Variational Bayesian Blind and Semiblind Channel Estimation

Samir-Mohamad Omar, Dirk T.M. Slock

***Abstract*— Blind and semiblind channel estimation is a topic that enjoyed explosive developments throughout the nineties, and then came to a standstill, probably because of perceived unsatisfactory performance. Blind channel estimation techniques were developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. Such blind channel estimates, especially those based on subspaces in the data, are often only partial and ill-conditioned. 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 recent years, such prior information on the channel has started to get exploited in pilot-based channel estimation, since often the pure pilot-based (deterministic) channel estimate is of limited quality due to limited pilots. In this paper we explore a Bayesian approach to (semi-)blind channel estimation, exploiting a priori information on fading channels. In the case of deterministic unknown input symbols, it suffices to augment the classical blind (quadratic) channel criterion with a quadratic criterion reflecting the Rayleigh fading prior. In the case of a Gaussian symbol model the blind criterion is more involved. The joint ML/MAP estimation of channels, deterministic unknown symbols, and channel profile parameters can be conveniently carried out using Variational Bayesian techniques. Variational Bayesian techniques correspond to alternating maximization of a likelihood w.r.t. subsets of parameters, but taking into account the estimation errors on the other parameters. To simplify exposition, we elaborate the details for the case of MIMO OFDM systems.**

I. Introduction

Blind and semiblind channel estimation techniques have been developed and are usually evaluated for a given channel realization, i.e. with a deterministic channel model, see [1] for an overview of such techniques. Such blind channel estimates, especially those based on subspaces in the data, are often only partial and ill-conditioned. Indeed, only part of the channel is blindly identifiable, especially in the case of MIMO channels. The type of blind channel estimation techniques we are mostly referring to here involve an FIR multichannel and are typically based on the second-order statistics of the received signal. Two types of techniques can be considered, treating the unknown input symbols as either deterministic unknowns or Gaussian white noise. In the first case, the techniques are often based on the subspace structure induced in the data by the multichannel aspect. The part of the channel that can be identified blindly is larger in the Gaussian input model case than in the deterministic input model case, but is in any case incomplete. Many of the deterministic input approaches are also quite sensitive to a number of hypotheses such as correct channel length (filter order) and no channel zeros. In general this means that these blind channel estimates can often become ill-conditioned, when the channel impulse response is tapered (e.g. due to a pulse shape filter) or when the channel is close to having zeros. In fact this means that the blind information on the channel can be substantial, but is limited to only part of the channel.

An overview of blind channel estimation techniques can be found in [1] for SIMO systems and in [2] for MIMO systems. Specific blind channel estimation techniques for Cyclic Prefix systems, as will be considered here, were introduced in [3], see also [4]. The concept of Bayesian blind channel estimation was introduced in [5], with in particular some considerations on identifiability issues. In this paper we focus on algorithms. The superscripts $.^*$, $.^T$ and $.^H$ denote complex conjugate, transpose and complex conjugate (Hermitian) transpose resp.

EURECOM's research is partially supported by its industrial members: Swisscom, Thales, SFR , Orange, STEricsson, Cisco, BMW Group , SAP, Monaco Telecom and Symantec. The research reported herein has also been partially supported by the European FP7 NoE NewCom++ and FET project CROWN, and by the French ANR projects APOGEE and SESAME.

Samir-Mohamad Omar and Dirk T.M. Slock are with the Mobile Communications Dept., EURECOM, BP 193, 06904 Sophia Antipolis Cdx, France. email: {omar, slock}@eurecom.fr.

II. MIMO Cyclic Prefix Block TX Systems

Consider a MIMO system with q inputs x_l, $p > q$ outputs y_i per (symbol/sample) period

$$\begin{aligned}\underbrace{\mathbf{y}[m]}_{p\times 1} &= \sum_{l=1}^{q}\sum_{j=0}^{L_l}\underbrace{\mathbf{h}^l[j]}_{p\times 1}\,\underbrace{x_l[m{-}j]}_{1\times 1}\\ &= \sum_{j=0}^{L}\underbrace{\mathbf{h}[j]}_{p\times q}\,\underbrace{\mathbf{x}[m{-}j]}_{q\times 1} = \underbrace{H(q)}_{p\times q}\,\underbrace{\mathbf{x}[m]}_{q\times 1}\end{aligned} \tag{1}$$

where $H(q) = \sum_{j=0}^{L}\mathbf{h}[j]\,q^{-j}$ is the MIMO system transfer function corresponding to the z transform of the impulse response $\mathbf{h}[.]$. Equation (1) mixes time domain and z transform domain notations to obtain a compact representation. In $H(q)$, z is replaced by q (not to be confused with the number of transmit antennas) to emphasize its function as an elementary time advance operator over one sample period. Its inverse corresponds to a delay over one sample period: $q^{-1}\mathbf{x}[n] = \mathbf{x}[n{-}1]$.

Consider a (OFDM or single-carrier) CP block transmission system with N samples per block. The introduction of a cyclic prefix of K samples means that the last K samples of the current block (corresponding to N samples) are repeated before the actual block. If we assume w.l.o.g. that the current block starts at time 0, then samples $\mathbf{x}[N{-}K]\cdots\mathbf{x}[N{-}1]$

are repeated at time instants $-K, \ldots, -1$. This means that the output at sample periods $0,\ldots,N-1$ can be written in matrix form as

$$\begin{bmatrix} \mathbf{y}[0] \\ \cdots \\ \mathbf{y}[N-1] \end{bmatrix} = \mathrm{Y}[0] = \mathrm{H}\,\mathrm{X}[0] + \mathrm{V}[0] \qquad (2)$$

where the matrix H is not only (block) Toeplitz but even (block) circulant: each row is obtained by a cyclic shift to the right of the previous row. Consider now applying an N-point FFT to both sides of (2) at block m:

$$F_{N,p}\mathrm{Y}[m] = F_{N,p}H\,F_{N,q}^{-1} \quad F_{N,q}\mathrm{X}[m] + F_{N,p}\mathrm{V}[m] \qquad (3)$$

or with new notations:

$$\mathrm{U}[m] \quad = \quad \mathcal{H} \qquad \mathrm{A}[m] \quad + \quad \mathrm{W}[m] \qquad (4)$$

where $F_{N,p} = F_N \otimes I_p$ (Kronecker product: $A \otimes B = [a_{ij}B]$), F_N is the N-point $N \times N$ DFT matrix, $\mathcal{H} = \mathrm{diag}\,\{\mathbf{h}_0,\ldots,\mathbf{h}_{N-1}\}$ is a block diagonal matrix with diagonal blocks $\mathbf{h}_k = \sum_{l=0}^{L}\mathbf{h}[l] \;\; e^{-j2\pi\frac{1}{N}kl}$, the $p \times q$ channel transfer function at tone k (frequency $= k/N$ times the sample frequency). In OFDM, the transmitted symbols are in $\mathrm{A}[m]$ and hence are in the frequency domain. The corresponding time domain samples are in $\mathrm{X}[m]$. The OFDM symbol period index is m. In Single-Carrier (SC) CP systems, the transmitted symbols are in $\mathrm{X}[m]$ and hence are in the time domain. The corresponding frequency domain data are in $\mathrm{A}[m]$. The components of V are considered white noise, hence the components of W are white also. At tone (subcarrier) $n \in \{0,\ldots,N-1\}$ we get the following input-output relation

$$\underbrace{\mathbf{u}_n[m]}_{p\times 1} = \underbrace{\mathbf{h}_n}_{p\times q}\;\underbrace{\mathbf{a}_n[m]}_{q\times 1} + \underbrace{\mathbf{w}_n[m]}_{p\times 1} \qquad (5)$$

where the symbol $a_n[m]$ belongs to some finite alphabet (constellation) in the case of OFDM.

III. Some Generalities for CP System Methods

In what follows, we shall see that for methods and performance analysis, we get a cost function or information at each tone for the channel response at that tone, and to get the cost function or information for the temporal channel response, it suffices to sum up the cost functions or informations over the tones after transforming back to the time domain. To be a bit more explicit, let $\overline{\mathbf{h}}_k = \mathrm{vec}\,(\mathbf{h}_k)$ and let $\mathbf{h}$ be the vectorized channel impulse response, i.e. $\mathbf{h} = \mathrm{vec}\,([\mathbf{h}^H[0]\cdots\mathbf{h}^H[L]]^H)$. Then there exists transformation matrices G_k (containing DFT portions) such that

$$\overline{\mathbf{h}}_k = G_k\,\mathbf{h}\,. \qquad (6)$$

Now, if at tone k we have a cost function of the form

$$\overline{\mathbf{h}}_k^H\,Q_k\,\overline{\mathbf{h}}_k \qquad (7)$$

then this induces a cost function for the overall channel impulse response of the form

$$\mathbf{h}^H\left[\sum_{k=0}^{N-1} G_k^H\,Q_k\,G_k\right]\mathbf{h} = \mathbf{h}^H\,Q\,\mathbf{h} \qquad (8)$$

and similarly for Fisher information matrices. So in what follows, we shall concentrate on the cost function for a given tone.

IV. Deterministic Symbols Case

Algorithms that fall under this category are

- subspace fitting (MIMO)
- Subchannel Response Matching (SRM)/ Cross Relation (CR) method (SIMO)
- DML, IQML, DIQML, PQML (SIMO)
- singular prediction parameters (MIMO): $P(z)\,H(z) = \mathbf{h}[0] \;\Rightarrow\; (\mathbf{h}^{\perp}[0]\,P(z))\,H(z) = 0$
- deterministic approach by itself of limited use in MIMO case unless channels of different sources of same length: the case of spatial multiplexing MIMO systems

V. Signal Subspace Fitting

Let us focus in particular on the signal subspace fitting method (noise subspace fitting can be similarly formulated for the SIMO case using the linear noise subspace parameterization in terms of the channel, considered in the next section). For the (spatiotemporally) white noise case (and assuming spatiotemporally white symbols for simplicity), the eigendecomposition of the covariance matrix of a block of signal in the time domain can in fact easily be computed from the eigendecompositions at each tone! Indeed

$$\begin{aligned} &R_{\mathrm{YY}} = \sigma_x^2\,\mathrm{H}\,\mathrm{H}^H + \sigma_v^2\,I_{Np} \\ &\Rightarrow \;\; F_{N,p}R_{\mathrm{YY}}F_{N,p}^{-1} \\ &= \sigma_x^2\,F_{N,p}\mathrm{H}F_{N,q}^{-1}F_{N,q}\mathrm{H}^H F_{N,p}^{-1} + \sigma_v^2\,F_{N,p}F_{N,p}^{-1} \\ &= \sigma_x^2\,\mathcal{H}\,\mathcal{H}^H + \sigma_v^2\,I_{Np} \end{aligned} \qquad (9)$$

where the matrix in square brackets is block diagonal. Hence the eigenvectors in the time domain are the IDFTs of the eigenvectors at each tone, and the eigenvalues are the same in time or frequency domain. This exact relationship no longer holds for the eigenvectors based on sample covariances in time and frequency domain due to the noise (it remains true in the absence of noise). Nevertheless this relationship encourages us to develop subspace fitting problems in the frequency domain, involving eigendecompositions of N $p\times p$ matrices instead of the eigendecomposition of one $Np \times Np$ matrix. Let $\widehat{\mathrm{E}}$ denote a sample average, then the details of the signal subspace fitting method are

- $\mathbf{r}_k = \mathrm{E}\,\mathbf{u}_k[n]\mathbf{u}_k^H[n] = \sigma_a^2\,\mathbf{h}_k\mathbf{h}_k^H + \sigma_{w_k}^2 I_p$
  $= V_{\mathcal{S},k}\Lambda_{\mathcal{S},k}V_{\mathcal{S},k}^H + \sigma_{w_k}^2\,V_{\mathcal{N},k}V_{\mathcal{N},k}^H$
- $\widehat{\mathbf{r}}_k = \widehat{\mathrm{E}}\,\mathbf{u}_k\mathbf{u}_k^H = \widehat{V}_{\mathcal{S},k}\widehat{\Lambda}_{\mathcal{S},k}\widehat{V}_{\mathcal{S},k}^H + \widehat{V}_{\mathcal{N},k}\widehat{\Lambda}_{\mathcal{N},k}\widehat{V}_{\mathcal{N},k}^H$
- signal subspace fitting cost function

$$\min_{\mathbf{h}}\;\sum_{k=0}^{N-1}\|\mathbf{h}_k^H\widehat{V}_{\mathcal{N},k}\|_F^2$$

- cost function per tone $\Rightarrow$ not very costly to introduce optimal weighting.

VI. More Deterministic Approaches

- Deterministic (symbols) ML (DML):

$$\max_{\mathbf{h}} \sum_{k=0}^{N-1} \text{tr}\,\{P_{\mathbf{h}_k}\widehat{\mathbf{r}}_k\} \Leftrightarrow \min_{\mathbf{h}} \sum_{k=0}^{N-1} \text{tr}\,\{P^{\perp}_{\mathbf{h}_k}\widehat{\mathbf{r}}_k\}$$

- IQML: in the SIMO case, we can introduce a linear (in the channel parameters) parameterization of the noise subspace, $\mathbf{h}^{\perp}_k$ so that $P^{\perp}_{\mathbf{h}} = P_{\mathbf{h}^{\perp}} \Rightarrow$

$$\min_{\mathbf{h}} \sum_{k=0}^{N-1} \text{tr}\,\{(\mathbf{h}_k^{\perp H}\mathbf{h}_k^{\perp})^{-1}\mathbf{h}_k^{\perp H}\widehat{\mathbf{r}}_k\mathbf{h}_k^{\perp}\}$$

- Subchannel Response Matching (SRM)/Cross Relation method (CR):

$$\min_{\mathbf{h}} \sum_{k=0}^{N-1} \text{tr}\,\{\mathbf{h}_k^{\perp H}\widehat{\mathbf{r}}_k\mathbf{h}_k^{\perp}\}$$

- Denoised IQML (DIQML):

$$\min_{\mathbf{h}} \sum_{k=0}^{N-1} \text{tr}\,\{(\mathbf{h}_k^{\perp H}\mathbf{h}_k^{\perp})^{-1}\mathbf{h}_k^{\perp H}(\widehat{\mathbf{r}}_k - \widehat{\sigma}^2_{w_k} I_p)\mathbf{h}_k^{\perp}\}$$

  Of course, one can now go further in denoising and replace $\widehat{\mathbf{r}}_k - \widehat{\sigma}^2_{w_k} I_p$ by its pure signal subspace part.
- WSSF/large sample Gaussian (symbols) ML (GML):

$$\max_{\mathbf{h}} \sum_{k=0}^{N-1} \text{tr}\,\left\{P_{\mathbf{h}_k}\widehat{V}_{\mathcal{S},k}\widehat{\widetilde{\Lambda}}^2_{\mathcal{S},k}\widehat{\Lambda}^{-1}_{\mathcal{S},k}\widehat{V}^H_{\mathcal{S},k}\right\}$$

VII. Bayesian Blind with Deterministic Symbols

Assume the Rayleigh fading channel has a prior distribution $\mathbf{h} \sim \mathcal{CN}(0, C^o_{\mathbf{h}})$, then a Bayesian blind criterion can be obtained straightforwardly by augmenting a classical blind criterion as follows

$$\mathbf{h}^H \frac{1}{\sigma_v^2\,\|\mathbf{h}\|^2}\, Q\,\mathbf{h} + \mathbf{h}^H\,(C^o_{\mathbf{h}})^{-1}\,\mathbf{h} \qquad (10)$$

which still remains (pseudo) quadratic in $\mathbf{h}$ ($\|\mathbf{h}\|^2$ refers to a separate estimate of $\|\mathbf{h}\|^2$). (10) would correspond to joint ML for the symbols and MAP for the channel if Q corresponds to one of the DML versions. Under a unit norm constraint, the minimization of (10) leads to $\mathbf{h} = \|\mathbf{h}\|\, V_{min}(\frac{1}{\sigma_v^2\|\mathbf{h}\|^2}\, Q + (C^o_{\mathbf{h}})^{-1})$ (which may need to be solved iteratively if Q depends on $\mathbf{h}$).

VIII. Gaussian Symbols Approaches

- Tone-wise covariance analysis

$$\mathbf{r}_k = \mathrm{E}\,\mathbf{u}_k[n]\mathbf{u}_k^H[n] = \sigma_a^2\,\mathbf{h}_k\mathbf{h}_k^H + \sigma^2_{w_k} I_p$$

  $\Rightarrow$ separate noise variance identifiable at every tone, this corresponds to a circulant noise covariance matrix in the time domain.
- GML: has the same gradient as WCM below.
- Weighted Covariance Matching (WCM):

$$\min_{\mathbf{h},\sigma^2} \sum_{k=0}^{N-1} \text{tr}\,\{\mathbf{r}_k^{-1}(\mathbf{r}_k - \widehat{\mathbf{r}}_k)\mathbf{r}_k^{-1}(\mathbf{r}_k - \widehat{\mathbf{r}}_k)\}$$

- Linear prediction based methods:

$$P(z)\,H(z) = \mathbf{h}[0]$$

  becomes

$$P_k\,\mathbf{h}_k = \mathbf{h}[0]$$

  tonewise.
- For MIMO, the proper exploitation of the Gaussian case is quite advantageous over the deterministic symbols approach.

In Gaussian ML (GML), since both symbols and noise are Gaussian, the received signal is Gaussian with a channel-dependent covariance matrix. This leads to the GML likelihood, in which the symbols are eliminated. Alternatively, it is quite straightforward to add the Bayesian Rayleigh channel prior to the likelihood for the joint estimation of channel and Gaussian symbols, leading to joint MAP estimation of channel and symbols:

$$\frac{1}{\sigma_w^2}\|\mathrm{U} - \mathcal{H}\,\mathrm{A}\|^2 + \frac{1}{\sigma_a^2}\|\mathrm{A}\|^2 + \mathbf{h}^H\,(C^o_{\mathbf{h}})^{-1}\,\mathbf{h} \qquad (11)$$

which is quadratic in $\mathbf{h}$ for fixed A, or in A for fixed $\mathbf{h}$. Knowing that (2) can be written in another form as follows:

$$\mathrm{Y} \;=\; \mathrm{H}\,\mathrm{X} + \mathrm{V} = \mathcal{X}\mathbf{h} + \mathrm{V} \qquad (12)$$

where $\mathcal{X} = \mathcal{X}' \otimes \mathbf{I_p}$ and $\mathcal{X}'$ is a circulant matrix filled with the elements (vectors) of X. Then after some manipulation we get the following solutions:

$$\hat{\mathbf{A}} = \left(\widehat{\mathcal{H}}^H\widehat{\mathcal{H}} + \frac{\sigma_w^2}{\sigma_a^2}I\right)^{-1}\widehat{\mathcal{H}}^H\mathrm{U} \qquad (13)$$

$$\hat{\mathbf{h}} = \left(\widehat{\mathcal{X}}^H\widehat{\mathcal{X}} + \sigma_w^2(C^o_{\mathbf{h}})^{-1}\right)^{-1}\widehat{\mathcal{X}}^H F^H_{N,p}\mathrm{U} \qquad (14)$$

Hence the Alternating GMAPGMAP (AGMAPGMAP) algorithm, by iteratively minimizing for $\mathbf{h}$ or A. Note that, although the (non-Gaussian) joint posterior density of $\mathbf{h}$ and A would be difficult to compute, the joint posterior is not required to obtain their MAP estimate, which is fairly simple to compute (at least when done iteratively as suggested here).

IX. Variational Bayesian Techniques

A recent tutorial on Variational Bayesian (VB) estimation techniques can be found in [6]. VB provides an approximate technique to determine the posterior probability density function (pdf) of the quantities to be estimated. Let θ denote the vector of all quantities to be estimated, including parameters and possibly signals (e.g. the "hidden variables" in EM terminology). And Y denotes the measurements. In many problems, the joint posterior pdf $f(\theta|Y)$ can be complicated to determine. Consider now a partition of θ into K subgroups of quantities that will get estimated per subgroup $\theta = \{\theta_k,\; k = 1,\ldots,K\}$. The idea of VB is to approximate $f(\theta|Y)$ by a product form $g(\theta|Y) = \prod_{k=1}^K g(\theta_k|Y)$ where the $g(\theta_k|Y)$ in general will differ from the true marginal pdfs $f(\theta_k|Y)$. The $g(\theta_k|Y)$ are determined by minimizing the Kullback-Leibler distance between $\prod_{k=1}^{K} g(\theta_k|Y)$ and $f(\theta|Y)$. This leads to the following implicit relations

$$\ln g(\theta_k|Y) = \mathrm{E}_{g(\theta_{\bar{k}}|Y)} \ln f(Y, \theta_k, \theta_{\bar{k}}) \; , k = 1, \ldots, K \quad (15)$$

where $\theta_{\bar{k}}$ is θ minus θ_k, hence $\theta = \{\theta_k, \theta_{\bar{k}}\}$. In practice, (15) needs to be solved iteratively by consecutively sweeping through $k = 1, \ldots, K$, at all times using for $g(\theta_{\bar{k}}|Y)$ the latest version available. This iterative process can be shown to converge montonically. Typically, when $f(Y|\theta)$ and the prior $f(\theta)$ are exponential family pdfs (typically Gaussian), then one can see from (15) that $g(\theta_k|Y)$ will also be of the exponential family. VB techniques have mainly been introduced to deal with hierarchical signal models: signals that contain parameters with a prior that depends itself on parameters with a prior etc. However, VB techniques can be useful in simpler problems also.

Note that Variational Bayesian techniques can also be applied in the presence of deterministic unknowns θ_D. There are two ways to think about deterministic unknowns:

(i) as truly deterministic, with prior $f(\theta_D) = \delta(\theta_D - \theta_D^o)$ where θ_D^o is the unknown true value of θ_D; in other words, $\theta_D \sim \mathcal{N}(\theta_D^o, C_{\theta_D}^o)$ where $C_{\theta_D}^o = 0\, I$.

(ii) as random with no prior information, hence $\theta_D \sim \mathcal{N}(\theta_D^o, C_{\theta_D}^o)$ where $C_{\theta_D}^o = \infty\, I$ and θ_D^o is unimportant (e.g. $\theta_D^o = 0$).

In case (i), VB becomes EM [6], in which case during the iterations the deterministic parameters are simply substituted by their current estimate. Case (ii) is closer to the VB spirit. If $\theta = \{\theta_D, \theta_S\}$ where θ_S are the stochastic parameters, then it suffices to replace $f(Y, \theta)$ in (15) by $f(Y, \theta_S|\theta_D) = f(Y|\theta)\, f(\theta_S)$. In this case also for the deterministic parameters not only their current estimates (posterior means) are accounted for but also their estimation error.

X. Variational Bayesian Blind Channel Estimation

We shall focus on the MIMO OFDM case with Rayleigh fading FIR channel and Gaussian symbol model. In this case $Y = \mathrm{U}$ and $\theta = \{\mathrm{A}, \mathbf{h}\}$. When applying the VB technique, $g(\mathrm{A}|\mathrm{U})$ factors as $g(\mathrm{A}|\mathrm{U}) = \prod_{n=1}^{N} g(\mathbf{a}_n|\mathrm{U})$. We have

$$\begin{aligned}\ln f(\mathbf{u}_n, \mathbf{a}_n, \mathbf{h}_n) = \mathrm{c}^{\mathrm{t}} + \tfrac{1}{\sigma_w^2}\{&-\mathbf{u}_n^H \mathbf{h}_n \mathbf{a}_n \\ &-\mathbf{a}_n^H \mathbf{h}_n^H \mathbf{u}_n + \mathbf{a}_n^H (\mathbf{h}_n^H \mathbf{h}_n + \tfrac{\sigma_w^2}{\sigma_a^2} I_q)\mathbf{a}_n\}\end{aligned} \quad (16)$$

where c^{t} denotes a constant. With (15) we hence get $g(\mathbf{a}_n|\mathrm{U}) \sim \mathcal{CN}(m_{\mathbf{a}_n}, C_{\mathbf{a}_n})$, where

$$\begin{aligned} C_{\mathbf{a}_n} &= \left(\tfrac{1}{\sigma_w^2} \mathrm{trb}\{R_{\overline{\mathbf{h}}_n}^T\} + \tfrac{1}{\sigma_a^2} I_q\right)^{-1} \\ m_{\mathbf{a}_n} &= \tfrac{1}{\sigma_w^2} C_{\mathbf{a}_n} m_{\mathbf{h}_n}^H \mathbf{u}_n \end{aligned} \quad (17)$$

where trb of a block matrix denotes a matrix obtained by taking trace of its blocks (e.g. $\mathbf{h}_n^H \mathbf{h}_n = \mathrm{trb}\{(\overline{\mathbf{h}}_n \overline{\mathbf{h}}_n^H)^T\}$), $\overline{\mathbf{h}}_n = \mathrm{vec}\,(\mathbf{h}_n)$, $m_{\overline{\mathbf{h}}_n} = G_n m_{\mathbf{h}}$, $C_{\overline{\mathbf{h}}_n} = G_n\, C_{\mathbf{h}}\, G_n^H$ and $m_{\mathbf{h}_n} = \mathrm{unvec}\,\{m_{\overline{\mathbf{h}}_n}\}$. In general, $R_{\mathbf{x}} = m_{\mathbf{x}} m_{\mathbf{x}}^H + C_{\mathbf{x}}$. The estimation of the symbols can be seen to correspond to the output of a MMSE linear equalizer in which the channel is not just replaced by its estimate, but the channel estimation error is taken into account also. On the other hand,

$$\begin{aligned}\ln f(\mathrm{U}, \mathrm{A}, \mathbf{h}) = \mathrm{c}^{\mathrm{t}} - \textstyle\sum_n \{ \tfrac{1}{\sigma_w^2} \|\mathbf{u}_n - (\mathbf{a}_n^T \otimes I_p) G_n \mathbf{h}\|^2 \\ \mathbf{h}^H G_n^H (G_n C_{\mathbf{h}}^o G_n^H)^{-1} G_n \mathbf{h}\}\end{aligned} \quad (18)$$

using e.g. $\mathbf{h}_n \mathbf{a}_n = (\mathbf{a}_n^T \otimes I_p) G_n \mathbf{h}$.
Hence with (15), $g(\mathbf{h}|\mathrm{U}) \sim \mathcal{CN}(m_{\mathbf{h}}, C_{\mathbf{h}})$ where

$$\begin{aligned} C_{\mathbf{h}} &= \left(\textstyle\sum_n G_n^H \{\tfrac{1}{\sigma_w^2} \left(R_{\mathbf{a}_n}^* \otimes I_p\right)\} G_n + (C_{\mathbf{h}}^o)^{-1}\right)^{-1} \\ m_{\mathbf{h}} &= C_{\mathbf{h}}\, \tfrac{1}{\sigma_w^2} \textstyle\sum_n G_n^H (m_{\mathbf{a}_n}^* \otimes I_p) \mathbf{u}_n \end{aligned} \quad (19)$$

The estimation of the channel can be seen to correspond to a Bayesian MMSE estimate, using all symbols as pilots, but taking into account that they have estimation error also. Upon convergence, the posterior means $m_{\mathbf{a}_n}$ and $m_{\mathbf{h}}$ are both MAP and MMSE estimates (due to Gaussianity) according to $g(\theta|Y) = g(\mathbf{h}|\mathrm{U}) \prod_n g(\mathbf{a}_n|\mathrm{U})$. However, they are neither MAP nor MMSE estimates according to the true posterior $f(\theta|Y)$. But it is intuitively clear that they should be reasonable approximations of the true MMSE estimates, which contrasts with the true MAP estimates of the AGMAPGMAP algorithm.

Remarks

The case of deterministic symbols can be handled also by just setting $\sigma_a^2 = \infty$ in (16), (17).

The extension to semiblind, with some symbols being pilots, hence being known exactly, is immediate. Their error covariance matrix will remain zero and their mean equals the pilot value.

Extensions of the methods presented can be considered to incorporate the estimation of e.g. $C_{\mathbf{h}}^o$ (or a set of parameters parameterizing $C_{\mathbf{h}}^o$), which would be especially meaningful in the scenario in which multiple realizations of $\mathbf{h}$ with the same $C_{\mathbf{h}}^o$ need to be estimated (e.g. in a sequence of OFDM symbols).

XI. Identifiability Considerations

Consider the joint estimation of the transmitted symbols in time domain X and the collective channel impulse response coefficients $\mathbf{h}$, making up together the parameter vector $\boldsymbol{\theta} = \left[\mathrm{X}^H\; \mathbf{h}^H\right]^H$. Then the Fisher Information matrix (FIM) on the basis of the data Y in (12) alone is

$$J_{\mathrm{Y}} = \frac{1}{\sigma_v^2} \left[\mathrm{H}\, \mathcal{X}\right]^H \left[\mathrm{H}\, \mathcal{X}\right] \; . \quad (20)$$

$\boldsymbol{\theta}$ is unidentifiable since indeed for a $q \times q$ mixing filter $\boldsymbol{\psi}(z)$, we have $H(q)\,\mathbf{x}[m] = (H(q)\boldsymbol{\psi}(q))\,(\boldsymbol{\psi}^{-1}(q)\,\mathbf{x}[m]) = \widetilde{H}(q)\,\widetilde{\mathbf{x}}[m]$. Hence it is impossible to distinguish $H(q)$ from $\widetilde{H}(q)$. If the delay spread of $H(q)$ is known and/or there are border conditions on the transmitted signal, then the frequency-selective mixture $\boldsymbol{\psi}(z)$ becomes a frequency-flat $\boldsymbol{\psi}$. In this case, J_{Y} has a null space which is the column space of $\left[\mathrm{X}^H\; -\mathbf{h}^H\right]^H$. Indeed $[\mathrm{H}\, \mathcal{X}]\, \left[\mathrm{X}^H\; -\mathbf{h}^H\right]^H = 0$. So the multiplicative ambiguity $\boldsymbol{\psi}$ translates into an (additive) singularity in the FIM.

In the case of Gaussian white symbols, the prior information on X translates into an additional FIM

$$J_{\mathrm{X}} = \frac{1}{\sigma_x^2}\left[\begin{array}{cc} I & 0 \\ 0 & 0 \end{array}\right] \qquad (21)$$

so at this point the overall FIM is $J_{\boldsymbol{\theta}} = J_{\mathrm{Y}} + J_{\mathrm{X}}$ which will have become non-singular. This would indicate identifiability. The ambiguity in this case is indeed reduced from an unconstrained $\boldsymbol{\psi}$ to a unitary $\boldsymbol{\psi}$. However, there is still ambiguity and hence unidentifiability. Actually, the proper treatement with Gaussian symbols does not allow presenting the FIM in the compact complex form presented here. In fact, $\boldsymbol{\theta}$ needs to be doubled in size by considering separately its real and imaginary components and the associated FIM needs to be considered, in order to see the FIM nullspace corresponding to a unitary ambiguity matrix.

When now furthermore (or alternatively) a Gaussian prior for the channel $\mathbf{h}$ is considered, then the FIM for $\boldsymbol{\theta}$ gets augmented with

$$J_{\mathbf{h}} = \left[\begin{array}{cc} 0 & 0 \\ 0 & (C^o_{\mathbf{h}})^{-1} \end{array}\right] \qquad (22)$$

which will again render the overall FIM $J_{\boldsymbol{\theta}} = J_{\mathrm{Y}} + J_{\mathbf{h}}(+J_{\mathrm{X}})$ nonsingular. So it would seem that the addition of prior information with an identical non-zero Power Delay Profile (PDP) for each of the antenna pair channels (corresponding to a nonsingular diagonal $C^o_{\mathbf{h}}$) renders $J_{\boldsymbol{\theta}}$ nonsingular and hence leads to (channel) identifiability. However this is not necessarily the case. In the case of Gaussian white symbols, and a unitary ambiguity matrix $\boldsymbol{\psi}$, if $C^o_{\mathbf{h}}$ is such that $(\boldsymbol{\psi} \otimes I_{p(L+1)})^H C^o_{\mathbf{h}} (\boldsymbol{\psi} \otimes I_{p(L+1)}) = C^o_{\mathbf{h}}$ (in which case the channel prior is insensitive to a unitary mixture), then still the Bayesian blind problem remains unidentifiable. The above condition occurs if $C^o_{\mathbf{h}}$ is of the form $C^o_{\mathbf{h}} = I_q \otimes C$ for any square matrix C of size $p(L+1)$. Hence the regularization of the blind channel estimation problem via prior information is a tricky issue due to the multiplicative nature of the ambiguity.

XII. Simulations

We simulate in this section both Bayesian and Variational Bayesian Blind (VBB) channel estimation techniques based on (13), (14) and (17), (19) appearing above. Moreover we simulate also a version of the Variational Bayesian approach where the channel parameters are deterministic unknowns, treated as random with no prior information, so $C^o_{\mathbf{h}} = \infty I$ in (19). We shall refer to this approach as Uniformed VBB (UVBB). In each MonteCarlo simulation we generate a Rayleigh fading channel with exponentially decaying power delay profile (PDP) for the channel between each transmitting and receiving antenna pair as follows: e^{-wn} where $n = 0 : L$ and $w = 2$ normally. Hence, $C^o_{\mathbf{h}}$ is the diagonal matrix $C^o_{\mathbf{h}} = I_q \otimes C \otimes I_p$ where $C = \mathrm{diag}\,\{e^{-wn}, n = 0 : L\}$. As for the symbols, we generate i.i.d. Gaussian symbols (which are hence i.i.d. Gaussian both in time and frequency domain). The performance of the different Bayesian channel estimators is evaluated by means of the Normalized MSE (NMSE) vs. SNR. The per receive antenna SNR is $\mathrm{SNR} = \frac{\sigma_x^2\,\mathrm{tr}\,\{C^o_{\mathbf{h}}\}}{p\,\sigma_v^2}$.

The NMSE is defined as $\frac{avg\,||\mathbf{h}-\widehat{\widehat{\mathbf{h}}}||^2}{avg\,||\mathbf{h}||^2}$ where $\widehat{\widehat{\mathbf{h}}} = \widehat{\mathbf{h}}\boldsymbol{\psi}$ is the channel estimate adjusted for blind channel estimation ambiguities. As we assume the channel length known here, ψ represents an instantaneous mixing matrix of size $(q \times q)$. The mixing matrix $\boldsymbol{\psi}$ can be obtained by minimizing the Frobenius norm of the following matrix error: $\min_{\boldsymbol{\psi}} \; ||\mathbf{h}' - \widehat{\mathbf{h}}'||^2_F$ where $\mathbf{h}' = (\mathbf{h}^H[0] \cdots \mathbf{h}^H[L])^H$ and $\widehat{\mathbf{h}}' = (\widehat{\mathbf{h}}^H[0] \cdots \widehat{\mathbf{h}}^H[L])^H$. For an unconstrained mixture, we get $\boldsymbol{\psi} = (\widehat{\mathbf{h}}'^H \widehat{\mathbf{h}}')^{-1} \widehat{\mathbf{h}}'^H \mathbf{h}' = U \Sigma V^H$ where the last expression represents the SVD of the resulting $\boldsymbol{\psi}$. In the case that $\boldsymbol{\psi}$ gets constrained to be a unitary matrix, the solution is $\boldsymbol{\psi} = UV^H$, see [2].

Both (B and VBB) algorithms are initialized by using (for $m_{\overline{\mathbf{h}}_n}$) noisy perturbations of the true channels. In the first iteration of (17) we use $R_{\overline{\mathbf{h}}_n} = m_{\overline{\mathbf{h}}_n} m^H_{\overline{\mathbf{h}}_n}$, hence $C_{\overline{\mathbf{h}}_n} = 0$. In Figure 1 we can notice how close the performance of both the Variational Bayesian and the Bayesian algorithms is since both fully exploit the prior information that exists about the channel and the symbols. However, we can notice that the UVBB method (with + marker, also called "Deterministic" in the legend) lags behind the normal Variational Bayesian (with * marker) where the prior information is taken into consideration. This is an expected result since the more information we exploit the better performance we get. However, at higher SNR the performance of the deterministic blind algorithm converges to that of the Bayesian blind algorithms. Also this result is expected since at very high SNR the contribution of prior information becomes negligible.

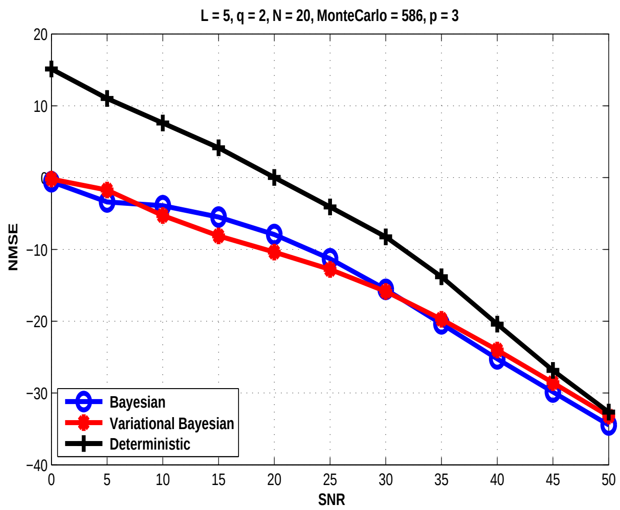

Fig. 1. NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 20$, unitary $\boldsymbol{\psi}$.

Whereas Fig. 1 uses a unitary $\boldsymbol{\psi}$, Fig. 2 uses an unconstrained $\boldsymbol{\psi}$, which leads to reduced NMSE since more prior information is exploited. At least, Fig. 1 shows that the exploitation of the white character of the symbols as we do here leads to a reduced unidentifiability of $\boldsymbol{\psi}$ to just a unitary mixing matrix.

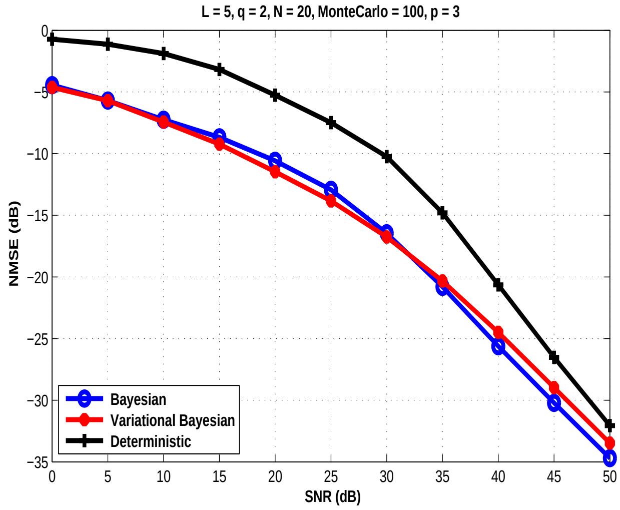

Fig. 2. NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 20$, unconstrained ψ.

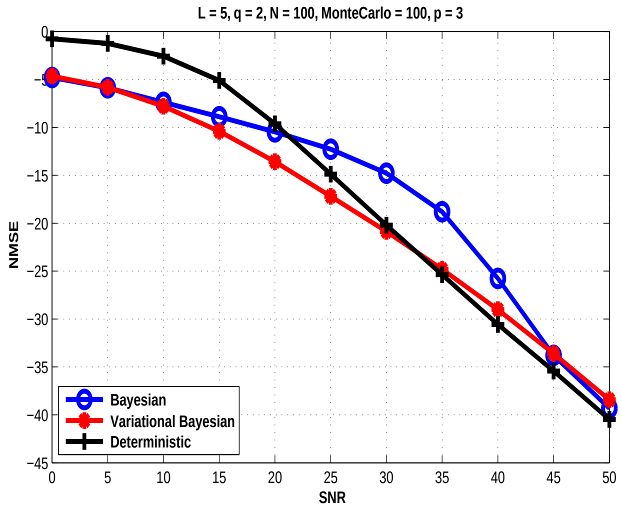

Fig. 3. NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 100$, unconstrained ψ.

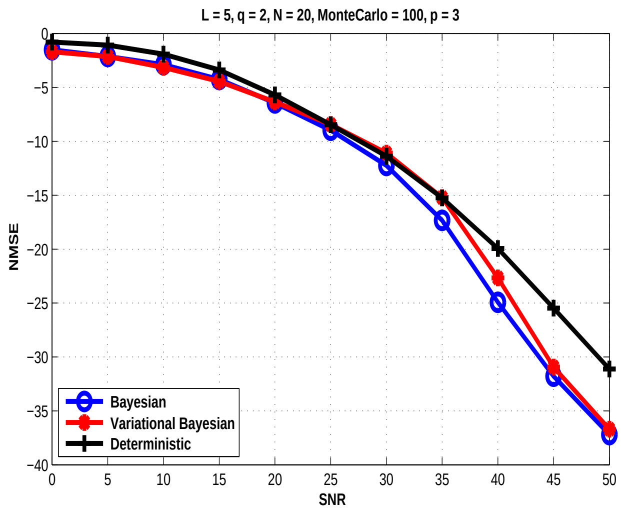

Fig. 4. NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 20$, unconstrained ψ, $w = 0.5$ in PDP.

In Fig. 3 the OFDM block length N gets increased from 20 (as in the previous two figures) to 100. The result is that the prior information introduced by a Bayesian approach only helps at low SNR, as could be expected. The other noticeable effect is that the Variational approach outperforms the non-Variational version over a wide SNR range.

In Fig. 4 finally, an exponential PDP with much shorter time constant ($w = 0.5$) is used, as compared to $w = 2$ in the previous three figures. The result is that the prior information only helps at very high SNR.

REFERENCES

[1] E. de Carvalho and D. Slock, "Semi–Blind Methods for FIR Multichannel Estimation," in *Signal Processing Advances in Wireless & Mobile Communications*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Prentice Hall, 2001.

[2] D. Slock and A. Medles, "Blind and Semiblind MIMO Channel Estimation," in *Space-Time Wireless Systems, From Array Processing to MIMO Communications*, H. Bölcskei, C. P. D.G̃esbert, and A.-J. van der Veen, Eds. Cambridge University Press, 2006.

[3] D. Slock, "Blind FIR Channel Estimation in Multichannel Cyclic Prefix Systems," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, July 2004.

[4] S. Omar and D. Slock, "Structured Spatio-Temporal Sample Covariance Matrix Enhancement with Application to Blind Channel Estimation in Cyclic Prefix Systems," in *Proc. IEEE Int'l Workshop on Signal Processing Advances in Wireless Comm's (SPAWC)*, Perugia, Italy, June 2009.

[5] D. Slock, "Bayesian Blind and Semiblind Channel Estimation," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, July 2004.

[6] D. Tzikas, A. Likas, and N. Galatsanos, "The Variational Approximation for Bayesian Inference, Life After the EM Algorithm," *IEEE Signal Processing Magazine*, Nov. 2008.