Stochastic analysis of frequency-domain adaptive filters

Feiran Yang[1*]

*Correspondence:
feiran@mail.ioa.ac.cn

[1] State Key Laboratory
of Acoustics, Institute
of Acoustics, Chinese Academy
of Sciences, Beijing 100190,
China

Abstract

This study investigates the convergence behaviors of a family of frequency-domain adaptive filters (FDAFs) under both exact- and under-modeling situations. The stochastic analysis is conducted by transforming the frequency-domain equations into their time-domain counterparts. We discuss the transient and steady-state convergence behaviors of four FDAF versions, i.e., the constrained FDAFs with and without step-normalization, the unconstrained FDAFs with and without step-normalization, and we also present the upper bounds of step size for mean stability and mean-square stability. Starting from the expression for the steady-state mean weight vector, this study investigates whether the FDAFs can converge to unknown system impulse responses and optimum Wiener solutions. Moreover, we provide the closed-form minimum mean-square error (MMSE) that each FDAF can achieve. The difference between the current work and our previous one is threefold. First, the presented time-domain analysis is much easier to handle and has a more explicit physical meaning than that in the frequency domain. Second, we here consider an arbitrary overlap factor between consecutive blocks, while our previous analysis only focuses on 50% overlap. Third, the presented MMSE expressions and excess mean-square error (EMSE) approximations have not been given before. Simulations reveal high consistency between the experimental and theoretical results.

Keywords: Adaptive filter, Frequency domain, Mean-square error, Transient analysis, Convergence behavior

1 Introduction

Frequency-domain adaptive filter (FDAFs) is originally proposed as a fast but exact realization of the block least mean-square (BLMS) algorithm [1, 2], and thus, they exhibit the same convergence characteristics. Subsequently, two strategies are presented to improve the convergence or gain computational efficiency. The first method is known as the step-normalization or self-orthogonalization procedure, in which the frequency-wise step size related to the corresponding signal power is adopted [3]. The second strategy is to remove the constraint on the weight vector in the time domain, and the corresponding algorithm is called the unconstrained FDAF [4, 5]. Thus, four variants of the FDAF algorithm are obtained, i.e., the constrained FDAFs with and without step-normalization, the unconstrained FDAFs with and without

step-normalization, which have found applications in a wide range of areas [6–13]. Recently, deep neural networks have been incorporated into the estimation of the gradient or the key parameters of the FDAFs [14, 15].

Besides the algorithm design, it is desirable to characterize the convergence of the adaptive filtering algorithms [16–22, 22, 24–27]. In [18], the second-order statistics for the BLMS are analyzed for Gaussian inputs. In [19], the influence of the windowing function on the FDAF convergence is analyzed. In [20], the mean convergence performance of several FDAF versions is extensively analyzed via inverse transformation of frequency-domain formulas into time-domain counterparts, but the steady-state mean-square error (MSE) expressions are provided without detailed derivations. In [21], the optimum Wiener solution of FDAFs is given in the frequency domain, and it is shown that the unconstrained FDAFs achieve a reduced steady-state MSE compared to constrained algorithms under the under-modeling scenario. In [22], the analysis of FDAFs is conducted in the frequency domain, and the eigenvalue spread of the matrix that controls the mean convergence is well studied. The steady-state performance of constrained FDAFs is analyzed under the under-modeling scenario in [23], but the input is assumed to be Gaussian. In [24], a full second-order statistical framework for unconstrained FDAFs is presented for noncircular Gaussian signals. Recently, we have conducted an extensive statistical analysis of the FDAFs [25–27]. We derive a unifying update equation for four versions of FDAF, which enables us to conduct the performance analysis in a unifying framework. In [28] and [29], we provide a detailed performance assessment of the partitioned-block FDAFs in the time domain. However, the analysis results cannot be applied to the FDAFs straightforwardly.

We comprehensively study the convergence behavior of FDAF algorithms in this paper. The frequency-domain formulas are transformed into time-domain ones, which are utilized to deduce the evolution of the weight-error vector. We then analyze the mean convergence and mean-square convergence of the FDAFs in detail. This study goes further than our previous work [25–27] in several aspects. First, the presented analysis is conducted completely in the time domain, while our previous work is in the frequency domain. We have found that compared to the frequency-domain analysis, the time-domain one is easier to follow, and the time-domain variables have an explicit physical meaning. Second, our previous work only focuses on 50% overlap, but we deal with an arbitrary overlap here. Third, we provide the analytical expressions for the attainable minimum MSE (MMSE) of four variants of the FDAF. Also, we derive the excess MSE (EMSE) approximations in the time domain, and we point out that the EMSE of the constrained FDAF algorithms with step-normalization given in [20] and [22] is biased even for small step sizes. The presented stochastic model is built in the under-modeling scenario, but the exact modeling scenario can be treated as a special case. Simulations are presented to validate the theoretical results.

The contributions of this paper are as follows:

- We describe the transient and steady-state convergence behavior of a family of the FDAFs with an arbitrary overlap in both exact- and under-modeling scenarios.

- We present a comprehensive performance comparison of the constrained and unconstrained FDAFs in terms of the convergence rate, the modeling ability, the steady-state MSE, and the attainable minimum MSE.
- We derive the approximated expressions of steady-state EMSE of four types of the FDAFs in exact modeling scenario, which are easier to follow than before.

2 FDAF

We begin our treatment by introducing the FDAF algorithms. Consider the linear time-invariant model in the framework of system identification

$$d(n) = \mathbf{x}^T(n)\mathbf{w}_{\mathrm{opt}} + v(n), \tag{1}$$

where $d(n)$ represents the desired response, n is the discrete-time index, T represents a transposition, $\mathbf{x}(n) = [x(n), \cdots, x(n-M+1)]^T$ denotes the input vector, $\mathbf{w}_{\mathrm{opt}} = [w_0, \cdots, w_{M-1}]^T$ signifies the impulse response of an unknown system with a length of M, and $v(n)$ accounts for the measurement noise signal that is independent with $x(n)$. The input and noise signals are the zero-mean and stationary random processes, respectively, with a variance of $E[x^2(n)] = \sigma_x^2$ and $E[v^2(n)] = \sigma_v^2$ where $E[\cdot]$ is the expectation. The linear model in (1) has been applied to various problems despite its simplicity.

The FDAF algorithms typically rely on a block-processing approach, i.e., the filtering out and the adaptive weights are calculated block by block. The weight vector remains the same in one block in FDAF algorithms, which is different from the sample-by-sample-based least mean-square (LMS) algorithm. In this paper, R and k denote the block length and the block index, respectively. We consider an adaptive transversal filter $\hat{\mathbf{w}}(k) = [\hat{w}_0(k), \cdots, \hat{w}_{L-1}(k)]^T$ that tries to imitate the unknown system response $\mathbf{w}_{\mathrm{opt}}$, where L is the adaptive filter's length. The relations $L = M$ and $L < M$ correspond to the exact and under-modeling situations, respectively. In this paper, we consider the under-modeling situation with $Q = M - L$ denoting the length of the under-modeling filter. However, we treat the exact modeling situation as a special case of the under-modeling situation, i.e., $Q = 0$, and hence, the presented analysis can cover both the exact and under-modeling situations.

We pad the estimated weight vector $\hat{\mathbf{W}}(k)$ with R zeros and obtain its frequency-domain representation

$$\hat{\mathbf{W}}(k) = \mathbf{F}\begin{bmatrix} \hat{\mathbf{w}}(k) \\ \mathbf{0}_{R\times 1} \end{bmatrix} = \mathbf{F}\Upsilon_{10}^T\hat{\mathbf{w}}(k), \tag{2}$$

where $N = L + R$ represents the DFT length, $\mathbf{F}$ denotes the $N \times N$ DFT matrix, $\Upsilon_{10} = [\mathbf{I}_L\ \mathbf{0}_{L\times R}]$, $\mathbf{I}_L$ represents an identity matrix with a size of $L \times L$, $\mathbf{0}_{L\times R}$ denotes a zero matrix with a size of $L \times R$, and $\mathbf{0}_{R\times 1}$ stands for the all-zero vector of length R. In our previous work [23–25], we focused on the special case with 50% overlap, i.e., $L = R$, while we now consider a general case for an arbitrary overlap.

The frequency-domain diagonal matrix $\boldsymbol{\mathcal{X}}(k)$ can be attained by converting an input signal block into the frequency domain

$$\boldsymbol{\mathcal{X}}(k) = \mathrm{diag}\{\mathbf{Fx}(k)\}, \tag{3}$$

where $\mathbf{x}(k) = [x(kR-L), \cdots, x(kR+R-1)]^T$ is the input vector with a length of N, and $\mathrm{diag}\{\cdot\}$ forms a diagonal matrix from inputs. This input vector contains R elements in the current block and L elements in the previous one.

The time-domain block error signal vector is expressed by [5]

$$\mathbf{e}(k) = \mathbf{d}(k) - \hat{\mathbf{y}}(k) = \mathbf{d}(k) - \Upsilon_{01}\mathbf{F}^{-1}\boldsymbol{\mathcal{X}}(k)\hat{\mathbf{W}}(k), \tag{4}$$

where $\mathbf{d}(k) = [d(kR), \cdots, d(kR+R-1)]^T$ is the block desired signal vector with a length of R, the block filtering out vector $\hat{\mathbf{y}}(k)$ and the block error vector $\mathbf{e}(k)$ are defined similarly, and the $R \times N$ matrix $\Upsilon_{01} = [\mathbf{0}_{R \times L}\ \mathbf{I}_{R \times R}]$ is utilized to retain the last R terms of the inverse DFT of $\boldsymbol{\mathcal{X}}(k)\hat{\mathbf{W}}(k)$ that corresponds exactly to the linear convolution. The time-domain error vector $\mathbf{e}(k)$ is initially padded with L zeros and then turned into the frequency-domain expression

$$\mathbf{E}(k) = \mathbf{F}\begin{bmatrix} \mathbf{0}_{L \times 1} \\ \mathbf{e}(k) \end{bmatrix} = \mathbf{F}\Upsilon_{01}^T\mathbf{e}(k). \tag{5}$$

The constrained FDAF is characterized by [22]

$$\hat{\mathbf{W}}(k+1) = \mathbf{G}_{10}[\hat{\mathbf{W}}(k) + \mu\boldsymbol{\Lambda}^{-1}\boldsymbol{\mathcal{X}}^H(k)\mathbf{E}(k)], \tag{6}$$

where superscript H means a complex-conjugate transpose, the constraining matrix $\mathbf{G}_{10} = \mathbf{F}\Upsilon_{10}^T\Upsilon_{10}\mathbf{F}^{-1}$ forces the last R components of the time-domain gradient $\mathbf{F}^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mathcal{X}}^H(k)\mathbf{E}(k)$ to zero, and the $N \times N$ diagonal matrix $\boldsymbol{\Lambda}$ determines how the step size is chosen. If we use $\boldsymbol{\Lambda} = \mathbf{I}_N$, a common step size parameter μ is then employed for all the frequencies, namely, the FDAF algorithm without step-normalization. This algorithm is nothing but a faithful frequency-domain representation of the BLMS algorithm. If we choose the diagonal matrix $\boldsymbol{\Lambda} = E[\boldsymbol{\mathcal{X}}^H(k)\boldsymbol{\mathcal{X}}(k)]$, the step size in each frequency bin is scaled by the power spectral density (PSD). The algorithm is referred to as the FDAF with step-normalization, which can greatly speed up the convergence [3–5].

The update equation of the unconstrained FDAF is [3]

$$\hat{\mathbf{W}}(k+1) = \hat{\mathbf{W}}(k) + \mu\boldsymbol{\Lambda}^{-1}\boldsymbol{\mathcal{X}}^H(k)\mathbf{E}(k). \tag{7}$$

Similarly, we obtain the unconstrained FDAF algorithm with or without step-normalization by using $\boldsymbol{\Lambda} = E[\boldsymbol{\mathcal{X}}^H(k)\boldsymbol{\mathcal{X}}(k)]$ or $\boldsymbol{\Lambda} = \mathbf{I}_N$. Unlike constrained FDAF algorithms, the gradient constraint matrix $\mathbf{G}_{10}$ is dropped out from (7), and hence (2) does not hold, which makes (4) implement a circular convolution for the unconstrained FDAF [30]. However, dropping the gradient constraint may have advantages for the under-modeling case, as the discussion will reveal.

3 Analysis of unconstrained FDAFs

This section analyzes the time-domain convergence behavior of unconstrained FDAF algorithms. The frequency-domain Eqs. (4) and (7) are changed to the time-domain representations, which are in turn used to evaluate the mean and mean-square performance and establish the stability bound.

3.1 Signal model

To pave the way for the performance analysis, we present an important result of the circulant matrix [4]

$$\boldsymbol{\mathcal{X}}^H(k) = \mathbf{F}\mathbf{X}_c^T(k)\mathbf{F}^{-1}, \boldsymbol{\mathcal{X}}(k) = \mathbf{F}\mathbf{X}_c(k)\mathbf{F}^{-1}, \tag{8}$$

where $\mathbf{X}_c(k)$ represents a circulant matrix with a size of $N \times N$ and the first column of $\mathbf{x}(k)$. We can decompose $\mathbf{X}_c(k)$ as

$$\mathbf{X}_c(k) = \begin{bmatrix} \hat{\mathbf{X}}_1(k) & \hat{\mathbf{X}}_2(k) \\ \mathbf{X}_1(k) & \mathbf{X}_2(k) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{X}}(k) \\ \bar{\mathbf{X}}(k) \end{bmatrix}, \tag{9}$$

where the matrices $\mathbf{X}_1(k), \mathbf{X}_2(k), \hat{\mathbf{X}}_1(k)$, and $\hat{\mathbf{X}}_2(k)$ have a size of $R \times L$, $R \times R$, $L \times L$, and $R \times R$, respectively, and the matrices $\bar{\mathbf{X}}(k) = [\mathbf{X}_1(k)\ \mathbf{X}_2(k)]$ and $\hat{\mathbf{X}}(k) = [\hat{\mathbf{X}}_1(k)\hat{\mathbf{X}}_2(k)]$ have a dimension of $R \times N$ and $L \times N$. For convenience, we define the matrix

$$\mathbf{P} = (\mathbf{F}^{-1}\boldsymbol{\Lambda}\mathbf{F})^{-1} = \mathbf{R}_c^{-1}, \tag{10}$$

where $\mathbf{R}_c = E[\mathbf{X}_c^T(k)\mathbf{X}_c(k)]$. Because $\mathbf{X}_c(k)$ is circulant, the matrices $\mathbf{R}_c$ and $\mathbf{P}$ are both circulant and symmetric. We then represent $\mathbf{P}$ in the block matrix form

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_2^T & \mathbf{P}_3 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{P}} \\ \hat{\mathbf{P}} \end{bmatrix}, \tag{11}$$

where $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ are $L \times L$, $L \times R$, and $R \times R$ submatrices, respectively, and $\bar{\mathbf{P}} = [\mathbf{P}_1\ \mathbf{P}_2]$ and $\hat{\mathbf{P}} = [\mathbf{P}_2^T\ \mathbf{P}_3]$ have a dimension of $L \times N$ and $R \times N$.

For the unconstrained algorithm, we define the time-domain weight vector $\hat{\mathbf{w}}_{un}(k) = \mathbf{F}^{-1}\hat{\mathbf{W}}(k) \triangleq [\hat{\mathbf{w}}_{un,L}^T(k), \hat{\mathbf{w}}_{un,R}^T(k)]^T$, where the subvectors $\hat{\mathbf{w}}_{un,L}(k)$ and $\hat{\mathbf{w}}_{un,R}(k)$ are of length L and R, respectively. Note that $\hat{\mathbf{w}}_{un}(k)$ has a length of N and the last R elements of $\hat{\mathbf{w}}_{un}(k)$ may be nonzero since the gradient constraint is removed. Using (8), we represent the block error vector as

$$\mathbf{e}(k) = \mathbf{d}(k) - \bar{\mathbf{X}}(k)\hat{\mathbf{w}}_{un}(k) \tag{12}$$

Calculating the filtering out signal may require future inputs due to the existence of $\hat{\mathbf{w}}_{un,R}(k)$ [20], and we hence call $\hat{\mathbf{w}}_{un,L}(k)$ and $\hat{\mathbf{w}}_{un,R}(k)$ the non-causal and causal parts, respectively.

Pre-multiplying both sides of (7) by $\mathbf{F}^{-1}$ and considering (8) and (10), we obtain the recursion of $\hat{\mathbf{w}}_{un}(k)$

$$\hat{\mathbf{w}}_{un}(k+1) = \hat{\mathbf{w}}_{un}(k) + \mu\mathbf{P}\bar{\mathbf{X}}^T(k)\mathbf{e}(k). \tag{13}$$

The true weight vector $\mathbf{w}_{opt}$ that we wish to predict can be split into two parts

$$\mathbf{w}_{opt} = \begin{bmatrix} \mathbf{w}_\dagger \\ \mathbf{w}_* \end{bmatrix}, \tag{14}$$

where the column vectors $\mathbf{w}_\dagger$ and $\mathbf{w}_*$ have a length of L and Q. The desired response vector $\mathbf{d}(k)$ can be expressed by

$$\mathbf{d}(k) = \bar{\mathbf{X}}(k)\bar{\mathbf{w}} + \mathbf{X}_3(k)\mathbf{w}_* + \mathbf{v}(k), \tag{15}$$

where $\mathbf{v}(k) = [\nu(kR), \cdots, \nu(kR + R - 1)]^T$ denotes the noise vector of length R, $\mathbf{X}_3(k) = [\mathbf{x}_3(kR - L) \cdots \mathbf{x}_3(kR - L + R - 1)]^T$ represents an $R \times Q$ input matrix with $\mathbf{x}_3(n) = [x(n), \cdots, x(n - Q + 1)]^T$, and $\bar{\mathbf{w}} = [\mathbf{w}_\dagger^T, \mathbf{0}_{1 \times R}]^T$ has a length of N.

3.2 Mean convergence

This section investigates initially the mean behavior of the unconstrained FDAF algorithms. For that, we introduce the time-domain weight-error vector $\tilde{\mathbf{w}}_{\mathrm{un}}(k) = \bar{\mathbf{w}} - \hat{\mathbf{w}}_{\mathrm{un}}(k) \overset{\Delta}{=} [\tilde{\mathbf{w}}_{\mathrm{un},L}^T(k), \tilde{\mathbf{w}}_{\mathrm{un},R}^T(k)]^T$, where $\tilde{\mathbf{w}}_{\mathrm{un},L}(k) = \mathbf{w}_\dagger - \hat{\mathbf{w}}_{\mathrm{un},L}(k)$ and $\tilde{\mathbf{w}}_{\mathrm{un},R}(k) = -\hat{\mathbf{w}}_{\mathrm{un},R}(k)$. We subtract $\bar{\mathbf{w}}$ from both sides of (13) and derive the update equation of $\tilde{\mathbf{w}}_{\mathrm{un}}(k)$

$$\tilde{\mathbf{w}}_{\mathrm{un}}(k + 1) = \tilde{\mathbf{w}}_{\mathrm{un}}(k) - \mu \mathbf{P}\bar{\mathbf{X}}^T(k)\mathbf{e}(k). \tag{16}$$

We incorporate (15) with (12) and have

$$\mathbf{e}(k) = \bar{\mathbf{X}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}(k) + \mathbf{X}_3(k)\mathbf{w}_* + \mathbf{v}(k). \tag{17}$$

We then substitute (17) into (16) and obtain

$$\tilde{\mathbf{w}}_{\mathrm{un}}(k + 1) = \big(\mathbf{I}_N - \mu \mathbf{A}_{\mathrm{un}}(k)\big)\tilde{\mathbf{w}}_{\mathrm{un}}(k) - \mu \mathbf{B}_{\mathrm{un}}(k)\mathbf{w}_* - \mu \mathbf{P}\bar{\mathbf{X}}^T(k)\mathbf{v}(k). \tag{18}$$

where $\mathbf{A}_{\mathrm{un}}(k) = \mathbf{P}\bar{\mathbf{X}}^T(k)\bar{\mathbf{X}}(k)$ and $\mathbf{B}_{\mathrm{un}}(k) = \mathbf{P}\bar{\mathbf{X}}^T(k)\mathbf{X}_3(k)$.

Taking the expectation of each side of (18) and invoking the independence assumption [4] yield

$$E[\tilde{\mathbf{w}}_{\mathrm{un}}(k + 1)] = \big(\mathbf{I}_N - \mu E[\mathbf{A}_{\mathrm{un}}(k)]\big)E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)] - \mu E[\mathbf{B}_{\mathrm{un}}(k)]\mathbf{w}_*, \tag{19}$$

where $E[\mathbf{A}_{\mathrm{un}}(k)] = \mathbf{P}\bar{\mathbf{R}}$ and $\bar{\mathbf{R}} = E[\bar{\mathbf{X}}^T(k)\bar{\mathbf{X}}(k)]$. The condition number of $E[\mathbf{A}_{\mathrm{un}}(k)]$ has a significant influence on the convergence speed of the unconstrained FDAFs. It is shown in [20] that the approximation $\mathbf{R}_{\mathrm{c}} \approx \frac{N}{R}\bar{\mathbf{R}}$ holds for a large N. For unconstrained FDAFs with step-normalization, the matrix controlling the mean convergence is thus proportional to the identity matrix, $E[\mathbf{A}_{\mathrm{un}}(k)] \approx \frac{R}{N}\mathbf{I}_N$. It turns out that the unconstrained FDAF with step-normalization only has one mode of convergence thanks to the self-orthogonalizing method, while that without step-normalization may have a rather slow convergence speed, particularly for a highly correlated input.

At steady state, (19) leads to

$$E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)] = -\bar{\mathbf{R}}^{-1}\bar{\mathbf{R}}_3\mathbf{w}_* \tag{20}$$

where $\bar{\mathbf{R}}_3 = E[\bar{\mathbf{X}}^T(k)\mathbf{X}_3(k)]$. We now investigate whether the unconstrained FDAFs converge to the Wiener solution using (20). Considering (17) and replacing $\tilde{\mathbf{w}}_{\mathrm{un}}(k)$ with a time-invariant vector, the Wiener solution is given by minimizing the MSE $E[\|\mathbf{e}(k)\|^2]$

$$\tilde{\mathbf{w}}_{\mathrm{un,opt}} = -\bar{\mathbf{R}}^{-1}\bar{\mathbf{R}}_3\mathbf{w}_*. \tag{21}$$

In the exact modeling situation, it has $E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)] = \tilde{\mathbf{w}}_{\mathrm{un,opt}} = \mathbf{0}_{N \times 1}$, and hence, the unconstrained FDAF algorithm converges to the optimum Wiener solution that is also

the true system impulse response. The non-causal part can converge to zero in mean, i.e., $E[\tilde{\mathbf{w}}_{\mathrm{un},R}(\infty)] = \mathbf{0}_{R \times 1}$, and hence, (4) implements an (approximate) linear convolution for sufficiently small step sizes.

In the under-modeling situation, it is observed from (20) and (21) that the unconstrained FDAF converges to the optimum Wiener solution in the mean sense regardless of input signal characteristics. In this case, it has $\bar{\mathbf{R}}_3 \neq \mathbf{0}_{N \times Q}$ for any inputs, and hence $E[\hat{\mathbf{w}}_{\mathrm{un}}(k)] = \bar{\mathbf{w}} - E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)] = \bar{\mathbf{w}} + \bar{\mathbf{R}}^{-1}\bar{\mathbf{R}}_3\mathbf{w}_* \neq \bar{\mathbf{w}}$, which indicates that the unconstrained FDAF algorithm approaches a biased estimate of the first L elements of the unknown impulse response. However, we will immediately show that we can indeed obtain more information about the coefficients of the unknown system from unconstrained FDAFs, which is not revealed in previous publications.

To proceed, we define $\bar{\mathbf{R}}^{-1}\bar{\mathbf{R}}_3 = [\boldsymbol{\beta}_1^T \, \boldsymbol{\beta}_2^T]^T$, where the matrices $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are of size $L \times Q$ and $R \times Q$, respectively. We then rewrite (20) as

$$E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)] = - \begin{bmatrix} \boldsymbol{\beta}_1\mathbf{w}_* \\ \boldsymbol{\beta}_2\mathbf{w}_* \end{bmatrix} = \begin{bmatrix} \mathbf{w}_\dagger - E[\hat{\mathbf{w}}_{\mathrm{un},L}(\infty)] \\ -E[\hat{\mathbf{w}}_{\mathrm{un},R}(\infty)] \end{bmatrix}. \tag{22}$$

We define $\boldsymbol{\beta}_2 = [\boldsymbol{\beta}_3 \, \boldsymbol{\beta}_4]$, where the matrices $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ are of size $R \times R$ and $R \times (Q - R)$. We then split $\mathbf{w}_*$ into two parts $\mathbf{w}_* = [\mathbf{w}_{*1}^T \, \mathbf{w}_{*2}^T]^T$, where lengths of the column vectors $\mathbf{w}_{*1}$ and $\mathbf{w}_{*2}$ are R and $Q - R$, respectively.

For white noise as input, we have $\boldsymbol{\beta}_1 = \mathbf{0}_{L \times Q}$, $\boldsymbol{\beta}_4 = \mathbf{0}_{R \times (Q-R)}$, and $\boldsymbol{\beta}_3 = \frac{1}{R}\mathrm{diag}\{[R, R-1, \cdots, 1]^T\}$. Using (22), it has

$$\mathbf{w}_\dagger = E[\hat{\mathbf{w}}_{\mathrm{un},L}(\infty)], \tag{23}$$

$$\mathbf{w}_{*1} = \boldsymbol{\beta}_3^{-1}E[\hat{\mathbf{w}}_{\mathrm{un},R}(\infty)]. \tag{24}$$

From (23) and (24), we observe that $\hat{\mathbf{w}}_{\mathrm{un},L}(k)$ can converge to the first L elements of the unknown plant in mean, and the first element of $\hat{\mathbf{w}}_{\mathrm{un},R}(k)$ is equal to w_L. The other $R-1$ components cannot approach to any part of the unknown plant. That is, the first $L+1$ coefficients of $E[\hat{\mathbf{w}}_{\mathrm{un}}(k)]$ converge to the first $L+1$ coefficients of the unknown plant. As a consequence, the unconstrained FDAF algorithm cannot directly model an unknown system having more than $L+1$ coefficients.[1] However, we can calculate $\mathbf{w}_{*1}$ using (24), and hence, a total of $L+R$ coefficients of the unknown system can be restored from the steady-state vector $E[\hat{\mathbf{w}}_{\mathrm{un}}(\infty)]$ of the unconstrained FDAF. This result holds without any restriction on the length of the system impulse response.

For correlated inputs, we have $\boldsymbol{\beta}_1 \neq \mathbf{0}_{L \times Q}$, and hence $E[\hat{\mathbf{w}}_{\mathrm{un},L}(\infty)] \neq \mathbf{w}_\dagger$, which indicates that the causal part of $\hat{\mathbf{w}}_{\mathrm{un}}(k)$ cannot converge to the first L coefficients of the unknown system in mean. When $Q > R$, we should solve $\boldsymbol{\beta}_2\mathbf{w}_* = E[\hat{\mathbf{w}}_{\mathrm{un},R}(\infty)]$ to obtain $\mathbf{w}_*$, which is however an underdetermined linear system of equations and a precise solution

---

[1] Only several papers have discussed the modeling capability of the unconstrained FDAF algorithm with a white noise input. In [20], it is observed from experiment that the first L weights of the unconstrained and constrained FDAFs are about the same in under-modeling case. In [30], it is pointed out that the unconstrained FDAF algorithm cannot accurately model an unknown system with greater than L coefficients. In our previous work [26], it is shown that the first L elements of $E[\hat{\mathbf{w}}_{\mathrm{un}}(\infty)]$ equal those of the true system response, but we did not point out that the $L+1$-th element of $E[\hat{\mathbf{w}}_{\mathrm{un}}(\infty)]$ is equal to that of true weight vector. The aforementioned discussion is limited to the case $N = L + R$. It is easy to verify that the unconstrained FDAF algorithm can directly model an unknown plant with $N - R + 1$ coefficients for any $N \geq L + R - 1$ for white noise input and more coefficients can be recovered using (22).

is not available. For $Q \leq R$, we can obtain $\mathbf{w}_*$ by solving the (over)-determined equation $\boldsymbol{\beta}_2 \mathbf{w}_* = E[\hat{\mathbf{w}}_{\mathrm{un},R}(\infty)]$ and then, we estimate the modeling part $\mathbf{w}_\dagger$ as

$$\mathbf{w}_\dagger = E[\hat{\mathbf{w}}_{\mathrm{un},L}(\infty)] - \boldsymbol{\beta}_1 \mathbf{w}_*. \tag{25}$$

Hence, all the coefficients of unknown plant can be recovered in this case.

Also, notice that the Wiener solution in the under-modeling case is not the first L coefficients of the system impulse response. In echo cancelation, we aim at canceling the echo signal and hence, we expect that the filter converges to the Wiener solution and an accurate estimation of the echo path is not the main concern. In applications of room impulse response identification, on the contrary, a precise estimation of the room impulse response is the objective. We can thus choose the proper variant of FDAFs according to the task at hand.

3.3 Mean-square convergence

This section studies the second-order convergence of the unconstrained FDAF. Using (17), we formulate the instantaneous MSE as

$$\begin{aligned}
J_{\mathrm{un}}(k) &= \frac{1}{R} E[\mathbf{e}^T(k)\mathbf{e}(k)] \\
&= \frac{1}{R}\Big\{ \mathrm{tr}(E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)]\bar{\mathbf{R}}) + 2E[\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)]\bar{\mathbf{R}}_3\mathbf{w}_* + \mathbf{w}_*^T\mathbf{R}_3\mathbf{w}_* \Big\} + \sigma_\nu^2,
\end{aligned} \tag{26}$$

where $\mathbf{R}_3 = E[\mathbf{X}_3^T(k)\mathbf{X}_3(k)]$, $J_{\mathrm{un,min}} = \sigma_\nu^2$ is the MMSE which can be achieved only in the exact modeling case, i.e., $M = L$, and $J_{\mathrm{un,ex}}(k) = \frac{1}{R}\big\{ \mathrm{tr}(E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)]\bar{\mathbf{R}})$ $+2E[\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)]\bar{\mathbf{R}}_3\mathbf{w}_* + \mathbf{w}_*^T\mathbf{R}_3\mathbf{w}_* \big\}$ is the EMSE. Note that in the under-modeling case, the MMSE is not $J_{\mathrm{un,min}} = \sigma_\nu^2$ given a fixed L as will be shown later. However, we still use the provided EMSE expression for convenience, which is reasonable if we treat the adaptive filter length as an independent variable.

Note that the calculation of the MSE $J_{\mathrm{un}}(k)$ requires evaluating $E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)]$ and $E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)]$. Since the former has been presented in (19), we now examine the evolution of $E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)]$. By post-multiplying (18) by its transpose and calculating its expectation, it holds that

$$\begin{aligned}
&E[\tilde{\mathbf{w}}_{\mathrm{un}}(k+1)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k+1)] \\
&\quad = E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)] \\
&\qquad - \mu E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)\mathbf{A}_{\mathrm{un}}^T(k)] \\
&\qquad - \mu E[\mathbf{A}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)] \\
&\qquad + \mu^2 E[\mathbf{A}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}^T\mathbf{A}_{\mathrm{un}}^T(k)] \\
&\qquad - \mu E[\tilde{\mathbf{w}}_{\mathrm{un}}(k)\mathbf{w}_*^T\mathbf{B}_{\mathrm{un}}^T(k)] \\
&\qquad - \mu E[\mathbf{B}_{\mathrm{un}}(k)\mathbf{w}_*\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)] \\
&\qquad + \mu^2 E[\mathbf{B}_{\mathrm{un}}(k)\mathbf{w}_*\tilde{\mathbf{w}}_{\mathrm{un}}^T(k)\mathbf{A}_{\mathrm{un}}^T(k)] \\
&\qquad + \mu^2 E[\mathbf{B}_{\mathrm{un}}(k)\mathbf{w}_*\mathbf{w}_*^T\mathbf{B}_{\mathrm{un}}^T(k)] \\
&\qquad + \mu^2 E[\mathbf{A}_{\mathrm{un}}(k)\tilde{\mathbf{w}}_{\mathrm{un}}(k)\mathbf{w}_*^T\mathbf{B}_{\mathrm{un}}^T(k)] \\
&\qquad + \mu^2 \delta_\nu^2 \mathbf{P}\mathbf{R}\bar{\mathbf{P}}.
\end{aligned} \tag{27}$$

We introduce the vector $\mathbf{z}_{un}(k) = \text{vec}\{E[\tilde{\mathbf{w}}_{un}(k)\tilde{\mathbf{w}}_{un}^T(k)]\}$ of length N^2, where $\text{vec}\{\cdot\}$ obtains column vectors by stacking all columns of the matrix argument [31]. Applying vectorization to (27), we set up a difference equation for $\mathbf{z}_{un}(k)$

$$\mathbf{z}_{un}(k+1) = \mathbf{H}_{un}\mathbf{z}_{un}(k) + \mathbf{\Theta}_{un}(k), \tag{28}$$

where

$$
\begin{aligned}
\mathbf{H}_{un} &= \mathbf{I}_{N^2} - \mu\mathbf{C}_{un} + \mu^2\mathbf{J}_{un}, \\
\mathbf{C}_{un} &= E[\mathbf{A}_{un}(k)] \otimes \mathbf{I}_N + \mathbf{I}_N \otimes E[\mathbf{A}_{un}(k)], \\
\mathbf{J}_{un} &= E[\mathbf{A}_{un}(k) \otimes \mathbf{A}_{un}(k)],
\end{aligned}
\tag{29}
$$

$$
\begin{aligned}
\mathbf{\Theta}_{un}(k) = {} & \mu^2\delta_\nu^2\text{vec}(\mathbf{P}\bar{\mathbf{R}}\mathbf{P}) - \mu(E[\mathbf{B}_{un}(k)] \otimes \mathbf{I}_N)\text{vec}(E[\tilde{\mathbf{w}}_{un}(k)]\mathbf{w}_*^T) \\
& + \mu^2(E[\mathbf{B}_{un}(k)] \otimes \mathbf{A}_{un}(k))\text{vec}(E[\tilde{\mathbf{w}}_{un}(k)]\mathbf{w}_*^T) \\
& - \mu(\mathbf{I}_N \otimes E[\mathbf{B}_{un}(k)])\text{vec}(\mathbf{w}_*E[\tilde{\mathbf{w}}_{un}^T(k)]) \\
& + \mu^2(E[\mathbf{A}_{un}(k)] \otimes \mathbf{B}_{un}(k))\text{vec}(\mathbf{w}_*E[\tilde{\mathbf{w}}_{un}^T(k)]) \\
& + \mu^2(E[\mathbf{B}_{un}(k)] \otimes \mathbf{B}_{un}(k))\text{vec}(\mathbf{w}_*\mathbf{w}_*^T).
\end{aligned}
\tag{30}
$$

The mean-square convergence behavior of the unconstrained FDAF is governed by the eigenvalues of $\mathbf{H}_{un}$. It should be mentioned that (28) does not resort to the Gaussian input assumption, and hence, the analysis in this paper is valid for arbitrary data distributions. We will use the same procedure to analyze the constrained FDAFs in the next section. In [32], the energy-conservation approach is used to derive the time evolution of the MSD and EMSE. Indeed, we can use the energy-conservation approach for analysis of the FDAFs and can obtain the same results. However, the method used here is easier to follow and understand than the energy-conservation approach.

The state-space model (30) depicts the mean-square behavior of the unconstrained FDAF. The learning curve can be evaluated by iterating recursion (30). The mean-square deviation (MSD), i.e., the system distance, is commonly used to measure the Euclidean distance between the estimated and true values. The instantaneous MSD of the unconstrained FDAFs is calculated by $\delta_{un}(k) = E(\|\tilde{\mathbf{w}}_{un}(k)\|^2) = \text{tr}(\text{vec}^{-1}(\mathbf{z}_{un}(k)))$. Notice that only the first L weights are involved in the MSD evaluation of the unknown system. At steady state, (28) leads to $\mathbf{z}_{un}(\infty) = (\mathbf{I}_{N^2} - \mathbf{H}_{un})^{-1}\mathbf{\Theta}_{un}(\infty)$, and hence, the steady-state MSD can be expressed by $\delta_{un}(k) = E(\|\tilde{\mathbf{w}}_{un}(\infty)\|^2) = \text{tr}(\text{vec}^{-1}(\mathbf{z}_{un}(\infty)))$. The steady-state EMSE then follows

$$J_{un,ex}(\infty) = \frac{1}{R}\left\{\text{tr}(E[\tilde{\mathbf{w}}_{un}(\infty)\tilde{\mathbf{w}}_{un}^T(\infty)]\bar{\mathbf{R}}) + 2E[\tilde{\mathbf{w}}_{un}^T(\infty)]\bar{\mathbf{R}}_3\mathbf{w}_* + \mathbf{w}_*^T\mathbf{R}_3\mathbf{w}_*\right\}. \tag{31}$$

At this point, we would like to identify the factors that influence the steady-state MSE of the deficient-length unconstrained FDAF, where we have assumed that M and L are fixed. For that, we define the difference between $E[\tilde{\mathbf{w}}_{un}(\infty)]$ and the optimum Wiener solution as $\mathbf{\Delta} = \tilde{\mathbf{w}}_{un,opt} - E[\tilde{\mathbf{w}}_{un}(\infty)]$. We then write the difference between $\hat{\mathbf{w}}_{un}(k)$ and its mean steady-state value $E[\hat{\mathbf{w}}_{un}(\infty)]$ as

$$\mathbf{q}_{un}(k) = \hat{\mathbf{w}}_{un}(k) - E[\hat{\mathbf{w}}_{un}(\infty)] = E[\tilde{\mathbf{w}}_{un}(\infty)] - \tilde{\mathbf{w}}_{un}(k) = \tilde{\mathbf{w}}_{un,opt} - \mathbf{\Delta} - \tilde{\mathbf{w}}_{un}(k). \tag{32}$$

Substituting (32) into (26), we may represent the steady-state MSE in terms of $\mathbf{q}_{\text{un}}(k)$

$$
\begin{aligned}
J_{\text{un}}(\infty) = {} & \frac{1}{R}\text{tr}(E[\mathbf{q}_{\text{un}}(\infty)\mathbf{q}_{\text{un}}^T(\infty)]\bar{\mathbf{R}}) \\
& + \frac{1}{R}\left\{\text{tr}(E\boldsymbol{\Delta}\boldsymbol{\Delta}^T\bar{\mathbf{R}}) - 2\text{tr}(\boldsymbol{\Delta}\tilde{\boldsymbol{w}}_{\text{un,opt}}^T\bar{\mathbf{R}}) + 2\boldsymbol{\Delta}^T\bar{\mathbf{R}}_3\mathbf{w}_*\right\} \\
& + \frac{1}{R}\left\{\text{tr}(\tilde{\mathbf{w}}_{\text{un,opt}}\tilde{\mathbf{w}}_{\text{un,opt}}^T\bar{\mathbf{R}}) + 2\tilde{\mathbf{w}}_{\text{un,opt}}^T\bar{\mathbf{R}}_3\mathbf{w}_* + \mathbf{w}_*^T\mathbf{R}_3\mathbf{w}_*\right\} + \sigma_\nu^2.
\end{aligned}
\tag{33}
$$

The first term on the right-hand side of (33) is directly related to $\mathbf{q}_{\text{un}}(\infty)$, i.e., the fluctuations of $\tilde{\mathbf{w}}_{\text{un}}(\infty)$ around its mean $E[\tilde{\mathbf{w}}_{\text{un}}(\infty)]$, and the second term is introduced by the bias $\boldsymbol{\Delta}$ between $E[\tilde{\mathbf{w}}_{\text{un}}(\infty)]$ and its optimal solution $\tilde{\mathbf{w}}_{\text{un,opt}}$. Because the two kinds of unconstrained FDAFs can converge to the optimal Wiener solution, i.e., $\boldsymbol{\Delta} = \mathbf{0}_{N\times 1}$, the MMSE of the unconstrained FDAFs with and without step-normalization can be expressed by

$$
J_{\min,1} = J_{\min,2} = \frac{1}{R}\left\{\text{tr}(\tilde{\mathbf{w}}_{\text{un,opt}}\tilde{\mathbf{w}}_{\text{un,opt}}^T\bar{\mathbf{R}}) + 2\tilde{\mathbf{w}}_{\text{un,opt}}^T\bar{\mathbf{R}}_3\mathbf{w}_* + \mathbf{w}_*^T\mathbf{R}_3\mathbf{w}_*\right\} + \delta_\nu^2.
\tag{34}
$$

3.4 Stability bound

We investigate the stability condition for the unconstrained FDAFs. Note that (19) is convergent if all eigenvalues of $\mathbf{I}_N - \mu E[\mathbf{A}_{\text{un}}(k)]$ are within the unit circle, i.e., $\rho(\mathbf{I}_N - \mu E[\mathbf{A}_{\text{un}}(k)]) < 1$ with $\rho(\cdot)$ denoting the spectral radius. The matrices $\mathbf{P}$ and $\bar{\mathbf{R}}$ are positive definite, so the eigenvalues of $E[\mathbf{A}_{\text{un}}(k)] = \mathbf{P}\bar{\mathbf{R}}$ are positive and real [31]. The condition $\rho(\mathbf{I}_N - \mu E[\mathbf{A}_{\text{un}}(k)]) < 1$ that guarantees the unconstrained FDAF's mean convergence is then equivalent to

$$
\mu < \frac{2}{\lambda_{\max}(E[\mathbf{A}_{\text{un}}(k)])},
\tag{35}
$$

where $\lambda_{\max}(\cdot)$ represents the maximum eigenvalue of the input matrix.

In light of (28), one observes that for the unconstrained FDAF, the mean-square convergence holds if and only if $\rho(\mathbf{H}_{\text{un}}) < 1$ holds. Because $\mathbf{C}_{\text{un}}$ may not be positive definite particularly in the step-normalization version, the method in [32] cannot be directly applicable to solve this inequality. We resort to the approach in [33] to deal with this problem. We decompose $\mathbf{P}$ as $\mathbf{P} = \mathbf{Q}\mathbf{S}\mathbf{Q}^T$, where the diagonal matrix $\mathbf{S}$ consists of the eigenvalues of $\mathbf{P}$ and the column vectors of the orthogonal matrix $\mathbf{Q}$ are the corresponding eigenvectors. We introduce a matrix $\boldsymbol{\alpha} = \mathbf{Q}\mathbf{S}^{\frac{1}{2}}\mathbf{Q}^T$. Pre- and post-multiplying (29) by $(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})^{-1}$ and $\boldsymbol{\alpha}\otimes\boldsymbol{\alpha}$, we arrive at

$$
\bar{\mathbf{H}}_{\text{un}} = \mathbf{I}_{N^2} - \mu\bar{\mathbf{C}}_{\text{un}} + \mu^2\bar{\mathbf{J}}_{\text{un}},
\tag{36}
$$

where $\quad\bar{\mathbf{H}}_{\text{un}} = (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})^{-1}\mathbf{H}_{\text{un}}(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha}),\quad \bar{\mathbf{C}}_{\text{un}} = (\boldsymbol{\alpha}\bar{\mathbf{R}}\boldsymbol{\alpha})\otimes\mathbf{I}_N + \mathbf{I}_N\otimes(\boldsymbol{\alpha}\bar{\mathbf{R}}\boldsymbol{\alpha}),\quad$ and $\bar{\mathbf{J}}_{\text{un}} = E[(\boldsymbol{\alpha}\bar{\mathbf{X}}^T(k)\bar{\mathbf{X}}(k)\boldsymbol{\alpha})\otimes(\boldsymbol{\alpha}\bar{\mathbf{X}}^T(k)\bar{\mathbf{X}}(k)\boldsymbol{\alpha})]$. The matrices $\mathbf{H}_{\text{un}}$ and $\bar{\mathbf{H}}_{\text{un}}$ have the same eigenvalues, which means that we can solve $\rho(\bar{\mathbf{H}}_{\text{un}}) < 1$ instead. We can easily infer that $\bar{\mathbf{C}}_{\text{un}}$ is positive definite and $\bar{\mathbf{J}}_{\text{un}}$ is non-negative definite. As a consequence, the condition on μ to guarantee $\rho(\bar{\mathbf{H}}_{\text{un}}) < 1$ can be given by [32]

$$0 < \mu < \min \left\{ \frac{1}{\lambda_{\max}\left(\bar{\mathbf{C}}_{\mathrm{un}}^{-1}\bar{\mathbf{J}}_{\mathrm{un}}\right)}, \frac{1}{\max\left(\lambda(\bar{\boldsymbol{\eta}}_{\mathrm{cn}}) \in R^+\right)} \right\}, \tag{37}$$

where $\bar{\boldsymbol{\eta}}_{\mathrm{un}} = \begin{pmatrix} \frac{1}{2}\bar{\mathbf{C}}_{\mathrm{un}} & -\frac{1}{2}\bar{\mathbf{J}}_{\mathrm{un}} \\ \mathbf{I}_{N^2} & \mathbf{0}_{N^2} \end{pmatrix}$.

4 Analysis of constrained FDAFs

We are now ready to study the statistical behavior of constrained FDAFs. The time-domain approach used above is adopted for the performance evaluation. However, as will be shown, the constrained algorithm exhibits some different convergence properties.

4.1 Mean convergence

We start by examining the mean convergence in the constrained case. For that, we define the time-domain weight vector as $\hat{\mathbf{w}}_{\mathrm{cn}}(k) = \Upsilon_{10}\mathbf{F}^{-1}\hat{\mathbf{W}}(k)$ of length L, where Υ_{10} extracts the first L components of the inverse DFT of $\hat{\mathbf{W}}(k)$. Using (4), we obtain the error vector

$$\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{X}_1(k)\hat{\mathbf{w}}_{\mathrm{cn}}(k). \tag{38}$$

Pre-multiply (6) by $\Upsilon_{10}\mathbf{F}^{-1}$, we obtain the time-domain update equation

$$\hat{\mathbf{w}}_{\mathrm{cn}}(k+1) = \hat{\mathbf{w}}_{\mathrm{cn}}(k) + \mu\bar{\mathbf{P}}\bar{\mathbf{X}}^T(k)\mathbf{e}(k). \tag{39}$$

Subtracting $\mathbf{w}_\dagger$ from (39) yields

$$\tilde{\mathbf{w}}_{\mathrm{cn}}(k+1) = \tilde{\mathbf{w}}_{\mathrm{cn}}(k) - \mu\bar{\mathbf{P}}\bar{\mathbf{X}}^T(k)\mathbf{e}(k), \tag{40}$$

where $\tilde{\mathbf{w}}_{\mathrm{cn}}(k) = \mathbf{w}_\dagger - \hat{\mathbf{w}}_{\mathrm{cn}}(k)$ denotes the weight-error vector of the constrained FDAF. We rewrite $\mathbf{d}(k)$ in terms of $\mathbf{w}_\dagger$ as

$$\mathbf{d}(k) = \mathbf{X}_1(k)\mathbf{w}_\dagger + \mathbf{X}_3(k)\mathbf{w}_* + \mathbf{v}(k). \tag{41}$$

Substituting (41) into (38) leads to

$$\mathbf{e}(k) = \mathbf{X}_1(k)\tilde{\mathbf{w}}_{\mathrm{cn}}(k) + \mathbf{X}_3(k)\mathbf{w}_* + \mathbf{v}(k). \tag{42}$$

Inserting (42) into (40), we obtain the recursion

$$\tilde{\mathbf{w}}_{\mathrm{cn}}(k+1) = \tilde{\mathbf{w}}_{\mathrm{cn}}(k) - \mu\mathbf{A}_{\mathrm{cn}}(k)\tilde{\mathbf{w}}_{\mathrm{cn}}(k) - \mu\mathbf{B}_{\mathrm{cn}}(k)\mathbf{w}_* - \mu\bar{\mathbf{P}}\bar{\mathbf{X}}^T(k)\mathbf{v}(k), \tag{43}$$

where $\mathbf{A}_{\mathrm{cn}}(k) = \bar{\mathbf{P}}\bar{\mathbf{X}}^T(k)\mathbf{X}_1(k)$ and $\mathbf{B}_{\mathrm{cn}}(k) = \bar{\mathbf{P}}\bar{\mathbf{X}}^T(k)\mathbf{X}_3(k)$.

Taking the mathematical expectation of (43) and using the independence assumption

$$E[\tilde{\mathbf{w}}_{\mathrm{cn}}(k+1)] = \left(\mathbf{I}_L - \mu E[\mathbf{A}_{\mathrm{cn}}(k)]\right)E[\tilde{\mathbf{w}}_{\mathrm{cn}}(k)] - \mu E[\mathbf{B}_{\mathrm{cn}}(k)]\mathbf{w}_*. \tag{44}$$

This recursion describes the time evolution of $E[\tilde{\mathbf{w}}_{\mathrm{cn}}(k)]$. The last term in (44) disappears in the exact modeling case. It is shown in [20] that for the constrained FDAF with step-normalization, we have the approximation $E[\mathbf{A}_{\mathrm{cn}}(k)] \approx \frac{R}{N}\mathbf{I}_L$ when N is large enough, and hence the algorithm only has a single convergence mode. However, the convergence speed of the constrained FDAF without step-normalization is linked to the eigenvalue spread of $E[\mathbf{A}_{\mathrm{cn}}(k)] = E[\mathbf{X}_1^T(k)\mathbf{X}_1(k)] = \mathbf{R}_{11}$, which may be significantly higher than that of the step-normalization version.

Using (9) and (11), the matrix governing the mean convergence of unconstrained FDAFs is represented as

$$E[\mathbf{A}_{\mathrm{un}}(k)] = \begin{bmatrix} \mathbf{P}_1\mathbf{R}_{11} + \mathbf{P}_2\mathbf{R}_{21} & \mathbf{P}_1\mathbf{R}_{21}^T + \mathbf{P}_2\mathbf{R}_{22} \\ \mathbf{P}_2^T\mathbf{R}_{11} + \mathbf{P}_3\mathbf{R}_{21} & \mathbf{P}_2^T\mathbf{R}_{21}^T + \mathbf{P}_3\mathbf{R}_{22} \end{bmatrix}. \tag{45}$$

where $\mathbf{R}_{21} = E[\mathbf{X}_2^T(k)\mathbf{X}_1(k)]$. It is then observed that the matrix that controls the mean convergence of the constrained version $E[\mathbf{A}_{\mathrm{cn}}(k)] = \bar{\mathbf{P}}E[\bar{\mathbf{X}}^T(k)\mathbf{X}_1(k)] = \mathbf{P}_1\mathbf{R}_{11} + \mathbf{P}_2\mathbf{R}_{21}$ is a submatrix of $E[\mathbf{A}_{\mathrm{un}}(k)]$. For the FDAF without step-normalization, it has been shown in [22] that

$$\frac{\lambda_{\max}(E[\mathbf{A}_{\mathrm{un}}(k)])}{\lambda_{\min}(E[\mathbf{A}_{\mathrm{un}}(k)])} \geq \frac{\lambda_{\max}(E[\mathbf{A}_{\mathrm{cn}}(k)])}{\lambda_{\min}(E[\mathbf{A}_{\mathrm{cn}}(k)])}. \tag{46}$$

Eq. (46) states that eigenvalue spread of $E[\mathbf{A}_{\mathrm{un}}(k)]$ is always larger or equal to that of $E[\mathbf{A}_{\mathrm{cn}}(k)]$, and hence, the constrained FDAF may converge faster than the unconstrained version. However, it is very challenging to verify if the relation (46) holds for the FDAF algorithms with step-normalization, which remains an open problem.

At steady state, the constrained FDAF converges to

$$\begin{aligned} E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)] &= -(E[\mathbf{A}_{\mathrm{cn}}(k)])^{-1}E[\mathbf{B}_{\mathrm{cn}}(k)]\mathbf{w}_* \\ &= -(\mathbf{P}_1\mathbf{R}_{11} + \mathbf{P}_2\mathbf{R}_{21})^{-1}(\mathbf{P}_1\mathbf{R}_{13} + \mathbf{P}_2\mathbf{R}_{23})\mathbf{w}_*, \end{aligned} \tag{47}$$

where $\mathbf{R}_{13} = E[\mathbf{X}_1^T(k)\mathbf{X}_3(k)]$, $\mathbf{R}_{21} = E[\mathbf{X}_2^T(k)\mathbf{X}_1(k)]$ and $\mathbf{R}_{23} = E[\mathbf{X}_2^T(k)\mathbf{X}_3(k)]$.

Let us investigate whether the constrained FDAF converges to the optimum Wiener solution. We present the optimum Wiener solution by minimizing the quadratic cost function $E[\|\mathbf{e}(k)\|^2]$ in (38)

$$\tilde{\mathbf{w}}_{\mathrm{cn,opt}} = -\mathbf{R}_{11}^{-1}\mathbf{R}_{13}\mathbf{w}_*. \tag{48}$$

The optimum Wiener solution of the unconstrained FDAF in (21) utilizes the future information, and hence, it is non-causal. However, the optimum Wiener solution of the constrained FDAF in (48) only requires the past inputs, which is causal. Also, Eq. (21) is an unconstrained solution since we do not impose any constraint on $\tilde{\mathbf{w}}_{\mathrm{un}}(k)$. However, Eq. (48) is a constrained solution since the last R components of the inverse DFT of $\hat{\mathbf{W}}(k)$ are always forced to be zero. Due to the additional constraint, the MMSE from (48) is larger than that from (21), which is already pointed out in [21] but the explicit MMSE expression is not given in [21].

In the exact modeling scenario, the optimum Wiener solution (48) is the true system impulse response, and hence, the constrained FDAF converges to the optimum Wiener solution, i.e., the true plant.

We now pay our attention to the under-modeling scenario.

- For constrained FDAFs without step-normalization, we have $\mathbf{P}_1 = \mathbf{I}_L, \mathbf{P}_2 = \mathbf{0}_{L \times R}$ and obtain the relation $E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)] = \tilde{\mathbf{w}}_{\mathrm{cn}}(k) = -\mathbf{R}_{11}^{-1}\mathbf{R}_{13}\mathbf{w}_*$. The constrained FDAF without step-normalization converges to the causal Wiener solution. For the white noise input, i.e., $\mathbf{R}_{13} = \mathbf{0}_{L \times Q}$, we have $E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)] = \mathbf{0}_{L \times 1}$, which means that the constrained FDAF without step-normalization converges to the first L coefficients of the

system impulse response. When correlated signals are used as input, however, it has $\mathbf{R}_{13} \neq \mathbf{0}_{L \times Q}$, and then $E[\tilde{\mathbf{w}}_{cn}(\infty)] \neq \mathbf{0}_{L \times 1}$, which means that the constrained FDAF algorithm without step-normalization cannot converge to the first L coefficients of the system impulse response.

- For constrained FDAFs with step-normalization, we have $\mathbf{P}_1 = N\sigma_x^2 \mathbf{I}_L, \mathbf{P}_2 = \mathbf{0}_{L \times R}, \mathbf{R}_{13} = \mathbf{0}_{L \times Q}$ for white noise input, and then $E[\tilde{\mathbf{w}}_{cn}(\infty)] = \tilde{\mathbf{w}}_{cn,opt} = \mathbf{0}_{L \times 1}$, which means that it can converge to the causal Wiener solution and the first L coefficients of the system impulse response. When correlated signals are used as input, we can verify that $E[\tilde{\mathbf{w}}_{cn}(\infty)] \neq \tilde{\mathbf{w}}_{cn,opt}$ and $E[\tilde{\mathbf{w}}_{cn}(\infty)] \neq \mathbf{0}_{L \times 1}$, which means that it cannot converge to the causal Wiener solution or the first L coefficients of the system impulse response.

4.2 Mean-square convergence

Next, we perform a detailed second-order analysis of the constrained FDAFs. Using (42), we obtain the instantaneous MSE of the constrained FDAFs

$$
\begin{aligned}
J_{cn}(k) &= \frac{1}{R} E[\mathbf{e}^T(k)\mathbf{e}(k)] \\
&= \frac{1}{R}\{\mathrm{tr}(E[\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)]\mathbf{R}_{11}) + 2\tilde{\mathbf{w}}_{cn}^T(k)\mathbf{R}_{13}\mathbf{w}_* + \mathbf{w}_*\mathbf{R}_3\mathbf{w}_*^T\} + \sigma_\nu^2.
\end{aligned}
\tag{49}
$$

where $J_{cn,min} = \sigma_\nu^2$ is the MMSE that can be realized only when the adaptive coefficients converge to the system impulse response, and $J_{cn,ex}(k) = \frac{1}{R}\{\mathrm{tr}(E[\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)]\mathbf{R}_{11}) + 2\tilde{\mathbf{w}}_{cn}^T(k)\mathbf{R}_{13}\mathbf{w}_* + \mathbf{w}_*\mathbf{R}_3\mathbf{w}_*^T\}$ is the EMSE.

We study the evaluation of the covariance $E[\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)]$ to obtain the MSE. We post-multiply (43) by its transpose and take the expectation

$$
\begin{aligned}
&E[\tilde{\mathbf{w}}_{cn}(k+1)\tilde{\mathbf{w}}_{cn}^T(k+1)] \\
&= E[\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)] \\
&\quad - \mu E[\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)\mathbf{A}_{cn}^T(k)] \\
&\quad - \mu E[\mathbf{A}_{cn}(k)\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)] \\
&\quad + \mu^2 E[\mathbf{A}_{cn}(k)\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)\mathbf{A}_{cn}^T(k)] \\
&\quad - \mu E[\tilde{\mathbf{w}}_{cn}(k)\mathbf{w}_*^T\mathbf{B}_{cn}^T(k)] \\
&\quad - \mu E[\mathbf{B}_{cn}(k)\mathbf{w}_*\tilde{\mathbf{w}}_{cn}^T(k)] \\
&\quad + \mu^2 E[\mathbf{B}_{cn}(k)\mathbf{w}_*\tilde{\mathbf{w}}_{cn}^T(k)\mathbf{A}_{cn}^T(k)] \\
&\quad + \mu^2 E[\mathbf{B}_{cn}(k)\mathbf{w}_*\mathbf{w}_*^T\mathbf{B}_{cn}^T(k)] \\
&\quad + \mu^2 E[\mathbf{A}_{cn}(k)\tilde{\mathbf{w}}_{cn}(k)\mathbf{w}_*^T\mathbf{B}_{cn}^T(k)] \\
&\quad + \mu^2 \delta_\nu^2 \bar{\mathbf{P}}\bar{\mathbf{R}}\bar{\mathbf{P}}^T.
\end{aligned}
\tag{50}
$$

Following the same argument used in the unconstrained FDAFs, we define the $L^2 \times 1$ vector $\mathbf{z}_{cn}(k) = \mathrm{vec}\{E[\tilde{\mathbf{w}}_{cn}(k)\tilde{\mathbf{w}}_{cn}^T(k)]\}$. Applying the operation of vectorization to (50), we obtain the following difference equation

$$
\mathbf{z}_{cn}(k+1) = \mathbf{H}_{cn}\mathbf{z}_{cn}(k) + \mathbf{\Theta}_{cn}(k),
\tag{51}
$$

where

$$
\begin{aligned}
\mathbf{H}_{\mathrm{cn}} &= \mathbf{I}_{L^2} - \mu \mathbf{C}_{\mathrm{cn}} + \mu^2 \mathbf{J}_{\mathrm{cn}} \\
\mathbf{C}_{\mathrm{cn}} &= E[\mathbf{A}_{\mathrm{cn}}(k)] \otimes \mathbf{I}_L + \mathbf{I}_L \otimes E[\mathbf{A}_{\mathrm{cn}}(k)] \\
\mathbf{J}_{\mathrm{cn}} &= E[\mathbf{A}_{\mathrm{cn}}(k) \otimes \mathbf{A}_{\mathrm{cn}}(k)]
\end{aligned}
\tag{52}
$$

$$
\begin{aligned}
\boldsymbol{\Theta}_{\mathrm{cn}}(k) = {}& \mu^2 \delta_\nu^2 \mathrm{vec}(\mathbf{P}\bar{\bar{\mathbf{R}}}\mathbf{P}^T) \\
& - \mu(E[\mathbf{B}_{\mathrm{cn}}(k)] \otimes \mathbf{I}_L)\mathrm{vec}(E[\tilde{\mathbf{w}}_{\mathrm{cn}}(k)]\mathbf{w}_*^T) \\
& + \mu^2(E[\mathbf{B}_{\mathrm{cn}}(k)] \otimes \mathbf{A}_{\mathrm{cn}}(k))\mathrm{vec}(E[\tilde{\mathbf{w}}_{\mathrm{cn}}(k)]\mathbf{w}_*^T) \\
& - \mu(\mathbf{I}_L \otimes E[\mathbf{B}_{\mathrm{cn}}(k)])\mathrm{vec}(\mathbf{w}_* E[\tilde{\mathbf{w}}_{\mathrm{cn}}^T(k)]) \\
& + \mu^2(E[\mathbf{A}_{\mathrm{cn}}(k)] \otimes \mathbf{B}_{\mathrm{cn}}(k))\mathrm{vec}(\mathbf{w}_* E[\tilde{\mathbf{w}}_{\mathrm{cn}}^T(k)]) \\
& + \mu^2(E[\mathbf{B}_{\mathrm{cn}}(k)] \otimes \mathbf{B}_{\mathrm{cn}}(k))\mathrm{vec}(\mathbf{w}_* \mathbf{w}_*^T).
\end{aligned}
\tag{53}
$$

Note that the eigenvalue spread of $\mathbf{H}_{\mathrm{cn}}$ determines the mean-square convergence behavior of constrained FDAFs. Once $\mathbf{z}_{\mathrm{cn}}(k)$ is recursively generated using (51), the instantaneous MSD of the constrained FDAFs is then given by $\delta_{\mathrm{cn}}(k) = E(\lVert \tilde{\mathbf{w}}_{\mathrm{cn}}(k) \rVert^2) = \mathrm{tr}(\mathrm{vec}^{-1}(\mathbf{z}_{\mathrm{cn}}(k)))$.

In the steady state, we have $\mathbf{z}_{\mathrm{cn}}(\infty) = (\mathbf{I}_{L^2} - \mathbf{H}_{\mathrm{cn}})^{-1}\boldsymbol{\Theta}_{\mathrm{cn}}(\infty)$ according to (51), and we immediately obtain the steady-state MSD $\delta_{\mathrm{cn}}(\infty) = E(\lVert \tilde{\mathbf{w}}_{\mathrm{cn}}(\infty) \rVert^2) = \mathrm{tr}(\mathrm{vec}^{-1}(\mathbf{z}_{\mathrm{cn}}(\infty)))$. Also, the steady-state EMSE follows:

$$
J_{\mathrm{cn,ex}}(\infty) = \frac{1}{R}\{\mathrm{tr}(E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)\tilde{\mathbf{w}}_{\mathrm{cn}}^T(\infty)]\mathbf{R}_{11}) + 2E[\tilde{\mathbf{w}}_{\mathrm{cn}}^T(\infty)]\mathbf{R}_{13}\mathbf{w}_* + \mathbf{w}_*\mathbf{R}_3\mathbf{w}_*^T\}. \tag{54}
$$

We have derived the steady-state EMSE expressions of four variants of the FDAF algorithm, which are accurate but somewhat complex. Some simplifications for the EMSE expressions in the exact modeling case can be found in the literature. Closed-form EMSE expressions are given in [20] without detailed derivation, and the same results are derived in the frequency domain in [4]. In Appendix A, we adopt a different approach to derive the approximate EMSE expressions in the time domain. As shown, our derivation is easy to understand and also we provide a more accurate solution for the constrained FDAF algorithm..

Next, we investigate how the MSE of the constrained FDAFs is generated and derive the attainable MMSE for the two versions. To this end, we define $\bar{\boldsymbol{\Delta}} = \tilde{\mathbf{w}}_{\mathrm{cn,opt}} - E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)]$. The difference between $\hat{\mathbf{w}}_{\mathrm{cn}}(k)$ and its steady-state mean value is

$$
\mathbf{q}_{\mathrm{cn}}(k) = \boldsymbol{\omega}_{\mathrm{cn}}(k) - E[\boldsymbol{\omega}_{\mathrm{cn}}(\infty)] = E[\boldsymbol{\omega}_{\mathrm{cn}}(\infty)] - \boldsymbol{\omega}_{\mathrm{cn}}(k) = \boldsymbol{\omega}_{\mathrm{cn,opt}} - \bar{\Delta} - \boldsymbol{\omega}_{\mathrm{cn}}(k). \tag{55}
$$

We rewrite the steady-state MSE terms of $\mathbf{q}_{\mathrm{cn}}(k)$

$$
\begin{aligned}
J_{\mathrm{cn}}(\infty) = {}& \frac{1}{R}\mathrm{tr}(E[\mathbf{q}_{\mathrm{cn}}(\infty)\mathbf{q}_{\mathrm{cn}}^T(\infty)]\mathbf{R}_{11}) \\
& + \frac{1}{R}\left\{ \mathrm{tr}(E\bar{\boldsymbol{\Delta}}\bar{\boldsymbol{\Delta}}^T\mathbf{R}_{11}) - 2\mathrm{tr}(\bar{\boldsymbol{\Delta}}\tilde{w}_{\mathrm{cn,opt}}^T\mathbf{R}_{11}) + 2\bar{\boldsymbol{\Delta}}^T\mathbf{R}_{13}\mathbf{w}_* \right\} \\
& + \frac{1}{R}\left\{ \mathrm{tr}(\tilde{\mathbf{w}}_{\mathrm{cn,opt}}\tilde{\mathbf{w}}_{\mathrm{cn,opt}}^T\mathbf{R}_{11}) + 2\tilde{\mathbf{w}}_{\mathrm{cn,opt}}^T\mathbf{R}_{13}\mathbf{w}_* + \mathbf{w}_*\mathbf{R}_3\mathbf{w}_*^T \right\} + \sigma_\nu^2
\end{aligned}
\tag{56}
$$

The first term in the right-hand side of (56) is related to the fluctuations of $\tilde{\mathbf{w}}_{cn}(\infty)$ around its mean $E[\tilde{\mathbf{w}}_{cn}(\infty)]$, and the second term is attributed to the bias between $E[\tilde{\mathbf{w}}_{cn}(\infty)]$ and its optimum solution $\tilde{\mathbf{w}}_{cn,opt}$.

Because the constrained FDAF algorithm without step-normalization for any inputs and the constrained FDAF algorithm with step-normalization for white noise input can converge to the optimum Wiener solution, i.e., $\bar{\boldsymbol{\Delta}} = \mathbf{0}_{L \times 1}$, the second term in (56) then becomes zero and the attainable MMSE is

$$J_{\min,3} = \frac{1}{R}\{\text{tr}(\tilde{\mathbf{w}}_{cn,opt}\tilde{\mathbf{w}}_{cn,opt}^T \mathbf{R}_{11}) + 2\tilde{\mathbf{w}}_{cn,opt}^T \mathbf{R}_{13}\mathbf{w}_* + \mathbf{w}_* \mathbf{R}_3 \mathbf{w}_*^T\} + \sigma_v^2. \tag{57}$$

which is the same as the LMS [34]. When the input is correlated, the constrained FDAF algorithm with step-normalization cannot converge to the Wiener solution, i.e., $\bar{\boldsymbol{\Delta}} \neq \mathbf{0}_{L \times 1}$. The attainable MMSE in this case is given by

$$J_{\min,4} = \frac{1}{R}\{\text{tr}(E[\tilde{\mathbf{w}}_{cn}(\infty)]E[\tilde{\mathbf{w}}_{cn}^T(\infty)]\mathbf{R}_{11}) + 2E[\tilde{\mathbf{w}}_{cn}^T(\infty)]\mathbf{R}_{13}\mathbf{w}_* + \mathbf{w}_* \mathbf{R}_3 \mathbf{w}_*^T\} + \sigma_v^2, \tag{58}$$

where $E[\tilde{\mathbf{w}}_{cn}(\infty)] = -(\mathbf{P}_1\mathbf{R}_{11} + \mathbf{P}_2\mathbf{R}_{21})^{-1}(\mathbf{P}_1\mathbf{R}_{13} + \mathbf{P}_2\mathbf{R}_{23})\mathbf{w}_*$.

We now compare the MMSE performance of four variants of FDAF algorithms. In the exact modeling scenario, all the variants can achieve the same MMSE, i.e., $J_{\min}(\infty) = \sigma_v^2$. The case is different for the under-modeling scenario. For white noise input, it has

$$J_{\min,1} = J_{\min,2} \leq J_{\min,3} = J_{\min,4}. \tag{59}$$

For correlated input, we have

$$J_{\min,1} = J_{\min,2} \leq J_{\min,3} < J_{\min,4}. \tag{60}$$

4.3 Stability bound

We now present the condition in which the constrained FDAF algorithm can be stable. For the constrained FDAF algorithm without step-normalization, we have $\mathbf{P}_1 = \mathbf{I}_L$, $\mathbf{P}_2 = \mathbf{0}_{L \times R}$, and hence, $E[\mathbf{A}_{cn}(k)] = \mathbf{R}_{11}$ is positive definite with real-valued eigenvalues. To guarantee the mean stability, we should have $\rho(\mathbf{I}_L - \mu E[\mathbf{A}_{cn}(k)]) < 1$ such that all eigenvalues of $\mathbf{I}_L - \mu E[\mathbf{A}_{cn}(k)]$ must be less than one in magnitude, and then, the step size satisfies

$$\mu < \frac{2}{\lambda_{\max}(E[\mathbf{A}_{cn}(k)])}. \tag{61}$$

According to (51), we should have $\rho(\mathbf{H}_{cn}) < 1$ such that the algorithm can be mean-square stable. We can infer that $\mathbf{C}_{cn}$ is positive definite and $\mathbf{J}_{cn}$ is non-negative definite, and then, the range of step size is given by [32]

$$0 < \mu < \min \left\{ \frac{1}{\lambda_{\max}(\mathbf{C}_{cn}^{-1}\mathbf{J}_{cn})}, \frac{1}{\max(\lambda(\boldsymbol{\eta}_{cn}) \in R^+)} \right\}, \tag{62}$$

where $\boldsymbol{\eta}_{cn} = \begin{pmatrix} \frac{1}{2}\mathbf{C}_{cn} & -\frac{1}{2}\mathbf{J}_{cn} \\ \mathbf{I}_{L^2} & \mathbf{0}_{L^2} \end{pmatrix}$.

For the constrained FDAF algorithm with step-normalization, $E[\mathbf{A}_{cn}(k)]$ is not a Hermitian matrix, and hence, the eigenvalues of $E[\mathbf{A}_{cn}(k)]$ may not be real. The condition that guarantees mean convergence is expressed as

$$0 < \mu < \frac{2\Re e(\lambda_i)}{|\lambda_i|^2}, i = 0, 1, \cdots, L - 1, \tag{63}$$

where λ_i is the i-th eigenvalue of $E[\mathbf{A}_{cn}(k)]$. Similarly, the constrained FDAF is mean-square stable, if and only if

$$\rho\left(\mathbf{I}_{L^2} - \mu\mathbf{C}_{cn} + \mu^2\mathbf{J}_{cn}\right) < 1. \tag{64}$$

However, the matrices $\mathbf{C}_{cn}$ and $\mathbf{J}_{cn}$ are not Hermitian in this case, and hence, we cannot use (62) to solve the inequality (64). In practice, a numerical solution may be used to handle this problem, see [25] for more details.

Given the complexity of the mathematical developments, some variables and matrices used in the formulations are summarized in Appendix B to enhance the readability, although they are explicitly defined.

5 Results and discussion

We conduct computer simulations to demonstrate and support our analysis. The unknown system impulse response has a length of $M = 16$ and is with the coefficients [0.01 0.02 −0.04 −0.08 0.15 −0.3 0.45 0.6 0.6 0.45 −0.3 0.15 −0.08 −0.04 0.02 0.01]. The adaptive filter has a length of $L = 16$ or $L = 14$, corresponding to the exact- and under-modeling situations, respectively. The input $x(n)$ is generated by filtering the white Gaussian noise with $H(z) = \sqrt{1 - a^2}/(1 - az^{-1})$, which is white noise for $a = 0$ and an AR(1) process for $0 < |a| < 1$. The signal-to-noise ratio (SNR) is defined as $10\log_{10}\left(E[(\mathbf{x}^T(n)\mathbf{w}_{opt})^2]/E[v^2(n)]\right)$. We estimate the required statistical moments by ensemble averaging.

Figure 1 presents the MSD and EMSE learning curves of the step-normalization FDAF algorithm. SNR = 10 dB and SNR = ∞ are used in this example. We here consider an exact modeling case, i.e., $L = 16$ and $R = 16$. The step size is $\mu = 0.1$. The theoretical learning curves of the unconstrained FDAF algorithm are generated recursively using (28) and (26), while that of the constrained version are obtained from (51) and (49). High consistency between theory and experimental results is witnessed in Fig. 1. Also, we find that the constrained and unconstrained FDAFs exhibit a similar initial convergence rate in this case. This is because the eigenvalue spread of $\mathbf{A}_{un}(k)$ is only slightly larger than that of $\mathbf{A}_{cn}(k)$, i.e., $\frac{\lambda_{\max}(E[\mathbf{A}_{un}(k)])}{\lambda_{\min}(E[\mathbf{A}_{un}(k)])} = 1.63$ and $\frac{\lambda_{\max}(E[\mathbf{A}_{cn}(k)])}{\lambda_{\min}(E[\mathbf{A}_{cn}(k)])} = 1.52$. The step-normalization procedure greatly reduces the eigenvalue spread of $\mathbf{A}_{un}(k)$ and $\mathbf{A}_{cn}(k)$, and the constraining matrix $\mathbf{G}_{10}$ further diagonalizes $\mathbf{A}_{cn}(k)$. The experimental MSD and EMSE for SNR = ∞ both achieve a floor due to the roundoff error. We conduct simulations in the under-modeling situation in Fig. 2, where we set $L = 14$ and $R = 28$ and other
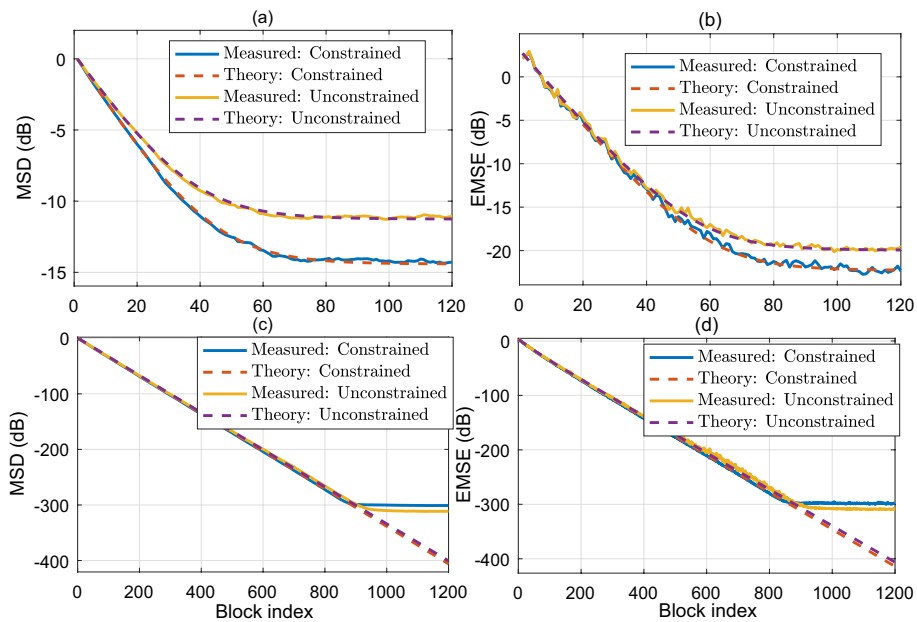
Fig. 1 Predicted and simulated learning curves in exact modeling situation: (**a**) MSD for SNR = 10 dB, (**b**) EMSE for SNR = 10 dB, (**c**) MSD for SNR = ∞, and (**d**) EMSE for SNR = ∞

conditions are the same as that in Fig. 1, and we observe that the presented model is capable of predicting the EMSE and MSD learning curves well.

Figure 3 presents the steady-state mean-square performance of the step-normalization FDAF algorithms for SNR = 30 dB. We here consider an exact modeling case with $L = 16$ and $R = 16$. An AR(1) process with $a = 0.9$ is used as input, and μ varies from 0.01 to 1.9. The stability bounds are also illustrated in Fig. 3. Note that the simulations are consistent with the theoretical predictions, i.e., Eqs. (31) and (54), even at large step sizes. We then investigate the accuracy of several approximations shown in Appendix A. To give a more intuitive observation, we present some theoretical and measured results in Table 1. Our approximated EMSE solutions of the unconstrained and constrained FDAFs, i.e., Eqs. (A5) and (A11), match the experimental values well for smaller step sizes, but the previous theory in [4] and [20] (i.e., Eq. (A12)) present the largest prediction errors. It is apparent that all the approximated results deviate from the simulations when the step size is large. Given the same step size, the constrained FDAF algorithm has lower steady-state MSD and EMSE values than the unconstrained algorithm.

We investigate the steady-state MSD and EMSE performance of the FDAFs in undermodeling in Fig. 4 for SNR = 30 dB. The experimental conditions are the same as Fig. 3 except that $L = 14$ and $R = 28$ are used. It is observed in this example that the theoretical predictions are similar to the simulations. The constrained FDAF achieves smaller
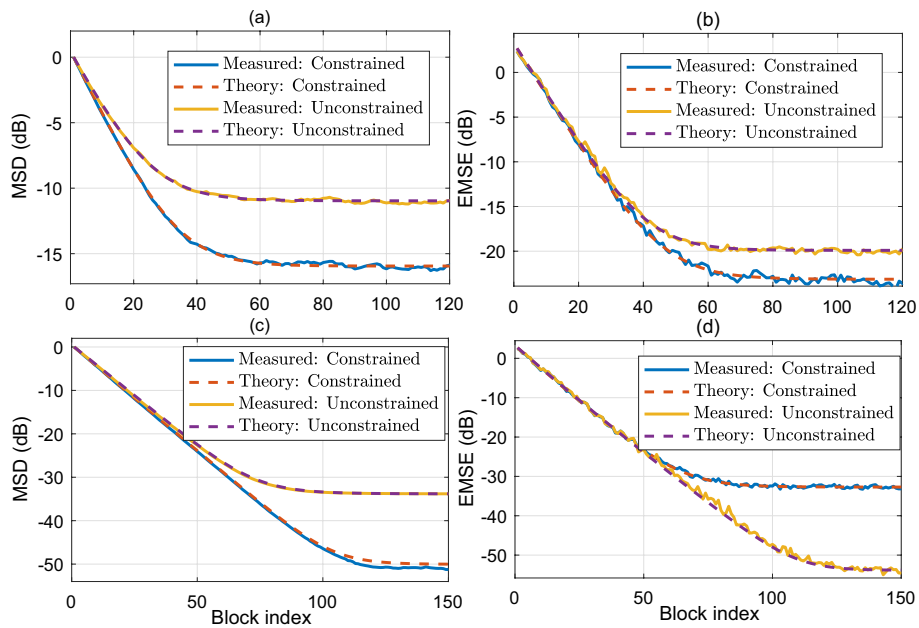
Fig. 2 Predicted and simulated learning curves in the under-modeling situation: (**a**) MSD for SNR = 10 dB, (**b**) EMSE for SNR = 10 dB, (**c**) MSD for SNR = ∞, and (**d**) EMSE for SNR = ∞
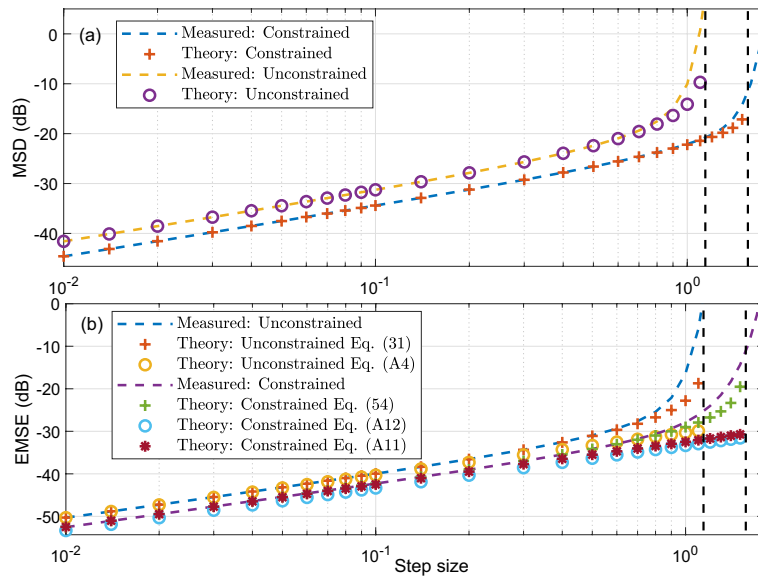

Fig. 3 Steady-state performance of the FDAF with step-normalization in the exact modeling situation: (**a**) MSD and (**b**) EMSE

steady-state MSD values than the unconstrained one in under-modeling case. However, the steady-state MSE in under-modeling case is quite different from that in the exact modeling case. When the step size is small enough, the EMSE related to the step size

Table 1 Measured and theoretical steady-state EMSEs of constrained FDAF with step-normalization in exact modeling case

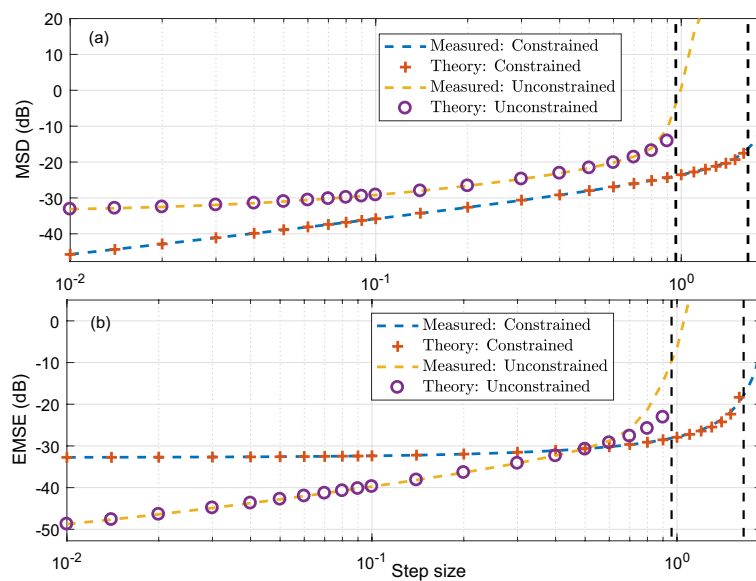| μ	Measured	Predicted from (54)	Predicted from (A11)	Predicted from (A12)
0.01	− 52.54	− 52.49	− 52.49	− 53.33
0.05	− 45.42	− 45.41	− 45.50	− 46.34
0.1	− 42.29	− 42.27	− 42.49	− 43.34
0.5	− 34.15	− 34.16	− 35.50	− 36.34
1	− 27.89	− 29.00	− 32.49	− 33.33
1.5	− 14.35	− 19.51	− 30.72	− 31.56

Fig. 4 Steady-state performance of the FDAF with step-normalization in the under-modeling situation: **a** MSD and **b** EMSE

can be ignored. Hence, the unconstrained FDAF has a lower steady-state EMSE when $\mu < 0.5$ since it converges to the non-causal Wiener solution. As the step size increases, the first term in (33) and (56) may dominate. Consequently, we found that for $\mu > 0.5$ the unconstrained FDAF algorithm has a larger EMSE compared to its constrained counterpart.

Figure 5 displays the MMSE of four FDAF algorithms with different a. The filter has a length of $L = 14$, and the block has a length of $R = 14$. Notice that an under-modeling situation has been considered here, whereas four variants of the FDAF have the same MMSE for a sufficient long filter length and the results are not shown. Because the unconstrained FDAF algorithm can always converge to the non-causal Wiener solution, it has a smaller MMSE than the constrained version. An intuitive explanation is that the unconstrained FDAF yields superior MSE performance by utilizing more data
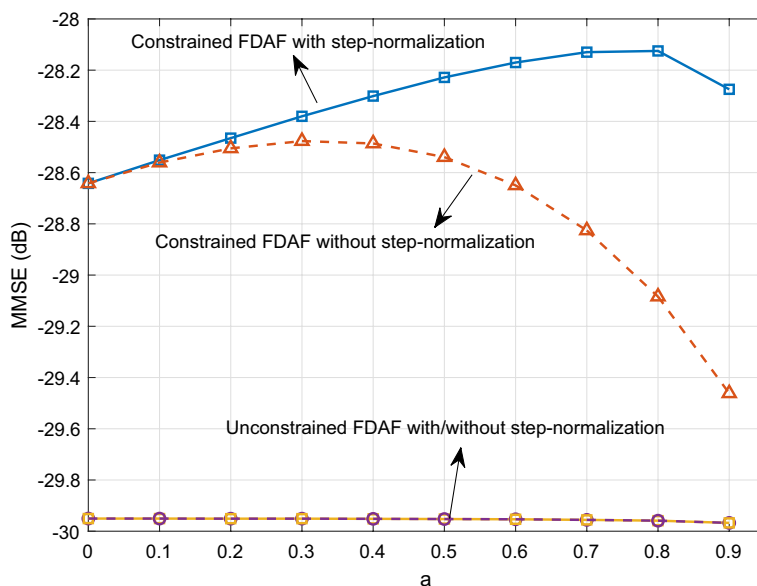
Fig. 5 Attainable MMSE of four FDAF algorithms as a function of the parameter *a*

than the constrained counterpart. This finding has already been investigated in [20] and [21], but they did not provide the analytical MMSE expressions. Also note that with or without step-normalization, the constrained FDAF attains similar MMSE values for *a* close to zero, i.e., the input is close to white noise. However, as the parameter *a* increases, the MMSE of the step-normalization FDAF turns out to be worse than that without step-normalization. The observation is consistent with our analysis, i.e., the constrained FDAF algorithm without step-normalization converges to the causal Wiener solution for an arbitrary input, while for the constrained FDAF algorithm with step-normalization, this can only happen for a white noise input.

6 Conclusion

We carried out the performance evaluation of a family of FDAF algorithms in the time domain. In particular, we have derived the recursive models for their mean and mean-square properties. We extensively discussed whether each variant of the FDAF converges to the system impulse response or the optimum Wiener solution. Closed-form expressions that characterize the MSD and MSE performance of the FDAFs were derived, and accurate and approximate steady-state EMSEs were given. The MMSEs of the four versions of the FDAF were compared. This work is more general and easier to follow than our previous study [25–27], which offers a thorough insight into the underlying learning theory of FDAFs. Computer simulations supported the theoretical model.

Appendix: A

This appendix examines the steady-state mean-square behavior of FDAFs that is easy to follow. Particularly, the derivations will rely on some simplifications.

For the unconstrained FDAF algorithm without step-normalization, it has $E[\mathbf{A}_{\mathrm{un}}(k)] = \bar{\mathbf{R}}$. Using (27) and ignoring the terms related to μ^2, we obtain

$$E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)\tilde{\mathbf{w}}_{\mathrm{un}}^T(\infty)]\bar{\mathbf{R}} + \bar{\mathbf{R}}E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)\tilde{\mathbf{w}}_{\mathrm{un}}^T(\infty)] = \mu\delta_v^2\bar{\mathbf{R}}. \tag{A1}$$

Taking the trace of both sides of (A1), it has

$$\mathrm{tr}\left(E[\mathbf{Q}_{\mathrm{un}}(\infty)\mathbf{Q}_{\mathrm{un}}^T(\infty)]\mathbf{N}\right) = \frac{1}{2}\mu\delta_v^2\mathrm{tr}(\mathbf{N}) = \frac{1}{2}\mu N R\delta_v^2\delta_x^2. \tag{A2}$$

Considering (31) and (A2), we then obtain the steady-state EMSE of the unconstrained FDAF algorithm without step-normalization

$$J_{\mathrm{un,ex}}(\infty) = \frac{1}{R}\mathrm{tr}(E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)\tilde{\mathbf{w}}_{\mathrm{un}}^T(\infty)]\bar{\mathbf{R}}) = \frac{1}{2}\mu N\delta_v^2\delta_x^2. \tag{A3}$$

For the step-normalization version, using the relation $E[\mathbf{A}_{\mathrm{un}}(k)] = \frac{R}{N}\mathbf{I}_N$ and (27), we have

$$E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)\tilde{\mathbf{w}}_{\mathrm{un}}^T(\infty)] = \frac{1}{2}\mu\delta_v^2\mathbf{P}. \tag{A4}$$

Using (31) and (A.4), the steady-state EMSE of the unconstrained FDAF algorithm with step-normalization then follows

$$J_{\mathrm{un,ex}}(\infty) = \frac{1}{R}\mathrm{tr}(E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)\tilde{\mathbf{w}}_{\mathrm{un}}^T(\infty)]\bar{\mathbf{R}}) = \frac{1}{2R}\mu\delta_v^2\mathrm{tr}(\mathbf{P}\bar{\mathbf{R}}) = \frac{1}{2}\mu\delta_v^2. \tag{A5}$$

For the constrained FDAF without step-normalization, it has $E[\mathbf{A}_{\mathrm{cn}}(k)] = \mathbf{R}_{11}$. Using (50) and ignoring the term related to μ^2, we get

$$E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)\tilde{\mathbf{w}}_{\mathrm{cn}}^T(\infty)]\mathbf{R}_{11} + \mathbf{R}_{11}E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)\tilde{\mathbf{w}}_{\mathrm{cn}}^T(\infty)] = \mu\delta_v^2\bar{\mathbf{P}}\bar{\mathbf{R}}\bar{\mathbf{P}}^T. \tag{A6}$$

Taking the trace of both sides of (A6) and considering the relation $\bar{\mathbf{P}}\bar{\mathbf{R}}\bar{\mathbf{P}}^T = \mathbf{R}_{11}$, one immediately obtains

$$\mathrm{tr}\left(E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)\tilde{\mathbf{w}}_{\mathrm{cn}}^T(\infty)]\mathbf{R}_{11}\right) = \frac{1}{2}\mu\delta_v^2\mathrm{tr}(\mathbf{R}_{11}). \tag{A7}$$

Considering $\mathrm{tr}(\mathbf{R}_{11}) = LR\delta_x^2$, we then obtain the EMSE of the constrained FDAF algorithm without step-normalization from (54)

$$J_{\text{cn,ex}}(\infty) = \frac{1}{R}\text{tr}(E[\tilde{\mathbf{w}}_{\text{cn}}(\infty)\tilde{\mathbf{w}}_{\text{cn}}^T(\infty)]\mathbf{R}_{11}) = \frac{1}{2}\mu L \delta_\nu^2 \delta_x^2. \tag{A8}$$

Note that Eqs. (A3), (A5), and (A8) agree with the results in in [4], [20], but our derivation is easier to understand.

For the constrained FDAF algorithm with step-normalization, it has an approximation $E[\mathbf{A}_{\text{cn}}(k)] = \frac{R}{N}\mathbf{I}_L$. Using (50) and ignoring the μ^2 terms, we get

$$E[\tilde{\mathbf{w}}_{\text{cn}}(\infty)\tilde{\mathbf{w}}_{\text{cn}}^T(\infty)] = \frac{1}{2}\frac{N}{R}\mu\delta_\nu^2\bar{\mathbf{P}}\bar{\mathbf{R}}\bar{\mathbf{P}}^T. \tag{A9}$$

Using the approximation $\mathbf{P}\bar{\mathbf{R}} = \frac{R}{N}\mathbf{I}_L$, we get $\bar{\mathbf{P}}\bar{\mathbf{R}} = [\mathbf{I}_L\ \mathbf{0}_{L\times R}]$, and hence $\bar{\mathbf{P}}\bar{\mathbf{R}}\bar{\mathbf{P}}^T \approx \frac{R}{N}\mathbf{P}_1$. Equation (A9) then becomes

$$E[\tilde{\mathbf{w}}_{\text{cn}}(\infty)\tilde{\mathbf{w}}_{\text{cn}}^T(\infty)] = \frac{1}{2}\mu\delta_\nu^2\mathbf{P}_1. \tag{A10}$$

The steady-state EMSE of the constrained FDAF algorithm with step-normalization can be expressed by

$$J_{\text{cn,ex}}(\infty) = \frac{1}{R}\text{tr}(E[\tilde{\mathbf{w}}_{\text{cn}}(\infty)\tilde{\mathbf{w}}_{\text{cn}}^T(\infty)]\mathbf{R}_{11}) = \frac{1}{2R}\mu\delta_\nu^2\text{tr}(\mathbf{P}_1\mathbf{R}_{11}). \tag{A11}$$

For white noise input, we have $\mathbf{P}_2\mathbf{R}_{21} = \mathbf{0}_L$, and hence $\mathbf{P}_1\mathbf{R}_{11} = E[\mathbf{A}_{\text{cn}}(k)] - \mathbf{P}_2\mathbf{R}_{21} \approx \frac{R}{N}\mathbf{I}_L$. Equation (A.11) then becomes

$$J_{\text{cn,ex}}(\infty) = \frac{1}{2}\frac{L}{N}\mu\delta_\nu^2. \tag{A12}$$

The approximation (A12) is the same as that in [4] and [20]. Note that (A12) has a good approximation when the input is white noise or close to white noise, but (A12) deviates from the simulated result for highly correlated inputs. However, the presented theoretical EMSE (A11) holds well for both white noise and correlated input.

It is observed that the sufficient-length FDAF algorithms with and without step-normalization can achieve the same steady-state EMSE result by step size adjustments, while their difference lies in the convergence speed. The EMSE gap between the two FDAF variants is determined by a factor of L/R, which becomes significant as the block length R increases.

Appendix: B

In this section, we show some variables and matrices used in the formulation for the convenience of the reader (Table 2).

Table 2 The symbol matrices and variables

Symbol	Dimension	Description
M	1×1	The length of the unknown plant
L	1×1	The length of the adaptive filter
Q	1×1	The length of the under-modeling filter
N	1×1	The DFT length
R	1×1	The block length
$x(n)$	1×1	The input signal
$d(n)$	1×1	The desired response
$\mathbf{x}(n)$	$M \times 1$	The input vector
$\mathbf{w}_{\mathrm{opt}}$	$M \times 1$	The impulse response vector of an unknown system
$\mathbf{w}_{\dagger}$	$L \times 1$	The column vector consisting of the first L elements of $\mathbf{w}_{\mathrm{opt}}$
$\mathbf{w}_{*}$	$Q \times 1$	The column vector consisting of the last Q elements of $\mathbf{w}_{\mathrm{opt}}$
$\hat{\mathbf{w}}(k)$	$L \times 1$	The estimated weight vector
$\hat{\mathbf{W}}(k)$	$N \times 1$	The frequency-domain weight vector
$\mathbf{F}$	$N \times N$	The DFT matrix
$\boldsymbol{\mathcal{X}}(k)$	$N \times N$	The frequency-domain input matrix
$\hat{\mathbf{y}}(k)$	$R \times 1$	The block filtering out vector
$\mathbf{d}(k)$	$R \times 1$	The block desired signal vector
$\mathbf{e}(k)$	$R \times 1$	The block error vector
$\mathbf{G}_{10}$	$N \times N$	The constraining matrix
$\mathbf{X}_{\mathrm{c}}(k)$	$N \times N$	A circulant matrix with the first column $\mathbf{x}(k)$
$\mathbf{R}_{\mathrm{c}}$	$N \times N$	A correlation matrix of the input signal $\mathbf{R}_{\mathrm{c}} = E[\mathbf{X}_{\mathrm{c}}^{T}(k)\mathbf{X}_{\mathrm{c}}(k)]$
$\mathbf{P}$	$N \times N$	The inverse of $\mathbf{R}_{\mathrm{c}}$
$\hat{\mathbf{w}}_{\mathrm{un}}(k)$	$N \times 1$	The time-domain weight vector of the unconstrained FDAF
$\tilde{\mathbf{w}}_{\mathrm{un}}(k)$	$N \times 1$	The time-domain weight-error vector of the unconstrained FDAF
$E[\mathbf{A}_{\mathrm{un}}(k)]$	$N \times N$	The matrix than controls the mean weight convergence of the unconstrained FDAF
$\mathbf{H}_{\mathrm{un}}$	$N^2 \times N^2$	The matrix than controls the mean-square convergence of the unconstrained FDAF
$\delta_{\mathrm{un}}(k)$	1×1	The instantaneous MSD of the unconstrained FDAF
$J_{\mathrm{un}}(k)$	1×1	The instantaneous MSE of the unconstrained FDAF
$\hat{\mathbf{w}}_{\mathrm{cn}}(k)$	$L \times 1$	The time-domain weight vector of the constrained FDAF
$\tilde{\mathbf{w}}_{\mathrm{cn}}(k)$	$L \times 1$	The time-domain weight-error vector of the constrained FDAF
$E[\mathbf{A}_{\mathrm{cn}}(k)]$	$L \times L$	The matrix than controls the mean weight convergence of the constrained FDAF
$\mathbf{H}_{\mathrm{cn}}$	$L^2 \times L^2$	The matrix than controls the mean-square convergence of the constrained FDAF
$\delta_{\mathrm{cn}}(k)$	1×1	The instantaneous MSD of the constrained FDAF
$J_{\mathrm{cn}}(k)$	1×1	The instantaneous MSE of the constrained FDAF
$\boldsymbol{\Delta}$	$N \times 1$	The difference between $E[\tilde{\mathbf{w}}_{\mathrm{un}}(\infty)]$ and the optimum Wiener solution
$\mathbf{q}_{\mathrm{un}}(k)$	$N \times 1$	The difference between $\hat{\mathbf{w}}_{\mathrm{un}}(k)$ and its mean steady-state value
$\bar{\boldsymbol{\Delta}}$	$L \times 1$	The difference between $E[\tilde{\mathbf{w}}_{\mathrm{cn}}(\infty)]$ and the optimum Wiener solution
$\mathbf{q}_{\mathrm{cn}}(k)$	$L \times 1$	The difference between $\hat{\mathbf{w}}_{\mathrm{cn}}(k)$ and its mean steady-state value

Declarations

References

1. E.R. Ferrara, Fast implementation of LMS adaptive filters. IEEE Trans. Acoust. Speech Signal Process. **28**(4), 474–475 (1980)
2. G.A. Clark, S.K. Mitra, S.R. Parker, Block implementation of adaptive digital filters. IEEE Trans. Circuits Syst. **28**(3), 584–592 (1981)
3. D. Mansour Jr., Unconstrained frequency-domain adaptive filter. IEEE Trans. Acoust. Speech Signal Process. **30**(5), 726–734 (1982)
4. B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications* (John Wiley & Sons, Hoboken, NJ, USA, 2013)
5. J.J. Shynk, Frequency-domain and multirate adaptive filtering. IEEE Signal Process. Mag. **9**(1), 14–37 (1992)
6. D. Comminiello, A. Nezamdoust, S. Scardapane, M. Scarpiniti, A. Hussain, A. Uncini, A new class of efficient adaptive filters for online nonlinear modeling. IEEE Trans. Syst. Man Cybern. Syst. **52**, 1384–1396 (2022)
7. J. Franzen, T. Fingscheidt, Improved measurement noise covariance estimation for N-channel feedback cancellation based on the frequency domain adaptive kalman filter. In: Proc. IEEE ICASSP, pp. 965–969 (2019)
8. J. Lorente, M. Ferrer, M.D. Diego, A. González, GPU implementation of multichannel adaptive algorithms for local active noise control. IEEE/ACM Trans. Audio Speech Lang. Process. **22**(11), 1624–1635 (2014)
9. M. Schneider, W. Kellermann, The generalized frequency-domain adaptive filtering algorithm as an approximation of the block recursive least-squares algorithm. EURASIP J. Adv. Signal Process. **2016**, 1–15 (2016)
10. J. Lu, K. Chen, X. Qiu, Convergence analysis of the modified frequency-domain block LMS algorithm with guaranteed optimal steady state performance. Signal Process. **132**, 165–169 (2017)
11. F. Yang, J. Yang, Mean-square performance of the modified frequency-domain block LMS algorithm. Signal Process. **163**(18–25), 596–604 (2019)
12. K. Mayyas, Performance analysis of the deficient length LMS adaptive algorithm. IEEE Trans. Signal Process. **53**(8), 2727–2734 (2005)
13. G. Enzner, S. Voit, Hybrid-frequency-resolution adaptive Kalman filter for online identification of long acoustic responses with low input-output latency. IEEE/ACM Trans. Audio Speech Lang. Process. **31**, 3550–3563 (2023)
14. T. Haubner, A. Brendel, W. Kellermann, End-to-end deep learning based adaptation control for frequency-domain adaptive system identification. In: Proc. IEEE ICASSP, pp. 766–770 (2022)
15. J. Casebeer, N. Bryan, P. Smaragdis, Meta-AF: meta-learning for adaptive filters. IEEE/ACM Trans. Audio Speech Lang. Process. **31**, 355–370 (2023)
16. A. Carini, G.L. Sicuranza, Transient and steady-state analysis of filtered-x affine projection algorithms. IEEE Trans. Signal Process. **54**(2), 665–678 (2006)
17. E. Eweda, N.J. Bershad, J.C.M. Bermudez, Stochastic analysis of the LMS and NLMS algorithms for cyclostationary white Gaussian and non-Gaussian inputs. IEEE Trans. Signal Process. **66**(18), 4753–4765 (2018)
18. A. Feuer, Performance analysis of the block least mean square algorithm. IEEE Trans. Circuits Syst. **32**(9), 960–963 (1985)
19. P.C.W. Sommen, P.J.V. Gerwen, H.J. Kotmans, A.J.E.M. Janssen, Convergence analysis of a frequency domain adaptive filter with exponential power averaging and generalized window function. IEEE Trans. Circuits Syst. **34**(7), 788–798 (1987)
20. J.C. Lee, C.K. Un, Performance analysis of frequency domain block LMS adaptive digital filters. IEEE Trans. Circuits Syst. **36**(2), 173–189 (1989)
21. A. Feuer, R. Cristi, On the steady state performance of frequency domain LMS algorithms. IEEE Trans. Signal Process. **41**(1), 419–423 (1993)
22. B. Farhang-Boroujeny, K.S. Chan, Analysis of the frequency-domain block LMS algorithm. IEEE Trans. Signal Process. **48**(8), 2332–2342 (2000)
23. M. Wu, J. Yang, Y. Xu, X. Qiu, Steady-state solution of the deficient length constrained FBLMS algorithm. IEEE Trans. Signal Process. **60**(12), 6681–6687 (2012)
24. X. Zhang, Y. Xia, C. Li, L. Yang, D.P. Mandic, Analysis of the unconstrained frequency-domain block LMS for second-order noncircular inputs. IEEE Trans. Signal Process. **67**(15), 3970–3984 (2019)
25. F. Yang, G. Enzner, J. Yang, A unified approach to the statistical convergence analysis of frequency-domain adaptive filters. IEEE Trans. Signal Process. **67**(7), 1785–1796 (2019)
26. F. Yang, J. Yang, Convergence analysis of deficient-length frequency-domain adaptive filters. IEEE Trans. Circuits Syst. I **66**(11), 4242–4255 (2019)
27. F. Yang, G. Enzner, J. Yang, New insights into convergence theory of constrained frequency-domain adaptive filters. Circuits Syst. Signal Process. **40**, 2076–2090 (2021)

28. F. Yang, G. Enzner, J. Yang, On the convergence behavior of partitioned-block frequency-domain adaptive filters. IEEE Trans. Signal Process. **69**, 4906–4920 (2021)
29. F. Yang, Analysis of deficient-length partitioned-block frequency-domain adaptive filters. IEEE/ACM Trans. Audio Speech Lang. Process. **30**, 456–467 (2022)
30. X. Li, W.K. Jenkins, The comparison of the constrained and unconstrained frequency-domain block-LMS adaptive algorithms. IEEE Signal Process. Lett. **44**(7), 1813–1816 (1996)
31. D.S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas* (Princeton University Press, Princeton, NJ, 2009)
32. Y. Al-Naffouri, A.H. Sayed, Transient analysis of data-normalized adaptive filters. IEEE Trans. Signal Process. **51**(3), 639–652 (2003)
33. F. Yang, J. Yang, Mean-square performance of the modified frequency-domain block LMS algorithm. Signal Process. **163**(18–25), 596–604 (2019)
34. K. Mayyas, Performance analysis of the deficient length LMS adaptive algorithm. IEEE Trans. Signal Process. **53**(8), 2727–2734 (2005)

Publisher's Note