

# Convergence and Consistency of Recursive Least Squares with Variable-Rate Forgetting

Adam L. Bruce<sup>a</sup>, Ankit Goel<sup>a</sup>, and Dennis S. Bernstein<sup>a</sup>

<sup>a</sup>*Department of Aerospace Engineering, The University of Michigan, Ann Arbor, Michigan 48109, United States*

---

## Abstract

A recursive least squares algorithm with variable rate forgetting (VRF) is derived by minimizing a quadratic cost function. Under persistent excitation and boundedness of the forgetting factor, the minimizer given by VRF is shown to converge to the true parameters. In addition, under persistent excitation and with noisy measurements, where the noise is uncorrelated with the regressor, conditions are given under which the minimizer given by VRF is a consistent estimator of the true parameters. The results are illustrated by a numerical example involving abruptly changing parameters.

*Key words:* system identification, identification methods, tracking and adaptation, adaptive systems, adaptive control

---

## 1 Introduction

Recursive least squares (RLS) is one of the foundational algorithms of systems and control theory, especially for signal processing, identification, and adaptive control [1, 2]. An early exposition of RLS is given in [3].

Standard RLS employs a constant forgetting factor  $\lambda$ , which enhances the importance of recent data over older data. Although  $\lambda$  can be set by the user, the performance of RLS is often extremely sensitive to the chosen value. Consequently, choosing a suitable value of  $\lambda$  is typically a trial and error process.

To remedy this problem, various techniques have been proposed to automatically vary the forgetting factor in response to the fit error. In particular, [4] reports a method for sequentially updating the forgetting factor to conserve the amount information used in the estimate, and [5] reports an update-based algorithm that uses noise statistics to control the forgetting factor. [6] gives a gradient-based algorithm for computing a forgetting factor that locally minimizes the mean-square error of the estimate, and [7] derives a Newton-type gradient-descent algorithm that combines sequential estimation with minimization of the mean-squared error. Finally, [8] gives a formula based on exponentiation of the squared residual.

---

*Email addresses:* [admb Bruce@umich.edu](mailto:admb Bruce@umich.edu) (Adam L. Bruce), [ankgoel@umich.edu](mailto:ankgoel@umich.edu) (Ankit Goel), [dsbaero@umich.edu](mailto:dsbaero@umich.edu) (Dennis S. Bernstein).

The present paper approaches the problem of varying the forgetting factor by deriving a generalization of RLS that includes time-dependent cost scaling and regularization. This formulation involves a growing-window cost function, and thus is distinct from the formulation of [9], which uses a sliding-window cost function. The growing-window cost function is advantageous since it directly generalizes traditional RLS and has the ability to weigh recent data more heavily than older data.

The first contribution of the paper is given by Theorem 1, which introduces RLS with variable-rate forgetting (VRF), a novel extension of RLS in which the role of the constant forgetting factor  $\lambda$  in RLS is replaced by a variable forgetting factor  $\beta_k$ . By setting  $\beta_k = \frac{1}{\lambda}$  for all  $k$ , VRF specializes to RLS with constant-rate-forgetting (CRF). The variable-rate-forgetting extensions of RLS given in [4–8] are special cases of Theorem 1 with specific choices of  $\beta_k$ . In addition, Theorem 1 refines the variable-rate weighting used in [1, pp. 17, 18]. In particular, we factor  $\alpha_k$  in [1, Eq. (2.12)] as  $\beta_k \cdots \beta_0$ , where  $1/\beta_k$  serves as the instantaneous forgetting factor at step  $k$ . This formulation allows the user to specify  $\beta_k$  at each step based on the current residual or knowledge of system changes. The second and third contributions of this paper are given by Theorems 2, 4, and Corollary 4, which prove conditions on  $\beta_k$  ensuring convergence under the assumption of persistency (Theorem 2) and consistency under the assumption of persistency and that the regressor and sensor noise are uncorrelated (Theorem 4, Corollary 4). Specific examples of  $\beta_k$  for consistent and non-consistent algorithms are given in Corollary 5. The

fourth contribution is two choices of  $\beta_k$  that may be useful in practice. In Section 6, we demonstrate these choices on an abruptly changing system with and without measurement noise and compare the performance of VRF and CRF for the given example.

The notation used throughout this paper is as follows. The symbols  $\mathbb{S}^n$ ,  $\mathbb{N}^n$ , and  $\mathbb{P}^n$  denote the sets of real  $n \times n$  symmetric, positive-semidefinite, and positive-definite matrices, respectively. For all  $A \in \mathbb{S}^n$ ,  $\lambda_i(A)$  denotes the  $i$ th largest eigenvalue of  $A$ ,  $\lambda_{\max}(A) \triangleq \lambda_1(A)$ , and  $\lambda_{\min}(A) \triangleq \lambda_n(A)$ .  $\lfloor x \rfloor$  denotes the greatest integer less than or equal to  $x \in \mathbb{R}$ . If  $X$  is a set, then the notation  $(x_k)_{k \geq 0} \subset X$  indicates that  $(x_k)_{k \geq 0}$  is a sequence in  $X$ . If  $(a_k)_{k \geq 0}, (b_k)_{k \geq 0} \subset \mathbb{R}$ , then the notation  $a_k \sim \mathcal{O}(b_k)$  indicates that there exists  $M > 0$  and  $K \geq 0$  such that, for all  $k \geq K$ ,  $a_k \leq Mb_k$ . Finally, for all  $k \geq 0$  and  $N \geq 0$ , we define  $\xi(k, N) \triangleq \lfloor \frac{k}{N+1} \rfloor$ .

## 2 Problem Formulation

Let  $\lambda \in (0, 1]$ ,  $\theta_0 \in \mathbb{R}^n$ , and  $P_0 \in \mathbb{P}^n$ . Furthermore, for all  $k \geq 0$ , let  $\phi_k \in \mathbb{R}^{p \times n}$ ,  $y_k \in \mathbb{R}^p$ ,  $e_k \triangleq y_k - \phi_k \theta$  and define  $J_k: \mathbb{R}^n \rightarrow [0, \infty)$  by

$$J_k(\theta) \triangleq \sum_{i=0}^k \lambda^{k-i} \|e_k\|^2 + \lambda^{k+1} (\theta - \theta_0)^T P_0^{-1} (\theta - \theta_0). \quad (1)$$

Equation (1) is the cost function for CRF, the minimization of which produces the least squares estimate of  $\theta$  given  $y_0, \dots, y_k$ . Since  $J_k$  is quadratic and strictly convex, it follows that its unique global minimizer,  $\theta_{k+1} \triangleq \operatorname{argmin}_{\theta \in \mathbb{R}^n} J_k(\theta)$ , is the only local minimizer. The following proposition gives the traditional RLS update equations for computing  $\theta_{k+1}$  [1, 2, 10].

**Proposition 1** *Under the notation and assumptions of the preceding paragraph, for all  $k \geq 0$ , define  $J_k: \mathbb{R}^n \rightarrow [0, \infty)$  by (1). Then*

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^T (y_k - \phi_k \theta_k), \quad (2)$$

where

$$P_{k+1} = \frac{1}{\lambda} P_k - \frac{1}{\lambda} P_k \phi_k^T (\lambda I_p + \phi_k P_k \phi_k^T)^{-1} \phi_k P_k. \quad (3)$$

In this paper, we introduce a generalization of (1) in which the forgetting factor is variable, prove a result analogous to Proposition 1 for the generalization, and analyze convergence and consistency for the family of algorithms thus obtained. To generalize (1), for all  $k \geq 0$ , let  $\beta_k > 0$ , define

$$\rho_k \triangleq \prod_{i=0}^k \beta_i, \quad \rho_{-1} \triangleq 1, \quad (4)$$

and define the cost function  $J_k: \mathbb{R}^n \rightarrow [0, \infty)$  by

$$J_k(\theta) \triangleq \sum_{i=0}^k \frac{\rho_i}{\rho_k} \|e_k\|^2 + \frac{1}{\rho_k} (\theta - \theta_0)^T P_0^{-1} (\theta - \theta_0). \quad (5)$$

Since (5) is quadratic and strictly convex, like (1), its unique global minimizer is the only local minimizer. Theorem 1 provides recursive update equations for this minimizer.

## 3 RLS with Variable-Rate Forgetting

Note that (5) can be written as

$$J_k(\theta) = \theta^T A_k \theta - 2b_k^T \theta + c_k, \quad (6)$$

where

$$A_k \triangleq \sum_{i=0}^k \frac{\rho_i}{\rho_k} \phi_i^T \phi_i + \frac{1}{\rho_k} P_0^{-1}, \quad (7)$$

$$b_k \triangleq \sum_{i=0}^k \frac{\rho_i}{\rho_k} \phi_i^T y_i + \frac{1}{\rho_k} P_0^{-1} \theta_0, \quad (8)$$

$$c_k \triangleq \sum_{i=0}^k \frac{\rho_i}{\rho_k} y_i^T y_i + \frac{1}{\rho_k} \theta_0^T P_0^{-1} \theta_0. \quad (9)$$

Since  $A_k$  is positive definite, we define the positive-definite matrix

$$P_k \triangleq A_k^{-1}, \quad (10)$$

where  $A_{-1} \triangleq P_0^{-1}$ .

The following result, *RLS with variable-rate forgetting (VRF)*, generalizes Proposition 1 to the minimizer of (5).

**Theorem 1** *Let  $\theta_0 \in \mathbb{R}^n$ ,  $P_0 \in \mathbb{P}^n$ , and, for all  $k \geq 0$ , let  $\phi_k \in \mathbb{R}^{p \times n}$ ,  $y_k \in \mathbb{R}^p$ , and  $\beta_k \in (0, \infty)$ . Then the minimizer  $\theta_{k+1}$  of (5) is given by*

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^T (y_k - \phi_k \theta_k), \quad (11)$$

and

$$P_{k+1} = L_k - L_k \phi_k^T (I_p + \phi_k L_k \phi_k^T)^{-1} \phi_k L_k, \quad (12)$$

$$L_k \triangleq \beta_k P_k. \quad (13)$$

The proof of Theorem 1 requires the following lemma.

**Lemma 1** *Let  $P_0 \in \mathbb{P}^n$  and, for all  $k \geq 0$ , let  $\beta_k > 0$ , define  $\rho_k$  by (4), and define  $P_k$  by (10). Then, for all  $k \geq 0$ ,*

$$P_{k+1}^{-1} = \frac{1}{\beta_k} P_k^{-1} + \phi_k^T \phi_k \quad (14)$$

$$= \frac{1}{\rho_k} \left( P_0^{-1} + \sum_{i=0}^k \rho_i \phi_i^T \phi_i \right). \quad (15)$$

**Proof.** Let  $k \geq 0$ . It follows from (7) that  $A_k = \frac{1}{\beta_k} A_{k-1} + \phi_k^T \phi_k$ , which, using (10), implies (14). Furthermore, (14) implies  $P_1^{-1} = \frac{1}{\rho_0} (P_0^{-1} + \rho_0 \phi_0^T \phi_0)$ , which confirms (15) for  $k = 0$ . Next, let  $k > 0$  and suppose for induction that (15) holds for  $k - 1$ . From (14) it follows that  $P_{k+1}^{-1} = \frac{1}{\beta_k} P_k^{-1} + \phi_k^T \phi_k = \frac{1}{\rho_k} (P_0^{-1} + \sum_{i=0}^{k-1} \rho_i \phi_i^T \phi_i) + \frac{\rho_k}{\rho_k} \phi_k^T \phi_k = \frac{1}{\rho_k} (P_0^{-1} + \sum_{i=0}^k \rho_i \phi_i^T \phi_i)$ .  $\square$

**Proof of Theorem 1.** Let  $k \geq 0$ . To prove (12), note that it follows from (13), (14), and the matrix inversion lemma that  $P_{k+1} = \left( \frac{1}{\beta_k} P_k^{-1} + \phi_k^T \phi_k \right)^{-1} = L_k - L_k \phi_k^T (I_p + \phi_k L_k \phi_k^T)^{-1} \phi_k L_k$ . To prove (11), note that (8), (10), and (14) imply that

$$\begin{aligned} \theta_{k+1} &= P_{k+1} \left( \sum_{i=0}^k \frac{\rho_i}{\rho_k} \phi_i^T y_i + \frac{1}{\rho_k} P_0^{-1} \theta_0 \right) \\ &= P_{k+1} \left( \phi_k^T y_k + \frac{\rho_{k-1}}{\rho_k} \left[ \sum_{i=0}^{k-1} \frac{\rho_i}{\rho_{k-1}} \phi_i^T y_i + \frac{1}{\rho_{k-1}} P_0^{-1} \theta_0 \right] \right) \\ &= P_{k+1} \left( \phi_k^T \phi_k + \frac{1}{\beta_k} P_k^{-1} \right) \theta_k + P_{k+1} \phi_k^T (y_k - \phi_k \theta_k) \\ &= \theta_k + P_{k+1} \phi_k^T (y_k - \phi_k \theta_k), \end{aligned} \quad \square$$

For all  $k \geq 0$ , let  $\beta_k = \frac{1}{\lambda}$ . Then (5) specializes to (1), and (11)–(13) specialize to (3) and (2). Theorem 1 thus includes Proposition 1 as a special case.

## 4 Convergence of VRF

### 4.1 Asymptotic Convergence

**Definition 1** A sequence  $(S_k)_{k \geq 0} \subset \mathbb{N}^n$  is persistent if there exist  $N \geq 1$  and  $\alpha > 0$  such that, for all  $j \geq 0$ ,

$$\alpha I_n \leq \sum_{i=0}^N S_{i+j}. \quad (16)$$

The numbers  $\alpha$  and  $N$  are, respectively, the lower bound and persistency window of  $(S_k)_{k \geq 0}$ . The sequence  $(\phi_k)_{k \geq 0} \subset \mathbb{R}^{n \times m}$  is persistent if  $(\phi_k^T \phi_k)_{k \geq 0}$  is persistent.

**Theorem 2** Let  $(\phi_k)_{k \geq 0} \subset \mathbb{R}^{n \times m}$ , be persistent, let  $\theta \in \mathbb{R}^n$ , and, for all  $k \geq 0$ , let  $y_k = \phi_k \theta$ . Furthermore, let  $a > 1$  and, for all  $k \geq 0$ , let  $\beta_k \geq 1$ . Finally, let  $\theta_0 \in \mathbb{R}^n$ , let  $P_0 \in \mathbb{P}^n$ , and, for all  $k \geq 0$ , define  $\theta_{k+1}$  by (11)–(13). Then  $\lim_{k \rightarrow \infty} \theta_k = \theta$ .

**Proof.** Let  $k \geq 0$  and define  $\tilde{\theta}_k \triangleq \theta_k - \theta$ . Using (11) and (14) it follows that  $\tilde{\theta}_{k+1} = (I_n - P_{k+1} \phi_k^T \phi_k) \tilde{\theta}_k = \frac{1}{\beta_k} P_{k+1} P_k^{-1} \tilde{\theta}_k$ , thus  $\tilde{\theta}_k = \frac{1}{\rho_{k-1}} P_k P_0^{-1} \tilde{\theta}_0$ . From (15), it fol-

lows that

$$\begin{aligned} \overline{\lim}_{k \rightarrow \infty} \|\tilde{\theta}_k\|^2 &\leq \overline{\lim}_{k \rightarrow \infty} \frac{\lambda_{\max}(P_k^2)}{\rho_{k-1}^2} \|P_0^{-1} \tilde{\theta}_0\|^2 \\ &\leq \overline{\lim}_{k \rightarrow \infty} \frac{\|P_0^{-1} \tilde{\theta}_0\|^2}{\lambda_{\max}^2 (P_0^{-1} + \sum_{i=0}^{k-1} \rho_i \phi_i^T \phi_i)} \\ &\leq \overline{\lim}_{k \rightarrow \infty} \frac{\|P_0^{-1} \tilde{\theta}_0\|^2}{[\lambda_{\max}(P_0^{-1}) + \xi(k, N+1)\alpha]^2} = 0. \quad \square \end{aligned}$$

### 4.2 Convergence Rate

**Definition 2** Let  $(S_i)_{i \geq 0} \subset \mathbb{N}^n$  be persistent with lower bound  $\alpha$  and window  $N$ . Then the upper bound  $\beta \in (0, \infty) \cup \{\infty\}$  of  $(S_i)_{i \geq 0}$  is

$$\beta \triangleq \sup_{j \geq 0} \lambda_{\max} \left( \sum_{i=j}^N S_j \right). \quad (17)$$

**Lemma 2** Let  $(S_i)_{i \geq 0} \subset \mathbb{N}^n$  be persistent with window  $N$ , lower bound  $\alpha$ , and upper bound  $\beta$ , and let  $(a_i)_{i \geq 0}$  be a nondecreasing sequence of nonnegative numbers. Then, for all  $k \geq 0$ ,

$$\alpha \ell_{\xi(k, N)-1} I_n \leq \sum_{i=0}^k a_i S_i \leq \beta r_{\xi(k, N)} I_n, \quad (18)$$

where  $\ell_j \triangleq \sum_{i=0}^j a_{i(N+1)}$  and  $r_j \triangleq \sum_{i=0}^j a_{i(N+1)+N}$ .

**Proof.** In the case where  $\beta = \infty$ , the upper bound of (18) is immediate. Hence, assume  $\beta < \infty$ . Let  $k \geq 0$ . Since  $(a_i)_{i \geq 0}$  is nondecreasing, for all  $j \geq 0$  and  $i \in \{0, \dots, N\}$ ,  $a_{i+j} \leq a_{N+j}$  and  $a_j \leq a_{i+j}$ . From (16) and (17) it follows that  $\alpha a_j I_n \leq a_j \sum_{i=0}^N S_{i+j} \leq \sum_{i=0}^N a_{i+j} S_{i+j}$ , thus

$$\alpha \ell_{\xi(k, N)-1} I_n \leq \sum_{q=0}^{\xi(k, N)-1} \sum_{i=0}^N a_{i+q(N+1)} S_{i+q(N+1)} \leq \sum_{i=0}^k a_i S_i.$$

Similarly,  $\sum_{i=0}^N a_{i+j} S_{i+j} \leq a_{N+j} \sum_{i=0}^N S_{i+j} \leq a_{N+j} \beta I_n$ , and thus

$$\sum_{i=0}^k a_i S_i \leq \sum_{q=0}^{\xi(k, N)-1} a_{q(N+1)+N} \beta I_n + a_k \beta I_n \leq \beta r_{\xi(k, N)} I_n. \quad \square$$

**Theorem 3** Under the assumptions and notation of Theorem 2,  $\|\tilde{\theta}_k\| \sim \mathcal{O}(1/\sum_{i=0}^{\xi(k, N)-1} \rho_{i(N+1)})$ .

**Proof.** Let  $M = \|P_0^{-1} \tilde{\theta}_0\|/\alpha$ . From Lemma 2, it follows that, for all  $k \geq 0$ ,  $\lambda_{\min}(P_0^{-1}) + \alpha \sum_{i=0}^{\xi(k, N)-1} \rho_{i(N+1)} \leq$

$\lambda_{\min}(P_0^{-1} + \sum_{i=0}^k \rho_i \phi_k^T \phi_k)$ , and therefore, for all  $k \geq 0$ ,

$$\begin{aligned} \|\tilde{\theta}_k\| &\leq \|P_k\| \|P_0^{-1} \tilde{\theta}_0\| \leq \frac{\|P_0^{-1} \tilde{\theta}_0\|}{\lambda_{\min} \|P_k^{-1}\|} \\ &= \frac{\|P_0^{-1} \tilde{\theta}_0\|}{\lambda_{\min}(P_0^{-1})} \left(1 + \frac{\alpha}{\lambda_{\min}(P_0^{-1})} \sum_{i=0}^{\xi(k,N)-1} \rho_{i(N+1)}\right)^{-1} \\ &\leq M \left(\sum_{i=0}^{\xi(k,N)-1} \rho_{i(N+1)}\right)^{-1}. \quad \square \end{aligned}$$

The following corollary shows that Theorem 3 can be used to prove convergence rates for RLS without forgetting as well as with CRF.

**Corollary 3** *Under the assumptions and notation of Theorem 2, assume that there exists  $\gamma \in [1, \infty)$  such that, for all  $k \geq 0$ ,  $\beta_k = \gamma$ . Then*

$$\|\tilde{\theta}_k\| \sim \begin{cases} \mathcal{O}(1/\xi(k, N)), & \gamma = 1, \\ \mathcal{O}(\gamma^{-(N+1)\xi(k, N)}), & \gamma > 1, \end{cases} \quad (19)$$

**Proof.** In the case where  $\gamma = 1$ , (19) is immediate from Theorem 3. Hence, suppose that  $\gamma > 1$ . From Theorem 3, it follows that there exist  $M_0 > 0$  and  $K_0 \geq 0$  such that, for all  $k \geq K$ ,  $\|\tilde{\theta}_k\| \leq M_0 / \sum_{i=0}^{\xi(k,N)-1} \rho_{i(N+1)}$ . Let  $M = M_0 \gamma^N$  and  $K = \max(K_0, N + 1)$ . Therefore, since, for all  $k \geq 0$ ,  $\rho_{k(N+1)} = \gamma^{k(N+1)+1}$ , it follows that, for all  $k \geq K$ ,

$$\begin{aligned} \|\tilde{\theta}_k\| &\leq \frac{M_0}{\sum_{i=0}^{\xi(k,N)-1} \gamma^{i(N+1)+1}} = \frac{M_0}{\gamma} \frac{\gamma^{N+1} - 1}{\gamma^{(N+1)\xi(k, N)} - 1} \\ &= \frac{M_0}{\gamma} \frac{\gamma^{N+1} - 1}{\gamma^{(N+1)\xi(k, N)} - 1} \frac{\gamma^{(N+1)\xi(k, N)}}{\gamma^{(N+1)\xi(k, N)}} \leq \frac{M}{\gamma^{(N+1)\xi(k, N)}}. \quad \square \end{aligned}$$

Since generally, for all  $k \geq 0$ ,  $\beta_k \geq 1$ , it follows that  $1/\sum_{i=0}^{\xi(k,N)-1} \rho_{i(N+1)} \leq 1/\xi(k, N)$ . Thus, this analysis suggests that, in the case where  $\beta_k > 1$  for an infinite set of indices, the asymptotic convergence rate of VRF is faster than the asymptotic convergence rate of RLS without forgetting.

## 5 Consistency of VRF

A sequence  $(X_k)_{k \geq 0}$  of vector-valued random variables on  $\Omega$  is a *consistent estimator* of  $\theta \in \mathbb{R}^n$  if, for all  $\varepsilon > 0$ ,

$$\lim_{k \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega: \|X_k(\omega) - \theta\| < \varepsilon\}) = 1. \quad (20)$$

When  $\theta$  is understood, for brevity, we call such sequences *consistent*. In this section we give conditions on  $\beta_k$  which are necessary and sufficient for the consistency of VRF when the measurements of  $\phi_k \theta$  are corrupted by noise.

**Theorem 4** *Let  $(\phi_k)_{k \geq 0}$  be a persistently exciting sequence with window  $N$ , lower bound  $\alpha$ , and upper bound  $\beta < \infty$ . Let  $\theta \in \mathbb{R}^n$ ,  $P_0 \in \mathbb{P}^n$ , and  $\theta_0 \sim \mathcal{N}(\theta, P_0)$ . Let  $(\nu_k)_{k \geq 0}$  be an  $\mathbb{R}^p$ -valued stationary Gaussian white-noise process with variance  $V$  and uncorrelated with  $\theta_0$ , and define  $y_k = \phi_k \theta + \nu_k$ . Furthermore, for all  $k \geq 0$ , let  $\beta_k \geq 1$ , and define  $\theta_{k+1}$  by (11)–(13). Then, for all  $k \geq 0$ ,  $\theta_k$  is a Gaussian random variable with mean  $\theta$ . Then*

$$\frac{\alpha \lambda_{\min}(V)}{\beta^2} \lim_{k \rightarrow \infty} \frac{q_{1, \xi(k, N)}}{s_{u, \xi(k, N)}^2} \leq \lim_{k \rightarrow \infty} \lambda_{\min}(\text{var}(\theta_k)) \quad (21)$$

$$\leq \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(\text{var}(\theta_k)) \leq \frac{\beta \lambda_{\max}(V)}{\alpha^2} \overline{\lim}_{k \rightarrow \infty} \frac{q_{u, \xi(k, N)}}{s_{1, \xi(k, N)}^2}, \quad (22)$$

where, for all  $j \geq 0$ ,  $s_{1, j} \triangleq \sum_{i=0}^{j-1} \rho_{i(N+1)}$ ,  $s_{u, j} \triangleq \sum_{i=0}^j \rho_{i(N+1)+N}$ ,  $q_{1, j} \triangleq \sum_{i=0}^{j-1} \rho_{i(N+1)}^2$ , and  $q_{u, j} \triangleq \sum_{i=0}^j \rho_{i(N+1)+N}^2$ .

**Proof.** With base case  $\theta_0 \sim \mathcal{N}(\theta, P_0)$ , suppose for induction that  $\theta_k \sim \mathcal{N}(\theta, \text{var}(\theta_k))$ . Define  $\tilde{\theta}_k \triangleq \theta_k - \theta$ . From (11), it follows that  $\tilde{\theta}_{k+1} = \beta_k^{-1} P_{k+1} P_k^{-1} \tilde{\theta}_k + P_{k+1} \phi_k^T \nu_k$ . Since  $\theta_k \sim \mathcal{N}(\theta, \text{var}(\theta_k))$ , it follows from Lemma A.1 that  $\tilde{\theta}_k \sim \mathcal{N}(0, \text{var}(\theta_k))$ . Next, define  $z_k \triangleq P_k^{-1} \tilde{\theta}_k$ . Since  $\tilde{\theta}_k \sim \mathcal{N}(0, \text{var}(\theta_k))$ , it follows from Lemma A.1 that  $z_k \sim \mathcal{N}(0, P_k^{-1} \text{var}(\theta_k) P_k^{-1})$ . Since  $\nu_k$  is uncorrelated with  $\nu_0, \dots, \nu_{k-1}, \theta_0$ , it follows that  $\nu_k$  and  $z_k$  are also uncorrelated. Furthermore,  $z_{k+1} = \beta_k^{-1} z_k + \phi_k^T \nu_k$ , and thus  $[z_k \ \nu_k]^T \sim \mathcal{N}(0_{2 \times 1}, \text{diag}(\text{var}(z_k), V))$ . Therefore, Lemma A.1 implies that  $z_{k+1} \sim \mathcal{N}(0, \text{var}(z_{k+1}))$  and  $\text{var}(z_{k+1}) = \beta_k^{-2} \text{var}(z_k) + \phi_k^T V \phi_k$ . Since  $\theta_{k+1} = P_{k+1} z_{k+1} + \theta$ , it follows from Lemma A.1 that  $\theta_{k+1} \sim \mathcal{N}(\theta, P_{k+1} \text{var}(z_{k+1}) P_{k+1}^{-1})$ . Thus, for all  $k \geq 0$ ,  $\theta_k$  is a Gaussian random variable with mean  $\theta$ . Since  $\text{var}(z_0) = P_0^{-1} P_0 P_0^{-1} = P_0^{-1}$ , it follows that  $\text{var}(z_{k+1}) = \rho_k^{-2} (P_0^{-1} + \sum_{i=0}^k \rho_i^2 \phi_i^T V \phi_i)$ . For convenience, define  $M_k \triangleq \sum_{i=0}^k \rho_i \phi_i^T \phi_i$ ,  $M_{\nu, k} \triangleq \sum_{i=0}^k \rho_i^2 \phi_i^T V \phi_i$ ,  $H_{0, k} \triangleq (P_0^{-1} + M_k)^{-1} P_0^{-1} (P_0^{-1} + M_k)^{-1}$ ,  $H_{\nu, k} \triangleq (P_0^{-1} + M_k)^{-1} M_{\nu, k} (P_0^{-1} + M_k)^{-1}$ . For all  $k \geq 0$ , it follows from Lemma 2 that

$$\alpha s_{1, \xi(k, N)} I_n \leq M_k \leq \beta s_{u, \xi(k, N)} I_n, \quad (23)$$

$$\alpha \lambda_{\min}(V) q_{1, \xi(k, N)} I_n \leq M_{\nu, k} \leq \beta \lambda_{\max}(V) q_{u, \xi(k, N)} I_n. \quad (24)$$

Since  $\beta_k \geq 1$ , it follows that  $q_{\ell, \xi(k, N)} \rightarrow \infty$  as  $k \rightarrow \infty$ , and thus  $\lambda_{\max}(M_k) \rightarrow \infty$  as  $k \rightarrow \infty$ . From this result and Lemma A.4 it follows that  $\overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(H_{0, k}) \leq \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(P_0^{-1}) / \lambda_{\max}(M_k)^2 = 0$ . Hence,  $\overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(H_{0, k}) = 0$ . Noting that  $\text{var}(\theta_k) = H_{0, k} + H_{\nu, k}$ , it follows from Lemmas A.3 and A.4, (23), and (24) that

$$\begin{aligned} \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(\text{var}(\theta_k)) &\leq \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(H_{0, k}) + \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(H_{\nu, k}) \\ &= \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(H_{\nu, k}) \leq \overline{\lim}_{k \rightarrow \infty} \frac{\lambda_{\max}(M_{\nu, k})}{\lambda_{\max}(P_0^{-1} + M_k)^2} \\ &\leq \overline{\lim}_{k \rightarrow \infty} \frac{\lambda_{\max}(M_{\nu, k})}{\lambda_{\max}(M_k)^2} \leq \frac{\beta \lambda_{\max}(V)}{\alpha^2} \overline{\lim}_{k \rightarrow \infty} \frac{q_{u, \xi(k, N)}}{s_{1, \xi(k, N)}^2}, \end{aligned}$$

Since  $H_{0, k} \in \mathbb{P}^n$  and  $\underline{\lim}_{k \rightarrow \infty} \lambda_{\min}(H_{0, k}) \leq \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(H_{0, k}) = 0$ , it follows that  $\underline{\lim}_{k \rightarrow \infty} \lambda_{\min}(H_{0, k}) = 0$ . Thus, from

Lemmas A.3, A.4, and A.5, [11, Fact 10.4.13], (23), and (24), it follows that

$$\begin{aligned} & \frac{\alpha \lambda_{\min}(V)}{\beta^2} \lim_{k \rightarrow \infty} \frac{q_{1,\xi(k,N)}}{s_{1,\xi(k,N)}^2} \leq \lim_{k \rightarrow \infty} \frac{\lambda_{\min}(M_{\nu,k})}{\lambda_{\min}(M_k)^2} \\ & = \lim_{k \rightarrow \infty} \frac{\lambda_{\min}(M_{\nu,k})}{[\lambda_{\max}(P_0^{-1}) + \lambda_{\min}(M_k)]^2} \leq \lim_{k \rightarrow \infty} \frac{\lambda_{\max}(M_{\nu,k})}{\lambda_{\min}(P_0^{-1} + M_k)^2} \\ & \leq \lim_{k \rightarrow \infty} [\lambda_{\min}(H_{0,k}) + \lambda_{\min}(H_{\nu,k})] = \lim_{k \rightarrow \infty} \lambda_{\min}(\text{var}(\theta_k)). \quad \square \end{aligned}$$

**Corollary 4** *Under the notation and assumptions of Theorem 4, consider the following statements: i)  $\overline{\lim}_{k \rightarrow \infty} q_{u,\xi(k,N)}/s_{1,\xi(k,N)}^2 = 0$ , ii)  $(\theta_k)_{k \geq 0}$  is consistent, iii)  $\overline{\lim}_{k \rightarrow \infty} q_{1,\xi(k,N)}/s_{u,\xi(k,N)}^2 = 0$ . Then i)  $\implies$  ii)  $\implies$  iii).*

**Proof.** To prove i)  $\implies$  ii), let  $\overline{\lim}_{k \rightarrow \infty} q_{u,\xi(k,N)}/s_{1,\xi(k,N)}^2 = 0$ . Then  $\lim_{k \rightarrow \infty} \lambda_{\max}(\text{var}(\theta_k)) = 0$ . Thus, from Lemma A.2, it follows that  $(\theta_k)_{k \geq 0}$  is consistent. To prove ii)  $\implies$  iii), suppose that  $(\theta_k)_{k \geq 0}$  is consistent. Then, from Lemma A.2, it follows that  $\underline{\lim}_{k \rightarrow \infty} \lambda_{\min}(\text{var}(\theta_k)) = 0$ , and therefore  $\underline{\lim}_{k \rightarrow \infty} q_{1,\xi(k,N)}/s_{u,\xi(k,N)}^2 = 0$ .  $\square$

**Corollary 5** *Under the notation and assumptions of Theorem 4, the following statements hold: i) assume that  $\prod_{k \geq 0} \beta_k$  is finite. Then  $(\theta_k)_{k \geq 0}$  is consistent; ii) let  $\beta_0 = 1$  and for all  $k > 0$ , let  $\beta_k = 1 + \frac{1}{k}$ . Then  $(\theta_k)_{k \geq 0}$  is consistent; iii) let  $\gamma \in [1, \infty)$ , and, for all  $k \geq 0$ , let  $\beta_k = \gamma$ . Then  $(\theta_k)_{k \geq 0}$  is consistent if and only if  $\gamma = 1$ .*

**Proof.** To prove i), suppose that  $\prod_{k \geq 0} \beta_k = \rho$  and let  $\varepsilon > 0$ . Thus there exists  $K > 0$  such that, for all  $i \geq K$ ,  $\rho - \varepsilon < \rho_i < \rho + \varepsilon$ . Let  $k_\varepsilon > 0$  be the smallest integer such that  $\xi(k_\varepsilon, N)(N+1) \geq K$ , and define  $B_\varepsilon \triangleq \sum_{i=0}^{\xi(k_\varepsilon, N)} \rho_{i(N+1)+N}$  and  $C_\varepsilon \triangleq \sum_{i=0}^{\xi(k_\varepsilon, N)} \rho_{i(N+1)}$ . Then, for all  $k > k_\varepsilon$ ,

$$\frac{q_{u,\xi(k,N)}}{s_{1,\xi(k,N)}^2} \leq \frac{B_\varepsilon + (\rho + \varepsilon)^2(\xi(k, N) - \xi(k_\varepsilon, N) - 1)}{(C_\varepsilon + (\rho - \varepsilon)(\xi(k, N) - \xi(k_\varepsilon, N) - 1))^2}. \quad (25)$$

Since the limit superior of the left-hand side of (25) is zero, it follows that  $(\theta_k)_{k \geq 0}$  is consistent. To prove ii), for all  $k \geq 0$ , let  $\beta_k = 1 + 1/k$ . Then, for all  $i \geq 0$ ,  $\rho_i = i+1$ , and thus  $q_{u,\xi(k,N)}$  and  $s_{1,\xi(k,N)}^2$  are polynomials of degree three and four, respectively. Hence the limit superior is zero, and therefore  $(\theta_k)_{k \geq 0}$  is consistent. To prove iii), suppose that  $\gamma = 1$ . Then  $\overline{\lim}_{k \rightarrow \infty} q_{u,\xi(k,N)}/s_{1,\xi(k,N)}^2 = \overline{\lim}_{k \rightarrow \infty} \xi(k, N)^{-1} = 0$ . Hence  $(\theta_k)_{k \geq 0}$  is consistent. Conversely, suppose  $\gamma > 1$ . Then, for all  $i \geq 0$ ,  $\rho_i = \gamma^{i+1}$ , and thus

$$\begin{aligned} \overline{\lim}_{k \rightarrow \infty} \frac{q_{1,\xi(k,N)}}{s_{u,\xi(k,N)}^2} & = \frac{1}{\gamma^{2N}} \frac{(1 - \gamma^{(N+1)})^2}{1 - \gamma^{2(N+1)}} \overline{\lim}_{k \rightarrow \infty} \frac{1 - \gamma^{2(N+1)(\xi(k,N)+1)}}{(1 - \gamma^{(N+1)(\xi(k,N)+1)})^2} \\ & = \frac{1}{\gamma^{2N}} \frac{\gamma^{(N+1)} - 1}{\gamma^{(N+1)} + 1}, \end{aligned}$$

which is positive because  $\gamma > 1$ . Therefore,  $(\theta_k)_{k \geq 0}$  is not consistent.  $\square$

Corollary 5 shows that if  $\prod_{k \geq 0} \beta_k$  converges, then VRF is consistent, but also that the converse is false, since  $\prod_{k > 0} 1 + \frac{1}{k} = \infty$ . Furthermore, CRF is consistent if and only if  $\lambda = 1$ , and thus RLS with a constant, nontrivial forgetting factor is not consistent.

## 6 Example: Abruptly Changing Parameters

Consider a mass-spring-damper system with  $m = 5$  kg,  $k = 1$  N/m, and  $b = 1$  N-sec/m sampled at 1 sample/sec, and suppose that at 100 samples the parameters of the system abruptly change to  $k = 10$  N/m and  $b = 0.01$  N-sec/m. This process is modeled by the time-varying discrete-time transfer function

$$G_k(\mathbf{q}) = \begin{cases} \frac{0.4606\mathbf{q} + 0.4307}{\mathbf{q}^2 - 1.64\mathbf{q} + 0.8187}, & k < 100, \\ \frac{0.4218\mathbf{q} + 0.4215}{\mathbf{q}^2 - 0.3116\mathbf{q} + 0.998}, & k \geq 100, \end{cases} \quad (26)$$

where  $\mathbf{q}$  is the forward shift operator. For all  $k \geq 0$ , let  $u_k \sim \mathcal{N}(0, 1)$ , and define

$$\beta_k \triangleq 1 + \eta \text{sat}_\gamma(\|y_k - \phi_k \theta_k\|), \quad (27)$$

where  $\eta, \gamma > 0$ , and  $\text{sat}_\gamma$  is the unit-slope saturation function with saturation level  $\gamma$ . Figure 1 shows the performance of VRF with  $\gamma = \eta = 1$  and CRF with  $\lambda = 0.99$ . VRF converges to the initial system parameters and re-converges to the modified parameters in about 10 samples, illustrating Theorem 2. In contrast, while CRF converges to the initial parameters, reconvergence to the modified parameters is still not achieved at 200 samples. Next, consider the same system with the output corrupted by additive noise  $\nu_k \sim \mathcal{N}(0, 0.05)$ , and define

$$\beta_k \triangleq \begin{cases} 1 + \eta \text{sat}_\gamma(E_\tau), & E_\tau > 1, \\ 1, & E_\tau \leq 1, \end{cases} \quad (28)$$

where  $\tau \in \mathbb{N}$  and  $E_\tau \triangleq (\frac{1}{\tau} \sum_{i=k-\tau}^k \|y_i - \phi_i \theta_i\|^2)^{1/2}$ . Figure 2 shows the performance of VRF with  $\eta = 1$ ,  $\gamma = 5$ , and  $\tau = 10$ , and CRF with  $\lambda = 0.99$ . VRF converges to the initial parameters and then reconverges to the new parameters in roughly 30 samples. As in the previous case, CRF converges to the initial parameters, but at 200 samples has still not reconverged to the modified parameters.  $\diamond$

## Acknowledgments

This research was partially supported by AFOSR under DDDAS grant FA9550-16-1-0071 (Dynamic Data-Driven Applications Systems).

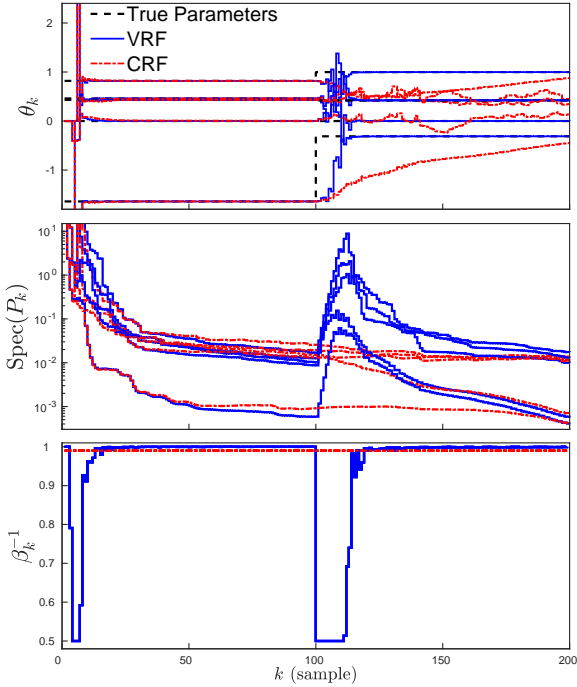


Fig. 1. The parameter estimate  $\theta_k$  given by VRF with  $\beta_k$  defined by (27) reconverges after an abrupt change in the system as guaranteed by Theorem 2. In contrast, The parameter estimate given by CRF with  $\lambda = 0.99$  requires many samples to reconverge.

## References

- [1] L. Ljung and T. Soderstrom. *Theory and Practice of Recursive Identification*. The MIT Press, 1983.
- [2] K. J. Astrom. *Adaptive Control*. Addison-Wesley, Reading, MA, second edition, 1995.
- [3] A. Albert and R. W. Sittler. A Method for Computing Least Squares Estimators that Keep Up with the Data. *SIAM J. Contr.*, 3(3):384–417, 1965.
- [4] T. R. Fortescue, L. S. Kershenbaum, and B. E. Ydstie. Implementation of Self-tuning Regulators with Variable Forgetting Factors. *Automatica*, 17(6):831–835, 1981.
- [5] C. Paleologu, J. Benesty, and C. Silviu. A Robust Variable Forgetting Factor Recursive Least-Squares Algorithm for System Identification. *IEEE Sig. Proc. Lett.*, 15:597–600, 2008.
- [6] S.-H. Leung and C. F. So. Gradient-Based Variable Forgetting Factor RLS Algorithm in Time-Varying Environments. *IEEE Trans. Sig. Proc.*, 53(8):3141–3150, 2005.
- [7] S. Song, J.-S. Lim, S. J. Baek, and K.-M. Sung. Gauss Newton variable forgetting factor recursive least squares for time varying parameter tracking. *Electron. Lett.*, 36(11):988–990, 2000.
- [8] D. J. Park, B. E. Jun, and Kim J. H. Fast tracking RLS algorithm using novel variable forgetting factor with unity zone. *Electron. Lett.*, 27(23):2150–2151, 1991.
- [9] A. A. Ali, J. B. Hoagg, M. Mossberg, and D. S. Bernstein. On the stability and convergence of a sliding-window variable-regularization recursive-least-squares algorithm. *Int. J. Adapt. Control Signal Process.*, 30:715–735, 2016.

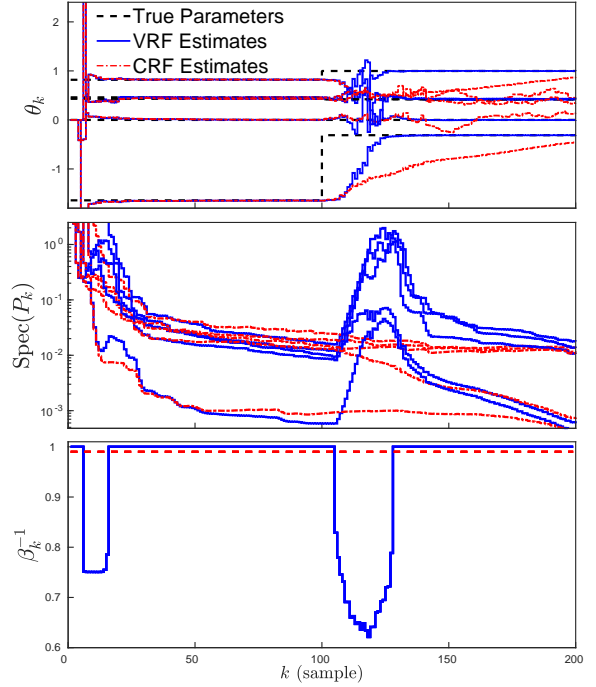


Fig. 2. The parameter estimate  $\theta_k$  given by VRF with  $\beta_k$  defined by (28) reconverges after an abrupt change in the system with noisy measurements. In contrast, the parameter estimate given by CRF with  $\lambda = 0.99$  requires many samples to reconverge.

- [10] S. A. Ul Islam and D. S. Bernstein. Recursive Least Squares for Real-Time Implementation. *IEEE Cont. Sys. Mag.*, 39(3):82–85, 2019.
- [11] D. S. Bernstein. *Scalar, Vector, and Matrix Mathematics: Theory, Facts, and Formulas - Revised and Expanded Edition*. Princeton University Press, 41 William Street, Princeton, NJ, 08540, 2018.

## A Lemmas

**Lemma A.1** Let  $A \in \mathbb{R}^{n \times n}$  and  $b \in \mathbb{R}^n$ . Let  $X \sim \mathcal{N}(\mu, P)$  and define  $Y \triangleq AX + b$ . Then  $Y \sim \mathcal{N}(A\mu + b, APA^T)$ .

**Lemma A.2** Let  $(\Omega, \Sigma, P)$  be a probability space, let  $\theta \in \mathbb{R}^n$ , and let  $(X_k: \Omega \rightarrow \mathbb{R}^n)_{k \geq 0}$  be a sequence of random variables such that, for all  $k \geq 0$ ,  $X_k \sim \mathcal{N}(\theta, \Sigma_k)$ . Then  $(X_k)_{k \geq 0}$  is a consistent estimator for  $\theta$  if and only if  $\lim_{k \rightarrow \infty} \Sigma_k = 0$ .

**Lemma A.3** Let  $(A_k)_{k \geq 0}, (B_k)_{k \geq 0} \subset (\mathbb{N}^n)$ . Then

$$\begin{aligned} \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(A_k + B_k) &\leq \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(A_k) + \overline{\lim}_{k \rightarrow \infty} \lambda_{\max}(B_k), \\ \underline{\lim}_{k \rightarrow \infty} \lambda_{\min}(A_k + B_k) &\geq \underline{\lim}_{k \rightarrow \infty} \lambda_{\min}(A_k) + \underline{\lim}_{k \rightarrow \infty} \lambda_{\min}(B_k). \end{aligned}$$

**Lemma A.4** Let  $A \in \mathbb{N}^n$  and  $B \in \mathbb{P}^n$ . Then, for all  $i = 1, \dots, n$ ,

$$\frac{\lambda_{\min}(A)}{\lambda_i(B)^2} \leq \lambda_i(B^{-1}AB^{-1}) \leq \frac{\lambda_{\max}(A)}{\lambda_i(B)^2}. \quad (\text{A.1})$$

Now assume that  $A \in \mathbb{P}^n$ . Then there exist  $0 < b_1 \leq b_2$  and  $0 < a_1 \leq a_2$  such that

$$a_1 I_n \leq A \leq a_2 I_n, \quad (\text{A.2})$$

$$b_1 I_n \leq B \leq b_2 I_n. \quad (\text{A.3})$$

Furthermore, for all  $a_1, a_2, b_1, b_2$  satisfying (A.2), (A.3),

$$\frac{a_1}{b_2^2} I_n \leq B^{-1} A B^{-1} \leq \frac{a_2}{b_1^2} I_n. \quad (\text{A.4})$$

**Lemma A.5** *Let  $a \in [0, \infty)$ , let  $(b_k)_{k \geq 0}, (c_k)_{k \geq 0} \subset [0, \infty)$ , and assume that  $\lim_{k \rightarrow \infty} b_k = \infty$ . Then, for all  $p \geq 0$ ,*

$$\lim_{k \rightarrow \infty} \frac{c_k}{(a + b_k)^p} = \lim_{k \rightarrow \infty} \frac{c_k}{b_k^p}. \quad (\text{A.5})$$