda^2 = E\{s[n]s^*[n]\}$ and $\sigma^2 = E\{\eta[n]\eta^*[n]\}$ respectively, we can derive the covariance matrix of $\boldsymbol{\epsilon}_k$ from (2) as $E\left\{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H\right\} \triangleq \mathbf{Q} = \lambda^2 \mathbf{H}_s \mathbf{H}_s^H + \sigma^2 \mathbf{I}$. If we additionally assume that there is a sufficient number of data symbols between two successive training sequences, (i.e. defining n_s as the length of the shortest sequence of data symbols, we assume $n_s \geqslant 2L$) we can say that $\boldsymbol{\epsilon}_k$ is white.

3. MAXIMUM LIKELIHOOD CHANNEL ESTIMATION

The noise part of $\boldsymbol{\epsilon}_k$ can be considered as purely Gaussian (which corresponds to the classical AWGN approximation), but this is not the case for the data part. However, we can apply the Gaussian likelihood function as a statistically sound fitting criterion. Relying on this approximation, we express (up to a constant term) the log-likelihood function of the system as:

$$-\mathcal{L} = \frac{K}{2} ln|\mathbf{Q}| + \frac{1}{2} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H \mathbf{Q}^{-1} (\mathbf{u}_k - \mathbf{T}\mathbf{h}). \tag{3}$$

Relying on the definition of $\mathbf{Q}$, the log-likelihood can be expressed as a direct function of the unknown parameters $\mathbf{h}$ and σ^2. This expression is quite complex and the minimization problem boils down to an $(L+1)$-dimensional nonlinear search.

To overcome this problem, we propose to disregard the structure of $\mathbf{Q}$, and ignore the relation that binds it to the parameters $\mathbf{h}$ and σ^2. We thus assume that the covariance matrix $\mathbf{Q}$ of the stochastic term $\boldsymbol{\epsilon}_k$ can be any symmetric positive definite matrix, regardless of $\mathbf{h}$ and σ^2. These assumptions (unstructured $\mathbf{Q}$ and gaussian $\boldsymbol{\epsilon}_k$) transform the initial ML problem into an optimization problem which is separable in its two variables $\mathbf{Q}$ and $\mathbf{h}$. We exploit this separation property in the next paragraphs in order to solve the ML problem in a less complex way than the $(L+1)$-dimensional nonlinear search.

4. ITERATIVE SOLUTION

When a minimization problem is separable in its variables, a solution that is commonly used is an iterative one. Each iteration consists in analytically minimizing the cost function with respect to one variable whilst keeping the other(s) fixed. The variable with respect to which the cost function is minimized is changed in each iteration. This procedure converges to a minimum of the cost function. If the starting point is accurate enough or if the surface is smooth enough, the point of convergence is the global minimum of the cost function, which corresponds to the ML estimate of $\mathbf{Q}$ and $\mathbf{h}$ in this case.

Assume that at the ith iteration an estimate $\hat{\mathbf{Q}}^i$ of the covariance matrix $\mathbf{Q}$ is available. We first seek the channel estimate $\hat{\mathbf{h}}^i$ that minimizes the cost function (3) with respect to $\mathbf{h}$ under the hypothesis that the available $\hat{\mathbf{Q}}^i$ is the true $\mathbf{Q}$, i.e. we compute $\hat{\mathbf{h}}^i = \mathbf{h}_{ML}(\mathbf{Q}^i)$, where $\mathbf{h}_{ML}(\mathbf{Q}) = \arg\min_{\mathbf{h}} -\mathcal{L}$, whose solution can be computed as:

$$\mathbf{h}_{ML}(\mathbf{Q}) = K^{-1} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \sum_{k=1}^{K} \mathbf{T}^H \mathbf{Q}^{-1} \mathbf{u}_k. \tag{4}$$

We then seek the covariance matrix $\hat{\mathbf{Q}}^{i+1}$ that minimizes (3) under the hypothesis that this new channel estimate $\hat{\mathbf{h}}^i$ is the true $\mathbf{h}$: $\hat{\mathbf{Q}}^{i+1} = \mathbf{Q}_{ML}(\hat{\mathbf{h}}^i)$, where $\mathbf{Q}_{ML}(\mathbf{h}) = \arg\min_{\mathbf{Q}} -\mathcal{L}$, whose solution can be computed as [5, pp. 200-202]:

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H. \tag{5}$$

$\hat{\mathbf{Q}}^{i+1}$ is then used as a starting point for the next iteration. The procedure is stopped when there is no significant difference between the estimates produced by two consecutive iterations. We still have to find an acceptable starting point for the iterations. In [6], we proposed an iterative method for channel identification in a similar context. The method proposed in that paper can be summarized as follows: when a noise covariance matrix estimate $\hat{\mathbf{Q}}$ is available, a weighted least squares channel estimate $\hat{\mathbf{h}}$ is derived. This

weighted least squares channel estimate is the same as the one we obtain by injecting $\hat{\mathbf{Q}}$ into (4): $\hat{\mathbf{Q}} = \mathbf{Q}_{ML}(\hat{\mathbf{h}})$. When a channel estimate $\hat{\mathbf{h}}$ is available, the noise covariance matrix estimate $\hat{\mathbf{Q}}$ is derived by averaging the K sample covariance matrices of the estimated noise. This estimate $\hat{\mathbf{Q}}$ is exactly equivalent to what we would obtain by injecting $\hat{\mathbf{h}}$ into (5):$\hat{\mathbf{h}} = \mathbf{h}_{ML}(\hat{\mathbf{Q}})$. The iterative method of [6] is initialized with a simple least squares channel estimate. Experimental results show that this initial channel estimate is good enough to converge to an accurate channel estimate. The experiments show as well that the method almost converges in one iteration. The improvements brought by the following iterations are marginal. Applying the method we propose here with an identity matrix as initial noise covariance matrix $\hat{\mathbf{Q}}^0 = \mathbf{I}$ yields exactly the same path through the iterations as in [6]. This discussion also shows that the method we proposed in [6] is indeed an iterative maximum likelihood one.

5. CLOSED FORM SOLUTION

An alternative strategy to the iterative procedure described above consists in directly finding an analytical expression for the global minimum of the likelihood function. The separation property of the cost function can be exploited again in order to find this global minimum. The idea is to analytically minimize the cost function with respect to one variable. This minimum is a function of the other variable. The first variable can be replaced by that function in the original cost function, which then becomes a single variable expression. Minimizing this new expression of the cost function with respect to the only variable left yields the global minimum.

We first minimize the cost function with respect to $\mathbf{Q}$, the solution of which is given by (5). Next, observe that the likelihood function (3) can be expressed as

$$-\mathcal{L} = \frac{K}{2} ln|\mathbf{Q}| + \frac{1}{2} tr\left(\mathbf{Q}^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h})(\mathbf{u}_k - \mathbf{T}\mathbf{h})^H\right).$$

The ML channel estimate $\mathbf{h}_{ML}$ is obtained replacing $\mathbf{Q}$ by $\mathbf{Q}_{ML}(\mathbf{h})$ in this expression of the likelihood function and minimizing the resulting expression with respect to $\mathbf{h}$:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} \left| K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H \right|. \quad (6)$$

Define the following:

$$\begin{aligned} \hat{\mathbf{R}} &\triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^H, \\ \bar{\mathbf{u}} &\triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k, \\ \hat{\mathbf{Q}} &\triangleq \hat{\mathbf{R}} - \bar{\mathbf{u}}\bar{\mathbf{u}}^H, \end{aligned} \quad (7)$$

where $\hat{\mathbf{Q}}$ is assumed to be positive definite[1]. Using these definitions, the ML problem (6) can be re-expressed as:

$$\begin{aligned} \mathbf{h}_{ML} &= \arg\min_{\mathbf{h}} \left|\hat{\mathbf{Q}}\right| \left|\mathbf{I} + \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H\right| \\ &= \arg\min_{\mathbf{h}} \left|\hat{\mathbf{Q}}\right| \left(1 + (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})\right). \end{aligned}$$

[1]We see from (7) that a necessary condition therefore is $K \geqslant n_t + L + 1$. When this condition is fulfilled, the randomness of the noise and the data ensures that $\hat{\mathbf{Q}}$ is a definite matrix with probability 1.

Since $\hat{\mathbf{Q}}$ is positive definite, $\left|\hat{\mathbf{Q}}\right|$ is positive. Additionally, this factor does not depend on $\mathbf{h}$. Our problem is thus equivalent to:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}).$$

The minimum is reached at

$$\mathbf{h}_{ML} = \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \mathbf{T}\right)^{-1} \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \bar{\mathbf{u}}\right). \quad (8)$$

This ML channel estimate is easy to compute as well as intuitively quite appealing for it shows that the ML channel estimate is simply a fit of $\mathbf{T}\mathbf{h}$ to $\bar{\mathbf{u}}$ in a weighted least squares sense.

Once the ML channel estimate is obtained, we can derive the corresponding ML covariance matrix estimate. First observe that, using the notations introduced in (7), the general expression (5) for the ML estimate of $\mathbf{Q}$ as a function of $\mathbf{h}$ can be rewritten as:

$$\begin{aligned} \mathbf{Q}_{ML}(\mathbf{h}) &= \hat{\mathbf{R}} - \bar{\mathbf{u}}\mathbf{h}^H\mathbf{T}^H - \mathbf{T}\mathbf{h}\bar{\mathbf{u}}^H + \mathbf{T}\mathbf{h}\mathbf{h}^H\mathbf{T}^H \\ &= \hat{\mathbf{Q}} + (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H. \end{aligned} \quad (9)$$

Hence, $\mathbf{Q}_{ML}$ is derived by inserting $\mathbf{h}_{ML}$ into this expression:

$$\mathbf{Q}_{ML} = \hat{\mathbf{Q}} + (\mathbf{T}\mathbf{h}_{ML} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h}_{ML} - \bar{\mathbf{u}})^H, \quad (10)$$

which differs from $\hat{\mathbf{Q}}$ by a rank-one term.

Remark: Observe the similarity between the closed form expression of the ML channel estimate $\mathbf{h}_{ML}$ (8) and the general expression of the ML channel estimate as a function of the covariance matrix $\mathbf{Q}$, $\mathbf{h}_{ML}(\mathbf{Q})$ (4). Using the definitions (7), (4) can indeed be rewritten as:

$$\mathbf{h}_{ML}(\mathbf{Q}) = \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \left(\mathbf{T}^H \mathbf{Q}^{-1} \bar{\mathbf{u}}\right).$$

This shows that (8) can be written as $\mathbf{h}_{ML} = \mathbf{h}_{ML}(\hat{\mathbf{Q}})$. On the other hand, when the ML estimate of the covariance matrix, $\mathbf{Q}_{ML}$, is fed into $\mathbf{h}_{ML}(\mathbf{Q})$, we should also obtain the true ML estimate of the channel, $\mathbf{h}_{ML}$. Performing the appropriate substitutions, it is indeed possible to check that

$$\mathbf{h}_{ML} = \mathbf{h}_{ML}(\hat{\mathbf{Q}}) = \mathbf{h}_{ML}(\mathbf{Q}_{ML}),$$

although we know from (10) that $\hat{\mathbf{Q}}$ and $\mathbf{Q}_{ML}$ differ. This shows that several estimates of the noise covariance matrix are acceptable in oder to derive the ML channel estimate $\mathbf{h}_{ML}$.

6. COMPARISON WITH THE ITERATIVE METHOD

As mentioned in [6], the iterative method seems to almost converge after a single iteration. In this section, we explain this by showing that the channel estimate obtained after one iteration is almost equal to the ML channel estimate.

Let us first detail the different steps of the iterative method. We initialize the algorithm with $\hat{\mathbf{Q}}^0 = \mathbf{I}$. We thus have:

$$\hat{\mathbf{h}}^0 = \mathbf{h}_{ML}(\mathbf{I}) = \left(\mathbf{T}^H\mathbf{T}\right)^{-1} \mathbf{T}^H \bar{\mathbf{u}}.$$

Exploiting (9), we can then write

$$\hat{\mathbf{Q}}^1 = \mathbf{Q}_{ML}(\hat{\mathbf{h}}^0) = \hat{\mathbf{Q}} + \left(\mathbf{T}\hat{\mathbf{h}}^0 - \bar{\mathbf{u}}\right) \left(\mathbf{T}\hat{\mathbf{h}}^0 - \bar{\mathbf{u}}\right)^H. \quad (11)$$

The channel estimate that is obtained after the first iteration is thus:

$$\hat{\mathbf{h}}^1 = \mathbf{h}_{ML}(\hat{\mathbf{Q}}^1).$$

Observing that $\hat{\mathbf{h}}^0$ is the least squares fit of $\mathbf{Th}$ to $\bar{\mathbf{u}}$, we may say that $\mathbf{T}\hat{\mathbf{h}}^0 - \bar{\mathbf{u}}$ is small. The second term of the right-hand side of (11) is a second-order function of this small term. We can thus safely neglect it and make the following approximation:

$$\hat{\mathbf{Q}}^1 \approx \hat{\mathbf{Q}}.$$

Using (8), we then observe that $\hat{\mathbf{h}}^1$ is close to $\mathbf{h}_{ML}$:

$$\hat{\mathbf{h}}^1 = \mathbf{h}_{ML}(\hat{\mathbf{Q}}^1) \approx \mathbf{h}_{ML}(\hat{\mathbf{Q}}) = \mathbf{h}_{ML}.$$

This explains why the iterative procedure almost converges in one iteration.

7. ASYMPTOTIC PROPERTIES

In this section, we study the asymptotic properties of the proposed ML channel estimator, i.e. the properties when the number of transmitted data blocks, K, is large. We first find out that the ML channel estimator $\mathbf{h}_{ML}$ is asymptotically unbiased. We next derive an analytical expression of its asymptotical variance, which is a good indicator of the accuracy of an unbiased estimator.

Let us first note that $\hat{\mathbf{Q}}$ can be rewritten as:

$$\begin{aligned}\hat{\mathbf{Q}} &= K^{-1}\sum_{k=1}^{K}\mathbf{u}_k\mathbf{u}_k^H - K^{-2}\sum_{k=1}^{K}\mathbf{u}_k\sum_{k=1}^{K}\mathbf{u}_k^H \\ &= K^{-1}\sum_{k=1}^{K}\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H - K^{-2}\sum_{i,j=1}^{K}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j^H.\end{aligned}$$

When discussing asymptotical properties, we replace $\hat{\mathbf{Q}}$ by its limit value, which is easy to derive:

$$\lim_{K\to\infty}\hat{\mathbf{Q}} = \mathbf{Q}.$$

Asymptotically, (8) becomes

$$\begin{aligned}\lim_{K\to\infty}\mathbf{h}_{ML} &= \lim_{K\to\infty}K^{-1}\sum_{k=1}^{K}(\mathbf{T}^H\mathbf{Q}^{-1}\mathbf{T})^{-1}\mathbf{T}^H\mathbf{Q}^{-1}\mathbf{u}_k \\ &= \mathbf{h} + \lim_{K\to\infty}K^{-1}\sum_{k=1}^{K}(\mathbf{T}^H\mathbf{Q}^{-1}\mathbf{T})^{-1}\mathbf{T}^H\mathbf{Q}^{-1}\boldsymbol{\epsilon}_k.\end{aligned}$$

If we remember that $\boldsymbol{\epsilon}_k$ is zero-mean we see that the ML estimator is asymptotically unbiased:

$$\lim_{K\to\infty}\mathbf{h}_{ML} = \mathbf{h}.$$

Exploiting the second order properties of $\boldsymbol{\epsilon}_k$, it is possible to prove that the asymptotical covariance of this estimator can be expressed as:

$$\lim_{K\to\infty}K(\mathbf{h}_{ML}-\mathbf{h})(\mathbf{h}_{ML}-\mathbf{h})^H = \left(\mathbf{T}^H\mathbf{Q}^{-1}\mathbf{T}\right)^{-1}.$$

This expression of the asymptotical covariance is equal to the Cramer-Rao bound for an unbiased estimator [5, pp. 564], which shows that our estimator is asymptotically optimal. We further see that the accuracy of the estimate will depend on both the channel realization and the chosen training sequence.

8. EXPERIMENTAL RESULTS

Since we have shown that the proposed iterative procedure converges to the closed-form ML estimate, we will only focus on this closed-form result when discussing the performance of the proposed ML algorithm. The performance metric that is used throughout this section is the Normalized Mean Squared Error (NMSE) of the proposed channel estimate: $NMSE(\hat{\mathbf{h}}) = \frac{||\hat{\mathbf{h}}-\mathbf{h}||^2}{||\mathbf{h}||^2}$. The experiments are performed on convolutive Rayleigh Fading channels of varying order L. The different taps of the channel are assumed to be identically distributed. The training sequences are random constant modulus sequences. The Signal to Noise Ratio (SNR) is defined as $SNR = \frac{||\mathbf{h}||^2\lambda^2}{\sigma^2}$.

8.1. Performance of the proposed method

In this section, we check how the proposed method performs and how the simulated results match the theoretical performance (asymptotical variance of the channel estimate). The presented results are always averaged over several channel realizations. For each of these channel realizations, the results are averaged over different training sequences.

In Fig. 1, we compare the simulated and the theoretical behavior of our method as a function of the SNR. This comparison is done for different values of the channel order and different values of the number of data blocks K. We see that the simulated and theoretical results match quite well, especially for large K. When the channel order is small, the NMSE keeps decreasing with a constant slope. When the channel order gets larger, the NMSE appears to saturate at higher SNRs. This floor appears when there is no exact solution to the channel identification problem in the noiseless case, this is when $n_t \geqslant 2L+1$ is not fulfilled. In Fig. 2, we evaluate the impact of the number of data blocks K on the NMSE. As expected, the NMSE decreases with a constant slope as K increases. We observe a good match between the theoretical and experimental results as soon as $K \geqslant 30$.

8.2. Comparison with Traditional ML Methods

Classical training-based ML channel estimation techniques discard the received samples that are corrupted by contributions of the unknown data symbols. They solely rely on the part of the received symbols that only contains contributions from the known training symbols. Such received symbols can be observed at the receiver only when $n_t \geqslant L+1$. In that case, the data model is changed into $\mathbf{u}'_k = \mathbf{T}'\mathbf{h} + \boldsymbol{\eta}'$, where $\mathbf{u}'_k = \mathbf{u}_k(L+1:n_t)$, $\mathbf{T}' = \mathbf{T}(L+1:nt,:)$ and $\boldsymbol{\eta}'_k$ is the AWGN term. Note that $\mathbf{T}'$ is an $(n_t-L)\times(L+1)$ Toeplitz matrix with $[t[L+1],\ldots,t[n_t]]^T$ on the first row and $[t[L+1],\ldots,t[1]]$ on the first column.

When the noise term is not colored as before, the solution of the ML channel identification problem is well known to be a simple least squares (LS) fit of $\mathbf{T}'\mathbf{h}$ to $\mathbf{u}'_k$. If we want to have a unique solution to our LS problem, we need to have as many equations as there are unknowns. This means that we need $\mathbf{u}'_k$ to have at least dimension $L+1$. This only happens when $n_t \geqslant 2L+1$. In that case, $\mathbf{T}'$ is tall and we can find a LS solution to derive our ML channel estimate: $\mathbf{h}'_{ML} = ({\mathbf{T}'}^H\mathbf{T}')^{-1}\sum_{k=1}^{K}\mathbf{T}'\mathbf{u}'_k$. Note that when the condition on L and n_t is fulfilled and the traditional method has a solution, the new method has a constant slope. When this condition is not fulfilled, the new method presents a floor in

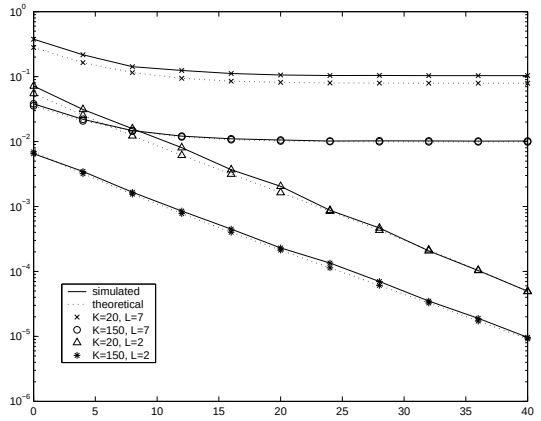

Fig. 1. Comparison of the simulated and theoretical NMSE vs. SNR for different channel orders when $n_T = 5$. The results are plotted for two different values of K, namely 20 and 150.

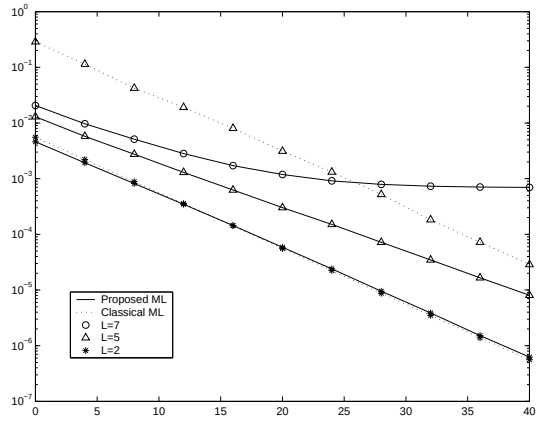

Fig. 3. Simulated NMSE vs. SNR for the proposed and traditional ML channel estimation methods for different channel orders when $n_T = 11$ and $K = 100$.

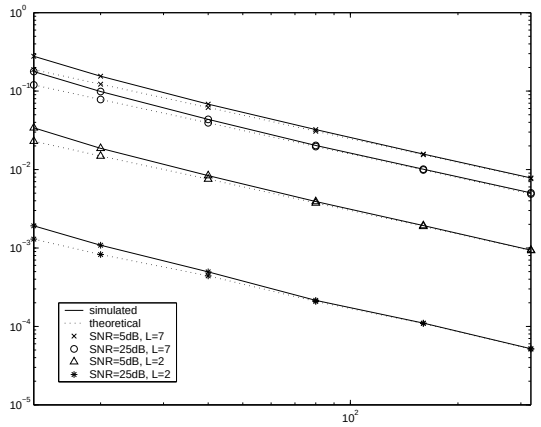

Fig. 2. Comparison of the simulated and theoretical NMSE vs. K for different channel orders when $n_T = 5$. The results are plotted for two different values of the SNR, namely 5 and 25dB.

the NMSE for high SNRs whereas the traditional method does not work.

In Fig. 3, we compare the results for different channel orders when the length of the training sequence n_t is fixed. We see that the new method and the traditional one yield an equivalent performance when the channel order is small. When the channel order increases, the new method outperforms the classical one, especially at low SNRs. When the channel order keeps growing and $n_t \geqslant 2L+1$ is not fulfilled anymore, the new method still provides reliable channel estimates, whilst the traditional method cannot be applied anymore.

9. CONCLUSIONS

In this paper, we presented a new training based ML channel identification method. We first proposed an iterative ML method and then derived closed form expressions for the ML channel estimate.

This new ML method outperforms classical training based ML estimation methods. The reason for this is that all the energy that is received from the known training symbols is exploited in order to estimate the channel, which is not the case for traditional methods. Furthermore, the new method is able to provide us with accurate channel estimates, even when the channel order increases to values that make it impossible to use the classical ML method.

10. REFERENCES

[1] H. Vikalo, B. Hassibi, B. Hochwald, T. Kailath, "Optimal Training for Frequency-Selective Fading Channels," in *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, Utah, May 2001.

[2] G. Leus and M. Moonen, "Semi-Blind Channel Estimation for Block Transmission with Non-Zero Padding," in *Proc. of the Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, Nov. 4-7 2001.

[3] L. Deneire, B. Gyselinckx and M. Engels., "Training Sequence vs. Cyclic Prefix: A New Look on Single Carrier Communication," *IEEE Communication Letters*, vol. 5, no. 7, pp. 292–294, 2001.

[4] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels (mobile radio)," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, November 1991.

[5] T. Söderström, P. Stoica, *System Identification*, International Series in Systems and Control Engineering. Prentice Hall, 1989.

[6] O. Rousseaux, G. Leus, P. Stoica and M. Moonen, "Channel Identification for Block Transmission," in *International Conference on Communications (ICC 2003)*, Anchorage, Alaska, May 2003, submitted.