A Maximum Likelihood Approach for Improved Training Based 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: {olivier.rousseaux}@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

* 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') and was partially sponsored by the Swedish Science Council

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

CONTENTS

I Introduction 3

II Data Model 5

III Maximum Likelihood Approach for Channel Identification 7

IV Iterative Procedure 8

V Closed Form Solution 9

V-A Constant Training Sequence 9

V-B Changing Training Sequence 12

V-C Noise Variance Estimate 14

V-D Comparison with the Iterative Method 14

VI Cramer Rao Bound and Asymptotic Properties 15

VI-A Cramer Rao Bounds 15

VI-B Asymptotic Properties of the Closed Form Channel Estimates 16

VI-B.1 Constant training 16

VI-B.2 Changing training 18

VII Experimental Results 19

VII-A Performance of the Proposed Method 20

VII-A.1 Comparison of the Cramer-Rao Bounds 20

VII-A.2 Changing Training Sequences 21

VII-A.3 Constant Training Sequences 23

VII-B Comparison with Traditional ML Methods 25

VIII Conclusions 26

References 27

Abstract

In this paper, we address the problem of identifying convolutive channels using a Maximum Likelihood (ML) approach when training sequences are 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 smaller than the coherence time of the channel). Several training sequences are used to produce a channel estimate.

Classical training-based identification methods require the use of training sequences that are longer than the channel impulse response, which ensures the presence of received symbols that only contain contributions from the known training symbols (and thus no contribution from the unknown data symbols). Only these symbols are then used in order to estimate the channel. Such methods are thus sub-optimal in the sense that they do not use all the received symbols that contain contributions from the training symbols. In the proposed method, there are no requirements on the length of the training sequences and all the received symbols that contain contributions from the training symbols are exploited. Making some reasonable approximations, we find a simplified expression of the likelihood function of the system. We then propose two approaches to maximize this expression of the likelihood function. The first method is iterative and converges to the channel estimate that maximizes the expression of the likelihood function. In a second approach, we derive low complexity closed-form expressions that maximize this likelihood function. We consider two different problem configurations: either the training sequence is changed after each block of data or the same training sequence is repeated after each block. Experimental results show that the proposed method outperforms classical ML channel identification methods.

I. INTRODUCTION

A major impediment of broadband communication 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]). Most existing methods require the training sequences to be at least as long as the channel impulse response. In this case, some of the received symbols contain contributions only from the known training symbols (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 Maximum Likelihood (ML) channel identification when only these data-free received symbols are used is equivalent to a least squares problem in the presence of Additive White Gaussian Noise (AWGN) at the receiver. However, this classical approach is sub-optimal since not all the received symbols that contain contributions from the training symbols are used in the identification procedure. A more efficient semi-blind technique has been proposed recently in the framework of Known Symbol Padding (KSP) transmission [2]. This technique is deterministic and exploits all the energy that is received from the training sequences. This method only works when a constant training sequence, which must be at least as long as the channel order, is periodically inserted in the transmitted sequence.

We present here a new method that does not impose a minimal length requirement on the training sequences and performs channel estimation exploiting all the received symbols that contain contributions from the training sequences using a ML approach. We consider a transmission scheme where training sequences are inserted between blocks of data symbols. For the sake of simplicity, we consider all the training sequences to have a constant length, but this is not necessary for the method to work. We consider quasi-static channels (the channel stays constant during the transmission of several blocks of data). We investigate both the situation where the same training sequence is repeated after each block of data and the situation where the training sequence is changed after each block of data. This method can be used to identify the channel for all transmission schemes where training symbols are inserted in the data stream, which covers classical training-based transmission but also KSP transmission [4] or Pilot Symbol Assisted Modulation (PSAM) [5].

The structure of the paper is as follows. In section II, we present our data model. In section III,

we derive an expression for the likelihood function of a channel estimate based on specific approximations. We then propose an iterative algorithm that converges to the channel estimate that maximizes this expression of the likelihood function (section IV). We next derive a closed form expression of the channel estimate that maximizes the approximated likelihood function, both for a constant (section V-A) and for a changing training sequence (section V-B). We experimentally test the proposed methods and compare them with classical ML methods in section VII, and finally draw some conclusions in section VIII.

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 $\mathbf{0}_{(M\times N)}$ is the all-zero matrix of size $M \times N$. The operator $(.)^*$ denotes the complex conjugate, $(.)^T$ the transpose and $(.)^H$ the complex conjugate transpose. Finally, $tr(.)$ denotes the trace of a matrix, and $|.|$ its determinant.

II. 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:

$$y[n] = \sum_{i=0}^{L} h[i]x[n-i] + \eta[n], \tag{1}$$

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

As mentioned in the introduction, we consider a transmission scheme where constant length training sequences are inserted between blocks of data symbols. There are two different possibilities: either the same training sequence is repeated after each block of data, or the training sequence is changed after each block. We refer to these two alternative schemes as the constant training sequences case and the changing training sequences case. As we will see later, these two alternative schemes yield different channel identification procedures. We describe the more general situation of a changing training sequence as often as possible, and only analyze the case of a constant training sequence when explicitly needed.

A total number of K training sequences is inserted in the stream of unknown data symbols. The kth training sequence, $\mathbf{t}_k = [t_k[1], \ldots, t_k[n_t]]^T$, starts at position n_k: $[x[n_k], \ldots, x[n_k+n_t-1]]^T = \mathbf{t}_k$. 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 $\mathbf{t}_k$ 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:

$$\mathbf{u}_k = \mathbf{T}_k \mathbf{h} + \boldsymbol{\epsilon}_k, \tag{2}$$

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

$$\boldsymbol{\epsilon}_k = \underbrace{\left[\begin{array}{ccc|ccc} h_L & \cdots & h_1 & & & \mathbf{0} \\ & \ddots & \vdots & h_0 & & \\ & & h_L & \vdots & \ddots & \\ \mathbf{0} & & & h_{L-1} & \cdots & h_0 \end{array}\right]}_{\mathbf{H}_s (n_t + L) \times (2L)} \mathbf{s}_k + \boldsymbol{\eta}_k, \tag{3}$$

with $\mathbf{s}_k = [s_k[1], \cdots, s_k[2L]] = [x[n_k - L], \ldots, x[n_k - 1], x[n_k + n_t], \ldots, x[n_k + n_t + L - 1]]^T$ the vector of surrounding data symbols, and $\boldsymbol{\eta}_k = [\eta[n_k], \cdots, \eta[n_k + n_t + L - 1]]$ is the corresponding AWGN term. Assuming that both the noise and the data are white and zero-mean ($E\{s_k[i]s_k[j]^*\} = E\{\eta[i]\eta[j]^*\} = 0, \quad \forall i, j, k,$ when $i \neq j$, and $E\{s_k[i]\} = E\{\eta[k]\} = 0$), we can say that $\boldsymbol{\epsilon}_k$ is zero-mean. Defining n_s as the length of the shortest sequence of data symbols ($n_s = \min_k\{n_{k+1} - (n_k + n_t - 1)\}$), we assume $n_s \geqslant 2L$. This ensures that the $\mathbf{s}_k$'s are uncorrelated, i.e. $E\{\mathbf{s}_k\mathbf{s}_l^*\} = \mathbf{0}$. Defining the signal and noise variances as $\lambda^2 = E\{s_k[i]s_k[i]^*\}$ and $\sigma^2 = E\{\eta[k]\eta[k]^*\}$ respectively, we can derive the covariance matrix of $\boldsymbol{\epsilon}_k$ from (3) as $E\left\{\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H\right\} \triangleq \mathbf{Q} = \lambda^2\mathbf{H}_s\mathbf{H}_s^H + \sigma^2\mathbf{I}$. The first and second order statistics of the stochastic term are thus as follows:

$$\begin{aligned} E\{\boldsymbol{\epsilon}_k\} &= \mathbf{0}_{(n_t+L)\times 1}, \\ E\left\{\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H\right\} \triangleq \mathbf{Q} &= \lambda^2\mathbf{H}_s\mathbf{H}_s^H + \sigma^2\mathbf{I}, \\ E\left\{\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_l^H\right\} &= \mathbf{0}_{(n_t+L)\times(n_t+L)}. \end{aligned} \tag{4}$$

III. Maximum Likelihood Approach for Channel Identification

An expression for the likelihood function is needed in order to perform Maximum Likelihood channel estimation. Such an expression is well known for systems where the disturbance has a Gaussian distribution. For the problem we are considering, where the disturbance is given by $\boldsymbol{\epsilon}_k$, this is not exactly the case. The noise part $\boldsymbol{\eta}_k$ has been considered as circularly Gaussian distributed (AWGN approximation), but this is not the case for the data symbols contribution $\mathbf{s}_k$. However, we can use the Gaussian likelihood function as a statistically sound fitting criterion if we assume that $\boldsymbol{\epsilon}_k$ are circularly Gaussian distributed. The longer the training sequences and the lower the SNR, the more accurate this approximation will be. Relying on this approximation, we can express (up to a constant term) the negative log-likelihood function of the system as:

$$-\mathcal{L} = K \ \ln|\mathbf{Q}| + \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H \mathbf{Q}^{-1} (\mathbf{u}_k - \mathbf{T}_k\mathbf{h}). \tag{5}$$

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. The corresponding ML channel estimate minimizes this expression w.r.t. $\mathbf{h}$ and σ^2. This minimization problem boils down to a computationally demanding $(L+2)$-dimensional nonlinear search.

To overcome this complexity 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. The corresponding ML channel estimate thus maximizes the likelihood function w.r.t $\mathbf{h}$ and $\mathbf{Q}$ (instead of $\mathbf{h}$ and $\boldsymbol{\sigma}^2$)

The approximations that were needed to approximate the ML problem are thus summarized as:

- $\boldsymbol{\epsilon}_k$ is Gaussian distributed
- $\mathbf{Q} = \mathbf{Q}^H > 0$ does not depend on the channel model

These assumptions transform the initial ML problem into an optimization problem which is separable in its two variables $\mathbf{Q}$ and $\mathbf{h}$. We exploit this separability property in the next sections in order to solve the optimization problem in a less complex way than the $(L+2)$-dimensional nonlinear search of the initial ML problem.

In the following, we will call the solution of the above optimization problem *the ML channel estimate*. This terminology is adopted for the sake of convenience as the presented results are actually approximations of the true ML channel.

IV. Iterative Procedure

When a minimization problem is separable in its variables, a common approach to find the solution is an iterative one. One 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 (see e.g. [6] where this approach is used to jointly estimate the transmitted data symbols and the channel). 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. In the sequel, we apply this approach to the likelihood function of the system, which leads us to the ML estimate of $\mathbf{Q}$ and $\mathbf{h}$.

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 (5) with respect to $\mathbf{h}$ for a fixed $\mathbf{Q} = \hat{\mathbf{Q}}^i$, i.e. we compute $\hat{\mathbf{h}}^i = \mathbf{h}_{ML}(\hat{\mathbf{Q}}^i)$, where $\mathbf{h}_{ML}(\mathbf{Q}) = \arg\min_{\mathbf{h}} -\mathcal{L}$. The solution to this optimization problem can be computed as:

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

We then seek the covariance matrix $\hat{\mathbf{Q}}^{i+1}$ that minimizes (5) with respect to $\mathbf{Q}$ for a fixed $\mathbf{h} = \hat{\mathbf{h}}^i$: $\hat{\mathbf{Q}}^{i+1} = \mathbf{Q}_{ML}(\hat{\mathbf{h}}^i)$, where $\mathbf{Q}_{ML}(\mathbf{h}) = \arg\min_{\mathbf{Q}} -\mathcal{L}$, the solution to this optimization problem can be computed as[1]:

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H. \tag{7}$$

$\hat{\mathbf{Q}}^{i+1}$ 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. Note that we still have to find an acceptable starting point for the iterations. In [8], we proposed an iterative method for channel identification in a similar context. It is straightforward to see

[1]The derivation of this result is detailed in Appendix A

that this method, which was presented as an iterative Weighted Least Squares (WLS) one, is actually similar to the one we propose here and is thus an iterative ML method. We show in [8] that initializing the iterative procedure with a simple Least Squares (LS) channel estimate yields good convergence properties. It is easy to show that applying the method we propose here with an identity matrix as initial covariance matrix yields exactly this LS channel estimate as initial channel estimate. We thus propose to initialize the iterative ML method with $\hat{\mathbf{Q}}^0 = \mathbf{I}$.

The iterative procedure is summarized as follows:

Algorithm 1 Iterative Maximum Likelihood Algorithm:

Initialization: $\hat{\mathbf{Q}}^0 = \mathbf{I}$

Iterations: $\hat{\mathbf{h}}^i = \left(\sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}^{i^{-1}} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}^{i^{-1}} \mathbf{u}_k$

$\hat{\mathbf{Q}}^{i+1} = K^{-1} \sum_{k=1}^{K} \left(\mathbf{u}_k - \mathbf{T}_k \hat{\mathbf{h}}^i \right) \left(\mathbf{u}_k - \mathbf{T}_k \hat{\mathbf{h}}^i \right)^H$

V. Closed Form Solution

An alternative strategy to the iterative procedure described above consists of directly finding an analytical expression for the global minimum of the likelihood function (5). The separability 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 then be eliminated in the original cost function, which then becomes a single variable expression. When the problem is separable in its two variables minimizing this new expression of the cost function with respect to the only variable left yields the global minimum. At this point, we need to distinguish between the constant training sequence case and the changing training sequence case.

A. Constant Training Sequence

In order to indicate that the training sequence after each block is the same, we simply omit the block index (subscript k) for the vectors $\mathbf{t}$ and the matrices $\mathbf{T}$. We first minimize the cost

function with respect to $\mathbf{Q}$, the solution of which is given by (7) with a constant $\mathbf{T}$ matrix:

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H. \tag{8}$$

Replacing $\mathbf{Q}$ by $\mathbf{Q}_{ML}(\mathbf{h})$ in the likelihood function as expressed in (5) yields the following:

$$-\mathcal{L} = K \mathrm{tr}\,(\mathbf{I}) + K \ln \left| K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H \right| + \text{cst.} \tag{9}$$

The ML channel estimate $\mathbf{h}_{ML}$ minimizes (9) with respect to $\mathbf{h}$:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} \left| K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H \right|. \tag{10}$$

Define

$$\begin{aligned} \hat{\mathbf{R}} &\triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^H, \\ \bar{\mathbf{u}} &\triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k, \\ \hat{\mathbf{Q}} &\triangleq \hat{\mathbf{R}} - \bar{\mathbf{u}} \bar{\mathbf{u}}^H, \end{aligned} \tag{11}$$

where $\hat{\mathbf{Q}}$ is assumed to be positive definite[2]. Using these definitions, the matrix in the minimization problem (10) can be re-expressed as:

$$\begin{aligned} & K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H \\ = \;& K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^H - K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \mathbf{h}^H \mathbf{T}^H - K^{-1} \sum_{k=1}^{K} \mathbf{T}\mathbf{h}\mathbf{u}_k^H + K^{-1} \sum_{k=1}^{K} \mathbf{T}\mathbf{h}\mathbf{h}^H \mathbf{T}^H \\ = \;& \hat{\mathbf{R}} - \bar{\mathbf{u}}\mathbf{h}^H \mathbf{T}^H - \mathbf{T}\mathbf{h}\bar{\mathbf{u}}^H + \mathbf{T}\mathbf{h}\mathbf{h}^H \mathbf{T}^H \\ = \;& \hat{\mathbf{R}} + (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H - \bar{\mathbf{u}}\bar{\mathbf{u}}^H \\ = \;& \hat{\mathbf{Q}} + \hat{\mathbf{Q}}\hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \\ = \;& \hat{\mathbf{Q}} \left(\mathbf{I} + \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \right), \end{aligned}$$

[2]We see from (11) that a necessary condition for this to hold is $K \geqslant n_t + L$. When this condition is fulfilled, the randomness of the noise and the data ensures that $\hat{\mathbf{Q}}$ is a positive definite matrix with probability 1.

Keeping in mind that $\hat{\mathbf{Q}}$ is positive definite, our minimization problem (10) is thus equivalent to:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} \left| \mathbf{I} + \hat{\mathbf{Q}}^{-1} \left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right) \left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right)^H \right|. \tag{12}$$

It can be shown[3] that

$$\left| \mathbf{I} + \hat{\mathbf{Q}}^{-1} \left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right) \left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right)^H \right| = 1 + \left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right)^H \hat{\mathbf{Q}}^{-1} \left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right).$$

Hence, the minimization problem (12) is equivalent to:

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

The solution is obtained by nulling the partial derivative of this expression with respect to $\mathbf{h}^H$, which yields:

$$\boxed{\mathbf{h}_{ML} = \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \mathbf{T}\right)^{-1} \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \bar{\mathbf{u}}\right).} \tag{13}$$

This ML channel estimate is easy to compute and also 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 (11), the expression (8) for the ML estimate of $\mathbf{Q}$ as a function of $\mathbf{h}$ can be rewritten as:

$$\begin{aligned} \mathbf{Q}_{ML}(\mathbf{h}) &= \hat{\mathbf{R}} - \bar{\mathbf{u}}\mathbf{h}^H\mathbf{T}^H - \mathbf{T}\mathbf{h}\bar{\mathbf{u}}^H + \mathbf{T}\mathbf{h}\mathbf{h}^H\mathbf{T}^H \\ &= \hat{\mathbf{Q}} + \left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right)\left(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}\right)^H. \end{aligned} \tag{14}$$

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

$$\boxed{\mathbf{Q}_{ML} = \mathbf{Q}_{ML}(\mathbf{h}_{ML}) = \hat{\mathbf{Q}} + \left(\mathbf{T}\mathbf{h}_{ML} - \bar{\mathbf{u}}\right)\left(\mathbf{T}\mathbf{h}_{ML} - \bar{\mathbf{u}}\right)^H,} \tag{15}$$

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

[3]We detail the proof in Appendix B.

B. Changing Training Sequence

We proceed in the same way as for the constant training sequence case. First observe that the likelihood function (5) can be expressed as:

$$-\mathcal{L} = K\ln|\mathbf{Q}| + \mathrm{tr}\left(\mathbf{Q}^{-1}\sum_{k=1}^{K}(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H\right). \tag{16}$$

We first minimize this cost function with respect to $\mathbf{Q}$ leading to $\mathbf{Q}_{ML}(\mathbf{h})$ as given by (7). Replacing $\mathbf{Q}$ by $\mathbf{Q}_{ML}(\mathbf{h})$ in (16) leaves us with an expression of the cost function that only depends on $\mathbf{h}$:

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

The ML channel estimate is thus computed as:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}}\left|\sum_{k=1}^{K}(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H\right|. \tag{17}$$

Although this problem seems similar to (10), the varying $\mathbf{T}_k$ forces us to adopt a different approach which will only lead us to an approximate solution. Let us first introduce the following notations:

$$\begin{aligned}
&\mathbf{h}_{LS} \triangleq \left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{T}_k\right)^{-1}\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{u}_k,\\
&\mathbf{g} \triangleq \mathbf{h} - \mathbf{h}_{LS},\\
&\mathbf{g}_{ML} \triangleq \mathbf{h}_{ML} - \mathbf{h}_{LS},\\
&\mathbf{e}_k \triangleq \mathbf{u}_k - \mathbf{T}_k\mathbf{h}_{LS},\\
&\hat{\mathbf{Q}}' \triangleq K^{-1}\sum_{k=1}^{K}\mathbf{e}_k\mathbf{e}_k^H.
\end{aligned} \tag{18}$$

where $\hat{\mathbf{Q}}'$ is assumed to be positive definite[4]. Using these notations, the minimization problem (17) can be rephrased as:

$$\mathbf{g}_{ML} = \arg\min_{\mathbf{g}}\left|\sum_{k=1}^{K}(\mathbf{e}_k - \mathbf{T}_k\mathbf{g})(\mathbf{e}_k - \mathbf{T}_k\mathbf{g})^H\right|. \tag{19}$$

[4]We see from (18) that a necessary condition therefore is $K \geqslant n_t + L$. When this condition is fulfilled, the randomness of the noise and the data ensures that $\hat{\mathbf{Q}}'$ is a positive definite matrix with probability 1.

The determinant in this last expression can be expressed as (up to a positive factor):

$$\left| \mathbf{I} + \hat{\mathbf{Q}}'^{-1} K^{-1} \sum_{k=1}^{K} \left[\mathbf{T}_k \mathbf{g}\mathbf{g}^H \mathbf{T}_k^H - \mathbf{T}_k \mathbf{g}\mathbf{e}_k^H - \mathbf{e}_k \mathbf{g}^H \mathbf{T}_k^H \right] \right|. \tag{20}$$

When K is large, both $\hat{\mathbf{h}}_{LS}$ and $\mathbf{h}_{ML}$ are close to the true $\mathbf{h}$. We can thus assume that $\mathbf{g}$, and consequently the second term in (20), are small in the vicinity of the solution. It is well known that, for $||\mathbf{\Delta}|| \ll 1$, $|\mathbf{I} + \mathbf{\Delta}| \approx 1 + tr\,(\mathbf{\Delta})$. Hence, for $K \gg 1$, the minimization problem in (19) can be approximated by:

$$\hat{\mathbf{g}}_{ML} = \arg\min_{\mathbf{g}} tr \left(\hat{\mathbf{Q}}'^{-1} \sum_{k=1}^{K} \left[\mathbf{T}_k \mathbf{g}\mathbf{g}^H \mathbf{T}_k^H - \mathbf{T}_k \mathbf{g}\mathbf{e}_k^H - \mathbf{e}_k \mathbf{g}^H \mathbf{T}_k^H \right] \right),$$

where $\hat{\mathbf{g}}_{ML}$ is an approximation of $\mathbf{g}_{ML}$. Exploiting the permutation property of the trace of a product, this problem can be rephrased as:

$$\hat{\mathbf{g}}_{ML} = \arg\min_{\mathbf{g}} \sum_{k=1}^{K} \left[\mathbf{g}^H \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{g} - \mathbf{e}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{g} - \mathbf{g}^H \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{e}_k \right]. \tag{21}$$

The solution to this minimization problem is obtained by nulling the partial derivate of this expression with respect to $\mathbf{g}^H$, and is given as:

$$\hat{\mathbf{g}}_{ML} = \left(\sum_{k=1}^{K} {\mathbf{T}_k}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{e}_k$$

We know from (18) that $\mathbf{h} = \mathbf{g} + \mathbf{h}_{LS}$. If we additionally replace $\mathbf{e}_k$ by $\mathbf{u}_k - \mathbf{T}_k \mathbf{h}_{LS}$, we obtain the following approximation $\hat{\mathbf{h}}_{ML}$ of the channel estimate $\mathbf{h}_{ML}$:

$$\hat{\mathbf{h}}_{ML} = \mathbf{h}_{LS} + \left(\sum_{k=1}^{K} {\mathbf{T}_k}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{u}_k - \left(\sum_{k=1}^{K} {\mathbf{T}_k}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{h}_{LS},$$

which effectively simplifies to:

$$\boxed{\hat{\mathbf{h}}_{ML} = \left(\sum_{k=1}^{K} {\mathbf{T}_k}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{u}_k.} \tag{22}$$

Based on this approximation of the ML channel estimate, it is possible to derive an approximation $\hat{\mathbf{Q}}_{ML}$ of the ML covariance matrix $\mathbf{Q}_{ML}$ inserting $\hat{\mathbf{h}}_{ML}$ into (7). First note that using (18), the expression (7) of $\mathbf{Q}_{ML}(\mathbf{h})$ can be rewritten as:

$$\mathbf{Q}_{ML}(\mathbf{h}) = \hat{\mathbf{Q}}' + K^{-1}\sum_{k=1}^{K}\left(\mathbf{T}_k\mathbf{h}-\mathbf{u}_k\right)\left(\mathbf{T}_k\mathbf{h}-\mathbf{u}_k\right)^H - K^{-1}\sum_{k=1}^{K}\left(\mathbf{T}_k\hat{\mathbf{h}}_{LS}-\mathbf{u}_k\right)\left(\mathbf{T}_k\hat{\mathbf{h}}_{LS}-\mathbf{u}_k\right)^H. \tag{23}$$

The approximate ML covariance matrix $\hat{\mathbf{Q}}_{ML}$ is obtained by replacing $\mathbf{h}$ by $\hat{\mathbf{h}}_{ML}$ in this last expression:

$$\boxed{\hat{\mathbf{Q}}_{ML} = \hat{\mathbf{Q}}' + K^{-1}\sum_{k=1}^{K}\left(\mathbf{T}_k\mathbf{h}_{ML}-\mathbf{u}_k\right)\left(\mathbf{T}_k\mathbf{h}_{ML}-\mathbf{u}_k\right)^H - K^{-1}\sum_{k=1}^{K}\left(\mathbf{T}_k\hat{\mathbf{h}}_{LS}-\mathbf{u}_k\right)\left(\mathbf{T}_k\hat{\mathbf{h}}_{LS}-\mathbf{u}_k\right)^H.} \tag{24}$$

C. Noise Variance Estimate

Besides an accurate channel estimate, which is helpful in many applications, an estimate of the noise variance can be worthwhile obtaining as well. This is the case for instance when Minimum Mean Square Error (MMSE) equalizers are sought. When both the ML estimates $\mathbf{h}_{ML}$ and $\mathbf{Q}_{ML}$ (or approximations thereof) are available, it is possible to estimate the noise variance σ^2 as outlined below.

Using the available ML channel estimate, we rely on the definition (3) of the matrix $\mathbf{H}_s$ in which we replace $\mathbf{h}$ by $\mathbf{h}_{ML}$ (or $\hat{\mathbf{h}}_{ML}$) to build $\hat{\mathbf{H}}_s$. If our estimates are accurate enough, subtracting $\lambda^2\hat{\mathbf{H}}_s\hat{\mathbf{H}}_s^H$ from $\mathbf{Q}_{ML}$ (or $\hat{\mathbf{Q}}_{ML}$) should leave us with a diagonal matrix with σ^2 repeated over the diagonal. We can thus use the following noise variance estimate:

$$\hat{\sigma}^2 = (n_T + L)^{-1}\, tr\left(\mathbf{Q}_{ML} - \lambda^2\hat{\mathbf{H}}_s\hat{\mathbf{H}}_s^H\right). \tag{25}$$

This estimate can be used in the case of both constant or changing training sequences (in which case $\mathbf{Q}_{ML}$ is replaced by $\hat{\mathbf{Q}}_{ML}$).

D. Comparison with the Iterative Method

As mentioned in [8], the iterative method seems to almost converge after one iteration. In this section, we explain this fast convergence by comparing the estimates provided by the iterative method with the closed form expressions derived above. We assume $K \gg 1$ throughout this discussion.

Let us first discuss the *constant training sequence* case. Initializing the algorithm with $\hat{\mathbf{Q}}^0 = \mathbf{I}$ yields $\hat{\mathbf{h}}^0 = \left(\mathbf{T}^H\mathbf{T}\right)^{-1}\mathbf{T}^H\bar{\mathbf{u}}$. Next, exploiting the expression (14) of $\mathbf{Q}_{ML}(\mathbf{h})$, we can write $\hat{\mathbf{Q}}^1 = \hat{\mathbf{Q}} + \left(\mathbf{T}\hat{\mathbf{h}}^0 - \bar{\mathbf{u}}\right)\left(\mathbf{T}\hat{\mathbf{h}}^0 - \bar{\mathbf{u}}\right)^H$. Observing that $\hat{\mathbf{h}}^0$ is the least squares fit of $\mathbf{Th}$ to $\bar{\mathbf{u}}$, we can say that $\mathbf{T}\hat{\mathbf{h}}^0 - \bar{\mathbf{u}}$ is small. The second term of the right-hand side of $\hat{\mathbf{Q}}^1$ 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 (13), we then observe that $\hat{\mathbf{h}}^1$ is close to $\mathbf{h}_{ML}$:

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

When *changing training sequences* are used, it is straightforward to see that initializing the iterative method with $\hat{\mathbf{Q}}^0 = \mathbf{I}$ yields $\hat{\mathbf{h}}^0 = \hat{\mathbf{h}}_{LS}$. It then follows that $\hat{\mathbf{Q}}^1 = \hat{\mathbf{Q}}'$ and, observing the similarity between (6) and (22), $\hat{\mathbf{h}}^1 = \hat{\mathbf{h}}_{ML}$. Hence the iterative procedure yields the approximate closed form ML channel estimate after one iteration.

This discussion shows that in both the constant and the changing training sequences cases, the channel estimate obtained after one iteration, $\hat{\mathbf{h}}^1$, is a good approximation of the ML channel estimate $\mathbf{h}_{ML}$. Keeping in mind that this $\mathbf{h}_{ML}$ is the convergence point of the iterative procedure explains the fast convergence properties of the iterative method.

VI. Cramer Rao Bound and Asymptotic Properties

A. Cramer Rao Bounds

The Cramer-Rao Bound (CRB) is a theoretical lower bound on the covariance matrix of an unbiased channel estimate. It can be shown (see e.g. [7, pp. 562]) that for any unbiased channel estimate $\hat{\mathbf{h}}$ of the channel $\mathbf{h}$, the following inequality holds:

$$\mathrm{cov}(\hat{\mathbf{h}}) \geqslant \boldsymbol{\mathcal{J}}^{-1},$$

where the covariance matrix of the channel estimate is defined as $\mathrm{cov}(\hat{\mathbf{h}}) = E\left\{(\hat{\mathbf{h}} - \mathbf{h})(\hat{\mathbf{h}} - \mathbf{h})^H\right\}$, and the Fisher information matrix is defined as $\boldsymbol{\mathcal{J}} = E\left\{\left(\frac{\partial \log(L)}{\partial \mathbf{h}}\right)^T\left(\frac{\partial \log(L)}{\partial \mathbf{h}}\right)\right\}$. Repeating the hypothesis of a circularly gaussian distributed noise term $\boldsymbol{\epsilon}_K$ allows us to use the expression (5) of the likelihood function. The dependence of $\mathbf{Q}$ on the channel parameters and the presence of $\mathbf{Q}^{-1}$ in this expression of the likelihood function make the derivation of this bound a quite complex problem. However, if we also repeat the hypotheses of independent $\mathbf{Q}$ and $\mathbf{h}$, we can

find an expression for the CRB of the channel identification problem under consideration in this paper.

In order to find this expression, we use an alternative formulation of the data model. Define $\mathbf{T}_{tot}$ as the collection of all the $\mathbf{T}_k$ matrices: $\mathbf{T}_{tot} = \left[\mathbf{T}_1^H, \cdots, \mathbf{T}_K^H\right]^H$, $\mathbf{u}_{tot}$ as the collection of all $\mathbf{u}_k$ vectors: $\mathbf{u}_{tot} = \left[\mathbf{u}_1^H, \cdots, \mathbf{u}_K^H\right]^H$, and $\boldsymbol{\epsilon}_{tot}$ as the collection of all $\boldsymbol{\epsilon}_k$ vectors: $\boldsymbol{\epsilon}_{tot} = \left[\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_K\right]^H$. Collecting the K equations from (2) allows us to express the data model as:

$$\mathbf{u}_{tot} = \mathbf{T}_{tot}\mathbf{h} + \boldsymbol{\epsilon}_{tot}. \tag{26}$$

It is clear from the statistics of $\boldsymbol{\epsilon}_k$ that $\boldsymbol{\epsilon}_{tot}$ is zero-mean ($E\{\boldsymbol{\epsilon}_{tot}\} = \mathbf{0}$) and that its covariance matrix $\mathbf{Q}_{tot} = E\{\boldsymbol{\epsilon}_{tot}\boldsymbol{\epsilon}_{tot}^H\}$ is a block-diagonal matrix with $\mathbf{Q}$ repeated along the main diagonal. In this context, the CRB, $\boldsymbol{\mathcal{J}}^{-1}$, is well known to be [7, pp. 564]: $\boldsymbol{\mathcal{J}}^{-1} = (\mathbf{T}_{tot}^H\mathbf{Q}_{tot}^{-1}\mathbf{T}_{tot})^{-1}$. Exploiting the block-diagonal structure of $\mathbf{Q}_{tot}$, this result can be written as:

$$\boxed{\boldsymbol{\mathcal{J}}^{-1} = \left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right)^{-1}.} \tag{27}$$

When constant training sequences are used, this expression simplifies into:

$$\boxed{\boldsymbol{\mathcal{J}}^{-1} = K^{-1}\left(\mathbf{T}^H\mathbf{Q}^{-1}\mathbf{T}\right)^{-1}.} \tag{28}$$

Note that this bound depends both on the channel realization (through the covariance matrix $\mathbf{Q}$) and on the chosen training sequences.

B. Asymptotic Properties of the Closed Form Channel Estimates

In this section, we study the asymptotic properties of the proposed closed form (approximate) ML channel estimates, that is their properties when the number of transmitted data blocks, K, is large.

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

$$\begin{aligned}\hat{\mathbf{Q}} &= \hat{\mathbf{R}} - \bar{\mathbf{u}}\bar{\mathbf{u}}^H \\ &= K^{-1}\sum_{k=1}^{K}\mathbf{u}_k\mathbf{u}_k^H - K^{-2}\sum_{k=1}^{K}\mathbf{u}_k\sum_{k=1}^{K}\mathbf{u}_k^H.\end{aligned}$$

Keeping in mind that $\mathbf{u}_k = \mathbf{T}\mathbf{h} + \boldsymbol{\epsilon}_k$, we have:

$$\hat{\mathbf{Q}} = K^{-1} \sum_{k=1}^{K} \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H - K^{-2} \sum_{i,j=1}^{K} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^H. \tag{29}$$

Using the central limit theorem, the above time averages can be replaced by their expected values when K tends to infinity:

$$\begin{aligned} \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H &= E\left\{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H\right\} = \mathbf{Q}, \\ \lim_{K\to\infty} K^{-2} \sum_{i,j=1}^{K} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^H &= \lim_{K\to\infty} K^{-1}\mathbf{Q} = 0. \end{aligned} \tag{30}$$

Therefore,

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

It follows that the ML channel estimate (13) is consistent:

$$\begin{aligned} \lim_{K\to\infty} \mathbf{h}_{ML} &= \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{Q}^{-1} \mathbf{u}_k & (31) \\ &= \mathbf{h} + \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{Q}^{-1} \boldsymbol{\epsilon}_k & (32) \\ &= \mathbf{h} & (33) \end{aligned}$$

The expected value of the scaled covariance matrix of the asymptotic channel estimate can be computed as:

$$\begin{aligned} &E\left\{\lim_{K\to\infty} K(\mathbf{h}_{ML} - \mathbf{h})(\mathbf{h}_{ML} - \mathbf{h})^H\right\} \\ &\quad = \lim_{K\to\infty} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{Q}^{-1} E\left\{K^{-1} \sum_{i,j=1}^{K} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^H\right\} \mathbf{Q}^{-1} \mathbf{T} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \\ &\quad = \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{Q}^{-1} \mathbf{Q} \mathbf{Q}^{-1} \mathbf{T} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \\ &\quad = \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1}. \end{aligned}$$

This expression is equal to the scaled CRB (see (28)).

2) Changing training: Let us first note that $\mathbf{e}_k$ can be rewritten as

$$\begin{aligned}\mathbf{e}_k &= \mathbf{u}_k - \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{u}_l \\ &= \boldsymbol{\epsilon}_k - \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \sum_{l=1}^{K} \mathbf{T}_l^H \boldsymbol{\epsilon}_l\end{aligned}$$

Based on this observation, it is possible to check that the shorthand notation $\hat{\mathbf{Q}}'$ defined in (18) converges to the true $\mathbf{Q}$ when K tends to infinity:

$$\begin{aligned}\lim_{K\to\infty} \hat{\mathbf{Q}}' &= \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \mathbf{e}_k \mathbf{e}_k^H \\ &= \lim_{K\to\infty} E\{\mathbf{e}_k \mathbf{e}_k^H\} \\ &= \lim_{K\to\infty} \mathbf{Q} - \mathbf{Q}\mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H - \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H \mathbf{Q}^H \\ &\quad + \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \left(\sum_{l=1}^{K} \mathbf{T}_l \mathbf{Q} \mathbf{T}_l^H \right)^{-1} \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H \\ &= \mathbf{Q}.\end{aligned}$$

Replacing $\mathbf{u}_k$ by its equivalent $\mathbf{T}_k \mathbf{h} + \boldsymbol{\epsilon}_k$ in (22) yields:

$$\hat{\mathbf{h}}_{ML} = \mathbf{h} + \left(\sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \boldsymbol{\epsilon}_k. \tag{34}$$

If the training sequences have a constant non-zero energy, it is clear that

$$\lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \neq \mathbf{0}.$$

Furthermore, it is clear from the statistics of $\boldsymbol{\epsilon}_k$ that

$$\lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \boldsymbol{\epsilon}_k = \mathbf{0}.$$

This shows that the approximate ML channel estimate is consistent:

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

Here also, it is possible to derive an expression for the expected value of the covariance matrix of the asymptotic channel estimate:

$$
\begin{aligned}
&E\left\{\lim_{K\to\infty}\left(\hat{\mathbf{h}}_{ML}-\mathbf{h}\right)\left(\hat{\mathbf{h}}_{ML}-\mathbf{h}\right)^{H}\right\} \\
&\quad=\lim_{K\to\infty}\left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right)^{-1}E\left\{\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\boldsymbol{\epsilon}_k\sum_{k=1}^{K}\boldsymbol{\epsilon}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right\}\left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right)^{-1} \\
&\quad=\lim_{K\to\infty}\left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right)^{-1}\left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{T}_k\right)\left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right)^{-1} \\
&\quad=\lim_{K\to\infty}\left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right)^{-1}
\end{aligned}
$$

The above expression of the expected value of the covariance matrix of the asymptotic channel estimate is equal to the CRB (see (27)).

VII. Experimental Results

The performance metric that is used throughout this section is the Normalized Mean Square Error (NMSE) of the proposed channel estimate:

$$\text{NMSE}=E\left\{\frac{||\hat{\mathbf{h}}-\mathbf{h}||^2}{||\mathbf{h}||^2}\right\}.$$

The results that are presented are obtained with the closed form channel estimates (13) and (22). When the iterative method results are investigated, we explicitly state it in the text. We use the CRB as a benchmark in the experiments. The CRB curves displayed on the graphs represent the NMSE of an estimator that achieves the CRB, which is $E\left\{\frac{\text{tr}(\boldsymbol{\mathcal{J}}^{-1})}{||\mathbf{h}||^2}\right\}$. The experiments are performed on convolutive Rayleigh Fading channels of varying order L. The different channel taps are independently identically distributed (Gaussian distribution). The training and data sequences are white QPSK sequences. The energy of the transmitted symbols (both data and training) is set to λ^2. Channels, data and training symbols are randomly generated. The presented results are obtained after averaging over a set of 100 channel realizations. For each of these channel realizations, the results are averaged over 100 different sets of training sequences in the changing training sequences case and over 100 different training sequences in the constant training sequences case. Note that this averaging is also done for the CRB results since the CRB depends both on the channel realization and the training sequences. The Signal to Noise Ratio (SNR) is defined as $SNR=E\left\{||\mathbf{h}||^2\right\}\frac{\lambda^2}{\sigma^2}$.

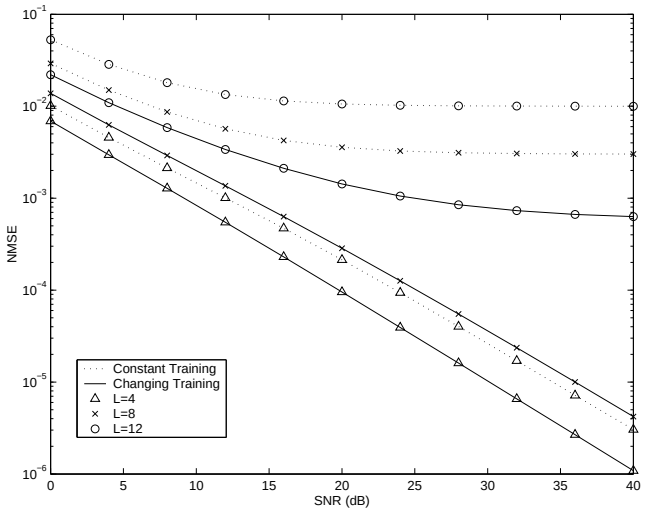

Fig. 1. Comparison of the CRB of the constant and changing training sequences cases vs. the SNR for different channel orders when $n_t = 10$ and $K = 100$.

A. Performance of the Proposed Method

In this section, we analyze and compare the algorithms proposed for the two situations that have been considered throughout this article: the constant and changing training sequence cases.

1) Comparison of the Cramer-Rao Bounds : To have a first insight on how these compare, we check the CRB performance for these two configurations. We consider a transmission scheme where the length of the training sequences is set to $n_t = 10$ and the number of observed training sequences is set to $K = 100$. The CRB for different channel orders in that context is presented in Fig. 1. We see that the use of changing training sequences systematically results in a reduced CRB for all channel orders. When the channel order is small, the CRB keeps decreasing with a constant slope as the SNR increases in both contexts, but the use of changing training sequences yields improved performance. For large channel orders, the CRB saturates at higher SNR in both contexts and the changing training curve again shows a better performance. For intermediate channel orders, we observe a saturation in the CRB curve when constant training is used whereas the use of changing training yields a constant slope in the CRB curve.

It is possible to show that the CRB decreases with a constant slope as the SNR increases when there is an exact solution to the channel identification problem in the noiseless case. The saturation effect in the CRB appears when there is no solution to the channel identification

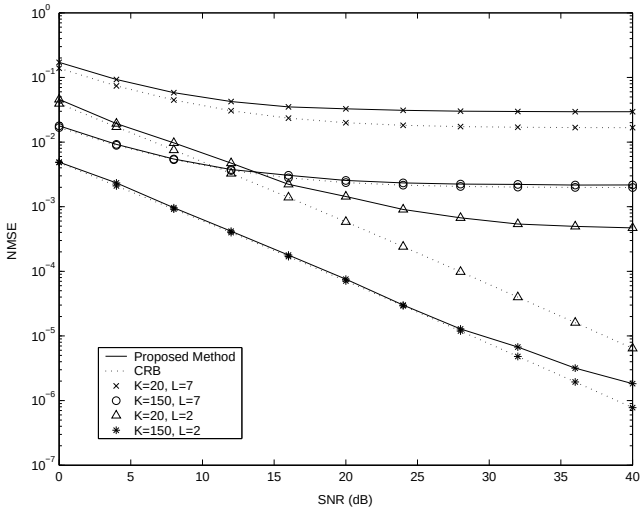

Fig. 2. Comparison of the simulated NMSE and the CRB vs. SNR for different channel orders when $n_t = 5$ and changing training sequences are used. The results are plotted for two different values of K, namely 20 and 150

problem in the noiseless case.

Observe that the number of received symbols that do not contain contributions from the unknown data symbols is equal to $n_t - L$. When constant training sequences are used, the channel identification problem in the noiseless case has an exact solution if there are at least $L+1$ such received symbols. That is when $n_t \geqslant 2L+1$. When changing training sequences are used, an exact solution exists in the noiseless case as soon as there is one such received symbol per transmitted training sequence, that is when $n_t \geqslant L+1$.

When n_t is fixed and the channel order is in the interval $\frac{n_t-1}{2} \leqslant L \leqslant n_t - 1$, using changing training sequences will yield a constant slope in the CRB for increasing SNRs whereas a floor will appear at high SNRs if a constant training sequence is used. For channel orders outside this interval, both methods show similar behaviors (constant slope for small channel orders and saturation for large channel orders), but there is still an advantage in using changing training sequences.

2) Changing Training Sequences: After this discussion on the CRB, we check how the proposed closed form channel estimates match this theoretical bound. We first check it for the approximate ML channel estimate proposed in the context of changing training sequences. In Fig. 2, we compare the simulated performance of our method with the corresponding CRB

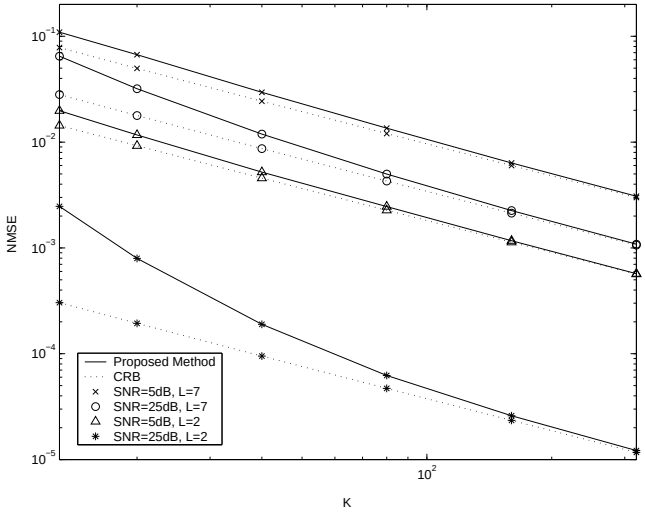

Fig. 3. Comparison of the simulated NMSE and the CRB vs. K for different channel orders when $n_t = 5$ and changing training sequences are used. The results are plotted for two different values of the SNR, namely 5 and $25dB$.

as a function of the SNR. We perform this comparison for two different channel orders: one for which the CRB has a constant slope, the other being large enough to have the CRB saturating at high SNRs. We repeat these experiments for two different values of K: a relatively small one and a larger one. We observe a relatively good match between the CRB and the experimental curves when the channel order is large and there is a floor in the CRB. The match is better for a larger K. When the channel order is small and there is no floor in the CRB, the theoretical and experimental curves match quite well at low SNRs but we see the emergence of a floor on the experimental NMSE for higher SNRs. The value of this floor decreases as the number of data blocks, K increases. In Fig. 3, we evaluate the impact of the number of data blocks K on the channel estimate NMSE. Here again, the simulations are done for two different channel orders. We further test the impact of the factor K for two different values of the SNR. Here again, we see that when the channel order is large, there is a relatively good match between the theoretical and experimental curves. When the channel order is smaller, there is a big difference between the theoretical and experimental values for small values of K when the SNR is large, which corresponds to the zone where the saturation occurs in Fig. 2. However, as K increases, the experimental NMSE and the CRB tend to merge. This difference between the CRB and the experimental results originates from the approximations we made in order to derive the

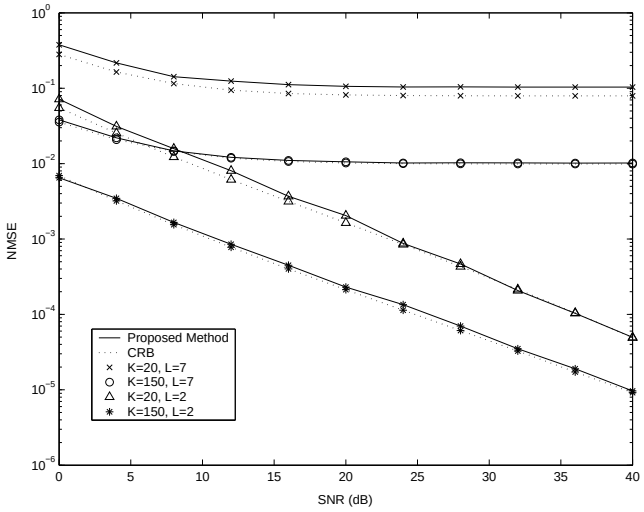

Fig. 4. Comparison of the simulated NMSE and the CRB vs. SNR for different channel orders when $n_t = 5$ and constant training sequences are used. The results are plotted for two different values of K, namely 20 and 150

approximate closed form ML channel estimate. These approximations do not hold when the SNR is large, K is small and the channel order is small.

However, when the iterative method is used, the channel estimate converges to $\mathbf{h}_{ML}$. It is indeed possible to check (not shown in the figures) that the gap between the CRB and the experimental results is closed after a few iterations. Hence, performing a few iterations allows us to avoid the saturation effect in the NMSE when the SNR is large and there are only a few training sequences available to perform channel estimation (small K).

3) Constant Training Sequences: In Fig. 5 and Fig. 4, we performed a similar analysis for the closed form ML channel estimate in the context of a constant training sequence. The figures show us that there is no significant difference between the CRB and the experimental results, except for very small values of K. This improved match between the CRB and the experimental results originates in the fact that we did not need to make any approximation when deriving the expression of the ML channel estimate in this case. Note that there is no point in using the iterative method in this context, since the closed form channel estimate corresponds to its convergence point, which is confirmed by experimental results (not shown in the figures).

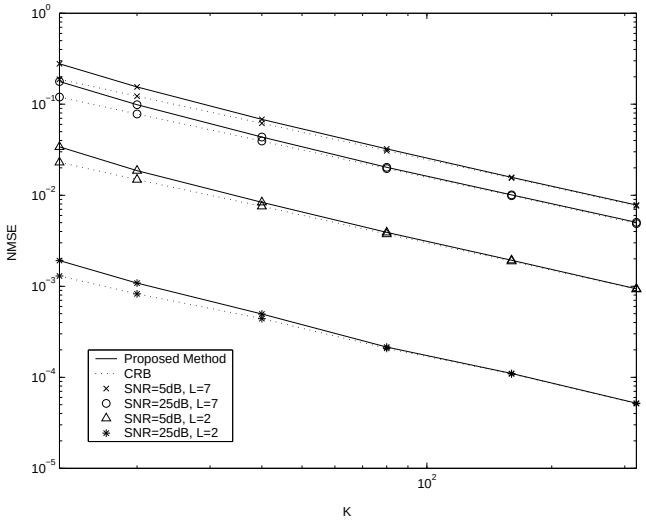

Fig. 5. Comparison of the simulated NMSE and the CRB vs. K for different channel orders when $n_t = 5$ and constant training sequences are used. The results are plotted for two different values of the SNR, namely 5 and $25dB$.

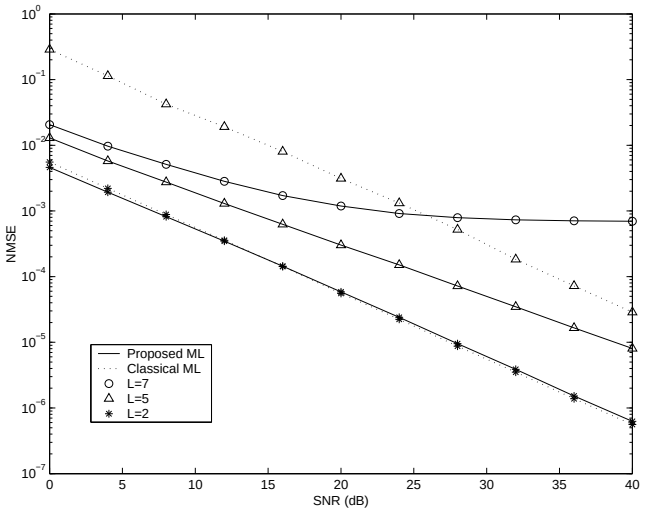

Fig. 6. Simulated NMSE vs. SNR for the proposed ML method and traditional ML channel estimation for different channel orders when constant training sequences of length $n_t = 11$ are used and the number of observed blocks is set to $K = 100$.

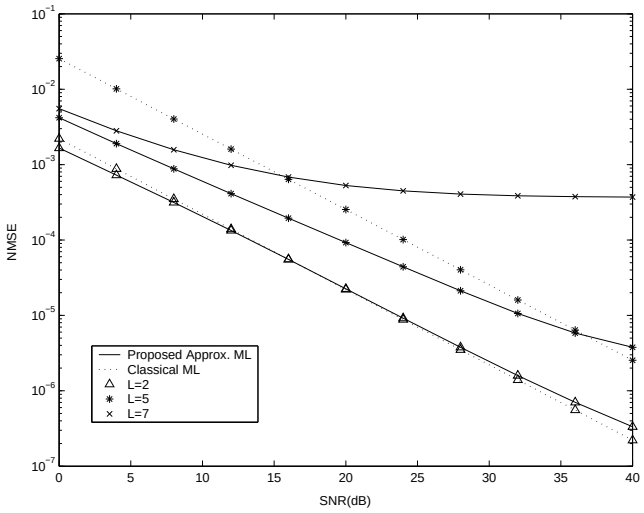

Fig. 7. Simulated NMSE vs. SNR for the proposed ML method and traditional ML channel estimation for different channel orders when changing training sequences of length $n_t = 6$ are used and the number of observed blocks is set to $K = 100$.

B. Comparison with Traditional ML Methods

Classical training-based ML channel estimation techniques solely rely on the part of the received symbols that only contain contributions from the known training symbols. They simply discard the received samples that are corrupted by contributions from the unknown data symbols. Such symbols can be observed at the receiver only when $n_t \geqslant L+1$. In that case, based on the data model we derived in section II, we can derive a data model that focuses on the received symbols that do not contain any contributions from the unknown data symbols. Define $\mathbf{u}'_k = \mathbf{u}_k(L+1 : n_t)$ and $\mathbf{T}'_k = \mathbf{T}_k(L+1 : nt, :)$. We then have:

$$\mathbf{u}'_k = \mathbf{T}'_k\mathbf{h} + \boldsymbol{\eta}'_k,$$

where $\boldsymbol{\eta}'_k$ is the AWGN term. Note that $\mathbf{T}'_k$ is an $(n_t - L) \times (L + 1)$ Toeplitz matrix with $[t_k[L + 1], \ldots, t_k[n_t]]^T$ as first column and $[t_k[L + 1], \ldots, t_k[1]]$ as first row.

When the noise term is not colored, the solution of the ML channel identification problem is well known to be a simple least squares fit of $\mathbf{T}'_k\mathbf{h}$ to $\mathbf{u}'_k$.

In the changing training sequences case, the classical ML channel estimate is thus given by

$$\mathbf{h}'_{ML} = \left(\sum_{k=1}^{K} {\mathbf{T}'}_k^H \mathbf{T}'_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}'_k \mathbf{u}'_k$$

Relying on the randomness of the training sequences, the inverse of the summation in this expression will exist with probability one as soon as $K \geqslant L+1$.

When a constant training sequence is used the condition on the channel order is more stringent. If we want to have a solution to our LS problem, we need to have as many equations as there are coefficients to identify in the channel model. 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, the constant $\mathbf{T}'$ matrix is tall and we can find a LS solution to derive our ML channel estimate:

$$\mathbf{h}'_{ML} = K^{-1} \left({\mathbf{T}'}^H \mathbf{T}' \right)^{-1} \sum_{k=1}^{K} \mathbf{T}' \mathbf{u}'_k.$$

Note that the conditions that relate the training sequences length n_t to the channel order L are quite stringent: it is simply impossible to identify the channel when these are not fulfilled. We can now compare the results obtained with these classical ML channel estimates with the proposed ML estimates. In Fig. 6, we consider the constant training sequences case and we analyze the changing training sequences case in Fig. 7. We compare the results for different channel orders when the length of the training sequences n_t is fixed.

In both situations, we see that the new method and the traditional one yield equivalent performance for small channel orders. When the channel order increases, the new method outperforms the classical one, especially at low SNRs. The only situation where this is not the case is for large SNR values when changing training sequences are used. We know however that increasing the number of observed blocks K or performing a few iterations would solve this problem. When the channel order keeps growing, the new method still provides reliable channel estimates whilst traditional methods cannot be applied anymore, which is a major advantage of the proposed method.

VIII. Conclusions

In this paper, we presented a new training based ML channel identification method. We analyzed two situations: the situation where the same training sequence is repeated at the end of each data block (constant training sequences case) and the situation where this training sequence

is changed at the end of each data block (changing training sequences case). We first proposed an iterative ML method and then derived closed form expressions for the ML channel estimates.

This new ML method clearly 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 channel estimation methods.

Experimental results also show that the use of changing training sequences yields more accurate channel estimates than constant training sequences.

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] S. Barbarossa, A. Scaglione and G. B. Giannakis, "Performance Analysis of a Deterministic Channel Estimator for Block Transmission Systems with Null Guard Intervals," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 684–695, March 2002.

[4] 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.

[5] 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.

[6] O. Rousseaux, G. Leus, M. Moonen, "A suboptimal Iterative Method for Maximum Likelihood Sequence Estimation in a Multipath Context," *EURASIP Journal on Applied Signal Processing (JASP)*, vol. 2002, no. 12, pp. 1437–1447, December 2002.

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

[8] O. Rousseaux, G. Leus, P. Stoica and M. Moonen, "A Stochastic Method for Training Based Channel Identification," in *Seventh International Symposium on Signal Processing and its Applications (ISSPA 2003)*, Paris, France, July 2003, submitted.

APPENDIX A

In this appendix, we show that

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H. \tag{35}$$

The proof is adapted from [7, pp.201-202]. Let us first define the sample covariance matrix:

$$\mathbf{R}_K(\mathbf{h}) \triangleq K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H.$$

Using this definition, the log-likelihood function (5) can be re-expressed as

$$-\mathcal{L} = \frac{K}{2} \left(\text{tr} \left(\mathbf{R}_K(\mathbf{h}) \mathbf{Q}^{-1} \right) + \ln|\mathbf{Q}| \right) + \text{cst} \tag{36}$$

The proposed ML estimate defined in (35) is equivalent to $\mathbf{Q}_{ML}(\mathbf{h}) = \mathbf{R}_K(\mathbf{h})$. Claiming that (36) is minimized with respect to $\mathbf{Q}$ for $\mathbf{Q} = \mathbf{R}_K(\mathbf{h})$ is equivalent to claiming that

$$\text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) + \ln|\mathbf{Q}| \geqslant \text{tr} \left(\mathbf{R}\mathbf{R}^{-1} \right) + \ln|\mathbf{R}|, \qquad \forall \;\; \mathbf{Q} = \mathbf{Q}^H > 0.$$

where $\mathbf{R}$ is a shorthand notation for $\mathbf{R}_K(\mathbf{h})$. The following equivalences are easily derived:

$$\begin{aligned}
& \text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) + \ln|\mathbf{Q}| \geqslant \text{tr} \left(\mathbf{I}_{n_t+L} \right) + \ln|\mathbf{R}| \\
\Leftrightarrow \quad & \text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) + \ln \left(|\mathbf{Q}| / |\mathbf{R}| \right) \geqslant n_t + L \\
\Leftrightarrow \quad & \text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) - \ln|\mathbf{R}\mathbf{Q}^{-1}| \geqslant n_t + L.
\end{aligned} \tag{37}$$

It is clear from its definition that $\mathbf{R}$ can be factorized in a full rank square matrix and its complex conjugate transpose: $\mathbf{R} = \mathbf{G}\mathbf{G}^H$. Define next the matrix $\mathbf{F} \triangleq \mathbf{G}^H \mathbf{Q}^{-1} \mathbf{G}$. This matrix $\mathbf{F}$ is symmetric and positive definite. Its eigenvalues $\lambda_1 \ldots \lambda_{n_t+L}$ clearly satisfy $\lambda_i > 0$. Using these new definitions, we can proceed and rephrase (37) as:

$$\begin{aligned}
& \text{tr} \left(\mathbf{G}\mathbf{G}^H \mathbf{Q}^{-1} \right) - \ln|\mathbf{G}\mathbf{G}^H \mathbf{Q}^{-1}| \geqslant n_t + L \\
\Leftrightarrow \quad & \text{tr} \left(\mathbf{G}^H \mathbf{Q}^{-1} \mathbf{G} \right) - \ln|\mathbf{G}^H \mathbf{Q}^{-1} \mathbf{G}| \geqslant n_t + L \\
\Leftrightarrow \quad & \text{tr}(\mathbf{F}) - \ln|\mathbf{F}| \geqslant n_t + L \\
\Leftrightarrow \quad & \sum_{i=1}^{n_t+L} \lambda_i - \ln \prod_{i=1}^{n_t+L} \lambda_i \geqslant n_t + L \\
\Leftrightarrow \quad & \sum_{i=1}^{n_t+L} \lambda_i - \ln\lambda_i - 1 \geqslant 0
\end{aligned}$$

Since $\forall \lambda > 0, \quad ln\lambda \leqslant \lambda - 1$, the last inequality is obviously true. This concludes the proof.

APPENDIX B

In this appendix, we show that

$$\left|\mathbf{I}+\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right)\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}\right| = 1+\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right).$$

Let us first perform an eigenvalue decomposition of the product $\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right)\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}$:

$$\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right)\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{H}, \tag{38}$$

where $\mathbf{U}$ is a unitary matrix and $\mathbf{\Sigma}$ is a diagonal matrix with the eigenvalues λ_i in decreasing order on the main diagonal.

Because of the term $\left(\mathbf{Th}-\bar{\mathbf{u}}\right)\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}$, (38) has only rank one and there is a single non-zero eigenvalue. Since (38) is definite positive, this only eigenvalue is positive, it is thus λ_1. Since the trace of a matrix is the sum of the eigenvalues of this matrix, λ_1 is equal to the trace of the initial product (38), or equivalently:

$$\lambda_1 = \left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right).$$

It follows that:

$$\left|\mathbf{I}+\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right)\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}\right| = \left|\mathbf{I}+\mathbf{\Sigma}\right| = 1+\lambda_1.$$

This concludes the proof.