



HAL
open science

Power Series for Noise Attenuation in Linear Regression Parameter Estimation *

Jian Wang, Alexey Bobtsov, Stanislav Aranovskiy, Denis Efimov, Anton Pyrkin

► **To cite this version:**

Jian Wang, Alexey Bobtsov, Stanislav Aranovskiy, Denis Efimov, Anton Pyrkin. Power Series for Noise Attenuation in Linear Regression Parameter Estimation *. 2025. <hal-04894795>

HAL Id: hal-04894795

<https://hal.science/hal-04894795v1>

Preprint submitted on 17 Jan 2025

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.



HAL Authorization

Power Series for Noise Attenuation in Linear Regression Parameter Estimation*

Jian Wang^{a,*}, Alexey Bobtsov^{b,a}, Stanislav Aranovskiy^c, Denis Efimov^d,
Anton Pyrkin^b

^aHangzhou Dianzi University, Hangzhou, China

^bITMO University, 49 av. Kronverkskiy, Saint Petersburg, 197101, Russia

^cCentraleSupélec – IETR, Avenue de la Boulaie, Cesson-Sévigné, 35576, France

^dInria, Univ. Lille, CNRS, UMR 9189 - CRISTAL, Lille, F-59000, France

Abstract

The constant parameter identification problem is considered for a linear regression model assuming that the noise is sufficiently small comparing to the regressor. Introducing a nonlinear transformation, the estimation is performed for an extended regression dependent on the powers of the unknown parameters and the disturbance, where the influence of the latter is attenuated. It is shown that such a transformation preserves the excitation of regressor under reasonable assumptions. The quality improvement is demonstrated in numerical experiments.

Keywords: parameter estimation, excitation, noise attenuation, linear regression

1. Introduction

2 The linear regression model represents the basic setup for estimation of
3 parameter in a noisy setting:

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

4 where $\tilde{y}(t) \in \mathbb{R}$ is the measured output, $\tilde{\phi}(t) \in \mathbb{R}^p$ is the measured vector
5 regressor, $\tilde{\theta} \in \mathbb{R}^p$ is unknown constant vector of parameters to be estimated,

*This work was supported by by the China 111 Project D17019.

*The corresponding author.

Email address: wangjian119@hdu.edu.cn (Jian Wang)

6 and $\tilde{d}(t) \in \mathbb{R}$ is a bounded measurement noise. There are many approaches
 7 providing a solution of this estimation problem (Ioannou and Kokotovic,
 8 1983; Ljung, 1987; Sastry and Bodson, 1989; Yan, 2009; Rencher and Chris-
 9 tensen, 2012) (to mention a few handbooks), ensuring different convergence
 10 and robustness qualities for evaluation of $\tilde{\theta}$, which are heavily based on the
 11 excitation characteristics of the regressor $\tilde{\phi}(t)$.

12 Interval excitation is necessary and sufficient for identification of $\tilde{\theta}$ in the
 13 noise-free case, while for $\tilde{d} \neq 0$, the persistence of excitation is commonly
 14 required (Morgan and Narendra, 1977; Narendra and Annaswamy, 1987).
 15 The latter property implies that collecting the measurements and continuing
 16 the observation it is possible to improve the estimation accuracy. Such a
 17 dependence on excitation of the regressor is limiting, and there are many
 18 works focusing on relaxation of this requirement (Chowdhary et al., 2012;
 19 Aranovskiy et al., 2017; Barabanov and Ortega, 2017; Efimov et al., 2019;
 20 Wang et al., 2019, 2020).

21 Among the performance characteristics of the identification algorithms,
 22 the time of convergence of the estimate to its ideal value and the estimation
 23 error robustness with respect to the perturbation \tilde{d} play important roles.
 24 Different filtering algorithms are often used to overcome the latter issue: for
 25 example, linear filters are applied to both sides of (1) improving the estima-
 26 tion accuracy in (Goldenshluger and Polyak, 1993; Granichin, 2004; Stepanov
 27 and Toropov, 2015). Tuning these filters requires the information about the
 28 frequency spectrum of the \tilde{d} or its stochastic characteristics. Moreover, the
 29 spectrum separation (independence) between the regressor $\tilde{\phi}$ and the pertur-
 30 bation \tilde{d} is often required. Unfortunately, these restrictions on \tilde{d} are rather
 31 difficult to guarantee in applications (since both, the output \tilde{y} and the re-
 32 gressor $\tilde{\phi}$, are measured with noises). Therefore, development of any other
 33 filtering approach independent on frequency properties of the measurement
 34 noise is of great importance.

35 In this work, following our preliminary investigations (Bobtsov et al.,
 36 2024), we propose a nonlinear filtering framework that is based on power
 37 operations and can be applied under smallness of the amplitude of noise in
 38 comparison with the excitation of the regressor. The proposed approach does
 39 not use any frequency or stochastic properties of \tilde{d} . The noise reduction is
 40 obtained at the price of augmentation of the size of the vector of unknown
 41 parameters and, hence, extension of the regressor. It will be shown that un-
 42 der reasonable hypotheses, the excitation characteristics of $\tilde{\phi}$ in (1) can be
 43 preserved in the new regressor, while application of the dynamic regression

44 extension and mixing (DREM) method (Aranovskiy et al., 2017) allows us to
 45 avoid the issues related with increased dimension of the vector of unknown
 46 parameters. The DREM method by itself has good robustness abilities pro-
 47 viding the estimates in a fixed time (Korotina et al., 2024), and supporting
 48 an independent estimation of any component of the vector of unknown pa-
 49 rameters (it is the property we will use).

50 The outline of this paper is as follows. Technical formulation of the
 51 considered problem is given in Section 2. Two approaches to the power
 52 transformation of the linear regression (which provides a nonlinear filtration)
 53 are introduced in Section 3. The estimation problem in the transformed
 54 regression is solved via the DREM method applying delay based extension
 55 in Section 4. Illustrative simulation results are shown in Section 5.

56 *Notation*

- 57 • Let \mathbb{R} and \mathbb{N} denote the real and natural numbers, $\mathbb{N}_* = \mathbb{N} \cup \{0\}$.
- 58 • Let $|x|$ be the Euclidean norm for a vector $x \in \mathbb{R}^n$ or the absolute value
 59 of a real number $x \in \mathbb{R}$.
- 60 • For a sequence of vectors $d : \mathbb{N}_* \rightarrow \mathbb{R}^m$ define the norm $\|d\|_\infty =$
 61 $\sup_{t \in \mathbb{N}_*} |d(t)|$, and the space of d with $\|d\|_\infty < +\infty$ we further denote
 62 as L_∞^m .
- 63 • The identity matrix of dimension $n \times n$ is denoted by I_n .
- 64 • The binomial coefficients ρ_i for $i = 1, \dots, n-1$ are the coefficients of
 65 the polynomial $(a+b)^n = a^n + \sum_{i=1}^{n-1} \rho_i a^{n-i} b^i + b^n$. For convenience,
 66 we also denote $\rho_0 = \rho_n = 1$.

67 **2. Problem statement**

68 The vector linear regression (1) can be transformed into a set of scalar
 69 linear regression models for each component of the vector $\tilde{\theta}$ separately us-
 70 ing the DREM approach (Aranovskiy et al., 2017). Thus, without loss of
 71 generality, for the problem statement, we consider the parameter estimation
 72 problem in a scalar noisy linear regression:

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

73 where $y(t) \in \mathbb{R}$ is the measured output, $\phi(t) \in \mathbb{R}$ is the measured regressor
 74 signal, $\theta \in \mathbb{R}$ is unknown constant parameter that we would like to find,
 75 and $d(t) \in \mathbb{R}$ is the measurement noise. For consistency of presentation, we
 76 limit the problem to the discrete-time case $t \in \mathbb{N}_*$; however, similar results
 77 can be obtained for continuous-time as well. We assume that $\phi, d \in L_\infty^1$
 78 and introduce the following definition of the widely known persistence of
 79 excitation property.

80 **Definition 1.** *The bounded signal φ is persistently exciting (PE) if there are*
 81 *$L > 0$ and $\varrho > 0$ such that*

$$\sum_{k=t}^{t+L} \varphi(k) \varphi^\top(k) \geq \varrho I_n \quad (3)$$

82 for all $t \in \mathbb{N}_*$.

83 As it has been shown in a recent work by Korotina et al. (2024), if re-
 84 gressor in (2) is persistently exciting (PE) (Narendra and Annaswamy, 1987;
 85 Ioannou and Kokotovic, 1983), the DREM method by itself is an identifi-
 86 cation approach providing a simple solution to the problem, and the main
 87 design choice next is to select and apply a noise filter to the obtained esti-
 88 mate. It was observed in the same work that nonlinear filtering algorithms
 89 demonstrate a remarkable performance comparing with conventional low-
 90 pass filtering. In the vein with this result, we would like to propose an ap-
 91 proach in the present note for introduction of nonlinear filtration in (2), with
 92 posterior recalling the DREM procedures to estimate the obtained extended
 93 vector of parameters containing the powers of θ .

94 To this end we consider a particular scenario when

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

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

99 Moreover, as it is shown by Aranovskiy et al. (2023), if the scalar model
 100 (2) is generated from (1) using the DREM procedure with a proper choice
 101 of a dynamic extension method, then $\tilde{\phi}$ being PE implies that ϕ is strictly
 102 separated from zero after an initial information accumulation interval, i.e.,
 103 we may assume $\phi(t) \geq \mu > 0$ for all $t \geq t_0$ for some $t_0 \in \mathbb{N}_*$.

104 **3. Extension to power linear regression**

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

111 *3.1. Power method*

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

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

113 which results in the new linear regression

$$\psi(t) = \varphi^\top(t)\Theta + v(t), \tag{4}$$

114 where

$$\begin{aligned} \psi(t) &= y^n(t), \quad v(t) = d^n(t), \\ \Theta &= [\theta^n \quad \theta^{n-1} \quad \dots \quad \theta]^\top \end{aligned} \tag{5}$$

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

$$\varphi(t) = \text{col}\left(\left(-1\right)^{n-i+1} \rho_i \phi^{n-i}(t) y^i(t)\right)_{i=0}^{n-1}, \tag{6}$$

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

118 Note that all elements of $\varphi(t)$ have a common multiplier $\phi(t)$ by construc-
 119 tion. Under certain conditions, this new extended regressor inherits PE of
 120 $\phi(t)$. This fact was proven by Bobtsov et al. (2024) for $n = 2$ and $n = 3$, and
 121 below we generalize these results to any $n \in \mathbb{N}$.

122 **Theorem 1.** *Let*

$$\phi(t) \geq \mu > 0, \quad \forall t \in \mathbb{N}_*, \tag{7}$$

123 *and for a given $n \in \mathbb{N}$, let there exist $L \in \mathbb{N}$ and $\kappa > 0$ such that there are*
 124 *$i_1, \dots, i_n \in \{t, \dots, t+L\}$ providing*

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

125 where

$$R_t(\phi, y) = \begin{bmatrix} c(\phi(i_1), y(i_1)) \\ \vdots \\ c(\phi(i_n), y(i_n)) \end{bmatrix} \quad (8)$$

126 and

$$c(\phi, y) = [\phi^{n-1} \quad \phi^{n-2}y \quad \phi^{n-3}y^2 \quad \dots \quad \phi y^{n-2} \quad y^{n-1}]. \quad (9)$$

127 Then the regressor φ is PE.

128 The requirement (7) implies PE of ϕ , but the converse is obviously not
 129 true. However, it was proven (Aranovskiy et al., 2023; Yi and Ortega, 2022)
 130 that if (1) has a PE vector regressor, then the DREM method with a properly
 131 chosen excitation scheme provides this separation property for ϕ . Hence,
 132 the property (7) can be assumed for (2) without much losing generality.
 133 Then the following simple and natural estimate can be derived in the non-
 134 transformed regression (2) (Korotina et al., 2024):

$$\hat{\theta}(t) = \frac{y(t)}{\phi(t)} \quad (10)$$

135 and

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

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

138 *Proof of Theorem 1.* The regressor φ is PE if the condition (3) is satisfied
 139 for all t . We will show that this holds under the conditions of the theorem.

140 Since $|\det R_t(\phi, y)| \geq \kappa$ and multiplication of a matrix's column by a
 141 constant implies the proportional change of its determinant, it follows

$$\left(\det \begin{bmatrix} \tilde{c}(\phi(i_1), y(i_1)) \\ \vdots \\ \tilde{c}(\phi(i_n), y(i_n)) \end{bmatrix} \right)^2 \geq \left(\kappa \prod_{i=0}^{n-1} \rho_i \right)^2,$$

142 where

$$\tilde{c}^\top(\phi, y) = \text{col} \left((-1)^{n-i+1} \rho_i \phi^{n-1-i} y^i \right)_{i=0}^{n-1}.$$

143 Define

$$\Psi_t = [\varphi(i_1) \ \cdots \ \varphi(i_n)].$$

144 From (6), it follows $\varphi = \phi \tilde{c}^\top(\phi, y)$ and thus, recalling (7),

$$\det(\Psi_t \Psi_t^\top) \geq \left(\mu^n \kappa \prod_{i=0}^{n-1} \rho_i \right)^2.$$

145 Since ϕ and d , and thus y , are bounded for all $t \in \mathbb{N}_*$, the maximum
146 eigenvalue of the $\Psi_t \Psi_t^\top$ matrix can be bounded as

$$\begin{aligned} \lambda_{\max}(\Psi_t \Psi_t^\top) &= \|\Psi_t^\top\|_2^2 \leq n \|\Psi_t^\top\|_\infty^2 \\ &\leq n \left(\sum_{i=0}^{n-1} \rho_i \varpi^{n-1} \right)^2 = n \varpi^{2(n-1)} \left(\sum_{i=0}^{n-1} \rho_i \right)^2, \end{aligned}$$

147 where $\varpi := \max(\|\phi\|_\infty, \|y\|_\infty)$.

148 Recalling the fact that the determinant of a matrix is the product of all
149 its eigenvalues, these inequalities yield

$$\Psi_t \Psi_t^\top \geq \varrho I$$

150 with

$$\varrho = \kappa^2 \frac{(\mu^n \prod_{i=0}^{n-1} \rho_i)^2}{(n \varpi^{2(n-1)} (\sum_{i=0}^{n-1} \rho_i)^2)^{n-1}}.$$

151 The proof is complete noting that

$$\sum_{k=t}^{t+L} \varphi(k) \varphi^\top(k) \geq \sum_{k=1}^n \varphi(i_k) \varphi^\top(i_k) = \Psi_t \Psi_t^\top.$$

152 □

153 **Remark 1.** *Theorem 1 is formulated using the available signals ϕ and y*
154 *allowing to verify the condition of the theorem at least for some t . On the*
155 *other hand, recalling $y = \phi\theta + d$ and that adding one matrix column to another*
156 *does not change the matrix's determinant, it is straightforward to show that*
157 *for the matrix R_t defined in (8), it holds for all $t \in \mathbb{N}_*$*

$$\det R_t(\phi, d) = \det R_t(\phi, y).$$

158 *This observation allows rewriting conditions of Theorem 1 in the terms of ϕ*
159 *and d , where d is not measurable but does not depend on ϕ .*

160 **Remark 2.** Let $n = 2$ and $L = 2$. Taking into account Remark 1, the
 161 condition of Theorem 1 reads

$$\min \left(\left(\det \begin{bmatrix} \phi(t) & d(t) \\ \phi(t+1) & d(t+1) \end{bmatrix} \right)^2, \right. \\
\left. \left(\det \begin{bmatrix} \phi(t) & d(t) \\ \phi(t+2) & d(t+2) \end{bmatrix} \right)^2, \right. \\
\left. \left(\det \begin{bmatrix} \phi(t+1) & d(t+1) \\ \phi(t+2) & d(t+2) \end{bmatrix} \right)^2 \right) \geq \kappa_1 > 0$$

162 for some $\kappa_1 > 0$. On the other hand, the condition for $n = 2$ developed by
 163 Bobtsov et al. (2024, Lemma 1) yields for $L = 2$

$$\begin{aligned} & (\phi(t)d(t+1) - \phi(t+1)d(t))^2 \\ & + (\phi(t)d(t+2) - \phi(t+2)d(t))^2 \\ & + (\phi(t+1)d(t+2) - \phi(t+2)d(t+1))^2 \geq \kappa_2 > 0 \end{aligned}$$

164 for some $\kappa_2 > 0$. Obviously, if one of these conditions is satisfied, then
 165 there exist such a pair κ_1, κ_2 that the second condition is satisfied as well. A
 166 similar observation can also be done for the case $n = 3$ (Bobtsov et al., 2024,
 167 Lemma 2).

168 3.2. Approximation via exponentiation

169 Another approach deals with exponentiation of $y(t)$ with posterior ap-
 170 proximation of the exponent by power series:

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

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

173 tion errors. A straightforward manipulation gives:

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

174 where

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

175 is the new noise, which we assume being small since $\epsilon_1(t)$ and $\epsilon_2(t)$ are so.
176 Again substituting $d(t) = y(t) - \phi(t)\theta$ from (2) the regression (4) is recovered
177 for

$$\psi(t) = e^{y(t)} - 1 - \sum_{i=1}^{N_2} \frac{y^i(t)}{i!}$$

178 and $n = N_1 + N_2$. Direct computations provide the following examples: if
179 $N_1 = N_2 = 1$ then

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

180 if $N_1 = N_2 = 2$ then

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

181 Let us now consider how to estimate the new vector of unknown pa-
182 rameters in (4), and to compare the obtained results with the conventional
183 solution (10).

184 4. Estimation in (4)

185 As we can see, the linear regression (4) can be derived using different
186 approaches: exactly for $v(t) = d^n(t)$ or with some additional approximation

187 items in $v(t)$ (in the latter case additional hypotheses are needed that the
188 neglected terms are sufficiently small), where the disturbance $v(t)$ is smaller
189 than in (2) but the vector of unknown parameters Θ has n elements con-
190 taining different powers θ^i for $i = 1, \dots, n$ of the unknown scalar parameter
191 of (2). Despite, as it was demonstrated above, the excitation of $\varphi(t)$ can be
192 preserved in this extended vector regression under reasonable assumptions,
193 it can be weak, and since it depends on the properties of $d(t)$, it is difficult
194 to predict it (inversely, a proper noise $d(t)$ may also enforce the excitation
195 of the regressor). However, developing estimation algorithms for (4) we can
196 utilize the fact that all elements of Θ are interrelated.

197 For example, the DREM approach can be applied to (4) allowing each
198 component of Θ to be estimated separately, then the noise $d(t)$ will be filtered
199 several times: first, by taking the power $d^n(t)$ of it, second, in the filter of
200 the extension part in DREM method.

201 Moreover, such a nonlinear filtering appears even if delays are used instead
202 of conventional convolution filters (which seem to be not yielding a filtering
203 ability). And this property we will demonstrate next by considering (4)
204 obtained by the power method.

205 Applying the delays in the DREM extension step we get an extended
206 version of (4):

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

207 where

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

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

208 where $\psi(t) \in \mathbb{R}$ and $v(t) \in \mathbb{R}$ are defined in (5), $\varphi(t) \in \mathbb{R}^n$ is defined in (6),
209 and $\tau \in \mathbb{N}$ is the elementary delay. If the new extended regressor Φ
210 nonsingular, then the direct estimates of Θ from (11) can be obtained:

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

211 where

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

212 is the implementable estimate that can be derived from (11) and $\mathcal{V}(t) =$
 213 $\Phi^{-1}(t)V(t)$ is the new noise proportional to d^n .

214 The following proposition shows that the first element of Θ corresponding
 215 to θ^n will be estimated as the geometric mean of the scalar evaluations $\hat{\theta}(t)$
 216 from (10) defined for non-zero regressor ϕ .

217 **Proposition 1.** *Consider the linear regression (2) where ϕ satisfies (7) and*
 218 *the related extended linear regression (11) for some $n \geq 2$ with a nonsingular*
 219 *matrix $\Phi(t)$. Then for the first component of $\hat{\Theta}$ in (12) it holds*

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

220 where $\hat{\theta}$ are defined by (10).

221 *Proof.* As in the proof of Theorem 1, recall that multiplication of a column or
 222 a row by a constant implies proportional change of the determinant. Recall
 223 the definition of φ in (6) and divide each i th column of the matrix Φ , $i =$
 224 $1, \dots, n$, by $(-1)^{n-i} \rho_{i-1}$. Next, divide each i th row of the matrix Φ , $i =$
 225 $1, \dots, n$, by $\phi(t - (i - 1)\tau)$ (due to (7) it is an admissible operation). Then

$$\det \Phi(t) = \prod_{i=1}^n \left((-1)^{n-i} \rho_{i-1} \phi(t - (i - 1)\tau) \right) \det \Phi_0(t),$$

226 where

$$\Phi_0(t) = \begin{bmatrix} c(\phi(t), y(t)) \\ c(\phi(t - \tau), y(t - \tau)) \\ \vdots \\ c(\phi(t - (n - 1)\tau), y(t - (n - 1)\tau)) \end{bmatrix}$$

227 and $c(\phi, y)$ was defined in (9) in Theorem 1. Due to Cramer's rule (Horn
 228 and Johnson, 2012), the first element of the vector $\hat{\Theta}$ is given by

$$\hat{\Theta}_1(t) = \frac{\det \bar{\Phi}_1(t)}{\det \Phi(t)},$$

229 where $\bar{\Phi}_1(t)$ denotes the matrix Φ with the first column replaced by the vector
 230 $Y(t)$.

231 Divide each i th column of the matrix $\bar{\Phi}_1(t)$ except the first one, $i =$
 232 $2, \dots, n$, by $(-1)^{n-i} \rho_{i-1}$. Next divide each i th row of $\bar{\Phi}_1(t)$, $i = 1, \dots, n$, by
 233 $y(t - (i - 1)\tau)$. Then

$$\begin{aligned} & \det \bar{\Phi}_1(t) \\ &= y(t) \prod_{i=2}^n \left((-1)^{n-i} \rho_{i-1} y(t - (i - 1)\tau) \right) \det \bar{\Phi}_{0,1}(t) \\ &= (-1)^{n-1} \prod_{i=1}^n \left((-1)^{n-i} \rho_{i-1} y(t - (i - 1)\tau) \right) \det \bar{\Phi}_{0,1}(t), \end{aligned}$$

234 where

$$\bar{\Phi}_{0,1}(t) = \begin{bmatrix} \bar{c}(\phi(t), y(t)) \\ \bar{c}(\phi(t - \tau), y(t - \tau)) \\ \vdots \\ \bar{c}(\phi(t - (n - 1)\tau), y(t - (n - 1)\tau)) \end{bmatrix}$$

235 and

$$\bar{c}(\phi, y) = [y^{n-1} \quad \phi^{n-1} \quad \phi^{n-2}y \quad \dots \quad \phi y^{n-2}].$$

236 Comparing $\bar{c}(\phi, y)$ and $c(\phi, y)$, it can be seen that the matrix $\bar{\Phi}_{0,1}(t)$ can be
 237 obtained from the matrix $\Phi_0(t)$ by performing $n - 1$ permutations of columns,
 238 and each permutation corresponds to the multiplication of the determinant
 239 of the matrix $\Phi_0(t)$ by -1 . Thus

$$\det \Phi_0(t) = (-1)^{n-1} \det \bar{\Phi}_{0,1}(t)$$

240 and

$$\begin{aligned} \hat{\Theta}_1(t) &= \frac{\det \bar{\Phi}_1(t)}{\det \Phi(t)} \\ &= \frac{\prod_{i=1}^n y(t - (i - 1)\tau)}{\prod_{i=1}^n \phi(t - (i - 1)\tau)} = \prod_{i=0}^{n-1} \hat{\theta}(t - i\tau) \end{aligned}$$

241 completing the proof. □

242 Note that the last element of Θ (that corresponds to θ) receives an esti-
 243 mate as the arithmetic mean of the same quantities:

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

244 The intermediate elements of Θ will be derived as respective products of
 245 $\hat{\theta}(t - i\tau)$ of given orders, i.e., for θ^2 it will be an averaged sum of $\hat{\theta}(t -$
 246 $i\tau)\hat{\theta}(t - k\tau)$ with $i \neq k = 0, \dots, n - 1$. The elements of the noise vector $\mathcal{V}(t)$
 247 will have corresponding dependence on $\delta(t) = \frac{d(t)}{\phi(t)}$, i.e., for the last element
 248 $\mathcal{V}_n(t) = \frac{1}{n} \sum_{i=0}^{n-1} \delta(t - i\tau)$.

249 Clearly, the geometric or arithmetic mean estimates, as well as others,
 250 can be derived directly, but it is interesting that they also can be obtained
 251 as particular solutions (for a special choice of delays as extension filters) of
 252 the DREM method applied to (4). Roughly speaking, it demonstrates what
 253 kind of nonlinear filtering can be derived by the proposed approach. Other
 254 filtering schemes can also be obtained, whose investigation is a direction of
 255 future research.

256 5. Example

257 Consider the linear regression (2), where

$$\phi(t) = 1 + \frac{1}{2} \sin(3T_s t),$$

258 where $T_s = 0.01$ second is the sampling period, and $t \in \mathbb{N}_*$. This scalar
 259 regression is PE and satisfies (7). The noise d is uniformly distributed in the
 260 interval from $-\bar{d}$ to \bar{d} with the positive constant $\bar{d} = 0.1$.

261 For comparison, we consider two estimates of θ :

- 262 • the direct estimate (10), which is further denoted as $\hat{\theta}^{\text{dir}}(t)$;
- 263 • the delay-based extension (11) and the estimate (12), where we use the
 264 result of Proposition 1 and compute

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

265 For this illustrative example, we do not consider other estimation meth-
 266 ods involving (implicitly or explicitly) dynamic noise filtration, such as low-
 267 pass filters, since our goal is to contrast the advantages of the new method
 268 (13) over the conventional approach (10) (at the same time focusing on the
 269 algebraic-like estimates (10) and (13)).

270 For numerical simulations, we set $\theta = 1$. For the power extension, we
 271 select $n = 3$. For the delay-based estimation, the delay is $\tau = 10$, which
 272 corresponds to 0.1 second.

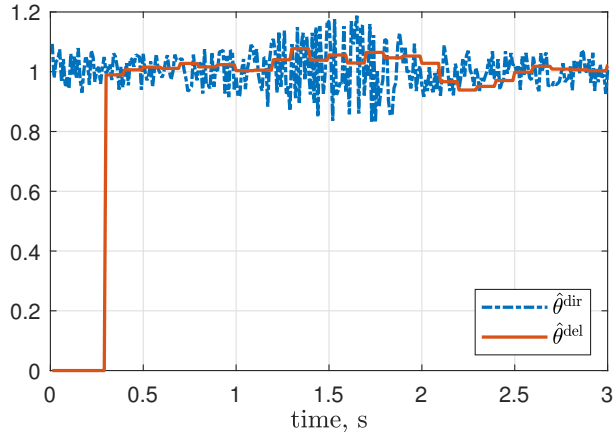


Figure 1: Transients of the direct estimate $\hat{\theta}^{\text{dir}}$ (10) and the delay-based estimate $\hat{\theta}^{\text{del}}$ (13). The true value is $\theta = 1$.

273 The transients are shown in Figure 1. The direct estimate $\hat{\theta}^{\text{dir}}$ provides
 274 an estimate after the first step; however, it has the worst noise sensitivity.
 275 The delay-based estimate converges after $n\tau T_s = 0.3$ seconds and show good
 276 noise attenuation; as shown in Proposition 1, such an estimate corresponds
 277 to the (nonlinear) geometric mean filtration.

278 6. Conclusion

279 An approach for noise attenuation in the linear regression was presented
 280 assuming that the perturbation is sufficiently small comparing to the re-
 281 gressor, or the regressor is sufficiently excited in relation with the noise
 282 amplitude. The method avoids the restrictions on stochastic or frequency
 283 properties of the noise. Two procedures for nonlinear transformation of a
 284 linear regression were presented. The obtained transformed linear regression
 285 has an augmented regressor and the vector of unknown parameters contain-
 286 ing the powers of θ and disturbance, while having attenuated the influence
 287 of the latter. It was demonstrated analytically that such a transformation
 288 of a regressor preserves the excitation under mild restrictions. The DREM
 289 method was applied to estimate the unknown parameters. The efficiency of
 290 the obtained estimators was illustrated and compared in numerical experi-
 291 ments.

292 **References**

- 293 Aranovskiy, S., Bobtsov, A., Ortega, R., Pyrkin, A., 2017. Performance
294 enhancement of parameter estimators via dynamic regressor extension and
295 mixing. *IEEE Transactions on Automatic Control* 62, 3546–3550.
- 296 Aranovskiy, S., Ushirobira, R., Korotina, M., Vedyakov, A., 2023. On
297 preserving-excitation properties of kreisselmeier’s regressor extension
298 scheme. *IEEE Transactions on Automatic Control* 68, 1296–1302.
- 299 Barabanov, N., Ortega, R., 2017. On global asymptotic stability of $\dot{x} =$
300 $-\phi(t)\phi^\top(t)x$ with $\phi(t)$ bounded and not persistently exciting. *Systems &*
301 *Control Letters* 109, 24–27.
- 302 Bobtsov, A., Aranovskiy, S., Efimov, D., Pyrkin, A., Vorobev, V., Wang, J.,
303 2024. Power noise filtration in drem, in: *Proc. European Control Confer-*
304 *ence, Stockholm*.
- 305 Chowdhary, G., Yucelen, T., Mühlegg, M., Johnson, E.N., 2012. Concurrent
306 learning adaptive control of linear systems with exponentially convergent
307 bounds. *International Journal of Adaptive Control and Signal Processing*
308 27, 280–301. doi:10.1002/acs.2297.
- 309 Efimov, D., Barabanov, N., Ortega, R., 2019. Robust stability under relaxed
310 persistent excitation conditions. *International Journal of Adaptive Control*
311 *and Signal Processing* doi:10.1002/acs.2997.
- 312 Goldenshluger, A., Polyak, B., 1993. Estimation of regression parameters
313 with arbitrary noise. *Mathematical Methods of Statistics* 2.
- 314 Granichin, O., 2004. Linear regression and filtering under nonstandard as-
315 sumptions (arbitrary noise). *IEEE Transactions on Automatic Control* 49,
316 1830–1837. doi:10.1109/TAC.2004.835585.
- 317 Horn, R., Johnson, C., 2012. *Matrix Analysis*. Cambridge University Press.
- 318 Ioannou, P., Kokotovic, P., 1983. *Adaptive Systems with Reduced Models*.
319 Springer Verlag, Secaucus, NJ.
- 320 Korotina, M., Aranovskiy, S., Ushirobira, R., Efimov, D., Wang, J., 2024. A
321 note on fixed- and discrete-time estimation via the drem method. *IEEE*
322 *Transactions on Automatic Control* , 1–6doi:10.1109/TAC.2024.3355803.

- 323 Ljung, L., 1987. System Identification: Theory for the User. Prentice-Hall,
324 Upper Saddle River, NJ.
- 325 Morgan, A., Narendra, K., 1977. On the uniform asymptotic stability of
326 certain linear nonautonomous differential equations. SIAM J. Control and
327 Optimization 15, 5–24.
- 328 Narendra, K., Annaswamy, A., 1987. Persistent excitation in adaptive
329 systems. International Journal of Control 45, 127–160. doi:10.1080/
330 00207178708933715.
- 331 Rencher, A., Christensen, W., 2012. Methods of Multivariate Analysis. vol-
332 ume 709 of *Wiley Series in Probability and Statistics*. 3rd ed., John Wiley
333 & Sons, New Jersey.
- 334 Sastry, S., Bodson, M., 1989. Adaptive Control: Stability, Convergence and
335 Robustness. Prentice-Hall, London.
- 336 Stepanov, O., Toropov, A., 2015. Nonlinear filtering for map-aided naviga-
337 tion. part 1. an overview of algorithms. Gyroscopy Navig , 324–337.
- 338 Wang, J., Efimov, D., Aranovskiy, S., Bobtsov, A., 2019. Fixed-time esti-
339 mation of parameters for non-persistent excitation. European Journal of
340 Control .
- 341 Wang, J., Efimov, D., Bobtsov, A., 2020. On robust parameter estimation
342 in finite-time without persistence of excitation. IEEE Transactions on
343 Automatic Control 65.
- 344 Yan, X., 2009. Linear Regression Analysis: Theory and Computing. World
345 Scientific, Singapore.
- 346 Yi, B., Ortega, R., 2022. Conditions for convergence of dynamic regressor
347 extension and mixing parameter estimators using lti filters. IEEE Trans-
348 actions on Automatic Control 68, 1253–1258.