Signal Processing 218 (2024) 109386

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

On the correlation between the noise and *a priori* error vectors for standard and augmented affine projection algorithms

Andrés Altieri [*,1]

Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes (UMR 8506), 91190, Gif-sur-Yvette, France
Centro de Simulación Computational, CSC-CONICET, 1425, Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Adaptive filter
a priori error
Affine projection algorithm
Variable step-size
Widely linear model
Non-circular signals

ABSTRACT

This paper analyzes the correlation matrix between the *a priori* error and measurement noise vectors for standard and augmented affine projection algorithms using a unified approach. This correlation stems from the dependence between the filter tap estimates and the noise samples, and has a strong influence on the mean square behavior of the algorithm. We show that the correlation matrix is upper triangular, and compute the diagonal elements in closed form, showing that they are independent of the input process statistics. Also, for white inputs we show that the matrix is fully diagonal. These results are valid in the transient and steady states, considering a possibly variable step-size. Our only assumption is that the filter order is large compared to its projection order and that the input signal is stationary. Using these results, we perform a steady-state analysis for small step size and provide a new simple closed-form expression for the mean-square error, which has comparable or better accuracy to many preexisting expressions, and is much simpler to compute. Finally, we also obtain expressions for the steady-state energy of the other components of the error vector.

1. Introduction and main contributions

Adaptive filters [1] have played a major role in many signal processing applications over the last few decades. The *normalized least mean-squares* (NLMS) algorithm is a widely used alternative, mainly due to its good performance, ease of implementation and low computational cost. As a downside, its rate of convergence is sensibly reduced when colored inputs are used [1]. In this context, the *affine projection algorithm* (APA) [2,3], provides an increase in convergence speed with a modest increase in computational complexity, while maintaining a robust behavior. The APA class is conceived for real or circular symmetric complex signals. However, in many applications, such as stereophonic echo cancellation [4,5], wind prediction [6], communications [7] and power grid estimation [8], among others, the signals of interest are complex and non-circular. This means that the standard APA is suboptimal for these cases. In this context, a new class of APAs called *widely linear* or *augmented* have been developed [9–12]. These algorithms are based on the widely linear model [13] and are therefore suited for both circular and non-circular signals. Among the class of widely linear algorithms the augmented APA (AAPA) [9] is one of the most representative and early members, since its mathematical structure and computational cost are similar to the standard APA. In addition, in recent years, new variants with a focus on robustness and reduced computational complexity have also emerged [14–16].

The analysis of the convergence behavior of both APA and AAPA is very involved due to the nonlinear dynamics of the update equations, which introduces a strong correlation between the magnitudes involved. In particular, the filter tap estimates are correlated with previous inputs and noise samples via the filter update equations. This induces a statistical dependence between previous noise samples and the error vector obtained from the reference signal and the estimates. This correlation is key to analyzing the algorithms but, for tractability reasons, some simplifying assumptions are usually invoked. In many cases, these assumptions include simplifications on the input signal model [17–19] or one or more independence assumptions [9,12,17,20–23] between filter tap estimates, filter inputs, noise, or functions of these magnitudes. For example, in [17], the authors perform a mean square (MS) analysis of APA considering a simplified model for the input process. In this model [24], the time-delayed input vector obtained from a stationary stochastic process is replaced by a sequence of independent random vectors which can take a finite number of orthogonal directions. In addition, they consider a strong hypothesis that past noise samples and filter coefficients are independent. In [9,20], energy

* Correspondence to: Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes (UMR 8506), 91190, Gif-sur-Yvette, France.
E-mail addresses: andres.altieri@centralesupelec.fr, andres.altieri@conicet.gov.ar.

1 This work was partially supported by the International Cooperation Project CNRS-CONICET MOSIME.

https://doi.org/10.1016/j.sigpro.2024.109386
Received 7 September 2023; Received in revised form 31 December 2023; Accepted 8 January 2024
Available online 10 January 2024

conservation arguments are used to study the MS behavior of both algorithms. Although there is not a simplified model for the input process, the independence between past noise samples and filter coefficients is maintained and other independence hypotheses are added. Later works extended [17,20] by attempting to consider the correlation between the filter coefficients and past noises. For example, [18] extends the analysis in [17] to consider the correlation between the noise and the error vector, but uses the same simplified model of the input signal.

In [22] the authors extend [20] considering the dependence on the noise and the filter tap estimates, but consider an independence assumptions between the tap estimates and a matrix obtained from input samples. In [23] the authors perform a MS deviation analysis for real signals in a system identification setup where the system inputs are white, keeping some independence assumptions from [20,22]. In [25] the authors improve on the analysis in [23] for an identification setup with white real input signals, by eliminating some of the independence assumptions of the previous work. In [19], they perform a MS analysis for real signals by developing recursive update equations for the correlation matrix of the filter tap error vector, considering its correlation with the noise. However, the analysis considers several simplifying assumptions, for example, simplifications for the input signal which are valid only asymptotically for white inputs. Finally, in [3] a MS analysis is presented using stochastic matrix theory and the assumption that certain data matrices have the so-called shuffling property.

In this paper, we provide further insight on the MS behavior of APA and AAPA by analyzing the correlation matrix between the additive noise vector $\mathbf{v}_i = [v_i, \dots, v_{i-K+1}]^T$ and the *a priori* error vector $\mathbf{e}_{a,i}$, which is the error between the signal estimate of the algorithms and the reference signal without the noise [20] (K is the (A)APA projection order). For APA we assume that the input signal is complex circular, while for the AAPA we assume it is possibly non-circular. This correlation matrix has a strong impact on the transient and steady state MS behavior of the error vector. When the tap estimates are assumed independent of the noise, this matrix is zero, which simplifies the analysis. In principle, this matrix is very complex and depends on the input signal statistics. However, our analysis shows that it is possible to compute its diagonal in closed form both in the transient and steady state of the algorithm. Furthermore, we show that this correlation depends only on the sequence of step-sizes and the noise variance, and is independent of the input process. In addition, we show that for a white input process the matrix is diagonal, and hence, we fully characterize it. For our analysis we consider the standard APA and AAPA with a variable step-size without regularization. In contrast to previous works, we do not take any assumptions on the input process (other than circularity for APA) and our only assumption is that the filter length M is large compared to the projection order K. Our simulations show that the obtained expressions are also valid in cases where this assumption does not hold. The previous works mentioned before have analyzed many aspects of the MS behavior (A)APA but have not analyzed the structure of this matrix. A related preceding work is [21], in which the authors find the trace of the correlation matrix for the case of a white input sequence for the variably regularized and fixed step-size APA. Our work is an extension because we consider any input process, we compute the diagonal of the matrix instead of the trace and give information on off-diagonal elements, and we also consider the AAPA.

Using the analysis of the correlation, we perform a new MS steady-state analysis of the behavior of the error vector for a small step-size. We provide a new simple closed-form expression for the steady-state MS error (MSE) of the algorithms which only depends on the noise variance, projection order and step-size. This formula is common to APA (with circular signals) and AAPA (with possibly non-circular signals); it predicts the same asymptotic MSE performance for both algorithms for a small asymptotic step-size, as observed in [9]. We show that, although very simple, this expression captures the behavior of the MSE compared to the well-known MSE formulas with greater complexity from [9,12,18,20,22]. Finally, using the analysis of the correlation matrix we provide a characterization of the steady-state energy of the other components of the error vector which, we believe, has not been done before.

The paper is organized as follows: in Section 2 we introduce the model and notation used. In Section 3 we present the analysis of the correlation between the noise and the *a priori* error vector and in Section 4 we perform the steady state analysis. Finally, in Section 5 we present simulation results, and in Section 6 we present some closing remarks. The proofs of the results are relegated to the Appendix to improve the flow of the manuscript.

1.1. Notation

We use boldface symbols for vectors (lower case) and matrices (upper case). $(\cdot)^T$ denotes transpose, $(\cdot)^H$ conjugate and transpose and $(\cdot)^*$ denotes conjugate. $\mathrm{diag}(\cdot)$ is the diagonal of a matrix, $\mathbb{E}$ is the expectation. For a matrix $\mathbf{R}$ we denote its (i,j) element as $[\mathbf{R}]_{i,j}$. # denotes the cardinality of a set.

2. System model and APA recursions

In this section we review the standard linear and widely linear models, which are commonly used. The former is usually considered under the hypotheses of complex circularly symmetric signals, while the latter is considered in the general case of non-circular signals. The standard APA [26] is developed for the linear model, while the AAPA [9] is considered for the widely linear model.

We review both models and show that a common representation can be used. We also show that in this common representation both algorithms can be written in a similar fashion, leading, for our purposes, to a unified treatment.

2.1. Widely linear model and AAPA recursion

We start from the widely linear model [13], in which a reference signal d_i is generated as:

$$d_i = \mathbf{u}_i^T \mathbf{h} + \mathbf{u}_i^H \mathbf{g} + v_i, \tag{1}$$

where $\mathbf{h}$ and $\mathbf{g}$ are complex vectors in $\mathbb{C}^{M\times 1}$. The input vector $\mathbf{u}_i = [u_i, \dots, u_{i-M+1}]^T$ contains samples of a non-circular complex second-order stationary process with zero mean and v_i is complex non-circular zero-mean white measurement noise of total (sum of real and imaginary) variance σ_v^2, independent of the input.

To characterize $\mathbf{u}_i$ we may separate the real and imaginary parts of $\mathbf{u}_i$ as $\mathbf{u}_i = \mathbf{u}_{R,i} + j\mathbf{u}_{I,i}$. The complex vector $\mathbf{u}_i$ is characterized by its $M \times M$ correlation and pseudocorrelation matrices:

$$\begin{aligned}
\mathbf{R}_{\mathbf{u}} &= \mathbb{E}\left[\mathbf{u}_i \mathbf{u}_i^H\right] &(2)\\
&= \mathbf{R}_{\mathbf{u}_R} + \mathbf{R}_{\mathbf{u}_I} + j\left[\mathbf{R}_{\mathbf{u}_I,\mathbf{u}_R} - \mathbf{R}_{\mathbf{u}_R,\mathbf{u}_I}\right], &(3)\\
\mathbf{P}_{\mathbf{u}} &= \mathbb{E}\left[\mathbf{u}_i \mathbf{u}_i^T\right] &(4)\\
&= \mathbf{R}_{\mathbf{u}_R} - \mathbf{R}_{\mathbf{u}_I} + j\left[\mathbf{R}_{\mathbf{u}_I,\mathbf{u}_R} + \mathbf{R}_{\mathbf{u}_R,\mathbf{u}_I}\right], &(5)
\end{aligned}$$

where $\mathbf{R}_{\mathbf{u}_R} = \mathbb{E}\left[\mathbf{u}_{R,i}\mathbf{u}_{R,i}^T\right]$, $\mathbf{R}_{\mathbf{u}_I} = \mathbb{E}\left[\mathbf{u}_{I,i}\mathbf{u}_{I,i}^T\right]$, $\mathbf{R}_{\mathbf{u}_I,\mathbf{u}_R} = \mathbb{E}\left[\mathbf{u}_{I,i}\mathbf{u}_{R,i}^T\right]$ and $\mathbf{R}_{\mathbf{u}_R,\mathbf{u}_I} = \mathbb{E}\left[\mathbf{u}_{R,i}\mathbf{u}_{I,i}^T\right]$ are the correlation and cross-correlation matrices of the real and imaginary parts. In the special case when u is circular we have $\mathbf{P}_{\mathbf{u}} = \mathbf{0}$ which implies $\mathbf{R}_{\mathbf{u}_R} = \mathbf{R}_{\mathbf{u}_I}$ and $\mathbf{R}_{\mathbf{u}_I,\mathbf{u}_R} + \mathbf{R}_{\mathbf{u}_R,\mathbf{u}_I} = \mathbf{0}$. In this case $\mathbf{R}_{\mathbf{u}}$ is sufficient to characterize the process.

The goal at time i is to obtain estimates $\mathbf{h}_i, \mathbf{g}_i$ of the unknown system vectors $\mathbf{h}, \mathbf{g}$ using the observations $\{d_n, \mathbf{u}_n\}_{n\le i}$. For the widely linear model the estimator is constructed as:

$$\hat{d}_i = \mathbf{u}_i^T \mathbf{h}_i + \mathbf{u}_i^H \mathbf{g}_i. \tag{6}$$

If we define the $M \times K$ data matrix $\mathbf{U}_i = \left[\mathbf{u}_i \ldots \mathbf{u}_{i-K+1}\right]$, and the $K \times 1$ reference data vector $\mathbf{d}_i = [d_i, \ldots, d_{i-K+1}]^T$, we can define an error vector

$$\mathbf{e}_i = \mathbf{d}_i - \mathbf{U}_i^T \mathbf{h}_i - \mathbf{U}_i^H \mathbf{g}_i, \tag{7}$$

which measures the error that $\mathbf{h}_i$ and $\mathbf{g}_i$ will produce estimating the reference signal for the time range $i, \ldots, i-K+1$. The AAPA [9] recursion can be expressed as:

$$\mathbf{h}_{i+1} = \mathbf{h}_i + \mathbf{U}_i^* \mathbf{S}_i \mathbf{e}_i, \tag{8}$$

$$\mathbf{g}_{i+1} = \mathbf{g}_i + \mathbf{U}_i \mathbf{S}_i \mathbf{e}_i, \tag{9}$$

where $\mathbf{S}_i = \mu(\mathbf{U}_i^H \mathbf{U}_i + \mathbf{U}_i^T \mathbf{U}_i^* + \beta \mathbf{I}_K)^{-1}$. The parameter $\beta > 0$ is a regularization which is included to improve the conditioning of the matrix $\mathbf{S}_i$, μ is a step size parameter which controls the convergence of the algorithm, and $\mathbf{I}_K$ is the $K \times K$ identity matrix. For tractability, in this work we consider that $\beta = 0$. The parameter $\mu > 0$ controls the speed, tracking and final error of the algorithm. A value of μ close to zero reduces the final error, while a value of μ closer to 1 improves the convergence speed and tracking of the algorithm. For this reason, a time-varying step-size μ_i can be used [12,27–29]. These time-varying step-sizes aim at accelerating convergence and/or increasing robustness by iteratively updating μ based on carefully chosen optimality constraints. In general, these constraints result in expressions in which the step-size is close to 1 at the beginning and decreases as convergence is achieved, to obtain a fast convergence and a small final error.

We also define a misalignment vectors $\tilde{\mathbf{h}}_i = \mathbf{h} - \mathbf{h}_i$ and $\tilde{\mathbf{g}}_i = \mathbf{g} - \mathbf{g}_i$, which measure the difference between the estimates and the true system vectors. With this we can define the *a priori* error vector as $\mathbf{e}_{a,i} = \mathbf{U}_i^T \tilde{\mathbf{h}}_i + \mathbf{U}_i^H \tilde{\mathbf{g}}_i$. This gives the estimation error of $\mathbf{d}_i$ without considering the additive noise component. Using the noise vector $\mathbf{v}_i = [v_i, \ldots, v_{i-K+1}]^T$, we can now write the error vector as $\mathbf{e}_i = \mathbf{e}_{a,i} + \mathbf{v}_i$.

The widely linear model and the AAPA recursion can also be written in a more compact manner. Defining the vectors $\mathbf{w} = [\mathbf{h}^T, \mathbf{g}^T]^T$ and $\mathbf{x}_i = [\mathbf{u}_i^H, \mathbf{u}_i^T]^T$ we may write the widely linear model (1) as:

$$d_i = \mathbf{x}_i^H \mathbf{w} + v_i. \tag{10}$$

Additionally, defining the estimator vector $\mathbf{w}_i = [\mathbf{h}_i^T, \mathbf{g}_i^T]^T$ we may rewrite the estimator (6) as:

$$\hat{d}_i = \mathbf{x}_i^H \mathbf{w}_i. \tag{11}$$

Finally, defining the $2M \times K$ extended data matrix $\mathbf{X}_i = \left[\mathbf{U}_i^H \ \mathbf{U}_i^T\right]^T$ the error vector (7) is written as:

$$\mathbf{e}_i = \mathbf{d}_i - \mathbf{X}_i^H \mathbf{w}_i. \tag{12}$$

The AAPA recursion given by (8) and (9) can be written compactly as:

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mathbf{X}_i \mathbf{S}_i \mathbf{e}_i, \tag{13}$$

where $\mathbf{S}_i$ is written as $\mathbf{S}_i = \mu(\mathbf{X}_i^H \mathbf{X}_i + \delta \mathbf{I}_K)^{-1}$. Finally, defining the extended misalignment vector $\tilde{\mathbf{w}}_i = \mathbf{w} - \mathbf{w}_i$ we may write the a priori error as $\mathbf{e}_{a,i} = \mathbf{X}_i^H \tilde{\mathbf{w}}_i$.

2.2. Linear model, APA and common notation

The standard linear model is normally used together with the assumption of complex circularly symmetric inputs. We can derive the linear model by considering that $\mathbf{h} = \mathbf{0}$ in the widely linear model (1). The estimator for the linear model and the expression of the error vector are then obtained by setting $\mathbf{h}_i = \mathbf{0}$ in (6) and (7), respectively. After doing this, we see that, in fact, the compact expressions for the widely linear model (10), (11) and (12) are also valid for the linear model, if we take:

$$\text{Linear Model} \begin{cases} \mathbf{w} = \mathbf{g} \\ \mathbf{w}_i = \mathbf{g}_i \\ \mathbf{x}_i = \mathbf{u}_i \\ \mathbf{X}_i = \mathbf{U}_i. \end{cases} \tag{14}$$

For the widely linear model on the other hand we used:

$$\text{Widely Linear Model} \begin{cases} \mathbf{w} = [\mathbf{h}^T, \mathbf{g}^T]^T \\ \mathbf{w}_i = [\mathbf{h}_i^T, \mathbf{g}_i^T]^T \\ \mathbf{x}_i = [\mathbf{u}_i^H \mathbf{u}_i^T]^T \\ \mathbf{X}_i = \left[\mathbf{U}_i^H \mathbf{U}_i^T\right]^T \end{cases} \tag{15}$$

Using the notation from (14), we may write the recursion of the standard APA for the linear model as [26]:

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mathbf{X}_i \mathbf{S}_i \mathbf{e}_i, \tag{16}$$

with $\mathbf{S}_i = \mu(\beta \mathbf{I}_K + \mathbf{X}_i^H \mathbf{X}_i)^{-1}$. Notice that (16) for APA is formally the same as (13) for AAPA, except that for AAPA we use (15).

2.3. Scenarios under consideration

In the previous sections we showed that the linear and widely linear models can be both characterized by (10), (11) and (13). For the widely linear model the magnitudes in these equations are defined by (15), while for the linear model, the magnitudes are given by (14). In our analysis we will consider two scenarios:

- **Scenario 1:** the signal d is generated using the widely linear model, with u and v possibly non-circular complex stationary processes. The AAPA (13) is used, with the notation given by (15).
- **Scenario 2:** the signal d is generated using the standard lineal model, with u and v circularly symmetric complex stationary processes. The standard APA (16) is used, with the notation in (14).

2.4. Structure of the matrix $\mathbf{X}_i^H \mathbf{X}_{i-m}$ for each model

In this section we briefly describe the structure of the matrices $\mathbf{X}_i^H \mathbf{X}_{i-m}$ which we will use in our analysis in the next section. For the widely linear case we have:

$$\mathbf{X}_i^H \mathbf{X}_{i-m} = \mathbf{U}_i^T \mathbf{U}_{i-m}^* + \mathbf{U}_i^H \mathbf{U}_{i-m} \tag{17}$$

$$= (\mathbf{U}_i^H \mathbf{U}_{i-m})^* + \mathbf{U}_i^H \mathbf{U}_{i-m} \tag{18}$$

$$= 2\Re(\mathbf{U}_i^H \mathbf{U}_{i-m}). \tag{19}$$

If we define $\tilde{\mathbf{u}}_i = [u_i, \ldots, u_{i-K+1}]^T$ as a vector containing K samples of u, it is straightforward to show from (19) that

$$\mathbb{E}\left[\mathbf{X}_i^H \mathbf{X}_{i-m}\right] = 2M\Re\mathbb{E}[\tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_{i-m}^H]^*, \tag{20}$$

$$= 2M\mathbb{E}\left[\tilde{\mathbf{u}}_{R,i}\tilde{\mathbf{u}}_{R,i-m}^T + \tilde{\mathbf{u}}_{I,i}\tilde{\mathbf{u}}_{I,i-m}^T\right]. \tag{21}$$

That is, the expectation of $\mathbf{X}_i^H \mathbf{X}_{i-m}$ is a real Toeplitz matrix proportional to the sum of the Kth order time-shifted correlation matrix of the real and imaginary components of u. When $m = 0$ we obtain the Kth order correlation matrix without time shifts (as defined in (3) for size M):

$$\mathbb{E}\left[\mathbf{X}_i^H \mathbf{X}_i\right] = 2M\left(\mathbf{R}_{\tilde{\mathbf{u}}_R} + \mathbf{R}_{\tilde{\mathbf{u}}_I}\right). \tag{22}$$

This matrix is both Toeplitz and symmetric.

For Scenario 2 with the linear model we have that:

$$\mathbb{E}\left[\mathbf{X}_i^H \mathbf{X}_{i-m}\right] = \mathbb{E}\left[\mathbf{U}_i^H \mathbf{U}_{i-m}\right] = M\mathbb{E}\left[\tilde{\mathbf{u}}_i \tilde{\mathbf{u}}_{i-m}^H\right]^*, \tag{23}$$

is the conjugate of the complex Toeplitz Kth order time-shifted correlation matrix of u. Again when $m = 0$ we obtain the Kth order autocorrelation matrix of u which is Hermitian and Toeplitz.

3. Analysis of the correlation between the a priori error and noise

In this section, we analyze the correlation matrix between the vectors $\mathbf{e}_{a,i}$ and $\mathbf{v}_i$ for both APA and AAPA. These vectors are correlated for $K > 1$ because the noise samples $v_{i-1}, \dots, v_{i-K+1}$ in $\mathbf{v}_i$ appear in previous updates of $\mathbf{w}_i$. We mention that $\mathbb{E}[\mathbf{e}_{a,i}\mathbf{v}_i^H]_{1,1} = 0$ because the first component of the *a priori* error vector is uncorrelated with the noise sample v_i. For $K > 1$ the remaining elements of $\mathbb{E}[\mathbf{e}_{a,i}\mathbf{v}_i^H]$ are unknown, except for the linear case and a white input signal for which the trace is known [21]. In what follows we analyze this matrix and provide some insights on its structure. We start from the recursion of the misalignment vector from (16):

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mathbf{X}_{i-1}\mathbf{S}_{i-1}\mathbf{e}_{i-1}, \tag{24}$$

where $\mathbf{S}_{i-1} = \mu_{i-1}(\mathbf{X}_{i-1}^H\mathbf{X}_{i-1})^{-1}$, now considers a variable step-size. By continuing the iteration into the past, the following recursion is obtained:

$$\tilde{\mathbf{w}}_i = \prod_{j=1}^{K}(\mathbf{I}-\mathbf{G}_j)\tilde{\mathbf{w}}_{i-K} - \sum_{j=1}^{K}\left(\prod_{k=1}^{j}(\mathbf{I}-\mathbf{G}_{k-1})\right)\mathbf{J}_j\mathbf{v}_{i-j}, \tag{25}$$

where $\mathbf{J}_j = \mathbf{X}_{i-j}\mathbf{S}_{i-j}$ and $\mathbf{G}_j = \mathbf{J}_j\mathbf{X}_{i-j}^H$, for $j > 0$, and $\mathbf{G}_0 = 0$. We can multiply (25) by $\mathbf{X}_i^H$ and $\mathbf{v}_i^H$ to obtain:

$$\begin{aligned}\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right] &= \mathbb{E}\left[\mathbf{X}_i^H\prod_{j=1}^{K}(\mathbf{I}-\mathbf{G}_j)\tilde{\mathbf{w}}_{i-K}\mathbf{v}_i^H\right] \\ &\quad -\mathbb{E}\left[\mathbf{X}_i^H\sum_{j=1}^{K}\left(\prod_{k=1}^{j}(\mathbf{I}-\mathbf{G}_{k-1})\right)\mathbf{J}_j\mathbf{v}_{i-j}\mathbf{v}_i^H\right].\end{aligned} \tag{26}$$

The first term on the right side vanishes because the noise vector $\mathbf{v}_i$ has zero mean and is independent of the rest. For $K = 1$ the second expectation is also zero as mentioned before. For $K > 1$ we then have:

$$\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right] = -\sum_{j=1}^{K}\mathbb{E}\left[\mathbf{X}_i^H\left(\prod_{k=1}^{j}(\mathbf{I}-\mathbf{G}_{k-1})\right)\mathbf{J}_j\right]\mathbb{E}\left[\mathbf{v}_{i-j}\mathbf{v}_i^H\right]. \tag{27}$$

We then write $\mathbb{E}\left[\mathbf{v}_{i-j}\mathbf{v}_i^H\right] = \sigma_v^2\tilde{\mathbf{I}}_{K,j}$ where $\tilde{\mathbf{I}}_{K,m} \in \mathbb{C}^{K\times K}$, is a matrix such that ($m \in \mathbb{Z}$):

$$\left[\tilde{\mathbf{I}}_{K,m}\right]_{q,p} = \begin{cases} 1 & \text{if } p-q = m \\ 0 & \text{otherwise.}\end{cases} \tag{28}$$

We notice that $\tilde{\mathbf{I}}_{K,K} = \mathbf{0}$, which means that the Kth term in the second expectation in (27) vanishes. Then we have:

$$\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right] = -\sigma_v^2 E\left[\mathbf{X}_i^H\mathbf{J}_1\right]\tilde{\mathbf{I}}_{2,1} \quad (K=2), \tag{29}$$

and for $K > 2$:

$$\begin{aligned}\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right] &= -\sigma_v^2\Bigg(\mathbb{E}\left[\mathbf{X}_i^H\mathbf{J}_1\right]\tilde{\mathbf{I}}_{K,1} + \\ &\quad + \sum_{j=2}^{K-1}\mathbb{E}\left[\mathbf{X}_i^H\left(\prod_{k=1}^{j-1}(\mathbf{I}-\mathbf{G}_k)\right)\mathbf{J}_j\right]\tilde{\mathbf{I}}_{K,j}\Bigg).\end{aligned} \tag{30}$$

The expression for $K > 2$ is involved because of matrix product $\prod_k(\mathbf{I}-\mathbf{G}_k)$ which is non-commutative.

We now introduce an approximation for products of matrices of the form $\mathbf{X}_i^H\mathbf{X}_{i-j}$ which appear in (30), under the assumption that $M \gg K$. These approximations are extensions of approximations that have been used for APA before [21,23,25], which we now also consider for AAPA:

- *Scenario 1, widely linear model with non-circular inputs:* we start by analyzing the structure of the matrix $\mathbf{X}_i^H\mathbf{X}_i$ and showing that when $M \gg K$ it is reasonable to take the approximation $\mathbf{X}_i^H\mathbf{X}_i \approx 2M\left(\mathbf{R}_{\tilde{\mathbf{u}}_R} + \mathbf{R}_{\tilde{\mathbf{u}}_I}\right)$. To see this, we consider the vector $\tilde{\mathbf{u}}_i = [u_i, \dots, u_{i-K+1}]^T$ which was defined in Section 2.4. Using this vector the matrix $\mathbf{U}_i$ can be written as $\mathbf{U}_i = \left[\tilde{\mathbf{u}}_i, \dots, \tilde{\mathbf{u}}_{i-M+1}\right]$ and we can write:

$$\frac{1}{M}\mathbf{X}_i^H\mathbf{X}_i = \frac{1}{M}\left(\mathbf{U}_i^T\mathbf{U}_i^* + \mathbf{U}_i^H\mathbf{U}_i\right) \tag{31}$$

$$= \frac{1}{M}\sum_{j=0}^{M-1}\tilde{\mathbf{u}}_{i-j}\tilde{\mathbf{u}}_{i-j}^H + \tilde{\mathbf{u}}_{i-j}^*\tilde{\mathbf{u}}_{i-j}^T. \tag{32}$$

  Then assuming that the process is stationary ergodic when $M \gg K$ we can approximate the time average on the right side of (32) by the expectation of one of the terms in the summation:

$$\begin{aligned}\frac{1}{M}\mathbf{X}_i^H\mathbf{X}_i &\approx \mathbb{E}\left[\tilde{\mathbf{u}}_{i-j}\tilde{\mathbf{u}}_{i-j}^H + \tilde{\mathbf{u}}_{i-j}^*\tilde{\mathbf{u}}_{i-j}^T\right] &(33)\\ &\approx 2\left(\mathbf{R}_{\tilde{\mathbf{u}}_R} + \mathbf{R}_{\tilde{\mathbf{u}}_I}\right), &(34)\end{aligned}$$

  which justifies the proposed approximation. In an analogous manner and using the same justification, we can introduce the approximation of the time shifted matrices:

$$\mathbf{X}_i^H\mathbf{X}_{i-j} \approx 2M\mathbb{E}\left[\tilde{\mathbf{u}}_{R,i}\tilde{\mathbf{u}}_{R,i-m}^T + \tilde{\mathbf{u}}_{I,i}\tilde{\mathbf{u}}_{I,i-m}^T\right], \tag{35}$$

  which is the sum of the time-shifted correlation matrices of the real and imaginary components of the input process, independent of i due to the stationarity condition. By comparing (35) with (21) we observe that it is equivalent to approximation $\mathbf{X}_i^H\mathbf{X}_{i-j}$ by its expectation.
- *Scenario 2, linear model with circular inputs:* using the same assumptions as in the previous case, we can also introduce approximations when $M \gg K$. In this case the approximation is $\mathbf{X}_i^H\mathbf{X}_i \approx M\mathbb{E}\left[\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_i^H\right]^*$, the correlation matrix of the circular input process. In the same way we obtain:

$$\mathbf{X}_i^H\mathbf{X}_{i-j} \approx M\mathbb{E}\left[\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_{i-m}^H\right]^*, \tag{36}$$

  which is a time shifted correlation matrix of the circular input process, also independent of i due to stationarity. As mentioned before, for APA these approximations have already been used for example in [21,23,25].

Using (30) as starting point and considering the approximations (35) and (36) we introduce our main result for this section:

Theorem 1. *When $M \gg K$, under the hypotheses of Scenarios 1 and 2, and assuming that (35) and (36) are valid approximations for Scenarios 1 and 2, respectively, then $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ is an upper triangular matrix, and its diagonal elements for $2 \le q \le K$ are:*

$$\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]_{q,q} = -\sigma_v^2\left(\mu_{i-1} + \sum_{j=2}^{q-1}\mu_{i-j}\prod_{k=1}^{j-1}(1-\mu_{i-k})\right). \tag{37}$$

Furthermore, if the input process is white then $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ is a diagonal matrix, that is, its only non-zero elements are given by (37).

Proof. It is presented in Appendix. □

From Theorem 1 we outline the following conclusions:

- The elements below the main diagonal of $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ are zero.
- The diagonal of $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ can be computed in closed form it depends only on the noise variance σ_v^2 and the sequence of step sizes $\{\mu_i\}$, that is, it does not depend on the statistics of the input process u. This is true in the transient and steady-state of the algorithms.
- In general the elements above the main diagonal will depend on the statistics of the input process u except for the white input case, when they are zero.

Considering a fixed step-size we have the following corollary which will be used in the following section:

Corollary 1. *Considering a constant step-size $\mu_i \equiv \mu$ in (37) we find that for $1 \le q \le K$:*

$$\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]_{q,q} = -\sigma_v^2\left[1-(1-\mu)^{q-1}\right]. \tag{38}$$

Notice that, in contrast to (37), this is also valid for $q = 1$. Summing the elements of the diagonal we find that:

$$\mathbb{E}\left[\mathbf{e}_{a,i}^H \mathbf{v}_i\right] = -\sigma_v^2 \left[K - \frac{1-(1-\mu)^K}{\mu}\right]. \tag{39}$$

In the following section we use these results to derive a new steady-state analysis of the algorithms.

4. A steady-state error analysis without the independence assumption

In this section we perform a small step-size analysis of the steady state behavior of the algorithms. We derive a new formula for the steady state value of the MSE and we also find the steady-state energy of the other components of the error vector $\mathbf{e}_i$. This is done under the conditions of Theorem 1 ($M \gg K$) and under a weak approximation on the energy of the error vector $\mathbf{e}_{a,i}$ in the steady state.

We start from the recursion of the error vector (24) and multiply by $\mathbf{X}_i^H$ to obtain:

$$\mathbf{X}_i^H \tilde{\mathbf{w}}_{i+1} = \mathbf{X}_i^H \tilde{\mathbf{w}}_i - \mu_i \mathbf{e}_i. \tag{40}$$

By taking the squared norm and the expectation we find that:

$$\mathbb{E}\left[\|\mathbf{X}_i^H \tilde{\mathbf{w}}_{i+1}\|^2\right] = \mathbb{E}\left[\|\mathbf{X}_i^H \tilde{\mathbf{w}}_i\|^2\right] + \mu_i^2 \mathbb{E}\left[\|\mathbf{e}_i\|^2\right] - \mu_i \mathbb{E}\left[\mathbf{e}_i^H \mathbf{e}_{a,i} + \mathbf{e}_{a,i}^H \mathbf{e}_i\right]. \tag{41}$$

As $i \to \infty$, in the steady state condition, the step-size is assumed to converge to a small steady-state value $\mu \ll 1$. Alternatively, we can assume that the step-size is fixed and small for all time-instants ($\mu_i \equiv \mu \ll 1$). As mentioned before, in general, a time varying step-size is preferred to benefit of the trade-off between high initial convergence speed (μ close to 1 at the initial iterations) and smaller final error (μ reduced as convergence is achieved) that we mention in Section 2.1. Practical step-size expressions have this behavior (see [12,29] for example), and, for this reason, it is reasonable to assume that under normal operating conditions, asymptotically the step size will have converged to a small value in a stationary environment. The exact final value achieved will depend on multiple factors such as the expression of the specific step-size and the color of the input signal, among others.

When the filter has converged and the step-size is small, the updates become small enough such that on average there is almost no change in the energy of the *a priori* error vector when filtering the same data before and after an update. This motivates the following approximation:

$$\mathbb{E}\left[\|\mathbf{X}_i^H \tilde{\mathbf{w}}_{i+1}\|^2\right] \approx \mathbb{E}\left[\|\mathbf{X}_i^H \tilde{\mathbf{w}}_i\|^2\right] \quad (\text{as } i \to \infty). \tag{42}$$

This means that, on average, there will be almost no reduction in the error after an update in the steady-state small-step size condition. Using this approximation in (41) we may obtain the following steady-state equation:

$$\mu \mathbb{E}\left[\|\mathbf{e}\|^2\right]_\infty = \mathbb{E}\left[\mathbf{e}^H \mathbf{e}_a\right]_\infty + \mathbb{E}\left[\mathbf{e}_a^H \mathbf{e}\right]_\infty, \tag{43}$$

where the subscript $(\cdot)_\infty$ denotes the steady-state values of the expectations as $i \to \infty$. This equation is valid in principle for any input process, provided that $\mu \ll 1$. We would like to compare (43) to the general steady-state equations that were presented in [20] for Scenario 2:

$$\mu \mathbb{E}\left[\mathbf{e}_i^H (\mathbf{X}_i^H \mathbf{X}_i)^{-1} \mathbf{e}_i\right] = \mathbb{E}\left[\mathbf{e}_{a,i}^H (\mathbf{X}_i^H \mathbf{X}_i)^{-1} \mathbf{e}_i\right] + \mathbb{E}\left[\mathbf{e}_i^H (\mathbf{X}_i^H \mathbf{X}_i)^{-1} \mathbf{e}_{a,i}\right] \quad (i \to \infty). \tag{44}$$

This equation is valid without any approximation whatsoever. It is interesting to notice that if in (44) we consider a white input and for $M \gg K$ we approximate $\mathbf{X}_i^H \mathbf{X}_i \approx M \sigma_x^2 \mathbf{I}$ we obtain (43). This means that (43) will be exact for a white input process for any value of μ, but our analysis shows it will be a good approximation for any other input process when $\mu \ll 1$, that is, all input processes will behave as white processes in the steady-state, provided that μ is small enough.

We can now use the results of Theorem 1 to characterize the steady-state value of the error vector without using an independence assumption. Using that $\mathbf{e}_i = \mathbf{e}_{a,i} + \mathbf{v}_i$ in (43) we can derive an expression for the energy of the error vector:

$$\mathbb{E}\left[\|\mathbf{e}\|^2\right]_\infty = \frac{2}{2-\mu} \left(\mathbb{E}\left[\|\mathbf{v}\|^2\right]_\infty + \mathbb{E}\left[\mathbf{e}_a^H \mathbf{v}\right]_\infty + \mathbb{E}\left[\mathbf{v}^H \mathbf{e}_a\right]_\infty\right). \tag{45}$$

Using (39) in (45), the steady-state value of the energy of the error vector evaluates to:

$$\mathbb{E}\left[\|\mathbf{e}\|^2\right]_\infty = \frac{2\sigma_v^2}{\mu(2-\mu)} \left(1-(1-\mu)^K\right). \tag{46}$$

Considering the discussion on the validity of (43), we can see that this expression will be valid for a white input for all values of μ and for general inputs for small values of μ. Also, for the energy of the a priori error vector we can find a closed form equation:

$$\mathbb{E}\left[\|\mathbf{e}_a\|^2\right]_\infty = \frac{\mu}{2-\mu} \mathbb{E}\left[\|\mathbf{v}\|^2\right]_\infty + \frac{\mu-1}{2-\mu} \left(\mathbb{E}\left[\mathbf{e}_a^H \mathbf{v}\right]_\infty + \mathbb{E}\left[\mathbf{v}^H \mathbf{e}_a\right]_\infty\right), \tag{47}$$

which can be evaluated using (39). In order to find the steady-state energy of the components of the error vector we introduce a simple assumption:

- (A1) In the steady-state, with a small step-size, the components of the a priori error vector have the same energy, that is, for $1 \le q \le K$:

$$\mathbb{E}\left[\mathbf{e}_a \mathbf{e}_a^H\right]_{q,q,\infty} \approx \frac{\mathbb{E}\left[\|\mathbf{e}_a\|^2\right]_\infty}{K}. \tag{48}$$

Assumption (A1) is not new, it is known in the literature [9,20]. In fact, this assumption is a much weaker version of the standard assumption [20] for the small step-size analysis, which assumes that $\mathbb{E}\left[\mathbf{e}_{a,i} \mathbf{e}_{a,i}^H\right]$ is a multiple of the identity matrix. Under this approximation we can write the excess mean square error (EMSE) as:

$$\mathbb{E}\left[\|\mathbf{e}_a\|^2\right]_\infty \approx K \mathbb{E}\left[|e_a|^2\right]_\infty \tag{49}$$

where $e_{a,i} = d_i - \mathbf{x}_i^H \tilde{\mathbf{w}}_i$ is the first component of $\mathbf{e}_{a,i}$. Now replacing (49) in (47) and noting that $\mathbb{E}\left[\mathbf{v}_i^H \mathbf{e}_{a,i}\right]$ is a real magnitude we can write the excess mean square error (EMSE) in the steady-state as:

$$\mathbb{E}\left[|e_a|^2\right]_\infty = \frac{\mu \sigma_v^2}{2-\mu} + \frac{2(\mu-1)}{K(2-\mu)} \mathbb{E}\left[\mathbf{e}_a^H \mathbf{v}\right]_\infty. \tag{50}$$

The first term in (50) is the formula for the EMSE for small μ presented in [9] for Scenario 1 and [20] for Scenario 2 using the independence assumption and others. This expression does not consider the important dependence of the EMSE with K. The second term is a new additional term which depends on K and under (49) allows us to correct the EMSE using the correlation between the noise and the *a priori* error vector. Finally, using (39) we can obtain a new and simple expression for the steady state EMSE as:

$$\mathbb{E}\left[|e_a|^2\right]_\infty = \sigma_v^2 \left[1 - \frac{2(1-\mu)\left(1-(1-\mu)^K\right)}{K\mu(2-\mu)}\right]. \tag{51}$$

Finally, the energy of the $1 \le q \le K$ components of the error vector is given by:

$$\mathbb{E}\left[\mathbf{e}_i \mathbf{e}_i^H\right]_{q,q} = \mathbb{E}\left[\mathbf{e}_{a,i} \mathbf{e}_{a,i}^H\right]_{q,q} + \mathbb{E}\left[\mathbf{e}_{a,i} \mathbf{v}_i^H\right]_{q,q} + \mathbb{E}\left[\mathbf{v}_i \mathbf{e}_{a,i}^H\right]_{q,q} + \sigma_v^2. \tag{52}$$

Under approximation (A1) all the components of the *a priori* error vector have energy given by (51). Then, using (51) and (38) we can obtain the energy of each component of the error vector as ($1 \le q \le K$):

$$\mathbb{E}\left[\mathbf{e}\mathbf{e}^H\right]_{q,q,\infty} = \max\left\{0, 2\sigma_v^2 \left[(1-\mu)^{q-1} - \frac{(1-\mu)\left(1-(1-\mu)^K\right)}{K(2-\mu)\mu}\right]\right\}. \tag{53}$$

The max is included because it can be seen that the expression will become negative as $\mu \to 1$ for $q > 1$. By setting $q = 1$ in this expression we obtain the steady state mean square error, which may be also obtained by adding σ_v^2 to (51):

$$\mathrm{MSE}_\infty = 2\sigma_v^2 \left[1 - \frac{(1-\mu)\left(1-(1-\mu)^K\right)}{K(2-\mu)\mu} \right]. \tag{54}$$

The analysis in this section shows that AAPA in Scenario 1 and APA in Scenario 2 achieve the same steady state MSE for a small asymptotic value of the step-size. In the case in which APA is used with the widely linear model, the MSE is larger than when using AAPA [9].

4.1. Comparison of the MSE formula with previous expressions

It is interesting to compare the new proposed expression for the steady-state MSE (54) with previous known formulas both for APA and AAPA. We focus on expressions that are proposed for arbitrary values of μ or for small values, excluding some formulas which exist that were proposed for large μ.

The simplest and well-known expression valid for small μ for APA with circular signals [20] is:

$$\mathrm{MSE} = \frac{2\sigma_v^2}{2-\mu}. \tag{55}$$

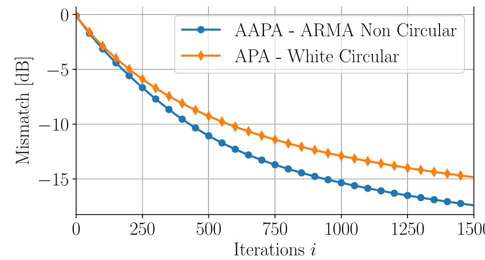
The complexity of this formula is comparable to (54) since it does not depend on the statistical properties of the input signal. However, in contrast to (54) it does not depend on the projection order K, and hence, some key behaviors are lost. In [18], another expression for steady-state MSE of APA was introduced, as a generalization of the one in [17]:

$$\mathrm{MSE} = \sigma_v^2 + \frac{\mu\sigma_v^2}{2-\mu}\mathbb{E}\left[\frac{1}{\|\mathbf{x}_i\|^2}\right]\mathrm{tr}(\mathbf{R}_x)(1+2\Gamma), \tag{56}$$

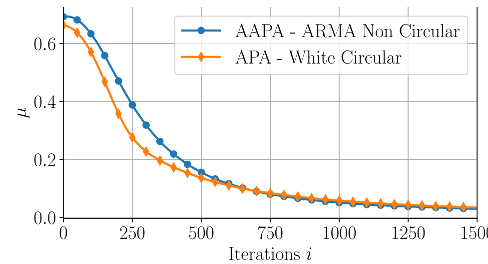
$$\Gamma = \sum_{i=1}^{M} p_i \sum_{q=1}^{K-1} \left[\left(1 - \frac{1-(1-p_i)^q}{1-(1-p_i)^K} \right) (1-\mu)^q \right], \tag{57}$$

$p_i = \lambda_i/\mathrm{tr}(\mathbf{R}_x)$, and λ_i are the eigenvalues of $\mathbf{R}_x$. This expression requires knowledge of the eigenvalues of the autocorrelation matrix of the input process and the expectation of the inverse of the norm of the input vector. A different expression, valid for any value of μ is (31) from [22], which is a generalization from [20]. This formula depends on the expectation of several matrices which are constructed from Kronecker products of matrices of size $M \times M$. Therefore, the final expression involves matrices of size $M^2 \times M^2$, which depend on the input process, and which need to be computed through Monte Carlo simulations. This makes the formula increasingly difficult to compute as the order of the filter M increases; in fact, for large values of M we were not able to compute it on a standard desktop computer. Comparing these formulas with (54), we see that (54) is simpler and requires less information to be computed, since it does not require specific knowledge of the characteristics of the input process or extensive Monte Carlo simulations. Furthermore in the Numerical Results section, we show that in the regime of small μ and moderately long filters, (54) provides comparable or sometimes better performance than the aforementioned alternatives.

For the case of AAPA it is shown in [9] that (55) is still valid, with the same mentioned limitations. A more exact formula for AAPA is (47) from [12]. This formula is presented in the analysis of the steady-state performance of VSS-WLCAPA but the hypotheses and approximations involved can be used for the analysis of other step-size expressions or for a constant step-size and Gaussian noise. Hence (47) from [12] can also be compared to (54). The expression from [12] is constructed from Kronecker products of expectation of certain matrices which depend on the input process. As a result, the computation of (47) from [12] involves Monte Carlo estimations of matrices of size $4M^2 \times 4M^2$ which depend on the input process, and these become increasingly difficult to compute as M increases.



(a) Mismatch obtained by using the average step-size (b), averaging over $J = 16000$ runs of APA and AAPA.



(b) Average step-size from [12] obtained by averaging 3200 independent runs of APA and AAPA.

Fig. 1. Example of mismatch for APA and AAPA using the averaged time step to test Theorem 1. The ARMA process is given by (58). $M = 64$, $K = 4$, SNR = 20 dB.

5. Numerical results

In this section we perform simulations to explore the validity of the proposed expressions. For both APA and AAPA we assume that the system vector $\mathbf{w}$ is drawn as a realization of a circular complex white Gaussian vector with unit variance taps, of length M and $2M$, respectively. For AAPA the inputs are chosen to correspond to Scenario 1 with some circular and non-circular inputs, while for APA they correspond to Scenario 2, with circular inputs.

The input process u_i for APA is assumed to be a white Gaussian circular process or a circular first order Gaussian autoregressive process (AR1) with a pole at 0.95. The input process u_i for AAPA is a complex white non-circular Gaussian process with unit variance for the real part and 1/2 variance for the imaginary part, or a circular first order Gaussian autoregressive process (AR1) with a pole at 0.95 (the same as for APA), or the complex non-circular autoregressive moving average (ARMA) from [9] with recursion:

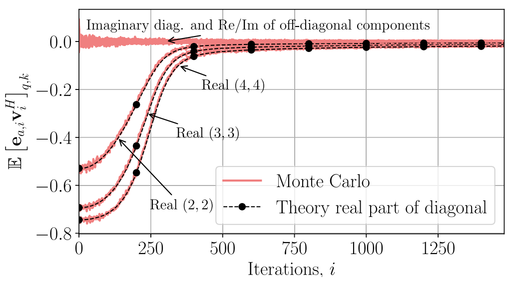
$$u_i = 0.5u_{i-1} + 2n_i + \frac{1}{2}n_i^* + n_{i-1} + 0.5n_{i-1}^*, \tag{58}$$

where n_i is a complex circular white Gaussian noise.

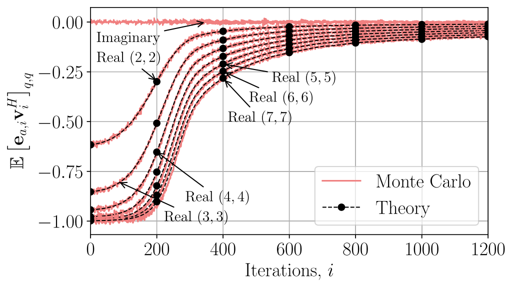
The additive noise v_i is complex circular white Gaussian noise, with variance σ_v^2, which is adjusted in each simulation to obtain the prescribed signal-to-noise ratio (SNR). We denote by J the number of averaged runs to obtain each plot. Finally, both algorithms are initialized with a zero initial condition.

5.1. Correlation between the a priori error and the noise

We first explore the validity of the results from Theorem 1 about the $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$. These results are valid for the transient and steady-state conditions of the algorithms, and for any deterministic time varying step-size sequence. It important to discuss how these results are to be validated. In general, variable step-size formulas are designed as deterministic sequences which depend on average values of different magnitudes of the algorithm. For example, the step size from [12] involves the magnitudes $\mathbb{E}\left[\|\mathbf{e}_i\|^2\right]$ and $\mathbb{E}\left[\|\mathbf{e}_{a,i}\|^2\right]$. If we knew these magnitudes precisely the sequence of step-sizes would be deterministic

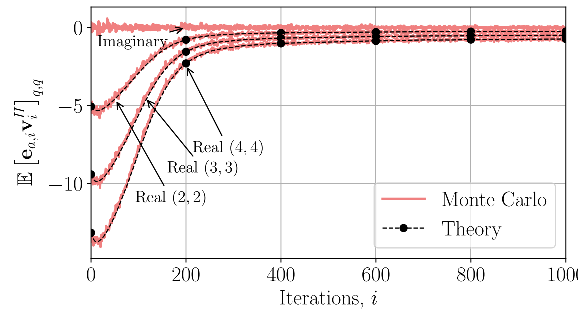


(a) White input, $K = 4$, $M = 64$. The matrix is diagonal, with real elements.

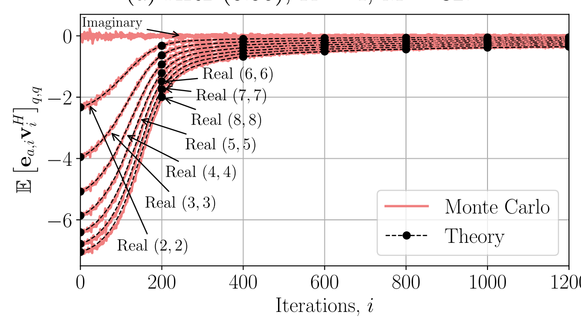


(b) AR1(0.95) input, $K = 8$, $M = 128$.

Fig. 2. Comparison of Monte Carlo simulations of $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ with the results of Theorem 1 for APA. SNR = 20 dB. $J = 16000$.



(a) AR1 (0.95), $K = 4$, $M = 32$.



(b) White input, $K = 8$, $M = 64$.

Fig. 3. APA. Comparison of Monte Carlo simulations of $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ compared to (37) and other results of Theorem 1 for different input processes. SNR = 10 dB. $J = 16000$.

and the results from Theorem 1 would apply directly. In practice these magnitudes are estimated using time-running averages of the instantaneous error magnitudes of the algorithm, which implies that the step size sequence is random and for each run there is a different realization. For this reason, to be theoretically consistent we cannot estimate $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ by averaging independent runs of the algorithm, because we would be considering different step-size sequences in each run. Given this observation, in order to test Theorem 1 we must consider a deterministic step-size sequence. In order to obtain one with a realistic behavior, we do as follows: for each simulation setup we perform 3200 runs of the algorithm (APA or AAPA) using the variable step size from [12] with configuration parameters $\alpha = 0.975$ and $\theta = 3$. The step-size sequences of the 3200 runs are then averaged to define a deterministic time sequence for each setup. Then $J = 16000$ runs of the algorithms are computed using this deterministic time sequence, which follows realistically the behavior of the step-size for each setup of each algorithm. It is worth mentioning that the step-size in [12] was designed with AAPA in mind but nevertheless it has a reasonable behavior and can be used for APA as means to prove the validity of Theorem 1. In Fig. 1 we can see an example plot of the average step-size sequence and mismatch (defined as $\|\tilde{\mathbf{w}}_i\|^2/\|\mathbf{w}\|^2$) attainable for two examples of APA and AAPA. We see that both the mismatch and the step-size exhibit the expected behavior. We note that the filter has not reached steady-state, but this is not important, since $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ converges when the step-size sequence does, which is sufficient for testing Theorem 1.

In Fig. 2 we compare the results of Theorem 1 for APA under a white ($K = 4$, $M = 64$) and AR1 ($K = 8$, $M = 128$) inputs for SNR = 20 dB (Scenario 2). For the white input $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ is diagonal and real, with only the (q, q) with $q > 1$ are nonzero and given by (37). For the AR1 the matrix is upper triangular with the same diagonal. We see that in both cases the simulations match the theory very well. In Fig. 3 we show results for APA in more demanding scenarios where the condition $M \gg K$ is strained and SNR= 10 dB. We consider the white ($K = 8$, $M = 64$) and AR1 ($K = 4$, $M = 32$) inputs as well. We can see that even in this scenario the simulations match the theoretical results well.

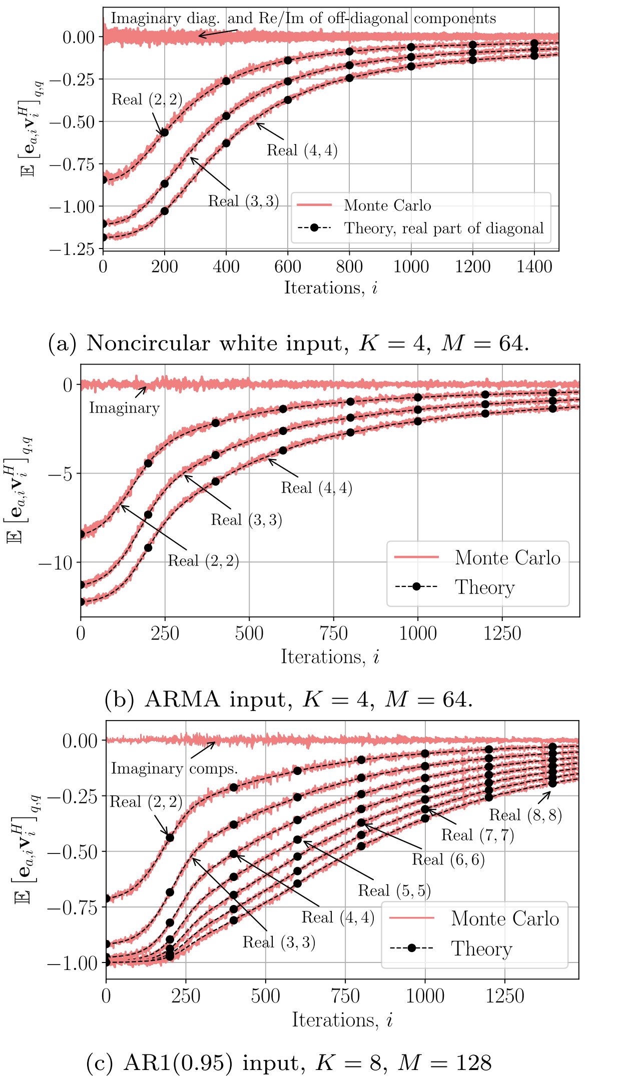
We now focus on the AAPA. In Fig. 4 we can see the result for non-circular white ($K = 4$, $M = 64$), circular AR1 ($K = 4$, $M = 64$) and non-circular ARMA ($K = 8$, $M = 128$) for SNR = 20 dB. In the three cases, we see that the simulations match the results accurately. Finally, in Fig. 5 we test the results of Theorem 1 in a more demanding scenario, where the hypotheses $M \gg K$ is strained, and the SNR = 10 dB. In particular we consider the white non-circular input($K = 8$, $M = 64$), non-circular ARMA ($K = 8$, $M = 64$) and non-circular AR1 ($K = 4$, $M = 32$). Again in this case we see that both simulations are in excellent agreement with the theoretical values, even in this regime.

To conclude this section, we see that both for the case of APA with circular inputs, and AAPA with circular and non-circular inputs, the results of Theorem 1 give a good fit the simulated values. We see however, that the variance of the Monte Carlo estimation is quite large, requiring a substantial number of runs as compared to for example, the mismatch curves.

5.2. Steady-state error analysis

In this section we evaluate the accuracy of the steady-state expression of Section 4. On one hand we want to validate Eq. (46) of $\mathbb{E}\left[\|\mathbf{e}\|^2\right]_\infty$, which is valid for a small step-size with general signals and for any step-size for white signals (provided that $M \gg K$). Then we want to validate Eq. (53) of the diagonal of the steady-state error covariance matrix $\mathbb{E}\left[\mathbf{e}\mathbf{e}^H\right]_\infty$, and in particular, the steady-state MSE.

We start by studying the validity of (46) for $\mathbb{E}\left[\|\mathbf{e}\|^2\right]_\infty$. We would like to mention that this expression was presented in [21] for circular white input signals for APA, but we present the white case for completeness. In Fig. 6 we show the results for APA with a circular white and AR1 input processes with SNR = 20 dB. In Fig. 7 we show the same simulations for AAPA, with white non-circular, circular AR1 (0.95) and for the non-circular ARMA inputs. Each curve was obtained by averaging $J = 400$ runs of the error to convergence and averaging the last 400 samples of the run. Since the formula is not dependent on M we do the simulations for different values of M. We see that similar conclusions can be drawn for APA and AAPA. For a white input the formula is exact for $0 \le \mu \le 1$, provided that $M \gg K$. The simulations

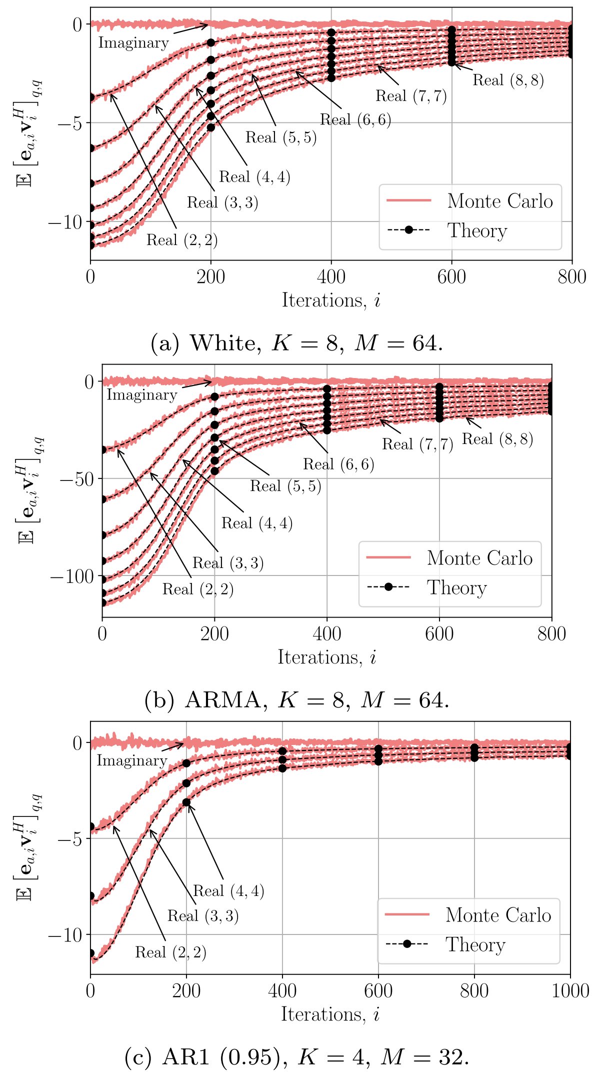


(a) Noncircular white input, $K = 4$, $M = 64$.

(b) ARMA input, $K = 4$, $M = 64$.

(c) AR1(0.95) input, $K = 8$, $M = 128$

Fig. 4. AAPA. Comparison of Monte Carlo simulations of $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ compared to (37) and other results of Theorem 1 for different input processes. SNR = 20 dB. $J = 16000$.



(a) White, $K = 8$, $M = 64$.

(b) ARMA, $K = 8$, $M = 64$.

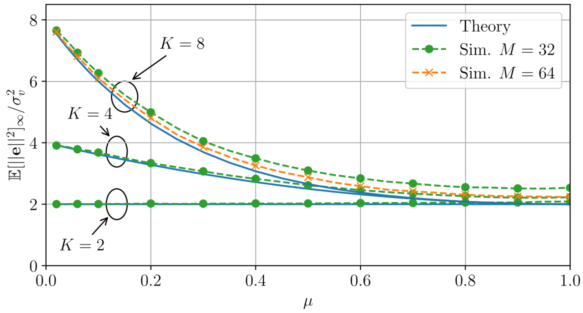
(c) AR1 (0.95), $K = 4$, $M = 32$.

Fig. 5. AAPA. Comparison of Monte Carlo simulations of $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ compared to (37) and other results of Theorem 1 for different input processes. SNR = 10 dB. $J = 16000$.

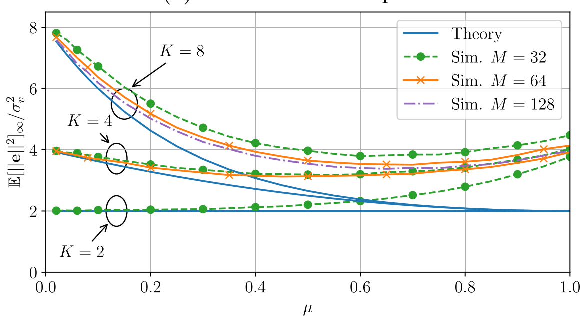
indeed support this both for APA with circular and AAPA for non-circular white noises: for $K = 2, 4$ $M = 32$ already provides excellent results, while for $K = 8$ we see that the accuracy improves with M. For the circular AR1 process, as expected the formula is valid for $\mu \ll 1$, as long as $M \gg K$. For example, for $K = 8$ the accuracy increases as M increases. Finally, for AAPA with the non-circular ARMA process we observe that the approximations are also good provided that μ is small. It is interesting to mention that the observed behavior on the validity of (46) for white signals for all μ and for small μ for arbitrary signals is compatible with the analysis presented in Section 4. Finally, we would like to point out that (46) depends on the approximation (42) and on (39). Eq. (39) has already shown to be accurate in the previous section, which means that by studying the validity of (46) we have also shown the accuracy of approximation (42) for small μ.

Next we evaluate the formula for the MSE (54) compared to the expressions discussed in Section 4.1. In Fig. 8 we plot the steady-state MSE for APA with circular white and AR1 (0.95) input for $K = 2, 4, 8$ and SNR = 20 dB. The theoretical expression (54) is valid when $M \gg K$, so the value of M is increased with K in a linear fashion, setting $M = 16K$. We compare (54) with (55), with (41) from [18] (reproduced as (56) in this paper), and with (31) from [22]. The characteristics of these formulas have already been discussed in Section 4.1. Eq. (55) is very simple, but shows the worst performance of all since it does not depend on K; hence it under-estimates the MSE in all cases. Comparing (54) with (31) from [18] we see that for white inputs both formulas give very similar results for all values of μ. This may be because in that case the eigenvalues of the input process are all the same. For the AR1 input, the expressions give different results for large μ but are close for small μ and close to the simulated values, making them useful to predict the behavior of the algorithm. Finally, neither give useful results for large μ; this is expected for (54) but less expected for (31), which depends on the statistics of the input process. On the other hand, (41) from [22] for $K = 2$ and $K = 4$ appears to under-estimate the MSE for white inputs, and for the AR1 input it gives useful results for small μ, although the other expressions seem to be more accurate. For the case of $K = 8$ and $M = 128$ it was not possible to this expression on a desktop computer due the large memory requirements. As a conclusion we see that for small μ and when $M \gg K$ (54) is as accurate as the alternatives, requires less information and is computationally less demanding.

In Fig. 9 we show the results for the MSE of AAPA with white non-circular, circular AR1 (0.95) and ARMA non-circular inputs the same setup as before. In this case we compare against (55) and also with (47) from [12]. Eq. (55) has the same limitations as in the case of APA, and in general underestimates the MSE. Comparing (54) with (47) from [12] for $K = 2$ and $K = 4$ we see that both predict similar

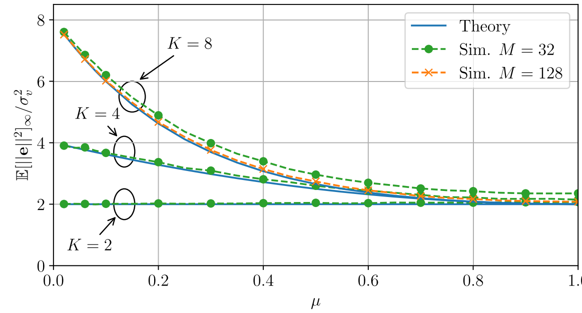


(a) Circular white input.

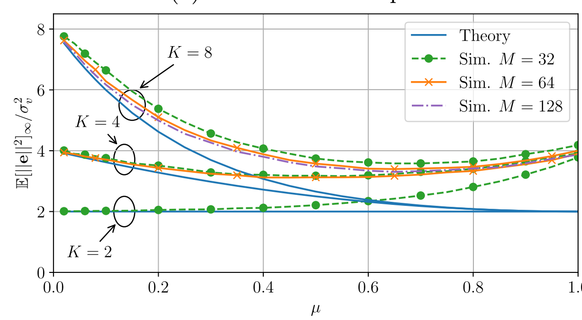


(b) Circular AR1 (0.95).

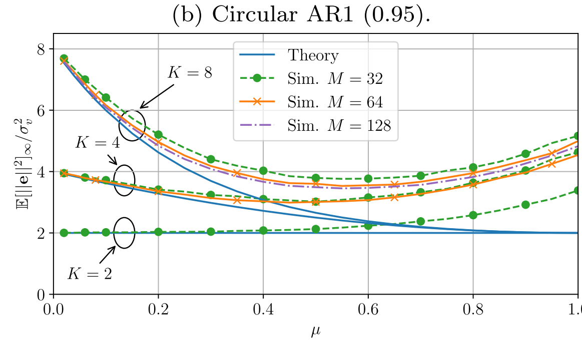
Fig. 6. APA: steady-state of the square norm of the error vector for different inputs. The theoretical expression is (46). SNR = 20 dB. $J = 400$.



(a) Circular white input.



(b) Circular AR1 (0.95).



(c) Non-circular ARMA.

Fig. 7. AAPA: steady-state of the square norm of the error vector for different inputs. The theoretical expression is (46). SNR = 20 dB. $J = 400$.

results for small μ and are very close to the Monte Carlo simulations. For $M = 128$ it was not possible to compute (47) on a desktop computer due to the large memory required. However, (54) is still straightforward to compute and provides a good approximation to the MSE. For large values of μ all the formulas become less accurate because in this regime the asymptotic MSE is strongly dependent on the characteristics of the input process; however, this is not a typical asymptotic operating condition for the algorithm.

We now show an example of application of the proposed expression (54) to predict the performance of the VSS-WLCAPA from [12]. In [12] the authors propose the following equation to estimate the asymptotic value of the step-size under Gaussian noise (erf is the standard error function):

$$\mathbb{E}[\mu_\infty] = \exp(-\theta) + \sqrt{\pi\theta}\left[\operatorname{erf}(\sqrt{\theta}) - 1\right]. \tag{59}$$

This formula depends only on a tuning parameter θ, typically in the range $(1,\ 4)$. Using (59) together with (47) from [12], the authors are able to predict the asymptotic value of the MSE as a function of θ. We now compare the predicted value of this approach with the predicted value obtained using (59) with the proposed formula (54). The results can be seen in Fig. 10 for $K = 2$ and $K = 4$, where we have selected values of M for which the assumption $M \gg K$ does not hold very well. We do not show the case $K = 8, M = 64$ because we could not compute the results from [12] using a desktop computer. The results from Fig. 10 show that (54) provides predicted performances which are compatible with the ones from [12] for practical values of θ at a much reduced computational cost and allows the estimation of the asymptotic MSE for longer filters with good performance. For large θ the formulas are less accurate of the dependence on the asymptotic MSE on the characteristics of the input process, but these values of θ are not in the practical range to achieve a good asymptotic performance.

To finish the section, we explore the validity of (53) to estimate steady-state energy of the other components of the error. The results for APA and AAPA can be seen in Figs. 11 and 12, respectively. We plot the estimated and simulated components of the energy of the errors for $K = 4$, $M = 32$ and SNR = 20 dB. In this scenario the condition $M \gg K$ does not hold very well, but we see that nevertheless the estimations are accurate are quite accurate for small values of μ even for colored signals. The same behavior is seen for larger values of K provided the values of M are raised proportionately.

6. Conclusion

In this paper we performed an analysis of correlation between the noise and *a priori* error vector of both APA (for circular signals) and AAPA (for general signals) using a unified framework. We showed that both algorithms share some common expressions of this correlation both in the transient and steady-state. We used this analysis to perform a new steady-state analysis of both algorithms in the small step-size regime. We have shown that under these conditions both algorithms satisfy the same closed-form formulas for the MSE, which does not depend on the statistics of the input signal. We have also provided approximate expressions for the energy of the other components of the error vector which are common to both algorithms.

CRediT authorship contribution statement

Andrés Altieri: Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

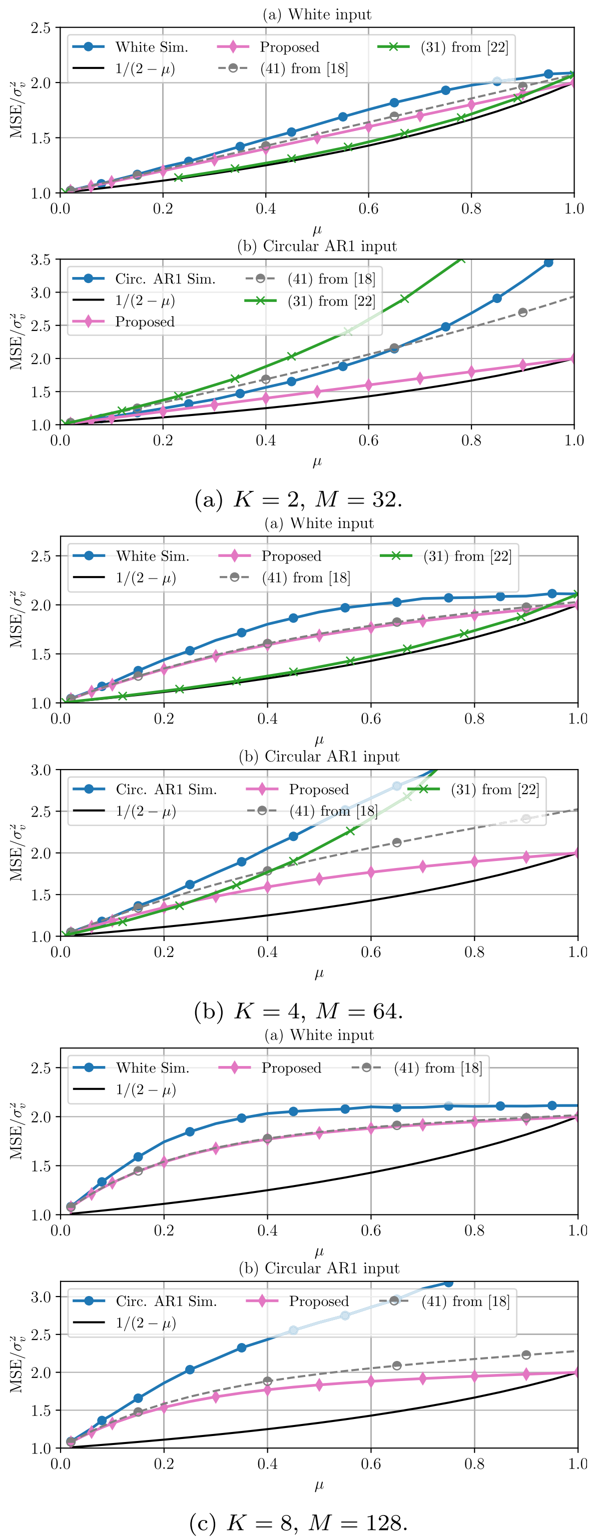


Fig. 8. APA: steady-state MSE for different inputs and values of K. The theoretical expression is (54). SNR = 20 dB. J = 400. (31) from [22] could not be computed for M = 128. 500 runs were used for [22].

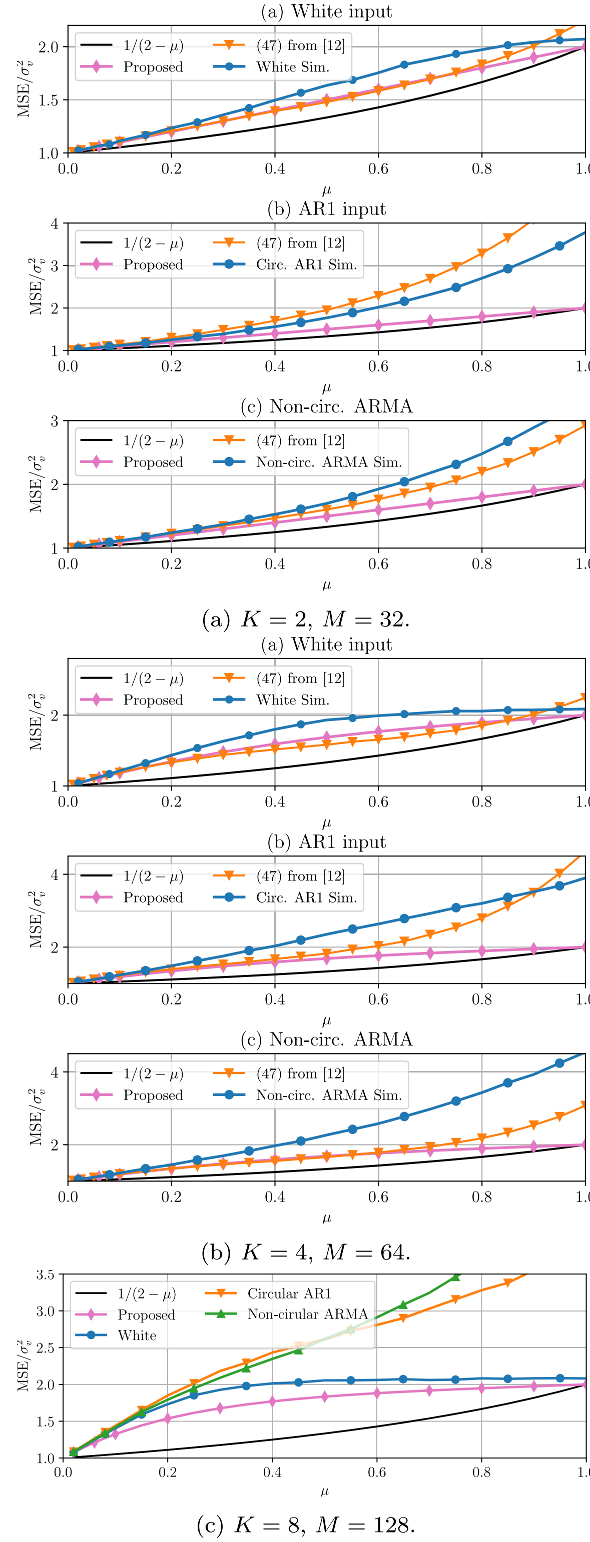


Fig. 9. AAPA: steady-state MSE for different inputs and values of K. The theoretical expression is (54). SNR = 20 dB. J = 400. It was not possible to compute (47) from [12] for M = 128. 450 runs where used for [12].

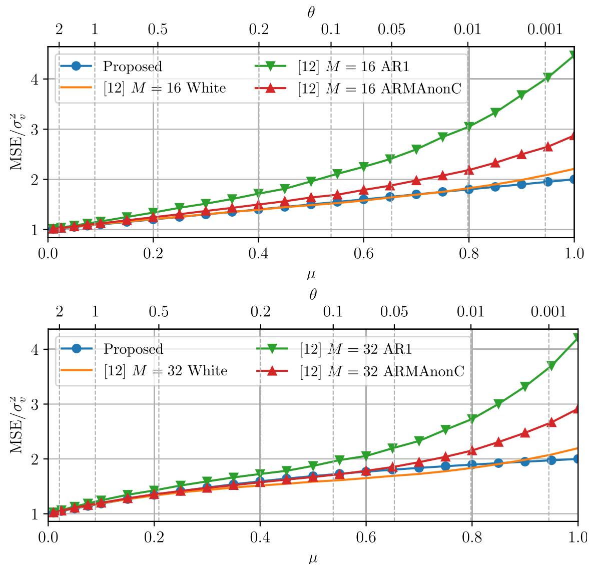


Fig. 10. Comparison of the asymptotic MSE for VSS-WLCAPA obtained from using (59) in (57) from [12] to using (59) in (54), as a function of setup parameter θ. 450 runs where used for the formulas from [12].

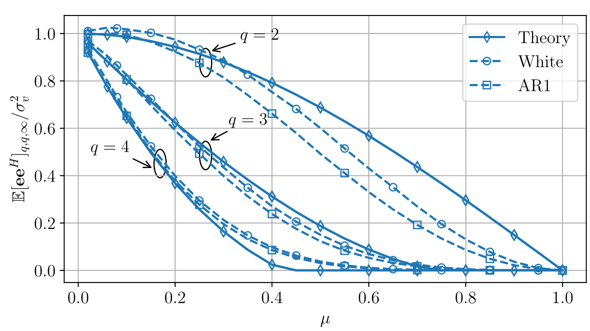


Fig. 11. APA: steady-state energy of the error components. The theoretical expression is (53). $K = 4$, $M = 32$, SNR = 20 dB. $J = 400$.

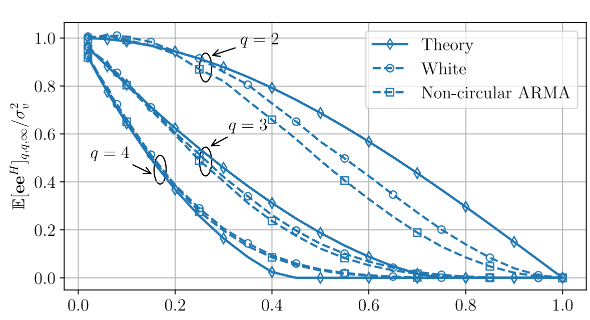


Fig. 12. AAPA: steady-state energy of the error components. The theoretical expression is given by (53). $K = 4$, $M = 32$, SNR = 20 dB. $J = 400$.

Data availability

Data will be made available on request.

Acknowledgment

The author would like to thank Dr. Hernan Rey for the discussions during the course of this work.

Appendix. Proof of Theorem 1

We first present two subsections with preliminary results, and the proof itself is in Appendix A.4.

A.1. Auxiliary linear algebra results

We present some auxiliary definitions and properties which will simplify the main proof of the theorem. The proofs in this section are straightforward and are omitted.

Definition A.1. Let $\mathcal{F}_m$ be the set of matrices in $\mathbb{C}^{K\times K}$ whose last $K-m$ rows are zero:

$$\mathcal{F}_m := \left\{\mathbf{A} \in \mathbb{C}^{K\times K} : [\mathbf{A}]_{i,j} = 0 \text{ if } m < i \le K\right\}. \tag{60}$$

Definition A.2. Let $\mathcal{C}_m$ be the set of matrices in $\mathbb{C}^{K\times K}$ whose first m columns are zero:

$$\mathcal{C}_m := \left\{\mathbf{A} \in \mathbb{C}^{K\times K} : [\mathbf{A}]_{i,j} = 0 \text{ if } 1 \le j \le m\right\}. \tag{61}$$

The following properties on the products of matrices hold:

Property A.1. *If $\mathbf{A} \in \mathcal{F}_m$ and $\mathbf{R} \in \mathbb{C}^{K\times K}$, then $\mathbf{AR} \in \mathcal{F}_m$.*

Property A.2. *If $\mathbf{A} \in \mathcal{C}_m$ and $\mathbf{R} \in \mathbb{C}^{K\times K}$ then $\mathbf{RA} \in \mathcal{C}_m$.*

Property A.3. *If $\mathbf{A} \in \mathcal{F}_n$, $\mathbf{B} \in \mathcal{C}_m$ and $\mathbf{R} \in \mathbb{C}^{K\times K}$, then from Properties A.1 and A.2 we have that $\mathbf{ARB} \in \mathcal{F}_n \cap \mathcal{C}_m$.*

Property A.4. *If $\mathbf{A} \in \mathcal{F}_m$ and $j \in \mathbb{Z}$, $j < 0$ then $\tilde{\mathbf{I}}_{K,j}\mathbf{A} \in \mathcal{F}_{m-j}$, with $\tilde{\mathbf{I}}_{K,j}$ given by (28).*

Property A.5. *If $\mathbf{R} \in \mathcal{F}_m \cap \mathcal{C}_m$ then $\mathbf{R}$ is an upper-triangular matrix, and* $\operatorname{diag}(\mathbf{R}) = \mathbf{0}$.

Definition A.3. Let $\mathbf{I}_{K,m}$ be a diagonal matrix such that the first m elements in the diagonal are 0, and the rest are 1:

$$[\mathbf{I}_{K,m}]_{i,j} = \begin{cases} 1 & \text{if } i = j \text{ and } i > m \\ 0 & \text{otherwise.} \end{cases} \tag{62}$$

Notice that $\mathbf{I}_{K,0}$ is the $K \times K$ identity matrix.

Consider the matrices $\tilde{\mathbf{I}}_{K,m}$ given by (28). When multiplying a matrix $\tilde{\mathbf{I}}_{K,m}$ with another matrix it will move the rows of a matrix up or down and pad with zeros. In addition, $\tilde{\mathbf{I}}_{K,0}$ is the $K \times K$ identity matrix.

We will use the following properties:

Property A.6. *For $1 \le m \le K-1$, $\tilde{\mathbf{I}}_{K,-m}\tilde{\mathbf{I}}_{K,m} = \mathbf{I}_{K,m}$.*

Property A.7. *If $m, n \in \{1,\dots,K-1\}$ then $\tilde{\mathbf{I}}_{K,m}\tilde{\mathbf{I}}_{K,n} = \tilde{\mathbf{I}}_{K,n+m}$.*

A.2. A matrix decomposition for the shifted correlation matrices

Consider a complex second order stationary process u_i and the vector $\tilde{\mathbf{u}}_i = [u_i,\dots,u_{i-K+1}]^T$ as defined in Section 2.4. Under scenario 1 (with u_i non-circular) let us define the time-shifted correlation matrix: $\mathbf{R}_{x,-m} = \mathbb{E}\left[\tilde{\mathbf{u}}_{R,i}\tilde{\mathbf{u}}_{R,i-m}^T + \tilde{\mathbf{u}}_{I,i}\tilde{\mathbf{u}}_{I,i-m}^T\right]$. Under scenario 2 (with u_i circular) let us define the conjugate time-shifted correlation matrix as $\mathbf{R}_{x,-m} = \mathbb{E}\left[\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_{i-m}^H\right]^*$. These two matrices appeared in Section 2.4 when computing $\mathbb{E}\left[\mathbf{X}_i^H\mathbf{X}_{i-m}\right]$ and are shown to be Toeplitz matrices. For $m = 0$ there is no time shift so we simply write $\mathbf{R}_{x,0} \equiv \mathbf{R}_x$.

Lemma A.1. *For $0 < m < K$ the matrix $\mathbf{R}_{x,-m}$ can be written in terms of $\mathbf{R}_x$ as:*

$$\mathbf{R}_{x,-m} = \tilde{\mathbf{I}}_{K,-m}\mathbf{R}_x + \mathbf{M}_m, \tag{63}$$

where $\mathbf{M}_m$ is some matrix in $\mathcal{F}_m$. When x is a white process, $\mathbf{M}_m$ is the zero matrix.

The proof is straightforward and is omitted. Essentially $\tilde{\mathbf{I}}_{K,-m}$ moves the rows of $\mathbf{R}_x$ downwards padding with rows of zeros from the top, while $\mathbf{M}_m$ completes the first m rows of $\mathbf{R}_{x,-m}$.

A.3. Expansion of a product of matrices as a sum

We now rewrite the product

$$\prod_{k=1}^{j-1}(\mathbf{I}-\mathbf{G}_k) = (\mathbf{I}-\mathbf{G}_1)(\mathbf{I}-\mathbf{G}_2)\dots(\mathbf{I}-\mathbf{G}_{j-1}), \tag{64}$$

which appears in (30) as a summation. Each of its terms can be obtained by selecting from each $(\mathbf{I}-\mathbf{G}_k)$ either the identity or the $\mathbf{G}_k$ matrix. Since the identity matrices do not affect the product, we can represent each term in the summation as a list of ordered indexes which correspond to the $\mathbf{G}_k$ matrices which appear in the summation term. Also, each term of the summation will be multiplied by (-1) if there is an odd number of $\mathbf{G}_k$ matrices involved. The pth term will be represented by an ordered list of indexes:

$$\mathcal{I}_p = [\sigma_{1,p},\dots,\sigma_{\#I_p,p}], \tag{65}$$

and the matrix product will be written as:

$$\prod_{k=1}^{j-1}(\mathbf{I}-\mathbf{G}_k) = \mathbf{I} + \sum_p (-1)^{\#\mathcal{I}_p}\prod_{n=1}^{\#I_p}\mathbf{G}_{i-\sigma_{n,p}} \tag{66}$$

$$= \mathbf{I} + \sum_p (-1)^{\#I_p}\prod_{n=1}^{\#\mathcal{I}_p}\mathbf{X}_{\sigma_{n,p}}\mathbf{S}_{\sigma_{n,p}}\mathbf{X}^H_{\sigma_{n,p}}. \tag{67}$$

The summation will contain $2^{j-1}-1$ terms and any indexed list $\mathcal{I}_p$ will satisfy the following properties:

1. $\sigma_{i,p}\in\mathbb{N} \;\; \forall i>K$.
2. $\#\mathcal{I}_p<j$.
3. $i>\sigma_{1,p}>\sigma_{2,p}>\dots>\sigma_{\#\mathcal{I}_p,p}>i-j$.

A.4. Proof of Theorem 1

In order to prove the theorem we need to start from (30). We focus on $K>2$, and leave $K=2$ for the reader. Using (67) in (30) and rearranging the terms we obtain ($K>2$):

$$\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right] = -\sigma_v^2\sum_{j=1}^{K-1}\mathbb{E}\left[\mathbf{X}_i^H\mathbf{J}_j\right]\tilde{\mathbf{I}}_{K,j}$$
$$-\sigma_v^2\sum_{j=2}^{K-1}\sum_p(-1)^{\#\mathcal{I}_p}\mathbb{E}\left[\mathbf{X}_i^H\left(\prod_{\sigma_n\in\mathcal{I}_p}\mathbf{X}_{\sigma_n}\mathbf{S}_{\sigma_n}\mathbf{X}^H_{\sigma_n}\right)\mathbf{J}_j\right]\tilde{\mathbf{I}}_{K,j}. \tag{68}$$

The coefficients σ_n are a function of p also, but to simplify the notation, we do not make this dependence explicit, that is, $\sigma_{n,p}\equiv\sigma_n$. Likewise we do not include the limits of the summation in p because it is not used. Now we simplify both terms of these expressions using the preliminary results of this appendix.

The first term on the right side of (68) can be written as:

$$-\sigma_v^2\sum_{j=1}^{K-1}\mathbb{E}\left[\mathbf{X}_i^H\mathbf{J}_j\right]\tilde{\mathbf{I}}_{K,j} \approx -\sigma_v^2\sum_{j=1}^{K-1}\mu_{i-j}\mathbf{R}_{x,-j}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j} \tag{69}$$

using (35) for Scenario 1 or (36) for Scenario 2. Now we have the following lemma:

Lemma A.2. *The right side of (69) can be written as:*

$$\sum_{j=1}^{K-1}\mu_{i-j}\mathbf{R}_{x,-j}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j} = \sum_{j=1}^{K-1}\mu_{i-j}\left(\mathbf{I}_{K,j}+\mathbf{M}_j\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}\right) \tag{70}$$

for certain matrices $\mathbf{M}_j\in\mathcal{F}_j$. In addition, $\mathbf{M}_j\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}\in\mathcal{C}_j\cap\mathcal{F}_j$.

Proof. To show this, we use Lemma A.1, to write $\mathbf{R}_{x,-j}=\tilde{\mathbf{I}}_{K,-j}\mathbf{R}_x+\mathbf{M}_j$, where $\mathbf{M}_j\in\mathcal{F}_j$. After replacing this in (69) we use Property A.6 to show that $\mathbf{I}_{K,j}=\tilde{\mathbf{I}}_{K,-j}\tilde{\mathbf{I}}_{K,j}$. Finally, notice that using Property A.3 we know that $\mathbf{M}_j\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}\in\mathcal{C}_j\cap\mathcal{F}_j$. □

Then for the second term on the right side of (68) we replace $\mathbf{S}_{\sigma_n}$ and $\mathbf{J}_j$ in terms of μ_{σ_n}, μ_{i-j}, $\mathbf{X}_{\sigma_n}$ and $\mathbf{X}_{i-j}$. We then define: $c(p)=\#\mathcal{I}_p$ and define $\tilde{\mathcal{I}}(p)=[\sigma_1,\dots,\sigma_{c(p)-1}]$ as the list in which the last element has been removed. Then, according to the scenario we use (35) or (36) to approximate:

- $\mathbf{X}_i^H\mathbf{X}_{\sigma_1}\approx M\mathbf{R}_{x,\sigma_1-i}$.
- $\mathbf{X}_{\sigma_n}^H\mathbf{X}_{\sigma_{n+1}}\approx M\mathbf{R}_{x,\sigma_{n+1}-\sigma_n}$.
- $\mathbf{X}_{\sigma_{c(p)}}^H\mathbf{X}_{i-j}\approx M\mathbf{R}_{x,i-j-\sigma_{c(p)}}$.

where $\mathbf{R}_{x,-m}$ has been defined in Appendix A.2. With this we can approximate:

$$\mathbb{E}\left[\mathbf{X}_i^H\left(\prod_{\sigma_n\in\mathcal{I}_p}\mathbf{X}_{\sigma_n}(\mathbf{X}^H_{\sigma_n}\mathbf{X}_{\sigma_n})^{-1}\mathbf{X}^H_{\sigma_n}\right)\mathbf{J}_j\right]\tilde{\mathbf{I}}_{K,j}\approx\mathbf{R}_{x,\sigma_1-i}$$
$$\times\mathbf{R}_x^{-1}\left[\prod_{\sigma_n\in\tilde{\mathcal{I}}(p)}\mathbf{R}_{x,\sigma_{n+1}-\sigma_n}\mathbf{R}_x^{-1}\right]\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}. \tag{71}$$

Although this expression is involved, it is possible to prove the following Lemma:

Lemma A.3. *The left side of (71) can be written as:*

$$\mathbf{R}_{x,\sigma_1-i}\mathbf{R}_x^{-1}\left(\prod_{\sigma_n\in\tilde{\mathcal{I}}(p)}\mathbf{R}_{x,\sigma_{n+1}-\sigma_n}\mathbf{R}_x^{-1}\right)\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}=\mathbf{I}_{K,j}+\mathbf{T}_{0,p,j}, \tag{72}$$

where $\mathbf{T}_{0,p,j}\in\mathcal{F}_j\cap\mathcal{C}_j$.[2] For a white input, $\mathbf{T}_{0,p,j}$ is the null matrix.

Proof. The proof proceeds by evaluating the product in (72) from right to left. To do this we proceed by induction, by showing that the following formula is valid:

$$\left(\prod_{\sigma_k\in\tilde{\mathcal{I}}(p):k\ge n}\mathbf{R}_{x,\sigma_{k+1}-\sigma_k}\mathbf{R}_x^{-1}\right)\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}=\tilde{\mathbf{I}}_{i-j-\sigma_n}\tilde{\mathbf{I}}_{K,j}+\mathbf{T}_n, \tag{73}$$

with $\mathbf{T}_n\in\mathcal{F}_{t_n}\cap\mathcal{C}_j$, $t_n=\max\left\{\sigma_n-(i-j),\sigma_n-\sigma_m\right\}$ and $n=1\dots(m-1)$. The induction proceeds backwards, starting from $k=c(p)-1$ to $k=1$. With this expression, (72) can be simplified to complete the proof.

First we factor the matrix $\mathbf{R}_{x,i-j-\sigma_{c(p)}}$ using Lemma A.1 and we simplify the left side of (73) as:

$$\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}=\left(\tilde{\mathbf{I}}_{i-j-\sigma_{c(p)}}\tilde{\mathbf{I}}_{K,j}+\mathbf{T}_{c(p)}\right), \tag{74}$$

where $\mathbf{T}_{c(p)}=\mathbf{M}_{c(p)}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}$. Since $\mathbf{M}_{c(p)}\in\mathcal{F}_{\sigma_{c(p)}-(i-j)}$ and $\tilde{\mathbf{I}}_{K,j}\in\mathcal{C}_j$ then using Property A.3 we have $\mathbf{T}_{c(p)}\in\mathcal{F}_{\sigma_{c(p)}-(i-j)}\cap\mathcal{C}_j$.

Now we prove that the result is valid for $n=c(p)-1$. In this case, the product (73) indexed by k only has the term $k=c(p)-1$. Applying the decomposition from Lemma A.1 to this term and using (74) we have:

$$\left(\mathbf{R}_{x,\sigma_{c(p)}-\sigma_{c(p)-1}}\mathbf{R}_x^{-1}\right)\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}=\left(\tilde{\mathbf{I}}_{\sigma_{c(p)}-\sigma_{c(p)-1}}+\mathbf{M}_{c(p)-1}\mathbf{R}_x^{-1}\right)$$
$$\times\left(\tilde{\mathbf{I}}_{i-j-\sigma_{c(p)}}\tilde{\mathbf{I}}_{K,j}+\mathbf{T}_{c(p)}\right). \tag{75}$$

We now expand and analyze the four terms of (75):

- Since $\sigma_{c(p)}-\sigma_{c(p)-1}<0$ and $i-j-\sigma_{c(p)}<0$ we apply Property A.7 to show that $\tilde{\mathbf{I}}_{\sigma_{c(p)}-\sigma_{c(p)-1}}\tilde{\mathbf{I}}_{i-j-\sigma_{c(p)}}\tilde{\mathbf{I}}_{K,j}=\tilde{\mathbf{I}}_{i-j-\sigma_{c(p)-1}}\tilde{\mathbf{I}}_{K,j}$.
- Since $\mathbf{T}_{c(p)}\in\mathcal{F}_{\sigma_{c(p)}-(i-j)}\cap\mathcal{C}_j$ we apply Property A.4 to show that $\tilde{\mathbf{I}}_{\sigma_{c(p)}-\sigma_{c(p)-1}}\mathbf{T}_{c(p)}\in\mathcal{F}_{\sigma_{c(p)-1}-(i-j)}\cap\mathcal{C}_j$.

[2] We write $\mathbf{T}_{0,p,j}$ to indicate they depend on the jth and pth indexes of the summations in (68). During the proof we do not explicit the indexes p,j.

- $\mathbf{M}_{c(p)-1}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{i-j-\sigma_{c(p)}}\tilde{\mathbf{I}}_{K,j} \in \mathcal{F}_{\sigma_{c(p)-1}-\sigma_c(p)} \cap \mathcal{C}_j$ since $\mathbf{M}_{c(p)-1} \in \mathcal{F}_{\sigma_m-\sigma_{m-1}}$ and $\mathbf{I}_{K,j} \in \mathcal{C}_j$ (Property A.3).
- $\mathbf{M}_{c(p)-1}\mathbf{R}_x^{-1}\mathbf{T}_{c(p)} \in \mathcal{F}_{\sigma_{c(p)-1}-\sigma_c(p)} \cap \mathcal{C}_j$ because $\mathbf{M}_{c(p)-1} \in \mathcal{F}_{\sigma_{c(p)}-\sigma_{c(p)-1}}$ and $\mathbf{T}_{c(p)} \in \mathcal{C}_j$ (Property A.3).

Thus, all the terms, except the first belong either to $(\mathcal{F}_{\sigma_{c(p)-1}-\sigma_c(p)} \cap \mathcal{C}_j)$ or $(\mathcal{F}_{\sigma_{c(p)-1}-(i-j)} \cap \mathcal{C}_j)$. We now define $t_{c(p)-1} \triangleq \max\left\{\sigma_{c(p)-1}-(i-j), \sigma_{c(p)-1} -\sigma_{c(p)}\right\}$ and:

$$\mathbf{T}_{c(p)-1} \triangleq \tilde{\mathbf{I}}_{\sigma_{c(p)}-\sigma_{c(p)-1}}\mathbf{T}_{c(p)} + \mathbf{M}_{c(p)-1}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{i-j-\sigma_{c(p)}}\tilde{\mathbf{I}}_{K,j} + \mathbf{M}_{c(p)-1}\mathbf{R}_x^{-1}\mathbf{T}_{c(p)}, \tag{76}$$

so that $\mathbf{T}_{c(p)-1} \in \mathcal{F}_{t_{c(p)-1}} \cap \mathcal{C}_j$. Then, (75) is written as:

$$\mathbf{R}_{x,\sigma_{c(p)}-\sigma_{c(p)-1}}\mathbf{R}_x^{-1}\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j} = \tilde{\mathbf{I}}_{i-j-\sigma_{c(p)-1}}\tilde{\mathbf{I}}_{K,j} + \mathbf{T}_{c(p)-1}, \tag{77}$$

where $\mathbf{T}_{c(p)-1} \in \mathcal{F}_{t_{c(p)-1}} \cap \mathcal{C}_j$ and $t_{c(p)-1} = \max\left\{\sigma_{c(p)-1}-(i-j), \sigma_{c(p)-1} -\sigma_m\right\} > 0$, which shows that the expression is valid for $n = c(p)-1$.

Now we assume that (73) is valid for $k = n+1$ and show that it is valid for $k = n$. This means that:

$$\prod_{\sigma_k \in \tilde{\mathcal{I}}(p):k\geq n}\left(\tilde{\mathbf{I}}_{\sigma_{k+1}-\sigma_k} + \mathbf{M}_k\mathbf{R}_x^{-1}\right)\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j} \tag{78}$$

$$= \left(\tilde{\mathbf{I}}_{\sigma_{n+1}-\sigma_n} + \mathbf{M}_n\mathbf{R}_x^{-1}\right)\left(\tilde{\mathbf{I}}_{i-j-\sigma_{n+1}}\tilde{\mathbf{I}}_{K,j} + \mathbf{T}_{n+1}\right), \tag{79}$$

where we have applied the inductive hypothesis. We have that $\mathbf{T}_{n+1} \in \mathcal{F}_{t_{n+1}} \cap \mathcal{C}_j$ with $t_{n+1} = \max\left\{\sigma_{n+1}-(i-j), \sigma_{n+1}-\sigma_{c(p)}\right\}$. Expanding the product and analyzing the terms we have:

- $\tilde{\mathbf{I}}_{\sigma_{n+1}-\sigma_n}\tilde{\mathbf{I}}_{i-j-\sigma_{n+1}}\tilde{\mathbf{I}}_{K,j} = \tilde{\mathbf{I}}_{i-j-\sigma_n}\tilde{\mathbf{I}}_{K,j}$.
- $\tilde{\mathbf{I}}_{\sigma_{n+1}-\sigma_n}\mathbf{T}_{t_{n+1}} \in \mathcal{F}_{t_{n+1}+(\sigma_n-\sigma_{n+1})} \cap \mathcal{C}_j$.
- $\mathbf{M}_n\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{i-j-\sigma_{n+1}}\tilde{\mathbf{I}}_{K,j} \in \mathcal{F}_{\sigma_{n+1}-\sigma_n} \cap \mathcal{C}_j$.
- $\mathbf{M}_n\mathbf{R}_x^{-1}\mathbf{T}_{t_{n+1}} \in \mathcal{F}_{\sigma_{n+1}-\sigma_n} \cap \mathcal{C}_j$.

But since $t_{n+1} > 0$ we have $\sigma_n - \sigma_{n+1} + t_{n+1} > \sigma_n - \sigma_{n+1}$. In addition:

$$t_n = \max\left\{\sigma_{n+1}-(i-j), \sigma_{n+1}-\sigma_m\right\} + \sigma_n - \sigma_{n+1} = \max\left\{\sigma_n-(i-j), \sigma_n-\sigma_m\right\}.$$

So we find that:

$$\mathbf{T}_n = \tilde{\mathbf{I}}_{\sigma_{n+1}-\sigma_n}\mathbf{T}_{t_{n+1}} + \mathbf{M}_n\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{i-j-\sigma_{n+1}}\tilde{\mathbf{I}}_{K,j} + \mathbf{M}_n\mathbf{R}_x^{-1}\mathbf{T}_{t_{n+1}} \in \mathcal{F}_{t_n} \cap \mathcal{C}_j, \tag{80}$$

with $t_n = \max\left\{\sigma_n-(i-j), \sigma_n-\sigma_m\right\}$. This proofs that (73) is valid.

We can now simplify (72) by replacing (73) with $k=1$:

$$\mathbf{R}_{x,\sigma_1-i}\mathbf{R}_x^{-1}\left(\prod_{\sigma_n \in \tilde{\mathcal{I}}(p)}\mathbf{R}_{x,\sigma_{n+1}-\sigma_n}\mathbf{R}_x^{-1}\right)\mathbf{R}_{x,i-j-\sigma_{c(p)}}\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j} = \mathbf{R}_{x,\sigma_1-i}\mathbf{R}_x^{-1} \times\left(\tilde{\mathbf{I}}_{i-j-\sigma_1}\tilde{\mathbf{I}}_{K,j} + \mathbf{T}_1\right). \tag{81}$$

Applying once more the decomposition of Lemma A.1 to $\mathbf{R}_{x,\sigma_1-i}$ we get:

$$\mathbf{R}_{x,\sigma_1-i}\mathbf{R}_x^{-1}\left(\tilde{\mathbf{I}}_{i-j-\sigma_1}\tilde{\mathbf{I}}_{K,j} + \mathbf{T}_1\right) = \tilde{\mathbf{I}}_{\sigma_1-i}\tilde{\mathbf{I}}_{i-j-\sigma_1}\tilde{\mathbf{I}}_{K,j} + \mathbf{T}_0, \tag{82}$$

with $\mathbf{T}_0 = \tilde{\mathbf{I}}_{\sigma_1-i}\mathbf{T}_1 + \mathbf{M}_0\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{i-j-\sigma_1}\tilde{\mathbf{I}}_{K,j} + \mathbf{M}_0\mathbf{R}_x^{-1}\mathbf{T}_1$.

Following a similar reasoning as with the other terms we can prove that $\mathbf{T}_0 \in \mathcal{F}_{t_0} \cap \mathcal{C}_j$ with $t_0 = \max\left\{i-\sigma_1, i-\sigma_1+\max\left\{\sigma_1-(i-j), \sigma_1 -\sigma_m\right\}\right\} = i-\sigma_1+\max\left\{\sigma_1-(i-j), \sigma_1-\sigma_m\right\} = \max\left\{j, i-\sigma_m\right\}$. But since $\sigma_m > i-j$ we have that $t_0 = j$, which shows that $\mathbf{T}_0 \in \mathcal{F}_j \cap \mathcal{C}_j$. Finally, we conclude the proof by noting that: $\tilde{\mathbf{I}}_{\sigma_1-i}\tilde{\mathbf{I}}_{i-j-\sigma_1}\tilde{\mathbf{I}}_{K,j} = \tilde{\mathbf{I}}_{K,-j}\tilde{\mathbf{I}}_{K,j} = \mathbf{I}_{K,j}$, which is obtained using Properties A.6 and A.7. The result for a white input follows by noting that all the $\mathbf{M}$ matrices from the decomposition (63) are zero for a white input. □

To conclude the proof of Theorem 1, we use the results of Lemmas A.2 and A.3. We replace (72) in (71) and this result together with (70) in (68) to obtain:

$$\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right] \approx -\sigma_v^2\sum_{j=1}^{K-1}\mu_{i-j}\left(\mathbf{I}_{K,j} + \mathbf{M}_j\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}\right) -\sigma_v^2\left[\sum_{j=2}^{K-1}\sum_p(-1)^{f(p)}\mu_{i-j}\left(\prod_{\sigma_n\in\mathcal{I}_p}\mu_{\sigma_n}\right)\left(\mathbf{I}_{K,j} + \mathbf{T}_{0,p,j}\right)\right]. \tag{83}$$

Using Property A.5 we have that $\mathbf{T}_{0,p,j}$ and $\mathbf{M}_j\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}$ are upper triangular matrices, so $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ also is. In particular for a white input, $\mathbf{T}_{0,p,j}$ and $\mathbf{M}_j$ are the zero matrix, so $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]$ is a diagonal matrix. In both cases, from Property A.5 we have that $\mathrm{diag}(\mathbf{T}_{0,p,j}) = \mathrm{diag}(\mathbf{M}_j\mathbf{R}_x^{-1}\tilde{\mathbf{I}}_{K,j}) = \mathbf{0}$ for all j, p. So we can simplify this expression to obtain:

$$\mathrm{diag}\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right] \approx -\sigma_v^2\mu_{i-1}\mathrm{diag}\left(\mathbf{I}_{K,1}\right) -\sigma_v^2\sum_{j=2}^{K-1}\mu_{i-j}\left[1+\sum_p(-1)^{f(p)}\left(\prod_{\sigma_n\in\mathcal{I}_p}\mu_{\sigma_n}\right)\right]\mathrm{diag}\left(\mathbf{I}_{K,j}\right). \tag{84}$$

Now we conclude the proof by observing that

$$1+\sum_p(-1)^{f(p)}\prod_{\sigma_n\in\mathcal{I}_p}\mu_{\sigma_n} = \prod_{k=1}^{j-1}\left(1-\mu_{i-k}\right). \tag{85}$$

that is, we revert the matrix factorization from Appendix A.2 which is also valid for scalars. $\mathbb{E}\left[\mathbf{e}_{a,i}\mathbf{v}_i^H\right]_{q,q}$ can be found by considering the individual elements of the diagonal.

References

[1] S. Haykin, Adaptive Filter Theory, Prentice Hall, Englewood Cliffs, NJ, 2000.
[2] K. Ozeki, T. Umeda, An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties, Electron. Commun. Japan I 67 (5) (1984) 19–27, http://dx.doi.org/10.1002/ecja.4400670503.
[3] K. Ozeki, Theory of Affine Projection Algorithms for Adaptive Filtering, Springer, Japan, 2016.
[4] J. Benesty, C. Paleologu, T. Gänsler, S. Ciochină, A Perspective on Stereophonic Acoustic Echo Cancellation, Springer-Verlag, Berlin, Germany, 2011.
[5] C. Stanciu, J. Benesty, C. Paleologu, T. Gänsler, S. Ciochină, A widely linear model for stereophonic acoustic echo cancellation, Signal Process. 93 (2) (2013) 511–516, http://dx.doi.org/10.1016/j.sigpro.2012.08.017.
[6] D. Mandic, S. Javidi, S. Goh, A. Kuh, K. Aihara, Complex-valued prediction of wind profile using augmented complex statistics, Renew. Energy 34 (1) (2009) 196–201, http://dx.doi.org/10.1016/j.renene.2008.03.022.
[7] L. Liu, R. Zhang, K.-C. Chua, Multi-antenna wireless powered communication with energy beamforming, IEEE Trans. Commun. 62 (12) (2014) 4349–4361, http://dx.doi.org/10.1109/TCOMM.2014.2370035.
[8] Y. Xia, S.C. Douglas, D.P. Mandic, Adaptive frequency estimation in smart grid applications: Exploiting noncircularity and widely linear adaptive estimators, IEEE Signal Process. Mag. 29 (5) (2012) 44–54, http://dx.doi.org/10.1109/MSP.2012.2183689.
[9] Y. Xia, C.C. Took, D.P. Mandic, An augmented affine projection algorithm for the filtering of noncircular complex signals, Signal Process. 90 (6) (2010) 1788–1799, http://dx.doi.org/10.1016/j.sigpro.2009.11.026.
[10] Y. Xia, C.C. Took, S. Javidi, D.P. Mandic, A widely linear affine projection algorithm, in: 2009 IEEE/SP 15th Workshop on Statistical Signal Processing, 2009, pp. 373–376, http://dx.doi.org/10.1109/SSP.2009.5278562.
[11] P. Wen, J. Zhang, S. Zhang, D. Li, Augmented complex-valued normalized subband adaptive filter: Algorithm derivation and analysis, J. Franklin Inst. B 356 (3) (2019) 1604–1622, http://dx.doi.org/10.1016/j.jfranklin.2018.11.013.
[12] L. Shi, H. Zhao, Y. Zakharov, B. Chen, Y. Yang, Variable step-size widely linear complex-valued affine projection algorithm and performance analysis, IEEE Trans. Signal Process. 68 (2020) 5940–5953, http://dx.doi.org/10.1109/TSP.2020.3029884.
[13] B. Picinbono, P. Chevalier, Widely linear estimation with complex data, IEEE Trans. Signal Process. 43 (8) (1995) 2030–2033, http://dx.doi.org/10.1109/78.403373.
[14] Z. Luo, J. Zhou, Y. Pu, A widely linear complex-valued affine projection sign algorithm with its steady-state mean-square analysis, Circuits Syst. Signal Process. 41 (6) (2022) 3446–3464.
[15] C. Qiu, G. Qian, S. Wang, Widely linear maximum complex correntropy criterion affine projection algorithm and its performance analysis, IEEE Trans. Signal Process. 70 (2022) 3540–3550, http://dx.doi.org/10.1109/TSP.2022.3185903.
[16] S. Lv, H. Zhao, W. Xu, Robust widely linear affine projection M-estimate adaptive algorithm: Performance analysis and application, IEEE Trans. Signal Process. 71 (2023) 3623–3636, http://dx.doi.org/10.1109/TSP.2023.3311880.
[17] S. Sankaran, A. Beex, Convergence behavior of affine projection algorithms, IEEE Trans. Signal Process. 48 (4) (2000) 1086–1096.
[18] T.K. Paul, T. Ogunfunmi, On the convergence behavior of the affine projection algorithm for adaptive filters, IEEE Trans. Circuits Syst. I. Regul. Pap. 58 (8) (2011) 1813–1826, http://dx.doi.org/10.1109/TCSI.2011.2106091.

[19] P. Park, C.H. Lee, J.W. Ko, Mean-square deviation analysis of affine projection algorithm, IEEE Trans. Signal Process. 59 (12) (2011) 5789–5799, http://dx.doi.org/10.1109/TSP.2011.2165709.
[20] H.-C. Shin, A. Sayed, Mean-square performance of a family of affine projection algorithms, IEEE Trans. Signal Process. 52 (1) (2004) 90–102.
[21] H. Rey, L.R. Vega, S. Tressens, J. Benesty, Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice, IEEE Trans. Signal Process. 55 (5) (2007) 2096–2109.
[22] S.-E. Kim, J.-W. Lee, W.-J. Song, A theory on the convergence behavior of the affine projection algorithm, IEEE Trans. Signal Process. 59 (12) (2011) 6233–6239, http://dx.doi.org/10.1109/TSP.2011.2168524.
[23] R.A. Dobre, V.A. Niţă, S. Ciochină, C. Paleologu, New insights on the convergence analysis of the affine projection algorithm for system identification, in: 2015 Int. Symposium on Signals, Circuits and Systems, ISSCS, 2015, pp. 1–4, http://dx.doi.org/10.1109/ISSCS.2015.7203988.
[24] D. Slock, On the convergence behavior of the LMS and the normalized LMS algorithms, IEEE Trans. Signal Process. 41 (9) (1993) 2811–2825, http://dx.doi.org/10.1109/78.236504.
[25] V.A. Niţă, R.A. Dobre, S. Ciochină, C. Paleologu, Improved convergence model of the affine projection algorithm for system identification, in: 2017 International Symposium on Signals, Circuits and Systems, ISSCS, 2017, pp. 1–4, http://dx.doi.org/10.1109/ISSCS.2017.8034877.
[26] L. Rey Vega, H. Rey, A Rapid Introduction to Adaptive Filtering, Springer-Verlag Berlin Heidelberg, 2013.
[27] M.Z.A. Bhotto, A. Antoniou, Affine-projection-like adaptive-filtering algorithms using gradient-based step size, IEEE Trans. Circuits Syst. I. Regul. Pap. 61 (7) (2014) 2048–2056, http://dx.doi.org/10.1109/TCSI.2014.2304665.
[28] L. Rey Vega, H. Rey, J. Benesty, A robust variable step-size affine projection algorithm, Signal Process. 90 (9) (2010) 2806–2810, http://dx.doi.org/10.1016/j.sigpro.2010.03.029.
[29] H.-C. Shin, A. Sayed, W.-J. Song, Variable step-size NLMS and affine projection algorithms, IEEE Signal Process. Lett. 11 (2) (2004) 132–135.