A new least squares parameter estimator for nonlinear regression equations with relaxed excitation conditions and forgetting factor

Romeo Ortega, Jose Guadalupe Romero, Stanislav Aranovskiy

A New Least Squares Parameter Estimator for Nonlinear Regression Equations with Relaxed Excitation Conditions and Forgetting Factor

Romeo Ortega[a,c], Jose Guadalupe Romero[a], Stanislav Aranovsky[b]

[a]*Departamento Académico de Sistemas Digitales, ITAM, Progreso Tizapán 1, Ciudad de México, 04100, México*
[b]*IETR–CentaleSupélec, Avenue de la Boulaie, 35576 Cesson-Sévigné, France*
[c]*Department of Control Systems and Robotics, ITMO University, Kronverkskiy av. 49, Saint Petersburg, 197101, Russia*

Abstract

In this note a new high performance least squares parameter estimator is proposed. The main features of the estimator are: (i) global exponential convergence is guaranteed for all *identifiable* linear regression equations; (ii) it incorporates a *forgetting factor* allowing it to preserve alertness to time-varying parameters; (iii) thanks to the addition of a mixing step it relies on a set of *scalar* regression equations ensuring a superior transient performance; (iv) it is applicable to nonlinearly parameterized regressions verifying a *monotonicity* condition and to a class of systems with *switched time-varying* parameters; (v) it is shown that it is *bounded-input-bounded-state* stable with respect to additive disturbances; (vi) continuous and discrete-time versions of the estimator are given. The superior performance of the proposed estimator is illustrated with a series of examples reported in the literature.

Keywords: Parameter estimation, Least Squares, Least squares identification algorithm Nonlinear regression model, Exponentially convergent identification.

1. Introduction

We have witnessed in the last few years an increasing interest in the analysis and design of new parameter estimators for linearly paramterized regression equations (LPRE) of the form $y(t) = \phi^\top(t)\theta$, with $y(t) \in \mathbb{R}$, $\phi(t) \in \mathbb{R}^q$ measurable signals and $\theta \in \mathbb{R}^q$ a constant vector of unknown parameters.[1] The main motivation of this research is to relax the highly restrictive assumption of *persistent excitation* (PE) imposed to guarantee global exponential convergence of classical gradient, least squares (LS) or Kalman-Bucy algorithms [16, 48, 52]. A second important motivation is to provide guaranteed good transient performance behavior since the one of the aforementioned schemes is highly unpredictable and only a weak monotonicity property of the *norm* of the vector of estimation errors can be insured.

1.1. Review of recent literature on LS estimators

It has recently been shown [3, 43] that global asymptotic convergence—but not exponential—of the error equation for standard continous-time (CT) gradient estimators is ensured under a strictly weaker condition of *generalized PE*—see [3, 55] for its definition and [13] for some robustness properties of the algorithm. Unfortunately, this condition is still extremely restrictive to be of practical use. In [7] it is shown that the classical discrete-time (DT) LS algorithm is asymptotically convergent if and only if the regressor $\phi(t)$ satisfies a new excitation property, called *weak PE*, that is strictly weaker than PE. This result is of limited interest because, on one hand, the definition is extremely technical and difficult to verify in applications. On the other hand, and more importantly, the analysis is limited to standard LS, without forgetting factor or covariance resetting that, as is well-known [16, 48], has a decreasing adaptation gain, loosing its alertness to track parameter variations, which is the main motivation for recursive algorithms. In [29] the underexcited scenario where the Gram matrix of the regressor has a q_0-dimensional kernel, with $q_0 \leq q$, is considered. It is shown that incorporating into a CT LS (or gradient) estimator the information of a basis expanding this kernel—the columns of the matrix $N \in \mathbb{R}^{q_0 \times q}$ in equation (14) or (24) whose columns $v_i \in \mathbb{R}^q$, $i = 1, \dots q_0$ satisfy equation (4)—it is possible to guarantee consistent estimation to its complementary space. Clearly, if the regressor is PE the dimension q_0 of the aforementioned kernel is zero and convergence of all the parameters is guaranteed, confirming the well-known result of the PE case. This result is related to the *partial convergence* property of [48, Theorem 2.7.4] where a similar fact is proven in the context of systems identification in underexcited situations. Although of theoretical interest, the result has little practical relevance because of the

Email addresses: romeo.ortega@itam.mx (Romeo Ortega), jose.romerovelazquez@itam.mx (Jose Guadalupe Romero), stanislav.aranovskiy@centralesupelec.fr (Stanislav Aranovsky)

[1] We consider the case of *scalar* $y(t)$ to simplify the notation—as will be seen below all results can be directly extended for the case of vector $y(t)$.

impossibilty to compute on-line the matrix N mentioned above. In [10] a slight modification of the DT LS algorithm is proposed to deal with *parameter variations* in the LPRE, however the main convergence (to a compact set) results still rely on PE assumptions. In [22], an LS version of the well-known *I&I estimator* [2] is proposed while in [28] a variation of LS that defines a *passive operator* is proposed.

In [49] an interesting generalization of the classical CT LS with forgetting factor is introduced, where the latter is allowed to be a time varying matrix. See [50] for a similar result in DT. The convergence analysis in both papers still relies on the PE assumption. In the very recent paper [5] a general structure to design and analyze DT LS-like recursive estimators parameterized in terms of some free functions is proposed—see equation (7). A new definition of excitation, called $t^\star$-excitation, is given in [5, Definition 1]. Interestingly, this definition involves not only the regressor but also two of the aforementioned functions. A particular choice of these functions yields the classical LS estimator and $t^\star$-excitation is *equivalent* to PE. However, other choices of these free parameters may yield weaker excitation conditions, for instance the choice given in (27). However, this selection has again the problem of driving the adaptation gain to zero, loosing the estimator alertness and it is not clear if there are other choices that do not suffer from this drawback. Another novelty of [5] is that it incorporates a very interesting analysis of robustness to additive disturbances in the LPRE, encrypted in the input-to-state-stability property. The interested reader is referred to [5, 49, 50] for a review of the extensive literature on LS with forgetting factors.

1.2. Relaxing the PE condition

A major breakthrough in the design of recurrent estimators is the proof that it is possible to establish global convergence under the extremely weak assumption of *interval excitation*[2] (IE) [20]—called initial excitation in [41] and excitation over a finite interval in [52]. To the best of the authors' knowledge the first estimators where such a result was established are the *concurrent* and the *composite learning* schemes reported in [9] and [40], respectively; see [34] for a recent survey on new estimators. These algorithms, which incorporate the monitoring of past data to build a stack of suitable regressor vectors, are closer in spirit to *off-line* estimators. See also [19, 31] for two early references where a similar idea is explored. As is well-known, the main drawback of off-line estimators is their inability to track parameter variations, which is very often the main objective in applications. This situation motivates the interest to develop *bona-fide* on-line estimators that relax the PE condition preserving the scheme's alertness [26].

New on-line estimators relying on the use of the *dynamic regressor extension and mixing* (DREM) technique with weaker excitation requirements have been recently proposed. DREM was first proposed in [1] for CT and in [4] for DT systems. In Appendix A it is recalled that the main step in the derivation of DREM estimators is the construction of a new *extended* LPRE $Y(t) = \Phi(t)\theta$, with $Y(t) \in \mathbb{R}^q$ and $\Phi(t) \in \mathbb{R}^{q\times q}$ a new *square matrix* regressor. Two procedures to construct the extended LPRE, reported in [21] and [24], respectively, were originally considered—for the sake of completeness both constructions are reviewed in Appendix A. The final—and critical—*mixing* step consists of the multiplication of this extended LPRE by the *adjugate* of $\Phi(t)$.[3] Clearly, this operation creates a new *scalar* LPRE of the form $\mathcal{Y}_i(t) = \Delta(t)\theta_i,\ i \in \bar{q}$, with $\mathcal{Y}(t) := \text{adj}\{\Phi(t)\}Y(t)$ and $\Delta(t) := \det\{\Phi(t)\}$ a *scalar* regressor, which is the essential feature of the approach.[4] DREM estimators have been successfully applied in a variety of identification and adaptive control problems, both, theoretical and practical ones, see [34, 38] for an account of some of these results.

The convergence properties of DREM-based estimators clearly depend on the scalar regressor $\Delta(t)$. Due to the scalar nature of $\Delta(t)$, it is clear that the parameter error converges if and only if $\Delta(t)$ is not square integrable (summable for DT systems) and convergence is exponential if and only if $\Delta(t)$ is PE, facts that were proven in [1]. In [14] a DREM-based algorithm using the extended regressor of [21] that ensures convergence in *finite-time* imposing the IE assumption on $\Delta(t)$ was proposed. An interesting open question was to establish the relation of the excitation of $\Delta(t)$ and the original regressor $\phi(t)$, which was studied in [17] and [55] for the extended regressors of [21] and [24], respectively. The *equivalence* between PE of $\Delta(t)$ and PE of the original regressor $\phi(t)$ was established for both extended regressors—proving that DREM-based estimators are at least as good as standard gradient or LS schemes for excited LPRE. On the other hand, in [17] it is shown that if $\phi(t)$ is IE then $\Delta(t)$ is also IE for the extended regressor [21], while in [55] it was shown that the scheme of [24] ensures the stronger property that $\Delta(t)$ is bounded away from zero in an open interval $[t_c, \infty)$ with $t_c > 0$. Finally, in [38, Proposition 3] a new extended regressor which guarantees exponential convergence under conditions that are *strictly weaker* than regressor PE was presented.

Three major developments in this line of research reported recently are:

(i) the proposal in [18] and [53] of two new, fully on-line, DREM-based estimators where exponential conver-

[2] It should be pointed out that IE is strictly weaker than the generalized PE of [55], the weak excitation PE property of [7] and the $t^\star$-excitation of [5].

[3] It is interesting to note that this operation was independently reported in [8] in the context of stochastic estimator convergence *analysis*.

[4] In Appendix A DREM is applied to nonlinearly parameterized regression equation (NLPRE) of the form (1), which is also considered in this paper.

gence is established imposing only the *IE condition* to $\phi(t)$;

(ii) the proof in [53] that IE of the regressor $\phi(t)$ is *equivalent* to identifiability of the LPRE. It should be recalled that identifiability of a LPRE is the existence of q linearly independent regressor vectors [53, Definition 2] $\phi(t_i)$, $i \in \bar{q}$, and is a *necessary and sufficient* condition for the on- or off-line estimation of the parameters [16];

(iii) the proof in [53] that the proposed estimator is applicable also to separable *NLPRE* of the form (1), provided the mapping $\mathcal{G}(\theta)$ verifies a monotonicity condition. Estimators for this kind of NLPRE were reported before in [36], but the convergence condition was expressed in terms of the scalar regressor $\Delta(t)$.

The estimators of [18] and [53] rely on the generation of new LPRE using the main idea of *generalized parameter estimation based observer* (GPEBO), which is a technique to design *state observers* for state-affine nonlinear systems, first proposed in [32] and latter generalized in [35]. GPEBO translates the problem of state-estimation into one of *parameter estimation* from a LPRE. The latter is generated exploiting the well-known property [47, Property 4.4] that the trajectories of an LTV system can be expressed as linear combinations of the columns of its fundamental matrix. Besides the addition of the computationally demanding calculation of the fundamental matrix, a potential drawback of GPEBO is that it essentially reconstructs the *initial conditions* of some error equation, an operation which may adversely affect the robustness of the estimator, [38, Remark 7] and [37], see also [54]. The procedure followed in the construction of the estimator of [18] is first the application of DREM and then invoke GPEBO, hence we refer to it as D+G. On the other hand, the estimator of [53] uses also GPEBO and DREM, but in the *opposite order*, so we refer to it in the sequel as G+D.

1.3. Contributions of the paper

In this paper we provide an alternative to the D+G and G+D estimators that also ensures global exponential convergence under the weak assumption of IE of the original regressor $\phi(t)$. The main features of this new estimator are summarized as follows.

F1 In contrast to the D+G and G+D estimators that implement a gradient descent search, we use the classical LS technique, hence we refer to it in the sequel as LS+D estimator. The superior convergence properties of LS estimators, as opposed to gradient-based, are widely recognized [16, 25, 45].

F2 We avoid the use of the GPEBO technique but instead exploit some structural properties of the LS estimator to construct the extended regressor. This fact removes the need to calculate the computationally demanding fundamental matrix.

F3 Similarly to the G+D scheme, the stability mechanism and, consequently, the stability analysis of the LS+D estimator is much more transparent than the one of the D+G estimator. There are two consequences of this fact, on one hand, the procedure of *tuning* the estimator to achieve a satisfactory transient performance, which is difficult for the D+G scheme, is straightforward for the LS+D one.

F4 A time-varying *forgetting factor* that allows the estimator to preserve its alertness to time-varying parameters is incorporated.

F5 Besides the case of LPRE we consider (separable and monotonic) *NLPRE*, with the associated estimator preserving all the properties of the case of LPRE. Also, we show that the proposed estimator is applicable to NLPRE with *switched time-varying* parameters.

F6 We show that the new estimator is *robust* with respect to additive disturbances, by proving that it defines a *bounded-input-bounded-state* (BIBS) stable system.

F7 The behaviour of many physical systems is described via CT models. On the other hand, DT implementations of estimators are of significant practical relevance. Therefore, similarly to [18, 38, 53], to comply with both scenarios we consider in the paper both kinds of LPREs. Interestingly, in contrast to [18], the construction and analysis tools of both cases are essentially the same—however, for the sake of clarity, they are presented in separate sections.

The remainder of the paper is organized as follows. In Section 2 we present the main result of the paper for CT systems, while the DT version is given in Section 3. For the sake of brevity we give both results for the general case of NLPRE, presenting the LPRE case as a corollary. Section 4 is devoted to the derivation of the proposed extended NLPRE applying directly the DREM construction procedure. Section 5 is devoted to the proof of *robustness* of the new estimator. Simulation results of some examples reported in the literature are given in Section 6 to illustrate the superior performance of the proposed LS+D estimator. The paper is wrapped-up with concluding remarks in Section 7.

Notation. I_n is the $n \times n$ identity matrix and $\mathbf{0}_{s\times r}$ is an $s \times r$ matrix of zeros. $\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0}$, $\mathbb{Z}_{>0}$ and $\mathbb{Z}_{\geq 0}$ denote the positive and non-negative real and integer numbers, respectively. For $q \in \mathbb{Z}_{>0}$ we defined the set $\bar{q} := \{1, 2, \ldots, q\}$. For $a \in \mathbb{R}^n$, we denote $|a|^2 := a^\top a$, and for any matrix A its induced norm is $\|A\|$. $\mathrm{vec} : \mathbb{R}^{p\times p} \to \mathbb{R}^{p^2}$ is an operator that piles up the columns of a matrix. CT signals $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ are denoted $x(t)$, while for DT sequences $x : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$ we use $x_k := x(kT_s)$, with $T_s \in \mathbb{R}_{>0}$ the sampling time. The action of an operator $\mathcal{H}$ on a CT signal $s(t)$ is denoted as $\mathcal{H}[s](t)$, and $\mathcal{H}[s](k)$ for a sequence s_k.

2. Main Result for Continuous-time Systems

In this section we present the proposed LS+D interlaced estimator for CT systems, with the first estimator being the LS with bounded-gain forgetting factor proposed in [51, Subsection 8.7.6]. First, we consider the case of NLPRE and then specialize to LPRE that, as expected, ensures stronger convergence properties.

2.1. Nonlinearly parameterized regression equations

Consider the following CT NLPRE

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

where $y(t) \in \mathbb{R}$, $\phi(t) \in \mathbb{R}^p$ and $\mathcal{G} : \mathbb{R}^q \to \mathbb{R}^p$, $q \leq p$, a smooth mapping verifying the following.

Assumption A1. [Monotonicity] There exists a matrix $Q \in \mathbb{R}^{q\times p}$ such that mapping $\mathcal{G}(\theta)$ verifies the linear matrix inequality

$$Q\nabla\mathcal{G}(\theta) + \nabla^\top\mathcal{G}(\theta)Q^\top \geq \rho I_q > 0,\ \forall\, \theta \in \mathbb{R}^q, \tag{2}$$

for some $\rho \in \mathbb{R}_{>0}$. Consequently [12, 42], The mapping $Q\mathcal{G}(\theta)$ is *strongly monotone*, that is,

$$(a-b)^\top\left[Q\mathcal{G}(a) - Q\mathcal{G}(b)\right] \geq \rho|a-b|^2 > 0,\ \forall\, a, b \in \mathbb{R}^q, \tag{3}$$

with $a \neq b$.

Assumption A2. [Interval Excitation] The regressor $\phi(t)$ is interval exciting (IE) [52, Definition 3.1]. That is, there exists constants $C_c > 0$ and $t_c > 0$ such that[5]

$$\int_0^{t_c} \phi(s)\phi^\top(s)ds \geq C_c I_p. \tag{4}$$

Proposition 1. Consider the NLPRE (1) with $\mathcal{G}(\theta)$ satisfying **Assumption A1** and $\phi(t)$ verifying **Assumption A2**. Define the LS+D interlaced estimator with time-varying forgetting factor

$$\dot{\hat{\eta}}(t) = \alpha F(t)\phi(t)[y(t) - \phi^\top(t)\hat{\eta}(t)],\ \hat{\eta}(0) =: \eta_0 \in \mathbb{R}^p \tag{5a}$$

$$\dot{F}(t) = -\alpha F(t)\phi(t)\phi^\top(t)F(t) + \beta(t)F(t),\ F(0) = \frac{1}{f_0}I_p \tag{5b}$$

$$\dot{\hat{\theta}}(t) = \gamma Q\Delta(t)[\mathcal{Y}(t) - \Delta(t)\mathcal{G}(\hat{\theta})],\ \hat{\theta}(0) =: \theta_0 \in \mathbb{R}^q \tag{5c}$$

$$\dot{z}(t) = -\beta(t)z(t),\ z(0) = 1, \tag{5d}$$

where we defined

$$\beta(t) := \beta_0\Big(1 - \frac{\|F(t)\|}{M}\Big) \tag{6a}$$

$$\Delta(t) := \det\{I_p - z(t)f_0F(t)\} \tag{6b}$$

$$\mathcal{Y}(t) := \operatorname{adj}\{I_p - z(t)f_0F(t)\}[\hat{\eta}(t) - z(t)f_0F(t)\eta_0], \tag{6c}$$

with tuning gains $\alpha > 0$, $f_0 > 0$, $\beta_0 > 0$, $M \geq \frac{1}{f_0}$ and $\gamma > 0$. Define the parameter estimation error $\tilde{\theta}(t) := \hat{\theta}(t) - \theta$. Then, for all $f_0 > 0$, $\eta_0 \in \mathbb{R}^p$ and $\theta_0 \in \mathbb{R}^q$, we have that

$$\lim_{t\to\infty} \tilde{\theta}(t) = 0,\ (exp), \tag{7}$$

with all signals *bounded.*

Proof. With some abuse of notation, define the signal

$$\tilde{\mathcal{G}}(t) := \hat{\eta}(t) - \mathcal{G}(\theta),$$

whose derivative is given by

$$\dot{\tilde{\mathcal{G}}}(t) = -\alpha F(t)\phi(t)\phi^\top(t)\tilde{\mathcal{G}}(t), \tag{8}$$

where we replaced (1) in (5a). Now, from

$$\frac{d}{dt}[F^{-1}(t)] = -F^{-1}(t)\dot{F}(t)F^{-1}(t),$$

we have that

$$\frac{d}{dt}[F^{-1}(t)] = -\beta(t)F^{-1}(t) + \alpha\phi(t)\phi^\top(t), \tag{9}$$

and consequently

$$\begin{aligned}\frac{d}{dt}[F^{-1}(t)\tilde{\mathcal{G}}(t)] &= F^{-1}(t)\dot{\tilde{\mathcal{G}}}(t) + \frac{d}{dt}[F^{-1}(t)]\tilde{\mathcal{G}}(t)\\ &= -\alpha\phi(t)\phi^\top(t)\tilde{\mathcal{G}}(t) + \alpha\phi(t)\phi^\top(t)\tilde{\mathcal{G}}(t)\\ &\quad - \beta(t)F^{-1}\tilde{\mathcal{G}}(t)\\ &= -\beta(t)F^{-1}(t)\tilde{\mathcal{G}}(t).\end{aligned}$$

This implies that

$$F^{-1}(t)\tilde{\mathcal{G}}(t) = z(t)f_0\tilde{\mathcal{G}}(0),\ \forall t \geq 0, \tag{10}$$

where we used the definition of the function $z(t)$ and the fact that $F^{-1}(0) = f_0 I_p$. The latter may be rewritten as[6]

$$\hat{\eta}(t) - \mathcal{G}(\theta) = z(t)f_0F(t)[\eta_0 - \mathcal{G}(\theta)],$$

from which we define the *extended NLPRE*

$$Y(t) = \Phi(t)\mathcal{G}(\theta) \tag{11a}$$

$$Y(t) := \hat{\eta}(t) - z(t)f_0F(t)\eta_0 \tag{11b}$$

$$\Phi(t) := I_p - z(t)f_0F(t). \tag{11c}$$

[5] In [52, Definition 3.1] there is an initial time in the integral that, for simplicity and without loss of generality, is taking here as zero.

[6] This key identity is given in [51, Equation (8.108)] for the case of LPRE.

Following the DREM procedure we multiply (11a) by $\text{adj}\{\Phi(t)\}$ to get the following NLPRE with scalar regressor

$$\mathcal{Y}(t) = \Delta(t)\mathcal{G}(\theta), \tag{12}$$

where we used (6b) and (6c). Replacing (12) in (5c) we get

$$\dot{\hat{\theta}}(t) = -\gamma\Delta^2(t)Q[\mathcal{G}(\hat{\theta}(t)) - \mathcal{G}(\theta)].$$

To analyse the stability of the latter system define the Lyapunov function candidate

$$V(\tilde{\theta}(t)) := \frac{1}{2\gamma}|\tilde{\theta}(t)|^2. \tag{13}$$

Computing its time derivative yields

$$\begin{aligned}\dot{V}(t) &= -\Delta^2(t)[\hat{\theta}(t) - \theta]^\top Q[\mathcal{G}(\hat{\theta}(t)) - \mathcal{G}(\theta)]\\ &\leq -\Delta^2(t)\rho|\tilde{\theta}(t)|^2\\ &= -2\rho\gamma\Delta^2(t)V(t),\end{aligned}$$

where we invoked the **Assumption A2** of strong monotonicity (3) of $Q\mathcal{G}(\theta)$ to get the first bound.

To complete the proof we first notice that in [51] it is shown that $F(t) \leq MI_p$ and that $\beta(t) \geq 0$—the latter implies that $z(t)$ is non-increasing and upper bounded by one. We consider two cases: when $z(t)F(t) \to 0$ and when $z(t)F(t) \geq \rho I_q > 0$. In the first case we notice that and if $z(t)F(t) \to 0$ then $\Delta(t) \to 1$ and, consequently, $\Delta(t)$ is PE. For the second case, we solve (9) to get

$$F^{-1}(t) = z(t)f_0 I_p + \alpha\int_0^t e^{-\int_\tau^t \beta(s)ds}\phi(\tau)\phi^\top(\tau)d\tau,\ \forall t \geq 0,$$

which may be rewritten as

$$\begin{aligned}I_p - z(t)f_0F(t) &= \alpha F(t)\int_0^t e^{-\int_\tau^t \beta(s)ds}\phi(\tau)\phi^\top(\tau)d\tau\\ &= \alpha F(t)z(t)\int_0^t \frac{1}{z(\tau)}\phi(\tau)\phi^\top(\tau)d\tau\\ &\geq \alpha F(t)z(t)\int_0^t \phi(\tau)\phi^\top(\tau)d\tau\\ &\geq \alpha\rho\int_0^t \phi(\tau)\phi^\top(\tau)d\tau,\ \forall t \geq 0.\end{aligned}$$

Now, from the implication

$$\phi(t) \in IE\ \Rightarrow\ \int_0^t \phi(\tau)\phi^\top(\tau)d\tau > 0,\ \forall t \geq t_c,$$

we conclude that the matrix $I_p - z(t)f_0F(t)$ is nonsingular for all $t \geq t_c$, which implies that $\Delta(t)$ is PE. This concludes the proof. □□□

Remark 1. Notice that, as indicated above, if $z(t) \to 0$ then $\Delta(t) \to 1$ and, in view of the extended NLPRE (11a), we have that $\mathcal{Y}(t) \to \mathcal{G}(\theta)$. Consequently the parameter update law (5c) will verify

$$\dot{\hat{\theta}}(t) \to \gamma Q[\mathcal{G}(\theta) - \mathcal{G}(\hat{\theta}(t))],$$

and the Lyapunov stability analysis still holds.

2.2. Linearly parameterized regression equations

In the next corollary we specialize the result of Proposition 1 for the case of LPRE. To streamline the presentation of the result we recall the following.

Definition 1. [16] The LPRE

$$y(t) = \phi^\top(t)\theta \tag{14}$$

where $y(t) \in \mathbb{R}$, $\phi(t) \in \mathbb{R}^q$ and $\theta \in \mathbb{R}^q$ is said to be *identifiable* if and only if there exists a set of time instants—$\{t_i\}_{i\in\bar{q}}$, $t_i \in \mathbb{R}_{>0}$, such that

$$\text{rank}\left\{\begin{bmatrix}\phi(t_1)|\phi(t_2)| & \cdots & |\phi(t_q)\end{bmatrix}\right\} = q.$$

For the sake of completeness we also recall the following result of [53]

Lemma 1. The LPRE (14) is identifiable *if and only if* the regressor vector ϕ is IE.

We are in position to present the main result of the subsection whose proof follows immediately from Lemma 1, the proof of Proposition 1 and [38, Proposition 2].

Corollary 1. Consider the LPRE (14) and assume it is *identifiable.* Define the LS+D interlaced estimator with time-varying forgetting factor

$$\begin{aligned}&\dot{\hat{\eta}}(t) = \alpha F(t)\phi(t)[y(t) - \phi^\top(t)\hat{\eta}(t)],\ \hat{\eta}(0) =: \eta_0 \in \mathbb{R}^p\\ &\dot{F}(t) = -\alpha F(t)\phi(t)\phi^\top(t)F(t) + \beta(t)F(t),\ F(0) = \frac{1}{f_0}I_p\\ &\dot{\hat{\theta}}(t) = \gamma\Delta(t)[\mathcal{Y}(t) - \Delta(t)\hat{\theta}],\ \hat{\theta}(0) =: \theta_0 \in \mathbb{R}^q\\ &\dot{z}(t) = -\beta(t)z(t),\ z(0) = 1,\end{aligned}$$

with tuning gains $\alpha > 0$, $f_0 > 0$, $\beta \geq 0$, $M \geq \frac{1}{f_0}$ and $\gamma > 0$, and we used the definitions (6). Then, for all $f_0 > 0$, $\eta_0 \in \mathbb{R}^q$ and $\theta_0 \in \mathbb{R}^q$, we have that (7) holds with all signals *bounded.* Moreover, the *individual* parameter errors verify the *monotonicity* condition

$$|\tilde{\theta}_i(t_a)| \leq |\tilde{\theta}_i(t_b)|,\ \forall\, t_a \geq t_b \geq 0,\ i \in \bar{q}.$$

3. Main Result for Discrete-time Systems

In this section we present the proposed LS+D estimator for DT systems. Similarly to the previous section, we consider first the case of NLPRE and then specialize to LPRE. As will be shown below, in the DT case we need an additional Lipschitz assumption on the mapping $\mathcal{G}(\theta)$, which is conspicuous by its absence in CT.

3.1. Nonlinearly parameterized regression equations

Consider the following DT NLPRE

$$y_k = \phi_k^\top\mathcal{G}(\theta) \tag{15}$$

where $y_k \in \mathbb{R}$, $\phi_k \in \mathbb{R}^p$ and $\mathcal{G} : \mathbb{R}^q \to \mathbb{R}^p$, $q \le p$. Assume $\mathcal{G}(\theta)$ verifies **Assumption A1** and the following.

Assumption A3. [Lipschitz] The mapping $\mathcal{G}(\theta)$ satisfies the *Lipschitz condition*

$$|\mathcal{G}(a) - \mathcal{G}(b)| \le \nu |a - b|, \ \forall a, b \in \mathbb{R}^q, \tag{16}$$

for some $\nu > 0$.

Moreover, assume the regressor ϕ_k satisfies.

Assumption A4. [Interval Excitation] [52, Definition 3.3] The regressor ϕ_k is IE. That is, there exists constants $C_d > 0$ and $k_c > 0$ such that

$$\Sigma_{j=0}^{k_c} \phi_j \phi_j^\top \ge C_d I_p. \tag{17}$$

Proposition 2. Consider the NLPRE (15) with $\mathcal{G}(\theta)$ satisfying **Assumption A1** and **A3** and ϕ_k verifying **Assumption A4**. Define the normalized LS+DREM interlaced estimator

$$\hat{\eta}_{k+1} = \hat{\eta}_k + \frac{1}{\beta + \phi_k^\top F_{k-1}\phi_k} F_{k-1}\phi_k(y_k - \phi_k^\top \hat{\eta}_k), \tag{18a}$$

$$F_k = \frac{1}{\beta}\left(I_p - \frac{1}{\beta + \phi_k^\top F_{k-1}\phi_k} F_{k-1}\phi_k\phi_k^\top\right) F_{k-1} \tag{18b}$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma Q \frac{\Delta_k}{1 + \Delta_k^2}[\mathcal{Y}_k - \Delta_k \mathcal{G}(\hat{\theta}_k)], \ \hat{\theta}_0 =: \theta_0 \in \mathbb{R}^q \tag{18c}$$

$$z_k = \beta z_{k-1}, \ z_{-1} = 1, \tag{18d}$$

with initial conditions $\hat{\eta}_0 =: \eta_0 \in \mathbb{R}^p$, $F_{-1} = \frac{1}{f_0} I_p$ and the definitions

$$\Delta_k := \det\{I_p - f_0 z_k F_{k-1}\} \tag{19a}$$

$$\mathcal{Y}_k := \operatorname{adj}\{I_p - f_0 z_k F_{k-1}\}(\hat{\eta}_k - f_0 z_k F_{k-1}\eta_0), \tag{19b}$$

with tuning parameters the initial condition $f_0 > 0$, the forgetting factor $\beta \in (0, 1]$ and the adaptation gain $\gamma > 0$, which is selected such that

$$\sigma := \rho - \frac{\gamma\nu^2}{2}\lambda_{\max}\{Q^\top Q\} > 0. \tag{20}$$

Define the parameter estimation error $\tilde{\theta}_k := \hat{\theta}_k - \theta$. Then, for all $f_0 > 0$, $\eta_0 \in \mathbb{R}^p$ and $\theta_0 \in \mathbb{R}^q$, we have that

$$\lim_{k\to\infty} |\tilde{\theta}_k| = 0, \ (exp), \tag{21}$$

with all signals *bounded.*

Proof. To simplify the notation we define the normalization sequence

$$m_k := \beta + \phi_k^\top F_{k-1}\phi_k. \tag{22}$$

With some abuse of notation, define the error signal

$$\tilde{\eta}_k := \hat{\eta}_k - \mathcal{G}(\theta), \tag{23}$$

whose dynamics is given by

$$\tilde{\eta}_{k+1} = \left(I_p - \frac{1}{m_k} F_{k-1}\phi_k\phi_k^\top\right)\tilde{\eta}_k, \tag{24}$$

where we used (15) and (18a). Now, direct application of the matrix inversion lemma to (18b) shows that,

$$F_k^{-1} = \beta F_{k-1}^{-1} + \phi_k \phi_k^\top. \tag{25}$$

Combining (24) and (25) we can prove the following fundamental property of LS

$$F_k^{-1}\tilde{\eta}_{k+1} = \beta F_{k-1}^{-1}\tilde{\eta}_k.$$

Solving this difference equation we get

$$\begin{aligned} F_{k-1}^{-1}\tilde{\eta}_k &= \beta^k F_{-1}^{-1}\tilde{\eta}_0 \\ &= z_k f_0 \tilde{\eta}_0, \ \forall k \ge 0, \end{aligned}$$

where we replaced the solution of (18d) and the initial condition choice $F_{-1} = \frac{1}{f_0} I_p$ to get the second identity. Using the definition (23), the equation above may be rewritten as the *extended LPRE*

$$(I_p - f_0 z_k F_{k-1})\mathcal{G}(\theta) = \hat{\eta}_k - f_0 z_k F_{k-1}\eta_0. \tag{26}$$

Following the DREM procedure we multiply (26) by $\operatorname{adj}\{I_p - f_0 z_k F_{k-1}\}$ to get the following NLPRE

$$\mathcal{Y}_k = \Delta_k \mathcal{G}(\theta), \tag{27}$$

where we used (19a) and (19b). Replacing (27) in (18c) we get the dynamics of the parameter error

$$\tilde{\theta}_{k+1} = \tilde{\theta}_k - \gamma \bar{\Delta}_k^2 Q[\mathcal{G}(\hat{\theta}_k) - \mathcal{G}(\theta)], \tag{28}$$

where, to simplify the notation, we defined the normalized scalar regressor sequence

$$\bar{\Delta}_k^2 := \frac{\Delta_k^2}{1 + \Delta_k^2} \le 1. \tag{29}$$

To analyze the stability of this equation define the Lyapunov function candidate

$$V_k = \frac{1}{2\gamma}|\tilde{\theta}_k|^2, \tag{30}$$

that satisfies

$$\begin{aligned} V_{k+1} =& V_k - \bar{\Delta}_k^2 \tilde{\theta}_k^\top Q[\mathcal{G}(\hat{\theta}_k) - \mathcal{G}(\theta)] \\ &+ \frac{\gamma}{2}\bar{\Delta}_k^4 [\mathcal{G}(\hat{\theta}_k) - \mathcal{G}(\theta)]^\top Q^\top Q[\mathcal{G}(\hat{\theta}_k) - \mathcal{G}(\theta)] \quad (31) \\ \le& V_k - \rho\bar{\Delta}_k^2|\tilde{\theta}_k|^2 + \frac{\gamma\nu^2}{2}\lambda_{\max}\{Q^\top Q\}\bar{\Delta}_k^4|\tilde{\theta}_k|^2 \\ \le& V_k - \left[\rho - \frac{\gamma\nu^2}{2}\lambda_{\max}\{Q^\top Q\}\right]\bar{\Delta}_k^2|\tilde{\theta}_k|^2 \quad (32) \\ =& V_k - \sigma\bar{\Delta}_k^2|\tilde{\theta}_k|^2. \end{aligned}$$

where we invoked **Assumption A2** and **Assumption A3** to get the first bound, (29) for the second one and used (20) in the last identity. Summing the inequality above we get

$$V_k - V_0 \le -\sum_{j=1}^{k} \sigma \bar{\Delta}_j^2 |\tilde{\theta}(t)_j|^2 \;\Rightarrow\; \frac{V_0}{\sigma} \ge \sum_{j=1}^{k} \bar{\Delta}_j^2 |\tilde{\theta}(t)_j|^2.$$

Taking the limit as $k \to \infty$ we conclude that $\Delta_k |\tilde{\theta}_k| \in \ell_2$, consequently

$$\lim_{k\to\infty} \bar{\Delta}_k |\tilde{\theta}_k| = 0. \tag{33}$$

Now, from the Algebraic Limit Theorem [46, Theorem 3.3] we know that the limit of the product of two convergent sequences is the product of their limits. On the other hand, from the fact that

$$V(k+1) \le V(k) \le V(0),\; \forall k \in \mathbb{Z}_{>0},$$

we have that $|\tilde{\theta}(k)|$ is a bounded monotonic sequence, hence it converges [46, Theorem 3.14]. Finally, if $\bar{\Delta}(k)$ converges to a *non-zero* limit, we conclude from (33) that $|\tilde{\theta}_k| \to 0$.

We will proceed now to prove that (17) of **Assumption A4** ensures this property of Δ_k, which together with the fact that if Δ_k converges to a non-zero limit, then $\bar{\Delta}_k$ also converges to a non-zero limit. Indeed, the solution of the difference equation (25) is given by

$$F_k^{-1} = \beta^{k+1} f_0 + \beta^k \sum_{j=0}^{k} \beta^{-j} \phi_j \phi_j^\top.$$

Evaluating this expression for $k = k_c$ yields

$$I_p - f_0 z_{k_c+1} F_{k_c} = \beta^{k_c} F_{k_c} \sum_{j=0}^{k_c} \beta^{-j} \phi_j \phi_j^\top.$$

The IE assumption ensures that the summation term is positive definite, since F_{k_c} is nonsingular this ensures that the matrix on the left hand side is nonsingular. The proof that this property holds for any $k > k_c$ stems from the observation that, for any $k_b > 0$ we have that

$$\sum_{j=0}^{k_c+k_b} \beta^{-j} \phi_j \phi_j^\top = \sum_{j=0}^{k_c} \beta^{-j} \phi_j \phi_j^\top + \sum_{j=k_c+1}^{k_c+k_b} \beta^{-j} \phi_j \phi_j^\top,$$

preserving the positivity property mentioned above. This completes the proof. □□□

3.2. Linearly parameterized regression equations

In this section, we use the result of Proposition 2 for the case of LPRE—obviously, in this linear case **Assumption A1** and **Assumption A3** are automatically satisfied. As a first step we recall [53] that Definition 1 and Lemma 1, given for continuous functions, are also valid for sequences.

As a second step notice that for the LPRE case (26) takes the form

$$(I_p - f_0 z_k F_{k-1})\theta = \hat{\eta}_k - f_0 z_k F_{k-1} \eta_0,$$

consequently (27) now becomes

$$\mathcal{Y}_i(k) = \Delta_k \theta_i,\; i \in \bar{q},$$

and the dynamics of the parameter error (28) is now given by

$$\tilde{\theta}_{k+1} = \tilde{\theta}_k - \gamma \bar{\Delta}_k^2 \tilde{\theta}_k,$$

whose stability is follows immediately from the IE assumption [53, Proposition 2] .

Corollary 2. Consider the LPRE and assume it is *identifiable*. Define the normalized LS+D interlaced estimator with forgetting factor

$$\begin{aligned}
\hat{\eta}_{k+1} &= \hat{\eta}_k + \frac{1}{\beta + \phi_k^\top F_{k-1} \phi_k} F_{k-1} \phi_k (y_k - \phi_k^\top \hat{\eta}_k),\\
F_k &= \frac{1}{\beta}\left(I_p - \frac{1}{\beta + \phi_k^\top F_{k-1}\phi_k} F_{k-1}\phi_k\phi_k^\top\right) F_{k-1}\\
\hat{\theta}_{k+1} &= \hat{\theta}_k + \gamma \frac{\Delta_k}{1+\Delta_k^2}[\mathcal{Y}_k - \Delta_k \hat{\theta}_k],\; \hat{\theta}_0 =: \theta_0 \in \mathbb{R}^q,
\end{aligned}$$

with initial conditions $\hat{\eta}_0 =: \eta_0 \in \mathbb{R}^q$, $F_{-1} = \frac{1}{f_0} I_p$, tuning gains $f_0 > 0$, $\beta \in (0,1]$ and $\gamma > 0$, and we used the definitions (19). Then, for all $f_0 > 0$, $\eta_0 \in \mathbb{R}^q$ and $\theta_0 \in \mathbb{R}^q$, we have that (33) holds with all signals *bounded*. Moreover, the *individual* parameter errors verify the monotonicity condition

$$|\tilde{\theta}_i(k_a)| \le |\tilde{\theta}_i(k_b)|,\; \forall\, k_a \ge k_b \ge 0,\; i \in \bar{q}.$$

Remark 2. The importance of the *element-by-element* monotonicity property of the parameter error can hardly be overestimated. It played a key role for the relaxation of the assumption of known sign of the high frequency in model reference adaptive control [14, 53] as well as in the solution of the adaptive pole placement problem [44].

3.3. Switching parameters case

In this section, we consider the case of switched parameters estimation. Whereas the results are presented in DT only, similar results can be formulated for the CT case in a straightforward manner. Rewrite (15) as

$$y_k = \phi_k^\top \mathcal{G}(\theta^*_{\sigma_k}), \tag{34}$$

where $\theta^*_{\sigma_k}$ denotes the switched unknown parameter vector with $\theta^*_{\sigma_k} \in \{\theta_1^*, \theta_2^*, \ldots, \theta_s^*\}$, $s \in \mathbb{Z}_{>0}$. The switching signal $\sigma_k : \mathbb{Z}_{\ge 0} \to \bar{s}$ is a *known*[7] piecewise-constant function defining the behavior of $\theta^*_{\sigma_k}$, i.e., $\theta^*_{\sigma_k} = \theta_i^*$ when $\sigma_k = i$,

[7] Such a scenario arises in several practical control scenarios, when the known switching signal σ_k characterizes known changes in operation regimes [23].

$i \in \bar{s}$. The known time instants when σ_k changes its value are further denoted as $t_{r,i}$, $i \in \mathbb{Z}_{\geq 0}$.

The estimator (18), (19) is not capable of estimating switched parameters as for $\beta < 1$ the sequence z_k converges to zero, and for $\beta = 1$ the LS estimator looses its alertness. To deal with switching parameters, we propose a resetting-based modification of the estimator (18), (19):

$$\hat{\eta}_{k+1} = \hat{\eta}_k + \frac{1}{1+\phi_k^\top F_{k-1}\phi_k} F_{k-1}\phi_k(y_k - \phi_k^\top \hat{\eta}_k), \quad (35a)$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \gamma Q \frac{\Delta_k}{1+\Delta_k^2}[\mathcal{Y}_k - \Delta_k \mathcal{G}(\hat{\theta}_k)],\ \hat{\theta}_0 =: \theta_0 \in \mathbb{R}^q, \quad (35b)$$

$$F_k = \begin{cases} \mathcal{N}_k F_{k-1} & \text{if } t_k \neq t_{r,i}\ \forall i, \\ \frac{1}{f_0} I_p & \text{otherwise,} \end{cases} \quad (35c)$$

$$\mathcal{N}_k = \left(I_p - \frac{1}{1+\phi_k^\top F_{k-1}\phi_k} F_{k-1}\phi_k\phi_k^\top \right),$$

$$\psi_{k+1} = \begin{cases} \psi_k & \text{if } t_k \neq t_{r,i}\ \forall i, \\ \hat{\eta}_{k+1} & \text{otherwise,} \end{cases}, \quad \psi_0 = \eta_0, \quad (35d)$$

where $\hat{\eta}_0 =: \eta_0 \in \mathbb{R}^p$, $F_{-1} = \frac{1}{f_0} I_p$, and

$$\Delta_k = \det\{I_p - f_0 F_{k-1}\}, \quad (36a)$$

$$\mathcal{Y}_k := \text{adj}\{I_p - f_0 F_{k-1}\}(\hat{\eta}_k - f_0 F_{k-1}\psi_k). \quad (36b)$$

Between the reseting instances $t_{r,i}$, the estimator (35), (36) reproduces the estimator (18), (19) with $\beta = 1$ and thus $z_k \equiv 1$. Then, at each reset instance $t_{r,i}$, the matrix F_k is reset to its initial condition F_{-1}, and the state ψ_k saves the value of $\hat{\eta}_k$. The state ψ_k thus plays the same role as η_0 in (26), compare (36b) and (19b). Following the properties of (18), (19), the proposed estimator ensures the boundedness of the states and is capable of estimating $\theta^*_{\sigma_k}$ if the following assumption holds.

Assumption A5. [Switching Interval Excitation]. The switching signal σ_k is such that the regressor ϕ_k is IE between two subsequent switching instants. That is, there exist constants $C_d > 0$ and $k_c > 0$ such that for any $i \in \mathbb{Z}_{\geq 0}$

$$t_{r,i} + k_c \leq t_{r,i+1}$$

and

$$\sum_{\ell=0}^{k_c} \phi_{t_{r,i}+\ell}\phi_{t_{r,i}+\ell}^\top \geq C_d I_p.$$

Remark 3. In words, **Assumption A5** means that the regressor satisfies the IE condition inside each subinterval $[t_{r,i}, t_{r,i+1}]$. For simplicity we have taken that the constants k_c and C_d that appear in the definition of IE are the *same* for all subintervals $[t_{r,i}, t_{r,i+1}]$, but this is clearly not necessary.

4. Derivation of the Extended NLPRE (11) via DRE

To simplify the reading of the material presented in this section we refer the reader to Appendix A where the procedure to derive DREM is recalled.

In Proposition 1 it is shown that the dynamic extension (5a) and (5b) generates the extended NLPRE (11) to which we apply the mixing step **S4** of Appendix A to generate the scalar NLPRE (12). In this section we prove that this extended NLPRE can also be derived directly applying the DREM step **S2** of Appendix A for a suitably defined LTV operator $\mathcal{H}$.[8] For the sake of brevity we only consider the CT case, with the DT case following *verbatim*.

Proposition 3. Define the state space realization of the LTV operator $\mathcal{H} : u \to U$ used in step **S2** of Appendix A as in (.1) with

$$A(t) := -\alpha F(t)\phi(t)\phi^\top(t),\ b(t) := \alpha F(t)\phi(t),$$

with $F(t)$ defined in (5b). Starting from the NLPRE $y(t) = \phi^\top(t)\mathcal{G}(\theta)$, construct $Y(t) \in \mathbb{R}^p$ and $\Phi(t) \in \mathbb{R}^{p\times p}$ via (.3) that is, as the solutions of the dynamic extension

$$\dot{Y}(t) = -\alpha F(t)\phi(t)\phi^\top(t)Y(t) + \alpha F(t)\phi(t)y(t), \quad (37a)$$

$$\dot{\Phi}(t) = -\alpha F(t)\phi(t)\phi^\top(t)\Phi(t) + \alpha F(t)\phi(t)\phi^\top(t), \quad (37b)$$

and initial conditions $Y(0) = \mathbf{0}_{p\times 1}$ and $\Phi(0) = \mathbf{0}_{p\times p}$.

i) The extended NLPRE $Y(t) = \Phi(t)\mathcal{G}(\theta)$ holds.

ii) The signals $Y(t)$ and $\Phi(t)$ satisfy (11b) and (11c), respectively, with $\hat{\eta}(t) \in \mathbb{R}^p$ and $F(t) \in \mathbb{R}^{p\times p}$ solutions of the differential equations (5a) and (5b), respectively.

Proof. The fact that the extended NLPRE $Y(t) = \Phi(t)\mathcal{G}(\theta)$ holds follows trivially from linearity of the operator $\mathcal{H}$.

To prove the claim (ii) we invoke (11b) and do the following calculations

$$\begin{aligned}\dot{Y}(t) &= \dot{\hat{\eta}}(t) - z(t)\dot{F}(t)\eta_0 - \dot{z}(t)f_0F(t)\eta_0 \\ &= \alpha F(t)\phi(t)[y(t) - \phi^\top(t)\hat{\eta}(t)] + \beta(t)z(t)f_0F(t)\eta_0 \\ &\quad + z(t)[\alpha F(t)\phi(t)\phi^\top(t)F(t) - \beta(t)F(t)]\eta_0 \\ &= \alpha F(t)\phi(t)y(t) - \alpha F(t)\phi(t)\phi^\top(t)[\hat{\eta}(t) - z(t)f_oF(t)\eta_0] \\ &= A(t)Y(t) + b(t)y(t).\end{aligned}$$

In the same spirit as above we compute the time derivative of $\Phi(t)$ as defined in (11c) to get

$$\begin{aligned}\dot{\Phi}(t) &= -\dot{z}(t)f_oF(t) - z(t)f_0\dot{F}(t) \\ &= \alpha z(t)f_oF(t)\phi(t)\phi^\top(t)F(t) \\ &= \alpha F(t)\phi(t)\phi^\top(t) - \alpha F(t)\phi(t)\phi^\top(t)[I_q - z(t)f_oF(t)] \\ &= A(t)\Phi(t) + b(t)\phi^\top(t).\end{aligned}$$

This completes the proof. □□□

[8] We refer the interested reader to [53, Proposition 3] where the DREM operator $\mathcal{H}$ for the G+D estimator reported in [53, Proposition 2] is identified.

Remark 4. It is important to note that the relation $Y(t) = \Phi(t)\mathcal{G}(\theta)$ imposes the constraint $Y(0) = \Phi(0)\mathcal{G}(\theta)$, which is satisfied with the zero initial conditions imposed in Proposition 3. As expected, this choice is consistent with the choice of initial conditions for $\hat{\eta}(t)$ and $F(t)$ given in Proposition 1.

Remark 5. The dynamic extension (5b)and (37) provides an alternative to the construction of the proposed estimator. The relationship between the two implementations boils down to a standard diffeomorphic change of coordinates. Indeed, while the state of the system in (5) and (5d) is given by $\mathrm{col}(\hat{\eta}(t), \mathrm{vec}(F(t)), z(t), \hat{\theta}(t)) \in \mathbb{R}^{(p+p^2+1+q)}$, the state of the system of Proposition 3 is $\mathrm{col}(Y(t), \mathrm{vec}(\Phi(t)), z(t), \hat{\theta}(t)) \in \mathbb{R}^{(p+p^2+1+q)}$, and the first two components are related by a simple invertible coordinate change

$$\begin{bmatrix} \hat{\eta}(t) \\ \mathrm{vec}(F(t)) \end{bmatrix} = \begin{bmatrix} Y(t) + [I_p - \Phi(t)]\eta_0 \\ \frac{1}{z(t)f_0}\mathrm{vec}(I_p - \Phi(t)) \end{bmatrix}.$$

However, the original implementation (5) clearly reveals the mechanism underlying the operation of the estimator, namely, the use of a classical LS update and the creation of the extended NLPRE exploiting the well-known property of LS (10).[9]

5. Robustness Analysis of the CT LS+D Estimator

In this section we analyze the robustness *vis-à-vis* additive perturbations of the CT LS+D estimator of Proposition 1. That is, we consider the *perturbed* NLPRE

$$y(t) = \phi^\top(t)\mathcal{G}(\theta) + d(t), \tag{38}$$

where $d(t)$ represents an additive perturbation signal. This signal may come from additive noise in the measurements of $y(t)$ and $\phi(t)$ or time variations of the parameters, that is, $d(t)$ may be decomposed as

$$d(t) = d_y(t) + d_\theta^\top(t)\phi(t) + d_\phi^\top(t)\mathcal{G}(\theta),$$

where $d_y(t) \in \mathbb{R}$ and $d_\phi(t) \in \mathbb{R}^p$ represent the measurement noise added to $y(t)$ and $\phi(t)$, respectively, and $d_\theta(t) \in \mathbb{R}^p$ captures *time variations* in the parameters. We make the reasonable assumption that these signals are all bounded and prove that the CT LS+D estimator defines a *bounded-input-bounded-state* (BIBS) stable system.

The main result is summarized in the proposition below.

Proposition 4. Consider the *perturbed* NLPLPRE (38) with $d(t)$ a bounded signal. Assume the regressor $\phi(t)$ is IE. The LS+D estimator of Proposition 1 applied to this NLPRE is *BIBS stable.*

[9] To the best of the authors' knowledge, this property was first reported in [11, equation (17)] and was widely used for the implementation of projections in indirect adaptive controllers [27].

Proof. In the light of Remark 3, to carry out the proof we rely on the use of the alternative implementation of the extended NLPRE of Proposition 3. Applying the operator $\mathcal{H}$ of Propositions 3 to the perturbed NLPRE (38) yields the perturbed version of the extended LPRE (11a) as

$$Y(t) = \Phi(t)\mathcal{G}(\theta) + \mathcal{H}[d](t), \tag{39}$$

where we exploited the property of linearity of $\mathcal{H}$. Next we proceed to show that the operator $\mathcal{H}$ is BIBO-stable. This is done by proving that, for all bounded $d(t)$, the signal $\mathcal{H}[d](t)$ is also bounded.

The signal $\mathcal{H}[d](t)$ is generated via the CT LTV system

$$\begin{aligned} \dot{x}_d(t) &= -\alpha F\phi(t)\phi^\top(t)x_d(t) + \alpha F(t)\phi(t)d(t) \\ \mathcal{H}[d](t) &= x_d(t). \end{aligned}$$

Defining $W(x_d) := \frac{1}{2}x_d^\top(t)F^{-1}(t)x_d(t)$, we have

$$\begin{aligned} \dot{W} &= -\beta(t)W(t) - \frac{\alpha}{2}[\phi^\top(t)x_d(t) - d(t)]^2 + \frac{\alpha}{2}d^2(t) \\ &\leq -\beta(t)W(t) + \frac{\alpha}{2}d^2(t). \end{aligned}$$

As $\beta(t) > 0$ and $d(t)$ is bounded, this proves that $x_d(t) = \mathcal{H}[d](t)$, is also bounded.

From the analysis above, we conclude that the operator $\mathcal{H}$ is BIBO-stable. Consequently, since $\phi(t)$ and $y(t)$ are bounded, it follows that $Y(t) = \mathcal{H}[y](t)$ and $\Phi(t) = \mathcal{H}[\phi^\top](t)$ are also bounded. It only remains to prove that and $\hat{\theta}(t)$ is bounded. Whence, multiplying (39) by $\mathrm{adj}\{\Phi(t)\}$ we get the following perturbed NLPRE

$$\mathcal{Y}(t) = \Delta(t)\mathcal{G}(\theta) + \xi(t), \tag{40}$$

where we defined the signal

$$\xi(t) := \mathrm{adj}\{\Phi(t)\}x_d(t). \tag{41}$$

We notice that this signal is bounded. Replacing (41) in the estimator (5c) yields

$$\dot{\hat{\theta}}(t) = -\gamma\Delta^2(t)Q[\mathcal{G}(\hat{\theta}(t)) - \mathcal{G}(\theta)] + \gamma\Delta(t)Q\xi(t).$$

Computing the derivative of the Lyapunov function candidate (13) we get

$$\begin{aligned} \dot{V}(t) &= -\Delta^2(t)[\hat{\theta}(t) - \theta]^\top Q[\mathcal{G}(\hat{\theta}(t)) - \mathcal{G}(\theta)] \\ &\quad + \Delta(t)\tilde{\theta}^\top(t)Q\xi(t) \\ &\leq -\rho\Delta^2(t)|\tilde{\theta}(t)|^2 + \Delta(t)|\tilde{\theta}(t)||Q\xi(t)| \\ &= -\frac{\rho}{2}\Delta^2(t)|\tilde{\theta}(t)|^2 - \frac{\rho}{2}[\Delta(t)|\tilde{\theta}(t)| - \frac{1}{\rho}|Q\xi(t)|]^2 \\ &\quad + \frac{1}{2\rho}|Q\xi(t)|^2 \\ &\leq -\rho\gamma\Delta^2(t)V(t) + \frac{1}{2\rho}|Q\xi(t)|^2. \end{aligned}$$

The proof of boundedness of $\tilde{\theta}(t)$ is completed recalling that in Proposition 1 it is shown that $\Delta(t)$ is PE. □□□

6. Simulation Examples

In this section we present simulations of the proposed CT and DT estimators using different examples recently reported in the literature.

6.1. Example 5 of [29]

Consider the second order stable, CT, linear system described by

$$\begin{aligned}\dot{x}_1(t) =& x_2(t)\\ \dot{x}_2(t) =& -\theta_1 x_1(t) - \theta_2 + \theta_3 u(t)\\ y(t) =& x_1(t),\end{aligned}$$

or equivalently

$$\ddot{x}_1(t) = -\theta_1 x_1(t) - \theta_2 \dot{x}_1(t) + \theta_3 u(t) \tag{42}$$

where θ_1, θ_2 and θ_3 are unknown parameters. Applying the filter

$$H(\mathfrak{p}) = \frac{1}{\mathfrak{p}+\lambda}$$

where $\mathfrak{p} := \frac{d}{dt}$, to both sides of (42) and rearranging the terms, we get the LPRE (14) with

$$y(t) = \mathfrak{p}H(\mathfrak{p})[x_2](t), \quad \phi(t) = H(\mathfrak{p})[\mathrm{col}(-x_1(t),\, -\mathfrak{p}x_1(t),\, u(t))], \tag{43}$$

and $\theta := \mathrm{col}(\theta_1,\, \theta_2,\, \theta_3)$.

To carry out the simulations we use the same conditions that [29], that is, we set to zero the initial conditions of the filters, as well as the initial value of the parameter estimation vector $\hat{\theta}(0) = 0$, $\hat{\eta}(0) = \mathrm{col}(0.1,\, 0.1,\, 0.1)$, $u(t) = 5$ and fix $\theta = \mathrm{col}(2, 3, 1)$. Besides, the tuning parameters of the proposed estimator of Corollary 1 were $\alpha = 20.3$, $f_0 = 4$, $\beta = 0.07$ and $\gamma = 700$. In Fig. 1 we appreciate the transient behavior of the estimated parameters, which clearly shows the estimation of the real values. This result should be contrasted with the non-converging behavior of the estimates reported in [29] with the gradient scheme and their modified gradient.

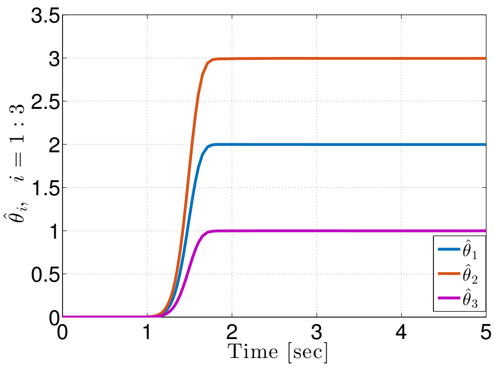


Figure 1: Transient behavior of the estimated parameters $\hat{\theta}_i(t)$ with $i = 1, 2, 3$.

To illustrate the use of the NLPRE (1), we notice that from the proposed values for θ, we have that θ_3 can be rewritten as $\theta_3 = \theta_2 - \theta_1$. Hence, after the application of the filter $H(\mathfrak{p})$, the system (42) can be written as the NLPRE (1) with $\mathcal{G}(\theta) := \mathrm{col}(\theta_1,\, \theta_2,\, \theta_2 - \theta_1)$, Thus, using the same initial conditions, estimator gains and verifying **Assumption A1** with

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

and $\rho = 1$, we carry out a simulation to estimate only θ_1 and θ_2 with the estimator of Proposition 1. Fig. 2 shows the transient behavior of the estimated parameters, showing again parameter convergence.

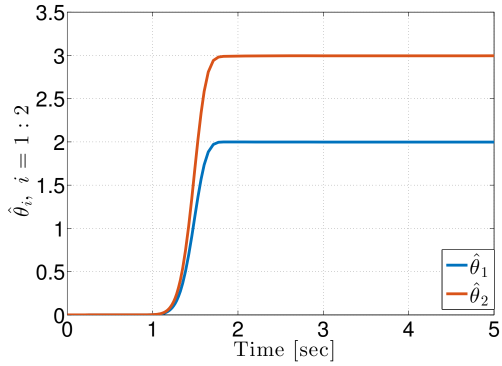


Figure 2: Transient behavior of the estimated parameters $\hat{\theta}_i(t)$ with $i = 1, 2$ using the NLPRE (1).

6.2. Example 4 of [15]

Consider the first order linear system

$$y_k = \frac{\theta_2}{\mathfrak{q} - \theta_1} u_k, \tag{44}$$

where $\mathfrak{q}$ is the forward-shiff operator and θ_1 and θ_2 are unknown parameters. After some simple calculations, we have that (44) can be written as a LPRE $y_k = \phi_k^\top \theta$ with

$$\phi_k = \mathrm{col}(y_{k-1},\, u_{k-1}), \quad \theta = \mathrm{col}(\theta_1,\, \theta_2).$$

To carry out the simulations we have also used the same initial conditions and parameters of [15], that is, $\theta_1 = 0.4$, $\theta_2 = 0.8$, $\theta_0 = \mathrm{col}(0,\, 0)$ and the input signal $u_k = 1$.[10] The tuning gains of the estimator of Corollary 2 were chosen as $\beta = 1$, $f_0 = 0.14$, $\gamma = 0.4$ and initial conditions $\hat{\eta}_0 = \mathrm{col}(1,\, 1)$. It is important to note that for this system (44), the estimator proposed in [15] only ensures the boundedness of $\hat{\theta}_i$, with $i = 1, 2$ (see Fig. 3 of [15]). This should be contrasted with our estimator, which, as can be seen in Fig. 3, converges to the real value.

[10] We notice that there is an unfortunate typo in the definition of u_k in [15, Example 4].

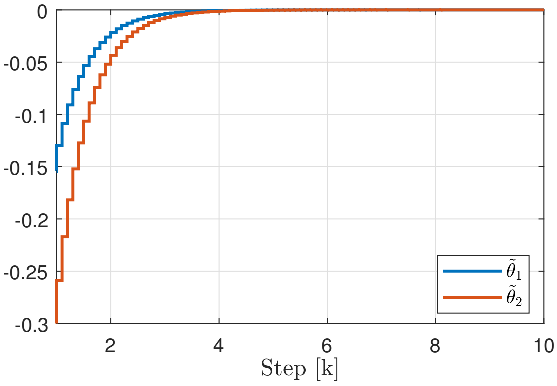


Figure 3: Transient behavior of the parameter estimates $\hat{\theta}_i$ with $i = 1, 2$.

6.3. *Example 8 of [30]*

We consider the DT system

$$y_k = -0.5y_{k-1} + 0.1y_{k-2} + u_{k-1} - 0.4u_{k-2},$$

which switches for $k \geq k_c$ to

$$y_k = 1.4y_{k-1} - 0.3y_{k-2} + u_{k-1} - 1.3u_{k-2}.$$

Note that for $k \geq k_c$ the plant is unstable and not minimum-phase. The initial conditions are $y_{-1} = -0.2$, $y_{-2} = 0.4$, and $u_{-1} = u_{-2} = 0$, and $k_c = 50$.

The system can be written in the form (34) setting $\mathcal{G}(\theta^*_{\sigma_k}) = \theta^*_{\sigma_k}$,

$$\begin{aligned}\phi_k &:= \begin{bmatrix}-y_{k-1} & -y_{k-2} & u_{k-1} & u_{k-2}\end{bmatrix}^\top,\\ \theta^*_1 &:= \begin{bmatrix}0.5 & -0.1 & 1 & -0.4\end{bmatrix}^\top,\\ \theta^*_2 &:= \begin{bmatrix}-1.4 & 0.3 & 1 & -1.3\end{bmatrix}^\top,\end{aligned}$$

and

$$\sigma_k = \begin{cases}1 & \text{for } k < k_c,\\ 2 & \text{for } k \geq k_c,\end{cases}$$

which corresponds to $t_{r,0} = 0$ and $t_{r,1} = k_c$.

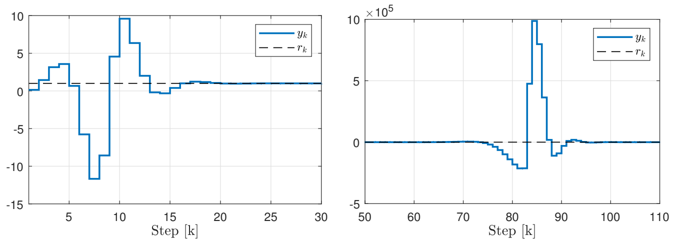
In this example, we consider the indirect adaptive poles placement for the reference tracking, where the reference signal is denoted as r_k. Then the control signal u is given by

$$u_k = c_{1,k}y_k + c_{2,k}y_{k-1} + c_{3,k}u_{k-1} + c_{4,k}r_k,$$

where the time-varying coefficients $c_{1,k}$, $c_{2,k}$, $c_{3,k}$, and $c_{4,k}$ are computed based on the current parameter estimate $\hat{\theta}_k$ to provide the desired poles and unit gain of the closed-loop system; if for a value of $\hat{\theta}_k$ the computations are ill-conditioned, then $u_k = 0$ is chosen. For this example, the desired poles for this are e^{-1}, $e^{-0.5+0.86\sqrt{-1}}$, and $e^{-0.5-0.86\sqrt{-1}}$, and the reference signal is $r_k \equiv 1$.

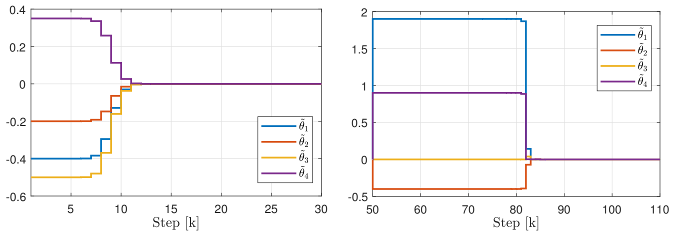
To estimate the parameters, we apply the resetting-based estimator (35), (36), where we set $f_0 = 0.4$, $\gamma = 500$, $\eta_0 = 0$, and $\theta_0 = \begin{bmatrix}0.1 & -0.3 & 0.5 & -0.05\end{bmatrix}$. Note that θ_0 cannot be chosen zero as such a choice yields zero input to the system and the regressor ϕ is not IE; for a nonzero choice of θ_0, the interval excitation is provided by the transients of the plant.

The simulation results are depicted in Fig. ?? for the output signal y and in Fig. 5 for the estimation errors $\tilde{\theta}_k = \hat{\theta}_k - \theta^*_{\sigma_k}$. It can be observed that after the switch, parameters estimation errors remain almost constant for approximately 30 steps, and then quickly converge. Further investigation shows that the regressor ϕ is not exciting on this initial interval, and thus the estimation does not progress. As soon as the IE condition is satisfied, the estimates $\hat{\theta}_k$ converges to the true value θ^*_2.



(a) Initial transients **(b)** Transients after the switch, $k_c = 50$

Figure 4: Transients of y_k.



(a) Initial transients **(b)** Transients after the switch, $k_c = 50$

Figure 5: Transients of $\tilde{\theta}$.

7. Concluding Remarks

We have presented in this paper a new robust DREM-based parameter estimator that proves global exponential convergence of the parameter errors with the weakest excitation assumption, namely, identifiability of the LPRE—which is, actually *necessary* for the off- or on-line estimation of the parameters. The main features of the estimator are: (i) it relies on the use of a high performance LS search, in contrast to the usually slower gradient descents; (ii) it ensures component-wise monotonicity of the parameter estimation errors; (iii) it incorporates a forgetting factor avoiding the well-known covariance wind-up problem of LS; (iv) it is applicable to NLPRE, which are separable and monotonic as well as to switching parameters; (v) it constructs the extended regressor avoiding the use of the computationally demanding GPEBO technique, exploiting instead the key structural property of the LS estimator captured in (10); and (vi) CT and DT implementations of the estimator are given. Several simulation results, borrowed from the literature, show the superior performance of the proposed estimator.

Acknowledgments

The first author is grateful to Dr. Michelangelo Bin for a detailed explanation of his important results reported in [5] and to Prof. Han-Fu Chen for bringing to his attention the use of the mixing step of DREM in [8]. He also thanks Dr. Lei Wang for his help in the derivation of the CT version of the LS+D with a forgetting factor.

References

[1] S. Aranovskiy, A. Bobtsov, R. Ortega and A. Pyrkin, Performance enhancement of parameter estimators via dynamic regressor extension and mixing, *IEEE Trans. Automatic Control*, vol. 62, pp. 3546-3550, 2017.

[2] A. Astolfi, D. Karagiannis and R. Ortega, *Nonlinear and Adaptive Control with Applications*, vol. 187, Springer, London, 2008.

[3] N. E. Barabanov and R. Ortega, On global asymptotic stability of $\dot{x} = \phi(t)\phi^\top(t)x$ with $\phi(t)$ bounded and not persistently exciting, *Systems and Control Letters*, vol. 109, pp. 24-27, 2017.

[4] A. Belov, R. Ortega and A. Bobtsov, Guaranteed performance adaptive identification scheme of discrete-time systems using dynamic regressor extension and mixing, *18th IFAC Symposium on System Identification, (SYSID 2018)*, Stockholm, Sweden, July 9-11, 2018.

[5] M. Bin, Generalized recursive least squares: Stability, robustness, and excitation, *Systems & Control Letters*, vol. 161, 105144, 2022.

[6] A. Bobtsov, B. Yi, R. Ortega and A. Astolfi, Generation of new exciting regressors for consistent on-line estimation of a scalar parameter, *IEEE Trans. Automatic Control*, `DOI: 10.1109/TAC.2022.3159568`, 2022.

[7] A. L. Bruce, A. Goel and D. S. Bernstein, Necessary and sufficient regressor conditions for the global asymptotic stability of recursive least squares, *Systems & Control Letters*, vol. 157, 10500527, 2021.

[8] H. F. Chen and W. Zhao, *Recursive Identification and Parameter Estimation*, CRC Press, 2014.

[9] G. Chowdhary, T. Yucelen, M. Muhlegg and E. N. Johnson, Concurrent learning adaptive control of linear systems with exponentially convergent bounds, *International Journal of Adaptive Control and Signal Processing*, vol. 27, no. 4, pp. 280-301, 2013.

[10] Y. Cui, J. E. Gaudio and A. M. Annaswamy, A new algorithm for discrete-time parameter estimation, `arXiv:2103.16653`, 2021.

[11] Ph. de Larminat, On the stabilizability condition in indirect adaptive control, *Automatica*, vol. 20, no, 6, pp. 793-795, 1984.

[12] B. P. Demidovich, Dissipativity of nonlinear systems of differential equations, *Vestnik Moscow State University, Ser. Mat. Mekh., Part I-6*, (1961) pp. 19-27; *Part II-1*, (1962), pp. 3-8, (in Russian).

[13] D. Efimov, N. Barabanov and R. Ortega, Robustness of linear time-varying systems with relaxed excitation, *Int. J. on Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1885-1900, 2019.

[14] D. Gerasimov, R. Ortega and V. Nikiforov, Adaptive control of multivariable systems with reduced knowledge of high frequency gain: Application of dynamic regressor extension and mixing estimators, *18th IFAC Symposium on System Identification, (SYSID 2018)*, Stockholm, Sweden, July 9-11, 2018.

[15] A. Goel, A. L. Bruce and D. S. Bernstein, Recursive least squares with variable-direction forgetting: Compensating for the loss of persistency, *IEEE Control Systems Magazine*, vol. 40, no, 4, pp. 80-102, 2020.

[16] G. Goodwin and K. Sin, *Adaptive Filtering Prediction and Control*, Prentice-Hall, 1984.

[17] M. Korotina, S. Aranovskiy, R. Ushirobira and A. Vedyakov, On parameter tuning and convergence properties of the DREM procedure, *2020 European Control Conference (ECC20)*, Saint Petersburg, Russia, May 12-15, 2020.

[18] M. Korotina, J. G. Romero, S. Aranovskiy, A. Bobtsov and R. Ortega, Persistent excitation is unnecessary for on-line exponential parameter estimation: a new algorithm that overcomes this obstacle, *Systems & Control Letters*, vol. 159, `doi.org/10.1016/j.sysconle.2021.105079`, 2022.

[19] J. Krause and P. Khargonekar, Parameter information content of measurable signals in direct adaptive control, *IEEE Trans. on Automatic Control*, vol. 32, no. 9, pp. 802-810. 1987.

[20] G. Kreisselmeier and G. Rietze-Augst, Richness and excitation on an interval—with application to continuous-time adaptive control, *IEEE Trans. Automatic Control*, vol. 35, no. 2, pp. 165-171, 1990.

[21] G. Kreisselmeier, Adaptive observers with exponential rate of convergence, *IEEE Trans. Automatic Control*, vol. 22, no. 1, pp. 2-8, 1977.

[22] M. Krstic, On using least-squares updates without regressor filtering in identification and adaptive control of nonlinear systems, *Automatica*, vol. 45, pp. 731-735, 2009.

[23] D. Liberzon, *Switching in Systems and Control*, vol 190, Springer, 2003.

[24] P.M. Lion, Rapid identification of linear and nonlinear systems, *AIAA Journal*, vol. 5, pp. 1835-1842, 1967.

[25] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, New Jersey, 1987.

[26] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, MIT Press, 1983.

[27] R. Lozano and X. Zhao, Adaptive pole placement without excitation probing signals, *IEEE Trans. Automatic Control*, vol. 39, no. 1, pp. 47-58, 1994.

[28] R. Lozano and C. Canudas, Passivity-based adaptive control of mechanical manipulators using LS-type estimation, *IEEE Trans. Automatic Control*, vol. 35, no. 12, pp. 1363-1365, 1990.

[29] R. Marino and P. Tomei, On exponentially convergent parameter estimation with lack of persistency of excitation, *Systems & Control Letters*, vol. 159, 105080, 2022.

[30] T. Nguyen, S. Islam, D. Bernstein and I. Kolmanovsky, Predictive cost adaptive control: A numerical investigation of persistency, consistency, and exigency, *IEEE Control Systems Magazine*, vol. 41, pp. 64-96, 2021.

[31] R. Ortega, An on-line least-squares parameter estimator with finite convergence time, *Proc. IEEE*, vol. 76, no. 7, 1988.

[32] R. Ortega, A. Bobtsov, A. Pyrkin and S. Aranovskiy, A parameter estimation approach to state observation of nonlinear systems, *Systems and Control Letters*, vol. 85, pp 84-94, 2015.

[33] R. Ortega, D. Gerasimov, N. Barabanov and V. Nikiforov, Adaptive control of linear multivariable systems using dynamic regressor extension and mixing estimators: Removing the high-frequency gain assumption, *Automatica*, vol. 110, 108589, 2019.

[34] R. Ortega, V. Nikiforov and D. Gerasimov, On modified parameter estimators for identification and adaptive control: a unified framework and some new schemes, *Annual Reviews in Control*, vol. 50, pp. 278-293, 2020.

[35] R. Ortega, A. Bobtsov, N. Nikolayev, J. Schiffer and D. Dochain, Generalized parameter estimation-based observers: Application to power systems and chemical-biological reactors, *Automatica*, vol. 129, 109635, 2021.

[36] R. Ortega, V. Gromov, E. Nuño, A. Pyrkin and J. G. Romero, Parameter estimation of nonlinearly parameterized regressions: application to system identification and adaptive control, *Automatica*, vol. 127, 109544, 2021.

[37] R. Ortega, Comments on recent claims about trajectories of con-

trol systems valid for particular initial conditions, *Asian Journal of Control*, DOI: 10.1002/asjc.2512, 2021.
[38] R. Ortega, S. Aranovskiy, A. Pyrkin, A Astolfi and A. Bobtsov, New results on parameter estimation via dynamic regressor extension and mixing: Continuous and discrete-time cases, *IEEE Trans. Automatic Control*, vol. 66, no. 5, pp. 2265-2272, 2021.
[39] R. Ortega, A. Bobtsov and N. Nikolayev, Parameter identification with finite-convergence time alertness preservation, *IEEE Control Systems Letters*, vol. 6, pp. 205-210, 2022
[40] Y. Pan and H. Yu, Composite learning robot control with guaranteed parameter convergence, *Automatica*, vol. 89, pp. 398-406, 2018.
[41] Y. Pan, S. Aranovskiy, A. Bobtsov, and H. Yu, Efficient learning from adaptive control under sufficient excitation, *International Journal of Robust and Nonlinear Control*, vol. 29, pp. 3111-3124, 2019.
[42] A. Pavlov, A. Pogromsky, N. van de Wouw and H. Nijmeijer, Convergence dynamics, a tribute to Boris Pavlovich Demidovich, *Systems & Control Letters*, vol. 52, pp. 257-261, 2004.
[43] L. Praly, Convergence of the gradient algorithm for linear regression models in the continuous and discrete-time cases, *Int. Rep. MINES ParisTech*, Centre Automatique et Systèmes, December 26, 2017.
[44] A. Pyrkin, R. Ortega, V. Gromov, A. Bobtsov and A. Vedyakov, A Globally convergent direct adaptive pole-placement controller for nonminimum phase systems with relaxed excitation assumptions, *Int. J. on Adaptive Control and Signal Processing*, vol. 33, pp 1491-1505, 2019.
[45] C. Rao and H. Toutenburg, H, *Linear Models: Least Squares and Alternatives*, Springer Series in Statistics (3rd ed.). Berlin, Springer, 2008.
[46] W. Rudin, *Principes of Mathematical Analysis*, 3rd ed., NY:McGraw-Hill, Inc, 1976.
[47] W.J. Rugh, *Linear Systems Theory*, 2nd ed., Prentice hall, NJ, 1996.
[48] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness*, Prentice-Hall, New Jersey, 1989.
[49] V. Shaferman, M. Schwegel, T. Gluck and A. Kugi, Continuous-time least-squares forgetting algorithms for indirect adaptive control, *European Journal of Control*, vol. 62, pp.105-112, 2021.
[50] H. Shin and H. Lee, A new exponential forgetting algorithm for recursive least-squares parameter estimation, arXiv:2004.03910, 2020.
[51] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice-Hall, New Jersey, USA, 1991.
[52] G. Tao, *Adaptive Control Design and Analysis*, vol. 37, John Wiley & Sons, New Jersey, 2003.
[53] L. Wang, R. Ortega, A. Bobtsov, J. G. Romero and B. Yi, Identifiability implies robust, globally exponentially convergent on-line parameter estimation: Application to model reference adaptive control, arXiv:2108.08436, 2021.
[54] Z. Wu, M. Ma, X. Xu, B. Liu and Z. Yu. Predefined-time parameter estimation via modified dynamic regressor extension and mixing, *Journal of the Franklin Institute*, doi.org/10.1016/j.jfranklin.2021.06.028, 2021.
[55] B. Yi and R. Ortega, Conditions for convergence of dynamic regressor extension and mixing parameter estimators using LTI filters, *IEEE Trans. Automatic Control*, 10.1109/TAC.2022.3149964, 2022.

Appendix A

In this appendix we briefly review the main steps in the construction of DREM-based estimators proceeding from the NLPRE (1). For the sake of brevity we restrict ourselves to CT versions, with the DT ones constructed *verbatim*. The interested reader is refered to [38] for further details on these constructions.

Derivation of classical DREM-based estimators

S1 Starting from the NLPRE $y(t) = \phi^\top(t)\mathcal{G}(\theta)$, with $y(t) \in \mathbb{R}$, $\phi(t) \in \mathbb{R}^p$ measurable signals, $\mathcal{G} : \mathbb{R}^q \to \mathbb{R}^p$, $p \geq q$ and $\theta \in \mathbb{R}^q$ a constant vector of unknown parameters.

S2 (Creation of the extended regressor) Inclusion of a *free, stable, linear operator* $\mathcal{H} : u(t) \to U(t)$, with $u(t) \in \mathbb{R}$ and $U(t) \in \mathbb{R}^p$, via its state space realization

$$\dot{U}(t) = A(t)U(t) + b(t)u(t), \tag{.1}$$

with $A(t) \in \mathbb{R}^{p\times p}$, $b(t) \in \mathbb{R}^p$. Upon application to the NLPRE above, create a new extended NLPRE

$$Y(t) = \Phi(t)\mathcal{G}(\theta) \tag{.2}$$

with

$$\begin{aligned} Y(t) &:= \mathcal{H}[y](t) \in \mathbb{R}^p \\ \Phi(t) &:= [\mathcal{H}[\phi_1](t) \mid \mathcal{H}[\phi_2](t) \mid \ \ldots \mid \mathcal{H}[\phi_p](t)] \in \mathbb{R}^{p\times p}. \end{aligned} \tag{.3}$$

We underscore the fact that the new extended regressor $\Phi(t)$ is a *square matrix*.

S3 (Lion's and Kreisselmeier REs) For Lion's RE [24] we select for $\mathcal{H}$ the LTI filter

$$A := \text{diag}\{-a_i\},\ b := \text{col}(b_1, \ldots, b_p),$$

with $b_i \neq b_j, a_i \neq a_j > 0,\ (i,j) \in \bar{p}$.
For Kreisselmeier RE we select LTV operators with

$$A := \text{diag}\{-a_i\},\ b(t) := \phi(t),\ a_i > 0,\ i \in \bar{p}.$$

S4 (Mixing step) Multiplication of the extended LPRE (.2) by the *adjugate* of $\Phi(t)$ to create the new NLPRE

$$\mathcal{Y}(t) = \Delta(t)\mathcal{G}(\theta), \tag{.4}$$

with

$$\begin{aligned} \mathcal{Y}(t) &:= \text{adj}\{\Phi(t)\}Y(t) \in \mathbb{R}^p \\ \Delta(t) &:= \det\{\Phi(t)\} \in \mathbb{R} \end{aligned}$$

and *scalar* regressor $\Delta(t)$. Notice that in the case of LPRE $y(t) = \phi^\top(t)\theta$ we obtain q scalar LPREs of the form

$$\mathcal{Y}_i(t) = \Delta(t)\theta_i,\ i \in \bar{q}.$$