Accelerated convergence with improved robustness for discrete-time parameter estimation

Stanislav Aranovskiy[a], Rosane Ushirobira[b], Denis Efimov[b], Jian Wang[c,*]

[a] *CentaleSupélec –IETR, Avenue de la Boulaie, 35576 Cesson-Sévigné, France.*
[b] *Inria, Univ. Lille, CNRS, UMR 9189 - CRIStAL, F-59000 Lille, France.*
[c] *Hangzhou Dianzi University, Hangzhou, China.*

Abstract

The dynamic regressor extension and mixing (DREM) method provides a fixed-time converging parameter estimator for persistently excited regressor under bounded measurement noises. This note aims to develop this approach for cases with weaker excitation and regressor constraints. Several nonlinear estimation schemes with fixed-time convergence rates and improved measurement noise robustness properties are proposed here.

1. Introduction

Identifying parameters in linear regression is a crucial theoretical and applied challenge in control engineering [1]. Many widely-known approaches [2], may be used to tackle this problem, generally with the condition that the regressor is suitably excited [3, 4]. Additional features are often demanded in real-world applications, including minimization of the convergence time, monotonicity of such a convergence, and robustness to uncertainties and noises. A solution recently proposed by the dynamic regressor extension and mixing (DREM) method [5] answered all these challenges and quickly became popular in the community [6, 7, 8]. In addition, the excitation constraints have been slightly relaxed with this technique.

[⋆]This work was supported by the China 111 Project D17019
[*]Corresponding author, wangjian119@hdu.edu.cn

The DREM method helps reduce the initial estimation problem for a vector of unknown parameters to a series of interrelated scalar linear regression problems, with a posterior application of various iterative schemes for asymptotic or finite/fixed-time evaluation of parameter values [6, 9, 7, 8, 10, 11]. Such a decomposition on individual scalar regressions for each parameter allows the accelerated convergence and monotonicity to be ensured, and at the same time, it also helps with a quantitative evaluation of robustness characteristics. For example, in [12], a practical fixed-time DREM-based scheme was proposed for a discrete-time setting, which used Kreisselmeier's filters to extend the dynamics. In such a case, the persistently excited regressor of the original vector estimation problem is transformed into a positive scalar one. This sign-definiteness property simplifies the analysis and tuning, directly connecting with the noise filtering framework.

The present work proposes the development of [12] considering the cases when the regressor may fail to be persistently excited or when Kreisselmeier's filters cannot be utilized, and the regressor after mixing, being certainly excited, is not strictly separated from zero. In these scenarios, the additional iterative linear or nonlinear estimation schemes have to be introduced after DREM, and some of them are designed and analyzed in this paper.

Notation

- The sets of real, nonnegative, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{N}$, respectively, $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$. The sets of real $m \times n$-matrices and real n-vectors are denoted by $\mathbb{R}^{m\times n}$ and $\mathbb{R}^n$, respectively. The identity matrix in $\mathbb{R}^{n\times n}$ is denoted by I_n.

- The rounding function to the biggest integer lower than $s \in \mathbb{R}$ is denoted by $\lfloor s \rfloor = \text{floor}(s)$.

- e denotes the exponential function.

- Definitions of input-to-state stability (ISS) and integral ISS (iISS) for discrete-time systems can be found in [13, 14] (it is worth highlighting that in such a setting, iISS follows from global asymptotic stability).

2. Problem statement

Consider a vector linear regression in discrete-time:

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

where $\overline{y}_k$, $\overline{v}_k \in \mathbb{R}$ are the output and the measurement noise, $\overline{\phi}_k \in \mathbb{R}^n$ is the regressor, and $\overline{\theta} = (\theta_1 \dots \theta_n)^\top \in \mathbb{R}^n$ is the vector of unknown parameters.

The DREM procedure applied to (1) (see the Appendix) yields the element-wise scalar linear regression

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

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.

Since the estimation of each θ_i is now independent of other components of $\overline{\theta}$, then in the sequel, the index i will be omitted, and the scalar linear regression

$$y_k = \phi_k \theta + v_k,\ k \in \mathbb{N} \tag{2}$$

with $y_k, \phi_k, \theta, v_k \in \mathbb{R}$ will be considered without losing generality.

Problem. Our goal is to estimate θ from the measurements of y_k and ϕ_k in the fastest way, taking into account the presence of the noise v_k.

For $k \in \mathbb{N}$, let $\widehat{\theta}_k$ denote an estimate of θ at the step k. Define $e_k := y_k - \phi_k \widehat{\theta}_k = \phi_k \widetilde{\theta}_k + v_k$ as the measured regression error, where $\widetilde{\theta}_k := \theta - \widehat{\theta}_k$ is the parameter estimation error.

Remark 1. Note that in the noise-free setting, the stated problem has a trivial solution

$$\widehat{\theta}_k = \frac{y_k}{\phi_k}$$

that can be applied as soon as $\phi_k \neq 0$ for some $k \in \mathbb{N}$. Hence, the noise robustness analysis is crucial.

Remark 2. Moreover, we can always assume that $\phi_k \neq 0$ for all $k \in \mathbb{N}$ since otherwise, $y_k = v_k$ (since ϕ_k is also available for measurements, such an event is detectable), and there is no sense in acting (such a situation with $\phi_k = 0$ can be interpreted as the absence of the measurements, and the next event k should not be generated).

According to the Excitation Preservation Lemma given in the Appendix, if $\overline{\phi}_k$ in (1) is (ℓ, μ)-persistently excited (in the sense of Definition 1 given in the Appendix) for some $\ell \in \mathbb{N}^*$ and $\mu > 0$, and Kreisselmeier's filters (17) with the parameter $\lambda \in (0, 1)$ are used in DREM, then $\phi_k \geq \alpha = \left(\mu\,(1-\lambda)\,\lambda^{\ell-1}\right)^n$ for all $k \geq \ell$ and $\liminf_{k\to\infty}(\phi_k) \geq \underline{\phi} = \frac{\alpha}{(1-\lambda^\ell)^n} > \alpha$. The solution to the stated problem for such a case was presented in [12]. In this work, we mainly focus our attention on the scenario when ϕ_k is not strictly separated from zero, it can change the sign, and it may present only interval excitation.

3. Linear time-varying estimator

For comparison purposes, let us recall the simplest and most popular solution to the problem mentioned above, the least-squares or gradient estimator:

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k + \gamma_k \phi_k (y_k - \phi_k \widehat{\theta}_k),\ \ k \in \mathbb{N}, \tag{3}$$

where $\gamma_k > 0$ is a time-varying adaptation gain, whose choice impacts the convergence rate and the robustness properties greatly.

Remark 3. For example, for $\gamma_k = \gamma h$ and $\gamma > 0$, the discrete-time equation (3) represents the explicit Euler discretization with step $h > 0$ of the continuous-time gradient descent estimation algorithm:

$$\dot{\widehat{\theta}}(t) = \gamma\phi(t)(y(t) - \phi(t)\widehat{\theta}(t)),\ \ t \in \mathbb{R}_+$$

where all signals keep their meaning, as continuous-time functions. For $\gamma_k = \frac{\gamma h}{1+\gamma h \phi_k^2}$, it corresponds to the implicit Euler discretization of the same continuous-time estimator. In other words, different kinds of discretization can be reduced to the choice of γ_k in the scalar case.

The parameter estimation error behavior is governed by

$$\widetilde{\theta}_{k+1} = (1 - \gamma_k \phi_k^2)\widetilde{\theta}_k - \gamma_k \phi_k v_k,\ k \in \mathbb{N}. \tag{4}$$

Denote

$$\sigma_j = \max_{0 \le i \le j-1} \gamma_i |\phi_i| \ \text{ and } \ \sigma_{\max} = \max_{j \ge 0} \sigma_j.$$

Also, in the sequel, we set:

$$\overline{v} := \max_{j \ge 0} |v_j|,$$

which we assume to be a finite constant.

The following results can be easily obtained (see also [15], whereas the continuous-time counterparts can be found in [16, 5, 9]):

Theorem 1. *For the estimation algorithm* (3) *with* $\gamma_k \in (0, \phi_k^{-2}]$ *for all* $k \in \mathbb{N}$*:*

a) if $\lim_{i \to +\infty} \prod_{j=0}^{i} 1 - \gamma_j \phi_j^2 = 0$ *and* $\lim_{i \to +\infty} \sum_{j=0}^{i} |v_j| < +\infty$*, then the dynamics of parameter estimation error* (4) *is iISS:*

$$|\widetilde{\theta}_k| \le \prod_{j=0}^{k-1} \left(1 - \gamma_j \phi_j^2\right) |\widetilde{\theta}_0| + \sigma_k \sum_{j=0}^{k-1} |v_j|,\ \forall k \in \mathbb{N}^*;$$

b) if there exist $\ell \in \mathbb{N}^*$ *and* $\delta \in (0,1)$ *such that* $\prod_{j=i+1}^{i+\ell} 1 - \gamma_j \phi_j^2 \le \delta$*, for any* $i \in \mathbb{N}$*, then the dynamics of parameter estimation error* (4) *is ISS with exponential convergence rate:*

$$|\widetilde{\theta}_k| \le \delta^{\lfloor \frac{k}{\ell} \rfloor} |\widetilde{\theta}_0| + \sigma_k \ell \frac{1 - \delta^{1 + \lfloor \frac{k-1}{\ell} \rfloor}}{1 - \delta} \max_{0 \le i \le k-1} |v_i|,\ \forall k \in \mathbb{N}^*.$$

Proof. Note that $0 \le 1 - \gamma_k \phi_k^2 \le 1$, for all $k \in \mathbb{N}$, under the imposed restriction on γ_k. Iteratively, using the error dynamics (4), we obtain for any $k \in \mathbb{N}^*$:

$$\widetilde{\theta}_k = \prod_{j=0}^{k-1} \left(1 - \gamma_j \phi_j^2\right) \widetilde{\theta}_0 - \sum_{j=0}^{k-1} \prod_{s=j+1}^{k-1} \left(1 - \gamma_s \phi_s^2\right) \gamma_j \phi_j v_j,$$

hence,

$$|\widetilde{\theta}_k| \le \prod_{j=0}^{k-1} \left(1 - \gamma_j \phi_j^2\right) |\widetilde{\theta}_0| + \sum_{j=0}^{k-1} \prod_{s=j+1}^{k-1} \left(1 - \gamma_s \phi_s^2\right) \gamma_j |\phi_j| |v_j|.$$

Since $\prod_{s=j+1}^{k-1} 1-\gamma_s\phi_s^2 \le 1$ for any $0 \le j \le k-1$, the property in the part *a*) can be derived directly. For the part *b*), note that

$$\prod_{j=0}^{k-1} 1-\gamma_j\phi_j^2 \le \delta^{\left\lfloor \frac{k}{\ell} \right\rfloor}$$

by the imposed restrictions, and

$$\begin{aligned}\sum_{j=0}^{k-1}\prod_{s=j+1}^{k-1}\left(1-\gamma_s\phi_s^2\right)\gamma_j|\phi_j||v_j| &\le \sigma_k \max_{0\le i\le k-1}|v_i| \sum_{j=0}^{k-1}\prod_{s=j+1}^{k-1}\left(1-\gamma_s\phi_s^2\right)\\ &\le \sigma_k \max_{0\le i\le k-1}|v_i| \sum_{j=0}^{k-1}\delta^{\max\left\{0,\left\lfloor\frac{k-j-1}{\ell}\right\rfloor\right\}} \le \sigma_k\ell \max_{0\le i\le k-1}|v_i| \sum_{j=0}^{\left\lfloor\frac{k-1}{\ell}\right\rfloor}\delta^j,\end{aligned}$$

which gives the desired estimate. □

Case *a*) is generic and it can be always ensured by the choice of γ_k provided that $\phi_k \neq 0$ for a non-accumulating subset of indexes $k \in \mathbb{N}$ (recall Remark 2). In case *b*), the asymptotic precision of estimation (asymptotic gain with respect to the measurement noise) is

$$\lim_{k\to+\infty}|\widetilde{\theta}_k| \le \frac{\sigma_{\max}\,\ell}{1-\delta}\,\overline{v},$$

and this scenario follows the PE property.

Remark 4. Indeed, let us demonstrate the last claim. By imposed restrictions, $\gamma_k\phi_k^2 \in (0,1]$ for all $k \in \mathbb{N}$. First, let the PE property be verified for $\sqrt{\gamma_k}\phi_k$ with $\ell \in \mathbb{N}^*$, then

$$\sum_{j=i+1}^{i+\ell}\gamma_j\phi_j^2 \ge \mu > 0$$

for all $i \in \mathbb{N}$. Direct computations show (note that $\ln(1-x) \le -x$, for $x \in [0,1)$):

$$\begin{aligned}0 < \prod_{j=i+1}^{i+\ell} 1-\gamma_j\phi_j^2 &= \mathrm{e}^{\ln\left(\prod_{j=i+1}^{i+\ell}1-\gamma_j\phi_j^2\right)} = \mathrm{e}^{\sum_{j=i+1}^{i+\ell}\ln\left(1-\gamma_j\phi_j^2\right)}\\ &\le \mathrm{e}^{-\sum_{j=i+1}^{i+\ell}\gamma_j\phi_j^2} \le \mathrm{e}^{-\mu} < 1,\end{aligned}$$

which implies the property stated in Theorem 1. Inversely, by the conditions of the theorem, $1 > \delta \ge \prod_{j=i+1}^{i+\ell} 1-\gamma_j\phi_j^2 \ge 0$ for all $i \in \mathbb{N}$. There are two possibilities. Either there is $r_i \in \{1, \dots, \ell\}$ such that $\gamma_{i+r_i}\phi_{i+r_i}^2 > 0.5$, and hence $\sum_{j=i+1}^{i+\ell} \gamma_j \phi_j^2 > 0.5$, or $\gamma_j \phi_j^2 \in [0, 0.5]$, for all $j \in \{i+1, \dots, i+\ell\}$, which implies (note that $\ln(1-x) \geq -2\ln(2)\, x$ for $x \in [0, 0.5]$):

$$0 > \ln(\delta) \geq \ln\left(\prod_{j=i+1}^{i+\ell} 1 - \gamma_j \phi_j^2\right) = \sum_{j=i+1}^{i+\ell} \ln\left(1 - \gamma_j \phi_j^2\right) \geq -2\ln(2) \sum_{j=i+1}^{i+\ell} \gamma_j \phi_j^2.$$

Therefore,

$$\sum_{j=i+1}^{i+\ell} \gamma_j \phi_j^2 \geq \frac{1}{2} \min\left\{1, \frac{\ln(\delta^{-1})}{\ln(2)}\right\},$$

and the claim is proven.

4. Fixed-time convergence

The conventional estimator given in the previous section demonstrates an asymptotic convergence of the parameter estimation error in the noise-free case. This implies a sensitive dependence of the estimation error transients on the initial deviation $|\widetilde{\theta}_0| = |\theta - \widehat{\theta}_0|$ (especially if $|\widetilde{\theta}_0|$ is significant). Due to the scalar nature of (2), this dependence on the initialization of the estimation algorithm (3) can be easily dropped. Indeed, if $\gamma_k = \phi_k^{-2}$ in (3), then

$$\widehat{\theta}_{k+1} = \frac{y_k}{\phi_k}$$

provided that $\phi_k \neq 0$ for some $k \in \mathbb{N}^*$, and obviously:

$$|\widetilde{\theta}_{k+1}| \leq \frac{|v_k|}{|\phi_k|} \leq \frac{\overline{v}}{|\phi_k|},$$

which implies the parameter estimation error $\widetilde{\theta}_{k+1}$ is independent of the initial deviation $\widetilde{\theta}_0$ at the price of a rather large noise gain if ϕ_k is small. This algorithm immediately solves the estimation problem without noise if the system is minimally excited (Remark 1). However, in the noisy scenario, applying such an estimate on each $k \in \mathbb{N}^*$ is not desirable, and it can be used only to cancel the influence of the initial guess $\widehat{\theta}_0$ or when the regressor ϕ_k takes maximal values. Consider $\mathfrak{b}_k$ the worst-case upper bound at each step, and recall that $\overline{v}$ is an upper bound on the noise amplitude. Thus, the following combination with (3) is proposed:

$$\mathfrak{b}_k = \delta^{\left\lfloor \frac{k-\mathfrak{n}_k}{\ell} \right\rfloor} \widetilde{\theta}_k^{\max} + \max_{\mathfrak{n}_k \leq j \leq k} \gamma_j |\phi_j| \; \ell \; \frac{1-\delta^{\left\lfloor \frac{k-\mathfrak{n}_k-1}{\ell} \right\rfloor}}{1-\delta} \overline{v} \tag{5}$$

and

$$\left(\widehat{\theta}_{k+1},\ \mathfrak{n}_{k+1},\ \widetilde{\theta}_{k+1}^{\max}\right) = \begin{cases} \left((1-\gamma_k\phi_k^2)\widehat{\theta}_k + \gamma_k\phi_k y_k,\ \mathfrak{n}_k,\ \widetilde{\theta}_k^{\max}\right), & \text{if } \mathfrak{b}_k \leq \beta\frac{\overline{v}}{|\phi_k|}, \\ \left(\frac{y_k}{\phi_k},\ k,\ \frac{\overline{v}}{|\phi_k|}\right) & \text{otherwise,} \end{cases}$$
$$\mathfrak{n}_0 = 0, \quad \widetilde{\theta}_0^{\max} \geq \widetilde{\theta}_0, \quad \beta > 0. \tag{6}$$

Here $\widetilde{\theta}_0^{\max}$ is the upper bound on initial estimation error, which can be computed if the set of admissible values for θ is known, in other cases, $\widetilde{\theta}_0^{\max} = +\infty$. The states $\mathfrak{n}_k$ and $\widetilde{\theta}_k^{\max}$ represent the number of the last step when a jump occurred and the corresponding jump-induced worst-case upper bound, respectively.

The idea of the estimation algorithm (6) is that, at each step, we compute $\mathfrak{b}_k$ while the estimator (3) is applied on (6) (see Theorem 1) and compare it with the worst-case upper bound if a jump is performed; the comparison threshold is tuned by the sensitivity coefficient β. Then, the next-step estimate $\widehat{\theta}_{k+1}$ is updated accordingly.

Theorem 2. *For the estimation algorithm* (5), (6) *with* $\gamma_k \in (0, \phi_k^{-2}]$, *for all* $k \in \mathbb{N}$ *and* $\widetilde{\theta}_0^{\max} \geq |\theta - \widehat{\theta}_0|$, *assume there exist* $\ell \in \mathbb{N}^*$ *and* $\delta \in (0,1)$ *such that*

$$\prod_{j=i+1}^{i+\ell} 1 - \gamma_j\phi_j^2 \leq \delta,\ \forall i \in \mathbb{N},$$

and $k^* \in \mathbb{N}$ *such that*

$$|\phi_{k^*}| > \delta^{-\left\lfloor \frac{k^*}{\ell} \right\rfloor} \frac{\overline{v}}{\widetilde{\theta}_{k^*}^{\max}}.$$

Then the dynamics of parameter estimation error is ISS with a fixed-time convergence rate.

Proof. Since all conditions of part *b*) of Theorem 1 are satisfied, and the algorithm (3) is replaced in (6) by $\widehat{\theta}_{k+1} = \frac{y_k}{\phi_k}$ only if the direct estimate $\frac{y_k}{\phi_k}$

outperforms (3) (it also corresponds to the substitution $\gamma_k = \phi_k^{-2}$ in (3)), the ISS property follows. The convergence time-dependence on the initial conditions is canceled once there exists $k \in \mathbb{N}$ such that (initially $\mathfrak{n}_0 = 0$)

$$\delta^{\lfloor \frac{k}{\ell} \rfloor} \widetilde{\theta}_k^{\max} + \max_{0 \le j \le k} \gamma_j |\phi_j| \ell \frac{1 - \delta^{1 + \lfloor \frac{k-1}{\ell} \rfloor}}{1 - \delta} \overline{v} > \frac{\overline{v}}{|\phi_k|},$$

i.e., once the value $\widehat{\theta}_{k+1} = \frac{y_k}{\phi_k}$ is assigned. Assume that for all $k \in \mathbb{N}$ the latter inequality is wrong:

$$\delta^{\lfloor \frac{k}{\ell} \rfloor} \widetilde{\theta}_k^{\max} \le \delta^{\lfloor \frac{k}{\ell} \rfloor} \widetilde{\theta}_k^{\max} + \max_{0 \le j \le k} \gamma_j |\phi_j| \ell \frac{1 - \delta^{1 + \lfloor \frac{k-1}{\ell} \rfloor}}{1 - \delta} \overline{v} \le \frac{\overline{v}}{|\phi_k|},$$

which implies that

$$|\phi_k| \le \delta^{-\lfloor \frac{k}{\ell} \rfloor} \frac{\overline{v}}{\widetilde{\theta}_k^{\max}}$$

for all $k \in \mathbb{N}$. The latter property cannot be true for all $k \in \mathbb{N}$ due to the conditions of the theorem, so we have a contradiction. □

If $\theta \in \mathbb{R}$ without any additional precision (*i.e.*, $\widetilde{\theta}_0^{\max} = +\infty$), then definitely applying (6) at the first instant $k \in \mathbb{N}$ such that $\phi_k \neq 0$ significantly improves the estimation performance in the case of a lousy initialization.

A similar extension can be formulated for part a) of the theorem establishing fixed-time iISS:

Theorem 3. *Assume* $\widetilde{\theta}_0^{\max} \ge |\theta - \widehat{\theta}_0|$, $\lim_{i \to +\infty} \prod_{j=0}^{i} 1 - \gamma_j \phi_j^2 = 0$ *and* $\lim_{i \to +\infty} \sum_{j=0}^{i} |v_j| < +\infty$. *Apply the estimation algorithm* (6) *with the worst-case upper bound* $\mathfrak{b}_k$ *computed as*

$$\mathfrak{b}_k = \prod_{j=\mathfrak{n}_k}^{k} \left(1 - \gamma_j \phi_j^2\right) \widetilde{\theta}_k^{\max} + \overline{v} \sum_{j=\mathfrak{n}_k}^{k} \prod_{s=j+1}^{k} \left(1 - \gamma_s \phi_s^2\right) \gamma_j |\phi_j|.$$

Then for $\gamma_k \in (0, \phi_k^{-2}]$, *the dynamics of parameter estimation error is iISS with a fixed-time convergence rate provided that there exists* $k^* \in \mathbb{N}$ *such that* $|\phi_{k^*}| > \frac{\overline{v}}{\widetilde{\theta}_{k^*}^{\max}} \left(\prod_{j=0}^{k^*} 1 - \gamma_j \phi_j^2\right)^{-1}$.

Proof. The proof follows the same lines as before by observing that we can assume without losing generality that $\prod_{j=0}^{k} 1 - \gamma_j \phi_j^2 > 0$ for any $k \in \mathbb{N}$ if the condition

$$\prod_{j=0}^{k}\left(1-\gamma_j\phi_j^2\right)\widetilde{\theta}_k^{\max}+\overline{v}\sum_{j=0}^{k}\prod_{s=j+1}^{k}\left(1-\gamma_s\phi_s^2\right)\gamma_j|\phi_j|>\frac{\overline{v}}{|\phi_k|}$$

is never satisfied. Indeed, $\gamma_k\phi_k^2=1$ implies that $\widehat{\theta}_{k+1}=\frac{y_k}{\phi_k}$, *i.e.*, the fixed-time jump is performed. □

Note that if excitation conditions are satisfied on a finite interval of time only, the the same (qualitatively) stability performance is kept for (5) and (6) once a jump is performed.

5. Nonlinear estimators

As it has been demonstrated in the previous section, a mild modification of any conventional estimation algorithm brings to it the fixed-time convergence rate (as (5) and (6) for (3)), making the error dynamics decay independent of the initial guess, and relaxing the excitation constraints. This shifts further the focus to the problem of noise attenuation and improvement of the respective gains for (3) or (5), (6). This section demonstrates how this can be achieved by introducing the nonlinearities in (3).

To this end, let $\rho:\mathbb{R}\to[0,1]$ be a continuous function such that $\rho(x)=0$ implies that $x=0$.

Example 1. For $\kappa\in(0,1)$, define

$$\rho(x)=\min\left\{|x|^{\kappa|x|},\ |x|^{-\frac{\kappa}{|x|}}\right\}, \tag{7}$$

then the function ρ has the following properties:

- $\rho(x)<1$ for $x\neq 0$ and $|x|\neq 1$;
- $\rho(x)\geq\rho_\kappa=\mathrm{e}^{-\kappa\mathrm{e}^{-1}}$ for all $x\in\mathbb{R}$, and $\lim_{\kappa\to 0}\rho_\kappa=1$;
- $\lim_{x\to 0}\rho(x)=\lim_{|x|\to\infty}\rho(x)=1$.

Example 2. For $a > 0$ and $\kappa > 0$, define

$$\rho(x) = \frac{|x|^\kappa}{a + |x|^\kappa}, \tag{8}$$

then the function ρ has the following properties:

- $\rho(x) < 1$ for $x \in \mathbb{R}$;
- $\lim_{|x|\to\infty} \rho(x) = 1$ and $\rho(0) = 0$.

Extending (3) consider:

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k + \gamma_k \phi_k e_k \rho(e_k),\ k \in \mathbb{N}, \tag{9}$$

for some $\gamma_k > 0$, where a nonlinear multiplicative gain $\rho(e_k) = \rho(\phi_k \widetilde{\theta}_k + v_k) = \rho(y_k - \phi_k \widehat{\theta}_k)$ is introduced. Clearly, this algorithm possesses performance properties similar to (3):

Theorem 4. *For the estimation algorithm* (9) *with* $\gamma_k \in (0, \phi_k^{-2} \frac{1}{\rho(e_k)}]$, *for all* $k \in \mathbb{N}$*:*

a) if $\lim_{i\to+\infty} \prod_{j=0}^{i} 1 - \gamma_j \phi_j^2 \rho(e_j) = 0$ *and* $\lim_{i\to+\infty} \sum_{j=0}^{i} |v_j| < +\infty$*, then the dynamics of parameter estimation error is iISS:*

$$|\widetilde{\theta}_k| \le \prod_{j=0}^{k-1} \left(1 - \gamma_j \phi_j^2 \rho(e_j)\right) |\widetilde{\theta}_0| + \sigma_k \sum_{j=0}^{k-1} |v_j|,\ \forall k \in \mathbb{N}^*;$$

b) if there exist $\ell \in \mathbb{N}^*$ *and* $\delta \in (0,1)$ *such that* $\prod_{j=i}^{i+\ell} 1 - \gamma_j \phi_j^2 \rho(e_j) \le \delta$ *for any* $i \in \mathbb{N}$*, then the dynamics of parameter estimation error is ISS with an exponential convergence rate:*

$$|\widetilde{\theta}_k| \le \delta^{\lfloor \frac{k-1}{\ell} \rfloor} |\widetilde{\theta}_0| + \sigma_k \ell \frac{1 - \delta^{1+\lfloor \frac{k-1}{\ell} \rfloor}}{1-\delta} \max_{0 \le i \le k-1} |v_i|,\ \forall k \in \mathbb{N}^*.$$

Proof. Note that $0 \le 1 - \gamma_k \phi_k^2 \rho(e_k) \le 1$, for all $k \in \mathbb{N}$, under the imposed restriction on γ_k and the properties of ρ. For the parameter estimation error the following equation can be obtained:

$$\widetilde{\theta}_{k+1} = (1 - \gamma_k \phi_k^2 \rho(e_k)) \widetilde{\theta}_k - \gamma_k \phi_k \rho(e_k) v_k,\ k \in \mathbb{N},$$

whose iterative application gives for any $k \in \mathbb{N}^*$:

$$\widetilde{\theta}_k = \prod_{j=0}^{k-1} \left(1 - \gamma_j \phi_j^2 \rho(e_j)\right) \widetilde{\theta}_0 - \sum_{j=0}^{k-1} \prod_{s=j+1}^{k-1} \left(1 - \gamma_s \phi_s^2 \rho(e_s)\right) \gamma_j \phi_j \rho(e_j) v_j,$$

and

$$|\widetilde{\theta}_k| \leq \prod_{j=0}^{k-1} \left(1 - \gamma_j \phi_j^2 \rho(e_j)\right) |\widetilde{\theta}_0| + \sum_{j=0}^{k-1} \prod_{s=j+1}^{k-1} \left(1 - \gamma_s \phi_s^2 \rho(e_s)\right) \gamma_j \rho(e_j) |\phi_j| |v_j|. \quad (10)$$

Since $\prod_{s=j+1}^{k-1} 1 - \gamma_s \phi_s^2 \rho(e_s) \leq 1$ and $\rho(e_j) \leq 1$ for any $j \in [0, k-1]$, the estimates in parts $a)$ and $b)$ can be derived directly as in Theorem 1. □

The obtained upper bounds on $|\widetilde{\theta}_k|$ are the same as in Theorem 1, then two questions arise:

1. Why do we need to introduce additional nonlinearities?

2. The imposed requirements on excitation in Theorem 4 seem to be more restrictive since everything is multiplied by the function ρ, which is smaller or equal to one?

The answers to the first question will be given below by considering particular expressions for ρ, while the last question has a negative answer and it can be reduced to a selection of the adaptation gain γ_k:

Proposition 1. *Let all conditions of Theorem 1 be satisfied, then there exist* $\gamma_k > 0$ *for* $k \in \mathbb{N}$ *such that the conditions of Theorem 4 are also verified.*

Proof. Let $\tilde{\gamma}_k \in (0, \phi_k^{-2}]$ be the adaptation gain for (3) making true the conditions of Theorem 1. Note that if $e_k = 0$, then in both algorithms, in (3) or (9), $\widehat{\theta}_{k+1} = \widehat{\theta}_k$, and these instants of time $k \in \mathbb{N}$ do not influence the transients. Hence, let $e_k \neq 0$, then taking $\gamma_k = \frac{\tilde{\gamma}_k}{\rho(e_k)}$ in (9) is admissible and implies that the conditions of Theorem 4 are true. □

Obviously, the algorithm (9) can substitute (3) in (6) replacing the update equation for $\widehat{\theta}$:

$$\widehat{\theta}_{k+1} = \begin{cases} \Big(1 - \gamma_k \phi_k^2 \rho(y_k - \phi_k \widehat{\theta}_k)\Big)\widehat{\theta}_k & \\ \quad + \gamma_k \phi_k \rho(y_k - \phi_k \widehat{\theta}_k) y_k & \text{if } \mathfrak{b}_k \leq \beta \frac{\overline{v}}{|\phi_k|}, \\ \frac{y_k}{\phi_k} & \text{otherwise,} \end{cases} \tag{11}$$

where all other variables and update equations are as in (5), (6). The estimator (11) allows us to profit the fixed-time convergence rate together with the nonlinear noise attenuation as we discuss in the next subsections.

Remark 5. Note that once $\mathfrak{b}_k > \beta \frac{\overline{v}}{|\phi_k|}$, the estimate $\widehat{\theta}_{k+1}$ is immediately settled in a vicinity of the true value θ whose radius is $\frac{\overline{v}}{|\phi_k|}$, and the further regressor excitation improves the estimation precision via the iterative filtering (9). Therefore, recall again that for this task, the PE property of ϕ_k is not needed, and even IE of the regressor could be enough for short-time filtering.

5.1. The case of Example 1

To formulate the result for this scenario, recall the Excitation Preservation Lemma given in the Appendix and Remark 2:

Corollary 1. *Let ρ be chosen as in* (7) *for* (9)*, $\phi_k \neq 0$, $\sup_{k\geq 0}|\phi_k| < +\infty$ and $\gamma_k = \phi_k^{-2}$ for all $k \in \mathbb{N}$, then*

a) in the case $v_k = 0$ for all $k \in \mathbb{N}$, the parameter estimation error admits a hyperexponential convergence rate at the origin, i.e.*, for any $\alpha \in (0,1)$, there exist $r_\alpha \leq 1$ such that*

$$|\widetilde{\theta}_k| \leq \alpha^k |\widetilde{\theta}_0|$$

for all $k \geq 0$ provided that $|\widetilde{\theta}_0| \leq r_\alpha$;

b) if $|\phi_k| \geq \underline{\phi} > 0$ for all $k \in \mathbb{N}$, then the parameter estimation error is ISS with a hyperexponential convergence rate, i.e.*, in the case $v_k = 0$ for all $k \in \mathbb{N}$, for any $\alpha \in (0,1)$, there exist $r_\alpha \leq 1$ and $R_\alpha \geq 1$ such that*

$$|\widetilde{\theta}_k| \leq \alpha^k |\widetilde{\theta}_0|$$

provided that $|\widetilde{\theta}_0| < r_\alpha$ *and* $k \geq 0$*, or* $|\widetilde{\theta}_0| > R_\alpha$ *and* $k \in \left[0, \frac{\ln\left(\frac{R_\alpha}{|\widetilde{\theta}_0|}\right)}{\ln(\alpha)}\right)$*, respectively. In addition, the asymptotic noise gain admits an estimate:*

$$\lim_{k\to+\infty} |\widetilde{\theta}_k| \leq \frac{1}{\underline{\phi}\rho_\kappa} \max_{i\geq 0} \rho(e_i)|v_i|.$$

Proof. In this case

$$|\widetilde{\theta}_{k+1}| \leq (1-\rho(e_k))|\widetilde{\theta}_k| + |\phi_k|^{-1}\rho(e_k)|v_k|$$

for all $k \in \mathbb{N}$. Since $\rho(e_k) \geq \rho_\kappa \in (0,1]$ for all $e_k \in \mathbb{R}$ by construction,

$$|\widetilde{\theta}_{k+1}| \leq (1-\rho_\kappa)|\widetilde{\theta}_k| + |\phi_k|^{-1}\rho(e_k)|v_k|$$

and the error $\widetilde{\theta}_k$ has an exponential convergence rate ρ_κ in the noise-free case (the same for the regression error e_k since ϕ_k has a bounded amplitude). Moreover, $\rho(e_k)$ tends to 1 as e_k tends to zero, and for any $\alpha \in (0,1)$, there exists T_1 such that $1-\rho(e_k) \leq \frac{\alpha}{2}$ for all $k \geq T_1$. Then for $k \geq T_1$ it holds

$$|\widetilde{\theta}_k| \leq \left(\frac{\alpha}{2}\right)^{k-T_1} (1-\rho_\kappa)^{T_1} |\widetilde{\theta}_0|.$$

Choose T such that

$$2^{T-T_1} > \left(\frac{1-\rho_\kappa}{\alpha}\right)^{T_1},$$

then for all $k \geq T$

$$|\widetilde{\theta}_k| \leq \left(\frac{\alpha}{2}\right)^{k-T_1} (1-\rho_\kappa)^{T_1} |\widetilde{\theta}_0| \leq \alpha^{k-T_1}\alpha^{T_1}|\widetilde{\theta}_0| = \alpha^k|\widetilde{\theta}_0|.$$

So, part a) is substantiated.

For part b), if the signal $|\phi_k|$ is lower bounded, then

$$|\widetilde{\theta}_{k+1}| \leq (1-\rho_\kappa)|\widetilde{\theta}_k| + \underline{\phi}^{-1}\rho(e_k)|v_k|,$$

for all $k \in \mathbb{N}$, and ISS follows since $\rho(e_k) \leq 1$. The hyperexponential convergence rate close to the origin can be substantiated as for part a), and it is required to show hyperexponential convergence in the noise-free case for big enough deviations of $|\widetilde{\theta}_k|$. Note that for any $\alpha \in (0,1)$, there exists $\overline{e}_\alpha > 0$ such that $1-\rho(e_k) \leq \alpha$, for all $|e_k| \geq \overline{e}_\alpha$, hence, $|\widetilde{\theta}_k| \leq \alpha^k|\widetilde{\theta}_0|$ while $|\widetilde{\theta}_k| \geq \underline{\phi}^{-1}\overline{e}_\alpha = R_\alpha$, *i.e.*, for $k \in \left[0, \frac{\ln\left(\frac{R_\alpha}{|\widetilde{\theta}_0|}\right)}{\ln(\alpha)}\right)$. The estimate of the asymptotic gain with respect to the noise v_k can be derived directly. □

It is worth highlighting that for part a) there is no condition imposed on the level of excitation of ϕ, and PE is imposed for case b) only. Nevertheless, for part b), the noise amplitude is filtered by ρ:

$$\rho(e_k)|v_k| \leq |v_k|$$

for all $k \in \mathbb{N}$, and this inequality becomes non-strict for $|e_k| \in \{0, 1\}$ only.

Remark 6. Due to the division by ϕ_k in γ_k, for $e_k = 1$, we have the algebraic one-shot estimation $\widehat{\theta}_{k+1} = \widehat{\theta}_k - \widetilde{\theta}_k = \theta$ in the noise-free case, similarly to (5), (6) or (11).

5.2. The case of Example 2

In this case, we can use the fact that the function ρ is monotonously growing in $|e_k|$ (in other words, $\rho(e_k)$ weights the adaptation gain γ_k proportionally to the regression error e_k).

Corollary 2. *Let ρ be chosen as in* (8) *for* (9) *and the conditions of Theorem 4 be valid, then the parameter estimation error admits the asymptotic gain upper bounds:*

$$\lim_{k\to+\infty} |\widetilde{\theta}_k| \leq \sigma_{\max}\ \rho\left([\sigma_{\max} \max_{i\geq 0} |\phi_i| + 1] \sum_{j=0}^{+\infty} |v_j|\right) \sum_{j=0}^{+\infty} |v_j|$$

or

$$\lim_{k\to+\infty} |\widetilde{\theta}_k| \leq \frac{\sigma_{\max}\ell}{1-\delta}\ \rho\left(\left(\frac{\sigma_{\max}\ell}{1-\delta} \max_{i\geq 0} |\phi_i| + 1\right) \max_{i\geq 0} |v_i|\right) \max_{i\geq 0} |v_i|$$

for parts a) or b), respectively.

Proof. For part a), note that the estimate obtained in the proof of Theorem 4 implies that

$$\lim_{k\to+\infty} |\widetilde{\theta}_k| \leq \sigma_{\max} \sum_{j=0}^{+\infty} |v_j|,$$

and

$$\rho(e_k) = \rho\left(\phi_k \widetilde{\theta}_k + v_k\right) \leq \rho\left(|\phi_k||\widetilde{\theta}_k| + |v_k|\right)$$

due to the shape of ρ, then from (10) we obtain:

$$\lim_{k\to+\infty} |\widetilde{\theta}_k| \leq \sigma_{\max} \sum_{j=0}^{+\infty} \rho\left(e_j\right)|v_j| \leq \sigma_{\max}\rho\left(\left(\sigma_{\max}\max_{i\geq 0}|\phi_i| + 1\right)\sum_{j=0}^{+\infty}|v_j|\right)\sum_{j=0}^{+\infty}|v_j|.$$

For part b), note that

$$\lim_{k\to+\infty} |\widetilde{\theta}_k| \leq \frac{\sigma_{\max}\ell}{1-\delta}\max_{i\geq 0}|v_i|$$

by the imposed restrictions, and

$$\sum_{j=0}^{k-1}\prod_{s=j+1}^{k-1}\left(1-\gamma_s\phi_s^2\rho(e_s)\right)\gamma_j\rho(e_j)|\phi_j||v_j| \leq \sigma_k \max_{0\leq i\leq k-1}\rho(e_i)|v_i|\sum_{j=0}^{k-1}\prod_{s=j+1}^{k-1}\left(1-\gamma_s\phi_s^2\rho(e_s)\right)$$
$$\leq \sigma_k \max_{0\leq i\leq k-1}\rho(e_i)|v_i|\sum_{j=0}^{k-1}\delta^{\max\{0,\left\lfloor\frac{k-j-1}{\ell}\right\rfloor\}} \leq \sigma_k\ell \max_{0\leq i\leq k-1}\rho(e_i)|v_i|\sum_{j=0}^{\left\lfloor\frac{k-1}{\ell}\right\rfloor}\delta^j,$$

which gives the desired estimate. □

The appearance of ρ in the worst-case asymptotic gain estimates implies that these gains for (9) are smaller than for (3).

5.3. Choice of adaptation gain

Since for the stability of (9), the condition $\gamma_k\phi_k^2\rho(e_k) \leq 1$ has to be satisfied for any $k \in \mathbb{N}$ ($\gamma_k\phi_k^2 \leq 1$ for (3)), a popular and logical selection is

$$\gamma_k = \frac{|\phi_k|^{\eta-2}}{b + \max_{0\leq i\leq k}|\phi_i|^\eta} \tag{12}$$

for some $b \geq 0$ and $\eta \geq 2$. For $\eta > 2$, the adaptation gain becomes proportional to the amplitude of ϕ_k and $\rho(e_k)$, *i.e.*, it becomes small if the regressor ϕ_k or the regression error e_k are small (in such a case, it makes no sense to make an estimation step due to a high noise sensitivity or established convergence, respectively).

6. Simulations

To illustrate the proposed algorithms, we consider the offset estimation problem for the measured signal

$$\overline{y}_k = A + B\sin(\omega_1 k + \psi)\cos(\omega_2 k) + \bar{v}_k, \tag{13}$$

where A, B, and ψ are unknown constant parameters, the frequencies $\omega_1 = \frac{2\pi}{14}$ and $\omega_2 = \frac{\pi}{21}$ are known, and $\bar{v}_k$ is the bounded measurement noise, $|\bar{v}_k| \leq 1$ for all $k \in \mathbb{N}$. The goal is to estimate the offset A. Equation (13) can be written in the form (1) choosing

$$\begin{aligned} \overline{\phi}_k &= \begin{bmatrix} 1 & \sin(\omega_1 k)\cos(\omega_2 k) & \cos(\omega_1 k)\cos(\omega_2 k) \end{bmatrix}^\top, \\ \overline{\theta} &= \begin{bmatrix} A & B\cos(\psi) & B\sin(\psi) \end{bmatrix}^\top. \end{aligned} \tag{14}$$

Applying the DREM procedure with Kreisselmeier's regressor extension (16), (17) with $\lambda = 0.966$, which corresponds to $\lambda^{20} \approx \frac{1}{2}$, we get the scalar linear regression in the form (2), where y_k is the first element of the extended vector Y_k defined in (17), ϕ_k is defined in (16), and θ is the first element of $\overline{\theta}$. Note that the DREM procedures yield a single scalar equation for the parameter of interest, namely $\theta = A$, and does not require estimating other elements of the vector $\overline{\theta}$.

It is straightforward to verify that $\overline{\phi}$ defined in (14) is (ℓ, μ)-PE in the sense of Definition 1 (given in Appendix) for $\ell = 42$ and $\mu = 10$, *i.e.*, for all $k \in \mathbb{N}$

$$\sum_{i=0}^{41} \overline{\phi}_{k+i}\overline{\phi}_{k+i}^\top \geq 10 I_3.$$

Applying the Excitation Preservation Lemma (see Appendix), we conclude that the new scalar regressor ϕ is also PE, and it is positive for all $k \geq \ell$. Thus for $\gamma_k \in (0, \phi_k^{-2}]$, it holds $\prod_{i=0}^{\ell-1}\left(1 - \gamma_{k+i}\phi_{k+i}^2\right) \leq \nu$ for some $\nu \in (0,1)$. The resulting signal ϕ_k is depicted in Fig. 1.

As a baseline for the following estimators' comparison, we consider the standard gradient estimator (3), where the gain γ_k is computed as in (12), with

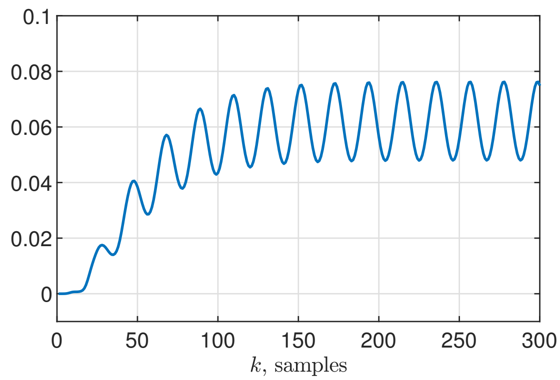

Figure 1: The new regressor ϕ_k computed as (16), (17) for the example (13), (14).

$b = 0.1$ and $\eta = \frac{5}{2}$. For all estimators, the initial value is chosen as $\hat{\theta}_0 = 0$. The true values of the parameters are $A = 5$, $B = 7$, and $\psi = \frac{\pi}{3}$.

First, we compare the baseline estimator with the fixed-time estimator (5), (6) satisfying the conditions of Theorem 2. As for the baseline estimator, the gain γ_k is computed as (12). We also assume that $|\theta| < 15$ yields for $\hat{\theta}_0 = 0$ the initial value $\widetilde{\theta}_0^{\max} = 15$. Finally, for the considered example, we have $\nu = 0.88$. The baseline estimator and the fixed-time estimator are compared in Fig. 2 and Fig. 3 for $\beta = 1$ and $\beta = 2$, respectively. The figures illustrate the superior performance of the fixed-time estimator. As soon as the signal ϕ becomes separated from zero, see Fig. 1, the estimator performs several jumps bringing the estimation error $\widetilde{\theta}$ to a neighborhood of zero. Then, the value of β is used to compare two worst-case error bounds as defined in (6). If not, a further jump can improve the worst-case bound, as in Fig. 3. Then the fixed-time estimator behaves as the baseline gradient estimator.

However, the fixed-time estimator requires the *a priori* information about the regressor ϕ, namely the excitation characteristics ν and ℓ. In the context of the DREM procedure, it is guaranteed that the regressor ϕ is strictly separated from zero after the first ℓ samples, and thus this drawback can be overcome by the nonlinear hyperexponential estimator (7), (9), where the gain γ_k is chosen following Corollary 1. For simulations, we set $\kappa = \frac{1}{2}$ in (7), and the comparison

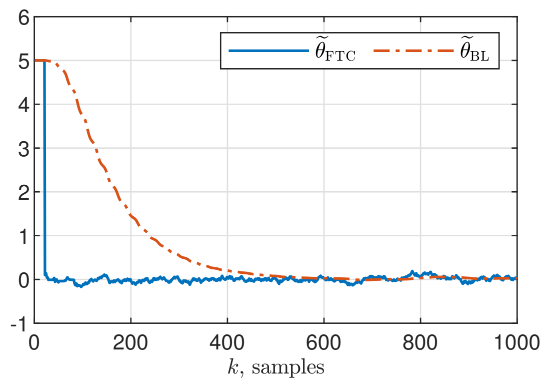

Figure 2: The estimation error $\widetilde{\theta}_{\mathrm{BL}}$ for the baseline estimator (3) and the estimation error $\widetilde{\theta}_{\mathrm{FTC}}$ for the fixed-time estimator (5), (6) with $\beta = 1$.

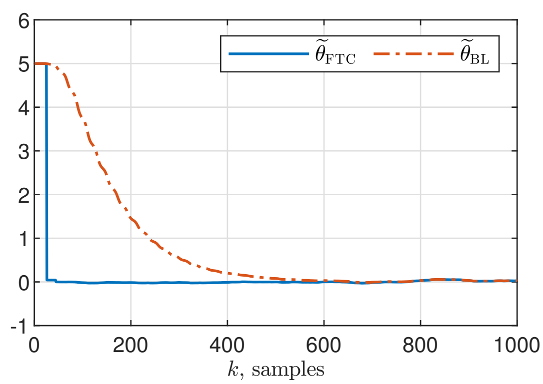

Figure 3: The estimation error $\widetilde{\theta}_{\mathrm{BL}}$ for the baseline estimator (3) and the estimation error $\widetilde{\theta}_{\mathrm{FTC}}$ for the fixed-time estimator (5), (6) with $\beta = 2$.

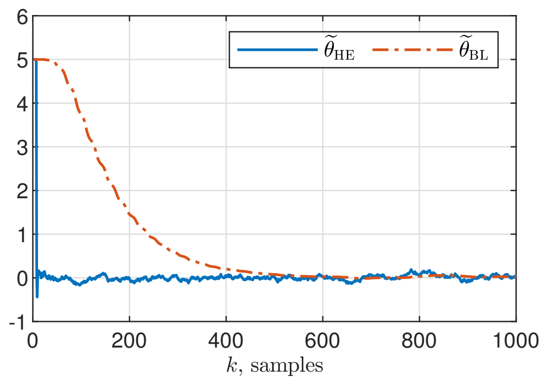

Figure 4: The estimation error $\widetilde{\theta}_{\mathrm{BL}}$ for the baseline estimator (3) and the estimation error $\widetilde{\theta}_{\mathrm{HE}}$ for the hyperexponential estimator (7), (9).

Estimator	BL (3)	FTC (5), (6) with $\beta = 2$	FTC (5), (6) with $\beta = 1$	HE (7), (9)	NL (8), (9)
MAE$\cdot 10^2$	2.8	2.8	5.9	5.9	1.6

Table 1: The steady-state MAE values for the considered estimators.

with the baseline estimator is given in Fig. 4. The hyperexponential estimator establishes fast convergence without requiring any information about the excitation characteristics of the regressor ϕ. However, the drawback is the slightly increased noise sensitivity observed in the steady-state oscillations.

Finally, we compare the baseline estimator with the nonlinear estimator (8), (9), where we set $\kappa = \frac{3}{4}$ and $a = 0.1$, and the gain γ_k is chosen for the baseline estimator following (12). The comparison is given in Fig. 5 and illustrates improved noise sensitivity in the steady-state, where the trade-off is the increased transient time. To illustrate the noise sensitivity, Table 1 summarizes the truncated mean absolute error (MAE) values computed in the steady-state over 30000 samples,

$$\mathrm{MAE}(\widetilde{\theta}) = \frac{1}{30000} \sum_{i=1001}^{31000} |\widetilde{\theta}_i|.$$

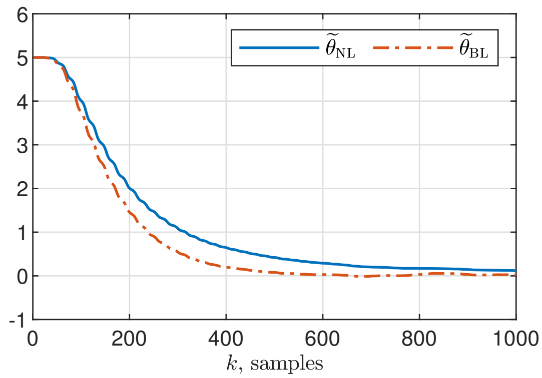

Figure 5: The estimation error $\widetilde{\theta}_{\mathrm{BL}}$ for the baseline estimator (3) and the estimation error $\widetilde{\theta}_{\mathrm{NL}}$ for the nonlinear estimator (8), (9).

7. Conclusion

For linear regression, this study developed numerous DREM-based parameter estimators of the form (11). are proposed for linear regression. These algorithms exhibited a fixed-time rate of convergence (uniform in the initial data) with enhanced noise gains (when compared to the linear estimator (3)). Parts a) of theorems 1, 4, corollaries 1, 2, and Theorem 3 assume a relaxed excitation of the regressor. Other parts of these results formulate additional constructive outcomes for the PE case. Since fixed-time convergence is ensured, the condition of PE of the regressor can be relaxed to IE. Furthermore, the iterative application of the estimation algorithms resulted in further noise filtering.

References

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

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

[3] S. Sastry, M. Bodson, Adaptive Control: Stability, Convergence and Robustness, Prentice-Hall, London, 1989.

[4] P. Ioannou, P. Kokotovic, Adaptive Systems with Reduced Models, Springer Verlag, Secaucus, NJ, 1983.

[5] S. Aranovskiy, A. Bobtsov, R. Ortega, A. Pyrkin, Performance enhancement of parameter estimators via dynamic regressor extension and mixing, IEEE Transactions on Automatic Control 62 (7) (2017) 3546–3550. `doi:10.1109/TAC.2016.2614889`.

[6] A. Belov, S. Aranovskiy, R. Ortega, N. Barabanov, A. Bobtsov, Enhanced parameter convergence for linear systems identification: The DREM approach, in: Proc. European Control Conference (ECC), Limassol, 2018, pp. 2794–2799.

[7] S. Aranovskiy, A. Belov, R. Ortega, N. Barabanov, A. Bobtsov, Parameter identification of linear time-invariant systems using dynamic regressor extension and mixing, International Journal of Adaptive Control and Signal Processing 33 (6) (2019) 1016–1030.

[8] A. Vediakova, A. Vedyakov, A. Pyrkin, A. Bobtsov, V. Gromov, Finite time frequency estimation for multi-sinusoidal signals, European Journal of Control 59 (2021) 38–46.

[9] J. Wang, D. Efimov, A. Bobtsov, On robust parameter estimation in finite-time without persistence of excitation, IEEE Transactions on Automatic Control 65 (4) (2020).

[10] R. Ortega, V. Nikiforov, D. Gerasimov, On modified parameter estimators for identification and adaptive control. a unified framework and some new schemes, Annual Reviews in Control 50 (2020) 278–293. `doi:https://doi.org/10.1016/j.arcontrol.2020.06.002`.

[11] J. Wang, D. Efimov, S. Aranovskiy, A. Bobtsov, Fixed-time estimation of parameters for non-persistent excitation, European Journal of Control (2019).

[12] M. Korotina, S. Aranovskiy, R. Ushirobira, D. Efimov, J. Wang, A note on fixed- and discrete-time estimation via the DREM method, AutomaticaSubmitted (2022).

[13] D. Angeli, Intrinsic robustness of global asymptotic stability, Systems & Control Letters 38 (4) (1999) 297–307. doi:https://doi.org/10.1016/S0167-6911(99)00077-8.

[14] Z.-P. Jiang, Y. Wang, Input-to-state stability for discrete-time nonlinear systems, Automatica 37 (6) (2001) 857–869. doi:https://doi.org/10.1016/S0005-1098(01)00028-0.

[15] L. Praly, Convergence of the gradient algorithm for linear regression models in the continuous and discrete time cases, int. Rep. MINES ParisTech, Centre Automatique et Systèmes (December 26 2016).
URL https://hal.archives-ouvertes.fr/hal-01423048

[16] A. Morgan, K. Narendra, On the uniform asymptotic stability of certain linear nonautonomous differential equations, SIAM J. Control and Optimization 15 (1) (1977) 5–24.

[17] K. S. Narendra, A. M. Annaswamy, Persistent excitation in adaptive systems, International Journal of Control 45 (1) (1987) 127–160. doi:10.1080/00207178708933715.

Appendix. DREM with Kreisselmeier's regressor extension

For the linear regression (1), the first *extension* step of the DREM procedure [5] 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. Denote

$$Y_k^i = H_i(z)\overline{y}_k,\ \Phi_k^i = H_i(z)\overline{\phi}_k,\ V_k^i = H_i(z)\overline{v}_k,\ i \in \{1, \dots, n\}, \tag{15}$$

$$Y_k = (Y_k^1 \dots Y_k^n)^\top,\ V_k = (V_k^1 \dots V_k^n)^\top \in \mathbb{R}^n,\ \Phi_k = (\Phi_k^1 \dots \Phi_k^n)^\top \in \mathbb{R}^{n \times n},$$

which leads to a new extended regression

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

where $\epsilon_k \in \mathbb{R}^n$ is an exponentially decaying term coming from the initialization of the filters (the filters can be applied directly to (1), or to the auxiliary regression problem $\overline{\phi}_k \overline{y}_k = \overline{\phi}_k \overline{\phi}_k^\top \overline{\theta} + \overline{\phi}_k \overline{v}_k$). Finally, realizing the second *mixing* step, for

$$\phi_k := \det(\Phi_k)\,,\ y_k := \mathrm{adj}(\Phi_k)\, Y_k,\ v_k = \mathrm{adj}(\Phi_k)\, \overline{V}_k \tag{16}$$

the method leads to the element-wise scalar linear regression:

$$y_{k,i} = \phi_k \theta_i + v_{k,i},\ k \in \mathbb{N},\ i \in \{1, \ldots, n\},$$

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 $\phi_k \in \mathbb{R}$.

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 choice of the filters H_i in (15). Let us recall a solution to this problem. To this end, let us characterize the admissible excitation levels [17]:

Definition 1. [PE] A bounded signal $\overline{\phi} : \mathbb{N} \to \mathbb{R}^n$ is called (ℓ, μ)-*persistently excited* (PE) if there exit $\ell \in \mathbb{N}^*$ 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 *interval excited* (IE) if the above inequality is satisfied for $k = 0$ only.

Within the context of the DREM procedure, Kreisselmeier's dynamic regressor extension for (15) applied to the auxiliary regression problem is given by

$$\begin{aligned} \Phi_{k+1} &= \lambda \Phi_k + (1 - \lambda)\, \overline{\phi}_k \overline{\phi}_k^\top\,, \\ Y_{k+1} &= \lambda Y_k + (1 - \lambda)\, \overline{\phi}_k \overline{y}_k, \end{aligned} \tag{17}$$

where $\lambda \in (0,1)$ is the low-pass filter tuning parameter, $Y_0 \in \mathbb{R}^n$ and $\Phi_0 \in \mathbb{R}^{n\times n}$, $\Phi_0 = \Phi_0^\top \geq 0$, and the choice $\Phi_0 = 0$, $Y_0 = 0$ yields $\epsilon_k \equiv 0$ for all $k \in \mathbb{N}$. The following result explains the advantages of the filter (17):

Lemma (Excitation Preservation)**.** *[12] Consider a (ℓ, μ)-PE signal $\overline{\phi}_k$ in* (1) *and the Kreisselmeier's dynamics extension* (17)*, then the signal ϕ_k defined in* (16) *is (ℓ, α)-PE, where*

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

Moreover,

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