Optimal and Suboptimal Approaches for Training Sequence Based Spatio-Temporal Channel Identification in Colored Noise

Hafedh Trigui   Dirk T.M. Slock
Institut EURECOM
2229 route des Crêtes, B.P. 193
06904 Sophia Antipolis Cedex, FRANCE
{trigui,slock}@eurecom.fr

Abstract

In wireless communications, spatial (via antenna arrays) and temporal (excess bandwidth) diversity may be exploited to simultaneously equalize a user of interest while canceling or reducing (cochannel) interfering users. This can be done using the Interference Canceling Matched Filter (ICMF) which we introduced previously. The ICMF depends on the channel for the user of interest, to be estimated with a training sequence, and a blind interference cancellation part. The critical part is the channel estimation. The usual least-squares method may lead to poor estimates in high interference environments. Significant improvements may result from the Maximum-Likelihood (ML) and suboptimal techniques investigated here.

1. Problem Formulation

We consider here linear digital modulation over a linear channel with additive noise. We consider furthermore a FIR multichannel model. The multiple FIR channels are due to oversampling of a single received signal and/or the availability of multiple received signals from an array of antennas (in the context of mobile digital communications) [1]. A third possibility for having multiple channels is when one-dimensional constellations (e.g. BPSK) are transmitted with modulation [2]. In that case, the channel impulse response and the received signal in baseband will be complex. The real and the imaginary parts of the channel (and the received signal) can be considered as two real channels through which the real symbols are received.

To further develop the case of oversampling, the cyclostationary received signal can be written as

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

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

$$\begin{gathered}
\boldsymbol{y}(k) \;=\; \sum_{i=0}^{N-1} \boldsymbol{h}(i)a(k-i) + \boldsymbol{v}(k) \;=\; \boldsymbol{H} A_N(k) + \boldsymbol{v}_k \;, \\
\boldsymbol{y}(k) = \begin{bmatrix} y_1(k) \\ \vdots \\ y_m(k) \end{bmatrix}, \boldsymbol{v}(k) = \begin{bmatrix} v_1(k) \\ \vdots \\ v_m(k) \end{bmatrix}, \boldsymbol{h}(k) = \begin{bmatrix} h_1(k) \\ \vdots \\ h_m(k) \end{bmatrix} \\
\boldsymbol{H}=[\boldsymbol{h}(N-1)\cdots\boldsymbol{h}(0)] \,, A_N(k)=\left[a^H(k-N+1)\cdots a^H(k)\right]^H
\end{gathered} \tag{2}$$

where the subscript i denotes the i^{th} channel and superscript ^{H} denotes Hermitian transpose. In the case of oversampling, $y_i(k)\,, i=1,\ldots,m$ represent the m phases of the polyphase representation of the oversampled signal: $y_i(k)=\mathrm{y}(t_0+(k+\frac{i}{m})T)$. In this representation, we get a discrete-time circuit in which the sampling rate is the symbol rate. Its output is a vector signal corresponding to a SIMO (Single Input Multiple Output) or vector channel consisting of m SISO discrete-time channels where m is the sum of the oversampling factors used for the possibly multiple antenna signals. 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. Assume we receive M samples: $\boldsymbol{Y}_M(k)=\mathcal{T}_M(\boldsymbol{H})\,A_{M+N-1}(k)+\boldsymbol{V}_M(k)$ 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 $[\boldsymbol{H}\;\; 0_{m\times(M-1)}]$ as first block row. We shall simplify the notation with $k=M-1$ to

$$\boldsymbol{Y} = \mathcal{T}(\boldsymbol{H})\,A + \boldsymbol{V}. \tag{3}$$

The term $\boldsymbol{v}(k)$ will be considered here to consist of both spatially and temporally correlated additive zero mean noise. In simulations, we often assume $\boldsymbol{v}(k)$ to consist of temporally and spatially i.i.d. noise plus co-channel multi-user interference. This would mean that the additive noise $v(t)$ is a combination of stationary and cyclostationary components with period T. When the noise consists of multiuser interference plus Gaussian noise, the optimal receiver performs joint detection of all users. However, the estimation of the matrix transfer function from all users to all antennas (and/or sampling phases) is a formidable and often prohibitive task. Furthermore, the complexity of MLSE can be enormous in this case. Therefore we shall concentrate on the detection of one user of interest, ignore the discrete distribution of the interferers and approximate them with a Gaussian distribution. We shall assume that the channel transfer functions for the interferers are also FIR and that their symbol sequences are uncorrelated. Hence we assume that $\boldsymbol{v}(k)$ is a multivariate MA($N^{'}$–1) process.

In the blind estimation problem on the basis of a burst of received data as in (3), the unknown parameters are the channel $\boldsymbol{H}$, the transmitted symbols A and the noise correlation sequence $r_{\boldsymbol{vv}}(0:N^{'}–1)$. However, this ensemble of unknown parameters is unidentifiable from the received data $\boldsymbol{Y}$. A training sequence, i.e. a subset of known transmitted symbols A_1, has to be available to enable estimation of all unknown parameters. In [3], which builds upon previous work as discussed in [3], a two-step procedure was proposed in which the training sequence was used to estimate the channel $\boldsymbol{H}$ via least-squares (as is usually done for training-sequence based channel estimation). A set of parameters equivalent to $r_{\boldsymbol{vv}}(0:N^{'}–1)$ in a filtering structure called the Interference Canceling Matched Filter (ICMF) was then estimated blindly. The remaining symbols A_2 can then be estimated using any of the existing receiver techniques that are based on known channel and noise statistics. The least-squares criterion for the channel identification used in [3] is optimal when the additive noise is white. When the noise is colored and furthermore contains multi-user interference, the channel estimate obtained by least-squares may not be good enough. We furthermore wish to consider also situations in which the training sequence length is long enough for identifiability but not long enough to permit a good quality channel estimate via least-squares. In this paper we propose two solutions of increasing complexity to obtain better channel estimates. In the first solution we estimate the channel from the training sequence via an optimal weighted least-squares criterion. The other solution involves again a weighted least-squares criterion, but with modified input and desired-response signals. We also consider the ML approach and investigate the Cramer-Rao performance bounds.

2. Gaussian Maximum Likelihood

We assume for the received signal $\boldsymbol{Y} = \mathcal{T}(\boldsymbol{H})\,A + \boldsymbol{V}$ that $\boldsymbol{V} \sim \mathcal{N}(0, R_{VV})$ and independent of $A \sim \mathcal{N}(A^o, C_{AA})$ (complex normal variables are assumed to be circular). A^o is the mean for the symbols A and C_{AA} their covariance matrix.

Let $A = \mathcal{P}\left[A_1^{oH}\; A_2^H\right]^H$ where A_1^o are the training symbols and $\mathcal{P}$ is a permutation matrix to account for the fact that the training sequence does not necessarily occur at the beginning. Hence we have

$$A^o = \mathcal{P}\begin{bmatrix} A_1^o \\ 0 \end{bmatrix}\;,\;\; C_{AA} = \mathcal{P}\begin{bmatrix} 0 & 0 \\ 0 & R_{A_2A_2} \end{bmatrix}\mathcal{P}^H \quad (4)$$

One obvious choice for $R_{A_2A_2}$ would be $R_{A_2A_2} = \sigma_a^2 I$. We also introduce the following notation

$$\mathcal{T}\,A = \mathcal{T}_1 A_1 + \mathcal{T}_2 A_2 = \mathcal{T}_1 A_1^o + \mathcal{T}_2 A_2. \quad (5)$$

We get for the correlation and covariance matrices

$$C_{YY} = \mathcal{T}_2 R_{A_2A_2}\mathcal{T}_2^H + R_{VV} = R_{YY} - (\mathcal{T}_1 A_1^o)\,(\mathcal{T}_1 A_1^o)^H\;. \quad (6)$$

We derive now the stochastic (or Gaussian) maximum likelihood approach (GML) as it is introduced in [4]. With the previous notation, we consider the received signal $\boldsymbol{Y} = \mathcal{T}(\boldsymbol{H})\,A + \boldsymbol{V}\,:\,\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}_1(\boldsymbol{H})A_1^o, C_{YY})$ and we shall maximize the complex probability density function (pdf)

$$f(\boldsymbol{Y}|\theta) = \frac{exp\left[-(\boldsymbol{Y}-\mathcal{T}_1(\boldsymbol{H})A_1^o)^H C_{YY}^{-1}(\boldsymbol{Y}-\mathcal{T}_1(\boldsymbol{H})A_1^o)\right]}{\pi^{mM}\,detC_{YY}} \quad (7)$$

with $\theta = \left[vec^H(\boldsymbol{H})tri^H(r_{\boldsymbol{vv}}(0))vec^H(r_{\boldsymbol{vv}}(1:N^{'}–1)\right]^H$ where the notation $vec(\boldsymbol{B})$ denotes a vector formed by stacking the columns of $\boldsymbol{B}$ and $tri(\boldsymbol{B})$ denotes a vector formed by stacking the lower triangular part of $\boldsymbol{B}$. If at least one interferer has an impulse response channel with the same length (or longer) than the user of interest (i.e $N^{'} \geq N$), then a more robust parameterization should be introduced. Thus, we can minimize the negative log-likelihood function

$$\begin{aligned}\mathcal{F}(\boldsymbol{Y}|\phi) = &(\boldsymbol{Y}-\mathcal{T}_1(\boldsymbol{H})A_1^o)^H(\bar{R}_{YY}-\sigma_a^2\mathcal{T}_1(\boldsymbol{H})\mathcal{T}_1(\boldsymbol{H})^H)^{-1}\\ &(\boldsymbol{Y}-\mathcal{T}_1(\boldsymbol{H})A_1^o)+ln(det(\bar{R}_{YY}-\sigma_a^2\mathcal{T}_1(\boldsymbol{H})\mathcal{T}_1(\boldsymbol{H})^H))\end{aligned} \quad (8)$$

w.r.t. $\phi = \left[vec^H(\boldsymbol{H})tri^H(\bar{r}_{\boldsymbol{yy}}(0))vec^H(\bar{r}_{\boldsymbol{yy}}(1:N^{'}–1)\right]^H$ where $\bar{R}_{YY} = \mathrm{E}_{A_1^o} R_{YY} = \sigma_a^2\mathcal{T}\mathcal{T}^H + R_{VV}$ which is banded block Toeplitz (removing now conditioning on A_1^o and assuming $R_{A_1^oA_1^o} = \sigma_a^2 I$ and A_1^o independent of the rest, and $R_{A_2A_2} = \sigma_a^2 I$). Such a parameterization is justified by the fact that the estimation of $\boldsymbol{H}$ or R_{VV} is not consistent. However, the estimation of $\bar{R}_{YY}$ is consistent. Moreover, we can neglect (asymptotically) the estimation errors if the burst is sufficiently long and suppose that $\bar{R}_{YY}$ is known. We consider in this paper only the case where $N^{'} = N$ which is of practical interest. We should emphasize the fact that only the number of the unknown symbols can be (asymptotically) infinite; the length of the training sequence is necessarily finite. The complex Fisher Information Matrix (FIM) can be defined as $J_{\underline{\phi}\underline{\phi}} = -\mathrm{E}_Y \frac{\partial}{\partial \underline{\phi}^*}\left(\frac{\partial \mathcal{F}(\boldsymbol{Y}|\phi)}{\partial \underline{\phi}^*}\right)^H$ where $\underline{\phi} = \left[\phi^H \phi^T\right]^H = \left[h^H \varphi^H\right]^H$ and $h = \left[\boldsymbol{h}^H(0)\cdots\boldsymbol{h}^H(N{-}1)\right]^H$ or more explicitly as

$$
\begin{aligned}
J_{hh}(i,j) &= \left(\mathcal{H}^H C_{YY}^{-1}\mathcal{H}\right)_{(i,j)} + \mathrm{tr}\left\{C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial h_i^*}\right)C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial h_j^*}\right)^H\right\}\\
J_{\underline{\phi}\varphi}(i,j) &= \mathrm{tr}\left\{\frac{\partial C_{YY}}{\partial \underline{\phi}_i^*}C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \varphi_j^*}\right)^H C_{YY}^{-1}\right\}
\end{aligned} \tag{9}
$$

where due to the commutativity of convolution $\mathcal{T}_1(\boldsymbol{H})A_1^o = \mathcal{H}(A_1^o)h = \mathcal{H}h$.

The Cramer-Rao Bound is the lower bound to the error covariance matrix of any unbiased estimator and is given by the inverse of the FIM (which is regular in our case for only one known symbol).

By considering some permutation matrix to push the training sequence at the begining of the burst and applying the partitioned matrix inversion lemma on the matrix C_{YY}, one can see that the additional blind information disappears when the length of the burst tends to infinity. Moreover, the CRB remains unchanged for a given training sequence embedded in bursts with various lengths. This shows that the whole information is concentrated around the training sequence part. Although the complex FIM is regular for only one known symbol, we can not apply a reasoning similar to [5] and refer to our problem as a Semi-Blind one. There is no blind method that allows estimation of the channel of the user of interest in the presence of arbitrary colored noise as considered here. In [6], the special case of a spatially correlated but temporally white noise is considered in which case blind identification can be possible. Another complication is the following. To be able to obtain a positive mS by mS block Toeplitz $\hat{\bar{R}}_{YY}$, by pre and post windowing, one can check easily that the burst should be longer than $(m-1)S+1$. In the multichannel case, we can only estimate structured matrices with smaller size than the burst length to garantee the positive definiteness of $\hat{\bar{R}}_{YY}$.

3. Optimally Weighted Least-Squares

To calculate the GML estimates, we can consider a Newton type algorithm to minimize (8). Since the computational cost can be large, we are interested here in

$$\min_{\boldsymbol{H}} ||\boldsymbol{Y} - \mathcal{T}_1 A_1^o||^2_{C_{YY}^{-1}} . \tag{10}$$

At this point, it is useful to consider the following facts $\forall\theta_o$

$$
\begin{aligned}
\arg\min_{\theta} ||Z{-}X\theta||^2_{[R-X\theta_o\theta_o^H X^H]^{-1}} &= \arg\min_{\theta} ||Z{-}X\theta||^2_{R^{-1}} ,\\
\min_{\theta} ||Z-X\theta||^2_{[R-X\theta_o\theta_o^H X^H]^{-1}} &= \min_{\theta} ||Z{-}X\theta||^2_{R^{-1}} .
\end{aligned} \tag{11}
$$

This means that we can replace the problem in (10) by

$$\min_{\boldsymbol{H}} ||\boldsymbol{Y} - \mathcal{T}_1 A_1^o||^2_{R_{YY}^{-1}} . \tag{12}$$

At this point we still have a complicated optimization problem since $R_{YY} = R_{YY}(\boldsymbol{H})$. We can approximate the problem in (12) with a simplified problem $\min_{\boldsymbol{H}} ||\boldsymbol{Y} - \mathcal{T}_1 A_1^o||^2_{\hat{\bar{R}}_{YY}^{-1}}$ in which $\hat{\bar{R}}_{YY}$ is a banded block Toeplitz estimate, obtained from $\boldsymbol{Y}\boldsymbol{Y}^H$, of $\bar{R}_{YY}$.

Compared to the original unweighted least-squares approach, (10) with C_{YY} fixed at the true value represents the optimally weighted least-squares criterion. A more exact solution to the problem (12) would be the following iterative solution

$$\min_{\boldsymbol{H}^{(i)}} \left\|\boldsymbol{Y} - \mathcal{T}_1(\boldsymbol{H}^{(i)})A_1^o\right\|^2_{\left[\hat{\bar{R}}_{YY} - \sigma_a^2 \mathcal{T}_1(\boldsymbol{H}^{(i-1)})\mathcal{T}_1^H(\boldsymbol{H}^{(i-1)})\right]^{-1}} \tag{13}$$

with $\boldsymbol{H}^{(-1)} = 0$.

4. Wiener Filtering

We propose now another sub-optimal training sequence based approach that we call Wiener Filtering (WF). It uses the fact that the MMSE linear receiver, the Wiener filter, has $\sigma_a^2 \mathbf{H}^\dagger(z) S_{\boldsymbol{yy}}^{-1}(z)$ as transfer function. A LS problem with all the known symbols as desired response can be formulated to estimate the matched filter. As a second step we can improve this estimation by an iterative optimally weighted LS by taking into account the covariance matrix of the errors (can be estimated).

4.1. Infinite Length Wiener Filtering

The difficulty here is how to approximate the IIR filter $S_{\boldsymbol{yy}}^{-1}(z)$ by a FIR one. This problem is discussed in [3] where a finite number of correlation lags of $\boldsymbol{y}_k$ are used to determine the FIR prediction filter $P(z)$ producing the prediction error $\boldsymbol{f}_k = P(z)\boldsymbol{y}_k$ and leading to the FIR model $S_{\boldsymbol{yy}}^{-1}(z) = P^\dagger(z)\sigma_{\boldsymbol{f}}^{-2}P(z)$ where the m by m matrix $\sigma_{\boldsymbol{f}}^2$ denotes the prediction error variance. With this AR approximation we focus on the information around the training sequence (assumed to be far from the burst edges) which makes sense as argued before.

If the prediction order is L and the training sequence contains K symbols, then we need only a portion of length $K{+}2L{+}N{-}1$ of the burst.

Let's express the portion of interest at instant k by $\boldsymbol{Y}_{K+2L+N-1}(k) = \tilde{\mathcal{P}}\boldsymbol{Y}_M(k)$, where $\tilde{\mathcal{P}}$ is a selection matrix. We introduce now the filtered version of the received signal by $\tilde{\boldsymbol{y}}_k = S^{-1}_{\boldsymbol{yy}}(z) \quad \boldsymbol{y}_k = P^{\dagger}(z)\sigma^{-2}_{\boldsymbol{f}}P(z) \quad \boldsymbol{y}_k$ or more conviently by a vectorial notation $\tilde{\boldsymbol{Y}}_{K+N-1}(k) = \mathcal{T}_{K+N-1}(\boldsymbol{P}^{\dagger})\left(\boldsymbol{I}_{K+L+N-1} \otimes \sigma^{-2}_{\boldsymbol{f}}\right)\mathcal{T}_{K+L+N-1}(\boldsymbol{P})\boldsymbol{Y}_{K+2L+N-1}(k)$. where $\otimes$ is the Kronecker product operator, $\boldsymbol{P} = [\boldsymbol{p}(L:-1:1)\ \boldsymbol{I}_m]$ the matrix of prediction coefficients and $\tilde{\boldsymbol{Y}}_{K+N-1}(k) = \left[\tilde{\boldsymbol{y}}^H_k \cdots \tilde{\boldsymbol{y}}^H_{k+K+N-2}\right]^H$. The subscripts are omitted in the following for simplicity of notation.

Formulate first the LS problem as a minimization w.r.t. the channel coefficients of

$$\sum_{j=1}^{K} n_j^2 = \sum_{j=1}^{K} |a_j - \sigma_a^2 h^H \tilde{Y}_j|^2 = \left\| A_1^o - \sigma_a^2 \mathcal{T}(h^H)\tilde{\boldsymbol{Y}} \right\|^2 \tag{14}$$

where $\tilde{Y}_j = \left[\tilde{\boldsymbol{y}}^H_{k+j-1} \cdots \tilde{\boldsymbol{y}}^H_{k+j+N-2}\right]^H$. This leads to the filter $\sigma_a^2 \hat{h} = (\sum_{j=1}^{K} \tilde{Y}_j \tilde{Y}_j^H)^{-1} \sum_{j=1}^{K} \tilde{Y}_j a_j^*$. By exploiting the fact that we use only a part of the received burst and observing that $\boldsymbol{N} = A_1^o - \sigma_a^2 \mathcal{T}(h^H)\tilde{\boldsymbol{Y}} = A_1^o - \sigma_a^2 \mathcal{Y}^T h^*$ consists of a mean part and a perturbation around it, the optimally weighted LS problem corresponding to (14)can be shown to be $\min_h \left\| A_1^o - \sigma_a^2 \mathcal{T}(h^H)\tilde{\boldsymbol{Y}} \right\|^2_{R^{*-1}_{NN}}$ where $R_{NN} = \boldsymbol{Q}A_1^o A_1^{oH}\boldsymbol{Q}^H + \boldsymbol{S}C_{YY}\boldsymbol{S}^H$, and

$$\begin{aligned}\boldsymbol{Q} &= \boldsymbol{I}_K - \sigma_a^2 \mathcal{T}(h^H)\mathcal{T}(\boldsymbol{P}^{\dagger})\left(\boldsymbol{I}_{K+L+N} \otimes \sigma^{-2}_{\boldsymbol{f}}\right)\mathcal{T}(\boldsymbol{P})\tilde{\mathcal{P}}\mathcal{T}_1(\boldsymbol{H}) \\ \boldsymbol{S} &= \sigma_a^2 \mathcal{T}(h^H)\mathcal{T}(\boldsymbol{P}^{\dagger})\left(\boldsymbol{I}_{K+L+N} \otimes \sigma^{-2}_{\boldsymbol{f}}\right)\mathcal{T}(\boldsymbol{P})\tilde{\mathcal{P}}.\end{aligned} \tag{15}$$

The solution to the Wiener filtering problem can be extended to a GML minimization of the negative log-likelihood function $\mathcal{F}(\tilde{\boldsymbol{Y}}|\phi) = N^H R^{*-1}_{NN} N + ln(det(R_{NN}))$ w.r.t ϕ. By using the fact that $\left(\boldsymbol{S}C_{YY}\boldsymbol{S}^H\right)^{-1}\boldsymbol{Q}A_1^o A_1^{oH}\boldsymbol{Q}^H \ll \boldsymbol{I}$ and $ln(det(\boldsymbol{I}+\delta)) \approx \mathrm{tr}(\delta)$ for small δ, we can extract some information from the determinant (and neglect the information in $det(\boldsymbol{S}C_{YY}\boldsymbol{S}^H)$) and get an approximated GML estimate

$$\begin{aligned}\hat{h} = &\left(\sigma_a^4 \mathcal{Y} R^{-1}_{NN}\mathcal{Y}^H + \mathcal{H}^H \boldsymbol{S}^H(\boldsymbol{S}C_{YY}\boldsymbol{S}^H)^{-1}\boldsymbol{S}\mathcal{H}\right)^{-1} \\ &\left(\mathcal{H}^H\boldsymbol{S}^H(\boldsymbol{S}C_{YY}\boldsymbol{S}^H)^{-1}A_1^o + \sigma_a^2 \mathcal{Y} R^{-1}_{NN} A_1^{o*}\right)\end{aligned} \tag{16}$$

4.2. FIR Wiener Filtering

To avoid the FIR approximations of the previous approach, we consider here the FIR Wiener Filter. As before, the LS problem boils, for $d \geq 0$, down to the minimisation w.r.t the channel coefficients of

$$\sum_{k=1}^{K} |a_k - \sigma_a^2 \left[0_{1,m(d+N-1)} h^H 0_{1,m(d-N+1)}\right] R^{-1}_{Y_{N+2d}Y_{N+2d}} Y_{N+2d}(k+N-d-1)|^2. \tag{17}$$

Though FIR approximations are avoided here, the MMSE and hence the estimation variance are slightly higher when d is finite.

5. Wiener Filtering at the ICMF output

If the ICMF is followed by a MMSE linear equalizer, a WF problem can be formulated at its output. The training sequence channel identification is done here as the first step. Since this structure allows only to identify $g(z) = \hat{\mathbf{H}}^{\dagger}(z)\mathbf{H}(z)$, we only have information on a $2N-1$ dimensional subset of the mN parameters. We can not identify the component of $\mathbf{H}(z)$ contained in the left null space of $\mathcal{T}_N(\hat{\mathbf{H}})$.

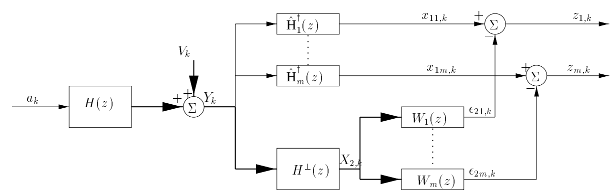

Figure 1. Expanded ICMF (EICMF)

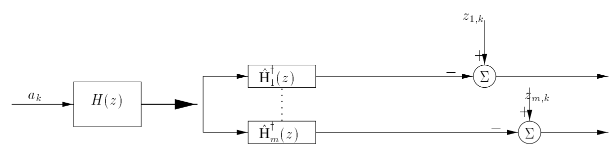

Figure 2. LS solution at the EICMF output

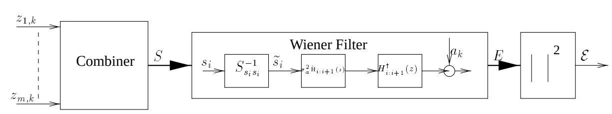

Figure 3. WF solution at the EICMF output

This problem appears also in the classical WF: we have a linear system of K equations and mN parameters. In the case of $mN > K$, we shall consider a combination of cost functions, each for a subset of channels. The combination of cost functions, weighted inversely proportionally to their respective MMSE's, should be minimized in one operation.

For the ICMF, we can create diversity by considering m scalar matched filters, one for each phase; the blocking equalizers $\mathbf{H}^{\perp}(z)$ ([7] and references therein) remains unchanged. However, a Wiener filter should be considered separately for each phase (see figure 1). The resulting ICMF is termed Expanded ICMF (EICMF). Thus, the number of equations becomes proportional to the number of channels: LS (figure 2) leads to $m(K-2N+2)$ equations while WF (figure 3) leads to mK equations if $m>2$ and K equations if $m=2$. The price paid for this additional diversity is the complexity of the receiver (in the interference cancellation part) since we pass for $W(z)$ from a vectorial filter to a matricial one.

6. Simulation Results

We consider two 5-tap random channel impulse responses for the user of interest and the interferer. The 4-QAM signals, assumed to be independent for the two users, are received by three antennas. The performance of the different algorithms are presented in figure 4 and compared to the Cramer-Rao Bound (for the ease of known $\bar{R}_{YY}$) by averaging 50 realisations. The training sequence of length 26 is taken in the middle of a 148 symbol burst length (as in GSM). We show the root normalized mean square error vs. SNR for SIR=10dB. WLS refers to (10). WF refers to (14) with prediction quantities estimated from a burst of 480, while EWF is with exact prediction. WWF corresponds to (16) with estimated prediction and EWWF with exact prediction.

7. Conclusions

The limited performance of the optimally WLS is shown. Even so, this performance is not realisable in practise since it requires the exact C_{YY}. The practical approximation presented in (13) does not work well because it is difficult to guarantee $\hat{\mathrm{C}}_{YY}>0$. Direct minimisation of (8) is very complex and shows similar difficulties. The WLS takes the information of the mean and doesn't consider the information of the covariance matrix. Its performance is far from the CRB. However, the WF approach is simple and can perform in the severe scenario (low SNRs and moderate SIR) as well as the WLS. If we use true correlation matrix $\bar{R}_{YY}$, the performance of WF will be very near the CRB which indicates that the proper estimation of $\bar{R}_{YY}$ is the key problem of this approach. The approximated GML criterion in (16) improves the performance of WF to some extent. The performance of training sequence LS or WF approaches at the EICMF output will be evaluated in a separate paper. It is expected to be good since the SINR at the EICMF output is maximized

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*, August 22-25 1994. Leuven, Belgium.

[2] M. Kristensson B.E. Ottersten and D.T.M. Slock. Blind Subspace Identification of a BPSK Communication Channel. In *Proc. 30th Asilomar Conf. on Signals, Systems and Computers*, 1996. Pacific Grove, CA.

[3] D.T.M. Slock. Spatio-Temporal Training-Sequence-Based Channel Equalization and Adaptive Interference Cancellation. In *Proc. ICASSP96 Conf.*, May 7-10 1996. Atlanta, GA.

[4] E. De Carvalho and D.T.M. Slock. Maximum Likelihood Blind FIR Multi-Channel Estimation with Gaussian Prior for the Symbols. In *Proc. ICASSP97 Conf.*, April 21-24 1997. Munchen, Germany.

[5] E. De Carvalho and D.T.M. Slock. Cramer-Rao Bounds for Semi-Blind, Blind and Training Sequence FIR Multichannel Estimation. *To be submitted to IEEE Trans. Signal Processing*.

[6] K. Abed Meraim P. Loubaton and E. Moulines. Subspace Method for Blind Identification of Multichannel FIR Filters in Noise Field with Unknown Spatial Covariance. In *Proc. 28th Asilomar Conf. on Signals, Systems and Computers*, 1994. Pacific Grove, CA.

[7] J. Ayadi and D.T.M Slock. Cramer-Rao Bounds and Methods for Knowledge Based Estimation of Multiple FIR Channels. In *Proc. SPAWC 97*, April 16-18 1997. Paris, France.

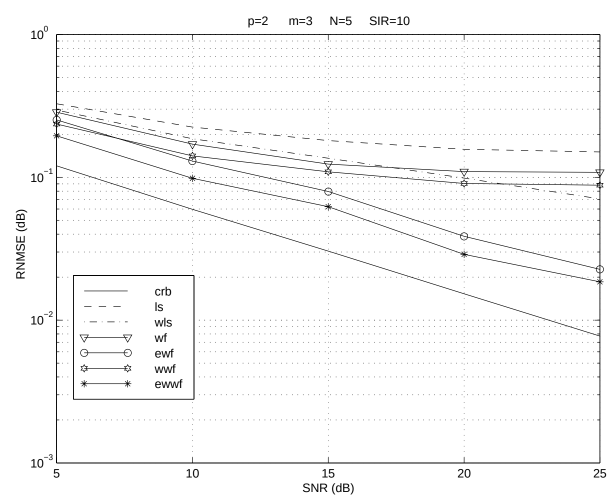

Figure 4. Performance Analysis