2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

Recursive Least Squares with Matrix Forgetting

Adam L. Bruce[1], Ankit Goel[2], and Dennis S. Bernstein[3]

Abstract— **This paper considers an extension of recursive least squares (RLS), where the cost function is modified to include a matrix forgetting factor. Minimization of the modified cost function provides a framework for combined variable-rate and variable-direction (RLS-VRDF) forgetting. This extension of RLS simultaneously addresses two key issues in standard RLS, namely, the need for variable-rate forgetting due to changing plant parameters as well as the need for variable-direction covariance updating due to the loss of persistency. The performance of RSL-VRDF is illustrated by an example with abrupt parameter changes and loss of persistency.**

I. Introduction

Recursive least squares (RLS) is an algorithm that is widely used in signal processing, identification, and adaptive control [1–4]. Standard RLS employs a forgetting factor λ that enhances the importance of recent data over older data, but unfortunately, the performance of RLS is often extremely sensitive to the choice of λ. To overcome this problem, various extensions of RLS have been developed to include variable-rate forgetting [5–9] or covariance resetting [10].

An additional weakness of RLS is that the use of a fixed- or variable-rate forgetting factor may cause the covariance to diverge when the input signal is not persistently exciting [11–15]. A variety of techniques have been developed for overcoming the divergence due to lack of persistency [11, 16, 17]. One approach to this problem is to restrict forgetting to the subspace in which the data provide new information about the parameters, [18–24]. Consequently, the direction of the forgetting is varied based on the information content of the measurements.

The main contribution of this paper is a modified cost function whose minimization yields a matrix forgetting RLS algorithm that can be specialized into a combined variable-rate and variable-direction RLS algorithm (RLS-VRDF). The resulting extension of RLS thus seeks to overcome both changing parameters and loss of persistency. In addition, since RLS-VRDF is obtained by minimizing a cost function, this modification of RLS has a known optimal interpretation, in contrast with extensions of RLS obtained by direct modification of the RLS update equations [16, 17]. Since this paper gives a general matrix forgetting RLS algorithm and a cost function which it minimizes, and then numerically investigates RLS-VRDF, while [25] only includes RLS-VRF and [26] only includes RLS-VDF, the contribution of this paper goes beyond [25, 26].

Matrix forgetting algorithms are also given in [27, 28]. However, [27] assumes an ARMA model of the system and develops matrix forgetting that only applies to the Instrumental Variable Method. In contrast, this paper makes no assumptions about the system and directly generalizes standard RLS. In [28], a matrix forgetting algorithm is derived by modifying the standard RLS cost function, but results in a covariance matrix that is not generally symmetric. The algorithm also assumes that there is a single output and restricts the forgetting matrix to be both constant and symmetric. In this paper, the covariance matrix is guaranteed to be symmetric, but the forgetting matrix need not be either symmetric or constant–allowing for a wide range of choices, such as RLS-VRDF. Furthermore, there is no assumption on the number of outputs.

The paper is organized as follows. In Section II, we introduce preliminary results on least squares optimization, including a recursive update algorithm for a general least squares cost which does not use the matrix inversion lemma (Lemma 1), and show that standard RLS can be obtained as a special case of this cost. In Section III, we specialize Proposition 2 in Section II to the case of matrix forgetting. Then, in Section IV, we specialize Theorem 1 in Section III further to the cases of RLS-VRF, RLS-VDF, and RLS-VRDF. Finally, in Section V, we show the performance of the different algorithms on a system identification example with both abruptly changing parameters and abrupt loss of persistency.

II. Preliminary Results on Least Squares Optimization

The following result on least squares optimization is an immediate consequence of Lemma 2 in the appendix.

Proposition 1: For all $k \geq 0$, let $y_k \in \mathbb{R}^p$, $\phi_k \in \mathbb{R}^{p\times n}$, and $\alpha_k \in \mathbb{R}^n$, let $Q_k \in \mathbb{R}^{p\times p}$ be positive semidefinite, let $R_k \in \mathbb{R}^{n\times n}$ be symmetric, define $J_k \colon \mathbb{R}^n \to [0,\infty)$ by

$$J_k(\theta) \triangleq \sum_{i=0}^{k}(y_i - \phi_i\theta)^{\mathrm{T}}Q_i(y_i-\phi_i\theta) + (\theta-\alpha_k)^{\mathrm{T}}R_k(\theta-\alpha_k), \quad (1)$$

[1]PhD Candidate, Dept. of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109. admbruce@umich.edu.
[2]Postdoctoral Researcher, Dept. of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109. ankgoel@umich.edu.
[3]Professor, Dept. of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109. dsbaero@umich.edu.

and assume that

$$A_k \triangleq \sum_{i=0}^{k} \phi_i^{\mathrm{T}} Q_i \phi_i + R_k \tag{2}$$

is positive definite. Then, for all $k \geq 0$, J_k is quadratic and strictly convex, and thus has a unique global minimizer, which is also the only local minimizer. For all $k \geq 0$, define

$$\theta_{k+1} \triangleq \operatorname{argmin}_{\theta \in \mathbb{R}^n} J_k(\theta). \tag{3}$$

Then, for all $k \geq 0$,

$$\theta_{k+1} = -A_k^{-1} b_k, \tag{4}$$

and the minimum value of J_k is given by

$$J_k(\theta_{k+1}) = c_k - b_k^{\mathrm{T}} A_k^{-1} b_k, \tag{5}$$

where

$$b_k \triangleq -\sum_{i=0}^{k} \phi_i^{\mathrm{T}} Q_i y_i - R_k \alpha_k, \tag{6}$$

$$c_k \triangleq \sum_{i=0}^{k} y_i^{\mathrm{T}} Q_i y_i + \alpha_k^{\mathrm{T}} R_k \alpha_k. \tag{7}$$

Note that in the case where R_k is positive definite (positive semidefinite) it follows that (5) is positive (nonegative). If however R_k is not positive semidefinite, (5) may be negative.

The next result is a recursive variation of Proposition 1.

Proposition 2: Under the notation and assumptions of Proposition 1, let $\theta_0 \in \mathbb{R}^n$ and define $R_{-1} \in \mathbb{R}^{n\times n}$ and $\alpha_{-1} \in \mathbb{R}^n$ such that $R_{-1}(\alpha_{-1} - \theta_0) = 0$. Then, for all $k \geq 0$,

$$A_k = A_{k-1} + \phi_k^{\mathrm{T}} Q_k \phi_k + R_k - R_{k-1}, \tag{8}$$

$$\begin{aligned}\theta_{k+1} = \theta_k &+ A_k^{-1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k) \\ &+ A_k^{-1}[R_k(\alpha_k - \theta_k) - R_{k-1}(\alpha_{k-1} - \theta_k)]. \end{aligned} \tag{9}$$

Proof. Since $A_{-1} = R_{-1}$, it follows that

$$\begin{aligned} A_0 &= \phi_0^{\mathrm{T}} Q_0 \phi_0 + R_0 \\ &= R_{-1} + \phi_0^{\mathrm{T}} Q_0 \phi_0 + R_0 - R_{-1} \\ &= A_{-1} + \phi_0^{\mathrm{T}} Q_0 \phi_0 + R_0 - R_{-1}, \end{aligned}$$

which confirms (8) for $k = 0$. Since A_0 is positive definite, Lemma 2 implies that

$$\begin{aligned} \theta_1 &= -A_0^{-1} b_0 \\ &= A_0^{-1}(\phi_0^{\mathrm{T}} Q_0 y_0 + R_0 \alpha_0) \\ &= A_0^{-1}(\phi_0^{\mathrm{T}} Q_0 y_0 + \phi_0^{\mathrm{T}} Q_0 \phi_0 \theta_0 - \phi_0^{\mathrm{T}} Q_0 \phi_0 \theta_0 + R_0 \alpha_0) \\ &= A_0^{-1}(\phi_0^{\mathrm{T}} Q_0 \phi_0 \theta_0 + R_0 \theta_0) + A_0^{-1} \phi_0^{\mathrm{T}} Q_0 (y_0 - \phi_0 \theta_0) \\ &\quad + A_0^{-1} R_0 (\alpha_0 - \theta_0) \\ &= \theta_0 + A_0^{-1} \phi_0^{\mathrm{T}} Q_0 (y_0 - \phi_0 \theta_0) \\ &\quad + A_0^{-1} R_0 (\alpha_0 - \theta_0) - A_0^{-1} R_{-1} (\alpha_{-1} - \theta_0), \end{aligned}$$

which confirms (9) for $k = 0$.

Now let $k \geq 1$. From (2) it follows that

$$\begin{aligned} A_k &= \sum_{i=0}^{k} \phi_i^{\mathrm{T}} Q_i \phi_i + R_k \\ &= \sum_{i=0}^{k-1} \phi_i^{\mathrm{T}} Q_i \phi_i + \phi_k^{\mathrm{T}} Q_k \phi_k + R_k \\ &= A_{k-1} - R_{k-1} + \phi_k^{\mathrm{T}} Q_k \phi_k + R_k, \end{aligned}$$

which confirms (8). Furthermore, b_k can be written recursively as

$$b_k = b_{k-1} - \phi_k^{\mathrm{T}} Q_k y_k + R_{k-1} \alpha_{k-1} - R_k \alpha_k.$$

It thus follows from Lemma 2 that

$$\begin{aligned} \theta_{k+1} &= -A_k^{-1} b_k \\ &= -A_k^{-1}(b_{k-1} - \phi_k^{\mathrm{T}} Q_k y_k + R_{k-1}\alpha_{k-1} - R_k \alpha_k) \\ &= -A_k^{-1}(b_{k-1} - \phi_k^{\mathrm{T}} Q_k y_k + \phi_k^{\mathrm{T}} Q_k \phi_k \theta_k - \phi_k^{\mathrm{T}} Q_k \phi_k \theta_k) \\ &\quad - A_k^{-1}(R_{k-1}\alpha_{k-1} - R_k \alpha_k) \\ &= -A_k^{-1}(b_{k-1} - \phi_k^{\mathrm{T}} Q_k \phi_k \theta_k) + A_k^{-1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k) \\ &\quad + A_k^{-1}(R_k \alpha_k - R_{k-1}\alpha_{k-1}) \\ &= A_k^{-1}(A_{k-1}\theta_k + \phi_k^{\mathrm{T}} Q_k \phi_k \theta_k) + A_k^{-1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k) \\ &\quad + A_k^{-1}(R_k \alpha_k - R_{k-1}\alpha_{k-1}) \\ &= A_k^{-1}(A_{k-1} + \phi_k^{\mathrm{T}} Q_k \phi_k + R_k - R_{k-1})\theta_k \\ &\quad + A_k^{-1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k) \\ &\quad + A_k^{-1}(R_k \alpha_k - R_{k-1}\alpha_{k-1}) - A_k^{-1}(R_k - R_{k-1})\theta_k \\ &= \theta_k + A_k^{-1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k) \\ &\quad + A_k^{-1}[R_k(\alpha_k - \theta_k) - R_{k-1}(\alpha_{k-1} - \theta_k)], \end{aligned}$$

which confirms (9). □

The following corollary of Proposition 2 is the classical least squares result, which we provide for comparison with the matrix forgetting result in Section IV.

Corollary 1: Let $\lambda \in (0, 1]$ and $\theta_0 \in \mathbb{R}^n$, and let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite. For all $k \geq 0$, let $y_k \in \mathbb{R}^p$ and $\phi_k \in \mathbb{R}^{p\times n}$, and define $J_k \colon \mathbb{R}^n \to [0, \infty)$ by

$$\begin{aligned}\bar{J}_k(\theta) \triangleq \sum_{i=0}^{k} &\lambda^{k-i}(y_i - \phi_i \theta)^{\mathrm{T}}(y_i - \phi_i \theta) \\ &+ \lambda^k (\theta - \theta_0)^{\mathrm{T}} P_0^{-1} (\theta - \theta_0). \end{aligned} \tag{10}$$

Then, for all $k \geq 0$, J_k is quadratic and strictly convex, and thus has a unique global minimizer, which is also the only local minimizer. For all $k \geq 0$, define

$$\theta_{k+1} \triangleq \operatorname{argmin}_{\theta \in \mathbb{R}^n} \bar{J}_k(\theta). \tag{11}$$

Then, for all $k \geq 0$,

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^{\mathrm{T}} (y_k - \phi_k \theta_k) \tag{12}$$

where

$$P_{k+1} = \frac{1}{\lambda} P_k - \frac{1}{\lambda} P_k \phi_k^{\mathrm{T}} (\lambda I_p + \phi_k P_k \phi_k^{\mathrm{T}})^{-1} \phi_k P_k. \tag{13}$$

III. RLS WITH MATRIX FORGETTING

Note that (9) requires computation of the $n \times n$ inverse A_k^{-1}. In Corollary 1, the matrix inversion lemma was used to replace the $n \times n$ inverse with a $p \times p$ inverse. As can be seen in the proof of Corollary 1, this reduction in complexity was because R_k is equal to a constant, namely P_0^{-1}, for all $k \geq 0$, and thus the term $R_k - R_{k-1}$ vanishes. An alternative approach is to choose a variable R_k that avoids the need for an $n \times n$ inverse. The following result uses a specific choice of R_k and α_k to obtain a version of RLS with matrix forgetting, thereby providing an explicit quadratic cost function which is minimized by matrix forgetting RLS. Hereafter, R_k^+ denotes the Moore-Penrose pseudoinverse of R_k.

Theorem 1: Let $\theta_0 \in \mathbb{R}^n$, and let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite. Furthermore, for all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$, let $Q_k \in \mathbb{R}^{p\times p}$ be positive definite, let $B_k \in \mathbb{R}^{n\times n}$ be nonsingular, and define $J_k \colon \mathbb{R}^n \to [0, \infty)$ by (1) with

$$R_k \triangleq \begin{cases} P_0^{-1}, & k = -1, \\ \sum_{i=0}^{k} B_i^{-\mathrm{T}} A_{i-1} B_i^{-1} - \sum_{i=1}^{k} A_{i-1}, & k \geq 0 \end{cases} \tag{14}$$

and

$$\alpha_k \triangleq \begin{cases} \theta_0, & k = -1, \\ R_k^+ S_k, & k \geq 0, \end{cases} \tag{15}$$

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

$$S_k \triangleq \sum_{i=0}^{k} \Pi_{k-1} \cdots \Pi_i B_i^{-\mathrm{T}} A_{i-1} B_i^{-1} \theta_i - \sum_{i=1}^{k} \Pi_{k-1} \cdots \Pi_i A_{i-1} \theta_i, \tag{16}$$

and, for all $i \geq 0$, $\Pi_i \triangleq R_i R_i^+$. Then, for all $k \geq 0$, A_k defined by (2) is positive definite. Furthermore, for all $k \geq 0$, let $y_k \in \mathbb{R}^p$, define $P_{k+1} \triangleq A_k^{-1}$, and define θ_{k+1} by (3). Then, for all $k \geq 0$, θ_{k+1} is given by

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

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k) + P_{k+1} (\Pi_k - I_n) \gamma_k, \tag{18}$$

where

$$L_k \triangleq B_k P_k B_k^{\mathrm{T}}, \tag{19}$$

$$\gamma_k \triangleq R_{k-1} \alpha_{k-1} + (B_k^{-\mathrm{T}} A_{k-1} B_k^{-1} - A_{k-1}) \theta_k. \tag{20}$$

Proof. Note that, for all $k \geq 0$, R_k satisfies

$$R_k = R_{k-1} + B_k^{-\mathrm{T}} A_{k-1} B_k^{-1} - A_{k-1}, \tag{21}$$

and α_k satisfies

$$\alpha_k = R_k^+ R_{k-1} \alpha_{k-1} + R_k^+ (B_k^{-\mathrm{T}} A_{k-1} B_k^{-1} - A_{k-1}) \theta_k. \tag{22}$$

Furthermore, note that A_{-1} is positive definite. Hence, suppose for induction that A_{k-1} is positive definite. Since $\alpha_{-1} = \theta_0$, it follows that $R_{-1}(\alpha_{-1} - \theta_0) = 0$ and therefore, from Lemma 1, it follows that, for all $k \geq 0$,

$$\begin{aligned} A_k &= A_{k-1} + \phi_k^{\mathrm{T}} Q_k \phi_k + R_k - R_{k-1}, \\ &= A_{k-1} + \phi_k^{\mathrm{T}} Q_k \phi_k + B_k^{-T} A_{k-1} B_k^{-1} - A_{k-1} \\ &= \phi_k^{\mathrm{T}} Q_k \phi_k + B_k^{-T} A_{k-1} B_k^{-1}. \end{aligned}$$

Since B_k is nonsingular and A_{k-1} is positive definite, it follows that $B_k^{-T} A_{k-1} B_k^{-1}$ is positive definite. Thus A_k is positive definite.

Next, define $L_k \triangleq B_k P_k B_k^{\mathrm{T}}$. From Lemma 1 with $A = B_k^{-T} A_{k-1} B_k^{-1}$, $U = \phi_k^{\mathrm{T}}$, $C = Q_k$, and $V = \phi_k$, it follows that, for all $k \geq 0$,

$$\begin{aligned} P_{k+1} &= (B_k^{-T} A_{k-1} B_k^{-1} + \phi_k^{\mathrm{T}} Q_k \phi_k)^{-1} \\ &= L_k - L_k \phi_k^{\mathrm{T}} (Q_k^{-1} + \phi_k L_k \phi_k^{\mathrm{T}})^{-1} \phi_k L_k, \end{aligned}$$

which confirms (17). Furthermore, from Lemma 1, it follows that, for all $k \geq 0$,

$$\begin{aligned} &\theta_{k+1} - \theta_k - P_{k+1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k) \\ &= P_{k+1} R_k (\alpha_k - \theta_k) - P_{k+1} R_{k-1} (\alpha_{k-1} - \theta_k) \\ &= P_{k+1} [R_k (R_k^+ R_{k-1} \alpha_{k-1} \\ &\quad + R_k^+ [B_k^{-\mathrm{T}} A_{k-1} B_k^{-1} - A_{k-1}] \theta_k - \theta_k) - R_{k-1} (\alpha_{k-1} - \theta_k)] \\ &= P_{k+1} [(\Pi_k R_{k-1} \alpha_{k-1} - R_{k-1} \alpha_{k-1}) \\ &\quad + \Pi_k (B_k^{-\mathrm{T}} A_{k-1} B_k^{-1} - A_{k-1}) \theta_k - (R_k - R_{k-1}) \theta_k] \\ &= P_{k+1} [(\Pi_k - I_n) R_{k-1} \alpha_{k-1} \\ &\quad + (\Pi_k - I_n)(B_k^{-\mathrm{T}} A_{k-1} B_k^{-1} - A_{k-1}) \theta_k] \\ &= P_{k+1} (\Pi_k - I_n) \gamma_k, \end{aligned}$$

which confirms (18). □

If R_k is nonsingular, then $\Pi_k = I_n$. Assuming that this is the case for all $k \geq 0$, Theorem 1 specializes to the following result.

Corollary 2: Under the notation and assumptions of Theorem 1, let $k \geq 0$ and assume that R_k is nonsingular. Then θ_{k+1} is given by

$$P_{k+1} = L_k - L_k \phi_k^{\mathrm{T}} (Q_k^{-1} + \phi_k L_k \phi_k^{\mathrm{T}})^{-1} \phi_k L_k, \tag{23}$$

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^{\mathrm{T}} Q_k (y_k - \phi_k \theta_k), \tag{24}$$

where L_k is defined by (19).

Note that, in Corollary 2, B_k can be viewed as a matrix forgetting factor. To see this, let $\lambda \in (0, 1]$ and let $B_k = \frac{1}{\sqrt{\lambda}} I_n$. Then (23)-(24) specialize to the traditional RLS equations with forgetting factor λ. Hereafter, we refer to this specialization as RLS with constant-rate forgetting (RLS-CRF).

IV. SPECIALIZATIONS

Note that the nonsingular matrix B_k in (19), (23), (24) can be chosen arbitrarily. In particular, the following specializations of Corollary 2 choose B_k in order to achieve variable-rate and variable-direction forgetting.

Variable-Rate Forgetting (VRF). For all $k \geq 0$, let $\beta_k \in (0,\infty)$, $B_k = \beta_k I_n$, and $Q_k = I_p$. Then (19), (17), and (18) are given by

$$L_k = \beta_k P_k, \tag{25}$$

$$P_{k+1} = L_k - L_k\phi_k^{\mathrm{T}}(I_p + \phi_k L_k \phi_k^{\mathrm{T}})^{-1}\phi_k L_k, \tag{26}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k). \tag{27}$$

Equations (25)–(27) give RLS with variable-rate forgetting (RLS-VRF) [25].

Variable-Direction Forgetting (VDF). Compute the singular value decomposition $P_k = U_k\Sigma_k U_k^{\mathrm{T}}$, where $U_k \in \mathbb{R}^{n\times n}$ is orthonormal, and define

$$\psi_k \triangleq \phi_k U_k. \tag{28}$$

Next, let $\varepsilon > 0$ be larger than the noise-to-signal ratio or, if no noise is present, larger than the machine zero. Finally, let $\lambda \in (0,1)$, and define

$$\overline{\Lambda}_k(i,i) \triangleq \begin{cases} \sqrt{\lambda}, & \|\psi_{k,i}\| > \varepsilon, \\ 1, & \text{otherwise}, \end{cases} \tag{29}$$

where $\psi_{k,i}$ is the ith column of ψ_k. Finally, define

$$B_k \triangleq U_k\overline{\Lambda}_k^{-1}U_k^{\mathrm{T}}, \tag{30}$$

Then, with B_k given by (30), equations (19), (23), (24) give RLS with variable-direction forgetting (RLS-VDF) [26].

Variable-Rate and -Direction Forgetting (VRDF). For all $k \geq 0$ and $i = 1,\ldots,p$, let $\beta_{i,k} \in (0,\infty)$, define

$$\overline{D}_k(i,i) \triangleq \begin{cases} \sqrt{\beta_{i,k}}, & \|\psi_{k,i}\| > \varepsilon, \\ 1, & \text{otherwise}, \end{cases} \tag{31}$$

with U_k, ψ_k, and $\psi_{k,i}$ defined as in (28), and let

$$B_k = U_k\overline{D}_k U_k^{\mathrm{T}}. \tag{32}$$

Then, with B_k given by (32), equations (19), (23), (24) give variable-rate and -direction forgetting (RLS-VRDF), which combines RLS-VRF and RLS-VDF.

V. Example: Abrupt Loss and Recovery of Persistency with Abruptly Changing Parameters.

Consider a mass-spring-damper system with $m = 5$ kg, $k = 1$ N/m, and $b = 1$ N·s/m sampled at 1 sample/s, and suppose that at 200 samples the parameters of the system abruptly change to $k = 10$ N/m and $b = 0.01$ N·s/m and then at 1200 samples the parameters of the system abruptly change again to $k = 0.1$ N/m and $b = 10$ N·s/m. This system is modeled by the time-varying discrete-time transfer function

$$G_k(\mathbf{q}) = \begin{cases} \dfrac{0.4606\mathbf{q} + 0.4307}{\mathbf{q}^2 - 1.64\mathbf{q} + 0.8187}, & k < 200, \\ \dfrac{0.4218\mathbf{q} + 0.4215}{\mathbf{q}^2 - 0.3116\mathbf{q} + 0.998}, & 200 \leq k \leq 1200, \\ \dfrac{0.2834\mathbf{q} + 0.1482}{\mathbf{q}^2 - 1.127\mathbf{q} + 0.1353}, & k > 1200, \end{cases} \tag{33}$$

where $\mathbf{q}$ is the forward shift operator. Let the input to (33) be given by

$$u_k = \begin{cases} \tilde{u}_k, & k < 100 \text{ or } k > 1000, \\ \sin(0.01k), & 100 \leq k \leq 1000, \end{cases} \tag{34}$$

where, for all $k \geq 0$,

$$\tilde{u}_k \triangleq \sin(0.01k) + \sin(0.1k) + \sin(k) + \sin(10k). \tag{35}$$

Note that, for all $100 < k < 1000$, (34) is not persistently exciting. Furthermore, suppose that the output is corrupted with additive Gaussian white noise with standard deviation $\sigma = 0.025$. Finally, define

$$\beta_k \triangleq \begin{cases} 1 + \eta \ \mathrm{sat}_\gamma(E_\tau), & E_\tau > 1, \\ 1, & E_\tau \leq 1, \end{cases} \tag{36}$$

where η and γ are positive numbers, sat_γ is the unit-slope saturation function with saturation level γ, τ is a positive integer, and

$$E_\tau \triangleq \left(\frac{1}{\tau}\sum_{i=k-\tau}^{k}\|y_i - \phi_i\theta_i\|^2\right)^{1/2}. \tag{37}$$

The forgetting factor produced by (36) increases when output measurements differ significantly from the predicted output, which occurs when the parameters change. To reduce sensitivity to noise, the output error is first fed through a moving average given by (37) and forgetting is only activated when the average output error exceeds unity.

Figure 1 shows the performance of RLS-CRF and RLS-VDF, both with $\lambda = 0.99$, and RLS-VRDF with β_k given by (36) with $\gamma = \eta = 1$. Note that, after both parameter changes, RLS-VRDF reconverges to the modified parameters more quickly than RLS-CRF and RLS-VDF. During the loss of persistency between $k = 100$ and $k = 1000$ samples, the covariance of RLS-CRF begins to diverge while the covariances of RLS-VDF and RLS-VRDF remain bounded. ◇

Conclusions and Future Research

Future research will focus on techniques for reducing the computational complexity of the singular value decomposition of P_k. A starting point for this objective is the recursive SVD presented in [29].

Acknowledgments

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

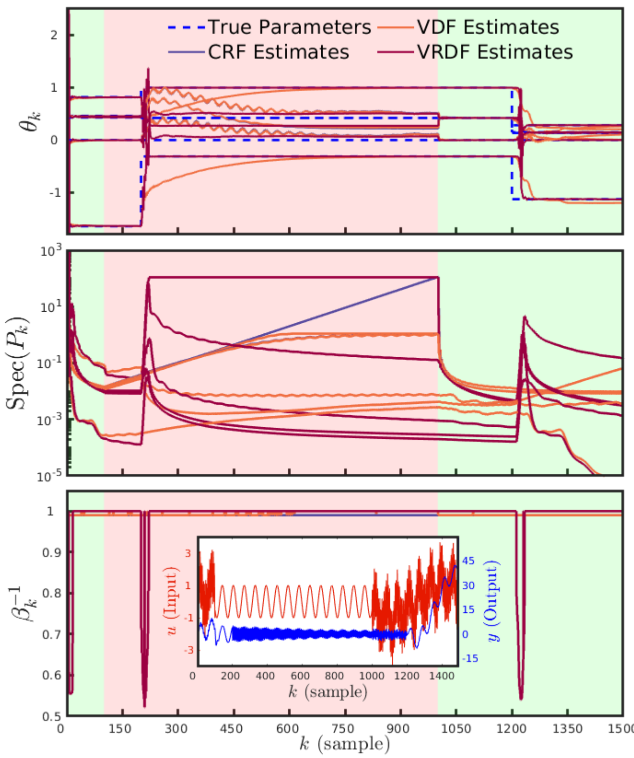

Fig. 1: Parameter estimates θ_k, spectrum of P_k, and spectrum of B_k given by RLS-CRF and RLS-VDF with $\lambda = 0.99$ and by RLS-VRDF with β_k defined by (36). Intervals where the input is persistently exciting are shaded in light green, while the interval where persistency is lost is shaded in light red. The input and output signals are shown in the middle of the bottom figure for reference.

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, 1995.

[3] S. A. U. Islam and D. S. Bernstein, "Recursive Least Squares for Real-Time Implementation," *IEEE Contr. Sys. Mag.*, vol. 39, pp. 82–85, 2019.

[4] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley, 2003.

[5] T. R. Fortescue, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of Self-tuning Regulators with Variable Forgetting Factors," *Automatica*, vol. 17, no. 6, pp. 831–835, 1981.

[6] C. Paleologu, J. Benesty, and C. Silviu, "A Robust Variable Forgetting Factor Recursive Least-Squares Algorithm for System Identification," *IEEE Sig. Proc. Lett.*, vol. 15, pp. 597–600, 2008.

[7] S.-H. Leung and C. F. So, "Gradient-Based Variable Forgetting Factor RLS Algorithm in Time-Varying Environments," *IEEE Trans. Sig. Proc.*, vol. 53, no. 8, pp. 3141–3150, 2005.

[8] 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.*, vol. 36, no. 11, pp. 988–990, 2000.

[9] D. J. Park, B. E. Jun, and K. J. H., "Fast tracking RLS algorithm using novel variable forgetting factor with unity zone," *Electron. Lett.*, vol. 27, no. 23, pp. 2150–2151, 1991.

[10] A. Osorio Cordero and D. Mayne, "Deterministic convergence of a self-tuning regulator with variable forgetting factor," *IEE Proceedings D Control Theory and Applications*, vol. 128, no. 1, p. 19, 1981.

[11] A. Goel and D. S. Bernstein, "A Targeted Forgetting Factor for Recursive Least Squares," *Proc. Conf. Dec. Contr., Miami, FL, December 2018.*, vol. 39, pp. 3899–3903, 2018.

[12] I. M. Mareels and M. Gevers, "Persistency of excitation criteria for linear, multivariable, time-varying systems," *Mathematics of Control, Signals, and Systems*, vol. 1, no. 3, pp. 203–226, 1988.

[13] I. M. Mareels, R. R. Bitmead, M. Gevers, C. R. Johnson, and R. L. Kosut, "How exciting can a signal really be?," *Sys. Contr. Lett.*, vol. 8, no. 3, pp. 197–204, 1987.

[14] I. Y. Mareels and M. Gevers, "Persistence of excitation criteria," in *Proc. Conf. Dec. Contr*, pp. 1933–1935, 1986.

[15] B. D. O. Anderson, "Adaptive systems, lack of persistency of excitation and bursting phenomena," *Automatica*, vol. 21, no. 3, pp. 247–258, 1985.

[16] M. E. Salgado, G. C. Goodwin and R. H. Middleton, "Modified least squares algorithm incorporating exponential resetting and forgetting," *Int. J. Contr.*, vol. 47, no. 2, pp. 477–491, 1988.

[17] A. L. Bruce and D. S. Bernstein, "A Modified Recursive Least Squares Algorithm with Forgetting and Bounded Covariance," *2019 American Control Conference (ACC), Philadelphia, PA, USA, July 10-12*, pp. 5340–5345, 2019.

[18] S. Bittanti, P. Bolzern, and M. Campi, "Exponential convergence of a modified directional forgetting identification algorithm," *Sys. Contr. Lett.*, vol. 14, no. 2, pp. 131–137, 1990.

[19] S. Bittanti, P. Bolzern, and M. Campi, "Convergence and exponential convergence of identification algorithms with directional forgetting factor," *Automatica*, vol. 26, no. 5, pp. 929–932, 1990.

[20] G. Kubin, "Stabilization of the RLS algorithm in the absence of persistent excitation," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1369–1372, 1988.

[21] L. Cao and H. Schwartz, "Directional forgetting algorithm based on the decomposition of the information matrix," *Automatica*, vol. 36, no. 11, pp. 1725–1731, 2000.

[22] G. Kreisselmeier, "Stabilized Leastsquares Type Adaptive Identifiers," *IEEE Transactions on Automatic Control*, vol. 35, pp. 306–310, mar 1990.

[23] R. Kulhavỳ and M. Kárnỳ, "Tracking of slowly varying parameters by directional forgetting," *IFAC Proc.*, vol. 17, no. 2, pp. 687–692, 1984.

[24] R. Kulhavỳ, "Restricted exponential forgetting in real-time identification," *Automatica*, vol. 23, no. 5, pp. 589–600, 1987.

[25] A. L. Bruce, A. Goel, and D. S. Bernstein, "Convergence and Consistency of Recursive Least Squares with Variable-Rate Forgetting," arXiv:2003.02737.

[26] A. Goel, A. L. Bruce, and D. S. Bernstein, "Recursive Least Squares with Variable-Direction Forgetting – Compensating for the loss of persistency," Accepted at Control Systems Magazine. arXiv:1001.1384.

[27] A. S. Poznyak and J. J. Medel Juarez, "Matrix forgetting factor," *Int. J. of Systems Science*, vol. 30, no. 2, pp. 165–174, 1999.

[28] J. Li and Y. Zheng, "Recursive idenntification of time-varying systems: Self-tuning and matrix RLS algorithms," *Systems Conrol Letters*, vol. 66, pp. 104–110, 2014.

[29] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, vol. 415, pp. 20–30, 2006.

Appendix

Lemma 1: (matrix inversion lemma) Let $A \in \mathbb{R}^{n\times n}$, $U \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{p\times p}$, and $V \in \mathbb{R}^{p\times n}$, and assume that A, C, and $A+UCV$ are nonsingular. Then

$$(A{+}UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \tag{38}$$

Lemma 2: Let $A \in \mathbb{R}^{n\times n}$, assume that A is positive definite, let $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$, and define $f : \mathbb{R}^n \to \mathbb{R}$ by

$$f(x) \triangleq x^{\mathrm{T}}Ax + 2b^{\mathrm{T}}x + c. \tag{39}$$

Then the unique minimizer of f is

$$x_{\mathrm{opt}} = -A^{-1}b, \tag{40}$$

and the minimum value of f is

$$f(x_{\mathrm{opt}}) = c - b^{\mathrm{T}}A^{-1}b. \tag{41}$$