aerospace

MDPI

Article

Momentum-Based Adaptive Laws for Identification and Control

Luke Somers and Wassim M. Haddad *

School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0150, USA; luke.somers@gatech.edu

* Correspondence: wm.haddad@aerospace.gatech.edu; Tel.: +1-404-894-1078

Abstract: In this paper, we develop momentum-based adaptive update laws for parameter identification and control to improve parameter estimation error convergence and control system performance for uncertain dynamical systems. Specifically, we introduce three novel continuous-time, momentum-based adaptive estimation and control algorithms and evaluate their effectiveness via several numerical examples. Our proposed adaptive architectures show faster parameter convergence rates as compared to the classical gradient descent and model reference adaptive control methods.

Keywords: parameter identification; online learning; MRAC; momentum-based learning; high-order tuners; exponential convergence

1. Introduction

One of the fundamental problems in feedback control design is the ability to address discrepancies between system models and real-world systems. To this end, adaptive control has been developed to address the problem of system uncertainty in control-system design [1–4]. In particular, adaptive control is based on constant linearly parameterized system uncertainty models of a known structure but unknown variation and, in the case of indirect adaptive control, combines an online parameter estimation algorithm with a control law to improve performance in the face of system uncertainties. More specifically, indirect adaptive controllers utilize parameter update laws to identify unknown system parameters and adjust feedback gains to account for system uncertainty.

The parameter estimation algorithms that have been typically used in adaptive control are predicated on two gradient descent type methods that update the system parameter errors in the direction that minimizes a prediction error for a specified cost function. Namely, for a *gradient descent flow algorithm*, the cost function involves an instantaneous prediction error between an estimated system output and the actual system output of the most recent data, whereas for an *integral gradient descent algorithm* the cost function captures an aggregate of the past prediction errors in an integral form while placing more emphasis on recent measurements [2].

Additionally, the recursive least squares (RLS) algorithm [2] has also been used for parameter estimation and has been shown to give superior convergence rates as compared to the classical gradient descent algorithms. RLS algorithms can be viewed as a gradient descent method with a time-varying learning rate [5]. For static (i.e., time-invariant) cost functions, *momentum-based methods*, also known as *higher-order tuners*, such as the Nesterov method, have been shown to achieve faster parameter error convergence as compared to traditional gradient descent algorithms [6]. Since faster convergence can lead to improved transient system performance, there has been a significant interest in extending momentum-based methods to time-varying cost functions, which naturally appear in adaptive control formulations [7–9].

Specifically, in [7] momentum-based architectures were merged with the standard gradient descent method using a variational approach to guarantee boundedness for the

check for updates

Citation: Somers, L.; Haddad, W.M. Momentum-Based Adaptive Laws for Identification and Control. *Aerospace* **2024**, *11*, 1017. https://doi.org/10.3390/aerospace11121017

Academic Editors: Nadjim Horri and Toufik Souanef

Received: 29 October 2024
Revised: 4 December 2024
Accepted: 9 December 2024
Published: 11 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

parameter estimation and model reference adaptive control problems involving time-varying regressors. Higher-order tuners were also shown to provide an additional degree of freedom in the design of adaptive control laws [10,11]. Given that integral gradient and recursive least squares algorithms provide superior noise rejection properties and improved system performance to that of standard gradient descent methods, there has been a recent interest in integrating momentum-based architectures into adaptive laws for parameter identification and control [5,7].

In this paper, we develop new continuous-time, momentum-based adaptive laws for identification and control by augmenting higher-order tuners into the integral gradient, recursive least squares and composite gradient algorithms. Specifically, in Section 2, we first review the existing gradient-based and momentum-based update laws and introduce three new higher-order tuner architectures. Next, in Section 3, we show how these update laws can be applied for parameter identification, and then, in Section 4, we introduce a momentum-based recursive least squares (MRLS) algorithm for model reference adaptive control. Finally, Section 5 provides several numerical examples that highlight the improved parameter error convergence rate of the proposed momentum-based update laws.

The notation used in this paper is standard. Specifically, we write $\mathbb{R}^n$ to denote the set of $n \times 1$ column vectors, $\mathbb{R}^{n\times m}$ to denote the set of $n \times m$ real matrices, and $(\cdot)^{\mathrm{T}}$ to denote transpose. The equi-induced 2-norm of a matrix $Q \in \mathbb{R}^{m\times n}$ is denoted as $\|Q\| \triangleq \sqrt{\lambda_{\max}(Q^{\mathrm{T}}Q)} = \sigma_{\max}(Q)$, where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue and $\sigma_{\max}(\cdot)$ denotes the maximum singular value. We write $\lambda_{\min}(\cdot)$ (resp., $\sigma_{\min}(\cdot)$) for the minimum eigenvalue (resp., singular value), $\frac{\partial V}{\partial x}(x)$ for the gradient of a scalar-valued function V with respect to a vector-valued variable x, and $\hat{f}(s)$ for the Laplace transform of $f(\cdot)$, that is $\hat{f}(s) \triangleq \mathcal{L}\{f(t)\} = \int_0^\infty e^{-st} f(t)\mathrm{d}t$. Finally, we write $\mathbb{R}(s)$ to denote the proper rational functions with coefficients in $\mathbb{R}$ (i.e., SISO proper rational transfer functions), $\mathbb{R}[s]$ to denote polynomials with real coefficients, $\mathcal{L}_2$ to denote the space of square-integrable Lebesgue measurable functions on $[0,\infty)$, and $\mathcal{L}_\infty$ to denote the space of bounded Lebesgue measurable functions on $[0,\infty)$.

2. From First-Order to Higher-Order Tuners for Parameter Estimation

Consider the system given by

$$y(t) = \theta^{*\mathrm{T}}\phi(t), \tag{1}$$

where, for $t \geq 0$, $y(t) \in \mathbb{R}$ is the system output, $\phi(t) \in \mathbb{R}^N$ is a time-varying regressor vector, and $\theta^* \in \mathbb{R}^N$ is an unknown system parameter vector. Here we assume that $\phi(\cdot)$ is bounded, and hence, can be normalized via the scaling factor $1 + \|\phi(t)\|$ so that, with a minor abuse of notation, $\|\phi(t)\| \leq 1$, $t \geq 0$. Since θ^* is unknown, we design an estimator of the form $y_{\mathrm{e}}(t) = \theta(t)^{\mathrm{T}}\phi(t)$, where $\theta(t)$, $t \geq 0$, is an estimate of θ^*, so that the prediction error between the estimated output $y_{\mathrm{e}}(t), t \geq 0$, and the actual system output $y(t)$, $t \geq 0$, is given by

$$e(t) \triangleq y_{\mathrm{e}}(t) - y(t) = \theta(t)^{\mathrm{T}}\phi(t) - y(t) = \tilde{\theta}(t)^{\mathrm{T}}\phi(t), \tag{2}$$

where $\tilde{\theta}(t) \triangleq \theta(t) - \theta^*$ is the parameter estimation error.

For the error system (2), consider the quadratic loss cost function

$$L(\theta, t) = \frac{1}{2}e^2(t) \tag{3}$$

and note that the gradient of L with respect to the parameter estimate θ is given by $\frac{\partial L(\theta,t)}{\partial \theta} = \phi(t)e(t)$. In this case, the gradient flow algorithm is expressed as

$$\dot{\theta}(t) = -\gamma\frac{\partial L(\theta, t)}{\partial \theta} = -\gamma\phi(t)e(t), \quad \theta(0) = \theta_0, \quad t \geq 0, \tag{4}$$

where $\gamma > 0$ is a weighting gain parameter. As shown in [7], $e \in \mathcal{L}_\infty \cap \mathcal{L}_2$ for all $\gamma > 0$.

It is well known that the rate of convergence for the gradient flow algorithm (4) can be slow [6]. However, if the regressor vector $\phi(\cdot)$ satisfies a persistency of the excitation condition, then $\theta(t)$ converges to θ^* exponentially with a convergence rate proportional to γ. Although in this case the convergence rate can improve significantly, large values of γ can result in a stiff initial value problem for (4) necessitating progressively smaller sampling times for implementing (4) on a digital computer. Furthermore, small sampling times can increase the sensitivity of the adaptive law to measurement noise and modeling errors [2]. To address these limitations, momentum-based accelerated algorithms predicated on higher-order tuner laws are introduced in [7] to expedite parameter estimation error convergence.

In [12], an array of accelerated methods for continuous-time systems are derived using a Lagrangian variational formulation. Building on this work, [7] extends the variational framework of [12] to address a time-varying instantaneous quadratic loss function of the form given by (3). Specifically, a Lagrangian of the form

$$\mathcal{L}(\theta, \dot{\theta}, t) = e^{\int_{t_0}^{t} \beta \mathcal{N}(s) \mathrm{d}s} \frac{1}{2\beta \mathcal{N}(t)} [\dot{\theta}^{\mathrm{T}}(t)\dot{\theta}(t) - \gamma \beta \mathcal{N}(t) e(t)^2] \tag{5}$$

is introduced, where $\mathcal{N}(t) \triangleq 1 + \mu \phi(t)^{\mathrm{T}} \phi(t)$, $\mu > 0$, $\beta > 0$ is a *friction coefficient*, and $\gamma > 0$, and the functional

$$J(\theta) = \int_{\mathbb{T}} \mathcal{L}(\theta, \dot{\theta}, t) \mathrm{d}s, \tag{6}$$

where $\mathbb{T}$ is a finite-time interval, is considered. A necessary condition for θ minimizing (6) is shown in [7] to satisfy

$$\ddot{\theta}(t) + \left[\beta \mathcal{N}(t) - \frac{\dot{\mathcal{N}}(t)}{\mathcal{N}(t)}\right] \dot{\theta}(t) = -\gamma \beta \mathcal{N}(t) \phi(t) e(t), \quad \dot{\theta}(0) = \dot{\theta}_0, \quad \theta(0) = \theta_0, \quad t \geq 0. \tag{7}$$

Note that (7) can be rewritten as

$$\dot{\mathcal{V}}(t) = -\gamma \phi(t) e(t), \quad \mathcal{V}(0) = \mathcal{V}_0, \quad t \geq 0, \tag{8}$$

$$\dot{\theta}(t) = -\beta [\theta(t) - \mathcal{V}(t)] \mathcal{N}(t), \quad \theta(0) = \theta_0, \tag{9}$$

where $\mathcal{V}$ is an auxiliary parameter and $\mathcal{V}_0 = \frac{1}{\mathcal{N}(0)\beta} \dot{\theta}_0 + \theta_0$. The stability properties of (8) and (9) are given in [7]. Taking the limiting case as $\beta \to \infty$, the gradient flow algorithm (4) can be recovered as a special case of (8) and (9).

Rather than relying solely on the instantaneous system measurement $y(t), t \geq 0$, the loss cost function (3) can be modified to incorporate a weighted sum of past system measurements. Specifically, we can consider loss cost functions in the form of (see [2])

$$L(\theta, t) = \frac{1}{2} \int_0^t e^{-\nu(t-\tau)} [y(\tau) - \theta^{\mathrm{T}}(t) \phi(\tau)]^2 \mathrm{d}\tau, \tag{10}$$

where $\nu > 0$ is a *forgetting factor* that prevents the degeneration of system information updates in some direction by placing more emphasis on the more recent measurements. In this case, the gradient of the loss cost function (10) is given by

$$\begin{aligned} \frac{\partial L(\theta, t)}{\partial \theta} &= \int_0^t e^{-\nu(t-\tau)} [y(\tau) - \theta^{\mathrm{T}}(t) \phi(\tau)] \phi(\tau) \mathrm{d}\tau \\ &= R(t)\theta(t) + Q(t), \end{aligned} \tag{11}$$

where

$$R(t) \triangleq \int_0^t e^{-\nu(t-\tau)} \phi(\tau) \phi(\tau)^{\mathrm{T}} \mathrm{d}\tau, \tag{12}$$

$$Q(t) \triangleq \int_0^t e^{-\nu(t-\tau)} y(\tau)\phi(\tau)^{\mathrm{T}} \mathrm{d}\tau. \tag{13}$$

Now, setting

$$\dot{\theta}(t) = -\gamma \frac{\partial L(\theta, t)}{\partial \theta}, \quad \theta(0) = \theta_0, \quad t \geq 0, \tag{14}$$

we obtain the *integral gradient algorithm* ([2])

$$\dot{\theta}(t) = -\gamma[R(t)\theta(t) + Q(t)], \quad \theta(0) = \theta_0, \quad t \geq 0, \tag{15}$$
$$\dot{R}(t) = -\nu R(t) + \phi(t)\phi(t)^{\mathrm{T}}, \quad R(0) = 0, \tag{16}$$
$$\dot{Q}(t) = -\nu Q(t) - y(t)\phi(t), \quad Q(0) = 0. \tag{17}$$

The stability properties of (15)–(17) are addressed in [2] ([Thm. 3.6.7]).

Next, motivated by the variational formulation of [7] we introduce a momentum-based integral gradient algorithm. Specifically, defining the Lagrangian

$$\mathcal{L}(\theta, \dot{\theta}, t) = e^{\int_{t_0}^t \beta \mathcal{N}_{\mathrm{i}}(s)\mathrm{d}s} \frac{1}{\beta \mathcal{N}_{\mathrm{i}}(t)} \left[\frac{1}{2} \dot{\theta}(t)^{\mathrm{T}} \dot{\theta}(t) - \gamma \beta \mathcal{N}_{\mathrm{i}}(t) \frac{1}{2} \int_0^t e^{-\nu(t-\tau)} (y(\tau) - \theta^{\mathrm{T}}(t)\phi(t))^2 \mathrm{d}\tau \right], \tag{18}$$

where $\mathcal{N}_{\mathrm{i}}(t) \triangleq 1 + \mu \|R(t)\|$, $\mu \geq 0$, and using the Euler-Lagrange equation, a necessary condition for θ minimizing (6) yields

$$\ddot{\theta}(t) + \left[\beta \mathcal{N}_{\mathrm{i}}(t) - \frac{\dot{\mathcal{N}}_{\mathrm{i}}(t)}{\mathcal{N}_{\mathrm{i}}(t)} \right] \dot{\theta}(t) = -\gamma \beta \mathcal{N}_{\mathrm{i}}(t)[R(t)\theta(t) + Q(t)], \quad \dot{\theta}(0) = \dot{\theta}_0, \quad \theta(0) = \theta_0, \quad t \geq 0, \tag{19}$$

where, for $t \geq 0$, $R(t)$ and $Q(t)$ are given by (12) and (13). Note that (19) can be rewritten as

$$\dot{\mathcal{V}}(t) = -\gamma[R(t)\theta(t) + Q(t)], \quad \mathcal{V}(0) = \mathcal{V}_0, \quad t \geq 0, \tag{20}$$
$$\dot{\theta}(t) = -\beta[\theta(t) - \mathcal{V}(t)]\mathcal{N}_{\mathrm{i}}(t), \quad \theta(0) = \theta_0, \tag{21}$$
$$\dot{R}(t) = -\nu R(t) + \phi(t)\phi(t)^{\mathrm{T}}, \quad R(0) = 0, \tag{22}$$
$$\dot{Q}(t) = -\nu Q(t) - y(t)\phi(t), \quad Q(0) = 0, \tag{23}$$

where $\mathcal{V}$ is an auxiliary parameter and $\mathcal{V}_0 = \frac{1}{\mathcal{N}_{\mathrm{i}}(0)\beta}\dot{\theta}_0 + \theta_0$. Furthermore, (20) and (21) can be rewritten in terms of the parameter estimation error $\tilde{\theta}(t)$, $t \geq 0$, and the auxiliary parameter estimation error $\tilde{\mathcal{V}}(t) \triangleq \mathcal{V}(t) - \theta^*$ as

$$\dot{\tilde{\mathcal{V}}}(t) = -\gamma R(t)\tilde{\theta}(t), \quad \tilde{\mathcal{V}}(0) = \tilde{\mathcal{V}}_0, \quad t \geq 0, \tag{24}$$
$$\dot{\tilde{\theta}}(t) = -\beta[\tilde{\theta}(t) - \tilde{\mathcal{V}}(t)]\mathcal{N}_{\mathrm{i}}(t), \quad \tilde{\theta}(0) = \tilde{\theta}_0. \tag{25}$$

The following definition and lemmas are needed for the main results of this section.

Definition 1. *A vector signal $\phi : \mathbb{R} \to \mathbb{R}^N$ is persistently excited (PE) if there exist a positive constant ρ and a finite-time $\delta > 0$ such that*

$$\rho I_N \leq \int_t^{t+\delta} \phi(\tau)\phi(\tau)^{\mathrm{T}} \mathrm{d}\tau, \quad t > 0. \tag{26}$$

Lemma 1. *Consider the system (22). Then, the following statements hold:*

(i)

$$\|R(t)\| \leq \frac{1}{\nu}, \quad t \geq 0. \tag{27}$$

(*ii*) *If* $\phi(\cdot)$ *is persistently excited, then*

$$e^{-\nu\delta}\rho \leq \|R(t)\|, \quad t \geq \delta. \tag{28}$$

Proof. To show (i), note that it follows from (12) and $\phi(t)\phi(t)^{\mathrm{T}} \leq I_N$, $t \geq 0$, that

$$R(t) \leq \int_0^t e^{-\nu(t-\tau)} I_N \mathrm{d}\tau \leq \frac{1}{\nu} I_N, \quad t \geq 0,$$

which proves (27). Next, to show (ii), note that if $\phi(\cdot)$ is PE, then

$$\begin{aligned} R(t) &= \int_0^{t-\delta} e^{-\nu(t-\tau)}\phi(\tau)\phi(\tau)^{\mathrm{T}}\mathrm{d}\tau + \int_{t-\delta}^{t} e^{-\nu(t-\tau)}\phi(\tau)\phi(\tau)^{\mathrm{T}}\mathrm{d}\tau \\ &\geq \int_{t-\delta}^{t} e^{-\nu(t-\tau)}\phi(\tau)\phi(\tau)^{\mathrm{T}}\mathrm{d}\tau \\ &\geq e^{-\nu\delta}\int_{t-\delta}^{t} \phi(\tau)\phi(\tau)^{\mathrm{T}}\mathrm{d}\tau \\ &\geq e^{-\nu\delta}\rho I_N, \quad t > \delta. \end{aligned} \tag{29}$$

This proves the lemma. □

Lemma 2 ([3])**.** *Let* $f : [0,\infty) \to \mathbb{R}^n$. *If* $f \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ *and* $\dot{f} \in \mathcal{L}_\infty$, *then* $\lim_{t\to\infty} f(t) = 0$.

Theorem 1. *Let* $\mu > \frac{2\gamma}{\beta}$, *and consider the momentum-based integral gradient algorithm* (20)–(23). *Then, the following statements hold:*

(*i*) $\theta \in \mathcal{L}_\infty, \dot{\theta}, \dot{\mathcal{V}}$, *and* $(\theta - \mathcal{V}) \in \mathcal{L}_\infty$, *and* $\lim_{t\to\infty}\dot{\theta}(t) = 0$.

(*ii*) *If* $\phi(\cdot)$ *is persistently excited, then* $\theta(t)$ *converges to* θ^* *exponentially as* $t \to \infty$ *with a convergence rate of* $\frac{a}{2}$, *where* $a \triangleq \min\{e^{-\nu\delta}\rho\gamma, 2\beta\}$.

Proof. To show (i), consider the positive-definite function $V(\tilde{\mathcal{V}}, \tilde{\theta}) \triangleq \tilde{\mathcal{V}}^{\mathrm{T}}\tilde{\mathcal{V}} + (\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}(\tilde{\theta} - \tilde{\mathcal{V}})$ and note that

$$\begin{aligned} \dot{V}(\tilde{\mathcal{V}}, \tilde{\theta}, t) &= 2\tilde{\mathcal{V}}^{\mathrm{T}}\dot{\tilde{\mathcal{V}}} + 2(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}(\dot{\tilde{\theta}} - \dot{\tilde{\mathcal{V}}}) \\ &= -2\gamma\tilde{\mathcal{V}}^{\mathrm{T}}R(t)\tilde{\theta} - 2\beta\|\tilde{\theta} - \tilde{\mathcal{V}}\|^2(1 + \mu\|R(t)\|) + 2\gamma(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}R(t)\tilde{\theta} \\ &\leq -2\gamma\tilde{\theta}^{\mathrm{T}}R(t)\tilde{\theta} - 2\beta\|\tilde{\theta} - \tilde{\mathcal{V}}\|^2(1 + \mu\|R(t)\|) + 4\gamma\|(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}R^{\frac{1}{2}}(t)\|\|R^{\frac{1}{2}}(t)\tilde{\theta}\| \\ &\leq -2\gamma\tilde{\theta}^{\mathrm{T}}R(t)\tilde{\theta} - 2\beta\|\tilde{\theta} - \tilde{\mathcal{V}}\|^2(1 + \mu\|R(t)\|) + 4\gamma\|(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}R^{\frac{1}{2}}(t)\|\|R^{\frac{1}{2}}(t)\tilde{\theta}\| \\ &\leq -\gamma\tilde{\theta}^{\mathrm{T}}R(t)\tilde{\theta} - 2\beta\|\tilde{\theta} - \tilde{\mathcal{V}}\|^2 - \gamma\left[\|R^{\frac{1}{2}}(t)\tilde{\theta}\| - 2\|(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}R^{\frac{1}{2}}(t)\|\right]^2 \\ &\leq -\gamma\tilde{\theta}^{\mathrm{T}}R(t)\tilde{\theta} - 2\beta\|\tilde{\theta} - \tilde{\mathcal{V}}\|^2, \end{aligned} \tag{30}$$

which, since $R(t) \geq 0, t \geq 0$, implies $\dot{V}(\tilde{\mathcal{V}}, \tilde{\theta}, t) \leq 0$, $t \geq 0$. Thus, $V, \tilde{\theta}$, and $(\tilde{\theta} - \tilde{\mathcal{V}}) \in \mathcal{L}_\infty$.

Next, integrating $\dot{V}$ over $[0,\infty)$ yields

$$\int_0^\infty \|R^{\frac{1}{2}}(\tau)\tilde{\theta}(\tau)\|^2\mathrm{d}\tau \leq V(\tilde{\mathcal{V}}_0, \tilde{\theta}_0, 0) - \lim_{t\to\infty} V(\tilde{\mathcal{V}}(t), \tilde{\theta}(t), t) < \infty,$$

and hence, $\|R^{\frac{1}{2}}\tilde{\theta}\| \in \mathcal{L}_2$. Similarly, $\|\tilde{\theta} - \tilde{\mathcal{V}}\|^2 \in \mathcal{L}_2$, and hence, $(\tilde{\theta} - \tilde{\mathcal{V}}) \in \mathcal{L}_\infty \cap \mathcal{L}_2$. Since (27) implies that $\|R\| \in \mathcal{L}_\infty$, it follows that $\|R^{\frac{1}{2}}\tilde{\theta}\| \in \mathcal{L}_\infty$. Now, it follows from (24) and (25) that $\dot{\tilde{\mathcal{V}}}, \dot{\tilde{\theta}} \in \mathcal{L}_\infty \cap \mathcal{L}_2$. Finally, since $\|\dot{R}\| \in \mathcal{L}_\infty$, it follows that $\ddot{\tilde{\mathcal{V}}}, \ddot{\tilde{\theta}} \in \mathcal{L}_\infty$, and, by Lemma 2, it follows that $\lim_{t\to\infty}\|\dot{\theta}(t)\| = 0$.

Next, to show (ii), it follows from (28) and (30) that

$$\dot{V}(\tilde{\mathcal{V}}, \tilde{\theta}, t) \leq -\min\{e^{-\nu\delta}\rho\gamma, 2\beta\}V(\tilde{\mathcal{V}}, \tilde{\theta}), \quad t \geq \delta,$$

which implies that

$$V(\tilde{\mathcal{V}}(t),\tilde{\theta}(t)) \leq V(\tilde{\mathcal{V}}_0,\tilde{\theta}_0)e^{-at}, \quad t \geq \delta. \tag{31}$$

Now, (31) implies that, for all $t \geq \delta$,

$$\|\tilde{\mathcal{V}}(t)\| \leq \sqrt{V(\tilde{\mathcal{V}}_0,\tilde{\theta}_0)}e^{-\frac{a}{2}t},$$
$$\|\tilde{\theta}(t)-\tilde{\mathcal{V}}(t)\| \leq \sqrt{V(\tilde{\mathcal{V}}_0,\tilde{\theta}_0)}e^{-\frac{a}{2}t},$$

and hence, using the triangle inequality $\|\tilde{\theta}\|-\|\tilde{\mathcal{V}}\| \leq \|\tilde{\theta}-\tilde{\mathcal{V}}\|$, it follows that

$$\|\tilde{\theta}(t)\| \leq 2\sqrt{V(\tilde{\mathcal{V}}_0,\tilde{\theta}_0)}e^{-\frac{a}{2}t}, \quad t \geq \delta, \tag{32}$$

which proves that $\tilde{\theta}(t)$ converges to 0 exponentially as $t \to \infty$ with a convergence rate of $\frac{a}{2}$. □

Remark 1. *Note that unlike gradient flow and recursive least squares algorithms, the momentum-based integral gradient algorithm* (20)–(23) *does not require the assumption that* $\dot{\phi}(\cdot) \in \mathcal{L}_\infty$ *in order to guarantee that* $\lim_{t\to\infty}\dot{\theta}(t) = 0$ *[7].*

Remark 2. *Note that a time-varying forgetting factor can also be employed in* (22) *and* (23)*. In this case, the PE condition can be replaced with a less restrictive excitation condition where* (26) *holds for a fixed-time t and not for all* $t \geq 0$ *[13].*

Next, we consider the cost function

$$J(\theta,t) = \int_0^t (y(\tau)-\phi(\tau)^{\mathrm{T}}\theta(\tau))^2 - (\theta(t)-\theta(\tau))^{\mathrm{T}}F(\tau)(\theta(t)-\theta(\tau))\mathrm{d}\tau$$
$$+ (\theta(t)-\theta(0))^{\mathrm{T}}\Gamma_0(\theta(t)-\theta(0)), \tag{33}$$

where for all $t \geq 0, F(t)$ is an $N \times N$ nonnegative-definite matrix function called the *general forgetting matrix* and Γ_0 is an $N \times N$ positive-definite matrix. It can be shown that a necessary condition for θ minimizing (33) gives the *recursive least squares* algorithm ([14])

$$\dot{\theta}(t) = -\Gamma(t)\phi(t)e(t), \quad \theta(0) = \theta_0, \quad t \geq 0, \tag{34}$$
$$\dot{\Gamma}(t) = -\Gamma(t)[\phi(t)\phi(t)^{\mathrm{T}} - F(t)]\Gamma(t), \quad \Gamma(0) = \Gamma_0. \tag{35}$$

Note that setting $F(t) \equiv 0$ and $F(t) = \nu H(t) = \nu\Gamma^{-1}(t)$, where $\Gamma(t)$ is positive-definite for all $t \geq 0$ (see [2]), we recover, respectively, the *pure recursive least squares* and the *recursive least squares with exponential forgetting* algorithms discussed in [2].

Note that the RLS algorithms (34) and (35) do not involve a gradient flow architecture, and the variational formulation cannot be used to generate a momentum-based recursive least squares (MRLS) algorithm. Despite this fact, higher-order tuner laws can still be incorporated into the RLS architecture to give the MRLS architecture

$$\dot{\mathcal{V}}(t) = -\Gamma(t)\phi(t)e(t), \quad \mathcal{V}(0) = \mathcal{V}_0, \quad t \geq 0, \tag{36}$$
$$\dot{\theta}(t) = -\beta(\theta(t)-\mathcal{V}(t))\mathcal{N}_{\mathrm{r}}(t), \quad \theta(0) = \theta_0, \tag{37}$$
$$\dot{\Gamma}(t) = -\Gamma(t)[\phi(t)\phi(t)^{\mathrm{T}} - F(t)]\Gamma(t), \quad \Gamma(0) = \Gamma_0, \tag{38}$$

where $\beta > 0$ and $\mathcal{N}_{\mathrm{r}}(t) \triangleq 1 + \mu\phi(t)^{\mathrm{T}}\Gamma(t)\phi(t) = 1 + \mu\|\Gamma^{\frac{1}{2}}(t)\phi(t)\|^2$, $\mu > 0$. Note that (36) and (37) can be rewritten as

$$\dot{\tilde{\mathcal{V}}}(t) = -\Gamma(t)\phi(t)e(t), \quad \tilde{\mathcal{V}}(0) = \tilde{\mathcal{V}}_0, \quad t \geq 0, \tag{39}$$

$$\dot{\tilde{\theta}}(t) = -\beta(\tilde{\theta}(t) - \tilde{\mathcal{V}}(t))\mathcal{N}_{\mathrm{r}}(t), \quad \tilde{\theta}(0) = \tilde{\theta}_0. \tag{40}$$

Noting that $H(t) = \Gamma^{-1}(t)$ and using the fact that

$$\frac{\mathrm{d}H(t)\Gamma(t)}{\mathrm{d}t} = \frac{\mathrm{d}I_n}{\mathrm{d}t} = 0 = \dot{H}(t)\Gamma(t) + H(t)\dot{\Gamma}(t), \tag{41}$$

it follows that

$$\dot{H}(t) = -F(t) + \phi(t)\phi(t)^{\mathrm{T}}, \quad H(0) = \Gamma_0^{-1}, \quad t \geq 0. \tag{42}$$

Lemma 3. *Consider the system* (42) *with* $F(t) = \nu H(t)$. *Then, the following statements hold:*
(i)

$$\|H(t)\| \leq \alpha_2, \quad t \geq 0, \tag{43}$$

where $\alpha_2 \triangleq \lambda_{\max}(\Gamma_0^{-1}) + \frac{1}{\nu}$.
(ii) *If* $\phi(\cdot)$ *is persistently excited, then*

$$\alpha_1 \leq \|H(t)\|, \quad t \geq 0, \tag{44}$$

where $\alpha_1 \triangleq \min\{e^{-\nu\delta}\rho, e^{-\nu\delta}\lambda_{\min}(\Gamma_0^{-1})\}$.

Proof. To show (i), note that it follows from (42) with $F(t) = \nu H(t)$ that

$$H(t) = e^{-\nu t}\Gamma_0^{-1} + \int_0^t e^{-\nu(t-\tau)}\phi(\tau)\phi(\tau)^{\mathrm{T}}\mathrm{d}\tau. \tag{45}$$

Now, using $\phi(t)\phi(t)^{\mathrm{T}} \leq I_N$, $t \geq 0$, yields

$$H(t) \leq \Gamma_0^{-1} + \int_0^t e^{-\nu(t-\tau)}I_N\mathrm{d}\tau \leq \lambda_{\max}(\Gamma_0^{-1})I_N + \frac{1}{\nu}I_N, \quad t \geq 0,$$

which proves (43).

Next, to show (ii), note that if $\phi(\cdot)$ is PE, then

$$H(t) \geq e^{-\nu t}\Gamma_0^{-1} \geq e^{-\nu\delta}\lambda_{\min}(\Gamma_0^{-1})I_N, \quad 0 \leq t < \delta, \tag{46}$$

and

$$H(t) \geq e^{-\nu\delta}\int_{t-\delta}^t \phi(\tau)\phi(\tau)^{\mathrm{T}}\mathrm{d}\tau \geq e^{-\nu\delta}\rho I_N, \quad t > \delta. \tag{47}$$

Thus,

$$\alpha_1 \leq \|H(t)\|, \quad t \geq 0. \tag{48}$$

This proves the lemma. □

Theorem 2. *Let* $\mu > \frac{2}{\beta}$, *and consider the momentum-based recursive least squares algorithm* (36)–(38). *Then,* $e(\cdot)$ *given by* (2) *satisfies* $e \in \mathcal{L}_2$. *Furthermore, the following statements hold:*
(i) *If there exists* $\alpha_1 > 0$ *such that* (44) *holds, then* $\mathcal{V}, \theta \in \mathcal{L}_\infty$. *If, in addition,* $\dot{\phi} \in \mathcal{L}_\infty$, *then* $\lim_{t\to\infty}\dot{\theta}(t) = 0$ *and* $\lim_{t\to\infty} e(t) = 0$.
(ii) *If* $F(t) = \nu H(t)$ *and* $\phi(\cdot)$ *is persistently excited, then the zero solution* $(\tilde{\mathcal{V}}(t), \tilde{\theta}(t)) \equiv (0, 0)$ *to* (39) *and* (40) *is exponentially stable.*

Proof. Consider the function $V(\tilde{\mathcal{V}},\tilde{\theta},t) \triangleq \tilde{\mathcal{V}}^{\mathrm{T}}H(t)\tilde{\mathcal{V}} + (\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta}-\tilde{\mathcal{V}})$ and note that

$$\begin{aligned}
&\dot{V}(\tilde{\mathcal{V}},\tilde{\theta},t)\\
&= 2\tilde{\mathcal{V}}^{\mathrm{T}}H(t)\dot{\tilde{\mathcal{V}}} + \tilde{\mathcal{V}}^{\mathrm{T}}\dot{H}(t)\tilde{\mathcal{V}} + 2(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\dot{\tilde{\theta}}-\dot{\tilde{\mathcal{V}}}) + (\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\dot{H}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\\
&= -2\tilde{\mathcal{V}}^{\mathrm{T}}H(t)\Gamma(t)\phi(t)e + \tilde{\mathcal{V}}^{\mathrm{T}}[-\nu H(t) + \phi(t)\phi(t)^{\mathrm{T}}]\tilde{\mathcal{V}} - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta}-\tilde{\mathcal{V}})\mathcal{N}_{\mathrm{r}}(t)\\
&\quad - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)\Gamma(t)\phi(t)e + (\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}[-\nu H(t) + \phi(t)\phi(t)^{\mathrm{T}}](\tilde{\theta}-\tilde{\mathcal{V}})\\
&= -\tilde{\mathcal{V}}^{\mathrm{T}}F(t)\tilde{\mathcal{V}} - \nu(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}F(t)(\tilde{\theta}-\tilde{\mathcal{V}}) - 2\tilde{\mathcal{V}}^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}\tilde{\theta} + \tilde{\mathcal{V}}^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}\tilde{\mathcal{V}}\\
&\quad - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta}-\tilde{\mathcal{V}})\mathcal{N}_{\mathrm{r}}(t) - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}\tilde{\theta}\\
&\quad + (\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}(\tilde{\theta}-\tilde{\mathcal{V}})\\
&= -\tilde{\mathcal{V}}^{\mathrm{T}}F(t)\tilde{\mathcal{V}} - \nu(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}F(t)(\tilde{\theta}-\tilde{\mathcal{V}}) + (\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}(\tilde{\theta}-\tilde{\mathcal{V}}) - 2e^2\\
&\quad - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta}-\tilde{\mathcal{V}})\mathcal{N}_{\mathrm{r}}(t) - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}\tilde{\theta}\\
&\quad + (\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}(\tilde{\theta}-\tilde{\mathcal{V}})\\
&= -\tilde{\mathcal{V}}^{\mathrm{T}}F(t)\tilde{\mathcal{V}} - \nu(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}F(t)(\tilde{\theta}-\tilde{\mathcal{V}}) + 2\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\|^2 - 2e^2\\
&\quad - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta}-\tilde{\mathcal{V}})\mathcal{N}_{\mathrm{r}}(t) - 2(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\phi(t)^{\mathrm{T}}\tilde{\theta}\\
&\leq -\tilde{\mathcal{V}}^{\mathrm{T}}F(t)\tilde{\mathcal{V}} - \nu(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}F(t)(\tilde{\theta}-\tilde{\mathcal{V}}) + 2\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\|^2 - 2e^2\\
&\quad - 2\beta(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta}-\tilde{\mathcal{V}})(1+\mu\|\Gamma^{\frac{1}{2}}(t)\phi(t)\|) + 2\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\||e|.
\end{aligned}$$

Now, by the Cauchy–Schwarz inequality,

$$\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\phi(t)\| = \|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H^{\frac{1}{2}}(t)\Gamma^{\frac{1}{2}}(t)\phi(t)\| \leq \|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|\|\Gamma^{\frac{1}{2}}(t)\phi(t)\|,$$

and hence,

$$\begin{aligned}
\dot{V}(\tilde{\mathcal{V}},\tilde{\theta},t) &\leq -\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 + 2(1-\beta\mu)\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2\|\Gamma^{\frac{1}{2}}(t)\phi(t)\|^2\\
&\quad - 2e^2 - 2\beta\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 + 2\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|\|\Gamma^{\frac{1}{2}}(t)\phi(t)\||e|\\
&\leq -\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 - 2\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2\|\Gamma^{\frac{1}{2}}(t)\phi(t)\|^2 - 2e^2\\
&\quad - 2\beta\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 - \|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2\|\Gamma^{\frac{1}{2}}\|^2\\
&\quad + 2\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|\|\Gamma^{\frac{1}{2}}(t)\phi(t)\||e|\\
&\leq -e^2 - \|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2\\
&\quad - 2\beta\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 - [\|H^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|\|\Gamma^{\frac{1}{2}}(t)\phi(t)\| - e]^2\\
&\leq -e^2 - \|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2\\
&\leq 0, \quad t \geq 0.
\end{aligned} \tag{49}$$

Next, integrating $\dot{V}$ over $[0,\infty)$ yields

$$\int_0^\infty e^2(\tau)\mathrm{d}\tau \leq V(\tilde{\mathcal{V}}_0,\tilde{\theta}_0,0) - \lim_{t\to\infty} V(\tilde{\mathcal{V}}(t),\tilde{\theta}(t),t) < \infty,$$

and hence, $e \in \mathcal{L}_2$.

To show *i*), note that (49) and (44) imply that $\tilde{\mathcal{V}} \in \mathcal{L}_\infty$ and $\tilde{\theta}-\tilde{\mathcal{V}} \in \mathcal{L}_\infty$. In addition, (37) implies that $\dot{\tilde{\theta}} \in \mathcal{L}_\infty$, and for $\dot{\phi} \in \mathcal{L}_\infty$, it follows from (2) and Lemma 2 that $\lim_{t\to\infty}\dot{\tilde{\theta}}(t) = 0$ and $\lim_{t\to\infty} e(t) = 0$.

Finally, to show *ii*), note that if $\phi(\cdot)$ is persistently excited, then (44) holds by Lemma 3. Now, using the fact that

$$\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\mathcal{V}}-\tilde{\theta}\|^2 = [\tilde{\mathcal{V}}^{\mathrm{T}},\tilde{\theta}^{\mathrm{T}}]^{\mathrm{T}}P[\tilde{\mathcal{V}},\tilde{\theta}], \tag{50}$$

where $P \triangleq \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ is a positive-definite matrix with eigenvalues $\{\frac{3-\sqrt{5}}{2}, \frac{3+\sqrt{5}}{2}\}$, it follows from the Schur decomposition that

$$\frac{3-\sqrt{5}}{2}(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2) \leq \|\tilde{\mathcal{V}}\|^2 + \|\tilde{\mathcal{V}} - \tilde{\theta}\|^2 \leq \frac{3+\sqrt{5}}{2}(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2), \tag{51}$$

and hence,

$$\left(\frac{3-\sqrt{5}}{2}\right)\alpha_1(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2) \leq V(\tilde{\mathcal{V}}, \tilde{\theta}, t) \leq \left(\frac{3+\sqrt{5}}{2}\right)\alpha_2(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2). \tag{52}$$

In this case, (49) gives

$$\begin{aligned} \dot{V}(\tilde{\mathcal{V}}, \tilde{\theta}, t) &\leq -e^2 - \|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta} - \tilde{\mathcal{V}})\|^2 \\ &\leq -\alpha_1\nu(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\mathcal{V}} - \tilde{\theta}\|^2) \\ &\leq -\alpha_1\nu\left(\frac{3-\sqrt{5}}{2}\right)(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2), \end{aligned} \tag{53}$$

which shows that the zero solution $(\tilde{\mathcal{V}}(t), \tilde{\theta}(t)) \equiv (0, 0)$ to (39) and (40) is exponentially stable [15] ([Theorem 4.6]). □

Remark 3. *Note that setting*

$$F(t) = \begin{cases} \nu H(t), & \text{if } \|H(t)\| \geq \alpha_1, \quad t \geq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{54}$$

guarantees that (44) *holds [14].*

The constraints on parameter μ in Theorems 1 and 2 limit the amount of momentum that can be applied, thereby limiting the potential advantage gained by using a momentum-based algorithm. In the context of static optimization, Lyapunov-like functions are typically constructed as the sum of two terms; one term representing the norm of the error of the parameter estimate and the other term corresponding to the loss cost function [12]. This generalized structure allows for adjustments to the parameter estimate $\tilde{\theta}$, that may not decrease $\|\tilde{\mathcal{V}}\|$ or $\|\tilde{\theta}\|$ but can reduce the loss cost function $L(\tilde{\theta}, t)$ as shown in Figure 1.

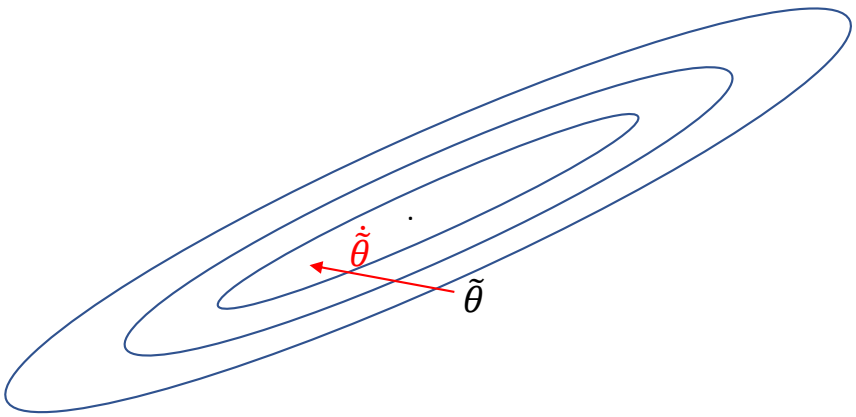

Figure 1. Visualization of momentum-based learning for an ill-conditioned problem; $\tilde{\theta}$ is updated in a direction that increases $\|\tilde{\theta}\|$ while still decreasing the loss cost function.

In order to incorporate (10) in our Lyapunov function, we introduce the *momentum-based composite gradient algorithm*

$$\dot{\mathcal{V}}(t) = -\Gamma(t)\phi(t)e(t) - \frac{\beta}{2}\Gamma(t)\mathcal{N}_{\mathrm{r}}(t)[R(t)\theta(t) + Q(t)], \quad \mathcal{V}(0) = \mathcal{V}_0, \quad t \geq 0, \tag{55}$$

$$\dot{\theta}(t) = -\beta[\theta(t) - \mathcal{V}(t)]\mathcal{N}_{\mathrm{r}}(t), \quad \theta(0) = \theta_0, \tag{56}$$

$$\dot{\Gamma}(t) = -\Gamma(t)[\phi(t)\phi(t)^{\mathrm{T}} - F(t)]\Gamma(t), \quad \Gamma(0) = \Gamma_0, \tag{57}$$

where $R(\cdot)$ and $Q(\cdot)$ are given by (12) and (13). Note that this composite update law includes an integral gradient term as well as a recursive least squares term in (55).

Theorem 3. *Let $\mu > \frac{2}{\beta}$, and consider the momentum-based composite gradient algorithm (55)–(57). Then, the following statements hold:*

(i) If there exists $\alpha_1 > 0$ such that (44) holds, then $\theta \in \mathcal{L}_\infty$, $\dot{\theta}, \dot{\mathcal{V}}$, and $(\theta - \mathcal{V}) \in \mathcal{L}_\infty$, and $\lim_{t\to\infty} \dot{\theta}(t) = 0$.

(ii) If $F(t) = \nu H(t)$ and $\phi(\cdot)$ is persistently excited, then the zero solution $(\tilde{\mathcal{V}}(t), \tilde{\theta}(t)) \equiv (0, 0)$ to (55)–(57) is exponentially stable.

Proof. This proof is similar to the proof of Theorem 2. Namely, consider the positive-definite function

$$V(\tilde{\mathcal{V}}, \tilde{\theta}, t) \triangleq \tilde{\mathcal{V}}^{\mathrm{T}} H(t)\tilde{\mathcal{V}} + (\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}} H(t)(\tilde{\theta} - \tilde{\mathcal{V}}) + 2L(\tilde{\theta}, t) \tag{58}$$

satisfying $V(0, 0, t) = 0$, $t \geq 0$, and note that

$$\dot{L}(\tilde{\theta}, t) = \frac{\partial L}{\partial \tilde{\theta}} \frac{\mathrm{d}\tilde{\theta}}{\mathrm{dt}} + \frac{\partial L}{\partial t} = \tilde{\theta}^{\mathrm{T}} R(t)\dot{\tilde{\theta}} - \nu L + \frac{1}{2}e(t)^2. \tag{59}$$

Now, using (49) and (59), note that

$$\begin{aligned}
\dot{V}(\tilde{\mathcal{V}}, \tilde{\theta}, t) &\leq -e(t)^2 - \|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta} - \tilde{\mathcal{V}})\|^2 - \beta\mathcal{N}_{\mathrm{r}}(t)\tilde{\mathcal{V}}^{\mathrm{T}} R(t)\tilde{\theta} \\
&\quad + \beta\mathcal{N}_{\mathrm{r}}(t)(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}} R(t)\tilde{\theta} - 2\tilde{\theta}^{\mathrm{T}} R(t)\beta(\tilde{\theta} - \tilde{\mathcal{V}})\mathcal{N}_{\mathrm{r}}(t) - 2\nu L + e(t)^2 \\
&\leq -\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta} - \tilde{\mathcal{V}})\|^2 - 2\beta\mathcal{N}_{\mathrm{r}}(t)\tilde{\theta}^{\mathrm{T}} R(t)\tilde{\theta} + 2\beta\mathcal{N}_{\mathrm{r}}(t)(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}} R(t)\tilde{\theta} \\
&\quad - 2\tilde{\theta}^{\mathrm{T}} R(t)\beta(\tilde{\theta} - \tilde{\mathcal{V}})\mathcal{N}_{\mathrm{r}}(t) - 2\nu L \\
&\leq -\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - \|F^{\frac{1}{2}}(t)(\tilde{\theta} - \tilde{\mathcal{V}})\|^2 - 2\tilde{\theta}^{\mathrm{T}} R(t)\tilde{\theta} - 2\nu L.
\end{aligned} \tag{60}$$

(i) now follows using similar arguments as in the proof of Theorem 2.

To show (ii), note that if $\phi(\cdot)$ is persistently excited, then (44) and (51) hold, and hence, using the fact that $L(\tilde{\theta}, t) \leq \frac{1}{\nu}\|\tilde{\theta}\|^2$, we have

$$\begin{aligned}
\left(\frac{3 - \sqrt{5}}{2}\right) e^{-\nu\delta}\rho(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2) \leq V(\tilde{\mathcal{V}}, \tilde{\theta}, t) &\leq \left(\frac{3 + \sqrt{5}}{2}\right)\frac{1}{\nu}(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2) + \frac{1}{\nu}\|\tilde{\theta}\|^2 \\
&\leq \left(\frac{5 + \sqrt{5}}{2\nu}\right)(\|\tilde{\mathcal{V}}\|^2 + \|\tilde{\theta}\|^2).
\end{aligned} \tag{61}$$

Now, using (49) and (60), it follows that the zero solution $(\tilde{\mathcal{V}}(t), \tilde{\theta}(t)) \equiv (0, 0)$ to (55) and (56) is exponentially stable [15] ([Theorem 4.6]). □

3. System Parameter Identification

In this section, we address the problem of parameter identification using the first- and higher-order tuner algorithms developed in Section 2 . First, the following definition is needed.

Definition 2 ([2]). *A signal $u : [0, \infty) \to \mathbb{R}$ is stationary and sufficiently rich of order n if the following statements hold:*

(i) The limit

$$\lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} u(\tau)u^{\mathrm{T}}(t+\tau)d\tau \tag{62}$$

exists uniformly in t_0.

(ii) The support of the spectral measure $S_u(\omega)$, $\omega \in \mathbb{R}$, of u contains at least n points.

Consider the stable single-input, single-output (SISO) plant, with output $y(\cdot)$ and input $u(\cdot)$ being the only available signals for measurement, given by

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0, \quad t \geq 0, \tag{63}$$

$$y(t) = C^{\mathrm{T}}x(t), \tag{64}$$

where $x(t) \in \mathbb{R}^n$, $t \geq 0$, $A \in \mathbb{R}^{n\times n}$, and $B, C \in \mathbb{R}^n$. Since there are an infinite number of triples (A, B, C) yielding the same input–output measurements, we cannot uniquely define the $n^2 + 2n$ coefficients in (A, B, C). However, we can determine the $n + m + 1$ parameters corresponding to the coefficients of the stable and strictly proper rational transfer function $G \in \mathbb{R}(s)$ given by

$$G(s) = C^{\mathrm{T}}(sI_n - A)^{-1}B = \frac{\hat{y}(s)}{\hat{u}(s)} = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_0}, \tag{65}$$

where, for $s \in \mathbb{C}$, $\hat{y}(s)$ is the system output, $\hat{u}(s)$ is the system input, and the coefficients a_i, $i = 0, \ldots, n-1$, and b_j, $j = 0, \ldots, m$, where $m \leq n-1$, are unknown system parameters. The goal of the system parameter identification problem is to identify the system parameter vector

$$\theta^* = [b_m, b_{m-1}, \ldots, b_0, a_{n-1}, \ldots, a_0] \tag{66}$$

containing the unknown coefficients of the plant transfer function that satisfies

$$s^n\hat{y}(s) = \theta^{*\mathrm{T}}\hat{Y}(s), \tag{67}$$

where $\hat{Y}(s) \triangleq [\alpha_m^{\mathrm{T}}(s)\hat{u}(s), -\alpha_{n-1}^{\mathrm{T}}(s)\hat{y}(s)]^{\mathrm{T}}$ and $\alpha_i(s) \triangleq [s^i, s^{i-1}, \ldots, 1]^{\mathrm{T}}$.

Note that (67) has the form of (1) in the frequency domain; however, in most applications the only signals that are available for measurement are the input $u(t)$, $t \geq 0$, and the output $y(t)$, $t \geq 0$, and not their derivatives. To address this, we filter (67) through a stable filter $\frac{1}{\Lambda(s)}$, where $\Lambda(s)$ is an arbitrary monic Hurwitz polynomial of degree n, to obtain, for $y(0) = 0$,

$$\hat{z}(s) = \theta^{*\mathrm{T}}\hat{\phi}(s), \tag{68}$$

where $\hat{z}(s) \triangleq \frac{s^n\hat{y}(s)}{\Lambda(s)}$ and $\hat{\phi}(s) \triangleq [\frac{\alpha_m^{\mathrm{T}}(s)}{\Lambda(s)}\hat{u}(s), -\frac{\alpha_{n-1}^{\mathrm{T}}(s)}{\Lambda(s)}\hat{y}(s)]^{\mathrm{T}}$. The time signals $\phi(t)$, $t \geq 0$, and $z(t)$, $t \geq 0$, can be generated by the state equations ([2])

$$\dot{\phi}_0(t) = \Lambda_c\phi_0(t) + b_c u(t), \quad \phi_0(0) = 0, \quad t \geq 0, \tag{69}$$

$$\phi_1(t) = P_0\phi_0(t), \tag{70}$$

$$\dot{\phi}_2(t) = \Lambda_c\phi_2(t) - b_c y(t), \quad \phi_2(0) = 0, \tag{71}$$

$$\phi(t) = [\phi_1^{\mathrm{T}}(t), \phi_2^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{72}$$

$$z(t) = y(t) + \lambda^{\mathrm{T}}\phi_2(t), \tag{73}$$

where, for $t \geq 0$, $\phi_0(t) \in \mathbb{R}^n$, $\phi_1(t) \in \mathbb{R}^{m+1}$, $\phi_2(t) \in \mathbb{R}^n$, and

$$\Lambda_c = \left[\frac{-\lambda^T}{I_{n-1} \mid 0_{(n-1)\times 1}}\right], \quad b_c = [1, 0, \ldots, 0]^T, \tag{74}$$

$$P_0 = [0_{(m+1)\times(n-m-1)} \mid I_{m+1}] \in \mathbb{R}^{(m+1)\times n}, \quad \lambda = [\lambda_{n-1}, \lambda_{n-2}, \ldots, \lambda_0]^T, \tag{75}$$

and where $\det(sI_n - \Lambda_c) = \Lambda(s) = s^n + \lambda^T \alpha_{n-1}(s)$.

Theorem 4. *Assume that $G \in \mathbb{R}(s)$ given by (65) has no pole-zero cancellations and $y(0) = 0$. If u is stationary and sufficiently rich of order $n + m + 1$, then the adaptive laws (20)–(23), (36)–(38), and (55)–(57) guarantee that $\theta(t)$ converges to θ^* as $t \to \infty$.*

Proof. The proof follows from Theorems 1, 2, and 3 by noting that if $u(\cdot)$ is stationary and sufficiently rich of order $n + m + 1$ and G has no pole-zero cancellations, then $\phi(\cdot)$ is PE [2] ([Thm. 5.2.4]). □

Remark 4. *Note that if $y(0) \neq 0$, then (68) would involve an additional bias error term. As shown in [2], this term converges to the origin exponentially fast, and hence, Theorem 4 remains valid.*

4. Momentum-Based Recursive Least Squares and Model Reference Adaptive Control

Recursive least squares type algorithms were first introduced in adaptive control in [16] and extended in [2,17,18]. In this section, we show how the MRLS algorithm proposed in Section 2 can be extended to the problem of model reference adaptive control for systems with relative degree one.

Consider the linear SISO system given by

$$\hat{y}_p(s) = G(s)\hat{u}_p(s), \tag{76}$$

where $G(s) = k_p \frac{N_p(s)}{D_p(s)} \in \mathbb{R}(s)$, $N_p(s), D_p(s) \in \mathbb{R}[s]$ are unknown polynomials, k_p is an unknown constant, $\hat{u}_p(s)$ is the control system input, and $\hat{y}_p(s)$ is the system output. As in [18], we make the following assumptions.

(i) $G \in \mathbb{R}(s)$ is minimum phase and has relative degree one.

(ii) The degree of $D_p(s)$ is n.

(iii) The sign of k_p is known.

The control objective is to design an appropriate control law $u_p(t), t \geq 0$, such that all the signals of the closed-loop system are bounded and $y_p(t)$, $t \geq 0$, tracks the output $y_r(t)$, $t \geq 0$, of a reference model given by

$$\hat{y}_r(s) = M(s)\hat{r}(s), \tag{77}$$

where $M(s) \triangleq \frac{k_r}{s+a_r} \in \mathbb{R}(s)$, $a_r > 0$ and k_r are known constants, and $\hat{r}(s)$ is the Laplace transform of a bounded piecewise continuous signal $r(t)$, $t \geq 0$.

To address this problem, consider the filter system ([2])

$$\dot{v}_1(t) = F_c v_1(t) + g_c u_p(t), \quad v_1(0) = v_{10}, \quad t \geq 0, \tag{78}$$

$$\dot{v}_2(t) = F_c v_2(t) + g_c y_p(t), \quad v_2(0) = v_{20}, \tag{79}$$

where, for $t \geq 0$, $v_1(t) \in \mathbb{R}^{n-1}$, $v_2(t) \in \mathbb{R}^{n-1}$,

$$F_c = \left[\frac{-\hat{\lambda}^T}{I_{n-2} \mid 0_{(n-2)\times 1}}\right] \in \mathbb{R}^{(n-1)\times(n-1)}, \quad \hat{\lambda} = [\lambda_{n-2}, \lambda_{n-3}, \ldots, \lambda_0]^T, \tag{80}$$

$g_c = [1, 0 \ldots, 0]^{\mathrm{T}} \in \mathbb{R}^{n-1}$, and $\det(sI_{n-1} - F_c) = s^{n-1} + \hat{\lambda}^{\mathrm{T}}\alpha_{n-2}(s) = F(s)$, where $F(s) \in \mathbb{R}[s]$ is an arbitrary monic Hurwitz polynomial of degree $n-1$. Here, we use $v_1(t), t \geq 0$, and $v_2(t), t \geq 0$, to form the regressor vector

$$\phi(t) = [v_1^{\mathrm{T}}(t), v_2^{\mathrm{T}}(t), y_{\mathrm{p}}(t), r(t)]^{\mathrm{T}} \in \mathbb{R}^{2n}. \tag{81}$$

Note that the existence of a constant parameter vector $\theta^* = [\theta_1^{*\mathrm{T}}, \theta_2^{*\mathrm{T}}, \theta_3^*, \theta_4^*]^{\mathrm{T}} \in \mathbb{R}^{2n}$ such that the transfer function of the SISO system (76) with $\hat{u}_{\mathrm{p}}(s) = \theta^{*\mathrm{T}}\hat{\phi}(s)$, where $\hat{\phi}(s) = [\frac{\alpha_{n-2}(s)}{F(s)}\hat{u}_{\mathrm{p}}(s), \frac{\alpha_{n-2}(s)}{F(s)}\hat{y}_{\mathrm{p}}(s), \hat{y}_{\mathrm{p}}(s), \hat{r}(s)]$, matching the reference model transfer function $M(s) \in \mathbb{R}(s)$ is guaranteed by the choice of $M(s)$ and Assumptions (i)–(iii).

Next, consider the control law

$$\hat{u}_{\mathrm{p}}(s) = L(s)\hat{\theta}^{\mathrm{T}}(s)\hat{\xi}(s), \tag{82}$$

where $\hat{\theta}(s) \in \mathbb{C}^{2n}$, $L(s) \triangleq s + \lambda_1,\ \lambda_1 > 0$, and $\hat{\xi}(s) \triangleq L^{-1}(s)\hat{\phi}(s) \in \mathbb{C}^{2n}$, and note that the tracking error $\hat{e}_1(s) \triangleq \hat{y}_{\mathrm{p}}(s) - \hat{y}_{\mathrm{r}}(s)$ satisfies

$$\hat{e}_1(s) = M(s)L(s)k^*[\hat{u}_{\mathrm{p}}(s) - \theta^{*\mathrm{T}}\hat{\phi}(s)] = M(s)L(s)k^*\tilde{\theta}^{\mathrm{T}}(s)\hat{\xi}(s), \tag{83}$$

where $k^* = \frac{k_{\mathrm{p}}}{k_{\mathrm{r}}}$. Note that (83) has a nonminimal state-space realization given by

$$\dot{e}_2(t) = A_m e_2(t) + B_m k^*\tilde{\theta}^{\mathrm{T}}(t)\xi(t), \quad e_2(0) = e_{20}, \quad t \geq 0, \tag{84}$$

$$e_1(t) = C_m e(t) + k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}(t)\xi(t), \tag{85}$$

where $e_2(t) \in \mathbb{R}^{3n-2}$, $A_m \in \mathbb{R}^{(3n-2)\times(3n-2)}$, $B_m = \mathbb{R}^{(3n-2)\times 1}$, and $C_m = \mathbb{R}^{1\times(3n-2)}$. Furthermore, note that

$$M(s)L(s) = \frac{k_{\mathrm{r}}(s+\lambda_1)}{s+a_{\mathrm{r}}} = \frac{k_{\mathrm{r}}(\lambda_1 - a_{\mathrm{r}})}{s+a_{\mathrm{r}}} + k_{\mathrm{r}}, \tag{86}$$

where $\lambda_1 > a_{\mathrm{r}}$, $\frac{k_{\mathrm{r}}(\lambda_1 - a_{\mathrm{r}})}{s+a_{\mathrm{r}}}$ is a strictly positive real transfer function. In this case, it follows from the Meyer–Kalman–Popov lemma [2] ([Lem. 3.5.4]) that there exist $(3n-2)\times(3n-2)$ positive-definite matrices $P > 0$ and $Q > 0$ such that

$$A_m^{\mathrm{T}}P + PA_m = -2Q, \tag{87}$$

$$PB_m = C_m^{\mathrm{T}}. \tag{88}$$

Since for every nonsingular matrix $S \in \mathbb{R}^{(3n-2)\times(3n-2)}$ and constant $\alpha > 0$, the realizations (A, B, C) and $(SAS^{-1}, \alpha SB, \alpha^{-1}CS^{-1})$ are equivalent, choosing $S = Q^{\frac{1}{2}}$ and $\alpha = \|C_m\|$ we can ensure that (A_m, B_m, C_m) is a realization that satisfies (87) and (88) with $Q = I_{3n-2}$ and C_m is normalized so that $\|C_m\| = 1$.

Next, consider the recursive least-squares algorithm given by

$$\dot{\theta}(t) = -\alpha\mathrm{sgn}(k^*)\Gamma(t)\xi(t)e_1(t), \quad \theta(0) = \theta_0, \quad t \geq 0, \tag{89}$$

$$\dot{\Gamma}(t) = \nu\Gamma(t) - \Gamma(t)\xi(t)\xi(t)^{\mathrm{T}}\Gamma(t), \quad \Gamma(0) = \Gamma_0, \tag{90}$$

where $\mathrm{sgn}(\sigma) \triangleq \sigma/|\sigma|$, $\sigma \neq 0$, and $\mathrm{sgn}(0) \triangleq 0$, $\alpha > 0$, and $\nu \geq 0$. Note that for $\nu = 0$, (89) and (90) recover the recursive least-squares algorithm of [19].

Here, we modify the recursive least-squares algorithm (89) and (90) to construct our MRLS algorithm as

$$\dot{\mathcal{V}}(t) = -\gamma\mathrm{sgn}(k^*)\Gamma(t)\xi(t)e_1(t), \quad \mathcal{V}(0) = \mathcal{V}_0, \quad t \geq 0, \tag{91}$$

$$\dot{\theta}(t) = -\beta(\theta(t) - \mathcal{V}(t))\mathcal{N}_{\mathrm{m}}(t), \quad \theta(0) = \theta_0, \tag{92}$$

$$\dot{\Gamma}(t) = \nu\Gamma(t) - \Gamma(t)\xi(t)\xi(t)^{\mathrm{T}}\Gamma(t), \quad \Gamma(0) = \Gamma_0, \tag{93}$$

where $\beta > 0$, $\gamma > 0$, $\alpha > 0$, $\nu > 0$,

$$\mathcal{N}_{\mathrm{m}}(t) \triangleq 1 + \mu\xi(t)^{\mathrm{T}}\Gamma(t)\xi(t) = 1 + \mu\|\Gamma^{\frac{1}{2}}(t)\xi(t)\|^2, \tag{94}$$

and $\mu > 0$. Note that (91) and (92) can be rewritten in terms of the error parameters $\tilde{\mathcal{V}}(t)$, $t \geq 0$, and $\tilde{\theta}(t)$, $t \geq 0$, as

$$\dot{\tilde{\mathcal{V}}}(t) = -\gamma\,\mathrm{sgn}(k^*)\Gamma(t)\xi(t)e_1(t), \quad \tilde{\mathcal{V}}(0) = \mathcal{V}_0 - \mathcal{V}^*, \quad t \geq 0, \tag{95}$$

$$\dot{\tilde{\theta}}(t) = -\beta(\tilde{\theta}(t) - \tilde{\mathcal{V}}(t))\mathcal{N}_{\mathrm{m}}(t), \quad \tilde{\theta}(0) = \theta_0 - \theta^*. \tag{96}$$

Theorem 5. *Consider the SISO system* (76) *with control law* (82) *and the MRLS algorithm given by* (91)–(93) *with* $\alpha > \frac{1}{|k_{\mathrm{r}}||k^*|}$ *and* $\mu > \frac{1}{\beta}(1 + \frac{2\alpha|k^*|}{\zeta})$, *where* $\zeta \leq \min\{1, \frac{1}{2k_{\mathrm{r}}}\}$. *Then* $e_2 \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ *and* $\tilde{\theta}^{\mathrm{T}}\xi \in \mathcal{L}_2$. *If, in addition,* $\nu = 0$, *then* $\tilde{\theta}, \tilde{\mathcal{V}} \in \mathcal{L}_\infty$ *and* $\lim_{t\to\infty} e_1(t) = 0$.

Proof. Consider the function

$$V(\tilde{\theta}, \tilde{\mathcal{V}}, e_2, t) \triangleq \alpha^{-1}|k^*|\tilde{\mathcal{V}}^{\mathrm{T}}H(t)\tilde{\mathcal{V}} + \alpha^{-1}|k^*|(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta} - \tilde{\mathcal{V}}) + e_2^{\mathrm{T}}Pe_2,$$

where $P > 0$ satisfies (87) and (88), and note that

$$\begin{aligned}
&\dot{V}(\tilde{\theta}, \tilde{\mathcal{V}}, e_2, t)\\
&\quad = e_2^{\mathrm{T}}P\dot{e}_2 + 2|k^*|\alpha^{-1}\tilde{\mathcal{V}}^{\mathrm{T}}H(t)\dot{\tilde{\mathcal{V}}} + |k^*|\alpha^{-1}\tilde{\mathcal{V}}^{\mathrm{T}}\dot{H}(t)\tilde{\mathcal{V}} + |k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\dot{H}(t)(\tilde{\theta} - \tilde{\mathcal{V}})\\
&\qquad + 2|k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\dot{\tilde{\theta}} - \dot{\tilde{\mathcal{V}}})\\
&\quad = -2\|e_2\|^2 + 2e_2^{\mathrm{T}}PB_m k^*\tilde{\theta}^{\mathrm{T}}\xi(t) - 2k^*\tilde{\mathcal{V}}^{\mathrm{T}}\xi(t)e_1 + |k^*|\alpha^{-1}\tilde{\mathcal{V}}^{\mathrm{T}}\dot{H}(t)\tilde{\mathcal{V}}\\
&\qquad + |k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\dot{H}(t)(\tilde{\theta} - \tilde{\mathcal{V}}) - 2\beta|k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta} - \tilde{\mathcal{V}})\mathcal{N}_{\mathrm{m}}(t)\\
&\qquad + 2k^*(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)e_1\\
&\quad = -2\|e_2\|^2 + 2(C_m e_2)^{\mathrm{T}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t) - 2k^*\tilde{\mathcal{V}}^{\mathrm{T}}\xi(t)e_1 + |k^*|\alpha^{-1}\tilde{\mathcal{V}}^{\mathrm{T}}\dot{H}(t)\tilde{\mathcal{V}}\\
&\qquad + |k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\dot{H}(t)(\tilde{\theta} - \tilde{\mathcal{V}}) - 2\beta|k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta} - \tilde{\mathcal{V}})\mathcal{N}_{\mathrm{m}}(t)\\
&\qquad + 2k^*(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)e_1\\
&\quad = -2\|e_2\|^2 + 2(e_1 - k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^{\mathrm{T}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t) - 2k^*\tilde{\mathcal{V}}^{\mathrm{T}}\xi(t)e_1\\
&\qquad + |k^*|\alpha^{-1}\tilde{\mathcal{V}}^{\mathrm{T}}[-F(t) + \xi(t)\xi(t)^{\mathrm{T}}]\tilde{\mathcal{V}} + |k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}[-F(t) +\\
&\qquad \xi(t)\xi(t)^{\mathrm{T}}](\tilde{\theta} - \tilde{\mathcal{V}}) - 2\beta|k^*|\alpha^{-1}(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta} - \tilde{\mathcal{V}})\mathcal{N}_{\mathrm{m}}(t) + 2k^*(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)e_1\\
&\quad = -2\|e_2\|^2 - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)(\tilde{\theta} - \tilde{\mathcal{V}})\|^2 - 2k_{\mathrm{r}}(k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^2\\
&\qquad + 4e_1k^*(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\xi(t) + |k^*|\alpha^{-1}|\tilde{\mathcal{V}}^{\mathrm{T}}\xi(t)|^2 + |k^*|\alpha^{-1}|(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)|^2\\
&\qquad - 2\beta|k^*|\alpha^{-1}\|(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}H^{\frac{1}{2}}(t)\|^2\mathcal{N}_{\mathrm{m}}(t), \quad t \geq 0.
\end{aligned} \tag{97}$$

Next, using the Cauchy–Schwarz and Young inequalities we obtain

$$\begin{aligned}
e_1^2 &= e_2^{\mathrm{T}}C_m^{\mathrm{T}}C_m e_2 + (k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^2 + 2k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}e_2^{\mathrm{T}}C_m\\
&\leq \|C_m\|^2\|e_2\|^2 + (k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^2 + 2k_{\mathrm{r}}\|k^*\tilde{\theta}^{\mathrm{T}}\xi(t)\|\|e_2^{\mathrm{T}}C_m\|\\
&\leq 2\|C_m\|^2\|e_2\|^2 + 2(k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^2\\
&\leq 2\|e_2\|^2 + 2(k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^2.
\end{aligned} \tag{98}$$

Now, using (98) along with triangle inequality $|\tilde{\mathcal{V}}^{\mathrm{T}}\xi(t)|^2 \leq |\tilde{\theta}^{\mathrm{T}}\xi(t)|^2 + |(\tilde{\theta} - \tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)|^2$, $t \geq 0$, it follows from (97) that

$$\dot{V}(\tilde{\theta}, \tilde{\mathcal{V}}, e_2, t)$$

$$
\begin{aligned}
&\leq -2(1-\zeta)\|e_2\|^2 - \zeta e_1^2 + 2\zeta(k_{\mathrm{r}}k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^2 - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - 2k_{\mathrm{r}}(k^*\tilde{\theta}^{\mathrm{T}}\xi(t))^2\\
&\quad - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 + 4|e_1||k^*|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\xi(t) + |k^*|\alpha^{-1}[|\tilde{\theta}^{\mathrm{T}}\xi(t)|^2\\
&\quad + \|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)\|^2] + |k^*|\alpha^{-1}|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)|^2\\
&\quad - 2\beta|k^*|\alpha^{-1}\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H^{\frac{1}{2}}(t)\|^2\mathcal{N}_{\mathrm{m}}(t)\\
&\leq -2(1-\zeta)\|e_2\|^2 + |k^*|(-2k_{\mathrm{r}}|k^*| + 2\zeta k_{\mathrm{r}}^2|k^*| + \alpha^{-1})|\tilde{\theta}^{\mathrm{T}}\xi(t)|^2\\
&\quad - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2\\
&\quad - [\sqrt{\zeta}|e_1| - 2\frac{|k^*|}{\sqrt{\zeta}}|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)|]^2 + \left(2|k^*|\alpha^{-1} + 4\frac{|k^*|^2}{\zeta}\right)|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}\xi(t)|^2\\
&\quad - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 - 2\beta|k^*|\alpha^{-1}\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H^{\frac{1}{2}}(t)\|^2\mathcal{N}_{\mathrm{m}}(t)\\
&\leq -2(1-\zeta)\|e_2\|^2 + |k^*|(-2k_{\mathrm{r}}|k^*| + 2\zeta k_{\mathrm{r}}^2|k^*| + \alpha^{-1})|\tilde{\theta}^{\mathrm{T}}\xi(t)|^2\\
&\quad - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2\\
&\quad + \left(2|k^*|\alpha^{-1} + \frac{4|k^*|^2}{\zeta} - 2\beta\mu|k^*|\alpha^{-1}\right)|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H^{\frac{1}{2}}(t)|^2\|\Gamma^{\frac{1}{2}}(t)\xi(t)\|\\
&\quad - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2 - 2\beta|k^*|\alpha^{-1}\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H^{\frac{1}{2}}(t)\|^2\\
&\leq -2(1-\zeta)\|e_2\|^2 + |k^*|(-2k_{\mathrm{r}}|k^*| + 2\zeta k_{\mathrm{r}}^2|k^*| + \alpha^{-1})|\tilde{\theta}^{\mathrm{T}}\xi(t)|^2\\
&\quad - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)\mathcal{V}\|^2 - |k^*|\alpha^{-1}\|F^{\frac{1}{2}}(t)(\tilde{\theta}-\tilde{\mathcal{V}})\|^2\\
&\quad - 2\beta|k^*|\alpha^{-1}\|(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H^{\frac{1}{2}}(t)\|^2\\
&\leq 0, \quad t \geq 0,
\end{aligned}
$$

which shows that $V(\tilde{\theta}, \tilde{\mathcal{V}}, e_2, t) \in \mathcal{L}_\infty$. Hence, $\tilde{\mathcal{V}}^{\mathrm{T}}H(t)\tilde{\mathcal{V}}$, $(\tilde{\theta}-\tilde{\mathcal{V}})^{\mathrm{T}}H(t)(\tilde{\theta}-\tilde{\mathcal{V}}) \in \mathcal{L}_\infty$, $t \geq 0$, $e_2^{\mathrm{T}}Pe_2 \in \mathcal{L}_\infty$. Next, integrating $\dot{V}$ over $[0,\infty)$ yields

$$
\int_0^\infty \|e_2(\tau)\|^2 \mathrm{d}\tau \leq V(\tilde{\mathcal{V}}_0, \tilde{\theta}_0, e_{20}, 0) - \lim_{t\to\infty} V(\tilde{\mathcal{V}}(t), \tilde{\theta}(t), e_2(t), t) < \infty,
$$

and hence, $e_2 \in \mathcal{L}_2$. Similarly, $\tilde{\theta}^{\mathrm{T}}\xi \in \mathcal{L}_2$ and, since $P > 0$, $e_2 \in \mathcal{L}_\infty$.

Finally, if $\nu = 0$, then $H(t)$ satisfies

$$
\dot{H}(t) = \xi(t)\xi(t)^{\mathrm{T}}, \quad H(0) = \Gamma_0^{-1}, \quad t \geq 0, \tag{99}
$$

or

$$
H(t) = \Gamma_0^{-1} + \int_0^t \xi(\tau)\xi(\tau)^{\mathrm{T}}\mathrm{d}\tau. \tag{100}
$$

Hence, $H(t) > \Gamma_0^{-1}$, $t \geq 0$, and thus, $\tilde{\mathcal{V}}, \tilde{\theta} \in \mathcal{L}_\infty$. Now, if $e_1, \tilde{\theta} \in \mathcal{L}_\infty$, then (85) implies that $\dot{e}_1 \in \mathcal{L}_\infty$ and, by Lemma 2, $\lim_{t\to\infty} e_1(t) = 0$. □

Remark 5. *Note that unlike many MRAC schemes (e.g., [20]), (91)–(93) does not necessitate a projection operator to guarantee boundedness of the tracking error. However, unlike [19], wherein only a lower bound is necessary for $|k^*|$, we require knowledge of a lower and upper bound for $|k^*|$.*

Remark 6. *It is important to note that there exists an alternative MRAC framework in the literature known as the normalized adaptive laws [2] whose design is not based on the tracking error $e_1(\cdot)$ but rather on a normalized estimation error of a particular parametrization of the plant. For this framework, the momentum-based integral gradient algorithm (20)–(23), the momentum-based recursive least squares algorithms (36)–(38), and momentum-based composite gradient algorithm (55)–(57) can be used directly for strictly proper plants without a relative degree restriction. In this case, the parametrization of the ideal controller is given by $u(s) = \theta^{*\mathrm{T}}\phi(s)$, where θ^* satisfies $\hat{z}(s) = \theta^{*\mathrm{T}}\hat{\phi}(s)$, $\hat{z}(s) = M(s)u_{\mathrm{p}}(s)$, and $\hat{\phi}(s) = [M(s)\hat{v}_1^{\mathrm{T}}(s), \hat{v}_2^{\mathrm{T}}(s), M(s)\hat{y}_{\mathrm{p}}(s), \hat{y}_{\mathrm{p}}(s)]$.*

5. Illustrative Numerical Examples

5.1. System Parameter Identification

Consider the third-order transfer function representing a servo control system for the pitch control of an aircraft given by

$$\hat{y}(s) = \frac{b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0} \hat{u}(s), \tag{101}$$

where $\hat{y}(s) \in \mathbb{C}$ is the pitch angle, $\hat{u}(s) \in \mathbb{C}$ is the system input, and b_1, b_0, a_2, a_1, $a_0 > 0$, and $b_0 \neq 0$ are the unknown system plant parameters. Note that, (68) gives $\hat{z}(s) = \frac{s^3}{\Lambda(s)}\hat{y}(s)$ with $\theta^* = [b_1, b_0, a_2, a_1, a_0]^{\mathrm{T}}$ and $\hat{\phi}(s) = [\frac{[s,1]}{\Lambda(s)}\hat{u}(s), \frac{-[s^2,s,1]}{\Lambda(s)}\hat{y}(s)]^{\mathrm{T}}$.

Let $a_0 = 0.1774$, $a_1 = 2.072$, $a_2 = 0.739$, $b_0 = 0.1774$, $b_1 = 1.151$, $u(t) = U[\cos(4t) + \cos(2t) + \sin(4t) + \sin(t)]$, $U > 0$, and choose $\Lambda(s) = (s+1)^3$. Note that $u(\cdot)$ is a stationary and sufficiently rich signal of order five, and hence, Theorem 4 guarantees that the estimated system parameters $\theta(t)$ will converge exponentially to θ^* as $t \to \infty$ using both the momentum-based integral gradient algorithm (20)–(23) and the MRLS algorithm (36)–(38).

First, we compare the performance of the RLS algorithm (34) and (35) with the MRLS algorithm (20)–(23). For this comparison, we set $U = 10$, $\Gamma(0) = 3000 I_5$, $\nu = 0.1$, $\beta = 0.1$, $\mu = \frac{2}{\beta}$, $\theta(0) = [0,0,0,0,0]^{\mathrm{T}}$, and $\mathcal{V}(0) = [0,0,0,0,0]^{\mathrm{T}}$. Figure 2 shows the system parameter estimate versus time. It can be seen that for both the RLS and the MRLS algorithms the parameter estimate $\theta(t)$ converges to θ^* as expected by Theorem 2. Moreover, as seen in Figure 3 the MRLS algorithm provides faster convergence of the system parameters to the true values as compared to the standard RLS algorithm.

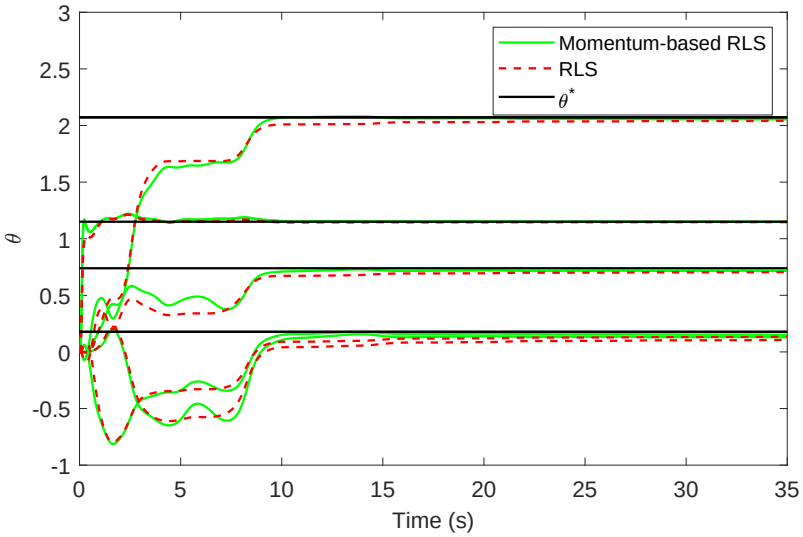

Figure 2. Parameter estimate values $\theta(t)$ versus time using the RLS and MRLS algorithms. In both cases the parameter estimate values converge to the true values with the MRLS algorithm converging faster.

Next, we compare the momentum-based composite gradient algorithm (55)–(57) with the *composite gradient algorithm* given by

$$\dot{\theta}(t) = -\Gamma(t)\phi(t)e(t) - \frac{\beta}{2}\Gamma(t)\mathcal{N}_{\mathrm{r}}(t)[R(t)\theta(t) + Q(t)], \quad \theta(0) = \theta_0, \quad t \geq 0. \tag{102}$$

Let $U = 1$, $\Gamma(0) = 1000 I_5$, $\nu = 0.2$, $\beta = 0.1$, $\mu = \frac{2}{\beta}$, $\theta(0) = [0,0,0,0,0]^{\mathrm{T}}$, and $\mathcal{V}(0) = [0,0,0,0,0]^{\mathrm{T}}$. Figure 4 shows the system parameter estimates versus time. It can be seen that for both the composite gradient and the momentum-based composite gradient algorithms the parameter estimate $\theta(t)$ converges to θ^* as expected by Theorem 3. However, as seen in

Figure 5, our proposed algorithm provides faster convergence of the system parameters to their true values. In particular, the momentum-based composite gradient settles around a value of 10^{-3} at $t = 16$ s as compared to 20 s for the composite gradient. Note that the convergence rate due to the addition of momentum is more pronounced in the case of the composite algorithm as compared to the RLS algorithm.

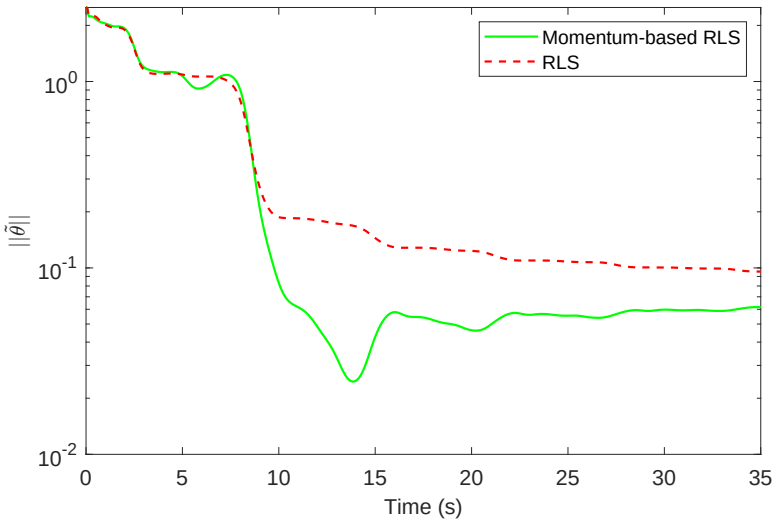

Figure 3. Norm of the parameter estimate error $\tilde{\theta}(t)$ versus time using the RLS and MRLS algorithms.

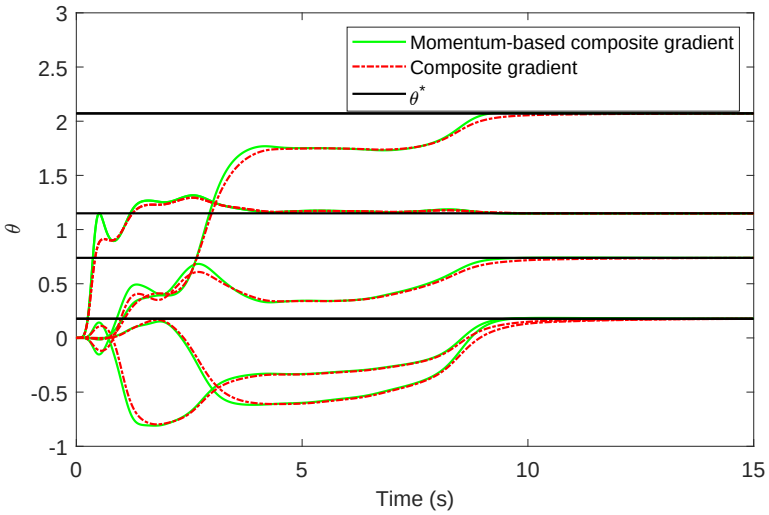

Figure 4. Parameter estimate values $\theta(t)$ versus time using the composite gradient and the momentum-based composite gradient algorithm. In both cases the parameter estimate values converge to the true values with the momentum-based integral algorithm converging slightly faster.

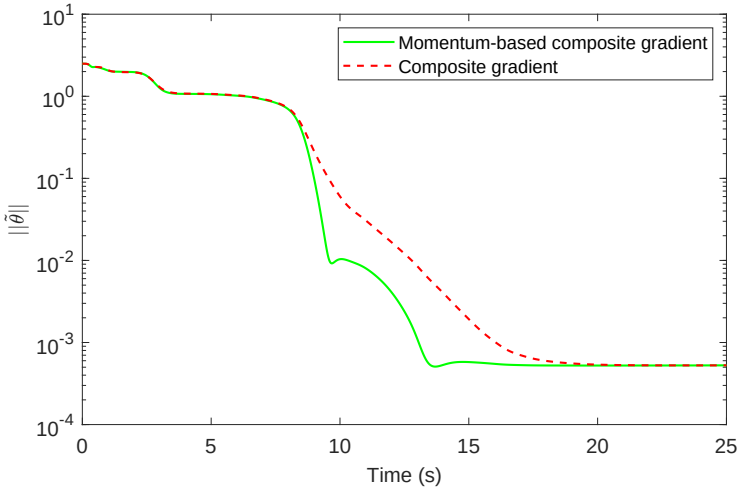

Figure 5. Norm of the parameter estimate error $\tilde{\theta}(t)$ versus time using the composite and the momentum-based composite algorithms.

5.2. Momentum-Based Recursive Least Squares and Model Reference Adaptive Control

Here, we consider the short-term dynamics of the aircraft as an example, as shown in Figure 6, where α is the angle of attack, γ is the flight path angle, and U_e is the equilibrium linear axial velocity. The transfer function describing the short-term dynamics of the reduced-order longitudinal state equation of the aircraft with angle of attack output and elevator deflection input η is given by [21]

$$\frac{\hat{\alpha}(s)}{\hat{\eta}(s)} = \frac{k_{\mathrm{p}}(s+b_0)}{s^2+a_1s+a_0}, \tag{103}$$

where $k_{\mathrm{p}} = \frac{z_\eta}{U_e}$, $b_0 = U_e\frac{m_\eta}{z_\eta}$, $a_1 = m_q + z_w$, and $a_0 = m_q z_w - m_w U_e$, and m_η, m_w, and m_q are the concise partial derivatives (see [21]) of the pitching moment with respect to the elevator angle η, the normal velocity w, and the pitch rate q. Moreover, z_η and z_w denote the concise partial derivatives of the normal force on the aircraft with respect to the elevator angle η and the normal velocity w. Here, we set $k_{\mathrm{p}} = -0.1601$, $b_0 = 71.9844$, $a_1 = 5.0101$, and $a_0 = 12.9988$ [21].



Figure 6. Schematic of an aircraft.

The desired system performance is to track the reference output

$$\hat{y}_{\mathrm{r}}(s) = \frac{5}{s+0.3}\hat{r}(s), \tag{104}$$

where $\hat{r}(s)$ is a reference command signal. Next, we use the filter system (78) and (79) with $F_{\mathrm{c}} = -2$ and $g_{\mathrm{c}} = 1$, and the control law (82) whose state space realization is given by

$$\dot{\xi}(t) = -\lambda_1\xi(t) + \phi(t), \quad \xi(0) = 0, \quad t \geq 0, \tag{105}$$

$$u_{\mathrm{p}}(t) = \theta^{\mathrm{T}}(t)\phi(t) + \dot{\theta}^{\mathrm{T}}(t)\xi(t), \tag{106}$$

where $\phi(t) = [v_1(t), v_2(t), y_{\mathrm{p}}(t), r(t)]^{\mathrm{T}}$. For our simulations, we select the reference signal to be a square wave with a frequency 2π and amplitude 1. Furthermore, the system initial conditions are set to $y(0) = 0$, $y_p(0) = 0$, $\dot{y}_p(0) = 0$, and $\Gamma(0) = 300I_4$.

Next, we compare the RLS algorithm (89) and (90) with our proposed MRLS algorithm (91)–(93). For this comparison, we set $\lambda_1 = 2$, $\alpha = 1$, $\beta = 1$, $\mu = 5$, $\gamma = 1$, $\nu = 0.3$, $\theta(0) = [0,0,0,0]^{\mathrm{T}}$, and $\mathcal{V}(0) = [0,0,0,0]^{\mathrm{T}}$. Note that all the conditions of Theorem 5 are satisfied. Figure 7 shows the system parameters versus time for the MRLS and the RLS algorithms. It can be seen that in both cases the system parameters converge to the true values with the MRLS providing faster convergence as compared to the standard RLS update law. Finally, Figure 8 shows the moving average (MA) of the absolute value of the tracking error $e_1(\cdot)$ with a sliding window of period 2π. Note that the MRLS algorithm provides slightly better tracking accuracy as compared to the RLS algorithm after the first 10 s. Finally, we note that the difference in runtime complexity between the different algorithms addressed in this paper is negligible.

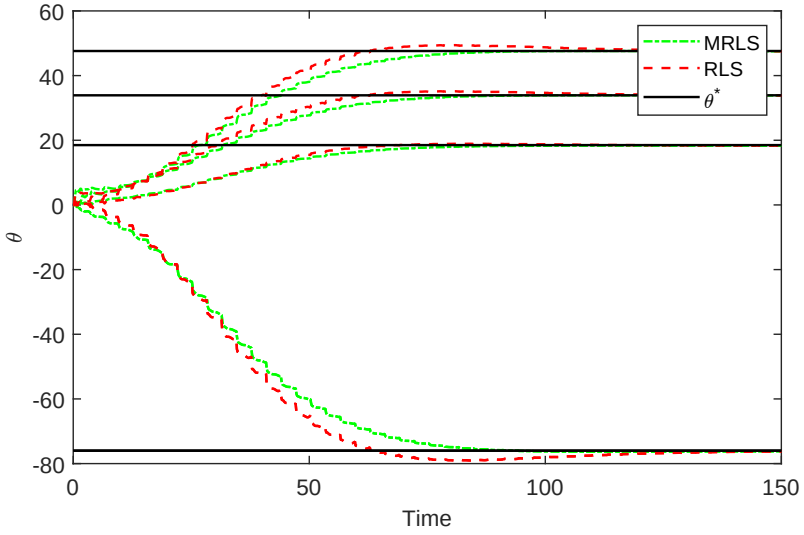

Figure 7. Parameter estimate values $\theta(t)$ versus time using the RLS and MRLS algorithms. In both cases the parameter estimate values converge to the true values with the MRLS algorithm converging faster.

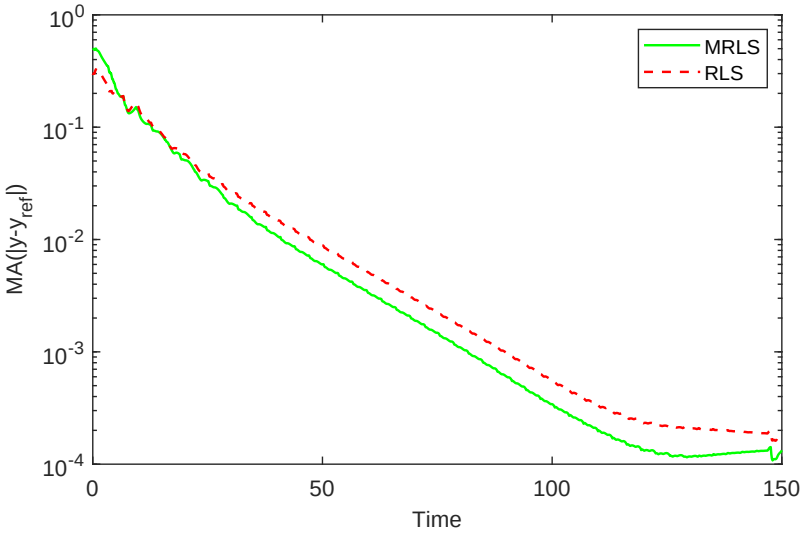

Figure 8. Moving average of the absolute value of the error $e_1(t) = y_{\mathrm{p}}(t) - y_{\mathrm{r}}(t)$ for the RLS and MRLS algorithms with a sliding window of 2π. The tracking error for the MRAC scheme predicated on the MRLS algorithm decreases faster than that predicated on the RLS algorithm.

6. Conclusions

In this paper, we developed three new momentum-based update laws for online parameter identification and model reference adaptive control. Specifically, we augmented higher-order tuner architectures into the integral gradient, recursive least squares, and composite gradient algorithms to achieve faster error convergence of the system parameters. Several numerical examples were provided to show the efficacy of the proposed approach. Future work will focus on developing new adaptive update laws for identification and control that guarantee finite time and fixed time convergence of the system parameters.

Author Contributions: L.S.: Conseptualization, Formal analysis, Software, Visualization, Writing-original draft. W.M.H.: Conceptualization, Formal analysis, Writing-review and editing, Supervision, Funding Acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Air Force Office of Scientific Research under Grant FA9550-20-1-0038.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Åström, K.J.; Wittenmark, B. *Adaptive Control*; Dover Publications: Mineola, NY, USA, 2008.
2. Ioannou, P.; Sun, J. *Robust Adaptive Control*; Dover Publications: Garden City, NY, USA, 2012.
3. Narendra, K.; Annaswamy, A.M. *Stable Adaptive Systems*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1989.
4. Krstic, M.; Kanellakopoulos, I.; Kokotovic, P. *Nonlinear and Adaptive Control Design*; Wiley: New York, NY, USA, 1995.
5. Cui, Y.; Annaswamy, A.M. Discrete-Time High Order Tuner with A Time-Varying Learning Rate. In Proceedings of the 2023 American Control Conference, San Diego, CA, USA, 31 May–2 June 2023; pp. 2993–2998.
6. Nesterov, Y. *Introductory Lectures on Convex Optimization*; Springer: New York, NY, USA, 2004.
7. Gaudio, J.E.; Gibson, T.E.; Annaswamy, A.M.; Bolender, M.A. Provably Correct Learning Algorithms in the Presence of Time-Varying Features Using a Variational Perspective. *arXiv* **2019**, arXiv:1903.04666. http://arxiv.org/abs/1903.04666.
8. Boffi, N.M.; Slotine, J.J.E. Implicit Regularization and Momentum Algorithms in Nonlinearly Parameterized Adaptive Control and Prediction. *Neural Comput.* **2021**, *33*, 590–673. [CrossRef] [PubMed]
9. Online accelerated data-driven learning for optimal feedback control of discrete-time partially uncertain systems. *Int. J. Adapt. Control Signal Process.* **2023**, *38*, 848–876.
10. Costa, R.R. Model-reference adaptive control with high-order parameter tuners. In Proceedings of the 2022 American Control Conference, Atlanta, GA, USA, 8–10 June 2022; pp. 3370–3375.
11. Costa, R.R. Least-squares model-reference adaptive control with high-order parameter tuners. *Automatica* **2024**, *163*, 111544. [CrossRef]
12. Wibisono, A.; Wilson, A.C.; Jordan, M.I. A variational perspective on accelerated methods in optimization. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7351–7358. [CrossRef] [PubMed]
13. Cho, N.; Shin, H.S.; Kim, Y.; Tsourdos, A. Composite Model Reference Adaptive Control with Parameter Convergence Under Finite Excitation. *IEEE Trans. Autom. Control* **2018**, *63*, 811–818. [CrossRef]
14. Shaferman, V.; Schwegel, M.; Glück, T.; Kugi, A. Continuous-time least-squares forgetting algorithms for indirect adaptive control. *Eur. J. Control* **2021**, *62*, 105–112. [CrossRef]
15. Haddad, W.M.; Chellaboina, V. *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*; Princeton University Press: Princeton, NJ, USA, 2008.
16. Goodwin, G.; Mayne, D. A parameter estimation perspective of continuous time model reference adaptive control. *Automatica* **1987**, *23*, 57–70. [CrossRef]
17. Gaudio, J.E.; Annaswamy, A.M.; Lavretsky, E.; Bolender, M.A. Parameter Estimation in Adaptive Control of Time-Varying Systems Under a Range of Excitation Conditions. *IEEE Trans. Autom. Control* **2022**, *67*, 5440–5447. [CrossRef]
18. Costa, R.R. Lyapunov design of least-squares model-reference adaptive control. *IFAC-PapersOnLine* **2020**, *53*, 3797–3802. [CrossRef]
19. Costa-Gomes, M.A.; Crawford, V.P.; Iriberri, N. Comparing models of strategic thinking in Van Huyck, Battalio, and Beil's coordination games. *J. Eur. Econ. Assoc.* **2009**, *7*, 365–376. [CrossRef]

20. Naik, S.; Kumar, P.; Ydstie, B. Robust continuous-time adaptive control by parameter projection. *IEEE Trans. Autom. Control* **1992**, *37*, 182–197. [CrossRef]
21. Cook, M. *Flight Dynamics Principles*; Elsevier: Oxford, UK, 2007.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.