Multi-Layered Recursive Least Squares for Time-Varying System Identification

Mohammad Towliat, Zheng Guo, Leonard J. Cimini, *Life Fellow, IEEE*, Xiang-Gen Xia, *Fellow, IEEE*, and Aijun Song, *Member, IEEE*

***Abstract*—Traditional recursive least squares (RLS) adaptive filtering is widely used to estimate the impulse responses (IR) of an unknown system. Nevertheless, the RLS estimator shows poor performance when tracking rapidly time-varying systems. In this paper, we propose a multi-layered RLS (m-RLS) estimator to address this concern. The m-RLS estimator is composed of multiple RLS estimators, each of which is employed to estimate and eliminate the misadjustment of the previous layer. It is shown that the mean squared error (MSE) of the m-RLS estimate can be minimized by selecting the optimum number of layers. We provide a method to determine the optimum number of layers. A low-complexity implementation of m-RLS is discussed and it is indicated that the complexity order of the proposed estimator can be reduced to $\mathcal{O}(M)$, where M is the IR length. Through simulations, we show that m-RLS outperforms the classic RLS and the RLS methods with a variable forgetting factor.**

***Index Terms*—Echo cancellation, mean square error, recursive least squares, system identification, time-varying systems.**

I. Introduction

RECURSIVE least squares (RLS) adaptive filtering is one of the most appealing frameworks for applications, such as system identification, inverse modeling, channel equalization, etc. [1]. The RLS algorithm, by using the Newton search, leads to a faster convergence than the steepest-decent based algorithms, e.g. least mean squares (LMS) [2]. It has been shown that low-complexity versions of the RLS algorithm, including the fast transversal RLS (FTRLS) [3]–[5], and the RLS lattice (RLSL) [6], [7] algorithms have almost the same convergence ability as the classic RLS, when tracking a time-varying system. It is noteworthy that the adaptability of RLS filtering is based on the fact that the estimation error is a zero-mean white Gaussian variable and, thus, the algorithm recursively minimizes a maximum likelihood objective function. Accordingly, a regularized objective function can be introduced to preserve the tracking ability of RLS in a non-Gaussian environment [8]–[11].

In system identification, RLS is used to estimate the impulse response (IR) of an unknown system, given the input and the output of the system as the desired signal. Speaking of the convergence behavior of the RLS system identifiers, it has been shown in [12] that the estimation error is caused by two types of factors, including the lag error (LE) and the estimation noise (EN). The LE is due to the time-variation of the system. Since the RLS estimation is based on the statistical averages, it inherits a latency when tracking a time-varying IR. Thus, the LE can be problematic when estimating a fast time-varying system. On the other hand, the EN is caused by the exponentially windowing nature of RLS. Using a forgetting factor limits the actual observation window size and, even under steady-state conditions, the IR estimate is contaminated by a misadjustment.

Regarding the forgetting factor, there is a trade-off between the LE and the EN. A small forgetting factor raises the RLS tracking agility (i.e., a lower LE) but also develops a higher steady-state misadjustment (i.e., a higher EN). Contrarily, a larger forgetting factor leads to poor adaptability, but a better steady-state performance [13]. As a result, numerous studies in this field are dedicated to infer an RLS algorithm with a variable forgetting factor (RLS-VFF). For instance, see [14]–[19]. The main idea in these works is to recognize the changes of the system's IR by measuring the changes in the error and desired signal's statistics, then setting a small forgetting factor during the IR changes, and a large forgetting factor during the steady-state periods.

In this paper, we look at the problem of low error system identification from another point of view. We propose the multi-layered RLS (m-RLS) algorithm that minimizes the sum of the LE and the EN, instead of minimizing either of the errors alone (what is performed in RLS-VFF approaches). The m-RLS approach is composed of multiple connected RLS estimators. At each time instance, the a posteriori error of the previous RLS is considered as the desired signal in the next RLS. From this perspective, m-RLS falls in the category of data-reuse adaptive algorithms [20].

Data-reusing idea was firstly introduced by Shaffer and Williams to improve the LMS convergence [21]. In data-reuse LMS (DR-LMS), multiple LMS blocks are cascaded. The filter weights of all blocks are updated over the same present data sample. However, it can be shown that the convergence improvement of DR-LMS method in [21] is not significant and

Manuscript received October 21, 2021; revised March 8, 2022 and April 16, 2022; accepted April 20, 2022. Date of publication April 29, 2022; date of current version May 16, 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ketan Rajawat. This work was supported in part by National Science Foundation under Grants CNS-1704097 and CNS-1704076. *(Corresponding author: Mohammad Towliat.)*

Mohammad Towliat, Leonard J. Cimini, and Xiang-Gen Xia are with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: mtowliat@udel.edu; cimini@udel.edu; xianggen@udel.edu).

Zheng Guo and Aijun Song are with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL 35487 USA (e-mail: zguo18@crimson.ua.edu; song@eng.ua.edu).

Digital Object Identifier 10.1109/TSP.2022.3170708

the performance lies between that of LMS and normalized LMS (NLMS) [22]. Therefore, modified DR-LMS algorithms are proposed to improve the performance of the original DR-LMS by engaging the present and the past data samples [23]–[27]. Nevertheless, all of these studies only cover data-reusing with LMS and, to the best of our knowledge, the proposed m-RLS method in this paper is the first attempt for implementing data-reusing idea with RLS.

We show that, in m-RLS, the total estimation error (sum of the LE and the EN) is a function of the number of layers. Therefore, by using the optimum number of layers, the estimation error of m-RLS can be less than that of RLS. The optimum number of layers tightly depends on the coherence length of the system's IR (the time range in which the IR approximately remains invariant) and signal to noise ratio (SNR). We provide a solution for determining the optimal number of layers. It must be mentioned that in all previous DR-LMS methods, the number of layers is fixed and there is no existing effort to determine the optimal one.

Next, the complexity of m-RLS is discussed by comparing the number of multiplications, additions, and divisions of two types of implementations, including the classic and the transversal dichotomous coordinate descent (DCD) [28] techniques. We show that using transversal DCD implementation significantly reduces the complexity of m-RLS and makes the proposed method a promising approach for the applications where superior adaptability is desired with the expense of a reasonable higher complexity.

In our simulations, we evaluate the performance of m-RLS and show that, in a rapidly time-varying system, the proposed estimator leads to a lower estimation error compared to that of the classic RLS and two different RLS-VFF techniques.

The rest of this paper is organized as follows. In Section II, we investigate the RLS system identification and the associated error when tracking a time-varying system. In Section III, we propose the m-RLS algorithm and discuss how its performance is dominated by the coherence length of the system. Then, we provide a solution for the optimum number of layers that minimizes the estimation error. Section IV discusses the implementation and complexity of m-RLS. The simulation results are brought in Section V. Finally, Section VI concludes this paper.

Notations: Matrices are denoted by boldface uppercase letters (e.g. $\mathbf{A}$), vectors are indicated by boldface lowercase letters (e.g. $\mathbf{a}$), and scalar quantities are presented by normal letters (e.g. a or A). $\mathbf{I}_M$ is the $M \times M$ identity matrix, and $\mathbb{E}$ is the mathematical expectation. $\|\mathbf{a}\|$ indicates the l_2 norm of vector $\mathbf{a}$. Finally, the superscripts $(.)^T$, $(.)^H$, and $(.)^*$ indicate transpose, conjugate transpose, and conjugate operators, respectively.

II. RLS System Identification

A. RLS Algorithm

Consider $x[n]$ as the input sequence to an unknown time-varying system with the IR vector $\mathbf{h}[n] = [h_0[n], \ldots, h_{M-1}[n]]^H$, where M is the length of the IR. The system's output is given as

$$d[n] = \mathbf{h}^H[n]\mathbf{x}[n] + w[n], \tag{1}$$

in which $\mathbf{x}[n] = [x[n], \ldots, x[n-M+1]]^T$ is the input signal vector, and $w[n]$ is an AWGN noise with variance σ_w^2. Given the input and the output signals, the system's IR can be estimated by using the well-known RLS algorithm as below [2]

$$e[n] = d[n] - \hat{\mathbf{h}}^H[n-1]\mathbf{x}[n] \tag{2a}$$

$$\mathbf{k}[n] = (\lambda + \mathbf{x}^H[n]\mathbf{P}[n-1]\mathbf{x}[n])^{-1}\mathbf{P}[n-1]\mathbf{x}[n] \tag{2b}$$

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + e^*[n]\mathbf{k}[n] \tag{2c}$$

$$\mathbf{P}[n] = \lambda^{-1}\left(\mathbf{I}_M - \mathbf{k}[n]\mathbf{x}^H[n]\right)\mathbf{P}[n-1], \tag{2d}$$

where $\hat{\mathbf{h}}[n]$ is the estimate of the IR, $d[n]$ is so called the desired signal, $0 < \lambda < 1$ is the forgetting factor, $e[n]$ denotes the a priori error, and $\mathbf{k}[n]$ is the gain vector. It can be shown that in this algorithm, $\mathbf{k}[n] = \mathbf{P}[n]\mathbf{x}[n]$, where $\mathbf{P}[n] = \mathbf{R}^{-1}[n]$, and $\mathbf{R}[n] = \sum_{k=0}^{n} \lambda^{n-k}\mathbf{x}[k]\mathbf{x}^H[k]$ [1].

B. Estimation Error in RLS

In this section, we inquire about the accuracy of the RLS estimate, $\hat{\mathbf{h}}[n]$, by comparing it to the varying IR vector, $\mathbf{h}[n]$. For this evaluation, we use the mean squared error (MSE) measure. By assuming that $\mathbb{E}\|\mathbf{h}[n]\|^2 = 1$, the normalized MSE is defined as

$$\mu_{\text{RLS}} = \mathbb{E}\left\|\mathbf{h}[n] - \hat{\mathbf{h}}[n]\right\|^2. \tag{3}$$

To facilitate analyzing μ_{RLS}, we hold the following assumptions: 1) the input signal is a zero-mean uncorrelated sequence, with a normalized power; 2) the IR estimate is independent of the input signal [12].

Since $\mathbb{E}\{\mathbf{x}[n]\mathbf{x}^H[n]\} = \mathbf{I}_M$, for a sufficiently large n, we have $\mathbf{R}[n] = \sum_{k=0}^{n} \lambda^{n-k} = \frac{1}{\varepsilon}\mathbf{I}_M$, where $\varepsilon = 1 - \lambda$. As a result, $\mathbf{P}[n] = \varepsilon\mathbf{I}_M$ and $\mathbf{k}[n] = \varepsilon\mathbf{x}[n]$. On the other hand, by replacing (1) in (2a), the a priori error is given as $e[n] = (\mathbf{h}[n] - \hat{\mathbf{h}}[n-1])^H\mathbf{x}[n] + w[n]$. Nesting $\mathbf{k}[n]$ and $e[n]$ in (2c), the IR estimate at time n becomes

$$\hat{\mathbf{h}}[n] = \mathbf{h}[n] - \mathbf{\Theta}[n]\left(\mathbf{h}[n] - \hat{\mathbf{h}}[n-1]\right) + \boldsymbol{\gamma}[n], \tag{4}$$

where $\mathbf{\Theta}[n] = \mathbf{I}_M - \varepsilon\mathbf{x}[n]\mathbf{x}^H[n]$ and $\boldsymbol{\gamma}[n] = \varepsilon w^*[n]\mathbf{x}[n]$. Similar to (4), the IR estimate at the past times can be written as $\hat{\mathbf{h}}[n-k] = \mathbf{h}[n-k] - \mathbf{\Theta}[n-k](\mathbf{h}[n-k] - \hat{\mathbf{h}}[n-k-1]) + \boldsymbol{\gamma}[n-k]$. By using this recursion, for $k = 0, \ldots, N-1$, one can expand (4) as

$$\hat{\mathbf{h}}[n] = \mathbf{A}[n]\mathbf{h}[n] + \mathbf{B}[n]\hat{\mathbf{h}}[n-N] + \mathbf{c}[n], \tag{5}$$

where

$$\mathbf{A}[n] = \varepsilon\mathbf{x}[n]\mathbf{x}^H[n] + \varepsilon\sum_{k=1}^{N-1}\left(\prod_{i=0}^{k-1}\mathbf{\Theta}[n-i]\right)\mathbf{x}[n-k]\mathbf{x}^H[n-k];$$

$$\mathbf{B}[n] = \prod_{k=0}^{N-1}\mathbf{\Theta}[n-k];$$

$$\mathbf{c}[n] = \boldsymbol{\gamma}[n] + \sum_{k=1}^{N-1}\left(\prod_{i=0}^{k-1}\mathbf{\Theta}[n-i]\right)\boldsymbol{\gamma}[n-k], \tag{6}$$

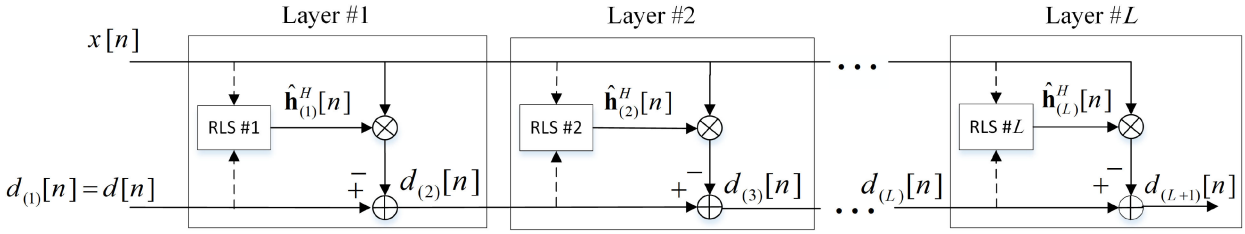


Fig. 1. The m-RLS system identifier structure.

and N is the coherence length of the IR (the time range for which the IR remains invariant), so that $\mathbf{h}[n-k] = \mathbf{h}[n]$, for $k = 0, \ldots, N-1$. Note that, for a relatively slow time-varying IR, N is large while, for a fast time-varying IR, N can be small.

By doing some algebra, it can be shown that $\mathbf{I}_M - \mathbf{A}[n] = \mathbf{B}[n]$. As a result, the difference between the true IR and its RLS estimate in (5) becomes

$$\mathbf{h}[n] - \hat{\mathbf{h}}[n] = \mathbf{B}[n](\mathbf{h}[n] - \hat{\mathbf{h}}[n-N]) - \mathbf{c}[n]. \tag{7}$$

On the right-hand side of (7), $\mathbf{c}[n]$ is the effect of the noise samples $w[n-k]$ for $k = 0, \ldots, N-1$ while, $\hat{\mathbf{h}}[n-N]$ depends on $w[n-k]$ for $k \geq N$. Since $\mathbf{c}[n]$ is uncorrelated with both $\hat{\mathbf{h}}[n-N]$ and $\mathbf{h}[n]$, replacing (7) in (3) leads to

$$\mu_{\text{RLS}} = \mathbb{E}\left\|\mathbf{B}[n](\mathbf{h}[n] - \hat{\mathbf{h}}[n-N])\right\|^2 + \mathbb{E}\|\mathbf{c}[n]\|^2. \tag{8}$$

We have shown in Appendix A that (8) can be simplified to

$$\begin{aligned}\mu_{\text{RLS}} &= \rho^N \mathbb{E}\|\mathbf{h}[n]\|^2 + \left(1-\rho^N\right)\psi\sigma_w^2 \\ &= \rho^N + \left(1-\rho^N\right)\psi\sigma_w^2,\end{aligned} \tag{9}$$

where $\rho = 1 - 2\varepsilon + \varepsilon^2 M$, and $\psi = \frac{\varepsilon^2 M}{1-\rho}$. It is noteworthy that, when the IR is time-invariant, (9) coincides with the known results in [1] and [12].

According to (9), the RLS estimation error is contributed by two terms [12], [29]. The first term is the LE caused by the RLS lag of tracking a time-varying IR. The second term is the EN, which is the effect of exponentially windowing nature of RLS, so that by letting $\lambda = 1$, this term is zero.

Speaking of the LE, we should mention that it originates from the fact that the RLS estimation is performed based on statistical averages which take several time samples to converge. In a time-varying system, the averaging process can be inadequate because of rapid variations. As mentioned in [3] and [13], for an stable convergence, the forgetting factor falls within the range $1 - \frac{2}{M} < \lambda < 1$. In this range, one can see that $0 < \rho < 1$. Accordingly, when the IR is relatively slow time-varying with a large N, the contribution of the LE in (9) is small and the MSE is mainly governed by the EN. On the other hand, for a rapidly time-varying IR with a small N, the LE is not negligible and can potentially be higher than that of the EN.

Addressing the problem of minimizing the MSE in tracking a rapidly time-varying system, we propose the m-RLS algorithm in the next section.

III. Multi-Layered RLS System Identification

In this section, we propose the m-RLS estimator to minimize the sum of the LE and the EN. The proposed m-RLS estimator is composed of multiple layers. At each layer, the effective IR (which is the estimation error of the previous layer) is estimated and eliminated from the desired signal by utilizing a separate RLS estimator. We show that, by using the optimum number of layers, the m-RLS strategy can lead to a lower MSE error than RLS.

A. Multi-Layered RLS Algorithm

Fig. 1 shows the m-RLS structure with L layers. Similar to the RLS estimator, m-RLS is used to estimate the IR given $d[n]$ and $\mathbf{x}[n]$.

In the first layer, the desired signal is $d[n]$, thus, we represent it with $d_{(1)}[n] = d[n]$, where the subscription denotes the layer's index. Accordingly, we can rewrite (1) as

$$d_{(1)}[n] = \mathbf{h}_{(1)}^H[n]\mathbf{x}[n] + w[n], \tag{10}$$

where $\mathbf{h}_{(1)}[n] = \mathbf{h}[n]$ is the effective IR at the first layer. RLS #1 estimates $\mathbf{h}_{(1)}[n]$ given $d_{(1)}[n]$ and $\mathbf{x}[n]$. We denote this estimate as $\hat{\mathbf{h}}_{(1)}[n]$. Then, the a posteriori error [1] in RLS #1 is given as

$$\begin{aligned}d_{(2)}[n] &= d_{(1)}[n] - \hat{\mathbf{h}}_{(1)}^H[n]\mathbf{x}[n] \\ &= \left(\mathbf{h}_{(1)}[n] - \hat{\mathbf{h}}_{(1)}[n]\right)^H \mathbf{x}[n] + w[n] \\ &= \mathbf{h}_{(2)}^H[n]\mathbf{x}[n] + w[n].\end{aligned} \tag{11}$$

where $\mathbf{h}_{(2)}[n] = \mathbf{h}_{(1)}[n] - \hat{\mathbf{h}}_{(1)}[n]$, is the error vector of estimating $\mathbf{h}_{(1)}[n]$ by using RLS #1. Due to the LE in RLS #1, $\mathbf{h}_{(2)}[n]$ is not a pure noise vector and, thus, $d_{(2)}[n]$ has a correlation with the input signal $\mathbf{x}[n]$.

At the second layer, RLS #2 is employed to provide an estimate of the effective IR $\mathbf{h}_{(2)}[n]$. To this end, $d_{(2)}[n]$ is taken as the desired signal. We denote the estimate of $\mathbf{h}_{(2)}[n]$ by $\hat{\mathbf{h}}_{(2)}[n]$. Similar to the first layer, the a posteriori error in RLS #2 becomes

$$\begin{aligned}d_{(3)}[n] &= d_{(2)}[n] - \hat{\mathbf{h}}_{(2)}^H[n]\mathbf{x}[n] \\ &= \mathbf{h}_{(3)}^H[n]\mathbf{x}[n] + w[n],\end{aligned} \tag{12}$$

where $\mathbf{h}_{(3)}[n] = \mathbf{h}_{(2)}[n] - \hat{\mathbf{h}}_{(2)}[n]$.

The same process is performed for all layers such that, at the lth layer (for $l = 1, \ldots, L$), the a posteriori error is

$$\begin{aligned}d_{(l+1)}[n] &= d_{(l)}[n] - \hat{\mathbf{h}}_{(l)}^H[n]\mathbf{x}[n] \\ &= \mathbf{h}_{(l+1)}^H[n]\mathbf{x}[n] + w[n],\end{aligned} \tag{13}$$

where $d_{(l)}[n] = \mathbf{h}_{(l)}^H[n]\mathbf{x}[n] + w[n]$, and

$$\mathbf{h}_{(l+1)}[n] = \mathbf{h}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n]. \tag{14}$$

Finally, the overall IR estimation by m-RLS is given as the sum of the estimates at all layers, that is

$$\hat{\mathbf{h}}[n] = \sum_{l=1}^{L} \hat{\mathbf{h}}_{(l)}[n]. \tag{15}$$

B. Estimation Error in Multi-Layered RLS

In this section, we evaluate the accuracy of the proposed m-RLS estimator. To facilitate the derivations, we hold the same assumptions as those in Section II-B.

The overall IR estimate by m-RLS is given in (15). The normalized MSE of this estimate is obtained as

$$\mu_{\mathrm{m-RLS}}(L) = \mathbb{E}\left\|\mathbf{h}[n] - \hat{\mathbf{h}}[n]\right\|^2. \tag{16}$$

In (16), we represent the MSE as a function of L because, according to (15), $\hat{\mathbf{h}}[n]$ is introduced based on the number of layers, L. By considering the recursions in (14), substituting (15) in (16) results in

$$\mu_{\mathrm{m-RLS}}(L) = \mathbb{E}\left\|\mathbf{h}_{(L+1)}[n]\right\|^2. \tag{17}$$

Eq. (17) indicates that the MSE of m-RLS is equivalent to the average power of the last effective IR, $\mathbf{h}_{(L+1)}[n]$. Thus, let us first evaluate the general case $\mathbb{E}\|\mathbf{h}_{(l+1)}[n]\|^2$, for $l = 1, \ldots L$. In the proposed estimator, each RLS operates separately; therefore, similar to (7), we can represent the difference between the effective IR and its estimate at the lth layer as

$$\begin{aligned}\mathbf{h}_{(l+1)}[n] &= \mathbf{h}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n] \\ &= \mathbf{B}_{(l)}[n](\mathbf{h}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n - N_{(l)}]) - \mathbf{c}_{(l)}[n], \end{aligned}\tag{18}$$

where

$$\mathbf{B}_{(l)}[n] = \prod_{k=0}^{N_{(l)}-1} \mathbf{\Theta}[n-k];$$

$$\mathbf{c}_{(l)}[n] = \boldsymbol{\gamma}[n] + \sum_{k=1}^{N_{(l)}-1} \left(\prod_{i=0}^{k-1} \mathbf{\Theta}[n-i]\right)\boldsymbol{\gamma}[n-k], \tag{19}$$

and $N_{(l)}$ is the coherence length of $\mathbf{h}_{(l)}[n]$.

In (18), $\mathbf{c}_{(l)}[n]$ is based on the noise samples $w[n-k]$ for $k = 0, \ldots, N_{(l)} - 1$, whereas $\hat{\mathbf{h}}_{(l)}[n - N_{(l)}]$ depends on $w[n-k]$ for $k \geq N_{(l)}$. Thus, $\mathbf{c}_{(l)}[n]$ is uncorrelated with $\hat{\mathbf{h}}_{(l)}[n - N_{(l)}]$. On the other hand, $\mathbf{c}_{(l)}[n]$ has a correlation with $\mathbf{h}_{(l)}[n]$, except for $l = 1$ (see (8)). The correlation between $\mathbf{c}_{(l)}[n]$ and $\mathbf{h}_{(l)}[n]$ for $l > 1$, leads to

$$\begin{aligned}\mathbb{E}\left\|\mathbf{h}_{(l+1)}[n]\right\|^2 = \mathbb{E}\left\|\mathbf{B}_{(l)}[n](\mathbf{h}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n - N_{(l)}])\right\|^2 \\ + \mathbb{E}\left\|\mathbf{c}_{(l)}[n]\right\|^2 + u(l),\end{aligned}\tag{20}$$

where $u(l) = -2\mathrm{Re}\mathbb{E}\{\mathbf{c}_{(l)}^H[n]\mathbf{B}_{(l)}[n]\mathbf{h}_{(l)}[n]\}$ is the cross-correlation term. Following the same steps as those in Appendix A, the average power of $\mathbf{h}_{(l+1)}[n]$ can be represented as

$$\begin{aligned}\mathbb{E}\left\|\mathbf{h}_{(l+1)}[n]\right\|^2 = \rho^{N_{(l)}}\mathbb{E}\left\|\mathbf{h}_{(l)}[n]\right\|^2 + \left(1 - \rho^{N_{(l)}}\right)\psi\sigma_w^2 \\ + u(l).\end{aligned}\tag{21}$$

Eq. (21) exhibits the recursive relation between the average powers of two consecutive effective IR vectors $\mathbf{h}_{(l+1)}[n]$ and $\mathbf{h}_{(l)}[n]$, for $l = 1, \ldots, L$. By using this recursion, one can expand $\mathbb{E}\|\mathbf{h}_{(L+1)}[n]\|^2$ based on $\mathbb{E}\|\mathbf{h}_{(1)}[n]\|^2 = 1$ and obtain $\mu_{\mathrm{m-RLS}}(L)$ in (17) as

$$\mu_{\mathrm{m-RLS}}(L) = \mathbb{E}\left\|\mathbf{h}_{(L+1)}[n]\right\|^2 = \prod_{l=1}^{L} \rho^{N_{(l)}} + v(L). \tag{22}$$

where $v(L) = \rho^{N_{(L)}} v(L-1) + (1 - \rho^{N_{(L)}})\psi\sigma_w^2 + u(L)$ is a recursive function initiated as $v(1) = (1 - \rho^{N_{(1)}})\psi\sigma_w^2$ and $u(1) = 0$.

According to (22), $\mu_{\mathrm{m-RLS}}(L)$ is the sum of two terms. $\prod_{l=1}^{L} \rho^{N_{(l)}}$ is the LE, originating from the system time-variations, and $v(L)$ is the EN, representing the noise effect. The values of the LE and the EN tightly depend on the coherence lengths of the effective IRs, $N_{(l)}$, for $l = 1, \ldots, L$. Thus, prior to discussing $\mu_{\mathrm{m-RLS}}(L)$, let us first investigate $N_{(l)}$ in the next section.

C. Coherence Lengths of Effective IRs

In this section, we investigate the coherence lengths of the effective IRs when SNR is high. Let $\varphi_{(l+1)}[m]$, for $m \geq 0$, be the normalized autocorrelation function (ACF) of $\mathbf{h}_{(l+1)}[n]$ calculated as

$$\varphi_{(l+1)}[m] = \frac{\mathbb{E}\{\mathbf{h}_{(l+1)}^H[n]\mathbf{h}_{(l+1)}[n-m]\}}{\mathbb{E}\left\|\mathbf{h}_{(l+1)}[n]\right\|^2}. \tag{23}$$

In (23), it is assumed that all tap-weights in the effective IR vector share the same ACF. When the SNR is high (i.e., σ_w^2 is small and negligible), replacing (18) in (23) gives us (see Appendix B)

$$\begin{aligned}&\varphi_{(l+1)}[m] = \\ &\left(2\varphi_{(l)}[m] - \varphi_{(l)}[m - N_{(l)}] - \varphi_{(l)}[m + N_{(l)}]\right) q_{(l)}[m],\end{aligned}\tag{24}$$

where

$$q_{(l)}[m] = \begin{cases} \left(\frac{\lambda^2}{\rho}\right)^m; \text{ for } 0 \leq m \leq N_{(l)} \\ \left(\frac{\lambda^2}{\rho}\right)^{N_{(l)}}; \text{ for } N_{(l)} < m, \end{cases} \tag{25}$$

and $\varphi_{(l)}[m]$ is the normalized ACF of $\mathbf{h}_{(l)}[n]$.

Given the normalized ACF of a random variable, the coherence length of the variable is defined as the interval for which, the normalized ACF is greater than 0.5 [30]. By using this fact, we show in Appendix C that, if $\varphi_{(l)}[m]$ has an exponential shape as $\varphi_{(l)}[m] = \exp(-\alpha m)$, with $\alpha > 0$, then (24) leads to the conclusion that $\varphi_{(l+1)}[m]$ also inherits an exponential form expressed as $\varphi_{(l+1)}[m] \approx \exp(-\beta m)$, for $0 \leq m \leq N_{(l)}$, where $\beta = 3\alpha + g$ and $g = \log(\frac{\rho}{\lambda^2})$.

By letting $\varphi_{(l)}[N_{(l)}] = 0.5$ and $\varphi_{(l+1)}[N_{(l+1)}] = 0.5$, the exponential shapes of $\varphi_{(l)}[m]$ and $\varphi_{(l+1)}[m]$ lead to $N_{(l)} = \frac{\log 2}{\alpha}$ and $N_{(l+1)} \approx \frac{\log 2}{\beta} = \frac{\log 2}{3\alpha + g}$, respectively. Combining the last two achievements gives

$$N_{(l+1)} \approx \left\lceil \frac{N_{(l)} \log 2}{N_{(l)} g + 3 \log 2} \right\rceil. \tag{26}$$

where $\lceil \cdot \rceil$ denotes the ceiling operator.

Eq. (26) highlights two important points about the coherence lengths. First, it shows $N_{(l+1)} \leq N_{(l)}$ indicating that $\mathbf{h}_{(l+1)}[n]$ is more time-varying than $\mathbf{h}_{(l)}[n]$. To physically explain this

achievement, consider (14), which demonstrates that the variations in $\mathbf{h}_{(l+1)}[n]$ is the superposition the variations in $\mathbf{h}_{(l)}[n]$ and $\hat{\mathbf{h}}_{(l)}[n]$. Because of the lag in $\hat{\mathbf{h}}_{(l)}[n]$, its variation can be considered almost noncoherent with that of $\mathbf{h}_{(l)}[n]$ in a specific sample time. As a result, $\mathbf{h}_{(l+1)}[n]$ becomes faster fluctuating than $\mathbf{h}_{(l)}[n]$.

The second point from (26) is that, by using this recursion, all coherence lengths can be approximately expressed based on the coherence lengths of the effective IR at the first layer, $N_{(1)} = N$ (note that the effective IR at the first layer, $\mathbf{h}_{(1)}[n]$, is equivalent to the system's IR, $\mathbf{h}[n]$, with the coherence length N).

D. Optimal Number of Layers

Eq. (22) shows that both the LE and the EN in m-RLS are functions of the number of layers, L. The optimum number of layers, $L_{\rm opt}$, is where the sum of the LE and the EN (i.e., the total MSE in (22)) is minimized, that is

$$L_{\rm opt} = \arg\min_{l} \quad \mathbb{E}\big\|\mathbf{h}_{(l+1)}[n]\big\|^2 \quad \text{s.t.} \quad 1 \le l \le L_{\max}, \tag{27}$$

where $L_{\max}$ is the maximum allowed number of layers to keep the complexity of m-RLS bounded. $L_{\max}$ should be selected so that $N_{(L_{\max})} \approx N_{(L_{\max}+1)}$. To do so, we can use the approximation in (26). This criterion implies that increasing the number of layers more than $L_{\max}$ has a trivial effect on the performance.

Note that by comparing (22) and (9), it is seen that $\mu_{\rm m-RLS}(1) = \mu_{\rm RLS}$. This indicates that m-RLS with one layer is identical to an RLS estimator. With this respect, one can see that $L_{\rm opt}$ in (27) leads to

$$\mu_{\rm RLS} \ge \mu_{\rm m-RLS}(L_{\rm opt}). \tag{28}$$

According to (28), by setting the number of layers to $L_{\rm opt}$, m-RLS has a more accurate estimate than RLS, if $L_{\rm opt} > 1$. The equality holds only for the case where $L_{\rm opt} = 1$.

As mentioned in Section III-C, $N_{(l)}$ in (22) can be introduced based on N. As a result, one can realize that $L_{\rm opt}$ in (27) mainly depends on two parameters, including the coherence length of the system's IR, N, and the noise power, σ_w^2. Later, in Section V, we infer that the solution of (27) ends to a larger $L_{\rm opt}$, for either a smaller N or a smaller σ_w^2. Alternatively stated, for either a relatively faster time-varying systems or a higher SNR, $L_{\rm opt}$ becomes larger.

Nevertheless, it is difficult to provide a closed-form solution for (27). Instead, we propose a numerical method to determine $L_{\rm opt}$. To this purpose, in Appendix D, we show that $\mathbb{E}\|\mathbf{h}_{(l+1)}[n]\|^2$ can be rewritten based on the average power of the corresponding a posteriori error as

$$\mathbb{E}\big\|\mathbf{h}_{(l+1)}[n]\big\|^2 = \mathbb{E}\big|d_{(l+1)}[n]\big|^2 + (1 - 2(1-\varepsilon M)^l)\sigma_w^2. \tag{29}$$

Given the noise power, the $(1 - 2(1-\varepsilon M)^l)\sigma_w^2$ term is known. The $\mathbb{E}|d_{(l+1)}[n]|^2$ term is the average power of the a posteriori error at the lth layer and can be numerically estimated as

$$\pi_{(l+1)}[n] = (1-z)\pi_{(l+1)}[n-1] + z\big|d_{(l+1)}[n]\big|^2, \tag{30}$$

where $\pi_{(l+1)}[n]$ is the real-time estimate of $\mathbb{E}|d_{(l+1)}[n]|^2$, $z = 1/(\mathcal{K}M)$ is a small weighting factor, and $\mathcal{K}$ is an integer [14].

By substituting (29) and (30) into (27), the problem of finding the optimum number of layers, at time n, can be reduced to

$$L_{\rm opt} = \arg\min_{l} \quad J_{(l+1)}[n], \quad \text{s.t.} \quad 1 \le l \le L_{\max} \tag{31}$$

where $J_{(l+1)}[n] = \pi_{(l+1)}[n] - 2(1-\varepsilon M)^l\sigma_w^2$. According to (31), at any time sample n, $J_{(l+1)}[n]$ needs to be calculated for $l = 1, \ldots, L_{\max}$. Then, $L_{\rm opt}$ corresponds to the smallest $J_{(l+1)}[n]$.

IV. Implementation and Complexity Analysis

Each layer of m-RLS is equipped with a separate RLS estimator as described in (2a)–(2d). However, since $\mathbf{k}[n]$ and $\mathbf{P}[n]$ only depend on the forgetting factor and the input signal, all layers in m-RLS share the same $\mathbf{k}[n]$ and $\mathbf{P}[n]$, which results in a simpler structure.

Similar to (2a), the a priori error of the RLS #l is given as $e_{(l)}[n] = d_{(l)}[n] - \hat{\mathbf{h}}_{(l)}^H[n-1]\mathbf{x}[n]$ and, similar to (2c), the effective IR is updated as $\hat{\mathbf{h}}_{(l)}[n] = \hat{\mathbf{h}}_{(l)}[n-1] + e_{(l)}^*[n]\mathbf{k}[n]$. After that, the a posteriori error is calculated as

$$\begin{aligned} d_{(l+1)}[n] &= d_{(l)}[n] - \hat{\mathbf{h}}_{(l)}^H[n]\mathbf{x}[n] \\ &= \Big(d_{(l)}[n] - \hat{\mathbf{h}}_{(l)}^H[n-1]\mathbf{x}[n]\Big) \\ &\quad - \Big(\hat{\mathbf{h}}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n-1]\Big)^H\mathbf{x}[n] = e_{(l)}[n]T[n], \end{aligned} \tag{32}$$

where $T[n] = 1 - \mathbf{k}^H[n]\mathbf{x}[n]$. As a result, the m-RLS algorithm can be summarized as shown in Algorithm 1, where $0 < \delta \ll 1$ is for initializing $\mathbf{P}[n]$.[1]

The time complexity is the number of operations (multiplications, additions, and divisions) that an adaptive algorithm performs to update the IR estimation for each time sample [31]. According to (2a)–(2d), the classic implementation of RLS algorithm requires $4M^2 + 3M + 1$ multiplications, $4M^2 - M$ additions, and M divisions at each sample time [13]. In this regard, Table I represents the complexity of different steps of m-RLS algorithm in Algorithm 1 by using the classic implementation of each RLS. As shown in this table, the total complexity of m-RLS is of order $\mathcal{O}(M^2)$ and higher than that of RLS.

It is worth mentioning that we can use low-complexity versions of RLS to implement m-RLS. For instance, by using the transversal DCD algorithm [28], the RLS complexity can be reduced to only $3M$ multiplications and $2MN_{\rm itr} + 6M$ additions, where $N_{\rm itr}$ is the number of iterations of the DCD algorithm. This approach can be used to implement Algorithm 1 for reducing the total complexity of steps 3, 4, and 10. In addition, when $z = 2^{-a}$ and a is an integer, the multiplication by z in step 12 is only a shift in a bit register. Table II shows the complexity of m-RLS by using the transversal DCD algorithm. Accordingly, the m-RLS complexity in Table II is of order $\mathcal{O}(M)$ and significantly lower than that of the classic implementation in Table I.

[1]The effect of $\mathbf{P}[-1]$ on the convergence behavior can be minimized by choosing a very small δ [1].

Algorithm 1: m-RLS Algorithm.

Initialization: $\mathbf{P}[-1]=\delta^{-1}\mathbf{I}_M$, $\mathbf{k}[-1]=\mathbf{0}$, $z=1/(M\mathcal{K})$, $\hat{\mathbf{h}}[-1]=\mathbf{0}$, $\hat{\mathbf{h}}_{(l)}[-1]=\mathbf{0}$, $\pi_{(l+1)}[-1]=0$, $r(l)=2(1-\varepsilon M)^l\sigma_w^2$, for $l=1,\ldots,L_{\max}$

for $n=1,2,3,\ldots$ **do**

$\mathbf{k}[n]=\left(\lambda+\mathbf{x}^H[n]\mathbf{P}[n-1]\mathbf{x}[n]\right)^{-1}\mathbf{P}[n-1]\mathbf{x}[n]$

$\mathbf{P}[n]=\lambda^{-1}\left(\mathbf{I}_M-\mathbf{k}[n]\mathbf{x}^H[n]\right)\mathbf{P}[n-1]$

$T[n]=1-\mathbf{k}^H[n]\mathbf{x}[n]$

$J_{\min}=\text{Inf}$

$L_{\text{opt}}=1$

for $l=1,\ldots,L_{\max}$ **do**

$e_{(l)}[n]=d_{(l)}[n]-\hat{\mathbf{h}}_{(l)}^H[n-1]\mathbf{x}[n]$

$\hat{\mathbf{h}}_{(l)}[n]=\hat{\mathbf{h}}_{(l)}[n-1]+e_{(l)}^*[n]\mathbf{k}[n]$

$d_{(l+1)}[n]=e_{(l)}[n]T[n]$

$\pi_{(l+1)}[n]=(1-z)\pi_{(l+1)}[n-1]+z\left|d_{(l+1)}[n]\right|^2$

$J=\pi_{(l+1)}[n]-r(l)$

if $J<J_{\min}$ **then**

$J_{\min}=J$

$L_{\text{opt}}=l$

$\hat{\mathbf{h}}[n]=\sum_{l=1}^{L_{\text{opt}}}\hat{\mathbf{h}}_{(l)}[n]$

TABLE I
COMPLEXITY OF THE PROPOSED ESTIMATOR IN ALGORITHM 1, BY USING CLASSIC IMPLEMENTATION

Step	Mult	Add	Div
3	$2M^2+M$	$2M^2-M$	M
4	$4M^2+3M+1$	$4M^2-M$	-
5	M	M	-
9	$L_{\max}M$	$L_{\max}M$	-
10	$L_{\max}M$	$L_{\max}M$	-
11	$L_{\max}$	-	-
12	$4L_{\max}$	$L_{\max}$	-
13	-	$L_{\max}$	-
14-16	-	$L_{\max}$	-
17	-	$(L_{\text{opt}}-1)M$	-
Total: $6M^2+(2L_{\max}+5)M+5L_{\max}+1$ Mult $6M^2+(2L_{\max}+L_{\text{opt}}-2)M+3L_{\max}$ Add M Div			

TABLE II
COMPLEXITY OF THE PROPOSED ESTIMATOR IN ALGORITHM 1, BY USING TRANSVERSAL DCD IMPLEMENTATION

Step	Mult	Add
3-4	M	$2M$
5	M	M
9	$L_{\max}M$	$L_{\max}M$
10	$L_{\max}M$	$L_{\max}(3M+2MN_{\text{itr}})$
11	$L_{\max}$	-
12	$L_{\max}$	$L_{\max}$
13	-	$L_{\max}$
14-16	-	$L_{\max}$
17	-	$(L_{\text{opt}}-1)M$
Total: $(2L_{\max}+2)M+L_{\max}$ Mult $([2N_{\text{itr}}+4]L_{\max}+L_{\text{opt}}+2)M+3L_{\max}$ Add		

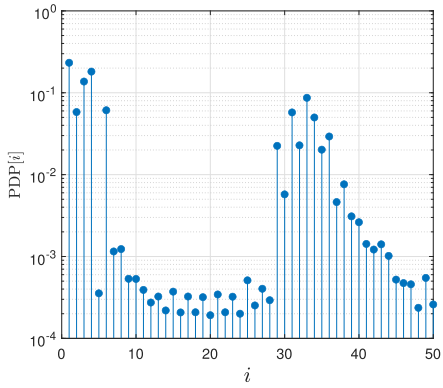


Fig. 2. The PDP of the unknown system's IR [33].

V. SIMULATION RESULTS

Digital self-interference cancellation in a full-duplex communication is one of the applications which requires accurate real-time tracking of the self-interference channel [32]. In this regard, for simulations, we consider that the unknown system is a self-interference channel, with length $M=50$ and the power-delay-profile (PDP) shown in Fig. 2 [33]. The time-varying IR is simulated based on the simulator discussed in [34]. The input signal is an uncorrelated binary sequence $x[n]\in\{\pm1\}$ with length 3000. The forgetting factor in the RLS and m-RLS algorithms is set to $\lambda=1-\frac{1}{2M}$. In m-RLS, we take $L_{\max}=5$, $\mathcal{K}=1$, and $\delta=0.001$. The reported results are the averages on 2000 Monte Carlo trials.

For a better explanation, we split the simulations into three parts. In the first part, the performance of m-RLS is assessed during the time. In the second part, we compare the simulation results for different coherence lengths and SNRs. Finally, in the third part, we evaluate the performances in a situation where the IR impulsively changes.

A. Performance Evaluation During the Time

In the first part of the simulations, we perform the simulation for the case where the coherence length of the IR taps is $N=200$, and SNR is 20 dB. Fig. 3-a compares the IR estimation error $||\mathbf{h}[n]-\hat{\mathbf{h}}[n]||^2$ for m-RLS and RLS for one trial; and Fig. 3-b shows the corresponding optimal number of layers, L_{opt}, in m-RLS. As it can be observed from Fig. 3-a, the estimation error in m-RLS is generally lower than that of RLS. This result shows that how the multi-layered idea can reduce the error of tracking a time-varying system. Fig. 3-b illustrates that, after the initial transience, L_{opt} mostly takes 3 and 4.

Fig. 4 shows the average of these results over 2000 Monte Carlo trials, where MSE is the average of $||\mathbf{h}[n]-\hat{\mathbf{h}}[n]||^2$, and $\bar{L}_{\text{opt}}$ is the average of L_{opt}. As seen, the MSE of m-RLS is around 1.5 dB lower than that of RLS. Based on Fig. 4-b, this supremacy is achieved by using $\bar{L}_{\text{opt}}\approx 3.5$.

To confirm our derivation about the ACFs of the effective IRs in m-RLS, Fig. 5 compares the theoretical $\varphi_{(l+1)}[m]$ in (24) with that inferred from simulation. Based on this figure, the theoretical ACFs roughly match the simulation results for $\varphi_{(l+1)}[m]\geq 0.5$. The slight mismatch, for $l=4,5$, comes from

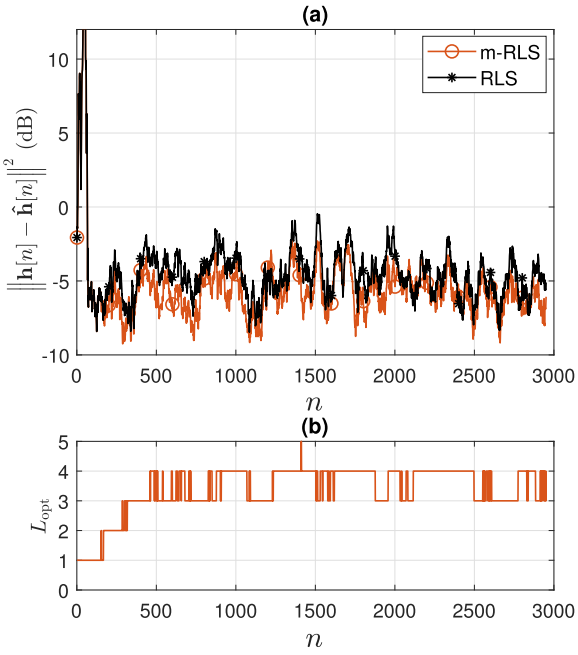


Fig. 3. When $N = 200$ and SNR is 20 dB, (a) the estimation errors of m-RLS and RLS in one trial, (b) the corresponding optimum number of layers in m-RLS.

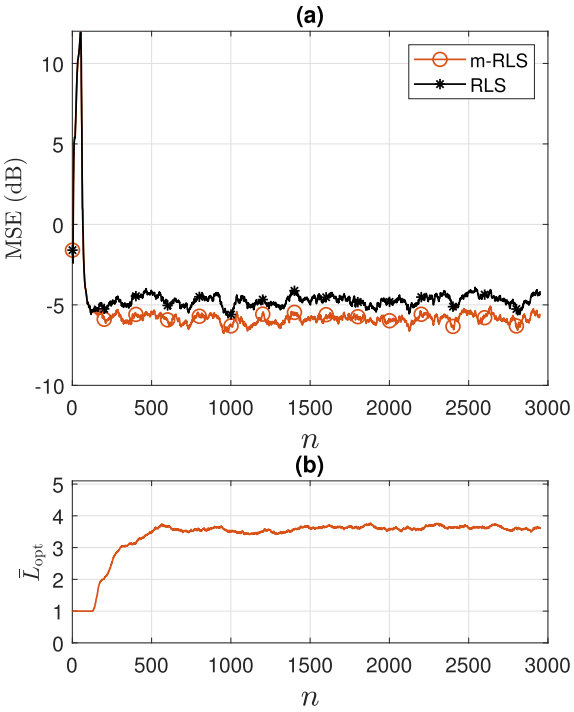


Fig. 4. When $N = 200$ and SNR is 20 dB, (a) MSE performances of m-RLS and RLS, (b) the average of the corresponding optimum number of layers in m-RLS.

the assumptions that we hold to simplify our derivation in Section III-C. It should be mentioned that, according to Fig. 5, the ACF of the effective IR at each layer is narrower than that of the previous layer bringing up this fact that each effective IR varies faster than the previous one.

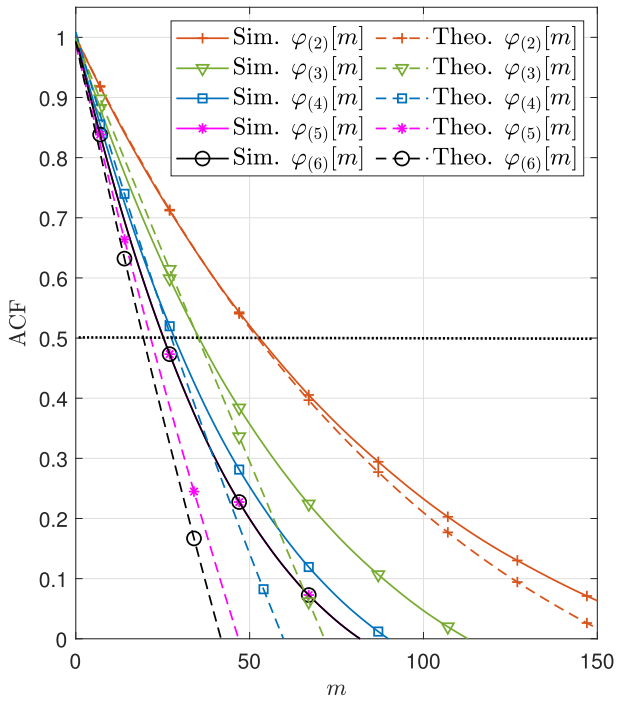


Fig. 5. The simulation and theoretical ACF of effective IRs in m-RLS, when $N = 200$ and SNR is 20 dB.

TABLE III
THE SIMULATION AND THEORETICAL COHERENCE LENGTHS OF EFFECTIVE IRS, WHEN $N = 200$

l	Sim. $N_{(l+1)}$	Theo. $N_{(l+1)}$
1	53	53
2	33	33
3	26	26
4	23	22
5	23	19

Furthermore, based on the plots in Fig. 5, we compare the simulation and theoretical coherence lengths $N_{(l+1)}$ in Table III. It is noteworthy that the simulated and theoretical coherence lengths are mutually close and justify the derivations in Section III-C.

B. Performance Evaluation for Different SNRs and Different Coherence Lengths

In the second part of the simulations, we inquire about the performance of m-RLS in different SNRs when the time-variation of the system is relatively fast, and slow with the coherence length $N = 200$, and 2000, respectively. In addition to the classic RLS, we also compare m-RLS with two types of RLS-VFF algorithms. We title these algorithms as RLS-VFF1 and RLS-VFF2 which are proposed in [15] and [14], respectively. The required parameters for RLS-VFF1 and RLS-VFF2 are set based on the original works.

Fig. 6-a shows the MSE results for $N = 200$. Fig. 6-b shows the corresponding $\bar{L}_{\text{opt}}$ in m-RLS. As seen, when the system is relatively fast time-varying, for the SNRs less than 10 dB, $\bar{L}_{\text{opt}} = 1$ and m-RLS is equivalent to RLS. However, for SNRs greater than 10 dB, $\bar{L}_{\text{opt}}$ is larger than one, and the performance of m-RLS becomes better than that of RLS. As the

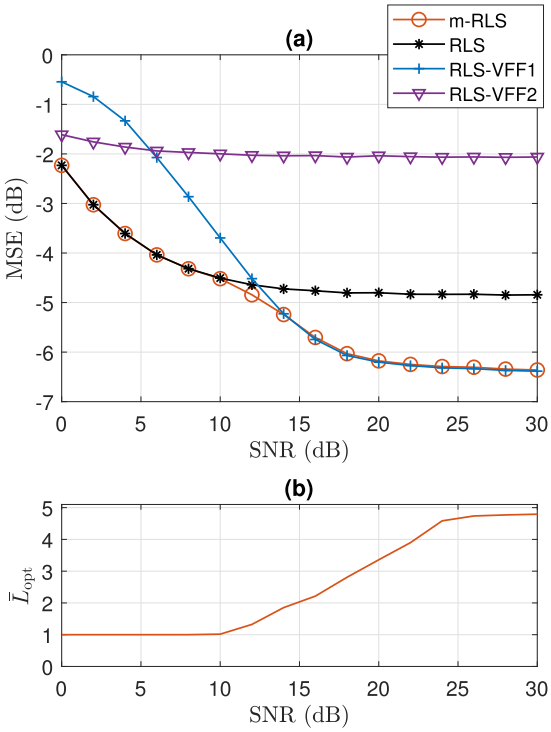


Fig. 6. When $N = 200$, (a) the MSE performances for different SNRs, (b) the average of the corresponding optimum number of layers in m-RLS.

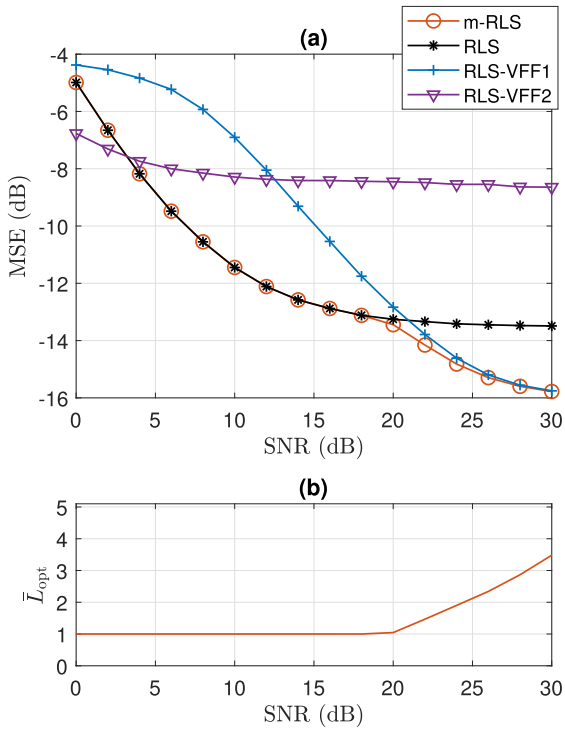


Fig. 7. When $N = 2000$, (a) the MSE performances for different SNRs, (b) the average of the corresponding optimum number of layers in m-RLS.

SNR increases, $\bar{L}_{\text{opt}}$ grows larger and the MSE performance of m-RLS becomes better than that of RLS. These results establish the fact that a higher SNR leads to a larger L_{opt}. The MSE of RLS-VFF1, for SNRs below 14 dB, is worst than that of m-RLS; however, for SNRs above 14 dB, it merges to the MSE of m-RLS. The performance of RLS-VFF2 does not remarkably change with SNR. This estimator has the highest MSE among the tested methods, except for very low SNRs, where it outperforms RLS-VFF1.

Fig. 7 shows the results when $N = 2000$. As seen, when the system is relatively slow time-varying, for SNR equal to 20 dB and higher, $\bar{L}_{\text{opt}}$ becomes larger than one and m-RLS outperforms RLS. Comparing Fig. 7-b to Fig. 6-b, one can see that $\bar{L}_{\text{opt}}$ is generally larger for the relatively faster time-varying IR.

We repeat the same simulation as that in Fig. 6 for the case where the input signal is a normalized and uncorrelated complex-valued Gaussian sequence. The results are represented in Fig. 8, which are fairly identical to those in Fig. 6. This shows that m-RLS is useful for any uncorrelated input signal.

It is worth noting that, for determining the optimal number of layers based on (31), the noise power must be known in the m-RLS algorithm. However, via numerical evaluations, we show that a moderate uncertainty about the noise power does not significantly affect the performance. To this purpose, Fig. 9 shows the performance of m-RLS with the same simulation parameters used in Fig. 6, except it is assumed the uncertain noise power is given to the algorithm as $(1 + u \times randn)\sigma_w^2$, where $randn$ is a zero-mean normalized Gaussian random variable and u is a parameter to control the extension of the uncertainty. As it can be seen from Fig. 9-a and Fig. 9-b, for $u = 0.1$ and 0.2, the MSE performance and $\bar{L}_{\text{opt}}$ are the same as those for $u = 0$ (no uncertainty). On the other hand, for $u = 0.5$, the MSE performance for SNRs lower than 14 dB becomes slightly worst than that of $u = 0$. In this SNR range, $\bar{L}_{\text{opt}}$ is higher than one. These results indicate that even up to 50% uncertainty on the noise power does not significantly affect the performance of m-RLS.

C. Performance Evaluation When an Impulsive Change Occurs

In this part of simulations, we investigate how an impulsive change, besides the continuous time-variations of the system, is handled in m-RLS. To simulate this scenario, we consider that the IR is time-varying with the coherence time N; in addition to this continuous variation, a significant impulsive change is also imposed to the IR at time $n = 1000$ by multiplying all taps of the IR vector by -1. When SNR is 10 dB, Fig. 10 and Fig. 11 show the results for $N = 200$ and 2000, respectively.

According to Fig. 10-a, when $N = 200$ and the IR is relatively fast time-varying, the MSE levels of m-RLS and RLS are almost the same, before the impulsive change at $n = 1000$. Within this period, the MSE of RLS-VFF1 is slightly higher than that of m-RLS and the MSE of RLS-VFF2 has the worst performance confirming the achievement observed in Fig. 6-a at SNR 10 dB. Nevertheless, once the impulsive change occurs, m-RLS has the fastest convergence in tracking this change. The MSE of RLS-VFF1, RLS, and RLS-VFF2 drop slower, respectively. After the transience interval (caused by the impulsive change), the MSE levels turn back to the same order as before the impulsive change.

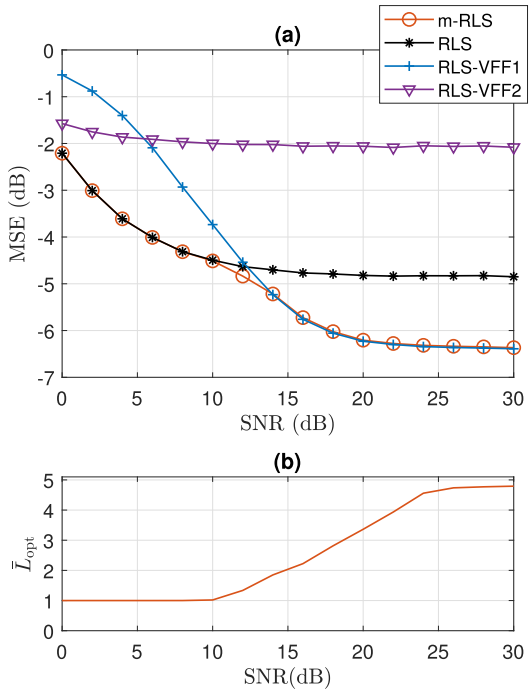


Fig. 8. When the input signal is uncorrelated complex-valued Gaussian and $N = 200$, (a) the MSE performances for different SNRs, (b) the average of the corresponding optimum number of layers in m-RLS.

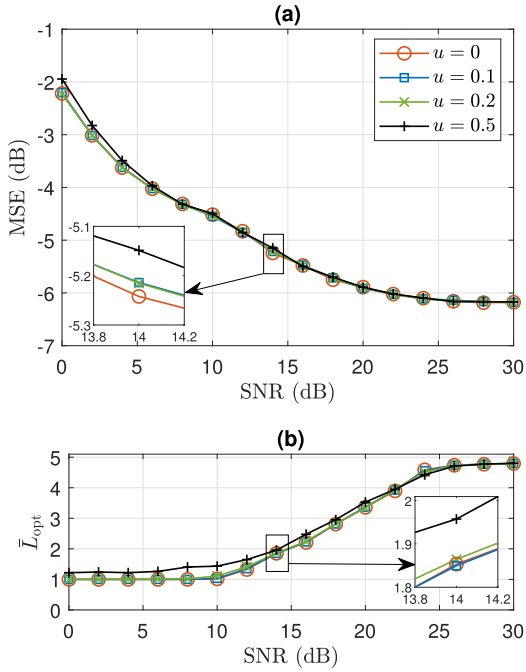


Fig. 9. When $N = 200$, and the noise power is uncertain in the m-RLS algorithm as $(1 + u \times randn)\sigma_w^2$, (a) the MSE performances of m-RLS for different SNRs, (b) the average of the corresponding optimum number of layers in m-RLS.

It is interesting to notice that, based on Fig. 10-b, $\bar{L}_{\text{opt}} \approx 1$ before the impulsive change. Once the IR is impulsively changed, $\bar{L}_{\text{opt}}$ raises up 3.7 and leads to the supremacy of m-RLS versus the other methods. Within the transience interval, the average of the optimum number of layers settles down to $\bar{L}_{\text{opt}} \approx 1$ again.

Fig. 11 shows the results when the IR is relatively slow time-varying with $N = 2000$. Based on Fig. 11-a, before and after the transient interval, m-RLS and RLS have the lowest MSE. The RLS-VFF2 and RLS-VFF1 methods have respectively the higher MSE levels verifying the results of Fig. 7-a at SNR 10 dB. Once the impulsive change happens, RLS-VFF2 and m-RLS attain the fastest convergence. RLS and RLS-VFF1 show slower drops, respectively. According to Fig. 11-b, $\bar{L}_{\text{opt}}$ jumps up to 3.6 once the impulsive change takes place and settles down to $\bar{L}_{\text{opt}} = 1$ when the transience interval ends.

Based on Fig. 10-b and Fig. 11-b, the optimum number of layers in m-RLS not only depends on the coherence length of the continuously time-variations of the IR but also can be affected by impulsive changes.

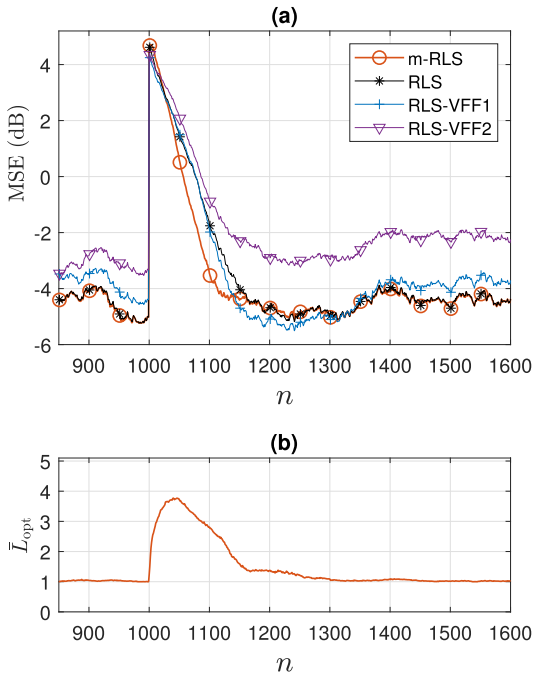


Fig. 10. When $N = 200$, SNR $= 10$ dB, and the impulsive change occurs at $n = 1000$, a) the MSE performances, (b) the average of the corresponding optimum number of layers in m-RLS.

VI. Conclusion

In this paper, we propose the m-RLS adaptive filtering to enhance the accuracy of tracking rapidly time-varying systems. We show that in m-RLS, the power of the lag error and the noise effect are functions of the number of layers. We provide a method to determine the optimum number of layers minimizing the sum of the lag error and the noise effect. The coherence length of the effective IRs in m-RLS are evaluated, and it is derived that each effective IR varies faster than that in the previous layer. The implementation complexity of m-RLS is studied and it is shown that m-RLS is more complex than RLS; however, the complexity order of the proposed approach can be reduced to $\mathcal{O}(M)$ by employing the transversal DCD algorithm, where M indicates the IR length.

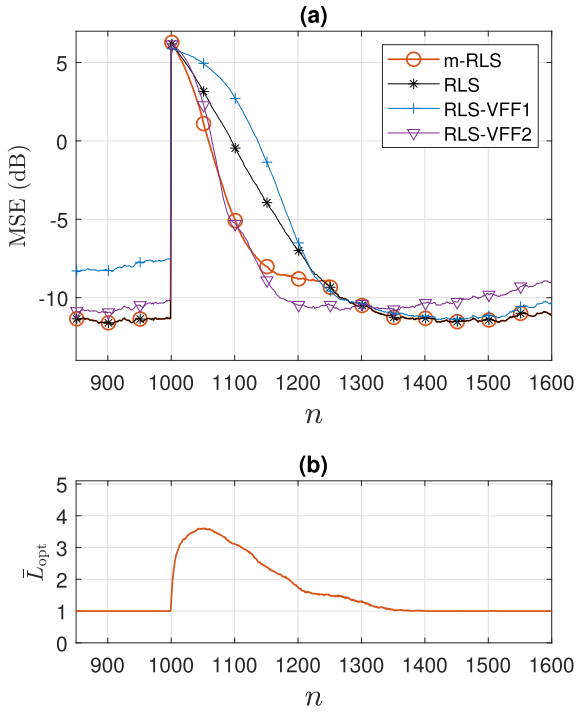


Fig. 11. When $N = 2000$, SNR $= 10$ dB, and the impulsive change occurs at $n = 1000$, a) the MSE performances, (b) the average of the corresponding optimum number of layers in m-RLS.

Based on the simulation results, we attain that, for either a relatively faster time-varying system or a higher SNR, the optimum number layers becomes larger. With the optimum number of layers, m-RLS outperforms the classic RLS and the investigated RLS methods with a variable forgetting factor especially when the system is rapidly time-varying. In addition, it is demonstrated that an uncertain knowledge about the noise power does not significantly deteriorate the performance of the proposed method.

APPENDIX A
PROOF OF (9)

We split this proof into three sub-proofs:

A. *Proof of* $\mathbb{E}||\mathbf{B}[n](\mathbf{H}[n] - \hat{\mathbf{H}}[n-N])||^2 = \rho^N \mathbb{E}||\mathbf{H}[n] - \hat{\mathbf{H}}[n-N]||^2$

Let us define $\mathbf{a} = \mathbf{h}[n] - \hat{\mathbf{h}}[n-N]$, which is assumed to be uncorrelated with $\mathbf{x}[n]$ (the second assumption in Section II-B). According to the definition of $\mathbf{\Theta}[n]$, one can see that

$$\mathbb{E}\|\mathbf{\Theta}[n]\mathbf{a}\|^2 = \sum_{i=0}^{M-1}\left[\mathbb{E}|a_i|^2 - 2\varepsilon \mathrm{Re}\sum_{j=0}^{M-1}\mathbb{E}\left\{x_i[n]x_j^*[n]a_i^*a_j\right\} + \varepsilon^2\mathbb{E}\left|\sum_{j=0}^{M-1}x_i[n]x_j^*[n]a_i\right|^2\right], \tag{33}$$

where $x_i[n]$ and a_i are the ith entries of $\mathbf{x}[n]$ and $\mathbf{a}$, respectively. Since $x_i[n]$'s and a_i's are uncorrelated, $\sum_{j=0}^{M-1}\mathbb{E}\{x_i[n]x_j^*[n]a_i^*a_j\} = \mathbb{E}|a_i|^2$, and $\mathbb{E}|\sum_{j=0}^{M-1}x_i[n]x_j^*[n]a_i|^2 = M\mathbb{E}|a_i|^2$. As a result,

$$\mathbb{E}\|\mathbf{\Theta}[n]\mathbf{a}\|^2 = \rho\sum_{i=0}^{M-1}\mathbb{E}|a_i|^2 = \rho\mathbb{E}\|\mathbf{a}\|^2, \tag{34}$$

in which $\rho = 1 - 2\varepsilon + \varepsilon^2 M$. In the next step, we define $\mathbf{b} = \mathbf{\Theta}[n]\mathbf{a}$. In the same way, since $\mathbf{\Theta}[n-1]$ and $\mathbf{b}$ are uncorrelated, $\mathbb{E}\|\mathbf{\Theta}[n-1]\mathbf{b}\|^2 = \rho\mathbb{E}\|\mathbf{b}\|^2 = \rho^2\|\mathbf{a}\|^2$. Continuing this procedure and recalling the definition of $\mathbf{B}[n]$ in (6), it is concluded that

$$\mathbb{E}\|\mathbf{B}[n]\mathbf{a}\|^2 = \mathbb{E}\left\|\prod_{k=0}^{N-1}\mathbf{\Theta}[n-k]\mathbf{a}\right\|^2 = \rho^N\mathbb{E}\|\mathbf{a}\|^2 = \rho^N\mathbb{E}\left\|\mathbf{h}[n] - \hat{\mathbf{h}}[n-N]\right\|^2. \tag{35}$$

B. *Proof of* $\mathbb{E}||\mathbf{H}[n] - \hat{\mathbf{H}}[n-N]||^2 = \mathbb{E}\|\mathbf{H}[n]\|^2$

We can write

$$\mathbb{E}\left\|\mathbf{h}[n] - \hat{\mathbf{h}}[n-N]\right\|^2 = \mathbb{E}\|\mathbf{h}[n]\|^2 + \mathbb{E}\left\|\hat{\mathbf{h}}[n-N]\right\|^2 - 2\mathrm{Re}\mathbb{E}\{\mathbf{h}^H[n]\hat{\mathbf{h}}[n-N]\}. \tag{36}$$

Assuming that the noise is insignificant, we can consider $\mathbb{E}\{\mathbf{h}^H[n]\hat{\mathbf{h}}[n-m]\} = \mathbb{E}\{\mathbf{h}^H[n]\mathbf{h}[n-m]\}$ for $m \in \mathbb{Z}$. As a result, we have

$$\mathbb{E}\left\|\hat{\mathbf{h}}[n-N]\right\|^2 = \mathbb{E}\|\mathbf{h}[n-N]\|^2 = \mathbb{E}\|\mathbf{h}[n]\|^2, \tag{37}$$

and

$$\mathbb{E}\{\mathbf{h}^H[n]\hat{\mathbf{h}}[n-N]\} = \sum_{i=0}^{M-1}\mathbb{E}\{h_i^*[n]h_i[n-N]\} = \varphi[N]\sum_{i=0}^{M-1}\mathbb{E}|h_i[n]|^2 = 0.5\mathbb{E}\|\mathbf{h}[n]\|^2, \tag{38}$$

where $h_i[n]$ is the ith entry of $\mathbf{h}[n]$, and $\varphi[n]$ is the normalized ACF (i.e., $\varphi[0] = 1$) of $h_i[n]$, for $i = 0, \ldots, M-1$ (note that it is assumed that all $h_i[n]$'s have the same ACF). In (38), we let $\varphi[N] = 0.5$ due to the fact that the ACF meets 0.5 at the coherence length [30]. Then, by considering (37) and (38) into (36), we have

$$\mathbb{E}||\mathbf{h}[n] - \hat{\mathbf{h}}[n-N]||^2 = \mathbb{E}\|\mathbf{h}[n]\|^2. \tag{39}$$

C. *Proof of* $\mathbb{E}\|\mathbf{C}[n]\|^2 = \varepsilon^2 M(\frac{1-\rho^N}{1-\rho})\sigma_w^2$

The definition of $\mathbf{c}[n]$ is brought in (6). Since the noise samples are uncorrelated,

$$\mathbb{E}\|\mathbf{c}[n]\|^2 = \mathbb{E}\|\boldsymbol{\gamma}[n]\|^2 + \sum_{k=1}^{N-1}\mathbb{E}\left\|(\prod_{i=0}^{k-1}\mathbf{\Theta}[n-i])\boldsymbol{\gamma}[n-k]\right\|^2. \tag{40}$$

On the other hand, from the first sub-section in Appendix A, we know that $\mathbb{E}||(\prod_{i=0}^{k-1}\mathbf{\Theta}[n-i])\boldsymbol{\gamma}[n-k]||^2 =$

$\rho^k \mathbb{E}\|\boldsymbol{\gamma}[n-k]\|^2 = \rho^k \varepsilon^2 M \sigma_w^2$. Replacing this in (40), we conclude that

$$\mathbb{E}\|\mathbf{c}[n]\|^2 = \varepsilon^2 M \sigma_w^2 \sum_{k=0}^{N-1} \rho^k = \varepsilon^2 M \frac{1-\rho^N}{1-\rho} \sigma_w^2 \tag{41}$$

Finally, considering the results in (35), (39), and (41) together leads to (9).

Appendix B
Proof of (24)

Since we assume the IRs and their estimates are uncorrelated with the input signal, discarding the noise in (18) leads to

$$\begin{aligned}&\mathbb{E}\{\mathbf{h}_{(l+1)}^H[n]\mathbf{h}_{(l+1)}[n-m]\}\\ &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \mathbb{E}\{Q_{i,j}\}\mathbb{E}\Big\{(h_{(l),i}[n] - \hat{h}_{(l),i}[n-N_{(l)}])^* \\ &\qquad (h_{(l),j}[n-m] - \hat{h}_{(l),j}[n-m-N_{(l)}])\Big\},\end{aligned} \tag{42}$$

where $Q_{i,j}$ is the (i,j)th entry of matrix $\mathbf{Q} = \mathbf{B}_{(l)}^H[n]\mathbf{B}_{(l)}[n-m]$. Also, $h_{(l),i}[n]$ and $\hat{h}_{(l),i}[n]$ are the ith entry of $\mathbf{h}_{(l)}[n]$ and $\hat{\mathbf{h}}_{(l)}[n]$, respectively. Considering that the tap-weights in an effective IR are mutually uncorrelated variables, we can rewrite (42) as

$$\begin{aligned}&\mathbb{E}\{\mathbf{h}_{(l+1)}^H[n]\mathbf{h}_{(l+1)}[n-m]\}\\ &= \sum_{i=0}^{M-1} \mathbb{E}\{Q_{i,i}\}\mathbb{E}\Big\{(h_{(l),i}[n] - \hat{h}_{(l),i}[n-N_{(l)}])^* \\ &(h_{(l),i}[n-m] - \hat{h}_{(l),i}[n-m-N_{(l)}])\Big\}.\end{aligned} \tag{43}$$

Based on the definition of $\mathbf{B}_{(l)}[n]$ in (19), and holding the assumption that the IR estimation is independent of the input signal, one can realize that

$$\begin{aligned}\mathbb{E}\{\mathbf{Q}\} = &\prod_{k=0}^{m-1} \mathbb{E}(\mathbf{I}_M - \varepsilon\mathbf{x}[n-k]\mathbf{x}^H[n-k]) \\ &\prod_{k=m}^{N_{(l)}-1} \mathbb{E}(\mathbf{I}_M + \varepsilon\mathbf{x}[n-m-k]\mathbf{x}^H[n-m-k])^2 \\ &\prod_{k=N_{(l)}}^{m+N_{(l)}-1} \mathbb{E}(\mathbf{I}_M - \varepsilon\mathbf{x}[n-k]\mathbf{x}^H[n-k]),\end{aligned} \tag{44}$$

when $m \le N_{(l)}$, and

$$\begin{aligned}\mathbb{E}\{\mathbf{Q}\} = &\prod_{k=0}^{N_{(l)}-1} \mathbb{E}(\mathbf{I}_M - \varepsilon\{\mathbf{x}[n-k]\mathbf{x}^H[n-k]\}) \\ &\prod_{k=m}^{m+N_{(l)}-1} \mathbb{E}(\mathbf{I}_M - \varepsilon\mathbf{x}[n-k]\mathbf{x}^H[n-k]),\end{aligned} \tag{45}$$

when $m > N_{(l)}$. Since the input signal $x[n]$ is normalized (i.e., $\mathbb{E}|x[n]|^2 = 1$), by using (44) and (45), the ith diagonal entry of $\mathbb{E}\{\mathbf{Q}\}$ becomes

$$\mathbb{E}\{Q_{i,i}\} = \begin{cases} \lambda^{2m}\rho^{N_{(l)}-m}; & \text{for } 0 \le m \le N_{(l)} \\ \lambda^{2N_{(l)}}; & \text{for } N_{(l)} < m. \end{cases} \tag{46}$$

As seen, $\mathbb{E}\{Q_{i,i}\}$ is independent of i, thus, replacing (46) in (43) results in

$$\begin{aligned}&\mathbb{E}\{\mathbf{h}_{(l+1)}^H[n]\mathbf{h}_{(l+1)}[n-m]\}\\ &= \mathbb{E}\Big\{(\mathbf{h}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n-N_{(l)}])^H \\ &\quad (\mathbf{h}_{(l)}[n-m] - \hat{\mathbf{h}}_{(l)}[n-m-N_{(l)}])\Big\}\mathbb{E}\{Q_{i,i}\} \\ &= \mathbb{E}\big\|\mathbf{h}_{(l+1)}[n]\big\|^2 \\ &\big(2\varphi_{(l)}[m] - \varphi_{(l)}[m-N_{(l)}] - \varphi_{(l)}[m+N_{(l)}]\big)\mathbb{E}\{Q_{i,i}\}.\end{aligned} \tag{47}$$

Then, substituting (47) and (21) in (23), when the noise is discarded, results in (24).

Appendix C
Proof of $\varphi_{(l+1)}[m] \approx \exp(-\beta M)$

From (24), we have $\varphi_{(l+1)}[m] = (f'[m] + f''[m])q_{(l)}[m]$, where $f'[m] = \varphi_{(l)}[m] - \varphi_{(l)}[m-N_{(l)}]$, and $f''[m] = \varphi_{(l)}[m] - \varphi_{(l)}[m+N_{(l)}]$. Since $\varphi_{(l)}[m] = \exp(-\alpha m)$, and $\varphi_{(l)}[N_{(l)}] = 0.5$, one can see that $f'[m] = 0.5\exp(-\alpha m)$, and $f''[m] = 0.5 - \frac{\alpha}{\log 2}m$, for $0 \le m \le N_{(l)}$. As a result, we can approximate that $f'[m] + f''[m] \approx \exp(-3\alpha m)$.

On the other hand, from (25), we have $q_{(l)}[m] = \exp(-gm)$, for $0 \le m \le N_{(l)}$, where $g = \log(\frac{\rho}{\lambda^2})$. Putting these results together, it is concluded that $\varphi_{(l+1)}[m] \approx \exp(-\beta m)$, where $\beta = 3\alpha + g$.

Appendix D
Proof of (29)

Since the input signal is normalized, based on (13), we have

$$\begin{aligned}\mathbb{E}\big|d_{(l+1)}[n]\big|^2 = &\mathbb{E}\big\|\mathbf{h}_{(l+1)}[n]\big\|^2 + \sigma_w^2 \\ &+ 2\mathrm{Re}\mathbb{E}\{\mathbf{h}_{(l+1)}^H[n]\mathbf{x}[n]w^*[n]\}.\end{aligned} \tag{48}$$

Let us first investigate $\mathbb{E}\{\mathbf{h}_{(l+1)}^H[n]\mathbf{x}[n]w^*[n]\}$ in (48). Similar to (4), for lth layer, we can write

$$\hat{\mathbf{h}}_{(l)}[n] = \mathbf{h}_{(l)}[n] - \boldsymbol{\Theta}[n]\Big(\mathbf{h}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n-1]\Big) + \boldsymbol{\gamma}[n]. \tag{49}$$

Replacing (14) in (49) results in

$$\mathbf{h}_{(l+1)}[n] = \boldsymbol{\Theta}[n]\Big(\mathbf{h}_{(l)}[n] - \hat{\mathbf{h}}_{(l)}[n-1]\Big) - \boldsymbol{\gamma}[n]. \tag{50}$$

By using (50), one can expand $\mathbf{h}_{(l+1)}[n]$ as

$$\begin{aligned}\mathbf{h}_{(l+1)}[n] = &\boldsymbol{\Theta}^l[n]\mathbf{h}_{(1)}[n] - \sum_{j=1}^{l}\boldsymbol{\Theta}^j[n]\hat{\mathbf{h}}_{(l-j+1)}[n-1] \\ &- \sum_{j=0}^{l-1}\boldsymbol{\Theta}^j[n]\boldsymbol{\gamma}[n].\end{aligned} \tag{51}$$

On the right-hand side of (51), the first and the second terms are uncorrelated with $w[n]$. The only correlated part is the third term because $\boldsymbol{\gamma}[n] = \varepsilon\mathbf{x}[n]w^*[n]$. Thus, since $\boldsymbol{\Theta}[n]$ is a Hermitian

matrix, we have

$$\mathbb{E}\{\mathbf{h}_{(l+1)}^{H}[n]\mathbf{x}[n]w^{*}[n]\} = -\varepsilon\sigma_w^2\mathbb{E}\{\mathbf{x}^{H}[n]\sum_{j=0}^{l-1}\mathbf{\Theta}^{j}[n]\mathbf{x}[n]\}. \tag{52}$$

Based on the definition of $\mathbf{\Theta}[n]$, and assuming that M is large enough so that $\mathbf{x}^{H}[n]\mathbf{x}[n] = M\mathbb{E}|x[n]|^2 = M$, it is easy to follow that $\mathbf{x}^{H}[n]\mathbf{\Theta}^{j}[n]\mathbf{x}[n] = M(1-\varepsilon M)^{j}$ is independent of n. As a result,

$$\mathbb{E}\{\mathbf{h}_{(l+1)}^{H}[n]\mathbf{x}[n]w^{*}[n]\} = -\left[1-(1-\varepsilon M)^{l}\right]\sigma_w^2. \tag{53}$$

Replacing (53) in (48) results in (29).

REFERENCES

[1] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. Chichester, U.K.: Wiley, 1998.

[2] S. Haykin, *Adaptive Filter Theory*. 4th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[3] D. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 92–114, Jan. 1991.

[4] J. Cioffi and T. Kailath, "Fast, recursive-least-squares transversal filters for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 2, pp. 304–337, Apr. 1984.

[5] J.-L. Botto, "Stabilization of fast recursive least-squares transversal filters for adaptive filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 12, 1987, pp. 403–406.

[6] D. Lee, M. Morf, and B. Friedlander, "Recursive least squares ladder estimation algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 3, pp. 627–641, Jun. 1981.

[7] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, no. 8, pp. 829–867, Aug. 1982.

[8] M. Z. A. Bhotto and A. Antoniou, "Robust recursive least-squares adaptive-filtering algorithm for impulsive-noise environments," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 185–188, Mar. 2011.

[9] S. Farahmand and G. B. Giannakis, "Robust RLS in the presence of correlated noise using outlier sparsity," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3308–3313, Jun. 2012.

[10] S.-C. Chan and Y.-X. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: Fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.

[11] Y. Zou, S. Chan, and T. Ng, "A recursive least m-estimate (RLM) adaptive filter for robust filtering in impulse noise," *IEEE Signal Process. Lett.*, vol. 7, no. 11, pp. 324–326, Nov. 2000.

[12] E. Eleftheriou and D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1097–1110, Oct. 1986.

[13] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ, USA: Wiley, 2003.

[14] C. Paleologu, J. Benesty, and S. Ciochina, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, 2008.

[15] M. Z. A. Bhotto and A. Antoniou, "New improved recursive least-squares adaptive-filtering algorithms," *IEEE Trans. Circuits Syst. I: Reg. Papers*, vol. 60, no. 6, pp. 1548–1558, Jun. 2013.

[16] S.-H. Leung and C. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, Aug. 2005.

[17] B. Toplis and S. Pasupathy, "Tracking improvements in fast RLS algorithms using a variable forgetting factor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-36, no. 2, pp. 206–227, Feb. 1988.

[18] S. Zhang and J. Zhang, "An RLS algorithm with evolving forgetting factor," in *Proc. 7th Int. Workshop Signal Des. Appl. Commun.*, 2015, pp. 24–27.

[19] Y. Chu and C. Mak, "A variable forgetting factor diffusion recursive least squares algorithm for distributed estimation," *Signal Process.*, vol. 140, pp. 219–225, 2017.

[20] J. Apolinario, M. Campos, and P. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3235–3242, Nov. 2000.

[21] S. Shaffer and C. Williams, "Comparison of LMS, alpha LMS, and data reusing LMS algorithms," in *Proc. Asilomar Conf. Circuits, Syst., Comput.*, 1983, pp. 260–264.

[22] J. Benesty and T. Gänsler, "On data-reuse adaptive algorithms," in *Proc. Int. Workshop Acoustic Echo Noise Control*, 2003, pp. 31–34.

[23] B. Schnaufer and W. Jenkins, "New data-reusing LMS algorithms for improved convergence," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 1993, pp. 1584–1588.

[24] C. Paleologu and J. Benesty, "A practical data-reuse adaptive algorithm for acoustic echo cancellation," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 2010–2014.

[25] L. F. O. Chamon, H. F. Ferro, and C. G. Lopes, "A data reusage algorithm based on incremental combination of LMS filters," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst. Comput.*, 2012, pp. 406–410.

[26] F. Tong, B. Benson, Y. Li, and R. Kastner, "Channel equalization based on data reuse LMS algorithm for shallow water acoustic communication," in *Proc. IEEE Int. Conf. Sensor Netw., Ubiquitous, Trustworthy Comput.*, 2010, pp. 95–98.

[27] M. T. Akhtar, "Binormalized data-reusing adaptive filtering algorithm for active control of impulsive sources," *Digit. Signal Process.*, vol. 49, pp. 56–64, 2016.

[28] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3150–3161, Jul. 2008.

[29] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2937–2944, Nov. 1994.

[30] R. Steele and L. Hanzo, *Mobile Radio Communications: Second and Third Generation Cellular and WATM Systems*, 2nd ed. Chichester, U.K.: IEEE Press - John Wiley, May 1999.

[31] J. Liu and Y. Zakharov, "Low complexity dynamically regularised RLS algorithm," *Electron. Lett.*, vol. 44, no. 14, p. 886–885, 2008.

[32] D. Wu, C. Zhang, S. Gao, and D. Chen, "A digital self-interference cancellation method for practical full-duplex radio," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput.*, 2014, pp. 74–79.

[33] M. Towliat, Z. Guo, L. J. Cimini, X.-G. Xia, and A. Song, "Self-interference channel characterization in underwater acoustic in-band full-duplex communications using OFDM," in *Proc. Global Oceans 2020: Singapore - U.S. Gulf Coast*, 2020, pp. 1–7.

[34] P. A. van Walree, T. Jenserud, and M. Smedsrud, "A discrete-time channel simulator driven by measured scattering functions," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 9, pp. 1628–1637, Dec. 2008.

Mohammad Towliat received the B.Sc. degree in electrical engineering from Birjand University, Birjand, Iran, in 2009, and the M.S. degree in electrical engineering from the Ferdowsi University of Mashhad, Mashhad, Iran, in 2012. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA. His general interests include MIMO multicarrier systems, space-time coding, underwater acoustic communication, and adaptive filtering.

Zheng Guo received the B.S. degree in electro-information engineering of Underwater Acoustics from Harbin Engineering University, Harbin, China, the M.Eng. degree in electronics and communication engineering from the University of Chinese Academy of Sciences, Beijing, China, and the Ph.D. degree in electrical engineering from the University of Alabama, Tuscaloosa, AL, USA, in 2014, 2017, and 2021, respectively.

He is currently an Engineer with NXP Semiconductors, San Jose, CA, USA. His research interests include signal processing in wireless systems and underwater acoustic communications.

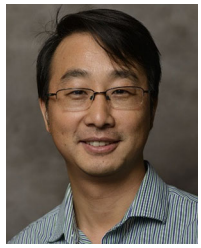
Leonard J. Cimini (Life Fellow, IEEE) received the Ph.D. degree from the University of Pennsylvania, Philadelphia, PA, USA, in 1982. He was with AT&T, first in Bell Laboratories and then at AT&T Laboratories, for 20 years. He has been a Professor with the University of Delaware, Newark, DE, USA, since 2002. He was elected as the IEEE Fellow for contributions to the theory and practice of high-speed wireless communications in 2000. For his pioneering work on OFDM for wireless systems, he was the recipient of the 2007 James R. Evans Avant Garde Award from the IEEE Vehicular Technology Society and the 2010 Innovators Award from the NJ Inventors Hall of Fame. He was the recipient of several other awards from the IEEE Communications Society, including the 2010 Stephen O. Rice Prize, the 2010 Donald W. McLellan Meritorious Service Award, the 2010 Recognition Award from the Wireless Communications Technical Committee, the 2011 Service and 2021 Technical Achievement Awards from the Communication Theory Technical Committee, the 2016 Joseph LoCicero Award for exemplary service to publications, and the 2021 Mentorship Award from the Women in Communications Engineering Standing Committee. He has held several publications and governance positions within the IEEE Communications Society, including as the Founding Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC) Wireless Communications Series.

Xiang-Gen Xia (Fellow, IEEE) received the B.S. degree in mathematics from Nanjing Normal University, Nanjing, China, and the M.S. degree in mathematics from Nankai University, Tianjin, China, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1983, 1986, and 1992, respectively.

He was a Senior/Research Staff Member with Hughes Research Laboratories, Malibu, CA, USA, during 1995–1996. In September 1996, he joined the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA, where he is the Charles Black Evans Professor.

His research interests include space-time coding, MIMO and OFDM systems, digital signal processing, and SAR and ISAR imaging. Dr. Xia is the author of the book *Modulated Coding for Intersymbol Interference Channels* (New York, Marcel Dekker, 2000).

Dr. Xia was the recipient of the National Science Foundation (NSF) Faculty Early Career Development (CAREER) Program Award in 1997, the Office of Naval Research (ONR) Young Investigator Award in 1998, and the Outstanding Overseas Young Investigator Award from the National Nature Science Foundation of China in 2001, the 2019 Information Theory Outstanding Overseas Chinese Scientist Award, The Information Theory Society of Chinese Institute of Electronics. Dr. Xia was an Associate Editor for numerous international journals including IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. Dr. Xia is the Technical Program Chair of the Signal Processing Symposium, Globecom 2007 in Washington D.C. and the General Co-Chair of ICASSP 2005 in Philadelphia.



Aijun Song (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Delaware, Newark, DE, USA, in 2005. From 2005 to 2008, he was a Postdoctoral Research Associate with the College of Earth, Ocean and Environment, University of Delaware, and also an Office of Naval Research Postdoctoral Fellow. He was a Research Professor with the University of Delaware from 2008 to 2015. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA. His research interests include digital communications and signal processing techniques for radio-frequency and underwater acoustic channels, ocean acoustics, sensor networks, and ocean monitoring and exploration.

Dr. Song was the General Co-Chair of the 2018 March NSF Workshop on Underwater Wireless Communications and Networking and the 2018 November NSF Workshop on Underwater Wireless Infrastructure. He was the General Co-Chair of the 14th International Conference on Underwater Networks & Systems. He was the recipient of the NSF CAREER Award (2021).