ELSEVIER

Signal Processing 80 (2000) 2049–2062

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

Blind channel estimation exploiting transmission filter knowledge

Jaouhar Ayadi[a], Dirk T.M. Slock[b,*]

[a]*Centre Suisse d'Electronique et de Microtechnique (CSEM), Jaquet-Droz 1, CH-2007 Neuchâtel, Switzerland*
[b]*Institut EURECOM, 2229 routes des Cretes, BP 193, 06904 Sophia Antipolis Cedex, France*

Received 5 January 1999; received in revised form 13 August 1999

Abstract

This contribution elaborates on the concept of blind identification of multiple FIR channels with transmission filter knowledge (WTXFK). This prior knowledge could, in fact, include not only the transmitter (TX) (pulse shaping) filter but also the receiver (RX) filter present in digital communication systems. Exploitation of this side information allows the estimation to concentrate on the impulse response of the actual propagation channel 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 directly by the channel and consider their extension to incorporate the prior knowledge. These methods include essentially subchannel response matching (SRM), subspace fitting and maximum likelihood (ML) techniques. All these methods are formulated for burst mode transmission. 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, a capability of the extended method to identify ill-conditioned channels, that one particular version SRM WTXFK often outperforms another one, and that ML methods can still further improve performance. © 2000 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Dieser Beitrag arbeitet das Konzept der blinden Identifizierung mehrerer FIR Kanäle mit Wissen über das Sendefilter ("with transmission filter knowledge", WTXFK) aus. Dieses Vorwissen könnte tatsächlich nicht nur den Sendefilter ("transmitter filter", TX) (Pulsformer), sondern auch das in digitalen Kommunikationssystemen vorhandene Empfangsfilter ("receiver filter", RX) beinhalten. Ausnutzen dieser Seiteninformation erlaubt es der Schätzung sich auf die Impulsantwort des tatsächlichen Ausbreitungskanals zu konzentrieren. Deshalb kann diese Schätzung genauer durchgeführt werden. Da das Vorwissen durch die Kanalimpulsantwort ausgedrückt wird, rekapitulieren wir eine Anzahl blinder Kanalschätzmethoden, die direkt durch den Kanal parametrisiert sind und betrachten ihre Erweiterung dahingehend, das Vorwissen einzubezichen. Diese Methoden beinhalten im wesentlichen "Subchannel Response Matching (SRM)", "Subspace Fitting" und "Maximum Likelihood (ML)" Techniken. All diese Methoden werden für die Übertragung im Burst-Modus formuliert. Wir diskutieren ebenfalls Grenzen der Leistungsfähigkeit durch Cramer–Rao Schranken ("Cramer–Rao Bounds" CRBs). Sowohl die Verfahren als auch die CRBs werden in einem deterministischen und einem Gaußschen Kontext für die unbekannten Sendesignale diskutiert. Simulationsergebnisse deuten an, dass das Ausnutzen von Vorwissen zu signifikanten Verbesserungen führen kann, dass die erweiterte Methode fähig ist, schlecht

* Corresponding author. Tel.: + 33-4-93-00-26-06.
E-mail address: slock@eurecom.fr (D.T.M. Slock).

PII: S0165-1684(00)00069-4

konditionierte Kanäle zu identifizieren, dass eine spezielle Version – SRM WTXFK – oftmals bessere Ergebnisse als andere Verfahren erzielt und dass ML Methoden die Leistungsfähigkeit noch weiter verbessern können.

Résumé

Cette contribution présente le concept de l'identification aveugle des canaux multiples à réponse impulsionnelle finie (RIF) avec connaîssance des filtres de transmission. Cette information a priori peut en réalité non seulement inclure le filtre de transmisssion (TX) (filtre de mise en forme) mais aussi le filtre de reception (RX), présents dans les systèmes de communications numériques. L'exploitation de cette information interne permet à l'opération d'estimation de se concentrer sur la réponse impulsionnelle de la vraie partie à estimer du canal (canal de propagation) en soi. Par conséquent, l'estimation peut être faite de façon plus précise. Comme l'information a priori est exprimée en terme de la réponse impulsionnelle du canal, nous reconsidérons une classe de méthodes d'estimation aveugle du canal qui sont paramètrisées directement par le canal et nous considérons leur extension pour incorporer l'information a priori. Ces méthodes incluent essentiellement le “Subchannel Response Matching” (SRM), l'ajustement de sous-espace et les techniques de Maximum de Vraisemblance (MV). Toutes ces méthodes sont formulées pour un mode de transmission par paquets. Nous discutons aussi les limites de performance sous forme de bornes de Cramer–Rao (BCR). Les méthodes et les BCRs sont discutées dans deux contextes des symboles inconnus transmis : déterministe et gaussien. Les résultats de simulation indiquent que l'exploitation de l'information a priori peut conduire à des améliorations significatives, à une capacité de la méthode étendue pour identifier les canaux mal-conditionnés, qu'une version particulière de SRM WTXFK est souvent supériuere (d'un point de vue de performance) à une autre version, et que les méthodes de MV peuvent encore améliorer la performance.

Keywords: Mobile communications; Transmission/reception filters knowledge; Blind multichannel identification; Cramer–Rao bounds

1. Introduction

In a mobile radio transmission context, channels are specific in that they may vary rapidly. Due to bandwidth limitations and multipath propagation, the transmission channel distorts the signal being transmitted, leading to inter-symbol interference (ISI). In order to recover the emitted data, the receiver needs to identify this channel distortion and equalize it. Classical system identification techniques require the use of both system input and output, which leads to the transmission of a training sequence, i.e. a set of fixed data (that do not carry information) that are known to both transmitter and receiver. The use of a training sequence reduces the transmission rate, especially when the training sequence has to be retransmitted often, due to the possibly fast channel variations that occur in mobile communications and consequently decreases the bandwidth efficiency. To the contrary, blind equalizers adapt without using a training sequence.

The identification of a non-minimum phase channel requires the use of higher-order statistics (HOS) in the case of stationary signals. When the signal is cyclostationary, which is the case for mobile communications signals, the channel can be identified using only second-order statistics (SOS) of the output. The sufficiency of SOS to identify the channel is due to an introduction of a linear multichannel formulation in the blind channel identification problem [29,30]. The multiple channels obtained in this formulation can arise in three different ways. The first way is perhaps artificial. It arises when using fractionally spaced equalizers. When the received signal is oversampled with a factor two w.r.t. the symbol rate, the even and odd received samples can be considered as two discrete-time received signals, corresponding to two symbol rate discrete-time channels that are the even and odd samples of the oversampled continuous-time channel response. Another possibility is to physically have multiple channels, which occurs when multiple antennas are used. The third way arises if the symbol constellation is one dimensional (e.g. PAM or BPSK) and the transmitted signal is modulated [14]. In that case, the baseband channel impulse response has a real and imaginary component, wheras the input is purely real. Hence working with real signals only, we get a one-input–two-output system.

A large proportion of recent research dealing with blind channel estimation and/or equalization has been devoted to techniques based on SOS of the received data. This attention is generally justified by the lower complexity of the methods based on SOS which makes the use of this class of blind channel estimation methods more desirable compared to, e.g. HOS-based techniques [3,9,10,20]. The fact that SOS can be sufficient for channel identification is due to the multichannel aspect. Our belief is that the fact that the channel can be estimated fairly accurately using relatively few data (a must for mobile communications) is due to the good estimability of the signal subspace with this few amount of data whereas the second-order moment cannot be well estimated. The first method for blindly identifying a linear channel from the SOS of the cyclostationary oversampled received signal was introduced by Tong et al. [29]. Since then, many researchers have investigated new blind identification/estimation techniques based on SOS. A first class of these SOS-based techniques deals with subspace-based algorithms [18,19,23,24,30]. The second class concerns multichannel linear prediction-based techniques [22,23]. However, channel identification using SOS alone is not sufficiently robust to be adoptable in standards. Therefore, it is necessary to exploit side information also. A first case of considering such side information could be the exploitation of a short training sequence [5]. Such a training sequence may be too short to allow reliable channel identification by itself. Hence, an extension of the blind multichannel techniques to incorporate the knowledge of a short training sequence can be considered. Such channel identification techniques are called semi-blind. In [5], Cramer–Rao performance bounds for semi-blind channel identification are investigated and a comparison with blind and training-sequence-based techniques is presented. In this paper, we consider the side information due to the prior knowledge corresponding to the transmitter (TX) and/or receiver (RX) filters present in digital communication systems, as illustrated in Fig. 1. This is motivated by the fact that in a digital communication framework, such as mobile communications, the receiver will have knowledge about the pulse shaping filter used at the transmitter. Hence, using this side information will simplify the channel estimation problem and reduce the number of parameters to be estimated which could require less signal power or less data length to solve the channel identification problem. Furthermore, in the case of ill-conditioned channels, SOS-based blind multichannel estimation techniques fail to estimate the channel because the matrix used in the optimization criterion is ill-conditioned. When transmission filter knowledge is used, the matrix of the new optimization criterion often becomes well conditioned and hence blind multichannel identification methods WTXFK can accurately identify the channel in those cases. Indeed, from an identifiability point of view, a SOS-based blind multichannel identification technique will not work if the monochannel, that corresponds to the factorization of common zeros from the subchannels, is non-minimum phase. However, if the monochannel corresponds to the TX and/or RX filters then the problem is fixed because we have prior knowledge of these filters.

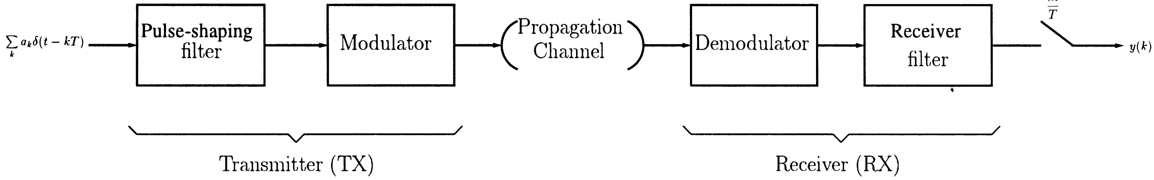

Fig. 1. A typical digital communication system.

This paper is organized as follows: in Section 2, we introduce the data model corresponding to the blind channel identification problem where we specify the multichannel aspect. In Section 3, we present the different blind channel estimation methods that are directly parameterized by the channel. In Section 4, the basic approach of incorporating the side information in the blind estimation methods is formulated based on the polyphase representation of the channel. This problem formulation is exploited in Section 5 where we reformulate the minimization criteria of the blind estimation methods and we specify the minimization constraints for the extended methods (methods incorporating the prior knowledge). The channel identifiability issues are discussed in Section 6, we show that methods using the side information can identify the channel for a certain channel class where methods without prior knowledge fail to achieve this channel identifiability. Section 7 deals with performance analysis through the Cramer–Rao bounds (with or without prior knowledge) for both deterministic and Gaussian symbols cases. This is followed in Section 8 by some numerical experiments illustrating a comparison between the different Cramer–Rao bounds (with and without prior knowledge) and the performance of the (extended) blind channel estimation methods.

2. Problem formulation

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 [21] and further explored in [7]. 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 [22,26]) or from a combination of both. For m channels the discrete-time input–output relationship can be written as

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

where $\boldsymbol{y}(k) = [y_1(k) \cdots y_m(k)]^{\mathrm{T}}, \boldsymbol{h}(i) = [h_1(i) \cdots h_m(i)]^{\mathrm{T}}$, $\boldsymbol{v}(k) = [v_1(k) \cdots v_m(k)]^{\mathrm{T}}$, $\boldsymbol{H} = [\boldsymbol{h}(N-1) \cdots \boldsymbol{h}(0)]$, $A_N(k) = [a(k-N-1) \cdots a(k)]^{\mathrm{T}}$ and superscript T denotes transpose. Let $\boldsymbol{H}(z) = \sum_{i=0}^{N-1} \boldsymbol{h}(i)z^{-i} = [H_1(z) \cdots H_m(z)]^{\mathrm{T}}$ be the SIMO channel transfer function, and $\boldsymbol{h} = [\boldsymbol{h}^{\mathrm{T}}(N-1) \cdots \boldsymbol{h}^{\mathrm{T}}(0)]^{\mathrm{T}}$. Consider the symbols i.i.d. if required and additive independent white Gaussian circular noise $\boldsymbol{v}(k)$ with $r_{\boldsymbol{vv}}(k-i) = E\boldsymbol{v}(k)\boldsymbol{v}(i)^{\mathrm{H}} = \sigma_v^2 I_m \,\delta_{ki}$ where superscript H denotes Hermitian transpose. Assume we receive M samples:

$$\boldsymbol{Y}_M(k) = \mathscr{T}_M(\boldsymbol{H})A_{M+N-1}(k) + \boldsymbol{V}_M(k), \quad (2)$$

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

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

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

3. Blind channel estimation

The channel can either be parameterized by its impulse response $\boldsymbol{h}$ or by the noise-free multivariate prediction error filter $\boldsymbol{P}(z)$ and $\boldsymbol{h}(0)$ which satisfy $\boldsymbol{P}(z)\boldsymbol{H}(z) = \boldsymbol{h}(0)$ [26]. 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 $\boldsymbol{h}$. Two approaches exist, depending on whether the symbols are considered deterministic or Gaussian unknowns.

3.1. Methods for deterministic symbols

3.1.1. Subchannel response matching (SRM)

The SRM approach was introduced in [2] and correponds also to Liu and Xu's deterministic approach [15–17]. The SRM approach was also presented in [11] and used as initialization method in Yingbo Hua's algorithm [13]. In order to explain the SRM technique, consider first the case of two channels: $m = 2$. One can observe that for noise-free signals, we have $H_2(z)y_1(k) - H_1(z)y_2(k) = 0$, which can be written in matrix form as $[H_2(z)\ -H_1(z)]\boldsymbol{y}(k) = \boldsymbol{H}^{\perp\dagger}(z)\boldsymbol{y}(k) = \boldsymbol{H}^{\perp\dagger}(z)\boldsymbol{H}(z)a(k) = 0$ where, e.g. $\boldsymbol{H}^{\dagger}(z) = \boldsymbol{H}^{\mathrm{H}}(1/z^*)$. Stacking these zeros into a vector for the signal $\{\boldsymbol{y}(k)\}_{k=0,\dots,M-1}$, we get an expression of the form $\mathscr{Y}\boldsymbol{h} = 0$ for some structured matrix $\mathscr{Y}$. Under the constraint $\|\boldsymbol{h}\|_2 = 1$, we find $\boldsymbol{h} = V_{\min}(\mathscr{Y}^{\mathrm{H}}\mathscr{Y})$ where $V_{\min}(A)$ denotes the eigenvector corresponding to the minimum eigenvalue of A. For $m > 2$, blocking equalizers $\boldsymbol{H}^{\perp\dagger}(z)$ can be constructed by considering the (sub)channels in pairs. The choice of $\boldsymbol{H}^{\perp\dagger}(z)$ is far from unique. To begin with, the number of pairs to be considered, which is the number of rows in $\boldsymbol{H}^{\perp\dagger}(z)$, is not unique. The minimum number is $m-1$ whereas the maximum number is $m(m-1)/2$, with corresponding $\boldsymbol{H}^{\perp}_{\min}(z)$ and $\boldsymbol{H}^{\perp}_{\max}(z)$. The choice of $\boldsymbol{H}^{\perp}_{\min}(z)$ is not unique. The convolution $\boldsymbol{H}^{\perp}(z)\boldsymbol{y}(k)$ involving $\{\boldsymbol{y}(k)\}_{k=0\cdots M-1}$ can be written in matrix form as $\mathscr{T}(\boldsymbol{h}^{\perp})\boldsymbol{Y}$. Since for the noise-free signal we get $\mathscr{T}(\boldsymbol{h}^{\perp})\boldsymbol{Y} = 0$, the SRM method minimizes the criterion $\|\mathscr{T}(\boldsymbol{h}^{\perp})\boldsymbol{Y}\|_2^2$. 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

$$
\begin{aligned}
\mathscr{J} &= \frac{1}{2\pi \mathrm{j}} \oint \operatorname{tr}\{\hat{\boldsymbol{H}}^{\perp\dagger} \mathrm{S}_{yy} \hat{\boldsymbol{H}}^{\perp}\} \frac{\mathrm{d}z}{z} \\
&= \frac{\sigma_a^2}{2\pi \mathrm{j}} \oint \boldsymbol{H}^{\dagger} \hat{\boldsymbol{H}}^{\perp} \hat{\boldsymbol{H}}^{\perp\dagger} \boldsymbol{H} \frac{\mathrm{d}z}{z} + \frac{\sigma_v^2}{2\pi \mathrm{j}} \oint \operatorname{tr}\{\hat{\boldsymbol{H}}^{\perp\dagger} \hat{\boldsymbol{H}}^{\perp}\} \frac{\mathrm{d}z}{z}.
\end{aligned}
\tag{4}
$$

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

$$
\min_{\|\hat{h}\|=1} \mathscr{J} = \alpha\sigma_v^2 \|\hat{\boldsymbol{h}}\|^2 + \frac{\sigma_a^2}{2\pi \mathrm{j}} \min_{\|\hat{h}\|=1} \oint \boldsymbol{H}^{\dagger} \hat{\boldsymbol{H}}^{\perp} \hat{\boldsymbol{H}}^{\perp\dagger} \boldsymbol{H} \frac{\mathrm{d}z}{z}
\tag{5}
$$

which leads to the correct value $\hat{\boldsymbol{h}} = \boldsymbol{h}$ (and hence an unbiased estimate!) apart from a scale factor (and assuming the channel order is chosen correctly). Here the motivation for chosing a balanced $\boldsymbol{H}^{\perp}(z)$ becomes apparent. The minimum number of rows in $\boldsymbol{H}^{\perp\dagger}_{\mathrm{bal}}(z)$ is m in which case $\alpha = 2$. The choice for such a $\boldsymbol{H}^{\perp}_{\mathrm{bal,min}}(z)$ is not unique. Note that $\boldsymbol{H}^{\perp}_{\max}(z)$ is balanced with $\alpha = m-1$. The choice of the noise subspace parameterization $\boldsymbol{H}^{\perp\dagger}(z)$ as $\boldsymbol{H}^{\perp\dagger}_{\min}(z)$ corresponds to Xu's deterministic least-squares channel identification approach [17]. In the literature, the SRM method is always proposed using $\boldsymbol{H}^{\perp}_{\max}(z)$. We get for instance

$$
\boldsymbol{H}^{\perp\dagger}_{\min}(z) = \begin{bmatrix} -H_2(z) & H_1(z) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -H_m(z) & 0 & \cdots & H_1(z) \end{bmatrix},
\tag{6}
$$

$$
\boldsymbol{H}^{\perp\dagger}_{\mathrm{bal,min}}(z) = \begin{bmatrix} -H_2(z) & H_1(z) & 0 & \cdots & 0 \\ 0 & -H_3(z) & H_2(z) & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ H_m(z) & 0 & \cdots & 0 & -H_1(z) \end{bmatrix}.
\tag{7}
$$

Continuing with this $\boldsymbol{H}^{\perp\dagger}_{\mathrm{bal}}(z)$, its ith row can be written as

$$
\boldsymbol{H}^{\perp\dagger}_{\mathrm{bal},i}(z) = \boldsymbol{H}^{\mathrm{T}}(z)\mathscr{P}_i,\ \mathscr{P}_{i+1} = \mathscr{C}\mathscr{P}_i\mathscr{C}^{\mathrm{H}}, \mathscr{P}_1 = \begin{bmatrix} 0 & 1 & 0 & \cdots \\ -1 & 0 & \cdots & \\ 0 & \vdots & \ddots & \\ \vdots & & & \end{bmatrix},\ \mathscr{C} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & 0 & 1 & 0 \end{bmatrix}.
\tag{8}
$$

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

$$\begin{aligned}\mathrm{tr}\{\mathscr{T}(\boldsymbol{h}^{\perp})\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}\mathscr{T}^{\mathrm{H}}(\boldsymbol{h}^{\perp})\} &= \mathrm{tr}\left\{\boldsymbol{h}^{\perp}\left(\sum_{k=N-1}^{M-1} \boldsymbol{Y}_N(k)\boldsymbol{Y}_N^{\mathrm{H}}(k)\right)\boldsymbol{h}^{\perp\mathrm{H}}\right\} \\ &= (M-N+1)\mathrm{tr}\{\boldsymbol{h}^{\perp}\hat{R}_{YY}\boldsymbol{h}^{\perp\mathrm{H}}\},\end{aligned} \tag{9}$$

where the ith row of $\boldsymbol{h}^{\perp}$ is $\boldsymbol{h}_i^{\perp} = \boldsymbol{h}^{\mathrm{T}}\mathscr{S}_i, \mathscr{S}_i = I_N \otimes \mathscr{P}_i$ and $\otimes$ denotes Kronecker product. Hence the SRM criterion in (9) becomes

$$\min_{\boldsymbol{h}} \boldsymbol{h}^{\mathrm{H}} B \boldsymbol{h}, \quad \text{where } B = \sum_{i=1}^{m} \mathscr{S}_i \hat{R}_{YY}^{*} \mathscr{S}_i^{\mathrm{H}}. \tag{10}$$

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

If the exact R_{YY} is used, then the noise contribution to criterion (10) is $2\sigma_v^2\|\boldsymbol{h}\|^2$. Hence the minimization of (10) subject to $\|\boldsymbol{h}\| = 1$ leads to the consistent SRM estimate $\boldsymbol{h} = V_{\min}(B)$, at least if the order is chosen correctly. Since $\sigma_v^2 = \lambda_{\min}(R_{YY})$, the minimum eigenvalue of R_{YY}, the noise contribution can be eliminated by replacing $\hat{R}_{YY}$ by $\hat{R}_{YY} - \lambda_{\min}(\hat{R}_{YY})I$ or, even better, by replacing B by $A = B - \lambda_{\min}(B)I$ (the former choice does not make B singular with a finite amount of data). With this modification, the criterion in (10) becomes (asymptotically) insensitive to the noise contribution and any normalization of $\boldsymbol{h}$ will lead to a consistent estimate.

3.1.2. *Signal subspace fitting* (*SSF*)

The structure of the covariance matrix of the received signal $\boldsymbol{Y}$ is

$$R_{YY} = E\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}} = \mathscr{T}(\boldsymbol{H})R_{AA}\mathscr{T}^{\mathrm{H}}(\boldsymbol{H}) + \sigma_v^2 I_{mM}, \tag{11}$$

where R_{AA} is the symbol covariance matrix $EAA^{\mathrm{H}} > 0$. The covariance matrix R_{YY} can be decomposed into signal and noise subspace contributions:

$$\begin{aligned}R_{YY} = E\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}} &= \sum_{i=1}^{M+N-1} \lambda_i V_i V_i^{\mathrm{H}} + \sum_{i=M+N}^{mM} \lambda_i V_i V_i^{\mathrm{H}} \\ &= \mathscr{V}_S \Lambda_S \mathscr{V}_S^{\mathrm{H}} + \mathscr{V}_N \Lambda_N \mathscr{V}_N^{\mathrm{H}}.\end{aligned} \tag{12}$$

In the eigen decomposition of the covariance matrix R_{YY} given in (12), the real non-negative eigenvalues λ_i are ordered in descending order, $\lambda_i > \sigma_v^2$ for $i = 1, \ldots, M+N-1$; $\Lambda_N = \sigma_v^2 I_{(m-1)M-N+1}$ and the sets of the eigenvectors $\mathscr{V}_S$ and $\mathscr{V}_N$ are orthonormal: $\mathscr{V}_S^{\mathrm{H}}\mathscr{V}_N = 0$.

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

$$\min_{\boldsymbol{h},T} \|\mathscr{T}(\boldsymbol{H}) - \mathscr{V}_S T\|_F, \tag{13}$$

where $\|X\|_F^2 = \mathrm{tr}\{X^{\mathrm{H}}X\}$. The optimal transformation matrix T can be found to be

$$T = \mathscr{V}_S^{\mathrm{H}}\mathscr{T}(\boldsymbol{H}). \tag{14}$$

Using (14), we obtain [26]

$$\min_{\|\boldsymbol{h}\|_2=1} \mathrm{tr}\{\mathscr{T}^{\mathrm{H}}(\boldsymbol{H})P_{\mathscr{V}_S}^{\perp}\mathscr{T}(\boldsymbol{H})\} = \min_{\|\boldsymbol{h}\|_2=1} \boldsymbol{h}^{\mathrm{H}} A \boldsymbol{h}, \tag{15}$$

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

3.1.3. *Noise subspace fitting* (*NSF*)

Similarly, since $\mathscr{V}_N^{\mathrm{H}}\mathscr{T}(\boldsymbol{H}) = 0$, $\mathscr{V}_N$ spans the noise subspace and $\mathscr{T}^{\mathrm{H}}(\boldsymbol{h}^{\perp})$ spans most of it. So we can introduce the following noise subspace fitting problem:

$$\min_{\boldsymbol{h},T} \|\mathscr{T}^{\mathrm{H}}(\boldsymbol{h}^{\perp}) - \mathscr{V}_N T\|_F. \tag{16}$$

After optimization w.r.t. T, we obtain

$$\begin{aligned}&\min_{\|\boldsymbol{h}\|_2=1} \mathrm{tr}\{\mathscr{T}(\boldsymbol{h}^{\perp})P_{\mathscr{V}_N}^{\perp}\mathscr{T}^{\mathrm{H}}(\boldsymbol{h}^{\perp})\} \\ &\quad = \mathrm{tr}\{\boldsymbol{h}^{\perp}B\boldsymbol{h}^{\perp\mathrm{H}}\} = \boldsymbol{h}^{\mathrm{H}} A \boldsymbol{h},\end{aligned} \tag{17}$$

where B can be determined from $P_{\mathscr{V}_N}^{\perp} = P_{\mathscr{V}_S}$ and $A = \sum_{i=1}^{m} \mathscr{S}_i B^{*} \mathscr{S}_i^{\mathrm{H}}$ for $\boldsymbol{H}_{\mathrm{bal}}^{\perp\dagger}(z)$ defined in (7).

3.1.4. *Deterministic ML* (*DML*)

Introduced in [22] for the case of $m = 2$ channels and extended in [26] to an arbitrary m, the DML method was adapted to the multi-user case in [24,25].

The considered likelihood is conditional on the transmitted symbols and the channel parameters, which are hence treated as deterministic unknowns. The stochastic part only comes from the additive noise. With the Gaussian white noise assumption, maximizing the likelihood reduces to

$$\min_{A,h} \|Y - \mathcal{T}(H)A\|^2. \tag{18}$$

This minimization problem is separable: for a fixed h, the optimal transmitted symbol estimates are

$$A = (\mathcal{T}^{\mathrm{H}}(H)\mathcal{T}(H))^{-1}\mathcal{T}^{\mathrm{H}}(H)Y, \tag{19}$$

eliminating A in terms of h via (19) from (18), we get $\min_h Y^{\mathrm{H}} P^{\perp}_{\mathcal{T}(H)} Y$, which means that the DML criterion boils down to adjusting the noise subspace via H. Now, we have approximately $P^{\perp}_{\mathcal{T}(H)} \approx P_{\mathcal{T}^{\mathrm{H}}(h^{\perp})}$ where the approximation error disappears asymptotically. Hence, we get

$$\begin{aligned}\min_{\|h\|=1} \; & Y^{\mathrm{H}} P_{\mathcal{T}^{\mathrm{H}}(h^{\perp})} Y \\ &= Y^{\mathrm{H}}\mathcal{T}^{\mathrm{H}}(h^{\perp})[\mathcal{T}(h^{\perp})\mathcal{T}^{\mathrm{H}}(h^{\perp})]^{+}\mathcal{T}(h^{\perp})Y \\ &= h^{\mathrm{H}}(\mathcal{Y}^{\mathrm{H}}[\mathcal{T}(h^{\perp})\mathcal{T}^{\mathrm{H}}(h^{\perp})]^{+}\mathcal{Y})h = h^{\mathrm{H}}Ah, \end{aligned} \tag{20}$$

where $\mathcal{T}(h^{\perp})Y = \mathcal{Y}h$ for some $\mathcal{Y}$. The choice of the noise subspace parameterization $H^{\perp\dagger}(z)$ using all pairs of channels leads to Yingbo Hua's ML method [12,13]. The optimization problem in (20) is non-linear. It can easily be solved iteratively in such a way that in each iteration, a quadratic problem appears. The iterative quadratic (IQ) strategy considers the quadratic "numerator" of the criterion, and for $h^{\perp}$ 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 $\mathcal{T}(h^{\perp})Y$ is $\mathcal{T}(h^{\perp})V$ with covariance matrix $\sigma_v^2\mathcal{T}(h^{\perp})\mathcal{T}^{\mathrm{H}}(h^{\perp})$. Asymptotically, any choice for $H^{\perp}(z)$ leads to the same performance (evry $H^{\perp\dagger}(z)$ contains a $H^{\perp\dagger}_{\min}(z)$) since

$$P_{H^{\perp}(z)} = P_{H^{\perp}_{\min}(z)} = P^{\perp}_{H(z)}. \tag{21}$$

3.1.5. Determination of $H(z)$ from $\bar{P}(z) = h^{\perp H}(0)P(z)$

This technique was proposed in [26] and uses a prediction-based noise subspace parameterization ($\bar{P}(z)$) which is not linear in terms of the channel impulse response). To begin with, consider first the linear prediction approach introduced in [22], let $P(z) = \sum_{i=0}^{L} p(i)\,z^{-i}$ with $p(0) = I_m$ be the MMSE multivariate prediction error filter of order L for the noise-free received signal $y(k)$. If $L \geqslant \underline{L} = \lceil (N-1)/(m-1) \rceil$, then it can be shown [26] that (with probability 1, for a completely random FIR channel impulse response)

$$P(z)H(z) = h(0). \tag{22}$$

From (22) it is clear that $H(z)$ and $P(z)$, $h(0)$ are equivalent parameterizations. Consider the full rank $m \times (m-1)$ matrix $h^{\perp}(0)$ defined such that $h^{\perp\mathrm{H}}(0)h(0) = 0$, then (22) implies that $\bar{P}(z) = h^{\perp\mathrm{H}}(0)P(z)$ is a $(m-1)\times m$ polynomial that satisfies

$$\bar{P}(z)H(z) = 0. \tag{23}$$

$\bar{P}(z)$ or equivalently $P(z)$ and $h(0)$ can be estimated using linear prediction or IQDML. If $\bar{P}(z)$ is estimated in a way that is robust to order overestimation, then the order of $H(z)$ is known and $H(z)$ can be estimated straightforwardly from $\bar{P}(z)$. This can be done using the following criterion:

$$\min_h \frac{1}{2\pi \mathrm{j}} \oint H^{\dagger}(z)\bar{P}^{\dagger}(z)\bar{P}(z)H(z)\frac{\mathrm{d}z}{z}. \tag{24}$$

Since $H(z) = Q(z)h = [z^{-(N-1)}I_m \quad z^{-(N-2)}I_m \quad \cdots \quad I_m]\, h$, the minimization problem given in (24) can be written as

$$\begin{aligned}\min_h \; & \frac{1}{2\pi \mathrm{j}} \oint h^{\mathrm{H}} Q^{\dagger}(z)\bar{P}^{\dagger}(z)\bar{P}(z)Q(z)h\frac{\mathrm{d}z}{z} \\ &= \min_h h^{\mathrm{H}}\left(\frac{1}{2\pi \mathrm{j}}\oint Q^{\dagger}(z)\bar{P}^{\dagger}(z)\bar{P}(z)Q(z)\frac{\mathrm{d}z}{z}\right)h \end{aligned} \tag{25}$$

which is again of the form $\min_h h^{\mathrm{H}}Ah$.

3.2. Methods for the Gaussian symbol model

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. One main approach in the Gaussian case is ML (GML). In this case $Y \sim \mathcal{N}(0, R_{YY})$ with

$R_{YY} = \sigma_a^2 \mathcal{T}(\boldsymbol{H})\mathcal{T}^{\mathrm{H}}(\boldsymbol{H}) + \sigma_v^2 I$. The negative log likelihood to be minimized is

$$\mathcal{L}(\boldsymbol{h}) = c^{\mathrm{t}} + \ln\det R_{YY} + \boldsymbol{Y}^{\mathrm{H}} R_{YY}^{-1} \boldsymbol{Y}. \tag{26}$$

Standard optimization techniques such as the Gauss–Newton or scoring methods can be applied. Another approach for the Gaussian symbol model is covariance matching [8].

4. TX/RX filter knowledge

The exploitation of TX/RX filter knowledge requires, in principle, that the sampling rate satisfy the Nyquist criterion and hence oversampling w.r.t. the symbol rate be used. Assume at first the case of a single receiver antenna. Consider now the overall impulse response $h(t) = c(t) * g(t)$ being the convolution of two systems: $g(t)$ is either the TX filter pulse shape or its convolution with the RX filter, and $c(t)$ is either the propagation channel convolved with the RX filter or just the propagation channel resp. In the frequency domain we get $H(f) = C(f)\,G(f)$. Consider now that $g(t)$ is band-limited. Then $G(f) = G(f)\,F(f)$ where $F(f)$ is an ideal lowpass filter (unit height rectangle) with bandwidth greater or equal to that of $G(f)$. Hence $H(f) = G(f)(F(f)C(f))$ where both factors $G(f)$ and $F(f)C(f)$ are now band-limited. With (sufficient) oversampling (essentially satisfying the Nyquist criterion for $G(f)$), the sampled version of $h(t)$ can be considered as the convolution of the sampled version of $g(t)$ with a certain discrete-time representation for $c(t)$, corresponding to a sampled version of $F(f)C(f)$. Note that if the sampling rate exceeds the bandwidth of $G(f)$, then $F(f)$ is not unique since its bandwidth can be chosen arbitrarily in between the bandwidth of $G(f)$ and the sampling frequency. So also the discrete-time representation for $c(t)$ is then non-unique.

Consider now a certain oversampling factor m and let the oversampled transfer function $H(z) = C(z)G(z)$ of the overall channel be the cascade of the actual channel $C(z)$ and the combined TX/RX filter $G(z)$. Each of these transfer functions can be decomposed into its polyphase components at the symbol rate, e.g. $H(z) = \sum_{i=0}^{m-1} z^{-i} H_i(z^m)$. These components can also be represented in vector form, $\boldsymbol{G}(z) = [G_1(z) \cdots G_m(z)]^{\mathrm{T}} = \sum_{k=0}^{K-1} \boldsymbol{g}(k) z^{-k}$ and $\boldsymbol{C}(z) = [C_1(z) \cdots C_m(z)]^{\mathrm{T}} = \sum_{k=0}^{L-1} \boldsymbol{c}(k) z^{-k}$ with $K + L - 1 = N$. The relations between the polyphase components can be obtained from

$$\sum_{i=0}^{m-1} z^{-i} H_i(z^m) = \left(\sum_{k=0}^{m-1} z^{-k} G_k(z^m)\right)\left(\sum_{l=0}^{m-1} z^{-l} C_l(z^m)\right). \tag{27}$$

In particular, for $m = 2$ we get

$$\begin{bmatrix} H_0(z) \\ H_1(z) \end{bmatrix} = \begin{bmatrix} G_0(z) & z^{-1}G_1(z) \\ G_1(z) & G_0(z) \end{bmatrix} \begin{bmatrix} C_0(z) \\ C_1(z) \end{bmatrix} = \begin{bmatrix} C_0(z) & z^{-1}C_1(z) \\ C_1(z) & C_0(z) \end{bmatrix} \begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} \tag{28}$$

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

$$\mathcal{T}_M(\boldsymbol{H}) = \mathcal{T}_M(\underline{\boldsymbol{G}})\mathcal{T}_{M+K-1}(\boldsymbol{C}), \tag{29}$$

where $\boldsymbol{C}$ is similar to $\boldsymbol{H}$ and

$$\underline{\boldsymbol{G}} = [\underline{\boldsymbol{g}}(K-1) \cdots \underline{\boldsymbol{g}}(0)],\; \underline{\boldsymbol{g}}(k) = \begin{bmatrix} g_0(k) & g_1(k-1) \\ g_1(k) & g_0(k) \end{bmatrix} \tag{30}$$

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

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 (see [1]). In that case we get $\boldsymbol{H}_i(z) = \underline{\boldsymbol{G}}_i(z)\boldsymbol{C}_i(z)$ for every antenna signal $i = 1, \ldots, q$ (where $\underline{\boldsymbol{G}}_i(z)$ may be independent of i) and $\boldsymbol{H}(z) = [\boldsymbol{H}_1^{\mathrm{T}}(z) \cdots \boldsymbol{H}_q^{\mathrm{T}}(z)]^{\mathrm{T}} = \mathrm{blockdiag}\{\underline{\boldsymbol{G}}_1(z) \cdots \underline{\boldsymbol{G}}_q(z)\}\boldsymbol{C}(z)$ where now $\boldsymbol{H}(z)$ and $\boldsymbol{C}(z)$ regroup mq channels.

5. Blind methods WTXFK

Prior TX/RX filter knowledge gets exploited by expressing $\boldsymbol{h} = \mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})\boldsymbol{c}$ and searching for $\boldsymbol{c}$. Since all five deterministic methods discussed above are of the form $\min_{\|\boldsymbol{h}\|=1} \boldsymbol{h}^{\mathrm{H}} A \boldsymbol{h}$, we get $\min_{\boldsymbol{c}} \boldsymbol{c}^{\mathrm{H}} \mathscr{T}^*(\underline{\boldsymbol{G}}^{\mathrm{t}}) A \mathscr{T}^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})\boldsymbol{c}$. In all methods except SRM, we can use $\|\boldsymbol{c}\| = 1$ 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 [27] is to translate $\|\boldsymbol{h}\|^2 = 1$ into the constraint $\boldsymbol{c}^{\mathrm{H}} \mathscr{T}^*(\underline{\boldsymbol{G}}^{\mathrm{t}}) \mathscr{T}^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})\boldsymbol{c} = 1$ which leads to a generalized eigenvalue problem that can alternatively be transformed into a regular eigenvalue 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 $\boldsymbol{H}^{\perp}$, the contribution of the noise to EA is a multiple of identity, whereas the contribution of the signal is singular. Hence, the noise contribution can be removed by considering $\min_{\|\boldsymbol{c}\|=1} \boldsymbol{c}^{\mathrm{H}} \mathscr{T}^*(\underline{\boldsymbol{G}}^{\mathrm{t}})(A - \lambda_{\min}(A)\, I) \mathscr{T}^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})\boldsymbol{c}$. For GML and covariance matching, one needs to reparameterize $\boldsymbol{h}$ in terms of $\boldsymbol{c}$.

6. Identifiability issues

Based on [4], the identifiability conditions when considering the overall channel were established: the subchannels must not share common zeros [29,30]. An interpretation of this identifiability condition is the following: if the subchannels have zeros in common, these terms can be factorized and the representation of the overall channel will introduce a monochannel (corresponding to the factorized terms) followed by a multichannel representation. For the deterministic symbol model, the monochannel cannot be separated from the input sequence. For the Gaussian symbol model, the monochannel aspect makes SOS insufficient to identify the channel (the factorized terms are not minimum phase in general).

Here, we discuss the identifiability issues when the estimation exploits the TX filter knowledge. It is clear that if either the unknown part or the known part of the channel is unidentifiable, or the product of the two corresponding transfer functions gives common zeros, the overall channel is unidentifiable for methods without prior knowledge. Taking into acount the prior knowledge, the blind estimation method can identify the channel even when the transmission filters are unidentifiable and the propagation channel is identifiable. To explain this, consider the case of two channels ($m = 2$). The subchannels $H_1(z)$ and $H_2(z)$ satisfy the identifiability conditions means that there exist two polynomials $F_1(z)$ and $F_2(z)$ such that

$$F_1(z)H_1(z) + F_2(z)H_2(z) = 1. \tag{31}$$

Replacing the expressions of $H_1(z)$ and $H_2(z)$ given by (28) in (31), gives

$$\begin{aligned} &F_0(z)(G_0(z)C_0(z) + z^{-1}G_1(z)C_1(z)) \\ &\quad + F_1(z)(G_1(z)C_0(z) + G_0(z)C_1(z)) = 1. \end{aligned} \tag{32}$$

This can be rewritten as follows:

$$\begin{aligned} &C_0(z)(F_0(z)G_0(z) + F_1(z)G_1(z)) \\ &\quad + C_1(z)(z^{-1}F_0(z)G_1(z) + F_1(z)G_0(z)) = 1. \end{aligned} \tag{33}$$

Equality (33) implies that there exist two polynomials $Q_1(z) = F_0(z)G_0(z) + F_1(z)G_1(z)$ and $Q_2(z) = z^{-1}F_0(z)G_1(z) + F_1(z)G_0(z)$ such that $Q_1(z)C_1(z) + Q_2(z)C_2(z) = 1$ which is the identifiability condition for the propagation channel ($C_1(z)$ and $C_2(z)$ are coprime). Observe that no conditions are required on $G_1(z)$ and $G_2(z)$ in (33), which means that even in the case where $G(z)$ is not identifiable, we can identify the channel.

7. Cramer–Rao bounds WTXFK

7.1. *Unconstrained deterministic Fisher information matrix (FIM)*

In [26], the deterministic FIM for the estimation of $\boldsymbol{h}$ from (2) was derived:

$$J(\boldsymbol{h}) = \sigma_v^{-2} \mathscr{A}_{M,N}^{\mathrm{H}} P_{\mathscr{T}(\boldsymbol{H})}^{\perp} \mathscr{A}_{M,N}, \tag{34}$$

where $\mathscr{A}_{M,N} = A_{M,N} \otimes I_m$ is such that $\mathscr{A}_{M,N}\boldsymbol{h} = \mathscr{T}_M(\boldsymbol{H})A$. The matrix $A_{M,N}$ is a Hankel matrix defined as

$$A_{M,N} = \begin{bmatrix} a(-N+1) & a(-N+2) & \cdots & a(0) \\ a(-N+2) & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & \vdots \\ a(M-N) & \cdots & \cdots & a(M-1) \end{bmatrix}. \quad (35)$$

7.2. Unconstrained Gaussian FIM

For the Gaussian case, things are a bit more intricate. Consider the estimation problem for $\boldsymbol{h}$. We have $\boldsymbol{Y} \sim \mathcal{N}(\mathscr{T}_M(\boldsymbol{H})A, R_{YY})$, $R_{YY} = \sigma_a^2 \mathscr{T}_M(\boldsymbol{H})\mathscr{T}_M^{\mathrm{H}}(\boldsymbol{H}) + \sigma_v^2 I$. The corresponding probability density function is

$$f(\boldsymbol{Y}/\boldsymbol{h}) = \frac{1}{\pi^{mM} \det R_{YY}} \mathrm{e}^{-(\boldsymbol{Y}-\mathscr{T}_M(\boldsymbol{H})A)^{\mathrm{H}} R_{YY}^{-1} (\boldsymbol{Y}-\mathscr{T}_M(\boldsymbol{H})A)}. \quad (36)$$

Let the Fisher information matrices (FIM) $J_{\varphi\psi}$ be defined as

$$J_{\varphi\psi} = E_{\boldsymbol{Y}/\boldsymbol{h}} \left(\frac{\partial \ln f(\boldsymbol{Y}/\boldsymbol{h})}{\partial \varphi^*} \right) \left(\frac{\partial \ln f(\boldsymbol{Y}/\boldsymbol{h})}{\partial \psi^*} \right)^{\mathrm{H}} \quad (37)$$

and we will consider $J_{\boldsymbol{hh}}$ and $J_{\boldsymbol{hh}^*}$. In the deterministic case, $J_{\boldsymbol{hh}^*} = 0$. In that case, $J_{\boldsymbol{hh}}$ can be considered as a complex FIM, and $C_{\tilde{\boldsymbol{h}}t} \geqslant J_{\boldsymbol{hh}}^+$, the complex CRB. If $J_{\boldsymbol{hh}^*} \neq 0$ as in the Gaussian case, $J_{\boldsymbol{hh}}^+$ is also a bound on $C_{\tilde{\boldsymbol{h}}}$, but not as tight as the actual CRB which we obtain by considering $\boldsymbol{h}_{\mathrm{R}} = [\mathrm{Re}(\boldsymbol{h})^{\mathrm{T}}\, \mathrm{Im}(\boldsymbol{h})^{\mathrm{T}}]^{\mathrm{T}}$, the associated real parameters. We get

$$J_{\mathrm{R}}(\boldsymbol{h}_{\mathrm{R}}) = 2\begin{bmatrix} \mathrm{Re}(J_{\boldsymbol{hh}}) & -\mathrm{Im}(J_{\boldsymbol{hh}}) \\ \mathrm{Im}(J_{\boldsymbol{hh}}) & \mathrm{Re}(J_{\boldsymbol{hh}}) \end{bmatrix} + 2\begin{bmatrix} \mathrm{Re}(J_{\boldsymbol{hh}^*}) & -\mathrm{Im}(J_{\boldsymbol{hh}^*}) \\ \mathrm{Im}(J_{\boldsymbol{hh}^*}) & \mathrm{Re}(J_{\boldsymbol{hh}^*}) \end{bmatrix} \quad (38)$$

and

$$J_{\boldsymbol{h}_i\boldsymbol{h}_j} = \mathrm{tr}\left\{ R_{YY}^{-1} \left(\frac{\partial R_{YY}}{\partial \boldsymbol{h}_i^*} \right) R_{YY}^{-1} \left(\frac{\partial R_{YY}}{\partial \boldsymbol{h}_j^*} \right)^{\mathrm{H}} \right\}, \quad (39)$$

$$J_{\boldsymbol{h}_i\boldsymbol{h}_j^*} = \mathrm{tr}\left\{ R_{YY}^{-1} \left(\frac{\partial R_{YY}}{\partial \boldsymbol{h}_i^*} \right) R_{YY}^{-1} \left(\frac{\partial R_{YY}}{\partial \boldsymbol{h}_j^*} \right) \right\}, \quad (40)$$

$$\frac{\partial R_{YY}}{\partial \boldsymbol{h}_i^*} = \sigma_a^2 \mathscr{T}_M(\boldsymbol{H}) \mathscr{T}_M^{\mathrm{H}} \left(\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{h}_i} \right). \quad (41)$$

7.3. Cramer–Rao bounds WTXFK

The unconstrained FIM or CRB can easily be transformed into the constrained CRB WTXFK. Since $\boldsymbol{h} = \mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})\boldsymbol{c}$, the $\boldsymbol{h}$ WTXFK satisfies the constraint $P^{\perp}_{\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})}\boldsymbol{h} = 0$. This leads to the CRB for the unbiased estimation error $\tilde{\boldsymbol{h}} = \boldsymbol{h} - \hat{\boldsymbol{h}}$ WTXFK [1,28,6]

$$C_{\tilde{\boldsymbol{h}}} = E\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^{\mathrm{H}} \geqslant [P_{\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})} J(\boldsymbol{h}) P_{\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})}]^+, \quad (42)$$

where superscript + denotes Moore–Penrose pseudo-inverse. The fact that the pseudo-inverse is used means that the component(s) of $\boldsymbol{h}$ in the null space of $P_{\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})} J(\boldsymbol{h}) P_{\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})}$ are known. This means first of all that the components of $\boldsymbol{h}$ outside of the column space of $\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})$ are known to be zero (due to the constraint). Note, furthermore, that $\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})$ may not be full column rank (if the sampling rate is higher than the Nyquist rate). Also, another singularity appears which is inherent in blind channel estimation. For the deterministic method, the channel can only be estimated up to a scale factor. We can adjust the scale factor by taking as final channel estimate $\hat{\hat{\boldsymbol{h}}} = \alpha\hat{\boldsymbol{h}}$ where α is obtained from $\min_\alpha \|\boldsymbol{h} - \alpha\hat{\boldsymbol{h}}\|$.

For the Gaussian symbol model, we need to work with the associated real parameters, which can be represented in the following vector form from $\boldsymbol{h} = \mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})\boldsymbol{c}$,

$$\boldsymbol{h}_{\mathrm{R}} = \begin{bmatrix} \mathrm{Re}(\boldsymbol{h}) \\ \mathrm{Im}(\boldsymbol{h}) \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})) & -\mathrm{Im}(\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})) \\ \mathrm{Im}(\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})) & \mathrm{Re}(\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})) \end{bmatrix} \boldsymbol{c}_{\mathrm{R}} = G_{\mathrm{R}}\, \boldsymbol{c}_{\mathrm{R}}. \quad (43)$$

The Gaussian CRB WTXFK is then the same as in (42) but with $\boldsymbol{h}$ and $\mathscr{T}_L^{\mathrm{T}}(\underline{\boldsymbol{G}}^{\mathrm{t}})$ replaced by $\boldsymbol{h}_{\mathrm{R}}$ and G_{R}. The estimation indeterminacy in the Gaussian case corresponds to a phase factor, $\hat{\hat{\boldsymbol{h}}} = \mathrm{e}^{\mathrm{j}\phi}\hat{\boldsymbol{h}}$, which can be adjusted by requiring that $\boldsymbol{h}^{\mathrm{H}}\hat{\hat{\boldsymbol{h}}}$ be real and positive.

Apart from a proper interpretation for the pseudo-inverse, these results for the Gaussian case were derived independently in [27], where some examples show that the Gaussian assumption improves the estimation quality considerably in certain cases.

8. Simulation results

In Fig. 2, the performance of the estimation of $\boldsymbol{h}$ for SRM and SRM WTXFK are compared to the corresponding deterministic CRB and CRB WTXFK. The data frame length is $M = 162$, oversampling factor $m = 2$ and the symbols are i.i.d. BPSK. The overall channel, presented in Fig. 3, is the convolution of a raised cosine pulse limited to $13T$ with rolloff factor $\alpha = 0.9$, and a two-ray multipath channel $c(t) = \delta(t) - 0.82\,\delta(t - T)$ (without exploitation of the TX filter, the overall channel is unidentifiable in this case [30]). The performance measure is the normalized MSE (NMSE) which is averaged over 100 Monte-Carlo runs:

$$\mathrm{NMSE} = \frac{1}{100}\sum_{i=1}^{100} \boldsymbol{h}^{\mathrm{H}} P^{\perp}_{\hat{\boldsymbol{h}}^{(i)}} \boldsymbol{h} / \|\boldsymbol{h}\|^2,$$

$$\text{where } \boldsymbol{h}^{\mathrm{H}} P^{\perp}_{\hat{\boldsymbol{h}}} \boldsymbol{h} = \min_{\alpha} \|\alpha\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2. \tag{44}$$

The signal-to-noise ratio (SNR) is defined as:

$$\mathrm{SNR} = \frac{\sigma_a^2 \|\boldsymbol{h}\|^2}{\sigma_v^2}. \tag{45}$$

The deterministic CRBs are normalized and computed as $\mathrm{tr}\{\mathrm{CRB}_{\boldsymbol{h}}\}/\|\boldsymbol{h}\|^2$. Our simulation results show that in terms of CRB, the approach WTXFK outperforms the one without this prior information. The difference between SRM and SRM WTXFK is even more spectacular: SRM on the complete channel suffers from channel zeros that are almost in common, whereas SRM WTXFK performs well. In order to illustrate that the exploitation of transmission filter knowledge does not simply result in a lower dimensional (LD) problem, we consider the performance of the SRM method weighted with the ratio of the number of estimated parameters: L/N. The obtained performance (curve corresponding to LD SRM) does not reach the performance of SRM WTXFK. Indeed, apart from the reduction of the number of parameters to be estimated, exploiting the transmission filter knowledge improves considerably the conditioning of the matrix in the optimization criterion. These two elements act simultaneously to improve the quality of the channel estimates. The same simulations (SSF, SSF WTXFK and LD SSF) were performed for the SSF technique, the results illustrated in

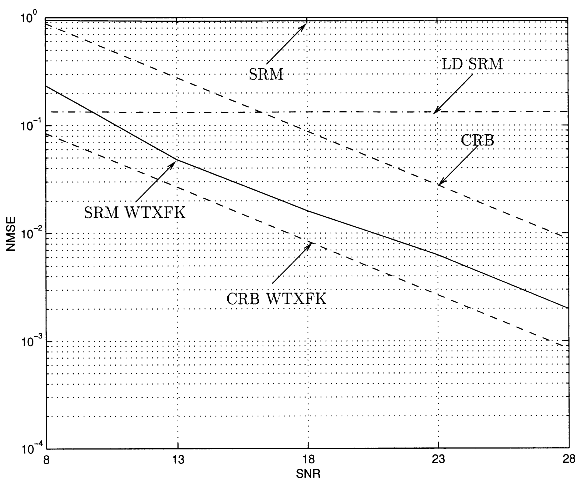

Fig. 2. Performance of SRM, LD SRM and SRM WTXFK.

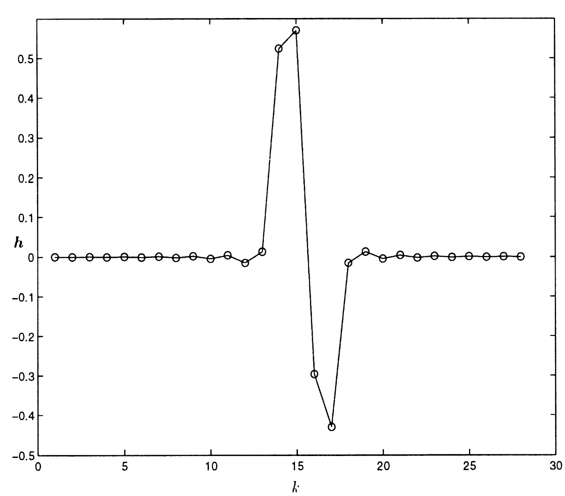

Fig. 3. The overall channel.

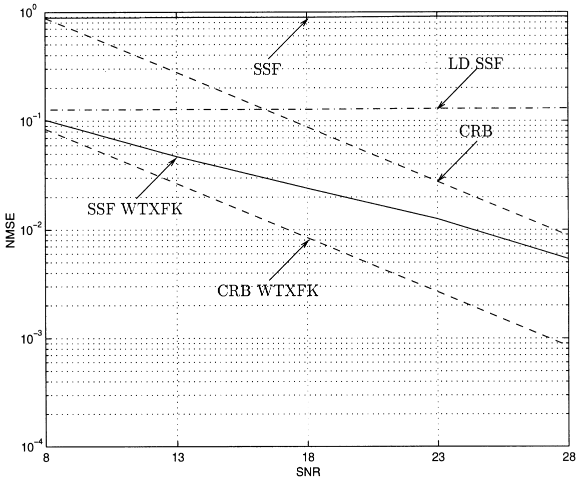

Fig. 4. Performance of SSF, LD SSF and SSF WTXFK.

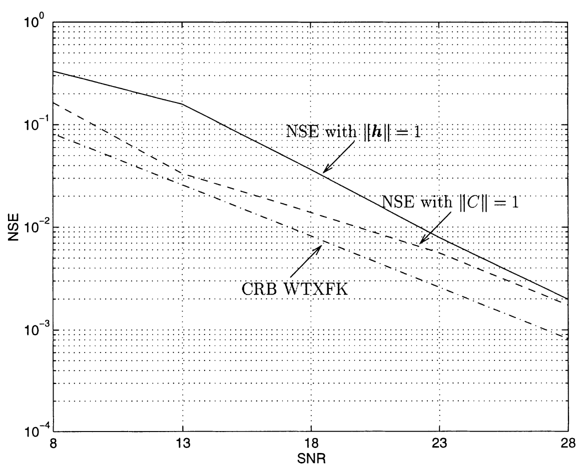

Fig. 5. Comparison of SRM WTXFK with $\|c\| = 1$ and $\|h\| = 1$.

Fig. 4 lead to analogous conclusions as those noted for the SRM method.

In Fig. 5, we used the same data and we compare the two unbiased forms of SRM WTXFK: the one using $\|h\| = 1$ and the one using $\|c\| = 1$ but with $A - \lambda_{\min}(A)I$: it is clear that the second approach outperforms the first one (by a factor of more than 5 at SNR = 13 dB).

In the simulation illustrated in Fig. 6, the idea is to study the behavior of the SRM WTXFK and IQML WTXFK (one iteration initialized with

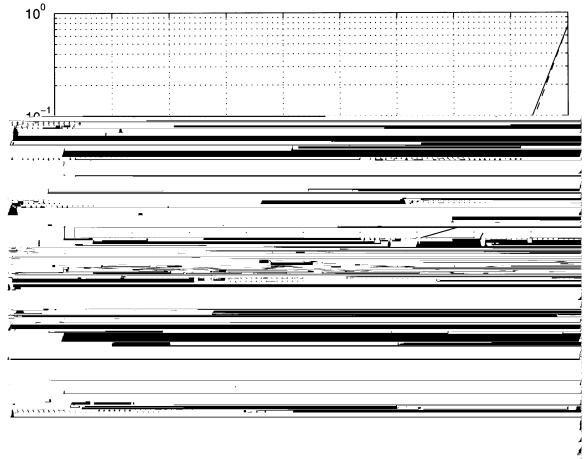

Fig. 6. Comparison of SRM WTXFK and IQML WTXFK.

SRM WTXFK) methods versus the conditioning of the propagation channel $\boldsymbol{c}$. We adopt the same pulse-shaping filter as before and we consider a propagation channel defined as

$$C = \begin{bmatrix} 1 & 1 \\ 1 & a \end{bmatrix}. \tag{46}$$

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 = 33 dB, and for values of a ranging from -1 to 0.8, show that 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.

9. Conclusions

Second-order blind channel estimation methods can give significantly improved performance by exploiting the knowledge of the TX and/or the RX filters: this introduces the concept of *blind channel estimation with side information*. A certain class of these blind methods: methods that are directly parameterized by the channel, can easily be extended to handle this side information. The extended methods outperform the original ones on two levels (our simulations were performed for SRM and SSF techniques but the conclusions hold for all the previously described blind multichannel estimation methods). The first one concerns the performance bounds (CRBs): the CRB corresponding to methods WTXFK can be orders of magnitude lower than the CRB corresponding to methods without the side information. The second level can be seen in terms of the used method itself: methods WTXFK give NMSE lower than the classical methods, and close to the CRB. An important issue that characterizes methods WTXFK is the capability to identify "difficult" channels (channels having almost common zeros). Notice that, apart from the improvement of the conditioning of the matrix of the criterion to be optimized, methods WTXFK reduce the number of parameters to be estimated and consequently often the computational complexity of the considered method.

References

[1] J. Ayadi, D.T.M. Slock, Blind multichannel identification in the stationary colored noise case, Proceedings of the Nineth IEEE Signal Processing Workshop on Statistical Signal and Array Processing: SSAP 98, Portland, OR, USA, September 1998.

[2] L.A. Baccala, S. Roy, A new blind time-domain channel identification method based on cyclostationarity, IEEE Signal Process. Lett. 1 (6) (June 1994) 89–91.

[3] A. Benveniste, M. Goursat, Blind equalizers, IEEE Trans. Comm. 32 (8) (August 1984) 871–883.

[4] R.R. Bitmead, S.Y. Kung, B.D.O. Anderson, T. Kailath, Greatest common divisers via generalized Sylvester and Bezout matrices, IEEE Trans. Automat. Control 23 (6) (December 1978) 1943–1947.

[5] E. de Carvalho, D.T.M. Slock, Cramer–Rao bounds for semi-blind, blind and training sequence based channel estimation, Proceedings of the SPAWC 97 Conference, Paris, France, April 1997.

[6] E. de Carvalho, D.T.M. Slock, Cramér–Rao bounds for blind multichannel estimation. IEEE Trans. Inform. Theory (2000), in preparation.

[7] Z. Ding, Z. Mao, Knowledge based identification of fractionnally sampled channels, Proceedings of the ICASSP 95 Conference, Detroit, USA, May 1995.

[8] G. Giannakis, S.D. Halford, Asymptotically optimal blind fractionally spaced channel estimation and performance analysis, IEEE Trans. Signal Process. 45 (7) (July 1997) 1815–1830.

[9] G.B. Giannakis, A. Swami, On estimating noncausal nonminimum phase ARMA models of non-Gaussian processes, IEEE Trans. Acoust. Speech Signal Process. 38 (3) (March 1990) 478–495.

[10] D.N. Godard, Self-recovering equalization and carrier tracking in two-dimensional data communication systems, IEEE Trans. Commun. 28 (11) (November 1984) 1867–1875.

[11] M.I. Gurelli, C.L. Nikias, EVAM: an eigenvector-based algorithm for multichannel blind deconvolution of input colored signals, IEEE Trans. Signal Process. 43 (1) (January 1995) 134–149.

[12] Y. Hua, Fast maximum likelihood for blind identification of multiple FIR channels, Proceedings of the 28th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October 31–November 2, 1994, pp. 415–419.

[13] Y. Hua, Fast maximum likelihood for blind identification of multiple FIR channels, IEEE Trans. Signal Process. 44 (3) (March 1996) 661–672.

[14] M. Kristensson, B.E. Ottersten, D.T.M Slock, Blind subspace identification of a BPSK communication channel, Proceedings of the 30th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, November 1996.

[15] H. Liu, G. Xu, A deterministic approach to blind symbol estimation, Proceedings of the 28th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October 31–November 2, 1994, pp. 1164–1168.

[16] H. Liu, G. Xu, A deterministic approach to blind symbol estimation, IEEE Signal Process. Lett. (December 1994) 205–207.

[17] G. Xu, H. Liu, L. Tong, T. Kailath, A least-squares approach to blind channel identification, IEEE Trans. Signal Process. 43 (12) (December 1995) 2982–2993.

[18] E. Moulines, P. Duhamel, J.-F. Cardoso, Subspace methods for the blind identification of multichannel FIR filters, Proceedings of ICASSP 94 Conference, Adelaide, Australia, April 1994.

[19] E. Moulines, P. Duhamel, J.-F. Cardoso, S. Mayrargue, Subspace methods for the blind identification of multichannel FIR filters, IEEE Trans. Signal Process. 43 (2) (February 1995) 516–526.

[20] Y. Sato, A method of self-recovering equalization for multilevel amplitude modulation systems, IEEE Trans. Comm. 23 (6) (June 1975) 679–682.

[21] S.V. Schell, D.L. Smith, W. Gardner, Blind channel identification using 2nd-order cyclostationary statistics, Proceedings of the EUSIPCO, Edinburgh, September 1994.

[22] D.T.M. Slock, Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction, Proceedings of the ICASSP 94 Conference, Adelaide, Australia, April 1994.

[23] D.T.M. Slock, Subspace techniques in blind mobile radio channel identification and equalization using fractional spacing and/or multiple antennas, Proceedings of the Third International Workshop on SVD and Signal Processing, Leuven, Belgium, August 22–25, 1994.

[24] D.T.M. Slock, Blind joint equalization of multiple synchronous mobile users using oversampling and/or multiple antennas, Proceedings of the 28th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October 31–November 2, 1994, pp. 1154–1158.

[25] D.T.M. Slock, Blind joint equalization of multiple synchronous mobile users for spatial division multiple access, in: E. Biglieri, M. Luise (Eds.), Signal Processing in Telecommunications, Proceedings of the Seventh Tyrrhenian International Workshop on Digital Communications, Springer, Berlin, 1995.

[26] D.T.M. Slock, C.B. Papadias, Blind fractionally-spaced equalization based on cyclostationarity, Proceedings of the Vehicular Technology Conference, Stockholm, Sweden, June 1994, pp. 1286–1290.

[27] D.L. Smith, Blind channel identification and equalization using second-order cyclostationarity, Ph.D. Thesis, Pennsylvania State University, August 1996.

[28] P. Stoica, B.C. Ng, On the Cramer–Rao bound under parametric constraints, IEEE Signal Process. Lett. 5 (7) (July 1998) 177–179.

[29] L. Tong, G. Xu, T. Kailath, A new approach to blind identification and equalization of multipath channels, Proceedings of the 25th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, November 1991, pp. 856–860.

[30] L. Tong, G. Xu, T. Kailath, Blind identification and equalization based on second order statistics: a time-domain approach, IEEE Trans. Inform. Theory 40 (2) (March 1994) 340–349.