BAYESIAN BLIND FIR CHANNEL ESTIMATION ALGORITHMS IN SIMO SYSTEMS

Samir-Mohamad Omar [*], *Dirk Slock* [*]

Mobile Communications Department, Eurecom
BP 193, 06904 Sophia Antipolis Cdx, France
omar@eurecom.fr, slock@eurecom.fr

Oussama Bazzi

Department of Physics and Electronics
Faculty of Sciences I, Lebanese University
Hadath ,Beirut, Lebanon
obazzi@ul.edu.lb

ABSTRACT

Blind channel estimation techniques were developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. On the other hand, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. In this paper we explore a Bayesian approach to blind channel estimation, exploiting a priori information on fading channels. We mainly focus on joint ML/MAP estimation of channels and symbols on one hand, and on ML/MAP estimation of channels with elimination of symbols on the other hand. As a consequence, a unified framework in addition to three new Bayesian estimators are introduced where their performance is compared by simulations to three existing non-Bayesian estimators. In the same context, we provide an insightful discussion of the accurate way of deriving the Bayesian Cramer Rao bound (BCRB) with an emphasis on its singularity.

1. INTRODUCTION

In the context of blind channel estimation, our main goal is to estimate the channel at the receiver and feed it to a certain equalizer (Linear, Decision Feedback ..) to detect the symbols. To accomplish this task blindly at the receiver we should exploit every piece of information exists related to any element of the transmission system. Moreover, sometimes assumptions are made and consequently the accuracy of the estimated parameters depends on how close those assumptions are for the reality. We will focus in this paper on the second order statistics and specifically on the maximum likelihood (ML) and/or maximum a posteriori methods (MAP). Two approaches exist in the literature on how to tackle the problem. The first approach is based on the fact that the symbols are considered as deterministic unknowns to be jointly estimated with the channel. Such algorithm is called Deterministic (or conditional) Maximum Likelihood (DML) method [1]. The second approach is based on treating the symbols as random quantities with known prior information to be eliminated or jointly estimated. When the symbols are eliminated, the method is called Gaussian (or unconditional) Maximum Likelihood (GML) [2] see also [3] for its implementation in sensor array processing. While when they are jointly estimated, the method is called GMAP-ML [4]. This is because we use maximum a posteriori (MAP) for symbols and ML for channels and noise variance. Furthermore, in all these approaches the channel was considered as deterministic unknown however, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. The concept of Bayesian blind channel estimation was introduced in [5], with in particular some considerations on identifiability issues. However, in [6] we discussed briefly some classical Bayesian algorithms and introduced the concept of variational Bayesian in the context of MIMO OFDM. Apart from the variational Bayesian techniques, we develop in this paper classical Bayesian algorithms that treat the channel as random with known prior information rather than deterministic. Once the channel is treated as random, we are within the framework of Bayesian blind channel estimation and there are are three cases to be handled. In the first case the symbols are considered as deterministic unknowns to be jointly estimated with channel. We call this method as ML-GMAP, for a similar reasoning discussed above. In the second case, the symbols are again to be jointly estimated with the channel but this time they are considered as random with known prior Gaussian distribution. We call this method GMAP-GMAP. In the third case, the symbols are again random with known prior Gaussian distribution but they are going to be eliminated rather than estimated. We call this method GMAP-Elm-GMAP where ”Elm” stands for elimination of symbols. Consequently, in section 3 we revisit three already existing deterministic estimators and develop three new Bayesian ones. Therefore, with the introduction of the Bayesian blind channel estimation algorithms the picture is broadened considerably and to sum up we depict the current picture in Table 1.

Algorithm	Symbols	Channel	Elm of Sym	Novel
ML-ML (DML)	Det	Det	No	No
GMAP-ML	Gaussian	Det	No	No
GMAP-Elm-ML	Gaussian	Det	Yes	No
ML-GMAP	Det	Gaussian	No	Yes
GMAP-GMAP	Gaussian	Gaussian	No	Yes
GMAP-Elm-GMAP	Gaussian	Gaussian	Yes	Yes

Table 1. Summary of all algorithms

2. SIMO FIR TX SYSTEM MODEL

In blind channel identification, a multichannel framework can be obtained from oversampling a received signal and leads to a Single Input Multiple Output (SIMO) vector channel representation. The multiple FIR channels we obtain in this representation can also be obtained from multiple received signals from an array of antennas or from a combination of both. To further develop the case of oversampling, consider a linear digital modulation over a linear channel with

[*] EURECOM's research is partially supported by its industrial members: BMW Group, Swisscom, Cisco, ORANGE, SFR, ST Ericsson, Thales, Symantec, SAP, Monaco Telecom. The research reported herein was also partially supported by the French ANR project SESAME, the EU FET project CROWN, the EU NoE Newcom++ and a PACA regional scholarship BDO550.

additive noise so that the received signal $y(t)$ has the following form

$$y(t) \ = \ \sum_k h(t-kT)a(k) \ + \ v(t). \tag{1}$$

In (1) $a(k)$ are the transmitted symbols, T is the symbol period and $h(t)$ is the channel impulse response. The channel is assumed to be FIR with length NT. If the received signal is oversampled at the rate $\frac{m}{T}$ (or if m different samples of the received signal are captured by m sensors every T seconds, or a combination of both), the discrete input-output relationship can be written as:

$$\boldsymbol{y}(k) = \sum_{i=0}^{N-1} \boldsymbol{h}(i)a(k-i) + \boldsymbol{v}(k) = \boldsymbol{H}A_N(k) + \boldsymbol{v}(k) \tag{2}$$

where $\boldsymbol{y}(k) = [y_1^H(k) \cdots y_m^H(k)]^H, \boldsymbol{h}(i) = \left[h_1^H(i) \cdots h_m^H(i)\right]^H$, with $h_m(i)$ denotes the ith tap of the mth receiving antenna, $\boldsymbol{v}(k) = [v_1^H(k) \cdots v_m^H(k)]^H$, $\boldsymbol{H} = [\boldsymbol{h}(N-1) \cdots \boldsymbol{h}(0)]$, $A_N(k) = \left[a(k-N+1)^H \cdots a(k)^H\right]^H$ and superscript ^{H} denotes Hermitian transpose. Let $\mathbf{H}(z) = \sum_{i=0}^{N-1} \boldsymbol{h}(i)z^{-i} = [\mathrm{H}_1^H(z) \cdots \mathrm{H}_m^H(z)]^H$ be the SIMO channel transfer function, and $\boldsymbol{h} = \left[\boldsymbol{h}^H(N-1) \cdots \boldsymbol{h}^H(0)\right]^H$. Consider additive independent white Gaussian circular noise $\boldsymbol{v}(k)$ with $r_{\boldsymbol{vv}}(k-i) = \mathrm{E}\,\boldsymbol{v}(k)\boldsymbol{v}(i)^H = \sigma_v^2 I_m\,\delta_{ki}$. Assume we receive M samples:

$$\boldsymbol{Y}_M(k) \ = \ \mathcal{T}_M(\boldsymbol{h})\,A_{M+N-1}(k) + \boldsymbol{V}_M(k) \tag{3}$$

where $\boldsymbol{Y}_M(k) = [\boldsymbol{y}^H(k-M+1) \cdots \boldsymbol{y}^H(k)]^H$ and similarly for $\boldsymbol{V}_M(k)$, and $\mathcal{T}_M(\boldsymbol{h})$ is a block Toeplitz matrix with M block rows and $[\boldsymbol{H} \quad 0_{m\times(M-1)}]$ as first block row. We shall simplify the notation in (3) with $k = M-1$ to

$$\boldsymbol{Y} \ = \ \mathcal{T}(\boldsymbol{h})\,A + \boldsymbol{V} = \mathcal{A}\boldsymbol{h} + \boldsymbol{V} \ . \tag{4}$$

where $\mathcal{A}$ is a block Toeplitz matrix filled with the elements of A. We assume that $mM > M+N-1$ in which case the channel convolution matrix $\mathcal{T}(\boldsymbol{h})$ has more rows than columns. If the $\mathrm{H}_i(z)$, $i = 1,\ldots,m$ have no zeros in common, then $\mathcal{T}(\boldsymbol{h})$ has full column rank.

3. A UNIFIED FRAMEWORK FOR DIFFERENT ALGORITHMS

As we have shown in Table 1 there are six possible estimators that can be classified into two categories. In the first category the subject of the estimators is to estimate the channel and the symbols jointly by making some assumptions on the channel and the symbols. If we denote by θ the unknown parameters to be estimated then it is given by:

$$\theta = [A^H, \boldsymbol{h}^H]^H \tag{5}$$

The likelihood function is given by:

$$f(Y,\theta) = f(Y/\theta)f(\theta) \tag{6}$$

where $f(\theta)$ stands for the probability density function (pdf) of θ, $f(Y,\theta)$ stands for the joint probability density function of Y and θ and $f(Y/\theta)$ stands for the pdf of Y conditioned on θ is given or known. Once we substitute θ in (6) by its elements we get:

$$f(Y, A, \boldsymbol{h}) = f(Y/A, \boldsymbol{h})f(A)f(\boldsymbol{h}) \tag{7}$$

Since the symbols and the channel are a priori independent of each other we can write $f(\theta) = f(A)f(\boldsymbol{h})$. Of course on the basis of how we treat the symbols and the channel both $f(A)$ and $f(\boldsymbol{h})$ differs from one estimator to another as we shall see in the sequel. Knowing that the cost function for the estimator is derived by maximizing the log-likelihood function, hence we apply the log function on both sides of (7) to get:

$$\ln[f(Y, A, \boldsymbol{h})] = \ln[f(Y/A, \boldsymbol{h})] + \ln[f(A)] + \ln[f(\boldsymbol{h})] \tag{8}$$

However, in the second category the subject of the estimators is to estimate the channel and the noise variance only while the symbols are supposed to be eliminated during the estimation process. Thus

$$\theta = [\boldsymbol{h}^H, \sigma_v^2]^H \tag{9}$$

Once we substitute θ in (6) by its elements we get:

$$f(Y, \boldsymbol{h}, \sigma_v^2) = f(Y/\boldsymbol{h}, \sigma_v^2)f(\boldsymbol{h})f(\sigma_v^2) \tag{10}$$

Again, since the cost function for the estimator is derived by maximizing the log-likelihood function, hence we apply the log function on both sides of (10) to get:

$$\ln[f(Y, \boldsymbol{h}, \sigma_v^2)] = \ln[f(Y/\boldsymbol{h}, \sigma_v^2)] + \ln[f(\boldsymbol{h})] + \ln[f(\sigma_v^2)] \tag{11}$$

We will develop in the following sections the cost functions of all the estimators that belong to both categories and provide a closed form formula for both the estimated channel and symbols where it is possible. It is worthy to note here that since the channel is treated as random rather than deterministic in some of the above mentioned estimators (ML-GMAP, GMAP-GMAP, GMAP-Elm-GMAP) in both categories, these estimators are considered as an example of the Bayesian blind channel estimation.

3.1. ML-ML (DML)

We start with ML-ML or what is called DML in the literature [1]. In this case both the symbols and the channel are considered as deterministic unknowns to be estimated. Hence it belongs to the first category and consequently the log-likelihood function is given by (8). Moreover, since both are deterministic we have $f(\boldsymbol{h}) = \boldsymbol{h}^o\delta(\boldsymbol{h}-\boldsymbol{h}^o)$ and $f(A) = A^o\delta(A-A^o)$ where $\boldsymbol{h}^o$ and A^o represent respectively the true values of the channel and the symbols. As the pdfs of both symbols and channel are constant, this means that they have no influence on the maximization of (8). Hence, we can derive the cost function by maximizing $\ln[f(Y/A, \boldsymbol{h})]$ directly where $f(Y/A, \boldsymbol{h}) = \frac{1}{(\pi\sigma_v^2)^{Mm}} exp[-\frac{1}{\sigma_v^2}(Y - \mathcal{T}(h)A)^H(Y - \mathcal{T}(h)A)]$. Thus the cost function is given by:

$$\min_{A,\boldsymbol{h}} \ ||Y - \mathcal{T}(h)A||^2 \tag{12}$$

The joint optimization of this cost function in both the channel ($\boldsymbol{h}$) and the symbols (A) is difficult. Fortunately, the observation is linear in both the channel and the symbols. In other words, we have a separable nonlinear LS problem, which allows us to reduce the complexity considerably. The nonlinear LS optimization can be done by iterating between minimization with respect to A and $\boldsymbol{h}$. By doing so, we get the following estimates:

$$\widehat{\boldsymbol{h}} = (\mathcal{A}^H\mathcal{A})^{-1}\mathcal{A}^H\boldsymbol{Y} \tag{13}$$

$$\widehat{A} = (\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h}))^{-1}\mathcal{T}^H(\boldsymbol{h})\boldsymbol{Y} \tag{14}$$

3.2. GMAP-ML

In this estimator [2],[4] we treat the symbols as random with Gaussian distribution while the channel is considered deterministic to be jointly estimated with the symbols. This estimator also belongs to the first category, thus the log-likelihood function is given by (8). Moreover, $f(A) = \frac{1}{(\pi\sigma_a^2)^{M+N-1}} exp[-\frac{A^HA}{\sigma_a^2}]$ and $f(\boldsymbol{h}) = \boldsymbol{h}^o\delta(\boldsymbol{h}-\boldsymbol{h}^o)$. It is obvious here that $\ln[f(\boldsymbol{h})]$ can be omitted without affecting the maximization of the log-likelihood function in (8). Hence, the cost function is given by:

$$\min_{A,\boldsymbol{h}} \frac{1}{\sigma_v^2}||Y - \mathcal{T}(h)A||^2 + \frac{||A||^2}{\sigma_a^2} \tag{15}$$

Following the same methodology used in ML-ML estimator we get:

$$\widehat{\boldsymbol{h}} = (\mathcal{A}^H\mathcal{A})^{-1}\mathcal{A}^H\boldsymbol{Y} \tag{16}$$

$$\widehat{A} = (\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2}I_{mN})^{-1}\mathcal{T}^H(\boldsymbol{h})\boldsymbol{Y} \tag{17}$$

3.3. GMAP-Elm-ML (GML)

This estimator belongs to the second category [2], hence we are interested in estimating the channel and the variance of the noise only while the symbols are supposed to be eliminated during the estimation process. Furthermore, the log-likelihood function is given by (11) where we consider the channel and the noise variance to be deterministic while the symbols have a Gaussian distribution. Here again, $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ have no influence on maximizing (11). Substituting $f(Y/\boldsymbol{h},\sigma_v^2) = \frac{1}{(\pi)^{Mm}|R_{YY}|}exp[-Y^H R_{YY}^{-1} Y)]$ where $R_{YY} = E\ \boldsymbol{YY}^H = \sigma_a^2 \mathcal{T}(\boldsymbol{h})\mathcal{T}(\boldsymbol{h})^H + \sigma_v^2 I_{Mm}$ in (11) after omitting $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ we get:

$$\min_{\boldsymbol{h},\sigma_v^2} \ \ln|R_{YY}| + \text{tr}\,(R_{YY}^{-1}\hat{R}_{YY}) \tag{18}$$

This cost function can be minimized by resorting to the method of scoring [7]. This method consists in an approximation of the Newton-Raphson algorithm which finds an estimate $\theta(i)$ at iteration i from $\theta(i-1)$, the estimate at iteration $i-1$, as:

$$\theta^{(i)} = \theta^{(i-1)} - \mu\left[\mathcal{F}^{''}|_{\theta^{(i-1)}}\right]^{-1}\ \mathcal{F}^{'}|_{\theta^{(i-1)}} \tag{19}$$

where $\mathcal{F}(\theta)$ is the cost function in (18), $\mathcal{F}^{''}$ is the hessian, $\mathcal{F}^{'}$ is the gradient of the cost function and μ is the step length that should be appropriately chosen to guarantee convergence to a local minimum. The method of scoring approximates the Hessian by its expected value, which is here the Gaussian Fisher Information Matrix (FIM). This approximation is justified by the law of large numbers as the number of data is generally large. In our case, the FIM is singular, and as a consequence formula (19) cannot be applied directly so we take the Moore-Penrose pseudo inverse of the FIM. On the other hand, it can be easily shown that this estimator is less sensitive to the common zeros problem. In fact by applying the matrix inversion lemma we can readily prove that $R_{YY}^{-1} = \frac{1}{\sigma_v^2}[I - \mathcal{T}(\boldsymbol{h})(\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2}I_{mN})^{-1}\mathcal{T}^H(\boldsymbol{h})]$.Therefore, even if $\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h})$ is rank deficient, the cost function may not blow up thanks to the regularization parameter $\frac{\sigma_v^2}{\sigma_a^2}I_{mN}$ introduced by the prior information on the symbols.

3.4. ML-GMAP

This is our first novel estimator where we introduce the concept of blind Bayesian channel estimation by treating the channel as random with Gaussian distribution $f(\boldsymbol{h}) = \frac{1}{(\pi)^{mN}|C_h^o|}exp[-\boldsymbol{h}^H C_h^{o-1}\boldsymbol{h}]$. However, the symbols are considered as deterministic unknowns to be jointly estimated with the channel hence this estimator belongs to the first category where the log-likelihood function is given by (8). Moreover, here again $\ln[f(A)]$ has no effect on maximizing (8) so it can be omitted. Therefore, the cost function is given by:

$$\min_{A,\boldsymbol{h}} \ \frac{1}{\sigma_v^2}||Y - \mathcal{T}(h)A||^2 + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{20}$$

Once again here, following the same methodology used in ML-ML estimator we get:

$$\widehat{\boldsymbol{h}} = (\mathcal{A}^H\mathcal{A} + \sigma_v^2 C_h^{o-1})^{-1}\mathcal{A}^H\boldsymbol{Y} \tag{21}$$

$$\widehat{A} = (\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h}))^{-1}\mathcal{T}^H(\boldsymbol{h})\boldsymbol{Y} \tag{22}$$

3.5. GMAP-GMAP

This is our second novel estimator where both the channels and the symbols are assumed random with Gaussian distribution and are supposed to be estimated jointly. Hence, this estimator in its turn belongs to the first category and its log-likelihood is given by (8). By substituting the terms in (8) by their corresponding functions we deduce the cost function as follows:

$$\min_{A,\boldsymbol{h}} \frac{1}{\sigma_v^2}||Y - \mathcal{T}(h)A||^2 + \frac{||A||^2}{\sigma_a^2} + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{23}$$

Also here, following the same methodology used in ML-ML estimator we get:

$$\widehat{\boldsymbol{h}} = (\mathcal{A}^H\mathcal{A} + \sigma_v^2 C_h^{o-1})^{-1}\mathcal{A}^H\boldsymbol{Y} \tag{24}$$

$$\widehat{A} = (\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2}I_{mN})^{-1}\mathcal{T}^H(\boldsymbol{h})\boldsymbol{Y} \tag{25}$$

3.6. GMAP-Elm-GMAP

This is our third novel estimator and it belongs to the second category since the symbols are supposed to be eliminated. It can be considered as an extension to GMAP-Elm-ML by exploiting the prior information exists about the channel. Its log-likelihood function is given by (11) but this time $\ln[f(\boldsymbol{h})]$ can't be omitted. Substituting the terms in (11) by their corresponding functions we get the cost function as follows:

$$\min_{\boldsymbol{h},\sigma_v^2} \ \ln|R_{YY}| + \text{tr}\,(R_{YY}^{-1}\hat{R}_{YY}) + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{26}$$

This cost function can be minimized using the scoring method discussed in case of GMAP-Elm-ML estimator.

4. BAYESIAN CRAMER RAO BOUND (BCRB)

It is well known that the channel can be estimated blindly up to a scalar ambiguity $\rho e^{j\phi}$ where ρ stands for the amplitude and ϕ stands for the phase. In [8] the Bayesian CRB in the context of cooperative OFDM was derived where the authors claimed (section III-B) that the knowledge of the prior information of the channel eliminates the ambiguity of the blind channel estimation. We will show in this short discussion that the prior information of the channel doesn't provide any information about the phase while it provides only a very limited information about the amplitude. Consequently, the ambiguity is not totally removed and the singularity persists. From the pdf of the channel shown before, we can easily notice that the prior Fisher information Matrix (FIM) is given by C_h^{o-1}. Usually the total FIM is the sum of the prior FIM and the FIM of the data. The latter is singular while the former has usually a full rank. Hence, the total FIM has a full rank. At the first glance this will lead to the same conclusion that was drawn in [8] namely, the prior information eliminates the blind channel ambiguity. However, a closer look at the problem will prove that this result is inaccurate at all.

Suppose we have a channel $\boldsymbol{h}^{'} = \rho e^{j\phi}\boldsymbol{h}$ then the FIM of this channel is given by $\frac{1}{\rho^2}C_h^{0-1}$ where we note that ϕ has been completely absorbed. This result shows that the prior FIM has no capability to provide any piece of information regarding the phase. If so, then the question is why the prior FIM has a full rank and doesn't admit any singularity? In order to answer this question and show that the prior FIM is singular we should reparametrize the problem between our hands. Moreover, we should also resort to splitting the complex channel parameters into their real and imaginary parts. When we accomplish the two previous steps and derive the FIM for the new reparametrized prior we will find it singular for sure. To commence with this task, lets take the first tap of the channel as a common factor we get $\boldsymbol{h} = \rho e^{j\phi}\boldsymbol{h}^{'}$ where $\boldsymbol{h}^{'} = [1\ \ \bar{\boldsymbol{h}}^H]^H$. Denote by $\theta = [\bar{\boldsymbol{h}}^{rT}, \bar{\boldsymbol{h}}^{sT}, \rho, \phi]$ the set of parameters to be estimated where $\bar{\boldsymbol{h}}^{rT}$ and $\bar{\boldsymbol{h}}^{sT}$ denotes respectively the real and the imaginary parts of $\bar{\boldsymbol{h}}$. Due to the lack of space we will not go into the detailed derivation nevertheless we will show below the resulting prior FIM (2mN x 2mN)which is given by:

$$FIM_{priro} = 2 \begin{bmatrix} C_h^o(1,1)\bar{C_h^o}^{-1} & 0 & 0 & 0 \\ 0 & C_h^o(1,1)\bar{C_h^o}^{-1} & 0 & 0 \\ 0 & 0 & C_h^{o-1}(1,1) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (27)$$

where $C_h^o(1,1)$ denotes the element that lies in the first row and first column of C_h^o and $\bar{C_h^o}^{-1}$ can be obtained from C_h^{o-1} by omitting the first row and the first column. It is evident now that the prior FIM admits one singularity that corresponds to the phase and it provides only the variance of the ambiguous amplitude $C_h^o(1,1)$ and not the amplitude itself. Hence, this information is considered limited and incomplete. Now to pursue the derivation of the BCRB we should play the same game with the FIM of the data. Doing so, we can show that the latter admits two singularities, one corresponds to the amplitude and the other corresponds to the phase. However, the total FIM which is the sum of the prior and the data FIMs will admit only one singularity that corresponds to the phase. This is because the prior FIM ameliorate only the singularity that corresponds to the amplitude which results from the FIM of the data. Therefore, the prior FIM only contributes to fix one singularity while it has no means to deal with the other. As a consequence, the resulting BCRB which is defined as the inverse of the total FIM is still singular and needs an additional constraint to fix the phase ambiguity.

5. SIMULATIONS

In this section we try to shed light by means of MonteCarlo simulations on the advantages of blind Bayesian compared to the blind non-Bayesian channel estimation. In each MonteCarlo simulation we generate a Rayleigh fading channel with exponentially decaying power delay profile (PDP)(assumed known) for the channel between each transmitting and receiving antenna pair as follows: e^{-wn} where $n = 0 : N-1$ and w is a constant that controls how steep is the decaying of the PDP. Hence, $C_{\mathbf{h}}^o$ is the diagonal matrix $C_{\mathbf{h}}^o = I_m \otimes C$ where $C = \text{diag}\{e^{-wn}, n = 0 : N-1\}$. As for the symbols, we generate random 8PSK symbols to reflect the real world case. The performance of the different channel estimators is evaluated by means of the Normalized MSE (NMSE) vs. SNR. The SNR is defined as: $\text{SNR} = \frac{||\mathcal{T}(h)A||^2}{mM\,\sigma_v^2}$ while The NMSE is defined as $\frac{avg\,||\boldsymbol{h}-\hat{\hat{\boldsymbol{h}}}||^2}{avg\,||\boldsymbol{h}||^2}$ where $\hat{\hat{\boldsymbol{h}}} = \frac{\hat{\boldsymbol{h}}^H\boldsymbol{h}}{||\hat{\boldsymbol{h}}||^2}\boldsymbol{h}$ is the channel estimate adjusted by the least squares constraint to fix the scalar ambiguity that results from the blind channel estimation. All the simulations are initialized by the Subchannel Response Matching (SRM) estimate [9]. In Figure 1, we take a look at the considerable gain (4dB) offered by both GMAP-GMAP and GMAP-Elm-GMAP over GMAP-ML at high SNR. Moreover, these two novel Bayesian estimators that we have introduced have the potential to exceed even GMAP-Elm-ML by couple of dBs. This emphasizes the indispensable role of exploiting the prior information of the channel in enhancing the estimation quality at the receiver. It is worthy to note that in simulating GMAP-Elm-GMAP and GMAP-Elm-ML in Figure 1, we consider σ_v^2 to be known hence, we only estimate the channel. This permits us to make a fair comparison between joint estimation of the channel and the symbols (GMAP-GMAP, GMAP-ML) and the estimation of the channel with marginalization of the symbols (GMAP-Elm-GMAP, GMAP-Elm-ML). Taking a close look at Figure 1 shows that, in such a scenario, the estimation of the channel with marginalization of the nuisance parameters (symbols) outperforms a little bit the joint estimation of the channel and the symbols. This holds true whatever is the assumption made for the channel namely, deterministic or Bayesian although it is more evident in the deterministic case.

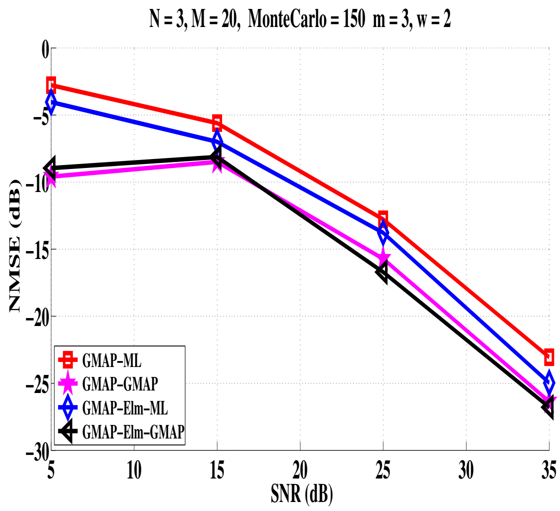

Fig. 1. NMSE vs. SNR for different estimators.

6. CONCLUSION

After we have introduced previously the concept of blind Bayesian channel estimation without providing any specific algorithm, the main message of this paper is to prove that there is yet a classical way to implement the blind Bayesian channel estimation apart form the variational Bayesian techniques introduced recently. This concept has been shown by augmenting the cost functions of some ML/MAP estimators that already exist in the literature. Moreover, the novel Bayesian estimators that we have derived in this paper show a considerable performance gain compared to the deterministic ones. Another aspect that has been addressed in this paper is the limited contribution of the prior information of the channel in fixing the ambiguities that result from the blind channel estimation problem. We have derived the reparametrized prior FIM of the channel showing that it is singular and is not capable of providing any information related to the ambiguous phase while it provides a limited information to fix the amplitude ambiguity.

7. REFERENCES

[1] D.T.M. Slock, "Subspace Techniques in Blind Mobile Radio Channel Identification and Equalization using Fractional Spacing and/or Multiple Antennas," in *Proc. 3rd International Workshop on SVD and Signal Processing*, Leuven, Belgium, Aug. 22-25 1994.

[2] E. de Carvalho and D.T.M. Slock, "Maximum-Likelihood FIR Multi-Channel Estimation with Gaussian Prior for the Symbols," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.

[3] A. Renaux, P. Forster, E. Boyer, and P. Larzabal, "Unconditional Maximum Likelihood Performance at Finite Number of Samples and High Signal-to-Noise Ratio," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2358 – 2364, 2007.

[4] D.T.M. Slock and C.B. Papadias, "Further Results on Blind Identification and Equalization of Multiple FIR Channels," in *Proc. ICASSP 95 Conf.*, Detroit, Michigan, May 1995.

[5] D.T.M. Slock, "Bayesian Blind and Semiblind Channel Estimation," in *Proc. IEEE SAM*, Barcelona, Spain, July 2004.

[6] S.-M. Omar and D.T.M. Slock, "Variational Bayesian Blind and Semiblind Channel Estimation," in *Proc. IEEE ISCCSP*, Limassol, Cyprus, March 2010.

[7] P.E. Kay, W. Murray, and M.H. Wright, *Practical Optimization*, Academic press, London, 1981.

[8] Wu Wei, Jia Xin, Cai Yueming, and Wu Xiaoxue ;, "Bayesian Cramer-rao Bound for channel estimation in cooperative OFDM," in *IEEE WCSP*, Suzhou, China, Oct. 2010.

[9] G. Xu, H. Liu, L. Tong, and T. Kailath, "A Least Squares Approach to Blind Channel Identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.