



# A note on the convergence of a class of adaptive optimal identifiers

Laurent Bako

## ► To cite this version:

Laurent Bako. A note on the convergence of a class of adaptive optimal identifiers. Automatica, 2023, 157, 10.1016/j.automatica.2023.111225 . hal-04571503

**HAL Id: hal-04571503**

**<https://hal.science/hal-04571503>**

Submitted on 7 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A note on the convergence of a class of adaptive optimal identifiers<sup>★</sup>

Laurent Bako<sup>a</sup>

<sup>a</sup>Univ Lyon, Ecole Centrale Lyon, INSA Lyon, Université Claude Bernard Lyon 1, CNRS, Ampère, UMR 5005, 69130 Ecully, France

---

## Abstract

This paper proposes a unifying framework for the convergence analysis of a class of adaptive optimal identifiers. The considered class of identifiers is constructed from the sequence of minimizing sets of a family of objective functions. For the purpose of the analysis we introduce a generalized version of the classical persistence of excitation condition. Based on this an Input-to-State-Stability (ISS)-type of property is derived for the studied class of adaptive estimators.

*Key words:* system identification, optimal adaptive identifier, online estimation.

---

## 1 Introduction

Estimation algorithms can be roughly classified into two main categories: batch methods and adaptive ones. While the first type of estimators produces a single estimate based on a finite collection of data, the second type generates a sequence of estimates by processing the data sequentially [2]. With an appropriate weighting of the data, adaptive estimators can be endowed with the capacity of tracking slow variations in the parameter being estimated. In some cases it is even possible to find a cheap and recursive updating mechanism which computes the estimate at a time step from the one obtained at the preceding time step and the new incoming information. In these particular cases, adaptive algorithms (e.g., the stochastic gradient descent algorithm) may present a much lower complexity than their batch counterparts, a property which is of practical interest when dealing with high dimensional or large amounts of data. Well-known adaptive identification algorithms include the recursive least squares (RLS with or without exponential forgetting factor) [7], the least mean squares and their respective variants [2].

This note considers a general adaptive estimation framework for linearly parameterized systems. Its contribution consists in the unified analysis of the convergence properties of a class of optimization-based adaptive identifiers. Our main result is the derivation of an Input-to-State (ISS)-type [3] of property under a new generalized persistency of excitation condition. An interesting feature of the pro-

posed analysis is that it relies (only) on the properties of the loss functions from which the estimating optimization problem is formed. That is, no explicit expression of the solution to the optimization problem is required for the analysis. This makes the approach very generic as it renders it applicable to a variety of estimators including those resulting from smooth/nonsmooth or convex/nonconvex loss functions. Many existing convergence analysis methods for adaptive optimal estimators apply to a single estimator and require in general an analytic expression of the solution of the optimization problem (e.g., the analysis of the RLS in [4]). In contrast, the current paper takes a more abstract analysis approach to adaptive optimal identifiers thereby revealing the precise convergence-inducing properties of the minimized loss function and of the training data.

Note that the focus of the paper is only to analyze the behavior of the sequence of estimates generated by the optimal adaptive identifier ; we do not consider implementation aspects such as computability of the estimates or numerical complexity.

**Notation.**  $\mathbb{R}$  (resp.  $\mathbb{R}_+$ ) denotes the set of real numbers (resp. nonnegative real numbers). We use  $\mathbb{R}^*$  to denote the set of nonzero real numbers and  $\mathbb{N}$  for the set of natural integers.  $\|\cdot\|$  refers to an arbitrary norm.  $I_n$  denotes the identity matrix of dimension  $n$ . For two symmetric matrices  $S_1$  and  $S_2$  the notation  $S_1 \preceq S_2$  means that  $S_2 - S_1$  is a positive definite matrix. A function  $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is said to be of class  $\mathcal{K}_\infty$  [5] if it is continuous, zero at zero, strictly increasing and satisfies  $\lim_{r \rightarrow +\infty} \alpha(r) = +\infty$ . We denote with  $\mathcal{K}_\infty$  the set of all such functions. Note that any  $\alpha \in \mathcal{K}_\infty$  is invertible and its inverse, which we will denote by  $\alpha^{-1}$ , is also a  $\mathcal{K}_\infty$  function.

---

<sup>★</sup> This paper was not presented at any IFAC meeting. Corresponding author L. Bako. Tel.: +33 472 186 452

Email address: laurent.bako@ec-lyon.fr (Laurent Bako).

## 2 Estimation setting

We consider a linearly parameterized data-generating system defined by

$$y_t = x_t^\top \theta^\circ + v_t, \quad (1)$$

where  $y_t \in \mathbb{R}$  is the output at discrete time  $t \in \mathbb{N}$ ,  $x_t \in \mathbb{R}^n$  is the predictor (regressor vector) and  $v_t \in \mathbb{R}$  represents modelling uncertainty.  $\theta^\circ \in \mathbb{R}^n$  is an unknown and constant parameter vector to be estimated.

Although the following analysis is, for simplicity reason, presented here for a single output model, the results are directly applicable to multivariate measurements model (i.e., situations where the to-be-estimated parameter vector  $\theta^\circ$  would be a matrix).

We consider the problem of finding a sequence  $\{\hat{\theta}_t\}$  of estimates for the parameter vector  $\theta^\circ$  in (1) such that each estimate  $\hat{\theta}_t$  relies on the data  $z^t = \{(x_\tau, y_\tau) : \tau = 1, \dots, t\}$  generated by the system up to time  $t$ . The focus of this note is to analyze the properties of a class of optimization-based estimators which generate the sequence of estimates such that

$$\hat{\theta}_t \in \arg \min_{\theta \in \mathbb{R}^n} V_t(\theta), \quad (2)$$

where the objective function  $V_t : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is taken to be of the form

$$V_t(\theta) = \sum_{k=1}^t \lambda^{t-k} \psi(y_k - x_k^\top \theta) + \lambda^t \psi_0(\theta - \hat{\theta}_0) \quad (3)$$

for  $t \geq 1$  and  $V_0(\theta) = \psi_0(\theta - \hat{\theta}_0)$ . Here,  $\hat{\theta}_0$  denotes some initial guess (or prior estimate) of the parameter vector  $\theta^\circ$ ,  $\lambda \in (0, 1)$  is called a forgetting factor and  $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $\psi_0 : \mathbb{R}^n \rightarrow \mathbb{R}_+$  are some loss functions. In words, Eq. (2) says that  $\hat{\theta}_t$  is selected as one minimizing point of the performance index  $\theta \mapsto V_t(\theta)$ .

Indeed, most adaptive schemes are recursive (greedy) heuristics in that they do not necessarily minimize at each time  $t$  a global objective  $V_t$  of the form (3). The approach taken here is to first setup as in (3) the cumulative performance function from which the estimates are generated. The question of whether the solution can then be computed exactly or approximately through a one step-ahead recursive update equation rather pertains to practical implementation aspects which are beyond the intended scope of the current paper. We rather focus on well-definedness of the sequence  $\{\hat{\theta}_t\}$  from (2) (i.e., existence of a minimum value for  $V_t$  at each time  $t$ ) and the convergence property.

For future reference, we consider the following properties. A positive function  $\ell : \mathbb{R}^n \rightarrow \mathbb{R}_+$  is said to

- A1 be positive-definite if  $\ell(x) = 0 \Leftrightarrow x = 0$ .
- A2 be symmetric if  $\ell(x) = \ell(-x) \forall x \in \mathbb{R}^n$ .

A3 satisfy a *generalized triangle inequality (GTI)* if there exists a constant number  $\alpha_\ell > 0$  such that

$$\ell(x - y) \geq \alpha_\ell \ell(x) - \ell(y) \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n. \quad (4)$$

A4 satisfy a *generalized homogeneity (GH)* property if there exists a  $\mathcal{K}_\infty$  function  $f_\ell$  such that

$$\ell(x) \geq f_\ell\left(\frac{1}{|r|}\right) \ell(rx) \quad \forall (x, r) \in \mathbb{R}^n \times \mathbb{R}^*. \quad (5)$$

We will assume throughout the paper that the loss functions  $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $\psi_0 : \mathbb{R}^n \rightarrow \mathbb{R}_+$  are *continuous* on their respective domains. Moreover, they will be required, according to the situation, to satisfy a subset of the properties A1–A4. Note that  $\psi$  and  $\psi_0$  need not be convex for the following reasoning to be valid. Examples of candidate functions for  $\psi$  may be all power functions of the form  $e \mapsto |e|^p/p$  with  $p$  being a strictly positive real number, or the scalar Huber loss defined on  $\mathbb{R}$ , for a given  $h > 0$ , by  $e \mapsto e^2/2$  if  $|e| \leq h$  and  $e \mapsto h(|e| - h/2)$  otherwise. In the special case where  $\psi(e) = e^2/2$  and  $\psi_0(\theta) = \theta^\top P_0^{-1} \theta$  (with  $P_0$  being a symmetric positive definite matrix), a closed form expression of the solution  $\hat{\theta}_t$  can indeed be obtained; moreover, that solution can be written in a recursive form (called the RLS algorithm, see, e.g., [7]).

**Lemma 1** *Let  $\ell : \mathbb{R}^n \rightarrow \mathbb{R}_+$  be a continuous function which has the properties A1–A2 and A4 with a  $\mathcal{K}_\infty$  function  $f_\ell$ . Then, for all norm  $\|\cdot\|$  on  $\mathbb{R}^n$ , there exist  $\xi_1, \xi_2 \in \mathcal{K}_\infty$  such that*

$$\xi_1(\|\theta\|) \leq \ell(\theta) \leq \xi_2(\|\theta\|) \quad \forall \theta \in \mathbb{R}^n. \quad (6)$$

A particular choice of  $\xi_1$  and  $\xi_2$  is given by  $\xi_1(r) = D_1 f_\ell(r)$  and  $\xi_2(r) = D_2 \frac{1}{f_\ell(1/r)}$  for  $r \neq 0$  and  $\xi_2(0) = 0$ , with  $D_1 = \min_{\|\theta\|=1} \ell(\theta) > 0$  and  $D_2 = \max_{\|\theta\|=1} \ell(\theta) > 0$ .

**PROOF.** The proof follows similar arguments as that of Lemma 2 in [6]. First, by the continuity of  $\ell$  we know that it admits minimum value  $D_1$  and maximum value  $D_2$  on the compact set  $\mathcal{S} = \{\theta \in \mathbb{R}^n : \|\theta\| = 1\}$ . Second, by the generalized homogenous property A4 of  $\ell$  we have

$$f_\ell(\|\theta\|) \ell\left(\frac{\theta}{\|\theta\|}\right) \leq \ell(\theta) \leq \frac{1}{f_\ell(1/\|\theta\|)} \ell\left(\frac{\theta}{\|\theta\|}\right) \quad (7)$$

for all  $\theta \neq 0$ . The result follows by noting that  $\theta/\|\theta\|$  lies in  $\mathcal{S}$ , which implies that  $D_1 \leq \ell(\theta/\|\theta\|) \leq D_2$ .  $\square$

## 3 General convergence analysis

We start the analysis by introducing a concept of persistence of excitation on the predictor sequences  $\{x_t\}$ .

**Definition 1** The sequence  $\{x_t\}$  is said to be persistently exciting (PE) with respect to the loss function  $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$  (subject to A1-A2 and A4) if there exist two  $\mathcal{K}_\infty$  functions  $\alpha$  and  $\beta$  and a fixed time horizon  $T > 0$  such that

$$\alpha(\|\theta\|) \leq \sum_{k=t+1}^{t+T} \psi(x_k^\top \theta) \leq \beta(\|\theta\|) \quad \forall (t, \theta) \in \mathbb{N} \times \mathbb{R}^n \quad (8)$$

for some norm  $\|\cdot\|$  on  $\mathbb{R}^n$ .

Definition 1 can be interpreted as a generalization of the classical PE condition, see e.g., [4]. For example, in the special case where  $\psi$  and the  $\mathcal{K}_\infty$  functions  $\alpha$  and  $\beta$  are defined respectively by  $\psi(e) = e^2$ ,  $\alpha(r) = ar^2$  and  $\beta(r) = br^2$  for some  $a > 0$  and  $b > 0$ , Eq.(8) becomes, when Euclidean norm is substituted for the generic norm  $\|\cdot\|$ ,

$$a\theta^\top \theta \leq \sum_{k=t+1}^{t+T} \theta^\top x_k x_k^\top \theta \leq b\theta^\top \theta \quad \forall (t, \theta) \in \mathbb{N} \times \mathbb{R}^n.$$

The latter is equivalent to  $aI_n \preceq \sum_{k=t+1}^{t+T} x_k x_k^\top \preceq bI_n \quad \forall t \in \mathbb{N}$ . The latter corresponds to the classical PE condition [4]. The PE condition (8) can indeed be equivalently reformulated as follows.

**Lemma 2** Assume that  $\psi$  is continuous and satisfies A1-A2 and A4. Then Eq. (8) holds true if and only if there exist positive constants  $\gamma_1$  and  $\gamma_2$  such that

$$\gamma_1 \leq \sum_{k=t+1}^{t+T} \psi\left(\frac{x_k^\top \theta}{\|\theta\|}\right) \leq \gamma_2 \quad \forall (t, \theta) \in \mathbb{N} \times \mathbb{R}^n \setminus \{0\}. \quad (9)$$

**PROOF.** The proof follows the same idea as that of Lemma 1. In effect, if (9) holds then by applying the inequality (7) with  $\ell(\theta) = \sum_{k=t+1}^{t+T} \psi(x_k^\top \theta)$  we immediately see that (8) holds with  $\mathcal{K}_\infty$  functions  $\alpha$  and  $\beta$  defined respectively by  $\alpha(r) = \gamma_1 f_\psi(r)$  and  $\beta(r) = \gamma_2 \frac{1}{f_\psi(1/r)}$  for  $r \neq 0$  and  $\beta(0) = 0$ .

Conversely, to see that (8) implies (9), it suffices to replace  $\theta$  with  $\theta/\|\theta\|$  in (8) and take  $\gamma_1 = \alpha(1)$  and  $\gamma_2 = \beta(1)$ .  $\square$

**Lemma 3** Let  $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$  and  $\psi_0 : \mathbb{R}^n \rightarrow \mathbb{R}_+$  be continuous functions satisfying the properties A1-A2 and A4. Consider a function  $G_t : \mathbb{R}^n \rightarrow \mathbb{R}_+$  defined by

$$G_t(\theta) = \alpha_\psi \sum_{k=1}^t \lambda^{t-k} \psi(x_k^\top \theta) + \alpha_{\psi_0} \lambda^t \psi_0(\theta), \quad (10)$$

where  $x_k \in \mathbb{R}^n \forall k$ ,  $\alpha_\psi > 0$ ,  $\alpha_{\psi_0} > 0$  and  $\lambda \in (0, 1)$ . If  $\{x_t\}$  satisfies the PE condition (8) with respect to the function  $\psi$ , then there exist time-invariant  $\mathcal{K}_\infty$  functions  $g_1$  and  $g_2$  such that

$$g_1(\|\theta\|) \leq G_t(\theta) \leq g_2(\|\theta\|) \quad \forall (t, \theta) \in \mathbb{N} \times \mathbb{R}^n. \quad (11)$$

**PROOF.** Since  $\{x_t\}$  is PE with respect to  $\psi$ , there exist  $\alpha \in \mathcal{K}_\infty$  and a number  $T > 0$  such that (8) holds. For any  $t \in \mathbb{N}$ , consider the pair  $(q_t, r_t)$  of integers denoting the quotient and the rest of the euclidean division of  $t$  by  $T$ , i.e., such that  $t = q_t T + r_t$  with  $0 \leq r_t < T$ .

Then by a similar reasoning as in [1], the following series of inequalities holds for  $t \geq T$ :

$$\begin{aligned} G_t(\theta) &\geq \lambda^t \sum_{i=1}^{q_t} \sum_{k=(i-1)T+1}^{iT} \lambda^{-k} \alpha_\psi \psi(x_k^\top \theta) \\ &= \lambda^{(q_t+1)T-1} \sum_{i=1}^{q_t} \lambda^{-(i-1)T} \sum_{k=(i-1)T+1}^{iT} \alpha_\psi \psi(x_k^\top \theta). \end{aligned}$$

Recalling that  $\sum_{k=(i-1)T+1}^{iT} \psi(x_k^\top \theta) \geq \alpha(\|\theta\|)$  by the PE condition, we can write

$$\begin{aligned} G_t(\theta) &\geq \lambda^{(q_t+1)T-1} \sum_{i=1}^{q_t} \lambda^{-(i-1)T} \alpha_\psi \alpha(\|\theta\|) \\ &= \lambda^{(q_t+1)T-1} \frac{1 - \lambda^{-q_t T}}{1 - \lambda^{-T}} \alpha_\psi \alpha(\|\theta\|) \geq g_{11}(\|\theta\|) \end{aligned}$$

with  $g_{11}(r) = \alpha_\psi \lambda^{2T-1} \alpha(r)$ . On the other hand, for  $t < T$ , we can write  $G_t(\theta) \geq \alpha_{\psi_0} \lambda^{T-1} \psi_0(\theta)$ . Since  $\psi_0$  is subject to the conditions A1-A2 and A4, we conclude from Lemma 1 that there exists  $g_{12} \in \mathcal{K}_\infty$  such that  $\alpha_{\psi_0} \lambda^{T-1} \psi_0(\theta) \geq g_{12}(\|\theta\|)$  for  $0 \leq t < T$ . Taking then  $g_1 = \min(g_{11}, g_{12})$ , it can be shown that  $g_1 \in \mathcal{K}_\infty$ . Hence, the lower bound in (11) is proved.

We now turn to proving the upper bound. For all  $t \geq T$ :

$$\begin{aligned} G_t(\theta) &\leq \lambda^t \sum_{i=1}^{q_t+1} \sum_{k=(i-1)T+1}^{iT} \lambda^{-k} \alpha_\psi \psi(x_k^\top \theta) + \alpha_{\psi_0} \lambda^t \psi_0(\theta) \\ &\leq \alpha_\psi \lambda^T \sum_{i=1}^{q_t+1} \lambda^{-iT} \sum_{k=(i-1)T+1}^{iT} \psi(x_k^\top \theta) + \alpha_{\psi_0} \lambda^T \psi_0(\theta) \\ &\leq \alpha_\psi \lambda^T \sum_{i=1}^{q_t+1} \lambda^{-iT} \beta(\|\theta\|) + \alpha_{\psi_0} \lambda^T \psi_0(\theta) \\ &\leq \alpha_\psi \frac{1 - \lambda^{-2T}}{1 - \lambda^{-T}} \beta(\|\theta\|) + \alpha_{\psi_0} \lambda^T \psi_0(\theta). \end{aligned}$$

The right hand side function satisfies the conditions of Lemma 1. Therefore, we can find a time-invariant  $g' \in \mathcal{K}_\infty$  such that  $G_t(\theta) \leq g'(\|\theta\|)$ . Consider now the case where  $0 \leq t < T$ . For each such  $t$ ,  $G_t(\theta)$  fulfills the conditions of Lemma 1. As a consequence, there exist  $\mathcal{K}_\infty$  functions  $g_{2,t}$ ,  $t = 0, \dots, T-1$ , such that  $G_t(\theta) \leq g_{2,t}(\|\theta\|)$ . It suffices then to take  $g_2 = \max(g', g_{2,0}, \dots, g_{2,T-1})$  (which is in  $\mathcal{K}_\infty$ ) for the upper bound in (11) to hold.  $\square$

The next lemma gives sufficient conditions for the the esti-

mates sequence  $\{\hat{\theta}_t\}$  to be well-defined.

**Lemma 4** Consider the performance index (3) with  $\psi$  and  $\psi_0$  satisfying all the properties A1–A4 along with continuity. If the regressor sequence  $\{x(t)\}$  satisfies the PE condition stated in (1), then the sequence of estimates (2) is well-defined, that is,  $\arg \min_{\theta} V_t(\theta) \neq \emptyset$  for all  $t \in \mathbb{N}$ .

**PROOF.** The idea of the proof is to show that  $V_t$  is coercive (i.e., continuous and radially unbounded) and then apply a result in [8] to conclude on the existence of a minimizer of  $V_t$  for all  $t$ . For this purpose, note that  $V_t$  is continuous as a consequence of  $\psi$  and  $\psi_0$  being continuous. Moreover, from the properties A2–A3 of  $\psi$  and  $\psi_0$ , one can write

$$\begin{aligned} V_t(\theta) &\geq G_t(\theta) - \sum_{k=1}^t \lambda^{t-k} \psi(y_k) - \lambda^t \psi_0(\hat{\theta}_0) \\ &\geq g_1(\|\theta\|) - \sum_{k=1}^t \lambda^{t-k} \psi(y_k) - \lambda^t \psi_0(\hat{\theta}_0) \end{aligned}$$

for some norm  $\|\cdot\|$  on  $\mathbb{R}^n$ . Here,  $g_1$  is defined as in (11). From the inequality above it is clear that  $\lim_{\|\theta\| \rightarrow \infty} V_t(\theta) = +\infty$ . Hence, we conclude that  $V_t$  is coercive. As a consequence,  $V_t$  admits a minimum value for all  $t$ . The sequence of estimates (2) is therefore well-defined as claimed.  $\square$

The claim of the lemma is that  $V_t$  admits a minimum value but the minimizer of  $V_t(\theta)$  need not be unique. The potential nonuniqueness of the sequence  $\{\hat{\theta}_t\}$  is not an issue as the derived convergence property holds for any sequence of estimate satisfying (2).

**Theorem 5** Consider a parameter sequence  $\{\hat{\theta}_t\}$  generated as in (2) with the assumption that the functions  $\psi$  and  $\psi_0$  in (3) are continuous and satisfy the conditions A1–A3. If the sequence  $\{x_t\}$  of regressors satisfies the PE condition as stated in Definition 1 for a given norm  $\|\cdot\|$  on  $\mathbb{R}^n$ , then there exists  $\xi \in \mathcal{K}_\infty$  such that

$$\|\hat{\theta}_t - \theta^\circ\| \leq \xi^{-1} \left[ 2\lambda^t \psi_0(\hat{\theta}_0 - \theta^\circ) + 2 \sum_{k=1}^t \lambda^{t-k} \psi(v_k) \right] \quad (12)$$

for all  $t \in \mathbb{N}$ .

**PROOF.** By invoking Eq. (1), it is easy to see that

$$V_t(\theta) = \sum_{k=1}^t \lambda^{t-k} \psi(v_k - x_k^\top (\theta - \theta^\circ)) + \lambda^t \psi_0((\theta - \theta^\circ) - (\hat{\theta}_0 - \theta^\circ)).$$

Now pose  $\eta_t = \hat{\theta}_t - \theta^\circ$  so that  $\eta_0 = \hat{\theta}_0 - \theta^\circ$ . Then we can observe that

$$V_t(\hat{\theta}_t) = \sum_{k=1}^t \lambda^{t-k} \psi(v_k - x_k^\top \eta_t) + \lambda^t \psi_0(\eta_t - \eta_0)$$

and

$$V_t(\theta^\circ) = \sum_{k=1}^t \lambda^{t-k} \psi(v_k) + \lambda^t \psi_0(\eta_0).$$

By the generalized triangle inequality property of  $\psi$  and  $\psi_0$ , we have  $\psi(v_k - x_k^\top \eta_t) \geq \alpha_\psi \psi(x_k^\top \eta_t) - \psi(v_k)$  and  $\psi_0(\eta_t - \eta_0) \geq \alpha_{\psi_0} \psi_0(\eta_t) - \psi_0(\eta_0)$ . Making use of these inequalities give

$$G_t(\eta_t) - V_t(\theta^\circ) \leq V_t(\hat{\theta}_t) \leq V_t(\theta^\circ),$$

where  $G_t$  is defined as in (10). In the above equation, the right hand side inequality follows from the definition of  $\hat{\theta}_t$  as a minimizer of  $V_t$ . Therefore, we have  $G_t(\eta_t) \leq 2V_t(\theta^\circ)$ . By applying Lemma 3 for a given norm  $\|\cdot\|$  on  $\mathbb{R}^n$ , there exists a time-invariant  $\mathcal{K}_\infty$  function  $\xi$  such that  $G_t(\eta_t) \geq \xi(\|\eta_t\|)$ . Combining this with the inequality above yields

$$\|\eta_t\| \leq \xi^{-1}(2V_t(\theta^\circ))$$

which is the stated result stated by the theorem.  $\square$

Theorem 5 states a property of ISS (see, e.g., [3] for a definition) for the system  $(\eta_0, \{v_t\}) \mapsto \{\eta_t\}$  which maps the initial parametric error  $\eta_0 = \hat{\theta}_0 - \theta^\circ$  and the measurement noise  $\{v_t\}$  to the estimation error sequence  $\{\eta_t\}$ . What it says is that the estimator forgets exponentially fast the initial condition and produces an error which is bounded by an increasing function of the amount of noise.

Assuming that the noise sequence  $\{v_t\}$  is bounded, we obtain an immediate consequence of Theorem 5 in the form of an asymptotic upper bound on the estimation error.

**Corollary 6** In addition to the conditions of Theorem 5, assume that the noise sequence  $\{v_t\}$  is bounded and  $\lambda \in (0, 1)$ . Then there exists a  $\mathcal{K}_\infty$  function  $\xi$  such that

$$\limsup_{t \rightarrow +\infty} \|\hat{\theta}_t - \theta^\circ\| \leq \xi^{-1} \left( \frac{2}{1-\lambda} \sup_{t \in \mathbb{N}} \psi(v_t) \right)$$

for any sequence  $\{\hat{\theta}_t\}$  generated according to (2).

A direct implication of this corollary is that in the noise-free case, we have  $\lim_{t \rightarrow +\infty} \hat{\theta}_t = \theta^\circ$ . More, it can be seen from Eq. (12) that the associated rate of convergence is exponential.

**Example 1.** If  $\psi$  and  $\psi_0$  are instantiated respectively by

$\psi(e) = |e|^p$  for a real  $p > 0$  and  $\psi_0(\theta) = \gamma_0 \|\theta\|^2$ , then the properties A1-A3 are satisfied. In particular, the constant  $\alpha_\psi$  in (4) can be taken equal to  $2^{1-1/p}$  if  $0 < p \leq 1$  and  $2^{1-p}$  otherwise. Moreover, we may take the functions  $\alpha$  and  $\beta$  in the PE condition (8) to be of the form  $\alpha(r) = ar^p$  and  $\beta(r) = br^p$  respectively for some positive constants  $a$  and  $b$ . Using now the proofs of Lemma 3 and Theorem 5 we determine a valid candidate for the  $\mathcal{K}_\infty$  function  $\xi$  in (12) as  $\xi(r) = \min(a\alpha_\psi\lambda^{2T-1}r^p, \gamma_0/2\lambda^{T-1}r^2)$ .

## 4 Conclusion

This paper has proposed a unified framework for convergence analysis of a class of adaptive optimization-based estimators. More precisely, an ISS-type of property has been derived for the parametric estimation error under a generalized condition of persistence of excitation. An interesting feature of the proposed analysis is that it relies only on the assumed properties of the loss function entering the definition of the to-be-minimized objective function, that is, no closed-form expression of the estimates is required in the analysis. This makes the analysis applicable to a wide range of optimal adaptive estimators. Moreover, it characterizes the precise convergence-inducing properties of the objective function and of the training data.

## References

- [1] L. Bako. Adaptive identification of linear systems subject to gross errors. *Automatica*, 67:192–199, 2016.
- [2] G. C. Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Dover Publications Inc., Mineola, New York, 2009.
- [3] Z.-P. Jiang and Y. Wang. Input-to-state stability for discrete-time nonlinear systems. *Automatica*, 37:857–869, 2001.
- [4] R. M. Johnstone, C. R. Johnson Jr, R. R. Bitmead, and B. D. Anderson. Exponential convergence of recursive least squares with exponential forgetting factor. *Systems & Control Letters*, 2:77–82, 1982.
- [5] C. M. Kellett. A compendium of comparison function results. *Mathematics of Control, Signals, and Systems*, 26:339–374, 2014.
- [6] A. Kircher, L. Bako, E. Blanco, and M. Benallouch. An optimization framework for resilient batch estimation in cyber-physical systems. *IEEE Transactions on Automatic Control*, 67:5246–5261, 2021.
- [7] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, 1983.
- [8] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Springer Verlag, 1997.