Subspace-based blind channel identification of SISO-FIR systems with improper random inputs

Donatella Darsena, Giacinto Gelli, Luigi Paura and Francesco Verde *

Abstract

In this paper, we consider the problem of blindly estimating the impulse response of a nonminimum phase single-input single-output (SISO) channel, by resorting only to the second-order statistics (SOS) of the channel output. On the basis of a general treatment of the problem, we show that a SOS-based subspace procedure can be applied to the problem at hand if the transmitted signal is an improper process, exhibiting some additional properties. After characterizing and discussing these properties, we show that they are exhibited by many improper digital modulation schemes of practical interest. Moreover, based on our unifying framework, we devise subspace-based algorithms for blind channel identification, by addressing in particular the related identifiability issues. Finally, the theoretical expression of the mean-squared error of the channel estimate is derived, and numerical simulations are carried out for its validation and for comparing the performance of the proposed algorithm with that of a conventional subspace-based method.

Keywords: Blind channel identification, conjugate cyclostationarity, improper random processes, perturbation analysis, single-input single-output systems, subspace methods.

1 Introduction

Blind channel identification and equalization techniques in digital communications avoid the transmission of training sequences and, thus, make more efficient use of the available bandwidth. Earlier works in blind channel identification consider a single-input single-output (SISO) model, obtained by sampling at the baud-rate the output of a linear time-invariant channel; in this case, it is well-known [1] that the time-invariant autocorrelation function of the discrete-time output does not contain enough information to identify nonminimum phase channels. For this reason, blind identification algorithms based on SISO models (see, e.g., [2] and references therein) explicitly or implicitly employ higher-order statistics (HOS) of the channel output. However, due to the large amount of data necessary to accurately estimate higher-order moments or cumulants, HOS-based algorithms suffer from a significant performance degradation when only a limited number of samples of the received signal is available (e.g., in packet-oriented transmission). It was recognized subsequently that, for a single-input

*This work was presented in part at the 6th Baiona Workshop on Signal Processing in Communications, Baiona (Spagna), pp. 109 – 114, September 2003, and it was supported by the MIUR 2002 Italian National Research Project. The authors are with Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni, Università degli Studi Federico II di Napoli, via Claudio 21, I-80125 Napoli, Italy. Tel.: +39-0817683154, Fax : +39-0817683149, E-mail: [darsena, paura, gelli, f.verde]@unina.it

multiple-output (SIMO) channel model, blind identification can be accomplished by resorting only to second-order statistics (SOS) of the channel outputs (see [1] for a comprehensive review). A SIMO model arises naturally when the receiver employs multiple sensors, or when the output of a SISO continuous-time channel is sampled at a rate higher than the baud rate, giving rise to multiple *virtual* subchannels. SIMO identification methods based on SOS have attracted a great research interest, because they are less data-consuming than HOS-based ones. Among them, the *subspace* approach originally proposed in [3] is perhaps the most popular one, since the channel is estimated as the solution of a *linear* systems of equations. Moreover, subspace-based methods are very robust against noise and perform channel identification under mild restrictions on subchannel zeros (i.e., the "no-common zeros" condition).

However, SIMO channel identification techniques based on oversampling cannot be used in all those situations where the excess bandwidth is limited [4]. On the other hand, the use of an antenna array in wireless communications is impractical in the downlink (from base station to mobile), while it can significantly increase the receiver cost in the uplink (from mobile to base station). Thus, some recent works have addressed the blind channel identification and equalization problem from a different point of view, by considering a baud-rate SISO model but attempting to exploit some additional information, so as to avoid the need to resort to oversampling or multiple sensors. In the *modulation-induced cyclostationarity* (MIC) approach [5] and [6], cyclostationarity is introduced in the transmitted data stream and, as a result, the autocorrelation function of the channel output turns out to be time-varying. This latter property is the key to develop blind channel identification and equalization algorithms, with no restrictions on the channel zeros and regardless of spectral properties of the additive stationary noise. A different approach, which is considered also in this paper, exploit rather the *improper* [8] nature of the transmitted signal. This choice is motivated by the fact that in many modulation formats of practical interest [9], such as ASK, differential BPSK (DBPSK), offset QPSK (OQPSK), offset QAM (OQAM), MSK and its variant Gaussian MSK (GMSK), the transmitted symbol sequence is an improper (possibly) *wide-sense conjugate cyclostationary* [10] process. When the information symbol sequence is an improper random process, complete characterization of the SOS of the channel output requires [8] the additional knowledge of its *conjugate correlation function* (see Section 2). In this case, both autocorrelation and conjugate autocorrelation functions of the received signal may contain sufficient information for blind identification of nonminimum phase SISO channels. Although from a slightly different perspective, the improper nature of the transmitted symbol sequence was perhaps first exploited for subspace-based blind channel identification in [11]. Indeed, with reference to a BPSK or MSK modulation format, it was recognized in [11] that, besides resorting to an antenna array or oversampling, a virtual SIMO model can be obtained from a SISO one also by separately processing the real and imaginary part of the received signal. Based on this observation, it was shown in [11] by simulation experiments that this virtual SIMO model allows one to improve estimation accuracy in subspace-based channel identification. A similar approach was independently followed in [12] with reference to GMSK modulation, where several SOS- and HOS-based channel identification and equalization techniques are proposed. Recently, an "analytical" method has been proposed [7] for SISO channel identification with improper sequences, which, starting from a set of samples of the autocorrelation and conjugate autocorrelation functions of the received signal in the absence of noise, builds a *nonlinear* system of equations, which can be solved for the unknown channel coefficients. However, the solution of such a system involves a quite complicated elimination procedure, which also requires a critical initialization. Moreover, when the statistics are estimated by using a short data record, the performances of the method are rather sensitive to the signal-to-noise ratio, and are difficult to characterize theoretically.

In this paper, we study the feasibility of performing *subspace-based* blind channel identification for a SISO channel model with an improper symbol sequence at its input. Unlike [11, 12], the blind identification algorithm is obtained in a general and unified framework, by linking its derivation to the basic theory of improper complex random vectors and processes [13]. This allows us to characterize the statistical properties that the symbol sequence must possess to allow for the use of a subspace-based method. In particular, we show that the channel impulse response cannot be estimated via a subspace-based algorithm even when the transmitted symbol sequence is an improper process. To allow for blind channel identification, the symbol sequence must be improper and satisfy a certain additional property: we refer to this subclass of improper random processes as *strongly improper* ones. By restricting attention to this family of processes, which however encompasses many digital modulation schemes of practical interest, channel identifiability conditions are derived in general, which are subsequently particularized and discussed in detail for the considered modulation schemes. Furthermore, the unified treatment considered in this paper allows us, unlike [7, 11, 12], to provide a detailed theoretical performance analysis of the proposed identification algorithm in a clear and concise manner.

The paper is organized as follows. In Section 2, we introduce notations and briefly recall some definitions about improper random processes that are used throughout the paper. In Section 3, the general system model, together with the basic assumptions, is introduced. The subspace-based algorithm for blind channel identification is presented in Section 4, and the identification conditions are derived. In Section 5, the application of the proposed method to some digital modulation schemes of practical interest is discussed. The theoretical performance analysis is carried out in Section 6, by using a first-order perturbative approach. In Section 7, comparative simulation results are presented to illustrate the performance of the algorithm and the validity of the theoretical analysis. Finally, conclusions are drawn in Section 8.

2 Notations and preliminaries

In the rest of the paper, we will use the following terminology and notations. Upper- and lower-case bold letters denote matrices and vectors; the superscripts $*$, T, H, -1 e $\dagger$ will denote the conjugate, the transpose, the hermitian (conjugate transpose), the inverse and the Moore-Penrose generalized inverse (pseudo-inverse) of a matrix; the subscripts R and I stand for real and imaginary parts of any complex-valued matrix, vector or scalar; $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{Z}$ are the fields of complex, real and integer numbers; $\mathbb{C}^n$ $[\mathbb{R}^n]$ denotes the vector-space of all n-column vectors with complex [real] coordinates; similarly, $\mathbb{C}^{n\times m}$ $[\mathbb{R}^{n\times m}]$ denotes the vector-space of all the $n\times m$ matrices with complex [real] elements; $\mathbf{0}_n$, $\mathbf{O}_{n\times m}$ and $\mathbf{I}_n$ denote the n-column zero vector, the $n\times m$ zero matrix and the $n\times n$ identity matrix; $\|\cdot\|$, $\det(\cdot)$ and $\text{rank}(\cdot)$ denote the Frobenius norm, the determinant and the rank of a matrix; $\otimes$ denotes the Kronecker product; for any $\boldsymbol{A}\in\mathbb{C}^{n\times m}$, $\mathcal{N}(\boldsymbol{A})$, $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{R}^{\perp}(\boldsymbol{A})$ denote the null space of $\boldsymbol{A}$, the column space of $\boldsymbol{A}$ and its orthogonal complement in $\mathbb{C}^n$; $\boldsymbol{A}=\text{diag}[\boldsymbol{A}_{11},\boldsymbol{A}_{22},\ldots,\boldsymbol{A}_{nn}]$ is the (block) diagonal matrix wherein $\{\boldsymbol{A}_{ii}\}_{i=1}^{n}$ are diagonal matrices; $\text{vec}(\boldsymbol{A})$ associates with any matrix $\boldsymbol{A}$ the vector obtained by stacking its columns; for any matrix $\boldsymbol{A}=[\boldsymbol{a}_0,\boldsymbol{a}_1,\ldots,\boldsymbol{a}_{L-1}]\in\mathbb{C}^{n\times L}$ and positive integer r, we define the *block-Toeplitz matrix*

$$\mathbf{T}_r(\boldsymbol{A})\triangleq\begin{bmatrix}\boldsymbol{a}_0 & \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_{L-1} & \mathbf{0}_n & \ldots & \mathbf{0}_n\\ \mathbf{0}_n & \boldsymbol{a}_0 & \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_{L-1} & \ldots & \mathbf{0}_n\\ \vdots & & \ddots & & \ddots & & \vdots\\ \mathbf{0}_n & \ldots & \mathbf{0}_n & \boldsymbol{a}_0 & \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_{L-1}\end{bmatrix}\in\mathbb{C}^{(nr)\times(L+r-1)}\,; \tag{1}$$

$\delta(\cdot)$ and $\text{E}[\cdot]$ denote the Kronecker delta and the statistical averaging operator; and, finally, for an arbitrary positive integer number r and $m\in\mathbb{Z}$, we denote with $\mathbf{\Gamma}_r(m)\in\mathbb{R}^{r\times r}$ the matrix whose $(k+1,h+1)$th entry is given by $\delta(m+h-k)$, for $k,h\in\{0,1,\ldots,r-1\}$.

Second-order moments of a zero-mean discrete-time complex random process $s(n)$ can be defined in terms of two functions: the autocorrelation function $R_{ss}(n,m)\triangleq\text{E}[s(n)\,s^*(n-m)]$ and the conjugate correlation function $R_{ss^*}(n,m)\triangleq\text{E}[s(n)\,s(n-m)]$. A complex random process $s(n)$ whose conjugate correlation function vanishes identically, i.e., $R_{ss^*}(n,m)\equiv 0$ for any $n,m\in\mathbb{Z}$, is called *proper*, otherwise it is said *improper* [8]. Although many complex processes arising in communications turn out to be proper (e.g., those obtained as complex envelopes of bandpass wide-sense stationary signals [8]), there are some interesting examples and applications of improper random processes (see [13] for a review). A great variety of random processes with applications to communication theory exhibit periodically or almost-periodically time-variant second-order moments (see [10] for a general treatment). A zero-mean random process $s(n)$ whose autocorrelation function is periodic in n with period $N_1 \geq 1$, i.e., $R_{ss}(n,m) = R_{ss}(n+N_1,m)$, is called *(wide-sense) cyclostationary* [10] of period N_1; similarly, if the conjugate correlation function is periodic in n with period $N_2 \geq 1$, i.e., $R_{ss^*}(n,m) = R_{ss}(n+N_2,m)$, the process $s(n)$ is called *(wide-sense) conjugate cyclostationary* [10] of period N_2. Observe that, when the previous properties hold for $N_1 = 1$ and/or $N_2 = 1$, the functions $R_{ss}(n,m)$ and/or $R_{ss^*}(n,m)$ do not depend on n. A proper process can be cyclostationary, whereas improper processes can be cyclostationary and/or conjugate cyclostationary.

3 The system model

Let us consider a digital communication system employing linear modulation with baud-rate $1/T_s$. The complex envelope of the received signal, after filtering, ideal carrier-frequency recovering and baud-rate sampling, can be expressed as

$$r(n) = \sum_{i=0}^{L_c-1} c(i)\, s(n-i) + w(n) \,, \tag{2}$$

where $s(n)$ is the sequence of the transmitted symbols, $c(n)$ denotes the *composite* impulse response (including transmitting filter, channel, receiving filter, and timing offset) of the linear time-invariant discrete-time signal channel, which is assumed to be a causal finite-impulse response (FIR) filter of order $L_c - 1 > 1$, and, finally, $w(n)$ represents additive noise at the output of the receiving filter.

The following assumptions will be considered throughout the paper:

A1) the symbol sequence $s(n)$ is a white zero-mean complex random process, whose autocorrelation function $R_{ss}(m) \stackrel{\triangle}{=} \mathrm{E}[s(n)\, s^*(n-m)] = \sigma_s^2\, \delta(m)$, with $\sigma_s^2 = \mathrm{E}[|s(n)|^2] > 0$, is independent of n; in addition, $s(n)$ is *improper* and *conjugate cyclostationary* of period N, with conjugate correlation function $R_{ss^*}(n,m) \stackrel{\triangle}{=} \mathrm{E}[s(n)\, s(n-m)] = \sigma_s^2\, f(n)\, \delta(m)$, where $f(n)$ is a *known* complex-valued periodic function of period N;

A2) $w(n)$ is a white zero-mean complex random process, statistically independent of $s(n)$, whose autocorrelation function $R_{ww}(m) \stackrel{\triangle}{=} \mathrm{E}[w(n)\, w^*(n-m)] = \sigma_w^2\, \delta(m)$, with $\sigma_w^2 = \mathrm{E}[|w(n)|^2]$, is independent of n; in addition, $w(n)$ is proper, i.e., $R_{ww^*}(n,m) \stackrel{\triangle}{=} \mathrm{E}[w(n)\, w(n-m)] \equiv 0$ for any $n, m \in \mathbb{Z}$;

A3) $c(n)$ is a *complex-valued* channel, that is, neither $c_R(n)$ nor $c_I(n)$ vanish identically; moreover, without loss of generality, it is assumed that $c(0) \neq 0$ and $c(L_c - 1) \neq 0$.

A large number of digital modulation schemes of practical interest [9] satisfy assumption A1, including ASK, differential BPSK (DBPSK), offset QPSK (OQPSK), offset QAM (OQAM), MSK and its variant Gaussian

MSK (GMSK) (see Section 5 for a detailed exposition and discussion). Assumption A2 is surely fulfilled if the continuous-time filter used at the receiving side has (approximatively) a square root raised-cosine impulse response; more generally, A2 holds if a whitened matched-filter [9] is employed at the receiver.

4 Subspace-based blind channel identification

In this Section we discuss the feasibility of using subspace-based techniques for estimating the unknown channel impulse response, and we derive the channel identification technique in a general framework. More specifically, in the following, we will exploit the second-order statistical properties of the improper received signal $r(n)$ to derive a general subspace method, which allows one to blindly estimate the unknown channel vector $\boldsymbol{c} \triangleq [c(0), c(1), \ldots, c(L_c - 1)]^T \in \mathbb{C}^{L_c}$. To this aim, we collect L_e consecutive samples of $r(n)$ in the vector $\boldsymbol{r}(n) \triangleq [r(n), r(n-1), \ldots, r(n - L_e + 1)]^T \in \mathbb{C}^{L_e}$, which, accounting for (2), can be written as

$$\boldsymbol{r}(n) = \boldsymbol{C}\, \boldsymbol{s}(n) + \boldsymbol{w}(n) \, , \tag{3}$$

where $\boldsymbol{C} \triangleq \mathbf{T}_{L_e}(\boldsymbol{c}^T) \in \mathbb{C}^{L_e \times K}$ represents the wide ($L_e < K$) channel Toeplitz matrix associated with the channel vector $\boldsymbol{c}$, with $K \triangleq L_e + L_c - 1$, $\boldsymbol{s}(n) \triangleq [\, s(n),\, s(n-1),\, \ldots,\, s(n-K+1)\,]^T \in \mathbb{C}^K$ and $\boldsymbol{w}(n) \triangleq [w(n), w(n-1), \ldots, w(n - L_e + 1)]^T \in \mathbb{C}^{L_e}$. Observe that, except for the trivial case when the channel impulse response $c(n)$ vanishes identically (which is in contrast to assumption A3), the channel matrix $\boldsymbol{C}$ turns out to be structurally full-row rank, i.e., rank$(\boldsymbol{C}) = L_e$. To conveniently exploit the improper nature of the transmitted symbols, let us consider the *augmented vector* $\widetilde{\boldsymbol{r}}(n) \in \mathbb{C}^{2L_e}$ obtained by stacking $\boldsymbol{r}(n)$ and its complex conjugate $\boldsymbol{r}^*(n)$. Accounting for (3), such a vector can be expressed as

$$\widetilde{\boldsymbol{r}}(n) \triangleq \begin{bmatrix} \boldsymbol{r}(n) \\ \boldsymbol{r}^*(n) \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{C} & \mathbf{O}_{L_e \times K} \\ \mathbf{O}_{L_e \times K} & \boldsymbol{C}^* \end{bmatrix}}_{\boldsymbol{C}_a \in \mathbb{C}^{(2L_e) \times (2K)}} \underbrace{\begin{bmatrix} \boldsymbol{s}(n) \\ \boldsymbol{s}^*(n) \end{bmatrix}}_{\widetilde{\boldsymbol{s}}(n) \in \mathbb{C}^{2K}} + \underbrace{\begin{bmatrix} \boldsymbol{w}(n) \\ \boldsymbol{w}^*(n) \end{bmatrix}}_{\widetilde{\boldsymbol{w}}(n) \in \mathbb{C}^{2L_e}} . \tag{4}$$

Note that, due to its block diagonal structure and the fact that $\boldsymbol{C}$ is full-row rank, also the wide matrix $\boldsymbol{C}_a$ is full-row rank, i.e., rank$(\boldsymbol{C}_a) =$ rank$(\boldsymbol{C}) +$ rank$(\boldsymbol{C}^*) = 2\, L_e$.

The proposed channel identification method is based on the eigenvalue decomposition (EVD) of the autocorrelation matrix $\widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}(n) \triangleq \mathrm{E}[\widetilde{\boldsymbol{r}}(n)\, \widetilde{\boldsymbol{r}}^H(n)] \in \mathbb{C}^{(2L_e) \times (2L_e)}$ of the augmented vector $\widetilde{\boldsymbol{r}}(n)$. Accounting for (4) and invoking assumptions A1-A2, such a matrix can be written as

$$\widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}(n) = \boldsymbol{C}_a\, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)\, \boldsymbol{C}_a^H + \sigma_w^2\, \mathbf{I}_{2L_e} \, , \tag{5}$$

where $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \triangleq \mathrm{E}[\widetilde{\boldsymbol{s}}(n)\, \widetilde{\boldsymbol{s}}^H(n)] \in \mathbb{C}^{(2K) \times (2K)}$ denotes the autocorrelation matrix of $\widetilde{\boldsymbol{s}}(n)$, whose elements can be expressed in terms of $R_{ss}(m)$ and $R_{ss^*}(n, m)$. Since, by assumption A1, the conjugate correlation function

$R_{ss^*}(n,m)$ is periodic in n with period N, the matrix $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ is periodic in n with the same period N; hence it is completely specified by its values for $n \in \mathcal{M} \stackrel{\triangle}{=} \{0, 1, \ldots, N-1\}$. A subspace-based blind identification procedure can be derived from (5) only if the *signal-dependent part* $\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H \in \mathbb{C}^{(2L_e)\times(2L_e)}$ of the autocorrelation matrix $\widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}$ is singular, i.e., it results $D(n) \stackrel{\triangle}{=} \text{rank}[\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H] < 2\,L_e$, at least for one value of $n \in \mathcal{M}$; in this case, the signal contribution is confined into a $D(n)$–dimensional subspace of $\mathbb{C}^{2L_e}$. Thus, it is crucial to investigate whether the previous condition is verified; as a first step towards this end, we provide the following Lemma.

Lemma 1. The matrix $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ is singular for any $n \in \mathcal{M}$ if and only if $N \leq K$ and there exists at least one value $\overline{n} \in \mathcal{M}$ such that $|f(\overline{n})| = 1$.

Proof. By invoking assumption A1, $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ can be partitioned as follows

$$\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) = \sigma_s^2 \begin{bmatrix} \mathbf{I}_K & \boldsymbol{F}(n) \\ \boldsymbol{F}^*(n) & \mathbf{I}_K \end{bmatrix} , \tag{6}$$

where the matrix $\boldsymbol{F}(n) \stackrel{\triangle}{=} \text{diag}[f(n), f(n-1), \ldots, f(n-K+1)] \in \mathbb{C}^{K\times K}$ is periodically time-varying in n with period N. By using the formula for the determinant of a 2-by-2 partitioned matrix [14], one obtains

$$\det[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)] = \sigma_s^2 \det(\mathbf{I}_K) \cdot \det[\mathbf{I}_K - \boldsymbol{F}^*(n)\,\boldsymbol{F}(n)] = \sigma_s^2 \prod_{\ell=0}^{K-1} \left(1 - |f(n-\ell)|^2\right) . \tag{7}$$

Hence, for a given n, $\det[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)] = 0$ if and only if there exists at least one $\overline{n} \in \{n, n-1, \ldots, n-K+1\}$ such that $|f(\overline{n})| = 1$. Remembering that $f(n)$ is a periodic function, if we make the simplifying assumption that its period $N \leq K$, the previous condition is satisfied *for any* $n \in \mathcal{M}$ if and only if there exists at least one value $\overline{n} \in \mathcal{M}$ such that $|f(\overline{n})| = 1$. □

Singularity of $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$, discussed in Lemma 1, represents a *necessary* condition in order to derive a subspace-based channel identification procedure on the basis of (5). In fact, since the matrix $\boldsymbol{C}_a$ is full-row rank, if $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ is nonsingular, it results [14] that $\text{rank}[\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H] = \text{rank}(\boldsymbol{C}_a) = 2\,L_e$; in this case, subspace-based channel identification based on (5) is impossible, since the columns of the matrix $\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H$ span the whole observation space $\mathbb{C}^{2L_e}$. It is interesting to note that, according to Lemma 1, $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ is surely nonsingular when the symbol sequence $s(n)$ is proper, i.e., $R_{ss^*}(n,m) \equiv 0$ for any $n, m \in \mathbb{Z}$: in fact, in this case it results $f(n) \equiv 0$ for any value of $n \in \mathcal{M}$. Therefore, *subspace-based identification based on* (5) *is impossible when* $s(n)$ *is proper*, which motivates *a posteriori* our choice to discuss the case of an improper random process $s(n)$. However, using an arbitrary improper random sequence $s(n)$ does not assure that its associated matrix $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ is singular, since it might happen that the conditions of Lemma 1 are not verified, that is, $|f(n)| \neq 1$ for any $n \in \mathcal{M}$. On the other hand, singularity of $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ represents only a necessary, but not sufficient, condition in order for $\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H$ to be singular: indeed, since $\boldsymbol{C}_a$ is wide ($L_e < K$), the signal-dependent matrix $\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H$ can be nonsingular even when $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}$ is singular. Thus, the following Lemma provides a *sufficient* condition on matrix $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$, which assures the singularity of the signal-dependent matrix $\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H$ *regardless of the channel coefficients.*

Lemma 2. The matrix $\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H$ is singular for any $n \in \mathcal{M}$ if $\text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)] < 2\, L_e$ for any $n \in \mathcal{M}$. Moreover, let $M(n)$ denote the number of diagonal entries of $\boldsymbol{F}(n)$ that lie on the unit circle in the complex plane and assume that $f(n) \neq 0$ for any $n \in \mathcal{M}$, it results that $\text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)] < 2\, L_e$ for any $n \in \mathcal{M}$ if and only if $M(n) > 2\,(L_c - 1)$ for any $n \in \mathcal{M}$.

Proof. By using straightforward rank inequalities [14], it results that

$$\text{rank}[\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H] \leq \min\{2\, L_e, \text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H]\} = \text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H] \leq \min\{2\, L_e, \text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)]\} \tag{8}$$

and, thus, if $\text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)] < 2\, L_e$, it follows that $\text{rank}[\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H] \leq \text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)] < 2\, L_e$, for any $n \in \mathcal{M}$; thus $\boldsymbol{C}_a \, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) \, \boldsymbol{C}_a^H$ is singular. Moreover, since $f(n) \neq 0$ for any $n \in \mathcal{M}$, the matrix $\boldsymbol{F}(n)$ is nonsingular for any $n \in \mathcal{M}$ and, thus, accounting for (6), the rank of $\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)$ is equal to the rank of the following matrix

$$\sigma_s^{-2} \begin{bmatrix} \mathbf{I}_K & \mathbf{O}_{K \times K} \\ \mathbf{O}_{K \times K} & \boldsymbol{F}(n) \end{bmatrix} \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n) = \begin{bmatrix} \mathbf{I}_K & \boldsymbol{F}(n) \\ \boldsymbol{F}(n) \, \boldsymbol{F}^*(n) & \boldsymbol{F}(n) \end{bmatrix}, \tag{9}$$

where $\boldsymbol{F}(n) \, \boldsymbol{F}^*(n) = \text{diag}[|f(n)|^2, |f(n-1)|^2, \ldots, |f(n-K+1)|^2]$. Owing to the particular structure of the matrix at the right-hand side of equation (9), let $M(n)$ denote the number of diagonal entries of $\boldsymbol{F}(n)$ that lie on the unit circle in the complex plane, it is apparent that $\text{rank}[\widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)] = 2\, K - M(n) < 2\, L_e$ if and only if $M(n) > 2\, L_e - 2\, K = 2\,(L_c - 1)$, for any $n \in \mathcal{M}$. □

Observe that Lemma 2 requires that P values of the function $f(n)$, say $f(n_1), f(n_2), \ldots, f(n_P)$, lie on the unit circle in the complex plane, where $\mathcal{P} \triangleq \{n_1, n_2, \ldots, n_P\} \subseteq \mathcal{M}$ must be chosen such that $M(n) > 2\,(L_c - 1)$ for any $n \in \mathcal{M}$. To simplify matters, we focus hereinafter our attention to the case where $\mathcal{P} \equiv \mathcal{M}$, that is, we require that $|f(n)| = 1$, for any $n \in \mathcal{M}$; in this case, it results that $M(n) = K$ and, thus, the condition of Lemma 2 is satisfied if: **A4)** $L_e > L_c - 1$, which in the following will be assumed to hold. In this case, Lemma 2 assures that the subclass of improper random processes for which $|f(n)| = 1$ for any $n \in \mathcal{M}$ allows one to perform subspace-based blind channel identification on the basis of (5). Motivated by this observation, since $f(n)$ is periodic of period N, we further restrict somehow assumption A1 by giving the following definition:

Definition 1. A random process $s(n)$ satisfying A1 is called *strongly improper* if $|f(n)| = 1$ for any $n \in \mathcal{M}$.

Let us observe that, according to A1 and the above definition, an arbitrary random process $s(n)$ is strongly improper if and only if $|\mathrm{E}[s^2(n)]| = \sigma_s^2 |f(n)| = \mathrm{E}[|s(n)|^2]$, for any $n \in \mathbb{Z}$; in its turn, by invoking the Cauchy-Schwartz inequality, this relation holds if and only if

$$\mathrm{E}\{|s^*(n) - \alpha(n)\, s(n)|^2\} = 0\,, \quad \text{for any } n \in \mathbb{Z}\,, \tag{10}$$

where $\alpha(n)$ is an arbitrary complex-valued deterministic function of n: thus, we can equivalently say that the random process $s(n)$ is strongly improper if and only if there exists a *mean-square* linear dependence between $s(n)$ and its conjugate $s^*(n)$, for any $n \in \mathbb{Z}$.

For the sake of simplicity, in the following, we will restrict our attention to the case where the periodic function $f(n)$, satisfying the constraint $|f(n)| = 1$, is a simple complex exponential, that is, we assume that $f(n) = e^{j\,2\pi\beta n}$, where $\beta = p/N$ and $p \in \mathcal{M}$; the value $0 \leq \beta < 1$ is called the *conjugate cycle frequency* [10] of the process $s(n)$. It is interesting to note that, in this case, the rank of the signal-dependent matrix $\boldsymbol{C}_a\, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)\, \boldsymbol{C}_a^H$ is independent of n, i.e., $D(n) = D$, for any $n \in \mathcal{M}$. In fact, in this case it can be seen that the periodically time-varying matrix $\boldsymbol{F}(n)$ in (6) can be written as $\boldsymbol{F}(n) = e^{j\,2\pi\beta n}\, \boldsymbol{J}$, with $\boldsymbol{J} \stackrel{\triangle}{=} \mathrm{diag}[1, e^{-j\,2\pi\beta}, \ldots, e^{-j\,2\pi\beta(K-1)}] \in \mathbb{C}^{K\times K}$ and, thus, accounting for the definition (4) of $\boldsymbol{C}_a$, the autocorrelation matrix $\widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}(n)$ given by (5) can be rewritten, after straightforward algebraic manipulations, as

$$\widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}(n) = \sigma_s^2\, \boldsymbol{\Omega}(n)\, \boldsymbol{\Psi}(\boldsymbol{c})\, \boldsymbol{\Psi}^H(\boldsymbol{c})\, \boldsymbol{\Omega}^*(n) + \sigma_w^2\, \mathbf{I}_{2L_e}\,, \tag{11}$$

where $\boldsymbol{\Omega}(n) \stackrel{\triangle}{=} \mathrm{diag}[\mathbf{I}_{L_e}, e^{-j\,2\pi\beta n}\mathbf{I}_{L_e}] \in \mathbb{C}^{(2L_e)\times(2L_e)}$ is a known block diagonal matrix, whereas the matrix $\boldsymbol{\Psi}(\boldsymbol{c}) \stackrel{\triangle}{=} [\boldsymbol{C}^T, \boldsymbol{J}^*\, \boldsymbol{C}^H]^T \in \mathbb{C}^{(2L_e)\times K}$ depends on the (unknown) channel vector $\boldsymbol{c} \in \mathbb{C}^{L_c}$. Since the diagonal matrix $\boldsymbol{\Omega}(n)$ is nonsingular for any $n \in \mathcal{M}$, it turns out that $D \stackrel{\triangle}{=} \mathrm{rank}[\boldsymbol{C}_a\, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)\, \boldsymbol{C}_a^H] = \mathrm{rank}[\boldsymbol{\Psi}(\boldsymbol{c})] \leq K = L_e + L_c - 1$, where the last inequality accounts for assumption A4; moreover, since the matrix $\boldsymbol{C}$ is full-row rank, it results that $D \geq L_e$. Therefore, we conclude that the value of the rank of the signal-dependent matrix $\boldsymbol{C}_a\, \widetilde{\boldsymbol{R}}_{\boldsymbol{ss}}(n)\, \boldsymbol{C}_a^H$ belongs to $D \in \{L_e, L_e+1, \ldots, L_e+L_c-1\}$ for any $n \in \mathcal{M}$.

The main difficult in using (11) to blindly estimate the unknown channel vector $\boldsymbol{c}$ lies in the periodically time-varying nature of $\widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}(n)$, which complicates its estimation from the received data. To remove the variability with n, we multiply $\widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}(n)$ by the *known* matrices $\boldsymbol{\Omega}^*(n)$ and $\boldsymbol{\Omega}(n)$, obtaining thus

$$\boldsymbol{R}_{\boldsymbol{zz}} \stackrel{\triangle}{=} \boldsymbol{\Omega}^*(n)\, \widetilde{\boldsymbol{R}}_{\boldsymbol{rr}}(n)\, \boldsymbol{\Omega}(n) = \sigma_s^2\, \boldsymbol{\Psi}(\boldsymbol{c})\, \boldsymbol{\Psi}(\boldsymbol{c})^H + \sigma_w^2\, \mathbf{I}_{2L_e}\,, \tag{12}$$

where the time-invariant matrix $\boldsymbol{R}_{\boldsymbol{zz}}$ can be interpreted as the autocorrelation matrix of the vector $\boldsymbol{z}(n) \stackrel{\triangle}{=} \boldsymbol{\Omega}^*(n)\, \widetilde{\boldsymbol{r}}(n) = [\boldsymbol{r}^T(n), e^{j\,2\pi\beta n}\, \boldsymbol{r}^H(n)]^T \in \mathbb{C}^{2L_e}$, i.e., $\boldsymbol{R}_{\boldsymbol{zz}} = \mathrm{E}[\boldsymbol{z}(n)\, \boldsymbol{z}^H(n)]$. On the basis of (12), the channel vector $\boldsymbol{c}$ can be estimated by resorting to the EVD of $\boldsymbol{R_{zz}}$, which in its turn can be consistently estimated, via batch or adaptive algorithms, from the received data. Following this approach, let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{2L_e}$ denote the eigenvalues of the autocorrelation matrix $\boldsymbol{R_{zz}}$, it follows that $\lambda_i > \sigma_w^2$, for $i \in \{1, 2, \ldots, D\}$, and $\lambda_i = \sigma_w^2$, for $i \in \{D+1, D+2, \ldots, 2\,L_e\}$. Moreover, let $\boldsymbol{e}_i \in \mathbb{C}^{2L_e}$ represent the eigenvector associated with the ith eigenvalue λ_i, the EVD of $\boldsymbol{R_{zz}}$ is given by

$$\boldsymbol{R_{zz}} = \boldsymbol{E}_{\mathcal{S}}\,\boldsymbol{\Lambda}_{\mathcal{S}}\,\boldsymbol{E}_{\mathcal{S}}^H + \boldsymbol{E}_{\mathcal{N}}\,\boldsymbol{\Lambda}_{\mathcal{N}}\,\boldsymbol{E}_{\mathcal{N}}^H\,, \tag{13}$$

where $\boldsymbol{\Lambda}_{\mathcal{S}} \triangleq \text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_D] \in \mathbb{R}^{D \times D}$ and $\boldsymbol{\Lambda}_{\mathcal{N}} \triangleq \sigma_w^2\,\mathbf{I}_{2L_e-D}$, whereas the columns of the matrices $\boldsymbol{E}_{\mathcal{S}} \triangleq [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_D] \in \mathbb{C}^{(2L_e)\times D}$ and $\boldsymbol{E}_{\mathcal{N}} \triangleq [\boldsymbol{e}_{D+1}, \boldsymbol{e}_{D+2}, \ldots, \boldsymbol{e}_{2L_e}] \in \mathbb{C}^{(2L_e)\times(2L_e-D)}$ span the D-dimensional subspace $\mathcal{R}[\boldsymbol{\Psi}(\boldsymbol{c})]$ (referred to as the *signal subspace*) and the $(2L_e - D)$-dimensional orthogonal complement $\mathcal{R}^{\perp}[\boldsymbol{\Psi}(\boldsymbol{c})]$ (referred to as the *noise subspace*) in $\mathbb{C}^{2L_e}$ to the signal subspace. By exploiting the orthogonality between signal and noise subspaces, the following relation holds

$$\boldsymbol{\Psi}(\boldsymbol{c})\,\boldsymbol{\Psi}^H(\boldsymbol{c})\boldsymbol{E}_{\mathcal{N}} = \mathbf{O}_{(2L_e)\times(2L_e-D)}\,, \tag{14}$$

which can be regarded as a matrix equation in $\boldsymbol{c}$; under appropriate conditions detailed in Theorem 1, the solution of (14) turns out to be essentially unique. Since discussion of the rank of $\boldsymbol{\Psi}(\boldsymbol{c})$ plays a key role, we first provide the following Lemma.

Lemma 3. Let $\zeta_1, \zeta_2, \ldots, \zeta_{L_c-1}$ denote the zeros of the $\mathcal{Z}$-transform of the channel vector $\boldsymbol{c}$. Under assumption A4, the matrix $\boldsymbol{\Psi}(\boldsymbol{c})$ turns out to be full-column rank, i.e., $D = \text{rank}[\boldsymbol{\Psi}(\boldsymbol{c})] = K = L_e + L_c - 1$, if and only if the following assumption holds: **A5)** $\zeta_\ell \neq \zeta_q^*\,e^{j\,2\pi\beta}$, for $\ell, q \in \{1, 2, \ldots, L_c - 1\}$.

Proof. See Appendix A. □

We defer to Section 5 a thorough discussion about implications of Lemma 3 and assumption A5 in practical cases, with specific reference to the considered modulation schemes. The following identification Theorem provides sufficient conditions for unique channel identification.

Theorem 1 (Identifiability condition)**.** Let $\boldsymbol{\Pi}_{\mathcal{N}} \in \mathbb{C}^{R\times 2L_e}$ denote any matrix verifying $\mathcal{N}(\boldsymbol{\Pi}_{\mathcal{N}}) = \mathcal{R}^{\perp}(\boldsymbol{E}_{\mathcal{N}})$, and consider the matrix equation

$$\boldsymbol{\Pi}_{\mathcal{N}}\,\boldsymbol{\Psi}(\boldsymbol{c}) = \mathbf{O}_{R\times(L_e+L_c-1)}\,. \tag{15}$$

Under assumptions A4 and A5, the following two statements are equivalent: (1) $\boldsymbol{c}' \neq \boldsymbol{c}$ is a solution of (15); (2) $\boldsymbol{c}' = \xi\,\boldsymbol{c}$, with $\xi \in \mathbb{R}$.

Proof. See Appendix B. □

The above theorem asserts that, under assumptions A4 and A5, the channel vector $\boldsymbol{c}$ can be uniquely determined, up to a scalar factor, from the $(L_e - L_c + 1)$-dimensional noise subspace $\mathcal{R}(\boldsymbol{E}_\mathcal{N})$ of the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{zz}}$; note that, unlike standard subspace-based channel identification, the scalar factor is *real* rather than complex. As to the choice of the matrix $\boldsymbol{\Pi}_\mathcal{N}$, several options may be pursued: for example, one can choose $\boldsymbol{\Pi}_\mathcal{N} = \boldsymbol{E}_\mathcal{N}^H \in \mathbb{C}^{(L_e - L_c + 1) \times 2L_e}$ or $\boldsymbol{\Pi}_\mathcal{N} = \boldsymbol{E}_\mathcal{N}\, \boldsymbol{E}_\mathcal{N}^H \in \mathbb{C}^{2L_e \times 2L_e}$. Although the former choice is undoubtedly the most obvious one, the latter one simplifies the performance analysis of the channel estimation method for a finite sample size (this issue is explored in Section 6), and thus we adopt the second choice hereinafter.

By exploiting the structure of $\boldsymbol{\Psi}(\boldsymbol{c})$, equation (15) can be equivalently written as a system of *linear* equations in the unknown channel coefficients. To this aim, observe that (15) holds if and only if $\text{vec}[\boldsymbol{\Pi}_\mathcal{N}\, \boldsymbol{\Psi}(\boldsymbol{c})] = \boldsymbol{0}_{2KL_e}$, with $K = L_e + L_c - 1$, which can be written as

$$\text{vec}[\boldsymbol{\Pi}_\mathcal{N}\, \boldsymbol{\Psi}(\boldsymbol{c})] = \boldsymbol{Q}\, \boldsymbol{c}_a = \boldsymbol{0}_{2KL_e} \,, \tag{16}$$

where the *parameterization matrix* $\boldsymbol{Q} \in \mathbb{C}^{(2KL_e) \times (2L_c)}$ depends on $\boldsymbol{J}$ and $\boldsymbol{E}_\mathcal{N}$ (see Appendix C for its detailed expression), whereas the vector $\boldsymbol{c}_a \triangleq [\boldsymbol{c}_R^T, \boldsymbol{c}_I^T]^T \in \mathbb{R}^{2L_c}$ is obtained by stacking the real $\boldsymbol{c}_R$ and imaginary $\boldsymbol{c}_I$ parts of the channel vector $\boldsymbol{c}$. Note that, according to Theorem 1, the unknown vector $\boldsymbol{c}_a$ can be uniquely determined up to a real scalar factor and, hence, $\text{rank}(\boldsymbol{Q}) = 2\, L_c - 1$ must hold.

In practice, we do not know the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{zz}}$ and, thus, sample estimates of the eigenvectors spanning the signal and noise subspaces must be obtained from the sample autocorrelation matrix

$$\widehat{\boldsymbol{R}}_{\boldsymbol{zz}} \triangleq \frac{1}{N_s} \sum_{n=0}^{N_s - 1} \boldsymbol{z}(n) \boldsymbol{z}^H(n) = \widehat{\boldsymbol{E}}_\mathcal{S}\, \widehat{\boldsymbol{\Lambda}}_\mathcal{S}\, \widehat{\boldsymbol{E}}_\mathcal{S}^H + \widehat{\boldsymbol{E}}_\mathcal{N}\, \widehat{\boldsymbol{\Lambda}}_\mathcal{N}\, \widehat{\boldsymbol{E}}_\mathcal{N}^H \,, \tag{17}$$

where N_s denotes the estimation sample size. For the problem at hand, the matrix $\widehat{\boldsymbol{R}}_{\boldsymbol{zz}}$ converges (in mean square and with probability one) to $\boldsymbol{R}_{\boldsymbol{zz}}$ as N_s approaches infinity. Let $\widehat{\boldsymbol{Q}}$ denote the parameterization matrix corresponding to $\widehat{\boldsymbol{\Pi}}_\mathcal{N} = \widehat{\boldsymbol{E}}_\mathcal{N}\, \widehat{\boldsymbol{E}}_\mathcal{N}^H$, equality (16) holds only approximately, that is, $\boldsymbol{\epsilon} \triangleq \text{vec}[\widehat{\boldsymbol{\Pi}}_\mathcal{N}\, \boldsymbol{\Psi}(\boldsymbol{c})] = \widehat{\boldsymbol{Q}}\, \boldsymbol{c}_a \simeq \boldsymbol{0}_{2KL_e}$. In this case, an estimate $\widehat{\boldsymbol{c}}_a$ of the unknown vector $\boldsymbol{c}_a$ can be obtained by solving the constrained minimization problem

$$\widehat{\boldsymbol{c}}_a = \arg \min_{\boldsymbol{x} \in \mathbb{R}^{2L_c}} \boldsymbol{x}^H \widehat{\boldsymbol{Q}}^H \widehat{\boldsymbol{Q}}\, \boldsymbol{x} \,, \quad \text{subject to} \quad \|\boldsymbol{x}\|^2 = 1 \,, \tag{18}$$

where the constraint is necessary to avoid the trivial solution $\widehat{\boldsymbol{c}}_a = \boldsymbol{0}_{2L_c}$. By invoking the Rayleigh-Ritz theorem [14], the solution of the optimization problem (18) is given by the eigenvector[1] associated with the smallest

[1] Note that, by construction [see the orthogonality relation (16)], all the eigenvectors corresponding to the smallest real eigenvalue of the matrix $\boldsymbol{Q}^H \boldsymbol{Q}$ must be real-valued. This structural property remains valid even when we consider $\widehat{\boldsymbol{Q}}^H \widehat{\boldsymbol{Q}}$, which is constructed by using $\widehat{\boldsymbol{\Pi}}_\mathcal{N}$.

eigenvalue of the matrix $\widehat{\boldsymbol{Q}}^H \widehat{\boldsymbol{Q}}$. Finally, observe that, similarly to many subspace-based algorithms (e.g., [3]), the described algorithm requires accurate channel-length determination, which can be obtained by performing rank estimation of the sample autocorrelation matrix $\widehat{\boldsymbol{R}}_{\boldsymbol{zz}}$.[2]

5 Application to some digital modulation schemes

A large number of digital modulation schemes fall into the considered framework, since they satisfy assumption A1 and agree with Definition 1; in the following, we briefly recall some of them and particularize the general results obtained in the previous section, discussing some specific issues.

Real-valued modulations

In M-ary amplitude-shift keying (M-ASK) modulation [9], the symbols $s(n)$ are modeled as a sequence of independent and identically distributed (i.i.d.) real-valued random variables, each assuming equiprobable values in $\mathcal{S} = \{2\,p - M - 1\}_{p=1}^{M}$. When $M = 2$, the 2-ASK modulation coincides with the well-known binary phase-shift keying (BPSK) modulation. A variation of BPSK is the differential BPSK (DBPSK) modulation scheme [9], wherein the information is encoded in the phase differences between successive signal transmissions, as opposed to absolute phase encoding used in the BPSK modulation. For M-ASK, BPSK, and DBPSK, it can be easily verified that $R_{ss^*}(n,m) = R_{ss}(m) = \sigma_s^2\,\delta(m)$, with $\sigma_s^2 = (M^2-1)/3$ for M-ASK and $\sigma_s^2 = 1$ for BPSK and DBPSK. Thus, the symbol sequence $s(n)$ is a strongly improper random process with conjugate cycle frequency $\beta = 0$. For the above-mentioned modulation techniques, assumption A5 (Lemma 3) simplifies to $\zeta_\ell \neq \zeta_q^*$, for $\ell, q \in \{1, 2, \ldots, L_c - 1\}$, which means that, according to Theorem 1, the channel vector $\boldsymbol{c}$ can be uniquely identified if *its $\mathcal{Z}$-transform has no real zeros and no conjugate pairs of zeros*. Interestingly, this identifiability condition also assures the existence of a *widely linear* FIR zero-forcing (ZF) equalizer [16], which achieves perfect ISI suppression in the absence of noise even when the signal is sampled at the baud-rate; this result contrasts with simple linear equalization, where ZF baud-rate equalization of a FIR channel can be attained only by resorting to infinite-impulse response (IIR) filters. Finally, note that the authors in [7] claim that their method is able to identify the channel vector $\boldsymbol{c}$ if its $\mathcal{Z}$-transform has no real zeros. However, as shown in Appendix D, this claim is based on a not entirely correct derivation, whereas the correct identifiability conditions of the method proposed in [7] are the same as those derived in this paper.

MSK-type modulations

[2]Recently, subspace-based algorithms have been developed [15], which are based on rank-revealing ULV decomposition and do not require rank estimation. The feasibility of applying these algorithms to our framework is an interesting research issue, which however is outside the scope of this paper.

The minimum-shift keying (MSK) modulation [9] is a binary continuous-phase modulation (CPM) with index $h = 1/2$. An important example of MSK modulation is the Gaussian MSK (GMSK), which is currently employed in the Global System for Mobile (GSM) communications standard. All CPM modulations belong to the class of non-linear modulation schemes and, thus, the received MSK signal cannot be exactly represented by the linear model considered in (2). However, following the development given by Laurent (see, e.g., [9]), for an arbitrary modulation pulse of (approximately) *finite* duration $L\,T_s$, the continuous-time MSK signal can be represented as a linear superposition of 2^{L-1} amplitude-modulated signals, where the first term (*principal component*) contains the most significant part of the signal energy[3]. Thus, by considering only the principal component, after baud-rate sampling, the received MSK signal can be satisfactorily *approximated* by (2), where the symbol sequence is expressed recursively [9] as $s(n) = j\,a(n)\,s(n-1)$, with $a(n)$ being a sequence of i.i.d. real-valued random variables taking equiprobable values in $\mathcal{S} = \{-1, 1\}$. Assuming, without loss of generality, that $s(0) = 1$, this recursive equation can be explicitly solved for $s(n)$, obtaining

$$s(n) = j^{\,n} \prod_{p=1}^{n} a(p)\,, \quad \text{for} \quad n = 1, 2, 3, \ldots\,. \tag{19}$$

Taking into account (19), it can be easily seen that $R_{ss}(m) = \delta(m)$ and $R_{ss^*}(n, m) = (-1)^n\,\delta(m)$. Thus, the symbol sequence $s(n)$ is a strongly improper process with conjugate cycle frequency $\beta = 1/2$. In this case, assumption A5 simplifies to $\zeta_\ell \neq -\zeta_q^*$, for $\ell, q \in \{1, 2, \ldots, L_c - 1\}$, which means that, according to Theorem 1, the channel vector $\boldsymbol{c}$ can be uniquely identified if *its $\mathcal{Z}$-transform has no purely imaginary zeros and no anti-conjugate*[4] *pairs of zeros*. Note that a similar identifiability condition appears in [12] with reference to blind channel equalization for GSM cellular systems.

Offset modulations

The term *offset* or *staggered* modulation refers to any complex scheme wherein the quadrature component (Q) of the modulated signal is shifted in time with respect to the in-phase components (I). Common examples of this family are the *offset quaternary phase-shift keying* (OQPSK) modulation and the *offset quadrature amplitude modulation* (OQAM) [9], which are the offset versions of QPSK and QAM, respectively. The main advantage of offset modulations, with respect to their non-offset counterparts, is the increased bandwidth efficiency, which motivated their use in wireless [17, 18] and cable modem systems [19].

[3]For instance, with reference to the GMSK case, it results [12] that the principal component of the Laurent series expansion of the transmitted signal contains the 99.5% of the overall GMSK signal energy. In other words, the effect of approximating the GMSK signal with a single amplitude-modulated pulse is to virtually generate an additive interference in the received signal, with a signal-to-interference ratio (SIR) of about 23 dB.

[4]We say that $a, b \in \mathbb{C}$ are anti-conjugate if $a = -b^*$.

The OQPSK transmitted signal can be regarded as composed by two BPSK signals in-quadrature for which the phase transitions are staggered in time by T_s seconds. After baud-rate sampling, the received OQPSK signal can be represented by (2), where the symbol sequence $s(n)$ is given by

$$s(n) = \begin{cases} a(n)\,, & n \text{ even}\,, \\ j\,a(n)\,, & n \text{ odd}\,, \end{cases} \tag{20}$$

and $a(n)$ is a sequence of i.i.d. random variables assuming equiprobable values in $\mathcal{S} = \{-1, 1\}$.

Rectangular[5] M-ary quadrature amplitude modulation (QAM) with rate $1/T_s$, where $M = 2^q$ (q even), can be interpreted [9] as the effect of simultaneously impressing, on two quadrature carriers, two half-rate mutually independent real i.i.d. sequences $a_I(n)$ and $a_Q(n)$ assuming equiprobable values in $\mathcal{S} = \{2\,p - \sqrt{M} - 1\}_{p=1}^{\sqrt{M}}$ An M-OQAM signal is obtained from an M-QAM one, by introducing between the two quadrature components a time offset of one symbol period T_s. In this case, it can be seen [16] that, after baud-rate sampling, the received OQAM signal can be represented by the model (2), where

$$s(n) = \begin{cases} a_I(\ell)\,, & n = 2\,\ell\,, \\ j\,a_Q(\ell)\,, & n = 2\,\ell + 1\,. \end{cases} \tag{21}$$

Due to the close similarity between (20) and (21), both OQPSK and OQAM signals satisfy assumption A1 and Definition 1: indeed, it results that $R_{ss}(m) = \sigma_s^2\,\delta(m)$ and $R_{ss^*}(n, m) = \sigma_s^2\,(-1)^n\,\delta(m)$, with $\sigma_s^2 = 1$ for OQPSK and $\sigma_s^2 = (M-1)/3$ for M-OQAM. Thus, similarly to MSK, the information sequence $s(n)$ for OQPSK and OQAM transmissions is a strongly improper process with conjugate cycle frequency $\beta = 1/2$. Since the conjugate correlation function has essentially the same form of the MSK one, the identifiability conditions for OQPSK and OQAM are the same already discussed for MSK-type modulations.

6 Performance analysis

In this section, we provide the theoretical performance analysis of the channel identification method described in Section 4, by deriving the analytical expression of the estimation mean square error (MSE). The basic tool for our analysis is the powerful first-order perturbation approach (see, e.g., [20, 21]), which allows us to relate, under the large sample size assumption, the properties of $\widehat{\mathbf{\Pi}}_{\mathcal{N}} = \widehat{\boldsymbol{E}}_{\mathcal{N}}\,\widehat{\boldsymbol{E}}_{\mathcal{N}}^H$ to those of the sample autocorrelation matrix $\widehat{\boldsymbol{R}}_{\boldsymbol{zz}}$ of the preprocessed received vector $\boldsymbol{z}(n) = \mathbf{\Omega}^*(n)\,\widetilde{\boldsymbol{r}}(n)$. Our derivation holds in particular for the modulation schemes discussed in Section 5, and the theoretical results will be validated by computer simulation experiments in Section 7.

[5] Although more complicated QAM constellation geometries can be considered, rectangular constellations are most frequently used in practice since they can be easily generated and demodulated [9].

When the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{zz}}$ is estimated from the received data as in (17), for a sufficiently large sample-size N_s, the estimate can be decomposed as $\widehat{\boldsymbol{R}}_{\boldsymbol{zz}} = \boldsymbol{R}_{\boldsymbol{zz}} + \boldsymbol{\delta R}_{\boldsymbol{zz}}$, where $\boldsymbol{\delta R}_{\boldsymbol{zz}}$ is a *small* additive perturbation (in the Frobenius norm sense). Consequently, the matrix $\widehat{\boldsymbol{Q}}$ can be written [20, 21] as $\widehat{\boldsymbol{Q}} = \boldsymbol{Q} + \boldsymbol{\delta Q}$, where $\boldsymbol{\delta Q}$ represents the resulting perturbation in the estimated parameterization matrix, whose norm is of the order of $\|\boldsymbol{\delta R}_{\boldsymbol{zz}}\|$. Since the channel vector $\boldsymbol{c}_a$ is estimated as the eigenvector of $\widehat{\boldsymbol{Q}}^H \widehat{\boldsymbol{Q}}$ [cfr. (18)] corresponding to its smallest eigenvalue, the estimated $\widehat{\boldsymbol{c}}_a$ can be decomposed in its turn as $\widehat{\boldsymbol{c}}_a = \boldsymbol{c}_a + \boldsymbol{\delta c}_a$, where $\boldsymbol{\delta c}_a$ represents the channel estimation error, whose norm is of the order of $\|\boldsymbol{\delta Q}\|$. Since the identification method of Section 4 is able to provide the channel impulse response only up to a real scalar factor, to carry out the performance analysis it is necessary to suitably normalize the estimated vector $\widehat{\boldsymbol{c}}_a$. We thus normalize $\widehat{\boldsymbol{c}}_a$ so that its dth entry $\widehat{c}_a(d)$ is exactly equal to the corresponding entry $c_a(d) \neq 0$ in the true channel vector $\boldsymbol{c}_a$, where $d = \arg\max_i |c_a(i)|$ and, hence, we define the *normalized* channel error as follows

$$(\boldsymbol{\delta c}_a)_n \triangleq \frac{c_a(d)}{\widehat{c}_a(d)} \widehat{\boldsymbol{c}}_a - \boldsymbol{c}_a \,. \tag{22}$$

Reasoning as in [22], we observe that, under the small perturbation assumption, one has

$$\frac{c_a(d)}{\widehat{c}_a(d)} = \frac{c_a(d)}{c_a(d) + \delta c_a(d)} = \frac{1}{1 + \frac{\delta c_a(d)}{c_a(d)}} \approx 1 - \frac{\delta c_a(d)}{c_a(d)} \,, \tag{23}$$

where here and in the sequel the symbol $\approx$ denotes *first-order equality*, i.e., we neglect all the summands that tend to zero, as the sample size N_s approaches infinity, faster than the perturbation norm $\|\boldsymbol{\delta R}_{\boldsymbol{zz}}\|$. By substituting (23) in (22), it follows that

$$(\boldsymbol{\delta c}_a)_n \approx \boldsymbol{\delta c}_a - \frac{\boldsymbol{c_a}}{c_a(d)}\, \delta c_a(d) = \underbrace{\left(\mathbf{I}_{2L_c} - \frac{\boldsymbol{c}_a\, \boldsymbol{f}_d^T}{c_a(d)}\right)}_{\boldsymbol{\Phi} \in \mathbb{C}^{(2L_c)\times(2L_c)}} \boldsymbol{\delta c}_a \,, \tag{24}$$

where $\boldsymbol{f}_d \in \mathbb{R}^{2L_c}$ is the dth canonical basis vector having all entries equal to zero expect for the dth one. Taking into account (24), the normalized channel estimation MSE is given by

$$\mathrm{E}[\|(\boldsymbol{\delta c}_a)_n\|^2] \approx \mathrm{E}[\boldsymbol{\delta c}_a^T \boldsymbol{\Phi}^T \boldsymbol{\Phi}\, \boldsymbol{\delta c}_a] = \mathrm{trace}\{\boldsymbol{\Phi}\, \mathrm{E}[\boldsymbol{\delta c}_a\, \boldsymbol{\delta c}_a^T]\, \boldsymbol{\Phi}^T\} \,, \tag{25}$$

where the last equality is obtained by using the properties of the trace operator. The channel perturbation $\boldsymbol{\delta c}_a$ can be expressed [20, 21] in terms of $\boldsymbol{\delta Q}$; more precisely, using (16), it turns out that $\widehat{\boldsymbol{Q}}\, \widehat{\boldsymbol{c}}_a \approx \boldsymbol{Q}\, \boldsymbol{\delta c}_a + \boldsymbol{\delta Q}\, \boldsymbol{c}_a \approx \mathbf{0}_{2KL_e}$, with $K = L_e + L_c - 1$, and, therefore, it follows that $\boldsymbol{\delta c}_a \approx -\boldsymbol{Q}^\dagger\, \boldsymbol{\delta Q}\, \boldsymbol{c}_a = -\boldsymbol{Q}^\dagger\, (\widehat{\boldsymbol{Q}} - \boldsymbol{Q})\boldsymbol{c}_a = -\boldsymbol{Q}^\dagger\, \widehat{\boldsymbol{Q}}\, \boldsymbol{c}_a$. Consequently, one obtains

$$\mathrm{E}[\boldsymbol{\delta c}_a\, \boldsymbol{\delta c}_a^T] \approx \boldsymbol{Q}^\dagger\, \mathrm{E}[(\widehat{\boldsymbol{Q}}\, \boldsymbol{c}_a)\, (\widehat{\boldsymbol{Q}}\, \boldsymbol{c}_a)^H](\boldsymbol{Q}^\dagger)^H = \boldsymbol{Q}^\dagger\, \boldsymbol{R}_{\boldsymbol{\epsilon\epsilon}}\, (\boldsymbol{Q}^\dagger)^H \,, \tag{26}$$

where $\boldsymbol{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} \stackrel{\triangle}{=} \mathrm{E}[\boldsymbol{\epsilon}\,\boldsymbol{\epsilon}^H]$ denotes the autocorrelation matrix of the vector $\boldsymbol{\epsilon} = \widehat{\boldsymbol{Q}}\,\boldsymbol{c}_a = \mathrm{vec}[\widehat{\boldsymbol{\Pi}}_{\mathcal{N}}\,\boldsymbol{\Psi}(\boldsymbol{c})] \in \mathbb{C}^{2KL_e}$. The next step consists of expressing $\boldsymbol{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$ in terms of the sample autocorrelation matrix $\widehat{\boldsymbol{R}}_{\boldsymbol{z}\boldsymbol{z}}$. To simplify matters, we assume that the mean-square linear dependence (10) between $s(n)$ and $s^*(n)$ holds *everywhere*, i.e., $s^*(n) = \alpha(n)\,s(n)$ for any $n \in Z$ and for any realization of the random process $s(n)$. To be consistent with A1, it turns out that $\alpha(n) = f(n)^{-1} = e^{-j\,2\pi\beta n}$, hence we are assuming explicitly that $s^*(n) = e^{-j\,2\pi\beta n}\,s(n)$. Observe that this subclass of strongly improper processes encompasses all the modulation formats recalled in Section 5. In these cases, it turns out that $\boldsymbol{s}^*(n) = \boldsymbol{J}^*\,\boldsymbol{s}(n)\,e^{-j\,2\pi\beta n}$ which, accounting for model (4) and the partitioned structure of $\boldsymbol{\Omega}^*(n)$ and $\boldsymbol{C}_a$, allows us to explicitly write $\boldsymbol{z}(n)$ as follows

$$\boldsymbol{z}(n) = \boldsymbol{\Psi}(\boldsymbol{c})\,\boldsymbol{s}(n) + \boldsymbol{v}(n)\,, \tag{27}$$

where $\boldsymbol{v}(n) \stackrel{\triangle}{=} \boldsymbol{\Omega}^*(n)\,\widetilde{\boldsymbol{w}}(n) \in \mathbb{C}^{2L_e}$. This alternative expression of $\boldsymbol{z}(n)$, together with the simplifying assumption that the additive noise $w(n)$ is Gaussian, allows one to obtain a coincise expression of the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$, which is valid up to first-order in $\|\boldsymbol{\delta R}_{\boldsymbol{z}\boldsymbol{z}}\|$.

Lemma 4. Assume, in addition to A2, that the additive noise $w(n)$ is a Gaussian process. Under the assumption that the signal-plus-noise eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_K$ of $\boldsymbol{R}_{\boldsymbol{z}\boldsymbol{z}}$ are distinct[6], the first-order approximation of the autocorrelation matrix of the residual vector $\boldsymbol{\epsilon}$ is given by

$$\boldsymbol{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} \approx \frac{1}{N_s} \sum_{m=-(L_e-1)}^{L_e-1} [\boldsymbol{\Psi}^H(\boldsymbol{c})\,\boldsymbol{\Upsilon}\,\boldsymbol{R}_{\boldsymbol{z}\boldsymbol{z}}(m)\,\boldsymbol{\Upsilon}\,\boldsymbol{\Psi}(\boldsymbol{c})]^* \otimes [\boldsymbol{\Pi}_{\mathcal{N}}\,\boldsymbol{R}_{\boldsymbol{v}\boldsymbol{v}}(m)\,\boldsymbol{\Pi}_{\mathcal{N}}]\,, \tag{28}$$

where we have defined the Hermitian matrix $\boldsymbol{\Upsilon} \stackrel{\triangle}{=} \boldsymbol{E}_{\mathcal{S}}\,(\boldsymbol{\Lambda}_{\mathcal{S}} - \sigma_w^2\,\mathbf{I}_K)^{-1}\boldsymbol{E}_{\mathcal{S}}^H \in \mathbb{C}^{(2L_e)\times(2L_e)}$, and the m-lag autocorrelation matrices $\boldsymbol{R}_{\boldsymbol{z}\boldsymbol{z}}(m) \stackrel{\triangle}{=} \mathrm{E}[\boldsymbol{z}(n)\,\boldsymbol{z}^H(n-m)]$ and $\boldsymbol{R}_{\boldsymbol{v}\boldsymbol{v}}(m) \stackrel{\triangle}{=} \mathrm{E}[\boldsymbol{v}(n)\,\boldsymbol{v}^H(n-m)]$. Moreover, it results that $\boldsymbol{R}_{\boldsymbol{z}\boldsymbol{z}}(m) = \sigma_s^2\,\boldsymbol{\Psi}(\boldsymbol{c})\,\boldsymbol{\Gamma}_K(m)\,\boldsymbol{\Psi}^H(\boldsymbol{c}) + \boldsymbol{R}_{\boldsymbol{v}\boldsymbol{v}}(m)$ and $\boldsymbol{R}_{\boldsymbol{v}\boldsymbol{v}}(m) = \sigma_w^2\,[\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{L_e}(m)]\,\boldsymbol{\Omega}^*(m)$.

Proof. The proof of (28) is lengthy but can be conducted by following the guidelines delineated in [23, 24]. As to the explicit expressions of matrices $\boldsymbol{R}_{\boldsymbol{z}\boldsymbol{z}}(m)$ and $\boldsymbol{R}_{\boldsymbol{v}\boldsymbol{v}}(m)$, observe that, taking into account the signal model (27) and invoking assumptions A1 and A2, it is easily seen that $\boldsymbol{R}_{\boldsymbol{z}\boldsymbol{z}}(m) = \boldsymbol{\Psi}(\boldsymbol{c})\,\boldsymbol{R}_{\boldsymbol{s}\boldsymbol{s}}(m)\,\boldsymbol{\Psi}^H(\boldsymbol{c}) + \boldsymbol{R}_{\boldsymbol{v}\boldsymbol{v}}(m)$, where $\boldsymbol{R}_{\boldsymbol{s}\boldsymbol{s}}(m) \stackrel{\triangle}{=} \mathrm{E}[\boldsymbol{s}(n)\,\boldsymbol{s}^H(n-m)] = \sigma_s^2\,\boldsymbol{\Gamma}_K(m)$. On the other hand, observing that, under assumption A2, it results that $\widetilde{\boldsymbol{R}}_{\boldsymbol{w}\boldsymbol{w}}(m) \stackrel{\triangle}{=} \mathrm{E}[\widetilde{\boldsymbol{w}}(n)\,\widetilde{\boldsymbol{w}}^H(n-m)] = \sigma_w^2\,[\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{L_e}(m)]$, one obtains that the autocorrelation matrix of $\boldsymbol{v}(n) = \boldsymbol{\Omega}^*(n)\,\widetilde{\boldsymbol{w}}(n)$ is given by $\boldsymbol{R}_{\boldsymbol{v}\boldsymbol{v}}(m) = \boldsymbol{\Omega}^*(n)\,\widetilde{\boldsymbol{R}}_{\boldsymbol{w}\boldsymbol{w}}(m)\,\boldsymbol{\Omega}^*(n-m) = \sigma_w^2\,[\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{L_e}(m)]\,\boldsymbol{\Omega}^*(m)$. □

At this point, we can state the main result of this section.

[6]This is a very mild assumption that holds generally in practice.

Theorem 2 (Normalized channel MSE)**.** Assume, in addition to A2, that the additive noise $w(n)$ is a Gaussian process. Under the assumption that the signal-plus-noise eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_K$ of $\boldsymbol{R_{zz}}$ are distinct, the first-order approximation of the normalized channel MSE is given by

$$\mathrm{E}[\|(\boldsymbol{\delta c}_a)_n\|^2] \approx \frac{\xi_1 + \xi_2}{N_s} \,, \tag{29}$$

where the summands

$$\xi_\ell \triangleq \sum_{m=-(L_e-1)}^{L_e-1} \mathrm{trace}\Big\{ \boldsymbol{\Phi}\, \boldsymbol{Q}^\dagger \Big\{ \big[\boldsymbol{\Psi}^H(\boldsymbol{c})\, \boldsymbol{E}_{\mathcal{S}}(\boldsymbol{\Lambda}_{\mathcal{S}} - \sigma_w^2\, \mathbf{I}_K)^{-1}\, \boldsymbol{\Theta}_\ell(m)\, (\boldsymbol{\Lambda}_{\mathcal{S}} - \sigma_w^2\, \mathbf{I}_K)^{-1} \boldsymbol{E}_{\mathcal{S}}^H\, \boldsymbol{\Psi}(\boldsymbol{c})\big]^* \otimes \big[\boldsymbol{E}_{\mathcal{N}}\, \boldsymbol{E}_{\mathcal{N}}^H\, (\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{L_e}(m))\, \boldsymbol{\Omega}^*(m)\, \boldsymbol{E}_{\mathcal{N}}\, \boldsymbol{E}_{\mathcal{N}}^H\big] \Big\} (\boldsymbol{Q}^\dagger)^H \boldsymbol{\Phi}^T \Big\} \,, \tag{30}$$

for $\ell \in \{1, 2\}$, with

$$\boldsymbol{\Theta}_1(m) \triangleq \sigma_s^2\, \sigma_w^2\, \boldsymbol{E}_{\mathcal{S}}^H\, \boldsymbol{\Psi}(\boldsymbol{c})\, \boldsymbol{\Gamma}_K(m)\, \boldsymbol{\Psi}^H(\boldsymbol{c})\, \boldsymbol{E}_{\mathcal{S}} \in \mathbb{C}^{K\times K} \tag{31}$$

and

$$\boldsymbol{\Theta}_2(m) \triangleq \sigma_w^4\, \boldsymbol{E}_{\mathcal{S}}^H\, [\mathbf{I}_2 \otimes \boldsymbol{\Gamma}_{L_e}(m)]\, \boldsymbol{\Omega}^*(m)\, \boldsymbol{E}_{\mathcal{S}} \in \mathbb{C}^{K\times K} \,, \tag{32}$$

are independent of the sample size N_s.

Proof. The proof of (29) can be obtained, after some algebraic manipulations, by substituting (28) in (26) and the result in (25). □

It can be observed from (29) that, for given nonzero values of the symbol and noise variances σ_s^2 and σ_w^2, the normalized channel MSE approaches zero as the sample size N_s tends to infinity.

7 Simulation results

In this section, the MSE performance of the proposed SISO subspace-based identification method, working at symbol spacing T_s, is investigated and compared with the reference SIMO subspace-based method proposed in [3] (referred to as the method of Moulines in the sequel), working at half-symbol spacing $T_s/2$. We validate our performance analysis of Section 6 by reporting both the analytical results and those obtained by Monte Carlo computer simulations. For comparison, in addition to its experimental results, the theoretical performance of the method of Moulines, derived in [22], is also plotted.

In all the experiments, the following simulation setting is assumed. The input stream $s(n)$ is drawn from an OQPSK constellation[7] and the additive noise $w(n)$ is a complex Gaussian process. The signal-to-noise

[7]We have also carried out simulations for a BPSK modulation, whose results are not reported here, since they are very similar to those of the OQPSK case.

ratio (SNR) is defined as SNR $\triangleq (\sigma_s^2/\sigma_w^2)\|\boldsymbol{c}\|^2$ and both the symbol and noise sequences are randomly and independently generated at the start of each Monte Carlo run. Since the proposed method and the method of Moulines employ a different discrete-time channel, to allow for a fair comparison, we considered for both methods the same continuous-time channel $c_a(t)$ which, with the exception of experiment 5, spans $L_c = 3$ symbol periods. More precisely, we started from the $T_s/2$-sampled version of $c_a(t)$, i.e., $\widetilde{c}(q) \triangleq c_a(q\,T_s/2)$, for $q \in \{0,1,\ldots,2L_c-1\}$, which can be expressed in terms of the two polyphase components $\widetilde{c}^{\,(0)}(n) \triangleq \widetilde{c}(2\,n)$ and $\widetilde{c}^{\,(1)}(n) \triangleq \widetilde{c}(2\,n+1)$, for $n \in \{0,1,\ldots,L_c-1\}$, which are the two channels employed by the SIMO method of Moulines. Thus, as a by-product, we obtain the unique symbol-spaced channel for the proposed SISO method as $c(n) = \widetilde{c}^{\,(0)}(n)$, $n \in \{0,1,\ldots,L_c-1\}$. The two channels $\widetilde{c}^{\,(\ell)}(n)$, for $\ell = 0, 1$, are assigned in terms of their $\mathcal{Z}$-transforms:

$$\widetilde{C}^{\,(\ell)}(z) = (1 - 0.5\,e^{j\,\theta_{1,\ell}}z^{-1})\,(1 - 1.2\,e^{j\,\theta_{2,\ell}}z^{-1})\,, \quad \text{for } \ell = 0, 1\,, \tag{33}$$

where $\theta_{1,0} = 0.5\,\pi + \gamma$, $\theta_{2,0} = \theta_{1,0} + \pi$, $\theta_{1,1} = \theta_{1,0} + \gamma$ and $\theta_{2,1} = \theta_{2,0} + \gamma$, and the angular separation γ can be adjusted so as to approach the situation where identification fails for both methods under comparison. Indeed, for $\gamma = 0$, the first polyphase component $\widetilde{c}^{\,(0)}(n) \equiv c(n)$ has two purely imaginary zeros, which is the identifiability limit for the proposed SISO method (see discussion in Section 5 for OQPSK modulation). Moreover, for $\gamma = 0$, the two subchannels $\widetilde{c}^{\,(0)}(n)$ and $\widetilde{c}^{\,(1)}(n)$ have common zeros, which is the identifiability limit for the SIMO method of Moulines. Unless otherwise specified, the angular separation is fixed to $\gamma = 0.2\,\pi$. The channel coefficients are scaled so that the two vector $\boldsymbol{c} \triangleq [\widetilde{c}^{\,(0)}(0), \widetilde{c}^{\,(0)}(1), \ldots, \widetilde{c}^{\,(0)}(L_c-1)]^T \in \mathbb{C}^{L_c}$ and $\widetilde{\boldsymbol{c}} \triangleq [\widetilde{c}(0), \widetilde{c}(1), \ldots, \widetilde{c}(2L_c-1)]^T \in \mathbb{C}^{2L_c}$ have unitary norm. As channel estimation performance measure, denoting with $\boldsymbol{x}$ the channel vector to be estimated for any of the two methods under comparison, we used the normalized MSE expressed in dB:

$$\text{MSE (dB)} \triangleq 10\,\log_{10}\left(\frac{1}{N_r}\sum_{i=1}^{N_r}\|\chi\,\widehat{\boldsymbol{x}}_i - \boldsymbol{x}\|^2\right)\,, \tag{34}$$

where $N_r = 1000$ is the number of independent Monte Carlo trials conducted for each experiments, $\widehat{\boldsymbol{x}}_i$ denotes the estimate of $\boldsymbol{x}$ obtained in the ith run and, denoting with $x(\ell)$ and $\widehat{x}_i(\ell)$ the ℓth entry of $\boldsymbol{x}$ and $\widehat{\boldsymbol{x}}_i$, the normalization coefficient χ is given by $\chi = x(d)/\widehat{x}_i(d)$, with $d = \arg\max_\ell |x(\ell)|$. This choice is consistent with the normalization carried out for performance analysis in Section 6.

Experiment 1 – Normalized channel MSE versus SNR. In this experiment, we evaluate the performances of the considered methods as a function of SNR ranging from 0 to 50 dB. The length of the temporal window is set to $L_e = 5$ and the sample size is equal to $N_s = 200$ symbols. Results of Fig. 1 evidence that the proposed method outperforms the method of Moulines for all values of SNR. Moreover, observe that the theoretical expression (29) of the MSE derived in the previous section agrees very well with the simulation results for all values of SNR whereas, in accordance with the experimental results conducted in [22, 23], the theoretical analysis for the method of Moulines is not accurate for small values of SNR (SNR $\leq$ 5 dB). Indeed, since our method assures better performances than the identification algorithm of Moulines, the small perturbation assumption is verified in our method for a wider range of SNR values.

Experiment 2 – Normalized channel MSE versus sample size. In this experiment, the performances of the considered methods are studied as a function of the sample size N_s ranging from 50 to 1000 symbols. The length of the temporal window is set to $L_e = 5$ and SNR $= 20$ dB. It can be seen from Fig. 2 that exploiting the improper nature of the transmitted symbol sequence allows one to obtain a more data-efficient channel estimation algorithm in comparison with the SIMO subspace-based method of Moulines. Note that the proposed method assures a normalized MSE equal to -30 dB for a sample size as short as 100 symbols, whereas the method of Moulines requires about 500 symbols to achieve the same accuracy. Also in this case, observe that the theoretical performance analysis for the proposed method agrees very well with Monte Carlo results.

Experiment 3 – Normalized channel MSE versus L_e. In this experiment, we evaluate the performances of the methods under comparison as a function of the length of the temporal window L_e ranging from 3 to 10. In this case, remembering that $L_c = 3$, the dimension $L_e - L_c + 1$ of the noise subspace $\mathcal{R}(\boldsymbol{E}_{\mathcal{N}})$ varies from 1 (corresponding to $L_e = 3$) to 8 (corresponding to $L_e = 10$). The sample size is equal to $N_s = 200$ symbols and SNR $= 20$ dB. Results of Fig. 3, besides corroborating again the accuracy of the theoretical analysis carried out in Section 6, show moreover that, compared with the algorithm of Moulines, the performance of the proposed method is less sensitive to the choice of the length L_e of the temporal window. Moreover, it is worthwhile to note that the proposed method is able to reliably estimate the channel parameters even when $L_e = 3$, i.e., only one noise eigenvector is considered. Finally, it is apparent that, for the proposed method, employing values of $L_e > 5$ does not lead to a significant improvement in channel estimation accuracy.

Experiment 4 – Normalized channel MSE versus γ. In this experiment, we compare the robustness of the considered methods to the location of channel zeros, by plotting the theoretical MSE as a function of the angular separation γ ranging from $0.005\,\pi$ to $0.3\,\pi$, and for different values of SNR (recall that as $\gamma \to 0$ the channel is not identifiable for both methods under comparison). Unlike all the other experiments, we chose not to report Monte Carlo results, due to the extreme variability of the curves when the channel approaches the non-identifiability condition. The length of the temporal window is set to $L_e = 5$ and the sample size is equal to $N_s = 200$ symbols. It can be noted from Fig. 4 that the proposed method is able to assure better performance than the algorithm of Moulines both in poor (i.e., small γ) as well as good (i.e., high γ) channel conditions. In particular, as γ approaches zero, the performance of the method of Moulines degrades faster than that of the proposed channel estimation technique.

Experiment 5 – Normalized channel MSE versus L_c. In the last experiment, we evaluate the performances of the considered methods as a function of the channel length L_c ranging from 2 to 7. In this case, the $\mathscr{Z}$-transforms of the polyphase components $\widetilde{c}^{(\ell)}(n)$, $\ell = 0, 1$, are given by

$$\widetilde{C}^{(\ell)}(z) = \prod_{i=1}^{L_c-1} (1 - \rho_i \, e^{j\,\theta_{i,\ell}} z^{-1}) \,, \quad \text{for } \ell = 0, 1 \,, \tag{35}$$

where $\theta_{1,0} = 0.5\,\pi + \gamma$ and $\rho_1 = 0.5$, $\theta_{2,0} = \theta_{1,0} + \pi$ and $\rho_2 = 1.2$, $\theta_{3,0} = -\theta_{1,0}$ and $\rho_3 = 0.8$, $\theta_{4,0} = -\theta_{2,0}$ and $\rho_4 = 1.5$, $\theta_{5,0} = -0.5\,\pi + \gamma$ and $\rho_5 = 1.5$, $\theta_{6,0} = -\theta_{5,0}$ and $\rho_6 = 1.8$, whereas $\theta_{i,1} = \theta_{i,0} + \gamma$, for $i \in \{1, 2, \ldots, L_c - 1\}$. The length of the temporal window is set to $L_e = 10$, the sample size is equal to $N_s = 200$ symbols and SNR $= 20$ dB. Results of Fig. 5 show that the normalized MSE for both methods increases as a function of the channel length L_c. However, the proposed method assure a significant performance gain with respect to the method of Moulines, for all the considered values of L_c.

8 Conclusion

We tackled in this paper the problem of blindly identifying a SISO-FIR channel via a subspace-based approach. We provided analytically that blind channel identification is possible if the transmitted symbol sequence belongs to a particular subclass of improper random processes, i.e., it is strongly improper, which is a property exhibited by many digital modulation schemes of practical interest. By exploiting the strongly improper nature of the transmitted symbol stream, we devised a general framework for subspace-based channel identification, which allowed us to address identifiability issues and carry out the theoretical performance analysis in a unifying manner. Results of computer simulations corroborated the theoretical performance analysis and showed that the proposed method performs significantly better than the conventional SIMO subspace-based algorithm originally proposed in [3]. Finally, it should be noted that our general framework allows one to extend the treatment to modulation techniques that are more sophisticated than those discussed in the paper, e.g., modulations schemes wherein conjugate cyclostationarity is deliberately introduced in transmission. This issue is the subject of our current research.

A Proof of Lemma 3

Let us define the nonsingular matrix $\mathbf{\Theta} \stackrel{\triangle}{=} \text{diag}[\overbrace{1, \ldots, 1}^{L_e+1}, e^{-j\,2\pi\beta}, e^{-j\,2\pi 2\beta}, \ldots, e^{-j\,2\pi(L_e-1)\beta}] \in \mathbb{C}^{(2L_e)\times(2L_e)}$. Observe[8] preliminarily that $\text{rank}(\mathbf{\Psi}) = \text{rank}(\mathbf{\Theta}\,\mathbf{\Psi}) = \text{rank}([\boldsymbol{C}^T, \boldsymbol{D}^T]^T)$, where $\boldsymbol{D} \stackrel{\triangle}{=} \mathbf{T}_{L_e}(\boldsymbol{d}^T) \in \mathbb{C}^{L_e \times K}$ denotes the Toeplitz matrix associated with $\overline{\boldsymbol{c}} \stackrel{\triangle}{=} [c^*(0), c^*(1)\,e^{j\,2\pi\beta}, \ldots, c^*(L_c-1)\,e^{j\,2\pi(L_c-1)\beta}]^T \in \mathbb{C}^{L_c}$. Thus, under assumption A4, applying standard results in blind channel identification (see, e.g., [25]) yields that $\mathbf{\Psi}$ is full-column rank, i.e., $\text{rank}(\mathbf{\Psi}) = K = L_e + L_c - 1$, *if and only if* the $\mathcal{Z}$-transforms of $\boldsymbol{c}$ and $\overline{\boldsymbol{c}}$ are *coprime*, that is, they do not share common zeros. Let

$$C(z) \stackrel{\triangle}{=} \mathcal{Z}[\boldsymbol{c}] = \sum_{q=0}^{L_c-1} c(q)\, z^{-q} = c(0) \prod_{q=1}^{L_c-1} (1 - \zeta_q\, z^{-1}) \tag{36}$$

be the $\mathcal{Z}$-transform of the vector $\boldsymbol{c}$, where $\zeta_1, \zeta_2, \ldots, \zeta_{L_c-1}$ denote the zeros of $C(z)$, it follows that

$$\begin{aligned} \overline{C}(z) \stackrel{\triangle}{=} \mathcal{Z}[\overline{\boldsymbol{c}}] &= \sum_{q=0}^{L_c-1} c^*(q)\, e^{j\,2\pi q\beta}\, z^{-q} = C^*(z^*\, e^{j\,2\pi\beta}) \\ &= c^*(0) \prod_{q=1}^{L_c-1} [1 - (\zeta_q^*\, e^{j\,2\pi\beta})\, z^{-1}]\,. \end{aligned} \tag{37}$$

Thus, from (36) and (37), it results that the polynomials $C(z)$ and $\overline{C}(z)$ do not share common zeros if and only if $\zeta_\ell \neq \zeta_q^*\, e^{j\,2\pi\beta}$, for $\ell, q \in \{1, 2, \ldots, L_c - 1\}$. □

B Proof of Theorem 1

To prove the theorem, we state beforehand the following lemma.

Lemma 5. Under assumptions A4 and A5, an arbitrary vector $\boldsymbol{c}' \stackrel{\triangle}{=} [c'(0), c'(1), \ldots, c'(L_c-1)]^T \in \mathbb{C}^{L_c}$ is a solution of (15) if and only if $\mathcal{R}(\mathbf{\Psi}') \subseteq \mathcal{R}(\mathbf{\Psi})$, where $\mathbf{\Psi}' \stackrel{\triangle}{=} [(\boldsymbol{C}')^T, \boldsymbol{J}^*\,(\boldsymbol{C}')^H]^T \in \mathbb{C}^{(2L_e)\times K}$, with $\boldsymbol{C}' \stackrel{\triangle}{=} \mathbf{T}_{L_e}[(\boldsymbol{c}')^T] \in \mathbb{C}^{L_e \times K}$ denoting the Toeplitz matrix associated with vector $\boldsymbol{c}'$ and $K = L_e + L_c - 1$.

Proof. Observe that, owing to (15) and accounting for the properties of $\mathbf{\Pi}_{\mathcal{N}}$, the columns of $\mathbf{\Psi}$ belong to $\mathcal{R}^\perp(\boldsymbol{E}_{\mathcal{N}})$, that is, $\mathcal{R}(\mathbf{\Psi}) \subseteq \mathcal{R}^\perp(\boldsymbol{E}_{\mathcal{N}})$. Due to the linear independence of the noise eigenvectors, the subspace $\mathcal{R}^\perp(\boldsymbol{E}_{\mathcal{N}})$ has dimension equal to K; on the other hand, by assumption, the dimension of $\mathcal{R}(\mathbf{\Psi})$ is K. Thus, it turns out that $\mathcal{R}(\mathbf{\Psi}) = \mathcal{R}^\perp(\boldsymbol{E}_{\mathcal{N}})$. On the basis of this equivalence, Lemma 5 is readily proven.

[8] To keep notation simple, throughout the Appendices, we replace $\mathbf{\Psi}(\boldsymbol{c})$ with $\mathbf{\Psi}$.

If: Let $\mathcal{R}(\boldsymbol{\Psi}') \subseteq \mathcal{R}(\boldsymbol{\Psi})$, the columns of $\boldsymbol{\Psi}'$ belong to the subspace $\mathcal{R}(\boldsymbol{\Psi})$ and, thus, to $\mathcal{R}^{\perp}(\boldsymbol{E}_{\mathcal{N}})$ which, accounting for the relation $\mathcal{N}(\boldsymbol{\Pi}_{\mathcal{N}}) = \mathcal{R}^{\perp}(\boldsymbol{E}_{\mathcal{N}})$, implies that $\boldsymbol{\Pi}_{\mathcal{N}}\,\boldsymbol{\Psi}' = \mathbf{O}_{R\times K}$.

Only If: Let $\boldsymbol{\Pi}_{\mathcal{N}}\,\boldsymbol{\Psi}' = \mathbf{O}_{R\times K}$, then the columns of $\boldsymbol{\Psi}'$ belong to the subspace $\mathcal{N}(\boldsymbol{\Pi}_{\mathcal{N}}) = \mathcal{R}^{\perp}(\boldsymbol{E}_{\mathcal{N}})$, that is, $\mathcal{R}(\boldsymbol{\Psi}') \subseteq \mathcal{R}^{\perp}(\boldsymbol{E}_{\mathcal{N}}) = \mathcal{R}(\boldsymbol{\Psi})$. $\square$

Given the nonsingular matrix $\boldsymbol{\Theta} \triangleq \text{diag}[\overbrace{1,\ldots,1}^{L_e+1}, e^{-j\,2\pi\beta}, e^{-j\,2\pi 2\beta}, \ldots, e^{-j\,2\pi(L_e-1)\beta}] \in \mathbb{C}^{(2L_e)\times(2L_e)}$, consider the matrices $\boldsymbol{\Upsilon} \triangleq \boldsymbol{\Theta}\,\boldsymbol{\Psi} = [\boldsymbol{C}^T, \boldsymbol{D}^T]^T \in \mathbb{C}^{(2L_e)\times K}$ and $\boldsymbol{\Upsilon}' \triangleq \boldsymbol{\Theta}\,\boldsymbol{\Psi}' = [(\boldsymbol{C}')^T, (\boldsymbol{D}')^T]^T \in \mathbb{C}^{(2L_e)\times K}$, where $\boldsymbol{D} \triangleq \mathbf{T}_{L_e}(\boldsymbol{d}^T) \in \mathbb{C}^{L_e\times K}$ and $\boldsymbol{D}' \triangleq \mathbf{T}_{L_e}[(\boldsymbol{d}')^T] \in \mathbb{C}^{L_e\times K}$ represent the Toeplitz matrices associated with the two vectors $\boldsymbol{d} \triangleq [c^*(0), c^*(1)\,e^{j\,2\pi\beta}, \ldots, c^*(L_c-1)\,e^{j\,2\pi(L_c-1)\beta}]^T \in \mathbb{C}^{L_c}$ and $\boldsymbol{d}' \triangleq [c'(0)^*, c'(1)^*\,e^{j\,2\pi\beta}, \ldots, c'(L_c-1)^*\,e^{j\,2\pi(L_c-1)\beta}]^T \in \mathbb{C}^{L_c}$, respectively. It can be shown that the following equivalence holds

$$\mathcal{R}(\boldsymbol{\Psi}') \subseteq \mathcal{R}(\boldsymbol{\Psi}) \quad \Longleftrightarrow \quad \mathcal{R}(\boldsymbol{\Upsilon}') \subseteq \mathcal{R}(\boldsymbol{\Upsilon})\,. \tag{38}$$

Let $\mathbf{T}_{L_e}(\boldsymbol{H}) \in \mathbb{C}^{(2L_e)\times K}$ and $\mathbf{T}_{L_e}(\boldsymbol{H}') \in \mathbb{C}^{(2L_e)\times K}$ be the *block-Sylvester* matrices associated with the matrices $\boldsymbol{H} \triangleq [\boldsymbol{h}(0), \boldsymbol{h}(1), \ldots, \boldsymbol{h}(L_c-1)] \in \mathbb{C}^{2\times L_c}$ and $\boldsymbol{H}' \triangleq [\boldsymbol{h}'(0), \boldsymbol{h}'(1), \ldots, \boldsymbol{h}'(L_c-1)] \in \mathbb{C}^{2\times L_c}$, respectively, where $\boldsymbol{h}(k) \triangleq [c(k), c^*(k)\,e^{j\,2\pi\beta k}]^T \in \mathbb{C}^2$ and $\boldsymbol{h}'(k) \triangleq [c'(k), (c'(k))^*\,e^{j\,2\pi\beta k}]^T \in \mathbb{C}^2$, for $k = 0, 1, \ldots, L_c - 1$. Since $\mathbf{T}_{L_e}(\boldsymbol{H})$ and $\mathbf{T}_{L_e}(\boldsymbol{H}')$ differ from $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Upsilon}'$, respectively, by interchange of rows, it follows that $\mathcal{R}(\boldsymbol{\Upsilon}) = \mathcal{R}[\mathbf{T}_{L_e}(\boldsymbol{H})]$ and $\mathcal{R}(\boldsymbol{\Upsilon}') = \mathcal{R}[\mathbf{T}_{L_e}(\boldsymbol{H}')]$. Thus, by virtue of Lemma 4 and (38), we can state that $\boldsymbol{c}'$ is a solution of (15) if and only if

$$\mathcal{R}[\mathbf{T}_{L_e}(\boldsymbol{H}')] \subseteq \mathcal{R}[\mathbf{T}_{L_e}(\boldsymbol{H})]\,. \tag{39}$$

By exploiting the shift property of the Toeplitz matrices and following the guidelines delineated in [3], it can be proven that, under assumptions A4 and A5, relation (39) holds if and only if $\boldsymbol{H}' = \xi\,\boldsymbol{H}$, with $\xi \in \mathbb{C}$, which implies that $c'(k) = \xi\,c(k)$ and $c'(k)^* = \xi\,c(k)^*$, for $k = 0, 1, \ldots, L_c - 1$: the last two equalities show that if $\boldsymbol{c}'$ is a solution of (15), then $\boldsymbol{c}'$ must differ from $\boldsymbol{c}$ up to a multiplicative scalar ξ, which must be a real number, i.e., $\xi^* = \xi$. $\square$

C Derivation of the parameterization matrix $\boldsymbol{Q}$

By performing the following partition $\boldsymbol{\Pi}_{\mathcal{N}} = [\boldsymbol{\Pi}'_{\mathcal{N}}, \boldsymbol{\Pi}''_{\mathcal{N}}]$, with $\boldsymbol{\Pi}'_{\mathcal{N}}, \boldsymbol{\Pi}''_{\mathcal{N}} \in \mathbb{C}^{(2L_e)\times L_e}$, and exploiting the structure of $\boldsymbol{\Psi}$, the matrix equation (15) becomes $\boldsymbol{\Pi}'_{\mathcal{N}}\,\boldsymbol{C} + \boldsymbol{\Pi}''_{\mathcal{N}}\,\boldsymbol{C}^*\,\boldsymbol{J}^* = \mathbf{O}_{(2L_e)\times K}$. This equation is equivalent to $\text{vec}(\boldsymbol{\Pi}_{\mathcal{N}}\,\boldsymbol{C} + \boldsymbol{\Pi}''_{\mathcal{N}}\,\boldsymbol{C}^*\,\boldsymbol{J}^*) = \mathbf{0}_{2KL_e}$, which, by using the properties of Kronecker product, can be written as $(\mathbf{I}_K \otimes \boldsymbol{\Pi}'_{\mathcal{N}})\,\text{vec}(\boldsymbol{C}) + (\boldsymbol{J}^* \otimes \boldsymbol{\Pi}''_{\mathcal{N}})\,\text{vec}(\boldsymbol{C}^*) = \mathbf{0}_{2KL_e}$. Observing that $\text{vec}(\boldsymbol{C}) = \text{vec}(\boldsymbol{C}_R) + j\,\text{vec}(\boldsymbol{C}_I)$ and

$\text{vec}(\boldsymbol{C}^*) = [\text{vec}(\boldsymbol{C})]^*$, we finally obtain

$$\text{vec}(\boldsymbol{\Pi}_{\mathcal{N}}\,\boldsymbol{\Psi}) = \boldsymbol{\Lambda}_1\,\text{vec}(\boldsymbol{C}_R) + \boldsymbol{\Lambda}_2\,\text{vec}(\boldsymbol{C}_I) = \mathbf{0}_{2KL_e}\,, \tag{40}$$

where $\boldsymbol{\Lambda}_1 \triangleq (\mathbf{I}_K \otimes \boldsymbol{\Pi}'_{\mathcal{N}}) + (\boldsymbol{J}^* \otimes \boldsymbol{\Pi}''_{\mathcal{N}}) \in \mathbb{C}^{(2KL_e)\times KL_e}$ and $\boldsymbol{\Lambda}_2 \triangleq j\,[(\mathbf{I}_K \otimes \boldsymbol{\Pi}'_{\mathcal{N}}) - (\boldsymbol{J}^* \otimes \boldsymbol{\Pi}''_{\mathcal{N}})] \in \mathbb{C}^{(2KL_e)\times KL_e}$. At this point, it is convenient to write the Toeplitz matrices $\boldsymbol{C}_R$ and $\boldsymbol{C}_I$ in the following form

$$\boldsymbol{C}_R = \sum_{\ell=0}^{L_c-1} c_R(\ell)\,\boldsymbol{F}^{(\ell)} \quad \text{and} \quad \boldsymbol{C}_I = \sum_{\ell=0}^{L_c-1} c_I(\ell)\,\boldsymbol{F}^{(\ell)}\,, \tag{41}$$

where $\boldsymbol{F}^{(\ell)} \triangleq \mathbf{T}_{L_e}(\boldsymbol{a}_\ell^T) \in \mathbb{R}^{L_e\times K}$ is the Toeplitz matrix associated with $\boldsymbol{a}_\ell \triangleq [\overbrace{0,\ldots,0}^{\ell},1,0,\ldots,0]^T \in \mathbb{R}^K$. Accounting for (41) and using again the properties of Kronecker product, it can be shown that $\text{vec}(\boldsymbol{C}_R) = \boldsymbol{\Lambda}_3\,\boldsymbol{c}_R$ and $\text{vec}(\boldsymbol{C}_I) = \boldsymbol{\Lambda}_3\,\boldsymbol{c}_I$, with $\boldsymbol{\Lambda}_3 \triangleq (\mathbf{I}_K \otimes \boldsymbol{F})\,\boldsymbol{G}$, and $\boldsymbol{F} \triangleq [\boldsymbol{F}^{(0)}, \boldsymbol{F}^{(1)}, \ldots, \boldsymbol{F}^{(L_c-1)}] \in \mathbb{R}^{L_e\times(KL_c)}$, where the matrix $\boldsymbol{G} \in \mathbb{R}^{(K^2L_c)\times L_c}$ (containing only zeros and ones) can be easily derived and, thus, its expression is omitted for the sake of brevity. On the basis of the above observations, equation (40) can be concisely written as $\text{vec}(\boldsymbol{\Pi}_{\mathcal{N}}\,\boldsymbol{\Psi}) = \boldsymbol{Q}\,\boldsymbol{c}_a = \mathbf{0}_{2KL_e}$, where $\boldsymbol{Q} \triangleq [\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2]\,(\mathbf{I}_2 \otimes \boldsymbol{\Lambda}_3) \in \mathbb{C}^{2KL_e\times(2L_c)}$ and $\boldsymbol{c}_a \triangleq [\boldsymbol{c}_R^T, \boldsymbol{c}_I^T]^T \in \mathbb{R}^{2L_c}$.

D Discussion on the identifiability condition derived in [7]

In this Appendix, we show that the identifiability conditions derived in [7, Theorem V.2, pag. 2201] are incomplete. Let us consider the case where the input sequence $s(n)$ is a stationary strongly *real* improper random process (e.g., ASK, BPSK or DBPSK), that is, $f(n) = 1$, for any $n \in \mathbb{Z}$; in this case, according to assumption A1, it results that $R_{ss}(m) = R_{ss^*}(m) = \sigma_s^2\,\delta(m)$. Using our notations, the proposition considered in [7, Theorem V.2] can be summarized as follows: prove that, in the absence of noise, the joint use of the autocorrelation $R_{rr}(m) \triangleq \text{E}[r(n)\,r^*(n-m)] = \sigma_s^2 \sum_{q=0}^{L_c-1} c(q)\,c^*(q-m)$ and the conjugate autocorrelation $R_{rr^*}(m) \triangleq \text{E}[r(n)\,r(n-m)] = \sigma_s^2 \sum_{q=0}^{L_c-1} c(q)\,c(q-m)$ functions of the received signal $r(n)$, given by (2), allows one to uniquely (up to a scalar factor) identify the channel impulse response $\{c(n)\}_{n=0}^{L_c-1}$. To study this problem, let us consider the $\mathcal{Z}$-transforms of the functions $R_{rr}(m)$ and $R_{rr^*}(m)$, which are given by[9]

$$S_{rr}(z) \triangleq \mathcal{Z}[R_{rr}(m)] = \sum_{m=-\infty}^{+\infty} R_{rr}(m)\,z^{-m} = C(z)\,C^*(1/z^*)\,, \tag{42}$$

$$S_{rr^*}(z) \triangleq \mathcal{Z}[R_{rr^*}(m)] = \sum_{m=-\infty}^{+\infty} R_{rr^*}(m)\,z^{-m} = C(z)\,C(1/z)\,, \tag{43}$$

[9]For the sake of simplicity, we assume that $\sigma_s^2 = 1$.

where $C(z) \stackrel{\triangle}{=} \mathcal{Z}[\boldsymbol{c}] = \sum_{q=0}^{L_c-1} c(q)\, z^{-q}$ denotes the $\mathcal{Z}$-transform of the channel vector $\boldsymbol{c}$. When the polynomials $S_{rr}(z)$ and $S_{rr^*}(z)$ are known, relations (42) and (43) can be used to identify the unknown channel $\mathcal{Z}$-transform $C(z)$. To this aim, it can be seen that the following two statements are equivalent: (1) an arbitrary polynomial $C'(z)$ satisfies both (42) and (43); (2) $C'(z) = A(z)\, C(z)$, where $A(z)$ is an arbitrary *rational* function verifying the two relations (i) $A(z)\, A^*(1/z^*) = 1$ and (ii) $A(z)\, A(1/z) = 1$. In addition, since the channel is a FIR system, the function $C'(z)$ cannot have poles except for $z = 0$: this fact imposes that (iii) the poles of $A(z)$ coincide with the zeros of $C(z)$. It is well-known that a polynomial $A(z)$ satisfying condition (i) admits the following factored form

$$A(z) = \eta \prod_{q=1}^{Q} \frac{1 - a_q^*\, z^{-1}}{a_q - z^{-1}} \,, \tag{44}$$

where η is an unit-modulus complex number. By virtue of (44), it follows that

$$A(z)\, A(1/z) = \eta^2 \prod_{q=1}^{Q} \frac{(1 - a_q^*\, z^{-1})\,(a_q^* - z^{-1})}{(a_q - z^{-1})\,(1 - a_q\, z^{-1})} \,. \tag{45}$$

The Authors in [7] claim that, *if* $C(z)$ *has no real zeros* (which, owing to condition (iii), implies that $A(z)$ has no real poles or, equivalently, zeros), the only function $A(z)$ satisfying both conditions (i) and (ii) is the zero-order polynomial $A(z) = \pm 1$. In this case, the unknown channel impulse response can be uniquely identified, that is, $C'(z)$ is a solution of both (42) and (43) if and only if $C'(z) = \pm C(z)$. As it is apparent from (45), this claim is *incomplete*: indeed, if $A(z)$ has no real poles while possessing a conjugate pair of zeros, say, $a_{q_1}^* = a_{q_2}$, with $q_1, q_2 \in \{1, 2, \ldots, Q\}$, there exists a function $A(z)$ verifying both conditions (i) and (ii), given by

$$A(z) = \pm \frac{(1 - a_{q_1}^*\, z^{-1})\,(1 - a_{q_2}^*\, z^{-1})}{(a_{q_1} - z^{-1})\,(a_{q_2} - z^{-1})} \,, \tag{46}$$

which clearly does not allow *unique* identification of the channel function $C(z)$. Based on this example and taking into account condition (iii), the condition derived by the Authors in [7] must be rephrased as follows: the channel vector $\boldsymbol{c}$ can be uniquely identified from (42) and (43) *if its* $\mathcal{Z}$*-transform has not real zeros and conjugate pairs of zeros.*

References

[1] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. of the IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.

[2] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Information Theory*, vol. 36, pp. 312–321, Mar. 1990.

[3] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[4] Z. Ding, "Characterictics of band-limited channels unidentifiable from second-order cyclostationary statistics," *IEEE Signal Processing Lett.*, vol. 3, pp. 150–152, May 1996.

[5] A. Chevreuil and Ph. Loubaton, "Blind second-order identification of FIR channels: forced cyclostationarity and structured subspace method," *IEEE Signal Processing Lett.,*, vol. 4, pp. 204–206, July 1997.

[6] E. Serpedin and G.B. Giannakis, "Blind channel identifcation and equalization with modulation-induced cyclostationarity," *IEEE Trans. Signal Processing*, vol. 46, pp. 1930–1944, July 1998.

[7] O. Grellier, P. Comon, B. Mourrain, and P. Trébuchet, "Analytical blind channel identification," *IEEE Trans. Signal Processing*, vol. 50, pp. 2196–2207, Sept. 2002.

[8] F.D. Neeser and J.L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Information Theory*, pp. 1293–1302, July 1993.

[9] J.G. Proakis, *Digital Communications*. New York: McGraw-Hill, 2001.

[10] W.A. Gardner, *Introduction to Random Processes*. New York: McGraw-Hill, 1990, ch.12.

[11] M. Kristensson, B. Ottersten and D. Slock, "Blind subspace identification of a BPSK communication channel," in *Proc. of 30th Asilomar Conference on Signals, Systems and Computers*, Monterey, USA, pp. 828-832, Nov. 1996.

[12] Z. Ding and G. Li, "Single-channel blind equalization for GSM cellular systems," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1493–1505, Oct. 1998.

[13] P.J. Schreier and L.L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Trans. Signal Processing*, vol. 51, pp. 714–725, March 2003.

[14] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge: Cambridge University Press, 1990.

[15] X. Li and H.H. Fan, "Blind channel identification: subspace tracking method without rank estimation," *IEEE Trans. Signal Processing*, vol. 49, pp. 2372–2382, Oct. 2001.

[16] W.H. Gerstacker, R. Schober and A. Lampe, "Receivers with widely linear processing for frequency-selective channels," *IEEE Trans. Commun.*, vol. 51, pp. 1512–1523, Sept. 2003.

[17] S. Benedetto and E. Biglieri, *Principles of Digital Communications with Wireless Applications*. New York: Kluwer Academic/Plenum Publishers, 1999.

[18] International Telecommunication Union (ITU), Radiocommunication Study Groups, *Study of the required geographic coordination distance between a HAPS TDMA system and adjacent system IMT-2000 terrestrial mobile stations operating in the same frequency band*. Document 8-1/368-E, May 1988.

[19] J.A.C. Bingham, *The Theory and Practice of Modem Design*. New York: J. Wiley & Sons, 1988.

[20] F. Li, H. Liu and R.J. Vaccaro "Performance analysis for DOA estimation algorithms: unification, simplification, and observations," *IEEE Trans. Aerospace and Electronic Systems*, vol. 29, pp. 1170–1184, Oct. 1993.

[21] G.W. Stewart, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Review*, vol. 15, pp. 727–764, Oct. 1973.

[22] W. Qiu and Y. Hua, "Performance analysis of the subspace method for blind channel identification," *Signal Processing*, vol. 50 (1996), pp. 71–81.

[23] M. Kristensson and B. Ottersten, "A statistical approach to subspace based blind identification," *IEEE Trans. Signal Processing*, vol. 46, pp. 1612–1623, June 1998.

[24] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies, " *IEEE Trans. Signal Processing*, vol. 39, pp. 1836–1847, Aug. 1991.

[25] E. Serpedin and G.B. Giannakis, "A simple proof of a known blind channel identifiability result," *IEEE Trans. Signal Processing*, vol. 47, pp. 591–593, Feb. 1999.

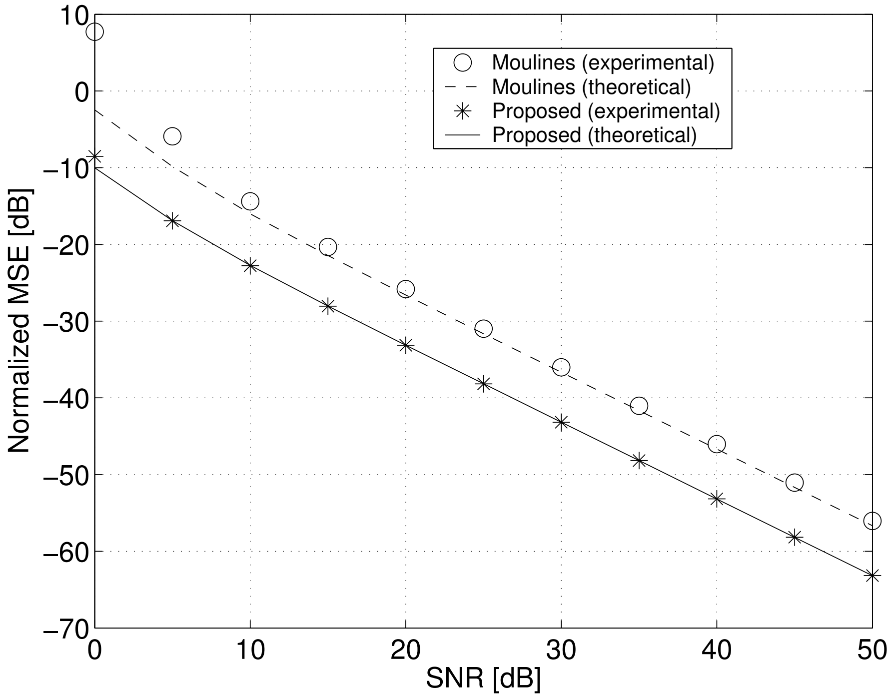

Figure 1: Normalized channel MSE versus SNR ($L_e = 5$ and $N_s = 200$ symbols).

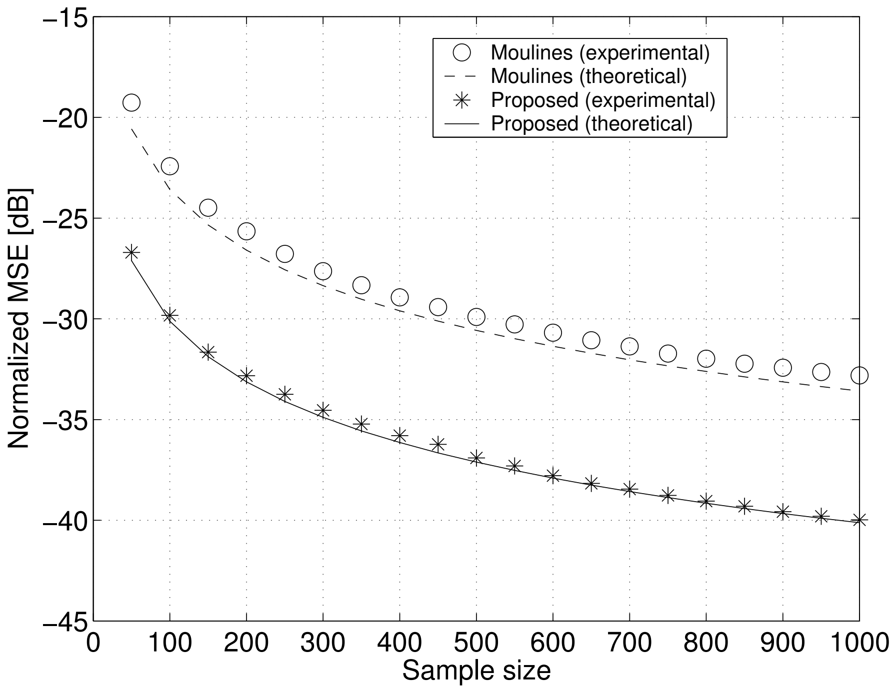

Figure 2: Normalized channel MSE versus N_s (in symbols) ($L_e = 5$ and SNR $= 20$ dB).

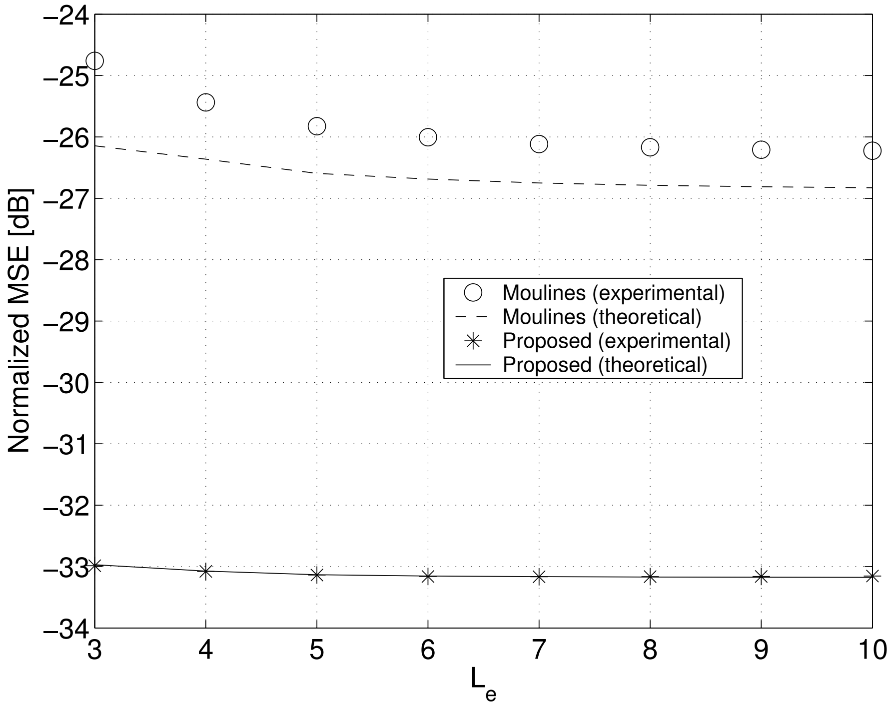

Figure 3: Normalized channel MSE versus L_e (in symbols) ($N_s = 200$ and SNR $= 20$ dB).

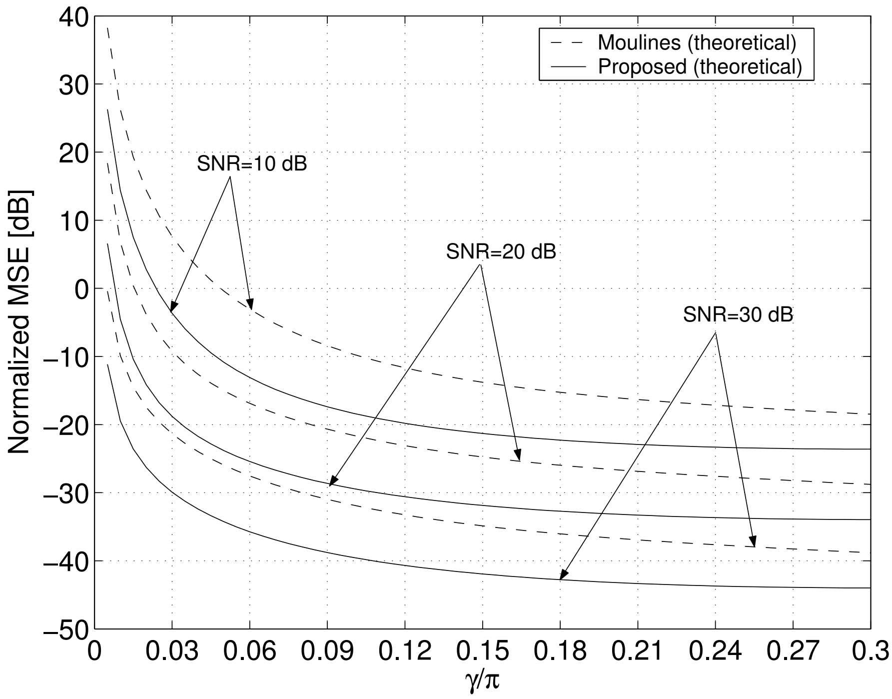

Figure 4: Normalized channel MSE versus γ/π for different values of SNR ($L_e = 5$ and $N_s = 200$).

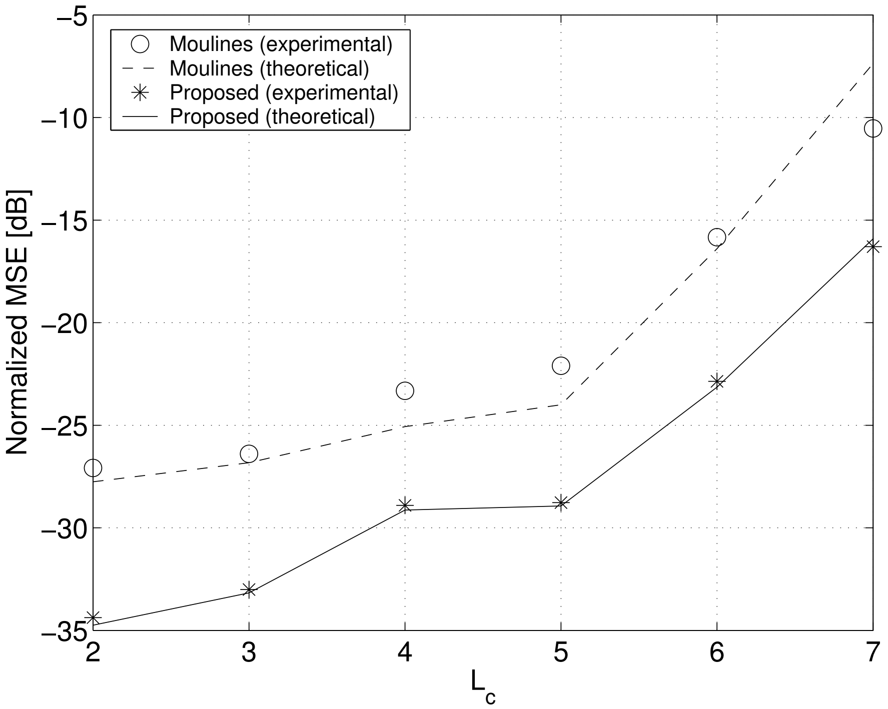

Figure 5: Normalized channel MSE versus L_c ($L_e = 10$, $N_s = 200$ and SNR $= 20$ dB).