

New Results on Parameter Estimation via Dynamic Regressor Extension and Mixing: Continuous and Discrete-time Cases

Romeo Ortega, Stanislav Aranovskiy, Anton Pyrkin, Alessandro Astolfi, Alexey Bobtsov

▶ **To cite this version:**

Romeo Ortega, Stanislav Aranovskiy, Anton Pyrkin, Alessandro Astolfi, Alexey Bobtsov. New Results on Parameter Estimation via Dynamic Regressor Extension and Mixing: Continuous and Discrete-time Cases. IEEE Transactions on Automatic Control, 2021, 66 (5), pp.2265-2272. 10.1109/TAC.2020.3003651 . hal-02948492

HAL Id: hal-02948492

https://univ-rennes.hal.science/hal-02948492

Submitted on 25 Sep 2020

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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

New Results on Parameter Estimation via Dynamic Regressor Extension and Mixing: Continuous and Discrete-time Cases

Romeo Ortega, *Fellow, IEEE*, Stanislav Aranovskiy, *Senior member, IEEE*, Anton A. Pyrkin, *Member, IEEE*, Alessandro Astolfi, *Fellow, IEEE*, Alexey A. Bobtsov, *Senior member, IEEE*

***Abstract*—We present some new results on the dynamic regressor extension and mixing parameter estimators for linear regression models recently proposed in the literature. This technique has proven instrumental in the solution of several open problems in system identification and adaptive control. The new results include: (i) a unified treatment of the continuous and the discrete-time cases; (ii) the proposal of two new extended regressor matrices, one which guarantees a quantifiable *transient performance improvement*, and the other exponential convergence under conditions that are *strictly weaker* than regressor persistence of excitation; and (iii) an alternative estimator ensuring convergence in *finite-time* whose adaptation gain, in contrast with the existing one, does not converge to zero. Simulations that illustrate our results are also presented.**

I. Introduction

Estimation of the parameters that describe an underlying physical setting is one of the central problems in control and systems theory that has attracted the attention of many researchers for several years. A typical scenario, which appears in system identification and adaptive control [8], [9], [16], [17], [22], is when the unknown parameters and the measured data are linearly related in a so-called *linear regression equation* (LRE). Classical solutions for this problem are gradient and least-squares (LS) estimators. The main drawback of these schemes is that convergence of the parameter estimates relies on the availability of signal excitation, a feature that is codified in the restrictive assumption of persistency of excitation (PE) of the regressor vector. Moreover, their transient performance is highly unpredictable and only a weak monotonicity property of the estimation errors can be guaranteed.

To overcome these two problems a new parameter estimation procedure, called dynamic regressor extension and mixing (DREM), has recently been proposed in [2] for continuous-time (CT) and in [5] for discrete-time (DT) systems. The construction of DREM estimators proceeds in two steps, first, the inclusion of a *free linear operator* that creates an extended, matrix LRE. Second, a *nonlinear* manipulation of the data that allow generating, out of an m-dimensional LRE, m *scalar*, and independent, LRE. DREM estimators have been successfully applied in a variety of identification and adaptive control problems. Interestingly, it has been shown in [18] that DREM can be reformulated as a functional Luenberger observer.

DREM estimators outperform classical gradient or LS estimators in the following precise aspects: independently of the excitation conditions, DREM guarantees monotonicity of *each* element of the parameter error vector that is much stronger than monotonicity of the vector *norm*, which is ensured with classical estimators. Moreover, parameter convergence in DREM is established *without* the PE condition. Instead of PE a non-square integrability condition on the determinant of a designer-dependent extended regressor matrix is imposed. A final interesting property of DREM that has been established in [19] is that it can be used to generate estimates with *finite-time convergence* (FTC), under interval excitation assumption.

The following new results on DREM are presented here:
(i) The unified treatment of the CT and the DT cases.
(ii) The definition of new linear operators that:
• ensure parameter error convergence under excitation conditions that are strictly weaker than regressor PE;
• guarantee a transient performance improvement;
• show that DREM contains, as a particular case, the extended LRE proposed in [11], which is used also in the adaptive controllers recently proposed in [6], [7], [21].
(iii) An alternative FTC estimator whose adaptation gain, in contrast with the existing one, does not converge to zero.

The remainder of the paper is organized as follows. To set up the notation a brief description of gradient and DREM estimators is given in Section II. In Section III we present the new version of DREM that ensures convergence under excitation conditions that are strictly weaker than regressor PE. In Section IV a general form of the free operator used in DREM is proposed to, on one hand, re-derive the extended regressor of [11] and, on the other hand, prove that transient performance is—quantifiably—improved. Section V presents a new DREM-based estimator with FTC. Simulation results are presented in Section VI. The paper is wrapped-up with future research in Section VII.

Notation. I_n is the $n \times n$ identity matrix. $\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0}$, $\mathbb{Z}_{>0}$ and $\mathbb{Z}_{\geq 0}$ denote the positive and non-negative real and integer numbers, respectively. For $x \in \mathbb{R}^n$, we denote $|x|^2 := x^\top x$. Continuous-time (CT) signals $s : \mathbb{R}_{\geq 0} \to \mathbb{R}$ are denoted $s(t)$, while for discrete-time (DT) sequences $s : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ we use $s(k) := s(kT_s)$, with $T_s \in \mathbb{R}_{>0}$ the sampling time. The action of an operator $\mathcal{H} : \mathcal{L}_\infty \to \mathcal{L}_\infty$ on a CT signal $u(t)$ is denoted $\mathcal{H}[u](t)$, while for an operator $\mathcal{H} : \ell_\infty \to \ell_\infty$ and a sequence $u(k)$ we use $\mathcal{H}[u](k)$. When a formula is applicable to CT

R. Ortega is with Laboratoire des Signaux et Systèmes, CNRS-SUPELEC, Plateau du Moulon, 91192, Gif-sur-Yvette, France and ITMO University, Kronverkskiy av. 49, St. Petersburg, 197101, Russia.
S. Aranovskiy is with CentraleSupélec – IETR, Avenue de la Boulaie, 35576 Cesson-Sévigné, France
A. Pyrkin is with Hangzhou Dianzi University, Hangzhou, 310018, China.
S. Aranovskiy, A. Pyrkin and A. Bobtsov are with the Faculty of Control Systems and Robotics, ITMO University, Kronverkskiy av. 49, St. Petersburg, 197101, Russia.
A. Astolfi is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK and the DICII, Universita di Roma "Tor Vergata", Via del Politecnico 1, 00133 Roma, Italy
A. Pyrkin is a corresponding author. E-mail: a.pyrkin@gmail.com

ACCEPTED MANUSCRIPT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

signals and DT sequences the time argument is omitted.

II. Background Material

We deal with the problem of on-line estimation of the unknown, constant parameters $\theta \in \mathbb{R}^m$ appearing in a LRE of the form

$$y = \phi^\top \theta + \varepsilon_t \tag{1}$$

where $y \in \mathbb{R}$ and $\phi \in \mathbb{R}^m$ are *measurable* CT or DT signals and ε_t is a (generic) exponentially decaying signal.[1] It is well-known that the availability of a LRE of the form (1) is instrumental for the development of most system identifiers and adaptive controllers [22]. Following standard practice, throughout the paper, the term ε_t is omitted.

A. Gradient estimator and the PE condition

In this subsection we recall the well-known gradient estimator, derive its parameter error equation (PEE) and recall its stability properties. Although this material is very well-known, it is included to make the document self-contained and set up the notation. First, we introduce the following.

Definition 1. A bounded signal $\phi \in \mathbb{R}^m$ is PE (denoted $\phi \in PE$) if there exist $\alpha \in \mathbb{R}_{>0}$ such that

$$\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau \geq \alpha I_m, \ \forall t \in \mathbb{R}_{\geq 0},$$

for some $T \in \mathbb{R}_{>0}$ in CT or

$$\sum_{j=k+1}^{k+K} \phi(j)\phi^\top(j) \geq \alpha I_m, \ \forall k \in \mathbb{Z}_{\geq 0},$$

for some $K \in \mathbb{Z}_{>0}$, with $K \geq m$, in DT. □□□

The following proposition is a milestone for systems theory and may be found in all identification and adaptive control textbooks, *e.g.*, [22].

Proposition 1. Consider the LRE (1).

(CT) The CT gradient-descent estimator

$$\dot{\hat{\theta}}(t) = \gamma\phi(t)[y(t) - \phi^\top(t)\hat{\theta}(t)], \tag{2}$$

with $\gamma > 0$ ensures the following.

• The *norm* of the parameter error vector $\tilde{\theta} := \hat{\theta} - \theta$ is monotonically non-increasing, that is,

$$|\tilde{\theta}(t_b)| \leq |\tilde{\theta}(t_a)|, \ \forall t_b \geq t_a \in \mathbb{R}_{\geq 0}. \tag{3}$$

• The CT PEE is given by

$$\dot{\tilde{\theta}}(t) = -\gamma\phi(t)\phi^\top(t)\tilde{\theta}(t),$$

and its zero equilibrium is globally exponentially stable (GES) *if and only if* $\phi(t) \in PE$. Moreover, there exist an optimal value of γ for which the rate of convergence is *maximum*.

(DT) The DT gradient-descent estimator

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\phi(k)}{\gamma + |\phi(k)|^2}[y(k) - \phi^\top(k)\hat{\theta}(k-1)],$$

ensures the following.

• The *norm* of the parameter error vector verifies

$$|\tilde{\theta}(k_b)| \leq |\tilde{\theta}(k_a)|, \ \forall k_b \geq k_a \in \mathbb{Z}_{\geq 0}. \tag{4}$$

• The DT PEE is given by

$$\tilde{\theta}(k) = \left[I_m - \frac{1}{\gamma + |\phi(k)|^2}\phi(k)\phi^\top(k)\right]\tilde{\theta}(k-1),$$

and its zero equilibrium is GES *if and only if* $\phi(k) \in PE$. □□□

In most applications, PE is an extremely restrictive condition, hence the interest of relaxing it. See [20] for a recent review of new estimators relaxing the PE condition, which include the ones reported in [6], [7], [21].

B. Generation of m scalar LRE via DREM

To overcome the limitation imposed by the PE condition and improve the transient performance of the estimator the DREM procedure, introduced in [2], [5], generates m new, one–dimensional, LRE to *independently* estimate each of the parameters. The first step in DREM is to introduce a *linear, single-input m-output, bounded-input bounded-output (BIBO)–stable* operator $\mathcal{H}$ and define the vector $Y \in \mathbb{R}^m$ and the matrix $\Phi \in \mathbb{R}^{m\times m}$ as

$$Y := \mathcal{H}[y], \ \Phi := \mathcal{H}[\phi^\top]. \tag{5}$$

Clearly, because of linearity and BIBO stability, these signals satisfy

$$Y = \Phi\theta. \tag{6}$$

At this point the key step of regressor "mixing" of the DREM procedure is used to obtain a set of m *scalar* equations as follows. First, recall that, for any (possibly singular) $m \times m$ matrix M we have [12] $\text{adj}\{M\}M = \det\{M\}I_m$, where $\text{adj}\{\cdot\}$ is the adjunct (also called "adjugate") matrix. Now, multiplying from the left the vector equation (6) by the *adjunct matrix* of Φ, we get

$$\mathcal{Y}_i = \Delta\theta_i, \quad i \in \{1, 2, \dots, m\} \tag{7}$$

where we have defined the scalar function $\Delta \in \mathbb{R}$

$$\Delta := \det\{\Phi\}, \tag{8}$$

and the vector $\mathcal{Y} \in \mathbb{R}^m$

$$\mathcal{Y} := \text{adj}\{\Phi\}Y. \tag{9}$$

Remark 1. In [13] an extended regressor like (6) has been constructed in CT using linear time-invariant (LTI) filters in the operator $\mathcal{H}$ used in (5)—see also [11], where this modification is also discussed. Unfortunately, besides some simulation evidence, no quantitative advantage—with respect to the gradient estimation—has been established for it.

Remark 2. As pointed out in [10] applying Cramer's law we have that $\mathcal{Y}_i = \det\{\Phi_{Y_i}\}$ where Φ_{Y_i} is the matrix obtained replacing the i-column of Φ with Y.

[1] This signal may be stemming from the effect of the initial conditions of various filters used to generate the LRE.

ACCEPTED MANUSCRIPT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

C. Properties of gradient parameter estimators in DREM

The availability of the scalar LREs (7) is the main feature of DREM that distinguishes it with respect to all other estimators. Indeed, as shown in the propostion below—the proof of which may be found in [2], [5]—it allows obtaining significantly stronger results using simple gradient estimators.

Proposition 2. Consider the scalar LREs (7).

(CT) The CT gradient-descent estimators[2]

$$\dot{\hat{\theta}}_i(t) = \gamma_i \Delta(t)[\mathcal{Y}_i(t) - \Delta(t)\hat{\theta}_i(t)], \tag{10}$$

with $\gamma_i \in \mathbb{R}_{>0}$ ensures the following.

• The CT PEEs are given by

$$\dot{\tilde{\theta}}_i(t) = -\gamma_i \Delta^2(t)\tilde{\theta}_i(t). \tag{11}$$

• The *individual* parameter errors are monotonically non-increasing, that is,

$$|\tilde{\theta}_i(t_b)| \leq |\tilde{\theta}_i(t_a)|, \ \forall t_b \geq t_a \in \mathbb{R}_{\geq 0}.$$

• The following *equivalence* holds

$$\lim_{t\to\infty} \tilde{\theta}_i(t) = 0 \quad \Leftrightarrow \quad \Delta(t) \notin \mathcal{L}_2,$$

and convergence can be made *arbitrarily fast* increasing γ_i.

• If $\Delta(t) \in PE$, the convergence is *exponential*.

(DT) The DT gradient-descent estimator

$$\hat{\theta}_i(k) = \hat{\theta}_i(k-1) + \frac{\Delta(k)}{\gamma_i + \Delta^2(k)}[\mathcal{Y}_i(k) - \Delta(k)\hat{\theta}_i(k-1)], \tag{12}$$

ensures the following.

• The DT PEEs are given by

$$\tilde{\theta}_i(k) = \frac{1}{1 + \frac{\Delta^2(k)}{\gamma_i}}\tilde{\theta}_i(k-1). \tag{13}$$

• The elements of the parameter error vector verify

$$|\tilde{\theta}_i(k_b)| \leq |\tilde{\theta}_i(k_a)|, \ \forall k_b \geq k_a \in \mathbb{Z}_{\geq 0}. \tag{14}$$

• The following *equivalence* holds

$$\lim_{t\to\infty} \tilde{\theta}_i(k) = 0 \quad \Leftrightarrow \quad \Delta(k) \notin \ell_2,$$

and convergence can be made *arbitrarily fast* decreasing γ_i.

• If $\Delta(k) \in PE$, the convergence is *exponential*. □□□

There are three important advantages of DREM over the standard gradient estimator.

P1 As shown in (14) the *individual* parameter errors are monotonically non-increasing, a property that is *strictly stronger* than monotonicity of their norm indicated in (3) and (4).

P2 Parameter convergence is established *without* the restrictive PE assumption—being replaced, instead, by a non square-integrability/summability assumption.

P3 Convergence rates of DREM can be made *arbitrarily fast* simply increasing γ_i in CT (or decreasing it in DT).

[2]In the sequel, the quantifier $i \in \{1, 2, \ldots, m\}$ is omitted for brevity.

Remark 3. Regarding the property **P2**, in [2] the relationship in CT between the conditions $\phi(t) \in PE$ and $\Delta(t) \notin \mathcal{L}_2$ is thoroughly discussed. In particular, in [2] it has been shown that, for *arbitrary* regressor vectors $\phi(t)$, these conditions are unrelated. On the other hand, for the case of identification of LTI systems, it has been shown in [4] that $\phi(t) \in PE$ if and only if $\Delta(t) \in PE$ for *almost all* LTI operators $\mathcal{H}$.

III. A DREM Estimator with Strictly Weaker Convergence Conditions

In this section we present a *new* version of DREM for which it is possible to show that its convergence conditions are *strictly weaker* than $\phi \in PE$. Since the construction, and the results, are very similar for CT and DT estimators, for brevity, we consider the latter case only.

Proposition 3. Consider the DT version of the LRE (1). Fix an integer $\bar{K} \geq m$ and define (5) using the LTV operator

$$\mathcal{H} := \begin{bmatrix}\phi(k-1) & \phi(k-2) & \cdots \phi(k-\bar{K})\end{bmatrix}\begin{bmatrix} q^{-1} \\ q^{-2} \\ \vdots \\ q^{-\bar{K}}\end{bmatrix}.$$

Assume $\phi(k) \in PE$ and $\bar{K} \geq K$, with K the size of the window given in Definition 1. The scalar, gradient-descent DT estimators (12), with $\Delta(k)$ and $\mathcal{Y}(k)$ defined in (8) and (9), ensure the following additional properties.

• The condition for parameter convergence of DREM, *i.e.*, $\Delta(k) \notin \ell_2$, is *strictly weaker* than $\phi(k) \in PE$. More precisely, the following implications hold:

$$\phi(k) \in PE \ \Rightarrow \ \Delta(k) \notin \ell_2, \tag{15}$$

$$\Delta(k) \notin \ell_2 \ \not\Rightarrow \phi(k) \in PE. \tag{16}$$

• The condition for *exponential* parameter convergence of DREM, *i.e.*, $\Delta(k) \in PE$, is also weaker than $\phi(k) \in PE$ in the following precise sense

$$\phi(k) \in PE \ \Rightarrow \ \Delta(k) \in PE, \tag{17}$$

$$\Delta(k) \in PE \ [K \geq 2] \not\Rightarrow \phi(k) \in PE \ [K \leq \bar{K}] \tag{18}$$

Proof. To prove the claims we make the key observation that

$$\Phi(k) = \sum_{j=k+1}^{k+\bar{K}} \phi(j - (1+\bar{K}))\phi^\top(j - (1+\bar{K})). \tag{19}$$

The implications (15) and (17) follow using the identity (19), Definition 1 and noting the obvious fact that if $\phi(k) \in PE$ in a window of size K, then it is also PE for any window of size $\bar{K} \geq K$.

The proof of (16) is established with the following scalar counterexample: $\phi(k) = (k+1)^{-\frac{1}{4}}$ with $\bar{K} = 1$. Since $\phi(k)$ tends to zero it is not PE, however, $\Delta(k) = (k+1)^{-\frac{1}{2}} \notin \ell_2$.

ACCEPTED MANUSCRIPT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

Finally, the proof of (18) is established with the following chain of implications:

$$\begin{aligned}
&\Delta(k) \in PE \text{ [with } K \geq 2] \Leftrightarrow \sum_{j=k+1}^{k+K} \Delta^2(j) > 0,\ \forall k \in \mathbb{Z}_{\geq 0} \\
&\Leftrightarrow \sum_{j=k+1}^{k+K} \prod_{i=1}^{m} \lambda_i^2\{\Phi(j)\} > 0,\ \forall k \in \mathbb{Z}_{\geq 0} \\
&\Leftrightarrow \prod_{i=1}^{m} \lambda_i^2\{\Phi(k+1)\} + \ldots + \prod_{i=1}^{m} \lambda_i^2\{\Phi(k+K)\} > 0, \forall k \in \mathbb{Z}_{\geq 0} \\
&\not\Leftarrow \lambda_i\{\Phi(k)\} > 0, \forall i \in \{1, \ldots, m\},\ \forall k \in \mathbb{Z}_{\geq 0} \\
&\Leftrightarrow \Phi(k) > 0,\ \forall k \in \mathbb{Z}_{\geq 0} \\
&\Leftrightarrow \phi(k) \in PE \text{ [with } K \leq \bar{K}],
\end{aligned}$$

where $\lambda_i\{\cdot\}$ denotes eigenvalues and in the third implication we have used the fact that $K > 1$. □□□

Remark 4. The qualifiers $K \geq 2$ and $K \leq \bar{K}$ in (18) are necessary to complete the proof. Actually, it can be shown that, without these qualifiers, $[\Delta(k) \in PE \Rightarrow \phi(k) \in PE]$.

IV. Some Specific Choices of the Operator $\mathcal{H}$

In the reported literature of DREM we have considered the use of simple first-order, LTI filters or pure delays in the vector operator $\mathcal{H}$, see [18] for a discussion on LTV operators. One of the main contributions of the paper is to propose a general form for these operators and give an *explicit* choice, that ensures a quantifiable transient performance improvement of the estimator. Another advantage of these general operators is that, as a particular case, we obtain the extended LRE proposed in [11] for adaptive state observation—referred in the sequel as Kreisselmeier's regressor extension (KRE).

A. A general LTV operator $\mathcal{H}$

In this subsection we propose to generate Y and Φ in (5) using, as elements of the operator $\mathcal{H}$, the single-input single-output (SISO) LTV operators

$$\mathcal{H}_i := c_i^\top (\mathfrak{d} I_{n_i} - A_i)^{-1} b_i + d_i + \nu_i \mathcal{D}_i, \tag{20}$$

where $A_i \in \mathbb{R}^{n_i \times n_i}$, $b_i, c_i \in \mathbb{R}^{n_i}$, $d_i, \nu_i \in \mathbb{R}$ are *time-varying*, $n_i \in \mathbb{Z}_{\geq 0}$ and the action of the operators $\mathfrak{d}$ and $\mathcal{D}_i$ is defined as

$$\mathfrak{d}[u] = \begin{cases} \frac{du(t)}{dt} =: p[u](t) & \text{in } CT \\ u(k+1) =: q[u](k) & \text{in } DT \end{cases}$$

and

$$\mathcal{D}_i[u] = \begin{cases} u(t - T_i),\ T_i \in \mathbb{R}_{\geq 0} & \text{in } CT \\ q^{-K_i} u(k),\ K_i \in \mathbb{Z}_{\geq 0} & \text{in } DT, \end{cases}$$

respectively. The triplets (A_i, b_i, c_i) should define BIBO stable systems and all matrices are bounded.

The state-space realizations of the SISO, BIBO stable subsystems $z = \mathcal{H}_i[u]$ are, clearly, given as

$$\begin{aligned} \dot{x}_i(t) &= A_i(t) x_i(t) + b_i(t) u(t) \\ z(t) &= c_i^\top(t) x_i(t) + d_i(t) u(t) + \nu_i(t) u(t - T_i), \end{aligned}$$

in CT, and

$$\begin{aligned} x_i(k+1) &= A_i(k) x_i(k) + b_i(k) u(k) \\ z(k) &= c_i^\top(k) x_i(k) + d_i(k) u(k) + \nu_i(k) u(k - K_i), \end{aligned}$$

in DT, with $x_i \in \mathbb{R}^{n_i}$ the corresponding state. In view of the equivalence between GES and BIBO-stability for LTV systems with bounded realization matrices, these state-space systems are GES.

Remark 5. The LTV operators (20) are a generalization of the first order LTI ones or simple delays considered in the reported literature of DREM. LTV operators are also considered in [18] to give a Luenberger observer interpretation of DREM.

B. Kreisselmeier's regressor extension

The construction of the KRE of [20] proceeds as follows. Premultiplying (1) by ϕ we obtain $\phi y = \phi\phi^\top \theta$, to which we can apply a *SISO*, linear, BIBO-stable operator $\mathcal{K}$ to obtain the new, *matrix* LRE

$$Z = \Omega \theta, \tag{21}$$

where we have defined

$$Z := \mathcal{K}[\phi y] \in \mathbb{R}^m,\ \Omega := \mathcal{K}[\phi\phi^\top] \in \mathbb{R}^{m \times m}. \tag{22}$$

Comparing (5), (6) with (21), (22) we see that the difference between DRE and KRE is that, in the first case, Y and Φ are obtained filtering—with m different filters—y and ϕ, and piling-up the filtered signals, while in the latter we filter ϕy and $\phi\phi^\top$ with one filter.

The proposition below, the proof of which is obtained via a direct calculation, shows that the KRE can be derived from the DRE construction using the generalized operators (20).[3]

Proposition 4. Define (5) using (20) with

$$n_i = 1,\ c_i = 1,\ A_i = -a,\ b_i = \phi_i,\ d_i = 0,\ \nu_i = 0.$$

Then, $Z = Y$ and $\Omega = \Phi$ as defined in (22). Proving that KRE is a *particular case* of DRE. □□□

Remark 6. The KRE construction was first proposed by Kreisselmeier in [11] for CT systems and the particular case

$$\mathcal{K}(p) = \frac{1}{p + a},\ a > 0, \tag{23}$$

the state-space realization of which is

$$\begin{aligned} \dot{\Omega}(t) &= -a\Omega(t) + \phi(t)\phi^\top(t), \\ \dot{Z}(t) &= -aZ(t) + \phi(t) y(t). \end{aligned} \tag{24}$$

The LRE (21) is used in the recently proposed MRACs [6], [21], see also [20] for a survey of the literature.

[3]The first author thanks Bowen Yi for bringing this fact to his attention.

ACCEPTED MANUSCRIPT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

C. An operator $\mathcal{H}$ with guaranteed transient performance improvement

One important feature of DREM is that it is possible to get the explicit solution of the PEEs, fully characterizing the time evolution of the parameter errors. Indeed, for the CT PEE (11) we have

$$\tilde{\theta}_i(t) = e^{-\gamma_i \int_0^t \Delta^2(s)ds}\tilde{\theta}_i(0), \tag{25}$$

Similarly, for the DT PEE (13) we have

$$\tilde{\theta}_i(k) = \prod_{j=0}^{k}\left[\frac{1}{1+\frac{\Delta^2(j)}{\gamma_i}}\right]\tilde{\theta}_i(0). \tag{26}$$

As seen from the two previous equations the transient performance of the DREM estimators is univocally determined by the "size" of Δ^2—with a faster convergence obtained with a "larger" Δ^2. To improve the transient behavior of the DREM estimator we propose in this subsection a particular selection of the feedforward gains d_i in the LTV operators $\mathcal{H}_i$ given in (20). Since the result is the same for CT and DT estimators, for brevity, we consider below the former case only.

To streamline the presentation of the result we define the matrix

$$\Phi_0(t) := \begin{bmatrix} c_1^\top(t)[pI_{n_1} - A_1(t)]^{-1}b_1(t) + \nu_1(t)\mathcal{D}_1 \\ \vdots \\ c_n^\top(t)[pI_{n_n} - A_n(t)]^{-1}b_n(t) + \nu_n(t)\mathcal{D}_n \end{bmatrix}[\phi^\top](t). \tag{27}$$

That is, the CT extended regressor matrix (6) generated with the operators (20) *with* $d_i(t) = 0$.

Proposition 5. Consider the CT DREM estimator (10) with the LRE (6) generated with the operators (20). Denote by $\tilde{\theta}_i^0(t)$ the parameter errors corresponding to the choice of $d_i(t) = 0$ and $\tilde{\theta}_i^{\mathtt{N}}(t)$ those corresponding to

$$d(t) = \text{adj}\{\Phi_0(t)\}\phi(t), \tag{28}$$

with $\Phi_0(t)$ defined in (27), all the remaining parameters of $\mathcal{H}_i$ and the estimators initial conditions the same for both cases. Then

$$|\tilde{\theta}_i^0(t)| > |\tilde{\theta}_i^{\mathtt{N}}(t)|,\ \forall t \in \mathbb{R}_{\geq 0}.$$

Proof. From the definitions of $\Phi(t)$ in (6), the operators (20) and $\Phi_0(t)$ in (27) we have that

$$\Phi(t) = \Phi_0(t) + d(t)\phi^\top(t). \tag{29}$$

In view of (25) the proof is completed showing that

$$\det\{\Phi_0(t)\} < \det\{\Phi(t)\}.$$

For, we apply Sylvester's determinant formula [12] to (29) to get

$$\begin{aligned}\det\{\Phi(t)\} &= \det\{\Phi_0(t)\} + d^\top(t)\text{adj}\{\Phi_0(t)\}\phi(t) \\ &= \det\{\Phi_0(t)\} + \phi^\top(t)\big[\text{adj}\{\Phi_0(t)\}\big]^\top\text{adj}\{\Phi_0(t)\}\phi(t) \\ &= \det\{\Phi_0(t)\} + |\text{adj}\{\Phi_0(t)\}\phi(t)|^2,\end{aligned}$$

where we have used (28) to obtain the second equation. The proof is completed noting that $\det\{\Phi_0(t)\} \neq 0$ implies that $\text{adj}\{\Phi_0(t)\}$ is full rank. Hence, if $\phi(t) \neq 0$, the second right hand term of the last identity above is positive. □□□

V. CT DREM Estimators with Alert Finite-Time Convergence

In [19] we have showed that CT DREM can be used to generate estimates that converge in *finite time* under the weakest interval excitation assumption.

A. An FTC DREM

For ease of reference, we recall the FTC result used in [19] to solve a mutivariable adaptive control problem.

Proposition 6. Consider the scalar CT LREs (7) and the gradient-descent estimator (10). Fix a constant $\nu_i \in (0,1)$ and assume there exists a time $t_c \in \mathbb{R}_{>0}$ such that

$$\gamma_i \int_0^{t_c} \Delta^2(s)ds \geq -\ln(\mu_i). \tag{30}$$

Define the FTC estimate

$$\hat{\theta}_i^{\mathtt{FTC}}(t) := \frac{1}{1-w_i^{\mathtt{c}}(t)}[\hat{\theta}_i(t) - w_i^{\mathtt{c}}(t)\hat{\theta}_i(0)], \tag{31}$$

where $w_i^{\mathtt{c}}(t)$ is defined via the clipping function

$$w_i^{\mathtt{c}}(t) = \begin{cases} \mu_i & \text{if} \quad w_i(t) \geq \mu_i \\ w_i(t) & \text{if} \quad w_i(t) < \mu_i, \end{cases} \tag{32}$$

with $w_i(t)$ given by

$$\dot{w}_i(t) = -\gamma\Delta^2(t)w_i(t),\ w_i(0) = 1. \tag{33}$$

The parameter estimation error converges to zero in *finite-time*. More precisely, $\hat{\theta}_i^{\mathtt{FTC}}(t) = \theta_i,\ \forall t \geq t_c$.

Proof. First, notice that the solution of (33) is

$$w_i(t) = e^{-\gamma_i \int_0^t \Delta^2(s)ds}.$$

The key observation is that, using the equation above in (25), and rearranging terms we get that

$$[1 - w_i(t)]\theta_i = \hat{\theta}_i(t) - w_i(t)\hat{\theta}_i(0). \tag{34}$$

Now, observe that $w_i(t)$ is a non-increasing function and, under the interval excitation assumption (30), we have that

$$w_i^{\mathtt{c}}(t) = w_i(t) < \mu_i,\ \forall t \geq t_c,$$

completing the proof. □□□

Remark 7. The FTC property established in this section is "trajectory-dependent", in the sense that it relates only to the trajectory generated for the initial condition $w_i(0) = 1$. This means that the flow of the closed-loop system contains other trajectories, and the appearance of a perturbation may drive our "good" trajectory towards a "bad" one. This is, of course, a robustness problem that needs to be further investigated. Notice, however, that if $\Delta(t) \in PE$ the system (33) is *contracting*, that is, all trajectories converge to one unique trajectory, and the problem is avoided.

ACCEPTED MANUSCRIPT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

B. New FTC DREM

The problem with the approach described above is that, *independently* from the behavior of $\Delta(t)$, the function $w(t)$ is monotonically non-increasing and, actually, it converges to zero if and only if $\Delta(t) \notin \mathcal{L}_2$. In this case, $\hat{\theta}_i^{\texttt{FTC}}(t) \to \hat{\theta}_i(t)$, hence, the new estimator reduces to the standard gradient one, losing its FTC feature. Therefore, to ensure that the adaptation gain does not converge to zero and is able to react upon the arrival of new excitation, it is necessary to reset the estimators (10) or (12)—a modification that is always problematic to implement. This property is known in the literature as *alertness* of the estimator, and has been thoroughly discussed for least-squares algorithm, see [22, Section 2.4.2].

In the proposition below we propose an alternative for the CT estimator of Proposition 6 that does not suffer from this practical drawback. For the sake of brevity, we present only the derivation of a relation similar to (34), from which we can easily construct the FTC estimator and prove that the new FTC estimator does not converge to the gradient one.

Proposition 7. Fix $T_{\texttt{D}} \in \mathbb{R}_{>0}$ and define

$$\dot{w}_i^{\texttt{D}}(t) = -\gamma_i \left[\Delta^2(t) - \Delta^2(t - T_{\texttt{D}})\right] w_i^{\texttt{D}}(t),\ w_i^{\texttt{D}}(0) = 1. \quad (35)$$

Then,

$$\left[1 - w_i^{\texttt{D}}(t)\right] \theta_i = \hat{\theta}(t) - w_i^{\texttt{D}}(t)\hat{\theta}_i(t - T_{\texttt{D}}).$$

Moreover, $w_i^{\texttt{D}}(t)$ is *bounded away from zero*.

Proof. Without loss of generality we assume that $\Delta(t-T_{\texttt{D}}) = 0$ for $t < T_{\texttt{D}}$. Then, the solution of (35) is

$$w_i^{\texttt{D}}(t) = e^{-\gamma_i \int_{t-T_{\texttt{D}}}^{t} \Delta^2(s)ds}. \quad (36)$$

From (36), and the fact that

$$\int_{t-T_{\texttt{D}}}^{t} \Delta^2(s)ds \leq \Delta_{\max}^2 T_{\texttt{D}},$$

where $\Delta_{\max} \geq \|\Delta(t)\|_\infty$, we conclude that

$$w_i^{\texttt{D}}(t) \geq e^{-\gamma_i \Delta_{\max}^2 T_{\texttt{D}}} > 0.$$

Now, from the solution of the PEE (25) in the interval $[t-T_{\texttt{D}}, t]$ we get

$$\tilde{\theta}_i(t) = e^{-\gamma_i \int_{t-t_{\texttt{D}}}^{t} \Delta^2(s)ds} \tilde{\theta}_i(t - T_{\texttt{D}}).$$

Hence, $\tilde{\theta}_i(t) = w_i^{\texttt{D}}(t)\tilde{\theta}_i(t - T_{\texttt{D}})$. The proof of the claim is established rearranging the terms of the equation above. □□□

The new signal $w_i^{\texttt{D}}(t)$, besides being bounded away from zero, *grows* if $\Delta(t)$ increases its value in an interval of length $T_{\texttt{D}}$, that is, if new excitation arrives to the system. In this way, the new FTC estimator preserves its FTC property if the parameters change. This fact is illustrated in the simulations of Subsection VI-B.

Remark 8. For the new FTC DREM estimator the interval excitation inequality becomes the existence of a time $t_c \geq T_{\texttt{D}}$ such that

$$\gamma_i \int_{t_c - T_{\texttt{D}}}^{t_c} \Delta^2(s)ds \geq -\ln(\mu_i). \quad (37)$$

Recalling (36), it has the same interpretation as (30).

Remark 9. The choice of the coefficients μ_i is, clearly, a compromise between high-gain injection—if it is close to 1—and the time where FTC is achieved. See [19] for additional details on this aspect. Notice also that, in the highly improbable situation, that Δ is periodic of period T_d, the signal $w^D(t) = 1,\ \forall t \geq 0$. In this case, the assumption of sufficient excitation of Proposition 6 is violated, invalidating the FTC claim.

VI. Simulations

In this section we present simulations illustrating the results of Propositions 5, 6 and 7.

A. Transient performance improvement of Proposition 5

To illustrate the performance improvement using the time-varying term $d(t)$ introduced in Proposition 5, we consider the problem of parameter estimation of the CT, first-order, LTI plant described by

$$\dot{y}(t) = ay(t) + bu(t), \quad (38)$$

where $u(t), y(t) \in \mathbb{R}$ are measurable signals and $a, b \in \mathbb{R}$ are uncertain parameters that should be estimated.

Following the standard LTI systems identification procedure [22], we first re-parametrize the model (38) to obtain the LRE (1). For, we apply the filters $\frac{1}{p+\lambda}$, with some $\lambda > 0$, to (38) to get the LRE (1) with

$$\phi(t) := \begin{bmatrix} \frac{1}{p+\lambda}[y](t) \\ \frac{1}{p+\lambda}[u](t) \end{bmatrix},\ \theta := \begin{bmatrix} a + \lambda \\ b \end{bmatrix}.$$

Two simulation scenarios have been considered: with a plant input that is sufficiently rich or not—that is, when the regressor $\phi(t)$ is PE or not. More precisely, we considered $u(t) = 15\sin(2.5t + 1)$ and $u(t) = 15$, respectively. For these two scenarios, we compare three different estimation schemes, namely, the standard gradient-descent (2), and the DREM scheme (10) with the operators $\mathcal{H}_i$ defined in (20) for $d = 0$ and $d(t)$ given by (28).

The following simulation parameters are used: $a = -0.4, b = 0.4, \lambda = 5, \gamma = 1$, with the coefficients

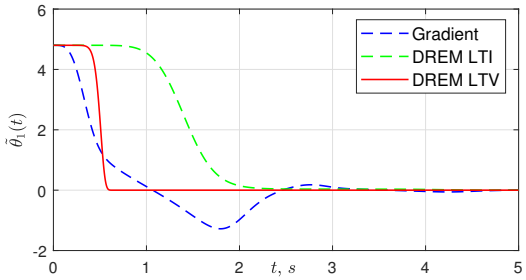
$$\mathcal{H}_1 :\ n_1 = 1,\ A_1 = -1,\ b_1 = 1,\ c_1 = 1,\ \mu_1 = 0$$
$$\mathcal{H}_2 :\ n_2 = 1,\ A_2 = -2,\ b_2 = 2,\ c_2 = 1,\ \mu_2 = 0,$$

for the LTI part of the operators $\mathcal{H}_i$.

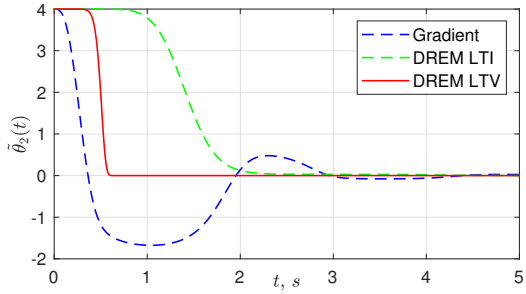
The transient behavior of the parameter estimation errors $\tilde{\theta}_1(t)$ and $\tilde{\theta}_2(t)$, for the three aforementioned estimators, is shown in Figs. 1 and 2. As predicted by the theory the gradient scheme yields a consistent estimate only for the case of sufficiently rich input, showing a significant steady state error for the constant plant input. On the other hand, both DREM schemes yield consistent estimates in both scenarios. Moreover, as expected from the analysis of Proposition 5, the addition of the feedforward term $d(t)$ given in (28) significantly improves the transient performance—achieving parameter convergence in less than a second, while the DREM scheme with $d = 0$ takes almost two seconds to converge. It should also be mentioned that both DREM schemes significantly outperform the standard gradient, even in the presence of a sufficiently rich input. This property stems from the fact that,

ACCEPTED MANUSCRIPT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

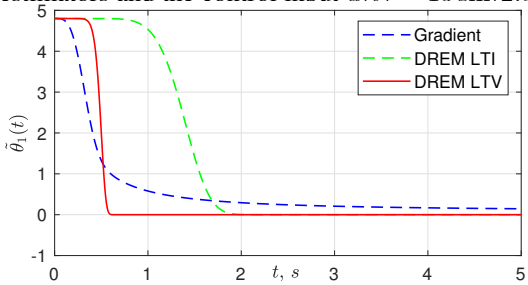


(a) Parameter estimation error $\tilde{\theta}_1(t)$

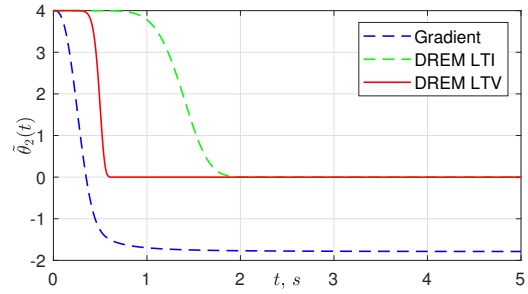


(b) Parameter estimation error $\tilde{\theta}_2(t)$

Fig. 1: Transients of the parameter estimation errors for different estimators and the control input $u(t) = 15\sin(2.5t+1)$.



(a) Parameter estimation error $\tilde{\theta}_1(t)$



(b) Parameter estimation error $\tilde{\theta}_2(t)$

Fig. 2: Transients of the parameter estimation errors for different estimators and the control input $u(t) = 15$.

as indicated in Proposition 2, DREM ensures monotonicity of each element of the parameter error vector, a fact that is clearly illustrated in the simulations.

B. Alertness preserving DREM with FTC of Proposition 7

In this subsection we compare the two FTC DREMs presented in Section V. Namely, the FTC DREM of Proposition 6, defined by (31), (33), and the new FTC DREM of Proposition 7 given by (35) and

$$\hat{\theta}_i^{\texttt{FTC}-\texttt{D}}(t) := \frac{1}{1-w_i^{\texttt{D}}(t)}[\hat{\theta}_i(t) - w_i^{\texttt{D}}(t)\hat{\theta}_i(0)],$$

which is computed as soon as $w_i^{\texttt{D}}(t) < \mu_i$. The objective of the simulation is to prove that the new FTC DREM is able to react when new excitation arrives. This is in contrast with the old FTC DREM estimator that, since $w(t) \to 0$, converges to the gradient estimator and loses its FTC alertness property.

We consider the simplest case of a scalar system $y(t) = \Delta(t)\theta$ and simulate the gradient estimator (10), that is,

$$\dot{\hat{\theta}}(t) = \gamma\Delta(t)[y(t) - \Delta(t)\hat{\theta}(t)],$$

together with (33) and (35), which are computed for $t \geq t_c$, with t_c defined via the interval excitation criteria (30) and (37), respectively.

We consider two scenarios: with and without excitation in $\Delta(t)$. For the first case we consider the PE signal $\Delta(t) = \sin(2\pi t)$, and for the second one $\Delta(t) = \frac{1}{t+1}$. Note that in the second case $\Delta(t) \to 0$, hence it is not PE. However, $\Delta(t) \notin \mathcal{L}_2$, hence it satisfies the conditions for convergence of the DREM estimator.

For simulations we set $\gamma = 2$, $\mu = 0.98$, and $T_{\texttt{D}} = 0.2$. These parameters have been chosen such that the transients of both FTC estimators coincide in the ideal case when θ is constant and the system is excited. To illustrate the FTC tracking capabilities of the estimators the unknown parameter θ is time-varying and given by

$$\theta(t) = \begin{cases} 10 \text{ for } 0 \leq t < 10, \\ 15 \text{ for } 10 \leq t < 20, \\ 15 - 0.5(t-20) \text{ for } 20 \leq t < 30, \\ 10 \text{ for } t > 30, \end{cases}$$

i.e., it starts at 10, jumps to 15 at $t = 10$, and then linearly returns to 10.

The transient of the estimators for $t \in [0, 3]$ and $\Delta(t) = \sin(2\pi t)$ are given in Fig. 3, where we plot the gradient estimate $\hat{\theta}(t)$, as well as the old and the new FTC estimates $\hat{\theta}^{\texttt{FTC}}(t)$ and $\hat{\theta}^{\texttt{FTC}-\texttt{D}}(t)$. We observe that, as expected, both FTC estimators are overlapped and converge in finite time, while the gradient converges only asymptotically.

The behavior of the estimators for $t \in [9, 40]$ is shown in Figure 4, where we also plot the time-varying parameter $\theta(t)$. As predicted by the theory, the old FTC behaves as the gradient estimator and their trajectories coincide. On the other hand, the new estimator preserves FTC alertness after the first parameter jump and achieves fast tracking of the linearly time-varying $\theta(t)$.

For the non-PE case of $\Delta(t) = \frac{1}{t+1}$, the transients of the estimators are given in Fig. 5. We observe that both FTC estimators, again, essentially coincide in the first few seconds and converge in finite time, while the gradient does it only asymptotically. After the first parameter change at $t = 10$ the old FTC and the gradient coincide, while the new FTC manages to track in finite time the parameter jump. However, during the ramp parameter change—because of the lack of excitation—neither one of the estimators can track the parameter variation but the new FTC estimator performs much better.

VII. Future Work

Current research is under way to derive some of the new results presented only for the CT time case, to the practically

ACCEPTED MANUSCRIPT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.3003651, IEEE Transactions on Automatic Control

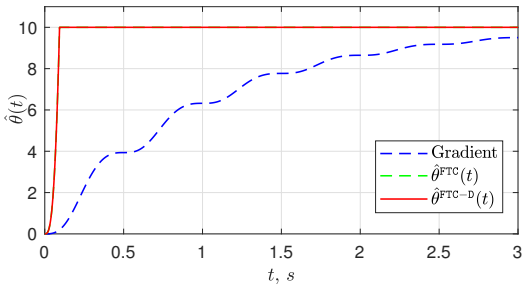


Fig. 3: Transients of the parameter estimates for $t \in [0, 3]$ with $\Delta(t) \in PE$.

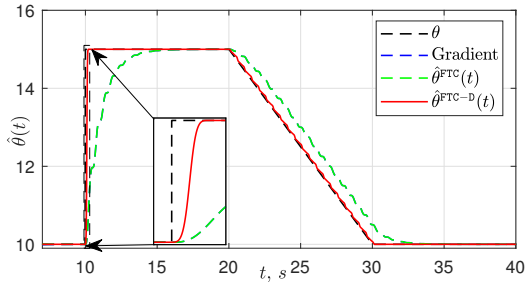


Fig. 4: Transients of the parameter estimates for $t \in [9, 40]$ with $\Delta(t) \in PE$.

important, DT case. Moreover, in the spirit of [4], we are further exploring the role of the operator $\mathcal{H}$ on the determinant of the extended regressor matrix Φ and we plan to study the effect of an additive signal in the LRE (1), to study its input-to-state stability properties.

A widely open, long-term research topic is how to deal with *nonlinear parameterizations*, that is, the case in which (1) is replaced by $y = F(\phi, \theta)$, where $F(\cdot, \cdot)$ is a nonlinear function. Some preliminary results exploiting convexity, concavity or monotonicity may be found in [1], [14], [15]. As pointed out in [2], DREM is directly applicable—without overparameterization—in the simplest case of separable nonlinearities, that is, when the regression is of the form $y = F_\phi(\phi)F_\theta(\theta)$. The more general case is a challenging open problem. Similarly, although the scheme of Proposition 7 preserves the estimator alertness, which is a necessary condition to track time-varying parameters, a formal treatment of this case is yet to be done.

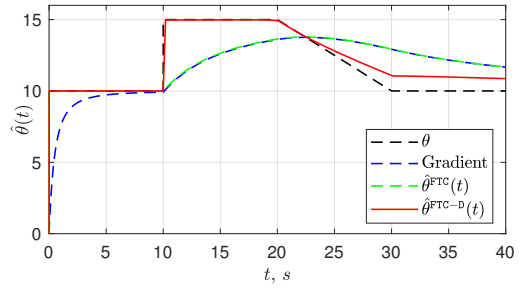


Fig. 5: Transients of the parameter estimates for $\Delta(t) = \frac{1}{t+1}$.

ACKNOWLEDGMENT

The authors would like to thank Vladimir Nikiforov and Dmitry Gerasimov for many useful discussions that helped us to improve the quality of our contribution.

This paper is partly supported by by Government of Russian Federation (GOSZADANIE 2.8878.2017/8.9, grant 08-08), the European Union's Horizon 2020 Research and Innovation Programme under Grant 739551 (KIOS CoE).

REFERENCES

[1] A.Annaswamy, F. Skantze and A. Loh, Adaptive control of continuous time systems with convex/concave parametrization, *Automatica*, vol. 34, no. 1, pp. 33-49, 1998.

[2] 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, no 7, pp.3546-3550, 2017.

[3] A. Astolfi, D. Karagiannis and R. Ortega, *Nonlinear and Adaptive Control with Applications*, Springer-Verlag, Berlin, Communications and Control Engineering, 2008.

[4] S. Aranovskiy, A. Belov, R. Ortega, N. Barabanov and A. Bobtsov, Parameter identification of linear time-invariant systems using dynamic regressor extension and mixing, *Int. J. on Adaptive Control and Signal Processing*, vol. 33, no. 6, pp. 1016-1030, 2019.

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

[6] N. Cho, H. Shin, Y. Kim and A. Tsourdos, Composite MRAC with parameter convergence under finite excitation, *IEEE Trans. Automatic Control*, vol. 63, no. 3, pp. 811-818, 2018.

[7] G. Chowdhary, T. Yucelen, M. Mhlegg and E. Johnson, Concurrent learning adaptive control of linear systems with exponentially convergent bounds, *Int. J. on Adaptive Control and Signal Processing*, vol. 27, no. 4, pp. 280-301, 2013.

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

[9] P. Ioannou and J. Sun, *Robust Adaptive Control*, Prentice–Hall, NJ, 1996.

[10] M. Korotina, S. Aranovskiy, R. Ushirobira, and A. Vedyakov, On parameter tuning and convergence properties of the DREM procedure, *in Proc. European Control Conference*, Saint-Petersburg, Russia, 2020.

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

[12] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Academic Press, NY, 1985.

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

[14] X. Liu, R. Ortega, H. Su and J. Chu. Immersion and invariance adaptive control of nonlinearly parameterized nonlinear systems, *IEEE Trans. Automatic Control*, vol. 55, no. 9, pp. 2209-2214, 2010.

[15] X. Liu, R. Ortega, H. Su and J. Chu, On adaptive control of nonlinearly parameterized nonlinear systems: towards a constructive procedure, *Systems and Control Letters*, vol. 10, pp. 36-43, 2011.

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

[17] K. Narendra and A. Annaswamy, *Stable Adaptive Systems*, Prentice Hall, New Jersey, 1989.

[18] R. Ortega, L Praly, S. Aranovskiy, B Yi and W. Zhang, On dynamic regressor extension and mixing parameter estimators: two Luenberger observers interpretations, *Automatica*, vol. 95, pp. 548-551, 2018.

[19] 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*, (to appear), (doi.org/10.1016/j.automatica.2019.108589.)

[20] R. Ortega, V. Nikiforov and D. Gerasimov, On modified parameter estimators for identification and adaptive control: a unified framework and some new schemes, *IFAC Annual Reviews in Control*, (to appear).

[21] S.B. Roy and S. Bhasin and I.N. Kar, Combined MRAC for unknown MIMO LTI systems with parameter convergence, *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 283-290, 2018,

[22] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness*, Prentice-Hall, New Jersey, 1989.

ACCEPTED MANUSCRIPT