

RESEARCH

Open Access



Analysis of the maximum likelihood channel estimator for OFDM systems in the presence of unknown interference

Azzouz Dermoune^{1*} and Eric Pierre Simon²

Abstract

This paper is a theoretical analysis of the maximum likelihood (ML) channel estimator for orthogonal frequency-division multiplexing (OFDM) systems in the presence of unknown interference. The following theoretical results are presented. Firstly, the uniqueness of the ML solution for practical applications, i.e., when thermal noise is present, is analytically demonstrated when the number of transmitted OFDM symbols is strictly greater than one. The ML solution is then derived from the iterative conditional ML (CML) algorithm. Secondly, 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. Thirdly, it is theoretically demonstrated that the channel estimator is not biased. The second and the third results are obtained by employing oblique projection theory. Furthermore, these results are confirmed by numerical results.

1 Introduction

Narrow band interference (NBI) arises in orthogonal frequency-division multiplexing (OFDM) systems in a number of transmission scenarios, such as Wi-Fi communications [1, 2] or cognitive radio, where different types of wireless services can use the same frequency band. The NBI can affect several subcarriers. It is well known that it greatly degrades the performance of the receiver if it is not treated [3, 4]. 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 NBI, neither the number of affected subcarriers nor their location in the spectrum is known. This brings about the need to estimate 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 methods have been proposed to estimate the channel in the presence of interference. In [5], channel estimation is investigated for OFDM systems in the presence of synchronous interference. However, in practical situations, the interferer's signals are in general asynchronous. Zhou et al.'s work [6] deals with the estimation of noise plus interference power together with successive soft data estimation at each subcarrier. Here, we focus on channel estimation based on pilots.

In [7], the authors proposed an estimator employing a specific pilot structure consisting of two types of pilot symbols with different pilot densities. More recently, in [8], a channel estimator is proposed based on a robust least-squares approach. However, the proposed method requires that the number of pilots be greater than twice the channel order, defined as the number of taps.

In [9] and [10], the NBI is assumed to be Gaussian distributed in the frequency domain, with zero mean and unknown power.

Under the same assumptions, channel estimation with NBI is investigated in the seminal paper [11]. The authors also consider the case where any possible correlation between the interference over adjacent subcarriers is neglected. This case can be considered as the worst

*Correspondence: Azzouz.Dermoune@math.univ-lille1.fr

¹Laboratoire Paul Painlevé, USTL-UMR-CNRS 8524. UFR de Mathématiques, Bât. M2, 59655 Villeneuve d'Ascq Cédex, France

Full list of author information is available at the end of the article

case, since correlation is additional information that could be used to improve the estimation. The authors perform pilot-aided channel estimation under the classical assumption that the number of pilots is greater than the channel order. Note that this assumption is also supposed to be verified in this paper. After formulating the maximum likelihood (ML) algorithm for the joint estimation of $\{\mathbf{h}, \sigma^2\}$, it is shown in [11] that the solution is non-unique when the channel order (denoted by L) is greater than the number of transmitted OFDM symbols (denoted by K), leading to ambiguous channel estimates. This is a severe limitation since $K \leq L$ is a common scenario in practice. For example in Wi-Fi scenarios, short frames like control frames are sent frequently. For this reason, the authors 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 off-line through exhaustive grid-search simulations. This can be a limitation for practical use.

In this paper, we first demonstrate that the ambiguities actually appear only when $K = 1$ unless the signal-to-noise ratio (SNR) value is extremely high, in which case ambiguities indeed appear if $K \leq L$. Thus, for typical SNR values corresponding to practical applications and $K > 1$, the joint ML technique can be used instead of the EM technique. This makes it possible to avoid a grid search. 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 technique.

For the case $K > 1$, the likelihood equations can be solved with the conditional ML (CML) algorithm. Note that the CML algorithm has been investigated for channel estimation in code division multiple access (CDMA) systems [12]. In this paper, we first present the CML equations for the considered OFDM system. Numerical simulations indicate that the CML is well defined in the SNR range corresponding to practical applications.

Then, we use a new formulation of the CML based on oblique projections to investigate the first moment. Oblique projections are well known for their applications in signal processing, especially in channel estimation [13–16]. With this formulation, it is proved that the channel estimator is unbiased. This result is 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.

This paper is organized as follows. Section 2 describes the system model. In Section 3, we discuss the joint ML estimation of $\{\mathbf{h}, \sigma^2\}$ and the question of the uniqueness of the solution. Then, Section 4 introduces the CML algorithm to find the solution. A theoretical study of the CML is provided in Section 5. The Cramer-Rao bound is derived in Section 6, and simulation results are presented in Section 7.

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 the matrix \mathbf{A} ; a_i indicates the i th entry of the vector \mathbf{a} . The 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 the operations of taking the transpose, Hermitian conjugate, complex conjugate, real part, and imaginary part. The mathematical expectation is denoted $\mathbb{E}[\cdot]$. The multivariate complex normal distribution of a P -dimensional random vector is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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 by χ_k^2 . The notations $\text{Range}(\mathbf{A})$ and $\text{Null}(\mathbf{A})$ indicate, respectively, the range space and the null space of \mathbf{A} .

Note on the notations in bold \mathbf{a}^2 and $|\mathbf{a}|^2$: let $\mathbf{a} = [a_1, \dots, a_N]^T$ be a vector of size $N \times 1$. We use the notation in bold \mathbf{a}^2 to denote the vector that is formed by taking the square of the entries of \mathbf{a} , i.e., $\mathbf{a}^2 = [a_1^2, \dots, a_N^2]^T$. Similarly, $|\mathbf{a}|^2 = [|a_1|^2, \dots, |a_N|^2]^T$.

2 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 subcarriers at the positions $\mathcal{P} = \{n_p, p = 1, \dots, P\}$. In our paper, 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, i.e., $|c_{p,k}|^2 = 1$, transmitted on the n_p th subcarrier, and $w_{p,k}$ is a 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 $\sigma_p^2 = \sigma_{\text{AWGN}}^2 + \sigma_{\text{NBI},p}^2$, where σ_{AWGN}^2 is the additive white Gaussian noise (AWGN) contribution and $\sigma_{\text{NBI},p}^2$ is the average NBI power, which is assumed constant over the transmission period. The channel frequency response H_p at the n_p th subcarrier is given by

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

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 $\exp\left(-j \frac{2\pi n_p(l-1)}{N}\right)$, $p = 1, \dots, P, l = 1, \dots, L$.

The ML estimation of the set of unknown parameters $\{\mathbf{h}, \boldsymbol{\sigma}^2\}$ is desired, where $\boldsymbol{\sigma}^2 = [\sigma_1^2, \dots, \sigma_P^2]^T$, 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 by $\bar{\mathbf{y}} = [\bar{y}_1, \dots, \bar{y}_P]^T$, 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 by $\mathbf{s}^2 = [s_1^2, \dots, s_P^2]^T$, where for $p = 1, \dots, P$,

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

3 Maximum likelihood estimation

In this section, the ML estimate of $\{\mathbf{h}, \boldsymbol{\sigma}^2\}$ will be investigated by following the approach presented in [11]. However, it will be shown that the ambiguities mentioned in [11] appear only when $K = 1$, unless the SNR value is extremely high. Hence, when $K \geq 2$ and practical SNR values are considered, 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 $\ell(\tilde{\mathbf{h}}, \tilde{\boldsymbol{\sigma}}^2)$ is given by [11]:

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

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 $\tilde{\mathbf{h}}$ are first calculated:

$$\hat{\sigma}_p^2(\tilde{\mathbf{h}}) = \frac{1}{K} \sum_{k=1}^K |y_{p,k} - \mathbf{F}_p \tilde{\mathbf{h}}|^2, \quad p = 1, \dots, P. \quad (8)$$

Then, substituting $\hat{\sigma}_p^2(\tilde{\mathbf{h}})$ for $\tilde{\sigma}_p^2$ in (7) yields

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

Finally, the ML estimate of the CIR vector \mathbf{h} is the one that minimizes $\Lambda(\tilde{\mathbf{h}})$. Special treatment is required due to the presence of the logarithm function in (9). Indeed, the values of $\tilde{\mathbf{h}}$ for which $\hat{\sigma}_p^2(\tilde{\mathbf{h}}) = 0$ make $\Lambda(\tilde{\mathbf{h}})$ tend to $-\infty$. The consequences for the uniqueness of the solution are explained in more detail in [11], where it is shown that the minimization leads to ambiguous channel estimates if $K \leq L$. At this point, it is important to precise the context leading to this result. It was assumed that no prior knowledge about the noise variances was available, which implied that theoretically they could be zero. However, in practice, the noise variance is never zero due to the presence of thermal noise. We will now revisit this ambiguity issue by taking into account the fact that the noise variance is not zero. Let us first observe that the equation $\hat{\sigma}_p^2(\tilde{\mathbf{h}}) = 0$ (8) for a given p yields a linear system of K equations with L unknowns:

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

where the $K \times L$ matrix $\tilde{\mathbf{A}}$ 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 $\tilde{\mathbf{A}}$ has to be taken into account when solving the system. On the one hand, the rows of $\tilde{\mathbf{A}}$ are all identical. On the other hand, in the presence of AWGN, the samples $y_{p,1}, \dots, y_{p,K}$ are all different since they are Gaussian and independent. Consequently, it is found that (10) has no solution. Then, it becomes obvious that for $K > 1$, $\hat{\sigma}_p^2(\tilde{\mathbf{h}}) > 0$ for all p and for all $\tilde{\mathbf{h}}$. Therefore, the ML estimate of the CIR vector \mathbf{h} is well defined for $K > 1$. This is true in the presence of AWGN, but at extremely high values of SNR, the difference between the samples tends to disappear and the problem of ambiguities still remains. However, simulation results show that the system is well

conditioned if the SNR is less than 150 dB, which is the case of practical applications.

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

$$\hat{\sigma}_p^2(\tilde{\mathbf{h}}) = s_p^2 + |\bar{y}_p - \mathbf{F}_p \tilde{\mathbf{h}}|^2, \quad (11)$$

when noticing that if $K = 1$, the sample variances $s_p^2 = 0$ for all p . However, if $K > 1$, $s_p^2 \neq 0$ for all p , making it impossible to set $\hat{\sigma}_p^2(\tilde{\mathbf{h}})$ 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 for the ML of σ^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 [17]

$$0 < \delta_p^2 \leq \hat{\sigma}_p^2 \quad (12)$$

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

$$\hat{\sigma}_p^2(\tilde{\mathbf{h}}, \delta_p^2) = |y_p - \mathbf{F}_p \tilde{\mathbf{h}}|^2, \quad \text{if } |y_p - \mathbf{F}_p \tilde{\mathbf{h}}|^2 \geq \delta_p^2, \quad (13)$$

$$\hat{\sigma}_p^2(\tilde{\mathbf{h}}, \delta_p^2) = \delta_p^2, \quad \text{if } |y_p - \mathbf{F}_p \tilde{\mathbf{h}}|^2 < \delta_p^2. \quad (14)$$

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

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

and then the CIR estimate $\hat{\mathbf{h}}$ is the one that minimizes $\Lambda(\tilde{\mathbf{h}}, \delta^2)$ with respect to $\tilde{\mathbf{h}}$.

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

4 Conditional ML

The CML algorithm is an iterative algorithm for solving the ML problem. The CML is the result of two nested minimizations. First, (7) is minimized given the channel $\tilde{\mathbf{h}}$, yielding the estimation of σ^2 given by (8) or (11):

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

where $|\bar{\mathbf{y}} - \mathbf{F}\tilde{\mathbf{h}}|^2 = [|\bar{y}_1 - \mathbf{F}_1 \tilde{\mathbf{h}}|^2, \dots, |\bar{y}_P - \mathbf{F}_P \tilde{\mathbf{h}}|^2]^T$ (see the note on the notation in bold $|\cdot|^2$ in the *Notations*

section). Conversely, (7) is minimized given $\tilde{\sigma}^2$, yielding the following estimate for \mathbf{h} :

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

We obtain the following iterative algorithm, with I the number of iterations:

-
- 1: **Initialization:** $\hat{\sigma}_p^{2(0)} = 1, p = 1, \dots, P$
 - 2: **for** $i = 0$ to $I - 1$ **do**
 - 3: $\hat{\mathbf{h}}^{(i+1)} = (\mathbf{F}^H \mathbf{D}^{-1}(\hat{\sigma}^{2(i)}) \mathbf{F})^{-1} \mathbf{F}^H \mathbf{D}^{-1}(\hat{\sigma}^{2(i)}) \bar{\mathbf{y}}$
 - 4: $\hat{\sigma}^{2(i+1)} = \mathbf{s}^2 + |\bar{\mathbf{y}} - \mathbf{F}\hat{\mathbf{h}}^{(i+1)}|^2$
 - 5: **end for**
-

We recall that we have used the notations in bold defined in the notation section. Note that with this initialization, $\hat{\mathbf{h}}^{(1)}$ is the ordinary least-squares estimate of \mathbf{h} . Note also that the expectation-maximization algorithm of [11, equations (28),(29)] corresponds to the CML algorithm for $\lambda = 0$.

This algorithm is known as the scoring method [18, 19] or the conditional maximum likelihood (CML) algorithm [20].

The CML's important properties include:

1. Given $\hat{\sigma}^{2(i)}$, the vector $\hat{\mathbf{h}}^{(i+1)}$ maximizes the likelihood. Given $\hat{\mathbf{h}}^{(i)}$, the vector $\hat{\sigma}^{2(i+1)}$ maximizes the likelihood.
2. The logarithmic means $\frac{1}{P} \sum_{p=1}^P \ln(\hat{\sigma}_p^{2(i)})$, for $i = 0, 1, \dots$, are non-increasing. In other words, for all i ,

$$\sum_{p=1}^P \ln(\hat{\sigma}_p^{2(i+1)}) \leq \sum_{p=1}^P \ln(\hat{\sigma}_p^{2(i)}). \quad (18)$$

In addition, $\sum_{p=1}^P \ln(\hat{\sigma}_p^{2(i)})$ converges to some constant $\ln(c^*) \geq \sum_{p=1}^P \ln(s_p^2)$.

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.

5 Theoretical analysis of the CML algorithm

In this section, some properties of the CML algorithm are studied. In particular, the first moment of the estimators is investigated. To do so, a formulation of the CML based on oblique projections is established. It will be shown that the CML algorithm can be viewed as successive oblique projections.

5.1 The CML algorithm and oblique projections

First, we briefly present 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 into two subspaces: the range space $\text{Range}(\Pi(\Sigma)) = \Pi(\Sigma) (\mathcal{C}^P)$ of $\Pi(\Sigma)$ and its null space $\text{Null}(\Pi(\Sigma)) = [\mathbf{I} - \Pi(\Sigma)] (\mathcal{C}^P)$. 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 onto $\text{Range}(\mathbf{F})$. If $\Sigma = \mathbf{I}$, then $\Pi(\mathbf{I}) = \mathbf{F} (\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H$ is the orthogonal projection on $\text{Range}(\mathbf{F})$. For simplicity of notation, $\Pi(\mathbf{I})$ is now denoted by Π .

Now, we go back to the CML algorithm defined in the previous Section.

At the first iteration, the orthogonal projection Π splits the sample mean into the following two components (see Fig. 1 for a geometrical interpretation):

$$\begin{aligned} \hat{\mathbf{F}}\mathbf{h}^{(1)} &= \Pi \bar{\mathbf{y}}, \\ \mathbf{b}^{(1)} &:= (\mathbf{I} - \Pi) \bar{\mathbf{y}}. \end{aligned} \quad (19)$$

The vector $\mathbf{b}^{(1)}$ is the orthogonal projection of $\bar{\mathbf{y}}$ onto $\text{Null}(\Pi)$.

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

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

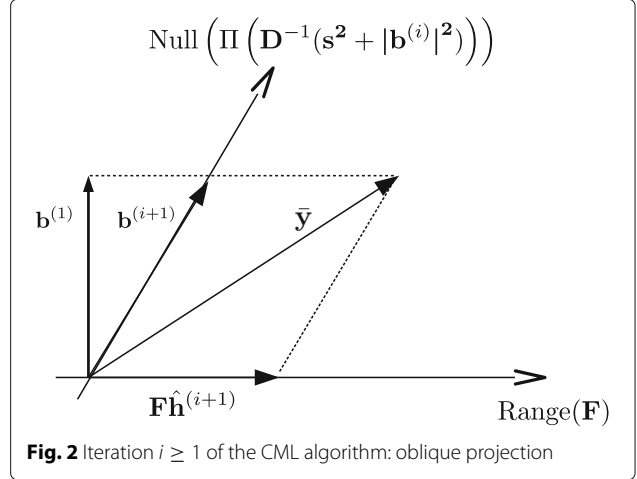
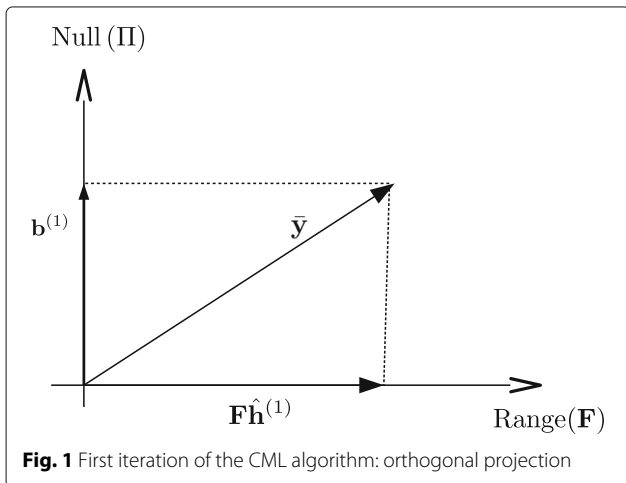
where the column vector $|\mathbf{b}^{(1)}|^2 = [|b_1^{(1)}|^2, \dots, |b_p^{(1)}|^2]^T$.

At the $i + 1$ th iteration, we have

$$\begin{aligned} \hat{\mathbf{h}}^{(i+1)} &= \left(\mathbf{F}^H \mathbf{D}^{-1} (\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2) \mathbf{F} \right)^{-1} \\ &\quad \mathbf{F}^H \mathbf{D}^{-1} (\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2) \bar{\mathbf{y}}. \end{aligned}$$

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

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



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

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

From this discussion, it can be concluded that the solution provided by the CML algorithm can be sought either in $\hat{\mathbf{F}}\mathbf{h}^{(i)}$ or in the variable $\mathbf{b}^{(i)} = \bar{\mathbf{y}} - \hat{\mathbf{F}}\mathbf{h}^{(i)}$. In the perspective of investigating the CML properties, it will be seen that it is more convenient to consider $\mathbf{b}^{(i)}$, which yields an equivalent algorithm for solving the CML:

$$\mathbf{b}^{(i+1)} = (\mathbf{I} - \Pi (\mathbf{D}^{-1} (\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2))) \bar{\mathbf{y}}. \quad (20)$$

Now, let us study the properties of Eq. (20). To do so, let us first write $\mathbf{b}^{(i+1)}$ as a function of the three following variables $\mathbf{b}^{(i+1)} = \varphi (\bar{\mathbf{y}}, \mathbf{s}^2, \mathbf{b}^{(i)})$ where

$$\varphi (\bar{\mathbf{y}}, \mathbf{s}^2, \mathbf{b}^{(i)}) := (\mathbf{I} - \Pi (\mathbf{D}^{-1} (\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2))) \bar{\mathbf{y}}, \quad (21)$$

is a function of size $P \times 1$. From now on, φ_p denotes the p th entry of φ . Now, the following properties are stated in Proposition 1. They will be used to derive Proposition 2.

Proposition 1 1) For $i \geq 1$,

$$\varphi (\bar{\mathbf{y}}, \mathbf{s}^2, \mathbf{b}^{(i)}) = \varphi (\mathbf{b}^{(1)}, \mathbf{s}^2, \mathbf{b}^{(i)}), \quad (22)$$

and

$$\begin{aligned} \varphi (\mathbf{b}^{(1)}, \mathbf{s}^2, \mathbf{b}^{(i)}) &= (\mathbf{I} - \Pi \mathbf{D} (\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2)) \\ &\quad \times \Pi \mathbf{D}^{-1} (\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2) \mathbf{b}^{(1)}. \end{aligned} \quad (23)$$

2) The maps

$$\varphi_p (\mathbf{b}^{(1)}, \mathbf{s}^2, \mathbf{b}) = (\mathbf{I} - \Pi \mathbf{D} (\mathbf{s}^2 + |\mathbf{b}|^2) \Pi \mathbf{D}^{-1} (\mathbf{s}^2 + |\mathbf{b}|^2))_p \mathbf{b}^{(1)} \quad (24)$$

for $p = 1, \dots, P$ are linear with respect to $\mathbf{b}^{(1)}$ and rational with respect to the variables $s_1^2, \dots, s_p^2, |b_1|^2, \dots, |b_p|^2$.

Recall that the subscript p applied to a matrix means taking the p th row of the matrix. More precisely,

$$\varphi_p(\mathbf{b}^{(1)}, \mathbf{s}^2, \mathbf{b}) = b_p^{(1)} - \sum_{k_1=1}^P \sum_{k_2=1}^P \frac{s_{k_1}^2 + |b_{k_1}|^2}{s_{k_2}^2 + |b_{k_2}|^2} \times \left[\prod \mathbf{D}^{(k_1)} \prod \mathbf{D}^{(k_2)} \right]_p \mathbf{b}^{(1)}, \quad (25)$$

where $\mathbf{D}^{(k)} = \mathbf{D}(0, \dots, 0, 1, 0, \dots, 0)$ and the k th entry is 1.

3) We have for $p = 1, \dots, P$,

$$\begin{aligned} b_p^{(i+1)} &= \varphi_p(\mathbf{b}^{(1)}, \mathbf{s}^2, \mathbf{b}^{(i)}) \\ &= \varphi_p(\mathbf{b}^{(1)}, \mathbf{s}^2, \varphi(\mathbf{b}^{(1)}, \mathbf{s}^2, \mathbf{b}^{(i-1)})) \\ &\vdots \\ &= \varphi_p^{(i)}(\mathbf{b}^{(1)}, \mathbf{s}^2) \end{aligned}$$

where the map $\varphi_p^{(i)}(\mathbf{b}^{(1)}, \mathbf{s}^2)$ is odd with respect to the variable $b_k^{(1)}$ for all $k = 1, \dots, P$.

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

$$\mathbf{b} - \varphi(\mathbf{b}^{(1)}, \mathbf{s}^2, \mathbf{b}).$$

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

The proofs are given in Appendix 1. Note that (22) is straightforward with the geometrical interpretation of Fig. 2, where it can be observed that the projections of $\mathbf{b}^{(1)}$ and $\bar{\mathbf{y}}$ onto $\text{Null}(\Pi(\mathbf{D}^{-1}(\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2)))$ are the same.

5.2 The mean of the CML estimator

The distributions of the sample mean and the sample variance are derived in Appendix 2. Upon the convergence of the algorithm (20), we obtain $\hat{\mathbf{h}}$, $\hat{\mathbf{b}}$, and $\hat{\sigma}^2$. From $\mathbf{F}\hat{\mathbf{h}} = \bar{\mathbf{y}} - \hat{\mathbf{b}}$ and (38) in Appendix 2, we obtain

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

From $\hat{\sigma}^2 = \mathbf{s}^2 + |\hat{\mathbf{b}}|^2$ and (39) in Appendix 2, we obtain

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

using the property that the mean of a chi-square random variable of n degrees of freedom is n . Now, from Proposition 1, the following can be shown (see Appendix 3 for the proof).

Proposition 2 The vector $\hat{\mathbf{b}}$ is zero mean, i.e.,

$$\mathbb{E}[\hat{b}_p] = 0, \quad p = 1, \dots, P. \quad (28)$$

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

5.3 Complexity of the CML algorithm

In this Section, the complexity of the CML algorithm (20) is investigated. First, it is noteworthy that it is more interesting in terms of complexity to use the formulation (23) instead of (20). Indeed, in (20), the matrix inversion $(\mathbf{F}^H (\mathbf{D}^{-1}(\mathbf{s}^2 + |\mathbf{b}^{(i)}|^2) \mathbf{F}))^{-1}$ is required at each iteration, whereas in (23), the computation of matrix Π can be done off-line and stored in a memory. In this way, the operations for computing (23) only consist in matrix products, which is less demanding. Moreover, the calculation of \mathbf{s}^2 is carried out just once at the beginning of the algorithm. Therefore, the algorithm requires $O(P^3)$ floating point operations in total for each iteration i .

6 The Cramer-Rao bound

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

$$\boldsymbol{\theta} = [h_1^R, \dots, h_L^R, h_1^I, \dots, h_L^I, \sigma_1^2, \dots, \sigma_P^2]^T, \quad (29)$$

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

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

$$\text{cov}(\hat{\boldsymbol{\theta}}) \geq \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{J}^{-1} \frac{\partial \psi(\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}}, \quad (30)$$

where

$$\psi(\boldsymbol{\theta}) = \mathbb{E}[\hat{\boldsymbol{\theta}}] \quad (31)$$

and $\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is the Jacobian matrix, whose (n, m) th element is given by $\frac{\partial \psi_n(\boldsymbol{\theta})}{\partial \theta_m}$. Note that in [11], the CRB is calculated by assuming perfect knowledge of the variance. In this Section, both channel and variance parameters are being considered. The 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 can be seen from (31) that knowing the moments of the estimator is required in order to calculate the CRB. The results of Section 5.2 will be exploited to do so.

The results of the calculation of \mathbf{J} are given below, and the details are in Appendix 4. The matrix \mathbf{J} can be written as

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

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 (41), (42), and (43) in Appendix 4. 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 of the blocks, as long as the submatrices are invertible, we have

$$\mathbf{J}^{-1} = \begin{bmatrix} \mathbf{J}_h^{-1} & \mathbf{0}_{2L \times P} \\ \mathbf{0}_{P \times 2L} & \mathbf{J}_{\sigma^2}^{-1} \end{bmatrix}. \quad (33)$$

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

$$\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathbf{I}_{2L} & \mathbf{0}_{2L \times P} \\ \mathbf{0}_{P \times 2L} & \mathbf{M} \end{bmatrix} \quad (34)$$

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

$$\frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{J}^{-1} \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T = \begin{bmatrix} \mathbf{J}_h^{-1} & \mathbf{0}_{2L \times P} \\ \mathbf{0}_{P \times 2L} & \mathbf{M} \mathbf{J}_{\sigma^2}^{-1} \mathbf{M}^T \end{bmatrix}. \quad (35)$$

Note that the calculation of $\mathbb{E} \left[\frac{\partial |\hat{b}_{p_1}|^2}{\partial \sigma_{p_2}^2} \right]$ is not feasible since there is no analytical expression for $\hat{\mathbf{b}}$. Therefore, the bound for the variance estimation cannot be found. However, the bound for the channel estimation is given from (29) and (35) by

$$\text{cov} \left(\left[\hat{h}_1^R, \dots, \hat{h}_L^R, \hat{h}_1^I, \dots, \hat{h}_L^I \right]^T \right) \geq \mathbf{J}_h^{-1}. \quad (36)$$

Recall that this bound has been derived by using the result of Proposition 2 showing that $\hat{\mathbf{h}}$ was unbiased.

7 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. The system parameters used for the simulation are as follows: $N = 64$ subcarriers, 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 eight subcarriers, yielding $P = 8$. A frame of $K = 4$ OFDM symbols is assumed. Note that $K < L$ with these considered values.

It is also assumed that two contiguous pilot subcarriers are affected by NBI, by adding a Gaussian disturbance of variance σ_{NBI}^2 to both subcarriers. The signal-to-noise ratio (SNR) is defined as $10 \log \frac{1}{\sigma_{\text{AWGN}}^2}$ where the power of the signal is normalized to one, and the signal-to-interference ratio (SIR) is defined as $10 \log \frac{1}{\sigma_{\text{NBI}}^2}$. The accuracy of the channel estimates is measured in terms of the mean square error (MSE), which is defined as $\frac{1}{L} \mathbb{E} \left[(\hat{\mathbf{h}} - \mathbf{h})^H (\hat{\mathbf{h}} - \mathbf{h}) \right]$, where the expectation is estimated via Monte Carlo simulations.

Figure 3 shows the MSE as a function of the SNR when the SIR is fixed to -5 dB. Four iterations are employed. Let us recall that the first iteration is the ordinary least-squares estimate (OLSE). For reference, the CRB calculated in Section 6 is added. It is seen that the algorithm converges after four iterations for the SNR ranging from 0

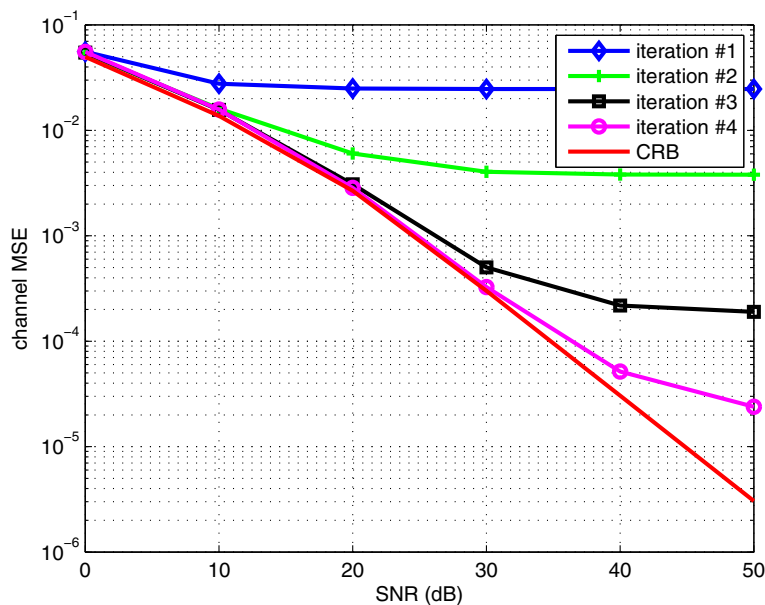


Fig. 3 Performance of the CML estimator as a function of the SNR for SIR = -5 dB, $K = 4$

to 30 dB and nearly attains the CRB, whereas the MSE for the OLSE has a floor at 2.5×10^{-2} . This shows that the ML algorithm is well conditioned for the considered SNR range [0, 50] dB.

Next, 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 can be observed that the performance of the OLSE approaches that of the second iteration when the SIR increases. This makes sense, as the OLSE is designed to perform well without NBI.

To investigate the impact of the number of OFDM symbols on the performance, K is now fixed to 8 and the other parameters are the same. The results are shown in Fig. 6 for an SIR = -5 dB. Compared to $K = 4$, it can be observed that the convergence is faster, since only three iterations are needed to converge. This is understood since it is a more favorable scenario (Fig. 6).

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 Figs. 7 and 8, 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 Proposition 2).

8 Conclusions

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 is strictly greater than one. For this case, we proposed using the conditional

maximum likelihood (CML) algorithm to obtain the estimates. New theoretical developments of the CML algorithm in this context have been presented. It was proved that the solution provided by the CML is an algebraic function of the data. Furthermore, it was also proved that the channel estimator is unbiased.

Appendix 1: Proof of Proposition 1

The proof of Proposition 1 1) is a consequence of the following general results. We have for any invertible matrix Σ that

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

The latter equality is equivalent to

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

Now, it can be easily shown that

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

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 this, we have

$$\Pi(\Sigma^{-1}) = \mathbf{F} \mathbf{F}^+ \Sigma (\mathbf{F}^H)^+ \mathbf{F}^H \Sigma^{-1} = \Pi \Sigma \Pi \Sigma^{-1}. \quad (37)$$

The proof of the remaining assertions is straightforward.

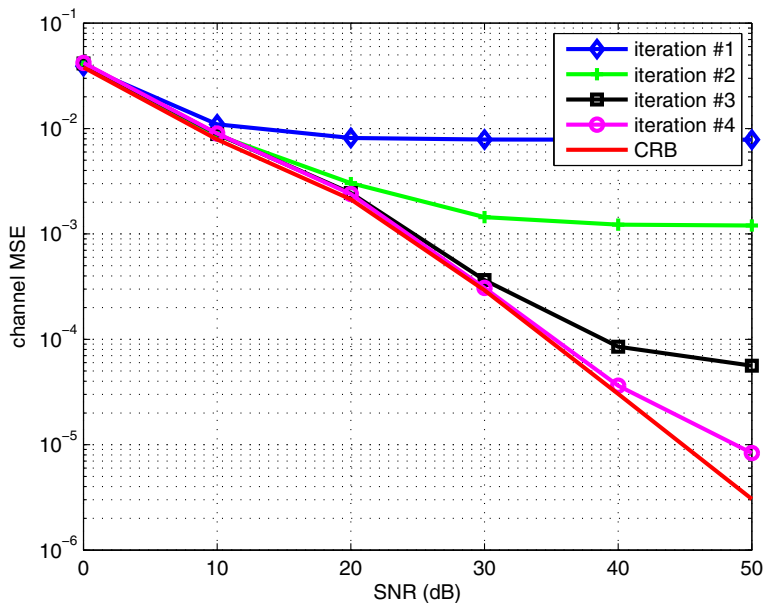


Fig. 4 Performance of the CML estimator as a function of the SNR for SIR = 0 dB, K = 4

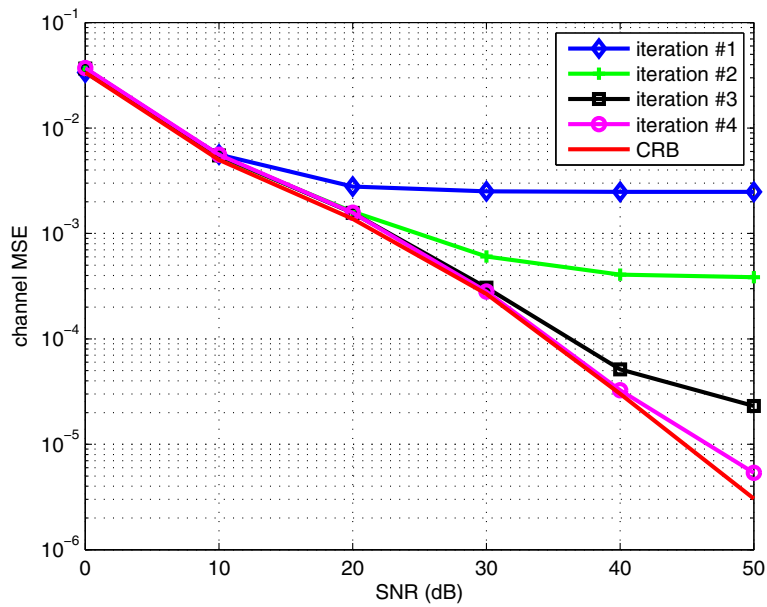


Fig. 5 Performance of the CML estimator as a function of the SNR for SIR = 5 dB, K = 4

Appendix 2: Probability distribution function of the sample mean and the sample variance

Cochran’s theorem [21] 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 down the distribution for s_p^2 . The distribution for \bar{y}_p is straightforward:

$$\bar{y}_p \sim \mathcal{CN}\left(\mathbf{F}_p \mathbf{h}, \frac{\sigma_p^2}{K}\right), \tag{38}$$

$$s_p^2 \sim \frac{\sigma_p^2}{2K} \chi_{2(K-1)}^2. \tag{39}$$

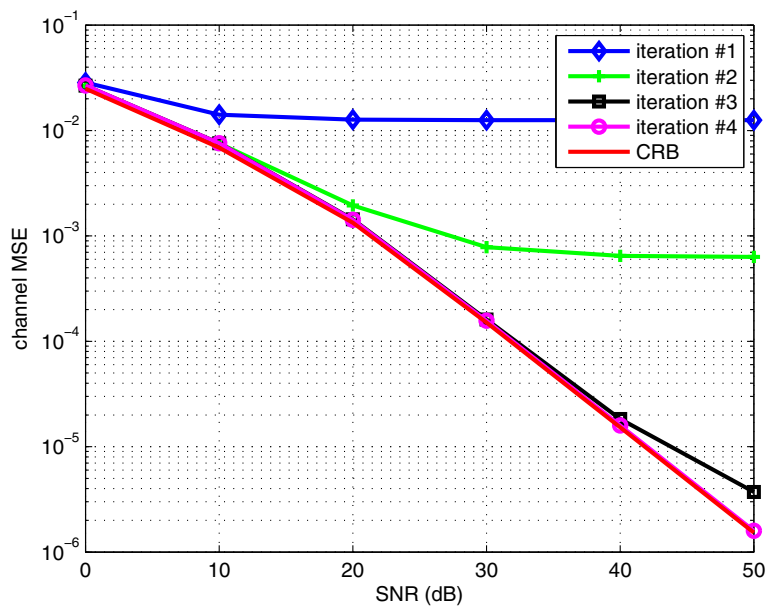
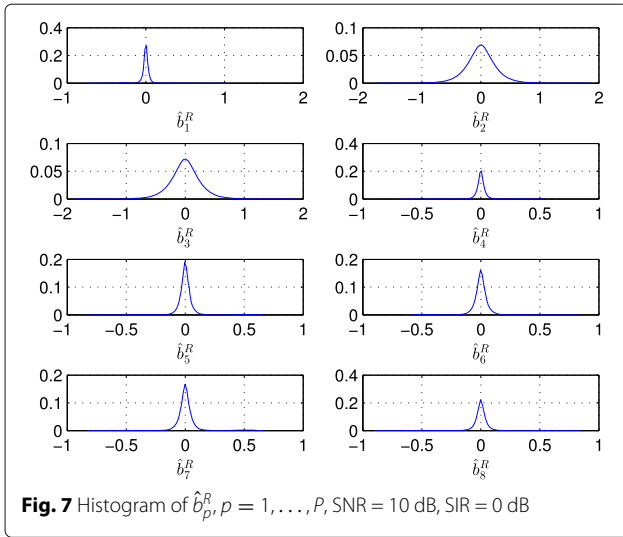


Fig. 6 Performance of the CML estimator as a function of the SNR for SIR = -5 dB, K = 8



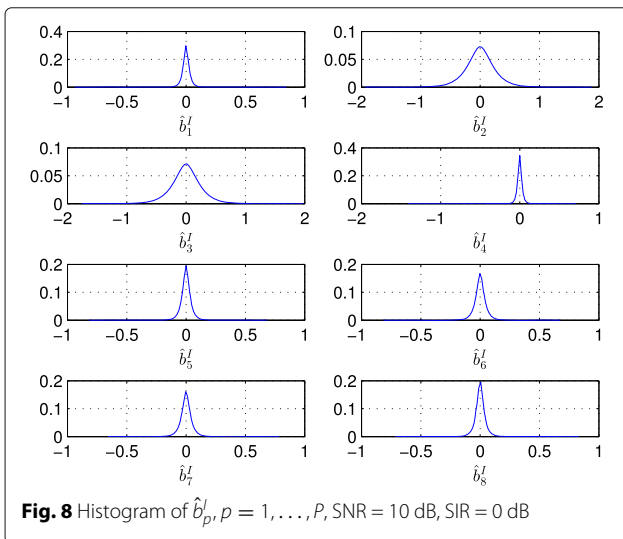
Here we find $2(K - 1)$ degrees of freedom since the considered random variables are complex.

Appendix 3: Proof of Proposition 2

The Gaussian vector $\mathbf{b}^{(1)}$, defined in (19), is zero mean with the covariance matrix $K^{-1}(\mathbf{I} - \Pi)\mathbf{D}(\sigma^2)$, i.e., $(\mathbf{I} - \Pi)\bar{\mathbf{y}} \sim (\mathbf{I} - \Pi)K^{-1/2}\mathbf{D}(\sigma)\mathcal{N}(0, \mathbf{I})$.

The components of the vector $2K\mathbf{s}^2$ are independent with the distribution $(\sigma_1^2\chi_{2(K-1)}^2(1), \dots, \sigma_P^2\chi_{2(K-1)}^2(P))$. 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 and $\bar{\mathbf{y}}$ are independent. From Proposition 1, the random vector $\mathbf{b}^{(i+1)} = \varphi^{(i)}(\mathbf{b}^{(1)}, \mathbf{s}^2)$ is a rational function having a positive denominator. Therefore,

$$\begin{aligned} \mathbb{E}(\mathbf{b}^{(i+1)}|\mathbf{s}^2) &= \mathbb{E}(\varphi^{(i)}(\mathbf{b}^{(1)}, \mathbf{s}^2) | \mathbf{s}^2) \\ &= \int \varphi^{(i)}(\mathbf{b}, \mathbf{s}^2) f_{\mathbf{b}^{(1)}}(\mathbf{b}) d\mathbf{b} = 0 \end{aligned}$$



because $\mathbf{b}^{(1)}$ is zero mean which implies that its probability density function $\mathbf{b} \rightarrow f_{\mathbf{b}^{(1)}}(\mathbf{b})$ is even and $\mathbf{b} \rightarrow \varphi^{(i)}(\mathbf{b}, \mathbf{s}^2)$ is odd. Then, using the basic property $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$, where X and Y are random variables, we obtain (28).

Appendix 4: The Fisher information matrix

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

$$\begin{aligned} \ell(\mathbf{h}, \sigma^2) &= K \sum_{p=1}^P \ln(\sigma_p^2) + \sum_{k=1}^K \sum_{p=1}^P \\ &\quad \times \frac{(y_{p,k}^R - H_p^R)^2 + (y_{p,k}^I - H_p^I)^2}{\sigma_p^2}. \end{aligned}$$

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

$$\mathbf{G} = \begin{bmatrix} \mathbf{F}^R & -\mathbf{F}^I \\ \mathbf{F}^I & \mathbf{F}^R \end{bmatrix}, \quad (40)$$

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

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

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 $\ell(\mathbf{h}, \sigma^2)$ with respect to $\mathbf{h}^R, \mathbf{h}^I$, and σ^2 are

$$\begin{aligned} \partial_{h_1^R} \ell(\mathbf{h}, \sigma^2) &= -2K \sum_{p=1}^P \frac{g_{p,1} b_p^R + g_{p+p,1} b_p^I}{\sigma_p^2}, \\ \partial_{h_1^I} \ell(\mathbf{h}, \sigma^2) &= -2K \sum_{p=1}^P \frac{g_{p,L+1} b_p^R + g_{p+p,L+1} b_p^I}{\sigma_p^2}, \\ \partial_{\sigma_p^2} \ell(\mathbf{h}, \sigma^2) &= \frac{K}{\sigma_p^2} - K \frac{s_p^2 + |b_p^R|^2 + |b_p^I|^2}{\sigma_p^4}, \end{aligned}$$

which yields

$$\partial_{h_1^R h_1^R}^2 \ell(\mathbf{h}, \sigma^2) = 2K \sum_{p=1}^P \frac{g_{p,1} g_{p,1} + g_{p+p,1} g_{p+p,1}}{\sigma_p^2} \quad (41)$$

$$\partial_{h_1^R h_1^I}^2 \ell(\mathbf{h}, \sigma^2) = 2K \sum_{p=1}^P \frac{g_{p,1} g_{p,L+1} + g_{p+p,1} g_{p+p,L+1}}{\sigma_p^2} \quad (42)$$

$$\partial_{h_1^R, h_1^I, h_2^I}^2 \ell(\mathbf{h}, \sigma^2) = 2K \sum_{p=1}^P \times \frac{g_{p,L+l_1} g_{p,L+l_2} + g_{p+P,L+l_1} g_{p+P,L+l_2}}{\sigma_p^2} \tag{43}$$

$$\partial_{h_1^R, \sigma_p^2}^2 \ell(\mathbf{h}, \sigma^2) = 2K \frac{g_{p,l} b_p^R + g_{p+P,l} b_p^I}{\sigma_p^4} \tag{44}$$

$$\partial_{\sigma_{p_1}^2, \sigma_{p_2}^2}^2 \ell(\mathbf{h}, \sigma^2) = \left(-\frac{K}{\sigma_{p_1}^4} + 2K \frac{s_{p_1}^2 + |b_{p_1}^R|^2 + |b_{p_1}^I|^2}{\sigma_{p_1}^6} \right) \delta_{p_1}^{p_2} \tag{45}$$

Now, the expectation needs to be taken to find \mathbf{J} . The expectations of (41), (42), and (43) are unchanged. From the definition of $b_p = \bar{y}_p - H_p$ and (38), the expectation of (44) is 0, and from (39), the expectation of (45) is $\frac{K}{\sigma_{p_1}^4} \delta_{p_1}^{p_2}$.

Acknowledgements

This work has been carried out in the framework of The ELSAT2020 project which is co-financed by the European Union with the European Regional Development Fund, the French state, and the Hauts de France Region Council. This work has been supported in part (or partially) by IRCICA USR 3380 CNRS-Univ (project: connected objects), F-59000 Lille, France (http://www.ircica.univ-lille1.fr).

Funding

We confirm that we do not have a funding source.

Authors' contributions

AD and EPS both worked on the calculation of the theoretical results. EPS made the matlab programs. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Laboratoire Paul Painlevé, USTL-UMR-CNRS 8524. UFR de Mathématiques, Bât. M2, 59655 Villeneuve d'Ascq Cédex, France. ²IEMN lab, TELICE group, University of Lille, Lille, France.

Received: 8 February 2017 Accepted: 19 September 2017

Published online: 29 September 2017

References

1. K Ohno, T Ikegami, Interference mitigation technique for coexistence of pulse-based UWB and OFDM. *EURASIP J. Wirel. Commun. Netw.* **2008**(1), 285683 (2008)
2. J-W van Bloem, R Schiphorst, T Kluwer, C Slump, in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on*. Interference measurements in IEEE 802.11 communication links due to different types of interference sources, (Shanghai, 2012), pp. 1–6
3. A Coulson, Narrowband interference in pilot symbol assisted OFDM systems. *IEETrans, EWireless Commun.* **3**(6), 2277–2287 (2004)
4. AJ Coulson, Bit error rate performance of OFDM in narrowband interference with excision filtering. *IEEE Trans Wireless Commun.* **5**(9), 2484–2492 (2006)

5. A Jeremic, T Thomas, A Nehorai, OFDM channel estimation in the presence of interference. *IEEE Trans. Signal Process.* **52**(12), 3429–3439 (2004)
6. J Zhou, J Qin, Y-C Wu, Variational inference-based joint interference mitigation and OFDM equalization under high mobility. *IEE Signal Signal Process. Lett.* **22**(11), 1970–1974 (2015)
7. S Lee, K Kwak, J Kim, D Hong, Channel estimation approach with variable pilot density to mitigate interference over time-selective cellular OFDM systems. *IEEE Trans. Wireless Commun.* **7**(7), 2694–2704 (2008)
8. Y Zhang, X Zhang, D Yang, in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*. A robust least square channel estimation algorithm for OFDM systems under interferences, (Shanghai, 2013), pp. 3122–3127
9. T Li, W-H Mow, V Lau, M Siu, R Cheng, R Murch, Robust joint interference detection and decoding for OFDM-based cognitive radio systems with unknown interference. *IEEE J. Sel. Areas Commun.* **25**(3), 566–575 (2007)
10. M Morelli, M Moretti, Robust frequency synchronization for OFDM-based cognitive radio systems. *IEEE Trans. Wireless Commun.* **7**(12), 5346–5355 (2008)
11. M Morelli, M Moretti, Channel estimation in OFDM systems with unknown interference. *IEEE Trans. Wireless Commun.* **8**(10), 5338–5347 (2009)
12. X Mestre, JR Fonollosa, ML approaches to channel estimation for pilot-aided multirate DS/CDMA systems. *IEEE Trans. Signal Process.* **50**(3), 696–709 (2002)
13. RT Behrens, LL Scharf, Signal processing applications of oblique projection operators. *IEEE Trans. Signal Process.* **42**(6), 1413–1424 (1994)
14. RT Behrens, LL Scharf, Corrections to "signal processing applications of oblique projection operators" [correspondence]. *IEEE Trans. Signal Process.* **44**(5) (1300)
15. B Cao, Q-Y Zhang, L Jin, N-T Zhang, Oblique projection polarization filtering-based interference suppressions for radar sensor networks. *EURASIP J. Wirel. Commun. Netw.* **2010**(1), 605103 (2010)
16. W Qiu, SK Saleem, E Skafidas, Identification of MIMO systems with sparse transfer function coefficients. *EURASIP J. Adv. Signal Process.* **2012**(1), 104 (2012)
17. H Hartley, K Jayatilake, Estimation for linear models with unequal variances. *J. Am. Stat. Assoc.* **68**(341), 189–192 (1973)
18. EB Andersen, Asymptotic properties of conditional maximum-likelihood estimators. *J. R. Stat. Soc. Ser. B Methodol.* 283–301 (1970)
19. TH Szatrowski, Necessary and sufficient condition for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Ann. Stat.* **8**, 802–810 (1980)
20. AP Dempster, NM Laird, DB Rubin, Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Methodol.* 1–38 (1977)
21. WG Cochran, in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 30, no. 02. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance (Cambridge Univ Press, 1934), pp. 178–191

Submit your manuscript to a SpringerOpen journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com