2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

On DREM regularization and unexcited linear regression estimation

Stanislav Aranovskiy[1], Rosane Ushirobira[2], and Denis Efimov[2]

***Abstract*— The problem of estimating unknown constant parameters in linear regression with measurement noise is considered in this paper. To analyze different levels of excitation of the regressor, two notions of partial and feeble excitation are introduced. The former implies the absence of persistent or interval excitation, while the latter property states that the excitation is insufficient for an efficient estimation in a noisy setting. The dynamic extension and mixing method (DREM) is used for the problem solution, and to improve its estimation performance, regularization is proposed, and the resulting improvement is investigated analytically. The theoretical findings are illustrated in the simulations.**

I. Introduction

We consider the linear regression equation (LRE)

$$y(t) = \phi^\top(t)\theta + v(t),\ t \in \mathbb{R}_+ \tag{1}$$

where $y(t) \in \mathbb{R}^\ell$ is the output signal, $\phi(t) \in \mathbb{R}^{n\times\ell}$ is the regressor, $v(t) \in \mathbb{R}^\ell$ is an additive distortion, *e.g.*, a measurement noise, and $\theta \in \mathbb{R}^n$ is the vector of unknown constant parameters. The signals y and ϕ are available, and the goal is to estimate the vector of parameters θ. We assume that the regressor ϕ and the distortion v are bounded and that the regressor ϕ is piecewise continuous.

The crucial property defining whether the vector of parameters θ can be estimated (uniformly in time) is the excitation of the regressor ϕ. The common types of excitation are persistent and interval ones, as given in Section II.

The classic result in adaptive control and parameter estimation states that for ϕ persistently exciting, the vector θ in LRE (1) can be estimated exponentially fast using, for example, least squares algorithms, and the estimation error is input-to-state stable with respect to v; see [1], [2]. In contrast, the interval excitation is not uniform in time, precluding noise filtering and robustness. However, it was shown in [3]–[5] that the interval excitation is, in fact, sufficient for estimation being the identifiability condition. If the regressor ϕ is not exciting over any interval in the sense of Definition 1 given in Section II, then there exist θ_a, $\theta_b \in \mathbb{R}^n$, $\theta_a \neq \theta_b$, such that $\phi(t)\theta_a = \phi(t)\theta_b$, for all t, and the vector θ in (1) cannot be reconstructed from the measurements of y and ϕ, even in the absence of v. On the other hand, if the regressor ϕ is exciting on an interval, the vector θ can be estimated in the absence of noise, e.g., using finite-/fixed-time estimators [4] or concurrent/composite learning [6], [7] (where the interval excitation is also known as *sufficient* excitation [3]).

Besides the compromised identifiability of θ, the lack of excitation of ϕ is also connected to the numerical implementation and tuning of estimation algorithms. For example, concurrent learning estimators use a historical data stack that can be seen as an accumulation of samples $\phi(t_k)\phi^\top(t_k) \in \mathbb{R}^{n\times n}$ at time instances t_k, $k \in \mathbb{N}$. This data stack is meant to keep the past information to be used with new measurements of ϕ, allowing for parameter estimation under interval excitation. Suppose that ϕ is interval excited as in Defintion 1, but the corresponding value μ is small regarding the magnitudes of signals and chosen numerical accuracy. Then, the accumulated data matrix may be ill-conditioned, requiring high gains and making the estimation prone to numerical errors.

A similar situation may arise in the Dynamic Regressor Extension and Mixing (DREM) estimation [8]. The *extension* step of this procedure transforms the LRE (1) into a novel extended LRE with a square regressor matrix $\Phi(t) \in \mathbb{R}^{n\times n}$ whose adjugate matrix is further used to decouple the vector problem (1) into a set of scalar LRE for each element of θ independently. If the original regressor ϕ is PE but with a relatively small value μ, then the extended matrix $\Phi(t)$ may be close to singular, complicating the practical implementation of a DREM estimator.

Nevertheless, even if the regressor ϕ is neither PE nor IE or is PE/IE with a small μ value, the regressor may still contain certain information, being exciting in *specific directions*. This concept is used in least-squares estimation with regressor projection [9] and directional forgetting [10], where the covariance matrix is updated only in those directions where the regressor ϕ contains new information.

Recently, a modification of the standard gradient and least-squares algorithms was proposed in [11], allowing for exponential estimation of a particular projection of θ in (1) under the lack of PE; the authors also propose a definition of order of PE lack. A similar problem was addressed in [12] in the context of the DREM procedure. The authors proposed a matrix update algorithm in the vein of directional forgetting and introduced a definition of semi-persistent excitation.

Novelty and Contribution. This research is motivated by these recent advances. We consider the LRE (1) under a deficiency of excitation of ϕ, and propose two notions quantifying the lack of persistent excitation as discussed above. Next, we study several regularization tools that allow us to improve the estimation accuracy for the vector of constant unknown parameters θ. Admissible bounds on the regularization matrix are evaluated.

The rest of the paper is organized as follows. Definitions of different excitation levels for a regressor are given in Section II, together with two new concepts of partial and feeble excitation. The problem statement is presented in Section III. Regularization for DREM is introduced and analyzed in

[1]CentraleSupélec–IETR, 35576 Cesson-Sévigné, France.
[2]Inria, Univ. Lille, CNRS, UMR 9189 - CRIStAL, F-59000 Lille, France.

Section IV. The results of computer experiments illustrating the efficiency of this regularization are shown in Section V.

Notation

- The sets of nonnegative real and nonnegative 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 $x \in \mathbb{R}^n$, $\|x\|$ denotes its Euclidean norm, and $\|A\|$ corresponds to the induced norm for $A \in \mathbb{R}^{n\times n}$.

II. Preliminaries

Classical definitions of persistent and interval excitation are recalled below:

Definition 1: Let $\phi : \mathbb{R}_+ \to \mathbb{R}^{n\times \ell}$ be a bounded signal.

(i) We call ϕ *persistently exciting* if there exist $T > 0$ and $\mu > 0$ such that for all $t \in \mathbb{R}_+$,

$$\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau \geq \mu I_n. \tag{2}$$

We write ϕ is PE or (T, μ)-PE to mention specific values of T and μ.

(ii) We call ϕ *interval exciting* if it is exciting over an interval (so (2) is satisfied only for a specific value of t, *e.g.*, $t = 0$). We say ϕ is IE. □

A. What is lack of excitation?

In contrast with persistent or interval excitation, a lack of excitation is a less common concept. It is often considered merely the absence of the PE/IE properties. A more sophisticated interpretation was suggested in the recent work [11], where the authors define the lack of persistence of excitation of order p as follows.

Definition 2 (Lack of persistence of excitation, [11]): For $0 \leq p \leq n$, a piecewise continuous uniformly bounded matrix function $\phi : \mathbb{R}_+ \to \mathbb{R}^{n\times \ell}$ has a *lack of persistency of excitation of order* p, if there exist $T > 0$, $k_T > 0$ and linearly independent orthogonal, unitary norm vectors $v_i \in \mathbb{R}^n$, $1 \leq i \leq n$, such that for all $t \in \mathbb{R}_+$,

$$v_i^\top \left(\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau \right) v_i = 0,\ 1 \leq i \leq p \tag{3}$$

and

$$v_i^\top \left(\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau \right) v_i \geq k_T,\ p+1 \leq i \leq n. \quad \square$$

A few remarks regarding Defintion 2 are given below.

1) An example of vectors v_i are the orthonormal eigenvectors of the symmetric $n\times n$ matrix $\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau$. The first p vectors correspond to zero eigenvalues.
2) Definition 2 operates with constant vectors v_i, meaning that the lack of excitation has *constant nature*; the kernel of the matrix $\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau$ is time-invariant.

For this paper, we consider the lack of excitation from another point of view by introducing the following definition of *partial excitation*.

Definition 3 (Partial excitation): A bounded signal $\phi : \mathbb{R}_+ \to \mathbb{R}^{n\times \ell}$ is *partially persistently (interval) exciting* of degree q, if there exist constant matrices $C \in \mathbb{R}^{n\times(n-q)}$ and $Z \in \mathbb{R}^{n\times q}$ such that

- $\operatorname{rank}\left(\begin{bmatrix} C & Z \end{bmatrix}\right) = n$,
- (degeneracy of regressor)

$$\phi^\top(t)C = 0,\ \forall t \in \mathbb{R}_+, \tag{4}$$

- the signal $\tilde{\phi} := \left(\phi^\top Z\right)^\top$ is persistently (interval) exciting. □

Definition 3 is more general than the definition of lack of persistent excitation of degree $p = n - q$ in [11]. It also includes the case of interval excitation and does not impose a particular structure of the matrices C and Z. However, it is easy to see that partial persistent excitation of degree q is equivalent to a lack of persistent excitation of degree $n - q$ for a special choice of C and Z.

Consider now the LRE (1), where the regressor ϕ is only partially exciting of degree q. Due to the invertibility of the matrix $\begin{bmatrix} C & Z \end{bmatrix}$, there exist $\tilde{\theta}_1 \in \mathbb{R}^{n-q}$ and $\tilde{\theta}_2 \in \mathbb{R}^q$ providing

$$\theta = C\tilde{\theta}_1 + Z\tilde{\theta}_2. \tag{5}$$

The component $C\tilde{\theta}_1$ of the vector θ is orthogonal to the regressor ϕ and cannot be estimated from (1). Thus, only the vector $\tilde{\theta}_2$, excited by $\tilde{\phi}$, can be reconstructed. To this end, substituting (5) to (1) yields the reduced-order LRE

$$y(t) = \tilde{\phi}^\top(t)\tilde{\theta}_2 + v(t) \tag{6}$$

allowing the estimation of $\tilde{\theta}_2$ from y and $\tilde{\phi}$ using any existing parameter estimation techniques.

Remark 1: The matrices C and Z in Definition 3 are not unique and may be written on a different basis. For instance, for a pair C, Z satisfying the conditions, for any invertible matrix R, the pair CR, ZR also fulfills the requirements. Thus, $\tilde{\theta}_2$ in (5) is defined by a specific choice of Z. Moreover, without loss of generality, we may assume that Z is orthogonal to C. Otherwise, the matrix Z can be written as $Z = CA + C^\perp B$, where $C^\perp \in \mathbb{R}^{n\times q}$ is a full-rank matrix in the orthogonal complement of C, and $A \in \mathbb{R}^{(n-q)\times q}$, $B \in \mathbb{R}^{q\times q}$ are any matrices. Then $\tilde{\phi}^\top(t) = \phi^\top(t)Z = \phi(t)C^\perp B$, i.e., only the projection of Z on the subspace orthogonal to C affects the measurements $y(t)$ and the reduced-order regressor $\tilde{\phi}(t)$.

B. Feeble Excitation

Both Definitions 2 and 3 operate with the exact equality to zero of the products in (3) and (4). Recalling that the LRE (1) contains the measurement noise v, from a practical point of view, it is beneficial to consider the cases when the regressor ϕ is formally exciting, but the level of excitation is feeble with respect to the noise magnitude or numeric precision of the system. We formulate feeble excitation as follows.

Definition 4: Let $\phi : \mathbb{R}_+ \to \mathbb{R}^{n\times \ell}$ be (T, μ)-PE (or IE) in the sense of Definition 1. Let $\lambda_1(t) \leq \lambda_2(t) \leq \ldots \leq \lambda_n(t)$ be the eigenvalues of the matrix $\int_t^{t+T} \phi(s)\phi^\top(s)ds$. We say that the excitation is *feeble of order* p if there exist $\mu_1, \mu_2 \in$

$\mathbb{R}_+$, $\mu_1 \ll \mu_2$, such that $\lambda_p(t) \leq \mu_1$, $\lambda_{p+1}(t) \geq \mu_2$, for all $t \in \mathbb{R}_+$ (for a specific $t \in \mathbb{R}_+$). □

Definition 4 is related to the ill-conditioning of the estimation problem. For instance, if the integral is taken on a larger time domain from t to $t+kT$ for $k \in \mathbb{N}^*$ to accumulate more information, the resulting matrix is not well-conditioned (it may also result in the augmentation of the noise influence if ϕ and v are interrelated). From a practical point of view, we are also interested in the case when μ_1 is of the order of magnitude of the noise v impacting the estimation accuracy.

C. Dynamic Regressor Extension and Mixing (DREM)

In a nutshell, the DREM procedure is a nonlinear dynamic transformation applied to the LRE (1) and transforming it into a set of n scalar LREs for each element of the vector θ. Thus, the elements of θ can be estimated independently, enhancing the transient performance; see [8] for more details.

The DREM procedure consists of the *extension* and *mixing* steps. First, a dynamic transformation is applied, leading to a new extended LRE

$$Y(t) = \Phi(t)\theta + V(t),\ V(t) = \overline{V}(t) + \epsilon(t),\ t \in \mathbb{R}_+, \quad (7)$$

where $Y(t) \in \mathbb{R}^n$ and $\Phi(t) \in \mathbb{R}^{n\times n}$ are measured signals generated by the extension of the dynamics, $\overline{V}(t) \in \mathbb{R}^n$ results from the noise v propagation, and $\epsilon(t) \in \mathbb{R}^n$ is an exponentially decaying term arising due to the initialization of the filters in the extending dynamics. Such an extension can be performed, for instance, via delay operators or linear filters, whose auxiliary role is also to filter the noise; see [13].

One particular choice yielding the extended model (7) is the Kreisselmeier's approach [14] given by: $\forall t \in \mathbb{R}_+$,

$$\begin{aligned}\dot{\Phi}(t) &= -a\Phi(t) + \phi(t)\phi^\top(t),\\ \dot{Y}(t) &= -aY(t) + \phi(t)y(t),\end{aligned} \quad (8)$$

where $\Phi(0) = \Phi_0 \geq 0$, $Y(0) = Y_0$, and $a > 0$ is a scalar tuning parameter. Notably, the choice $\Phi_0 = 0$, $Y_0 = 0$ yields $\epsilon \equiv 0$ in (7). The work [15] shows that (8) preserves the persistent/interval excitation providing the quantitative evaluation of the excitation level of Φ, and [11] shows that it also preserves the lack of excitation in the sense of Definition 2.

The second step of the DREM procedure, *mixing*, is next applied to derive a set of n scalar equations. Multiplying (7) by the adjugate matrix of Φ, denoted as $\mathrm{adj}(\Phi(t))$, on the left and setting $\mathcal{Y}(t) := \mathrm{adj}(\Phi(t))\,Y(t)$, $\mathcal{V}(t) := \mathrm{adj}(\Phi(t))\,V(t)$, $\forall t \in \mathbb{R}_+$, we get

$$\mathcal{Y}_i(t) = \Delta(t)\theta_i + \mathcal{V}_i(t), \quad (9)$$

where $\mathcal{Y}_i$, θ_i, and $\mathcal{V}_i$ are the ith elements of the vectors $\mathcal{Y}$, θ, and $\mathcal{V}$, respectively, $i \in \overline{1,n}$, and the scalar function $\Delta : \mathbb{R}_+ \to \mathbb{R}$ is the determinant of Φ, $\Delta(t) := \det(\Phi(t))$, $\forall t \in \mathbb{R}_+$. The set of n scalar LRE (9) sharing the same bounded scalar regressor Δ is the main result of the DREM procedure. It is worth noting that for a bounded regressor ϕ, the vector $\mathcal{V}$ is also bounded, and $v \equiv 0$ implies $\mathcal{V} \equiv 0$.

III. Problem statement

As discussed in Section II-A, if the excitation is partial of the degree q, then only a reduced-order q-dimensional problem (6) can be solved. As it was explained in the introduction, it also implies that the LRE is not identifiable for the whole θ. Therefore, it is difficult to hope in getting better with a reliable estimation of a projection of θ_1 by using any tool.

At the same time, a feeble excitation theoretically allows the vector θ to be identified, but such degeneracy of the regressor makes the estimation more sensitive to noise. A conventional approach in such a setting is to introduce a regularization in the estimation algorithm, which is our main goal in the context of DREM in this work.

IV. Regularization of the DREM procedure

Consider the DREM procedure for the LRE (1), where a dynamic extension, *e.g.*, the Kreisselmeier's approach (8), yields the extended LRE (7). If the original regressor ϕ is partially (feeble) excited, then according to quantitative estimates from [11], [15], the extended regressor Φ is singular (ill-conditioned). To this end, we intend to revise the mixing step of DREM, introducing a regularization.

Let $H(t) \in \mathbb{R}^{n\times n}$ be a time-varying regularization matrix and define $\forall t \in \mathbb{R}_+$

$$\widetilde{\Phi}(t) := \Phi(t) + H(t),\ \widetilde{\Delta}(t) := \det\Big(\widetilde{\Phi}(t)\Big).$$

Multiplying (7) by $\mathrm{adj}\Big(\tilde{\Phi}(t)\Big)$ on the left and defining

$$\widetilde{\mathcal{Y}}(t) := \mathrm{adj}\Big(\widetilde{\Phi}(t)\Big)\,Y(t),\ \widetilde{\mathcal{V}}(t) := \mathrm{adj}\Big(\widetilde{\Phi}(t)\Big)\,V(t),$$

for all $t \in \mathbb{R}_+$, yields

$$\widetilde{\mathcal{Y}}(t) = \widetilde{\Delta}(t)\theta + \widetilde{\mathcal{V}}(t) - \mathrm{adj}\Big(\widetilde{\Phi}(t)\Big)\,H(t)\theta.$$

Suppose that the matrix $H(t)$ is such that $\widetilde{\Phi}(t)$ is invertible and $\widetilde{\Delta}(t)$ is strictly separated from zero for all $t \geq t_0 \in \mathbb{R}_+$, where t_0 defines the interval of the initial information accumulation.

As discussed in [16], an estimate of θ can be obtained via a pointwise (algebraic) estimator

$$\hat{\theta}(t) = \Big(\widetilde{\Delta}(t)\Big)^{-1}\widetilde{\mathcal{Y}}(t), \forall t \geq t_0, \quad (10)$$

followed by a (probably nonlinear) filter. The paper [16] also shows that such a scheme operates as well (or better) as estimators with asymptotic convergence, e.g., the gradient and least-squares ones. Then the parameter estimation error is given by, $\forall t \geq t_0$,

$$\hat{\theta}(t) - \theta = \Big(\widetilde{\Delta}(t)\Big)^{-1}\widetilde{\mathcal{V}}(t) - \tilde{d}(t). \quad (11)$$

where $\tilde{d}(t) := \Big(\widetilde{\Phi}(t)\Big)^{-1} H(t)\theta$ is the regularization error. The first term on the right-hand side corresponds to the measurement noise, and the second term, $\tilde{d}$, is the distortion introduced by the regularization.

Let us evaluate the properties of $\tilde{d}$ for different scenarios according to the invertibility of Φ and H. For completeness, the case of partial excitation (with a singular matrix Φ) is also investigated. These formulas[1] will also be helpful later in implementing the regularization and derivation of the matrix H.

A. H is invertible

First, suppose that Φ is invertible, e.g., in the case of feeble excitation, and that H is also invertible (non-singular regularization). It holds $\widetilde{\Phi}^{-1} = \Phi^{-1} - \left(\Phi + \Phi H^{-1}\Phi\right)^{-1}$, and the following expression can be derived:

$$\tilde{d} = \widetilde{\Phi}^{-1} H\theta = \left(I + H^{-1}\Phi\right)^{-1}\theta. \quad (12)$$

This equation shows that as the regularization matrix H goes to zero, the regularization-induced distortion $\tilde{d}$ also goes to zero. However, decreasing H also reduces the lower bound $\widetilde{\Delta}$, making the estimation more sensitive to the noise due to the $\widetilde{\Delta}^{-1}\widetilde{\mathcal{V}}$ term in (11). Thus, a choice of the regularization matrix H is subject to the trade-off between noise sensitivity and distortion, similar to the classic bias-variance trade-off.

Next, suppose the matrix Φ is singular, e.g., partially excited. To get $\widetilde{\Delta}$ separated from zero, we introduce an invertible regularization matrix H. Then it holds $\widetilde{\Phi}^{-1} = \left(I + H^{-1}\Phi\right)^{-1} H^{-1}$ and $\tilde{d}$ obeys the same equation (12) as above.

Therefore, as we can conclude from these two cases, if H is invertible, we have the same shape of $\tilde{d}$ independently of the invertibility of Φ.

B. Invertible Φ and singular H

As the matrix H yields a perturbation, reducing it can be of interest. Suppose that Φ is invertible, e.g., in the case of feeble excitation, and that H is singular, i.e., the regularization is introduced in some specific directions only, then similarly to the previous case, we obtain

$$\tilde{d} = \widetilde{\Phi}^{-1} H\theta = \left(I + \Phi^{-1}H\right)^{-1}\Phi^{-1}H\theta.$$

By analogy with the case of an invertible matrix H, a smaller matrix H implies a smaller distortion and a higher noise sensitivity. Again this obvious trade-off is observed.

For a more constructive result, consider a block structure of Φ. For feeble excitation, the first p eigenvalues are small, then we can assume that

$$\Phi = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix},$$

where $A \in \mathbb{R}^{p\times p}$ (or with other dimensions) should be improved by the regularization, and from $\Phi > 0$ we conclude that its Schur complement $S = D - B^\top A^{-1}B$ is also a positive definite matrix. Then using the block matrix inversion formula, we obtain:

$$\Phi^{-1} = \begin{bmatrix} J & -A^{-1}BS^{-1} \\ -S^{-1}B^\top A^{-1} & S^{-1} \end{bmatrix},$$
$$J = A^{-1} + A^{-1}BS^{-1}B^\top A^{-1}.$$

[1]These derivations use various formulas for the inverse of a sum of matrices, where the details are omitted due to the lack of space; see [17].

For $\widetilde{\Delta}$ to be separated from zero, we need $\Phi + H$ with sufficiently big eigenvalues. One possible choice of H is

$$H = \begin{bmatrix} H_1 & 0 \\ 0 & 0 \end{bmatrix}, \; H_1 > 0, \; H_1 \in \mathbb{R}^{p\times p}$$

Then $\widetilde{\Phi} = \begin{bmatrix} A + H_1 & B \\ B^\top & D \end{bmatrix}$ is no more feebly excited for a properly selected H_1. Therefore, we have

$$\Phi^{-1}H = \begin{bmatrix} JH_1 & 0 \\ -S^{-1}B^\top A^{-1}H_1 & 0 \end{bmatrix},$$

and

$$\tilde{d} = \begin{bmatrix} \Sigma J H_1 & 0 \\ S^{-1}B^\top A^{-1}H_1(\Sigma J H_1 - I) & 0 \end{bmatrix}\theta,$$

where $\Sigma = (I + JH_1)^{-1}$.

Consequently, the components of the parameter vector θ, corresponding to the part of Φ that is sufficiently excited (the matrix D), do not introduce a perturbation in the regularization error $\tilde{d}$, and their estimates are not much influenced provided that the vector $\begin{bmatrix} S^{-1}B^\top A^{-1}H_1(\Sigma J H_1 - I) & 0 \end{bmatrix}\theta$ is sufficiently small.

C. Φ and H are simultaneously diagonalizable

Let us consider the special case when the regularization matrix H is simultaneously diagonalizable with Φ. Such a matrix H may be time-varying and computed online based on the spectral decomposition of Φ, which is symmetric by design. The spectral decomposition of Φ for all $t \geq t_0$ is given by

$$\Phi(t) = Q^\top(t)\Lambda_\Phi(t)Q(t), \quad (13)$$

where Q is a time-varying orthogonal matrix of orthonormal eigenvectors vectors (cf. with constant v_i in Definition 2 or matrix C in Definition 3), and $\Lambda_\Phi(t)$ is a diagonal matrix of eigenvalues of $\Phi(t)$. Let the matrix H be symmetric and have the same eigenspace, i.e.,

$$H(t) = Q^\top(t)\Lambda_H(t)Q(t), \quad (14)$$

where $\Lambda_H(t)$ is a diagonal matrix of eigenvalues of $H(t)$. Then

$$\widetilde{\Phi}(t) = Q^\top(t)\left(\Lambda_\Phi(t) + \Lambda_H(t)\right)Q(t).$$

Suppose that $\Lambda_H(t)$ is chosen such that $\widetilde{\Phi}(t)$ is invertible, $\forall t \geq t_0$. Then

$$\tilde{d}(t) = Q^\top(t)\left(\Lambda_\Phi(t) + \Lambda_H(t)\right)^{-1}\Lambda_H(t)Q(t)\theta.$$

Recalling Definition 4, $0 \leq \lambda_1(t) \leq \ldots \leq \lambda_n(t)$ are the eigenvalues of $\Phi(t)$, i.e., the diagonal elements of $\Lambda_\Phi(t)$; suppose for simplicity that $\lambda_i(t)$ is the i-th diagonal element, $i = 1, \ldots, n$. Let the diagonal elements of $\Lambda_H(t)$ be $\sigma_i(t)$. Then

$$\tilde{d}(t) = G(t)\theta, \; G(t) = Q^\top(t)\Lambda_G(t)\, Q(t), \quad (15)$$

where

$$\Lambda_G(t) = \operatorname{diag}\left(\frac{\sigma_i(t)}{\lambda_i(t) + \sigma_i(t)}\right)_{i=1}^{n},$$

and all eigenvalues of $G(t)$ are not greater than 1.

Suppose that Φ is partially (feeble) exciting of degree q (of order p), i.e., the first $p = n - q$ eigenvalues of Φ are zeros (close to zero): $\lambda_i(t) = 0$ ($\lambda_i(t) \leq \mu_1$) for all $t \geq t_0$ for $i = 1, \ldots, p$. To ensure that the DREM-generated regressor $\widetilde{\Delta}(t)$ is strictly separated from zero, it is reasonable to choose $\sigma_i(t) = \sigma_0 > 0$ for $i = 1, \ldots, p$, where σ_0 is a tuning parameter. Setting $\sigma_i = 0$ for $i = p+1, \ldots, n$, then

$$\Lambda_H(t) = \begin{bmatrix} \sigma_0 I & 0 \\ 0 & 0 \end{bmatrix}, \; \Lambda_G(t) = \begin{bmatrix} E(t) & 0 \\ 0 & 0 \end{bmatrix},$$

where $E(t) = \operatorname{diag}\left(\frac{\sigma_0}{\lambda_i(t)+\sigma_0}\right)_{i=1}^{p}$.

The eigenvectors of $\Phi(t)$ associated with zero (small) eigenvalues are the first p columns of the matrix $Q(t)$; denote these columns by $Q_0(t) \in \mathbb{R}^{n\times p}$ and the rest columns as $Q_1(t) \in \mathbb{R}^{n\times q}$, then $Q(t) = \begin{bmatrix} Q_0(t) & Q_1(t) \end{bmatrix} = \begin{bmatrix} Q_{00}(t) & Q_{01}(t) \\ Q_{01}^\top(t) & Q_{11}(t) \end{bmatrix}$, where $Q_{00}(t) \in \mathbb{R}^{p\times p}$, $Q_{11}(t) \in \mathbb{R}^{q\times q}$ and $Q_{01}(t) \in \mathbb{R}^{p\times q}$ are respective sub-blocks. At the same time, the columns of $Q_1(t)$ are the eigenvectors corresponding zero eigenvalues, $\sigma_i = 0$, of the matrix $G(t)$ defined in (15).

In the partial excitation case, the columns of $Q_0(t)$ span the same subspace of $\mathbb{R}^n$ as the columns of C defined in Definition 3, corresponding to the kernel of $\Phi(t)$. Moreover, if Z is orthogonal to C, as discussed in Remark 1, then $Q_1(t)$ spans the same subspace as Z. Thus, Z is orthogonal to $G(t)$ for all $t \geq t_0$, and

$$Z^\top \tilde{d}(t) = Z^\top G(t)\theta = 0. \tag{16}$$

Equation (16) implies that for the specific choice of the regularization matrix $H(t)$ as (14) with $\sigma_i = 0$ for $i = p+1, \ldots, n$, the distortion $\tilde{d}(t)$ is orthogonal to Z. Thus, in the absence of noise, the projection of the estimate (10) under this regularization on the subspace orthogonal to C, i.e., where the regressor ϕ is exciting, is not affected by the distortion. Let $\hat{\theta}_{ss}$ be the steady-state value of $\hat{\theta}(t)$ in the absent of noise. Then

$$\hat{\theta}_{ss} = C\left(\tilde{\theta}_1 + \tilde{d}_1\right) + Z\tilde{\theta}_2,$$

where $\tilde{d}_1$ it the projection of $\tilde{d}$ on C. Thus this representation can be used as an alternative to the reduced-order LRE (6).

In the case of feeble excitation,

$$\tilde{d}(t) = \begin{bmatrix} Q_{00}(t)E(t)Q_{00}(t) & Q_{00}(t)E(t)Q_{01}(t) \\ Q_{01}(t)E(t)Q_{00}(t) & Q_{01}(t)E(t)Q_{01}(t) \end{bmatrix} \theta$$

which has a generic structure. In general case the regularization always perturb the estimation accuracy and optimization of its value in the presence of noise is desirable.

D. Trade-off estimation

Let us provide an estimate on H providing the desired trade-off between the noise filtering and the bias error.

Define the estimation error for the algorithm (10):

$$\begin{aligned} e_H(t) = \hat{\theta}(t) - \theta &= \tilde{\Phi}^{-1}(t)Y(t) - \theta \\ &= \tilde{\Phi}^{-1}(t)\left(V(t) - H(t)\theta\right), \end{aligned} \tag{17}$$

then our goal is to prove that under certain restrictions on the matrix $H(t)$, and for a sufficiently big amplitude of the noise $V(t)$ in comparison with the norm of θ, the estimation accuracy after regularization is better, i.e.,

$$\|e_H(t)\| \leq \|e_0(t)\|, \; \forall t \geq t_0, \tag{18}$$

where e_0 corresponds to e_H for $H = 0$, i.e., to the absence of regularization. The necessity of introduction of a relation between $V(t)$ and θ is intuitively clear from the expression of e_H: the last term $V(t) - H(t)\theta$ contains all variables of interest, which is multiplied on a common gain $\tilde{\Phi}^{-1}(t)$, and obviously, if $H(t)$ is very big, the bias error $H(t)\theta$ starts to dominate and corrupt the estimation error, as has already been discussed.

Theorem 1: Assume that $\|\theta\| \leq \theta_{\max}$ for a given upper bound $\theta_{\max} > 0$, the matrix $\tilde{\Phi}(t)$ is nonsingular for all $t \geq t_0$, and there exists a constant $\beta > 0$ such that

$$\|\tilde{\Phi}^{-1}(t)V(t)\| \geq \|\tilde{\Phi}^{-1}(t)\|\beta$$

for all $t \geq t_0$. Then for

$$\frac{\beta}{\sqrt{2}\theta_{\max}} \geq \|H\|, \; \Phi^{-1}H + H\Phi^{-1} + H\Phi^{-2}H \geq 2I_n$$

the desired relation (18) is satisfied.[2]

The existence of β implies that the noise $V(t)$ is sufficiently rich and not zero, providing an excitation of the signal $\tilde{\Phi}^{-1}(t)V(t)$. If the noise is zero, it is straightforward that regularization is the only source of inaccuracy in the estimation.

Next, we illustrate these results with an academic example.

V. Example

For an illustrative example, we consider the LRE (1) with $n = 3$, $\ell = 1$, and

$$\phi(t) = \begin{bmatrix} 1 & \sin(t) & \sin(t+s) \end{bmatrix} \tag{19}$$

with constant $s \in [0, \; \frac{\pi}{2}]$. Choosing $T = 2\pi$, we obtain

$$\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau = \begin{bmatrix} 2\pi & 0 & 0 \\ 0 & \pi & \pi\cos(s) \\ 0 & \pi\cos(s) & \pi \end{bmatrix},$$

where the eigenvalues of this matrix are $\lambda_1 = \pi\left(1 - \cos(s)\right)$, $\lambda_2 = \pi\left(1 + \cos(s)\right)$, and $\lambda_3 = 2\pi$. The excitation properties of ϕ are summarized below.

- For any $s > 0$, the regressor (19) is $(\mu, 2\pi)$-PE with $\mu = 2\pi^3\left(1 - \cos^2(s)\right)$, see Definition 1.
- A value of s close to zero makes the excitation feeble of order 1 in the sense of Definition 4, i.e., the eigenvalue λ_1 can be made arbitrary close to zero making the ratio $\frac{\lambda_2}{\lambda_1}$ arbitrary large.
- For $s = 0$, the regressor is partially persistently exciting of degree 2 in the sense of Definition 3. This can be seen choosing $C = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}^\top$, $Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}^\top$ and verifying that $\phi^\top(t)C = 0$ for all $t \in \mathbb{R}_+$ and

[2] The proof of Theorem 1 is omitted due to the lack of space.

$\tilde{\phi}(t) = Z^\top \phi(t)$ is PE. Note also that Z is orthogonal to C as discussed in Remark 1.

Choose the vector of unknown parameters as $\theta = \begin{bmatrix}1 & -3 & 2\end{bmatrix}^\top$ and let $v(t)$ be a uniform bounded noise, $|v(t)| \leq \bar{v}$, where $\bar{v} = 0.3$.

For the extension step of the DREM procedure, we apply (8), where $a = \frac{1}{2}$, $\Phi_0 = 0$ and $Y_0 = 0$. It yields the extended LRE (7), where the regularization and mixing can be performed as discussed in Section IV.

Persistently and Partially excited cases. For $s = \frac{\pi}{2}$, both eigenvalue λ_1 and λ_2 equal to π, and all parameters can be estimated without a regularization. The estimation accuracy thus depends only on the noise filtering.

For $s = 0$, the smallest eigenvalue is zero, and the vector θ cannot be reconstructed. Then, two options are available. The first is to compute the new regressor $\tilde{\phi} = Z^\top \phi$ as discussed in Section II-A, and perform estimation based on the reduced-order LRE (6). Note that if the matrix Z is not known *a priori*, it can be found based on the spectral decomposition of Φ, as discussed in Section IV-C. It is also interesting to see that due to the special structure of the matrix Z, the first element of $\tilde{\theta}_2$ coincides with the first element of θ, allowing for its estimation.

Another option is to apply the regularization (13). Then, due to (15) and the form of the matrix Z, the first element of $\tilde{d}$ is zero allowing for reconstruction of the first element of θ.

Feeble excitation. Choose $s = 0.1$ making λ_1 small and the matrix $\int_t^{t+T} \phi(\tau)\phi^\top(\tau)d\tau$ poorly conditioned. Then, to improve the estimation accuracy, we introduce the constant regularization matrix $H(\sigma) = \sigma\,\mathrm{diag}(0, I_2)$, and $\sigma > 0$. Such a structure is motivated by the results presented in Section IV-B.

We are interested to see how the value of σ impacts the estimation accuracy. Towards this end, for a given value of σ we compute $e_{H(\sigma)}(t)$ defined in (17) and find the mean value of $|e_{H(\sigma)}(t)|^2$ over the simulation horizon after the transients (approximately $3 \cdot 10^5$ samples). These values, computed for σ varying from 0 (no regularization) to 0.01 are given in Figure 1. For no regularization, the noise impacts the accuracy; as σ increases, the noise sensitivity is improved, but for high σ the regularization-induced bias becomes dominating.

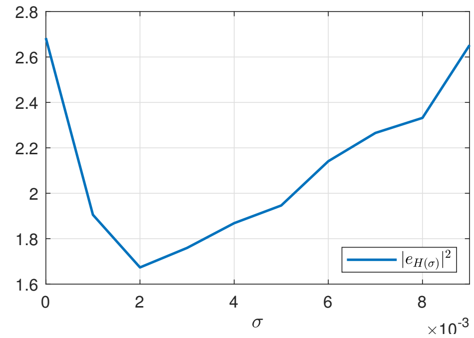

Fig. 1. Mean value of $|e_{H(\sigma)}(t)|^2$ as a function of σ.

VI. Conclusions

The LRE problem with measurement noise was investigated. Two concepts of partial and feeble excitation were introduced, which determine the conditions of a lack of persistent or interval excitation or that the excitation is insufficient for a reliable estimation with perturbations. Regularization was introduced for DREM, and the conditions for improving accuracy were evaluated. An extension to the case of time-varying parameters is left for future research.

References

[1] D. Efimov and A. Fradkov, "Design of impulsive adaptive observers for improvement of persistency of excitation," *International Journal of Adaptive Control and Signal Processing*, vol. 29, no. 6, pp. 765–782, jul 2014.

[2] D. Efimov, N. Barabanov, and R. Ortega, "Robustness of linear time-varying systems with relaxed excitation," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1885–1900, apr 2019.

[3] Y. Pan and H. Yu, "Composite learning robot control with guaranteed parameter convergence," *Automatica*, vol. 89, pp. 398–406, mar 2018.

[4] J. Wang, D. Efimov, S. Aranovskiy, and A. A. Bobtsov, "Fixed-time estimation of parameters for non-persistent excitation," *European Journal of Control*, vol. 55, pp. 24–32, sep 2020.

[5] L. Wang, R. Ortega, and A. Bobtsov, "Observability is sufficient for the design of globally exponentially stable state observers for state-affine nonlinear systems," *Automatica*, vol. 149, p. 110838, mar 2023.

[6] R. Kamalapurkar, B. Reish, G. Chowdhary, and W. E. Dixon, "Concurrent learning for parameter estimation using dynamic state-derivative estimators," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3594–3601, jul 2017.

[7] 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, apr 2019.

[8] R. Ortega, S. Aranovskiy, A. A. Pyrkin, A. Astolfi, and A. A. Bobtsov, "New results on parameter estimation via dynamic regressor extension and mixing: Continuous and discrete-time cases," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2265–2272, may 2021.

[9] A. Zolghadri, J. Cieslak, P. Goupil, and R. Dayre, "Turning theory to practice in model-based FDI: Successful application to new generation airbus aircraft," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 12 773–12 778, jul 2017.

[10] 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*, vol. 40, no. 4, pp. 80–102, aug 2020.

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

[12] A. Glushchenko and K. Lastochkin, "Relaxation of conditions for convergence of dynamic regressor extension and mixing procedure," *Automation and Remote Control*, vol. 84, pp. 16–47, 2023.

[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, jul 2017.

[14] G. Kreisselmeier, "Adaptive observers with exponential rate of convergence," *IEEE Transactions on Automatic Control*, vol. 22, no. 1, pp. 2–8, feb 1977.

[15] S. Aranovskiy, R. Ushirobira, M. Korotina, and A. Vedyakov, "On preserving-excitation properties of kreisselmeier's regressor extension scheme," *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 1296–1302, feb 2023.

[16] M. Korotina, S. Aranovskiy, R. Ushirobira, D. Efimov, and J. Wang, "Fixed-time parameter estimation via the discrete-time drem method," in *Proc. 22st IFAC World Congress*, Yokohama, Japon, 2023.

[17] R. Horn and C. Johnson, *Matrix Analysis.* Cambridge University Press, 2012.