Statistical analysis of some second-order methods for blind channel identification/equalization with respect to channel undermodeling

Jean-Pierre Delmas, Houcem Gazzah, Athanasios P. Liavas and Phillip A. Regalia
Institut National des Télécommunications,
9 rue Charles Fourier, 91011 Evry Cedex, France.
tel: (33)-1-60 76 46 32.
fax: (33)-1-60 76 44 33.
Email: delmas@int-evry.fr

Abstract

Many second-order approaches have been proposed recently for blind FIR channel identification in a single-input/multi-output context. In practical conditions, the measured impulse responses usually possess "small" leading and trailing terms, the second-order statistics are estimated from finite sample size and there is additive white noise. This paper, based on a functional methodology, develops a statistical performance analysis of any second-order approach under these practical conditions. We study two channel models. In the first model, the channel tails are considered as deterministic. We derive expressions for the asymptotic bias and covariance matrix (when the sample size tends to ∞) of the m-th order estimated significant part of the impulse response. In the second model, the tails are treated as zero mean Gaussian random variables. Expressions for the asymptotic covariance matrix of the estimated significant part of the impulse response are then derived when the sample size tends to ∞ and the variance of the tails tends to 0. Furthermore, some asymptotic statistics are given for the estimated zero-forcing equalizer, the combined channel-equalizer impulse response and some byproducts, such as the open eye measure. This allows one to assess the influence of the limited sample size and the size of the tails, respectively, on the performance of identification and equalization of the algorithms under study. Closed form expressions of these statistics are given for the least-squares, the subspace, the linear prediction and the outer-product decomposition methods, as examples. Finally, the accuracy of the asymptotic analysis is checked by numerical simulations; the results are found to be valid in a very large domain of the sample size and the size of the tails.

Final version of SP-10608

1 Introduction

The recent development of second-order statistics (SOS)-based blind identification/equalization methods in a single-input/multi-output channel setting, derived either from fractional sampling in the receiver or from the use of an array of sensors, has been considered a major breakthrough and has spawned intensive research in the area. When the order of the channel is known and the second-order statistics are exact, the SOS-based blind identification methods are able to identify the channel under the so-called length and zero conditions. These same conditions ensure the existence of a finite length equalizer achieving perfect channel equalization in the absence of noise. The behavior of these methods may change dramatically however, under practically inevitable "less ideal" conditions that often occur together, such as

- second-order statistics estimated from finite sample observations;
- non-negligible additive channel noise;
- long tails of "small" leading and/or trailing impulse response terms.

Physical microwave radio channel impulse responses often possess weak leading and/or trailing terms [1] [2]. This is because the global impulse response models the transmitter shaping filter, the propagation through the channel, and the receiver filter, with each contributing to leading and/or trailing impulse response terms. In this context, it often proves convenient to partition the true channel impulse response into the *significant part* and the *tails*. By significant part is meant that part usually found near the middle of the impulse response, containing all the "large terms" and possibly some "small" intermediate terms as well; the "small" leading and trailing terms compose the tails.

The robustness of SOS-based blind identification methods with respect to the presence of tails has been studied in [3], [4], but assuming exact signal statistics are available and that channel noise is negligible. In this context, each second-order method attempts to fit a finite length (m, say) impulse response to the true channel impulse response whose actual length (including the tails) is $M > m$. Worst-case bounds are derived for the channel estimation error, and reveal that the successful application of second-order methods hinges critically on matching the assumed channel length m to the *effective* impulse response length, i.e., the length of the significant part. If the assumed length m exceeds the effective length, then the second-order methods are tacitly attempting to identify parts of the tails; this give rise to an ill-conditioned identification problem [3], [4], and should thus be avoided. Similarly, choosing m smaller than the effective length imposes a lower bound on the identification error in terms of the norm of the significant terms of the true impulse response which are excluded, irrespective of the method employed [3], [4]. This underscores the relevance of efficient methods for effective length detection [5],[6].

Here we pursue robustness aspects for any second-order method with respect to finite sample size statistics and additive white channel noise by developing a functional approach. We assume, however, that the effective channel length is correctly detected using, e.g., the method of [5],[6]; otherwise, the resulting length mismatch can result in such poor mean asymptotic performance as to render subsequent variance analyses of little interest. Two channel models are considered. In the first model, these channel tails are considered as deterministic. We derive the asymptotic bias and covariance of the estimated significant part of the impulse response when the sample size tends to infinity. The results show a similar flavor to the effect of source number underestimation on MUSIC location estimates studied in [7], in which the presence of weaker sources exerts a bias on the estimated stronger sources. In the second model, since the terms of the tails are much less stable than the significant terms, they are modeled as zero mean Gaussian random variables. We derive the asymptotic covariance of the estimated significant part of the impulse response in the limit as the sample size tends to infinity and the variance of the tails

tends to zero. These asymptotes are motivated by the fact that the tails are often one or two orders of magnitude smaller than the significant terms. General closed form expressions are given for these statistics, then derived for the least-squares (LS) [8], the subspace [9], the linear prediction (LP) [10] and the outer-product decomposition (OPD) [1] methods as examples. We note that our analysis does not fit the method by Pozidis and Petropulu [11] which relies on a spectrum estimation based on periodograms of the data, but that our performance analysis encompasses the previous statistical studies [12], [13], [14], [15] if the channel impulse reponse has no tail.

The paper is organized as follows: In Section 2, for convenience of the reader and in order to fix notations, we review the channel model and the main steps of the LS, SS, LP and OPD methods with exact second-order statistics and exact order model. In Section 3, a functional statistical analysis methodology is given for the two models of the tails. For notational simplicity, the analysis is given in the real case, as it may be straightforwardly extended to the complex case. Using a functional approach, we give the asymptotic bias and covariance matrix of the estimated m-th order significant part of the channel impulse response, the zero-forcing equalizer and the combined channel-equalizer impulse response, for any second-order method. In Section 4, we assess the performance of the LS, SS, LP and OPD methods by deriving the explicit formulas of the previous asymptotic statistics, then analyzed and compared to previous results. Finally, in Section 5 we present some simulations in which the significant part of the channel impulse response has either good or poor diversity. We examine the accuracy of the expressions of the bias and the mean square error of our estimators for the LS, SS, LP and OPD methods. In addition, we investigate the sample size and the tails size domains for which our asymptotic results remain valid.

The following notations are used throughout the paper. Matrices and vectors are represented by bold upper case and bold lower case characters, respectively. Vectors are by default in column orientation, while T, H, $*$ and $(\cdot)^\sharp$ stand for transpose, transconjugate, conjugate and Moore Penrose pseudoinverse, respectively. $\mathbf{e}_{k,i}$ is the ith unit vector in $\mathcal{R}^k$. $\mathbf{J}_k$ and $\mathbf{Z}_k$ are, respectively, the k-th order antidiagonal matrix and the shift matrix with 1's above the principal diagonal. $\mathrm{E}(\cdot), \mathrm{Cov}(\cdot), \mathrm{Tr}(\cdot)$ and $\|\cdot\|_{\mathrm{Fro}}$ denote the expectation, the covariance, the trace and the Frobenius matrix norm, respectively. $\mathrm{Vec}(\cdot)$ is the "vectorization" operator that turns a matrix into a vector by stacking the columns of the matrix one below another. It is used in conjunction with the Kronecker product $\mathbf{A}\otimes\mathbf{B}$ as the block matrix whose (i,j) block element is $a_{i,j}\mathbf{B}$ with the vec-permutation matrix [16] $\mathbf{K}_{r,s}$ which transforms $\mathrm{Vec}(\mathbf{A})$ to $\mathrm{Vec}(\mathbf{A}^T)$ for any $r \times s$ matrix $\mathbf{A}$ and with the following properties ($\mathbf{B}$ is any $p \times q$ matrix in the third relation)

$$\mathrm{Vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{Vec}(\mathbf{B}) \tag{1.1}$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD} \tag{1.2}$$

$$\mathbf{K}_{r,p}(\mathbf{A} \otimes \mathbf{B})\mathbf{K}_{q,s} = \mathbf{B} \otimes \mathbf{A} \tag{1.3}$$

2 Some second-order methods: Exact order case

For convenience of the reader and in order to fix notations, we recall the basic steps of the LS, SS, LP and OPD methods, based on exact second-order statistics for the single-input/two-output channel setting presented in Fig.1. This setting is obtained by channel oversampling by a factor of 2 or by using a 2-sensor receiver. We have chosen to treat only this setting because it is both quite common in telecommunications and it leads to very simple results; in particular, for the LS and SS methods there is simple relationship between the minimal covariance matrix eigenvector and the estimated impulse response.

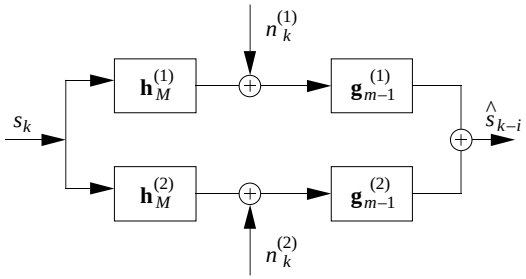


Fig.1 Noisy two-channel equalization setting.

2.1 Two-channel model

If the true channel order is M, the output of the i-th channel, $x_k^{(i)}$, for $i = 1, 2$, is given by

$$x_k^{(i)} = \sum_{l=0}^{M} h_l^{(i)} s_{k-l} + n_k^{(i)}. \tag{2.1}$$

The input sequence s_k is assumed to be i.i.d., zero mean, and of unit variance, $h_k^{(i)}$ is the impulse response of the i-th channel, $i = 1, 2$, and $n_k^{(i)}$ is additive zero mean Gaussian white channel noise with power σ_n^2. We assume that the two channels do not share common zeros, guaranteeing their identifiability. By stacking the $L+1$ most recent samples of each channel, we obtain the representation.

$$\mathbf{x}_L(k) \stackrel{\text{def}}{=} \left(\mathbf{x}_k^T, \ldots, \mathbf{x}_{k-L}^T\right)^T = \mathcal{T}_L(\mathbf{h}_M)\mathbf{s}_{L+M}(k) + \mathbf{n}_L(k)$$

with $\mathbf{x}_k \stackrel{\text{def}}{=} \left(x_k^{(1)}, x_k^{(2)}\right)^T$, $\mathbf{h}_M \stackrel{\text{def}}{=} \left(\mathbf{h}_{(0)}^T, \ldots, \mathbf{h}_{(M)}^T\right)^T$, $\mathbf{n}_k \stackrel{\text{def}}{=} \left(n_k^{(1)}, n_k^{(2)}\right)^T$, $\mathbf{s}_{L+M}(k) \stackrel{\text{def}}{=} (s_k, \ldots, s_{k-L-M})^T$ and where $\mathcal{T}_L(\mathbf{h}_M)$ is the $2(L+1) \times (L+M+1)$ Sylvester resultant matrix with $\mathbf{h}_{(k)} \stackrel{\text{def}}{=} \left(h_k^{(1)}, h_k^{(2)}\right)^T$:

$$\mathcal{T}_L(\mathbf{h}_M) = \begin{pmatrix} \mathbf{h}_{(0)} & \cdots & \cdots & \mathbf{h}_{(M)} & & \\ & \ddots & & & \ddots & \\ & & \mathbf{h}_{(0)} & \cdots & \cdots & \mathbf{h}_{(M)} \end{pmatrix}.$$

In the sequel, we recall briefly the second-order methods under study, in the exact second-order statistics case, assuming that the true channel is the m-th order channel $\mathbf{h}_m$.

2.2 LS and SS methods

The LS and SS estimates of $\mathbf{h}_m$, which coincide in the two-channel case with $L = m$ [17], defined up to a constant scale factor, are given by the relation $\begin{bmatrix} h_k^{(1)} \\ h_k^{(2)} \end{bmatrix} = \begin{bmatrix} v_k^{(2)} \\ -v_k^{(1)} \end{bmatrix}$, with $\mathbf{v}_{2(m+1)} = \left(v_0^{(1)}, v_0^{(2)}, \ldots, v_m^{(1)}, v_m^{(2)}\right)^T$ being the eigenvector associated with the unique smallest eigenvalue of $\mathbf{R}_m \stackrel{\text{def}}{=} \mathrm{E}\left(\mathbf{x}_m(k)\mathbf{x}_m^T(k)\right)$, i.e.,

$$\mathbf{h}_m = \mathbf{T}_m \mathbf{v}_{2(m+1)} \tag{2.2}$$

where $\mathbf{T}_m$ is the antisymmetric orthogonal matrix $\mathbf{I}_{m+1} \otimes \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

2.3 LP method

The basic steps of the LP method in the two-channel case are sketched in the sequel. First, the coefficients $[\mathbf{A}_1, \ldots, \mathbf{A}_m]$ of a $2 \times 2m$ predictor filter are given by

$$[\mathbf{A}_1, \ldots, \mathbf{A}_m] = -[\mathbf{r}_1, \ldots, \mathbf{r}_m]\Big(\mathbf{R}'_{m-1}\Big)^{-1}$$

with

$$\mathbf{R}'_{m-1} \stackrel{\text{def}}{=} \mathrm{E}\left(\mathbf{x}_{m-1}(k)\mathbf{x}^T_{m-1}(k)\right) - \sigma_n^2\mathbf{I}_{2m} \quad \text{and} \quad \mathbf{r}_i \stackrel{\text{def}}{=} \mathrm{E}(\mathbf{x}_k\mathbf{x}^T_{k-i}), \qquad i = 1, \ldots, m.$$

Then the rank-one innovation covariance matrix $\mathbf{D} = \mathbf{h}_{(0)}\mathbf{h}^T_{(0)}$ is given by $\mathbf{D} = \mathbf{r}'_0 + \sum_{k=1}^m \mathbf{A}_k{\mathbf{r}_k}^T$, with $\mathbf{r}'_0 = \mathrm{E}(\mathbf{x}_k\mathbf{x}^T_k) - \sigma_n^2\mathbf{I}_2$. If λ and $\mathbf{v}$ are, respectively, the nonzero eigenvalue of the rank-one matrix $\mathbf{D}$, and its associated eigenvector, then an m-th order zero-forcing zero-delay equalizer[1] is given by $\mathbf{g}_m = \frac{1}{\sqrt{\lambda}}[\mathbf{I}_2, \mathbf{A}_1, \ldots, \mathbf{A}_m]^T\mathbf{v}$ and the impulse response $\mathbf{h}_m$ is identified as

$$\mathbf{h}_m = \mathbf{S}_m\mathbf{g}_m \qquad \text{with} \qquad \mathbf{S}_m \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{r}'_0 & \mathbf{r}_1 & \cdots & \mathbf{r}_m \\ \mathbf{r}_1 & \mathbf{r}_2 & \ddots & \\ \vdots & \ddots & & \\ \mathbf{r}_m & & & \end{pmatrix}_{2(m+1)\times 2(m+1)} .$$

2.4 OPD method

The outer-product decomposition method [1] is based on the rank-one outer-product matrix $\mathbf{h}_m\mathbf{h}^T_m$, which is shown to be equal to $\mathbf{D}_3 \stackrel{\text{def}}{=} \mathbf{D}_1 - \mathbf{D}_2$, where:

$$\mathbf{D}_1 = \mathbf{S}_m\mathbf{R}'_m{}^{\natural}{\mathbf{S}_m}^T = \begin{bmatrix} \times & \times \\ \times & \mathbf{D}'_2 \end{bmatrix}, \qquad \mathbf{D}_2 = \begin{bmatrix} \mathbf{D}'_2 & \mathbf{O}_{2m,2} \\ \mathbf{O}_{2,2m} & \mathbf{O}_{2,2} \end{bmatrix} \tag{2.3}$$

m+2

The OPD estimate of $\mathbf{h}_m$, defined up to scale factor, is the eigenvector $\mathbf{v}$ associated with the unique nonzero eigenvalue of $\mathbf{D}_3$:

$$\mathbf{h}_m = \mathbf{v}. \tag{2.4}$$

3 Statistical analysis methodology

3.1 m-th order effective channel identification

We denote by *effective order* of the channel the order detected by a suitable rank detection procedure (see e.g. [5] [2]). Our principal concern in this section is deducing the asymptotic performance of any second-order algorithm that assumes that the effective order of the impulse response is detected beforehand is m. We call these methods m-th order SOS-based methods. To this end, we partition the true impulse response $\mathbf{h}_M$ into the zero-padded m-th order significant part $\mathbf{h}^z_{m,M}$ and tails $\mathbf{d}^z_{m,M}$ as follows:

$$\mathbf{h}_M = \mathbf{h}^z_{m,M} + \mathbf{d}^z_{m,M}, \tag{3.1}$$

[1] We note that this equalizer has no reason to be minimum norm.

[2] The procedure introduced in [5] gives as byproduct (through the existence of a gap between two consecutive eigenvalues of the estimated covariance matrix), an indication of the stability of the blind channel approximation problem. In the stable case, existence of a significant part of order m which gathers most of energy is ensured.

$$\mathbf{h}^z_{m,M} \stackrel{\text{def}}{=} [\underbrace{\mathbf{0}^T \ \cdots \ \mathbf{0}^T}_{m_1}, \underbrace{\mathbf{h}^T_{(m_1)} \ \cdots \ \mathbf{h}^T_{(m_1+m)}}_{m+1}, \underbrace{\mathbf{0}^T \ \cdots \ \mathbf{0}^T}_{m_2}]^T,$$

$$\mathbf{d}^z_{m,M} \stackrel{\text{def}}{=} [\underbrace{\mathbf{h}^T_{(0)} \ \cdots \ \mathbf{h}^T_{(m_1-1)}}_{m_1}, \underbrace{\mathbf{0}^T \ \cdots \ \mathbf{0}^T}_{m+1}, \underbrace{\mathbf{h}^T_{(m_1+m+1)} \ \cdots \ \mathbf{h}^T_{(M)}}_{m_2}]^T.$$

We denote the non zero-padded vectors $\mathbf{h}_m$ and $\mathbf{d}_{m,M}$ as follows:

$$\mathbf{h}_m \stackrel{\text{def}}{=} [\,\mathbf{h}^T_{(m_1)} \ \cdots \ \mathbf{h}^T_{(m_1+m)}]^T, \qquad \mathbf{d}_{m,M} \stackrel{\text{def}}{=} [\,\mathbf{h}^T_{(0)} \ \cdots \ \mathbf{h}^T_{(m_1-1)}, \mathbf{h}^T_{(m_1+m+1)} \ \cdots \ \mathbf{h}^T_{(M)}]^T.$$

Here $\mathbf{h}_m$ denotes the m-th order significant part of the channel. $M = m_1 + m + m_2$ where m_1 and m_2 denote respectively the length of the leading and trailing parts of the channel impulse response $\mathbf{h}_M$. We note that this partitioning of $\mathbf{h}_M$ remains valid if the two significant parts of each subchannel are not aligned because the size ("large" or "small") is taken as the norm $\|\mathbf{h}_{(k)}\|$. Also M, m_1 and m_2 are considered to be known for analysis purposes, but of course, they are unknown from the algorithmic point of view.

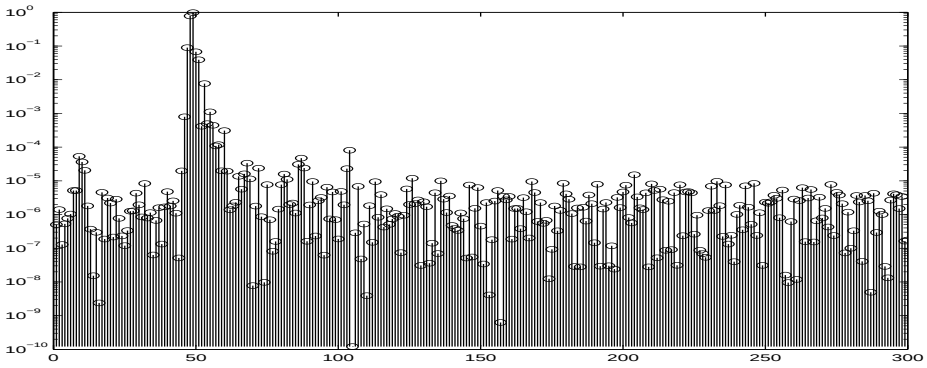


Fig.2 Square magnitude of the real part of microwave radio channel, oversampled, by a factor of 2.

In Fig.2, we plot the real part of the *chan1.mat* oversampled, by a factor of 2, complex-valued microwave radio channel, found at *http://spib.rice.edu/spib/microwave.html*. Here, the "small" terms are about two orders of magnitude smaller than the significant terms but the partitioning of the impulse response into the "large" and "small" terms is not perfectly clear.

To study the performance of such m-th order SOS-based algorithms, we introduce two $2(m+1) \times 2(m+1)$ spatio-temporal covariance matrices. The first is the estimated spatio-temporal covariance matrix of the data

$$\widehat{\mathbf{R}}^M_m \stackrel{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_m(k)\mathbf{x}^T_m(k) = \frac{1}{N}\sum_{k=1}^{N} \left[(\mathcal{T}_m(\mathbf{h}_M)\,\mathbf{s}_{m+M}(k) + \mathbf{n}_m(k)\right][\cdot]^T,$$

whose expectation yields

$$\mathbf{R}^M_m \stackrel{\text{def}}{=} \mathrm{E}(\widehat{\mathbf{R}}^M_m) = \mathcal{T}_m(\mathbf{h}_M)\mathcal{T}^T_m(\mathbf{h}_M) + \sigma^2_n \mathbf{I}_{2(m+1)}.$$

The second is the spatio-temporal covariance associated with the m-th order significant part of the channel impulse response:

$$\mathbf{R}^m_m \stackrel{\text{def}}{=} \mathcal{T}_m(\mathbf{h}_m)\mathcal{T}^T_m(\mathbf{h}_m) + \sigma^2_n \mathbf{I}_{2(m+1)}. \tag{3.2}$$

To consider the asymptotic performance of an m-th order SOS-based algorithm, we adopt a functional approach which consists of recognizing that the whole process of constructing an estimate $\widehat{\mathbf{h}}_m$ of $\mathbf{h}_m$ is equivalent to defining a functional relation linking the estimate $\widehat{\mathbf{h}}_m$ to the sample statistics $\widehat{\mathbf{R}}_m^M$ from which it is inferred. This functional dependence is denoted $\widehat{\mathbf{h}}_m = \mathrm{alg}(\widehat{\mathbf{R}}_m^M)$. Clearly, $\mathbf{h}_m = \mathrm{alg}(\mathbf{R}_m^m)$, so the different algorithms $\mathrm{alg}(\cdot)$ constitute distinct extensions of the mapping $\mathbf{R}_m^m \to \mathbf{h}_m$ generated by (3.2) to any unstructured real symmetric $\widehat{\mathbf{R}}_m^M$. We consider two models of tails.

3.1.1 Deterministic model of tails

In the deterministic model of tails, the tails are considered as deterministic and we are interested in the asymptotic bias and asymptotic covariance matrix of $\widehat{\mathbf{h}}_m$ when the sample number N tends to ∞. $\widehat{\mathbf{R}}_m^M$ may be considered as a perturbation of $\mathbf{R}_m^M$:

$$\widehat{\mathbf{R}}_m^M = \mathbf{R}_m^M + \delta\mathbf{R}_m^M \tag{3.3}$$

where $\delta\mathbf{R}_m^M$ is the finite sample size error, verifying $\mathrm{E}(\delta\mathbf{R}_m^M) = \mathbf{O}$ and $\mathrm{Cov}(\delta\mathbf{R}_m^M) = O\left(\frac{1}{N}\right)$ [18, §7.3]. Because the mapping $\mathrm{alg}(\cdot)$ is sufficiently regular in a neighborhood of $\mathbf{R}_m^M$ for most algorithms (if necessary, regularization techniques are employed), we have from (3.3),

$$\widehat{\mathbf{h}}_m = \mathrm{alg}(\mathbf{R}_m^M) + (\mathbf{D}_{\mathrm{alg}}, \delta\mathbf{R}_m^M)\Big|_{\mathbf{R}_m^M} + O(\|\delta\mathbf{R}_m^M\|^2), \tag{3.4}$$

where $(\mathbf{D}_{\mathrm{alg}}, \delta\mathbf{R}_m^M)\Big|_{\mathbf{R}_m^M}$ denotes the differential of the mapping $\mathrm{alg}(\cdot)$ evaluated at point $\mathbf{R}_m^M$ applied to $\delta\mathbf{R}_m^M$. Taking expectations, we obtain:

$$\mathrm{E}(\widehat{\mathbf{h}}_m) = \mathrm{alg}(\mathbf{R}_m^M) + O\left(\frac{1}{N}\right). \tag{3.5}$$

The matrix $\mathbf{R}_m^M$ may be considered as a perturbation of $\mathbf{R}_m^m$:

$$\mathbf{R}_m^M = \mathbf{R}_m^m + \delta\mathbf{R}_d, \tag{3.6}$$

where $\delta\mathbf{R}_d$ is due to the tails, i.e.,

$$\delta\mathbf{R}_d \stackrel{\mathrm{def}}{=} \mathcal{T}_m(\mathbf{h}_{m,M}^z)\mathcal{T}_m^T(\mathbf{d}_{m,M}^z) + \mathcal{T}_m(\mathbf{d}_{m,M}^z)\mathcal{T}_m^T(\mathbf{h}_{m,M}^z) + O(\|\mathbf{d}_{m,M}\|^2). \tag{3.7}$$

The vectorization of $\delta\mathbf{R}_d$ yields:

$$\mathrm{Vec}(\delta\mathbf{R}_d) = \mathbf{C}_1\mathbf{d}_{m,M} + O(\|\mathbf{d}_{m,M}\|^2),$$

with

$$\mathbf{C}_1 \stackrel{\mathrm{def}}{=} \left[\mathbf{K}_{2(m+1),2(m+1)} + \mathbf{I}_{4(m+1)^2}\right]\left[\mathbf{K}_{m+1,2} \otimes \mathcal{T}_m(\mathbf{h}_{m,M}^z)\right]\left[\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{1}_m \otimes \begin{pmatrix} \mathbf{I}^\dagger \\ \mathbf{O}_{m+1,M-m} \end{pmatrix} \\ \mathbf{I}^\dagger \end{pmatrix}\right]\mathbf{K}_{2,M-m}$$

where

$$\mathbf{I}^\dagger \stackrel{\mathrm{def}}{=} \begin{bmatrix} \mathbf{I}_{m_1,m_1} & \mathbf{O}_{m_1,m_2} \\ \mathbf{O}_{m+1,m_1} & \mathbf{O}_{m+1,m_2} \\ \mathbf{O}_{m_2,m_1} & \mathbf{I}_{m_2,m_2} \end{bmatrix} \tag{3.8}$$

is defined from the linear relation linking $\mathbf{d}_{m,M}^z$ and $\mathbf{d}_{m,M}$ ($\mathbf{d}_{m,M}^z \stackrel{\mathrm{def}}{=} \mathbf{I}^\ddagger\mathbf{d}_{m,M} = (\mathbf{I}^\dagger \otimes \mathbf{I}_2)\mathbf{d}_{m,M}$) and where the vec-permutation matrix $\mathbf{K}_{r,s}$ is defined in the Introduction section and the classical property (1.1)

has been used. Let $\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^m_m}$ denote the matrix associated with the differential $\mathbf{D}_{\text{alg}}$ at point $\mathbf{R}^m_m$; precise expressions for each algorithm will be given in Section 4. Using (3.6), the first-order perturbation analysis of an m-th order SOS-based algorithm acting on $\mathbf{R}^M_m$ evaluated at point $\mathbf{R}^m_m$ gives

$$\begin{aligned}\text{alg}(\mathbf{R}^M_m) &= \text{alg}(\mathbf{R}^m_m) + (\mathbf{D}_{\text{alg}}, \delta\mathbf{R}_d)|_{\mathbf{R}^m_m} + O(\|\mathbf{d}_{m,M}\|^2) \\ &= \mathbf{h}_m + \mathbf{A}^{\text{alg}}_{h,\mathbf{R}^m_m}\text{Vec}(\delta\mathbf{R}_d) + O(\|\mathbf{d}_{m,M}\|^2) \\ &= \mathbf{h}_m + \mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}\mathbf{d}_{m,M} + O(\|\mathbf{d}_{m,M}\|^2), \end{aligned} \tag{3.9}$$

with $\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m} \stackrel{\text{def}}{=} \mathbf{A}^{\text{alg}}_{h,\mathbf{R}^m_m}\mathbf{C}_1$. So, from (3.5) and (3.9), the following result holds:

Result 1 *The asymptotic bias in the deterministic model of tails is given by*

$$\text{E}(\widehat{\mathbf{h}}_m) - \mathbf{h}_m = \mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}\mathbf{d}_{m,M} + O(\|\mathbf{d}_{m,M}\|^2) + O\left(\frac{1}{N}\right), \tag{3.10}$$

and when $N \to \infty$ and $\|\mathbf{d}_{m,M}\| \to 0$.

$$\frac{\|\text{E}(\widehat{\mathbf{h}}_m) - \mathbf{h}_m\|}{\|\mathbf{d}_{m,M}\|} \leq \sigma_1(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}) \tag{3.11}$$

where $\sigma_1(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m})$ is the largest singular value of $\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}$ and equality prevails for tails $\mathbf{d}_{m,M}$ colinear with the right singular vector of $\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}$.

Then, from (3.4) and (3.5), the mapping alg$(\cdot)$ gives the deviation from the asymptotic mean $\text{E}(\widehat{\mathbf{h}}_m)$:

$$\begin{aligned}\widehat{\mathbf{h}}_m - \text{E}(\widehat{\mathbf{h}}_m) &= (\mathbf{D}_{\text{alg}}, \delta\mathbf{R}^M_m)\Big|_{\mathbf{R}^M_m} + O(\|\delta\mathbf{R}^M_m\|^2) + O\left(\frac{1}{N}\right) \\ &= \mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}\text{Vec}(\delta\mathbf{R}^M_m) + O(\|\delta\mathbf{R}^M_m\|^2) + O\left(\frac{1}{N}\right). \end{aligned} \tag{3.12}$$

It thus follows that

$$\lim_{N\to\infty} N\text{Cov}(\widehat{\mathbf{h}}_m) = \mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}\mathbf{C}_{R^M_m}\left(\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}\right)^T$$

with $\mathbf{C}_{R^M_m} = \lim_{N\to\infty} N\text{Cov}(\text{Vec}(\widehat{\mathbf{R}}^M_m)) = \lim_{N\to\infty} N\text{E}(\text{Vec}(\delta\mathbf{R}^M_m)\text{Vec}(\delta\mathbf{R}^M_m)^T)$. And since $\text{Vec}(\delta\mathbf{R}^M_m)$ is asymptotically Gaussian (see Appendix A) i.e., $\sqrt{N}(\text{Vec}(\widehat{\mathbf{R}}^M_m) - \text{Vec}(\mathbf{R}^M_m)) \stackrel{\mathcal{L}}{\to} \mathcal{N}(\mathbf{0}, \mathbf{C}_{R^M_m})$, we have, thanks to a continuity theorem [19, Th. 6.2a, p. 387] applied to the differentiable mapping alg$(\cdot)$, the following asymptotic distribution result:

Result 2 *In the deterministic model of tails, $\widehat{\mathbf{h}}_m$ is asymptotically Gaussian when $N \to \infty$:*

$$\sqrt{N}(\widehat{\mathbf{h}}_m - \text{E}(\widehat{\mathbf{h}}_m)) \stackrel{\mathcal{L}}{\to} \mathcal{N}\left(\mathbf{0}, \mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}\mathbf{C}_{R^M_m}\left(\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}\right)^T\right), \tag{3.13}$$

where

$$\lim_{N\to\infty} N\text{Cov}(\widehat{\mathbf{h}}_m) = \mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}\mathbf{C}_{R^M_m}\left(\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}\right)^T. \tag{3.14}$$

3.1.2 Statistical model of tails

In the statistical model of tails, the components of $\mathbf{d}_{m,M}$ are assumed to be independent, zero mean, Gaussian with the same variance and $\mathrm{E}\|\mathbf{d}_{m,M}\|^2 \stackrel{\text{def}}{=} \sigma_d^2$. Here $\mathbf{d}_{m,M}$ is assumed to be independent from s_k and $n_k^{(i)}, i = 1, 2$. The matrix $\widehat{\mathbf{R}}_m^M$ may be considered here as a perturbation of $\mathbf{R}_m^m$:

$$\widehat{\mathbf{R}}_m^M = \mathbf{R}_m^m + \delta\mathbf{R} \text{ with } \delta\mathbf{R} \stackrel{\text{def}}{=} \delta\mathbf{R}_m^m + \delta\mathbf{R}_d + O(\|\mathbf{d}_{m,M}\|^2) + O(\|\delta\mathbf{R}_m^m\|^2) + O(\|\delta\mathbf{R}_m^m\mathbf{d}_{m,M}\|),$$

where $\delta\mathbf{R}_m^m$ is the finite sample size error:

$$\delta\mathbf{R}_m^m \stackrel{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N}\left[(\mathcal{T}_m(\mathbf{h}_m)\,\mathbf{s}_{2m}(k) + \mathbf{n}_m(k)\right][\cdot]^T - \mathbf{R}_m^m.$$

A first-order perturbation analysis of an m-th order SOS-based algorithm acting on $\widehat{\mathbf{R}}_m^M$ evaluated at point $\mathbf{R}_m^m$ gives the estimate:

$$\widehat{\mathbf{h}}_m = \mathbf{h}_m + \delta\mathbf{h}_m,$$

with

$$\begin{aligned}\delta\mathbf{h}_m &= \mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\mathrm{Vec}(\delta\mathbf{R}) + O(\|\delta\mathbf{R}\|^2)\\ &= \mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\mathrm{Vec}(\delta\mathbf{R}_m^m) + \mathbf{B}_{h,\mathbf{R}_m^m}^{\text{alg}}\mathbf{d}_{m,M} + O(\|\delta\mathbf{R}\|^2). \qquad (3.15)\end{aligned}$$

So, when $N \to \infty$ and $\sigma_d^2 \to 0$,

$$\mathrm{E}(\widehat{\mathbf{h}}_m) = \mathbf{h}_m + \mathrm{E}(O(\|\delta\mathbf{R}\|^2)) = \mathbf{h}_m + O(\|\mathbf{d}_{m,M}\|^2) + O\left(\frac{1}{N}\right), \qquad (3.16)$$

and since $\mathbf{d}_{m,M}$ and $\delta\mathbf{R}_m^m$ are independent random variables, the following result holds:

Result 3

$$\mathrm{Cov}(\widehat{\mathbf{h}}_m) = \frac{1}{N}\mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\mathbf{C}_{R_m^m}\left(\mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\right)^T + \frac{\sigma_d^2}{2(M-m)}\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{alg}}\left(\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{alg}}\right)^T + O(\|\mathbf{d}_{m,M}\|^4) + O\left(\frac{1}{N^2}\right) + O\left(\frac{\|\mathbf{d}_{m,M}\|^2}{N}\right), \qquad (3.17)$$

$$\mathrm{E}\|\widehat{\mathbf{h}}_m - \mathbf{h}_m\|^2 \sim \frac{1}{N}\mathrm{Tr}\left(\mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\mathbf{C}_{R_m^m}\left(\mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\right)^T\right) + \frac{\sigma_d^2}{2(M-m)}\mathrm{Tr}\left(\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{alg}}\left(\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{alg}}\right)^T\right), \qquad (3.18)$$

when $N \to \infty$ and $\sigma_d^2 \to 0$, where the expression of $\mathbf{C}_{R_m^m} = \lim_{N\to\infty} N\mathrm{E}(\mathrm{Vec}(\delta\mathbf{R}_m^m)\mathrm{Vec}(\delta\mathbf{R}_m^m)^T)$ is given in Appendix A.

Two particular cases can be deduced, (1) exact channel order and finite sample size (no tails in the impulse response [3]) and (2) exact second-order statistics and tails in the impulse response, where we have respectively:

$$\sqrt{N}(\widehat{\mathbf{h}}_m - \mathbf{h}_m) \stackrel{\mathcal{L}}{\to} \mathcal{N}\left(\mathbf{0}, \mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\mathbf{C}_{R_m^m}\left(\mathbf{A}_{h,\mathbf{R}_m^m}^{\text{alg}}\right)^T\right) \quad \text{when } N \to \infty, \qquad (3.19)$$

$$\frac{1}{\sigma_d}(\widehat{\mathbf{h}}_m - \mathbf{h}_m) \stackrel{\mathcal{L}}{\to} \mathcal{N}\left(\mathbf{0}, \frac{1}{2(M-m)}\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{alg}}\left(\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{alg}}\right)^T\right) \quad \text{when } \sigma_d^2 \to 0. \qquad (3.20)$$

Thus, the influence of (1) the finite sample size and (2) the tails, can be analyzed in the same framework. The identified channel by any m-th order SOS-based algorithm is close to the m-th order significant part $\mathbf{h}_m$ of the impulse response $\mathbf{h}_M$. This closeness depends on the diversity of $\mathbf{h}_m$, as will be seen in Sections 4 and 5 and also, depending on the case considered, on the sample size N or on the size of the tails measured by σ_d^2. Upper bounds with a similar flavor have previously been obtained in [3] and [4] for the LS/SS and LP methods, with respect to the presence of tails.

[3] We note, that in the absence of tails, the deterministic tail model gives $\mathbf{R}_m^M = \mathbf{R}_m^m$. So, the estimates $\widehat{\mathbf{h}}_m$ are asymptotically unbiased with asymptotic distribution (3.13) identical to (3.19).

3.2 Zero-forcing equalization

Having "identified" the m-th order channel $\widehat{\mathbf{h}}_m$, we can equalize it perfectly in the noiseless case, by using the zero-forcing equalizers of order $m-1$, for delays $i=0,\ldots,2m-1$, given by

$$\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i} = \mathcal{T}^{-T}_{m-1}(\widehat{\mathbf{h}}_m)\mathbf{e}_{2m,i+1}. \tag{3.21}$$

In the presence of additive channel noise, the output of the equalizer $\widehat{s}_{k-i}$ is corrupted by additive noise of power $\|\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}\|^2\sigma_n^2$. Of course, $\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}$ is not a zero-forcing equalizer for the true channel $\mathbf{h}_M$. To gauge the equalization error, we introduce the combined channel-equalizer impulse response, denoted $\widehat{\mathbf{f}}_{m+M-1,i} \stackrel{\mathrm{def}}{=} (\hat{f}_{0,i},\ldots,\hat{f}_{m+M-1,i})^T$, according to:

$$\widehat{\mathbf{f}}_{m+M-1,i} = \mathcal{T}^{T}_{m-1}(\mathbf{h}_M)\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}. \tag{3.22}$$

Adopting the functional approach of Section 3.1 to $\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}$ and $\widehat{\mathbf{f}}_{m+M-1,i}$, we have[4]:

$$\widehat{\mathbf{R}}^M_m \stackrel{\mathrm{alg}}{\longmapsto} \widehat{\mathbf{h}}_m \longmapsto \widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i} \longmapsto \widehat{\mathbf{f}}_{m+M-1,i}.$$

But since

$$\mathbf{R}^m_m \stackrel{\mathrm{alg}}{\longmapsto} \mathbf{h}_m \longmapsto \mathbf{g}^{\mathrm{ZF}}_{m-1,i} \longmapsto \mathbf{f}_{m+M-1,i} = \mathcal{T}^{T}_{m-1}(\mathbf{h}_M)\mathbf{g}^{\mathrm{ZF}}_{m-1,i} \neq \mathbf{e}_{m+M,p+i+1},$$

we can extend the results of Section 3.1 to estimates $\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}$ and $\widehat{\mathbf{f}}_{m+M-1,i}$. Naturally, this approach could be applied to any L-th order Wiener equalizer:

$$\widehat{\mathbf{g}}^{\mathrm{W}}_{L,i} = \left(\widehat{\mathbf{R}}^M_L\right)^{-1}\mathcal{T}_L(\widehat{\mathbf{h}}_m)\mathbf{e}_{2(L+1),i+1}.$$

3.2.1 Zero-forcing equalizer

The mapping $\widehat{\mathbf{h}}_m \longmapsto \widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}$ given by (3.21) is differentiable at $\mathbf{h}_m$, with differential:

$$\begin{aligned}\delta\mathbf{g}_i &= -\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\delta\mathcal{T}^{T}_{m-1}(\mathbf{h}_m)\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathbf{e}_{2m,i+1} = -\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathcal{T}^{T}_{m-1}(\delta\mathbf{h}_m)\mathbf{g}^{\mathrm{ZF}}_{m-1,i}\\ &= -\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathcal{T}^{T}_{m}(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\delta\mathbf{h}_m,\end{aligned} \tag{3.23}$$

where the commutativity of the convolution product has been used in the third equality. So, since

$$\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i} = \mathbf{g}^{\mathrm{ZF}}_{m-1,i} + \delta\mathbf{g}_i + O\|\delta\mathbf{h}_m\|^2,$$

applying the chain differential rule to the deterministic model of tails gives the asymptotic bias:

$$\mathrm{E}(\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}) - \mathbf{g}^{\mathrm{ZF}}_{m-1,i} = \mathbf{B}^{\mathrm{alg}}_{g_i,\mathbf{R}^m_m}\mathbf{d}_{m,M} + O(\|\mathbf{d}_{m,M}\|^2) + O\left(\frac{1}{N}\right),$$

with

$$\mathbf{B}^{\mathrm{alg}}_{g_i,\mathbf{R}^m_m} \stackrel{\mathrm{def}}{=} -\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathcal{T}^{T}_{m}(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\mathbf{B}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}.$$

Results 1, 2, and 3 hold for $\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}$ upon using $\mathbf{A}^{\mathrm{alg}}_{g_i,\mathbf{R}^M_m}$ in place of $\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^M_m}$:

$$\mathbf{A}^{\mathrm{alg}}_{g_i,\mathbf{R}^M_m} \stackrel{\mathrm{def}}{=} -\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathcal{T}^{T}_{m}(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^M_m}. \tag{3.24}$$

Furthermore, in the statistical model of tails, results (3.16), (3.17), (3.18), (3.19) and (3.20) hold provided $\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}$ and $\mathbf{B}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}$ are replaced respectively by $\mathbf{A}^{\mathrm{alg}}_{g_i,\mathbf{R}^m_m}$ and $\mathbf{B}^{\mathrm{alg}}_{g_i,\mathbf{R}^m_m}$ with:

$$\mathbf{A}^{\mathrm{alg}}_{g_i,\mathbf{R}^m_m} \stackrel{\mathrm{def}}{=} -\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathcal{T}^{T}_{m}(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}.$$

[4]We note that this last mapping is defined only for analysis purposes as $\mathbf{h}_M$ is unknown to the receiver.

3.2.2 Combined channel-equalizer

From (3.1), the combined channel-equalizer impulse response (3.22) reads:

$$\begin{aligned}\widehat{\mathbf{f}}_{m+M-1,i} &= \mathcal{T}^T_{m-1}(\mathbf{h}^z_{m,M})\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i} + \mathcal{T}^T_{m-1}(\mathbf{d}^z_{m,M})\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i},\\ &= \begin{bmatrix} \mathbf{0}_{m_1} \\ \mathcal{T}^T_{m-1}(\mathbf{h}_m)\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i} \\ \mathbf{0}_{m_2} \end{bmatrix} + \mathcal{T}^T_{m-1}(\mathbf{d}^z_{m,M})\widehat{\mathbf{g}}^{\mathrm{ZF}}_{m-1,i}. \end{aligned} \tag{3.25}$$

The mapping $\widehat{\mathbf{h}}_m \longmapsto \widehat{\mathbf{f}}_{m+M-1,i}$ given by (3.25) is differentiable at $\mathbf{h}_m$ with the differential function:

$$\delta\mathbf{f}_i = -\begin{bmatrix} \mathbf{0}_{m_1} \\ \mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\delta\mathbf{h}_m \\ \mathbf{0}_{m_2} \end{bmatrix} - \mathcal{T}^T_{m-1}(\mathbf{d}^z_{m,M})\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\delta\mathbf{h}_m, \tag{3.26}$$

thanks to $\mathcal{T}^T_{m-1}(\mathbf{h}_m)\delta\mathbf{g}_i = -\mathcal{T}^T_{m-1}(\delta\mathbf{h}_m)\mathbf{g}^{\mathrm{ZF}}_{m-1,i} = -\mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\delta\mathbf{h}_m$, where the first and the second equalities come respectively from the differentiation of $\mathcal{T}^T_{m-1}(\mathbf{h}_m)\mathbf{g}^{\mathrm{ZF}}_{m-1,i} = \mathbf{e}_{2m,i+1}$ and the commutativity of the convolution product for the first term of (3.26) and thanks to (3.23) for its second part. Because $\mathbf{g}^{\mathrm{ZF}}_{m-1,i}$ equalizes perfectly $\mathbf{h}_m$, we may use the commutativity of the convolution product to obtain

$$\begin{aligned}\mathbf{f}_{m+M-1,i} &= \mathcal{T}^T_{m-1}(\mathbf{h}^z_{m,M})\mathbf{g}^{\mathrm{ZF}}_{m-1,i} + \mathcal{T}^T_{m-1}(\mathbf{d}^z_{m,M})\mathbf{g}^{\mathrm{ZF}}_{m-1,i},\\ &= \mathbf{e}_{m+M,m_1+i+1} + \mathcal{T}^T_M(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\mathbf{d}^z_{m,M} = \mathbf{e}_{m+M,m_1+i+1} + \mathcal{T}^T_M(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\mathbf{I}^{\ddagger}\mathbf{d}_{m,M},\end{aligned}$$

So, using Result 1, we have for the deterministic model of tails the following asymptotic bias:

$$\mathrm{E}(\widehat{\mathbf{f}}_{m+M-1,i}) - \mathbf{e}_{m+M,m_1+i+1} = \left(\mathcal{T}^T_M(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\mathbf{I}^{\ddagger} - \begin{bmatrix} \mathbf{0}_{m_1} \\ \mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i}) \\ \mathbf{0}_{m_2} \end{bmatrix} \mathbf{B}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}\right)\mathbf{d}_{m,M} + O\|\mathbf{d}_{m,M}\|^2 + O\left(\frac{1}{N}\right).$$

Results (3.12), (3.13) and (3.14) hold for $\widehat{\mathbf{f}}_{m+M-1,i}$ provided $\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^M_m}$ is replaced by $\mathbf{A}^{\mathrm{alg}}_{f_i,\mathbf{R}^M_m}$

$$\mathbf{A}^{\mathrm{alg}}_{f_i,\mathbf{R}^M_m} = -\left(\begin{bmatrix} \mathbf{0}_{m_1} \\ \mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i}) \\ \mathbf{0}_{m_2} \end{bmatrix} + \mathcal{T}^T_{m-1}(\mathbf{d}^z_{m,M})\mathcal{T}^{-T}_{m-1}(\mathbf{h}_m)\mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i})\right)\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^M_m}.$$

In the statistical model of tails, relations (3.16), (3.17), (3.18), (3.19) and (3.20) hold for $\mathbf{f}_{m,i}$ thanks to (3.26) and (3.15), provided $\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}$ and $\mathbf{B}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}$ are respectively replaced by $\mathbf{A}^{\mathrm{alg}}_{f_i,\mathbf{R}^m_m}$ and $\mathbf{B}^{\mathrm{alg}}_{f_i,\mathbf{R}^m_m}$:

$$\mathbf{A}^{\mathrm{alg}}_{f_i,\mathbf{R}^m_m} \stackrel{\mathrm{def}}{=} -\begin{bmatrix} \mathbf{0}_{m_1} \\ \mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i}) \\ \mathbf{0}_{m_2} \end{bmatrix} \mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^m_m} \quad \text{and} \quad \mathbf{B}^{\mathrm{alg}}_{f_i,\mathbf{R}^m_m} \stackrel{\mathrm{def}}{=} -\begin{bmatrix} \mathbf{0}_{m_1} \\ \mathcal{T}^T_m(\mathbf{g}^{\mathrm{ZF}}_{m-1,i}) \\ \mathbf{0}_{m_2} \end{bmatrix} \mathbf{B}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}$$

3.2.3 Open eye measure

Results concerning byproducts such as the open eye measure (OEM) can be deduced. Following our functional approach, the chain of operations

$$\widehat{\mathbf{R}}^M_m \longmapsto \widehat{\mathbf{f}}_{m+M-1,i} \longmapsto \mathrm{OEM}(\widehat{\mathbf{f}}_{m+M-1,i}) \stackrel{\mathrm{def}}{=} \frac{\sum_{k\neq i}\widehat{f}^2_{k,i}}{\widehat{f}^2_{i,i}},$$ [5]

[5] We suppose here that the term $\widehat{f}_{i,i}$ is the dominant term of the combined channel-equalizer response $\widehat{\mathbf{f}}_{m+M-1,i}$

allows the asymptotic statistics of $\mathrm{OEM}(\widehat{\mathbf{f}}_{m+M-1,i})$ to be deduced. The mapping $\widehat{\mathbf{f}}_{m+M-1,i} \longmapsto \mathrm{OEM}(\widehat{\mathbf{f}}_{m+M-1,i})$ is differentiable to the second-order at the point $\mathbf{e}_{m+M,i+1}$, with a zero first-order derivative and a second-order derivative $\mathbf{\Delta} = 2(\mathbf{I}_{m+M} - \mathbf{e}_{m+M,m_1+i+1}\mathbf{e}_{m+M,m_1+i+1}^T)$, so that

$$\mathrm{OEM}(\widehat{\mathbf{f}}_{m+M-1,i}) = 0 + 0 + \frac{1}{2}\delta\mathbf{f}_i^T\mathbf{\Delta}\delta\mathbf{f}_i + o\|\delta\mathbf{f}_i\|^2 = \frac{1}{2}\mathrm{Tr}(\delta\mathbf{f}_i\delta\mathbf{f}_i^T\mathbf{\Delta}) + o\|\delta\mathbf{f}_i\|^2.$$

So, in the statistical model of tails, we have from

$$\delta\mathbf{f}_i = \mathbf{A}^{\mathrm{alg}}_{f_i,\mathbf{R}_m^m}\mathrm{Vec}(\delta\mathbf{R}_m^m) + \mathbf{B}^{\mathrm{alg}}_{f_i,\mathbf{R}_m^m}\mathbf{d}_{m,M} + O\|\delta\mathbf{R}\|^2,$$

the asymptotic mean open eye measure when $N \to \infty$ and $\sigma_d^2 \to 0$:

$$\mathrm{E}\left(\frac{\sum_{k\neq i}\hat{f}_{k,i}^2}{\hat{f}_{i,i}^2}\right) \sim \frac{1}{N}\mathrm{Tr}\left(\frac{1}{2}\mathbf{A}^{\mathrm{alg}}_{f_i,\mathbf{R}_m^m}\mathbf{C}_{R_m^m}\left(\mathbf{A}^{\mathrm{alg}}_{f_i,\mathbf{R}_m^m}\right)^T\mathbf{\Delta}\right) + \frac{\sigma_d^2}{2(M-m)}\mathrm{Tr}\left(\frac{1}{2}\mathbf{B}^{\mathrm{alg}}_{f_i,\mathbf{R}_m^m}\left(\mathbf{B}^{\mathrm{alg}}_{f_i,\mathbf{R}_m^m}\right)^T\mathbf{\Delta}\right).$$

4 Application to the LS, SS, LP and OPD methods

We proceed with the derivation of the matrices $\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}_m^m}$ associated with the differential of the mappings $\mathrm{alg}(\cdot)$ at point $\mathbf{R}_m^m$, as all other quantities defined in the previous section are derived from it. In particular, the matrices $\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}_m^M}$ are deduced from $\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}_m^m}$ by replacing respectively $\mathbf{S}_m(\mathbf{h}_m)$, $(\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m))^\natural$ and $(\mathcal{T}_{m-1}(\mathbf{h}_m)\mathcal{T}_{m-1}^T(\mathbf{h}_m))^{-1}$ by $\mathbf{S}_m(\mathbf{h}_M)$, $(\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M))^\natural$ [6] and $(\mathcal{T}_{m-1}(\mathbf{h}_M)\mathcal{T}_{m-1}^T(\mathbf{h}_M))^{-1}$. As usual, the mapping $\widehat{\mathbf{R}}_m^M \overset{\mathrm{alg}}{\longmapsto} \widehat{\mathbf{h}}_m$ is built by replacing respectively $\mathbf{R}_m^M$ and σ_n^2 by $\widehat{\mathbf{R}}_m^M$ and $\widehat{\sigma}_n^2 = \widehat{\mathbf{v}}_{2(m+1)}^T\widehat{\mathbf{R}}_m^M\widehat{\mathbf{v}}_{2(m+1)}$ (where $\widehat{\mathbf{v}}_{2(m+1)}$ is the eigenvector of $\widehat{\mathbf{R}}_m^M$ associated with its smallest eigenvalue), in the relations given in Sections 2.2, 2.3 and 2.4 relating $\mathbf{R}_m$ to $\mathbf{h}_m$. In this Section, $\mathbf{R}_m^m$ is denoted as $\mathbf{R}_m$ for simplicity.

4.1 LS and SS methods

Thanks to a perturbation result [20], concerning the eigenvector associated with the unique smallest eigenvalue of $\mathbf{R}_m$, (2.2) gives the differential $\delta\mathbf{h}_m$:

$$\begin{aligned}
\delta\mathbf{h}_m &= -\mathbf{T}_m{\mathbf{R}'_m}^\natural\delta\mathbf{R}_m\mathbf{v}_{2(m+1)} + O(\|\delta\mathbf{R}_m\|^2),\\
&= -\mathbf{T}_m(\mathbf{v}_{2(m+1)}^T \otimes {\mathbf{R}'_m}^\natural)\mathrm{Vec}(\delta\mathbf{R}_m) + O(\|\delta\mathbf{R}_m\|^2),\\
&= -\mathbf{T}_m(\mathbf{h}_m^T\mathbf{T}_m \otimes [\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural)\mathrm{Vec}(\delta\mathbf{R}_m) + O(\|\delta\mathbf{R}_m\|^2),
\end{aligned}$$

Thus,

$$\mathbf{A}^{\mathrm{SS}}_{h,\mathbf{R}_m^m} = -\mathbf{T}_m(\mathbf{h}_m^T\mathbf{T}_m \otimes [\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural). \tag{4.1}$$

4.2 LP method

After some modifications, [14, rel.(33)] reads:

$$\delta\mathbf{h}_m = \delta\mathbf{S}_m\mathbf{g}_m - \frac{1}{2\lambda}\mathrm{Tr}(\delta\mathbf{D}\mathbf{v}\mathbf{v}^T)\mathbf{h}_m - \mathbf{S}_m\begin{bmatrix}\mathbf{O} & \mathbf{O}\\ \mathbf{O} & {\mathbf{R}'^{-1}}_{m-1}\end{bmatrix}\delta\mathbf{R}'_m\mathbf{g}_m + O(\|\delta\mathbf{R}_m\|^2). \tag{4.2}$$

From [4], it can be shown that:

$$\mathrm{Tr}(\delta\mathbf{D}\mathbf{v}\mathbf{v}^T) = \lambda\mathbf{g}_m^T\delta\mathbf{R}'_m\mathbf{g}_m. \tag{4.3}$$

[6] where $(.)^\natural$ denotes here the operation that consists in forcing to zero the smallest eigenvalue of $(.)$ and then inverting the truncated version of $(.)$ in its range space.

The matrix $\mathbf{S}_m$ is a linear function of the matrix $\mathbf{R}'_m$:

$$\mathrm{Vec}(\mathbf{S}_m) = \mathbf{C}_2 \mathrm{Vec}(\mathbf{R}'_m), \tag{4.4}$$

with

$$\mathbf{C}_2 \stackrel{\text{def}}{=} \begin{bmatrix} & \mathbf{I}_2 \otimes (\mathbf{Z}_{m+1}^0 \mathbf{J}_{m+1}) \\ \mathbf{O}_{2(m+1)^2,2(m+1)m} & \mathbf{I}_2 \otimes (\mathbf{Z}_{m+1}^1 \mathbf{J}_{m+1}) \\ & \vdots \\ & \mathbf{I}_2 \otimes (\mathbf{Z}_{m+1}^m \mathbf{J}_{m+1}) \end{bmatrix} \otimes \mathbf{I}_2.$$

Since the differential $\delta \mathbf{v}_{2(m+1)}$ is orthogonal to $\mathbf{v}_{2(m+1)}$ (because $\|\mathbf{v}_{2(m+1)}\| = 1$), we have:

$$\begin{aligned} \delta\sigma_n^2 &= \delta(\mathbf{v}_{2(m+1)}^T \mathbf{R}_m \mathbf{v}_{2(m+1)}) = \mathbf{v}_{2(m+1)}^T \delta\mathbf{R}_m \mathbf{v}_{2(m+1)} + O(\|\delta\mathbf{R}_m\|^2) \\ &= (\mathbf{v}_{2(m+1)}^T \otimes \mathbf{v}_{2(m+1)}^T)\mathrm{Vec}(\delta\mathbf{R}_m) + O(\|\delta\mathbf{R}_m\|^2). \end{aligned}$$

Thus

$$\mathrm{Vec}(\mathbf{R}'_m) = \mathbf{C}_3 \mathrm{Vec}(\mathbf{R}_m), \tag{4.5}$$

with $\mathbf{C}_3 \stackrel{\text{def}}{=} \mathbf{I}_{4(m+1)^2} - \mathrm{Vec}(\mathbf{I}_{2(m+1)})(\mathbf{v}_{2(m+1)}^T \otimes \mathbf{v}_{2(m+1)}^T)$. Finally, putting together (4.2), (4.3), (4.4) and (4.5), we get $\delta\mathbf{h}_m = \mathbf{A}_{h,\mathbf{R}_m^m}^{\mathrm{LP}} \mathrm{Vec}(\delta\mathbf{R}_m) + O\|\delta\mathbf{R}_m\|^2$ with:

$$\mathbf{A}_{h,\mathbf{R}_m^m}^{\mathrm{LP}} = \left((\mathbf{g}_m^T \otimes \mathbf{I}_{2(m+1)})\mathbf{C}_2 - \left(\mathbf{g}_m^T \otimes \left[\frac{1}{2}\mathbf{h}_m \mathbf{g}_m^T + \mathbf{S}_m \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & (\mathcal{T}_{m-1}(\mathbf{h}_m)\mathcal{T}_{m-1}^T(\mathbf{h}_m))^{-1} \end{pmatrix} \right] \right) \right) \mathbf{C}_3. \tag{4.6}$$

4.3 OPD method

We note that the mapping $\widehat{\mathbf{R}}_m^M \stackrel{\mathrm{OPD}}{\longmapsto} \widehat{\mathbf{h}}_m$ is a composition of differentiable mappings. Thus is differentiable at point $\mathbf{R}_m$ despite the pseudo-inverse $[\widehat{\mathbf{R}}_m^M - \widehat{\sigma}_n^2\mathbf{I}]^\natural$ being included in relation (2.3) of the OPD algorithm, because $\widehat{\mathbf{R}}_m^M - \widehat{\sigma}_n^2\mathbf{I}$ is singular with rank $2m+1$, as with $\mathbf{R}_m - \sigma_n^2\mathbf{I}_{2(m+1)}$.

Thanks to a perturbation result [20], concerning the eigenvector associated with the unique biggest eigenvalue λ of $\mathbf{D}_3$, (2.4) gives the perturbation $\delta\mathbf{h}_m$:

$$\begin{aligned} \delta\mathbf{h}_m &= -(\mathbf{h}_m\mathbf{h}_m^T - \mathbf{I}_{2(m+1)})^\natural \delta\mathbf{D}_3 \mathbf{h}_m + O(\|\delta\mathbf{D}_3\|^2) \\ &= -(\mathbf{h}_m^T \otimes (\mathbf{h}_m\mathbf{h}_m^T - \mathbf{I}_{2(m+1)})^\natural)\mathrm{Vec}(\delta\mathbf{D}_3) + O(\|\delta\mathbf{D}_3\|^2), \\ &= -(\mathbf{h}_m^T \otimes (\mathbf{h}_m\mathbf{h}_m^T - \mathbf{I}_{2(m+1)})^\natural)\mathbf{C}_4\mathrm{Vec}(\delta\mathbf{D}_1) + O(\|\delta\mathbf{D}_3\|^2), \end{aligned} \tag{4.7}$$

with $\mathrm{Vec}(\delta\mathbf{D}_3) = \mathbf{C}_4\mathrm{Vec}(\delta\mathbf{D}_1)$, where, from (2.3), $\mathbf{C}_4$ is given by:

$$\mathbf{C}_4 \stackrel{\text{def}}{=} \mathbf{I}_{4(m+1)^2} - (\mathbf{Z}_{m+1} \otimes \mathbf{I}_2) \otimes (\mathbf{Z}_{m+1} \otimes \mathbf{I}_2). \tag{4.8}$$

From (2.3), we get:

$$\delta\mathbf{D}_1 = \delta\mathbf{S}_m\mathbf{R}_m'^\natural\mathbf{S}_m^T + \mathbf{S}_m\mathbf{R}_m'^\natural\delta\mathbf{S}_m^T + \mathbf{S}_m\delta\mathbf{R}_m'^\natural\mathbf{S}_m^T + O(\|\delta\mathbf{R}_m\|^2) \tag{4.9}$$

with $\delta\mathbf{R}_m'^\natural = -\mathbf{R}_m'^\natural\delta\mathbf{R}'_m\mathbf{R}_m'^\natural$. So, by vectorization, we get:

$$\begin{aligned} \mathrm{Vec}(\delta\mathbf{D}_1) &= (\mathbf{S}_m\mathbf{R}_m'^\natural \otimes \mathbf{I}_{2(m+1)})\mathrm{Vec}(\delta\mathbf{S}_m) + (\mathbf{I}_{2(m+1)} \otimes \mathbf{S}_m\mathbf{R}_m'^\natural)\mathbf{K}_{2(m+1),2(m+1)}\mathrm{Vec}(\delta\mathbf{S}_m) \\ &\quad -(\mathbf{S}_m\mathbf{R}_m'^\natural \otimes \mathbf{S}_m\mathbf{R}_m'^\natural)\mathrm{Vec}(\delta\mathbf{R}'_m) + O(\|\delta\mathbf{R}_m\|^2). \end{aligned} \tag{4.10}$$

Finally, putting together (4.7), (4.10), (4.4) and (4.5), we get $\delta\mathbf{h}_m = \mathbf{A}_{h,\mathbf{R}_m^m}^{\mathrm{OPD}}\mathrm{Vec}(\delta\mathbf{R}_m) + O\|\delta\mathbf{R}_m\|^2$ with:

$$\begin{aligned} \mathbf{A}_{h,\mathbf{R}_m^m}^{\mathrm{OPD}} &= -\left(\mathbf{h}_m^T \otimes (\mathbf{h}_m\mathbf{h}_m^T - \mathbf{I}_{2(m+1)})^\natural\right)\mathbf{C}_4(((\mathbf{S}_m[\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural \otimes \mathbf{I}_{2(m+1)}) \\ &\quad +(\mathbf{I}_{2(m+1)} \otimes \mathbf{S}_m[\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural)\mathbf{K}_{2(m+1),2(m+1)})\mathbf{C}_2\mathbf{C}_3 - (\mathbf{S}_m\mathbf{R}'_m \otimes \mathbf{S}_m\mathbf{R}'_m)\mathbf{C}_3). \end{aligned} \tag{4.11}$$

4.4 Analysis of the results

As shown in Section 3, the performance in terms of asymptotic bias and variance in the deterministic model of tails and mean square errors in the statistical model of tails are directly related to matrices $\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^m_m}$ and $\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}$, but relations (4.1), (4.6) and (4.11) are lacking engineering insight and as such are complicated to analyze. However, we see in the following that these performance depend on the significant part $\mathbf{h}_m$ of the impulse response $\mathbf{h}_M$ through its *diversity* and on the sensitivity of this diversity adapted to each algorithm.

Influence of diversity. The significant part $\mathbf{h}_m$ of the impulse response $\mathbf{h}_M$ acts upon $\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^m_m}$ through $\mathbf{R}'^{\natural}_m = [\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}^T_m(\mathbf{h}_m)]^{\natural}$ for the SS/LS and the OPD algorithms and through $\mathbf{R}'^{-1}_{m-1} = [\mathcal{T}_{m-1}(\mathbf{h}_m)\mathcal{T}^T_{m-1}(\mathbf{h}_m)]^{-1}$ for the LP algorithm if $\mathbf{h}_m$ is normalized. In fact, the behaviors of the terms $[\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}^T_m(\mathbf{h}_m)]^{\natural}$ and $[\mathcal{T}_{m-1}(\mathbf{h}_m)\mathcal{T}^T_{m-1}(\mathbf{h}_m)]^{-1}$ are very close because they are respectively dominated by the inverse of the square of the singular values

$$\delta_m \stackrel{\text{def}}{=} \sigma_{2m+1}(\mathcal{T}_m(\mathbf{h}_m)) \qquad \text{and} \qquad \delta'_m \stackrel{\text{def}}{=} \sigma_{2m}(\mathcal{T}_{m-1}(\mathbf{h}_m))$$

which are not orderable but practically very close to each other. These singular values may be interpreted as a measure of *diversity* of $\mathbf{h}_m$ [3] as they measure respectively the distance in the matrix 2-norm of $\mathcal{T}_m(\mathbf{h}_m)$ and $\mathcal{T}_{m-1}(\mathbf{h}_m)$ from the matrices of rank $2m$ and $2m-1$, violating thus the rank assumptions. So, the performance (asymptotic bias in the deterministic model of tails and mean square error in the statistical model of tails) of the algorithms degrade when this diversity decreases. This diversity of the significant part $\mathbf{h}_m$ of the impulse reponse $\mathbf{h}_M$ acts upon $\mathbf{A}^{\text{alg}}_{h,\mathbf{R}^M_m}$ as well, because $\sigma_{2m+1}(\mathcal{T}_m(\mathbf{h}_M)) \approx \sigma_{2m+1}(\mathcal{T}_m(\mathbf{h}_m))$ and $\sigma_{2m}(\mathcal{T}_{m-1}(\mathbf{h}_M)) \approx \sigma_{2m}(\mathcal{T}_{m-1}(\mathbf{h}_m))$. Thus the variances of the estimates given in the deterministic model of tails degrade too, when this diversity decreases.

Influence of the sensitivity of this diversity adapted to each algorithm. Concerning the bias performance, $\sigma_1(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m})$ can be considered as a better measure of *diversity sensitivity* of $\mathbf{h}_m$ adapted to each algorithm than δ_m and δ'_m which do not depend on the algorithm used. We note that the bias norm upper bound $\sigma_1(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m})\|\mathbf{d}_{m,M}\|$ given in the Result 1 is *attainable* for the worst-case tail $\mathbf{d}_{m,M}$ (i.e. the tail which maximizes the bias norm $\|\mathrm{E}(\widehat{\mathbf{h}}_m) - \mathbf{h}_m\|$ for fixed $\|\mathbf{d}_{m,M}\|$) which is colinear with the right singular vector of $\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}$ associated with its largest singular value. This worst-case tail is of the form (proved in Appendix B for the LS/SS algorithm but only observed by computer for the LP and OPD algorithms): $\mathbf{d}_{m,M} = (\mathbf{d}^T_{m_1}, \mathbf{d}^T_{m_2})^T$ with:

$$\mathbf{d}_{m_1} = [\underbrace{\mathbf{0}^T \cdots \mathbf{0}^T}_{m_1-m}, \mathbf{d}^{l\,T}_1 \cdots \mathbf{d}^{l\,T}_m]^T \text{ if } m_1 \geq m \text{ and } \mathbf{d}_{m_2} = [\mathbf{d}^{r\,T}_1 \cdots \mathbf{d}^{r\,T}_m, \underbrace{\mathbf{0}^T \cdots \mathbf{0}^T}_{m_2-m}]^T \text{ if } m_2 \geq m \quad (4.12)$$

and $\mathbf{d}^l_1, \ldots, \mathbf{d}^l_m$ and $\mathbf{d}^r_1, \ldots, \mathbf{d}^r_m$ do not depend on m_1 and m_2, but depend on the algorithm, given $\mathbf{h}_m$. Thus, the "worst" tail gathers on both side of the significant part along a length equal to the order of this significant part. Concerning the mean square error in the statistical model of tails, the part $\frac{\sigma^2_d}{2(M-m)}\mathrm{Tr}\left(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}\left(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}\right)^T\right)$, attributed to the tails in the Result 3, depends only on the tail energy per term if $m_1, m_2 \geq m$. More precisely, it is proved in Appendix B for the LS/SS algorithm, but observed by computer only for the LP and OPD algorithms that:

$$\begin{aligned} \mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}\left(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}^m_m}\right)^T &= \mathbf{\Psi}(\mathbf{h}_m, m, m_2), \qquad m_1 \geq m \\ &= \mathbf{\Psi}(\mathbf{h}_m, m_1, m), \qquad m_2 \geq m \\ &= \mathbf{\Psi}(\mathbf{h}_m, m, m), \qquad m_1, m_2 \geq m. \end{aligned} \quad (4.13)$$

Furthermore, it is shown by simulation that $\sigma_1^2(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}_m^m})$ is the dominant term of $\text{Tr}\left(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}_m^m}\left(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}_m^m}\right)^T\right) = \sum_{i=1}^{2(M-m)} \sigma_i^2(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}_m^m})$. So $\sigma_1(\mathbf{B}^{\text{alg}}_{h,\mathbf{R}_m^m})$ can also be interpretated as a measure of *diversity sensitivity* of the algorithm with respect to the tails, for the mean square error of $\widehat{\mathbf{h}}_m^{\text{alg}}$ in the statistical model of tails.

Relation with previous works. The bias norm upper bound given in the Result 1 to the first order can be compared to the upper bound of the errors of the estimates $\widehat{\mathbf{h}}_m$ given in [3] and [4] for the SS/LS and LP methods, respectively, in exact statistics situation:

$$\| \lim_{N\to\infty} \text{E}(\widehat{\mathbf{h}}_m^{\text{SS}}) - \mathbf{h}_m\| \leq 2\sqrt{2(m+1)}\frac{\|\mathbf{d}_{m,M}\|}{\delta_m} + \|\mathbf{d}_{m,M}\|^2 \tag{4.14}$$

$$\| \lim_{N\to\infty} \text{E}(\widehat{\mathbf{h}}_m^{\text{LP}}) - \mathbf{h}_m\| \leq \sqrt{m+1}\frac{\|\mathbf{d}_{m,M}\|}{\delta'_m} + (2m+3)\frac{\|\mathbf{d}_{m,M}\|}{\|\mathbf{h}_{(m_1)}\|}\sqrt{1+\frac{1}{{\delta'_m}^2}}, \tag{4.15}$$

where $\mathbf{h}_{(m_1)}$ is the first term of the significant part of $\mathbf{h}_M$. These upper bounds are proportional to $\|\mathbf{d}_{m,M}\|$ in the first-order of $\|\mathbf{d}_{m,M}\|$ and are respectively inversely proportional to δ_m and δ'_m (when $\delta'_m \ll 1$), while the bounds (3.11) are dominated respectively by the inverse of the square of δ_m and δ'_m for respectively the SS/LS and the LP methods. The bounds (4.14) and (4.15) are shown to be rather loose in the following Section, as compared to the bound given by the Result 1.

5 Simulations

In this Section, we examine through examples of the performance of the LS/SS, LP and OPD methods, the accuracy of the expressions of the bias, the mean square error of our estimators, and we investigate the sample size and the tails size domains for which our asymptotic approach is valid. We consider throughout this section an impulse response $\mathbf{h}_M$ with $M = 12$, where the order of the significant part is $m = 2$. We present two types of significant part $\mathbf{h}_2$ with $\|\mathbf{h}_2\|^2 = 1$. One offers "great" diversity:

$$\mathbf{h}_2 = \left[\ -.6804 \quad .4281; \quad .1770 \quad -.2446; \quad -.0902 \quad -.5043\ \right]^T \tag{5.1}$$

with $\sigma_5(\mathcal{T}_2(\mathbf{h}_2)) = 0.4157$ and $\sigma_4(\mathcal{T}_1(\mathbf{h}_2)) = 0.4165$, and the other "poor" diversity:

$$\mathbf{h}_2 = \left[\ -.6804 \quad .5902; \quad .1770 \quad -.2656; \quad -.0902 \quad -.2803\ \right]^T \tag{5.2}$$

with $\sigma_5(\mathcal{T}_2(\mathbf{h}_2)) = 0.2369$ and $\sigma_4(\mathcal{T}_1(\mathbf{h}_2)) = 0.2354$. In all the simulations the order m is correctly detected beforehand by the procedure described in [5],[6]. For each experiment, 1000 independent Monte Carlo simulations are performed. The signal to noise ratio $SNR = 10\log\frac{\|\mathbf{h}_M\|^2}{2\sigma_n^2}$ is fixed to 17dB except in Fig.4 and Fig.6.

The first experiment presents the deterministic model of tails and examines the performance of the different second-order algorithms. Table 1 compares the theoretical asymptotic bias of $\widehat{\mathbf{h}}_2$ given by Result 1, with the estimated bias given by simulation for $N = 300$ and $N = 1000$ and for two proportional tails: $\|\mathbf{d}_{2,12}\| = 0.05$ (for which $\sigma_5(\mathcal{T}_2(\mathbf{h}_{12})) = 0.4195$ and $\sigma_4(\mathcal{T}_1(\mathbf{h}_{12})) = 0.4220$) and $\|\mathbf{d}_{2,12}\| = 0.1$ (for which $\sigma_5(\mathcal{T}_2(\mathbf{h}_{12})) = .4253$ and $\sigma_4(\mathcal{T}_1(\mathbf{h}_{12})) = .4300$) for the channel with significant part $\mathbf{h}_2$ shown in (5.1). This table shows a good agreement between theoretical and estimated values. The difference between these values increases with increasing $\|\mathbf{d}_{2,12}\|$ and decreases with increasing N, which is explained by the second-order term in $\|\mathbf{d}_{2,12}\|$ and the first-order term in $1/N$. The numerical values of the attainable bounds (3.11) for the LS/SS and LP algorithms (for $\|\mathbf{d}_{2,12}\| = 0.1$) are respectively 0.2282 and 0.1917.

Compared with upper bounds (4.14) and (4.15), 1.1885 and 2.6809, these latter bounds are not very tight. Figures 3 and 4 plot the theoretical mean square error (MSE) of $\widehat{\mathbf{h}}_2$:

$$\mathrm{E}\|\widehat{\mathbf{h}}_2 - \mathbf{h}_2\|^2 = \|\mathrm{bias}(\widehat{\mathbf{h}}_2)\|^2 + \mathrm{Tr}(\mathrm{Cov}\widehat{\mathbf{h}}_2) \sim \|\mathbf{B}^{\mathrm{alg}}_{h,\mathbf{R}^m_m}\mathbf{d}_{m,M}\|^2 + \frac{1}{N}\mathrm{Tr}\left(\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^M_m}\mathbf{C}_{R^M_m}{\mathbf{A}^{\mathrm{alg}}_{h,\mathbf{R}^M_m}}^T\right)$$

and the estimated MSE given by simulation versus the sample size and the signal to noise ratio, respectively. (N ranges from 30 to 1000 and the signal to noise ratio from 8dB to 37db). We observe that the SS/LS algorithm outperforms the LP and OPD algorithms. Furthermore the LP and the OPD mean square errors are almost equivalent, with a slight superiority of the OPD algorithm.

The second experiment presents the statistical model of tails and examines the performance of the different second-order algorithms. Fig. 5 exhibits the theoretical MSE (3.18) of $\widehat{\mathbf{h}}_2$ and its estimated MSE obtained by simulation for a signal to noise ratio of 17dB in two situations: in no channel tail situation versus the sample size and in exact statistics situation versus the energy of the tails $\mathrm{E}(\|\mathbf{d}_{2,12}\|^2)$. We observe that if we separate the effects of the tails and of the finite sample size, the three algorithms under study are almost equivalent with respect to the tail sensitivity, but the LS/SS algorithm outperforms the other algorithms with respect to the finite sample size sensitivity. In Figures 6 and 7, the finite sample size and the tail contributions are simultaneously present. The two figures compare the theoretical MSE of $\widehat{\mathbf{h}}_2$ and its estimated MSE obtained by simulation, for $N = 300$, versus the signal to noise ratio for $\mathrm{E}\|\mathbf{d}_{2,12}\|^2 = 0.01$ fixed and versus the energy of the tails $\mathrm{E}\|\mathbf{d}_{2,12}\|^2$ for a signal to noise ratio fixed at 17dB. The adequation between the theoretical and the estimated MSE is good except for the LP algorithm for which 300 samples is too small (see Fig.8). Furthermore, in Fig. 7, the two channels given by (5.1) and (5.2) are exhibited. Naturally these MSEs increase when the diversity decreases. But we note that unlike the LS/SS algorithm, the LP and OPD algorithms are less sensitive to the diversity of the significant part $\mathbf{h}_m$. Fig. 8 compares the theoretical MSE of the estimated significant part $\widehat{\mathbf{h}}_2$, the estimated zero-forcing equalizer $\widehat{\mathbf{g}}^{\mathrm{ZF}}_{2,2}$ and the combined channel-equalizer $\widehat{\mathbf{f}}_{2,2}$, and the theoretical mean of $\mathrm{OEM}(\widehat{\mathbf{f}}_{2,2})$ with the estimated MSE and estimated mean obtained by simulation. We note good agreement between the theoretical and estimated MSE and mean open eye measure for $N > 300$. The performance of the LP and OPD algorithms are equivalent but the LS/SS algorithm outperforms the other algorithms in presence of finite sample sizes and channel tails. Naturally, the conclusions of these two simulations must be mitigated because a thorough comparison between the studied algorithms would need a large quantity of scenarios (various channels, m, M, N and SNR), but is beyond the scope of this paper.

To see that our analysis breaks down when a partition between significant part and tails is ambiguous, we consider the popular multipath transfer function in raised-cosine. Unlike preceding papers (e.g. [1]), we retain most of the terms of the infinite length impulse response ($M = 40$). The so computed impulse reponse $\mathbf{h}_M$ is inevitably ill-conditioned. However its effective part is better conditioned and consequently it may be blindly identified. We choose the three ray multipath channel $c(t) = \delta(t) + 0.43\delta(t - 0.41T) + 0.41\delta(t - 0.89T)$ with a roll-off factor of 0.4. In this situation, the procedure given in [5] gives $m = 2$ and by forcing the value of m to 1, 2 3, 4 the theoretical and estimated MSE of $\widehat{\mathbf{h}}_m$ defined as

$$\mathrm{E}(\min_{m_1+m_2=M-m} \|\widehat{\mathbf{h}}_m - \left[\mathbf{O}_{2(m+1),2m_1}, \mathbf{I}_{2(m+1)}, \mathbf{O}_{2(m+1),2m_2}\right]\mathbf{h}_M\|^2)$$

with $\|\mathbf{h}_M\| = 1$ given for the SS algorithm are shown in Table 2. We see that our analysis based on a deterministic model of tails is valid for $m = 2$. We observe that a correct detection of the significant order m is critical. For $m > 2$, the diversity of $\mathbf{h}_m$ is very small, so the estimated and theoretical variance of $\widehat{\mathbf{h}}_m$ degrades considerably. We note that the corresponding theoretical values are large. In fact from (4.1), (4.6) and (4.11), the algorithms derivative involves the inversion of the channel covariance matrix,

which in this case is poorly conditionned. Our first order perturbation analysis is no longer valid. Only the SS algorithm is able to identify the effective response of our three ray multipath channel for roll-off factor 0.4 thanks to its better sensibility to the diversity of $\mathbf{h}_m$ (see Section 4.4). Furthermore we note that for weaker roll off factors, we are not in the context of an effective response clearly distinct from small tails.

6 Conclusion

We built a general functional methodology for studying the statistical performance of second-order methods for blind channel identification/equalization in practical situations, i.e., in presence of estimated second-order statistics from finite samples observation, non-negligible additive channel noise and long tails of "small" leading and/or trailing impulse response terms, we built . We proposed two models for the channel tails and we obtained general asymptotic statistics of the estimated significant part of the channel, the zero-forcing equalizer, the combined channel-equalizer impulse response and the open eye measure. These asymptotic statistics are valid in a large domain of tail size and sample size. It is shown that these performance measures are related to the diversity of the significant part of the channel, but also to a diversity sensitivity of this significant part adapted to each algorithm. Finally we applied our functional approach to the LS/SS, LP and OPD algorithms as examples.

almost equivalent algorithms, are less reponse, if the beforehand.

A Asymptotic Normality of $\widehat{\mathbf{R}}_m^M$

In this appendix, $\widehat{\mathbf{R}}_m^M$ and $\mathbf{R}_m^M$ are denoted $\widehat{\mathbf{R}}_N$ and $\mathbf{R}$ for simplicity and to specify the dependence on N of $\widehat{\mathbf{R}}_m^M$. Since x_k^i, $i = 1, 2$ associated with the impulse response $\mathbf{h}_M$ are M-dependent processes, we can apply the asymptotic normality results of [21, theorem 14, p.228] and [18, theorems 6.4.2 and 7.2.1]. Adapting these results to the multivariate process $\mathbf{x}_k$ and using some properties of the vec-permutation matrix, the vec-operator and Kronecker products given in [16], $\mathrm{Vec}(\widehat{\mathbf{R}}_N)$ is asymptotically Gaussian

$$\sqrt{N}(\mathrm{Vec}(\widehat{\mathbf{R}}_N) - \mathrm{Vec}(\mathbf{R})) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{C}_R)$$

with asymptotic covariance matrix given by $\mathbf{C}_R = \lim_{N\to\infty} N\mathrm{Cov}(\mathrm{Vec}(\widehat{\mathbf{R}}_N))$ with

$$\mathbf{C}_R = \int_{-1/2}^{+1/2} \mathbf{E}_m(f)^* \otimes \mathbf{E}_m(f) df + \int_{-1/2}^{+1/2} [\mathbf{E}_m(f)^* \otimes \mathbf{E}_m(f)]\mathbf{K}_{2(m+1),2(m+1)} df + \kappa \mathbf{Q}_m \tag{A.1}$$

where

$$\mathbf{E}_m(f) \stackrel{\mathrm{def}}{=} \mathbf{e}_m(f)\mathbf{e}_m^H(f) \otimes \mathbf{S}(f), \tag{A.2}$$

and

$$\mathbf{Q}_m \stackrel{\mathrm{def}}{=} \mathrm{Vec}(\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M))\mathrm{Vec}^T(\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M)). \tag{A.3}$$

$\mathbf{e}_m(f)$ denotes the vector $(1, e^{-i2\pi f}, \ldots, e^{-i2\pi m f})^T$. $\mathbf{S}(f)$ is the spectral density matrix of the 2-dimensional vector process $\mathbf{x}_k$. From (2.1) is easily seen that:

$$\mathbf{S}(f) = \left(\sum_{k=0}^{M} \mathbf{h}_{(k)} e^{-i2\pi k f}\right)\left(\sum_{k=0}^{M} \mathbf{h}_{(k)} e^{-i2\pi k f}\right)^H + \sigma_n^2 \mathbf{I}_2, \tag{A.4}$$

κ is the cumulant $\mathrm{cum}(s_k, s_k, s_k, s_k)$. Naturally the asymptotic normality of $\widehat{\mathbf{R}}_m^m$ is obtained in the same way by replacing $\mathbf{h}_M$ by $\mathbf{h}_m$ in (A.2), (A.3) and (A.4).

We note that the performance of the LS/SS method is insensitive to the distribution of the input s_k because the last term of (A.1) does not affect the asymptotic covariance of the estimates given by the LS/SS method. This is immediately shown because

$$\begin{aligned}\mathbf{A}^{\mathrm{SS}}_{h,\mathbf{R}^M_m}\mathbf{Q}_m{\mathbf{A}^{\mathrm{SS}}_{h,\mathbf{R}^M_m}}^T &= \left(\mathbf{T}_m(\mathbf{h}_m^T\mathbf{T}_m\otimes[\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M)]^\natural)\mathrm{Vec}(\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M))\right)(.)^T\\ &= \left(\mathbf{T}_m\mathrm{Vec}([\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M)]^\natural\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M)\mathbf{T}_m\mathbf{h}_m)\right)(.)^T\\ &= \left(\mathbf{T}_m\mathrm{Vec}([\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M)]^\natural\mathcal{T}_m(\mathbf{h}_M)\mathcal{T}_m^T(\mathbf{h}_M)\mathbf{v}_{2(m+1)}\right)(.)^T=\mathbf{O}.\end{aligned}$$

Second equality uses relation (1.1) and third equality is due to the orthogonality of $\mathbf{v}_{2(m+1)}$ to the column space of $\mathcal{T}_m(\mathbf{h}_M)$. This extends a result given in [15].

B Proof of relations (4.12) and (4.13) for the LS/SS algorithm

For the SS algorithm, (4.12) and (4.13) are proved thanks to the following simplification of $\mathbf{B}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}$ defined in (3.9) by $\mathbf{A}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}\mathrm{Vec}(\delta\mathbf{R}_d)=\mathbf{B}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}\mathbf{d}_{m,M}+O(\|\mathbf{d}_{m,M}\|^2)$. From (3.7),

$$\begin{aligned}&\mathbf{A}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}\mathrm{Vec}(\delta\mathbf{R}_d)\\ &\quad=-\mathbf{T}_m\left(\mathbf{h}_m^T\mathbf{T}_m\otimes[\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural\right)\mathrm{Vec}\left(\mathcal{T}_m(\mathbf{h}^z_{m,M})\mathcal{T}_m^T(\mathbf{d}^z_{m,M})+\mathcal{T}_m(\mathbf{d}^z_{m,M})\mathcal{T}_m^T(\mathbf{h}^z_{m,M})\right)+O(\|\mathbf{d}_{m,M}\|^2\\ &\quad=\mathbf{T}_m\mathrm{Vec}\left([\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural\left[\mathcal{T}_m(\mathbf{h}^z_{m,M})\mathcal{T}_m^T(\mathbf{d}^z_{m,M})+\mathcal{T}_m(\mathbf{d}^z_{m,M})\mathcal{T}_m^T(\mathbf{h}^z_{m,M})\right]\mathbf{T}_m\mathbf{h}_m\right)+O(\|\mathbf{d}_{m,M}\|^2)\end{aligned}$$

Then thanks to (2.2), as $-\mathbf{T}_m\mathbf{h}_m$ is left orthogonal to the Sylvester resultant matrix $\mathcal{T}_m(\mathbf{h}_m)$ and hence to $\mathcal{T}_m(\mathbf{h}^z_{m,M})$, $\mathbf{A}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}\mathrm{Vec}(\delta\mathbf{R}_d)$ reduces to:

$$\mathbf{A}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}\mathrm{Vec}(\delta\mathbf{R}_d)=-\mathbf{T}_m[\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural\mathcal{T}_m(\mathbf{h}^z_{m,M})\mathcal{T}_m^T(\mathbf{d}^z_{m,M})\mathbf{v}_m+O(\|\mathbf{d}_{m,M}\|^2).$$

Using the commutativity of the convolution product and the selection matrix $\mathbf{I}^\ddagger$ (3.8) that links $\mathbf{d}^z_{m,M}$ to $\mathbf{d}_{m,M}$, it holds that

$$\mathbf{B}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}=-\mathbf{T}_m[\mathcal{T}_m(\mathbf{h}_m)\mathcal{T}_m^T(\mathbf{h}_m)]^\natural\mathcal{T}_m(\mathbf{h}^z_{m,M})\mathcal{T}_M^T(\mathbf{v}_m)\mathbf{I}^\ddagger.$$

Proof of relation (4.13) In $\mathbf{B}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}\left(\mathbf{B}^{\mathrm{SS}}_{h,\mathbf{R}^m_m}\right)^T$, a priori only $\mathcal{T}_m(\mathbf{h}^z_{m,M})\mathcal{T}_M^T(\mathbf{v}_m)\mathbf{I}^\ddagger{\mathbf{I}^\ddagger}^T\mathcal{T}_M(\mathbf{v}_m)\mathcal{T}_m^T(\mathbf{h}^z_{m,M})$ depends on m_1 and m_2. First, consider $\mathcal{T}_m(\mathbf{h}^z_{m,M})\mathcal{T}_M^T(\mathbf{v}_m)$. Because any row of the Sylvester resultant matrix $\mathcal{T}_l(\mathbf{a}_u)$ can be permuted to give $\mathcal{T}_l(\mathbf{a}_u)=\mathbf{K}_{l+1,2}\begin{bmatrix}\mathcal{T}_l(\mathbf{a}_u^{(1)})\\ \mathcal{T}_l(\mathbf{a}_u^{(2)})\end{bmatrix}$ (where $\mathbf{a}_u$ and $\mathbf{a}_u^{(i)}$ are defined similarly as $\mathbf{h}_m$ and $\mathbf{h}_m^{(i)}$, $i=1,2$), it holds

$$\begin{aligned}\mathcal{T}_m(\mathbf{h}^z_{m,M})\mathcal{T}_M^T(\mathbf{v}_m) &= \mathbf{K}_{m+1,2}\begin{bmatrix}\mathcal{T}_m(\mathbf{h}^{z^{(1)}}_{m,M})\\ \mathcal{T}_m(\mathbf{h}^{z^{(2)}}_{m,M})\end{bmatrix}\left[\mathcal{T}_M^T(\mathbf{v}_m^{(1)})\mathcal{T}_M^T(\mathbf{v}_m^{(2)})\right]\mathbf{K}_{2,M+1}\\ &\overset{\mathrm{def}}{=} \mathbf{K}_{m+1,2}\begin{bmatrix}\mathbf{U}_{11} & \mathbf{U}_{12}\\ \mathbf{U}_{21} & \mathbf{U}_{22}\end{bmatrix}\mathbf{K}_{2,M+1}\end{aligned}\tag{B.1}$$

where $\mathbf{U}_{ij}\overset{\mathrm{def}}{=}\mathcal{T}_m(\mathbf{h}^{z^{(i)}}_{m,M})\mathcal{T}_M^T(\mathbf{v}_m^{(j)})$, $i,j=1,2$. Consequently, the product of $\mathcal{T}_m(\mathbf{h}^{z^{(i)}}_{m,M})$ by a column of $\mathcal{T}_M^T(\mathbf{v}_m^{(j)})$ can be interpreted, if entries of this column are regarded as input data, as the output of the channel $\mathbf{h}^{z^{(i)}}_{m,M}$ driven by this input. As columns of $\mathcal{T}_M^T(\mathbf{v}_m^{(j)})$ are shifted versions of each other, the columns of $\mathcal{T}_m(\mathbf{h}^{z^{(i)}}_{m,M})\mathcal{T}_M^T(\mathbf{v}_m^{(j)})$ are also shifted versions of each other. Therefore, each matrix $\mathbf{U}_{ij}$ has a Toeplitz structure and an entry $(\mathbf{U}_{ij})_{a,b}$, $a=1,\ldots,m+1$ and $b=1,\ldots,M+1$, is the scalar product between the

$m+M+1$-dimensional vectors $\left[\underbrace{0\,..\,0}_{a-1+m_1}\ h_0^{(i)}\cdots h_m^{(i)}\ \underbrace{0\,..\,0}_{m+1-a+m_2}\right]^T$ and $\left[\underbrace{0\,..\,0}_{b-1}\ v_0^{(j)}\cdots v_m^{(j)}\ \underbrace{0\,..\,0}_{M+1-b}\right]^T$. If $m_1=m+m_1'\geq m$, it is straightforwardly to prove that

$$\mathbf{U}_{ij}=\underbrace{\begin{bmatrix}0&.&.&0\\.&&&.\\.&&&.\\0&.&.&0\end{bmatrix}}_{\mathbf{O}_{m+1,m_1'}}\underbrace{\begin{bmatrix}u_{i,j}^1&u_{i,j}^2&.&.&u_{i,j}^{2m+1}&0&.&0\\0&u_{i,j}^1&.&.&&&.&.\\.&.&&&&&.&.\\0&0&.&0&u_{i,j}^1&.&.&.\end{bmatrix}}_{\mathbf{U}_{ij}'}\qquad\text{where } \mathbf{U}_{ij}' \text{ does not depend on } m_1.$$

Then

$$\begin{aligned}&\mathcal{T}_m(\mathbf{h}_{m,M}^z)\mathcal{T}_M^T(\mathbf{v}_m)\mathbf{I}^\ddagger\mathbf{I}^{\ddagger^T}\mathcal{T}_M(\mathbf{v}_m)\mathcal{T}_m^T(\mathbf{h}_{m,M}^z)=\\&\mathbf{K}_{m+1,2}\begin{bmatrix}\mathbf{U}_{11}&\mathbf{U}_{12}\\\mathbf{U}_{21}&\mathbf{U}_{22}\end{bmatrix}\mathbf{K}_{2,M+1}\mathbf{I}^\ddagger\mathbf{I}^{\ddagger^T}\mathbf{K}_{M+1,2}\begin{bmatrix}\mathbf{U}_{11}^T&\mathbf{U}_{21}^T\\\mathbf{U}_{12}^T&\mathbf{U}_{22}^T\end{bmatrix}\mathbf{K}_{2,m+1}\mathbf{K}_{m+1,2}\begin{bmatrix}\tilde{\mathbf{U}}_1&\tilde{\mathbf{U}}_2\\\tilde{\mathbf{U}}_3&\tilde{\mathbf{U}}_4\end{bmatrix}\mathbf{K}_{2,m+1}.\end{aligned}$$

Using the definition of $\mathbf{I}^\ddagger$, (3.8) and property (1.3), it is straightforwardly proved that

$$\mathbf{K}_{2,M+1}\mathbf{I}^\ddagger\mathbf{I}^{\ddagger^T}\mathbf{K}_{M+1,2}=\left(\mathbf{I}_2\otimes\begin{bmatrix}\mathbf{I}_{m_1'}&&&\\&\mathbf{I}_m&&\\&&\mathbf{O}_{m+1}&\\&&&\mathbf{I}_{m_2}\end{bmatrix}\right)\stackrel{\text{def}}{=}\mathbf{I}_2\otimes\mathbf{I}^+,$$

so that any of the blocks $\tilde{\mathbf{U}}_{ij}$ is given by $\mathbf{U}_{i_1i_2}\,\mathbf{I}^+\,\mathbf{U}_{j_1j_2}^T+\mathbf{U}_{k_1k_2}\,\mathbf{I}^+\,\mathbf{U}_{l_1l_2}^T$ for $i_1,i_2,j_1,j_2,k_1,k_2,l_1,l_2=1,2$. It remains to see that each term of this sum does not depend on m_1. It is proved, as

$$\mathbf{U}_{i_1i_2}\,\mathbf{I}^+\,\mathbf{U}_{j_1j_2}^T\stackrel{\text{def}}{=}[\mathbf{O}_{m+1,m_1'}\mathbf{U}_{i_1i_2}']\mathbf{I}^+\begin{bmatrix}\mathbf{O}_{m_1',m+1}\\\mathbf{U}_{j_1j_2}'^T\end{bmatrix}=\mathbf{U}_{i_1i_2}'\begin{bmatrix}\mathbf{I}_m&&\\&\mathbf{O}_{m+1}&\\&&\mathbf{I}_{m_2}\end{bmatrix}\mathbf{U}_{j_1j_2}'^T.\qquad\blacksquare$$

Proof of relation (4.12) Using (B.1) and (1.3) (which implies $\mathbf{K}_{2,M+1}\mathbf{I}^\ddagger=(\mathbf{I}_2\otimes\mathbf{I}^\dagger)\mathbf{K}_{2,M+1}$), we have

$$\mathcal{T}_m(\mathbf{h}_{m,M}^z)\mathcal{T}_M^T(\mathbf{v}_m)\mathbf{I}^\ddagger=\mathbf{K}_{m+1,2}\begin{bmatrix}\mathbf{U}_{11}\mathbf{I}^\dagger&\mathbf{U}_{12}\mathbf{I}^\dagger\\\mathbf{U}_{21}\mathbf{I}^\dagger&\mathbf{U}_{22}\mathbf{I}^\dagger\end{bmatrix}\mathbf{K}_{2,M+1}.$$

If $m_1=m+m_1'\geq m$ and $m_2=m+m_2'\geq m$, then $\mathbf{U}_{ij}\mathbf{I}^\dagger=[\,\mathbf{O}_{m+1,m_1'}\mathbf{U}_{ij}'',\mathbf{O}_{m+1,m_2'}\,]\,\mathbf{I}^\dagger=\mathbf{U}_{ij}''\mathbf{I}^{++}$ where $\mathbf{U}_{ij}''$ does not depend on m_1 and m_2 and

$$\mathbf{I}^{++}\stackrel{\text{def}}{=}\begin{bmatrix}\mathbf{O}_{m,m_1'}&\mathbf{I}_m&\mathbf{O}_m&\mathbf{O}_{m,m_2'}\\\mathbf{O}_{m+1,m_1'}&\mathbf{O}_{m+1,m}&\mathbf{O}_{m+1,m}&\mathbf{O}_{m+1,m_2'}\\\mathbf{O}_{m,m_1'}&\mathbf{O}_m&\mathbf{I}_m&\mathbf{O}_{m,m_2'}\end{bmatrix}.$$

Therefore, $\mathcal{T}_m(\mathbf{h}_{m,M}^z)\mathcal{T}_M^T(\mathbf{v}_m)\mathbf{I}^\ddagger=\mathbf{K}_{m+1,2}\begin{bmatrix}\mathbf{U}'_{11}&\mathbf{U}'_{12}\\\mathbf{U}'_{21}&\mathbf{U}'_{22}\end{bmatrix}(\mathbf{I}_2\otimes\mathbf{I}^{++})\,\mathbf{K}_{2,M+1}$. Consequently $\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{SS}}=\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{SS}^\dagger}(\mathbf{I}_2\otimes\mathbf{I}^{++})\,\mathbf{K}_{2,M+1}$ where $\mathbf{B}_{h,\mathbf{R}_m^m}^{\text{SS}^\dagger}$ do not depend on m_1 and m_2. Let $\mathbf{d}=\left[d_1^{(1)}d_1^{(2)}\ldots d_{m_1+m_2}^{(1)}d_{m_1+m_2}^{(2)}\right]^T$ be a unit norm vector. It is uniquely expressed as the sum of two orthogonal vectors $\mathbf{d}=\mathbf{d}_a+\mathbf{d}_b$ where

$$\begin{aligned}\mathbf{d}_a&\stackrel{\text{def}}{=}\left[\mathbf{0}_{2m_1'}^T\ d_{m_1'+1}^{(1)}d_{m_1'+1}^{(2)}\cdots d_{m_1'+2m}^{(1)}d_{m_1'+2m}^{(2)}\ \mathbf{0}_{2m_2'}^T\ \right]^T\\\mathbf{d}_b&\stackrel{\text{def}}{=}\left[d_1^{(1)}d_2^{(2)}\ldots d_{m_1'}^{(1)}d_{m_1'}^{(2)}\ \mathbf{0}_{4m}^T\ d_{m_1+m+1}^{(1)}d_{m_1+m+1}^{(2)}\cdots d_{m_1+m_2}^{(1)}d_{m_1+m_2}^{(2)}\right]^T\end{aligned}$$

so that $\mathbf{I}^{++}\mathbf{d}_b^{(i)} = \mathbf{0}$ and $\mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}}\mathbf{d} = \mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}^\dagger}\left(\mathbf{I}_2 \otimes \mathbf{I}^{++}\right)\mathbf{K}_{2,M+1}\mathbf{d}_a = \mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}^\dagger}\left(\mathbf{I}_2 \otimes \mathbf{I}^{++}\right)\begin{bmatrix}\mathbf{d}_a^{(1)}\\ \mathbf{d}_a^{(2)}\end{bmatrix}$ where $\mathbf{d}_a^{(i)} = \left[\mathbf{0}_{m_1'}^T \; d_{m_1'+1}^{(i)} \cdots d_{m_1'+2m}^{(i)} \; \mathbf{0}_{m_2'}^T\right]^T, i = 1, 2$. Consequently, a unit norm vector $\mathbf{d}$ that maximizes $\|\mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}}\mathbf{d}\|$ satisfies $\mathbf{d} = \mathbf{d}_a$ with $\mathbf{d}_a^{(i)} = \left[\mathbf{0}_{m_1'}^T \mathbf{d}_l^{(i)^T} \mathbf{d}_r^{(i)^T} \mathbf{0}_{m_2'}^T\right]^T, i = 1, 2$ where $\mathbf{d}_l^{(i)}$ and $\mathbf{d}_r^{(i)}$ are m-dimensional vectors. Furthermore

$$\mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}}\mathbf{d} = \mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}^\dagger}\begin{bmatrix}\mathbf{d}_l^{(1)}\\ \mathbf{0}_{m+1}\\ \mathbf{d}_r^{(1)}\\ \mathbf{d}_l^{(2)}\\ \mathbf{0}_{m+1}\\ \mathbf{d}_r^{(2)}\end{bmatrix} = \mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}^\dagger}\left(\mathbf{I}_2 \otimes \begin{bmatrix}\mathbf{I}_m & \mathbf{O}_m\\ \mathbf{O}_{m+1,m} & \mathbf{O}_{m+1,m}\\ \mathbf{O}_m & \mathbf{I}_m\end{bmatrix}\right)\begin{bmatrix}\mathbf{d}_l^{(1)}\\ \mathbf{d}_r^{(1)}\\ \mathbf{d}_l^{(2)}\\ \mathbf{d}_r^{(2)}\end{bmatrix} \stackrel{\mathrm{def}}{=} \mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}^\ddagger}\begin{bmatrix}\mathbf{d}_l^{(1)}\\ \mathbf{d}_r^{(1)}\\ \mathbf{d}_l^{(2)}\\ \mathbf{d}_r^{(2)}\end{bmatrix}$$

where $\mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}^\ddagger}$ does not depend on m_1 and m_2. Its right singular vectors and singular values also do not depend on m_1 and m_2. $\mathbf{d}$ is a right singular vector associated with the largest singular value of $\mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}}$ iff $\left[\mathbf{d}_l^{(1)^T}\mathbf{d}_r^{(1)^T}\mathbf{d}_l^{(2)^T}\mathbf{d}_r^{(2)^T}\right]$ is a right singular vector associated with the largest singular value of $\mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}^\ddagger}$. So, the right singular vector associated with the largest singular value of $\mathbf{B}_{h,\mathbf{R}_m^m}^{\mathrm{SS}}$ does not depend on m_1 and m_2. ■

References

[1] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Processing*, vol. 45, no. 12, pp. 3053-3061, Dec. 1997.

[2] J.R. Treichler, I. Fijalkow and C.R. Johnson, "Fractionally spaced equalizers. How long should they really be?," *IEEE Signal Processing Magazine*, pp. 65-81, May 1996.

[3] A. P. Liavas, P. A. Regalia and J.P. Delmas, "Robustness of least squares and subspace methods for blind channel identification/equalization with respect to effective channel undermodeling/overmodeling," *IEEE Trans. Signal Processing*, vol. 47, no. 6, pp. 1636-1645, June 1999.

[4] A. P. Liavas, P. A. Regalia and J.P. Delmas, "On the robustness of the linear prediction method for blind channel identification with respect to effective channel undermodeling/overmodeling," *to appear in IEEE Trans. Signal Processing.*

[5] A. P. Liavas, P. A. Regalia and J.P. Delmas, "Blind channel approximation: effective channel order determination," *IEEE Trans. Signal Processing*, vol. 47, no. 12, pp. 3336-3344, December 1999.

[6] A. P. Liavas, P. A. Regalia and J.P. Delmas, "Blind channel approximation: effective channel order determination," in *Proc. Asilomar Conf. Signals, Syst.,Comput.*, Pacific Grove, pp. 1153-1156, November 1998.

[7] B. M. Radich and K. M. Buckley, "The effect of source number underestimation on MUSIC location estimates," *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 233-236, Jan. 1994.

[8] G. Xu, H. Liu, L. Tong and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982-2993, Dec. 1995.

[9] 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.

[10] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multilinear prediction," in *IEEE Proc. ICASSP'94*, Adalaide, Australia, April 1994.

[11] H. Pozidis and A.P. Petropulu, "Cross-spectrum based blind channel identification" *IEEE Trans. Signal Processing*, vol. 45, no. 12, pp. 2977-2992, Dec. 1997.

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

[13] Y. Hua and H.H. Yang "On performance of cross-relation method for blind channel identification" *Signal Processing* (Elsevier), vol. 62, pp. 187-205, 1997.

[14] K. Abed-Meraim, E. Moulines and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 694-705, Mar. 1997.

[15] K. Abed-Meraim, J.F. Cardoso, A. Y. Gorokhov, P. Loubaton and E. Moulines, "On subspace methods for blind identification of single-input multiple-ouput FIR systems," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 42-55, Jan. 1997.

[16] H.V. Henderson and S. R. Searle, "The vec-permutation matrix, the vec-operator and Kronecker products: a review," *Linear and Multilinear Algebra*, vol. 9, pp. 271-288, 1981.

[17] H. Zeng and L. Tong, "Connections between the least squares and subspace approaches to blind channel estimation," *IEEE Trans. Signal Processing*, vol. 44, pp. 1593-1596, June 1996.

[18] P. J. Brockwell and R. Davis, *Times series: Theory and methods*, New York: Springer Verlag, 1990.

[19] C.R. Rao, *Linear statistical inference and its applications*, New York: Wiley, 1973.

[20] C. D. Meyer and G. W. Stewart, "Derivatives and perturbations of eigenvectors," *Siam J. Numer. Anal.*, vol. 25, no. 3, pp. 679-691, June 1988.

[21] E. J. Hannan, *Multiple times series*, New York: Wiley, 1970.

(a)

Theoretical bias components			Estimated bias components		
SS/LS	LP	OPD	SS/LS	LP	OPD
0.0006	0.0045	0.0067	0.0007	0.0142	0.0093
-0.0086	-0.0103	-0.0099	-0.0083	-0.0165	-0.0126
0.0010	0.0106	0.0119	-0.0002	0.0074	0.0122
-0.0079	-0.0294	-0.0261	-0.0088	-0.0253	-0.0257
-0.0114	-0.0030	-0.0032	-0.0121	-0.0019	-0.0066
-0.0052	-0.0038	-0.0001	-0.0048	-0.0016	0.0019
Theoretical bias norm			Estimated bias norm		
0.0172	0.0336	0.0312	0.0178	0.0343	0.0332

(b)

Theoretical bias components			Estimated bias components		
SS/LS	LP	OPD	SS/LS	LP	OPD
0.0013	0.0090	0.0134	0.0008	0.0195	0.0173
-0.0172	-0.0206	-0.0199	-0.0158	-0.0297	-0.0260
0.0020	0.0212	0.0237	-0.0018	0.0192	0.0245
-0.0157	-0.0588	-0.0522	-0.0195	-0.0549	-0.0507
-0.0229	-0.0059	-0.0065	-0.0259	-0.0039	-0.0070
-0.0104	-0.0076	-0.0002	-0.0083	-0.0080	0.0002
Theoretical bias norm			Estimated bias norm		
0.0343	0.0671	0.0625	0.0370	0.0688	0.0647

(c)

Theoretical bias components			Estimated bias components		
SS/LS	LP	OPD	SS/LS	LP	OPD
0.0006	0.0045	0.0067	0.0004	0.0074	0.0074
-0.0086	-0.0103	-0.0099	-0.0080	-0.0128	-0.0113
0.0010	0.0106	0.0119	0.0000	0.0098	0.0119
-0.0079	-0.0294	-0.0261	-0.0087	-0.0281	-0.0254
-0.0114	-0.0030	-0.0032	-0.0122	-0.0030	-0.0035
-0.0052	-0.0038	-0.0001	-0.0046	-0.0032	0.0003
Theoretical bias norm			Estimated bias norm		
0.0172	0.0336	0.0312	0.0176	0.0335	0.0314

Table 1 Theoretical asymptotic bias of $\widehat{\mathbf{h}}_2$, compared with estimated bias in the deterministic model of tails for respectively (a) $N = 300$, $\|\mathbf{d}_{2,12}\| = 0.05$; (b) $N = 300$, $\|\mathbf{d}_{2,12}\| = 0.1$; (c) $N = 1000$, $\|\mathbf{d}_{2,12}\| = 0.05$.

m	1	2	3	4
Theoretical MSE	0.0634			
Estimated MSE	0.3061	0.038	0.5189	0.8889

Table 2 Theoretical and estimated MSE of $\widehat{\mathbf{h}}_m$ given by the SS algorithm, with $N = 300$ and a signal to noise ratio of 17dB for different values of m.

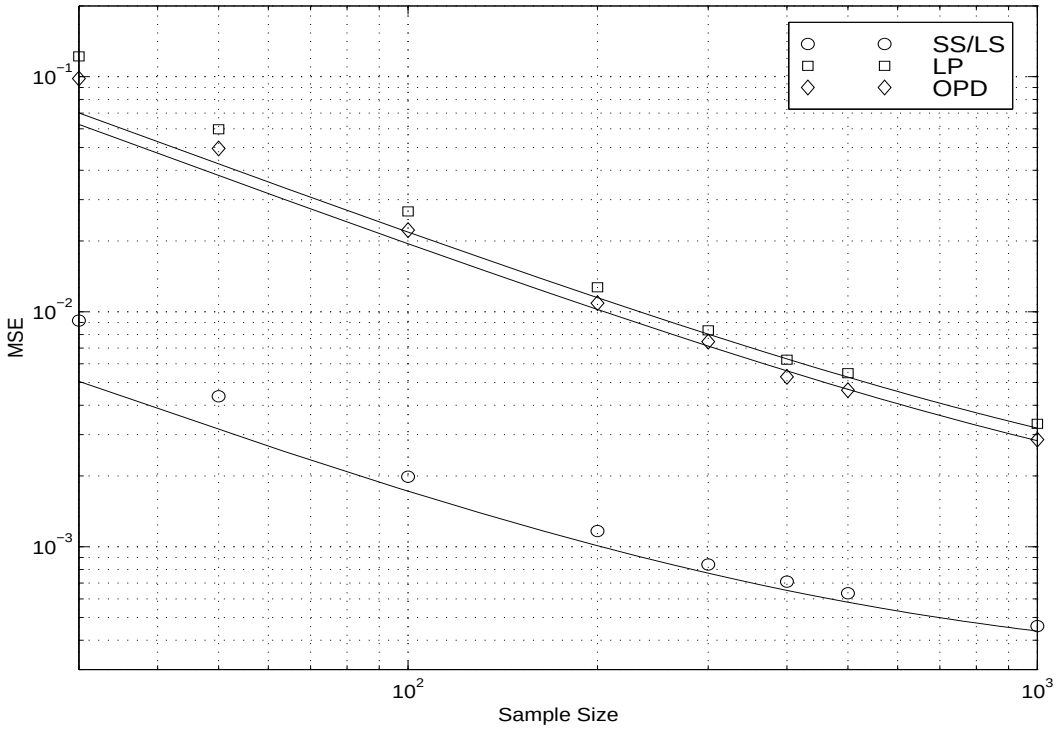


Fig.3 MSE of $\widehat{\mathbf{h}}_2$ versus the sample size for $\mathbf{h}_2$ given by (5.1), a signal noise ratio of 17dB and for $\|\mathbf{d}_{2,12}\| = 0.05$.

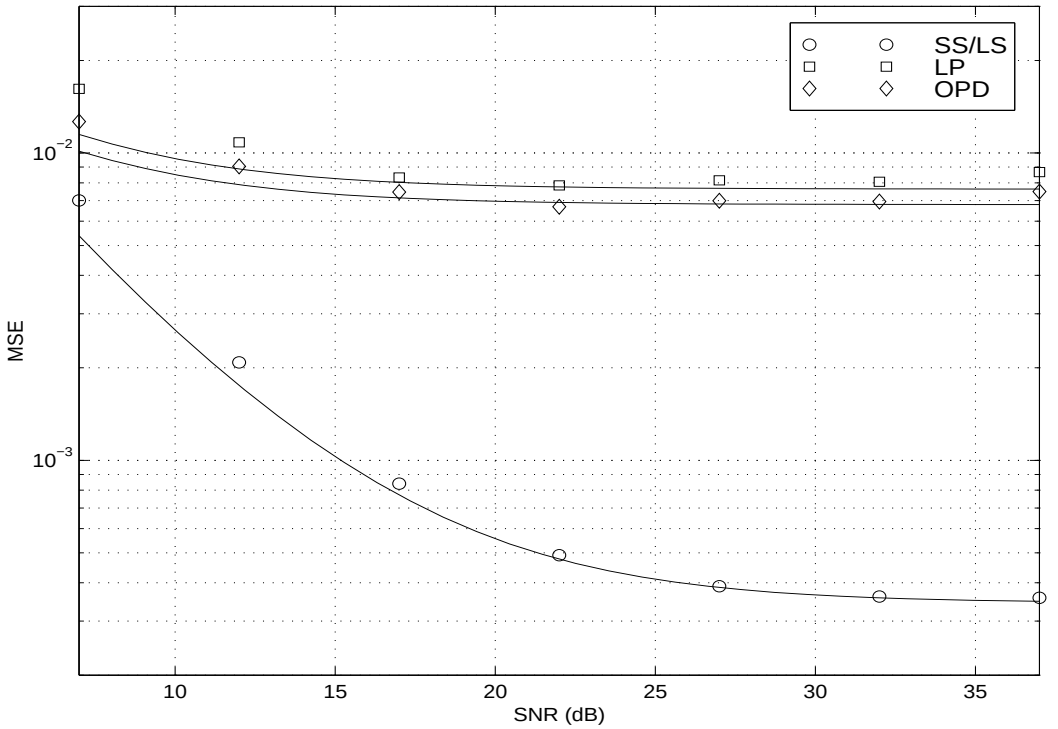


Fig.4 MSE of $\widehat{\mathbf{h}}_2$ versus the signal to noise ratio, for $N = 300$ and $\mathbf{h}_2$ given by (5.1) and for $\|\mathbf{d}_{2,12}\| = 0.05$.

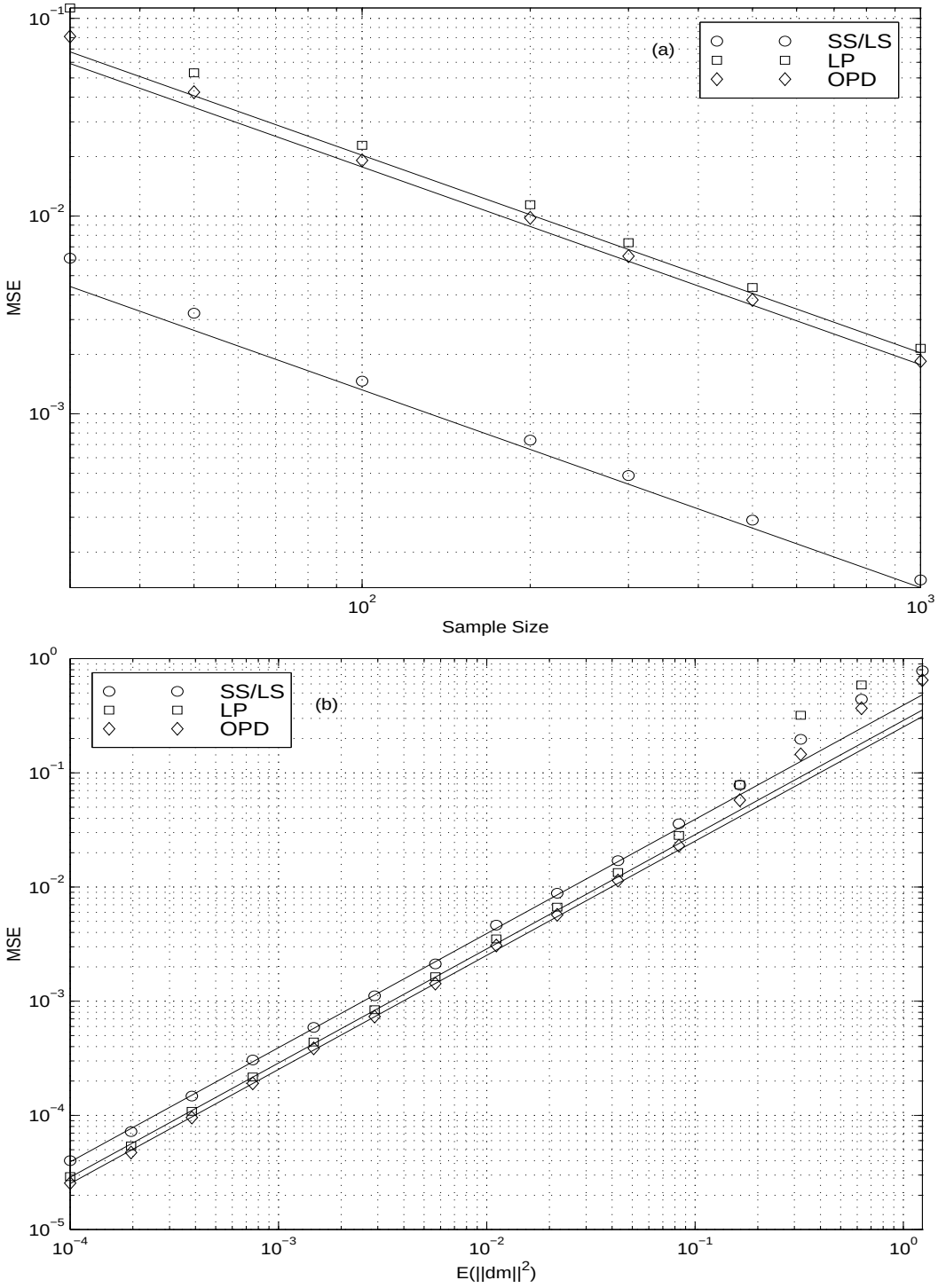


Fig.5 MSE of $\widehat{\mathbf{h}}_2$, for no tails, versus the sample size (a), and for exact statistics versus $\sigma_d^2 = \mathrm{E}(\|\mathbf{d}_{2,12}\|^2)$ (b).

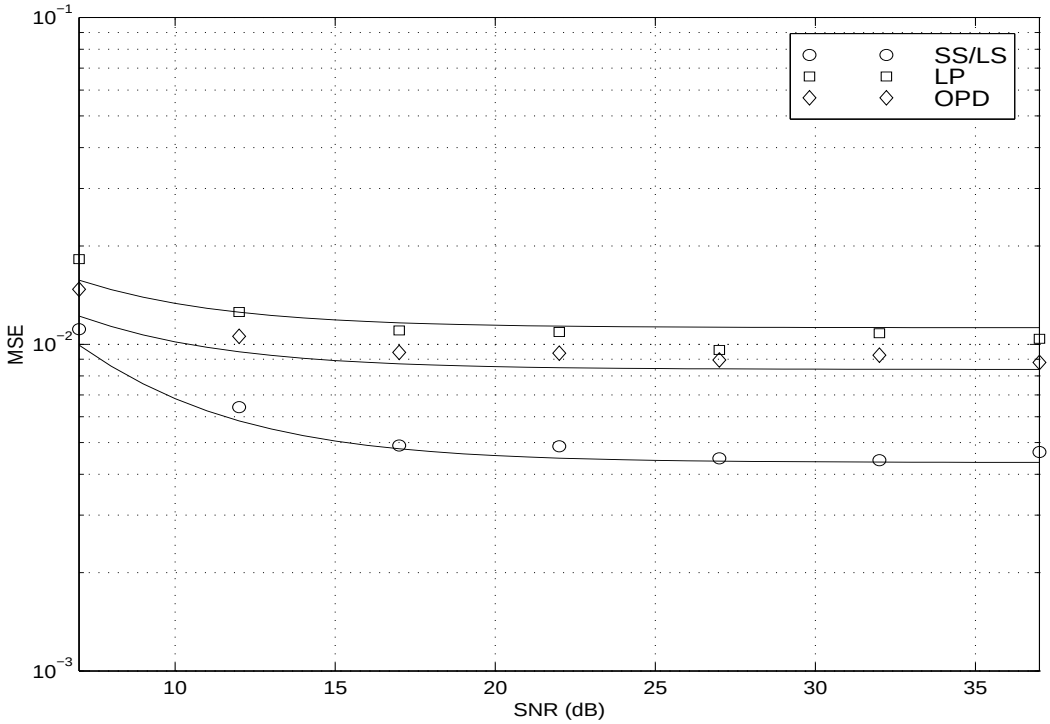


Fig.6 MSE of $\widehat{\mathbf{h}}_2$, for $N = 300$ and $\|\mathbf{d}_{2,12}\|^2 = 0,01$, versus the signal to noise ratio.

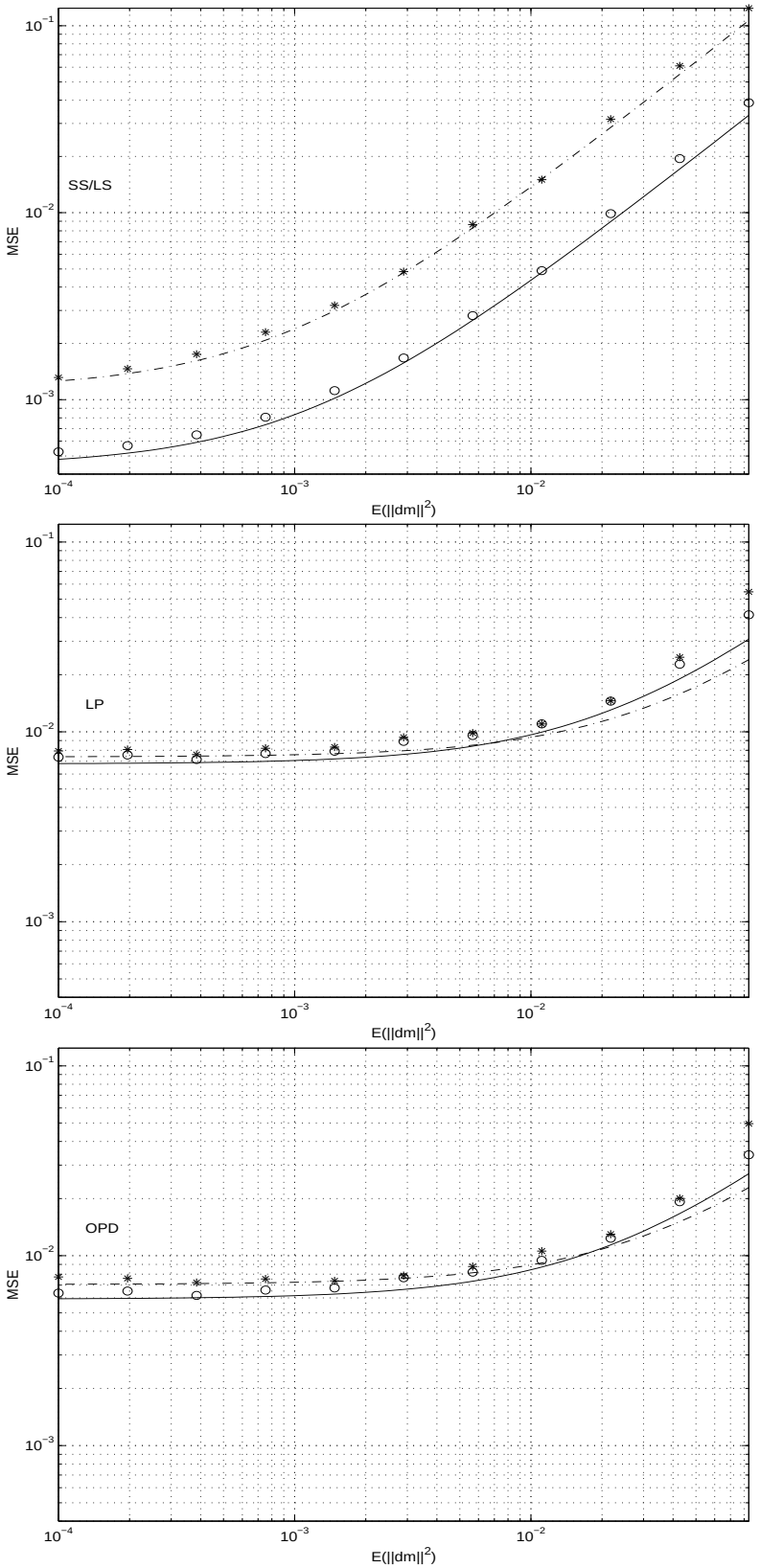


Fig.7 MSE of $\widehat{\mathbf{h}}_2$, for $N = 300$ and SNR=17dB, versus σ_d^2 for respectively $\mathbf{h}_2$ given by (5.1) (o)(—) and (5.2) (*- - -).

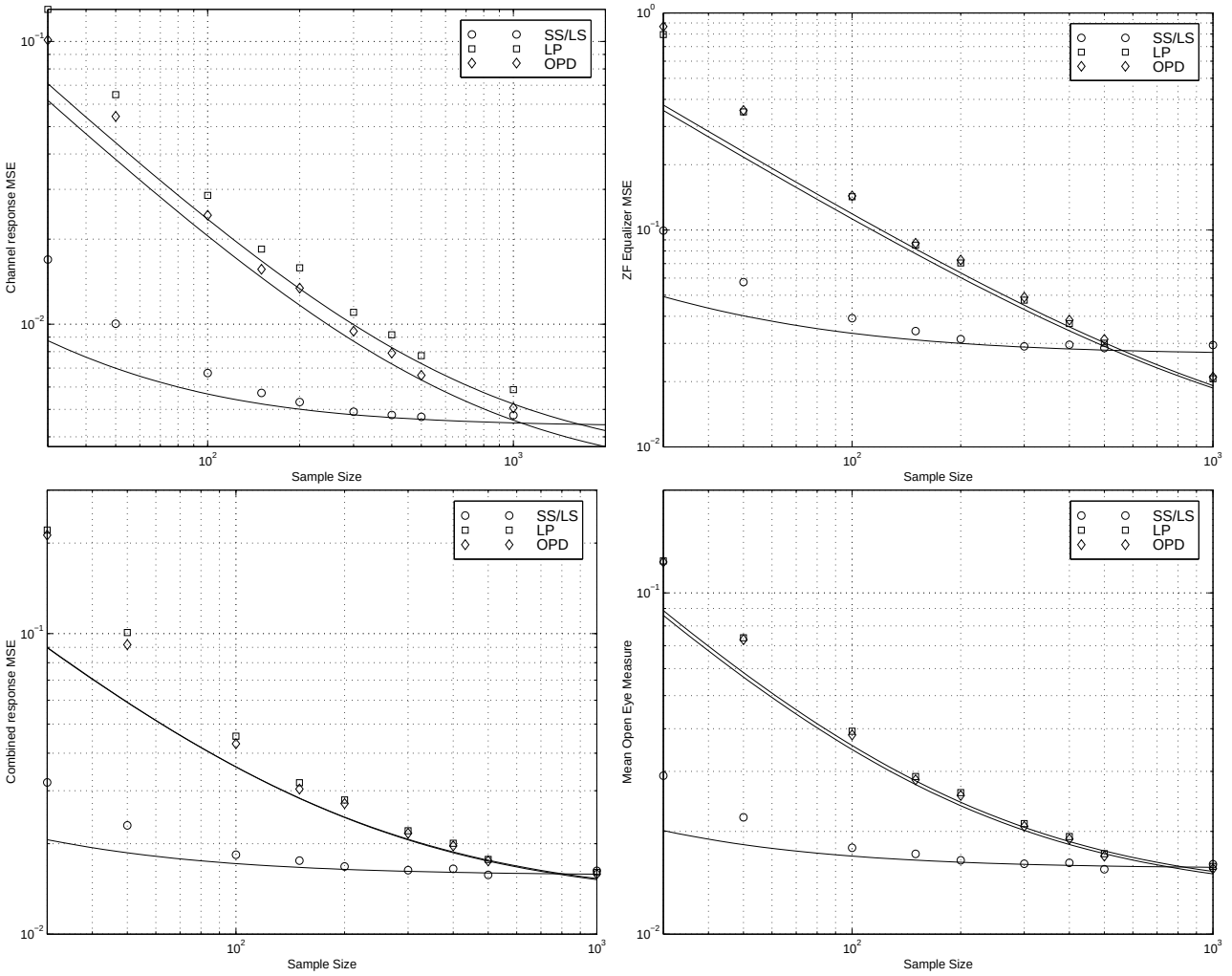


Fig.8 MSE of $\widehat{\mathbf{h}}_2$, $\widehat{\mathbf{g}}_{2,2}^{\mathrm{ZF}}$, $\widehat{\mathbf{f}}_{2,2}$ and OEM($\widehat{\mathbf{f}}_{2,2}$) for SNR=17dB, versus the sample size for $\sigma_d^2 = 0.0111$ and $\mathbf{h}_2$ given by (5.1).