

RESEARCH

Open Access



# On the use of the channel second-order statistics in MMSE receivers for time- and frequency-selective MIMO transmission systems

Manuel A. Vázquez<sup>1,3\*</sup>  and Joaquín Míguez<sup>2,3</sup>

## Abstract

Equalization of unknown frequency- and time-selective multiple input multiple output (MIMO) channels is often carried out by means of decision feedback receivers. These consist of a channel estimator and a linear filter (for the estimation of the transmitted symbols), interconnected by a feedback loop through a symbol-wise threshold detector. The linear filter is often a minimum mean square error (MMSE) filter, and its mathematical expression involves second-order statistics (SOS) of the channel, which are usually ignored by simply assuming that the channel is a known (deterministic) parameter given by an estimate thereof. This appears to be suboptimal and in this work we investigate the kind of performance gains that can be expected when the MMSE equalizer is obtained using SOS of the channel process. As a result, we demonstrate that improvements of several dBs in the signal-to-noise ratio needed to achieve a prescribed symbol error rate are possible.

**Keywords:** MIMO, MMSE, Joint channel and data estimation, Second-order statistics

## 1 Introduction

The main appeal in using a multiple input multiple output (MIMO) wireless communication system stems from the fact that the channel capacity increases linearly with the minimum between the number of transmitting antennas and that of receiving antennas [1]. Unfortunately, the complexity of optimal MIMO detectors (which minimize the probability of either symbol or sequence detection errors) grows exponentially with the number of input streams and the order of the channel, if the latter is frequency-selective [2]. Therefore, suboptimal equalization algorithms that avoid this computational burden are needed in order to take advantage, in a practical setup, of the increase in capacity that a MIMO channel can offer. Additionally, in most real-world scenarios, the channel is unknown and must be estimated prior to data detection. A decision

feedback equalizer (DFE) type of receiver [3–7] is then an appealing choice due to its ease of implementation and the good trade-off between computational complexity and performance that it achieves.

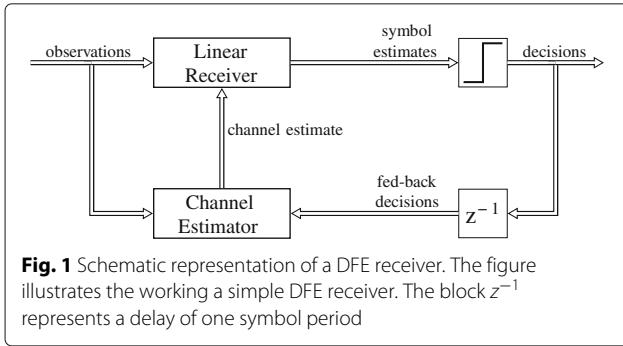
Figure 1 shows a simple DFE scheme. The main blocks are an adaptive channel estimation algorithm and a linear filter. The latter takes a channel estimate and the observations at the receiver front-end to produce linear estimates of the transmitted symbols. These estimates are either real or complex numbers, depending on the modulation format. A threshold detector is used to convert them into hard symbol decisions, i.e., discrete estimates chosen from the symbol alphabet based on a minimum distance rule. We assume that the detector operates symbol-wise in order to keep the computational effort limited. The decisions at the output of the detector are fed back to the channel estimation block, so that they can be used to improve the subsequent channel estimates. Usually, the detected symbols are also employed to cancel intersymbol interference (see, e.g., [8] and Section 3 in this article), but this is omitted in the figure for simplicity. We

\*Correspondence: mvazquez@tsc.uc3m.es

<sup>1</sup>Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Avenida de la Universidad 30, Leganés, Spain

<sup>3</sup>Gregorio Marañón Health Research Institute, Calle del Dr. Esquerdo, 46, 28007 Madrid, Spain

Full list of author information is available at the end of the article



remark that the receiver is nonlinear, due to the thresholding operation, yet its computational complexity is similar to that of a linear receiver combined with an adaptive channel estimation algorithm [9], ([7] Chapter 16).

Any linear filter is amenable to be used in a DFE structure. However, the use of the minimum mean square error (MMSE) filter has become widespread because it offers an attractive trade-off between noise amplification and interference cancellation [10]. Many joint data detection and channel estimation algorithms (see, e.g., [11–14]) rely on the linear MMSE filter to carry out the equalization of an unknown MIMO channel. In every case, and to the best of our knowledge, the *point estimates* of the channel impulse response (CIR) provided by the corresponding estimator are used by the MMSE filter as if they were the *true* CIR. However, the estimates are actually statistics of the true channel and, as such, have a mean and a covariance matrix, the latter measuring the uncertainty we have about their accuracy. When the Kalman filter (KF) [15, 16] is used to estimate the channel, both statistics, the mean and the covariance matrix of the channel become available, but even then the usual approach (e.g., in [11, 17]) consists in taking the mean of the estimate as if it were the true channel and ignore the information provided by the covariance matrix. In this paper, we argue that significant performance gains can be expected by taking advantage of the second-order statistics (SOS) of the channel, with a low impact on the computational complexity of the receiver. To be specific, we show that reductions of several dBs in the signal-to-noise ratio (SNR) needed to attain a prescribed symbol error rate (SER) can be achieved using the proposed scheme. The main contributions of this work are the design and implementation of the new linear MMSE equalizer that exploits the second-order statistics of the channel, as well as extensive computer simulations showing the gains that can be expected from the proposed method as compared to the conventional one.

The remaining of this paper is organized as follows. In Section 2, the discrete-time baseband equivalent signal model of a MIMO transmission system with frequency- and time-selective channel is described. The standard

linear MMSE equalizer is briefly reviewed in Section 3. Our extension thereof using the channel SOS is introduced in Section 4. Section 5 outlines the DFE schemes resulting from the investigated equalizers. In Section 6, we show and discuss the results of extensive computer simulations to compare the performance of the conventional DFE MMSE receiver (that ignores the channel SOS) and the proposed DFE SOS-MMSE scheme. Finally, Section 7 is devoted to the conclusions.

### 1.1 Summary of notation

Given a time-indexed sequence of (column) vectors,  $\mathbf{x}_a, \mathbf{x}_{a+1}, \dots, \mathbf{x}_b$ , we denote by  $\mathbf{x}_b^a$  the column vector constructed by stacking, in order, all the vectors between  $\mathbf{x}_a$  and  $\mathbf{x}_b$  (including both), i.e.,

$$\mathbf{x}_b^a = \begin{bmatrix} \mathbf{x}_a^\top & \mathbf{x}_{a+1}^\top & \dots & \mathbf{x}_b^\top \end{bmatrix}^\top.$$

An identity matrix of order  $k$  is denoted by  $\mathbf{I}_k$ , whereas  $\mathbf{0}_{N,M}$  is an  $N \times M$  all-zeros matrix. If  $N = M$ , we simply write  $\mathbf{0}_N$ . For a vector  $\mathbf{x}$ ,  $x(i)$  represents its  $i$ th element, and given a matrix  $\mathbf{A}$ ,  $\mathbf{a}(i)$  refers to  $i$ th column.

## 2 Signal model

### 2.1 Time- and frequency-selective MIMO channel

We consider a MIMO communication system with  $N_t$  transmitting antennas and  $N_r$  receiving antennas separated by a time- and frequency-selective (MIMO) channel. The discrete-time baseband-equivalent model describing the transmission can be written as (see, e.g., [18])

$$\mathbf{y}_t = \sum_{i=0}^{m-1} \mathbf{H}_t(i) \mathbf{s}_{t-i} + \mathbf{g}_t, \quad (1)$$

where  $\mathbf{y}_t$  is a  $N_r \times 1$  vector containing the observations collected at time  $t$ ,  $m$  is the number of taps (usually referred to as the order) of the frequency-selective MIMO channel,  $\mathbf{H}_t(i)$  is the (time-varying)  $N_r \times N_t$  channel matrix associated with the  $i$ th tap,  $\mathbf{s}_t$  is a vector of size  $N_t \times 1$  comprising the symbols transmitted at time  $t$ , and  $\mathbf{g}_t$  is an  $N_r \times 1$  vector of independent additive white Gaussian noise (AWGN) components with zero mean and variance  $\sigma_g^2$ .

Grouping the matrices associated with the different taps of the channel,  $\mathbf{H}_t(i), i = 0, \dots, m - 1$ , in a single *overall* channel matrix,

$$\mathbf{H}_t = [\mathbf{H}_t(m-1) \quad \mathbf{H}_t(m-2) \quad \dots \quad \mathbf{H}_t(0)], \quad (2)$$

of size  $N_r \times N_t m$ , allows Eq. (1) to be written in a more compact form as

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{s}_{t-m+1} + \mathbf{g}_t, \quad (3)$$

where  $\mathbf{s}_{t-m+1}$  is a  $N_t m \times 1$  vector that stacks all the symbol vectors involved in the  $t$ th observation,

$$\mathbf{s}_{t-m+1} = \begin{bmatrix} \mathbf{s}_{t-m+1}^\top & \mathbf{s}_{t-m+2}^\top & \dots & \mathbf{s}_t^\top \end{bmatrix}^\top. \quad (4)$$

While non-standard, the notation in (4) shows explicitly that vector  $\mathbf{s}_{t-m+1}$  is constructed by stacking simpler vectors in order and indicates the time indexes of the first ( $t - m + 1$  in this case) and last ( $t$  here) elements to be stacked. These features should ease the understanding of some formulas in the sequel.

The evolution of the channel is modeled by means of an autoregressive (AR) process driven by white Gaussian noise<sup>1</sup> [11]. For the sake of generality, we consider an AR process of order  $R$ , whose analytical description is given by

$$\mathbf{H}_t = \sum_{r=1}^R a_r \mathbf{H}_{t-r} + \mathbf{V}_t, \quad (5)$$

where  $a_r, r = 1, \dots, R$ , are the coefficients of the process and  $\mathbf{V}_t$  is a  $N_r \times N_t m$  matrix with independent and identically distributed (i.i.d.) Gaussian random variables (r.v.) with zero mean and variance  $\sigma_v^2$ .

Equations (3) and (5) can be seen, respectively, as the observation and state equations of a random dynamic system in state-space form. Since both equations are linear and the corresponding noise processes are Gaussian, the Kalman filter (KF) can be applied to exactly compute the posterior probability distribution of the time-varying MIMO channel when the symbols are available.

In order to do so while using the standard KF equations, we first need to gather the whole state of the system (here, the channel at the last  $R$  time instants) in a single vector and rewrite the state and observation equations in terms of it. Matrix  $\mathbf{H}_t$  can be represented as a vector in a straightforward manner by, e.g., stacking all its columns one upon another. In particular, if we let  $\mathbf{h}_t(j)$  denote the  $j$ th column of matrix  $\mathbf{H}_t$ , then the  $N_r N_t m \times 1$  vector

$$\mathbf{h}_t = \left[ \mathbf{h}_t(1)^\top \quad \mathbf{h}_t(2)^\top \quad \dots \quad \mathbf{h}_t(N_t m)^\top \right]^\top \quad (6)$$

contains the same coefficients as matrix  $\mathbf{H}_t$ . Using this vectorial notation and taking into account that, according to Eq. (5), the channel at a certain time instant depends on the channel at the  $R$  previous time instants, the state of the system at time  $t$  can be represented by the vector

$$\mathbf{h}_{t-R+1} = \left[ \mathbf{h}_{t-R+1}^\top \quad \mathbf{h}_{t-R+2}^\top \quad \dots \quad \mathbf{h}_t^\top \right]^\top. \quad (7)$$

The state equation of the system can then be written in terms of this *augmented* channel vector as

$$\mathbf{h}_{t-R+1} = \mathbf{Q} \mathbf{h}_{t-1} + \mathbf{v}_t, \quad (8)$$

where  $\mathbf{v}_t$  is a  $N_r N_t m R \times 1$  vector with i.i.d. Gaussian r.v.'s of zero mean and variance  $\sigma_v^2$  in the last  $N_r N_t m$  positions and zeros in the rest, and the state transition matrix,  $\mathbf{Q}$ , is defined as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{0}_{N_r N_t m} & \mathbf{I}_{N_r N_t m} & \mathbf{0}_{N_r N_t m} & \dots & \mathbf{0}_{N_r N_t m} \\ \mathbf{0}_{N_r N_t m} & \mathbf{0}_{N_r N_t m} & \mathbf{I}_{N_r N_t m} & \mathbf{0}_{N_r N_t m} & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0}_{N_r N_t m} \\ \mathbf{0}_{N_r N_t m} & \mathbf{0}_{N_r N_t m} & \dots & \mathbf{0}_{N_r N_t m} & \mathbf{I}_{N_r N_t m} \\ a_1 \mathbf{I}_{N_r N_t m} & a_2 \mathbf{I}_{N_r N_t m} & \dots & \dots & a_R \mathbf{I}_{N_r N_t m} \end{bmatrix}. \quad (9)$$

with  $\mathbf{0}_{N_r N_t m}$  denoting an  $N_r N_t m \times N_r N_t m$  all-zeros matrix, and  $\mathbf{I}_{N_r N_t m}$  an identity matrix of order  $N_r N_t m$ .

The observation (Eq. (3)) can also be easily rewritten in terms of the augmented channel vector,  $\mathbf{h}_{t-R+1}$ , as

$$\mathbf{y}_t = \mathbf{S}_t \mathbf{h}_{t-R+1} + \mathbf{g}_t \quad (10)$$

where

$$\mathbf{S}_t = \begin{bmatrix} \mathbf{0}_{N_r \times N_r N_t m (R-1)} & s_{t-m+1}(1) \mathbf{I}_{N_r} & s_{t-m+1}(2) \mathbf{I}_{N_r} & \dots \\ & & & & s_{t-m+1}(N_t m) \mathbf{I}_{N_r} \end{bmatrix}. \quad (11)$$

We use the dynamic system in state-space form specified by Eqs. (10) and (8) (which is equivalent to that given by Eqs. (3) and (5)) to track the unknown time-varying MIMO channel by means of a KF.

## 2.2 Stacked model

When a channel is time dispersive, a reliable detection of the transmitted symbols usually requires smoothing. It entails taking into account the observations  $\mathbf{y}_{t:t+d}$  (the parameter  $d \geq 1$  being the smoothing lag) in order to detect the vector  $\mathbf{s}_t$  containing the symbols transmitted at time  $t$ . In such case, it is useful to consider an equation that relates a tall vector of stacked observations with the transmitted symbols, namely,

$$\mathbf{y}_{t+d}^i = \bar{\mathbf{H}}_{t,d} \mathbf{s}_{t-m+1} + \mathbf{g}_{t+d}^i, \quad (12)$$

where  $\mathbf{y}_{t+d}^i = \left[ \mathbf{y}_t^\top, \mathbf{y}_{t+1}^\top, \dots, \mathbf{y}_{t+d}^\top \right]^\top$ , and the  $N_r(d+1) \times N_t(m+d)$  composite channel matrix is defined as

$$\bar{\mathbf{H}}_{t,d} = \begin{bmatrix} \mathbf{H}_t(m-1) & \mathbf{H}_t(m-2) & \dots & \mathbf{H}_t(0) & \mathbf{0}_{N_r \times N_t} & \dots & \mathbf{0}_{N_r \times N_t} \\ \mathbf{0}_{N_r \times N_t} & \mathbf{H}_{t+1}(m-1) & \dots & \mathbf{H}_{t+1}(1) & \mathbf{H}_{t+1}(0) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \mathbf{0}_{N_r \times N_t} \\ \mathbf{0}_{N_r \times N_t} & \dots & \mathbf{0}_{N_r \times N_t} & \mathbf{H}_{t+d}(m-1) & \dots & \dots & \mathbf{H}_{t+d}(0) \end{bmatrix}. \quad (13)$$

Equation (12) involves symbol vectors  $\mathbf{s}_{t-m+1}, \dots, \mathbf{s}_{t-1}$ , which, at time  $t$ , have already been detected. It is convenient to identify their contribution to the stacked observations vector,  $\mathbf{y}_{t+d}^i$ . Let us decompose the overall channel matrix  $\bar{\mathbf{H}}_{t,d}$  as

$$\bar{\mathbf{H}}_{t,d} = \left[ \mathbf{H}_{t,d}^\ddagger \quad \mathbf{H}_{t,d} \right], \quad (14)$$

where the submatrices  $\mathbf{H}_{t,d}^\ddagger$  and  $\mathbf{H}_{t,d}$  encompass, respectively, the first  $N_t(m - 1)$  and last  $N_t(d + 1)$  columns of  $\overline{\mathbf{H}}_{t,d}$ . Then, the vector of stacked observations can be rewritten as

$$\begin{aligned} \mathbf{y}_{t+d} &= \begin{bmatrix} \mathbf{H}_{t,d}^\ddagger & \mathbf{H}_{t,d} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{t-m+1} \\ \mathbf{s}_{t+d} \end{bmatrix} + \mathbf{g}_{t+d} \\ &= \mathbf{H}_{t,d}^\ddagger \mathbf{s}_{t-m+1} + \mathbf{H}_{t,d} \mathbf{s}_{t+d} + \mathbf{g}_{t+d}, \end{aligned} \quad (15)$$

where the term  $\mathbf{H}_{t,d}^\ddagger \mathbf{s}_{t-m+1}$  contains the contribution of the symbols transmitted up to time  $t - 1$ , and can be treated as causal inter-symbol interference.

### 2.3 Kalman filtering

The KF [15] provides the optimal solution to the problem of estimating the state of a dynamic system in state-space form when its state and observation equations are linear and their corresponding noises are Gaussian.

In the problem at hand, the state of the system at time  $t$  is given by the augmented channel vector,  $\mathbf{h}_{t-R+1}$ , and Eqs. (8) and (10) can be seen as, respectively, the state and observation equations of a dynamic system in state-space form. Since the above constraints of linearity and Gaussianity are met, the KF can be used to compute the probability density function of the state conditional on the available observations,  $p(\mathbf{h}_{t-R+1} | \mathbf{y}_0, \dots, \mathbf{y}_t)$ . However, the observation equation involves knowing, at time  $t$ , matrix  $\mathbf{S}_t$ , which includes all the symbols transmitted between time instants  $t - m + 1$  and  $t$ . In practice, only estimates of the symbols transmitted up to time  $t - 1$  are available at time  $t$ , and hence we aim at the (*predictive*) distribution of the state conditional on all the past observations and previously detected symbols, i.e.,  $p(\mathbf{h}_{t-R+1} | \mathbf{y}_0, \dots, \mathbf{y}_{t-1}, \tilde{\mathbf{s}}_0, \dots, \tilde{\mathbf{s}}_{t-1})$  with  $\tilde{\mathbf{s}}_t$  denoting the vector containing the hard estimates of the symbols in  $\mathbf{s}_t$ . Every expectation in the remaining of the paper is also (implicitly) conditional on the same information, and we denote it as  $\mathbb{E}_{t-1}[\cdot]$ . For example, the posterior mean of the CIR at time  $t + k$  ( $k \geq 0$ ) conditional on  $\mathbf{y}_0, \dots, \mathbf{y}_{t-1}$  and  $\tilde{\mathbf{s}}_0, \dots, \tilde{\mathbf{s}}_{t-1}$  is written as  $\mathbb{E}_{t-1}[\mathbf{h}_{t+k}]$  and the posterior cross-covariance between  $\mathbf{h}_{t+k}$  and  $\mathbf{h}_{t+k'}$ ,  $k, k' \geq 0$ , is denoted  $\mathbb{E}_{t-1}[\mathbf{h}_{t+k} \mathbf{h}_{t+k'}^H]$ .

Notice that here the KF only yields an approximate solution insofar as it depends on the goodness of the previously detected symbols fed to it.

### 3 Linear MMSE smoothing

In order to detect the symbols transmitted at time  $t$  over a frequency-selective channel, it is usually a good approach to first remove the contribution of the already detected symbols from the observations vector [8]. At the sight of

Eq. (15), we can obtain causal-interference-free observations as

$$\mathbf{z}_{t+d} := \mathbf{y}_{t+d} - \mathbf{H}_{t,d}^\ddagger \mathbf{s}_{t-m+1} \quad (16)$$

$$= \mathbf{H}_{t,d} \mathbf{s}_{t+d} + \mathbf{g}_{t+d}. \quad (17)$$

Computing  $\mathbf{z}_{t+d}$  from  $\mathbf{y}_{t+d}$  entails knowing vector  $\mathbf{s}_{t-m+1}$ , which encompasses symbol vectors  $\mathbf{s}_{t-m+1}, \dots, \mathbf{s}_{t-1}$ . These are unknown but previous estimates thereof are available at time  $t$  and can be used as a surrogate. Hence, in practice, the stacked symbols vector  $\mathbf{s}_{t-m+1}$  in Eq. (16) is replaced with vector,  $\tilde{\mathbf{s}}_{t-m+1}$  that contains hard estimates of the same symbols. This is a common approximation for the design of DFES, and it usually makes sense under the assumption that the receiver is operating with a sufficiently low symbol error probability. Throughout the paper, we rely on this approximation, which amounts to taking the previously detected symbols as if they were the truly transmitted symbols, i.e.,

$$\tilde{\mathbf{s}}_{t-m+1} = \mathbf{s}_{t-m+1}. \quad (18)$$

Assuming the causal interference is properly canceled, the linear MMSE estimation of the symbols transmitted at time  $t$  considering the observations up to time  $t + d$  can be easily derived from Eq. (17) (see, e.g., [19]). In particular, let the  $N_r(d + 1) \times N_t(d + 1)$  matrix  $\mathbf{F}_t$  represent the response of a linear system. Then, estimates of the transmitted symbols are computed as <sup>2</sup>

$$\hat{\mathbf{s}}_{t+d} = \mathbf{F}_t^H \mathbf{z}_{t+d}, \quad (19)$$

and, in order to minimize the mean square error of these estimates, the response matrix can be computed by solving the optimization problem

$$\mathbf{F}_t = \arg \min_{\mathbf{F}_t} \mathbb{E}_{t-1} \left[ \left| \mathbf{F}_t^H \mathbf{z}_{t+d} - \mathbf{s}_{t+d} \right|^2 \right]. \quad (20)$$

Since the ultimate aim is to estimate  $\mathbf{s}_t$  but we are using observations up to time  $t + d$ , we refer to the linear system whose response is given by  $\mathbf{F}_t$  in (20) as an MMSE smoother.

Equation (20) poses a quadratic optimization problem and it is straightforward to obtain the closed-form solution (see, e.g., [2])

$$\mathbf{F}_t^H = \mathbb{E}_{t-1} \left[ \begin{bmatrix} \mathbf{s}_{t+d} & \mathbf{z}_{t+d}^H \end{bmatrix} \left( \mathbb{E}_{t-1} \left[ \begin{bmatrix} \mathbf{z}_{t+d} & \mathbf{z}_{t+d}^H \end{bmatrix} \right] \right)^{-1} \right]. \quad (21)$$

Again, if we assume that the causal inter-symbol interference has been completely removed from the

observations, so that these are given by Eq. (17), then the expectations on the right-hand side of Eq. (21) can be shown to be

$$\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{s}_t & \mathbf{z}_t^H \\ \mathbf{z}_{t+d} & \mathbf{z}_{t+d}^H \end{bmatrix} = \sigma_s^2 \mathbb{E}_{t-1} [\mathbf{H}_{t,d}^H] \quad (22)$$

$$\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{z}_t & \mathbf{z}_t^H \\ \mathbf{z}_{t+d} & \mathbf{z}_{t+d}^H \end{bmatrix} = \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d} \mathbf{s}_t & \mathbf{s}_t^H \mathbf{H}_{t,d}^H \\ \mathbf{H}_{t,d} \mathbf{s}_{t+d} & \mathbf{s}_{t+d}^H \mathbf{H}_{t,d}^H \end{bmatrix} + \sigma_g^2 \mathbf{I}_{N_r(d+1)}, \quad (23)$$

where  $\sigma_s^2$  denotes the variance of the symbols and it has been taken into account that the noise at time  $t$  is white and independent of the channel process and the symbols transmitted up to  $t - 1$ , and that the channel and the symbols are a priori independent.

The expectation on the right-hand side of Eq. (23) is usually approximated, to the best of our knowledge, by dealing with the channel matrix  $\mathbf{H}_{t,d}$  as if it were a known given (deterministic) parameter, and hence

$$\begin{aligned} \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d} \mathbf{s}_t & \mathbf{s}_t^H \mathbf{H}_{t,d}^H \\ \mathbf{H}_{t,d} \mathbf{s}_{t+d} & \mathbf{s}_{t+d}^H \mathbf{H}_{t,d}^H \end{bmatrix} &= \mathbf{H}_{t,d} \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{s}_t & \mathbf{s}_t^H \\ \mathbf{s}_{t+d} & \mathbf{s}_{t+d}^H \end{bmatrix} \mathbf{H}_{t,d}^H \\ &= \sigma_s^2 \mathbf{H}_{t,d} \mathbf{H}_{t,d}^H. \end{aligned} \quad (24)$$

Substituting (24) into Eq. (23) yields

$$\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{z}_t & \mathbf{z}_t^H \\ \mathbf{z}_{t+d} & \mathbf{z}_{t+d}^H \end{bmatrix} = \sigma_s^2 \mathbf{H}_{t,d} \mathbf{H}_{t,d}^H + \sigma_g^2 \mathbf{I}_{N_r(d+1)}, \quad (25)$$

and combining Eqs. (22) and (25) in Eq. (21) yields the final expression for the response of the conventional linear MMSE smoother,

$$\mathbf{F}_t^H = \sigma_s^2 \mathbf{H}_{t,d}^H \left( \sigma_s^2 \mathbf{H}_{t,d} \mathbf{H}_{t,d}^H + \sigma_g^2 \mathbf{I}_{N_r(d+1)} \right)^{-1}. \quad (26)$$

So far, we have assumed the channel is a known (deterministic) parameter. However, this is not usually the case in practice, and the common approach to tackle this problem consists in replacing, whenever necessary, the (true) channel matrix with its expectation. Notice that here this entails a twofold approximation. On one hand, even when assuming that the symbols up to time  $t - 1$  have been detected exactly, at best one can only obtain *approximate* causal-interference-free observations as

$$\mathbf{z}_{t+d} \approx \mathbf{y}_{t+d} - \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d}^\dagger \\ \mathbf{H}_{t,d}^\dagger \end{bmatrix} \mathbf{s}_{t-m+1}. \quad (27)$$

On the other hand, taking the true channel matrix in Eq. (26) to be equal to its expectation results in the following approximation for the linear MMSE filter

$$\mathbf{F}_t^H \approx \sigma_s^2 \mathbb{E}_{t-1} [\mathbf{H}_{t,d}^H] \left( \sigma_s^2 \mathbb{E}_{t-1} [\mathbf{H}_{t,d}] \mathbb{E}_{t-1} [\mathbf{H}_{t,d}^H] + \sigma_g^2 \mathbf{I}_{N_r(d+1)} \right)^{-1}. \quad (28)$$

At the sight of Eqs. (13) and (14), in order to obtain  $\mathbb{E}_{t-1} [\mathbf{H}_{t,d}^\dagger]$  and  $\mathbb{E}_{t-1} [\mathbf{H}_{t,d}^H]$  on the right-hand side of

Eqs. (27) and (28), respectively, we need the expectations of the matrices  $\mathbf{H}_t, \mathbf{H}_{t+1}, \dots, \mathbf{H}_{t+d}$ . At time  $t$ , the expectation of the channel matrix  $\mathbf{H}_t$  is given by the predictive distribution of the KF, which takes into account the observations and symbols vectors up to time  $t - 1$ . However, the expectations of the matrices  $\mathbf{H}_{t+1}, \dots, \mathbf{H}_{t+d}$  have to be computed as well. In order to do so, we simply use Eq. (5) (the state equation of the system) to expand the expected channel matrix at time  $t$ , i.e.,

$$\mathbb{E}_{t-1} [\mathbf{H}_{t+k}] = \sum_{r=1}^R a_r \mathbb{E}_{t-1} [\mathbf{H}_{t+k-r}], \quad k = 1, \dots, d. \quad (29)$$

#### 4 MMSE smoothing using the channel SOS

The proposed MMSE detector treats the channel as an unknown (multidimensional) random variable (as opposed to a deterministic known parameter), and takes advantage of its second-order statistics rather than just its expectation. Additionally, it avoids performing *explicit* interference cancelation, since this cannot be performed exactly. In order to do so, it aims to detect the transmitted symbols by solving the optimization problem

$$\mathbf{F}_t = \arg \min_{\mathbf{F}_t} \mathbb{E}_{t-1} \left[ \left| \mathbf{F}_t^H \mathbf{y}_{t+d} - \mathbf{s}_{t+d} \right|^2 \right], \quad (30)$$

which is exactly the same as that posed by Eq. (20) replacing  $\mathbf{z}_{t+d}$  with  $\mathbf{y}_{t+d}$ . Hence, the solution is

$$\mathbf{F}_t^H = \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{s}_{t+d} & \mathbf{y}_{t+d} \end{bmatrix} \left( \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{y}_{t+d} & \mathbf{y}_{t+d}^H \end{bmatrix} \right)^{-1}. \quad (31)$$

Through straightforward algebraic manipulation, one can show

$$\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{s}_{t+d} & \mathbf{y}_{t+d}^H \end{bmatrix} = \sigma_s^2 \mathbb{E}_{t-1} [\mathbf{H}_{t,d}^H] \quad (32)$$

$$\begin{aligned} \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{y}_{t+d} & \mathbf{y}_{t+d}^H \end{bmatrix} &= \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d} \mathbf{s}_t & \mathbf{s}_t^H \mathbf{H}_{t,d}^H \\ \mathbf{H}_{t,d} \mathbf{s}_{t+d} & \mathbf{s}_{t+d}^H \mathbf{H}_{t,d}^H \end{bmatrix} \\ &+ \mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d}^\dagger \mathbf{s}_{t-m+1} & \mathbf{s}_{t-m+1}^H \mathbf{H}_{t,d}^{\dagger H} \\ \mathbf{H}_{t,d}^\dagger \mathbf{s}_{t-m+1} & \mathbf{s}_{t-m+1}^H \mathbf{H}_{t,d}^{\dagger H} \end{bmatrix} \\ &+ \sigma_g^2 \mathbf{I}_{N_r(d+1)}, \end{aligned} \quad (33)$$

where we have used that vector  $\mathbf{s}_{t-m+1}$  is known because the expectations are conditional on the previously detected symbols and we are assuming these match the truly transmitted ones (see Eq. (18) and the surrounding discussion).

##### 4.1 The observation autocorrelation matrix

From Eq. (33), computing  $\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{y}_{t+d} & \mathbf{y}_{t+d}^H \end{bmatrix}$  actually amounts to the calculation of  $\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d} \mathbf{s}_t & \mathbf{s}_t^H \mathbf{H}_{t,d}^H \\ \mathbf{H}_{t,d} \mathbf{s}_{t+d} & \mathbf{s}_{t+d}^H \mathbf{H}_{t,d}^H \end{bmatrix}$  and

$\mathbb{E}_{t-1} \left[ \mathbf{H}_{t,d}^\dagger \mathbf{s}_{t-m+1} \mathbf{s}_{t-m+1}^H \mathbf{H}_{t,d}^{\dagger H} \right]$ . Regarding the first expectation, if we let  $\mathbf{h}_{t,d}(j)$  denote the  $j$ th column of matrix  $\mathbf{H}_{t,d}$  and  $s_{t+d}(j)$  denote the  $j$ th element within vector  $\mathbf{s}_{t+d}$ , then the expectation  $\mathbb{E}_{t-1} \left[ \mathbf{H}_{t,d} \mathbf{s}_{t+d} \mathbf{s}_{t+d}^H \mathbf{H}_{t,d}^H \right]$  can be rewritten as

$$\begin{aligned} \mathbb{E}_{t-1} \left[ \mathbf{H}_{t,d} \mathbf{s}_{t+d} \mathbf{s}_{t+d}^H \mathbf{H}_{t,d}^H \right] &= \mathbb{E}_{t-1} \left[ \left( \mathbf{H}_{t,d} \mathbf{s}_{t+d} \right) \left( \mathbf{H}_{t,d} \mathbf{s}_{t+d} \right)^H \right] \\ &= \mathbb{E}_{t-1} \left[ \left( \sum_{i=1}^{N_t(d+1)} \mathbf{h}_{t,d}(i) s_{t+d}(i) \right) \right. \\ &\quad \left. \times \left( \sum_{j=1}^{N_t(d+1)} \mathbf{h}_{t,d}^H(j) s_{t+d}^*(j) \right) \right] \\ &= \sum_{i=1}^{N_t(d+1)} \sum_{j=1}^{N_t(d+1)} \mathbb{E}_{t-1} [\mathbf{h}_{t,d}(i) \mathbf{h}_{t,d}^H(j)] \\ &\quad \mathbb{E}_{t-1} \left[ s_{t+d}(i) s_{t+d}^*(j) \right] \\ &= \sigma_s^2 \sum_{i=1}^{N_t(d+1)} \mathbb{E}_{t-1} [\mathbf{h}_{t,d}(i) \mathbf{h}_{t,d}^H(i)], \end{aligned} \quad (34)$$

where the third equality follows because the symbols from time  $t$  onwards are a priori independent of the channel at time  $t$  and subsequent time instants, while the fourth equality holds because of the (also a priori) independence between different symbols,

$$\mathbb{E}_{t-1} \left[ s_{t+d}(i) s_{t+d}^*(j) \right] = \begin{cases} \sigma_s^2, & i = j \\ 0, & i \neq j \end{cases}. \quad (35)$$

Similarly, if  $\mathbf{h}_{t,d}^\dagger(i)$  refers to the  $i$ th column in matrix  $\mathbf{H}_{t,d}^\dagger$ , and  $s_{t-m+1}(i)$  to the  $i$ th symbol within vector  $\mathbf{s}_{t-m+1}$  (assumed known), we have

$$\begin{aligned} \mathbb{E}_{t-1} \left[ \mathbf{H}_{t,d}^\dagger \mathbf{s}_{t-m+1} \mathbf{s}_{t-m+1}^H \mathbf{H}_{t,d}^{\dagger H} \right] &= \mathbb{E}_{t-1} \left[ \left( \mathbf{H}_{t,d}^\dagger \mathbf{s}_{t-m+1} \right) \left( \mathbf{H}_{t,d}^\dagger \mathbf{s}_{t-m+1} \right)^H \right] \\ &= \mathbb{E}_{t-1} \left[ \left( \sum_{i=1}^{N_t(m-1)} \mathbf{h}_{t,d}^\dagger(i) s_{t-m+1}(i) \right) \right. \\ &\quad \left. \times \left( \sum_{j=1}^{N_t(m-1)} s_{t-m+1}^*(j) \mathbf{h}_{t,d}^{\dagger H}(j) \right) \right] \\ &= \sum_{i=1}^{N_t(m-1)} \sum_{j=1}^{N_t(m-1)} s_{t-m+1}(i) s_{t-m+1}^*(j) \\ &\quad \mathbb{E}_{t-1} \left[ \mathbf{h}_{t,d}^\dagger(i) \mathbf{h}_{t,d}^{\dagger H}(j) \right]. \end{aligned} \quad (36)$$

where, once again, we have used that the expectation is conditional on all the previously detected symbols and

hence, assuming these were exactly detected, vector  $\mathbf{s}_{t-m+1}$  is known.

#### 4.2 Channel cross-correlation matrices

Equations (34) and (36) involve computing the cross-correlation between different columns of matrices  $\mathbf{H}_{t,d}$  and  $\mathbf{H}_{t,d}^\dagger$ , respectively. These are submatrices of  $\bar{\mathbf{H}}_{t,d}$  (see Eq. (14)), and hence their columns are ultimately columns from  $\bar{\mathbf{H}}_{t,d}$ . In particular, if we let  $\bar{\mathbf{h}}_{t,d}(j)$  denote the  $j$ th column of matrix  $\bar{\mathbf{H}}_{t,d}$ , then

$$\mathbf{h}_{t,d}(j) = \bar{\mathbf{h}}_{t,d}(j + N_t(m - 1)), \quad 1 \leq j \leq N_t(d + 1) \quad (37)$$

$$\mathbf{h}_{t,d}^\dagger(j) = \bar{\mathbf{h}}_{t,d}^\dagger(j), \quad 1 \leq j \leq N_t(m - 1), \quad (38)$$

and every required cross-correlation is ultimately between columns of  $\bar{\mathbf{H}}_{t,d}$ . The structure of a column from the latter can be inferred from Eq. (13). Specifically, the  $j$ th column of  $\bar{\mathbf{H}}_{t,d}$  is given by

$$\bar{\mathbf{h}}_{t,d}(j) = \begin{bmatrix} \check{\mathbf{h}}_t(j) \\ \check{\mathbf{h}}_{t+1}(j) \\ \vdots \\ \check{\mathbf{h}}_{t+d}(j) \end{bmatrix}, \quad (39)$$

where

$$\check{\mathbf{h}}_{t+i}(j) = \begin{cases} \mathbf{h}_{t+i}(j - iN_t), & 0 \leq i \leq d, \quad iN_t < j \leq (i + m)N_t \\ \mathbf{0}_{N_r \times 1}, & \text{otherwise,} \end{cases} \quad (40)$$

is a  $N_r \times 1$  column vector.

We compute the cross-correlation between any pair of columns in  $\bar{\mathbf{H}}_{t,d}$ , by way of their means and cross-covariance matrix, as

$$\begin{aligned} \mathbb{E}_{t-1} \left[ \bar{\mathbf{h}}_{t,d}(i) \bar{\mathbf{h}}_{t,d}^H(j) \right] &= \mathbb{E}_{t-1} \left[ \check{\mathbf{h}}_{t,d}(i) \right] \mathbb{E}_{t-1} \left[ \check{\mathbf{h}}_{t,d}^H(j) \right] \\ &\quad + \Sigma_{\bar{\mathbf{h}}_{t,d}(i), \bar{\mathbf{h}}_{t,d}(j)}, \end{aligned} \quad (41)$$

where  $\Sigma_{\bar{\mathbf{h}}_{t,d}(i), \bar{\mathbf{h}}_{t,d}(j)}$  stands for the cross-covariance matrix between  $N_r(d + 1) \times 1$  (column) vectors  $\bar{\mathbf{h}}_{t,d}(i)$  and  $\bar{\mathbf{h}}_{t,d}(j)$ .

The expectation of every column,  $\check{\mathbf{h}}_{t,d}(i)$ ,  $i = 1, \dots, N_t(m + d)$ , in matrix  $\bar{\mathbf{H}}_{t,d}$  is readily available from the expectation of the entire matrix,  $\bar{\mathbf{H}}_{t,d}$ , which can be obtained in a straightforward manner as explained at the end of Section 3. As for the cross-covariance  $\Sigma_{\bar{\mathbf{h}}_{t,d}(i), \bar{\mathbf{h}}_{t,d}(j)}$ , it is given, at the sight of Eq. (39), by

$$\Sigma_{\bar{\mathbf{h}}_{t,d}(i), \bar{\mathbf{h}}_{t,d}(j)} = \begin{bmatrix} \Sigma_{\check{\mathbf{h}}_t(i), \check{\mathbf{h}}_t(j)} & \Sigma_{\check{\mathbf{h}}_t(i), \check{\mathbf{h}}_{t+1}(j)} & \cdots & \Sigma_{\check{\mathbf{h}}_t(i), \check{\mathbf{h}}_{t+d}(j)} \\ \Sigma_{\check{\mathbf{h}}_{t+1}(i), \check{\mathbf{h}}_t(j)} & \Sigma_{\check{\mathbf{h}}_{t+1}(i), \check{\mathbf{h}}_{t+1}(j)} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{\check{\mathbf{h}}_{t+d}(i), \check{\mathbf{h}}_t(j)} & \cdots & \cdots & \Sigma_{\check{\mathbf{h}}_{t+d}(i), \check{\mathbf{h}}_{t+d}(j)} \end{bmatrix}, \quad (42)$$

where  $\check{\Sigma}_{\check{\mathbf{h}}_{t+k}(i),\check{\mathbf{h}}_{t+l}(j)}$ ,  $k, l = 0, \dots, d$  is the cross-covariance between vectors  $\check{\mathbf{h}}_{t+k}(i)$  and  $\check{\mathbf{h}}_{t+l}(j)$ . In particular the entry at the  $(u, w)$  position of  $\check{\Sigma}_{\check{\mathbf{h}}_{t+k}(i),\check{\mathbf{h}}_{t+l}(j)}$  is the covariance between the  $u$ th element of  $\check{\mathbf{h}}_{t+k}(i)$  and the  $w$ th element of  $\check{\mathbf{h}}_{t+l}(j)$ . Notice that when both vectors in the subscript are the same, this yields the (self-)covariance matrix  $\check{\Sigma}_{\check{\mathbf{h}}_{t+k}(i),\check{\mathbf{h}}_{t+k}(i)} = \check{\Sigma}_{\check{\mathbf{h}}_{t+k}(i)}$ .

Recall from Eq. (40) that  $\check{\mathbf{h}}_{t+k}(i)$  for  $k = 0, \dots, d$ , and  $i = 1, \dots, N_r(d + 1)$ , is either an all-zeros (column) vector or a column from matrix  $\mathbf{H}_{t+k}$ . Thus, when computing the cross-covariance between vectors  $\check{\mathbf{h}}_{t+k}(i)$  and  $\check{\mathbf{h}}_{t+l}(j)$ , if one of the vectors (or both) is all-zeros, then  $\check{\Sigma}_{\check{\mathbf{h}}_{t+k}(i),\check{\mathbf{h}}_{t+l}(j)} = \mathbf{0}_{N_r}$ . On the contrary, if both vectors are different from the all-zeros column vector, then  $\check{\mathbf{h}}_{t+k}(i)$  must be a column from matrix  $\mathbf{H}_{t+k}$ , and  $\check{\mathbf{h}}_{t+l}(j)$  one from matrix  $\mathbf{H}_{t+l}$ . In that case, let us assume  $\check{\mathbf{h}}_{t+k}(i) = \mathbf{h}_{t+k}(n)$  and  $\check{\mathbf{h}}_{t+l}(j) = \mathbf{h}_{t+l}(q)$  for some  $n, q \in \{1, \dots, N_t m\}$ , so that  $\check{\Sigma}_{\check{\mathbf{h}}_{t+k}(i),\check{\mathbf{h}}_{t+l}(j)} = \Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l}(q)}$ .

In order to compute  $\Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l}(q)}$  for  $k, l \in \{0, \dots, d\}$  and  $n, q \in \{1, \dots, N_t m\}$ , we consider two different cases. On one hand, if  $k = l$  then  $\Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+k}(q)}$  is ultimately the cross-covariance between columns  $n$  and  $q$  of matrix  $\mathbf{H}_{t+k}$  and can be obtained from the KF. Indeed, we can make the KF evolve from time  $t$  up to time  $t + k$  when no new information is available by taking  $k$  predictive steps. This yields predictive statistics for  $\mathbf{h}_{t+k}$  (a vectorial representation of  $\mathbf{H}_{t+k}$ ), and from its covariance matrix it is straightforward to obtain the cross-covariance matrix,  $\Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+k}(q)}$  between any pair of columns  $\mathbf{h}_{t+k}(n)$  and  $\mathbf{h}_{t+k}(q)$  within  $\mathbf{H}_{t+k}$ . On the other hand, if  $k \neq l$  then

$$\Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l}(q)} = \mathbb{E}_{t-1} \left[ (\mathbf{h}_{t+k}(n) - \mathbb{E}_{t-1}[\mathbf{h}_{t+k}(n)]) (\mathbf{h}_{t+l}(q) - \mathbb{E}_{t-1}[\mathbf{h}_{t+l}(q)])^H \right] \quad (43)$$

$$= \mathbb{E}_{t-1} \left[ \mathbf{h}_{t+k}(n) \mathbf{h}_{t+l}(q)^H \right] - \mathbb{E}_{t-1}[\mathbf{h}_{t+k}(n)] \mathbb{E}_{t-1}[\mathbf{h}_{t+l}(q)^H], \quad (44)$$

and from Eqs. (43) and (44), it can be shown (see Appendix 7 for details) that, for  $k < l$ ,

$$\Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l}(q)} = \sum_{r=1}^R a_r \Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l-r}(q)}, \quad (45)$$

which allows for the recursive computation of the cross-covariance between any given two (different) columns in matrices  $\mathbf{H}_t, \mathbf{H}_{t+1}, \dots, \mathbf{H}_{t+d}$ . Notice that when  $k > l$ , we can still use the above formula since

$$\Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l}(q)} = \Sigma_{\mathbf{h}_{t+l}(q),\mathbf{h}_{t+k}(n)}^H.$$

It is interesting to note that, when  $R = 1$ , Eq. (45) becomes

$$\begin{aligned} \Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l}(q)} &= a_1 \Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+l-1}(q)} \\ &= \dots = a_1^{l-k} \Sigma_{\mathbf{h}_{t+k}(n),\mathbf{h}_{t+k}(q)}. \end{aligned} \quad (46)$$

Since a KF estimating the augmented channel vector defined in Eq. (7) yields, at time  $t$ , the cross-covariance matrices  $\Sigma_{\mathbf{h}_k(i),\mathbf{h}_l(j)}$  with  $k, l = t - R + 1, \dots, t$  and  $i, j = 1, \dots, N_t m$ , Eq. (45) allows the recursive computation of all the cross-covariance matrices needed to obtain, according to Eq. (42), the cross-covariance matrix  $\Sigma_{\bar{\mathbf{h}}_{t,d}(i),\bar{\mathbf{h}}_{t,d}(j)}$  for any pair of columns in  $\bar{\mathbf{H}}_{t,d}$ . To summarize,

- Equations (45), (42), and (41) together yield the channel cross-correlation matrices  $\mathbb{E}_{t-1} \left[ \bar{\mathbf{h}}_{t,d}(i) \bar{\mathbf{h}}_{t,d}(j)^H \right]$  in closed form;
- Equations (41), (34), (37), (36), (38), and (33) yield the observation autocorrelation matrix  $\mathbb{E}_{t-1} \left[ \mathbf{y}_{t+d} \mathbf{y}_{t+d}^H \right]$ .

### 4.3 The SOS-MMSE smoother

Having computed the stacked observations autocorrelation matrix given by Eq. (33), it is straightforward to plug it into Eq. (31), along with right-hand side of (32), to obtain the final expression for the proposed MMSE smoother that exploits the channel SOS explicitly,

$$\begin{aligned} \mathbf{F}_t^H &= \sigma_s^2 \mathbb{E}_{t-1} \left[ \mathbf{H}_{t,d}^H \right] \left( \sigma_s^2 \sum_{i=1}^{N_t(d+1)} \mathbb{E}_{t-1} \left[ \mathbf{h}_{t,d}(i) \mathbf{h}_{t,d}^H(i) \right] + \right. \\ &\quad + \sum_{i=1}^{N_t(m-1)} \sum_{j=1}^{N_t(m-1)} s_{t-1}^{t-m+1}(i) s_{t-1}^{t-m+1}(j) \mathbb{E}_{t-1} \left[ \mathbf{h}_{t,d}^\ddagger(i) \mathbf{h}_{t,d}^{\ddagger H}(j) \right] \\ &\quad \left. + \sigma_g^2 \mathbf{I}_{N_r(d+1)} \right)^{-1}. \end{aligned} \quad (47)$$

Notice that, at the sight of the right-hand side of Eq. (47), the proposed MMSE detector takes advantage of the previously detected symbols. Hence, causal-interference cancelation is still being performed, although in an implicit manner (as opposed to the explicit causal-interference cancelation carried out by the conventional MMSE detector).

In order to ease the implementation of the proposed scheme, Pseudocode 1 gives an overview of the necessary steps to obtain the MMSE smoother at time  $t$  when the order of the AR process used to model the channel dynamics is  $R = 1$ . The extension to higher order AR processes is straightforward (the procedure is essentially the same), though some care is needed to build up the involved cross-covariance matrices in an adequate order.

Assuming the smoothing lag,  $d$ , is approximately equal to the memory of the channel<sup>3</sup>,  $m$ , the complexity of the proposed receiver when  $R = 1$  is  $\mathcal{O}(N_r^3 N_t^3 m^4)$ . This is

**Pseudocode 1** Proposed SOS-MMSE smoother at time  $t$  when  $R = 1$ 


---

```

1: for  $k = 1, \dots, d$  do
2:   use KF equations over the dynamic system described by (10) and (8) with  $\mathbf{S}_t = \mathbf{0}_{N_r \times N_t m}$  and  $\mathbf{y}_t = \mathbf{0}_{N_r \times 1}$  to get the predictive mean,  $\hat{\mathbf{H}}_{t+k} := \mathbb{E}_{t-1} [\mathbf{H}_{t+k}]$ , and covariance matrix,  $\Sigma_{\hat{\mathbf{h}}_{t+k}}$ , of  $\mathbf{H}_{t+k}$  given observations,  $\mathbf{y}_0, \dots, \mathbf{y}_{t-1}$ , and the previously detected symbols,  $\tilde{\mathbf{s}}_0, \dots, \tilde{\mathbf{s}}_{t-1}$ .
3: use  $\hat{\mathbf{H}}_t, \hat{\mathbf{H}}_{t+1}, \dots, \hat{\mathbf{H}}_{t+d}$  to build an estimate,  $\hat{\bar{\mathbf{H}}}_{t,d}$ , of  $\bar{\mathbf{H}}_{t,d}$  using (13)
   // computation of  $\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d} \mathbf{s}_{t+d}^T & \mathbf{s}_{t+d}^H & \mathbf{H}_{t,d}^H \end{bmatrix}$ 
4: set  $\mathbf{S} \leftarrow \mathbf{0}$ 
5: for  $i = 1, \dots, N_t(d+1)$  do
6:   from Eq. (37) obtain the index,  $i'$ , of the column within matrix  $\bar{\mathbf{H}}_{t,d}$  such that  $\mathbf{h}_{t,d}(i) = \bar{\mathbf{h}}_{t,d}(i')$ 
7:   initialize the covariance matrix for the  $i'$ -th column of matrix  $\bar{\mathbf{H}}_{t,d}$ :  $\Sigma_{\hat{\mathbf{h}}_{t,d}(i'), \hat{\mathbf{h}}_{t,d}(i')} \leftarrow \mathbf{0}_{N_r(d+1)}$ 
8:   for  $k = 0, \dots, d$  do
9:     from Eq. (40) obtain the index  $n$  such that  $\check{\mathbf{h}}_{t+k}(i') = \mathbf{h}_{t+k}(n)$ 
10:    for  $l = k+1, \dots, d$  do
11:      from Eq. (40) obtain the index  $q$  such that  $\check{\mathbf{h}}_{t+l}(i') = \mathbf{h}_{t+l}(q)$ 
12:      if  $\check{\mathbf{h}}_{t+k}(i')$  or  $\check{\mathbf{h}}_{t+l}(i')$  is  $\mathbf{0}_{N_r \times 1}$  then
13:        continue to the next iteration of loop starting in line 10
14:      obtain  $\Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+k}(q)}$ , the cross-covariance matrix between columns  $n$  and  $q$  of  $\hat{\mathbf{H}}_{t+k}$  by selecting the appropriate coefficients from  $\Sigma_{\hat{\mathbf{h}}_{t+k}}$ 
15:      compute, according to Eq. (46),  $\Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+l}(q)} = a_1^{l-k} \Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+k}(q)}$ 
16:      set  $\Sigma_{\check{\mathbf{h}}_{t+k}(i'), \check{\mathbf{h}}_{t+l}(i')} \leftarrow \Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+l}(q)}$ 
17:      insert  $\Sigma_{\check{\mathbf{h}}_{t+k}(i'), \check{\mathbf{h}}_{t+l}(i')}$  (and its transpose if  $k \neq l$ ) into the adequate position of matrix  $\Sigma_{\hat{\mathbf{h}}_{t,d}(i'), \hat{\mathbf{h}}_{t,d}(i')}$  according to (42)
18:   obtain  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i')]$  as the  $i'$ -th column of  $\hat{\bar{\mathbf{H}}}_{t,d}$ 
19:   from  $\Sigma_{\hat{\mathbf{h}}_{t,d}(i'), \hat{\mathbf{h}}_{t,d}(i')}$  and  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i')]$  compute  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i') \bar{\mathbf{h}}_{t,d}^H(i')]$  according to (41)
20:   set  $\mathbf{S} \leftarrow \mathbf{S} + \mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i') \bar{\mathbf{h}}_{t,d}^H(i')]$ 
   // computation of  $\mathbb{E}_{t-1} \begin{bmatrix} \mathbf{H}_{t,d}^\dagger \mathbf{s}_{t-m+1}^T & \mathbf{s}_{t-m+1}^H & \mathbf{H}_{t,d}^{\dagger H} \end{bmatrix}$ 
21: set  $\mathbf{S}^\ddagger \leftarrow \mathbf{0}$ 
22: for  $i = 1, \dots, N_t(m-1)$  do
23:   for  $j = 1, \dots, N_t(m-1)$  do
24:     initialize the cross-covariance matrix between the  $i$ -th and  $j$ -th columns of matrix  $\bar{\mathbf{H}}_{t,d}$ :  $\Sigma_{\hat{\mathbf{h}}_{t,d}(i), \hat{\mathbf{h}}_{t,d}(j)} \leftarrow \mathbf{0}_{N_r(d+1)}$ 
25:     for  $k = 0, \dots, d$  do
26:       from Eq. (40) obtain the index  $n$  such that  $\check{\mathbf{h}}_{t+k}(i) = \mathbf{h}_{t+k}(n)$ 
27:       for  $l = 0, \dots, d$  do
28:         from Eq. (40) obtain the index  $q$  such that  $\check{\mathbf{h}}_{t+l}(j) = \mathbf{h}_{t+l}(q)$ 
29:         if  $\check{\mathbf{h}}_{t+k}(i)$  or  $\check{\mathbf{h}}_{t+l}(j)$  is  $\mathbf{0}_{N_r \times 1}$  then
30:           continue to the next iteration of loop starting in line 27
31:         if  $k \leq l$  then
32:           obtain  $\Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+k}(q)}$ , the cross-covariance matrix between columns  $n$  and  $q$  of  $\hat{\mathbf{H}}_{t+k}$  by selecting the appropriate coefficients from  $\Sigma_{\hat{\mathbf{h}}_{t+k}}$ 
33:         else
34:           obtain  $\Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+k}(q)}$  as the transpose of  $\Sigma_{\hat{\mathbf{h}}_{t+l}(q), \hat{\mathbf{h}}_{t+l}(n)}$ , the cross-covariance matrix between columns  $q$  and  $n$  of  $\hat{\mathbf{H}}_{t+l}$  by selecting the appropriate coefficients from  $\Sigma_{\hat{\mathbf{h}}_{t+l}}$ 
35:         compute, according to Eq. (46),  $\Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+l}(q)} = a_1^{|l-k|} \Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+k}(q)}$ 
36:         set  $\Sigma_{\check{\mathbf{h}}_{t+k}(i), \check{\mathbf{h}}_{t+l}(j)} \leftarrow \Sigma_{\hat{\mathbf{h}}_{t+k}(n), \hat{\mathbf{h}}_{t+l}(q)}$ 
37:         insert  $\Sigma_{\check{\mathbf{h}}_{t+k}(i), \check{\mathbf{h}}_{t+l}(j)}$  into the adequate position of matrix  $\Sigma_{\hat{\mathbf{h}}_{t,d}(i), \hat{\mathbf{h}}_{t,d}(j)}$ 
38:     obtain  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i)]$  and  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(j)]$  as, respectively, the  $i$ -th and  $j$ -th columns of  $\hat{\bar{\mathbf{H}}}_{t,d}$ 
39:     from  $\Sigma_{\hat{\mathbf{h}}_{t,d}(i), \hat{\mathbf{h}}_{t,d}(j)}$ ,  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i)]$  and  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(j)]$  compute  $\mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i) \bar{\mathbf{h}}_{t,d}^H(j)]$  according to (41)
40:     set  $\mathbf{S}^\ddagger \leftarrow \mathbf{S}^\ddagger + s_{t-m+1}(i) s_{t-m+1}^*(j) \mathbb{E}_{t-1} [\bar{\mathbf{h}}_{t,d}(i) \bar{\mathbf{h}}_{t,d}^H(j)]$ 
41: return  $\mathbf{F}_t^H = \sigma_s^2 \hat{\mathbf{H}}_{t,d} (\sigma_s^2 \mathbf{S} + \mathbf{S}^\ddagger + \sigma_g^2 \mathbf{I}_{N_r(d+1)})^{-1}$ 

```

---



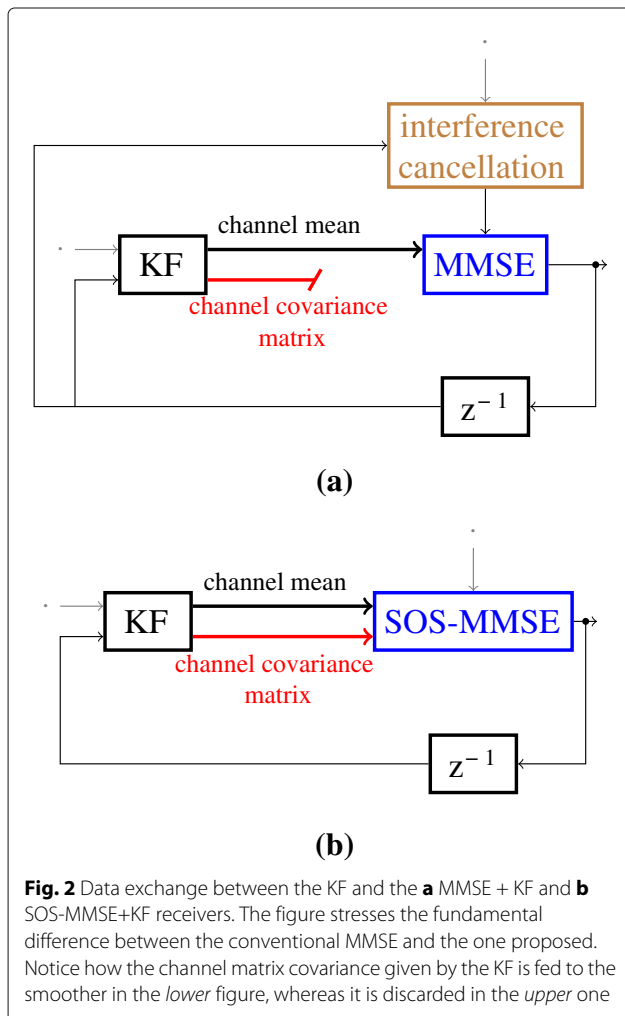
ultimately the complexity of running  $m$  KFs, each one estimating a state vector of length  $N_r N_t m$  (see lines 1–2 of Pseudocode 1). Therefore, the computational complexities of the SOS-MMSE smoother and the conventional MMSE smoother of Section 3 are of the same order.

### 5 DFE schemes

The two MMSE smoothers described in Sections 3 and 4 can be readily used in a DFE scheme that relies on a Kalman filter for the channel tracking. In particular, we aim at comparing the performance of

- a conventional MMSE DFE that neglects the SOS generated by the KF, termed “MMSE + KF” in the sequel, and
- the proposed SOS-based MMSE DFE, termed “SOS-MMSE + KF”.

Figure 2 illustrates schematically the fundamental differences between the MMSE + KF- and SOS-MMSE + KF-based receivers. The KF yields the mean and the



covariance matrix of the channel impulse response, and both receivers make use of the former (along with the observations) to obtain estimates of the symbols transmitted. However, the proposed SOS-MMSE filter also takes advantage of the covariance matrix whereas the conventional MMSE neglects the information contained within this statistic. Also notice that the proposed receiver does not perform explicit interference cancelation (since this cannot be carried out exactly), as opposed to the conventional MMSE.

### 6 Simulation results

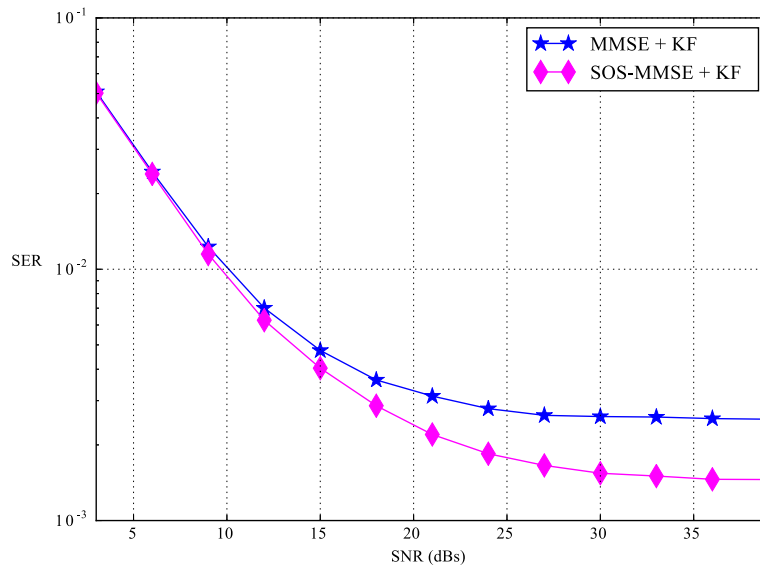
In order to assess the performance of the proposed algorithm, we have carried out computer simulations considering a system with  $N_t = 4$  transmitting antennas and  $N_r = 7$  receiving antennas. The modulation format is BPSK and transmission is carried out in frames of  $K = 300$  symbol vectors (i.e., 1200 binary symbols overall), including a training sequence of length  $T = 30$  comprising symbols known to the receiver. This last parameter has been selected empirically, after observing that an increase thereof does not yield any noticeably performance improvement while decreasing it has indeed a negative impact. The training sequence is used at the beginning of each data frame to obtain a rough estimate of the channel impulse response. However, extending the method to use pilot symbols instead of, or in addition to, a training preamble is straightforward.

A flat power profile is assumed for the channel, and every coefficient is initially (and independently) drawn from a Gaussian distribution with zero mean and unit variance. As for the channel model, an AR process of order 1 has been considered, i.e.,  $R = 1$ . The coefficient of the AR process is  $a_1 = 1 - 10^{-5}$ , and we evaluate the performance of the MMSE + KF and SOS-MMSE + KF receivers in terms of the symbol error rate (SER) considering two different values for the variance of the channel noise,  $\sigma_v^2$ , each one studied in a section of its own<sup>4</sup>. Furthermore, different values for the channel order are explored in every case.

In all the simulations, each data frame is generated independently of all others (including the transmitted data, the MIMO channel realization and the noise terms), and the lag for the MMSE smoothers is set to  $d = m - 1$ . The latter condition guarantees that every symbol is detected using all the related observations, and values of the smoothing lag above  $m - 1$  do not seem to yield a noticeable performance gain. The results are averaged over 60, 000 data frames.

#### 6.1 Slow fading channel ( $\sigma_v^2 = 5 \times 10^{-3}$ )

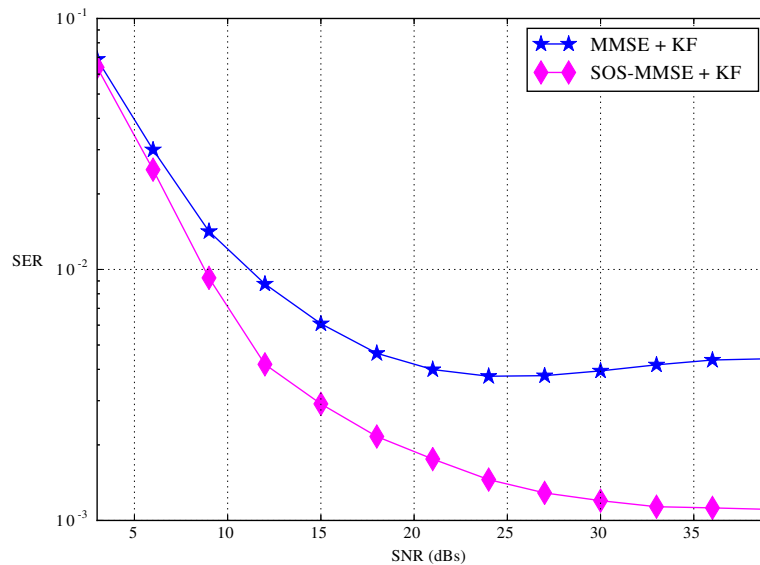
In this scenario, we consider a value of  $\sigma_v^2 = 0.005$  for the variance of the noise in the AR process governing the evolution of the channel (see Eq. (5)).



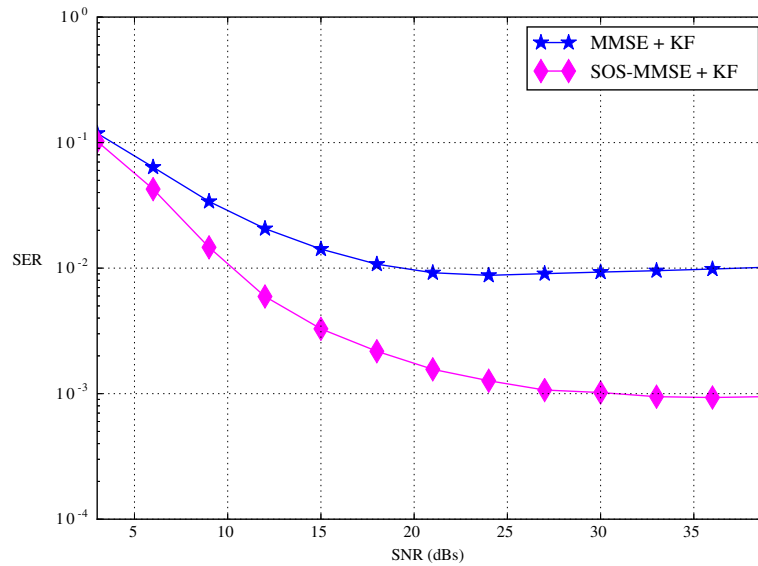
**Fig. 3** SER for several values of the SNR (dB) with  $\sigma_v^2 = 0.005$  and  $m = 1$

Figure 3 compares the SER achieved by the algorithms MMSE + KF and SOS-MMSE + KF for different values of the SNR when the channel is flat ( $m = 1$ ). In order to reach a SER of  $10^{-2}$ , the method using the proposed MMSE DFE requires roughly 0.4 dBs less SNR than the method using the conventional MMSE DFE. This gap largely widens as the SNR increases: for a SER of  $5 \times 10^{-3}$ , the curve for the MMSE + KF is more than 1 dB away from that of the SOS-MMSE + KF. Notice that both methods exhibit an error floor, but the one associated with the DFE scheme introduced in this paper is lower.

When the channel is flat, only the *present* state of the channel is needed (along with the observations) in order to detect the symbols transmitted. On the other hand, when the channel is dispersive, i.e.,  $m > 1$ , not only does the number of channel coefficients to be estimated increase, but predictions of *future* channel states are also necessary in order to perform both causal-interference cancelation and smoothing. Overall, this results in less reliable channel estimates being employed, and hence accounting for their uncertainty becomes important. This is illustrated in Fig. 4 that shows the performance of the



**Fig. 4** SER for several values of the SNR (dB) with  $\sigma_v^2 = 0.005$  and  $m = 3$



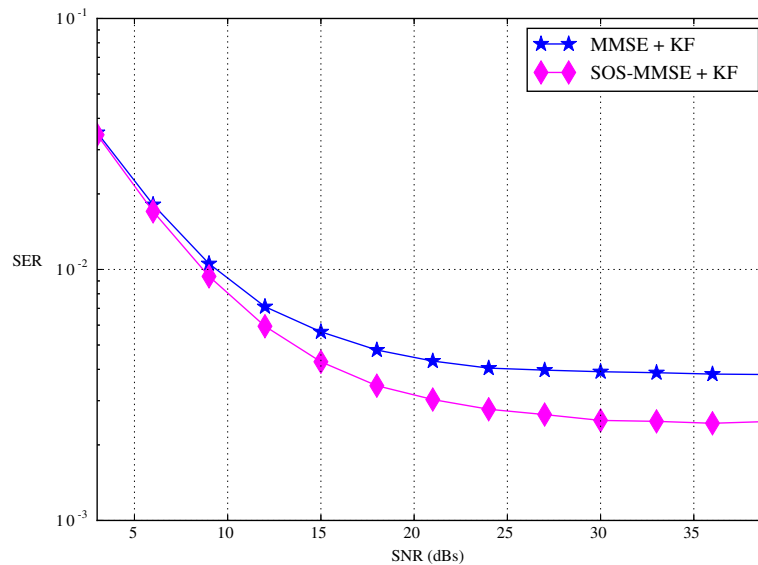
**Fig. 5** SER for several values of the SNR (dB) with  $\sigma_v^2 = 0.005$  and  $m = 5$

algorithms when  $m = 3$ . When comparing these results to those obtained for  $m = 1$ , the SER achieved by the MMSE + KF degrades over the whole range of SNRs on account of this algorithm neglecting the channel SOS. On the other hand, the SOS-MMSE + KF is able to successfully cope with the uncertainty in the channel estimates and even sees a performance boost due to the increase in diversity given by a higher channel order. As a consequence, the proposed receiver now exhibits a more clear advantage over the conventional one. Indeed, in order to attain a SER of  $10^{-2}$ , the latter requires around 2.4 dBs more SNR than the former. Figure 5 shows that when

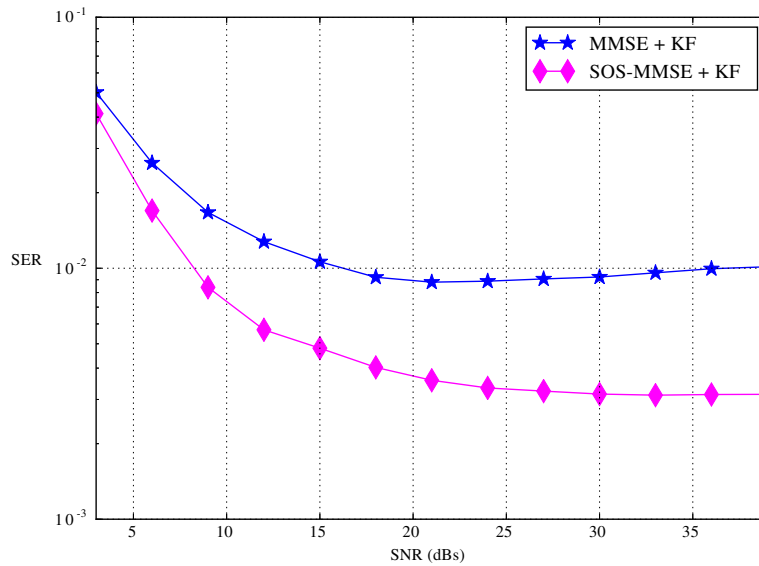
the channel order is  $m = 5$ , the gap between the curves of the SOS-MMSE + KF and MMSE + KF algorithms widens even more. Moreover, the performance of the SOS-based MMSE DFE improves slightly whereas that of the conventional MMSE DFE further deteriorates.

**6.2 Fast fading channel ( $\sigma_v^2 = 0.01$ )**

Increasing the variance of the channel noise has a twofold effect. On one side, since the channel now changes more rapidly and hence it is harder to track, the performance of any receiver is expected to worsen. On the other side, the channel evolving faster means that predictions about



**Fig. 6** SER for several values of the SNR (dB) with  $\sigma_v^2 = 0.01$  and  $m = 1$



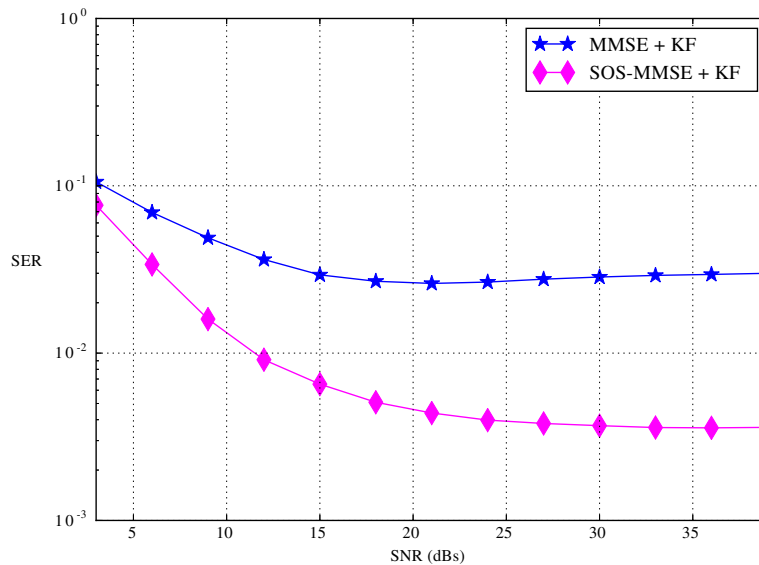
**Fig. 7** SER for several values of the SNR (dB) with  $\sigma_v^2 = 0.01$  and  $m = 3$

its future state are less reliable, and hence accounting for their uncertainty is even more important (which should benefit the SOS-based MMSE DFE). From a mathematical point of view, if the predicted channel estimates are not accurate, then the elements in the covariance matrices that enter Eq. (41) are non-negligible and so is their contribution to the computation of the proposed MMSE DFE in Eq. (47).

Figure 6 shows the performance of the algorithms in a flat channel ( $m = 1$ ). The SER of both DFEs degrades in the medium-high SNR region, as compared to the previous scenario, but this penalty is larger in the case of

the MMSE + KF. Thus, the proposed MMSE DFE now exhibits a more pronounced advantage over the conventional MMSE (about 0.65 dBs for a SER of  $10^{-2}$  and more than 3 dBs for a SNR of  $5 \times 10^{-3}$ ).

Increasing the channel order in a fast fading channel has a negligible effect on the performance of the SOS-MMSE + KF<sup>5</sup> but seriously harms that of the MMSE + KF. In Fig. 7, it can be seen that, when the channel order is  $m = 3$ , a SER of  $10^{-2}$  requires  $\approx 8.3$  dBs less SNR in the former than in the latter (recall that in the previous scenario this gap was approximately 2.4 dBs).



**Fig. 8** SER for several values of the SNR (dB) with  $\sigma_v^2 = 0.01$  and  $m = 5$

Results for the case in which  $m = 5$  are shown in Fig. 8. Again, the advantage of the proposed method over the conventional one is much more clear as the channel order increases.

### 7 Conclusions

In this work, we have introduced an enhanced version of the conventional MMSE equalizer for time-selective MIMO channels that takes advantage of the posterior second-order statistics of the channel provided by the KF. Computer simulations show that the proposed SOS-MMSE DFE yields significant performance gains (in terms of SER) over the conventional MMSE in the medium-high SNR region. In highly dispersive channels, the SNR required for the proposed SOS-MMSE receiver to achieve a certain SER can be several dBs lower than that required by the conventional MMSE. This is especially true for fast-varying channels, in which the uncertainty about the channel estimates becomes important. Indeed, a measure of this uncertainty is given by the second-order statistics of the channel, which are dismissed by the conventional MMSE, but handled by the one introduced in this paper.

It is important to point out that the application of the SOS-aided MMSE filter introduced in this work is not restricted to DFE receivers. On the contrary, the key idea is very general and can be integrated into any MMSE-based scheme as long as SOS of an unknown random variable that is relevant for the filter (here, the channel) are available.

One last remark is that the computational complexity of the proposed SOS-MMSE DFE is of the same order as that of the conventional scheme that neglects the channel SOS.

### Endnotes

<sup>1</sup> For all practical purposes, any model that is linear and affected by Gaussian noise is amenable to be used here.

<sup>2</sup> Notice that (for the sake of mathematical convenience) estimates of the symbol vectors  $\mathbf{s}_{t+1}, \mathbf{s}_{t+2}, \dots, \mathbf{s}_{t+d}$  are obtained at time  $t$  (using observations up to time  $t + d$ ) since they are also included in  $\mathbf{s}_{t+d}$ . However, they are dismissed at that time and the actual estimates of the symbols in  $\mathbf{s}_{t+i}$  are computed at time  $t + i$  using the observations up to time  $t + i + d$ .

<sup>3</sup> This is a reasonable hypothesis since the smoothing lag should be selected to account for, at least, the observations containing all the energy of the symbols transmitted at time  $t$ . This obviously depends on the length of the CIR,  $m$ , and it is common to set the smoothing lag to  $d = m - 1$ . This is also the case in the experiments whose results are presented in Section 6.

<sup>4</sup> Notice that the higher the variance of the channel noise, the more rapidly the channel coefficients fluctuate. A faster varying channel can also be obtained by decreasing the coefficient of the AR process,  $a_1$ , which determines the correlation between a channel coefficient and itself at a different time instant.

<sup>5</sup> Here, the increase in diversity due to a higher channel order is not enough to compensate for the rapid variation of the channel coefficients.

### Appendix

#### Computation of the cross-covariance $\Sigma_{\mathbf{h}_{t+k}(n), \mathbf{h}_{t+l}(q)}$ with $k < l$

The  $q$ th column in matrix  $\mathbf{H}_{t+l}$ ,  $\mathbf{h}_{t+l}(q)$ , can be expressed, according to Eq. (5), as

$$\mathbf{h}_{t+l}(q) = \sum_{r=1}^R a_r \mathbf{h}_{t+l-r}(q) + \mathbf{v}_{t+l}, \tag{48}$$

where  $\mathbf{v}_{t+l}$  is a  $N_r \times 1$  vector of i.i.d. Gaussian random variables (r.v.) with zero mean and variance  $\sigma_v^2$ . Its expectation is then given by

$$\mathbb{E}_{t-1} [\mathbf{h}_{t+l}(q)] = \sum_{r=1}^R a_r \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}(q)]. \tag{49}$$

Notice that this last equation is just Eq. (29) restated column-wise.

Substituting Eqs. (48) and (49) into (43), we obtain

$$\begin{aligned} \Sigma_{\mathbf{h}_{t+k}(n), \mathbf{h}_{t+l}(q)} = & \mathbb{E}_{t-1} \left[ \mathbf{h}_{t+k}(n) \sum_{r=1}^R a_r (\mathbf{h}_{t+l-r}^H(q) \right. \\ & \left. - \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)]) \right] + \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n) \mathbf{v}_{t+l}^H] \\ & - \mathbb{E}_{t-1} \left[ \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \sum_{r=1}^R a_r (\mathbf{h}_{t+l-r}^H(q) \right. \\ & \left. - \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)]) \right] - \\ & - \mathbb{E}_{t-1} \left[ \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \mathbf{v}_{t+l}^H \right] \end{aligned} \tag{50}$$

Since we are assuming  $k < l$ , it is clear from Eq. (5) that vectors  $\mathbf{h}_{t+k}(n)$  and  $\mathbf{v}_{t+l}^H$  are independent (the channel at time  $t+k < t+l$  is independent of the noise affecting the channel in the future), and hence  $\mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n) \mathbf{v}_{t+l}^H] = \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \mathbb{E}_{t-1} [\mathbf{v}_{t+l}^H] = 0$  due the channel noise,  $\mathbf{v}_{t+l}$ , having zero mean. For the same reason, the other expectation involving the channel noise, i.e.,  $\mathbb{E}_{t-1} [\mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \mathbf{v}_{t+l}^H]$ , is also zero

since  $\mathbb{E}_{t-1}[\mathbf{h}_{t+k}(n)]$  is a constant. Therefore, Eq. (50) becomes

$$\begin{aligned} \Sigma_{\mathbf{h}_{t+k}(n), \mathbf{h}_{t+l}(q)} &= \mathbb{E}_{t-1} \left[ \mathbf{h}_{t+k}(n) \sum_{r=1}^R a_r (\mathbf{h}_{t+l-r}^H(q)) \right. \\ &\quad \left. - \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)] \right] - \\ &\quad - \mathbb{E}_{t-1} \left[ \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \sum_{r=1}^R a_r (\mathbf{h}_{t+l-r}^H(q)) \right. \\ &\quad \left. - \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)] \right] \end{aligned}$$

(and through straightforward algebraic manipulation)

$$\begin{aligned} &= \sum_{r=1}^R a_r (\mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n) \mathbf{h}_{t+l-r}^H(q)] - \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)]) \\ &\quad - \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \sum_{r=1}^R a_r (\mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)] - \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)]) \end{aligned} \tag{51}$$

Since all the terms in the second summation of (51) are zero, only the first summation is left and the equation can be rewritten as

$$\begin{aligned} \Sigma_{\mathbf{h}_{t+k}(n), \mathbf{h}_{t+l}(q)} &= \sum_{r=1}^R a_r (\mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n) \mathbf{h}_{t+l-r}^H(q)] \\ &\quad - \mathbb{E}_{t-1} [\mathbf{h}_{t+k}(n)] \mathbb{E}_{t-1} [\mathbf{h}_{t+l-r}^H(q)]) \\ &= \sum_{r=1}^R a_r \Sigma_{\mathbf{h}_{t+k}(n), \mathbf{h}_{t+l-r}(q)}, \end{aligned} \tag{52}$$

where Eq. (44) has been used in the last step of the derivation.

**Abbreviations**

AWGN: Additive white Gaussian noise; AR: Autoregressive; BPSK: Binary phase-shift keying; CIR: Channel impulse response; DFE: Decision feedback equalizer; i.i.d.: Independent and identically distributed; KF: Kalman filter; MIMO: Multiple input multiple output; MMSE: Minimum mean square error; r.v.: Random variable; SOS: Second-order statistics; SER: Symbol error rate

**Acknowledgements**

This work was supported by *Ministerio de Economía y Competitividad* of Spain (project COMPREHENSION TEC2012-38883-C02-01), *Comunidad de Madrid* (project CASH-CAM-CM S2013/ICE-2845), and the Office of Naval Research Global (award no. N62909-15-1-2011).

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Avenida de la Universidad 30, Leganés, Spain. <sup>2</sup>School of Mathematical Sciences, Queen Mary University of London, Mile End Road, E1 4NS, London, UK. <sup>3</sup>Gregorio Maraón Health Research Institute, Calle del Dr. Esquerdo, 46, 28007 Madrid, Spain.

Received: 26 April 2016 Accepted: 8 November 2016  
Published online: 29 November 2016

**References**

1. E Telatar, Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun.* **10**, 585–595 (1999)
2. S Verdú, *Multuser Detection*. (Cambridge University Press, Cambridge, 1998)
3. TS Rappaport, *Wireless Communications: Principles and Practice*, (2nd Edition). (Prentice-Hall, Upper Saddle River, 2001)
4. C Tidestav, A Ahlen, M Sternad, Realizable MIMO decision feedback equalizers: structure and design. *IEEE Trans. Signal Process.* **49**(1), 121–133 (2001)
5. CA Belfiore, JH Park Jr., Decision feedback equalization. *Proc. IEEE.* **67**(8), 1143–1156 (1979)
6. AH Sayed, *Adaptive Filters*. (Wiley-IEEE Press, Hoboken, 2008)
7. AF Molisch, *Wireless Communications*, vol. 15. (John Wiley & Sons, West Sussex, 2010)
8. JG Andrews, Interference cancellation for cellular systems: a contemporary overview. *Wireless Commun. IEEE.* **12**(2), 19–29 (2005)
9. AM Chan, GW Wornell, A new class of efficient block-iterative interference cancellation techniques for digital communication receivers. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **30**(1-3), 197–215 (2002)
10. E Biglieri, AR Calderbank, AG Constantinides, A Goldsmith, A Paulraj, *MIMO Wireless Communications*. (Cambridge University Press, Cambridge, 2010), p. 1323
11. C Komnikakis, C Fragouli, AH Sayed, RD Wesel, Multi-input multi-output fading channel tracking and equalization using Kalman estimation. *IEEE Trans. Signal Process.* **50**(5), 1065–1076 (2002)
12. J Choi, Equalization and semi-blind channel estimation for space-time block coded signals over a frequency-selective fading channel. *IEEE Trans. Signal Process.* **52**(3), 774–785 (2004)
13. MA Vázquez, MF Bugallo, J Míguez, Sequential Monte Carlo methods for complexity-constrained MAP equalization of dispersive MIMO channels. *Signal Process.* **88**, 1017–1034 (2008)
14. G Yanfei, H Zishu, in *Communications, Circuits and Systems, 2005. Proceedings. 2005 International Conference On*. MIMO channel tracking based on Kalman filter and MMSE-DFE, vol. 1 (IEEE, 2005), pp. 223–226
15. RE Kalman, A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960)
16. BDO Anderson, JB Moore, *Optimal filtering*. (Englewood Cliffs, New Jersey, 1979)
17. N Al-Dhahir, AH Sayed, The finite-length multi-input multi-output MMSE-DFE. *Signal Process. IEEE Trans.* **48**(10), 2921–2936 (2000)
18. D Guo, X Wang, Blind detection in MIMO systems via sequential Monte Carlo. *IEEE J. Selected Areas Commun.* **21**(3), 464–473 (2003)
19. J Míguez, L Castedo, Space-time channel estimation and soft detection in time-varying multiaccess channels. *Signal Process.* **83**(2), 389–411 (2003)

Submit your manuscript to a SpringerOpen journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)