



HAL
open science

Analysis of the Maximum-likelihood Channel Estimator for OFDM Systems with Unknown Interference

Azzouz Dermoune, Eric Pierre Simon

► **To cite this version:**

Azzouz Dermoune, Eric Pierre Simon. Analysis of the Maximum-likelihood Channel Estimator for OFDM Systems with Unknown Interference. 2015. hal-01169591

HAL Id: hal-01169591

<https://hal.science/hal-01169591>

Preprint submitted on 29 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analysis of the Maximum-likelihood Channel Estimator for OFDM Systems with Unknown Interference

Azzouz Dermoune¹, Eric Pierre Simon²

Abstract—In this paper, a comprehensive theoretical analysis of the maximum-likelihood (ML) channel estimator for OFDM systems with unknown interference is presented. The uniqueness of the solution is analytically demonstrated when the number of transmitted OFDM symbols is strictly greater than one. This solution is derived from the iterative conditional ML (CML) algorithm. It is shown that the channel estimate can be described as an algebraic function whose inputs are the initial value and the means and variances of the received samples. It is also theoretically demonstrated that the channel estimator is not biased, and that the second moment exists as long as the number of transmitted OFDM symbols is strictly greater than three. Furthermore, this is confirmed by numerical results.

Index Terms—OFDM, interference, channel, estimation, cognitive-radio.

I. INTRODUCTION

The narrow band interference (NBI) arises in OFDM systems for a number of transmission scenarios, such as Wi-Fi communications [1] or cognitive radio where different types of wireless services may use the same frequency band. The NBI can affect several subcarriers. It is well known that it strongly degrades the performance of the receiver if it is not treated [2] [3]. When the transmission is NBI-free, the noise consists only of thermal noise, yielding a uniform noise variance for all subcarriers, resulting in the estimation of a single scalar parameter. However, in the presence of NBI, the noise originates from both thermal noise and interference. Due to the nature of the NBI, neither the number of affected subcarriers nor their location in the spectrum is known. This prompts the need for estimating the noise variance for each subcarrier, yielding a vector estimation, denoted σ^2 , rather than a scalar. The objective in the presence of NBI is therefore to estimate the set of parameters $\{\mathbf{h}, \sigma^2\}$, where \mathbf{h} is the vector containing the taps of the channel impulse response.

Several articles have been proposed to estimate the channel in the presence of NBI. In [4], the authors proposed an estimator employing a specific pilot structure which consists of two types of pilot symbols with different pilot density. More recently, in [5] the authors proposed a channel estimator based on a robust least-square approach. However, the proposed method imposes a number of pilots greater than twice the channel order defined as the number of taps.

In [6] and [7], the NBI is assumed to be Gaussian distributed in the frequency domain with zero-mean and unknown power.

Following the same assumption, Morelli and Moretti investigate the channel estimation with NBI in their seminal paper [8]. They also consider the case where any possible correlation between the interference over adjacent subcarriers is neglected. This case can be considered as the worst case since the correlation is an additional information that could be used to improve the estimation performance. After formulating the maximum likelihood (ML) algorithm for the joint estimation of $\{\mathbf{h}, \sigma^2\}$, they show that the solution is non-unique when the channel order (noted L) is greater than the number of transmitted OFDM symbols (noted K), leading to ambiguous channel estimates. This constraint is very restrictive since $K < L$ is a very common scenario in practice. For this reason, they suggest resorting to another algorithm, the expectation maximization (EM) algorithm with the complete data set $\{\mathbf{X}, \sigma^2\}$, where \mathbf{X} contains the received signal. This amounts to treating the noise variances as a nuisance random vector. The drawback of this approach is that it imposes the selection of a distribution for the random vector, the inverse gamma, and then the fixing of the distribution parameter through simulations. This can be a limitation for practical use.

In this paper, we first demonstrate that the ambiguities only appear when $K = 1$ and not $K < L$, which is much less restrictive. Hence if $K > 1$, the joint ML approach can be followed and there is no need to use the EM algorithm. But even the case $K = 1$ can be handled with a specific approach briefly outlined in this paper. Thus, these results open a wider field of application for the joint ML approach.

For the case $K > 1$, the likelihood equations are solved with the conditional ML (CML) algorithm ([9], [10]). Then, we develop a new theoretical framework for the CML by using an original formulation based on oblique projections. Precisely, within this framework, we are able to obtain the following novel results. It is proved that the channel estimator is unbiased. It is also shown that the second moment exists as long as $K > 3$. These results are of importance, in particular, for deriving the Cramer-Rao bound of this channel estimation problem. Moreover, the channel estimate is proved to be an algebraic function whose inputs are the initial value and the means and variances of the received samples.

Note that the authors of [11] also perform EM, but with the complete data set $\{\mathbf{X}, \mathbf{C}\}$ where \mathbf{C} is the set of data symbols. Here, as in [8], we focus on channel estimation based only on pilots, but we stick to the original joint ML estimation of

¹Laboratoire Paul Painlevé, USTL-UMR-CNRS 8524. UFR de Mathématiques, Bât. M2, 59655 Villeneuve d'Ascq Cédex, France. ² IEMN lab, TELICE group, University of Lille, FRANCE Corresponding Author: Azzouz.Dermoune@math.univ-lille1.fr

$\{\mathbf{h}, \sigma^2\}$, i.e. it is not assumed an a priori distribution for σ^2 .

To sum up, our contributions are as follows:

- we prove that the joint ML estimation of $\{\mathbf{h}, \sigma^2\}$ gives a unique solution as long as the number of transmitted OFDM symbols is strictly greater than one,
- we prove that the joint ML of $\{\mathbf{h}, \sigma^2\}$ is an algebraic function with the initial value, the sample means and the sample variances as inputs,
- we prove that the CML algorithm yields an unbiased channel estimator, and that the second moment exists as long as the number of transmitted OFDM symbols is strictly greater than three.

The paper is organized as follows. Section II describes the system model. In Section III we discuss the joint ML estimation of $\{\mathbf{h}, \sigma^2\}$ and the question of the uniqueness of the solution. Then, Section IV introduces the CML algorithm to find the solution. A theoretical study of the CML is provided in Section V. The Cramer Rao bound is derived in Section VI and simulation results are presented in Section VII.

Notations: The field of complex numbers is denoted \mathcal{C} . Matrices [vectors] are denoted with upper [lower] case boldface letters (e.g. \mathbf{A} , or \mathbf{a}). The complex number $a_{i,j}$ indicates the (i,j) th entry of matrix \mathbf{A} ; a_i indicates the i th entry of vector \mathbf{a} . Vector \mathbf{a}_i is the i th row vector of matrix \mathbf{A} . The $N \times N$ identity matrix is denoted by \mathbf{I}_N and $\mathbf{0}_{M,N}$ is the $M \times N$ matrix of zeros. The matrix $\mathbf{D}(\mathbf{x})$ is a diagonal matrix with vector \mathbf{x} on its main diagonal. The superscripts $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, $(\cdot)^R$ and $(\cdot)^I$ stand respectively for transpose, Hermitian, complex conjugate, real part and imaginary part operators. The mathematical expectation is denoted $\mathbb{E}[\cdot]$. The multivariate complex normal distribution of a P-dimensional random vector is denoted as $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is the P-dimensional mean vector and $\boldsymbol{\Sigma}$ the $P \times P$ covariance matrix. The chi-square distribution with k degrees of freedom is denoted as χ_k^2 . The notations $\mathcal{R}(\mathbf{A})$ and $\mathcal{N}(\mathbf{A})$ indicate respectively the range space and the null space of \mathbf{A} .

II. SYSTEM MODEL

Let us consider an OFDM system with N subcarriers, and a cyclic prefix length N_g . We assume that the channel between the transmitter and the receiver is modelled as a frequency-selective fading channel with a channel impulse response (CIR) vector \mathbf{h} of order L , $\mathbf{h} = [h_1, \dots, h_L]^T$. The CIR \mathbf{h} is assumed to be static over the transmission of K OFDM symbols. To estimate the channel, P pilot symbols with constant energy are inserted into the N sub-carriers at the positions $\mathcal{P} = \{n_p, p = 1, \dots, P\}$. In our work, we do not consider a particular pilot scheme \mathcal{P} , and all our derivations could be applied to any \mathcal{P} . The only constraint is that $L < P$. The received frequency-domain pilot sample of the k th OFDM symbol at the n_p subcarrier is:

$$x_{p,k} = c_{p,k}H_p + w_{p,k}, \quad (1)$$

where $c_{p,k}$ is the pilot symbol with normalized power transmitted on the n_p th subcarrier and $w_{p,k}$ is the disturbance term that takes into account the background noise plus any possible interference. The random complex number $w_{p,k}$ is assumed

to be Gaussian distributed with zero mean and unknown variance $\sigma_p^2 = \sigma_{TN}^2 + \sigma_{NBI,p}^2$, where σ_{TN}^2 is the thermal noise contribution and $\sigma_{NBI,p}^2$ is the average NBI power assumed constant over the transmission period. The channel frequency response H_p at the n_p th subcarrier is given by:

$$H_p = \sum_{l=1}^L h_l \exp(-j \frac{2\pi n_p(l-1)}{N}), \quad p = 1, \dots, P. \quad (2)$$

This yields the model for the P pilot subcarriers of the k th received OFDM block:

$$\mathbf{x}_k = \mathbf{C}_k \mathbf{F} \mathbf{h} + \mathbf{w}_k, \quad k = 1, \dots, K, \quad (3)$$

with $\mathbf{x}_k = [x_{1,k}, \dots, x_{P,k}]^T$, $\mathbf{w}_k = [w_{1,k}, \dots, w_{P,k}]^T$, $\mathbf{C}_k = \mathbf{D}([c_{1,k}, \dots, c_{P,k}])$ where $\mathbf{D}(\mathbf{u})$ is the diagonal matrix with the entries of vector \mathbf{u} on its diagonal, and \mathbf{F} is the $P \times L$ matrix with the (p,l) th entry defined as $\exp(-j \frac{2\pi n_p(l-1)}{N})$, $p = 1, \dots, P, l = 1, \dots, L$.

ML estimation of the set of unknown parameters $\{\mathbf{h}, \sigma^2\}$ is desired, where $\sigma^2 = [\sigma_1^2, \dots, \sigma_P^2]^T$, based on the set of received samples $\{\mathbf{y}_k = \mathbf{C}_k^{-1} \mathbf{x}_k, k = 1, \dots, K\}$.

Let us now define the sample means and the sample variances of the received samples, which will be used in the rest of the paper. The sample mean vector is denoted as $\bar{\mathbf{y}} = [\bar{y}_1, \dots, \bar{y}_P]^T$, where for $p = 1, \dots, P$,

$$\bar{y}_p = \frac{1}{K} \sum_{k=1}^K y_{p,k}. \quad (4)$$

The sample variance vector is denoted as $\mathbf{s}^2(\mathbf{y}) = [s_1^2(\mathbf{y}), \dots, s_P^2(\mathbf{y})]^T$, where for $p = 1, \dots, P$,

$$s_p^2(\mathbf{y}) = \frac{1}{K} \sum_{k=1}^K |y_{p,k} - \bar{y}_p|^2. \quad (5)$$

III. MAXIMUM LIKELIHOOD ESTIMATION

In this section, the ML estimate of $\{\mathbf{h}, \sigma^2\}$ will be investigated by following the approach presented in [8]. However, it will be shown that the ambiguities mentioned in [8] appear only when $K = 1$. Hence, when $K \geq 2$, it will be possible to get the ML solution without ambiguities. The estimates when $K \geq 2$ will be derived through the conditional ML and their properties studied in the next section.

Recall that the K independent observations $\mathbf{y}_1, \dots, \mathbf{y}_K$ are drawn from the following p -variate normal regression model

$$\mathcal{CN}(\mathbf{F}\mathbf{h}, \mathbf{D}(\boldsymbol{\sigma}^2)). \quad (6)$$

Then, the negative log-likelihood function $\ell(\mathbf{h}, \boldsymbol{\sigma}^2)$ is given by [8]:

$$\ell(\mathbf{h}, \boldsymbol{\sigma}^2) = K \sum_{p=1}^P \ln(\pi \sigma_p^2) + \sum_{k=1}^K \sum_{p=1}^P \frac{|y_{p,k} - \mathbf{f}_p \mathbf{h}|^2}{\sigma_p^2}. \quad (7)$$

Here, \mathbf{f}_p denotes the p th row of the matrix \mathbf{F} .

The approach to derive the ML solution is summed up below. The variances which minimize (7) for a given \mathbf{h} are first calculated:

$$\sigma_p^2(\mathbf{h}) = \frac{1}{K} \sum_{k=1}^K |y_{p,k} - \mathbf{f}_p \mathbf{h}|^2. \quad (8)$$

Then, substituting $\sigma_p^2(\mathbf{h})$ for σ_p^2 in (7) yields:

$$\Lambda(\mathbf{h}) := \ell(\mathbf{h}, \boldsymbol{\sigma}^2(\mathbf{h})) = K \sum_{p=1}^P \ln(\pi \sigma_p^2(\mathbf{h})) + KP. \quad (9)$$

Finally, the ML estimate of the CIR vector \mathbf{h} is the one that minimizes $\Lambda(\mathbf{h})$. Special treatment is required due to the presence of the logarithm function in (9). Indeed, the values of \mathbf{h} for which $\sigma_p^2(\mathbf{h}) = 0$ make $\Lambda(\mathbf{h})$ tend to $-\infty$. The consequences on the uniqueness of the solution are explained in more detail in [8] where the authors show that the minimization leads to ambiguous channel estimates if $K < L$. Let us show that their assertion is true when $K = 1$, but is too restrictive when $2 \leq K \leq L$. To understand why, it can be observed that the equation $\sigma_p^2(\mathbf{h}) = 0$ (8) for a given p yields a linear system of K equations with L unknowns:

$$\mathbf{A}\mathbf{h} = [y_{p,1}, \dots, y_{p,K}]^T, \quad (10)$$

where the $K \times L$ matrix \mathbf{A} is built by stacking the row vectors \mathbf{f}_p . If $K = 1$, then the system (10) is underdetermined (one equation and L unknowns), yielding an infinite number of solutions. However, for $K > 1$, the specific structure of matrix \mathbf{A} has to be taken into account when solving the system. Thus, it is found that (10) has no solution because on the one hand the samples $y_{p,1}, \dots, y_{p,K}$ are all different since they are Gaussian and independent and on the other hand the rows of \mathbf{A} are all identical. Then it becomes obvious that for $K > 1$, $\sigma_p^2(\mathbf{h}) > 0$ for all p and for all \mathbf{h} . Therefore, the ML estimate of the CIR vector \mathbf{h} is well defined for $K > 1$.

This ambiguity issue appears more clearly with the formulation of $\sigma_p^2(\mathbf{h})$ based on the sample means and variances:

$$\sigma_p^2(\mathbf{h}) = s_p^2(\mathbf{y}) + |\bar{y}_p - \mathbf{f}_p \mathbf{h}|^2, \quad (11)$$

and when noticing that if $K = 1$, the sample variances $s_p^2(\mathbf{y}) = 0$ for all p . However, if $K > 1$, $s_p^2(\mathbf{y}) \neq 0$ for all p , making it impossible to set $\sigma_p^2(\mathbf{h})$ to zero.

This article is concerned with the case $K > 1$. However, it is worth mentioning that the case of $K = 1$ can still be handled with the following approach. It has been shown that it is meaningless to search the ML of $\boldsymbol{\sigma}^2$ in the domain $(0, +\infty)^P$. A possible solution is to restrict the parameter space by imposing a priori lower bounds of the form [12]

$$0 < \delta_p^2 \leq \sigma_p^2 \quad (12)$$

on the variances σ_p^2 . Let us define the vector $\boldsymbol{\delta}^2 = [\delta_1^2, \dots, \delta_P^2]^T$. The variances $\boldsymbol{\sigma}^2 \in \prod_{p=1}^P [\delta_p^2, +\infty)$ which minimize (7) for given \mathbf{h} are given by

$$\sigma_p^2(\mathbf{h}, \delta_p^2) = |y_p - \mathbf{f}_p \mathbf{h}|^2, \quad \text{if } |y_p - \mathbf{f}_p \mathbf{h}|^2 \geq \delta_p^2, \quad (13)$$

$$\sigma_p^2(\mathbf{h}, \delta_p^2) = \delta_p^2, \quad \text{if } |y_p - \mathbf{f}_p \mathbf{h}|^2 < \delta_p^2. \quad (14)$$

The vector $\boldsymbol{\sigma}^2(\mathbf{h}, \boldsymbol{\delta}^2)$ is substituted for $\boldsymbol{\sigma}^2$ in $\ell(\mathbf{h}, \boldsymbol{\sigma}^2)$ to obtain

$$\Lambda(\mathbf{h}, \boldsymbol{\delta}^2) = \sum_{p=1}^P \left(\ln(\pi \sigma_p^2(\mathbf{h}, \delta_p^2)) + \frac{|y_p - \mathbf{f}_p \mathbf{h}|^2}{\sigma_p^2(\mathbf{h}, \delta_p^2)} \right), \quad (15)$$

and then the CIR estimate is the one that minimizes $\Lambda(\mathbf{h}, \boldsymbol{\delta}^2)$.

As previously stated, this article will focus on $K > 1$. As there is no closed-form solution for the minimization of $\Lambda(\mathbf{h})$, we suggest using the conditional ML in the next section to find an iterative solution and study the properties of this solution.

IV. CONDITIONAL ML (CML)

The CML is the result of two nested minimizations. First, (7) is minimized given the channel \mathbf{h} , yielding the estimation of $\boldsymbol{\sigma}^2$ given by (8) or (11):

$$\boldsymbol{\sigma}^2(\mathbf{h}) = \mathbf{s}^2(\mathbf{y}) + |\bar{\mathbf{y}} - \mathbf{F}\mathbf{h}|^2, \quad (16)$$

where the vector $|\bar{\mathbf{y}} - \mathbf{F}\mathbf{h}|^2 := [|\bar{y}_1 - \mathbf{f}_1 \mathbf{h}|^2, \dots, |\bar{y}_P - \mathbf{f}_P \mathbf{h}|^2]^T$. Conversely, (7) is minimized given $\boldsymbol{\sigma}^2$, yielding the estimation of \mathbf{h} :

$$\mathbf{h}(\boldsymbol{\sigma}^2) = (\mathbf{F}^H \mathbf{D}^{-1}(\boldsymbol{\sigma}^2) \mathbf{F})^{-1} \mathbf{F}^H \mathbf{D}^{-1}(\boldsymbol{\sigma}^2) \bar{\mathbf{y}}, \quad (17)$$

Hence the CML algorithm: from an initial estimate $\boldsymbol{\sigma}^{2(i)}$ of $\boldsymbol{\sigma}^2$, the estimate of \mathbf{h} is

$$\mathbf{h}^{(i+1)} = (\mathbf{F}^H \mathbf{D}^{-1}(\boldsymbol{\sigma}^{2(i)}) \mathbf{F})^{-1} \mathbf{F}^H \mathbf{D}^{-1}(\boldsymbol{\sigma}^{2(i)}) \bar{\mathbf{y}}, \quad (18)$$

yielding

$$\boldsymbol{\sigma}^{2(i+1)} = \mathbf{s}^2(\mathbf{y}) + |\bar{\mathbf{y}} - \mathbf{F}\mathbf{h}^{(i+1)}|^2. \quad (19)$$

This algorithm is known as the scoring method [9] [10] or the conditional maximum likelihood algorithm (CML) [13]. The CML's important properties, which will be used in appendix B-C, include:

1) Given $\boldsymbol{\sigma}^{2(i)}$, the vector $\mathbf{h}^{(i+1)}$ maximizes the likelihood. Given $\mathbf{h}^{(i)}$, the vector $\boldsymbol{\sigma}^{2(i+1)}$ maximizes the likelihood.

2) The logarithmic means $\frac{1}{P} \sum_{p=1}^P \ln(\sigma_p^{2(i)})$, for $i = 0, 1, \dots$ is non-increasing. In other words, for all i ,

$$\sum_{p=1}^P \ln(\sigma_p^{2(i+1)}) \leq \sum_{p=1}^P \ln(\sigma_p^{2(i)}). \quad (20)$$

And $\sum_{p=1}^P \ln(\sigma_p^{2(i)})$ converges to some constant $\ln(c^*) \geq \sum_{p=1}^P \ln(s_p^2(\mathbf{y}))$.

In this section, the CML algorithm has been presented with some of its well known properties. However, to our knowledge, no work has been carried out about the moments of the CML solution in this particular context. The next section will address this topic.

V. CML ALGORITHM AND OBLIQUE PROJECTIONS

In this section, the first and the second moments of the estimators are investigated. To do so, an original formulation of the CML based on oblique projections is established. It will be shown that the CML algorithm can be viewed as a succession of oblique projections.

First, let us recall a few preliminaries about projections. For any invertible matrix Σ , observe that the matrix $\Pi(\Sigma) := \mathbf{F}(\mathbf{F}^H \Sigma \mathbf{F})^{-1} \mathbf{F}^H \Sigma$ splits the space \mathcal{C}^P on two subspaces: the range space $\mathcal{R}(\Pi(\Sigma)) = \Pi(\Sigma)(\mathcal{C}^P)$ of $\Pi(\Sigma)$ and its nullspace $\mathcal{N}(\Pi(\Sigma)) = [\mathbf{I} - \Pi(\Sigma)](\mathcal{C}^P)$. Note that the range of $\Pi(\Sigma)$ is the range of \mathbf{F} . The linear operator defined by $\Pi(\Sigma)$ is known as an oblique projection on $\mathcal{R}(\mathbf{F})$. If $\Sigma = \mathbf{I}$, then

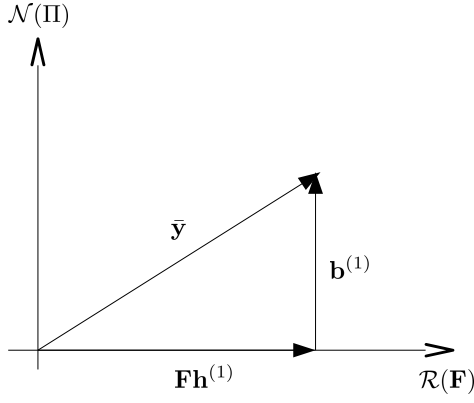


Fig. 1. First iteration of the CML algorithm : orthogonal projection

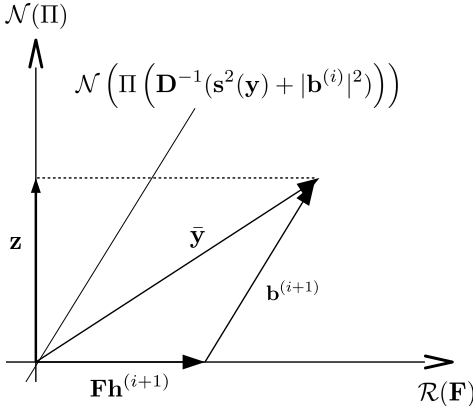


Fig. 2. Iteration i of the CML algorithm : oblique projection

$\Pi(\mathbf{I}) = \mathbf{F}(\mathbf{F}^H\mathbf{F})^{-1}\mathbf{F}^H$ is the orthogonal projection on $\mathcal{R}(\mathbf{F})$. For simplicity of notation, $\Pi(\mathbf{I})$ is now denoted Π .

Now, let us go back to the CML algorithm. If the initialization $\sigma_p^{2(0)} = 1$ for all p , then the first iteration of CML algorithm is given from (18) by:

$$\mathbf{h}^{(1)} = (\mathbf{F}^H\mathbf{F})^{-1}\mathbf{F}^H\bar{\mathbf{y}}. \quad (21)$$

Note that $\mathbf{h}^{(1)}$ is the ordinary least square estimate of \mathbf{h} . The orthogonal projection Π splits the sample mean in the following two components (see Fig. 1 for a geometrical interpretation):

$$\mathbf{F}\mathbf{h}^{(1)} = \Pi\bar{\mathbf{y}}, \quad \mathbf{b}^{(1)} := (\mathbf{I} - \Pi)\bar{\mathbf{y}}.$$

The vector $\mathbf{b}^{(1)}$ is the orthogonal projection of $\bar{\mathbf{y}}$ on $\mathcal{N}(\Pi)$.

Given $\mathbf{h}^{(1)}$, the ML of the variance vector is given by

$$\sigma^{2(1)} = \mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(1)}|^2,$$

where the column vector $|\mathbf{b}^{(1)}|^2 = [|b_1^{(1)}|^2, \dots, |b_P^{(1)}|^2]^T$. We have at the $i+1$ -th iterations

$$\mathbf{h}^{(i+1)} = (\mathbf{F}^H\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2)\mathbf{F})^{-1} \mathbf{F}^H\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2)\bar{\mathbf{y}}.$$

The oblique projection $\Pi(\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2))$ splits the sample mean in the following two components (see Fig. 2):

$$\begin{aligned} \mathbf{F}\mathbf{h}^{(i+1)} &= \Pi(\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2))\bar{\mathbf{y}}, \\ \mathbf{b}^{(i+1)} &= \bar{\mathbf{y}} - \mathbf{F}\mathbf{h}^{(i+1)} \\ &= (\mathbf{I} - \Pi(\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2)))\bar{\mathbf{y}}. \end{aligned}$$

Given $\mathbf{h}^{(i+1)}$, the ML of the variance vector is given by

$$\sigma^{2(i+1)} = \mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i+1)}|^2.$$

From this discussion, it can be concluded that the solution of the CML algorithm can be searched either in $\mathbf{F}\mathbf{h}^{(i)}$ or in the variable $\mathbf{b}^{(i)} = \bar{\mathbf{y}} - \mathbf{F}\mathbf{h}^{(i)}$. In the perspective of calculating the bias, it will be more convenient to consider $\mathbf{b}^{(i)}$ (see (26) and (27)), which yields the equivalent algorithm for solving the CML:

$$\mathbf{b}^{(i+1)} = (\mathbf{I} - \Pi(\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2)))\bar{\mathbf{y}}. \quad (22)$$

A. Properties of algorithm (22)

First, let us define $\mathbf{z} = (\mathbf{I} - \Pi)\bar{\mathbf{y}}$, the projection of $\bar{\mathbf{y}}$ on $\mathcal{N}(\Pi)$ (see Fig. 2). Unlike $\bar{\mathbf{y}}$, \mathbf{z} is zero-mean, which will be used for the calculation of the bias. We also define function φ from (22) as follows:

$$\varphi(\bar{\mathbf{y}}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}^{(i)}) := (\mathbf{I} - \Pi(\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2)))\bar{\mathbf{y}}, \quad (23)$$

and let $\Pi = [\pi_{ij} : i, j = 1, \dots, P]$ be the entries of the matrix Π and $c_{ljp} = \pi_{pl}\pi_{lj}$ for $l, j, p = 1, \dots, P$.

Now, the following properties of algorithm (22) are derived in order to subsequently calculate the bias.

Proposition 5.1:

1) For $i \geq 0$,

$$\mathbf{b}^{(i+1)} = \varphi(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}^{(i)}). \quad (24)$$

2) The map φ from \mathcal{C}^P to \mathcal{C}^P is rational. More precisely, for $p = 1, \dots, P$, the component φ_p of φ

$$\begin{aligned} \mathbf{b} \in \mathcal{C}^P \rightarrow \varphi_p(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}) &= z_p - \sum_{l=1}^P c_{lpp} z_l \\ &\quad - \sum_{l=1, j=1, l \neq j}^P c_{ljp} z_j \frac{s_l^2(\mathbf{y}) + |b_l|^2}{s_j^2(\mathbf{y}) + |b_j|^2} \\ &= \frac{\sum_{j=1}^P z_j f_{p,j}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}|^2)}{\prod_{l=1}^P (s_l^2(\mathbf{y}) + |b_l|^2)}, \end{aligned} \quad (25)$$

where $f_{p,j}$ are polynomials of degree P .

3) For $p = 1, \dots, P$,

$$\begin{aligned} \mathbf{b}^{(i+1)} &= [\varphi(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \cdot) \circ \dots \circ \varphi(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \cdot)](\mathbf{b}^{(i)}) \\ &:= \varphi^{(i)}(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}^{(i)}) \\ &= \frac{\sum_{j=1}^P z_j f_{p,j}^{(i)}(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), |\mathbf{b}^{(i)}|^2)}{g_p^{(i)}(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), |\mathbf{b}^{(i)}|^2)}, \end{aligned}$$

where $f_{p,j}^{(i)}, g_p^{(i)}$ are polynomials, even with respect to \mathbf{z} and $g_p^{(i)} > 0$.

The proofs are given in appendix A. Note that proposition 5.1 1) is straightforward with the geometrical interpretation of Fig. 2, where it can be observed that the projections of \mathbf{z} and $\bar{\mathbf{y}}$ on $\mathcal{N}(\Pi(\mathbf{D}^{-1}(\mathbf{s}^2(\mathbf{y}) + |\mathbf{b}^{(i)}|^2)))$ are the same.

B. Asymptotic Properties of algorithm (22)

Proposition 5.2: 1) Any limit (say) $\omega(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}^{(1)})$ of the sequence $\mathbf{b}^{(i+1)}$ is a root of the rational map

$$\mathbf{b} - \varphi(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}).$$

2) The map $(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}^{(1)}) \rightarrow \omega(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}^{(1)})$ is an algebraic function, i.e. for $p = 1, \dots, P$, there exist polynomial functions Q_p such that $Q_p(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \omega(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{b}^{(1)})) = 0$.

The proofs are a consequence of Proposition 5.1.

C. Mean and variance of CML estimator

At this stage, it is important to recall that our objective is now to calculate the bias of the CML solution. At the convergence of the algorithm, we obtain $\hat{\mathbf{h}}$, $\hat{\mathbf{b}}$ and $\hat{\sigma}^2$. From $\mathbf{F}\hat{\mathbf{h}} = \hat{\mathbf{y}} - \hat{\mathbf{b}}$ and (37), we have:

$$\mathbb{E}[\hat{\mathbf{h}}] = \mathbf{h} - (\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H \mathbb{E}[\hat{\mathbf{b}}]. \quad (26)$$

From $\hat{\sigma}^2 = \mathbf{s}^2(\mathbf{y}) + |\hat{\mathbf{b}}|^2$ and (38), we have:

$$\mathbb{E}[\hat{\sigma}^2] = \frac{K-1}{K} \sigma^2 + \mathbb{E}[|\hat{\mathbf{b}}|^2], \quad (27)$$

using the property that the mean of a chi-square random variable of n degrees of freedom is n . Now, from Proposition (5.1), the following can be shown (see Appendix B for the proof).

Corollary 5.3:

1) Given $\mathbf{s}^2(\mathbf{y})$, and if $\mathbf{b}^{(1)} = (\mathbf{I} - \Pi)\bar{\mathbf{y}}$, then the vector $\mathbf{b}^{(i+1)}$ is zero-mean, i.e.

$$\mathbb{E}[\mathbf{b}^{(i+1)} | \mathbf{s}^2(\mathbf{y})] = 0.$$

Therefore, from (26), $\hat{\mathbf{h}}$ is an unbiased estimator.

2) If $\mathbf{b}^{(1)} = (\mathbf{I} - \Pi)\bar{\mathbf{y}}$, and the sample size $K \geq 4$, then, for $i \geq 1$, $\mathbf{b}^{(i)}$ and its limit $\hat{\mathbf{b}}$ have a second moment.

VI. CRAMER-RAO BOUND

Let us define the $2L + P \times 1$ vector of the real parameters to be estimated:

$$\boldsymbol{\theta} = [h_1^R, \dots, h_L^R, h_1^I, \dots, h_L^I, \sigma_1^2, \dots, \sigma_P^2]^T, \quad (28)$$

where h_l^R, h_l^I are respectively the real and imaginary parts of h_l .

The Cramer-Rao bound (CRB) states that the covariance matrix of $\hat{\boldsymbol{\theta}}$ satisfies

$$\text{cov}(\hat{\boldsymbol{\theta}}) \geq \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{J}^{-1} \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T, \quad (29)$$

where

$$\psi(\boldsymbol{\theta}) = \mathbb{E}[\hat{\boldsymbol{\theta}}] \quad (30)$$

and $\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is the Jacobian matrix whose (n, m) th element is given by $\frac{\partial \psi_n(\boldsymbol{\theta})}{\partial \theta_m}$. Matrix \mathbf{J} is the $2L + P \times 2L + P$ Fisher information matrix. Its (m, n) th entry is defined as $\mathbb{E} \left[\frac{\partial^2}{\partial \theta_n \partial \theta_m} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) \right]$, where $\ell(\mathbf{h}, \boldsymbol{\sigma}^2)$ is the negative log likelihood defined in (7).

Therefore, it is seen from (30) that knowing the moments of the estimator is required in order to calculate the CRB. The results of Section V-C will be exploited to do so.

The results of the calculation of \mathbf{J} are given below, and the details are in appendix C. Matrix \mathbf{J} can be written as:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_h & \mathbf{0}_{2L, P} \\ \mathbf{0}_{P, 2L} & \mathbf{J}_{\sigma^2} \end{bmatrix} \quad (31)$$

where $\mathbf{J}_{\sigma^2} = K \mathbf{D} \left(\frac{1}{\sigma_1^4}, \dots, \frac{1}{\sigma_P^4} \right)$ and where the entries of \mathbf{J}_h are defined by (43), (44) and (45). To compute the CRB, the calculation of the inverse of \mathbf{J} is required. Since the inverse of a block diagonal matrix is the block diagonal matrix of the inverses, as long as the submatrices are invertible, we have

$$\mathbf{J}^{-1} = \begin{bmatrix} \mathbf{J}_h^{-1} & \mathbf{0}_{2L, P} \\ \mathbf{0}_{P, 2L} & \mathbf{J}_{\sigma^2}^{-1} \end{bmatrix}. \quad (32)$$

From (26), (27) and corollary 5.3 we express $\psi(\boldsymbol{\theta})$ as a function of \mathbf{h} and $\boldsymbol{\sigma}^2$ and we calculate the derivative:

$$\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{I}_{2L} & \mathbf{0}_{2L \times P} \\ \mathbf{0}_{P \times 2L} & \mathbf{M} \end{bmatrix} \quad (33)$$

where the (p_1, p_2) th entry of \mathbf{M} is defined as $\frac{K-1}{K} \delta_{p_1}^{p_2} + \mathbb{E} \left[\frac{\partial |\hat{b}_{p_2}|^2}{\partial \sigma_{p_1}^2} \right]$. Therefore, we have:

$$\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{J}^{-1} \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T = \begin{bmatrix} \mathbf{J}_h^{-1} & \mathbf{0}_{2L \times P} \\ \mathbf{0}_{P \times 2L} & \mathbf{M} \mathbf{J}_{\sigma^2}^{-1} \mathbf{M}^T \end{bmatrix}. \quad (34)$$

Note that the calculation of $\mathbb{E} \left[\frac{\partial |\hat{b}_{p_2}|^2}{\partial \sigma_{p_1}^2} \right]$ is not feasible since there is no analytical expression for $\hat{\mathbf{b}}$. Therefore, the bound for the variance estimation can not be found. However, the bound for the channel estimation is given from (28) and (34) by:

$$\text{cov} \left(\left[\hat{h}_1^R, \dots, \hat{h}_L^R, \hat{h}_1^I, \dots, \hat{h}_L^I \right]^T \right) \geq \mathbf{J}_h^{-1}. \quad (35)$$

Recall that this bound has been derived by using the result of corollary 5.3 showing that $\hat{\mathbf{h}}$ was unbiased.

VII. SIMULATION RESULTS

In order to validate the results, computer simulations were carried out in accordance with the IEEE 802.11g standard with a carrier frequency equal to 2.4 GHz as in [8]. System parameters used for the simulation are as follows: $N = 64$, a bandwidth of 20 MHz and a cyclic prefix of length 16. The discrete-time channel is assumed to have $L = 6$ channel taps modelled with a Rayleigh channel with an exponentially decaying power such that: $\mathbb{E}[|h_l|^2] = \sigma_h^2 \cdot \exp(-l)$ with $l = 1, 2, \dots, L = 6$. The constant σ_h^2 is chosen to normalize the channel power to one. For the simulation, the pilots are evenly inserted every 8 subcarriers, yielding $P = 8$. A frame of $K = 4$ OFDM symbols is considered.

It is also assumed that 2 contiguous pilot sub-carriers are affected by NBI by adding a Gaussian disturbance of variance σ_I^2 . The signal-to-noise ratio (SNR) is defined as $10 \log \frac{1}{\sigma_I^2}$ where the power of the signal is normalized to one, and the signal-to interference ratio (SIR) is defined as $10 \log \frac{1}{\sigma_I^2}$. The accuracy of the channel estimates is measured

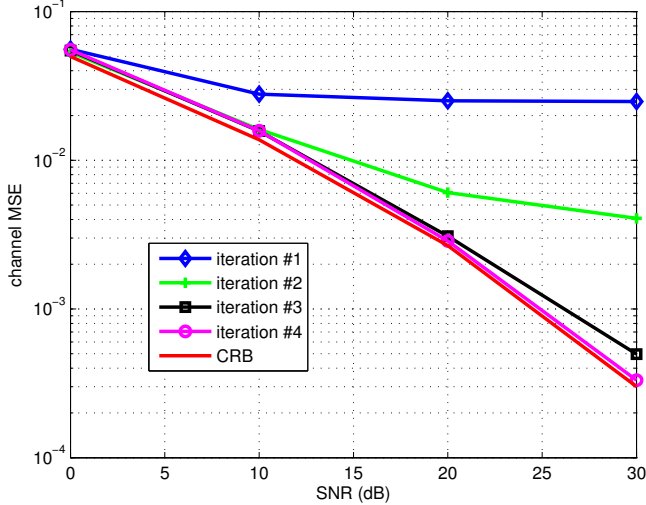


Fig. 3. Performance of the CML estimator as a function of the SNR for SIR = -5 dB

in terms of mean square error (MSE) which is defined as $\frac{1}{L} \mathbb{E} [(\hat{\mathbf{h}} - \mathbf{h})^H (\hat{\mathbf{h}} - \mathbf{h})]$ where the expectation is estimated via monte-carlo simulations in the following.

Fig. 3 shows the MSE as a function of the SNR when the SIR is fixed to -5 dB. Four iterations are considered. Let us recall that the first iteration is the ordinary least square estimate (OLSE) (see (21)). For reference, the CRB calculated in section VI is added. It is seen that the algorithm converges after four iterations, and nearly attains the CRB, whereas the MSE for the OLSE has a floor at 2.5×10^{-2} .

Then, the SIR is fixed to 0 dB in Fig. 4 and to 5 dB in Fig. 5. In these two cases, the algorithm converges after only three iterations. Moreover, it is observed that the performance of the OLSE approaches that of the second iteration when the SIR increases. It makes sense, as the OLSE is designed to perform well without NBI.

Finally, the bias of $\hat{\mathbf{h}}$ is studied. We plot the histogram of the real part and imaginary part of \hat{b}_p , $p = 1, \dots, P$ in Fig. 6 and Fig. 7, respectively. The SNR is fixed to 10 dB and the SIR to 0 dB. It can be observed that the mean is zero for all p , which leads to an unbiased estimator for $\hat{\mathbf{h}}$. This confirms corollary 5.3 1).

VIII. CONCLUSION

This article has addressed the problem of maximum likelihood channel estimation for OFDM systems in the presence of unknown interference. First, it was proved that the solution is without ambiguities as long as the number of transmitted OFDM symbols, denoted K , is strictly greater than one. For this case, we have proposed to use the conditional maximum likelihood (CML) algorithm to obtain the estimates. New theoretical developments to the CML algorithm in this context have been brought. It was proved that the solution of the algorithm is an algebraic function of the data. Furthermore, it was also proved that the channel estimator is unbiased and that the second moment exists as long as $K > 3$.

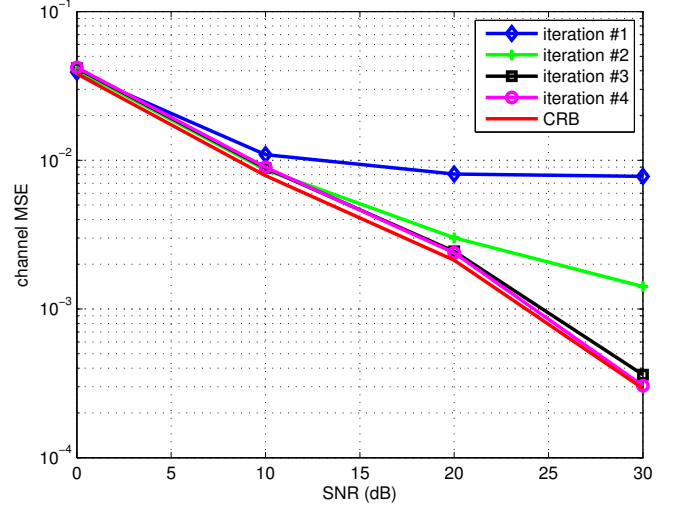


Fig. 4. Performance of the CML estimator as a function of the SNR for SIR = 0 dB

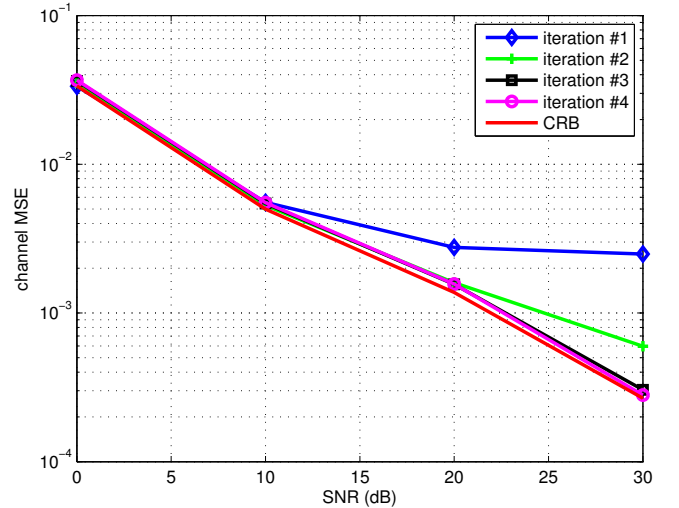


Fig. 5. Performance of the CML estimator as a function of the SNR for SIR = 5 dB

APPENDIX A PROOF OF PROPOSITION 5.1

The proof of proposition 5.1 1) is a consequence of the following general results. Let Σ be an invertible matrix. The equality

$$(\mathbf{I} - \Pi(\Sigma)) = (\mathbf{I} - \Pi(\Sigma))(\mathbf{I} - \Pi)$$

is equivalent to

$$\Pi = \Pi(\Sigma)\Pi.$$

Now, it can easily be shown that

$$\begin{aligned} \Pi(\Sigma)\Pi &= \mathbf{F}(\mathbf{F}^H \Sigma \mathbf{F})^{-1} \mathbf{F}^H \Sigma \mathbf{F} (\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H \\ &= \Pi. \end{aligned}$$

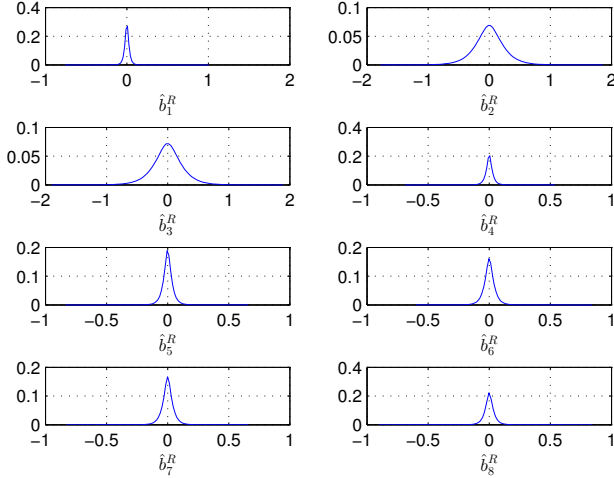


Fig. 6. Histogram of \hat{b}_p^R , $p = 1, \dots, P$, SNR = 10 dB, SIR = 0 dB

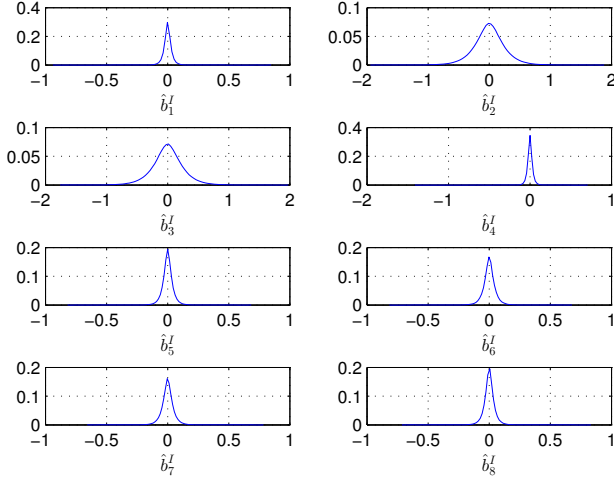


Fig. 7. Histogram of \hat{b}_p^I , $p = 1, \dots, P$, SNR = 10 dB, SIR = 0 dB

The proof of 2) is a consequence of the following general result:

$$(\mathbf{F}^H \Sigma^{-1} \mathbf{F})^{-1} = \mathbf{F}^+ \Sigma (\mathbf{F}^H)^+$$

where $\mathbf{F}^+ = (\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H$ and $(\mathbf{F}^H)^+ = \mathbf{F} (\mathbf{F}^H \mathbf{F})^{-1}$. From that we have

$$\Pi (\Sigma^{-1}) = \mathbf{F} \mathbf{F}^+ \Sigma (\mathbf{F}^H)^+ \mathbf{F}^H \Sigma^{-1} = \Pi \Sigma \Pi \Sigma^{-1}. \quad (36)$$

The proof of the remaining assertions is straightforward.

APPENDIX B PROOF OF COROLLARY (5.3)

A. Distribution probability of the sample mean and the sample variance

Cochran's theorem states that the sample mean and variance are two independent random variables. Moreover, it is also

stated that the sample variance of K independent normally distributed real random variables with mean 0 and standard deviation 1 has a chi-square distribution with $K - 1$ degrees of freedom. Therefore, we can write the distributions for $s_p^2(\mathbf{y})$. The distribution for \bar{y}_p is straightforward:

$$\bar{y}_p \sim \mathcal{CN}(\mathbf{f}_p \mathbf{h}, \frac{\sigma_p^2}{K}), \quad (37)$$

$$s_p^2(\mathbf{y}) \sim \frac{\sigma_p^2}{2K} \chi_{2(K-1)}^2. \quad (38)$$

Here we find $2(K-1)$ degrees of freedom since the considered random variables are complex.

B. Proof of assertion 1) of Corollary (5.3)

From Cochran's theorem, it can be deduced that the Gaussian vector $\mathbf{z} := (\mathbf{I} - \Pi)\bar{\mathbf{y}}$ is zero-mean with the covariance matrix $K^{-1}(\mathbf{I} - \Pi)\mathbf{D}(\sigma^2)$, i.e. $(\mathbf{I} - \Pi)\bar{\mathbf{y}} \sim (\mathbf{I} - \Pi)K^{-1/2}\mathbf{D}(\sigma)\mathcal{CN}(0, \mathbf{I})$. The components of the vector $2K\mathbf{s}^2(\mathbf{y})$ are independent with the distribution probability $(\sigma_1^2 \chi_{2(K-1)}^2(1), \dots, \sigma_P^2 \chi_{2(K-1)}^2(P))$. Here $\chi_{2(K-1)}^2(1), \dots, \chi_{2(K-1)}^2(P)$ are i.i.d. with the common distribution $\chi_{2(K-1)}^2$. From Cochran's theorem $\mathbf{s}^2(\mathbf{y})$ and $\bar{\mathbf{y}}$ are independent. From Proposition (5.1), the random vector $\mathbf{b}^{(i+1)} = \varphi^{(i)}(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{z})$ is rational function having a positive denominator. Therefore,

$$\mathbb{E}(\mathbf{b}^{(i+1)} | \mathbf{s}^2(\mathbf{y})) = \mathbb{E}(\varphi^{(i)}(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{z})) = 0$$

because \mathbf{z} is zero-mean and $\mathbf{z} \rightarrow \varphi^{(i)}(\mathbf{z}, \mathbf{s}^2(\mathbf{y}), \mathbf{z})$ is odd.

C. Proof of assertion 2) of Corollary (5.3)

Here, $\mathbf{b}^{(1)} = \mathbf{z}$ is again used. We wish to prove that the second moment exists, i.e. $\mathbb{E}[|b_p^{(i)}|^2] < +\infty$ for $p = 1, \dots, P$. The proof starts with:

$$\begin{aligned} \left(\mathbb{E}[|b_p^{(i)}|^2] \right)^2 &= \left(\mathbb{E} \left[\frac{|b_p^{(i)}|^2}{s_p^2(\mathbf{y})} s_p^2(\mathbf{y}) \right] \right)^2 \\ &\leq \mathbb{E} \left[\frac{|b_p^{(i)}|^4}{s_p^4(\mathbf{y})} \right] \mathbb{E}[s_p^4(\mathbf{y})] \end{aligned}$$

from Cauchy-Schwartz inequality. Since $s_p^2(\mathbf{y})$ has a chi-square distribution, all its moments exist, so $\mathbb{E}[s_p^4(\mathbf{y})] < +\infty$. It remains to prove that $\mathbb{E} \left[\frac{|b_p^{(i)}|^4}{s_p^4(\mathbf{y})} \right] < +\infty$. Note first that, by definition, the CML algorithm increases the log likelihood function from one iteration to the next. So, from (20) we get:

$$\prod_{p=1}^P (s_p^2(\mathbf{y}) + |b_p^{(i+1)}|^2) \leq \prod_{p=1}^P (s_p^2(\mathbf{y}) + |b_p^{(1)}|^2).$$

We derive that

$$\prod_{p=1}^P \left(1 + \frac{|b_p^{(i+1)}|^2}{s_p^2(\mathbf{y})} \right) \leq \prod_{p=1}^P \left(1 + \frac{|b_p^{(1)}|^2}{s_p^2(\mathbf{y})} \right).$$

For all $p = 1, \dots, P$,

$$\frac{|b_p^{(i+1)}|^2}{s_p^2(\mathbf{y})} \leq \prod_{p=1}^P \left(1 + \frac{|b_p^{(1)}|^2}{s_p^2(\mathbf{y})} \right). \quad (39)$$

Now taking the square, calculating the expectation and using $\mathbf{b}^{(1)} = \mathbf{z}$ yields:

$$\mathbb{E} \left[\frac{|b_p^{(i+1)}|^4}{s_p^4(\mathbf{y})} \right] \leq \mathbb{E} \left[\prod_{p=1}^P \left(1 + \frac{|z_p|^2}{s_p^2(\mathbf{y})} \right)^2 \right]. \quad (40)$$

At this stage, it remains to be proved that the right side of inequality (40) is finite in order to finish the proof.

Recall that z_p and $s_p^2(\mathbf{y})$ are independent from Cochran's theorem, and that $s_p^2(\mathbf{y})$ and $s_q^2(\mathbf{y})$ are independent if $p \neq q$. Moreover, since \mathbf{z} is Gaussian, all its moments exist. It remains to be shown that $\mathbb{E} \left[\frac{1}{s_p^2(\mathbf{y})} \right]$ and $\mathbb{E} \left[\frac{1}{s_p^4(\mathbf{y})} \right]$ are finite. From (38) we get the probability density function of $s_p^2(\mathbf{y})$, yielding:

$$\mathbb{E} \left[\frac{1}{s_p^2(\mathbf{y})} \right] = \int_0^{+\infty} \frac{1}{s} \cdot \left(\frac{1}{a2^{K-1}\Gamma(K-1)} \left(\frac{s}{a} \right)^{K-2} e^{-\frac{s}{2a}} \right) ds \quad (41)$$

where $a = \frac{\sigma_p^2}{2K}$ is used to simplify notations. This integral is finite as long as $K \geq 3$. In a similar way, we find that $\mathbb{E} \left[\frac{1}{s_p^4(\mathbf{y})} \right]$ is finite as long as $K \geq 4$, which concludes the proof.

APPENDIX C

FISHER INFORMATION MATRIX

To facilitate the calculations, the negative log-likelihood is rewritten using real numbers:

$$\ell(\mathbf{h}, \boldsymbol{\sigma}^2) = K \sum_{p=1}^P \ln(\sigma_p^2) + \sum_{k=1}^K \sum_{p=1}^P \frac{(y_{p,k}^R - H_p^R)^2 + (y_{p,k}^I - H_p^I)^2}{\sigma_p^2} \quad (42)$$

First, we define the $2P \times 2L$ matrix \mathbf{G} :

$$\mathbf{G} = \begin{bmatrix} \mathbf{F}^R & -\mathbf{F}^I \\ \mathbf{F}^I & \mathbf{F}^R \end{bmatrix}, \quad (42)$$

and we write H_p^R and H_p^I as a function of $[h_1^R, \dots, h_L^R, h_1^I, \dots, h_L^I]^T$:

$$\begin{aligned} H_p^R &= \mathbf{g}_p \cdot [h_1^R, \dots, h_L^R, h_1^I, \dots, h_L^I]^T, \\ H_p^I &= \mathbf{g}_{P+p} \cdot [h_1^R, \dots, h_L^R, h_1^I, \dots, h_L^I]^T. \end{aligned}$$

where \mathbf{g}_p means the p th row of \mathbf{G} . Let $g_{p,q}$ be the (p, q) th entry of \mathbf{G} . The derivatives of $\ell(\mathbf{h}, \boldsymbol{\sigma}^2)$ with respect to \mathbf{h}^R , \mathbf{h}^I and $\boldsymbol{\sigma}^2$ are equal to:

$$\begin{aligned} \partial_{h_i^R} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) &= -2K \sum_{p=1}^P \frac{g_{p,i} b_p^R + g_{P+p,i} b_p^I}{\sigma_p^2}, \\ \partial_{h_i^I} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) &= -2K \sum_{p=1}^P \frac{g_{p,L+i} b_p^R + g_{P+p,L+i} b_p^I}{\sigma_p^2}, \\ \partial_{\sigma_p^2} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) &= \frac{K}{\sigma_p^2} - K \frac{s_p^2(\mathbf{y}) + |b_p^R|^2 + |b_p^I|^2}{\sigma_p^4}, \end{aligned}$$

which yields:

$$\partial_{h_{i_1}^R h_{i_2}^R} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) = 2K \sum_{p=1}^P \frac{g_{p,i_1} g_{p,i_2} + g_{P+p,i_1} g_{P+p,i_2}}{\sigma_p^2} \quad (43)$$

$$\partial_{h_{i_1}^R h_{i_2}^I} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) = 2K \sum_{p=1}^P \frac{g_{p,i_1} g_{p,L+i_2} + g_{P+p,i_1} g_{P+p,L+i_2}}{\sigma_p^2} \quad (44)$$

$$\partial_{h_{i_1}^I h_{i_2}^I} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) = 2K \sum_{p=1}^P \frac{g_{p,L+i_1} g_{p,L+i_2} + g_{P+p,L+i_1} g_{P+p,L+i_2}}{\sigma_p^2} \quad (45)$$

$$\partial_{h_i^R \sigma_p^2} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) = 2K \frac{g_{p,i} b_p^R + g_{P+p,i} b_p^I}{\sigma_p^4} \quad (46)$$

$$\partial_{\sigma_{p_1}^2 \sigma_{p_2}^2} \ell(\mathbf{h}, \boldsymbol{\sigma}^2) = \left(-\frac{K}{\sigma_{p_1}^4} + 2K \frac{s_{p_1}^2(\mathbf{y}) + |b_{p_1}^R|^2 + |b_{p_1}^I|^2}{\sigma_{p_1}^6} \right) \delta_{p_1}^{p_2}. \quad (47)$$

Then, the expectation needs to be calculated to find \mathbf{J} . The expectation of (43), (44) and (45) is unchanged. From the definition of $b_p = \bar{y}_p - H_p$ and (37), the expectation of (46) is found to be 0, and from (38) the expectation of (47) is found to be $\frac{K}{\sigma_{p_1}^4} \delta_{p_1}^{p_2}$.

REFERENCES

- [1] J.-W. van Bloem, R. Schiphorst, T. Kluwer, and C. Slump, "Interference Measurements in IEEE 802.11 Communication Links Due to Different Types of Interference Sources," in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on*, Sept 2012, pp. 1–6.
- [2] A. Coulson, "Narrowband interference in pilot symbol assisted OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2277–2287, Nov 2004.
- [3] —, "Bit error rate performance of OFDM in narrowband interference with excision filtering," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2484–2492, September 2006.
- [4] S. Lee, K. Kwak, J. Kim, and D. Hong, "Channel estimation approach with variable pilot density to mitigate interference over time-selective cellular OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2694–2704, July 2008.
- [5] Y. Zhang, X. Zhang, and D. Yang, "A robust least square channel estimation algorithm for ofdm systems under interferences," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, April 2013, pp. 3122–3127.
- [6] T. Li, W.-H. Mow, V. Lau, M. Siu, R. Cheng, and R. Murch, "Robust joint interference detection and decoding for OFDM-based cognitive radio systems with unknown interference," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 566–575, 2007.
- [7] M. Morelli and M. Moretti, "Robust frequency synchronization for OFDM-based cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5346–5355, December 2008.
- [8] —, "Channel estimation in OFDM systems with unknown interference," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5338–5347, 2009.
- [9] T. W. Anderson, "Asymptotically efficient estimation of covariance matrices with linear structure," *The Annals of Statistics*, pp. 135–141, 1973.
- [10] T. H. Szatrowski, "Necessary and sufficient condition for explicit solutions in the multivariate normal estimation problem for patterned means and covariances," *The Annals of Statistics*, vol. 8, pp. 802–810, 1980.
- [11] T.-H. Pham and Y.-C. Liang, "Joint Channel Information Estimation and Data Detection for OFDM-Based Systems under Unknown Interference," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, May 2012, pp. 1–5.
- [12] H. Hartley and K. Jayatilake, "Estimation for linear models with unequal variances," *Journal of the American Statistical Association*, vol. 68, no. 341, pp. 189–192, 1973.
- [13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.