



Parameters Estimation via Dynamic Regressor Extension and Mixing

Stanislav Aranovskiy, Alexey Bobtsov, Romeo Ortega, Anton Pyrkin

► **To cite this version:**

Stanislav Aranovskiy, Alexey Bobtsov, Romeo Ortega, Anton Pyrkin. Parameters Estimation via Dynamic Regressor Extension and Mixing. American Control Conference (ACC), Jul 2016, Boston, United States. hal-02378513

HAL Id: hal-02378513

<https://hal-centralesupelec.archives-ouvertes.fr/hal-02378513>

Submitted on 17 Jun 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.

Parameters Estimation via Dynamic Regressor Extension and Mixing*

Stanislav Aranovskiy¹, Alexey Bobtsov², Romeo Ortega³, Anton Pyrkin²

Abstract—A new way to design parameter estimators with enhanced performance is proposed in the paper. The procedure consists of two stages, first, the generation of new regression forms via the application of a dynamic operator to the original regression. Second, a suitable mix of these new regressors to obtain the final desired regression form. For classical linear regression forms the procedure yields a new parameter estimator whose convergence is established without the usual requirement of regressor persistency of excitation. The technique is also applied to nonlinear regressions with “partially” monotonic parameter dependence—giving rise again to estimators with enhanced performance. Simulation results illustrate the advantages of the proposed procedure in both scenarios.

I. INTRODUCTION

A new procedure to design parameter identification schemes is proposed in this article. The procedure, called Dynamic Regressor Extension and Mixing (DREM), consists of two stages, first, the generation of new regression forms via the application of a dynamic operator to the data of the original regression. Second, a suitable mix of these new data to obtain the final desired regression form to which standard parameter estimation techniques are applied.

The DREM procedure is applied in two different scenarios. First, for linear regression systems, it is used to generate a new parameter estimator whose convergence is ensured *without a persistency of excitation* (PE) condition on the regressor. It is well known that standard parameter estimation algorithms applied to linear regressions give rise to a linear time-varying system, which is exponentially stable if and only if a certain PE condition is imposed—this fundamental result constitutes one of the main building blocks of identification and adaptive control theories [2], [3]. Relaxation of the PE condition is a challenging theoretical problem and many

*Due to a lack of space the proofs were not included. The interested reader is referred to [1] to examine them.

This article is supported by Russian Science Foundation, project 16-11-00049.

¹Stanislav Aranovskiy is with Non-A team, Inria – Lille Nord-Europe, Parc Scientifique de la Haute Borne 40, avenue Halley Bat.A, Park Plaza, 59650 Villeneuve d’Ascq, France, and with the Department of Control Systems and Informatics, ITMO University, Kronverkskiy av. 49, Saint Petersburg, 197101, Russia : stanislav.aranovskiy@inria.fr

²Alexey Bobtsov and Anton Pyrkin are with the Department of Control Systems and Informatics, ITMO University, Kronverkskiy av. 49, Saint Petersburg, 197101, Russia.

³Romeo Ortega is with the LSS-Supelec, 3, Rue Joliot-Curie, 91192 Gif-sur-Yvette, France : ortega@lss.supelec.fr

research works have been devoted to it in various scenarios, see *e.g.*, [4]–[11] and references therein. Due to its practical importance research on this topic is of great current interest.

The second parameter estimation problem studied in this article is when the parameters enter *nonlinearly* in the regression form. It is well known that nonlinear parameterizations are inevitable in any realistic practical problem. On the other hand, designing parameter identification algorithms for nonlinearly parameterized regressions is a difficult poorly understood problem. An interesting case that has recently been explored in the literature is when nonlinearities can be factorisable and the dependence with respect to the parameters exhibit some *monotonicity* properties; see [12]–[14]. Unfortunately, it is often the case that this property holds true *only for some* of the functions entering in the regression stymying the application of the proposed techniques. Our second contribution is the use of the DREM technique to “isolate” the good nonlinearities and be able to exploit the monotonicity to achieve consistent parameter estimation for nonlinearly parameterised regressions with factorisable nonlinearities—*not imposing PE conditions*.

The remaining of the paper is organized as follows. The DREM technique is first explained with its application to linear regressions in Section II. In Section III DREM is used for nonlinear factorisable regressions with “partially” monotonic parameter dependence. In both sections representative simulation examples are presented. Some concluding remarks and future research are given in Section IV.

Notation. For $x \in \mathbb{R}^n$, $|x|$ is the Euclidean norm. All functions and mappings in the paper are assumed sufficiently smooth. For a function of scalar argument $g : \mathbb{R} \rightarrow \mathbb{R}^s$, g' denotes its first order derivative. For functions $V : \mathbb{R}^n \rightarrow \mathbb{R}$ we define the operator $\nabla V := (\frac{\partial V}{\partial x})^\top$. Also, for mappings $\Phi : \mathbb{R}^n \times \mathbb{R}^q \rightarrow \mathbb{R}^n$ we define its (transposed) Jacobian matrix $\nabla_x \Phi(x, \theta) := [\nabla \Phi_1(x, \theta), \dots, \nabla \Phi_n(x, \theta)]$. For the distinguished element $x_* \in \mathbb{R}^n$ and any mapping $F : \mathbb{R}^n \rightarrow \mathbb{R}^s$ we denote $F_* := F(x_*)$.

II. CONSISTENT ESTIMATION FOR LINEAR REGRESSIONS WITHOUT PE

In this section the DREM technique is applied to classical linear regressions. The main contribution is the removal of

the—often overly restrictive—assumption of regressor PE to ensure parameter convergence.

A. Standard procedure and the PE condition

Consider the basic problem of on–line estimation of the constant parameters of the q –dimensional linear regression

$$y(t) = m^\top(t)\theta, \quad (1)$$

where¹ $y : \mathbb{R}_+ \rightarrow \mathbb{R}$ and $m : \mathbb{R}_+ \rightarrow \mathbb{R}^q$ are known, bounded functions of time and $\theta \in \mathbb{R}^q$ is the vector of unknown parameters. The standard gradient estimator

$$\dot{\hat{\theta}} = \Gamma m(y - m^\top \hat{\theta}), \quad (2)$$

with a positive definite adaptation gain $\Gamma \in \mathbb{R}^{q \times q}$ yields the error equation

$$\dot{\tilde{\theta}} = -\Gamma m(t)m^\top(t)\tilde{\theta}, \quad (3)$$

where $\tilde{\theta} := \hat{\theta} - \theta$ are the parameter estimation errors. It is well–known [2], [3] that the zero equilibrium of the linear time–varying system (3) is (uniformly) exponentially stable if and only if the regressor vector m is PE, that is, if

$$\int_t^{t+T} m(s)m^\top(s)ds \geq \delta I_q, \quad (4)$$

for some $T, \delta > 0$ and for all $t \geq 0$, which will be denoted as $m(t) \in \text{PE}$. If $m(t) \notin \text{PE}$, which happens in many practical circumstances, very little can be said about the asymptotic stability of (3), hence about the convergence of the parameter errors to zero.

Remark 1: It is well known that the PE conditions for the gradient estimator presented above and more general estimators—like (weighted) least squares—exactly coincide [15]. Since the interest in the paper is to relax the PE condition, attention is restricted to the simple gradient estimator.

Remark 2: To simplify the notation it has been assumed above that the measurement signal y is one–dimensional. As will become clear below DREM is applicable also for the vector case.

B. Proposed dynamic regressor extension and mixing procedure

To overcome the limitation imposed by the PE condition the DREM procedure generates q new, one–dimensional, regression models to independently estimate each of the parameters under conditions on the regressor m that differ from the PE condition (4).

The first step in DREM is to introduce $q - 1$ linear, \mathcal{L}_∞ –stable operators $H_i : \mathcal{L}_\infty \rightarrow \mathcal{L}_\infty$, $i \in \{1, 2, \dots, q - 1\}$,

¹When clear from the context, in the sequel the arguments of the functions are omitted.

whose output, for any bounded input, may be decomposed as

$$(\cdot)_{f_i}(t) := [H_i(\cdot)](t) + \epsilon_t, \quad (5)$$

with ϵ_t is a (generic) exponentially decaying term. For instance, the operators H_i may be simple, exponentially stable LTI filters of the form $H_i(p) = \frac{\alpha_i}{p + \beta_i}$, with $p := \frac{d}{dt}$ and $\alpha_i \neq 0$, $\beta_i > 0$; in this case ϵ_t accounts for the effect of the initial conditions of the filters. Another option of interest are delay operators, that is $[H_i(\cdot)](t) := (\cdot)(t - d_i)$, where $d_i \in \mathbb{R}_+$.

Now, we apply these operators to the regressor equation (1) to get the filtered regression²

$$y_{f_i} = m_{f_i}^\top \theta.$$

Piling up the original regressor equation (1) with the $q - 1$ filtered regressors we can construct the extended regressor system

$$Y_e(t) = M_e(t)\theta, \quad (6)$$

where we defined $Y_e : \mathbb{R}_+ \rightarrow \mathbb{R}^q$ and $M_e : \mathbb{R}_+ \rightarrow \mathbb{R}^{q \times q}$ as

$$Y_e := \begin{bmatrix} y \\ y_{f_1} \\ \vdots \\ y_{f_{q-1}} \end{bmatrix}, \quad M_e := \begin{bmatrix} m^\top \\ m_{f_1}^\top \\ \vdots \\ m_{f_{q-1}}^\top \end{bmatrix}. \quad (7)$$

Note that, because of the \mathcal{L}_∞ –stability assumption of H_i , Y_e and M_e are bounded. Premultiplying (6) by the adjugate matrix of M_e we get q scalar regressors of the form

$$Y_i(t) = \phi(t)\theta_i \quad (8)$$

with $i \in \bar{q} := \{1, 2, \dots, q\}$, where Y_i is the i –th component of the vector $Y : \mathbb{R}_+ \rightarrow \mathbb{R}^q$

$$Y(t) := \text{adj}\{M_e(t)\}Y_e(t), \quad (9)$$

and we defined the determinant of M_e as

$$\phi(t) := \det\{M_e(t)\}. \quad (10)$$

The estimation of the parameters θ_i from the scalar regression form (8) can be easily carried out via

$$\dot{\hat{\theta}}_i = \gamma_i \phi(Y_i - \phi \hat{\theta}_i), \quad i \in \bar{q}, \quad (11)$$

with adaptation gains $\gamma_i > 0$. From (8) it is clear that the latter equations are equivalent to

$$\dot{\tilde{\theta}}_i = -\gamma_i \phi^2 \tilde{\theta}_i, \quad i \in \bar{q}. \quad (12)$$

Solving this simple scalar differential equation we conclude that

$$\phi(t) \notin \mathcal{L}_2 \implies \lim_{t \rightarrow \infty} \tilde{\theta}_i(t) = 0. \quad (13)$$

²To simplify the presentation in the sequel we will neglect the ϵ_t terms, which will be incorporated in the analysis later.

The derivations above establish the following proposition.

Proposition 1: Consider the q -dimensional linear regression (1) where $y : \mathbb{R}_+ \rightarrow \mathbb{R}$ and $m : \mathbb{R}_+ \rightarrow \mathbb{R}^q$ are known, bounded functions of time and $\theta \in \mathbb{R}^q$ is the vector of unknown parameters. Introduce $q - 1$ linear, \mathcal{L}_∞ -stable operators $H_i : \mathcal{L}_\infty \rightarrow \mathcal{L}_\infty$, $i \in \{1, 2, \dots, q - 1\}$ verifying (5). Define the vector Y_e and the matrix M_e as given in (7). Consider the estimator (11) with ϕ and Y_i defined in (10) and (9), respectively. The implication (13) holds. $\square\square\square$

Remark 3: It is important to underscore that for any matrix $A \in \mathbb{R}^{q \times q}$ holds $\text{adj}\{A\}A = \det\{A\}I_q$, even if A is not full rank, [16].

Remark 4: If we take into account the presence of the exponentially decaying terms ϵ_t in the filtering operations the error equation (12) becomes $\dot{\tilde{\theta}}_i = -\gamma_i \phi^2 \tilde{\theta}_i + \epsilon_t$, $i \in \bar{q}$. The analysis of this equation, which establishes (13), may be found in Lemma 1 of [4].

C. Discussion

Two natural questions arise at this point.

- Q1. Is the condition $\phi(t) \notin \mathcal{L}_2$ weaker than $m(t) \in \text{PE}$?
 Q2. Given a regressor $m(t) \notin \text{PE}$ how to select the operators H_i to enforce the condition $\phi(t) \notin \mathcal{L}_2$?

Regarding the question Q1, the following remarks are in order.

Remark 5: In Remark 7 the example of $m(t) \notin \text{PE}$ is given, such that $\phi(t) \notin \mathcal{L}_2$. Next we present an example, such that $m(t) \in \text{PE}$, but $\phi(t) \in \mathcal{L}_2$.

Consider the regressor $m(t) := [\sin(t) \ \cos(t)]^\top$ and the operator $H(p) = \frac{c(p+1)}{p^2+p+2}$, where $c > 0$. Note that for unit frequency the operator H provides zero phase shift and the magnitude gain c . Thus $m_{1f}(t) = c \sin(t)$, $m_{2f}(t) = c \cos(t)$ and

$$M_e(t) = \begin{bmatrix} \sin(t) & \cos(t) \\ c \sin(t) & c \cos(t) \end{bmatrix}.$$

Obviously, $m(t) \in \text{PE}$, but $\det\{M_e(t)\} \equiv 0$ and $\phi(t) \in \mathcal{L}_2$.

Remark 6: From definition (4) it is clear that the PE condition is a requirement imposed on the *minimal* eigenvalue of the matrix as illustrated by the equivalence $\lambda_{\min} \left\{ \int_t^{t+T} m(s)m^\top(s)ds \right\} \geq \delta > 0 \iff m(t) \in \text{PE}$, where $\lambda_{\min}\{\cdot\}$ denotes the minimal eigenvalue. On the other hand, the condition $\phi(t) \notin \mathcal{L}_2$ is a restriction on *all* eigenvalues of the matrix M_e . Indeed, this is clear recalling that the determinant of a matrix is the product of all its eigenvalues and that for any two bounded signals $a, b : \mathbb{R}_+ \rightarrow \mathbb{R}$ we have $a(t)b(t) \notin \mathcal{L}_2 \implies a(t) \notin \mathcal{L}_2$ and $b(t) \notin \mathcal{L}_2$. Consequently, a necessary condition for parameter convergence of the

estimators (11) is that all eigenvalues of the matrix M_e are not square integrable.

To provide a (partial) answer to the question Q2 above let us consider the simplest case of $q = 2$ with $m = \text{col}(m_1, m_2)$. In this case

$$\phi = m_1 m_{2f} - m_{1f} m_2. \quad (14)$$

The proposition below identifies a class of regressors $m(t) \notin \text{PE}$ but $\phi(t) \notin \mathcal{L}_2$ for the case of H a simple LTI filter.

Proposition 2: Define the set of differentiable functions

$$\mathcal{G} := \{g : \mathbb{R}_+ \rightarrow \mathbb{R} \mid g(t) \in \mathcal{L}_\infty, \dot{g}(t) \in \mathcal{L}_\infty, \dot{g}(t) \notin \mathcal{L}_2, \lim_{t \rightarrow \infty} g(t) = \lim_{t \rightarrow \infty} \dot{g}(t) = 0\}$$

For all $g \in \mathcal{G}$ the regressor $m(t) = [1, g + \dot{g}]^\top \notin \text{PE}$. Let the operator H be defined as

$$[H(\cdot)](t) = \left[\frac{1}{p+1}(\cdot) \right](t).$$

The function ϕ defined in (14) verifies $\phi(t) \notin \mathcal{L}_2$.

Remark 7: An example of a function $g \in \mathcal{G}$ is $g(t) = \sin(t)(1+t)^{-\frac{1}{2}}$. The corresponding regressor is

$$m(t) = \begin{bmatrix} 1 \\ \frac{\sin t + \cos t}{(1+t)^{\frac{1}{2}}} - \frac{\sin t}{2(1+t)^{\frac{3}{2}}} \end{bmatrix}. \quad (15)$$

Remarks 5, 6, and 7 illustrate the fact that the new convergence condition $\phi(t) \notin \mathcal{L}_2$ proposed in the paper radically differ from the standard PE requirement $m(t) \in \text{PE}$.

D. Simulation results

We first evaluate the performance of the classical parameters estimator (2) with $m(t)$ given by (15). From the analysis of Subsection II-A we know that the LTV system (3) is stable, but it is not *exponentially* stable since $m(t) \notin \text{PE}$, and PE is a necessary condition for exponential stability.

The transient behavior of the parameter errors $\tilde{\theta}(t)$ with $\Gamma = \gamma I_2$ and $\theta = \text{col}(-3, 3)$ is shown in Fig. 1 for $\tilde{\theta}(0) = \text{col}(3, -3)$, $\gamma = 3$ and $\gamma = 10$. It is worth noting that it is not possible to conclude from the simulations whether $\tilde{\theta}(t)$ converges to zero asymptotically or not. The plots show that convergence has not been achieved even after a reasonably long period of 500. The graphs also show that increasing γ that, in principle, should speed-up the convergence, makes the situation even worse, cf. Fig. 1 (a) and (b). If the adaptation gain is taken as $\Gamma = \text{diag}\{\gamma_1, \gamma_2\}$ it is possible to improve the transient performance, but this requires a time-consuming, trial-and-error tuning stage that is always undesirable.

Next we study performance of the DREM estimator (11) with the same $m(t)$, H defined in Proposition 2 and $\theta = \text{col}(-3, 3)$. The transient behavior of $\tilde{\theta}(t)$ is given in Fig.

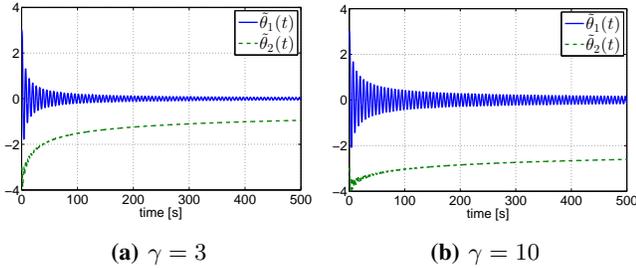


Fig. 1: Transient performance of the parameter errors $\tilde{\theta}(t)$ for the gradient estimator (2) with $m(t)$ given by (15), $\tilde{\theta}(0) = \text{col}(3, -3)$ and $\Gamma = \gamma I_2$, $\gamma = 3, 10$.

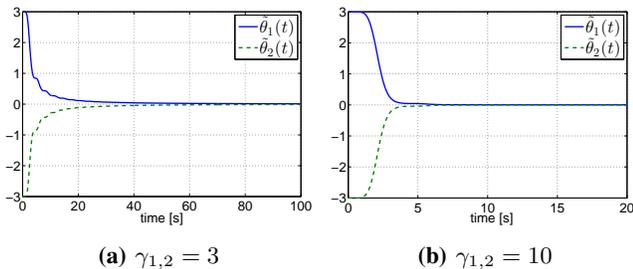


Fig. 2: Transient performance of the parameter errors $\tilde{\theta}(t)$ for the DREM estimator (11) with $m(t)$ given by (15), H defined in Proposition 2, $\tilde{\theta}(0) = \text{col}(3, -3)$ and $\gamma_{1,2} = 3, 10$.

2 for $\tilde{\theta}(0) = \text{col}(3, -3)$, $\gamma_{1,2} = 3$ and $\gamma_{1,2} = 10$. The simulations illustrate significant performance improvement both in oscillatory behavior and in convergence speed—notice the difference in time scales. Moreover, tuning of the gains γ_i in the DREM estimator is straightforward, cf. Fig. 2 (a) and (b).

III. PARAMETER ESTIMATION OF “PARTIALLY” MONOTONIC REGRESSIONS

In this section we propose to use the DREM technique for *nonlinearly* parameterised regressions with factorisable nonlinearities. In contrast with [13], we consider the case where *some*—but not all—of the functions verify a monotonicity condition. The main objective is to generate a new regressor that contains only these “good” nonlinearities.

We consider factorisable regressions of the form

$$\mathbf{y}(t) = \mathbf{m}(t)\psi(\theta), \quad (16)$$

where $\mathbf{y} : \mathbb{R}_+ \rightarrow \mathbb{R}^n$ and $\mathbf{m} : \mathbb{R}_+ \rightarrow \mathbb{R}^{n \times p}$ contain measurable functions, the mapping $\psi : \mathbb{R}^q \rightarrow \mathbb{R}^p$ is known and $\theta \in \mathbb{R}^q$ is the *unknown* parameter vector. It is clear that the nonlinear regression (16) can be “transformed” into a linear one defining the vector $\eta := \psi(\theta)$ to which the standard gradient estimator

$$\dot{\hat{\eta}} = \Gamma \mathbf{m}^\top (\mathbf{y} - \mathbf{m}\hat{\eta}) \quad (17)$$

can be applied. However, overparametrization suffers from well-known shortcomings, cf. [2], [3], [13].

A. Main result

To state the main result of this section we make the following assumption.

Assumption 1: Consider the regression form (16), where

$$n < p. \quad (18)$$

There are q functions ψ_i that, reordering the outputs y_i , we arrange in a vector $\psi_g : \mathbb{R}^q \rightarrow \mathbb{R}^q$,

$$q < p, \quad (19)$$

verifying

$$P \nabla \psi_g(\theta) + [\nabla \psi_g(\theta)]^\top P \geq \rho_0 I_q > 0, \quad (20)$$

for some positive definite matrix $P \in \mathbb{R}^{q \times q}$. In this case, the function ψ_g verifies [17]

$$(\mathbf{a} - \mathbf{b})^\top P [\psi_g(\mathbf{a}) - \psi_g(\mathbf{b})] \geq \rho_1 |\mathbf{a} - \mathbf{b}|^2, \quad \forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^q, \quad (21)$$

for some $\rho_1 > 0$.

Consistent with Assumption 1 we rewrite (16) as

$$\mathbf{y}_N(t) = \begin{bmatrix} \mathbf{m}_g(t) & \mathbf{m}_b(t) \end{bmatrix} \begin{bmatrix} \psi_g(\theta) \\ \psi_b(\theta) \end{bmatrix}, \quad (22)$$

where $\mathbf{y}_N : \mathbb{R}_+ \rightarrow \mathbb{R}^n$ is the reordered output vector, $\mathbf{m}_g : \mathbb{R}_+ \rightarrow \mathbb{R}^{n \times q}$, $\mathbf{m}_b : \mathbb{R}_+ \rightarrow \mathbb{R}^{n \times (p-q)}$, $\psi_g : \mathbb{R}^q \rightarrow \mathbb{R}^q$ and $\psi_b : \mathbb{R}^q \rightarrow \mathbb{R}^{p-q}$.

As will become clear below DREM must accomplish two tasks, on one hand, generate a regression without \mathbf{m}_b . On the other hand, to be able to relax the PE condition, the new regressor matrix should be square (or tall). If (19) does not hold, *all* functions ψ_i , $i = 1, \dots, p$, satisfy the monotonicity condition and there is no need to eliminate any one of them. On the other hand, if (18) is not satisfied a square regressor without the “bad” part of the regressor ψ_b can be created without the introduction of the operators H_i . Indeed, if $n = p$ the matrix \mathbf{m}_b is tall and it admits a full-rank left annihilator $\mathbf{m}_b^\perp : \mathbb{R}_+ \rightarrow \mathbb{R}^{q \times n}$. Moreover, the new regressor matrix $\mathbf{m}_b^\perp \mathbf{m}_g$ is square. A similar situation arises if $n > p$.

Following DREM we introduce n_f operators, apply them to some rows of (22) and pile all the regression forms to get

$$\begin{bmatrix} \mathbf{y}_N \\ \mathbf{y}_{N_f} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_g & \mathbf{M}_b \end{bmatrix} \begin{bmatrix} \psi_g(\theta) \\ \psi_b(\theta) \end{bmatrix}. \quad (23)$$

where we defined the matrices $\mathbf{M}_g : \mathbb{R}_+ \rightarrow \mathbb{R}^{(n+n_f) \times q}$, $\mathbf{M}_b : \mathbb{R}_+ \rightarrow \mathbb{R}^{(n+n_f) \times (p-q)}$

$$\mathbf{M}_g := \begin{bmatrix} \mathbf{m}_g \\ \mathbf{m}_{gf} \end{bmatrix}, \quad \mathbf{M}_b := \begin{bmatrix} \mathbf{m}_b \\ \mathbf{m}_{bf} \end{bmatrix}. \quad (24)$$

To select the number n_f of operators we notice that the matrix to be eliminated, that is \mathbf{M}_b , is of dimension $(n + n_f) \times (p - q)$. Therefore, to have a left annihilator for it with q rows, which is needed to make the new regressor square, we must fix $n_f = p - n$. Define

$$\Phi := \mathbf{M}_b^\perp \mathbf{M}_g. \quad (25)$$

Multiplying on the left by $\text{adj}\{\Phi\}\mathbf{M}_b^\perp$ the equation (23) yields the desired regressor form

$$\mathbf{Y} = \det\{\Phi\}\psi_g(\theta), \quad (26)$$

where

$$\mathbf{Y} := \text{adj}\{\Phi\}\mathbf{M}_b^\perp \begin{bmatrix} \mathbf{y}_N \\ \mathbf{y}_{N_f} \end{bmatrix}. \quad (27)$$

We propose the estimator

$$\dot{\hat{\theta}} = \det\{\Phi\}\Gamma P[\mathbf{Y} - \det\{\Phi\}\psi_g(\hat{\theta})], \quad (28)$$

with $\Gamma \in \mathbb{R}^{q \times q}$, $\Gamma > 0$, and we are in position to present the main result of this section

Proposition 3: Consider the nonlinearly parameterised factorisable regression (22) satisfying Assumption 1. Introduce $p-n$ linear, \mathcal{L}_∞ -stable operators $H_i : \mathcal{L}_\infty \rightarrow \mathcal{L}_\infty$, $i \in \{1, 2, \dots, p-n\}$ verifying (5). Define the matrices \mathbf{M}_g , \mathbf{M}_b as given in (24). Consider the estimator (28) with Φ and \mathbf{Y} defined in (25), (27) and $\mathbf{M}_b^\perp : \mathbb{R}_+ \rightarrow \mathbb{R}^{q \times p}$ a full-rank left annihilator of \mathbf{M}_b . The following implication holds

$$\det\{\Phi(t)\} \notin \mathcal{L}_2 \implies \lim_{t \rightarrow \infty} |\tilde{\theta}(t)| = 0.$$

Moreover, if $\det\{\Phi(t)\} \geq \kappa > 0$, then $|\tilde{\theta}(t)|$ tends to 0 exponentially fast.

Remark 8: In [13] it is shown that the local verification of the monotonicity condition (20) reduces to a linear matrix inequality (LMI) test provided some prior knowledge on the parameters is available. More precisely, assume $\theta \in \Theta \subset \mathbb{R}^q$, with

$$\Theta := \{\theta \in \mathbb{R}^q \mid \theta_i \in [\theta_i^m, \theta_i^M] \subset \mathbb{R}\}.$$

The quadratic approximation of the mapping $\psi_g(\theta)$ verifies (20) if and only if the LMI

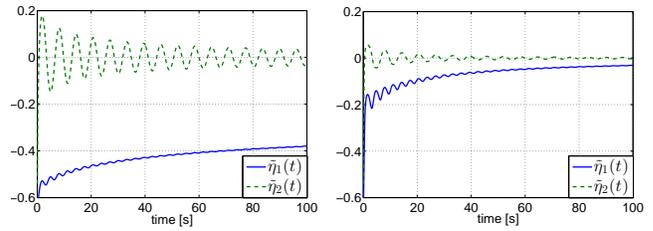
$$P\nabla\psi_g(v_i) + [\nabla\psi_g(v_i)]^\top P > 0, \quad i = 1, \dots, (2^q)^q$$

is feasible, where the vectors $v_i \in \Theta$ are computable from the vertices of Θ .

B. An example

Consider the simplest scalar case of $n = 1$, $p = 2$ and $q = 1$. The regression (16) becomes

$$\mathbf{y}(t) = \begin{bmatrix} \mathbf{m}_1(t) & \mathbf{m}_2(t) \end{bmatrix} \begin{bmatrix} \psi_1(\theta) \\ \psi_2(\theta) \end{bmatrix}, \quad (29)$$



$$(a) \Gamma = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

$$(b) \Gamma = \begin{bmatrix} 50 & 0 \\ 0 & 5 \end{bmatrix}$$

Fig. 3: Transient behaviour of the errors $\tilde{\eta}(t)$ for the over-parameterized parameter estimator (17) for different gains; $\tilde{\eta}(0) = 0$ and $\tilde{\eta}(0) = -\psi(\theta) = -\psi(1) \approx [-0.63, -0.54]^\top$.

where $\mathbf{y} : \mathbb{R}_+ \rightarrow \mathbb{R}$, $\mathbf{m}_i : \mathbb{R}_+ \rightarrow \mathbb{R}$ and $\psi_i : \mathbb{R} \rightarrow \mathbb{R}$, for $i = 1, 2$. Assume that $\psi_1(\theta)$ is *strongly monotonically increasing*, that is, $\psi_1'(\theta) \geq \rho_0 > 0$, and that inequality (21) is satisfied with $P = 1$ and ψ_g chosen as ψ_1 .

Following the DREM procedure we apply an operator H to (29) and pile-up the two regressions as

$$\begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}_f(t) \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1(t) & \mathbf{m}_2(t) \\ \mathbf{m}_{1f}(t) & \mathbf{m}_{2f}(t) \end{bmatrix} \begin{bmatrix} \psi_1(\theta) \\ \psi_2(\theta) \end{bmatrix}.$$

Multiplying on the left the equation above by the row vector $[\mathbf{m}_{2f} \quad -\mathbf{m}_2]$ we get the desired regression involving only ψ_1 , namely, $\mathbf{Y}(t) = \Phi(t)\psi_1(\theta)$, where we defined the signals

$$\mathbf{Y} := \mathbf{m}_{2f}\mathbf{y} - \mathbf{m}_2\mathbf{y}_f, \quad \Phi := \mathbf{m}_{2f}\mathbf{m}_1 - \mathbf{m}_2\mathbf{m}_{1f}. \quad (30)$$

From Proposition 3 we conclude that the estimator

$$\dot{\hat{\theta}} = \gamma\Phi[\mathbf{Y} - \Phi\psi_1(\hat{\theta})], \quad \gamma > 0 \quad (31)$$

ensures that $\tilde{\theta}(t) \rightarrow 0$ as $t \rightarrow \infty$ if $\Phi(t) \notin \mathcal{L}_2$.

The proposition below identifies a class of regressors $\mathbf{m}(t) \notin \text{PE}$ but $\Phi(t) \notin \mathcal{L}_2$ for a simple delay operator.

Proposition 4: The regressor

$$\mathbf{m}(t) = \begin{bmatrix} \frac{\sin(t)}{\sqrt{t+2\pi}} & 1 \end{bmatrix} \notin \text{PE}.$$

Let the operator H be the delay operator, that is, $(\cdot)_f(t) = (\cdot)(t-d)$ and $d \in [\frac{\pi}{2}, \frac{3\pi}{2}]$. The function Φ defined in (30) verifies $\Phi(t) \notin \mathcal{L}_2$.

As an example consider the regression

$$\mathbf{y}(t) = \mathbf{m}_1(t)(\theta - e^{-\theta}) + \mathbf{m}_2(t)\cos(\theta),$$

which clearly satisfies condition (21) with $\psi_g(\theta)$ chosen as $\psi_1(\theta) := \theta - e^{-\theta}$. Simulations of the overparametrized estimator (17) with $\theta = 1$ are given in Fig. 3 while simulations of the DREM estimator (31) with the delay operator H defined in Proposition 4 with $d = 2$ are shown in Fig. 4.

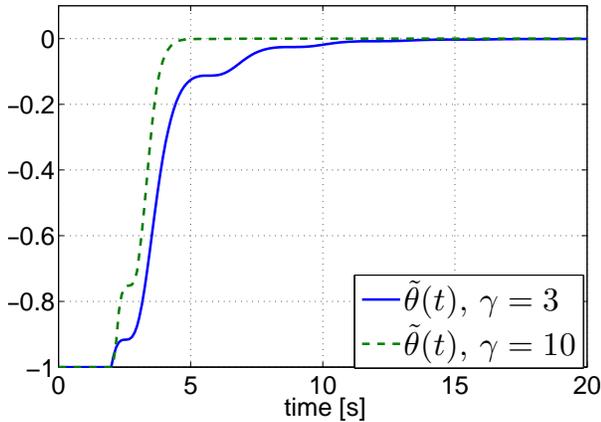


Fig. 4: Transients of the error $\tilde{\theta}(t)$ of the DREM parameters estimator (31) with different gains; $\tilde{\theta}(0) = -1$.

IV. CONCLUDING REMARKS AND FUTURE RESEARCH

A procedure to generate new regression forms for which we can design parameter estimators with enhanced performance has been proposed. The procedure has been applied to linear regressions yielding new estimators whose parameter convergence can be established without invoking the usual, hardly verifiable, PE condition. Instead, it is required that the new regressor vector is not square integrable, which is different than PE of the original regressor. For nonlinearly parameterised regressions with monotonic nonlinearities the procedure allows to treat cases when only some of the nonlinearities verify this monotonicity condition. Similarly to the case of linear regressions, convergence is ensured if the determinant of the new regressor is not square integrable.

The design procedure includes many degrees of freedom to verify the aforementioned convergence condition. Current research is under way to make more systematic the choice of this degrees of freedom. It seems difficult to achieve this end at the level of generality presented in the paper. Therefore, we are currently considering more “structured” situations, for instance, when the original regression form comes from classes of physical dynamical systems or for a practical application. Preliminary calculations for the problem of current-voltage characteristic of photovoltaic cells—which depend nonlinearly on some unknown parameters—are encouraging and we hope to be able to report the results soon.

REFERENCES

- [1] S. Aranovskiy, B. A., R. Ortega, and P. A., “Performance enhancement of parameter estimators via dynamic regressor extension and mixing,” *arXiv*, 2015, <http://arxiv.org/abs/1509.02763>.
- [2] L. Ljung, *System identification: theory for the user*. New Jersey: Prentice Hall, 1987.
- [3] S. Sastry and M. Bodson, *Adaptive control: stability, convergence and robustness*. Courier Dover Publications, 2011.
- [4] S. Aranovskiy, A. Bobtsov, A. Pyrkin, R. Ortega, and A. Chaillet, “Flux and position observer of permanent magnet synchronous motors with relaxed persistency of excitation conditions,” in *Proc. 1st IFAC Conf. on Modelling, Identification and Control of Nonlinear Systems*, Saint-Petersburg, Russia, Jun. 2015, pp. 311–316.
- [5] G. Chowdhary and E. Johnson, “Concurrent learning for convergence in adaptive control without persistency of excitation,” in *Proc. 49th IEEE Conference on Decision and Control*, Atlanta, GA, Dec. 2010, pp. 3674–3679.
- [6] D. Efimov and A. Fradkov, “Design of impulsive adaptive observers for improvement of persistency of excitation,” *International Journal of Adaptive Control and Signal Processing*, vol. 29, no. 6, pp. 765–782, 2015.
- [7] Y. Pan and H. Yu, “Composite learning from adaptive dynamic surface control,” *IEEE Transactions on Automatic Control*, vol. PP, no. 99, pp. 1–1, 2015.
- [8] H. Jain, “Graduated persistent excitation and steady state margins for adaptive systems,” Master’s thesis, Massachusetts Institute of Technology, Cambridge, 2007.
- [9] R. Mishkov and S. Darmonski, “Exact parameter estimation without persistent excitation in nonlinear adaptive control systems,” in *Proc. 7th National Conference on Process Automation in the Food and Biotechnology Industries*, vol. 60, Plovdiv, Bulgaria, Oct. 2013, pp. 100–106.
- [10] Y. Pan, L. Pan, and H. Yu, “Composite learning control with application to inverted pendulums,” *arXiv:1507.07844*, Jul. 2015.
- [11] A. Pyrkin, A. Bobtsov, S. Kolyubin, A. Vedyakov, O. Borisov, V. Gromov, A. Margun, and D. Bazylev, “Fast compensation of unknown multiharmonic disturbance for nonlinear plant with input delay,” in *Proc. 11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, Caen, France, Jul. 2013, pp. 546–551.
- [12] X. Liu, R. Ortega, H. Su, and J. Chu, “Immersion and invariance adaptive control of nonlinearly parameterized nonlinear systems,” *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2209–2214, 2010.
- [13] —, “On adaptive control of nonlinearly parameterized nonlinear systems: Towards a constructive procedure,” *Systems & Control Letters*, vol. 60, no. 1, pp. 36 – 43, 2011.
- [14] I. Tyukin, D. Prokhorov, and C. van Leeuwen, “Adaptation and parameter estimation in systems with unstable target dynamics and nonlinear parametrization,” *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1543–1559, 2007.
- [15] S. Dasgupta and Y.-F. Huang, “Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise,” *IEEE Trans. Inf. Theory*, vol. 33, no. 3, pp. 383–392, 1987.
- [16] P. Lancaster and M. Tismenetsky, *The theory of matrices: with applications*. Academic press, 1985.
- [17] A. Pavlov, A. Pogromsky, N. van de Wouw, and H. Nijmeijer, “Convergent dynamics, a tribute to Boris Pavlovich Demidovich,” *Systems & Control Letters*, vol. 52, no. 3–4, pp. 257 – 261, 2004.