Systems & Control Letters 161 (2022) 105144

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Generalized recursive least squares: Stability, robustness, and excitation

Michelangelo Bin

Department of Electrical and Electronic Engineering, Imperial College London, London, UK

A R T I C L E   I N F O

Article history:
Received 14 July 2021
Received in revised form 7 December 2021
Accepted 11 January 2022
Available online 4 February 2022

Keywords:
Least squares
Robust adaptive control
Errors-in-variables identification

A B S T R A C T

We study a class of recursive least-squares estimators in an errors-in-variables setting where disturbances affect both the regressor and the regressand variables. We prove the existence and stability of an optimal steady state and robustness with respect to the disturbances in form of input-to-state and input–output stability relative to the unperturbed steady-state trajectories. Depending on the choice of some design parameters, different specific estimators can be realized within the considered class, each of which is associated with a different underlying optimization problem and with different excitation requirements for the unperturbed regressor. As expected, we find that persistence of excitation is associated with uniform, in fact exponential, convergence. In addition, we also show that choices of the design parameters are possible for which convergence and robustness hold without persistence of excitation and with the same asymptotic gain, the only difference being a loss of uniformity in the convergence rate.

1. Introduction

Many adaptive control problems reduce to find, asymptotically, an unknown parameter $\theta^\star$ in a linear regression of the form

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

given some measurements of $y^\star$ and $\phi^\star$. Typically, in control applications, $y^\star$ and $\phi^\star$ are not directly accessible and, instead, one measures the perturbed signals

$$y(t) := y^\star(t) + d_y(t), \quad \phi(t) := \phi^\star(t) + d_\phi(t), \tag{2}$$

in which $d := (d_y, d_\phi)$ is an unmeasured disturbance term. In fact, in practice measurements are never perfect and parameter estimators are often components of a larger control scheme where d is given by a combination of exogenous disturbances and other closed-loop signals.

When $d_\phi = 0$ and d_y is a purely exogenous stochastic process, estimating $\theta^\star$ in (1) is probably the most well-studied problem in statistical learning and system identification [1–4], where the focus is on the statistical characterization of the residual identification error and on the unbiasedness or consistency properties of the estimator. In control applications, however, d_ϕ is typically non-zero, making the estimation of $\theta^\star$ a considerably more challenging *errors-in-variables* problem [3]. Moreover, d is hardly characterizable in purely statistical terms, and typical assumptions (e.g., that d is white) do not hold. Indeed, d is, at least partly, endogenous, deterministic, and correlated with other closed-loop signals and even with the present and past estimates of $\theta^\star$ (this is typical of closed-loop operation [5–8]). As a consequence, classical results and methods of system identification are not directly applicable in the analysis of closed-loop control systems.

In this article, we approach the problem of estimating $\theta^\star$ under a different control-oriented perspective. We consider generic disturbance terms d on which we make no statistical assumption, and we shift the focus from unbiasedness and consistency to *robustness* of the estimate with respect to d, formalized in terms of *input-to-state stability* (ISS) [9,10]. In this way, as in [5–8], we enable the use of canonical nonlinear control techniques applying to ISS systems, such as *small-gain* methods [11], for the analysis of interconnections between controlled systems and identifiers. In this connection, the branch of system identification most related to this work is *set membership identification* [12,13], which deals with generic bounded disturbances. Nevertheless, to the best of the author's knowledge, asymptotic stability and ISS properties are not addressed in such literature, while they constitute the main focus of this article.

Furthermore, motivated again by control applications, in this article we limit our focus to estimation algorithms admitting a *recursive* formulation [14], i.e. methods where the estimate, say $\theta(t)$, of $\theta^\star$ is given by the output of a differential/difference equation of the form

$$\begin{aligned} \mathrm{D}x(t) &= f(x(t), y(t), \phi(t), t), \\ \theta(t) &= g(x(t), y(t), \phi(t), t), \end{aligned} \tag{3}$$

E-mail address: m.bin@imperial.ac.uk.

https://doi.org/10.1016/j.sysconle.2022.105144

in which $x(t)$ lies in some finite-dimensional normed space, and D represents the differentiation operator in continuous time and the one-step difference operator in discrete time.

The use of deterministic update laws of the form (3) in adaptive control can be traced back at least to Whitaker's heuristic gradient law [15], known as the "MIT rule" [16]. Estimation laws with stability guarantees started to be developed shortly thereafter based on Lyapunov analysis [17–19]. These are laws of the form (3) with $x = \theta$ and f chosen to guarantee that a given Lyapunov candidate has a negative derivative along the closed-loop solutions. Typically, the stability analysis focuses on the whole closed-loop system, and possible internal stability properties of the update laws per se are instead not studied. By the end of the 70s, Lyapunov-based designs were applied to many linear observation [20–22] and control [19,20,23,24] problems. Extensions to nonlinear systems pervaded most of the 80s and 90s [16,25–33], eventually leading to the development of adaptive backstepping [28–31,34,35] and to the idea of *modularization* [31,34,35], aimed at separating the roles of the estimation law and the rest of the controller. With some notable exceptions, such as the *immersion and invariance* approach [36], the *DREM* method [37,38], and the observer designs of [8,39–41], Lyapunov theory still constitutes the main way of design of estimation laws in control applications. See [33] for a comprehensive overview.

The vast majority of the aforementioned estimation schemes concern ideal cases where no disturbance is present (i.e., $d = 0$). Nevertheless, it is well-known since the 70s that adaptive controllers may suffer from critical robustness problems when disturbances add to the measured signals, to the point that bounded and vanishing disturbances may destabilize the closed-loop system [23,33,42–44]. The "weak link" is usually the estimation scheme, and the reason is that, as mentioned earlier, the estimation law is typically designed to impose a desirable condition on the entire closed-loop system, but the resulting system (3) does not typically possess any internal stability property necessary to cope with disturbances. To obviate this problems, many "robustification" measures have been developed in the years (see, e.g., [33,42,45–50]). However, most of them deal with prediction-error models where only d_y is present (i.e., $d_\phi = 0$). In turn, dealing with errors-in-variables problems where $d_\phi \neq 0$ is considerably more complex than dealing with just d_y (as a trivial example, notice that $d_\phi = -\phi^\star$ would produce a null regressor ϕ).

The last two decades saw also the emergence of adaptive *output regulation* [5–7,51–53]. Unlike adaptive stabilization, where the target steady state is typically a simple equilibrium and adaptation aims at finding a stabilizing controller, in adaptive output regulation the target steady state is generally a set where a non-trivial residual dynamics takes place (for instance, the controller must oscillate to counteract disturbances), and a robust stabilizer is generally assumed to be given. Adaptation, indeed, rather concerns the controller's *internal model*, with the goal of learning at run time the right "feedforward" control action coping with exogenous disturbances and tracking objectives. In regulation, some kind of residual excitation of the regressor is typically obtained also at the steady state, since the residual dynamics is not an equilibrium. Yet, especially in the recent identification-based designs of [6,7,53], the underlying adaptation task is an errors-in-variables problem of the kind (1)–(2), where neither d_y nor d_ϕ can be assumed zero nor exogenous stochastic processes with known characteristics. Dealing with such disturbances poses additional challenges, but also leads to robustness with respect to exogenous disturbances, unmodeled dynamics, and persistent prediction errors due to the fact that even optimal models may not be perfect. Moreover, it also supports an analysis based on small-gain conditions between the identifier and the rest of the closed-loop system that, as in [31,34,35], allows modularity in the choice of the identifier and stabilizer, which can be designed independently [6–8,53]. Yet, contrary to [31,34,35] where modularity requires a robustification of the controller, the approaches [6–8,53] require robustness is also for the identifier.

With the aim of fulfilling such robustness conditions, [6–8,53] proposed a robustified variation of Kreisselemeier's weighted least-squares estimator [22] possessing ISS-like properties with respect to both d_y and d_ϕ. In this paper, we revise and extend further this approach. Specifically, we study a class of generalized recursive least-squares schemes containing the identifier of [7,8] as a specific case. We prove robustness, in ISS terms, with respect to both d_y and d_ϕ. Moreover, we show that, for different choices of some design parameters, one can either achieve a uniform exponential convergence rate of the estimate by requiring usual *persistence of excitation* (PE) [54] of the regressor $\phi^\star$ or, rather, a non-uniform convergence rate by requiring, however, an excitation condition on $\phi^\star$ that is strictly weaker than PE. In particular, we find that the excitation assumptions necessary for convergence and robustness depend on the choice of some of the parameters only affecting the convergence rate and not the asymptotic gain. In turn, this implies the notable fact that PE is not necessary to obtain strong robustness properties such as ISS. In this connection, this article aligns with the recent trend in adaptive control studying parameter adaptation without PE (see, for instance, [38,55,56]).

The paper is organized as follows. Section 2 presents some preliminary notions. In Section 3, we formalize the problem. In Section 4, we describe the considered least-squares estimators and, in Section 5, we state and prove their main properties. In Section 6, we then discuss some choices of the degrees of freedom left open, and their connection with convergence rate and excitation requirements. Finally, Section 7 reports some concluding remarks.

2. Notation and preliminaries

We denote by $\mathbb{R}$ and $\mathbb{N}$ the set of real and natural numbers respectively ($0 \in \mathbb{N}$). If $\sim$ is a relation on a set S, for an $s \in S$ we let $S_{\sim s} := \{c \in S \ : \ c \sim s\}$. For $a, b \in \mathbb{N}$, we let $[a, b] := \{a, \dots, b\}$. The symbol $\subset$ denotes non-strict inclusion, and $A \setminus B$ the set difference between A and B. We denote by $|\cdot|_p$ the vector or induced matrix p-norm. We drop the subscript when $p = 2$ or the specific normed space is clear. If X is a subset of a normed vector space $\mathcal{X}$, $|X| := \sup\{|x| \ : \ x \in X\}$, and $d(a, X) := \inf_{x\in X} |x - a|$ denotes the distance of $a \in \mathcal{X}$ to X. If $f : \mathcal{X} \to \mathcal{Y}$, with $\mathcal{Y}$ a normed vector space, we let $|f|_X := |f(X)|$. If A is a square matrix, $A^\dagger$ denotes its Moore–Penrose pseudoinverse, and $\underline{\lambda}(A)$ its smallest eigenvalue. Given two square matrices A and B of the same dimension, we write $A \geq B$ if $A - B$ is positive semidefinite. If $A_1, \dots, A_n$ are matrices, we denote by $\operatorname{col}(A_1, \dots, A_n)$ and $\operatorname{diag}(A_1, \dots, A_n)$ their column and diagonal concatenations whenever they make sense.

A continuous function $\rho : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be of class-$\mathcal{K}$ ($\rho \in \mathcal{K}$) if $\rho(0) = 0$ and it is strictly increasing. A continuous function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be of class-$\mathcal{KL}$ ($\beta \in \mathcal{KL}$) if $\beta(\cdot, t) \in \mathcal{K}$ for all $t \geq 0$, and $\beta(s, \cdot)$ is strictly decreasing to zero for all $s \geq 0$. Given a function f, we denote by $\operatorname{dom} f$ its domain and by $\operatorname{ran} f$ its range.

Let $\mathcal{X}$, $\mathcal{Y}$ be normed vector spaces, $g : \mathcal{X} \to \mathcal{Y}$, and $X \subset \mathcal{X}$. A function $s : \mathcal{X} \to \mathcal{Y}$ is said to be of class $\mathcal{S}(X, g)$ ($s \in \mathcal{S}(X, g)$) if it satisfies $s(x) = g(x)$ for all $x \in X$, it is continuous everywhere if g is continuous on X, and $\operatorname{ran} s$ is bounded in $\mathcal{Y}$ if X is bounded in $\mathcal{X}$. Class-$\mathcal{S}(X, g)$ functions are generalized "saturated versions" of g. For instance, If $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}$, g is the identity, and $X = [-a, a]$ for some $a > 0$, the function s defined as $s(x) = x$ if $|x| \leq a$, and $s(x) = a \operatorname{sign}(x)$ otherwise, is of class $\mathcal{S}(X, g)$.

If X is bounded, the following property holds.

Lemma 1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be normed vector spaces, and let $X, X' \subset \mathcal{X}$ be bounded sets satisfying $X' \subset X$ and with the property that there exists $\nu > 0$ such that $d(x, X') \geq \nu$ for all $x \in \mathcal{X} \setminus X$. If $s \in \mathcal{S}(X, g)$ and g is Lipschitz on X with Lipschitz constant $L > 0$, then s satisfies $|s(x_1) - s(x_2)| \leq \max\{2c/\nu, L\}|x_1 - x_2|$ for all $x_1 \in X'$ and $x_2 \in \mathcal{X}$, where $c = |\operatorname{ran} s|$.*

Proof. Pick $x_1 \in X'$ and $x_2 \in \mathcal{X}$ arbitrarily. If $x_2 \in X$, then the claim trivially holds as g is Lipschitz on X with Lipschitz constant L. Otherwise, $x_2 \in \mathcal{X} \setminus X$, and the claim follows by the same arguments of [57, Lemma A.7]: by assumption, $d(x_2, X') \geq \nu$. As X is bounded, $c := |\operatorname{ran} s| < \infty$. Thus, $|s(x_1) - s(x_2)| \leq 2c \leq (2c/\nu)d(x_2, X') \leq (2c/\nu)|x_2 - x_1|$. □

3. Problem description

From a system-theoretic viewpoint, in absence of disturbances, identifying $\theta^\star$ in (1) from the *nominal data*[1] $y^\star$ and $\phi^\star$ is an observation problem for the initial state z of the linear time-varying system

$$y^\star(t) = \phi^\star(t)^\top z(t), \quad \mathrm{D}z(t) = 0, \quad z(0) = \theta^\star, \tag{4}$$

from the output $y^\star(t)$, and given the knowledge of the system's model represented by $\phi^\star$. In turn, a necessary and sufficient condition for identifiability, is that the Gramian matrix (equal to $\int_0^t \phi^\star(s)\phi^\star(s)^\top ds$ in continuous time and $\sum_{s=0}^t \phi^\star(s)\phi^\star(s)^\top$ in discrete time) is full-rank for some $t > 0$. This may suggest a basic recursive estimation law of the kind (3) obtained with $x = (x_1, x_2)$ and

$$\begin{aligned}
\mathrm{D}x_1(t) &= \phi(t)\phi(t)^\top, && x_1(0) = 0,\\
\mathrm{D}x_2(t) &= \phi(t)y(t), && x_2(0) = 0,\\
\theta(t) &= x_1(t)^\dagger x_2(t). &&
\end{aligned} \tag{5}$$

When $d = 0$, this law guarantees $\theta(t) = \theta^\star$ for all t for which the previously-defined Gramian is invertible. Nevertheless, (5) suffers from several drawbacks making it uninteresting for applications. First, it needs initialization. Indeed, if $x(0) \neq 0$, the estimate is biased. This means that, in case $\theta^\star$ changes even slightly during operation, (5) is not be able to track the new value. This problem is due to the lack of global convergence of (5). If, moreover, disturbances enter into play, then (5) provides unstable estimates even with d bounded and arbitrarily small. Therefore, even if (5) requires the weakest form of excitation on the regressor $\phi^\star$, being invertibility of the Gramian necessary for identifiability, it nevertheless provides a fragile estimation law unable to cope with practical needs. At the price of strengthen the requirements on $\phi^\star$, one is rather interested in estimation laws that are globally convergent and well-behaved in presence of disturbances. In particular, in this paper we consider the following requirements:

- **P1.** *Nominal Performance:* when $d = 0$, the estimate of $\theta^\star$ must converge to $\theta^\star$.
- **P2.** *Robustness:* when $d \neq 0$, the deviation of the resulting estimate from that attained with $d = 0$ must vary continuously with the asymptotic size of d.

In **P1**, convergence means $\lim_{t\to\infty} |\theta(t) - \theta^\star| = 0$, where $\theta(t)$ denotes the estimate of $\theta^\star$ at time t, although in the following we shall also consider approximate convergence in presence of regularization (see Proposition 3 and Section 6.1). In **P2**, on the other hand, robustness may be seen as an *asymptotic gain* property [10]. Both properties, together, can be grouped within a stronger ISS condition. Specifically, if the estimator is described by (3), then we ask for the existence of a *nominal steady-state pair* (x_{ss}, θ_{ss}), solution of (3) for $y = y^\star$ and $\phi = \phi^\star$ (i.e., when $d = 0$), such that θ_{ss} satisfies **P1**, and for every solution pair (x, d) to (3) with $d \neq 0$ the following hold

$$\begin{aligned}
|x(t) - x_{ss}(t)| &\leq \beta\big(|x(0) - x_{ss}(0)|, t\big) + \kappa\big(|d|_{[0,t]}\big)\\
|\theta(t) - \theta_{ss}(t)| &\leq \alpha(|x(t) - x_{ss}(t)|) + \rho\big(|d|_{[0,t]}\big)
\end{aligned}$$

for all t, where $\kappa, \alpha, \rho \in \mathcal{K}$ and $\beta \in \mathcal{KL}$. In turn, with $\omega = \alpha \circ \kappa + \rho \in \mathcal{K}$, this implies [10]

$$\limsup_{t\to\infty} |\theta(t) - \theta_{ss}(t)| \leq \omega\Big(\limsup_{t\to\infty} |d(t)|\Big), \tag{6}$$

which is **P2**. Hence, designing a recursive estimator of the kind (3) achieving **P1** and **P2** boils down to design a pair (f, g) guaranteeing the existence of a nominal steady state for the estimator (3) when driven by the nominal inputs $y^\star$ and $\phi^\star$, that is robustly stable (in the ISS sense) when (3) is driven by the corrupted inputs y and ϕ.

In the following, we present a generalized class of least-squares algorithm of the form (3) fulfilling **P1** and **P2**. For simplicity, we limit to the scalar case where $y^\star, d_y : \mathbb{R}_{\geq 0} \to \mathbb{R}$, $\phi^\star, d_\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_\theta}$, and $\theta^\star \in \mathbb{R}^{n_\theta}$ for some $n_\theta \in \mathbb{N}_{\geq 1}$. The case in which $y^\star(t) \in \mathbb{R}^m$ with $m > 1$ can be easily handled by the "concatenation" of m single-variable problems of the kind treated here. Moreover, we focus on discrete-time laws, as they are more interesting in applications. However, most of what said has a straightforward continuous-time counterpart.

Finally, we make the following assumption.

Assumption 1. $\phi^\star$ and $y^\star$ are bounded, i.e. there exist $\bar{\phi}, \bar{y} > 0$ such that $|\phi^\star(t)| \leq \bar{\phi}$ and $|y^\star(t)| \leq \bar{y}$ for all $t \in \mathbb{N}$.

4. Generalized recursive least squares

4.1. Basic definitions

In the following, we study a class of recursive estimators of the kind (3) having the form

$$\begin{aligned}
x_1(t+1) &= \mu(t)x_1(t) + \lambda(t)\sigma\big(\phi(t)\big)\\
x_2(t+1) &= \mu(t)x_2(t) + \lambda(t)\delta\big(\phi(t), y(t)\big)\\
\theta(t) &= \gamma\big(x_1(t), x_2(t), t\big)
\end{aligned} \tag{7}$$

in which $x_1(t) \in \mathbb{S}_{n_\theta}$, being $\mathbb{S}_{n_\theta}$ the set of symmetric positive semidefinite n_θ-by-n_θ matrices, $x_2(t), \theta(t) \in \mathbb{R}^{n_\theta}$, $\mu(t) \in [0, 1)$ and $\lambda(t) \in \mathbb{R}_{\geq 0}$ are such that[2]

$$\sum_{s=0}^{t-1}\left(\prod_{\tau=s+1}^{t-1}\mu(\tau)\right)\lambda(s) \leq 1, \qquad \forall t \in \mathbb{N}, \tag{8}$$

the inputs ϕ and y are given by (2), with $d = (d_y, d_\phi) : \mathbb{N} \to \mathbb{R} \times \mathbb{R}^{n_\theta}$, and σ, δ and γ are functions to be designed later. In the following, we let $x := (x_1, x_2) \in \mathbb{X} := \mathbb{S}_{n_\theta} \times \mathbb{R}^{n_\theta}$ and $|x(t)| := \max\{|x_1(t)|, |x_2(t)|\}$. Moreover, we call (x, d) a *solution pair* to (7) if either $\operatorname{dom}(x, d) = [0, n]$ for some $n \in \mathbb{N}$ or $\operatorname{dom}(x, d) = \mathbb{N}$, and if x solves (7) when ϕ and y are given by (2) with $(d_y, d_\phi) = d$.

[1] Throughout the article, the word "nominal" is used in reference to quantities associated with $d = 0$. Specifically, "nominal data" refers to $(y^\star, \phi^\star)$, representing the measured data (2) when $d = 0$.

[2] There is no loss of generality in assuming that the right-hand side of (8) equals 1 as long as the left-hand side is finite, as one can always rescale λ to obtain (8).

4.2. Design of σ and δ, and robust stability

In this section, we detail the design of the functions σ and δ in (7) and we prove that System (7) is ISS with respect to the disturbance d and relative to a time-varying steady state $x_{ss} = (x_{1,ss}, x_{2,ss})$ defined by the nominal data $(y^\star, \phi^\star)$. As a first step, we define two constants $c_1, c_2 \in \mathbb{R}_{>0} \cup \{\infty\}$ in such a way that

$$\bar{\phi} \le c_1, \qquad \bar{y} \le c_2. \tag{9}$$

We stress that c_1 and c_2 do not have to equal $\bar{\phi}$ and $\bar{y}$, and they are allowed to be ∞. As clarified later (see Proposition 1, and Remarks 1 and 2), choosing $c_1 = c_2 = \infty$ allows a simpler choice of σ and δ but leads to a quadratic asymptotic gain in the aforementioned ISS property. Instead, choosing c_1 and c_2 finite (e.g., $c_1 = \bar{\phi}$ and $c_2 = \bar{y}$) requires a more complex design of σ and δ but leads to a linear asymptotic gain.

Next, we construct σ and δ and prove two results (Lemmas 2 and 3) bounding the differences $\sigma(\phi^\star(t)) - \sigma(\phi(t))$ and $\delta(\phi^\star(t), y^\star(t)) - \delta(\phi(t), y(t))$ by a function of the norm of the disturbance d. These bounds are instrumental for the ISS results that follows. Define $B_1 := \{\phi \in \mathbb{R}^{n_\theta} : |\phi| \le c_1 + 1\}$, and let

$$g_1 : \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_\theta \times n_\theta}, \quad \phi \mapsto g_1(\phi) := \phi\phi^\top.$$

Then, we fix $\sigma : \mathbb{R}^{n_\theta} \to \mathbb{S}_{n_\theta}$ as any arbitrary function of class $\mathcal{S}(B_1, g_1)$ (see Section 2). If $c_1 < \infty$, then $\operatorname{ran}\sigma$ is bounded. In such case, we define $\bar{\sigma} := |\operatorname{ran}\sigma|$ and $\ell_1 := \max\{2\bar{\sigma}, 2c_1 + 2\}$. Then, σ satisfies the following property.

Lemma 2. *Suppose that Assumption 1 holds, and let $\rho_1 \in \mathcal{K}$ be defined as*

$$\rho_1(s) := \begin{cases} 2\bar{\phi}s + s^2 & \text{if } c_1 = \infty \\ \ell_1 s & \text{otherwise.} \end{cases}$$

Then, for every $t \in \mathbb{N}$, $|\sigma(\phi^\star(t)) - \sigma(\phi(t))| \le \rho_1(|d_\phi(t)|)$.

Lemma 2 is proved in the Appendix. Similarly, let $B_2 := \{(\phi, y) \in \mathbb{R}^{n_\theta} \times \mathbb{R} : |\phi|, |y| \le \max\{c_1, c_2\} + 1\}$ and define

$$g_2 : \mathbb{R}^{n_\theta} \times \mathbb{R} \to \mathbb{R}^{n_\theta}, \quad (\phi, y) \mapsto g_2(\phi, y) := \phi y.$$

Then, we fix δ as any arbitrary function of class $\mathcal{S}(B_2, g_2)$. As before, if $c_1, c_2 < \infty$, we let $\bar{\delta} := |\operatorname{ran}\delta|$ and $\ell_2 := \{2\bar{\delta}, 2\max\{c_1, c_2\} + 2\}$. Then, the following holds.

Lemma 3. *Suppose that Assumption 1 holds, and let $\rho_2 \in \mathcal{K}$ be defined as*

$$\rho_2(s) := \begin{cases} (\bar{\phi} + \bar{y})s + s^2 & \text{if } c_1 = \infty \text{ or } c_2 = \infty \\ \ell_2 s & \text{otherwise.} \end{cases}$$

Then, for every $t \in \mathbb{N}$, $|\delta(\phi^\star(t), y^\star(t)) - \delta(\phi(t), y(t))| \le \rho_2(|d(t)|)$.

The proof of Lemma 3 follows the same arguments of that of Lemma 2, and it is thus omitted.

Remark 1 (*Implementation of σ and δ*)**.** If $c_1 = c_2 = \infty$, then necessarily $\sigma = g_1$ and $\delta = g_2$, since in such case $B_1 = \mathbb{R}^{n_\theta}$ and $B_2 = \mathbb{R}^{n_\theta} \times \mathbb{R}$. Otherwise, σ and δ can be implemented, for instance, as component-wise saturation functions. Namely, with $\operatorname{sat}_k(\cdot) := \min\{\max\{-k, \cdot\}, k\}$, $\sigma(\phi)$ can be defined as the matrix with (i, j)th entry $\sigma(\phi)_{ij} := \operatorname{sat}_{c_1+1}(\phi_i)\operatorname{sat}_{c_1+1}(\phi_j)$, and $\delta(\phi, y)$ as the vector with ith component $\delta(\phi, y)_i := \operatorname{sat}_{(\max\{c_1,c_2\}+1)^2}(y\phi_i)$.

With these definitions in mind, we now show that there exists a steady-state trajectory $x_{ss} = (x_{1,ss}, x_{2,ss})$ which is robustly stable for (7). In particular, we define $x_{1,ss} : \mathbb{N} \to \mathbb{R}^{n_\theta \times n_\theta}$ and $x_{2,ss} : \mathbb{N} \to \mathbb{R}^{n_\theta}$ as

$$\begin{aligned} x_{1,ss}(t) &:= \sum_{s=0}^{t-1} \left(\prod_{\tau=s+1}^{t-1} \mu(\tau) \right) \lambda(s)\phi^\star(s)\phi^\star(s)^\top, \\ x_{2,ss}(t) &:= \sum_{s=0}^{t-1} \left(\prod_{\tau=s+1}^{t-1} \mu(\tau) \right) \lambda(s)\phi^\star(s)y^\star(s). \end{aligned} \tag{10}$$

Under Assumption 1, and by construction of σ and δ, $\sigma(\phi^\star(t)) = \phi^\star(t)\phi^\star(t)^\top$ and $\delta(\phi^\star(t), y^\star(t)) = \phi^\star(t)y^\star(t)$ for all $t \in \mathbb{N}$. Hence, it is easy to check that (10) is a solution of (7) starting at the origin and corresponding to $d = 0$. Moreover, we underline that, once μ and λ are fixed, the trajectories (10) only depend on the nominal data $\phi^\star$ and $y^\star$. Define the functions $\beta \in \mathcal{KL}$ and $\kappa \in \mathcal{K}$ as

$$\beta(s, t) := s \prod_{\tau=0}^{t-1} \mu(\tau), \quad \kappa(s) := \max\{\rho_1(s), \rho_2(s)\}, \tag{11}$$

where ρ_1 is given by Lemma 2, and ρ_2 by Lemma 3. Then, the following proposition establishes ISS of (7) relative to the steady-state trajectory x_{ss} and with respect to the disturbance d.

Proposition 1. *Suppose that Assumption 1 holds. Then, for every solution pair (x, d) of (7) and every $t \in \operatorname{dom}(x, d)$,*

$$|x(t) - x_{ss}(t)| \le \beta\big(|x(0) - x_{ss}(0)|, t\big) + \kappa\big(|d|_{[0,t-1]}\big). \tag{12}$$

In addition, if d is bounded and $\operatorname{dom}(x, d) = \mathbb{N}$, then

$$\limsup_{t\to\infty} |x(t) - x_{ss}(t)| \le \kappa\left(\limsup_{t\to\infty} |d(t)|\right). \tag{13}$$

The proof of Proposition 1 is given in the Appendix. In the next section, we show that x_{ss} is also optimal with respect to a specific weighted least-squares cost function.

Remark 2 (*Linear Gains*)**.** When $c_1, c_2 < \infty$, Lemmas 2 and 3 guarantee that the differences $\sigma(\phi^\star(t)) - \sigma(\phi(t))$ and $\delta(\phi^\star(t), y^\star(t)) - \delta(\phi(t), y(t))$ can be bounded by a *linear* function of the size of the disturbance d at time t, rather than a quadratic bound obtained when $c_1 = c_2 = \infty$. In turn, this implies that the asymptotic gain κ in (12) and (13) is linear. This "linear gain property" is used, for instance, in [7,53] to enforce a closed-loop small-gain condition by using a linear "high-gain" stabilization method.

4.3. Least-squares optimality of the steady state

From (1) and (10) we deduce

$$x_{1,ss}(t)\theta^\star = x_{2,ss}(t), \quad \forall t \in \mathbb{N}. \tag{14}$$

Hence, whenever $x_{1,ss}(t)$ is non-singular, $\theta^\star$ can be univocally identified from the sole knowledge of $x_{ss}(t)$.

More in general, the steady state x_{ss} univocally characterizes the optimal solutions to the least-squares optimization problem described hereafter. Consider the functional $J : \mathbb{N} \times \mathbb{R}^{n_\theta} \to \mathbb{R}_{\ge 0}$ mapping (t, θ) to

$$\begin{aligned} J(t, \theta) := &\sum_{s=0}^{t-1} \left(\prod_{\tau=s+1}^{t-1} \mu(\tau) \right) \lambda(s)\Big(y^\star(s) - \phi^\star(s)^\top\theta\Big)^2 \\ &+ (\theta - \theta_0(t))^\top \Omega(t)(\theta - \theta_0(t)), \end{aligned} \tag{15}$$

in which $\theta_0 : \mathbb{N} \to \mathbb{R}^{n_\theta}$ and $\Omega : \mathbb{N} \to \mathbb{S}_{n_\theta}$ are arbitrary bounded functions. In (15), the first term is a sum of the squares of the historical prediction errors $y^\star(s) - \phi^\star(s)^\top\theta$ produced by θ, weighted by the terms $\prod_{\tau=s+1}^{t-1} \mu(\tau)\lambda(s)$. The second is instead a *regularization* term penalizing the weighted distance of θ to $\theta_0(t)$, where the weights are defined by the *regularization matrix* $\Omega(t)$. We underline that $\theta_0 = 0$ and $\Omega = 0$ are feasible choices, leading to a pure unregularized least squares functional.

For each $t \in \mathbb{N}$, let

$$\Theta^\circ(t) := \left\{\theta \in \mathbb{R}^{n_\theta} \,:\, \forall \theta' \in \mathbb{R}^{n_\theta},\, J(t,\theta') \geq J(t,\theta)\right\}$$

be the set of all minimizers of $J(t,\cdot)$. Then, the following proposition (proved in the Appendix) shows that $\Theta^\circ(t)$ is univocally identified by $x_{ss}(t)$.

Proposition 2. *For all* $t \in \mathbb{N}$,

$$\Theta^\circ(t) = \left\{\theta \in \mathbb{R}^{n_\theta} : \big(x_{1,ss}(t) + \Omega(t)\big)\theta = x_{2,ss}(t) + \Omega(t)\theta_0(t)\right\}.$$

Clearly, the equality $(x_{1,ss}(t) + \Omega(t))\theta = x_{2,ss}(t) + \Omega(t)\theta_0(t)$ reduces to (14) when $\Omega(t) = 0$. Indeed, in such case $\theta^\star \in \Theta^\circ(t)$, while this is in general not true if $\Omega(t) \neq 0$. Indeed, a non-zero $\Omega(t)$ induces a bias between the optimal solutions minimizing $J(t,\theta)$ and the parameter $\theta^\star$ in (1) or, more in general, the quantity $x_{1,ss}(t)^{-1}x_{2,ss}(t)$ providing the unique least-squares solution to (15) when $\Omega = 0$ and $x_{1,ss}(t)$ is invertible. This bias is quantified by the following proposition.

Proposition 3. *For every* $(x_1,x_2) \in \mathbb{X}$ *such that* x_1 *is invertible, every* $\Omega \in \mathbb{S}_{n_\theta}$, *and every* $\theta_0 \in \mathbb{R}^{n_\theta}$

$$\begin{aligned} &\left|(x_1+\Omega)^{-1}(x_2+\Omega\theta_0) - x_1^{-1}x_2\right| \\ &\quad \leq |x_1^{-1}|\left(|x_2\,\|\,x_1^{-1}| + |\theta_0|\right)|\Omega| + |\theta_0\,\|\,x_1^{-1}|^2|\Omega|^2. \end{aligned} \tag{16}$$

Thus, in particular, $\lim_{\Omega\to 0}(x_1+\Omega)^{-1}(x_2+\Omega\theta_0) = x_1^{-1}x_2$.

Proof. Since $x_1, \Omega \in \mathbb{S}_{n_\theta}$ and x_1 is invertible, then $x_1 + \Omega \in \mathbb{S}_{n_\theta}$ is invertible. By adding and subtracting $x_1^{-1}(x_2+\Omega\theta_0)$ and using $(a^{-1}-b^{-1}) = b^{-1}(b-a)a^{-1}$ we obtain

$$\begin{aligned} &\left|(x_1+\Omega)^{-1}(x_2+\Omega\theta_0) - x_1^{-1}x_2\right| \\ &\quad \leq |x_2+\Omega\theta_0|\cdot|(x_1+\Omega)^{-1}|\cdot|x_1^{-1}|\cdot|\Omega| + |x_1^{-1}|\cdot|\theta_0|\cdot|\Omega|, \end{aligned}$$

which implies (16) since $|(x_1+\Omega)^{-1}| \leq |x_1^{-1}|$ being $x_1, \Omega, x_1 + \Omega \in \mathbb{S}_{n_\theta}$. Indeed, for every invertible $M \in \mathbb{S}_{n_\theta}$, $|M^{-1}| = 1/\underline{\lambda}(M)$. Then, by taking a unitary v such that $(x_1+\Omega)v = \underline{\lambda}(x_1+\Omega)v$, one obtains $1/|(x_1+\Omega)^{-1}| = \underline{\lambda}(x_1+\Omega) = v^\top(x_1+\Omega)v \geq v^\top x_1 v \geq \underline{\lambda}(x_1) = 1/|x_1^{-1}|$, implying $|(x_1+\Omega)^{-1}| \leq |x_1^{-1}|$. □

Using a non-zero $\Omega(t)$ in spite of this bias is nevertheless a well-established practice in system identification [58] as, in general, it makes numerically ill-posed problems treatable and it may be used to penalize deviations from a given value $\theta_0(t)$ of interest, as typical for instance of *continual learning*. Moreover, as discussed later in Section 6.1, it also ensures stability and convergence (although necessarily approximate) in absence of sufficient excitation.

4.4. Design of γ and state-to-output stability

In view of Propositions 1 and 2, one could be tempted to take γ in (7) as

$$\gamma(x_1,x_2,t) = \big(x_1+\Omega(t)\big)^\dagger\big(x_2+\Omega(t)\theta_0(t)\big). \tag{17}$$

Indeed, when computed along the steady-state trajectory x_{ss}, this would give

$$\begin{aligned} \theta_{ss}(t) &:= \gamma(x_{1,ss}(t), x_{2,ss}(t), t) \\ &= \big(x_{1,ss}(t)+\Omega(t)\big)^\dagger\big(x_{2,ss}(t)+\Omega(t)\theta_0(t)\big) \in \Theta^\circ(t). \end{aligned}$$

Nevertheless, the choice (17) does not allow us to establish a relation between the difference $\theta(t) - \theta_{ss}(t)$ and $x(t) - x_{ss}(t)$ since, in general, the mapping $x_1 + \Omega \mapsto (x_1+\Omega)^\dagger$ is not continuous. This, in turn, implies that a uniform bound of the form (6) cannot be obtained in general with the choice (17). By following [8,53], we solve this problem by designing γ as a robustified version of (17).

With $k_1 := \sup_{t\in\mathbb{N}}|\Omega(t)|$, $k_2 := \sup_{t\in\mathbb{N}}|\Omega(t)\theta_0(t)|$, and $\varepsilon > 0$ an arbitrary constant, we define the set

$$\begin{aligned} \Gamma_\varepsilon := \Big\{(x_1,x_2) \in \mathbb{X} : |x_1| \leq \bar\phi^2 + k_1 + 1,\ \underline{\lambda}(x_1) \geq \varepsilon/2, \\ |x_2| \leq \bar\phi\bar y + k_2 + 1\Big\}. \end{aligned}$$

Then, we define the map

$$g_\theta : \mathbb{R}^{n_\theta\times n_\theta} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_\theta}, \quad (x_1,x_2) \mapsto x_1{}^\dagger x_2,$$

we pick an arbitrary $h \in \mathcal{S}(\Gamma_\varepsilon, g_\theta)$, and we finally define $\gamma : \mathbb{X}\times\mathbb{N} \to \mathbb{R}^{n_\theta}$ as

$$\gamma(x_1,x_2,t) := h\big(x_1+\Omega(t), x_2+\Omega(t)\theta_0(t)\big). \tag{18}$$

With $\bar h := |\operatorname{ran} h|$, let

$$\rho_\theta := \max\left\{\frac{4(\bar\phi\bar y + k_2 + 1) + 2\varepsilon}{\varepsilon^2}, \frac{4\bar h}{\min\{2,\varepsilon\}}\right\}. \tag{19}$$

Then, γ satisfies the following *state-to-output stability* property (the proof of Lemma 4 is given in the Appendix).

Lemma 4. *Suppose that Assumption 1 holds. Then, for all* $t \in \mathbb{N}$ *such that* $\underline{\lambda}(x_{1,ss}(t)+\Omega(t)) \geq \varepsilon$, $|\gamma(x_1(t),x_2(t),t) - \gamma(x_{1,ss}(t),x_{2,ss}(t),t)| \leq \rho_\theta|x(t)-x_{ss}(t)|$.

Remark 3 (*Implementation of* γ)**.** The function γ is fixed by (18) once h is chosen. A possible choice of h satisfying the properties above is to pick $h(x_1,x_2)$ as the vector with ith component $h(x_1,x_2)_i := \mathrm{sat}_{\bar\theta}(u_i)$ with u_i the ith component of $x_1{}^\dagger x_2$, $\bar\theta := 2(\bar\phi\bar y + k_2 + 1)/\varepsilon$, and sat defined as in Remark 1. Indeed, if $(x_1,x_2) \in \Gamma_\varepsilon$, then $|u_i| \leq |x_1^{-1}x_2| \leq \bar\theta$ and thus $h(x_1,x_2) = g_\theta(x_1,x_2)$.

5. Main result

In the previous section, we constructed a class of recursive estimation laws of the form (7), under the assumption (Assumption 1) that $\phi^\star$ and $y^\star$ in (1) are bounded with known bounds. The possible choices for the degrees of freedom $\mu, \lambda, \sigma, \delta, \gamma$ in (7) are summarized below:

- I1. $\mu(t) \in [0,1)$ and $\lambda(t) \geq 0$ satisfy (8).
- I2. $\sigma \in \mathcal{S}(B_1,g_1)$ and $\delta \in \mathcal{S}(B_2,g_2)$, in which the constants $c_1, c_2 \in \mathbb{R}_{>0}\cup\{\infty\}$ defining the sets B_1 and B_2 are arbitrary provided that (9) hold (Section 4.2). See also Remark 1.
- I3. γ is defined in (18), with $h \in \mathcal{S}(\Gamma_\varepsilon, g_\theta)$, $\Omega : \mathbb{N} \to \mathbb{S}_{n_\theta}$ and $\theta_0 : \mathbb{N} \to \mathbb{R}^{n_\theta}$ arbitrary bounded signals, and $\varepsilon > 0$ arbitrary (Section 4.4). See also Remark 3.

We associate with the quantities μ, λ, Ω, and ε the following *excitation condition* on the unperturbed regressor $\phi^\star$.

Definition 1. With $t^\star \in \mathbb{N}$, $\phi^\star$ is said to be $t^\star$-exciting with respect to $(\mu,\lambda,\Omega,\varepsilon)$ if, for all $t \geq t^\star$,

$$\Omega(t) + \sum_{s=0}^{t-1}\left(\prod_{\tau=s+1}^{t-1}\mu(\tau)\right)\lambda(s)\phi^\star(s)\phi^\star(s)^\top \geq \varepsilon I. \tag{20}$$

Remarks on Definition 1 and details on how it relates to typical excitation properties, such as PE, are postponed to Section 6, while here we limit to notice that (20) only involves the unperturbed regressor $\phi^\star$, despite the presence of the disturbance d_ϕ in the measurements. Define

$$\theta_{ss}(t) := \gamma(x_{1,ss}(t), x_{2,ss}(t), t). \tag{21}$$

Then, the following theorem summarizes the main properties of the considered estimator (7).

Theorem 1. *Suppose that Assumption 1 holds, and consider System (7) with $\mu, \lambda, \sigma, \delta, \gamma$ chosen according to Items I1–I3. Let $x_{ss} = (x_{1,ss}, x_{2,ss})$ and θ_{ss} be defined, respectively, as in (10) and (21). Then, for every solution pair (x, d) of (7), the following hold:*

C1. *For all $t \in \operatorname{dom}(x, d)$, (12) holds. If, in addition, d is bounded and $\operatorname{dom}(x, d) = \mathbb{N}$, then also (13) holds.*

C2. *If $\phi^\star$ is $t^\star$-exciting with respect to $(\mu, \lambda, \Omega, \varepsilon)$, then*

$$|\theta(t) - \theta_{ss}(t)| \le \rho_\theta \beta\big(|x(0) - x_{ss}(0)|, t\big) + \rho_\theta \kappa\big(|d|_{[0,t-1]}\big) \tag{22}$$

holds for all $t \in \operatorname{dom}(x, d)_{\ge t^\star}$, with β and κ defined in (11), and ρ_θ in (19). In particular, if d is bounded and $\operatorname{dom}(x, d) = \mathbb{N}$, then

$$\limsup_{t\to\infty} |\theta(t) - \theta_{ss}(t)| \le \rho_\theta \kappa \left(\limsup_{t\to\infty} |d(t)| \right). \tag{23}$$

C3. *If, in addition to C2, $\lim_{t\to\infty} \Omega(t) = 0$, then*

$$\limsup_{t\to\infty} |\theta(t) - \theta^\star| \le \rho_\theta \kappa \left(\limsup_{t\to\infty} |d(t)| \right).$$

Proof. Claim C1 is Proposition 1.

Next, notice that the left-hand side of (20) equals $\Omega(t) + x_{1,ss}(t)$. Hence, if $\phi^\star$ is $t^\star$-exciting with respect to $(\mu, \lambda, \Omega, \varepsilon)$, then $\underline{\lambda}(x_{1,ss}(t) + \Omega(t)) \ge \varepsilon$ for all $t \ge t^\star$. Thus, Claim C2 follows from Lemma 4 and Proposition 1.

Regarding Claim C3, first notice that, since $\mu(t), \lambda(t) \ge 0$ for all $t \in \mathbb{N}$, then from (8) and (10) we get $|x_{1,ss}(t)| \le \bar{\phi}^2$ and $|x_{2,ss}(t)| \le \bar{\phi}\bar{y}$ for all $t \in \mathbb{N}$. Thus, if $\phi^\star$ is $t^\star$-exciting with respect to $(\mu, \lambda, \Omega, \varepsilon)$, then $(x_{1,ss}(t) + \Omega(t), x_{2,ss}(t) + \Omega(t)\theta_0(t)) \in \Gamma_\varepsilon$ for all $t \ge t^\star$ and, by definition of γ,

$$\theta_{ss}(t) = \big(x_{1,ss}(t) + \Omega(t)\big)^{-1}\big(x_{2,ss}(t) + \Omega(t)\theta_0(t)\big), \ \forall t \ge t^\star. \tag{24}$$

Next, notice that, if $\Omega(t) \to 0$ and $\phi^\star$ is $t^\star$-exciting with respect to $(\mu, \lambda, \Omega, \varepsilon)$, then there exists $\bar{t} \in \mathbb{N}$ such that $\underline{\lambda}(x_{1,ss}(t)) \ge \varepsilon/2$ for all $t \ge \bar{t}$. Indeed, $\Omega(t) + x_{1,ss}(t) \ge \varepsilon I$ implies $\varepsilon \le v^\top\big(\Omega(t) + x_{1,ss}(t)\big)v \le |\Omega(t)| + v^\top x_{1,ss}(t) v$ for all unitary v. Taking v as a unitary eigenvector of $x_{1,ss}(t)$ associated with $\underline{\lambda}(x_{1,ss}(t))$ yields $\underline{\lambda}(x_{1,ss}(t)) \ge \varepsilon - |\Omega(t)|$. As $\Omega \to 0$, we thus obtain the existence of such $\bar{t}$.

The previous analysis, in particular, implies that $x_{1,ss}(t)$ is nonsingular for all $t \ge \bar{t}$ and, since $x_{1,ss}(t) \in \mathbb{S}_{n_\theta}$, that $|x_{1,ss}(t)^{-1}| = \underline{\lambda}(x_{1,ss}(t))^{-1} \le 2/\varepsilon$. Moreover, in view of (14), it also implies that $\theta^\star = x_{1,ss}(t)^{-1} x_{2,ss}(t)$ for all $t \ge \bar{t}$. Then, from (24) and Proposition 3, we obtain

$$\begin{aligned} |\theta(t) - \theta^\star| &\le |\theta(t) - \theta_{ss}(t)| + |\theta_{ss}(t) - \theta^\star| \\ &\le |\theta(t) - \theta_{ss}(t)| + \left(\frac{4\bar{\phi}\bar{y}}{\varepsilon^2} + \frac{2|\theta_0(t)|}{\varepsilon} \right) |\Omega(t)| \\ &\quad + \frac{4|\theta_0(t)|}{\varepsilon^2} |\Omega(t)|^2 \end{aligned}$$

for all $t \ge \bar{t}$. As θ_0 is bounded and $\Omega \to 0$, then Claim C3 follows from (23). □

6. Excitation and convergence rate

Condition (20) in Definition 1 is a *joint* property linking the unperturbed regressor $\phi^\star$ and the degrees of freedom $\varepsilon, \Omega, \mu, \lambda$. As these quantities also affect the estimate's convergence properties, such as bias (Proposition 3) and convergence rate (Proposition 1 and (11)), then it is licit to expect that the excitation level of $\phi^\star$ and the convergence properties of (7) are related to one another. In this section, we further discuss Definition 1 and provide insights about the relation between excitation and convergence rate. In particular, we find that different choices of μ and λ are associated with different excitation requirements on $\phi^\star$ and different convergence rates. However, notably, robustness with respect to disturbances is not directly touched by the particular choice of μ and λ, since the function κ in (11) does not depend on μ or λ.

6.1. On the role of Ω

The regularization matrix Ω can be used to guarantee that *every* $\phi^\star$ is $t^\star$-exciting with respect to $(\mu, \lambda, \Omega, \varepsilon)$ with the same $t^\star$. Indeed, if for some $t^\star \in \mathbb{N}$, $\Omega(t) \ge \varepsilon I$ for all $t \ge t^\star$, then every $\phi^\star$ is automatically $t^\star$-exciting with respect to $(\mu, \lambda, \Omega, \varepsilon)$. This can be seen by noticing that, in view of (10), the excitation condition (20) can be written as $\Omega(t) + x_{1,ss}(t) \ge \varepsilon I$, and $x_{1,ss}(t)$ is symmetric and positive semidefinite for every $\phi^\star$ and $t \in \mathbb{N}$. Thus, for every $v \in \mathbb{R}^{n_\theta}$ and $t \ge t^\star$, $v^\top(\Omega(t) + x_{1,ss}(t))v \ge v^\top \Omega(t) v \ge \varepsilon|v|^2$.

As (20) plays a role in the construction of γ (Lemma 4), and in particular in guaranteeing that the quantity $(x_{1,ss}(t) + \Omega(t), x_{2,ss}(t) + \Omega(t)\theta_0(t))$ is inside the set Γ_ε, then this property of Ω is consistent with the usual interpretation of regularization that it makes the problem numerically well-posed. However, we remark once again that a non-zero $\Omega(t)$ induces a bias in the parameter estimate (Proposition 3 and Theorem 1), so as it represents a trade-off between robustness with respect to excitation properties of $\phi^\star$ and bias. Finally, we underline that, although they fall beyond the scope of this paper, feedback techniques to adapt $\Omega(t)$ from data are a viable option. Indeed, Ω is a free parameter and, for instance, in absence of disturbances the following is a feasible a choice

$$\begin{aligned} &\Omega(t) = \varepsilon I && \text{if } \underline{\lambda}(x_1(t)) < \varepsilon \\ &\Omega(t) \in \{0, \varepsilon I\} && \text{if } \underline{\lambda}(x_1(t)) = \varepsilon \\ &\Omega(t) = 0 && \text{otherwise} \end{aligned}$$

aimed at taking $\Omega(t)$ non-zero only when needed.

6.2. PE and uniform exponential convergence

One says that $\phi^\star$ is PE if there exist $T \in \mathbb{N}_{\ge 1}$, $t^\star \in \mathbb{N}$, and $k > 0$ such that [54]

$$\sum_{s=t-T}^{t-1} \phi^\star(s)\phi^\star(s)^\top \ge kI, \qquad \forall t \ge \max\{t^\star, T\}. \tag{25}$$

It is a well-known general property of adaptive systems [33,54], that persistence of excitation is associated with exponential convergence, and System (7) makes no exception. Specifically, take $\mu(t)$ constant (we shall refer to this constant as $\mu \in [0, 1)$ to avoid introducing further notations), and $\lambda = 1 - \mu$. Then, the equations of x_1 and x_2 in (7) become moving averages, and (8) holds, since

$$\begin{aligned} \sum_{s=0}^{t-1} \left(\prod_{\tau=s+1}^{t-1} \mu(\tau) \right) \lambda(s) &= (1-\mu) \sum_{s=0}^{t-1} \mu^{t-s-1} \\ &= (1-\mu) \sum_{\tau=0}^{t-1} \mu^\tau \le (1-\mu) \sum_{\tau=0}^{\infty} \mu^\tau = 1. \end{aligned}$$

Moreover, in view of (11), the class-$\mathcal{KL}$ function β governing the convergence rate (see Proposition 1 and Theorem 1) becomes

$$\beta(s, t) = s\,\mu^t,$$

which is exponentially decaying with rate μ.

Regarding the excitation condition of Definition 1, by letting $\Omega(t) = 0$, Condition (20) becomes

$$\sum_{s=0}^{t-1} \mu^{t-s-1}\phi^\star(s)\phi^\star(s)^\top \geq \frac{\varepsilon}{1-\mu} I, \qquad \forall t \geq t^\star, \tag{26}$$

which is equivalent to PE in the following sense.

Lemma 5. *Suppose that Assumption 1 holds. If $\phi^\star$ satisfies (26) for some $t^\star \in \mathbb{N}$ and $\varepsilon > 0$, then it satisfies (25) with the same $t^\star$, with $k \leq \varepsilon/(2-2\mu)$, and $T \geq \max\{1, \log(\varepsilon/(2\bar\phi^2))/\log\mu\}$. Conversely, if $\phi^\star$ satisfies (25) for some $k > 0$, $T \in \mathbb{N}_{\geq 1}$, and $t^\star \in \mathbb{N}_{\geq T}$, then it satisfies (26) with the same $t^\star$ and $\varepsilon \leq k\mu^{T-1}(1-\mu)$.*

The proof of Lemma 5 is omitted for reason of space. Finally, we remark that this choice of μ and λ is associated with the usual exponentially weighted least-squares cost functional. In particular, (15) reads

$$J(t,\theta) = (1-\mu)\sum_{s=0}^{t-1}\mu^{t-s-1}\big(y^\star(s) - \phi^\star(s)^\top\theta\big)^2 + (\theta - \theta_0(t))^\top \Omega(t)(\theta - \theta_0(t)).$$

6.3. Non-uniform convergence without PE

A different choice of μ and λ may be associated with a weaker excitation condition on $\phi^\star$, but with a weaker convergence rate. We show this with the following choice

$$\mu(t) = \frac{t+a}{t+a+1}, \qquad \lambda(t) = \frac{1}{t+a+1}, \tag{27}$$

where $a \in \mathbb{N}$ is arbitrary. With this choice, we have

$$\sum_{s=0}^{t-1}\left(\prod_{\tau=s+1}^{t-1}\mu(\tau)\right)\lambda(s) = \sum_{s=0}^{t-1}\frac{1}{t+a} = \frac{t}{t+a} \leq 1,$$

so as (8) holds. With $\Omega = 0$, Condition (20) now becomes

$$\frac{1}{t+a}\sum_{s=0}^{t-1}\phi^\star(s)\phi^\star(s)^\top \geq \varepsilon I, \qquad \forall t \geq t^\star. \tag{28}$$

It turns out that (28) is *strictly weaker* than PE, as established by the following proposition.

Proposition 4. *If $\phi^\star$ is PE, i.e. it satisfies (25) for some $T \in \mathbb{N}_{\geq 1}$, $t^\star \in \mathbb{N}$, and $k > 0$, then it also satisfies (28) for some $t^\star \in \mathbb{N}$ and with $\varepsilon = k/(2T + a)$. The converse implication, instead, does not hold in general. Namely, there exist bounded signals $\phi^\star$ satisfying (28) that are not PE, i.e., for which (25) does not hold for any choice of $T \in \mathbb{N}_{\geq 1}$, $t^\star \in \mathbb{N}$ and $k > 0$.*

Proposition 4 is proved at the end of the section, we first discuss some relevant consequences. First, we underline that, compared to the previous case, with this choice for μ the convergence rate of x towards x_{ss} is hyperbolic instead of exponential. Indeed, the function β in (11) now reads as

$$\beta(s,t) = s\frac{a}{t+a}.$$

In turn, this implies that convergence is not uniform in the initial time. This is consistent with well-known facts about necessity of PE for uniform convergence [54]. Nevertheless, we underline that the ISS condition of Proposition 1 is unchanged. In particular, the asymptotic gain relating the disturbance d and the differences $x(t) - x_{ss}(t)$ and $\theta(t) - \theta_{ss}(t)$ (Theorem 1) is not touched by the choice of μ. Therefore, PE is not only unnecessary for convergence, but also for robustness, with the only difference that the result of Proposition 1, as well as the limits in the claims of Theorem 1, are no more uniform in the initial time. Hence, to summarize, PE provides uniformity of convergence (at an exponential rate) in absence of disturbances and uniform ISS and asymptotic gain properties in presence of disturbances. Without PE, uniformity in the initial time is lost, but ISS and asymptotic gain properties are maintained.

Finally, we notice that the choice (27) is associated with the following *unweighted* least-squares functional

$$J(t,\theta) = \frac{1}{a+t}\sum_{s=0}^{t-1}\big(y^\star(s) - \phi^\star(s)^\top\theta\big)^2 + (\theta - \theta_0(t))^\top \Omega(t)(\theta - \theta_0(t)).$$

Proof (*Proposition 4*)**.** Assume $\phi^\star$ is PE. Without loss of generality, we prove the first implication of the proposition with $t^\star = 0$ in (25). Pick $t \geq T$, and let $n := \max\{m \in \mathbb{N} : t \geq mT\}$. Then, $n \geq 1$, $t \geq nT$, and

$$\begin{aligned}\frac{1}{t+a}\sum_{s=0}^{t-1}\phi^\star(s)\phi^\star(s)^\top &\geq \frac{1}{t+a}\sum_{s=0}^{nT-1}\phi^\star(s)\phi^\star(s)^\top \\ &= \frac{1}{t+a}\sum_{h=0}^{n-1}\sum_{s=hT}^{(h+1)T-1}\phi^\star(s)\phi^\star(s)^\top \geq \frac{1}{t+a}\sum_{h=0}^{n-1}kI \\ &= \frac{nk}{a+nT+(t-nT)}I = \frac{k}{T + \dfrac{a+t-nT}{n}}I \geq \frac{k}{2T+a}I,\end{aligned}$$

where, in the last inequality, we used the fact that $t - nT \leq T$ and $n \geq 1$, which imply $(a+t-nT)/n \leq a+T$.

The second part of the proposition is proved with a counterexample. Namely, we construct a bounded $\phi^\star$ that is not PE but satisfies (28). Consider a partition of $\mathbb{N}$ of the form $\mathbb{N} = \cup_{n=1}^{\infty} I_n$ with $I_n = \left\{\sum_{k=1}^{n-1}2^k, \ \dots, \ \sum_{k=1}^{n-1}2^k + 2^n - 1\right\}$. Thus, in particular, $I_1 = \{0,1\}$, $I_2 = \{2,3,4,5\}$, $I_3 = \{6,\dots,13\}$ etc., and the nth interval I_n has cardinality 2^n. Let $\phi^\star$ be defined in such a way that, for all $n \in \mathbb{N}$ and all $t \in I_n$,

$$\phi^\star(t) = \begin{cases} 0 & \text{if } t < \min I_n + 2^{n-1} \\ 1 & \text{otherwise.}\end{cases} \tag{29}$$

In particular, $\phi^\star(t)$ equals 0 on the first half of each I_n and 1 in the second half.

Clearly, $\phi^\star$ is not PE. Indeed, for every $T \in \mathbb{N}$ and $t^\star \in \mathbb{N}$, there exist n and $t > \max\{T, t^\star\}$ such that $[t-T, t-1] \subset I_n$ and $t-1 < \min I_n + 2^{n-1}$. In turn, this implies that $\sum_{t-T}^{t-1}\phi^\star(s)\phi^\star(s)^\top = 0$. Hence, (25) does not hold. However, Condition (28) holds with

$$t^\star \geq 3, \qquad \varepsilon = \frac{1}{4}\left(1 - \frac{a+2}{a+t^\star}\right).$$

To see this, pick $t^\star \geq 3$ and $t \geq t^\star$ arbitrarily, and let $m = m(t)$ be the largest integer such that $\tau(m) := 2 + 4 + \cdots + 2^m \leq t$. Since, in view of (29),

$$\sum_{s=0}^{\tau(m)-1}\phi^\star(s) = \sum_{k=1}^{m}\sum_{s\in I_k}\phi^\star(s) = \frac{1}{2}\sum_{k=1}^{m}2^k = \frac{\tau(m)}{2}$$

and $\phi^\star(s)\phi^\star(s)^\top = \phi^\star(s) \geq 0$ for all $s \in \mathbb{N}$, then

$$\begin{aligned}\frac{1}{t+a}\sum_{s=0}^{t-1}\phi^\star(s)\phi^\star(s)^\top &\geq \frac{1}{t+a}\sum_{s=0}^{\tau(m)-1}\phi^\star(s) = \frac{\tau(m)}{2(t+a)} \\ &\geq \frac{1}{4}\frac{t-2}{t+a} \geq \frac{1}{4}\left(1 - \frac{a+2}{a+t^\star}\right) = \varepsilon,\end{aligned}$$

where we used the fact that, by definition of m, $t \leq \tau(m+1) = \tau(m) + 2^{m+1} = 2\tau(m) + 2$. $\square$

7. Conclusions

In this paper, we studied a class of robustified least-squares estimators of the form (7) aimed to extract from the perturbed measurements (2) the parameter $\theta^\star$ relating the unperturbed regressor $\phi^\star$ and regressand $y^\star$ according to (1). We considered disturbances acting on all the variables and, instead of studying the statistical properties of the estimates when the disturbances are random processes, as typical of statistical learning, we approached the estimation problem in control theoretic terms. Specifically, we proved existence and stability of an optimal steady-state defined by the unperturbed variables, and we proved robustness with respect of disturbances in terms of ISS.

While robust stability always holds (Proposition 1), input–output stability of the estimation error holds if the unperturbed regressor carries enough excitation, where how much is "enough" is formalized in Definition 1 and, when $\Omega = 0$, depends on the degrees of freedom μ and λ. In this connection, we showed that some choices of these degrees of freedom require PE of $\phi^\star$ and confer uniform exponential convergence on the produced estimates. Moreover, we also showed that other choices are possible that yield a non-uniform convergence rate but require excitation conditions on $\phi^\star$ that are strictly weaker than PE. Notably, the asymptotic gain property relating the asymptotic estimation error to the size of the disturbances is the same in the two cases. Indeed, in general, all choices of μ and λ satisfying (8) do not affect the function κ in (12). Finally, we showed that taking $\Omega(t)$ positive definite ensures that every $\phi^\star$ is PE, but introduces a bias in the parameter estimate. This is, therefore, a degree of freedom that must be traded off in applications.

Overall, our findings suggest that robustification is a good way to achieve robustness with respect to disturbances, and that robustness is not necessarily related to excitation, which instead mainly affects the convergence rate.

CRediT authorship contribution statement

Michelangelo Bin: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Technical proofs

Proof of Lemma 2. If $c_1 = \infty$, then $B_1 = \mathbb{R}^{n_\theta}$ and, hence, $\sigma = g_1$. Therefore, for every $a, b \in \mathbb{R}^{n_\theta}$, $\sigma(a) - \sigma(b) = aa^\top - bb^\top = (a-b)a^\top + b(a-b)^\top$. With $a = \phi(t) = \phi^\star(t) + d_\phi(t)$ and $b = \phi^\star(t)$, this yields $|\sigma(\phi(t)) - \sigma(\phi^\star(t))| = |d_\phi(t)\phi(t)^\top + \phi^\star(t)d_\phi(t)^\top| \le 2\bar\phi|d_\phi(t)| + |d_\phi(t)|^2 = \rho_1(|d_\phi(t)|)$.

Next, suppose that $c_1 < \infty$. From the previous computations, we deduce that g_1 is Lipschitz on B_1 with Lipschitz constant $2c_1 + 2$. Let $X' := \{\phi \in \mathbb{R}^{n_\theta} : |\phi| \le c_1\}$. Then, $\phi \in \mathbb{R}^{n_\theta} \setminus B_1$, implies $d(\phi, X') \ge 1$. Hence, we apply Lemma 1 with $X = B_1$ to obtain $|\sigma(a) - \sigma(b)| \le \ell_1|a - b| = \rho_1(|a - b|)$ for all $a \in X'$ and $b \in \mathbb{R}^{n_\theta}$, which concludes the proof since, by (9), $\phi^\star(t) \in X'$. □

Proof of Proposition 1. Pick a solution pair (x, d) of (7) and a $t \in \mathrm{dom}(x, d)$. From the first of (7), we obtain

$$x_1(t) = \left(\prod_{s=0}^{t-1}\mu(s)\right)x_1(0) + \sum_{s=0}^{t-1}\left(\prod_{\tau=s+1}^{t-1}\mu(\tau)\right)\lambda(s)\sigma\big(\phi(s)\big)$$

for all $t \in \mathbb{N}$. By Assumption 1, $\phi^\star(t) \in B_1$. Since $\sigma \in \mathcal{S}(B_1, g_1)$, then $\sigma(\phi^\star(t)) = \phi^\star(t)\phi^\star(t)^\top$. Hence, from (10) and by using Lemma 2, Inequality (8), and the fact that $x_{1,\mathrm{ss}}(0) = 0$ by construction, we get

$$\begin{aligned}|x_1(t) - x_{1,\mathrm{ss}}(t)| &\le \left(\prod_{s=0}^{t-1}\mu(s)\right)|x_1(0) - x_{1,\mathrm{ss}}(0)| \\ &\quad + \sum_{s=0}^{t-1}\left(\prod_{\tau=s+1}^{t-1}\mu(\tau)\right)\lambda(s)\big|\sigma\big(\phi(s)\big) - \sigma\big(\phi^\star(s)\big)\big| \\ &\le \beta(|x_1(0) - x_{1,\mathrm{ss}}(0)|, t) + \rho_1\big(|d|_{[0,t-1]}\big),\end{aligned}$$

where β is given in (11). By means of the same arguments, one shows that

$$|x_2(t) - x_{2,\mathrm{ss}}(t)| \le \beta(|x_2(0) - x_{2,\mathrm{ss}}(0)|, t) + \rho_2\big(|d|_{[0,t-1]}\big).$$

By construction, $\max\{\beta(s_1, t), \beta(s_2, t)\} = \beta(\max\{s_1, s_2\}, t)$ for all $s_1, s_2, t \ge 0$. Then, (12) follows by the fact that $|x(t) - x_{\mathrm{ss}}(t)| = \max\{|x_1(t) - x_{1,\mathrm{ss}}(t)|,\ |x_2(t) - x_{2,\mathrm{ss}}(t)|\}$.

Finally, if d is bounded, then $\limsup_{t\to\infty}|d(t)| < \infty$, and Inequality (13) follows from (12) by means of standard ISS arguments (see, e.g., [10]). □

Proof of Proposition 2. Call for brevity $S(t) := \{\theta \in \mathbb{R}^{n_\theta} : \big(x_{1,\mathrm{ss}}(t) + \Omega(t)\big)\theta = x_{2,\mathrm{ss}}(t) + \Omega(t)\theta_0(t)\}$. We have to show that $\Theta^\circ(t) = S(t)$. From (10), we obtain $x_{1,\mathrm{ss}}(t) = \Phi^\top W\Phi$ and $x_{2,\mathrm{ss}}(t) = \Phi^\top WY$, where $Y := \mathrm{col}(y^\star(0), \dots, y^\star(t-1))$, $\Phi := \mathrm{col}(\phi^\star(0)^\top, \dots, \phi^\star(t-1)^\top)$, and $W := \mathrm{diag}(\prod_{\tau=1}^{t-1}\mu(\tau)\lambda(0), \dots, \lambda(t-1))$. This implies

$$S(t) = \left\{\theta \in \mathbb{R}^{n_\theta} : \big(\Phi^\top W\Phi + \Omega(t)\big)\theta = \Phi^\top WY + \Omega(t)\theta_0(t)\right\}.$$

Moreover, we can write $J(t, \theta) = (Y - \Phi\theta)^\top W(Y - \Phi\theta) + (\theta - \theta_0(t))^\top\Omega(t)(\theta - \theta_0(t))$. Then, by substitution, one can show that, for each $\bar\theta \in S(t)$,

$$J(t, \theta) = J(t, \bar\theta) + (\theta - \bar\theta)^\top(\Phi^\top W\Phi + \Omega(t))(\theta - \bar\theta) \tag{A.1}$$

holds for all $\theta \in \mathbb{R}^{n_\theta}$. Then, since $\Phi^\top W\Phi + \Omega(t)$ is positive semidefinite, $J(t, \theta) \ge J(t, \bar\theta)$. For the arbitrariness of $\theta \in \mathbb{R}^{n_\theta}$ and $\bar\theta \in S(t)$, we thus conclude that $S(t) \subset \Theta^\circ(t)$.

To prove the converse, pick $\theta \in \Theta^\circ(t)$ and $\bar\theta \in S(t)$ arbitrarily. In view of (A.1), we must have $(\theta - \bar\theta)^\top(\Phi^\top W\Phi + \Omega(t))(\theta - \bar\theta) = 0$. As $\Phi^\top W\Phi + \Omega(t)$ is symmetric and positive semidefinite, this implies $(\Phi^\top W\Phi + \Omega(t))(\theta - \bar\theta) = 0$. Hence, $(\Phi^\top W\Phi + \Omega(t))\theta = (\Phi^\top W\Phi + \Omega(t))\bar\theta = \Phi^\top WY + \Omega(t)\theta_0(t)$, which shows that $\theta \in S(t)$. □

Proof of Lemma 4. Define the set $\Gamma' := \left\{(x_1, x_2) \in \mathbb{X} : |x_1| \le \bar\phi^2 + k_1,\ \underline{\lambda}(x_1) \ge \varepsilon,\ |x_2| \le \bar\phi\bar y + k_2\right\}$. Then, the following lemma, proved below, holds true.

Lemma 6. *$\Gamma' \subset \Gamma_\varepsilon$ and $x \in (\mathbb{R}^{n_\theta \times n_\theta} \times \mathbb{R}^{n_\theta}) \setminus \Gamma_\varepsilon$ implies $d(x, \Gamma') \ge \min\{1, \varepsilon/2\}$.*

Since $\mu(t), \lambda(t) \ge 0$, from (8) and (10) we get $|x_{1,\mathrm{ss}}(t)| \le \bar\phi^2$ and $|x_{2,\mathrm{ss}}(t)| \le \bar\phi\bar y$. Thus, if $\underline{\lambda}(x_{1,\mathrm{ss}}(t) + \Omega(t)) \ge \varepsilon$, then $(x_{1,\mathrm{ss}}(t) + \Omega(t), x_{2,\mathrm{ss}}(t) + \Omega(t)\theta_0(t)) \in \Gamma'$. Moreover, for arbitrary $(a_1, b_1), (a_2, b_2) \in \Gamma_\varepsilon$, we have

$$\begin{aligned}|g_\theta(a_1, b_1) - g_\theta(a_2, b_2)| &= |{a_1}^\dagger b_1 - {a_2}^\dagger b_2 \pm {a_2}^\dagger b_1| \\ &\le |{a_1}^\dagger - {a_2}^\dagger||b_1| + |{a_2}^\dagger||b_1 - b_2|.\end{aligned}$$

Since for both $i = 1, 2$, $(a_i, b_i) \in \Gamma_\varepsilon$, then $|b_i| \le \bar\phi\bar y + k_2 + 1$, $a_i \in \mathbb{S}_{n_\theta}$, and $\underline{\lambda}(a_i) \ge \varepsilon/2$. Hence, a_i is invertible, ${a_i}^\dagger = a_i^{-1}$, and $|{a_i}^\dagger| = |a_i^{-1}| = \underline{\lambda}(a_i)^{-1} \le 2\varepsilon^{-1}$. Thus, ${a_1}^\dagger - {a_2}^\dagger = a_1^{-1} - a_2^{-1} =$

${a_2}^{-1}(a_2 - a_1){a_1}^{-1}$, which yields $|{a_1}^{\dagger} - {a_2}^{\dagger}| \leq |{a_1}^{-1}| \cdot |{a_2}^{-1}| \cdot |a_1 - a_2| \leq (4/\varepsilon^2)|a_1 - a_2|$ and

$$\begin{aligned}
&|g_\theta(a_1, b_1) - g_\theta(a_2, b_2)| \\
&\quad \leq \frac{4(\bar{\phi}\bar{y} + k_2 + 1)}{\varepsilon^2}|a_1 - a_2| + \frac{2}{\varepsilon}|b_1 - b_2| \\
&\quad \leq \frac{4(\bar{\phi}\bar{y} + k_2 + 1) + 2\varepsilon}{\varepsilon^2}|(a_1, b_1) - (a_2, b_2)|,
\end{aligned}$$

where we used $|(a_1, b_1) - (a_2, b_2)| = \max\{|a_1 - a_2|, |b_1 - b_2|\}$. Namely, g_θ is Lipschitz on Γ_ε.

Since $h \in \mathcal{S}(\Gamma_\varepsilon, g_\theta)$ and, as proved earlier, $(x_{1,\mathrm{ss}}(t) + \Omega(t), x_{2,\mathrm{ss}}(t) + \Omega(t)\theta_0(t)) \in \Gamma'$, there follows from Lemmas 6 and 1 that

$$\begin{aligned}
&|\gamma(x_{1,\mathrm{ss}}(t), x_{2,\mathrm{ss}}(t), t) - \gamma(x_1(t), x_2(t), t)| \\
&\quad = |h(x_{1,\mathrm{ss}}(t) + \Omega(t), x_{2,\mathrm{ss}}(t) + \Omega(t)\theta_0(t)) \\
&\qquad\quad - h(x_1(t) + \Omega(t), x_2(t) + \Omega(t)\theta_0(t))| \\
&\quad \leq \rho_\theta |x(t) - x_{\mathrm{ss}}(t)|,
\end{aligned}$$

with ρ_θ given by (19), which is the claim. □

Proof of Lemma 6. That $\Gamma' \subset \Gamma_\varepsilon$ is obvious. If $(a, b) \notin \Gamma_\varepsilon$, then *(i)* $|a| > \bar{\phi}^2 + k_1 + 1$, or *(ii)* $|b| > \bar{\phi}\bar{y} + k_2 + 1$, or *(iii)* $\underline{\lambda}(a) < \varepsilon/2$. In case *(i)* or *(ii)* hold, for each $(a', b') \in \Gamma'$ we have $|(a, b) - (a', b')| = \max\{|a - a'|, |b - b'|\} \geq \max\{\big||a| - |a'|\big|, \big||b| - |b'|\big|\} > 1$. Thus, $d((a, b), \Gamma') \geq 1$.

Consider now Case *(iii)*. Let v be a unitary eigenvector of a corresponding to $\underline{\lambda}(a)$. Then, for all $(a', b') \in \Gamma'$

$$|a - a'| = \sup_{|z|=1} |(a' - a)z| \geq |(a' - a)v| \geq \big||a'v| - \underline{\lambda}(a)\big|. \tag{A.2}$$

Since $\underline{\lambda}(a') \geq \varepsilon$, a' is invertible. Moreover, since $a' \in \mathbb{S}_{n_\theta}$, $|(a')^{-1}| = 1/\underline{\lambda}(a')$, which implies $\underline{\lambda}(a') = \underline{\lambda}(a')|v| = |(a')^{-1}a'v|/|(a')^{-1}| \leq |a'v|$. Thus, $|a'v| \geq \underline{\lambda}(a') \geq \varepsilon > 2\underline{\lambda}(a)$, and (A.2) yields $|a - a'| \geq |a'v| - \underline{\lambda}(a) \geq \varepsilon - \varepsilon/2 = \varepsilon/2$. Hence, in all cases the claim holds. □

References

[1] P. Van Overschee, B. De Moor, Subspace Identification for Linear Systems, Springer, 1996.

[2] L. Ljung, System Identification: Theory for the User, Pearson Education, 1998.

[3] T. Söderström, Errors-in-variables methods in system identification, Automatica 43 (2007) 939–958.

[4] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed., Springer, 2009.

[5] M. Bin, L. Marconi, A.R. Teel, Adaptive output regulation for linear systems via discrete-time identifiers, Automatica 105 (2019) 422–432.

[6] M. Bin, L. Marconi, "Class-Type" identification-based internal models in multivariable nonlinear output regulation, IEEE Trans. Automat. Control 65 (10) (2020) 4369–4376.

[7] M. Bin, P. Bernard, L. Marconi, Approximate nonlinear regulation via identification-based adaptive internal models, IEEE Trans. Automat. Control 66 (8) (2021) 3534–3549.

[8] M. Bin, L. Marconi, Model identification and adaptive state observation for a class of nonlinear systems, IEEE Trans. Automat. Control 66 (12) (2021) 5621–5636.

[9] E. Sontag, Smooth stabilization implies coprime factorization, IEEE Trans. Automat. Control 34 (4) (1989) 435–443.

[10] E. Sontag, Y. Wang, New characterizations of input-to-state stability, IEEE Trans. Automat. Control 41 (9) (1996) 1283–1294.

[11] Z.P. Jiang, A.R. Teel, L. Praly, Small-gain theorem for ISS systems and applications, Math. Control Signals Systems 7 (2) (1994) 95–120.

[12] M. Milanese, C. Novara, Set Membership identification of nonlinear systems, Automatica 40 (6) (2004) 957–975.

[13] A. Vicino, G. Zappa, Sequential approximation of feasible parameter sets for identification with set membership uncertainty, IEEE Trans. Automat. Control 41 (6) (1996) 774–785.

[14] L. Ljung, T. Söderström, Theory and Practice of Recursive Identification, MIT Press, 1983.

[15] H. Whitaker, An Adaptive System for Control of the Dynamics Performances of Aircraft and Spacecraft, Inst. Aeronautical Sciences, 1959, Paper no. 59–100.

[16] I.M. Mareels, B.D. Anderson, R.R. Bitmead, M. Bodson, S.S. Sastry, Revisiting the MIT rule for adaptive control, in: Adaptive Systems in Control and Signal Processing 1986, Elsevier, 1987, pp. 161–166.

[17] D.D. Donalson, C.T. Leondes, A model referenced parameter tracking technique for adaptive control systems II-stability analysis by the second method of Lyapunov, IEEE Trans. Appl. Ind. 82 (68) (1963) 252–262.

[18] P. Parks, Liapunov redesign of model reference adaptive control systems, IEEE Trans. Automat. Control 11 (3) (1966) 362–367.

[19] P.H. Phillipson, Design methods for model-reference adaptive systems, Proc. Inst. Mech. Eng. 183 (1) (1968) 695–706.

[20] K. Narendra, P. Kudva, Identification and adaptation using Lyaponov's direct method, in: IEEE Conf. Decision Control, 1973, pp. 521–527.

[21] G. Luders, K. Narendra, A new canonical form for an adaptive observer, IEEE Trans. Automat. Control 19 (2) (1974) 117–119.

[22] G. Kreisselmeier, Adaptive observers with exponential rate of convergence, IEEE Trans. Automat. Control 22 (1) (1977) 2–8.

[23] B. Egardt, Stability of Adaptive Controllers, Vol. 20, Springer, 1979.

[24] G. Goodwin, P. Ramadge, P. Caines, Discrete-time multivariable adaptive control, IEEE Trans. Automat. Control 25 (3) (1980) 449–456.

[25] G. Bastin, M. Gevers, Stable adaptive observers for nonlinear time-varying systems, IEEE Trans. Automat. Control 33 (7) (1988) 650–658.

[26] D. Taylor, P. Kokotović, R. Marino, I. Kannellakopoulos, Adaptive regulation of nonlinear systems with unmodeled dynamics, IEEE Trans. Automat. Control 34 (4) (1989) 405–412.

[27] R. Marino, Adaptive observers for single output nonlinear systems, IEEE Trans. Automat. Control 35 (9) (1990) 1054–1058.

[28] I. Kanellakopoulos, P. Kokotovic, A. Morse, Systematic design of adaptive controllers for feedback linearizable systems, IEEE Trans. Automat. Control 36 (11) (1991) 1241–1253.

[29] Z.-P. Jiang, L. Praly, Iterative designs of adaptive controllers for systems with nonlinear integrators, in: 30th IEEE Conf. Decision and Control, Vol. 3, 1991, pp. 2482–2487.

[30] M. Krstić, I. Kanellakopoulos, P.V. Kokotović, Adaptive nonlinear control without overparametrization, Systems Control Lett. 19 (3) (1992) 177–185.

[31] M. Krstić, P.V. Kokotović, I. Kanellakopoulos, Nonlinear and Adaptive Control Design, John Wiley & Sons, 1995.

[32] P. Kokotović, M. Arcak, Constructive nonlinear control: a historical perspective, Automatica 37 (5) (2001) 637–662.

[33] P.A. Ioannou, J. Sun, Robust Adaptive Control, Courier Corporation, 2012.

[34] M. Krstić, P. Kokotović, Adaptive nonlinear design with controller-identifier separation and swapping, IEEE Trans. Automat. Control 40 (3) (1995) 426–440.

[35] M. Krstić, P.V. Kokotović, Modular approach to adaptive nonlinear stabilization, Automatica 32 (4) (1996) 625–629.

[36] A. Astolfi, D. Karagiannis, R. Ortega, Nonlinear and Adaptive Control with Applications, Springer, 2007.

[37] S. Aranovskiy, A. Bobtsov, R. Ortega, A. Pyrkin, Performance enhancement of parameter estimators via dynamic regressor extension and mixing, IEEE Trans. Automat. Control 62 (7) (2017) 3546–3550.

[38] R. Ortega, S. Aranovskiy, A.A. Pyrkin, A. Astolfi, A.A. Bobtsov, New results on parameter estimation via dynamic regressor extension and mixing: Continuous and discrete-time cases, IEEE Trans. Automat. Control 66 (5) (2021) 2265–2272.

[39] I.Y. Tyukin, E. Steur, H. Nijmeijer, C. van Leeuwen, Adaptive observers and parameter estimation for a class of systems nonlinear in the parameters, Automatica 49 (8) (2013) 2409–2423.

[40] C. Afri, V. Andrieu, L. Bako, P. Dufour, State and parameter estimation: A nonlinear luenberger observer approach, IEEE Trans. Automat. Control 62 (2) (2017) 973–980.

[41] M. Buisson-Fenet, V. Morgenthaler, S. Trimpe, F. Di Meglio, Joint state and dynamics estimation with high-gain observers and Gaussian process models, IEEE Control Syst. Lett. 5 (5) (2021) 1627–1632.

[42] P.A. Ioannou, P.V. Kokotović, Instability analysis and improvement of robustness of adaptive control, Automatica 20 (5) (1984) 583–594.

[43] C. Rohrs, L. Valavani, M. Athans, G. Stein, Robustness of continuous-time adaptive control algorithms in the presence of unmodeled dynamics, IEEE Trans. Automat. Control 30 (9) (1985) 881–889.

[44] Z.H. Li, C. Wen, C.B. Soh, T.Y. Chai, Robustness of Krstic's new adaptive control scheme, in: 3rd IFAC Symposium on Nonlinear Control Systems Design, 1995, pp. 115–120.

[45] P. Ioannou, K. Tsakalis, A robust direct adaptive controller, IEEE Trans. Automat. Control 31 (11) (1986) 1033–1043.

[46] R.A. Freeman, M. Krstić, P.V. Kokotović, Robustness of adaptive nonlinear control to bounded uncertainties, Automatica 34 (10) (1998) 1227–1230.

[47] F. Ikhouane, M. Krstić, Adaptive backstepping with parameter projection: Robustness and asymptotic performance, Automatica 34 (4) (1998) 429–435.

[48] Z.-P. Jiang, L. Praly, Design of robust adaptive controllers for nonlinear systems with dynamic uncertainties, Automatica 34 (7) (1998) 825–840.
[49] F. Mazenc, M. de Queiroz, M. Malisoff, Uniform global asymptotic stability of a class of adaptively controlled nonlinear systems, IEEE Trans. Automat. Control 54 (5) (2009) 1152–1158.
[50] E. Lavretsky, K.A. Wise, Robust and Adaptive Control: With Aerospace Applications, Springer, 2013.
[51] A. Serrani, A. Isidori, L. Marconi, Semi-global nonlinear output regulation with adaptive internal model, IEEE Trans. Automat. Control 46 (8) (2001) 1178–1194.
[52] F. Delli Priscoli, L. Marconi, A. Isidori, A new approach to adaptive nonlinear regulation, SIAM J. Control Optim. 45 (3) (2006) 829–855.
[53] P. Bernard, M. Bin, L. Marconi, Adaptive output regulation via nonlinear Luenberger observer-based internal models and continuous-time identifiers, Automatica 122 (2020) 109261.
[54] K.S. Narendra, A.M. Annaswamy, Persistent excitation in adaptive systems, Internat. J. Control 45 (1) (1987) 127–160.
[55] Y. Song, K. Zhao, M. Krstić, Adaptive control with exponential regulation in the absence of persistent excitation, IEEE Trans. Automat. Control 62 (5) (2017) 2589–2596.
[56] N. Barabanov, R. Ortega, On global asymptotic stability of $\dot{x} = -\phi(t)\phi^\top(t)x$ with ϕ not persistently exciting, Systems Control Lett. 109 (2017) 24–29.
[57] P. Bernard, Observer Design for Nonlinear Systems, Springer, 2019.
[58] J. Sjöberg, T. McKelvey, L. Ljung, On the use of regularization in system identification, in: IFAC 12th Triennial World Congress, 1993, pp. 75–80.