2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Parameter Identification in Linear Error Equations: Guaranteeing Output Error Boundedness

Gilberto Pin, Yizhou Gong, Yang Wang and Andrea Serrani

***Abstract*— In this paper, we tackle the classical problem of estimating the parameters of an algebraic linear parameter model with the objective of solving the long-standing problem of guaranteeing boundedness of the output error independently from the growth of the regressors. Two solutions are presented. The first solution provides global results under the assumption that the time derivative of the regressor is available. The other solution disposes of the knowledge of the derivative of the regressor, and yields results that are valid in a semi-global sense, under the assumption that the regressor has a bounded growth. Simulation results provides an illustration of the proposed techniques in comparison with standard unnormalized and normalized gradient laws.**

I. INTRODUCTION

The estimation of the parameters of an algebraic linear model is one of the most basic subjects in adaptive identification, learning, and adaptive control monographs (see [1]–[5], to name a few). Of particular note is the gradient-based parameter estimator, which is quite standard yet powerful, and has received continued attention in the optimization community [6], and neural network learning community [7]. Tremendous efforts have been devoted to achieving faster learning by modifications of gradient-based estimation methods, (for instance, DREM-based estimation methods [8], [9], high-order tuner-based estimation methods [10], [11]), and various fixes, such as projections, deadzones, and integrator leakages [3], have been introduced in the gradient-based estimation methods to enhance robustness to model uncertainties. However, across these techniques, boundedness of the output error of the estimation model is often overlooked. Indeed, unless the regressor signals are assumed to be bounded, the output error can be persistently increasing in magnitude, which is undesirable in applications. Besides, boundedness of the output error may be critical for the stability of the closed-loop system if the parameter estimates are employed in a control law. Even if normalization is included in the adaptation law, the normalized output error is bounded while the output error is still not ensured to be bounded. Moreover, normalization methods tend to slow down the adaptation, leading to performance degradation [12].

Based on above considerations, an interesting question arises: *how to guarantee boundedness of the output error without requiring boundedness of the regressor, ensuring at the same time boundedness of the parameter estimates and asymptotic convergence of the output error?*

This question is answered in this paper, under certain assumptions. By modifying the conventional gradient-based parameter estimator, we achieve boundedness directly from Lyapunov arguments without invoking signal chasing arguments and/or asking for the boundedness of the regressor. To be specific, we incorporate the output error term into the designed Lyapunov-like function, with the consequence that the bound on the output error can be made independent of the regressor's magnitude and basically dependent only on the initial mismatch. Further, the estimated parameters under the proposed adaptation laws are able to exponentially converge to their true values whenever *Persistent Excitation* (PE) of the regressor holds, which makes our results self-contained.

The remainder of the paper is organized as follows. Section II formulates the linear output error equation for parameter learning and presents the most common methodologies. In Section III, two different adaptation laws are presented. The first solution is based on the accessibility of the regressor's derivative, which provides global results. The second solution is applicable in absence of direct knowledge of the derivative of the regressor, but requires the assumption of bounded growth of the regressor itself. As opposed to the first method, the second solution provides results that are valid in a semi-global sense, that is, for arbitrary (but fixed) compact sets of initial conditions. The convergence properties of the proposed update laws are also characterized. Numerical examples are provided in Section IV to illustrate the performance of the adaptation laws compared with the conventional gradient-based adaptation laws. Conclusions and an outlook on future research plans are offered in Section V.

II. Problem Formulation

We consider the problem of estimating the parameter vector θ of the following paradigmatic algebraic linear parameter model

$$y(t) = \xi^\top(t)\theta \tag{1}$$

where $y \in \mathbb{R}$, $\xi \in \mathbb{R}^n$ denote known output and regressor signals respectively, and $\theta \in \mathbb{R}^n$ is a vector of unknown constant parameters. By making use of the known regressor $\xi(t)$, it is immediate to build a conventional output-estimation model in the form

$$\hat{y}(t) = \xi^\top(t)\hat{\theta}(t), \tag{2}$$

where $\hat{\theta}(t)$ are the estimated parameters, that are possibly time-varying being subjected to some form of adaptation

Gilberto Pin is with the Digital Technology Org. of Electrolux Italia S.p.A., Italy (e-mail: gilberto.pin@electrolux.com).
Yizhou Gong and Yang Wang are with the School of Information Science and Technology, Shanghai Technology University, China. This work was supported in part by the Yangfan Program of Shanghai, China, under Grant 21YF1429600.
Andrea Serrani is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus OH – USA.
Corresponding author: Y.Wang, wangyang4@shanghaitech.edu.cn

laws. We consider the possibility of "learning" the parameter without using batch process, but modifying dynamically $\hat{\theta}(t)$ on-line, by using the measurements $y(t)$ and $\xi(t)$ as they are collected.

The usual approach adopted to identify the parameter vector θ consists of resorting to a *Certainty Equivalence* (CE) adaptation law (namely the conventional gradient-based adaptation law), which takes the following form

$$\dot{\hat{\theta}}(t) = -\gamma \xi(t) \tilde{y}(t), \quad \gamma > 0 \tag{3}$$

or is equipped with a normalization that permits to discount the magnitude of the regressor $\xi(t)$:

$$\dot{\hat{\theta}}(t) = -\gamma \frac{\xi(t)}{1 + \xi^\top(t)\xi(t)} \tilde{y}(t). \tag{4}$$

In (3) and (4), the parameter error is denoted by $\tilde{\theta}(t) := \hat{\theta}(t) - \theta$, and the output error $\tilde{y}(t) := \hat{y}(t) - y(t)$ is described by

$$\tilde{y}(t) = \xi^\top(t) \tilde{\theta}(t). \tag{5}$$

The stability of the aforementioned adaptation laws (3), (4) has been thoroughly studied by means of Lyapunov-like arguments that permit to prove boundedness of the parameter estimation error $\tilde{\theta}(t)$ and asymptotic convergence of $\tilde{y}(t)$ to zero, under mild assumption on the regressor $\xi(t)$ and its first time-derivative, but do not allow to guarantee boundedness of $\tilde{y}(t)$ in the general case (for instance, regressor is unbounded). Indeed, using the classic candidate Lyapunov-like function

$$V(\tilde{\theta}) = \frac{1}{2} \tilde{\theta}^\top \tilde{\theta}.$$

It is readily seen that the Lie derivative along (3) is

$$\dot{V}(t) = -\gamma \tilde{\theta}^\top(t) \xi(t) \tilde{y}(t) = -\gamma \tilde{y}(t)^2 \leq 0, \ \ \forall t \geq 0,$$

and along (4) is

$$\dot{V}(t) = -\gamma \frac{\tilde{y}(t)^2}{1 + \xi^\top(t)\xi(t)} \leq 0, \ \ \forall t \geq 0,$$

which implies that V along system's trajectories is bounded by its initial condition (i.e., the norm of the parameter error vector $\tilde{\theta}(t)$ is bounded), but no conclusion can be drawn on the boundedness of $\tilde{y}(t)$ from the analysis of the Lyapunov function alone, as $\tilde{y}(t)$ does not appear as a constituent of $V(\tilde{\theta})$. Obviously, from the error model (5), we can readily obtain the bound

$$|\tilde{y}(t)| \leq \|\xi(t)\| \|\tilde{\theta}(t)\| \leq \|\xi(t)\| \sqrt{2V\left(\tilde{\theta}(0)\right)},$$

which clearly shows that the output error bound depends on the magnitude of the regressor. Surprisingly, the authors are unaware of any results that establish independence of the output error from the magnitude (and growth) of the regressor. Available results are mainly focused on the asymptotic convergence properties of the estimators and on the sole boundedness of the parameters. Asymptotic convergence of $\tilde{y}(t)$ follows by application of *Barbalat's Lemma* [3], whenever $\tilde{y}(t)$ is uniformly continuous, which in turn can be established by assuming boundedness of both the regressor vector $\xi(t)$ and its time-derivative. Note that asymptotic convergence of the output error follows without any assumption of PE on the regressor, which instead is required to guarantee asymptotic convergence of the parameter estimation error.

To ensure boundedness of the output error independently from the magnitude of the regressor, we will adopt a Lyapunov-like function similar to the one commonly used in the relative-degree-unitary adaptive control problems [3], that does not only take as a constituent block the parameter error, but also the output error:

$$V(\tilde{y}, \tilde{\theta}) = \frac{1}{2} \tilde{y}^2 + \frac{1}{2\mu} \tilde{\theta}^\top \tilde{\theta}, \quad \mu > 0. \tag{6}$$

Clearly, if one can prove that $\dot{V}(t) \leq 0, \ \forall t \geq 0,$ then $V(\tilde{y}(t), \tilde{\theta}(t)) \leq V(\tilde{y}(0), \tilde{\theta}(0)), \ \forall t \geq 0$, and in turn one can conclude that both $\tilde{y}(t)$ and $\tilde{\theta}(t)$ are bounded by their initial conditions. In this context, we will present in the sequel two parameter adaptation laws fitting the above requirements. The key ingredient of both adaptation laws is a direct feedthrough from the measured output, thus the estimated parameter will not be the classical ones based on pure-integration. The first adaptation law will require the knowledge of $\dot{\xi}(t)$ (measurable), which is available in some applications. The second adaptation law will only require that $\dot{\xi}(t)$ is bounded with an upper bound on its norm known conservatively and that a bound on the parameter estimation error is also known conservatively. Note that this last requirement is not restrictive compared to the standard assumptions formulated in the literature about the boundedness of the regressor's time-derivative, which is usually invoked to establish the uniform continuity condition required to use the *Barbalat's lemma* to prove asymptotic convergence of the output error. Thus, we will prove that the exploitation of a conservative norm bound on the regressor's time-derivative in the proposed non-CE adaptation law is able to ensure boundedness of the output error, with a bound that does not depend on the norm of the regressor. Indeed, with conventional adaptive laws (3), (4), known bounds on the output error are in turn dependent on the norm of the regressor. As opposed to the first method, which yields global results, the second approach provides a guaranteed domain of validity for arbitrary fixed compact sets of initial conditions.

III. ADAPTATION LAWS WITH DIRECT OUTPUT FEEDTHROUGH

In this section, we present two adaptation laws characterized by the fact that the estimated parameter vector $\hat{\theta}(t)$ contains a direct feedthrough from the observed output $y(t)$. Consequently, the update laws depart from the classical dynamic structure obtained from certainty-equivalence design.

A. The case of known derivative of the regressor

When the derivative of the regressor, $\dot{\xi}(t)$, is known, then it is possible to exploit this additional information to modify the usual CE adaptive law to enforce boundedness of the output error. Consider the following adaptation law comprising these key ingredients: a direct output-to-parameter-estimate feedthrough ($y(t)$ to $\hat{\theta}(t)$), a regressor-derivative injection

($\dot{\xi}(t)$), and a filtered regressor's auto-covariance matrix $\Xi(t)$:

$$
\begin{aligned}
&\hat{\theta}(t) = \left(I + \Xi(t) + \mu\xi(t)\xi^\top(t)\right)^{-1}\left(\mu\xi(t)y(t) + \hat{\vartheta}(t)\right)\\
&\dot{\Xi}(t) = -\phi\Xi(t) + (1-\phi)\,\xi(t)\xi^\top(t)\\
&\dot{\hat{\vartheta}}(t) = \dot{\Xi}(t)\hat{\theta}(t) + \mu\dot{\xi}(t)\tilde{y}(t) - \mu\gamma\xi(t)\tilde{y}(t)\\
&\Xi(0) = 0_{n,n} \in \mathbb{R}^{n\times n}\\
&\hat{\vartheta}(0) = \hat{\vartheta}_0 \in \mathbb{R}^n
\end{aligned} \tag{7}
$$

where $\mu > 0$ is an arbitrary positive constant parameter adaptation gain, $\phi : 0 < \phi < 1$ is a user-defined forgetting factor, $\gamma > 0$ is a positive constant gain useful to tune the convergence speed of the output error, and $\hat{\vartheta}_0$ is an initial parameter guess.

The convergence properties of the proposed adaptation law are stated in the following theorem:

Theorem 3.1: Consider the model (1), where the regressor $\xi(\cdot) : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is a known continuously differentiable signal, with known derivative. Let the gain γ in (7) be selected so that $\gamma > (1-\phi)/(2\mu)$. Then, the adaptation law (7) applied to the linear output error equation (5) guarantees the following properties, for all $\hat{\vartheta}_0 \in \mathbb{R}^n$ and all $\theta \in \mathbb{R}^n$:

(i) $\hat{\theta}(\cdot) \in \mathcal{L}_\infty, \tilde{y}(\cdot) \in \mathcal{L}_\infty \cap \mathcal{L}_2$;

(ii) if $\xi(\cdot), \dot{\xi}(\cdot) \in \mathcal{L}_\infty$, then in addition to property (i), $\tilde{y}(t) \to 0$ as $t \to \infty$;

(iii) if $\xi(\cdot), \dot{\xi}(\cdot) \in \mathcal{L}_\infty$ and $\xi(\cdot)$ is PE, then in addition to properties (i), (ii), $\tilde{y}(t), \tilde{\theta}(t)$ converge to zero exponentially fast.

Proof: Fix, arbitrarily, $\hat{\vartheta}_0 \in \mathbb{R}^n$ and $\theta \in \mathbb{R}^n$. By left-multiplying both sides of the first identity in (7) by the nonsingular matrix $I + \mu\xi(t)\xi^\top(t) + \Xi(t)$ and rearranging terms, one obtains the following simplified expression for $\hat{\theta}(t)$:

$$
\hat{\theta}(t) = -\mu\xi(t)\tilde{y}(t) - \Xi(t)\hat{\theta}(t) + \hat{\vartheta}(t). \tag{8}
$$

The time-derivative of $\hat{\theta}(t)$ in (8) reads as

$$
\begin{aligned}
\dot{\hat{\theta}}(t) &= \cancel{-\mu\dot{\xi}(t)\tilde{y}(t)} - \mu\xi(t)\dot{\tilde{y}}(t) - \cancel{\dot{\Xi}(t)\hat{\theta}(t)} - \Xi(t)\dot{\hat{\theta}}(t)\\
&\quad + \cancel{\dot{\Xi}(t)\hat{\theta}(t)} + \cancel{\mu\dot{\xi}(t)\tilde{y}(t)} - \mu\gamma\xi(t)\tilde{y}(t)\\
&= -\mu\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t) - \mu\xi(t)\xi^\top(t)\dot{\tilde{\theta}}(t) - \Xi(t)\dot{\hat{\theta}}(t)\\
&\quad - \mu\gamma\xi(t)\tilde{y}(t)\\
&= \left(\frac{1}{\mu}(I + \Xi(t)) + \xi(t)\xi^\top(t)\right)^{-1} \times\\
&\qquad \left(-\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t) - \gamma\xi(t)\tilde{y}(t)\right).
\end{aligned} \tag{9}
$$

To prove the stability properties, we will employ the following quadratic form, that will serve as a candidate Lyapunov function in the PE case to show the exponential convergence of the estimator, as well as a Lyapunov-like function in absence of PE:

$$
V_\Xi(t, \tilde{y}, \tilde{\theta}) = \frac{1}{2}\tilde{y}^2 + \frac{1}{2\mu}\tilde{\theta}^\top\left(I + \Xi(t)\right)\tilde{\theta}, \quad \mu > 0. \tag{10}
$$

This function is clearly inspired by (6), used in relative-degree one adaptive control problems, with the addition of time-varying matrix $\Xi(t)$ as a weighting term for the parameter error. The Lie derivative of (10) along the solutions of (7) and along the trajectory (5) reads as

$$
\begin{aligned}
\dot{V}_\Xi(t) &= \tilde{y}(t)\dot{\tilde{y}}(t) + \frac{1}{\mu}\tilde{\theta}^\top(t)(I + \Xi(t))\dot{\tilde{\theta}}(t)\\
&\quad + \frac{1}{\mu}\frac{\tilde{\theta}^\top(t)\dot{\Xi}(t)\tilde{\theta}(t)}{2}\\
&= \tilde{\theta}^\top(t)\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t) + \tilde{\theta}^\top(t)\xi(t)\xi^\top(t)\dot{\tilde{\theta}}^\top(t)\\
&\quad + \frac{1}{\mu}\tilde{\theta}^\top(t)\left(I + \Xi(t)\right)\dot{\tilde{\theta}}(t)\\
&\quad - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t) + \frac{1}{\mu}\frac{1-\phi}{2}\tilde{y}(t)^2\\
&= \tilde{\theta}^\top(t)\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t)\\
&\quad + \tilde{\theta}^\top(t)\left(\frac{1}{\mu}(I + \Xi(t)) + \xi(t)\xi^\top(t)\right)\dot{\tilde{\theta}}(t)\\
&\quad - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t) + \frac{1}{\mu}\frac{1-\phi}{2}\tilde{y}(t)^2
\end{aligned} \tag{11}
$$

Upon substitution of (9) in (11), we obtain the following Lie derivative of (10) along the trajectories of (7):

$$
\begin{aligned}
\dot{V}_\Xi(t) &= \cancel{\tilde{\theta}^\top(t)\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t)} - \cancel{\tilde{\theta}^\top(t)\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t)}\\
&\quad - \gamma\tilde{\theta}^\top(t)\xi(t)\tilde{y}(t)\\
&\quad - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t) + \frac{1}{\mu}\frac{1-\phi}{2}\tilde{y}(t)^2\\
&= -\left(\gamma - \frac{1}{\mu}\frac{1-\phi}{2}\right)\tilde{y}(t)^2 - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t).
\end{aligned} \tag{12}
$$

Choosing $\gamma > (1-\phi)/(2\mu)$, then $\dot{V}_\Xi(t) \leq 0,\ \forall t \geq 0$, hence $V_\Xi(t)$, along the system's trajectories, is bounded by its initial condition. Therefore, $\tilde{y}(\cdot), \tilde{\theta}(\cdot) \in \mathcal{L}_\infty$, thus $\hat{\theta}(\cdot) \in \mathcal{L}_\infty$. Integrating both sides of the last identity in (12) one immediately obtains that $\tilde{y}(\cdot) \in \mathcal{L}_2$. Moreover, if $\xi(t)$ and $\dot{\xi}(t)$, $t \geq 0$, are bounded, $\dot{\hat{\theta}}(t)$ is bounded from (9), which implies that $\tilde{y}(t)$, $t \geq 0$, is uniformly continuous. Hence, by *Barbalat Lemma*, it can be concluded that $\tilde{y}(t)$ converges asymptotically to zero. We are left to show that $\tilde{\theta}(t)$ converges exponentially to zero if $\xi(t)$ is PE. According to Lemma 6.1, proven in Appendix, the PE condition

$$
\int_{t-T_{PE}}^{t} \xi(\tau)\xi^\top(\tau)d\tau \geq \epsilon_{PE}I, \ \ \forall t \geq T_{PE}, \tag{13}
$$

for some $T_{PE}, \epsilon_{PE} > 0$, implies that there exists $\epsilon_\Xi > 0$ such that $\Xi(t) \geq \epsilon_\Xi I$ for all $t \geq T_{PE}$. This latter result, in turn, implies that

$$
\dot{V}_\Xi(t) \leq -\left(\gamma - \frac{1}{\mu}\frac{1-\phi}{2}\right)\tilde{y}(t)^2 - \frac{\epsilon_\Xi}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\tilde{\theta}(t), \tag{14}
$$

for all $t \geq T_{PE}$. Hence, assuming that $\xi(\cdot)$ is bounded and PE, it follows that $\Xi(t)$ is bounded for all $t \geq 0$ and uniformly positive definite, rendering $V_\Xi(\cdot)$ a quadratic, positive definite, radially unbounded and decrescent Lyapunov function, which is also strict. In view of (14) we conclude that in case of PE the function $V_\Xi(t)$ converges to zero exponentially fast along trajectories. Consequently, both $\tilde{y}(t)$ and $\tilde{\theta}(t)$ converge to zero exponentially fast. ■

Remark 3.1: Note that (8) is not implementable in this very form (as opposed to the equivalent expression (7), which instead is computable from available signals) as $\tilde{y}(t)$ implicitly needs $\hat{\theta}(t)$ to be available. Therefore, the adaptation law (7) is equivalent to the not-realizable direct-output-error feedthrough law (8). As such, matrix normalization is used in the present solution to overcome the lack of realizability for (8), not to enforce signal boundedness as instead is the case of the usual scalar normalization (see, e.g., (4)).

B. Unknown derivative of the regressor, with known bound

When the derivative of the regressor is unknown, one can exploit a known bound on the norm of the time-derivative of the regressor to introduce a discontinuous injection aimed at "dominating" the regressor's derivative. The price to pay, however unsurprisingly, is a semi-global domain of validity of the ensuing results.

More specifically, let $\ell > 0$ be a known constant such that $\sup_{t\geq 0} \|\dot{\xi}(t)\| \leq \ell$. Assume that the unknown parameter θ in the model (1) ranges within a known compact set $\Theta \subset \mathbb{R}^n$. Furthermore, we assume that the initial value of the regressor, $\xi(0)$, is taken within a known compact set $\mathcal{K}_\xi \subset \mathbb{R}^n$. It is stressed that both Θ and $\mathcal{K}_\xi$ can be selected arbitrarily, but must remain fixed to be used in the selection of the gains of the update law to be presented below. It is also noted that Θ and $\mathcal{K}_\xi$ implicitly determine a compact set $\mathcal{K}_y \subset \mathbb{R}$ for the initial condition $y(0)$, that is, $y(0) \in \mathcal{K}_y$, where

$$\mathcal{K}_y = \{y \in \mathbb{R} : |y| \leq \max |\xi^\top \theta|,\ \xi \in \mathcal{K}_\xi,\ \theta \in \Theta\} \quad (15)$$

In this scenario, the update law (7) is replaced by

$$\begin{aligned}
&\hat{\theta}(t) = \left(I + \Xi(t) + \mu\xi(t)\xi^\top(t)\right)^{-1} \left(\mu\xi(t)y(t) + \hat{\vartheta}(t)\right) \\
&\dot{\Xi}(t) = -\phi\Xi(t) + (1-\phi)\,\frac{\xi(t)\xi^\top(t)}{1+\xi^\top(t)\xi(t)} \\
&\dot{\hat{\vartheta}}(t) = \dot{\Xi}(t)\hat{\theta}(t) - \mu\delta\xi(t)\mathrm{sign}(\tilde{y}(t)) - \mu\gamma\xi(t)\tilde{y}(t) \\
&\Xi(0) = 0_{n,n} \in \mathbb{R}^{n\times n} \\
&\hat{\vartheta}(0) = \hat{\vartheta}_0 \in \Theta_\vartheta \subset \mathbb{R}^n
\end{aligned} \quad (16)$$

where $\delta > 0$ is a scalar and $\Theta_\vartheta \subset \mathbb{R}^n$ is a compact set, both to be determined. The compact set Θ_ϑ shall be selected "large enough" so that the set

$$\begin{aligned}
\hat{\Theta} := \Big\{&\hat{\theta} \in \mathbb{R}^n : \hat{\theta} = \left(I + \mu\xi\xi^\top\right)^{-1} \left(\mu\,\xi y + \hat{\vartheta}\right),\ \xi \in \mathcal{K}_\xi, \\
&y \in \mathcal{K}_y,\ \hat{\vartheta} \in \Theta_\vartheta\Big\},
\end{aligned} \quad (17)$$

which is the set of initial conditions $\hat{\theta}(0)$ for $\hat{\theta}(t)$, contains the parameter set Θ in its interior.

The convergence properties of the adaptation law (16) are summarized as follows:

Theorem 3.2: Let $\ell > 0$ be an upper bound on $\|\dot{\xi}(t)\|$, $t \geq 0$. Fix, arbitrarily, compact sets $\Theta, \mathcal{K}_\xi \subset \mathbb{R}^n$ for θ and $\xi(0)$, respectively, and let $\mathcal{K}_y$ and be defined as in (15). Select the compact set Θ_ϑ such that $\hat{\Theta}$ in (17) satisfies $\hat{\Theta} \supseteq \Theta$. Finally, as in Theorem 3.1, choose $\gamma > (1-\phi)/(2\mu)$.

Then, there exists a positive constant $\overline{\delta}$ such that for all $\delta \geq \overline{\delta}$, the adaptation law (16) guarantees that:

(i) $\hat{\theta}(\cdot) \in \mathcal{L}_\infty$, $\tilde{y}(\cdot) \in \mathcal{L}_\infty \cap \mathcal{L}_2$;

(ii) if $\dot{\xi}(\cdot) \in \mathcal{L}_\infty$, then in addition to property (i), $\tilde{y}(t) \to 0$ as $t \to \infty$;

(iii) if $\xi(\cdot) \in \mathcal{L}_\infty$ and $\xi(\cdot)$ is PE, then in addition to properties (i), (ii), $\tilde{y}(t), \tilde{\theta}(t)$ converge to zero exponentially fast.

Proof: Let $\theta \in \Theta$ and $\xi(\cdot)$ be such that $\xi(0) \in \mathcal{K}_\xi$ and $\|\dot{\xi}(t)\| \leq \ell$ for all $t \geq 0$, and fix, arbitrarily, $\hat{\vartheta}_0 \in \Theta_\vartheta$. Notice that, in correspondence with Θ, $\mathcal{K}_\xi$, and the sets defined in (15) and (17), compact sets $\tilde{\mathcal{K}}_y \subset \mathbb{R}$ and $\tilde{\Theta} \subset \mathbb{R}^n$ for the initial values $\tilde{y}(0)$ and $\tilde{\theta}(0)$, respectively, are determined via the relations

$$\begin{aligned}
\tilde{\mathcal{K}}_y = \big\{&\tilde{y} \in \mathbb{R} : |\tilde{y}| \leq \max |\xi^\top \theta| + \max |y|,\ \xi \in \mathcal{K}_\xi, \\
&\theta \in \hat{\Theta},\ y \in \mathcal{K}_y\big\} \\
\tilde{\Theta} := \big\{&\tilde{\theta} \in \mathbb{R}^n : \tilde{\theta} = \hat{\theta} - \theta,\ \hat{\theta} \in \hat{\Theta},\ \theta \in \Theta\big\}
\end{aligned} \quad (18)$$

Consequently, the differential equations for the output and parameter estimation errors

$$\dot{\tilde{y}}(t) = \dot{\xi}^\top(t)\tilde{\theta}(t) + \xi^\top(t)\dot{\hat{\theta}}(t)\,, \quad \dot{\tilde{\theta}}(t) = \dot{\hat{\theta}}(t) \quad (19)$$

take initial conditions $(\tilde{y}(0), \tilde{\theta}(0))$ in the compact set $\tilde{\mathcal{K}}_y \times \tilde{\Theta}$ for all $\theta \in \Theta$, all $\xi(0) \in \mathcal{K}_\xi$ and all $\hat{\vartheta}(0) \in \Theta_\vartheta$. For the time-varying system (19), consider again the Lyapunov function candidate (10). Due to normalization of the regressor on the right-hand of the dynamic equation for $\Xi(t)$ in (16), the trajectory $\Xi(t)$ is uniformly bounded. As a result, $V_\Xi(t, \tilde{y}, \tilde{\theta})$ is positive definite, radially unbounded and decrescent, as there exist globally quadratic and positive definite functions $W_1(\tilde{y}, \tilde{\theta})$ and $W_2(\tilde{y}, \tilde{\theta})$ satisfying

$$W_1(\tilde{y}, \tilde{\theta}) \leq V_\Xi(t, \tilde{y}, \tilde{\theta}) \leq W_2(\tilde{y}, \tilde{\theta})$$

for all $(\tilde{y}, \tilde{\theta}) \in \mathbb{R} \times \mathbb{R}^n$ and all $t \geq 0$. For $c > 0$, let $\Omega_c \subset \mathbb{R} \times \mathbb{R}^n$ denote the level set of the function $W_2(\tilde{y}, \tilde{\theta})$

$$\Omega_c := \big\{(\tilde{y}, \tilde{\theta}) \in \mathbb{R} \times \mathbb{R}^n : W_2(\tilde{y}, \tilde{\theta}) \leq c\big\}, \quad c > 0$$

and select $c > 0$ such that $\Omega_c \subset \tilde{\mathcal{K}}_y \times \tilde{\Theta}$. Since the matrix $I + \Xi(t) + \mu\xi(t)\xi^\top(t)$ is nonsingular, the first identity of (16) can be rewritten as $\hat{\theta}(t) = -\Xi(t)\hat{\theta}(t) - \mu\xi(t)\tilde{y}(t) + \hat{\vartheta}(t)$, whose time-derivative is

$$\begin{aligned}
\dot{\hat{\theta}}(t) &= -\cancel{\dot{\Xi}(t)\hat{\theta}(t)} - \Xi(t)\dot{\hat{\theta}}(t) - \mu\dot{\xi}(t)\tilde{y}(t) - \mu\xi(t)\dot{\tilde{y}}(t) \\
&\quad + \cancel{\dot{\Xi}(t)\hat{\theta}(t)} - \mu\delta\xi(t)\mathrm{sign}(\tilde{y}(t)) - \mu\gamma\xi(t)\tilde{y}(t) \\
&= -\Xi(t)\dot{\hat{\theta}}(t) - \mu\dot{\xi}(t)\tilde{y}(t) - \mu\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t) \\
&\quad - \mu\xi(t)\xi^\top(t)\dot{\hat{\theta}}(t) - \mu\delta\xi(t)\mathrm{sign}(\tilde{y}(t)) - \mu\gamma\xi(t)\tilde{y}(t) \\
&= \left(\frac{1}{\mu}(I + \Xi(t)) + \xi(t)\xi^\top(t)\right)^{-1} \times \\
&\qquad \Big(-\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t) - \gamma\xi(t)\tilde{y}(t) \\
&\qquad\quad - \dot{\xi}(t)\tilde{y}(t) - \delta\xi(t)\mathrm{sign}(\tilde{y}(t))\Big).
\end{aligned} \quad (20)$$

Upon substitution of (20) in (11), we obtain the following expression of the Lie derivative of (10) along the trajectories of the system:

$$
\begin{aligned}
\dot{V}_{\Xi}(t) &= \cancel{\tilde{\theta}^\top(t)\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t)} - \cancel{\tilde{\theta}^\top(t)\xi(t)\dot{\xi}^\top(t)\tilde{\theta}(t)} \\
&\quad - \gamma\tilde{\theta}^\top(t)\xi(t)\tilde{y}(t) - \tilde{\theta}^\top(t)\dot{\xi}(t)\tilde{y}(t) \\
&\qquad - \delta\tilde{\theta}^\top(t)\xi(t)\mathrm{sign}(\tilde{y}(t)) \\
&\qquad\quad - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t) + \frac{1}{\mu}\frac{1-\phi}{2}\tilde{y}(t)^2 \\
&= -\left(\gamma - \frac{1}{\mu}\frac{1-\phi}{2}\right)\tilde{y}(t)^2 - \tilde{\theta}^\top(t)\dot{\xi}(t)\tilde{y}(t) - \delta|\tilde{y}(t)| \\
&\quad - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t) \\
&\le -\left(\gamma - \frac{1}{\mu}\frac{1-\phi}{2}\right)\tilde{y}(t)^2 + |\tilde{\theta}^\top(t)\dot{\xi}(t)||\tilde{y}(t)| - \delta|\tilde{y}(t)| \\
&\quad - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t) \\
&\le -\left(\gamma - \frac{1}{\mu}\frac{1-\phi}{2}\right)\tilde{y}(t)^2 - \left(\delta - |\tilde{\theta}^\top(t)\dot{\xi}(t)|\right)|\tilde{y}(t)| \\
&\quad - \frac{1}{\mu}\frac{\phi}{2}\tilde{\theta}^\top(t)\Xi(t)\tilde{\theta}(t). \qquad (21)
\end{aligned}
$$

Choosing δ to satisfy the following inequality

$$\delta \ge \bar{\delta} := \ell \max_{\tilde{\theta}\in\Omega_c} |\tilde{\theta}| \qquad (22)$$

and recalling that $\gamma > (1-\phi)(2\mu)$, one obtains

$$\dot{V}_{\Xi}(t,\tilde{y},\tilde{\theta}) \le -\left(\gamma - \frac{1-\phi}{2\mu}\right)\tilde{y}^2 - \frac{\phi}{2\mu}\tilde{\theta}^\top\Xi(t)\tilde{\theta} \le 0 \qquad (23)$$

for all $(\tilde{y},\tilde{\theta}) \in \Omega_c$ and all $t \ge 0$. As a result, all solutions $\big(\tilde{y}(t),\tilde{\theta}(t)\big)$, $t \ge 0$, originating from $\tilde{\mathcal{K}}_y \times \tilde{\Theta}$ are bounded, i.e., $\tilde{y}(\cdot),\tilde{\theta}(\cdot) \in \mathcal{L}_\infty$. Integrating both sides of (23) we immediately obtain that $\tilde{y}(\cdot) \in \mathcal{L}_2$. Moreover, if $\xi(\cdot),\dot{\xi}(\cdot)$ are bounded, then from (20), $\dot{\hat{\theta}}(\cdot) \in \mathcal{L}_\infty$ as well, which implies that $\tilde{y}(t)$, $t \ge 0$, is uniformly continuous. Thus, by *Barbalat's Lemma* one can conclude that $\tilde{y}(t)$ converges to zero asymptotically. To prove exponential convergence of both $\tilde{y}(t)$ and $\tilde{\theta}(t)$ in case of PE, one can proceed according to the same line of reasoning used to prove Theorem 3.1. ■

IV. ILLUSTRATIVE SIMULATION STUDY

In this section, we conduct numerical experiments to validate the theoretical result of the proposed adaptive law. Consider a second-order SISO system in the form of

$$z(s) = \frac{b_1 s + b_0}{s^2 + a_1 s + a_0}\,[\![u(t)]\!]$$

where a_i, b_i, $i = 0,1$ are unknown parameters to be estimated. $\mathcal{F}[\![u]\!]$ denotes the output of the operator $\mathcal{F}(s)$ (in the Laplace domain) with the input $u(t)$. It is explicitly seen that the system is expressed as a dynamic model. Following the standard procedures of transferring the dynamic model into algebraic model, we first rewrite the system into

$$s^2[\![z(t)]\!] = -a_1 s[\![z(t)]\!] - a_0[\![z(t)]\!] + b_1 s[\![u(t)]\!] + b_0[\![u(t)]\!]$$

and filter both sides by the operator $\Lambda(s) = (s+1)^2$, which gives the targeted algebraic linear parameter model in the form of (1):

$$y(t) = \xi^\top(t)\theta,$$

where $y(s) := \frac{s^2}{\Lambda(s)}[\![z(t)]\!]$, and

$$
\begin{aligned}
\xi(s) &:= \left(\tfrac{s}{\Lambda(s)}[\![z(t)]\!] \quad \tfrac{1}{\Lambda(s)}[\![z(t)]\!] \quad \tfrac{s}{\Lambda(s)}[\![u(t)]\!] \quad \tfrac{1}{\Lambda(s)}[\![u(t)]\!]\right)^\top \\
\theta &:= \begin{pmatrix} -a_1 & -a_0 & b_1 & b_0 \end{pmatrix}^\top.
\end{aligned}
$$

We compare the proposed adaptive laws (7), (16) with the conventional ones (3), (4), under different choices of u. The unknown plant parameters are chosen as $a_1 = 1, a_0 = 1, b_1 = 1, b_0 = -1$. For the sake of a fair comparison, the tuning parameter γ in all four methods are the same, specifically, $\gamma = 10$ and the other parameters μ,ϕ,δ in our method are selected as $\mu = 0.8$, $\phi = 0.5$, and $\delta = 20$, which is sufficiently large. $\hat{\vartheta},\Xi$ and the plant are initialized to be $\hat{\vartheta}(0) = 0_{4,1}$, $\Xi(0) = 0_{4,4}$, and $z(0) = 2$. All the simulations are conducted under the Runge-Kutta integration method with fixed sampling interval $10^{-3}s$.

A. *Case 1:* $u(t) = 10\,t$

In the first case of $u(t) = 10\,t$, $\xi(t)$ is not PE and monotonically increasing, but has a bounded time-derivative, thus the requirement for the design of adaptation law (16) is fulfilled. As depicted in Fig.1, the output error $\tilde{y}(t)$ under the four adaptation laws, though converging to zero at the very beginning, behaves differently afterward. Since we choose a monotonically increasing signal $u(t)$, the regressor $\xi(t)$ is also monotonically increasing, which leads to the unboundedness of the output error under the CE adaptation laws without/with normalization. Consequently, the output error $\tilde{y}(t)$ under the conventional adaptation laws (3), (4) diverges, whereas the proposed adaptation laws guarantee the boundedness of the output error. Thanks to the output feedthrough, the output error remains bounded, which achieves the main objective of the proposed adaptation laws.

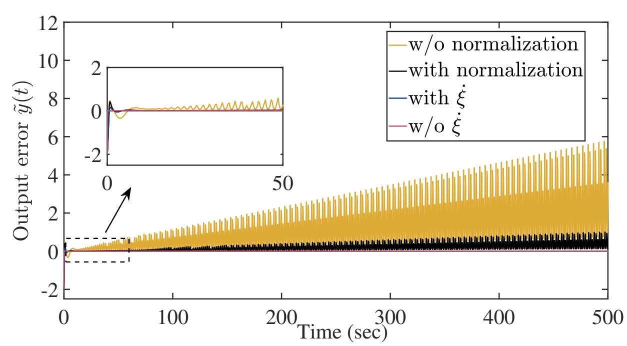

Fig. 1. Comparison of the time history of output error $\tilde{y}(t)$.

B. *Case 2:* $u(t) = \omega(t) + \nu(t)$

In the second case, we choose a sufficiently rich and bounded signal $\omega(t) = \sin(t) + \sin(3t)$ to verify the convergence properties of $\tilde{y}(t)$ and $\hat{\theta}(t)$. Also, we show the robustness of the proposed algorithm by selecting a random noise $\nu(t)$ with uniform distribution within the interval $[-0.1, 0.1]$. The linear parameter model now becomes $y(t) = \xi^\top(t)\theta + \omega_f(t)$, with $\omega_f(s) := \frac{s^2}{\Lambda(s)}[\![\omega(t)]\!]$. Utilizing [13, Corollary 5.2], we can compute the norm-bound of $\omega_f(t)$ as $\sup_{t\ge 0}|\omega_f(t)| \le 2.7066$. The outcomes of the adaptation laws in the noise-free and noisy cases are shown in Figs. 2-3. Figs. 2-3 demonstrate the exponential convergence of the estimated parameter error to zero without noise $\nu(t)$ and to a small residual set in the presence of noise $\nu(t)$.

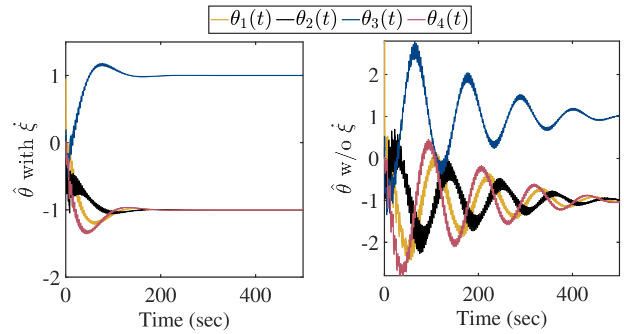

Fig. 2. Time history of $\hat{\theta}(t)$ converging to $\theta_1 = -1, \theta_2 = -1, \theta_3 = 1, \theta_4 = -1$ by adaptation laws (7), (16) without noise $\nu(t)$.

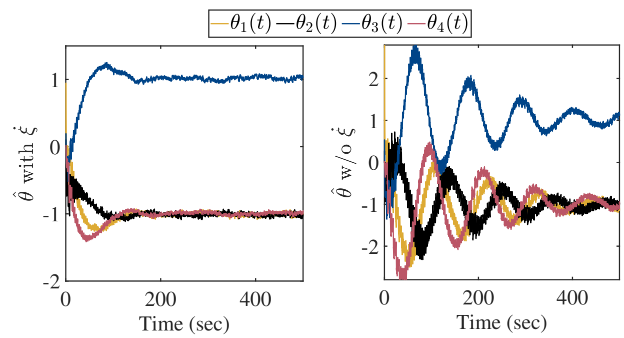

Fig. 3. Time history of $\hat{\theta}(t)$ converging to $\theta_1 = -1, \theta_2 = -1, \theta_3 = 1, \theta_4 = -1$ by adaptation law (7), (16) with noise $\nu(t)$.

V. CONCLUSION

In this paper, we have tackled the long-standing issue of guaranteeing, in the learning of algebraic linear-in-the parameter error models, a bound on the output error independent from the magnitude of the regressors, but depending only on the initial mismatch. We have presented two novel adaptation laws that yield Lyapunov-certified output error boundedness, even in absence of PE. As for conventional adaptation laws, in case of PE the proposed algorithms yield exponential convergence of the output and parameter estimation errors to zero. The outcomes of the simulations are consistent with the theoretical results and show that the proposed adaptation laws outperform the state of the art. Future research will be directed towards including the proposed adaptation laws in adaptive control problems.

VI. APPENDIX

Lemma 6.1: Let $\Xi(t)$ be a filtered regressor's autocovariance matrix evolving according to (7). If the PE condition (13) holds, then $\exists \epsilon_\Xi > 0$:

$$\Xi(t) \geq \epsilon_\Xi I, \ \ \forall t \geq T_{PE}.$$

Proof: Let $v \in \mathbb{R}^n$ be an arbitrary constant n-dimensional vector. From the dynamics of $\Xi(t)$ we obtain

$$v^\top \Xi(t) v = \int_0^t e^{-\phi(t-\tau)}(1-\phi) v^\top \xi(\tau)\xi(\tau)^\top v d\tau \qquad (24)$$

Given $0 < \phi < 1$, all terms in (24) are non-negative, therefore $v^\top \Xi(t) v \geq 0, \forall t \geq 0$, for any possible $v \in \mathbb{R}^n$. Moreover, for any $t_1, t_2 : 0 \leq t_1 < t_2$ we have

$$v^\top \Xi(t_2) v - v^\top \Xi(t_1) v \geq \int_{t_1}^{t_2} e^{-\phi(t_2-\tau)}(1-\phi) v^\top \xi(\tau)\xi(\tau)^\top v d\tau.$$

Considering that $v^\top \Xi(t_1) v \geq 0$, and that, for any $\phi : 0 < \phi < 1$ the following inequality holds:

$$e^{-\phi(t_2-\tau)}(1-\phi) \geq e^{-\phi(t_2-t_1)}(1-\phi), \ \ \forall \tau \in [t_1, t_2],$$

then, we get the lower bound

$$v^\top \Xi(t_2) v \ \geq e^{-\phi(t_2-t_1)}(1-\phi) \times v^\top \left(\int_{t_1}^{t_2} \xi(\tau)\xi(\tau)^\top d\tau \right) v$$

Finally, taking $t_2 = t$ and $t_1 = t - T_{PE}$, with $t \geq T_{PE}$, and using the PE assumption (13), we obtain

$$v^\top \Xi(t) v \geq e^{-\phi T_{PE}}(1-\phi) v^\top \epsilon_{PE} I v$$

for any possible $v \in \mathbb{R}^n$, which yields to the lower bound

$$\Xi(t) \geq \epsilon_\Xi I, \ \ \forall t \geq T_{PE}$$

with $\epsilon_\Xi = \epsilon_\Xi(\epsilon_{PE}) := e^{-\phi T_{PE}}(1-\phi)\epsilon_{PE}$. ■

REFERENCES

[1] K. Narendra and A. Annaswamy, *Stable Adaptive Systems*. Englewood Cliffs, NJ: Dover, 2005.

[2] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[3] P. Ioannou and J. Sun, *Robust Adaptive Control*. Englewood Cliffs, NJ: Prentice Hall, 1996.

[4] A. M. Annaswamy and A. L. Fradkov, "A historical perspective of adaptive control and learning," *Annual Reviews in Control*, vol. 52, pp. 18–41, 2021.

[5] D. Soudbakhsh, A. M. Annaswamy, Y. Wang, S. L. Brunton, J. Gaudio, H. Hussain, D. Vrabie, J. Drgona, and D. Filev, "Data-driven control: Theory and applications," in *2023 American Control Conference (ACC)*, 2023, pp. 1922–1939.

[6] D. Bertsekas, *Dynamic Programming and Optimal Control: Volume I*. Athena Scientific, 2012, vol. 4.

[7] F. Lewis, S. Jagannathan, and A. Yesildirak, *Neural network control of robot manipulators and non-linear systems*. CRC press, 1998.

[8] R. Ortega, S. Aranovskiy, A. A. Pyrkin, A. Astolfi, and A. A. Bobtsov, "New results on parameter estimation via dynamic regressor extension and mixing: Continuous and discrete-time cases," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2265–2272, 2021.

[9] M. Korotina, S. Aranovskiy, R. Ushirobira, D. Efimov, and J. Wang, "A note on fixed- and discrete-time estimation via the DREM method," *IEEE Transactions on Automatic Control*, pp. 1–6, 2024.

[10] A. S. Morse, "High-order parameter tuners for the adaptive control of nonlinear systems," in *Robust Control: Proceedings of a Workshop Held in Tokyo, Japan, June 23–24, 1991*. Springer, 1992, pp. 138–145.

[11] J. E. Gaudio, A. M. Annaswamy, M. A. Bolender, E. Lavretsky, and T. E. Gibson, "A class of high order tuners for adaptive systems," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 391–396, 2021.

[12] A. S. Morse, "A comparative study of normalized and unnormalized tuning errors in parameter adaptive control," *International Journal of Adaptive Control and Signal Processing*, vol. 6, no. 4, pp. 309–318, 1992.

[13] H. Khalil, *Nonlinear Systems*. Englewood Cliffs, NJ: Prentice Hall, 2001.