CRAMER-RAO BOUNDS AND METHODS FOR KNOWLEDGE BASED ESTIMATION OF MULTIPLE FIR CHANNELS

Jaouhar Ayadi *Dirk T.M. Slock*
Institut EURECOM
B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
{ayadi,slock}@eurecom.fr

ABSTRACT

This contribution elaborates on the concept of blind identification of multiple FIR channels with prior knowledge (WPK). The prior knowledge considered here corresponds to the transmitter (TX) (pulse shaping) and/or receiver (RX) filters present in digital communication systems. Exploitation of this prior knowledge allows the estimation to concentrate on the impulse response of the actual channel part itself. Hence this estimation can be done more accurately. Since the prior information is expressed in terms of the channel impulse response, we review a number of blind channel estimation methods that are parameterized by the channel and consider their extension to incorporate the prior knowledge. These methods include Subchannel Response Matching (SRM), subspace fitting and Maximum Likelihood (ML) techniques. We also discuss performance limits in the form of Cramer-Rao bounds (CRBs). Both the methods and the CRBs are discussed in a deterministic and a Gaussian context for the unknown transmitted symbols. Simulation results indicate that the exploitation of the prior knowledge can lead to significant improvements, that one particular SRM WPK method often outperforms another one, and that ML methods can still further improve performance.

1. INTRODUCTION

The goal of blind identification is to identify the unknown channel using the received signal only. Most of the work on blind identification considers the entire channel which includes the shaping filter, the actual propagation channel and the receiver filter. However, usually the only unknown quantity is the multipath, the 'propagation channel'. Blind channel identification exploiting the prior knowledge of TX/RX filters has been introduced in [1] and further explored in [2]. These blind techniques exploit a multichannel formulation corresponding to a Single Input Multiple Output (SIMO) vector channel. The channel is assumed to have a finite delay spread NT. The multiple FIR channels can be obtained by oversampling a single received signal, but can also be obtained from multiple received signals from an array of antennas (in the context of mobile digital communications [3],[4]) or from a combination of both. For m channels the discrete-time input-output relationship can be written as:

$$\mathrm{y}(k)=\sum_{i=0}^{N-1}\mathrm{h}(i)a(k-i)+\mathrm{v}(k)=\mathrm{H}\,A_N(k)+\mathrm{v}(k) \qquad (1)$$

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

$$\mathrm{Y}_M(k) \;=\; \mathcal{T}_M(\mathrm{H}\,)\,A_{M+N-1}(k)+\mathrm{V}_M(k) \qquad (2)$$

where $\mathrm{Y}_M(k) = [\mathrm{y}^H(k-M+1)\cdots\mathrm{y}^H(k)]^H$ and similarly for $\mathrm{V}_M(k)$. $\mathcal{T}_M(\mathrm{X}\,)$ is a block Toepliz matrix with M block rows and $[\mathrm{X}\ \ 0_{p\times(M-1)q}]$ as first block row, X being considered as a block row vector with $p\times q$ blocks. We shall simplify the notation in (2) with $k=M-1$ to

$$\mathrm{Y} \;=\; \mathcal{T}(\mathrm{H}\,)\,A+\mathrm{V}\;. \qquad (3)$$

We assume that $mM > M+N-1$ in which case the channel convolution matrix $\mathcal{T}(\mathrm{H}\,)$ has more rows than columns. If the $\mathrm{H}_i(z)$, $i=1,\ldots,m$ have no zeros in common, then $\mathcal{T}(\mathrm{H}\,)$ has full column rank (which we will henceforth assume). For obvious reasons, the column space of $\mathcal{T}(\mathrm{H}\,)$ is called the signal subspace and its orthogonal complement the noise subspace. The signal subspace is parameterized linearly by h.

2. BLIND CHANNEL ESTIMATION

The channel can either be parameterized by its impulse response h or by the noisefree multivariate prediction error filter $\mathbf{P}(z)$ and $\mathrm{h}(0)$ which satisfy $\mathbf{P}(z)\mathbf{H}(z) = \mathrm{h}(0)$ [4]. However, it is not clear how to express prior information on the TX/RX filters in terms of the prediction filter. Hence we stick here to blind methods that are parameterized in terms of h. Two approaches exist, depending on whether the symbols are considered deterministic or Gaussian unknowns.

2.1. Methods for Deterministic Symbols

2.1.1. Subchannel Response Matching (SRM)

In order to explain the SRM technique [5], consider first the case of two channels: $m=2$. One can observe that for noise-free signals, we have $\mathrm{H}_2(z)y_1(k) - \mathrm{H}_1(z)y_2(k) = 0$, which can be written in matrix form as $[\mathrm{H}_2(z)\ -\mathrm{H}_1(z)]\,\mathrm{y}(k) = \mathbf{H}^{\perp\dagger}(z)\mathrm{y}(k) = \mathbf{H}^{\perp\dagger}(z)\mathbf{H}(z)a(k) = 0$ where e.g. $\mathbf{H}^{\dagger}(z) = \mathbf{H}^H(1/z^*)$. Stacking these zeros into a vector for the signal $\{\mathrm{y}(k)\}_{k=0\cdots M-1}$, we get an expression of the form $\mathcal{Y}h = 0$ for some structured matrix $\mathcal{Y}$. Under the constraint $\|\mathrm{h}\|_2 = 1$, we find $\mathrm{h} = V_{min}(\mathcal{Y}^H\mathcal{Y})$ where $V_{min}(A)$ denotes the eigenvector corresponding to the minimum eigenvalue of A. For $m>2$, blocking equalizers $\mathbf{H}^{\perp\dagger}(z)$ can be constructed by considering the (sub)channels in pairs. The choice of $\mathbf{H}^{\perp\dagger}(z)$ is far from unique. To begin with, the number of pairs to be considered, which is the number of rows in $\mathbf{H}^{\perp\dagger}(z)$, is not unique. The minimum number is $m{-}1$ whereas the maximum number is $\frac{m(m-1)}{2}$, with corresponding $\mathbf{H}^{\perp}_{min}(z)$ and $\mathbf{H}^{\perp}_{max}(z)$. The choice of $\mathbf{H}^{\perp}_{min}(z)$ is not unique. The convolution $\mathbf{H}^{\perp}(z)\mathrm{y}(k)$ involving $\{\mathrm{y}(k)\}_{k=0\cdots M-1}$ can be written in matrix form as $\mathcal{T}(\mathrm{h}^{\perp})\mathrm{Y}$. Since for the noise-free signal we get $\mathcal{T}(\mathrm{h}^{\perp})\mathrm{Y}=0$, the SRM method minimizes the criterion $\left\|\mathcal{T}(\mathrm{h}^{\perp})\mathrm{Y}\right\|_2^2$. By the law of large numbers, asymptotically this criterion can be replaced by its expected value, which can be rewritten in the frequency domain as

$$\begin{array}{l}\mathcal{J}=\frac{1}{2\pi j}\oint \mathrm{tr}\,\{\hat{\mathbf{H}}^{\perp\dagger}\mathrm{S_{yy}}\hat{\mathbf{H}}^{\perp}\}\frac{dz}{z}=\\ \frac{\sigma_a^2}{2\pi j}\oint \mathbf{H}^{\dagger}\hat{\mathbf{H}}^{\perp}\hat{\mathbf{H}}^{\perp\dagger}\mathbf{H}\frac{dz}{z}+\frac{\sigma_v^2}{2\pi j}\oint \mathrm{tr}\,\{\hat{\mathbf{H}}^{\perp\dagger}\hat{\mathbf{H}}^{\perp}\}\frac{dz}{z}\end{array} \qquad (4)$$

We shall call $\mathbf{H}^{\perp}_{bal}(z)$ balanced if $\mathrm{tr}\,\{\mathbf{H}^{\perp\dagger}(z)\mathbf{H}^{\perp}(z)\}=\alpha\mathbf{H}^{\dagger}(z)\mathbf{H}(z)$ for some real scalar α. In that case

$$\min_{\|\hat{\mathrm{h}}\|=1}\ \mathcal{J}=\alpha\sigma_v^2+\frac{\sigma_a^2}{2\pi j}\min_{\|\hat{\mathrm{h}}\|=1}\oint \mathbf{H}^{\dagger}\hat{\mathbf{H}}^{\perp}\hat{\mathbf{H}}^{\perp\dagger}\mathbf{H}\frac{dz}{z} \qquad (5)$$

which leads to the correct value $\hat{\mathrm{h}}=\mathrm{h}$ (and hence an unbiased estimate!) apart from a scale factor (and assuming the channel order is chosen correctly). The minimum number of rows in $\mathbf{H}^{\perp\dagger}_{bal}(z)$ is m in which case $\alpha=2$. The choice for such a $\mathbf{H}^{\perp}_{bal,min}(z)$ is not unique. Note that $\mathbf{H}^{\perp}_{max}(z)$ is balanced with $\alpha=m{-}1$. In the literature, the SRM method (which has been reinvented several times) is always proposed using $\mathbf{H}^{\perp}_{max}(z)$. We get for instance

$$\mathbf{H}^{\perp\dagger}_{min}(z)=\left[\begin{array}{cccc}-\mathrm{H}_2(z) & \mathrm{H}_1(z) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ -\mathrm{H}_m(z) & 0 & \cdots & \mathrm{H}_1(z)\end{array}\right] \qquad (6)$$

$$\mathbf{H}^{\perp\dagger}_{bal,min}(z)=\left[\begin{array}{ccccc}-\mathrm{H}_2(z) & \mathrm{H}_1(z) & 0 & \cdots & 0\\ 0 & -\mathrm{H}_3(z) & \mathrm{H}_2(z) & \cdots & \vdots\\ \vdots & & \ddots & \ddots & 0\\ \mathrm{H}_1(z) & 0 & \cdots & 0 & -\mathrm{H}_m(z)\end{array}\right] \qquad (7)$$

Continuing with this $\mathbf{H}^{\perp\dagger}_{bal}(z)$, its i^{th} row can be written as

$$\mathbf{H}^{\perp\dagger}_{bal,i}(z)=\mathbf{H}^{T}(z)\mathcal{P}_i,\ \mathcal{P}_{i+1}=\mathcal{C}\mathcal{P}_i\mathcal{C}^H,$$

$$\mathcal{P}_1=\left[\begin{array}{cccc}0 & 1 & 0 & \cdots\\ -1 & 0 & \cdots & \\ 0 & \vdots & \ddots & \\ \vdots & & & \end{array}\right],\ \mathcal{C}=\left[\begin{array}{cccc}0 & \cdots & 0 & 1\\ 1 & 0 & \cdots & 0\\ 0 & \ddots & & \vdots\\ \vdots & 0 & 1 & 0\end{array}\right]. \qquad (8)$$

For this $H^{\perp\dagger}_{bal}(z)$, the SRM criterion $\left\|\mathcal{T}(\mathrm{h}^{\perp})\mathrm{Y}\right\|_2^2$ can be written as the minimization w.r.t. h of

$$\begin{array}{l}\mathrm{tr}\,\{\mathcal{T}(\mathrm{h}^{\perp})\mathrm{Y}\mathrm{Y}^H\mathcal{T}^H(\mathrm{h}^{\perp})\}\\ =\ \mathrm{tr}\,\{\mathrm{h}^{\perp}\left(\sum_{k=N-1}^{M-1}\mathrm{Y}_N(k)\mathrm{Y}_N^H(k)\right)\mathrm{h}^{\perp H}\}\\ =(M{-}N{+}1)\ \mathrm{tr}\,\{\mathrm{h}^{\perp}\widehat{R}_{YY}\mathrm{h}^{\perp H}\}\end{array} \qquad (9)$$

where the i^{th} row of $\mathrm{h}^{\perp}$ is $\mathrm{h}^{\perp}_i=\mathrm{h}^T\mathcal{S}_i$, $\mathcal{S}_i=I_N\otimes\mathcal{P}_i$. Hence the SRM criterion in (9) becomes

$$\min_{\mathrm{h}}\mathrm{h}^H A\,\mathrm{h}\ ,\ \text{where } A=\sum_{i=1}^{m}\mathcal{S}_i\widehat{R}^*_{YY}\mathcal{S}_i^H\ . \qquad (10)$$

It is expected that the use of a $\mathbf{H}^{\perp\dagger}_{bal}(z)$ with more rows leads to improved performance.

2.1.2. Signal Subspace Fitting (SSF)

The covariance matrix of the received signal can be decomposed into signal and noise subspace contributions:

$$\begin{array}{rcl}R_{YY}=\mathrm{E}\mathrm{Y}\mathrm{Y}^H & = & \sum_{i=1}^{M+N-1}\lambda_i V_i V_i^H\ +\sum_{i=M+N}^{mM}\lambda_i V_i V_i^H\\ & = & \mathcal{V}_S\Lambda_S\mathcal{V}_S^H\ +\ \mathcal{V}_N\Lambda_N\mathcal{V}_N^H\end{array} \qquad (11)$$

Since both $\mathcal{T}(\mathrm{H})$ and $\mathcal{V}_S$ should span the signal subspace, we can introduce the following signal subspace fitting problem:

$$\min_{\mathrm{h},T}\|\mathcal{T}(\mathrm{H})-\mathcal{V}_S T\|_F \qquad (12)$$

where $\|X\|_F^2=\mathrm{tr}\,\{X^H X\}$. After optimization w.r.t. T, we obtain [4]

$$\min_{\|\mathrm{h}\|_2=1}\ \mathrm{tr}\,\{\mathcal{T}^H(\mathrm{H})P^{\perp}_{\mathcal{V}_S}\mathcal{T}(\mathrm{H})\}=\min_{\|\mathrm{h}\|_2=1}\mathrm{h}^H A\,\mathrm{h} \qquad (13)$$

where $P^{\perp}_X=I-P_X=I-X(X^HX)^{+}X^H$ and $^{+}$ denotes Moore-Penrose pseudo-inverse. A can be determined from $P^{\perp}_{\mathcal{V}_S}=P_{\mathcal{V}_N}$. The solution is again $\mathrm{h}=V_{min}(A)$.

2.1.3. Noise Subspace Fitting (SSF)

Similarly, $\mathcal{V}_N$ spans the noise subspace and $\mathcal{T}^H(\mathrm{h}^{\perp})$ spans most of it. So we can introduce the following signal subspace fitting problem:

$$\min_{\mathrm{h},T}\left\|\mathcal{T}^H(\mathrm{h}^{\perp})-\mathcal{V}_N T\right\|_F\ . \qquad (14)$$

After optimization w.r.t. T, we obtain $\min_{\|\mathrm{h}\|_2=1}$ of

$$\mathrm{tr}\,\{\mathcal{T}(\mathrm{h}^{\perp})P^{\perp}_{\mathcal{V}_N}\mathcal{T}^H(\mathrm{h}^{\perp})\}=\ \mathrm{tr}\,\{\mathrm{h}^{\perp}B\mathrm{h}^{\perp H}\}=\mathrm{h}^H A\,\mathrm{h} \qquad (15)$$

where B can be determined from $P^{\perp}_{\mathcal{V}_N}=P_{\mathcal{V}_S}$ and $A=\sum_{i=1}^{m}\mathcal{S}_i B\mathcal{S}_i^H$.

2.1.4. Deterministic ML (DML)

With the Gaussian noise assumption, maximizing the likelihood reduces to $\min_{A,\mathrm{h}}\|\mathrm{Y}-\mathcal{T}(\mathrm{H})A\|^2$. After optimization w.r.t. A we get $\min_{\mathrm{h}}\mathrm{Y}^H P^{\perp}_{\mathcal{T}(\mathrm{H})}\mathrm{Y}$. Now we have approximately $P^{\perp}_{\mathcal{T}(\mathrm{H})}\approx P_{\mathcal{T}^H(\mathrm{h}^{\perp})}$ where the approximation error disappears asymptotically. Hence we get $\min_{\|\mathrm{h}\|=1}$ of

$$\mathrm{Y}^H P_{\mathcal{T}^H(\mathrm{h}^{\perp})}\mathrm{Y}=\mathrm{h}^H\left(\mathcal{Y}^H[\mathcal{T}(\mathrm{h}^{\perp})\mathcal{T}^H(\mathrm{h}^{\perp})]^{+}\mathcal{Y}\right)\mathrm{h}=\mathrm{h}^H A\mathrm{h} \qquad (16)$$

where $\mathcal{T}(\mathrm{h}^{\perp})\mathrm{Y}=\mathcal{Y}\mathrm{h}$ for some $\mathcal{Y}$. The iterative quadratic (IQ) strategy considers the quadratic "numerator" of the criterion, and for $\mathrm{h}^{\perp}$ in the "denominator" the value from the previous iteration is used. If the initialization is consistent, then only one iteration leads to an ABC estimate. Note that interpreting the SRM method as a least-squares (LS) problem, the DML criterion is the corresponding optimally weighted LS problem: the noise in $\mathcal{T}(\mathrm{h}^{\perp})\mathrm{Y}$ is $\mathcal{T}(\mathrm{h}^{\perp})\mathrm{V}$ with covariance matrix $\sigma_v^2\mathcal{T}(\mathrm{h}^{\perp})\mathcal{T}^H(\mathrm{h}^{\perp})$. Asymptotically, any choice for $\mathbf{H}^{\perp}(z)$ leads to the same performance since $P_{\mathbf{H}^{\perp\dagger}(z)} = P_{\mathbf{H}^{\perp\dagger}_{min}(z)}$.

2.2. Methods for Gaussian Symbols

Whereas with deterministic symbols the channel can only be determined blindly up to an arbitrary complex scale factor, in the Gaussian symbols case also the norm of the channel gets estimated. The main approach in the Gaussian case is ML (GML). In this case $\mathrm{Y} \sim \mathcal{N}(0, R_{YY})$ with $R_{YY} = \sigma_a^2\mathcal{T}(\mathrm{H})\mathcal{T}^H(\mathrm{H}) + \sigma_v^2 I$. The negative log likelihood to be minimized is

$$\mathcal{L}(\mathrm{h}) = c^t + \ln\det R_{YY} + \mathrm{Y}^H R_{YY}^{-1}\mathrm{Y}\,. \tag{17}$$

Standard optimization techniques such as the Gauss-Newton or scoring methods can be applied.

3. TX/RX FILTER KNOWLEDGE

Consider a certain oversampling factor m and let the oversampled transfer function $\mathrm{H}(z) = \mathrm{C}(z)\mathrm{G}(z)$ of the overall channel be the cascade of the actual channel $\mathrm{C}(z)$ and the combined TX/RX filter $\mathrm{G}(z)$. Each of these transfer functions can be decomposed into its polyphase components at the symbol rate, e.g. $\mathrm{H}(z) = \sum_{i=0}^{m-1} z^{-i}\mathrm{H}_i(z^m)$. These components can also be represented in the SIMO form, $\mathbf{G}(z) = [\mathrm{G}_1^H(z)\cdots\mathrm{G}_m^H(z)]^H = \sum_{k=0}^{K-1}\mathrm{g}(k)z^{-k}$ and $\mathbf{C}(z) = [\mathrm{C}_1^H(z)\cdots\mathrm{C}_m^H(z)]^H = \sum_{k=0}^{L-1}\mathrm{c}(k)z^{-k}$ with $K+L-1 = N$. The relations between the polyphase components can be obtained from

$$\sum_{i=0}^{m-1} z^{-i}\mathrm{H}_i(z^m) = \left(\sum_{k=0}^{m-1} z^{-k}\mathrm{G}_k(z^m)\right)\left(\sum_{l=0}^{m-1} z^{-l}\mathrm{C}_l(z^m)\right) \tag{18}$$

In particular for $m = 2$ we get

$$\begin{aligned}\begin{bmatrix}\mathrm{H}_0(z)\\ \mathrm{H}_1(z)\end{bmatrix} &= \begin{bmatrix}\mathrm{G}_0(z) & z^{-1}\mathrm{G}_1(z)\\ \mathrm{G}_1(z) & \mathrm{G}_0(z)\end{bmatrix}\begin{bmatrix}\mathrm{C}_0(z)\\ \mathrm{C}_1(z)\end{bmatrix}\\ &= \begin{bmatrix}\mathrm{C}_0(z) & z^{-1}\mathrm{C}_1(z)\\ \mathrm{C}_1(z) & \mathrm{C}_0(z)\end{bmatrix}\begin{bmatrix}\mathrm{G}_0(z)\\ \mathrm{G}_1(z)\end{bmatrix}\end{aligned} \tag{19}$$

or $\mathbf{H}(z) = \underline{\mathbf{G}}(z)\mathbf{C}(z) = \underline{\mathbf{C}}(z)\mathbf{G}(z)$. In the time domain, we get

$$\mathcal{T}_M(\mathrm{H}) = \mathcal{T}_M(\underline{\mathrm{G}})\mathcal{T}_{M+K-1}(\mathrm{C}) \tag{20}$$

where C is similar to H and

$$\underline{\mathrm{G}} = \left[\underline{\mathrm{g}}(K{-}1)\cdots\underline{\mathrm{g}}(0)\right],\; \underline{\mathrm{g}}(k) = \begin{bmatrix} g_0(k) & g_1(k-1)\\ g_1(k) & g_0(k)\end{bmatrix} \tag{21}$$

and we assume $g_1(K{-}1) = 0$. The relation between h and c is $\mathrm{h} = \mathcal{T}_L^T(\underline{\mathrm{G}}^t)\mathrm{c}$ where ^{t} denotes transposition of the blocks: $\underline{\mathrm{G}}^t = \left[\underline{\mathrm{g}}^T(K{-}1)\cdots\underline{\mathrm{g}}^T(0)\right]$.

In CDMA applications, large excess bandwidth exists and hence large oversampling factors can be used. In TDMA applications, only a small excess bandwidth is available and the oversampling factor will usually be limited to $m = 2$. However, more channels can be obtained by e.g. exploiting multiple antenna signals. In that case we get $\mathbf{H}_i(z) = \underline{\mathbf{G}}_i(z)\mathbf{C}_i(z)$ for every antenna signal $i = 1\ldots q$ (where $\underline{\mathbf{G}}_i(z)$ may be independent of i) and $\mathbf{H}(z) = [\mathbf{H}_1^H(z)\cdots\mathbf{H}_q^H(z)]^H = \mathrm{blockdiag}\{\underline{\mathbf{G}}_1(z)\cdots\underline{\mathbf{G}}_q(z)\}\mathbf{C}(z)$ where now $\mathbf{H}(z)$ and $\mathbf{C}(z)$ regroup mq channels.

4. BLIND METHODS WPK

Prior TX/RX filter knowledge gets exploited by expressing $\mathrm{h} = \mathcal{T}_L^T(\underline{\mathrm{G}}^t)\mathrm{c}$ and searching for c. Since all four deterministic methods discussed above are of the form $\min_{\|\mathrm{h}\|=1}\mathrm{h}^H A\mathrm{h}$, we get $\min_{\mathrm{c}}\mathrm{c}^H\mathcal{T}^*(\underline{\mathrm{G}}^t)A\mathcal{T}^T(\underline{\mathrm{G}}^t)\mathrm{c}$. In all methods except SRM, we can use $\|\mathrm{c}\| = 1$ as non-triviality constraint. For SRM however, the noise contribution has to be taken into account properly in order to avoid bias. One solution as proposed independently in [6] is to translate $\|\mathrm{h}\|^2 = 1$ into the constraint $\mathrm{c}^H\mathcal{T}^*(\underline{\mathrm{G}}^t)\mathcal{T}^T(\underline{\mathrm{G}}^t)\mathrm{c} = 1$ which leads to a generalized eigenvalue problem that can alternatively be transformed into a regular V_{min} problem. This solution consists again in constraining the filter in such a way that it has no influence on the noise component. A second solution consists of (asymptotically) removing the noise contribution altogether. For a balanced $\mathbf{H}^{\perp}$, the contribution of the noise to $\mathrm{E}A$ is a multiple of identity, whereas the contribution of the signal is singular. Hence, the noise contribution can be removed by considering $\min_{\|\mathrm{c}\|=1}\mathrm{c}^H\mathcal{T}^*(\underline{\mathrm{G}}^t)(A - \lambda_{min}(A)\,I)\mathcal{T}^T(\underline{\mathrm{G}}^t)\mathrm{c}$. For GML, one needs to introduce (20) in (17).

5. CRAMER-RAO BOUNDS WPK

In [4], the deterministic Cramer-Rao Bound (CRB) for the estimation of h from (2) was derived. Its extension to the estimation of c can be shown to be

$$CRB_{\mathrm{c}} = \sigma_v^2\left[\mathcal{A}_{M+K-1,L}^H\mathcal{T}_M^H(\mathrm{G})P_{\mathcal{T}(\mathrm{H})}^{\perp}\mathcal{T}_M(\mathrm{G})\mathcal{A}_{M+K-1,L}\right]^+ \tag{22}$$

where $\mathcal{A}_{M+K-1,L} = A_{M+K-1,L}\otimes I_m$ is such that $\mathcal{A}_{M+K-1,L}\mathrm{c} = \mathcal{T}_{M+K-1}(\mathrm{C})A$. The CRB_{c} for c can be transformed into a bound for the unbiased estimation error $\widetilde{\mathrm{h}} = \mathrm{h} - \widehat{\mathrm{h}}$ WPK on the overall channel $\mathrm{h} = \mathcal{T}_L^T(\underline{\mathrm{G}}^t)\mathrm{c}$:

$$C_{\widetilde{\mathrm{h}}} = \mathrm{E}\widetilde{\mathrm{h}}\widetilde{\mathrm{h}}^H \geq \mathcal{T}_L^T(\underline{\mathrm{G}}^t)CRB_{\mathrm{c}}\mathcal{T}_L^*(\underline{\mathrm{G}}^t)\,. \tag{23}$$

For the Gaussian case, things are a bit more intricate. Let the Fisher information matrices (FIM) $J_{\varphi\psi}$ be defined as:

$$J_{\varphi\psi} = E_{Y/\mathrm{c}}\left(\frac{\partial\ln f(Y/\mathrm{c})}{\partial\varphi^*}\right)\left(\frac{\partial\ln f(Y/\mathrm{c})}{\partial\psi^*}\right)^H \tag{24}$$

and we will consider J_{cc} and J_{cc^*}. In the deterministic case, $J_{\mathrm{cc}^*} = 0$. In that case, J_{cc} can be considered as a complex FIM, and $C_{\widetilde{\mathrm{c}}} \geq J_{\mathrm{cc}}^{-1}$, the complex CRB. If $J_{\mathrm{cc}^*} \neq 0$ as in the Gaussian case, J_{cc}^{-1} is also a bound on $C_{\widetilde{\mathrm{c}}}$, but not as tight as the actual CRB which we obtain by considering $\mathrm{c}_R = \left[\mathrm{Re}(\mathrm{c})^T\;\; \mathrm{Im}(\mathrm{c})^T\right]^T$, the associated real parameters. We get:

$$J_R(\mathrm{c}_R) = 2\begin{bmatrix}Re(J_{\mathrm{cc}}) & -Im(J_{\mathrm{cc}})\\ Im(J_{\mathrm{cc}}) & Re(J_{\mathrm{cc}})\end{bmatrix} + 2\begin{bmatrix}Re(J_{\mathrm{cc}^*}) & -Im(J_{\mathrm{cc}^*})\\ Im(J_{\mathrm{cc}^*}) & Re(J_{\mathrm{cc}^*})\end{bmatrix} \tag{25}$$

and

$$J_{\mathrm{c}_i\mathrm{c}_j} = \mathrm{tr}\left\{R_{YY}^{-1}\left(\frac{\partial R_{YY}}{\partial\mathrm{c}_i^*}\right)R_{YY}^{-1}\left(\frac{\partial R_{YY}}{\partial\mathrm{c}_j^*}\right)^H\right\}, \tag{26}$$

$$J_{\mathrm{c}_i\mathrm{c}_j^*} = \mathrm{tr}\left\{R_{YY}^{-1}\left(\frac{\partial R_{YY}}{\partial\mathrm{c}_i^*}\right)R_{YY}^{-1}\left(\frac{\partial R_{YY}}{\partial\mathrm{c}_j^*}\right)\right\}, \tag{27}$$

$$\frac{\partial R_{YY}}{\partial\mathrm{c}_i^*} = \sigma_a^2\mathcal{T}_M(\mathrm{H})\,\mathcal{T}_{M+K-1}^H\left(\frac{\partial\mathrm{C}}{\partial\mathrm{c}_i}\right)\mathcal{T}_M^H(\underline{\mathrm{G}}) \tag{28}$$

These results were derived independently in [6], where some examples show that the Gaussian assumption improves the estimation quality considerably in certain cases.

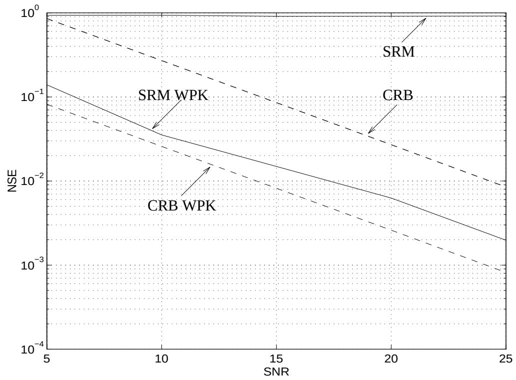

Figure 1. Performance of SRM and SRM WPK

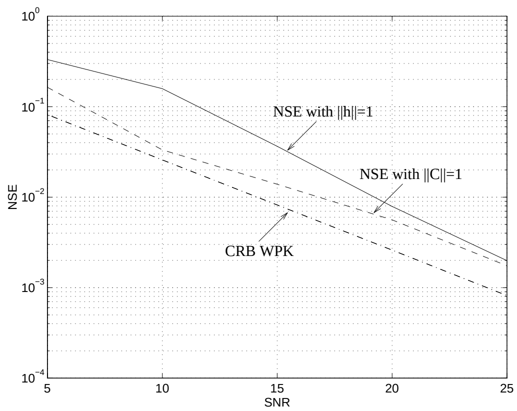

Figure 2. Comparison of SRM WPK with $\|\mathrm{c}\| = 1$ and $\|\mathrm{h}\| = 1$

6. SIMULATION RESULTS

In Figure.1, the performance of SRM and SRM WPK are compared to the corresponding deterministic CRB and CRB WPK. The data frame length is $M = 162$, oversampling factor $m = 2$ and the symbols are i.i.d. BPSK. The overall channel is the convolution of a raised cosine pulse limited to $13T$ with rolloff factor $\alpha = 0.9$, and a two ray multipath channel $c(t) = \delta(t) - 0.82\,\delta(t-T)$. The performance measure is the Normalized MSE (NMSE) which is averaged over 100 Monte-Carlo runs: NMSE$= \frac{1}{100}\sum_{i=1}^{100} \|\widehat{\mathrm{h}}^{(i)} - \mathrm{h}\|^2/\|\mathrm{h}\|^2$. The CRBs are normalized and computed as $\mathrm{tr}\ \{CRB_{\mathrm{h}}\}/\|\mathrm{h}\|^2$ and $\mathrm{tr}\ \{\mathcal{T}_L^T(\underline{\mathrm{G}}^t) CRB_{\mathrm{C}} \mathcal{T}_L^*(\underline{\mathrm{G}}^t)\}/\|\mathrm{h}\|^2$. Our simulation results show that in terms of CRB, the approach WPK outperforms the one without this prior information. The difference between SRM and SRM WPK is even more spectacular: SRM on the complete channel suffers from channel zeros that are almost in common, whereas SRM WPK performs well.

In Figure.2, we used the same data and we compare the two unbiased forms of SRM WPK: the one using $\|\mathrm{h}\| = 1$ and the one using $\|\mathrm{c}\| = 1$ but $A - \lambda_{min}(A)I$: it is clear that the second approach outperforms the first one (by a factor of more than 5 at SNR=10dB).

In the simulation illustrated in Figure.3, the idea is to study the behavior of the SRM WPK and IQML WPK (one iteration initizalized with SRM WPK) methods versus the conditioning of the propagation channel C. We adopt

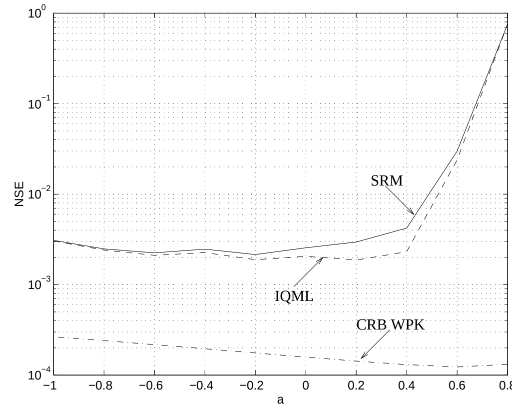

Figure 3. Comparaison of SRM WPK and IQML

the same pulse-shaping filter as before and we consider a propagation channel defined as:

$$\mathrm{C} = \begin{bmatrix} 1 & 1 \\ 1 & a \end{bmatrix}. \tag{29}$$

When $a = 1$, the two subchannels are parallel (zero in common), and when $a = -1$ the two subchannels are orthogonal (channel well conditioned). Simulation results, at SNR=30dB, and for values of a ranging from -1 to 0.8, show that for on the average IQML does not perform drastically better than SRM (the best improvement is about a factor of two, obtained for $a = 0.4$); but we have noted that for some realizations IQML outperforms SRM significantly.

REFERENCES

[1] Stephan V. Schell, Dale L. Smith and W. Gardner. "Blind Channel Identification Using 2nd-Order Cyclostationnary Statistics". In *Proc. EUSIPCO*, Edinburgh, Sept. 1994.

[2] Zhi Ding and Z. Mao. "Knowledge Based Identification of Fractionnally Sampled Channels". In *Proc. ICASSP 95 Conf.*, Detroit, USA, May 1995.

[3] D.T.M. Slock. "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction". In *Proc. ICASSP 94 Conf.*, Adelaide, Australia, April 1994.

[4] D.T.M. Slock and C.B. Papadias. "Blind Fractionally-Spaced Equalization Based on Cyclostationarity". In *Proc. Vehicular Technology Conf.*, pages 1286–1290, Stockholm, Sweden, June 1994.

[5] L.A. Baccala and S. Roy. "A New Time-Domain Blind Channel Identification Method". *IEEE Signal Processing Letters*, 1(6):89 91, June 1994.

[6] D.L. Smith. *Blind Channel Identification and Equalization Using Second-Order Cyclostationarity.* PhD thesis, Pennsylvania State University, Aug. 1996.