Performance Enhancement of Parameter Estimators via Dynamic Regressor Extension and Mixing

Stanislav Aranovskiy, Alexey Bobtsov, Romeo Ortega, Anton Pyrkin

Performance Enhancement of Parameter Estimators via Dynamic Regressor Extension and Mixing*

Stanislav Aranovskiy[1,2], *Member, IEEE,* Alexey Bobtsov[2], *Senior Member, IEEE,* Romeo Ortega[3], *Fellow Member, IEEE,* Anton Pyrkin[2], *Member, IEEE*

***Abstract*—A new procedure to design parameter estimators with enhanced performance is proposed in the paper. For classical linear regression forms it 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.**

***Index Terms*—Estimation, persistent excitation, nonlinear regressor, monotonicity**

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]. To the best of the authors' knowledge there is no systematic way to conclude asymptotic stability for this system without this assumption, which is rarely verified in applications. Relaxation of the PE condition is a challenging theoretical problem and many research works have been devoted to it in various scenarios, see *e.g.*, [4]–[9] 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 the dependence with respect to the parameters exhibit some *monotonicity* properties; see [10]–[12]. 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*.

Notation For $x \in \mathbb{R}^n$, $|x|^2 = x^\top x$. All functions are assumed sufficiently smooth. For functions of scalar argument $g : \mathbb{R} \to \mathbb{R}^s$, g' denotes its first order derivative. For functions $V : \mathbb{R}^n \to \mathbb{R}$ we define the operator $\nabla V := (\frac{\partial V}{\partial x})^\top$.

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

$$y(t) = m^\top(t)\theta, \tag{1}$$

[1]Stanislav Aranovskiy is with the School of Automation, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou, Zhejiang 310018, P.R.China

[2]Stanislav Aranovskiy, Alexey Bobtsov and Anton Pyrkin are with the Department of Control Systems and Informatics, ITMO University, Kronverkskiy av. 49, Saint Petersburg, 197101, Russia : aranovskiysv@niuitmo.ru

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

*An abridged version of the paper was presented at American Control Conference 2016 [1].

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

where[1] $y:\mathbb{R}_+\to\mathbb{R}$ and $m:\mathbb{R}_+\to\mathbb{R}^q$ are known, bounded functions of time and $\theta\in\mathbb{R}^q$ is the vector of unknown parameters. The standard gradient estimator

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

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

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

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) globally exponentially stable if and only if the regressor vector m is PE, that is, if

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

for some $T,\delta>0$ and for all $t\geq 0$, which will be denoted as $m(t)\in$ PE. If $m(t)\notin$ 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. In spite of some erroneous claims [13], it is well known that the PE conditions for the gradient estimator presented above and more general estimators—like (weighted) least squares—exactly coincide [14]. Since the interest in the paper is to relax the PE condition, and in the interest of brevity, attention is restricted to the simple gradient estimator.

B. 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, $\mathcal{L}_\infty$–stable* operators $H_i:\mathcal{L}_\infty\to\mathcal{L}_\infty,\ i\in\{1,2,\ldots,q-1\}$, whose output, for any bounded input, may be decomposed as

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

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 regression[2]

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

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

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

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

$$Y_e(t)=M_e(t)\theta, \tag{6}$$

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

$$Y_e:=\mathrm{col}(y,y_{f_1},\cdots,y_{f_{q-1}}),\ M_e^\top:=\begin{bmatrix} m & m_{f_1} & \cdots & m_{f_{q-1}}\end{bmatrix}. \tag{7}$$

Premultiplying (6) by the *adjunct matrix* of M_e we get q scalar regressors of the form

$$Y_i(t)=\phi(t)\theta_i \tag{8}$$

with $i\in\bar{q}:=\{1,2,\ldots,q\}$, where we defined the determinant of M_e as

$$\phi(t):=\det\{M_e(t)\}. \tag{9}$$

and the vector $Y:\mathbb{R}_+\to\mathbb{R}^q$

$$Y(t):=\mathrm{adj}\{M_e(t)\}Y_e(t). \tag{10}$$

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

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

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

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

Solving this simple scalar differential equation we conclude that

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

The derivations above establish the following proposition.

Proposition 1. Consider the q–dimensional linear regression (1) where $y:\mathbb{R}_+\to\mathbb{R}$ and $m:\mathbb{R}_+\to\mathbb{R}^q$ are known, bounded functions of time and $\theta\in\mathbb{R}^q$ is the vector of unknown parameters. Introduce $q-1$ linear, $\mathcal{L}_\infty$–stable operators $H_i:\mathcal{L}_\infty\to\mathcal{L}_\infty,\ i\in\{1,2,\ldots,q-1\}$ 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 (9) and (10), respectively. The *equivalence* (13) holds.

□□□

Remark 2. It is important to underscore that for any matrix $A\in\mathbb{R}^{q\times q}$ we have that $\mathrm{adj}\{A\}A=\det\{A\}I_q$, even if A is *not full rank* [15].

Remark 3. 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 may be found in Lemma 1 of [4] where it is shown that (13) still holds.

C. Discussion

Two natural questions arise at this point.

Q1. Is the condition $\phi(t) \notin \mathcal{L}_2$ weaker than $m(t) \in$ PE?

Q2. Given a regressor $m(t) \notin$ PE is it possible to select operators H_i to enforce the condition $\phi(t) \notin \mathcal{L}_2$?

Regarding question Q2 notice that

$$M_e^\top = \begin{bmatrix} 1 & H_1 & \cdots & H_{q-1} \end{bmatrix} m,$$

which is a linear operation. However, computing the determinant of M_e is nonlinear—hence the question is far from obvious. A (partial) answer to it is given in the subsection II-D. Regarding the question Q1 we underline the following observation that underscores the different nature of the two conditions. 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 $\lambda_{\min}\left\{\int_t^{t+T} m(s)m^\top(s)ds\right\} \geq \delta > 0 \iff m(t) \in$ 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 $a, b : \mathbb{R}_+ \to \mathbb{R}$ 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.

D. An example

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

$$\phi = m_1 m_{2f} - m_{1f} m_2. \tag{14}$$

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

Fact 1. Define the set of differentiable functions

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

For all $g \in \mathcal{G}$ the regressor $m(t) = [1,\ g + \dot{g}]^\top \notin$ 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 $\phi(t) \notin \mathcal{L}_2$.

Proof. The fact that $m(t) \notin$ PE is obvious because $\lim_{t\to\infty} m_2(t) = 0$. Now, we have that $m_{1f} = 1 + \epsilon_t$ and from the filter equations we get $\dot{m}_{2f} = -m_{2f} + m_2$. On the other hand, from the definition of m we have $\dot{g} = -g + m_2$. Substracting these two equations we get

$$\frac{d}{dt}(m_{2f} - g) = -(m_{2f} - g),$$

consequently $m_{2f} = g + \epsilon_t$. Replacing these expressions in (14) yields

$$\begin{aligned} \phi &= m_{2f} - (1 + \epsilon_t) m_2 \\ &= (g + \epsilon_t) - (1 + \epsilon_t)(g + \dot{g}) = -\dot{g} + \epsilon_t, \end{aligned}$$

where we have used the fact that $g(t) \in \mathcal{L}_\infty$ and $\dot{g}(t) \in \mathcal{L}_\infty$ to obtain the last equation. This completes the proof. □□□

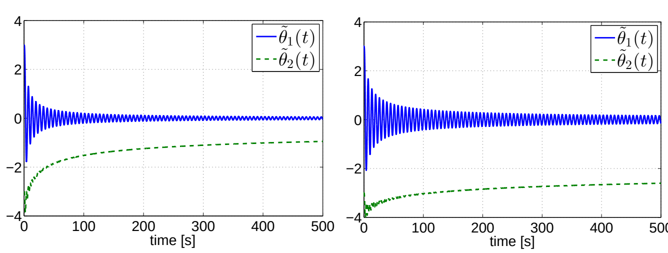
Remark 4. An example of a function $g \in \mathcal{G}$ is $g(t) = \sin(t)(1+t)^{-\frac{1}{2}}$. The corresponding regressor is

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

E. 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 $m(t) \notin$ 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.

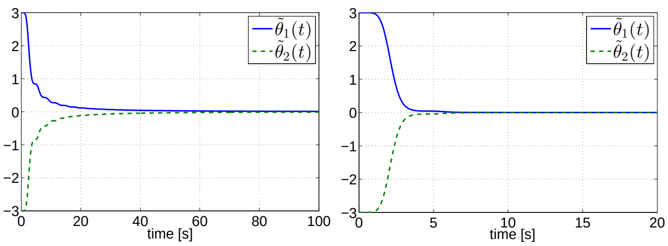


(a) $\gamma = 3$ **(b)** $\gamma = 10$

Fig. 1: Transient performance of the parameter errors $\tilde{\theta}(t)$ for the gradient estimator (2) with $m(t)$ given by (15).

Next we study performance of the DREM estimator (11) with the same $m(t)$ and $\theta = \text{col}(-3, 3)$. The transient behavior of $\tilde{\theta}(t)$ is given in Fig. 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, since the role of the gains γ_i in the DREM estimator is obvious, their tuning is straightforward, cf. Fig. 2 (a) and (b).



(a) $\gamma_{1,2} = 3$ **(b)** $\gamma_{1,2} = 10$

Fig. 2: Transient performance of the parameter errors $\tilde{\theta}(t)$ for the DREM estimator (11) with $m(t)$ given by (15).

III. Parameter Estimation of "Partially" Monotonic Regressions

In this section we propose to use the DREM technique for *nonlinearly* parameterised regressions with factorisable non-linearities. In contrast with [11], 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

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

where $\mathbf{y} : \mathbb{R}_+ \to \mathbb{R}^n$ and $\mathbf{m} : \mathbb{R}_+ \to \mathbb{R}^{n\times p}$ contain measurable functions, the mapping $\psi : \mathbb{R}^q \to \mathbb{R}^p$ 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

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

can be applied. However, overparametrization suffers from well–known shortcomings, *cf.* [2], [3], [11].

A. Main result

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

Assumption 1. Consider the regression form (16). There are q functions ψ_i that, reordering the outputs y_i, we arrange in a vector $\psi_g : \mathbb{R}^q \to \mathbb{R}^q$, verifying

$$P\nabla\psi_g(\theta) + [\nabla\psi_g(\theta)]^\top P \geq \rho_0 I_q > 0, \tag{18}$$

for some positive definite matrix $P \in \mathbb{R}^{q\times q}$.

Consistent with Assumption 1 we rewrite (16) as

$$\mathbf{y}_N(t) = \begin{bmatrix} \mathbf{m}_g(t) & \mathbf{m}_b(t) \end{bmatrix} \begin{bmatrix} \psi_g(\theta) \\ \psi_b(\theta) \end{bmatrix}, \tag{19}$$

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

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). Given these tasks, to obtain a sensible problem formulation the following assumption is imposed.

Assumption 2. The regression (19) satisfies

$$q < p \tag{20}$$
$$n < p. \tag{21}$$

If (20) does not hold *all* functions ψ_i, $i = 1, \ldots, p$, satisfy the monotonicity condition and there is no need to eliminate any one of them. On the other hand, if (21) 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 $\mathbf{m}_b^\perp : \mathbb{R}_+ \to \mathbb{R}^{q\times n}$. 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 (19) and pile all the regression forms to get

$$\begin{bmatrix} \mathbf{y}_N \\ \mathbf{y}_{Nf} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_g & \mathbf{M}_b \end{bmatrix} \begin{bmatrix} \psi_g(\theta) \\ \psi_b(\theta) \end{bmatrix}. \tag{22}$$

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

$$\mathbf{M}_g := \begin{bmatrix} \mathbf{m}_g \\ \mathbf{m}_{gf} \end{bmatrix}, \; \mathbf{M}_b := \begin{bmatrix} \mathbf{m}_b \\ \mathbf{m}_{bf} \end{bmatrix}. \tag{23}$$

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

$$\Phi := \mathbf{M}_b^\perp \mathbf{M}_g. \tag{24}$$

Multiplying on the left by $\text{adj}\{\Phi\}\mathbf{M}_b^\perp$ the equation (22) yields the desired regressor form

$$\mathbf{Y} = \det\{\Phi\}\psi_g(\theta), \tag{25}$$

where

$$\mathbf{Y} := \text{adj}\{\Phi\}\mathbf{M}_b^\perp \begin{bmatrix} \mathbf{y}_N \\ \mathbf{y}_{Nf} \end{bmatrix}. \tag{26}$$

We propose the estimator

$$\dot{\hat{\theta}} = \det\{\Phi\}\Gamma P[\mathbf{Y} - \det\{\Phi\}\psi_g(\hat{\theta})], \tag{27}$$

with $\Gamma \in \mathbb{R}^{q\times q}$, $\Gamma > 0$. Using (25) the error equation is

$$\dot{\tilde{\theta}} = -\det{}^2\{\Phi\}\Gamma P[\psi_g(\hat{\theta}) - \psi_g(\theta)].$$

To analyse its stability define the Lyapunov function candidate $V(\tilde{\theta}) = \frac{1}{2}\tilde{\theta}^\top\Gamma^{-1}\tilde{\theta}$, whose derivative yields

$$\begin{aligned}\dot{V} &= -\det{}^2\{\Phi\}(\hat{\theta}-\theta)^\top P[\psi_g(\hat{\theta}) - \psi_g(\theta)]\\ &\leq -\det{}^2\{\Phi\}\frac{2\rho_1}{\lambda_{\max}\{\Gamma\}}V.\end{aligned}$$

If the matrix $\Phi(t)$ is full rank and $\det^2\{\Phi(t)\} \geq \kappa > 0$, then

$$\dot{V} \leq -\frac{2\kappa\rho_1}{\lambda_{\max}\{\Gamma\}}V,$$

and *exponential* stability of the error equation is ensured. Otherwise, integrating the inequality yields

$$V(t) \leq e^{-\frac{2\rho_1}{\lambda_{\max}\{\Gamma\}}\int_0^t \det^2\{\Phi(s)\}ds}V(0),$$

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

We are in position to present the main result of this section, whose proof follows from the derivations above.

Proposition 2. Consider the nonlinearly parameterised factorisable regression (19) satisfying Assumptions 1 and 2. Introduce $p-n$ linear, $\mathcal{L}_\infty$–stable operators $H_i : \mathcal{L}_\infty \to \mathcal{L}_\infty,\ i \in \{1,2,\ldots,p-n\}$ verifying (5). Define the matrices $\mathbf{M}_g,\ \mathbf{M}_b$ as given in (23). Consider the estimator (27) with Φ and $\mathbf{Y}$ defined in (24), (26) and $\mathbf{M}_b^\perp : \mathbb{R}_+ \to \mathbb{R}^{q\times p}$ a full–rank left annihilator of $\mathbf{M}_b$. The following implication holds

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

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

B. An example

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

$$\mathbf{y}(t) = \begin{bmatrix}\mathbf{m}_1(t) & \mathbf{m}_2(t)\end{bmatrix}\begin{bmatrix}\psi_1(\theta)\\ \psi_2(\theta)\end{bmatrix}, \tag{28}$$

where $\mathbf{y} : \mathbb{R}_+ \to \mathbb{R}$, $\mathbf{m}_i : \mathbb{R}_+ \to \mathbb{R}$ and $\psi_i : \mathbb{R} \to \mathbb{R}$, for $i = 1, 2$. Assume that $\psi_1(\theta)$ is *strongly monotonically increasing*, that is, $\psi_1'(\theta) \geq \rho_0 > 0$. In this case, the function ψ_1 verifies [16]

$$(a-b)[\psi_1(a) - \psi_1(b)] \geq \rho_1(a-b)^2, \quad \forall a,b \in \mathbb{R}, \tag{29}$$

for some $\rho_1 > 0$.

Following the DREM procedure we apply an operator H to (28) 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 $[\mathbf{m}_{2f}\ \ -\mathbf{m}_2]$ we get the desired regression involving only ψ_1, namely, $\mathbf{Y}(t) = \Phi(t)\psi_1(\theta)$, where we defined the signals

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

From Proposition 2 we conclude that the estimator

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

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

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

Fact 2. 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,

$$(\cdot)_f(t) = (\cdot)(t-d), \quad d \in \left[\frac{\pi}{2},\ \frac{3\pi}{2}\right].$$

The function Φ defined in (30) verifies $\Phi(t) \notin \mathcal{L}_2$.

Proof. The fact that $\mathbf{m}(t) \notin$ PE is obvious because $\mathbf{m}_1(t) \to 0$. Now, the function Φ defined in (30) takes the form

$$\Phi(t) = \frac{\sin(t)}{\sqrt{t+2\pi}} - \frac{\sin(t-d)}{\sqrt{t+2\pi-d}}.$$

Whence

$$\begin{aligned}\Phi^2(t) = &\frac{\sin^2(t)}{t+2\pi} - 2\cos(d)\frac{\sin^2(t)}{\sqrt{t+2\pi}\sqrt{t+2\pi-d}}\\ &+ \frac{\sin^2(t-d)}{t+2\pi-d} + \sin(d)\frac{\sin(2t)}{\sqrt{t+2\pi}\sqrt{t+2\pi-d}},\end{aligned}$$

where some basic trigonometric identities have been used to derive the identity. Note that the first three right hand terms of the last identity are not integrable. Since $\cos(d) \leq 0$ in the admissible range of d the sum of these terms is also not integrable. On the other hand, the last right term verifies

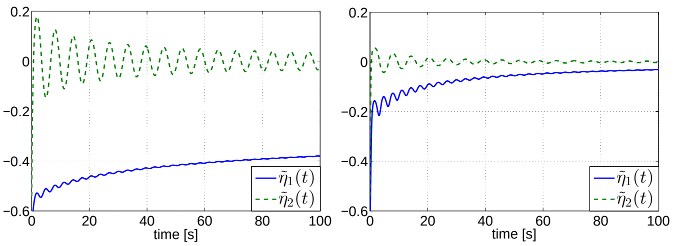
$$\sin(d)\int_0^\infty \frac{\sin(2t)}{\sqrt{t+2\pi}\sqrt{t+2\pi-d}}dt < \infty.$$

Thus, $\Phi(t) \notin \mathcal{L}_2$. □□□

As an example consider the nonlinear regression

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

which clearly satisfies condition (29). Simulations of the overparametrized estimator (17) with $\theta = 1$ are given in Fig. 3 while simulations of the DREM estimator (31) are shown in Fig. 4.



(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 adaptation gains and $\hat{\eta}(0) = 0$.

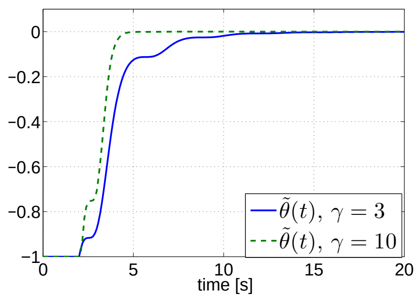


Fig. 4: Transient behaviour of the error $\tilde{\theta}(t)$ for the DREM estimator (31) with different gains and $\hat{\theta}(0) = 0$.

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 from 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. In [17] it is shown that DREM yields a convergent position observer for synchronous motors even in the absence of PE, and in [18] the DREM is used to improve transients performance in a multiple frequency identification task.

References

[1] S. Aranovskiy, A. Bobtsov, R. Ortega, and A. Pyrkin, "Parameters estimation via dynamic regressor extension and mixing," in *American Control Conference*, Boston, MA, USA, 2016.

[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] H. Jain, "Graduated persistent excitation and steady state margins for adaptive systems," Master's thesis, Massachusetts Institute of Technology, Cambridge, 2007.

[8] Y. Pan, L. Pan, and H. Yu, "Composite learning control with application to inverted pendulums," *arXiv:1507.07844*, Jul. 2015.

[9] 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.

[10] 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.

[11] ——, "On adaptive control of nonlinearly parameterized nonlinear systems: Towards a constructive procedure," *Systems & Control Letters*, vol. 60, no. 1, pp. 36 – 43, 2011.

[12] 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.

[13] M. Arteaga and Y. Tang, "Adaptive control of robots with an improved transient performance," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1198–1202, 2002.

[14] 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.

[15] P. Lancaster and M. Tismenetsky, *The theory of matrices: with applications*. Academic press, 1985.

[16] 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.

[17] A. Bobtsov, D. Bazylev, A. Pyrkin, S. Aranovskiy, and R. Ortega, "A robust nonlinear position observer for synchronous motors with relaxed excitation conditions," *International Journal of Control*, no. n/a, p. n/a, 2016, accepted.

[18] S. Aranovskiy, A. Bobtsov, R. Ortega, and A. Pyrkin, "Improved transients in multiple frequencies estimation via dynamic regressor extension and mixing," in *IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, Eindhoven, The Netherlands, 2016.