INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2016; **30**:715–735
Published online 7 October 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/acs.2634

On the stability and convergence of a sliding-window variable-regularization recursive-least-squares algorithm

Asad A. Ali[1], Jesse B. Hoagg[2,*,†], Magnus Mossberg[3] and Dennis S. Bernstein[1]

[1]*Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, USA*
[2]*Department of Mechanical Engineering, University of Kentucky, Lexington, KY 40506, USA*
[3]*Department of Engineering and Physics, Karlstad University, SE-65188 Karlstad, Sweden*

SUMMARY

A sliding-window variable-regularization recursive-least-squares algorithm is derived, and its convergence properties, computational complexity, and numerical stability are analyzed. The algorithm operates on a finite data window and allows for time-varying regularization in the weighting and the difference between estimates. Numerical examples are provided to compare the performance of this technique with the least mean squares and affine projection algorithms. Copyright © 2015 John Wiley & Sons, Ltd.

Received 11 March 2015; Revised 27 July 2015; Accepted 11 September 2015

KEY WORDS: variable regularization; sliding-window RLS; digital signal processing

1. INTRODUCTION

Recursive-least-squares (RLS) and gradient-based algorithms are widely used in signal processing, estimation, identification, and control [1–9]. Under ideal conditions, that is, noiseless measurements and persistency of the data, these techniques are guaranteed to converge to the minimizer of a quadratic function [2, 5]. In practice, the accuracy of the estimates depends on the level of noise and the persistency of the data.

The standard RLS algorithm operates on a growing window of data, where new data are added to the RLS cost function as they become available, and past data are progressively discounted through the use of a forgetting factor. In contrast, sliding-window RLS algorithms [10–14] require no forgetting factor because they operate on a finite data window of fixed length, where new data replace past data in the RLS cost function. Sliding-window least squares techniques are available in both batch and recursive formulations. As shown in [11], sliding-window RLS algorithms have enhanced tracking performance compared with standard RLS algorithms in the presence of time-varying parameters.

In standard RLS, the positive definite initialization of the covariance matrix is the inverse of the weighting on a regularization term in a quadratic cost function. This regularization term compensates for the potential lack of persistency, ensuring that the cost function has a unique minimizer at each step. Traditionally, the regularization term is fixed for all steps of the recursion. An optimally regularized adaptive filtering algorithm with constant regularization is presented in [15]. However, variants of RLS with time-varying regularization have been developed in the context of adaptive filtering, echo cancelation, and affine projection [16–21].

*Correspondence to: Jesse B. Hoagg, Department of Mechanical Engineering, University of Kentucky, 151 Ralph G. Anderson Building, Lexington, KY 40507, USA.
†E-mail: jhoagg@engr.uky.edu

In the present work, we derive a novel sliding-window variable-regularization RLS (SW-VR-RLS) algorithm, where the weighting on the regularization term can change at each step. An additional extension presented in this paper also involves the regularization term. Specifically, the regularization term in standard RLS weights the difference between the next estimate and the initial estimate, while the regularization term in sliding-window RLS weights the difference between the next estimate and the estimate at the beginning of the sliding window. In the present paper, the regularization term weights the difference between the next estimate and an arbitrarily chosen time-varying vector. As a special case, the time-varying vector can be the current estimate or a recent estimate. These variable-regularization extensions of sliding-window RLS can facilitate trade-offs among transient error, rate of convergence, and steady-state error.

We derive the SW-VR-RLS equations and analyze their convergence properties in the absence of noise. While standard RLS entails the update of the estimate and the covariance matrix, sliding-window RLS involves the update of an additional symmetric matrix of size $n \times n$, where n is the dimension of the estimate. Furthermore, SW-VR-RLS requires updating of one more symmetric matrix of size $n \times n$ to account for the time-varying regularization.

A preliminary version of the SW-VR-RLS algorithm appeared in the conference proceedings [22] without any analysis of convergence or numerical stability. The goal of the present paper is to provide a more complete development of the SW-VR-RLS algorithm, including an analysis of convergence and numerical stability.

In this paper, a matrix $A \in \mathbb{R}^{n\times n}$ is positive semidefinite ($A \geqslant 0$) if it is symmetric and has nonnegative eigenvalues, and a matrix $A \in \mathbb{R}^{n\times n}$ is positive definite ($A > 0$) if it is symmetric and has positive eigenvalues. Furthermore, if $A \in \mathbb{R}^{n\times n}$, then $||A||$ denotes the maximum singular value of A, and if $x \in \mathbb{R}^n$, then $||x||$ denotes the Euclidean norm of x.

2. THE NON-RECURSIVE SOLUTION

Let r be a non-negative integer. For all integers $i \geqslant -r$, let $b_i \in \mathbb{R}^n$ and $A_i \in \mathbb{R}^{n\times n}$, where A_i is positive semidefinite. For all integers $i \geqslant 0$, let $\alpha_i \in \mathbb{R}^n$ and $R_i \in \mathbb{R}^{n\times n}$, where R_i is positive semidefinite. Assume that, for all $k \geqslant 0$, $\sum_{i=k-r}^{k-1} A_i + R_k$ is positive definite. In practice, the matrix A_k and the vector b_k depend on data, whereas the regularization weighting R_k and regularization parameter α_k are chosen by the user. For all $k \geqslant 0$, the sliding-window, variable-regularization quadratic cost is defined by

$$J_k(x) \triangleq \sum_{i=k-r}^{k} x^{\mathrm{T}} A_i x + b_i^{\mathrm{T}} x + (x-\alpha_k)^{\mathrm{T}} R_k (x-\alpha_k), \tag{1}$$

where $x \in \mathbb{R}^n$ and $x_0 = -\frac{1}{2}\left(\sum_{i=-r}^{0} A_i + R_0\right)^{-1}\left(\sum_{i=-r}^{0} b_i - 2R_0\alpha_0\right)$ is the minimizer of $J_0(x)$. Note that the regularization term $(x-\alpha_k)^{\mathrm{T}} R_k (x-\alpha_k)$ in (1) contains weighting R_k and parameter α_k, which are potentially time varying. For all $k \geqslant 0$, the unique minimizer x_k of (1) is

$$x_k = -\frac{1}{2}\left(\sum_{i=k-r}^{k} A_i + R_k\right)^{-1}\left(\sum_{i=k-r}^{k} b_i - 2R_k\alpha_k\right). \tag{2}$$

Example 1

Consider the weighted regularized least squares cost function

$$\mathcal{J}_k(x) \triangleq \sum_{i=k-r}^{k} \left(y_i - F_i^{\mathrm{T}} x\right)^{\mathrm{T}} W_i \left(y_i - F_i^{\mathrm{T}} x\right) + (x-\alpha_k)^{\mathrm{T}} R_k (x-\alpha_k),$$

where $x \in \mathbb{R}^n$. Let r be a non-negative integer; and, for all $i \geqslant -r$, let $y_i \in \mathbb{R}^l$, $\alpha_i \in \mathbb{R}^n$, $F_i \in \mathbb{R}^{n\times l}$, $R_i \in \mathbb{R}^{n\times n}$, and $W_i \in \mathbb{R}^{l\times l}$, where W_i is positive definite. Furthermore, for all $i \geqslant -r$, define $A_i \triangleq F_i W_i F_i^{\mathrm{T}}$ and $b_i \triangleq -2F_i W_i y_i$. Then, for all $k \geqslant 0$ and $x \in \mathbb{R}^n$, $\mathcal{J}_k(x) = J_k(x) +$

DOI: 10.1002/acs

$\sum_{i=k-r}^{k} y_i^{\mathrm{T}} W_i y_i$. Thus, the minimizer of $J_k(x)$ is also the minimizer of $\mathcal{J}_k(x)$. Moreover, it follows from (2) that the minimizer of $\mathcal{J}_k(x)$ is given by

$$x_k = \left(\sum_{i=k-r}^{k} F_i W_i F_i^{\mathrm{T}} + R_k \right)^{-1} \left(\sum_{i=k-r}^{k} F_i W_i y_i + R_k \alpha_k \right).$$

Example 2
Let n and r be positive integers, for $i \in \{1, \ldots, n\}$, let $a_i, c_i \in \mathbb{R}$, and, for all $i \geqslant -r-n$, let u_i, $y_i \in \mathbb{R}$. Furthermore, for all $k \geqslant 0$, let y_k satisfy the infinite impulse response model

$$y_k = \sum_{i=1}^{n} a_i y_{k-i} + \sum_{i=1}^{n} c_i u_{k-i}.$$

Next, for all $i \geqslant -r$, define $\psi_i \triangleq [u_{i-1} \quad \cdots \quad u_{i-n} \quad y_{i-1} \quad \cdots \quad y_{i-n}]^{\mathrm{T}}$, and consider the cost (1), where $A_i \triangleq \psi_i \psi_i^{\mathrm{T}}$ and $b_i \triangleq -2 y_i \psi_i$. Define $x_* \triangleq [a_1 \quad \cdots \quad a_n \quad c_1 \quad \cdots \quad c_n]^{\mathrm{T}}$. The objective is to choose the regularization parameters R_k and α_k such that the sequence of minimizers $\{x_k\}_{k=0}^{\infty}$ of (1) converges to x_*. Note that, for all $k \geqslant -r$, rank $A_k \leqslant 1$. As shown in Section 4, the rank of A_k affects the computational complexity of the recursive formulation of (2).

3. THE SW-VR-RLS SOLUTION

Defining

$$P_k \triangleq \left(\sum_{i=k-r}^{k} A_i + R_k \right)^{-1}, \tag{3}$$

it follows that (2) can be written as

$$x_k = -\frac{1}{2} P_k \left(\sum_{i=k-r}^{k} b_i - 2 R_k \alpha_k \right). \tag{4}$$

To express P_k recursively, consider the decomposition

$$A_k = \psi_k \psi_k^{\mathrm{T}}, \tag{5}$$

where $\psi_k \in \mathbb{R}^{n \times n_k}$ and $n_k \triangleq \operatorname{rank} A_k$. Next, for all $k \geqslant 1$, define

$$Q_k \triangleq \left(\sum_{i=k-r}^{k-1} A_i + R_k \right)^{-1} = \left(P_k^{-1} - A_k \right)^{-1}. \tag{6}$$

It follows from (5) and (6) that

$$P_k = \left(Q_k^{-1} + \psi_k \psi_k^{\mathrm{T}} \right)^{-1}.$$

Using the matrix inversion lemma

$$(X + UCV)^{-1} = X^{-1} - X^{-1} U \left(C^{-1} + V X^{-1} U \right)^{-1} V X^{-1}, \tag{7}$$

with $X = Q_k^{-1}$, $U = \psi_k$, $C = I_{n_k}$, and $V = \psi_k^{\mathrm{T}}$, where I_{n_k} is the $n_k \times n_k$ identity matrix, implies that

$$P_k = Q_k - Q_k \psi_k \left(I_{n_k} + \psi_k^{\mathrm{T}} Q_k \psi_k \right)^{-1} \psi_k^{\mathrm{T}} Q_k.$$

DOI: 10.1002/acs

To express Q_k recursively, for all $k \geqslant 1$, define

$$L_k \triangleq \left(\sum_{i=k-r-1}^{k-1} A_i + R_k \right)^{-1} = \left(Q_k^{-1} - A_{k-r-1}\right)^{-1} = \left(Q_k^{-1} - \psi_{k-r-1}\psi_{k-r-1}^{\mathrm{T}}\right)^{-1}. \tag{8}$$

Using (7) with $X = L_k^{-1}$, $U = \psi_{k-r-1}$, $C = -I_{n_{k-r-1}}$, and $V = \psi_{k-r-1}^{\mathrm{T}}$, it follows that

$$Q_k = L_k - L_k \psi_{k-r-1} \left(-I_{n_{k-r-1}} + \psi_{k-r-1}^{\mathrm{T}} L_k \psi_{k-r-1}\right)^{-1} \psi_{k-r-1}^{\mathrm{T}} L_k.$$

To express L_k recursively, we substitute (3) into itself to obtain

$$P_k^{-1} = \sum_{i=k-r}^{k} A_i + R_k = P_{k-1}^{-1} + A_k - A_{k-r-1} + R_k - R_{k-1}. \tag{9}$$

Thus, it follows from (6), (8), and (9) that

$$L_k = \left(P_{k-1}^{-1} + R_k - R_{k-1}\right)^{-1}. \tag{10}$$

Next, we factor $R_k - R_{k-1}$ as

$$R_k - R_{k-1} = \phi_k S_k \phi_k^{\mathrm{T}}, \tag{11}$$

where $\phi_k \in \mathbb{R}^{n \times m_k}$, $m_k \triangleq \operatorname{rank}(R_k - R_{k-1})$, and $S_k \in \mathbb{R}^{m_k \times m_k}$ has the form $S_k \triangleq \operatorname{diag}(\pm 1, \ldots, \pm 1)$. Using (7) with $X = P_{k-1}^{-1}$, $U = \phi_k$, $C = S_k$, and $V = \phi_k^{\mathrm{T}}$, it follows from (10) that

$$L_k = P_{k-1} - P_{k-1}\phi_k \left(S_k + \phi_k^{\mathrm{T}} P_{k-1} \phi_k\right)^{-1} \phi_k^{\mathrm{T}} P_{k-1}.$$

We now summarize the SW-VR-RLS algorithm.

Algorithm 1
For each $k \geqslant 1$, the unique minimizer x_k of (1) is given by

$$L_k = P_{k-1} - P_{k-1}\phi_k \left(S_k + \phi_k^{\mathrm{T}} P_{k-1} \phi_k\right)^{-1} \phi_k^{\mathrm{T}} P_{k-1}, \tag{12}$$

$$Q_k = L_k - L_k \psi_{k-r-1} \left(-I_{n_{k-r-1}} + \psi_{k-r-1}^{\mathrm{T}} L_k \psi_{k-r-1}\right)^{-1} \psi_{k-r-1}^{\mathrm{T}} L_k, \tag{13}$$

$$P_k = Q_k - Q_k \psi_k \left(I_{n_k} + \psi_k^{\mathrm{T}} Q_k \psi_k\right)^{-1} \psi_k^{\mathrm{T}} Q_k, \tag{14}$$

$$x_k = -\frac{1}{2} P_k \left(\sum_{i=k-r}^{k} b_i - 2R_k \alpha_k \right), \tag{15}$$

where $P_0 = \left(\sum_{i=-r}^{0} A_i + R_0\right)^{-1}$ and $x_0 = -\frac{1}{2} P_0 \left(\sum_{i=-r}^{0} b_i - 2R_0\alpha_0\right)$.

As an alternative to Algorithm 1, the equation for x_k can be expressed using the recursion matrix P_k. First, it follows from (15) that

$$\sum_{i=k-r-1}^{k-1} b_i = -2P_{k-1}^{-1} x_{k-1} + 2R_{k-1}\alpha_{k-1}. \tag{16}$$

Using (9) and (16), it follows that (15) can be written as

$$
\begin{aligned}
x_k &= -\frac{1}{2} P_k \left(\sum_{i=k-r-1}^{k-1} b_i + b_k - b_{k-r-1} - 2R_k\alpha_k \right) \\
&= -\frac{1}{2} P_k \left(-2P_{k-1}^{-1} x_{k-1} + 2R_{k-1}\alpha_{k-1} + b_k - b_{k-r-1} - 2R_k\alpha_k \right) \\
&= -\frac{1}{2} P_k \left[-2(P_k^{-1} - A_k + A_{k-r-1} - R_k + R_{k-1}) x_{k-1} + 2R_{k-1}\alpha_{k-1} + b_k - b_{k-r-1} - 2R_k\alpha_k \right] \\
&= x_{k-1} - P_k \left[(A_k - A_{k-r-1}) x_{k-1} + (R_k - R_{k-1}) x_{k-1} + R_{k-1}\alpha_{k-1} + \frac{1}{2}(b_k - b_{k-r-1}) - R_k\alpha_k \right].
\end{aligned}
$$

We now summarize the alternative SW-VR-RLS algorithm.

Algorithm 2
For each $k \geqslant 1$, the unique minimizer x_k of (1) is given by

$$L_k = P_{k-1} - P_{k-1}\phi_k \left(S_k + \phi_k^{\mathrm{T}} P_{k-1} \phi_k \right)^{-1} \phi_k^{\mathrm{T}} P_{k-1}, \tag{17}$$

$$Q_k = L_k - L_k \psi_{k-r-1} \left(-I_{n_{k-r-1}} + \psi_{k-r-1}^{\mathrm{T}} L_k \psi_{k-r-1} \right)^{-1} \psi_{k-r-1}^{\mathrm{T}} L_k, \tag{18}$$

$$P_k = Q_k - Q_k \psi_k \left(I_{n_k} + \psi_k^{\mathrm{T}} Q_k \psi_k \right)^{-1} \psi_k^{\mathrm{T}} Q_k, \tag{19}$$

$$
\begin{aligned}
x_k = x_{k-1} - P_k &\left[(A_k - A_{k-r-1}) x_{k-1} + \frac{1}{2}(b_k - b_{k-r-1}) \right] \\
&- P_k \left[(R_k - R_{k-1}) x_{k-1} + R_{k-1}\alpha_{k-1} - R_k\alpha_k \right],
\end{aligned} \tag{20}
$$

where $P_0 = \left(\sum_{i=-r}^{0} A_i + R_0 \right)^{-1}$ and $x_0 = -\frac{1}{2} P_0 \left(\sum_{i=-r}^{0} b_i - 2R_0\alpha_0 \right)$.

The theoretical properties and computational complexity of Algorithms 1 and 2 are identical, but their numerical properties are different, which will be discussed in Section 7.

If, for all $i \in \{-r, \ldots, 0\}$, $A_i = 0$ and $b_i = 0$, then $x_0 = \alpha_0$ and $P_0 = R_0^{-1}$. Furthermore, if the regularization weighting R_k is constant, that is, for all $k \geqslant 0$, $R_k = R_0 > 0$, then (11) implies that $\phi_k = 0$ and (17) simplifies to $L_k = P_{k-1}$, and thus, computation of L_k is not required.

4. COMPUTATIONAL COMPLEXITY

First, consider Algorithm 1. The computational complexity of the matrix products and inverse in (12) is $\mathcal{O}\left(n^2 m_k\right)$ and $\mathcal{O}\left(m_k^3\right)$, respectively, where $m_k = \operatorname{rank}\left(R_k - R_{k-1}\right) \leqslant n$. Hence, (12) is $\mathcal{O}\left(n^2 m_k\right)$. In particular, if, for all $k \geqslant 0$, $m_k = 1$, then the inverse in (12) is a scalar inverse, and (12) is $\mathcal{O}(n^2)$.

The matrix products and inverse in (14) are $\mathcal{O}\left(n^2 n_k\right)$ and $\mathcal{O}\left(n_k^3\right)$, respectively, where $n_k = \operatorname{rank} A_k \leqslant n$. Hence, (14) is $\mathcal{O}\left(n^2 n_k\right)$. Similarly, (13) is $\mathcal{O}\left(n^2 n_{k-r-1}\right)$. In particular, if, for all $k \geqslant 0$, $n_k = 1$, then the inverses in (13) and (14) are scalar inverses, and (13) and (14) are $\mathcal{O}(n^2)$.

Finally, note that (15) is $\mathcal{O}(n^2)$. Therefore, if, for all $k \geqslant 0$, $\operatorname{rank}(R_k - R_{k-1}) = 1$ and $\operatorname{rank} A_k = 1$, then the computational complexity of Algorithm 1 is $\mathcal{O}(n^2)$.

Now, consider Algorithm 2. Because (17), (18), and (19) are identical to (12), (13), and (14), respectively, and (20) is $\mathcal{O}(n^2)$, it follows that the computational complexity of Algorithm 2 is identical to the computational complexity of Algorithm 1.

5. CONVERGENCE ANALYSIS OF SW-VR-RLS

Definition 1 ([23]*)*
Let $x_{\mathrm{eq}} \in \mathbb{R}^n$, and consider the nonlinear time-varying system

$$x_{k+1} = f(x_k, k), \tag{21}$$

where $f : \mathbb{R}^n \times \{0, 1, 2, \ldots\} \to \mathbb{R}^n$ is a continuous function such that, for all $k \geqslant 0$, $f(x_{\mathrm{eq}}, k) = x_{\mathrm{eq}}$. The equilibrium solution $x_k \equiv x_{\mathrm{eq}}$ of (21) is *Lyapunov stable* if, for every $\varepsilon > 0$ and $k_0 \geqslant 0$, there exists $\delta(\varepsilon, k_0) > 0$ such that $||x_{k_0} - x_{\mathrm{eq}}|| < \delta$ implies that, for all $k \geqslant k_0$, $||x_k - x_{\mathrm{eq}}|| < \varepsilon$. The equilibrium solution $x_k \equiv x_{\mathrm{eq}}$ of (21) is *uniformly Lyapunov stable* if, for every $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that, for all $k_0 \geqslant 0$, $||x_{k_0} - x_{\mathrm{eq}}|| < \delta$ implies that, for all $k \geqslant k_0$, $||x_k - x_{\mathrm{eq}}|| < \varepsilon$. The equilibrium solution $x_k \equiv x_{\mathrm{eq}}$ of (21) is *globally asymptotically stable* if it is Lyapunov stable and, for all $k_0 \geqslant 0$ and $x_{k_0} \in \mathbb{R}^n$, $\lim_{k\to\infty} x_k = x_{\mathrm{eq}}$.

The following result provides boundedness properties of the SW-VR-RLS algorithm. The proof is in Appendix A. This result applies to both SW-VR-RLS implementations, specifically, Algorithms 1 and 2.

Theorem 1
For all $k \geqslant 0$, let $T_k \in \mathbb{R}^{n\times n}$ be positive definite, and assume that there exist $\varepsilon_1, \varepsilon_2 \in (0, \infty)$ such that, for all $k \geqslant 0$,

$$\varepsilon_1 I_n \leqslant T_{k+1} \leqslant T_k \leqslant \varepsilon_2 I_n. \tag{22}$$

Furthermore, for all $k \geqslant 0$, let $\xi_k \in \mathbb{R}$; assume that $0 < \inf_{k\geqslant 0} \xi_k \leqslant \sup_{k\geqslant 0} \xi_k < \infty$; and define $R_k \triangleq \xi_k T_k$. Then, the following statements hold:

(i) $\{L_k\}_{k=1}^{\infty}$, $\{Q_k\}_{k=1}^{\infty}$, and $\{P_k\}_{k=0}^{\infty}$ are bounded.
(ii) Assume that $\{\alpha_k\}_{k=0}^{\infty}$ and $\{b_k\}_{k=0}^{\infty}$ are bounded. Then, $\{x_k\}_{k=0}^{\infty}$ is bounded.

For all $k \geqslant 0$, define $\Phi_k \triangleq [\psi_k \ \ \cdots \ \ \psi_{k-r}] \in \mathbb{R}^{n\times q_k}$, where $q_k \triangleq \sum_{i=0}^{r} n_{k-i}$, so that $\sum_{i=k-r}^{k} A_i = \Phi_k \Phi_k^{\mathrm{T}}$. Furthermore, using the matrix inversion lemma, it follows from (3) that

$$P_k = R_k^{-1} - R_k^{-1}\Phi_k \left(I_{q_k} + \Phi_k^{\mathrm{T}} R_k^{-1} \Phi_k\right)^{-1} \Phi_k^{\mathrm{T}} R_k^{-1}. \tag{23}$$

Next, let ν be a positive integer, for all $k \geqslant \nu$, let $\alpha_k = x_{k-\nu}$, for all $k > \nu - 1$, define $\chi_k \triangleq \left[x_k^{\mathrm{T}} \ \ x_{k-1}^{\mathrm{T}} \ \ \cdots \ \ x_{k-\nu+1}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{n\nu}$, and, for all $i \in \{1, \ldots, \nu\}$, let $\chi_{k,i} \triangleq x_{k-i+1}$. Then, it follows from (15) that, for all $k > \nu - 2$,

$$\begin{bmatrix} \chi_{k+1,1} \\ \chi_{k+1,2} \\ \vdots \\ \chi_{k+1,\nu} \end{bmatrix} = \begin{bmatrix} -P_{k+1}\left(\displaystyle\sum_{i=k+1-r}^{k+1} \frac{1}{2} b_i - R_{k+1}\chi_{k,\nu}\right) \\ \chi_{k,1} \\ \vdots \\ \chi_{k,\nu-1} \end{bmatrix}. \tag{24}$$

The following result provides stability and convergence properties of the SW-VR-RLS algorithm. The proof is in the Appendix. This result applies to both SW-VR-RLS implementations.

Theorem 2
For all $k \geqslant 0$, let $T_k \in \mathbb{R}^{n\times n}$ be positive definite, and assume that there exist $\varepsilon_1, \varepsilon_2 \in (0, \infty)$ such that, for all $k \geqslant 0$, (22) holds. Furthermore, for all $k \geqslant 0$, let $\xi_k \in \mathbb{R}$; assume that $0 < \inf_{k\geqslant 0} \xi_k \leqslant \sup_{k\geqslant 0} \xi_k < \infty$; and define $R_k \triangleq \xi_k T_k$. Let ν be a positive integer; let $\eta \in \mathbb{R}^n$; for all $0 \leqslant k \leqslant \nu - 1$, define $\alpha_k \triangleq \eta$; and, for all $k \geqslant \nu$, define $\alpha_k \triangleq x_{k-\nu}$, where $x_{k-\nu}$ satisfies (4).

DOI: 10.1002/acs

Let $P_0 = \left(\sum_{i=-r}^{0} A_i + R_0\right)^{-1}$ and $x_0 = -\frac{1}{2}P_0\left(\sum_{i=-r}^{0} b_i - 2R_0\eta\right)$, assume that there exists a unique $x_* \in \mathbb{R}^n$ such that, for all $k \geqslant 0$,

$$A_k x_* + \frac{1}{2}b_k = 0, \tag{25}$$

and define $\chi_* \triangleq \begin{bmatrix} x_*^{\mathrm{T}} & x_*^{\mathrm{T}} & \cdots & x_*^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{n\nu}$. Then, the following statements hold:

(i) $\chi_k \equiv \chi_*$ is an equilibrium solution of (24).
(ii) The equilibrium solution $\chi_k \equiv \chi_*$ of (24) is uniformly Lyapunov stable, and, for all $x_0 \in \mathbb{R}^n$, $\{x_k\}_{k=0}^{\infty}$ is bounded.
(iii) $\sum_{j=\nu}^{\infty}(x_{j-\nu} - x_*)^{\mathrm{T}}\Phi_j\left[\xi_j I_{q_j} + \Phi_j^{\mathrm{T}} T_j^{-1}\Phi_j\right]^{-1}\Phi_j^{\mathrm{T}}(x_{j-\nu} - x_*)$ and $\sum_{j=\nu}^{\infty}||x_j - x_{j-\nu}||^2$ exist.
(iv) Assume that $\{A_k\}_{k=0}^{\infty}$ is bounded. Then, $\lim_{k\to\infty}\psi_k^{\mathrm{T}}(x_{k-\nu} - x^*) = 0$ and $\lim_{k\to\infty}\left(A_k x_k + \frac{1}{2}b_k\right) = 0$.
(v) Assume that $\{A_k\}_{k=0}^{\infty}$ is bounded, and assume that there exists $c > 0$ and a non-negative integer l such that, for all $k \geqslant \nu l - r$, $cI_n \leqslant \sum_{i=0}^{l} A_{k-\nu i}$. Then, for all $x_0 \in \mathbb{R}^n$, $\lim_{k\to\infty} x_k = x_*$, and $\chi_k \equiv \chi_*$ is globally asymptotically stable.

6. SIMULATIONS

In this section, we study the effect of R_k and r on SW-VR-RLS and compare SW-VR-RLS with the proportionate affine projection algorithm (PAPA) [24] and the proportionate normalized least mean squares (PNLMS) algorithm [25] for systems, where x_* changes abruptly.

Let ℓ be the number of data points, and, for any sequence $\{p_k\}_{k=1}^{\ell}$, define the root-mean-square value

$$\sigma_p \triangleq \sqrt{\frac{1}{\ell}\sum_{k=1}^{\ell} p_k^2}.$$

Let n be a positive integer, and, for $i \in \{0, \ldots, n-1\}$, let $h_i \in \mathbb{R}$. Define $x_* \triangleq [h_0 \quad h_1 \quad \cdots \quad h_{n-1}]^{\mathrm{T}}$. For all $k \geqslant 1$, let $u_k \in \mathbb{R}$, and, for all $-r-n+1 \leqslant k \leqslant 0$, let $u_k = 0$ and $y_k = 0$. Furthermore, for all $k \geqslant 1$, let y_k satisfy the finite impulse response

$$y_k = \sum_{i=0}^{n-1} h_i u_{k-i}. \tag{26}$$

Next, for all $k \geqslant -r-n+1$, define the noisy output $\bar{y}_k \triangleq y_k + w_k$, where, for all $-r-n+1 \leqslant k \leqslant 0$, $w_k = 0$, and, for all $k \geqslant 1$, $w_k \in \mathbb{R}$ is sampled from a white noise process with a zero-mean Gaussian distribution with variance σ_w^2. Define the signal-to-noise ratio (SNR) SNR $\triangleq \sigma_y/\sigma_w$.

Let $x \in \mathbb{R}^n$, for all $k \geqslant -r$, define $\psi_k \triangleq [u_k \quad \cdots \quad u_{k-n+1}]^{\mathrm{T}}$, and, for all $k \geqslant 0$, define the cost function

$$\mathcal{J}_k(x) \triangleq \sum_{i=k-r}^{k}\left(\bar{y}_k - \psi_k^{\mathrm{T}}x\right)^{\mathrm{T}}\left(\bar{y}_k - \psi_k^{\mathrm{T}}x\right) + (x-\alpha_k)^{\mathrm{T}}R_k(x-\alpha_k). \tag{27}$$

For all $k \geqslant -r$, define $A_k \triangleq \psi_k\psi_k^{\mathrm{T}}$ and $b_k \triangleq -2\bar{y}_k\psi_k$. It follows from (27) that

$$\mathcal{J}_k(x) = \sum_{i=k-r}^{k}\left(x^{\mathrm{T}}A_k x + b_k x_k\right) + (x-\alpha_k)^{\mathrm{T}}R_k(x-\alpha_k) + \sum_{i=k-r}^{k}\bar{y}_k^{\mathrm{T}}\bar{y}_k. \tag{28}$$

DOI: 10.1002/acs

Then, for all $k \geqslant 0$, $\mathcal{J}_k(x) = J_k(x) + \sum_{i=k-r}^{k} \bar{y}_i^{\mathrm{T}} \bar{y}_i$, and thus, the minimizer x_k of (28) is given by the minimizer (2) of the SW-VR-RLS cost function (1).

Next, it follows from (26) that the for all $k \geqslant 1$, $y_k = \phi^{\mathrm{T}} x_*$, where $x_* \triangleq [h_0 \quad h_1 \quad \cdots \quad h_{n-1}]^{\mathrm{T}}$ is a vector of the unknown impulse response coefficients. For all examples, $n = 15$, and

$$x_* = \begin{cases} z_1, & \text{if } 0 \leqslant k \leqslant 999, \\ z_2, & \text{if } k \geqslant 1000, \end{cases} \tag{29}$$

where z_1 and z_2 are randomly selected. In all examples, they are

$$\begin{aligned} z_1 \triangleq \big[&-1.0667\ 0.9337\ 0.3503\ -0.0290\ 0.1825\ -1.5651\ -0.0845\ 1.6039 \\ &0.0983\ 0.0414\ -0.7342\ -0.0308\ 0.2323\ 0.4264\ -0.3728 \big]^{\mathrm{T}} \in \mathbb{R}^{15}, \\ z_2 \triangleq \big[&-0.0835\ 0.8205\ -1.3594\ 1.4417\ 0.8726\ 0.4442\ -0.2222\ -0.8215 \\ &0.5131\ -0.6638\ 0.1265\ -0.0155\ -0.1581\ 0.6957\ -0.8379 \big]^{\mathrm{T}} \in \mathbb{R}^{15}. \end{aligned}$$

For all examples, we use Algorithm 1, where $\alpha_0 = x_0$, and for all $k \geqslant 1$, $\alpha_k = x_{k-1}$. Define the performance $\varepsilon_k \triangleq 20\log_{10}(\|x_* - x_k\|/\|x_*\|)$. We compute the ensemble average of ε_k based on 100 simulations with independent realizations of u_k and w_k.

6.1. *Effect of R_k*

First, we examine the effect of R_k on the performance of SW-VR-RLS, where $R_k \equiv R$ is constant and the coefficients of (26) change abruptly at $k = 1000$. Let $r = 60$, for all $k \geqslant 0$, let u_k be sampled from a white noise process with a zero-mean Gaussian distribution with variance 10, and let x_* be given by (29).

We test SW-VR-RLS for three values of R_k and three values of SNR. Specifically, $R = 1000 I_n$, $R = 10000 I_n$, and $R = 30000 I_n$. Figure 1 shows that, for this example, not only a smaller value of R yields faster convergence of ε_k but also a larger asymptotic mean value of ε_k. Furthermore, for each R, a larger value of SNR yields a smaller asymptotic mean value of ε_k.

To understand why a smaller value of R yields a larger asymptotic mean value of ε_k in the case of noisy data, first, note that a smaller R makes the regularization term $(x_k - x_{k-1})^{\mathrm{T}} R (x_k - x_{k-1})$ of (1) smaller. Because the regularization term has the effect of opposing movement of the estimate x_k away from x_{k-1}, smaller R makes x_k more sensitive to noise. Furthermore, as k increases, $||x_k - x_{k-1}||$ tends to decrease to its asymptotic mean value, and thus, the regularization term

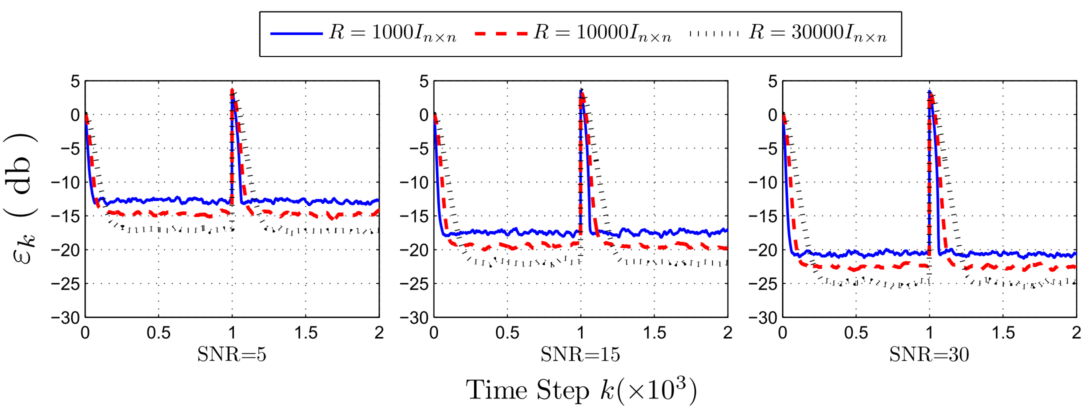

Figure 1. Effect of R_k on convergence of ε_k to its asymptotic mean value, where $R_k \equiv R$ is a constant. For this example, a smaller value of R yields faster convergence of ε_k to its asymptotic mean value but also a larger asymptotic mean value of ε_k. Furthermore, for each value of R, a larger value of SNR yields a smaller asymptotic mean value of ε_k.

DOI: 10.1002/acs

$(x_k - x_{k-1})^{\mathrm{T}} R (x_k - x_{k-1})$ decreases. Thus, a larger value of R means that the regularization term contributes more asymptotically to the cost function (1). Thus, more regularization (i.e., larger R) can make the estimate x_k asymptotically less sensitive to noise in y_k, which in turn can yield smaller asymptotic mean values of ε_k.

Next, we consider a time-varying R_k. First, define the residual $v_k \triangleq ||\bar{y}_k - \psi_k^{\mathrm{T}} x_k||$ and the filtered residual $\bar{v}_k = \gamma \bar{v}_{k-1} + (1-\gamma) v_k$, where $\gamma \in (0, 1)$ is a smoothing factor. Furthermore, let

$$R_k = \begin{cases} R_{\min} I_n, & \bar{v}_k \leqslant \rho, \\ R_{\max} I_n, & \bar{v}_k > \rho. \end{cases} \tag{30}$$

For this example, $\gamma = 0.05$, $R_{\min} = 10000$, $R_{\max} = 50000$, $\rho = 2.5$, and SNR $= 20$. Note that we allow only rank 1 modifications in R_k so that the computational complexity of SW-VR-RLS is $\mathcal{O}(n^2)$. Therefore, in order to modify R_k from $R_{\min} I_n$ to $R_{\max} I_n$, we modify the first diagonal entry of R_k at the current time step and change the next diagonal entry at the next time step and so on. Figure 2 shows that (30) yields a smaller asymptotic mean value of ε_k than $R_k \equiv 10000 I_n$ and faster convergence of ε_k to its asymptotic mean value than $R_k \equiv 50000 I_n$.

6.2. *Effect of window size*

For all $k \geqslant 0$, let u_k be sampled from a zero-mean Gaussian white noise process with variance 10, let SNR$=$ 20, let x_* be given by (29), and, for all $k \geqslant 0$, let $R_k = 1000 I_n$. We test SW-VR-

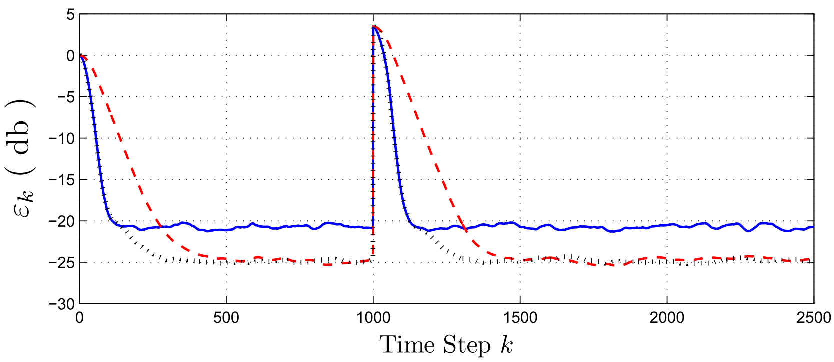

Figure 2. Effect of R_k on convergence of ε_k to its asymptotic mean value when R_k is time-varying. The solid line, dashed line, and dotted line indicate SW-VR-RLS with $R_k \equiv 10000 I_n$, $R_k \equiv 50000 I_n$, and R_k given by (30), respectively. For this example, R_k given by (30) yields a smaller asymptotic mean value of ε_k than $R_k \equiv 10000 I_n$ and yields faster convergence of ε_k to its asymptotic mean value than $R_k \equiv 50000 I_n$.

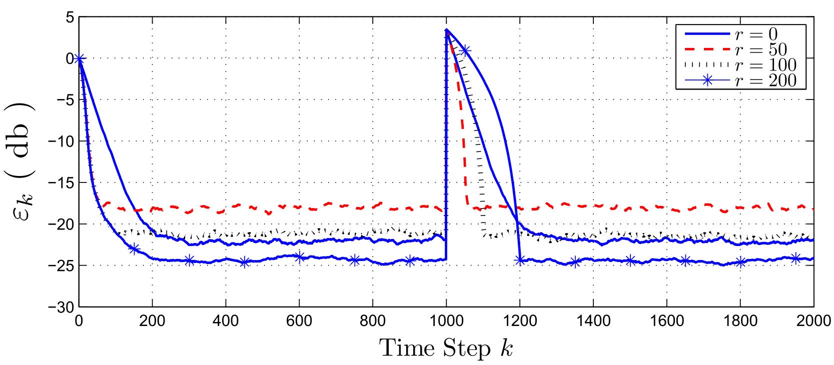

Figure 3. Effect of r on convergence of ε_k to its asymptotic mean value. This plot shows that, as r is increased from 0, the asymptotic mean value of ε_k and the speed of convergence of ε_k to its asymptotic mean value first increase and then decrease.

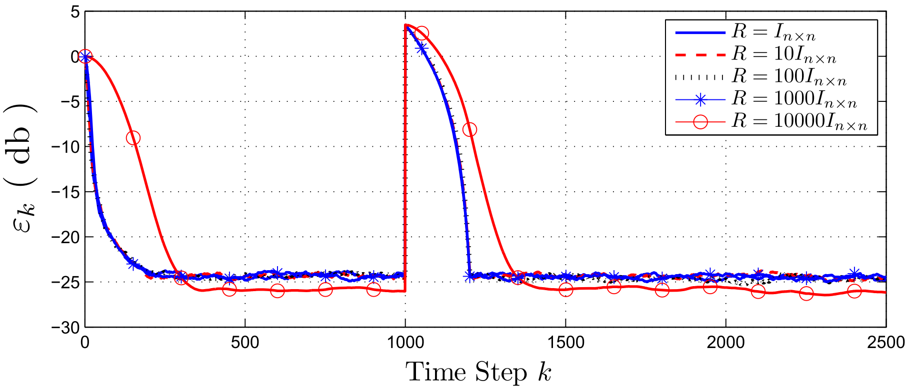

Figure 4. Effect of constant R on convergence of ε_k to its asymptotic mean value when $r = 200$. This plot shows that decreasing the value of R from $1000I_n$ to I_n does not increase either the speed of convergence or the asymptotic mean value of ε_k.

RLS with $r = 0$, $r = 50$, $r = 100$, and $r = 200$. Figure 3 shows that, as r is increased from 0, the asymptotic mean value of ε_k and the speed of convergence of ε_k to its asymptotic mean value initially increase and then decrease.

To gain further insight into how to choose r, we fix $r = 200$ and test SW-VR-RLS when $R_k \equiv R$ is constant. We test five different values of R, specifically, $R = I_n$, $R = 10I_n$, $R = 100I_n$, $R = 1000I_n$, and $R = 10000I_n$. For this simulation, Figure 4 shows that decreasing the value of R from $1000I_n$ to I_n does not increase the speed of convergence of ε_k to its asymptotic mean value. This suggests that, as R is decreased beyond a certain value, it no longer affects the speed of convergence or asymptotic mean value of ε_k, and r must be decreased in order to increase the speed of convergence of ε_k.

6.3. Comparison with PAPA and PNLMS

In this section, we compare SW-VR-RLS with PAPA [24] and PNLMS [25], which are widely used and hence chosen for comparison. PAPA and PNLMS include regularization terms that weight the difference between the current estimate x_k and previous estimate x_{k-1}. This aspect of PAPA and PNLMS is similar to SW-VR-RLS, where $\alpha_k = x_{k-1}$.

For all $k \geqslant 0$, let u_k be sampled from a white noise process with a zero-mean Gaussian distribution with variance 10, let x_* be given by (29), and let SNR $= 20$. For SW-VR-RLS, we use $r = 60$ and R_k specified by (30) with $R_{\min} = 6000$, $R_{\max} = 25\,000$, $\rho = 2.5$, and $\gamma = 0.1$. For PNLMS [25], we set δ(PNLMS) $= 0.01$, ρ(PNLMS) $= 15/(n+1)$, μ(PNLMS) $= 0.2$; and for the PAPA [24], we set δ_ρ(PAPA)$=$ 0.01, ρ(PAPA)$=$ $15/n$, μ(PAPA) $= 0.2$, and δ(PAPA)$=$ $100/n$. Note that for these parameters, all three algorithms have approximately the same mean steady-state error. Figure 5 shows that, for $k \leqslant 999$, SW-VR-RLS yields faster convergence of ε_k to its asymptotic mean value than PNLMS and PAPA. Furthermore, at $k = 1000$, $x_* \neq z_1$; and SW-VR-RLS yields faster convergence of ε_k to its new asymptotic mean value than PNLMS and PAPA.

Next, we consider the case where u_k is colored. Because convergence of SW-VR-RLS, PAPA, and PNLMS is slower in the presence of colored inputs as compared with white inputs, we consider

$$x_* = \begin{cases} z_1, & \text{if } 0 \leqslant k \leqslant 3999, \\ z_2, & \text{if } k \geqslant 4000. \end{cases}$$

Let SNR $= 20$, $\bar{u}_k$ be sampled from a white noise process with a zero-mean Gaussian distribution with variance 10, and let

$$u_k = 0.9u_{k-1} + \bar{u}_k.$$

For SW-VR-RLS, we use $r = 800$ and R_k specified by (30) with $R_{\min} = 5 \times 10^4$, $R_{\max} = 35 \times 10^4$, $\rho = 3.5$, and $\gamma = 0.01$. For PNLMS [25], we set δ(PNLMS) $= 0.05$, ρ(PNLMS) $=$

DOI: 10.1002/acs

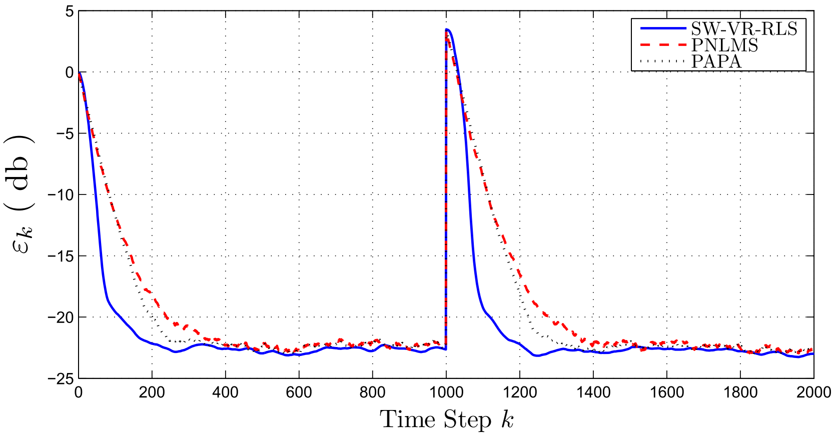

Figure 5. This plot compares SW-VR-RLS with PAPA and PNLMS when the input signal is white. For $k \leqslant 1000$, SW-VR-RLS yields faster convergence of ε_k to its asymptotic mean value than PNLMS and PAPA. Furthermore, at $k = 1000$, $x_* \neq z_1$; and SW-VR-RLS yields faster convergence of ε_k to its new asymptotic mean value than PNLMS and PAPA.

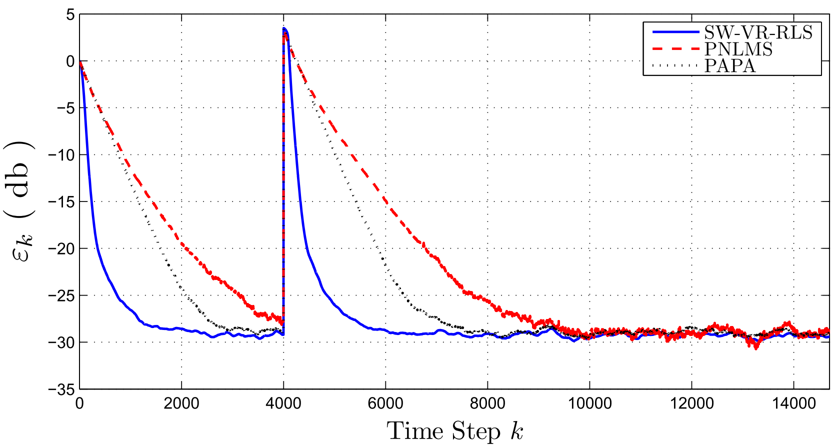

Figure 6. This plot compares SW-VR-RLS with PAPA and PNLMS when the input signal u_k is colored. For $k \leqslant 4000$, SW-VR-RLS yields faster convergence of ε_k to its asymptotic mean value than PNLMS and PAPA. Furthermore, at $k = 4000$, $x_* \neq z_1$; and SW-VR-RLS yields faster convergence of ε_k to its new asymptotic mean value than PNLMS and PAPA.

$15/(n+1)$, and $\mu(\text{PNLMS}) = 0.085$; and, for PAPA [24], we set $\delta_\rho(\text{PAPA}) = 0.01$, $\rho(\text{PAPA}) = 15/n$, $\mu(\text{PAPA}) = 0.02$, and $\delta(\text{PAPA}) = 5/n$. Note that we have chosen these parameters such that all three algorithms have approximately the steady-state mean error. Figure 6 shows that, for this example, and for $k \leqslant 3999$, SW-VR-RLS yields faster convergence of ε_k to its asymptotic mean value than PNLMS and PAPA. Furthermore, at $k = 4000$, $x_* \neq z_1$; and SW-VR-RLS yields faster convergence of ε_k to its asymptotic mean value than PNLMS and PAPA.

7. NUMERICAL STABILITY

In this section, we investigate the numerical stability of SW-VR-RLS to account for the effects of round-off and quantization errors in x_k and P_k. Throughout this section, we assume that, for all $0 \leqslant k \leqslant \nu - 1$, $\alpha_k \triangleq x_0$, and, for all $k \geqslant \nu$, $\alpha_k \triangleq x_{k-\nu}$, where ν is a positive integer.

DOI: 10.1002/acs

7.1. Numerical errors in x_k

To examine the numerical stability of Algorithms 1 and 2, we perturb x_{k_0} at step k_0 by the amount $\gamma \in \mathbb{R}^n$ and analyze the propagation of this error, assuming that all subsequent calculations are performed with infinite-precision arithmetic.

We first analyze the numerical stability of Algorithm 1, that is, we analyze the propagation of a perturbation in x_{k_0} at step k_0 assuming that, for all $k > k_0$, x_k is updated using (15). For all $k > k_0$, let $\bar{x}_k$ denote the SW-VR-RLS minimizer given by Algorithm 1, where the initial condition is $\bar{x}_{k_0} \triangleq x_{k_0} + \gamma$, where x_{k_0} is the SW-VR-RLS minimizer given by Algorithm 1 at step k_0. Thus, it follows from (15) that, for all $k \geqslant k_0$, $\bar{x}_k$ satisfies

$$\bar{x}_k = -\frac{1}{2} P_k \left(\sum_{i=k-r}^{k} b_i - 2R_k \bar{\alpha}_k \right), \tag{31}$$

where, for all $k_0 \leqslant k \leqslant k_0 + \nu - 1$, $\bar{\alpha}_k \triangleq \alpha_k$ and, for all $k \geqslant k_0 + \nu$, $\bar{\alpha}_k \triangleq \bar{x}_{k-\nu}$. For all $k \geqslant k_0$, define $\delta_k \triangleq \bar{x}_k - x_k$ and note that $\delta_{k_0} = \gamma$. It follows from (31) and (15) that, for all $k > k_0$,

$$\delta_k = P_k R_k (\bar{\alpha}_k - \alpha_k) = P_k R_k \delta_{k-\nu}, \tag{32}$$

where, for all $k_0 - \nu + 1 \leqslant k \leqslant k_0 - 1$, we define $\delta_k \triangleq 0$. For all $k > k_0 + \nu - 1$, define $\Delta_k \triangleq \begin{bmatrix} \delta_k^{\mathrm{T}} & \delta_{k-1}^{\mathrm{T}} & \cdots & \delta_{k-\nu+1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{n\nu}$ and, for all $i \in \{1, \ldots, \nu\}$, let $\Delta_{k,i} \triangleq \delta_{k-i+1}$. Then, it follows from (32) that, for all $k > k_0 + \nu - 2$,

$$\begin{bmatrix} \Delta_{k+1,1} \\ \Delta_{k+1,2} \\ \vdots \\ \Delta_{k+1,\nu} \end{bmatrix} = \begin{bmatrix} P_{k+1} R_{k+1} \Delta_{k,\nu} \\ \Delta_{k,1} \\ \vdots \\ \Delta_{k,\nu-1} \end{bmatrix}. \tag{33}$$

Note that $\Delta_k \equiv 0$ is an equilibrium solution of (33). The following result shows that, under the assumptions of Theorem 1, the equilibrium solution $\Delta_k \equiv 0$ of (33) is globally asymptotically stable. The proof is in the Appendix.

Theorem 3

Consider the error systems (17), (18), (19), and (32). For all $k \geqslant k_0$, let $T_k \in \mathbb{R}^{n\times n}$ be positive definite, and assume that there exist $\varepsilon_1, \varepsilon_2 \in (0, \infty)$ such that, for all $k \geqslant k_0$, (22) holds. Furthermore, for all $k \geqslant k_0$, let $\xi_k \in \mathbb{R}$; assume that $0 < \inf_{k \geqslant k_0} \xi_k \leqslant \sup_{k \geqslant k_0} \xi_k < \infty$; and define $R_k \triangleq \xi_k T_k$. Then, the following statements hold:

(i) $\{L_k\}_{k=k_0+1}^{\infty}$, $\{Q_k\}_{k=k_0+1}^{\infty}$, and $\{P_k\}_{k=k_0}^{\infty}$ are bounded.
(ii) The equilibrium solution $\Delta_k \equiv 0$ of (33) is uniformly Lyapunov stable, and, for all $\delta_{k_0} \in \mathbb{R}^n$, $\{\delta_k\}_{k=k_0}^{\infty}$ is bounded.
(iii) Assume that $\{A_k\}_{k=k_0}^{\infty}$ is bounded, and assume that there exists $c > 0$ and a non-negative integer l such that, for all $k \geqslant k_0 + \nu l - r$, $cI_n \leqslant \sum_{i=0}^{l} A_{k-\nu i}$. Then, for all $\delta_{k_0} \in \mathbb{R}^n$, $\lim_{k\to\infty} \delta_k = 0$, and $\Delta_k \equiv 0$ is globally asymptotically stable.

We now examine the numerical stability of Algorithm 2, that is, we analyze the propagation of a perturbation in x_{k_0} at step k_0 assuming that, for all $k > k_0$, x_k is updated using (20). For all $k > k_0$, let $\bar{x}_k$ denote the SW-VR-RLS minimizer given by Algorithm 2, where the initial condition is $\bar{x}_{k_0} \triangleq x_{k_0} + \gamma$, where x_{k_0} is the SW-VR-RLS minimizer given by Algorithm 2 at step k_0. Thus, it follows from (20) that, for all $k \geqslant k_0$, $\bar{x}_k$ satisfies

$$\begin{aligned} \bar{x}_k = {} & [I_n - P_k(A_k - A_{k-r-1} + R_k - R_{k-1})]\, \bar{x}_{k-1} - \frac{1}{2} P_k (b_k - b_{k-r-1}) \\ & + P_k R_k \bar{\alpha}_k - P_k R_{k-1} \bar{\alpha}_{k-1}, \end{aligned} \tag{34}$$

DOI: 10.1002/acs

where, for all $k_0 \leqslant k \leqslant k_0 + \nu - 1$, $\bar{\alpha}_k \triangleq \alpha_k$, and, for all $k \geqslant k_0 + \nu$, $\bar{\alpha}_k \triangleq \bar{x}_{k-\nu}$. For all $k \geqslant k_0$, define $\delta_k \triangleq \bar{x}_k - x_k$ and note that $\delta_{k_0} = \gamma$. Subtracting (20) from (34), and using (9), it follows that, for all $k > k_0$,

$$\begin{aligned}\delta_k &= \left[I_n - P_k(A_k - A_{k-r-1} + R_k - R_{k-1})\right](\bar{x}_{k-1} - x_{k-1}) \\ &\quad + P_k R_k (\bar{\alpha}_k - \alpha_k) - P_k R_{k-1} (\bar{\alpha}_{k-1} - \alpha_{k-1}) \\ &= \left(I_n - P_k \left(P_k^{-1} - P_{k-1}^{-1}\right)\right)(\bar{x}_{k-1} - x_{k-1}) + P_k R_k (\bar{\alpha}_k - \alpha_k) - P_k R_{k-1} (\bar{\alpha}_{k-1} - \alpha_{k-1}) \\ &= P_k P_{k-1}^{-1} \delta_{k-1} + P_k R_k \delta_{k-\nu} - P_k R_{k-1} \delta_{k-\nu-1},\end{aligned} \tag{35}$$

where, for all $k_0 - \nu \leqslant k \leqslant k_0 - 1$, we define $\delta_k \triangleq 0$. Note that the error dynamics (35) for Algorithm 2 are different from the error dynamics (32) for Algorithm 1. We show numerically that there exists $\delta_{k_0} \in \mathbb{R}^n$ such that δ_k for Algorithm 2 given by (35) does not decay to zero. In contrast, Theorem 3 implies that δ_k for Algorithm 1 given by (32) does decay to zero.

We now numerically test the stability of the single error propagation dynamics for x_k given by (35) and (32). Let $n = 10$, $r = 5$, and $\nu = 1$; and, for all $k \geqslant -r$, let the entries of ψ_k be generated from a zero-mean Gaussian distribution with unit variance. Furthermore, for all $k \geqslant -r$, let $A_k = \psi_k \psi_k^{\mathrm{T}}$, and, for all $k \geqslant 0$, let $R_k = I_n$. Moreover, let $\delta_{-1} = 0$, and let δ_0 be generated from a zero-mean Gaussian distribution with unit variance. Finally, for all $k \geqslant 0$, let P_k be given by (3). Figure 7 shows δ_k for (35) and (32) and shows that, for this example, δ_k given by (35) does not decay to zero, whereas δ_k given by (32) decays to zero.

Next, we test Algorithms 1 and 2 using the same setup as in Section 6.1 but with no noise, $x_* = z_1$, and a perturbation in x_k at step $k = 500$. Figure 8 shows ε_k for Algorithms 1 and 2 with perturbation (dashed line) and without perturbation (solid line) in x_k and shows that, after $k = 500$, for Algorithm 1 with perturbation, ε_k converges to the unperturbed value of ε_k, but for Algorithm 2 with perturbation, ε_k does not converge the unperturbed value of ε_k.

Since the x_k update for Algorithm 2 is derived from the x_k update for Algorithm 1, Figure 8 suggests that the derivation of the x_k update for Algorithm 2 introduces the equivalent of a pole on the unit circle at 1 of a linear time-invariant discrete-time system, due to which a perturbation in x_k does not decay. To illustrate this, let $\kappa \in \mathbb{R}$, for all $k \geqslant 0$, let $a_k \in \mathbb{R}$ be sampled from a white noise process with a zero-mean Gaussian distribution and variance 0.0025, let $b_k = a_k + 0.5\sin(0.01k)$, and, for all $k \geqslant 0$, define the asymptotically stable linear system

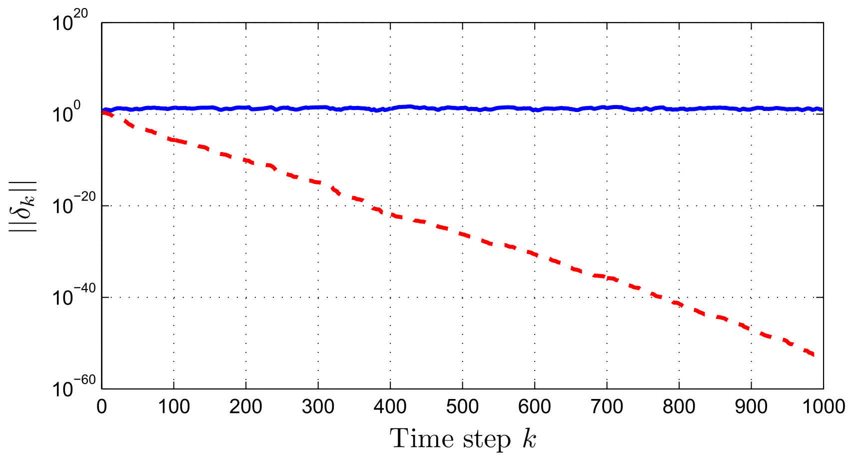

Figure 7. This plot shows the solution δ_k of the error-propagation systems for x_k given by (35) and (32). The solid line indicates the solution to (35), whereas the dashed line indicates the solution to (32). This plot shows that δ_k given by (35) does not decay to zero, whereas δ_k given by (32) decays to zero.

DOI: 10.1002/acs

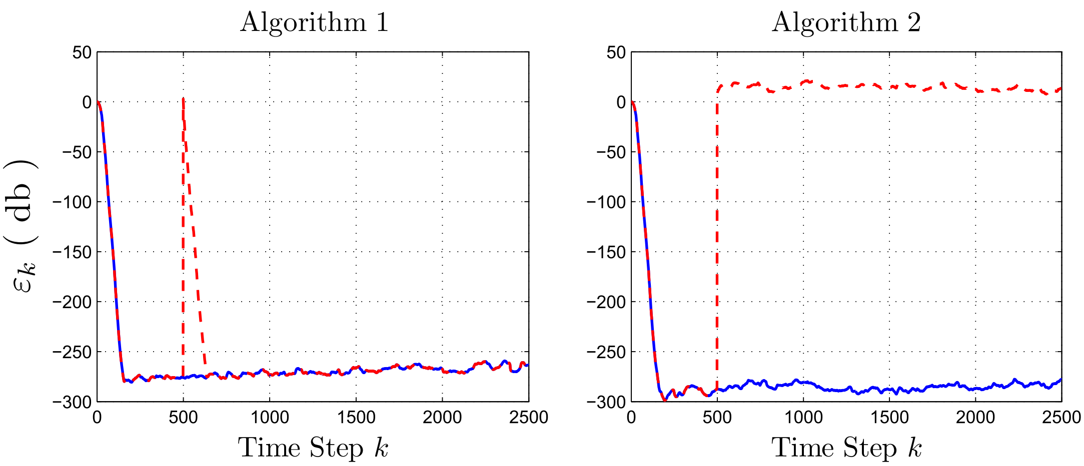

Figure 8. This plot shows ε_k for Algorithms 1 and 2 with perturbation (dashed line) and without perturbation (solid line) in x_k and shows that, after $k = 500$, for Algorithm 1 with perturbation, ε_k converges to the unperturbed value of ε_k, but for Algorithm 2 with perturbation, ε_k does not converge the unperturbed value of ε_k.

$$x_{k+1} = 0.5x_k + b_{k+1} + b_k, \tag{36}$$

with the initial condition $x_0 = \kappa$. It follows from (36) that $x_k = 0.5x_{k-1} + b_k + b_{k-1}$, and thus,

$$b_k = x_k - 0.5x_{k-1} - b_{k-1}. \tag{37}$$

Using (37) in (36) yields, for all $k \geqslant 0$,

$$x_{k+2} = 1.5x_{k+1} - 0.5x_k + b_{k+2} - b_k, \tag{38}$$

with the initial conditions $x_0 = \kappa$ and $x_1 = 0.5\kappa + b_1 + b_0$. Note that (38) has a pole at 1. Note that using (37) in (36) is similar to using (9) and (16) in (15) to obtain

$$\begin{aligned} x_k &= -\frac{1}{2}P_k\left(\sum_{i=k-r-1}^{k-1} b_i + b_k - b_{k-r-1} - 2R_k\alpha_k\right) \\ &= -\frac{1}{2}P_k\left(-2P_{k-1}^{-1}x_{k-1} + 2R_{k-1}\alpha_{k-1} + b_k - b_{k-r-1} - 2R_k\alpha_k\right), \end{aligned}$$

which is one of the steps in deriving Algorithm 2 from Algorithm 1.

Figure 9 shows x_k given by (36) and (38) with a perturbation at step $k = 200$ (dashed line) and without perturbation (solid line). After $k = 200$, for (36) with perturbation, x_k converges to the unperturbed value of x_k, but for (38) with perturbation, x_k does not converge the unperturbed value of x_k.

7.2. *Numerical errors in P_k*

We now consider the effect of round-off and quantization errors in P_k. As in the case of x_k, we perturb P_{k_0} at step k_0, and analyze the propagation of this error, assuming that all subsequent calculations are performed with infinite-precision arithmetic. Let $\Gamma \in \mathbb{R}^{n\times n}$. For all $k > k_0$, let $\bar{P}_k$ be given by Algorithm 1, where the initial conditions are $\bar{P}_{k_0} = P_{k_0} + \Gamma$, $\bar{Q}_{k_0} = Q_{k_0}$, and $\bar{L}_{k_0} = L_{k_0}$, where P_{k_0}, Q_{k_0}, and L_{k_0} are given by Algorithm 1 at step k_0. Thus, it follows that, for all $k \geqslant k_0$, $\bar{P}_k$, $\bar{Q}_k$, and $\bar{L}_k$ satisfy

DOI: 10.1002/acs

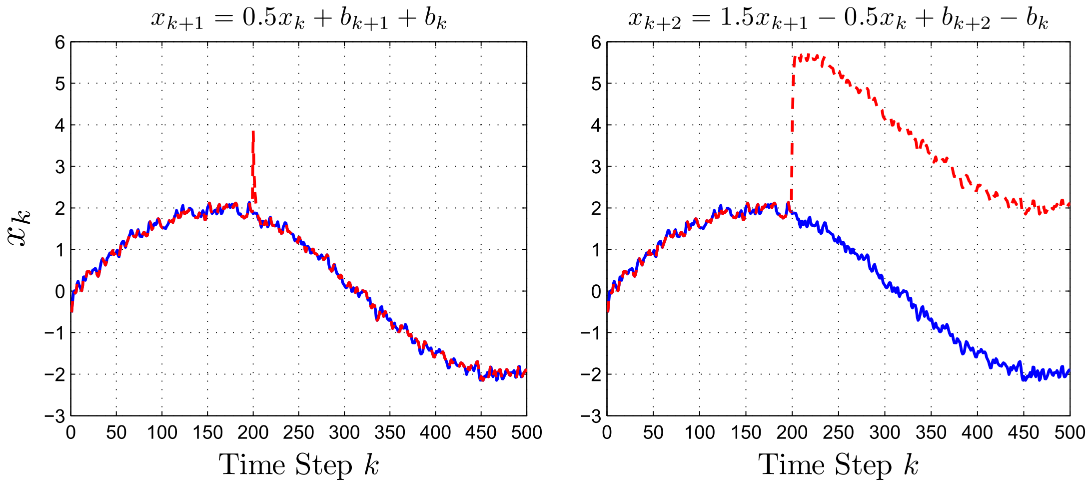

Figure 9. This plot shows x_k given by (36) and (38) with perturbation at step $k = 200$ (dashed line) and without perturbation (solid line). After $k = 200$, for (36) with perturbation, x_k converges to the unperturbed value of x_k, but for (38) with perturbation, x_k does not converge the unperturbed value of x_k.

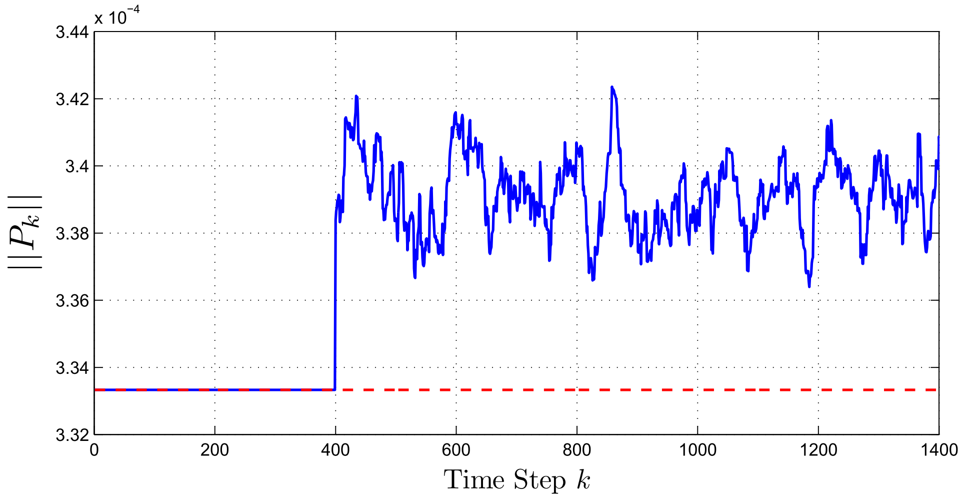

Figure 10. This figure shows $||P_k||$ for SW-VR-RLS with P_k perturbed at $k = 400$ (solid line) and SW-VR-RLS with unperturbed P_k (dashed line). This figure shows that, after P_k is perturbed at $k = 400$, the error between SW-VR-RLS with perturbed P_k and SW-VR-RLS with unperturbed P_k does not decay.

$$\bar{L}_k = \bar{P}_{k-1} - \bar{P}_{k-1}\phi_k \left(S_k + \phi_k^{\mathrm{T}} \bar{P}_{k-1}\phi_k\right)^{-1} \phi_k^{\mathrm{T}} \bar{P}_{k-1},$$
$$\bar{Q}_k = \bar{L}_k - \bar{L}_k\psi_{k-r-1} \left(-I_{n_{k-r-1}} + \psi_{k-r-1}^{\mathrm{T}} \bar{L}_k\psi_{k-r-1}\right)^{-1} \psi_{k-r-1}^{\mathrm{T}} \bar{L}_k,$$
$$\bar{P}_k = \bar{Q}_k - \bar{Q}_k\psi_k \left(I_{n_k} + \psi_k^{\mathrm{T}} \bar{Q}_k\psi_k\right)^{-1} \psi_k^{\mathrm{T}} \bar{Q}_k.$$

For all $k \geqslant k_0$, define $\delta P_k \triangleq \bar{P}_k - P_k$ and note that $\delta P_{k_0} = \Gamma$. We now show numerically that δP_k does not decay to zero. In this paper, we mitigate this by resetting SW-VR-RLS at regular intervals.

We consider the same setup as in Example 6.3, where the input is white except, for all $k \geqslant 0$, $R_k = 3000 I_n$ and $w_k = 0$. We compare SW-VR-RLS with P_{400} perturbed by a positive definite matrix $\Gamma = \delta P_{400}$ and SW-VR-RLS with no perturbation. Figure 10 shows that the error δP_k does not decay.

DOI: 10.1002/acs

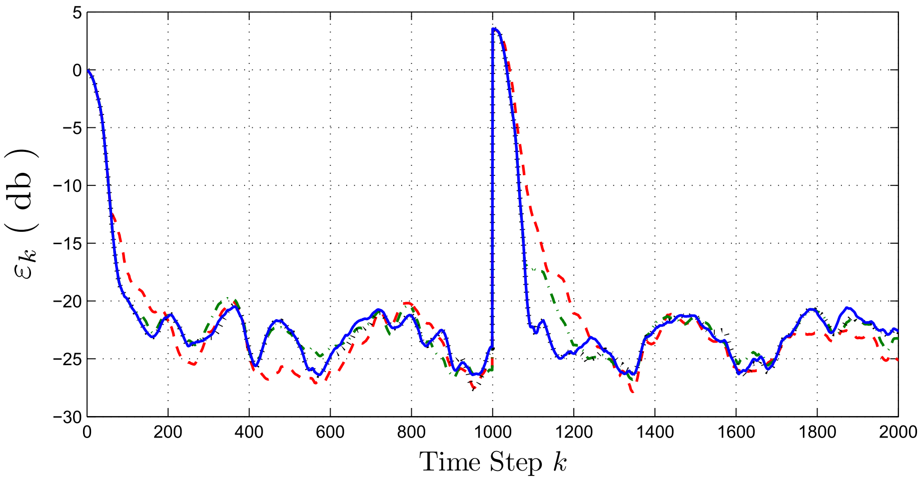

Figure 11. Effect of resetting on SW-VR-RLS for $k_s = 60$ (dashed line), $k_s = 120$ (dash-dotted line), $k_s = 300$ (dotted line), and no resetting (solid line). This plot shows that, after ε_k reaches its asymptotic value and $R_k = R_{\max}$, then ε_k for SW-VR-RLS with covariance resetting does not deviate significantly from SW-VR-RLS without resetting.

We now numerically investigate the effect of resetting Algorithm 2 at regular intervals. The following procedure resets SW-VR-RLS at time step k.

1. x_k is unchanged.
2. For all $i < k$, set $x_i = 0$.
3. Set $\alpha_k = x_k$.
4. For all $i \leqslant k$, set $A_i = 0$ and $b_i = 0$.
5. Set $P_k = R_k^{-1}$.

Note that the resetting procedure is the same for Algorithms 1 and 2 as the Q_k, L_k, and P_k update equations are identical for both algorithms. Furthermore, note that if R_k is a diagonal matrix, then the inverse in step 5 is $\mathcal{O}(n)$. We now investigate the effect of periodically resetting SW-VR-RLS after k_s steps. For this example, we consider the same setup as in Example 6.3, where the input is white. We compare SW-VR-RLS without resetting and SW-VR-RLS with $k_s = 60$, $k_s = 120$ steps, and $k_s = 300$ steps. We show ε_k for a single trial. Figure 11 shows that if ε_k reaches its asymptotic value and $R_k = R_{\max}$, then ε_k for SW-VR-RLS with covariance resetting does not deviate significantly from SW-VR-RLS without resetting. However, resetting SW-VR-RLS when $R_k = R_{\min}$ and ε_k is adapting quickly yields slower convergence of ε_k to its asymptotic value as compared with SW-VR-RLS without resetting. Note that in all cases, resetting SW-VR-RLS does not introduce large transients in ε_k.

8. CONCLUSIONS

A sliding-window variable-regularization recursive-least-squares algorithm has been presented. This algorithm allows for a cost function that has a time-varying regularization term, which provides the ability to vary the weighting in the regularization as well as what is being weighted. The convergence properties of the algorithm in the absence of noise were proved, and the effects of window size and regularization were investigated numerically. Furthermore, SW-VR-RLS was numerically compared with PAPA and PNLMS for white and colored input noises. Numerical examples demonstrated that time-varying regularization can have a positive impact on the convergence properties. Numerical and experimental comparisons to other algorithms, such as those in [16–21], are areas for further investigation. The numerical stability of the algorithm was analyzed analytically and numerically, and it was proved that numerical errors in x_k decay to zero. Furthermore, the numerical errors in P_k were mitigated using resetting, and the effect of resetting on SW-VR-RLS was investigated numerically.

DOI: 10.1002/acs

APPENDIX A: PROOFS OF THEOREMS 1, 2, AND 3

Proof of Theorem 1
To show (i), it follows from the first inequality in (22) that, for all $k \geqslant 0$, $R_k \geqslant c_1 I_n$, where $c_1 \triangleq \varepsilon_1 \inf_{k\geqslant 0} \xi_k > 0$. Because, for all $k \geqslant 0$, A_k is positive semidefinite, it follows from (3) that $P_k^{-1} \geqslant c_1 I_n$, which implies that $0 \leqslant P_k \leqslant \frac{1}{c_1} I_n$. Thus, $\{P_k\}_{k=0}^{\infty}$ is bounded. Similarly, it follows from (6) and (8) that, for all $k \geqslant 1$, $Q_k^{-1} \geqslant c_1 I_n$ and $L_k^{-1} \geqslant c_1 I_n$, which imply that $0 \leqslant Q_k \leqslant \frac{1}{c_1} I_n$ and $0 \leqslant L_k \leqslant \frac{1}{c_1} I_n$. Thus, $\{Q_k\}_{k=1}^{\infty}$ and $\{L_k\}_{k=1}^{\infty}$ are bounded.

To show (ii), note that because $\{b_k\}_{k=0}^{\infty}$ is bounded, it follows that $\kappa_1 \triangleq \sup_k ||b_k|| < \infty$. Additionally, because $\{\alpha_k\}_{k=0}^{\infty}$ is bounded, it follows that $\kappa_2 \triangleq \sup_k ||\alpha_k|| < \infty$. Furthermore, it follows from the last inequality in (22) that, for all $k \geqslant 0$, $R_k \leqslant c_2 I_n$, where $c_2 \triangleq \varepsilon_2 \sup_{k\geqslant 0} \xi_k < \infty$. Hence, it follows from (4) that, for all $k \geqslant 0$,

$$\begin{aligned} ||x_k|| &= \left\| \frac{1}{2} P_k \left(\sum_{i=k-r}^{k} b_i - 2R_k \alpha_k \right) \right\| \\ &\leqslant \frac{1}{2} ||P_k|| \left\| \sum_{i=k-r}^{k} b_i - 2R_k \alpha_k \right\| \\ &\leqslant \frac{1}{2} ||P_k|| \left(\left\| \sum_{i=k-r}^{k} b_i \right\| + 2||R_k|| ||\alpha_k|| \right) \\ &\leqslant \frac{1}{c_1} \left((r+1)\kappa_1 + 2c_2\kappa_2 \right). \end{aligned}$$

Therefore, $\{x_k\}_{k=0}^{\infty}$ is bounded. □

Proof of Theorem 2
To show (i), let $\chi_{\nu-1} = \chi_*$. Then, it follows from (24) and (25) that $\chi_{\nu,2} = \chi_{\nu,3} = \chi_{\nu,\nu} = \cdots = x_*$, and

$$\chi_{\nu,1} = -P_\nu \left(\sum_{i=\nu-r}^{\nu} \frac{1}{2} b_i - R_\nu x_* \right) = -P_\nu \left(-\sum_{i=\nu-r}^{\nu} \frac{1}{2} A_i - R_\nu \right) x_* = x_*,$$

and thus, $\chi_\nu = \chi_*$. Similarly, for $k = \nu$, it follows from (24) and (25) that $\chi_{\nu+1,2} = \chi_{\nu+1,3} = \chi_{\nu+1,\nu} = \cdots = x_*$, and

$$\chi_{\nu+1,1} = -P_{\nu+1} \left(\sum_{i=\nu+1-r}^{\nu+1} \frac{1}{2} b_i - R_{\nu+1} x_* \right) = -P_{\nu+1} \left(-\sum_{i=\nu+1-r}^{\nu+1} \frac{1}{2} A_i - R_{\nu+1} \right) x_* = x_*,$$

and thus, $\chi_{\nu+1} = \chi_*$. It follows that, for all $k > \nu - 2$, $\chi_k = \chi_*$, and thus, $\chi_k \equiv \chi_*$ is an equilibrium solution of (24).

To show (ii), because, for all $k \geqslant \nu$, $\alpha_k = x_{k-\nu}$, it follows from (4) that, for all $k \geqslant \nu$,

$$x_k = -P_k \left(\sum_{i=k-r}^{k} \frac{1}{2} b_i - R_k x_{k-\nu} \right),$$

where, for all $j \in \{0, 1, \ldots, \nu - 1\}$, the initial conditions are $P_j = \left(\sum_{i=j-r}^{j} A_i + R_j \right)^{-1}$ and $x_j = -\frac{1}{2} P_0 \left(\sum_{i=j-r}^{j} b_i - 2R_j \eta \right)$. It follows that

DOI: 10.1002/acs

$$
\begin{aligned}
x_k &= P_k\left(\sum_{i=k-r}^{k} A_i + R_k\right)x_{k-\nu} - P_k\sum_{i=k-r}^{k}\left(A_i x_{k-\nu} + \frac{1}{2}b_i\right)\\
&= x_{k-\nu} - P_k\sum_{i=k-r}^{k}\left(A_i x_{k-\nu} + \frac{1}{2}b_i\right).
\end{aligned} \tag{39}
$$

Define $\tilde{x}_k \triangleq x_k - x_*$. Subtracting x_* from (39), and using (23) and (25), yields, for all $k \geqslant \nu$,

$$
\begin{aligned}
\tilde{x}_k &= \tilde{x}_{k-\nu} - P_k\sum_{i=k-r}^{k} A_i\tilde{x}_{k-\nu}\\
&= \tilde{x}_{k-\nu} - P_k\Phi_k\Phi_k^{\mathrm{T}}\tilde{x}_{k-\nu}\\
&= P_k\left(P_k^{-1} - \Phi_k\Phi_k^{\mathrm{T}}\right)\tilde{x}_{k-\nu}\\
&= P_k R_k\tilde{x}_{k-\nu}\\
&= \left[R_k^{-1} - R_k^{-1}\Phi_k\left(I_{q_k} + \Phi_k^{\mathrm{T}}R_k^{-1}\Phi_k\right)^{-1}\Phi_k^{\mathrm{T}}R_k^{-1}\right]R_k\tilde{x}_{k-\nu}\\
&= \tilde{x}_{k-\nu} - T_k^{-1}\Phi_k\left(\xi_k I_{q_k} + \Phi_k^{\mathrm{T}}T_k^{-1}\Phi_k\right)^{-1}\Phi_k^{\mathrm{T}}\tilde{x}_{k-\nu}\\
&= \tilde{x}_{k-\nu} - T_k^{-1}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}\tilde{x}_{k-\nu},
\end{aligned} \tag{40}
$$

where $\Omega_k \triangleq \xi_k I_{q_k} + \Phi_k^{\mathrm{T}}T_k^{-1}\Phi_k$. Define $\tilde{\chi}_k \triangleq \chi_k - \chi_*$, and, for all $i \in \{1,\ldots,\nu\}$, define $\tilde{\chi}_{k,i} \triangleq \tilde{x}_{k-i+1}$. Then, it follows from (24) and (40) that, for all $k > \nu - 2$,

$$
\begin{bmatrix}\tilde{\chi}_{k+1,1}\\ \tilde{\chi}_{k+1,2}\\ \vdots\\ \tilde{\chi}_{k+1,\nu}\end{bmatrix} = \begin{bmatrix}\left(I - T_{k+1}^{-1}\Phi_{k+1}\Omega_{k+1}^{-1}\Phi_{k+1}^{\mathrm{T}}\right)\tilde{\chi}_{k,\nu}\\ \tilde{\chi}_{k,1}\\ \vdots\\ \tilde{\chi}_{k,\nu-1}\end{bmatrix}. \tag{41}
$$

Note that $\tilde{\chi}_k \equiv 0$ is an equilibrium solution of (41). For all $z \in \mathbb{R}$, define the strictly increasing functions $\alpha(z) \triangleq \varepsilon_1 z^2$ and $\beta(z) \triangleq \varepsilon_2 z^2$, and, for all $k \geqslant \nu - 1$, define the Lyapunov function

$$
V(\tilde{\chi}_k, k) \triangleq \sum_{i=1}^{\nu}\tilde{\chi}_{k,i}^{\mathrm{T}}T_{k+1-i}\tilde{\chi}_{k,i}.
$$

The difference $\Delta V_k \triangleq V(\tilde{\chi}_k, k) - V(\tilde{\chi}_{k-1}, k-1)$ is given by

$$
\begin{aligned}
\Delta V_k &= \tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\left(T_k - T_{k-\nu}\right)\tilde{\chi}_{k-1,\nu} - 2\tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu}\\
&\quad + \tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}T_k^{-1}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu}\\
&\leqslant -2\tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu} + \tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}T_k^{-1}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu}\\
&= -\tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\left(I_{q_k} + I_{q_k} - \Phi_k^{\mathrm{T}}T_k^{-1}\Phi_k\Omega_k^{-1}\right)\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu}\\
&= -\tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\left[I_{q_k} + \left(\Omega_k - \Phi_k^{\mathrm{T}}T_k^{-1}\Phi_k\right)\Omega_k^{-1}\right]\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu}\\
&= -\tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\left(I_{q_k} + \xi_k\Omega_k^{-1}\right)\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu}\\
&\leqslant -\tilde{\chi}_{k-1,\nu}^{\mathrm{T}}\Phi_k\Omega_k^{-1}\Phi_k^{\mathrm{T}}\tilde{\chi}_{k-1,\nu}.
\end{aligned} \tag{42}
$$

Because, for all $k \geqslant \nu - 1$ and $\tilde{\chi}_k \in \mathbb{R}^{n\nu}$, $\alpha(||\tilde{\chi}_k||) \leqslant V(\tilde{\chi}_k, k) \leqslant \beta(||\tilde{\chi}_k||)$ and $\Delta V_k \leqslant 0$, it follows from [23, Theorem 13.11] that the equilibrium solution $\tilde{\chi}_k \equiv 0$ of (41) is uniformly Lyapunov stable. Furthermore, because $\alpha(z) \to \infty$ as $z \to \infty$, it follows from [23, Corollary 13.4] that, for each $\tilde{\chi}_{\nu-1} \in \mathbb{R}^{n\nu}$, the sequence $\{\tilde{\chi}_k\}_{k=\nu-1}^{\infty}$ is bounded. Hence, for each $x_0 \in \mathbb{R}^n$, $\{\tilde{x}_k\}_{k=0}^{\infty}$ is bounded, and thus, $\{x_k\}_{k=0}^{\infty}$ is bounded.

To show (iii), it follows from (42) that

$$0 \leqslant \sum_{j=\nu}^{k} \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu} \leqslant -\sum_{j=\nu}^{k} \Delta V_j = V(\tilde{\chi}_{\nu-1}, \nu-1) - V(\tilde{\chi}_k, k) \leqslant V(\tilde{\chi}_{\nu-1}, \nu-1).$$

Hence, the nondecreasing sequence $\left\{\sum_{j=\nu}^{k} \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu}\right\}_{k=\nu}^{\infty}$ is bounded, and thus, $\sum_{j=\nu}^{\infty} \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu}$ exists.

Next, for all $k \geqslant \nu$, define $\mathcal{M}_k \triangleq \sum_{j=\nu}^{k} ||x_j - x_{j-\nu}||^2$, and it follows from (40) that

$$\mathcal{M}_k = \sum_{j=\nu}^{k} ||T_j^{-1} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu}||^2 \leqslant \sum_{j=\nu}^{k} \|T_j^{-1}\| \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} T_j^{-1} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu}.$$

Note that, for all $k \geqslant \nu$, $\|T_k^{-1}\| \leqslant \|\frac{1}{\varepsilon_1} I_n\| = \frac{1}{\varepsilon_1}$. Therefore,

$$\begin{aligned}
\mathcal{M}_k &\leqslant \frac{1}{\varepsilon_1} \sum_{j=\nu}^{k} \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-1} \left(\xi_j I_{q_j} + \Phi_j^{\mathrm{T}} T_j^{-1} \Phi_j - \xi_j I_{q_j}\right) \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu} \\
&= \frac{1}{\varepsilon_1} \sum_{j=\nu}^{k} \tilde{x}_{j-\nu}^{T} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu} - \frac{1}{\varepsilon_1} \sum_{j=\nu}^{k} \xi_j \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-2} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu} \\
&\leqslant \frac{1}{\varepsilon_1} \sum_{j=\nu}^{k} \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu}.
\end{aligned}$$

Because $\sum_{j=\nu}^{\infty} \tilde{x}_{j-\nu}^{\mathrm{T}} \Phi_j \Omega_j^{-1} \Phi_j^{\mathrm{T}} \tilde{x}_{j-\nu}$ exists, it follows that the nondecreasing sequence $\{\mathcal{M}_k\}_{k=\nu}^{\infty}$ is bounded, and thus, $\lim_{k\to\infty} \mathcal{M}_k$ exists, which verifies (iii).

To show (iv), because $\{A_k\}_{k=0}^{\infty}$ is bounded, it follows that $\{\Phi_k\}_{k=0}^{\infty}$ is bounded. Because, in addition, $\{\xi_k\}_{k=0}^{\infty}$ and $\{T_k^{-1}\}_{k=0}^{\infty}$ are bounded, it follows that there exists $c_3 > 0$ such that, for all $k \geqslant 0$, $c_3 I_{q_k} \leqslant \sigma_{\min}\left(\Omega_k^{-1/2}\right) I_{q_k} \leqslant \Omega_k^{-1/2}$, which implies that

$$0 \leqslant c_3 ||\Phi_k^{\mathrm{T}} \tilde{x}_{k-\nu}|| \leqslant \sigma_{\min}\left(\Omega_k^{-1/2}\right) ||\Phi_k^{\mathrm{T}} \tilde{x}_{k-\nu}|| \leqslant ||\Omega_k^{-1/2} \Phi_k^{\mathrm{T}} \tilde{x}_{k-\nu}||.$$

Therefore, because (iii) implies that $\lim_{k\to\infty} \Omega_k^{-1/2} \Phi_k^{\mathrm{T}} \tilde{x}_{k-\nu} = 0$, it follows that $\lim_{k\to\infty} \Phi_k^{\mathrm{T}} \tilde{x}_{k-\nu} = 0$, which implies that $\lim_{k\to\infty} \psi_k^{\mathrm{T}} \tilde{x}_{k-\nu} = 0$.

Next, because $\{A_k\}_{k=0}^{\infty}$ is bounded, it follows that $\kappa \triangleq \sup_{k\geqslant 0} \sigma_{\max}(\psi_k) < \infty$. Thus,

$$\begin{aligned}
||A_k x_k + \frac{1}{2} b_k|| &= ||A_k x_k - A_k x_*|| \\
&= ||\psi_k \psi_k^{\mathrm{T}} \tilde{x}_k|| \\
&\leqslant \kappa ||\psi_k^{\mathrm{T}} \tilde{x}_k|| \\
&= \kappa ||\psi_k^{\mathrm{T}} \tilde{x}_{k-\nu} + \psi_k^{\mathrm{T}} \tilde{x}_k - \psi_k^{\mathrm{T}} \tilde{x}_{k-\nu}|| \\
&\leqslant \kappa \left(||\psi_k^{\mathrm{T}} \tilde{x}_{k-\nu}|| + ||\psi_k|| ||\tilde{x}_k - \tilde{x}_{k-\nu}||\right) \\
&\leqslant \kappa \left(||\psi_k^{\mathrm{T}} \tilde{x}_{k-\nu}|| + \kappa ||\tilde{x}_k - \tilde{x}_{k-\nu}||\right) \\
&= \kappa ||\psi_k^{\mathrm{T}} \tilde{x}_{k-\nu}|| + \kappa^2 ||\tilde{x}_k - \tilde{x}_{k-\nu}||.
\end{aligned} \tag{43}$$

Because $\lim_{k\to\infty} \psi_k^{\mathrm{T}} \tilde{x}_{k-\nu} = 0$, and (iii) implies that $\lim_{k\to\infty}(\tilde{x}_k - \tilde{x}_{k-\nu}) = 0$, it follows from (43) that $\lim_{k\to\infty}(A_k x_k + \frac{1}{2} b_k) = 0$, which confirms (iv).

To show (v), it follows from (43) that $\lim_{k\to\infty} A_k \tilde{x}_k = 0$ and $\lim_{k\to\infty} \psi_k^{\mathrm{T}} \tilde{x}_k = 0$. Next, using arguments similar to those used in (43), we obtain

$$\begin{aligned} ||A_{k-\nu}\tilde{x}_k|| &\leqslant \kappa||\psi_{k-\nu}^{\mathrm{T}}\tilde{x}_k|| \\ &= \kappa||\psi_{k-\nu}^{\mathrm{T}}\tilde{x}_{k-\nu} + \psi_{k-\nu}^{\mathrm{T}}\tilde{x}_k - \psi_{k-\nu}^{\mathrm{T}}\tilde{x}_{k-\nu}|| \\ &\leqslant \kappa||\psi_{k-\nu}^{\mathrm{T}}\tilde{x}_{k-\nu}|| + \kappa^2||\tilde{x}_k - \tilde{x}_{k-\nu}||. \end{aligned} \tag{44}$$

Because $\lim_{k\to\infty}\psi_k^{\mathrm{T}}\tilde{x}_k = 0$, and (iii) implies that $\lim_{k\to\infty}(\tilde{x}_k - \tilde{x}_{k-\nu}) = 0$, it follows from (44) that $\lim_{k\to\infty} A_{k-\nu}\tilde{x}_k = 0$ and $\lim_{k\to\infty}\psi_{k-\nu}^{\mathrm{T}}\tilde{x}_k = 0$. Again, using arguments similar to those used in (43), we obtain

$$\begin{aligned} ||A_{k-2\nu}\tilde{x}_k|| &\leqslant \kappa||\psi_{k-2\nu}^{\mathrm{T}}\tilde{x}_k|| \\ &= \kappa||\psi_{k-2\nu}^{\mathrm{T}}\tilde{x}_{k-\nu} + \psi_{k-2\nu}^{\mathrm{T}}\tilde{x}_k - \psi_{k-2\nu}^{\mathrm{T}}\tilde{x}_{k-\nu}|| \\ &\leqslant \kappa||\psi_{k-2\nu}^{\mathrm{T}}\tilde{x}_{k-\nu}|| + \kappa^2||\tilde{x}_k - \tilde{x}_{k-\nu}||. \end{aligned} \tag{45}$$

Because $\lim_{k\to\infty}\psi_{k-\nu}^{\mathrm{T}}\tilde{x}_k = 0$, and (iii) implies that $\lim_{k\to\infty}(\tilde{x}_k - \tilde{x}_{k-\nu}) = 0$, it follows from (45) that $\lim_{k\to\infty} A_{k-2\nu}\tilde{x}_k = 0$ and $\lim_{k\to\infty}\psi_{k-2\nu}^{\mathrm{T}}\tilde{x}_k = 0$. Repeating this argument shows that, for all $i \in \{0, 1, 2, \ldots, l\}$, $\lim_{k\to\infty} A_{k-\nu i}\tilde{x}_k = 0$. Because, for all $k \geqslant \nu l - r$, $cI_n \leqslant \sum_{i=0}^{l} A_{k-\nu i}$, it follows that

$$||\tilde{x}_k|| \leqslant \frac{1}{c}\left|\left|\sum_{i=0}^{l} A_{k-\nu i}\tilde{x}_k\right|\right| \leqslant \frac{1}{c}\sum_{i=0}^{l}||A_{k-\nu i}\tilde{x}_k||,$$

which implies that $\lim_{k\to\infty}\tilde{x}_k = 0$. Thus, $\lim_{k\to\infty}\chi_k = \chi_*$, and the equilibrium solution $\chi_k \equiv \chi_*$ of (24) is globally asymptotically stable. □

Proof of Theorem 3

To show (i), note that the update equations for L_k, Q_k, and P_k are identical to those in SW-VR-RLS. Thus, (i) follows directly from Theorem 1.

To show (ii) and (iii), it follows from (32) and (23) that, for all $k \geqslant k_0 + \nu$,

$$\begin{aligned} \delta_k &= P_k R_k \delta_{k-\nu} \\ &= \left[R_k^{-1} - R_k^{-1}\Phi_k\left(I_{q_k} + \Phi_k^{\mathrm{T}} R_k^{-1}\Phi_k\right)^{-1}\Phi_k^{\mathrm{T}} R_k^{-1}\right] R_k\delta_{k-\nu} \\ &= \delta_{k-\nu} - T_k^{-1}\Phi_k\left(\xi_k I_{q_k} + \Phi_k^{\mathrm{T}} T_k^{-1}\Phi_k\right)^{-1}\Phi_k^{\mathrm{T}}\delta_{k-\nu}. \end{aligned}$$

The remainder of the proof of (ii) and (iii) is analogous to the proof of Theorem 2 from (40) onwards with x_k replaced by δ_k, x_* replaced by 0, $\tilde{x}_k$ replaced by δ_k, and $\tilde{\chi}_k$ replaced by Δ_k. □

REFERENCES

1. Ljung L, Söderström T. *Theory and Practice of Recursive Identification*. The MIT Press: Cambridge, MA, 1983.
2. Goodwin GC, Sin KS. *Adaptive Filtering, Prediction, and Control*. Prentice Hall: Englewood Cliffs, NJ, 1984.
3. Söderström T, Stoica P. *System Identification*. Prentice–Hall: Upper Saddle River, NJ, 1989.
4. Juang JN. *Applied System Identification*. Prentice-Hall: Upper Saddle River, NJ, 1993.
5. Åström KJ, Wittenmark B. *Adaptive Control* (2nd edn). Addison-Wesley: MA, 1995.
6. Ioannou P, Sun J. *Robust Adaptive Control*. Prentice Hall: Upper Saddle River, NJ, 1996.
7. Ljung L. *System Identification: Theory for the User* (2nd edn). Prentice-Hall Information and Systems Sciences: Englewood Cliffs, NJ, 1999.
8. Tao G. *Adaptive Control Design and Analysis*. Wiley: Hoboken, NJ, 2003.
9. Sayed AH. *Adaptive Filters*. John Wiley and Sons, Inc.: Hoboken, NJ, 2008.
10. Gustafsson F. *Adaptive Filtering and Change Detection*. Wiley: Hoboken, NJ, 2000.
11. Jiang J, Zhang Y. A novel variable-length sliding window blockwise least-squares algorithm for on-line estimation of time-varying parameters. *International Journal of Adaptive Control and Signal Processing* 2004; **18**:505–521.
12. Belge M, Miller EL. A sliding window RLS-like adaptive algorithm for filtering alpha-stable noise. *IEEE Signal Processing Letters* 2000; **7**(4):86–89.
13. Choi BY, Bien Z. Sliding-windowed weighted recursive least-squares method for parameter estimation. *Electronics Letters* 1989; **25**(20):1381–1382.

DOI: 10.1002/acs

14. Liu H, He Z. A sliding-exponential window RLS adaptive algorithm: properties and applications. *Signal Processing* 1995; **45**(3):357–368.
15. van Waterschoot T, Rombouts G, Moonen M. Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement. *Signal Processing* 2008; **88**:594–611.
16. Houacine A, Demoment G. Chandrasekhar adaptive regularizer for adaptive filtering. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, 1986; 2967–2970.
17. Gay SL. Dynamically regularized fast RLS with application to echo cancellation. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, 1996; 957–960.
18. Choi YS, Shin HC, Song WJ. Affine projection algorithms with adaptive regularization matrix. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, 2006; 201–204.
19. Stokes JW, Platt JC. Robust RLS with round robin regularization including application to stereo acoustic echo cancellation. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, 2006; 736–739.
20. Chen K, Lu J, Xu B. A method to adjust regularization parameter of fast affine projection algorithm. *Proceedings of 8th IEEE International Conference on Signal Processing*, Beijing, China, 2006.
21. Challa D, Grant SL, Mohammad A. Variable regularized fast affine projections. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, HI, 2007; 89–92.
22. Hoagg JB, Ali A, Mossberg M, Bernstein DS. Sliding window recursive quadratic optimization with variable regularization. *Proceedings of the American Control Conference*, San Francisco, CA, 2011; 3275–3280.
23. Haddad WM, Chellaboina V. *Nonlinear Dynamical Systems and Control: A Lyapunov-based Approach*. Princeton University Press: Princeton, New Jersey, 2008.
24. Gansler T, Benesty J, Gay SL, Sondhi MM. A robust proportionate affine projection algorithm for network echo cancellation. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, 2000; 793–796.
25. Duttweiler DL. Proportionate normalized least-mean-squares adaptation in echo cancelers. *IEEE Transactions on Speech and Audio Processing* 2000; **8**(5):508–518.

DOI: 10.1002/acs