2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

A Modified Recursive Least Squares Algorithm with Forgetting and Bounded Covariance

Adam L. Bruce and Dennis S. Bernstein

***Abstract*— Recursive least squares (RLS) is widely used in identification and estimation. An unfortunate weakness of RLS is the divergence of its covariance matrix in cases where the data are not sufficiently persistent. To solve this problem, [1] introduced the exponential forgetting and resetting algorithm (EFRA), whose covariance update equation is modified so that the covariance matrix remains bounded. Unfortunately, EFRA does not include RLS as a special or limiting case, and cannot easily approximate RLS estimates. In this paper, we derive a modified RLS variant of EFRA that includes RLS without forgetting as a limiting case, and that can closely approximate RLS with forgetting. An additional advantage of MRLS relative to EFRA is greater ease in choosing parameters to set the covariance bounds.**

I. INTRODUCTION

The recursive least squares (RLS) algorithm is one of the key fundamental tools of identification, signal processing, estimation, and control [2, Section 2.2], [3, Chapter 13], [4], [5], [6, Chapter 12]. RLS provides a recursive technique for minimizing the least-squares cost function $J(x) = (Ax - b)^{\mathrm{T}}(Ax-b)$, where each row of A provides an additional data point that can be used to update the previous estimate of x. A unique minimizer of A exists if and only if $A^{\mathrm{T}}A$ is positive definite; an equivalent condition is that A is left invertible. Since A may not be left invertible for a limited amount of data, the RLS cost function includes a regularization term.

The simplest approach to deriving RLS is to define a recursion for the quadratic term in x appearing in $J(x)$ in terms of a covariance matrix. The matrix inversion can then be used to arrive at the final update equations. The term "covariance matrix" arises from the relationship between RLS and the Kalman filter [7], [8]. In particular, by defining the state update $x_{k+1} = x_k$, which models the assumption that x is constant, $Ax = b$ can be viewed as the measurement equation $b = Ax$, where A is the "C" matrix for use in the Kalman filter. With these dynamics and measurement equation, the Riccati difference equation with initial condition given by the regularization term yields the RLS covariance update equation. Note that the state update $x_{k+1} = x_k$ does not include disturbance noise, and thus the Riccati covariance update lacks a constant driving term. Consequently, the solution of the Riccati difference equation is monotonically decreasing.

A useful variation of RLS is obtained by modifying the cost function to include a forgetting factor $\lambda \in (0,1)$. By using λ, older data are discounted relative to more recent data. Consequently, RLS can respond more quickly to changes in x. In terms of the Kalman filter, the forgetting factor corresponds to the state equation $x_{k+1} = \frac{1}{\sqrt{\lambda}}x_k$, which, since $\lambda < 1$, are unstable. This instability allows the covariance to increase, which explains the ability of RLS to respond more quickly to changes in x.

An unfortunate side effect of the forgetting factor occurs when the data are not persistently exciting, which is reflected by the situation where $(\frac{1}{\sqrt{\lambda}}I, A)$ is not observable. In this case, sensor noise causes the covariance to diverge, leading to divergence of the estimate of x [9].

In order to overcome covariance divergence, a modified covariance equation is given in [1]. This modified covariance equation includes terms that bound the covariance. Most importantly, the modified covariance equation allows the covariance to increase, thus providing the ability to adapt to changes in x. Consequently, the *exponential forgetting and resetting algorithm* (EFRA) of [1] provides forgetting while preventing divergence. Unfortunately, EFRA does not include RLS as a special or limiting case.

The contribution of the present paper is to derive a modified RLS (MRLS) variation of [1] that includes RLS without forgetting as a limiting case and can closely approximate RLS with forgetting. Like EFRA, MRLS provides forgetting action while bounding the covariance. An additional benefit of MRLS relative to EFRA is greater simplicity in setting the upper and lower covariance bounds.

The contents of the paper are as follows. Section II reviews recursive least squares. Section III introduces EFRA and shows that it does not have an RLS limit. Section IV gives the modified RLS algorithm and its proof. Section V shows that MRLS can approximate RLS. Section VI gives some results which help simplify MRLS coefficient selection. Finally, Sections VII-IX give examples that illustrate the performance of MRLS, RLS, and EFRA in different scenarios.

II. REVIEW OF RECURSIVE LEAST SQUARES

In this section, we briefly review of recursive least squares (RLS) with forgetting factor λ.

Theorem 2.1: For all $k \geq 1$, let $\phi(k) \in \mathbb{R}^{p\times n}$ and $y(k) \in \mathbb{R}^p$. Furthermore, let $\theta_0 \in \mathbb{R}^n$, let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite, let $\lambda \in (0,1]$, and, for all $k \geq 0$, denote

978-1-5386-7928-9/$31.00 ©2019 AACC

the minimizer of the function

$$J_k(\theta) \triangleq \sum_{i=1}^{k} \lambda^{k-i}(y(i)-\phi(i)\theta)^{\mathrm{T}}(y(i)-\phi(i)\theta) + \lambda^k(\theta-\theta_0)^{\mathrm{T}}P_0^{-1}(\theta-\theta_0) \quad (1)$$

by

$$\theta_k \triangleq \mathrm{argmin}_{\theta\in\mathbb{R}^n} J_k(\theta). \quad (2)$$

Then, for all $k \geq 1$, θ_k is given by

$$\theta_k = \theta_{k-1} + P_{k-1}\phi(k)^{\mathrm{T}}(\lambda I + \phi(k)P_{k-1}\phi(k)^{\mathrm{T}})^{-1} \cdot (y(k)-\phi(k)\theta_{k-1}), \quad (3)$$

$$P_k = \frac{1}{\lambda}P_{k-1} - \frac{1}{\lambda}P_{k-1}\phi(k)^{\mathrm{T}} \cdot \left[\lambda I + \phi(k)P_{k-1}\phi(k)^{\mathrm{T}}\right]^{-1}\phi(k)P_{k-1}. \quad (4)$$

It can be seen from (4) that the current covariance matrix is given by the sum of a positive-definite matrix and a negative-semidefinite matrix. In the case where $\lambda = 1$, the positive-definite matrix is the previous covariance matrix, and thus the sequence of covariance matrices is nonincreasing with respect to the positive-semidefinite matrix partial ordering. In the case where $\lambda < 1$, the sequence of covariance matrices is not necessarily nonincreasing. By allowing eigenvalues of the covariance matrix to increase, the effect of the forgetting factor is to discount prior information and facilitate future learning.

III. Exponential Forgetting and Resetting Algorithm

Although the use of the forgetting factor allows eigenvalues of the covariance to increase and thus facilitate learning, an undesirable side effect is that, in the absence of persistent excitation and in the presence of noise, the covariance may diverge (see Section IX). An extension of RLS given by the exponential forgetting and resetting algorithm (EFRA) [1] overcomes this problem.

Theorem 3.1: Let $\alpha \in (0,1)$, $\gamma \in (0,\alpha)$, $\beta > 0$, and $\delta > 0$, and assume that

$$(\alpha-\gamma)^2 + 4\beta\delta < (1-\alpha)^2. \quad (5)$$

Furthermore, define

$$\sigma \triangleq \frac{\alpha-\gamma}{2\delta}\left(\sqrt{1+\frac{4\beta\delta}{(\alpha-\gamma)^2}}-1\right), \quad (6)$$

$$\nu \triangleq \frac{\gamma}{2\delta}\left(1+\sqrt{1+\frac{4\beta\delta}{\gamma^2}}\right), \quad (7)$$

let $P_0 \in \mathbb{R}^{n\cdot n}$ be positive semidefinite, and assume that

$$\sigma I \leq P_0 \leq \nu I. \quad (8)$$

Furthermore, let $\theta_0 \in \mathbb{R}^n$, for all $k \geq 0$, let $y(k) \in \mathbb{R}$ and $\phi(k) \in \mathbb{R}^{1\times n}$, and consider the update equations

$$\theta_{k+1} = \theta_k + \frac{\alpha}{1+\phi(k)P_k\phi(k)^{\mathrm{T}}}P_k\phi(k)^{\mathrm{T}}(y(k)-\phi(k)\theta_k) \quad (9)$$

$$P_{k+1} = (1+\gamma)P_k - \frac{\alpha}{1+\phi(k)P_k\phi(k)^{\mathrm{T}}}P_k\phi(k)\phi(k)^{\mathrm{T}}P_k + \beta I - \delta P_k^2. \quad (10)$$

Then the following statements hold:

i) For all $k \geq 0$, $\sigma I \leq P_k \leq \nu I$.
ii) For all $k \geq 0$, $\beta I + \gamma P_k - \delta P_k^2 \geq 0$.
iii) If there exists $k_0 \geq 0$ such that, for all $k \geq k_0$, $\phi(k) = 0$, then $\lim_{k\to\infty} P_k = \nu I$.

Note that (10) includes the terms βI and δP_k^2, which do not appear in (4). These terms enforce the bounds given by $i)$. Note, in addition, that γ plays the role of $1/\lambda - 1$ in RLS.

Comparing (9), (10) with (3), (4), it can be seen that these results coincide in the case where $\beta = \delta = 0$ and $1/\lambda = 1+\gamma = \alpha$. Since $0 < \alpha < 1$, it follows that $-1 < \gamma < 0$, which contradicts the assumption that $0 < \gamma < \alpha$. Therefore, there is no choice of parameters α, γ, β, and δ for which RLS is a special or limiting case of EFRA. Finally, for each choice of α and γ, it can be shown that the ratio ν/σ cannot be set arbitrarily. For example, let $\alpha = 1/2$ and $\gamma = 1/4$. From (5), it follows that $\beta\delta < 0.046875$. Furthermore, (6) and (7) imply that

$$\frac{\nu}{\sigma} = \frac{\frac{1}{4}+\sqrt{\frac{1}{16}+4\beta\delta}}{\sqrt{\frac{1}{16}+4\beta\delta}-\frac{1}{4}}. \quad (11)$$

Thus, for all $\beta > 0$ and $\delta < 0.046875/\beta$, it follows that $\nu/\sigma > 3$.

IV. Modified RLS

Inspired by EFRA, we now derive modified RLS (MRLS), which, like EFRA, has bounded covariance, but, unlike EFRA, can approximate RLS. Furthermore, attaining an arbitrary choice of covariance bounds is simpler for MRLS than for EFRA.

Theorem 4.1: Let $\beta \in (0,\infty)$, $\delta \in (0,\infty)$, and $\gamma \in [1, \frac{3}{2})$, assume that

$$\gamma + 2\beta\delta < \tfrac{3}{2}, \quad (12)$$

and define

$$\bar{\alpha} \triangleq \frac{2[\sqrt{(\gamma-1)^2+4\beta\delta}(2-\gamma-\sqrt{(\gamma-1)^2+4\beta\delta})+\gamma-1]}{1-(2-\gamma-\sqrt{(\gamma-1)^2+4\beta\delta})^2}. \quad (13)$$

Then, $\bar{\alpha} \in (0,1)$. Next, let $\alpha \in (0,\bar{\alpha})$, and define

$$\sigma(\alpha) \triangleq \frac{\gamma-1-\alpha+\sqrt{(\gamma-1-\alpha)^2+4\beta\delta}}{2\delta}. \quad (14)$$

Then,

$$0 < \sigma(\alpha) < (1-\alpha)\sigma(0) < \sigma(0). \tag{15}$$

Furthermore, let $\theta_0 \in \mathbb{R}^n$ and $P_0 \in \mathbb{R}^{n\times n}$, and assume that P_0 is positive definite and satisfies

$$\sigma(\alpha)I \le P_0 \le \sigma(0)I. \tag{16}$$

For all $k \ge 0$, let $\varepsilon_k \in (0,\infty)$, $\eta_k \in (0,\infty)$, $y(k) \in \mathbb{R}^p$, and $\phi(k) \in \mathbb{R}^{p\times n}$, and consider the update equations

$$\theta_{k+1} = \theta_k + \eta_k P_k \phi(k)^{\mathrm{T}}[\varepsilon_k I + \phi(k)P_k\phi(k)^{\mathrm{T}}]^{-1} \cdot (y(k) - \phi(k)\theta_k), \tag{17}$$

$$P_{k+1} = \gamma P_k - \alpha P_k\phi(k)^{\mathrm{T}} \cdot [\varepsilon_k I + \phi(k)P_k\phi(k)^{\mathrm{T}}]^{-1}\phi(k)P_k + \beta I - \delta P_k^2. \tag{18}$$

Then, for all $k \ge 1$,

$$\sigma(\alpha)I \le P_k \le \sigma(0)I. \tag{19}$$

Proof. To prove that $\bar\alpha \in (0,1)$, note that (12) implies that

$$\begin{aligned}(1-\gamma)^2 + 4\beta\delta &= 1 - 2\gamma + \gamma^2 + 4\beta\delta \\ &< 1 - 2\gamma + \gamma^2 + 3 - 2\gamma \\ &= (2-\gamma)^2.\end{aligned}$$

Hence, the numerator of (13) is positive. Furthermore, since $1 \le \gamma < \frac{3}{2}$, it follows that

$$\begin{aligned}1 &\le \sqrt{1+2(\gamma-1)(2-\gamma)} = \sqrt{(2-\gamma)^2 + (1-\gamma)^2} \\ &< 2 - \gamma + \sqrt{(1-\gamma)^2 + 4\beta\delta}.\end{aligned}$$

Since $2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta} > 0$ and $2-\gamma+\sqrt{(1-\gamma)^2+4\beta\delta} > 1$, it follows that

$$\begin{aligned}&2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta} \\ &< (2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta})(2-\gamma+\sqrt{(1-\gamma)^2+4\beta\delta}) \\ &= (2-\gamma)^2 - (1-\gamma)^2 - 4\beta\delta \\ &= 1 - (2\gamma - 2 + 4\beta\delta) \\ &< 1.\end{aligned}$$

Hence, the denominator of (13) is positive and therefore, $\bar\alpha > 0$. Furthermore,

$$\begin{aligned}&2[\sqrt{(1-\gamma)^2+4\beta\delta}(2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta}) + \gamma - 1] \\ &= -(2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta})^2 \\ &\quad + (2-\gamma)^2 - (1-\gamma)^2 - 4\beta\delta + \gamma - 1 \\ &= 1 - 2(\gamma-1) + (\gamma-1)^2 \\ &\quad - (\gamma-1)^2 - 4\beta\delta + \gamma - 1 - (2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta})^2 \\ &= 1 - (\gamma-1) - 4\beta\delta - (2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta})^2 \\ &< 1 - (2-\gamma-\sqrt{(1-\gamma)^2+4\beta\delta})^2.\end{aligned}$$

Hence, $\bar\alpha < 1$. Therefore, $\bar\alpha \in (0,1)$.

To show that $\sigma(\alpha) > 0$, note that since $\beta > 0$ and $\delta > 0$ it follows that

$$|\gamma - \alpha - 1| < \sqrt{(\gamma-\alpha-1)^2 + 4\beta\delta}, \tag{20}$$

and thus $\gamma - \alpha - 1 + \sqrt{(\gamma-\alpha-1)^2+4\beta\delta} > 0$. Since $\alpha > 0$ it follows that $\gamma - \alpha - 1 + \sqrt{(\gamma-\alpha-1)^2+4\beta\delta} < \gamma - 1 + \sqrt{(\gamma-1)^2+4\beta\delta}$, hence $\sigma(\alpha) < \sigma(0)$.

To show that $\sigma(\alpha) < (1-\alpha)\sigma(0)$, define the positive numbers

$$b \triangleq \gamma - 1, \quad c \triangleq 4\beta\delta, \tag{21}$$

$$f \triangleq \sqrt{b^2+c} = \sqrt{(\gamma-1)^2+4\beta\delta}, \tag{22}$$

$$g \triangleq 1 - b = 2 - \gamma, \tag{23}$$

and note that

$$\bar\alpha = \frac{2[f(g-f)+b]}{1-(g-f)^2}. \tag{24}$$

Furthermore, note that, since $\alpha < \bar\alpha$ and $g - f < 1$, it follows that $0 < 2[f(g-f)+b] - \alpha[1-(g-f)^2]$. Hence,

$$\begin{aligned}0 &< 2\alpha[f(g-f)+b] - \alpha^2[1-(g-f)^2] \\ &= \alpha^2g^2 - 2\alpha f^2 + \alpha^2 f^2 + 2\alpha g f - 2\alpha^2 g f - \alpha^2 + 2b\alpha \\ &= \alpha^2g^2 - 2\alpha f^2 + \alpha^2 f^2 + 2\alpha g f - 2\alpha^2 g f \\ &\quad - \alpha^2 + 2b\alpha + f^2 - f^2 \\ &= \alpha^2 g^2 + (1-\alpha)^2 f^2 + 2\alpha(1-\alpha) g f - \alpha^2 + 2b\alpha - f^2 \\ &= [\alpha g + (1-\alpha) f]^2 - [\alpha^2 - 2b\alpha + f^2] \\ &= [\alpha g + (1-\alpha) f]^2 - [(b-\alpha)^2 + c] \\ &= [\alpha g + (1-\alpha) f - \sqrt{(b-\alpha)^2+c}] \\ &\quad \cdot [\alpha g + (1-\alpha) f + \sqrt{(b-\alpha)^2+c}].\end{aligned}$$

Since $\alpha < 1$, it follows that $\alpha g + (1-\alpha) f + \sqrt{(b-\alpha)^2+c} > 0$. Therefore,

$$\begin{aligned}0 &< \alpha g + (1-\alpha) f - \sqrt{(b-\alpha)^2+c} \\ &= \alpha(1-b) + (1-\alpha)\sqrt{b^2+c} - \sqrt{(b-\alpha)^2+c} + b - b \\ &= (1-\alpha)b + (1-\alpha)\sqrt{b^2+c} - [b - \alpha + \sqrt{(b-\alpha)^2+c}] \\ &= (1-\alpha)[b + \sqrt{b^2+c}] - [b - \alpha + \sqrt{(b-\alpha)^2+c}],\end{aligned}$$

Hence,

$$\begin{aligned}\sigma(\alpha) &= \frac{b-\alpha+\sqrt{(b-\alpha)^2+c}}{2\delta} \\ &< (1-\alpha)\frac{b+\sqrt{b^2+c}}{2\delta} = (1-\alpha)\sigma(0).\end{aligned}$$

Next, let $k \ge 0$, suppose that P_k is positive semidefinite, and define

$$M_k = \begin{bmatrix} P_k & P_k\phi(k)^{\mathrm{T}} \\ \phi(k)P_k & \phi(k)P_k\phi(k)^{\mathrm{T}} \end{bmatrix}. \tag{25}$$

Since P_k is positive semidefinite, M_k can be written as

$$M_k = \begin{bmatrix} P_k^{1/2} \\ \phi(k)P_k^{1/2} \end{bmatrix} \begin{bmatrix} P_k^{1/2} & P_k^{1/2}\phi(k)^{\mathrm{T}} \end{bmatrix}, \tag{26}$$

and thus M_k is positive semidefinite. Since $\varepsilon_k > 0$, it follows that

$$N_k = M_k + \begin{bmatrix} 0 & 0 \\ 0 & \varepsilon_k I \end{bmatrix} \tag{27}$$

is also positive semidefinite. Therefore, since $\phi(k)P_k\phi(k)^{\mathrm{T}}+\varepsilon_k I$ is positive definite, it follows that the Schur complement of N_k is positive semidefinite, and thus

$$P_k\phi(k)^{\mathrm{T}}[\varepsilon_k+\phi(k)P_k\phi(k)^{\mathrm{T}}]^{-1}\phi(k)P_k \le P_k. \tag{28}$$

Next, we show that, for all $k\ge 1$, $P_k\le\sigma(0)I$. Let $k\ge 1$. By (16), $P_0\le\sigma(0)I$. Hence, suppose that $P_{k-1}\le\sigma(0)I$. Now, define

$$a \stackrel{\triangle}{=} 1-\sqrt{(1-\gamma)^2+4\beta\delta}, \tag{29}$$

and note that

$$\begin{aligned} P_k-\sigma(0)I &= a(P_{k-1}-\sigma(0)I)\\ &\quad -\delta(P_{k-1}-\sigma(0)I)^2-\alpha G_{k-1}. \end{aligned} \tag{30}$$

From (12) it follows that

$$(\gamma-1)^2+4\beta\delta=\gamma^2-4\gamma+1+2\gamma+4\beta\delta<(\gamma-2)^2<1,$$

and thus $\sqrt{(\gamma-1)^2+4\beta\delta}<1$, which implies that $a>0$. Since a is positive, it follows from $P_{k-1}\le\sigma(0)I$ and (30) that $P_k-\sigma(0)I\le 0$. Hence, $P_k\le\sigma(0)I$.

Next, we show that, for all $k\ge 1$, $P_k\ge\sigma(\alpha)I$. Let $k\ge 1$. By assumption, $P_0\ge\sigma(\alpha)I$. Hence, suppose that $P_{k-1}\ge\sigma(\alpha)I$. Now, define

$$\begin{aligned} G_{k-1} &\stackrel{\triangle}{=} P_{k-1}\phi(k-1)^{\mathrm{T}}[\varepsilon_{k-1}I+\phi(k-1)P_{k-1}\phi(k-1)^{\mathrm{T}}]^{-1}\\ &\quad\cdot\phi(k-1)P_{k-1}, \end{aligned} \tag{31}$$

$$f(\mu) \stackrel{\triangle}{=} (\gamma-\alpha)\mu+\beta-\delta\mu^2. \tag{32}$$

Since P_{k-1} is positive definite, it follows that $P_{k-1}\ge G_{k-1}$, and thus

$$P_k=\gamma P_{k-1}-\alpha G_{k-1}+\beta I-\delta P_{k-1}^2, \tag{33}$$

$$\ge(\gamma-\alpha)P_{k-1}+\beta I-\delta P_{k-1}^2. \tag{34}$$

Since $\sigma(\alpha)I\le P_{k-1}\le\sigma(0)I$, it follows that $\mathrm{spec}(P_{k-1})\subset[\sigma(\alpha),\sigma(0)]$, and thus

$$\min_{[\sigma(\alpha),\sigma(0)]} f(\mu)\le\min_{\mathrm{spec}(P_{k-1})} f(\mu)\le\lambda_{\min}(P_k), \tag{35}$$

where the minimum in (35) exists because f is continuous and $[\sigma(\alpha),\sigma(0)]$ is compact. Since f is concave, its unique stationary point is its maximizer, and thus the minimizer of f over $[\sigma(\alpha),\sigma(0)]$ is either $\sigma(\alpha)$ or $\sigma(0)$. Since $f(\sigma(0))=\sigma(0)$, $f(\sigma(\alpha))=(1-\alpha)\sigma(\alpha)$, and $\sigma(\alpha)<(1-\alpha)\sigma(0)$, it follows that $f(\sigma(0))<f(\sigma(\alpha))$ and thus $\sigma(\alpha)$ is the unique minimizer of f over $[\sigma(\alpha),\sigma(0)]$. Therefore, $\lambda_{\min}(P_k)\ge\sigma(\alpha)$, and hence $P_k\ge\sigma(\alpha)I$. □

V. Approximation of RLS by MRLS

In this section, we show that, in the case where $\lambda=1$, MRLS approximates RLS as a limiting case. Although the same statement cannot be made in the case where $\lambda<1$, we show that the MRLS estimates approximate the RLS estimates in the case where $\lambda\approx 1$.

First, we show that RLS is a limiting case of MRLS in the case where $\lambda=1$. To see this, note that, for all $\gamma\in(1,\frac{3}{2})$,

$$\lim_{\beta\downarrow 0}\bar\alpha=\frac{\gamma^2-3\gamma+2}{\gamma^2-3\gamma+2}=1. \tag{36}$$

Hence, by L'Hôpital's rule,

$$\lim_{\gamma\downarrow 1}\lim_{\beta\downarrow 0}\bar\alpha=\lim_{\gamma\downarrow 1}\frac{\gamma^2-3\gamma+2}{\gamma^2-3\gamma+2}=1. \tag{37}$$

Hence, letting $\beta\to 0$, $\gamma\to 1$, $\delta\to 0$, $\alpha\to 1$, $\varepsilon_k=1$, $\eta_k=1$, and setting $\lambda=\lim_{\gamma\downarrow 1}1/\gamma=1$, it follows that (17), (18) approximate (3), (4).

Next, let $\gamma\in(1,\frac{3}{2})$. Letting $\beta\to 0$, $\delta\to 0$, $\alpha\to 1$, $\varepsilon_k=1$, $\eta_k=1/\gamma$, and setting $\lambda=\lim_{\gamma\downarrow 1}1/\gamma$, it follows that (17), (18) are given by

$$\begin{aligned} \theta_{k+1} &= \theta_k+P_k\phi(k)^{\mathrm{T}}[\lambda I+\phi(k)P_k\phi(k)^{\mathrm{T}}]^{-1}\\ &\quad\cdot(y(k)-\phi(k)\theta_k), \end{aligned} \tag{38}$$

$$\begin{aligned} P_{k+1} &= \frac{1}{\lambda}P_k-P_k\phi(k)^{\mathrm{T}}\\ &\quad\cdot[\lambda I+\phi(k)P_k\phi(k)^{\mathrm{T}}]^{-1}\phi(k)P_k. \end{aligned} \tag{39}$$

Comparing (39) and (4), we see the only difference is in the second term. For $\lambda\approx 1$, this difference is small, and the examples in Section VIII show that (38), (39) numerically approximate (3), (4).

VI. Coefficient Selection for MRLS

In this section we give a heuristic procedure for choosing MRLS parameters that yield an approximate forgetting factor and specified covariance bounds. The procedure is easy to apply but is not guaranteed to be successful in every case.

Let λ, x, and y be the desired forgetting factor, upper covariance bound, and lower covariance bound, respectively. Next, consider the following steps:

i) Let $\gamma=\frac{1}{\lambda}$.

ii) Choose $\alpha\in(0,1)$ such that

$$\frac{x}{y}<\frac{\gamma-1}{\gamma-\alpha-1}. \tag{40}$$

iii) Compute

$$\delta \stackrel{\triangle}{=} \frac{(\gamma-1)(x-y)+\alpha y}{x^2-y^2}, \tag{41}$$

$$\beta \stackrel{\triangle}{=} \frac{xy[\alpha x-(\gamma-1)(x-y)]}{x^2-y^2}. \tag{42}$$

iv) Compute $\bar\alpha$ using (13).

v) If $\alpha<\bar\alpha$, then the coefficients γ, α, δ, β yield an MRLS filter with $\sigma(0)=x$ and $\sigma(\alpha)=y$, where $\sigma(\alpha)$ is defined by (14). If $\alpha>\bar\alpha$, then decrease α and repeat steps $ii)-v)$.

vi) Choose ε_k and η_k.

To obtain an estimator that mimics RLS but has bounded covariance, choose $\varepsilon_k=\lambda$ and $\eta_k=1$. If β is small then

$\bar{\alpha} \approx 1$. In this case, let $\alpha \approx 1$ to most effectively approximate RLS.

The following examples illustrate the behavior of MRLS and its relationship to EFRA and RLS.

VII. Example 1: MRLS Approximation of RLS

Consider the system

$$G(\mathbf{q}) = \frac{\mathbf{q} - \frac{1}{12}}{\mathbf{q}^2 - \frac{1}{12}\mathbf{q} - \frac{1}{12}}. \tag{43}$$

with input

$$u(k) = 1 + \sin\frac{2\pi k}{10} + \sin\frac{2\pi k}{20} + \sin\frac{2\pi k}{100}. \tag{44}$$

Suppose that the measurement of the output is corrupted by noise with standard deviation $\sigma = 0.01$. Let $\lambda = 0.999$, $P_0 = 100$, and $\theta_0 = 0$. Consider the MRLS parameters

$$\begin{aligned} &\alpha = 0.99998, \quad \gamma = 1.001, \quad \delta = 10^{-7}, \\ &\beta = 0.001, \quad \eta = 1, \quad \varepsilon = 0.999. \end{aligned} \tag{45}$$

The covariance bounds corresponding to these parameters are $\sigma(\alpha) = 0.001$ and $\sigma(0) = 1.011 \times 10^4$, which closely approximate RLS. Figure 1 shows the similarity between the responses for RLS and MRLS.

Now consider the MRLS parameters

$$\begin{aligned} &\alpha = 0.999998, \quad \gamma = 1.001, \quad \delta = 1, \\ &\beta = 0.001, \quad \eta = 1, \quad \varepsilon = 0.999. \end{aligned} \tag{46}$$

The covariance bounds corresponding to these parameters are $\sigma(\alpha) = 0.001$ and $\sigma(0) = 0.0321$. With these parameters, Figure 1 shows the difference between the responses for RLS and MRLS.

VIII. Example 2: Sudden Change of Parameters

Consider the system described by the time-dependent transfer function

$$G_k(\mathbf{q}) = \begin{cases} \frac{\mathbf{q}+0.2}{\mathbf{q}^2-0.6\mathbf{q}+0.08}, & k \le 50000, \\ \frac{2\mathbf{q}+0.5}{\mathbf{q}^2+0.4\mathbf{q}-0.05}, & k > 50000, \end{cases} \tag{47}$$

which has a sudden change of parameters at $k = 50,000$. The driving signal is given by

$$u(k) = 1 + \sin\frac{2\pi k}{10} + \sin\frac{2\pi k}{20} + \sin\frac{2\pi k}{100}. \tag{48}$$

Suppose that the output measurement is noise free. Let $\lambda = 0.999$, $P_0 = 100$, and $\theta_0 = 0$. Consider the EFRA parameters

$$\alpha = 0.375, \quad \gamma = 0.001, \quad \delta = 0.05, \quad \beta = 1.2525, \tag{49}$$

and the MRLS parameters

$$\begin{aligned} &\alpha = 0.991, \quad \gamma = 1.001, \quad \delta = 0.00001, \\ &\beta = 0.001, \quad \eta = 1, \quad \varepsilon = 0.999. \end{aligned} \tag{50}$$

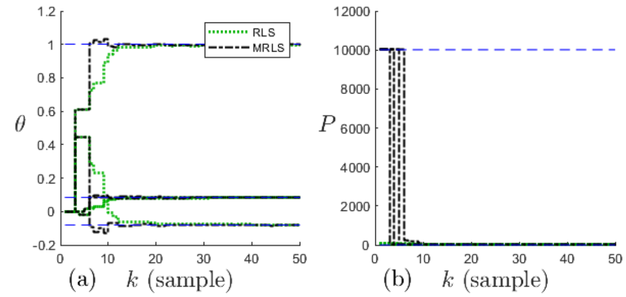

Fig. 1. Example 1: MRLS with coefficients that closely approximate RLS. (a) shows the parameter estimates and their true values; (b) shows the eigenvalues of the covariance matrix with the MRLS covariance bounds.

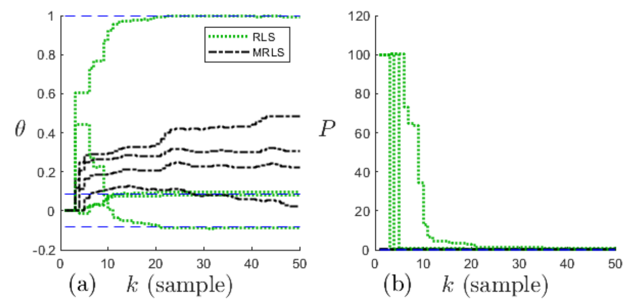

Fig. 2. Example 1: MRLS with coefficients that differ significantly from RLS. (a) shows the parameter estimates and their true values; (b) shows the eigenvalues of the covariance matrix with the MRLS covariance bounds.

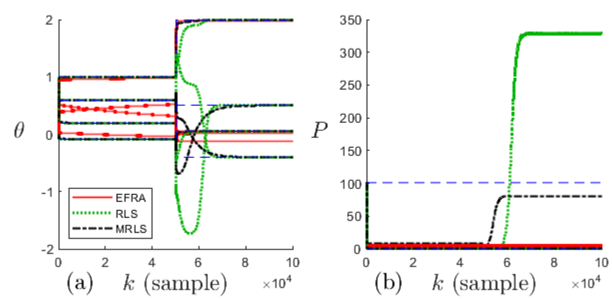

Fig. 3. Example 2: Sudden change of parameters. (a) shows the parameter estimates and their true values; (b) shows the eigenvalues of the covariance matrix with the EFRA and MRLS covariance bounds.

Figure 3 shows the parameter estimates and covariance eigenvalues for RLS, EFRA, and MRLS, along with the EFRA and MRLS covariance bounds. The choice of constraint on the covariance prevents convergence of EFRA to the true value both before and after the parameter change. In contrast, RLS and MRLS quickly converge to the true parameter values, and then reconverge to the new values after approximately $20,000$ steps.

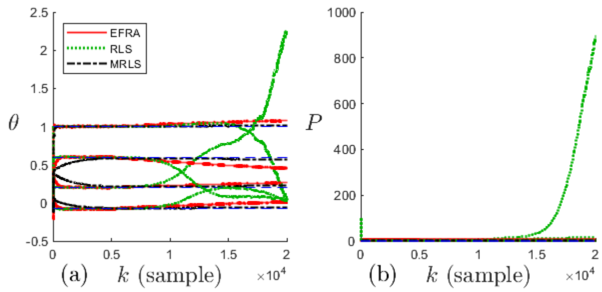

Fig. 4. Example 3: Loss of persistency at $k = 10,000$. (a) shows the parameter estimates with the true values shown by blue dashed lines; (b) shows the eigenvalues of the covariance matrix with EFRA covariance bounds (red dashed) and MRLS covariance bounds (blue dashed).

IX. Example 3: Sudden Loss of Persistency with Sensor Noise

Consider the system

$$G(\mathbf{q}) = \frac{\mathbf{q} + 0.2}{\mathbf{q}^2 - 0.6\mathbf{q} + 0.08} \tag{51}$$

with the driving signal

$$u(k) = \begin{cases} \nu(k) & k \le 5000, \\ \sin\frac{k}{10} & k > 5000, \end{cases} \tag{52}$$

where, for all $k \ge 1$, $\nu(k) \sim \mathcal{N}(0,1)$. Note that u suddenly loses persistency at $k = 5000$. Suppose that the measurement of the output is corrupted by noise with standard deviation $\sigma = 0.01$. Let $\lambda = 0.999$, $P_0 = 100$, and $\theta_0 = 0$. Consider the EFRA parameters

$$\alpha = 0.375, \quad \gamma = 0.001, \quad \delta = 0.05, \quad \beta = 1.2525, \tag{53}$$

and the MRLS parameters

$$\alpha = 0.991, \quad \gamma = 1.001, \quad \delta = 1, \tag{54}$$

$$\beta = 0.001, \quad \eta = 1, \quad \varepsilon = 0.999. \tag{55}$$

Figure 4 shows the parameter estimates and covariance eigenvalues for RLS, EFRA, and MRLS, along with the EFRA and MRLS covariance bounds. When the driving signal loses persistency at $k = 5000$, the RLS covariance diverges, and the RLS parameter estimates diverge from the true parameter values. Since EFRA and MRLS both have bounded covariance, neither can diverge. The MRLS covariance is bounded to be close to zero, causing a lag in the initial estimate convergence, but also maintaing the estimates close to the converged values, even when persistency is lost.

X. Conclusions

In this paper we derived a modified recursive least squares (MRLS) algorithm with forgetting and bounded covariance. Unlike EFRA [1], it is possible to select the parameters of MRLS such that MRLS approximates RLS as a limiting case in the absence of forgetting, and approximately in the case of forgetting. In addition, the simpler constraints of MRLS enable a straightforward process for choosing the MRLS parameters that yield specified covariance bounds.

Examples were given to compare the performance of RLS, MRLS, and EFRA. These examples showed that, by suitably choosing the covariance bounds, it is possible to use MRLS effectively in various situations, including sudden changes in parameters and lack of persistency.

The derivation of MRLS raises questions for future research. The highest priority is to investigate whether or not there is a bounded covariance algorithm that exactly yields RLS with forgetting as a limiting case. Another question is whether or not analytical bounds can be found for the error between RLS with forgetting and the closest MRLS approximation to RLS. The numerical examples in Section VII suggest that these bounds are tight.

XI. Acknowledgments

This research was supported This research was supported by AFOSR under DDDAS grant FA9550-16-1-0071 (Dynamic Data-Driven Applications Systems http://www.1dddas.org/), by the Office of Naval Research under grant N00014-18-1-2211 of the BRC Program, and by the National Science Foundation under grant CMMI-1634709.

References

[1] 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.
[2] K. J. Astrom, *Adaptive Control*. Reading, MA: Addison-Wesley, 2 ed., 1995.
[3] S. Haykin, *Adaptive Filter Theory*. Prentice-Hall, third ed., 1996.
[4] C. R. Johnson, *Lectures on Adaptive Parameter Estimation*. Prentice-Hall, 1988.
[5] B. Lindoff and J. Holst, "Convergence analysis of the RLS identification algorithm with exponential forgetting in stationary ARX-structures," *Int. J. Adapt. Contr. Sig. Proc.*, vol. 13, pp. 1–22, 1999.
[6] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley, 2003.
[7] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Sig. Proc. Mag.*, vol. 13, pp. 18–60, 1994.
[8] L. Cao. and H. M. Schwartz, "Exponential convergence of the kalman filter based parameter estimation algorithm," *Int. J. Adapt. Control Signal Process.*, vol. 17, pp. 763–783, 2003.
[9] S. Dasgupta and Y.-F. Huang, "Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise," *IEEE Trans. Info. Theory*, vol. 33, no. 3, pp. 383–392, 1987.