

## Power noise filtration in DREM

Alexey Bobtsov<sup>1,2</sup>, Stanislav Aranovskiy<sup>3</sup>, Denis Efimov<sup>4</sup>, Anton Pyrkin<sup>1</sup>, Vladimir Vorobev<sup>1</sup>, and Jian Wang<sup>2</sup>

**Abstract**—The problem of estimation in the linear regression model is studied under the hypothesis that the noise is sufficiently small comparing to the regressor. Then the estimation solution is searched for a new regression containing the powers of the unknown parameters and disturbance, where the influence of the latter is attenuated. It is shown that power transformation of a regressor can preserve the excitation under mild assumptions. Possibilities of evaluation of powers of parameters are investigated using the dynamic regression extension and mixing (DREM) method. The performance of the estimators is demonstrated in numerical experiments.

### I. INTRODUCTION

We consider the problem of parameter estimation in a noisy linear regression:

$$\tilde{y}(t) = \tilde{\phi}^\top(t)\tilde{\theta} + \tilde{d}(t), \quad (1)$$

where  $\tilde{y}(t) \in \mathbb{R}$  is the measured output,  $\tilde{\phi}(t) \in \mathbb{R}^p$  is the measured vector regressor,  $\tilde{\theta} \in \mathbb{R}^p$  is unknown constant vector of parameters to be estimated, and  $\tilde{d}(t) \in \mathbb{R}$  is a bounded measurement noise. There are many methods solving this standard estimation problem, see [1]–[5] (to mention a few handbooks), providing different performances of evaluation of the value of  $\tilde{\theta}$ , and being crucially dependent on the excitation properties of the regressor.

Interval excitation is sufficient for identification of  $\tilde{\theta}$  in the noise-free setting, while in the presence of disturbance  $\tilde{d}$ , the persistence excitation requirement is commonly imposed [6], [7], which roughly speaking implies that collecting the measurements and continuing the observation it is possible to improve the estimation precision. The latter condition can be restrictive, and there are many works oriented on its relaxation [8]–[13].

Among the quality requirements for identification algorithms, it is worth to highlight the time of convergence of the estimate to its ideal value, and the estimation error robustness with respect to the perturbation  $\tilde{d}$ . Dealing with the latter issues, different filtering algorithms, often linear, are applied to both sides of (1), in order to improve the estimation accuracy [14]–[16]. Usually, this requires the information about the frequency spectrum of the  $\tilde{d}$  or its stochastic characteristics, and ideally the spectrum separation

(independence) between the regressor  $\tilde{\phi}$  and the perturbation  $\tilde{d}$  is used. Meeting these constraints on  $\tilde{d}$  can be rather difficult in applications, hence, development of any other filtering approach is of great importance.

In this work we propose a nonlinear filtering framework that is based on power operations and can be applied under smallness of the amplitude of noise in comparison with the excitation of the regressor, but it does not use any frequency or stochastic properties of  $\tilde{d}$ . The noise reduction is obtained at the price of augmentation of the size of the vector of unknown parameters and, hence, extension of the regressor. Under mild assumptions, we will show that the excitation characteristics of  $\tilde{\phi}$  in (1) can be preserved for the new regressor, while application of the dynamic regression extension and mixing (DREM) method [9] allow us to avoid the issues related with increased dimension of the vector of unknown parameters. The DREM method by itself has good robustness abilities providing the estimates in a fixed time [17], and supporting an independent estimation of any component of the vector of unknown parameters (it is the property we will use).

The outline of this paper is as follows. Technical formulation of the considered problem is given in Section II. Two approaches to the power transformation of the linear regression (which provides a nonlinear filtration) are introduced in Section III. The estimation problem in the transformed regression is solved via the DREM method applying delay based extension in Section IV. Illustrative simulation results are shown in Section V.

### Notation

- Let  $\mathbb{R}$  and  $\mathbb{N}$  denote the real and natural numbers,  $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$  and  $\mathbb{N}_* = \mathbb{N} \cup \{0\}$ .
- Let  $|x|$  be the Euclidean norm for a vector  $x \in \mathbb{R}^n$  or the absolute value of a real number  $x \in \mathbb{R}$ .
- For a Lebesgue measurable function of time  $d : \mathbb{R}_+ \rightarrow \mathbb{R}^m$  define the norm  $\|d\|_\infty = \text{ess sup}_{t \in \mathbb{R}_+} |d(t)|$ , and the space of  $d$  with  $\|d\|_\infty < +\infty$  we further denote as  $\mathcal{L}_\infty^m$ .
- For a sequence of vectors  $d : \mathbb{N}_* \rightarrow \mathbb{R}^m$  define the norm  $\|d\|_\infty = \sup_{t \in \mathbb{N}_*} |d(t)|$ , and the space of  $d$  with  $\|d\|_\infty < +\infty$  we further denote as  $L_\infty^m$ .
- The identity matrix of dimension  $n \times n$  is denoted by  $I_n$ .
- The binomial coefficients  $\rho_i$  for  $i = 1, \dots, n-1$  are the coefficients of the polynomial  $(a+b)^n = a^n + \sum_{i=1}^{n-1} \rho_i a^{n-i} b^i + b^n$ . For convenience, we also denote  $\rho_0 = \rho_n = 1$ .

<sup>1</sup>ITMO University, 49 av. Kronverkskiy, 197101, Saint Petersburg, Russia.

<sup>2</sup>Hangzhou Dianzi University, Hangzhou, China.

<sup>3</sup>CentraleSupélec – IETR, Avenue de la Boulaie, 35576 Cesson-Sévigné, France. Corresponding author, stanislav.aranovskiy@centralesupelec.fr

<sup>4</sup>Inria, Univ. Lille, CNRS, UMR 9189 - CRISTAL, F-59000 Lille, France.

This work was supported by the Ministry of Science and Higher Education of Russian Federation, goszadanie no. 2019-0898.

## II. PROBLEM STATEMENT

Consider the parameter estimation problem in a scalar noisy linear regression:

$$y(t) = \phi(t)\theta + d(t), \quad (2)$$

where  $y(t) \in \mathbb{R}$  is the measured output,  $\phi(t) \in \mathbb{R}$  is the measured regressor signal,  $\theta \in \mathbb{R}$  is unknown constant parameter that we would like to find, and  $d(t) \in \mathbb{R}$  is the measurement noise;  $t \in \mathbb{R}_+$  is continuous or  $t \in \mathbb{N}_*$  is discrete. Assume that  $\phi, d \in \mathcal{L}_\infty^1$  or  $\phi, d \in L_\infty^1$  for continuous or discrete time, respectively.

*Remark 1:* The scalar case is considered without losing generality since it can be always obtained from (1) using DREM approach [9].

As it has been shown in a recent work [17], if regressor in (2) is persistently excited (PE) [1], [7], DREM method by itself is an identification approach providing a simple solution to the problem, and the main technical difficulty next is to apply a noise filter to the obtained estimate. It was observed in the same work that nonlinear filtering algorithms demonstrate a remarkable performance comparing with conventional low-pass filtering. In the vein with this result, we would like to propose an approach in the present note for introduction of nonlinear filtration in (2), with posterior recalling the DREM procedures to estimate the obtained extended vector of parameters containing the powers of  $\theta$ .

To this end we consider a particular scenario when

- the regressor  $\phi$  is PE;
- the noise is sufficiently small, i.e.,  $\|d\|_\infty \ll 1$ .

Formally the latter condition can be always provided by scaling the available for measurements output  $y(t)$  and regressor  $\phi(t)$ .

Moreover, as it is shown in [18], if the scalar model (2) is generated from (1) using the DREM procedure with a proper choice of a dynamic extension method, then  $\tilde{\phi}$  being PE implies that  $\phi$  is strictly separated from zero after an initial information accumulation interval, i.e., we may assume  $\phi(t) \geq \mu > 0$  for all  $t \geq t_0$  for some  $t_0 \in \mathbb{R}_+$ .

## III. EXTENSION TO POWER LINEAR REGRESSION

Our main idea is based on the following simple observation. Since the noise is small,  $\|d\|_\infty \ll 1$ , if a suitable transformation of (2) yields a new linear regression with the  $d^n(t)$  term for some  $n \in \mathbb{N}$  with  $n > 1$ , then the noise influence will be attenuated. There are several approaches to obtain such a new linear regression and a price to pay for that, which we present and discuss below.

### A. Power method

One straightforward way is to take the  $n$ th power of both sides in (2):

$$(y(t) - \phi(t)\theta)^n = d^n(t),$$

which results in the new linear regression

$$\psi(t) = \varphi^\top(t)\Theta + v(t), \quad (3)$$

where

$$\begin{aligned} \psi(t) &= y^n(t), \quad v(t) = d^n(t), \\ \Theta &= [\theta^n \quad \theta^{n-1} \quad \dots \quad \theta]^\top \end{aligned} \quad (4)$$

and the coefficients of  $\varphi(t) \in \mathbb{R}^n$  are defined respectively as functions of powers of  $\phi(t)$ ,  $y(t)$  and their products:

$$\varphi(t) = \text{col} \left( (-1)^{n-i+1} \rho_i \phi^{n-i}(t) y^i(t) \right)_{i=0}^{n-1}, \quad (5)$$

where  $\rho_i$  for  $i = 1, \dots, n-1$  are the binomial coefficients, and  $\rho_0 = 1$ .

Note that all elements of  $\varphi(t)$  have a common multiplier  $\phi(t)$  by construction. Under certain conditions, this new extended regressor inherits PE of  $\phi(t)$ . Indeed, the following simple result can be derived in discrete time:

*Lemma 1:* Let  $n = 2$ ,

$$\phi(t) > \mu > 0, \quad \forall t \in \mathbb{N}_* \quad (6)$$

and there exist  $L \in \mathbb{N}$  and  $\kappa > 0$  such that

$$\sum_{k=t}^{t+L-1} \sum_{j=k+1}^{t+L} (\phi(k)d(j) - \phi(j)d(k))^2 \geq \kappa, \quad \forall t \in \mathbb{N}_*,$$

then the regressor  $\varphi$  is PE.

The requirement (6) implies PE of  $\phi$ , but the converse is obviously not true. However, it was proven in [18], [19] that if (1) has a PE vector regressor, then there is a choice of filters in DREM method providing this separation property for  $\phi$ . Note also that for any PE signal  $\phi$ ,  $\phi(t)$  cannot be identically zero for all  $t \in \mathbb{N}_*$ , therefore, by skipping the time instants  $t' \in \mathbb{N}_*$  where  $\phi(t') = 0$  (i.e.,  $y(t') = d(t')$ ), we can construct a subsequence of time instants where separation of  $\phi$  from zero is satisfied. Hence, the property (6) can be assumed for an excited (2) without much losing generality (we just exclude the case  $\mu = 0$  from consideration). Then the following simple and natural estimate can be derived in the non-transformed regression (2) [17]:

$$\hat{\theta}(t) = \frac{y(t)}{\phi(t)} \quad (7)$$

and

$$\theta = \hat{\theta}(t) - \delta(t),$$

where  $\delta(t) = \frac{d(t)}{\phi(t)}$  is the related estimation error. We will use these quantities later for comparison.

*Proof:* [Proof of Lemma 1] By definition, the signal  $\varphi$  is PE if there are  $L > 0$  and  $\varrho > 0$  such that

$$\sum_{k=t}^{t+L} \varphi(k)\varphi^\top(k) \geq \varrho I_n \quad (8)$$

for all  $t \in \mathbb{N}_*$ . The direct computations show:

$$\begin{aligned} \sum_{k=t}^{t+L} \varphi(k)\varphi^\top(k) &= \sum_{k=t}^{t+L} \phi^2(k) \begin{bmatrix} \phi^2(k) & -2\phi(k)y(k) \\ -2\phi(k)y(k) & 4y^2(k) \end{bmatrix} \\ &\geq \mu^2 R_t \end{aligned}$$

where

$$R_t = \begin{bmatrix} \sum_{k=t}^{t+L} \phi^2(k) & r_t^{12} \\ r_t^{12} & r_t^{22} \end{bmatrix},$$

$$r_t^{12} = -2 \sum_{k=t}^{t+L} (\phi^2(k)\theta + \phi(k)d(k)),$$

$$r_t^{22} = 4\theta^2 \sum_{k=t}^{t+L} \phi^2(k) + 8\theta \sum_{k=t}^{t+L} \phi(k)d(k) + 4 \sum_{k=t}^{t+L} d^2(k).$$

Since

$$\det R_t = 4 \left( \sum_{k=t}^{t+L} \phi^2(k) \sum_{k=t}^{t+L} d^2(k) - \left( \sum_{k=t}^{t+L} \phi(k)d(k) \right)^2 \right)$$

and

$$\begin{aligned} & \sum_{k=t}^{t+L} \phi^2(k) \sum_{k=t}^{t+L} d^2(k) - \left( \sum_{k=t}^{t+L} \phi(k)d(k) \right)^2 \\ &= \frac{1}{2} \sum_{k=t}^{t+L} \sum_{j=t}^{t+L} (\phi(k)d(j) - \phi(j)d(k))^2 \\ &= \sum_{k=t}^{t+L-1} \sum_{j=k+1}^{t+L} (\phi(k)d(j) - \phi(j)d(k))^2 \end{aligned}$$

the claim follows from conditions of the proposition:

$$\sum_{k=t}^{t+L} \varphi(k) \varphi^\top(k) \geq \mu \tilde{\kappa} I_n > 0, \quad \forall t \in \mathbb{N}_*$$

and some  $\tilde{\kappa} \in (0, \sqrt{\kappa})$ . ■

The restriction given on the signal  $\phi(t)d(t)$  in this lemma can be verified if the variables  $\phi$  and  $d$  are independent and  $d$  is zero-mean, then the PE property may be preserved in (3): in such a case  $\lim_{L \rightarrow +\infty} \sum_{k=t}^{t+L} \phi(k)d(k) = 0$  as the mean value of an unbiased signal and  $\sum_{k=t}^{t+L} \phi^2(k) \sum_{k=t}^{t+L} d^2(k)$  is positive ensuring existence of  $\kappa$ .

*Lemma 2:* Let  $n = 3$ , (6) be true and there exist  $L \in \mathbb{N}$  and  $\kappa > 0$  such that

$$\begin{aligned} & \sum_{k=t}^{t+L} \sum_{j=t}^{t+L} \sum_{s=t}^{t+L} (\phi(k)d(j) - \phi(j)d(k))^2 (\phi(k)d(s) - \phi(s)d(k))^2 \\ & \quad \times (\phi(s)d(j) - \phi(j)d(s))^2 \geq \kappa, \quad \forall t \in \mathbb{N}_*, \end{aligned}$$

then the regressor  $\varphi$  is PE.

*Proof:* For  $n = 3$ , repeating the same computations we

get

$$\begin{aligned} \det R_t &= 3^4 \left[ 2 \sum_{k=t}^{t+L} \phi^2(k) d^2(k) \sum_{k=t}^{t+L} \phi^3(k) d(k) \sum_{k=t}^{t+L} \phi(k) d^3(k) \right. \\ & \quad + \sum_{k=t}^{t+L} \phi^4(k) \sum_{k=t}^{t+L} d^4(k) \sum_{k=t}^{t+L} \phi^2(k) d^2(k) \\ & \quad - \sum_{k=t}^{t+L} d^4(k) \left( \sum_{k=t}^{t+L} \phi^3(k) d(k) \right)^2 - \sum_{k=t}^{t+L} \phi^4(k) \left( \sum_{k=t}^{t+L} \phi(k) d^3(k) \right)^2 \\ & \quad \left. - \left( \sum_{k=t}^{t+L} \phi^2(k) d^2(k) \right)^3 \right] \\ &= \frac{3^3}{2} \sum_{k=t}^{t+L} \sum_{j=t}^{t+L} \sum_{s=t}^{t+L} (\phi(k)d(j) - \phi(j)d(k))^2 (\phi(k)d(s) - \phi(s)d(k))^2 \\ & \quad \times (\phi(s)d(j) - \phi(j)d(s))^2, \end{aligned}$$

which implies the result. ■

For  $n = 3$  the restriction on  $\phi$  and  $d$  has a similar spirit, saying that if excitation of the signal  $\phi(t)d(t)$  is sufficiently rich, then PE is kept in (3). The results of these lemmas can be generalized to any  $n \in \mathbb{N}$ :

*Theorem 1:* Let (6) be true and there exist  $L \in \mathbb{N}$  and  $\kappa > 0$  such that there are  $i_1, \dots, i_n \in \{t, \dots, t+L\}$  providing

$$|\det R_t(\phi, d)| \geq \kappa, \quad \forall t \in \mathbb{N}_*,$$

where

$$R_t(\phi, d) = \begin{bmatrix} c(\phi(i_1), d(i_1)) \\ \vdots \\ c(\phi(i_n), d(i_n)) \end{bmatrix} \quad (9)$$

and

$$c(\phi, d) = [\phi^{n-1} \quad \phi^{n-2}d \quad \phi^{n-3}d^2 \quad \dots \quad \phi d^{n-2} \quad d^{n-1}],$$

then the regressor  $\varphi$  is PE.

The proof of the theorem is omitted due to the lack of space. Note that the same conclusions can be obtained for interval excitation.

*Remark 2:* It is straightforward to see that Lemma 1 and Lemma 2 define equivalent conditions to Theorem 1. E.g., for  $n = 2$  and  $L = 2$ , the condition of Lemma 1 reads

$$\begin{aligned} & (\phi(t)d(t+1) - \phi(t+1)d(t))^2 \\ & + (\phi(t)d(t+2) - \phi(t+2)d(t))^2 \\ & + (\phi(t+1)d(t+2) - \phi(t+2)d(t+1))^2 \geq \kappa_1 > 0, \end{aligned}$$

and the condition of Theorem 1 for  $n = 2$  and  $L = 2$  reads

$$\begin{aligned} & \min \left( \left( \det \begin{bmatrix} \phi(t) & d(t) \\ \phi(t+1) & d(t+1) \end{bmatrix} \right)^2, \right. \\ & \left( \det \begin{bmatrix} \phi(t) & d(t) \\ \phi(t+2) & d(t+2) \end{bmatrix} \right)^2, \\ & \left. \left( \det \begin{bmatrix} \phi(t+1) & d(t+1) \\ \phi(t+2) & d(t+2) \end{bmatrix} \right)^2 \right) \geq \kappa_2 > 0 \end{aligned}$$

for some  $\kappa_1, \kappa_2$ . Obviously, if one of these conditions is satisfied, then there exist such a pair  $\kappa_1, \kappa_2$  that the second condition is satisfied as well.

## B. Approximation of exponent

Another approach deals with exponentiation of  $y(t)$  with posterior approximation of the exponent by power series.

1) *Continuous time:* Denote  $\xi(t) = e^{y(t)} = e^{\phi(t)\theta} e^{d(t)}$ , then using representation of the exponent function by an infinite power series of the argument, the expression for this new variable can be rewritten as follows:

$$\xi(t) = e^{\phi(t)\theta} \sum_{i=0}^{+\infty} \frac{d^i(t)}{i!} = e^{\phi(t)\theta} \left( \sum_{i=0}^N \frac{d^i(t)}{i!} + \epsilon(t) \right),$$

where  $N \in \mathbb{N}$  defines the order of truncation of the series and  $\epsilon(t)$  is an approximation error, which should be of order  $d^{N+1}(t)$  and sufficiently small for  $\|d\|_\infty \ll 1$ . Taking derivative of  $\xi(t)$  or applying any linear filter we get:

$$\begin{aligned} \dot{\xi}(t) &= e^{\phi(t)\theta} \left[ \dot{\phi}(t)\theta \left( \sum_{i=0}^N \frac{d^i(t)}{i!} + \epsilon(t) \right) \right. \\ &\quad \left. + \dot{d}(t) \sum_{i=0}^{N-1} \frac{d^i(t)}{i!} + \dot{\epsilon}(t) \right] \\ &= \xi(t) \left[ \dot{\phi}(t)\theta + \frac{\dot{d}(t) \sum_{i=0}^{N-1} \frac{d^i(t)}{i!} + \dot{\epsilon}(t)}{\sum_{i=0}^N \frac{d^i(t)}{i!} + \epsilon(t)} \right] \end{aligned}$$

or equivalently

$$\left( \dot{\xi}(t) - \xi(t)\dot{\phi}(t)\theta \right) \sum_{i=0}^N \frac{d^i(t)}{i!} = v(t), \quad (10)$$

where

$$\begin{aligned} v(t) &= \xi(t) \left( \dot{d}(t) \sum_{i=0}^{N-1} \frac{d^i(t)}{i!} + \dot{\epsilon}(t) \right) \\ &\quad - \epsilon(t) \left( \dot{\xi}(t) - \xi(t)\dot{\phi}(t)\theta \right) \end{aligned}$$

contains all terms proportional to uncertain variables  $\epsilon(t)$ ,  $\dot{\epsilon}(t)$  and  $\dot{d}(t)$  that we assume to be negligible. Finally, using  $\dot{\xi}(t) = \dot{y}(t)\xi(t)$ , substituting  $d(t) = y(t) - \phi(t)\theta$  from (2) in the left-hand side of (10) and opening the brackets, we obtain (3) for  $n = N + 1$ , the respective  $\Theta$ , and for  $\psi(t)$  and  $\varphi(t)$  dependent on  $y(t)$ ,  $\phi(t)$ ,  $\dot{y}(t)$ , and  $\dot{\phi}(t)$ .

*Remark 3:* To implement the proposed continuous-time approximation, the derivatives of  $y$  and  $\phi$  should be available. This shortcoming can be alleviated, e.g., applying an LTI filter to the signals  $y$  and  $\phi$  of the original scalar equation (2) and using the filtered signals and their derivatives (available as internal filters' variables).

2) *Discrete time:* In such a case differentiation is not possible, then we can use the series directly:

$$\begin{aligned} e^{y(t)} &= e^{\phi(t)\theta} e^{d(t)} = \sum_{i=0}^{+\infty} \frac{(\phi(t)\theta)^i}{i!} \sum_{i=0}^{+\infty} \frac{d^i(t)}{i!} \\ &= \left( \sum_{i=0}^{N_1} \frac{(\phi(t)\theta)^i}{i!} + \epsilon_1(t) \right) \left( \sum_{i=0}^{N_2} \frac{d^i(t)}{i!} + \epsilon_2(t) \right), \end{aligned}$$

where  $N_1, N_2 \in \mathbb{N}$  are the orders of approximation of related exponents by finite-length power series and  $\epsilon_1(t), \epsilon_2(t) \in \mathbb{R}$

are corresponding approximation errors. A straightforward manipulation gives:

$$\begin{aligned} e^{y(t)} - 1 &= \sum_{i=1}^{N_1} \frac{(\phi(t)\theta)^i}{i!} + \sum_{i=1}^{N_2} \frac{d^i(t)}{i!} \\ &\quad + \sum_{i=1}^{N_1} \frac{(\phi(t)\theta)^i}{i!} \sum_{i=1}^{N_2} \frac{d^i(t)}{i!} + v(t), \end{aligned}$$

where

$$v(t) = \epsilon_1(t) \sum_{i=0}^{N_2} \frac{d^i(t)}{i!} + \epsilon_2(t) \sum_{i=0}^{N_1} \frac{(\phi(t)\theta)^i}{i!} + \epsilon_1(t)\epsilon_2(t)$$

is the new noise, which we assume being small since  $\epsilon_1(t)$  and  $\epsilon_2(t)$  are so. Again substituting  $d(t) = y(t) - \phi(t)\theta$  from (2) the regression (3) is recovered for  $\psi(t) = e^{y(t)} - 1 - \sum_{i=0}^{N_2} \frac{y^i(t)}{i!}$  and  $n = N_1 + N_2$ . Direct computations provide the following examples: if  $N_1 = N_2 = 1$  then

$$\psi(t) = e^{y(t)} - 1 - y(t), \quad \varphi(t) = [-\phi^2(t) \quad \phi(t)y(t)]^\top,$$

if  $N_1 = N_2 = 2$  then

$$\begin{aligned} \psi(t) &= e^{y(t)} - 1 - \frac{y^2(t)}{2} - y(t), \\ \varphi(t) &= \frac{1}{2} \begin{bmatrix} \frac{1}{2}\phi^4(t) \\ -\phi^3(t)y(t) \\ \frac{1}{2}\phi^2(t)y^2(t) - \phi^2(t)y(t) \\ \phi(t)y^2(t) \end{bmatrix}. \end{aligned}$$

Let us now consider how to estimate the new vector of unknown parameters in (3), and to compare the obtained results with the conventional solution (7).

## IV. ESTIMATION IN (3)

As we can see, the linear regression (3) can be derived using different approaches: exact with  $v(t) = d^n(t)$  or with some additional approximation items in  $v(t)$  (in the latter case additional hypotheses are needed that the neglected terms are sufficiently small), where the disturbance  $v(t)$  is smaller than in (2) but the vector of unknown parameters  $\Theta$  has  $n$  elements containing different powers  $\theta^i$  for  $i = 1, \dots, n$  of the unknown scalar parameter of (2). Despite, as it was demonstrated above, the excitation of  $\varphi(t)$  can be preserved in this extended vector regression under reasonable assumptions, it can be weak, and since it depends on the properties of  $d(t)$ , it is difficult to predict it (inversely, a proper noise  $d(t)$  may also enforce the excitation of the regressor). However, developing estimation algorithms for (3) we can utilize the fact that all elements of  $\Theta$  are interrelated.

For example, DREM approach can be applied to (3) allowing each component of  $\Theta$  to be estimated separately, then the noise  $d(t)$  will be filtered several times: first, by taking the power  $d^n(t)$  of it, second, in the filter of the extension part in DREM method.

Moreover, such a nonlinear filtering appears even if delays are used instead of conventional convolution filters (which seem to be not yielding a filtering ability). And this property

we will demonstrate next by considering (3) obtained by the power method.

Applying the delays in the DREM extension step we get an extended version of (3):

$$Y(t) = \Phi(t)\Theta + V(t), \quad (11)$$

where

$$Y(t) = \begin{bmatrix} \psi(t) \\ \psi(t-\tau) \\ \vdots \\ \psi(t-(n-1)\tau) \end{bmatrix}, \quad V(t) = \begin{bmatrix} v(t) \\ v(t-\tau) \\ \vdots \\ v(t-(n-1)\tau) \end{bmatrix},$$

$$\Phi(t) = \begin{bmatrix} \varphi^\top(t) \\ \varphi^\top(t-\tau) \\ \vdots \\ \varphi^\top(t-(n-1)\tau) \end{bmatrix},$$

where  $\psi(t) \in \mathbb{R}$  and  $v(t) \in \mathbb{R}$  are defined in (4),  $\varphi(t) \in \mathbb{R}^n$  is defined in (5), and  $\tau \in \mathbb{R}_+$  (in continuous time or  $\tau \in \mathbb{N}$  in the discrete-time case) is the elementary delay. If the new extended regressor  $\Phi$  is nonsingular, then the direct estimates of  $\Theta$  from (11) can be obtained:

$$\Theta = \hat{\Theta}(t) - \mathcal{V}(t),$$

where

$$\hat{\Theta}(t) = \Phi^{-1}(t)Y(t) \quad (12)$$

is the implementable estimate that can be derived from (11) and  $\mathcal{V}(t) = \Phi^{-1}(t)V(t)$  is the new noise proportional to  $d^n$ .

The following proposition shows that the first element of  $\Theta$  corresponding to  $\theta^n$  will be estimated as the geometric mean of the scalar evaluations  $\hat{\theta}(t)$  from (7) defined for non-zero regressor  $\phi$ .

*Proposition 1:* Consider the linear regression (2) where  $\phi$  satisfies (6) and the related extended linear regression (11) for some  $n \geq 2$  with a non-singular matrix  $\Phi(t)$ . Then for the first component of  $\hat{\Theta}$  in (12) it holds

$$\hat{\Theta}_1(t) = \prod_{i=0}^{n-1} \hat{\theta}(t - i\tau),$$

where  $\hat{\theta}$  are defined by (7).

The proof of Proposition is omitted due to the lack of space.

Note that the last element of  $\Theta$  (that corresponds to  $\theta$ ) receives an estimate as the arithmetic mean of the same quantities:

$$\hat{\Theta}_n(t) = \frac{1}{n} \sum_{i=0}^{n-1} \hat{\theta}(t - i\tau).$$

The intermediate elements of  $\Theta$  will be derived as respective products of  $\hat{\theta}(t - i\tau)$  of given orders, i.e., for  $\theta^2$  it will be an averaged sum of  $\hat{\theta}(t - i\tau)\hat{\theta}(t - k\tau)$  with  $i \neq k = 0, \dots, n-1$ . The elements of the noise vector  $\mathcal{V}(t)$  will have corresponding dependence on  $\delta(t) = \frac{d(t)}{\phi(t)}$ , i.e., for the last element  $\mathcal{V}_n(t) = \frac{1}{n} \sum_{i=0}^{n-1} \delta(t - i\tau)$ .

Clearly, the geometric or arithmetic mean estimates, as well as others, can be derived directly, but it is interesting that they also can be obtained as particular solutions (for a special choice of delays as extension filters) of DREM method applied to (3). Roughly speaking, it demonstrates what kind of nonlinear filtering can be derived by the proposed approach. Other filtering schemes can also be obtained, whose investigation is a direction of future research.

## V. EXAMPLE

Consider the linear regression (2), where

$$\phi(t) = \mu + a(\sin(\omega t))^2$$

for some known positive constants  $\mu$ ,  $a$ , and  $\omega$ . This scalar regression is PE and satisfies (6). The noise  $d$  is uniformly distributed in the interval from  $-\bar{d}$  to  $\bar{d}$  for some positive constant  $\bar{d}$ .

For comparison, we consider three estimates of  $\theta$ :

- the direct estimate (7), which is further denoted as  $\hat{\theta}^{\text{dir}}(t)$ ;
- the delay-based extension (11) and the estimate (12), where we use the result of Proposition 1 and compute

$$\hat{\theta}^{\text{del}}(t) = \left( \hat{\Theta}_1(t) \right)^{\frac{1}{n}}; \quad (13)$$

- the DREM procedure applied to (3) with Kreisselmeier's dynamics extension, as described in [17], [18]:

$$\begin{aligned} \frac{d}{dt} \mathcal{F}(t) &= -\alpha \mathcal{F}(t) + \beta \varphi(t) \varphi^\top(t) \\ \frac{d}{dt} \mathcal{Y}(t) &= -\alpha \mathcal{Y}(t) + \beta \varphi(t) \psi(t), \\ \Delta(t) &= \det \mathcal{F}(t), \\ \bar{\mathcal{Y}}(t) &= \text{adj}(\mathcal{F}(t)) \mathcal{Y}(t), \\ \hat{\theta}^{\text{drem}} &= \frac{1}{\max(\Delta(t), \epsilon_\Delta)} \bar{\mathcal{Y}}_n(t), \end{aligned} \quad (14)$$

where  $\alpha > 0$  and  $\beta > 0$  are scalar tuning parameters,  $\epsilon_\Delta > 0$  is the threshold applied during the initial information accumulation,  $\text{adj}$  denotes the adjugate matrix, and  $\bar{\mathcal{Y}}_n(t)$  denotes the  $n$ th element of the vector  $\bar{\mathcal{Y}}(t)$ .

For numerical simulations, we set  $\mu = 0.3$ ,  $a = 0.15$ ,  $\omega = \pi$ ,  $\bar{d} = 0.1$ , and  $\theta = 2$ . For the power extension, we set  $n = 3$ . For the delay-based estimation, the delay is  $\tau = 0.1$ , and for the DREM procedure we set  $\alpha = 0.5$ ,  $\beta = 100$ , and  $\epsilon_\Delta = 10^{-5}$ .

The transients are shown in Figure 1. The direct estimate  $\hat{\theta}^{\text{dir}}$  provides an estimate after the first step; however, it has the worst noise sensitivity. The delay-based estimate converges after  $n\tau = 0.3$  seconds and show good noise attenuation; as shown in Proposition 1, such an estimate corresponds to the (nonlinear) geometric mean filtration. Finally, the DREM-based estimate has the slowest convergence and the best noise attenuation; note that the dynamics extension behaves as a low-pass filter.

For numerical comparison of the estimates, the mean absolute error (MAE) and the mean square error (MSE) are

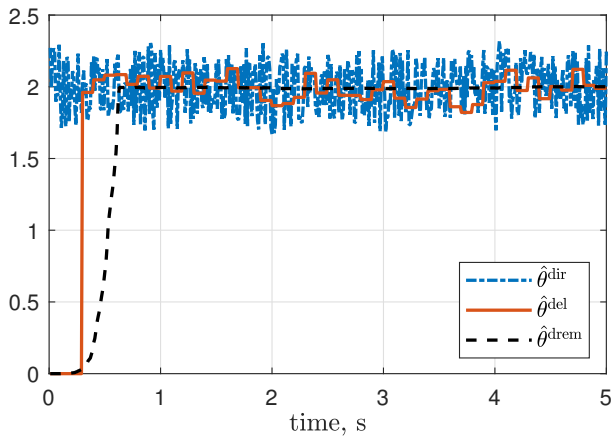


Fig. 1. Transients of the direct estimate  $\hat{\theta}^{\text{dir}}$  (7), the delay-based estimate  $\hat{\theta}^{\text{del}}$  (13) and the DREM-based estimate  $\hat{\theta}^{\text{drem}}$  (14). The true value is  $\theta = 2$ .

TABLE I  
MSE AND MAE OF THE ESTIMATES

	$\hat{\theta}^{\text{dir}}$	$\hat{\theta}^{\text{del}}$	$\hat{\theta}^{\text{drem}}$
MAE · 10 <sup>3</sup>	136.1	74.0	3.0
MSE · 10 <sup>3</sup>	25.2	8.5	0.2

given in Table I; these values are computed after the transient period over  $10^6$  samples.

As we increase the order  $n$ , the DREM procedure requires for extra tuning of  $\alpha$  and  $\beta$ ; otherwise, the value of  $\Delta$  remains small. In contrast, the delay-based estimation does not require any extra tuning as the order  $n$  increases.

## VI. CONCLUSION

The problem of noise attenuation in the linear regression estimation model was studied. Avoiding the restrictions on stochastic of frequency properties of the measurement perturbations, it is only assumed that the noise is sufficiently small comparing to the regressor (the regressor is sufficiently excited in relation with the noise amplitude). Two methods of nonlinear transformation of the linear regression were proposed. The obtained transformed linear regression has an augmented regressor and the vector of unknown parameters containing the powers of  $\theta$  and disturbance, where the influence of the latter is attenuated. It was proven that the presented power transformation of a regressor keeps the excitation under mild restrictions. The DREM method was applied to estimate the unknown parameters. The performance of the estimators was demonstrated and compared in numerical experiments.

## REFERENCES

- [1] P. Ioannou and P. Kokotovic, *Adaptive Systems with Reduced Models*. Secaucus, NJ: Springer Verlag, 1983.
- [2] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1987.
- [3] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness*. London: Prentice-Hall, 1989.
- [4] X. Yan, *Linear Regression Analysis: Theory and Computing*. Singapore: World Scientific, 2009.

- [5] A. Rencher and W. Christensen, *Methods of Multivariate Analysis*, 3rd ed., ser. Wiley Series in Probability and Statistics. New Jersey: John Wiley & Sons, 2012, vol. 709.
- [6] A. Morgan and K. Narendra, "On the uniform asymptotic stability of certain linear nonautonomous differential equations," *SIAM J. Control and Optimization*, vol. 15, no. 1, pp. 5–24, 1977.
- [7] K. Narendra and A. Annaswamy, "Persistent excitation in adaptive systems," *International Journal of Control*, vol. 45, no. 1, pp. 127–160, 1987.
- [8] G. Chowdhary, T. Yucelen, M. Mühlegg, and E. N. Johnson, "Concurrent learning adaptive control of linear systems with exponentially convergent bounds," *International Journal of Adaptive Control and Signal Processing*, vol. 27, no. 4, pp. 280–301, 2012.
- [9] S. Aranovskiy, A. Bobtsov, R. Ortega, and A. Pyrkin, "Performance enhancement of parameter estimators via dynamic regressor extension and mixing," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3546–3550, jul 2017.
- [10] N. Barabanov and R. Ortega, "On global asymptotic stability of  $\dot{x} = -\phi(t)\phi^\top(t)x$  with  $\phi(t)$  bounded and not persistently exciting," *Systems & Control Letters*, vol. 109, pp. 24–27, 2017.
- [11] D. Efimov, N. Barabanov, and R. Ortega, "Robust stability under relaxed persistent excitation conditions," *International Journal of Adaptive Control and Signal Processing*, 2019.
- [12] J. Wang, D. Efimov, S. Aranovskiy, and A. Bobtsov, "Fixed-time estimation of parameters for non-persistent excitation," *European Journal of Control*, 2019.
- [13] J. Wang, D. Efimov, and A. Bobtsov, "On robust parameter estimation in finite-time without persistence of excitation," *IEEE Transactions on Automatic Control*, vol. 65, no. 4, 2020.
- [14] A. Goldenshluger and B. Polyak, "Estimation of regression parameters with arbitrary noise," *Mathematical Methods of Statistics*, vol. 2, 01 1993.
- [15] O. Granichin, "Linear regression and filtering under nonstandard assumptions (arbitrary noise)," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1830–1837, 2004.
- [16] O. Stepanov and A. Toropov, "Nonlinear filtering for map-aided navigation. part 1. an overview of algorithms," *Gyroscopy Navig*, p. 324–337, 6 2015.
- [17] M. Korotina, S. Aranovskiy, R. Ushirobira, D. Efimov, and J. Wang, "Fixed-time parameter estimation via the discrete-time drem method," in *Proc. 22st IFAC World Congress*, Yokohama, Japan, 2023.
- [18] S. Aranovskiy, R. Ushirobira, M. Korotina, and A. Vedyakov, "On preserving-excitation properties of kreisselmeier's regressor extension scheme," *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 1296–1302, feb 2023.
- [19] B. Yi and R. Ortega, "Conditions for convergence of dynamic regressor extension and mixing parameter estimators using lti filters," *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 1253–1258, 2022.