A note on fixed-and discrete-time estimation via the DREM method

M. Korotina, Stanislav Aranovskiy, Rosane Ushirobira, Denis Efimov, Jian Wang

▶ **To cite this version:**

A note on fixed- and discrete-time estimation via the DREM method

Marina Korotina[1,2], Stanislav Aranovskiy[1], Rosane Ushirobira[3], Denis Efimov[3], and Jian Wang[4]

Abstract—A simple fixed-time converging estimation algorithm is presented for a linear regression using the dynamic regressor extension and mixing (DREM) method within a discrete-time setting, with a persistently excited regressor and bounded measurement noises. The solution is based on Kreisselmeier's filters, and it is computationally simpler than the existing analogs.

I. INTRODUCTION

In this note, we consider the parameter estimation problem for a discrete-time linear regression equation (LRE)

$$\overline{y}_k = \overline{\phi}_k^\top \overline{\theta} + \overline{v}_k, \ k \in \mathbb{N}, \tag{1}$$

where $\overline{y}_k, \overline{v}_k \in \mathbb{R}$ are the output and the bounded measurement noise, $\overline{\phi}_k$, $\overline{\theta} = (\theta_1 \ldots \theta_n)^\top \in \mathbb{R}^n$ are the bounded regressor and the unknown parameter vectors. The goal is to estimate $\overline{\theta}$ given the online measurements $\overline{y}_k$ and $\overline{\phi}_k$.

The parameter estimation problem for dynamic or static systems is an essential issue in many scientific disciplines, and many well-matured approaches to solving this problem are available [1], [2]. Despite various batch methods based on post-processing of collected measurements, two popular iterative parameter estimation methods are the recursive least-squares and gradient-descent estimators [3]. The fundamental result in parameter estimation claims that if the regressor $\overline{\phi}$ is persistently exciting (PE), that is exciting uniformly in time (see Section III for a formal definition of this property), then the recursive least-squares and gradient-descent iterative estimators converge exponentially, they are also input-to-state stable (ISS), and provide filtering of the noise $\overline{v}_k$.

Numerous recent studies address possible relaxation of the PE requirement in both continuous- and discrete-time settings. These results include the composite [4] and concurrent [5] learning, directional forgetting [6], fixed-time estimation [7], and the recent regularization-based results [8], [9]. However, it is widely understood that if the regressor $\overline{\phi}$ is not exciting uniformly in time, but only on a particular interval (interval excitation), then the additive noise cannot be efficiently filtered, and the PE-relaxation methods typically consider the noise-free scenario, $\overline{v} \equiv 0$. In this note, due to the presence of noise $\overline{v}_k$ in (1), we are focused on the uniform in time excitation, i.e., when the regressor $\overline{\phi}$ is PE.

Standard iterative estimators also have some known shortages. Among them, it is necessary to mention the difficulty of accelerating the convergence process (this property is mainly predefined by the level of excitation in the regressor) and

the non-monotonicity of the convergence [10]–[12]. The latter fact means that despite a decreasing norm of the parameter vector estimation error, the respective discrepancies for each parameter may demonstrate a complex oscillatory behavior before settling down in a vicinity of the true value, which may postpone a real-time utilization of the obtained estimates. That is why a solution to these drawbacks recently proposed by the dynamical regressor extension and mixing (DREM) method [13] has quickly gained popularity [14]–[17].

The DREM procedure consists of two steps: extension and mixing. In the first step, the model (1) is extended to get a new linear regression with a square $n \times n$ regressor matrix. Then, at the mixing step, the single extended LRE for a vector of unknown parameters is transformed into a set of n scalar LREs for each element of the vector θ independently. Then an iterative estimator, e.g., the gradient-descent one, is applied to each of the new scalar LREs separately, yielding the element-wise improved transients with simple and transparent estimator tuning rules.

However, in this note, we argue that *the use of iterative estimators in the DREM procedure is more a tradition than a need*. We show that a proper choice of an LRE extension at the first step of the DREM procedure ensures that the novel regressor is strictly positive allowing for point-wise (algebraic) estimation. A point-wise estimator can be of limited practical interest due to its noise sensitivity, even if it provides the fixed-time (independent in initial conditions) convergence. To this end, we study the worst-case bound on the parameter estimation error and propose tuning guidelines minimizing this bound.

To summarize, the contribution of this paper is as follows:
- we show that a particular choice for LRE extension ensures strict positiveness of the novel regressor;
- we analyze noise propagation for the point-wise (algebraic) estimator and derive the worst-case bound on estimation error.

Finally, we address the argument that standard iterative estimators provide a good trade-off between the transient time and the steady-state noise attenuation (filtering). We show that we obtain faster transients with a comparable steady-state performance by putting the point-wise estimator in a serial connection with a (non-)linear filter. Together, these contributions support our claim that under the PE condition, the DREM procedure can function as a fixed-time estimator without being followed by gradient-type algorithms.

The rest of the paper is organized as follows. In Section II we briefly describe the DREM procedure. In Section III we discuss how the initial excitation can be preserved by the DREM procedure. In Section IV we present the main result of the note, and Section V contains simulations illustrating our results. Finally, Section VI contains some concluding remarks.

[1]CentraleSupélec–IETR, 35576 Cesson-Sévigné, France.
[2]ITMO University, 197101 Saint Petersburg, Russia.
[3]Inria, Univ. Lille, CNRS, UMR 9189 - CRIStAL, F-59000 Lille, France.
[4]Hangzhou Dianzi University, Hangzhou, China. Corresponding author, wangjian119@hdu.edu.cn

Notation

- The sets of nonnegative real and integer numbers are denoted by $\mathbb{R}_+$ and $\mathbb{N}$, respectively. Also, $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$. The set of real $n \times m$-matrices is denoted by $\mathbb{R}^{n \times m}$. The n-identity matrix is denoted by I_n.
- For a vector $x \in \mathbb{R}^n$, $\|x\|$ denotes its Euclidean norm.
- The rounding function to the greatest integer smaller than $s \in \mathbb{R}$ is denoted by $\lfloor s \rfloor = \text{floor}(s)$.

II. DYNAMICAL REGRESSOR EXTENSION AND MIXING

The original DREM procedure in [13] consists in finding n stable causal filters $H_i(z)$, where $i \in \{1, \dots, n\}$, and z is the time shift operator (*i.e.*, $z^h \overline{y}_k = \overline{y}_{k-h}$ for any $k, h \in \mathbb{N}$, $k \geq h$), whose auxiliary role is also to filter the noise. For $i \in \{1, \dots, n\}$, denote

$$Y_k^i = H_i(z)\overline{y}_k, \ \Phi_k^i = H_i(z)\overline{\phi}_k, \ V_k^i = H_i(z)\overline{v}_k, \quad (2)$$
$$Y_k = (Y_k^1 \dots Y_k^n)^\top, \ V_k = (V_k^1 \dots V_k^n)^\top \in \mathbb{R}^n,$$
$$\Phi_k = (\Phi_k^1 \dots \Phi_k^n)^\top \in \mathbb{R}^{n \times n}.$$

That leads to a new extended regression

$$Y_k = \Phi_k \overline{\theta} + \overline{V}_k, \ \overline{V}_k = V_k + \epsilon_k, \ k \in \mathbb{N}, \quad (3)$$

where $\epsilon_k \in \mathbb{R}^n$ is an exponentially decaying term coming from the initialization of the filters. Finally, for

$$\phi_k = \det(\Phi_k), \ y_k = \text{adj}(\Phi_k) Y_k, \ v_k = \text{adj}(\Phi_k) \overline{V}_k \quad (4)$$

this yields the element-wise scalar linear regression:

$$y_{k,i} = \phi_k \theta_i + v_{k,i}, \ k \in \mathbb{N}, \ i \in \{1, \dots, n\}, \quad (5)$$

where $y_{k,i} \in \mathbb{R}$ and $\phi_k \in \mathbb{R}$ are known signals, $\theta_i \in \mathbb{R}$ is the unknown constant parameter to be estimated, and $v_{k,i} \in \mathbb{R}$ is an unknown bounded measurement distortion. Note that the estimation of each θ_i is now explicitly independent of other components of $\overline{\theta}$, and the interconnection is hidden in $\phi_k \in \mathbb{R}$.

For $k \in \mathbb{N}$, let $\widehat{\theta}_k \in \mathbb{R}^n$ denote an estimate of $\overline{\theta}$ at time k. Define $e_k := y_k - \phi_k \widehat{\theta}_k = \phi_k \widetilde{\theta}_k + v_k$ as the measured regression error, where $\widetilde{\theta}_k := \overline{\theta} - \widehat{\theta}_k$ is the parameter estimation error. It remains now to design an algorithm to compute $\widehat{\theta}_k$, which is a simpler problem than (1), since in (5) the estimation of each θ_i can be treated separately. Moreover, in scalar linear regression, the convergence of $\widetilde{\theta}_k$ is element-wise monotone and can be accelerated by tuning the estimation procedures [13].

One of the main issues of the DREM methodology is the excitation of the common regressor ϕ_k, and its relation with the excitation of $\overline{\phi}_k$ in the original problem statement, which is obviously predefined by the choice of the filters H_i in (2). Let us present a solution to this problem.

III. KREISSELMEIER'S REGRESSOR EXTENSION

Let us show how the DREM procedure can keep the excitation level of the original regression for (5). To this end, let us characterize the admissible excitation levels [10]:

Definition 1 (PE). *A bounded signal* $\overline{\phi} : \mathbb{N} \to \mathbb{R}^n$ *is called* (ℓ, μ)-*persistently excited (PE) if there exist* $\ell \in \mathbb{N}^*$ *and* $\mu > 0$ *such that*

$$\sum_{i=0}^{\ell-1} \overline{\phi}_{k+i} \overline{\phi}_{k+i}^\top \geq \mu I_n, \ \forall k \in \mathbb{N}.$$

It is said to be intervally excited (IE) *if the above inequality is satisfied for* $k = 0$ *only*.

It worth noting that in the noise-free case, interval excitation can be related with the conventional identifiability/observability condition for $\overline{\theta}$ in (1).

To keep the excitation, we proceed as follows. Instead of applying filters H_i directly to (1), consider the auxiliary regression problem

$$\overline{\phi}_k \overline{y}_k = \overline{\phi}_k \overline{\phi}_k^\top \overline{\theta} + \overline{\phi}_k \overline{v}_k$$

and apply $H(z) = \frac{1-\lambda}{z^{-1}-\lambda}$, where $\lambda \in (0, 1)$. Then for

$$Y_k = H(z)\overline{\phi}_k \overline{y}_k, \ \Phi_k = H(z)\overline{\phi}_k \overline{\phi}_k^\top, \ V_k = H(z)\overline{\phi}_k \overline{v}_k,$$

we obtain the extended LRE (3). The state-space implementation of this extension is known as Kreisselmeier's dynamic regressor extension and is given by

$$\Phi_{k+1} = \lambda \Phi_k + (1 - \lambda) \overline{\phi}_k \overline{\phi}_k^\top, \quad (6)$$
$$Y_{k+1} = \lambda Y_k + (1 - \lambda) \overline{\phi}_k \overline{y}_k,$$

where $Y_0 \in \mathbb{R}^n$ and $\Phi_0 \in \mathbb{R}^{n \times n}$, $\Phi_0 = \Phi_0^\top \geq 0$. After the mixing step, it yields the decoupled linear regression equation (5). Specifically, the choice $\Phi_0 = 0$ and $Y_0 = 0$ guarantees that $\epsilon_k \equiv 0$ in (3) for all $k \in \mathbb{N}$.

The next Lemma shows that the proposed extension keeps the excitation properties of the original regressor $\overline{\phi}$.

Lemma 1. *Consider a* (ℓ, μ)-*PE signal* $\overline{\phi}$ *and the Kreisselmeier's dynamic extension* (6). *Then the signal* ϕ *defined in* (4) *is* (ℓ, α)-*PE, with*

$$\alpha = \left(\mu (1 - \lambda) \lambda^{\ell-1} \right)^n.$$

Moreover,

$$\phi_k \geq \alpha, \ \forall k \geq \ell; \quad \liminf_{k \to \infty} \phi_k \geq \alpha \left(\frac{1}{1 - \lambda^\ell} \right)^n.$$

Proof: The solution of (6) is

$$\Phi_k = \lambda^k \Phi_0 + (1 - \lambda) \sum_{i=0}^{k-1} \psi(k, i), \quad (7)$$

with $\psi(k, i) = \lambda^{k-1-i} \overline{\phi}_i \overline{\phi}_i^\top$. Set $q = \lfloor \frac{k}{\ell} \rfloor$ and $v = k - q\ell$, so $k = q\ell + v$. Then for all $k \geq \ell$

$$\sum_{i=0}^{k-1} \psi(k, i) = \sum_{j=1}^{q} \left(\sum_{i=k-j\ell}^{k-1-(j-1)\ell} \psi(k, i) \right) + \sum_{i=0}^{v-1} \psi(k, i).$$

Since $\psi(k, i) \geq 0$ for all $k, i \in \mathbb{N}$ and $\Phi_0 \geq 0$, it follows from (7) that for all $k \geq \ell$

$$\Phi_k \geq (1 - \lambda) \sum_{j=1}^{q} \left(\sum_{i=k-j\ell}^{k-1-(j-1)\ell} \psi(k, i) \right).$$

Recalling that $\lambda < 1$, for all $k \geq \ell$ and $j \in \{1, \ldots q\}$, it holds

$$\sum_{i=k-j\ell}^{k-1-(j-1)\ell} \psi(k,i) \geq \lambda^{j\ell-1} \sum_{i=k-j\ell}^{k-1-(j-1)\ell} \overline{\phi}_i \overline{\phi}_i^\top$$
$$= \lambda^{j\ell-1} \sum_{i=0}^{\ell-1} \overline{\phi}_{k-j\ell+i} \overline{\phi}_{k-j\ell+i}^\top,$$

and using that $\overline{\phi}$ is (ℓ, μ)-PE, we obtain

$$\sum_{i=k-j\ell}^{k-1-(j-1)\ell} \psi(k,i) \geq \lambda^{j\ell-1} \mu I_n.$$

Thus

$$\Phi_k \geq \mu(1-\lambda) \sum_{j=1}^{q} \lambda^{j\ell-1} I_n = \mu(1-\lambda)\lambda^{\ell-1} \frac{1-\lambda^{q\ell}}{1-\lambda^\ell} I_n.$$

Then for $\phi_k = \det(\Phi_k)$, we obtain for all $k \geq \ell$ (and since $q \geq 1$)

$$\phi_k \geq \left(\mu(1-\lambda)\lambda^{\ell-1}\right)^n = \alpha.$$

Moreover, $\lambda^{q\ell} \to 0$ as $k \to \infty$, and consequently

$$\liminf_{k \to \infty} \phi_k \geq \left(\mu(1-\lambda)\frac{\lambda^{\ell-1}}{1-\lambda^\ell}\right)^n.$$

∎

In the continuous-time setting, a similar result on the choice of filters is given in [18]. Of course, another filter selection is possible, providing the excitation properties of (1) in (5).

Remark 1. *If $\overline{\phi}$ in (1) is IE, then ϕ resulting from (4), (6) cannot be identically zero, and if $\overline{\phi}$ is PE, then ϕ is strictly positive for all $k \geq \ell$. Thus we further consider only those steps $k \in \mathbb{N}$ when $\phi_k \neq 0$. Indeed, otherwise the linear regression (5) reads $y_k = v_k$ and there is no reason to update the estimate for such samples, i.e., $\widehat{\theta}_k = \widehat{\theta}_{k-1}$ for $\phi_k = 0$ (recall that the signal ϕ is known).*

IV. NOISE PROPAGATION

In the noise-free case, after applying the DREM method and obtaining the decoupled linear regression (5), the estimation problem has a trivial point-wise solution

$$\widehat{\theta}_k = \frac{y_k}{\phi_k}, \tag{8}$$

which can be unconditionally applied as soon as $\phi_k \neq 0$ for some $k \in \mathbb{N}$. Such a condition is always verified with an IE-signal ϕ.

In the presence of noise, the noise propagation can be studied using the established properties of Kreisselmeier's filters (6), and the following simple result can be obtained constituting the main outcome of this note:

Theorem 1. *Let the regressor signal $\overline{\phi}$ in (1) be (ℓ, μ)-PE and the Kreisselmeier's dynamic extension filters (6) be used in (2). Then the estimate $\widehat{\theta}_k$ can be computed by (4), (8), for all $k \geq \ell$, and*

$$\sup_{k \geq \ell} \|\widetilde{\theta}_k\| \leq \frac{\sup_{k \geq \ell} \|v_k\|}{\alpha},$$
$$\limsup_{k \to +\infty} \|\widetilde{\theta}_k\| \leq (1-\lambda^\ell)^n \frac{\limsup_{k \to +\infty} \|v_k\|}{\alpha}.$$

Proof: According to Lemma 1, for all $k \geq \ell$, $\phi_k \geq \alpha = \left(\mu(1-\lambda)\lambda^{\ell-1}\right)^n$, hence the division in (8) is well defined. By definition,

$$\widetilde{\theta}_k = \overline{\theta} - \frac{y_k}{\phi_k} = \frac{v_k}{\phi_k}$$

giving the required estimates by the properties of ϕ_k. ∎

Thus, the DREM-based algorithm (6), (4), (8) guarantees a *fixed-time* convergence of the estimate $\widehat{\theta}_k$ to the true value $\overline{\theta}$ (the convergence time $\ell \geq n$ is independent of initial conditions), while the estimation error is proportional to the noise v amplitude.

Remark 2. *According to the estimates given in Theorem 1, the choice $\lambda^* = \frac{\ell-1}{\ell}$ minimizes the parameter estimation error gain α^{-1} with respect to the noise. Augmenting the value of λ can improve the asymptotic precision of the proposed algorithm.*

Several final comments are in order:
1) Roughly speaking, the DREM method provides a direct measurement of the unknown parameter vector $\overline{\theta}$ for $k \geq \ell$:

$$\mathcal{Y}_k = \overline{\theta} + \mathcal{V}_k, \quad \text{with } \mathcal{Y}_k = \phi_k^{-1} y_k, \ \mathcal{V}_k = \phi_k^{-1} v_k.$$

Note also that (4), (6), (8) implicitly realize exponentially-weighted least-squares estimation with a forgetting factor λ.

2) Any additional noise filtering can be applied to $\mathcal{Y}_k$ to attenuate the influence of $\mathcal{V}_k$. In the existing literature, gradient algorithms are frequently used for (5) [7], [16], [19], [20], which serve merely for implicit noise-canceling, being probably not the best choice for such a purpose. Indeed, the conventional selection is the least-squares algorithm:

$$\widehat{\theta}_k = \widehat{\theta}_{k-1} + \gamma_k \phi_k (y_k - \phi_k \widehat{\theta}_{k-1}), \ \gamma_k \in (0, 2\phi_k^{-2}).$$

For $\gamma_k = \phi_k^{-2}$, it is reduced to (8), while for $\gamma_k = \varrho \phi_k^{-2}$ with $\varrho \in (0,1)$, to a low-pass filtering of (8). In general, many more options are available, as the moving-average filter or nonlinear designs that cannot be directly applied to the linear regression (5), e.g., the median filter. See Section V for an illustrative comparison of selected methods.

3) Note that Kreisselmeier's dynamic extension (6) filters the noise $\overline{v}_k$ already, or other noise filters can be applied to get (1).

V. SIMULATIONS

A. General description

We consider the standard example where we estimate parameters of a sinusoidal signal:

$$\overline{y}_k = A_0 + A\sin(\omega k + \psi) + \overline{v}_k, \ k \in \mathbb{N}, \tag{9}$$

which can be written as (1) with

$$\overline{\phi}_k = \begin{pmatrix} 1 & \sin(\omega k) & \cos(\omega k) \end{pmatrix}^\top,$$
$$\overline{\theta} = \begin{pmatrix} A_0 & A\cos(\psi) & A\sin(\psi) \end{pmatrix}^\top,$$

where $\omega = \frac{2\pi}{T}$, $T \in \mathbb{N}$, $T \geq 3$, is the known period, and $A_0 \in \mathbb{R}$, $A \in \mathbb{R}_+$, $\psi \in [0, 2\pi)$ are unknown constant parameters. For the conciseness of simulations, we focus below on estimating the parameter $\theta_1 = A_0$ only.

The DREM procedure (6) with zero initial conditions yields the element-wise scalar linear regression (5), where we are interested in the equation corresponding to $i = 1$ only, namely

$$y_{k,1} = \phi_k \theta_1 + v_{k,1}. \tag{10}$$

Note that the DREM procedure allows for the single parameter estimation; otherwise, the whole vector $\overline{\theta}$ has to be estimated.

It is straightforward to show that the regressor $\overline{\phi}$ is (ℓ, μ)-PE with $\ell = T$ and $\mu = \frac{T}{2}$, and Lemma 1 applies. The signal ϕ is thus also PE and it is strictly separated from zero for $k \geq T$. Then the direct estimation (8) is applied,

$$\hat{\theta}_{k,1}^{\mathrm{dir}} = \frac{y_{k,1}}{\max\{\phi_k, \varepsilon_\phi\}}, \tag{11}$$

where $\varepsilon_\phi > 0$ is a small constant introduced to ensure the feasibility of implementing the direct estimation for the initial transients of (6).

In the sequel, we consider various filtering techniques applied to the direct estimate (11) and compare with estimators applied to the element-wise linear regression model (10). The results illustrate that fixed-time direct estimation followed by filtering performs better than estimators with asymptotic convergence.

B. Used approaches

By abuse of notation, we use y_k, θ, and $\hat{\theta}_k$ to denote $y_{k,1}$, θ_1, and $\hat{\theta}_{k,1}$, resp., since we are interested only in $\hat{\theta}_1$.

For the element-wise linear regression (10), we consider the following two estimators:

- the gradient estimator given by

$$\hat{\theta}_k^{\mathrm{gr}} = \hat{\theta}_{k-1}^{\mathrm{gr}} + \frac{\gamma_0}{1 + \gamma_0 \phi_k^2} \left(y_k - \phi_k \hat{\theta}_{k-1}^{\mathrm{gr}} \right), \tag{12}$$

  where $\gamma_0 > 0$ is the tuning coefficient;
- the least-squares estimator with forgetting given by

$$p_k = \frac{p_{k-1}}{\lambda_{ls} + \phi_k^2 p_{k-1}}, \ \ \hat{\theta}_k^{\mathrm{ls}} = \hat{\theta}_{k-1}^{\mathrm{ls}} + p_k \phi_k \left(y_k - \phi_k \hat{\theta}_{k-1}^{\mathrm{ls}} \right), \tag{13}$$

  with $p_0 > 0$ and $\lambda_{ls} \in (0, 1)$ is the forgetting factor.

For the direct estimate, we consider the following filtering methods applied to $\hat{\theta}_k^{\mathrm{dir}}$ given in (11):

- the low-pass filtering given by

$$\hat{\theta}_k^{\mathrm{low}} = \varrho \hat{\theta}_{k-1}^{\mathrm{low}} + (1 - \varrho) \hat{\theta}_k^{\mathrm{dir}}, \tag{14}$$

  where $\varrho \in (0, 1)$ is the tuning coefficient;
- the moving-average filtering given by

$$\hat{\theta}_k^{\mathrm{ma}} = \frac{1}{N} \sum_{i=0}^{N-1} \hat{\theta}_{k-i}^{\mathrm{dir}}, \tag{15}$$

  where $N = \min\{k, N_{\mathrm{ma}}\}$, and N_{ma} is the length of the averaging window;
- the median filter given by

$$\hat{\theta}_k^{\mathrm{med}} = \mathrm{median}\left\{ \hat{\theta}_{k-N+1}^{\mathrm{dir}}, \ldots, \hat{\theta}_k^{\mathrm{dir}} \right\}, \tag{16}$$

where $N = \min\{k, N_{\mathrm{med}}\}$, and N_{med} is the number of samples used in the median computation,

- the Kalman filter given by

$$\hat{\theta}_k^{\mathrm{kal}} = \hat{\theta}_{k-1}^{\mathrm{kal}} + \frac{P_{k-1} + Q}{P_{k-1} + Q + R_k} \left(\frac{y_k}{\phi_k} - \hat{\theta}_{k-1}^{\mathrm{kal}} \right),$$

$$P_k = \frac{R_k}{P_{k-1} + Q + R_k} \left(P_{k-1} + Q \right), \tag{17}$$

$$R_k = \frac{R_0}{\max\left\{ \phi_k^2, \varepsilon_\phi^2 \right\}},$$

where $Q \geq 0$ and $R_0 > 0$ are the process and observation noise covariances. When $Q \neq 0$, the case of time-varying parameters can be treated.

C. Simulation results

Consider the signal (9), where $A_0 = -3$, $A = 5$, $\psi = \frac{\pi}{3}$, and $T = 117$. The measurement noise $\overline{v}$ is chosen as a uniform random variable, $\overline{v} \sim U(-1, 1)$. Following Remark 2, the value λ in (6) is chosen as $\lambda = \frac{T-1}{T}$ and the initial conditions in (6) are zero. Then the lower bound α defined in Lemma 1 is $\alpha = 0.0063$ and the asymptotic lower bound of ϕ is 0.025.

For the direct estimation, the value ε_ϕ in (11) is chosen as $\varepsilon_\phi = 10^{-8}$. The initial values in the estimators (12), (13) and filters (14), (17) are zero; the initial values of p_0 in (13) and P_0 in (17) equal to one. The parameters of the estimators (12), (13) and filters (14), (15), (17) are chosen to provide equal transient time in the ideal noise-free scenario, where the transient time is measured as the step number when the estimation errors decays to 5% of the initial value. The values are $\gamma_0 = 10^3$, $\lambda_{ls} = e^{-\frac{3}{28}}$, $\varrho = e^{-\frac{3}{73}}$, $N_{\mathrm{ma}} = 76$, and $Q = 2.5 \cdot 10^{-6}$, $R_0 = 1.75 \cdot 10^{-4}$. For the median filter (16), the number of used samples is equal to the length of the averaging window in (15), $N_{\mathrm{med}} = N_{\mathrm{ma}}$.

Fig. 1 depicts the estimation error transients of the considered methods in the ideal noise-free scenario. Denote the step number when the value ϕ_k overcomes the threshold ε_ϕ as T_{fix}, where $T_{\mathrm{fix}} = 14$ in the considered example. The direct estimate (11) converges in the fixed time T_{fix}. The median filter also has the fixed-time convergence property with the convergence time $\min\left\{ 2T_{\mathrm{fix}}, T_{\mathrm{fix}} + \left\lfloor \frac{N_{\mathrm{med}}}{2} \right\rfloor \right\}$. The moving-average filter converges in the fixed time $T_{\mathrm{fix}} + N_{\mathrm{ma}} - 1$, and other estimates converge asymptotically.

Fig. 2 depicts the estimation error transients of the considered methods for noisy measurements. The direct estimate exhibits sensitivity to the noise in the initial transient when the value of ϕ is small; this sensitivity can be reduced by increasing ε_ϕ. Nevertheless, all considered methods perform well in alleviating the noise, where filtering of the direct estimate converges to a vicinity of the true value faster than the estimators applied to the linear regression model.

The steady-state performance of the considered methods is summarized in Table I, where the mean squared error (MSE) and the mean absolute error (MAE) are given; these values are computed over 10^6 samples after the transients.

To summarize, under equivalent tuning, the direct estimate (8) provides fixed-time convergence, and followed by a filter
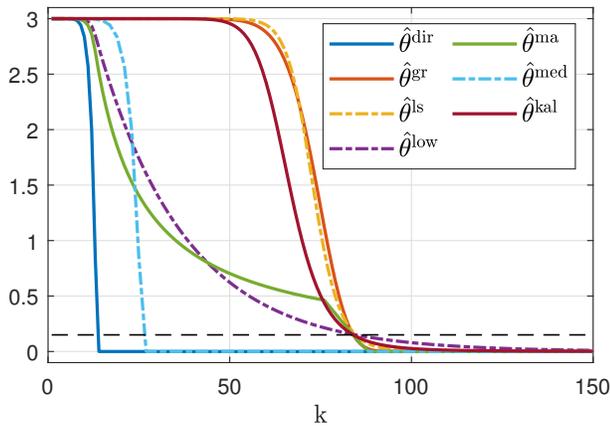
Fig. 1. Estimation error transients in the noise-free case: the direct estimation $\hat{\theta}^{\mathrm{dir}}$ (11), the gradient estimator $\hat{\theta}^{\mathrm{gr}}$ (12), the least-squares estimator $\hat{\theta}^{\mathrm{ls}}$ (13), the low-pass filtering $\hat{\theta}^{\mathrm{low}}$ (14), the moving-average filtering $\hat{\theta}^{\mathrm{ma}}$ (15), the median filtering $\hat{\theta}^{\mathrm{med}}$ (16), the Kalman filter $\hat{\theta}^{\mathrm{kal}}$ (17). The dashed line corresponds to 5% of the initial error value; the estimators are tuned to have approximately equal transient time. For the median filter, the number of samples is the same as for the moving average.
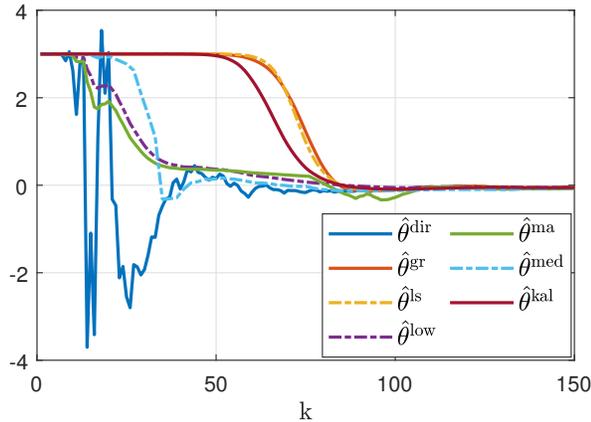

Fig. 2. Estimation error transients for noised measurements: the direct estimation $\hat{\theta}^{\mathrm{dir}}$ (11), the gradient estimator $\hat{\theta}^{\mathrm{gr}}$ (12), the least-squares estimator $\hat{\theta}^{\mathrm{ls}}$ (13), the low-pass filtering $\hat{\theta}^{\mathrm{low}}$ (14), the moving-average filtering $\hat{\theta}^{\mathrm{ma}}$ (15), the median filtering $\hat{\theta}^{\mathrm{med}}$ (16), the Kalman filter $\hat{\theta}^{\mathrm{kal}}$ (17).

provides parameter estimation with an improvement in the noise attenuation, and it should be considered as a reasonable alternative to the standard parameter estimation approaches applied to the element-wise scalar linear regression (5).

VI. CONCLUSION

This work has established that the DREM procedure provides a direct parameter estimation by transforming the initial vector linear regression into one with unknown parameters directly measured with noise. We have also proved that Kreisselmeier's filters preserve the initial excitation of the

TABLE I
MSE AND MAE OF THE SELECTED METHODS.

	$\hat{\theta}^{\mathrm{gr}}$	$\hat{\theta}^{\mathrm{ls}}$	$\hat{\theta}^{\mathrm{low}}$	$\hat{\theta}^{\mathrm{ma}}$	$\hat{\theta}^{\mathrm{med}}$	$\hat{\theta}^{\mathrm{kal}}$
MSE $\cdot 10^3$	1.4	1.3	1.2	1.2	1.2	1.1
MAE $\cdot 10^2$	3.0	3.0	2.7	2.7	2.7	2.6

regressor in the DREM method. Noise filters can be effectively coupled with DREM. In the future, we shall investigate other estimation algorithms with accelerated convergence rates in discrete time, especially when it is impossible to obtain a sign-definite regressor after DREM or to improve robustness to measurement noise.

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1987.

[2] X. Yan, *Linear Regression Analysis: Theory and Computing*. Singapore: World Scientific, 2009.

[3] I. D. Landau, R. Lozano, M. M'Saad, and A. Karimi, *Adaptive control: algorithms, analysis and applications*. Springer Science & Business Media, 2011.

[4] 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.

[5] Y. Pan, S. Aranovskiy, A. Bobtsov, and H. Yu, "Efficient learning from adaptive control under sufficient excitation," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 10, pp. 3111–3124, 2019.

[6] A. Goel, A. L. Bruce, and D. S. Bernstein, "Recursive least squares with variable-direction forgetting: Compensating for the loss of persistency [lecture notes]," *IEEE Control Systems Magazine*, vol. 40, no. 4, pp. 80–102, 2020.

[7] J. Wang, D. Efimov, S. Aranovskiy, and A. Bobtsov, "Fixed-time estimation of parameters for non-persistent excitation," *European Journal of Control*, 2019.

[8] R. Marino and P. Tomei, "On exponentially convergent parameter estimation with lack of persistency of excitation," *Systems & Control Letters*, vol. 159, p. 105080, 2022.

[9] M. Bin, "Generalized recursive least squares: Stability, robustness, and excitation," *Systems & Control Letters*, vol. 161, p. 105144, 2022.

[10] K. S. Narendra and A. M. Annaswamy, "Persistent excitation in adaptive systems," *International Journal of Control*, vol. 45, no. 1, pp. 127–160, 1987.

[11] D. Efimov and A. Fradkov, "Design of impulsive adaptive observers for improvement of persistency of excitation," *Int. J. Adaptive Control and Signal Processing*, vol. 29, no. 6, pp. 765–782, 2015.

[12] D. Efimov, N. Barabanov, and R. Ortega, "Robust stability under relaxed persistent excitation conditions," *International Journal of Adaptive Control and Signal Processing*, 2019.

[13] 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, 2017.

[14] N. Barroso, R. Ushirobira, D. Efimov, M. Sow, and J.-C. Massabuau, "Model-based adaptive filtering of harmonic perturbations applied to high-frequency noninvasive valvometry," in *IFAC World Congress*, Berlin, Germany, 2020.

[15] X. Shao, Q. Hu, Y. Shi, and B. Yi, "Data-driven immersion and invariance adaptive attitude control for rigid bodies with double-level state constraints," *IEEE Transactions on Control Systems Technology*, 2021.

[16] A. Vediakova, A. Vedyakov, A. Pyrkin, A. Bobtsov, and V. Gromov, "Finite time frequency estimation for multi-sinusoidal signals," *European Journal of Control*, vol. 59, pp. 38–46, 2021.

[17] J. Schiffer, P. Aristidou, and R. Ortega, "Online estimation of power system inertia using dynamic regressor extension and mixing," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4993–5001, 2019.

[18] S. Aranovskiy, R. Ushirobira, M. Korotina, and A. Vedyakov, "On preserving-excitation properties of kreisselmeiers regressor extension scheme," *IEEE Transactions on Automatic Control*, 2022.

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

[20] R. Ortega, V. Nikiforov, and D. Gerasimov, "On modified parameter estimators for identification and adaptive control. a unified framework and some new schemes," *Annual Reviews in Control*, vol. 50, pp. 278–293, 2020.