Generalized Forgetting Recursive Least Squares: Stability and Robustness Guarantees

Brian Lai and Dennis S. Bernstein, *Life Fellow, IEEE*

***Abstract*—This work presents generalized forgetting recursive least squares (GF-RLS), a generalization of RLS that encompasses many extensions of RLS as special cases. First, sufficient conditions are presented for the 1) Lyapunov stability, 2) uniform Lyapunov stability, 3) global asymptotic stability, and 4) global uniform exponential stability of parameter estimation error in GF-RLS when estimating fixed parameters without noise. Second, robustness guarantees are derived for the estimation of time-varying parameters in the presence of measurement noise and regressor noise. These robustness guarantees are presented in terms of global uniform ultimate boundedness of the parameter estimation error. A specialization of this result gives a bound to the asymptotic bias of least squares estimators in the errors-in-variables problem. Lastly, a survey is presented to show how GF-RLS can be used to analyze various extensions of RLS from literature.**

***Index Terms*—Identification, recursive least squares (RLS), robustness, stability analysis, errors in variables.**

I. Introduction

RECURSIVE least squares (RLS) is a foundational algorithm in systems and control theory for the online identification of fixed parameters [1], [2], [3]. A property of RLS is the eigenvalues of the covariance matrix which are monotonically decreasing over time and may become arbitrarily small [4, Sec. 2.3.2], [5], resulting in eventual sluggish adaptation and inability to track time-varying parameters [6], [7]. Numerous extensions of RLS have been developed to improve identification of time-varying parameters, including exponential forgetting [3], [8], variable-rate forgetting [9], [10], [11], [12], [13], [14], directional forgetting [15], [16], [17], resetting [7], [18], [19], and multiple forgetting [20], among others. Hence, we use the general term *forgetting* to describe the processes in extensions of RLS which break the monotonicity of the covariance matrix. Furthermore, several general frameworks have been developed which include extensions of RLS as special cases, for example [21] in discrete-time and [22] in continuous-time. Bin [23] developed a much more general framework of recursive estimators which contains RLS extensions as a special case. These frameworks help to unify various RLS extensions and provide overarching analysis. In discussing RLS extensions, we highlight three important points: 1) cost function, 2) stability, and 3) robustness.

1) Cost Function: The RLS update equations are derived as a recursive method to find the minimizer of a least-squares cost function [3]. While some RLS extensions can be derived from modified least-squares cost functions (e.g., exponential forgetting [3] and variable-rate forgetting [9]), many have been developed as ad-hoc modifications to the RLS update equations, without an associated cost function (e.g., [7], [8], [11], [12], [15], [16], [17], [19], [20]). Therefore, there is an interest in developing a cost function from which extensions of RLS can be derived. In [22], a continuous-time cost functional is presented, from which continuous-time least-squares algorithms are derived. However, the least-squares algorithms derived from the cost function in [22] are continuous-time versions of RLS extensions, not the original RLS extensions developed in discrete time.

2) Stability: Many RLS extensions give conditions which guarantee stability of parameter estimation error to zero when estimating constant parameters [8], [9], [17], [18]. General frameworks [21], [22], [23] all present stability guarantees which apply to various RLS extensions. The stability analyses in [21] and [22] consider RLS extensions with scalar measurements and present exponential stability guarantees for constant-parameter estimation. While the stability analysis in [23] encompasses RLS extensions with vector measurements, the sufficient conditions for stability may be overly restrictive when applied to RLS extensions as a much more general class of recursive estimators is analyzed.

3) Robustness: Several RLS extensions further analyze robustness to time-varying parameters and to bounded measurement noise [11], [24], [25]. If there is noise in the measurement and regressor, this is known as the *errors-in-variables* problem, which is significantly more challenging than the problem of robustness to measurement noise alone [26], [27], [28]. In particular, if the measurement noise and regressor noise are correlated, then, with the exception of some special cases, the least squares estimator is asymptotically biased [29, p. 205] and it is often desirable to bound such bias. Similarly to stability, while the authors in [23] addressed the errors-in-variables problem, the required sufficient conditions may be overly restrictive and the bounds on asymptotic bias overly loose when applied specifically to RLS extensions.

Manuscript received 23 February 2024; accepted 21 April 2024. Date of publication 26 April 2024; date of current version 25 October 2024. This work was supported by the NSF Graduate Research Fellowship under Grant DGE 1841052. Recommended by Associate Editor A. Bobtsov. *(Corresponding author: Brian Lai.)*

The authors are with the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: brianlai@umich.edu; dsbaero@umich.edu).

Digital Object Identifier 10.1109/TAC.2024.3394351

A. Contributions

The contributions of this article are summarized as follows.

1) We derive generalized forgetting RLS (GF-RLS), which is a discrete-time version of the continuous-time RLS generalization developed in [22], see Section II. While [22] presents continuous-time versions of RLS algorithms that fit into their framework, we will show in Section V how various RLS algorithms from literature, without any modifications, can be derived from the GF-RLS cost function as special cases. Hence, our subsequent analysis directly applies the original discrete-time RLS algorithms which have been widely used in their original discrete-time form.
2) For constant-parameter estimation, we use Lyapunov methods to develop stability guarantees for GF-RLS. These guarantees extend results obtained in [21] by generalizing to vector measurements and by providing weaker stability guarantees when not all the conditions for exponential stability are met, see Section III. The sufficient conditions for stability we present are similar to the heuristic conditions in [7] for design of RLS algorithms. Thus, we believe our sufficient conditions may serve as a more precise guideline in the design of future RLS algorithms.
3) In addition, we develop robustness guarantees of GF-RLS to bounded parameter variation, bounded measurement noise, and bounded regressor noise. In particular, we obtain sufficient conditions for the global uniform ultimate boundedness of the parameter-estimation error. None of these three additional factors were considered in [21] or [22]. A specialization of this result provides a bound on the asymptotic bias of parameter estimation error in the context of the errors-in-variables problem, see Section IV.

B. Notation and Terminology

$\mathbb{N}_0$ denotes the set of nonnegative integers $\{0,1,2,\ldots\}$. I_n denotes the $n\times n$ identity matrix, and $0_{m\times n}$ denotes the $m\times n$ zero matrix. For symmetric $A\in\mathbb{R}^{n\times n}$, let the n real eigenvalues of A be denoted by $\boldsymbol{\lambda}_{\mathbf{min}}(A)\triangleq\boldsymbol{\lambda}_{\boldsymbol{n}}(A)\leq\cdots\leq\boldsymbol{\lambda}_{\mathbf{max}}(A)\triangleq\boldsymbol{\lambda}_{\mathbf{1}}(A)$. For $B\in\mathbb{R}^{m\times n}$, $\boldsymbol{\sigma}_{\mathbf{max}}(B)$ denotes the largest singular value of B, and $\boldsymbol{\sigma}_{\mathbf{min}}(B)$ denotes the smallest singular value of B.

For symmetric $P,Q\in\mathbb{R}^{n\times n}$, $P\prec Q$ (respectively, $P\preceq Q$) denotes that $Q-P$ is positive definite (respectively, positive semidefinite). For all $x\in\mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm, that is $\|x\|\triangleq\sqrt{x^{\mathrm{T}}x}$. For positive-semidefinite $R\in\mathbb{R}^{n\times n}$ and $x\in\mathbb{R}^n$, $\|x\|_R\triangleq\sqrt{x^{\mathrm{T}}Rx}$. For symmetric $S\in\mathbb{R}^{n\times n}$, $\|x\|_S^2\triangleq x^{\mathrm{T}}Sx$. Note that the notation $\|x\|_S^2$ is used only for convenience and that $\|x\|_S$ is not defined when S is not positive semidefinite. For $\varepsilon>0$ and $x_e\in\mathbb{R}^n$, define the closed ball $\bar{\mathcal{B}}_\varepsilon(x_e)\triangleq\{x\in\mathbb{R}^n:\|x-x_e\|\leq\varepsilon\}$.

Definition 1: A sequence $(\phi_k)_{k=k_0}^\infty\subset\mathbb{R}^{p\times n}$ is *persistently exciting* if there exist $N\geq1$ and $\alpha>0$ such that, for all $k\geq k_0$

$$\alpha I_n\preceq\sum_{i=k}^{k+N-1}\phi_i^{\mathrm{T}}\phi_i. \quad (1)$$

Furthermore, α and N are, respectively, the *lower bound* and *persistency window* of $(\phi_k)_{k=k_0}^\infty$.

Definition 2: A sequence $(\phi_k)_{k=k_0}^\infty\subset\mathbb{R}^{p\times n}$ is *bounded* if there exists $\beta\in(0,\infty)$ such that, for all $k\geq k_0$

$$\phi_k^{\mathrm{T}}\phi_k\preceq\beta I_n. \quad (2)$$

Furthermore, β is the *upper bound* of $(\phi_k)_{k=k_0}^\infty$.

II. Generalized Forgetting Recursive Least Squares (GF-RLS)

The following theorem presents GF-RLS, which is a discrete-time RLS generalization derived from minimizing a least-squares cost function.

Theorem 1: For all $k\geq0$, let $\Gamma_k\in\mathbb{R}^{p\times p}$ be positive definite, let $\phi_k\in\mathbb{R}^{p\times n}$, and let $y_k\in\mathbb{R}^p$. Furthermore, let $P_0\in\mathbb{R}^{n\times n}$ be positive definite, and let $\theta_0\in\mathbb{R}^n$. For all $k\geq0$, let $F_k\in\mathbb{R}^{n\times n}$ be symmetric and satisfy

$$F_k\prec P_0^{-1}+\sum_{i=0}^{k-1}\left(-F_i+\phi_i^{\mathrm{T}}\Gamma_i^{-1}\phi_i\right). \quad (3)$$

For all $k\geq0$, define $J_k:\mathbb{R}^n\to\mathbb{R}$ by

$$J_k(\hat\theta)\triangleq J_{k,\mathrm{loss}}(\hat\theta)-J_{k,\mathrm{forget}}(\hat\theta)+J_{k,\mathrm{reg}}(\hat\theta) \quad (4)$$

where

$$J_{k,\mathrm{loss}}(\hat\theta)\triangleq\sum_{i=0}^{k}\|y_i-\phi_i\hat\theta\|_{\Gamma_i^{-1}}^2 \quad (5)$$

$$J_{k,\mathrm{forget}}(\hat\theta)\triangleq\sum_{i=0}^{k}\|\hat\theta-\theta_i\|_{F_i}^2 \quad (6)$$

$$J_{k,\mathrm{reg}}(\hat\theta)\triangleq\|\hat\theta-\theta_0\|_{P_0^{-1}}^2. \quad (7)$$

Then, J_k has a unique global minimizer, denoted

$$\theta_{k+1}\triangleq\arg\min_{\hat\theta\in\mathbb{R}^n}J_k(\hat\theta) \quad (8)$$

which, for all $k\geq0$, is given by

$$P_{k+1}^{-1}=P_k^{-1}-F_k+\phi_k^{\mathrm{T}}\Gamma_k^{-1}\phi_k \quad (9)$$

$$\theta_{k+1}=\theta_k+P_{k+1}\phi_k^{\mathrm{T}}\Gamma_k^{-1}(y_k-\phi_k\theta_k). \quad (10)$$

Proof: See Appendix C. ■

For all $k\geq0$, We call $y_k\in\mathbb{R}^p$ the *measurement*, $\phi_k\in\mathbb{R}^{p\times n}$ the *regressor*, and $\theta_k\in\mathbb{R}^n$ the *parameter estimate*. Moreover, we call $\Gamma_k\in\mathbb{R}^{p\times p}$ the *weighting matrix*, $F_k\in\mathbb{R}^{n\times n}$ the *forgetting matrix*, and $P_k\in\mathbb{R}^{n\times n}$ the *covariance matrix*. Furthermore, (9) and (10) are the GF-RLS update equations.

Notice that condition (3) guarantees that J_k has a unique global minimizer, as shown in the proof of Theorem 1. Corollary 1 gives an important interpretation to (3).

Corollary 1: Consider the notation and assumptions of Theorem 1. Then, for all $k\geq0$

$$P_k^{-1}=P_0^{-1}+\sum_{i=0}^{k-1}\left(-F_i+\phi_i^{\mathrm{T}}\Gamma_i^{-1}\phi_i\right). \quad (11)$$

Hence, for all $k\geq0$, (3) holds if and only if

$$P_k^{-1}-F_k\succ0. \quad (12)$$

Proof: Equation (11) follows directly from repeated substitution of (9). Next, (12) follows from substituting (11) into (3). ∎

Corollary 1 shows that to ensure that the GF-RLS cost (4) has a unique global minimizer [i.e., (3) is satisfied], it suffices to, for all $k \geq 0$, choose F_k such that $P_k^{-1} - F_k \succ 0$.

Definition 3: GF-RLS is *proper* if, for all $k \geq 0$, $F_k \in \mathbb{R}^{n\times n}$ is positive semidefinite. GF-RLS is *improper* if it is not proper.

Note that the GF-RLS cost $J_k(\hat{\theta})$ is composed as the sum of three terms, namely, the *loss term* $J_{k,\text{loss}}(\hat{\theta})$, the *forgetting term* $-J_{k,\text{forget}}(\hat{\theta})$, and the *regularization term* $J_{k,\text{reg}}(\hat{\theta})$. Note that, if GF-RLS is proper, then, for all $\hat{\theta} \in \mathbb{R}^n$, the forgetting term $-J_{k,\text{forget}}(\hat{\theta})$ is nonpositive. In practice, if GF-RLS is proper, then the forgetting term rewards the difference between the estimate θ_{k+1} and θ_i for previous steps $0 \leq i \leq k$. This reward is weighted by the forgetting matrix $F_k \in \mathbb{R}^{n\times n}$. It is shown in Section V that for particular choices of the forgetting matrix, we recover extensions of RLS with forgetting from GF-RLS.

III. Stability of Fixed Parameter Estimation

For the analysis of this section, we make the assumption that there exist fixed parameters $\theta \in \mathbb{R}^n$ such that, for all $k \geq 0$

$$y_k = \phi_k \theta. \tag{13}$$

Furthermore, for all $k \geq 0$, we define the parameter estimation error $\tilde{\theta}_k \in \mathbb{R}^n$ by

$$\tilde{\theta}_k \triangleq \theta_k - \theta. \tag{14}$$

Substituting into (10), it then follows that

$$\tilde{\theta}_{k+1} = M_k \tilde{\theta}_k \tag{15}$$

where for all $k \geq 0$, $M_k \in \mathbb{R}^{n\times n}$ is defined

$$M_k \triangleq I_n - P_{k+1}\phi_k^{\mathrm{T}}\Gamma_k^{-1}\phi_k. \tag{16}$$

Hence, (15) is a linear time-varying (LTV) system with an equilibrium $\tilde{\theta}_k \equiv 0$.

Next, for all $k \geq 0$, let $\Gamma_k^{-\frac{1}{2}} \in \mathbb{R}^{p\times p}$ be the unique positive-definite matrix such that

$$\Gamma_k^{-1} = \Gamma_k^{-\frac{1}{2}\mathrm{T}}\Gamma_k^{-\frac{1}{2}}. \tag{17}$$

Furthermore, define the *weighted regressor* $\bar{\phi}_k \in \mathbb{R}^{p\times n}$ by

$$\bar{\phi}_k \triangleq \Gamma_k^{-\frac{1}{2}}\phi_k. \tag{18}$$

Substituting (18) into (16), it follows that, for all $k \geq 0$

$$M_k = I_n - P_{k+1}\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k. \tag{19}$$

Finally, let $k_0 \geq 0$ and consider the following conditions.

A1) For all $k \geq k_0$, $F_k \succeq 0$.

A2) There exists $b \in (0,\infty)$ such that, for all $k \geq k_0$, $(P_k^{-1} - F_k)^{-1} \preceq bI_n$.

A3) There exists $a > 0$ such for all $k \geq k_0$, $aI_n \preceq P_k$.

A4) The sequence of weighted regressors $(\bar{\phi}_k)_{k=k_0}^{\infty}$ is persistently exciting with lower bound $\bar{\alpha} > 0$ and persistency window $N \geq 1$ and bounded with upper bound $\bar{\beta} \in (0,\infty)$.

We now present Theorem 2 which gives sufficient conditions for the stability of the equilibrium $\tilde{\theta}_k \equiv 0$ of (15). See Appendix B for relevant discrete-time stability definitions.

Theorem 2: For all $k \geq 0$, let $\Gamma_k \in \mathbb{R}^{p\times p}$ be positive definite, let $\phi_k \in \mathbb{R}^{p\times n}$, let $y_k \in \mathbb{R}^p$, and let $F_k \in \mathbb{R}^{n\times n}$ be symmetric and satisfy (3). Let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite, and let $\theta_0 \in \mathbb{R}^n$. For all $k \geq 1$, let $P_k \in \mathbb{R}^{n\times n}$ and $\theta_k \in \mathbb{R}^n$ be recursively updated by (9) and (10). Furthermore, assume there exists $\theta \in \mathbb{R}^n$ such that, for all $k \geq 0$, (13) holds. Then the following statements hold.

1) If there exists $k_0 \geq 0$ such that A1) and A2), then the equilibrium $\tilde{\theta}_k \equiv 0$ of (15) is Lyapunov stable.
2) If there exists $k_0 \geq 0$ such that A1), A2), and A3) then the equilibrium $\tilde{\theta}_k \equiv 0$ of (15) is uniformly Lyapunov stable.
3) If there exists $k_0 \geq 0$ such that A1), A2), and A4), then the equilibrium $\tilde{\theta}_k \equiv 0$ of (15) is globally asymptotically stable.[1]
4) If there exists $k_0 \geq 0$ such that A1), A2), A3), and A4), then the equilibrium $\tilde{\theta}_k \equiv 0$ of (15) is globally uniformly exponentially stable.[1]

Proof: In the case $k_0 = 0$, see Appendix D for a proof of statements 1) and 2), and Appendix E for a proof of statements 3) and 4). The case $k_0 \geq 1$ can be shown similarly. Also see Fig. 2 for a proof roadmap of Theorem 2. ∎

A. Discussion of Conditions A1) Through A4)

This subsection gives a brief discussion of conditions A1) through A4) used in Theorem 2.

1) Condition A1): Note that by Definition 3, this condition is equivalent to GF-RLS being proper. Furthermore, whether or not GF-RLS is proper is a direct consequence of the algorithm design. We will show in Section V how ten different extensions of RLS are all proper (some requiring minor assumptions), and hence satisfy condition A1).

2) Conditions A2) and A3): In 1988, Salgado et al. [7] qualitatively proposed that RLS extensions should guarantee a (nonzero) lower bound and a (noninfinite) upper bound of the covariance matrix P_k for good performance. That is, $a > 0$ and $b \in (0,\infty)$ such that, for all $k \geq 0$, $aI_n \preceq P_k \preceq bI_n$. Many RLS extensions since have provided analysis which guarantee an upper and lower bound of the covariance matrix [5], [15], [17], [19]. A lower bound on P_k is equivalent to condition A3). However, an upper bound on P_k does not guarantee condition A2).

Nevertheless, for many choices of F_k from literature, condition A2) follows easily from an upper bound on the covariance matrix. A future area of interest is whether similar stability and robustness guarantees exist if condition A2) is replaced with an upper bound on P_k.

3) Condition A4): Persistent excitation and boundedness of the sequence of regressors $(\phi_k)_{k=k_0}^{\infty}$ is an important requirement for convergence in RLS extensions [7], [8]. While work has been

[1] Since (15) is LTV, it suffices to say that (15) is asymptotically (resp. uniformly exponentially) stable, as asymptotic (respectively, uniform exponential) stability implies global asymptotic (respectively, uniform exponential) stability for LTV systems [30], [31, Sec. 5.5].

done to relax the persistent excitation condition [32], [33], [34], it has been shown that *weak persistent excitation* is necessary for the global asymptotic stability of RLS [35].

Note that condition A4) requires persistent excitation and boundedness of the the sequence of weighted regressors $(\bar{\phi}_k)_{k=k_0}^{\infty}$, rather than the sequence of regressors $(\phi_k)_{k=k_0}^{\infty}$. Corollary 2 gives a sufficient condition for when persistent excitation and boundedness of the sequence of regressors $(\phi_k)_{k=k_0}^{\infty}$ implies persistent excitation and boundedness of the sequence of weighted regressors $(\bar{\phi}_k)_{k=k_0}^{\infty}$. Note, however, that the bounds guaranteed by Corollary 2 are often loose and it is preferable in practice to directly analyze the sequence of weighted regressors.

Corollary 2: Assume there exist $k_0 \geq 0$ and $0 < \gamma_{\min} < \gamma_{\max}$ such that, for all $k \geq k_0$

$$\gamma_{\min} I_p \preceq \Gamma_k \preceq \gamma_{\max} I_p. \tag{20}$$

Furthermore, let $(\phi_k)_{k=k_0}^{\infty}$ be persistently exciting with lower bound $\alpha > 0$ and persistency window N and bounded with upper bound $\beta \in (0, \infty)$. Then, $(\bar{\phi}_k)_{k=k_0}^{\infty}$ is persistently exciting with lower bound $\alpha/\gamma_{\max}$ and persistency window N and bounded with upper bound $\beta/\gamma_{\min}$.

Proof: Note that, for all $k \geq k_0$, $\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k = \phi_k^{\mathrm{T}}\Gamma_k^{-1}\phi_k \succeq \frac{1}{\gamma_{\max}}\phi_k^{\mathrm{T}}\phi_k$, and hence $\sum_{i=k}^{k+N-1}\bar{\phi}_i^{\mathrm{T}}\bar{\phi}_i \succeq \sum_{i=k}^{k+N-1}\frac{1}{\gamma_{\max}}\phi_i^{\mathrm{T}}\phi_i \succeq \frac{\alpha}{\gamma_{\max}}I_n$. ■

IV. Robustness to Time-Varying Parameters, Measurement Noise, and Regressor Noise

For the analysis of this section, we make the assumption that, for all $k \geq 0$, the parameters $\theta_{\mathrm{true},k} \in \mathbb{R}^n$ are time varying and satisfy

$$\theta_{\mathrm{true},k+1} = \theta_{\mathrm{true},k} + \delta_{\theta,k} \tag{21}$$

where, for all $k \geq 0$, $\delta_{\theta,k} \in \mathbb{R}^n$ is the *change in the parameters*. Note that no model of how the parameters evolve is known. Furthermore, assume that, for all $k \geq 0$

$$y_k = (\phi_k + \delta_{\phi,k})\theta_{\mathrm{true},k} + \delta_{y,k} \tag{22}$$

where $\delta_{y,k} \in \mathbb{R}^p$ is the *measurement noise* and $\delta_{\phi,k} \in \mathbb{R}^{p\times n}$ is the *regressor noise*. Furthermore, for all $k \geq 0$, we define the *weighted measurement noise* at step k, $\bar{\delta}_{y,k} \in \mathbb{R}^p$, and the *weighted regressor noise*, $\bar{\delta}_{\phi_k} \in \mathbb{R}^{p\times n}$, by

$$\bar{\delta}_{y,k} \triangleq \Gamma_k^{-\frac{1}{2}}\delta_{y,k} \tag{23}$$

$$\bar{\delta}_{\phi,k} \triangleq \Gamma_k^{-\frac{1}{2}}\delta_{\phi,k}. \tag{24}$$

Note that, for all $k \geq 1$, the parameter estimate θ_k is based on measurements up to step $k-1$, that is, $\{y_0, \ldots, y_{k-1}\}$. To compensate for this one-step delay, we define, for all $k \geq 0$, the parameter estimation error $\breve{\theta}_k \in \mathbb{R}^n$ by

$$\breve{\theta}_k \triangleq \theta_k - \theta_{\mathrm{true},k-1}. \tag{25}$$

Substituting (22) and (25) into (10) implies that, for all $k \geq 0$,

$$\breve{\theta}_{k+1} = M_k(\breve{\theta}_k - \delta_{\theta,k-1}) + P_{k+1}\phi_k^{\mathrm{T}}\Gamma_k^{-1}(\delta_{\phi,k}\theta_{\mathrm{true},k} + \delta_{y,k})$$

and then substituting (18), (23), and (24), it follows that

$$\breve{\theta}_{k+1} = M_k(\breve{\theta}_k - \delta_{\theta,k-1}) + P_{k+1}\bar{\phi}_k^{\mathrm{T}}(\bar{\delta}_{\phi,k}\theta_{\mathrm{true},k} + \bar{\delta}_{y,k}). \tag{26}$$

Hence, (26) is a nonlinear system $\breve{\theta}_{k+1} = \breve{f}(k, \breve{\theta}_k)$.

Finally, let $k_0 \geq 0$ and consider the following conditions.

A5) There exists $\delta_\theta \geq 0$ such that, for all $k \geq k_0$, $\|\delta_{\theta,k}\| \leq \delta_\theta$.

A6) There exists $\bar{\delta}_y \geq 0$ such that, for all $k \geq k_0$, $\|\bar{\delta}_{y,k}\| \leq \bar{\delta}_y$.

A7) There exists $\bar{\delta}_\phi \geq 0$ such that the sequence $(\bar{\delta}_{\phi,k})_{k=k_0}^{\infty}$ is bounded with upper bound $\bar{\delta}_\phi$.

A8) There exists $\theta_{\max} \geq 0$ such that, for all $k \geq k_0$, $\|\theta_{\mathrm{true},k}\| \leq \theta_{\max}$.

We now present Theorem 3 which gives sufficient conditions for the global uniform ultimate boundedness of (26). Please see Appendix B for the definition of global uniform ultimate boundedness.

Theorem 3: For all $k \geq 0$, let $\Gamma_k \in \mathbb{R}^{p\times p}$ be positive definite, let $\phi_k \in \mathbb{R}^{p\times n}$, let $y_k \in \mathbb{R}^p$, and let $F_k \in \mathbb{R}^{n\times n}$ be symmetric and satisfy (3). Let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite, and let $\theta_0 \in \mathbb{R}^n$. For all $k \geq 1$, let $P_k \in \mathbb{R}^{n\times n}$ and $\theta_k \in \mathbb{R}^n$ be recursively updated by (9) and (10). Furthermore, for all $k \geq 0$, let $\theta_{\mathrm{true},k} \in \mathbb{R}^n$, $\delta_{\theta,k} \in \mathbb{R}^n$, $\delta_{y,k} \in \mathbb{R}^p$, and $\delta_{\phi,k} \in \mathbb{R}^{p\times n}$ satisfy (21) and (22). Finally, let $k_0 \geq 0$ be such that all conditions A1)–A8) hold. Then, system (26) is globally uniformly ultimately bounded with bound ε given by

$$\varepsilon = \varepsilon^*\left[\delta_\theta + b\bar{\beta}^{\frac{1}{2}}\left(\bar{\delta}_\phi^{\frac{1}{2}}\theta_{\max} + \bar{\delta}_y\right)\right] \tag{27}$$

where

$$\varepsilon^* \triangleq \max\left\{1, \frac{1}{\sqrt{a}}\right\}\left(\Delta_N + \sqrt{\Delta_N + \Delta_N^2}\right)N \tag{28}$$

$$\Delta_N \triangleq \frac{N}{a\bar{\alpha}}\left(1 + b\bar{\beta}\right)\left[1 + \frac{N-1}{2}\left(b\bar{\beta}\right)^2\right] - 1. \tag{29}$$

Proof: We prove the case $k_0 = 0$. The case $k_0 \geq 1$ can be shown similarly. Note that, for all $k \geq 0$, (26) can be written as

$$\breve{\theta}_{k+1} = M_k(\breve{\theta}_k - \delta_{\theta,k-1} + M_k^{-1}P_{k+1}\bar{\phi}_k^{\mathrm{T}}(\bar{\delta}_{\phi,k}\theta_{\mathrm{true},k} + \bar{\delta}_{y,k})). \tag{30}$$

Moreover, it follows from (9) and (16) that, for all $k \geq 0$, $M_k = P_{k+1}(P_k^{-1} - F_k)$. It follows from Corollary 1 that, for all $k \geq 0$, $(P_k^{-1} - F_k)$ is nonsingular, and hence

$$M_k^{-1} = (P_k^{-1} - F_k)^{-1}P_{k+1}^{-1}. \tag{31}$$

Substituting (31) into (30) then gives, for all $k \geq 0$

$$\breve{\theta}_{k+1} = M_k(\breve{\theta}_k - \zeta_k)$$

where $\zeta_k \in \mathbb{R}^n$ is defined

$$\zeta_k \triangleq \delta_{\theta,k-1} - (P_k^{-1} - F_k)^{-1}\bar{\phi}_k^{\mathrm{T}}(\bar{\delta}_{\phi,k}\theta_{\mathrm{true},k} + \bar{\delta}_{y,k}). \tag{32}$$

Next, it follows from applying triangle inequality and norm submultiplicativity to (32) and using the bounds in conditions A2), A4), A5), A6), A7), and A8) that, for all $k \geq 0$,

$$\|\zeta_k\| \leq \delta_\theta + b\bar{\beta}^{\frac{1}{2}}\left(\bar{\delta}_\phi^{\frac{1}{2}}\theta_{\max} + \bar{\delta}_y\right) \triangleq \zeta.$$

Finally, it follows Lemma 11 in Appendix F that the system (26) is globally uniformly ultimately bounded with bound $\varepsilon^*\zeta$. For further details, see Fig. 2 for a proof roadmap of Theorem 3. ■

As a sanity check, note that, for all $N \geq 1$, $\Delta_N > \frac{Nb\bar{\beta}}{a\bar{\alpha}} - 1 \geq N-1 \geq 0$, and hence $\varepsilon^* > 0$. Therefore, if $\delta_\theta > 0$, $\bar{\delta}_y > 0$, or $\bar{\delta}_\phi^{\frac{1}{2}}\theta_{\max} > 0$, then $\varepsilon > 0$.

A. Discussion of Conditions A5) Through A8)

This subsection gives a discussion of conditions A5) through A8) used in Theorem 3. See Section III-A for a discussion of conditions A1) through A4).

1) Condition A5): Condition A5) is a bound on how quickly the parameters being estimated, $\theta_{\text{true},k}$, can change. While these parameters are not known, in practice, this bound can be estimated from data.

2) Conditions A6) and A7): Conditions A6) and A7) are, respectively, bounds on the weighted measurement noise and weighted regressor noise. While noise from certain distributions has no guaranteed bound (e.g., Gaussian noise), in practice these bounds can be approximated from data.

Corollary 3 gives a sufficient condition for when bounded measurement noise and bounded regressor noise imply, respectively, bounded weighted measurement noise and bounded weighted regressor noise. Note, however, that the bounds guaranteed by Corollary 2 are often loose and it is preferable in practice to directly analyze the weighted measurement noise and weighted regressor noise.

Corollary 3: Assume there exist $k_0 \geq 0$ and $0 < \gamma_{\min} < \gamma_{\max}$ such that, for all $k \geq k_0$, (20) holds. Then, the following statements hold.

1) If there exists $\delta_y \geq 0$ such that, for all $k \geq k_0$, $\|\delta_{y,k}\| \leq \delta_y$, then, for all $k \geq k_0$, $\|\bar{\delta}_{y,k}\| \leq \delta_y/\sqrt{\gamma_{\min}}$.
2) If there exists $\delta_\phi \geq 0$ such that $(\delta_{\phi,k})_{k=k_0}^\infty$ is bounded with upper bound δ_ϕ, then $(\bar{\delta}_{\phi,k})_{k=k_0}^\infty$ is bounded with upper bound $\frac{\delta_\phi}{\gamma_{\min}}$.

Proof: To show 1), note that, for all $k \geq k_0$, $\|\Gamma_k^{-\frac{1}{2}}\xi_k\| \leq \boldsymbol{\sigma}_{\mathbf{max}}(\Gamma_k^{-\frac{1}{2}})\|\xi_k\| \leq \frac{\xi}{\sqrt{\gamma_{\min}}}$. Lastly, to show 2), note that, for all $k \geq k_0$, $\bar{\delta}_{\phi,k}^{\mathrm{T}}\bar{\delta}_{\phi,k} = \delta_{\phi,k}^{\mathrm{T}}\Gamma_k^{-1}\delta_{\phi,k} \preceq \frac{1}{\gamma_{\min}}\delta_{\phi,k}^{\mathrm{T}}\delta_{\phi,k} \preceq \frac{\delta_\phi}{\gamma_{\min}}I_n$. ■

3) Condition A8): Condition A8) is a bound on the magnitude of the parameters being estimated. While the parameters $\theta_{\text{true},k}$ are not known, this bound can also be approximated in practice.

B. Specialization to Errors in Variables

An important specialization of Theorem 3 is the case of fixed parameters (i.e., $\delta_\theta = 0$). In this case, only the effect of measurement noise and regressor noise is considered, a problem known as errors in variables [26]. Note that the measurement noise and regressor noise may be correlated, resulting in an asymptotically biased least squares estimator [29, p. 205]. If parameters are fixed ($\delta_\theta = 0$), it follows from Theorem 3 that (26) is globally uniformly ultimately bounded with bound $\varepsilon = \varepsilon^* b\bar{\beta}^{\frac{1}{2}}(\bar{\delta}_\phi^{\frac{1}{2}}\theta_{\max} + \bar{\delta}_y)$.

More generally, Theorem 3 can be specialized to assume fixed parameters by setting $\delta_\theta = 0$ and/or to assume no measurement noise by setting $\bar{\delta}_y = 0$ and/or to assume no regressor noise by setting $\bar{\delta}_\phi = 0$. As a sanity check, note that if $\delta_\theta = \bar{\delta}_y = \bar{\delta}_\phi = 0$, then (27) simplifies to $\varepsilon = 0$.

V. RLS Extensions as Special Cases of GF-RLS

This section shows how several extensions of RLS with forgetting are special cases of GF-RLS. For simplicity, we assume that, for all $k \geq 0$, $\Gamma_k = I_p$ in GF-RLS. This uniform weighting is in accordance with the RLS extensions we present as originally published. However, these methods can easily be extended to nonuniform weighting by, for all $k \geq 0$, selecting positive-definite $\Gamma_k \in \mathbb{R}^{p\times p}$. Thereafter, only the forgetting matrix F_k needs to be specified for all $k \geq 0$. Furthermore, the stability results presented in Theorem 2 and robustness results presented in Theorem 3 apply to any algorithm that is a special case of GF-RLS.

For all the following methods, for all $k \geq 0$, let $\phi_k \in \mathbb{R}^{p\times n}$ and $y_k \in \mathbb{R}^p$. Furthermore let $P_0 \in \mathbb{R}^{n\times n}$ be positive definite and $\theta_0 \in \mathbb{R}^n$. If an extension of RLS is a special case of proper GF-RLS, we say that extension is proper. Note that we have made minor notational changes to some RLS extensions in order to present all algorithms with the same notation. Otherwise, we have done our best to present all algorithms as originally published. A flowchart summary of this section is given in Fig. 1.

A. Recursive Least Squares

RLS [3] is derived by denoting the minimizer of the cost function

$$J_k(\hat{\theta}) = \sum_{i=0}^{k} \|y_i - \phi_i\hat{\theta}\|^2 + \|\theta - \theta_0\|^2_{P_0^{-1}} \tag{33}$$

by $\theta_{k+1} \triangleq \arg\min_{\hat{\theta}\in\mathbb{R}^n} J_k(\hat{\theta})$. It follows that, for all $k \geq 0$, θ_{k+1} is given by

$$P_{k+1}^{-1} = P_k^{-1} + \phi_k^{\mathrm{T}}\phi_k \tag{34}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k). \tag{35}$$

Comparing (34) and (35) to (9) and (10), it follows that RLS is a special case of GF-RLS where, for all $k \geq 0$, $\Gamma_k = I_p$ and

$$F_k = 0_{n\times n}. \tag{36}$$

Note that, for all $k \geq 0$, $P_k^{-1} \succ 0$, hence $P_k^{-1} - F_k \succ 0$ and $F_k \succeq 0$. Therefore RLS is proper.

B. Exponential Forgetting

A classical method to introduce forgetting in RLS is called *exponential forgetting*, where a forgetting factor $0 < \lambda \leq 1$ is introduced which provides exponentially higher weighting to more recent measurements and data [3], [5]. Exponential forgetting RLS is derived by denoting the minimizer of the cost function

$$J_k(\hat{\theta}) = \sum_{i=0}^{k} \lambda^{k-i}\|y_i - \phi_i\hat{\theta}\|^2 + \lambda^{k+1}\|\theta - \theta_0\|^2_{P_0^{-1}} \tag{37}$$

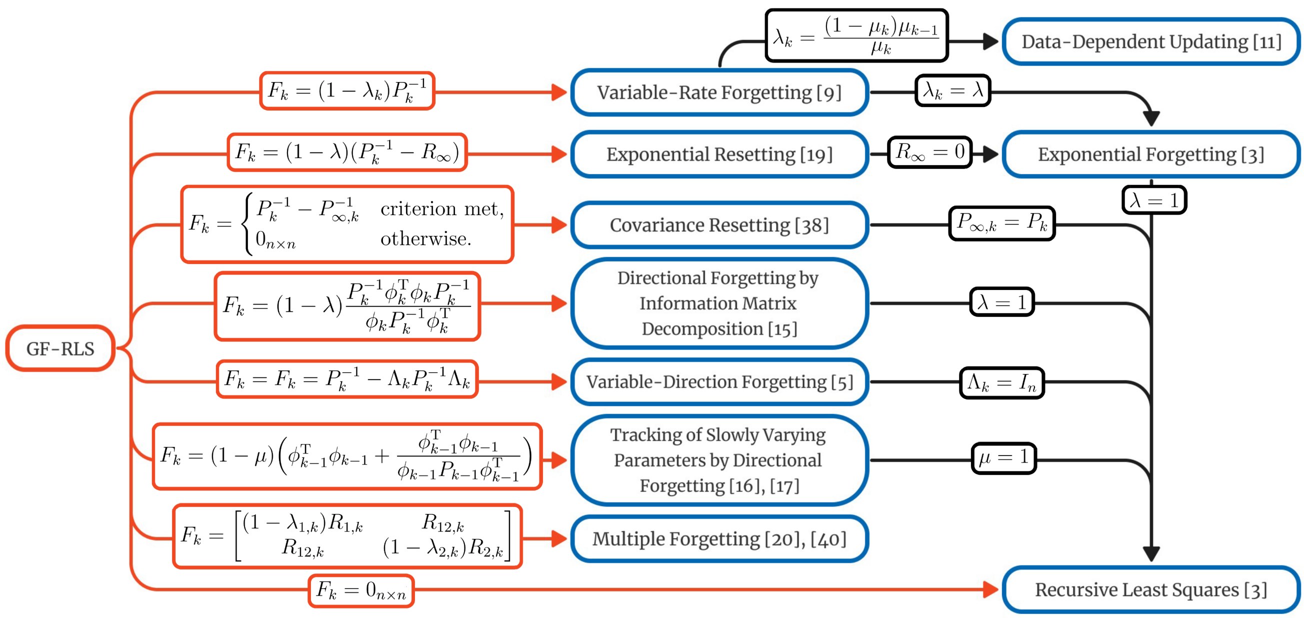

Fig. 1. This flowchart summarizes how different extensions of RLS (blue) can be derived as special cases of GF-RLS (red). Furthermore, this chart summarizes how certain RLS extensions are special cases of other RLS extensions (black).

by $\theta_{k+1} \triangleq \arg\min_{\hat\theta\in\mathbb{R}^n} J_k(\hat\theta)$. It follows that, for all $k \ge 0$, θ_{k+1} is given by

$$P_{k+1}^{-1} = \lambda P_k^{-1} + \phi_k^{\rm T}\phi_k \tag{38}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\rm T}(y_k - \phi_k\theta_k). \tag{39}$$

Comparing (38) and (39) to (9) and (10), it follows that exponential forgetting is a special case of GF-RLS, where, for all $k \ge 0$, $\Gamma_k = I_p$ and

$$F_k = (1-\lambda)P_k^{-1}. \tag{40}$$

Note that, for all $k \ge 0$, $P_k^{-1} \succ 0$, hence $P_k^{-1} - F_k = \lambda P_k^{-1} \succ 0$ and $F_k \succeq 0$. Therefore, exponential forgetting is proper.

C. Variable-Rate Forgetting

An extension of exponential forgetting is *variable-rate forgetting*, in which a time-varying forgetting factor, $0 < \lambda_k \le 1$, is selected at each step $k \ge 0$, in place of the constant forgetting factor of exponential forgetting. Variable-rate forgetting is derived in [9] by defining the cost function

$$J_k(\hat\theta) = \sum_{i=0}^{k} \frac{\rho_k}{\rho_i}\|y_i - \phi_i\hat\theta\|^2 + \rho_k\|\theta - \theta_0\|^2_{P_0^{-1}} \tag{41}$$

where, for all $k \ge 0$, $\rho_k \triangleq \prod_{i=0}^{k}\lambda_i$. If, for all $k \ge 0$, the minimizer of $J_k(\hat\theta)$ is denoted by $\theta_{k+1} \triangleq \arg\min_{\hat\theta\in\mathbb{R}^n} J_k(\hat\theta)$, it follows that, for all $k \ge 0$, θ_{k+1} is given by

$$P_{k+1}^{-1} = \lambda_k P_k^{-1} + \phi_k^{\rm T}\phi_k \tag{42}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\rm T}(y_k - \phi_k\theta_k). \tag{43}$$

Comparing (42) and (43) to (9) and (10), it follows that variable-rate forgetting is a special case of GF-RLS, where, for all $k \ge 0$, $\Gamma_k = I_p$ and

$$F_k = (1-\lambda_k)P_k^{-1}. \tag{44}$$

Note that, for all $k \ge 0$, $P_k^{-1} \succ 0$, hence $P_k^{-1} - F_k = \lambda_k P_k^{-1} \succ 0$ and $F_k \succeq 0$. Therefore, variable-rate forgetting is proper.

Many methods exist to design this time-varying forgetting factor including methods assuming known noise variance [10], online estimation of noise power [12], gradient-based methods [13], and statistical methods [14].

D. Data-Dependent Updating

Data-dependent updating was developed in [11] and was inspired as a way to prevent instabilities in the presence of bounded output disturbances. Data-dependent updating can be summarized by the update equations

$$P_{k+1}^{-1} = (1-\mu_k)P_k^{-1} + \mu_k\phi_k^{\rm T}\phi_k \tag{45}$$

$$\theta_{k+1} = \theta_k + \mu_k P_{k+1}\phi_k^{\rm T}(y_k - \phi_k\theta_k) \tag{46}$$

where, for all $k \ge 0$, $0 \le \mu_k < 1$. Next, for all $k \ge 0$, define $\bar P_k \in \mathbb{R}^{n\times n}$ by

$$\bar P_k \triangleq \mu_{k-1}P_k \tag{47}$$

where $\mu_{-1} \triangleq 1$. It then follows that, for all $k \ge 0$, (45) and (46) can be written as

$$\bar P_{k+1}^{-1} = \frac{(1-\mu_k)\mu_{k-1}}{\mu_k}\bar P_k^{-1} + \phi_k^{\rm T}\phi_k \tag{48}$$

$$\theta_{k+1} = \theta_k + \bar P_{k+1}\phi_k^{\rm T}(y_k - \phi_k\theta_k). \tag{49}$$

Comparing (48) and (49) to (42) and (43), it follows that data-dependent updating is simply a special case of variable-rate forgetting where, for all $k \geq 0$,

$$\lambda_k = \frac{(1-\mu_k)\mu_{k-1}}{\mu_k}. \tag{50}$$

For connections to GF-RLS, see Section V-C on variable-rate forgetting.

E. Exponential Resetting

Exponential resetting was developed in [19] and can be summarized by the update equations

$$P_{k+1}^{-1} = \lambda P_k^{-1} + (1-\lambda)R_\infty + \phi_k^{\mathrm{T}}\phi_k \tag{51}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k) \tag{52}$$

where $R_\infty \in \mathbb{R}^{n\times n}$ is positive semidefinite. Note that while [19] assumes that R_∞ is positive definite, it is simple to extend the results of [19] to positive-semidefinite R_∞. It is shown in [19] that, for all $k \geq 0$, P_k is positive definite. Furthermore, Lai and Bernstein [19] showed that the exponential resetting property is satisfied, namely that if there exists $M \geq 0$ such that, for all $k \geq M$, $\phi_k = 0_{p\times n}$, then $\lim_{k\to\infty} P_k^{-1} = R_\infty$.

Comparing (51) and (52) to (9) and (10), it follows that exponential resetting is a special case of GF-RLS, where, for all $k \geq 0$, $\Gamma_k = I_p$ and

$$F_k = (1-\lambda)(P_k^{-1} - R_\infty). \tag{53}$$

Note that, for all $k \geq 0$, $P_k^{-1} - F_k = \lambda P_k^{-1} + (1-\lambda)R_\infty \succ 0$. Furthermore, [19, Proposition 6] shows that, for all $k \geq 0$, $P_k^{-1} \succeq \lambda^k P_0^{-1} + (1-\lambda^k)R_\infty$. Note that if $P_0^{-1} \succeq R_\infty$, then, for all $k \geq 0$, $P_k^{-1} \succeq \lambda^k R_\infty + (1-\lambda^k)R_\infty = R_\infty$ implying that $F_k \succeq 0$. Therefore, if $P_0^{-1} \succeq R_\infty$, then exponential resetting is proper.

F. Covariance Resetting

A simple ad-hoc extension of RLS is *covariance resetting* [36], where, if a criterion for resetting is met at step k, then the covariance matrix P_k is reset to a desired positive-definite matrix, $P_{\infty,k} \in \mathbb{R}^{n\times n}$. Covariance resetting gives, for all $k \geq 0$, the update equations

$$P_{k+1}^{-1} = \begin{cases} P_{\infty,k}^{-1} + \phi_k^{\mathrm{T}}\phi_k, & \text{criterion is met} \\ P_k^{-1} + \phi_k^{\mathrm{T}}\phi_k, & \text{otherwise} \end{cases} \tag{54}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k). \tag{55}$$

Comparing (54) and (55) to (9) and (10), it follows that covariance resetting is a special case of GF-RLS where, for all $k \geq 0$, $\Gamma_k = I_p$ and

$$F_k = \begin{cases} P_k^{-1} - P_{\infty,k}^{-1}, & \text{criterion is met} \\ 0_{n\times n}, & \text{otherwise.} \end{cases} \tag{56}$$

Note that, for all $k \geq 0$,

$$P_k^{-1} - F_k = \begin{cases} P_{\infty,k}^{-1}, & \text{criterion is met} \\ P_k^{-1}, & \text{otherwise} \end{cases} \tag{57}$$

and hence $P_k^{-1} - F_k \succeq 0$. Moreover, note that when a criterion for resetting is met, $F_k \succeq 0$ if and only of $P_k \preceq P_{\infty,k}$. Thus, if $P_k \preceq P_{\infty,k}$ whenever a criterion for resetting is met, then covariance resetting is proper.

Covariance resetting can similarly be applied to any RLS extension, resetting the covariance when a criterion is met, and following the nominal algorithm otherwise. Such an algorithm would also be a special case of GF-RLS.

G. Directional Forgetting by Information Matrix Decomposition

A directional forgetting algorithm based on the decomposition of the information matrix (i.e., inverse of the covariance matrix) is presented in [15]. This method was developed in the special case of scalar measurements ($p = 1$) and can be summarized by the update equations

$$R_{k+1} = \bar{R}_k + \phi_k^{\mathrm{T}}\phi_k \tag{58}$$

$$P_{k+1} = \bar{P}_k - \frac{\bar{P}_k\phi_k^{\mathrm{T}}\phi_k\bar{P}_k}{1+\phi_k\bar{P}_k\phi_k^{\mathrm{T}}} \tag{59}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k) \tag{60}$$

where

$$\bar{R}_k \triangleq \begin{cases} R_k - (1-\lambda)\frac{R_k\phi_k^{\mathrm{T}}\phi_k R_k}{\phi_k R_k \phi_k^{\mathrm{T}}}, & \|\phi_k\| > \varepsilon \\ R_k, & \|\phi_k\| \leq \varepsilon \end{cases} \tag{61}$$

$$\bar{P}_k \triangleq \begin{cases} P_k + \frac{1-\lambda}{\lambda}\frac{\phi_k^{\mathrm{T}}\phi_k}{\phi_k R_k \phi_k^{\mathrm{T}}}, & \|\phi_k\| > \varepsilon \\ P_k, & \|\phi_k\| \leq \varepsilon \end{cases} \tag{62}$$

and where $\varepsilon > 0$, $0 < \lambda \leq 1$ and, for all $k \geq 0$, $R_k = P_k^{-1}$ and $\bar{R}_k = \bar{P}_k^{-1}$.

Comparing (58) and (60) to (9) and (10), it follows that directional forgetting by information matrix decomposition is a special case of GF-RLS, where, for all $k \geq 0$, $\Gamma_k = I_p$ and $F_k = R_k - \bar{R}_k$. If $\|\phi_k\| > \varepsilon$, then F_k can be expressed

$$F_k = (1-\lambda)\frac{P_k^{-1}\phi_k^{\mathrm{T}}\phi_k P_k^{-1}}{\phi_k P_k^{-1}\phi_k^{\mathrm{T}}} \tag{63}$$

otherwise, $F_k = 0_{n\times n}$. It is shown in [15] that, for all $k \geq 0$, $\bar{R}_k = P_k^{-1} - F_k \succ 0$ and $F_k \succeq 0$. Therefore, directional forgetting by information matrix decomposition is proper.

H. Variable-Direction Forgetting

Variable-direction forgetting was developed in [5] and is based on the singular value decomposition of the inverse covariance matrix P_k^{-1}. For all $k \geq 0$, a positive-definite $\Lambda_k \in \mathbb{R}^{n\times n}$ is constructed for the update equations

$$P_{k+1}^{-1} = \Lambda_k P_k^{-1}\Lambda_k + \phi_k^{\mathrm{T}}\phi_k \tag{64}$$

$$\theta_{k+1} = \theta_k + P_{k+1}\phi_k^{\mathrm{T}}(y_k - \phi_k\theta_k). \tag{65}$$

Details on constructing Λ_k can be found in [5, eqs. (67) and (68)]. Comparing (64) and (65) to (9) and (10), it follows that variable-direction forgetting is a special case of GF-RLS where, for all $k \geq 0$, $\Gamma_k = I_p$ and

$$F_k = P_k^{-1} - \Lambda_k P_k^{-1}\Lambda_k. \tag{66}$$

Note that, for all $k \geq 0$, $P_k^{-1} - F_k = \Lambda_k P_k^{-1} \Lambda_k \succ 0$. Moreover, it is shown in the proof of [5, Proposition 9] that, for all $k \geq 0$, $P_k^{-1} - \Lambda_k P_k^{-1} \Lambda_k \succeq 0$. Therefore, variable-direction forgetting is proper.

I. Tracking of Slowly Varying Parameters by Directional Forgetting

Another directional forgetting method, developed in [16] and analyzed in [17], was designed to track slowly varying parameters. A simulation study of this method can also be found in [37]. This method was developed in the special case of scalar measurements ($p = 1$) and can be summarized by the update equations

$$P_{k+1}^{-1} = P_k^{-1} + \beta_k \phi_k^{\mathrm{T}} \phi_k \tag{67}$$

$$\theta_{k+1} = \theta_k + \frac{1}{1 + \phi_k P_k \phi_k^{\mathrm{T}}} P_k \phi_k^{\mathrm{T}} (y_k - \phi_k \theta_k) \tag{68}$$

where, for all $k \geq 0$,

$$\beta_k \triangleq \begin{cases} \mu - \frac{1-\mu}{\phi_k P_k \phi_k^{\mathrm{T}}}, & \phi_k P_k \phi_k^{\mathrm{T}} > 0 \\ 1, & \phi_k P_k \phi_k^{\mathrm{T}} = 0 \end{cases} \tag{69}$$

and $0 < \mu \leq 1$ is the forgetting factor. To show this method is a special case of GF-RLS, first note that, for all $k \geq 0$, (85) of Lemma 2 can be used to rewrite (68) as

$$\theta_{k+1} = \theta_k + (P_k^{-1} + \phi_k^{\mathrm{T}} \phi_k)^{-1} \phi_k (y_k - \phi_k \theta_k). \tag{70}$$

Next, defining $\bar{P}_0 \triangleq P_0$ and, for all $k \geq 0$, $\bar{P}_{k+1}^{-1} \triangleq P_k^{-1} + \phi_k^{\mathrm{T}} \phi_k$. It then follows that, for all $k \geq 0$, (70) and (67) can be rewritten as

$$\bar{P}_{k+1}^{-1} = \bar{P}_k^{-1} - (1 - \beta_{k-1}) \phi_{k-1}^{\mathrm{T}} \phi_{k-1} + \phi_k^{\mathrm{T}} \phi_k \tag{71}$$

$$\theta_{k+1} = \theta_k + \bar{P}_{k+1} \phi_k^{\mathrm{T}} (y_k - \phi_k \theta_k) \tag{72}$$

where $\beta_{-1} \triangleq 0$ and $\phi_{-1} \triangleq 0_{1 \times n}$.

Comparing (71) and (72) to (9) and (10), it follows that this direction forgetting method is a special case of GF-RLS where, for all $k \geq 0$, $\Gamma_k = I_p$ and $F_k = (1 - \beta_{k-1}) \phi_{k-1}^{\mathrm{T}} \phi_{k-1}$. If $\phi_k P_k \phi_k^{\mathrm{T}} > 0$, then F_k simplifies to

$$F_k = (1 - \mu) \left(\phi_{k-1}^{\mathrm{T}} \phi_{k-1} + \frac{\phi_{k-1}^{\mathrm{T}} \phi_{k-1}}{\phi_{k-1} P_{k-1} \phi_{k-1}^{\mathrm{T}}} \right) \tag{73}$$

otherwise, $F_k = 0_{n \times n}$. It is shown in [17] that, for all $k \geq 0$, $P_k \succ 0$. Therefore, for all $k \geq 0$, $\bar{P}_k^{-1} - F_k = \bar{P}_{k+1}^{-1} - \phi_k^{\mathrm{T}} \phi_k = P_k^{-1} \succ 0$. Furthermore, $\mu \leq 1$, and hence, for all $k \geq 0$, $F_k \succeq 0$. Therefore, this direction forgetting method is proper.

J. Multiple Forgetting

Multiple forgetting was developed in [20] for the special case $n = 2$ and $p = 1$ to allow for different forgetting factors for the two parameters being estimated. To introduce multiple forgetting, we write, for all $k \geq 0$, $\phi_k \in \mathbb{R}^{1 \times 2}$ as

$$\phi_k = \begin{bmatrix} \phi_{1,k} & \phi_{2,k} \end{bmatrix}. \tag{74}$$

Then, multiple forgetting can be summarized by the update equations

$$R_{1,k+1} = \lambda_{1,k} R_{1,k} + \phi_{1,k}^2 \tag{75}$$

$$R_{2,k+1} = \lambda_{2,k} R_{2,k} + \phi_{2,k}^2 \tag{76}$$

$$\theta_{k+1} = \theta_k + L_{\mathrm{new},k} (y_k - \phi_k \theta_k) \tag{77}$$

where, for all $k \geq 0$, $\lambda_{1,k}, \lambda_{2,k} \in (0, 1]$, $R_{1,k}, R_{2,k} \in (0, \infty)$, and

$$L_{\mathrm{new},k} \triangleq \frac{1}{1 + \frac{\phi_{1,k}^2}{\lambda_{1,k} R_{1,k}} + \frac{\phi_{2,k}^2}{\lambda_{2,k} R_{2,k}}} \begin{bmatrix} \frac{\phi_{1,k}}{\lambda_{1,k} R_{1,k}} \\ \frac{\phi_{2,k}}{\lambda_{2,k} R_{2,k}} \end{bmatrix}. \tag{78}$$

It was further shown in [38] that (75) through (78) are equivalent to the update equations

$$R_{k+1} = \begin{bmatrix} \lambda_{1,k} R_{1,k} & 0 \\ 0 & \lambda_{2,k} R_{2,k} \end{bmatrix} + \phi_k^{\mathrm{T}} \phi_k \tag{79}$$

$$\theta_{k+1} = \theta_k + P_{k+1} \phi_k^{\mathrm{T}} (y_k - \phi_k \theta_k) \tag{80}$$

where, for all $k \geq 0$, $R_k \in \mathbb{R}^{2 \times 2}$ is positive definite and $P_k \triangleq R_k^{-1} \in \mathbb{R}^{2 \times 2}$. Furthermore, for all $k \geq 0$, denote R_k as

$$R_k \triangleq \begin{bmatrix} R_{1,k} & R_{12,k} \\ R_{12,k} & R_{2,k} \end{bmatrix}. \tag{81}$$

Note that (79) and (80) are equivalent to the GF-RLS update (9) and (10) where, for all $k \geq 0$, $\Gamma_k = I_p$ and

$$F_k = \begin{bmatrix} (1 - \lambda_{1,k}) R_{1,k} & R_{12,k} \\ R_{12,k} & (1 - \lambda_{2,k}) R_{2,k} \end{bmatrix}. \tag{82}$$

Furthermore, note that, for all $k \geq 0$,

$$P_k^{-1} - F_k = \begin{bmatrix} \lambda_{1,k} R_{1,k} & 0 \\ 0 & \lambda_{2,k} R_{2,k} \end{bmatrix} \succ 0 \tag{83}$$

since the diagonal elements of positive-definite R_k are positive. Note that, for all $k \geq 0$, F_k is not necessarily positive semidefinite. However, for all $k \geq 0$, since R_k is positive definite, there exist $\lambda_{1,k}, \lambda_{2,k} \in (0, 1]$ small enough such that F_k is positive semidefinite. Hence, if, for all $k \geq 0$, $\lambda_{1,k}, \lambda_{2,k}$ are chosen sufficiently small, then multiple forgetting is proper.

VI. Conclusion

This article develops GF-RLS, a general framework for RLS extensions derived from minimizing a least-squares cost function. Several RLS extensions are shown to be special cases of GF-RLS, and hence, can be derived from the GF-RLS cost function. It is important to note that while the update equations of an RLS extension may not, at face value, seem to be a special case of the GF-RLS update equations, they may still be a special case with some redefinitions. For example, see Sections V-D, V-I, and V-J. This connects a cost function to many RLS extensions which were originally developed as ad-hod modifications to the RLS update equations (e.g., [7], [8], [11], [12], [15], [16], [17], [19], [20]).

Further, stability and robustness guarantees are presented for GF-RLS. These guarantees facilitate stability and robustness analysis for various RLS extension that are a special cases of

GF-RLS. Furthermore, a specialization of the robustness result presented gives a bound to the asymptotic bias of the least squares estimator in the errors-in-variables problem. Applications of this analysis include RLS-based adaptive control [39], [40] and online transfer function identification [41], [42]. Similar analysis may be used to derive tighter bounds if specialized to a single extension of RLS.

A practical use of this work is that conditions A1)–A3) provide a general guideline for the future design of RLS extensions, backed by theoretical analysis. Satisfying conditions A1)–A3) is entirely dependent on the design of the RLS algorithm while conditions A4) through A8) only concern the data being collected. Hence, if an extension of RLS is designed to satisfy A1)–A3), then the theoretical guarantees of Theorems 2 and 3 follow under the assumption of conditions A4) through A8).

Appendix A
Useful Lemmas

Lemma 1: Let $A \in \mathbb{R}^{n\times n}$ be positive definite, let $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$, and define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) \triangleq x^{\mathrm{T}}Ax + 2b^{\mathrm{T}}x + c$. Then, f has a unique global minimizer given by $\arg\min_{x\in\mathbb{R}^n} f(x) = -A^{-1}b$.

Lemma 2 (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 $V \in \mathbb{R}^{p\times n}$. If A, C, and $A + \mathrm{UCV}$ are nonsingular, then $C^{-1} + VA^{-1}U$ is nonsingular, and

$$(A + \mathrm{UCV})^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \quad (84)$$

$$(A + \mathrm{UCV})^{-1}UC = A^{-1}U(C^{-1} + VA^{-1}U)^{-1}. \quad (85)$$

Lemma 3: Let $A \in \mathbb{R}^{n\times m}$ be the partitioned matrix

$$A \triangleq \begin{bmatrix} A_{11} & \dots & A_{1l} \\ \vdots & \ddots & \vdots \\ A_{k1} & \dots & A_{kl} \end{bmatrix} \quad (86)$$

where, for all $i \in \{1, \dots, k\}$ and $j \in \{1, \dots, l\}$, $A_{ij} \in \mathbb{R}^{n_i\times m_j}$. Then,

$$\boldsymbol{\sigma}_{\mathbf{max}}(A)^2 \le \sum_{i=1}^{k}\sum_{j=1}^{l} \boldsymbol{\sigma}_{\mathbf{max}}(A_{ij})^2. \quad (87)$$

Proof: See [43, Th. 1]. ■

Appendix B
Discrete-Time Stability Theory

Let $f : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}^n$ and consider the system

$$x_{k+1} = f(k, x_k) \quad (88)$$

where, for all $k \ge 0$, $x_k \in \mathbb{R}^n$, and $f(k, \cdot)$ is continuous.

Definition 4: For $x_{\mathrm{eq}} \in \mathbb{R}^n$, $x_k \equiv x_{\mathrm{eq}}$ is an *equilibrium* of system (88) if, for all $k \ge 0$, $f(k, 0) = 0$.

The following definition is given by [44, Definition 13.7 pp. 783 and 784].

Definition 5: If $x_k \equiv 0$ is an equilibrium of (88), then define the following.

1) The equilibrium $x_k \equiv 0$ of (88) is *Lyapunov stable* if, for all $\varepsilon > 0$, $k_0 \ge 0$, and $x_{k_0} \in \mathbb{R}^n$, there exists $\delta > 0$ such that, if $\|x_{k_0}\| < \delta$, then, for all $k \ge k_0$, $\|x_k\| < \varepsilon$.
2) The equilibrium $x_k \equiv 0$ of (88) is *uniformly Lyapunov stable* if, for all $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $k_0 \ge 0$ and $x_{k_0} \in \mathbb{R}^n$, if $\|x_{k_0}\| < \delta$, then, for all $k \ge k_0$, $\|x_k\| < \varepsilon$.
3) The equilibrium $x_k \equiv 0$ of (88) is *globally asymptotically stable* if it is Lyapunov stable and, for all $k_0 \ge 0$ and $x_{k_0} \in \mathbb{R}^n$, $\lim_{k\to\infty} x_k = 0$.
4) The equilibrium $x_k \equiv 0$ of (88) is *globally uniformly exponentially stable* if there exist $\alpha > 0$ and $\beta > 1$ such that, for all $k_0 \ge 0$, $x_{k_0} \in \mathbb{R}^n$, and $k \ge k_0$, $\|x_k\| \le \alpha\|x_{k_0}\|\beta^{-k}$.

The following result is a specialization of [44, Th. 13.11, pp. 784, 785].

Theorem 4: Let $\mathcal{D} \subset \mathbb{R}^n$ be an open set such that $0 \in \mathcal{D}$ and let $x_k \equiv 0$ be an equilibrium of (88). Furthermore, let $V : \mathbb{N}_0 \times \mathbb{R} \to \mathbb{R}$ and assume that, for all $k \in \mathbb{N}_0$, $V(k, \cdot)$ is continuous. Then, the following statements hold.

1) If there exists $\alpha > 0$ such that, for all $k \ge 0$ and $x \in \mathcal{D}$

$$V(k, 0) = 0 \quad (89)$$

$$\alpha\|x\|^2 \le V(k, x) \quad (90)$$

$$V(k + 1, f(k, x)) - V(k, x) \le 0 \quad (91)$$

then the equilibrium $x_k \equiv 0$ of (88) is Lyapunov stable.

2) If there exist $\alpha > 0$, and $\beta > 0$ such that, for all $k \ge 0$ and $x \in \mathcal{D}$, (90), (91), and

$$V(k, x) \le \beta\|x\|^2 \quad (92)$$

then the equilibrium $x_k \equiv 0$ of (88) is uniformly Lyapunov stable.

3) If there exist $\alpha > 0$, and $\gamma > 0$ such that, for all $k \ge 0$ and $x \in \mathbb{R}^n$, (89), (90), and

$$V(k + 1, f(k, x)) - V(k, x) \le -\gamma\|x\|^2 \quad (93)$$

then the equilibrium $x_k \equiv 0$ of (88) is globally asymptotically stable.

4) If there exist $\alpha > 0$, $\beta > 0$, $\gamma > 0$ such that, for all $k \ge 0$ and $x \in \mathbb{R}^n$, (90), (92), and (93), then the equilibrium $x_k \equiv 0$ of (88) is globally uniformly exponentially stable.

The following definition is given by [44, Definition 13.9, pp. 789, 790].

Definition 6: The system (88) is *globally uniformly ultimately bounded with bound* ε if, for all $\delta \in (0, \infty)$, there exists $K > 0$ such that, for all $k_0 \ge 0$ and $x_{k_0} \in \mathbb{R}^n$, if $\|x_{k_0}\| < \delta$, then, for all $k \ge k_0 + K$, $\|x_k\| < \varepsilon$.

The following result is a specialization of [44, Corollary 13.5, pp. 790, 791].

Theorem 5: Let $V : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}$ and assume that, for all $k \in \mathbb{N}_0$, $V(k, \cdot)$ is continuous. Furthermore, assume that, for all $k \ge 0$ and $x \in \mathbb{R}^n$,

$$\alpha\|x\|^2 \le V(k, x) \le \beta\|x\|^2. \quad (94)$$

Furthermore, assume there exist $\mu > 0$ and a continuous function $W : \mathbb{R}^n \to \mathbb{R}$ such that, for all $k \ge 0$ and $\|x\| > \mu$, $W(x) >$

0 and

$$V(k+1, f(k,x)) - V(k,x) \le -W(x). \tag{95}$$

Finally, assume that $\sup_{(k,x)\in\mathbb{N}_0\times\bar{\mathcal{B}}_\mu(0)} V(k+1, f(k,x))$ exists, where $\bar{\mathcal{B}}_\mu(0) \triangleq \{x\in\mathbb{R}^n : \|x\| \le \mu\}$. Then, for all ε such that[2]

$$\varepsilon \ge \max\left\{\mu, \sqrt{\sup_{(k,x)\in\mathbb{N}_0\times\bar{\mathcal{B}}_\mu(0)} V(k+1, f(k,x))}\right\} \tag{96}$$

the system (88) is globally uniformly ultimately bounded with bound ε.

Next, $k \ge 0$, define $f_k : \mathbb{R}^n \to \mathbb{R}^n$ by, for all $x \in \mathbb{R}^n$

$$f_k(x) = f(k,x). \tag{97}$$

Furthermore, let $N \ge 1$ and, for all $l = 0, 1, \ldots, N-1$, define $f_l^N : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}^n$ by, for all $j \ge 0$ and $x \in \mathbb{R}^n$

$$f_l^N(j,x) \triangleq (f_{jN+l+N-1} \circ \cdots \circ f_{jN+l+1} \circ f_{jN+l})(x) \tag{98}$$

and note that, for all $j \ge 0$, $f_l^N(j,\cdot)$ is continuous. Also note that, for all $j \ge 0$,

$$x_{(j+1)N+l} = f_l^N(j, x_{jN+l}). \tag{99}$$

In other words, f_l^N can be used to evolve the states of (88) at time steps $\{l, N+l, 2N+l, \ldots\}$. Finally, for all $l = 0, 1, \ldots, N-1$ and $j \ge 0$, define $x_{l,j}^N \in \mathbb{R}^n$ by

$$x_{l,j}^N \triangleq x_{jN+l} \tag{100}$$

which gives the system

$$x_{l,j+1}^N = f_l^N(j, x_{l,j}^N). \tag{101}$$

Lemma 4: Let $N \ge 1$, and assume that, for all $l = 0, \ldots, N-1$, the system (101) is globally uniformly ultimately bounded with bound ε. Then, (88) is globally uniformly ultimately bounded with bound ε.

Proof: Let $\delta_0 \in (0,\infty)$, let $k_0 \ge 0$, and let $x_{k_0} \in \mathbb{R}^n$. Assume that $\|x_{k_0}\| < \delta_0$. Note that there exist $j_0 \ge 0$ and $l_0 \in \{0, \ldots, N-1\}$ such that $k_0 \triangleq j_0 N + l_0$, and it follows from the assumption that the system $x_{l_0,j+1}^N = f^N(j, x_{l_0,j}^N)$ is globally uniformly ultimately bounded with bound ε. Hence, there exists $J_0 \ge 0$ such that, for all $j \ge j_0 + J_0$, $\|x_{jN+l_0}\| < \varepsilon$. Equivalently, for all $j \ge J_0$, $\|x_{k_0+jN}\| < \varepsilon$.

Next, for all $i = 1, 2, \ldots, N-1$, note that there exists $\delta_i \in (0,\infty)$ such that $\|x_{k_0+i}\| < \delta_i$. By similar reasoning as before, there exists $J_i \ge 0$ such that, for all $j \ge J_i$, $\|x_{k_0+i+jN}\| < \varepsilon$.

Finally, let $K \triangleq (\max\{J_0, \ldots, J_{N-1}\} + 1)N$. Note that, for all $k \ge k_0 + K$, there exist $i \in \{0, 1, \ldots, N-1\}$ and $j \ge J_i$ such that $k = k_0 + i + jN$, and hence $\|x_k\| < \varepsilon$. ■

Appendix C
Proof of Theorem 1

Proof of Theorem 1: First note that it follows from (4) that $J_0(\hat{\theta})$ can be written as $J_0(\hat{\theta}) = \hat{\theta}^{\mathrm{T}} H_0 \hat{\theta} + 2b_0^{\mathrm{T}}\hat{\theta} + c_0$, where

$$H_0 \triangleq \phi_0^{\mathrm{T}}\Gamma_0^{-1}\phi_0 + P_0^{-1} - F_0$$

$$b_0 \triangleq -\phi_0^{\mathrm{T}}\Gamma_0^{-1}y_0 - (P_0^{-1} - F_0)\theta_0$$

$$c_0 \triangleq y_0^{\mathrm{T}}\Gamma_0^{-1}y_0 + \theta_0^{\mathrm{T}}(P_0^{-1} - F_0)\theta_0.$$

Defining $P_1 \triangleq H_0^{-1}$, it follows that (9) holds for $k = 0$. Furthermore, it follows from (3) with $k = 0$ that $P_0^{-1} - F_k \succ 0$, and hence $H_0 \succeq P_0^{-1} - F_k \succ 0$. Therefore, Lemma 1 implies that J_0 has the unique minimizer $\theta_1 \in \mathbb{R}^n$ given by

$$\begin{aligned}\theta_1 &= -H_0^{-1}b_0 = P_1[\phi_0^{\mathrm{T}}\Gamma_0^{-1}y_0 + (P_0^{-1} - F_0)\theta_0]\\ &= P_1[\phi_0^{\mathrm{T}}\Gamma_0^{-1}y_0 + (P_0^{-1} - F_0 + \phi_0^{\mathrm{T}}\Gamma_0^{-1}\phi_0)\theta_0 - \phi_0^{\mathrm{T}}\Gamma_0^{-1}\phi_0\theta_0]\\ &= P_1[\phi_0^{\mathrm{T}}\Gamma_0^{-1}y_0 + P_1^{-1}\theta_0 - \phi_0^{\mathrm{T}}\Gamma_0^{-1}\phi_0\theta_0]\\ &= \theta_0 + P_1\phi_0^{\mathrm{T}}\Gamma_0^{-1}(y_0 - \phi_0\theta_0).\end{aligned}$$

Hence, (10) is satisfied for $k = 0$.

Now, let $k \ge 1$. Note that $J_k(\hat{\theta})$, given by (4), can be written as $J_k(\hat{\theta}) = \hat{\theta}^{\mathrm{T}} H_k \hat{\theta} + 2b_k^{\mathrm{T}}\hat{\theta} + c_k$, where

$$H_k = \sum_{i=0}^{k}\left(\phi_i^{\mathrm{T}}\Gamma_i^{-1}\phi_i - F_i\right) + P_0^{-1}$$

$$b_k = \sum_{i=0}^{k}\left(-\phi_i^{\mathrm{T}}\Gamma_i^{-1}y_i + F_i\theta_i\right) - P_0^{-1}\theta_0$$

$$c_k = \sum_{i=0}^{k}\left(y_i^{\mathrm{T}}\Gamma_i^{-1}y_i - \theta_i^{\mathrm{T}}F_i\theta_i\right) + \theta_0^{\mathrm{T}}P_0^{-1}\theta_0.$$

Furthermore, H_k and b_k can be written recursively as

$$H_k = H_{k-1} - F_k + \phi_k^{\mathrm{T}}\Gamma_k^{-1}\phi_k$$

$$b_k = b_{k-1} - \phi_k^{\mathrm{T}}\Gamma_k^{-1}y_k + F_k\theta_k.$$

Defining $P_{k+1} \triangleq H_k^{-1}$, it follows that (9) is satisfied. Furthermore, it follows from (3) that H_k is positive definite. Therefore, Lemma 1 implies that J_k has the unique minimizer θ_{k+1} given by

$$\begin{aligned}\theta_{k+1} &= -H_k^{-1}b_k = -P_{k+1}b_k\\ &= -P_{k+1}(b_{k-1} - \phi_k^{\mathrm{T}}\Gamma_k^{-1}y_k + F_k\theta_k)\\ &= P_{k+1}(P_k^{-1}\theta_k + \phi_k^{\mathrm{T}}\Gamma_k^{-1}y_k - F_k\theta_k)\\ &= P_{k+1}\left[(P_{k+1}^{-1} - \phi_k^{\mathrm{T}}\Gamma_k^{-1}\phi_k + F_k)\theta_k + \phi_k^{\mathrm{T}}\Gamma_k^{-1}y_k - F_k\theta_k\right]\\ &= P_{k+1}\left[P_{k+1}^{-1}\theta_k - \phi_k^{\mathrm{T}}\Gamma_k^{-1}\phi_k\theta_k + \phi_k^{\mathrm{T}}\Gamma_k^{-1}y_k\right]\\ &= \theta_k + P_{k+1}\phi_k^{\mathrm{T}}\Gamma_k^{-1}(y_k - \phi_k\theta_k).\end{aligned}$$

Hence, (10) is satisfied. ■

Appendix D
Proof of 1) and 2) of Theorem 2

Lemma 5: For all $k \ge 0$, let $F_k \succeq 0$ and assume there exists $b \in (0,\infty)$ such that $(P_k^{-1} - F_k)^{-1} \preceq bI_n$. Then, $P_k \preceq bI_n$.

Proof: It follows from $(P_k^{-1} - F_k)^{-1} \preceq bI_n$ that $P_k^{-1} - F_k \succeq \frac{1}{b}I_n$, and hence $P_k^{-1} \succeq \frac{1}{b}I_n + F_k \succeq \frac{1}{b}I_n$. Therefore, $P_k \preceq bI_n$. ■

[2]Note that [44, Corollary 13.5] writes $\sup_{(k,x)\in\cdots} V(k, f(k,x))$ which is a typo that has been verified with the author W. M. Haddad of [44].

For all $k \geq 0$, define $\Delta V_k \in \mathbb{R}^{n\times n}$ by

$$\Delta V_k \triangleq -M_k^{\mathrm{T}} P_{k+1}^{-1} M_k + P_k^{-1} \tag{102}$$

where, for all $k \geq 0$, M_k is defined in (16).

Lemma 6: For all $k \geq 0$,

$$\Delta V_k \succeq F_k + \frac{\bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k}{1 + \boldsymbol{\lambda}_{\mathbf{max}}(\bar{\phi}_k \bar{\phi}_k^{\mathrm{T}}) \boldsymbol{\lambda}_{\mathbf{max}}((P_k^{-1} - F_k)^{-1})}. \tag{103}$$

Proof: Let $k \geq 0$. It follows from substituting (19) into (102) that ΔV_k can be expanded as

$$\Delta V_k = -P_{k+1}^{-1} + 2\bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k - \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k P_{k+1} \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k + P_k^{-1}. \tag{104}$$

It then follows from substituting (9) into (104) that

$$\begin{aligned} \Delta V_k &= F_k + \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k - \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k P_{k+1} \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k \\ &= F_k + \bar{\phi}_k^{\mathrm{T}} (I_p - \bar{\phi}_k P_{k+1} \bar{\phi}_k^{\mathrm{T}}) \bar{\phi}_k. \end{aligned} \tag{105}$$

Next, define $G_k \in \mathbb{R}^{p\times p}$ by

$$G_k \triangleq I_p + \bar{\phi}_k (P_k^{-1} - F_k)^{-1} \bar{\phi}_k^{\mathrm{T}}. \tag{106}$$

It follows from (3) and (11) that $P_k^{-1} - F_k \succ 0_{n\times n}$. Therefore, $\bar{\phi}_k (P_k^{-1} - F_k)^{-1} \bar{\phi}_k^{\mathrm{T}} \succeq 0$. It then follows that $G_k \succeq I_p$ and hence G_k is nonsingular. Next, it follows from substituting (9) into (106) that:

$$G_k = I_p + \bar{\phi}_k [P_{k+1}^{-1} - \bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k]^{-1} \bar{\phi}_k^{\mathrm{T}}. \tag{107}$$

Finally, applying (84) of Lemma 2 to (107) gives that

$$G_k^{-1} = I_p - \bar{\phi}_k P_{k+1} \bar{\phi}_k^{\mathrm{T}}. \tag{108}$$

Substituting (108) into (105), it follows that:

$$\Delta V_k = F_k + \bar{\phi}_k^{\mathrm{T}} G_k^{-1} \bar{\phi}_k. \tag{109}$$

Finally, it follows from (106) that:

$$G_k \preceq \left[1 + \boldsymbol{\lambda}_{\mathbf{max}}(\bar{\phi}_k^{\mathrm{T}} \bar{\phi}_k) \boldsymbol{\lambda}_{\mathbf{max}}\left((P_k^{-1} - F_k)^{-1}\right)\right] I_p. \tag{110}$$

Combining (109) and (110) yields (103). ■

Proof of Statements 1) and 2) of Theorem 2: Define $V : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}$ by $V(k, \tilde{\theta}) \triangleq \tilde{\theta}^{\mathrm{T}} P_k^{-1} \tilde{\theta}$. Note that, for all $k \geq 0$,

$$V(k, 0) = 0. \tag{111}$$

Next, from (15), we define $f : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}^n$ by $f(k, \tilde{\theta}) \triangleq M_k \tilde{\theta}$. Note that, for all $k \geq 0$ and $\tilde{\theta} \in \mathbb{R}^n$,

$$V(k+1, f(k, \tilde{\theta})) - V(k, \tilde{\theta}) = -\tilde{\theta}^{\mathrm{T}} \Delta V_k \tilde{\theta}.$$

Then, for all $k \geq 0$ and $\tilde{\theta} \in \mathbb{R}^n$, it follows from Lemma 6 and condition A1) that:

$$V(k+1, f(k, \tilde{\theta})) - V(k, \tilde{\theta}) \leq -\tilde{\theta}^{\mathrm{T}} F_k \tilde{\theta} \leq 0. \tag{112}$$

We now prove statements 1) and 2) as follows.

1) By Lemma 5, conditions A1) and A2) imply that, for all $k \geq 0$ and $\tilde{\theta} \in \mathbb{R}^n$

$$\frac{1}{b} \|\tilde{\theta}\|^2 \leq V(k, \tilde{\theta}). \tag{113}$$

Equations (111)–(113) imply that (89)–(91) are satisfied. It then follows from part 4) of Theorem 4 that the equilibrium $\tilde{\theta}_k \equiv 0$ of (15) is Lyapunov stable.

2) A3) further implies that, for all $k \geq 0$ and $\tilde{\theta} \in \mathbb{R}^n$

$$V(k, \tilde{\theta}) \leq \frac{1}{a} \|\tilde{\theta}\|^2. \tag{114}$$

Equations (112)–(114) imply that (90)–(92) are satisfied. By part 2) of Theorem 4, it follows that the equilibrium $\tilde{\theta}_k \equiv 0$ of (15) is uniformly Lyapunov stable. ■

Appendix E
Proof of 3) and 4) of Theorem 2

For all $k \geq 0$ and $N \geq 1$, define $\Delta^N V_k \in \mathbb{R}^{n\times n}$ by

$$\Delta^N V_k \triangleq -M_k^{\mathrm{T}} \cdots M_{k+N-1}^{\mathrm{T}} P_{k+N}^{-1} M_{k+N-1} \cdots M_k + P_k^{-1}. \tag{115}$$

Next, all $k \geq 0$, $i \geq 1$, define $W_{k,i} \in \mathbb{R}^{p\times p}$ by

$$W_{k,i} \triangleq \bar{\phi}_{k+i} P_{k+1} \bar{\phi}_k^{\mathrm{T}}. \tag{116}$$

For all $k \geq 0$, define $\Phi_{k,1} \triangleq \bar{\phi}_k$, $\Psi_{k,1} \triangleq \bar{\phi}_k$, and $\mathcal{W}_{k,N} = I_p$. Furthermore, for all $k \geq 0$ and $N \geq 2$, define $\Phi_{k,N} \in \mathbb{R}^{Np\times n}$, $\Psi_{k,N} \in \mathbb{R}^{Np\times Np}$, and $\mathcal{W}_{k,N} \in \mathbb{R}^{Np\times Np}$ by

$$\Phi_{k,N} \triangleq \begin{bmatrix} \bar{\phi}_k \\ \vdots \\ \bar{\phi}_{k+N-1} \end{bmatrix} \tag{117}$$

$$\Psi_{k,N} \triangleq \begin{bmatrix} \bar{\phi}_k \\ \bar{\phi}_{k+1} M_k \\ \vdots \\ \bar{\phi}_{k+N-1} M_{k+N-2} \cdots M_{k+1} M_k \end{bmatrix} \tag{118}$$

$$\mathcal{W}_{k,N} \triangleq \begin{bmatrix} I_p & 0 & 0 & \cdots & 0 \\ W_{k,1} & I_p & 0 & \cdots & 0 \\ W_{k,2} & W_{k+1,1} & I_p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_{k,N-1} & W_{k+1,N-2} & W_{k+2,N-3} & \cdots & I_p \end{bmatrix}. \tag{119}$$

Lemma 7: For all $k \geq 0$ and $N \geq 1$,

$$\mathcal{W}_{k,N} \Psi_{k,N} = \Phi_{k,N} \tag{120}$$

Proof: For all $k \geq 0$, and $N = 1$, (120) simplifies to $\bar{\phi}_k = \bar{\phi}_k$. Next, note that, for all $k \geq 0$ and $N \geq 2$, $\mathcal{W}_{k,N} \Psi_{k,N}$ can be written as

$$\mathcal{W}_{k,N} \Psi_{k,N} = \begin{bmatrix} \bar{\phi}_k \\ \begin{bmatrix} W_{k,1} & I_p \end{bmatrix} \Psi_{k,2} \\ \begin{bmatrix} W_{k,2} & W_{k+1,1} & I_p \end{bmatrix} \Psi_{k,3} \\ \vdots \\ \begin{bmatrix} W_{k,N-1} & \cdots & W_{k+N-2,1} & I_p \end{bmatrix} \Psi_{k,N} \end{bmatrix}. \tag{121}$$

Next, (116) implies that, for all $N \geq 2$ and $0 \leq i \leq N-2$,

$$\begin{aligned} W_{k+i,N-1-i} \bar{\phi}_{k+i} &= \bar{\phi}_{k+N-1} P_{k+i+1} \bar{\phi}_{k+i}^{\mathrm{T}} \bar{\phi}_{k+i} \\ &= \bar{\phi}_{k+N-1} (I_p - M_{k+i}). \end{aligned}$$

Substituting this identity into the left-hand side of (122), it follows that, for all $N \geq 2$,

$$\begin{aligned}&\left[W_{k,N-1} \quad W_{k+1,N-2} \quad \cdots \quad W_{k+N-2,1}\right] \Psi_{k,N-1}\\&= W_{k,N-1}\bar{\phi}_k + \sum_{i=0}^{N-2} W_{k+i,N-1-i}\bar{\phi}_{k+i}M_{k+i-1}\cdots M_k\\&= \bar{\phi}_{k+N-1}\left(I_p - M_k + \sum_{i=1}^{N-2}(I - M_{k+i})M_{k+i-1}\cdots M_k\right).\end{aligned}$$

Note that this forms a telescoping series, which, by cancellation of successive terms, simplifies to

$$\begin{aligned}&\left[W_{k,N-1} \quad W_{k+1,N-2} \quad \cdots \quad W_{k+N-2,1}\right] \Psi_{k,N-1}\\&\qquad = \bar{\phi}_{k+N-1}(I_p - M_{k+N-2}\cdots M_{k+1}M_k).\end{aligned} \tag{122}$$

Next, adding $\bar{\phi}_{k+N-1}M_{k+N-2}\cdots M_{k+1}M_k$ to both sides of (122) implies that, for all $N \geq 2$,

$$\begin{aligned}&\left[W_{k,N-1} \quad W_{k+1,N-2} \quad \cdots \quad W_{k+N-2,1} \quad I_p\right] \Psi_{k,N}\\&\qquad = \bar{\phi}_{k+N-1}.\end{aligned} \tag{123}$$

Hence, applying (123) to the row partitions of (121) yields (120). ∎

Lemma 8: Assume that, for all $k \geq 0$, $F_k \succeq 0_{n\times n}$. Also assume there exists $b \in (0,\infty)$ such that, for all $k \geq 0$, $P_k \preceq bI_n$. Furthermore, assume $(\bar{\phi}_k)_{k=0}^{\infty}$ is persistently exciting with lower bound $\bar{\alpha} > 0$ and persistency window N and bounded with upper bound $\bar{\beta} \in (0,\infty)$. Then, for all $k \geq 0$ and $N \geq 1$

$$\Delta^N V_k \succeq c_N I_n \succ 0_{n\times n} \tag{124}$$

where

$$c_N \triangleq \frac{\bar{\alpha}}{N}(1+b\bar{\beta})^{-1}\left[1 + \frac{N-1}{2}\left(b\bar{\beta}\right)^2\right]^{-1}. \tag{125}$$

Proof: To begin, we show that, for all $k \geq 0$ and $N \geq 1$

$$\Delta^N V_k \succeq (1+b\bar{\beta})^{-1}\Psi_{k,N}^{\mathrm{T}}\Psi_{k,N}. \tag{126}$$

For brevity, we define $\nu \triangleq (1+b\bar{\beta})^{-1}$. Proof of (126) follows by induction on N. First, let $k \geq 0$ and consider the base case $N = 1$. Note that $\Delta^1 V_k = \Delta V_k$ and $\Psi_{k,1} = \bar{\phi}_k$. Hence, it follows from Lemma 6 that $\Delta^1 V_k \succeq F_k + \nu\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k \succeq \nu\Psi_{k,1}^{\mathrm{T}}\Psi_{k,1}$. Next, let $N \geq 2$. Note that $\Delta^N V_k$, given by (115) can be expressed recursively as

$$\begin{aligned}\Delta^N V_k &= M_k^{\mathrm{T}}(\Delta^{N-1}V_{k+1} - P_{k+1}^{-1})M_k + P_k^{-1}\\&= M_k^{\mathrm{T}}\Delta^{N-1}V_{k+1}M_k + \Delta V_k.\end{aligned}$$

It follows from inductive hypothesis that $\Delta^{N-1}V_{k+1} \succeq \nu\Psi_{k+1,N-1}^{\mathrm{T}}\Psi_{k+1,N-1}$. Substituting into the previous equation gives

$$\begin{aligned}\Delta^N V_k &\succeq \nu M_k^{\mathrm{T}}\Psi_{k+1,N-1}^{\mathrm{T}}\Psi_{k+1,N-1}M_k + \nu\bar{\phi}_k^{\mathrm{T}}\bar{\phi}_k\\&= \nu\left[\bar{\phi}_k^{\mathrm{T}} \quad M_k^{\mathrm{T}}\Psi_{k+1,N-1}^{\mathrm{T}}\right]\begin{bmatrix}\bar{\phi}_k\\ \Psi_{k+1,N-1}M_k\end{bmatrix}.\end{aligned}$$

Note that (118) implies that $[\bar{\phi}_k^{\mathrm{T}} \quad M_k^{\mathrm{T}}\Psi_{k+1,N-1}^{\mathrm{T}}] = \Psi_{k,N}^{\mathrm{T}}$ and (126) is proven.

Next, note that, for all $k \geq 0$ and $N \geq 1$, $\mathcal{W}_{k,N}$ is lower triangular with all ones on the main diagonal, and hence is nonsingular. Thus, Lemma 7 implies that, for all $k \geq 0$ and $N \geq 1$, $\Psi_{k,N} = \mathcal{W}_{k,N}^{-1}\Phi_{k,N}$, and thus

$$\Psi_{k,N}^{\mathrm{T}}\Psi_{k,N} \succeq \frac{\Phi_{k,N}^{\mathrm{T}}\Phi_{k,N}}{\boldsymbol{\sigma}_{\mathbf{max}}(\mathcal{W}_{k,N})^2}. \tag{127}$$

Note that if $N = 1$ then, for all $k \geq 0$, $\mathcal{W}_{k,1} = I_p$ and hence

$$\boldsymbol{\sigma}_{\mathbf{max}}(\mathcal{W}_{k,1})^2 = 1. \tag{128}$$

Next, if $N \geq 2$ then Lemma 3 implies that $\boldsymbol{\sigma}_{\mathbf{max}}(\mathcal{W}_{k,N})^2 \leq N\boldsymbol{\sigma}_{\mathbf{max}}(I_p)^2 + \sum_{i=1}^{N-1}\sum_{j=1}^{N-i}\boldsymbol{\sigma}_{\mathbf{max}}(W_{k-1+i,j})^2$. Note that, for all $1 \leq i \leq N-1$ and $1 \leq j \leq N-i$, it follows from (116) that $\boldsymbol{\sigma}_{\mathbf{max}}(W_{k-1+i,j}) \leq b\bar{\beta}$. Using this inequality, it follows that, for all $k \geq 0$ and $N \geq 2$, $\boldsymbol{\sigma}_{\mathbf{max}}(\mathcal{W}_{k,N})^2 \leq N + \sum_{i=1}^{N-1}\sum_{j=1}^{N-i}(b\bar{\beta})^2$, which simplifies to

$$\boldsymbol{\sigma}_{\mathbf{max}}(\mathcal{W}_{k,N})^2 \leq N + \frac{N(N-1)}{2}\left(b\bar{\beta}\right)^2. \tag{129}$$

It then follows from (128) and (129) that, for all $k \geq 0$ and $N \geq 1$

$$\frac{1}{\boldsymbol{\sigma}_{\mathbf{max}}(\mathcal{W}_{k,N})^2} \geq \frac{1}{N}\left[1 + \frac{N-1}{2}\left(b\bar{\beta}\right)^2\right]^{-1}. \tag{130}$$

Furthermore, persistent excitation of $(\bar{\phi}_k)_{k=0}^{\infty}$ implies that, for all $k \geq 0$ and $N \geq 1$

$$\Phi_{k,N}^{\mathrm{T}}\Phi_{k,N} \succeq \bar{\alpha}I_n. \tag{131}$$

Finally, substituting (127), (130), and (131) into (126) yields (124). ∎

Proof of Statements 3) and 4) of Theorem 2: Note that repeated substitution of (15) gives, for all $j \geq 0$

$$\tilde{\theta}_{(j+1)N} = M_{(j+1)N-1}\cdots M_{jN}\tilde{\theta}_{jN}.$$

Hence, we define $f^N : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}^n$ by, for all $j \geq 0$ and $\tilde{\theta} \in \mathbb{R}^n$,

$$f^N(j,\tilde{\theta}) \triangleq M_{jN+N-1}\cdots M_{jN+1}M_{jN}\tilde{\theta}.$$

Further, for all $j \geq 0$, define $\tilde{\theta}_j^N \in \mathbb{R}^n$ by $\tilde{\theta}_j^N \triangleq \tilde{\theta}_{jN}$, which yields the system

$$\tilde{\theta}_{j+1}^N = f^N(j,\tilde{\theta}_j^N). \tag{132}$$

Next, define $V^N : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}$ by

$$V^N(j,\tilde{\theta}) \triangleq \tilde{\theta}^{\mathrm{T}}P_{jN}^{-1}\tilde{\theta}$$

and note that, for all $j \geq 0$,

$$V^N(j,0) = 0. \tag{133}$$

Also, note that, for all $j \geq 0$,

$$V^N\left(j+1, f^N(j,\tilde{\theta})\right) - V^N\left(j,\tilde{\theta}\right) = -\tilde{\theta}^{\mathrm{T}}\Delta^N V_{jN}\tilde{\theta}$$

where $\Delta^N V_{jN} \in \mathbb{R}^{n\times n}$ is defined in (115). It then follows from Lemma 8 that, for all $j \geq 0$,

$$V^N\left(j+1, f^N(j,\tilde\theta)\right) - V^N\left(j,\tilde\theta\right) \leq -c_N\|\tilde\theta\|^2 \quad (134)$$

where $c_N > 0$ is defined in (125). Next, it follows from Lemma 5 that, for all $k \geq 0$, $P_k \preceq bI_n$ and it follows then from (15) and (19) that, for all $k \geq 0$

$$\begin{aligned}\|\tilde\theta_{k+1}\| &\leq (\boldsymbol{\sigma}_{\mathbf{max}}(I_n) + \boldsymbol{\sigma}_{\mathbf{max}}(P_{k+1})\boldsymbol{\sigma}_{\mathbf{max}}(\bar\phi_k^{\mathrm{T}}\bar\phi_k))\|\tilde\theta_k\| \\ &\leq (1+b\bar\beta)\|\tilde\theta_k\|. \end{aligned}\quad (135)$$

Hence, (135) implies that, for all $j \geq 0$ and $l = 1,\ldots,N-1$

$$\|\tilde\theta_{jN+l}\| \leq \left(1+b\bar\beta\right)^{N-1}\|\tilde\theta_{jN}\|. \quad (136)$$

We now prove statements 3) and 4) of Theorem 2 as follows.

3) By Lemma 5, conditions A1) and A2) implies that, for all $j \geq 0$ and $\tilde\theta \in \mathbb{R}^n$

$$\frac{1}{b}\|\tilde\theta\|^2 \leq V^N(j,\tilde\theta). \quad (137)$$

Equations (133), (134), and (137) imply that (89), (90), and (93) are satisfied. Hence, by part 3) of Theorem 4, the equilibrium $\tilde\theta_j^N \equiv 0$ of (132) is globally asymptotically stable.

We now show that the equilibrium $\tilde\theta_k \equiv 0$ of (15) is also globally asymptotically stable. Let $\varepsilon > 0$ and $k_0 \geq 0$. Write $k_0 = j_0 N + l_0$, where $j_0 \geq 0$, and $0 \leq l_0 \leq N-1$. Since the equilibrium $\tilde\theta_j^N \equiv 0$ of (132) is Lyapunov stable, we can choose δ such that $\|\tilde\theta_{j_0 N}\| < \delta$ implies that, for all $j \geq j_0$,

$$\|\tilde\theta_{jN}\| < \varepsilon(1+b\bar\beta)^{\frac{1}{N-1}}. \quad (138)$$

Let $\|\tilde\theta_{k_0}\| < \delta$. For all $k^* \geq k_0$, write $k^* = j^*N + l^*$ and note that $j^* \geq j_0$. It then follows from (136) and (138) that, for all $k^* \geq k_0$,

$$\begin{aligned}\|\tilde\theta_{k^*}\| &\leq (1+b\bar\beta)^{N-1}\|\tilde\theta_{j^*N}\| \\ &< (1+b\bar\beta)^{N-1}\varepsilon(1+b\bar\beta)^{\frac{1}{N-1}} = \varepsilon.\end{aligned}$$

Hence, the equilibrium $\tilde\theta_k \equiv 0$ of (15) is Lyapunov stable. A similar argument using (136) can be used to show that $\lim_{j\to\infty}\tilde\theta_{jN} = 0$ implies that $\lim_{k\to\infty}\tilde\theta_k = 0$. Therefore, the equilibrium $\tilde\theta_k \equiv 0$ of (15) is globally asymptotically stable.

4) Condition A3) further implies that, for all $j \geq 0$ and $\tilde\theta \in \mathbb{R}^n$,

$$V^N(j,\tilde\theta) \leq \frac{1}{a}\|\tilde\theta\|^2. \quad (139)$$

(134), (137), and (139) imply that (90), (92), and (93) are satisfied. Hence, by part 4) of Theorem 4, the equilibrium $\tilde\theta_j^N \equiv 0$ of (132) is globally uniformly exponentially stable. Using (136) in an argument similar to the proof of statement 3), it can be shown that the equilibrium $\tilde\theta_k \equiv 0$ of (15) is also globally uniformly exponentially stable. ■

Appendix F
Lemmas Used in the Proof of Theorem 3

Suppose, for all $k \geq 0$, there exists $\zeta_k \in \mathbb{R}^n$ such that,

$$\breve\theta_{k+1} = M_k(\breve\theta_k - \zeta_k) \quad (140)$$

where $\breve\theta_k$ is defined in (25) and M_k is defined in (16). Next, for all $k \geq 0$ and $i \geq 1$, define $\mathcal{M}_{k,i} \in \mathbb{R}^{n\times n}$ by

$$\mathcal{M}_{k,i} \triangleq \begin{cases} M_k & i = 1 \\ M_{k+i-1}\cdots M_{k+1}M_k & i \geq 2. \end{cases} \quad (141)$$

Moreover, for all $k \geq 0$, let $P_k^{-\frac{1}{2}} \in \mathbb{R}^{n\times n}$ be the unique positive-semidefinite matrix such that $P_k^{-1} = P_k^{-\frac{1}{2}\mathrm{T}} P_k^{-\frac{1}{2}}$.

Lemma 9: Assume, for all $k \geq 0$, there exists $\zeta_k \in \mathbb{R}^n$ such that (140) holds. Then, for all $N \geq 1$ and $k \geq 0$,

$$\breve\theta_{k+N} = \mathcal{M}_{k,N}\breve\theta_k - \overline{\mathcal{M}}_{k,N}\bar\zeta_{k,N} \quad (142)$$

where $\overline{\mathcal{M}}_{k,N} \in \mathbb{R}^{n\times Nn}$ and $\bar\zeta_{k,N} \in \mathbb{R}^{Nn\times 1}$ are defined

$$\overline{\mathcal{M}}_{k,N} \triangleq \begin{bmatrix}\mathcal{M}_{k,N} & \mathcal{M}_{k+1,N-1} & \cdots & \mathcal{M}_{k+N-1,1}\end{bmatrix} \quad (143)$$

$$\bar\zeta_{k,N} \triangleq \begin{bmatrix}\zeta_k^{\mathrm{T}} & \zeta_{k+1}^{\mathrm{T}} & \cdots & \zeta_{k+N-1}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}. \quad (144)$$

Proof: Let $k \geq 0$ and proof follows by induction on $N \geq 1$. First, consider the base case $N = 1$. Note that $\overline{\mathcal{M}}_{k,1} = \mathcal{M}_{k,1} = M_k$ and $\bar\zeta_{k,1} = \zeta_k$. Hence, (142) follows immediately from (140). Next, let $N \geq 2$. By inductive hypothesis, $\breve\theta_{k+N-1} = \mathcal{M}_{k,N-1}\breve\theta_k - \overline{\mathcal{M}}_{k,N-1}\bar\zeta_{k,N-1}$. Furthermore, it follows from (140) that $\breve\theta_{k+N} = M_{k+N-1}(\breve\theta_{k+N-1} - \zeta_{k+N-1})$. Combining these two equalities gives

$$\begin{aligned}\breve\theta_{k+N} &= M_{k+N-1}\mathcal{M}_{k,N-1}\breve\theta_k \\ &\quad + \begin{bmatrix} M_{k+N-1}\overline{\mathcal{M}}_{k,N-1} & M_{k+N-1}\end{bmatrix}\begin{bmatrix}\bar\zeta_{k,N-1} \\ \zeta_{k+N-1}\end{bmatrix}\end{aligned}$$

which can be rewritten as (142). ■

Note that from (140) and Lemma 9, it follows that, for all $l = 0,1,\ldots,N-1$ and $j \geq 0$,

$$\breve\theta_{(j+1)N+l} = \mathcal{M}_{jN+l,N}\breve\theta_{jN+l} - \overline{\mathcal{M}}_{jN+l,N}\bar\zeta_{jN+l,N}. \quad (145)$$

Next, for all $l = 0,1,\ldots,N-1$ and $j \geq 0$, define $\breve\theta_{l,j}^N \in \mathbb{R}^n$ by

$$\breve\theta_{l,j}^N \triangleq \breve\theta_{jN+l} \quad (146)$$

which, for all $l = 0,1,\ldots,N-1$, gives the system

$$\breve\theta_{l,j+1}^N = \mathcal{M}_{jN+l,N}\breve\theta_{l,j}^N - \overline{\mathcal{M}}_{jN+l,N}\bar\zeta_{jN+l,N}. \quad (147)$$

Lemma 10: Assume, for all $k \geq 0$, there exists $\zeta_k \in \mathbb{R}^n$ such that (140) holds. Furthermore, assume conditions A1)–A4) hold and assume there exists $\zeta \geq 0$ such that, for all $k \geq 0$,

$$\|\zeta_k\| \leq \zeta. \quad (148)$$

Then, for all $l = 0,1,\ldots,N-1$, the system (147) is globally uniformly ultimately bounded with bound $\varepsilon^*\zeta$, where ε^* is given by (28).

Proof: For brevity, we prove the case $l = 0$. The cases $l = 1,\ldots,N-1$ can be shown similarly.

Define $\check{f}^N : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\check{f}^N(j, \check{\theta}) \triangleq \mathcal{M}_{jN,N}\check{\theta} - \overline{\mathcal{M}}_{jN,N}\bar{\zeta}_{jN,N} \tag{149}$$

and note that, for all $j \geq 0$, $\check{\theta}^N_{0,j+1} = f^N(j, \check{\theta}^N_{0,j})$. Furthermore, define $V^N : \mathbb{N}_0 \times \mathbb{R}^n \to \mathbb{R}$ by

$$V^N(j, \check{\theta}) \triangleq \check{\theta}^{\mathrm{T}} P^{-1}_{jN}\check{\theta}. \tag{150}$$

Note that, for all $j \geq 0$ and $\check{\theta} \in \mathbb{R}^n$,

$$\frac{1}{b}\|\check{\theta}\|^2 \leq V^N(j, \check{\theta}) \leq \frac{1}{a}\|\check{\theta}\|^2 \tag{151}$$

where the lower bound follows from Lemma 5 and conditions A1) and A2) and where the upper bound follows from condition A3).

Next, by substituting (149) into (150), it follows that, for all $j \geq 0$ and $\check{\theta} \in \mathbb{R}^n$,

$$\begin{aligned} &V^N(j+1, \check{f}^N(j, \check{\theta})) - V^N(j, \check{\theta}) \\ &= -\check{\theta}^{\mathrm{T}}\Delta^N V_{jN}\check{\theta} - 2\check{\theta}^{\mathrm{T}}\mathcal{M}^{\mathrm{T}}_{jN,N}P^{-1}_{(j+1)N}\overline{\mathcal{M}}_{jN,N}\bar{\zeta}_{jN,N} \\ &\quad + \bar{\zeta}^{\mathrm{T}}_{jN,N}\overline{\mathcal{M}}^{\mathrm{T}}_{jN,N}P^{-1}_{(j+1)N}\overline{\mathcal{M}}_{jN,N}\bar{\zeta}_{jN,N} \end{aligned} \tag{152}$$

where $\Delta^N V_{jN} \in \mathbb{R}^{n\times n}$ is defined in (115).

Next, since conditions A1)–A4) hold, it then follows from Lemma 8 that, for all $k \geq 0$ and $N \geq 1$,

$$-\Delta^N V_k = \mathcal{M}^{\mathrm{T}}_{k,N}P^{-1}_{k+N}\mathcal{M}_{k,N} - P^{-1}_k \preceq -c_N I_n \tag{153}$$

where $c_N > 0$ is defined in (125). It then follows from (153) that, for all $k \geq 0$ and $N \geq 1$,

$$\begin{aligned} \mathcal{M}^{\mathrm{T}}_{k,N}P^{-1}_{k+N}\mathcal{M}_{k,N} &= \left(P^{-\frac{1}{2}}_{k+N}\mathcal{M}_{k,N}\right)^{\mathrm{T}}(P^{-\frac{1}{2}}_{k+N}\mathcal{M}_{k,N}) \\ &\preceq P^{-1}_k - c_N I_n \preceq \left(\frac{1}{a} - c_N\right)I_n. \end{aligned} \tag{154}$$

Therefore, (154) implies that, for all $k \geq 0$ and $N \geq 1$

$$\boldsymbol{\sigma}_{\mathbf{max}}(P^{-\frac{1}{2}}_{k+N}\mathcal{M}_{k,N}) \leq \sqrt{\frac{1}{a} - c_N}. \tag{155}$$

Next, it follows from (143) that, for all $j \geq 0$

$$\begin{aligned} &P^{-\frac{1}{2}}_{(j+1)N}\overline{\mathcal{M}}_{jN,N} \\ &= \left[P^{-\frac{1}{2}}_{(j+1)N}\mathcal{M}_{jN,N} \quad \cdots \quad P^{-\frac{1}{2}}_{(j+1)N}\mathcal{M}_{jN+N-1,1}\right]. \end{aligned} \tag{156}$$

Applying Lemma 3 to (156) and substituting the bound (155) then gives

$$\boldsymbol{\sigma}_{\mathbf{max}}\left(P^{-\frac{1}{2}}_{(j+1)N}\overline{\mathcal{M}}_{jN,N}\right) \leq \sqrt{\sum_{i=1}^{N}\left(\frac{1}{a} - c_i\right)}. \tag{157}$$

Furthermore, it is easy to verify from definition (125) that, for all $i \geq 1$, $c_{i+1} < c_i$. Therefore, for all $j \geq 0$,

$$\boldsymbol{\sigma}_{\mathbf{max}}\left(P^{-\frac{1}{2}}_{(j+1)N}\overline{\mathcal{M}}_{jN,N}\right) \leq \sqrt{N\left(\frac{1}{a} - c_N\right)}. \tag{158}$$

Bounds (154) and (158) then imply that, for all $j \geq 0$,

$$\mathcal{M}^{\mathrm{T}}_{jN,N}P^{-1}_{(j+1)N}\overline{\mathcal{M}}_{jN,N} \preceq \sqrt{N}\left(\frac{1}{a} - c_N\right)I_n \tag{159}$$

$$\overline{\mathcal{M}}^{\mathrm{T}}_{jN,N}P^{-1}_{(j+1)N}\overline{\mathcal{M}}_{jN,N} \preceq N\left(\frac{1}{a} - c_N\right)I_n. \tag{160}$$

Finally, it follows from (148) and (144) that, for all $j \geq 0$,

$$\|\bar{\zeta}_{jN,N}\| \leq \sqrt{N}\zeta. \tag{161}$$

Substituting (153), (159), (160), and (161) into (152) and applying the Cauchy Schwarz inequality gives, for all $j \geq 0$ and $\check{\theta} \in \mathbb{R}^n$, the bound

$$\begin{aligned} &V^N\left(j+1, \check{f}^N\left(j, \check{\theta}\right)\right) - V^N\left(j, \check{\theta}\right) \\ &\leq -c_N\|\check{\theta}\|^2 + 2N\zeta\left(\frac{1}{a} - c_N\right)\|\check{\theta}\| + N^2\zeta^2\left(\frac{1}{a} - c_N\right). \end{aligned} \tag{162}$$

Next, note that $\Delta_N \in \mathbb{R}$, defined in (29), can be written as

$$\Delta_N = \frac{1}{ac_N} - 1. \tag{163}$$

It then follows from (162) that, for all $j \geq 0$ and $\check{\theta} \in \mathbb{R}^n$,

$$\begin{aligned} &\frac{1}{c_N}\left[V^N(j+1, \check{f}^N(j, \check{\theta})) - V^N(j, \check{\theta})\right] \\ &\leq -\|\check{\theta}\|^2 + 2N\zeta\left(\frac{1}{ac_N} - 1\right)\|\check{\theta}\| + N^2\zeta^2\left(\frac{1}{ac_N} - 1\right) \\ &= -\|\check{\theta}\|^2 + 2N\zeta\Delta_N\|\check{\theta}\| + N^2\zeta^2\Delta_N. \end{aligned} \tag{164}$$

Next, we define $\mu_N \in \mathbb{R}$ by

$$\mu_N \triangleq \left(\Delta_N + \sqrt{\Delta_N + \Delta_N^2}\right)N\zeta. \tag{165}$$

It follows from solving the quadratic equation (164) that, for all $j \geq 0$ and $\check{\theta} \in \mathbb{R}^n$, such that $\|\check{\theta}\| > \mu_N$,

$$V^N(j+1, \check{f}^N(j, \check{\theta})) - V^N(j, \check{\theta}) < 0. \tag{166}$$

Next, it follows from substituting (149) into (150) that, for all $j \geq 0$ and $\check{\theta} \in \mathbb{R}^n$,

$$\begin{aligned} &\sqrt{V^N(j+1, \check{f}^N(j, \check{\theta}))} \\ &\quad = \|\mathcal{M}_{jN,N}\check{\theta} - \overline{\mathcal{M}}_{jN,N}\bar{\zeta}_{jN,N}\|_{P^{-1}_{(j+1)N}} \\ &\quad = \|P^{-\frac{1}{2}}_{(j+1)N}\mathcal{M}_{jN,N}\check{\theta} - P^{-\frac{1}{2}}_{(j+1)N}\overline{\mathcal{M}}_{jN,N}\bar{\zeta}_{jN,N}\| \\ &\quad \leq \boldsymbol{\sigma}_{\mathbf{max}}(P^{-\frac{1}{2}}_{(j+1)N}\mathcal{M}_{jN,N})\|\check{\theta}\| \\ &\qquad + \boldsymbol{\sigma}_{\mathbf{max}}(P^{-\frac{1}{2}}_{(j+1)N}\overline{\mathcal{M}}_{jN,N})\|\bar{\zeta}_{jN,N}\| \end{aligned} \tag{167}$$

where the last inequality follows from triangle inequality. Applying inequalities (155), (158), and (161)–(167), it follows that, for all $j \geq 0$ and $\check{\theta} \in \mathbb{R}^n$,

$$\sqrt{V^N(j+1, \check{f}^N(j, \check{\theta}))} \leq \sqrt{\frac{1}{a} - c_N}\|\check{\theta}\| + \sqrt{\frac{1}{a} - c_N}N\zeta. \tag{168}$$

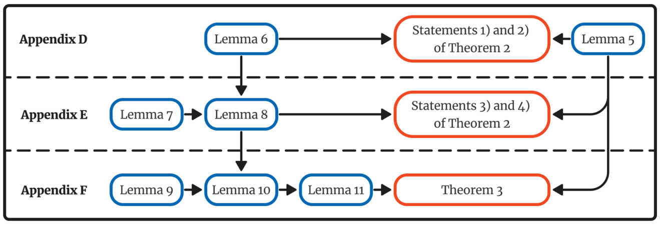

Fig. 2. Proof "roadmap" of Theorem 2 and Theorem 3.

For brevity, denote

$$\sqrt{\sup V} \triangleq \sqrt{\sup_{(j,\check{\theta})\in\mathbb{N}_0\times\bar{\mathcal{B}}_{\mu_N}(0)} V^N(j+1,\check{f}^N(j,\check{\theta}))}.$$

It then follows from (168) that

$$\sqrt{\sup V} \le \sqrt{\frac{1}{a}-c_N}\,(\mu_N+N\zeta)\,. \tag{169}$$

Substituting (165) into (169), it follows that

$$\sqrt{\sup V} \le \sqrt{\frac{1}{a}-c_N}\left(1+\Delta_N+\sqrt{\Delta_N+\Delta_N^2}\right)N\zeta. \tag{170}$$

Next, note that (163) implies that

$$\frac{1}{a}-c_N=\frac{1}{a}\frac{\Delta_N}{1+\Delta_N}. \tag{171}$$

Substituting (171) into (170) and simplifying, it follows that

$$\sqrt{\sup V} \le \frac{1}{\sqrt{a}}\left(\Delta_N+\sqrt{\Delta_N+\Delta_N^2}\right)N\zeta. \tag{172}$$

Applying Theorem 5, it follows from (151), (165), (166), and (172) that the system (147) with $l=0$ is globally uniformly ultimately bounded with bound $\varepsilon^*\zeta$, where ε^* is given by (28). ■

Lemma 11: Assume, for all $k\ge 0$, there exists $\zeta_k\in\mathbb{R}^n$ such that (140) holds. Furthermore, assume conditions A1)–A4) hold. Finally, assume there exists $\zeta\ge 0$ such that, for all $k\ge 0$, $\|\zeta_k\|\le\zeta$. Then, the system (140) is globally uniformly ultimately bounded with bound $\varepsilon^*\zeta$, where ε^* is given by (28).

Proof: It follows from Lemma 10 that, for all $l=0,1,\ldots,N-1$, the system (147) is globally uniformly ultimately bounded with bound $\varepsilon^*\zeta$, where ε^* is given by (28). It then follows from Lemma 4 that the system (140) is also globally uniformly ultimately bounded with bound $\varepsilon^*\zeta$, where ε^* is given by (28). ■

References

[1] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA, USA: MIT Press, 1983.
[2] K. J. Åström and B. Wittenmark, *Adaptive Control*. North Chelmsford, MA, USA: Courier Corporation, 2013.
[3] S. A. U. Islam and D. S. Bernstein, "Recursive least squares for real-time implementation," *IEEE Control Syst. Mag.*, vol. 39, no. 3, pp. 82–85, Jun. 2019.
[4] S. Sastry, M. Bodson, and J. F. Bartram, *Adaptive Control: Stability, Convergence, and Robustness*. Mineola, NY, USA: Dover Publications, 1990.
[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 Syst. Mag.*, vol. 40, no. 4, pp. 80–102, Aug. 2020.
[6] R. Ortega, V. Nikiforov, and D. Gerasimov, "On modified parameter estimators for identification and adaptive control. a unified framework and some new schemes," *Annu. Rev. Control*, vol. 50, pp. 278–293, 2020.
[7] M. E. Salgado, G. C. Goodwin, and R. H. Middleton, "Modified least squares algorithm incorporating exponential resetting and forgetting," *Int. J. Control*, vol. 47, no. 2, pp. 477–491, 1988.
[8] R. M. Johnstone, C. R. Johnson Jr, R. R. Bitmead, and B. D. Anderson, "Exponential convergence of recursive least squares with exponential forgetting factor," *Syst. Control Lett.*, vol. 2, no. 2, pp. 77–82, 1982.
[9] A. L. Bruce, A. Goel, and D. S. Bernstein, "Convergence and consistency of recursive least squares with variable-rate forgetting," *Automatica*, vol. 119, 2020, Art. no. 109052.
[10] T. 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.
[11] 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. Inf. Theory*, vol. 33, no. 3, pp. 383–392, May 1987.
[12] C. Paleologu, J. Benesty, and S. Ciochina, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, 2008.
[13] S.-H. Leung and C. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3141–3150, Aug. 2005.
[14] N. Mohseni and D. S. Bernstein, "Recursive least squares with variable-rate forgetting based on the f-test," in *Proc. Amer. Control Conf.*, 2022, pp. 3937–3942.
[15] L. Cao and H. Schwartz, "A directional forgetting algorithm based on the decomposition of the information matrix," *Automatica*, vol. 36, no. 11, pp. 1725–1731, 2000.
[16] R. Kulhavy and M. Kárny, "Tracking of slowly varying parameters by directional forgetting," *IFAC Proc. Volumes*, vol. 17, no. 2, pp. 687–692, 1984.
[17] 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.
[18] H.-S. Shin and H.-I. Lee, "A new exponential forgetting algorithm for recursive least-squares parameter estimation," 2020, *arXiv:2004.03910*.
[19] B. Lai and D. S. Bernstein, "Exponential resetting and cyclic resetting recursive least squares," *IEEE Contr. Syst. Lett.*, vol. 7, pp. 985–990, 2023.
[20] A. Vahidi, A. Stefanopoulou, and H. Peng, "Recursive least squares with forgetting for online estimation of vehicle mass and road grade: Theory and experiments," *Veh. System Dyn.*, vol. 43, no. 1, pp. 31–55, 2005.
[21] J. Parkum, N. K. Poulsen, and J. Holst, "Recursive forgetting algorithms," *Int. J. Control*, vol. 55, no. 1, pp. 109–128, 1992.
[22] V. Shaferman, M. Schwegel, T. Glück, and A. Kugi, "Continuous-time least-squares forgetting algorithms for indirect adaptive control," *Eur. J. Control*, vol. 62, pp. 105–112, 2021.
[23] M. Bin, "Generalized recursive least squares: Stability, robustness, and excitation," *Syst. Control Lett.*, vol. 161, 2022, Art. no. 105144.
[24] C. Samson, "Stability analysis of adaptively controlled systems subject to bounded disturbances," *Automatica*, vol. 19, no. 1, pp. 81–86, 1983.
[25] J. Milek and F. J. Kraus, "Time-varying stabilized forgetting for recursive least squares identification," *IFAC Proc. Volumes*, vol. 28, no. 13, pp. 137–142, 1995.
[26] T. Söderström, "Errors-in-variables methods in system identification," *Automatica*, vol. 43, no. 6, pp. 939–958, 2007.
[27] Z. Griliches and J. A. Hausman, "Errors in variables in panel data," *J. Econometrics*, vol. 31, no. 1, pp. 93–118, 1986.
[28] J. Gillard, "An overview of linear structural models in errors in variables regression," *REVSTAT-Stat. J.*, vol. 8, no. 1, pp. 57–80, 2010.
[29] L. Ljung, *System Identification*. Berlin, Germany: Springer, 1998.
[30] B. Zhou and T. Zhao, "On asymptotic stability of discrete-time linear time-varying systems," *IEEE Trans. Autom. Control*, vol. 62, no. 8, pp. 4274–4281, Aug. 2017.
[31] R. P. Agarwal, *Difference Equations and Inequalities: Theory, Methods, and Applications*. Boca Raton, FL, USA: CRC Press, 2000.

[32] D. Efimov, N. Barabanov, and R. Ortega, "Robust stability under relaxed persistent excitation conditions," in *Proc. IEEE Conf. Decis. Control*, 2018, pp. 7243–7248.
[33] J. Wang, D. Efimov, S. Aranovskiy, and A. A. Bobtsov, "Fixed-time estimation of parameters for non-persistent excitation," *Eur. J. Control*, vol. 55, pp. 24–32, 2020.
[34] K. G. Vamvoudakis, M. F. Miranda, and J. P. Hespanha, "Asymptotically stable adaptive– optimal control algorithm with saturating actuators and relaxed persistence of excitation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2386–2398, Nov. 2016.
[35] A. L. Bruce, A. Goel, and D. S. Bernstein, "Necessary and sufficient regressor conditions for the global asymptotic stability of recursive least squares," *Syst. Control Lett.*, vol. 157, 2021, Art. no. 105005.
[36] G. Goodwin, E. Teoh, and H. Elliott, "Deterministic convergence of a self-tuning regulator with covariance resetting," *IEE Proc. D-Control Theory Appl.*, vol. 1, no. 130, pp. 6–8, 1983.
[37] D. Bertin, S. Bittanti, and P. Bolzern, "Tracking of nonstationary systems by means of different prediction error direction forgetting techniques," in *Adaptive Systems in Control and Signal Processing*. Amsterdam, Netherlands: Elsevier, 1987, pp. 185–190.
[38] F. Fraccaroli, A. Peruffo, and M. Zorzi, "A new recursive least squares method with multiple forgetting schemes," in *Proc. IEEE 54th Conf. Decis. Control*, 2015, pp. 3367–3372.
[39] T. W. Nguyen, S. A. U. Islam, D. S. Bernstein, and I. V. Kolmanovsky, "Predictive cost adaptive control: A numerical investigation of persistency, consistency, and exigency," *IEEE Control Syst. Mag.*, vol. 41, no. 6, pp. 64–96, Dec. 2021.
[40] S. A. U. Islam, T. W. Nguyen, I. V. Kolmanovsky, and D. S. Bernstein, "Data-driven retrospective cost adaptive control for flight control applications," *J. Guid. Control, Dyn.*, vol. 44, no. 10, pp. 1732–1758, 2021.
[41] D. Muller and W. W. Wei, "Iterative least squares estimation and identification of the transfer function model," *J. Time Ser. Anal.*, vol. 18, no. 6, pp. 579–592, 1997.
[42] J. C. Akers and D. S. Bernstein, "Armarkov least-squares identification," in *Proc. Amer. Control Conf.*, 1997, pp. 186–190.
[43] R. Bhatia and F. Kittaneh, "Norm inequalities for partitioned operators and an application," *Mathematische Annalen*, vol. 287, no. 1, pp. 719–726, 1990.
[44] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

Brian Lai received the B.S. degree in mechanical engineering and mathematics from Rutgers University, New Brunswick, NJ, USA, in 2020, and the M.S. degree in aerospace engineering, in 2022, from the University of Michigan, Ann Arbor, MI, USA, where he is currently working toward the Ph.D. degree.

His research interests include data-driven identification, estimation, and adaptive control for aerospace applications.

Mr. Lai was the recipient of the National Science Foundation Graduate Research Fellowship in 2021.

Dennis S. Bernstein (Life Fellow, IEEE) received the Sc.B. degree in applied mathematics from Brown University, Providence, RI, USA, in 1977, and the M.S.E. and Ph.D. degrees in computer, information, and control engineering from the University of Michigan, Ann Arbor, MI, USA, in 1979 and 1982, respectively.

He is currently the James E. Knott Professor with the Aerospace Engineering Department, University of Michigan. He has authored the book titled *Scalar, Vector, and Matrix Mathematics*, third edition (Princeton University Press, 2018). His research interests include identification, estimation, and control for aerospace applications.

Dr. Bernstein was the Editor-in-Chief of the *IEEE Control Systems Magazine* from 2003 to 2011.