Theoretical Analysis of the Performance of the Data-Reuse RLS Algorithm

Wei Gao, Jie Chen, Cédric Richard

Theoretical Analysis of the Performance of the Data-Reuse RLS Algorithm

Wei Gao, *Member, IEEE*, Jie Chen, *Senior Member, IEEE*, Cédric Richard *Senior Member, IEEE*, Wentao Shi, *Member, IEEE*, and Qunfei Zhang, *Member, IEEE*

***Abstract*—Data-reuse RLS (DR-RLS) algorithm is a computationally efficient technique that has been recently introduced to improve the tracking capabilities of the RLS. Nevertheless, no analysis of its convergence has been proposed so far. The aim of that work is to fill this gap with theoretical analyzes in the mean and mean-square error sense. Transient and steady-state models are provided. They are validated through numerical simulations.**

***Index Terms*—Adaptive filtering, data-reuse, RLS algorithm, transient convergence analysis, steady-state analysis.**

I. Introduction

In terms of both convergence rate and steady-state misadjustment error, the RLS algorithm significantly outperforms the LMS algorithm even with colored input signals [1]–[3]. Nevertheless, the intrinsic shortcomings of the RLS include its high computational complexity and poor tracking capability. This may constraint its practical applicability in a variety of applications. Several variants of the RLS have been developed to further improve its performance. Specifically, the fast RLS (FRLS) algorithm was devised to reduce computational cost, which then becomes proportional to the order of the filter [4]–[7]. The variable forgetting factor RLS (VFF-RLS) algorithm was devised to reach a trade-off between small misadjustment error and fast convergence rate [8]–[12]. An efficient version of the RLS algorithm utilizing dichotomous coordinate descent iterations was also proposed to reduce its computational cost while preserving the estimation performance [13]–[16].

Inspired by data-reuse LMS-type algorithms [17]–[19], the data-reuse RLS (DR-RLS) algorithm was recently introduced in order to improve its misadjustment and tracking capability by reusing the same input data several times in a computationally efficient manner [20]. The rational behind DR-RLS is similar to that of the VFF-RLS. The fast DR-RLS algorithm was proposed to tackle the numerical instability of the FRLS, for instance when dealing with nonstationary inputs in acoustic echo cancellation problems [21]. Based on data-reused principle, the low-complexity DR-RLS algorithm with dichotomous coordinate descent iterations was introduced [22].

No theoretical analysis of the performance of the DR-RLS algorithm has been proposed so far in the literature, essentially because of the difficulty in evaluating statistically the effects of the data-reused term. Here we address this issue by carrying out a detailed analysis of the performance of the DR-RLS in the mean and mean-square error sense. Simulation results are provided to validate the resulting theoretical transient and steady-state models. This work therefore contributes to the general comprehension of the stochastic convergence characteristics of the DR-RLS algorithm.

Notation: The transpose of a vector or matrix is represented by $(\cdot)^\top$. $\mathbb{E}\{\cdot\}$ is the expected value of its argument. Superscript $(\cdot)^{-1}$ refers to the inverse of a square matrix, and $\text{tr}\{\cdot\}$ denotes its trace. The vectorization operator that stacks the columns of a matrix is refered to as $\text{vec}\{\cdot\}$. Kronecker product is denoted by $\otimes$. Matrix $\mathbf{I}_N$ is the $N\times N$ identity matrix. Notation $\binom{n}{k}$ refers to the number of subsets of size k of a n-element set. The squared norm of column vector $\mathbf{x}$ weighted by any positive definite matrix $\mathbf{\Sigma}$ is given by $\|\mathbf{x}\|^2_{\mathbf{\Sigma}} = \mathbf{x}^\top\mathbf{\Sigma}\mathbf{x}$.

II. Preliminaries and DR-RLS algorithm

Consider an unknown system defined by:

$$y_n = \mathbf{x}_n^\top \mathbf{w}^\star + z_n \tag{1}$$

where $\mathbf{x}_n = [x(n), \dots, x(n-L+1)]^\top$ is the real-valued regression vector at instant n, and $\mathbf{w}^\star \in \mathbb{R}^L$ is an unknown parameter vector to be estimated. We assume that the correlation matrix $\mathbf{R}_\mathbf{x} = \mathbb{E}\{\mathbf{x}_n\mathbf{x}_n^\top\}$ is positive definite. The observation noise z_n is modeled as a white Gaussian noise with zero-mean and variance σ_z^2. We assume that z_n is statistically independent of all other signals.

DR-RLS algorithm allows to achieve both a reasonably low misadjustment error and good tracking ability [20], [21]:

$$\mathbf{P}_n = \lambda^{-1}\left(\mathbf{P}_{n-1} - \frac{\lambda^{-1}\mathbf{P}_{n-1}\mathbf{x}_n\mathbf{x}_n^\top\mathbf{P}_{n-1}}{1+\lambda^{-1}\mathbf{x}_n^\top\mathbf{P}_{n-1}\mathbf{x}_n}\right), \tag{2}$$

$$\alpha_n = 1 - \mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n, \tag{3}$$

$$e_n = y_n - \mathbf{x}_n^\top\mathbf{w}_{n-1}, \tag{4}$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + e_n\frac{1-\alpha_n^M}{1-\alpha_n}\mathbf{P}_n\mathbf{x}_n, \tag{5}$$

with $0 \ll \lambda < 1$ the forgetting factor, and M the data-reuse parameter. Here matrix $\mathbf{P}_n$ is an estimate of the inverse of the

Manuscript received March 28, 2023; revised May 31, 2023.

This work was supported in part by the National NSFC under Grants 62171205 and 62171380.

Wei Gao is with the School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China (email: wei_gao@ujs.edu.cn).

Jie Chen, Wentao Shi, and Qunfei Zhang are with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China (email: dr.jie.chen@ieee.org; swt@nwpu.edu.cn; zhangqf@nwpu.edu.cn).

Cédric Richard is with the CNRS, OCA, Université Côte d'Azur, 06108 Nice, France (e-mail: cedric.richard@unice.fr).

time-averaged correlation matrix $\mathbf{\Phi}_n$ of input data [1], [2], defined as:

$$\mathbf{\Phi}_n = \sum_{i=0}^{n} \lambda^{n-i} \mathbf{x}_i \mathbf{x}_i^\top + \delta \lambda^{n+1} \mathbf{I}_L = \lambda \mathbf{\Phi}_{n-1} + \mathbf{x}_n \mathbf{x}_n^\top. \quad (6)$$

Setting $M = 1$, the DR-RLS algorithm reduces to the RLS. Replacing (3) into (5), the recursion of the DR-RLS can be reformulated as:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \left(\mathbf{x}_n^\top \mathbf{P}_n \mathbf{x}_n\right)^{i-1} e_n \mathbf{P}_n \mathbf{x}_n. \quad (7)$$

III. Transient performance

We now analyze the convergence behavior of the DR-RLS in the mean and mean-square error sense. Defining the weight error vector $\widetilde{\mathbf{w}}_n$ at time instant n between $\mathbf{w}_n$ and $\mathbf{w}^\star$ as:

$$\widetilde{\mathbf{w}}_n = \mathbf{w}_n - \mathbf{w}^\star, \quad (8)$$

we aim to study the recursion of $\mathbb{E}\{\widetilde{\mathbf{w}}_n\}$ and its correlation matrix $\mathbf{K}_n = \mathbb{E}\big\{\widetilde{\mathbf{w}}_n \widetilde{\mathbf{w}}_n^\top\big\}$ over time.

Before proceeding, for mathematical tractability, we need to introduce the following simplifying assumption.

Assumption A1: The weight error vector $\widetilde{\mathbf{w}}_{n-1}$ is statistically independent of the regression input vector $\mathbf{x}_n$.

The above independence assumption A1 is commonly used for analyzing the transient behavior of adaptive filters [1], [2].

A. Mean error transient behavior model

We start with the error analysis of the DR-RLS in the mean error sense. By taking the expected value of (6), we obtain:

$$\mathbb{E}\{\mathbf{\Phi}_n\} = \lambda \mathbb{E}\{\mathbf{\Phi}_{n-1}\} + \mathbf{R}_\mathbf{x} \quad (9)$$

with the initial condition $\mathbf{\Phi}_0 = \delta \mathbf{I}_L$ and $\delta > 0$ a parameter. Note that (9) will be utilized to derive the analytical models below. Substituting (1) into (4), then using definition (8), the instantaneous estimation error (4) becomes:

$$e_n = z_n - \mathbf{x}_n^\top \widetilde{\mathbf{w}}_{n-1}. \quad (10)$$

Subtracting the optimal weight vector $\mathbf{w}^\star$ from both sides of (7), with (8) and (10) it follows that:

$$\begin{aligned}\widetilde{\mathbf{w}}_n = \widetilde{\mathbf{w}}_{n-1} &+ \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \left(\mathbf{x}_n^\top \mathbf{P}_n \mathbf{x}_n\right)^{i-1} z_n \mathbf{P}_n \mathbf{x}_n \\ &- \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \left(\mathbf{x}_n^\top \mathbf{P}_n \mathbf{x}_n\right)^{i-1} \mathbf{P}_n \mathbf{x}_n \mathbf{x}_n^\top \widetilde{\mathbf{w}}_{n-1}. \quad (11)\end{aligned}$$

Pre-multiplying both sides of the above equation by $\mathbf{P}_n^{-1}$, and using the definition of $\mathbf{\Phi}_n = \mathbf{P}_n^{-1}$, we have:

$$\begin{aligned}\mathbf{\Phi}_n \widetilde{\mathbf{w}}_n = \mathbf{\Phi}_n \widetilde{\mathbf{w}}_{n-1} &+ \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \left(\mathbf{x}_n^\top \mathbf{P}_n \mathbf{x}_n\right)^{i-1} z_n \mathbf{x}_n \\ &- \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \left(\mathbf{x}_n^\top \mathbf{P}_n \mathbf{x}_n\right)^{i-1} \mathbf{x}_n \mathbf{x}_n^\top \widetilde{\mathbf{w}}_{n-1}. \quad (12)\end{aligned}$$

Taking the expected value of (12), then considering the statistical independence of noise z_n with any other signal, and $\mathbb{E}\{z_n\} = 0$, yields:

$$\begin{aligned}\mathbb{E}\big\{\mathbf{\Phi}_n \widetilde{\mathbf{w}}_n\big\} &= \mathbb{E}\big\{\mathbf{\Phi}_n \widetilde{\mathbf{w}}_{n-1}\big\} \quad (13) \\ &- \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \mathbb{E}\left\{\left(\mathbf{x}_n^\top \mathbf{P}_n \mathbf{x}_n\right)^{i-1} \mathbf{x}_n \mathbf{x}_n^\top \widetilde{\mathbf{w}}_{n-1}\right\}.\end{aligned}$$

The second term on the r.h.s. of (13) can be approximated as follows:

$$\begin{aligned}&\sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \mathbb{E}\left\{\left(\mathbf{x}_n^\top \mathbf{P}_n \mathbf{x}_n\right)^{i-1} \mathbf{x}_n \mathbf{x}_n^\top \widetilde{\mathbf{w}}_{n-1}\right\} \quad (14) \\ &\overset{(a)}{\approx} \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \mathbb{E}\left\{\left(\|\mathbf{x}_n\|^2_{\mathbf{P}_n}\right)^{i-1} \mathbf{x}_n \mathbf{x}_n^\top\right\} \mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\} \\ &\overset{(b)}{\approx} \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \mathbb{E}\left\{\left(\|\mathbf{x}_n\|^2_{\mathbf{P}_n}\right)^{i-1}\right\} \mathbf{R}_\mathbf{x} \mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\} \\ &\overset{(c)}{\approx} \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \mathrm{tr}\big\{\mathbf{R}_\mathbf{x} \mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\big\}^{i-1} \mathbf{R}_\mathbf{x} \mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}.\end{aligned}$$

Approximation (a) results from assumption A1. The rational behind approximation (b) is as follows. First, we observe that the entries of matrix $\mathbf{x}_n \mathbf{x}_n^\top$ are defined by $x(n-i)x(n-j)$. A common approximation that works well for reasonably large regression vector $\mathbf{x}_n$ is to ignore the correlation between its square norm $\|\mathbf{x}_n\|^2_{\mathbf{P}_n}$ and $x(n-i)x(n-j)$ and $\|\mathbf{x}_n\|^2_{\mathbf{P}_n}$, since the former tends to vary much slowly around its mean value $(1-\lambda)L$ [23]–[25]. Approximation (c) results from a similar argument as above. The rational of the above approximations is experimentally confirmed in the simulations section. Finally, with approximation $\mathbb{E}\{\mathbf{\Phi}_n \widetilde{\mathbf{w}}_n\} \approx \mathbb{E}\{\mathbf{\Phi}_n\}\mathbb{E}\{\widetilde{\mathbf{w}}_n\}$ successfully utilized in [26]–[28], and $\mathbb{E}\{\mathbf{\Phi}_n \widetilde{\mathbf{w}}_{n-1}\} \approx \mathbb{E}\{\mathbf{\Phi}_n\}\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}$ studied in Appendix A, and then using (14), (13) becomes:

$$\begin{aligned}\mathbb{E}\{\mathbf{\Phi}_n\}\mathbb{E}\{\widetilde{\mathbf{w}}_n\} &= \mathbb{E}\{\mathbf{\Phi}_n\}\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\} \quad (15) \\ &- \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \mathrm{tr}\big\{\mathbf{R}_\mathbf{x} \mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\big\}^{i-1} \mathbf{R}_\mathbf{x} \mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}.\end{aligned}$$

Pre-multiplying equation (15) by $\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}$, the mean weight error behavior for the DR-RLS algorithm is given by:

$$\begin{aligned}\mathbb{E}\{\widetilde{\mathbf{w}}_n\} = \Big[\mathbf{I}_L &- \sum_{i=1}^{M} \binom{M}{i} (-1)^{i+1} \mathrm{tr}\big\{\mathbf{R}_\mathbf{x} \mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\big\}^{i-1} \\ &\times \mathbb{E}\{\mathbf{\Phi}_n\}^{-1} \mathbf{R}_\mathbf{x}\Big] \mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\} \quad (16)\end{aligned}$$

where (9) needs to be utilized to implement recursion (16).

B. Mean-square error transient behavior model

We now focus on the mean-square error of the DR-RLS. By squaring the estimation error (10), taking its expectation, and then utilizing A1 and the statistics of z_n, the mean-square error (MSE) at time instant n can be approximated as:

$$\mathrm{MSE}_n = \mathbb{E}\{e_n^2\} \approx \sigma_z^2 + \mathrm{tr}\big\{\mathbf{R}_\mathbf{x} \mathbf{K}_{n-1}\big\}. \quad (17)$$

Note that the last term on the r.h.s. of equation (17) denotes the instantaneous excess MSE (EMSE) [1], [2]. The mean-square-deviation (MSD) at time instant n is defined by [2]:

$$\text{MSD}_n = \mathbb{E}\{\|\widetilde{\mathbf{w}}_n\|^2\} = \text{tr}\{\mathbf{K}_n\}. \tag{18}$$

To evaluate (17) and (18), we have to devise a recursion for calculating $\mathbf{K}_n$ in the following. Post-multiplying recursive relation (12) by its transpose, taking the expected value, then considering the statistics of z_n, yields:

$$\begin{aligned}\mathbb{E}\{\mathbf{\Phi}_n\widetilde{\mathbf{w}}_n\widetilde{\mathbf{w}}_n^\top\mathbf{\Phi}_n\} = \mathbb{E}\{\mathbf{\Phi}_n\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Phi}_n\}\\ + \mathbf{T}_1 + \mathbf{T}_2 - (\mathbf{T}_3 + \mathbf{T}_3^\top)\end{aligned} \tag{19}$$

where

$$\begin{aligned}\mathbf{T}_1 = \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\mathbb{E}\Big\{\left(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n\right)^{i+j-2}\\ \times z_n^2\mathbf{x}_n\mathbf{x}_n^\top\Big\},\end{aligned} \tag{20}$$

$$\begin{aligned}\mathbf{T}_2 = \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\mathbb{E}\Big\{\left(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n\right)^{i+j-2}\\ \times \mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{x}_n\mathbf{x}_n^\top\Big\},\end{aligned} \tag{21}$$

$$\begin{aligned}\mathbf{T}_3 = \sum_{i=1}^{M}\binom{M}{i}(-1)^{i+1}\mathbb{E}\Big\{\left(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n\right)^{i-1}\\ \times \mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Phi}_n\Big\}.\end{aligned} \tag{22}$$

We shall now focus on calculating matrices $\mathbf{T}_1$ to $\mathbf{T}_3$. Applying similar approximations as in (14), and using the statistical property of z_n, matrix $\mathbf{T}_1$ can be approximated by:

$$\begin{aligned}\mathbf{T}_1 \approx \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\text{tr}\{\mathbf{R}_\mathbf{x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\}^{i+j-2}\\ \times \sigma_z^2\mathbf{R}_\mathbf{x}.\end{aligned} \tag{23}$$

Using A1 and Isserlis' theorem, matrix $\mathbf{T}_2$ can be determined as follows [1], [29]:

$$\begin{aligned}\mathbf{T}_2 &\approx \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\text{tr}\{\mathbf{R}_\mathbf{x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\}^{i+j-2}\\ &\quad\times \mathbb{E}\{\mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{x}_n\mathbf{x}_n^\top\}\\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\text{tr}\{\mathbf{R}_\mathbf{x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\}^{i+j-2}\\ &\quad\times \left(\mathbf{R}_\mathbf{x}\text{tr}\{\mathbf{R}_\mathbf{x}\mathbf{K}_{n-1}\} + 2\,\mathbf{R}_\mathbf{x}\mathbf{K}_{n-1}\mathbf{R}_\mathbf{x}\right).\end{aligned} \tag{24}$$

To evaluate matrix $\mathbf{T}_3$, we introduce the following approximation studied in Appendix A:

$$\begin{aligned}\mathbb{E}\Big\{\left(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n\right)^{i-1}\mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Phi}_n\Big\}\\ \approx \mathbb{E}\Big\{\left(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n\right)^{i-1}\mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\Big\}\mathbb{E}\{\mathbf{\Phi}_n\}.\end{aligned} \tag{25}$$

Using (25) and considering A1, matrix $\mathbf{T}_3$ is given by:

$$\mathbf{T}_3 \approx \sum_{i=1}^{M}\binom{M}{i}(-1)^{i+1}\text{tr}\{\mathbf{R}_\mathbf{x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\}^{i-1}\mathbf{R}_\mathbf{x}\mathbf{K}_{n-1}\mathbb{E}\{\mathbf{\Phi}_n\}. \tag{26}$$

Replacing (23), (24), and (26) into (19), using two necessary approximations $\mathbb{E}\{\mathbf{\Phi}_n\widetilde{\mathbf{w}}_n\widetilde{\mathbf{w}}_n^\top\mathbf{\Phi}_n\} \approx \mathbb{E}\{\mathbf{\Phi}_n\}\mathbf{K}_n\mathbb{E}\{\mathbf{\Phi}_n\}$ and $\mathbb{E}\{\mathbf{\Phi}_n\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Phi}_n\} \approx \mathbb{E}\{\mathbf{\Phi}_n\}\mathbf{K}_{n-1}\mathbb{E}\{\mathbf{\Phi}_n\}$, then pre-multiplying and post-multiplying it by $\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}$ simultaneously, we arrive at the mean-square error behavior for DR-RLS algorithm given by (27), shown at the top of next page.

IV. Steady-state performance

We shall now investigate the error of the DR-RLS at steady-state from the transient models devised in the last section. Because $0 \ll \lambda < 1$ and using (6), the steady-state expectation of matrix $\mathbf{P}_n$ can be approximately computed as follows [2], [30], [31]:

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\{\mathbf{P}_n\} &\approx \lim_{n\to\infty}\mathbb{E}\{\mathbf{P}_n^{-1}\}^{-1}\\ &= \lim_{n\to\infty}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1} = (1-\lambda)\,\mathbf{R}_\mathbf{x}^{-1}.\end{aligned} \tag{28}$$

Theorem 1 (Mean stability): Assume that model (1) and A1 hold. The weight error vector of DR-RLS algorithm converges to zeros vector as $n \to \infty$, if the forgetting factor λ satisfies:

$$1 - \frac{2}{(M-1)L} < \lambda < 1. \tag{29}$$

Proof: Assume (28) is approximately held when n is sufficiently large. Replacing (28) into (16) as $n \to \infty$, it results in the mean weight error of the DR-RLS at steady-state:

$$\mathbb{E}\{\widetilde{\mathbf{w}}_n\} = \Big[1 - \sum_{i=1}^{M}\binom{M}{i}(-1)^{i+1}(1-\lambda)^iL^{i-1}\Big]\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\}, \tag{30}$$

as $n \to \infty$. To guarantee the convergence in the mean, (30) implies that:

$$0 < \sum_{i=1}^{M}\binom{M}{i}(-1)^{i+1}(1-\lambda)^iL^{i-1} < 1. \tag{31}$$

Since $0 \ll \lambda < 1$, the r.h.s. of (31) holds on. Hence, we only need to prove the l.h.s. of (31). When M is odd, we find that the last term is positive. Consider now the case when M is even. Then, we split all the terms defined by (31) into $M/2$ sub-terms, namely,

$$\begin{aligned}\frac{M!}{(M-1)!}(1-\lambda) - \frac{M!}{2!(M-2)!}(1-\lambda)^2L > 0, \ldots,\\ \frac{M!}{(M-1)!}(1-\lambda)^{M-1}L^{M-2} - (1-\lambda)^ML^{M-1} > 0,\end{aligned}$$

which are positive provided that (29) is satisfied. Therefore, we conclude that DR-RLS solution is asymptotically unbiased if condition is satisfied (29). ■

Let us define the steady-state correlation matrix of $\widetilde{\mathbf{w}}_n$ by $\mathbf{K}_\infty = \lim_{n\to\infty}\mathbf{K}_n$. Substituting (28) into (27), and assuming that the DR-RLS converges as $n \to \infty$, yields:

$$2(\rho_1 - \rho_2)\mathbf{K}_\infty = \rho_2\sigma_z^2\mathbf{R}_\mathbf{x}^{-1} + \rho_2\text{tr}\{\mathbf{R}_\mathbf{x}\mathbf{K}_\infty\}\mathbf{R}_\mathbf{x}^{-1} \tag{32}$$

with

$$\rho_1 = \sum_{i=1}^{M}\binom{M}{i}(-1)^{i+1}\big[L(1-\lambda)\big]^{i-1}(1-\lambda), \tag{33}$$

$$\mathbf{K}_n = \mathbf{K}_{n-1} + \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\mathrm{tr}\{\mathbf{R_x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\}^{i+j-2}\sigma_z^2\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\mathbf{R_x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}$$
$$+ \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\mathrm{tr}\{\mathbf{R_x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\}^{i+j-2}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\big(\mathbf{R_x}\mathrm{tr}\{\mathbf{R_x}\mathbf{K}_{n-1}\} + 2\mathbf{R_x}\mathbf{K}_{n-1}\mathbf{R_x}\big)\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}$$
$$- \sum_{i=1}^{M}\binom{M}{i}(-1)^{i+1}\mathrm{tr}\{\mathbf{R_x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\}^{i-1}\left(\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\mathbf{R_x}\mathbf{K}_{n-1} + \mathbf{K}_{n-1}\mathbf{R_x}\mathbb{E}\{\mathbf{\Phi}_n\}^{-1}\right) \tag{27}$$

$$\rho_2 = \sum_{i=1}^{M}\sum_{j=1}^{M}\binom{M}{i}\binom{M}{j}(-1)^{i+j+2}\big[L(1-\lambda)\big]^{i+j-2}(1-\lambda)^2. \tag{34}$$

Consider the following properties:

$$\mathrm{tr}\{\mathbf{\Sigma}\mathbf{B}\} = \big[\mathrm{vec}(\mathbf{B}^\top)\big]^\top\boldsymbol{\sigma}, \tag{35}$$
$$\mathrm{vec}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}) = (\mathbf{B}^\top\otimes\mathbf{A})\boldsymbol{\sigma}, \tag{36}$$

with $\boldsymbol{\sigma} = \mathrm{vec}(\mathbf{\Sigma})$. Post-multiplying both sides of (32) by $\mathbf{R_x}$, and applying property (35), we have:

$$2(\rho_1-\rho_2)\mathbf{K}_\infty\mathbf{R_x} = \rho_2\sigma_z^2\mathbf{I}_L + \rho_2\big(\big[\mathrm{vec}(\mathbf{R_x})\big]^\top\mathbf{k}_\infty\big)\mathbf{I}_L \tag{37}$$

with $\mathbf{k}_\infty = \mathrm{vec}(\mathbf{K}_\infty)$. By vectorizing both sides of (37), and using property (36), it results that:

$$2(\rho_1-\rho_2)(\mathbf{R_x}\otimes\mathbf{I}_L)\mathbf{k}_\infty = \rho_2\sigma_z^2\mathrm{vec}(\mathbf{I}_L) + \rho_2\mathrm{vec}(\mathbf{I}_L)\big[\mathrm{vec}(\mathbf{R_x})\big]^\top\mathbf{k}_\infty. \tag{38}$$

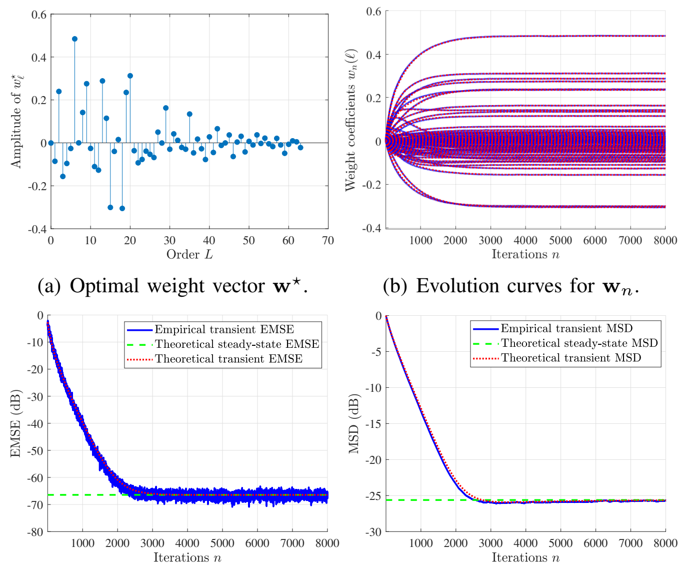
Then, we obtain:

$$\mathbf{k}_\infty = \rho_2\sigma_z^2\big[2(\rho_1-\rho_2)(\mathbf{R_x}\otimes\mathbf{I}_L) - \rho_2\mathrm{vec}(\mathbf{I}_L)\big[\mathrm{vec}(\mathbf{R_x})\big]^\top\big]^{-1}\mathrm{vec}(\mathbf{I}_L) \tag{39}$$

which allows to recover the steady-state matrix $\mathbf{K}_\infty$. Finally, we can compute the steady-state EMSE or MSE, and MSD of the DR-RLS according to (17) and (18) as $n\to\infty$.

V. Simulation Results

The overall performance of the DR-RLS has already been validated thoroughly in [20], [21]. We now verify the correctness and precision of the presented theoretical models with simulations. All empirical curves were averaged over 200 Monte Carlo runs. The input signal was obtained by filtering a zero-mean white Gaussian noise $s(n)$ through an AR(1) model, namely, $x(n) = 0.6x(n-1) + s(n)$, where variances σ_s^2 and σ_x^2 were set to 0.64 and 1, respectively. The noise variance σ_z^2 was set to 0.02. Figure 1(a) depicts $\mathbf{w}^\star \in \mathbb{R}^{64}$ generated from a standard normal distribution and scaled by an exponential decay factor 0.5. The data-reuse parameter M was set to 4. The forgetting factor λ and the parameter δ were set to 0.9995 and 1×10^3, respectively. Weight vector $\mathbf{w}_0$ was initialized to zero. Figure 1(b) illustrates the mean weight behavior of the DR-RLS. We can find that all the theoretical curves (dotted red) of weight coefficients $w_n(\ell)$ predicted by (16) are consistent with the empirical curves (solid blue), which validates the asymptotic unbiasedness of the algorithm. Figure 1(c) illustrates the good consistency between the empirical learning curves of the EMSE and its theoretical prediction obtained from (17), (27), and (39) in transient and steady-state phases. Figure 1(d) shows that the empirical learning curves of the MSD generally match well with their transient and steady-state theoretical predictions based on (18), (27), and (39). To conclude, Fig. 1 illustrates the accuracy of our models and the rationality of all approximations used in the analysis.



(a) Optimal weight vector $\mathbf{w}^\star$. (b) Evolution curves for $\mathbf{w}_n$. (c) EMSE curves. (d) MSD curves.

Fig. 1. Comparisons of empirical results and theoretical predictions.

Appendix A
Proofs of Approximations

To justify the consistency of some approximations introduced in this paper, we decompose $\mathbf{\Phi}_n = \mathbb{E}\{\mathbf{\Phi}_n\} + \mathbf{\Delta}_n$ with $\mathbf{\Delta}_n$ some random fluctuations. We have:

$$\mathbb{E}\{\mathbf{\Phi}_n\widetilde{\mathbf{w}}_{n-1}\} = \mathbb{E}\{\mathbf{\Phi}_n\}\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\} + \mathbb{E}\{\mathbf{\Delta}_n\widetilde{\mathbf{w}}_{n-1}\}, \tag{40}$$

$$\begin{aligned}&\mathbb{E}\{(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n)^{i-1}\mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Phi}_n\}\\ &= \mathbb{E}\{(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n)^{i-1}\mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\}\mathbb{E}\{\mathbf{\Phi}_n\}\\ &\quad + \mathbb{E}\{(\mathbf{x}_n^\top\mathbf{P}_n\mathbf{x}_n)^{i-1}\mathbf{x}_n\mathbf{x}_n^\top\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Delta}_n\},\end{aligned} \tag{41}$$

$$\begin{aligned}\mathbb{E}\{\mathbf{\Phi}_n\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Phi}_n\} &= \mathbb{E}\{\mathbf{\Phi}_n\}\mathbf{K}_{n-1}\mathbb{E}\{\mathbf{\Phi}_n\}\\ &+ \mathbb{E}\{\mathbf{\Phi}_n\}\mathbb{E}\{\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Delta}_n\} + \mathbb{E}\{\mathbf{\Delta}_n\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\mathbf{\Delta}_n\}\\ &+ \mathbb{E}\{\mathbf{\Delta}_n\widetilde{\mathbf{w}}_{n-1}\widetilde{\mathbf{w}}_{n-1}^\top\}\mathbb{E}\{\mathbf{\Phi}_n\}.\end{aligned} \tag{42}$$

We assume that the entries of $\mathbf{\Delta}_n$ are small in comparison with those of $\mathbb{E}\{\mathbf{\Phi}_n\}$ because equation (6) means that $\mathbf{\Phi}_n$ is a low-pass filtering of $\mathbf{x}_n^\top\mathbf{x}_n$ [26]–[28]. Therefore, the first term on the r.h.s. of (40)–(42) dominates the remaining ones, leading to the approximations.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 2nd ed. New Jersey: Prentice-Hall, 1991.

[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[3] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, 4th ed. New York, USA: Springer, 2013.

[4] J. Cioffi and T. Kailath, "Fast, recursive-least-squares transversal filters for adaptive filtering," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 304–337, 1984.

[5] A. Carini and E. Mumolo, "A numerically stable fast RLS algorithm for adaptive filtering and prediction based on the UD factorization," *IEEE Trans. on Signal Process.*, vol. 47, no. 8, pp. 2309–2313, 1999.

[6] J. Benesty and T. Gänsler, "New insights into the RLS algorithm," *EURASIP J. Adv. Signal Process.*, vol. 3, pp. 331–339, 2004.

[7] L. Rey Vega, H. Rey, J. Benesty, and S. Tressens, "A fast robust recursive least-squares algorithm," *IEEE Trans. on Signal Process.*, vol. 57, no. 3, pp. 1209–1216, Mar. 2009.

[8] B. Toplis and S. Pasupathy, "Tracking improvements in fast RLS algorithms using a variable forgetting factor," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 206–227, 1988.

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

[10] 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.

[11] Y. Cai, R. C. de Lamare, M. Zhao, and J. Zhong, "Low-complexity variable forgetting factor mechanism for blind adaptive constrained constant modulus algorithms," *IEEE Trans. on Signal Process.*, vol. 60, no. 8, pp. 3988–4002, 2012.

[12] Y. J. Chu and S. C. Chan, "A new local polynomial modeling-based variable forgetting factor RLS algorithm and its acoustic applications," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 2059–2069, 2015.

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

[14] Y. V. Zakharov and V. H. Nascimento, "DCD-RLS adaptive filters with penalties for sparse identification," *IEEE Trans. on Signal Process.*, vol. 61, no. 12, pp. 3198–3213, 2013.

[15] C. Elisei-Iliescu, C. Paleologu, J. Benesty, C. Stanciu, C. Anghel, and S. Ciochină, "Recursive least-squares algorithms for the identification of low-rank systems," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 5, pp. 903–918, May 2019.

[16] Y. Yu, L. Lu, Z. Zheng, W. Wang, Y. Zakharov, and R. C. de Lamare, "DCD-based recursive adaptive algorithms robust against impulsive noise," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 7, pp. 1359–1363, 2020.

[17] J. Apolinário, M. L. R. Campos, and P. S. R. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," *IEEE Trans. on Signal Process.*, vol. 48, no. 11, pp. 3235–3242, 2000.

[18] P. S. R. Diniz and S. Werner, "Set-membership binormalized data-reusing LMS algorithms," *IEEE Trans. on Signal Process.*, vol. 51, no. 1, pp. 124–134, 2003.

[19] R. A. Soni, K. A. Gallivan, and W. K. Jenkins, "Low-complexity data reusing methods in adaptive filtering," *IEEE Trans. on Signal Process.*, vol. 52, no. 2, pp. 394–405, 2004.

[20] C. Paleologu, J. Benesty, and S. Ciochină, "Data-reuse recursive least-squares algorithms," *IEEE Signal Process. Lett.*, vol. 29, pp. 752–756, 2022.

[21] L. M. Dogariu, C. Paleologu, J. Benesty, and S. Ciochina, "On the performance of a data-reuse fast RLS algorithm for acoustic echo cancellation," in *IEEE BlackSeaCom*, 2022, pp. 135–140.

[22] I. D. Fîciu, C. L. Stanciu, C. Elisei-Iliescu, C. Anghel, and R. M. Udrea, "Low-complexity implementation of a data-reuse RLS algorithm," in *45th International Conference on Telecommunications and Signal Processing (TSP)*, 2022, pp. 289–293.

[23] C. Samson and V. Reddy, "Fixed point error analysis of the normalized ladder algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 5, pp. 1177–1191, 1983.

[24] S. de Almeida, J. Bermudez, N. Bershad, and M. Costa, "A statistical analysis of the affine projection algorithm for unity step size and autoregressive inputs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 7, pp. 1394–1405, 2005.

[25] J. Chen, C. Richard, J.-C. M. Bermudez, and P. Honeine, "Variants of non-negative least-mean-square algorithm and convergence analysis," *IEEE Trans. on Signal Process.*, vol. 62, no. 15, pp. 3990–4005, 2014.

[26] E. Eweda, N. J. Bershad, and J. C. M. Bermudez, "Stochastic analysis of the recursive least squares algorithm for cyclostationary colored inputs," *IEEE Trans. on Signal Process.*, vol. 68, pp. 676–686, 2020.

[27] W. Gao, J. Chen, and C. Richard, "Transient theoretical analysis of diffusion RLS algorithm for cyclostationary colored inputs," *IEEE Signal Process. Lett.*, vol. 28, pp. 1160–1164, 2021.

[28] W. Gao, J. Chen, C. Richard, W. Shi, and Q. Zhang, "Transient performance analysis of the ℓ_1-RLS," *IEEE Signal Process. Lett.*, vol. 29, pp. 90–94, 2022.

[29] J. Chen, C. Richard, Y. Song, and D. Brie, "Transient performance analysis of zero-attracting LMS," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1786–1790, 2016.

[30] A. Bertrand, M. Moonen, and A. H. Sayed, "Diffusion bias-compensated RLS estimation over adaptive networks," *IEEE Trans. on Signal Process.*, vol. 59, no. 11, pp. 5212–5224, Nov. 2011.

[31] R. Arablouei, K. Dogancay, S. Werner, and Y. Huang, "Adaptive distributed estimation based on recursive least-squares and partial diffusion," *IEEE Trans. on Signal Process.*, vol. 62, no. 14, pp. 3510–3522, Jul. 2014.