2022 American Control Conference
Atlanta, USA, June 8-10, 2022

Recursive Least Squares with Variable-Rate Forgetting Based on the F-Test

Nima Mohseni[1], Dennis S. Bernstein[1]

***Abstract*— A variable-rate forgetting factor for recursive least squares is developed for parameter identification of time-varying systems. The variable-rate forgetting factor uses the *F*-test to compare short- and long-term variances of the one-step prediction errors of recursive least squares (RLS). If the short-term error variance is statistically larger than the long-term error variance, then it is assumed that the underlying parameters have changed and forgetting is required. The level of forgetting is proportional to how far the ratio of the short-term and long-term error variances deviate from the expected ratio given by the *F*-distribution. RLS with *F*-test variable-rate forgetting (RLS/FTVRF) is shown to generalize an existing variable-rate forgetting factor that uses a ratio of the root-mean-square (RMS) performance error and noise standard deviation. The approach is applied to a parameter identification task and is compared to a constant-rate forgetting factor and the RMS performance error and noise standard-deviation-based forgetting factor.**

I. Introduction

Recursive least squares (RLS) is widely used for parameter estimation and adaptive control [1], [2]. To track time-varying parameters, RLS can incorporate a forgetting factor λ, which discounts past data. Choosing an appropriate λ is typically done through trial and error or, when the identification is performed offline, maximum likelihood methods. Typical values of λ are between 0.98 and 1 [3], [4].

When a parameter change occurs and forgetting is not enabled, RLS converges slowly to the new parameter values. On the other hand, the use of forgetting when parameters do not change and the data is not persistently exciting can lead to divergence of the singular values of the RLS covariance matrix [5]. In the context of adaptive control, instability of RLS leads to instability of the controller and catastrophic blow-up.

In contrast with constant-rate forgetting, variable-rate forgetting (VRF) allows the forgetting factor to change during operation. VRF versions of RLS are given in [6], [7], [8], These formulations were extended in [9] to include criteria for setting the level of forgetting at each step while maintaining convergence and consistency. Increasing interest in VRF is reflected in [10], [11], [12].

As the above discussion suggests, a key aspect of VRF is the criterion used to vary the forgetting level. In the presence of sensor noise, VRF may set the forgetting factor to be lower than necessary, thereby leading to loss of information when such loss is not warranted. Within the context of adaptive control, a deadzone variation of the approach of [9] was thus proposed in [13]. The formulation of [13] entails the ratio of root-mean-square (RMS) performance error to noise standard deviation.

Inspired by the statistical analysis of the ratio of sample variances, the present paper proposes a variable-rate forgetting method using the F-test. The F-test is typically used to compare whether two sample variances are statistically equal in analysis of variance tests [14], [15]. The ratio of sample variances taken from normally distributed random variables follows an F-distribution, and if the ratio exceeds or is below what is expected for a given significance level, then the two variances are determined to not be equal. The main idea is to use the F-test to determine the level of forgetting to use in RLS based on the the ratio of the variance of two sliding windows of prediction errors of differing length. When the variance of recent predictions increases relative to earlier predictions, we expect that the parameters have changed, and forgetting is warranted relative to the increase in variance. The goal is to prevent forgetting when sufficiently exciting data is not available while also enabling forgetting during parameter changes to allow RLS to quickly learn new parameters.

Section II gives an overview of the RLS-VRF algorithm given in [9]. In Section III we describe the proposed variable-rate forgetting method using the F-test. Section IV shows that RLS with the RMS performance error VRF (RLS/RMSVRF) used in [13] and [16] is a special case of the proposed method. Section V shows an example of the proposed method in an identification task under noisy measurements and nonpersistently exciting inputs in comparison to a constant-rate forgetting factor and RLS/RMSVRF.

II. Recursive Least Squares with Variable-Rate Forgetting

Define the forgetting factor $\lambda \in (0,1]$, initial parameter vector $\theta_0 \in \mathbb{R}^{n\times 1}$, initial positive-definite parameter covariance $P_0 \in \mathbb{R}^{n\times n}$, the regressor $\phi_k \in \mathbb{R}^{p\times n}$, measurement $y_k \in \mathbb{R}^{p\times 1}$, and the prediction error $e_k(\theta) \triangleq y_k - \phi_k\theta$. The cost function for recursive least squares (RLS) is then given by

$$J_k(\theta) \triangleq \sum_{i=0}^{k} \lambda^{k-i} e_i^{\mathrm{T}}(\theta) e_i(\theta) + \lambda^{k+1}(\theta-\theta_0)^{\mathrm{T}} P_0^{-1}(\theta-\theta_0). \tag{1}$$

To generalize (1) such that λ can vary as a function of k,

[1]Department of Aerospace Engineering, The University of Michigan, Ann Arbor, MI 48109, USA

let $\beta_k > 0$ and define

$$\rho_k \triangleq \prod_{i=0}^{k} \beta_i. \tag{2}$$

As shown in [9], unique global minimizer of the cost function

$$J_k(\theta) \triangleq \sum_{i=0}^{k} \frac{\rho_i}{\rho_k} e_i^{\mathrm{T}}(\theta) e_i(\theta) + \frac{1}{\rho_k}(\theta - \theta_0)^{\mathrm{T}} P_0^{-1}(\theta - \theta_0) \tag{3}$$

is given by

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k) \tag{4}$$

where

$$P_{k+1} = L_k - L_k\phi_k^{\mathrm{T}}(I_p + \phi_k L_k \phi_k^{\mathrm{T}})^{-1}\phi_k L_k \tag{5}$$

$$L_k \triangleq \beta_k P_k \tag{6}$$

The variable rate forgetting factor is then defined as $\lambda_k \triangleq \beta_k^{-1}$. Equations (4)-(6) are the recursive least squares with variable rate forgetting (RLS-VRF).

III. Variable Rate Forgetting Using the F-Test

When using RLS, we would like to use forgetting to place a higher weight on more recent data while also suspending forgetting when new data is not available. To accomplish this, let β_k be given by

$$\beta_k \triangleq 1 + \eta g(e_0(\theta_0), \ldots, e_k(\theta_k))\mathbf{1}[g(e_0(\theta_0), \ldots, e_k(\theta_k))] \tag{7}$$

where $\mathbf{1}\colon \mathbb{R} \to \{0, 1\}$ is the unit step function, $\eta > 0$, and $g(e_0(\theta_0), \ldots, e_k(\theta_k))$ is a function of past RLS prediction errors. From (7) it follows that, if $g(e_0(\theta_0), \ldots, e_k(\theta_k)) \leq 0$, then forgetting is suspended, otherwise the level of forgetting is proportional to the magnitude of $g(e_0(\theta_0), \ldots, e_k(\theta_k))$ scaled by η. The objective is to determine an appropriate $g(e_0(\theta_0), \ldots, e_k(\theta_k))$. We expect that, if the true θ has changed relative to the current estimate of θ, then the variance of the prediction errors will increase [17]. When this occurs, we wish to forget older data in order to adjust the parameter estimate quickly. Therefore, we wish to compare the variance of a long and short window of past RLS prediction errors to determine whether or not the variance has increased, and, if so, enable forgetting. For $p = 1$, the F-test will facilitate this comparison.

Given two sample variances $\sigma_{\tau_{\mathrm{n}}}^2$ from $\tau_{\mathrm{n}} + 1$ samples and $\sigma_{\tau_{\mathrm{d}}}^2$ from $\tau_{\mathrm{d}} + 1$ samples, where $\tau_{\mathrm{d}} > \tau_{\mathrm{n}} \geq p$, and $\sigma_{\tau_{\mathrm{n}}}^2 \geq \sigma_{\tau_{\mathrm{d}}}^2$, the variance $\sigma_{\tau_{\mathrm{n}}}^2$ is greater than $\sigma_{\tau_{\mathrm{d}}}^2$ with significance level α if

$$F_{\tau_{\mathrm{n}},\tau_{\mathrm{d}}}^{-1}(1-\alpha) < \frac{\sigma_{\tau_{\mathrm{n}}}^2}{\sigma_{\tau_{\mathrm{d}}}^2}, \tag{8}$$

where $F_{\tau_{\mathrm{n}},\tau_{\mathrm{d}}}^{-1}(x)$ is the inverse cumulative distribution function of the F-distribution with degrees of freedom τ_{n} and τ_{d} [15]. The larger the variance ratio is from $F_{\tau_{\mathrm{n}},\tau_{\mathrm{d}}}^{-1}(1-\alpha)$, the stronger the evidence that $\sigma_{\tau_{\mathrm{n}}}^2$ is greater than $\sigma_{\tau_{\mathrm{d}}}^2$. The F-test can also be written as

$$\sqrt{F_{\tau_{\mathrm{n}},\tau_{\mathrm{d}}}^{-1}(1-\alpha)} < \sqrt{\frac{\sigma_{\tau_{\mathrm{n}}}^2}{\sigma_{\tau_{\mathrm{d}}}^2}}. \tag{9}$$

This leads to a $g(e_0(\theta_0), \ldots, e_k(\theta_k))$ for the case $p = 1$. Given sample variances of the past RLS prediction errors $\sigma_{\tau_{\mathrm{n}}}^2(e_{k-\tau_n}(\theta_{k-\tau_n}), \ldots, e_k(\theta_k))$ and $\sigma_{\tau_{\mathrm{d}}}^2(e_{k-\tau_d}(\theta_{k-\tau_d}), \ldots, e_k(\theta_k))$, for $p = 1$, the function $g(e_0(\theta_0), \ldots, e_k(\theta_k))$ is defined as

$$g \triangleq \sqrt{\frac{\sigma_{\tau_{\mathrm{n}}}^2}{\sigma_{\tau_{\mathrm{d}}}^2}} - \sqrt{F_{\tau_{\mathrm{n}},\tau_{\mathrm{d}}}^{-1}(1-\alpha)}, \tag{10}$$

where the error terms $e_k(\theta_k)$ are dropped for notational convenience. Using (10), forgetting is enabled when $\sigma_{\tau_{\mathrm{n}}}^2$ is statistically larger than $\sigma_{\tau_{\mathrm{d}}}^2$. The magnitude of the forgetting factor λ_k is inversely proportional to the difference between the square roots of the variance ratio and $F_{\tau_{\mathrm{n}},\tau_{\mathrm{d}}}^{-1}(1-\alpha)$, thereby increasing the level of forgetting when there is more evidence that $\sigma_{\tau_{\mathrm{n}}}^2$ is larger $\sigma_{\tau_{\mathrm{d}}}^2$. A large value of α will cause the level of forgetting to be more sensitive to changes in the ratio of $\sigma_{\tau_{\mathrm{n}}}^2$ to $\sigma_{\tau_{\mathrm{d}}}^2$ compared to a smaller one.

For the case $p \geq 1$, the variances $\sigma_{\tau_{\mathrm{n}}}$ and $\sigma_{\tau_{\mathrm{d}}}$ are now covariance matrices $\Sigma_{\tau_{\mathrm{n}}}$ and $\Sigma_{\tau_{\mathrm{d}}}$, and the ratio of the two covariance matrices is given by $\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1}$. In this case, the ratio must be converted into a scalar test statistic. Four commonly used test statistics are

- Wilks's Lambda: $|I + \frac{\tau_{\mathrm{n}}}{\tau_{\mathrm{d}}}\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1}|^{-1} = \frac{|\tau_{\mathrm{d}}\Sigma_{\tau_{\mathrm{d}}}|}{|\tau_{\mathrm{n}}\Sigma_{\tau_{\mathrm{n}}} + \tau_{\mathrm{d}}\Sigma_{\tau_{\mathrm{d}}}|} = \prod_{i=1}^{n} \frac{1}{1+\mu_i}$,
- Lawley-Hotelling Trace: $\frac{\tau_{\mathrm{n}}}{\tau_{\mathrm{d}}}\mathrm{tr}(\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1}) = \sum_{i=1}^{n} \mu_i$,
- Pillai's Trace: $\mathrm{tr}\left(\frac{\tau_{\mathrm{n}}}{\tau_{\mathrm{d}}}\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1}(I + \frac{\tau_{\mathrm{n}}}{\tau_{\mathrm{d}}}\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1})^{-1}\right) = \sum_{i=1}^{n} \frac{\mu_i}{1+\mu_i}$,
- Roy's Greatest Root: $\max_i(\mu_i)$,

where $\mu_i,\ i = 1, \ldots, n$ are the eigenvalues of $\frac{\tau_{\mathrm{n}}}{\tau_{\mathrm{d}}}\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1}$ [15]. The Lawley-Hotelling trace is chosen due to its ease of use, similarity to the variance ratio in the F-Test, and availability of simple approximations. Using the Lawley-Hotelling trace with the approximation given by [18], $\Sigma_{\tau_{\mathrm{n}}}$ is greater than $\Sigma_{\tau_{\mathrm{d}}}$ with significance level α if

$$F_{p\tau_{\mathrm{n}},b}^{-1}(1-\alpha) < \frac{\tau_{\mathrm{n}}}{\tau_{\mathrm{d}}}\frac{\mathrm{tr}(\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1})}{c}, \tag{11}$$

where

$$a \triangleq \frac{(\tau_{\mathrm{n}} + \tau_{\mathrm{d}} - p - 1)(\tau_{\mathrm{d}} - 1)}{(\tau_{\mathrm{d}} - p - 3)(\tau_{\mathrm{d}} - p)},$$

$$b \triangleq 4 + \frac{(p\tau_{\mathrm{n}} + 2)}{(a-1)}, \quad c \triangleq \frac{p\tau_{\mathrm{n}}(b-2)}{b(\tau_{\mathrm{d}} - p - 1)}. \tag{12}$$

For $p = 1$, (11) is equivalent to the F-test. Given sample covariances of the past RLS prediction errors $\Sigma_{\tau_{\mathrm{n}}}$ and $\Sigma_{\tau_{\mathrm{d}}}$, g is defined as

$$g \triangleq \sqrt{\frac{\tau_{\mathrm{n}}}{\tau_{\mathrm{d}}}\frac{\mathrm{tr}\left(\Sigma_{\tau_{\mathrm{n}}}\Sigma_{\tau_{\mathrm{d}}}^{-1}\right)}{c}} - \sqrt{F_{p\tau_{\mathrm{n}},b}^{-1}(1-\alpha)}. \tag{13}$$

The resulting RLS update using the F-test for variable-rate forgetting (RLS/FTVRF) is given in Algorithm 1.

Some recommendations for the parameters $\tau_{\rm n}$, $\tau_{\rm d}$, η, and α are given as follows. A small $\tau_{\rm n}$ will cause forgetting to occur sooner and is recommended for fast changing systems while a larger $\tau_{\rm n} < 100$ will delay forgetting and is recommended for slowly changing systems. $\tau_{\rm d}$ should be 5-10 times larger than $\tau_{\rm n}$. The parameter η adjusts the amount of forgetting and is recommended to set $0 < \eta \leq 1$, with lower values of η if lack of input persistency is expected. The variable α adjusts the sensitivity of the forgetting factor and should be less than 0.1 with smaller values reducing the sensitivity. Smaller values are recommended if the system has noisy measurements.

Algorithm 1 RLS-VRF using the F-test (RLS/FTVRF)

Initialize: $\theta_0 \in \mathbb{R}^{n\times 1}$, $P_0 \in \mathbb{R}^{n\times n}$ positive-definite, $\tau_{\rm d} > \tau_{\rm n} \geq p$, $\eta > 0$, $\alpha > 0$, $k = 0$, and a buffer of $\tau_{\rm d}+1$ previous errors
while $k \geq 0$ **do**
  Measure $y_k \in \mathbb{R}^{p\times 1}$
  $e_k \leftarrow y_k - \phi_k\theta_k$
  Add e_k to error buffer and remove oldest entry
  Compute sample covariance matrices $\Sigma_{\tau_{\rm n}} \in \mathbb{R}^{p\times p}$, $\Sigma_{\tau_{\rm d}} \in \mathbb{R}^{p\times p}$ from previous $\tau_{\rm n}+1$ and $\tau_{\rm d}+1$ errors from buffer
  if $k \geq \tau_{\rm d}+1$ **then**
    Compute a, b, and c using (12)
    $g \leftarrow$ (13)
  else
    $g \leftarrow 0$
  end if
  $\beta_k \leftarrow 1 + \eta g \mathbf{1}[g]$
  $L_k \leftarrow \beta_k P_k$
  $\theta_{k+1} \leftarrow \theta_k + P_{k+1}\phi_k^{\rm T}(y_k - \phi_k\theta_k)$
  $P_{k+1} \leftarrow L_k - L_k\phi_k^{\rm T}(I_p + \phi_k L_k \phi_k^{\rm T})^{-1}\phi_k L_k$
  $\phi_{k+1} \leftarrow$ Update regressor ϕ_k with current measurement and input
  $k \leftarrow k+1$
end while

IV. Equivalence to the RLS/RMSVRF Function

In this section, we show that RLS/RMSVRF used in [13], [16] is a special case of RLS/FTVRF.

Proposition 4.1: Assume $\mathbb{E}[e(\theta)] = 0$. For $p = 1$ and $\alpha = 1 - F_{\tau_{\rm n},\tau_{\rm d}}(1)$, the variable-rate forgetting function given in [13], [16] is a special case of (13).

Proof: From (13), it follows that

$$\begin{aligned} g &\triangleq \sqrt{\frac{\tau_{\rm n}}{\tau_{\rm d}}\frac{\operatorname{tr}\left(\Sigma_{\tau_{\rm n}}\Sigma_{\tau_{\rm d}}^{-1}\right)}{c}} - \sqrt{F^{-1}_{p\tau_{\rm n},b}(1-\alpha)},\\ &= \sqrt{\frac{\sigma^2_{\tau_{\rm n}}}{\sigma^2_{\tau_{\rm d}}}} - \sqrt{F^{-1}_{\tau_{\rm n},\tau_{\rm d}}(1-\alpha)},\\ &= \sqrt{\frac{\frac{1}{\tau_{\rm n}}\sum_{i=k-\tau_n}^{k} e_i^2(\theta_i)}{\frac{1}{\tau_{\rm d}}\sum_{i=k-\tau_d}^{k} e_i^2(\theta_i)}} - \sqrt{F^{-1}_{\tau_{\rm n},\tau_{\rm d}}(1-\alpha)},\\ &= \sqrt{\frac{\frac{1}{\tau_{\rm n}}\sum_{i=k-\tau_n}^{k} e_i^2(\theta_i)}{\frac{1}{\tau_{\rm d}}\sum_{i=k-\tau_d}^{k} e_i^2(\theta_i)}} - 1. \end{aligned}$$

■

For the suggested values of $\tau_{\rm n}$ and $\tau_{\rm d}$ given in [16], the forgetting function is equivalent to using a significance level of $\alpha \approx 0.5$, and is equal to 0.5 in the limit of the window sizes $\lim_{\tau_{\rm n},\tau_{\rm d}\to\infty} F_{\tau_{\rm n},\tau_{\rm d}}(1) = 0.5$. A significance level of $\alpha = 0.5$ means that 50% of the time, we conclude that $\sigma^2_{\tau_{\rm n}} > \sigma^2_{\tau_{\rm d}}$ when it is not true, causing forgetting to occur when it is not needed. This may lead to instability of the RLS/RMSVRF algorithm if the forgetting were to occur under nonpersistent excitation.

V. Examples

To demonstrate RLS/FTVRF, we use a similar example to the one used in [9]. Consider a mass-spring-damper system with $M = 5\,\mathrm{kg}$, $K = 1\,\frac{\mathrm{N}}{\mathrm{m}}$, and $C = 1\,\frac{\mathrm{N\,s}}{\mathrm{m}}$. After 100 steps, the system parameters change to $K = 10\,\frac{\mathrm{N}}{\mathrm{m}}$, and $C = 0.01\,\frac{\mathrm{N\,s}}{\mathrm{m}}$. The discrete-time transfer function is given by

$$G_k(\mathbf{q}) = \begin{cases} \frac{0.4606\mathbf{q}+0.4307}{\mathbf{q}^2-1.64\mathbf{q}+0.8187}, & k < 100 \\ \frac{0.4218\mathbf{q}+0.4215}{\mathbf{q}^2-0.3116\mathbf{q}+0.998}, & k \geq 100, \end{cases} \tag{14}$$

where $\mathbf{q}$ is the forward shift operator. We compare RLS/FTVRF to RLS with a constant-rate forgetting factor (RLS/CRF) of $\lambda = 0.99$ under noiseless measurements, noisy measurements, and nonpersistent inputs. We also compare to RLS/RMSVRF under noisy measurements. For all cases, $\theta_0 = 0_{5\times 1}$, $P_0 = 100 I_5$, $\tau_{\rm n} = 10$, $\tau_{\rm d} = 80$, $\eta = 1$, and $\alpha = 0.001$. The regressor is implemented as $\phi_k = \begin{bmatrix} y_{k-1} & y_{k-2} & u_k & u_{k-1} & u_{k-2}\end{bmatrix}$ so that the coefficients of the transfer function (14) are identified in the parameter vector θ_k.

A. Noiseless Measurements

Let the input into (14) be $u_k \sim \mathcal{N}(0,1)$. First, comparing to RLS/CRF, from Figure 1 the parameters for RLS/FTVRF converge in less than 20 steps after the parameter change at $k = 100$. For RLS/CRF, the parameters are still not converged after 200 steps. Notice for RLS/FTVRF that once the prediction error increases, forgetting is enabled and $\operatorname{tr}(P_k)$ increases as past data is given lower weight. Figure 2 shows the median, and 5th and 95th percentiles of 1000 Monte Carlo simulations of RLS/FTVRF. Notice that all simulations converge to the true parameters in less than 20 steps.

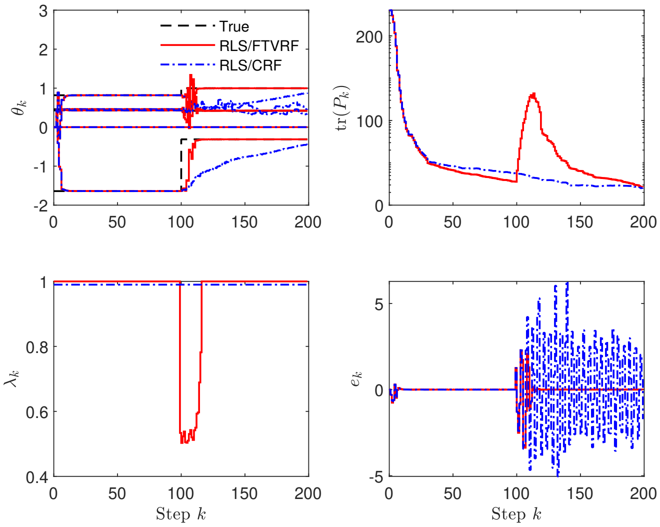

Fig. 1. Noiseless measurements. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/FTVRF and RLS/CRF.

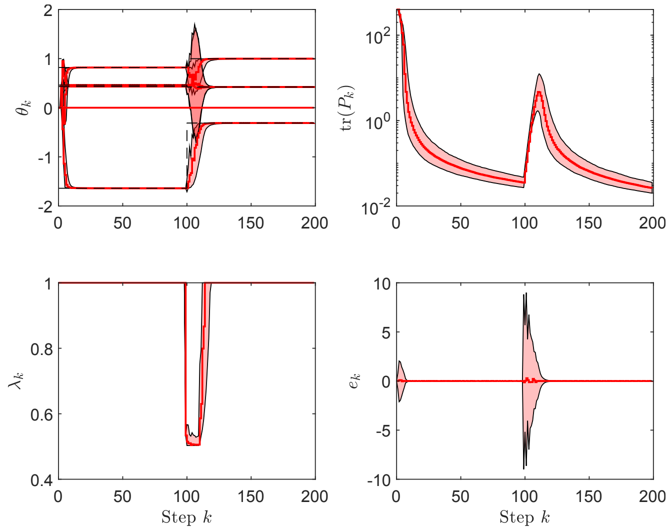

Fig. 2. Noiseless measurements. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/FTVRF for 1000 simulations. The red line is the median and the upper and lower bounds are the 95th and 5th percentiles, respectively.

B. Noisy Measurements

Let the input into (14) be $u_k \sim \mathcal{N}(0,1)$. Now the measurements are corrupted by noise $\nu_k \sim \mathcal{N}(0, 0.05)$. From Figure 3 the parameters for RLS/FTVRF converge in less than 20 steps after the parameters change at $k = 100$ despite the noisy measurements. For RLS/CRF, the parameters are still not converged after 200 steps. Notice for RLS/FTVRF that once the prediction error increases, forgetting is enabled and $\mathrm{tr}(P_k)$ increases as past data is given lower weight. The forgetting factor takes longer to reach its minimum value than in Figure 1, suggesting that the F-test limits the level of forgetting due to the uncertainty in whether the variance of errors has increased due to a parameter change or just temporarily due to noise. Figure 4 shows the median, and 5th and 95th percentiles of 1000 Monte Carlo simulations of RLS/FTVRF. Note that most of the simulation runs converge to the true parameters in less than 50 steps.

Figure 5 compares RLS/FTVRF to RLS/RMSVRF. Notice that RLS/RMSVRF enables forgetting at step 88 which is before any of the model parameters change and that forgetting also occurs at near the end of the simulation due to noise. RLS/FTVRF takes 8 more steps for its error to converge than RLS/RMSVRF, but convergence speed can be improved by increasing the significance level to $\alpha = 0.01$ without substantially risking forgetting before the parameter change or due to noise. Figure 6 shows the median, and 5th and 95th percentiles of 1000 Monte Carlo simulations of the RLS/RMSVRF method. At step 80, 26% of simulations forget when it is not needed. Many simulations also forget long after the parameter change suggesting sensitivity to noise.

C. Nonpersistently Exciting Input

Let the input into (14) be

$$u_k = \begin{cases} \mathcal{N}(0,1), & \text{if } k < 100 \\ 0, & \text{if } k \geq 100, \end{cases} \tag{15}$$

with measurements corrupted by noise $\nu_k \sim \mathcal{N}(0, 0.05)$. The input is no longer persistently exciting once the parameters change although the system may still be oscillating. In Figure 7 notice that over many steps for RLS/CRF, $\mathrm{tr}(P_k)$ continuously increases and will eventually cause RLS to 'blow-up' while for RLS/FTVRF, $\mathrm{tr}(P_k)$ stays bounded. Figure 8 shows the median, 5th and 95th percentiles of 1000 Monte Carlo simulations of RLS/FTVRF. Notice that all simulations keep $\mathrm{tr}(P_k)$ bounded.

VI. Conclusion

This paper developed and investigated the performance of RLS/FTVRF, which uses a variable-rate forgetting factor for recursive least squares based on the F-test. The variable-rate forgetting method uses a ratio of covariances of errors from a short and long moving horizon to determine if the underlying parameters have changed. A multivariate approximation of the F-test is used to extend the method to the multi-output case. The method was compared to a constant-rate forgetting factor in noiseless, noisy, and nonpersistently exciting input situations and also compared to the RLS/RMSVRF method used in [13], [16]. It was shown that RLS/FTVRF enabled forgetting when parameter parameter changes occurred and prevented forgetting from occurring due to noise. In the nonpersistent input case, the method kept RLS from forgetting and kept the eigenvalues of the RLS covariance matrix bounded.

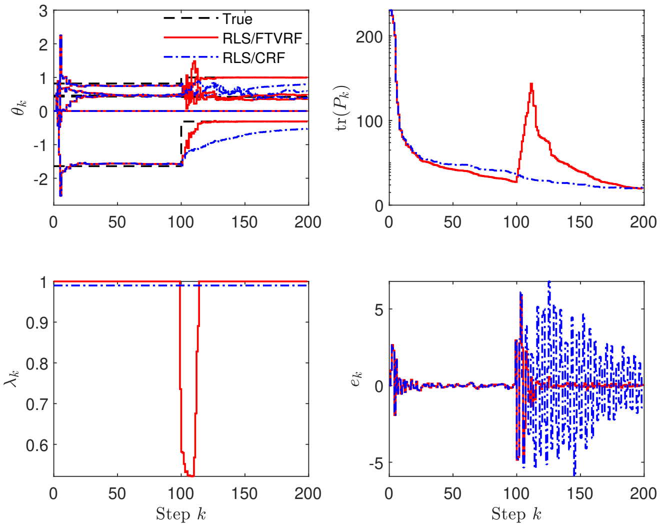

Fig. 3. Noisy measurements. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/FTVRF and RLS/CRF.

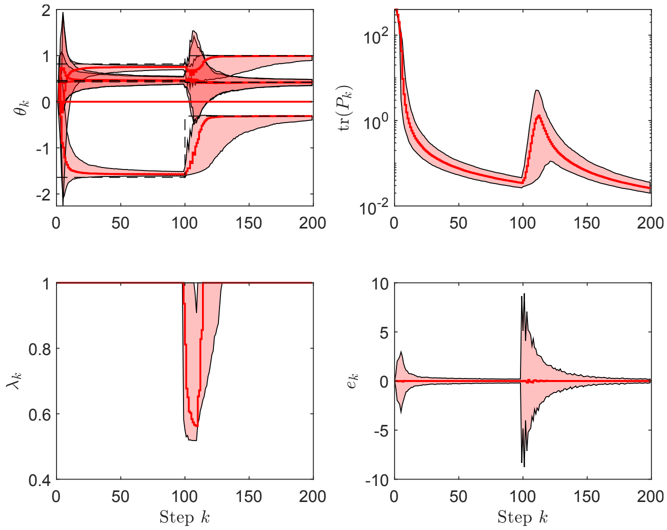

Fig. 4. Noisy measurements. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/FTVRF for 1000 simulations. The red line is the median, and the upper and lower bounds are the 95th and 5th percentiles, respectively.

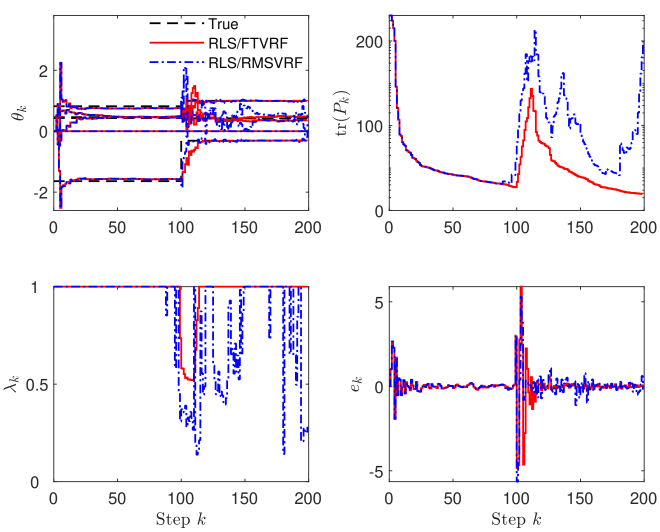

Fig. 5. Noisy measurements. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/FTVRF and RLS/RMSVRF.

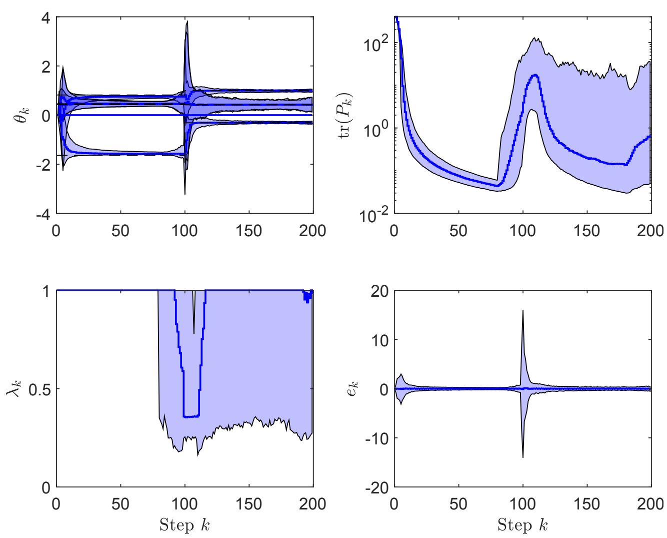

Fig. 6. Noisy measurements. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/RMSVRF for 1000 simulations. The blue line is the median and the upper and lower bounds are the 95th and 5th percentiles, respectively. Notice how forgetting sometimes occurs even before the parameter change at 100 steps. Forgetting also sometimes occurs long after the parameter change due to noise.

The F-test is known to be sensitive to non-normality of the data used to compute the sample variances. Situations such as nonwhite noise sources will need to be investigated in order to determine the extent of this sensitivity in the context of variable-rate forgetting. Additionally it is preferred to use a method of computing sample variances such that no buffer of past errors is needed for memory and computational performance reasons. Incorporating a weighted moving average of errors with a weighted sum of squares of errors as in [19] with a modified version of the F-test would allow faster computation of the forgetting factor.

Acknowledgments

This work was supported by a NASA Space Technology Graduate Research Opportunity under grant number 80NSSC20K1164.

References

[1] K. J. Astrom and B. Wittenmark, *Adaptive Control*, 2nd ed. Addison-Wesley, 1995.

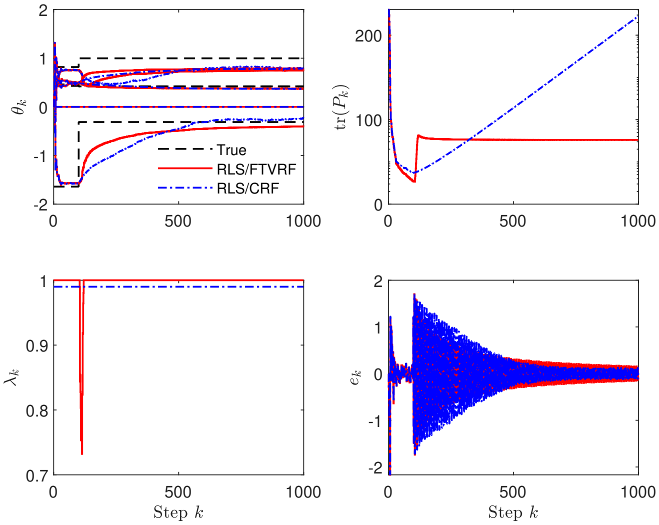

Fig. 7. Nonpersistently exciting input. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/FTVRF and RLS/CRF.

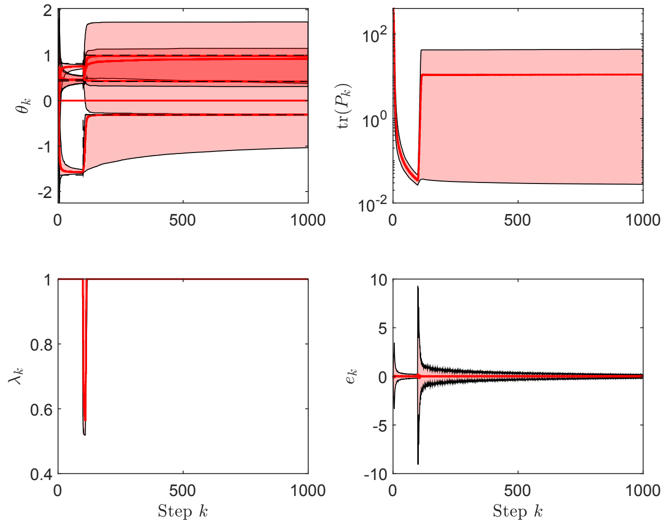

Fig. 8. Nonpersistently exciting measurements. Estimated parameters θ_k, trace of RLS covariance $\mathrm{tr}(P_k)$, forgetting factor λ_k, and prediction error e_k for RLS/FTVRF for 1000 simulations. The red line is the median and the upper and lower bounds are the 95th and 5th percentiles, respectively.

[2] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Prentice-Hall, 1996.
[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New Jersey: Wiley, 2003.
[4] L. Ljung, *System Identification: Theory for the User*. USA: Prentice-Hall, Inc., 1986.
[5] A. Goel, A. L. Bruce, and D. S. Bernstein, "Recursive least squares with variable-direction forgetting: Compensating for the loss of persistency [lecture notes]," *IEEE Control Systems Magazine*, vol. 40, no. 4, pp. 80–102, 2020.
[6] T. Fortescue, L. Kershenbaum, and B. Ydstie, "Implementation of self-tuning regulators with variable forgetting factors," *Automatica*, vol. 17, no. 6, pp. 831–835, 1981.
[7] S.-H. Leung and C. So, "Gradient-based variable forgetting factor rls algorithm in time-varying environments," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3141–3150, 2005.
[8] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, ser. The MIT Press Series in Signal Processing, Optimization, and Control. MIT Press, 1983, no. 4.
[9] A. L. Bruce, A. Goel, and D. S. Bernstein, "Convergence and consistency of recursive least squares with variable-rate forgetting," *Automatica*, vol. 119, p. 109052, 2020.
[10] C. Elisei-Iliescu, C. Paleologu, J. Benesty, C. Stanciu, C. Anghel, and S. Ciochină, "A decomposition-based rls algorithm with variable forgetting factors," in *13th International Conference on Communications*, 2020, pp. 43–48.
[11] H. Shin and H. Lee, "A new exponential forgetting algorithm for recursive least-squares parameter estimation," *CoRR*, vol. abs/2004.03910, 2020.
[12] S. Draŝković, Ẑ. Đurović, and V. Petrović, "Absolute finite differences based variable forgetting factor rls algorithm," *IET Signal Processing*, 2021.
[13] S. A. U. Islam, T. Nguyen, I. Kolmanovsky, and D. S. Bernstein, "Data-Driven Retrospective Cost Adaptive Control for Flight Control Applications," *J. Guid. Contr. Dyn.*, 2021, available online.
[14] G. W. Snedecor, *Statistical Methods*, 4th ed. Wiley, 1948.
[15] R. Mickey and A. C. Rencher, *Methods of Multivariate Analysis*. Wiley, 1998, vol. 93, no. 443.
[16] T. Nguyen, S. A. U. Islam, D. S. Bernstein, and I. Kolmanovsky, "Predictive cost adaptive control: A numerical investigation of persistency, consistency, and exigency," *IEEE Control Systems Magazine*, 2021, to appear.
[17] R. R. Rhinehart, "Automated steady and transient state identification in noisy processes," in *American Control Conference*, 2013, pp. 4477–4493.
[18] J. J. McKeon, "F approximations to the distribution of Hotelling's T_0^2," *Biometrika*, vol. 61, no. 2, pp. 381–383, 08 1974.
[19] Songling Cao and R. Russell Rhinehart, "An efficient method for online identification of steady state," *Journal of Process Control*, vol. 5, no. 6, pp. 363–374, 1995.